

# **THESE**

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

par

Laetitia FOUILLEN

## **Universalité des méthodologies protéomiques pour l'identification et la caractérisation des protéines**

Soutenue le 25 juin 2009 devant la commission d'examen :

Dr. Alain VAN DORSSELAER

Directeur de thèse

Prof. Sandrine BOSCHI-MULLER

Rapporteur

Dr. Nicolas SOMMERER

Rapporteur

Dr. Hélène ROGNIAUX

Examineur

Prof. Michel BURGARD

Examineur interne







“Le commencement de toutes les sciences,  
c'est l'étonnement de ce que les choses sont ce qu'elles sont.”

Aristote

“Les grands accomplissements sont réussis non par la force,  
mais par la persévérance.”

Samuel Johnson



## Merci !

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

Je souhaite en premier lieu adresser ma profonde reconnaissance à Alain Van Dorsselaer pour m'avoir accueillie dans son laboratoire. Je tiens à remercier chaleureusement Sarah Sanglier pour m'avoir encadrée, soutenue et encouragée tout au long de cette thèse et Fabrice Bertile pour son encadrement et pour m'avoir appris à différencier les manchots des pingouins.

J'adresse mes remerciements à M. Michel Burgard qui a accepté de présider mon jury de thèse ainsi qu'à Mmes Sandrine Boschi-Muller et Hélène Rogniaux ainsi qu'à M. Nicolas Sommerer, qui ont consacré de leur temps à l'évaluation de ce travail.

J'adresse mes remerciements à la région Alsace et à la société AliX pour le financement de cette thèse.

Je remercie également l'ensemble des personnes avec lesquelles j'ai été amenée à collaborer et qui m'ont entraînée dans diverses thématiques dans le désordre : C Branlant, V. Marchand et M.Santerre, S. Maenner, A.Vautrin et I. Behm-Ansmat, A. Bar, J-M. Saliou, A. Abdulrahman et A. Poterzsmann, D. Quenet et F. Dantzer, M. Zimmermann et J. Haiech, T. Raclot...

Mes remerciements s'adressent également aux membres du laboratoire que j'ai pu découvrir et rencontrer au cours de ces années : François, Véronique D., Christine S. et Cyril pour votre disponibilité et vos conseils, Véronique T., guide dans le monde de l'administration, Jean-Michel, pour tes cours sur le monde des ARNs, Chrystel, merci de m'avoir accompagnée sur la découverte des manchots, le bureau des pouilleux : Thierry, Nicolas, Daniel, non je ne t'ai pas toujours martyrisé!, Cédric, pour m'avoir permis de mieux comprendre le monde du non-covalent, Audrey et Sébastien (grand chef pouilleux!), pour les pauses et les réflexions scientifiques ou pas, Danièle, pour m'avoir emmenée et soutenue pendant la gym, Dimitri, pour la découverte de phosphorylation, Christine C. et Laurent, pour m'avoir intégrée au labo et soutenue, Guillaume, pour tes sujets de discussions divers (de la masse à la politique!), Jean- Marc, pour la découverte du monde merveilleux des grues et ton soutien technique de tous les instants, Fabrice V., pour soigner les maux de mon ordi "ma petite truite" et m'avoir fait beaucoup rire ! Flavie, pour tes discussions et la découverte de la Suisse, Hélène et Jennifer, pour avoir cru en moi et m'avoir poussée quand il le fallait, Agnès, pour ton amitié et ton soutien quotidien mais aussi pour tout le reste!!

Enfin, je tiens à remercier ma famille, mes amis et mes amours qui ont jalonné et enrichi ma vie jusqu'à aujourd'hui....





## PLAN GENERAL

<b>INTRODUCTION GÉNÉRALE</b> .....	<b>9</b>
<b>INTRODUCTION BIBLIOGRAPHIQUE</b> .....	<b>13</b>
CHAPITRE I – LES STRATEGIES GENERALES EN PROTEOMIQUE.....	17
CHAPITRE II – LES OUTILS UTILISES EN PROTEOMIQUE .....	33
CONCLUSION .....	51
REFERENCES BIBLIOGRAPHIQUES .....	53
<b>PARTIE I – IDENTIFICATION DES PROTEINES IMPLIQUEES DANS DES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE</b> .....	<b>61</b>
CHAPITRE I. SYNTHÈSE BIBLIOGRAPHIQUE ET STRATEGIE DEVELOPPEE POUR L'ANALYSE DES PROTEINES IMPLIQUEES DANS LES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE .....	63
CHAPITRE II. IDENTIFICATION DE PARTENAIRES PROTEIQUES DE LA POLY(ADP-RIBOSE) POLYMERASE-2 ..	71
CHAPITRE III. IDENTIFICATION DE PARTENAIRES PROTEIQUES DE COMPLEXES ARN/PROTEINE .....	77
CONCLUSION GENERALE DE LA PARTIE "IDENTIFICATION DES PROTEINES IMPLIQUEES DANS DES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE" .....	101
RÉFÉRENCES BIBLIOGRAPHIQUES .....	103
<b>PARTIE II – CARACTERISATION FINE DE PROTEINES PURIFIEES PHOSPHORYLEES</b> .....	<b>107</b>
CHAPITRE I - INTRODUCTION SUR L'ANALYSE DES PHOSPHORYLATIONS PAR SPECTROMETRIE DE MASSE ET STRATEGIE DEVELOPPEE. ....	109
CHAPITRE II - ÉTUDE DES PHOSPHORYLATIONS DU DOMAINE KINASE DE LA PROTEINE DAPK1 .....	123
CHAPITRE III - ÉTUDE DES PHOSPHORYLATIONS DU COMPLEXE CAK .....	135
CONCLUSION GENERALE DE LA PARTIE "CARACTERISATION FINE DE PROTEINES PURIFIEES PHOSPHORYLEES" .....	159
RÉFÉRENCES BIBLIOGRAPHIQUES .....	161
<b>PARTIE III- APPORTS DE L'APPROCHE PROTEOMIQUE POUR L'IDENTIFICATION DES PROTEINES PLASMATIQUES DIFFERENTIELLEMENT EXPRIMEES LORS DU JEUNE PROLONGE CHEZ LE MANCHOT ADELIE, UN ORGANISME DONT LE GENOME N'EST PAS SEQUENCE</b> .....	<b>167</b>
CHAPITRE I : SYNTHÈSE BIBLIOGRAPHIQUE : STRATEGIES ET METHODES PROTEOMIQUE .....	169
CHAPITRE II : APPLICATION : ANALYSE DIFFERENTIELLE DU PROTEOME PLASMATIQUE DU MANCHOT ADELIE EN REPOSE AU JEUNE PROLONGE.....	179
CHAPITRE III : CONCLUSION : APPROCHE PROTEOMIQUE ET BIOLOGIE DES ORGANISMES DONT LE GENOME N'EST PAS CONNU .....	199
RÉFÉRENCES BIBLIOGRAPHIQUES .....	203
<b>CONCLUSION GENERALE</b> .....	<b>213</b>
<b>PARTIE EXPERIMENTALE</b> .....	<b>219</b>
<b>ANNEXES</b> .....	<b>231</b>



## PLAN DETAILLE

<b>INTRODUCTION GÉNÉRALE .....</b>	<b>9</b>
<b>INTRODUCTION BIBLIOGRAPHIQUE .....</b>	<b>13</b>
<b>CHAPITRE I – LES STRATEGIES GENERALES EN PROTEOMIQUE .....</b>	<b>17</b>
<b>1. Les stratégies classiquement utilisées .....</b>	<b>17</b>
1.a. L’empreinte peptidique massique .....	17
1.b. La stratégie nanoLC-MS/MS .....	19
1.b.1). Les règles de fragmentation.....	21
1.c.2). L’identification des protéines par peptide Fragment FingerPrinting (PFF).....	22
<b>2. Des alternatives aux approches protéomiques classiques .....</b>	<b>23</b>
2.a. La chromatographie multidimensionnelle .....	23
2.b. L’approche <i>de novo</i> .....	26
2.b.1) Le séquençage de novo.....	26
2.b.2) La recherche par homologie de séquence .....	27
2.c L’approche "error-tolerant" .....	28
<b>3. La bioinformatique est indissociable de l’analyse protéomique .....</b>	<b>28</b>
3.a. Les banques de données protéiques.....	28
3.b. L’évaluation statistique des données.....	30
3.c. Automatisation des identifications.....	30
<b>CHAPITRE II – LES OUTILS UTILISES EN PROTEOMIQUE .....</b>	<b>33</b>
<b>1. Les techniques de séparation.....</b>	<b>33</b>
1.a. L’électrophorèse.....	33
1.a.1) Le gel d’électrophorèse 2D .....	33
1.a.2) Le gel d’électrophorèse 1D .....	36
1.a.3) Le Off-Gel .....	36
1.b. La chromatographie liquide .....	37
1.b.1) Le système Chip .....	37
1.b.2) La chromatographie UHPLC .....	38
<b>2. La spectrométrie de masse .....</b>	<b>38</b>
2.a. Les sources d’ionisation .....	39
2.a.1). L’ionisation Electrospray .....	39
2.a.2). L’ionisation MALDI .....	40
2.b. Les analyseurs .....	41
2.b.1) Généralités.....	42
2.b.2). Le quadripôle .....	43
2.b.3) L’analyseur à temps de Vol (TOF) .....	43
2.b.4) L’analyseur Trappe ionique.....	45
2.b.5). Des analyseurs en tandem : l’exemple du Q-TOF .....	47
<b>3. Les modes de fragmentation .....</b>	<b>48</b>
3.a. La méthode de fragmentation CID.....	48
3.b. La méthode de fragmentation ECD/ETD .....	49
<b>CONCLUSION .....</b>	<b>51</b>
<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>53</b>

## **PARTIE I – IDENTIFICATION DES PROTEINES IMPLIQUEES DANS DES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE..... 61**

<b>CHAPITRE I. SYNTHÈSE BIBLIOGRAPHIQUE ET STRATEGIE DEVELOPPEE POUR L’ANALYSE DES PROTEINES IMPLIQUEES DANS LES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE .....</b>	<b>63</b>
<b>1. Synthèse bibliographique .....</b>	<b>63</b>
1.a Purification des complexes protéiques .....	63
1.a.1) Les complexes protéine/protéine .....	63
1.a.2) Les complexes ARN/protéine.....	64
1.b. Analyse par spectrométrie de masse des complexes purifiés .....	66
<b>2. Stratégie développée pour l’identification des protéines .....</b>	<b>68</b>
2.a. Purification du complexe et séparation des protéines. ....	69
2.b. Analyse par spectrométrie de masse .....	69

CHAPITRE II. IDENTIFICATION DE PARTENAIRES PROTEIQUES DE LA POLY(ADP-RIBOSE) POLYMERASE-2 ..	71
1. Contexte biologique .....	71
2. Mise en œuvre de l'approche protéomique développée .....	72
3. Résultats .....	73
4. Conclusion .....	76
CHAPITRE III. IDENTIFICATION DE PARTENAIRES PROTEIQUES DE COMPLEXES ARN/PROTEINE .....	77
1. Choix du couplage utilisé pour les études ARN/Protéine .....	77
2. Etude des partenaires protéiques d'interaction du site A7 de l'ARN de VIH-1 .....	79
2.a. Contexte biologique .....	79
2.b. Application de la stratégie à la recherche de partenaires protéiques d'interaction de l'ARN SIs2 contenant le site A7 .....	81
2.c. Résultats .....	83
2.d. Conclusion .....	86
3. Etudes de partenaires de l'ARN Xist impliqué dans l'inactivation. ....	89
3.a. Contexte biologique .....	89
3.b. Etude de l'influence du nombre de "A-repeats" sur les protéines recrutées .....	92
3.c. Etude de la composition du complexe selon le degré de différenciation .....	94
3.d. Conclusion .....	99
CONCLUSION GENERALE DE LA PARTIE "IDENTIFICATION DES PROTEINES IMPLIQUEES DANS DES COMPLEXES PROTEINE/PROTEINE ET ARN/PROTEINE" .....	101
RÉFÉRENCES BIBLIOGRAPHIQUE .....	103

## **PARTIE II – CARACTERISATION FINE DE PROTEINES PURIFIEES PHOSPHORYLEES..... 107**

CHAPITRE I - INTRODUCTION SUR L'ANALYSE DES PHOSPHORYLATIONS PAR SPECTROMETRIE DE MASSE ET STRATEGIE DEVELOPPEE .....	109
1. Les difficultés liées à l'analyse des phosphorylations par spectrométrie de masse .....	109
1.a. Intérêts biologiques des phosphorylations .....	109
1.b. Généralités sur les phosphorylations .....	110
1.c. Enrichissement des phosphopeptides .....	110
1.d. Identification des phosphopeptides. ....	111
1.e. Localisation des sites de phosphorylations .....	113
1.e.1) Les approches par fragmentation CID .....	114
1.e.2) Les approches par fragmentation ECD/ETD .....	117
1.f. Conclusion .....	117
2. Stratégie développée pour l'identification de phosphorylations de protéines purifiées.....	118
2.a. Détermination de l'état de phosphorylation sur les protéines entières.....	119
2.b. Localisation des phosphorylations sur les peptides .....	121
CHAPITRE II - ETUDE DES PHOSPHORYLATIONS DU DOMAINE KINASE DE LA PROTEINE DAPK1 .....	123
1. Contexte biologique .....	123
2. Etude des protéines entières .....	124
3. Phosphorylation sur DAPK1 WT .....	126
4. Phosphorylation sur Delta-DAPK1 .....	130
5. Conclusion .....	132
CHAPITRE III - ETUDE DES PHOSPHORYLATIONS DU COMPLEXE CAK .....	135
1. Contexte Biologique .....	135
2. Mesure de masse et déphosphorylation des protéines entières.....	135
3. Comparaison du taux de phosphorylation des protéines selon la lignée cellulaire utilisée pour la production du CAK et après l'autophosphorylation du complexe .....	139
3.a. Production du complexe avec une autre lignée cellulaire .....	139
3.b. Autophosphorylation du complexe.....	140
4. Localisation des sites de phosphorylations du complexe CAK.....	142
4.a. Phosphorylation de CDK7 .....	142
4.b. Phosphorylation de la Cycline H.....	147
4.c. Phosphorylation de la protéine MAT1 .....	149
4.d. Les sites de phosphorylations après autophosphorylation du complexe .....	152
4.e. Les fragmentations des peptides phosphorylés.....	154
4.e.1) Comparaison des spectres de fragmentation d'un peptide phosphorylé et non phosphorylé ..	154
4.e.2) Comparaison des spectres de fragmentation d'un peptide phosphorylé obtenus sur un Q-TOF et une trappe ionique .....	155
5. Conclusion et perspectives .....	158
CONCLUSION GENERALE DE LA PARTIE "CARACTERISATION FINE DE PROTEINES PURIFIEES PHOSPHORYLEES" .....	159

RÉFÉRENCES BIBLIOGRAPHIQUE .....	161
<b>PARTIE III- APPORTS DE L'APPROCHE PROTEOMIQUE POUR L'IDENTIFICATION DES PROTEINES PLASMATIQUES DIFFERENTIELLEMENT EXPRIMEES LORS DU JEUNE PROLONGE CHEZ LE MANCHOT ADELIE, UN ORGANISME DONT LE GENOME N'EST PAS SEQUENCE.....</b>	<b>167</b>
CHAPITRE I : SYNTHÈSE BIBLIOGRAPHIQUE : STRATEGIES ET METHODES PROTEOMIQUE .....	169
1. <i>Identification de protéines d'organismes dont le génome n'est pas séquencé : quelle     stratégie ?</i> .....	169
2. <i>Intérêts et limitations des approches protéomiques à partir d'échantillons plasmatiques</i> .....	171
3. <i>La préparation des échantillons, une étape incontournable des approches protéomiques</i> ....	173
4. <i>Analyse protéomique différentielle : notion de quantification relative</i> .....	177
CHAPITRE II : APPLICATION : ANALYSE DIFFERENTIELLE DU PROTEOME PLASMATIQUE DU MANCHOT ADELIE EN REPONSE AU JEUNE PROLONGE.....	179
1. <i>Contexte scientifique général</i> .....	179
2. <i>Le modèle du jeûne prolongé : adaptations métaboliques</i> .....	179
3. <i>Le cas du manchot Adélie</i> .....	181
4. <i>Problématique et objectifs de l'étude</i> .....	182
5. <i>Méthodes et stratégie de notre approche protéomique</i> .....	184
6. <i>Résultats et discussion</i> .....	186
6.a. Paramètres globaux et statut métabolique des animaux .....	186
6.b. Préparation des échantillons et mise au point des conditions de séparations des protéines .....	187
6.c. Succès de l'approche protéomique pour l'identification des protéines plasmatiques de manchot Adélie .....	188
6.d. Identification des protéines différentiellement exprimées entre les manchots P2C et P3C .....	190
6.e. Identification des protéines différentiellement exprimées entre les manchots P2, P3-A et P3+A ....	195
6.f. Comparaison des effets du jeûne prolongé sur le protéome plasmatique de rat et de manchot Adélie .....	197
CHAPITRE III : CONCLUSION : APPROCHE PROTEOMIQUE ET BIOLOGIE DES ORGANISMES DONT LE GENOME N'EST PAS CONNU .....	199
RÉFÉRENCES BIBLIOGRAPHIQUE .....	203
<b>CONCLUSION GENERALE .....</b>	<b>213</b>
<b>PARTIE EXPERIMENTALE.....</b>	<b>219</b>
1. <i>Préparation d'échantillon</i> .....	221
1.a. Production des protéines recombinantes .....	221
1.a.1) Production de DAPK1 WT et Delta-DAPK1 .....	221
1.a.2) Production du complexe CAK .....	221
1.b. Préparation d'échantillon et séparation des protéines par électrophorèse .....	221
1.b.1) Déplétion d'albumine et précipitation des protéines.....	221
1.b.2) Préparation des gels d'électrophorèses .....	222
1.b.3) Découpe et traitement des morceaux de gel.....	222
1.c. Traitements enzymatiques.....	222
1.c.1) Digestion enzymatique .....	222
1.c.2) Traitement pour la déphosphorylation.....	223
2. <i>Analyses par spectrométrie de masse:</i> .....	224
2.a. Analyses MS:.....	224
2.a.1) Analyses LC-MS (MicroLC-MicroTOF).....	224
2.a.2) Analyses MALDI –MS (Ultraflex™ MALDI-TOF/TOF) .....	224
2.b. Analyses MS/MS .....	225
2.b.1) Analyses nanoLC-MS/MS réalisées avec le couplage CapLC-QTOF.....	225
2.b.2) Analyses nanoLC-MS/MS réalisées avec le couplage NanoAcquity-Synapt .....	227
2.b.3) Analyses nanoLC-MS/MS réalisées avec le couplage nanoLC-ChipCube–Trappe ionique (HCT plus ou HCT Ultra) .....	228
<b>ANNEXES .....</b>	<b>231</b>



## Liste des principales abréviations

ACN :	Acetonitrile
ADN :	Acide désoxyribonucléique
ARN :	Acide ribonucléique
BLAST :	Basic Local Alignment Search Tool
BSA :	Bovine Serum Albumin
CID :	Collision Induced Dissociation
Da :	Dalton
DHB	Acide 2,5-dihydrobenzoïque
DTT :	Dithiothreitol
ECD :	Electron Capture Dissociation
EDTA :	EthyleneDiamineteTraacetic Acid
ESI :	Electrospray Ionisation
ETD :	Electron Transfer Dissociation
FT-ICR :	Fourier Transform Ion Cyclotron Resonance
FWHM :	Full Width at Half Maximum
HCCA :	Acide $\alpha$ -cyano-4-hydrocinnamique
HPLC :	High Performance Liquid Chromatography
IEF :	Iso-Electrofocalisation
IT :	Ion Trap
LC :	Liquid Chromatography
MALDI :	Matrix-Assisted Laser Desorption/Ionisation
MS :	Spectrométrie de Masse
MS/MS :	Spectrométrie de masse en tandem
MudPIT :	Multi-Dimensional Protein Identification Technology
Nano-ESI :	Nano-electrospray
NanoLC-MS/MS :	Nano liquid chromatography tandem mass spectrometry
PMF :	Peptide Mass Fingerprinting
Q :	Quadripôle
RP :	Reverse Phase
SCX :	Strong Cation Exchange
SDS-PAGE :	Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis
TCA :	Trifluoroacetic acid
TOF :	Time of Flight





## **Introduction générale**



L'activité biologique, l'adressage et la demi-vie des protéines peuvent varier en fonction de leurs structures (primaire, secondaire ou tertiaire), de leurs niveaux d'expression et/ou de leurs interactions avec d'autres molécules (protéines, ARN...). L'analyse protéomique, qui étudie l'ensemble des protéines exprimées dans un tissu ou une cellule à un moment donné et dans des conditions données, apparaît donc essentielle pour permettre une meilleure compréhension de nombreux processus biologiques.

Cependant la grande diversité des problèmes biologiques abordée ces dernières années par l'analyse protéomique a montré que celle-ci devait être encore développée. Il est clair que les méthodologies protéomiques doivent être développées et adaptées en fonction du type de question posée (Identifications de protéines, caractérisation fine de la structure primaire, quantification du niveau d'expression de protéines...) mais aussi en fonction des contraintes techniques imposées par les échantillons (complexité, stabilité, solubilité...). Elles doivent permettre de résoudre des problèmes aussi variés que l'identification de partenaires d'interactions, l'identification de protéines d'organismes dont le génome n'est pas encore séquencé, et la caractérisation fine de la structure primaire.

L'objectif de ce travail est de développer de méthodologies d'analyse protéomique adapté à des problématiques particulière. Les approches protéomiques sont composées de plusieurs étapes mettant en jeu la préparation d'échantillon, la séparation des protéines et peptides et enfin l'analyse des protéines et peptides par spectrométrie de masse suivie de l'interprétation des données générées. Les études présentées dans cette thèse ont nécessité d'adapter ces différentes étapes pour répondre au mieux aux diverses questions biologiques auxquelles j'ai été confrontée. Celles-ci consistaient en :

- l'identification de partenaires d'interaction,
- la caractérisation fine de modifications post-traductionnelles,
- l'identification avec une grande confiance des protéines d'organismes dont le génome n'est pas connu.

Ce manuscrit se compose de 4 parties :

La première partie de ce manuscrit est une **partie bibliographique**. Un premier chapitre est consacré à la présentation des stratégies protéomiques d'identification des protéines par spectrométrie de masse. Le second chapitre décrit les techniques de séparation et les différents types de spectromètres de masse utilisés lors ma thèse.

Les résultats obtenus sont ensuite présentés en 3 parties distinctes:

- La première partie concerne l'identification des protéines impliquées dans des complexes protéine/protéine et ARN/protéine.

- **Un premier chapitre** introduit les stratégies protéomiques utilisées pour l'identification de partenaires protéiques de protéines ou d'ARNs. Ce chapitre présente également la stratégie que j'ai

développée au cours de mon travail de thèse pour l'étude de complexes multi-protéiques et protéine/ARN.

- **Le second chapitre** porte sur l'étude des partenaires d'interaction de la poly(ADP-ribose) polymérase-2, protéine assurant l'intégrité du génome.

- **Le troisième chapitre** décrit deux études portant sur l'identification de protéines impliquées dans des complexes ARN/protéine : la recherche des protéines nucléaires impliquées dans la régulation des sites d'épissage A7 de l'ARN VIH-1 et l'identification des protéines recrutées par l'ARN Xist impliqué dans la mise en place de l'inactivation du chromosome X.

- La deuxième partie de ce travail de thèse concerne la caractérisation fine de protéines phosphorylées purifiées.

- **Le premier chapitre** fait le point sur les difficultés liées à l'analyse de protéines phosphorylées et à la localisation précise des sites de phosphorylation par spectrométrie de masse. Ce chapitre présente également la stratégie développée lors de mon travail de thèse pour la caractérisation des sites de phosphorylation d'une protéine purifiée (DAPK) et d'un complexe de trois protéines potentiellement phosphorylées (CAK).

- **Le second chapitre** est consacré à la caractérisation fine des sites de phosphorylation de la partie kinase de la protéine DAPK1 et de son mutant.

- **Le troisième chapitre** décrit l'étude du complexe CAK (cdk-activating kinase) composé de trois protéines. Le taux relatif de phosphorylation du complexe ainsi que la localisation fine des sites de phosphorylation ont été réalisés.

- La troisième partie de mon manuscrit concerne les apports de l'approche protéomique pour l'identification des protéines plasmatiques différentiellement exprimées lors du jeûne prolongé chez le manchot Adélie, un organisme dont le génome n'est pas séquencé.

- **Le premier chapitre** présente les stratégies analytiques qui permettent l'identification de protéines plasmatiques issues d'organismes peu ou pas représentés dans les banques de données, ainsi que les intérêts et les limitations de ces approches. Ce chapitre présente également plus en détails une des étapes clés de ce type d'analyse, i.e. la préparation d'échantillons plasmatiques qui est cruciale pour permettre la meilleure détection et l'identification des protéines. Enfin, les méthodes qui permettent la quantification relative du niveau d'expression des protéines entre des situations variées est présentée puisque ce type d'informations est indispensable dans une approche différentielle afin de pouvoir en déduire quels sont les mécanismes moléculaires susceptibles d'être impliqués dans la réponse biologique observée.

- **Le second chapitre** porte sur l'application de l'approche protéomique différentielle à l'identification des protéines plasmatiques différentiellement exprimées (quantification des variations d'expression protéique) selon le statut métabolique de manchots Adélie (dont le génome n'est pas séquencé) qui jeûnent spontanément pendant une période plus ou moins prolongée lors de leur reproduction.

## **Introduction bibliographique**

Chapitre I – Les stratégies générales en protéomique

Chapitre II – Les outils utilisés en protéomique



Une protéine est une macromolécule, formée d'une ou plusieurs chaînes d'acides aminés, qui résulte de la traduction de l'ARN messager lui-même issu de la transcription de l'ADN. La synthèse de protéines résulte donc directement de l'information génétique fournie par l'organisme. Selon leur structure, leur fonction va varier. Les protéines peuvent remplir des rôles dans les phénomènes tels que la catalyse, le transport, la communication, la signalisation, ou encore la reconnaissance. Selon l'état physiologique des cellules et des tissus, le type de protéines exprimées, leur abondance, leur état de modification peut varier, leur caractérisation est devenue primordiale à ce jour pour mieux comprendre les processus biologiques [Wilkins et al., 1996; Aebersold & Goodlett, 2001].

Le terme "protéomique" fut alors introduit pour définir l'analyse de ces protéines exprimées dans une cellule ou un tissu à un moment donné [Anderson & Anderson, 1998]. Toutefois, la définition du terme "protéomique" a été longtemps discutée et aujourd'hui une définition adéquate est la suivante : "l'étude de l'ensemble des protéines exprimées dans une cellule incluant leurs isoformes, les modifications qu'elles peuvent subir, leurs interactions avec d'autres protéines, leur description structurale et les complexes qu'elles peuvent former, donc de ce fait tout ce qui est post-génomique" [Tyers & Mann, 2003].

Les nouveaux besoins en analyse protéomique résident donc dans une identification toujours plus fine des protéines. Le véritable objectif est une caractérisation complète des protéines ce qui inclut entre autre la mise en évidence de modifications post-traductionnelles, de mutations, leurs interactions avec d'autres molécules mais également l'étude de l'hétérogénéité de ces protéines. Pour atteindre cet objectif ambitieux, il est nécessaire de développer des méthodologies protéomiques adaptées à chaque question mais aussi de disposer d'instruments protéomiques adaptés.

Le premier chapitre de cette partie est consacré à la description des différentes stratégies en protéomique. Les outils nécessaires à ces diverses stratégies sont présentés dans le second chapitre.





## **Chapitre I – les stratégies générales en protéomique**

### **1. Les stratégies classiquement utilisées**

Deux approches sont classiquement utilisées en protéomique. Les protéines sont séparées par gel d'électrophorèse monodimensionnel (1D) ou bidimensionnel (2D). Après excision des bandes (gel 1D) ou spots (gel 2D), les protéines sont digérées enzymatiquement. Les peptides sont ensuite analysés par l'approche "empreinte peptidique massique" ou par l'approche nanoLC-MS/MS.

#### **1.a. L'empreinte peptidique massique**

L'analyse par spectrométrie de masse permettant de travailler sur des échantillons complexes, le concept de l'identification par "empreinte peptidique massique" (PMF, Peptide Mass Fingerprint) fût décrit dans les années 1990 par cinq groupes de chercheurs [Henzel et al., 1993; James et al., 1993; Mann et al., 1993; Pappin et al., 1993; Yates et al., 1993].

Généralement associée à une étape de séparation des protéines sur gel d'électrophorèse, cette stratégie consiste à :

- i) Digérer les protéines en peptides avec une enzyme spécifique,
- ii) Mesurer précisément, pour un échantillon donné, l'ensemble des masses des peptides présents dans cet échantillon,
- iii) Comparer via des moteurs de recherche la liste de masses mesurées aux listes de masses théoriques générées *in silico* à partir des banques protéiques (Figure 1). L'interprétation de ces données se fait à l'aide de moteurs de recherches automatisés.

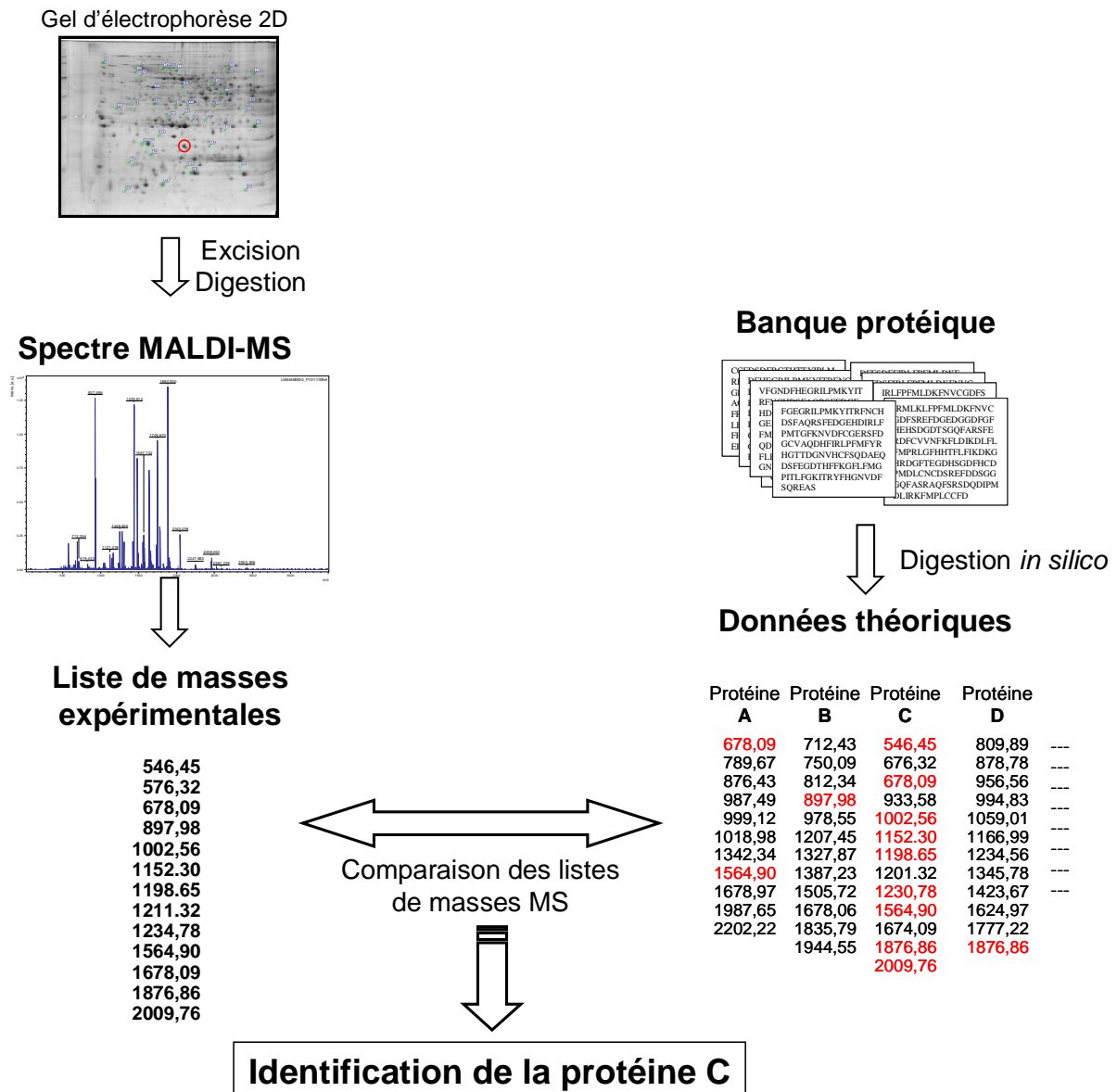


Figure 1 : Illustration de la stratégie d'identification des protéines, séparées par gel d'électrophorèse 2D, par empreinte peptidique massique à partir d'analyse MALDI-TOF-MS.

Classiquement, les identifications de protéines par PMF sont réalisées sur des extraits de peptides de spots de gels 2D analysés par MALDI-TOF-MS. Les analyses par MALDI-TOF-MS sont adaptées pour le PMF car elles sont rapides et facilement automatisables, tolérantes aux sels et aux détergents, très sensibles et présentent une bonne résolution et une bonne précision de masse (10 ppm) [Horn et al., 2004]. Les identifications par PMF peuvent aussi être réalisées à partir de données de LC-MS obtenues sur des instruments permettant de mesurer des masses avec une bonne précision (ESI-MS ou FT-ICR).

Afin de valider une identification par PMF, plusieurs critères sont à prendre en compte :

- Le pourcentage de recouvrement de séquence obtenu ou plus précisément le nombre de peptides identifiés car le pourcentage de recouvrement dépend de la taille de la protéine et peut donc être élevé même si le nombre de peptides identifiés est faible dans le cas de petites protéines.
- Les erreurs observées sur les mesures de masse.
- La comparaison entre le pI et le poids moléculaire théorique de la protéine identifiée à sa position sur le gel lorsqu'il s'agit d'analyses de spots de gel 2D.

Les critères d'acceptation des identifications par PMF sont inclus dans les directives de publication de données protéomiques [Carr et al., 2004; Bradshaw et al., 2006; Wilkins et al., 2006].

Malgré ses avantages qui en font une technique de choix pour l'analyse protéomique à haut débit des protéines contenues dans des spots de gels 2D, cette stratégie d'identification présente de nombreuses limites :

- Les échantillons analysés doivent être relativement "simples" car les mélanges complexes vont donner lieu à des superpositions d'empreintes peptidiques difficiles à interpréter et à des phénomènes de suppression de signal [Clauser et al., 1999].
- La digestion de protéines de faible poids moléculaire ou faiblement abondantes va produire un petit nombre de peptides pouvant être identifiés ce qui influera sur la qualité et la confiance accordée à l'identification de ces protéines.
- La précision des mesures de masse doit être très élevée pour limiter les identifications de faux positifs [Clauser et al., 1999].

Avec la croissance exponentielle de la taille des banques de données, la spécificité du PMF est de plus en plus limitée et l'identification de faux positifs de plus en plus probable. En effet, plusieurs compositions en acides aminés peuvent avoir des masses identiques. Si on ne dispose que de la masse du peptide, rien ne permettra de valider l'une ou l'autre des séquences. En augmentant le nombre de séquences dans les banques, le nombre de masses théoriques augmente et donc également le nombre de masses redondantes.

### **1.b. La stratégie nanoLC-MS/MS**

L'utilisation du couplage nanoLC-MS/MS pour l'identification de protéines est une des techniques les plus utilisées en analyse protéomique. Cette stratégie (Figure 2) est souvent composée de différentes étapes :

- i) Séparation des protéines sur gel d'électrophorèse,
- ii) Digestion des protéines avec une enzyme spécifique,
- iii) Séparation et fragmentation des peptides par nanoLC-MS/MS,

iv) Interprétation des spectres de fragmentation et comparaison aux banques de séquences à l'aide de moteurs de recherche dédiés.

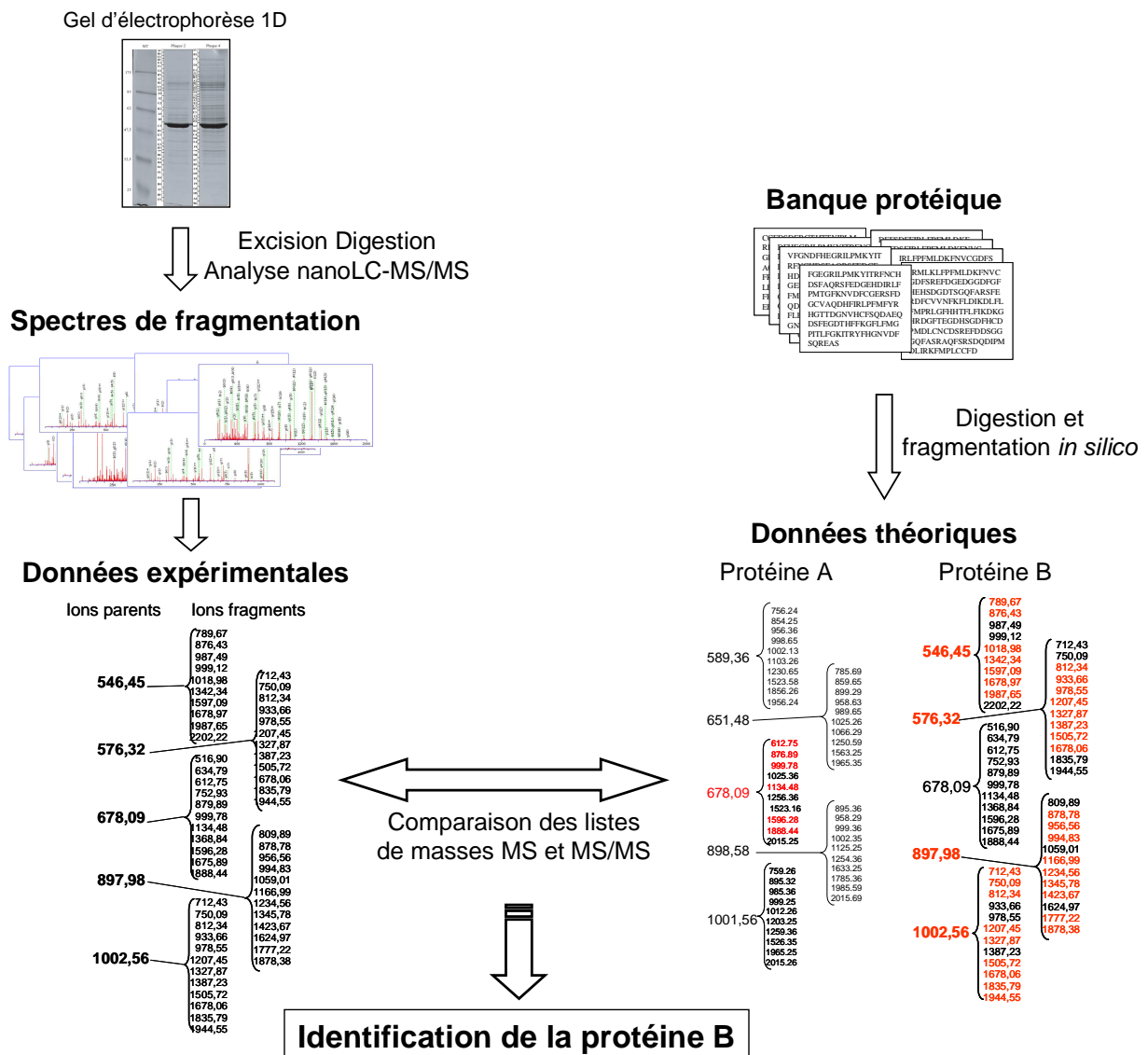


Figure 2 : Illustration de la stratégie d'identification des protéines, séparées par gel d'électrophorèse 1D, à partir d'analyses nanoLC-MS/MS.

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

De même, cette étape de séparation permet d'effectuer de la fragmentation sur chacun des peptides protonés séparés, contenant les informations de séquences du peptide, et de pouvoir travailler en automatique. L'ensemble des spectres de fragmentation permettra alors d'identifier des

protéines plus aisément mais également de mettre en évidence la présence de modifications post-traductionnelles ou des mutations sur les séquences.

### 1.b.1). Les règles de fragmentation

Les règles de fragmentations peptidiques permettent d'interpréter les données de fragmentation obtenues et ainsi d'identifier les protéines présentes dans l'échantillon.

L'observation de la fragmentation d'un grand nombre de peptides a permis à Roepstorff et Fohlmann d'établir une première nomenclature de fragmentation peptidique [Roepstorff & Fohlman, 1984] qui a été complétée et simplifiée par Biemann quelques années plus tard [Biemann, 1990] : cette nomenclature fait toujours référence aujourd'hui (Figure 3).

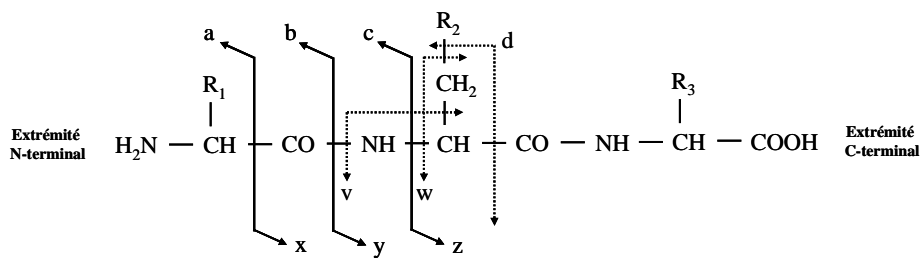


Figure 3 : Nomenclature de fragmentation d'un peptide constitué de trois acides aminés (R1, R2 et R3) selon Biemann [Biemann, 1990].

Deux types d'ions sont observés :

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

Les différences de masse entre des ions consécutifs d'une même série permettent de déterminer l'identité des acides aminés et donc de déduire la séquence du peptide (Figure 4).

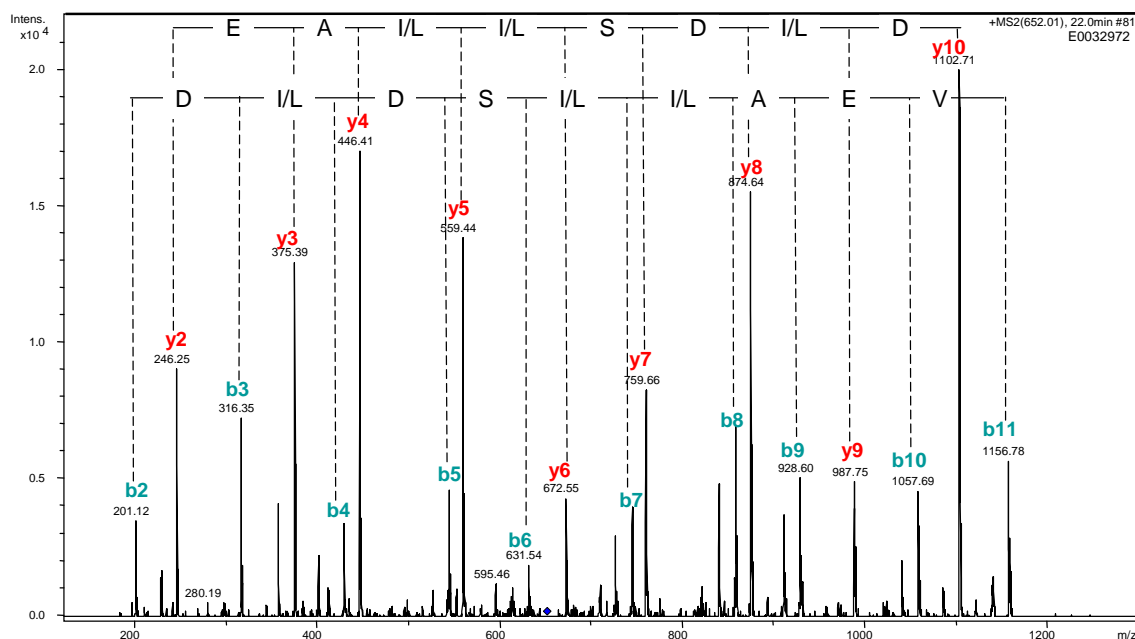


Figure 4 : Illustration de la détermination de la séquence à partir d'un spectre de fragmentation. (Spectre de fragmentation de l'ion  $m/z$  652.0, 2+ ; Séquence SLDLDSIIAEVK).

Toutefois, dans le cas de certains acides aminés, il ne sera pas possible de trancher. C'est le cas par exemple de la leucine (L) et de l'isoleucine (I) qui sont des isomères et ont donc la même masse. De même, selon la précision de mesure de masse du spectromètre de masse, il ne sera pas possible de trancher dans le cas d'une lysine (K) et d'une glutamine (Q) qui ne diffère en masse que de 0,036 Da. Néanmoins, ces difficultés peuvent être contournées en utilisant des spectromètres de masse ayant des résolutions élevées.

Des doubles fragmentations peuvent générer ce que l'on appelle des fragments internes : les fragments internes ne contenant qu'une seule chaîne latérale et formés par la combinaison d'une coupure de type a et d'une coupure de type y sont appelés ions immoniums [De Hoffmann et al., 1994]. La présence de ces ions immoniums dans les basses masses sur les spectres MS/MS est très informative pour l'interprétation des spectres et confirme la présence de certains acides aminés dans la séquence (voir application pour l'identification de phosphopeptides Partie III - Chap. II).

### 1.c.2). L'identification des protéines par peptide Fragment FingerPrinting (PFF)

Le terme PFF fut introduit pour la première fois en 2004 par Blueggel *et al.* [Blueggel et al., 2004]. Cette approche est la plus couramment utilisée pour l'interprétation des spectres de fragmentation.

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

- Mascot ([www.matrixscience.com](http://www.matrixscience.com))
- Sequest ([www.thermo.com](http://www.thermo.com))
- X!Tandem ([www.thegpm.org](http://www.thegpm.org))
- OMSSA (<http://pubchem.ncbi.nlm.nih.gov/omssa/>)

L'algorithme de Mascot, disponible au laboratoire, est basé sur une approche probabilistique (Probability-based matching) [Perkins et al., 1999]. Les masses des fragments MS/MS calculées à partir des peptides présents dans la banque de données sont comparées aux masses expérimentales des fragments mesurées. Un score est calculé qui reflète la significativité de la correspondance entre les spectres théoriques et expérimentaux.

## **2. Des alternatives aux approches protéomiques classiques**

### **2.a. La chromatographie multidimensionnelle**

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

Ainsi, en 1999, Yates *et al.* proposent une nouvelle stratégie d'identification pour l'analyse directe d'échantillons complexes, la chromatographie multidimensionnelle couplée à la spectrométrie de masse : la DALPC "Direct Analysis of Large Protein Complex" [Link et al., 1999]. On parle de stratégie MDLC pour "Multi-Dimensional Liquid Chromatography" [Giddings, 1984]. Elle a ouvert la voie à la "protéomique shotgun" qui consiste à obtenir un profil global des protéines présentes dans un mélange par l'analyse directe de celui-ci [Wolters et al., 2001]. Le mélange complexe de protéines est directement digéré en peptides qui sont séparés par des étapes de chromatographies multiples avant d'être analysés en MS et fragmentés en MS/MS [Swanson & Walshburn, 2005].

L'approche la plus utilisée consiste à coupler une première étape de séparation des peptides par une chromatographie d'échange de cations (SCX : Strong Cation Exchange) suivie d'une étape de chromatographie de phase inverse. Trois approches ont été décrites dans la littérature pour la réalisation des analyses SCX/RP/MS/MS :

- La stratégie off-line (Figure 5-a) : L'étape de chromatographie SCX est réalisée de façon indépendante, les fractions sont collectées avant d'être analysées individuellement en nanoLC-MS/MS avec séparation sur colonne de phase inverse classique. N'étant pas directement couplée à la colonne RP, cette approche est très flexible quant aux solvants utilisés pour la SCX (chlorure de

sodium et chlorure de potassium pourront être utilisés malgré leur incompatibilité avec l'analyse par MS).

- La stratégie on-line (Figure 5-b) : les deux colonnes sont couplées via un système de valve contenant une colonne de chargement de phase inverse C18. Cette étape permet d'éliminer les éluants utilisés incompatibles avec la MS avant l'élution des peptides de la colonne d'enrichissement sur la colonne RP couplée au spectromètre de masse [Nagele et al., 2004]. Néanmoins, l'architecture des valves dans cette configuration est très complexe, ce qui rend cette configuration délicate à mettre en œuvre. Par contre, une variante de cette configuration utilisant des solvants compatibles avec la MS comme l'acétate d'ammonium pour la SCX et ne nécessitant donc pas de colonne d'enrichissement a également été décrite [Xiang et al., 2004].

- La stratégie MudPIT (Multidimensional Protein Identification Technologies) (Figure 5-c) : Les peptides sont chargés sur une colonne microcapillaire bi-phasique remplie en série par une résine SCX et par une résine de phase inverse. La colonne chargée est ensuite intégrée dans le système et les peptides sont élués de la résine SCX par paliers avec un gradient en sels pour se déposer en tête de la phase RP. Ils sont ensuite élués une seconde fois en fonction de leur hydrophobicité croissante vers le spectromètre de masse [Link et al., 1999; Washburn et al., 2001; Wolters et al., 2001; Gonzalez-Begne et al., 2009].



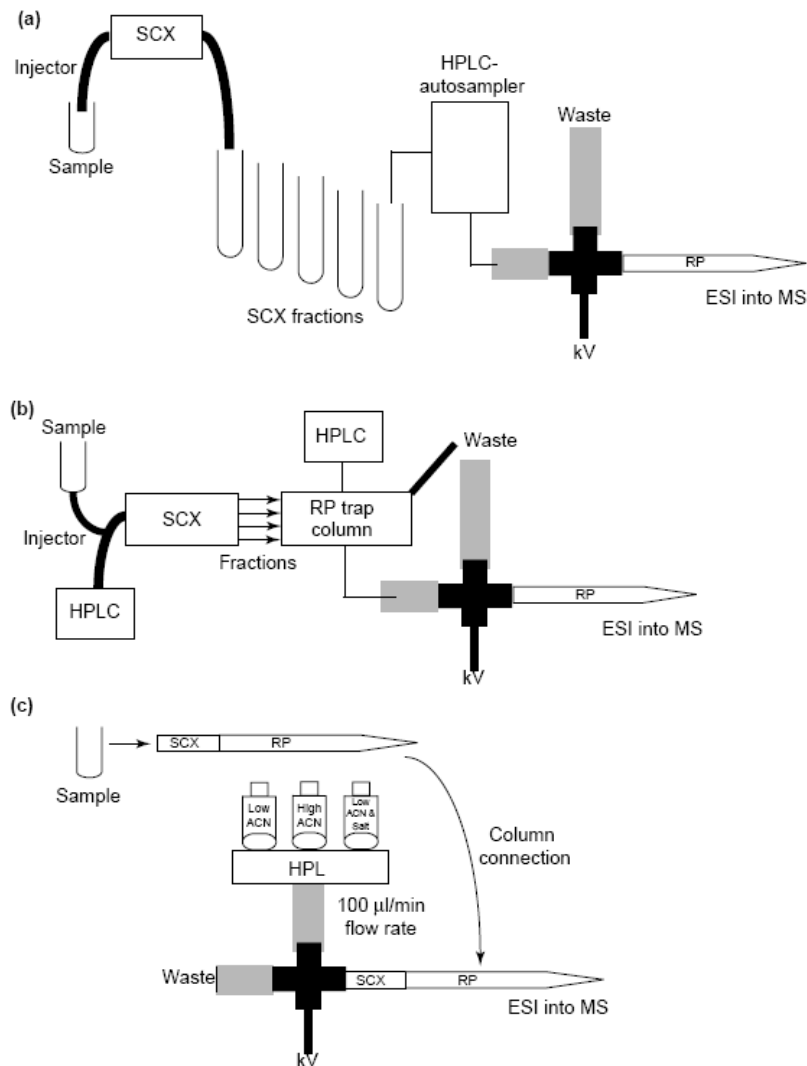


Figure 5 : Les différentes possibilités de chromatographie multidimensionnelle [Swanson & Walshburn, 2005].

Une des autres approches utilisées consiste à coupler une première étape de séparation des peptides par une chromatographie de phase inverse à pH élevé (pH~ 10) suivie d'une étape de chromatographie de paire d'ion sur phase inverse à faible pH (pH~ 2) (IP-RP). La séparation des peptides à haut puis bas pH est possible grâce à la structure amphiphile des peptides [Toll et al., 2005].

La comparaison entre les stratégies SCX/RP/MS/MS et RP/IP-RP/MS/MS a permis en mettre en évidence que ces deux méthodes sont complémentaires. En effet, des résultats équivalents en terme de nombre de peptides identifiés lors de l'analyse du protéome de *Corynebacterium glutamicum*, 65.3 % des protéines ont été identifiées par les deux méthodes, 14.5 % par la méthode SCX/RP/MS/MS et 20.2 % par la méthode RP/IP-RP/MS/MS [Delmotte et al., 2009].

Une fois les protéines analysées par MS/MS, le traitement des données et la recherche dans les banques de données vont permettre d'identifier de nombreuses protéines. L'inconvénient majeur de la stratégie de digestion en solution réside dans l'interprétation des données générées par

l'analyse directe de mélanges complexes de peptides. En effet, le nombre de peptides générés est considérable et certaines identifications de protéines ne se font qu'avec un seul peptide. Il peut s'avérer difficile d'attribuer avec certitude un peptide à une protéine notamment dans le cas d'identifications de protéines ayant de fortes homologues de séquences [Nesvizhskii & Aebersold, 2005]. La validation des interprétations et des identifications de protéines doit se faire avec la plus grande prudence en vérifiant la qualité des spectres de fragmentation.

## **2.b. L'approche *de novo***

L'approche *de novo* consiste en l'interprétation des spectres de masse de fragmentation afin de déduire directement de ces spectres des morceaux de séquence en acides aminés. Les morceaux de séquence d'acides aminés déterminés peuvent ensuite être soumis à des algorithmes de recherche de similarités de séquences dans les banques de données.

Cette approche permet d'identifier des protéines non présentes dans les banques de données par homologie avec des protéines ayant des séquences proches, présentes dans les banques. Les possibilités de travailler et d'identifier des protéines d'organismes peu étudiés, dont le génome n'a pas encore été séquencé, sont donc beaucoup plus vastes qu'avec les programmes classiquement utilisés. De plus, contrairement à l'approche stricte d'identification dans les banques de données (PFF), est tolérante aux "erreurs" de séquence, i.e. les mutations, les délétions d'acides aminés...

### **2.b.1) Le séquençage *de novo***

L'étude des règles de fragmentation des peptides a rendu l'interprétation des fragments des peptides possible. Le séquençage peut se faire manuellement, i.e. avec la calculatrice ou bien avec des outils d'aide à l'annotation manuelle présents dans les logiciels d'analyses des spectromètres de masse. Ces méthodes sont longues et fastidieuses. Cependant, l'interprétation manuelle complète des spectres peut être nécessaire pour la recherche de certaines modifications post-traductionnelles ou mutations.

Les algorithmes d'aide au séquençage *de novo* sont nécessaires pour les études à plus grande échelle notamment lors d'études concernant l'identification de protéines provenant d'organismes dont le génome n'est pas séquencé. Plusieurs logiciels sont disponibles [Nesvizhskii et al., 2007] :

- Peaks ([www.bioinformaticssolutions.com](http://www.bioinformaticssolutions.com))
- PepNovo ([www.peptide.ucsd.edu/pepnovo.py](http://www.peptide.ucsd.edu/pepnovo.py))
- Lutefisk ([www.hairyfarguy.com/lutefisk](http://www.hairyfarguy.com/lutefisk))

Chaque algorithme possède ses spécificités. Seul le logiciel Peaks (Bioinformatics Solution) est décrit, car il s'agit du logiciel disponible au laboratoire, qui fut utilisé durant cette thèse.

L'algorithme de Peaks repose sur un modèle mathématique qui est décrit par Ma *et al.* [Ma *et al.*, 2003] et fonctionne en 4 étapes :

- i) Pré-traitement des données MS/MS brutes incluant le filtrage du bruit, le centrage des pics et la déconvolution des ions multi-chargés,
- ii) Calcul d'un grand nombre de séquences candidates possibles,
- iii) Sélection des meilleurs candidats,
- iv) Calcul d'un score de confiance global pour la séquence du peptide et d'un score de confiance individuel pour chacun des acides aminés.

Le principal avantage de cet algorithme est que l'absence d'un pic dans la succession d'une série y (ou b) ne pose pas de problème majeur. En effet, Peaks pourra trouver deux morceaux de séquence dans un même peptide avec un niveau très élevé de confiance entrecoupés par un morceau de faible confiance.

Cependant, un point faible peut être relevé : l'algorithme se base toujours sur la masse globale du peptide et va donc toujours proposer une séquence complète du peptide allant jusqu'à la masse du parent. Ce point est quelquefois limitant car souvent le spectre de fragmentation est très beau sur une partie seulement de la gamme de masse alors que l'algorithme va toujours jusqu'au bout de la séquence au détriment parfois de la bonne détermination de la belle région du spectre uniquement.

De plus, les spectres MS/MS doivent être de très bonne qualité pour pouvoir en déduire un morceau de séquence suffisamment long (5-7 acides aminés consécutifs). Cette exigence de qualité est beaucoup plus importante que pour des recherches classiques par comparaison de listes de masses. En effet, si les séries des ions fragments permettant de lire les différences de masses correspondant à des acides aminés sont interrompues, l'interprétation devient rapidement impossible par séquençage *de novo*.

### **2.b.2) La recherche par homologie de séquence**

Les séquences déterminées à partir des spectres MS/MS ne sont en général que des fragments de séquences de peptides et ne dépassent pas 10 à 15 acides aminés dans les meilleurs cas. En 2001, Shevchenko *et al.* ont développé un protocole de recherche adapté pour ce type de données qu'ils ont appelé MS-BLAST ([Shevchenko *et al.*, 2001]; <http://dove.embl-heidelberg.de/Blast2/msblast.html>). L'algorithme MS-BLAST ne tolère pas d'interruption à l'intérieur d'un peptide mais en revanche les interruptions entre peptides ne sont aucunement pénalisées et peuvent être de n'importe quelle longueur. Cela permet de soumettre l'ensemble des morceaux de séquence obtenus pour l'ensemble des spectres MS/MS traités en une seule requête.

MS-BLAST identifie des HSPs (high-scoring segment pairs) entre des peptides soumis et des peptides présents dans les banques et leur attribue un score individuel. Lorsque plusieurs HSPs sont identifiées sur une même protéine, le score global tiendra compte de la somme des scores des HSPs

individuelles. Ces scores seront comparés à des "scores seuils" calculés et fixés au préalable par l'algorithme pour attester du degré de confiance de l'identification. Le calcul du score ne tient pas compte du nombre de peptides soumis dans une requête ce qui présente le grand avantage de pouvoir soumettre l'ensemble des peptides déterminés par séquençage *de novo* et même d'y inclure des peptides redondants lorsqu'il y a doute et hésitation entre plusieurs séquences pour un même spectre.

Une évaluation des performances de l'algorithme MS-BLAST a été réalisée par Habermann *et al.* [Habermann *et al.*, 2004] et les résultats de plus de 20000 requêtes MS-BLAST ont conduit à la détermination d'un taux d'identification de faux positifs de moins de 3 %.

### **2.c L'approche "error-tolerant"**

L'approche "error-tolerant" de Mascot est une technique permettant de faire des identifications en autorisant la possibilité d'"erreurs" par rapport à la banque protéique dans les séquences identifiées. Il s'agit de réaliser les identifications en deux étapes, tout d'abord par l'approche PFF. Cette approche va permettre d'identifier un certain nombre de protéines. Puis, une seconde recherche est lancée dans la banque constituée des protéines précédemment identifiées. Durant cette recherche, l'ensemble des modifications/mutations est testé dans les limites de la précision de masse demandées [Creasy & Cottrell, 2002].

## **3. La bioinformatique est indissociable de l'analyse protéomique**

### **3.a. Les banques de données protéiques**

Il existe une grande variété de banques protéiques disponibles qui diffèrent selon leur exhaustivité, leur degré de redondance et la qualité de leurs annotations [Apweiler *et al.*, 2004]. Le degré de vérification des séquences, de corrections et d'ajouts d'annotations est un critère important de distinction des différentes banques protéiques accessibles.

On peut ainsi distinguer deux catégories de banques protéiques :

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

ii) Les banques protéiques couramment utilisées en protéomique sont des banques pour lesquelles un travail de nettoyage, de tri, de documentation et d'annotation a été réalisé par des

experts. La liste de ces banques est longue et seules les banques les plus couramment utilisées en analyse protéomique sont brièvement détaillées ici :

- Le consortium UniProt :

En 2003, le Swiss Institute of Bioinformatics (SIB), l'European Bioinformatics Institute (EBI) et le groupe de la Protein Information Resource (PIR) ont décidé d'unir leurs forces et de former le consortium UniProt centralisant les banques PIR-PSD, SwissProt et TrEMBL [Apweiler et al., 2004; Bairoch et al., 2005]. Ces banques sont accessibles sur le site <http://www.uniprot.org/>

La banque PIR-PSD (Protein Information Resource Protein Sequence Database) :

Il s'agit de la plus ancienne des banques "corrigées" établie dès 1984. La caractéristique de cette banque est qu'elle est organisée selon une classification par familles et super-familles de protéines et les annotations sont également basées sur cette organisation. Cette banque contient des annotations fonctionnelles, structurales, bibliographiques et génétiques.

La banque Swiss-Prot, ou UniProtKB/Swiss-Prot :

Créée en 1986 par Amos Bairoch, la banque Swiss-Prot est l'une des banques corrigées de référence. Il s'agit de la banque de séquences protéiques la moins redondante, la mieux annotée et la plus intégrée par des références croisées vers d'autres banques. Le processus d'intégration d'une nouvelle séquence dans Swiss-Prot comprend plusieurs étapes visant à contrôler la pertinence de son entrée dans la banque. Ces étapes sont réalisées par un groupe d'experts en annotations. Un soin particulier est apporté pour indiquer la nature, expérimentale ou bioinformatique, des informations fonctionnelles, ainsi que le niveau de confiance qui est accordé à ces informations.

La banque TrEMBL (Translation from EMBL) ou UniProtKB/TrEMBL :

La banque TrEMBL est une banque complémentaire à Swiss-Prot, créée pour donner un accès rapide aux utilisateurs à de nouvelles séquences protéiques. Elle contient les séquences annotées *in silico* d'après la traduction de l'ensemble des séquences codantes des banques nucléotidiques hormis celles déjà présentes dans Swiss-Prot mais également des séquences protéiques issues de publications ou déposées directement par les utilisateurs. Le principal travail de cette banque est de combiner informatiquement toutes les séquences identiques afin d'éviter des redondances, et d'améliorer automatiquement les informations contenues dans TrEMBL. Ce processus est basé sur un système standardisé de transfert des annotations des protéines en comparant ces séquences non annotées aux séquences annotées dans Swiss-Prot. Une fois ces vérifications faites, ces séquences pourront être traitées par des experts qui jugeront de la pertinence de ces pré-annotations et qui pourront se réserver le droit de les modifier ou encore de les supprimer.

- La banque “International Protein Index” (IPI) [Kersey et al., 2004] :

Cette banque protéique non redondante est focalisée sur le protéome humain, de rat et de souris. Elle contient à la fois les séquences annotées de banques protéiques issues de Swiss-Prot et TrEMBL mais aussi des séquences issues des prédictions automatiques du génome (Ensembl et RefSeq). Elle est accessible sur le site de l’European Bioinformatics Institute : [www.ebi.ac.uk/IPI/](http://www.ebi.ac.uk/IPI/)

- La banque “Entrez Protein” :

Cette banque généraliste se trouve à la jonction des deux types de banques car elle contient à la fois les séquences issues de la traduction automatique des banques nucléotidiques mais aussi des séquences annotées de banques protéiques (Swiss-Prot, PIR, de RefSeq et de la PDB (Protein Data Bank)). Cette banque est donc volumineuse et redondante mais par contre elle présente l’avantage d’être relativement complète. Elle est accessible sur le site du NCBI : <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>.

### **3.b. L’évaluation statistique des données**

Afin d’augmenter la confiance dans les identifications obtenues par PFF, une estimation du taux de faux positifs peut être réalisée sur les résultats issus des recherches dans les banques protéiques à l’aide des moteurs de recherche. Pour cela, deux approches sont proposées : l’approche “Target-decoy” et l’approche empirique de Bayes.

L’approche “Target-Decoy” [Elias & Gygi, 2007] : Il s’agit de donner la possibilité au moteur de recherche d’identifier une fausse séquence en lui permettant de chercher dans une banque de données target-decoy, i.e. une banque de données protéiques vraies (target) combinée avec une banque de données protéiques fausse (decoy). Cette banque decoy, créée à partir de la banque target est soit une banque inversée (les séquences des entrées sont lues du C-ter vers le N-ter) soit une banque aléatoire (les séquences sont mélangées). Le taux de faux positif est alors estimé grâce aux nombres d’identifications dans la banque decoy par rapport aux nombres d’identification total.

L’approche empirique de Bayes : Les scores de recherche observés pour toutes les identifications sont modélisés. Cela permet d’identifier deux distributions, celles des identifications correctes et celles des identifications incorrectes. Une probabilité d’identification correcte est attribuée à un spectre de fragmentation en fonction du score. Ce type d’approche est utilisé notamment par l’utilisation du logiciel PeptideProphet [Keller et al., 2002].

### **3.c. Automatisation des identifications**

La recherche dans les banques de données et l’évaluation statistiques des résultats peut être long, des logiciels d’automatisation ont donc été développés [Palagi et al., 2006] :

- ProteinLynx Global Server ([www.water.com](http://www.water.com))
- ProteinScape™ ([www.proteinscape.com](http://www.proteinscape.com))

- Scaffold ([www.proteomesoftware.com](http://www.proteomesoftware.com))

- Trans-Proteomic Pipeline (<http://tools.proteomecenter.org/software.php>)

Ces logiciels peuvent intégrer les résultats de plusieurs moteurs de recherche et une validation statistique des données obtenues. Le Trans-Proteomic Pipeline [Keller et al., 2005] intègre notamment plusieurs moteurs de recherche et l'algorithme de validation statistique Peptideprophet. Scaffold peut intégrer les résultats de recherche issues de plusieurs moteurs de recherche dont Mascot et permet d'utiliser l'approche Target-decoy ou l'approche empirique de Bayes pour la validation statistique.

L'avantage de ces logiciels est la possibilité de combiner les moteurs de recherche différents et de valider statistiquement ces nombreux résultats de façon identique et rapide. Le fait d'utiliser plusieurs moteurs de recherche peut augmenter le nombre de peptides identifiés et a pour conséquence d'obtenir des identifications plus sûres [Kapp et al., 2005; Searle et al., 2008]. De plus, le taux de faux positif est considérablement diminué par l'utilisation de plusieurs moteurs de recherche [Jones et al., 2009].





## **Chapitre II – les outils utilisés en protéomique**

L'ensemble des stratégies protéomiques nécessite des outils adaptés. Les séparations de protéines et de peptides doivent être compatibles avec l'analyse par spectrométrie de masse. Les instruments doivent être sensibles et précis, donc résolutif, afin de permettre une analyse protéomique apportant le maximum d'informations.

### **1. Les techniques de séparation**

#### **1.a. L'électrophorèse**

##### ***1.a.1) Le gel d'électrophorèse 2D***

En 1975, O'Farrel reporte la détection de 1000 spots distincts sur un gel 2D [O'Farrell, 1975]. En 2004, Görg *et al.* annoncent qu'un gel 2D peut résoudre jusqu'à 5000 protéines simultanément (2000 protéines en routine) et que 1 ng de protéine peut être détecté et quantifié dans un spot [Gorg *et al.*, 2004].

La première étape de séparation des protéines aussi appelée première dimension correspond à une isoélectrofocalisation (IEF) des protéines en fonction de leur point isoélectrique (Figure 6). Cette séparation se réalise avec un gel dans lequel est immobilisé un gradient de pH (IPG) [Bjellqvist *et al.*, 1982]. Les protéines, entraînées par un champ électrique, vont se déplacer dans ce gel en fonction de leur charge nette. Quand les protéines arrivent au point de pH qui correspond à leur point isoélectrique, leur charge globale devient nulle, et leur migration s'arrête.

Une deuxième dimension (Figure 6) permet de séparer les protéines en fonction de leurs masses moléculaires (électrophorèse SDS-PAGE) [Kenrick & Margolis, 1970]. Une fois les protéines séparées selon leur charge sur un strip IPG, celui-ci est récupéré pour être placé au sommet d'un gel de polyacrylamide de granulométrie variable en présence d'un agent dénaturant anionique, le sodium dodecylsulfate (SDS). Le SDS interagit avec les protéines en entraînant leur dénaturation et en leur donnant une charge négative [Laemmli, 1970]. Cette charge négative va permettre d'accélérer les protéines dans un champ électrique, tandis que la masse moléculaire détermine la distance de migration selon la réticulation du gel.

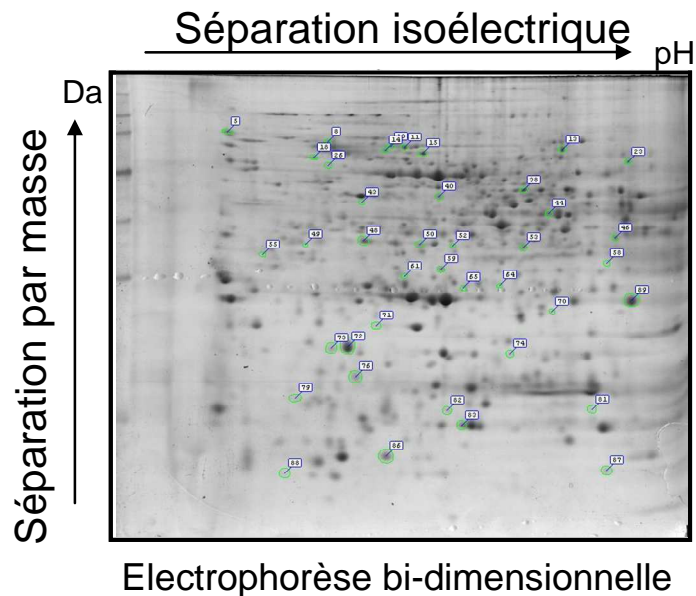


Figure 6 : Exemple de gel bidimensionnel (2D). La première dimension permettant de séparer les protéines en fonction de leur point isoélectrique et la deuxième dimension permettant une séparation en fonction de la masse.

La sensibilité de détection des gels passe par la sensibilité des colorants utilisés (Tableau 1). Les colorations les plus couramment utilisées en analyse protéomique sont :

- le bleu de Coomassie : le marquage se fait par des interactions non-covalentes entre le colorant et les groupements amino ou thiols des protéines [Neuhoff et al., 1988]. Malgré la faible sensibilité de cette coloration, elle présente le grand avantage d'être linéaire et donc très adaptée pour les études comparatives.

- la coloration à l'argent : le marquage se fait par affinité entre la protéine et le cation métallique. Sa limite de détection est 100 fois plus faible que la coloration au bleu de Coomassie [Shevchenko et al., 1996]. Cependant, elle ne témoigne pas de l'exacte stœchiométrie des protéines en présence. Au-delà de certaines concentrations, la coloration devient non linéaire et des incompatibilités avec l'analyse par spectrométrie de masse MALDI et ESI ont été mises en évidence [Gevaert & Vandekerckhove, 2000]. Toutefois, des études menées au laboratoire ont permis de mettre au point un protocole améliorant sa compatibilité avec la spectrométrie de masse [Richert et al., 2004].

- les colorations fluorescentes : le principe de ce type de marquage consiste à lier un fluorophore à la protéine par des liaisons covalentes ou non. L'inconvénient de ce type de marquage réside dans la détection, qui nécessite l'utilisation d'outils optiques très performants. Cependant, il est beaucoup plus sensible et permet de détecter des protéines jusqu'à l'attomole. Les colorations Sypro Ruby® ou le Rhutenium II [Rabilloud et al., 2001] présentent une sensibilité intermédiaire entre les colorations au nitrate d'argent et au bleu de Coomassie.

Il existe des colorations spécifiques qui permettent d'identifier des protéines particulières :

- La coloration ProQ Diamond : cette coloration est spécifique aux protéines portant une phosphorylation sur une sérine, thréonine ou tyrosine [Steinberg et al., 2003],
- La coloration ProQ Emerald, spécifique aux glycoprotéines [Steinberg et al., 2001],
- La coloration Cyanines : elle consiste à marquer les protéines avec des marqueurs fluorescents sur les cystéines (Cy2, Cy3 et Cy5). Ce marquage est utilisé pour réaliser des différentiels (DIGE : Difference Gel Electrophoresis) en marquant deux échantillons différents par un marqueur spécifique et en réalisant un seul gel d'électrophorèse, les deux marqueurs n'étant pas visible à la même longueur d'onde [Tonge et al., 2001].

Tableau 1 : Caractéristiques des colorants généralement utilisée en protéomique [Riederer, 2008]

Colorant	Spécificité	Sensibilité	Gamme linéaire
<b>Détection visible</b>			
Bleu de comassie		30-100 ng	2-3 log
Bleu de comassie colloïdale		8-10 ng	2-3 log
Nitrate d'argent		80 pg-2 ng	1.5 log
<b>Détection Fluorescente</b>			
SyproRuby		30-60 pg	3-4 log
Ruthénium II		1 ng	>3 log
Deep Purple		<1 ng	>3 log
ProQ Diamond	Phosphoprotéines	1-16 ng	>3 log
ProQ Emerald	Glycoprotéines	0.3-2 ng	>3 log
Cyanines (Cy2, Cy3 et Cy5)	Protéines marquées en amont du gel	4-8 ng	4-5 log

Bien que les gels 2D demandent un grand savoir faire, cette technique est reproductible et suffisamment rapide pour le haut débit étant donné qu'il est tout à fait possible de faire plusieurs électrofocalisations, ou plusieurs migrations SDS-PAGE en même temps. De plus, cette technique est extrêmement performante et permet d'observer plusieurs milliers de protéines ainsi que certaines modifications post-traductionnelles qui entraînent des changements au niveau du pI des protéines ou de grandes variations de masses qui seront alors visibles sur la deuxième dimension.

Toutefois, cette technique présente certaines limitations [Ong & Pandey, 2001] tels que :

- La résolution des gels : face à l'énorme diversité des protéines cellulaires, la résolution des gels 2D n'est pas toujours suffisante pour séparer l'ensemble des protéines et la co-migration de plusieurs protéines dans un spot n'est pas rare [Gygi et al., 2000; Campostrini et al., 2005]. Pour

améliorer la résolution des gels, des gels à gamme de pH réduite ont été développés [Hoving et al., 2000].

- La gamme dynamique de détection : celle-ci est de  $10^3$ - $10^4$  alors que dans une cellule on a des protéines minoritaires présentes à 100 copies par cellule alors que les plus abondantes sont présentes à 100 000 000 copies par cellule pour une gamme dynamique de  $10^6$  [Lopez, 2007]. A plus forte raison encore, dans des sérums, la gamme dynamique peut atteindre  $10^9$  [Rabilloud, 2002]. Des méthodes de pré-fractionnement sont donc nécessaires afin de visualiser les protéines minoritaires.

- La visualisation des protéines membranaires ou hydrophobes : les protéines membranaires sont largement sous-représentées sur les gels 2D et ce principalement pour 3 raisons : elles sont en général peu abondantes, leurs pI sont alcalins et elles sont faiblement solubles dans les milieux aqueux utilisés pour l'IEF [Santoni et al., 2000]. D'importantes améliorations ont été apportées en incorporant des agents de solubilisation, surfactants et détergents, tels que la thiourée ou le CHAPS. Cependant, malgré les progrès réalisés dans ce domaine, les problèmes d'extraction et de solubilisation des protéines membranaires resteront vraisemblablement les plus limitants pour les approches par gel 2D [Rabilloud, 2002].

### **1.a.2) Le gel d'électrophorèse 1D**

Le gel SDS-PAGE (sodium dodecylsulfate-PolyAcrylamide Gel Electrophoresis) permet de séparer les protéines en fonction de leurs masses apparentes quelles que soient leurs caractéristiques physico-chimiques. Facile et rapide à mettre en œuvre, la séparation par SDS-PAGE peut être utilisée seule (sans séparation par IEF) pour pré-fractionner des échantillons complexes de protéines avant leur analyse par spectrométrie de masse. Néanmoins, son pouvoir de séparation reste relativement faible et la présence de nombreuses protéines par bande de gel 1D implique un couplage avec un système de nanoLC-MS/MS. Dans cette stratégie, deux étapes de séparations sont donc réalisées [Rabilloud, 2002; Miguet, 2006] :

- Une étape de séparation des protéines par gel SDS-PAGE. Les bandes de gel 1D contenant une ou plusieurs protéines sont découpées puis les protéines sont digérées in-gel.

- Une étape de séparation des peptides par nanoLC. Les peptides sont séparés sur des nano-colonnes de phase inverse, élués par un gradient eau/acétonitrile avant d'entrer dans le spectromètre de masse.

### **1.a.3) Le Off-Gel**

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

protéines en fonction de leur  $pI$  à 0,1 unité de pH près, tout en récupérant les échantillons en solution. Ainsi, il est ensuite possible de refractionner les extraits par d'autres techniques (gel 1D par exemple). Enfin, des études ont montré son utilisation pour l'analyse de mélanges de protéines [Ros et al., 2002] et de peptides [Horth et al., 2006].

### **1.b. La chromatographie liquide**

Les techniques de séparation par chromatographie liquide ont fait l'objet de nombreuses adaptations pour rendre le couplage LC-spectrométrie de masse ESI (compatible avec l'introduction des échantillons en solution) robuste et sensible. En ESI-MS, l'intensité du signal est concentration-dépendante et non débit-dépendante [Ikonomou et al., 1990]. La recherche d'une sensibilité toujours meilleure a donc naturellement conduit au développement de systèmes nanoLC permettant de travailler à des débits de l'ordre de 100 à 300 nl/min. Les avancées techniques réalisées dans le domaine des nano-colonnes de chromatographie et de l'instrumentation de nanoLC ont ainsi amené le couplage nanoLC au cœur de l'analyse protéomique.

La nanoLC de phase inverse est classiquement utilisée pour séparer et concentrer un mélange plus ou moins complexe de peptides. Cependant, la complexité des échantillons en analyse protéomique est souvent telle que la nanoLC ne peut pas être utilisée seule, sans une autre technique de séparation en amont. En effet, des extraits protéiques cellulaires peuvent contenir plusieurs milliers de protéines qui généreront des centaines de milliers de peptides tryptiques qui ne pourront pas être séparés efficacement sur nano-colonnes. Cependant selon la complexité des échantillons, la détection peut rapidement être limitée. C'est pourquoi à ce jour, le défi réside dans le développement de nouveaux systèmes de chromatographie liquide permettant de gagner en sensibilité, en résolution mais également en temps d'analyse.

#### ***1.b.1) Le système Chip***

Un des inconvénients du système nanoLC classique réside dans le nombre de connections qui augmentent considérablement les volumes morts et entraînent une perte en résolution chromatographique mais également une perte en sensibilité. Un autre inconvénient majeur est le système d'aiguille nanospray sur certains spectromètres de masse en sortie de colonne qui nécessite d'être changé régulièrement.

De ce fait, des améliorations ont été apportées en mettant au point un nouveau système chromatographique : la Chip et son interface, le cube (Agilent technologies) [Yin et al., 2005; Yin & Killeen, 2007]. La Chip intègre une pré-colonne d'enrichissement, une colonne de séparation, le nébulisateur nanospray et l'ensemble des connections nécessaires (Figure 7). Ces chips sont constituées de films de polyimide laminés à l'intérieur desquels des canaux sont creusés par ablation-laser. Le volume des capillaires est limité dans ce système, ce qui évite les possibilités de fuites et élimine les volumes morts post-colonne. La résolution chromatographique ainsi que la sensibilité sont considérablement améliorées par l'utilisation de ce système [Carapito, 2006; Hardouin et al., 2006].

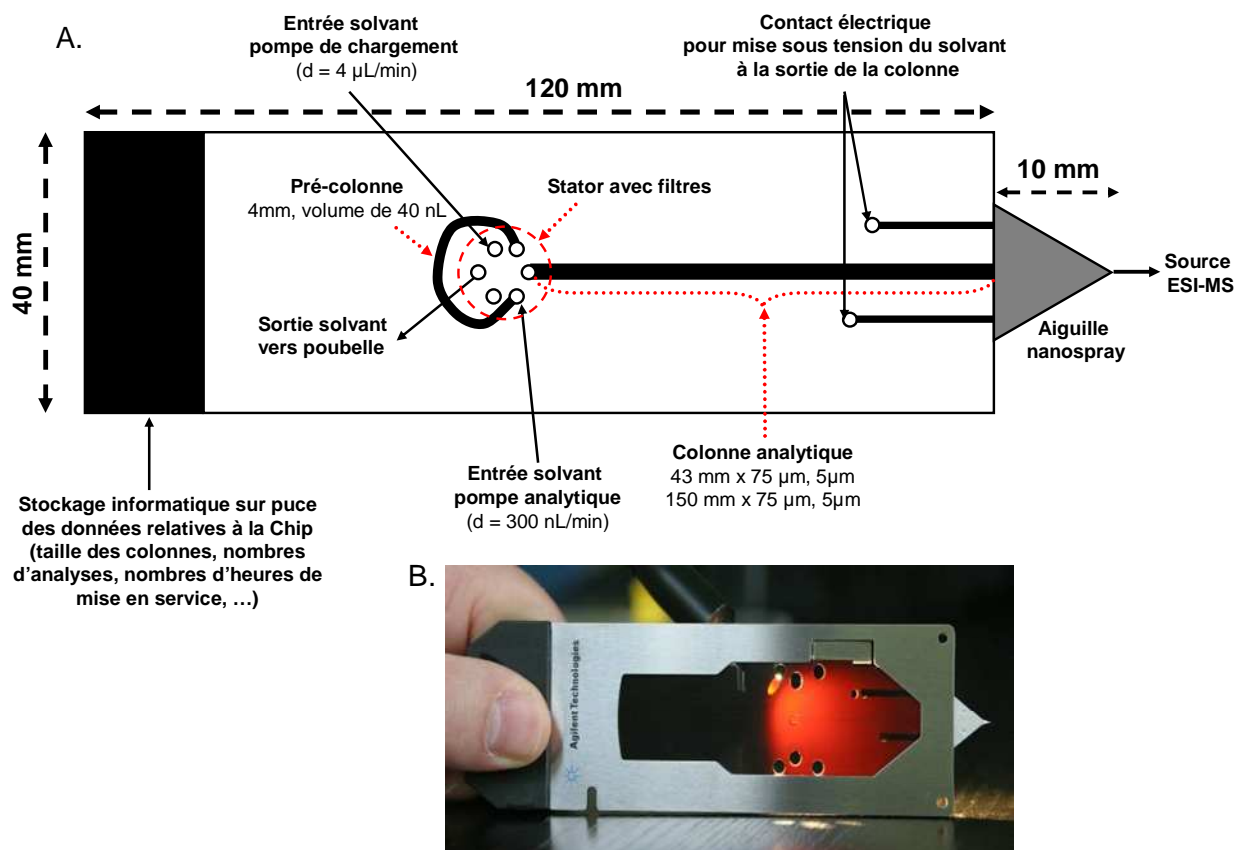


Figure 7 : A) Représentation schématique de la Chip et de ces composantes. Elle intègre la pré-colonne, la colonne analytique, l'aiguille nanospray, le stator avec les 6 ports et des filtres dans un sandwich en feuilles de polyimide protégé par une carcasse métallique. L'ensemble mesure 120 mm de long et 40 mm de large. B) Photo d'une chip.

### 1.b.2) La chromatographie UHPLC

Le développement de colonnes avec des tailles de particules toujours plus petites permettant d'augmenter la résolution et la sensibilité commencent à montrer leur utilité pour l'analyse protéomique [Motoyama et al., 2006; Kay et al., 2007]. Toutefois, ce type de colonne, permettant de travailler à des débits très élevés entraîne une augmentation de pression sur les systèmes. Afin de pouvoir palier à ce problèmes, des systèmes de chromatographie permettant de tenir des pressions très élevées furent développés (Ultra High performances Liquid Chromatography, UHPLC<sup>TM</sup>).

## 2. La spectrométrie de masse

La spectrométrie de masse est une méthode d'analyse basée sur la séparation en phase gazeuse de molécules chargées en fonction de leur rapport masse/charge ( $m/z$ ). Un spectromètre de masse est constitué de trois parties (Figure 8) :

- La source d'ionisation, qui consiste à volatiliser et à ioniser les molécules,
- L'analyseur, qui permet de collecter et de séparer les molécules ionisées selon leur rapport masse sur charge ( $m/z$ ),

- Le détecteur, qui va permettre de transformer les ions en courant qui sera mesuré.

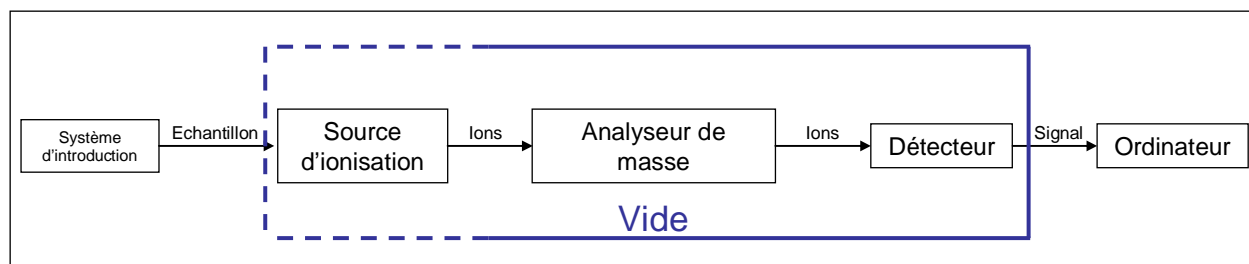


Figure 8 : Schéma des différents constituants d'un spectromètre de masse.

## **2.a. Les sources d'ionisation**

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

### ***2.a.1). L'ionisation Electrospray***

L'analyse par spectrométrie de masse à source électrospray de molécules de poids moléculaires élevés fût pour la première fois suggérée dans les années 1960 par Dole [Dole et al., 1968]. Cependant, ce n'est que quelques années plus tard, lorsque Fenn reprit ces travaux que la source ESI montra tout son intérêt pour l'étude de protéines, notamment en mettant en évidence les états de charges multiples des molécules. Ainsi, ces sources peuvent être utilisées avec des analyseurs de gammes de  $m/z$  limitées et permettent ainsi l'analyse de molécules de haut poids moléculaires [Yamashita & Fenn, 1984]. Enfin, la source électrospray permet de générer des ions en phase gazeuse à partir d'analytes qui se trouvent en solution et à pression atmosphérique. Cette technique a donc pu être naturellement couplée à la chromatographie liquide (LC-MS) [Whitehouse et al., 1985].

Le phénomène ESI se produit à pression atmosphérique, sous l'effet d'un fort champ électrique. Le mécanisme macroscopique de l'ESI est aujourd'hui bien décrit : il assure le passage des ions présents en solution vers la phase gazeuse. Cependant, quelques divergences persistent sur la désorption des ions de la solution vers la phase gazeuse (le mécanisme de la charge résiduelle [Dole et al., 1968] et le mécanisme de l'évaporation ionique [Iribarne, 1976]).

Le processus d'ionisation-désorption ESI se divise en 3 étapes majeures (Figure 9):

- i) La production des gouttelettes chargées à partir de l'électrolyte en solution,
- ii) La fission des gouttelettes chargées par explosions coulombiennes,

iii) Le transfert des ions désolvatés en phase gazeuse.

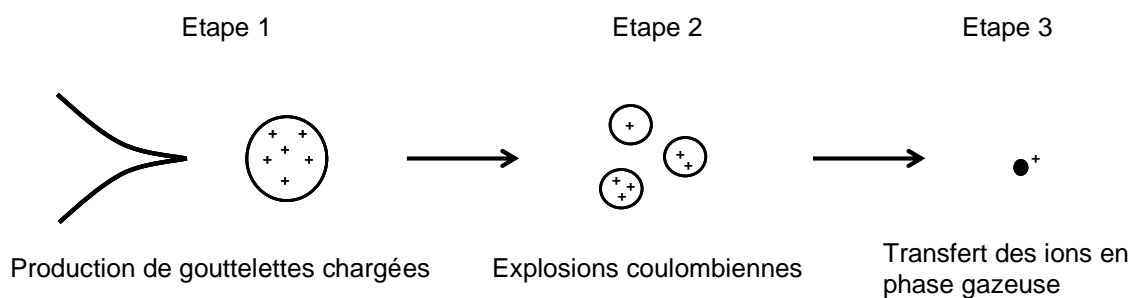


Figure 9 : Représentation schématique de la production des ions par le mode électrospray

En mode d'ionisation ESI, le courant d'ion détecté par le spectromètre de masse dépend de la concentration de l'analyte et non du débit avec lequel l'échantillon est infusé [Ikonomou et al., 1991]. Pour des faibles quantités d'échantillon, il devient donc intéressant de travailler à des débits plus faibles mais sur des solutions plus concentrées afin de gagner en sensibilité. De ces observations sont nées les sources micro-électrospray atteignant des débits inférieurs à 1  $\mu\text{l}/\text{min}$  [Emmett & Caprioli, 1994] et en 1996 les sources nanospray (20  $\text{nL}/\text{min}$ ) [Wilm & Mann, 1996].

La miniaturisation des sources ESI présente de nombreux avantages pour l'analyse : tolérance aux sels, augmentation de la sensibilité et augmentation de la précision de la mesure de masse des protéines en moyennant le signal sur une plus longue durée [Wilm & Mann, 1996]. Les diverses applications du nanospray, notamment pour l'analyse protéomique montrent aujourd'hui l'importance d'un tel dispositif.

### **2.a.2). L'ionisation MALDI**

La technique MALDI a été introduite en 1988 par Karas et Hillenkamp qui décrivaient l'analyse de protéines de poids moléculaires dépassant 10 kDa [Karas & Hillenkamp, 1988].

Cette technique consiste à irradier avec un faisceau laser pulsé de longueur d'onde donnée (généralement dans l'UV) un dépôt cristallin contenant une matrice organique et l'échantillon à analyser. Les matrices les plus couramment utilisées sont l'acide  $\alpha$ -cyano-4-hydroxycinnamique (peptides et protéines), l'acide sinapinique (protéines) et l'acide 2,5-dihydrobenzoïque (phosphopeptides et oligosaccharides). La qualité des spectres dépend de plusieurs facteurs liés à la préparation de l'échantillon comme le choix de la matrice, la méthode de dépôt pour la co-cristallisation matrice-analyte, le pH des solutions, ...

Le processus d'ionisation par MALDI s'effectue sous vide ( $\approx 10^{-7}$  mbar). Les phénomènes ayant lieu ne sont, à l'heure actuelle, ni entièrement ni unanimement établis. Plusieurs modèles ont été proposés mais de nombreuses ambiguïtés subsistent concernant certains aspects du processus.



Cependant, selon les hypothèses les plus communément admises, le processus est constitué de trois événements majeurs (Figure 10) [De Hoffmann & Stroobant, 2003; Knochenmuss & Zenobi, 2003] :

- i) L'excitation des molécules de matrice par les photons du laser,
- ii) L'émission des molécules cibles dans la phase gazeuse,
- iii) L'ionisation des molécules en phase gazeuse.

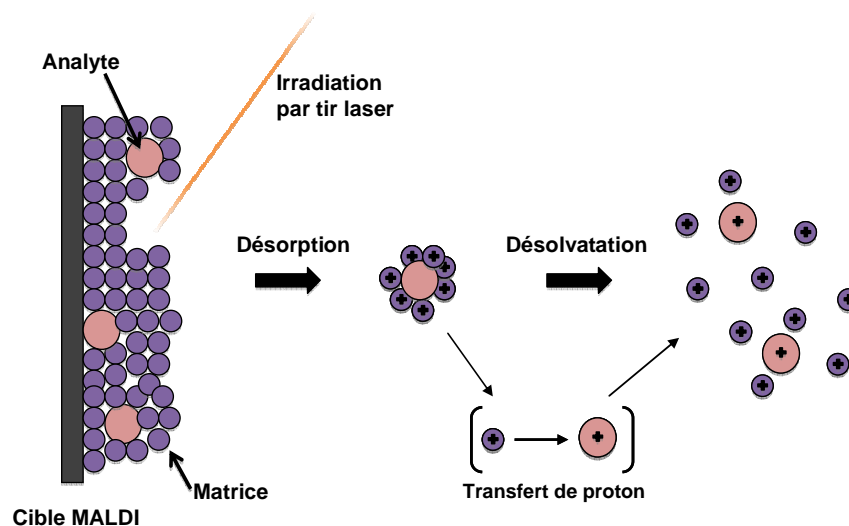


Figure 10 : Mécanisme d'ionisation MALDI [De Hoffmann & Stroobant, 2003]

La technique MALDI est particulièrement tolérante aux sels (par rapport à l'ESI). Cette tolérance est notamment due à la nature solide du dépôt, ainsi qu'à la faible solubilisation des matrices dans l'eau. Le processus MALDI est indépendant des propriétés d'absorption et de la taille des composés à analyser et il permet donc de ce fait la désorption et l'ionisation d'analytes de très haut poids moléculaire (>100000 Da).

## **2.b. Les analyseurs**

L'analyseur a pour rôle de mesurer un rapport masse/charge ( $m/z$ ). Le principe physique d'une telle détermination va dépendre du type d'analyseur utilisé par le spectromètre de masse. On dénombre plusieurs types d'analyseurs tels que les analyseurs quadripolaire (Q), à Temps de Vol (TOF), de type trappe à ion (IT), à résonance cyclotronique (ICR), ainsi qu'à secteur magnétique.

Ces différents analyseurs sont tous compatibles avec le mode d'ionisation électrospray, alors que le couplage le plus fréquemment utilisé avec le mode d'ionisation MALDI est l'analyseur à temps de vol. Dans certains cas, plusieurs analyseurs peuvent être couplés entre eux afin d'obtenir des appareils hybrides [Glish & Burinsky, 2008], tels que des spectromètres de masse de type Q-TOF ou TOF-TOF.

### 2.b.1) Généralités

Les performances d'un analyseur sont définies par quelques paramètres clés : sa résolution (Figure 11), sa précision de mesure de masse, sa gamme de masse (Tableau 2), sa sensibilité et sa capacité à réaliser de la spectrométrie de masse en tandem, MS/MS voire de la MS multiple.

La résolution  $R$  d'un analyseur définit la capacité de distinguer une masse  $M$  par rapport à une masse  $M + \Delta M$ . On admet que deux pics sont résolus quand l'intensité de la vallée entre ces pics égale 10 % de l'intensité du pic le plus faible (on parlera de résolution à 10 %). La résolution pour un pic isolé se calcule en prenant la largeur du pic  $\delta m$  à 10 % de son maximum (Figure 11). On parle maintenant couramment de résolution FWHM (Full Width at Half Maximum) dans ce cas, la largeur  $\delta m$  est prise à 50 % de son maximum.

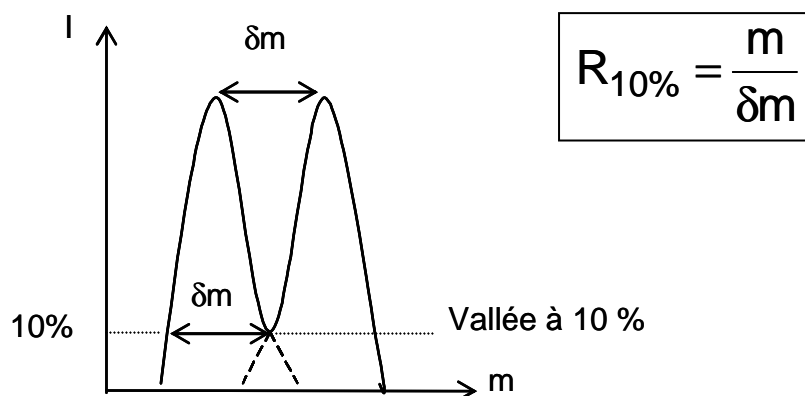


Figure 11 : La résolution est le pouvoir de séparation du spectromètre de masse, elle est définie par la relation de l'encadré qui dans ce cas se situe à 10 % de vallée.

Tableau 2 : Performance des différents analyseurs

Analyseurs	Résolution FWHM	Gamme $m/z$
Quadripôle (Q)	2 000	8 000
Magnétique	20 000	20 000
Temps de vol (TOF)	20 000	500 000
Trappe ionique	5 000	6 000
Cyclotron à résonance des ions (FT-ICR)	1 000 000	4 000 à 100 000

Durant ce travail de thèse, différents instruments ont été utilisés : un MALDI-TOF-TOF (Ultraflex, Bruker Daltonics), des Q-TOF (Q-TOF2 et Synapt, Waters Corps.) et des trappes ioniques (HCTPlus et HCTUltra, Bruker Daltonics). Seuls les principes des analyseurs constituant ces différents instruments sont présentés.

### 2.b.2). Le quadripôle

L'analyseur quadripolaire, couramment utilisé en protéomique, est généralement associé à une source électrospray et peut être couplé à d'autres analyseurs (Q-Q-Q et Q-TOF). Son principe a été décrit par Paul et Steinweger en 1953 [Paul & Steinwedel, 1953]. Il est constitué de 4 barres métalliques parallèles : les barres opposées, connectées électriquement, sont portées au même potentiel  $\Phi_0$  (ou  $-\Phi_0$ ) (Figure 12-A). Le potentiel  $\Phi_0$  comporte une composante sinusoïdale  $V\cos\omega t$ , où  $V$  est l'amplitude de la radio-fréquence, de fréquence  $\omega$  et une composante continue  $U$ .

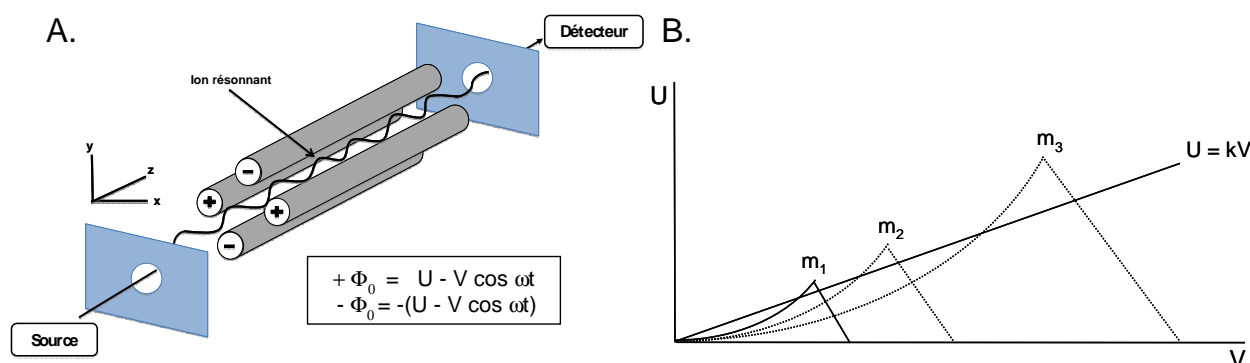


Figure 12 : Schéma d'un analyseur quadripolaire. La trajectoire des ions se fait selon l'axe  $z$  [Campana, 1980].

Le quadripôle est un analyseur à balayage qui utilise la stabilité de la trajectoire des ions entre les barres pour les séparer en fonction de leur rapport  $m/z$ . Les ions accélérés suivant l'axe  $z$  pénètrent entre les barres du quadripôle et conservent leur vitesse suivant cet axe. Cependant, ils sont soumis aux accélérations résultantes des forces dues aux champs électriques suivant  $x$  et  $y$ . Les équations régissant la trajectoire des ions dans le quadripôle ont été établies en 1866 par le physicien Mathieu. La résolution de ces équations permet de déterminer les zones de stabilité des ions d'une masse donnée en fonction de  $U$  et de  $V$  (Figure 12-B). D'après ce diagramme, un balayage maintenant un rapport  $U/V$  constant permettra de transmettre successivement des masses différentes.

### 2.b.3) L'analyseur à temps de Vol (TOF)

Le principe de séparation des ions selon leur rapport  $m/z$  en fonction de leur temps de vol date des années 50 [Wiley & McLaren, 1955] mais ce n'est qu'avec l'apparition de la source MALDI en 1987 [Karas et al., 1987] que l'utilisation de cet analyseur fut enfin reconnue. Classiquement couplé à des sources d'ionisation MALDI (pulsé donc compatible avec une mesure de temps), l'apparition de l'injection orthogonale a permis son couplage avec des sources ESI.

L'analyseur à temps de vol est constitué de deux zones : une zone d'accélération des ions par application de potentiels élevés et une zone libre de champ dans lesquelles règne un vide poussé de l'ordre de  $2 \cdot 10^{-7}$  mbar (Figure 13). Les ions accélérés par un champ électrique  $E$  vont pénétrer dans le tube de vol avec une certaine énergie cinétique  $E_c$  donc une certaine vitesse qui dépend du nombre de charges  $z$  de l'ion ainsi que de sa masse  $m$ . Par exemple, pour un même état de charge, un ion de

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

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

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

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

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

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

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

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

Toutefois, avec le développement de l'extraction retardée et du réflecteur électrostatique, la résolution et la sensibilité du TOF ont pu être considérablement améliorées. Nous décrirons brièvement leurs principes qui sont également illustré dans la figure 13 :

- L'extraction retardée (Delayed Extraction DE) consiste à ajouter une lentille supplémentaire avant l'analyseur, la lentille DE. Cette lentille sur laquelle on applique un potentiel légèrement plus élevé que celui présent dans la source va freiner les ions avant qu'ils n'entrent dans le TOF. Ceci permet de refocaliser les ions et de leur faire prendre un départ simultané dans le tube de vol [Vestal et al., 1995].

- Le réflecteur électrostatique (Reflectron) consiste à ajouter une série de lentilles à l'extrémité du tube de vol qui vont agir comme un miroir électrostatique. En appliquant sur ces lentilles des potentiels croissants, les ions vont être progressivement freinés jusqu'à faire demi-tour pour voyager vers le détecteur. La distance à parcourir étant doublée, les ions seront mieux séparés. De même, on va pouvoir refocaliser les ions de mêmes  $m/z$  qui lors de leur entrée dans le tube de vol n'ont pas forcément la même énergie cinétique. En augmentant le temps de vol, on diminue cet écart entre deux mêmes ions. Cela a pour effet d'augmenter considérablement la résolution [Mamyrin et al., 1973].

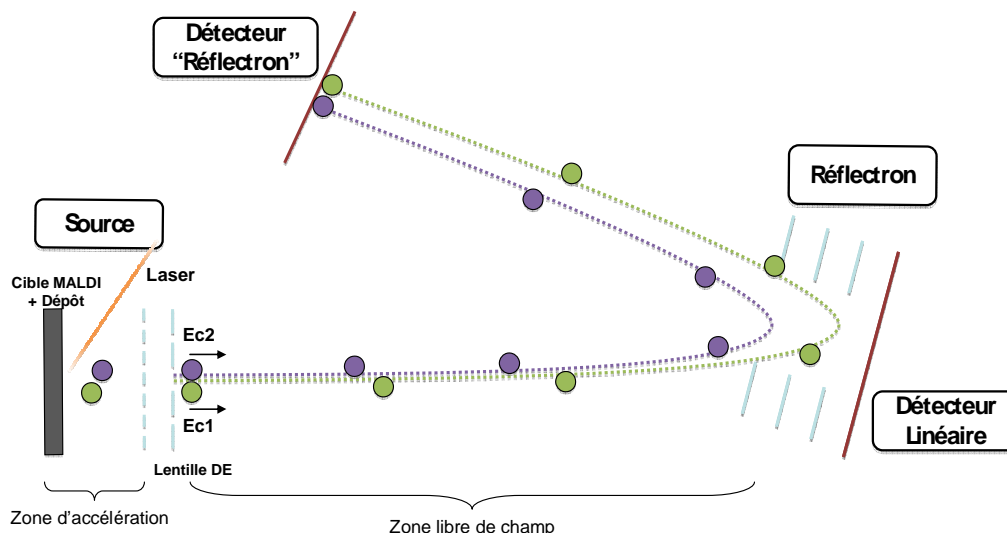


Figure 13 : Schéma du MALDI-TOF en mode réflectron comportant un système d'extraction retardé (DE). Deux ions de même rapport  $m/z$  et d'énergie cinétique différente ( $E_{c1} > E_{c2}$ ) pénètrent plus ou moins profondément dans le réflectron, sont réfléchis et arrivent simultanément au détecteur.

