

THESE

présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE LOUIS PASTEUR

DE STRASBOURG

Par

Laurent MIGUET

Nouvelles méthodologies dans l'analyse protéomique par spectrométrie de masse. Application à la recherche de biomarqueurs dans le cadre des leucémies.

Soutenue le 8 septembre 2006 devant la commission d'examen :

Dr. Alain VAN DORSSELAER

Directeur de thèse

Prof. Marcel HIBERT

Rapporteur interne

Prof. Jean-Marie SCHMITTER

Rapporteur externe

Dr. Thierry RABILLOUD

Rapporteur externe

Dr. Noelle POTIER

Examineur

Dr. Laurent MAUVIEUX

Examineur

A ma famille, mes amis

A Vanessa

« Si vous ne pouvez expliquer un concept à un enfant de six ans,
c'est que vous ne le comprenez pas complètement. »
« Un problème sans solution est un problème mal posé. »

A. Einstein

« Ce que l'on conçoit bien s'énonce clairement,
et les mots pour le dire arrivent aisément. »

N. Boileau

Remerciements

Pour commencer, je souhaite remercier Alain Van Dorsselaer pour m'avoir accueilli dans son laboratoire et surtout pour m'avoir soutenu et encouragé tout au long de ce travail de thèse.

Merci pour ses conseils et ses recommandations qui m'ont permis de développer un esprit critique ; merci de m'avoir appris mon métier : la spectrométrie de masse.

Je tiens aussi à remercier l'ARC pour m'avoir financé lors de ma dernière année de thèse.

J'adresse mes plus vifs remerciements au Professeur Marcel Hibert qui a accepté de présider ce jury de thèse. Je remercie également le professeur Jean-Marie Schmitter et le Docteur Thierry Rabilloud qui ont bien voulu prendre le temps de juger ce travail.

Un grand merci également à Noëlle Potier pour m'avoir encadré lors de ce travail de thèse, merci pour les discussions « remonte moral » qui m'auront permis de « recharger les batteries », merci pour m'avoir fait confiance sur les différents projets que tu m'as confiés.

Ce travail de thèse n'aurait pas pu aboutir sans la profonde collaboration qui s'est mise en place avec le docteur Laurent Mauvieux. Je tiens tout particulièrement à le remercier pour nos longues discussions et pour la dynamique qu'il a su insuffler aux différents projets que nous avons menés ensemble.

Je tiens aussi à remercier l'équipe d'Isabelle Schalk pour l'étude sur *Pseudomonas aeruginosa*.

Enfin un grand merci à tout le LSMBO avec, dans le désordre :

Guillaume C. qui m'aura appris à faire mes premiers pas dans la spectrométrie de masse et pour nos nombreuses soirées passées ensemble à se serrer les coudes ou à s'entretuer sur internet. Je te souhaite le plus beau mariage du monde (après le mien...) ; Christine C. qui aura eu le mérite de me supporter dans le bureau pendant toutes ces années ; Haiko, compagnon de repas qui est le prochain sur la liste (tu vas voir ça va bien se passer) ; Philippe avec qui le SELDI aura été une belle aventure.

Merci à tous ceux qui prennent la relève : Seb pour les discussions et les manip sur la chromato, mais aussi pour nos longues conversations footballistiques (Allez Racing...) ; Audrey pour le peps et la bonne humeur quotidienne (tu as encore de la marge avant de ressembler au robot d'Actarus).

Merci à la Jeunesse : Laetitia, n'oublie pas de te faire ton emploi du temps et tu verras ça ira tout seul, Agnès et son Hercule j'espère que le canapé va tenir, Julien qui va découvrir la Hollande pendant que je découvrirai l'Angleterre (merci pour les matchs), Cédric qui, bien que supporter de l'OM, ira loin en SM, et Guillaume B., merci pour avoir fait un super travail sur la repro ; je te souhaite beaucoup de réussite pour le reste.

Merci aux post-docs, nouveaux et anciens, Flavie et Dimitri pour leurs réponses aux questions biologiques, Andy pour ses cours d'anglais qui vont me servir de l'autre côté de la Manche, et à Fabrice B., un merci spécial pour ses corrections, pour tout le temps qu'il aura passé à m'aider et pour ses discussions sur les statistiques. J'espère que tu feras ton trou dans le monde de la spectro de masse.

Merci aux permanents : Danièle pour avoir pris le temps de m'apprendre à travailler proprement et à faire des gels droits, Sarah et ses rédactions de projets incessantes, Christine S. et tous ses conseils sur la chromatographie, JMS pour ses nombreuses photos au cours de toutes nos péripéties, Manue pour ses discussions sur la protéomique, Véro et ses attentions permanentes, Raymond et ses conseils aussi bien sur les machines que sur la Vie, et René pour ses conseils bricolages.

Enfin un merci spécial aux deux occupants de mon « deuxième bureau » Hélène et Fabrice (version alpha), ce n'est qu'une page qui se tourne mais il en reste beaucoup à écrire, n'oubliez pas de me rendre visite à Oxford...

Et à tous ceux que j'ai oubliés, Merci.

PLAN GENERAL

Introduction générale 13

Les objectifs de ce travail de thèse15

Partie I : Introduction à la protéomique et à la peptidomique, bibliographie et mise en œuvre expérimentale 17

Chapitre I : Objectifs de l'approche protéomique.....19

Chapitre II : L'outil Spectrométrie de Masse pour l'étude protéomique. Principes et description..... 26

Chapitre III : Mise en œuvre expérimentale et méthodologie : De la préparation de l'échantillon à l'analyse par spectrométrie de masse, le cas des protéines membranaires51

Partie II : Résultats 81

Chapitre I : Analyse différentielle et quantification par spectrométrie de masse de protéines issues de mélanges complexes 83

Chapitre II : Elaboration de nouveaux outils et de nouvelles stratégies analytiques appliqués à l'analyse des protéines membranaires par spectrométrie de masse 103

Chapitre III : Recherche de nouveaux biomarqueurs dans le cadre des hémopathies lymphoïdes chroniques 159

Conclusions générales..... 199

Partie Expérimentale.....205

Annexes 219

PLAN DETAILLE

Introduction générale.....	13
Les objectifs de ce travail de thèse.....	15
Introduction à la Protéomique et à la peptidomique, Bibliographie et Mise en œuvre Expérimentale	17
.....	
Chapitre I : Objectifs de l'approche protéomique.....	19
1) Définition de la protéomique	19
2) Approche protéomique descriptive : Obtenir un catalogue	20
3) Approche protéomique différentielle et quantification.....	22
4) Recherche de modifications post-traductionnelles	22
5) Protéomique Clinique et recherche de biomarqueurs	23
6) La complexomique : étude de la machinerie cellulaire.....	24
7) Utilité de la spectrométrie de masse pour l'étude protéomique.....	25
Chapitre II : L'outil Spectrométrie de Masse pour l'étude protéomique. Principes et descriptions.....	26
1) Les sources.....	26
1.1) Source par ionisation électrospray (ESI).....	27
1.1.1) La production de gouttelettes chargées à partir d'une solution.....	27
1.1.2) Fission des gouttelettes chargées, l'explosion coulombienne	28
1.1.3) Emission des ions en phase gazeuse.....	29
1.1.3.1) Modèle de Dole (la charge résiduelle).....	29
1.1.3.2) Modèle d'Iribarne et de Thomson	30
1.1.4) Miniaturisation des sources électrospray.....	30

1.1.5) L'interface	31
1.1.5.1) Les géométries des interfaces	31
1.1.5.2) La fragmentation en source	33
1.2) Source MALDI.....	33
1.2.1) Le rôle de la matrice	33
1.2.2) Le principe d'ionisation.....	34
1.2.2.1) Le laser	34
1.2.2.2) Le choix de la matrice	34
1.2.2.3) Le mode de dépôt	35
a) La méthode de la goutte séchée	35
b) La méthode en couche mince	35
c) La méthode en sandwich.....	35
1.2.2.4) Les cibles utilisées pour les dépôts.....	36
1.2.2.5) Le mécanisme d'ionisation MALDI.....	36
2) Les Analyseurs.....	37
2.1) Performance des analyseurs et résolution	37
2.1.1) Vitesse de balayage	38
2.1.2) La résolution.....	38
2.2) L'analyseur Quadripolaire.....	39
2.3) L'analyseur à temps de vol.....	41
2.3.1) Principe général.....	41
2.3.2) Le mode linéaire	42
2.3.3) L'extraction retardée	43
2.3.4) Le réflecteur	43
2.3.5) La source d'ion orthogonale.....	44
2.3.6) La fragmentation en MALDI-TOF.....	45
2.3.6.1) Le mode PSD.....	45
2.3.6.2) Le mode Lift.....	45
2.4) L'hybride Q-TOF	46
2.5) La trappe ionique.....	46
2.5.1) Description	46
2.5.2) Principe de fonctionnement.....	47
2.5.2.1) Analyse en mode MS.....	47
2.5.2.2) Analyse en mode MS/MS.....	49
3) Le SELDI (Surface Enhanced Laser Desorption Ionisation), une technique alternative pour l'étude des biomolécules	49

3.1) Présentation de la technologie SELDI-TOF	49
3.2) Les différentes surfaces chromatographiques	50
Chapitre III : Mise en œuvre expérimentale et méthodologie : De la préparation de l'échantillon à l'analyse par spectrométrie de masse, le cas des protéines membranaires	51
1) Les méthodes de séparation des protéines. Le problème des protéines membranaires	51
1.1) Le Gel 2D, historique et possibilités	51
1.1.1) La première dimension	52
1.1.2) La préparation de l'échantillon	53
1.1.3) Les colorations	53
1.1.4) Les gels 2D à haut débit et la standardisation	53
1.1.5) La méthode d'électrophorèse Off-Gel™ (OGE)	54
1.2) Le couplage chromatographie liquide-spectrométrie de masse	55
1.2.1) La chromatographie liquide off-line	55
1.2.2) La chromatographie liquide on-line	56
1.3) Le gel 1D, une alternative aux gels 2D, application aux protéines membranaires	57
2) Simplification des échantillons	57
2.1) Augmenter le nombre d'étapes de séparations	57
2.2) Limiter le nombre de protéines : l'étude de sous protéomes	58
2.3) Un sous protéome particulier : la membrane plasmique	58
2.3.1) Les différents types de préparations de membranes plasmiques	58
2.3.2) Les microparticules	59
3) Digestion des protéines et extraction des peptides	60
3.1) Digestion des protéines	60
3.2) Extraction des peptides	60
4) Analyse protéomique par spectrométrie de masse	60
4.1) Présentation de l'approche protéomique par spectrométrie de masse	60
4.2) L'empreinte peptidique massique (masse fingerprint) / approche MALDI-MS	61
4.3) Information de séquence, approche MS/MS	62
4.3.1) La fragmentation peptidique	62
4.3.2) Le séquençage <i>De Novo</i>	63
4.4) La quantification des protéines	63

5) Traitement informatique et recherche dans les bases de données.....	64
5.1) Recherche dans les banques de données	65
5.2) Les critères d'identifications	65
5.3) Archivage et bases de données.....	66
6) Conclusions de cette étude bibliographique et perspectives d'études	66
Références bibliographiques	69

RESULTATS81

Chapitre I : Analyse différentielle et quantification par spectrométrie de masse de protéines issues de mélanges complexes..... 83

1) Etude des protéines membranaires impliquées dans le transport du fer chez *Pseudomonas aeruginosa*. Approche différentielle par électrophorèse bidimensionnelle et analyse par MALDI-MS..... 84

1.1) Introduction et intérêt biologique..... 84

1.2) Approche protéomique différentiel par Gel 2D et analyse par empreinte peptidique massique (MALDI-MS)

 1.2.1) Préparation des échantillons de protéines membranaires..... 85

 1.2.2) Electrophorèse 2D

 1.2.3) Analyse MALDI..... 86

1.3) Résultats

1.4) Conclusions et perspectives

2) Quantification par marquage isotopique, développement de nouveaux marqueurs sans étape de chromatographie d'affinité (TAG). Application à des mélanges complexes

2.1) Introduction..... 90

2.2) Choix du marqueur et voie de synthèse

2.3) Validation du marqueur pour des études par spectrométrie de masse..... 93

 2.3.1) Spécificité et efficacité du marqueur..... 93

 2.3.2) Compatibilité du marqueur vis-à-vis des analyses protéomiques

 2.3.3) Compatibilité du marqueur vis-à-vis de la chromatographie liquide

2.4) Application à des cas biologiques

 2.4.1) Cellules Hela

 2.4.1.1) Préparation des cellules Hela et marquage des échantillons

2.4.1.2) Résultats	98
2.4.2) Protéines membranaires issues de <i>Acinetobacter baumannii</i>	99
2.4.2.1) Préparation des échantillons et marquage	99
2.4.2.2) Résultats	100
2.4.3) Complexes pré-ribosomique.....	100
2.4.3.1) Préparation des échantillons et marquage	100
2.4.3.2) Résultats	101
2.5) Conclusions.....	101
2.6) Perspectives : développement d'un marqueur C13	101

Chapitre II : Elaboration de nouveaux outils et de nouvelles stratégies analytiques appliqués à l'analyse des protéines membranaires par spectrométrie de masse..... 103

1) Intérêt biologique des protéines membranaires 103

2) Développement d'un nouvel outil pour l'étude des protéines membranaires : le cas des microparticules..... 105

2.1) Introduction aux microparticules 105

2.1.1) Génération des microparticules 105

2.1.2) Stratégie employée pour l'étude des microparticules..... 106

2.2) Validation de l'outil « microparticules »..... 106

2.2.1) Composition protéique des microparticules 107

2.2.1.1) Différence dans l'expression des protéines en fonction du mode de stimulation..... 107

2.2.1.2) Reproductibilité de la composition protéique des microparticules.. 107

2.2.2) Comparaison de la composition en protéines des microparticules par rapport à une préparation membranaire classique..... 108

2.3) Publication..... 109

2.4) Conclusions..... 129

3) Combinaison des données MALDI et ESI, et reproductibilité de l'analyse nanoLC-MS/MS 130

3.1) Introduction 130

3.2) Les données MALDI..... 130

3.3) Les données nanoLC-MS/MS : un catalogue de protéines 131

3.4) La reproductibilité de l'analyse nanoLC-MS/MS..... 132

3.4.1) Stratégie expérimentale	134
3.4.2) Résultats	135
3.4.2.1) Effets mémoires	135
3.4.2.2) Nombre d'analyses pour identifier « toutes » les protéines de l'échantillon.....	135
3.4.2.3) Estimation du nombre minimum d'analyses à réaliser.....	136
3.5) Utilisation combinée des données LC-MALDI et LC-ESI	137
3.5.1) LC-MALDI off-line	137
3.5.2) Génération d'un spectre unique.....	139
3.5.3) Combinaison des données MALDI et ESI	142
3.6) Conclusions et perspectives	143
3.7) Publication.....	143
4) Conclusions générales.....	160

Chapitre III : Recherche de nouveaux biomarqueurs dans le cadre des hémopathies lymphoïdes chroniques 159

1) Contexte biologique	160
1.1) Définitions.....	160
1.2) Problématique	160
2) Recherche et identification de biomarqueurs de leucémies : méthodes classiques.....	161
2.1) Les différents biomarqueurs connus	162
2.2) Classification des différentes leucémies	162
3) Recherche et identification de biomarqueurs de leucémies : approche protéomique par analyse des protéines membranaires	164
3.1) Production des microparticules et préparation de l'échantillon	165
3.2) Analyse par spectrométrie de masse	166
3.3) Résultats et comparaison entre les différentes pathologies.....	167
3.4) Proposition de nouveaux candidats	169
3.5) Validation de ces nouveaux candidats	170
3.6) Conclusion et perspectives	171
4) Utilisation de l'approche peptidomique par SELDI-TOF-MS sur des sérums de patients	173
4.1) Description de l'approche	173
4.2) Résultats : profiling et analyse statistique.....	175
4.3) Purification des peptides d'intérêts et identification par spectrométrie de masse.....	176
4.4) Publication.....	176

4.5) Conclusions et perspectives	189
5) Conclusions générales.....	190
Références bibliographiques	191
Conclusions générales	199
Partie expérimentale	205
1) Synthèse du marqueur NEIAM.....	207
2) Préparation des gels d'électrophorèse.....	207
3) Protéomique	208
3.1) Réduction-alkylation-digestion	208
3.2) L'extraction des peptides	209
4) La chromatographie liquide	209
4.1) HPLC Waters Alliance (purification des protéines dans le cadre de l'étude SELDI).....	209
4.2) HPLC Agilent 1100 et collecteur de fraction Agilent 1100 (fractionnement des peptides pour les expériences LC-MALDI-MS).....	210
4.3) CapLC Micromass	210
5) Les spectromètres de masse	211
5.1) Le MALDI-TOF.....	211
5.2) Les spectromètres de masse électrospray	214
5.2.1) Le spectromètre de masse LC-TOF.....	214
5.2.2) Le spectromètre de masse Q-TOF	214
5.3) Le SELDI	215
5.3.1) Préparation des échantillons.....	215
5.3.2) Acquisition des données SELDI	216
5.3.3) Traitement statistique des données.....	216
6) Recherche dans les banques de données en utilisant l'algorithme MASCOT.....	217
Annexes	219

Liste des abréviations

ActD:	actinomycineD
CID :	dissociation induite par collision
Da :	dalton
DTT :	dithiothreitol
ES :	électrospray
ESI-MS :	spectrométrie de masse électrospray
ICAT :	isotope coded affinity TAG
ICNAT :	isotope coded non affinity TAG
IPG :	gradient de pH immobilisé
LAL :	leucémie aigue lymphoblastique
LC-MS :	couplage chromatographie liquide spectrométrie de masse
LLC :	leucémie lymphoïde chronique
LNH :	lymphome non Hodgkinien à petites cellules
m/z :	rapport masse sur charge
MALDI :	ionisation/désorption laser assisté par matrice
MCL :	lymphome du manteau
ME :	membrane externe
MI :	membrane interne
MPs :	microparticules
MS :	spectrométrie de masse
MS/MS :	spectrométrie de masse en tandem
NEIAM :	N-éthyl-iodoacétamide
PHA :	phyto-hémagglutinine
Q :	quadripôle
SELDI :	ionisation/désorption laser activé par surface
SM:	spectrométrie de masse
TCEP:	tris(2-carboxyethyl)phosphine hydrochloride
TOF:	temps de vol (Time of flight)
Tris :	tris-(hydroxyméthyl)-aminoéthane

Introduction générale

Introduction générale

Les objectifs de ce travail de thèse

Ces dernières années, de nombreux génomes ont été séquencés dont celui de l'homme. Le séquençage de ces génomes ainsi que les récents développements techniques dans le domaine de la spectrométrie de masse (rapidité d'analyse et sensibilité) ont permis d'étudier à grande échelle les protéines exprimées, par analyse protéomique et spectrométrie de masse.

L'une des applications majeures de la protéomique consiste en l'analyse différentielle de l'expression protéique au niveau d'une cellule, d'un organe ou d'un organisme, et ceci en fonction des différentes conditions de son environnement.

Dans ce contexte, la recherche de biomarqueurs liés à une pathologie particulière, ou tout au moins de protéines spécifiques de l'état observé, est devenu un des points centraux de l'analyse protéomique. Dans notre étude, nous nous sommes intéressés à la recherche de biomarqueurs spécifiques de pathologies qui touchent les cellules sanguines circulantes. Pour cela, nous avons décidé de nous intéresser plus particulièrement aux protéines issues de la membrane plasmique. En effet, ces protéines sont faciles d'accès lorsque l'on travaille sur des cellules entières. De plus, il existe potentiellement un anticorps spécifique de chacune de ces protéines qui pourrait être utilisé à des fins diagnostiques (cytométrie en flux, Elisa...), voire même thérapeutiques.

Malheureusement, les différentes techniques protéomiques, qui vont de la préparation de l'échantillon à l'analyse par spectrométrie de masse, sont très peu compatibles avec l'étude des protéines membranaires. Tout d'abord, il est difficile d'obtenir une préparation membranaire que l'on pourrait qualifier de propre, à savoir non contaminée par d'autres protéines cytosolubles ou originaires des différents organites de la cellule. On peut aussi mentionner la difficulté de séparer des mélanges protéiques complexes sur des gels d'électrophorèse bidimensionnels, ou enfin celle de digérer ces protéines hydrophobes et d'analyser les peptides qui en résultent.

Dans ce contexte général, l'objectif de ce travail de thèse est de développer de nouvelles méthodes pour la recherche, l'identification et la caractérisation de nouveaux biomarqueurs spécifiques de pathologies par spectrométrie de masse et analyse protéomique. De plus, nous nous sommes attachés à développer de nouvelles méthodes pour la préparation de l'échantillon et l'étude des protéines membranaires par spectrométrie de masse.

Ce manuscrit est divisé en deux parties :

La première partie de ce mémoire est bibliographique et a pour but de définir les différents objectifs de la protéomique, de présenter les techniques de spectrométrie de masse utilisées (MALDI-MS, ESI-MS/MS), et de décrire les diverses étapes qui vont de la préparation de l'échantillon à l'analyse par spectrométrie de masse.

La deuxième partie de ce manuscrit vise à présenter les résultats obtenus lors de ce travail de thèse et est subdivisée en trois chapitres :

- **La quantification par spectrométrie de masse de protéines issues de mélanges complexes.**

Tout d'abord, nous avons utilisé une stratégie « classique » par comparaison de gels d'électrophorèse bidimensionnels et analyses par MALDI-MS sur des protéines issues de la membrane externe de bactéries. Ensuite, nous nous sommes attachés à développer et à valider un nouveau marqueur isotopique qui ne présente pas d'étape de chromatographie d'affinité. Nous avons appliqué ce marqueur à trois mélanges protéiques complexes.

- **L'élaboration de nouveaux outils et de nouvelles stratégies analytiques appliqués à l'analyse des protéines membranaires par spectrométrie de masse.**

Cette étude a pour but de caractériser pour la première fois le contenu protéique de microparticules issues de la membrane plasmique par analyse nanoLC-MS/MS. De plus, nous avons développé une méthode qui combine les modes d'ionisations MALDI et ESI en vue d'obtenir un meilleur pourcentage de recouvrement des protéines membranaires identifiées.

- **La recherche de nouveaux biomarqueurs dans le cadre des hémopathies lymphoïdes chroniques.**

Nous nous sommes intéressés à la recherche de biomarqueurs spécifiques des hémopathies lymphoïdes chroniques. Pour cela, nous avons mis en place deux approches. La première approche vise à identifier des biomarqueurs de surface en tirant avantage de l'enrichissement naturel en protéines issues de la membrane plasmique. La deuxième approche que nous avons employée cherche à identifier des biomarqueurs circulants.

Une partie annexe qui reprend la liste des protéines identifiées, ainsi que la présentation de quatre publications qui ne seront pas détaillées dans ce manuscrit. Ces publications rentrent dans le cadre de l'étude du complexome : interaction protéine/métal et analyse de complexes protéiques à partir de purification Tap-TAG [Helmlinger et al. 2004; Desplancq et al. 2005; Hublitz et al. 2005; Nomine et al. 2005].

Partie I :

Introduction à la Protéomique et à la Peptidomique, Bibliographie et Mise en œuvre Expérimentale

Chapitre 1 : Objectifs de l'approche protéomique

Chapitre 2 : L'outil Spectrométrie de masse pour l'analyse protéomique. Principes et description.

Chapitre 3 : Mise en œuvre expérimentale et méthodologie. De la préparation de l'échantillon à l'analyse par spectrométrie de masse, le cas des protéines membranaires.

Objectifs de l'approche protéomique

1) Définition de la protéomique

La protéomique s'intéresse à l'étude du protéome, c'est-à-dire à l'ensemble des protéines constituant un organisme vivant dans sa globalité, un tissu, une cellule ou un compartiment cellulaire (ex: les protéines nucléaires, les protéines mitochondriales, les protéines membranaires...) [Kenyon et al. 2002]. L'approche protéomique vise aussi à établir l'identité, la quantité et la fonction de ces protéines, et à en déterminer l'expression en fonction des conditions environnementales [Patterson et Aebersold 2003].

Lors de ces dernières années, les biologistes ont séquencé plusieurs génomes, aussi bien bactériens et végétaux que mammifères, dont l'un des plus attendu est le génome humain annoncé dans les années 2000 et terminé en 2003 [Venter et al. 2001]. Mais le fait d'avoir séquencé ces différents génomes ne permet pas de connaître les protéines exprimées dans une cellule à un instant donné de sa vie. De plus, le faible nombre de gènes eucaryotes (environ 30000 pour l'homme), et qui reste constant dans tous les moments de la vie de la cellule, laisse à penser que la complexité des différents organismes va être induite par une variation de l'expression des protéines. Cependant, la complexité de ces systèmes biologiques ne repose pas uniquement sur l'expression des gènes codants pour les diverses protéines, mais aussi sur leur quantité, les modifications post-traductionnelles, l'épissage différentiel ainsi que les différentes interactions des protéines entre elles expliquent entre autre la complexité de ces systèmes biologiques (Figure 1).

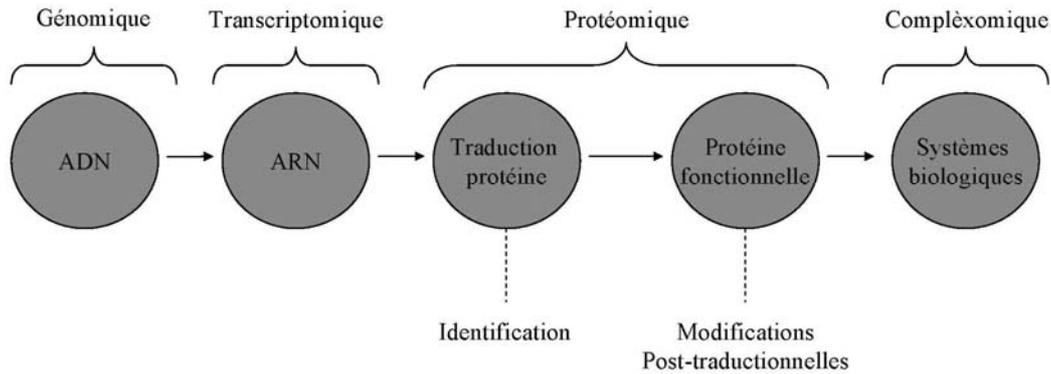


Figure 1 : De la génomique à la protéomique. D'après [Patterson et Aebersold 2003].

De toutes ces modalités, résulte l'affirmation : « *Même génome, différents protéomes* », titre de la revue de Friedrich Lottspeich qui prend comme exemple la chenille qui se transforme en papillon [Lottspeich 1999]. Ainsi, l'analyse protéomique propose de donner des cartes protéiques les plus globales d'un système biologique à un moment donné de son existence en fonction des conditions physiopathologiques.

Mais ce type d'approche n'aurait pas été possible sans le développement simultané de différentes techniques allant de la préparation de l'échantillon (gel d'électrophorèse, chromatographie liquide,...) à l'étude bio-informatique (banques de données génomique et protéique, clusters d'ordinateurs,...), tout en passant par l'analyse par spectrométrie de masse (sensibilité, résolution et précision des mesures de masse, nouvelles génération de spectromètres,...) [Pandey et Mann 2000].

2) Approche protéomique descriptive : Obtenir un catalogue

Le premier point d'étude de la protéomique est l'identification exhaustive des protéines exprimées par un organisme [Malmstrom et al. 2004] [McGregor et Dunn 2006]. Ceci a été possible grâce au développement simultané des techniques séparatives ainsi que des différents appareillages de spectrométrie de masse. De plus, cette approche a considérablement été dopée par tous les programmes de séquençage de génomes ainsi que des outils et moteurs de recherche bioinformatique. D'une manière traditionnelle, cette approche consiste à séparer un mélange protéique complexe sur un gel d'électrophorèse, à digérer les protéines les analyser par spectrométrie de masse (figure 2) [Tonella et al. 2001].

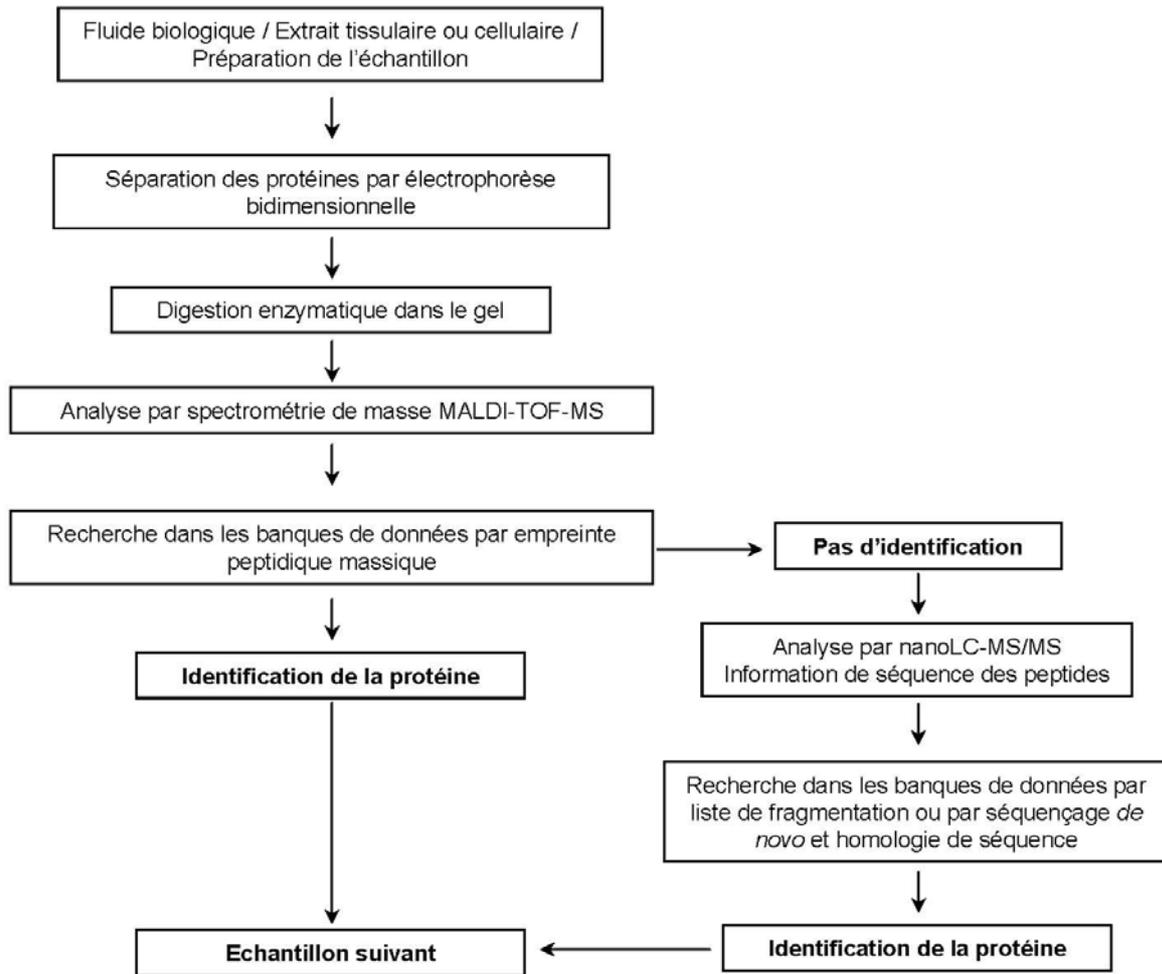


Figure 2 : Représentation schématique de l'approche protéomique descriptive.

L'application d'une telle technologie de séparation des protéines, couplée à la spectrométrie de masse a permis à un grand nombre de groupes de se lancer dans l'analyse systématique afin de réaliser des cartes protéiques de gel 2D, et donc la mise en place de bases de données construites autour des gels 2D [Toda et al. 1998] [Mollenkopf et al. 1999].

D'autres techniques ont aussi été développées par la suite afin de pouvoir identifier un plus grand nombre de protéines. Ainsi le groupe de Yates a mis en place une stratégie basée sur des chromatographies successives afin d'augmenter au maximum le nombre de protéines identifiées [Washburn et al. 2001]. Cette technique permet, par une suite de chromatographies successives (en générale une échangeuse ionique suivie d'une phase inverse), de séparer des mélanges très complexes de protéines avant de les identifier par spectrométrie de masse.

3) Approche protéomique différentielle et quantification

Le deuxième aspect important de l'analyse protéomique est l'analyse différentielle qui consiste par exemple à comparer les protéomes de deux états distincts (malade/sain ; traitement/pas de traitement ;...) en observant l'apparition, la disparition ou les variations des quantités des protéines. L'approche de quantification la plus classique consiste à séparer les mélanges protéiques sur des gels d'électrophorèse bidimensionnelle et après coloration à comparer les différences entre les gels. Les spots contenant les protéines exprimées de façon différentielle seront analysés par spectrométrie de masse (Figure 3).

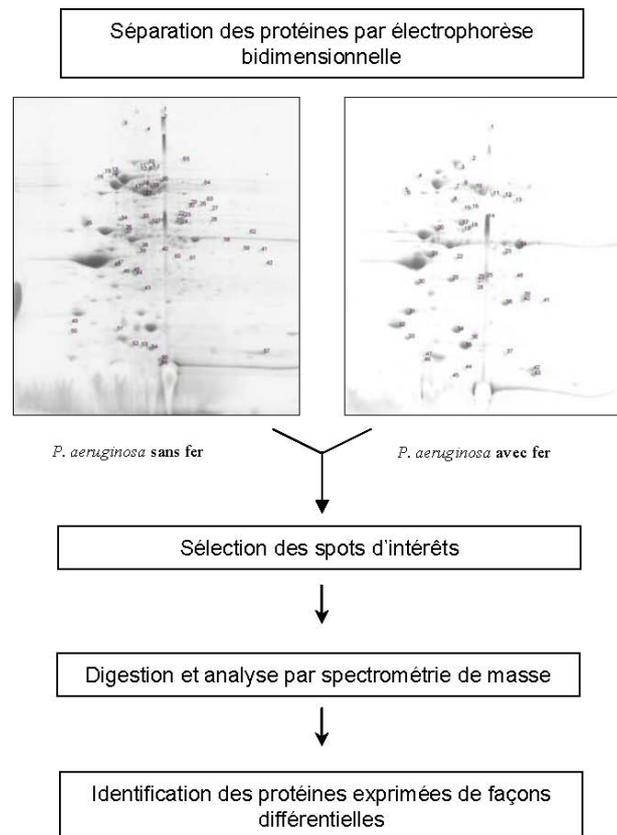


Figure 3 : Illustration schématique de l'approche protéomique différentielle.

4) Recherche de modification post-traductionnelles

Un troisième point très important pour l'étude protéomique par spectrométrie de masse est la recherche des modifications post-traductionnelles portées par les protéines d'intérêts. En effet, le faible nombre de gènes eucaryotes laisse à penser que la complexité des organismes va résulter entre autre de la maturation et des modifications post-traductionnelles des protéines. Ces modifications post-traductionnelles regroupent les modifications chimiques covalentes ainsi que la perte d'acides aminés sous l'action d'exopeptidases [Biemann 1990]. La spectrométrie de masse est aujourd'hui l'outil de

choix pour répondre aux questions des modifications post- traductionnelles portées par les protéines. Elle peut en effet identifier le type de modification (phosphorylation, glycosylation, oxydation...), déterminer son emplacement sur la séquence des protéines, voir sa structure [Mann et Jensen 2003] [Heintz 2004].

5) Protéomique Clinique et recherche de biomarqueurs

Une des thématiques les plus en vogue à l'heure actuelle est ce que l'on appelle la protéomique clinique et la recherche de biomarqueurs dans les différents fluides biologiques (sérum, urine, liquide céphalo-rachidien,...) [Veenstra et al. 2005]. Cette approche a pour rôle de rechercher des protéines ou des peptides de dégradations spécifiques d'une pathologie, permettant ainsi de mettre au point un diagnostic des différentes pathologies, facile, rapide et peu onéreux. Ces techniques consistent à simplifier le contenu protéique d'un échantillon en fonction des propriétés physico-chimiques des protéines. Une telle simplification va permettre un enrichissement des différents composés les uns par rapport aux autres, ce qui va permettre de pouvoir les analyser plus facilement par spectrométrie de masse. En effet, une telle simplification d'un mélange protéique permet de diminuer les effets de suppression de signal, et donc d'augmenter le nombre de protéines observées. Ainsi les technologies SELDI (Surface Enhanced Laser Desorption Ionisation) ou ClinProt développées respectivement par Ciphergen et Bruker, ainsi que les nombreux kits de déplétion de l'albumine ou maintenant des 20 protéines majoritaires du sérum développé par Sigma, montrent bien cet engouement [Gong et al. 2006].

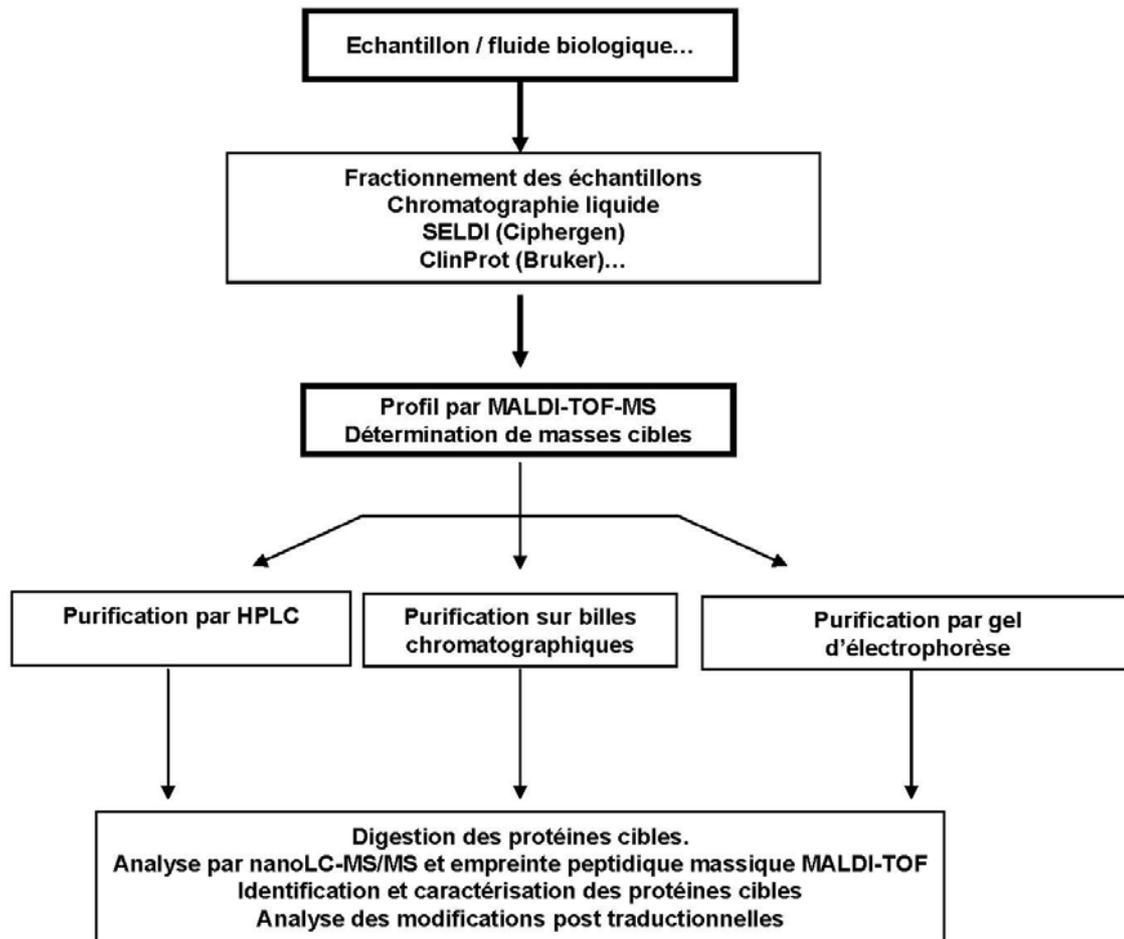


Figure 4 : Illustration schématique de l'approche protéomique clinique.

6) La complexomique : étude de la machinerie cellulaire

La complexomique est le dernier point pour lequel la spectrométrie de masse joue un rôle important. Les protéines n'agissent pas indépendamment dans les organismes, elles interagissent entre elles. Le terme complexomique représente donc l'interaction de ces différentes protéines pour réaliser leur fonction biologique [Deshaies et al. 2002]. L'étude par spectrométrie de masse intervient à deux niveaux. Tout d'abord, la spectrométrie de masse peut, après purification d'un complexe protéique (colonne d'affinité) identifier les partenaires qui interagissent avec la protéine d'intérêt, et donc de caractériser le complexe. Dans un deuxième temps, la spectrométrie de masse supramoléculaire, qui permet de garder les interactions non covalentes à l'intérieur du spectromètre de masse peut donner des informations sur les interactions protéine/protéine protéine/ligand ou protéine/métal.

7) Spectrométrie de masse et approche protéomique

La spectrométrie de masse permet donc, depuis un peu plus d'une dizaine d'années de répondre à de nombreuses questions biologiques en identifiant des protéines ainsi qu'en les caractérisant (modifications post-traductionnelles, chimiques, mutations d'acides-aminés,...). En effet, les techniques d'ionisation MALDI et ESI couplées à des analyseurs de plus en plus résolutifs et de plus en plus sensibles rendent possible l'analyse de quantités très faibles (quelques femtomoles à sub-femtomoles) de produits [Aebersold et Mann 2003].

Aujourd'hui, la protéomique a montré qu'elle était devenue indispensable afin de mieux comprendre les processus biologiques normaux ou pathologiques, pour le dépistage de nouveaux marqueurs diagnostiques ou thérapeutiques [Hanash 2003], ainsi que le développement de nouvelles molécules pharmacologiques [Loo et al. 1999].

L'outil Spectrométrie de Masse pour l'analyse protéomique. Principes et description

Ce deuxième chapitre est consacré aux principales techniques de spectrométrie de masse utilisées dans le cadre de ce travail de thèse pour l'analyse protéomique.

D'une manière générale, un spectromètre de masse est divisé en quatre régions distinctes qui sont :

- La source qui est le lieu où les ions sont générés et émis dans la phase gazeuse.
- L'interface qui assure la transmission et la désolvatation des ions entre la source et l'analyseur.
- L'analyseur qui sert à mesurer le rapport masse/charge des ions générés
- Le détecteur.

1) Les sources

Jusqu'au milieu des années 80 les processus d'ionisation étaient des processus d'ionisation très énergétiques et destructeurs car communiquant une grande quantité d'énergie aux différents analytes. On peut ainsi parler des modes d'ionisation par impact électronique ou chimique, ou bombardement par atomes, mais ces techniques, généralement destructives, ne permettent pas d'analyser, d'une manière générale, les biomolécules. En effet, la majorité des molécules analysées par ces sources d'ionisation se trouvent pyrolysées à cause du bombardement ou de la température avant d'être ionisées. Il aura fallu attendre les années 80-90 pour voir apparaître des techniques d'ionisation que l'on va qualifier de douces qui vont permettre l'étude des biomolécules. Ces techniques sont les techniques d'ionisation par électrospray (ESI) [Fenn et al. 1989] et d'ionisation laser assisté par matrice (MALDI) [Karas et al. 1987].

1.1) Source par ionisation électrospray (ESI)

Le phénomène d'ionisation par électrospray a été décrit par les travaux de Dole en 1968 [Dole et al. 1968], et il faudra attendre le milieu des années 80 pour coupler ce mode d'ionisation à un analyseur quadripolaire par le groupe de Fenn, ce qui lui vaudra le prix Nobel en 2002 [Yamashita et Fenn 1984]. Il démontre lors de ces expériences que des molécules non volatiles peuvent être ionisées et transmises en phase gazeuse sans fragmentation ou destruction de la molécule. Cette découverte va entraîner l'application à l'ionisation de biomolécules par électrospray. En effet, cette technique s'avère parfaitement compatible pour l'étude de ces composés (molécules polaires et thermosensibles). De plus cette technique d'ionisation permet d'obtenir des ions multichargés, permettant alors d'analyser des composés biologiques de plus en plus lourds, de plus en plus complexes et avec une précision de mesure de masse remarquable.

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 en phase liquide (LC-MS) [Whitehouse et al. 1985].

Ce phénomène de génération des ions produit sous l'effet d'un champ électrique intense peut être décomposé en trois étapes distinctes (Figure 5) :

- La production de gouttelettes chargées à partir de l'analyte en solution par l'application d'une haute tension sur un capillaire métallique.
- La fissions des gouttelettes chargées en gouttelettes de plus en plus petites appelées gouttelettes filles par explosions coulombiennes.
- L'émission d'ions désolvatés en phase gazeuse.

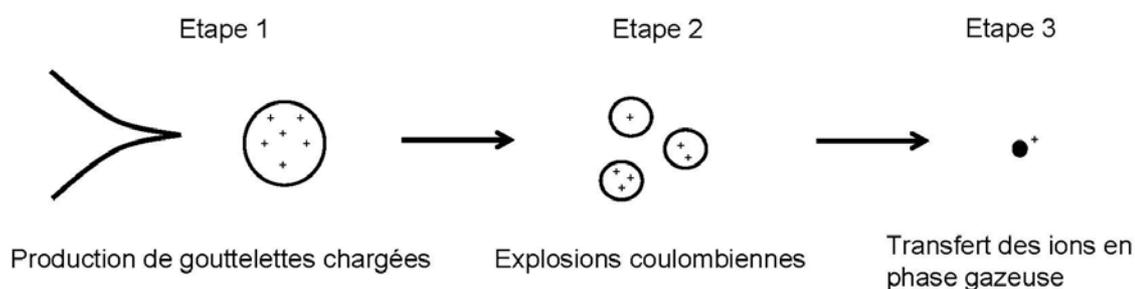


Figure 5 : Illustration de la production des ions par le mode électrospray.

1.1.1) La production de gouttelettes chargées à partir d'une solution

Le processus électrospray est obtenu par l'application d'un champ électrique intense (10^6 V/m) sur la solution d'électrolytes à analyser. Ce champ électrique va entraîner la polarisation du

liquide et la séparation des charges positives et négatives. En mode d'ionisation positif (mode le plus couramment utilisé pour l'étude des biomolécules mis à part les acides nucléiques), le capillaire métallique chargé positivement va jouer le rôle d'anode, va attirer et neutraliser les charges négatives par oxydation de l'électrolyte, du solvant ou du capillaire métallique. Il va en résulter que les charges positives vont se concentrer à la pointe du capillaire. Le liquide va alors avoir tendance à s'allonger et former un cône appelé « le cône de Taylor ». Ce cône va s'étirer jusqu'à éclater en minces gouttelettes enrichies en ions positifs (figure 6). C'est pour cette raison que l'électrospray peut être comparé à une cellule électrochimique (respect de l'équilibre des charges) [Blades 1991].

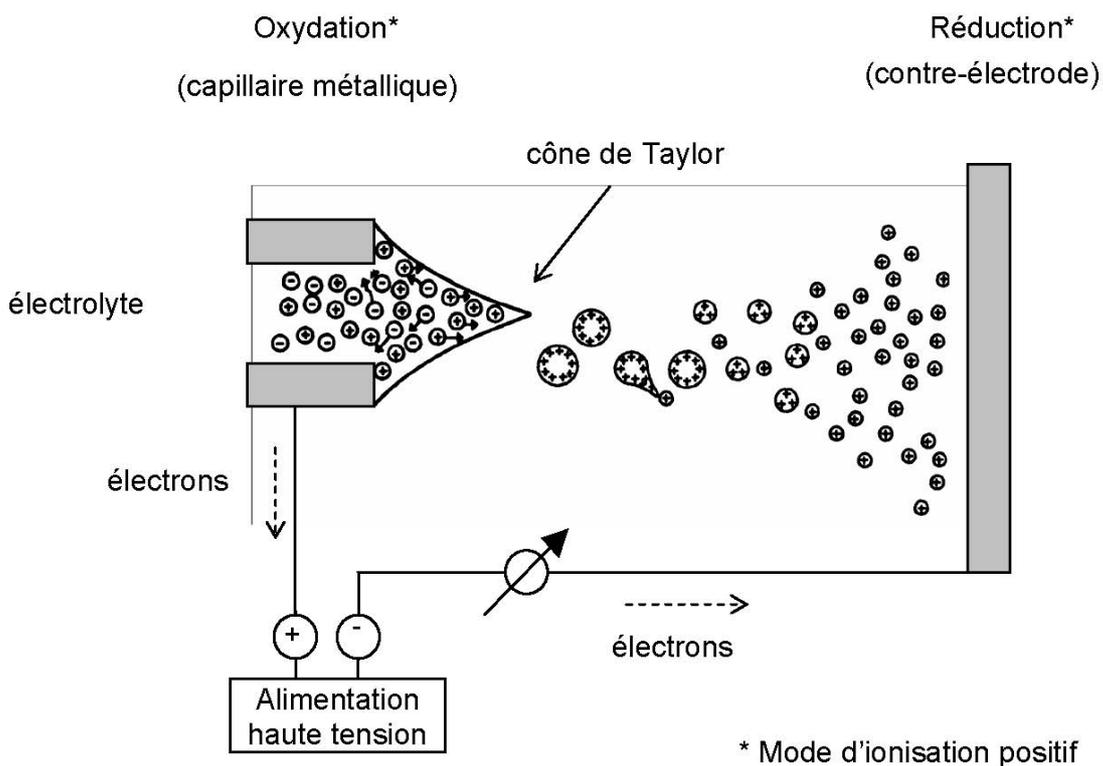


Figure 6 : Schéma d'une source électrospray.

En pratique l'électrospray standard est obtenu par l'infusion d'électrolytes à des débits compris entre 1 et 100 $\mu\text{l}/\text{min}$. Ce phénomène est en règle générale assisté par un gaz coaxial (Gaz de nébulisation, en règle générale de l'azote) afin de stabiliser la production des gouttelettes.

1.1.2) Fission des gouttelettes chargées, l'explosion coulombienne

Le solvant contenu dans les gouttelettes chargées va progressivement s'évaporer entraînant une diminution de la taille des gouttelettes. La densité de charge de la gouttelette va augmenter jusqu'à atteindre le rayon critique appelé le rayon de Rayleigh (Figure 7), moment où les forces de répulsion électrostatiques sont égales aux forces de tension de surface.

$$Q = 8 \pi (\epsilon_0 \gamma R_r^3)^{0.5}$$

Avec:

Q: la charge de la gouttelette
 Rr: le rayon critique de rayleigh
 E0: la permittivité du vide
 Y: la tension de surface

Figure 7 : Equation régissant le rayon de Rayleigh.

Au moment où les forces de répulsion coulombiennes deviennent supérieures aux forces de cohésion du liquide, la gouttelette alors instable va exploser en gouttelettes filles, qui subiront à leur tour le même processus et ceci sur plusieurs générations de gouttelettes filles de plus en plus petites jusqu'à donner des ions complètement désolvatés.

1.1.3) Emission des ions en phase gazeuse

Encore maintenant, 20 ans après les travaux de Fenn, le principe du passage des ions en phase gazeuse est un processus controversé, et deux théories sont proposées pour en expliquer le mécanisme. La première théorie est le modèle de la charge résiduelle proposé par Dole [Dole et al. 1968], la deuxième est le principe de l'évaporation ionique d'Iribarne et de Thomson [Iribarne 1976] [Thomson 1979].

1.1.3.1) Modèle de Dole (la charge résiduelle)

Ce mécanisme propose une succession d'explosions coulombiennes jusqu'à obtenir une gouttelette fille ultime ne contenant qu'un seul ion. Les dernières molécules de solvant comprises dans cette gouttelette vont s'évaporer afin d'aboutir à un ion complètement désolvaté (Figure 8).

Modèle de la charge résiduelle

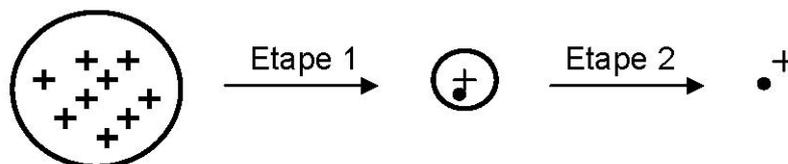


Figure 8 : Modèle de la charge résiduelle (modèle de Dole).

1.1.3.2) Modèle d'Iribarne et de Thomson (Modèle de l'évaporation des ions)

Ce deuxième modèle propose que lors du processus d'évaporation du solvant, la densité de charge est suffisante pour émettre directement les ions en phase gazeuse (Figure 9).

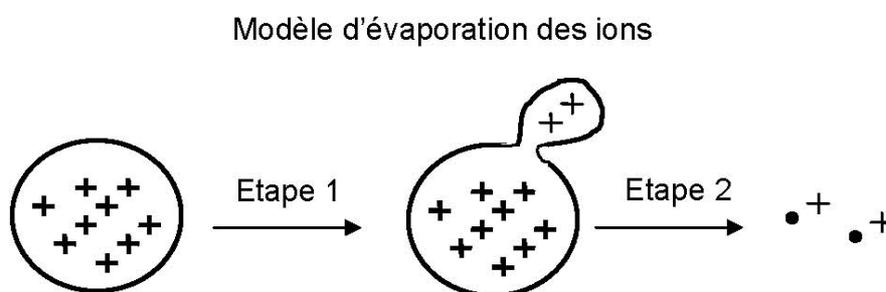


Figure 9 : Modèle de l'évaporation des ions (Iribarne et Thomson).

A l'heure actuelle il est admis que le modèle de Dole expliquerait le processus de désorption des gros ions (macromolécules, protéines,...) alors que le modèle d'Iribarne et de Thomson s'appliquerait plutôt aux molécules de petite taille [Kearle 2000].

1.1.4) Miniaturisation des sources électrospray

Comme nous l'avons mentionné précédemment l'ionisation par électrospray se réalise avec des débits assez importants (de 4 à 50 microlitres par minute). En outre, le spectromètre de masse mesure un courant d'ions total, qui est directement lié à la concentration des ions présents dans l'échantillon et non pas au débit auquel l'analyte est infusé [Ikononou 1991; Tang et Kearle 1993]. Une variante à l'électrospray va donc être de former des gouttelettes plus petites et ceci en infusant des quantités d'analytes beaucoup plus petites [Wilm et Mann 1996]. Le fait de diminuer le débit va permettre de diminuer la taille des gouttelettes émises au niveau du cône de Taylor, ce qui va augmenter la densité de charge de ces gouttelettes permettant ainsi une meilleure ionisation/désorption des molécules, ce qui équivaut à un gain substantiel de sensibilité. Dans ce cas l'assistance pneumatique n'est plus obligatoire.

Cette technique d'ionisation est donc particulièrement bien adaptée pour l'étude protéomique grâce à sa plus grande sensibilité et à sa plus grande précision en mesure de masse. Le nanospray va donc travailler avec des débits de l'ordre de quelques centaines de nanolitres par minute, permettant donc de diminuer la consommation d'analytes. On distingue deux types d'ionisation nanospray, le nanospray off-line, qui est une technique lente qui nécessite l'analyse échantillon par échantillon (non automatisable) et le nanospray on-line qui lui est couplé à un système de chromatographie liquide, permettant l'automatisation de la technique et donc l'analyse à haut débit.

1.1.5) L'interface

L'interface est la partie du spectromètre de masse qui permet de désolvater les différents ions produit par le processus électrospray et de les transmettre vers l'analyseur. L'interface va donc avoir pour but d'assurer la transmission des ions de la pression atmosphérique au niveau de la source à un vide poussé (10^{-6} à 10^{-7} mbar) au niveau de l'analyseur. L'interface aura aussi pour rôle de refocaliser le faisceau d'ions divergent ainsi que de terminer le processus de désolvatation déjà entrepris au niveau de la source.

Les ions vont être dirigés vers l'analyseur du spectromètre de masse sous l'action combinée de champs électriques et de gradients de pression, et le processus de refocalisation des ions va être obtenu grâce à l'utilisation de multipôles (quadripôles, hexapôles ou octopôles) à l'entrée de l'analyseur.

1.1.5.1) Les géométries des interfaces

Au cours de ces dernières années, de nombreuses configurations d'interfaces ont été développées et améliorées. Trois d'entre elles sont schématisées sur les figures 10, 11 et 12. Il s'agit respectivement de l'interface d'un triple quadripôle (BioQ, Waters), d'un appareil hybride de type ESI-Q-TOF (Q-TOF II Waters) et d'une trappe à ions (HCT ultra, Bruker). Ces trois interfaces présentent les trois types de géométrie les plus répandus à l'heure actuelle, à savoir une orientation dans l'axe de la source, une orientation à 90° et enfin la géométrie dite « Z-spray » et qui correspond à un enchaînement de deux coudes successifs, ce qui la rend très permissive aux sels.

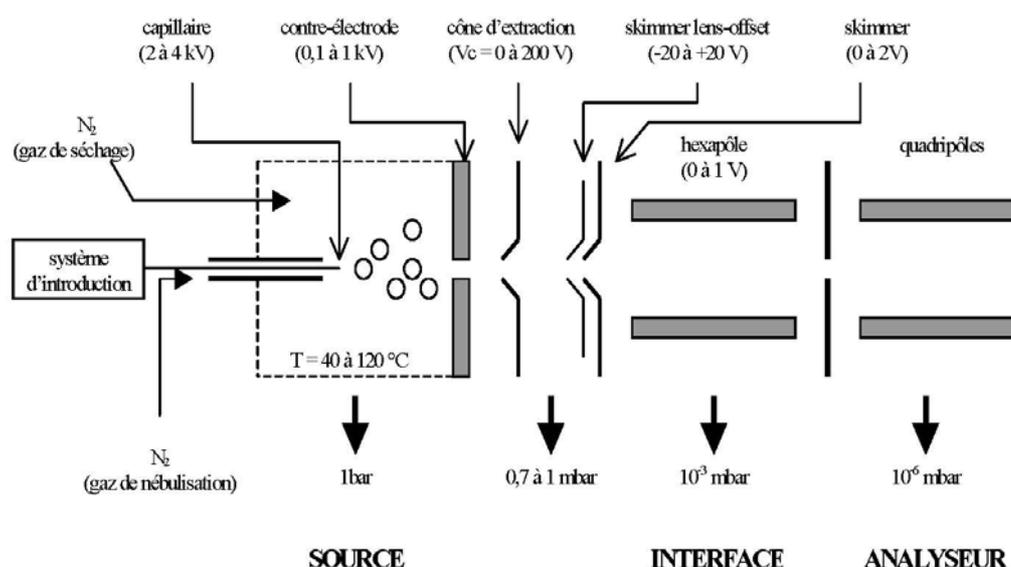


Figure 10 : Représente d'une source linéaire et de l'interface du BioQ (Waters).

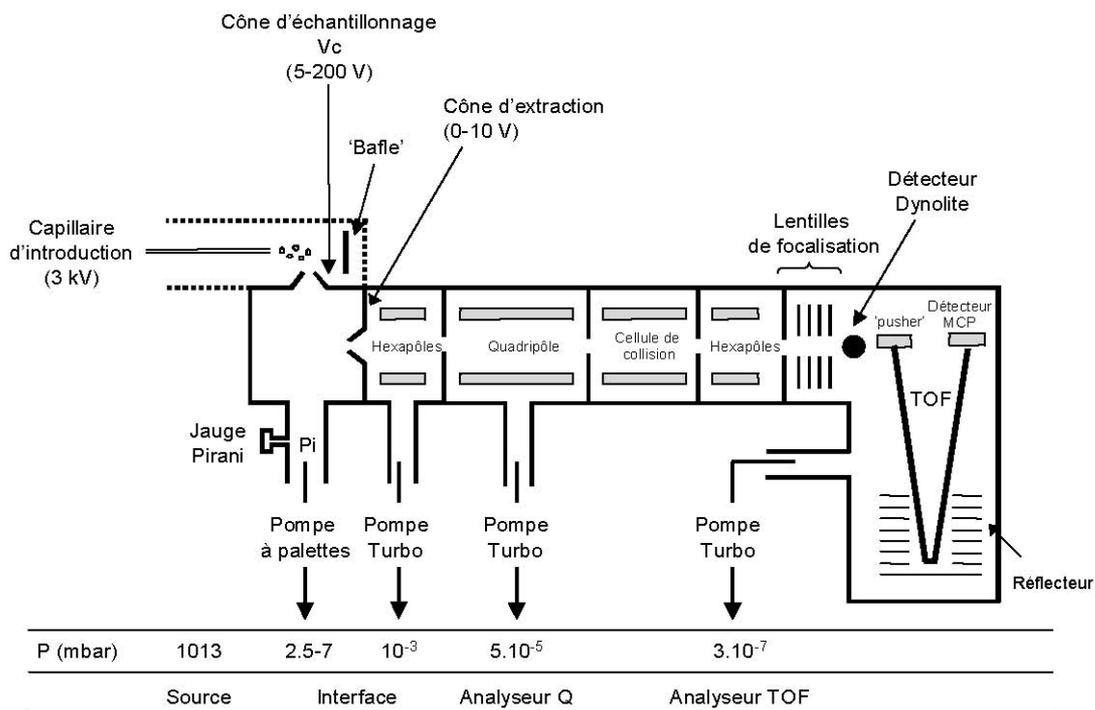


Figure 11 : Représentation schématique d'une source de type « z-spray » et de l'interface de l'ESI-Q-TOF II (Waters).

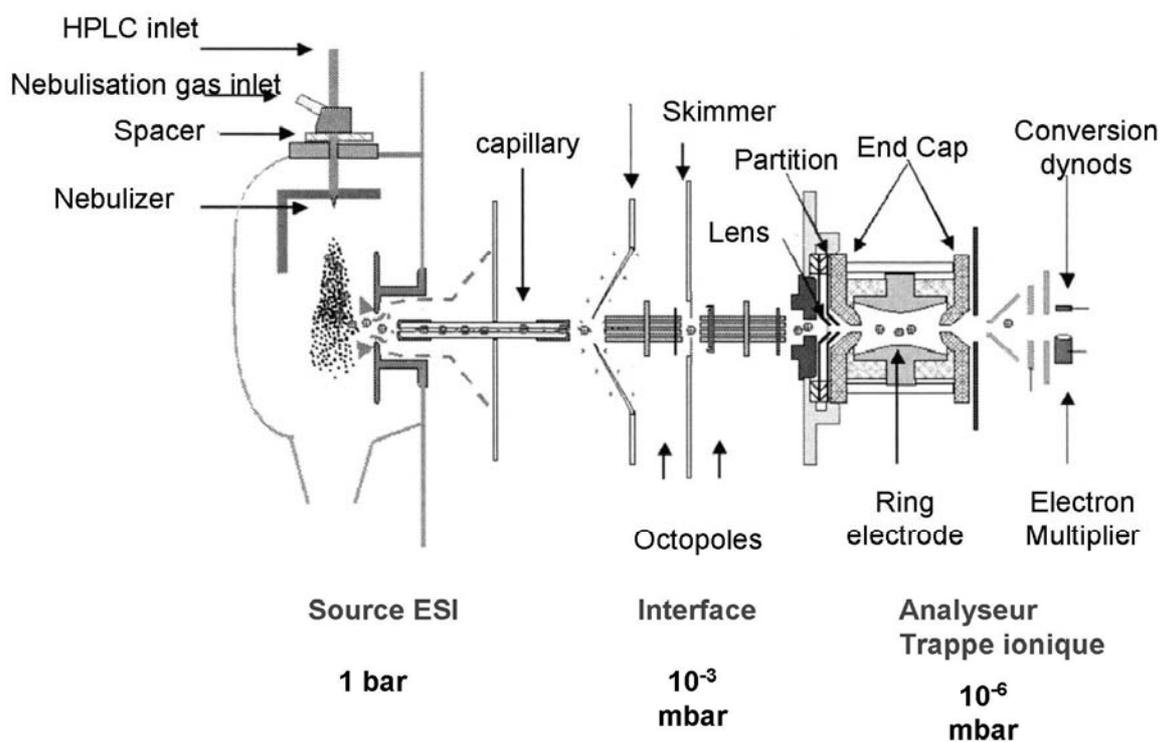


Figure 12 : Représentation schématique d'une source orthogonale et de l'interface d'une trappe ionique de HCT (Bruker).

1.1.5.2) La fragmentation en source

Les différences de potentiel appliquées entre la source et les différentes lentilles successives de l'interface vont accélérer les ions. Cette accélération va se traduire par une augmentation de l'énergie cinétique. Or il règne au niveau de l'interface une pression résiduelle de l'ordre du millibar, ce qui est suffisamment faible pour que les ions puissent accélérer, mais ce qui est suffisamment élevé pour générer un nombre élevé de collisions. Lors de ces collisions, l'énergie cinétique des différents ions se transforme en énergie interne. Cette augmentation d'énergie interne est très importante pour garantir une bonne désolvatation des ions, mais si elle devient trop importante, on va assister à des fragmentations. Il est donc nécessaire de bien contrôler les paramètres de pressions et de tensions appliquées au niveau de l'interface pour avoir la meilleure transmission et la meilleure désolvatation possible. Ces différents paramètres sont cruciaux pour l'étude des complexes non covalents [Chevreux 2005] [Sanglier 2002].

1.2) Source MALDI

La spectrométrie de masse par Désorption et Ionisation Laser Assisté par Matrice (MALDI) a été introduite à la fin des années 80 par Karas et Hillenkamp [Karas et al. 1987; Karas et Hillenkamp 1988], ainsi que par l'équipe Japonaise de Tanaka et qui lui aura valu le prix Nobel en même temps que Fenn en 2002. La technique MALDI consiste à irradier un dépôt cristallin d'une matrice organique co-cristallisant avec l'analyte. L'excitation de la matrice par le laser génère des ions moléculaires qui sont généralement analysés par un analyseur de type temps de vol (TOF), même si de plus en plus de constructeurs développent des appareils hybrides comme le MALDI-Q-TOF, ou le MALDI-Q-Trap...

1.2.1) Le rôle de la matrice

La première étape d'une analyse par MALDI-MS consiste à co-cristalliser une faible quantité de molécules à analyser avec une grande quantité de molécules de matrice. La matrice MALDI est d'une manière générale une petite molécule capable d'absorber des radiations UV/IR.

- Le premier rôle de la matrice est de co-cristalliser avec l'analyte à étudier. La matrice va être déposée en large excès par rapport au composé et va ainsi pouvoir le protéger.
- La matrice et son pouvoir dénaturant va empêcher la formation d'agrégats entre les différentes molécules à analyser, permettant ainsi l'analyse des ions moléculaires.
- Etant donné que la matrice va absorber à la longueur d'onde du laser, la désorption ionisation des molécules va être indépendante des propriétés intrinsèques de la molécule.
- Absorbant à la longueur d'onde du laser, l'adaptation des longueurs d'onde par rapport à l'analyte devient inutile.

1.2.2) Le principe d'ionisation

La complexité de l'ionisation MALDI vient de la multiplicité des phénomènes qui en sont à l'origine. Les trois paramètres les plus critiques étant le laser, le choix de la matrice et le mode de dépôt de l'échantillon avec la matrice.

1.2.2.1) Le laser

Le premier point important est de connaître le laser utilisé et la longueur d'onde employée. En effet, du choix du laser va dépendre le choix de la matrice et de l'analyte. Le laser qui à l'heure actuelle est le plus communément répandu est le laser à azote qui irradie à 337 nm.

1.2.2.2) Le choix de la matrice

La matrice est donc le point central qui va transférer l'énergie des photons à l'analyte. Son choix va se faire en fonction de la molécule à analyser ainsi que de la longueur d'onde du laser utilisé. La matrice la plus couramment utilisée pour les analyses protéomiques est l'alpha-cyano-4-hydroxynamique (CHCA) [Kussmann 1997].

Pour les protéines de plus haut poids moléculaire, la matrice de prédilection va être l'acide sinapinique, alors que pour observer les phosphorylations les oligosaccharides les lipides, le choix se portera préférentiellement sur l'acide 2,5 dihydroxybenzoïque (DHB).

Il existe aussi des matrices basiques qui vont se préparer sans solvant organique, ce qui va permettre de garder intact les complexes non covalents préformés en solution. Les principales matrices sont répertoriées dans le tableau 1.

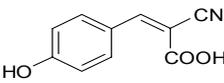
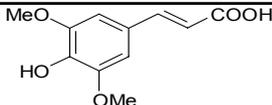
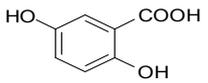
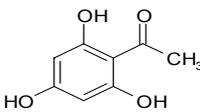
Nom de la matrice	Structure	Longueur d'onde d'absorption en nm	Application
acide α -cyano-4-hydroxynamique (α -CHCA)		337, 355 et 266 (UV)	Matrice polyvalente, très utilisée en protéomique pour l'analyse de peptides et de petite protéines
acide 3,5-dimethoxy-4hydroxycinnamique (acide sinapinique)		337, 355 et 266 (UV)	Surtout utilisée pour l'étude des protéines entières
acide 2,5-dihydrobenzoïque (DHB)		337, 355 et 266 (UV) 2,94 et 10,6 μ m (IR)	Matrice polyvalente, utilisée pour l'étude des phosphopeptides
2,4,6-Trihydroxyacetophenone (THAP)		337, 355 et 266 (UV)	surtout utilisée pour l'étude des acides nucléiques

Tableau 1 : Exemple des matrices employées en spectrométrie de masse MALDI.

1.2.2.3) Le mode de dépôt

Parmi les modes de dépôts les plus usuels, on en distingue trois :

a) La méthode de la goutte séchée

Cette technique consiste à mélanger préalablement à la solution contenant l'échantillon une solution de matrice. Ce mélange est alors déposé sur la cible et évaporé sous vide ou à l'air afin qu'ils co-cristallisent. Généralement les cristaux produits par cette technique sont assez grossiers, peu réguliers et donc très hétérogènes, ce qui est un point négatif quand à la reproductibilité ainsi que la résolution des données générées. En contre partie cette technique présente la mise en œuvre la plus rapide et la plus facile et est très bien adaptée pour l'étude des petites molécules et des peptides.

b) La méthode en couche mince

La méthode en couche mince se réalise en deux étapes. La première consiste à réaliser une solution de matrice dans un solvant très volatil (type acétone) et d'en déposer une goutte sur la cible. Ainsi, l'évaporation du solvant va être très rapide, et l'on obtiendra alors une couche très mince de cristaux de matrice qui cette fois ci seront homogènes. L'analyte est alors déposé sur cette couche mince et évaporé. Cette technique est plus sensible étant donné que les peptides sont répartis de manière uniforme, mais en contre partie le dépôt est beaucoup plus fragile, et le nombre de tirs laser sur le dépôt est donc limité.

c) La méthode en sandwich

Cette méthode dérive directement de la méthode en couche mince. Elle présente par contre une étape supplémentaire dans laquelle on recouvre la solution d'analytes par une solution saturée de matrice en faisant attention d'utiliser des solvants qui ne vont pas redissoudre la première couche de matrice. Cette méthode va générer des cristaux de petite taille très homogènes, ce qui va permettre d'augmenter notablement la résolution et la reproductibilité des différentes analyses. Cette technique est très adaptée pour l'analyse des macromolécules et les protéines.

De plus, 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 de la plupart des matrices dans l'eau. On pourra donc facilement laver un dépôt MALDI en ajoutant après la cristallisation une goutte d'eau sur le dépôt qui pourra être chassée par un flux d'air ou ré-aspirée.

1.2.2.4) Les cibles utilisées pour les dépôts

Les supports sur lesquels sont réalisés les dépôts sont appelés cibles. Elles sont en général formées d'acier inoxydable, conductrices et traitées en surface de manière à être chimiquement inerte par rapport à l'échantillon. Il existe également des cibles en verre, et plus récemment en plastique et qui sont alors recouvertes d'un film conducteur. La surface des dépôts (en goutte séchée) sur ces cibles varie de 5 à 15 mm², et seule une très faible proportion de cette surface va être balayée par le laser. Le fait de ne balayer qu'une petite partie du spot et donc de n'ioniser qu'une petite partie de l'échantillon déposé, va entraîner une difficulté au niveau de la reproductibilité de l'analyse des spots, et donc rendre très difficile l'automatisation de la technique, ainsi que la comparaison de différents profils entre eux.

Une nouvelle technologie dite « AnchorChip » développée par Bruker Daltonics, permet d'augmenter considérablement la sensibilité et la reproductibilité des analyses. Le principe est de recouvrir une plaque métallique par un composé hydrophobe à base de téflon sur lequel se répartissent des sites à caractère hydrophile appelés ancres qui présentent un petit diamètre. En séchant, les solvants organiques s'évaporent plus vite, laissant l'analyte dans une goutte d'eau qui va se concentrer spontanément sur la surface hydrophile de la plaque. Ce phénomène, indépendant du volume de l'échantillon, va permettre de concentrer celui-ci sur une surface beaucoup plus petite que sur une surface classique, et donc augmenter considérablement la quantité d'ions désorbés au cours des différents tirs laser. On va donc obtenir une reproductibilité bien supérieure, ce qui va permettre l'automatisation de la technique ainsi que la comparaison de différents spectres entre eux.

1.2.2.5) Le mécanisme d'ionisation MALDI

Le processus d'ionisation MALDI à la différence du processus électrospray s'effectue sous vide (10^{-7} mbar). Les ions sont produits en phase gazeuse sous l'effet d'une impulsion laser qui irradie la surface cristalline de la matrice qui a piégé l'analyte. Ce faisceau laser va exciter et ioniser la matrice par l'absorption d'un photon. La matrice va alors se dissocier et se sublimer, et l'analyte va être entraîné dans ce plasma d'expansion. C'est au niveau de ce plasma qu'il va y avoir transfert de charge entre la matrice et l'analyte [Glückmann 1999].