#### 2.b.4) L'analyseur Trappe ionique

Dans les années 50, Paul et Steinwedel ont été les premiers à décrire des analyseurs de type trappe ionique [Paul & Steinwedel, 1953]. Ce travail a été reconnu par un prix Nobel décerné en 1989 à Wolfgang Paul [Paul, 1990]. La trappe ionique est constituée de 3 électrodes, une électrode annulaire en forme de diabol et deux électrodes chapeaux quasi-hyperboliques. Les ions entrent et sortent de la trappe par des orifices situés dans les électrodes chapeaux (Figure 14). L'électrode chapeau d'entrée des ions contient un orifice unique tandis que l'électrode chapeau de sortie des ions présente plusieurs petits trous par lesquels les ions vont pouvoir sortir pour atteindre le détecteur. A la différence des analyseurs décrits précédemment, les ions ne traversent pas l'analyseur mais restent confiné à l'intérieure de cette trappe qui a un volume d'environ  $1\text{cm}^3$ . Il règne au sein de cette cellule une pression résiduelle d'hélium de l'ordre de  $5 \cdot 10^{-5}$  mbar qui va permettre d'augmenter l'efficacité de piégeage des ions et d'assurer le rôle de gaz de collision lors de l'excitation des ions en vue de la fragmentation.

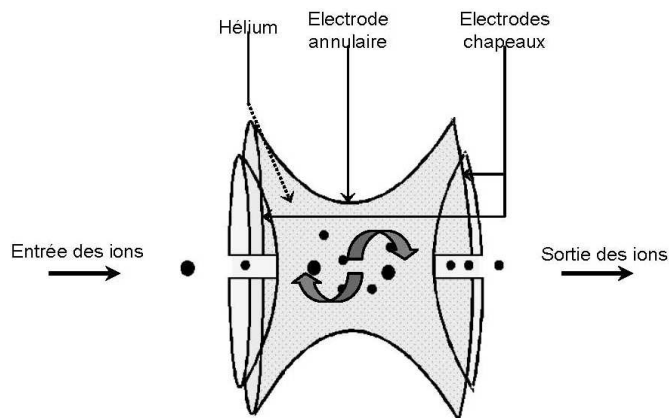


Figure 14 : Représentation schématique d'une trappe ionique.

Comme dans le cas du quadripôle, lorsque les ions pénètrent dans la trappe, ils vont être soumis à un champ quadripolaire généré en appliquant une radiofréquence  $V \cos \omega t$  sur l'électrode annulaire qui va permettre de les piéger. Ce champ quadripolaire peut être assimilé à un puits de potentiel de largeur  $2z_0$  au fond duquel les ions sont piégés. La profondeur du puits de potentiel est définie par l'équation suivante :

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

$D_z$  : profondeur du puits de potentiel  
 $V$  : amplitude de radiofréquence  
 $\omega_{RF}$  : fréquence angulaire

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

La profondeur du puits de potentiel est directement liée à l'amplitude de la radiofréquence  $V$ . Plus celle-ci va augmenter, plus le puits de potentiel sera profond et meilleur sera le piégeage des ions. De même, la profondeur du puits étant inversement proportionnelle au rapport  $m/z$  d'un ion, les ions de  $m/z$  élevés seront mieux piégés que les ions de  $m/z$  plus faibles. L'amplitude de la radiofréquence  $V$  a donc une influence directe sur la gamme de  $m/z$  analysée. En effet, pour une valeur de  $V$  fixe, certains ions de faibles  $m/z$  auront une fréquence trop élevée pour qu'ils soient piégés, on appelle cette limite le cut-off. Leur trajectoire tridimensionnelle résulte également des équations de Mathieu et suit une courbe de Lissajou [De Hoffmann et al., 1994].

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

- Ejection simple : celle-ci consiste à effectuer un balayage croissant de la radiofréquence afin d'éjecter les ions par ordre croissant de  $m/z$ . Or étant limité par la valeur de la tension applicable (20kV), il y aura des ions de  $m/z$  élevés pour lesquels l'amplitude maximale de la tension alternative ne sera pas suffisante pour les éjecter. Cela a pour effet de limiter la gamme de  $m/z$  analysable.

- Ejection résonante [March, 1997] : afin d'augmenter la gamme de  $m/z$  analysable, on va mettre les ions en résonance pour forcer leur éjection. Pour cela, on transfère de l'énergie cinétique aux ions suivant l'axe  $z$  en appliquant une tension alternative sur les électrodes chapeaux. Les ions qui étaient alors confinés au centre de la trappe verront leur mouvement oscillatoire périodique augmenter et seront de ce fait éjectés. De même chaque ion ayant une fréquence d'oscillation propre, il sera possible de mettre en résonance chacun d'eux individuellement. Enfin, l'éjection résonante permet également d'augmenter la résolution et la précision de masse. En effet, lorsque la quantité d'ions accumulés est trop importante, le champ quadripolaire est perturbé entraînant une éjection retardée des ions, on appelle cela les phénomènes d'espace-charge. Ces phénomènes se traduisent par une baisse de la résolution et de la précision de mesure de l'instrument.

La trappe permettant de contrôler les trajectoires des ions et de ce fait de les éjecter sélectivement, cet analyseur peut être utilisé pour fragmenter des ions soit en deux étapes MS/MS ou

soit encore en de multiples étapes  $MS^3$ ,  $MS^4$ , ...,  $MS^n$ . En effet, il est possible d'isoler sélectivement un ion dans la trappe puis de le fragmenter. Pour cela, il faudra éjecter les ions de  $m/z$  inférieurs et de  $m/z$  supérieurs à l'ion de  $m/z$  sélectionné. On procède de la manière suivante :

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

- Ejection des ions de  $m/z > m/z$  de l'ion précurseur : on effectue un balayage en fréquence en partant d'une fréquence proche de la fréquence de résonance de l'ion précurseur. Afin d'affiner l'isolation de l'ion précurseur, on effectue des balayages en fréquences proches de la fréquence de résonance de celui-ci.

La fragmentation est provoquée par l'ajustement de la fréquence de la tension alternative sur l'électrode chapeau de manière à ce qu'elle coïncide avec la fréquence propre de résonance des ions précurseurs : cette amplitude est réglée afin d'éviter l'éjection des ions de la trappe (l'amplitude de fragmentation est plus faible que l'amplitude d'éjection). Ainsi, dopés en énergie cinétique, les ions précurseurs entrent en collision avec les molécules d'hélium. Au fur et à mesure des collisions, leur énergie cinétique est convertie en énergie interne vibrationnelle et l'ion parent finit par se fragmenter. L'énergie de fragmentation est exclusivement transmise aux ions parents. Ainsi, les ions fils issus d'une fragmentation ne peuvent donc, théoriquement, se fragmenter à nouveau pour donner des ions petits-fils, l'énergie qui est transmise aux ions parents n'étant délivrée qu'à leur seule fréquence de résonance.

Par ailleurs, un avantage de l'analyseur à trappe ionique est la possibilité de faire de la  $MS^n$ . En effet, le processus à plusieurs étapes, piégeage-isolation-fragmentation, peut être répété plusieurs fois et ceci s'avère très utile pour l'élucidation structurale des composés analysés.

### **2.b.5). Des analyseurs en tandem : l'exemple du Q-TOF**

Cet appareil est constitué de l'assemblage de deux analyseurs, un premier analyseur quadripolaire suivi d'un tube de vol, les deux étant séparés par une cellule de collision. Dans cette configuration le quadripôle peut fonctionner soit comme analyseur, soit comme guide d'ions ( $U=0$ ).

Lorsque le quadripôle fonctionne comme guide d'ion ( $U=0$ ), les ions sont analysés par le TOF, le quadripôle fonctionne alors en mode "RF-only". Cependant, la transmission des ions dans le quadripôle sera dépendante de l'amplitude de la radiofréquence : la fonction "MS profile" permet de modifier la valeur de l'amplitude de la radiofréquence au cours d'une analyse de façon à laisser passer successivement différentes gammes de rapports  $m/z$ .

Lorsque le quadripôle fonctionne comme analyseur, un ion peut être sélectionné dans le quadripôle, transmis vers la cellule de collision dans laquelle il sera fragmenté puis les fragments générés seront analysés dans le second analyseur (le TOF). On parle alors de spectromètre de masse en tandem permettant de faire de la  $MS/MS$  sur des ions sélectionnés.

Dans cette configuration, le Q-TOF est particulièrement adapté pour la réalisation d'analyses protéomiques en couplage nanoLC-MS/MS : durant ces analyses, des cycles MS suivis de une ou plusieurs MS/MS sur les peptides sélectionnés permettront d'obtenir des informations sur les peptides élués de la colonne chromatographique. C'est dans cette configuration que le Q-TOF2 (Waters Corps.) a été utilisé durant ce travail (Figure 15).

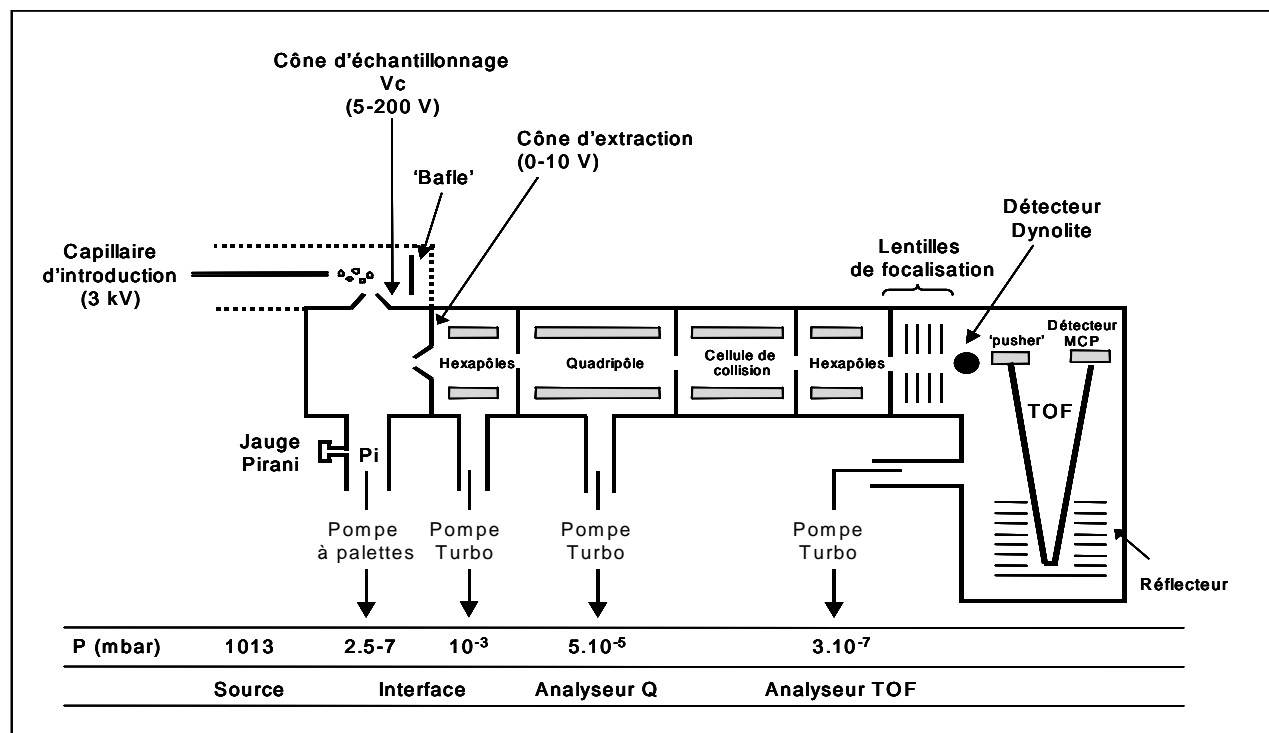


Figure 15 : Représentation schématique du Q-TOF2 (Waters Corps.)

### 3. Les modes de fragmentation

#### 3.a. La méthode de fragmentation CID

Le transfert des ions en phase gazeuse par ESI n'est pas un processus hautement énergétique : les ions entrant dans le spectromètre de masse ont donc des énergies internes faibles [Gaskell, 1997]. On peut distinguer deux types de fragmentations par CID (Collision Induced Dissociation), les fragmentations basse énergie et les fragmentations haute énergie. Les spectres MS/MS obtenus sur des analyseurs de type Q-TOF ou trappe ionique mettent en jeu des énergies de fragmentation de quelques eV, les collisions sont donc de basse énergie.

L'énergie requise pour la fragmentation des peptides est apportée par activation suite à des collisions avec des molécules de gaz. L'addition d'énergie par activation collisionnelle altère la population initiale des formes protonées du peptide et entraîne une augmentation de la population des formes ayant des énergies plus élevées que les formes les plus stables : dans ces molécules excitées, le proton est délocalisé à différents endroits sur le squelette peptidique. C'est la présence de ce proton qui va initier la fragmentation au niveau des liaisons peptidiques pour former des ions b et y



: on dit que ces fragmentations basses énergies sont dirigées par la charge (“charge-directed fragmentation”) et suivent le modèle du “proton mobile” décrit en 1996 par Dongré *et al.* [Dongré et al., 1996].

A côté de cela, la spectrométrie de masse en tandem obtenue sur des analyseurs de type TOF/TOF ou magnétique fait intervenir des énergies de fragmentation de quelques keV et est appelée collision à haute énergie. Les fragmentations des chaînes latérales générant des ions des séries d, v et w n’apparaissent que lors de fragmentations à haute énergie [De Hoffmann et al., 1994].

### **3.b. La méthode de fragmentation ECD/ETD**

La fragmentation des peptides est réalisée par apport d’électron. Celle-ci peut se faire par capture d’électrons par les peptides ionisés traversés par un flux d’électrons (fragmentation ECD). Cette technique est réalisée dans les instruments de type FT-ICR (Fournier Transform – Ion Cyclotron Resonance)[Zubarev et al., 1998]. L’instrumentation nécessaire à la mise en œuvre de cette fragmentation possède des contraintes techniques importantes (vide poussé, champ statique..). La fragmentation peut aussi être réalisée par transfert d’électrons (fragmentation ETD). Des radicaux anions sont mis en contact avec les peptides ionisés et le transfert de l’électron de l’anion avec un peptide ionisé permet la fragmentation [Syka et al., 2004]. Elle peut être réalisée dans des trappes ioniques, spectromètres de masse plus présents dans les laboratoires.

Ce processus de fragmentation est dit “non ergodique”, i.e. il ne suit pas le chemin énergétique préférentiel pour la dissociation des molécules. Il n’y a pas de redistribution d’énergie sur le peptide, la fragmentation est donc “statistique” et ne dépend pas de la séquence en acides aminés (contrairement à la fragmentation CID). Les ions résultants de cette fragmentation sont des ions c et z.



## **Conclusion**

L'analyse protéomique consiste à identifier de façon fine les protéines pour une meilleure compréhension des processus biologiques. Il est nécessaire de développer une méthodologie protéomique pour répondre à ces questions.

Pour choisir et définir la stratégie et les instruments les plus adaptées pour répondre à une question biologique, il faut prendre en compte plusieurs paramètres :

La définition précise de la question posée (identification, caractérisation...)

Le type d'échantillons (solubilisation, complexité, stabilité...).

Le type de séparation (au niveau des protéines ou des peptides)

Le spectromètre adapté (analyse MS ou analyse MS/MS)

L'interprétation des données (choix de la banque, choix des critères et type de validations...)



## **Références bibliographiques**

### **A.**

- Aebersold, R. and Goodlett, D. R. (2001). "Mass spectrometry in proteomics." Chem Rev **101**(2): 269-95.
- Anderson, N. L. and Anderson, N. G. (1998). "Proteome and proteomics: new technologies, new concepts, and new words." Electrophoresis **19**(11): 1853-61.
- Apweiler, R., Bairoch, A. and Wu, C. H. (2004). "Protein sequence databases." Curr Opin Chem Biol **8**(1): 76-80.

### **B.**

- Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. and Yeh, L. S. (2005). "The Universal Protein Resource (UniProt)." Nucleic Acids Res **33**(Database issue): D154-9.
- Biemann, K. (1990). "Appendix 5. Nomenclature for peptide fragment ions (positive ions)." Methods Enzymol **193**: 886-7.
- Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. and Postel, W. (1982). "Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications." J Biochem Biophys Methods **6**(4): 317-39.
- Blueggel, M., Chamrad, D. and Meyer, H. E. (2004). "Bioinformatics in proteomics." Curr Pharm Biotechnol **5**(1): 79-88.
- Bradshaw, R. A., Burlingame, A. L., Carr, S. and Aebersold, R. (2006). "Reporting protein identification data: the next generation of guidelines." Mol Cell Proteomics **5**(5): 787-8.

### **C.**

- Campana, J. E. (1980). "Elementary theory of the quadrupole mass filter." Int J Mass Spectrom Ion Proc **33**: 101-117.
- Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C., Dondi, F., Pastorino, F., Ponzoni, M. and Righetti, P. G. (2005). "Spot overlapping in two-dimensional maps: a serious problem ignored for much too long." Proteomics **5**(9): 2385-95.
- Carapito, C. (2006) Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique Thèse de l'Université Louis Pasteur
- Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K. and Nesvizhskii, A. (2004). "The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data." Mol Cell Proteomics **3**(6): 531-3.

Clauser, K. R., Baker, P. and Burlingame, A. L. (1999). "Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching." Anal Chem **71**(14): 2871-82.

Creasy, D. M. and Cottrell, J. S. (2002). "Error tolerant searching of uninterpreted tandem mass spectrometry data." Proteomics **2**(10): 1426-34.

#### **D.**

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

De Hoffmann, E. and Stroobant, V. (2003). Mass spectrometry: Principles and applications.

Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E., van Dorsselaer, A. and Huber, C. G. (2009). "Repeatability of peptide identifications in shotgun proteome analysis employing off-line two-dimensional chromatographic separations and ion-trap MS." J Sep Sci **32**(8): 1156-64.

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

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

#### **E.**

Elias, J. E. and Gygi, S. P. (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Methods **4**(3): 207-14.

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

#### **F.**

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

#### **G.**

Gaskell, S. J. (1997). "Electrospray: Principles and Practice." Journal of Mass Spectrometry **32**: 677-688.

Gevaert, K. and Vandekerckhove, J. (2000). "Protein identification methods in proteomics." Electrophoresis **21**(6): 1145-54.

Giddings, J. C. (1984). "Two-dimensional separations: concept and promise." Anal Chem **56**(12): 1258A-1260A, 1262A, 1264A passim.

Glish, G. L. and Burinsky, D. J. (2008)."Hybrid mass spectrometers for tandem mass spectrometry." J Am Soc Mass Spectrom **19**(2): 161-72.

Gonzalez-Begne, M., Lu, B., Han, X., Hagen, F. K., Hand, A. R., Melvin, J. E. and Yates, J. R. (2009)."Proteomic Analysis of Human Parotid Gland Exosomes by Multidimensional Protein Identification Technology (MudPIT)." J Proteome Res.

Gorg, A., Weiss, W. and Dunn, M. J. (2004)."Current two-dimensional electrophoresis technology for proteomics." Proteomics **4**(12): 3665-85.

Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. and Aebersold, R. (2000)."Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology." Proc Natl Acad Sci U S A **97**(17): 9390-5.

## **H.**

Habermann, B., Oegema, J., Sunyaev, S. and Shevchenko, A. (2004)."The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches." Mol Cell Proteomics **3**(3): 238-49.

Hardouin, J., Duchateau, M., Joubert-Caron, R. and Caron, M. (2006)."Usefulness of an integrated microfluidic device (HPLC-Chip-MS) to enhance confidence in protein identification by proteomics." Rapid Commun Mass Spectrom **20**(21): 3236-44.

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

Horn, D. M., Peters, E. C., Klock, H., Meyers, A. and Brock, A. (2004)."Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting." Int J Mass Spectrom **238**: 189-186.

Horth, P., Miller, C. A., Preckel, T. and Wenz, C. (2006)."Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis." Mol Cell Proteomics **5**(10): 1968-74.

Hoving, S., Voshol, H. and van Oostrum, J. (2000)."Towards high performance two-dimensional gel electrophoresis using ultrazoom gels." Electrophoresis **21**(13): 2617-21.

Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E. and Engelhard, V. H. (1992)."Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry." Science **255**(5049): 1261-3.

## **I.**

Ikonomou, M. G., Blades, A. T. and Kebarle, P. (1990)."Investigations of the electrospray interface for liquid chromatography/mass spectrometry " Anal Chem **62**(9): 957-967.

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

Iribarne, J. V. a. T., B.A. (1976)."On the evaporation of small ions from charged droplets." The Journal of Chemical Physics **64**(6): 2287-2294.

## **J.**

James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993)."Protein identification by mass profile fingerprinting." Biochem Biophys Res Commun **195**(1): 58-64.

Jones, A. R., Siepen, J. A., Hubbard, S. J. and Paton, N. W. (2009)."Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines." Proteomics **9**(5): 1220-9.

## **K.**

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. (2005)."An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis." Proteomics **5**(13): 3475-90.

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

Karas, M. and Hillenkamp, F. (1988)."Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Anal Chem **60**(20): 2299-301.

Kay, R. G., Gregory, B., Grace, P. B. and Pleasance, S. (2007)."The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum." Rapid Commun Mass Spectrom **21**(16): 2585-93.

Keller, A., Eng, J., Zhang, N., Li, X.-j. and Aebersold, R. (2005)."A uniform proteomics MS/MS analysis platform utilizing open XML file formats." Molecular Systems Biology **1**: 2005.0017.

Keller, A., Nesvizhskii, A. I., Kolker, E. and Aebersold, R. (2002)."Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search." Anal Chem **74**(20): 5383-92.

Kenrick, K. G. and Margolis, J. (1970)."Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique." Anal Biochem **33**(1): 204-7.

Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E. and Apweiler, R. (2004)."The International Protein Index: an integrated database for proteomics experiments." Proteomics **4**(7): 1985-8.

Knochenmuss, R. and Zenobi, R. (2003)."MALDI ionization: the role of in-plume processes." Chem Rev **103**(2): 441-52.

## **L.**

Laemmli, U. K. (1970)."Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.



Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. and Yates, J. R., 3rd (1999). "Direct analysis of protein complexes using mass spectrometry." Nat Biotechnol **17**(7): 676-82.

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

## **M.**

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. and Lajoie, G. (2003). "PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry." Rapid Commun Mass Spectrom **17**(20): 2337-42.

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

Mann, M., Hojrup, P. and Roepstorff, P. (1993). "Use of mass spectrometric molecular weight information to identify proteins in sequence databases." Biol Mass Spectrom **22**(6): 338-45.

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

Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J., Girault, H. H. and Rossier, J. S. (2003). "Protein fractionation in a multicompartiment device using Off-Gel isoelectric focusing." Electrophoresis **24**(1-2): 3-11.

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

Motoyama, A., Venable, J. D., Ruse, C. I. and Yates, J. R., 3rd (2006). "Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples." Anal Chem **78**(14): 5109-18.

## **N.**

Nagele, E., Vollmer, M. and Horth, P. (2004). "Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome." J Biomol Tech **15**(2): 134-43.

Nesvizhskii, A. I. and Aebersold, R. (2005). "Interpretation of shotgun proteomic data: the protein inference problem." Mol Cell Proteomics **4**(10): 1419-40.

Nesvizhskii, A. I., Vitek, O. and Aebersold, R. (2007). "Analysis and validation of proteomic data generated by tandem mass spectrometry." Nat Methods **4**(10): 787-97.

Neuhoff, V., Arold, N., Taube, D. and Ehrhardt, W. (1988). "Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250." Electrophoresis **9**(6): 255-62.

## **O.**

O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." J Biol Chem **250**(10): 4007-21.

Ong, S. E. and Pandey, A. (2001). "An evaluation of the use of two-dimensional gel electrophoresis in proteomics." Biomol Eng **18**(5): 195-205.

## **P.**

Palagi, P. M., Hernandez, P., Walther, D. and Appel, R. D. (2006). "Proteome informatics I: bioinformatics tools for processing experimental data." Proteomics **6**(20): 5435-44.

Pappin, D. J., Hojrup, P. and Bleasby, A. J. (1993). "Rapid identification of proteins by peptide-mass fingerprinting." Curr Biol **3**(6): 327-32.

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

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

Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-67.

## **R.**

Rabilloud, T. (2002). "Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains." Proteomics **2**(1): 3-10.

Rabilloud, T., Strub, J. M., Luche, S., van Dorsselaer, A. and Lunardi, J. (2001). "A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels." Proteomics **1**(5): 699-704.

Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E. and Rabilloud, T. (2004). "About the mechanism of interference of silver staining with peptide mass spectrometry." Proteomics **4**(4): 909-16.

Riederer, B. M. (2008). "Non-covalent and covalent protein labeling in two-dimensional gel electrophoresis." J Proteomics **71**(2): 231-44.

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

Ros, A., Faupel, M., Mees, H., Oostrum, J., Ferrigno, R., Reymond, F., Michel, P., Rossier, J. S. and Girault, H. H. (2002). "Protein purification by Off-Gel electrophoresis." Proteomics **2**(2): 151-6.

## S.

- Santoni, V., Molloy, M. and Rabilloud, T. (2000)."Membrane proteins and proteomics: un amour impossible?" Electrophoresis **21**(6): 1054-70.
- Searle, B. C., Turner, M. and Nesvizhskii, A. I. (2008)."Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies." J Proteome Res **7**(1): 245-53.
- Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W. and Standing, K. G. (2001)."Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching." Anal Chem **73**(9): 1917-26.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996)."Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels." Anal Chem **68**(5): 850-8.
- Steinberg, T. H., Agnew, B. J., Gee, K. R., Leung, W. Y., Goodman, T., Schulenberg, B., Hendrickson, J., Beechem, J. M., Haugland, R. P. and Patton, W. F. (2003)."Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology." Proteomics **3**(7): 1128-44.
- Steinberg, T. H., Pretty On Top, K., Berggren, K. N., Kemper, C., Jones, L., Diwu, Z., Haugland, R. P. and Patton, W. F. (2001)."Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots." Proteomics **1**(7): 841-55.
- Swanson, S. K. and Walshburn, M. P. (2005)."The continuing evolution of shotgun proteomics." Drug Discov Today **10**: 719-725.
- Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F. (2004)."Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." Proc Natl Acad Sci U S A **101**(26): 9528-33.

## T.

- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y. and Yoshida, T. (1988)."Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **2**: 151-153.
- Toll, H., Oberacher, H., Swart, R. and Huber, C. G. (2005)."Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH." J Chromatogr A **1079**(1-2): 274-86.
- Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I. and Davison, M. (2001)."Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology." Proteomics **1**(3): 377-96.
- Tyers, M. and Mann, M. (2003)."From genomics to proteomics." Nature **422**(6928): 193-7.

## V.

- Vestal, M. C., Juhasz, P. and Martin, S. A. (1995)."Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **9**(11): 1044-1050.

## **W.**

Washburn, M. P., Wolters, D. and Yates, J. R., 3rd (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nat Biotechnol **19**(3): 242-7.

Whitehouse, C. M., Dreyer, R. N., Yamashita, M. and Fenn, J. B. (1985). "Electrospray interface for liquid chromatographs and mass spectrometers." Anal Chem **57**(3): 675-9.

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

Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W. and Dunn, M. J. (2006). "Guidelines for the next 10 years of proteomics." Proteomics **6**(1): 4-8.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L. and Hochstrasser, D. F. (1996). "From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis." Biotechnology (N Y) **14**(1): 61-5.

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

Wolters, D. A., Washburn, M. P. and Yates, J. R., 3rd (2001). "An automated multidimensional protein identification technology for shotgun proteomics." Anal Chem **73**(23): 5683-90.

## **X.**

Xiang, R., Shi, Y., Dillon, D. A., Negin, B., Horvath, C. and Wilkins, J. A. (2004). "2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474." J Proteome Res **3**(6): 1278-83.

## **Y.**

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

Yates, J. R., 3rd, Speicher, S., Griffin, P. R. and Hunkapiller, T. (1993). "Peptide mass maps: a highly informative approach to protein identification." Anal Biochem **214**(2): 397-408.

Yin, H. and Killeen, K. (2007). "The fundamental aspects and applications of Agilent HPLC-Chip." J Sep Sci **30**(10): 1427-34.

Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M. and van de Goor, T. (2005). "Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip." Anal Chem **77**(2): 527-33.

## **Z.**

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

## **PARTIE I – Identification des protéines impliquées dans des complexes protéine/protéine et ARN/protéine.**

**Chapitre I. Synthèse bibliographique et stratégie développée pour l'analyse des protéines impliquées dans les complexes protéine/protéine et ARN/protéine**

**Chapitre II. Identification de partenaires protéiques de la poly(ADP-ribose) polymérase-2**

**Chapitre III. Identification de partenaires protéiques de complexes ARN/protéine**



## **Chapitre I. Synthèse bibliographique et stratégie développée pour l'analyse des protéines impliquées dans les complexes protéine/protéine et ARN/protéine**

Les processus biologiques sont orchestrés et régulés par les interactions dynamiques entre les protéines ou entre les protéines et d'autres molécules, tel que des ligands, des métaux...[Gingras et al., 2007]. Des fonctions cellulaires importantes telles que la réplication de l'ADN, la transcription et la traduction demandent une action coordonnée d'un grand nombre de protéines, assemblées en complexes multiprotéiques qui diffèrent, selon leur action, en composition et structure. L'identification de ces complexes protéiques, les interactions qui ont lieu entre les protéines et leur évolution au cours du temps et en fonction de leur environnement est centrale dans la compréhension des mécanismes cellulaires [Hakes et al., 2008].

### **1. Synthèse bibliographique**

#### **1.a Purification des complexes protéiques**

Les complexes protéine/protéine ou ARN/protéine doivent être purifiés, afin que seules les protéines étant en interaction avec la protéine ou l'ARN d'intérêt soient co-purifiées et analysées.

##### ***1.a.1) Les complexes protéine/protéine***

Les méthodes d'affinités pour l'étude des complexes protéiques sont basées sur l'immunoprécipitation du complexe [Kocher & Superti-Furga, 2007].

Le complexe protéique est précipité du milieu dans lequel il se trouve (lysate cellulaire...) grâce à un anticorps dirigé contre une protéine connue (généralement la protéine d'intérêt) faisant partie du complexe (Figure 1-a). L'avantage de cette technique est de pouvoir isoler le complexe de façon endogène à partir de diverses sources biologiques (tissus, fluides circulants...). Cependant, cette technique nécessite de disposer d'un anticorps efficace et spécifique de la protéine cible.

Deux méthodes de purification peuvent être utilisées pour des complexes produits *in vitro* (Figure 1-b et c). Le principe est de produire une des protéines appartenant au complexe avec une étiquette (tag) spécifique. Un enrichissement du complexe contenant cette protéine marquée est réalisé à l'aide d'une colonne ayant une forte affinité pour l'étiquette (Figure 1-b). Cette technique peut être réalisée avec de nombreuses étiquettes différentes [Terpe, 2003]. Pour purifier un complexe protéine/protéine de manière plus efficace, Rigaut *et al.* ont développé une méthode de double purification mettant en jeu le marquage d'une protéine avec deux étiquettes (TAP-tag) (Figure 1-c).

Ces deux étiquettes sont le peptide calmodulin-binding (CBP) et la protéine A issus de *Staphylococcus aureus* séparés par un site de clivage [Rigaut et al., 1999].

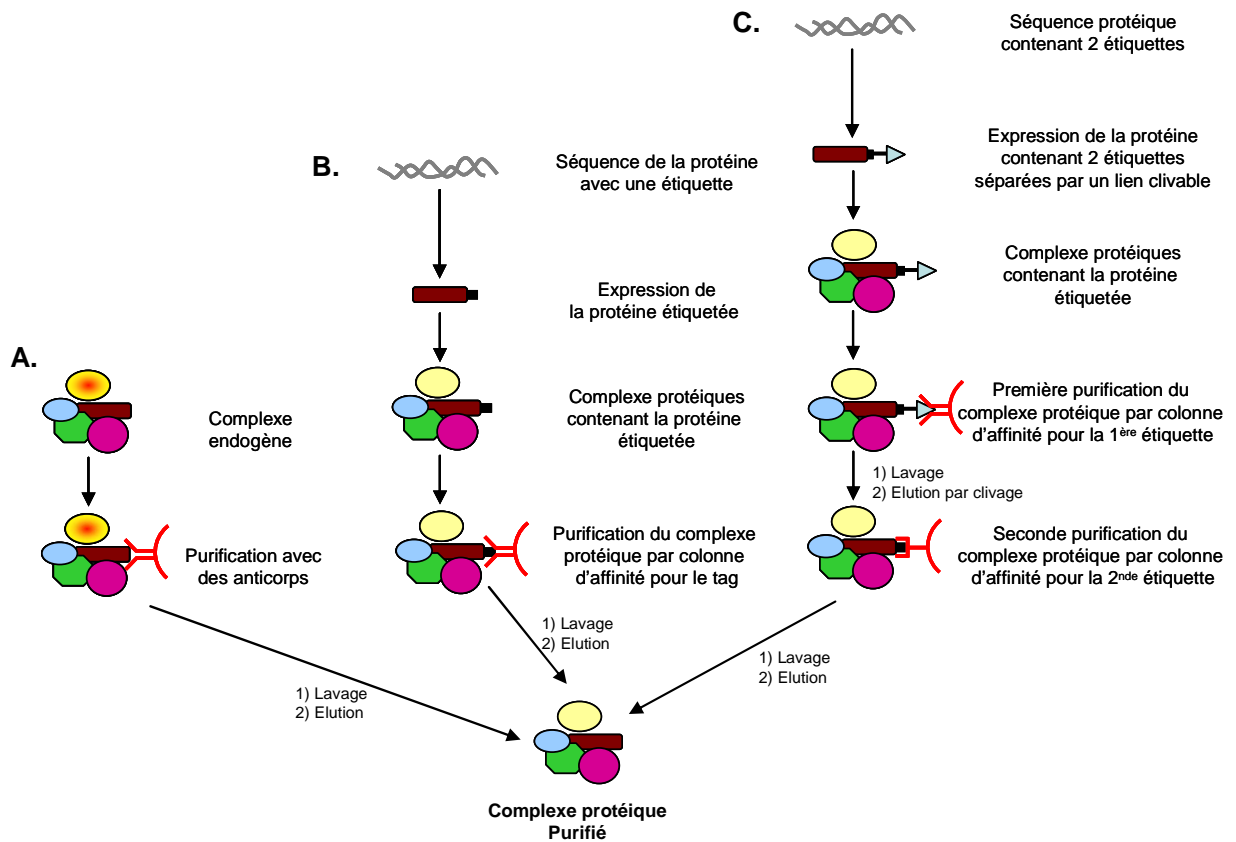


Figure 1 : Méthodes de purifications de complexes protéiques [Kocher & Superti-Furga, 2007].

La purification choisie dépendra de chaque complexe et des ses propriétés. Une purification en une seule étape conservera un grand nombre de candidats, contrairement à la méthode TAP-tag. Par contre, celle-ci peut engendrer la perte de certaines interactions spécifiques, notamment les plus faibles. Un équilibre doit être trouvé entre minimiser les interactions non spécifiques, par la stringence du lavage, et conserver un maximum d'interactions spécifiques faibles.

### 1.a.2) Les complexes ARN/protéine

Pour la purification des complexes ARN/protéine, plusieurs méthodes de purification par affinité existent, qui consistent à fixer l'ARN sur des billes d'une colonne de chromatographie puis d'incuber ce système en présence d'extrait protéique afin de constituer un complexe ARN/protéine (Figure 2). Il existe plusieurs types de billes et de façons de fixer l'ARN. Chacune influencent la formation du complexe ARN/protéine et la présence de contaminants. Ainsi pas moins de 6 techniques de fixation de l'ARN sont présentes dans la littérature:

- L'ARN peut être dérivé puis directement lié covalamment sur des billes d'agarose [Venables et al., 2005; Goina et al., 2008],



- Un autre moyen consiste en l'utilisation de billes de sepharose-oligo(dT) capables de s'hybrider à une queue poly(A) sur l'ARN d'intérêt [Bose et al., 2006],

- Une troisième technique est basée sur l'interaction biotine/streptavidine. Dans ce cas, les billes d'agarose sont dérivées avec la streptavidine et la biotine est greffée sur l'ARN d'intérêt [Chen et al., 2008],

- Sur le même modèle, deux autres techniques utilisant un antibiotique fixé aux billes (sepharose – tobramycine ou sepharose – dihydrostreptomycine). L'ARN d'intérêt est transcrit à un ARN aptamère capable de se lier à l'antibiotique (ARN tobratag ou ARN streptotag) [Bachler et al., 1999; Hartmuth et al., 2002],

- Enfin la technique MS2 emploie une protéine de fusion dont un domaine fixe les billes d'amylose et l'autre fixe une série de tige-boucle qui sont cotranscrit avec l'ARN d'intérêt [Zhou et al., 2002].

Les points critiques de ces techniques sont la capacité des billes et l'élément de jonction (antibiotiques, protéine de fusion...) à ne pas fixer de contaminants et les conditions de maintien du système d'affinité (pH non physiologique, faible concentration saline...). Cela joue donc sur la formation du complexe et sa purification en limitant la stringence des lavages.

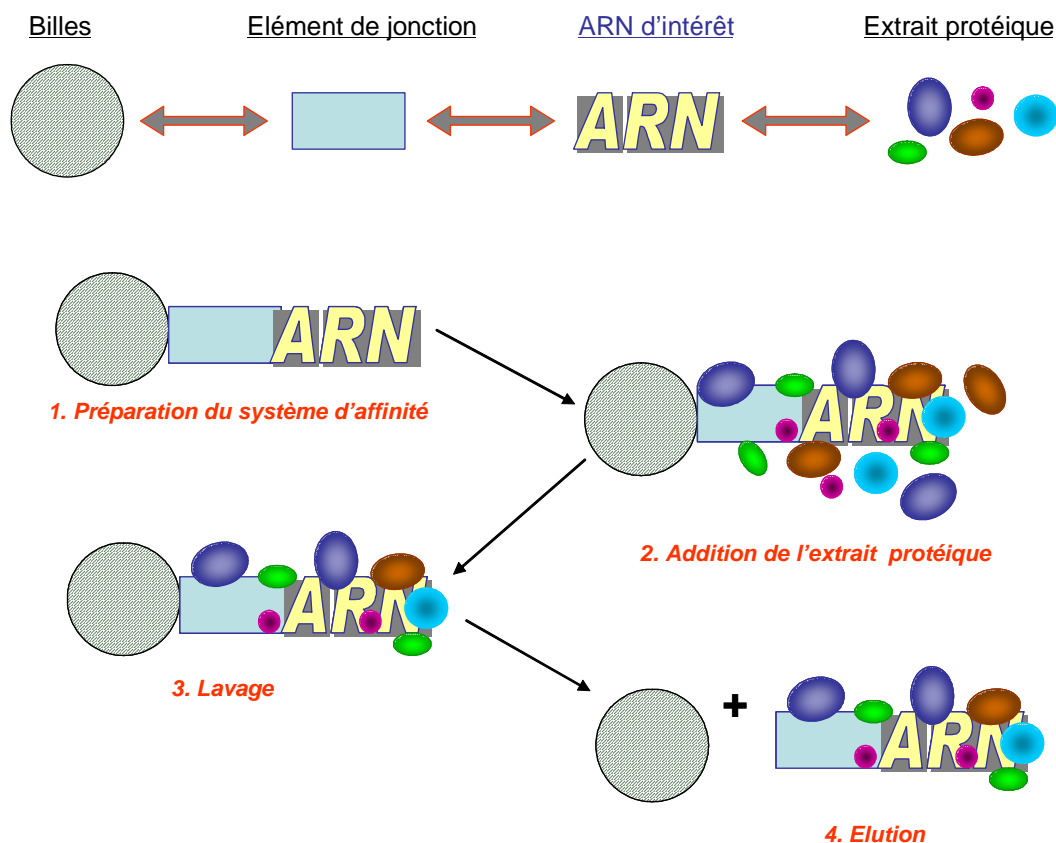


Figure 2 : Représentation schématique du système d'affinité utilisé pour la purification d'un complexe ARN/protéine.

## 1.b. Analyse par spectrométrie de masse des complexes purifiés

Les complexes protéiques sont composés de plusieurs protéines dont l'identification peut être réalisée par spectrométrie de masse. Deux types d'approches sont utilisés (Figure 3) :

- L'analyse indirecte du complexe, c'est à dire l'analyse après séparation des protéines ou des peptides,
- L'analyse directe du complexe intact, c'est à dire en conservant la structure du complexe dans des conditions non-dénaturantes.

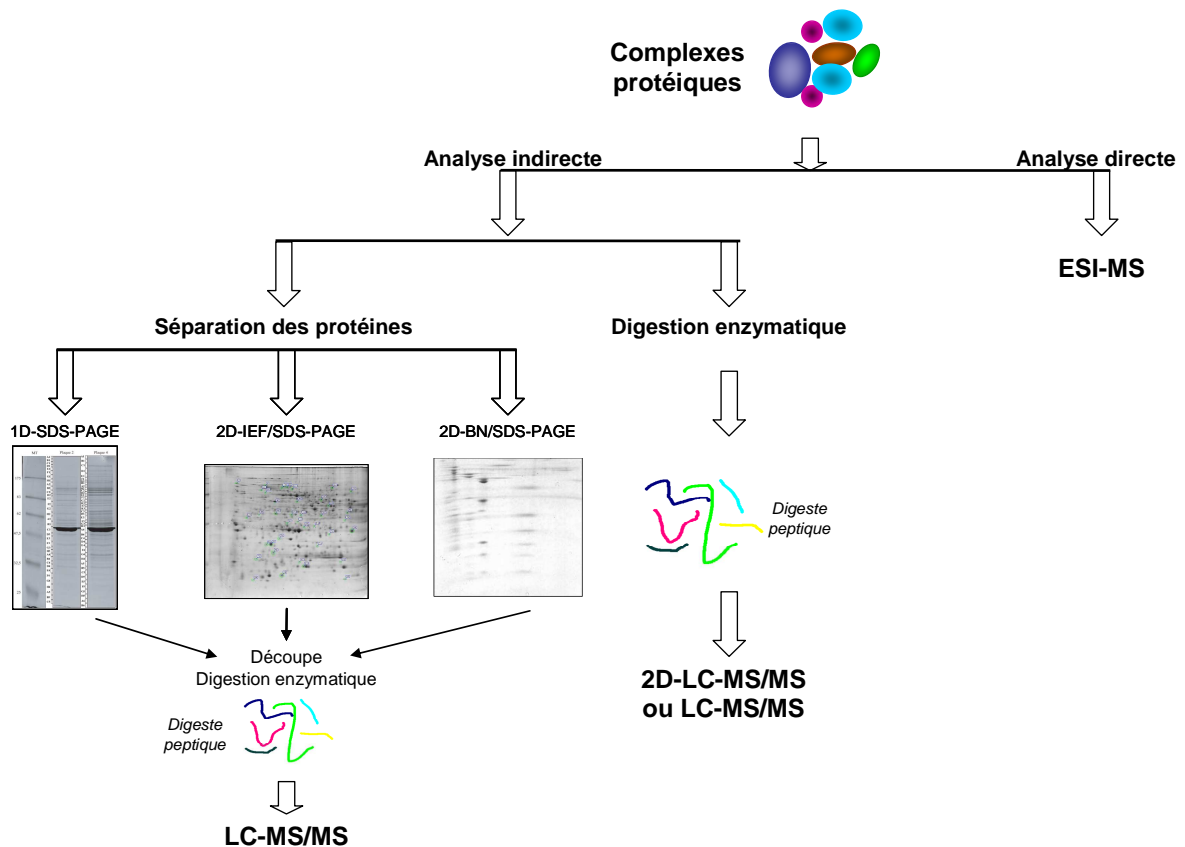


Figure 3 : Stratégies d'analyses pour les complexes protéiques après purification des complexes

L'analyse indirecte des complexes peut se faire après séparation des protéines suivies d'une digestion enzymatique ou après la digestion des complexes entiers et séparation des peptides générés.

La première approche est basée sur la séparation des protéines par gel d'électrophorèse. Les protéines ainsi séparées sont ensuite soumises à une digestion enzymatique avant une analyse par nanoLC-MS/MS. Les protéines peuvent être séparées par des gels d'électrophorèses 1D-SDS-PAGE, 2D-IEF/SDS-PAGE ou 2D BN/SDS-PAGE.

Les gels 1D-SDS-PAGE, 2D-IEF/SDS-PAGE sont des gels qui ne conservent, à aucun moment, l'intégrité du complexe, et qui séparent les protéines selon leur taille ou leur charge et

leur taille, respectivement (voir Partie I – Chap. II – I.1.). Lorsque la complexité du mélange est relativement réduite, de part la purification, les protéines sont séparées par gel d'électrophorèse 1D-SDS-PAGE [Lamond & Mann, 1997]. L'électrophorèse 1D est plus rapide et limite la perte de matériel. Elle est certes moins résolutive que l'électrophorèse 2D mais elle permet de séparer des protéines issues de l'échantillon et du contrôle en parallèle et des protéines membranaires. Cependant lorsqu'il s'agit des complexes très importants, comme l'identification de protéines issues du complexe du spliceosome [Neubauer et al., 1998], la séparation par électrophorèse 2D-IEF/SDS-PAGE peut être utile.

Le gel 2D-BN/SDS-PAGE consiste à séparer les complexes protéiques dans une 1<sup>ère</sup> dimension. L'intégrité des complexes protéiques est conservée durant la séparation, qui se fait en fonction de la charge, la taille et la forme du complexe. Chaque complexe protéique isolé est ensuite dénaturé et ses sous-unités sont séparés en fonction de leur taille dans une seconde dimension SDS-PAGE. Les protéines alignées selon la seconde dimension sont issues du même complexe [Schagger & von Jagow, 1991; Lasserre et al., 2006]. La séparation 2D-BN/SDS-PAGE permet de séparer des complexes membranaires [Fandino et al., 2005] ou des complexomes (ensemble des complexes présents dans une cellule à un moment et dans des conditions données) [Lasserre et al., 2006].

La seconde approche est basée sur la digestion enzymatique de l'ensemble du complexe purifié avant identification des protéines par 2D-LC-MS/MS [Link et al., 1999] (voir Partie I – Chap. I – I.3.). L'avantage de la technique basée sur la séparation des protéines avant la digestion permet d'éliminer des contaminants tels que les tampons issus de la purification du complexe, pouvant gêner l'ionisation des peptides lors de l'analyse. La technique de digestion en solution permet d'avoir la possibilité d'analyser l'ensemble des peptides possibles, mais ceci à la condition que les contaminants issus de la purification permettent d'obtenir une digestion totale. Cette stratégie est très utile pour des protéines pouvant être difficilement séparées par gel d'électrophorèse, tel que les protéines membranaires [Fischer et al., 2006; Lu et al., 2008].

Une fois les protéines analysées par MS/MS, le traitement des données et la recherche dans les banques de données vont permettre d'identifier certaines protéines appartenant au complexe analysé. L'inconvénient majeur de la stratégie de digestion en solution réside dans l'interprétation des données générées par l'analyse directe de mélanges complexes de peptides. En effet, le nombre de peptides générés est considérable et certaines identifications de protéines ne se font qu'avec un peptide. Il peut s'avérer difficile d'attribuer avec certitude un peptide à une protéine notamment dans le cas d'identifications de protéines ayant de fortes homologies de séquences [Nesvizhskii & Aebersold, 2005]. La validation des interprétations et des identifications de protéines doit donc se faire avec la plus grande prudence en vérifiant la qualité des spectres de fragmentation.

L'analyse directe des complexes peut être réalisée par une analyse par spectrométrie de masse supramoléculaire. En effet, si certaines précautions expérimentales et instrumentales sont respectées, le processus électrospray permet de conserver l'intégrité structurale de systèmes non

covalents lors de leur transfert de la solution à la phase gazeuse [Sanglier-Cianferani, 2002]. La mesure de masse de ces édifices permet d'obtenir des informations d'états d'oligomérisation et des stœchiométries d'interaction, en particulier dans le cas de complexes ARN/protéine [Gordiyenko & Robinson, 2008; Sanglier et al., 2008].

Le choix du type de purification dépend des propriétés des complexes recherchés. Plusieurs choix d'approches protéomiques sont possibles, qui permettront d'obtenir des identifications de protéines partenaires. Le choix de l'approche se fait en fonction de la question biologique posée et de la complexité de l'échantillon. Les partenaires protéiques identifiés par approche protéomique sont des candidats à une interaction avec la protéine ou l'ARN cible.

## 2. Stratégie développée pour l'identification des protéines

Pour les études d'identification des complexes protéine/protéine et ARN/protéine, la stratégie que nous avons développée peut être résumée par la figure 4.

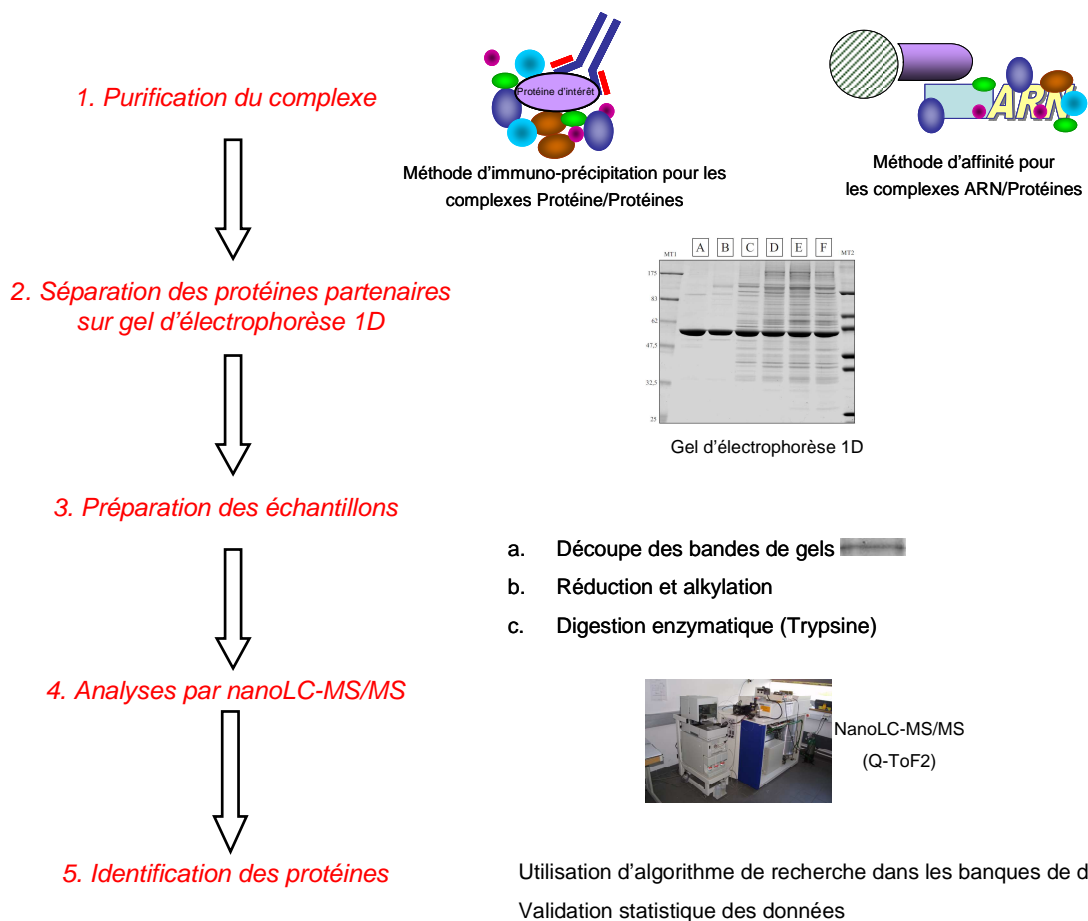


Figure 4 : Stratégie d'analyse protéomique développée pour l'identification de protéines partenaires d'une protéine ou d'un ARN.

## **2.a. Purification du complexe et séparation des protéines.**

Les complexes protéine/protéine ont été purifiés par immuno-précipitation. Pour les complexes ARN/protéine, nos collaborateurs ont utilisé les méthodes d'affinité sépharose – tobramycine et la technique MS2. Un contrôle, où seul l'ARN support est utilisé, est réalisé en parallèle pour chaque série d'expérience, afin de pouvoir identifier les protéines interagissant avec l'ARN support seul.

Les protéines constitutives des complexes sont ensuite séparées par gel 1D. Pour les exemples présentés par la suite, les différentes protéines présentes sur le gel ont été révélées par coloration soit avec le bleu de Coomassie soit la coloration Sypro Ruby®. Chaque piste du gel d'électrophorèse 1D est découpée, soit de façon ciblée en découpant uniquement les bandes les plus intenses ou d'intérêt, soit de façon systématique, tous les 2mm. Le choix de la découpe est fait en fonction du nombre de protéines séparées par le gel. Les découpes sont faites de la même façon pour l'échantillon comme pour le contrôle. La digestion des protéines présentes dans les bandes de gel se fait classiquement avec la trypsine.

## **2.b. Analyse par spectrométrie de masse**

L'analyse des mélanges peptidiques obtenus est réalisée en nanoLC-MS/MS de façon automatique, les analyses pour les études présentées dans cette partie ont été réalisées sur un système CapLC-Q-TOF2 (Waters Corps.). La séparation des peptides est réalisée par chromatographie liquide sur une colonne de phase inverse C18 (Colonne LCPacking C18, 75 mm x 150mm). Les peptides sont introduits dans le spectromètre de masse par une source électrospray et analysés en mode MS/MS. Les trois ions les plus abondants du spectre MS (préférentiellement les doublement et triplement chargés) sont sélectionnés et fragmentés en mode MS/MS. Un ion parent ne peut être sélectionné pour fragmentation que deux fois consécutives, au-delà il est exclu de la sélection durant 30 s, afin de pouvoir fragmenter d'autres ions et obtenir plus d'informations sur le mélange peptidique analysé. Les données brutes issues de l'analyse sont traitées par le logiciel ProteinLynx (Waters Corps.) et mises en forme pour pouvoir être soumises à des algorithmes de recherche.

Les données MS/MS sont soumises à une recherche dans une banque de données protéiques à l'aide du moteur de recherche Mascot. Le choix de la banque protéique et des critères d'acceptation des identifications sont spécifiques à chaque étude. Certaines études ayant été réalisées au début de mon travail de thèse, les interprétations ont été validées manuellement. Par la suite, le logiciel Scaffold (Proteome software, version 01-07-00) nous a permis de valider les identifications de façon automatique. Ce type de stratégie va permettre d'identifier des candidats protéiques pour l'interaction protéine/protéine ou ARN/protéine recherchée. Pour confirmer l'interaction entre la protéine ou l'ARN cible avec un candidat protéique, une validation biologique doit être menée.

Cette stratégie d'identification des protéines impliquées dans des complexes a été utilisée pour de nombreux projets durant ma thèse.

Le projet d'identification des partenaires protéiques spécifiques de la poly(ADP-ribose)polymérase-2 (Parp-2) permet d'illustrer l'application de cette stratégie aux complexes protéine/protéine.

De nombreuses collaborations m'ont permis d'utiliser la stratégie pour des complexes ARN/protéine pour diverses thématiques :

- La Dystrophie Myotonique de type 1 (DM1) est une des dystrophies musculaires les plus fréquentes chez l'adulte, avec une incidence mondiale d'une personne atteinte sur 8000. La DM1 est caractérisée par des répétitions de triplets CUG de l'ARNm du gène de la myotonine protéine kinase (DMPK). Les séquences répétées pourraient induire des modifications de la disponibilité des facteurs impliqués dans la régulation de l'épissage alternatif. L'objectif est donc d'identifier les facteurs susceptibles de se lier aux répétitions CUG.
- La production de protéines de VIH-1 dépend de l'épissage alternatif du transcript viral primaire. La multiplication du VIH-1 est dépendante de l'utilisation des différents sites d'épissage. Le but de ce travail est d'utiliser l'approche protéomique pour la recherche des protéines nucléaires impliquées dans la régulation des sites d'épissage A7 et A3 de l'ARN VIH-1.
- Le virus du sarcome de Rous (RSV pour Rous Sarcomas Virus) est un rétrovirus oncogène simple. L'épissage de son ARN implique un seul site donneur et deux sites accepteurs alternatifs. La partie NRS (Negative Regulator of Splicing) est une séquence ARN impliquée dans l'inhibition de l'épissage de l'ARN viral. L'objectif est d'identifier les facteurs protéiques interagissant avec l'ARN NRS donc les protéines impliquées dans l'inhibition de l'épissage de l'ARN viral.
- Les enfants atteints de progéria (maladie du vieillissement prématuré) sont majoritairement porteurs d'une mutation du gène LMNA (codant pour la lamine A). Cette mutation fait apparaître au niveau de l'ARN pré-messager LMNA un site donneur d'épissage. L'utilisation de ce site induit un événement d'épissage alternatif résultant en une délétion de 50 acides aminés sur la protéine produite la lamine A. L'objectif est l'identification des partenaires d'interaction protéique de l'ARN LMNA.
- La présence de deux chromosomes X dans les cellules de femelles de vertébrés par rapport à l'unique chromosome X de cellules mâles impose un mécanisme de compensation du dosage des gènes portés par le chromosome X supplémentaire, qui passe notamment par l'inactivation de ce chromosome. L'objectif de cette étude est l'identification des protéines recrutées par l'ARN Xist impliqué dans la mise en place de l'inactivation du chromosome X.

Parmi ces divers projets, je vais développer trois études : l'identification des partenaires protéiques spécifiques de la poly(ADP-ribose)polymérase-2, la recherche des protéines interagissant avec le site A7 de l'ARN de VIH-1 et l'identification des partenaires protéiques de l'ARN Xist dans l'inactivation du chromosome X.

## **Chapitre II. Identification de partenaires protéiques de la poly(ADP-ribose) polymérase-2**

Le but de cette étude, réalisée en collaboration avec Delphine Quénet et Françoise Dantzer du département "Intégrité du génome" de l'ESBS à Strasbourg, était d'identifier les partenaires protéiques spécifiques de la poly(ADP-ribose)polymérase-2 (Parp-2).

### **1. Contexte biologique**

La cellule eucaryote dispose d'un réseau de surveillance sur l'état de l'intégrité de son génome, qui lui permet soit de survivre en réparant les lésions présentes dans l'ADN soit de déclencher la mort cellulaire par apoptose. Les poly(ADP-ribose)polymérases (Parp) DNA-dépendantes (Parp-1 et -2) font partie du réseau de surveillance du génome qui garantit son intégrité [Oei et al., 2005]. Elles catalysent la poly(ADP-ribosyl)ation, une modification post-traductionnelle des protéines nucléaires induite par la présence de cassures dans l'ADN.

Des études menées par nos collaborateurs ont mis en évidence une dérégulation importante du processus de spermatogenèse chez des souris mâles déficientes en Parp-2 [Dantzer et al., 2006]. La spermatogenèse se définit comme la transformation des cellules germinales diploïdes appelées spermatogonies en gamètes masculins haploïdes appelés spermatozoïdes. Ce processus biologique se déroule en 3 étapes (Figure 5) : la prolifération cellulaire, la méiose et la différenciation de spermatides en spermatozoïdes encore appelée spermiogenèse.

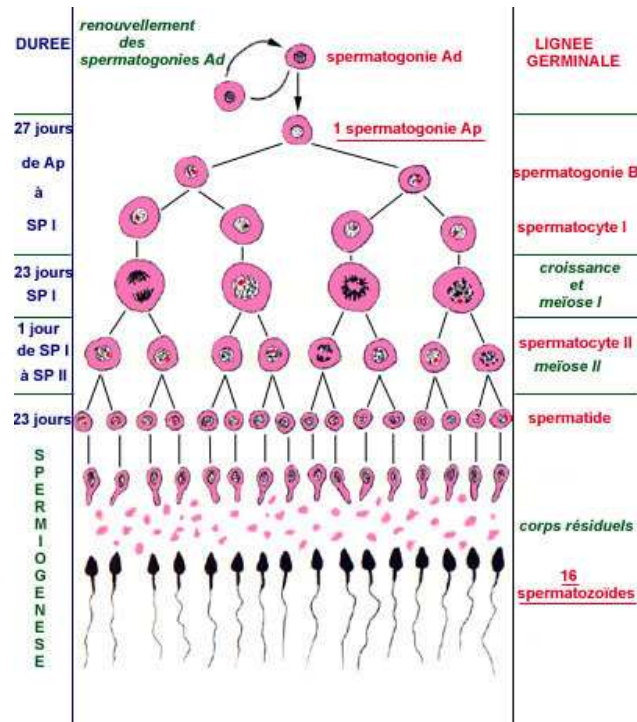


Figure 5 : La spermatogénèse.

Les deux étapes primordiales de la spermatogénèse sont la méiose et la spermiogénèse car une dérégulation de l'une et/ou l'autre de ces étapes provoque un avortement de la spermatogénèse et donc un phénotype de stérilité. Or, les études menées par nos collaborateurs ont montré que Parp-2 joue un rôle essentiel au niveau de la spermiogénèse [Dantzer et al., 2006]. Pour mieux comprendre les mécanismes mis en jeu dans le processus de spermatogénèse et d'identifier le rôle fonctionnel de Parp-2, la recherche de partenaires spécifiques de Parp-2 est réalisée par la méthode précédemment décrite.

## 2. Mise en œuvre de l'approche protéomique développée

Pour l'identification des protéines partenaires spécifiques de Parp-2, un extrait cellulaire total est préparé à partir de 30 testicules de souris. Des immuno-précipitations avec soit un anticorps purifié anti-Parp-2 soit un anticorps contrôle sont réalisées sur l'extrait protéique total. L'identification des partenaires immuno-précipités est effectuée selon la stratégie protéomique précédemment décrite. Les protéines sont séparées sur gel 1D SDS-PAGE et détectées par une coloration au SyproRuby® (Figure 6). La protéine Parp-2 recombinante purifiée a été également déposée sur le gel, afin de visualiser la migration de la protéine seule.



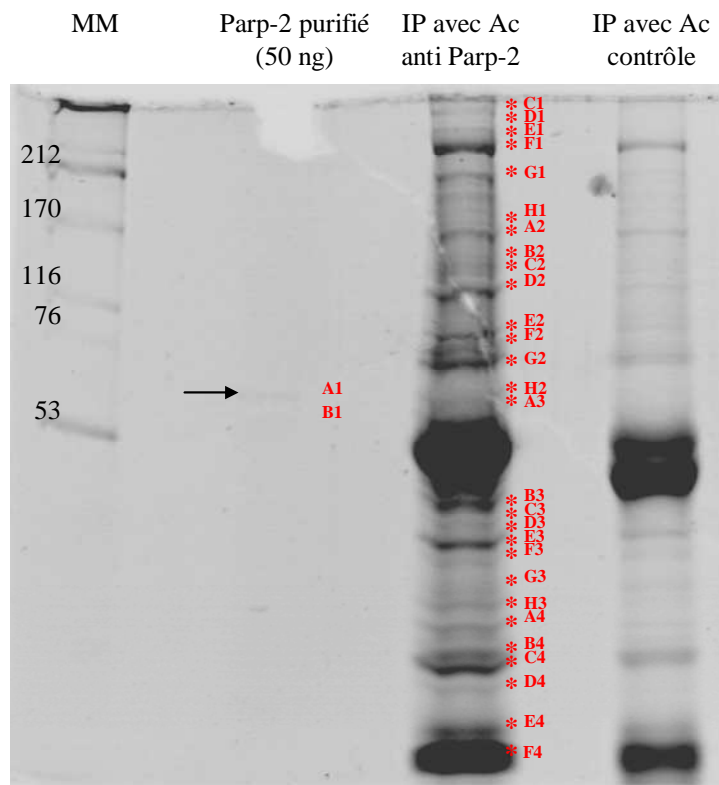


Figure 6 : Gel d'électrophorèse monodimensionnel, coloré au SyproRuby®

Le profil protéique obtenu avec l'anticorps purifié anti-Parp-2 révèle une grande quantité de partenaires co-purifiés avec Parp-2. Au contraire, peu de protéines ont été co-immuno-précipitées aspécifiquement avec l'anticorps contrôle. Les bandes protéiques visibles et la protéine purifiée ont été découpées (comme indiqué figure 6) et analysés par nanoLC-MS/MS. La quantité de protéines présentes sur le gel est peu importante, donc nous avons choisi l'instrumentation la plus sensible au laboratoire au moment où ce projet a débuté : le couplage CapLC-QTOF2 (Waters Corps.)

### 3. Résultats

Pour l'interprétation des données MS/MS, les recherches ont été faites dans la banque Swiss-Prot (version du 04/04/2006, 215741 séquences). Cette étude a été réalisée avant l'installation au laboratoire des logiciels permettant l'automatisation pour la validation des interprétations issues des données MS/MS. Celles-ci ont donc été faites manuellement avec comme critères de validation deux peptides correctement séquencés par protéines avec des spectres de masse vérifiés manuellement. Cela a permis d'identifier 113 protéines uniques pouvant être des candidats à l'interaction avec Parp-2. (Annexe 1 : Liste des protéines identifiées dans les bandes de gel correspondant à l'immuno-précipitation anti-Parp-2).

La protéine cible, Parp-2 (O88554; 3 peptides) a été identifiée lors de l'analyse de la bande H2, qui est située au même niveau que la protéine recombinante purifiée déposé sur le gel. De

nombreuses protéines ont été identifiées, parmi lesquelles Parp-1 (P27008; 13 peptides) et l'ADN polymérase  $\beta$  (Q8K409; 6 peptides). Les protéines Parp-1 et ADN polymérase  $\beta$  sont connues pour être des partenaires d'interaction de Parp-2 [Schreiber et al., 2002]. L'identification de ces deux protéines confirme l'efficacité de l'immuno-précipitation réalisée pour la purification du complexe.

Chez les eucaryotes, les histones servent à l'empaquetage de l'ADN mais participent aussi à la vérification de l'intégrité de cet assemblage. Les isoformes de l'histone H1 [H1.1 (P43275; 13 peptides), H1.2 (P15864; 9 peptides), H1.3 (P43277; 7 peptides), H1.4 (P43274; 5 peptides), H1.5 (P43276; 3 peptides)] et H1t (Q07133; 3 peptides), isoforme s'exprimant spécifiquement dans les testicules ont été identifiées. L'histone Macro-H2A1.2. (Q9QZQ8) a été identifiée avec 11 peptides repartis sur l'ensemble de la séquence (36.6 % de couverture de séquence) (Figure 7-A). Macro-H2A1.2. est connue pour être un partenaire de Parp-1 [Nusinow et al., 2007]. Pour confirmer l'interaction entre Parp-2 et Macro-H2A1.2., une immuno-précipitation du complexe Parp-2/protéine a été réalisée à l'aide d'un anticorps anti-Parp-2 (Figure 7-B), le complexe obtenu a été déposé sur gel. La révélation du gel a été faite par un anticorps contre Parp-2 et un anticorps contre Macro-H2A1.2. Les deux protéines sont donc bien dans le même complexe, nos collaborateurs poursuivent les études pour caractériser cette interaction.

### A

```

1   SSRGGKKKST KTSRSAKAGV IFPVGRMLRY IKKGHPKYRI GVGAPVYMAA
51  VLEYLTAEIL ELAGNAARDN KKGRVTPRHI LLAVANDEEL NQLLKGVTTIA
101 SGGVLPNIHP ELLAKKRGSK GKLEAIITPP PAKKAKSPSQ KKPVAKKTGG
151 KKGARKSKKK QGEVSKAASA DSTTEGTPTD GFTVLSTKSL FLGQKLNLIH
201 SEISNLAGFE VEAIINPTNA DIDLKDDLGN TLEKKGKKEF VEAVLELRKK
251 NGPLEVAGAA ISAGHGLPAK FVIHCNSPVW GADKCEELLE KTVKNCLALA
301 DDRKLLKSIAP PSIGSGRNGF PKQTAAQLIL KAISYFVST MSSSIKTVYF
351 MLFDSESIGI YVQEMAKLDA N

```

### B

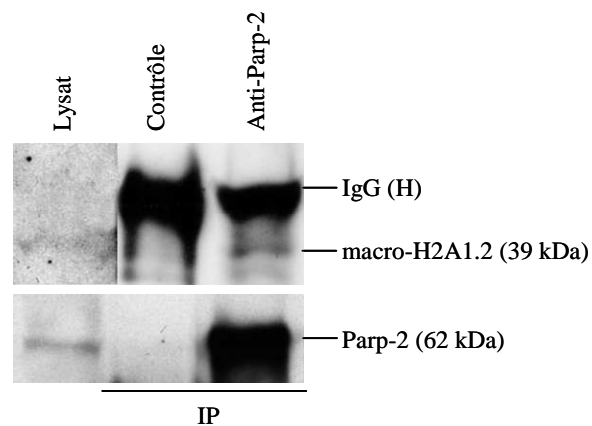


Figure 7 : Mise en évidence d'une interaction entre Parp-2 et macro-H2A1.2. A) Distribution des peptides identifiés sur la séquence de macro-H2A1.2. B) Immuno-précipitation de Parp-2 par l'anticorps purifié anti-Parp-2 sur un extrait protéique total et analyse en Western Blot avec un anticorps anti-macroH2A1.2.

Parmi les protéines identifiées, des protéines impliquées dans la régulation de la transcription ont été détectées. On trouve Cullin Associated NEDD8 Dissociated protein 1 (Q6ZQ38; 10 peptides)

qui régule la transcription de certains promoteurs, DNA Damage binding protein 1 (Q16531; 5 peptides) composé d'un complexe intervenant dans la réponse aux dommages induits par les UV, l'ADN hélicase II (Q08211; 5 peptides) intervenant dans l'initiation de la transcription et le facteur de transcription intermédiaire 1 beta (TIFβ) (Q62318 ; 6 peptides). TIFβ fait partie des protéines candidates ayant retenu l'attention de nos collaborateurs et dont l'interaction avec Parp-2 a été étudiée plus en détails (Figure 8).

**A**

Peptide	Mobs (m/z)	Charge	Séquence
297-304	480.22	2	MAILQIMK + 2 Ox(M)
408-427	674.27	3	VAERPGTNSTGPGPMAPPR + Ox(M)
472-482	552.24	2	GEGEVSGLLR
<b>492-506</b>	<b>837.84</b>	<b>2</b>	<b>LDLDTSDSQPPVFK</b>
774-789	595.92	3	LTEDKADVQSIIGLQR
779-789	600.27	2	ADVQSIIGLQR

**B**

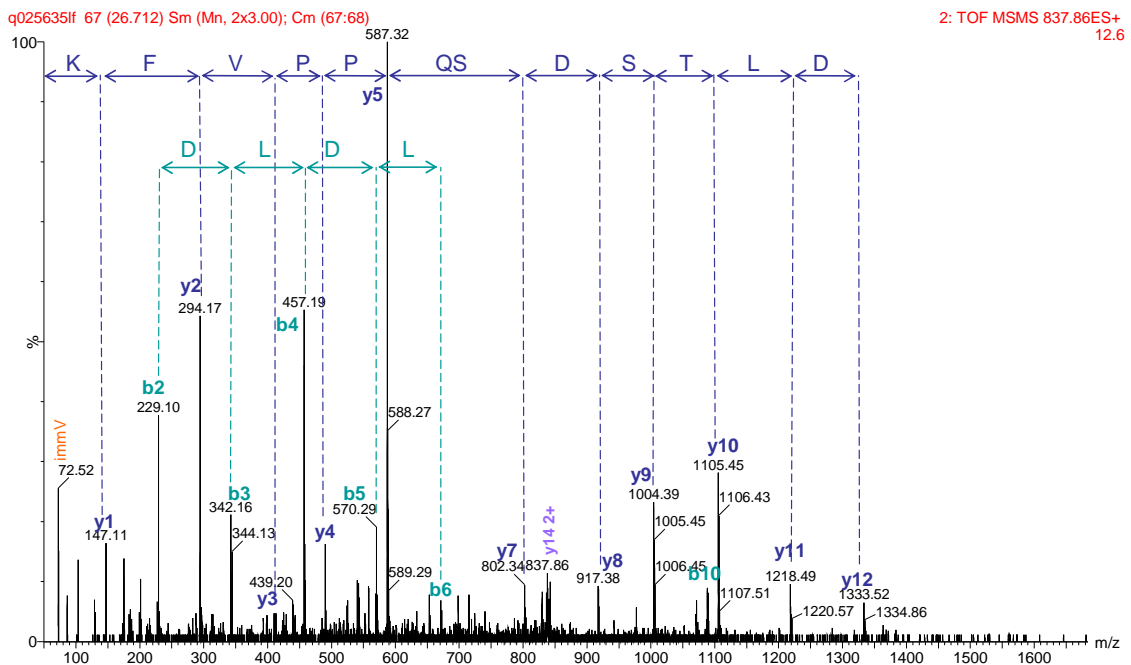


Figure 8 : Identification de TIFβ. A) Liste de peptides identifiés. B) Spectre de fragmentation de l'ion m/z 837.8, doublement chargé, correspondant au peptide 495-506 LDLDTSDSQPPVFK.

La protéine TIFβ est identifiée avec 6 peptides repartis sur l'ensemble de la séquence (8.3% de couverture de séquence) (Figure 8-A). Les spectres de fragmentation ont permis une identification sans ambiguïté de chacun de peptides, tel que le spectre de fragmentation de l'ion m/z 837.8 où la série des ions fragments y permet de séquencer le peptide LDLDTSDSQPPVFK (Figure 8-B). L'interaction entre TIFβ et Parp-2 a été mise en évidence *in vitro* mais aussi *in vivo* par nos collaborateurs [Quenet et al., 2008].

L'analyse protéomique a également permis d'identifier des protéines de la famille Hsp (Heat shock protein). Parmi celles-ci, la plus intéressante est probablement Hsp70.2 (P14659; 25 peptides), protéine chaperone dont l'absence chez la souris conduit à la stérilité des mâles. De plus, le

phénotype de stérilité observé sur les souris KO Hsp70.2 est très semblable au phénotype de stérilité des souris KO Parp-2. Les importines [Importin beta-1 subunit (P70168; 4 peptides), Importin beta-3 (Q8BKC5; 7 peptides) et Importin-alpha re-exporter (Q9ERK4; 4 peptides)] peuvent être des protéines intéressantes car impliquées dans le transport cytoplasme - noyau des histones. Nous avons aussi identifié des protéines s'exprimant uniquement dans les testicules comme la protéine d'interaction Zona pellucida sperm binding protein (Q60736; 3 peptides).

#### **4. Conclusion**

Les analyses protéomiques du complexe protéine/protéine dans lequel intervient Parp-2 a permis de mettre en évidence une liste importante de candidats partenaires d'interaction. Les protéines Macro-H2A1.2. et Hsp70.2 sont les protéines actuellement en cours d'étude par nos collaborateurs. L'interaction entre la protéine Parp-2 et la protéine TIF $\beta$ , identifiée par approche protéomique, a été étudié en détail par nos collaborateurs. Ce travail a donné lieu à une publication dont la version ci-jointe a été acceptée en juin 2008 dans le Journal of the Federation of American Societies of Experimental Biology.

*Signalement bibliographique ajouté par le :*

**UNIVERSITÉ DE STRASBOURG**  
**Service Commun de Documentation**

**The histone subcode: poly(ADP-ribose) polymerase-1 (Parp-1) and Parp-2 control cell differentiation by regulating the transcriptional intermediary factor TIF1 $\beta$  and the heterochromatin protein HP1 $\alpha$**

Delphine QUENET, Véronique GASSER, Laetitia FOUILLEN, Florence CAMMAS, Sarah SANGLIER-CIANFERANI, Régine LOSSON, and Françoise DANTZER

**The FASEB Journal, 2008, vol. 22, pages 3853-3865**

**Publication 1 : pages 76-... :**

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'UdS peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1096/fj.08-113464>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>



## **Chapitre III. Identification de partenaires protéiques de complexes ARN/protéine**

De nombreuses collaborations avec l'équipe de Christiane Branlant du laboratoire AREMS m'ont permis d'appliquer cette stratégie pour des complexes ARN/protéine pour diverses thématiques. Deux thématiques sont développées dans ce chapitre:

- Etude des partenaires protéiques d'interaction du site A7 de l'ARN de VIH-1, réalisée en collaboration avec Virginie Marchand.
- Etude des partenaires protéiques de l'ARN Xist impliqués dans l'inactivation, réalisée en collaboration avec Sylvain Maenner.

### **1. Choix du couplage utilisé pour les études ARN/Protéine**

Au début des collaborations concernant les études pour les identifications des partenaires d'interaction de complexes ARN/protéine, la question du choix de l'instrument a été posée. Les échantillons protéiques dont nous disposions étaient des gels d'électrophorèse 1D, colorés au bleu colloïdal, issus de la purification de complexe ARN/protéine. A ce moment, le laboratoire disposait de deux couplages nanoLC-MS/MS avec des sensibilités différentes : le couplage CapLC-QTOF2 (Waters Corp.) et le couplage nanoLC-Chip (Agilent Technologies) - Trappe ionique HCT plus (Bruker Daltonics). Afin de choisir le système convenant le mieux pour répondre aux questions posées, i.e. l'identification des protéines partenaires d'un ARN, une dizaine de bandes de gels ont été analysées sur les deux systèmes et interprétés de façon identique (Figure 9).

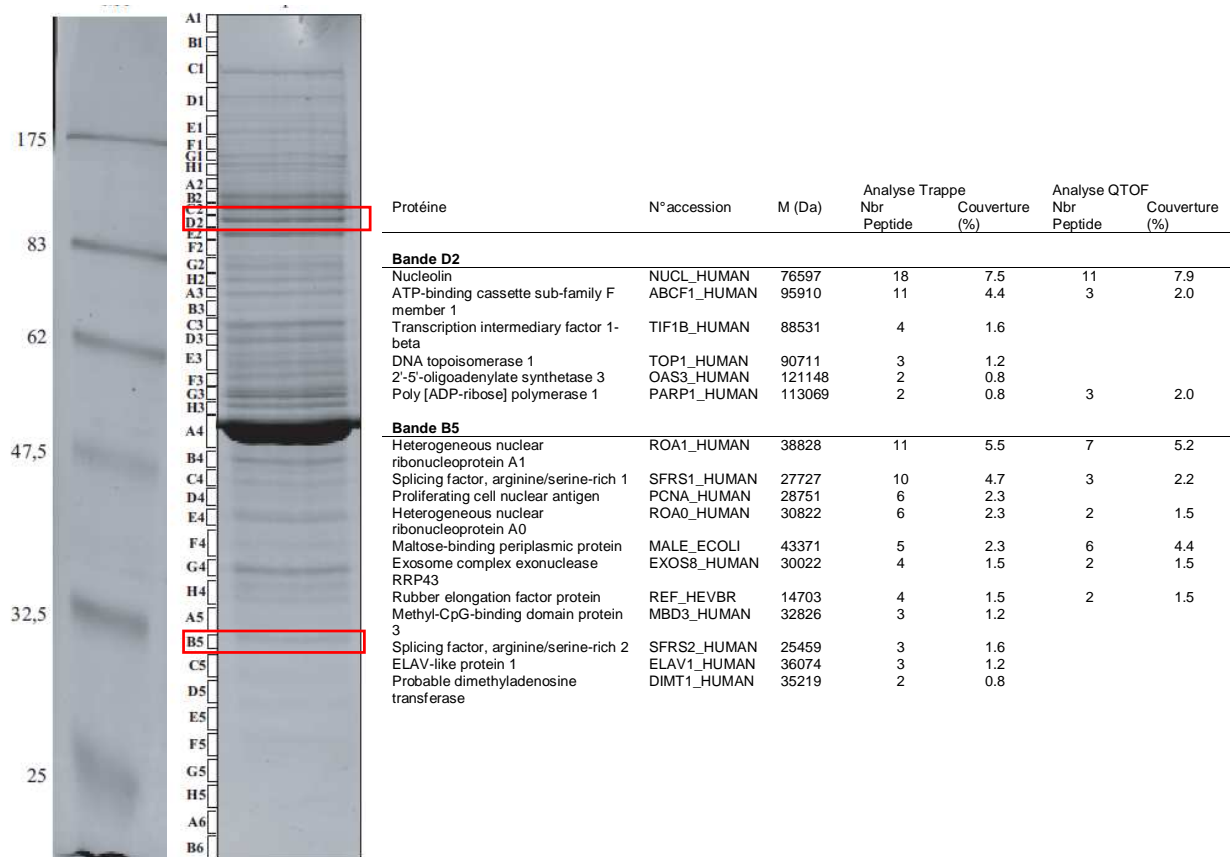


Figure 9 : Comparaison des identifications obtenues pour les bandes D2 et B5 d'un gel d'électrophorèse 1D, coloré au bleu colloïdal, entre des analyses nanoLC-MS/MS réalisées sur le couplage CapLC-QTOF2 (Waters Corp.) et le couplage nanoLC-Chip (Agilent Technologies) - Trappe ionique HCT plus (Bruker Daltonics). L'échantillon présenté correspond à un échantillon préliminaire de l'étude concernant les identifications de partenaires protéiques de l'ARN Xist (voir III.), il s'agit de l'incubation de l'ARN XIM avec un extrait nucléaire de cellules HeLa, purifié par la méthode MS2.

Deux exemples sont présentés, il s'agit de l'analyse d'une bande intense (bande D2) ainsi que d'une bande un peu plus faiblement colorée (bande B5). Les protéines majoritaires ont été identifiées par les deux systèmes. Plusieurs protéines minoritaires ont été identifiées uniquement par le couplage nanoLC-Chip - Trappe ionique HCT plus. Après discussion avec nos collaborateurs, le choix a été fait de travailler sur le couplage CapLC-QTOF2. Les protéines majoritaires après la purification sont considérées comme les protéines formant des complexes importants avec l'ARN cible.



## 2. Etude des partenaires protéiques d'interaction du site A7 de l'ARN de VIH-1

### 2.a. Contexte biologique

Le virus VIH-1 ou Virus de l'Immunodéficience Humaine de type 1, responsable du Syndrome d'ImmunoDéficience Acquise (SIDA) a été isolé en 1983 par le Pr. Luc Montagnier (Prix Nobel de Médecine 2008). Depuis le virus s'est répandu et est responsable d'une épidémie dont souffrent aujourd'hui 33 millions d'individus dans le monde (Chiffres ONUSIDA 2008). Le SIDA conduit à une grave immunodépression exposant le malade à des infections opportunistes qui lui sont fatales. Le virus est transmis de la mère à l'enfant et par voies sanguine ou sexuelle. Après contamination d'un individu, le virus est transporté en moins de 48 h dans les ganglions lymphatiques, où il infecte les lymphocytes T et les macrophages [Sleasman & Goodenow, 2003]. Une fois l'ADN viral intégré dans la cellule infectée, il ne peut être éliminé ou détruit, sauf s'il y a destruction de la cellule elle-même. L'infection se propage selon le cycle viral du virus VIH-1, comme décrit dans la figure 10.

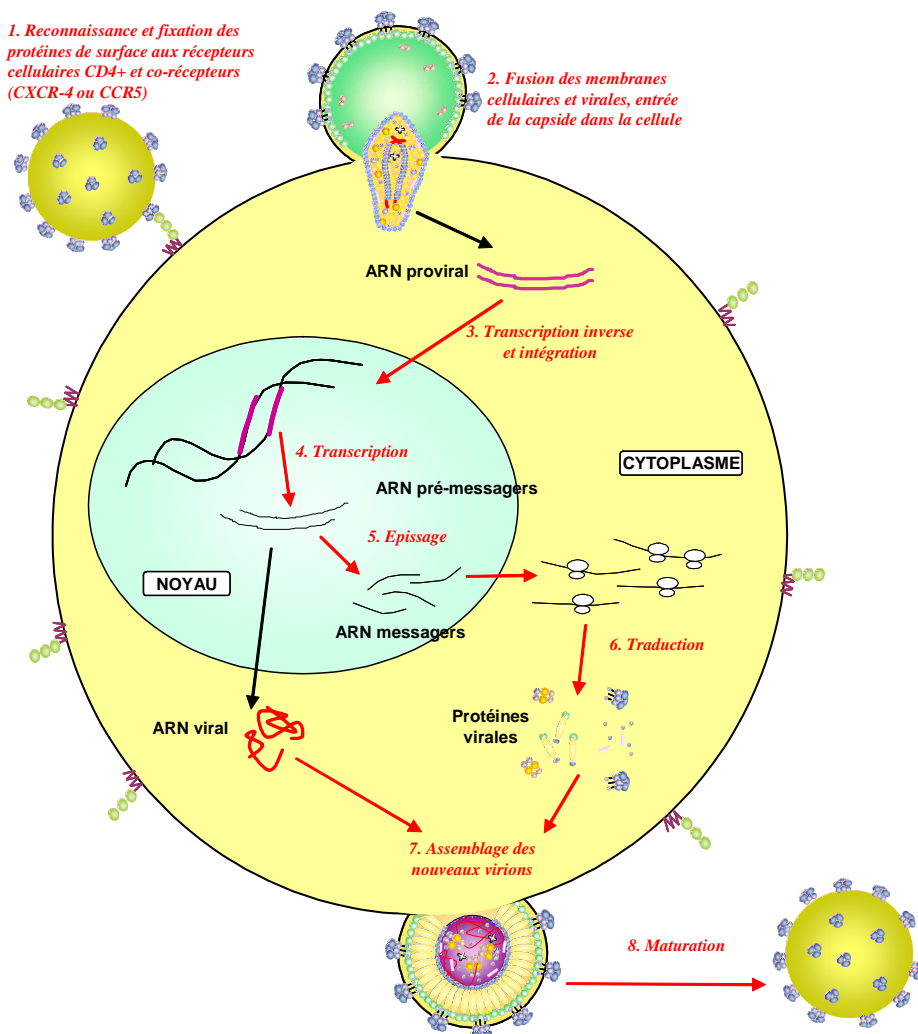


Figure 10 : Cycle Viral du virus VIH-1 [Saliou, 2008]. L'infection d'une cellule par le virus VIH-1 débute par la fixation spécifique au récepteur CD4 (1) et la fusion des membranes pour permettre l'entrée des virions dans la

cellule (2). La capside renfermant l'ARN viral pénètre dans la cellule. Dans le cytoplasme de la cellule infectée, l'ARN génomique est converti en ADN proviral par transcription inverse (3). L'ADN proviral est importé dans le noyau de la cellule hôte et intégré dans son génome. Le virus utilise l'ensemble de la machinerie cellulaire pour la transcription de l'ADN proviral (4). Le transcrite primaire de VIH-1 sert à la fois d'ARN génomique pour les nouveaux virions produits et d'ARN messager pour la synthèse des protéines virales, réalisée par la machinerie de traduction de la cellule hôte (5-7). Les particules virales sont enfin assemblées à partir des différents partenaires viraux et cellulaires, puis subissent une ultime étape de maturation (8) pour devenir infectieux et se propager.

Une des étapes clé du cycle viral est l'épissage alternatif de l'ARN pré-messager du VIH-1. Le virus utilise la machinerie cellulaire composée pour l'épissage du spliceosome, complexe de macronucléoprotéines, qui catalyse la réaction d'épissage, et de facteurs cellulaires, qui la régule. L'épissage a lieu dans le noyau et transforme donc les ARN pré-messagers (pré-ARNm) issus de la transcription, en ARN messagers (ARNm) prêt pour la traduction. Classiquement, les séquences retenues sont appelées exon et les séquences excisées introns. La réaction d'épissage s'établit entre un site donneur d'épissage (5') et un site accepteur (3') (Figure 11). L'épissage alternatif permet de produire un grand nombre de protéines différentes à partir d'un même pré-ARNm.

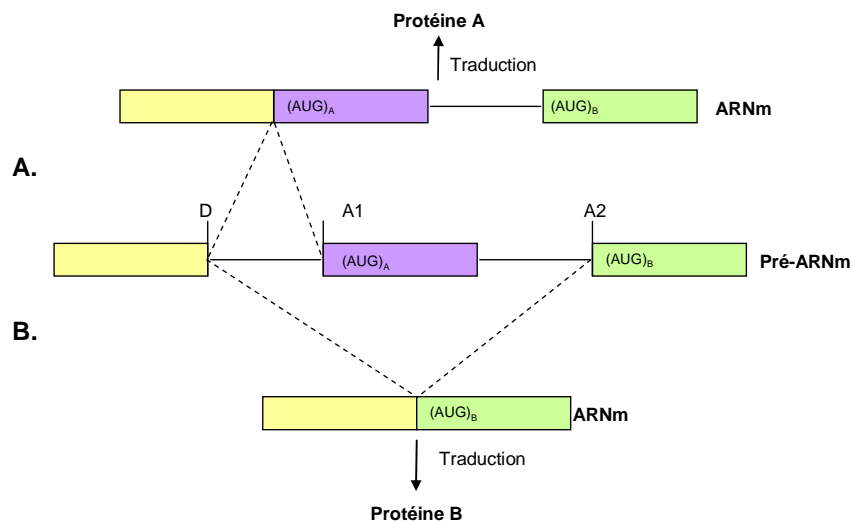


Figure 11 : Représentation schématique de l'épissage alternatif. A) Epissage entre le site donneur D et le site accepteur A1 permettant d'exprimer la protéine A. B) Epissage entre le site D et le site accepteur A2 permettant d'exprimer la protéine B.

Dans le cas de l'ARN pré-messager de VIH-1, 5 sites donneurs (D1, D1A, D2, D3 et D4) et 9 sites accepteurs (A1, A1A, A2, A3, A4c, A4a, A4b, A5 et A7) d'épissage (Figure 12) sont utilisés en combinaison pour générer les différents ARNm.

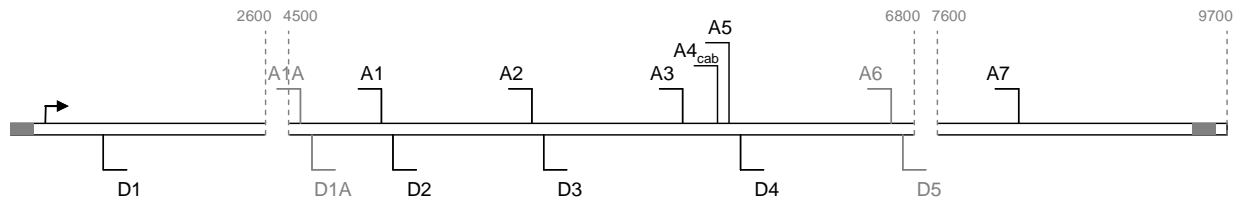


Figure 12 : Organisation des sites d'épissage de l'ARN pré-messager de VIH-1 [Saliou, 2008]. L'ARN de certaines souches virales peut contenir un ou deux autres sites fonctionnels A6 et D5.

Une quarantaine d'ARN messagers viraux de différentes tailles générés codent pour les 6 protéines et 3 précurseurs protéiques nécessaires et suffisants à la formations de nouveaux virions. Comme tous les autres sites, l'utilisation du site accepteur A7 est soumise à plusieurs régulations. Afin de mieux comprendre ces régulations, nos collaborateurs cherchent à identifier les partenaires protéiques du site A7 de l'ARN de VIH-1, notamment par une approche protéomique.

Un des facteurs de régulation de l'épissage est aussi l'accessibilité sur l'ARN des sites d'épissage et des sites de fixation des facteurs de régulation. Ainsi la séquence ARN contenant le site A7 a été réalisé de sorte que l'ARN puisse adopter sa structure secondaire (Figure 13).

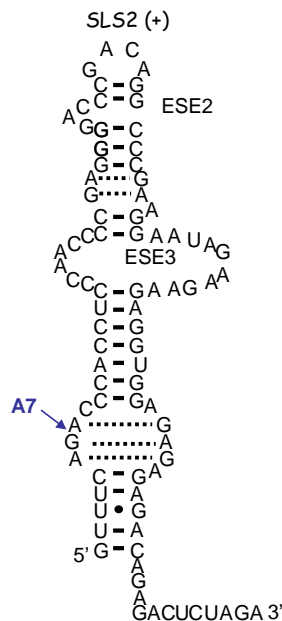


Figure 13 : Structure secondaire de l'ARN Sls2, contenant le site d'épissage A7.

## **2.b. Application de la stratégie à la recherche de partenaires protéiques d'interaction de l'ARN Sls2 contenant le site A7.**

L'ARN Sls2 a été incubé avec un extrait nucléaire de cellules HeLa et les complexes obtenus sont analysés selon la stratégie d'identification présentée précédemment. Pour la purification du complexe ARN/protéine, deux méthodes sont utilisées : la méthode tobramycine et la méthode MS2 (Figure 14).

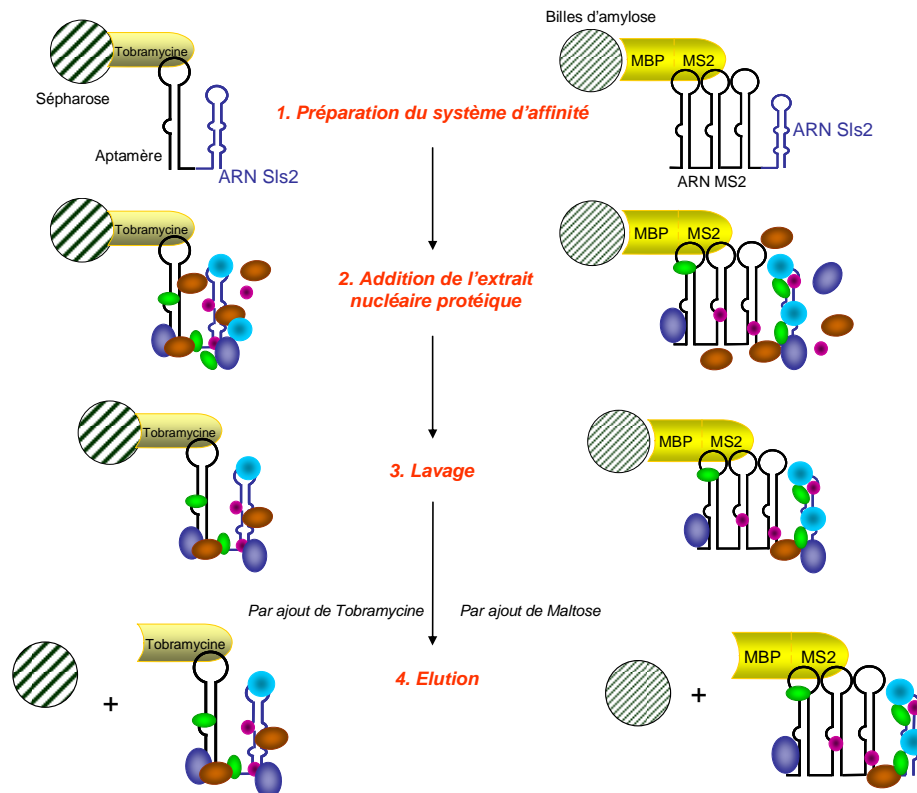


Figure 14 : Les méthodes de purification des complexes ARN/protéine utilisées. A) Méthode tobramycine. B) Méthode MS2.

Dans le cas de la technique tobramycine [Hartmuth et al., 2002], l'élément de jonction est la tobramycine fixé sur les billes de sépharose et capable de se lier l'ARN aptamère. Dans le cas de la technique MS2 [Zhou et al., 2002], l'élément de jonction est la protéine de fusion MBP-MS2 capable de fixer aux billes d'amylose et à l'ARN MS2. L'incubation est réalisée avec un extrait nucléaire de cellules HeLa et l'éluion avec un excès de tobramycine ou par un ajout de maltose. Quelle que soit la méthode de purification, les protéines issues de l'éluion sont fractionnées enfin sur un gel d'électrophorèse mono-dimensionnel (Figure 15).

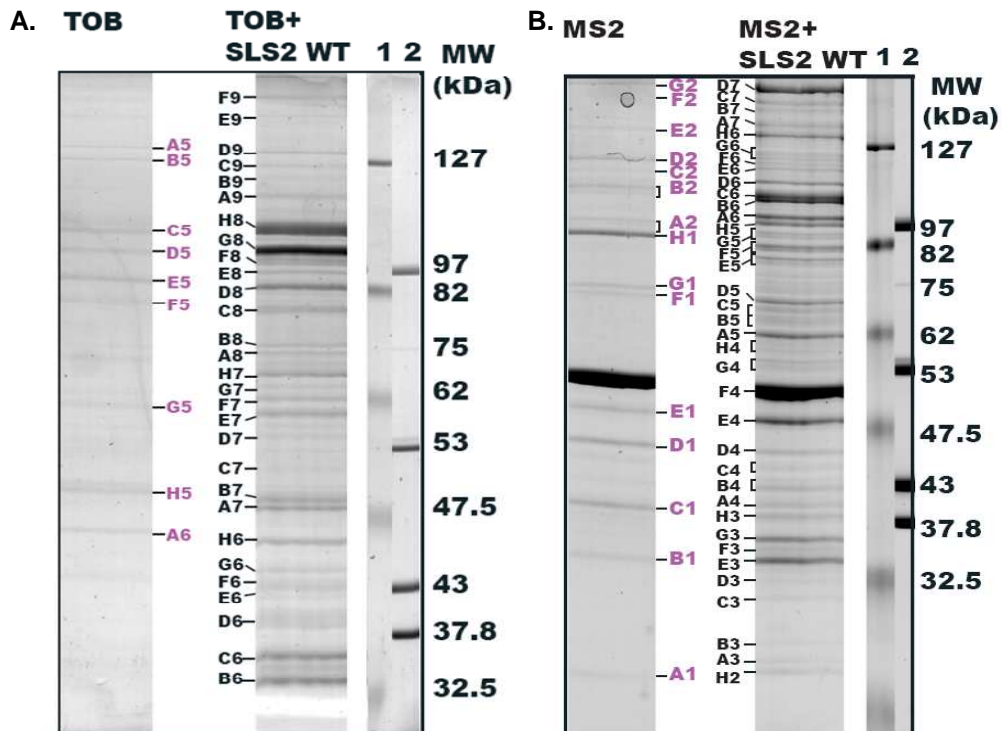


Figure 15 : Gels d'électrophorèse 1D, colorés au bleu de colloïdal. A) Gel 1D des protéines issues de la purification avec la méthode Tobramycine. B) Gel 1D des protéines issues de la purification avec la méthode MS2.

L'ensemble des bandes visibles et signalées sur les gels sont découpées et traitées avant d'être analysées par nanoLC-MS/MS. L'ensemble de données MS/MS est ensuite soumis à une recherche dans la banque de données protéique Swiss-Prot (version du 03/04/2007, 263525 séquences) à l'aide de l'algorithme de recherche Mascot. Les interprétations sont effectuées manuellement avec pour critères de validation d'une identification un minimum de 3 peptides ayant des spectres de fragmentation vérifiés manuellement. Les critères de validations choisis sont stringents pour identifier avec confiance les protéines majoritaires partenaires de l'ARN.

## 2.c. Résultats

Les deux gels d'électrophorèses (Figure 15) présentent un grand nombre de bandes visibles tout le long du gel. Les profils obtenus sont plutôt similaires, bien que plus de bandes soient visibles au delà de 127 kDa sur la piste MS2+Sls2WT que sur la piste TOB+Sls2WT. Pour chaque méthode, un contrôle est effectué en parallèle avec soit l'aptamère seul, soit l'ARN MS2 seul selon la méthode employée. Quelques protéines sont présentes dans les contrôles, comme nous le montrent les bandes visibles sur ces pistes. La bande intense visible sur le gel MS2 dans le contrôle et dans l'échantillon correspond à la protéine de fusion MBP-MS2, protéine impliquée dans la méthode de purification MS2.

Un grand nombre de protéines ont été identifiées par cette méthode dont 33 protéines sont communes aux deux méthodes (Figure 16). 22 protéines sont identifiées uniquement par la méthode tobramycine et 58 uniquement par la méthode MS2 (Annexe 2). La formation des complexes est faite à pH physiologique 7.5 pour la méthode MS2, contrairement à la méthode tobramycine où cette formation est réalisée à pH 9.1 [Marchand, 2005]. On peut donc supposer qu'un plus grand nombre de complexes sont formés à pH 7.5, ce qui explique le plus grand nombre de protéines identifiées pour la méthode MS2 que par la méthode tobramycine. De plus, la quantité de complexes purifiés obtenues par la méthode MS2 est plus importante que celle obtenues par la méthode tobramycine, à partir des mêmes quantités d'ARN et d'extraits nucléaires utilisées.

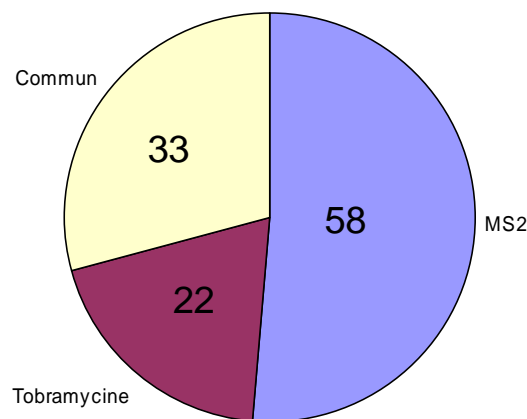


Figure 16 : Répartition des identifications selon la méthode de purification.

Plusieurs catégories de protéines sont identifiées. On remarque d'abord les protéines associées à l'épissage, tels que les hnRNP (Heterogeneous nuclear ribonucleoprotein), les snRNP (Small nuclear ribonucleoprotein) et les protéines SR ou SR-like (Splicing Related). Il y a également les protéines associées à la traduction (Translations related factors) et les protéines liées à l'ARN (RNA binding protein) qui potentiellement peuvent aussi réguler l'épissage.

Plusieurs protéines connues comme partenaires du site A7 [Saliou et al., 2009] ont été identifiées par l'approche protéomique après purification :

- Par les deux méthodes : les protéines "Splicing factor U2AF 35kDa Subunit" (Q01081; Tobramycine : 4 peptides, MS 2 : 4 peptides) et "Splicing factor U2AF 65kDa Subunit" (P26368; Tobramycine : 16 peptides, MS2 : 7 peptides).
- Uniquement par la méthode MS2 : la protéine "Splicing factor, arginine/serine 1" ou ASF/FS2 (Q07955; 6 peptides).

La protéine hnRNP A1 a également été identifiée par les deux méthodes et est en interaction avec l'ARN Sls2 d'après des travaux antérieurs réalisés par nos collaborateurs [Marchand et al., 2002]. Cependant elle est identifiée dans les purifications issues de la méthode tobramycine (30 peptides) et de la méthode MS2 (29 peptides) mais également dans le contrôle de la méthode MS2 (13 peptides). L'élément de jonction peut fixer des protéines contaminants du système mais

également des protéines en interaction avec l'ARN purifié. Les identifications par l'approche protéomique apportent donc des candidats à l'interaction avec l'ARN d'intérêt qui nécessitent d'être confirmés par des méthodes biologiques

Trois protéines (Nucléoline ; P19338, hnRNP H ; P31943, hnRNP K ; P61978), identifiées par les 2 méthodes avec des couvertures de séquences similaires (Tableau 1) ont particulièrement intéressé nos collaborateurs. Ces trois candidats sont actuellement en cours d'études. Des expériences de 'gel-retards' ou des immuno-précipitations (Figure 17) ont déjà permis de valider le fait que ces protéines sont en interactions directes avec l'ARN SIs2.

Tableau 1 : Identification des protéines Nucléoline, hnRNP H et hnRNP K par les méthodes de purification Tobramycine et MS2.

Protéine	N°Accession	Méthode Tobramycine		Méthode MS2	
		Nbr peptide	Couverture %	Nbr peptide	Couverture %
Nucléoline	P19338	17	25	20	31
hnRNP H	P31943	6	19	3	10
hnRNP K	P61978	15	40	14	38

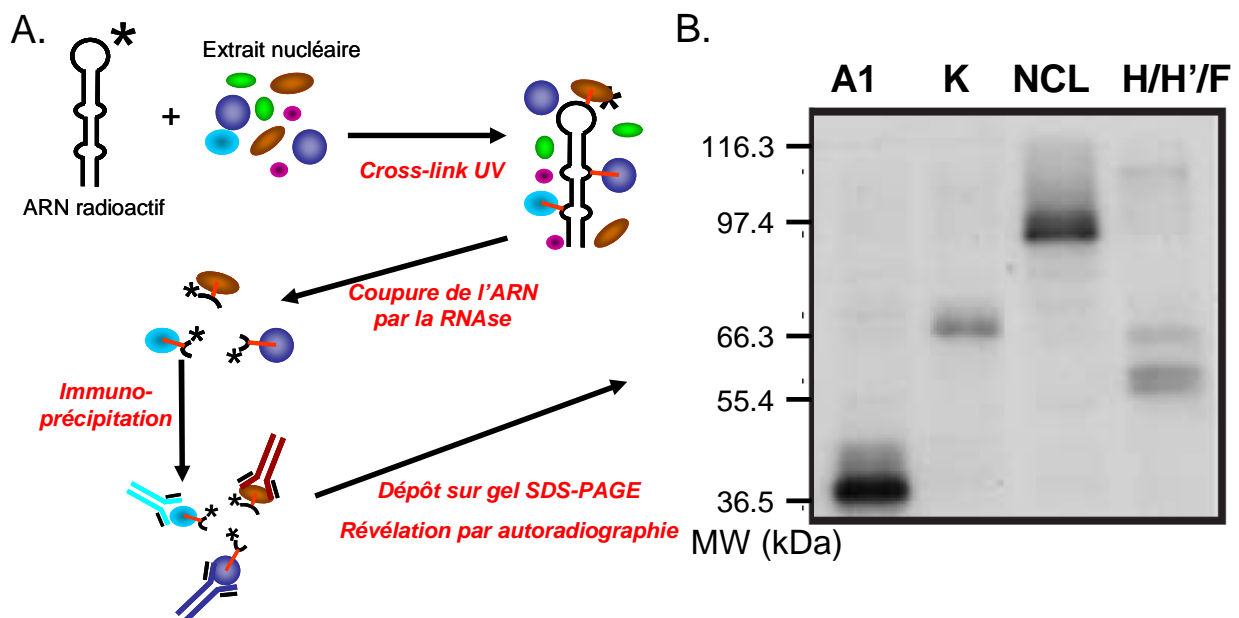


Figure 17 : Mise en évidence de l'interaction entre l'ARN SIs2 et les protéines hnRNP A1 [A1], hnRNP K [K], nucléoline [NCL] et hnRNP H [H/H'/F]. A) Représentation schématique de l'expérience d'immuno-précipitation. B) Gel SDS-PAGE résultant de l'expérience.

L'ARN SIs2, uniformément radioactif, a été incubé avec l'extrait nucléaire de cellules HeLa puis soumis à une réaction de pontage par UV. Une liaison covalente a donc été réalisée entre l'ARN

et les protéines en interaction direct avec l'ARN. Après coupure de l'ARN par une RNase, chaque protéine recherchée a été immuno-précipitée puis déposée sur gel SDS-PAGE. Seuls les complexes contenant de l'ARN radioactif peuvent être révélés, cela permet de mettre en évidence les interactions directes entre la protéine cible et l'ARN. Pour vérifier la validité de la méthode employée, une immunoprécipitation de la protéine hnRNP A1 est réalisée. Cette protéine est en interaction avec Sls2 d'après des travaux antérieurs réalisés par nos collaborateurs [Marchand et al., 2002]. Grâce notamment à cette expérience, une interaction directe entre l'ARN Sls2, contenant le site A7, et les protéines hnRNP H, hnRNP K et la nucléoline a été mise en évidence.

## **2.d. Conclusion**

L'approche protéomique a donc permis de mettre en évidence trois potentiels candidats régulateurs de l'épissage du site A7, les protéines hnRNP H, hnRNP K et la nucléoline. Les études continuent pour localiser les sites de fixations de ces protéines sur l'ARN et surtout pour évaluer l'effet de ces protéines au niveau de l'épissage. L'utilisation de deux méthodes de purification a permis de limiter le nombre de candidats protéiques pouvant être en interaction forte avec l'ARN Sls2 portant le site A7.

Ces résultats ont fait l'objet d'une communication par affiche (version jointe en page suivante) présenté lors des 23èmes journées françaises de Spectrométrie de Masse qui se sont déroulées à Nantes en septembre 2006.



# A proteomic approach for the identification of nuclear proteins involved in the regulation of HIV-1 RNA splicing



Laetitia Fouillen<sup>1</sup>, Virginie Marchand<sup>2</sup>, Christelle Aigueperse<sup>2</sup>, Maryline Santerre<sup>2</sup>  
Iouri Motorine<sup>2</sup>, Christiane Branlant<sup>2</sup>, Sarah Sanglier<sup>1</sup>, Alain Van Dorsselaer<sup>1</sup>



<sup>1</sup>Laboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien UMR7178 CNRS/Université Louis Pasteur, ECPM, 25 rue Becquerel - 67087 Strasbourg Cedex 2 - France

<sup>2</sup>Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR7567 CNRS/UHP, Faculté des Sciences, 54506 Vandœuvre Les Nancy Cedex - France

## INTRODUCTION

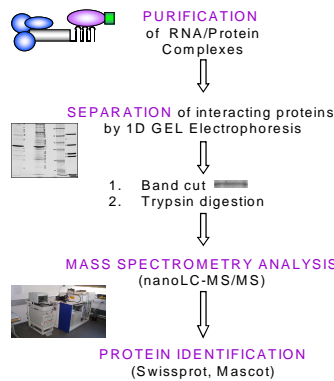
Production of HIV-1 proteins depends upon a complex alternative splicing of the viral primary transcript. HIV-1 multiplication is strictly dependent on the use of some of the splicing sites (A2, A3, A4, A5, A7). It is thus highly important to study the molecular mechanisms involved in the regulation of HIV-1 RNA splicing.

The aim of this work is to use a systematic proteomic approach for the search of nuclear proteins involved in the regulation of HIV-1 RNA splicing. The proteomic approach developed is based on the fractionation of the complexes formed between a fragment of the viral RNA (SLS2) containing the splicing site A7 and a HeLa cell nuclear extract.

### Overall Strategy To Identify Proteins that Interact with a specific RNA

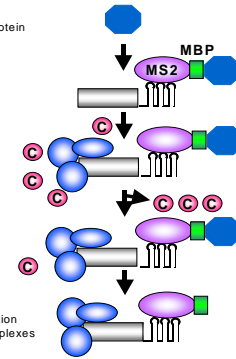
Our approach is based on the fractionation of the complexes followed by a 1D gel electrophoresis and mass spectrometry analyses.

The complexes were formed between a fragment of the viral RNA (SLS2) containing the splicing site A7 and a HeLa cell nuclear extract. Protein complexes were isolated using the MS2 (RNA construction) purification procedure [Zhou *et al.*, 2002].



### MS2 Purification Procedure

- Binding of MS2-MBP protein on amylose beads
- Binding of the targeted RNA
- Incubation with the HeLa nuclear extract
- Washing steps to remove bound non-specific proteins
- Elution by maltose addition of the RNA/Protein complexes



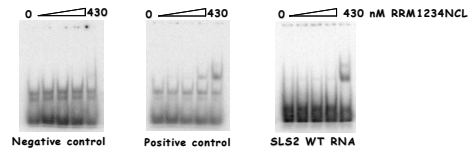
## RESULTS

### Main proteins involved in the regulation of the splicing site A7

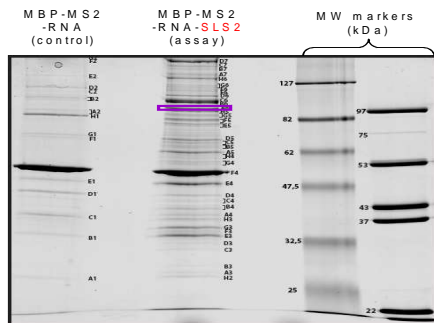
- 93 proteins were identified to interact specifically with SLS2 containing the splicing site A7
- Major proteins involved in the regulation of HIV-1 RNA splicing might be **hnRNP A1**, **nucleolin** and other **hnRNP (K, R, Q, U)**.
- The hnRNP proteins bind to pre-mRNA concomitant with transcription and form ribonucleoprotein complex essential for post-transcription events. In agreement with previous studies [Marchand *et al.*, 2002], **site A7 is downregulated by hnRNP A1**.

### Nucleolin is directly bound to the SLS2 RNA containing the splicing site A7

- Nucleolin is a RNA-binding protein, highly abundant in nucleolus
- Nucleolin is a new partner of site A7 which is thought to be required for the correct folding and packaging of pre-rRNA
- Gel shift assays confirm that nucleolin (NCL) is directly bound to the SLS2 RNA containing the splicing site A7



RNAs containing either the winner SELEX sequence for nucleolin or a negative control or the SLS2 WT RNA were incubated with increasing concentrations of purified nucleolin (RRM1,2,3,4). The RNP complexes were further analyzed by gel-shift assays. No complex was detected with the negative control RNA. However, RNP complexes were both detected with the positive control RNA as well as the SLS2 RNA.



### Proteins identified by nanoLC-MS/MS analysis in band A6

Name	Swissprot Accession number	Sequence coverage	Number of peptides
Nucleolin	P19338	32%	28
Transcription intermediary factor 1-beta (TIF1-beta)	Q13263	16%	14
Anaphase promoting complex subunit 2 (APC2)	Q9UJX6	8%	5
Serine/threonine-protein kinase SRPK1	Q96SB4	4%	3
UDP-N-acetylglucosamine - peptide N-acetylglucosaminyl transferase 110kDa subunit	Q15294	6%	6
DNA Mismatch repair protein Msh2	P43246	5%	4
Ubiquitin-like containing PHD and RING finger domains protein 1	Q96T88	2%	2
Heterogeneous nuclear ribonucleoprotein U (hnRNP U)	Q00839	2%	2

Zhou Z., Sim J., Griffith J., and Reed R. PNAS, 2002, Vol. 99, No. 19, 12203-07.  
Marchand V., Méreau A., Jacquenet S., Thomas D., Mougín A., Gattori R., Stevénin J., and Branlant C. J. Mol. Biol. 2002, 323, 629-52.



### 3. Etudes de partenaires de l'ARN Xist impliqué dans l'inactivation.

#### 3.a. Contexte biologique

Chez les espèces animales, où le nombre des chromosomes X diffère entre les deux sexes, des mécanismes de compensation du dosage des gènes ont été mis en place pour assurer un taux similaire dans les deux sexes des transcrits des gènes liés au chromosome X. Ces mécanismes de compensation sont très importants, car une défaillance dans leur mise en place conduit à la mort de l'embryon très tôt dans le développement. Suivant les espèces, des stratégies différentes sont utilisées pour gérer ce problème de compensation. Chez les mammifères, il y a mise en place de l'inactivation d'un des deux chromosomes X dans les cellules femelles (Figure 18).

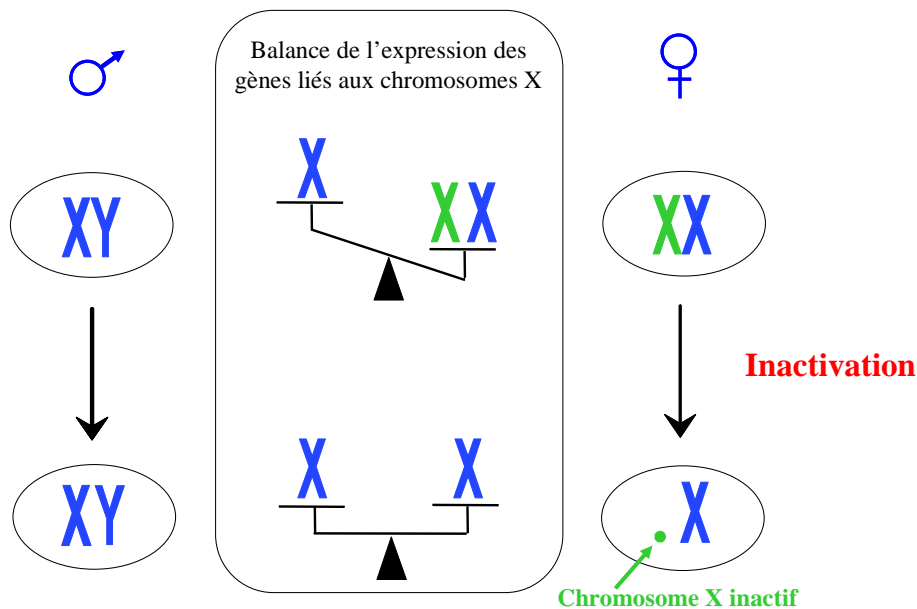


Figure 18 : Représentation schématique de la compensation du dosage de l'expression des gènes liés aux chromosomes X chez les mammifères.

La première étape de cette inactivation aléatoire consiste en l'accumulation d'un très grand ARN non codant (17 à 19 kb) appelé Xist (X inactive specific transcript) sur le futur chromosome X inactif, ceci à un stade très précoce du développement. Le gène codant cet ARN est localisé dans une région appelée Xic (X inactivation center) présente sur les deux chromosomes X. L'inactivation du chromosome X est aléatoire et l'accumulation de Xist commence au début de la différenciation des cellules de la masse interne de cellules au stade blastocyste et continue pendant les premiers jours de la différenciation chez l'embryon (Figure 19).

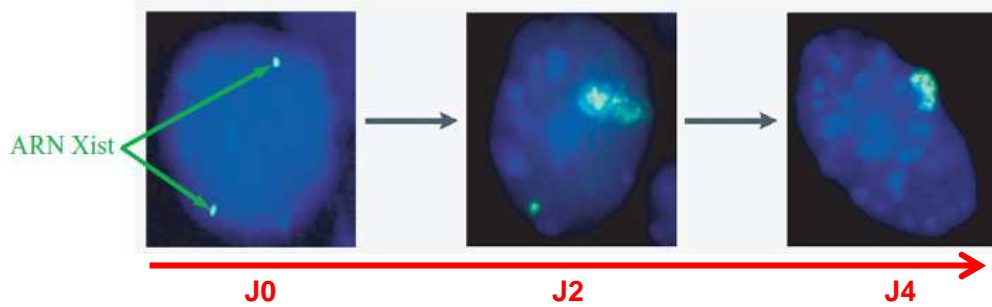


Figure 19 : Evolution de la quantité et de la localisation de l'ARN Xist au cours de la différenciation des cellules ES. L'ARN Xist a été détecté par emploi d'une sonde spécifique marquée par un fluorochrome [Avner & Heard, 2001].

La première étape de cette inactivation aléatoire du chromosome X consiste en l'accumulation d'un grand nombre de molécules de l'ARN Xist sur le chromosome X à inactiver.

Etant donné la complexité d'étude chez l'embryon, un modèle cellulaire est utilisé: il s'agit des cellules ES murines (cellules souches embryonnaires) qui permettent de récapituler les différentes étapes de l'inactivation. Ainsi, à J0, les cellules ES indifférenciées présentent deux signaux qui mettent en évidence l'expression de l'ARN Xist sur les deux chromosomes X. A J2 les cellules ES en début de différenciation présentent une accumulation de Xist sur un des deux chromosomes X. A J4, dans les cellules ES différenciées, l'ARN Xist habille le chromosome inactif et l'expression du gène Xist est réprimée sur le chromosome actif.

Cette accumulation de l'ARN Xist est suivie du recrutement des gènes jusque là actifs dans un territoire hétérochromatinien inactif. L'apposition des marques épigénétiques (modifications des régions N-terminales des histones en l'occurrence méthylation des lysines 27 des histones H3) permet la transmission de l'état inactif à la descendance. Enfin, la méthylation des séquences CpG de l'ADN rend l'inactivation irréversible. Le chromosome inactivé forme alors une structure compacte appelée corps de Barr. Nous venons de voir que ce processus est complexe et qu'il est soumis à l'unique expression du gène Xist. En d'autres termes l'ARN Xist est le pré-requis à une extinction transcriptionnelle efficace ce qui souligne l'importance du rôle d'un ARN dans un processus de régulation génique. Il s'agit d'un concept intéressant dont les différents mécanismes moléculaires sont encore très mal connus.

L'ARN Xist est présent chez tous les mammifères à placenta. Ces ARN Xist ont beaucoup divergés toutefois seules un certain nombre de régions comportant des répétitions de séquences sont conservées. En particulier, une région dite des "A-repeats" qui comporte 7 à 8 éléments répétés, et qui est localisée à l'extrémité 5' de l'ARN Xist, est très conservée chez les mammifères. Cette région des "A-repeats" joue un rôle essentiel pour la fonction de l'ARN Xist, elle est absolument nécessaire pour la mise en place de l'inactivation [Wutz et al., 2002].

Le structure secondaire d'un ARN conditionne l'activité d'un ARN. Or aucune structure secondaire des régions "A-repeat" n'a été décrite dans la littérature. Seule une structure secondaire prédite par informatique existe [Wutz et al., 2002]. Nos collaborateurs ont donc entrepris de réaliser l'analyse expérimentale de cette région et ont ainsi pu proposer une structure secondaire des régions "A-repeat" murine (Figure 20).

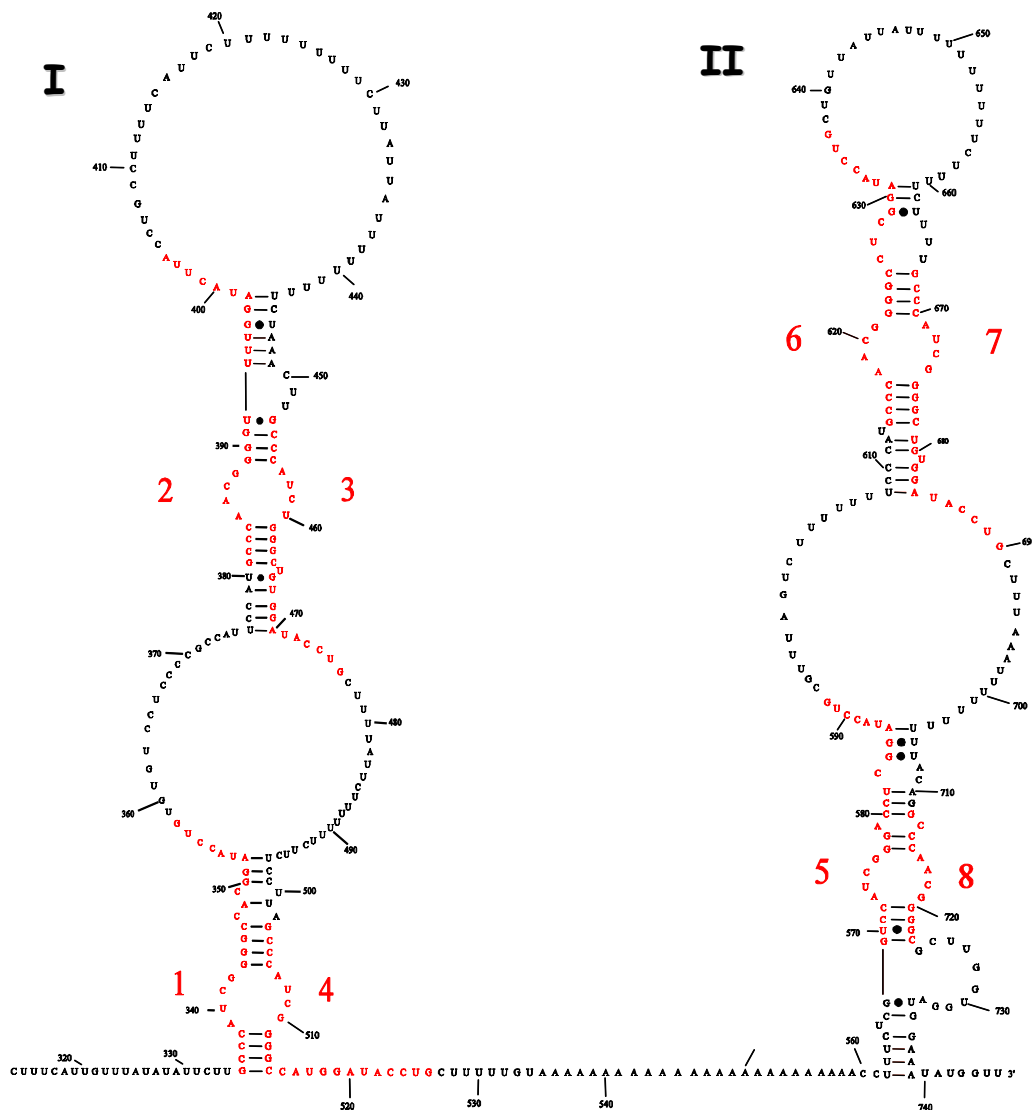


Figure 20 : Modèle de structure secondaire proposé pour les régions des "A-repeat" murine. Les régions des "A-repeat" sont en rouge.

L'objectif de l'étude est d'identifier les facteurs protéiques susceptibles de se lier à la région des "A-repeats".

- La première partie de l'étude consiste à comprendre si le nombre de régions "A-repeats" a une influence sur les protéines recrutées,
- La seconde partie consiste à identifier les partenaires protéiques à plusieurs stades de différenciation pour une structure donnée.

### **3.b. Etude de l'influence du nombre de "A-repeats" sur les protéines recrutées**

L'inactivation du chromosome X est initiée dans les cellules très rapidement après le déclenchement de la différenciation. Pour cette étude, des extraits nucléaires de cellules souches embryonnaires de souris (cellules ES) avant différenciation (J0) ont été préparées. La première partie de l'étude a consisté à comprendre si le nombre de régions "A-repeats" a une influence sur la composition du complexe protéique formé. Pour cela, des ARN possédant 1, 2, 4 et 8 répétitions ont été réalisés (Figure 21), puis incubés en présence d'extrait nucléaire de cellules ES avant différenciation (J0). Les complexes ARN/protéine ont été purifiés par la méthode d'affinité MS2 [Zhou et al., 2002].

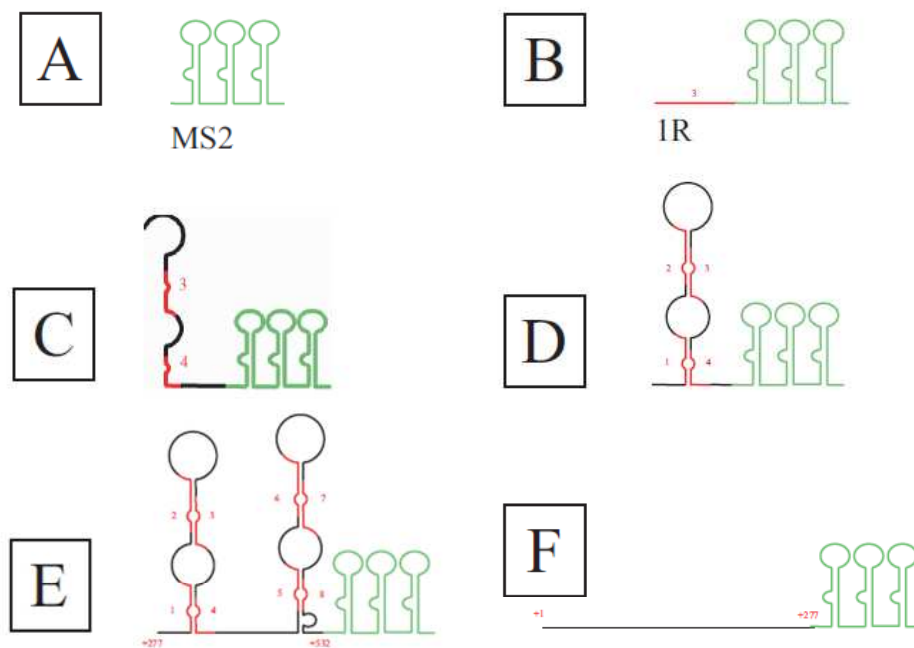


Figure 21 : Représentation schématisques des structures secondaires des différents ARN utilisés A) Partie MS2, B) 1 répétition "A-repeat" fusionné à l'ARN MS2, C) 2 répétitions "A-repeat" fusionnées à l'ARN MS2, D) 4 répétitions "A-repeat" fusionnées à l'ARN MS2, E) 8 répétitions "A-repeat" fusionnées à l'ARN MS2, F) ARN situé en amont des parties "A-repeat" fusionné à l'ARN MS2.

L'ARN E correspond à l'ARN Xist murin, et sera nommé XIM dans la suite du chapitre. Deux contrôles sont également utilisés (Figure 21) : le contrôle MS2 et l'ARN F correspondant à un morceau d'ARN Xist, composé de 277 bases et situé en amont de la région contenant les "A-repeats". Ces contrôles nous permettront de supprimer de la liste des protéines se liant spécifiquement. Les

protéines purifiées par la méthode d'affinité MS2 sont séparées sur un gel d'électrophorèse 1D (Figure 22).

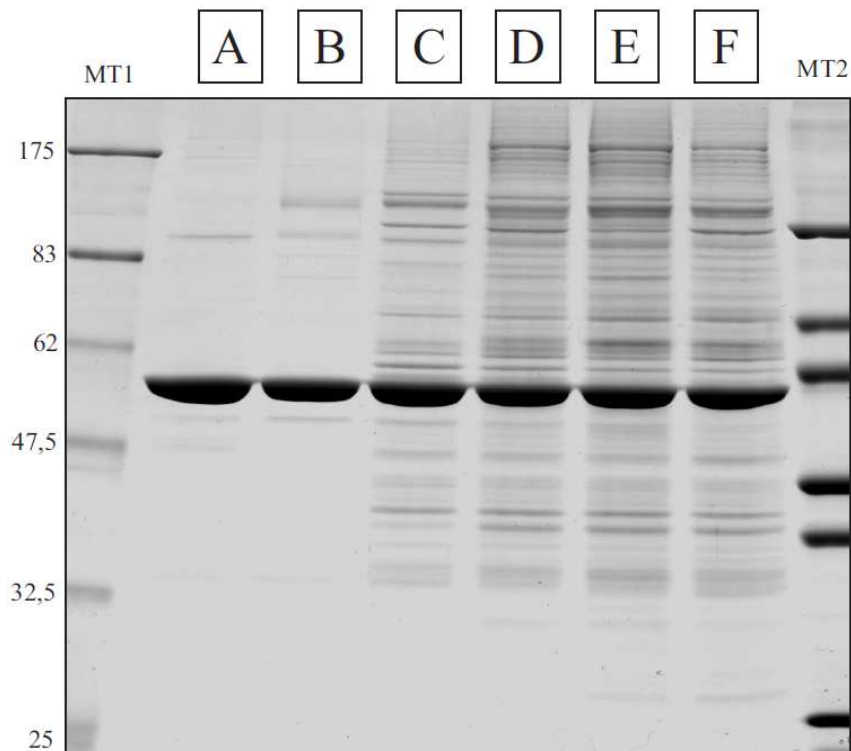


Figure 22 : Gel d'électrophorèse 1D, coloré au bleu colloïdal. Chaque piste correspond aux protéines issues de la purification MS2 du complexe obtenu par incubation d'un extrait nucléaire de cellules ES à J0 réalisée avec l'ARN présenté en figure 21.

L'étude visuelle de ce gel permet de dire que les compositions de chaque éluat ne sont pas identiques. En effet, les pistes B (1 répétition) et C (2 répétitions) présentent un nombre de bandes moins important que pour les pistes D (4 répétitions) et E (8 répétitions, ARN XIM). Pour avoir une vue la plus exhaustive du contenu protéique de ces échantillons, la découpe de ce gel a été faite de façon systématique.

Pour l'interprétation des données MS/MS, nous avons choisi de travailler avec une banque protéique target-decoy restreinte aux taxonomies *Rattus* et *Mus*, les données protéiques sont issues de la banque UniprotKB (version du 07/08/2008, 157674 séquences). Le logiciel Scaffold est utilisé pour cette étude afin d'automatiser les interprétations et permettre le calcul du taux de faux positifs obtenu. Les identifications sont réalisées avec 1 peptide dont le score minimal est de 35. Cela permet d'obtenir un taux de faux positifs inférieur à 2%. La banque et les critères de validation ont été choisis afin d'obtenir un maximum d'identifications pour être en accord avec les critères actuels de publication [Carr et al., 2004; Bradshaw et al., 2006].

77 protéines uniques sont présentes seulement dans les pistes D et E et non identifiées dans les contrôles A et F (Annexe 3). Le contrôle A (10 protéines) n'a pas permis d'identifier un grand nombre de protéines contrairement au contrôle F (87 protéines). Les identifications des pistes B et C ne sont pas présentées puisqu'il s'agit majoritairement de protéines également présentes dans les pistes D et E.

Parmi le grand nombre de protéines identifiées, on peut trouver plusieurs protéines ayant un rapport avec l'inactivation transcriptionnelle. Il s'agit par exemple du complexe corépresseur transcriptionnel composé de la protéine Sin3a (Paired amphipathic helix protein Sin3a, Q60520; 1 peptide), la protéine BRM1L (Breast cancer metastasis-suppressor 1-like protein, Q3U1T3; 1 peptide) et la protéine HDAC1 (Histone deacetylase 1, O0910; 1 peptide). Seuls les construits ARN ayant au minimum 1 tige-boucle (4 répétitions) ont permis d'identifier ces protéines. Cette structure secondaire semble être nécessaire pour l'inactivation. Deux protéines, identifiées dans le cas où l'ARN possède 8 répétitions, font partie du complexe Polycomb 2 (PRC2). Il s'agit de la protéine EZH2 (Enhancer of zeste homolog 2, Q61188; 2 peptides) et de la protéine RBBP4 (Retinoblastoma binding protein 4, Q60972; 2 peptides). Ce complexe protéique catalyse la méthylation des histones, marques épigénétiques qui apparaissent juste après l'accumulation de l'ARN Xist sur le chromosome X en voie d'inactivation [Schoeftner et al., 2006]. Cela met en évidence que le nombre d'éléments répétés et la structure 2D qu'ils peuvent adopter ensemble sont cruciaux pour la fixation du complexe PRC2 auquel ces protéines appartiennent.

### **3.c. Etude de la composition du complexe selon le degré de différenciation**

La structure secondaire en tige-boucle et les 8 répétitions apparaissent comme nécessaires pour la mise en place de l'inactivation du chromosome X. La seconde partie de l'étude consiste à identifier les partenaires protéiques impliqués dans l'inactivation à des stades de différenciation différents avec un ARN contenant 8 régions "A-repeats" (ARN XIM). Pour cela, des extraits nucléaires de cellules ES à différents stades de différenciation (J0, J2 et J4), sont incubés avec l'ARN XIM et avec l'ARN MS2. Le gel 1D obtenu est présenté en figure 23.



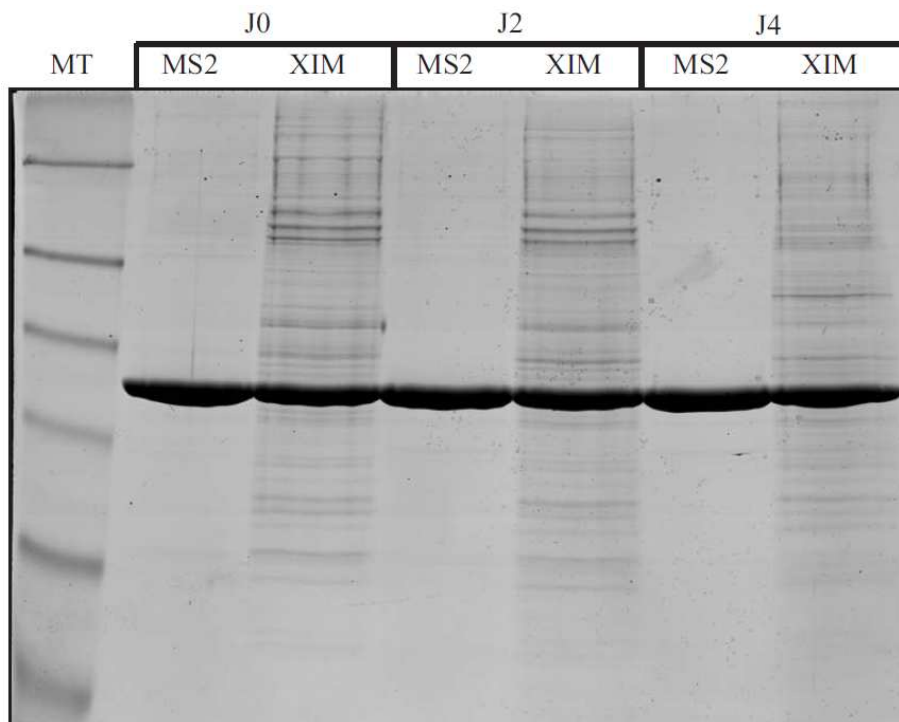


Figure 23 : Gel d'électrophorèse 1D, coloré au bleu colloïdal. Chaque piste correspond aux protéines issues de la purification MS2 du complexe obtenus par incubation d'un extrait nucléaire de cellules ES à J0, J2 et J4 réalisée avec l'ARN MS2 ou XIM.

Les profils des pistes correspondant aux éluions de purification réalisés avec XIM J0, XIM J2 et XIM J4 sont similaires, bien que dans les hautes masses on puisse observer quelques bandes plus intenses pour XIM J0 et XIM J2. Les pistes ont été découpées de façon systématique, et traitées selon la méthode précédemment décrite. L'interprétation des données est réalisée strictement de la même façon et avec les mêmes critères que pour la première partie. 182 protéines uniques, absentes des contrôles, ont été identifiées pour l'ensemble des pistes XIM J0, XIM J2 et XIM J4 (Annexe 4).

Nous nous sommes intéressées particulièrement aux protéines identifiées à J0 car elles sont les plus nombreuses et sont potentiellement impliquées dans l'initiation de l'inactivation. La protéine Sin3a (Q60520 ; 2 peptides), ayant une activité corépresseur transcriptionnel, est retrouvée dans ces échantillons. Elle avait été également identifiée dans la première partie de l'étude. D'autres protéines intéressantes sont également identifiées : Lin28A (Lin-28 homolog A, Q8K3Y3 ; 1 peptide), impliquée dans la régulation temporelle du développement, KSRP (Far upstream element binding protein 1, Q91WJ8 ; 7 peptides), protéine capable de se lier à des séquences riches en résidus U (régions situées entre les "A-repeats") et HRNPD-L (Heterogeneous nuclear ribonucleoprotein D-like, Q9Z130 ; 5 peptides), protéine promouvant la répression de la transcription.

Parmi les protéines identifiées, deux protéines RBBP7 (Retinoblastoma binding protein 7, Q60973, 2 peptides) et SUZ12 (Suppressor of zeste 12 protein homolog, Q80U70, 1 peptide) appartiennent au complexe PRC2. Ce complexe protéique, important pour l'inactivation, est composé de 5 protéines : SUZ12, EZH2, RBBP4, RBBP7, Polycomb protein EED. Deux de ces protéines (EZH2 et RBBP4) ont été identifiées dans la première partie de l'étude dans la piste correspondant à l'incubation de l'ARN XIM avec des cellules ES J0 (Figure 24).

Les deux expériences nous permettent de mettre en évidence la présence de 4 des 5 protéines du complexe PRC2. Les 4 protéines SUZ12, EZH2, RBBP4, RBBP7 sont présentes uniquement à J0. Aucun peptide ne correspondant à l'une des 5 protéines du complexe PRC2 n'est identifié dans les pistes XIM J2 et XIM J4. Elles ont été identifiées avec peu de peptides (Tableau 2). Les protéines ont été identifiées dans des bandes de gel contenant un grand nombre de protéines : 14 protéines dans la bande contenant SUZ12 et 6 protéines dans celle correspondant à EZH2.

Les protéines RBBP4 et RBBP7, respectivement 47655 Da et 47790 Da, sont élués au même niveau que la protéine de fusion MBP-MS2 (Figure 24). Cette protéine, support du système d'affinité MS2 utilisé pour ces expériences, est en large excès par rapport aux protéines de l'extrait nucléaire. Il est donc difficile de détecter un grand nombre de peptides pour identifier ces protéines. De plus, RBBP4 et RBBP7 ont des séquences très peu différentes (Figure 25-A). Trois peptides ont été identifiés. Le peptide 144-156 TPSSDVLVFSYTK (Figure 25-B) est un peptide commun aux deux séquences. Les deux autres peptides (Figure 25-C et D) ont permis de différencier les deux protéines.

Le complexe PRC2, se localisant sur le chromosome X, est identifié uniquement pour un ARN possédant 8 répétitions de "A-repeats" et pour un extrait nucléaire de cellule ES murin en tout début de différenciation (J0).

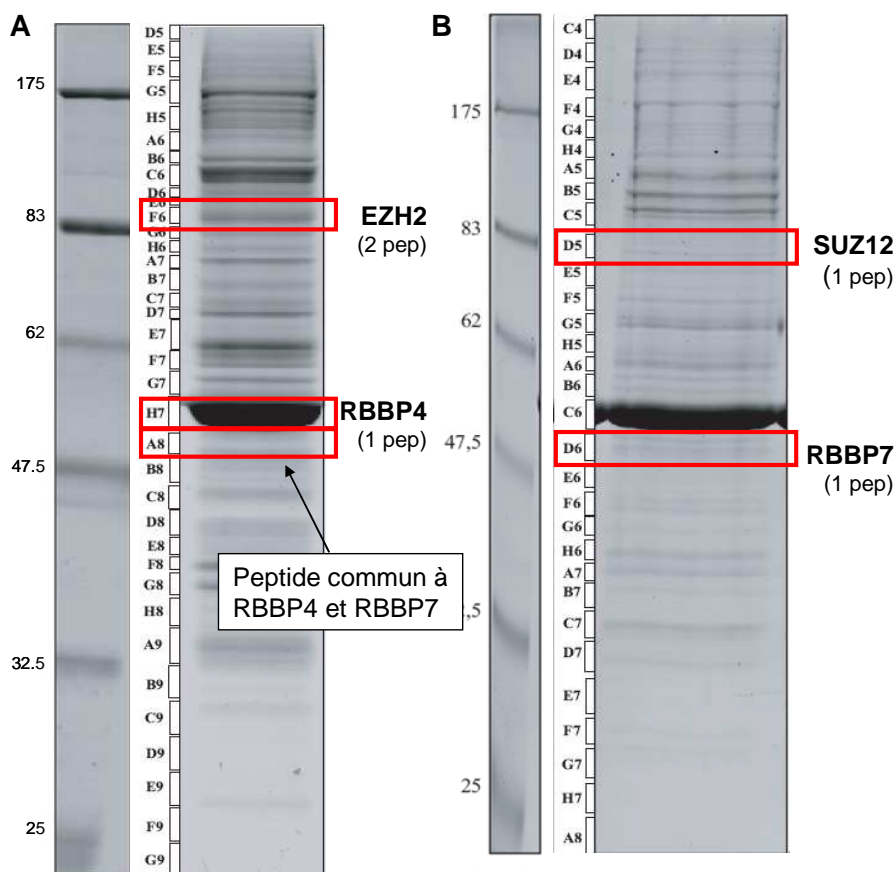


Figure 24 : Localisation sur les gels 1D des protéines du complexe PRC2 identifiées par spectrométrie de masse (incubation de l'ARN XIM avec l'extrait nucléaire de cellules ES J0 suivi d'une purification MS2). Les gels A et B sont des répliqués biologiques.

Tableau 2 : Peptides identifiés appartenant aux protéines composant le complexe PRC2.

Protéine	N° Accession	Peptide	M Obs. (m/z)	Charge	Séquence
SUZ12	Q80U70	269-306	643.8	2	TFVAQMTVFDK
EZH2	Q61188	82-99	1010.0	2	EC(Cam)SVTDLDFPAQVIPLK
		442-456	888.4	2	VLIGTYDYNFC(Cam)AIAR
RBBP4	Q60972	2-15	804.3	2	A(Ac)DKEAAFDDAVEER
		144-156	736.3	2	TPSSDVLVFDTK
RBBP7	Q60973	5-14	650.7	2	EM(Ox)FEDTVEER
		144-156	736.3	2	TPSSDVLVFDTK

A

```

RBBP4_MOUSE MADKEAAFDDAVEERVINEEYKIWKNTPFPLYDLVMTHALEWPSLTAQWLDPVTRPEGKDFSIHRLVLGTH
RBBP7_MOUSE MASKEM-FEDTVEERVINEEYKIWKNTPFPLYDLVMTHALQWPSLTVQWLPEVTKPEGKDVALHWLVLGTH

RBBP4_MOUSE TSDEQNHLVIASVQLPNDDAQFDASHYDS EKGEFGGFGSVSGKIEIEIKINHEGEVNRARYMPQNPCIAT
RBBP7_MOUSE TSDEQNHLVVARVHIPNDDAQFDASHCDS DKGEFGGFGSVTGKIECEIKINHEGEVNRARYMPQNPHIAT

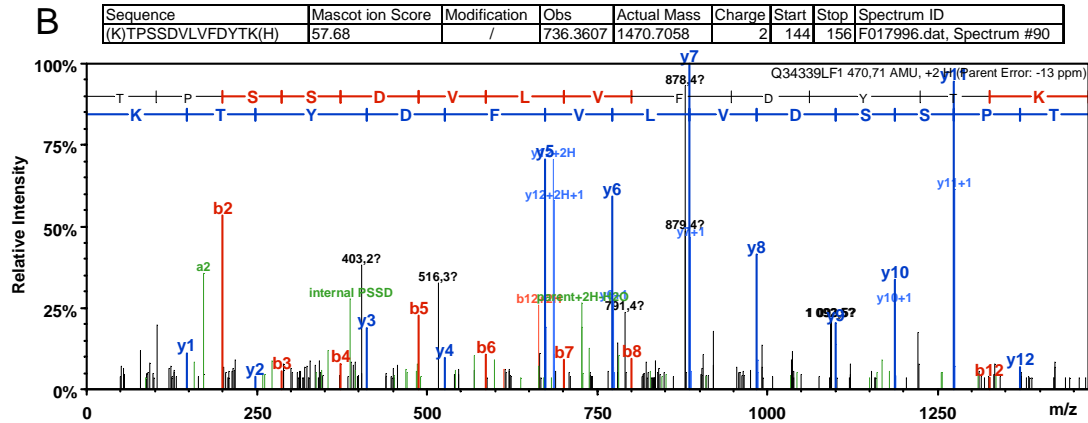
RBBP4_MOUSE KTPSSDVLVFDYTKHPKPPDPSGECNPDRLRGRHQKEGYGLSWNPNSLGHLLSASDDHTICLWDIASAVPKK
RBBP7_MOUSE KTPSSDVLVFDYTKHPKPPDPSGECNPDRLRGRHQKEGYGLSWNSNLGHLLSASDDHTVCLWDINAGPKE

RBBP4_MOUSE GKVVDAKTIPTGHTAVVEDVSWHLLHESLFGSVADDQKLMIDTRSNNTSKPSSHVDAHTAEVNCLSFNPY
RBBP7_MOUSE GKIVDAKAIPTGHSVAVVEDVAWHLHESLFGSVADDQKLMIDTRSNNTSKPSHLVDAHTAEVNCLSFNPY

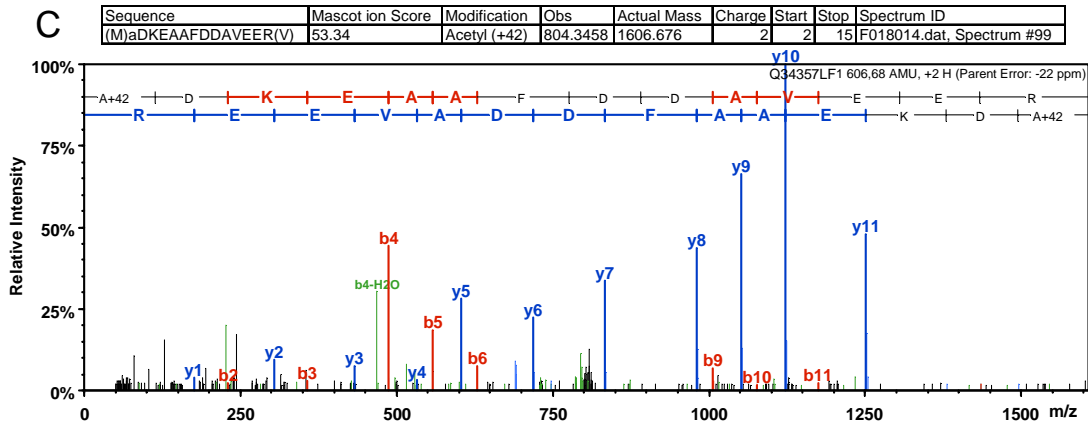
RBBP4_MOUSE SEFIIATGSADKTVALWDLRNLKLLKLSFESHKDEIFQVQWSPHNETILASSGTRRLNVDLSKIGEEQS
RBBP7_MOUSE SEFIIATGSADKTVALWDLRNLKLLKLSFESHKDEIFQVHWSPHNETILASSGTRRLNVDLSKIGEEQS

RBBP4_MOUSE PEDAEADGPELLFIH GGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDEDDPEGSDP-EGQGS
RBBP7_MOUSE AEDAEADGPELLFIH GGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDESDVTASELEGQGS
  
```

B



C



D

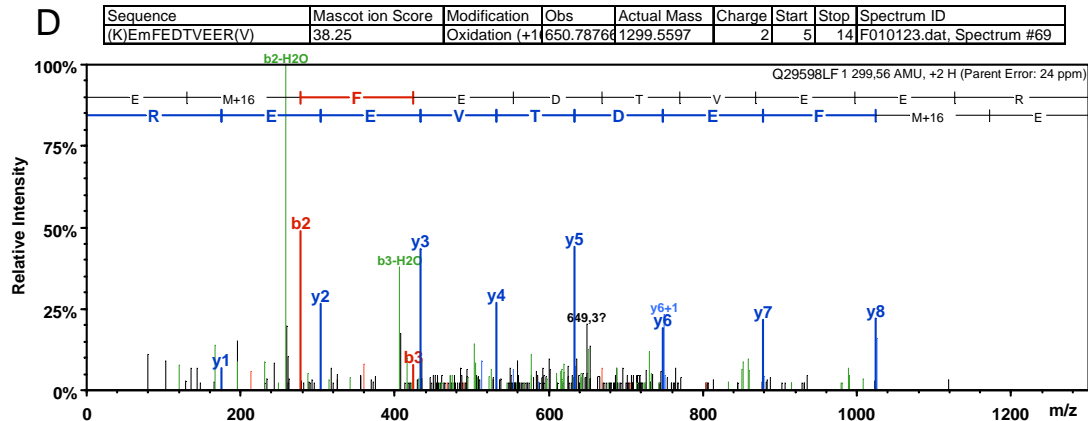


Figure 25 : Identification des protéines RBBP4 et RBBP7. A) Alignement des séquences de RBBP4 et RBBP7. B) Spectre de fragmentation de l'ion m/z 736.3, doublement chargé, correspondant au peptide commun aux deux protéines. C) Spectre de fragmentation de l'ion m/z 804.3, doublement chargé, correspondant au peptide spécifique à RBBP4. D) Spectre de fragmentation de l'ion m/z 650.7, doublement chargé, correspondant au peptide spécifique à RBBP7.

### **3.d. Conclusion**

L'approche protéomique a donc permis de mettre en évidence le complexe PRC2, composé de 5 protéines, partenaires d'interactions des régions " A-repeats" de l'ARN Xist. L'importance de la structure secondaire de l'ARN et du nombre d'éléments "A-repeat" a été montré pour la constitution des complexes protéiques nécessaires pour l'inactivation.

L'approche protéomique a permis d'identifier un grand nombre de protéines candidates à l'interaction des régions " A-repeats" de l'ARN Xist selon les stades de différenciation de la cellule. De nombreuses interactions restent donc à étudier.

Ce travail a donné lieu à une publication dont la version ci-jointe a été soumise dans le journal PLoS Biology.

## 2D structure of the A region of Xist RNA and its implication for PRC2 association

Maenner Sylvain<sup>1</sup>, Magali Blaud<sup>1</sup>, Laetitia Fouillen<sup>3</sup>, Anne Savoye<sup>1</sup>, Marchand Virginie<sup>1\*</sup>, Agnès Dubois<sup>2</sup>, Sarah Sanglier-Cianférani<sup>3</sup>, Alain Van Dorselaer<sup>3</sup>, Philippe Clerc<sup>2</sup>, Philip Avner<sup>2</sup>, Athanase Visvikis<sup>1</sup> and Christiane Branlant<sup>1</sup>

*1 AREMS, Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France*

*2 Génétique Moléculaire Murine, Institut Pasteur and CNRS2578 25 Rue du Docteur Roux, 75015 Paris, France*

*3 Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubert Curien, Département des Sciences Analytiques, Université de Strasbourg, CNRS UMR 7178, ECPM, 25, rue Becquerel, 67087 Strasbourg, France*

*\* Present adress : Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse, 1, 69117 Heidelberg, Germ*

+ Corresponding author

Christiane BRANLANT

ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale,  
Nancy université, UMR 7214 CNRS-UHP 1, Faculté des sciences et techniques, BP  
70239

54506 Vandoeuvre-lès-Nancy Cedex, France

Tel: 33 3 83 68 43 03 Fax: 33 3 83 68 43 07

E-mail: [christiane.branlant@maem.uhp-nancy.fr](mailto:christiane.branlant@maem.uhp-nancy.fr)

## **ABSTRACT**

Xist RNA is crucial for X chromosome inactivation in the female cells of placental mammals. The conserved Xist A region which contains eight repeated elements plays an essential role in this process. Little however is known on its structure and mechanism of action. We have performed the first experimental analysis of its structure in solution, using chemical and enzymatic probes and FRET experiments. Both mouse and human A regions were found to form two long stem-loop structures each containing 4 repeats. In contrast to previous predictions, interactions take place between repeats and between repeats and spacers. Mass spectrometry and antibody based analyses of affinity-purified RNP complexes showed that the 4-repeat structures can recruit components of the PRC2 complex upon incubation with mouse ES cell nuclear extract. However, association of Suz12 was more efficient when the entire A region was used.

## Introduction

In mammals, the transcriptional silencing of one of the two X chromosomes in female cells (X-chromosome inactivation, XCI) ensures sex chromosome dosage compensation<sup>1</sup>. Once acquired early in development, the inactivated state is faithfully inherited through successive cell divisions. XCI initiation is associated with increased Xist RNA transcription. Whilst, first retained near its transcription site, Xist RNA then spreads along the entire X chromosome from which it has been transcribed<sup>2-5</sup> whilst a series of epigenetic marks, which include the repressive histone modifications H3K27me3, H3K9me3, are recruited to the presumptive inactive X chromosome. Xist RNA is a long non-coding RNA, (17 kb in length in the mouse) which is capped, spliced and polyadenylated. Little is known about its structure and mechanism of action.

The Xist gene has a complex origin. It includes degenerated pieces of an ancient protein gene *Lnx3* as well as genomic repeat elements derived at least in part from transposon integration events<sup>6,7</sup>. The most conserved Xist RNA regions correspond to repeat elements (denoted A to E in mouse<sup>8</sup>) which are organized as tandem arrays. The A region (positions 292 to 713 in mouse, accession no. gi|37704378|ref|NR\_001463.2|<sup>2</sup>) and 350 to 770 in human (accession no. gi|340393|gb|M97168.1|<sup>5</sup>), is the most highly conserved of the repeat regions and is critical for initiation of XCI. The observation that female mouse embryos carrying a mutated Xist<sup>ΔA</sup> gene inherited from males are selectively lost during embryogenesis underlines the importance of this element<sup>9</sup>. Recent data has shown that an early event in silencing is the formation of a Xist RNA compartment and that the A region whilst not necessary for formation of this compartment is needed for relocation of X linked genes into this territory<sup>10</sup>. Over-expression of a Xist<sup>ΔA</sup> RNA in transgenic mouse ES cells indicates that the A region whilst not necessary for Xist coating, is implicated in the recruitment of the PRC2 complex<sup>11-16</sup>. The PRC2 complex contains the Suz12, Eed, Ezh2 and Rbap46-48 proteins<sup>17,18</sup>. Eed and Suz12 have been proposed to bind nucleic acids<sup>19,20</sup>, whereas, Rbap46-48 may interact with nucleosome protein components<sup>17</sup>. Lysine 27 tri-methylation of histone H3 is catalysed by Ezh2<sup>12,14</sup> and both Eed and Suz12 are required for this activity<sup>20</sup>. Recently a short 1600 nucleotide-long RNA which contains the A region at its 5'



extremity was suggested to be expressed early in XCI initiation and to bind the PRC2 complex<sup>21</sup>.

Since the function of Xist RNA is expected to depend on its 2D structure, studies aimed at establishing the 2D structure of the Xist A region have considerable interest. Based on nucleotide sequence of the A region and computer prediction, Wutz and colleagues have proposed that each repeat forms two short stem-loop structures<sup>11</sup>. Recent NMR analysis confirmed that one of these stem-loop structures can be formed *in vitro* by an RNA molecule bearing a single copy of the mouse repeat A sequences<sup>22</sup>. In Xist RNA, the repeat sequences are however separated by long spacer regions (21 to 48 nt long for mouse). Since current models fail to take account of this sequence complexity an experimental analysis of the entire A region was thought likely to provide valuable information on the structure of the A region. As conventional probing experiments are however, hindered by the presence of the repeated sequences and long U tracks, we applied a combined approach exploiting both chemical and enzymatic probing of RNA structure in solution and FRET experiments using fluorescent oligonucleotide probes complementary to different parts of the A region.

Using this dual approach we could show that repeats in the A region interact with each other to form long irregular stem-loop structures. Such inter-repeat interactions appear to be required for the binding of the various components of the PRC2 complex. We identified the minimal number of repeats necessary for such binding. The implications of our results within the wider context of X-inactivation and of the XCI mechanism(s) underlying silencing are discussed.

## Results

### Probing of mouse and human A region 2D structures

We analysed in parallel both the entire mouse and human A regions, as the sequence divergence of the inter-repeat linking sequence between mouse and human was expected to provide insight into how the spacer regions might influence the A repeat structure. The specific primers for the two A regions are listed in **Supplementary Table 1**. To test if the A region interacts with neighbouring Xist RNA sequences, an RNA that contained only the mouse A region (positions 277 to 760 – in mouse Xist RNA) and a larger RNA including the sequence extending from positions 1 to 1137 were studied in parallel by limited enzymatic digestion. Very similar digestion profiles (**Fig. 1a and b**) were obtained for the two RNAs when digestions were performed on the T7 RNA transcripts after folding under the conditions outlined in Materials and Methods. We conclude that the A region probably folds on itself without major interaction with other upstream and downstream Xist sequences. Hence, our subsequent analyses of the 2D structure of the Xist A region were carried out in the absence of flanking sequences (positions 227 to 760 and 330 to 796 for the mouse and human A regions, respectively). Each enzymatic digestion and chemical modification assay was carried out in duplicate using different transcript preparations and each extension analysis repeated 2 or 3 times for each primer. Representative examples of the primer extension analyses are provided in Figure 1 and **Supplementary Fig 1a.-e**.

### M-fold assisted modelling of the mouse A region 2D structure

The structure proposed by Wutz and colleagues, in which each of the repeats fold into a double stem-loop structure, could not explain the numerous V1 RNase cleavages that we observed with both the mouse and human A regions<sup>11</sup> (**Fig. 2**). We explored the possibility that each repeat folds into a unique longer stem-loop structure. Such folding was similarly unable to explain V1 RNase cleavages (**Supplementary Fig. 2**). We conclude that the 2D structure may involve interactions between repeats and spacers and inter-repeat interactions. There is however a multitude of potential ways for duplex formation between repeats. Our design of the putative structure was orientated by the detection of six successive strong V1 RNase cleavages in the central poly A sequence (positions 550 to 555), suggesting the

involvement of this segment in a helical structure. The strong modification by DMS of a sequence immediately downstream (positions 556 to 561), was an indication for a single-stranded state. One possible explanation for these data was the formation of a central stem-loop structure called SLS2M with a U track on one strand and an A track on the other strand (**Fig. 3c**). Formation of this central stem-loop structure was subsequently imposed as a constraint when exploring the possible folding of the mouse A region. This excluded structures in which, two successive repeats would interact with each other (Model 1, **Fig. 3a**), since in this case, the entire polyA region would interact with the polyU track located upstream of repeat 3, which was not in agreement with the probing data (**Fig. 3a**). Another possible structure involved formation of an interaction between the 5' and 3' halves of the A region. This would generate a very long irregular stem-loop structure with an A rich terminal loop (Model 2, **Fig. 3b**). An alternative structure involved the folding on themselves of the 5' and 3' parts of the A region, with SLS2M in between (Model 3, **Fig. 3c**). Several other alternative pairings of the repeats were also explored - none fitted the chemical and enzymatic data perfectly.

The notion of independent folding of the 5' part of the A region (positions 318 to 521 in mouse RNA), was supported by M-Fold analysis of this segment which identified a highly stable long irregular stem-loop structure, SLS1M ( $\Delta G = -41.96$  kcal/mol at 0°C in 3 M NaCl), in which repeat 1 interacts with repeat 4 and repeat 2 interacts with repeat 3. It is the most thermodynamically stable structure proposed for this 5' segment, and was predicted irrespective of whether the experimental data were introduced as a constraint in the M-Fold search. In SLS1M, each repeat interacts both with another repeat and with a spacer segment, increasing the stability of the overall structure. Similarly, M-fold analysis of the 3' part of the mouse A region, suggested that repeat 5 interacts with repeat 8 and a spacer region and repeat 6 with repeat 7 and a spacer region. The resulting SLS3M structure was predicted as the most favourable structure by M-Fold when the experimental data were incorporated as a constraint. The overall predicted three stem-loop structure (Model 3) has a low calculated free energy (-77.76 kcal/mol) and has a better fit to the experimental data than Models 1 and 2 (**Fig. 3a and b**) suggesting that, in solution, Model 3 is the most likely structure among the numerous possibilities.

**The possibility to form structure 3 is phylogenetically conserved**

If a structure has biological relevance, it is generally conserved throughout evolution. Therefore, we tested whether the most favourable structures identified for the mouse A region were relevant to the human A region in solution. The sequence of the human A region differs from that of mouse by the presence of an additional repeat 5 and the absence of a long central polyA region. Experimental data (**Fig. 4 and 5** and **supplementary Fig. 3 a-e**), suggested that the central repeat 5 forms a central stem-loop structure (SLS2H). Based on this, structures similar to mouse Models 2 and 3 could be proposed for the human A region which involve either a long irregular stem-loop structure including all the repeats (Model 2), or a three stem-loop structure (Model 3) with repeats 1 to 4 forming a first stem-loop structure (SLS1H), repeat 5 alone a second stem loop (SLS2H) and repeats 6 to 9 a third stem-loop structure (SLS3H). As for the mouse A region: i) A 2D structure in which each repeat interacts with its immediate downstream repeat (repeats 1, 3, 6 and 8 with repeats 2, 4, 7 and 9, respectively, Model 1) was not supported by the probing data (segment 688 to 696) (**Supplementary Fig. 4**), ii) M-Fold analysis of the 5' portion of the human A region (positions 370 to 530) either with or without the experimental data as a constraint, identified SLS1H as the most stable structure ( $\Delta G = -42.70$  kcal/mol) (**Fig. 3c**), iii) SLS3H was retained as the most stable structure for the 3' part of the A region, when the experimental data were added as a constraint to an M-Fold search, iv) the 3 stem-loop structure corresponding to Model 3 ( $\Delta G = -86.6$  kcal/mol) had the best fit with probing data compared to the other 2D models. Further support for Model 3 was provided by our observation of identical patterns of enzymatic cleavage for the entire human A region and for the isolated SLS1H portion (**Fig. 4**).

The maintenance of interactions between both spacers and repeats during mammalian evolution of the A region implies that the nucleotide sequences involved in these interactions were either conserved or subjected to compensatory base changes. This was confirmed by the alignment of the mouse, human, orangutan, baboon, lemur, dog, rabbit, cow, elephant A region sequences (**Supplementary Fig. 5 and 6**). Nucleotide sequence conservation extends out beyond the repeats themselves for the majority of the repeats.

### **FRET experiments bring additional data in favour of Model 3**

Three oligonucleotide pairs (P1-P5, P2-P4 and P6-P7) were retained in order to test Model 3 by FRET experiments (**Fig. 6**). This Model predicts that the P1-P5 and P2-

P4 pairs of oligonucleotides interact with the single-stranded segments which border the helix formed by repeats 1 and 4, whilst the P6-P7 pair of oligonucleotides should interact with the single-stranded segments bordering the helix formed by repeats 5 and 8. A marked FRET effect would therefore be expected for these three oligonucleotide pairs if the A region was folded as in Model 3. The distance between the fluorophores of these 3 pairs of oligonucleotides would on the other hand, be expected to be much larger if region A was folded as in structures 1 or 2. Whilst tertiary structural interactions might decrease the distances, a lower level of FRET would still be expected to be observed for the three pairs of oligonucleotides if the A region folded according to structures 1 or 2 (**Fig. 6**). The P1 and P6 oligonucleotides bind to two single-stranded segments which flank the helix formed by repeats 1 and 8 in structure 2. A strong FRET effect for P1 and P6 would therefore be expected if the A region was folded according to structure 2. To monitor the level of FRET obtained for oligonucleotides bordering a helix, we used the short R1-2 transcript containing repeats 1 and 2 and their bordering sequences which adopts a single unique 2D structure and the P1-P3' oligonucleotide pair (**Supplementary Fig. 7**). Other controls exploited the P3-P6 and P3-P5 pairs which were not expected to be in close proximity in any of the 3 proposed structures (**Fig. 6**). The oligonucleotide pairs used are shown in Fig. 6, along with examples of typical fluorescence intensity spectra recorded in FRET experiments for the P2-P4 and P3-P6 pairs (**Fig. 6d**). High FRET signals in the range of 50% were obtained for the P1/P5, P2/P4 and P7/P6 oligonucleotide pairs, whilst lower FRET signals were observed for the P1-P6 (35%) and especially the P3-P6 and P3-P5 oligonucleotide pairs (25 and 22 %, respectively). This is compatible with a large part of the molecules being folded in solution into structure 3. Folding of a large number of molecules into structure 2 would have led to a strong FRET signal for the P1-P6 pair and lower signals for the 5 other pairs, which was not observed. The strong FRET effects obtained for the P1/P5, P2/P4 and P7/P6 oligonucleotide pairs argues strongly against folding according to structure 1. Based on our FRET data, we conclude that folding predominantly occurs according to Model 3.

### **Recruitment of the PRC2 complex by the A region**

Previous studies have shown that the PRC2 complex interacts with the Xi<sup>12,14,15,23</sup> and the A region has been proposed to recruit the PRC2 complex through the Ezh2

subunit which would act as an RNA-binding subunit<sup>21</sup>. We wished to explore further the binding of the PRC2 complex to the A region in the light of our structural data. In particular, we were interested in determining how many A region repeats were required to bind the individual Eed, Ezh2, RbAp46, RbAp48 and Suz12 components of the PRC2 complex.

We initiated a proteomic approach based on affinity chromatography purification of complexes formed upon incubation of *in vitro* transcribed Xist A region RNAs with nuclear extracts, followed by protein identification by mass spectrometry and Western blot analysis. Mouse ES cells are a widely exploited model for the study of XCI initiation and we reasoned that as Xist RNA acts as an initiator of XCI, proteins which have to interact with this RNA to ensure early Xist functions, should already be present in the nuclear extract of ES female mouse cells prior to differentiation. We used a control RNA containing only the three MS2 protein binding sites and tested five RNAs containing different segments of the mouse Xist A region flanked by the three MS2 binding sites at their 3' end (**Fig. 7a**). These RNAs denoted as 1R/MS2, 2R/MS2, 4R/MS2, Aregion/MS2 and HIV/MS2 contained respectively: repeat 1 without any neighbouring sequence, repeats 3 and 4 and their bordering spacers (positions 401 to 552 in mouse Xist RNA), the SLS1M stem-loop structure, the entire A region and a fragment of HIV-1 RNA (positions 5338 to 5514 in the BRU RNA) used as a negative control.

In order to get an idea of the proteins capable of associating with the entire A region, the proteins bound to purified complexes formed on the Aregion/MS2 RNA were analysed by mass spectrometry. Among numerous proteins detected, were protein PTB and components of the PRC2 complex (Ezh2, RbAp46, RbAp48 and Suz12) (**Supplementary Fig. 8**). We then evaluated by western blot experiments, the relative amounts of each of the PRC2 components in RNPs formed by the various RNAs tested. Whilst Eed, Ezh2 and PTB were detected in complexes formed on RNAs containing two or more repeats, binding of RbAp46 and RbAp48 was detected only when using RNAs with at least 4 repeats and Suz12 when the entire A region was used (**Fig. 7c**). The control HIV-1 RNA bound none of these proteins (**Fig. 7b**).

To explore further these data, we performed a series of experiments in which fragments of the A region were transcribed *in vitro* as radio-labelled RNA without MS2 fusion and these RNAs were incubated with mouse ES nuclear extracts. Three

distinct RNAs (the complete A region, 4R and 2R RNAs, **Fig. 7 d and e**) were used for these experiments.

In confirmation of the possible interaction of Eed with an RNA containing only two repeats, trace amounts of 2R RNA were retained on the beads when an anti-Eed antibody was used. Only complexes containing the 4R RNA or the entire A region were retained when anti-Suz12, anti-Ezh2, anti-RbAp46 and anti-RbAp48 antibodies were used. These observations confirmed the importance of the corresponding regions for association of these proteins. Clearly, however, higher amounts of the entire A region compared to 4R RNA were bound when the anti-Suz12 and Ezh2 antibodies were used.

We conclude that whilst some segments of the A region allow the binding of particular PRC2 components, the entire A region is required for efficient association of this complex.

## Discussion

The A region of Xist RNA plays an essential role in the X inactivation process. Here, we show that *in vitro*, the repeated elements in the A region of both mouse and human Xist RNAs interact together and with the intervening spacer regions. This leads to the formation of peculiar long irregular stem-loop structures containing 4 repeats and long U rich terminal and internal loops. Our proteomic analyses suggests that these 4-repeat structures may correspond to functional modules initiating the assembly of the PRC2 complex.

### A new conception of the 2D structure of the A repeat region

Until now, both computer<sup>11</sup> and experimental analyses<sup>22</sup> of the possible 2D structure of the A region of Xist RNA have privileged the individual repeat as the unit of folding. However, the presence of long intervening spacer sequences between the repeats, suggests that these spacer sequences may participate in 2D structure formation, and points to the potential inadequacy of previous models. Our detailed chemical and enzymatic probing of the A region structure in solution involving the design of specific primers for reverse transcriptase extension analysis enabled us to identify for the first time the double-stranded and single-stranded segments making up the A region structure in solution. The data obtained clearly demonstrate that the repeats do not fold on themselves but rather fold one with the other.

Chemical and enzymatic probing of an RNA structure in solution often allows the building of a unique 2D structure model in agreement with the experimental data. Studies on the A region were however complicated by the high degree of sequence redundancy. Use of a recently proposed biophysical approach, based on FRET assays<sup>24</sup>, helped overcome these difficulties by providing information on the relative distances between the sequences flanking the various repeats. To our knowledge, up to now, this approach has only been used to confirm the 2D structure model of telomerase RNA<sup>24</sup>. This method, which involves the utilization of oligonucleotides carrying donor and acceptor fluorescent dyes complementary to single-stranded segments in the studied RNA, proved particularly well suited to the study of the A region, since our probing data identified several long single-stranded segments which were able to bind the oligonucleotide probes. Among the possible 2D structures for the A region, only one, structure 3, showed perfect agreement with the FRET data. Structure 3 which contains two long irregular stem-loop structures, each involving 4



repeats (4-repeat structure) also shows the best agreement with the chemical and enzymatic data. The two long stem-loop structures are separated by a short stem-loop structure, corresponding to a divergent region between mouse and human Xist RNAs. One repeat in this segment is common to all the sequenced Xist RNAs (**Supplementary Fig. 5**), except mouse RNA. In the latter, it is replaced by a polyA sequence forming a short stem loop with a polyU sequence. Interestingly, nucleotide sequence conservation in the A region extends to the spacer extremities, which contribute to the possibility of forming the 4-repeat structures (**Supplementary Fig. 5**).

Although, the presence of large internal and terminal loops decreases the stability of stem-loop structures, the predicted free energies of the two 4-repeat structures in both mouse and human RNAs have strongly negative values (between -33 and -45 kcal/mol), explaining why they are proposed by the M-Fold software. In addition, these 4-repeat structures may be stabilized *in cellulo* by RNA-protein interactions. Interestingly, protein PTB which contains 4 RNA Recognition Motifs (RRMs), that are each able to interact with UCUU(C), UUCUCU or CUCUCU sequences, showed high affinity for the A region in nuclear extract binding experiments<sup>25</sup> (**Fig. 7**). As UCUU motifs are present on each side of the large internal loops in the 4-repeat structures and in the terminal loop, one might imagine that interactions of the RRM of a single PTB molecule with these various segments may stabilize the 4-repeat structure as suggested by previously proposed models for protein PTB-RNA interaction<sup>26</sup>.

### **Possible functional implication of the 4-repeat structure**

Our adaptation of the affinity purification chromatography, originally developed for purifying spliceosome complexes<sup>27</sup>, to complexes formed upon incubation of different fragments of the A region with nuclear extracts prepared from undifferentiated mouse ES cells, coupled with mass spectrometry and Western blot analyses was powerful. Together with immunoselection assays performed on assembled RNP complexes, it revealed the capability of four components of the PRC2 complex to associate with an RNA corresponding to one of the 4-repeat structures formed by the A region. However, our observation that the entire A region is needed for efficient association of the Suz12 protein suggests a putative additional functional role for the entire A repeat region in either binding Suz12 or in stabilising the binding of Suz12 to the 4 repeat structure.

Whilst UV cross-linking of the RNP complexes formed with ES nuclear extract using the entire A region, has confirmed the direct binding of PTB to this RNA region, direct binding of components of the PRC2 complex was not detected (data not shown). Neither Ezh2 nor Eed, which were previously proposed to be recruited by Xist in an A region dependent manner<sup>21</sup>, were cross-linked in significant amounts, suggesting that their association with the A region is mediated via association with other nuclear proteins. Therefore, the peculiar SLS1 and SLS3 structures in the A region may be needed to recruit nuclear proteins which have an affinity for components of the PRC2 complex or to reinforce the RNA affinity for these components. Mass spectrometry analysis of RNP complexes formed with the entire A region showed that, in addition to components of the PRC2 complex, a large number of other nuclear proteins can associate with this RNA region. In further studies, it will be important to identify which of these proteins are required for PRC2 association and which ones bind directly to the A region structure.

Our results show that Suz12 has distinct association properties compared to other components of the PRC2 complex and suggest that it may require the presence of an additional RNA motif, which is formed only by the entire A region, or more simply require more than 4 repeats for efficient association with the RNA. The observation of Wutz and colleagues, (2002) that the presence of at least 5.5 repeats was needed to initiate inactivation<sup>11</sup> may be related to this finding. Additional support for the functional significance of the 4-repeat model comes from a reworking of data obtained by Wutz and colleagues, who tested the effect of a series of mutations within the A region on XCI initiation<sup>11</sup>. Our structural studies show that all the variants (XR, XSR, XCR), classed by Wutz *et al.* as active are able to form the 4-repeat structure, whereas the two inactive variants (XS1 and XNX) cannot (**Supplementary Fig. 9**).

### **Possible implication of PRC2 and other nuclear protein association with the A region**

Although it is clear that the A region is essential for the X inactivation process, the precise role and mechanisms involved in the action of the A region remain unclear. Its deletion was shown to block silencing but not the coating of the X chromosome by Xist<sup>11</sup>, an observation in agreement with a possible role of the A region in PRC2 recruitment. PRC2 is needed for apposition of some, but not all, of the epigenetic

marks which are specific features of silenced chromatin in general and the inactive X in particular (methylation of histone H3 at position 27)<sup>20</sup>. The association of PRC2 with long ncRNAs before transfer of the PRC2 complex to chromatin may be a general mechanism for chromatin silencing processes that depend on long ncRNAs. Both the HOTAIR and Kcnq1ot1 long ncRNAs, which are involved in gene silencing, were recently found to bind the PRC2 complex<sup>19,28</sup>.

Recruitment of PRC2 is a relatively early event in X inactivation<sup>14</sup> in agreement with a possible early association of this complex with Xist RNA prior to extensive Xist coating of chromatin. One could imagine that, PRC2 is associated with the chromatin upon Xist coating through its interaction with proteins bound to Xist RNA. Alternatively coating of the Xist RNP may facilitate PRC2 transfer to chromatin by interaction of some of the RNP components with proteins of the chromatin structure. Lee and colleagues recently reported the existence of the 1600 nucleotide long RepA RNA carrying the A region at its 5' extremity, which may be expressed prior to expression of the entire Xist RNA and has been reported to recruit the PRC2 complex in a very early step of XCI<sup>21</sup>. Independent confirmation of these findings will be of major importance to the field.

Screening of the numerous other proteins that we have found to be capable of association with the entire A region by mass spectrometry, for their eventual specific involvement in the recruitment of genes to the X-inactivation domain<sup>10</sup> or other early events characterising the onset of X-inactivation and silencing will be of potential major importance to our understanding of X-inactivation.

## **METHODS**

### **RNA preparation**

DNA fragments coding for the entire A regions of mouse and human Xist RNAs and their subfragments were PCR amplified using mouse or HeLa cell genomic DNA, and cloned into plasmid pUC18 under the control of a T7 promoter. RNAs were generated by run-off transcription with T7 RNA polymerase as previously described<sup>29</sup>. DNA templates were digested with RNase-free DNase I and RNA transcripts were purified on denaturing 3 to 8% polyacrylamide gels.

### **Enzymatic and chemical probing of RNA secondary structure**

RNA 2D structures in solution were probed according as follows<sup>29</sup>: 200 ng of transcripts dissolved at a 80 nM concentration in buffer D (20 mM Hepes-KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol) were renatured by 10 min heating at 65°C, followed by slow cooling at room temperature with addition of 1 µl of 62.5 mM MgCl<sub>2</sub>. After 10 min preincubation at room temperature, RNase T1 (0.02 or 0.0375 U/µl) or T2 (0.025 or 0.0375 U/µl) were added under conditions such that they cleaved single-stranded segments. V1 RNase (2.5x10<sup>-3</sup> or 5x10<sup>-3</sup> U/µl) was used to cleave double stranded and stacked residues. DMS [1 µl of a 1/4 or 1/8 (V/V) DMS/EtOH solution] was employed to modify single stranded A and C residues and CMCT (4 or 5 µl of a 180 mg/ml solution) to modify single stranded U and to a lower extent G residues. Reactions were stopped as described in<sup>29</sup>. Cleavage and modification positions were identified by primer extension<sup>29</sup>. Stable secondary structures having the best fit with experimental data, were identified with the Mfold software, version 8.1<sup>30</sup>. Probing data were introduced as a constraint in the search.

### **Steady-state fluorescence measurements**

Fluorescence spectra were recorded at 4°C, with an excitation wavelength of 515 nm and scanning from 500 to 750 nm (excitation and emission bandwidth of 3 nm). The procedure used was derived from<sup>24</sup>. The RNA and Cy3-oligonucleotide were mixed at a 1:1 molar ratio in 160 µl of 150 mM NaCl, 3.25 mM MgCl<sub>2</sub> and 15 mM Na citrate (pH 7.0), to a final concentration (0.38 µM) superior to the K<sub>d</sub>, incubated at 85°C for 5

min and slowly cooled at room temperature for 15 min. After 4 h of incubation at 4°C, the yield of oligonucleotide association was determined by electrophoresis in a non-denaturing gel. Fluorescence in the gel was measured with a Typhoon (9410) Healthcare scanner. When a satisfying yield of association was detected, the emission of the Cy3-labeled complex was measured on a fix spectrofluorometer (SAFAS). Ten spectra were averaged. Then, the Cy5-labeled oligonucleotide was added at a 1:1 molar ratio, and incubation carried out at 4°C for 4 h. Ten spectra were recorded and the Fluorescence Resonance Energy Transfer (FRET) for the Cy3-Cy5 pair was calculated taking account of the bound/unbound ratio of Cy3-oligonucleotide. Each FRET experiment was repeated three times using different batches of transcripts.

### **Purification of RNP by MS2 selection affinity**

The entire mouse A region and several fragments were cloned 3' to a T7 promoter and 5' to the MS2 tag present in plasmid pAdML3<sup>27,31</sup>.

Nuclear extracts from undifferentiated female ES cells (LF2) were prepared according to<sup>32</sup>, and dialyzed against buffer D. One hundred pmol of MS2-tagged RNAs were denatured, renatured, as described above, and incubated with a 5 fold molar excess of purified MS2-MBP fusion protein<sup>31</sup> at 4°C for 15 min. The RNA-MS2-MBP complexes formed were incubated with amylose beads (40 µl, GE Healthcare) equilibrated in buffer D for 2h at 4°C. After three washes with 500 µl of buffer D, 1 mg of nuclear extract in 150 µl of buffer D containing 5 µM of yeast tRNAs was added. After 15 min of incubation at 4°C with constant agitation, three successive washes were performed in Buffer D and RNP complexes eluted by incubation with 80 µl of Buffer D containing 10 mM maltose (30 min at 4°C). Half of the eluted RNP complex formed with the entire A region was fractionated by 10% SDS-PAGE for mass spectrometry analyses. For all the purified RNP complexes, 10% of the eluted material was used for Western blot analysis performed according to<sup>33</sup>.

### **Mass spectrometry analysis**

Each lane of the SDS-PAGE was cut into 2 mm sections, and proteins submitted to in-gel trypsin digestion. Analysis of extracted peptides was performed using nano-LC-MS-MS on a CapLC capillary LC system coupled to a QTOF2 mass spectrometer (Waters) according to standard protocols (**supplementary Fig. 8**). The MS/MS data

were analyzed using the MASCOT 2.2.0. algorithm (Matrix Science) for search against an in-house generated protein database composed of protein sequences of Rattus and Mus downloaded from UniprotKB <http://beta.uniprot.org/> (August 07, 2008) and protein sequences of known contaminant proteins such as porcine trypsin and human keratins concatenated with reversed copies of all sequences. Spectra were searched with a mass tolerance of 0.3 Da for MS and MS/MS data, allowing a maximum of 1 missed cleavage with trypsin and with carbamidomethylation of cysteines, oxidation of methionines and N-acetyl protein specified as variable modifications. Protein identifications were validated when one peptide had a Mascot ion score above 35. Evaluations were performed using the peptide validation software Scaffold (proteome Software).

### **Immunoselection of RNP**

RNA transcripts were dephosphorylated, 5'-end labelled with [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol), purified and quantified according to<sup>34</sup>. About 70 pmol of the RNA were denatured, renatured as described above and incubated with 30  $\mu$ g of nuclear extract for 30 min at room temperature with constant agitation. About 40  $\mu$ l of Protein G-sepharose beads suspension blocked with BSA (2  $\mu$ g) and coated for 2 h at 4°C with 10  $\mu$ l of each antibodies were incubated with the RNP complexes for 2 h at 4°C in 300  $\mu$ l of immunoselection buffer (150 mM NaCl, 10 mM Tris-HCl, pH 8.0, NP40 0.1%). Beads were washed three times for 10 min at 4°C with 750  $\mu$ l of immunoselection buffer containing 0.5% NP40. RNAs were phenol extracted, ethanol precipitated, fractionated on 7% polyacrylamide gel and analysed by autoradiography.

## **Acknowledgments**

Professor R. Lührman (Max Planck Institut, Goettingen) is thanked for his generous gift of plasmids pMBP-MS2 and pAdML3 and for help in the use of the MS2-MBP RNP purification approach. Professors I Motorine and C Aigueperse (AREMS, Nancy) are acknowledged for their advice in implanting this approach at Nancy. V Ségault (AREMS, Nancy) is thanked for advice concerning the immunopurification assays. S Mazerès (IPBS, Toulouse) is thanked for helping us to define the FRET experimental protocol. S Maenner, A Savoye and V Marchand were PhD fellows of the French Ministry “Enseignement Supérieur et la Recherche” and M Blaud was supported by a post-doctoral fellowship from the “Centre National de la Recherche Scientifique” (CNRS). L Fouillen acknowledges the CNRS, Région Alsace and AliX for financial support. Research funds to CB were obtained from the CNRS and the French Ministry of “Enseignement Supérieur et la Recherche”. PA was supported by recurrent funding from the CNRS and the Institut Pasteur and by an EU contract from the Epigenome NoE. Both CB and PA were supported by the French National Agency for Research (ANR) (contract N°ANR 07 BLAN 0047 102).

## References

1. Lyon, M.F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372-3 (1961).
2. Brockdorff, N. et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-26 (1992).
3. Borsani, G. et al. Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-9 (1991).
4. Cohen, H.R. & Panning, B. XIST RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1. *Chromosoma* **116**, 373-83 (2007).
5. Brown, C.J. et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-42 (1992).
6. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653-5 (2006).
7. Shevchenko, A.I. et al. Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res* **15**, 127-36 (2007).
8. Brockdorff, N. X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* **18**, 352-8 (2002).
9. Hoki, Y. et al. A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* **136**, 139-46 (2009).
10. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-37 (2006).
11. Wutz, A., Rasmussen, T.P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167-74 (2002).
12. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-95 (2003).
13. Plath, K. et al. Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol* **167**, 1025-35 (2004).
14. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-5 (2003).
15. Mak, W. et al. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Curr Biol* **12**, 1016-20 (2002).
16. Kohlmaier, A. et al. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* **2**, E171 (2004).
17. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-43 (2002).
18. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-905 (2002).



19. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-23 (2007).
20. Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**, 155-64 (2004).
21. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-6 (2008).
22. Duszczuk, M.M., Zanier, K. & Sattler, M. A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res* **36**, 7068-77 (2008).
23. Kalantry, S. et al. The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* **8**, 195-202 (2006).
24. Gavory, G., Symmons, M.F., Krishnan Ghosh, Y., Klenerman, D. & Balasubramanian, S. Structural analysis of the catalytic core of human telomerase RNA by FRET and molecular modeling. *Biochemistry* **45**, 13304-11 (2006).
25. Perez, I., McAfee, J.G. & Patton, J.G. Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* **36**, 11881-90 (1997).
26. Auweter, S.D. & Allain, F.H. Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol Life Sci* **65**, 516-27 (2008).
27. Deckert, J. et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**, 5528-43 (2006).
28. Pandey, R.R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232-46 (2008).
29. Mougin, A. et al. Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *Rna* **2**, 1079-93 (1996).
30. Jaeger, J.A., Turner, D.H. & Zuker, M. Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A* **86**, 7706-10 (1989).
31. Zhou, Z., Sim, J., Griffith, J. & Reed, R. Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A* **99**, 12203-7 (2002).
32. Dignam, J.D., Martin, P.L., Shastry, B.S. & Roeder, R.G. Eukaryotic gene transcription with purified components. *Methods Enzymol* **101**, 582-98 (1983).
33. Jacquenet, S. et al. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J Biol Chem* **276**, 40464-75 (2001).
34. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular cloning. A laboratory manual*, (Cold Spring Harbor Laboratory Press, New York, 1989).

## Figure legends

### **Figure 1 : Comparative analysis of enzymatic cleavages of mouse A region alone and in the 1137-nt long 5'-terminal segment of mouse Xist RNA, respectively.**

The two RNAs (a. for 5' terminal 1137-nt long RNA, b. for A region) were *in vitro* transcribed and renatured as described in Material and Methods, before being subjected to limited digestion with T1, T2 or V1 RNases under the conditions described in Materials and Methods. Extension analyses were performed using oligonucleotide 3866 (Table 1) as the primer. The resulting cDNAs were fractionated by electrophoresis on 7% denaturing polyacrylamide gel. Lanes U, G, C and A correspond to the sequencing ladder obtained with the same primer. Lane marked by Contr corresponds to primer extension analysis of the undigested RNA transcript. Nucleotide numbering on the lefthand side of the autoradiogram takes the first residue of mouse Xist RNA as residue 1. The sequences corresponding to repeats 3 and 4 are indicated by vertical bars on the right hand side of the autoradiograms.

### **Figure 2 : Schematic superposition of data obtained by enzymatic digestion and chemical modification of the mouse A region onto the two stem-loop structure models previously proposed for each A repeats<sup>11</sup>.**

Each of the 7 repeats as well as the 8<sup>th</sup> half repeat from the mouse A repeat region were folded according to the previously predicted two stem-loop structure<sup>11</sup>. T1, T2 and V1 RNase cleavages were represented by arrows surmounted by circles, triangles and squares, respectively. Nucleotides modified by DMS or CMCT are circled. Colours of circles and arrows give an indication of the yields of modifications and cleavages, red, yellow and green for strong, medium and low modification or cleavage, respectively. The V1 RNase cleavages and chemical modifications that cannot be explained by the two stem-loop structure models are encircled by blue lanes.

### **Figure 3: Schematic superposition of data obtained by enzymatic digestion and chemical modification of the mouse A region onto three possible 2D structure models involving interaction between repeats.**

In Model 1, the various stem-loop structures all involve two successive repeats, in Model 2 repeats in the 5' half of the A region interact with repeats in the 3' half of this region, in Model 3 two stem-loop structures containing 4 repeats are separated by a

small stem-loop corresponding to polyA and polyU sequences. The repeats are indicated by red lines and are numbered from 1 to 8. Representations of chemical and enzymatic data are as in Figure 2. The free energies of each stem-loop structures at 0°C and in 3M NaCl were calculated with the M-fold software.

**Figure 4: Comparative analysis of enzymatic cleavages in the entire human A region and RNA SLS1H containing the human repeats 1, 2, 3 and 4 by primer extension**

Legend as in Figure 1, except, that the transcripts correspond to the entire human A repeat region (human A RNA) and the 5' half of the human A region (SLS1H RNA), respectively.

**Figure 5: Schematic superposition of data obtained by enzymatic digestion and chemical modification of the human A region onto a 2D structure model corresponding to the mouse Model 3.**

Repeat 5 is located in a short central stem-loop structure flanked by two larger stem-loop structures that each contains 4 repeats. Representation of the enzymatic cleavages and chemical modifications are as in Figures 2 and 3. The free energies of the stem-loop structures were calculated with the M-fold software.

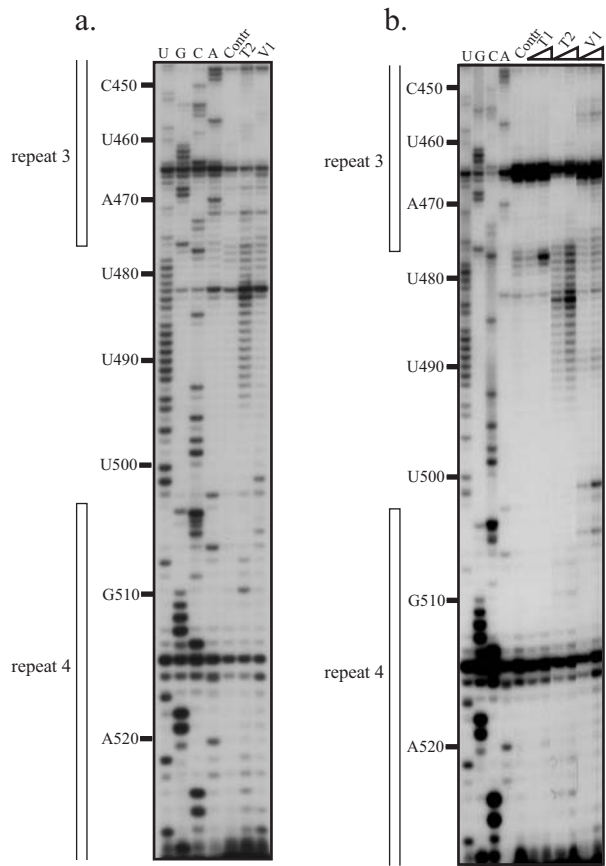
**Figure 6: Secondary structure analysis of the mouse A region by steady-state fluorescence provides additional support to Model 3.**

In Panel (a), the binding sites of oligonucleotides P1 to P6 used in the FRET experiments are shown for the three 2D structures of the A region corresponding to Models 1, 2 and 3. The identity of the chromophore present in each oligonucleotide (donor Cy3 or acceptor Cy5) is indicated in green and blue, respectively. Cy3 and Cy5-labeled oligonucleotides were purchased from Eurogentec. As illustrated in Panel (b), the emission fluorescence spectra from 530 to 745 nm of the donor oligonucleotide bound alone to the RNA (green curve) was collected, as well as the emission spectra obtained in the presence of the donor and acceptor oligonucleotides (violin curve). No energy transfer between Cy3 and Cy5-labeled oligonucleotides was detected in solution. The FRET efficiency for each pair of oligonucleotides was defined as the decrease in fluorescence of the donor at 564 nm in the presence of the acceptor. Two representative examples of FRET assays

(oligonucleotide pairs P2/P4 and P3/P6) are shown in Panel (b). The FRET efficiencies measured for the six pairs of oligonucleotides are provided in Panel (c) (mean values of three independent experiments). Standard deviations ( $\sigma$ ) are shown. The relative efficiencies of the FRET obtained for each oligonucleotide pairs are schematically represented in Panel (a) by lines joining the oligonucleotides. The thickness of the lines reflects the efficiency of the FRET effect.

**Figure 7: The PRC2 complex assembles on fragments of the A region containing at least 4 repeats.**

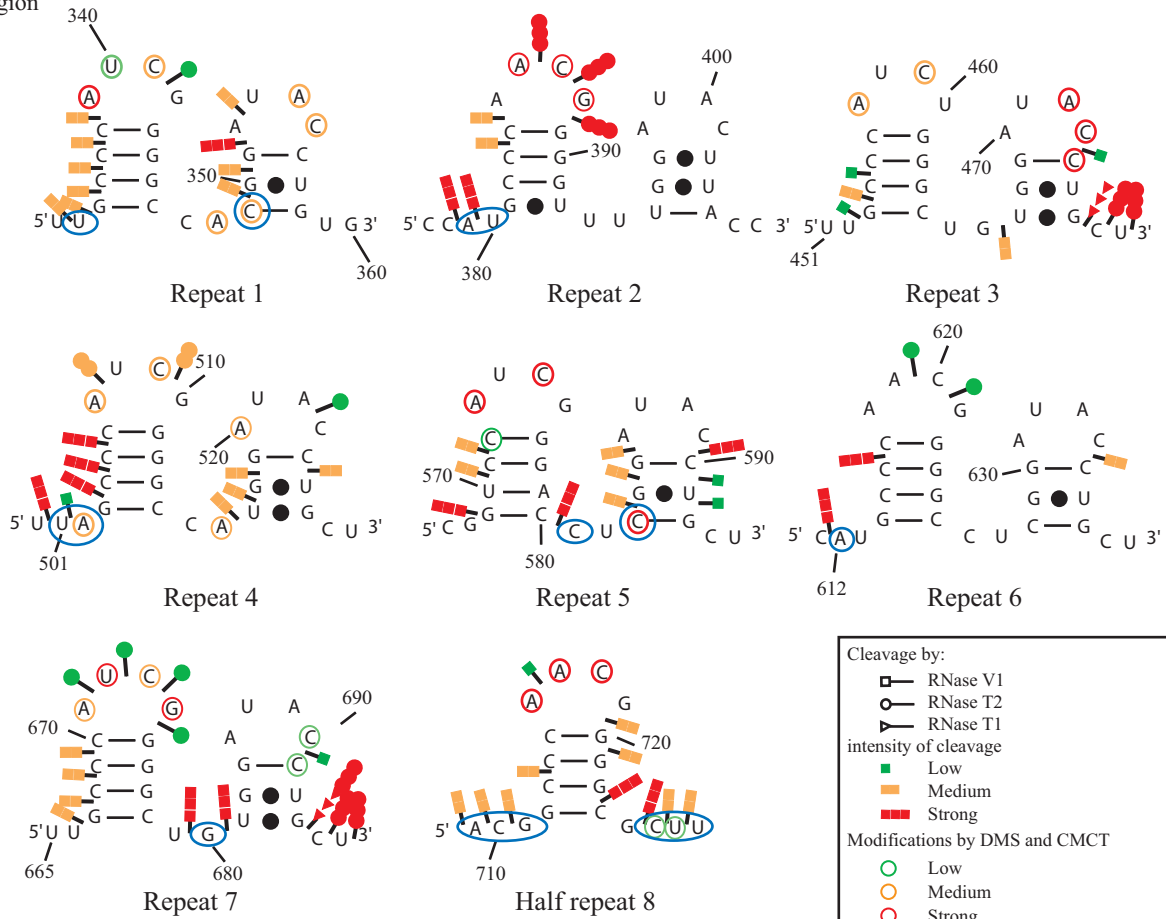
(a) Representation of the fusion RNAs used for formation of RNP complexes with ES cell nuclear extracts. Retention of RNA containing three MS2 coat protein binding sites (MS2) on amylose beads was mediated by the MS2/MBP fusion protein<sup>31</sup>. Analysis of the protein content of the RNP complexes formed on the A region/MS2 RNA was achieved by mass spectrometry (**Supplementary Fig. 8**). (b and c) Western blot assays using antibodies specific for the PTB, Suz12, RbAp46, RbAp48, Eed and Ezh2 protein were used to evaluate the relative amounts of these proteins in the purified complexes formed with the various RNAs shown in panel (a). Antibodies were purchased from Santa Cruz (Sc), Abcam (ab) and Calbiochem (cb) : anti-Ezh2 (anti-ENX-1 H-80, sc-25383), anti-RbAp46 (ab3535), anti-RbAp48 (ab488), anti-Eed (ab4469), anti-Suz12 (ab12073) and anti-PTB (cb NA63). (d and e) Test of the association of radiolabelled fragments of the A region with components of the PRC2 complex present in nuclear extracts of ES cells. The three RNAs tested are represented in Panel (d). RNAs bound to the G sepharose beads were fractionated by electrophoresis on 7% denaturing gels. Autoradiograms of the gels are shown in Panel (e). Input corresponds to 10% of the material incubated with the beads.



5' terminal 1137-nt long RNA A repeat region

**Figure 1**

Mouse A region



**Figure 2**

Cleavage by:

■ RNase V1

○ RNase T2

▲ RNase T1

intensity of cleavage

■ Low

■ Medium

■ Strong

Modifications by DMS and CMCT

○ Low

○ Medium

○ Strong

○ Cleavages and modifications in contradiction with structure model

Mouse A region

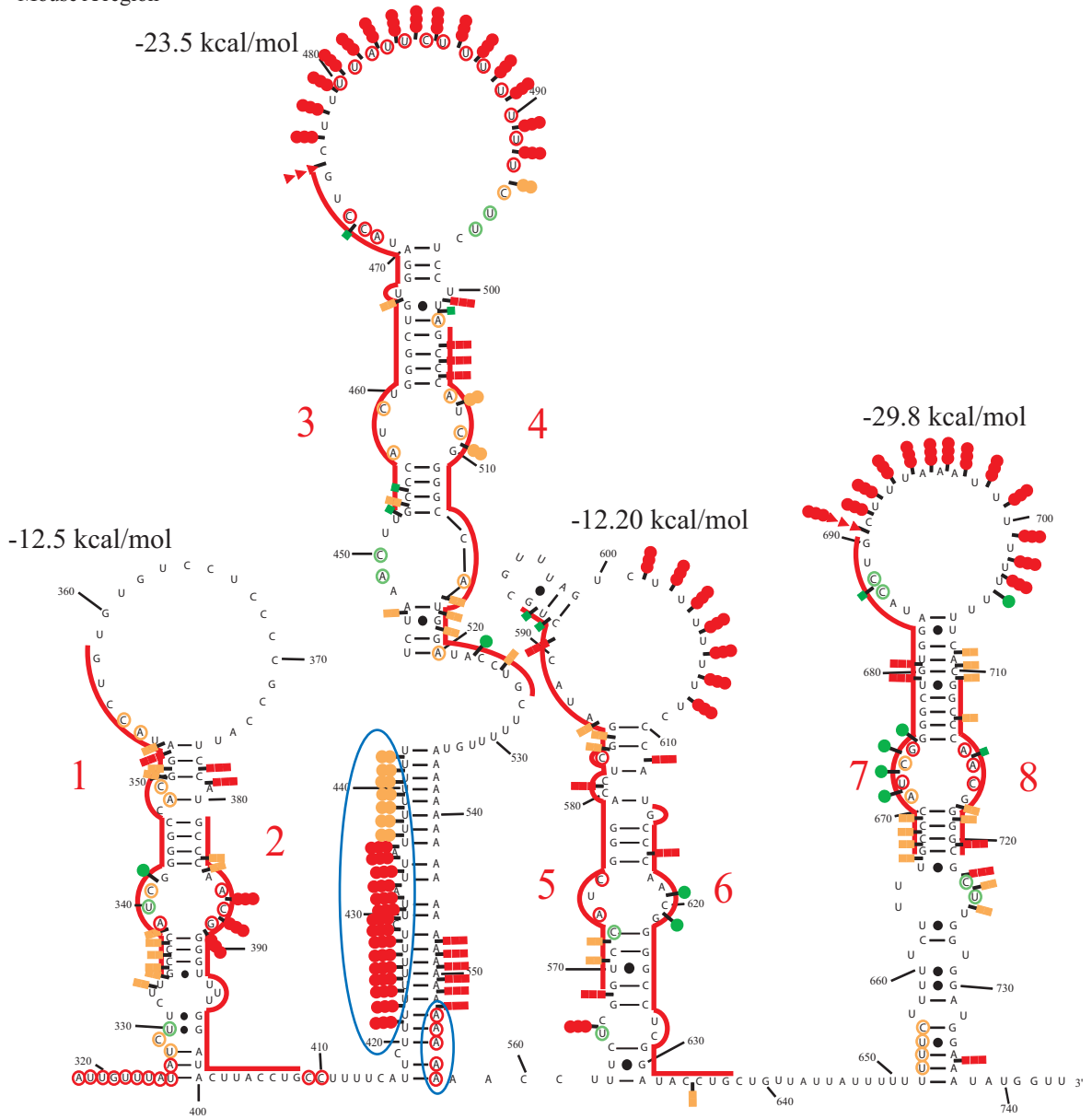


Figure 3a

Mouse A region

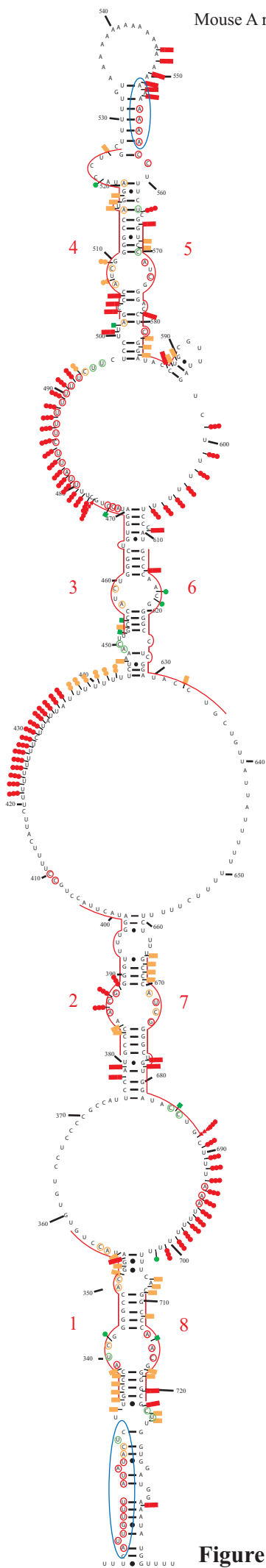


Figure 3b

Mouse A region

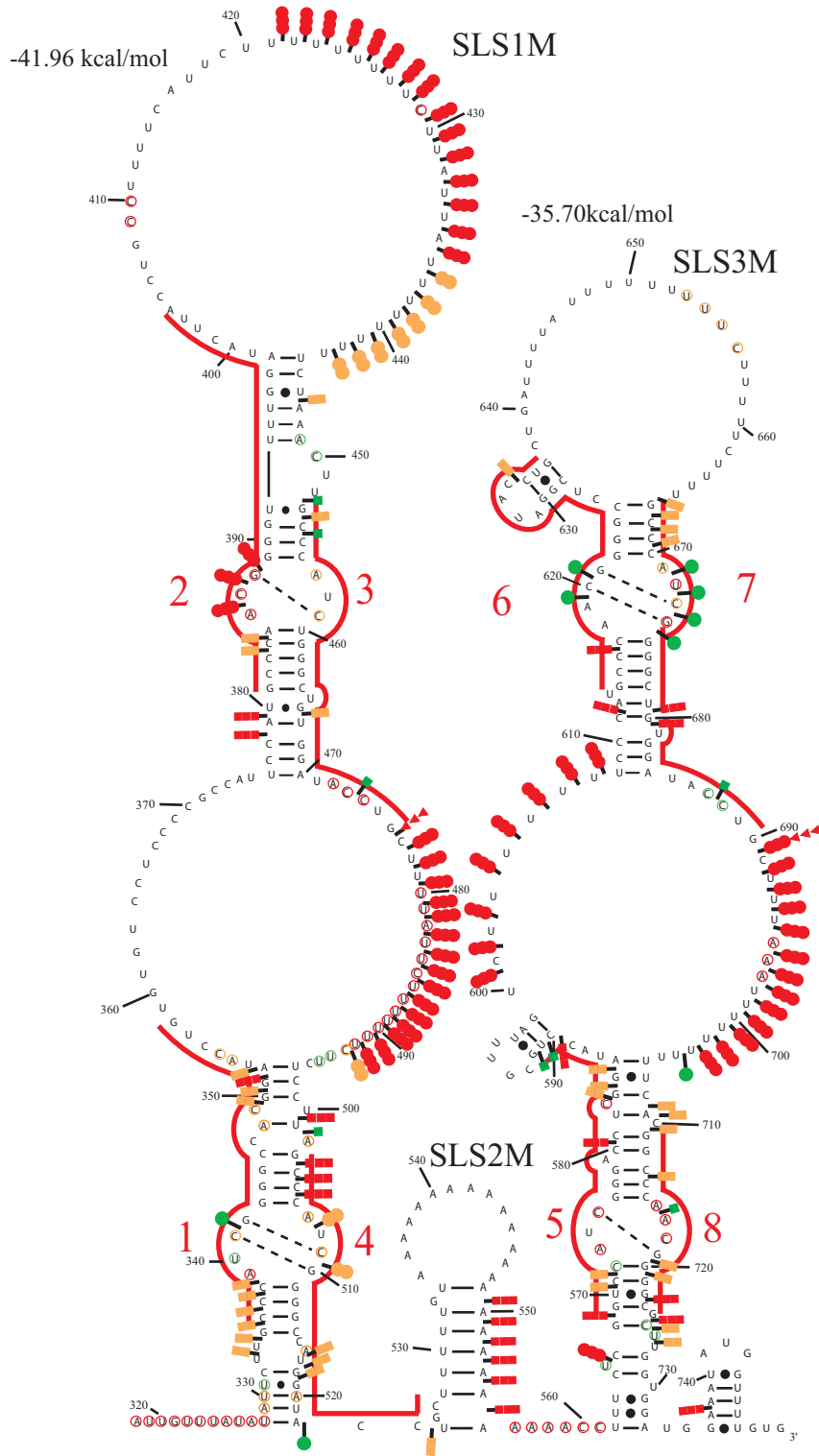
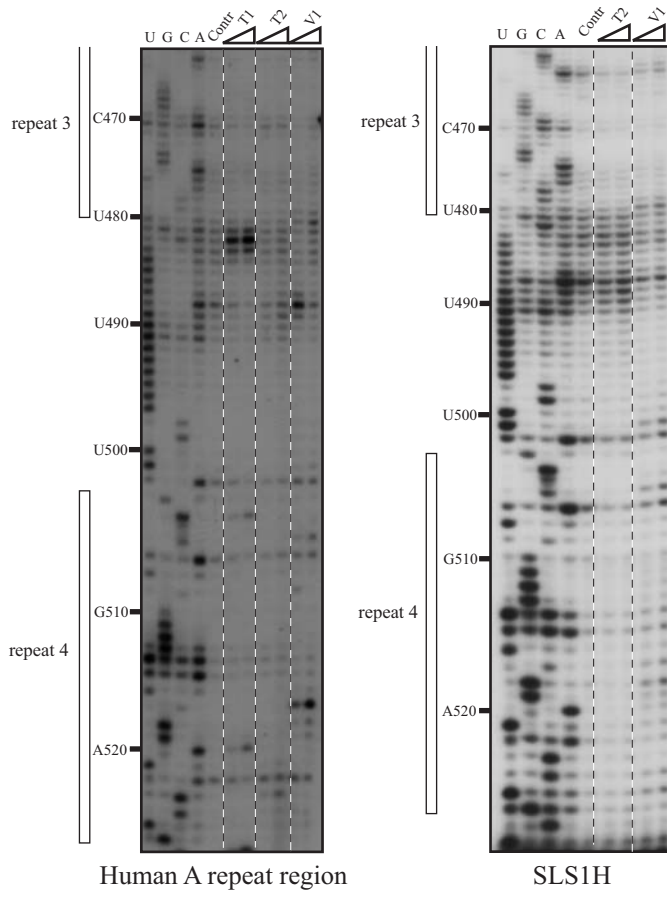
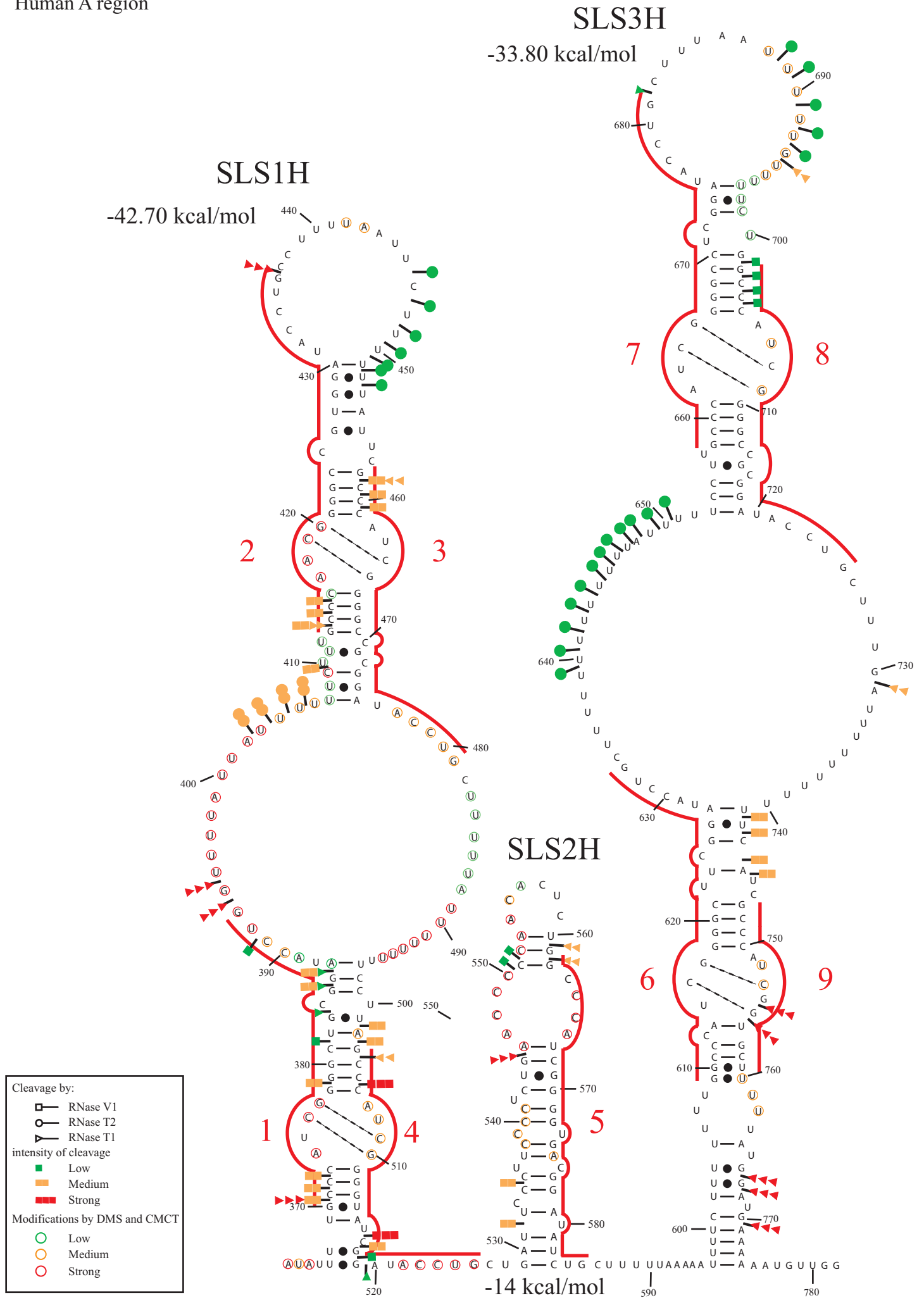


Figure 3c





**Figure 4**



**Figure 5**

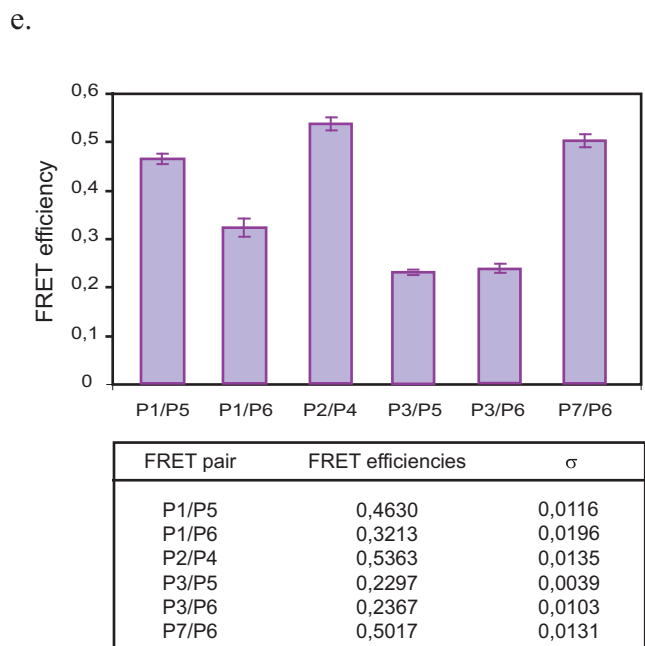
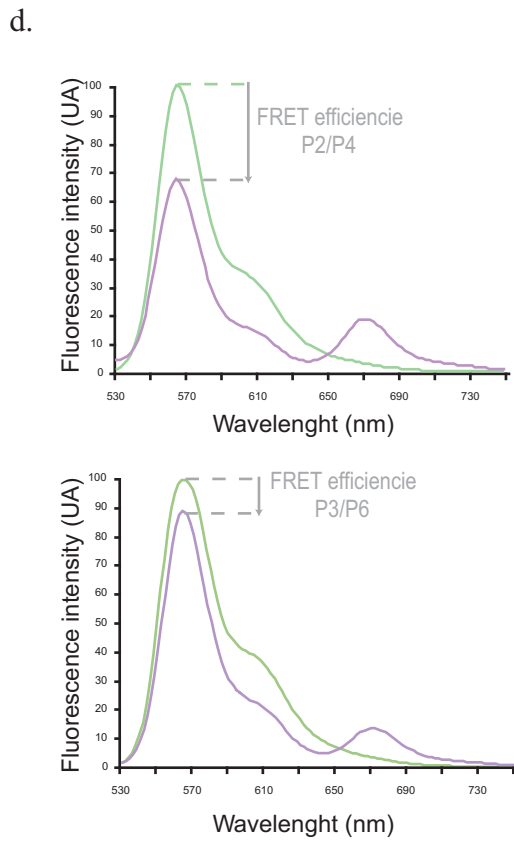
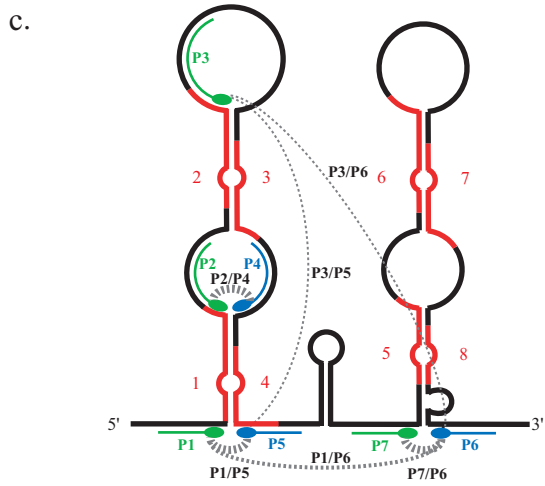
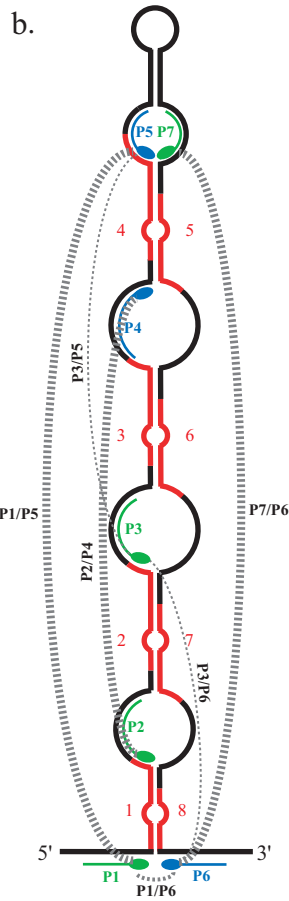
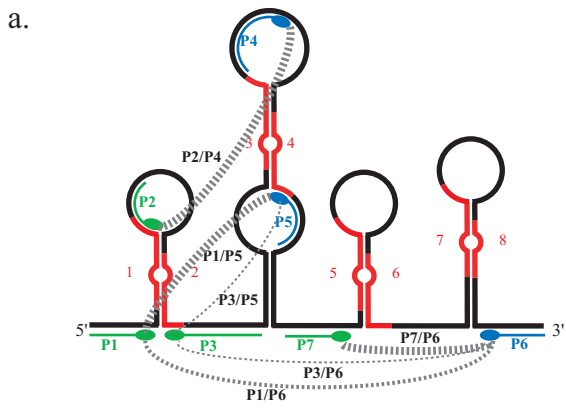
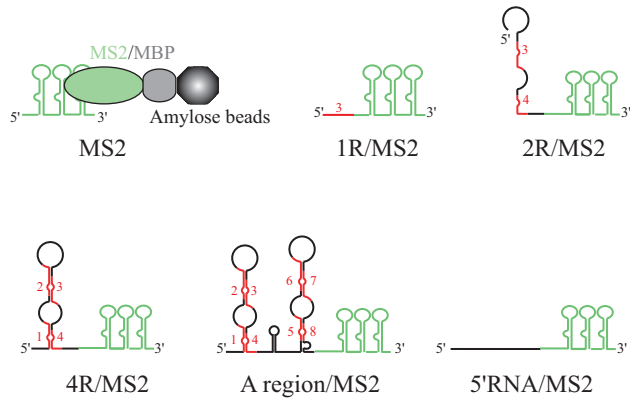
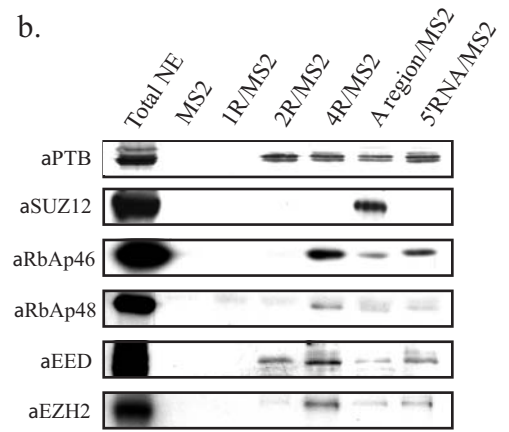


Figure 6

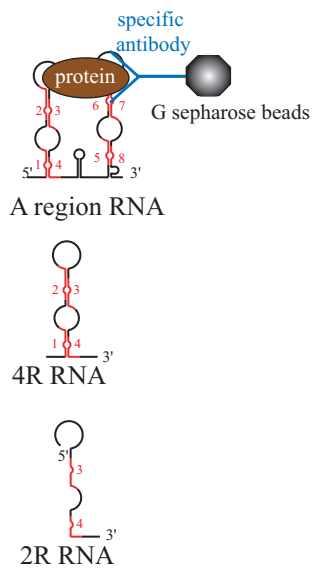
a.



b.



c.



d.