Ce phénomène d'ionisation comporte deux phases. La première phase correspond à l'ionisation de la matrice au niveau du plasma d'expansion. Cette ionisation résulte de l'absorption d'un photon émis par le laser et de la température élevée régnant dans le plasma. Le processus d'ionisation secondaire correspond au transfert d'un proton en phase gazeuse entre la matrice et l'analyte [Niu et al. 1998; Knochenmuss et Zenobi 2003] (Figure 13).

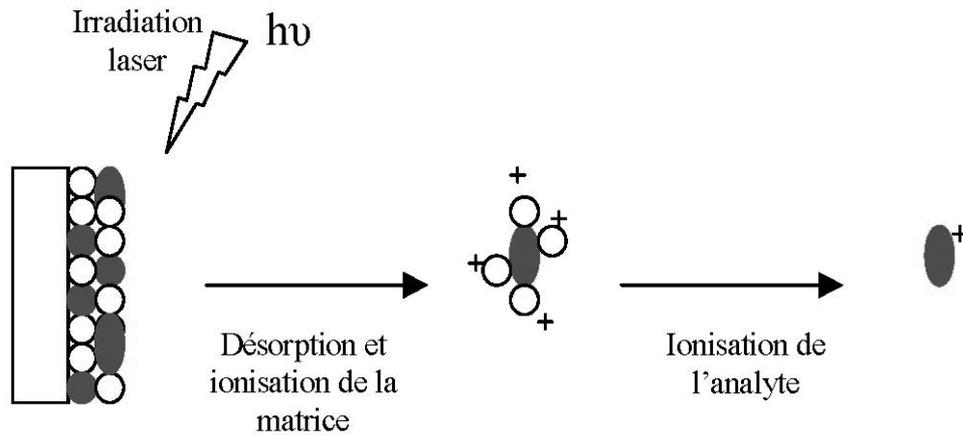


Figure 13 : Schéma du processus de désorption/ionisation par MALDI.

Les ions sont ensuite accélérés par l'application d'un champ électrique intense entre la cible et la première lentille, appelée lentille d'extraction, avant de gagner l'analyseur.

2) 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) [Yamashita et Fenn 1984], à Temps de Vol (TOF) [Olthoff et al. 1988], de type trappe à ion (IT) [Stafford et al. 1984], à résonance cyclotronique (ICR) [Henry et al. 1989], ainsi qu'à secteur magnétique [Gallagher 1990].

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 tels que des spectromètres de masse de type Q-TOF ou TOF-TOF pour ne citer que les plus courants.

2.1) Performances des analyseurs et résolution

Un analyseur se définit aujourd'hui selon deux principes, à savoir sa vitesse de balayage, c'est-à-dire le temps d'analyse qu'il va lui falloir pour balayer la gamme de masse voulue, et sa résolution.

2.1.1) Vitesse de balayage

La vitesse d'analyse varie énormément en fonction des différents types d'analyseurs. On parle de secondes pour quelques milliers d'unités de masses quand on travaille avec un quadripôle, à quelques micro secondes lorsque l'on travaille avec des analyseurs à temps de vol ou avec les dernières générations de trappes à ions qui arrivent maintenant à balayer 26000 m/z par seconde, ce qui représente des temps d'analyses de l'ordre de la milli seconde.

2.1.2) La résolution

La résolution R d'un analyseur définit la capacité de distinguer une masse M par rapport à une masse M + ΔM. 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 δm à 10% de son maximum (figure14). On parle maintenant couramment de résolution FWHM (Full Width at Half Maximum), dans ce cas, la largeur δm est prise à 50% de son maximum.

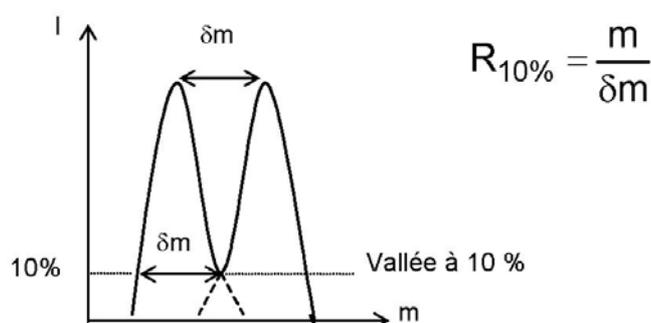


Figure 14 : La résolution est le pouvoir de séparation du spectromètre de masse, elle est définie par la relation de l'encadré qui dans ce cas se situe à 10% de vallée.

Analyseurs	Gamme de masse (m/z)	Résolution
Quadripôles	4000	Unitaire
Trappe ionique	8000	Unitaire
Temps de Vol	Illimitée	5000-20000
Magnétique	20000	20000
Résonance cyclotronique	4000	100000

Tableau 2 : Performance des différents analyseurs précédemment cités.

2.2) L'analyseur quadripolaire (Q)

Le quadripôle, développé par Paul [Paul et Raether 1955] est l'analyseur le plus couramment utilisé en couplage avec les sources électrospray. Cet analyseur est capable d'analyser un courant continu d'ions qui mesurent le rapport m/z en fonction de la trajectoire des ions dans le quadripôle. Cet analyseur quadripolaire est constitué de quatre barres métalliques parallèles [Friedman et al. 1982]. Les quatre barres sont reliées entre elles deux à deux et portées à un potentiel opposé Φ_0 et $-\Phi_0$, de sorte que deux barres adjacentes soient de potentiels opposés. Le potentiel Φ_0 comprend une composante continue U ainsi qu'une composante sinusoïdale $V \cos \omega t$ tel que : $\Phi_0 = U - V \cos \omega t$, et que $-\Phi_0 = -U + V \cos \omega t$, où ω représente une radiofréquence (quelques centaines de KHz) (figure 15) [Campana 1980].

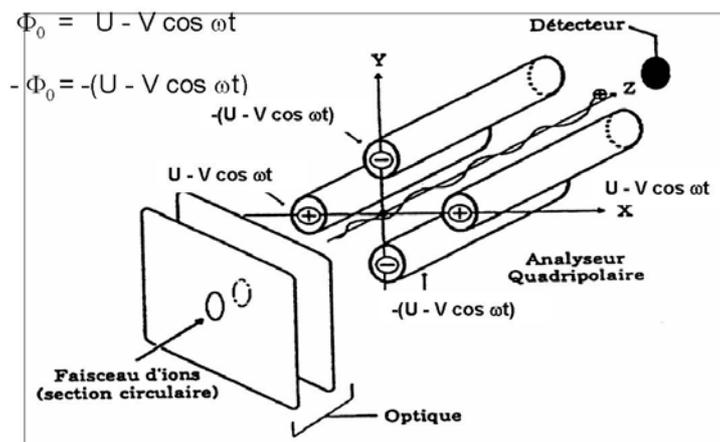


Figure 15 : Schéma d'un analyseur quadripolaire avec la représentation de la trajectoire des ions selon l'axe du spectromètre [Campana 1980].

Le quadripôle est un analyseur à balayage, il va utiliser 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 valeurs de balayage des composantes U et V selon les axes x et y et qui vont permettre de laisser passer au travers du quadripôle un m/z selon une trajectoire stable. Les équations régissant la trajectoire des ions dans un quadripôle ont été établies en 1866 par le physicien Mathieu pour décrire la propagation d'ondes dans des membranes et qui présentent deux types de solutions : les trajectoires

stables avec des valeurs U et V qui permettent aux ions de traverser le quadripôle et les trajectoires instables, qui vont entraîner la déflexion des ions entre les barres du quadripôle.

Pour un quadripôle donné, r_0 et ω sont constants, U et V sont les variables. Pour une masse donnée, il est possible de tracer le diagramme de stabilité en fonction des valeurs U et V . Avec cette représentation, les zones de stabilité d'une masse ont une forme de triangle (figure 16). On voit dans un tel diagramme qu'un balayage maintenant le rapport U/V constant permet de transmettre successivement des masses différentes.

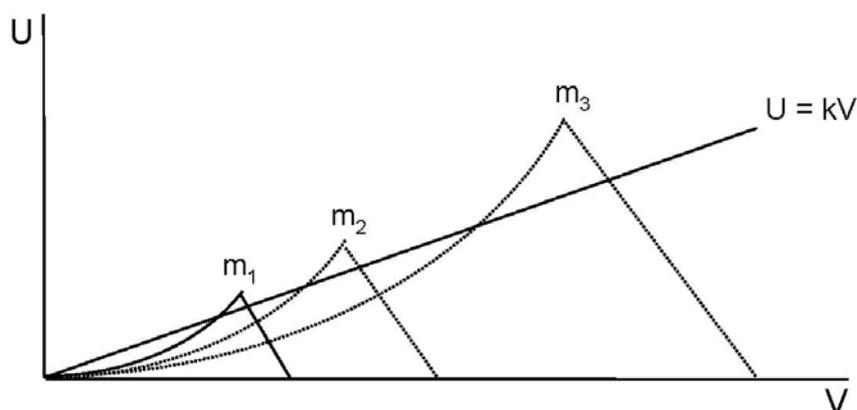


Figure 16 : diagramme de stabilité des ions en fonction des valeurs U et V pour des masses différentes m_1 , m_2 et m_3 .

En l'absence de tension continue ($U = 0$), la résolution devient nulle. Cependant, la valeur de V ne permet pas la transmission de tous les ions. En effet, si de 0 on augmente V de manière à dépasser tout juste la zone de stabilité m_1 , tous les ions de masses inférieures ou égales à m_1 auront une trajectoire instable et tous ceux de masse supérieure à m_1 auront une trajectoire stable.

Dans le cas d'un spectromètre de masse hybride de type ESI-Q-TOF, ou les triples quadripôles, le premier quadripôle peut fonctionner soit comme un analyseur (avec $U > 0$), soit comme un guide d'ions. Lorsque le quadripôle fonctionne comme guide d'ions ($U = 0$), les ions sont analysés par le second détecteur. Dans ce cas, on revient au schéma classique d'un spectromètre de masse usuel, on dit que le quadripôle fonctionne en « RF only ». Il est cependant important de noter que la transmission des ions le long du quadripôle discriminera les masses suivant la valeur V de la tension alternative. Pour moyenniser cet effet, il est possible de modifier la valeur V au cours d'une analyse afin de laisser passer successivement certaines catégories d'ions.

La gamme de m/z de la plupart des analyseurs quadripolaires commerciaux est limitée (typiquement < 4000 m/z) mais il est possible de l'étendre en diminuant la fréquence liée à la tension alternative. Une telle modification de la fréquence a permis au groupe de Robinson [Sobott et al. 2002] de sélectionner un ion dans le quadripôle à 22000 m/z , bien sûr, la diminution de fréquence entraîne malheureusement

une diminution de la résolution de l'analyseur [Winger et al. 1993]. Enfin, le quadripôle a en théorie une résolution unitaire sur toute la gamme de masse.

2.3) L'analyseur à temps de vol (TOF)

2.3.1) Principe général

L'analyseur à temps de vol est un cylindre vide dans lequel règne une basse pression avec d'un côté la source (ESI ou MALDI), et de l'autre côté le détecteur. Le principe de base de l'analyseur à temps de vol consiste à mesurer le temps que mettent les ions à parcourir la distance du tube de vol après avoir été accélérés par un champ électrique. Cet analyseur fonctionne en fait comme un chronomètre mesurant le temps de vol des différents ions.

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 champs (Figure 17).

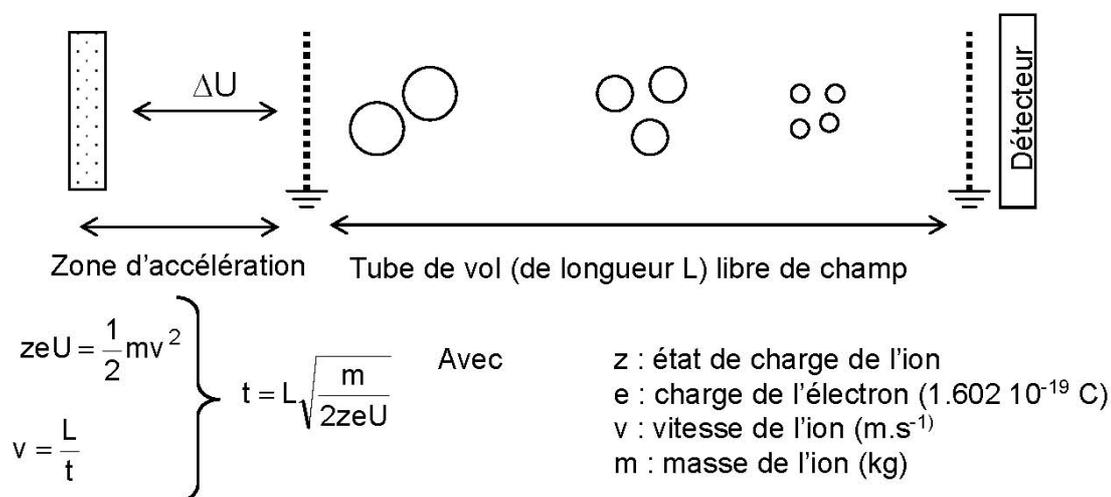


Figure 17 : Principe de l'analyseur à temps de vol.

Les ions produits en source vont être accélérés sous l'effet d'un champ électrique fixe, ce qui aura pour effet de donner la même énergie cinétique à tous les ions de même charge (sachant que $E_c = zeU$). La vitesse des ions sera alors de : $V = (2Ec/m)^{1/2}$.

Le rapport m/z des différents ions pourra alors être déduit de la mesure du temps nécessaire pour traverser le tube de vol. De plus, il faut retenir que le rapport m/z évolue proportionnellement au carré du temps.

2.3.2) Le mode linéaire

Ces analyseurs ont un mode de fonctionnement assez simple et en règle générale, couplés à des sources de type MALDI. Les ions sont générés sous forme de paquets qui correspondent aux impulsions laser utilisées par le processus MALDI. Ces paquets d'ions vont être accélérés et expulsés de la source par différents gradients de potentiel décroissant. Ces ions ainsi générés auront donc une énergie cinétique similaire, et donc des vitesses différentes en fonction de la masse des ions (figure 18).

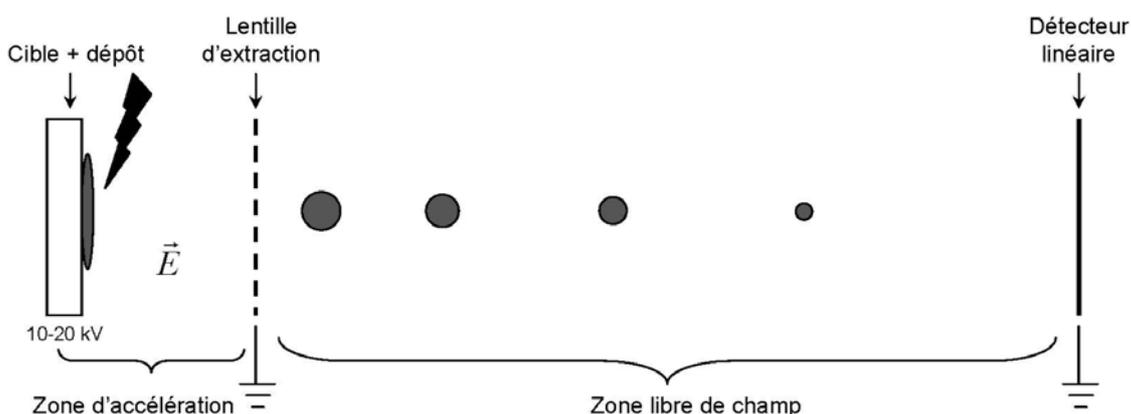


Figure 18 : Schéma du MALDI-TOF linéaire.

L'analyse de molécules en mode linéaire est particulièrement bien adaptée pour l'étude des molécules de haut poids moléculaire et présente une bonne sensibilité. Par contre, cette technique ne permet pas d'atteindre les valeurs de résolution précédemment citées, et présente une faible résolution de l'ordre de 1000. Ce phénomène est surtout dû à deux causes qui sont :

- Une distribution en énergie cinétique non homogène.
- Une hétérogénéité dans les différents paquets d'ions formés au moment de l'impulsion laser.

2.3.3) L'extraction retardée

L'extraction retardée, aussi appelée « time lag focusing » a pour rôle de diminuer au maximum la dispersion spatiale des ions générés par le processus MALDI [Wiley et McLaren 1955]. Cette refocalisation est permise par l'ajout d'une lentille sur laquelle est appliqué un potentiel légèrement supérieur au potentiel appliqué sur la cible. Ce potentiel légèrement supérieur va arrêter la progression des ions dans l'analyseur, ce qui va permettre de refocaliser le nuage d'ions et de les placer sur une même « ligne de départ ». Ce potentiel est appliqué pendant quelques centaines de nanosecondes (ce temps va dépendre de la gamme de masse que l'on souhaite analyser). A la fin de ce temps cet excès de potentiel est abaissé et les ions vont donc pénétrer en même temps dans le tube de vol. Cette refocalisation au niveau de la source MALDI va permettre d'augmenter considérablement la résolution [Cotter 1989] [Colby 1994].

2.3.4) Le réflecteur

Une seconde amélioration, qui consiste à appliquer un miroir ou réflecteur électrostatique à l'extrémité de l'analyseur, a été apportée afin d'augmenter la résolution [Mamyrin 1973]. Le réflecteur applique un champ électrique qui s'oppose à la progression des ions dans l'analyseur. Les ions vont pénétrer plus ou moins profondément dans le champ électrique du réflecteur en fonction de leur énergie cinétique initiale. Quand les ions atteignent le zéro d'énergie, ils font demi-tour et sont réaccélérés en sens inverse (figure 19).

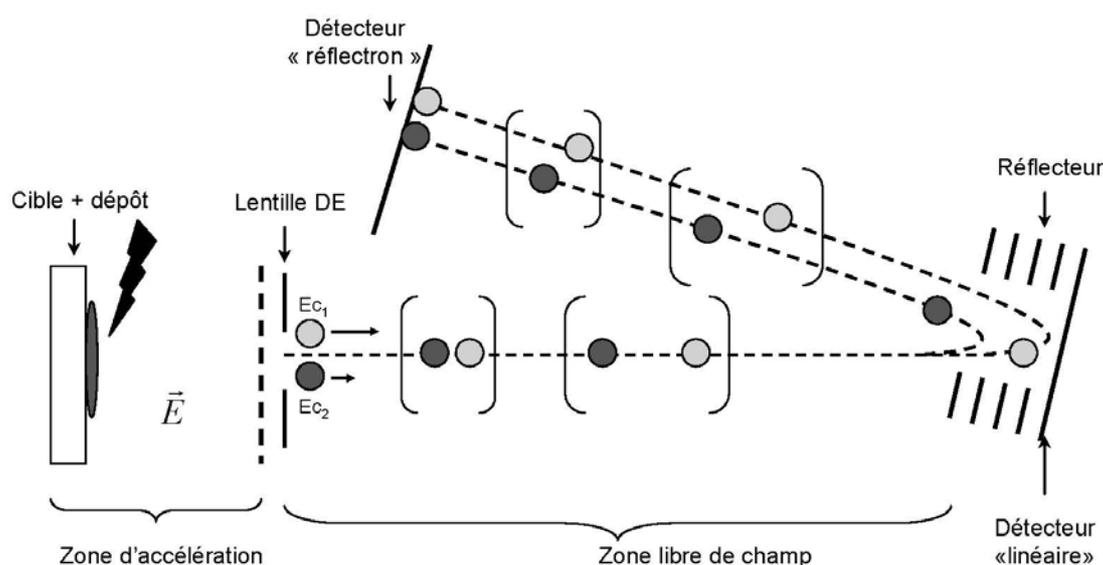


Figure 19 : Schéma du MALDI-TOF en mode réflecteur comportant un système d'extraction retardé (DE). Deux ions de même rapport m/z et d'énergie cinétique différente ($E_{c1} > E_{c2}$) pénètrent plus ou moins profondément dans le réflecteur sont réfléchis et arrivent simultanément au détecteur [Wagner 2004].

Cette stratégie peut être de nos jours conduite plusieurs fois afin d'augmenter la résolution au détriment bien sûr de la sensibilité (ultrTOF-Q, Bruker Daltonics, résolution > 40000 avec la technologie MultiPass™).

Les analyseurs TOF sont désormais des analyseurs ayant de bonnes résolutions et présentent l'avantage unique de pouvoir analyser en théorie des composés de rapport m/z infini. Le mode d'ionisation électrospray couplé au TOF semble donc particulièrement bien adapté à l'analyse de molécules biologiques de très hauts poids moléculaires [Sanglier 2002].

2.3.5) La source d'ion orthogonale

Historiquement les analyseurs à temps de vol décrits dans les années 50 [Wiley et McLaren 1955], ont surtout été couplés aux sources MALDI, car ils étaient tout à fait compatibles avec l'injection pulsée des ions par le processus MALDI. Le couplage entre le mode d'ionisation électrospray et l'analyseur TOF nécessita une injection orthogonale pulsée des ions dans l'analyseur, afin de limiter la perte de résolution due à la dispersion transversale en énergie cinétique. L'injection orthogonale se fait par l'intermédiaire d'une lentille (le pusher) qui émet un signal carré très bref, ce qui a pour effet d'envoyer dans le tube de vol des petits paquets d'ions (figure 20).

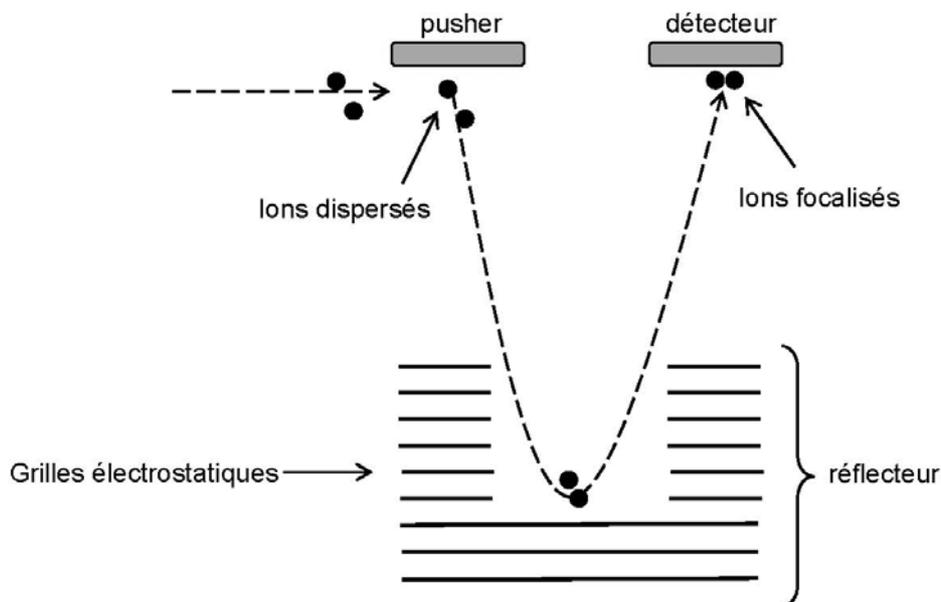


Figure 20 : Représentation schématique d'un analyseur à temps de vol avec une source d'ion orthogonale [Chevreux 2005].

Les analyseurs à temps de vol sont désormais des analyseurs ayant une bonne résolution et une bonne sensibilité. De plus, avec la source d'ion orthogonale, l'analyseur TOF est tout à fait adapté au mode d'ionisation électrospray et donc à l'analyse de différentes molécules biologiques et de peptides.

2.3.6) La fragmentation en MALDI-TOF

2.3.6.1) Le mode PSD (Post source Decay)

Lors du processus d'ionisation MALDI les ions reçoivent une grande quantité d'énergie interne (température du plasma, choc entre les différentes molécules...). Cette accumulation d'énergie interne peut provoquer la fragmentation de ces ions, mais les différents fragments ne peuvent être observables par un détecteur linéaire étant donné qu'ils ont moins d'énergie cinétique mais qu'ils conservent la même vitesse que les ions parents. Dans cette approche le réflecteur va jouer le rôle de filtre en énergie cinétique et va permettre de refocaliser les ions en énergie cinétique, ce qui va permettre de les séparer en fonction de leur rapport m/z . Cependant le réflecteur n'est pas capable de refocaliser tous les ions fragments de façon correcte. Il est donc nécessaire d'appliquer différentes tensions décroissantes au niveau du réflecteur pour pouvoir analyser tous les ions fragments. Le spectre global sera alors ensuite reconstitué en combinant les différents spectres générés. Cette approche est donc longue et fastidieuse, en plus de demander une grande quantité d'échantillon, approche donc peu compatible avec l'étude protéomique.

2.3.6.2) Le mode LIFT

La technologie TOF-TOF a été décrite pour la première fois en 2000 [Medzihradszky et al. 2000], et est adaptée à toutes les études protéomiques [Yergey et al. 2002]. Comme dans le cadre du mode PSD, les ions dans la source sont accélérés à un potentiel de 8kV et sont fragmentés dans la première région du tube de vol. A ce moment, l'ion précurseur et ses fragments volent à la même vitesse. La sélection de ces différents ions est assurée par une grille de déflexion qui se base sur le temps de vol. La cellule LIFT, placée après cette grille de déflexion a pour but de ré-acceler l'ion parent et les ions fragments avec un nouveau potentiel de l'ordre de 19 kVs. Dans la seconde partie du tube de vol, les ions fragments vont évoluer avec une vitesse dépendante de leur masse (les plus légers allant plus vite que les ions plus lourds).

Le principe de cette stratégie est résumé sur la figure figure 21.

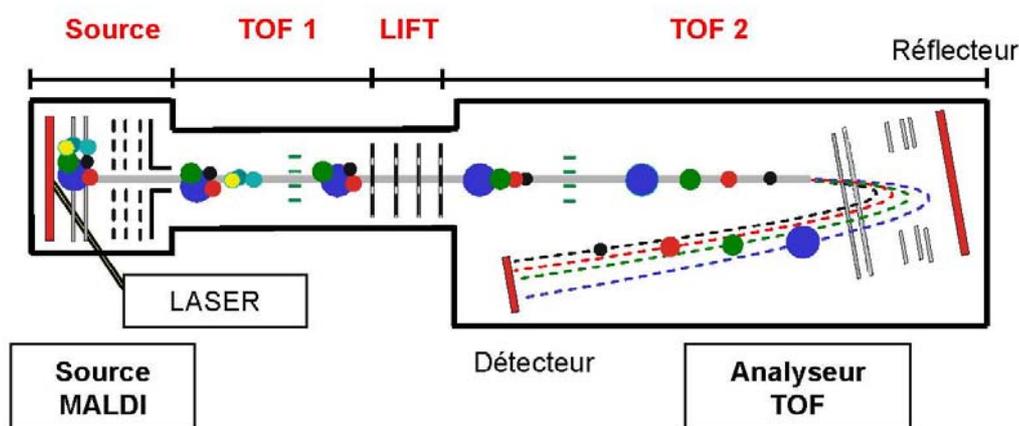


Figure 21 : Schéma du MALDI TOF/TOF ultraflex de Bruker qui présente une cellule Lift.

2.4) l'hybride Q-TOF

Le Q-TOF est un appareil hybride composé de deux analyseurs, un quadripôle couplé à un temps de vol orienté perpendiculairement le plus souvent couplé à une source électrospray (figure 9). Entre les deux analyseurs on trouve une cellule de collision, où règne une pression résiduelle de gaz (le plus généralement de l'argon). Les ions sont sélectionnés par le premier quadripôle avant d'être accélérés et fragmentés dans la cellule de collision par une succession de chocs avec les molécules de gaz résiduelles. Les différents ions fragmentés vont être refocalisés grâce à un hexapôle et une succession de lentilles qui vont ralentir le flux d'ions avant d'arriver au niveau du pusher et d'être réaccélérés par paquets vers l'analyseur à temps de vol.

Le gros avantage de ce couplage entre un quadripôle et un TOF est l'augmentation de sensibilité et de résolution de tels instruments hybrides.

2.5) La trappe ionique

2.5.1) Description

La trappe ionique est constituée de deux électrodes chapeaux de forme hyperbolique et d'une électrode annulaire [Stafford et al. 1984]. A la différence des analyseurs décrits précédemment, les ions ne traversent pas l'analyseur mais restent confinés à l'intérieur de cette trappe qui a un volume d'environ 1cm^3 . Il règne au sein de cette cellule une pression résiduelle d'hélium de l'ordre de 1.10^{-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 22).

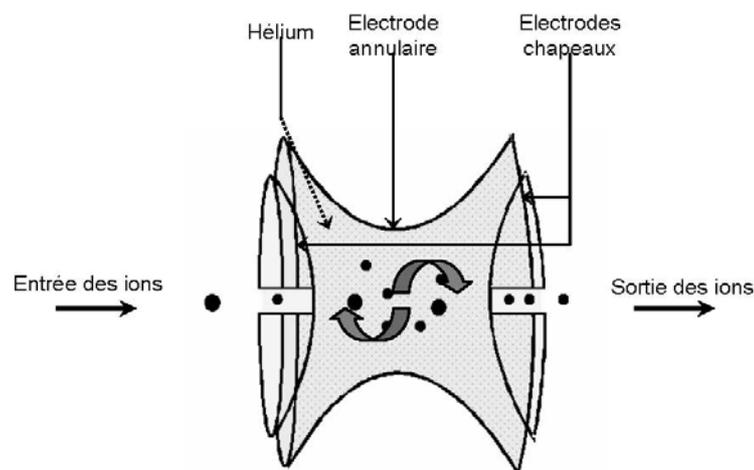


Figure 22 : Représentation schématique d'une trappe ionique.

2.5.2) principe de fonctionnement

Le fonctionnement de la trappe à ions se décompose en deux étapes lors de l'analyse MS :

- Piégeage des ions dans la trappe
- Ejection des ions

et en quatre étapes lors de l'analyse MS/MS :

- Piégeage des ions
- Isolement des ions parents
- Excitation et fragmentation
- Ejection des ions fragments

2.5.2.1) Analyse en mode MS

Les ions issus de l'interface sont piégés dans l'analyseur sous l'effet du champ quadripolaire généré par l'électrode annulaire soumise à une radiofréquence. Les différents ions piégés vont suivre une trajectoire qui va correspondre au champ quadripolaire et ainsi suivre le modèle d'une courbe de Lissajous. La stabilité des ions piégés dans la trappe va dépendre de la radiofréquence appliquée sur l'électrode annulaire. Plus la radiofréquence sera grande, plus le puits de potentiel sera important et donc meilleur sera le piégeage. Par contre, la profondeur du puits de potentiel est inversement proportionnel au rapport m/z des différents ions piégés (plus la valeur de V augmente plus les ions de faible rapport m/z auront une trajectoire instable et ne pourront pas être piégés dans la trappe) (Figure 23).

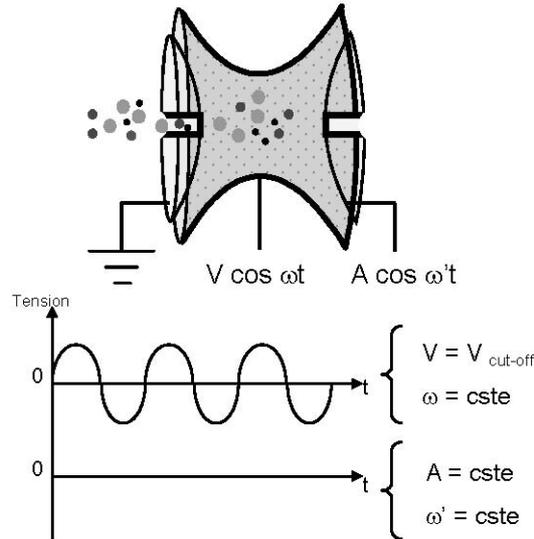


Figure 23 : Etape de piégeage des ions.

L'éjection des ions et la mesure du rapport m/z se fait selon deux manières :

La première manière consiste à réaliser un balayage croissant en amplitude de la radiofréquence appliquée sur l'électrode annulaire. Ce balayage vers des valeurs de radiofréquence de plus en plus élevées va entraîner l'éjection des ions selon des valeurs de m/z croissantes, mais cette approche n'est applicable qu'à des valeurs de m/z faibles du à la génération de décharges électriques au niveau de la trappe lorsque l'on atteint des valeurs de l'ordre de 20KV.

La deuxième méthode développée pour l'éjection des ions de la trappe est d'appliquer une tension alternative sur les électrodes chapeau, ce qui va déstabiliser les ions présents dans la trappe et entraîner leur éjection successive en fonction de leur rapport m/z et de la tension appliquée sur les électrodes (figure 24).

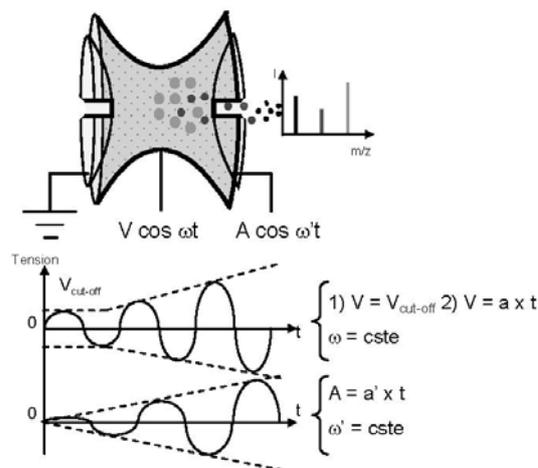


Figure 24 : Etape d'éjection des ions de la trappe.

2.5.2.2) Analyse en mode MS/MS

L'analyse en mode MS/MS d'une trappe à ion comporte deux étapes supplémentaires par rapport à l'analyse en mode MS qui sont l'isolement de l'ion parent que l'on désire fragmenter et sa fragmentation proprement dite.

La maîtrise de la trajectoire des ions dans une trappe permet d'isoler spécifiquement des ions de m/z définis correspondant à l'ion que l'on désire fragmenter. Les ions de plus faible m/z que l'ion précurseur seront éjectés de la trappe par l'application d'une radiofréquence croissante jusqu'à une amplitude proche de celle correspondant à l'éjection de l'ion parent et les ions de m/z plus élevés vont être expulsés de la trappe par l'application d'une fréquence croissante de la tension alternative sur les électrodes chapeaux. Une fois que les ions parents sont isolés au sein de la trappe, ils vont être déstabilisés par l'application d'une tension sur les électrodes chapeaux dont l'amplitude correspondra à la fréquence de résonance de ces ions. Les ions vont donc quitter leur trajectoire de stabilité et vont accumuler de l'énergie cinétique. Les ions qui vont rentrer en collision avec les molécules de gaz résiduelles vont se fragmenter. Les différents ions fils générés ne seront plus déstabilisés (la fréquence appliquée ne correspondra plus au rapport m/z des différents ions fils). Les différents ions fils seront alors éjectés de la trappe comme décrit précédemment.

Enfin ce processus d'isolement-fragmentation peut potentiellement être reproduit un grand nombre de fois. Dans ce cas on parle de MS^2 , MS^3 ... MS^n .

3) Le SELDI (Surface Enhanced Laser Desorption Ionisation), une technique alternative pour l'étude des biomolécules.

Le SELDI-TOF est une variante très proche du MALDI-TOF. En effet, la divergence ne porte que sur la surface sur laquelle les échantillons vont être déposés avant d'être analysés.

3.1) Présentation de la technologie SELDI-TOF

Le principe de la technologie SELDI-TOF [Hutchens 1993] repose sur le fait que la surface (ou Chip) sur laquelle les échantillons biologiques (peptides, mélanges de protéines, fluides biologiques complexes) sont déposés est en fait constituée d'une surface chromatographique active. Cette surface chromatographique a pour but de simplifier les différents mélanges protéiques en fonction de leurs propriétés physico-chimiques [Carlson et al. 2005] [Merchant et Weinberger 2000] [Weinberger et al. 2000]. Après différents lavages des dépôts, afin de ne retenir que les protéines spécifiques de chaque surface, on rajoute de la matrice comme décrit précédemment (§ 1.2.2.3), les peptides et/ou les protéines sont alors désorbées sous l'action d'un tir laser et transmises vers un analyseur à temps de vol linéaire (§ 2.3).

Cette technologie a été très rapidement appliquée afin de réaliser des profils peptidiques et protéiques de différents fluides biologiques pour découvrir des biomarqueurs spécifiques de différentes pathologies [Coombes 2005].

3.2) Les différentes surfaces chromatographiques

Il existe à l'heure actuelle une dizaine de surfaces chromatographiques différentes. Elles regroupent les échangeuses d'ions (anionique et cationique), les surfaces hydrophobes et hydrophiles ainsi que différentes surfaces présentant une réactivité chimique pour différents couplages covalents (Figure 25).