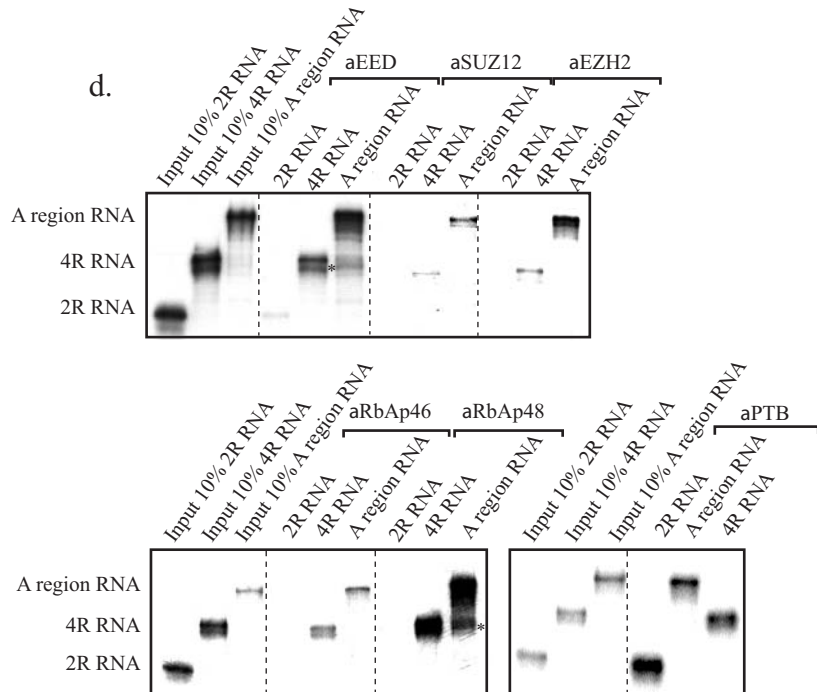


Figure 7

## **Conclusion générale de la partie "Identification des protéines impliquées dans des complexes protéine/protéine et ARN/protéine"**

L'identification des protéines partenaires d'interaction d'une autre protéine ou d'un ARN nécessite de maîtriser plusieurs paramètres :

- La purification des complexes : La comparaison des 2 méthodes de purification de complexe ARN/protéine (Tobramycine/MS2) a permis de mettre en avant l'importance du choix de la purification pour identifier les protéines en interaction avec l'ARN cible et non les protéines en interaction avec l'ARN support.

- La sensibilité de l'instrumentation : En fonction de la quantité de matériel disponible et de la question posée, il est important de choisir un instrument avec une sensibilité adaptée.

- Le traitement des données de spectrométrie de masse : l'évaluation statistique des données obtenues est nécessaire afin d'augmenter la confiance dans les identifications obtenues. Pour cela le choix de la banque de données protéiques et le choix des critères de validations des données est crucial.

L'analyse protéomique apporte des candidats potentiellement impliqués dans une interaction avec la protéine ou l'ARN cible. Ces candidats protéiques doivent être validés par des méthodes biologiques (Western Blot, Gel retard..) afin de confirmer l'interaction.



## **Références bibliographique**

### **A.**

Avner, P. and Heard, E. (2001). "X-chromosome inactivation: counting, choice and initiation." Nat Rev Genet **2**(1): 59-67.

### **B.**

Bachler, M., Schroeder, R. and von Ahsen, U. (1999). "StreptoTag: a novel method for the isolation of RNA-binding proteins." Rna **5**(11): 1509-16.

Bose, S. K., Sengupta, T. K., Bandyopadhyay, S. and Spicer, E. K. (2006). "Identification of Ebp1 as a component of cytoplasmic bcl-2 mRNP (messenger ribonucleoprotein particle) complexes." Biochem J **396**(1): 99-107.

Bradshaw, R. A., Burlingame, A. L., Carr, S. and Aebersold, R. (2006). "Reporting protein identification data: the next generation of guidelines." Mol Cell Proteomics **5**(5): 787-8.

### **C.**

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K. and Nesvizhskii, A. (2004). "The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data." Mol Cell Proteomics **3**(6): 531-3.

Chen, H. H., Chang, J. G., Lu, R. M., Peng, T. Y. and Tarn, W. Y. (2008). "The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene." Mol Cell Biol **28**(22): 6929-38.

### **D.**

Dantzer, F., Mark, M., Quenet, D., Scherthan, H., Huber, A., Liebe, B., Monaco, L., Chicheportiche, A., Sassone-Corsi, P., de Murcia, G. and Menissier-de Murcia, J. (2006). "Poly(ADP-ribose) polymerase-2 contributes to the fidelity of male meiosis I and spermiogenesis." Proc Natl Acad Sci U S A **103**(40): 14854-9.

### **F.**

Fandino, A. S., Rais, I., Vollmer, M., Elgass, H., Schagger, H. and Karas, M. (2005). "LC-nanospray-MS/MS analysis of hydrophobic proteins from membrane protein complexes isolated by blue-native electrophoresis." J Mass Spectrom **40**(9): 1223-31.

Fischer, F., Wolters, D., Rogner, M. and Poetsch, A. (2006). "Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection." Mol Cell Proteomics **5**(3): 444-53.

## **G.**

Gingras, A. C., Gstaiger, M., Raught, B. and Aebersold, R. (2007). "Analysis of protein complexes using mass spectrometry." Nat Rev Mol Cell Biol **8**(8): 645-54.

Goina, E., Skoko, N. and Pagani, F. (2008). "Binding of DAZAP1 and hnRNP A1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant." Mol Cell Biol **28**(11): 3850-60.

Gordiyenko, Y. and Robinson, C. V. (2008). "The emerging role of MS in structure elucidation of protein-nucleic acid complexes." Biochem Soc Trans **36**(Pt 4): 723-31.

## **H.**

Hakes, L., Pinney, J. W., Robertson, D. L. and Lovell, S. C. (2008). "Protein-protein interaction networks and biology--what's the connection?" Nat Biotechnol **26**(1): 69-72.

Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M. and Luhrmann, R. (2002). "Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method." Proc Natl Acad Sci U S A **99**(26): 16719-24.

## **K.**

Kocher, T. and Superti-Furga, G. (2007). "Mass spectrometry-based functional proteomics: from molecular machines to protein networks." Nat Methods **4**(10): 807-15.

## **L.**

Lamond, A. I. and Mann, M. (1997). "Cell biology and the genome projects a concerted strategy for characterizing multiprotein complexes by using mass spectrometry." Trends Cell Biol **7**(4): 139-42.

Lasserre, J. P., Beyne, E., Pyndiah, S., Lapailleur, D., Claverol, S. and Bonneau, M. (2006). "A complexomic study of Escherichia coli using two-dimensional blue native/SDS polyacrylamide gel electrophoresis." Electrophoresis **27**(16): 3306-21.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. and Yates, J. R., 3rd (1999). "Direct analysis of protein complexes using mass spectrometry." Nat Biotechnol **17**(7): 676-82.

Lu, B., McClatchy, D. B., Kim, J. Y. and Yates, J. R., 3rd (2008). "Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry." Proteomics **8**(19): 3947-55.

## **M.**

Marchand, V. (2005) Etude des mécanismes de régulation de l'épissage des deux rétrovirus, le virus du Sarcome de Rous (RSV) et le virus responsable de l'immunodéficience humaine (VIH-1). Thèse de l'Université Henri Poincaré

Marchand, V., Méreau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stévenin, J. and Branlant, C. (2002). "A Janus Splicing Regulatory Element Modulates HIV-1 tat and rev mRNA Production by Coordination of hnRNP A1 Cooperative Binding." J. Mol. Biol. **323**: 629-652.



## **N.**

Nesvizhskii, A. I. and Aebersold, R. (2005). "Interpretation of shotgun proteomic data: the protein inference problem." Mol Cell Proteomics **4**(10): 1419-40.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998). "Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex." Nat Genet **20**(1): 46-50.

Nusinow, D. A., Hernandez-Munoz, I., Fazio, T. G., Shah, G. M., Kraus, W. L. and Panning, B. (2007). "Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome." J Biol Chem **282**(17): 12851-9.

## **O.**

Oei, S. L., Keil, C. and Ziegler, M. (2005). "Poly(ADP-ribosylation) and genomic stability." Biochem Cell Biol **83**(3): 263-9.

## **Q.**

Quenet, D., Gasser, V., Fouillen, L., Cammas, F., Sanglier-Cianferani, S., Losson, R. and Dantzer, F. (2008). "The histone subcode: poly(ADP-ribose) polymerase-1 (Parp-1) and Parp-2 control cell differentiation by regulating the transcriptional intermediary factor TIF1beta and the heterochromatin protein HP1alpha." Faseb J **22**(11): 3853-65.

## **R.**

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999). "A generic protein purification method for protein complex characterization and proteome exploration." Nat Biotechnol **17**(10): 1030-2.

## **S.**

Saliou, J.-M. (2008) Etude sur le complexe TAR/Tat/cycline T1 et étude des régulations de l'épissage de l'ARN pré-messager du virus HIV-1: effet global des protéines virales et analyse fine du rôle des protéines SR ASF/SF2 et 9G8 au site accepteur A3. Thèse de l'Université Henri Poincaré

Saliou, J. M., Bourgeois, C. F., Ayadi-Ben Mena, L., Ropers, D., Jacquenet, S., Marchand, V., Stevenin, J. and Branlant, C. (2009). "Role of RNA structure and protein factors in the control of HIV-1 splicing." Front Biosci **14**: 2714-29.

Sanglier-Cianferani, S. (2002) La spectrométrie de masse: un nouvel outil pour l'étude des interactions fiables en biologie Thèse de l'Université Louis Pasteur

Sanglier, S., Atmanene, C., Chevreux, G. and Dorselaer, A. V. (2008). "Nondenaturing mass spectrometry to study noncovalent protein/protein and protein/ligand complexes: technical aspects and application to the determination of binding stoichiometries." Methods Mol Biol **484**: 217-43.

Schagger, H. and von Jagow, G. (1991). "Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form." Anal Biochem **199**(2): 223-31.

Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. and Wutz, A. (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing." Embo J. **25**: 3111-3122.

Schreiber, V., Ame, J. C., Dolle, P., Schultz, I., Rinaldi, B., Fraulob, V., Menissier-de Murcia, J. and de Murcia, G. (2002). "Poly(ADP-ribose) polymerase-2 (PARP-2) is required for efficient base excision DNA repair in association with PARP-1 and XRCC1." J Biol Chem **277**(25): 23028-36.

Sleasman, J. W. and Goodenow, M. M. (2003). "13. HIV-1 infection." J Allergy Clin Immunol **111**(2 Suppl): S582-92.

#### **T.**

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **60**(5): 523-33.

#### **V.**

Venables, J. P., Bourgeois, C. F., Dalglish, C., Kister, L., Stevenin, J. and Elliott, D. J. (2005). "Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon." Hum Mol Genet **14**(16): 2289-303.

#### **W.**

Wutz, A., Rasmussen, T. P. and Jaenisch, R. (2002). "Chromosomal silencing and localization are mediated by different domains of Xist RNA." Nat Genet **30**(2): 167-74.

#### **Z.**

Zhou, Z., Licklider, L. J., Gygi, S. P. and Reed, R. (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**(6903): 182-5.

## **PARTIE II – Caractérisation fine de protéines purifiées phosphorylées**

**Chapitre I. Introduction sur l'analyse des phosphorylations par spectrométrie de masse et stratégie développée.**

**Chapitre II. Etude des phosphorylations du domaine kinase de la protéine DAPK1**

**Chapitre III. Etude des phosphorylations du complexe CAK**



## **Chapitre I - Introduction sur l'analyse des phosphorylations par spectrométrie de masse et stratégie développée.**

### **1. Les difficultés liées à l'analyse des phosphorylations par spectrométrie de masse**

#### **1.a. Intérêts biologiques des phosphorylations**

Les modifications post-traductionnelles sont des événements chimiques qui modifient les propriétés des protéines après leur traduction, soit par clivage protéolytique, soit par addition de groupements chimiques sur un ou plusieurs acides aminés. La protéine ainsi modifiée adopte une nouvelle structure et acquiert des propriétés physicochimiques très différentes de la molécule directement codée par le gène. Plus de 200 modifications post-traductionnelles différentes ont été décrites. Elles ont toutes des rôles importants dans la cellule et son fonctionnement [Walsh et al., 2005].

Les deux plus importantes modifications sont la glycosylation et la phosphorylation. Les protéines glycosylées représentent la majorité des marqueurs de surface cellulaires ainsi que des protéines sécrétées. Elles sont critiques dans la détermination de la stabilité de la protéine, sa conformation ou son affinité avec d'autres protéines. De plus, elles sont impliquées dans un très grand nombre de processus biologiques tels que la croissance cellulaire, la réponse immunitaire ou la communication intercellulaire [Budnik et al., 2006].

Environ 30% des protéines présentent une ou plusieurs phosphorylations [Schweppe et al., 2003]. La phosphorylation régule la translocation des protéines entre divers compartiments cellulaires, les interactions protéine/protéine et l'activation ou l'inactivation de certaines fonctions cellulaires. La phosphorylation est souvent une clé régulatrice cruciale pour la signalisation cellulaire. Une dérégulation de phosphorylation peut être impliquée dans diverses maladies, notamment parmi certains cancers [Schmelzle & White, 2006].

Les sites de phosphorylations sont nombreux et varient selon le rôle biologique de la protéine à un moment donné [Mann et al., 2002]. De plus, la part de protéines phosphorylées est peu importante par rapport à la part de protéines disponibles à un moment donné. Tous ces éléments impliquent que l'identification et la caractérisation des protéines phosphorylées ne sont pas évidentes.

## **1.b. Généralités sur les phosphorylations**

La phosphorylation consiste en l'addition d'un groupement phosphate sur un acide aminé. C'est une modification réversible et labile. Il existe plusieurs types de phosphorylations [Sickmann & Meyer, 2001].

- Les O-phosphorylations : elles consistent en l'addition d'un groupement phosphate sur la fonction alcool des sérines, des thréonines et des tyrosines, par formation d'une liaison O-P (Figure 1).

- Les N-phosphorylations : elles consistent en l'addition d'un groupement phosphate sur la fonction amine des arginines, des histidines et des lysines.

- Les Acylphosphorylations (Ajout d'une phosphorylation sur l'acide aspartique et l'acide glutamique), et les S-phosphorylations (Ajout d'une phosphorylation sur les cystéines) sont plus rares.

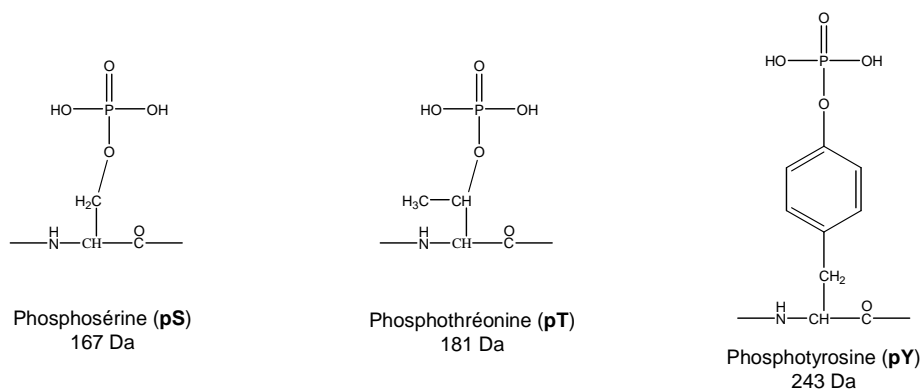


Figure 1 : Les acides aminés O-phosphorylés

Les phosphopeptides ont une efficacité d'ionisation plus faible que celle d'un peptide ne portant pas de modification, il est donc plus difficile de détecter l'ion correspondant au phosphopeptide [Janek et al., 2001; Gropengiesser et al., 2009]. Cette particularité induit donc que l'identification et la caractérisation des phosphorylations par spectrométrie de masse restent toujours un défi et demandent des stratégies adaptées. Durant ma thèse, mon travail s'est limité à l'étude de O-Phosphorylations, donc seules les méthodologies pour caractériser ces phosphorylations seront décrites dans ce chapitre.

## **1.c. Enrichissement des phosphopeptides**

Les phosphorylations étant peu abondantes, des méthodes d'enrichissement ont été développées pour les phosphoprotéines mais surtout pour les phosphopeptides.

L'enrichissement des phosphoprotéines et phosphopeptides par immunoprécipitation à l'aide d'anticorps est principalement utilisé pour les phosphotyrosines [Steen et al., 2002]. Pour les phosphosérines et les phosphothréonines, ce système est peu utilisé car les anticorps pS et pT sont peu spécifiques [Mann et al., 2002].

Les principales méthodes d'enrichissement utilisés sont basées sur l'affinité des phosphopeptides avec certains métaux :

- la chromatographie IMAC (Immobilized Metal Ion Affinity Chromatography) avec  $\text{Fe}^{3+}$ ,  $\text{Al}^{3+}$ ,  $\text{Ga}^{3+}$  ou  $\text{Co}^{2+}$  [Nuhse et al., 2003; Kange et al., 2005].

- la chromatographie basée sur l'affinité avec le  $\text{TiO}_2$ . Cette méthode d'enrichissement peut être couplée directement avec le système nanoLC-MS/MS [Pinkse et al., 2004; Klemm et al., 2006].

Les modifications chimiques des phosphorylations sont des méthodes utilisées pour enrichir en phosphopeptides. Il en existe deux grands types :

- la  $\beta$ -élimination suivie d'une addition de Michael d'un groupement nucléophile portant une molécule tag, qui permettra de réaliser une chromatographie d'affinité contre ce tag et les peptides ainsi modifiés sont analysés par MS. Cette méthode n'est pas applicable pour les phosphotyrosines, car la  $\beta$ -élimination des phosphotyrosines n'est pas possible [Meyer et al., 1986; Oda et al., 2001; Jalili & Ball, 2008].

- la méthode PhosphoAmidate Chemistry (PAC) : les phosphopeptides sont modifiés et liés à des groupements iodoacetyles attachés à un support polymérique. Après enrichissement, les phosphopeptides sont élués sous leur forme phosphorylée originale et analysés par MS. Cette méthode, applicable aux phosphotyrosines, nécessite beaucoup d'étapes ce qui peut engendrer une perte d'échantillon [Zhou et al., 2001; Tao et al., 2005].

Ces méthodes permettent de simplifier l'échantillon et d'enrichir en phosphopeptides, elles peuvent être utilisées séparément ou bien en combinaison: par exemple la méthode SIMAC qui associe les méthodes IMAC et  $\text{TiO}_2$  [Thingholm et al., 2008] ou l'association d'une méthode de modification chimique ( $\beta$ -élimination suivie d'une addition de Michael) avec une colonne IMAC [Thompson et al., 2003].

#### **1.d. Identification des phosphopeptides.**

L'empreinte peptidique massique est utilisée pour identifier les peptides phosphorylés. La comparaison des masses théoriques et des masses expérimentales des peptides permet de mettre en évidence les peptides ayant un adduit de masse de + 80 Da. Ceci peut permettre d'identifier les peptides phosphorylés [D'Ambrosio et al., 2007]. Cependant, un peptide portant une sulfatation (adduit de +79.95 Da) peut être confondu avec un peptide phosphorylé du fait du même adduit de masse (+79.96 Da) [Yu et al., 2007]. Il est donc nécessaire de pouvoir différencier les peptides phosphorylés de ceux qui ne le sont pas, soit par détection d'ion métastable correspondant au peptide ayant perdu l'acide phosphorique, soit par comparaison entre l'échantillon non déphosphorylé et l'échantillon déphosphorylé.

- La labilité des groupements phosphates des phosphopeptides permet de générer des ions métastables spécifiques de la phosphorylation lors d'une analyse MALDI-MS [Annan & Carr, 1996; Schnölzer & Lehmann, 1997]. Ces ions diagnostiques permettent d'identifier les peptides phosphorylés en MALDI avant de le fragmenter en MS/MS pour identifier leurs sites de phosphorylations [Hem et al., 2007].

- L'empreinte peptidique massique entre l'échantillon non déphosphorylé et l'échantillon déphosphorylé permet de détecter les peptides phosphorylés [Liao et al., 1994]. Sur le spectre MALDI-MS après la déphosphorylation, seuls les peptides portant une phosphorylation ne sont plus présents, et apparaissent à un m/z initial - 80 Da (Figure 2). Les peptides portant une sulfatation ne sont pas affectés par la déphosphorylation et n'ont donc pas disparu du spectre MALDI-MS. La déphosphorylation des peptides est réalisée à l'aide de la phosphatase alcaline, l'incubation est réalisée en solution ou bien directement sur la cible MALDI [Larsen et al., 2001].

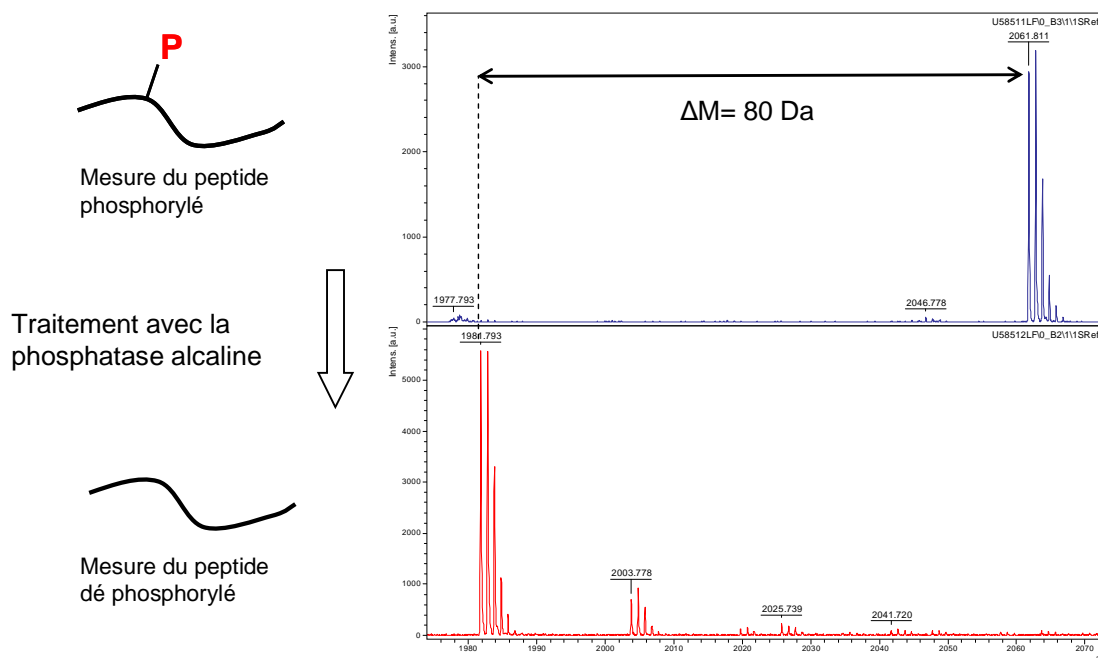


Figure 2 : Détails de spectres de masse MALDI-TOF-MS du peptide 48-63 FQpSEEQQTEDELQDK de la  $\beta$ -caséine avant et après traitement avec la phosphatase alcaline.

Les peptides portant une phosphorylation sont difficilement détectés par une analyse MALDI-MS. En effet, la phosphorylation diminue l'ionisation MALDI des peptides phosphorylés par rapport aux peptides non phosphorylés [Gropengiesser et al., 2009]. La déphosphorylation permet de mettre en évidence des peptides déphosphorylés, non détectés sous leur forme phosphorylée.

Pour augmenter la détection des phosphopeptides, l'utilisation d'une matrice permettant de conserver plus facilement les groupements labiles peut être utilisée telle que l'acide 2,5 dihydrobenzoïque (DHB) [Strupat et al., 1991]. Des études ont montré que l'ajout d'acide à cette matrice permet d'augmenter la détection des phosphopeptides : ajout de 1% d'acide phosphorique [Kjellstrom & Jensen, 2004] ou de 1% d'acide méthanediphosphonique (MDPNA) [Kuyama et al.,



2008]. L'utilisation d'une matrice liquide ionique formée à partir de DHB, de pyridine et de 1% d'acide phosphorique permet de mieux détecter les peptides phosphorylés notamment par le fait que le dépôt MALDI obtenu est plus homogène que celui obtenu avec la matrice sous forme cristalline [Tholey & Heinzle, 2006].

Une modification chimique des phosphopeptides permet également d'augmenter leur efficacité d'ionisation en MALDI-MS. Ces modifications sont basées sur les réactions de  $\beta$ -élimination/addition de Michael et sont donc applicable seulement aux phosphosérines et phosphothréonines. La modification consiste à remplacer le groupement phosphate chargé négativement par un groupement stable chargé positivement [Klemm et al., 2004; Arrigoni et al., 2006; Tsumoto et al., 2008].

### **1.e. Localisation des sites de phosphorylations**

Pour caractériser complètement un peptide phosphorylé, il est nécessaire de pouvoir situer la phosphorylation sur un acide aminé. La localisation de la phosphorylation peut être réalisée par la fragmentation du peptide phosphorylé lors d'une analyse MS/MS. Plusieurs approches sont possibles, soit par fragmentation CID, soit par fragmentation ECD/ETD (Figure 3).

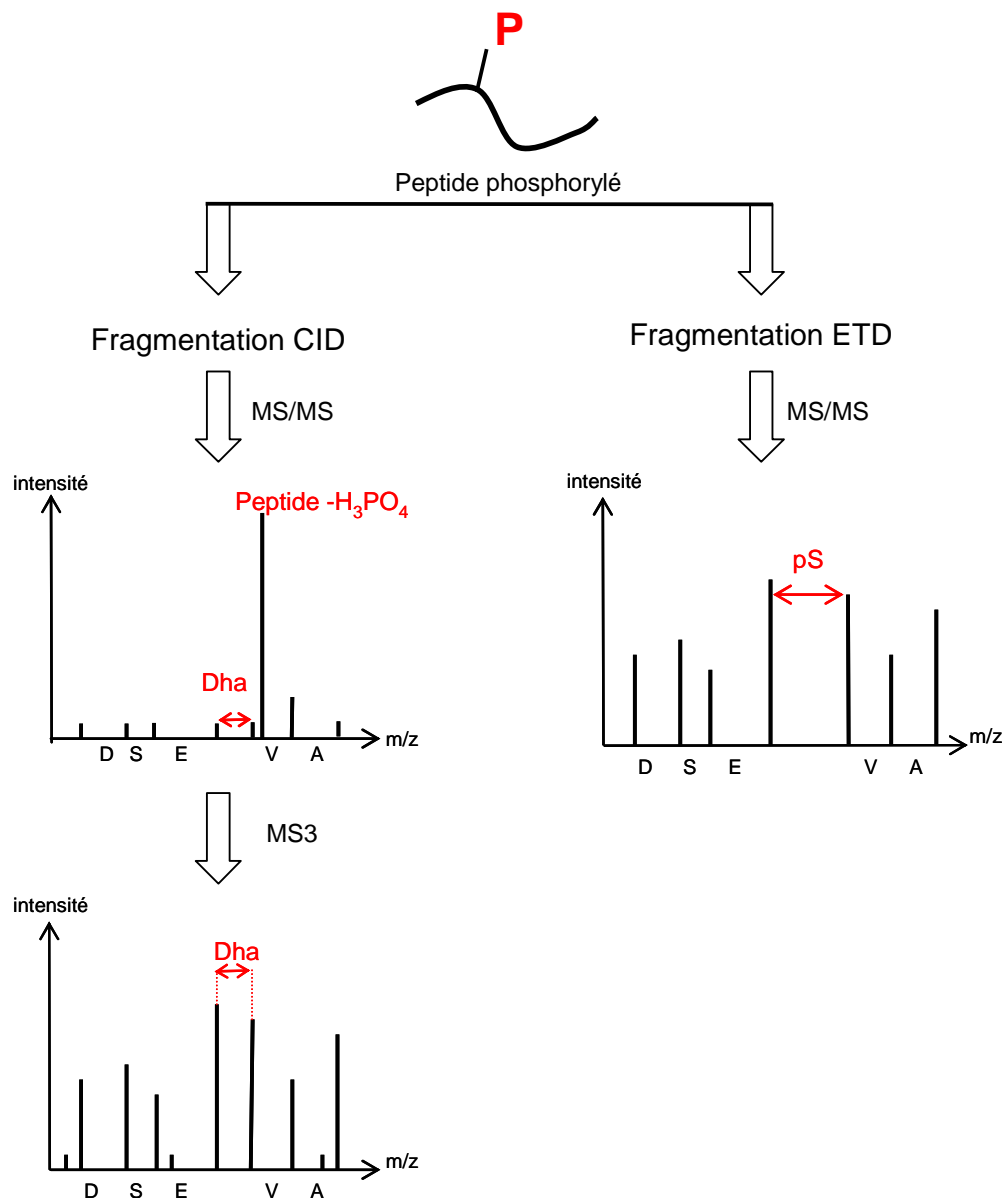


Figure 3 : Illustration des différentes fragmentations utilisées pour la localisation d'une phosphorylation [dehydroalanine (Dha) et phosphosérine (pS)].

### 1.e.1) Les approches par fragmentation CID

- Fragmentation MS/MS :

Lors d'une fragmentation CID, les premières liaisons à fragmenter sont les liaisons les plus faibles. Or, la liaison O-P est une liaison covalente plus fragile que les liaisons peptidiques. Donc, la fragmentation de la liaison O-P génère donc un ion souvent intense, qui dépend de la séquence du peptide et de l'acide aminé portant la phosphorylation [DeGnore & Qin, 1998]. La perte de la phosphorylation est une réaction de  $\beta$ -élimination de  $H_3PO_4$ . A l'issue de cette réaction, la phosphosérine et la phosphothréonine génèrent l'acide dehydroalanine (Dha) et l'acide dehydro

aminobutyrique (Dhb) respectivement (**Erreur ! Source du renvoi introuvable.** 4). La réaction de  $\beta$ -élimination de  $H_3PO_4$  est peu probable avec la phosphotyrosine du fait de la stabilisation du groupement phosphate et de l'encombrement stérique du groupement aromatique de la tyrosine.

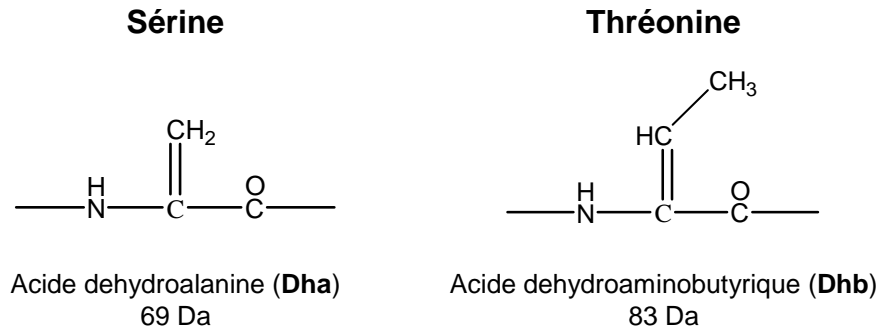


Figure 4 : Les acides aminés spécifiques issus la  $\beta$ -élimination de la phosphosérine et la phosphothréonine.

La fragmentation MS/MS donne parfois des informations suffisantes pour le séquençage et la localisation des phosphorylations, malgré la présence de l'ion résultant de la perte de l'acide phosphorique [Merrick et al., 2001; D'Ambrosio et al., 2006].

Dans le cas où la phosphorylation est située sur la tyrosine, la perte de phosphorylation due à la  $\beta$ -élimination n'étant pas présente, la fragmentation MS/MS est suffisante pour obtenir des informations de séquence. L'ion phosphotyrosine étant plus stable, l'ion immonium pY ( $m/z$  216.045) est relativement stable et peut être utilisé pour détecter les peptides possédant une phosphotyrosine (Figure 5) [Steen et al., 2006].

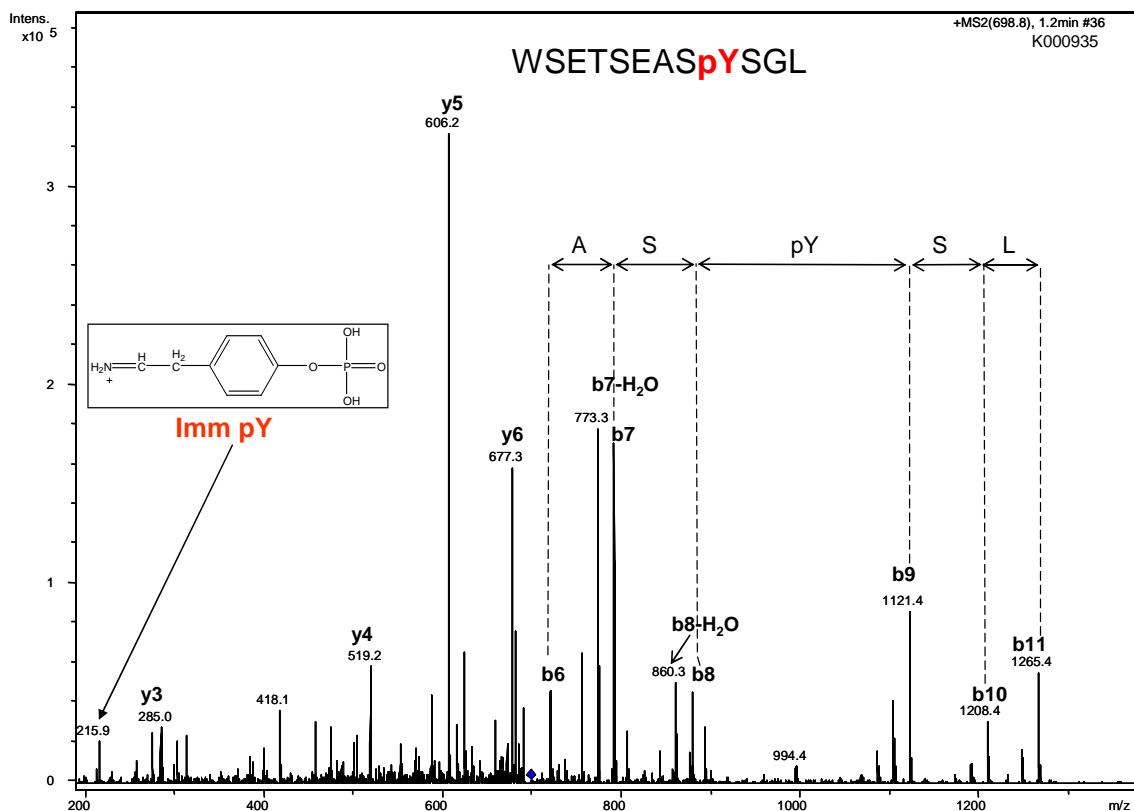


Figure 5 : Spectre de fragmentation de l'ion  $m/z$  698.8, doublement chargé, du peptide phosphorylé WSETSEASpYSGI. L'interprétation du spectre permet de mettre en évidence la phosphotyrosine par la présence des ions  $b_8$  et  $b_9$  et par la détection de l'ion immonium phosphotyrosine à  $m/z$  215.9.

#### - Fragmentation MS3 :

Lors d'une fragmentation CID MS/MS, la perte de l'acide phosphorique est souvent dominante provoquant ainsi une faible fragmentation du peptide. Le séquençage du peptide peut donc être difficile ainsi que la localisation de la phosphorylation (Figure 3). Le fragment généré par perte de neutre en MS/MS est utilisé par une nouvelle fragmentation (fragmentation MS3). Le spectre résultant MS3 (Figure 3) permet d'obtenir des informations de séquences et de localiser la phosphorylation grâce aux ions spécifiques créés par la perte de neutre (Dha et Dhb) [Beausoleil et al., 2004; Jin et al., 2004].

La détection de la perte de neutre lors de la fragmentation MS/MS est utilisée comme marqueur spécifique des peptides phosphorylés. Les spectres de fragmentation MS3 peuvent être obtenus de manière automatique dès qu'un ion très intense du spectre MS/MS est détecté comme correspondant à une perte de phosphorylation. Cela peut induire des faux positifs lors des identifications, notamment lors d'identification de phosphopeptides par soumission à des moteurs de recherches, l'état de charge de l'ion précurseur n'étant pas forcément indiqué dans le fichier soumis.

En effet, la fragmentation d'un peptide contenant une méthionine oxydée génère une perte d'acide méthanesulfénique. Une perte d'acide phosphorique issu d'un ion triplement chargé (98/3) est similaire en masse à une perte d'acide méthanesulfénique issu d'un ion doublement chargé (64/2), si

on considère une fenêtre d'isolation de 2 Da (correspondant à celle utilisée classiquement pour une trappe ionique). De plus, la perte de l'acide méthanesulfénique d'un peptide contenant une méthionine oxydée (Figure 6) génère également l'acide aminé acide dehydroaminobutyrique (Dhb) [Wolschin & Weckwerth, 2006]. Il est donc important de prendre en compte tous les paramètres et de vérifier chaque spectre de fragmentation de phosphopeptides.

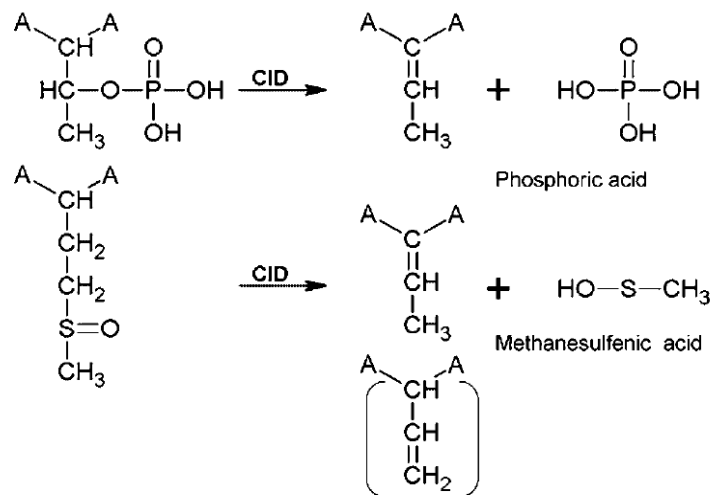


Figure 6 : Formation de l'acide dehydroaminobutyrique par la thréonine phosphorylée et de la méthionine oxydée lors d'une fragmentation CID [Wolschin & Weckwerth, 2006].

### 1.e.2) Les approches par fragmentation ECD/ETD

Le processus de fragmentation par capture ou transfert d'électron ne suivant pas le chemin énergétique préférentiel, les liaisons les plus faibles ne sont pas rompues préférentiellement. Lors de la fragmentation d'un peptide portant une phosphorylation, la perte de l'acide phosphorique n'est pas favorisée. L'ion correspondant à la perte de  $H_3PO_4$  n'est pas majoritaire sur un spectre de fragmentation ETD, contrairement à une fragmentation CID. Les ions générés sont issus de la fragmentation du squelette peptidique. Cela permet de séquencer le peptide tout en gardant l'acide aminé phosphorylé intact (Figure 3) [Mikesh et al., 2006; Chi et al., 2007]. Le séquençage et la localisation du phosphopeptide peuvent donc être aisément réalisés à partir d'un spectre MS/MS ETD.

### 1.f. Conclusion

Les phosphorylations sont des modifications post-traductionnelles importantes mais souvent peu abondantes. De plus, le caractère labile et hydrophile du groupement phosphate rend sa détection et sa caractérisation complexes. Plusieurs difficultés sont donc à prendre en compte pour caractériser des phosphorylations:

- l'abondance des protéines/peptides phosphorylés,
- l'efficacité d'ionisation des phosphopeptides,

- la fragmentation des phosphopeptides souvent dominée par la perte du groupement phosphate.

Pour l'identification et la caractérisation des phosphorylations, des solutions peuvent être apportées à ces problèmes tels que l'enrichissement ou des méthodes de fragmentations spécifiques (CID-MS3 ou ETD). Différentes stratégies peuvent être envisagées pour identifier et caractériser les phosphorylations selon les échantillons. Cependant, malgré la multiplicité des approches, aucune méthode n'étant universelle, la caractérisation des phosphorylations reste un défi.

Dans le cadre de mon travail de thèse, une stratégie a été développée pour l'identification et de la caractérisation de phosphorylations de protéines purifiées.

## 2. Stratégie développée pour l'identification de phosphorylations de protéines purifiées

Une méthodologie en deux étapes a été développée pour l'étude des protéines phosphorylées purifiées. Cette approche (Figure 7) permet tout d'abord la détermination de l'état de phosphorylation de la protéine entière (i.e. le nombre de phosphorylations), puis la localisation du site de phosphorylation sur le squelette peptidique.

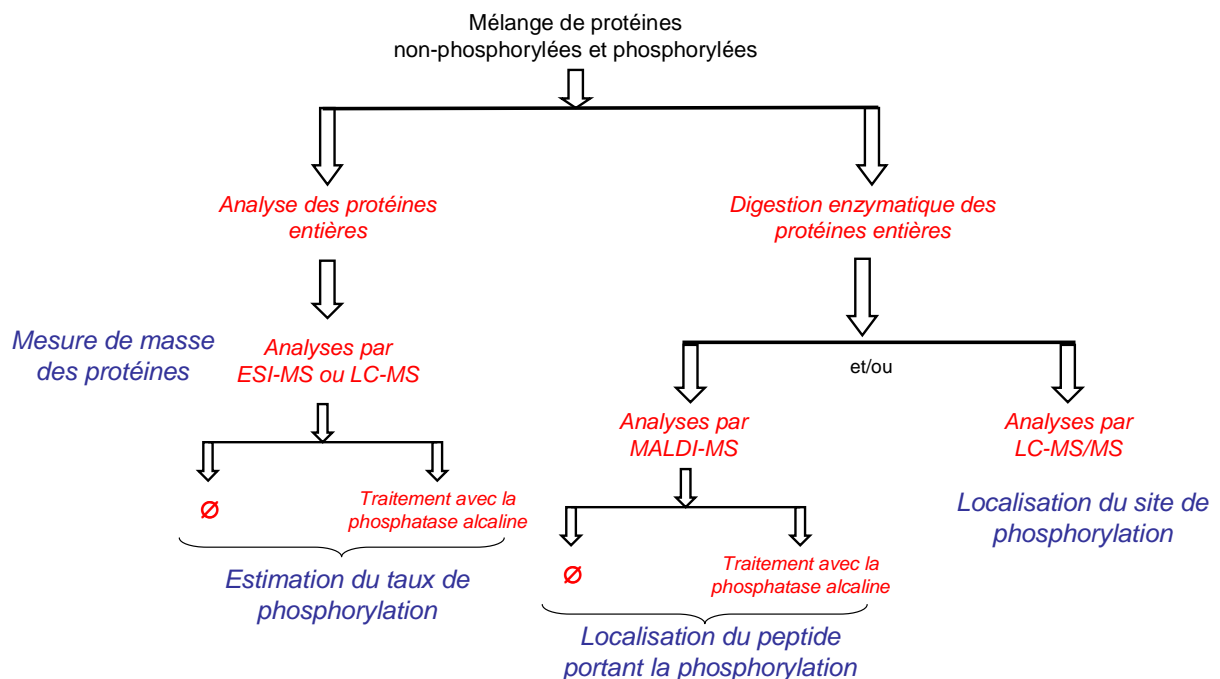


Figure 7 : Stratégie générale utilisée pour la détermination de l'état de phosphorylation de protéines purifiées.

## 2.a. Détermination de l'état de phosphorylation sur les protéines entières

L'état de phosphorylation des protéines entières est réalisé en 2 étapes (Figure 8) :

- i) Mesure de masse de la protéine ou du mélange de protéines,
- ii) Déphosphorylation du mélange de protéines à l'aide de la phosphatase alcaline.

La déphosphorylation de protéines entières est une méthodologie peu courante, qui permet de connaître l'état de phosphorylation de la protéine [Garnier et al., 2001].

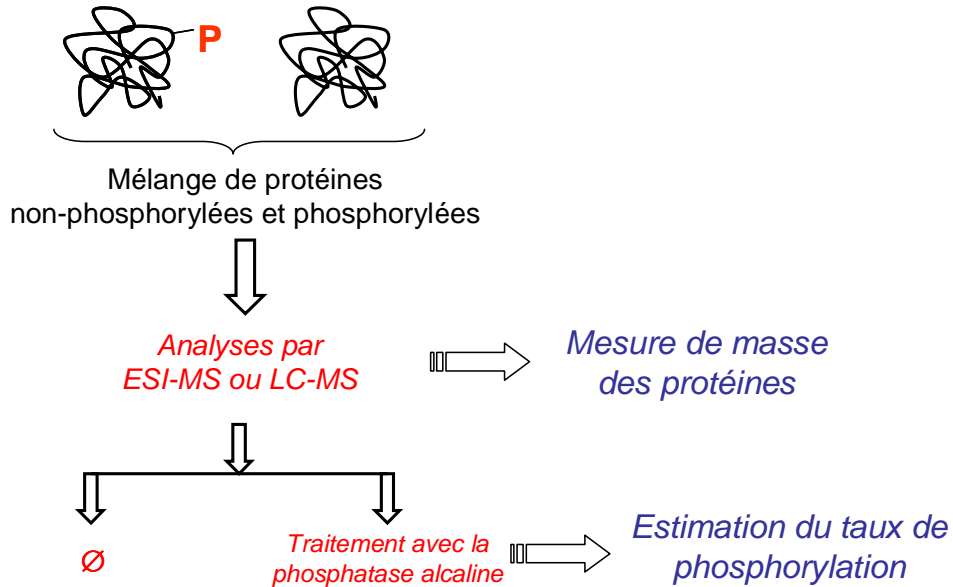


Figure 8 : Stratégie d'analyse des phosphoprotéines entières permettant de déterminer leur état de phosphorylation .

La mesure de masse de la protéine recombinante est déterminée par une analyse ESI-MS ou LC-MS (Figure 9). Différentes populations présentant des incréments de masse de + 80 Da permettent d'émettre l'hypothèse de la présence de formes phosphorylées.

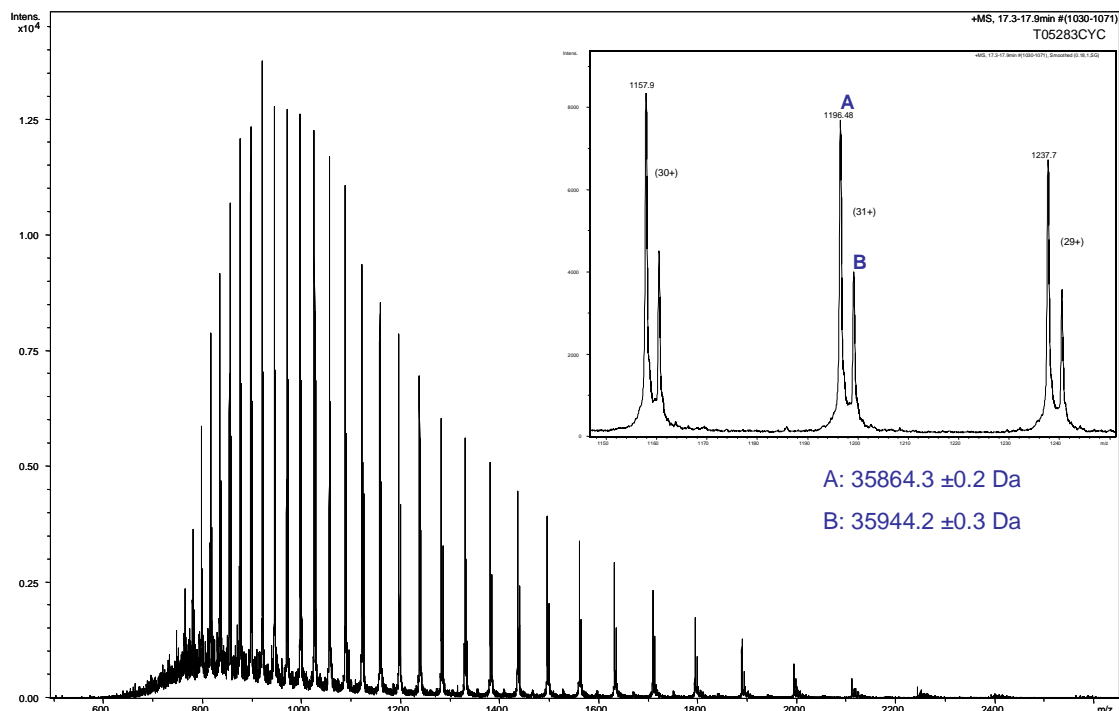


Figure 9 : Spectre de masse ESI-MS d'une protéine entière MAT1 ( $M_A$ :  $35864.3 \pm 0.2$  Da et  $M_B$ :  $35944.2 \pm 0.3$  Da). En insert, zoom sur les états de charges 29 + à 31+.

Pour confirmer la présence d'une phosphorylation, la protéine est incubée en présence de phosphatase alcaline. Cette enzyme va déphosphoryler la protéine. Après ce traitement, seule la forme non phosphorylée devrait être présente, et cela va ainsi confirmer la présence de phosphorylation sur la protéine (Figure 10).

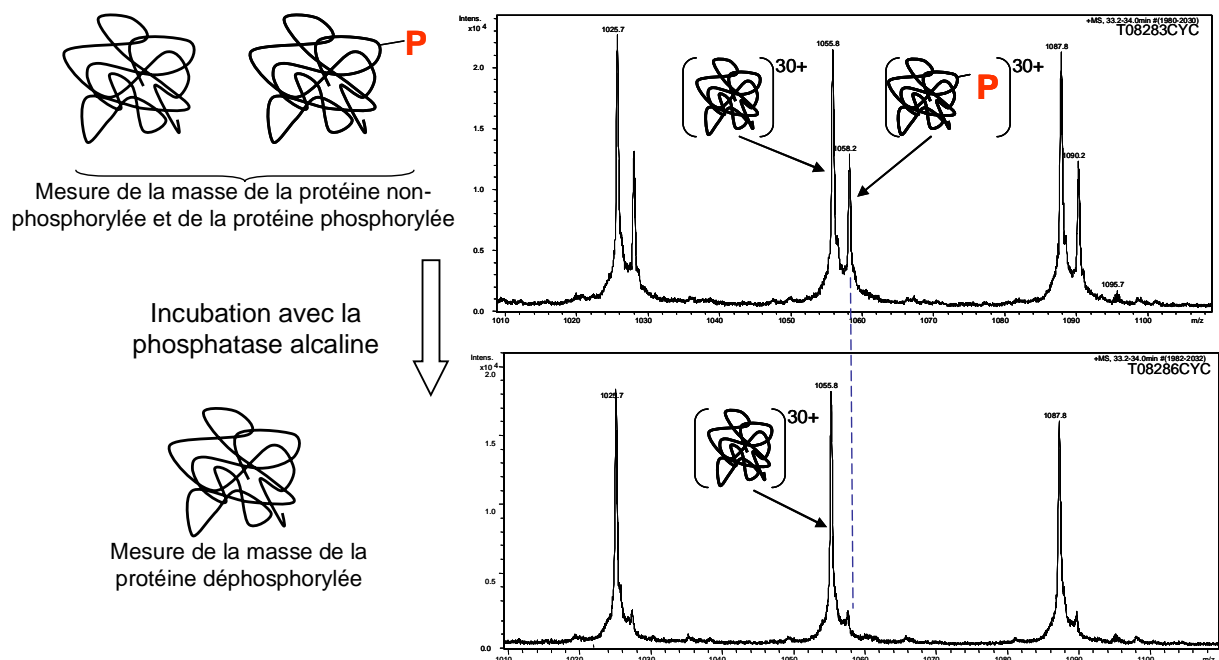


Figure 10 : Détails de spectres de masse ESI-MS de la protéine MAT1 entière ( $M$  :  $35865.1 \pm 0.2$  Da et  $M_{+P}$  :  $35945.7 \pm 0.3$  Da) avant et après déphosphorylation par la phosphatase alcaline. Zoom sur les états de charges 29 + à 31+.



En terme d'efficacité d'ionisation, la phosphorylation, i.e. l'ajout d'un groupement phosphate sur un acide aminé, est considérée comme ayant un impact important au niveau des peptides (jusqu'à 20 acides aminés) [Gropengiesser et al., 2009]. Cela permet donc de considérer l'impact de la phosphorylation sur l'ionisation des protéines comme étant négligeable : l'efficacité d'ionisation est par conséquent la même pour la protéine non phosphorylée que pour la protéine phosphorylée. Aussi il est possible d'estimer le taux de phosphorylation relatif de la protéine en utilisant les intensités (hauteur de pics) correspondants à chaque espèce sur le spectre de masse déconvolué.

$$\text{taux relatif de la protéine } n \text{ fois phosphorylée} = \frac{\text{Intensité de l'espèce } n \text{ fois phosphorylée}}{\sum \text{Intensités des espèces présentes}} \times 100$$

## **2.b. Localisation des phosphorylations sur les peptides**

Afin de pouvoir localiser la phosphorylation au niveau du squelette peptidique, la protéine est digérée de façon enzymatique. Les peptides obtenus sont analysés, par MALDI-MS et/ou par nanoLC-MS/MS. L'analyse MALDI-MS permet d'identifier le peptide phosphorylé, tandis que l'analyse nanoLC-MS/MS permet de localiser le site de phosphorylation au niveau de l'acide aminé (Figure 11).

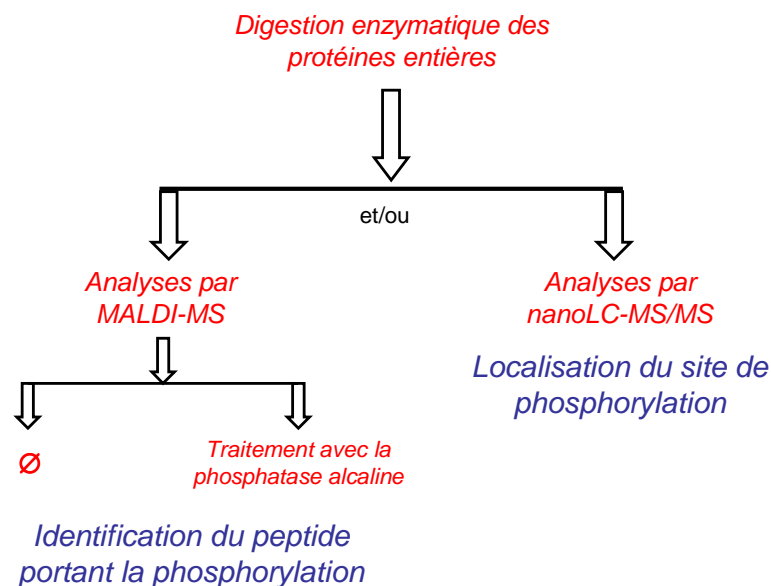


Figure 11 : Stratégie pour la localisation du site de phosphorylation sur les peptides.

Les analyses MALDI-MS sont réalisées avec un dépôt de type goutte séchée et avec deux matrices différentes: l'acide  $\alpha$ -cyano-4-hydrocinnamique (HCCA), utilisé couramment pour les analyses MALDI-MS et l'acide 2,5 dihydrobenzoïque (DHB), qui devrait permettre de mieux visualiser les phosphopeptides. Une partie du mélange peptidique est déphosphorylée par incubation avec la phosphatase alcaline. Une comparaison entre les analyses du mélange peptidique "phosphorylé" et

"déphosphorylé" est réalisée afin d'identifier le peptide portant la phosphorylation. Une fois le peptide portant la phosphorylation repéré, la localisation du site de phosphorylation au niveau de l'acide aminé se fait par analyse nanoLC-MS/MS par fragmentation CID.

Cette méthodologie, développée pour déterminer l'état de phosphorylation, a été appliquée avec succès à deux études, présentées dans les chapitres suivants :

- Etude des phosphorylations du domaine kinase de la protéine DAPK1.
- Etude des phosphorylations du complexe CAK.

## Chapitre II - Etude des phosphorylations du domaine kinase de la protéine DAPK1

Cette étude a été réalisée en collaboration avec Michael Zimmermann et Jacques Haiech de la Faculté de Pharmacie de Strasbourg ainsi que Cédric Atmanene au sein du laboratoire pour l'analyse des complexes non covalents.

### 1. Contexte biologique

La protéine DAPK1 (Death-associated protein kinase ; N° Accession P53355) est une protéine dont les fonctions biologiques sont associées à la mort cellulaire. Elle a été identifiée aussi bien comme un facteur pro-apoptotique, en régulant la suppression de tumeurs, que comme un facteur anti-apoptotique, en maintenant l'équilibre cellulaire par un rôle protecteur [Schneider-Stock et al., 2005]. DAPK1 fait partie de la famille des  $Ca^{2+}$ /calmodulin regulated serine/threonine kinases, qui est composée de 4 autres membres (ZIP(Dlk)-kinase, DRP-1, DRAK1 et DRAK2 [Shohat et al., 2002]). La protéine DAPK1 est composée de nombreux domaines (Figure 12) dont un domaine kinase, ayant une forte homologie de séquence avec les autres membres de la famille.

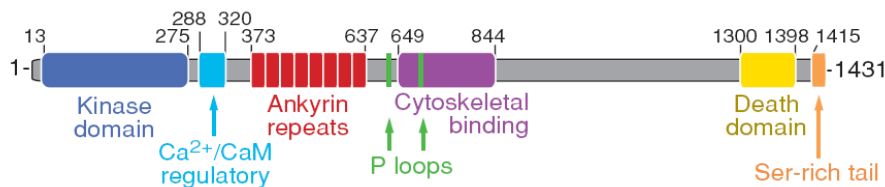


Figure 12 : DAPK1 et ses différents domaines [Bialik & Kimchi, 2006]

Le domaine kinase, commun aux différentes protéines de cette famille, semble être important pour les fonctions cellulaires des  $Ca^{2+}$ /calmodulin regulated serine/threonine kinases [Shohat et al., 2002]. Une étude de cristallisation de ce domaine montre qu'il possède une boucle basique (K46-V56), qui semble être impliquée de façon cruciale dans les mécanismes de régulation de mort ou de survie cellulaire [Tereshko et al., 2001].

Afin d'identifier le rôle de cette boucle basique, deux formes différentes du domaine kinase de DAPK1 ont été produites : la forme entière (résidus 1 à 285), nommé DAPK1 WT et la forme délétée de 6 acides aminés, nommée Delta-DAPK1 (Figure 13). Les 6 acides aminés délétés (S52-S57) font partie de la boucle basique.

```

DAPK-WT      MTVFRQENV D YYDTGEELG SGQFAVVKC REKSTGLQYA AKFIKKRRTK SRRGVS RED
DeltaDAPK    MTVFRQENV D YYDTGEELG SGQFAVVKC REKSTGLQYA AKFIKKRRTK S-----RED

DAPK-WT      IEREVSILKE IQHPNVITLH EVYENKTDVI LILELVAGGE LDFDLAEKES LTEEEATEFL
DeltaDAPK    IEREVSILKE IQHPNVITLH EVYENKTDVI LILELVAGGE LDFDLAEKES LTEEEATEFL

DAPK-WT      KQILNGVYYL HSLQIAHFDL KPENIMLLDR NVPKPRIKII DFGLAHKIDF GNEFKNIFGT
DeltaDAPK    KQILNGVYYL HSLQIAHFDL KPENIMLLDR NVPKPRIKII DFGLAHKIDF GNEFKNIFGT

DAPK-WT      PEFVAPEIVN YEPLGLEADM WSIGVITYIL LSGASPFLGD TKQETLANVS AVNYEFEDEY
DeltaDAPK    PEFVAPEIVN YEPLGLEADM WSIGVITYIL LSGASPFLGD TKQETLANVS AVNYEFEDEY

DAPK-WT      FSNTSALAKD FIRRLLVKDP KKRMTIQDSL QHPWIKPKDT QQALSSAWSH PQFEK
DeltaDAPK    FSNTSALAKD FIRRLLVKDP KKRMTIQDSL QHPWIKPKDT QQALSSAWSH PQFEK

```

Figure 13 : Alignement de séquences de DAPK1 WT et Delta-DAPK1 permettant de visualiser la délétion des 6 acides aminés.

## 2. Etude des protéines entières

Les protéines sont analysées par ESI-MS en conditions dénaturantes. Ces analyses ont montré la présence de deux espèces différentes pour chaque protéine (Figure 14).

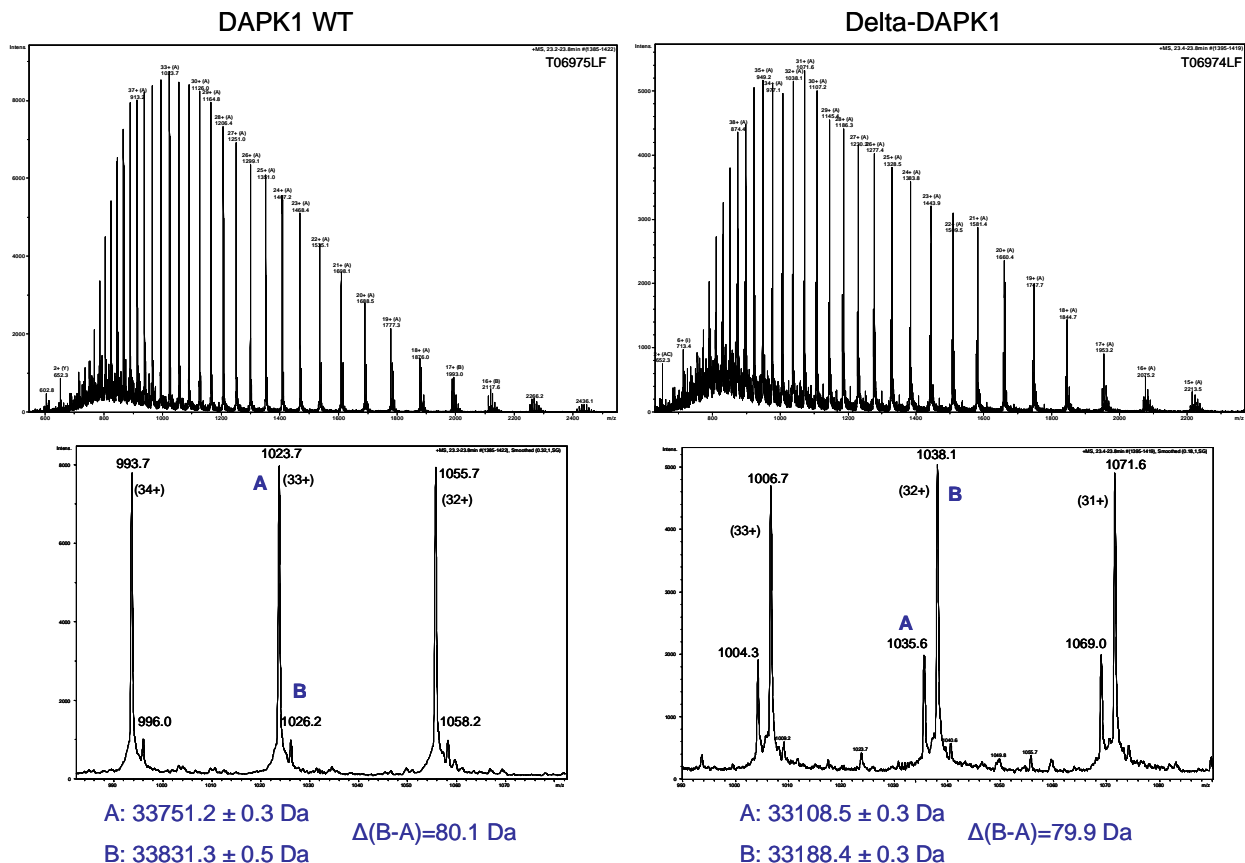


Figure 14 : Spectres de masse ESI-MS des protéines DAPK1 WT et Delta-DAPK1 dans des conditions dénaturantes et zoom sur les états de charges 31+ à 34+.

Les masses mesurées ( $33108.5 \pm 0.3$  Da pour Delta-DAPK1 et  $33751.2 \pm 0.3$  Da pour DAPK1 WT) correspondent aux masses théoriques soustraites de la masse de la méthionine N-terminale (-131 Da). La deuxième population observée pour les deux protéines correspond à une différence de masse de +79.9 Da pour Delta-DAPK1 et +80.1 Da pour DAPK1 WT, ce qui peut suggérer l'existence d'une forme phosphorylée. Ces résultats ont permis d'évaluer un taux de phosphorylation important (78%) pour Delta-DAPK1 tandis qu'un taux plus faible est observé pour DAPK1 WT (24%). Les protéines ont été soumises à une déphosphorylation à l'aide de la phosphatase alcaline pour confirmer la présence de phosphorylation. Aucune déphosphorylation n'a été observée: soit l'hypothèse de la phosphorylation est fausse, soit la phosphatase alcaline n'a pas eu accès aux sites de phosphorylation.

Les protéines entières ont également été analysées par ESI-MS en conditions natives, afin de déterminer leur état d'oligomérisation. Ces analyses ont permis de mettre en évidence que dans les mêmes conditions expérimentales et instrumentales (pH, force ionique du tampon, concentration en protéine et réglages intrinsèques du spectromètre de masse), la protéine DAPK1 WT existe en équilibre monomère/dimère, alors que la protéine Delta-DAPK1 n'est présente que sous forme monomérique (Figure 15).

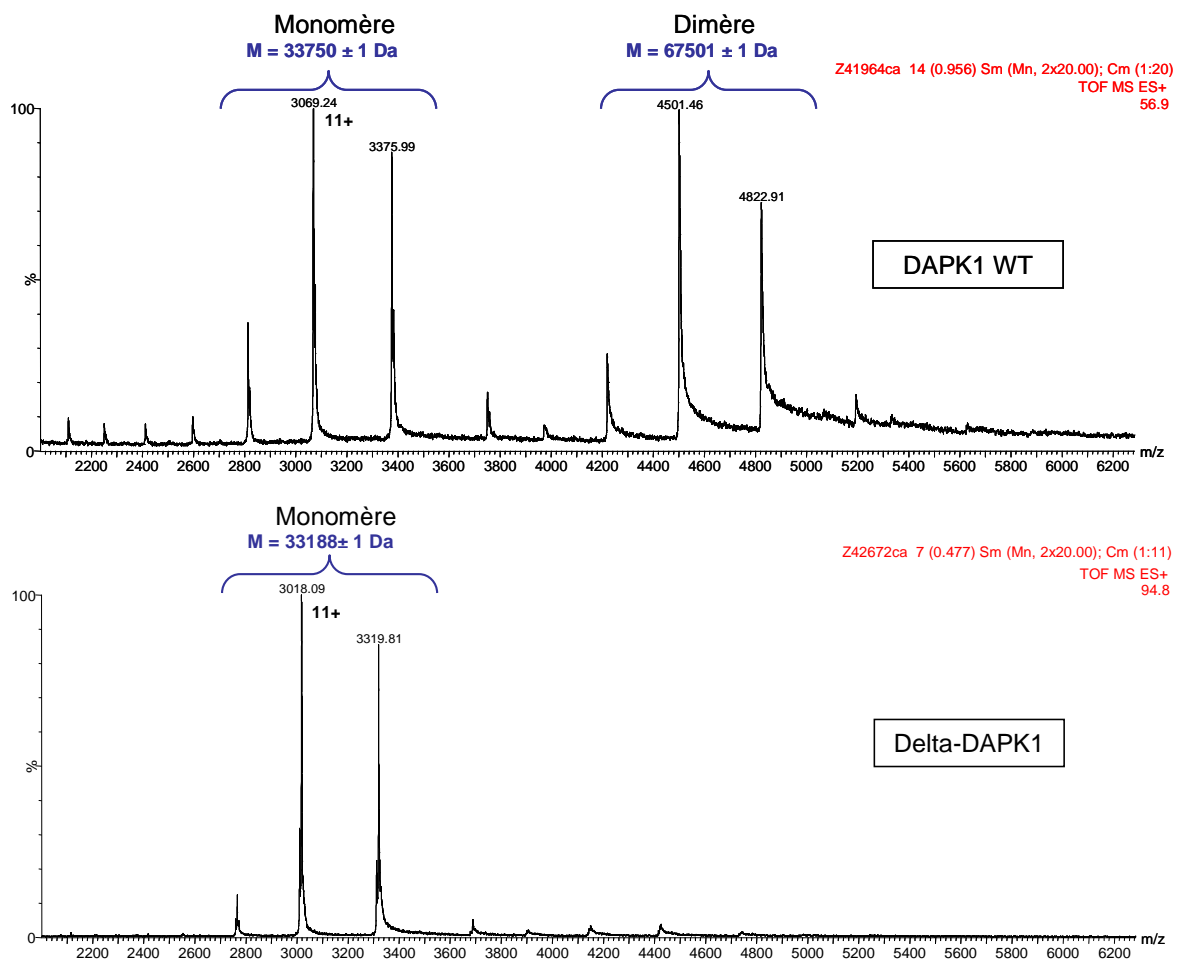


Figure 15 : Etat d'oligomérisation de DAPK1 WT et Delta-DAPK1. Spectres ES-MS de DAPK1 WT (en haut) et Delta-DAPK1 (en bas) obtenus en conditions non-dénaturantes ( $\text{NH}_4\text{Ac}$  50mM, pH 8.8).

Une des hypothèses émise concerne l'influence des phosphorylations sur l'état d'oligomérisation des protéines. Ces phosphorylations sont-elles situées sur le même acide aminé ou sur des acides aminés différents pour les deux protéines ? Il nous a donc semblé intéressant de caractériser plus précisément ces phosphorylations.

Pour déterminer les sites de phosphorylations, chaque protéine a été digérée enzymatiquement soit avec la trypsine soit avec la chymotrypsine. Les peptides obtenus ont été uniquement analysés par nanoLC-MS/MS par couplage chipLC (Agilents Technologies) – trappe ionique HCT Plus (Bruker Daltonics).

### 3. Phosphorylation sur DAPK1 WT

La combinaison des peptides identifiés par les deux digestions a permis d'obtenir un recouvrement de séquence pour DAPK1 WT est de 78% (Figure 16). La digestion avec la trypsine permet d'obtenir une couverture de séquence de 61%, et la digestion avec la chymotrypsine une couverture de séquence de 56%.

MTVFRQENVD DYYDTGEELG SGQFAVVKKC REKSTGLQYA AKFIKKRRTK  
 SSRRGVSRED IEREVSILKE IQHPNVITLH EVYENKTDVI LILELVAGGE  
LFDFLAEKES LTEEEATEFL KQILSGVYYL HSLQIAHFDL KPENIMLLDR  
 NVPKPRIKII DFGLAHKIDF GNEFKNIFGT PEFVAPEIVN YEPLGLEADM  
WSIGVITYIL LSGASPFLGD TKQETLANVS AVNYDFEDEV FSNTSALAKD  
FIRRLLVKDP KKRMTIQDSL QHPWIKPKDT QQALSSAWSH PQFEK

Taux de couverture: 78%

Figure 16 : Recouvrement de séquence pour DAPK1 WT. Les peptides identifiés issus de la digestion trypsique sont soulignés en traits pleins, et ceux issus de la digestion chymotrypsique sont soulignés en pointillés.

L'analyse nanoLC-MS/MS du digest trypsique permet d'observer un ion doublement chargé à m/z 509.9 pouvant potentiellement correspondre au peptide 34-42 STGLQYAAK portant une phosphorylation (Figure 17-A). La fragmentation de ce ion présente un fragment très intense à m/z 460.9 (doublement chargé), ce qui correspond à une perte de H<sub>3</sub>PO<sub>4</sub> (- 49 Da, doublement chargé). Pour interpréter les fragments peptidiques, moins intenses, un zoom a été réalisé (Figure 17-B). L'annotation du spectre permet d'identifier un morceau de séquence "GLQYAA" grâce aux ions y et b-H<sub>3</sub>PO<sub>4</sub>. D'après l'interprétation du spectre de fragmentation, il reste deux acides aminés pouvant porter la phosphorylation : la sérine S34 et la thréonine T35. En s'intéressant aux ions présents sur le spectre de fragmentation, un ion situé à m/z 154.2 (mono-chargé) est visible. Cela pourrait correspondre à l'ion immonium phosphothréonine.

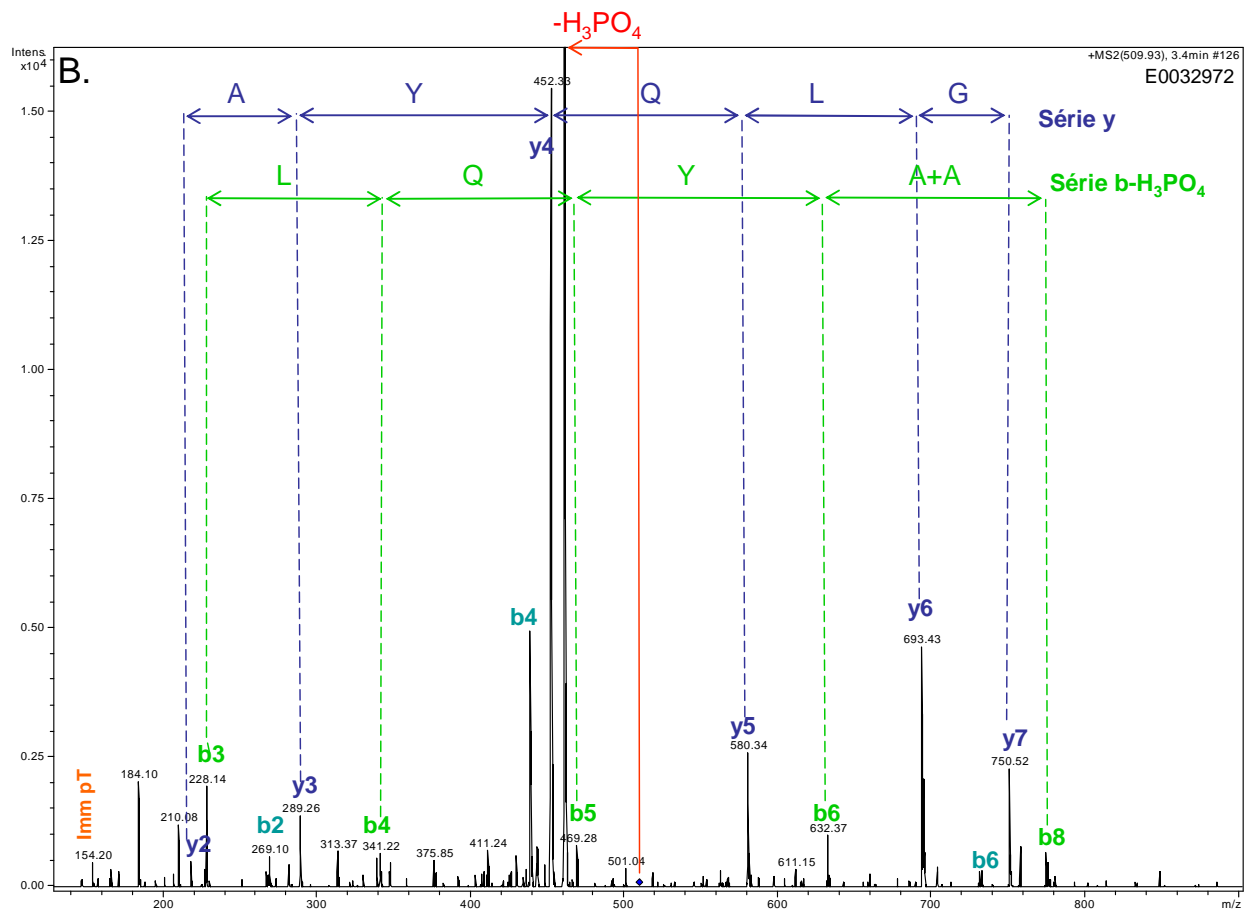
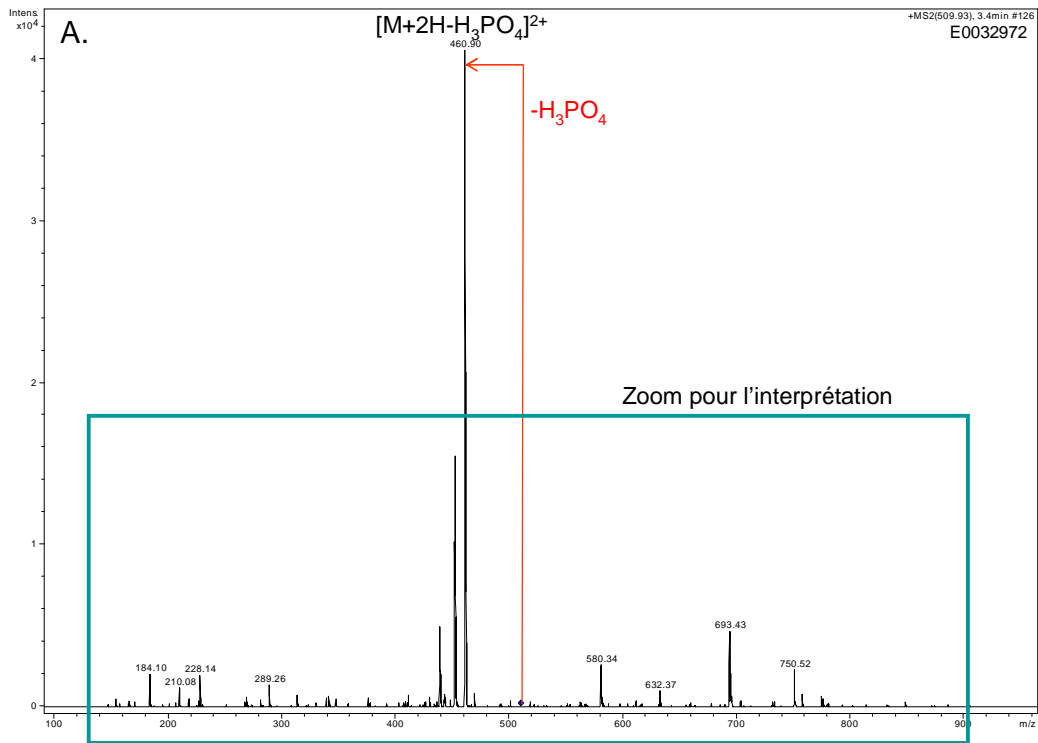


Figure 17 : Localisation du site de phosphorylation situé sur le peptide 34-42 STGLQYAAK de DAPK1 WT.  
 A) Spectre de fragmentation de l'ion  $m/z$  509.9 (2+) issu de l'analyse de la digestion trypsique de DAPK1 WT  
 B) Détail du spectre de fragmentation annoté.

Afin de confirmer l'aspect non artefactuel de cet ion, une nouvelle analyse nanoLC-MS/MS du même échantillon a été réalisée avec le même couplage nanoLC-MS/MS en optimisant la méthode pour la détection de cet ion. Pour cette analyse, les paramètres de sélection de l'ion parent et la gamme de masse utilisée pour la fragmentation ont été modifiés. Pour la sélection de l'ion parent, la gamme de masse a été réduite de 200-2500 m/z à 505-515 m/z, l'ion parent recherché ayant un m/z de 509.9. La gamme de masse de fragmentation a été paramétrée de 50 à 350 m/z au lieu de 50 à 2800 m/z afin de pouvoir visualiser l'ion immonium phosphothréonine recherché (m/z 154.0). Cette méthode a permis d'obtenir un spectre de fragmentation de l'ion m/z 510.0 (Figure 18) ce qui permet de confirmer la présence de l'ion m/z 154.0 correspondant à l'ion immonium phosphothréonine.

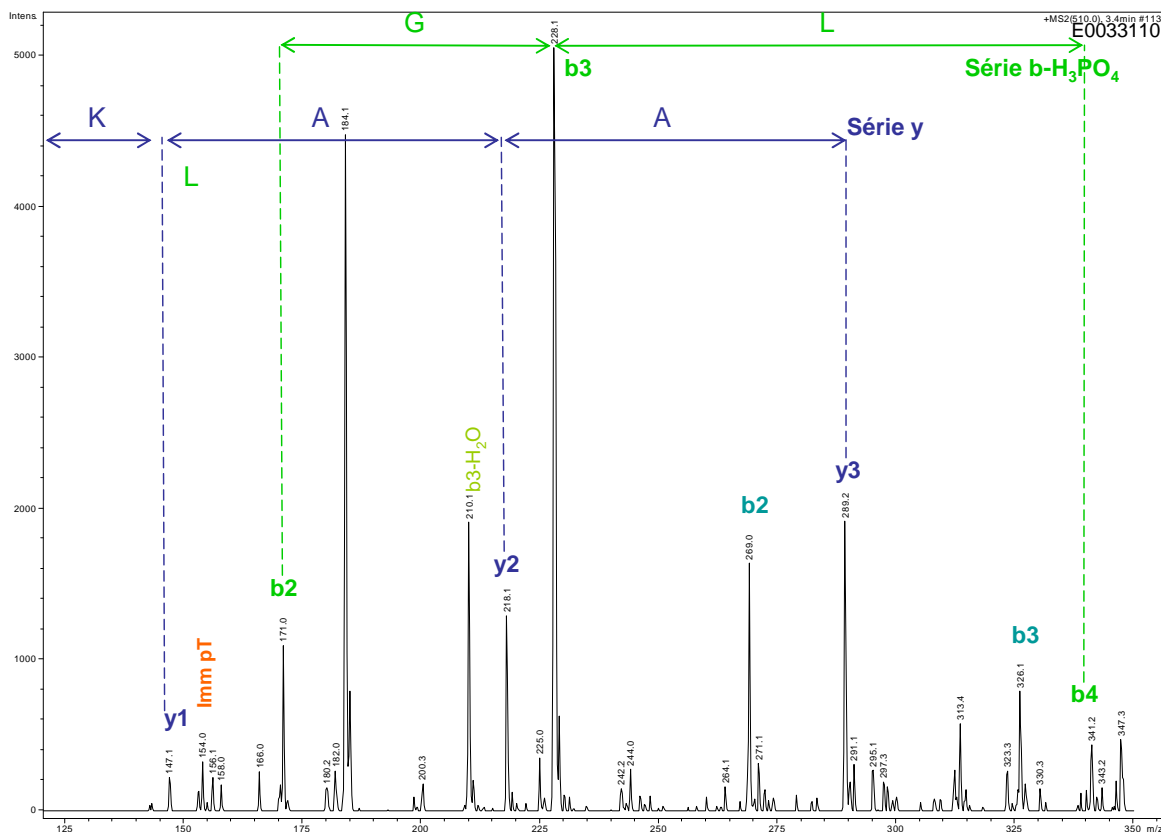


Figure 18 : Confirmation de l'ion immonium à m/z 154.0. Spectre de fragmentation de l'ion m/z 510.0 (2+) issu de l'analyse MS/MS (gamme de fragmentation réduite) de la digestion trypsique de DAPK1 WT correspondant au peptide 34-42 STGLQYAAK.

Les deux analyses nanoLC-MS/MS ont permis de mettre en évidence une phosphorylation sur de la thréonine T35 du peptide 34-42 de la protéine DAPK1 WT.

L'analyse LC-MS de la protéine entière a permis de montrer que 24 % de DAPK1 WT sont phosphorylés, donc on peut supposer que le même peptide non-phosphorylé est présent à environ 76%. Donc la détection du phosphopeptide est complexe dû à sa faible proportion au sein du mélange peptidique. C'est pourquoi, l'analyse nanoLC-MS/MS de DAPK1 WT permet de mettre en évidence le peptide 34-42 "SpTGLQYAAK" (m/z 509.9, doublement chargé) mais aussi le peptide 34-42



"STGLQYAAK" non-phosphorylé (m/z 496.9, doublement chargé). L'extraction du courant d'ion de ces ions permet de mettre en évidence l'importance du peptide non-phosphorylé par rapport au peptide phosphorylé (Figure 19). Cela peut s'expliquer, non seulement par la proportion du peptide non phosphorylé par rapport au peptide phosphorylé après digestion enzymatique, mais aussi par la plus faible efficacité d'ionisation des peptides phosphorylés par rapport aux peptides non-phosphorylés. L'ensemble de ces faits montre la difficulté pour détecter un phosphopeptide peu présent dans le mélange peptidique.

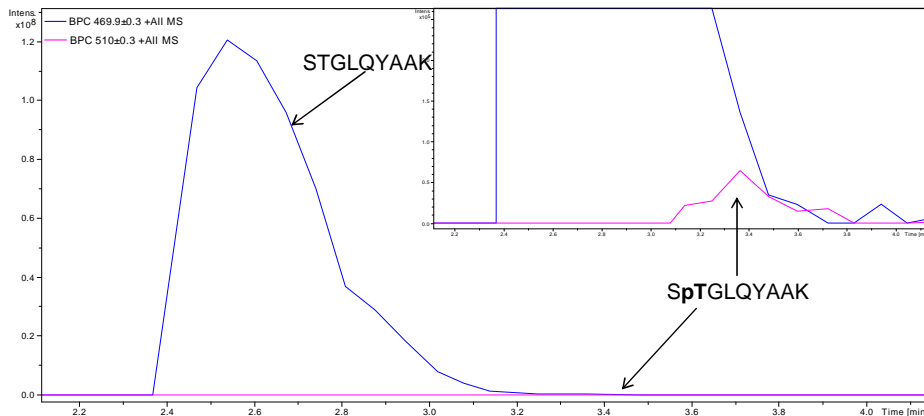


Figure 19 : Extraction des ions à m/z 496.9 correspondant au peptide 34-42 "STGLQYAAK" non-phosphorylé et m/z 510.0 correspondant au peptide 34-42 "SpTGLQYAAK" portant une phosphorylation située sur la thréonine T35.

On peut remarquer que dans notre cas, le peptide non phosphorylé est élué avant le peptide phosphorylé. Le fait d'ajouter un groupement phosphate sur un peptide diminue son hydrophobicité : on s'attend donc à ce que des peptides phosphorylés soient élués avant les peptides non-phosphorylés correspondants lors d'une séparation sur colonne de phase inverse. Bien que pour la majorité des cas l'élution se déroule dans cet ordre, Steen *et al.* ont montré que les peptides phosphorylés pouvaient être élués juste après les peptides non-phosphorylés correspondants, dû à une compensation de l'hydrophobicité par une neutralisation apparente des charges du peptide [Steen *et al.*, 2006]. Dans notre cas, le peptide possède un seul résidu basique (la lysine), donc le peptide sans phosphorylation a donc une charge apparente 2+, alors que une fois phosphorylé, donc avec un groupement chargé négativement, la charge apparente du peptide est de 1+. Ceci peut expliquer l'ordre d'élution inhabituel du peptide 34-42 STGLQYAAK.

#### 4. Phosphorylation sur Delta-DAPK1

Le même travail a ensuite été effectué sur Delta-DAPK1. La combinaison des peptides identifiés par les deux digestions a permis d'obtenir un recouvrement de séquence pour Delta-DAPK1 de 91% (Figure 20). La digestion avec la trypsine permet d'obtenir une couverture de séquence de 66%, et la digestion avec la chymotrypsine une couverture de 65%.

MTVFRQENVD DYYDTGEELG SGQFAVVKKC REKSTGLQYA AKFIKKRRTK  
-----  
SREDIEREVS ILKEIQHPNV ITLHEVYENK TDVILILELV AGGELFDFLA  
-----  
EKESLTEEEA TEFLKQILNG VYYLHSLQIA HFDLKPENIM LLDRNVPKPR  
-----  
IKIIDFGLAH KIDFGNEFKN IFGTPEFVAP EIVNYEPLGL EADMWSIGVI  
-----  
TYILLSGASP FLGDTKQETL ANVSAVNYEF EDEYFSNTSA LAKDFIRLL  
-----  
VKDPKKRMTI QDSLQHPWIK PKDTQQALSS AWSHPOFEK

Taux de couverture: 91%

Figure 20 : Recouvrement de séquence pour Delta-DAPK1 Les peptides identifiés issus de la digestion trypsique sont soulignés en traits pleins, et ceux issus de la digestion chymotrypsique sont soulignés en pointillés.

Un peptide phosphorylé ( $m/z$  687.2, doublement chargé) a été identifié par analyse nanoLC-MS/MS issue de la digestion avec la trypsine (Figure 21-A). La fragmentation de l'ion  $m/z$  687.2 (doublement chargé) présente un fragment très intense à  $m/z$  654.5 (doublement chargé), ce qui correspond à une perte de  $H_3PO_4$  (- 49 Da, doublement chargé). Pour pouvoir interpréter les autres fragments, moins intenses, un zoom a été réalisé (Figure 21-B). L'annotation du spectre permet d'identifier un morceau de séquence "pTIQDSLQHPWIK" grâce aux ions  $y$  et  $b-H_3PO_4$ . Le site de phosphorylation est déterminé par la présence des ions  $b_2-H_3PO_4$  et  $b_3-H_3PO_4$  correspondant à l'acide aminé "acide dehydroaminobutyrique" (Dhb). L'acide aminé "phosphothréonine" (pT) est également visible grâce aux ions  $b_2-H_3PO_4$  et  $b_3$ . Ce spectre MS/MS permet donc d'identifier le peptide 257-272 RMpTIQDSLQHPWIKPK portant une phosphorylation située sur la T259.

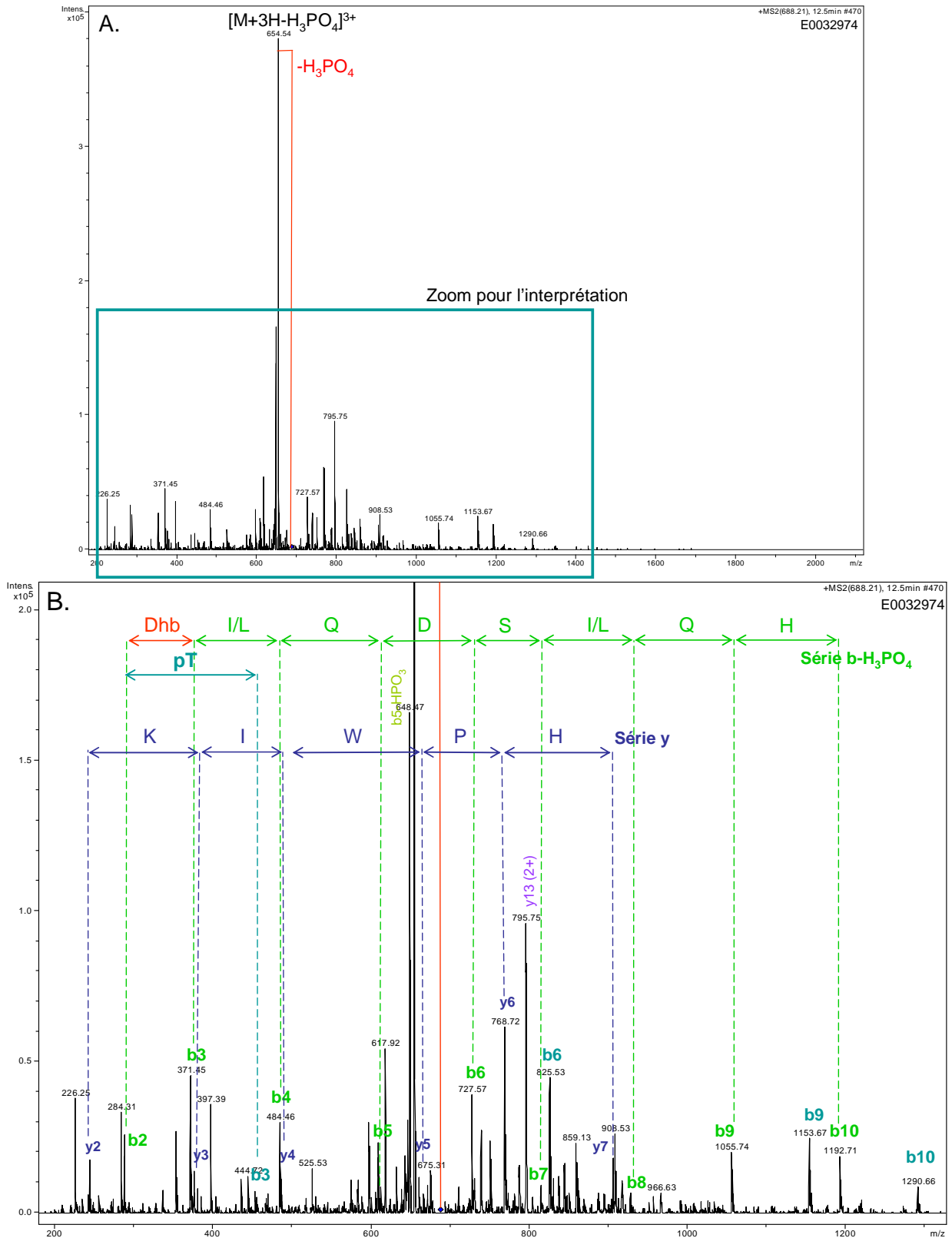


Figure 21 : Localisation du site de phosphorylation situé sur le peptide 257-272 RMTIQDSLQHPWIKPK de Delta-DAPK1. A) Spectre de fragmentation de l'ion  $m/z$  687.2 (2+) issu de l'analyse de la digestion trypsique de Delta-DAPK1 B) Détail du spectre de fragmentation annoté.

Pour savoir si le peptide **RMpTIQDSLQHPWIKPK** portant une phosphorylation située sur la T259 est détecté dans l'analyse du digest de la protéine DAPK1 WT, nous avons vérifié l'absence de l'ion  $m/z$  687.2 correspondant à l'ion doublement chargé du peptide et de l'ion  $m/z$  654.5 correspondant à l'ion doublement chargé de la perte de phosphorylation de l'ion **RMpTIQDSLQHPWIKPK** (Figure 22). Le peptide **SpTGLQYAAK** présent dans la protéine DAPK1 WT n'était pas détecté dans la protéine Delta-DAPK1. De la même façon, le peptide **SpTGLQYAAK** présent dans la protéine DAPK1 WT n'est pas détecté pour la protéine Delta-DAPK1.