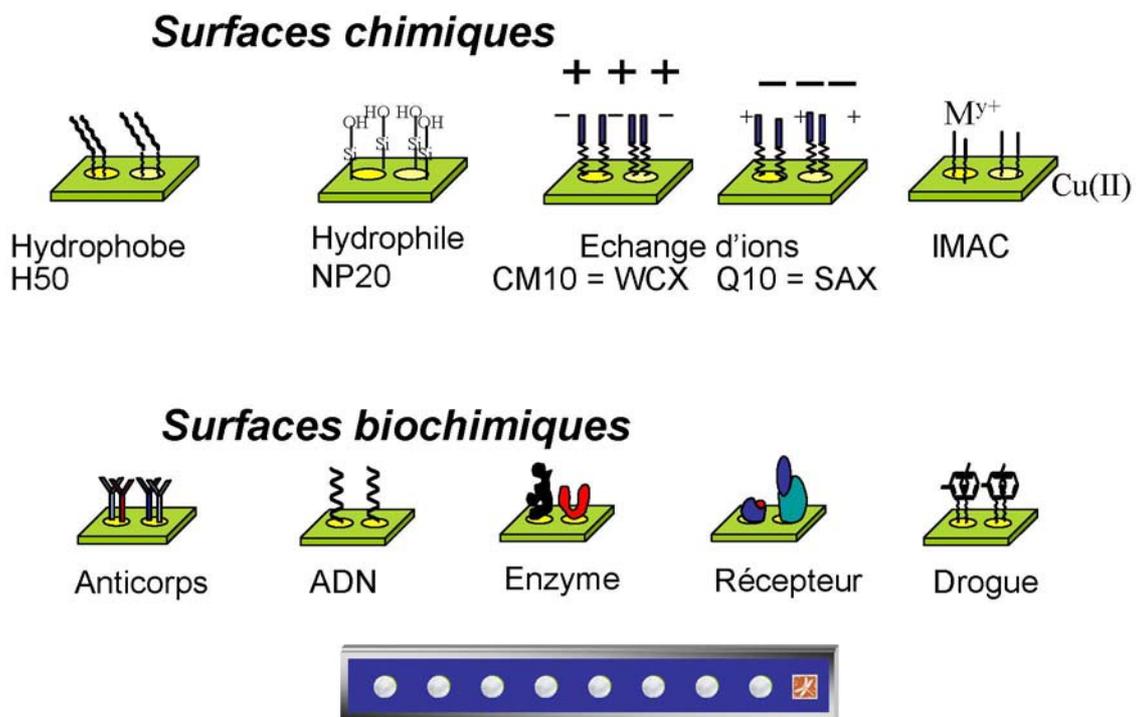


Figure 25 : représentation des différentes surfaces chromatographiques disponibles.

Mise en œuvre expérimentale et méthodologie. De la préparation de l'échantillon à l'analyse par spectrométrie de masse, le cas des protéines membranaires

1) Les méthodes de séparations des protéines. Le problème des protéines membranaires

1.1) Le gel 2D. Historique et possibilités

L'outil de séparation des protéines le plus couramment utilisé en analyse protéomique est le gel d'électrophorèse bidimensionnel, plus communément appelé gel 2D. La première dimension permet de séparer les protéines en fonction de leur point isoélectrique, et la deuxième dimension est un gel SDS-PAGE classique qui va séparer les protéines en fonction de leurs masses moléculaires [Kenrick et Margolis 1970]. Cette technique est donc appropriée pour l'étude des protéines comprises entre 10 et 300 kDa et avec un point isoélectrique compris entre 4 et 10 (figure 26).

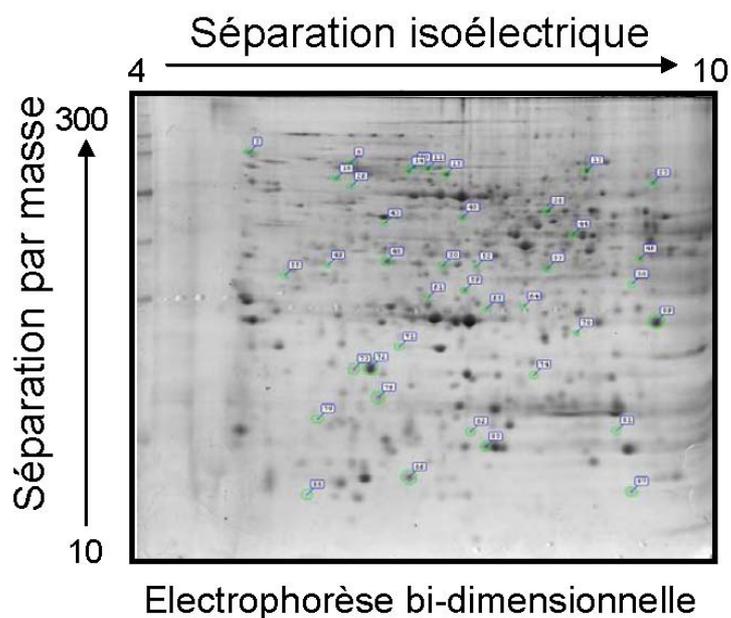


Figure 26 : Exemple de gel bidimensionnel. La première dimension permettant de séparer les protéines en fonction de leur point isoélectrique et la deuxième dimension permettant une séparation en fonction de la masse.

Bien que le gel 2D ait fait ses apparitions dans le début des années 70 [Kenrick et Margolis 1970], il n'en reste pas moins irremplaçable. En effet, dès les années 75, O'Farrell reporte la détection de 1000 spots distincts sur un gel 2D [O'Farrell 1975], et depuis ces années, un grand nombre de développements ont été entrepris afin de continuer à augmenter les performances des gels 2D [Tonella et al. 2001].

1.1.1) La première dimension

La première étape de séparation des protéines aussi appelée première dimension correspond à une isoélectrofocalisation des protéines en fonction de leur point isoélectrique. 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 donc elles ne migrent plus. Cette technique a été très largement améliorée afin d'en augmenter la résolution et la sensibilité. De nombreux travaux ont permis de développer entre autre les techniques de solubilisation des protéines ou les gradients de pH restreints, c'est-à-dire de réaliser des zooms de pH afin d'augmenter la résolution et le nombre de protéines détectées sur les gels 2D.

1.1.2) La préparation de l'échantillon

Ce que l'on entend par préparation de l'échantillon représente surtout la solubilisation des différentes protéines avant de les déposer sur le strip de la première dimension. La difficulté de cette étape réside dans l'importance de solubiliser toutes les protéines sans en modifier les propriétés physico-chimiques, ce qui la rend incompatible avec les différents détergents ioniques (forts) comme le SDS par exemple. Des études ont toutefois montré que l'utilisation de détergents doux comme le CHAPS ou le Triton était possible et pouvait donc aider à la solubilisation des différents échantillons protéiques [Rabilloud et al. 1997]. De la qualité de cette préparation va dépendre le résultat de l'électrophorèse. On peut noter, en plus de la difficulté de solubilisation des échantillons, l'importance de la dénaturation des protéines en réduisant par exemple les ponts disulfures et en alkylant les résidus cystéines libres, ou encore la phase de réhydratation du strip avec le mélange protéique.

1.1.3) Les colorations

La coloration est l'étape qui permet de révéler les différentes protéines présentes sur le gel. Le choix du colorant va déterminer la sensibilité de la technique. Les colorants les plus communément utilisés et compatibles avec l'analyse par spectrométrie de masse sont :

- Le bleu de Coomassie : bien que peu sensible il présente le grand avantage d'être considéré comme linéaire et donc d'être parfaitement adapté pour toutes les études comparatives.
- La coloration à l'argent : beaucoup plus sensible que la coloration au bleu, elle dépend des propriétés intrinsèques de la protéine et ne peut donc pas être utilisée pour comparer la quantité de protéines différentes les unes par rapport aux autres.
- La coloration par fluorescence (coloration Sypro® ruby par exemple) : est à l'heure actuelle la technique la plus sensible mais elle nécessite un éclairage UV afin de visualiser les différents spots.

D'autres types de colorations sont développées en parallèle afin d'être plus sensibles ou plus compatibles avec la spectrométrie de masse comme la coloration au ruthénium ou des colorations à base d'argent mais qui ne présentent pas d'agent fixant incompatible avec la spectrométrie de masse comme le formaldéhyde [Richert et al. 2004], ou encore spécifiques des protéines membranaires [Bertsch et Kassner 2003].

1.1.4) Le gel 2D à haut débit et la standardisation

Bien que les gels 2D demandent un grand savoir faire, le développement de nouveaux protocoles ou de gels pré-coulés par exemple, rendent cette technique reproductible en fonction du

type d'échantillon 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.

Les éléments qui freinent cette standardisation viennent surtout de la préparation des différents échantillons et de leur complexité. En effet, un mélange protéique trop complexe ou qui présente des protéines très majoritaires entrainera une mauvaise séparation des protéines et donc une résolution médiocre et les protéines de faible abondance ne pourront pas être analysées par cette approche [Gygi et al. 2000].

Enfin, cette technique n'est pas appropriée pour l'étude d'un certain nombre de protéines comme les protéines hydrophobes ou très basiques [Bunai et Yamane 2005] [Galeva et Altermann 2002] [Santoni et al. 2000]. En effet, étant donné que les détergents ioniques ne sont pas compatibles avec la première dimension des gels, il est presque impossible de solubiliser correctement les différents mélanges protéiques [Adessi et al. 1997] [Santoni et al. 1999] même si des efforts sont développés dans ce sens [Luche et al. 2003].

1.1.5) La méthode d'électrophorèse Off-Gel™ (OGE)

L'OGE est une technique de fractionnement isoélectrique qui permet la purification d'une ou plusieurs protéines dans une solution placée au-dessus d'un gel contenant un gradient de pH immobilisé (IPG) (figure 27).

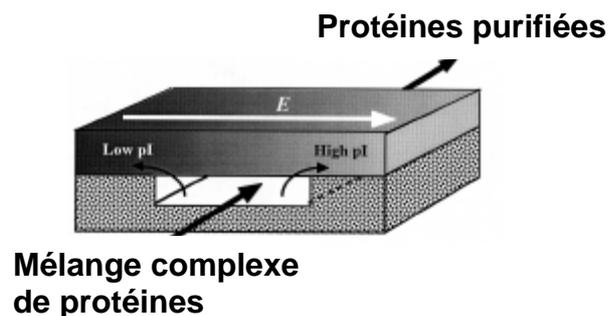


Figure 27 : Représentation d'un système off-gel.

Cette méthode permet l'extraction de protéines dans une certaine gamme de pI (restreinte) à partir d'échantillons complexes, tout en les gardant en solution, sans aucun besoin d'ampholytes ou autre tampon.

Le système se compose d'une chambre fine dont un des côtés est constitué d'un gel à gradient de pH immobilisé. Le mélange complexe de protéines est introduit dans la chambre et une recirculation est

mise en place grâce à une minipompe péristaltique. L'application d'un champ électrique (E) de façon tangentielle au gel (réduit les risques de précipitation des protéines) entraîne toutes les espèces chargées, qui vont migrer dans le gel (d'où la séparation qui dure jusqu'à une heure).

Le champ électrique sépare donc les protéines comme si elles étaient dans le gel lui-même. Les protéines chargées positivement migrent vers la cathode et pénètrent le gel en suivant les lignes de courant alors que les protéines chargées négativement migrent vers l'anode.

Seules les espèces neutres, dont celles ayant un pI proche du pH du gel en contact restent en solution dans la chambre. Ce système, adapté au traitement d'échantillons biologiques complexes, permet de traiter jusqu'à plusieurs mg de protéines. Il peut être aisément combiné avec d'autres méthodes de séparation : l'E2D, la LC/MS, le microarray, la cristallisation,

De plus, il est possible de faire plusieurs OGE de suite pour affiner la séparation en resserrant la gamme de pI. De bons résultats sont obtenus sans déplétion des protéines majoritaires au préalable [Arnaud et al. 2002] [Ros et al. 2002].

1.2) Le couplage chromatographie liquide spectrométrie de masse

Face à la complexité croissante des échantillons, la chromatographie liquide à haute pression se trouve être une technique de choix pour la simplification des différents mélanges protéiques et peptidiques. On distingue deux types de chromatographies liquides associées à la spectrométrie de masse : la chromatographie off-line qui demande la collecte des différentes fractions éluées pour être analysées ultérieurement et la chromatographie dite on-line qui elle est directement couplée à un spectromètre de masse [Hunt et al. 1992].

1.2.1) La chromatographie liquide off-line

Cette technique consiste à réaliser une chromatographie liquide et de collecter les différentes fractions d'élutions. Ces fractions pourront être par la suite analysées par spectrométrie de masse MALDI ou par infusion électrospray. Ces séparations permettent de purifier de grandes quantités de protéines ou de peptides qui pourront servir à d'autres expériences comme l'insertion de modifications chimiques utilisées lors de l'approche COFRADIC [Gevaert et al. 2003] ou des analyses de séquence comme la dégradation d'Edman par exemple, ou afin de réaliser des analyses par MALDI qui ne peut pas être directement couplé à la chromatographie liquide.

De plus les constructeurs ont développé ces dernières années des systèmes de collecte automatique des différentes fractions d'élution d'HPLC. Il existe aussi des systèmes capables de déposer directement une petite quantité de la fraction collectée sur des plaques MALDI afin d'en réaliser l'étude par spectrométrie de masse (figure 28).

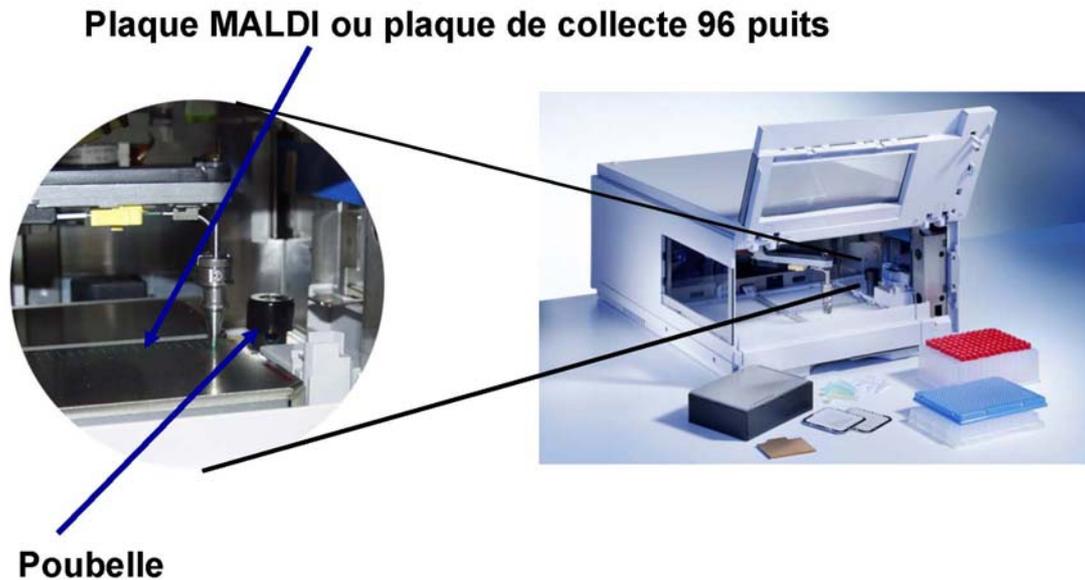


Figure 28 : photographie du collecteur de fraction Agilent.

1.2.2) La chromatographie liquide on-line

A ses débuts, l'HPLC était couplée avec des détecteurs UV-visible et n'était pas compatible avec l'analyse par spectrométrie de masse. En effet, les molécules en solution devaient être vaporisées et ionisées pour être détectées. Il aura donc fallu attendre l'apparition de l'électrospray dans les années 80 pour avoir une méthode d'ionisation véritablement compatible avec la spectrométrie de masse et l'étude des biomolécules. De plus, la diminution des diamètres des différentes colonnes et donc la diminution des débits a permis la mise en place de couplages avec la spectrométrie de masse de plus en plus efficaces, ce qui permet de travailler aujourd'hui avec des débits de l'ordre de quelques centaines de nanolitres par minute et d'avoir des gammes de détection de l'ordre de la femtomole injectée.

Ce couplage entre la chromatographie liquide et les spectromètres de masses permet aujourd'hui de réaliser entre autre ce que l'on appelle la LC-MS/MS en mode automatique. Les différents ions élués au cours de l'analyse chromatographique sont sélectionnés, fragmentés et analysés de façon systématique et automatique permettant d'avoir des informations de séquence indispensables pour l'analyse protéomique.

Cette approche de LC-MS ou de LC-MS/MS est toutefois insuffisante lors de l'étude de mélanges très complexes et qui nécessitent des étapes de simplifications supplémentaires.

1.3) Le gel 1D une alternative aux gels 2D, application aux protéines membranaires

Comme nous l'avons déjà mentionné le gel 2D est très peu permissif quand à la préparation et la solubilisation de l'échantillon. Il est donc très difficile de réaliser des séparations de préparations protéiques membranaires sur gel 2D [Galeva et Altermann 2002]. En effet, vu que seule l'utilisation de détergents doux est possible, les protéines membranaires ne peuvent pas être solubilisées et donc être correctement déposées sur gels 2D. La solution pour l'étude des protéines membranaires est de remplacer l'analyse par gel 2D et MALDI-MS par une électrophorèse monodimensionnelle, beaucoup moins résolutive (quelques dizaines voir centaines de bandes) mais compatible avec l'étude de protéines hydrophobes. Les différentes bandes de ce gel sont ensuite découpées et digérées pour être analysées par une approche nanoLC-MS/MS. Cette approche va donc séparer dans un premier temps les protéines en fonction de leur masse (gel SDS) et dans un deuxième temps selon l'hydrophobicité des peptides issus de la digestion. Bien que cette approche soit extrêmement lourde à mettre en place elle est actuellement l'une des mieux adaptées pour l'étude des protéines membranaires.

2) Simplification des échantillons

Un des points cruciaux de la protéomique est la préparation de l'échantillon. En effet, les résultats obtenus par spectrométrie de masse vont dépendre de cet échantillon, et plus l'échantillon à analyser est complexe, plus il va être difficile à analyser. Un des effets les plus gênants sera la suppression de signal engendrée par les protéines ou les peptides majoritaires. Il est donc nécessaire de travailler sur des sous-protéomes pour limiter le nombre de protéines présentes dans l'échantillon ou d'augmenter le nombre d'étapes de simplification des différents mélanges.

2.1) Augmenter le nombre d'étapes de séparation

Le premier point de simplification des échantillons dans l'étude protéomique est d'augmenter le nombre d'étapes de séparation des protéines et des peptides. La première façon d'augmenter les étapes de séparation est de coupler le gel d'électrophorèse 2D avec une chromatographie liquide couplée à la spectrométrie de masse. Cette approche permet l'identification de protéines et ceci même si le génome de l'organisme étudié n'est pas séquencé [Carapito et al. 2005].

D'autres approches pour augmenter le nombre de protéines identifiées (ce qui correspond à simplifier le contenu d'un échantillon pour limiter les effets de suppression de signal) ont été développées dont l'une des plus importante est celle du groupe de J. R. Yates qui se base sur des chromatographies multidimensionnelles et appelée Mudpit [Washburn et al. 2001]. Le mélange protéique est digéré et le mélange peptidique est alors injecté sur une colonne échangeuse de cation suivie par une colonne de phase inverse. Les différents peptides issus de la digestion seront donc séparés dans un premier temps

par leur charge avant d'être séparés en fonction de leur hydrophobicité. Cette approche permet d'identifier 4 fois plus de protéines que lors d'approches plus classiques (Figure 29).

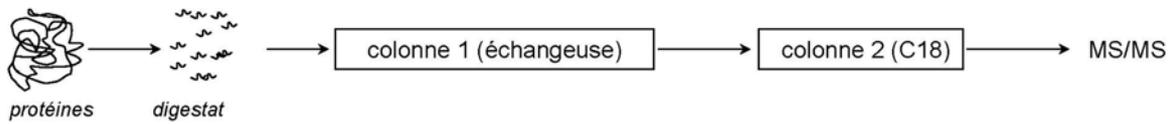


Figure 29 : Schéma de l'approche LC-LC-MS/MS.

2.2) Limiter le nombre de protéines : l'étude de sous-protéomes

Le deuxième point qui peut faciliter l'analyse protéomique est le fait d'étudier les différents sous-protéomes de la cellule. On va entendre par sous-protéome les différents organites de la cellule comme la mitochondrie ou le noyau [Brugiere et al. 2004] [Dreger 2003]. De plus on peut se limiter aux différents partenaires protéiques d'une protéine cible. Le principe de cette approche est de créer in-vivo le complexe protéique et de le purifier par des chromatographies d'affinité (immunoprécipitation, Tap-Tag...). Ce complexe protéique sera alors déposé sur un gel monodimensionnel et analysé par spectrométrie de masse, ce qui permettra de connaître spécifiquement les différentes protéines qui interagissent avec la protéine d'intérêt [Rigaut et al. 2005].

2.3) Un sous-protéome particulier : la membrane plasmique

Bien qu'elles ne représentent qu'un tiers des protéines codées par le génome [Wallin et von Heijne 1998], les protéines membranaires ont une importance cruciale en biologie étant donné qu'elles se situent à l'interface entre deux compartiments cellulaires (cytoplasme/mitochondrie ; cytoplasme/noyau ; cellule/milieu extérieur...). Plus spécifiquement, les protéines de la membrane plasmique ont un rôle primordial tant du point de vue biologique (transmission des signaux), que d'un point de vue pharmacologique (marqueurs de pathologie faciles d'accès, cibles thérapeutiques...) [Jang et Hanash 2003].

2.3.1) Les différents types de préparations de membranes plasmiques

Il existe un grand nombre d'approches différentes pour réaliser des préparations membranaires. Les plus fréquemment rencontrées sont les préparations par centrifugations différentielles, centrifugation sur gradient de sucrose [Klein et al. 2005] ou encore la biotinylation des molécules de surface [Zhao et al. 2004] [Peirce et al. 2004]. Ces différentes techniques sont difficiles à

mettre en place et bien qu'enrichies en protéines membranaires ces préparations sont souvent contaminées par des protéines membranaires issues des autres organites cellulaires.

2.3.2) Les microparticules

Lors de conditions de stress telle que l'activation de la division cellulaire ou l'apoptose, la répartition en acides gras de la membrane entre le feuillet externe et le feuillet interne est perturbée, ce qui entraîne l'émission spontanée de vésicules qui sont appelées microparticules (Figure 30). Ces microparticules ont des tailles variables qui peuvent aller jusqu'à 1µm de diamètre. Elles sont donc facilement différenciables des autres vésicules membranaires comme les exosomes par exemple.

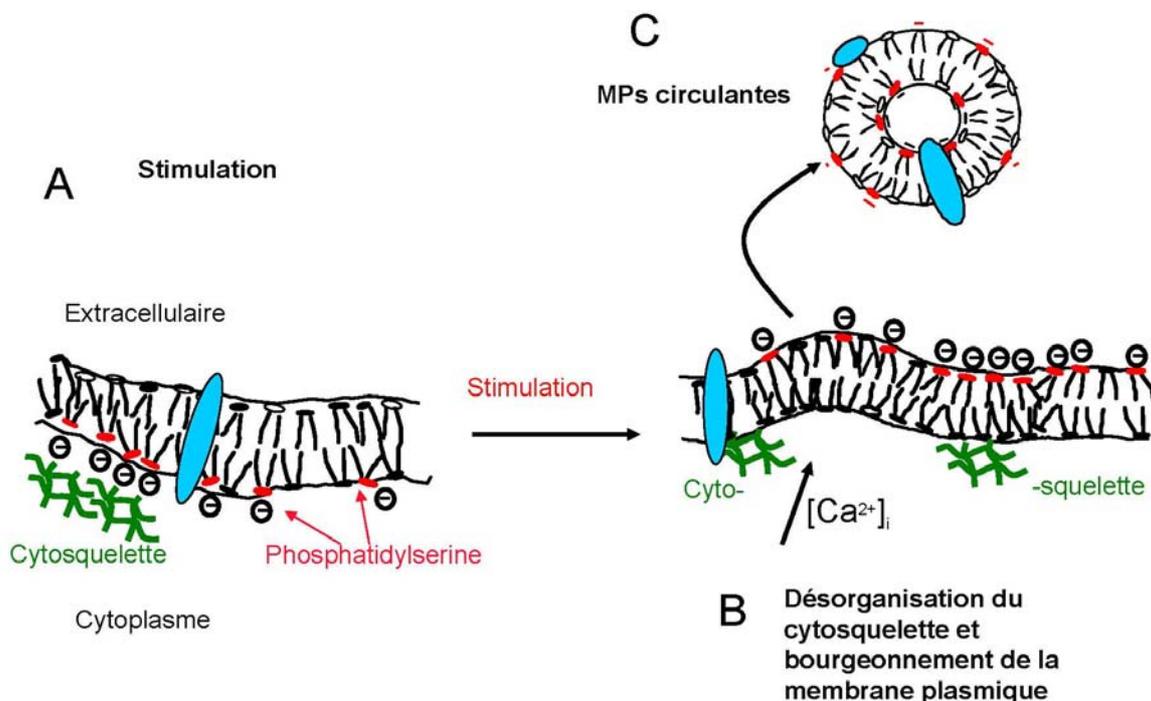


Figure 30 : Schématisation de la production de microparticules.

Au vu de leur taille les microparticules sont facilement séparées du surnageant par des centrifugations lentes (15000g pendant 30 minutes) [Freyssinet 2003]. Etant donné que ces structures proviennent directement de la membrane plasmique, elles doivent être naturellement enrichies en protéines issues de la membrane plasmique. Au début de ce travail de thèse, il n'y avait eu aucune analyse protéomique basée sur l'étude des microparticules. Cette étude est l'une des thématiques de ce travail de thèse qui est discuté plus en détail dans la partie II, chapitre II de ce manuscrit.

3) Digestion des protéines et extraction des peptides

3.1) Digestion des protéines

La digestion in gel est basée sur la procédure de Rosenfeld [Rosenfeld et al. 1992]. Les différentes étapes pour éliminer les différents sels et le SDS du gel qui peut interférer avec la digestion enzymatique sont dans un premier temps : deux lavages successifs avec une solution d'hydrogénocarbonate d'ammonium et d'acétonitrile, suivi par les étapes de réduction alkylation des ponts disulfures des protéines. Les morceaux de gel sont alors déshydratés par l'ajout d'acétonitrile à température ambiante ce qui entraîne une diminution importante de leur volume. Lors de l'ajout du tampon contenant l'enzyme, les morceaux de gel vont réagir comme des éponges en aspirant le liquide, ce qui va augmenter l'efficacité de la digestion.

Différentes petites modifications de protocoles ont été apportées à cette étape de digestion surtout en ce qui concerne les protéines membranaires à savoir l'utilisation de concentrations plus importantes de solvants organiques (acétonitrile, isopropanole, ...) ou l'ajout de détergents (CHAPS, SDS, ...) [Lu et Zhu 2005]. Il existe aussi des développements sur l'utilisation de la combinaison de différentes enzymes protéolytiques, mais là encore, même si les résultats semblent favorables, aucun protocole standardisé n'est pour l'instant proposé [Fischer et Poetsch 2006].

3.2) Extraction des peptides

L'approche classique pour l'extraction des différents peptides générés par la digestion se fait par l'ajout d'une solution d'eau/acétonitrile avec quelques pourcents d'acide formique. Les peptides vont alors diffuser de façon passive entre les morceaux de gel et le tampon. L'acétonitrile en excès sera évaporé (soit à l'air libre soit au speedvac) avant d'être analysé par spectrométrie de masse MALDI ou ESI.

Là encore, il n'existe pas de protocole standardisé pour l'étude des protéines membranaires (utiliser plus de solvant organique ou ajouter des détergents par exemple) [Richert 2003].

4) Analyse protéomique par spectrométrie de masse

4.1) Présentation de l'approche protéomique par spectrométrie de masse

L'identification des protéines est une technique bien maîtrisée par notre laboratoire et ceci depuis plusieurs années [Carte 2001]. Une illustration de la stratégie protéomique par spectrométrie de masse est représenté en figure 31.

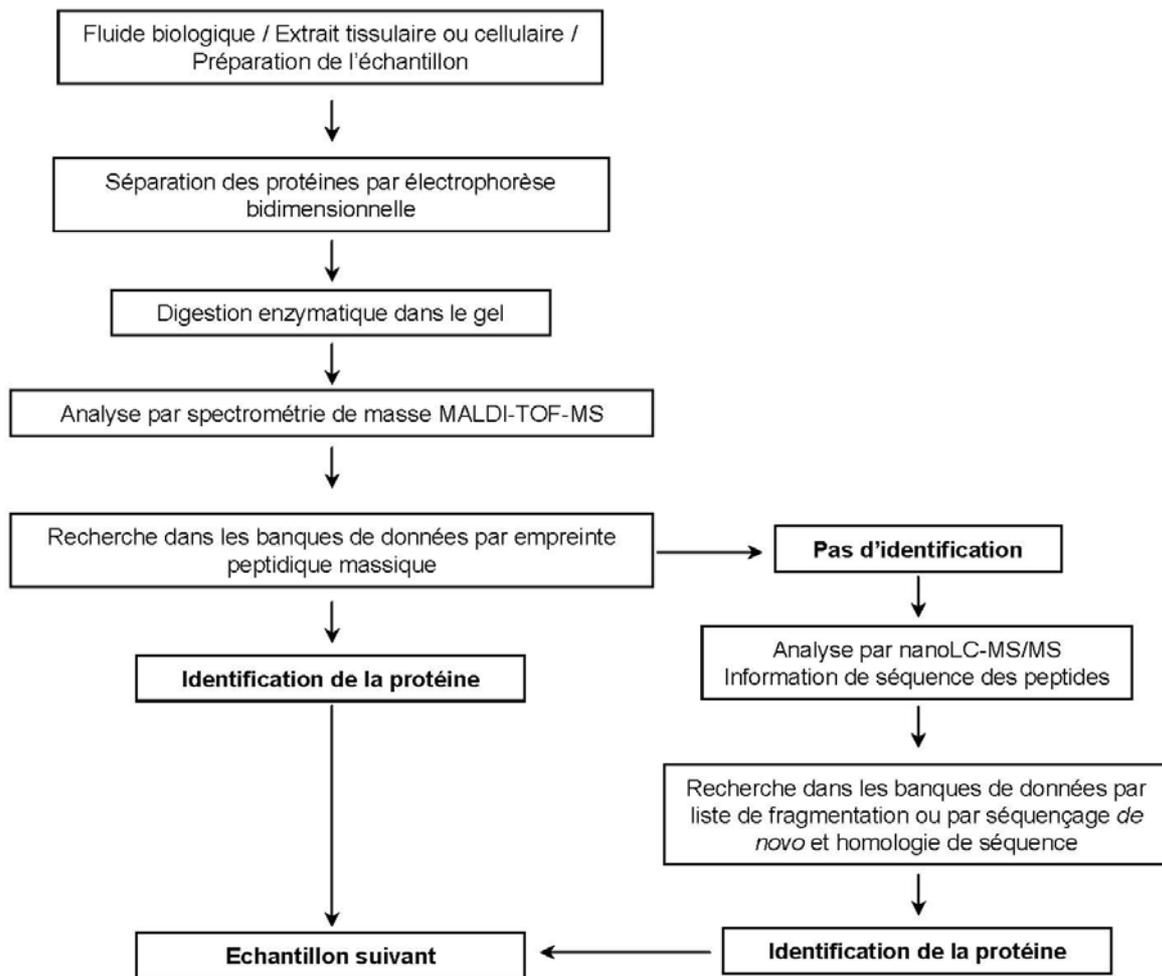


Figure 31 : Illustration schématique de l'approche protéomique par spectrométrie de masse.

4.2) L'empreinte peptidique massique (mass fingerprint) / approche MALDI-MS

Les protéines séparées en règle générale par électrophorèse 2D sont digérées et analysées par MALDI-TOF-MS. Cette analyse génère une liste de masses expérimentales des différents peptides générés par la protéolyse de l'échantillon qui vont pouvoir être comparées aux masses théoriques des peptides issues de digestions *in silico*. L'identification des protéines va se faire en fonction du nombre de masses peptidiques qui correspondent à la protéine, le pourcentage de recouvrement de la séquence protéique et enfin la précision de mesure des différentes masses peptidiques.

Cette technique qui est facile à mettre en œuvre et qui permet d'obtenir rapidement des résultats présente toutefois des limites quand à l'identification des protéines. En effet, les banques protéiques devenant de plus en plus complexes, les identifications par empreinte peptidique massique sont de plus en plus difficiles à valider. Les critères de validité sont : le pourcentage de recouvrement de la protéine (ou le nombre de peptides identifiés) et la précision de la mesure de masse [Carr et al. 2004]

[Bradshaw 2005]. On peut admettre qu'une protéine est identifiée par PMF lorsque le pourcentage de recouvrement dépasse 30% (ou 7 peptides) avec une précision en masse meilleure que 50 ppm. De plus, si l'organisme étudié n'est pas ou peu séquencé, les différentes protéines ne seront pas disponibles dans les banques de données, et il faudra avoir recours à la spectrométrie de masse en tandem afin de générer des fragmentations peptidiques afin d'obtenir des informations de séquence.

4.3) Information de séquence, approche MS/MS

Cette approche va permettre de sélectionner un ion appelé alors ion précurseur, de le fragmenter, et ainsi de pouvoir obtenir des informations de séquence (MALDI-TOF-TOF, ESI-Q-TOF, ESI-Trap).

4.3.1) La fragmentation peptidique

La fragmentation des peptides se fait au niveau de la liaison peptidique. A partir de cette constatation une nomenclature a été introduite en 1984 [Roepstorff et Fohlman 1984] puis complétée et simplifiée en 90 [Biemann 1990]. Cette nomenclature est représentée en figure 32.

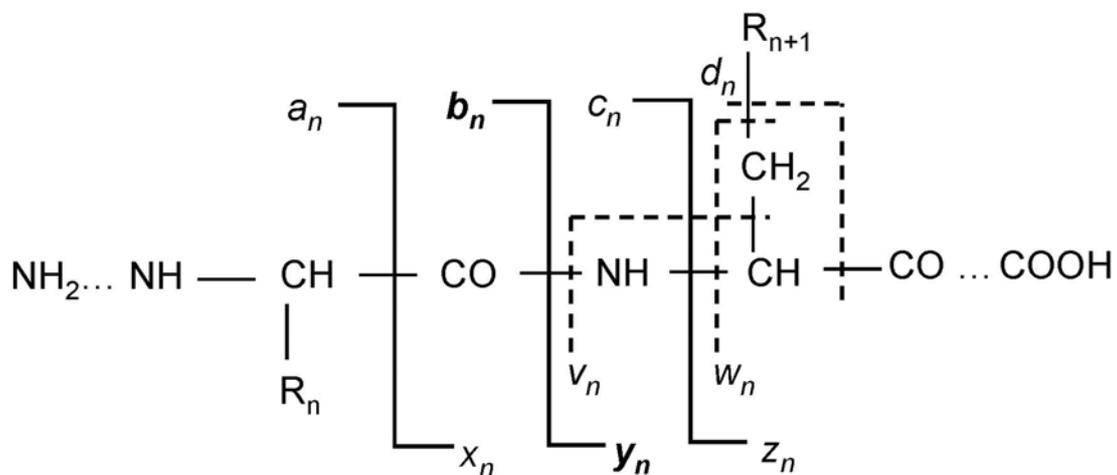


Figure 32 : Nomenclature de la fragmentation peptidique en spectrométrie de masse.

Les ions parents sont sélectionnés soit par le quadripôle dans le cas de l'utilisation d'un Q-TOF, soit par l'isolement de l'ion parent dans le cas de l'utilisation d'une trappe à ions. Ces ions seront alors fragmentés et les différents fragments générés seront alors analysés. La liste de masse des fragments pourra alors être, tout comme la liste de masse MALDI, soumise à interrogation dans les banques de données et donc comparée à une liste de masses de fragmentations théoriques. Cette recherche va permettre d'identifier la protéine dont le peptide a été sélectionné et fragmenté.

4.3.2) Le séquençage *De Novo*

Le séquençage *De Novo* consiste à reconstituer la séquence peptidique d'un ion fragmenté à partir de son spectre MS/MS. A partir de chaque spectre MS/MS une séquence en acide aminés sera déduite, partielle ou complète, et qui pourra servir à des recherches dans les banques de données par homologie de séquence et non plus par homologie de masse. Cette approche est indispensable lorsque l'on étudie des protéines issues d'organismes non séquencés et que les différentes protéines ne se trouvent pas dans les différentes banques protéiques.

4.4) La quantification des protéines

La technique de quantification la plus classique consiste à déposer les mélanges protéiques des différents états sur des gels bidimensionnels distincts, de les colorer avec des colorants compatibles avec la quantification, et de comparer les intensités des différents spots présents respectivement sur chacun des gels. Les spots contenant les protéines exprimées de façon différentielle sur les gels sont alors quantifiés, sélectionnés et analysés par spectrométrie de masse. On peut alors réaliser une carte présentant le changement d'expression des protéines en fonctions des différentes conditions de stimulations extérieures.

Différentes techniques se basant sur le gel 2D ont depuis été développées comme l'approche baptisée DIGE pour « Difference Gel Electroporesis » [Unlu et al. 1997]. Cette technique se propose de marquer les échantillons par des colorants fluorescents spécifiques avant d'être déposés sur un gel 2D. Lors de la révélation, deux images vont être générées correspondant à chacun des échantillons. Ces deux images pourront alors être comparées l'une par rapport à l'autre.

Une autre approche très couramment utilisée est celle développée par le groupe d'Aebersold. Cette technique repose sur l'utilisation de marqueurs chimiques qui diffèrent par l'incorporation d'isotope stables (en règle générale l'hydrogène est remplacé par des atomes de Deutérium, et les marqueurs les plus récents utilisent le carbone 12 et le carbone 13) [Gygi et al. 1999]. Les échantillons protéiques à comparer seront donc marqués par des réactifs qui ont des masses différentes. L'analyse par spectrométrie de masse pourra alors mettre en évidence la différence de masse entre les peptides venant des échantillons différents. Les intensités de ces peptides pourront être comparées afin de réaliser la quantification relative entre les différents échantillons (Figure 33). Malheureusement, cette approche qui présente une étape de chromatographie d'affinité entraîne la perte de tous les peptides qui ne sont pas marqués. Les protéines identifiées au cours de cette approche le seront donc avec très peu de peptides, et cette stratégie ne sera donc pas applicable dans le cas où le génome de l'organisme étudié est en cours ou n'est pas séquencé.

Une variante de cette technique de quantification a été développée par le groupe de M. Mann et consiste à incorporer le marqueur isotopique lors de la croissance des cellules en utilisant un acide

aminé marqué (en règle générale une leucine marquée CH3 et une leucine marquée CD3) [Ong et al. 2002], ou encore plus récemment l'approche ISIL qui combine l'étude différentielle sur gel d'électrophorèse avec un marqueur isotopique [Asara et al. 2006]. Pour revue voir aussi [Lill 2003] [Wu et al. 2006].

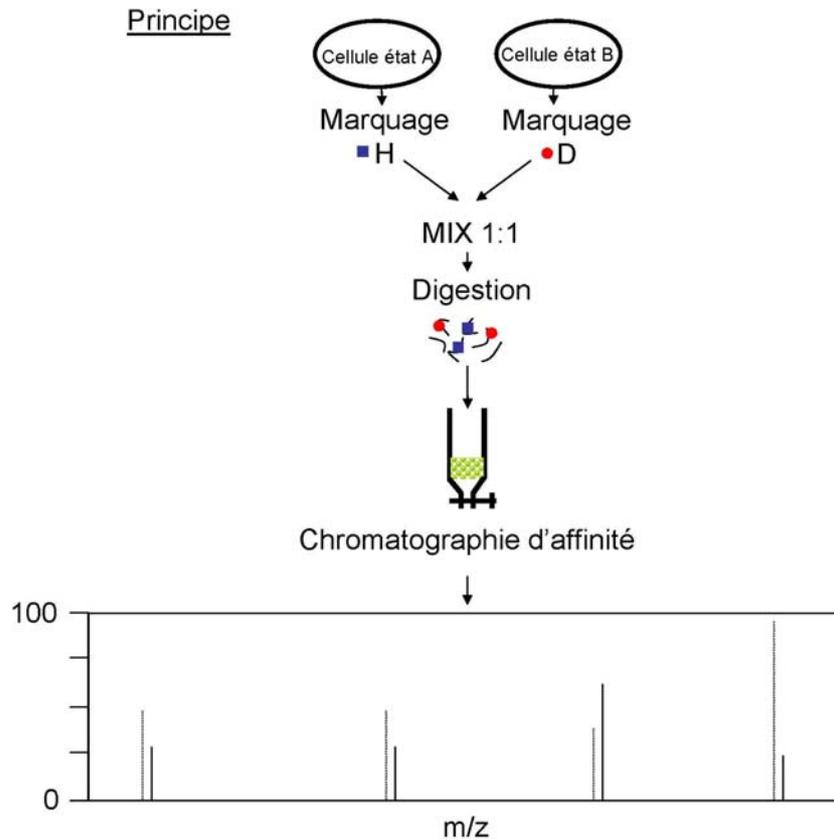


Figure 33 : représentation schématique du principe de quantification par l'approche ICAT.

Ces différentes techniques de quantification sont efficaces sur des échantillons concentrés et riches en protéines, mais ne sont encore que très peu applicables à des mélanges protéiques membranaires par exemple.

5) Traitement informatique et recherche dans les bases de données

De plus en plus de génomes sont séquencés, de plus en plus de protéines sont étudiées. Les banques protéiques et génomiques deviennent de plus en plus importantes. Pour exemple la banque Swiss Prot qui est l'une des banques les plus sélectives est passée de 115106 entrées protéiques en octobre 2002 à 212425 entrées en mars 2006. Cette augmentation additionnée à l'analyse systématique et à haut débit des différents échantillons par spectrométrie de masse représente aussi un challenge informatique.

5.1) recherche dans les banques de données.

Le premier point à souligner est la recherche des différentes données générées par spectrométrie de masse dans les différentes banques de données protéiques. Comme nous l'avons mentionné les banques protéiques deviennent de plus en plus complexes (NCBI : 3 479 934 séquences), ce qui nécessite l'utilisation d'ordinateurs de plus en plus performants. Par exemple nous disposons au laboratoire d'une station de travail biprocesseur xeon uniquement dédiée à la recherche dans les banques de données.

Les recherches dans ces banques de données se font par l'intermédiaire de quelques algorithmes commerciaux ou libres d'accès sur internet comme MASCOT développé par Matrix Science (<http://www.matrixscience.com>), Protein Prospector (<http://prospector.ucsf.edu/>) ou encore des sites comme ExPasy qui regroupent différents moteurs de recherche comme Aldente ou Phenyx. Ces différents algorithmes permettent l'identification de protéines qui sera basée sur un score qui dépendra entre autre du nombre de peptides ou de fragments MS/MS qui sont attribuées à la protéine, du nombre d'entrées de la banque. Ces différents logiciels sont aussi applicables pour rechercher directement dans les différents génomes, ce qui permet de s'affranchir des problèmes d'annotation ou de protéines issues de familles multigéniques [Delalande et al. 2005].

Enfin il faut aussi mentionner les logiciels de recherche par alignement de séquence de BLAST ou FASTA (<http://www.ncbi.nlm.nih.gov/blast/>). Ces logiciels vont permettre la comparaison de séquences peptidiques par rapport aux différentes séquences des banques. Les différents alignements de séquence ainsi réalisés permettront l'identification des protéines par homologie de séquence, même si celles-ci présentent des différences entre les différentes espèces.

5.2) Les critères d'identifications

L'identification de protéines issues d'un mélange complexe séparées sur gel 1D se fait très souvent avec peu de peptides. Il est donc nécessaire d'établir des critères d'identifications stricts. Ces critères sont à l'heure actuelle assez controversés. Pour cela, les différents journaux de protéomique ont décidé de définir des critères d'identifications [Bradshaw 2005; Wilkins et al. 2006]. D'une façon générale, les différents peptides qui servent à l'identification des protéines sont inspectés manuellement. Dans les cas litigieux, c'est-à-dire peu de peptides ou des MS/MS de mauvaise qualité, les spectres sont interprétés manuellement afin d'en déterminer la séquence peptidique. Différents logiciels ont été développés dans le but de faciliter cette étape telle que PEAKS studioTM (BSI) ou PepSeq (MassLynxTM, Waters). Ne seront considérées comme identifiées que les protéines qui présentent au moins un peptide dont le spectre de fragmentation permet de déterminer sa séquence partielle ou totale. La séquence déterminée est alors soumise à une recherche par alignement de séquence afin de déterminer la spécificité du peptide vis-à-vis de la protéine. On considérera donc une

protéine comme identifiée si lors de l'analyse nanoLC-MS/MS au moins un peptide a été séquencé et que le spectre MS/MS permet de reconstituer la séquence ou une partie de la séquence de ce peptide, et qu'il est spécifique de la protéine.

5.3) Archivages et bases de données

Même si une analyse MALDI ne représente que quelques centaines de Ko, une analyse LC-MS/MS représente plus de 100 Mo. En plus des différents espaces disques alloués pour la sauvegarde de ces différentes données, il est important de développer des bases de données qui regroupent ces différentes informations. Différents logiciels ont donc été développés pour stocker les différentes données générées par spectrométrie de masse. On peut citer ProteinScape développé par Bruker ou SQL-LIMS développé par Applied Biosystem. Des solutions alternatives peuvent aussi être envisagées, ainsi au laboratoire, nous développons une base de données qui regroupe tous les projets et qui rattache à ces projets, tous les échantillons et toutes les analyses correspondantes. De tels systèmes informatiques sont très importants à mettre en place afin de centraliser les données pour des utilisations ultérieures comme lancer de nouvelles recherches dans de nouvelles banques ou dans des banques complétées.

6) Conclusions de cette étude bibliographique et perspectives d'études.

Les données actuelles montrent que l'analyse protéomique permet l'identification, la quantification et la caractérisation d'un grand nombre de protéines, et ceci que ce soit par des analyses MALDI-MS ou nanoLC-MS/MS. Cependant, il subsiste des difficultés pour l'étude de certaines familles de protéines, dont la plus importante est la grande famille des **protéines membranaires**. En effet, ces protéines présentent, de par leur caractère hydrophobe, une forte incompatibilité vis-à-vis de l'analyse protéomique.

Comme nous l'avons déjà mentionné, les préparations de protéines membranaires sont longues et laborieuses, et souvent contaminées par un grand nombre de protéines cytosoliques. De plus, ces protéines sont difficilement séparables sur des gels d'électrophorèse 2D. Enfin, les protéines issues de la membrane plasmique, en plus de leur rôle biologique sont très importantes du point de vue médical, étant donné qu'elles représentent **des cibles diagnostiques et thérapeutiques** très importantes.

Durant ce travail de thèse, nous avons :

- Développé de nouvelles méthodologies de **quantification** pour les mélanges protéiques complexes
- Validé de **nouveaux outils biologiques** pour l'étude des protéines membranaires

- Mis au point une nouvelle stratégie combinant les informations MALDI et ESI pour **augmenter le pourcentage de recouvrement des protéines membranaires**
- Appliqué ces stratégies pour la **recherche de nouveaux biomarqueurs** membranaires dans le cadre des hémopathies lymphoïdes chroniques.

Références bibliographiques

A

- Adessi, C., Miede, C., Albrieux, C. and Rabilloud, T. (1997). "Two-dimensional electrophoresis of membrane proteins: a current challenge for immobilized pH gradients." *Electrophoresis* **18**(1): 127-35.
- Aebersold, R. and Mann, M. (2003). "Mass spectrometry-based proteomics." *Nature* **422**(6928): 198-207.
- Arnaud, I. L., Josserand, J., Rossier, J. S. and Girault, H. H. (2002). "Finite element simulation of Off-Gel trade mark buffering." *Electrophoresis* **23**(19): 3253-61.
- Asara, J. M., Zhang, X., Zheng, B., Christofk, H. H., Wu, N. and Cantley, L. C. (2006). "In-Gel Stable-Isotope Labeling (ISIL): a strategy for mass spectrometry-based relative quantification." *J Proteome Res* **5**(1): 155-63.

B

- Bertsch, M. and Kassner, R. J. (2003). "Selective staining of proteins with hydrophobic surface sites on a native electrophoretic gel." *J Proteome Res* **2**(5): 469-75.
- Biemann, K. (1990). "Appendix 5. Nomenclature for peptide fragment ions (positive ions)." *Methods Enzymol* **193**: 886-7.
- Biemann, K. (1990). "Peptides and proteins: overview and strategy." *Methods Enzymol* **193**: 351-60.
- 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.
- Blades, A., Michael G. Ikonou, and Paul Kebarle (1991). "Mechanism of electrospray mass spectrometry. Electrospray as an electrolysis cell." *Anal Chem* **63**: 2109-2114.
- Bradshaw, R. A. (2005). "Revised draft guidelines for proteomic data publication." *Mol Cell Proteomics* **4**(9): 1223-5.

Brugiere, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., d'Herin, P., Garin, J., Bourguignon, J., Joyard, J. and Rolland, N. (2004). "The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions." *Phytochemistry* **65**(12): 1693-707.

Bunai, K. and Yamane, K. (2005). "Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives." *J Chromatogr B Analyt Technol Biomed Life Sci* **815**(1-2): 227-36.

C

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

Carapito, C., Muller, D., Turlin, E., Koechler, S., Danchin, A., Van Dorsselaer, A., Leize-Wagner, E., Bertin, P. N. and Lett, M. C. (2005). "Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caenibacter arsenoxydans*, a metalloresistant beta-proteobacterium with an unsequenced genome." *Biochimie* **5**: 5.

Carlson, S. M., Najmi, A., Whitin, J. C. and Cohen, H. J. (2005). "Improving feature detection and analysis of surface-enhanced laser desorption/ionization-time of flight mass spectra." *Proteomics* **5**(11): 2778-88.

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K. and Nesvizhskii, A. (2004). "The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data." *Mol Cell Proteomics* **3**: 531-533.

Carte, N. (2001). "La trappe ionique et l'ionisation électrospray : un nouveau potentiel pour la caractérisation de biomolécules." thèse de l'université Louis Pasteur.

Chevreux, G. (2005). "Etude des interactions moléculaires spécifiques par spectrométrie de masse: application à la chimie du vivant." Thèse de l'université Louis Pasteur.

Colby, S. M., King, T. B., Reilly, J. P., Lubman, D. M. (1994). "Improving the resolution of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry by exploiting the correlation between ion position and velocity." *Rapid Communications in Mass Spectrometry* **8**(11): 865-868.

Coombes, K. R. M., J. S.; Hu, J.; Edmonson, S. R.; Baggerly, K. A. (2005). "Serum proteomics profiling--a young technology begins to mature." *Nat Biotechnol* **23**(3): 291-2.

Cotter, R. J. (1989). "Time-of-flight mass spectrometry: an increasing role in the life sciences." *Biomed Environ Mass Spectrom* **18**(8): 513-32.

D

- Delalande, F., Carapito, C., Brizard, J. P., Brugidou, C. and Van Dorsselaer, A. (2005). "Multigenic families and proteomics: extended protein characterization as a tool for paralog gene identification." *Proteomics* **5**(2): 450-60.
- Deshaies, R. J., Seol, J. H., McDonald, W. H., Cope, G., Lyapina, S., Shevchenko, A., Shevchenko, A., Verma, R. and Yates, J. R., 3rd (2002). "Charting the protein complexome in yeast by mass spectrometry." *Mol Cell Proteomics* **1**(1): 3-10.
- Desplancq, D., Bernard, C., Sibler, A. P., Kieffer, B., Miguet, L., Potier, N., Van Dorsselaer, A. and Weiss, E. (2005). "Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120." *Biotechniques* **39**(3): 405-11.
- Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D. and Alice, M. B. (1968). "Molecular Beams of Macroions." *J Chem Phys* **49**(5): 2240-9.
- Dreger, M. (2003). "Subcellular proteomics." *Mass Spectrom Rev* **22**(1): 27-56.

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.
- Fischer, F. and Poetsch, A. (2006). "Protein cleavage strategies for an improved analysis of the membrane proteome." *Proteome Sci* **4**(1): 2.
- Freyssinet, J. M. (2003). "Cellular microparticles: what are they bad or good for?" *J Thromb Haemost* **1**(7): 1655-62.
- Friedman, M. H., Yergey, A. L. and Campana, J. E. (1982). "Fundamentals of ion motion in electric radio-frequency multipole fields." *J. Phys. E: Sci. Instrum.* **15**: 53-56.

G

- Galeva, N. and Altermann, M. (2002). "Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins." *Proteomics* **2**(6): 713-22.
- Gallagher, R. T., Chapman, J. R., Mann M. (1990). "Design and performance of an electrospray ionization source for a doubly-focusing magnetic sector mass spectrometer." *Rapid Communications in Mass Spectrometry* **4**(10): 369-372.

- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. and Vandekerckhove, J. (2003). "Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides." *Nat Biotechnol* **21**(5): 566-9.
- Glückmann, M., Michael Karas (1999). "The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization." *Journal of Mass Spectrometry* **34**(5): 467-477.
- Gong, Y., Li, X., Yang, B., Ying, W., Li, D., Zhang, Y., Dai, S., Cai, Y., Wang, J., He, F. and Qian, X. (2006). "Different immunoaffinity fractionation strategies to characterize the human plasma proteome." *J Proteome Res* **5**(6): 1379-87.
- 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.
- 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

- Hanash, S. (2003). "Disease proteomics." *Nature* **422**(6928): 226-32.
- Heintz, D. W., V.; High, A. A.; Van Dorsselaer, A.; Reski, R.; Sarnighausen, E. (2004). "An efficient protocol for the identification of protein phosphorylation in a seedless plant, sensitive enough to detect members of signalling cascades." *Electrophoresis* **25**(7-8): 1149-59.
- Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguët, L., Potier, N., Van-Dorsselaer, A., Wurtz, J. M., Mandel, J. L., Tora, L. and Devys, D. (2004). "Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes." *Hum Mol Genet* **13**(12): 1257-65.
- Henry, K. D., Williams, E. R., Wang, B. H., McLafferty, F. W., Shabanowitz, J. and Hunt, D. F. (1989). "Fourier-transform mass spectrometry of large molecules by electrospray ionization." *Proc Natl Acad Sci* **86**(23): 9075-8.
- Hublitz, P., Kunowska, N., Mayer, U. P., Müller, J. M., Heyne, K., Yin, N., Fritzsche, C., Poli, C., Miguët, L., Schupp, I. W., van Grunsven, L. A., Potiers, N., van Dorsselaer, A., Metzger, E., Roemer, K. and Schule, R. (2005). "NIR is a novel INHAT repressor that modulates the transcriptional activity of p53." *Genes Dev* **19**(23): 2912-24.

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.

Hutchens, T. W., Tai-Tung Yip, T. (1993). "New desorption strategies for the mass spectrometric analysis of macromolecules." *Rapid Communications in Mass Spectrometry* **7**(7): 576-580.

I

Ikonomou, M. G., Blades A. T., Kebarle, P. (1991). "Electrospray-Ion Spray: A comparison of mechanisms and performance." *Anal Chem* **63**: 1989-98.

Iribarne, J. V., Thomson B. A. (1976). "On the evaporation of small ions from charged droplets." *J Chem Phys* **64**(6): 2287-2294.

J

Jang, J. H. and Hanash, S. (2003). "Profiling of the cell surface proteome." *Proteomics* **3**(10): 1947-54.

K

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

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

Kebarle, P. (2000). "A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry." *J Mass Spectrom* **35**(7): 804-17.

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

Kenyon, G. L., DeMarini, D. M., Fuchs, E., Galas, D. J., Kirsch, J. F., Leyh, T. S., Moos, W. H., Petsko, G. A., Ringe, D., Rubin, G. M. and Sheahan, L. C. (2002). "Defining the mandate of proteomics in the post-genomics era: workshop report." *Mol Cell Proteomics* **1**(10): 763-80.

Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F. and Oesterhelt, D. (2005). "The membrane proteome of *Halobacterium salinarum*." *Proteomics* **5**(1): 180-97.

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

L

Lill, J. (2003). "Proteomic tools for quantitation by mass spectrometry." *Mass Spectrom Rev* **22**(3): 182-94.

Loo, J. A., DeJohn, D. E., Du, P., Stevenson, T. I. and Ogorzalek Loo, R. R. (1999). "Application of mass spectrometry for target identification and characterization." *Med Res Rev* **19**(4): 307-19.

Lottspeich, F. (1999). "Proteome Analysis: A Pathway to the Functional Analysis of Proteins." *Angew Chem Int Ed Engl* **38**(17): 2476-2492.

Lu, X. and Zhu, H. (2005). "Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins." *Mol Cell Proteomics* **4**(12): 1948-58.

Luche, S., Santoni, V. and Rabilloud, T. (2003). "Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis." *Proteomics* **3**(3): 249-53.

M

Malmstrom, J., Larsen, K., Malmstrom, L., Tufvesson, E., Parker, K., Marchese, J., Williamson, B., Hattan, S., Patterson, D., Martin, S., Graber, A., Juhasz, H. P., Westergren-Thorsson, G. and Marko-Varga, G. (2004). "Proteome annotations and identifications of the human pulmonary fibroblast." *J Proteome Res* **3**(3): 525-37.

Mamyrin, B. A., Karataev, V. I., Shmikk, D. V., Zagulin, V. A. (1973). "Mass reflectron. New nonmagnetic time-of-flight high-resolution mass spectrometer." *Sov. phys. JETP* **64**: 82-89.

Mann, M. and Jensen, O. N. (2003). "Proteomic analysis of post-translational modifications." *Nat Biotechnol* **21**(3): 255-61.

Martin Kussmann (1997). "Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes." *Journal of Mass Spectrometry* **32**(6): 593-601.

- McGregor, E. and Dunn, M. J. (2006). "Proteomics of the heart: unraveling disease." *Circ Res* **98**(3): 309-21.
- Medzihradzky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L. and Burlingame, A. L. (2000). "The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer." *Anal Chem* **72**(3): 552-8.
- Merchant, M. and Weinberger, S. R. (2000). "Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry." *Electrophoresis* **21**(6): 1164-77.
- Mirgorodskaya, O. A., Shevchenko, A. A., Chernushevich, I. V., Dodonov, A. F. and Miroshnikov, A. I. (1994). "Electrospray-ionization time-of-flight mass spectrometry in protein chemistry." *Anal Chem* **66**: 99-107.
- Mollenkopf, H. J., Jungblut, P. R., Raupach, B., Mattow, J., Lamer, S., Zimny-Arndt, U., Schaible, U. E. and Kaufmann, S. H. (1999). "A dynamic two-dimensional polyacrylamide gel electrophoresis database: the mycobacterial proteome via Internet." *Electrophoresis* **20**(11): 2172-80.

N

- Niu, S., Zhang, W. and Chait, B. T. (1998). "Direct comparison of infrared and ultraviolet wavelength matrix-assisted laser desorption/ionization mass spectrometry of proteins." *J Am Soc Mass Spectrom* **9**(1): 1-7.
- Nomine, Y., Charbonnier, S., Miguet, L., Potier, N., Van Dorsselaer, A., Atkinson, R. A., Trave, G. and Kieffer, B. (2005). "¹H and ¹⁵N resonance assignment, secondary structure and dynamic behaviour of the C-terminal domain of human papillomavirus oncoprotein E6." *J Biomol NMR* **31**(2): 129-41.

O

- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." *J Biol Chem* **250**(10): 4007-21.
- Olthoff, J. K., Lys, I. A. and Cotter, R. J. (1988). "A pulsed time-of-flight mass spectrometer for liquid secondary ion mass spectrometry." *Rapid Commun Mass Spectrom* **2**(9): 171-75.
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." *Mol Cell Proteomics* **1**(5): 376-86.

P

- Pandey, A. and Mann, M. (2000). "Proteomics to study genes and genomes." *Nature* **405**(6788): 837-46.
- Patterson, S. D. and Aebersold, R. H. (2003). "Proteomics: the first decade and beyond." *Nat Genet* **33 Suppl**: 311-23.
- Paul, W. and Raether M. (1955). "Das elektrische Massenfilter." *Z Phys* **140**: 262-73.
- Peirce, M. J., Wait, R., Begum, S., Saklatvala, J. and Cope, A. P. (2004). "Expression profiling of lymphocyte plasma membrane proteins." *Mol Cell Proteomics* **3**(1): 56-65. Epub 2003 Oct 21.

R

- Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J. (1997). "Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients." *Electrophoresis* **18**(3-4): 307-16.
- Richert, S. (2003). "De l'identification à la caractérisation des protéines: Développement de la spectrométrie de masse dans le cadre de l'approche protéomique." Thèse de l'université Louis Pasteur.
- 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.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Seraphin, B. (1999). "A generic protein purification method for protein complex characterization and proteome exploration" *Nat Biotechnol* **17**(10): 1030-1032
- 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.
- Rosenfeld, J., Capdevielle, J., Guillemot, J. C. and Ferrara, P. (1992). "In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis." *Anal Biochem* **203**(1): 173-9.

S

- Sanglier, S. (2002). "La spectrométrie de masse: un nouvel outil pour l'étude des interactions faibles en biologie." Thèse de l'université Louis Pasteur.
- Santoni, V., Molloy, M. and Rabilloud, T. (2000). "Membrane proteins and proteomics: un amour impossible?" *Electrophoresis* **21**(6): 1054-70.
- Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D., Mansion, M., Kieffer, S., Garin, J. and Rossignol, M. (1999). "Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels." *Electrophoresis* **20**(4-5): 705-11.
- Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. and Robinson, C. V. (2002). "A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies." *Anal Chem* **74**(6): 1402-7.
- Stafford, J. G. C., Kelley, P. E., Syka, J. E. P., Reynolds, W. E. and Todd, J. F. J. (1984). "Recent improvements in and analytical applications of advanced ion trap technology." *International Journal of Mass Spectrometry and Ion Processes* **60**(1): 85-98.

T

- Tang, L. and Kebarle, P. (1993). "Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analyte in the Electrosprayed Solution." *Anal Chem* **65**: 3654-3668.
- Thomson, B. A., Iribarne, J. V. (1979). "Field induced ion evaporation from liquid surfaces at atmospheric pressure." *J Chem Phys* **71**(11): 4451-63.
- Toda, T., Kajii, K. and Kimura, N. (1998). "TMIG-2DPAGE: a new concept of two-dimensional gel protein database for research on aging." *Electrophoresis* **19**(2): 344-8.
- Tonella, L., Hoogland, C., Binz, P. A., Appel, R. D., Hochstrasser, D. F. and Sanchez, J. C. (2001). "New perspectives in the Escherichia coli proteome investigation." *Proteomics* **1**(3): 409-23.

U

- Unlu, M., Morgan, M. E. and Minden, J. S. (1997). "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts." *Electrophoresis* **18**(11): 2071-7.

V

Veenstra, T. D., Conrads, T. P., Hood, B. L., Avellino, A. M., Ellenbogen, R. G. and Morrison, R. S. (2005). "Biomarkers: mining the biofluid proteome." *Mol Cell Proteomics* **4**(4): 409-18.

Venter, J. C. et al (2001). "The sequence of the human genome." *Science* **291**(5507): 1304-51.

W

Wagner, E. (2004). "Nouvelles approches de la spectrométrie de masse en tandem pour la caractérisation de biomolécules. Application à la protéomique." Thèse de l'université Louis Pasteur.

Wallin, E. and von Heijne, G. (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." *Protein Sci* **7**(4): 1029-38.

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.

Weinberger, S. R., Morris, T. S. and Pawlak, M. (2000). "Recent trends in protein biochip technology." *Pharmacogenomics* **1**(4): 395-416.

Wiley, W. C. and McLaren, I. H. (1955). "Time-of-Flight Mass Spectrometer with Improved Resolution." *Rev Sci Instr* **16**(12): 1150-7.

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.

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

Winger, B. E., Light-Wahl, K. J., Ogorzalek Loo, R. R., Udseth, H. R. and Smith, R. D. (1993). "Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry." *J Am Soc Mass Spectrom* **4**: 536-545.

Wu, W. W., Wang, G., Baek, S. J. and Shen, R. F. (2006). "Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF." *J Proteome Res* **5**(3): 651-8.

Y

Yamashita, M. and Fenn, J. B. (1984). "Electrospray Ion Source. Another Variation on the Free-Jet Theme." *J. Phys. Chem.* **88**: 4451-4460.

Yergey, A. L., Coorsen, J. R., Backlund, P. S., Jr., Blank, P. S., Humphrey, G. A., Zimmerberg, J., Campbell, J. M. and Vestal, M. L. (2002). "De novo sequencing of peptides using MALDI/TOF-TOF." *J Am Soc Mass Spectrom* **13**(7): 784-91.

Z

Zhao, Y., Zhang, W. and Kho, Y. (2004). "Proteomic analysis of integral plasma membrane proteins." *Anal Chem* **76**(7): 1817-23.

Partie II :

Résultats

- Chapitre 1 : Analyse différentielle et quantification par spectrométrie de masse de mélanges complexes.
- Chapitre 2 : Elaboration de nouveaux outils et de nouvelles stratégies analytiques appliqués à l'analyse des protéines membranaires par spectrométrie de masse.
- Chapitre 3 : Recherche de nouveaux biomarqueurs dans le cadre des hémopathies lymphoïdes chroniques

Analyse différentielle et quantification par spectrométrie de masse de protéines issues de mélanges complexes.

Conclusions majeures :

- Nous avons identifié des protéines impliquées dans le transport du fer chez *Pseudomonas aeruginosa* par une approche gel 2D et analyse MALDI-MS.
- Nous avons développé un nouveau marqueur isotopique qui ne présente pas d'étape de chromatographie d'affinité. Ce marqueur a été validé pour les différentes étapes de l'analyse protéomique et appliqué à différents mélanges protéiques complexes.

La première partie de ce travail de thèse est basée sur la quantification relative des protéines issues de mélanges complexes. Ce chapitre s'articule autour de deux points (figure 1). Le premier qui illustre l'approche classique par étude différentielle de **gels d'électrophorèse bidimensionnelle**, avec pour exemple l'étude de l'expression des protéines membranaires de *Pseudomonas aeruginosa*, et le deuxième qui reporte le **développement et la mise au point d'un marqueur isotopique** qui ne présente pas d'étape de chromatographie d'affinité.

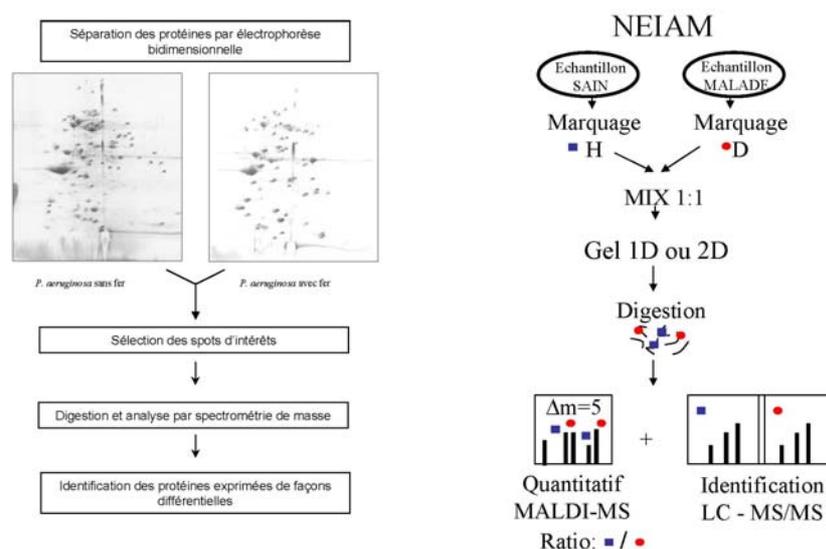


Figure 1 : Représentation des deux approches de quantification, par gel d'électrophorèse bidimensionnelle et par marquage isotopique sans étape de chromatographie d'affinité.

1) Etude des protéines membranaires impliquées dans le transport du fer chez *Pseudomonas aeruginosa*. Approche différentielle par électrophorèse bidimensionnelle et analyse MALDI-MS.

Cette étude se situe dans le cadre d'une collaboration avec le département des Récepteurs et Protéines membranaires de Frank Pattus et plus particulièrement en collaboration avec Isabelle Schalk (UMR7175 LC-1 CNRS/ULP, Institut Gilbert-Laustriat). Cette étude vise à identifier les protéines membranaires exprimées de façon différentielle entre une souche de *Pseudomonas aeruginosa* cultivée en présence ou en absence de fer. L'importance de l'électrophorèse bidimensionnelle couplée à l'analyse par MALDI-MS pour réaliser l'identification et la quantification des différentes protéines observables sur gel 2D est ici soulignée.

1.1) Introduction et intérêt biologique

Pseudomonas aeruginosa est une bactérie Gram-négatif (une membrane interne – une paroi - une membrane externe (figure 2)) omniprésente et responsable d'un grand nombre d'affections nosocomiales. Elle représente plus de 90% des infections dues au genre *Pseudomonas*. Cette bactérie, de quelques microns de diamètre, est capable de s'adapter très rapidement aux variations environnementales, et a développé une résistance à un grand nombre d'antibiotiques [Obritsch et al. 2005].

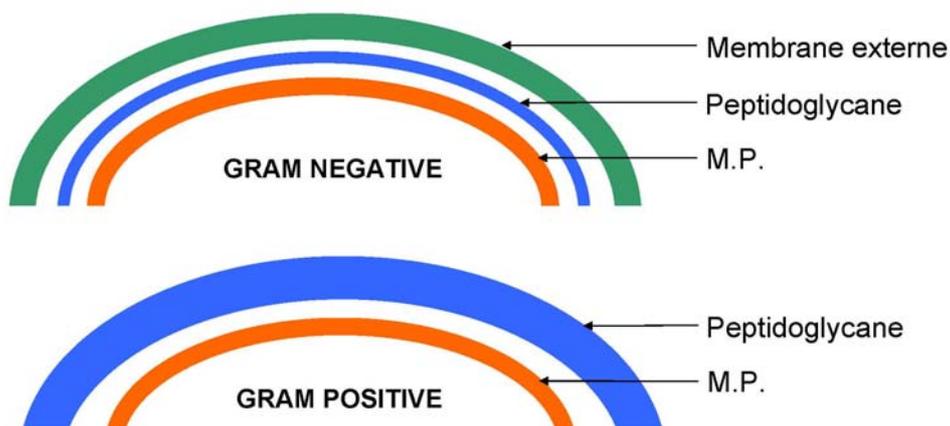


Figure 2 : Représentation schématique des membranes bactériennes en fonction de leur origine GRAM positive ou négative. Dans nos différentes études, nous avons travaillé sur des bactéries GRAM négatives. Ces bactéries présentent une membrane interne ou plasmique et une membrane externe, séparées par une fine paroi de peptidoglycane.

Le fer, bien que peu soluble dans les phases aqueuses à pH physiologique, est un oligoélément essentiel pour la vie. Comme tous les organismes *Pseudomonas aeruginosa* a besoin de fer pour sa

croissance et son développement. De plus, l'incorporation de fer par la bactérie est cruciale pour sa virulence [Hassett et al. 1995]. Lors d'infections, ces bactéries vont synthétiser toute une série de protéines qui vont lui permettre d'incorporer le fer et cela d'une manière très efficace [Schalk et al. 1999] [Schalk et al. 2002]. Bien connaître les différentes protéines impliquées dans le transport du fer entre le milieu extérieur et la bactérie pourrait permettre de trouver de nouvelles cibles thérapeutiques efficaces [Miller et Mahenthiralingam 2003].

1.2) Approche protéomique différentielle par Gel 2D et analyse par empreinte peptidique massique (MALDI-MS)

1.2.1) Préparation des échantillons de protéines membranaires

La structure des protéines membranaires est différente selon qu'elles proviennent de la membrane interne ou de la membrane externe de la bactérie. En effet, les protéines membranaires ont en règle générale des structures en hélice α qui les rendent très peu solubles dans tous les types de solvant qui ne contiennent pas de détergent fort. A l'inverse, les protéines de la membrane externe des bactéries ont en règle générale une structure en tonneau (repliement de feuillet β), structure beaucoup plus soluble dans les différents tampons classiquement utilisés dans l'analyse protéomique. Elles ne nécessitent pas la présence d'un détergent fort pour être solubilisées. Dans un premier temps nous ne nous occuperons donc que des protéines issues de la membrane externe de *Pseudomonas aeruginosa*. Les différentes préparations de membranes externes, issues des souches cultivées en présence et en l'absence de fer, ont été obtenues après lyse des cellules et centrifugations multiples sur gradient de sucrose [Schalk et al. 1999]. Cette méthode permet de séparer assez facilement les membranes internes des membranes externes et d'obtenir des préparations membranaires peu polluées par des contaminants issus de la membrane interne ou d'autres débris bactériens.

1.2.2) Electrophorèse 2D

580 μ g de protéines membranaires issues des deux souches bactériennes (déterminé par dosage en utilisant un kit BioRad compatible avec les tampons de solubilisation) [Chevallet et al 1998] ont été solubilisés (Urée et Thiourée, ASB 14, DTT, inhibiteurs de protéases, Triton X100, ampholytes) et déposées sur des gels à gradient de pH immobilisé (BioRad ready strip IPG 18 cm pH 3-10). La première dimension (isoélectrofocalisation) a été suivie d'un gel SDS PAGE 5-20% avant d'être coloré au bleu colloïdal. Les gels ont été réalisés en double. Après inspection visuelle, les spots présentant des différences d'intensité ont été découpés et analysés par spectrométrie de masse (Figure 3).