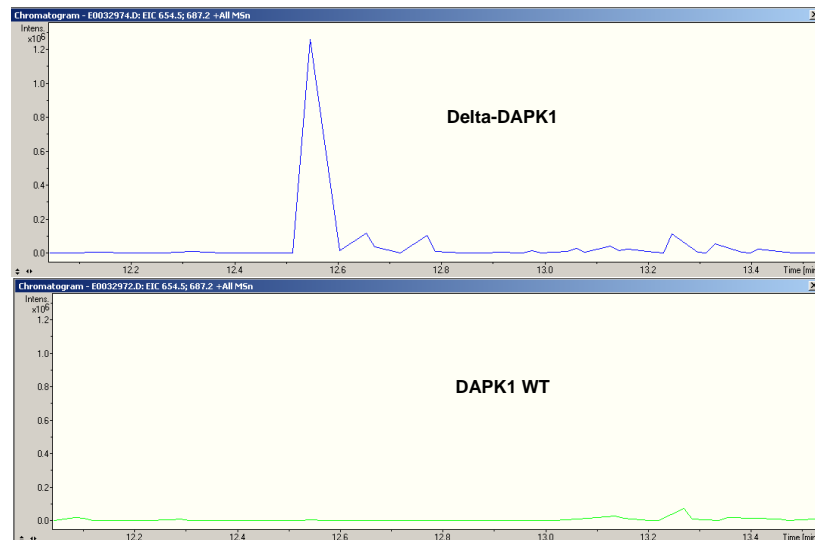


Figure 22 : Extraction des ions à  $m/z$   $687.2 \pm 0.5$  et  $654.5 \pm 0.5$  correspondant à l'ion doublement chargé du peptide 257-272 **RMpTIQDSLQHPWIKPK** portant une phosphorylation située sur la T259 et à l'ion doublement chargé de l'ion **RMDhbiQDSLQHPWIKPK**.

## 5. Conclusion

La caractérisation des protéines Delta-DAPK1 et DAPK1 WT a été réalisée par analyse nanoLC-MS/MS. Les digestions trypsique et chymotrypsique ont permis d'obtenir un bon recouvrement de séquence pour les deux protéines.

L'étude de la protéine DAPK1 WT permet de mettre en évidence une phosphorylation située sur la thréonine T35, tandis que l'étude de Delta-DAPK1 montre une phosphorylation située sur la thréonine T259. Les deux phosphorylations ne sont pas situées dans la même zone sur la structure tridimensionnelle du dimère de DAPK1 [Tereshko et al., 2001] : la phosphorylation T35 semble être située dans l'interface de dimérisation (en vert dans la figure 23), tandis que la T259 semble être plus éloignée de cette interface (en rouge dans la figure 23). Des études complémentaires devront être menées pour comprendre l'influence de la localisation des phosphorylations.

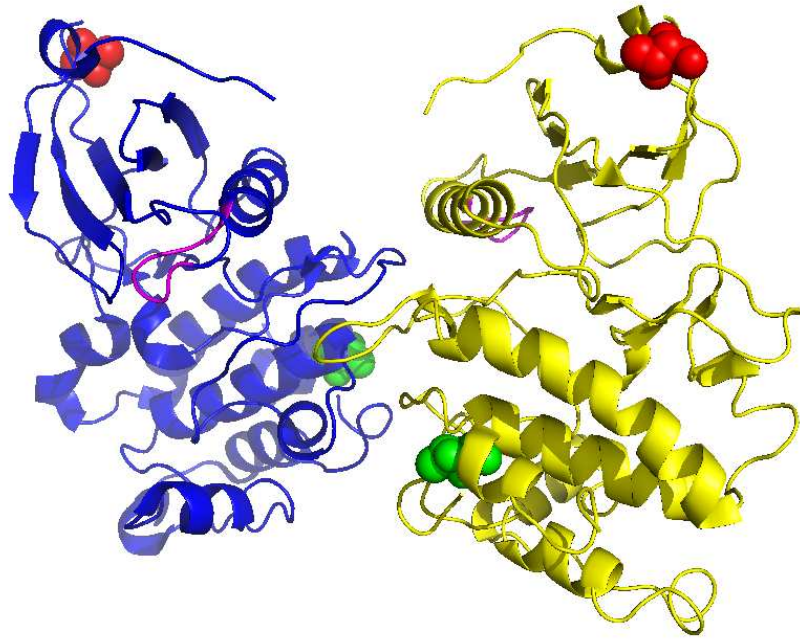


Figure 23 : Représentation tridimensionnelle du dimère de DAPK, avec en vert la thréonine T35 identifiée comme phosphorylée pour DAPK1 WT et en rouge la thréonine T256 identifiée comme phosphorylée pour DeltaDAPK1.

Ces résultats feront l'objet d'une publication dont la version provisoire est jointe à ce manuscrit en page suivante.



# Death-Associated Protein Kinase: the crucial role of a specific basic loop in DAPK's homodimerization process

Michael Zimmermann<sup>1</sup>, Cédric Atmanene<sup>2</sup>, Laetitia Fouillen<sup>2</sup>, Alain Van Dorsselaer<sup>2</sup>, Sarah Sanglier-Cianferani<sup>2</sup>, Jacques Haiech<sup>1</sup>

<sup>1</sup>School of Biotechnology and School of Pharmacy 74, route du rhin, 67401 ILLKIRCH, France

<sup>2</sup>Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC-DNA, Université de Strasbourg, CNRS, UMR7178 ; 25 rue Becquerel, 67 087 STRASBOURG, France

## Summary:

Death-Associated Protein Kinase (DAPK) is a member of the Ca<sup>2+</sup>/calmodulin regulated Serine/Threonine protein kinases. Its biological function has been associated with induced cell death, which makes this protein an interesting biological target for research related to neuronal injuries and oncology. In the here presented in vitro study we show the possibility of homodimerization of DAPK's catalytic domain and the crucial role of a specific basic loop in this process. Furthermore, two new auto-phosphorylation sites within the kinase domain have been localized and related to the oligomerization state of DAPK's catalytic domain. In a chemical biology approach, these findings were then exploited in order to develop a specific screening assay to analyze the molecular mechanisms that regulate the process of homodimerization.

Based on our experimental findings, structural analysis of the protein and previously reported characteristics of the DAPK-family proteins, a comprehensive model of the regulation and activation of DAPK's core domain is proposed in the present report. For the first time, we suggest an induced monomerization process as being a part of the molecular switch between cellular survival and the different forms of induced cell death mediated by the auto-inhibited and the activated form of DAPK, respectively.

## Keywords:

Death Associated Protein Kinase (DAPK); Homodimerization; Induced Cell Death; Fluorescent Anisotropy; Noncovalent mass spectrometry

## Abbreviations:

DAPK: Death Associated Protein Kinase

CaM: Calmodulin

HTS: High Throughput Screening

nanoESI-MS: Nanoelectrospray ionization mass spectrometry

huMLCK: Human Myosin Light Chain Kinase

NH<sub>4</sub>Ac: Ammonium acetate

nanoLC-MS/MS: nanoliquid chromatography coupled tandem mass spectrometry

## Introduction:

Death-Associated Protein Kinase (DAPK) was first identified in a functional screening assay that aimed at identifying genes involved in interferon- $\gamma$  induced cell death [i]. Besides its C-terminal kinase domain the gene codes for a calmodulin(CaM)-binding sequence, ankyrin repeats, P-loops, a cytoskeleton binding region, and a death domain (for review see [ii]). Various studies were performed to investigate the role and the activation mechanism mediated by the interplay of CaM and the auto-phosphorylable binding region of DAPK, which directly follows C-terminal to its catalytic core [iii, iv]. The heterologous expression of the isolated N-terminal region of the *dapk*-gene in *E. coli* results in a constitutively active kinase, whose structure has been resolved [v] and which was earlier exploited for the identification of a first synthetic substrate peptide [vi], as well as for the discovery of inhibitory small molecules [vii,viii].

Even though the biological function of DAPK always centers on induced cell death, its observed implications in this process are diverse in terms of upstream and downstream occurring events. These eventually lead to different cellular outputs that might even be converse among themselves. In this way, a growing dataset describes DAPK as an anti-apoptotic factor [ix] or even a growth factor [x] rather than a pro-apoptotic trigger under given conditions. Another decision, which at least partially seems to be taken on the DAPK level, is the nature of death a cell is undertaking after according induction. Recent evidence describes autophagy (type II cell death [xi]) as a further cellular destiny procured by DAPK besides the well-established apoptotic pathways (type I cell death).

All these observations led to an understanding of DAPK's role as being a "molecular switch" or "integration point" guiding a cell's fate into one or the other direction depending on the prevailing environment [xii]. The presence of a "death-associated multi-protein complex" is thought to be required for the determination of the downstream endpoint of the signaling cascade. Consequently, a lot of effort was put in deciphering various members of such putative protein complexes. Recent results resulting from systematic studies such as investigations on the proteomic level [xiii] or screening assays of peptide libraries [xiv, xv, xvi] led to the identification of physical interaction partners of DAPK in addition to the known affiliated proteins [ii, vi].

The documented formation of functional heterodimers between the catalytic cores of DAPK and its close family member ZIPK [xvii], which share a sequence identity of about 80% [xviii], let us hypothesize that DAPK could self-associate to form a homodimer. Noncovalent nanoelectrospray ionization mass spectrometry (nanoESI-MS) revealed the possible formation of such homodimer as well as the previously delineated importance of the DAPK-family's specific basic loop in such a protein-protein interaction [xvii]. We furthermore evidenced a direct correlation between DAPK homodimerization propensity and the phosphorylation rate of two newly identified autophosphorylation sites within the catalytic domain.



We subsequently screened a fluorophore-labeled peptide-like library in order to identify traceable binding partners of DAPK's catalytic domain, which then served as a reporter of the protein dimerization state. In the screening for such molecular tools we used a high throughput approach based on fluorescent anisotropy measurements. The same technique was also used in competitive assays that evidenced the spatial independency of the probe's binding site from the one of the peptide substrate [vi] as well as the one of ATP/ADP.

We finally propose a comprehensive model of the molecular mechanisms that might play an important role in DAPK's switch between various cellular pathways and its related fates.

## **Experimental Procedure:**

### *Protein expression and purification*

The protein expression plasmids pASK-IBA3 (Sigma) carrying the cDNA of DAPK's catalytic core (residues 1-285) as well as the ones resulting in the described mutated protein DAPKdel ( $\Delta$ S52-S57), were kindly provided by D. Martin Watterson (Northwestern University, Chicago IL). The proteins were produced and purified in *Escherichia Coli* (DH5 $\alpha$ ) as previously described [xix] except for some modifications that resulted from various optimization steps. These contained expression assays at 37 and 22°C for expression durations of 1 to 7 hours after the induction with 0.2 or 0.4 $\mu$ g/mL anhydrotetracyclin (IBA Göttingen, Germany). Thereafter, the expression yield of each condition was analyzed by a colorimetric dot blot assay (material by Biorad, Hercules, CA; nitocellulose membrane 0.2mm by BDH, Poole, England) with the HRP-tagged anti-Streptag antibody (IBA Göttingen, Germany) and 4-chloro-1-naphtol (Sigma Aldrich, St. Louis, MO). The intensities of the colored dots were quantified by the program ImageJ (<http://rsbweb.nih.gov>) and the conditions, which resulted in the maximal intensity were chosen. This led to the expression of the proteins at 22°C for 4h in tryptone soy broth (Biorad, Hercules, CA). After lysis, sonication and centrifugation, as described in [xix] we applied the clarified lysate directly to a high capacity streptactin superflow resin (IBA Göttingen, Germany). In addition to the manufacturer's recommendations a salt wash of 3CV of the high salt washing buffer (100mM TrisHCl pH 8, 500mM NaCl) and the subsequent regeneration by 2CV of washing buffer (100mM TrisHCl pH 8; 150mM NaCl) resulted in an increased purity of the proteins. Using a 10kDa cut off concentrator (Satorius, Göttingen, Germany) the elution buffer (100mM TrisHCl pH 8; 150mM NaCl, 1mM EDTA; 2.5mM desthiobiotin) was exchanged by the storing buffer (20mM TrisHCl pH 7.5, 250mM NaCl, 1mM EDTA, 1mM DTT) and the volume was finally adjusted in order to achieve protein concentrations between 200 and 700 $\mu$ M. The concentrations were determined by a NanoDrop spectrometer (Thermo Scientific, Waltham MA). Proteins integrity was checked by nanoESI-MS mass measurement under denaturing conditions as described below and their purity was analyzed by capillary electrophoresis (Bioanalyzer, Agilent Technologies Santa Clara CA). The homogeneity of the protein solutions was investigated by dynamic light scattering measurements (Dynapro Model 801, Wyatt Technology Europe GmbH, Dernbach, Germany).

### *nanoESI-MS for the study of noncovalent complexes*

Prior to any mass spectrometry experiments, proteins buffer was exchanged against a 10 mM ammonium acetate (NH<sub>4</sub>Ac) solution at pH 8.8 using microcentrifuge gel filtration columns (Zeba 0.5 ml, Thermo Scientific, Rockford, IL). Proteins concentration was determined spectrophotometrically.

ADP solution was desalted using a 1 ml anionic exchange column (HiTrap Q-Sepharose, GE Healthcare, Little Chalfont, UK) in order to exchange sodium counter-ions against volatile ammonium ions. ADP concentration was then determined spectrophotometrically and magnesium acetate was subsequently added to this solution to reach a 1:1 ADP:Mg<sup>2+</sup> molar ratio.

NanoESI-MS analyses were carried out on an electrospray time-of-flight mass spectrometer (LCT, Waters, Manchester, UK) equipped with an automated chip-based nanoESI source (Triversa Nanomate, Advion Biosciences, Ithaca, NY) operating in the positive ion mode. External calibration was performed with the multiply charged ions produced by a 2 µM horse heart myoglobin solution diluted in a 1:1 (v/v) water : acetonitrile mixture acidified with 1% (v/v) formic acid.

Purity and homogeneity of DAPKwt and DAPKdel were first checked in denaturing conditions by diluting proteins to 2 µM in a 1:1 (v/v) water : acetonitrile mixture acidified with 1% (v/v). Analyses in non-denaturing conditions were then performed by diluting proteins to 5 µM in NH<sub>4</sub>Ac buffer at pH 8.8 (adjusted with ammonia). Different NH<sub>4</sub>Ac concentrations were used to check their influence on proteins oligomerization state. DAPKwt autophosphorylation experiments were carried out by mixing the protein at 30 µM with 10 mM ATP and 10 mM MgCl<sub>2</sub> in 50 mM NH<sub>4</sub>Ac pH 8.8. Following a 10 h incubation period at 37°C, the protein was buffer exchanged against 50 mM NH<sub>4</sub>Ac (pH 8.8) using microcentrifuge gel filtration columns (Zeba micro, Thermo Scientific, Rockford, IL). In order to investigate the influence of ADP-Mg on DAPK's oligomerization state, titration experiments were performed using a fixed concentration of DAPKwt (5 µM) and increasing amounts of ADP-Mg. Experiments were realized after a careful optimization of instrumental parameters. Particularly, the pressure in the first pumping stage was raised up to 6 mbar using a throttling valve and the acceleration voltage applied on the sample cone was set to 100 V. In these conditions, optimal ion desolvation and ion transmission are achieved while the integrity of noncovalent complexes is preserved during their transfer into the gas phase. Data analysis was performed with MassLynx 4.1 (Waters, Manchester, UK).

#### *Identification of phosphorylation sites by mass spectrometry*

Enzymatic digestion and subsequent MS analysis for the identification of phosphopeptides were performed in the same manner for DAPKwt and DAPKdel.

Briefly, proteins (60 pmol) were first denatured with 1µl ACN during 1h. Enzymatic digestion was performed at 25°C for 16 h after addition of either porcine trypsin (5µl at 12.5ng/µl, Promega, Madison, WI, USA) or bovine α-chymotrypsin (5µl at 25ng/µl, Roche, Indianapolis, IN).

All nanoLC-MS/MS analyses were performed using an Agilent 1100 series HPLC-Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to an HCT plus ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The microfluidic chip consisted of a Zorbax 300SB-C18 enrichment column (40 nL, 5µm) followed by a Zorbax 300SB-C18 analytical column (150mm x 75µm, with a

5 $\mu$ m particle size). Details concerning the nanoLC gradient and MS settings are given in supplementary data.

The MS and the MS/MS data were searched using a local Mascot server (version MASCOT 2.2.0, MatrixScience, UK) against the Swiss-Prot database (v56.8, 410518 sequences). Searches were performed with a mass tolerance of 0.5 Da in MS mode and 0.5 Da in MS/MS mode and with the following parameters: trypsin specificity, one missed cleavage and chymotrypsin specificity, two missed cleavages, methionine oxidation, protein amino-terminal acetylation and phosphorylation (STY) as variable modifications, and without any taxonomic restriction.

#### *Production and HTS of the fluorescent library*

Designing and combinatorial synthesis of the fluorescent compound library followed the previously described procedures [i]. Fluorescence polarization assays were performed using a Victor 3 apparatus (Perkin-Elmer Waltham MA). Lissamine-tagged compounds of the CHPO Compounds Library (Institut Gilbert Laustriat, Strasbourg, France) were diluted in the assay buffer (50mM HEPES, 150mM KCl, 1mM MgCl<sub>2</sub>, pH 7.5) to a working concentration of 0.2 $\mu$ M. 15 $\mu$ l of these dilutions were transferred to each well of the assay plate (Corning Costar 96-well black polystyrene plates, Model 3686, Corning, Acton, MA). All pipetting of the HTS was done on Biomek<sup>®</sup> 2000 (Beckman Coulter, Fullerton CA). The fluorescent polarization degrees (mFP) were measured at an excitation and emission wavelength of 530 and 610nm respectively. 15 $\mu$ l of the assay buffer with DAPK, at a concentration of 4 $\mu$ M, were added to each of the wells. For each plate blank control wells were measured, which allowed a background correction by subtracting blank parallel and perpendicular values from the corresponding intensities of each of the assay wells.

#### *ATP, ADP, substrate peptide and p21 titrations*

All titration measurements were performed on a FlexStation 3 (Molecular Devices Union City CA). The other experimental conditions were described in section “*HTS of the fluorescent library*“. A twofold dilution cascade of the protein between 200 $\mu$ M and 0.8 $\mu$ M was prepared and 10 $\mu$ l of these solutions were dispatched on a 96-well assay plates. 10 $\mu$ l of the probe solution (0.2 $\mu$ M of CHPO 187-3-H11-para in assay buffer) were added to the wells and the fluorescent polarization was recorded as specified before. Subsequently, 1 $\mu$ l of concentrated ATP, ADP or substrate peptide [vi] (Tocris Bristol, UK) in assay buffer with 0.1 $\mu$ M of the fluorescent probe were pipetted to the wells in order to subsequently obtain the following final concentrations of ATP and ADP: 0.4, 0.8, 1.6, 3.2, 6.3, 12.5, 25, 50, and 1000 $\mu$ M. The highest concentration was not produced for the peptide substrate.

The titrations with the two p21 peptide fragments of 20 amino acids (peptide number 8 and 14 [xiv]) were performed analogously to the one with the substrate peptide.

### *Titration under monomeric and dimeric conditions*

The performed titrations followed the same protocol as described under “ATP, ADP and substrate peptide titrations”, but under buffer conditions analogous to the ones listed under “nanoESI-MS for the study of noncovalent complexes“. For investigations of the monomer and the dimer the NH<sub>4</sub>Ac concentrations were kept constant at 250mM and 5mM, respectively. To induce gradual monomerization of the DAPK’s catalytic core, the NH<sub>4</sub>Ac concentrations was stepwise augmented (5, 10, 25, 50, 100 and 250mM) by the addition of highly concentrated stock solution (1mM).

To determine the K<sub>d</sub>-values of  $\gamma$ -[6-Aminohexyl]-ATP-Atto495 (Jena Bioscience, Jena, Germany) its concentration was set to 0.1 $\mu$ M, while the rest of the conditions were chosen analogously to the ones described above. Fluorescent polarization was recorded at an excitation and emission wavelength of 493nm and 520nm respectively.

## Results and Discussion:

### *Homodimerization of DAPK's catalytic core*

The question of DAPK oligomerization state was addressed using automated chip-based nanoESI-MS. DAPKwt was first analyzed under denaturing conditions highlighting the presence of a first species (figure 1a), which molecular weight ( $33751.1 \pm 0.1$  Da) is in agreement with the molecular weight calculated from the theoretical amino acid sequence without the N-terminal methionine ( $33750.4$  Da). A second less intense ion distribution is also observed with a molecular weight of  $33830.7 \pm 0.3$  Da, displaying thus a mass increase of  $+79.6$  Da relative to the first species, which can be attributed to a putative phosphorylation of DAPKwt. In a second step, DAPKwt was analyzed under non-denaturing conditions, revealing the detection of two ion distributions detected in the mass ranges  $m/z$  3000-4000 and  $m/z$  4000-5500, respectively (figure 1c). The former corresponds to the 10+ to 12+ charge states of monomeric DAPKwt (measured MW =  $33750 \pm 1$  Da) while the latter is assigned to the 14+ to 16+ charge states of dimeric DAPKwt (measured MW =  $67501 \pm 1$  Da). These results evidence thus a monomer / dimer equilibrium of DAPKwt.

Interestingly, this monomer / dimer equilibrium was shown to be sensitive to the buffer ionic strength. Indeed, DAPK's oligomerization state is progressively shifted towards the monomeric form of the protein when buffer concentration is increased from 10 to 250 mM (figures 2a to 2d). These observations could be corroborated by DLS-measurements. Despite polydisperse solutions arising from high protein concentrations required by DLS experiments (500  $\mu$ M), results confirmed that monomer / dimer ratio increases with increasing ammonium acetate concentrations (data not shown). The shift observed from dimeric to monomeric DAPKwt with increasing  $\text{NH}_4\text{Ac}$  concentrations suggests that this protein-protein association involves electrostatic interactions which are sensitive to the ionic strength of the medium.

### *Involvement of DAPK's basic loop in the homodimerization process*

The ionic strength dependent monomer / dimer equilibrium of DAPKwt prompted us to investigate the role in the dimerization process of a basic loop corresponding to the DAPK-family fingerprint region. This loop is located in close proximity to the ATP binding pocket and putative binding site of the substrate peptide [xix, xxiv] and has been previously associated with the ability to form DAPK-ZIPK heterodimer [xvii].

In the present study a truncated protein was produced by deleting residues S52 to S57 involved in this basic loop (DAPKdel). The size of this deletion was based on DAPK's sequence alignment to the human myosin light chain kinase (huMLCK), whose catalytic domain shows a high degree of sequence identity [i]. Furthermore, although a dimerization process was earlier described for huMLCK [i], it was associated with the CaM-binding domain and not directly with the catalytic domain. Therefore we assumed that the capacity to form homodimers lays in differences of the sequence of the

two proteins. Hence, the deletion S52-S57, which corresponds to the part of the basic loop that is present in DAPK, but completely missing in huMLCK, was produced in the sequence of DAPK's core domain. Even though, DAPK's basic loop is located in close proximity to the ATP pocket and putative binding site of the substrate peptide [xix, ii], its partial absence does not lead to a significant change of the catalytic rate [xix].

DAPKdel was first analyzed in denaturing conditions, evidencing by this way the presence of the protein under two forms (figure 1b). The first species which is the less intense one has a molecular weight of  $33108.8 \pm 0.7$  Da in agreement with the expected mass of the mutated protein after deletion of its N-terminal methionine (33107.7 Da). The second species displays a molecular weight of  $33188.7 \pm 0.4$  Da corresponding to a +79.9 Da mass increase and can putatively be attributed to the phosphorylated protein. In order to determine the oligomerization state of DAPKdel, noncovalent nanoESI-MS analyses were performed in strictly identical experimental and instrumental conditions as those applied to analyze DAPKwt. Interestingly, DAPKdel is only detected as a monomer (MW =  $33108 \pm 1$  Da – figure 1d) even in low ionic strength conditions which were shown to promote DAPKwt dimerization (figures 2e to 2h). These results enlighten thus that this basic loop, which is a unique feature of the kinases belonging to the DAPK family, plays a crucial role in the formation of homodimers. These findings are well supported by the very first crystallographic work that was performed on DAPK's catalytic core by Tereshko et al. Indeed, two different crystal organizations were described and one of which is suggestive for homodimers [v]. Furthermore, the resolved structure of the putative dimers also localizes the basic loop within the interface of the two catalytic cores.

#### *Spatial and quantitative dependence of DAPK's autophosphorylation on its homodimerization process*

Mass spectrometry analyses of DAPKwt and DAPKdel under denaturing conditions revealed that both proteins are present in two distinct forms which differ by about 80 Da (figures 1a and 1b). This mass difference, together with the fact that DAPK displays a kinase activity, led us to consider the possibility of these proteins to be autophosphorylated. Interestingly, figures 1a and 1b evidence a lower extent of phosphorylated protein for DAPKwt compared to DAPKdel (24% and 78% relative intensity, respectively)

In order to confirm the presence of a phosphorylation and to localize the phosphorylation site, both trypsin- and chymotrypsin-digested DAPKwt and DAPKdel were analyzed by nanoLC-MS/MS analysis. Combination of both enzymatic digestions led to 77.9% and 91.3% sequence coverage for DAPKwt and DAPKdel, respectively. Among the identified tryptic peptides, S34-K42 from DAPKwt and R257-K272 from DAPKdel were found to be phosphorylated as evidenced by the loss of  $H_3PO_4$  in tandem mass spectrometry experiments (figure 3). Moreover, interpretation of MS/MS spectra of these peptides allows to precisely localize the phosphorylation sites on T35 for DAPKwt and T259 for DAPKdel (figures 3a and 3b, respectively). As there are no endogenous Ser/Thr-protein kinases

known in *E. coli* and as the sequences around the mentioned sites show high identity to the ones of DAPK's known substrates, we suggest autophosphorylation as the underlying mechanism of the observed phosphorylations of the protein.

The hypothetical autophosphorylation ability of DAPKwt catalytic core was investigated by incubating the protein overnight at 37°C in presence of ATP and Mg<sup>2+</sup>. Subsequent mass measurement under denaturing conditions revealed a significant increase of the phosphorylated form, which relative intensity raises from 18% to 41% (figures 4a and 4b). Moreover, DAPKwt analyses in non-denaturing conditions clearly show that a higher phosphorylation rate is accompanied by a decrease of the dimer relative abundance (figures 4c and 4d). These observations corroborate the autophosphorylation hypothesis of DAPK's catalytic core and suggest an inverse relationship between dimerization and autophosphorylation. To our knowledge, both phosphorylation sites are described for the first time and it will take some further investigations to put their role in the regulation of DAPK in its proper context.

#### *High Throughput Screening (HTS) of fluorescent compound library*

We have screened a chemical library of 1388 compounds that are tagged with either the ortho or the para isomer of the fluorophore lissamine [xx] in order to identify molecular probes for further investigations of DAPK's regulatory system. The library was screened for molecules that bind to DAPK's catalytic domain by measurements of the fluorescent polarization. After bulk analysis, the 14 compounds (1%) showing the highest binding affinity were selected as hits which could be grouped in two distinct clusters of molecules. Within the same cluster, the structures mainly differed in the length of the linker between the lissamine and the compound. These observations provide evidence that some common features of the identified molecules, such as their bi-cationic character as well as the interjacent aromatic residue might be responsible for their binding abilities to DAPK. Another outcome of the screening assay was the fact that the same compound did not show comparable binding affinity to the protein when either linked to the ortho or the para isomer of lissamine. Detailed investigations during the secondary screening confirmed this observation and the modeled K<sub>d</sub>-values of the ortho-compounds proved to be considerably higher than the ones of the corresponding para-analogues (data not shown).

In the validation phase, DAPK's purified catalytic core was titrated by the identified hit-molecules and the resulting fluorescence polarization was recorded at each concentration. The results allowed to fit binding curves and to determine K<sub>d</sub>-values for each of the hits. Based on the determined affinity toward DAPK and on the compound's convenience to be re-synthesized, the molecule CHPO 187-3-H11-para (further on called probe) was selected for all subsequent investigations (figure 5A).



### *Interaction of monomeric and dimeric DAPK with the selected probe*

In order to assess the fluorescent probe's binding affinity to DAPK, above-described secondary screening assay was repeated under buffer conditions favoring either dimeric or monomeric DAPKwt. Titration experiments were thus performed for both DAPKwt and DAPKdel at 5, 100 and 250 mM NH<sub>4</sub>Ac. Monitoring fluorescence polarization evidences that the binding affinity of the probe for DAPKwt increases with higher NH<sub>4</sub>Ac concentrations (figure 5B). Indeed, K<sub>d</sub>-values calculated from these binding curves are  $16.5 \pm 1.2 \mu\text{M}$  and  $1.5 \pm 0.2 \mu\text{M}$  under buffer conditions favoring the dimer (5 mM NH<sub>4</sub>Ac) and the monomer (250 mM NH<sub>4</sub>Ac), respectively. In contrast, DAPKdel which was shown to be exclusively monomeric regardless of NH<sub>4</sub>Ac concentration, do not display such an ionic strength dependent binding affinity (Figure 5B) Altogether these results clearly show a modified interaction between the probe and the protein depending on its oligomerization state. Therefore, the fluorescent probe appears to be a convenient tool to explore the dimerization state of the DAPK catalytic core.

### *Localization of the probe's binding site*

In order to address the question of the probe's binding site, competitive titrations were carried out to check whether it shares the same binding pocket as one of the known substrates or products of the kinase, i.e., ATP, ADP or the peptide substrate [vi].

A sequential multilayered assay corresponding to a total of 960 different conditions per 96-well plate was thus developed. For the assay's first layer, 0.1  $\mu\text{M}$  of the probe was incubated with a geometrical concentration range (0 to 100  $\mu\text{M}$ ) of DAPK and the fluorescent polarization of lissamine was recorded. This titration experiment was then repeated in presence of increasing amounts of ATP (0.4, 0.8, 1.6, 3.2, 6.3, 12.5, 25, 50 and 1000  $\mu\text{M}$ ). This multilayered titration matrix was subsequently applied to the substrate peptide and ADP replacing ATP.

Figures 6A and 6B show the graphical representation of the procedure's output for the titrations of ATP and substrate peptide, respectively. Results clearly show that neither ATP nor the substrate peptide is able to eject the probe away from its binding site, as this would lead to a flattening of the titration curves with increasing concentrations of ATP or substrate peptide. Identical results were obtained with ADP replacing ATP (data not shown). These observations led us to the conclusion that the probe binds to a third site, different from the ones occupied by the ATP/ADP and the substrate peptide. Moreover, figure 6A shows that increasing fluorescence polarization levels are obtained in presence of higher ATP concentrations, which indicates an improved probe's binding affinity in presence of ATP. Such an observation, together with the one of the site's spatial independency, is a strong hint for an involved allosteric mechanism.

In the attempt to represent the described data in a clear overview and to quantify the allosteric impact, we used the measured data to model the developing of the probe's affinity for the protein as it was described by Roehrl et al [i]. We fitted binding curves through data points that represent the same

effector to protein ratios. As an index of the probe's affinity we assigned a  $K_d$ -values to each of the thusly obtained curves. Figure 6C represents the probe's calculated  $K_d$ -values in function of the stoichiometrical ratio of either ATP or ADP to DAPK. The graphic clearly shows the similarity of the impacts of ATP and ADP, as corresponding  $K_d$ -values of both titrations are comparable. This representation also permits to visualize the magnitude of the induced effect. In the presence of neither ATP nor ADP the probe's  $K_d$ -values was determined to  $30.9 \pm 2.8 \mu\text{M}$ . A maximal effect was achieved at ATP/ADP to DAPK ratios higher than 2, which led to  $K_d$ -value of the probe of  $6.40 \pm 0.54 \mu\text{M}$ . Hence, both ATP and ADP possess the ability to increase the probe's affinity roughly five times.

#### *ATP/ADP induced monomerization of DAPK's catalytic core*

As ATP/ADP effect on probe's affinity for DAPK<sub>wt</sub> is similar to ionic strength induced DAPK<sub>wt</sub> monomerization, the question of whether ATP and ADP directly act on DAPK monomer / dimer equilibrium was raised. Considering such an influence of ATP/ADP, one would expect an increased affinity of ATP and ADP for the monomerized compared to the dimerized form of the protein. In order to test our hypothesis  $K_d$ -values were determined for a fluorophor labeled ATP ( $\gamma$ -[6-Aminohexyl]-ATP-Atto495) under the buffer conditions promoting the dimeric (5 mM  $\text{NH}_4\text{Ac}$ ) or the monomeric form (250 mM  $\text{NH}_4\text{Ac}$ ) of DAPK<sub>wt</sub>. The results confirm our assumption, as the  $K_d$ -value for the dimer was determined to be  $3.98 \pm 0.22 \mu\text{M}$ , while the one for the monomer decreases to  $0.290 \pm 0.011 \mu\text{M}$  (Figure 7A)

To confirm this model of ATP/ADP induced monomerization, three further types of experiments were performed. In a first approach, noncovalent ESI-MS experiments were carried out with DAPK<sub>wt</sub> in presence of increasing amounts of ADP-Mg. Figure 7B shows that incubation of the protein with five to twenty molar equivalents of ADP-Mg leads to the detection of 1:1, 2:1 and 2:2 DAPK:ADP-Mg complexes ( $34199 \pm 1 \text{ Da}$ ,  $67964 \pm 4 \text{ Da}$ ,  $68411 \pm 6 \text{ Da}$ , respectively). Interestingly, ADP binding is accompanied by a shift of DAPK's oligomerization equilibrium towards the monomeric form, which corroborates our model of ATP/ADP induced monomerization (figure 7B).

Then, we investigated whether the deletion of the basic loop, which impedes dimerization of the catalytic core, would modify the probe's binding behavior upon the addition of ATP or ADP. Therefore, multilayered titration matrices were performed with DAPK<sub>del</sub>. Strikingly, we observed that the partial absence of the basic loop leads to a converse effect of both ATP and ADP compared to the one observed for DAPK<sub>wt</sub>. At ATP and ADP to DAPK<sub>del</sub> ratios higher than two, the  $K_d$ -values of the probe get maximal for both effectors and were calculated to approximately  $30 \mu\text{M}$  (figure 7C).

Finally, ATP and ADP titrations were repeated under experimental conditions favoring either the presence of the monomer or the dimer of DAPK<sub>wt</sub>. Figure 7D shows the probe's determined  $K_d$ -values to either the monomer or the dimer of DAPK<sub>wt</sub> in function of the ratio of ATP or ADP to the protein. As under the chosen conditions ( $\text{NH}_4\text{Ac}$ , 5mM) the dimer is the dominant form of the protein

present, but the monomeric form still co-exists at a significant level, the determined  $K_d$ -value at low ATP/ADP to DAPK ratios, is somewhat lower than observed in Figure 6C. The developing of the  $K_d$ -values to the monomer (Figure 7D) resembles a lot the ones observed for the DAPKdel (Figure 6C), only that the level of affinity is decreased for the mutated protein.

The obtained results (Figure 7) clearly consolidate our model of a constant equilibrium of monomers and homodimers of DAPK's catalytic domain, which can be shifted towards the monomer by the presence of either ATP or ADP.

#### *Significance of DAPK's regulatory switch between monomers and dimers*

The previously reported crystallographic structure in the cubic form suggests that both basic loops and peptide binding regions of the two subunits building the homodimer face each other [v]. Our results that highlight the importance of the basic loop in the process of dimer formation and the structural observation that the loop is neighboring the active site strengthen these previous findings. Therefore, the simple fact that the access to the catalytic cleft of homodimers is sterically limited for substrate proteins, let us suspect the monomer as the catalytically more active form. This hypothesis gets supported by the fact that ATP, which is a reactant of the phosphotransfer reaction shifts the equilibrium towards the monomer. In this spirit we also re-analyzed the results of DAPK's multilayered substrate peptide titration matrix in order to quantify the impact of substrate peptide on the dimerization. Analogous to the described procedure of the ATP/ADP-titrations we modeled  $K_d$ -values for each of the concentrations of the peptide substrate at which a titration with DAPKwt was performed. In the absence of the peptide, the probes affinity was modeled to  $26 \pm 1.2 \mu\text{M}$ , which corroborates the initial affinity measurement of the ATP/ADP titrations. The  $K_d$ -values decrease with increasing peptide to DAPK ratios down to  $16.2 \pm 1.9 \mu\text{M}$  for a ratio of 10. From these data it becomes obvious that the impact of ATP/ADP on the equilibrium of monomers and dimers is significantly stronger than the one of the substrate peptide. Nevertheless, these results imply a stronger affinity of the substrate peptide for the monomer than for the dimer. This might be a further hint for a higher enzymatic activity of the monomer than of the dimer. The decreased magnitude of the effect compared to the one induced by the ATP might be explained by the fact that the substrate peptide is in competition with the regulatory pseudo-substrate, which is part of the enzyme's inhibitory mechanism. However, additional investigations will be required to unravel the potential participation of this regulatory domain in the here proposed mechanism of "activating monomerization".

## **Conclusions**

In our study we demonstrated the utility of a chemical biology approach for the identification of fluorescently marked ligands of proteins that bind to not priori defined binding sites. As a consequence of the non-competition of these compounds with the known endogenous partners and their fluorescent traceability, they are perfectly suited as small probes to sense the state of molecular systems.

We have identified such a molecular tool that is sensitive to the oligomerization state of the catalytic core of the human Death-Associated Protein Kinase, which allowed us to clearly evidence the presence of homodimers of this protein domain. DAPK is seen as an important actor in the integration of various signal transduction pathways, which ultimately decides about either cellular survival or induced cell death.

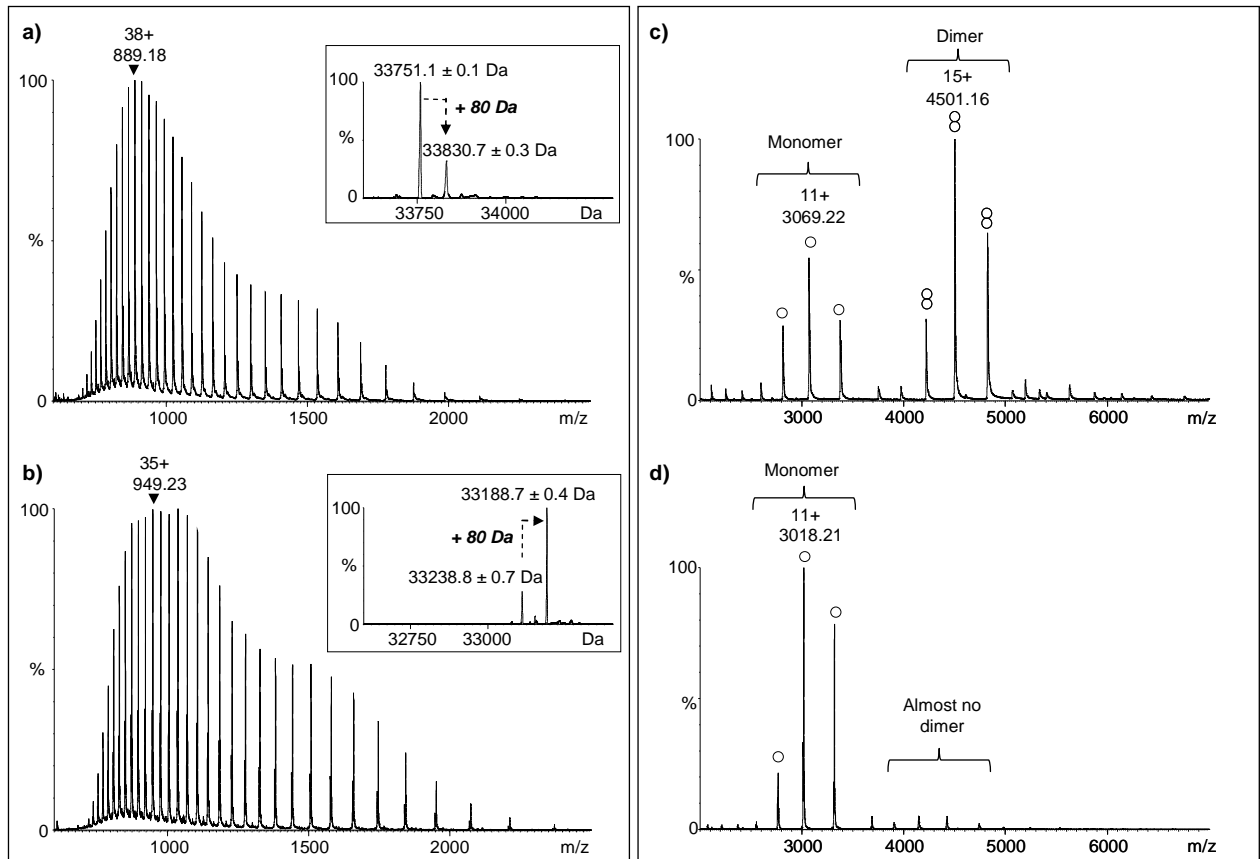
Fraser et al [xiv] showed the induction of an allosteric effect on DAPK's catalytic activity by peptide fragments derived from the protein p21. We performed a competitive assay between the fluorescent probe and the peptide fragments number 8 and 14, which were shown to be allosterical effectors, in order to investigate whether the peptides and our probe share a common site of interaction on DAPK. The assay was set up as the ones for ATP and the substrate peptide. Although the ratio between these two peptides to the probe was increased up to 500, there was no decrease of the fluorescent signal observable. This leads us to the conclusion that the identified third binding site is not the same as the one of p21 and the question of the biological significance of the probe's binding pocket remains yet unanswered.

Our here presented results, together with structural considerations lead us to believe that the homodimerization of DAPK's core domain is an important regulatory mechanism for this enzyme's catalytic activity. We propose a model of induced monomerization, which assigns the sterically hindered active cleft of dimers and accessible one of monomers to cellular survival and induced cell death respectively.

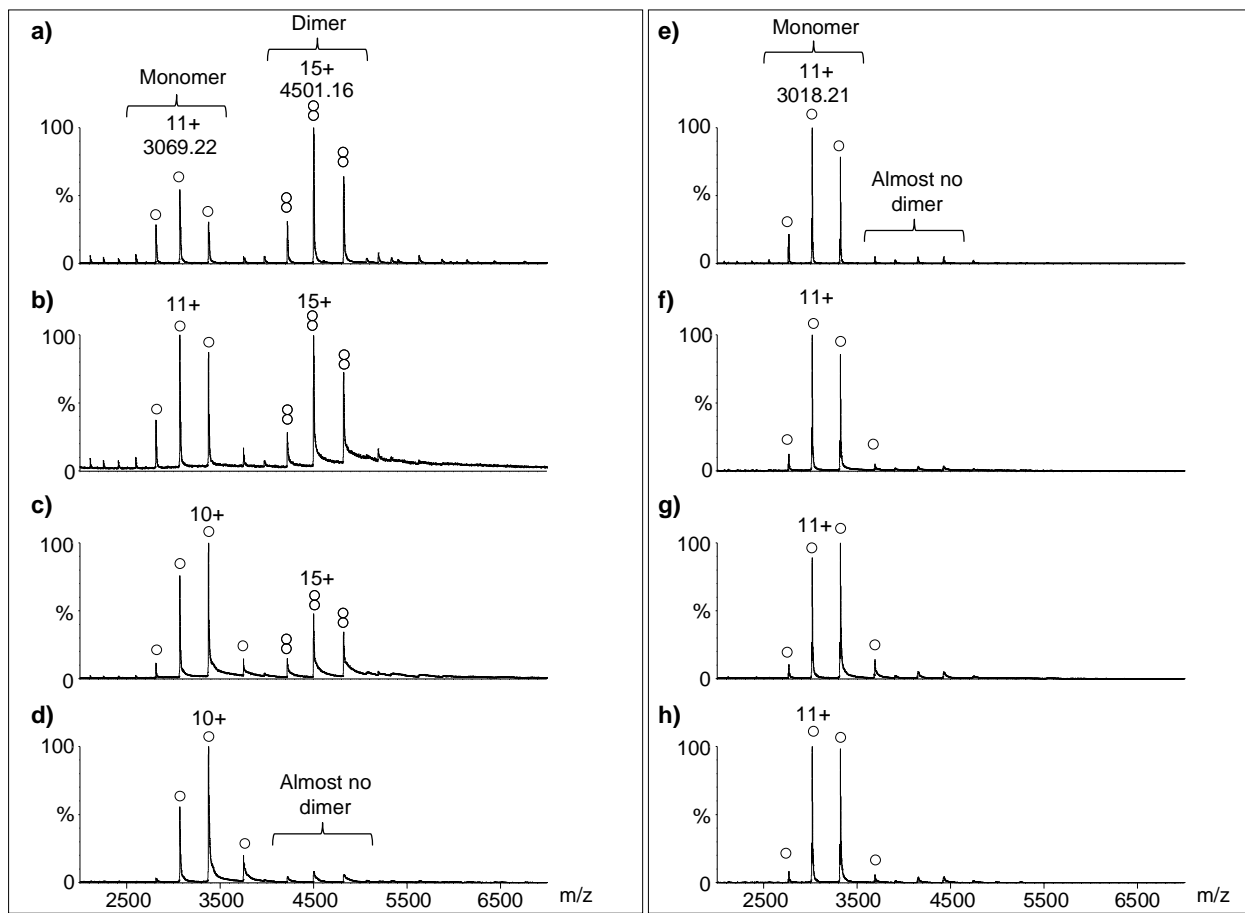
We believe, that future investigations should aim at bringing into context the here described regulatory mechanism with the ones that were earlier described for this kinase, namely autophosphorylation, pseudosubstrate interference and the activation through CaM.

## **Acknowledgement:**

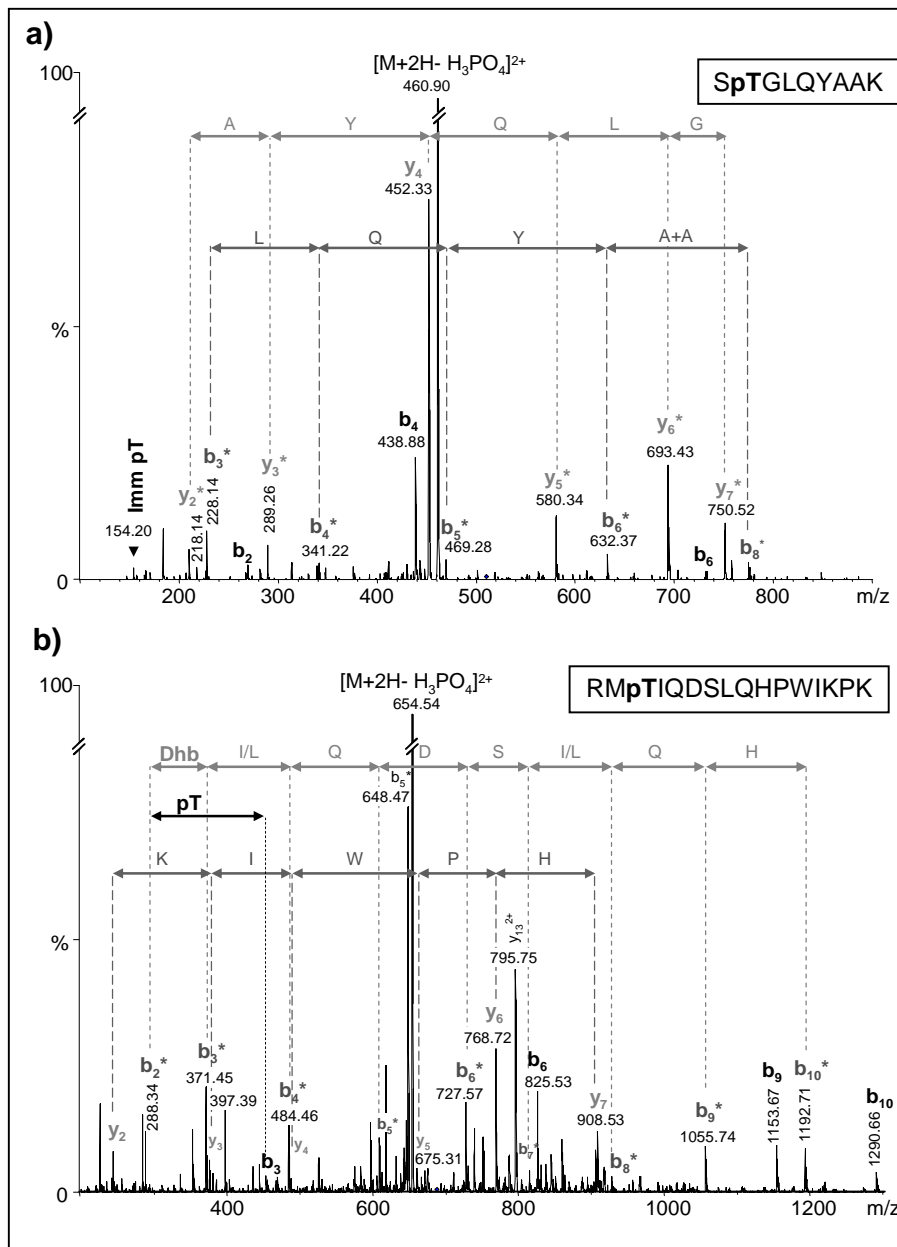
This work was funded by the CNRS and the University of Strasbourg (UdS). CA was supported by a studentship from the French Ministère de la Recherche. LF gratefully acknowledges the region Alsace for financial support.



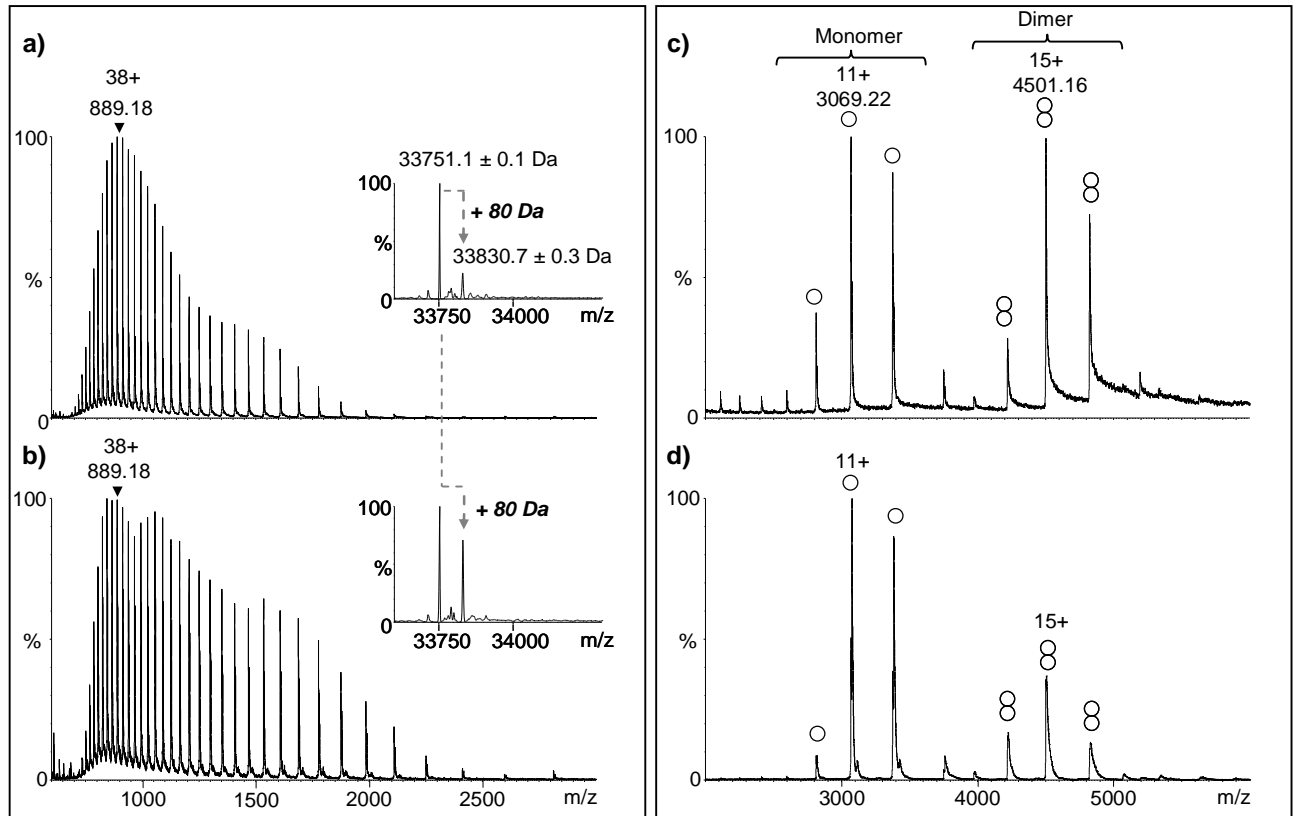
**Fig. 1:** Determination of DAPK oligomerization state. (a) DAPKwt and (b) DAPKdel were diluted to 2  $\mu$ M in a 1:1 (v/v) water : acetonitrile mixture acidified with 1% (v/v) formic acid analysis. (c) DAPKwt and (d) DAPKdel were diluted to 5  $\mu$ M in 5 mM  $\text{NH}_4\text{Ac}$  buffer pH 8.8. (c) and (d) are related to monomeric and dimeric DAPK, respectively. Inserts correspond to deconvoluted mass spectra obtained using the MaxEnt 1 algorithm from MassLynx 4.1 (Waters, Manchester, UK)



**Fig. 2:** Noncovalent mass spectrometry analysis of (a-d) DAPKwt and (e-h) DAPKdel. Proteins were diluted to 5  $\mu$ M in  $\text{NH}_4\text{Ac}$  buffer at (a, e) 5 mM, (b, f) 50 mM, (c, g) 120 mM and (d, h) 250 mM. (○) and (⊗) are related to monomeric and dimeric DAPK, respectively.

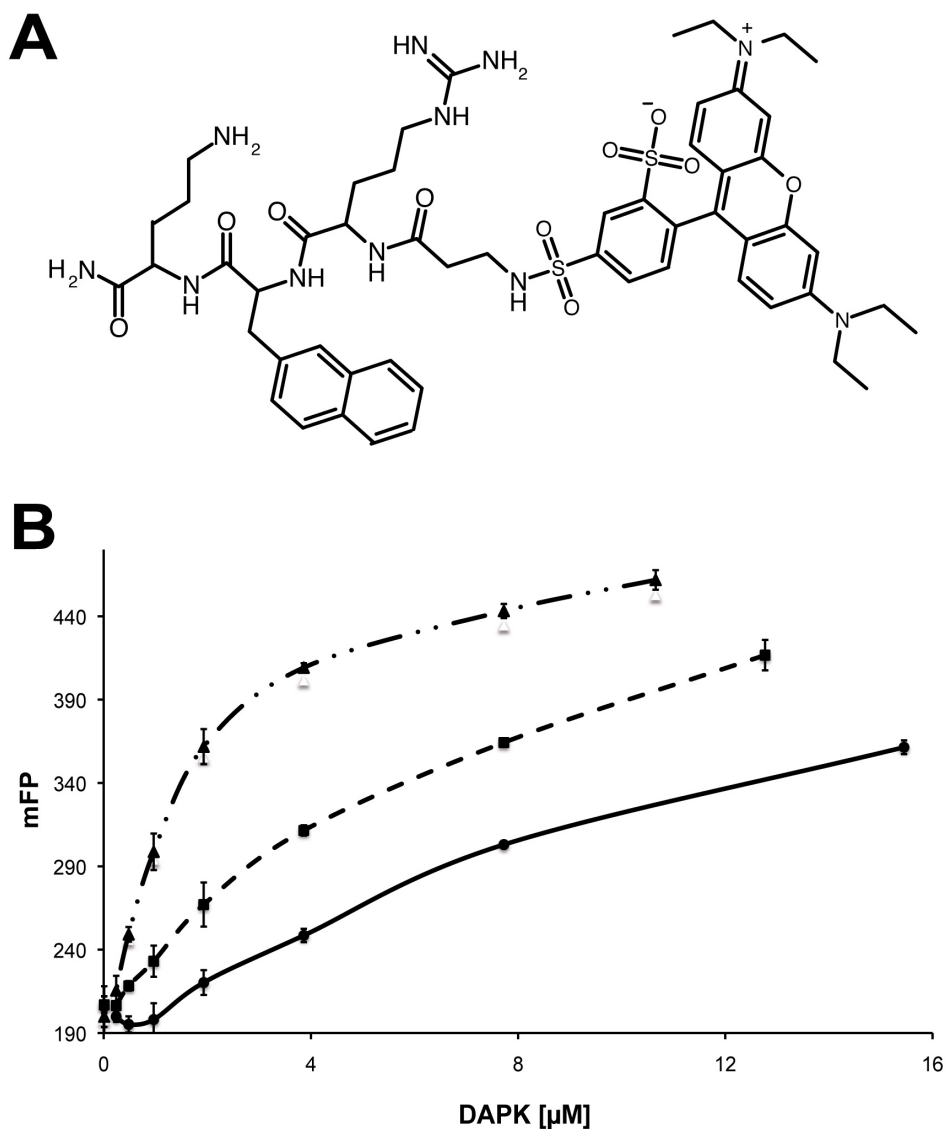


**Fig. 3:** Localization of the phosphorylation sites of (a) DAPKwt and (b) DAPKdel. MS/MS spectra correspond to the tryptic doubly charged phosphopeptides a) 34-SpTGLQYAAK-42 (509.9  $m/z$ ) from DAPKwt and b) 257-RMpTIQDSLQHPWIKPK-272 (688.2  $m/z$ ) from DAPKdel. The most intense peak of each mass spectrum corresponds to the neutral loss of phosphoric acid from the precursor ion ( $-49$   $m/z$ ). Phosphorylation sites are assigned by the detection of (a) phosphothreonine immonium ion (Imm pT) and (b) dehydrobutyric acid (Dhb) and phosphothreonine (pT). \* are related to diagnostic fragment ions after loss of phosphoric acid.

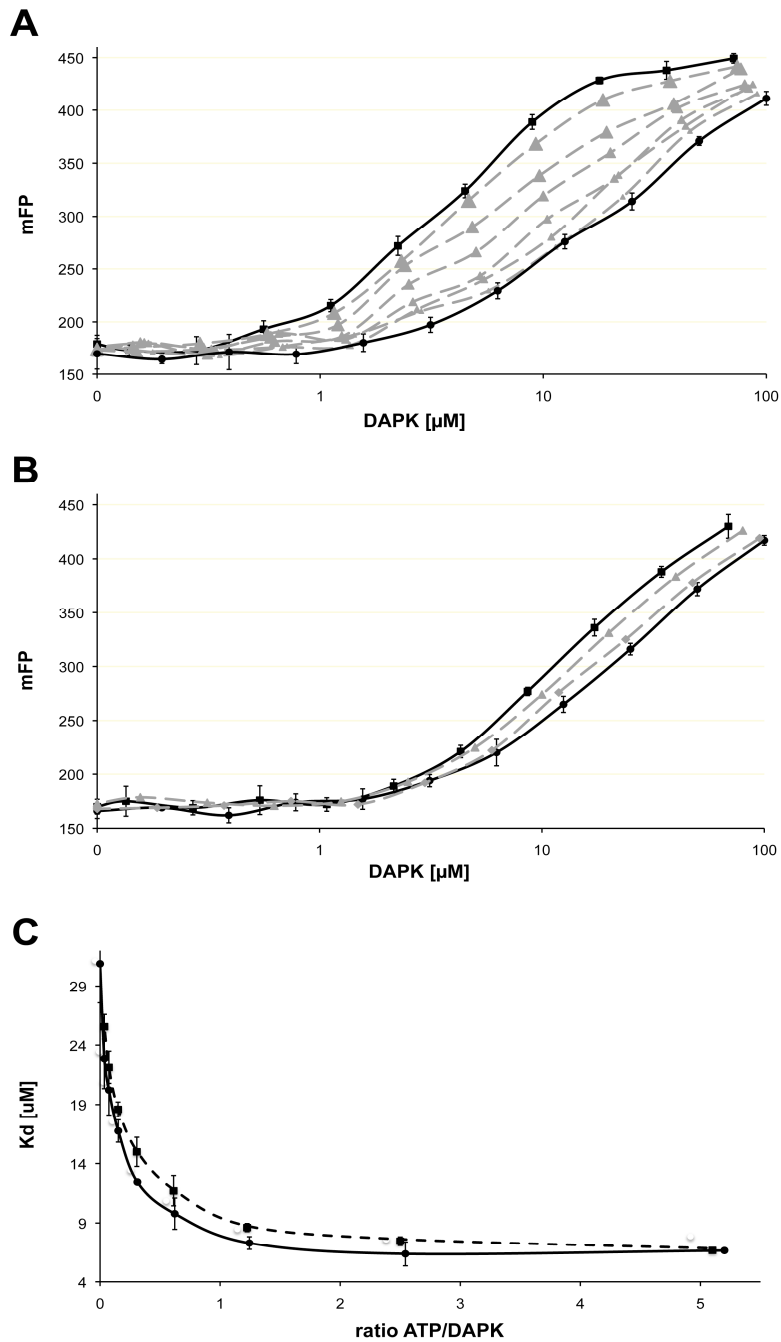


**Fig. 4:** DAPKwt autophosphorylation experiment. The protein was incubated overnight at 37°C in presence of a 333-fold molar excess of ATP-Mg. Analyses were performed (a, c) before and (b, d) after ATP-Mg treatment either (a, b) in denaturing conditions (2 μM DAPKwt) or (c, d) in 50 mM ammonium acetate buffer at pH 8.8 (5 μM DAPK). Inserts correspond to deconvoluted mass spectra obtained using the MaxEnt 1 algorithm from MassLynx 4.1 (Waters, Manchester, UK). (○) and (⊗) are related to monomeric and dimeric DAPK, respectively.

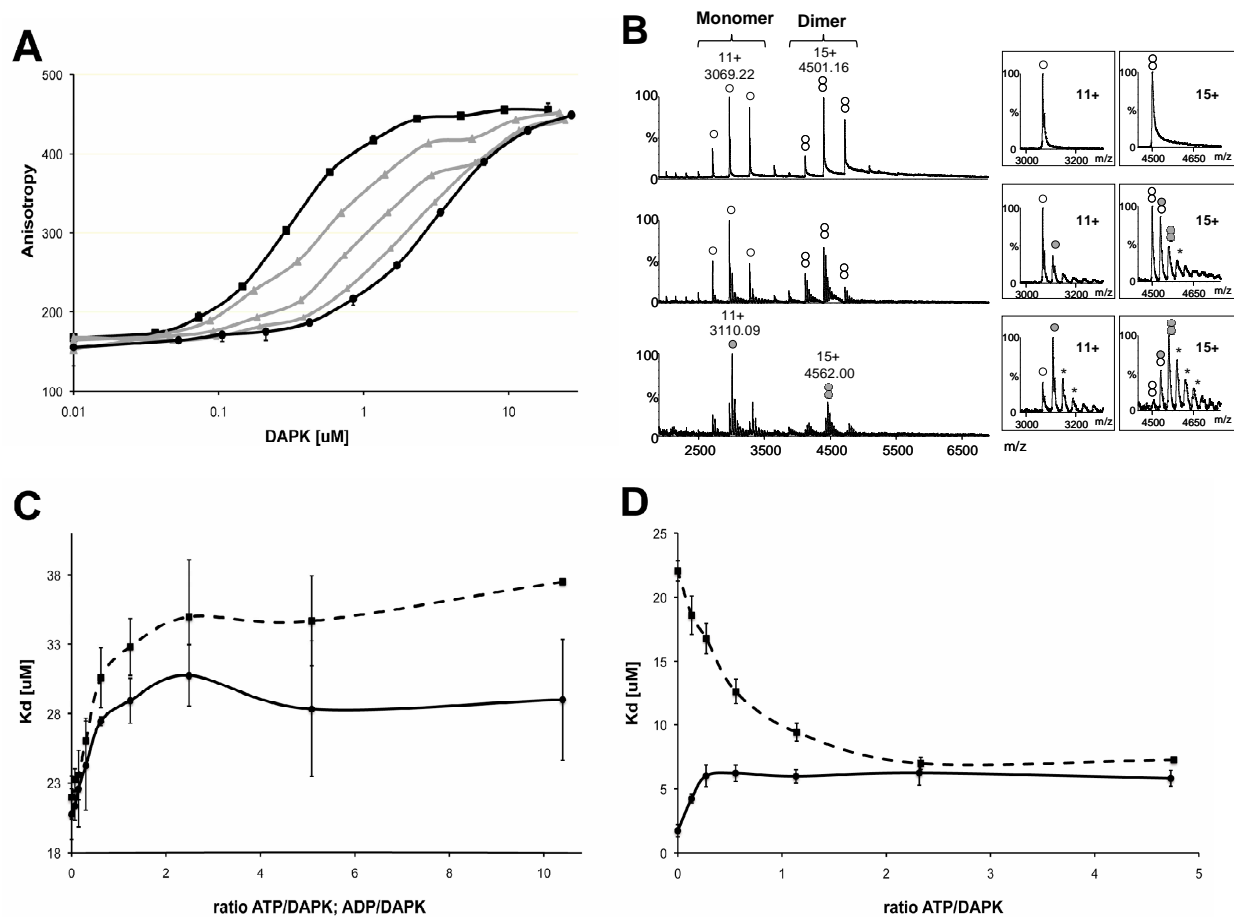




**Fig. 5:** Fluorescent molecular probe and dimerization of DAPK's catalytic core A: Chemical structure of CHPO 187-3 H11-para, which was selected as best hit in the primary and secondary screening assays. B: Binding of CHPO 187-3 H11-para (0.1 $\mu$ M) to DAPKwt at different concentration of NH<sub>4</sub>Ac (● 5mM, ■ 100mM and ▲ 250mM). The recorded fluorescent polarization (mFP) is shown in dependence of the protein concentration. Error bars indicate the standard deviation of four independent titrations. Higher NH<sub>4</sub>Ac concentrations that lead to the predominance of the monomeric form of DAPK result in an increased affinity for the fluorescent probe CHPO 187-3 H11-para.



**Fig. 6:** Titration curves showing mFP in function of the concentration of DAPKwt. Error bars indicating the standard deviation were calculated from four independent titrations. Error bars are only shown for the most out-layer curves in order to keep the representation clear A: ATP-titration at the following total concentrations of ATP: ● 0 $\mu$ M; ▲ to ▲ 0.4, 0.8, 1.6, 3.2, 6.3, 12.5 $\mu$ M ■ 25 $\mu$ M. The graphs for an ATP concentration of 50 $\mu$ M and 1000 $\mu$ M are not shown, as they are congruent with the one at 25 $\mu$ M. B: Peptide substrate-titration. Analogous to the ATP-titration for peptide concentrations of: ● 0 $\mu$ M; ◆ 0.8 $\mu$ M ▲ 6.3 $\mu$ M; ■ 50 $\mu$ M. Only a representative number of titration curves are shown to keep the graphs well distinguishable from each other. C: Fitted  $K_d$ -values of the fluorescent probe in function of the stoichiometrical ratio of ATP/DAPK and ADP/DAPK respectively. For both ATP (—●—) and ADP (- - ■ - -) the  $K_d$ -value is approaching a constant  $K_d$ -value when the ratio of either of which to DAPK increases.



**Fig. 7:** ATP and ADP induced monomerization of DAPK's catalytic domain. **A:** Binding of the fluorescently tagged ATP (Atto-495-ATP) to the catalytic domain of DAPK. The binding curve of ATP shifts toward the left when the monomeric form is induced by higher  $\text{NH}_4\text{Ac}$  concentrations. One binding curve corresponds to a constant concentration of  $\text{NH}_4\text{Ac}$ : ● 0mM; ▲ 25, 50 and 100mM, ■ 250mM. **B:** Noncovalent mass spectrometry analysis of DAPKwt at 5  $\mu\text{M}$  in 50 mM  $\text{NH}_4\text{Ac}$  buffer at pH 8.8 either alone (top) or in presence of 25  $\mu\text{M}$  ADP-Mg (middle) and 100  $\mu\text{M}$  ADP-Mg (bottom). Inserts correspond to enlargements of 11+ and 15+ charge state of monomeric and dimeric DAPKwt, respectively. (○), (●), (⊗), (⊗) and (⊗) are related to 1:0, 1:1, 2:0, 2:1 and 2:2 DAPKwt:ADP-Mg complexes, respectively. \* represent non-specific ADP-Mg adducts. **C:** Fitted  $K_d$ -values of the fluorescent probe in function of the stoichiometrical ratio of ATP/DAPKdel (—●—) and ADP/DAPKdel (- - ■ - -). The developing of the  $K_d$ -values with increasing ratios of ATP/ADP to DAPK is contrary to the one observed in Fig. 6c. **D:** DAPKwt is assayed under conditions favoring monomers (—●—) and dimers (- - ■ - -). The probe's binding compartment to the dimer is comparable to the one observed in the physiological buffer (Fig. 6c), while the one to the monomer shows high similarity to the observations with DAPKdel (Fig. 7c).

## References:

- i. L P Deiss, E Feinstein, H Berissi, O Cohen, A Kimchi. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes Dev* 1995 vol. 9 (1) pp. 15-30
- ii. Galit Shohat, Gidi Shani, Miriam Eisenstein, Adi Kimchi, The DAP-kinase family of proteins: study of a novel group of calcium-regulated death-promoting kinases. *Biochim Biophys Acta* 2002 vol. 1600 (1-2) pp. 45-50
- iii. O Cohen, E Feinstein, A Kimchi, DAP-kinase is a Ca<sup>2+</sup>/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity. *EMBO J* 1997 vol. 16 (5) pp. 998-1008
- iv. G Shohat, T Spivak-Kroizman, O Cohen, S Bialik, G Shani, H Berrisi, M Eisenstein, A Kimchi, The pro-apoptotic function of death-associated protein kinase is controlled by a unique inhibitory autophosphorylation-based mechanism. *J Biol Chem* 2001 vol. 276 (50) pp. 47460-7
- v. V Tereshko, M Teplova, J Brunzelle, D M Watterson, M Egli Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression *Nat Struct Biol* 2001 vol. 8 (10) pp. 899-907
- vi. Andrew M Schumacher, James P Schavocky, Anastasia V Velentza, Salida Mirzoeva, D Martin Watterson, A calmodulin-regulated protein kinase linked to neuron survival is a substrate for the calmodulin-regulated death-associated protein kinase. *Biochemistry* 2004 vol. 43 (25) pp. 8116-24
- vii. Anastasia V Velentza, Andrew M Schumacher, D Martin Watterson, Structure, activity, regulation, and inhibitor discovery for a protein kinase associated with apoptosis and neuronal death. *Pharmacol Ther* 2002 vol. 93 (2-3) pp. 217-24
- viii. Anastasia V Velentza, Mark S Wainwright, Magdalena Zasadzki, Salida Mirzoeva, Andrew M Schumacher, Jacques Haiech, Pamela J Focia, Martin Egli, D Martin Watterson, An aminopyridazine-based inhibitor of a pro-apoptotic protein kinase attenuates hypoxia-ischemia induced acute brain injury. *Bioorg Med Chem Lett* 2003 vol. 13 (20) pp. 3465-70
- ix. Y Jin, E K Blue, S Dixon, L Hou, R B Wysolmerski, P J Gallagher, Identification of a new form of death-associated protein kinase that promotes cell survival. *J Biol Chem* 2001 vol. 276 (43) pp. 39667-78
- x. Shani Bialik, Adi Kimchi, The death-associated protein kinases: structure, function, and beyond. *Annu Rev Biochem* 2006 vol. 75 pp. 189-210
- xi. P G Clarke, Developmental cell death: morphological diversity and multiple mechanisms. *Anat Embryol* 1990 vol. 181 (3) pp. 195-213
- xii. Devrim Gozuacik, Adi Kimchi, DAPk protein family and cancer. *Autophagy* 2006 vol. 2 (2) pp. 74-9

- xiii. Shani Bialik, Hanna Berissi, Adi Kimchi, A high throughput proteomics screen identifies novel substrates of death-associated protein kinase. *Mol Cell Proteomics* 2008 vol. 7 (6) pp. 1089-98
- xiv. Jennifer A Fraser, Ted R Hupp Chemical genetics approach to identify peptide ligands that selectively stimulate DAPK-1 kinase activity. *Biochemistry* 2007 vol. 46 (10) pp. 2655-73
- xv. Ben Harrison, Michaela Kraus, Lindsay Burch, Craig Stevens, Ashley Craig, Phillip Gordon-Weeks, Ted R Hupp, DAPK-1 binding to a linear peptide motif in MAP1B stimulates autophagy and membrane blebbing. *J Biol Chem* 2008 vol. 283 (15) pp. 9999-10014
- xvi. Craig Stevens, Ted R Hupp, Novel insights into DAPK autophagic signalling using peptide aptamer combinatorial protein-interaction screens. *Autophagy* 2008 vol. 4 (4) pp. 531-3
- xvii. Gidi Shani, Lea Marash, Devrim Gozuacik, Shani Bialik, Lior Teitelbaum, Galit Shohat, Adi Kimchi, Death-associated protein kinase phosphorylates ZIP kinase, forming a unique kinase hierarchy to activate its cell death functions. *Mol Cell Biol* 2004 vol. 24 (19) pp. 8611-26
- xviii. T Kawai, M Matsumoto, K Takeda, H Sanjo, S Akira, ZIP kinase, a novel serine/threonine kinase which mediates apoptosis. *Mol Cell Biol* 1998 vol. 18 (3) pp. 1642-51
- xix. A V Velentza, A M Schumacher, C Weiss, M Egli, D M Watterson (2001) *J. Biol. Chem.* 276, 38956-65
- xx. Dagher R, Pigault C, Bonnet D, Boeglin D, Pourbaix C, Kilhoffer MC, Villa P, Wermuth CG, Hibert M, Haiech J. Use of a fluorescent polarization based high throughput assay to identify new Calmodulin ligands. *Biochimica et Biophysica.* 2006; 1763; 1250-55.
- xxi. E B Babiyuchuk, V S Babiyuchuk, A Sobieszek, Modulation of smooth muscle myosin light chain kinase activity by Ca<sup>2+</sup>/calmodulin-dependent, oligomeric-type modifications. *Biochemistry* 1995 vol. 34 (19) pp. 6366-72
- xxii. Roehrl et al. A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein- protein interactions by fluorescence polarization. *Biochemistry* (2004) vol. 43 (51) pp. 16056-66



## **Chapitre III - Etude des phosphorylations du complexe CAK**

Cette étude a été réalisée en collaboration avec Wassim Abdulrahman et Arnaud Poterzsmann du "Département de Biologie et Génomique Structurales" de l'IGBMC, Strasbourg.

### **1. Contexte Biologique**

Le complexe CAK est composé de trois protéines : CDK7 (cyclin-dépendent kinase), Cycline H associée à MAT1 ("*ménage à trois*"). Ce complexe est présent dans la cellule dans deux états différents : soit libre, soit associé au complexe "core" du facteur de transcription TFIIH. Ce complexe est impliqué dans trois des plus importantes fonctions au sein de la cellule : la réparation de l'ADN, le contrôle du cycle cellulaire et la transcription [Zurita & Merino, 2003].

L'activité du CAK est contrôlée par de nombreux mécanismes, dont la liaison avec MAT1 et la phosphorylation du CDK7 au niveau de sa "T-loop" [Larochelle et al., 2001]. Deux résidus sont décrits comme étant phosphorylés S164 et T170. Les Cyclines sont également des protéines phosphorylées : la Cycline H peut être phosphorylée par l'action de CDK8/CyclinC en S5 et en S304 [Akoulitchev et al., 2000] ou en T315 par l'action de CK2 [Akoulitchev et al., 2000; Schneider et al., 2002]. Une étude à haut débit d'identification de phosphoprotéines a mis en évidence récemment des résidus phosphorylés sur la Cycline H (S132, S307, S322) [Daub et al., 2008; Dephoure et al., 2008] et sur MAT1 (T51, S205 et S279) [Daub et al., 2008], le rôle biologique de ces phosphorylations n'étant pas encore identifié.

Seules les protéines CDK7 mono-phosphorylée en T170 [Lolli et al., 2004] et Cycline H non-phosphorylée [Kim et al., 1996] ont été cristallisées séparément. Les interactions entre les trois protéines n'étant pas bien définies, il importe donc de les identifier. Pour cela, la cristallisation du complexe peut apporter des réponses. Celle-ci nécessite des protéines homogènes, i.e. complètement phosphorylées ou complètement non-phosphorylées. L'objectif de cette étude est de définir l'état de phosphorylation (taux de phosphorylation et sites de phosphorylations) du complexe CAK, afin de guider la production du complexe CAK totalement phosphorylé.

### **2. Mesure de masse et déphosphorylation des protéines entières**

Le complexe CAK a été produit dans la lignée cellulaire d'insecte Sf9 (lignée cellulaire issue du légionnaire d'automne "*Spodoptera frugiperda*", ravageur du maïs, coton et riz). L'analyse des protéines a été réalisée à l'aide d'un couplage LC-MS (MicroLC (Agilent Technologies) – MicroTOF

(Bruker Daltonics)). Les protéines ont été séparées sur une colonne Vydac Grace C4 (300Å – 0.5x150mm- 5µm) (Figure 24-A), et analysées en ESI-MS (Figure 24-B et C).

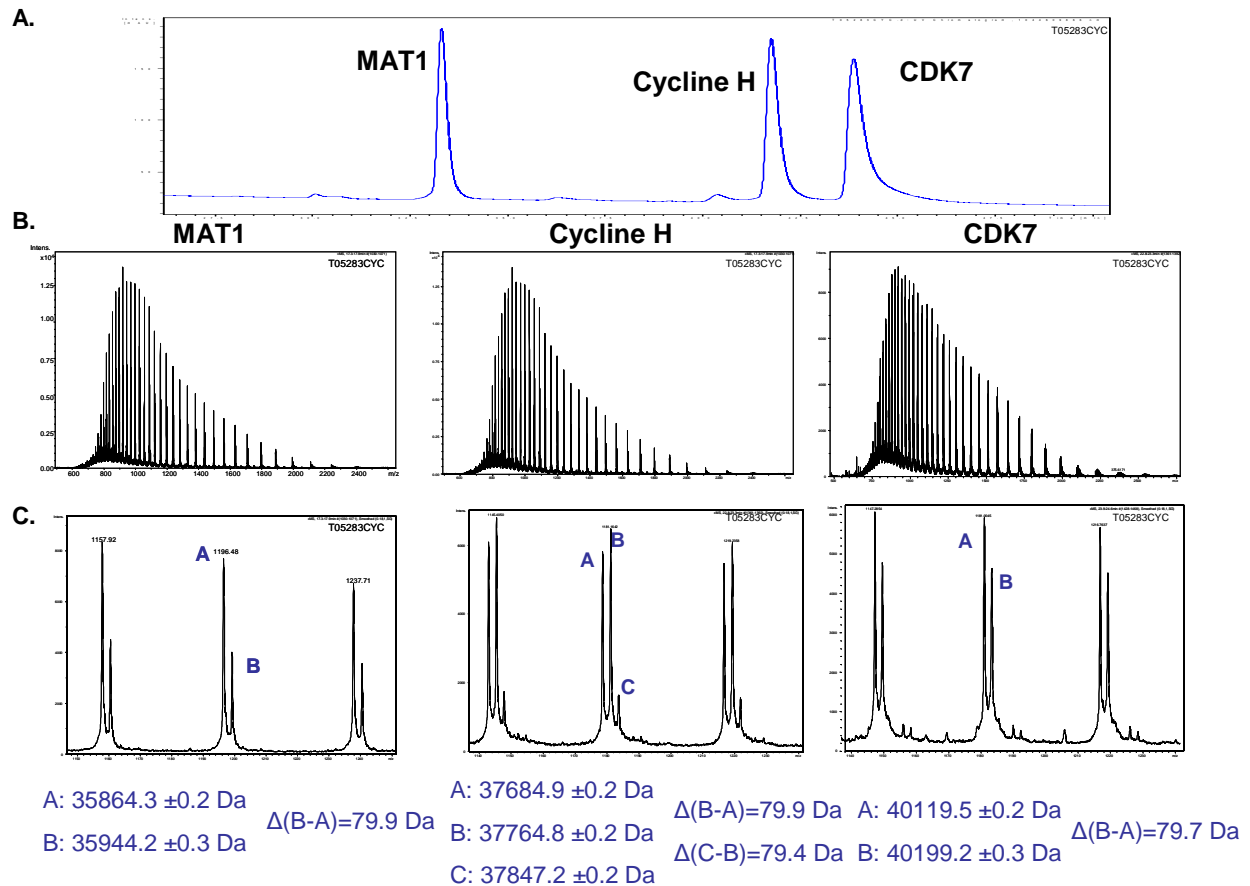


Figure 24 : Mesure des masses du complexe CAK par LC-MS. A) Chromatogramme UV de la séparation des 3 protéines. B) Spectres de masse ESI-MS de chacune des protéines. C) Détails des spectres de masse ESI-MS : zoom sur les états de charges 30+ à 32+.

Le spectre de masse ESI-MS de MAT1 révèle deux populations, ayant une différence de masse de 79.9 Da, qui pourrait correspondre à une phosphorylation. De la même façon, on trouve trois espèces sur le spectre de masse correspondant à la Cycline H et deux séries d'ions pour CDK7 ayant des différences de masse du même ordre (Figure 24). Afin de vérifier que les populations observées à + 80 Da sont bien des espèces phosphorylées, le complexe a été déphosphorylé avec la phosphatase alcaline. Afin de vérifier l'activité de la phosphatase alcaline, un contrôle a été réalisé sur 1 nmol de β-caséine (possédant 5 phosphorylations) incubée avec 5 unités de phosphatase alcaline. Une différence de masse de 400 Da, soit l'équivalent de 5 phosphorylations, est observée après déphosphorylation. La déphosphorylation du complexe CAK a nécessité une mise au point afin de pouvoir déphosphoryler les trois protéines simultanément. Les trois protéines sont considérées comme étant dans des conditions équimolaires à l'issue de leur production. 5, 15 et 20 unités de phosphatase alcaline sont incubées avec 127 pmol de complexe CAK. L'analyse des trois conditions de traitement à la phosphatase alcaline permet de montrer que la déphosphorylation de la Cycline H est complète pour 15 unités de phosphatase alcaline (Figure 26).



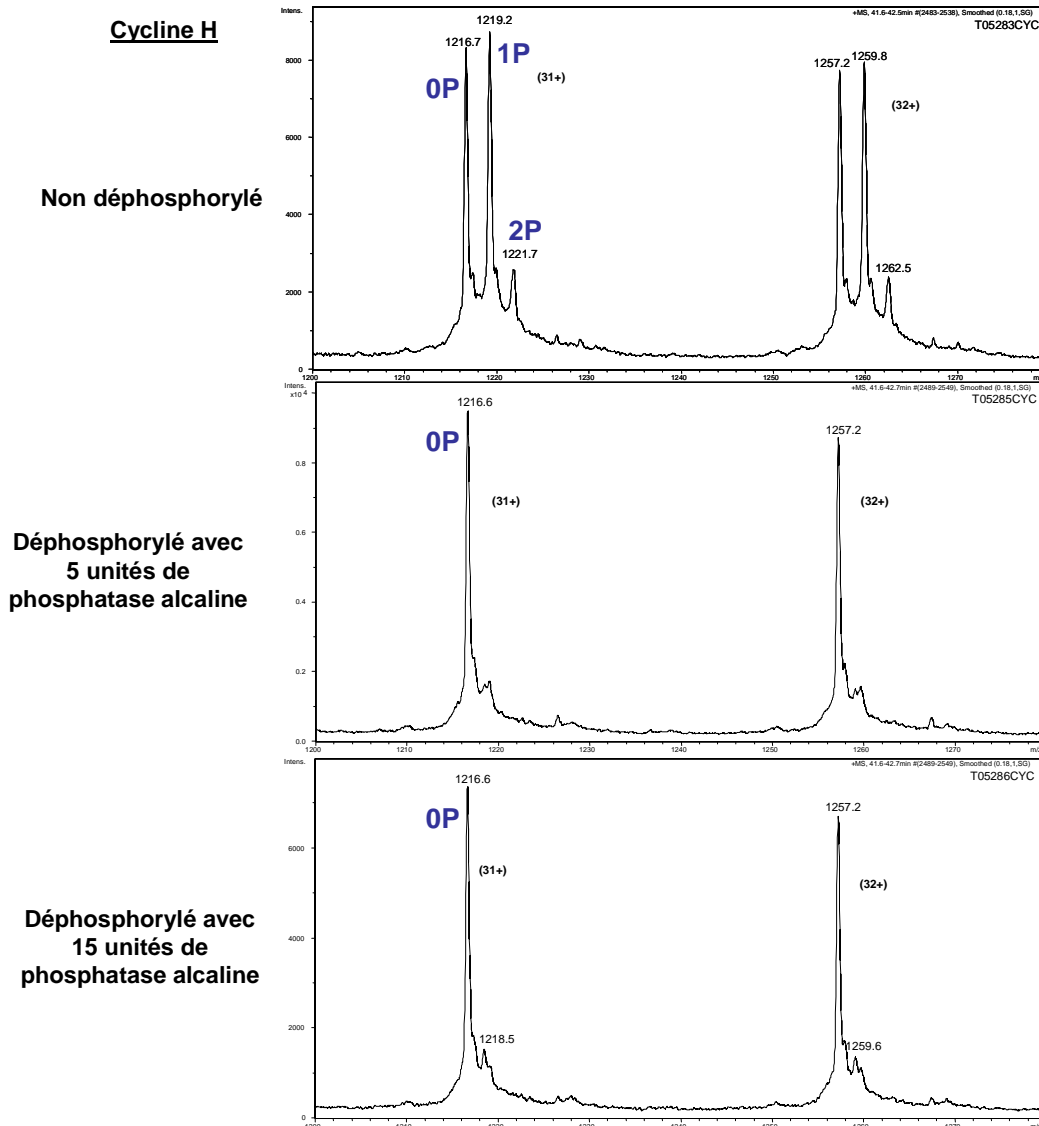


Figure 26 : Mise au point des conditions de déphosphorylations du complexe CAK. Exemple de la Cycline H ( $M_{+0P}$  :  $37684.9 \pm 0.2$  Da;  $M_{+1P}$  :  $37764.8 \pm 0.2$  Da;  $M_{+2P}$  :  $37847.2 \pm 0.2$  Da). Zoom des spectres ESI-MS sur les états de charges 31+ et 32+.

La déphosphorylation complète a été observée pour les protéines MAT1 et Cycline H, mais n'a pas été observée pour CDK7 (Figure 26). La phosphatase alcaline n'a pas eu d'effet de déphosphorylation sur CDK7. Le ratio phosphatase alcaline/complexe a été augmenté, mais aucune déphosphorylation de CDK7 n'a été observée. Une dénaturation du complexe a été tentée avant traitement avec la phosphatase alcaline. Malheureusement, les protéines se sont agrégées et l'analyse devient impossible. Plusieurs hypothèses peuvent expliquer la non-déphosphorylation de CDK7 : soit la phosphorylation pourrait se situer au sein du complexe donc inaccessible à la phosphatase alcaline [Garnier et al., 2001], soit l'hypothèse de phosphorylation est fausse.

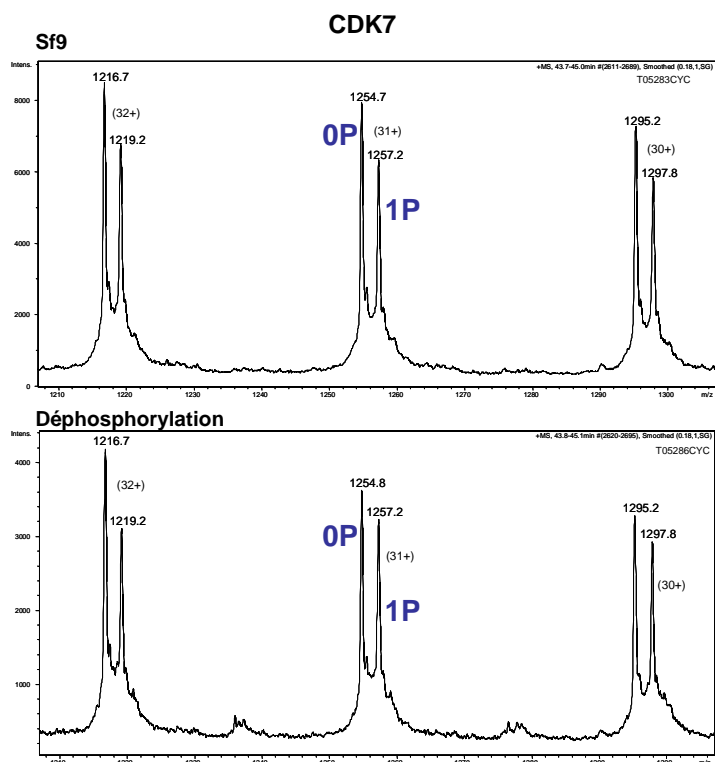


Figure 26 : Détails des spectres de masse ESI-MS de CDK7 ( $M_{+0P}$  :  $40119.5 \pm 0.2$  Da;  $M_{+1P}$  :  $40199.2 \pm 0.3$  Da) avant et après déphosphorylation avec 15 unités de phosphatase alcaline. Zoom sur les états de charges 30+ à 32+.

Malgré cela et dans l'hypothèse où les trois protéines sont phosphorylées, un taux relatif de phosphorylation a pu être évalué (Tableau 1). L'analyse du complexe, produit par Sf9, a été réalisée trois fois (échantillons provenant de la même purification). Cela a permis d'estimer l'erreur sur la précision de mesure du taux de phosphorylation. Elle est identique pour toutes les phosphorylations ( $\pm 1$  %). Le complexe CAK n'est pas homogène du point de vue de la phosphorylation et ne peut être cristallisé directement après production du complexe.

Tableau 1 : Taux de phosphorylations avant et après déphosphorylation.

Protéine	Nombre de phosphorylation	Avant déphosphorylation	Après déphosphorylation
MAT1	0P	65 %	89 %
	1P	35 %	11 %
Cycline H	0P	43 %	100 %
	1P	45 %	0 %
	2P	12 %	0 %
CDK7	0P	56 %	56 %
	1P	44 %	44 %

### 3. Comparaison du taux de phosphorylation des protéines selon la lignée cellulaire utilisée pour la production du CAK et après l'autophosphorylation du complexe

Dans le but d'homogénéiser le complexe en terme de phosphorylation, la production du complexe CAK dans une autre lignée cellulaire ou son autophosphorylation sont testées et leurs taux de phosphorylation des différentes protéines sont estimés.

#### 3.a. Production du complexe avec une autre lignée cellulaire

Le complexe CAK a été produit dans une lignée cellulaire d'insecte (HighFive™). Cette lignée, nommée Hi5, est issue d'une chenille, dévoreuse de choux "*Trichoplusia Ni*" et est couramment utilisée pour la production de protéines. De la même façon que précédemment, les protéines sont analysées par LC-MS. Le complexe produit contient également des phosphorylations mais les protéines sont phosphorylées dans des proportions différentes que celles estimées pour le complexe CAK produit par Sf9. Le taux de phosphorylation de la Cycline H ne varie pas selon la lignée de production (Tableau 2), par contre les taux de phosphorylation de CDK7 et MAT1 ne sont pas identiques à ceux obtenus par la production du complexe par Sf9 (Figure 27).

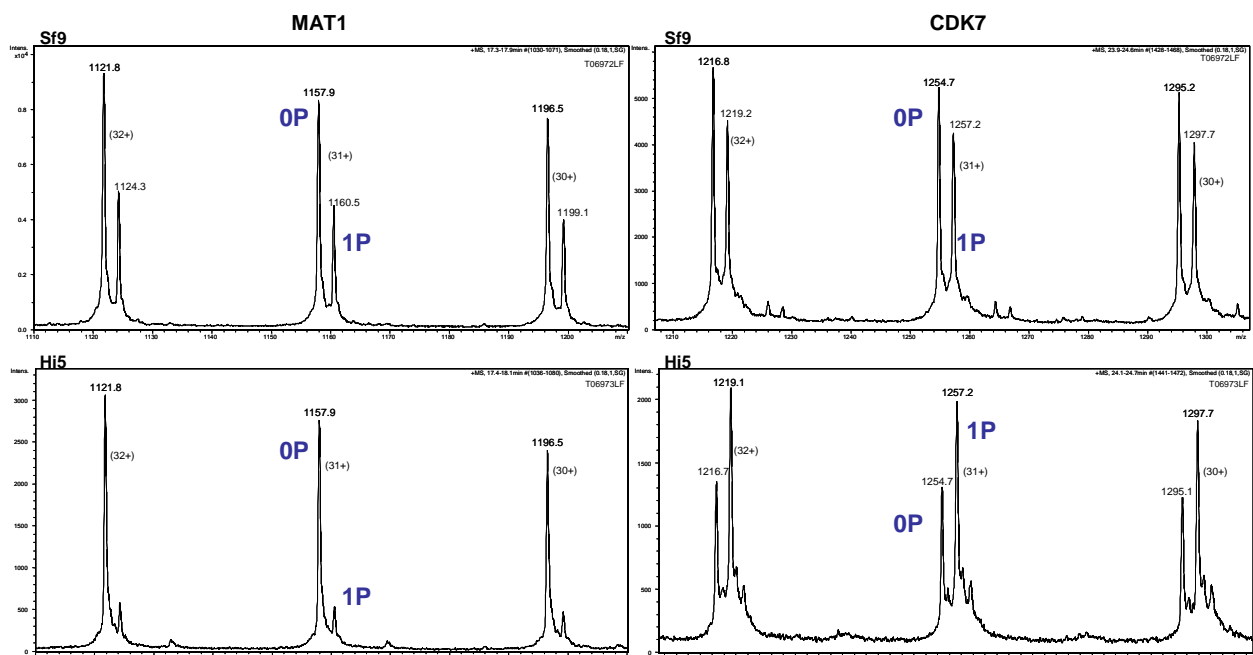


Figure 27 : Détails des spectres de masse ESI-MS de MAT1 ( $M_{+1P} : 35864.3 \pm 0.2$  Da et  $M_{+2P} : 35944.2 \pm 0.3$  Da) et CDK7 ( $M_{+0P} : 40119.5 \pm 0.2$  Da;  $M_{+1P} : 40199.2 \pm 0.3$  Da) issus du complexe produit par Sf9 et Hi5. Zoom sur les états de charges 30+ à 32+.

Tableau 2 : Taux de phosphorylation obtenus avec la lignée cellulaire Sf9 et Hi5

Protéine	Nombre de phosphorylation	Lignée Sf9	Lignée Hi5
MAT1	0P	65 %	84 %
	1P	35 %	16 %
Cycline H	0P	43 %	33 %
	1P	45 %	47 %
	2P	12 %	20 %
CDK7	0P	56 %	42 %
	1P	44 %	58 %

De cette première série d'expériences, nous pouvons conclure que le complexe CAK n'est pas totalement phosphorylé qu'ils s'agissent du complexe produit par les lignées cellulaires Sf9 ou Hi5.

### **3.b. Autophosphorylation du complexe**

Une solution envisagée pour augmenter l'homogénéité de CAK en terme de phosphorylation est l'autophosphorylation. Le complexe s'autophosphoryle en présence d'ATP et au cours du temps (Figure 28). L'autophosphorylation du complexe a été réalisée par incubation du complexe avec de l'ATP (1 mM) et MgCl<sub>2</sub> (10 mM) à température ambiante.

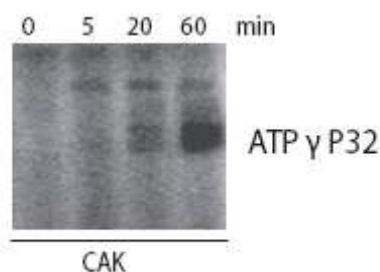


Figure 28 : Gel 1D montrant l'incorporation de ATP P32 au cours du temps dans le complexe CAK, révélé par autoradiographie.

Après autophosphorylation du complexe, les protéines sont analysées par LC-MS comme précédemment pour déterminer quels sont les taux de phosphorylations de chaque protéine, afin de déterminer si l'autophosphorylation est réalisée sur l'ensemble du complexe ou sur une ou deux sous-unités.

Le taux de phosphorylation ne varie pas pour MAT1 ou la Cycline H (Tableau 3). Seule la protéine CDK7 est autophosphorylée (Figure 29). La proportion de la forme CDK7 non-phosphorylée diminue fortement. Par contre, une forme possédant trois phosphorylations apparaît.

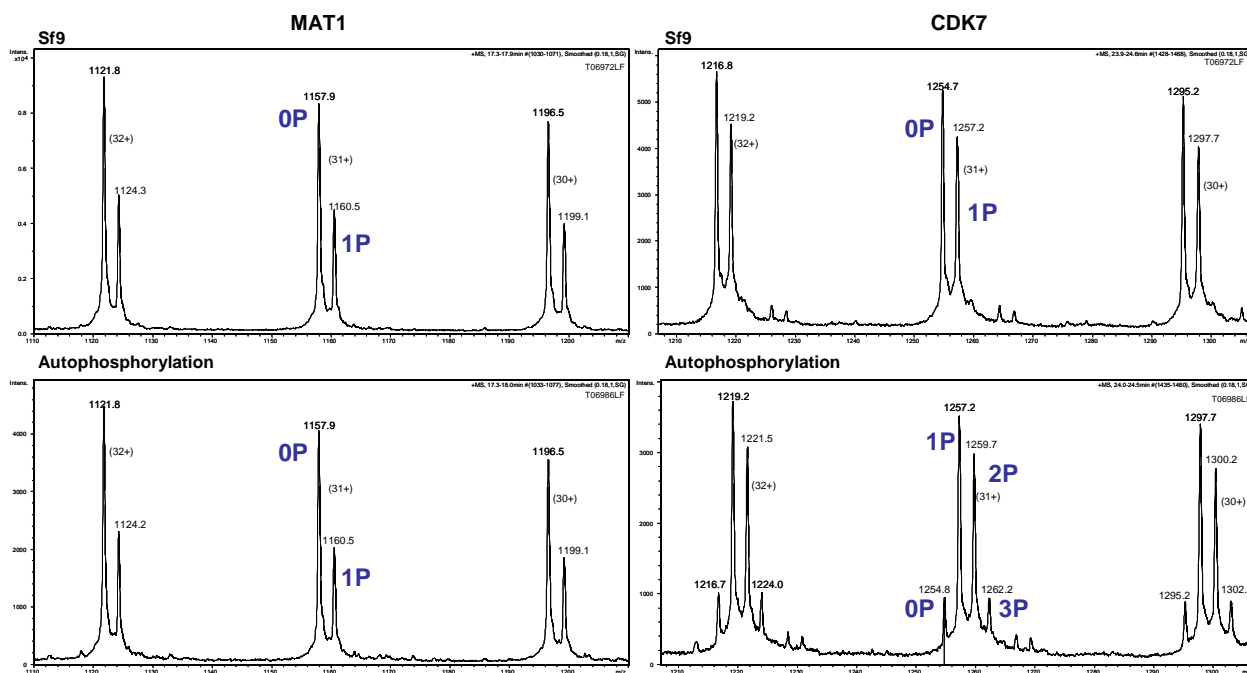


Figure 29 : Détails des spectres de masse ESI-MS de MAT1 ( $M_{+1P}$  :  $35864.3 \pm 0.2$  Da et  $M_{+2P}$  :  $35944.2 \pm 0.3$  Da) et CDK7 ( $M_{+0P}$  :  $40119.5 \pm 0.2$  Da;  $M_{+1P}$  :  $40199.2 \pm 0.3$  Da;  $M_{+2P}$  :  $40280.6 \pm 0.3$  Da et  $M_{+3P}$  :  $40360.9 \pm 0.4$  Da) issus du complexe produit par Sf9 et par Sf9 après autophosphorylation. Zoom sur les états de charges 30+ à 32+.

Tableau 3 : Taux de phosphorylation obtenus avec la lignée cellulaire Sf9 avant et après autophosphorylation.

Protéine	Nombre de phosphorylation	Avant Auto - phosphorylation	Après Auto - phosphorylation
MAT1	0P	65 %	67 %
	1P	35 %	33 %
Cycline H	0P	43 %	40 %
	1P	45 %	47 %
	2P	12 %	13 %
CDK7	0P	56 %	13 %
	1P	44 %	42 %
	2P	0 %	33 %
	3P	0 %	12 %

L'homogénéité en terme de phosphorylation du complexe CAK n'a pas encore pu être obtenue. Cependant, diverses pistes pour parvenir à ce but sont envisageables. Ainsi l'enrichissement en phosphoprotéines semble être une piste intéressante mais est peu fréquemment utilisé dans la

littérature [Dubrovskaja & Souchelnytskyi, 2005]. Plusieurs techniques d'enrichissement peuvent être envisagées tel que :

i) la purification avec des anticorps des résidus phosphorylés du complexe. Cette technique nécessite d'obtenir des anticorps spécifiques aux résidus phosphorylés, ce qui reste difficile pour les phosphosérines et phosphothréonines.

ii) l'enrichissement basé sur l'affinité des phosphorylations avec les métaux. Wolschin et al. ont réalisés des enrichissement de phosphoprotéines à partir d'extraits de feuilles d'arabette des dames "*Arabidopsis thaliana*" par chromatographie d'affinité  $Al(OH)_3$  [Wolschin et al., 2005].

La difficulté de l'enrichissement du complexe CAK est la volonté de ne pas le dénaturer et donc de garder la conformation tridimensionnelle du complexe afin pouvoir de le cristalliser.

#### **4. Localisation des sites de phosphorylations du complexe CAK**

Nous avons ensuite cherché à localiser plus précisément les différentes phosphorylations du complexe CAK, issu de la lignée cellulaire Sf9. Pour cela et afin de simplifier les analyses, une collecte des protéines issues de la chromatographie liquide a été réalisée. Mais les protéines du complexe CAK séparées ne sont pas stables et les protéines s'agrègent. Nous avons donc envisagé de réaliser les digestions enzymatiques sur le complexe CAK directement en solution sans étape chromatographique préalable, dans son tampon de purification. Les protéines ont été digérées à l'aide de la trypsine ou de la chymotrypsine, les deux digestions étant analysées en MALDI-MS et en nanoLC-MS/MS. Les analyses MALDI-MS ont été réalisées à l'aide de deux matrices HCCA et DHB en dépôt "goutte séchée".

##### **4.a. Phosphorylation de CDK7**

La protéine CDK7 a été observée en LC-MS sous deux formes: non-phosphorylée ( $M : 40119.5 \pm 0.2$  Da) et mono-phosphorylée ( $M : 40199.2 \pm 0.3$  Da). La séquence de la protéine CDK7 a été identifiée avec un recouvrement de séquence est de 58 % par les analyses MALDI-MS et en nanoLC-MS/MS (Figure 30).

MALDVKSRAK RYEKLDFLGE QFATVYKAR DKNTNQIVAI KKIKLGHRSE  
AKDGINRTAL REIKLLQELS HPNIIGLLDA FGHKSSNISLV FDFMETDLEV  
 IIKDNSLVLT PSHIKAYMLM TLQGLEYLHQ HWILHRDLKP NNLLLDENG  
LKLADFGLAK SFGSPNRAYT HOVVTRWYRA PELLFGARMY GVGADMWAVG  
 CILAELLLGV PFLPGDSDDL QLTRIFETLG TPTEEQWPD CSLPDYVTFK  
 SFPGIPLHHI FSAAGDLLD LIQGLFLFNP CARITATQAL KMKYFSNRPG  
 PTPGCQLPRP NCPVETLKEQ SNPALAIKRK RTEALEQGL PKKLIFSAWS  
 HPQFEK

Figure 30 : Recouvrement de séquence de CDK7 après analyses MALDI-MS (Traits pleins) et en nanoLC-MS/MS (Pointillés) du complexe CAK digéré soit par la trypsine, soit par la chymotrypsine.

Tout l'intérêt de combiner plusieurs enzymes de digestion et plusieurs modes d'ionisation est illustré au travers de la caractérisation de CDK7.

- Les analyses MALDI-MS permettent d'obtenir une couverture de séquence de 24% alors que les analyses nanoLC-MS/MS permettent d'obtenir une couverture de séquence de 55% (Figure 30). La couverture de séquence est meilleure par l'analyse nanoLC-MS/MS car les peptides sont séparés donc analysés séquentiellement, il y a donc moins de suppression d'ionisation que lors d'une analyse MALDI-MS.

- L'utilisation des deux enzymes de digestions est nécessaire car elle permet d'augmenter sensiblement la couverture de séquence. En effet, seuls trois peptides en commun aux deux digestions sont détectés (Encadrés bleus, figure 31). La digestion avec la trypsine permet d'obtenir une couverture de séquence de 54%, et la digestion avec la chymotrypsine une couverture de 26% (Figure 31). Les autres enzymes de digestions (type GluC, LysC) n'ont pas été utilisées car elles génèrent des peptides de taille élevée qui ne sont pas facilement fragmentés en MS/MS.