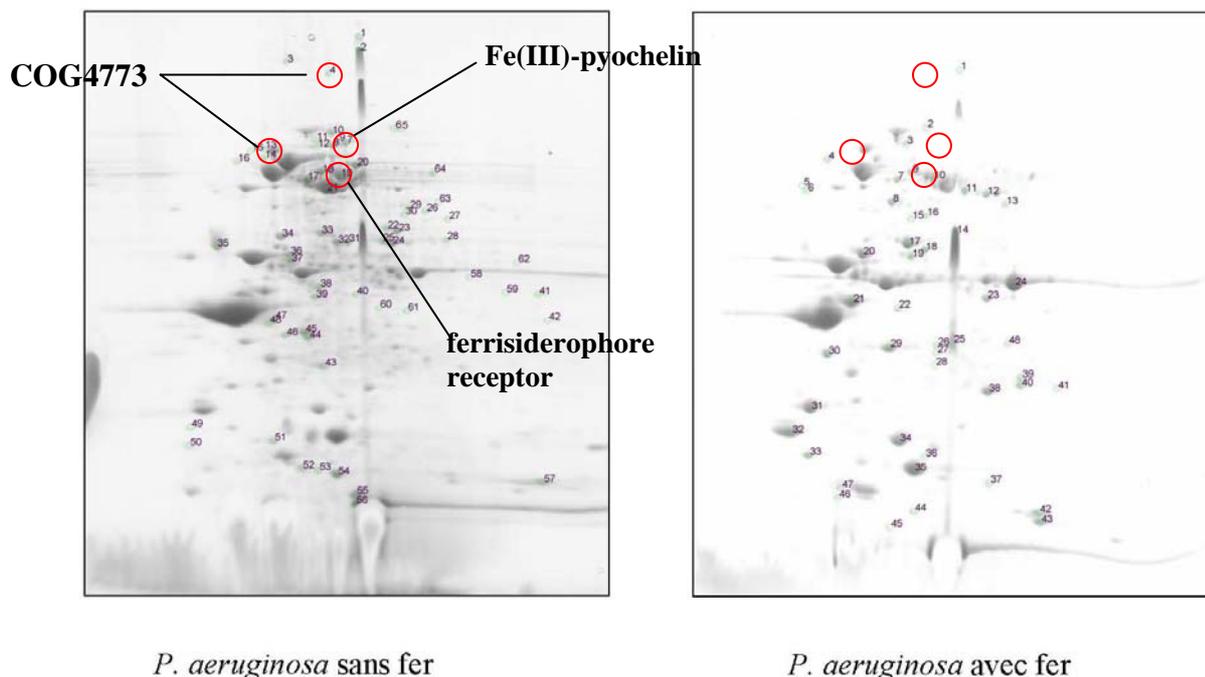


Figure 3 : gels 2D obtenus à partir de préparation de membranes externes de *P. aeruginosa* en présence et en absence de fer. Les protéines cerclées représentent les protéines identifiées impliquées dans le transport du fer.

1.2.3) Analyse MALDI

L'analyse des gels a permis de mettre en évidence 132 spots qui présentaient une différence au niveau de l'intensité de la coloration. Après excision, les protéines contenues dans ces spots ont été réduites et alkylées avant d'être digérées à la trypsine. Les peptides extraits sont analysés par spectrométrie de masse MALDI-TOF. La recherche des différentes listes de masses générées à partir du logiciel FlexAnalysis™ dans la banque de données NCBIInr a permis l'identification de 109 protéines. Ces 109 identifications représentent en fait 75 protéines différentes, puisque une même protéine peut être retrouvée dans plusieurs spots (isoformes, protéines paralogues...). Les différents résultats obtenus sont reportés dans les tableaux 1 et 2.

Spot avec fer	Variation	Protéine	Localisation	Couverture	Nombre de peptides
1	down	Pas d'identification			
2	up	COG0653: Preprotein translocase subunit SecA	Membrane	63%	60
3	up	organic solvent tolerance protein OstA precursor	Membrane	55%	57
4	up	twitching motility protein PilJ	Membrane	30%	22
		COG1629: Outer membrane receptor proteins, mostly Fe transport	Membrane	28%	16
		COG0840: Methyl-accepting chemotaxis protein	Membrane	29%	20
5	up	COG3240: Phospholipase/lecithinase/hemolysin	inconnue	54%	30
6	up	glycine betaine transmethylase	Cytoplasme	39%	26
7	down	Pas d'identification			

8	up	Type 4 fimbrial biogenesis outer membrane protein PilQ precursor	Membrane	49%	35
9	egal	Type 4 fimbrial biogenesis outer membrane protein PilQ precursor	Membrane	67%	49
		hypothetical protein PA0434	Membrane	49%	34
10	down	probable outer membrane protein precursor	Membrane	41%	25
11	up	COG1629: Outer membrane receptor proteins, mostly Fe transport	Membrane	65%	46
12	up	succinate dehydrogenase (A subunit)	Membrane	53%	32
		probable TonB-dependent receptor	Membrane	42%	20
13	up	Pas d'identification			
14	up	Pas d'identification			
15	up	hypothetical protein PA4624	inconnue	54%	29
16	up	electron transfer flavoprotein-ubiquinone oxidoreductase	Cytoplasme	66%	38
17	up	ATP synthase alpha chain	Membrane	54%	32
18	up	probable outer membrane protein precursor	Membrane	81%	37
19	up	probable outer membrane protein precursor	Membrane	76%	34
		COG4625: Uncharacterized protein with a C-terminal OMP (outer membrane protein) domain	Membrane	29%	27
20	up	ATP synthase beta chain	Membrane	81%	38
21	up	COG2885: Outer membrane protein and related peptidoglycan-associated (lipo)proteins	Membrane	62%	24
22	egal	hypothetical protein PA4372	inconnue	74%	25
23	up	COG1686: D-alanyl-D-alanine carboxypeptidase	Cytoplasme	58%	21
24	up	Pas d'identification			
25	up	hypothetical protein PA4441	inconnue	74%	12
26	up	hypothetical protein PA4441	inconnue	74%	12
27	up	competence protein ComL	Membrane	54%	20
28	up	probable outer membrane protein precursor	Membrane	63%	15
		ATP synthase beta chain	Membrane	43%	15
29	up	COG2067: Long-chain fatty acid transport protein	Membrane	53%	21
30	up	probable peptidyl-prolyl cis-trans isomerase, FkbP-type	Cytoplasme	41%	9
31	up	COG3047: Outer membrane protein W	Membrane	52%	18
32	up	COG2885: Outer membrane protein and related peptidoglycan-associated (lipo)proteins	Membrane	35%	10
33	up	COG3047: Outer membrane protein W	Membrane	49%	15
34	up	Pas d'identification			
35	egal	Pas d'identification			
36	up	conserved hypothetical protein	inconnue	68%	12
37	up	Pas d'identification			
38	up	probable iron-sulfur protein	Membrane	76%	13
39	up	probable ATP-binding component of ABC transporter	Membrane	47%	15
		COG0479: Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	Membrane	42%	11
40	up	succinate dehydrogenase (B subunit)	Membrane	57%	14
41	egal	probable ATP-binding component of ABC transporter	Membrane	63%	13
42	up	Pas d'identification			
43	up	Pas d'identification			
44	down	conserved hypothetical protein	inconnue	86%	12
45	up	Pas d'identification			
46	up	Pas d'identification			
47	up	hypothetical protein PA3062	inconnue	68%	9

Tableau 1 : Identification des protéines de la membrane externe de *Pseudomonas aeruginosa* en conditions enrichies en fer.

Spot sans fer	Variation	Protéine	Localisation	Couverture	Nombre de peptides
1	up	probable ATP-binding component of ABC transporter	Membrane	70%	19
2	up	Pas d'identification			
3	up	Pas d'identification			
4	up	COG4773: Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	Membrane	58%	42
5	up	Fe(III)-pyochelin outer membrane receptor precursor	Membrane	56%	34
6	up	2-oxoglutarate dehydrogenase (E1 subunit)	Cytoplasme	31%	33
7	up	Fe(III)-pyochelin outer membrane receptor precursor	Membrane	42%	28
8	up	COG1629: Outer membrane receptor proteins, mostly Fe transport	Membrane	45%	45
		probable ClpA/B-type protease	Cytoplasme	26%	20
9	up	probable TonB-dependent receptor	Membrane	47%	38
10	up	COG1629: Outer membrane receptor proteins, mostly Fe transport	Membrane	42%	41
11	up	COG0542: ATPases with chaperone activity, ATP-binding subunit	Cytoplasme	59%	66
12	up	organic solvent tolerance protein OstA precursor	Membrane	46%	37
13	up	phenylalanyl-tRNA synthetase, beta subunit	Cytoplasme	44%	34
		probable hydroxamate-type ferrisiderophore receptor	Membrane	42%	31
		probable outer membrane protein precursor	Membrane	36%	23
		probable TonB-dependent receptor	Membrane	29%	19
14	up	probable outer membrane protein precursor	Membrane	41%	27
15	up	COG4773: Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	Membrane	56%	41
16	down	COG1629: Outer membrane receptor proteins, mostly Fe transport	Membrane	46%	34
17	up	twitching motility protein PilJ	Membrane	27%	18
18	egal	COG4773: Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	Membrane	54%	38
19	up	hypothetical protein PA0434	Membrane	18%	19
20	up	probable outer membrane protein precursor	Membrane	78%	56
21	up	Haem/Haemoglobin uptake outer membrane receptor PhuR precursor	Membrane	60%	39
22	up	Pas d'identification			
23	up	probable class III aminotransferase	Cytoplasme	68%	27
24	up	probable class III aminotransferase	Cytoplasme	75%	31
25	up	L-ornithine N5-oxygenase	Cytoplasme	75%	28
26	up	COG3486: Lysine/ornithine N-monooxygenase	Cytoplasme	70%	26
27	up	acetolactate synthase large subunit	Cytoplasme	48%	35
28	up	Chain B, Dihydrolipoamide Dehydrogenase (E.C.1.8.1.4) Complex With Flavin-Adenine-Dinucleotide (Fad	Cytoplasme	43%	19
29	up	probable class III aminotransferase	Cytoplasme	50%	18
30	up	conserved hypothetical protein	Inconnue	52%	26
		hypothetical protein Paer03004127	Inconnue	31%	19
31	up	conserved hypothetical protein	Inconnue	65%	33
32	egal	PA2391	Membrane	62%	23
		probable outer membrane protein precursor	Membrane	47%	18
33	down	probable outer membrane protein precursor	Membrane	78%	35
34	up	ATP synthase alpha chain	Cytoplasme	54%	33
		conserved hypothetical protein	Membrane	53%	24
35	up	COG1344: Flagellin and related hook-associated proteins	Cytoplasme	59%	22
36	up	Insulin-cleaving metalloproteinase outer membrane protein precursor	Membrane	50%	28

37	up	COG3659: Carbohydrate-selective porin	Membrane	61%	22
		isocitrate dehydrogenase	Cytoplasme	52%	21
		Basic amino acid, basic peptide and imipenem outer membrane porin OprD precursor	Membrane	44%	15
38	up	isocitrate dehydrogenase	Cytoplasme	66%	33
39	egal	COG2355: Zn-dependent dipeptidase, microsomal dipeptidase homolog	Membrane	50%	24
		hypothetical protein PA2033	Inconnue	49%	15
40	up	hypothetical protein PA4372	Inconnue	74%	23
41	up	ORF_16; similar to UDP-N-acetylglucosamine 2-epimerase	Cytoplasme	64%	23
		ketol-acid reductoisomerase	Cytoplasme	49%	14
42	up	RNA 3'-terminal phosphate cyclase	Cytoplasme	64%	20
43	up	COG1652: Uncharacterized protein containing LysM domain	Inconnue	52%	15
44	up	hypothetical protein PA4836	Inconnue	85%	21
45	down	pyoverdine synthetase F	Cytoplasme	50%	18
46	up	probable outer membrane protein precursor	Membrane	53%	19
		pyoverdine synthetase F	Cytoplasme	37%	14
47	up	COG1192: ATPases involved in chromosome partitioning	Cytoplasme	56%	12
48	up	COG2885: Outer membrane protein and related peptidoglycan-associated (lipo)proteins	Membrane	62%	23
49	up	COG2885: Outer membrane protein and related peptidoglycan-associated (lipo)proteins	Membrane	45%	15
50	up	Pas d'identification			
51	up	Pas d'identification			
52	up	Chain A, Nmr Structure Of P. Aeruginosa Protein Pa1324: Northeast Structural Genomics Consortium Ta	cytoplasme	54%	10
53	up	Pas d'identification			
54	up	conserved hypothetical protein	Inconnue	87%	12
55	egal	conserved hypothetical protein	Inconnue	80%	16
56	egal	conserved hypothetical protein	Inconnue	70%	9
57	up	Pas d'identification			
58	up	COG3755: Uncharacterized protein conserved in bacteria	Inconnue	60%	11
59	up	Pas d'identification			
60	up	RNA 3'-terminal phosphate cyclase	Cytoplasme	65%	21
61	up	Pas d'identification			
62	up	Pas d'identification			
63	up	Pas d'identification			
64	down	pyochelin biosynthesis protein PchD	Cytoplasme	45%	27
65	up	probable outer membrane protein precursor	Membrane	51%	34
		probable TonB-dependent receptor	Membrane	54%	34

Tableau 2 : Identification des protéines de la membrane externe de *Pseudomonas aeruginosa* en conditions carencées en fer.

1.3) Résultats

Sur les 75 protéines identifiées par cette approche MALDI-TOF, 38 sont localisées au niveau de la membrane (51%). Parmi ces dernières, 12 sont liées au transport de métaux. 3 protéines impliquées dans le transport du fer ont été identifiées comme surexprimées dans la préparation issue de la souche de *Pseudomonas Aeruginosa* cultivée dans un milieu carencé en fer. Ces différentes

protéines (COG4773: Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid [Sauer et al. 1987] spot 4 et spot 15; Fe(III)-pyochelin outer membrane receptor precursor [Tseng et al. 2006] spot 5 et 7; probable hydroxamate-type ferrisiderophore receptor [James et al. 2005] spot 13) sont mises en évidence sur la figure 1. Ces protéines possèdent une très forte affinité pour le fer (de l'ordre du nano-molaire) et sont connues pour jouer un rôle important dans son transport au travers de la membrane [Schalk et al. 2002]. Cette première approche confirme donc la surexpression de ces protéines dans *Pseudomonas* lorsque ces bactéries se développent dans des milieux carencés en fer. De plus cette étude a permis de mettre en évidence l'existence d'isoformes provenant de modifications post-traductionnelles, ainsi que des variants issus de différents gènes paralogues (protéines identifiées dans plusieurs spots présentant des points isoélectriques et des masses différentes), mais ces différentes espèces protéiques n'ont pas été caractérisées.

1.4) Conclusions et Perspectives

Cette première étude descriptive montre la possibilité d'identifier et de quantifier un grand nombre de protéines membranaires par une approche d'électrophorèse bidimensionnelle et analyse par empreinte peptidique massique (analyse MALDI-MS). Elle devra être complétée par des analyses nanoLC-MS/MS, pour pouvoir identifier les différents gènes paralogues exprimés lors du transport du fer par *Pseudomonas aeruginosa*. Cette étude peut être considérée comme une première étape pour une meilleure compréhension des mécanismes impliqués dans le transport du fer à travers la membrane bactérienne, ce qui pourrait permettre de trouver de nouvelles cibles thérapeutiques contre des souches bactériennes qui deviennent de plus en plus résistantes aux antibiotiques.

2) Quantification par marquage isotopique, développement de nouveaux marqueurs sans étape de chromatographie d'affinité. Application à des mélanges complexes.

Au sein du laboratoire nous avons entrepris la synthèse et le développement d'un marqueur isotopique spécifique des cystéines permettant la quantification de mélanges protéiques complexes sans perte d'information peptidique. Ce travail a été initié par Christelle Guillier ancienne Post-Doctorante du laboratoire.

2.1) Introduction

Comme nous l'avons déjà mentionné dans l'introduction de ce travail de thèse, les marqueurs commerciaux (type ICAT : isotope coded affinity tag) présentent une étape de chromatographie d'affinité qui engendre la perte de tous les peptides qui n'ont pas réagi avec le marqueur (Figure 4).

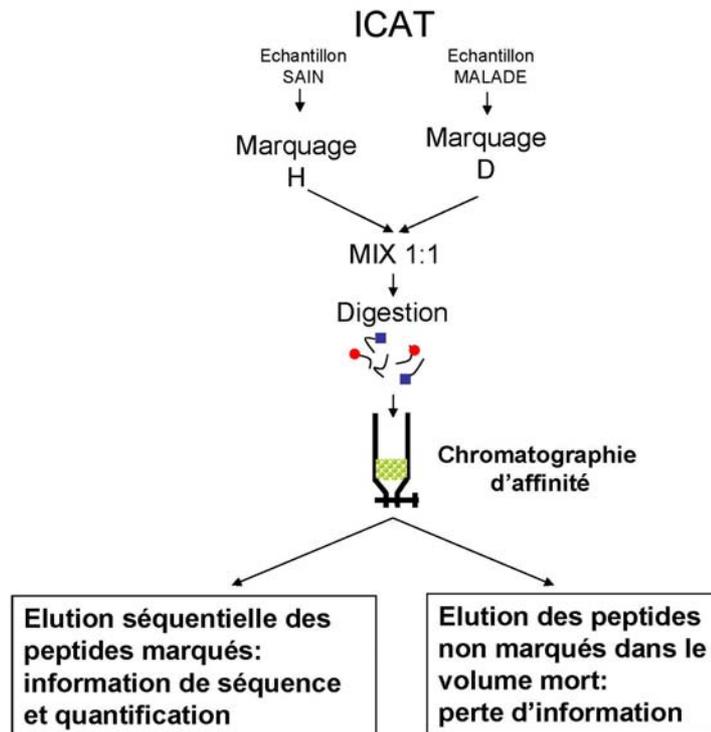


Figure 4 : Les peptides qui ne sont pas marqué sont élués dans le volume mort de la colonne et sont donc perdus pour le reste de l'analyse.

Cette étape de sélection est particulièrement restrictive lorsque l'on travaille sur des organismes dont le génome n'est pas séquencé. En effet, dans ce genre de cas il est nécessaire d'avoir le maximum d'informations liées aux différents peptides tryptiques issues de la digestion.

Nous avons donc décidé de développer au sein du laboratoire une approche utilisant un réactif spécifique des cystéines, et dans laquelle la chromatographie d'affinité est remplacée par une séparation sur gel d'électrophorèse, et ceci avant la digestion des protéines.

Le réactif choisi doit pouvoir être synthétisé de manière simple et rapide et exister sous deux formes, une forme légère et une forme lourde, où les hydrogènes sont remplacés par des deutériums. La différence de masse entre les deux espèces doit être d'au moins quatre Daltons (Figure 5). Cette différence de masse est nécessaire pour que les massifs isotopiques des peptides marqués avec le réactif léger par rapport aux peptides marqués avec le réactif lourd ne soient pas confondus.

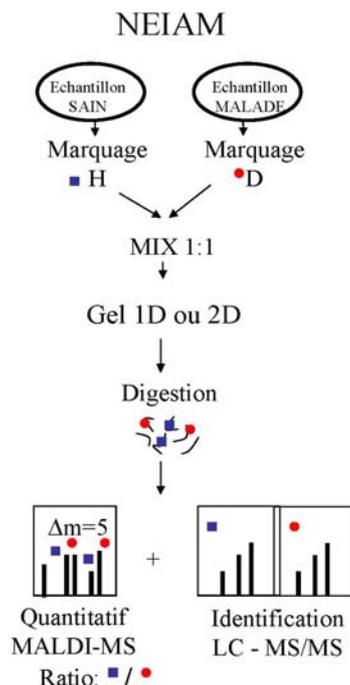


Figure 5 : Schéma de la stratégie développée au laboratoire pour la quantification de mélanges protéiques complexes. Le NEIAM (N-éthyl iodoacétamide) existe sous deux formes, une forme légère enrichie en hydrogène et une forme lourde enrichie en deutérium. Les protéines sont séparées sur gel d'électrophorèse, digérées et analysées par spectrométrie de masse. La quantification se fait par comparaison des intensités relatives des peptides cystéines portant un marqueur D0 par rapport aux peptides cystéines portant un marqueur D5.

2.2) Choix du marqueur et voie de synthèse

Le marqueur choisi est le N-éthyl iodoacétamide (NEIAM) précédemment décrit par Guo et coll. en 2002 [Guo et al. 2002]. Ce marqueur est très facilement synthétisé en faisant réagir en milieu basique de l'anhydride iodoacétique sur un sel d'éthylamine (figure 6). Ce sel existe commercialement sous une forme légère où le groupement éthyl ne présente que des hydrogènes et sous une forme lourde où les 5 hydrogènes sont remplacés par des deutériums. On peut ainsi synthétiser de façon très rapide un marqueur lourd et un marqueur léger dont la différence de masse est de 5 Da, pour un moindre coût. De plus, la purification de ce composé se fait par simple précipitation éther/hexane.

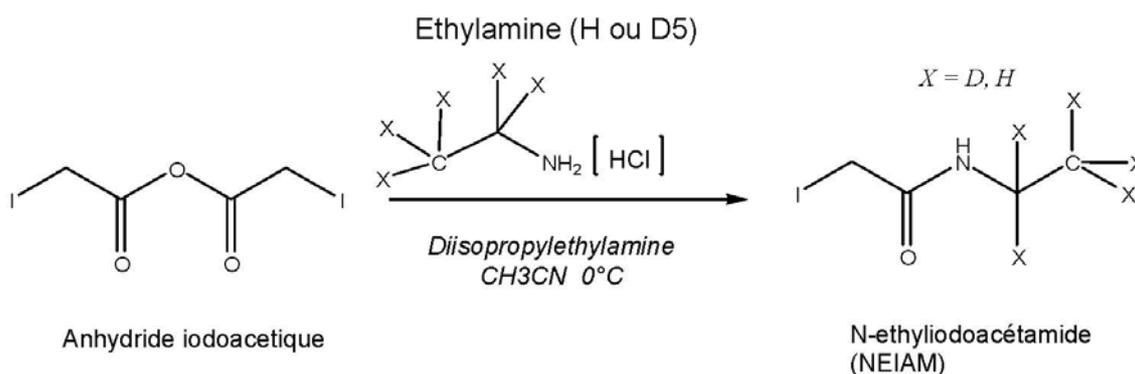


Figure 6 : Voie de synthèse du marqueur. La réaction se fait par l'addition d'anhydride iodoacétique sur de la diisopropyléthylamine marquée D0 ou D5 à 0°C (Cf annexes).

2.3) Validation du marqueur pour des études protéomiques par spectrométrie de masse

La première étape de ce travail a été de montrer la compatibilité de ce marqueur vis-à-vis des analyses protéomiques, et ceci selon trois points distincts. Tout d'abord, le marqueur doit être spécifique des résidus cystéines, de plus il doit être compatible avec les analyses par spectrométrie de masse MS et MS/MS et enfin il doit être compatible avec la chromatographie liquide.

2.3.1) Spécificité et efficacité du marqueur.

Le premier point testé est la spécificité du marqueur vis-à-vis des groupements cystéines. Ces tests ont été réalisés sur une protéine modèle : la RNase A qui comporte 8 cystéines formant 4 ponts disulfures. Après la réduction des ponts disulfures par un ajout soit de DTT, soit de TCEP (Tris(2-carboxyethyl)phosphine). L'intérêt du TCEP par rapport au DTT est qu'il ne présente pas de fonction thiol, et donc est un composé inerte vis-à-vis de notre marqueur. Le NEIAM a été ajouté et une cinétique de marquage a été réalisée entre 2 et 120 minutes. Le suivi de la réaction était effectué par ESI-MS. Au bout de 30 minutes la réaction peut être considérée comme totale, étant donné que la masse de la protéine n'évolue plus (figure 7).

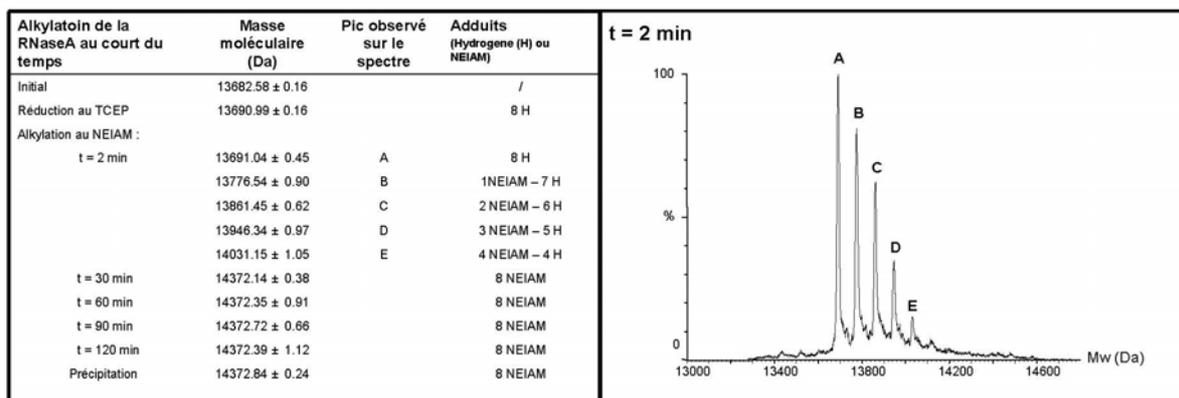


Figure 7 : Cinétique de réaction du NEIAM sur une protéine modèle la RNaseA. La figure présente un spectre ESI déconvolué de la RNaseA au bout de 2 minutes de réaction avec le NEIAM. 5 formes sont visibles correspondant à l'adduit de 1, 2, 3, 4 ou 5 NEIAM.

En effet, on peut remarquer sur la figure qu'au bout de 30 minutes la masse de la RNaseA n'évolue plus (14372 Da). La différence de masse par rapport à la RNase A réduite (13690 Da) est de 681 Da. Cette différence de masse correspond exactement à l'addition de 8 groupements NEIAM ($85,1 \times 8 = 680,8$ Da) correspondant aux 8 cystéines de la protéine. On peut donc en déduire que la réaction est totale au bout de 30 minutes et qu'elle est raisonnablement spécifique des résidus cystéines (uniquement 8 adduits même avec un temps de réaction de 2 heures). Afin de valider la spécificité des sites d'alkylation, nous avons digéré la protéine afin d'analyser les différents peptides par spectrométrie de masse, et d'identifier les peptides qui présentent une modification. Les masses des peptides observés ont été comparées aux masses théoriques attendues (digestion in silico de la RNase A). Seules les masses des peptides présentant une cystéine dans la séquence se sont révélées modifiées par le marqueur indiquant que la réaction d'alkylation se fait de façon spécifique sur les résidus cystéine. On peut donc en conclure que le marqueur développé au laboratoire est spécifique des résidus cystéine des protéines et est efficace (marquage de tous les résidus).

2.3.2) Compatibilité du marqueur vis-à-vis des analyses protéomiques.

Le deuxième point qui devait être validé est la compatibilité du marqueur avec les différentes étapes de protéomique à savoir l'électrophorèse et la linéarité de la réponse lors de l'analyse par spectrométrie de masse. Pour cela, une protéine (l'albumine) a été traitée avec le NEIAM et déposée sur gel d'électrophorèse. La masse globale de la protéine modifiée étant très proche de la masse de la protéine de départ, la résolution du gel n'est pas suffisante pour voir la différence. La linéarité du marquage (obtenir une intensité proportionnelle à la quantité de marqueur) a été testée en réalisant différents ratios de marqueurs avant d'être analysés par MALDI-TOF. Ces résultats sont reportés sur la (figure 8).

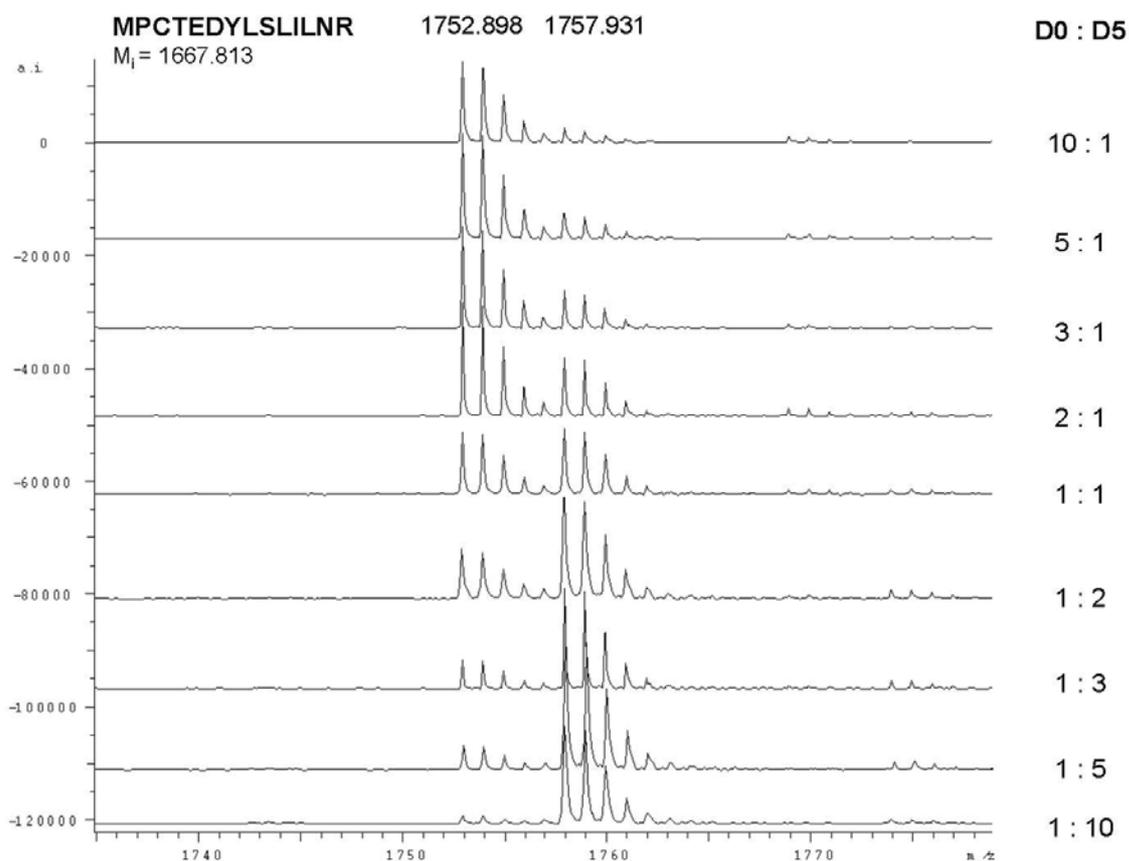


Figure 8 : Analyse MALDI-MS de différents ratios D0/D5 d'un peptide cystéine de la BSA.

Comme le montre le spectre, on peut observer la linéarité de la réponse en fonction de la quantité de protéine analysée. En effet, si on regarde le ratio 1/1 entre le NEIAM D0 et D5, on voit que la réponse observée par MALDI-MS est linéaire, étant donné que les intensités des deux formes sont égales.

2.3.3) Compatibilité du marqueur vis-à-vis de la chromatographie liquide.

Le troisième point était de vérifier la compatibilité du marqueur vis-à-vis de la chromatographie liquide et de la fragmentation par spectrométrie de masse en tandem. En effet, il est connu que des changements de composition entre les atomes d'hydrogènes et les atomes de deutérium entraînent des différences au niveau des temps de rétention. Ces différences entre les temps de rétention rendent la quantification très difficile étant donné que les peptides à comparer (quantifier) ne seront pas élués dans les mêmes fractions chromatographiques [Lill 2003]. Nous avons pu montrer qu'avec notre marqueur, le temps de rétention entre les peptides marqués avec le NEIAM-D0 et le NEIAM-D5 était inférieur à 5 secondes (figure 9).

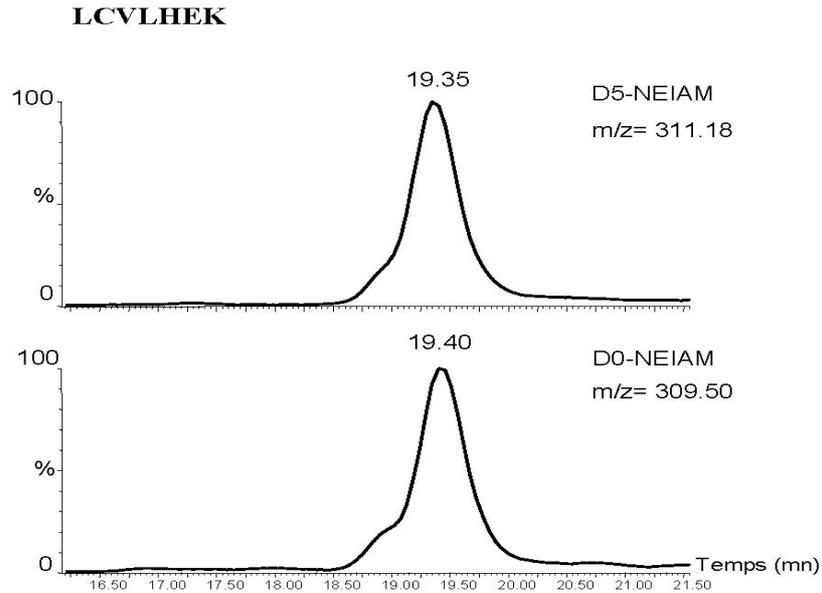


Figure 9 : Shift dans les temps de rétentions chromatographiques d'un peptide labellé au NEIAM D0 et au NEIAM D5. La différence des temps d'éluion est de 5 secondes.

De plus, les différents peptides marqués au NEIAM ne présentent pas de difficulté à se fragmenter, et les différents spectres MS/MS sont d'aussi bonne qualité que les spectres obtenus sur des peptides non marqués. Un exemple de spectre de fragmentation est présenté en figure 10.

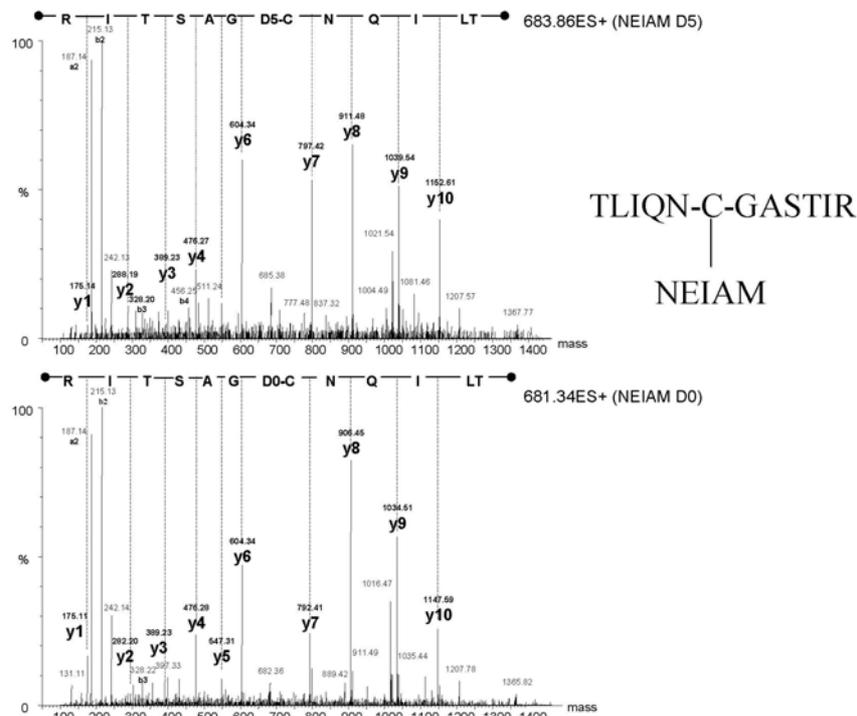


Figure 10 : Spectre de fragmentation MS/MS d'un peptide trypsinique issue d'une protéine chaperonne marqué soit au NEIAM D0 soit au NEIAM D5. La fragmentation ne pose aucun problème, et la cystéine alkylée est clairement identifiable.

On peut donc en conclure que le NEIAM est compatible avec l'analyse protéomique et peut donc être appliqué à la quantification de différents mélanges protéiques.

2.4) Application à des cas biologiques.

Ce marqueur a été utilisé pour quantifier les protéines issues de trois systèmes biologiques différents :

- Quantifier des protéines cytosoliques issues de cellules Hela différemment exprimées lors du stress oxydatif.
- Quantifier des protéines membranaires de deux souches d'*Acinetobacter baumannii*, l'une virulente et l'autre non.
- Quantifier les protéines impliquées dans la formation des complexes pré-ribosomiques obtenus à partir d'une souche sauvage et d'une souche mutante (délétion du gène NOG1).

2.4.1) Cellules Hela.

Dans cette étude nous avons déterminé les changements d'expression protéique entre des cellules Hela contrôles et des cellules qui ont été soumises à un stress oxydatif.

2.4.1.1) Préparation des cellules Hela et marquage des échantillons.

Les cellules Hela ont été mises en culture dans un milieu DMEM contenant 1mM de pyruvate, 10mM de tampon NaOH-Hepes pH 7.5 avec 10% de sérum de veau fœtal. Le stress oxydatif a été réalisé par l'ajout dans le milieu de culture de 150µM d'hydroperoxyde de tertio-butyl pendant deux heures. Les cellules ont ensuite été lavées et reprises dans un tampon de lyse (PMSF et CHAPS). Les gros débris cellulaires et les cellules non lysées ont été éliminés par centrifugation.

La réaction de marquage a été réalisée sur 1mg de mélange protéique dans un tampon NH_4HCO_3 pH8. Les protéines ont été réduites par l'ajout de 25mM de TCEP et alkylées par l'ajout de 25m de NEIAM D0 ou D5. Après une heure la réaction d'alkylation a été arrêtée par précipitation à l'acétone et les protéines ont été culottées et récupérées par centrifugation (10000g pendant 20 minutes). 300µg de chaque extrait protéique (contrôle et stress oxydatif) ont alors été mélangés dans un rapport 1/1. Ce mélange a ensuite été chargé sur gel SDS-Page. Après coloration le gel a été découpé tous les millimètres, les bandes ont été digérées et les peptides extraits et analysés par nanoLC-MS/MS.

2.4.1.2) Résultats

L'analyse nanoLC-MS/MS a permis d'identifier 230 protéines différentes parmi lesquelles 130 (57%) ont pu être quantifiées dont 15 présentaient un caractère basique, et n'auraient donc pas été détectées sur un gel 2D. Le calcul de la quantification a été réalisé en comparant les intensités des peptides marqués au D0 (contrôle) par rapport aux mêmes peptides marqués au D5 (stress oxydatif). Parmi les 130 protéines qui ont été quantifiées, environ 30 présentaient une variation d'intensité supérieure à 10%. Ces différentes protéines sont reportées dans le Tableau 3 et la figure 11.

SwissProt	Identification	Variation D5/D0
		Diminution après stress oxydatif
P30044	Peroxiredoxin 5	-2.94
Q00839	Heterogenous nuclear ribonucleoprotein U	-2.63
P49588	Alanyl t-RNA synthetase	-2.32
P04406	Glyceraldehyde 3 phosphate dehydrogenase (GADPH)	-1.96
Q92598	HSP 105 kDa	-1.96
P04645	60S ribosomal protein L30	-1.81
P12750	40S ribosomal protein S4	-1.81
P04720	Elongation factor 1-alpha	-1.61
P04083	Annexin I	-1.56
P13639	Elongation factor 2	-1.47
P00558	Phosphoglycerate kinase 1 (PGK1)	-1.42
P46782	40S ribosomal protein S5	-1.38
P00938	Triosephosphate isomerase	-1.38
P25388	Guanine nucleotide binding protein	-1.37
P23528	Cofilin	-1.29
P08578	Small nuclear ribonucleoprotein E	-1.25
P06733	Alpha-enolase	-1.21
P29312	14-3-3 protein zeta/delta	-1.20
P00338	Lactate dehydrogenase A chain	-1.20
Q00610	Clathrin heavy chain	-1.14
		Augmentation après stress oxydatif
O00410	Importin beta-3 subunit	+3.50
P05092	Peptidyl prolyl cis trans isomerase	+2.39
P49327	Fatty acid synthase	+2.22
P30041	Antioxidant protein 2	+1.84
P49736	DNA replication licensing factor MCM2	+1.79
Q9Y617	Phosphoserine aminotransferase	+1.45
P31947	14-3-3 protein sigma	+1.36
P13667	Protein disulfide isomerase A4 precursor	+1.28
P02570	Actin cytoplasmic 1	+1.20
Q06830	Peroxiredoxin 1	+1.10

Tableau 3 : Récapitulatif des protéines quantifiées par notre approche. La variation est calculée en fonction des intensités des peptides marqués au NEIAM D5 (stress oxydatif) par rapport aux peptides marqués au NEIAM D0 (contrôle). Un rapport négatif représente une diminution de l'expression protéique, et un rapport positif une augmentation.

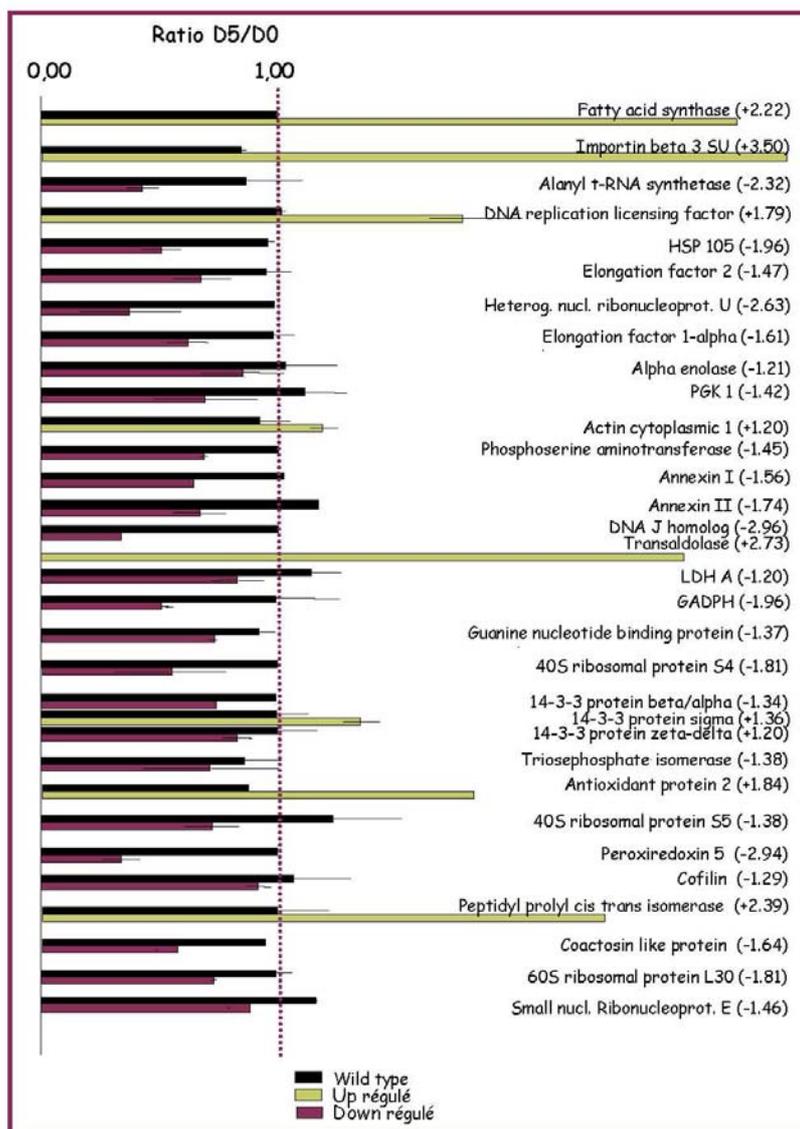


Figure 11 : Représentation schématique de l'augmentation et de la diminution de l'expression des protéines dans le cas d'un stress oxydatif.

2.4.2) Protéines membranaires issues de *Acinetobacter baumannii*.

Dans cette étude nous avons cherché à déterminer l'expression différentielle des protéines membranaires d'*Acinetobacter baumannii*, entre deux souches, l'une sensible à l'action des antibiotiques (A26) et l'autre résistante (ATCC).

2.4.2.1) Préparation des échantillons et marquage

Les préparations membranaires ont été obtenues par centrifugation sur gradient de sucrose comme décrit dans la publication de Siroy et al. [Siroy et al. 2005]. La réaction d'alkylation a été

réalisée sur 400µg d'extrait protéique, en présence d'une grande quantité de solvant organique (50% acétonitrile ou 60% méthanol ou 60% isopropanol...). Cette présence de solvant organique doit servir à solubiliser correctement le mélange protéique pour que la réaction d'alkylation puisse se faire. Nos différents tests n'ont pas permis d'obtenir une bonne solubilisation des protéines, mis à part des essais avec le DMSO. Malheureusement, ce solvant est difficile à éliminer et il nous a posé de nombreux problèmes lors de l'étape d'électrophorèse.

Le reste du protocole (gel d'électrophorèse, digestion...) est identique à celui décrit pour les cellules Hela. L'étude a porté sur les protéines issues de la membrane interne et de la membrane externe des bactéries.

2.4.2.2) Résultats

L'étude de protéines membranaires d'*Acinetobacter baumannii* a permis l'identification de plus de 400 protéines distinctes. Sur ces 400 protéines identifiées, seulement 66 ont pu être quantifiées (17%). Ceci est dû en grande partie aux peu de peptides identifiées pour chacune des protéines (hydrophobicité des protéines, problèmes de digestion et d'extraction des peptides hydrophobes).

Parmi les protéines qui ont été identifiées avec un marqueur, aucune n'était exprimée de façon différentielle entre les deux souches bactériennes. Bien que cette approche ne nous ait pas permis de quantifier beaucoup de protéines (problèmes de solubilisation des protéines lors de la réaction d'alkylation), notre approche a permis l'identification de 400 protéines différentes. On a donc un compromis entre l'identification des différentes protéines du mélange et la quantification de celles-ci.

2.4.3) Complexes pré-ribosomique.

Dans cette étude nous avons cherché à déterminer l'expression différentielle des protéines ribosomales impliquées dans la formation du complexe pré-ribosomique RLP24 soit pour une souche de levure sauvage, soit pour une souche mutante (délétion du gène NOG1 impliqué dans la synthèse des protéines ribosomales) [Kallstrom et al. 2003] [Jensen et al. 2003].

2.4.3.1) Préparation des échantillons et marquage

Les complexes protéiques ont été purifiés par précipitation au méthanol/chloroforme. La quantité de protéine disponible est très faible (15 µg par échantillon). Le protocole de marquage et de séparation sur gel est le même que celui utilisé pour l'étude des protéines Hela.

2.4.3.2) Résultats

106 protéines différentes impliquées dans la formation des complexes pré-ribosomiques ont été identifiées. Sur ces différentes protéines identifiées, moins d'une dizaine ont pu être quantifiées, et aucune ne présentait de variation d'expression. Le faible rendement de quantification est dû à la trop faible quantité de matériel biologique disponible. En effet, avec notre stratégie il faut obtenir des informations de séquence afin d'identifier les différentes protéines du mélange, mais il faut aussi arriver à digérer, extraire et enfin analyser un peptide qui comporte une cystéine marquée au NEIAM. Ceci est possible à partir du moment où les protéines sont présentes en suffisamment grande quantité, pour avoir la chance d'analyser un peptide cystéine.

2.5) Conclusions

Nous avons développé au sein de notre laboratoire une stratégie de marquage afin de quantifier des mélanges de protéines complexes, et ceci sans étape d'enrichissement des peptides marqués. Cette étape est remplacée par une électrophorèse 1D SDS-PAGE.

Nous avons pu montrer que le marqueur développé est tout à fait compatible avec les différentes étapes de l'analyse protéomique (gel, digestion, analyses...). Nous avons aussi montré la linéarité du marquage et ceci pour des ratios allant jusqu'à 10 fois. De plus, nous avons appliqué cette méthode de quantification à différents mélanges protéiques complexes. Bien que cette méthode de quantification ne soit que peu efficace vis-à-vis des protéines hydrophobes (problèmes lors de la réaction d'alkylation, digestion et extraction des peptides hydrophobes) et qu'elle nécessite une grande quantité de matériel (complexes pré-ribosomiques) elle est très efficace pour des mélanges protéiques complexes disponibles en grande quantité. De plus, étant donné que cette stratégie ne présente pas d'étape de chromatographie d'affinité cette approche est très adaptée pour l'étude de mélanges protéiques issus d'espèces dont le génome n'est que partiellement voire pas du tout séquencé. En effet, une étape de chromatographie d'affinité entraîne la perte de tous les peptides non marqués, ce qui entraîne des identifications avec très peu de peptides. Les identifications par homologie de séquence sont alors rendues presque impossible car pas suffisamment informatives.

2.6) Perspectives : développement d'un marqueur C13

Le problème des différents marqueurs qui utilisent le deutérium est que les formes légères et les formes lourde ne présentent pas le même comportement vis-à-vis de la chromatographie liquide [Zhang et al. 2002]. Les temps de rétention entre les différentes formes peuvent varier de plusieurs secondes. Cette différence dans les différents temps de rétention entraîne une difficulté quant à la réalisation d'une quantification précise. Afin de réduire complètement les différences de temps de

réretention lors de la chromatographie liquide, nous avons entrepris la synthèse de marqueurs utilisant l'incorporation de C13 à la place du deutérium [Hansen et al. 2003]. Dans cette optique, nous avons décidé de synthétiser un dérivé de la proline. En effet, il existe commercialement une proline enrichie en C13 (Figure 12).

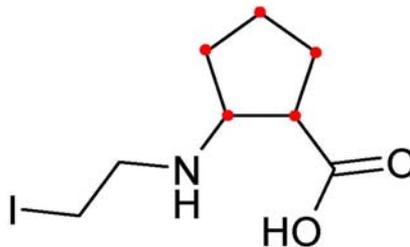


Figure 12 : Structure du dérivé de la proline. Les cinq atomes de carbone impliqué dans le cycle peuvent être enrichis en C¹³.

L'application de ce marqueur à différentes protéines modèles (BSA et RNase A) n'a pas donné satisfaction à ce jour. En effet, bien que la réaction d'alkylation ne pose pas de problème, l'extraction et l'analyse des peptides par spectrométrie de masse n'a permis d'identifier qu'un seul peptide présentant une cystéine (figure 13). De plus ce peptide qui possède une cystéine ne présente aucune modification (carbamidométhyl ou marqueur proline).

```

1  MKWVTFISLL  LLFSSAYSRG  VFRRDTHKSE  IAHRFKDLGE  EHFKGLVLIA
51  FSQYLQQCPF  DEHVKLVNEL  TEFAKTCVAD  ESHAGCEKSL  HTLFGDELCK
101 VASLRETYGD  MADCCEKQEP  ERNECFLSHK  DDSPDLPKLK  PDPNTLCDEF
151 KADEKKFWGK  YLYEIARRHP  YFYAPELLYY  ANKYNGVFQE  CCQAEDKGAC
201 LLPKIETMRE  KVLASSARQR  LRCASIQKFG  ERALKAWSVA  RLSQKFPKAE
251 FVEVTKLVTD  LTKVHKECCH  GDLLECADDR  ADLAKYICDN  QDTISSKLKE
301 CCDKPLLEKS  HCIAEVEKDA  IPENLPPLTA  DFAEDKDVCK  NYQEAKDAFL
351 GSFLYEYSRR  HPEYAVSVLL  RLAKYEYATL  EECCAADDPH  ACYSTVFDKL
401 KHLVDEPQNL  IKQNCDFEK  LGYGFQNAL  IVRYTRKVPQ  VSTPTLVEVS
451 RSLGKVGTRC  CTKPESERMC  CEDYLSLIL  NRLCVLHEKT  PVSEKVTKCC
501 TESLVNRRPC  FSALTPDETY  VPKAFDEKLF  TFHADICTLP  DTEKQIKKQT
551 ALVELLKHKP  KATEEQKTV  MENFVAFVVK  CCAADDKEAC  FAVEGPKLVV
601 STQTALA

```

Figure 13 : Séquence de la BSA. La BSA a été alkylée avec notre dérivé de proline, digéré et analysé par spectrométrie de masse nanoLC-MS/MS. Les peptides en vert ont été identifiés. Seulement un peptide contenant une cystéine a été sélectionné et séquencé, mais ce peptide n'avait pas réagi avec le marqueur.

Différentes variations ont été réalisées sur la base du marqueur proline (blocage de la fonction acide...) mais aucun de ces dérivés n'a permis d'obtenir des peptides marqués. Il faudra donc développer une autre voie de synthèse que la proline afin de réaliser un marqueur enrichi en C13.

Elaboration de nouveaux outils et de nouvelles stratégies analytiques appliqués à l'analyse des protéines membranaires par spectrométrie de masse.

Conclusions majeures :

- Nous avons démontré que les microparticules sont des structures naturellement enrichies en protéines issues de la membrane plasmique, et qu'elles en reflètent la composition.
- De plus, nous avons développé une nouvelle stratégie analytique qui combine les données ESI et MALDI pour augmenter le pourcentage de recouvrement des différentes protéines identifiées.

Le but de cette étude était de mettre au point un nouvel outil pour étudier (identifier) les protéines issues de la membrane plasmique, et de développer une approche permettant d'augmenter les informations de séquence (pourcentage de recouvrement, modifications post-traductionnelles...) des différentes protéines identifiées par spectrométrie de masse. Cette étude a été réalisée en collaboration avec le Dr Laurent Mauvieux qui a fourni les différentes préparations membranaires, et échantillons biologiques.

1) Intérêt des protéines membranaires

La membrane plasmique a longtemps été considérée comme une simple barrière entre le milieu extérieur et l'intérieur de la cellule. Il est maintenant communément reconnu que la membrane plasmique joue un rôle prépondérant dans un grand nombre de processus physiologiques ainsi que dans les processus de communication entre la cellule et le milieu environnant. Bien que la structure de base de la membrane plasmique (et de toute membrane biologique) soit constituée d'une bicouche phospholipidique, la plupart des fonctions spécifiques de la membrane plasmique sont portées par les protéines qui sont y enchâssées.

Les fonctions principales des protéines membranaires sont :

- Les échanges sélectifs de la matière (transporteurs membranaires, canaux ioniques et protéines impliquées dans l'exocytose et l'endocytose...)
- L'adhérence de la cellule à la matrice extracellulaire et aux cellules adjacentes (intégrines et cadhérines...)
- Les connexions avec le cytosquelette (vinculine associée avec les intégrines et la membrane plasmique...)
- La réception des signaux extracellulaires (récepteur du facteur de croissance EGF...)
- La transduction du signal par des molécules effectrices (récepteurs couplés aux protéines-G...)
- Diverses activités enzymatiques (protéine kinase C...)

Les protéines membranaires représentent 1/3 des protéines codées par le génome. Ceci suggère qu'elles représentent donc une classe importante des protéines exprimées dans la cellule [Wallin et von Heijne 1998]. Le fait que cette classe de protéines se situe à l'interface entre l'extérieur et l'intérieur de la cellule les rend aussi très importantes d'un point de vue pharmacologique. En effet, elles représentent des marqueurs protéiques faciles d'accès lorsque l'on souhaite travailler sur des cellules entières (cryométrie en flux par exemple) qui peuvent servir de marqueurs diagnostics voir prédictifs, ainsi que de cibles thérapeutiques [Jang et Hanash 2003].

Les protéines membranaires présentent des domaines transmembranaires à caractère très hydrophobe, qui permettent de traverser la bicouche lipidique. Ces domaines qui ont entre 20 et 30 résidus, peuvent former soit des hélices α (cas le plus fréquent chez les eucaryotes) soit des feuilletts β (surtout dans la membrane externe des bactéries). Ces structures à caractère hydrophobe rendent les protéines peu solubles dans les différents solvants aqueux utilisés dans les analyses par spectrométrie de masse, ce qui les rend donc très difficilement analysables dès qu'elles sont sorties de leur membrane d'origine. De plus, les préparations de protéines membranaires plasmiques, comme nous l'avons déjà mentionné dans la première partie de ce manuscrit, sont en règle générale très hétérogènes et contaminées par un grand nombre de protéines cytosoliques ou issues des autres compartiments cellulaires. Il est donc nécessaire de développer de nouveaux outils pour obtenir des préparations membranaires plasmiques facilement utilisables en spectrométrie de masse.

2) Développement d'un nouvel outil pour l'étude des protéines membranaires : le cas des microparticules.

D'une manière générale, il est très difficile d'obtenir une préparation de protéines spécifiques de la membrane plasmique d'une façon propre à partir de cellules eucaryotes. Ces préparations sont en règle générale très contaminées par les protéines issues des organites intracellulaires. Différentes méthodes ont été développées lors de ces dernières années afin d'enrichir au maximum ces préparations tel que la biotinylation des protéines de surface ou une succession de centrifugations sur gradient de sucrose [Klein et al. 2005] [Zhao et al. 2004] [Peirce et al. 2004], mais ces préparations restent malgré tout insatisfaisantes.

Dans le but de trouver un outil d'étude qui ne présente pas les inconvénients des préparations membranaires classiques, nous avons décidé de nous intéresser à l'étude de vésicules membranaires plasmiques naturelles : les microparticules.

2.1) Introduction aux microparticules

2.1.1) Génération des microparticules

En conditions de stress ou lorsqu'elles se divisent, les cellules émettent spontanément des vésicules appelées microparticules et qui proviennent directement de la membrane plasmique (Cf introduction).

Les microparticules ont une taille pouvant aller de 0.1 μ m jusqu'à 1 μ m de diamètre, ce qui les différencie des autres vésicules comme les exosomes qui sont des structures beaucoup plus petites. Les microparticules sont donc très facilement purifiées par simple centrifugation lente (de 15 à 20000g pendant 30 minutes).

Etant donné que ces structures sont directement émises par la membrane plasmique, nous avons émis l'hypothèse qu'elles étaient très fortement enrichies en protéines associées ou issues de la membrane plasmique, et pouvaient donc servir comme préparation de protéines membranaires plasmiques.

De plus, les cellules sous l'action de différentes substances chimiques telles que des antibiotiques comme l'actinomycineD1 (Act) ou des produits stimulant la division cellulaire comme la phyto-hémagglutinine (PHA) sont capables de produire des microparticules en grande quantité, ce qui rend l'étude protéomique plus aisée (préparation d'une quantité plus importante de protéines membranaires).

2.1.2) Stratégie employée pour l'étude des microparticules.

Etant donné le caractère hydrophobe des protéines membranaires, leur séparation à partir d'un mélange complexe a été réalisée sur gel SDS-PAGE 1D à gradient (8-15%). La révélation a été réalisée par coloration au bleu colloïdal. Le gel 1D a alors été découpé systématiquement tous les deux millimètres. Chacune des bandes a été lavée, et les protéines qu'elles contenaient ont été réduites (DTT) et alkylés (iodoacétamide) avant digestion à la trypsine. Les peptides générés ont été extraits et analysés par nanoLC-MS/MS (figure 1).

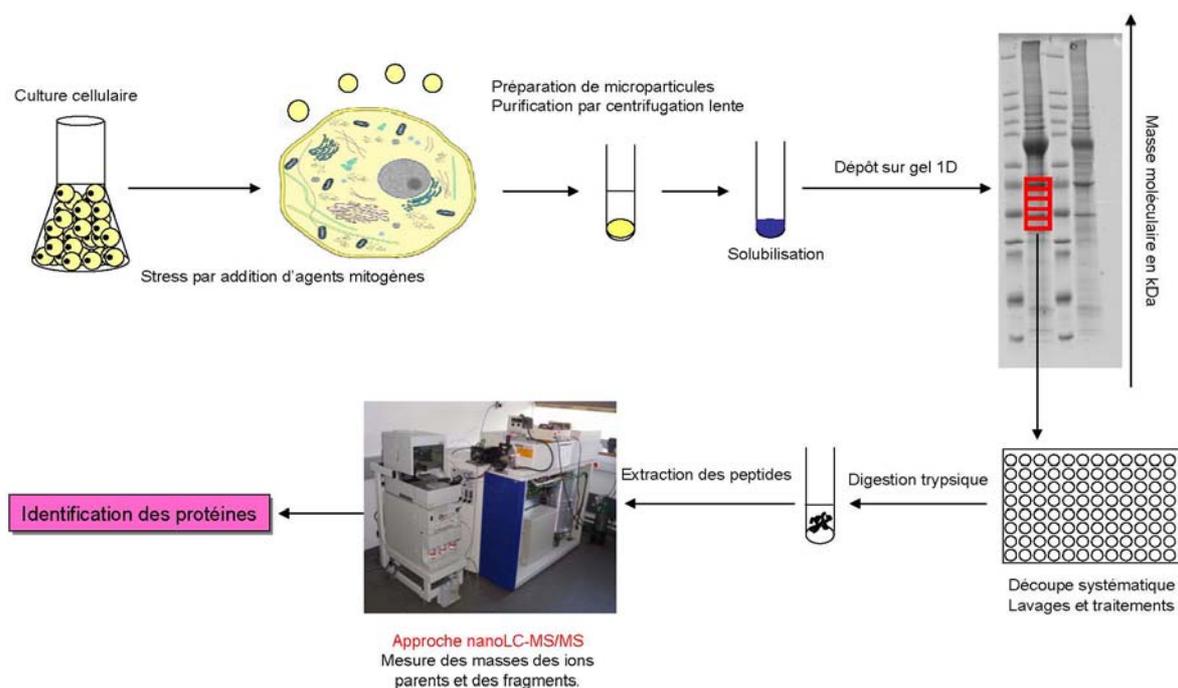


Figure 1 : Représentation schématique de l'approche protéomique pour l'étude des microparticules.

2.2) Validation de l'outil « microparticules »

Au moment où nous avons entrepris l'étude des microparticules, aucune analyse protéomique n'avait été réalisée et publiée dans la littérature sur ces structures. Afin de valider notre hypothèse selon laquelle les microparticules présentent un enrichissement naturel en protéines membranaires, nous avons décidé d'analyser le protéome de ces structures. Pour cela nous avons travaillé sur une lignée cellulaire de lymphocytes T (CEM). Les cellules ont été stimulées soit par de l'actinomycineD1 (apoptose) soit par de la phyto-hémagglutinine (PHA), et les protéines contenues dans les microparticules ont été analysées par spectrométrie de masse. Dans un deuxième temps, le protéome obtenu a été comparé au protéome obtenu à partir d'une préparation membranaire classique.

2.2.1) Composition protéique des microparticules

Nous avons déterminé la composition protéique des microparticules tous modes de stimulation confondus. Les différents extraits peptidiques obtenus par digestion trypsique des bandes de gel 1D des différents modes de stimulation ont été analysés par nanoLC-MS/MS sur un Q-TOF II (Waters). Les recherches dans la banque de données SwissProt des différentes listes de masses générées lors de l'analyse ont permis d'identifier de façon cumulée 390 protéines distinctes, dont 34% étaient des protéines issues ou associées à la membrane plasmique. De plus, parmi ces protéines membranaires, nous avons pu identifier des clusters de différenciation spécifiques du type cellulaire étudié tel que CD1c, CD98 ou CD99. Au total 17 clusters de différenciation ont été identifiés par cette approche. La localisation des protéines est basée sur les données des banques de données SwissProt et Bioinformatic harvester (<http://www.expasy.org/> ; <http://harvester.embl.de>).

2.2.1.1) Différence dans l'expression des protéines en fonction du mode de stimulation.

Afin de déterminer si la composition protéique des microparticules obtenues après stimulation par un agent mitogène et après stimulation par un antibiotique est différente, nous avons comparé entre eux les différents protéomes obtenus.

Très peu de différences ont été observées, et aucune classe ou famille de protéines ne semble différer d'un mode de préparation à l'autre. Plus de 70% des protéines identifiées étaient communes aux deux préparations. Les variations entre les deux protéomes portent principalement sur la composition en protéines ribosomales ou sur les protéines issues des différents organites de la cellule (golgi, réticulum et noyau). Ces protéines sont légèrement plus abondantes lors de l'utilisation de l'actinomycine D1. En effet, il est compréhensible que l'utilisation d'agents antimitotiques soit plus agressive que l'utilisation d'agents mitotiques (la PHA), et donc qu'une plus grande proportion de protéines cytoplasmiques soit identifiée.

2.2.1.2) Reproductibilité de la composition protéique des microparticules.

En vue de montrer la bonne reproductibilité des préparations de microparticules en termes de structures, nous avons répété cette approche (stimulation ; gel 1D ; analyse nanoLC-MS/MS) trois fois. Ces expériences se sont montrées très reproductibles. En comparant les différentes protéines identifiées pour chacune des analyses, plus de 70% d'entre elles étaient communes aux différentes analyses (si on ne tient pas compte des différentes formes de protéines ribosomales).

On peut donc en déduire que le contenu protéique de ces différentes structures est indépendant du type de stimulation et est spécifique du type cellulaire étudié.

2.2.2) Comparatif de la composition en protéines des microparticules par rapport à une préparation membranaire classique.

La composition en protéines membranaires des microparticules a été comparée à celle obtenue par l'analyse d'une préparation membranaire classique. La préparation classique a été obtenue à partir de la même lignée cellulaire (CEM T), en utilisant un protocole d'enrichissement des protéines de la membrane plasmique [Aleksandrov et al. 2001] (centrifugation à 100000g sur sucrose).

La fraction protéique a été séparée comme pour les microparticules sur un gel à gradient 1D SDS-PAGE. Après coloration, les bandes ont été découpées tous les deux millimètres. Les différentes protéines contenues dans les bandes de gel ont été digérées et identifiées par spectrométrie de masse. Cette approche a permis d'identifier environ 300 protéines distinctes. La proportion de protéines membranaires contenues dans cette préparation est de l'ordre de 30%, mais seulement 24% de ces protéines peuvent être localisées au niveau de la membrane plasmique (figure 2). Les microparticules présentent donc un enrichissement naturel en protéines issues de la membrane plasmique. De plus, comme on peut le voir sur le diagramme C de la figure 2, seulement 9 protéines membranaires plasmiques (6%) ont été identifiées spécifiquement dans la préparation classique, alors que 52% ont été identifiées spécifiquement dans les microparticules. Ces 9 protéines ne sont pas structurellement ou fonctionnellement reliées. On peut donc en conclure que les microparticules sont des structures enrichies en protéines issues de la membrane plasmique, et qu'elles comprennent à peu près toutes les protéines membranaires plasmiques que l'on pourrait trouver dans des préparations classiques, mais elles sont moins contaminées.

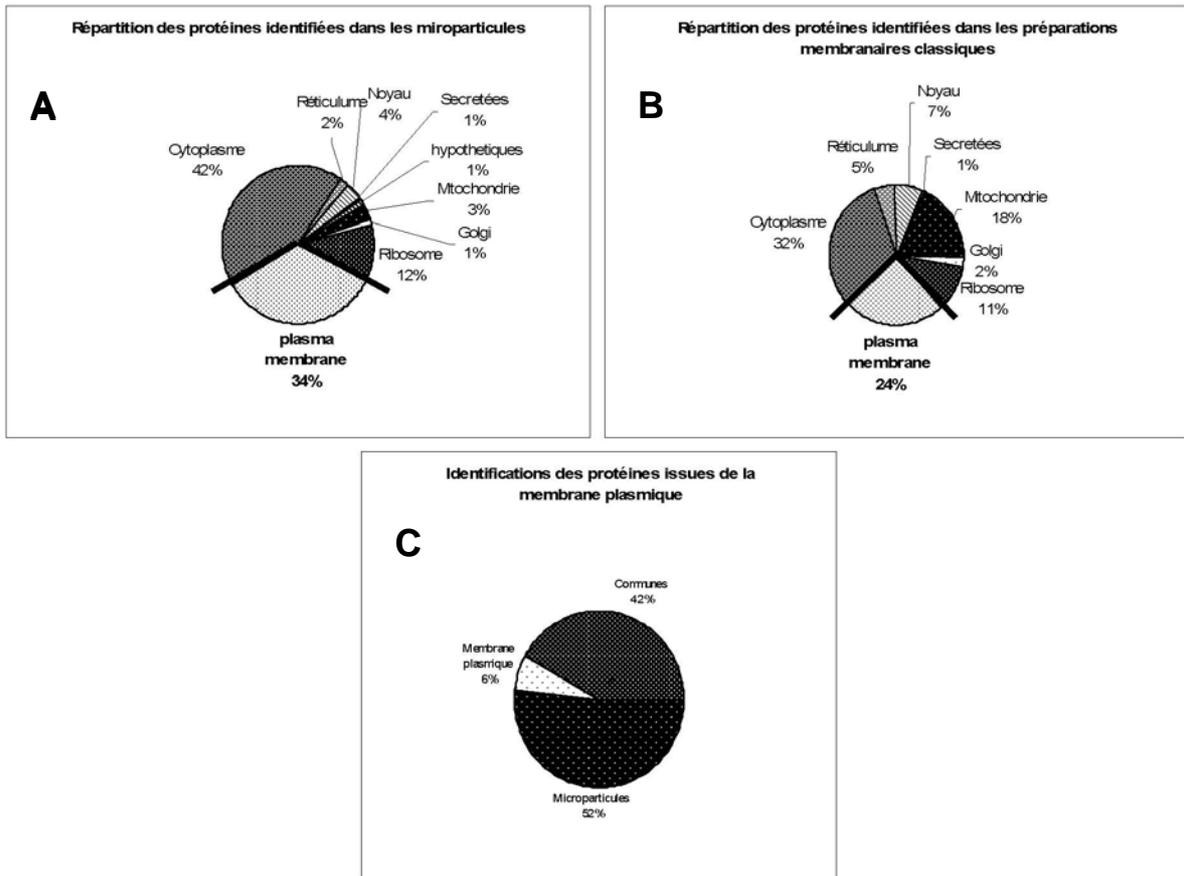


Figure 2 : Localisation des protéines identifiées dans les préparations de microparticules et dans la préparation membranaire « classique ». La figure C représente la proportion de protéines issues de la membrane plasmique identifiée de façon spécifique dans les différentes préparations.

2.3) Publication

Ces travaux ont fait l'objet d'une publication dans Proteomics.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization

Laurent Miguet, Karine Pacaud, Claire Felden, Bénédicte Hugel, M. Carmen Martinez, Jean-Marie Freyssinet, Raoul Herbrecht, Noelle Potier, Alain van Dorsselaer, Laurent Mauvieux

Proteomics, 2006, Vol. 6, Pages 153-171

Pages 153 à 171 :

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

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <http://www3.interscience.wiley.com/cgi-bin/fulltext/112190470/PDFSTART>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

2.4) conclusions

Lors de cette étude, nous avons montré le potentiel de l'approche protéomique des microparticules pour l'étude des protéines issues de la membrane plasmique. En effet, les protéines membranaires plasmiques s'y trouvent fortement enrichies, alors que très peu de contaminants issus des différents organites de la cellule sont présents (moins de 10%). Nous avons montré lors de cette étude que les microparticules sont des structures faciles d'accès (préparation facile et rapide) et dont la purification est réalisée par simples centrifugations. De plus, ce sont des structures potentiellement produites par un très grand nombre de type cellulaire. Enfin ces structures sont naturellement enrichies en protéines issues de la membrane plasmique et ne présentent que très peu de contaminants issus de des différents organites de la cellule.

Différents groupes ont commencé à s'intéresser à l'analyse du contenu protéique des microparticules [Banfi et al. 2005] [Garcia et al. 2005]. Les résultats présentés par ces différents groupes diffèrent de ceux que nous venons de présenter (proportion en protéines membranaires beaucoup plus faible). La raison majeure de cette différence pourrait être due à la vitesse de centrifugation lors de la préparation des microparticules. En effet, comme nous l'avons mentionné, les microparticules ont été culottées par une centrifugation lente (environ 15000g), alors que dans ces différents travaux, les microparticules ont été centrifugées avec des vitesses comprises entre 100 000 et 250 000 g. Ces différences de protocoles ont été récemment soulignées par Jin M. et al, mais sans les discuter [Jin et al. 2005]. Etant donnée la taille relativement importante des microparticules (jusqu'à 1µm de diamètre), nous pensons qu'une centrifugation lente ($\leq 20000g$) est cruciale pour enrichir spécifiquement la préparation en microparticules et pour ne pas culotter dans le même temps des structures plus petites comme des débris cellulaires ou des exosomes.

3) Combinaison des données MALDI et ESI, et reproductibilité de l'analyse nanoLC-MS/MS.

3.1) Introduction

L'analyse de mélanges protéiques complexes par spectrométrie de masse, préalablement séparés sur des gels d'électrophorèse et après digestion enzymatique, permet l'identification de plusieurs centaines de protéines différentes. Lorsque l'on travaille sur des protéines solubles, la stratégie classique est de les séparer sur un gel 2D qui est assez résolutif (plusieurs milliers de spots sont observables), avant de les analyser par MALDI-MS et/ou par nanoLC-ESI-MS/MS. Les identifications réalisées par cette approche se font généralement avec un grand nombre de peptides et donc un pourcentage de recouvrement important. En effet, les protéines sont assez bien individualisées dans le gel, et donc tous les peptides analysés proviennent d'un nombre très réduit de protéines. De plus, il est assez facile d'accéder aux différentes isoformes de la protéine qui sont soit directement visibles sur le gel 2D (modifications entraînant une différence dans le point isoélectrique de la protéine), soit par l'analyse et l'identification d'un peptide discriminant [Delalande et al. 2005].