MALDVKSRAK RYEKLDFLGE QFATVYKAR DKNTNQIVAI KKIKLGHRSE  
AKDGINRTAL REIKLLQELS HPNIIGLLDA FGHKSSNISLV FDFMETDLEV  
 IIKDNSLVLT PSHIKAYMLM TLQGLEYLHQ HWILHRDLKP NNLLLDENG  
LKLADFGLAK SFGSPNRAYT HOVVTRWYRA PELLFGARMY GVGADMWAVG  
 CILAELLLGV PFLPGDSDDL QLTRIFETLG TPTEEQWPD CSLPDYVTFK  
 SFPGIPLHHI FSAAGDLLD LIQGLFLFNP CARITATQAL KMKYFSNRPG  
 PTPGCQLPRP NCPVETLKEQ SNPALAIKRK RTEALEQGL PKKLIFSAWS  
 HPQFEK

Figure 31 : Recouvrement de séquence de CDK7 après analyses MALDI-MS et en nanoLC-MS/MS du complexe CAK digéré soit par la trypsine (Traits pleins), soit par la chymotrypsine (Pointillés). Les peptides encadrés sont les peptides détectés commun aux deux digestions.

Deux peptides portant une phosphorylation chacun sur CDK7 ont été détectés par analyse MALDI-MS :

L'ion à  $m/z$  844.26 présent avant traitement avec la phosphatase alcaline, n'est plus visible après la déphosphorylation (Figure 31). L'ion issu de cette déphosphorylation est l'ion  $m/z$  764.29 (-80 Da). Cet ion correspond au peptide 161-167 SFGSPNR, déjà présent lors de l'analyse MALDI-MS avant déphosphorylation. Deux sites de phosphorylations potentiels sont présents dans ce peptide (S161 et S164).

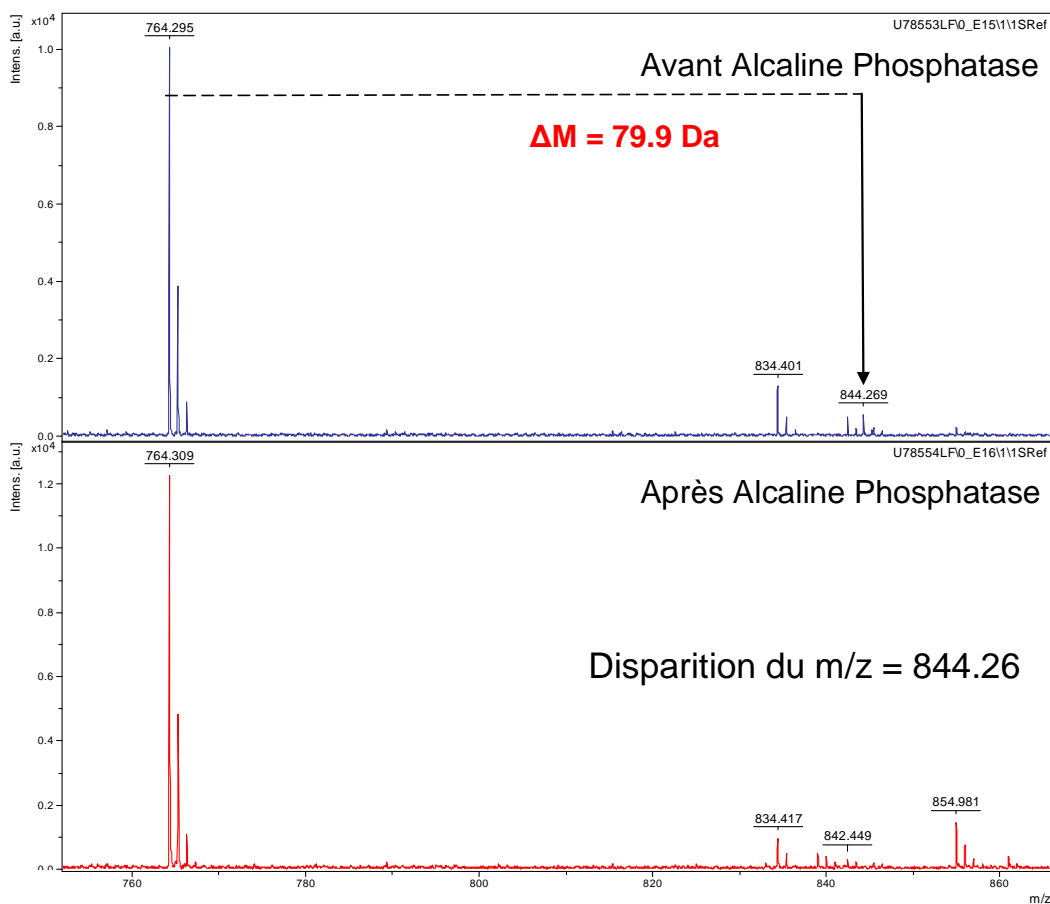


Figure 32 : Détails des spectres de masse MALDI-MS avant et après traitement avec la phosphatase alcaline. Zoom sur la gamme de  $m/z$  750-810 (Matrice HCCA, digestion trypsine).



L'ion à  $m/z$  1154.61 présent avant traitement avec la phosphatase alcaline, n'est plus visible après la déphosphorylation (Figure 33). L'ion issu de cette déphosphorylation correspond à l'ion  $m/z$  1074.65 (-80 Da). Cet ion correspond au peptide 168-176 AYTHQVVTR ayant trois sites de phosphorylations potentiels (Y169, T170 et T175), déjà présent lors de l'analyse MALDI-MS avant déphosphorylation.

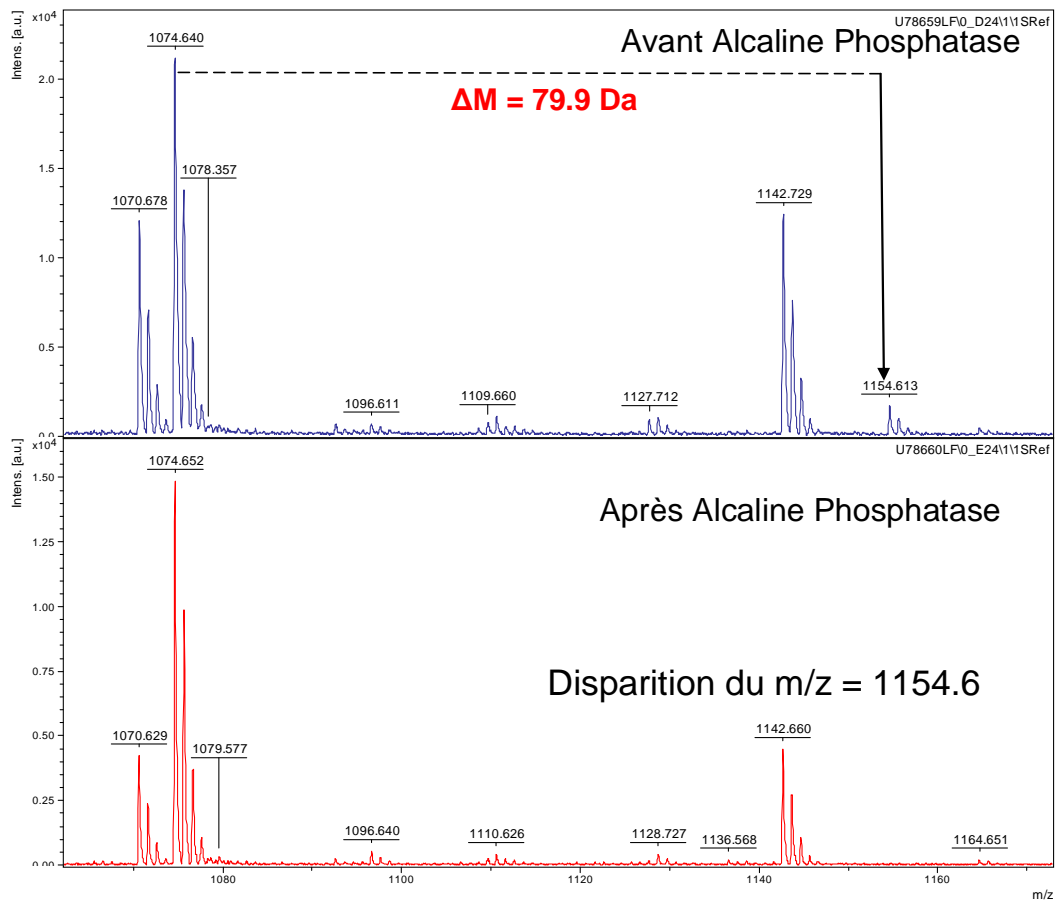
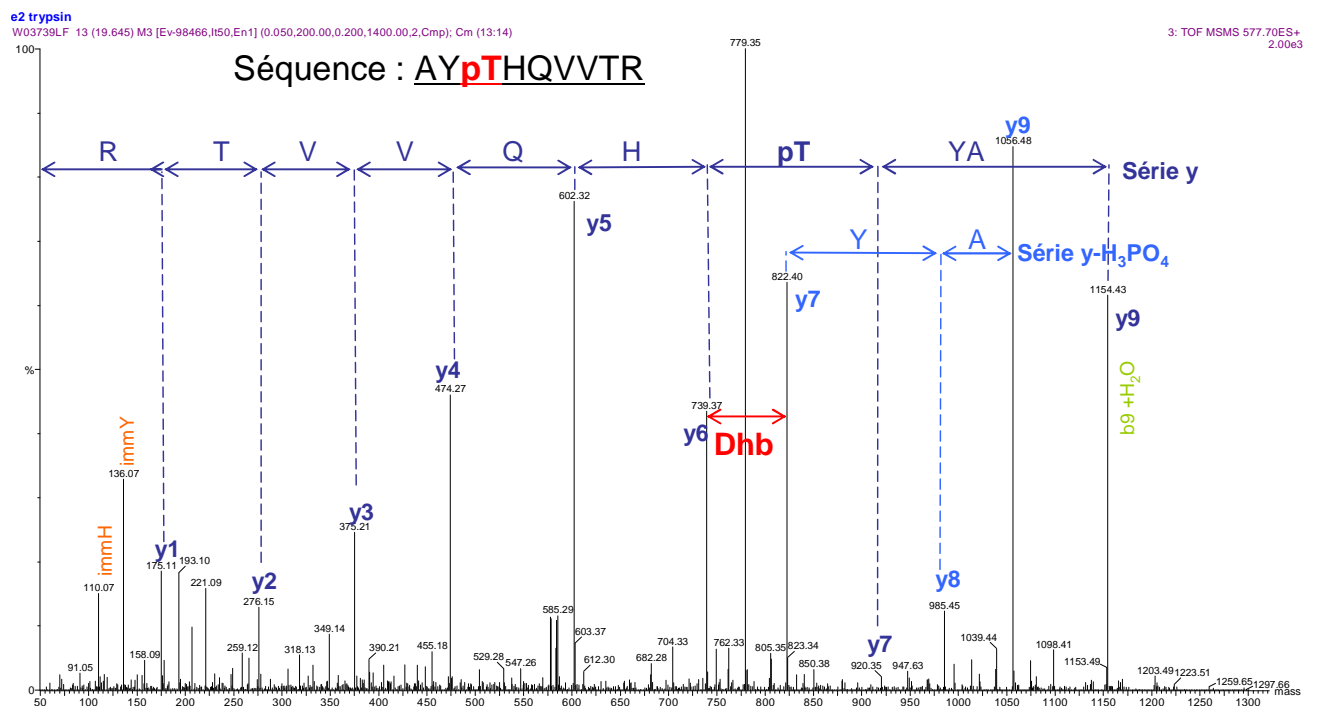
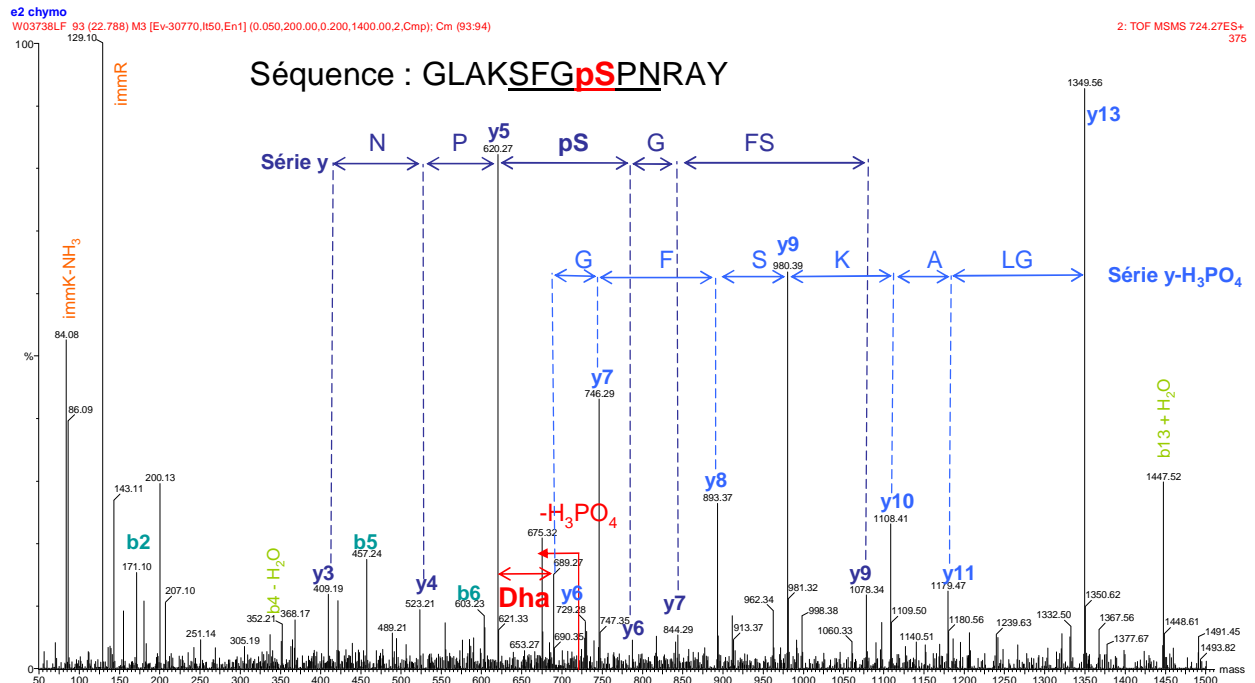


Figure 33 : Détails des spectres de masse MALDI-MS avant et après traitement avec la phosphatase alcaline. Zoom sur la gamme de  $m/z$  1060-1165 (Matrice DHB, digestion trypsine).

Deux peptides ont été fragmentés en nanoLC-MS/MS et correspondent aux peptides observés par l'analyse MALDI-MS : l'ion m/z 724.2, doublement chargé (Figure 34) et l'ion m/z 577.7, doublement chargé (Figure 35).



L'ion  $m/z$  724.2 (Figure 34) correspond au peptide phosphorylé 157-169 GLAKSFGSPNRAY. L'annotation du spectre permet d'identifier un morceau de séquence "SFGSPN" grâce aux ions  $y$  et  $y\text{-H}_3\text{PO}_4$ . Le site de phosphorylation est déterminé par la présence des ions  $y_5$  et  $y_6\text{-H}_3\text{PO}_4$  correspondant à l'acide aminé acide dehydroalanine (Dha). L'acide aminé phosphosérine (pS) est également visible grâce aux ions  $y_5$  et  $y_6$ . L'interprétation de ce spectre permet de situer la phosphorylation sur la sérine S164 du peptide 157-169 "GLAKSFGSPNRAY".

L'ion  $m/z$  577.7 (Figure 35) correspond au peptide 168-176 AYTHQVVTR. L'annotation du spectre de fragmentation permet d'identifier la séquence entière du peptide "AYTHQVVTR" grâce aux ions  $y$  et  $y\text{-H}_3\text{PO}_4$ . Du fait de la présence de l'ion  $m/z$  136.0 correspondant à l'ion immonium tyrosine, la phosphorylation n'est pas située sur cet acide aminé. Le site de phosphorylation est déterminé par la présence des ions  $y_6$  et  $y_7\text{-H}_3\text{PO}_4$  correspondant à l'acide aminé acide dehydroaminobutyrique (Dhb). L'acide aminé phosphothréonine (pT) est également visible grâce aux ions  $y_6$  et  $y_7$ . Le peptide 168-176 "AYTHQVVTR" porte donc une phosphorylation sur la thréonine T170.

Deux phosphorylations ont été mises en évidence, or l'analyse LC-MS a montré que CDK7 été présents sous deux formes non-phosphorylée et mono-phosphorylée, le fait de ne voir d'un seul ajout de + 80 Da ne signifie pas qu'il y a une seule phosphorylation, mais cela peut aussi correspondre à un mélange de formes monophosphorylées sur des acides aminés différents. Dans ce cas, la détection des phosphopeptides est encore plus complexe, car l'abondance de chaque phosphopeptide est très faible et celle du peptide non-phosphorylé correspondant relativement élevée. Ceci pourrait permettre d'expliquer pourquoi les peptides non-phosphorylés sont déjà présents de façon importante dans les spectres MALDI-MS présentés (Figure 32 et 33).

Les analyses MALDI-MS et nanoLC-MS/MS permettent de localiser deux phosphorylations au niveau de la sérine S164 et de la thréonine T170 sur la protéine CDK7, ce qui correspond aux résidus décrits comme étant phosphorylés [Larochelle et al., 2001].

#### **4.b. Phosphorylation de la Cycline H**

La Cycline H a été détectée préalablement par LC-MS sous trois formes: non-phosphorylée ( $M : 37684.9 \pm 0.2$  Da), mono-phosphorylée ( $M : 37764.8 \pm 0.2$  Da) et bi-phosphorylée ( $M : 37847.2 \pm 0.2$  Da). Les analyses MALDI-MS et en nanoLC-MS/MS ont permis d'obtenir un taux de recouvrement de séquence de 74 % (Figure 36). La digestion avec la trypsine permet d'obtenir une couverture de séquence de 51% par l'analyse MALDI-MS et 48% par l'analyse nanoLC-MS/MS, et la digestion avec la chymotrypsine une couverture de 38% par l'analyse MALDI-MS et 30% par l'analyse nanoLC-MS/MS.

MYHNSSQKRH WTFSSSEQLA RLRADANRKF RCKAVANGKV LPNDPVFLEP  
HEEMTLCKYY EKRLLEFCSV FKPAMPRSVV GTACMYFKRF YLNNSVMEYH  
PRIIMLTCAF LACKVDEFNV SSPQFVGNLR ESPLGQEKAL EQILEYELLL  
IQQLNFHLIV HNPYRPFEGF LIDLKTRYPI LENPEILRKT ADDFLNRIAL  
TDAYLLYTPS QIALTAILSS ASRAGITMES YLSESLMLKE NRTCLSQLLD  
IMKSMRNLVK KYEPPRSEEV AVLKQKLERC HSAELALNVI TKKRKGYEDD  
DYVSKKSKHE EEWTDDDLVLV ESL

Figure 36 : Recouvrement de séquence de la Cycline H après analyses MALDI-MS (Traits pleins) et en nanoLC-MS/MS (Pointillés) du complexe CAK digéré soit par la trypsine, soit par la chymotrypsine.

Un peptide portant la phosphorylation sur la Cycline H a été repéré par analyse MALDI-MS (Matrice DHB, digestion trypsine) à  $m/z$  1845.78 après le traitement avec la phosphatase alcaline (Figure 37). L'ion issu de cette déphosphorylation correspond au peptide 309-323 HEEEEWTDDLVEVL. Deux sites de phosphorylations potentiels sont présents dans ce peptide (T315 et S323). Aucun autre peptide n'est détecté comme pouvant être phosphorylé parmi les peptides théoriques issus des digestions de la Cycline H.

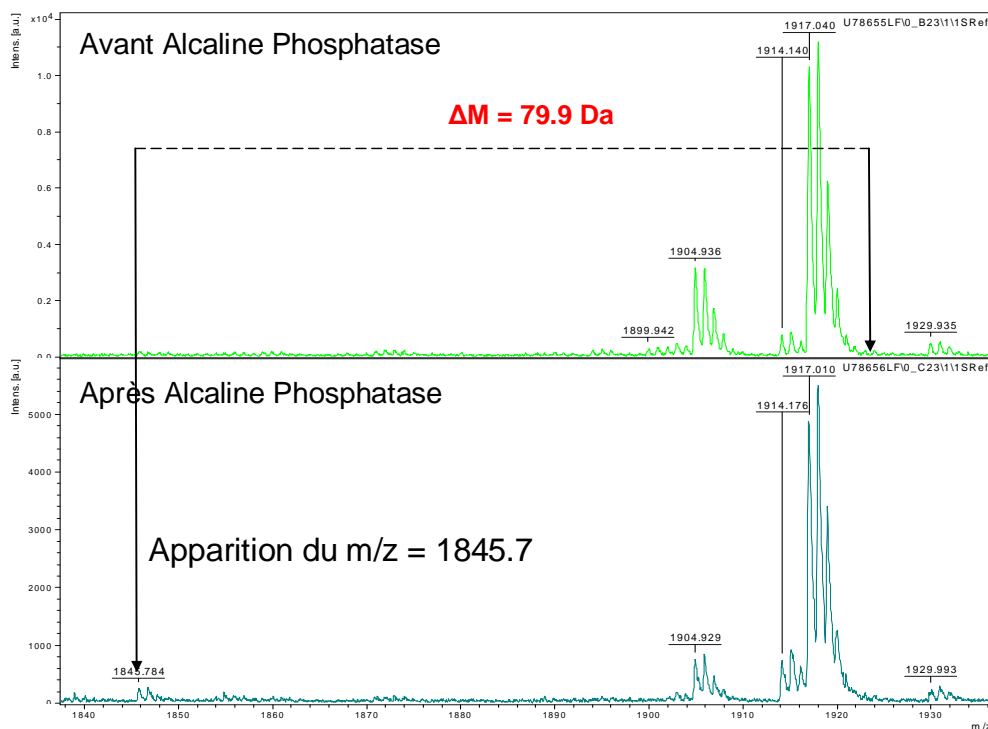


Figure 37 : Détails des spectres de masse MALDI-MS (Matrice DHB, digestion trypsine) avant et après traitement avec la phosphatase alcaline, zoom sur la gamme de  $m/z$  1835-1935.

L'observation du peptide phosphorylé par MALDI-MS a permis de rechercher les ions correspondants aux différents peptides (issus des deux digestions avec au maximum un site de coupure manqué) dans les analyses nanoLC-MS/MS. Deux spectres de fragmentation des ions à  $m/z$  1071.4 (doublement chargé) et  $m/z$  963.8 (triplement chargé), répondant à ces critères, pourraient correspondre respectivement au peptide 307-323 SKHEEEWTDDLVEVL portant une

phosphorylation (Figure 38) et au peptide 309-323 HEEEEWTDDDLVESL avec une phosphorylation. L'ion majoritaire du spectre de fragmentation de l'ion parent à m/z 1071.4 est la perte de H<sub>3</sub>PO<sub>4</sub> (m/z 1022.2). L'annotation du spectre permet d'identifier un morceau de séquence "EEEEWpTDDDLVESL" grâce aux ions y et b. Le site de phosphorylation est déterminé par la présence des ions b<sub>8</sub> et b<sub>9</sub>-H<sub>3</sub>PO<sub>4</sub> correspondant à l'acide aminé acide dehydroaminobutyrique (Dhb). L'acide animé "phosphothréonine" (pT) est également visible grâce aux ions b<sub>8</sub> et b<sub>9</sub> ainsi que les ions y<sub>8</sub> et y<sub>9</sub>.

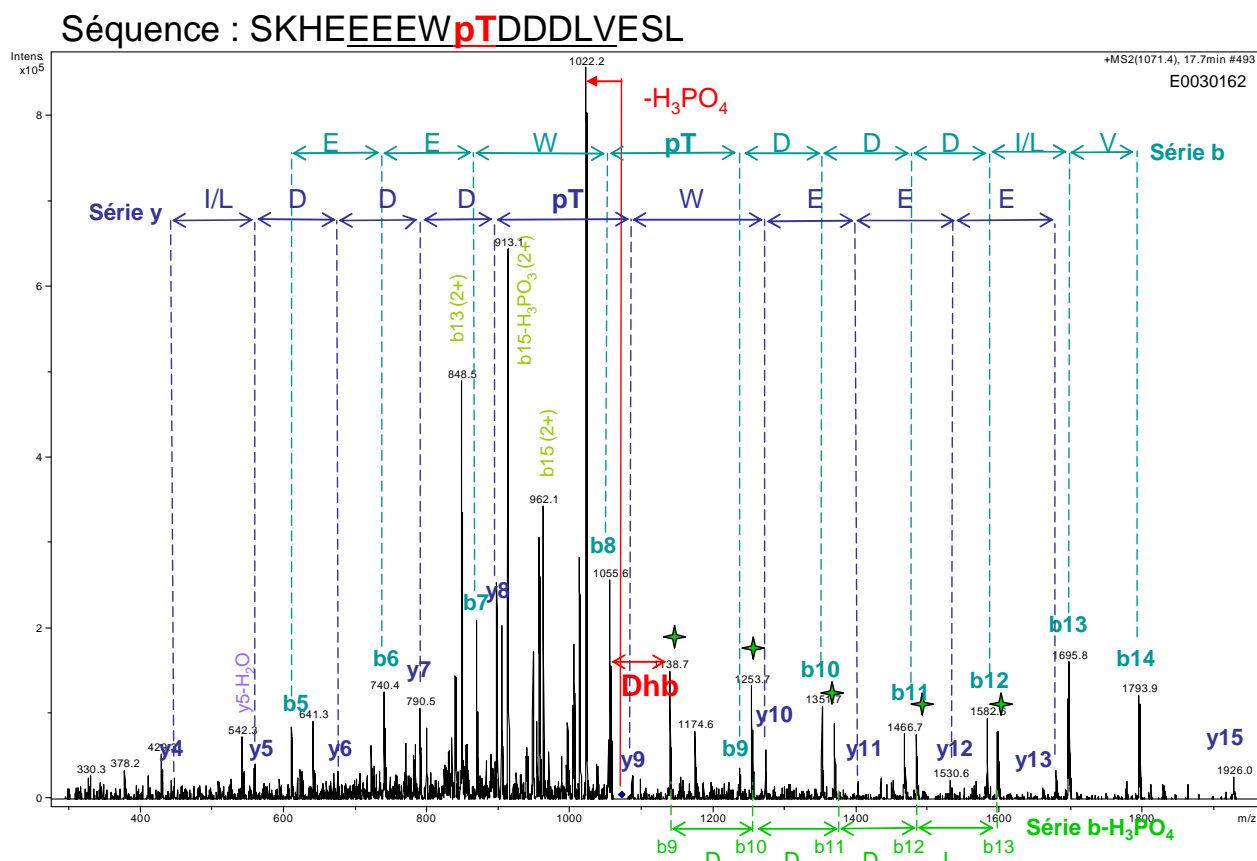


Figure 38 : Spectre de fragmentation du m/z 1071.4 (2+) d'un peptide issu de la digestion avec la trypsine (nanoChip - trappe ionique HCT Plus) permettant la localisation de la phosphothréonine T315 de Cycline H.

Plusieurs phosphorylations ont été décrites dans la littérature pour la Cycline H (S5, S132, S304, S307, T315, S322). Les analyses MALDI-MS et nanoLC-MS/MS ont permis de localiser une phosphorylation au niveau de la thréonine T315 sur la protéine Cycline H.

#### **4.c. Phosphorylation de la protéine MAT1**

L'analyse LC-MS avait permis d'identifier la protéine MAT1 sous deux formes: non-phosphorylée (M : 35864.3 ± 0.2 Da) et mono-phosphorylée (M : 35944.2 ± 0.3 Da). A l'issue des analyses MALDI-MS et en nanoLC-MS/MS, la séquence de la protéine MAT1 a été presque entièrement recouverte avec un taux de couverture de séquence est 86 % (Figure 39). La digestion avec la trypsine permet d'obtenir une couverture de séquence de 69% par l'analyse MALDI-MS et 59% par l'analyse nanoLC-MS/MS, et la digestion avec la chymotrypsine une couverture de 25% par l'analyse MALDI-MS et 46% par l'analyse nanoLC-MS/MS.

MDDQGCPRCK TTKYRNPSLK LMVNVCGHTL CESCVDLLFV RGAGNCPECG  
TPLRKSNFRV QLFEDPTVDK EVEIRKKVLK IYNKREEDFP SLREYNDFLE  
EVEEIVFNLT NNVDLDNTTK KMEIYQKENK DVIQKNKLKL TREQEELEEA  
LEVERQENEQ RRLFIQKEEQ LQQILKRRKNK QAFLELESS DLPVALLLAQ  
HKDRSTQLEM QLEKPKPVKP VTFSTGIKMG QHISLAPIHK LEEALYEQP  
LQIETYGPHV PELEMLGRLG YLNHVRAASP QDLAGGYTSS LACHRALQDA  
FSGLFWQPS

Figure 39 : Recouvrement de séquence de MAT1 après analyses MALDI-MS (Traits pleins) et en nanoLC-MS/MS (Pointillés) du complexe CAK digéré soit par la trypsine, soit par la chymotrypsine.

Un peptide portant la phosphorylation sur MAT1 a été identifié par analyse MALDI-MS (Matrice DHB, digestion trypsine) (Figure 40). L'ion à m/z 1984.83 présent avant traitement avec la phosphatase alcaline, n'est plus visible après la déphosphorylation. L'ion issu de cette déphosphorylation correspond à l'ion m/z 1904.88 (-80 Da). Cet ion correspond au peptide 277-295 AASPQDLAGGYTSSLACHR, déjà présent lors de l'analyse MALDI-MS avant déphosphorylation. Les sites de phosphorylations potentiels sont nombreux dans ce peptide (S279, Y287, T288, S289 et S290).

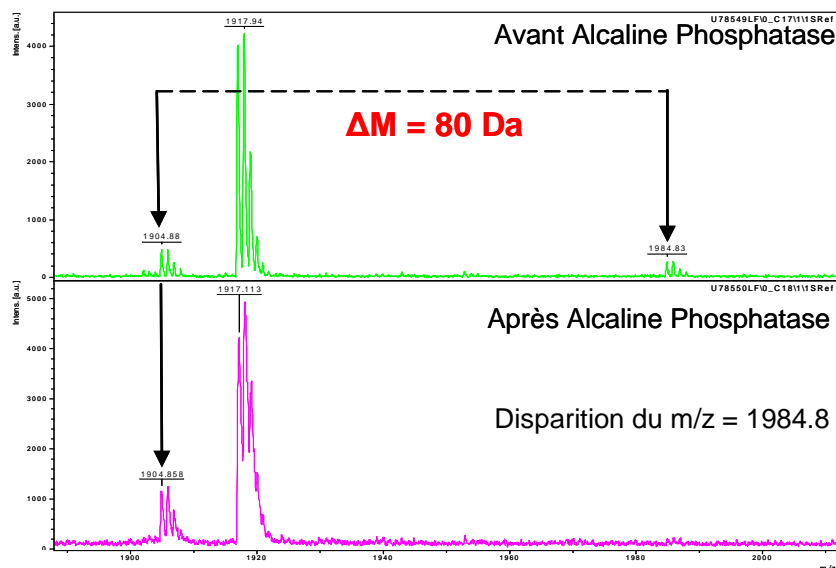


Figure 40 : Détails des spectres de masse MALDI-MS (Matrice DHB, digestion trypsine) avant et après traitement avec la phosphatase alcaline, zoom sur la gamme de m/z 1895-2010.

L'observation du peptide phosphorylé par MALDI-MS a permis de rechercher les ions correspondants aux différents peptides (issus des deux digestions avec au maximum un site de coupure manqué) dans les analyses nanoLC-MS/MS. Le spectre de fragmentation de l'ion m/z 818.3 (doublement chargé), répondant à ces critères, pourrait correspondre au peptide 273-287 NHVRAASPQDLAGGY portant une phosphorylation (Figure 41). L'annotation du spectre de masse MS/MS permet d'identifier un morceau de séquence "VRAApSPQDLAGGY" grâce aux ions y, b et b<sub>7</sub>-H<sub>3</sub>PO<sub>4</sub>. Le site de phosphorylation est déterminé par la présence des ions b<sub>6</sub> et b<sub>7</sub>-H<sub>3</sub>PO<sub>4</sub>.

correspondant à l'acide aminé acide dehydroalanine (Dha). L'acide animé "phosphosérine" (pS) est également visible grâce aux ions  $b_6$  et  $b_7$ . Ce spectre MS/MS permet donc d'identifier le peptide 273-287 NHVRAApSQDLAGGY portant une phosphorylation située sur la S279.

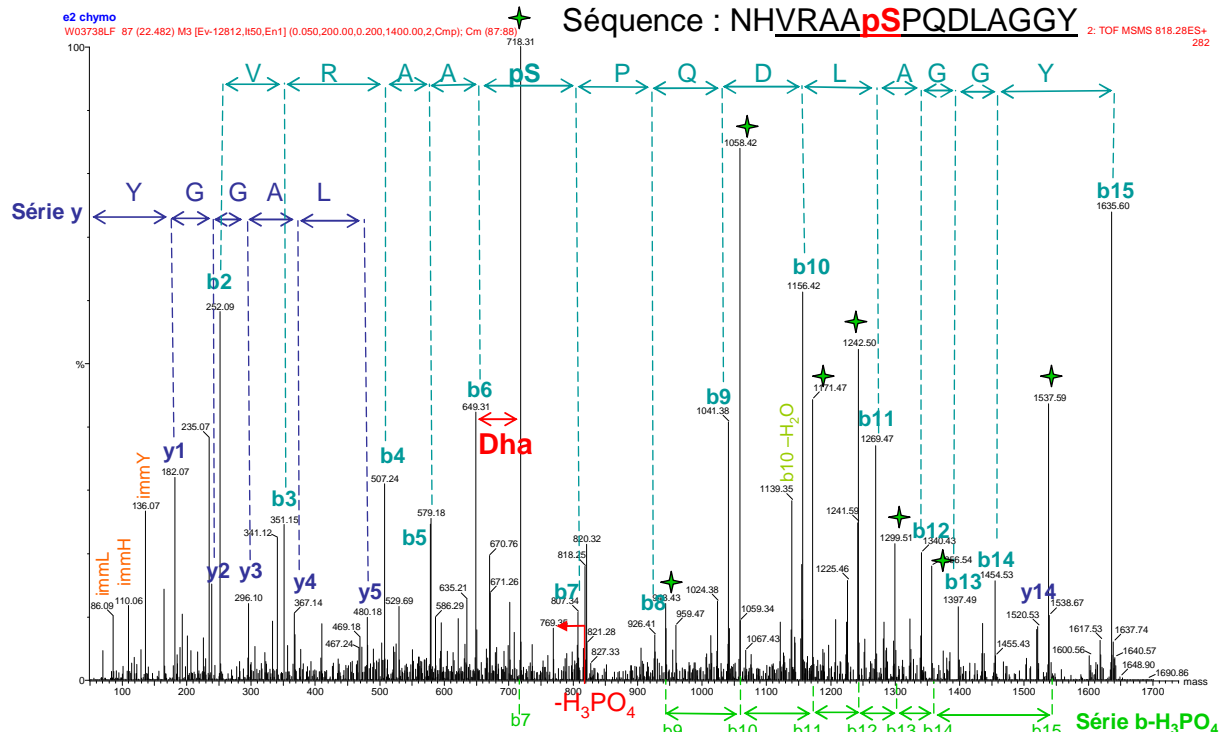


Figure 41 : Spectre de fragmentation du  $m/z$  818.3 (2+) d'un peptide issu de la digestion avec la chymotrypsine (Couplage nanoAquty-QTOF Synapt) permettant la localisation de la phosphosérine S279 de MAT1.

Le peptide 272-287 phosphorylé LNHVRAApSQDLAGGY a également été fragmenté ( $m/z = 874.7$ , 2+ et  $m/z = 583.6$ , 3+). La phosphorylation sur MAT1 se situe donc sur la sérine S279. Les deux autres phosphorylations décrites dans la littérature (T51 et S205) n'ont pas été observées dans nos échantillons.

#### **4.d. Les sites de phosphorylations après autophosphorylation du complexe**

Les analyses LC-MS du complexe après autophosphorylation montrent que les taux de phosphorylations n'ont pas variés pour MAT1 et Cycline H, mais deux espèces phosphorylés supplémentaires ont été observées pour CDK7. Afin d'identifier ces phosphorylations, nous avons réalisé le même travail sur les protéines du complexe autophosphorylé.

Les digestions avec la trypsine ou la chymotrypsine n'ont pas été aussi complètes que celles réalisées sur le complexe CAK non autophosphorylé. Ceci est peut être dû à des adduits aspécifiques d'ATP sur les protéines, présents malgré le dessalage du complexe CAK réalisé à l'issue de l'incubation avec l'ATP et le  $MgCl_2$  et avant digestion. Les enzymes de digestion n'ont donc pas pu accéder à toutes les protéines de la même façon que précédemment.

Dans une analyse nanoLC-MS/MS du complexe CAK autophosphorylé, le peptide 309-323 HEEEEWTDDDLVESL a été observé sous sa forme doublement phosphorylé. Lors de l'autophosphorylation, les taux de phosphorylations de la Cycline H n'ont pas variés, un seul site de phosphorylation a pu être caractérisé. Le spectre de fragmentation de l'ion à  $m/z$  1003.4, doublement chargé comporte deux ions correspond à des pertes de  $H_3PO_4$ : l'ion  $m/z$  954.6, majoritaire, qui représente la perte d'une phosphorylation et l'ion  $m/z$  905.5, représentant la perte de deux phosphorylations (Figure 42-A). Deux séries d'ion  $y$  et  $y-H_3PO_4$  permettent de lire la séquence "EEW**p**TDDDLV" (Figure 42-B). La sérine S322 étant l'autre acide aminé pouvant porter une phosphorylation, donc la seconde phosphorylation est supposé être localisée sur la sérine. La phosphorylation sur la thréonine T315 est confirmée par les ions  $y_8-H_3PO_4$  et  $y_9-2(H_3PO_4)$ . Cela permet de mettre en évidence deux phosphorylations (T315 et S322) sur le peptide 309-323 HEEEEWTDDDLVESL de la Cycline H.

Le peptide 309-323 HEEEEWTDDDLVESL portant uniquement une phosphorylation sur la sérine S322 n'a pas été détectée. Ce peptide sous cette forme n'est peut être pas présent dans la protéine, il est peut être nécessaire que la thréonine T315 soit phosphorylée pour pouvoir phosphoryler la sérine S322. En effet, cette phosphorylation a toujours été observée uniquement avec la thréonine T315 phosphorylé [Dephoure et al., 2008].



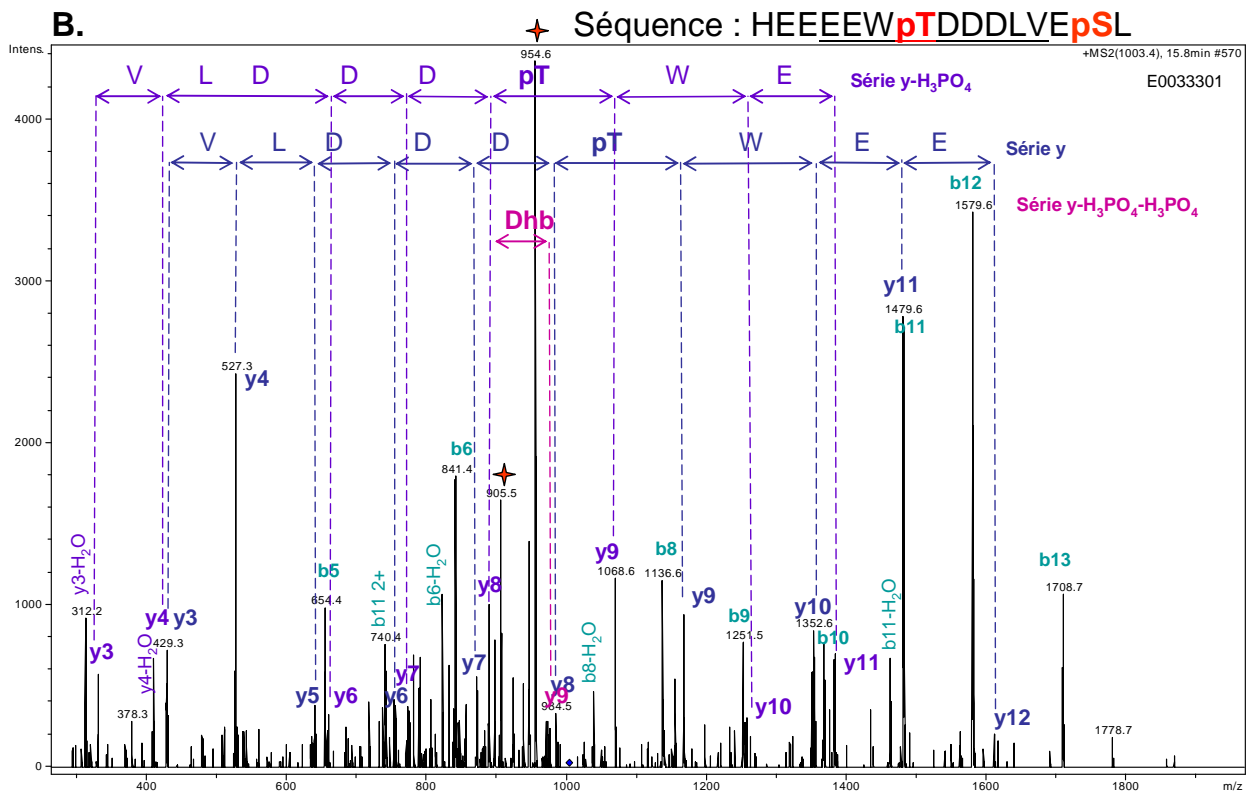
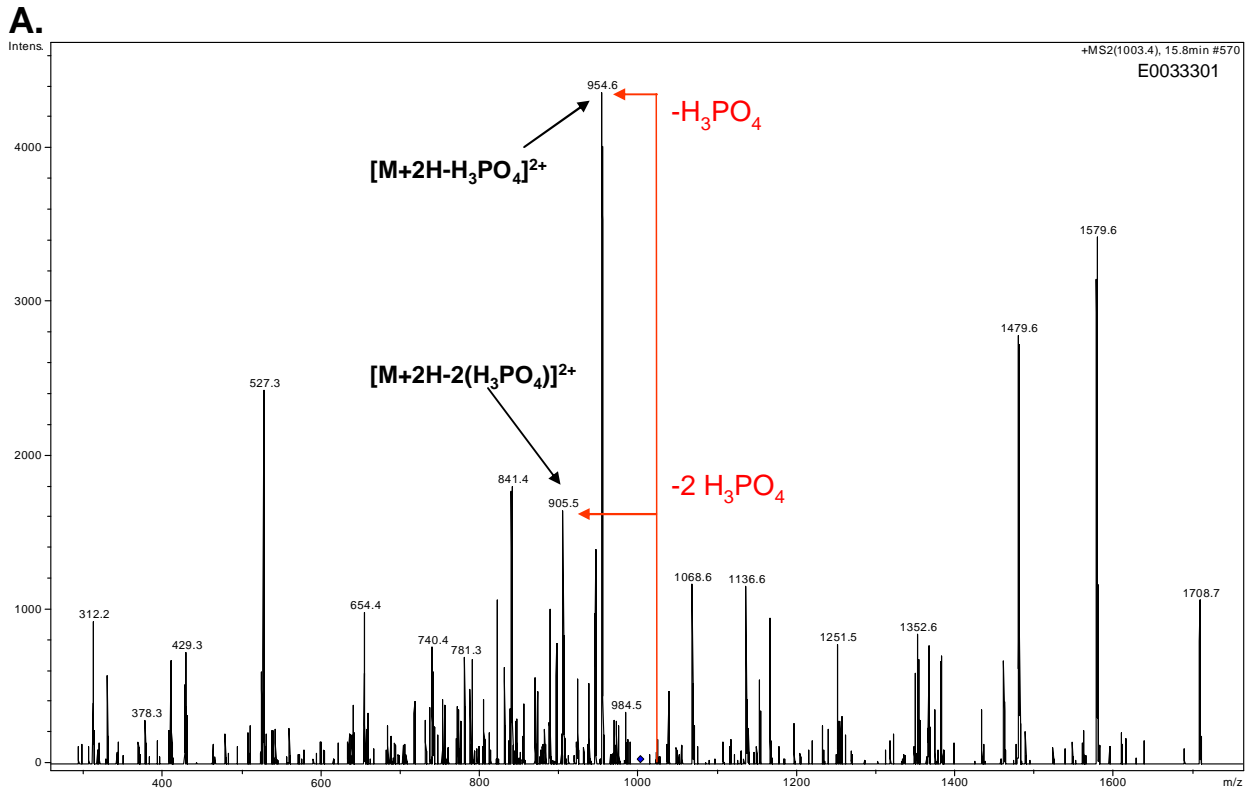


Figure 42 : Spectre de fragmentation du  $m/z$  1003.4 ( $2+$ ) d'un peptide issu de la digestion avec la trypsine (nanoChip - trappe ionique HCT Plus) permettant la localisation de la phosphothréonine T315 et de la phosphosérine S322 de Cycline H.

Des espèces di-phosphorylée ( $M : 40280.6 \pm 0.3$  Da) et une forme tri-phosphorylée ( $M : 40360.9 \pm 0.4$  Da) de CDK7 ont été identifiées par LC-MS après autophosphorylation, on suppose donc

qu'un nouveau site de phosphorylation est présent sur CDK7. Aucune des analyses nanoLC-MS/MS réalisées après autophosphorylation n'a permis de mettre en évidence un site de phosphorylation supplémentaire. Une partie des peptides ne sont pas identifiés, dû au faible recouvrement de CDK7 et il est possible que le site de phosphorylation puisse se situer sur ces peptides. Il importe donc d'augmenter le taux de recouvrement de CDK7 en combinant différentes enzymes de digestion pour obtenir des peptides de taille à pouvoir être analysé en nanoLC-MS/MS.

La caractérisation des sites de phosphorylations après autophosphorylation du complexe fera l'objet d'une nouvelle étude.

#### **4.e. Les fragmentations des peptides phosphorylés**

##### ***4.e.1) Comparaison des spectres de fragmentation d'un peptide phosphorylé et non phosphorylé***

La comparaison des spectres de fragmentation d'un peptide non-phosphorylé et d'un peptide phosphorylé permet de mettre en évidence la complexité de l'interprétation. Le spectre de fragmentation du peptide non phosphorylé 272-287 "LNHVRAASPQDLGGY" (m/z 835.3, doublement chargé) est comparé au spectre de fragmentation du peptide phosphorylé 272-287 LNHVRAApSPQDLGGY (m/z 874.7, doublement chargé) portant une phosphorylation située sur la S279 (Figure 43).

Les ions  $y_1$  à  $y_7$  et les ions  $b_1$  à  $b_7$  ont des masses identiques pour les spectres de fragmentations du peptide non-phosphorylé et du peptide phosphorylé (Figure 43, ions en noirs). Au contraire, les ions  $b_8$  à  $b_{16}$  et  $y_8$  à  $y_{16}$  seront affectés par la présence de la phosphorylation. Certains ions formés portent la phosphorylation, ils auront donc un m/z de +80 Da que les mêmes ions non phosphorylés (Figure 43, ions en vert) et certains ions formés correspondent au peptide ayant subi une  $\beta$ -élimination, les ions (Figure 43, ions en rouge) auront une différence de m/z de 98 Da avec ceux portant la phosphorylation.

Par exemple, les ions  $b_8$  formés ne sont donc pas identiques et permettent de localiser la phosphorylation:

- l'ion  $b_8$  (m/z 849.38) du peptide non phosphorylé contient la sérine,
- l'ion  $b_8$  (m/z 831.39) du peptide phosphorylé contient la dehydroalanine, acide aminé résultant de la  $\beta$ -élimination de la phosphosérine,
- l'ion  $b_8$  (m/z 929.38) du peptide phosphorylé contient la phosphosérine.

L'ensemble des ions permet de situer de séquencer le peptide et de localiser la phosphorylation si les ions issus de la fragmentation peptidique sont détectés.

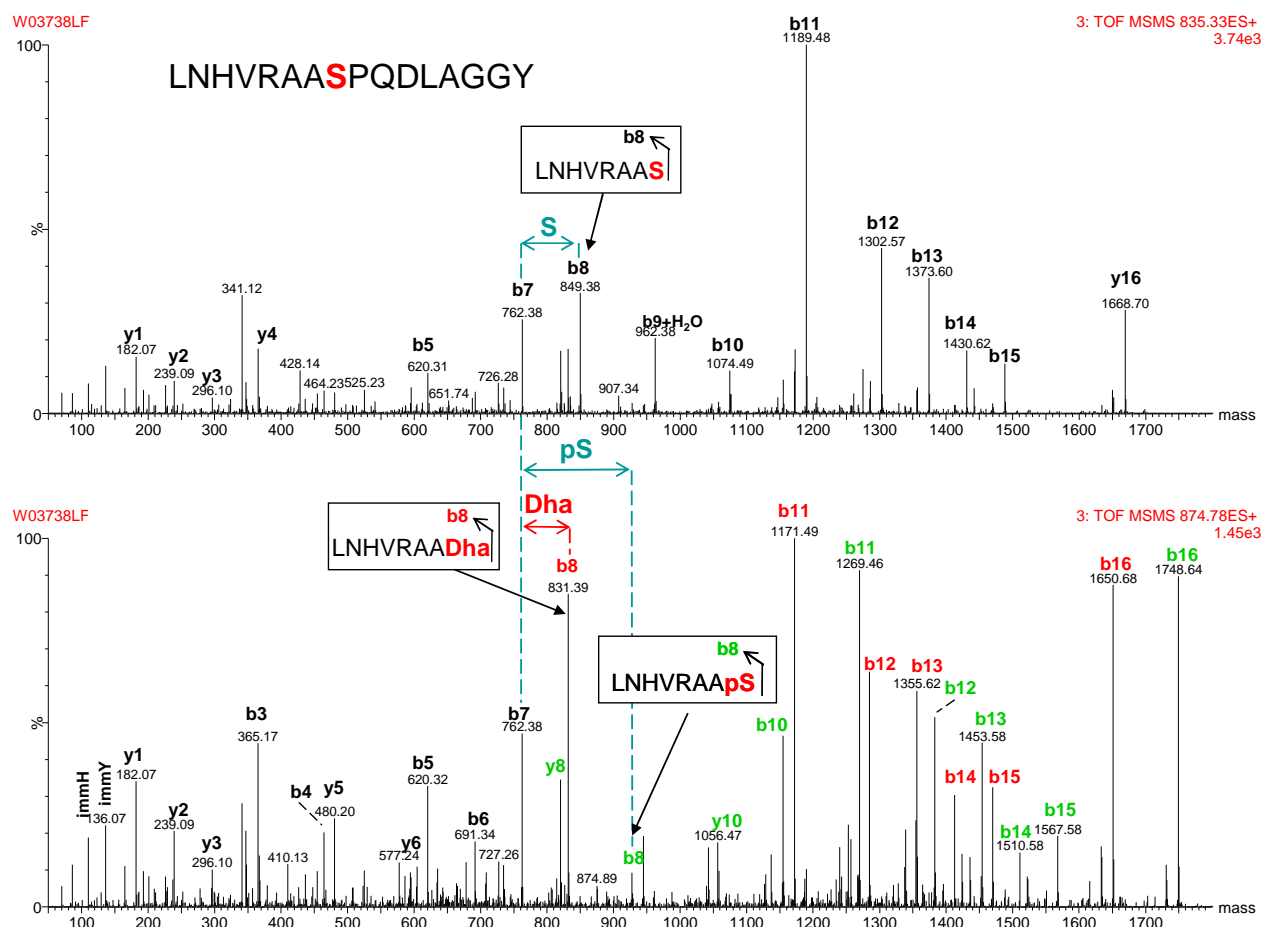


Figure 43 : Spectres de fragmentation de l'ion m/z 835.3, doublement chargé correspondant au peptide 272-287 "LNHVRAASPQDLAGGY" non-phosphorylé (en haut) et de l'ion m/z 874.7, doublement chargé correspondant au peptide 272-287 "LNHVRAApSPQDLAGGY" portant une phosphorylation située sur la S279 (en bas). Les ions notés en vert portent la phosphorylation, ceux en rouge correspondent aux ions ayant subi la perte de H<sub>3</sub>PO<sub>4</sub>.

#### 4.e.2) Comparaison des spectres de fragmentation d'un peptide phosphorylé obtenus sur un Q-TOF et une trappe ionique

Pour la nanoLC-MS/MS, les deux couplages disponibles au laboratoire ont été utilisés dans cette étude: le système de chromatographie sur support microfluidique nanoChip (Agilent Technologies) couplé à la trappe ionique HCT Plus (Bruker Daltonics) et le couplage chromatographie ultra haute performance nanoAquity-QTOF Synapt (Waters Corp.).

La chromatographie liquide a été réalisée différemment sur les deux couplages (longueur de colonne, gradient) et la vitesse de scan des spectromètres de masse est différente. Les géométries de

cellules de collision (trappe ionique 3D / "T-Wave") et les énergies de fragmentation sont différentes. Les informations de fragmentation du même ion ne sont donc pas identiques, notamment au niveau des peptides ayant une modification post-traductionnelle labile [Nita-Lazar et al., 2008]. L'intensité de l'ion correspondant à la perte de l'acide phosphorique est diminuée par les multiples collisions réalisées dans la cellule de collision du QTOF, ce qui augmente le rendement de fragmentation des ions issus du squelette peptidique et permet ainsi un séquençage du phosphopeptide plus aisé.

L'ion  $m/z$  874.7 (doublement chargé), correspondant au peptide LNHVRAApSPQDLAGGY, a été fragmenté sur la trappe ionique (Figure 44-A) et sur le Synapt (Figure 44-B).

L'ion majoritaire ( $m/z$  826.1) sur le spectre MS/MS issu de la trappe ionique (Figure 44-A) est l'ion correspondant à la perte de  $H_3PO_4$ . Cet ion n'est pas présent dans le spectre MS/MS du Synapt. Lors de l'analyse dans la trappe ionique, l'ion  $m/z$  826.1, correspondant au peptide déphosphorylé, est majoritaire. Dans ce cas, les fragments présents sont suffisants pour identifier le morceau de séquence "VRAApSPQDLA" et localiser la phosphorylation grâce aux acides aminés "phosphosérine" (pS) et "acide dehydroalanine" (Dha).

Dans le spectre de fragmentation issu de la fragmentation avec le Synapt (Figure 44-B), l'ion correspondant à la perte de  $H_3PO_4$  n'est pas présent. Les fragments permettent donc d'identifier une plus grande partie de la séquence "VRAApSPQDLAGGY" et de situer également la phosphorylation.

La perte de l'acide phosphorique est majoritairement observée sur les spectres de fragmentation issus de la trappe ionique, cette perte spécifique permet de repérer facilement les phosphopeptides au sein d'une analyse complexe. Cependant, cette perte masque souvent des informations de séquences, nécessaire pour la localisation de la phosphorylation sur un acide aminé. La fragmentation avec le Synapt permet d'obtenir plus facilement ces informations, puisque l'ion correspondant à la perte du groupement phosphate est peu ou pas présent dans les spectres de fragmentation. Une analyse du même digeste de peptides sur les deux instruments permet d'obtenir globalement plus d'informations de séquence complémentaires, les ions de fragmentations détectés n'étant pas tous identiques.

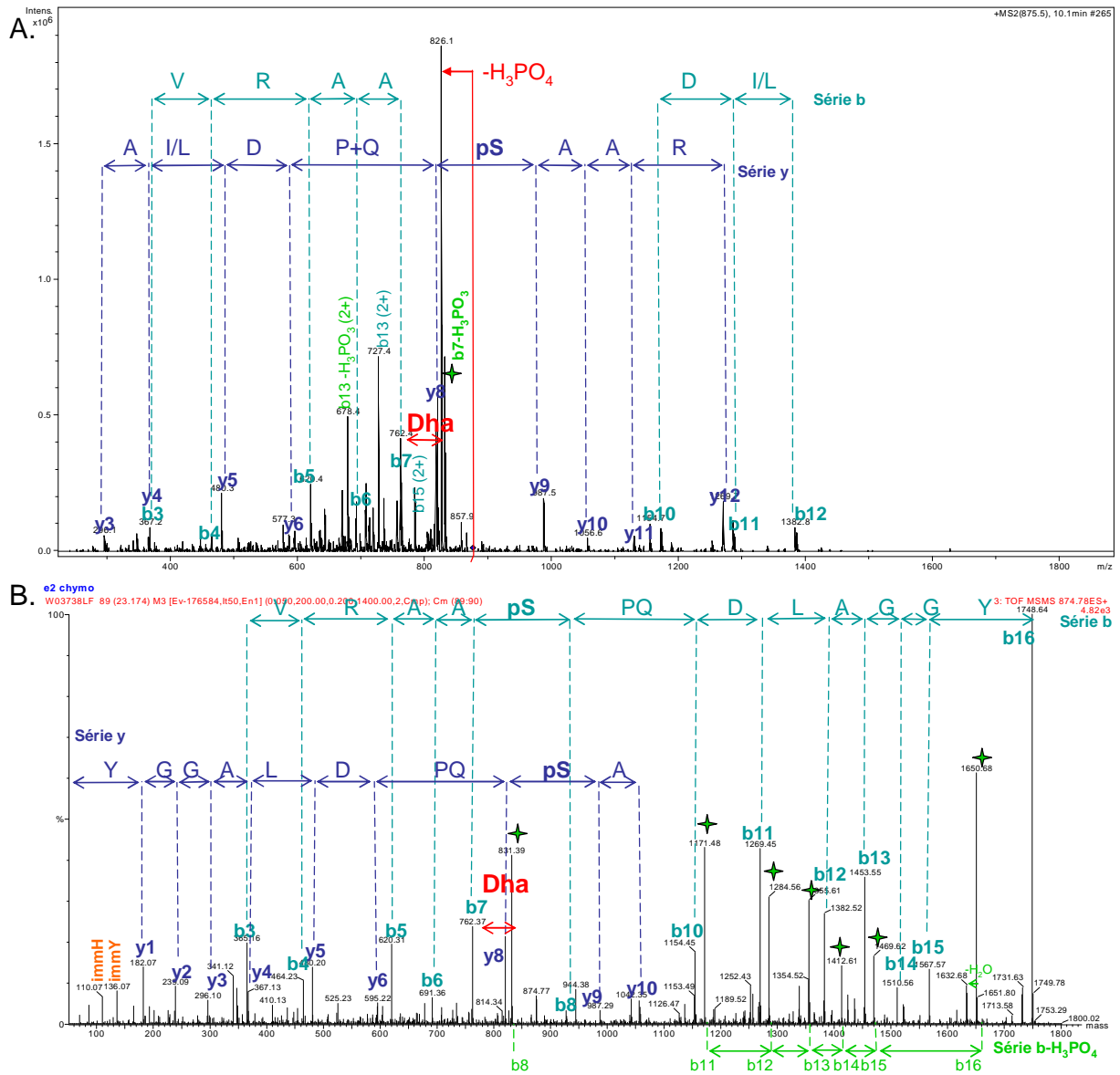


Figure 44 : Spectres de fragmentation de l'ion m/z = 874.7; 2+ correspondant au peptide LNHVRApSPQDLGGY. A) réalisé avec le couplage nanoChip couplé à la trappe ionique HCT Plus B) réalisé avec le couplage nanoAquity-QTOF Synapt.

## **5. Conclusion et perspectives**

Le but de cette étude était de caractériser l'état de phosphorylation (taux de phosphorylation et site de phosphorylation) du complexe CAK, afin de guider la production du complexe CAK totalement phosphorylé. Les taux de phosphorylation ont pu être estimés par des analyses LC-MS. Les différents essais (lignées cellulaires et autophosphorylation) n'ont pour l'instant pas permis d'obtenir un complexe homogène en phosphorylation, mais l'enrichissement en complexe phosphorylé pour atteindre ce but est envisagé.

Des sites de phosphorylation sur les trois protéines composant le complexe CAK ont été identifiés grâce à une combinaison astucieuses de deux digestions enzymatique et à la réalisation combinée d'analyses MALDI-MS et nanoLC-MS/MS. Les phosphorylations ont été identifiées sur des protéines recombinantes produites dans des lignées cellulaires d'insectes. La question de la pertinence biologique de la présence de ces phosphorylations reste posée : le complexe CAK endogène possède-t-il les mêmes phosphorylations?

Les perspectives de ce travail sont de caractériser les phosphorylations sur le complexe CAK endogène et de localiser les phosphorylations de CDK7 à l'issue de l'autophosphorylation du complexe.

## **Conclusion générale de la partie "Caractérisation fine de protéines purifiées phosphorylées"**

L'identification et la caractérisation des phosphorylations n'est pas facile, du fait de la nature très labile de cette modification, qui rend sa détection difficile et sa fragmentation complexe à interpréter. La stratégie mise en place pour l'étude des phosphorylations de protéines purifiées en deux étapes a permis de mettre en évidence :

- Des taux de phosphorylations différents selon la lignée cellulaire utilisée pour la production des protéines.
- L'influence de l'autophosphorylation sur le taux de phosphorylation des protéines.
- La localisation des sites de phosphorylation de chaque protéine.





## **Références bibliographique**

### **A.**

Akoulitchev, S., Chuikov, S. and Reinberg, D. (2000). "TFIIH is negatively regulated by cdk8-containing mediator complexes." Nature **407**(6800): 102-6.

Annan, R. S. and Carr, S. A. (1996). "Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry." Anal Chem **68**(19): 3413-21.

Arrigoni, G., Resjo, S., Levander, F., Nilsson, R., Degerman, E., Quadroni, M., Pinna, L. A. and James, P. (2006). "Chemical derivatization of phosphoserine and phosphothreonine containing peptides to increase sensitivity for MALDI-based analysis and for selectivity of MS/MS analysis." Proteomics **6**(3): 757-66.

### **B.**

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C. and Gygi, S. P. (2004). "Large-scale characterization of HeLa cell nuclear phosphoproteins." Proc Natl Acad Sci U S A **101**(33): 12130-5.

Bialik, S. and Kimchi, A. (2006). "The death-associated protein kinases: structure, function, and beyond." Annu Rev Biochem **75**: 189-210.

Budnik, B. A., Lee, R. S. and Steen, J. A. (2006). "Global methods for protein glycosylation analysis by mass spectrometry." Biochim Biophys Acta **1764**(12): 1870-80.

### **C.**

Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., Syka, J. E., Bai, D. L., Shabanowitz, J., Burke, D. J., Troyanskaya, O. G. and Hunt, D. F. (2007). "Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry." Proc Natl Acad Sci U S A **104**(7): 2193-8.

### **D.**

D'Ambrosio, C., Arena, S., Fulcoli, G., Scheinfeld, M. H., Zhou, D., D'Adamio, L. and Scaloni, A. (2006). "Hyperphosphorylation of JNK-interacting protein 1, a protein associated with Alzheimer disease." Mol Cell Proteomics **5**(1): 97-113.

D'Ambrosio, C., Salzano, A. M., Arena, S., Renzone, G. and Scaloni, A. (2007). "Analytical methodologies for the detection and structural characterization of phosphorylated proteins." J Chromatogr B Analyt Technol Biomed Life Sci **849**(1-2): 163-80.

Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oppermann, F. S., Korner, R., Greff, Z., Keri, G., Stemmann, O. and Mann, M. (2008). "Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle." Mol Cell **31**(3): 438-48.

DeGnore, J. P. and Qin, J. (1998). "Fragmentation of phosphopeptides in an ion trap mass spectrometer." J Am Soc Mass Spectrom **9**(11): 1175-88.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J. and Gygi, S. P. (2008). "A quantitative atlas of mitotic phosphorylation." Proc Natl Acad Sci U S A **105**(31): 10762-7.

Dubrovskaya, A. and Souchelnytskyi, S. (2005). "Efficient enrichment of intact phosphorylated proteins by modified immobilized metal-affinity chromatography." Proteomics **5**(18): 4678-83.

#### **G.**

Garnier, C., Lafitte, D., Jorgensen, T. J., Jensen, O. N., Briand, C. and Peyrot, V. (2001). "Phosphorylation and oligomerization states of native pig brain HSP90 studied by mass spectrometry." Eur J Biochem **268**(8): 2402-7.

Gropengiesser, J., Varadarajan, B. T., Stephanowitz, H. and Krause, E. (2009). "The relative influence of phosphorylation and methylation on responsiveness of peptides to MALDI and ESI mass spectrometry." J Mass Spectrom **44**(5): 821-831.

#### **H.**

Hem, S., Rofidal, V., Sommerer, N. and Rossignol, M. (2007). "Novel subsets of the Arabidopsis plasmalemma phosphoproteome identify phosphorylation sites in secondary active transporters." Biochem Biophys Res Commun **363**(2): 375-80.

#### **J.**

Jalili, P. R. and Ball, H. L. (2008). "Novel reversible biotinylated probe for the selective enrichment of phosphorylated peptides from complex mixtures." J Am Soc Mass Spectrom **19**(5): 741-50.

Janek, K., Wenschuh, H., Bienert, M. and Krause, E. (2001). "Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry." Rapid Commun Mass Spectrom **15**(17): 1593-9.

Jin, W. H., Dai, J., Zhou, H., Xia, Q. C., Zou, H. F. and Zeng, R. (2004). "Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry." Rapid Commun Mass Spectrom **18**(18): 2169-76.

#### **K.**

Känge, R., Selditz, U., Granberg, M., Lindberg, U., Ekstrand, G., Ek, B. and Gustafsson, M. (2005). "Comparison of different IMAC techniques used for enrichment of phosphorylated peptides." J Biomol Tech **16**(2): 91-103.

Kim, K. K., Chamberlin, H. M., Morgan, D. O. and Kim, S. H. (1996). "Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase." Nat Struct Biol **3**(10): 849-55.

Kjellstrom, S. and Jensen, O. N. (2004). "Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins." Anal Chem **76**(17): 5109-17.

Klemm, C., Otto, S., Wolf, C., Haseloff, R. F., Beyermann, M. and Krause, E. (2006). "Evaluation of the titanium dioxide approach for MS analysis of phosphopeptides." J Mass Spectrom **41**(12): 1623-32.

Klemm, C., Schroder, S., Gluckmann, M., Beyermann, M. and Krause, E. (2004). "Derivatization of phosphorylated peptides with S- and N-nucleophiles for enhanced ionization efficiency in matrix-

assisted laser desorption/ionization mass spectrometry." Rapid Commun Mass Spectrom **18**(22): 2697-705.

Kuyama, H., Sonomura, K. and Nishimura, O. (2008)."Sensitive detection of phosphopeptides by matrix-assisted laser desorption/ionization mass spectrometry: use of alkylphosphonic acids as matrix additives." Rapid Commun Mass Spectrom **22**(8): 1109-16.

## **L.**

Larochelle, S., Chen, J., Knights, R., Pandur, J., Morcillo, P., Erdjument-Bromage, H., Tempst, P., Suter, B. and Fisher, R. P. (2001)."T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity." Embo J **20**(14): 3749-59.

Larsen, M. R., Sorensen, G. L., Fey, S. J., Larsen, P. M. and Roepstorff, P. (2001)."Phosphoproteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis." Proteomics **1**(2): 223-38.

Liao, P. C., Leykam, J., Andrews, P. C., Gage, D. A. and Allison, J. (1994)."An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry." Anal Biochem **219**(1): 9-20.

Lolli, G., Lowe, E. D., Brown, N. R. and Johnson, L. N. (2004)."The crystal structure of human CDK7 and its protein recognition properties." Structure **12**(11): 2067-79.

## **M.**

Mann, M., Ong, S. E., Gronborg, M., Steen, H., Jensen, O. N. and Pandey, A. (2002)."Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome." Trends Biotechnol **20**(6): 261-8.

Merrick, B. A., Zhou, W., Martin, K. J., Jeyarajah, S., Parker, C. E., Selkirk, J. K., Tomer, K. B. and Borchers, C. H. (2001)."Site-specific phosphorylation of human p53 protein determined by mass spectrometry." Biochemistry **40**(13): 4053-66.

Meyer, H. E., Hoffmann-Posorske, E., Korte, H. and Heilmeyer, L. M., Jr. (1986)."Sequence analysis of phosphoserine-containing peptides. Modification for picomolar sensitivity." FEBS Lett **204**(1): 61-6.

Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J. and Hunt, D. F. (2006)."The utility of ETD mass spectrometry in proteomic analysis." Biochim Biophys Acta **1764**(12): 1811-22.

## **N.**

Nita-Lazar, A., Saito-Benz, H. and White, F. M. (2008)."Quantitative phosphoproteomics by mass spectrometry: past, present, and future." Proteomics **8**(21): 4433-43.

Nuhse, T. S., Stensballe, A., Jensen, O. N. and Peck, S. C. (2003)."Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry." Mol Cell Proteomics **2**(11): 1234-43.

**O.**

Oda, Y., Nagasu, T. and Chait, B. T. (2001). "Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome." Nat Biotechnol **19**(4): 379-82.

**P.**

Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B. and Heck, A. J. (2004). "Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns." Anal Chem **76**(14): 3935-43.

**S.**

Schmelzle, K. and White, F. M. (2006). "Phosphoproteomic approaches to elucidate cellular signaling networks." Curr Opin Biotechnol **17**(4): 406-14.

Schneider-Stock, R., Roessner, A. and Ullrich, O. (2005). "DAP-kinase--protector or enemy in apoptotic cell death." Int J Biochem Cell Biol **37**(9): 1763-7.

Schneider, E., Kartarius, S., Schuster, N. and Montenarh, M. (2002). "The cyclin H/cdk7/Mat1 kinase activity is regulated by CK2 phosphorylation of cyclin H." Oncogene **21**(33): 5031-7.

Schnölzer, M. and Lehmann, W. (1997). "Identification of modified peptides by metastable fragmentation in MALDI mass spectrometry." Int J Mass Spectrom Ion Proc **169/170**: 263-271.

Schwappe, R. E., Haydon, C. E., Lewis, T. S., Resing, K. A. and Ahn, N. G. (2003). "The characterization of protein post-translational modifications by mass spectrometry." Acc Chem Res **36**(6): 453-61.

Shohat, G., Shani, G., Eisenstein, M. and Kimchi, A. (2002). "The DAP-kinase family of proteins: study of a novel group of calcium-regulated death-promoting kinases." Biochim Biophys Acta **1600**(1-2): 45-50.

Sickmann, A. and Meyer, H. E. (2001). "Phosphoamino acid analysis." Proteomics **1**(2): 200-6.

Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N. and Kirschner, M. W. (2006). "Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements." Mol Cell Proteomics **5**(1): 172-81.

Steen, H., Kuster, B., Fernandez, M., Pandey, A. and Mann, M. (2002). "Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway." J Biol Chem **277**(2): 1031-9.

Strupat, K., Karas, M. and Hillenkamp, F. (1991). "2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry." Int J Mass Spectrom Ion Proc **111**: 89-102.

**T.**

Tao, W. A., Wollscheid, B., O'Brien, R., Eng, J. K., Li, X. J., Bodenmiller, B., Watts, J. D., Hood, L. and Aebersold, R. (2005). "Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry." Nat Methods **2**(8): 591-8.

Tereshko, V., Teplova, M., Brunzelle, J., Watterson, D. M. and Egli, M. (2001). "Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression." Nat Struct Biol **8**(10): 899-907.

Thingholm, T. E., Jensen, O. N., Robinson, P. J. and Larsen, M. R. (2008). "SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides." Mol Cell Proteomics **7**(4): 661-71.

Tholey, A. and Heinze, E. (2006). "Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives." Anal Bioanal Chem **386**(1): 24-37.

Thompson, A. J., Hart, S. R., Franz, C., Barnouin, K., Ridley, A. and Cramer, R. (2003). "Characterization of protein phosphorylation by mass spectrometry using immobilized metal ion affinity chromatography with on-resin beta-elimination and Michael addition." Anal Chem **75**(13): 3232-43.

Tsumoto, H., Ra, M., Samejima, K., Taguchi, R. and Kohda, K. (2008). "Chemical derivatization of peptides containing phosphorylated serine/threonine for efficient ionization and quantification in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **22**(7): 965-72.

#### **W.**

Walsh, C. T., Garneau-Tsodikova, S. and Gatto, G. J., Jr. (2005). "Protein posttranslational modifications: the chemistry of proteome diversifications." Angew Chem Int Ed Engl **44**(45): 7342-72.

Wolschin, F. and Weckwerth, W. (2006). "Methionine oxidation in peptides--a source for false positive phosphopeptide identification in neutral loss driven MS3." Rapid Commun Mass Spectrom **20**(16): 2516-8.

Wolschin, F., Wienkoop, S. and Weckwerth, W. (2005). "Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC)." Proteomics **5**(17): 4389-97.

#### **Y.**

Yu, Y., Hoffhines, A. J., Moore, K. L. and Leary, J. A. (2007). "Determination of the sites of tyrosine O-sulfation in peptides and proteins." Nat Methods **4**(7): 583-8.

#### **Z.**

Zhou, H., Watts, J. D. and Aebersold, R. (2001). "A systematic approach to the analysis of protein phosphorylation." Nat Biotechnol **19**(4): 375-8.

Zurita, M. and Merino, C. (2003). "The transcriptional complexity of the TFIID complex." Trends Genet **19**(10): 578-84.



**Partie III- Apports de l'approche protéomique pour l'identification des protéines plasmatiques différentiellement exprimées lors du jeûne prolongé chez le manchot Adélie, un organisme dont le génome n'est pas séquencé**

**Chapitre I. Synthèse bibliographique : Stratégies et méthodes protéomiques.**

**Chapitre II. Application : Analyse différentielle du protéome plasmatique du manchot Adélie en réponse au jeûne prolongé**

**Chapitre III. Conclusion : Approche protéomique et biologie des organismes dont le génome n'est pas connu**





## **Chapitre I : Synthèse bibliographique : Stratégies et méthodes protéomique**

### **1. Identification de protéines d'organismes dont le génome n'est pas séquencé : quelle stratégie ?**

Aujourd'hui, l'achèvement du séquençage du génome d'un nombre croissant d'espèces marque le début de la meilleure compréhension de la fonction précise des gènes dans la cellule (génomique fonctionnelle). Le séquençage et l'annotation d'un nombre croissant de génomes, ainsi que la prédiction des séquences protéiques qui en dérivent, assurent un continu et rapide enrichissement des banques de données nucléotidiques et protéiques. Ainsi, depuis le séquençage des premiers génomes complets, i.e. ceux d'*Haemophilus influenzae* (1,8 Mbps) en 1995 et de la levure *Saccharomyces cerevisiae* (14 Mbps) en 1996 [Goffeau et al., 1996], une véritable course au séquençage a été lancée. La "Genomes Online Database" (GOLD ; <http://www.genomesonline.org/>), qui compile les informations relatives à tous les projets de séquençage [Liolios et al., 2008], rapporte en mai 2009 que sur plus de 4800 projets de séquençage, presque 1000 sont déjà complets et publiés, alors que parmi les quelques 3800 projets en cours plus de 2500 concernent des bactéries (procaryotes) et environ 1000 des organismes eucaryotes. Il en résulte une extraordinaire expansion des banques de données comme par exemple celles du NCBI qui réunissait, en mai 2009, plus de 5 millions de séquences géniques et près de 20 millions de séquences protéiques.

Pour un nombre de plus en plus important d'espèces, la connaissance de leurs séquences nucléotidiques et peptidiques permet le développement d'"outils moléculaires" de type sondes nucléotidiques ou anticorps, ce qui permet l'étude de nombreux mécanismes moléculaires en suivant les niveaux d'expression de nombreux gènes et/ou protéines et leurs variations en réponse à un traitement, des conditions environnementales, une maladie...

Cependant, de nombreuses questions d'ordre biologique nécessitent de pouvoir être étudiées chez des espèces dont le génome n'est pas encore séquencé. Il peut s'agir d'espèces qui, dans leur milieu naturel, vivent dans des conditions environnementales non reproductibles au laboratoire et associées à des processus physiologiques particuliers (notre étude, Partie III – Chap. II). La biologie des espèces scientifiquement (e.g. *Arabidopsis thaliana*) et agronomiquement (e.g. *Vitis vinifera*, *Triticum aestivum*) importantes nécessite aussi d'être étudiée. Pour de telles espèces, si les "outils moléculaires" ne sont pas (encore) disponibles, les approches protéomiques basées sur l'analyse par spectrométrie de masse des biomolécules sont devenues incontournables.

Par exemple, une approche différentielle basée sur la séparation des protéines par électrophorèse bidimensionnelle couplée à des analyses en spectrométrie de masse (nanoLC-MS/MS), a permis d'étudier les effets d'un herbicide de plus en plus utilisé, le flumioxazin, sur le protéome de différents tissus d'une espèce majeure dans l'agriculture européenne, *Vitis vinifera*

[Castro et al., 2005]. Ceci a pu être réalisé alors même que très peu d'informations quant aux séquences protéiques codées par le génome de cette espèce étaient disponibles. Ainsi, le flumioxazin altère l'expression de protéines de réponse au stress, notamment oxydant, et de protéines impliquées dans le processus photosynthétique et le métabolisme énergétique de la plante. L'approche protéomique a donc permis dans ce cas de préciser les mécanismes moléculaires de réponse au stress chimique, suggérant que l'herbicide pourrait agir de façon systémique dans les divers tissus de la plante, fournissant de possibles marqueurs de l'exposition à cet herbicide, et soulignant peut être la nécessité d'utiliser de tels herbicides de façon plus respectueuse de l'environnement.

Un autre exemple nous est donné par la détermination des mécanismes moléculaires associés au développement de la betterave à sucre, dont le génome n'est toujours pas séquencé [Catusse et al., 2008]. Dans ce cas, l'approche protéomique a permis, en identifiant plusieurs centaines de protéines, de reconstruire de façon détaillée les voies métaboliques de la graine de betterave en développement, et de montrer que l'activité métabolique y est compartimentée entre les racines, les cotylédons et le péricarpe, ce qui indique une répartition des différentes tâches métaboliques entre les divers tissus. Ces informations nouvelles revêtent une importance essentielle en suggérant des cibles pour l'amélioration de la qualité et de la vigueur des graines, i.e. leur capacité à produire des plantes vigoureuses, ce qui est directement lié au rendement des cultures.

Le succès de l'approche protéomique pour l'identification de protéines d'organismes dont le génome n'est pas connu repose en grande partie sur les stratégies expérimentales mises en œuvre, du point de vue de la préparation des échantillons, des analyses en spectrométrie de masse et des modes d'interprétation des données ainsi acquises. Ainsi, la préparation des échantillons doit permettre de "décomplexifier" au maximum les extraits protéiques, non seulement pour améliorer la détection des espèces protéiques dans le spectromètre de masse, mais aussi pour en simplifier l'analyse (Voir Introduction bibliographique – Chap. II – I.1.). Par définition, les organismes dont le génome n'est pas séquencé ne sont pas ou que peu représentés dans les banques protéiques. Dans ce contexte, les recherches basées sur des similarités strictes de masse entre données expérimentales et données théoriques (approche "non tolérantes aux erreurs", voir Introduction bibliographique – Chap. I – 1.c.2) requièrent la présence dans ces banques de données de séquences identiques à celles de l'espèce étudiée. Dans la plupart des cas, on s'attend plutôt à ce que bon nombre des séquences protéiques de l'organisme non séquencé qui fait l'objet d'une analyse protéomique soient sensiblement différentes (mutations, inversion, délétion...) de celles de leurs homologues chez d'autres espèces qui sont déjà implémentées dans les banques de données. L'information de séquence apportée par la MS/MS devient alors indispensable, et les approches par LC-MS/MS doivent être privilégiées. Les modes d'interprétation des données de spectrométrie de masse basés sur des homologies de séquences non obligatoirement strictes entre données expérimentales et données théoriques (approches "tolérantes aux erreurs", voir Introduction bibliographique – Chap. I – 2.b et 2.c.) peuvent alors être privilégiés.

La combinaison et la complémentarité de ces diverses approches sont souvent la clé du succès pour l'identification des protéines d'une espèce dont le génome n'est pas connu. Après une première étape de recherches PFF à l'aide d'un algorithme de type Mascot puis une seconde étape à l'aide du mode "tolérant aux erreurs" que propose certains algorithmes comme par exemple Mascot (mode "error-tolerant") ou une approche de séquençage *de novo*, plus chronophage, laborieuse et requérant un grand savoir faire et une très bonne connaissance des règles de fragmentation en spectrométrie de masse (Voir Introduction bibliographique – Chap. I –1.b.1) peut alors être entreprise. Dans tous les cas, la présence dans les banques de séquences d'organismes phylogénétiquement proches de celui étudié peut constituer une aide non-négligeable.

## **2. Intérêts et limitations des approches protéomiques à partir d'échantillons plasmatiques**

La réponse des diverses cellules de l'organisme à une situation physiopathologique telle que, par exemple, la malnutrition, la dénutrition ou le jeûne prolongé peut être définie par l'ensemble des gènes qui s'y expriment à un instant donné, i.e. par son transcriptome (ensemble des ARNm) et par son protéome (ensemble des protéines). Les protéines d'un fluide biologique, tel le plasma, qui est dépourvu de génome comme de transcriptome spécifique, sont classiquement considérées comme reflétant l'ensemble des productions des tissus présents au sein d'un organisme. Dans ce contexte, le plasma et/ou le sérum ont été considérés comme les échantillons les mieux adaptés pour la découverte de biomarqueurs [Hu et al., 2006; Rajcevic et al., 2009].

Un biomarqueur peut être défini comme étant une variation détectable et informative sur un processus biologique donné, ce qui est essentiel d'un point de vue clinique mais aussi en regard de la recherche fondamentale en biologie et en médecine. Les biomarqueurs rassemblent donc toutes les molécules, dont les peptides et les protéines, dont les variations de niveaux indiquent une maladie particulière (diagnostic, pronostic) et/ou permettent de mieux comprendre les mécanismes moléculaires possiblement impliqués dans des conditions physiopathologiques données. Le maigre succès dans la découverte et la validation de biomarqueurs d'utilité clinique a donné lieu à la proposition de stratégies alternatives, comme par exemple l'introduction d'une étape initiale de recherche des biomarqueurs au niveau cellulaire ou tissulaire, ou encore, si un biofluide doit être analysé, l'implémentation de stratégies hautement sélectives de sélection (séparations, enrichissements) des protéines [Lescuyer et al., 2007]. Néanmoins, la protéomique des fluides corporels continue de produire de nombreuses avancées [Good et al., 2007].

L'intérêt de l'étude du protéome et du peptidome plasmatiques repose essentiellement sur la possibilité de mettre en évidence, à partir d'un seul compartiment corporel dont le prélèvement est facile, des signaux répondant à des conditions pathophysiologiques données, et :

- émis à partir de sources distinctes et donc caractérisant la réponse globale d'un organisme,
- directement disponibles pour agir à distance,

- dont les effets pourraient être complémentaires dans une situation physiologique donnée.

Le caractère essentiel de telles études est par exemple souligné par les efforts internationaux s'appliquant à caractériser le plasma et le sérum dans le cadre du "Plasma Proteome Project" (PPP) qui est organisé par la "Human Proteome Organization" (HUPO). Le but de ces études est de mettre au point des méthodes d'analyses fiables, efficaces et reproductibles pour l'identification des protéines circulantes humaines [Omenn, 2004; Omenn et al., 2005; Omenn et al., 2009].

Ainsi, la première phase de ce projet a :

i) évalué les avantages et les limitations de nombreuses techniques de déplétion, de fractionnement, et analyses par spectrométrie de masse,

ii) comparé les prélèvements de plasma sur différents anticoagulants,

iii) créé une base de données accessible au public et réunissant l'ensemble des données d'identifications obtenues : PRIDE ("PRoteomics IDentifications database" ; [www.ebi.ac.uk/pride](http://www.ebi.ac.uk/pride)) [Martens et al., 2005; Jones et al., 2006; Jones et al., 2008].

Les recommandations issues de cette première phase sont de privilégier les échantillons plasmatiques collectés sur EDTA par rapport aux échantillons sériques. Pour améliorer la résolution, la sensibilité et la reproductibilité dans l'identification des peptides et des protéines, la combinaison de différentes technologies de déplétion, de fractionnement, et par spectrométrie de masse est recommandée, avec la définition de critères stringents d'évaluation des spectres de masse. En résumé, cette phase a jeté les fondations pour le développement et la validation de "biomarqueurs" circulants dans le domaine de la biologie et de la médecine.

Les objectifs de la seconde phase du PPP sont maintenant :

i) de stimuler la soumission de données protéomiques de qualités, obtenues à partir d'échantillons de plasma humain,

ii) d'établir des ressources bioinformatiques robustes, impliquant PRIDE et le PeptideAtlas [Deutsch et al., 2005; Zhang et al., 2008],

iii) de collaborer avec d'autres initiatives de la HUPO afin de faire de l'approche protéomique des échantillons plasmatiques la voie commune de découverte de "biomarqueurs".

Le succès des analyses protéomiques à partir d'échantillons plasmatiques est conditionné par l'attention particulière qui doit être portée sur les étapes de préparation des échantillons, puis après découverte d'un candidat biomarqueur, sur sa validation à partir de cohortes de patients par exemple [Conrads et al., 2006; Good et al., 2007; Gerszten et al., 2008].

En effet, les extraits protéiques d'origine biologique sont par nature hautement complexes, comprenant plusieurs milliers de protéines et peptides (quantités de l'ordre de  $10^{10}$  -  $10^{12}$ ). Or, les instruments actuels utilisés en protéomique ne permettent d'analyser que des protéines présentes dans une gamme de quantité de l'ordre de  $10^3$  -  $10^4$ . Par ailleurs, toutes les protéines contenues dans

un extrait d'origine biologique tel le plasma ne sont pas présentes en quantité comparable. Seules quelques protéines sont majoritairement abondantes. Par exemple :

- l'albumine représente jusqu'à 50% des protéines plasmatiques totales [Anderson & Anderson, 2002],
- les 99% des protéines plasmatiques totales ne sont constitués que des 22 protéines les plus abondantes.

Ainsi, les concentrations sont extrêmement hétérogènes entre les protéines majoritaires (albumine :  $35-50 \text{ mg.ml}^{-1}$ ) et minoritaires (interleukine :  $6-10 \text{ pg.ml}^{-1}$ ) d'un fluide biologique comme le plasma. Si de nombreuses méthodes permettent l'étude de l'expression des protéines majoritaires, le défi actuel est de pouvoir étudier les composés minoritaires contenus dans ce type d'extraits [Adkins et al., 2002; Tirumalai et al., 2003]. Un pré-fractionnement doit donc être réalisé préalablement à toute analyse protéomique et/ou peptidomique pour éliminer sélectivement les protéines majoritaires et séparer au mieux les espèces protéiques d'un extrait dans le but de détecter plus de composés minoritaires potentiellement d'intérêt.

### **3. La préparation des échantillons, une étape incontournable des approches protéomiques**

Aujourd'hui, des approches très variées peuvent être utilisées pour éliminer les protéines majoritaires qui risquent de masquer la détection des composés plus minoritaires, et pour fractionner les extraits protéiques. Ceci est généralement réalisé grâce à des techniques d'immuno-déplétion. L'élimination sélective des protéines majoritaires d'un extrait protéique complexe utilise ici des anticorps dirigés spécifiquement contre ces protéines. Il existe divers kits de déplétion [Bjorhall et al., 2005] permettant la déplétion de l'albumine plasmatique ou la déplétion de plusieurs protéines majoritaire du plasma. Ainsi, il existe des colonnes "multi-affinité" permettant d'éliminer jusqu'à 14 des protéines majoritaires du plasma (~92% de déplétion des protéines totales du plasma) (colonnes MARS, Agilent Technologies) [Zolotarjova et al., 2008]. Des développements ont aboutit récemment à la commercialisation de kits d'immuno-déplétion qui permettent d'éliminer 20 protéines plasmatiques parmi les plus abondantes chez l'homme (jusqu'à 97-99% des protéines totales, ProteoPrep 20, Sigma Aldrich).

L'élimination des protéines majoritaires de haut poids moléculaire peut aussi être obtenue par filtration sur membrane, qui est basée sur la rétention sur une membrane des protéines dont le poids moléculaire dépasse une limite donnée, alors que les protéines dont le poids moléculaire est inférieur à cette limite seront éluées. L'efficacité d'une telle méthode est notamment illustrée par des résultats montrant qu'après ultrafiltration il est possible de détecter 5 à 20 fois plus de peptides sériques [Orvisky et al., 2006].

Enfin, par l'ajout de solvants, les protéines plasmatiques majoritaires de haut poids moléculaire précipitent plus facilement que des protéines plus légères et moins abondantes, ce qui permet d'enrichir l'extrait en protéines plus faiblement abondantes. L'ajout de solvant permet aussi de rompre certaines interactions entre les protéines faiblement et fortement abondantes, ce qui libère davantage de protéines faiblement abondantes. Kay *et al.* ont montré qu'une précipitation avec l'acétonitrile des protéines du sérum permet d'éliminer approximativement 99.6% des protéines, avec une élimination complète des protéines ayant une masse moléculaire supérieure à 75kDa et une grande partie de l'albumine [Kay *et al.*, 2008].

Cependant, la précipitation des protéines est une méthode peu reproductible et qui est souvent associée à d'autres méthodes de fractionnement pour augmenter son efficacité [Michopoulos *et al.*, 2009]. Parmi toutes ces méthodes de déplétion des protéines majoritaires, l'immuno-déplétion est la plus utilisée du fait de la spécificité, reproductibilité et efficacité qu'elle procure en général.

Pourtant, l'immuno-déplétion est dans une large mesure impuissante lorsqu'il s'agit d'éliminer les protéines majoritaires à partir d'un échantillon collecté chez une espèce pour laquelle les anticorps d'intérêt ne sont pas disponibles, et à plus forte raison pour les organismes dont le génome n'a pas encore été séquencé. Dans ce cas, la méthodologie "Equalizer beads" semble des plus prometteuses [Righetti *et al.*, 2006; Sennels *et al.*, 2007]. Elle utilise des ligands hexapeptidiques greffés sur des billes, chaque ligand étant présent en nombre équivalent mais limité. Il y aurait statistiquement des sites de liaison pour toutes les protéines d'un échantillon plasmatique. Le nombre limité des sites de liaison sera responsable d'une compression de la gamme dynamique des concentrations protéiques. Une grande partie des protéines abondantes est éliminée et les protéines faiblement abondantes sont enrichies, alors même que des représentants de toutes les protéines sont retenus dans l'échantillon. Ce nivellement des abondances des protéines est directement responsable de l'enrichissement en protéines faiblement abondantes dans l'échantillon. L'efficacité de cette méthode est notamment illustrée par des résultats montrant que par l'utilisation des billes "Equalizer beads" il est possible d'identifier des protéines plasmatiques humaines présentes dans des concentrations de l'ordre du µg/l au g/l, montrant ainsi la possibilité par cette technologie de détecter des protéines fortement mais également faiblement abondantes [Sihlbom *et al.*, 2008]. Même si l'on doit rester prudent vis-à-vis de cette méthode puisque des doutes ont été émis quant à sa reproductibilité, il a tout de même été montré que son utilisation permet d'obtenir des données quantitatives fiables à partir de protéines modèles [Roux-Dalvai *et al.*, 2008]. Cette même démonstration à partir d'échantillons biologiques complexes n'a pas encore été produite.

L'élimination des protéines majoritaires d'un mélange complexe permet en général une meilleure détection des protéines peu abondantes en électrophorèse bidimensionnelle et en chromatographie liquide [Zolotarjova *et al.*, 2008]. Cependant, une des limitations reste l'élimination aspécifique de certaines protéines qui peut venir fausser les résultats [Granger *et al.*, 2005]. Certaines protéines faiblement abondantes peuvent en effet être éliminées de l'échantillon lorsqu'elles sont par exemple en interaction avec des protéines telle que l'albumine [Granger *et al.*, 2005]. Pour limiter ces

inconvenients, il est recommandé de rompre les interactions hydrophobes entre albumine et autres protéines en travaillant dans des conditions dénaturantes avec par exemple l'ajout de solvant organique [Tirumalai et al., 2003; Harper et al., 2004; Merrell et al., 2004].

Les méthodes de fractionnement des extraits protéiques rassemblent différentes techniques séparatives, aussi bien électrophorétiques [Righetti et al., 2005] que chromatographiques [Lescuyer et al., 2004; Linke et al., 2004; Baumann et al., 2005].

L'électrophorèse bidimensionnelle (2-DE) et l'approche 2D-DIGE (Difference Gel Electrophoresis) sont considérées comme de puissantes méthodes qui permettent de séparer plusieurs centaines (milliers) de protéines simultanément, et d'en déterminer les niveaux d'expression dans les fluides biologiques [Gorg et al., 2004] (voir Introduction bibliographique – Chap. II – 1.a). Les gels d'électrophorèses permettent de réaliser des séparations reproductibles et l'analyse des images de gels issus de deux échantillons différents permet de quantifier les variations d'expression protéiques. Cette quantification est réalisée par comparaison entre les intensités de spots détectés sur les gels à l'aide de logiciels dédiés (voir Partie III – Chap. I - 4). L'efficacité de ces méthodes, combinées à des analyses en spectrométrie de masse, a notamment permis de montrer les variations du niveau d'expression de plusieurs protéines plasmatiques [Bertile et al., 2009], et de plusieurs protéines intestinales [Lenaerts et al., 2006] chez des animaux soumis à un jeûne prolongé. Ainsi, la réduction marquée des niveaux circulants d'apolipoprotéine A-IV en P3 a été supposée comme possiblement impliquée dans un signal d'alarme qui induit les animaux à se réalimenter (Voir Partie III – Chap. II - 2). De plus, il a pu être déduit de ces analyses que les voies métaboliques intestinales sont sélectivement régulées en fonction des phases métaboliques. La 2-DE est donc une méthode très bien adaptée à l'étude des désordres métaboliques.

Les chromatographies sont des méthodes de choix pour le fractionnement et la séparation de molécules biologiques. De nombreux types de fractionnements différents existent, qui dépendent du caractère physico-chimique des protéines à séparer et du type de phase utilisée pour la séparation (voir Introduction bibliographique – Chap. II – 1.b). De plus, la possibilité de coupler la chromatographie liquide à la spectrométrie de masse permet de travailler avec de faibles quantités d'échantillon en minimisant les pertes et les contaminations éventuelles.

L'approche consistant à combiner plusieurs dimensions de chromatographie liquide est aujourd'hui très populaire [Link et al., 1999]. La séparation des protéines de sérum murin par chromatographie d'échange d'anions (WAX) en couplage avec une chromatographie d'échange de cations (WCX) a par exemple permis, après digestion enzymatique de l'ensemble des fractions obtenues puis analyse des peptides par chromatographie bidimensionnelle couplée à une analyse par spectrométrie de masse (SCX/RP/MS/MS), d'identifier 12389 peptides uniques correspondant à 4567 protéines [Hood et al., 2005]. La combinaison de deux dimensions de chromatographie de phase inverse à différents pH permet aussi d'améliorer le fractionnement des protéines dans des extraits

protéiques complexes (voir Introduction bibliographique – Chap. I – 2.a). L'efficacité d'une telle séparation a été illustrée par l'identification, après une première dimension de chromatographie de phase inverse à pH 11.5 suivie d'une deuxième dimension à pH 2 couplée à la spectrométrie de masse, de 56 peptides précurseurs de neuropeptides, dont 17 nouveaux dans des échantillons protéiques extraits de l'hypothalamus et de striatum de rat [Dowell et al., 2006].

Parmi les techniques d'extraction de peptides, l'extraction sur phase solide (SPE) représente une approche attractive [Hennion, 1999; Souverain et al., 2004; Koomen et al., 2005]. Rapide, sélective et peu consommatrice de solvants, la SPE est aujourd'hui une des méthodes privilégiées pour le prétraitement des échantillons biologiques. Le principe de base de la SPE est de piéger sélectivement les molécules d'intérêt sur un adsorbant, tout en éliminant les protéines majoritaires d'un mélange complexe et/ou les macromolécules biologiques, pour ensuite les éluer à l'aide d'un solvant spécifique. Les analyses du protéome plasmatique combinent souvent plusieurs méthodes. La combinaison de l'ultrafiltration, de la SPE et de la chromatographie liquide est un protocole efficace et reproductible qui a permis :

- i) d'obtenir plus de 99.9% de déplétion d'albumine à partir d'échantillons de plasma de rat,
- ii) d'enrichir fortement la fraction des peptides dont rapport m/z est inférieur à 3000,
- iii) de détecter plus de 450 peptides (<3000 m/z),
- iv) d'identifier des peptides peu abondants ne représentant en masse que  $10^{-6}$  % de toutes les protéines [Bertile et al., 2007].

Tout l'intérêt d'un tel protocole réside dans cette possibilité d'analyser la fraction peptidique peu abondante, puisque la majorité des protéines et peptides circulants sont faiblement concentrés.

Basée sur le principe de la SPE, une des méthodes récentes les plus attractives réside dans l'utilisation de phases à accès restreint ou RAM (Restricted Access Material) qui permettent de cibler les analytes de bas poids moléculaire [Souverain et al., 2004]. Les RAM combinent deux modes de séparation, les chromatographies d'exclusion et d'adsorption.

Les propriétés physiques et chimiques des RAM permettent :

- i) d'exclure les macromolécules en limitant leur accès aux sites d'adsorption,
- ii) de les séparer rapidement des petits analytes.

Directement intégrées dans un système HPLC, les colonnes RAM permettent de purifier en ligne les échantillons biologiques. Ceci permet leur analyse sans étape préalable de précipitation des protéines et limite donc la perte d'information due à la co-précipitation des analytes et/ou au phénomène de suppression de signal en spectrométrie de masse. Les RAM ont déjà révélé des potentialités élevées pour le fractionnement d'échantillons et l'extraction de peptides [Racaiyte et al., 2000]. L'efficacité de la méthode est notamment illustrée par des résultats montrant qu'à partir d'une solution d'albumine à 0.4  $\mu$ M les colonnes RAM permettent d'en éliminer 99.7-99.8 % [Rieux et al., 2007].



#### 4. Analyse protéomique différentielle : notion de quantification relative

La quantification des protéines contenues dans un extrait protéique est un des grands défis actuels de l'analyse protéomique [Domon & Aebersold, 2006; Bantscheff et al., 2007; Kuhner & Gavin, 2007; Tamvakopoulos, 2007].

Plusieurs stratégies par marquage isotopique ont été développées. L'idée clé est que deux formes d'une même molécule qui ne diffèrent que par leur composition isotopique vont se comporter de façon identique lors de l'analyse MS et ne seront différenciées que par un écart de masse. Ainsi, le ratio entre les deux pics correspondant à ces deux isoformes indiquera précisément les quantités relatives des deux espèces. L'incorporation d'isotopes stables (principalement  $^2\text{H}$ ,  $^{18}\text{O}$ ,  $^{13}\text{C}$  et  $^{15}\text{N}$ ) peut être réalisée de plusieurs manières:

- par réactions chimiques : stratégie ICAT (Isotope-coded affinity tag) [Gygi et al., 1999],
- par marquage métabolique utilisant des acides aminés marqués : stratégie SILAC (Stable isotope labelling with amino acids in cell culture) [Blagoev et al., 2004],
- par incorporation enzymatique d' $^{18}\text{O}$  durant la protéolyse [Mirgorodskaya et al., 2000; Fenselau & Yao, 2007].

La quantification relative peut également être réalisée par marquage des extrémités N-terminales et des chaînes latérales des lysines de tous les peptides par réactions chimiques à l'aide de 4 réactifs isobares (méthode iTRAQ). La quantification est basée sur l'utilisation des ions diagnostics obtenus après fragmentation [Ross et al., 2004].

D'autres stratégies de quantification ont récemment été développées, dites sans marquage ou méthodes "label-free". Ces méthodes regroupent essentiellement :

- le comptage du nombre de spectres ("spectral counting"), qui suppose que plus une protéine est abondante dans l'échantillon, plus le nombre de spectres permettant son identification sera grand [Liu et al., 2004; Sadygov et al., 2004],
- l'analyse de l'abondance des peptides à partir des chromatogrammes d'ions extraits correspondants aux peptides identifiés [Higgs et al., 2005],
- la mesure du courant d'ion total (TIC) moyen de tous les spectres MS/MS qui ont permis l'identification d'une protéine [Asara et al., 2008].

La quantification des protéines peut également être réalisée par une approche différentielle basée sur leur séparation par électrophorèse bidimensionnelle. La détermination des variations d'expression protéique associées à un traitement ou une situation physiopathologique donnée est dans ce cas basée sur l'analyse de l'intensité ou de la densité de la coloration des spots protéiques, à l'aide de logiciels experts, e.g. PDQuest (BioRad), SameSpot (Nonlinear Dynamics), Melanie (Swiss Institute of Bioinformatics)... [Rosengren et al., 2003; Palagi et al., 2006]. Cependant, certains

colorants manquent de sensibilité et ne permettent pas d'obtenir des données quantitatives (voir Introduction bibliographique – Chap. II – 1.a). La méthode la plus couronnée de succès est aujourd'hui la méthode 2D DIGE ("two-dimensional difference gel electrophoresis"). En règle générale, il est nécessaire de réaliser des répliquats techniques et biologiques, afin de s'affranchir de toute variation non induite par le traitement étudié. Les répliquats biologiques sont essentiels quelle que soit la méthodologie employée, tandis que les répétitions techniques permettent de montrer la reproductibilité de l'expérience et sont surtout requises dans une méthodologie non "DIGE".

L'analyse des images de gels 2D comportent plusieurs étapes [Gorg et al., 2004; Berth et al., 2007]. Après élimination du bruit de fond et détection automatique des spots protéiques, l'image doit être "nettoyée", souvent manuellement, en fonction du logiciel utilisé, afin d'éliminer tout "spot" qui ne serait pas protéique (marqueur de poids moléculaire, tâche de colorants...). A l'inverse, certains spots non détectés devront être rajoutés manuellement. Il s'agit ensuite de "recaler" les gels les uns entre les autres afin que les spots protéiques correspondants soient marqués comme tels ("matching"). Bien qu'automatisée, cette étape nécessite aussi une intervention manuelle, surtout si une distorsion importante s'est produite sur un des gels par rapport aux autres.

L'étape suivante consiste en l'intégration automatisée de l'intensité ou de la densité de la coloration de chacun des spots sur chacun des gels. Les valeurs obtenues sont normalisées, surtout afin de corriger les variations de quantité de protéines déposées, l'efficacité de coloration... Cette normalisation est très importante et peut être réalisée par rapport à l'intensité moyenne de tous les spots détectés sur un gel (méthode la plus fréquemment employée) ou à partir de ratios entre l'intensité d'un spot protéique sur un gel donné et celle du spot correspondant sur un gel de référence ("mastergel") qui est une image de gel fictif représentant le profil commun à tous les gels. Après comparaison des répliquats technique d'un point de vue statistique, la comparaison va concerner les différences entre les groupes qui sont dues à l'effet du traitement ou de la situation étudiés. Dans ce cas, selon le nombre de protéomes qui sont comparés, on utilisera essentiellement le test t de Student ou l'analyse de variance (ANOVA) suivie d'un test post-hoc de type test de Tukey.

## **Chapitre II : Application : Analyse différentielle du protéome plasmatique du manchot Adélie en réponse au jeûne prolongé**

### **1. Contexte scientifique général**

La masse et la composition corporelle d'un animal peuvent varier considérablement au cours de son cycle annuel en fonction de contraintes écophysiologiques comme la reproduction, l'hibernation et la mue. Le stockage des réserves énergétiques et leur mobilisation ultérieure sont donc des étapes essentielles à la survie de l'animal dans son milieu naturel, puisqu'elles peuvent subvenir aux besoins de l'organisme dans des situations où les apports alimentaires sont espacés dans le temps.

Les réserves énergétiques sont constituées de glucides et surtout de lipides. Les glucides, qui sont stockés sous forme de glycogène dans le foie et le muscle, constituent une réserve de glucose qui offre l'avantage de pouvoir être mobilisée très rapidement. Cette réserve est cependant épuisée en seulement quelques heures. Les triacylglycérols (triesters d'acides gras et de glycérol) constituent la principale forme de stockage à long terme des lipides dans le tissu adipeux blanc.

A long terme, la masse et la composition corporelle sont peu variables. La balance énergétique est en effet finement régulée, les ajustements de la prise alimentaire et de la dépense énergétique conditionnant le niveau des réserves. Toute modification durable de cet état d'équilibre aboutit, selon la situation, à l'obésité ou à l'amaigrissement de l'individu. Cependant, les mécanismes de régulation de la balance et des réserves énergétiques sont encore largement méconnus.

### **2. Le modèle du jeûne prolongé : adaptations métaboliques**

Le suivi de l'évolution de la masse corporelle permet de diviser le jeûne prolongé en trois phases métaboliques distinctes (Figure 1) [Le Maho et al., 1976; Goodman et al., 1980; Robin et al., 1988]. Ces mêmes trois phases sont retrouvées chez tous les organismes privés de nourriture, et ceci à l'échelle de temps près.

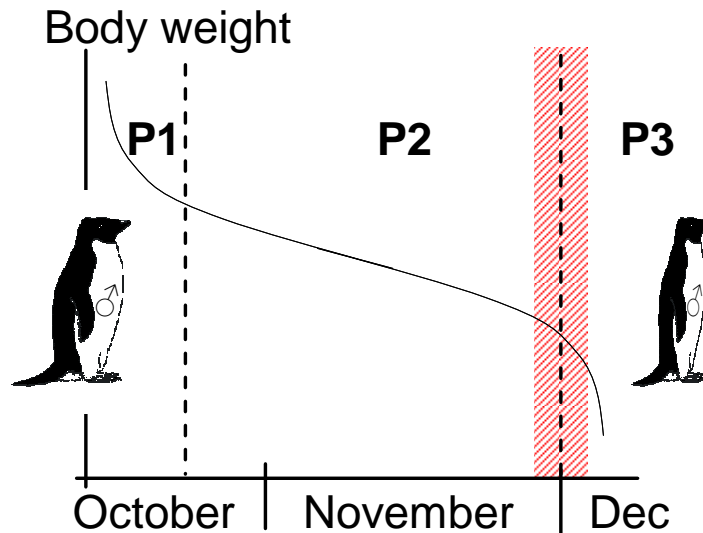


Figure 1 : Evolution de la masse corporelle au cours du jeûne prolongé.

La phase 1 (P1), dite d'adaptation, est courte et caractérisée par une diminution rapide de la masse corporelle. L'organisme utilise alors principalement ses réserves glucidiques, qui auront été totalement épuisées à l'entrée en phase 2 (P2). La P1 est de plus marquée par l'augmentation de la mobilisation des réserves lipidiques et la diminution de l'utilisation des protéines. La vidange du tractus digestif contribue également à la forte perte de masse.

La P2 est caractérisée par une diminution lente et régulière de la masse corporelle. Les lipides de réserve assurent alors la plus grande part de la dépense énergétique, une part minime provenant des protéines tissulaires [Belkhou et al., 1991; Robin et al., 1998].

L'atteinte d'un seuil critique de déplétion lipidique correspond à l'entrée en phase 3 (P3) [Goodman et al., 1980; Robin et al., 1998]. Il reste alors 20 à 25% du stock lipidique initial. L'utilisation des protéines corporelles est alors augmentée et celle des lipides diminuée, la masse corporelle diminuant en conséquence à nouveau rapidement. Cette situation est certes une étape critique, mais elle est réversible puisqu'à la réalimentation on assiste à une restauration des réserves [Belkhou et al., 1991; Cherel & Le Maho, 1991].

A la transition entre la P2 et la P3, les animaux deviennent beaucoup plus actifs [Koubi et al., 1991; Robin et al., 1998]. L'abandon spontané d'œufs en incubation a également été rapporté chez le manchot royal [Olsson, 1997]. Ces changements comportementaux sont supposément activés par un "signal d'alarme" endogène, qui induit les animaux à se réalimenter avant une déplétion létale de leur réserves énergétiques [Le Maho et al., 1988; Robin et al., 1998]. La nature et l'origine de ces signaux sont encore inconnues.

### 3. Le cas du manchot Adélie

De nombreuses espèces animales peuvent être amenées à jeûner de façon spontanée pendant des périodes prolongées. C'est notamment le cas lorsque la disponibilité en nourriture est limitée ou lorsque la recherche alimentaire entre en compétition avec d'autres activités, comme par exemple la migration ou la reproduction [Mrosovsky & Sherry, 1980]. Le jeûne à long terme se rencontre communément chez des espèces d'oiseaux marins en période de reproduction. Les manchots par exemple, qui se nourrissent exclusivement en mer, doivent néanmoins rester sur le continent pour de longues périodes lorsqu'ils se reproduisent, particulièrement durant l'incubation de leurs œufs [Williams, 1995].

Le manchot Adélie (*Pygoscelis adeliae*), fait partie des trois espèces de manchots (avec le manchot à jugulaire *Pygoscelis antarctica* et le manchot empereur *Aptenodytes forsteri*) qui ont colonisé les côtes du continent Antarctique. Il vient se reproduire en bordure du continent pendant l'été austral. Sur l'archipel de Pointe Géologie en Terre Adélie, les mâles sont les premiers à arriver sur la colonie mi-octobre, suivis des femelles environ une semaine plus tard (Figure 2).

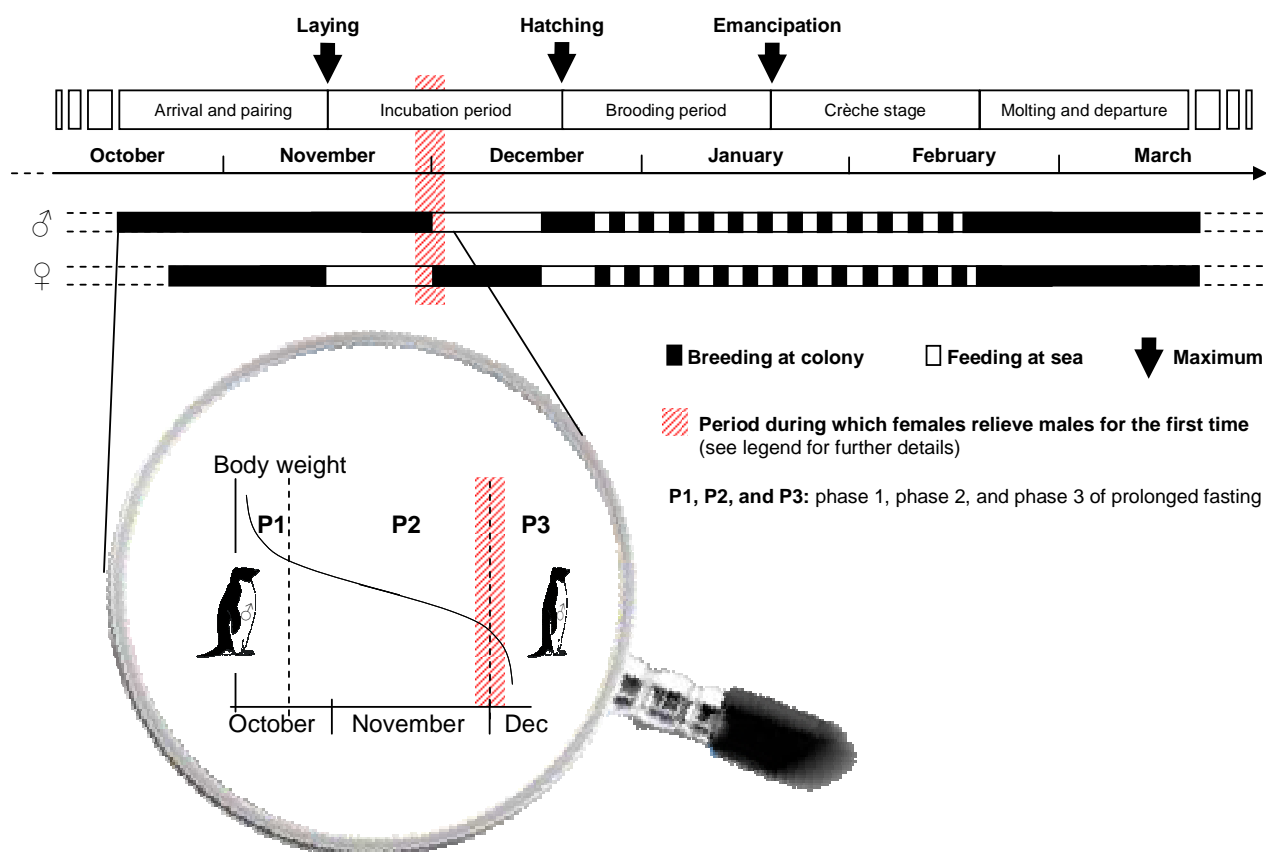


Figure 2 : Cycle de reproduction du manchot Adélie à Dumont d'Urville avec les alternances de séjours à terre et en mer.

Après s'être accouplés, dans la première quinzaine de novembre, la femelle pond un à deux œufs puis elle retourne s'alimenter en mer, et confie les œufs au mâle qui les couve pendant environ 2 semaines. Le mâle assurant la première partie de l'incubation, il est donc amené au total à jeûner 35-40 jours en moyenne. La femelle, de retour sur la colonie va alors relayer, puis les deux membres du couple vont ainsi alterner les voyages alimentaires en mer et les séjours à terre pour couvrir puis, à partir de mi-décembre, nourrir les poussins. Les voyages alimentaires post-éclosion sont courts, de 1 à 2 jours. Enfin, en février, les jeunes manchots qui ont été abondamment et régulièrement alimentés, ont terminé leur croissance. Les parents les laissent alors à leur propre destin et ils n'ont plus qu'à attendre la fin de leur mue pour partir à leur tour en mer.

Ainsi, ce cycle de reproduction estival impose toute une suite de contraintes énergétiques aux adultes, qui doivent alterner des phases de jeûne et des phases d'alimentation en mer. Notre étude s'est focalisée sur la période de jeûne la plus longue du mâle (octobre-novembre) qui, lorsque la femelle arrive pour le relayer et incuber à son tour, présente un état métabolique à la transition entre la P2 et la P3 du jeûne prolongé (Figure 2).

#### **4. Problématique et objectifs de l'étude**

La nature du signal ou des signaux de réalimentation qui induisent les animaux en P3 à rechercher de la nourriture n'est pas connue. En effet, peu de molécules ont été rapportées comme susceptibles de jouer un rôle clé dans l'induction du comportement alimentaire. Brièvement, il pourrait s'agir de la corticostérone [Challet et al., 1995], de certaines protéines de la famille des sirtuines [Masoro, 2004], et de molécules directement produites par le tissu adipeux [Barzilai & Gabriely, 2001; Bertile & Raclot, 2004; Bertile & Raclot, 2007]. Cependant, le rôle "signal" de la corticostérone a été largement remis en question chez le canard Eider (*Somateria mollissima*) [Criscuolo et al., 2005] et le rôle des sirtuines n'est pas clairement établi [Masoro, 2004]. Parmi les nombreux facteurs sécrétés par le tissu adipeux et capables de réguler les réserves énergétiques de l'organisme, la leptine est un des candidats majeurs possiblement impliqué dans la réponse au jeûne prolongé [Bertile et al., 2003]. Cependant, bien qu'essentiels, les effets de la leptine n'expliquent pas à eux seuls les réponses métaboliques à la privation de nourriture.

En résumé, aucune molécule n'explique à ce jour à elle seule, ou conjointement avec d'autres, l'étiologie des transitions métaboliques associées au jeûne prolongé, ni l'induction de la réalimentation en fin de jeûne. L'identification et le rôle exact de facteurs régulateurs restent donc à déterminer. On s'attend à ce que de tels signaux soient circulants pour pouvoir informer le cerveau de la situation critique pour la survie de l'organisme dénutri. Ces signaux pourraient correspondre à des variations du niveau plasmatique des certaines protéines durant le jeûne.

Dans le contexte de la recherche des facteurs régulant la réponse au jeûne prolongé, notre travail était basé sur l'étude du manchot Adélie en période de reproduction, et plus précisément sur la première période pendant laquelle la femelle relève le mâle (Figure 2). En effet, à ce stade la biologie de cet oiseau présente une particularité unique : le statut métabolique est, dans une certaine mesure,

découplé de la réponse comportementale. En effet, l'entrée en P3 n'est pas obligatoirement associée au départ du mâle pour aller se réalimenter. Ceci est illustré sur la figure suivante (Figure 3):

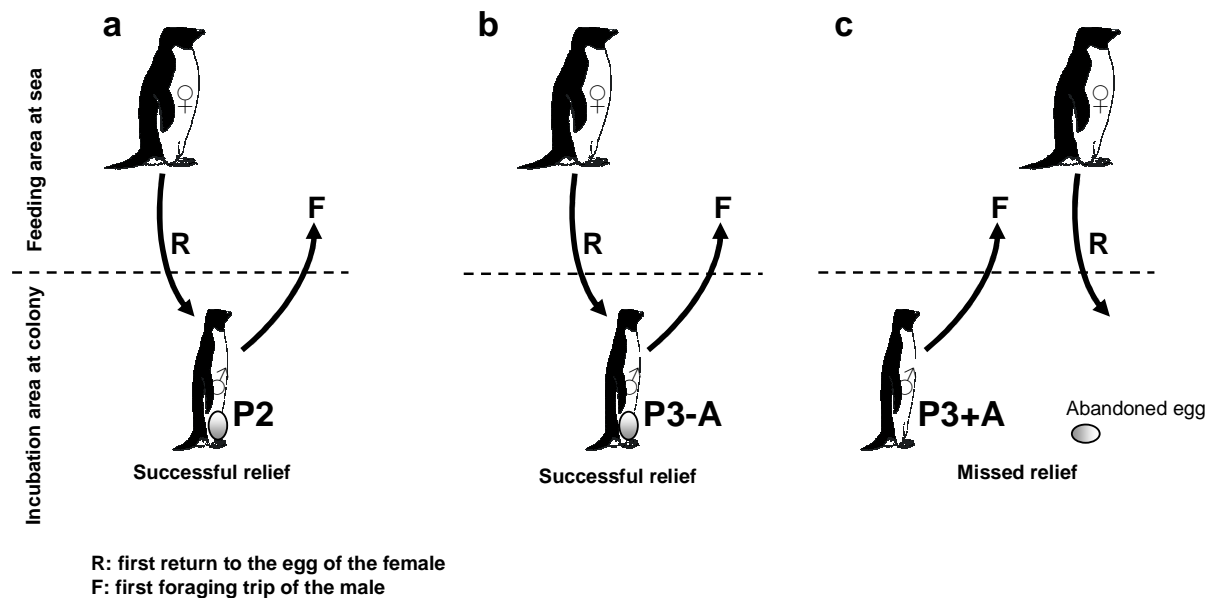


Figure 3 : 1<sup>ère</sup> relève entre mâle et femelle de manchot Adélie pour l'incubation des œufs en relation avec le métabolisme du mâle.

Vers la mi-décembre (Figure 2), la femelle relève généralement le mâle alors qu'il est en fin de phase 2 (Figure 3-a). Dans ce cas, la relève est habituellement couronnée de succès, la femelle va incuber les œufs alors que le mâle va partir en mer pour se réalimenter. Dans de rares cas, la femelle va relayer le mâle qui bien que déjà entré en P3 n'est pas encore parti se réalimenter (P3-A ; Figure 3-b). Dans ce cas, le découplage entre statut métabolique et comportement alimentaire permet la poursuite de la reproduction du couple de manchots, la femelle poursuivant l'incubation des œufs alors que le même part se réalimenter en mer. Enfin, toujours dans de rares cas, le mâle part se réalimenter en mer avant le retour de la femelle (P3+A ; Figure 3-c). Ce départ est associé à un statut métabolique qui correspond à une P3 de jeûne prolongé. Un tel couple de manchots va donc échouer dans sa reproduction.

Ces situations écopysiologiques particulières offrent un très bon modèle d'étude pour rechercher les facteurs circulants qui contrôlent les limites physiologiques du jeûne de longue durée. Les objectifs de notre étude étaient par conséquent de déterminer le protéome plasmatique différentiel de mâles de manchots Adélie P3-A, de mâles P3+A et d'animaux contrôles en P2 (Figure 3). Afin de préciser les changements du protéome plasmatique qui sont liés au statut métabolique indépendamment de l'investissement parental et donc de la motivation à incuber, nous avons tout d'abord comparé le protéome plasmatique d'animaux captifs, non incubants en P2 et en P3.

## 5. Méthodes et stratégie de notre approche protéomique

Le suivi des manchots, une partie des mesures et la collecte des échantillons ont été réalisés en antarctique, près de la base scientifique Dumont d'Urville. Après prélèvement sanguin, le plasma a été préparé et conservé dans l'azote liquide jusqu'à analyse. Seule la stratégie protéomique déployée dans cette étude sera détaillée ci-après :

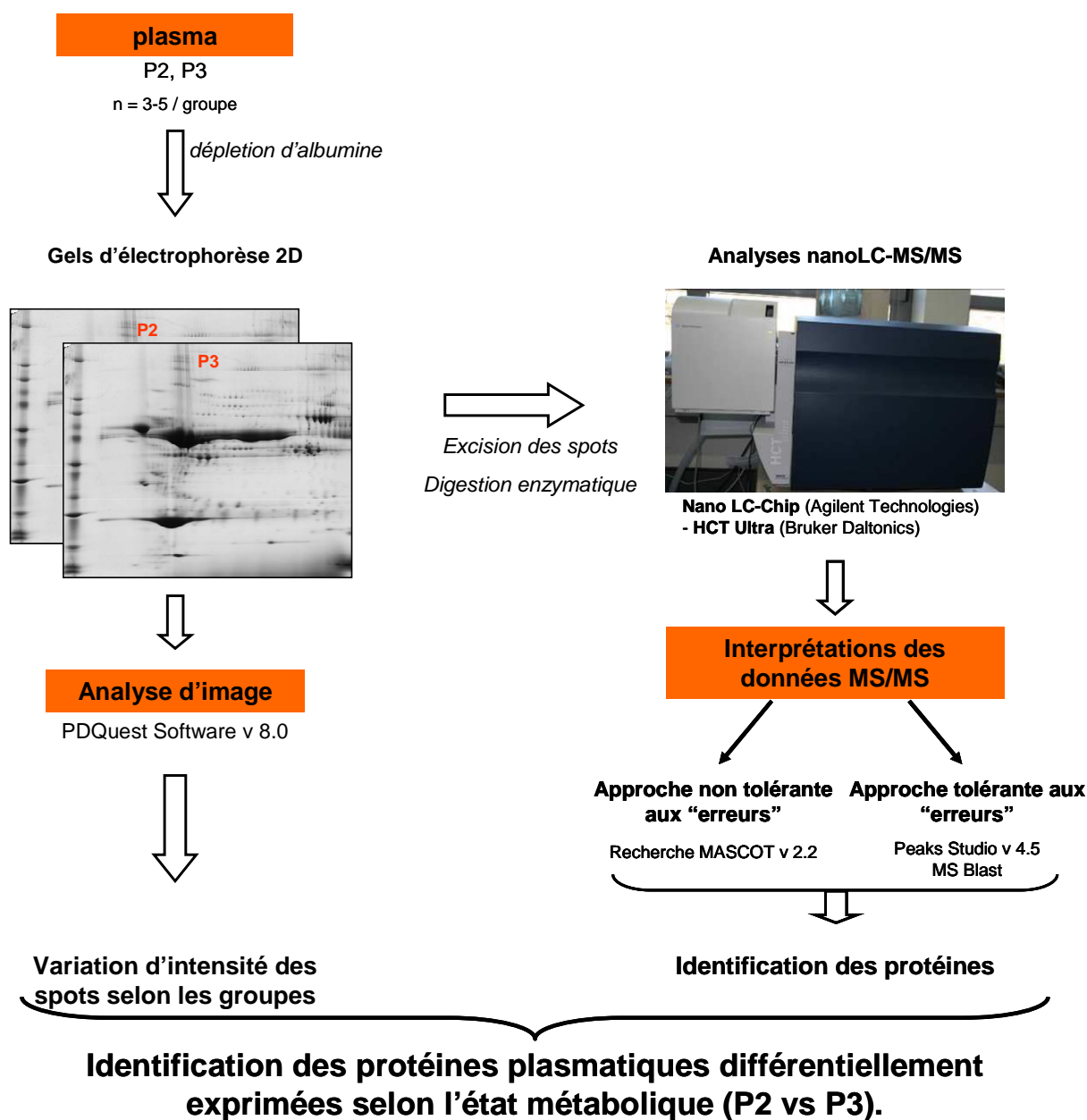


Figure 4 : Analyse protéomique différentielle du plasma de manchot Adélie selon son état métabolique au cours du jeûne d'incubation.



Après immunodéplétion de l'albumine contenue dans nos échantillons plasmatiques, les protéines ont été séparées par 2-DE et les spots protéiques ont été révélés au bleu de coomassie. Tous les gels ont été réalisés en duplicats (1 réplicat technique par échantillon), et 3 à 5 réplicats biologiques ont pu être réalisés selon le groupe considéré et les contraintes de prélèvements liées au travail sur le terrain.

Les analyses d'image ont été réalisées à l'aide du logiciel PDQuest v8.0 (Biorad). Les spots protéiques ont tout d'abord été détectés et leur intensité normalisée. Après vérification de la distribution selon une loi Normale de toutes nos valeurs, les intensités normalisées ont tout d'abord été comparées entre les réplicats techniques pour tester la reproductibilité des extraits séparés (paired-samples t test). En l'absence de différence significative entre réplicats techniques, les intensités normalisées ont été moyennées. Les différents états métaboliques (P2C vs. P3C ; P2 vs. P3-A vs. P3+A) ont ensuite été comparés à partir de ces moyennes, afin de déterminer les spots dont l'intensité variait selon l'état métabolique (t test ou test ANOVA + Tukey test).

L'identification des protéines a été réalisée à l'aide du couplage nanoLC-Chip (Agilent Technologies) – trappe ionique HCT Ultra (Bruker Daltonics). Les spots protéiques ont tout d'abord été excisés, et les protéines contenues dans ces spots ont été réduites, alkylées et digérées à la trypsine avant analyse des peptides par nanoLC-MS/MS.

Les données obtenues ont été interprétées à l'aide de l'algorithme de recherche Mascot (recherche non tolérante aux "erreurs" par similarité stricte de masse, voir Introduction bibliographique – Chap. I –1.c.2). Lorsque les recherches Mascot n'ont donné lieu à aucune identification, les données de spectrométrie de masse ont alors été interprétées par l'approche *de novo* (recherche tolérante aux "erreurs", voir Introduction bibliographique – Chap. I – 2.b). L'approche *de novo* a également été utilisée pour augmenter la couverture de séquence des protéines différentielles identifiées dans cette étude. Pour ce faire, nous avons généré une banque de données protéique restreinte à la taxonomie "Oiseaux" (téléchargée depuis le site internet de NCBI version du 02/08/2008, 160400 séquences). Afin d'augmenter la confiance dans les identifications de protéines par spectrométrie de masse, en estimant notamment un taux de faux positif, une stratégie "target-decoy" a été développée. Pour cela, une banque decoy a été créée par inversion des séquences de la banque téléchargée et concaténée avec cette même banque. Les identifications de protéines ont été réalisées à l'aide de cette nouvelle banque renommée "target-decoy".

Les recherches par similarité stricte de masse avec Mascot ont été réalisées avec les paramètres suivants :

- Enzyme de digestion Trypsine et une seule coupure manquée autorisée
- Tolérance de masse MS : 0.5 Da et MS/MS : 0.5 Da
- Modifications possibles : Carbamidomethylation sur les cystéines, oxydation sur les méthionines et acétylation en N-terminal de la protéine.

L'ensemble des résultats générés par les recherches Mascot ont été chargés dans le logiciel Scaffold (Proteome software, version 01-07-00). Ce logiciel permet de fixer des critères stringents

pour l'identification des protéines avec un taux de faux positif estimé reste faible (en général < 5 % ; < 1 % dans cette étude).

- Pour les protéines identifiées avec minimum 2 peptides: un seuil minimal de score Mascot que les scores individuels de chaque peptide doivent dépasser a été défini (> 36 dans cette étude)

- Pour les protéines identifiées avec 1 seul peptide: le score du peptide doit excéder le score "identity" Mascot, score calculé par Mascot pour identification (au risque  $p < 0.05$ ), de 13 dans cette étude.

Le séquençage de novo a été réalisé à l'aide du logiciel Peaks Studio v4.5 (Bioinformatics Solution), en retenant les mêmes paramètres que ceux utilisés pour la recherche Mascot. Les séquences déduites ont été soumises au logiciel MS-Blast (voir Introduction bibliographique – Chap. I – 2.b.2) pour des recherches par homologie de séquence dans notre banque "target-decoy". Les résultats MS-Blast ont été validés pour une séquence en acides aminés ayant un score de 64 pour 1 HSP, 102 pour 2 HSPs, 150 pour 3 HSPs et 200 pour 4 HSPs.

## 6. Résultats et discussion

### 6.a. Paramètres globaux et statut métabolique des animaux

Afin de caractériser le statut métabolique des animaux, leur masse corporelle, activité locomotrice et divers paramètres plasmatiques ont été mesurés. Les résultats de ces mesures sont reportés dans les tableaux suivants.

Tableau 1: Profils des manchots Adélie captifs en P2 et en P3 (NEFA: acides gras non esterifiés,  $\beta$ OHB:  $\beta$ -hydroxybutyrate). Les valeurs significativement différents ( $p < 0.05$ ) ne partagent pas la même lettre.

	P2C	P3C
Body mass, Kg	4.40 $\pm$ 0.01 <sup>a</sup>	3.37 $\pm$ 0.02 <sup>b</sup>
Locomotor activity, au	0.87 $\pm$ 0.08 <sup>a</sup>	1.56 $\pm$ 0.15 <sup>b</sup>
Plasma NEFA, mM	0.85 $\pm$ 0.09	0.70 $\pm$ 0.07
Plasma $\beta$ OHB, mM	1.18 $\pm$ 0.06 <sup>a</sup>	1.00 $\pm$ 0.04 <sup>b</sup>
Plasma glucose, mM	12.2 $\pm$ 0.6	12.8 $\pm$ 0.4
Plasma uric acid, mM	0.28 $\pm$ 0.03 <sup>a</sup>	1.12 $\pm$ 0.13 <sup>b</sup>

Tableau 2 : Profils des manchots Adélie en reproduction (NEFA: acides gras non esterifiés,  $\beta$ OHB:  $\beta$ -hydroxybutyrate). Les valeurs significativement différents ( $p < 0.05$ ) ne partagent pas la même lettre.

	P2	P3-A	P3+A
Body mass, Kg	4.39 $\pm$ 0.11 <sup>a</sup>	3.51 $\pm$ 0.07 <sup>b</sup>	3.40 $\pm$ 0.03 <sup>b</sup>
Plasma NEFA, mM	0.84 $\pm$ 0.09	0.81 $\pm$ 0.07	0.70 $\pm$ 0.04
Plasma $\beta$ OHB, mM	1.19 $\pm$ 0.07 <sup>a</sup>	0.74 $\pm$ 0.05 <sup>b</sup>	0.92 $\pm$ 0.16 <sup>a,b</sup>
Plasma glucose, mM	12.9 $\pm$ 0.7	13.6 $\pm$ 0.2	13.2 $\pm$ 0.9
Plasma uric acid, mM	0.25 $\pm$ 0.02 <sup>a</sup>	0.43 $\pm$ 0.04 <sup>b</sup>	0.69 $\pm$ 0.15 <sup>b</sup>

La pesée des animaux a permis de montrer que les manchots P3C avaient une masse corporelle ~ 23 % inférieure à celle des manchots P2C, et que les manchots P3-A et P3+A avaient une masse corporelle ~ 20 % et ~ 22.5 % inférieure à celle des manchots P2, respectivement. Ces diminutions de masse corporelle sont essentiellement dues à la diminution de la masse grasse, les lipides de réserves ayant été largement mobilisés pour satisfaire les besoins énergétiques de l'organisme. Chez les animaux captifs, la pose de podomètres a permis de suivre l'activité locomotrice journalière des animaux et de montrer que celle-ci est ~ 1.8 fois plus élevée chez les P3C que les P2C.

La validation du statut métabolique des animaux a été obtenue par le dosage des niveaux circulants de  $\beta$ -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. Enfin, les niveaux plasmatiques de glucose ont été mesurés. La diminution de la concentration du  $\beta$ -hydroxybutyrate et en acides gras libres entre les groupes P2C et P3C et entre les groupes P2 et P3-A/P3+A 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, aussi bien chez les P3C comparés aux P2C que chez les P3-A/P3+A comparés aux P2. Enfin, l'utilisation des substrats lipidiques et protéiques en P2 et en P3 du jeûne est accompagnée d'un maintien de la glycémie, essentiellement due à la néo-synthèse de glucose à partir des acides aminés, et à l'utilisation préférentielle de substrats alternatifs au glucose (surtout acides gras et corps cétoniques) par de nombreux tissus.

#### **6.b. Préparation des échantillons et mise au point des conditions de séparations des protéines**

Afin d'éliminer au mieux l'albumine, protéine plasmatique majoritaire, de nos échantillons de plasma de manchot et de mieux séparer et détecter les autres protéines plasmatiques, nous avons testé un kit d'immuno-déplétion utilisant des anticorps dirigés contre l'albumine humaine (Vivascience). Nous avons ainsi pu dépler environ 20 % des protéines totales, ce qui ne représente qu'une faible proportion de l'albumine totale. La qualité de cette déplétion n'a pas été améliorée par l'utilisation d'une quantité variable de résine d'affinité. Cependant, si visuellement la déplétion d'albumine n'est pas flagrante sur les images de gels, la quantité de spots détectés après coloration au bleu de Coomassie était environ deux fois plus grande après déplétion que sans déplétion (Figure 5). Nous avons également testé, en combinaison avec l'étape d'immuno-déplétion, plusieurs méthodes de précipitation des protéines et, sur la base de la qualité des images de gels d'électrophorèse bidimensionnelle que nous avons obtenues, l'utilisation de TCA/acétone (détection d'environ 400 spots par gels) s'est révélée meilleure que celle d'acétone seul (détection d'environ 300 spots par gels).

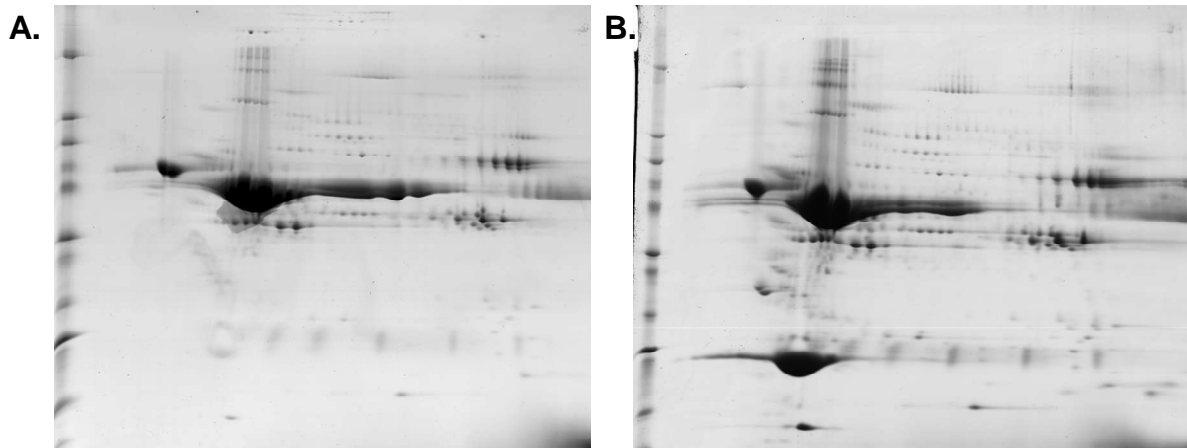


Figure 5 : Images représentatives de gels d'électrophorèse réalisés à partir d'échantillons de plasma de manchot Adélie sans (A) et après déplétion d'albumine (B) et précipitation au TCA/acétone.

Sans grande surprise, nous avons donc constaté que les anticorps utilisés dans les kits d'immuno-déplétion semblent lier l'albumine de manchot moins efficacement que l'albumine humaine, ce qui souligne les différences de séquences de cette protéine entre les espèces. Bien que largement incomplète, la déplétion de l'albumine nous a tout de même permis d'obtenir de bonnes conditions de séparation et la visualisation de nombreuses protéines.

### **6.c. Succès de l'approche protéomique pour l'identification des protéines plasmatiques de manchot Adélie**

La figure 6 présente une image représentative des protéines plasmatiques de manchot Adélie séparées sur gel 2D. L'analyse d'image (logiciel PDQuest<sup>TM</sup>, version 8.0, Biorad) a permis de détecter environ 400 spots par gel, visualisés au bleu de Coomassie.

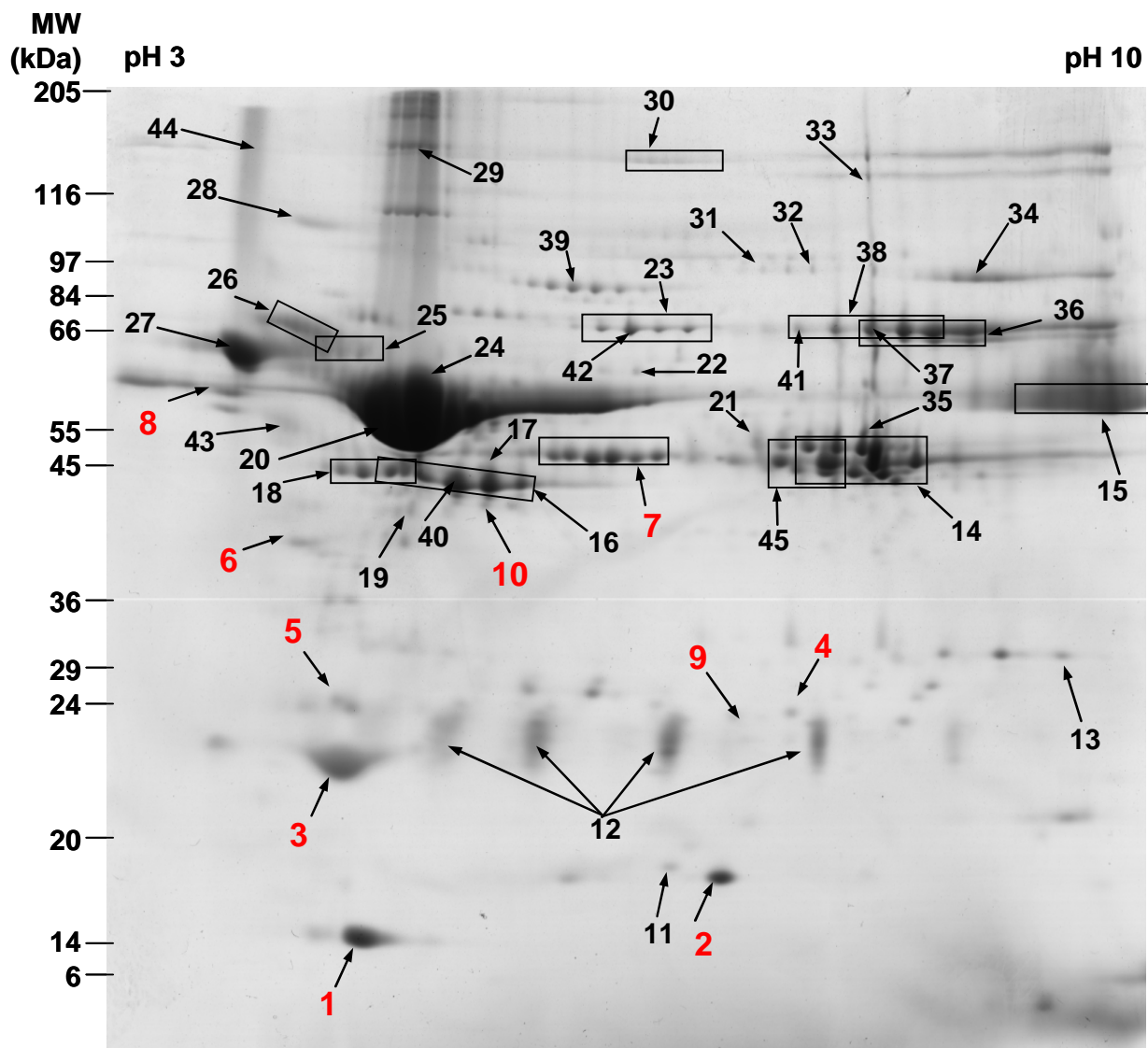


Figure 6 : Image représentative des spots excisés à partir des gels d'électrophorèse réalisés à partir d'échantillons de plasma de manchot Adélie. Les nombres font référence aux numéros des spots que l'on retrouve dans le tableau 3 et dans l'annexe 5. Les nombres en rouge indiquent les spots dont l'intensité était différentielle entre les groupes de manchots. MW : masse moléculaire.

L'analyse nanoLC-MS/MS a permis l'identification de 50 protéines aviaires uniques (Annexe 5) contenues dans 80 spots excisés, avec un taux de faux positifs inférieur à 1 %. Pour les protéines contenues dans de multiples spots sur les gels, un seul numéro est utilisé (Figure 6). L'utilisation de la banque de données restreinte à la taxonomie "Oiseaux" s'est révélée un choix judicieux puisque les séquences protéiques de 12 espèces d'oiseaux différentes nous ont permis ces identifications, même si les données présentes dans les banques sont bien entendu plus riches pour le poulet (*Gallus gallus*).

Comme on peut le constater sur la figure 6, les identifications ont pu être réalisées sur l'ensemble du gel, pour des protéines de haute comme de basse masse moléculaire, et pour des

protéines plutôt acides ou plutôt basiques. Les masses moléculaires et pI théoriques des protéines identifiées étaient en accord avec les valeurs observées sur les gels. En moyenne, les identifications ont été obtenues à partir de plus 7 peptides uniques en moyenne et pour une couverture de séquence de 15 % en moyenne, ce qui souligne l'intérêt et la puissance de l'approche protéomique pour l'étude du protéome d'organismes dont le génome n'est pas connu. Néanmoins, 18 protéines n'ont été identifiées qu'avec 1 seul peptide.

#### **6.d. Identification des protéines différentiellement exprimées entre les manchots P2C et P3C**

L'analyse d'image n'a révélé aucune différence significative entre les réplicats techniques que nous avons réalisés, ce qui montre la reproductibilité de l'expérience. Par conséquent, les valeurs d'intensité normalisées pour un spot donné ont été moyennées entre ces réplicats techniques et c'est à partir de cette nouvelle intensité moyenne que nous avons comparés les groupes d'animaux entre eux. Ainsi, l'intensité de 10 spots distincts s'est révélée significativement différente entre les manchots captifs ayant jeûné jusqu'en phase 2 (P2C) et phase 3 (P3C) du jeûne prolongé. Les numéros correspondants à ces 10 spots sont en rouge sur la figure 6. Les 10 protéines contenues dans ces 10 spots et donc différentiellement exprimées entre les manchots P2C et P3C comprenaient la transthyrétine, la "retinol binding protein", les apolipoprotéines A-I et A-IV, l'anhydrase carbonique 2, l'adiponectine, la "vitamin D binding protein", la nectinepsine, le complément C3 et l'alpha 1 antitrypsine. Le tableau 3 résume les données obtenues à l'aide de l'algorithme Mascot et à l'aide de l'approche par séquençage *de novo*.

Tableau 3 : Liste des protéines plasmatisques différentielle entre les manchots P2C et P3C et leurs annotations. Toutes les protéines ont été identifiées dans la taxonomie *Gallus gallus* excepté la nectinepsine dans la taxonomie *Cortunix cortunix*.

N°	Protein	NCBI Accession N°	pI / MW (Da)	unique peptides / seq coverage	Peptide sequences	MS/MS ion score	MS-Blast HSP score
1	Tranthyretin	gi 1633502	5.1 / 14142.3	6 / 84.6	APLVSHGSDSK / VEFDTSSYWK / GLGLSPFHEYADVVFANDSGHR / GSPAANVAVK / HYTIAALLSPFSYSTTAVVSDPQE / VLDAVR APLVSHGSDSK / GSPAANVAVK / KAADGSWQDFATGK / TAVVSDPE / VEGLYR / VFTANDSGHR	42.8/63/55.4/5 1.4/113/36.7	56/67/81/41/9 7/50/39/74
2	Retinol-binding protein	gi 1865583 3	5.62/22497.7	10 / 72.9	ELNEDGTCADSYVFSR / VIVHNGFCS / GNDDHWVVDYDITYALHYSR / VKENFDK / KDPEGLFLQDNVVAQFTVDENGQMSATAK / QIDLCLDR / QIDLCLDRK / LFNNWVDCADMIGSFTDTEPAK / YSGTWYAMAK / YWGVASFLOK	64.7/46.5/60.5 /41.2/45/54.8/ 39.5/92.2/52.2 /54.5	
3	Apolipoprotein A-I	gi 211146	5.97/30663.1	18 / 50.3	ASGKDAIAQFESSAVGK / DAIAQFESSAVGK / DMVDVYLETVK / IRDMVDVYLETVK / LADNLDLSAAAAK / LTPVAEEAR / NLAPYSDELRLADNLDLSAAAAK / LTPVAEEAR / GIPOAAEYQAK / VVEQLSNLR / WTELEQFR / EDPQAPLER / LAPVAKDLK / EDAMPYYK / LTPVAE / YYQEVRLPVQQYYER / YKQRLLP / TPLVKDF VVDALNSIQTK	37.2/71.7/75.1 /63.1/122/38.3 /43.9	89/56/65/54/6 6/42/36/41/41/ 39/34/35
4	Carbonic anhydrase 2	gi 115454	6.56/28989.2	1 / 4.2		53.8	
5	Adiponectin	gi 4555120 7	5.45/26443.8	6 / 18.4	SAFSVGLTER / APHPNPLR SAFSVGLTER / APHPNPLR / EQSHYDA / GAFVYR / VFTFMQFQ / GHNGL	51.8/56.2	68/67/49/39/3 4/34
6	Apolipoprotein A-IV	gi 3645997	4.8/40835.1	6 / 16.1	ETVDLQQAQAEITK / LAPLAQLQDER FAQEVKDGILNR / FLSTAEP / QEAAPL / YFTELGNAK ACSR / AGCCAQEGLEKK / APLPVLLGSTR / CCPNNSP / CCPNNSPPLYC / CLSPAPT / CLSPAPTSCFLK / CSNGSPFPAHPGT / DAGSSALGTK / DFFDRFL / ECCSAK / EGLEQK / EPSNEEL / EPSNEELCKAFK / FADCCSGK / FASTCCPA / FASTCCPPNS / FELAR / FNQCCSDVAEDC / FSYLTLAQLK / GEAPLPVLL / HPGTA / HPPQQLP / HTTPDAFLGK / LCGQYALNLFDFK / LCLAAL / LPEAQPTNEQLC / LPPFLEK / LPVLLGSTR / MESTCC / NFLSM / PCCEDAGSSALS / PEASPELL / PEASPELLGKLVQR / PSNEELCKAFK / QQPTNEQLC / RETMPEASPELLANLVEQR / RLVHE / RLAAY / RRHRNV / RYLFELAR / SACLDG / SPFPAHPGT / STCCPPNS / STCCPPNSPPLYC / TCCAD / TLLANSR / TLLSLLLSNR / VAEDCFQ / VCQEFQALGK / YDASE / YLFELAR / YLKPSNE	85/79.2	63/37/34/64 35/43/71/48/8 1/44/69/74/40/ 39/47/35/51/7 9/52/62/68/36/ 68/58/51/36/4 4/62/95/37/62/ 40/57/38/39/6 2/48/52/71/47/ 76/34/35/34/5 4/35/67/53/94/ 41/34/62/45/5 1/38/46/36
8	Nectinepsin	gi 1480970	4.91/40053.1	7 / 17.0	FDDGALDPGYPR / LDAAMAGR / LISDVWIEGPIIDAAFR HEPTQWEC / EPEELCSR / DYSTVCK / DGALDTESPR CDNFDLR / GLCVADPYEITVMK / GTMTVVTVYKAK / LNEFDTVK	87.5/52.8/87/8	38/64/55/34
9	complement component 3	gi 4538230 3	6.69/18406.8	4 / 2.5		50/68.5/37.6/5 9.8	
10	PREDICTED: similar to alpha-1-antitrypsin	gi 1180921 31	5.7/47563.1	1 / 2.3	ATSTTLFMGK	64	

Parmi ces 10 protéines différentielles, l'anhydrase carbonique 2 et l'alpha 1 antitrypsine n'ont été identifiées qu'avec 1 seul peptide. Conformément aux règles de publications des données de spectrométrie de masse [Carr et al., 2004; Bradshaw et al., 2006; Wilkins et al., 2006] et pour montrer la bonne qualité des spectres de masse ayant permis ces identifications, ceux-ci ont été annotés et sont présentés ci-après dans les figures 7 et 8.

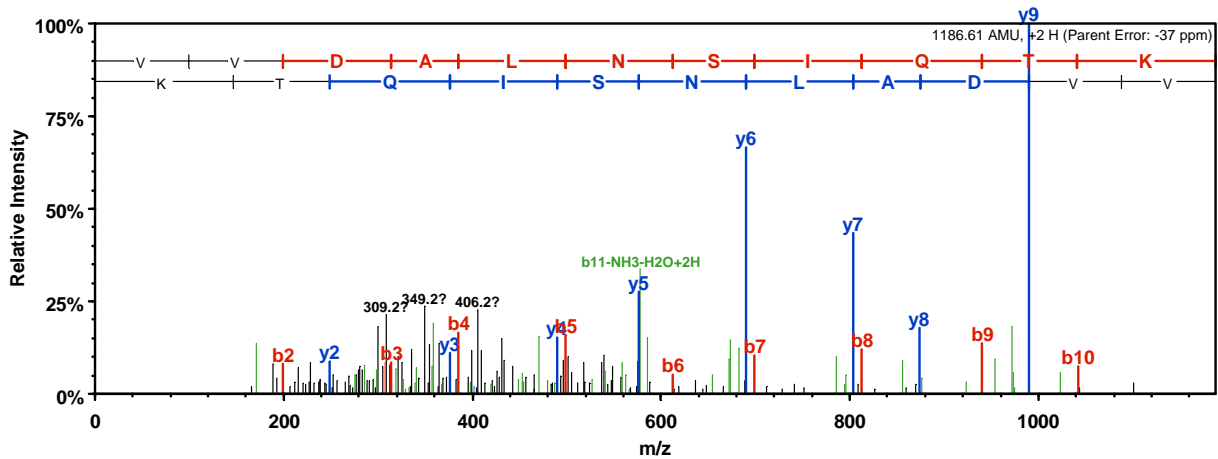


Figure 7 : Spectre de fragmentation de l'ion 594.3, doublement chargé, correspondant au peptide 159-169 VVDALNSIQTK de l'anhydrase carbonique 2.

Les recherches par similarité stricte de masse à l'aide de l'algorithme Mascot ont permis d'identifier un peptide dont la séquence VVDALNSIQTK a été attribuée à l'anhydrase carbonique 2 avec un score d'ion de 53.8, i.e. passant nos critères de validation. L'annotation du spectre de fragmentation correspondant à ce peptide (figure 7) montre que l'identification a été obtenue à partir des séries d'ions b et y, présentes sur le spectre de fragmentation de l'ion m/z 594.3, doublement chargé.

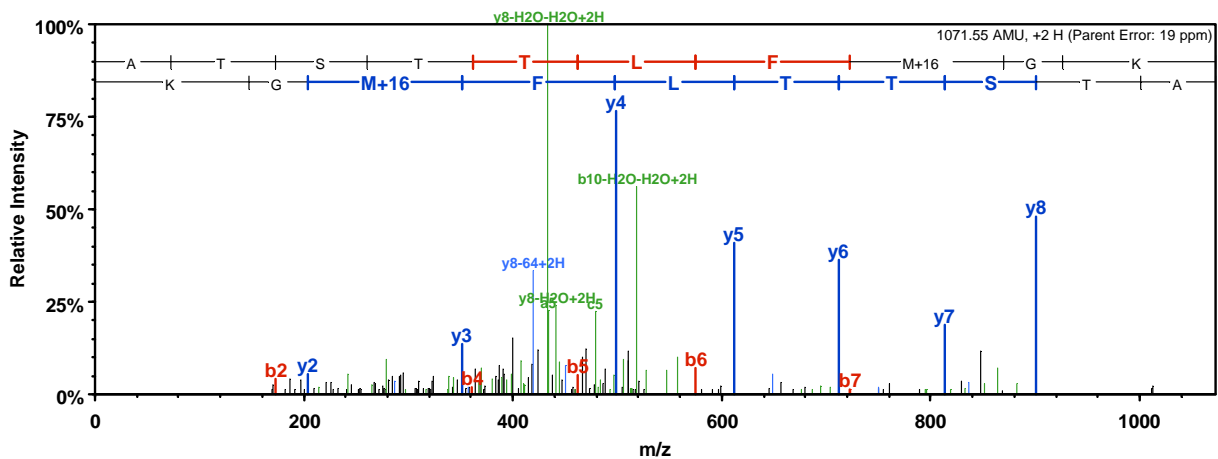


Figure 8 : Spectre de fragmentation de l'ion 536.8, doublement chargé, correspondant au peptide 403-412 ATSTTLFMoxGK de l'alpha-1-antitrypsine.

Les recherches par similarité stricte de masse à l'aide de l'algorithme Mascot ont permis d'identifier un peptide dont la séquence ATSTTLFMoxGK, sur laquelle la méthionine est oxydée, a été attribuée à l'alpha-1-antitrypsine avec un score d'ion de 64.0, i.e. passant nos critères de validation. L'annotation du spectre de fragmentation correspondant à ce peptide (Figure 8) montre que l'identification a été obtenue à partir de la série d'ions y, présente sur le spectre de fragmentation de l'ion m/z 536.8, doublement chargé.



L'intensité des spots contenant les 10 protéines différemment exprimées entre les manchots P2C et P3C est illustrée dans la figure 9. Pour la "vitamin D binding protein", qui est contenue dans plusieurs spots sur le gel (voir le spot N°7 sur la figure 6), la somme des intensités des spots des isoformes a été utilisée pour l'analyse.

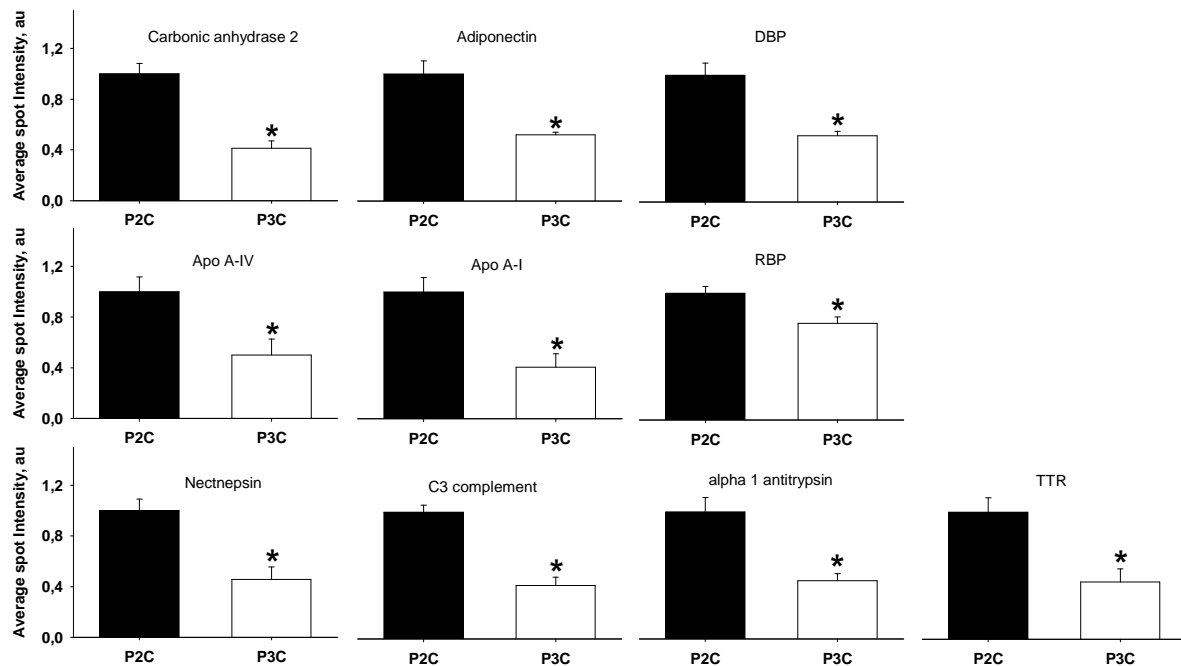


Figure 9 : Niveaux plasmatiques relatifs, mesurés par analyse des images de gels 2D, des protéines différemment exprimées et identifiées chez les manchots P2C et P3C. Les résultats sont les moyennes  $\pm$  ESM,  $n = 5$  / groupe. Les unités sont arbitraires et les valeurs obtenues chez les P3C sont exprimées relativement à celles obtenues chez les P2C qui ont été normalisées à 1. DBP : vitamin D binding protein ; Apo : apolipoprotéine ; RBP : retinol binding protein ; TTR : transthyrétin. P2C et P3C : manchots captifs ayant jeûné totalement jusqu'en phase 2 et phase 3. Les différences significatives ( $p < 0.005$  sauf :  $p < 0.02$  pour nectinpsin et TTR,  $p < 0.05$  pour RBP et apo A-IV) sont indiquées par le symbole \*.

Par rapport aux manchots P2C, les intensités des spots de l'anhydrase carbonique 2, de l'apolipoprotéine A-I, de la nectinpsine, du complément C3, de l'alpha 1 antitrypsine, et de la transthyrétine étaient diminuées significativement chez les animaux P3C de 54-58 %, celles de l'adiponectine, de la "vitamin D binding protein" et de l'apolipoprotéine A-IV de 48-50 % et celle de la "retinol binding protein" de seulement 24 %.

Parmi ces protéines différentielles, l'adiponectine est une adipokine qui stimule l'oxydation des acides gras, et régule aussi bien le métabolisme lipidique que glucidique [Havel, 2004; Nedvidkova et al., 2005]. De plus, les niveaux circulants d'adiponectine sont en général inversement reliés à l'adiposité des individus, et bien corrélés avec la sensibilité à l'insuline. La diminution des niveaux circulants d'adiponectine chez les manchots P3C par rapport aux P2C est en accord avec la littérature

[Bertile & Raclot, 2006], et supporte un rôle de cette protéine comme signal du statut nutritionnel et métabolique des animaux. La réduction des niveaux d'adiponectine pourrait donc notamment être liée à la diminution de l'utilisation des substrats lipidiques en P3. L'adiponectine étant aussi connue pour stimuler la dépense d'énergie [Qi et al., 2004], la diminution de l'adiponectinémie observée ici pourrait aussi favoriser une économie d'énergie en période de privation de nourriture.

Les apolipoprotéines A-I et A-IV sont des composants essentiels des lipoprotéines et jouent un rôle majeur dans le métabolisme des lipides. De plus, l'apolipoprotéine A-IV joue un rôle important dans la régulation à long terme de la prise alimentaire et de la masse corporelle, et agit comme signal de satiété [Tso & Liu, 2004]. L'apolipoprotéine A-IV est en effet connue pour réduire la prise alimentaire et la masse corporelle [Fujimoto et al., 1993]. Les niveaux circulants diminués d'apolipoprotéine A-IV en P3 pourraient signifier la levée de l'inhibition exercée par cette protéine sur la prise alimentaire. Ainsi, les niveaux bas d'apolipoprotéine A-IV en P3 pourraient contribuer à la stimulation du comportement alimentaire lorsque la durée du jeûne se prolonge.

La "retinol binding protein" et la transthyréline sont des protéines circulantes qui transportent le rétinol et la thyroxine, respectivement. La "retinol binding protein" est impliquée dans le développement de la résistance à l'insuline [von Eynatten & Humpert, 2008], ce qui est encore sujet à débat chez l'homme [Lewis et al., 2008]. Elle est associée à l'accumulation des dépôts adipeux viscéraux [Tschoner et al., 2008]. La réduction des niveaux circulants de "retinol binding protein" en P3 pourrait donc être liée à la diminution de l'adiposité des manchots à ce stade. La transthyréline a été indirectement associée au statut nutritionnel, agissant comme un indicateur de la privation de nourriture [Fuhrman et al., 2004]. Les niveaux circulants diminués de transthyréline pourraient donc constituer un bon marqueur de la P3 du jeûne prolongé.

La "vitamin D binding protein" est impliquée dans le transport des métabolites de la vitamine D, et dans le métabolisme lipidique [Gomme & Bertolini, 2004]. Les variations des niveaux de cette protéine sont de plus connus pour être liées aux variations de concentrations en insuline [Nyomba et al., 1985]. Les niveaux plasmatiques diminués de "vitamin D binding protein" chez les animaux P3C pourraient donc être liés à des variations du "signaling" de l'insuline.

L'alpha 1 antitrypsine est essentiellement impliquée dans la protection des tissus vis-à-vis des dommages protéolytiques [Ranes & Stoller, 2005]. La diminution de son expression en P3 par rapport à la P2 chez les animaux captifs pourrait donc favoriser la mobilisation des protéines corporelles à des fins énergétiques à ce stade.

Ainsi, nos résultats montrent essentiellement que les niveaux plasmatiques réduits d'apolipoprotéine A-IV pourraient jouer un rôle dans le signal d'alarme qui induit les animaux à se réalimenter en P3 du jeûne. Les niveaux différentiels des autres protéines plasmatiques sont plutôt liés au statut métabolique et nutritionnel des oiseaux, et reflètent essentiellement des variations du métabolisme lipidique et du "signaling" de l'insuline au cours du jeûne prolongé.

## 6.e. Identification des protéines différemment exprimées entre les manchots P2, P3-A et P3+A

L'intensité des spots contenant les 10 protéines différemment exprimées entre les manchots P2, P3-A et P3+A est illustrée dans la figure 10. Pour la "vitamine D binding protein", qui est contenue dans plusieurs spots sur le gel (voir le spot N°7 sur la figure 6), la somme des intensités des spots des isoformes a été utilisée pour l'analyse.

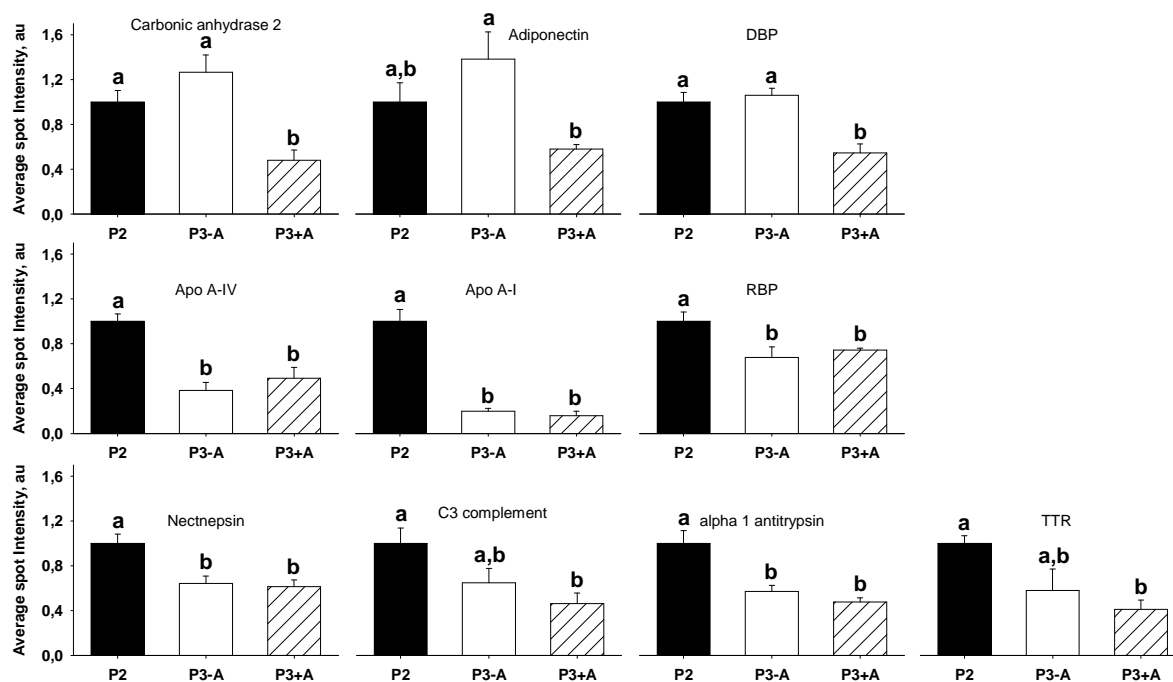


Figure 10 : Niveaux plasmatiques relatifs, mesurés par analyse des images de gels 2D, des protéines différemment exprimées et identifiées chez les manchots P2, P3-A et P3+A. Les résultats sont les moyennes  $\pm$  ESM,  $n = 3-5$  / groupe. Les unités sont arbitraires et les valeurs obtenues chez les P3-A et P3+A sont exprimées relativement à celles obtenues chez les P2 qui ont été normalisées à 1. DBP : vitamine D binding protein ; Apo : apolipoprotéine ; RBP : retinol binding protein ; TTR : transthyretin. P2C et P3C : manchots captifs ayant jeûné totalement jusqu'en phase 2 et phase 3. Les valeurs significativement différentes ( $p < 0.05$ ) ne partagent pas la même lettre.

Par rapport aux manchots P2, les intensités des spots de l'anhydrase carbonique 2, de l'adiponectine, de la "vitamine D binding protein", du complément C3 et de la transthyretine n'étaient diminuées significativement que chez les animaux P3+A (de 52, 42, 45, 54 et 59 %, respectivement). Par contre, par rapport aux intensités des spots mesurées chez les manchots P2, celles des apolipoprotéines A-IV et A-I, de la "retinol binding protein", de la nectinepsine et de l'alpha 1 antitrypsine étaient diminuées significativement aussi bien chez les animaux P3-A (de 62, 80, 32, 36 et 43 %, respectivement) que P3+A (de 51, 84, 26, 39 et 52 %, respectivement).

Les situations écophysiologiques originales (P3-A et P3+A) auxquelles nous avons été confrontés permettent de répartir les protéines différentielles identifiées dans cette étude en deux groupes distincts : celles dont les niveaux d'expression sont comparables entre les groupes P2 et P3-A et celles dont les niveaux d'expression sont comparables entre les groupes P3-A et P3+A. Ceci a permis d'apporter des précisions quant aux conclusions apportées par la première partie de l'expérience (voir paragraphe 6.d). Les manchots P3-A (qui n'ont pas abandonné leur œuf pour partir se réalimenter) ayant des niveaux plasmatiques d'apolipoprotéine A-IV similaires à ceux des manchots P3+A (qui ont abandonné leur œuf pour partir se réalimenter), les niveaux bas de cette protéine ne semblent donc pas pouvoir être directement impliqués dans le signal de réalimentation, mais plutôt dans les orientations du métabolisme (lipidique) des oiseaux. La régulation de l'expression de l'anhydrase carbonique 2, de l'adiponectine, et de la "vitamine D binding protein" ne semble pas strictement liée au statut métabolique des manchots, puisque la motivation à se reproduire semble ici jouer un rôle important. Certaines hormones sont bien connues pour être impliquées dans le comportement parental et l'attention au nid, comme par exemple la prolactine [Buntin, 1996]. Chez le manchot Adélie, l'élévation de la prolactinémie qui stimule l'investissement dans la reproduction semble se maintenir pendant une période dépassant celle pendant laquelle les animaux sont contraints à jeûner [Vleck et al., 1999]. De plus, la prolactinémie ne semble pas diminuée pas une augmentation artificielle de la durée d'incubation, et ne diminue que très lentement en cas de perte de la nichée [Vleck et al., 2000]. La régulation de la seule prolactinémie ne semble donc pas pouvoir diriger le comportement des manchots P3-A ou P3+A. Sans perte de la nichée, l'abandon du nid n'intervient en général que lorsque l'oiseau au jeûne atteint un seuil de déplétion critique avant le retour de sa partenaire [Ainley et al., 1983; Ainley et al., 1986]. L'élévation des niveaux circulants de corticostérone (hormone de stress par excellence chez les oiseaux) en P3 du jeûne chez le manchot Adélie [Cockrem et al., 2006] pourrait également jouer un rôle en "signalant" l'état de stress nutritionnel. Les niveaux de corticostérone et de prolactine seront donc prochainement mesurés chez nos animaux. Il se pourrait que les animaux P3-A soient dans une phase 3 moins prononcée que les manchots P3+A. Ceci est notamment supporté par nos résultats montrant que les manchots P3-A ont une moindre perte (de 2.5%) de masse corporelle que les manchots P3+A par rapport aux P2, ainsi qu'une élévation 1.6 fois moindre de leurs niveaux circulants d'acide urique (Tableau 2). Ceci pourrait possiblement signifier que le signal d'alarme induisant les animaux à se réalimenter n'intervient pas dès la transition entre la phase 2 et la phase 3 du jeûne, mais plutôt lorsque la phase 3 est bien établie. Ceci ne pouvait être déterminé que chez des animaux très adipeux au début de leur jeûne puisque dans ce cas l'installation des événements et transitions métaboliques, qui prend davantage de temps que chez des rongeurs de laboratoires, peut être plus facilement appréhendée.

Le modèle animal choisi pour cette étude a donc permis de préciser les mécanismes moléculaires de réponse au jeûne prolongé, en apportant des informations originales par rapport à ce qui était connu chez le rongeur de laboratoire. Les déterminants exacts du signal de réalimentation restent à élucider et impliquent probablement de nombreux acteurs impliqués aussi bien dans le métabolisme énergétique, que dans le contrôle central de la prise alimentaire et la régulation des

processus cognitifs qui conduisent un animal à abandonner son investissement dans l'incubation de son œuf pour favoriser sa survie.

#### **6.f. Comparaison des effets du jeûne prolongé sur le protéome plasmatique de rat et de manchot Adélie**

Les résultats apportés par cette étude chez le manchot Adélie captif confirment en grande partie ceux obtenus par l'analyse protéomique du plasma de rats soumis à un jeûne plus ou moins prolongé [Bertile et al., 2009]. Des informations sont néanmoins nouvelles puisque des protéines différentielles chez le manchot n'avaient pas été rapportées chez le rat (anhydrase carbonique 2, nectinepsine, complément C3, alpha 1 antitrypsine, et adiponectine). Cependant, l'originalité de cette étude réside essentiellement dans l'analyse du protéome plasmatique des manchots en reproduction. En effet, les situations écophysiologiques originales (P3-A et P3+A) ont permis l'obtention de précieuses informations pour la meilleure compréhension des mécanismes moléculaires de réponse au jeûne prolongé. Ces mécanismes n'étaient pas accessibles par l'étude chez le rongeur de labo [Bertile et al., 2009] ou chez des manchots captifs soumis à un jeûne prolongé (première partie de cette étude, paragraphe 6.d). La composante "spontanée" de l'engagement des manchots dans un jeûne prolongé lors de leur reproduction était essentielle pour préciser ces mécanismes et suppose l'existence d'autres mécanismes qui restent à élucider et qui pourrait impliquer des voies centrales impliquées dans la balance entre motivation à se reproduire et motivation alimentaire. Enfin, la balance entre corticostéronémie et prolactinémie devra être étudiée pour préciser les facteurs régulant le comportement des manchots en incubation qui décident ou non de partir se réalimenter.



### **Chapitre III : Conclusion : Approche protéomique et biologie des organismes dont le génome n'est pas connu**

En conclusion, l'approche protéomique s'est révélée couronnée de succès pour ce qui est de l'identification de protéines de manchot Adélie, une espèce dont le génome n'est pas séquencé. Des protéines plasmatiques différenciellement exprimées selon le statut métabolique et comportemental des manchots ont ainsi été identifiées, suggérant des mécanismes moléculaires de réponse au jeûne prolongé. Notre approche protéomique a donc permis de résoudre la question biologique posée.

Un des aspects importants du succès de notre approche analytique réside dans la duplicité des méthodes d'interprétation des données de spectrométrie de masse. En effet, le recours à l'approche *de novo* pour l'identification des protéines de manchot Adélie, dont le génome n'est pas séquencé, et plus généralement pour l'identification des protéines de tout organisme peu ou faiblement représenté dans les banques de données est essentiel.

En effet, l'approche *de novo* a permis la déduction de séquences en acides aminés et par suite l'identification de protéines à partir des séquences déduites issues des spectres MS/MS qui, bien que de bonne qualité, n'ont pas été attribués lors des recherches Mascot. Ceci reflète l'absence de séquences strictement similaires dans les banques de données. Par exemple, la "Vitamin D binding protein" n'a pu être identifiée qu'en ayant recours au séquençage *de novo*. Ainsi, cette identification a tout de même pu être réalisée à l'aide de 53 peptides uniques recouvrant la séquence de la protéine de poulet (*Gallus gallus*) à 58.6%, ce qui est un très bon résultat compte tenu sa grande taille (Figure 11) :

MKAAVTFLLL	LSVATHAEHR	GRPYVRDKVC	QEFKTMGKDD	FRAMTLIMNS
		<b>VC</b>	<b>QEFQALGK</b>	<b>TLIMNS</b>
RKFSNATFEE	ISHLVHEMVS	LAETCCADGV	DPSCYDTGSS	ALSAKSCSPD
<b>R</b>		<b>TCCAD</b>	<b>PCCEDAGSS</b>	<b>ALSAK CSNG</b>
SPFPAHPGTA	ACCLHQGLEQ	KLCLAALHP	PRQLPHYVEP	SNEELCEAFK
<b>SPFPAHPGTA</b>	<b>GCCAQEGLEK</b>	<b>KLCLAAL HP</b>	<b>PQQLP YVEP</b>	<b>SNEELCKAFK</b>
KDPKDFADRF	LHEYVSSYGQ	APLPVLLGST	RNFLSMVSTC	CISPSPTVCF
<b>DFFDRF</b>	<b>VHE</b>	<b>GE</b>	<b>APLPVLLGST</b>	<b>RNFLSMESTC</b>
LKEKLQRKTL	SLLTLMNSRA	CSRLAAYGKD	KMKFSYLTML	AQKIPSASFE
<b>LK</b>	<b>TL</b>	<b>SLLTLLSNRA</b>	<b>CSRLAAY</b>	<b>FSYLTSL</b>
DLSPLAEDAA	EMFSQCCDSV	AEDCIQQKLS	EHTTKICATL	SAKDKRFADC
	<b>FNQCCSDV</b>	<b>AEDCFQ</b>		<b>FADC</b>
CEGKNVMQNY	FCISSLQPAV	APKLPPELQKP	TNKQLCDDNG	ALHTRYMF
<b>CSGK</b>		<b>LPEAQQP</b>	<b>TNEQLC</b>	<b>RYLFE</b>
LARRHTNVPD	VFLGKLYDAS	ENVIKECCSA	KDASSCLDGK	RQQMGTELPA
<b>LARRHTTVPD</b>	<b>AFLGK YDAS</b>	<b>E</b>	<b>ECCSA</b>	<b>K</b>
FLEKTDQLCG	QYTELNFLDF	KKRLRDSIRQ	TRPDASPELL	TQLTDQRADF
<b>FLEK</b>	<b>LCG</b>	<b>QYTALNFLDF</b>	<b>K</b>	<b>PEASPELL</b>
ASTCCPANSP	PLYCVAQVTA	ELASTCKQGS	CMLV	<b>GKLVEQR</b>
<b>ASTCCPPNSP</b>	<b>PLYC</b>			<b>F</b>

Figure 11 : Alignement des séquences des peptides identifiés avec la séquence complète (en noir) de la "vitamin D binding protein". La séquence complète (en noir) correspond à la protéine de poulet (*Gallus gallus*, NP\_990213.1). Les peptides (en bleu) ont été identifiés par approche de novo. Les acides aminés en italique sont ceux qui étaient différents entre les séquences peptidiques déterminées expérimentalement par séquençage de novo et la séquence théorique chez le poulet.

Par ailleurs, l'approche *de novo* a aussi permis de largement renforcer l'identification des protéines différentielles identifiées dans cette étude, en ajoutant un ou plusieurs nouveaux peptides uniques. Par exemple, l'apolipoprotéine A-I avait été identifiée à l'aide de 7 peptides à l'issue de la recherche Mascot, attribués à la séquence de la protéine de poulet (*Gallus gallus*), ce qui correspondait à une couverture de 27.2 %. En plus de ces 7 peptides, 2 autres (GIPQAAEYQAK et VVEQLSNLR) ont été attribués par Mascot à la séquence de l'apolipoprotéine A-I de canard colvert (*Anas platyrhynchos*, gi|113990). Restreindre la banque de données que nous avons téléchargée pour cette étude à la taxonomie "oiseaux" s'est donc révélé un choix judicieux puisque permettant des identifications à partir des homologues de plusieurs espèces. 9 peptides supplémentaires viennent renforcer cette identification grâce à l'approche *de novo* et permettent d'augmenter la couverture de séquence à 50.8 % (Figure 11).



```

MRGVLVTLAV LFLTGTQARS FWQHDEPQTP LDRIRDMVDV YLETVKASGK
                EDPQAP LERIRDMVDV YLETVKASGK
DAIAQFESSA VGKQLDLKLA DNLDTLSAAA AKLREDMAPY YKEVREMWLK
DAIAQFESSA VGK LA DNLDTLSAAA AK EDMAPY YKEVR
DTEALRAELT KDLEEVKEKI RPFLDQFSAK WTEELEQYRQ RLTPVAQELK
                WTEELEQFR LTPVAQELK
ELTKQKVELM QAKLTPVAEE ARDRLRGHVE ELRKNLAPYS DELRQKLSQK
                LTPVAEE AR NLAPYS DELR
LEEIREKGIP QASEYQAKVM EQLSNLREKM TPLVQEFRRER LTPYAENLKN
                GIP QASEYQAKVV EQLSNLR PLVQQYVER
RLISFLDELQ KSV

```

Figure 12 : Alignement des séquences des peptides identifiés avec la séquence complète (en noir) de l'apolipoprotéine A-I de poulet. Les peptides en rouge ont été attribués par Mascot à la séquence d'apolipoprotéine A-I de poulet. Les peptides en vert ont été attribués par Mascot à la séquence d'apolipoprotéine A-I de canard colvert. Les peptides en bleu ont été identifiés par approche de novo. Les acides aminés en italique sont ceux qui étaient différents entre les séquences peptidiques déterminées expérimentalement par séquençage de novo ou celles attribués par Mascot à une espèce autre que le poulet et la séquence théorique chez le poulet.

Il est à noter que nos résultats montrant des différences dans l'expression de certaines protéines entre les groupes de manchots n'ont pas pu être confirmés par une autre méthode de type ELISA puisqu'aucun anticorps dirigé spécifiquement contre ces protéines chez le manchot n'est disponible. Cependant, en déterminant un faible taux de faux positifs (< 1 %), la mise en place de notre stratégie "target-decoy" a permis de renforcer la confiance et la non ambiguïté de nos identifications. De plus, le fait que certaines des différences d'expressions protéiques rapportées ici étaient également retrouvées chez une autre espèce, en l'occurrence le rat (*Rattus norvegicus*), lors du jeûne prolongé [Bertile et al., 2009] permet en quelque sorte de valider de nos résultats de quantification relative.

Ce travail a fait l'objet de communications par affiche (versions jointes en pages suivantes) présentées:

- Première version: Lors du 56ème congrès de l'"American Society Mass Spectrometry" qui s'est déroulé en juin 2008 à Denver, CO, USA.

- Seconde version: Lors du "HUPO 7th Annual World Congress" qui s'est déroulé en aout 2008 à Amsterdam, Pays-Bas, lors du 25èmes Journées Françaises de Spectrométrie de Masse qui se sont déroulées en septembre 2008 à Grenoble et lors du 25èmes Congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique qui s'est déroulés en octobre 2008 à Tous.



# Biodiversity exploration using MS-based proteomics for the improved understanding of extreme physiopathological situations



Laetitia Fouillen<sup>1</sup>, Thierry Raclot<sup>2</sup>, Alain Van Dorsselaer<sup>1</sup>, Fabrice Bertile<sup>1</sup>  
<sup>1</sup>IPHC-DSA, ULP, CNRS, Strasbourg, France; <sup>2</sup>IPHC-DEPE, ULP, CNRS, Strasbourg, France

## OVERVIEW

- Purpose:** Identification of the plasma proteins differentially expressed between the fed state, P2 and P3 of fasting
- Methods:** Plasma sample pre-treatment techniques and two-dimensional gel electrophoresis (2-DE) Image analysis MS identifications and target-decoy strategy adapted to a non-sequenced organism
- Results:** Using albumin depletion and protein precipitation with TCA, more than 400 spots were detected. Among 65 unique proteins identified (with little false positive identifications), 7 proteins were found to be differentially expressed between groups
- Conclusion:** The combination of plasma sample pre-treatment techniques, 2-D gels, and MS identification revealed potential adaptive mechanisms to prolonged fasting

## INTRODUCTION

During the reproduction period, Adelle penguins (*Pygoscelis adeliae*) spontaneously fast for up to 6-7 weeks. Such a prolonged fasting is characterized by three distinct metabolic phases (P1 to P3). P1 is a short period of adaptation followed by P2, during which energy requirements are mostly derived from lipid fuels. P3 is characterized by an increase in body protein utilization and enhanced drive for refeeding. The mechanisms mediating metabolic, hormonal and behavioural adaptations to prolonged fasting are not fully understood.



Therefore, our aim was to identify the plasma proteins that are differentially expressed between the fed state, P2 and P3 of fasting. To this end, we used a differential proteomic approach based on 2-DE and MS analyses. As no molecular tool, such as antibodies, is available for penguin proteins, no immunological validation of the proteomics results can be performed. Thus, we developed a target-decoy search strategy in order to increase confidence in protein identification by MS

## METHODS

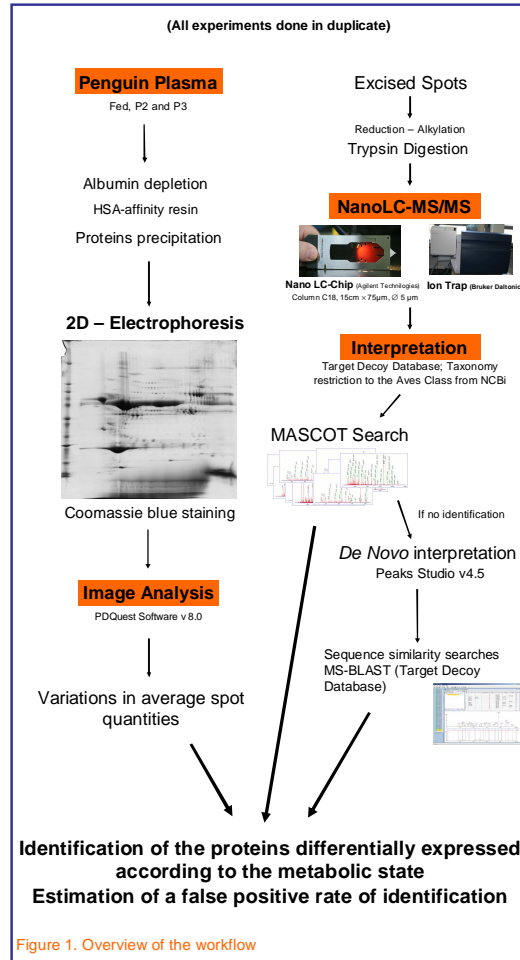


Figure 1. Overview of the workflow

Table 1. Profiles of fed and fasted penguins

Metabolic state	FED	P2	P3
Duration of fasting (days)	-	35-40	40-45
Body mass (Kg)	5.69 ± 0.15	3.90 ± 0.11	3.37 ± 0.12
Body mass loss (%)	-	31 ± 1	41 ± 2
Plasma β-OHB (mM)	0.85 ± 0.08	1.18 ± 0.06	0.74 ± 0.05
Plasma glucose (mM)	12.8 ± 0.7	12.2 ± 0.6	13.2 ± 0.9
Plasma uric acid (mM)	0.45 ± 0.02	0.28 ± 0.03	1.12 ± 0.13

P2 and P3: phase 2 and phase 3 of fasting.

## RESULTS

### Comparison of plasma sample pre-treatments

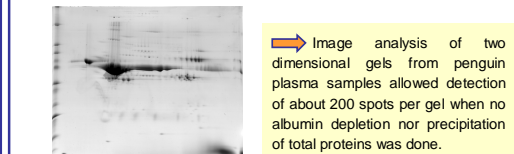


Figure 2. 2D gel of crude plasma sample

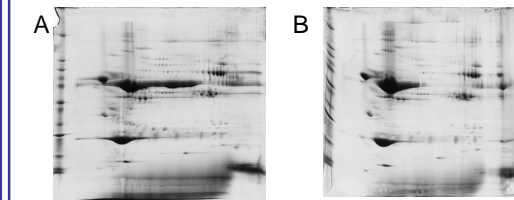


Figure 3. 2D gels after albumin depletion using 1 volume of HSA-affinity resin followed by precipitation with TCA (A) and acetone (B). Protein precipitation was better with TCA than with acetone. Image analysis allowed detection of about 300 spots per gel after acetone precipitation and 400 spots per gel after TCA precipitation.

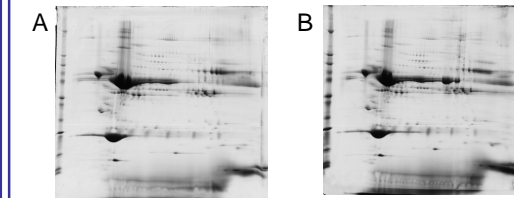


Figure 4. 2D gels after albumin depletion using 2 (A) and 3 (B) volumes of HSA-affinity resin followed by precipitation with TCA.

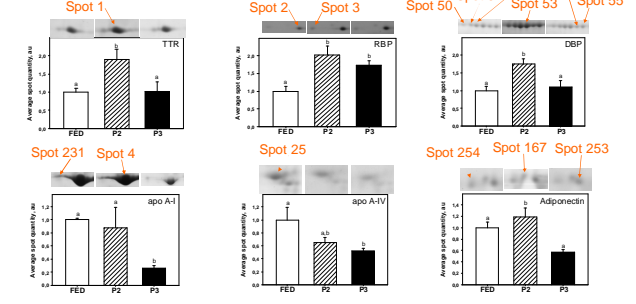
Albumin depletion and the number of spots detected were not improved when increasing HSA-affinity resin quantity.

Loading of 700µg of total plasma proteins per gel was enough to detect a maximum of spots (~400) without artefacts due to the albumin spot. HSA-affinity resin for human albumin did not prove enough efficiency to remove penguin albumin, whatever the quantity of resin. Nevertheless, the number of spots visualized after Coomassie blue staining was greater after depletion. The differential analysis between fed, P2 and P3 penguins was then performed after albumin depletion with 1 volume of HSA-affinity resin followed by precipitation with TCA

## CONCLUSION

Plasma sample pre-treatment techniques and a target-decoy strategy for protein identification by MS were adapted to a non-sequenced organism to study poorly known biological processes. The variations in protein levels reflect essentially fasting-induced alterations in lipid metabolism and insulin signaling. Interestingly, the reduction of apolipoprotein A-IV levels could contribute to the alarm signal that triggers refeeding in late fasting.

### Image analysis and identification of proteins



The amount of protein in gel spots was determined by calculating the average intensity of gel spots using the PDQuest™ software. From preliminary image analysis, 7 proteins were found to be differentially expressed between groups (see identifications in Table 2). For those proteins with multiple spots in the gel, the average of spot quantities of individual spots from isoforms was used in the analysis. Values are means ± SEM, n=6-8 per group. Bars that do not share the same superscript letter are significantly different from each other (ANOVA and Bonferroni post hoc tests; P<0.05).

Table 2. Differentially expressed proteins during fasting : identification by MS

Spots	Accession Number	Protein	Taxonomy	Number of unique peptides
1	gi 1633502	Chain A, Transthyretin	Gallus gallus	6
2	gi 1865833	Chain A, Chicken Plasma Retinol-Binding Protein	Gallus gallus	10
3	gi 1865833	Chain A, Chicken Plasma Retinol-Binding Protein	Gallus gallus	8
4	gi 113990	Apolipoprotein A-I precursor	Gallus gallus	7
	gi 2286119	Apolipoprotein A-I precursor	Anas platyrhynchos	2
231	gi 113990	Apolipoprotein A-I precursor	Gallus gallus	4
	gi 2286119	Apolipoprotein A-I precursor	Anas platyrhynchos	2
25	gi 3645997	apolipoprotein AIV	Gallus gallus	2
50	NP_990213.1	Vitamin D binding protein	Gallus gallus	23
51	NP_990213.1	Vitamin D binding protein	Gallus gallus	25
52	NP_990213.1	Vitamin D binding protein	Gallus gallus	22
53	NP_990213.1	Vitamin D binding protein	Gallus gallus	29
54	NP_990213.1	Vitamin D binding protein	Gallus gallus	20
55	NP_990213.1	Vitamin D binding protein	Gallus gallus	14
84	gi 45382302	complement component 3	Gallus gallus	8
85	gi 45382303	complement component 3	Gallus gallus	14
170	gi 45382303	complement component 3	Gallus gallus	1
167	gi 45551207	adiponectin	Gallus gallus	2
253	gi 45551208	adiponectin	Gallus gallus	1
254	gi 45551209	adiponectin	Gallus gallus	2

NanoLC-Chip/MS/MS data from ~200 excised spots were searched using a target-decoy strategy. Using the Mascot program, 56 unique avian proteins were identified. 9 additional proteins were identified using de novo analyses and MS-Blast searches. Only 1 protein was identified in the decoy database. Our results then seem to contain little false positive identifications.



## INTRODUCTION

During the reproduction period, Adelie penguins (*Pygoscelis adeliae*) spontaneously fast for up to 6-7 weeks. Such a **prolonged fasting** is characterized by three distinct metabolic phases (P1 to P3). P1 is a short period of adaptation followed by P2, during which energy requirements are mostly derived from lipid fuels. P3 is characterized by an increase in body protein utilization and enhanced drive for refeeding. The mechanisms mediating metabolic, hormonal and behavioural adaptations to prolonged fasting are not fully understood.



Therefore, our aim was to **identify the plasma proteins that are differentially expressed between the fed state, P2 and P3 of fasting**. To this end, we used a differential proteomic approach based on 2-DE and MS analyses. As no molecular tool, such as antibodies, is available for penguin proteins, no immunological validation of the proteomic results can be performed. Thus, we developed a **target-decoy search strategy** in order to increase confidence in protein identification by MS.

A similar analysis has already been performed in experimentally fasted rats. The **results from penguin and rat samples are compared**.

## METHODS

(All experiments done in duplicate)

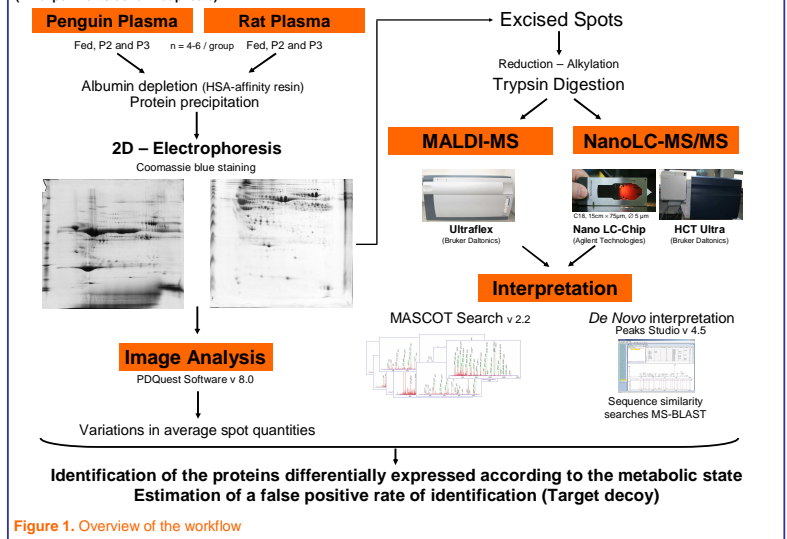


Figure 1. Overview of the workflow

## RESULTS

### Penguin plasma sample pre-treatment

Human serum albumin-affinity resin did not prove enough efficiency to remove penguin albumin, whatever the quantity of resin. Nevertheless, the number of spots visualized after Coomassie blue staining was greater after depletion. The differential analysis between fed, P2 and P3 penguins was then performed after albumin depletion with HSA-affinity resin followed by precipitation with TCA/acetone which was better than with acetone.

Loading of 700 µg of total plasma proteins per gel was enough to **detect a maximum of spots (~400 / gel) without artefacts due to the albumin spot**.

### Penguin metabolic states

Table 1. Profiles of fed and fasted penguins

Metabolic state	FED	P2	P3
Duration of fasting (d)	-	35-40	40-45
Body mass (Kg)	5.69 ± 0.15 <sup>a</sup>	3.90 ± 0.11 <sup>b</sup>	3.37 ± 0.12 <sup>c</sup>
Body mass loss (%)	-	31 ± 1 <sup>a</sup>	41 ± 2 <sup>b</sup>
Plasma β-OHB (mM)	0.85 ± 0.08 <sup>a</sup>	1.18 ± 0.06 <sup>b</sup>	0.74 ± 0.05 <sup>a</sup>
Plasma glucose (mM)	12.8 ± 0.7	12.2 ± 0.6	13.2 ± 0.9
Plasma uric acid (mM)	0.45 ± 0.02 <sup>a</sup>	0.28 ± 0.03 <sup>b</sup>	1.12 ± 0.13 <sup>c</sup>

Plasma parameters allowed to validate the metabolic state of all penguins (Table 1) and rats (data not shown).

Values in a same row that do not share the same superscript are significantly different ( $p < 0.05$ ).

### Gel image analysis

#### A (Penguins)

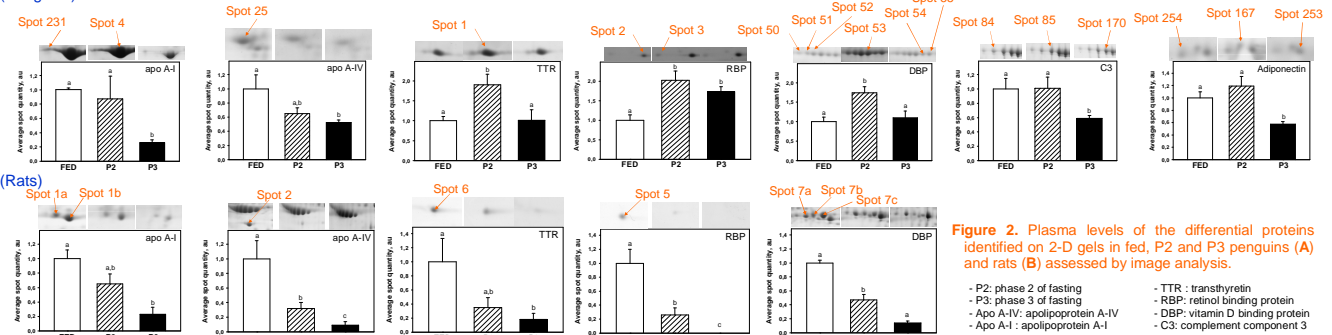


Figure 2. Plasma levels of the differentially expressed proteins identified on 2-D gels in fed, P2 and P3 penguins (A) and rats (B) assessed by image analysis.

- P2: phase 2 of fasting  
- P3: phase 3 of fasting  
- Apo A-I: apolipoprotein A-I  
- TTR: transthyretin  
- RBP: retinol binding protein  
- DBP: vitamin D binding protein  
- C3: complement component 3

From preliminary image analysis, 7 proteins were found to be differentially expressed between penguin groups (see identifications in Table 2). Among these 7 differential proteins, 5 were also differentially expressed between rat groups (see identifications in Table 3). The variations according to the metabolic state were similar between penguins and rats for apolipoproteins (A-I and A-IV), but quite different for TTR, RBP and DBP. Adiponectin and C3 were not detected in rats samples.

For those proteins with multiple spots in the gel, the average of spot quantities of individual spots from isoforms was used in the analysis. Values are means ± SEM,  $n = 4-6$  per group. Bars that do not share the same superscript letter are significantly different from each other (ANOVA and Bonferroni post hoc tests;  $p < 0.05$ ).

### Identification of proteins by MS

Table 2. Identification of differentially expressed penguin proteins during fasting

Spots	Protein	Taxonomy	Number of unique peptides
4	Apolipoprotein A-I precursor	Gallus gallus	7
	Apolipoprotein A-I precursor	Anas platyrhynchos	2
231	Apolipoprotein A-I precursor	Gallus gallus	4
	Apolipoprotein A-I precursor	Anas platyrhynchos	2
25	apolipoprotein A-IV	Gallus gallus	2
167	adiponectin	Gallus gallus	2
253	adiponectin	Gallus gallus	1
254	adiponectin	Gallus gallus	2
1	Chain A, Transthyretin	Gallus gallus	6
2	Chain A, Chicken Plasma Retinol-Binding Protein	Gallus gallus	10
3	Chain A, Chicken Plasma Retinol-Binding Protein	Gallus gallus	8
50	Vitamin D binding protein	Gallus gallus	23
51	Vitamin D binding protein	Gallus gallus	25
52	Vitamin D binding protein	Gallus gallus	22
53	Vitamin D binding protein	Gallus gallus	29
54	Vitamin D binding protein	Gallus gallus	20
55	Vitamin D binding protein	Gallus gallus	14
84	complement component 3	Gallus gallus	8
85	complement component 3	Gallus gallus	14
170	complement component 3	Gallus gallus	1

Table 3. Identification of differentially expressed rat proteins during fasting

Spots	Protein	Taxonomy	Number of unique peptides
1a	Apolipoprotein A-I precursor	Rattus norvegicus	18
1b	Apolipoprotein A-I precursor	Rattus norvegicus	18
2	apolipoprotein A-IV	Rattus norvegicus	31
6	Chain A, Transthyretin	Rattus norvegicus	8
5	Plasma Retinol-Binding Protein	Rattus norvegicus	12
7a	Vitamin D binding protein	Rattus norvegicus	22
7b	Vitamin D binding protein	Rattus norvegicus	23
7c	Vitamin D binding protein	Rattus norvegicus	37

⇒ **Penguins** : NanoLC-Chip/MS/MS data from ~200 excised spots were searched using a target-decoy strategy. Using the Mascot program, **56 unique avian proteins** were identified. 9 additional proteins were identified using *de novo* analysis and MS-Blast searches. Only 1 protein was identified in the decoy database. Our results then seem to contain **little false positive identifications**.

⇒ **Rats** : MALDI/MS data from ~30 excised spots were searched using the Mascot program. **27 unique rat proteins** were identified. The variations in protein levels were confirmed by ELISA (data not shown).

## CONCLUSION

Plasma sample pre-treatment techniques and a target-decoy strategy for protein identification by MS were adapted to a non-sequenced organism to study poorly known biological processes. The variations in protein levels reflect essentially **fasting-induced alterations in lipid metabolism and insulin signaling**. The variations observed in penguins were not strictly similar to those observed in rats, which could reflect **specific adaptations** when a given organism (Penguin = "fasting professional") is compelled to regularly fast for long periods. Interestingly, the reduction of **apolipoprotein A-IV** (satiety and adiposity signal) levels in both species could contribute to the alarm signal that triggers refeeding in late fasting.



## **Références Bibliographique**

### **A.**

Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L. and Pounds, J. G. (2002). "Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry." Mol Cell Proteomics **1**(12): 947-55.

Ainley, D. G., LeResche, R. E. and Sladen, W. J. L. (1983). "Breeding Biology of the Adélie Penguin." Univ. of California Press, Berkeley.

Ainley, D. G., LeResche, R. E. and Sladen, W. J. L. (1986). "Survival analysis of eggs and chicks in Adelie Penguins (*Pygoscelis adeliae*). ." Auk **103**: 379-388.

Anderson, N. L. and Anderson, N. G. (2002). "The human plasma proteome: history, character, and diagnostic prospects." Mol Cell Proteomics **1**(11): 845-67.

Asara, J. M., Christofk, H. R., Freemark, L. M. and Cantley, L. C. (2008). "A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen." Proteomics **8**(5): 994-9.

### **B.**

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-31.

Barzilai, N. and Gabriely, I. (2001). "The role of fat depletion in the biological benefits of caloric restriction." J Nutr **131**(3): 903S-906S.

Baumann, S., Ceglarek, U., Fiedler, G. M., Lembcke, J., Leichtle, A. and Thiery, J. (2005). "Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." Clin Chem **51**(6): 973-80.

Belkhou, R., Cherel, A., Heitz, A., Robin, J. P. and Le Maho, Y. (1991). "Energy contribution of proteins and lipids during prolonged fasting in the rat." Nutrition research **11**: 365-74.

Berth, M., Moser, F. M., Kolbe, M. and Bernhardt, J. (2007). "The state of the art in the analysis of two-dimensional gel electrophoresis images." Appl Microbiol Biotechnol **76**(6): 1223-43.

Bertile, F., Oudart, H., Maho, Y. L. and Raclot, T. (2003). "Recombinant leptin in the hypothalamic response to late fasting." Biochem Biophys Res Commun **310**(3): 949-55.

Bertile, F. and Raclot, T. (2004). "Differences in mRNA expression of adipocyte-derived factors in response to fasting, refeeding and leptin." Biochim Biophys Acta **1683**(1-3): 101-9.

Bertile, F. and Raclot, T. (2006). "Adipose-Derived Factors During Nutritional Transitions." Curr Nutr Food Sci **2**(2): 127-139.

Bertile, F., Robert, F., Delval-Dubois, V., Sanglier, S., Schaeffer, C. and Van Dorselaer, A. (2007). "Endogenous Plasma Peptide Detection and Identification in the Rat by a Combination of Fractionation Methods and Mass Spectrometry." Biomarker Insights **2**: 385-401.

Bertile, F., Schaeffer, C., Le Maho, Y., Raclot, T. and Van Dorsselaer, A. (2009). "A proteomic approach to identify differentially expressed plasma proteins between the fed and prolonged fasted states." Proteomics **9**(1): 148-58.

Bjorhall, K., Miliotis, T. and Davidsson, P. (2005). "Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples." Proteomics **5**(1): 307-17.

Blagoev, B., Ong, S. E., Kratchmarova, I. and Mann, M. (2004). "Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics." Nat Biotechnol **22**(9): 1139-45.

Bradshaw, R. A., Burlingame, A. L., Carr, S. and Aebersold, R. (2006). "Reporting protein identification data: the next generation of guidelines." Mol Cell Proteomics **5**(5): 787-8.

Buntin, J. D. (1996). "Neural and hormonal control of parental behavior in birds." Advances in the study of behavior **25**: 161-213.

### **C.**

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K. and Nesvizhskii, A. (2004). "The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data." Mol Cell Proteomics **3**(6): 531-3.

Castro, A. J., Carapito, C., Zorn, N., Magne, C., Leize, E., Van Dorsselaer, A. and Clement, C. (2005). "Proteomic analysis of grapevine (*Vitis vinifera* L.) tissues subjected to herbicide stress." J Exp Bot **56**(421): 2783-95.

Catusse, J., Strub, J. M., Job, C., Van Dorsselaer, A. and Job, D. (2008). "Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression." Proc Natl Acad Sci U S A **105**(29): 10262-7.

Challet, E., le Maho, Y., Robin, J. P., Malan, A. and Cherel, Y. (1995). "Involvement of corticosterone in the fasting-induced rise in protein utilization and locomotor activity." Pharmacol Biochem Behav **50**(3): 405-12.

Cherel, Y. and Le Maho, Y. (1991). "Refeeding after the late increase in nitrogen excretion during prolonged fasting in the rat." Physiol Behav **50**(2): 345-9.

Cockrem, J. F., Potter, M. A. and Candy, E. J. (2006). "Corticosterone in relation to body mass in Adelie penguins (*Pygoscelis adeliae*) affected by unusual sea ice conditions at Ross Island, Antarctica." Gen Comp Endocrinol **149**(3): 244-52.

Conrads, T. P., Hood, B. L. and Veenstra, T. D. (2006). "Sampling and analytical strategies for biomarker discovery using mass spectrometry." Biotechniques **40**(6): 799-805.

Criscuolo, F., Chastel, O., Bertile, F., Gabrielsen, G., Le Maho, Y. and Raclot, T. (2005). "Corticosterone alone does not trigger a short term behavioural shift in incubating female Common Eiders, but does modify long term reproductive success." Journal of Avian Biology **36**: 306-312.



**D.**

Deutsch, E. W., Eng, J. K., Zhang, H., King, N. L., Nesvizhskii, A. I., Lin, B., Lee, H., Yi, E. C., Ossola, R. and Aebersold, R. (2005). "Human Plasma PeptideAtlas." Proteomics **5**(13): 3497-500.

Domon, B. and Aebersold, R. (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-7.

Dowell, J. A., Heyden, W. V. and Li, L. (2006). "Rat neuropeptidomics by LC-MS/MS and MALDI-FTMS: Enhanced dissection and extraction techniques coupled with 2D RP-RP HPLC." J Proteome Res **5**(12): 3368-75.

**F.**

Fenselau, C. and Yao, X. (2007). "Proteolytic labeling with <sup>18</sup>O for comparative proteomics studies: preparation of <sup>18</sup>O-labeled peptides and the <sup>18</sup>O/<sup>16</sup>O peptide mixture." Methods Mol Biol **359**: 135-42.

Fuhrman, M. P., Charney, P. and Mueller, C. M. (2004). "Hepatic proteins and nutrition assessment." J Am Diet Assoc **104**(8): 1258-64.

Fujimoto, K., Fukagawa, K., Sakata, T. and Tso, P. (1993). "Suppression of food intake by apolipoprotein A-IV is mediated through the central nervous system in rats." J Clin Invest **91**(4): 1830-3.

**G.**

Gerszten, R. E., Accurso, F., Bernard, G. R., Caprioli, R. M., Klee, E. W., Klee, G. G., Kullo, I., Laguna, T. A., Roth, F. P., Sabatine, M., Srinivas, P., Wang, T. J. and Ware, L. B. (2008). "Challenges in translating plasma proteomics from bench to bedside: update from the NHLBI Clinical Proteomics Programs." Am J Physiol Lung Cell Mol Physiol **295**(1): L16-22.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S. G. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.

Gomme, P. T. and Bertolini, J. (2004). "Therapeutic potential of vitamin D-binding protein." Trends Biotechnol **22**(7): 340-5.

Good, D. M., Thongboonkerd, V., Novak, J., Bascands, J. L., Schanstra, J. P., Coon, J. J., Dominiczak, A. and Mischak, H. (2007). "Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future." J Proteome Res **6**(12): 4549-55.

Goodman, M. N., Larsen, P. R., Kaplan, M. M., Aoki, T. T., Young, V. R. and Ruderman, N. B. (1980). "Starvation in the rat. II. Effect of age and obesity on protein sparing and fuel metabolism." Am J Physiol **239**(4): E277-E286.

Gorg, A., Weiss, W. and Dunn, M. J. (2004). "Current two-dimensional electrophoresis technology for proteomics." Proteomics **4**(12): 3665-85.

Granger, J., Siddiqui, J., Copeland, S. and Remick, D. (2005). "Albumin depletion of human plasma also removes low abundance proteins including the cytokines." Proteomics **5**(18): 4713-8.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999). "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." Nat Biotechnol **17**(10): 994-9.

## **H.**

Harper, R. G., Workman, S. R., Schuetzner, S., Timperman, A. T. and Sutton, J. N. (2004). "Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry." Electrophoresis **25**(9): 1299-306.

Havel, P. J. (2004). "Update on adipocyte hormones: regulation of energy balance and carbohydrate/lipid metabolism." Diabetes **53 Suppl 1**: S143-51.

Hennion, M. C. (1999). "Solid-phase extraction: method development, sorbents, and coupling with liquid chromatography." J Chromatogr A **856**(1-2): 3-54.

Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P. and Hale, J. E. (2005). "Comprehensive label-free method for the relative quantification of proteins from biological samples." J Proteome Res **4**(4): 1442-50.

Hood, B. L., Zhou, M., Chan, K. C., Lucas, D. A., Kim, G. J., Issaq, H. J., Veenstra, T. D. and Conrads, T. P. (2005). "Investigation of the mouse serum proteome." J Proteome Res **4**(5): 1561-8.

Hu, S., Loo, J. A. and Wong, D. T. (2006). "Human body fluid proteome analysis." Proteomics **6**(23): 6326-53.

## **J.**

Jones, P., Cote, R. G., Cho, S. Y., Klie, S., Martens, L., Quinn, A. F., Thorneycroft, D. and Hermjakob, H. (2008). "PRIDE: new developments and new datasets." Nucleic Acids Res **36**(Database issue): D878-83.

Jones, P., Cote, R. G., Martens, L., Quinn, A. F., Taylor, C. F., Derache, W., Hermjakob, H. and Apweiler, R. (2006). "PRIDE: a public repository of protein and peptide identifications for the proteomics community." Nucleic Acids Res **34**(Database issue): D659-63.

## **K.**

Kay, R., Barton, C., Ratcliffe, L., Matharoo-Ball, B., Brown, P., Roberts, J., Teale, P. and Creaser, C. (2008). "Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis." Rapid Commun Mass Spectrom **22**(20): 3255-60.

Koomen, J. M., Li, D., Xiao, L. C., Liu, T. C., Coombes, K. R., Abbruzzese, J. and Kobayashi, R. (2005). "Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery." J Proteome Res **4**(3): 972-81.

Koubi, H. E., Robin, J. P., Dewasmes, G., Le Maho, Y., Frutoso, J. and Minaire, Y. (1991). "Fasting-induced rise in locomotor activity in rats coincides with increased protein utilization." Physiol Behav **50**(2): 337-43.

Kuhner, S. and Gavin, A. C. (2007). "Towards quantitative analysis of proteome dynamics." Nat Biotechnol **25**(3): 298-300.

## **L.**

Le Maho, Y., Delclitte, P. and Chatonnet, J. (1976). "Thermoregulation in fasting emperor penguins under natural conditions." Am J Physiol **231**(3): 913-22.

Le Maho, Y., Robin, J. P. and Cherel, A. (1988). "Starvation as a treatment for obesity: the need to conserve body protein." News Physiol Sci **3**: 21-24.

Lenaerts, K., Sokolovic, M., Bouwman, F. G., Lamers, W. H., Mariman, E. C. and Renes, J. (2006). "Starvation induces phase-specific changes in the proteome of mouse small intestine." J Proteome Res **5**(9): 2113-22.

Lescuyer, P., Hochstrasser, D. and Rabilloud, T. (2007). "How shall we use the proteomics toolbox for biomarker discovery?" J Proteome Res **6**(9): 3371-6.

Lescuyer, P., Hochstrasser, D. F. and Sanchez, J. C. (2004). "Comprehensive proteome analysis by chromatographic protein prefractionation." Electrophoresis **25**(7-8): 1125-35.

Lewis, J. G., Shand, B. I., Frampton, C. M., Elder, P. A. and Scott, R. S. (2008). "Plasma retinol-binding protein is not a marker of insulin resistance in overweight subjects: a three year longitudinal study." Clin Biochem **41**(13): 1034-8.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. and Yates, J. R., 3rd (1999). "Direct analysis of protein complexes using mass spectrometry." Nat Biotechnol **17**(7): 676-82.

Linke, T., Ross, A. C. and Harrison, E. H. (2004). "Profiling of rat plasma by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry, a novel tool for biomarker discovery in nutrition research." J Chromatogr A **1043**(1): 65-71.

Liolios, K., Mavromatis, K., Tavernarakis, N. and Kyripides, N. C. (2008). "The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata." Nucleic Acids Res **36**(Database issue): D475-9.

Liu, H., Sadygov, R. G. and Yates, J. R., 3rd (2004). "A model for random sampling and estimation of relative protein abundance in shotgun proteomics." Anal Chem **76**(14): 4193-201.

## **M.**

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J. and Apweiler, R. (2005). "PRIDE: the proteomics identifications database." Proteomics **5**(13): 3537-45.

Masoro, E. J. (2004). "Role of sirtuin proteins in life extension by caloric restriction." Mech Ageing Dev **125**(9): 591-4.

Merrell, K., Southwick, K., Graves, S. W., Esplin, M. S., Lewis, N. E. and Thulin, C. D. (2004). "Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry." J Biomol Tech **15**(4): 238-48.

Michopoulos, F., Lai, L., Gika, H., Theodoridis, G. and Wilson, I. (2009). "UPLC-MS-Based Analysis of Human Plasma for Metabonomics Using Solvent Precipitation or Solid Phase Extraction." J Proteome Res.

Mirgorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Korner, R., Sonksen, C. P. and Roepstorff, P. (2000). "Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards." Rapid Commun Mass Spectrom **14**(14): 1226-32.

Mrosovsky, N. and Sherry, D. F. (1980). "Animal anorexias." Science **207**(4433): 837-42.

## **N.**

Nedvidkova, J., Smitka, K., Kopsky, V. and Hainer, V. (2005). "Adiponectin, an adipocyte-derived protein." Physiol Res **54**(2): 133-40.

Nyomba, B. L., Bouillon, R., Lissens, W., Van Baelen, H. and De Moor, P. (1985). "1,25-Dihydroxyvitamin D and vitamin D-binding protein are both decreased in streptozotocin-diabetic rats." Endocrinology **116**(6): 2483-8.

## **O.**

Olsson, O. (1997). "Clutch abandonment: a state-dependent decision in king penguin." Journal of Avian Biology **28**(3): 264-267.

Omenn, G. S. (2004). "The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses." Proteomics **4**(5): 1235-40.

Omenn, G. S., Aebersold, R. and Paik, Y. K. (2009). "7(th) HUPO World Congress of Proteomics: launching the second phase of the HUPO Plasma Proteome Project (PPP-2) 16-20 August 2008, Amsterdam, The Netherlands." Proteomics **9**(1): 4-6.

Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W. and Hanash, S. M. (2005). "Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database." Proteomics **5**(13): 3226-45.

Orvisky, E., Drake, S. K., Martin, B. M., Abdel-Hamid, M., Resson, H. W., Varghese, R. S., An, Y., Saha, D., Hortin, G. L., Loffredo, C. A. and Goldman, R. (2006). "Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma." Proteomics **6**(9): 2895-902.

**P.**

Palagi, P. M., Hernandez, P., Walther, D. and Appel, R. D. (2006). "Proteome informatics I: bioinformatics tools for processing experimental data." Proteomics **6**(20): 5435-44.

**Q.**

Qi, Y., Takahashi, N., Hileman, S. M., Patel, H. R., Berg, A. H., Pajvani, U. B., Scherer, P. E. and Ahima, R. S. (2004). "Adiponectin acts in the brain to decrease body weight." Nat Med **10**(5): 524-9.

**R.**

Racaityte, K., Lutz, E. S. M., Unger, K. K., Lubda, D. and Boos, K. S. (2000). "Analysis of neuropeptide Y and its metabolites by high-performance liquid chromatography-electrospray ionization mass spectrometry and integrated sample clean-up with a novel restricted-access sulphonic acid cation exchanger." J Chromatogr A **890**(1): 135-44.

Rajcevic, U., Niclou, S. P. and Jimenez, C. R. (2009). "Proteomics strategies for target identification and biomarker discovery in cancer." Front Biosci **14**: 3292-303.

Ranes, J. and Stoller, J. K. (2005). "A review of alpha-1 antitrypsin deficiency." Semin Respir Crit Care Med **26**(2): 154-66.

Rieux, L., Bischoff, R., Verpoorte, E. and Niederlander, H. A. (2007). "Restricted-access material-based high-molecular-weight protein depletion coupled on-line with nano-liquid chromatography-mass spectrometry for proteomics applications." J Chromatogr A **1149**(2): 169-77.

Righetti, P. G., Boschetti, E., Lomas, L. and Citterio, A. (2006). "Protein Equalizer Technology : the quest for a "democratic proteome"." Proteomics **6**(14): 3980-92.

Righetti, P. G., Castagna, A., Herbert, B. and Candiano, G. (2005). "How to bring the "unseen" proteome to the limelight via electrophoretic pre-fractionation techniques." Biosci Rep **25**(1-2): 3-17.

Robin, J. P., Boucontet, L., Chillet, P. and Groscolas, R. (1998). "Behavioral changes in fasting emperor penguins: evidence for a "refeeding signal" linked to a metabolic shift." Am J Physiol **274**(3 Pt 2): R746-53.

Robin, J. P., Frain, M., Sardet, C., Groscolas, R. and Le Maho, Y. (1988). "Protein and lipid utilization during long-term fasting in emperor penguins." Am J Physiol **254**(1 Pt 2): R61-8.

Rosengren, A. T., Salmi, J. M., Aittokallio, T., Westerholm, J., Lahesmaa, R., Nyman, T. A. and Nevalainen, O. S. (2003). "Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels." Proteomics **3**(10): 1936-46.

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., Bartlett-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004). "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." Mol Cell Proteomics **3**(12): 1154-69.

Roux-Dalvai, F., Gonzalez de Peredo, A., Simo, C., Guerrier, L., Bouyssie, D., Zanella, A., Citterio, A., Burlet-Schiltz, O., Boschetti, E., Righetti, P. G. and Monsarrat, B. (2008). "Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry." Mol Cell Proteomics **7**(11): 2254-69.

## **S.**

Sadygov, R. G., Liu, H. and Yates, J. R. (2004). "Statistical models for protein validation using tandem mass spectral data and protein amino acid sequence databases." Anal Chem **76**(6): 1664-71.

Sennels, L., Salek, M., Lomas, L., Boschetti, E., Righetti, P. G. and Rappsilber, J. (2007). "Proteomic analysis of human blood serum using peptide library beads." J Proteome Res **6**(10): 4055-62.

Sihlbom, C., Kanmert, I., Bahr, H. and Davidsson, P. (2008). "Evaluation of the combination of bead technology with SELDI-TOF-MS and 2-D DIGE for detection of plasma proteins." J Proteome Res **7**(9): 4191-8.

Souverain, S., Rudaz, S. and Veuthey, J. L. (2004). "Restricted access materials and large particle supports for on-line sample preparation: an attractive approach for biological fluids analysis." J Chromatogr B Analyt Technol Biomed Life Sci **801**(2): 141-56.

## **T.**

Tamvakopoulos, C. (2007). "Mass spectrometry for the quantification of bioactive peptides in biological fluids." Mass Spectrom Rev **26**(3): 389-402.

Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P. and Veenstra, T. D. (2003). "Characterization of the low molecular weight human serum proteome." Mol Cell Proteomics **2**(10): 1096-103.

Tschoner, A., Sturm, W., Engl, J., Kaser, S., Laimer, M., Laimer, E., Weiss, H., Patsch, J. R. and Ebenbichler, C. F. (2008). "Retinol-binding protein 4, visceral fat, and the metabolic syndrome: effects of weight loss." Obesity (Silver Spring) **16**(11): 2439-44.

Tso, P. and Liu, M. (2004). "Apolipoprotein A-IV, food intake, and obesity." Physiol Behav **83**(4): 631-43.

## **V.**

Vleck, C. M., Bucher, T. L., Reed, W. L. and Kristmundsdottir, A. Y. (1999). "Changes in reproductive hormones and body mass through the reproductive cycle in the Adélie Penguin (*Pygoscelis adeliae*),

with associated data on courting-only individuals." Proceedings of the 22nd International Ornithology Congress: 1210–1223.

Vleck, C. M., Ross, L. L., Vleck, D. and Bucher, T. L. (2000). "Prolactin and parental behavior in Adelle penguins: effects of absence from nest, incubation length, and nest failure." Horm Behav **38**(3): 149-58.

von Eynatten, M. and Humpert, P. M. (2008). "Retinol-binding protein-4 in experimental and clinical metabolic disease." Expert Rev Mol Diagn **8**(3): 289-99.

## **W.**

Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W. and Dunn, M. J. (2006). "Guidelines for the next 10 years of proteomics." Proteomics **6**(1): 4-8.

Williams, A. (1995). "The Penguins." Oxford University press.

## **Z.**

Zhang, Q., Menon, R., Deutsch, E. W., Pitteri, S. J., Faca, V. M., Wang, H., Newcomb, L. F., Depinho, R. A., Bardeesy, N., Dinulescu, D., Hung, K. E., Kucherlapati, R., Jacks, T., Politi, K., Aebersold, R., Omenn, G. S., States, D. J. and Hanash, S. M. (2008). "A mouse plasma peptide atlas as a resource for disease proteomics." Genome Biol **9**(6): R93.

Zolotarjova, N., Mrozinski, P., Chen, H. and Martosella, J. (2008). "Combination of affinity depletion of abundant proteins and reversed-phase fractionation in proteomic analysis of human plasma/serum." J Chromatogr A **1189**(1-2): 332-8.





## **Conclusion générale**



L'objectif de ce travail de thèse était de développer des méthodologies d'analyse protéomique adaptées à des problématiques particulières et cela dans le but d'étendre les possibilités de l'analyse protéomique pour améliorer nos connaissances en biologie.

En effet, pour une partie des problèmes abordés par les biologistes et nécessitant des analyses protéomiques, il est apparu que les méthodologies de l'analyse protéomique actuelles devaient être finement adaptées au cas par cas pour qu'elles puissent apporter des réponses satisfaisantes. Ces méthodologies doivent donc évoluer et s'adapter au type de question posée (identifications de protéines constitutives de complexes, identification de protéines d'organismes dont le génome n'est pas encore séquencé, caractérisation fine de leur structure primaire, caractérisation fine de leur modifications post-traductionnelles, quantification du niveau d'expression de protéines, ...) mais aussi aux contraintes techniques imposées par la nature et le mode de préparation des échantillons (complexité, stabilité, solubilité...).

Nous avons donc sélectionné trois types de problématiques souvent rencontrées en analyse protéomique et qui posent presque toujours des problèmes de réalisation et d'interprétation :

- l'identification de protéines en interaction dans des complexes multiprotéiques, mais aussi dans des complexes plus originaux tels que les complexes mixtes protéines/ARN
- la caractérisation fine de modifications post-traductionnelles et plus précisément les phosphorylations,
- l'identification avec une grande confiance des protéines plasmatique d'organismes dont le génome n'est pas connu.

Les résultats suivants ont été obtenus pour ces trois types de problématiques :

**Dans la partie I : l'identification des protéines partenaires d'interaction d'une protéine ou d'un ARN cible.**

Une stratégie protéomique, impliquant une purification du complexe d'intérêt et des analyses protéomiques par nanoLC-MS/MS visant à la plus grande sensibilité, a pu être développée. Cette approche a permis de réduire l'impact des protéines non spécifiques tout en permettant l'identification de protéines faiblement abondantes au sein de mélanges complexes. Cette approche a été mise en œuvre dans 3 applications :

- Pour l'étude de partenaires d'interaction de la protéine Parp-2 (poly(ADP-ribose)polymérase-2) avec l'identification de trois protéines (Macro-H2A1.2., Hsp70.2 et TIF $\beta$ ), potentiellement intéressantes pour une meilleure compréhension du rôle fonctionnel de Parp-2 dans les mécanismes de spermatogenèse.
- Pour l'étude du mécanisme d'épissage relatif au site A7 de l'ARN de VIH-1, trois protéines identifiées par protéomique (hnRNP H, hnRNP K et la nucléoline) sont en cours de validation biologique en tant que régulateur de l'épissage alternatif.

- Pour la recherche de partenaires d'interaction protéiques de l'ARN Xist à partir d'extraits nucléaires de cellules ES murines à différents stades de différenciation, l'interaction entre l'ARN Xist cible et le complexe Polycomb 2 (SUZ12, EZH2, RBBP4, RBBP7) dont l'implication dans les mécanismes d'inactivation du chromosome X a été confirmée par les biologistes.

**Dans la partie II : développement d'une approche protéomique permettant de caractériser l'état de phosphorylation de protéines purifiées.**

Nous avons d'abord pu déterminer le taux de phosphorylation global des protéines grâce à des mesures de masses moléculaires réalisées sur les protéines entières. Dans un deuxième temps, nous avons déterminé le site exact de phosphorylation. Deux exemples sont présentés :

- Le domaine kinase de la protéine DAPK1 (Death – associated protein kinase), impliquée dans plusieurs voies de signalisation menant à la mort cellulaire, a été produit sous sa forme native sauvage et modifiée (délétion d'une boucle de 6 acides aminés). La protéine native forme un homodimère tandis que la protéine délétée reste monomérique. L'approche utilisée nous a permis de montrer un taux de phosphorylation global plus important pour la forme délétée. De plus, nous avons mis en évidence que les deux protéines ne sont pas modifiées au niveau du même acide aminé. La forme entière est phosphorylée en T35 et la forme délétée en T259.

- Le complexe CAK (pour cdk-activating kinase) est composé de trois protéines (CDK7, Cycline H et MAT1). La méthodologie développée nous a permis de connaître les variations du taux de phosphorylation au niveau des protéines entières en fonction du type de cellules utilisées pour la production du complexe. Le taux de phosphorylation diffère pour les deux lignées cellulaires comparées (Sf9 et HighFive). Les sites de phosphorylations des trois protéines ont été localisés sur MAT1 (S279), sur la Cycline H (T315 et S323) ainsi que sur CDK7 (S164 et T170).

**Dans la partie III : identification des protéines plasmatiques différemment exprimées selon l'état métabolique et nutritionnel du manchot Adélie dont le génome n'a pas été séquencé.**

Deux difficultés majeures ont été rencontrées dans ce projet. La première concerne la complexité bien connue des échantillons plasmatiques que nous avons résolue grâce à un prétraitement adapté. La deuxième est liée à l'absence de séquences génomiques annotées du manchot. Nous avons montré que des logiciels experts permettaient de réaliser à haut débit du séquençage *de novo* (séquençage automatisé) et donc l'identification de nombreuses protéines par homologie de séquence. Les analyses d'images des gels d'électrophorèse d'extraits plasmatiques ont permis de visualiser plus de 400 spots et 50 protéines uniques ont été identifiées. Dix protéines ont été mises en évidence comme différemment exprimées selon l'état métabolique de l'animal durant le jeûne. Ces modifications pourraient refléter des adaptations spécifiques qui permettraient au manchot de supporter la privation totale de nourriture pendant de longues périodes et ceci de façon répétée tout au long de sa vie, avec le minimum d'effets délétères. Notre travail a donc permis une meilleure compréhension des mécanismes moléculaires médiant l'adaptation au jeûne prolongé chez

le manchot Adélie, ce qui va constituer les bases de futures études des liens possibles entre jeune prolongé, réalimentation d'organismes dénutris, obésité et diabète chez l'homme.

L'ensemble de mes travaux de thèse montre l'universalité des approches protéomiques développées dans la littérature. Cependant, elles devront toujours être adaptées très finement pour garder leur potentiel pour l'analyse d'échantillons provenant d'origines diverses et de complexités variables. Ceci nécessite de la part des laboratoires spécialisés en analyse protéomique une maîtrise de l'ensemble des outils de la protéomique : préparation de l'échantillon, séparation des protéines, analyse par spectrométrie de masse, analyse bioinformatique des données MS et MS/MS pour une identification la plus sûre et complète des protéines.



## **Partie Expérimentale**





## **1. Préparation d'échantillon**

### **1.a. Production des protéines recombinantes**

*Protéines utilisées pour les études de phosphorylations – Partie II.*

#### **1.a.1) Production de DAPK1 WT et Delta-DAPK1**

Les protéines DAPK1 WT et Delta-DAPK1 sont produites à partir du plasmide pASK-IBA3 (Sigma) et d'un plasmide mutant réalisé par D. Martin Watterson (Northwestern University, Chicago IL.) dans des cellules d'*Escherichia Coli* (DH5 $\alpha$ ). Après production et lyses des cellules, les protéines sont purifiées à l'aide d'une résine streptactin (High capacity streptactin superflow resin - IBA Göttingen). Les tampons sont ensuite échangés par filtration sur membranes avec une limite de poids de 10kDa (Satorius).

#### **1.a.2) Production du complexe CAK**

La production du complexe est réalisée à l'aide du système de vecteur d'expression baculovirus (BEVES). Le vecteur utilisé est un vecteur poly-cistronique contenant les séquences des protéines CDK7-Strep, Cycline H et MAT1. Les lignées cellulaires utilisées pour la production sont des lignées d'insectes : *Spodoptera frugiperda* Sf9 et BTI-TN5B1-4, nommé Hi5 dans le manuscrit. Après production et lyses des cellules, le complexe est purifié à l'aide d'une résine streptactin (IBA Göttingen).

### **1.b. Préparation d'échantillon et séparation des protéines par électrophorèse**

#### **1.b.1) Déplétion d'albumine et précipitation des protéines**

*Préparation réalisée pour les échantillons de plasma de manchot – Partie III.*

La déplétion d'albumine est réalisée avec une colonne contenant 400  $\mu$ l de résine HSA-Affinity sépharose (Vivascience). 50  $\mu$ l de plasma et 150  $\mu$ l de tampon (Binding buffer PBS) sont déposés sur la résine et incubés 30 min à température ambiante sous agitation. Après centrifugation (2 min, 400 g), le surnageant est récupéré. La résine est lavée avec 200  $\mu$ l de tampon PBS pendant 30 min à température ambiante sous agitation et centrifugée 2 min à 400 g. Les deux surnageants sont réunis. Les protéines sont ensuite précipitées avec 50  $\mu$ l d'acide trifluoroacétique (TCA) + 50  $\mu$ l d'eau puis incubées pendant 30 min à température ambiante sous agitation. Après centrifugation (10 min, 1500 g), 500  $\mu$ l d'acétone sont ajoutés au culot obtenu et laissés à -20°C toute la nuit. Le culot est dissout dans une solution de Chaps 4 %, DTT 1 %, Urée 8 M, Thiourée 2 M, TLCK 0.5  $\mu$ g/ $\mu$ l, triton 0.5 %, Ampholytes 3-10 2 % et un cocktail d'inhibiteur de protéase.

### **1.b.2) Préparation des gels d'électrophorèses**

*Préparation réalisée pour les échantillons de plasma de manchot – Partie III.*

La première dimension est réalisée sur un strip de pH 3 à pH 10 non linéaire. Le strip est réhydraté avec l'échantillon d'abord passivement, puis activement en appliquant 50 V pendant 16 h. La focalisation est ensuite réalisée pendant 28 h et avec un total de 85000 Vh.

Les protéines sont réduites en présence du tampon d'équilibration additionné de DTT 1 % pendant 30 min puis alkylées pendant 30 min avec le tampon d'équilibration additionné d'iodoacétamide 2.5 %. Le tampon d'équilibration contient du SDS 2 %, de l'urée 6 M, du TrisHCl pH 8.8 50 mM et du Glycérol 30 %.

Un gel d'électrophorèse en gradient continu 7-18 % en polyacrylamide est réalisé. Le strip contenant les protéines est déposé en haut du gel, ainsi que des marqueurs protéiques (Sigmamarker). La migration du gel est réalisée avec un ampérage de 9mA jusqu'à ce que la migration arrive à l'extrémité du gel. La coloration des gels est réalisée au bleu de coomassie colloïdal G250 (Sigma).

### **1.b.3) Découpe et traitement des morceaux de gel**

*Traitements réalisés pour les identifications des protéines issues de complexes - Partie I - et des protéines de plasma de manchot – Partie III.*

Les spots ou bandes de gel contenant les protéines sont découpés. Le robot de préparation MassPREP Station (Micromass Ltd.) est utilisé pour le lavage, la réduction et l'alkylation des morceaux de gel. Les morceaux de gel sont lavés par 100 µL avec une solution ACN/NH<sub>4</sub>HCO<sub>3</sub> 25 mM pendant 10 min. Cette opération est répétée 2 fois. Les spots sont ensuite déshydratés par 50 µL d'ACN pendant 5 min. La solution réductrice est une solution de dithiothreitol 10 mM dans NH<sub>4</sub>HCO<sub>3</sub> 25 mM. 50 µL de DTT sont déposés sur les morceaux de gel pendant 1 h à 57°C. 50 µL d'une solution alkylante d'iodoacétamide 55 mM dans NH<sub>4</sub>HCO<sub>3</sub> 25 mM, sont ajoutés. Après 20 min de réaction d'alkylation, l'excédent est éliminé, puis les morceaux de gels sont lavés pour éliminer les restes de DTT et d'iodoacétamide et ils sont ensuite déshydratés pour recevoir l'enzyme de digestion.

## **1.c. Traitements enzymatiques**

### **1.c.1) Digestion enzymatique**

- Enzymes de digestion

Deux enzymes de digestions ont été utilisées durant ma thèse la trypsine (Promega) et la chymotrypsine (Roche). Les enzymes sont reprises dans NH<sub>4</sub>CO<sub>3</sub> 25 mM pH 7 pour arriver à une concentration de 12.5 ng/µl pour la trypsine et 25 ng/µl pour la chymotrypsine. Les digestions se déroulent pendant 12 h avec la trypsine à 37°C et à température ambiante pour la chymotrypsine.

- Digestion en solution (*réalisées pour les études de phosphorylation - Partie II*)

Les protéines sont dénaturées en ajoutant un volume d'acétonitrile correspondant à 10 % du volume final de l'échantillon pendant 1 h. L'enzyme de digestion est ajouté à l'échantillon en respectant un ratio enzyme/substrat de 1/50.

- Digestion "in gel" (*réalisées pour les identifications des protéines issues de complexes - Partie I - et des protéines de plasma de manchot – Partie III*).

Les morceaux de gel déshydratés contenant les protéines vont aspirer la solution d'enzyme de digestion. L'extraction des différents peptides générés par la digestion se fait par l'ajout d'une solution H<sub>2</sub>O/ACN/HCCOH (39/60/1 v/v/v). Les peptides vont diffuser de façon passive des différents morceaux de gel. L'excès d'acétonitrile est évaporé et le mélange peptidique est analysé par spectrométrie de masse.

### **1.c.2) Traitement pour la déphosphorylation**

*Traitement réalisé pour les études de phosphorylation - Partie II.*

Le traitement avec la phosphatase alcaline (Cat. No. 713 023, Roche Diagnostics) permet de déphosphoryler les peptides ou les protéines.

- Déphosphorylation des protéines:

Pour la déphosphorylation des protéines du complexe CAK, la phosphatase alcaline a été utilisée en ajoutant 5 µl de 1 U/µL de phosphatase alcaline, pH 7.8 à 10 µl de 0.4 mg/ml de complexe CAK. La réaction de déphosphorylation se déroule pendant 8 h à 37°C.

- Déphosphorylation des peptides:

Le mélange peptidique (5 µl) est incubé en eppendorf en présence de 2 µl / 0.4 U de phosphatase alcaline (solution de 5 µl (1 U/µL) d'alcaline phosphatase est reprise par 20 µl NH<sub>4</sub>CO<sub>3</sub> 25 mM pH 7) pendant une heure à 37°C. L'échantillon est ensuite acidifié avec une solution à 0.1 % TFA jusqu'à pH 3.

## 2. Analyses par spectrométrie de masse

### 2.a. Analyses MS

#### **2.a.1) Analyses LC-MS (MicroLC-MicroTOF)**

*Analyses réalisées pour les déterminations du taux de phosphorylations du complexe CAK - Partie II.*

Les analyses LC-MS sont réalisées à l'aide d'un système HPLC Agilent 1100 series HPLC (Agilent Technologies) couplé à un spectromètre de masse hybride quadrupole temps de vol MicroTOF (Bruker Daltonics). Les échantillons sont piégés sur d'une pré-colonne C4 de taille 10 mm x 1 mm, d'une porosité de 5 µm et les protéines sont séparées à l'aide d'une colonne C4 de taille 150 mm x 0.5 mm de granulométrie 300 Å et d'une porosité de 5 µm (Grace Vydac). Le système de solvants est composé par le solvant A (eau acidifié par 0.1 % TFA) et le solvant B (acétonitrile acidifié par 0.067 % de TFA). Le piégeage des protéines est réalisé durant 3 minutes à un débit de 15 µl/min avec 100 % de solvant A. La chromatographie des protéines est réalisée à un débit de 100 µl/min à 40°C avec un gradient linéaire allant de 20 % à 85 % de solvant B en 40 minutes suivi de 85 % de solvant B durant 10 minutes.

Le spectromètre de masse (MicroTOF) est étalonné par le mélange de composés de synthèse contenu dans le "Tuning Mix " LC/MSD ion trap G2431A (Agilent Technologies). La tension optimale appliquée à l'entrée du capillaire est de -4500 volts et la gamme de masse est fixée de 300 m/z à 3000 m/z. Le système de chromatographie ainsi que le spectromètre de masse sont contrôlés par le logiciel HyStar version 3.2.44.0 (Bruker Daltonics). Les données obtenues au cours de l'analyse sont traitées par le logiciel DataAnalysis 3.4 Build 192 (Bruker Daltonics). La déconvolution des spectres de masse des protéines est réalisée selon les paramètres par défaut du logiciel DataAnalysis 3.4 Build 192. Le lissage des spectres de masse est réalisé à l'aide de l'algorithme de Savitzky Golay avec une largeur de lissage de 0.2 m/z pour 1 cycle.

#### **2.a.2) Analyses MALDI –MS (Ultraflex™ MALDI-TOF/TOF)**

*Analyses réalisées pour les identifications des phosphopeptides du complexe CAK - Partie II.*

Les mesures de masse sont réalisées à l'aide de spectromètre de masse MALDI-TOF (Ultraflex™ MALDI-TOF/TOF, Bruker Daltonics). L'instrument travaille en mode réflectron positif avec une tension de 25 kV appliquée à la cible et une tension 26 kV appliqué au réflectron. La valeur de l'extraction retardée a été optimisée à 110 ns afin d'obtenir la meilleure résolution sur les peptides servant à la calibration. Le laser utilisé est un laser à azote (337 nm) travaillant à une fréquence de 20 Hz. L'acquisition des données est réalisée dans une gamme de masse allant de 400 m/z à 4000 m/z, la valeur de suppression de matrice étant fixée à 500 m/z. Le spectromètre de masse est étalonné à

l'aide d'un mélange contenant 7 peptides (Bruker Peptide Calibration Standard #206196, Bruker Daltonics) (bradykinin 1-7 ( $m/z=757.400$ ), human angiotensin II ( $m/z=1046.542$ ), human angiotensin I ( $m/z=1296.685$ ), substance P ( $m/z=1347.735$ ), bombesin ( $m/z=1619.822$ ), renin ( $m/z=1758.933$ ), ACTH 1-17 ( $m/z=2093.087$ ) and ACTH 18-39 ( $m/z=2465.199$ )). Les spectres de masses obtenus sont traités par le logiciel FlexAnalysis 2.4 build 11 (Bruker Daltonics) utilisant les paramètres par défaut et l'algorithme SNAP pour la détection du pic monoisotopique.

0.5  $\mu$ l d'échantillon et 0.5  $\mu$ l de matrice sont déposés, mélangés sur la cible et séchés à pression atmosphérique selon la technique de la goutte séchée. Deux matrices sont utilisées, la  $\alpha$ -cyano-4-hydroxycinnamic (HCCA) (Bruker #203072) et le 2,5 acide di-hydrobenzoïque (DHB) (Fluka #37550). Une solution saturée de matrice HCCA dans 50/50 H<sub>2</sub>O/ACN est diluée trois fois par une solution H<sub>2</sub>O/ACN 50/50 acidifiée par 0.1 % de TFA. La matrice DHB est préparée à la concentration de 10 mg/ml dans une solution H<sub>2</sub>O/ACN 50/50.

## **2.b. Analyses MS/MS**

Afin de contrôler le bon fonctionnement du système de chromatographie ainsi que celui du spectromètre de masse, des paramètres de mesures des performances des 2 instruments ont été définies. Ces tests sont réalisés avec un mélange de peptides issu de la digestion de l'albumine sérique bovine (BSA) par de la trypsine. Ces tests permettent de contrôler et de suivre l'état de la colonne, la sensibilité et la résolution du spectromètre de masse. Les paramètres de validations sont différents pour chaque couplage.

### **2.b.1) Analyses nanoLC-MS/MS réalisées avec le couplage CapLC-QTOF**

*Analyses réalisées pour les identifications des protéines issus de complexes protéine/protéine et ARN/protéine - Partie I.*

Les analyses nanoLC-MS/MS sont réalisées à l'aide d'une chaîne CapLC (Micromass Ltd.) couplé à un Q-TOF 2 (Micromass Ltd.).

- Système nanoLC CapLC :

Le système HPLC utilisé est une CapLC (Micromass Ltd.) qui permet l'injection automatique des échantillons et leur concentration sur une pré-colonne. L'échantillon injecté est concentré et dessalé sur une pré-colonne de 1mm de long et 300  $\mu$ m de diamètre interne remplie d'une phase stationnaire de type PepMap C18 de granulométrie 5  $\mu$ m et de porosité de 100 Å (LC-Pakings) à un débit de 15  $\mu$ l/min durant 3 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

La pré-colonne est ensuite mise automatiquement à l'aide du "Stream Select" (Micromass Ltd.) en ligne avec la colonne analytique. Les peptides retenus sur la pré-colonne sont élués à contre sens

vers la colonne analytique. Les peptides sont séparés sur la colonne analytique de phase inverse de 15 cm long et 75 µm de diamètre interne remplie d'une phase stationnaire de type PepMap C18 de granulométrie 3 µm et de porosité 100 Å (LC-Packings).

La phase mobile A est une solution d'eau acidifiée par 0,1 % d'acide formique, la phase mobile B est une solution d'acétonitrile acidifiée par 0,1 % d'acide formique. L'élution des peptides est réalisée à un débit de 200 nl/min par un gradient de 5 à 45 % de solvant B durant 35 minutes, puis 95 % de solvant B durant 15 minutes. La colonne est ré-équilibrée 20 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

- Spectromètre de masse Q-TOF2 :

Les analyses MS et MS/MS sont réalisées sur un spectromètre de masse de type Q-TOF 2 (Micromass Ltd.) qui se compose d'une source à ionisation électrospray (Z-Spray) en mode positif, d'un quadrupole jouant un rôle de filtre de masse, d'un hexapôle qui est la cellule de collision et d'un tube de vol placé de manière orthogonal qui est l'analyseur. Une tension de 3500 V est appliquée au capillaire et une tension de 35 V au cône. La calibration du spectromètre de masse est réalisée à l'aide d'une solution d'acide phosphorique à 0.01 % dans l'eau sur une gamme de masse de 50 à 2000 m/z. L'acquisition des données est réalisée selon un programme de basculement automatique du mode MS au mode MS/MS (MS 1 sec/scan de 250 à 1500 m/z et MS/MS 3 sec/scan de 50 à 2000 m/z). Le programme sélectionne les trois peptides doublement ou triplement chargés ayant les signaux les plus intenses afin de les isoler et les fragmenter. La fragmentation dans la cellule de collision est réalisée avec de l'argon. Le système est piloté par le logiciel MassLynx 4.0 (Micromass Ltd.).

La vérification des performances du système HPLC et du spectromètre de masse est réalisée avec 200 fmol de digeste de BSA. Le système est considéré comme fonctionnel si il répond aux caractéristiques suivantes en mode MS :

m/z	507,8	582,3
Intensité BPI	> 400	> 400
Largeur du pic à mi-hauteur	18 ± 5 s	15 ± 5 s
Résolution isotopique	> 3000	> 3000

En mode MS/MS, le critère bon fonctionnement du système de masse est un pourcentage de couverture de séquence de la BSA supérieur à 40 % selon les paramètres suivants :

Recherche dans la banque de données Swiss-Prot à partir du moteur de recherche Mascot avec les critères de recherche suivants :

- Enzyme : trypsine/P
- Miss cleavage : 1
- Modifications : carbamidomethyl (C) et oxydation (M)
- Etat de charge : +1, +2, +3
- Tolérance de masse MS et MS/MS : ± 0.3 Da

### **2.b.2) Analyses nanoLC-MS/MS réalisées avec le couplage NanoAcquity-Synapt**

*Analyses réalisées pour les identifications des phosphopeptides de DAPK et du complexe CAK - Partie II.*

Les analyses nanoLC-MS/MS sont réalisées à l'aide d'une chaîne nanoAcquity (Waters Corp.) couplée à un Synapt (Waters Corp.).

- Système nanoLC NanoAcquity :

Le système HPLC utilisé est une nanoAcquity (UPLC, Waters Corp.) qui permet de travailler à une pression maximale de 10000 psi avec des colonnes contenant des particules de petit diamètre (< 2 µm). L'échantillon injecté est concentré et dessalé sur une pré-colonne de 20 mm de long et 180 µm de diamètre interne remplie d'une phase stationnaire de type Symetry C18 de granulométrie 5 µm et de porosité de 100 Å (Waters Corp.) à un débit de 5 µl/min durant 3 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

La pré-colonne est ensuite mise automatiquement en ligne avec la colonne analytique. Les peptides retenus sur la pré-colonne sont élués vers la colonne analytique. Les peptides sont séparés sur la colonne analytique de 20 cm long et 75 µm de diamètre interne remplie d'une phase stationnaire de type C18 ACQUITY UPLC® BEH130 de granulométrie 1,7 µm et de porosité 130 Å (Waters Corp.). La phase mobile A est une solution d'eau acidifiée par 0,1 % d'acide formique, la phase mobile B est une solution d'acétonitrile acidifiée par 0,1 % d'acide formique. L'éluion des peptides est réalisée à un débit de 400 nl/min par un gradient de 1 à 40 % de solvant B durant 35 minutes à la température de 45°C, puis 65 % de solvant B durant 5 minutes. La colonne est ré-équilibrée 10 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

- Spectromètre de masse Synapt :

Les analyses MS et MS/MS sont réalisées sur un spectromètre de masse de type Synapt HDMS (Waters Corp.) qui se compose d'une source à ionisation électrospray (Z-Spray) en mode positif, d'un quadropole, d'une cellule de collision ("T-waves") et d'un tube de vol placé de manière orthogonal. Une tension de 3500 V est appliquée au capillaire et une tension de 35 V au cône. La calibration du spectromètre de masse est réalisée à l'aide d'une solution d'acide phosphorique à 0,01 % dans l'eau sur une gamme de masse de 50 à 2000 m/z. Le spectromètre de masse est équipé d'un système dit de "Lock Mass" qui permet d'injecter à des temps définis une référence. La référence choisie est le [Glu<sup>1</sup>]-fibrinopeptide B (GFP) L'ion (M+2H)<sup>2+</sup> de m/z 785,8426 est utilisé pour étalonner les données MS et l'ion (M+H)<sup>+</sup> de m/z 684,3469 est utilisé pour les données MS/MS.

L'acquisition des données est réalisée selon un programme de basculement automatique du mode MS au mode MS/MS (MS 0,5 sec/scan de 250 à 1500 m/z et MS/MS 0,6 sec/scan de 50 à

2000 m/z). Le programme sélectionne les trois peptides doublement ou triplement chargés ayant les signaux les plus intenses afin de les isoler et les fragmenter. La fragmentation dans la cellule de collision est réalisée avec de l'argon. Le système est piloté par le logiciel MassLynx 4.1 (SCN 566, Waters Corp.).

Les données obtenues au courant de l'analyse sont traitées et converties en liste de valeur de masse (pkI) exploitable par des logiciels de recherches dans les banques de données. Le traitement est réalisé par ProteinLynx Browser 2.3 (Waters Corp.). Le bruit de fond est soustrait selon un algorithme avec un seuil de 5 %.

La vérification des performances du système HPLC et du spectromètre de masse est réalisée avec 50 fmol de digeste de BSA. Le système est considéré comme fonctionnel si il répond aux caractéristiques suivantes en mode MS :

m/z	<b>722,33</b>	<b>464,25</b>
Intensité du pic BPI	> 10 <sup>3</sup>	> 10 <sup>3</sup>
Largeur à mi-hauteur du pic BPI	< 0.15 min	< 0.18 min
Résolution isotopique	> 9000	> 9000

En mode MS/MS, le critère de bon fonctionnement du système de masse est un pourcentage de couverture de séquence de la BSA supérieur à 45 % selon les paramètres suivants :

Recherche dans la banque de données Swiss-Prot à partir du moteur de recherche Mascot avec les critères de recherche suivants :

- Enzyme : trypsine/P
- Miss cleavage : 1
- Modifications : carbamidomethyl (C) et oxydation (M)
- Etat de charge : +1, +2, +3
- Tolérance de masse MS et MS/MS : 25 ppm et 0.05 Da

### ***2.b.3) Analyses nanoLC-MS/MS réalisées avec le couplage nanoLC-ChipCube–Trappe ionique (HCT plus ou HCT Ultra)***

*Analyses réalisées pour les caractérisations des phosphopeptides de DAPK et du complexe CAK - Partie II - et les identifications des protéines plasmatiques de Manchot Adélie - Partie III.*

Les analyses nanoLC-MS/MS sont réalisées à l'aide d'une chaîne HPLC 1100 équipé d'un système nanoLC-ChipCube (Agilent Technologies) couplé à une trappe ionique HCT plus ou HCT Ultra (Bruker Daltonics).



- Système nanoLC-ChipCube :

Le système Chip cube (Agilent Technologies) est constitué d'une pré-colonne d'enrichissement et d'une colonne intégrée sur un support, placé dans le cube fixé au spectromètre de masse.

La pré-colonne d'enrichissement contient une phase inverse C18 Zorbax d'une granulométrie de 5 µm, de porosité 300 Å et d'un volume de 40 nL. La colonne de séparation des peptides est composée par la même phase que la pré-colonne d'enrichissement, mais existe en 2 dimensions 43 mm x 75 µm ou 150 mm x 75 µm.

*Les chips utilisées pour la caractérisation des phosphorylations sont des chip de 150 mm et celles pour l'identification des protéines plasmatiques de manchot de 43 mm.*

L'échantillon est chargé sur la pré-colonne à un débit de 3,75 µl/min durant 3 minutes à l'aide d'une solution aqueuse avec 2 % d'acétonitrile acidifiée par 0,1 % d'acide formique.

La phase mobile A est une solution aqueuse avec 2 % d'acétonitrile acidifiée par 0,1 % d'acide formique, la phase mobile B est une solution d'acétonitrile avec 2 % d'eau acidifiée par 0,1 % d'acide formique. La séparation des peptides est réalisée à un débit de 300 nl/min pour les 2 types de colonnes, un gradient linéaire de 8 % à 40% de solvant B en 7 minutes pour la colonne de 43 mm et un gradient linéaire 8 % à 40 % en 30 minutes pour la colonne de 150 mm.

- Spectromètre de masse Trappe ionique (HCT plus ou HCT Ultra) :

La trappe ionique HCT plus ou HCT ultra (Bruker Daltonics) opère en mode électrospray positif. La tension appliquée à l'extrémité extérieure du capillaire de transmission est de -1750 V. L'étalonnage est réalisé à l'aide du mélange "TuneMix" G2431A (Agilent Technologies). L'acquisition des données MS est réalisée à une vitesse de 8100 m/z par seconde dans une gamme de masse de 250 m/z à 2500 m/z. La quantité d'ions dans la trappe (Ion Charge Control) est fixée à 100000 avec un temps d'accumulation maximal de 200 ms et un spectre de masse MS correspond à la moyenne de 4 balayages (rolling average 2).

L'acquisition des données est réalisée selon un programme de basculement automatique du mode MS au mode MS/MS. Le programme sélectionne les trois peptides doublement ou triplement chargés ayant les signaux les plus intenses avec une fenêtre d'isolation de 4 Da afin de les fragmenter. L'acquisition des données MS/MS est réalisée à une vitesse de 26000 m/z par seconde dans une gamme de masse de 50 m/z à 2800 m/z. La quantité d'ions dans la trappe (Ion Charge Control) est fixée à 300000 et un spectre de masse MS/MS correspond à la moyenne de 6 balayages. La rampe d'énergie utilisée pour la fragmentation est de 1.5 V dont l'amplitude varie de 30 à 200 %. Le système est piloté par le logiciel EsquireControl 5.3 Build 11.0 (Bruker Daltonics). Les données obtenues au cours de l'analyse sont traitées et converties en liste de valeur de masse (mgf) exploitable par des logiciels de recherches dans les banques de données. Le traitement des données est réalisé par le logiciel DataAnalysis 3.3 Build 146 (Bruker Daltonics). Le bruit de fond est soustrait

selon l'algorithme de Savitzky Golay avec une largeur de lissage de 0,2 m/z pour 1 cycle et un seuil de 60000.

La vérification des performances du système HPLC et du spectromètre de masse est réalisée avec 200 fmol de digeste de BSA. Le système est considéré comme fonctionnel si il répond aux caractéristiques suivantes en mode MS :

Chip 43 mm

A partir des données MS : il faut obtenir les caractéristiques suivantes pour les 2 ions demandés :

m/z	722.4	464,3
Temps de rétention	1,5 ± 1 min	6,5 ± 1 min
Intensité	> 10 <sup>7</sup>	> 10 <sup>7</sup>
Largeur à mi-hauteur	0.2 ± 0.1min	0.2 ± 0.1min
Résolution isotopique	> 2000	> 1500

Chip 150mm

A partir des données MS : il faut obtenir les caractéristiques suivantes pour les 2 ions demandés :

m/z	722.4	464,3
Temps de rétention	3.5 ± 1 min	12 ± 1 min
Intensité	> 10 <sup>7</sup>	> 10 <sup>7</sup>
Largeur à mi-hauteur	0.2 ± 0.1min	0.2 ± 0.1min
Résolution isotopique	> 2000	> 1500

En mode MS/MS le critère de bon fonctionnement du système de masse est un pourcentage de couverture de séquence de la BSA d'au moins 40 % pour les colonnes de 43 mm et 45 % ou plus pour les colonnes de 150 mm selon les paramètres suivants :

Recherche dans la banque de données Swiss-Prot à partir du moteur de recherche Mascot avec les critères de recherche suivants :

- Enzyme : trypsine/P
- Miss cleavage : 1
- Modifications : carbamidomethyl (C) et oxydation (M)
- Etat de charge : +1, +2, +3
- Tolérance de masse MS et MS/MS : ± 0.3 Da

## **Annexes**



## Annexe 1 :

Liste des protéines identifiées dans la bande de gel correspondant à l'IP anti-Parp-2

Protéines	Taxonomie	N° Accession	M (Da)	Nbr de peptides
<b>Parp</b>				
Poly [ADP-ribose] polymerase 1 (PARP-1)	Rattus norvegicus	P27008	112458	13
Poly [ADP-ribose] polymerase 2 (PARP-2)	Mus musculus	O88554	63356	3
<b>Histones et variants</b>				
Histone H1.1 (H1 VAR.3) (H1a)	Mus musculus	P43275	21641	13
Histone H1.2 (H1 VAR.1) (H1c)	Mus musculus	P15864	21123	9
Histone H1.3 (H1 VAR.4) (H1c)	Mus musculus	P43277	21955	7
Histone H1.4 (H1 VAR2) (H1e)	Mus musculus	P43274	21833	5
Histone H1.5 (H1 VAR.5) (H1b)	Mus musculus	P43276	22431	3
Histone H1t (Testicular H1 histone)	Mus musculus	Q07133	21396	3
Core histone macro-H2A.1 (Histone macroH2A1)	Mus musculus	Q9QZQ8	39579	11
<b>Protéines impliquées dans la régulation transcriptionnelle</b>				
Transcription intermediary factor 1-beta (TIF1-beta)	Mus musculus	Q62318	88791	6
DNA damage-binding protein 1 (DDB1)	Mus musculus	Q16531	126887	5
ATP-dependent RNA helicase A (Nuclear DNA helicase II)	Mus musculus	Q08211	140792	5
Cullin-associated NEDD8-dissociated protein 1	Mus musculus	Q6ZQ38	136245	10
<b>Protéines de la famille des Heat Shock protein (Hsp)</b>				
Heat shock 70 kDa protein 4	Mus musculus	Q61316	94073	8
Heat shock 70 kDa protein 4L	Mus musculus	P48722	94322	13
Heat shock protein 70 A1	Anopheles albimanus	P41825	70208	4
60 kDa heat shock protein, mitochondrial precursor (Hsp60)	Mus musculus	P63038	60917	6
Heat shock protein HSP 90-alpha (HSP 86)	Mus musculus	P07901	84604	18
Heat shock-related 70 kDa protein 2 (Hsp 70.2)	Rattus norvegicus	P17156	69698	5
Heat shock-related 70 kDa protein 2 (Hsp 70.2)	Mus musculus	P14659	69486	25
78 kDa glucose-regulated protein precursor (GRP 78)	Mus musculus	P20029	72377	17
Stress-70 protein, mitochondrial precursor (GRP 75)	Mus musculus	P38647	73483	4
Endoplasmic precursor (GRP 94)	Mus musculus	P08113	92418	12
<b>Autres protéines</b>				
DNA polymerase beta (EC 2.7.7.7)	Mus musculus	Q8K409	38133	6
Importin beta-1 subunit	Mus musculus	P70168	97090	4
Importin beta-3	Mus musculus	Q8BKC5	123511	7
Importin-alpha re-exporter	Mus musculus	Q9ERK4	110384	4
Tubulin alpha-1 chain	Mus musculus	P68369	50104	4
Tubulin alpha-2 chain	Mus musculus	P05213	50120	2
Tubulin beta-5 chain	Mus musculus	P99024	49639	7
Myosin Id	Mus musculus	Q5SYD0	116007	6

Protéines	Taxonomie	N° Accession	M (Da)	Nbr de peptides
Myosin I $\alpha$	Mus musculus	Q9WTI7	118082	3
Myosin-9	Mus musculus	Q8VDD5	226086	19
Myosin-11	Mus musculus	O08638	226888	44
High affinity immunoglobulin gamma Fc receptor I precursor	Mus musculus	P26151	44860	3
Ig gamma-1 chain C region, membrane-bound form	Mus musculus	P01869	43359	4
Ig gamma-2B chain C region, membrane-bound form	Mus musculus	P01867	44231	3
Ig kappa chain C region	Mus musculus	P01837	11771	3
Ig kappa chain V-II region 26-10	Mus musculus	P01631	12265	4
Ig kappa chain V-II region MIL	Homo sapiens	P01616	12048	3
Ig kappa chain V-III region CBPC 101	Mus musculus	P01664	11957	3
Ig kappa chain V-III region PC 2880/PC 1229	Mus musculus	P01654	11973	3
Ig kappa chain V-III region PC 7175	Mus musculus	P01671	12003	4
Ig kappa chain V-III region TEPC 124	Mus musculus	P01629	12331	3
Ig kappa chain V-V region K2 precursor (Fragment)	Mus musculus	P01635	12573	2
Ig kappa chain V-V region MPC11 precursor	Mus musculus	P01633	16423	3
14-3-3 protein epsilon (14-3-3E)	Mus musculus	P62259	29155	3
14-3-3 protein zeta/delta	Mus musculus	P63101	24454	6
26S proteasome non-ATPase regulatory subunit 2	Mus musculus	Q8VDM4	100139	3
2-oxoglutarate dehydrogenase E1 component, mitochondrial precursor (EC 1.2.4.2)	Mus musculus	Q60597	116043	3
40S ribosomal protein S2 (S4) (LLRep3 protein)	Mus musculus	P25444	31212	3
40S ribosomal protein SA (p40)	Mus musculus	P14206	32567	2
52 kDa Ro protein (Sjogren syndrome type A antigen)	Mus musculus	Q62191	54140	3
60S acidic ribosomal protein P0 (L10E)	Mus musculus	P14869	34195	5
60S ribosomal protein L6 (TAXREB107)	Mus musculus	P47911	33358	3
Actin, cytoplasmic 1 (Beta-actin)	Mus musculus	P60710	41710	14
Aldose reductase (EC 1.1.1.21)	Mus musculus	P47911	33358	8
Annexin A2	Mus musculus	P07356	38521	11
Annexin A4	Mus musculus	P97429	35836	2
Annexin A5	Mus musculus	P48036	35730	4
Annexin A11	Mus musculus	P97384	54077	2
Clathrin heavy chain	Mus musculus	Q68FD5	191435	28
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	Rattus norvegicus	P08461	58727	6
Dipeptidase 3 precursor (EC 3.4.13.19)	Mus musculus	Q9DA79	54213	6
DnaJ homolog subfamily C member 9	Mus musculus	Q91WN1	30040	4
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 67 kDa subunit precursor	Mus musculus	Q91YQ5	68486	5
Dr1-associated corepressor	Mus musculus	Q9D6N5	22133	2
Dynein heavy chain, cytosolic (DYHC)	Rattus norvegicus	P38650	531916	18
Elongation factor 1-gamma (EF-1-gamma)	Mus musculus	Q9D8N0	49898	3
Elongation factor 2 (EF-2)	Mus musculus	P58252	95122	4
Eukaryotic initiation factor 4A-I	Mus musculus	P60843	46128	3
Glutathione S-transferase Mu 1 (EC 2.5.1.18)	Mus musculus	P10649	25822	7
Glutathione S-transferase Mu 2 (EC 2.5.1.18)	Mus musculus	P15626	25569	6
Glutathione S-transferase Mu 5 (EC 2.5.1.18)	Mus musculus	P48774	26617	7
Glyceraldehyde-3-phosphate dehydrogenase	Mus musculus	P16858	35656	8
Glycylpeptide N-tetradecanoyltransferase 1	Homo sapiens	P30419	56770	3

Protéines	Taxonomie	N° Accession	M (Da)	Nbr de peptides
Guanine nucleotide-binding protein G(i), alpha-2 subunit	Mus musculus	P08752	40314	2
Heterogeneous nuclear ribonucleoprotein K	Mus musculus	P61979	50944	2
Heterogeneous nuclear ribonucleoproteins A2/B1	Mus musculus	Q88569	35971	4
Interleukin enhancer-binding factor 2 (Nuclear factor of activated T-cells 45 kDa)	Mus musculus	Q9CXY6	43035	4
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase	Mus musculus	P53395	53127	3
L-lactate dehydrogenase A chain (LDH muscle subunit)	Mus musculus	P06151	36344	6
L-lactate dehydrogenase B chain (LDH heart subunit)	Mus musculus	P16125	36418	4
L-lactate dehydrogenase C chain (LDH testis subunit)	Mus musculus	P00342	35758	8
Lysozyme C precursor (EC 3.2.1.17)	Gallus gallus	P00698	16228	14
Mitochondrial carrier homolog 2	Mus musculus	Q791V5	33477	3
Mitochondrial inner membrane protein	Mus musculus	Q8CAQ8	83848	4
NADP-dependent malic enzyme (EC 1.1.1.40)	Mus musculus	P06801	63958	4
Peroxiredoxin-6 (EC 1.11.1.15)	Mus musculus	O08709	24724	4
Phosphoglycerate kinase, testis specific	Mus musculus	P09041	44723	6
Protein disulfide-isomerase A3 precursor	Mus musculus	P27773	56586	7
Protein disulfide-isomerase precursor	Mus musculus	P09103	57108	3
Pyruvate dehydrogenase E1 component alpha subunit, testis-specific form	Mus musculus	P35487	43385	6
Pyruvate dehydrogenase E1 component beta subunit, mitochondrial precursor	Mus musculus	Q9D051	38912	7
Pyruvate kinase isozyme M2 (EC 2.7.1.40)	Mus musculus	P52480	57719	11
Serine/threonine-protein kinase VRK3	Mus musculus	Q8K3G5	50798	4
Solute carrier family 2, facilitated glucose transporter member 3	Mus musculus	P32037	53444	3
Spectrin alpha chain, brain	Mus musculus	P16546	167450	2
Spectrin beta chain, brain 1	Mus musculus	Q62261	274252	6
SPFH domain-containing protein 2 precursor	Mus musculus	Q8BFZ9	37849	4
T-complex protein 1 subunit epsilon	Mus musculus	P80316	59586	4
T-complex protein 1 subunit gamma	Mus musculus	P80318	50591	3
T-complex protein 1 subunit theta	Mus musculus	P42932	59386	8
T-complex protein 1 subunit zeta	Mus musculus	P80317	57837	3
Transitional endoplasmic reticulum ATPase	Mus musculus	Q01853	89121	13
Tropomodulin-3 (Ubiquitous tropomodulin)	Mus musculus	Q9JHJ0	39478	4
Tropomyosin alpha-3 chain (Tropomyosin-3)	Mus musculus	P21107	32843	4
Voltage-dependent anion-selective channel protein 2	Mus musculus	Q60930	31713	5
Voltage-dependent anion-selective channel protein 3	Mus musculus	Q60931	30733	3
Zona pellucida sperm-binding protein 3 receptor precursor	Mus musculus	Q60736	64908	3
Ovalbumin	Gallus gallus	P01012	42722	15
Ovotransferrin precursor	Gallus gallus	P02789	77726	4
Serum albumin precursor	Gallus gallus	P07724	68648	8





## Annexe 2 :

Liste des protéines et nombre de peptides identifiés par les deux méthodes sur l'ARN SIs2.

Protéines	N° Accession	Masse (Da)	Pep Tob	Pep MS2
Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2 / hnRNP B1)	P22626	37407	59	22
Heterogeneous nuclear ribonucleoprotein A3 (hnRNP A3)	P51991	39571	15	19
Heterogeneous nuclear ribonucleoproteins D0 (hnRNP D0)	Q14103	38410	14	15
Heterogeneous nuclear ribonucleoprotein G (hnRNP G)	P38159	42306	3	4
Heterogeneous nuclear ribonucleoprotein H (hnRNP H)	P31943	49067	16	8
Heterogeneous nuclear ribonucleoprotein K (hnRNP K)	P61978	50944	25	24
Heterogeneous nuclear ribonucleoprotein Q (hnRNP Q)	O60506	69590	21	15
Heterogeneous nuclear ribonucleoprotein R (hnRNP R)	O43390	70899	23	17
Splicing factor U2AF 35 kDa subunit	Q01081	27723	4	4
Splicing factor U2AF 65 kDa subunit	P26368	53467	16	7
Histone deacetylase 2 (HD2)	Q92769	55329	8	9
Histone-binding protein RBBP4 (RBBP-4)	Q09028	47495	5	4
Probable ATP-dependent RNA helicase DDX5 (DEAD box protein 5)	P17844	69105	11	23
Nucleolar RNA helicase II (DEAD-box protein 21)	Q9NR30	87302	14	6
General transcription factor 3C polypeptide 3	Q9Y5Q9	101208	6	16
WD-repeat protein 74 (NOP seven-associated protein 1)	Q6RFH5	42415	4	6
Nucleolin (Protein C23)	P19338	76437	72	28
RNA-binding protein FUS (Oncogene FUS)	P35637	53394	3	6
RNA-binding region-containing protein 2	Q14498	59343	4	5
Selenocysteine-specific elongation factor	P57772	65279	4	11
Vigilin (High density lipoprotein-binding protein)	Q00341	141352	28	9
Chromodomain helicase-DNA-binding protein 4	Q14839	217853	15	45
Importin alpha-2 subunit	P52292	57826	16	7
GTP-binding protein 1 (G-protein 1)	P62826	24408	11	3
HBS1-like protein (ERFS)	Q9Y450	75426	24	6
Actin-like protein 6A (53 kDa BRG1-associated factor A)	O96019	47430	5	6
N-acylneuraminatase cytidyltransferase	Q8NFW8	48349	5	5
Nuclease sensitive element-binding protein 1 (Y-box binding protein 1)	P67809	35903	17	5
Possible global transcription activator SNF2L4	P51532	184471	7	22
Ras GTPase-activating-like protein IQGAP1 (p195)	P46940	189434	23	23
Ras-GTPase-activating protein binding protein 1	Q13283	52132	9	8
Protein RCC2 (Telophase disk protein of 60 kDa)	Q9P258	565049	15	9
ATP-dependent RNA helicase A (Nuclear DNA helicase II)	Q5R874	140788	25	30

Liste des protéines et nombre de peptides identifiés par soit la méthode Tobramycine soit la méthode MS2 sur l'ARN SIs2.

Protéines	N° Accession	Masse (Da)	Peptides
<b>Protéines identifiées uniquement par la méthode Tobramycine</b>			
U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa)	P08621	51526	5
Histone deacetylase 6 (HD6)	Q9UBN7	131348	3
Heat shock protein HSP 90-beta (HSP 84) (HSP 90)	P08238	83081	6
Chromodomain helicase-DNA-binding protein 3 (CHD-3)	Q12873	220554	8
Cisplatin resistance-associated overexpressed protein	O95232	51435	3
Developmentally regulated GTP-binding protein 2 (DRG 2)	P55039	40720	10
Eukaryotic translation initiation factor 2 subunit 2	P20042	38364	6
DNA-binding protein A (Cold shock domain protein A)	P16989	40036	3
DNA-directed RNA polymerase I 40 kDa polypeptide	O15160	39225	4
DNA-directed RNA polymerase I associated factor 53 kDa subunit	Q9GZS1	53928	6
DNA-directed RNA polymerase I largest subunit	O95602	194068	9
Eukaryotic translation initiation factor 2 subunit 3	P41091	50946	10
Eukaryotic translation initiation factor 5B (eIF-5B)	O60841	138715	10
T-complex protein 1, alpha subunit (TCP-1-alpha) (CCT-alpha)	P17987	60306	4
T-complex protein 1, gamma subunit (TCP-1-gamma) (CCT-gamma)	P49368	60364	3
Transcriptional activator protein PUR-alpha	Q00577	34889	3
tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit TRM6	Q9UJA5	55764	9
UNR protein (N-ras upstream gene protein)	O75534	88829	24
Serine-threonine kinase receptor-associated protein (UNR-interacting protein)	Q9Y3F4	38414	8
Activator 1 40 kDa subunit (Replication factor C 40 kDa subunit)	P35250	39132	5
NAD-dependent malic enzyme, mitochondrial precursor	P23368	65402	5
Ras-GTPase-activating protein binding protein 2	Q9UN86	54078	17
G-rich sequence factor-1 (GRSF-1)	Q12849	50138	9
ATP-dependent RNA helicase A (Nuclear DNA helicase II)	Q5R874	140788	25
<b>Protéines identifiées uniquement par la méthode MS2</b>			
Heterogeneous nuclear ribonucleoprotein L (hnRNP L)	P14866	60149	3
RNA 3'-terminal phosphate cyclase	O00442	39311	6
U4/U6 small nuclear ribonucleoprotein Prp4 (U4/U6 snRNP 60 kDa protein)	O43172	58412	9
Splicing factor, arginine/serine-rich 1	Q07955	27597	6
Splicing factor, proline- and glutamine-rich	P23246	76102	7
Heat shock 70 kDa protein 1 (HSP70.1)	P08107	70009	7
THO complex subunit 4 (Tho4)	Q86V81	26741	7
Cleavage and polyadenylation specificity factor 5	O43809	26211	6
Histone deacetylase 1 (HD1)	Q13547	55068	6
ATP-dependent DNA helicase II 80 kDa subunit	P13010	82521	37
ATP-dependent RNA helicase DDX1 (DEAD box protein 1)	Q92499	82380	31
ATP-dependent RNA helicase DDX24 (DEAD box protein 24)	Q9GZR7	96271	5
DNA (cytosine-5)-methyltransferase 1	P26358	183050	13
DNA mismatch repair protein Msh2 (MutS protein homolog 2)	P43246	104686	4
DNA topoisomerase 2-alpha	P11388	174276	12
DNA topoisomerase 2-beta	Q02880	183152	11

Protéines	N° Accession	Masse (Da)	Peptides
Cell division cycle protein 23 homolog (Anaphase-promoting complex subunit 8)	Q9UJX2	68242	6
Cell division cycle protein 27 homolog (CDC27Hs) (H-NUC)	P30260	91809	9
Anaphase promoting complex subunit 2 (APC2)	Q9UJX6	93769	5
Anaphase promoting complex subunit 7 (APC7)	Q9UJX3	63121	10
General transcription factor 3C polypeptide 1	Q12789	238145	11
General transcription factor 3C polypeptide 4	Q9UKN8	91943	3
General transcription factor 3C polypeptide 5	Q9Y5Q8	59533	7
General transcription factor II-I (GTFII-I)	P78347	112346	7
Transcription intermediary factor 1-beta (TIF1-beta)	Q13263	88493	14
Transcriptional repressor p66 beta (p66/p68)	Q8WXI9	65220	4
Structural maintenance of chromosome 1-like 1 protein (SMC-protein)	O97593	143159	5
Structural maintenance of chromosome 3 (Chondroitin sulfate proteoglycan 6)	Q9UQE7	141454	13
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member	Q92922	122676	10
Possible global transcription activator SNF2L1	P28370	114659	16
Eukaryotic translation initiation factor 2 subunit 1	P05198	35958	7
Cullin-4A (CUL-4A)	Q13619	76772	3
DNA-binding protein RFX5 (Regulatory factor X subunit 5)	P48382	65283	4
Gamma-interferon-inducible protein Irf-16	P16666	88218	8
Guanine nucleotide binding protein-like 3 (Nucleolar GTP-binding protein 3)	Q9BVP2	61958	6
Testis-specific Y-encoded-like protein 1 (TSPY-like 1)	Q9H0U9	49261	4
GTP-binding protein era homolog (hERA)	O75616	48319	6
Tubulin beta-2C chain (Tubulin beta-2 chain)	P68371	88493	3
Ubiquitin-like containing PHD and RING finger domains protein 1	Q96T88	89757	8
UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit	O15294	116850	6
Kinesin-like protein KIF2	O00139	76876	4
Microtubule-associated protein 4 (MAP 4)	P27816	120944	4
Negative elongation factor C/D (NELF-C/D) (TH1-like protein)	Q8IXH7	66204	3
Non-POU domain-containing octamer-binding protein (NonO protein)	Q15233	54197	12
Probable ribosome biogenesis protein NEP1 (C2f protein)	Q92979	26572	10
Protein C14orf166	Q9Y224	28051	12
Septin-2 (Protein NEDD5)	Q15019	41461	5
Serine/threonine-protein kinase SRPK1	Q96SB4	92349	3
Glyceraldehyde-3-phosphate dehydrogenase	P04406	35899	3



### Annexe 3 :

Protéines et nombre de peptides identifiés dans les bandes D et E et non identifiés dans les contrôles A et F.

Protéines	N°Accession	D	E
40S ribosomal protein S3a	P97351 RS3A_MOUSE	1	
40S ribosomal protein S10	P63325 RS10_MOUSE	1	1
40S ribosomal protein S23	P62267 RS23_MOUSE		1
40S ribosomal protein S26	P62855 RS26_MOUSE		1
40S ribosomal protein S3	P62908 RS3_MOUSE		1
60S ribosomal protein L27a	P14115 RL27A_MOUSE		1
Activating signal cointegrator 1 complex subunit 3-like 1	Q6P4T2 Q6P4T2_MOUSE	3	
Actin-like protein 6A	Q9Z2N8 ACL6A_MOUSE	1	
Breast cancer metastasis-suppressor 1-like protein	Q3U1T3 BRM1L_MOUSE	1	
Bromodomain adjacent to zinc finger domain protein 1A	O88379 BAZ1A_MOUSE	1	
Casein kinase II subunit alpha	Q60737 CSK21_MOUSE	1	
Cleavage and polyadenylation specificity factor subunit 1	Q9EPU4 CPSF1_MOUSE	2	3
Cleavage and polyadenylation specificity factor 3	Q8CIM0 Q8CIM0_MOUSE	1	2
Cleavage and polyadenylation specificity factor subunit 6	Q6NVF9 CPSF6_MOUSE	1	
Cleavage stimulation factor, 3' pre-RNA, subunit 1	Q99LC2 CSTF1_MOUSE	1	
Cold-inducible RNA-binding protein	P60824 CIRBP_MOUSE	1	1
DNA (cytosine-5)-methyltransferase 1	P13864 DNMT1_MOUSE	1	3
DNA-directed RNA polymerase I subunit RPA1	O35134 RPA1_MOUSE	3	4
DNA-directed RNA polymerases I, II, and III subunit RPABC3	Q923G2 RPAB3_MOUSE		1
DEAD (Asp-Glu-Ala-Asp) box polypeptide 47	Q9CWX9 DDX47_MOUSE		1
Enhancer of zeste homolog 2	Q61188 EZH2_MOUSE		2
Eukaryotic translation elongation factor 1 gamma	Q4FZK2 Q4FZK2_MOUSE		1
Eukaryotic translation initiation factor 2 subunit 1	Q6ZWX6 IF2A_MOUSE		3
Eukaryotic translation initiation factor 2, subunit 2	Q99L45 IF2B_MOUSE		2
Eukaryotic translation initiation factor 2A	Q8BJW6 EIF2A_MOUSE		1
Elongation factor 1-alpha 1	P10126 EF1A1_MOUSE		1
Elongation factor 1-beta	O70251 EF1B_MOUSE		1
Elongation factor 1-delta	P57776 EF1D_MOUSE		1
Glomulin	Q8BZM1 GLMN_MOUSE		1
Gtf2h1 protein	Q7TPY0 Q7TPY0_MOUSE	1	
Heterogeneous nuclear ribonucleoprotein K	P61979 HNRPK_MOUSE	3	8
Heterogeneous nuclear ribonucleoproteins C1/C2	Q9Z204 HNRPC_MOUSE	2	2
Heat shock 70kD protein 5	A2AUF6 A2AUF6_MOUSE	1	
High density lipoprotein binding protein -	Q3U4Z7 Q3U4Z7_MOUSE	1	2
Hist2h4 protein	Q3U4Z7 Q3U4Z7_MOUSE		2
Histone deacetylase 1	O09106 HDAC1_MOUSE	1	
Histone H2B (Fragment)	A0JLV3 A0JLV3_MOUSE	1	1
Laminin receptor (Fragment)	B2CY77 B2CY77_MOUSE	2	1
Methyltransferase 10 domain containing	Q5SW15 Q5SW15_MOUSE	1	
MKIAA1676 protein (Fragment)	Q6ZPK2 Q6ZPK2_MOUSE	1	
Origin recognition complex subunit 5	Q9WUV0 ORC5_MOUSE	1	1
Paired amphipathic helix protein Sin3a	Q60520 SIN3A_MOUSE	1	1

Protéines	N° Accession	D	E
Poly(rC)-binding protein 1	P60335 PCBP1_MOUSE	1	
p21-activated protein kinase-interacting protein 1	Q9DCE5 PK11P_MOUSE		1
Pre-mRNA 3'-end-processing factor FIP1	Q9D824 FIP1_MOUSE	1	
Pre-mRNA-processing factor 6	Q91YR7 PRP6_MOUSE		1
Probable ATP-dependent RNA helicase DDX49	Q4FZF3 DDX49_MOUSE	1	
Protein DEK	Q7TNV0 DEK_MOUSE	1	1
Protein FRG1	P97376 FRG1_MOUSE		1
Prp19 beta protein	Q4ADG5 Q4ADG5_MOUSE	1	
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DDX15	O35286 DHX15_MOUSE	1	
Putative RNA-binding protein 3	O89086 RBM3_MOUSE		1
Rap1 interacting factor 1 homolog	Q6PR54 RIF1_MOUSE	4	3
Retinoblastoma binding protein 4	Q60972 RBBP4_MOUSE	1	2
Replication factor C subunit 4	Q99J62 RFC4_MOUSE	1	
Ribonuclease P protein subunit p30	O88796 RPP30_MOUSE	2	
RNA binding motif protein, X chromosome	A2AFI3 A2AFI3_MOUSE	1	1
Rps16 protein	A4FUS1 A4FUS1_MOUSE		1
RuvB-like 1	P60122 RUVB1_MOUSE		1
RuvB-like protein 2	Q9WTM5 RUVB2_MOUSE		1
Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	P62137 PP1A_MOUSE	2	2
Sister chromatid cohesion protein PDS5 homolog A	Q6A026 PDS5A_MOUSE		2
Small subunit processome component 20 homolog	Q5XG71 UTP20_MOUSE		1
Splicing factor 3a, subunit 3	Q9D554 SF3A3_MOUSE	1	
Splicing factor, arginine/serine-rich 3	P84104 SFRS3_MOUSE	1	2
Structure specific recognition protein 1 (Fragment)	A2AW05 A2AW05_MOUSE	1	
THO complex subunit 4	O08583 THOC4_MOUSE	1	1
Thyroid hormone receptor interactor 12	Q497C1 Q497C1_RAT	3	4
Treacle protein	O08784 TCOF_MOUSE	1	
Tripartite motif-containing protein 71	Q1PSW8 LIN41_MOUSE	1	6
U6 snRNA-associated Sm-like protein LSM4	Q9QXA5 LSM4_MOUSE		1
Ubiquitin carboxyl-terminal hydrolase 7	Q6A4J8 UBP7_MOUSE	1	
UDP-N-acetylglucosamine—peptide N-acetylglucosaminyltransferase 110 kDa subunit	Q8CGY8 OGT1_MOUSE		1
Upstream binding transcription factor, RNA polymerase I	A2AWT5 A2AWT5_MOUSE		1
Y box protein 2 -	P62960 YBOX1_MOUSE	1	1
Zinc finger RNA binding protein	B2RUG7 B2RUG7_MOUSE	1	1

## **Annexe 4 :**

Protéines et nombre de peptides identifiés à J0, J2 et J4 et non identifiés dans les contrôles.

<b>Protéines</b>	<b>N° Accession</b>	<b>J0</b>	<b>J2</b>	<b>J4</b>
40S ribosomal protein S10	P63325 RS10_MOUSE			1
40S ribosomal protein S3	P62908 RS3_MOUSE	2	1	
40S ribosomal protein S3a	P97351 RS3A_MOUSE		1	
40S ribosomal protein S4, X isoform	P62702 RS4X_MOUSE	1	1	
40S ribosomal protein S6	P62754 RS6_MOUSE		1	
40S ribosomal protein S7	P62082 RS7_MOUSE	1	1	2
40S ribosomal protein S9	Q6ZWN5 RS9_MOUSE		1	1
60S ribosomal protein L6	P47911 RL6_MOUSE	1	1	1
60S ribosomal protein L7a	P12970 RL7A_MOUSE		1	
60S ribosomal protein L9	P51410 RL9_MOUSE			1
60S ribosomal protein L13	P47963 RL13_MOUSE	1		
60S ribosomal protein L18	P35980 RL18_MOUSE		2	
60S ribosomal protein L23a	P62751 RL23A_MOUSE			1
60S ribosomal protein L31	P62900 RL31_MOUSE			2
60 kDa heat shock protein, mitochondrial precursor	P63038 CH60_MOUSE	1		
Activating signal cointegrator 1 complex subunit 3-like 1	Q6P4T2 Q6P4T2_MOUSE	15		
Anaphase-promoting complex subunit 7	Q9WVM3 APC7_MOUSE	1		
ATP-dependent RNA helicase DDX1	Q91VR5 DDX1_MOUSE	2		
ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase) 1	Q921K2 Q921K2_MOUSE	2		
Budding uninhibited by benzimidazoles 3 homolog	Q6ZWM5 Q6ZWM5_MOUSE	1		
Cell division cycle 5-like	Q6A068 CDC5L_MOUSE	1		
Cell growth-regulating nucleolar protein	Q08288 LYAR_MOUSE	1	1	
Chromodomain helicase DNA binding protein 5	A2A8L1 A2A8L1_MOUSE	1		
Cisplatin resistance-associated overexpressed protein	Q5SUF2 CROP_MOUSE	2		
Cleavage and polyadenylation specificity factor subunit 1	Q9EPU4 CPSF1_MOUSE	7		1
Cleavage and polyadenylation specificity factor subunit 2	O35218 CPSF2_MOUSE	1		
Cleavage and polyadenylation specificity factor 3	Q8CIM0 Q8CIM0_MOUSE	2		2
Cleavage and polyadenylation specificity factor subunit 5	Q9CQF3 CPSF5_MOUSE	4	4	2
Cleavage and polyadenylation specificity factor subunit 6	Q6NVF9 CPSF6_MOUSE			1
Cleavage and polyadenylation specificity factor subunit 7	Q8BTV2 CPSF7_MOUSE			1
Cleavage stimulation factor, 3' pre-RNA, subunit 1	A2AP97 A2AP97_MOUSE	2		1
Cleavage stimulation factor, 3' pre-RNA, subunit 2	A2AEK0 A2AEK0_MOUSE	1		
Cleavage stimulation factor, 3' pre-RNA, subunit 3	A2AQH9 A2AQH9_MOUSE	3		5
DAZ associated protein 1	Q9JII5 DAZP1_MOUSE	4		
Developmentally-regulated GTP-binding protein 1	P32233 DRG1_MOUSE	1		
DNA (cytosine-5)-methyltransferase 1	P13864 DNMT1_MOUSE	2		
DNA-directed RNA polymerase I subunit RPA1	O35134 RPA1_MOUSE	1		
DNA-directed RNA polymerases I, II, and III subunit RPABC1	Q3V214 Q3V214_MOUSE		1	
Double-strand-break repair protein rad21 homolog	Q61550 RAD21_MOUSE	1		
ELAV-like protein 1	P70372 ELAV1_MOUSE	4	3	4
Elongation factor 1-beta	O70251 EF1B_MOUSE	1		
Elongation factor Tu GTP binding domain containing 2	A2AH85 A2AH85_MOUSE	8		
Eukaryotic initiation factor 4A-II	P10630 IF4A2_MOUSE	1		
Eukaryotic translation elongation factor 1 gamma	Q4FZK2 Q4FZK2_MOUSE	1		

Protéines	N° Accession	J0	J2	J4
Eukaryotic translation initiation factor 2 subunit 1	Q6ZWX6 IF2A_MOUSE	5	1	
Eukaryotic translation initiation factor 2 subunit 2	Q99L45 IF2B_MOUSE	1		
Eukaryotic translation initiation factor 2 subunit 3, X-linked	Q9Z0N1 IF2G_MOUSE	2		
Exosome complex exonuclease MTR3	Q8BTW3 EXOS6_MOUSE	1		
Exosome component 2 (Fragment)	A2AV19 A2AV19_MOUSE	1		
Exosome component 7	Q4VBW5 Q4VBW5_MOUSE	1		
Exosome complex exonuclease RRP41	Q921I9 EXOS4_MOUSE	1		
Far upstream element (FUSE) binding protein 1	Q91WJ8 FUBP1_MOUSE	7		2
Far upstream element (FUSE) binding protein 3	A2AJ72 A2AJ72_MOUSE	4	2	7
Far upstream element-binding protein 2	Q99PF5 FUBP2_RAT	9		6
General transcription factor 3C polypeptide 1	Q8K284 TF3C1_MOUSE	2		
General transcription factor IIIC, polypeptide 4	Q148Y1 Q148Y1_MOUSE	1		
Guanine nucleotide-binding protein subunit beta-2-like 1	P68040 GBLP_MOUSE	2	3	
H/ACA ribonucleoprotein complex subunit 4	Q9ESX5 DKC1_MOUSE	1		
Heterogeneous nuclear ribonucleoproteins C1/C2	Q9Z204 HNRPC_MOUSE	7	5	1
Heterogeneous nuclear ribonucleoprotein D-like	Q9Z130 HNRDL_MOUSE	5	3	6
Heterogeneous nuclear ribonucleoprotein G	P84586 HNRPG_RAT	2	1	3
Heterogeneous nuclear ribonucleoprotein K	P61979 HNRPK_MOUSE	5	2	5
Heterogeneous nuclear ribonucleoprotein L	Q8R081 HNRPL_MOUSE	2		5
Heterogeneous nuclear ribonucleoprotein M	Q9D0E1 HNRPM_MOUSE	3		2
Heterogeneous nuclear ribonucleoprotein Q	Q7TMK9 HNRPQ_MOUSE	6	3	10
Heterogeneous nuclear ribonucleoprotein R	Q8VHM5 Q8VHM5_MOUSE	3		9
HnRNP-associated with lethal yellow (Fragment)	A2AU61 A2AU61_MOUSE	1		
High density lipoprotein binding protein	Q3U4Z7 Q3U4Z7_MOUSE	5	1	
H1 histone family, member X	Q80ZM5 Q80ZM5_MOUSE		1	
Histone H1.1	P43275 H11_MOUSE		1	3
Histone H1.2	P15864 H12_MOUSE	1		4
Histone H1.4	P43274 H14_MOUSE			3
Histone H2A (Fragment)	A0AUV1 A0AUV1_MOUSE			1
Histone H2B (Fragment)	A0JLV3 A0JLV3_MOUSE			1
Histone H3 (Fragment)	A1L0U3 A1L0U3_MOUSE			1
Hmmr protein	Q00547 HMMR_MOUSE	3		
Importin subunit beta-1	P70168 IMB1_MOUSE	2		
Integrator complex subunit 3	Q7TPD0 INT3_MOUSE	1		
Integrator complex subunit 4	Q8CIM8 INT4_MOUSE	1		
Integrator complex subunit 6	Q6PCM2 INT6_MOUSE	1		
INO80 complex subunit C	Q8BHA0 CR037_MOUSE	2		
Kelch repeat and BTB domain-containing protein 3	Q8BHI4 KBTB3_MOUSE			1
KH domain containing, RNA binding, signal transduction associated 1	B2KG38 B2KG38_MOUSE		1	
Kinesin family member 2C	A2AE71 A2AE71_MOUSE	1		
Kinesin-like protein KIF2A	P28740 KIF2A_MOUSE	1		
Leucine-rich PPR motif-containing protein, mitochondrial precursor	Q6PB66 LPPRC_MOUSE	4		
Lin-28 homolog A	Q8K3Y3 LN28A_MOUSE	1	2	
Lysozyme C type P precursor	P17897 LYSCP_MOUSE	1		
Matrin-3	Q8K310 MATR3_MOUSE	3		7
Microtubule-associated protein 4	P27546 MAP4_MOUSE	1		
Myelin basic protein expression factor 2, repressor	A2ATP5 A2ATP5_MOUSE	2		
Novel protein	A2ATZ4 A2ATZ4_MOUSE	1		



Protéines	N° Accession	J0	J2	J4
Nucleolysin TIAR	P70318 TIAR_MOUSE	2	1	2
Paired amphipathic helix protein Sin3a	Q60520 SIN3A_MOUSE	2		
Paraspeckle component 1	Q8R326 PSPC1_MOUSE	2	1	
PC4 and SFRS1 interacting protein 1	A2B110 A2B110_MOUSE	2		
Poly(RC) binding protein 2	B2M1R7 B2M1R7_MOUSE	1		
Poly(rC)-binding protein 1	P60335 PCBP1_MOUSE	1	1	
Poly(U)-binding-splicing factor PUF60	Q3UEB3 PUF60_MOUSE	5	3	6
Polyadenylate-binding protein 2	Q8CCS6 PABP2_MOUSE	1		1
Polycomb protein Suz12	Q80U70 SUZ12_MOUSE	1		
Polypyrimidine tract binding protein 1	Q922I7 Q922I7_MOUSE	8	7	13
Polypyrimidine tract-binding protein 2	Q91Z31 PTBP2_MOUSE	1	2	2
Pre-mRNA 3'-end-processing factor FIP1	Q9D824 FIP1_MOUSE			1
Pre-mRNA-splicing factor SPF27	Q9D287 SPF27_MOUSE	1	1	
Pre-mRNA-processing factor 6	Q91YR7 PRP6_MOUSE	2	1	
Probable ATP-dependent RNA helicase DDX47	Q9CWX9 DDX47_MOUSE		1	
Probable ribosome biogenesis protein NEP1	O35130 NEP1_MOUSE	1		
Processing of 1, ribonuclease P/MRP family	Q8K205 Q8K205_MOUSE	1		
Proliferation-associated 2G4	P50580 PA2G4_MOUSE	1		
Protein DEK	Q7TNV0 DEK_MOUSE	3		
Protein FRG1	P97376 FRG1_MOUSE	1		
Prp19 beta protein	Q99KP6 PRP19_MOUSE	2		1
PRP4 pre-mRNA processing factor 4 homolog	Q9DAW6 PRP4_MOUSE		1	
Prpf8 protein	A5D6Q4 A5D6Q4_MOUSE	9	1	
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	O35286 DHX15_MOUSE	2		
Putative uncharacterized protein	Q9DAE2 Q9DAE2_MOUSE	1		
Putative uncharacterized protein	Q3U2B5 Q3U2B5_MOUSE	3		1
Ras GTPase-activating protein-binding protein 1	P97855 G3BP1_MOUSE	2		
Ras GTPase-activating protein-binding protein 2	P97379 G3BP2_MOUSE	1		
Recombining binding protein suppressor of hairless	P31266 SUH_MOUSE	1		
Regulator of chromosome condensation	Q8VE37 RCC1_MOUSE	1		
Replication factor C (Activator 1) 2	Q9WUK4 RFC2_MOUSE	2		
Replication factor C subunit 4	Q99J62 RFC4_MOUSE	1		
Ribonuclease P protein subunit p30	O88796 RPP30_MOUSE	2		
Retinoblastoma binding protein 7	Q60973 RBBP7_MOUSE	1		
Ribosomal protein S5, isoform CRA_b	B0BN81 B0BN81_RAT		2	
Ribosomal protein L19	P84099 RL19_MOUSE	1	1	
Ribosomal protein S14	P62264 RS14_MOUSE			1
Ribosomal protein S2	P25444 RS2_MOUSE		1	
Ribosomal protein S8	P62242 RS8_MOUSE		2	
RNA binding motif protein 39 (Fragment)	Q0VGU9 Q0VGU9_MOUSE	2		
RNA granule protein 105	Q60865 CAPR1_MOUSE	2		
RNA-binding protein FUS	P56959 FUS_MOUSE	4	3	
Putative RNA-binding protein Luc7-like 1	Q9CY14 LUC7L_MOUSE	1		
Rpo1-1 protein	P52432 RPAC1_MOUSE	2	1	
Selenocysteine-specific elongation factor	Q9JHW4 SELB_MOUSE	1		
Small nuclear ribonucleoprotein B	P27048 RSMB_MOUSE	3	1	1
Small nuclear ribonucleoprotein D1	P62315 SMD1_MOUSE			1

Protéines	N° Accession	J0	J2	J4
Small nuclear ribonucleoprotein E	P62305 RUXE_MOUSE			3
Small nuclear ribonucleoprotein Sm D2	P62317 SMD2_MOUSE			3
Small proline-rich protein 2G	O70554 SPR2B_MOUSE			1
Splicing factor 1	Q64213 SF01_MOUSE	1		3
Splicing factor 3, subunit 1	Q8K4Z5 SF3A1_MOUSE	1		
Splicing factor 3a, subunit 3	Q9D554 SF3A3_MOUSE	1		
Splicing factor 3B, subunit 1	Q99NB9 SF3B1_MOUSE	2		
Splicing factor 3b, subunit 2	Q80W39 Q80W39_MOUSE	2		
Splicing factor 3b, subunit 3	Q921M3 SF3B3_MOUSE	3		
Splicing factor U2AF 65 kDa subunit	P26369 U2AF2_MOUSE	7		6
Splicing factor, arginine/serine-rich 2	Q62093 SFRS2_MOUSE	1	1	2
Splicing factor, arginine/serine-rich 3	P84104 SFRS3_MOUSE	2	2	2
Splicing factor, arginine/serine-rich 5	O35326 SFRS5_MOUSE	2		1
Splicing factor arginine/serine-rich 11	Q8CFF2 Q8CFF2_MOUSE	1		
Structural maintenance of chromosomes 1A	A2AFQ5 A2AFQ5_MOUSE	6		1
Structural maintenance of chromosomes protein 3	Q9CW03 SMC3_MOUSE	2		
Structure specific recognition protein 1 (Fragment)	Q08943 SSRP1_MOUSE		1	
Superkiller viralicidic activity 2-like 2	Q9CZU3 SK2L2_MOUSE	1		
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	Q91ZW3 SMCA5_MOUSE	1		
T-complex protein 1 subunit gamma	P80318 TCPG_MOUSE,	2		
T-complex protein 1 subunit theta	P42932 TCPQ_MOUSE	1		
T-complex protein 1, alpha subunit	P11984 TCPA1_MOUSE	1		
Telomere-associated protein RIF1	Q6PR54 RIF1_MOUSE	3		
THO complex subunit 4	O08583 THOC4_MOUSE	3	3	1
Transcription initiation factor TFIID subunit 5	Q8C092 TAF5_MOUSE	1		
Transformation/transcription domain-associated protein	Q80YV3 TRRAP_MOUSE	1		
Treacle protein	O08784 TCOF_MOUSE	1		
U1 small nuclear ribonucleoprotein A	Q62189 SNRPA_MOUSE	4	6	1
U1 small nuclear ribonucleoprotein polypeptide A	Q62376 RU17_MOUSE	2	2	
U1 small nuclear ribonucleoprotein A'	P57784 RU2A_MOUSE	4	5	2
U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1	Q9D883 U2AF1_MOUSE	3	5	3
U2 small nuclear ribonucleoprotein B	Q9CQI7 RU2B_MOUSE	1		
U4/U6.U5 tri-snRNP-associated protein 2	Q3TIX9 SNUT2_MOUSE	3	1	
UPF0568 protein C14orf166 homolog	Q9CQE8 CN166_MOUSE	2	1	
Upstream binding transcription factor, RNA polymerase I	P25976 UBF1_MOUSE	5		
WD repeat-containing protein 5	P61965 WDR5_MOUSE	1		
WD repeat-containing protein 57	Q6PE01 WDR57_MOUSE	1		
WD repeat-containing protein 82	Q8BFQ4 WDR82_MOUSE	1		
Zinc finger protein 207 (Fragment)	Q9Z326 Q9Z326_MOUSE	1		

**Annexe 5 :**

Protéines et peptides identifiés dans le plasma de manchot

N°	Protein	NCBI Accession N°	pI / MW (Da)	unique peptides / seq coverage	Peptide sequences	Mascot MS/MS ion score	MS-Blast HSP score
1	Tranthyretin	gij1633502	5.1 / 14142.3	6 / 84.6	APLVSHGSVDSK / VEFDTSSYWK / GLGLSPFHEYADVFTANDSGHR / GSPAANVAVK / HYTIAALLSPFSYSTTAVVSDPQE / VLDAVR	42.8/63/55.4 /51.4/113/36.7	
2	Retinol-binding protein	gij18655833	5.62/22 497.7	10 / 72.9	APLVSHGSVDSK / GSPAANVAVK / KAADGSWQDFATGK / TAVVSDPE / VEGLYR / VFTANDSGHR ELNEDGTCADSYFVFSR / VIVHNGFCS / GNDDHWVVDYDYALHYSCR / VKENFDK / KDPEGLFLQDNVVAQFTVDENGQMSATAK / QIDLCLDR / QIDLCLDRK / LFNNDVDCADMIGSFTDTEPAK / YSGTWYAMAK / YWGVASFLQK	64.7/46.5/60.5/41.2/45/5.4.8/39.5/92.2/52.2/54.5	56/67/81/41/97/50/39/74
3	Apolipoprotein A-I	gij211146	5.97/30 663.1	18 / 50.3	ASGKDAIAQFESSAVGK / DAIAQFESSAVGK / DMVDVYLETVK / IRDMVDVYLETVK / LADNLDLSAAAAK / LTPVAEEAR / NLAPYSDELRLADNLDLSAAAAK / LTPVAEEAR / GIPQAAEYQAK / VVEQLSNLR / WTEELEQFR / EDPQAPLER / LAPVAKDLK / EDAMPYYK / LTPVAE / YYQEVV / PLVQQYVER / YKQRLLP / TPLVKDF	37.2/71.7/75.1/63.1/122/38.3/43.9	89/56/65/54/66/42/36/41/41/39/34/35
4	Carbonic anhydrase 2	gij115454	6.56/28 989.2	1 / 4.2	VVDALNSIQTK	53.8	
5	Adiponectin	gij45551207	5.45/26 443.8	6 / 18.4	SAFSVGLTER / APHPNVPLR	51.8/56.2	
6	Apolipoprotein A-IV	gij3645997	4.8/408 35.1	6 / 16.1	SAFSVGLTER / APHPNVPLR / EQSHYDA / GAFVYR / VFTFMQFQ / GHNGL ETVDQLQQAETK / LAPLAQQLDER	85/79.2	68/67/49/39/34/34
7	Vitamine D binding protein	NP_990213.1	6.47/53 651.1	53 / 58.6	FAQEVKDGLNR / FLSTAEP / QEAAPL / YFTELGSNAK ACSRL / AGCCAQEGLEKK / APLPVLLGSTR / CCPNSP / CCPNSPPLYC / CLSPAPT / CLSPAPTSCFLK / CSNGSPFAHPGT / DAGSSALGTK / DFFDRFL / ECCSAK / EGGLEQK / EPSNEEL / EPSNEELCKAFK / FADCCSGK / FASTCCPA / FASTCCPPNS / FELAR / FNQCCSDVAEDC / FSYLTSLAQK / GEAPLPVLL / HPGTA / HPPQQLP / HTTVPDAFLGK / LCGQYTALNFLDFK / LCLAAL / LPEAQPTNEQLC / LPPFLEK / LPVLLGSTR / MESTCC / NFLSM / PCCEDAGSSALSAK / PEASPELL / PEASPELLGKLVQR / PSNEELCKAFK / QQPTNEQLC / RETMPEASPELLANLVEQR / RFVHE / RLAAY / RRHRNV / RYLFELAR / SACLDG / SPFPAHPGT / STCCPPNS / STCCPPNSPLYC / TCCAD / TLLANSR / TLLSLLTLLSNR / VAEDCFQ / VCQEFQALGK / YDASE / YLFELAR / YLKPSNE		63/37/34/64 35/43/71/48/81/44/69/74/40/39/47/35/51/79/52/62/68/36/68/58/51/36/44/62/95/37/62/40/57/38/39/62/48/52/71/47/76/34/35/34/54/35/67/53/94/41/34/62/45/51/38/46/36

8	Nectinepsin	gij1480970	4.91/40 053.1	7 / 17.0	FDDGALDPGYPR / LDAAMAGR / LISDVWGIEGPIDAAFTR	87.5/52.8/87 .8	
9	complement component 3	gij45382303	6.69/18 4068	4 / 2.5	HEPTQWEC / EPEELCSR / DYSTVCK / DGALDTESPR CDNFDLR / GLCVADPYEITVMK / GTMTVVTVYKAK / LNEEDFTVK	50/68.5/37.6 /59.8	38/64/55/34
10	PREDICTED: similar to alpha-1- antitrypsin	gij11809213 1	5.7/475 63.1	1 / 2.3	ATSTTLFMGK	64	
11	prostaglandin H2 D- isomerase	NP_989590. 1	6.3/208 43.6	6/29.7	LSATADGNLEVTGMPYK / GSSTMVLLYSR / ALVATQLSK / SYTSPR / ELLLLP / LPVQTD		79/68/57/39/ 36/35
12	Ig light chain C-region	gij212936	5.85/11 531	1/18.8	QSNNQYMASSYLTLXASDWK	81.9	
13	lactate dehydrogenas e A	gij11841993 3	7.87/36 496.4	1/13.0	VIGSGCNLDSAR	68.7	
14	fibrinogen beta chain	gij211780	7.18/52 661.7	5/10.8	YYWGGTYSWDMAK / DNLDNNIPSSLR / GSWYSMK / QDGSVNFGR / YQLSVSNYK	68.7/63.2/56 .9/44.2/46.6/ 71.2	
15	immunoglobuli n heavy chain variable region	gij16151324 7	7.89/13 130.5	1/10.9	ADDTATYYCAK	76.8	
16	Chain C, Crystal Structure Of Native Chicken Fibrinogen With Two Different Bound Ligands	gij21730885	5.32/46 843.9	14/29.5	DCQDIANK / DTAEIQETTGR / EFGHLSRDDTTEFWLGNEK / ENCCILDER / FGSYCPTTCGIADFFNK / GTADYAVFK / IELEDWSGK / IIEEIIR / KGTADYAVFK / KIIEIIR / KSGLYFIKPQK / LDGSEDFR / RLDGSEDFR / SGLYFIKPQK	39.7/85.6/78 .6/38.8/92.7/ 70.9/60.2/52 .1/64.1/54.2/ 50.4/62.1/44 .4/52.2	
17	PREDICTED: laminin, gamma 1	gij11809422 6	12.1/17 3534.5	1/0.6	ISISGXTTTR	51.6	
18	PREDICTED: similar to Serpina1d- prov protein	gij11809195 8	6.71/47 720.2	2/4.47	AALDVDER / EFFVDAETTVK	54/49.3	
19	haptoglobin- like protein	gij89274064	5.34/37 092.4	1/2.7	VGYSVSGWGR	49.8	
20	Serum albumin precursor	gij113575	5.51/69 900.7	8/13.8	DSYGAMADCCSK / DVVDLAQK / LVKDVVDLAQK / NECFLSFK / QSDINTCFGEEGANLIVQSR / RPCFTAMGVDTK / TIADGFTAMVDK / YNDLKEETFK	63.5/46.9/55 .9/50/50.2/6 2.6/86.5/68.	

21	histidine phosphatase of the endoplasmic reticulum	gij2795769	8.15/50 241.8	4/7.8	CLQSGAAFR / FPALFAAR / GYGYDINSR / VLEYLNDLK	5 63.8/38.1/45 .1/44.5	
	PREDICTED: similar to apolipoprotein H	gij11809976 7	8.6/387 64.6	1/3.6	WPLNTLLCLPK	65	
22	antithrombin	gij18140913	6.31/51 386.7	8/20.0	DNGENIFMSPLSISTAFAMTK / EAALNTIIFMGR / FAVVFYK / LPGIIAEGR / MGLEDLFPENAR / SSELVSNR / TSDQIHFFFAK / VQVLELPYK	85.6/62.9/36 .8/44.4/67.3/ 64/78.1/46.4	
23	Gelsolin precursor	gij34921412	5.93/85 814.3	11/16.0	AGGVASGFR / DAQEEKTEALK / DKDQTEGLGEAYISGHVAK / EPGLQIWR / FDLVPVPK / GAAAIFTVQMDDYLQGK / HTQVQVLPESGETPLFK / HVVPNEVTVQR / LFACSNK / TATDFIDK / YIETDPASR	62.2/55.1/35 .1/42.1/41.3/ 45.8/73.3/55 .4/40.2/39.3/ 49.2	
24	PREDICTED: similar to heparin cofactor II	gij11809835 9	6.0/557 99	5/11.2	DFSILNDFR / FIVDRPFLFIYEHR / LNILNANFGFNLYR / QITPQVVEK / RNFQYTLR	56.2/41.2/79 .5/51.1/38.5	
25	immunoglobulin heavy chain variable region	gij16151324 7	7.89/13 130.5	1/10.9	ADDTATYYCAK	91	
26	prothrombin	gij6683108	6.04/69 376.2	7/13.0	DIALMHLK / GDACEGDSGGPFVMK / LIENYCR / TTVEFTPR / VTDNMFCAGYSPEDSKR / WYQVGIVSWGEGCDR / YGFYTHVFR	39.2/91.3/43 .9/59.3/68.1/ 81.4/44.1	
27	PIT 54 protein	NP_997063.	4.61/50 821.5	20/35.1	GCGSPLSALGGAR / DLNDAQVVCK / GPDMLWLDDV / GTLCDDKWELR / NASAVCSDSGLTV / SLTECEAK / QLGCSSTAVSAP / VEVLHENR / HNCYH / QLGCSAAVSA / CDDQWD / AKVVCK / GELWLSL / VGGPNR / ASLTECEA / DADVCAVVK / ECQAR / GTEVPE / VSTSVR / AQVVCK		76/74/67/66/ 65/57/54/52/ 46/44/44/44/ 44/43/42/40/ 40/39/36/34
	PREDICTED: similar to deleted in malignant brain tumors 1	XP_424430. 2	5.28/20 1452.7	20/8.0	QLGCSSTAVSAP / NDGKDLNDAQVVC / GTLCDDKWELR / ASAVCSDSGL / GCGSPLSALGGAR / VEVLHENR / QLGCSAAVSA / NASAVCSDS / DAQVVCK / ECQAR / VELLHENR / AQVVC / WLENVE / AKVVC / VVVSAP / ASLTECEA / ASLTEC / SPCAG / GELWLSL / VGGPNR		67/60/59/47/ 47/43/43/41/ 40/40/39/38/ 37/37/37/36/ 36/36/36/36/ 35/35
28	PREDICTED: similar to inter-alpha-trypsin inhibitor heavy	gij11811760 0	9.82/14 148.4	3/20.0	HMEVDVR / IIEPQGIR / SNALDMENFK	37.6/52.9/41 .3	

	chain 2, partial PREDICTED: similar to Alpha 1 microglobulin/b ikunin	gij11809912 9	5.59/38 4402	1/3.1	DPGPCSGMLSR	61.5	
29	Alpha- fetoprotein precursor	P84407.1	6.26/70 680.4	5/7.7	CEGDMVECMDDR / RYAEFSTQLLLR / HCCSAMYAYR / FVVEYAR / FADCCSK		61/51/47/45/ 43
30	PREDICTED: similar to Alpha-2- macroglobulin precursor	gij11808328 2	6.05/16 2825.2	9/6.0	AIGYLVSGYQR / FEVTVK / FSLLLPSNVVQDSGR / MLSGFIPVK / QLLVEPEGVEK / SLIDDVVSEK / TGQLSEEVK / TVAWLVTPK / VITILDEK	60.9/38.6/57 .5/48.8/50.8/ 59.7/76.2/50 .4/72.3	
31	PREDICTED: similar to Complement C5 precursor	gij11809944 0	6.45/18 6288.8	1/0.5	LTEDNFVGR	52.8	
32	PREDICTED: similar to complement protein C7	gij50761596	6.48/50 207.2	1/2.9	LSESVLAYTFQVK	79.4	
33	PREDICTED: similar to complement regulator factor H	XP_426613. 2	6.73/13 9081.5	6/3.7	EDPASSPFR / YEPGETLR / DCEPPPELAN / WSPPP / TLYLAGA / CANNYVR		67/60/58/43/ 40/35
34	PREDICTED: similar to plasminogen	gij11808830 8	7.92/90 745.9	10/12.6	CTTPPPVPAPGR / DQQCITLAENTK / ENAVLGGR / ETGFPVIENK / GTASFITGK / KAWEYCDIPK / NPDGDVNGPWCYTTPR / TAVLFR / TNAVLYEK / TPENYPCK /	57.5/95.3/39 .6/46.5/61.4/ 37.9/64.1/38 .5/54.4/42.8 60.7	
35	Pyruvate kinase muscle isozyme	gij125608	7.29/57 997.4	1/3.1	FGVEQNVDMVFASFIR	60.7	
36	Chain A, Apo Duck Ovotransferrin	gij15782999 9	6.19/75 615.7	6/7.7	ATYLDCAIK / DEYELLCLDGSR / DGKGDVAFVK / GDVAFVK / HTTPQENAPEEKDEYELLCLDGSR / WCTISSAEEKK	40.9/38.1/59 .6/51.4/70.6 1.5	
37	melanoma- associated antigen p97	NP_990538. 1	5.92/80 912.4	14/14.7	QAVANFFSASCV / YYGYTGAFR / SYTAVV / FTLNNLQGK / KAVANFFSA / KGTDFTLNNLQG / WCTLSSAE / EFELLCTNG / NLPLGTLLRK / LCELCK / FFAVAVAR / VTLWRT / SGQNK / AWLGND		72/56/52/46/ 45/41/40/40/ 40/38/36/35/ 35/34 37/104/40/5 8/66/58/71/3
38	iron binding protein	gij62867684	6.82/77 704.3	54/60.2	ADALSLDSG / ADALSLDSGKVFEEAVAPYK / ANLMAYR / APGSPNSR / APYSGYGSFAFH / ATYLDCLK / AVANFFSASCV /		

					CAPGS / CAVGK / DCAVVDD / DEFELLCTNGQR / DGKGDVAFVK / DLLFK / DSALKLK / EAGLAPYK / EDMTQWC / EKDEYELLCLDGSR / EPASYFAVAVVK / FYGYTGAFR / GSALSYYTVAVVK / GYTGAF / HAVAVR / HSLVEEETNGK / KGTDFTLNNLQGK / KVDDLWSFLSK / LCQLC / LCTDGKR / LFGPPGK / LLCLDGSR / LLQVC / LQPLAAEVY / NFEEYFSEGCAPGSP / NLPLGTLLKR / NPSDLLQVCTFL / PLAAEVY / PLGTLLQR / QAVANFFSASCV / QVCTF / SALQSL / SYYTVAVVK / TSCHTGLGR / TTSTFELFGP / TTSTFQLFGPPGK / VALNCLQK / VMDFHE / VYEHTD / WCTLSSAEENK / WNNLQGK / WNVLSN / YFSEGCAPG / YGYTGAFR	6/44/35/68/6 1/34/35/56/3 5/95/80/44/6 5/37/35/58/7 4/66/38/44/4 9/53/35/52/4 40/52/86/50/ 37/79/41/36/ 59/64/40/67/ 34/34/43/55/ 44/34/35/52
					AGEPASYFAVAVVK / DGKGDVAFVK / GDVAFIK / YGYTGAFR	85.8/63.5/49 .8/40.2 53.8
39	PREDICTED: similar to centrosomal protein 164kDa	gij 11810197 4	5.54/12 9943.4	1/0.6	SKQEILK	
40	angiopoietin 2	NP_990148. 1	6.16/46 646.8	3/6.9	YCELDSYGNGWTVLKR / CTQMVT / RTWRAY	72/36/36
	PREDICTED: similar to IgG Fc binding protein, partial fibrinogen, gamma chain	gij 11810714 1	5.71/21 258.9	1/4.7	LPISLAQ GK	52.1
41	recombination activating protein 1	NP_990320. 1	5.48/49 642.5	3/6.9	VYCELDSYGNGWTVLKR / TLPYMLR / HCQQTC	115/38/34
		gij 12621250 0	8.68/10 9642.3	1/1.3	DMEEEILEXMKAK	
42	PREDICTED: similar to centrosomal protein 164kDa	gij 11810197 4	5.54/12 9943.1	1/0.6	SKQEILK	54.1
43	PREDICTED: similar to angiotensinogen	XP_419584. 1	5.25/51 346.8	6/15.5	LPLQSSAWLQQLSPR / DLDPSTNLLMVVDVR / EPKEFWVDSN / AVDFTNPGEAAK / QLNAFVEAK / VPFLSVTGTGFK	87/75/72/71/ 64/54
44	PREDICTED: similar to CNKSR1 protein	XP_417698. 2	5.71/33 279.1	1/3.7	CYCETQAEDPN	65
45	PREDICTED: pre-fibrinogen alpha subunit	gij 11808977 0	5.69/82 420.8	4/3.4	SFDYQVDK / DSNVPEHFK / LEVDIDIK / RLEVDIDIK	63.5/38/57.8 /64





Institut Pluridisciplinaire Hubert Curien

Service de reprographie  
Michel Schlée

23, rue du Loess  
67037 Strasbourg Cedex 2