Cette approche n'est toutefois pas applicable aux protéines membranaires qui ne peuvent pas être séparées sur un gel 2D. Nous sommes donc obligés de revenir au gel 1D qui est beaucoup moins résolutif (quelques centaines de bandes différentes). Il en résulte qu'il y a plusieurs dizaines de protéines différentes par bande de gel 1D, dont les différentes isoformes d'une même protéine (modifications post-traductionnelles) ou les différents variants protéiques (codés par des gènes paralogues). Cette surabondance de protéines entraîne une grande difficulté dans l'identification qui n'est obtenue qu'à partir de très peu de peptides différents pour chacune des protéines et donc une couverture de séquence très faible.

Notre objectif était donc ici de mettre en place une stratégie visant à améliorer l'identification et la caractérisation des protéines membranaires, en analysant un maximum de peptides pour augmenter la couverture de séquence et de passer de l'identification vers la caractérisation des protéines. Le matériel biologique utilisé lors de cette étude est en grande partie des préparations de microparticules issues de lymphocytes T, en dans une moindre mesure des préparations membranaires de cellules CacoII.

3.2) Les données MALDI

L'identification par empreinte peptidique massive a l'avantage de présenter une grande sensibilité et d'être extrêmement rapide du point de vue de l'analyse. Cette technique est très bien adaptée en ce qui concerne l'identification de protéines issues de gels 2D, mais ne l'est pas pour les gels 1D. En effet, la résolution de ces gels est mauvaise (une centaine de bandes par rapport aux milliers de spots des gels 2D). Cette mauvaise résolution entraîne la présence d'un grand nombre de protéines par bandes (plusieurs dizaines). Le mélange peptidique qui en résulte ne peut pas être

analysé par MALDI-MS. Cette approche ne permet pas d'identifier les protéines contenues dans la bande de gel, et ceci quel que soit le moteur de recherche utilisé figure 3. En effet, les analyses par spectrométrie de masse MALDI-MS sont soumises à des effets de compétition et de suppression de signal bien connues [Kratzer et al. 1998]. L'analyse par MALDI-MS et empreinte peptidique d'un mélange aussi complexe que le digest d'une bande de gel 1D génère plusieurs centaines de peptides différents qui ne peuvent pas tous être observés lors d'une seule analyse MALDI, ce qui entraîne la perte d'une grande quantité d'information.

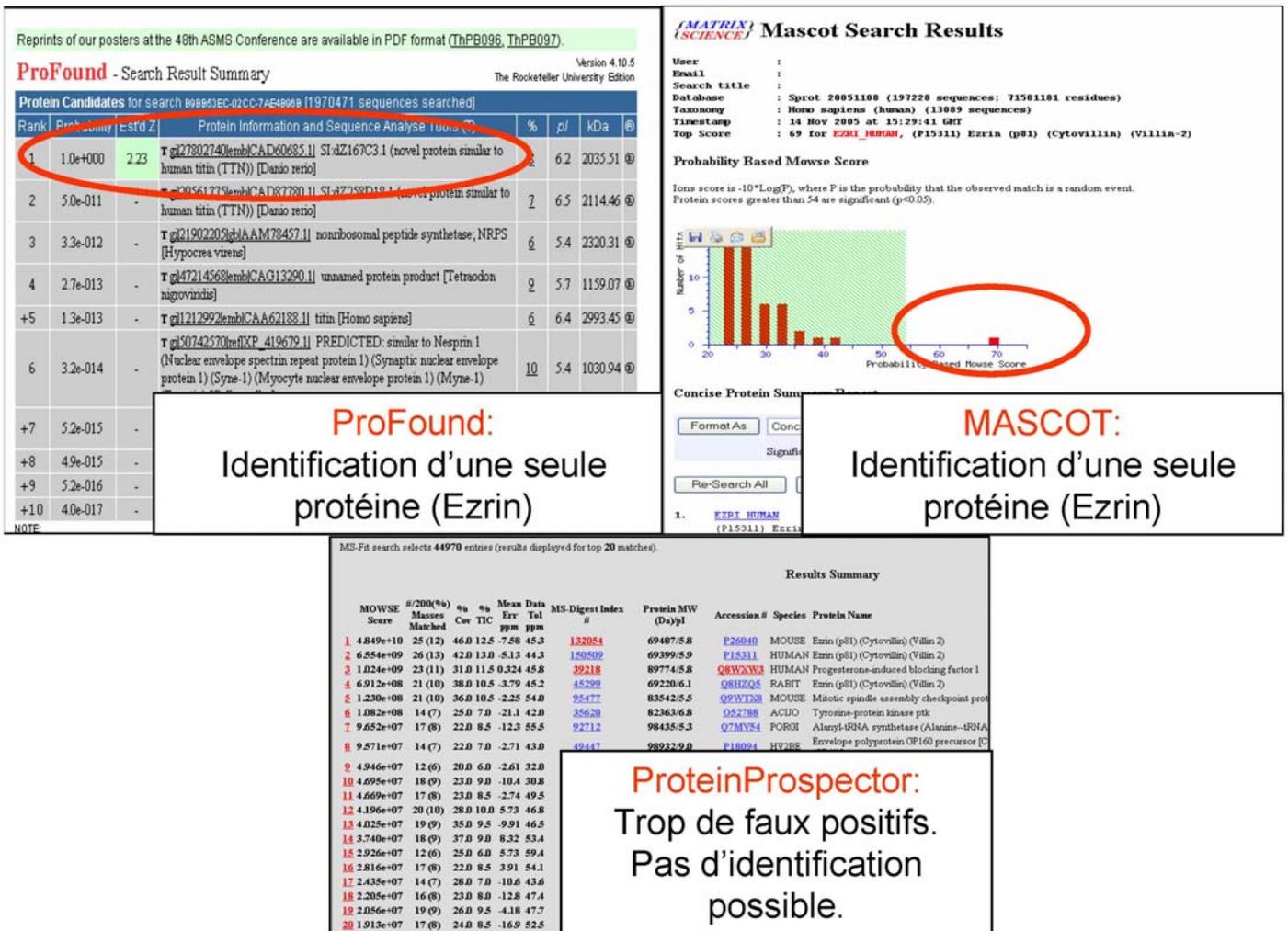


Figure 3 : Interrogation dans une banque de donnée via différents moteurs de recherche d'une liste de masse MALDI obtenue à partir de l'analyse d'une bande de gel 1D.

3.3) Les données nanoLC-MS/MS : un catalogue de protéines

Etant donné que l'analyse MALDI-MS n'est pas possible sur des mélanges peptidiques complexes issus de gel 1D, il est nécessaire de passer par une étape de simplification supplémentaire. En règle générale on introduit une étape de chromatographie liquide avant de réaliser l'analyse par spectrométrie de masse. Il en résulte que la stratégie la plus communément utilisée pour analyser des

mélanges complexes de protéines issus de gel 1D est l'analyse par nanoLC-MS/MS. Cette stratégie permet l'identification de plusieurs centaines de protéines différentes [Ahram et Springer 2004]. Cela est possible grâce à une étape de fractionnement du mélange peptidique par chromatographie liquide. Les différents peptides issus de la digestion sont élués au cours du gradient chromatographique et analysés par spectrométrie de masse en tandem. L'information de séquence obtenue lors de la fragmentation, ainsi que l'information de masse des peptides parents permet d'identifier un nombre très important de protéines différentes. Cependant les protéines identifiées à partir de cette stratégie le sont souvent avec peu de peptides.

Si on prend le même mélange peptique que celui analysé par MALDI-MS dans le point 3.2), on remarque que l'on est capable d'identifier beaucoup plus de protéines (tableau 1).

N° accession	Protéine	Masse	pl	Couverture	Nbre peptides
P15311	Ezrin (p81) (Cytovillin) (Villin-2)	69225	5.95	11%	8
P26038	Moesin (Membrane-organizing extension spike protein)	67647	6.09	6%	4
Q96QK1	Vacuolar protein sorting 35 (Vesicle protein sorting 35)	91649	5.32	7%	5
P08195	4F2 cell-surface antigen heavy chain (4F2hc)	57909	5.20	10%	4
P07900	Heat shock protein HSP 90-alpha (HSP 86)	84490	4.94	3%	2
P17858	6-phosphofructokinase, liver type	84863	7.08	4%	2
P35908	Keratin, type II cytoskeletal 2 epidermal	65825	8.07	2%	2
P13798	Acylamino-acid-releasing enzyme (EC 3.4.19.1)	81173	5.29	3%	2
Q15758	Neutral amino acid transporter B(0) (ATB(0))	56562	5.34	2%	1
Q5T9A4	ATPase family AAA domain containing protein 3B	72527	9.30	1%	2
P48960	CD97 antigen precursor (Leukocyte antigen CD97)	91781	6.40	1%	1
Q8IW75	Serpin A12 precursor (Visceral adipose-specific serpin)	47145	9.31	1%	1

Tableau 1 : Identifications, à partir d'une analyse nanoLC-MS/MS, des protéines contenues dans une bande de gel 1D. Les protéines sont identifiées avec peu de peptides.

Par cette approche nanoLC-MS/MS il est possible d'identifier à partir du même échantillon 12 protéines différentes, alors que l'analyse MALDI n'avait permis d'identifier qu'une seule protéine (l'ezrin cf figure 3 de ce chapitre). Par contre, le nombre de peptides permettant l'identification des ces protéines est vraiment très faible. On peut constater que 3 de ces protéines ont été identifiées avec seulement 1 peptide, et 1/3 des protéines identifiées le sont avec un ou deux peptides.

3.4) Reproductibilité de l'analyse nanoLC-MS/MS

L'approche que nous avons développée au sein de notre laboratoire pour analyser les protéines membranaires présente toutefois un caractère aléatoire intrinsèque. Les contraintes liées à l'instrument, les variations de temps de rétention au niveau de l'analyse chromatographique d'une analyse à l'autre et les temps morts nécessaires à l'acquisition électronique et aux calculs informatiques imposent à

l'utilisateur des choix de paramétrage du spectromètre de masse [Stewart et al. 2004] [Durr et al. 2004] (Figure 4).

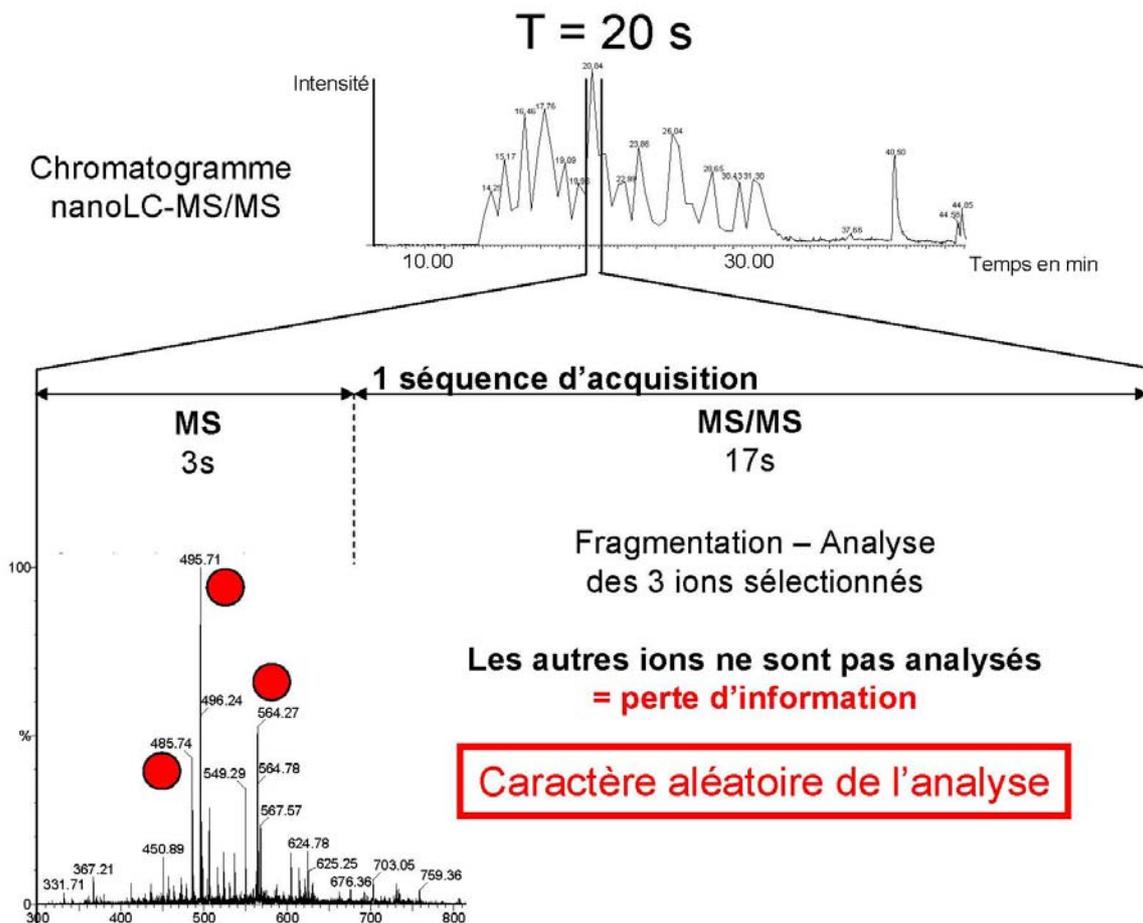


Figure 4 : Représentation schématique d'une séquence d'acquisition nanoLC-MS/MS sur un Q-TOF II (Waters). Lors de l'élution d'un pic chromatographique, plusieurs peptides sont coélus. Tous les peptides ne peuvent pas être sélectionnés et fractionnés. Il en résulte une perte d'information importante.

Lors d'une unique analyse nanoLC-MS/MS, le caractère aléatoire de l'analyse et la complexité de l'échantillon biologique, ne permettent d'avoir qu'une vision partielle de la composition protéique de l'échantillon. En effet, comme on peut le voir sur la figure 4, le chromatogramme obtenu lors d'une analyse est très complexe, et il en résulte qu'un pic chromatographique comporte beaucoup de peptides différents (spectre MS de la figure 4). Parmi ces masses peptidiques, seules les plus intenses vont être sélectionnées et fragmentées. Tous les autres ions ne sont pas sélectionnés. De plus, lors de l'analyse MS/MS, tous les ions qui continueront d'être élués par la chromatographie ne peuvent pas non plus être sélectionnés pour être fragmentés. Ce type d'analyse est donc sujet à une très grande

perte d'information. La question du nombre d'analyses à réaliser afin d'obtenir des informations de séquence sur un maximum de peptides et donc une image la plus complète possible du protéome doit donc être posée. Nous nous sommes par conséquent attachés à déterminer le nombre d'analyses nanoLC-MS/MS nécessaire pour identifier le maximum de protéines présentes dans l'échantillon. Cet aspect qui cherche à identifier toutes les protéines d'un échantillon est très important dans le cadre de la comparaison de différents protéomes puisqu'il ne faut dans ce cas « rater » aucune protéine afin de pouvoir affirmer qu'une protéine est exclusivement présente dans un échantillon (Cf chapitre III).

3.4.1) Stratégie expérimentale

Trois points ont été considérés :

- L'évaluation des effets mémoires possibles entre deux injections successives.
- La détermination du nombre d'analyses nécessaire pour identifier toutes les protéines présentes dans le mélange.
- Le problème posé par les effets d'ordres entre les analyses (calculer les $n!$ permutations possibles où n est le nombre d'analyses).

Pour cela, nous avons mis en place trois méthodes distinctes :

- Méthode 1 : Injections alternées entre l'échantillon et des blancs afin de déterminer les effets mémoires.
- Méthode 2 : Injections successives du même échantillon.
- Méthode 3 : Injections successives de différents échantillons.

Ces méthodes ont été appliquées à différents échantillons issus de trois gammes de masses de gel 1D. Les bandes découpées ont été lavées, et les protéines réduites et alkylées avant d'être digérées et analysées par spectrométrie de masse nanoLC-MS/MS. Ces différentes méthodes sont reportées dans la figure 5.

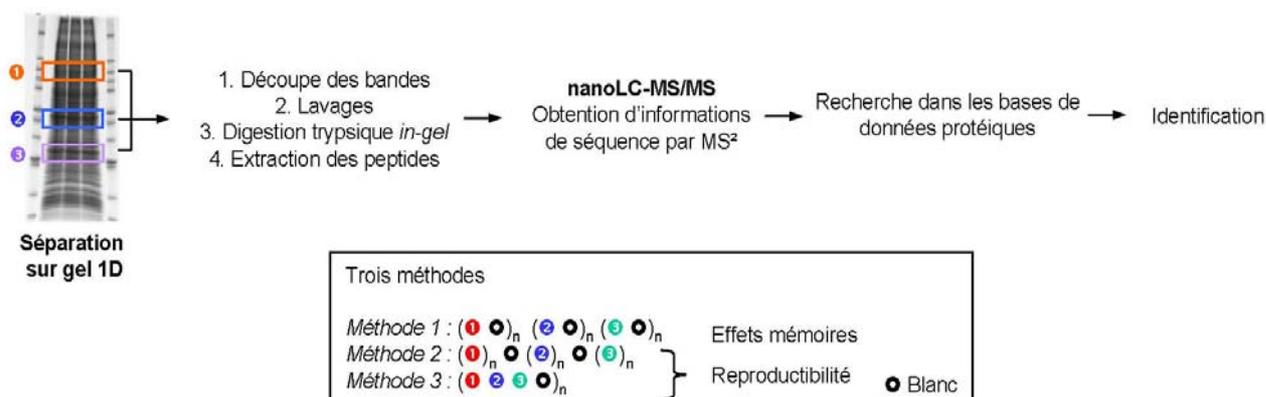


Figure 5 : Représentation de la stratégie expérimentale et des différentes méthodes.

3.4.2) Résultats

3.4.2.1) Effets mémoires

Nous avons vérifié par la méthode 1 (échantillon, blanc, échantillon...) qu'il n'y avait pas d'effet mémoire lors de nos différentes analyses. En effet, l'analyse des « blancs » entre deux injections n'a pas permis d'observer de peptides contaminant d'une analyse à l'autre.

3.4.2.2) Nombre d'analyses nécessaires pour identifier « toutes » les protéines de l'échantillon.

Le nombre d'analyses nécessaires correspond au nombre d'injections à effectuer pour que lors de la dernière analyse toutes les protéines qui y sont identifiées l'aient déjà été dans les analyses précédentes. Si plus aucune nouvelle protéine n'est identifiée, on peut considérer que « toutes » les protéines de l'échantillon ont alors été identifiées.

Dans un premier temps, nous avons réalisé trois injections successives du même échantillon avec les méthodes 2 et 3. Lors de la troisième injection 13% des protéines identifiées ne l'avaient pas été lors des 2 analyses précédentes. Donc trois injections successives ne sont pas suffisantes pour avoir une vision globale de l'échantillon protéique considéré.

Nous avons donc augmenté le nombre d'injections successives à 6. Après 5 analyses consécutives, l'information obtenue sur le contenu protéique de l'échantillon n'est pas significativement nouvelle. En effet, seuls 2% des protéines identifiées lors de la 6e analyse ne l'avaient pas été dans les précédentes. Par conséquent, nous admettons que toutes les protéines présentes dans l'échantillon sont identifiées par 6 analyses consécutives. On peut donc considérer raisonnablement que le nombre de

protéines identifiées à l'issu des 6 analyses représentent la totalité des protéines de l'échantillon, et donc 100% des protéines identifiables (figure 6).

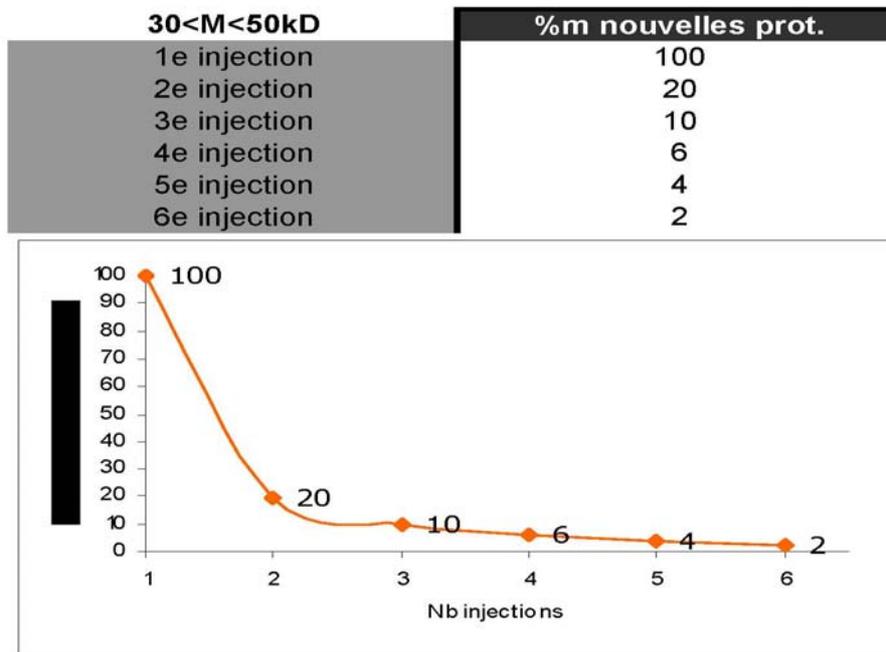


Figure 6 : Représentation du nombre de nouvelles protéines identifiées pour chaque analyse supplémentaire. Par exemple, lors de la troisième analyse, 10% des protéines identifiées ne l'avaient pas été lors des deux analyses précédentes. Lors de la sixième injection seulement 2% des protéines identifiées sont nouvelles.

3.4.2.3) Estimation du nombre minimum d'analyses à réaliser

En partant de l'hypothèse que 100% des protéines de l'échantillon sont identifiées au bout de 6 analyses, il est possible de déterminer la fraction du contenu protéique global de l'échantillon accessible en limitant le nombre d'analyses consécutives. Ainsi, on observe que 70% des protéines sont identifiées lors de la première analyse et que 90% le sont au bout de 3 (Figure 7).

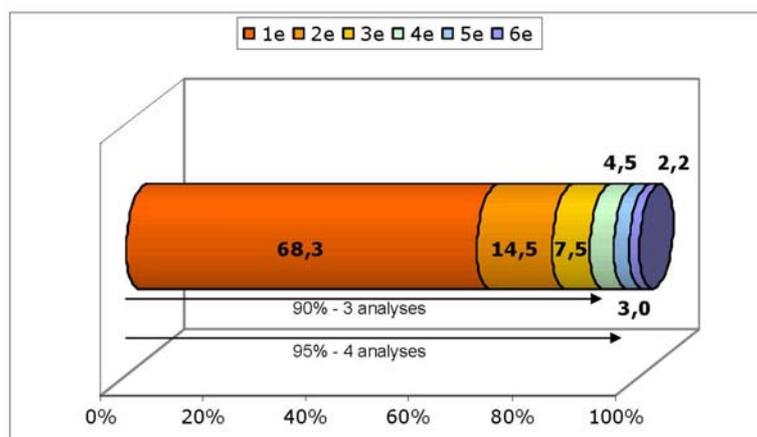


Figure 7 : Pourcentage de protéines contenues dans l'échantillon en fonction du nombre d'analyse.

Chaque essai supplémentaire, coûteux en échantillon et en temps, n'apporte que peu d'informations nouvelles. En effet, la quatrième analyse ne permet d'identifier que 5% e protéines supplémentaires.

3.5) Combinaison des données LC-MALDI et LC-ESI

3.5.1) LC-MALDI off-line

Comme nous l'avons mentionné précédemment les mélanges peptidiques issus de gel 1D sont trop complexes pour être analysés directement par MALDI-MS et empreinte peptidique massique. A l'instar d'une analyse nanoLC-MS/MS, nous avons donc ajouté une étape préalable de séparation des peptides (chromatographie liquide et collecte de fraction) avant analyse par MALDI-MS. Les mélanges peptidiques, issus de la digestion des bandes de gel 1D, ont été injectés sur une colonne de phase inverse C18 et les fractions ont été collectées systématiquement soit toutes les deux minutes, soit toutes les cinq minutes. Chacune de ces fractions a alors été analysée par MALDI-MS. Entre 10 et 20 spectres MALDI par bande de gel 1D ont ainsi été obtenus (Figure 8).

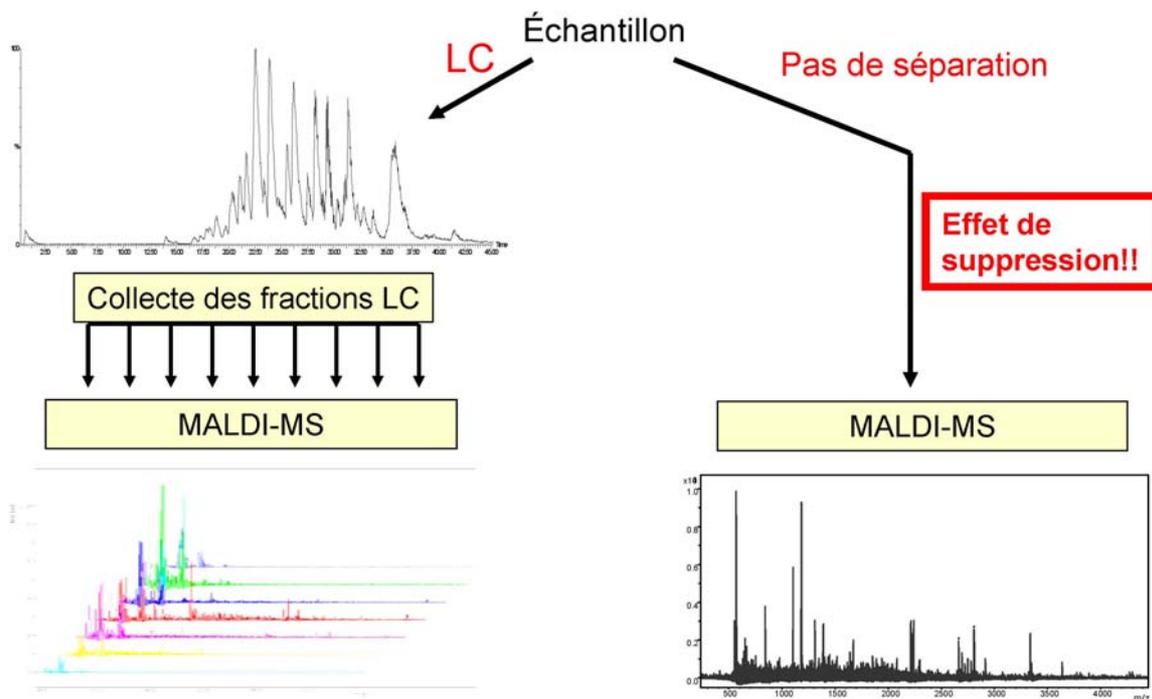


Figure 8 : Représentation schématique de l'approche LC-MALDI.

Cette étape permet de fractionner le mélange peptidique, ce qui va permettre de diminuer les différents effets de suppression de signal [Kratzer et al. 1998]. Un exemple de cette simplification est représenté sur les figure 9A et 9B.

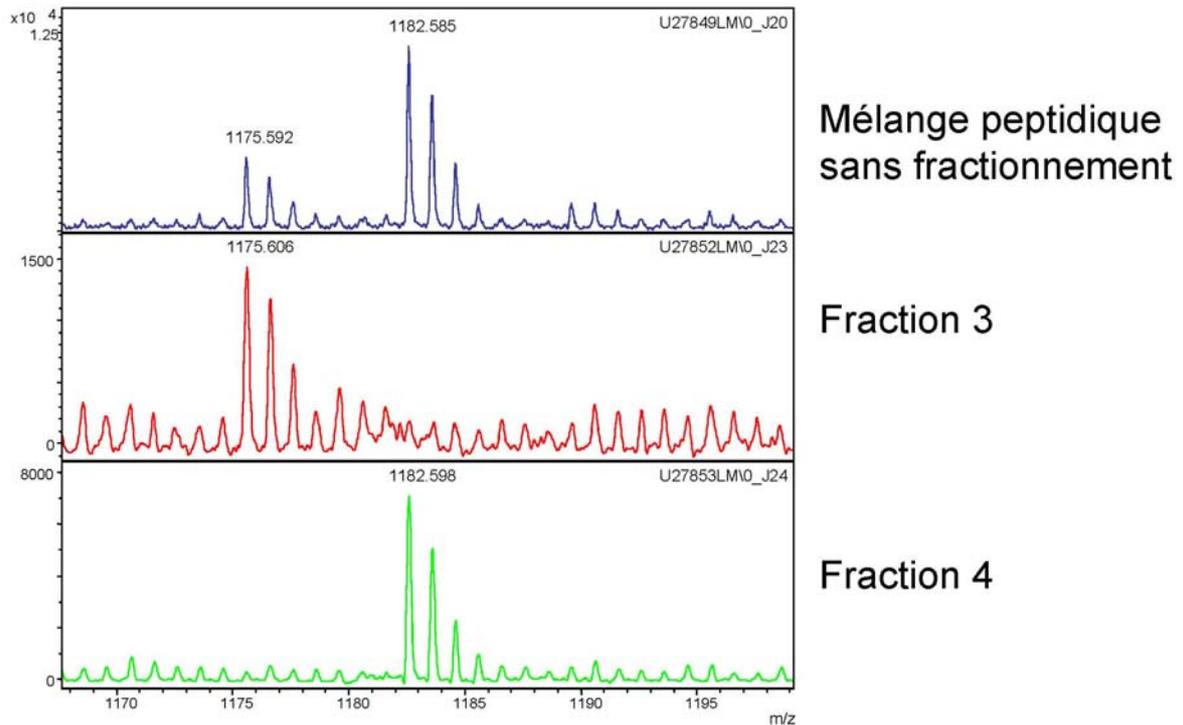


Figure 9A : Spectre MALDI obtenu après fractionnement chromatographique. Les différents peptides sont collectés dans différentes fractions.

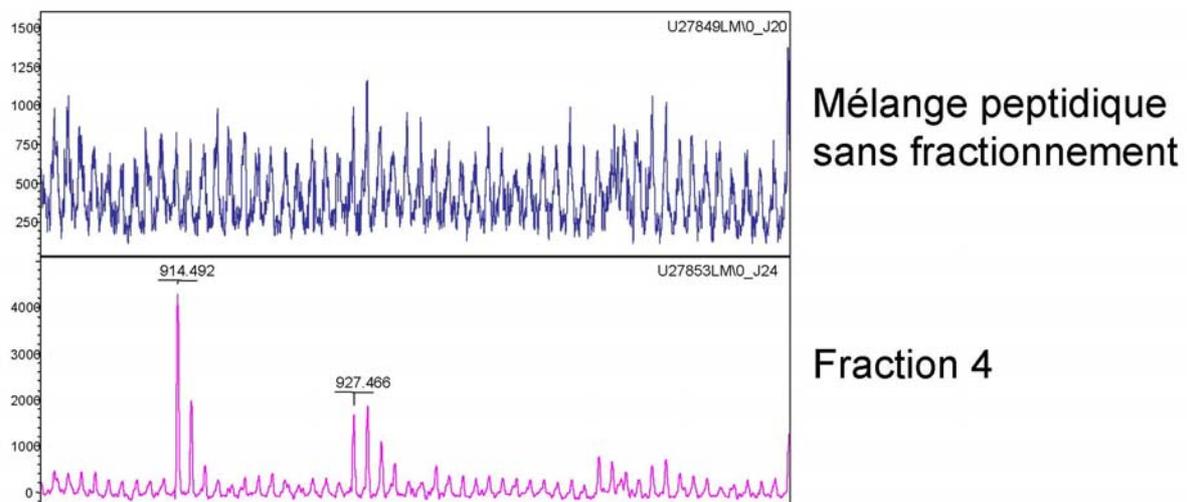


Figure 9B : La simplification de l'échantillon permet de diminuer l'effet de suppression de signal. L'analyse des spectres LC-MALDI permet d'identifier de « nouveaux » peptides.

On peut remarquer sur la figure 9A que les peptides qui étaient contenus dans les mélanges initiaux sont séparés entre les différentes fractions chromatographiques. En simplifiant ainsi le mélange peptidique entre les différentes fractions, on diminue de concert l'effet de suppression de signal. Cette diminution de l'effet de suppression de signal permet d'observer de nouveaux pics comme le montre

la figure 9B, où l'on voit deux nouvelles masses qui n'étaient pas visibles sur le spectre non fractionné.

3.5.2) Génération d'un spectre unique

Un problème posé par cette approche est la redondance des différentes masses observées dans les différents spectres. En effet, lors de la chromatographie liquide, les peptides peuvent être élués dans deux fractions successives. Un exemple de ce phénomène est reporté dans le tableau 2.

Fraction 3		Fraction 4		Fraction 5		Fraction 6	
m/z	Intensité.	m/z	Intensité.	m/z	Intensité.	m/z	Intensité.
1116.601	275.04	1110.572	471.45	1117.552	108.42	1108.562	908.64
1126.579	227.37	1112.563	293.54	1118.156	80.47	1111.561	839.69
1130.584	256.65	1115.575	588.12	1118.557	155.69	1117.565	1150.26
1133.573	329.15	1117.575	382.34	1120.544	99.30	1118.577	1442.84
1148.610	208.07	1120.558	2052.50	1136.571	186.66	1120.574	1706.19
1168.573	327.82	1126.563	361.93	1155.834	120.23	1126.573	1737.09
1170.598	225.73	1134.588	448.76	1166.554	92.99	1129.574	926.79
1175.645	1524.49	1136.573	704.28	1181.139	88.76	1149.577	812.84
1177.615	225.07	1148.585	732.38	1183.604	71.54	1153.582	1990.63
1179.619	410.39	1151.620	365.57	1195.565	61.95	1175.615	12046.31
1186.620	177.61	1153.571	1954.01	1223.992	55.98	1179.591	1145.48
1190.623	322.79	1155.548	349.59	1236.596	100.32	1182.606	14688.19
1192.636	233.04	1159.625	453.20	1243.542	63.35	1189.615	1898.71
1207.623	249.77	1170.640	852.53	1250.706	72.79	1190.614	1351.17
1216.627	134.80	1182.600	7497.94	1265.615	130.52	1194.593	1437.49
1233.650	341.45	1186.607	405.97	1271.634	68.13	1198.602	792.23
1235.637	296.62	1190.624	507.18	1277.678	233.97	1204.610	1950.08
1244.663	218.38	1194.617	362.21	1290.085	79.58	1219.605	1051.65
1248.626	354.71	1203.651	1342.87	1300.728	73.45	1230.668	938.24
1255.684	270.26	1205.610	401.77	1312.714	112.11	1233.637	1680.48
1263.637	199.07	1217.602	437.57	1320.731	100.63	1236.638	957.52
1269.653	262.62	1220.620	429.89	1336.678	58.58	1248.617	2086.23
1286.659	183.95	1224.648	293.57	1345.694	117.09	1250.637	945.01
1299.659	250.73	1226.631	419.34	1361.781	88.93	1252.639	732.79
1311.648	166.80	1230.699	357.27	1371.837	104.69	1254.645	856.00
1314.644	131.39	1231.682	594.95	1395.689	79.27	1264.633	1046.28
1332.694	315.29	1241.642	516.59	1400.664	148.12	1265.633	1395.10
1334.692	221.57	1251.651	627.73	1405.782	68.30	1276.666	1174.40
1336.670	183.17	1253.631	618.56	1416.695	70.26	1278.666	811.79
1344.692	242.97	1265.631	932.69	1422.761	67.46	1280.626	996.15
1366.656	132.78	1270.611	300.85	1428.697	104.58	1283.615	1263.09
1371.663	228.79	1273.655	381.61	1435.770	428.61	1303.677	2538.47
1394.710	181.08	1276.650	349.82	1447.811	217.99	1308.671	1737.63
1403.759	150.27	1283.607	759.59	1453.966	90.00	1310.702	5166.80
1412.760	283.89	1286.687	330.82	1458.802	114.35	1317.689	912.53
1417.696	240.28	1292.672	299.70	1467.821	55.92	1333.675	1128.34
1428.743	216.42	1310.698	2079.23	1473.798	128.16	1348.685	2184.37
1435.233	114.40	1314.709	338.16	1495.744	51.22	1354.665	910.23

1446.732	145.59	1327.704	571.24	1500.728	54.16	1371.798	1955.30
1452.745	147.34	1348.680	536.62	1505.776	106.57	1380.733	1153.42
1455.770	179.15	1353.688	362.38	1511.818	114.50	1386.767	4877.34
1459.793	128.36	1355.670	462.01	1522.098	51.45	1400.693	1506.76
1473.800	216.33	1366.728	340.20	1581.955	62.34	1412.741	1725.03
1484.702	143.26	1380.715	501.63	1607.053	52.76	1417.774	658.51
1488.818	771.28	1383.705	476.28	1614.317	52.63	1428.754	1427.67
1498.749	154.73	1386.755	1608.63	1625.877	66.49	1430.725	1221.95
1509.722	649.56	1393.729	786.98	1652.772	84.87	1435.774	2773.03
1512.719	253.69	1412.725	480.50	1669.818	203.78	1445.791	1446.16

Tableau 2 : Listes de masses obtenues par analyse MALDI des différentes fractions chromatographiques. Les masses surlignées en couleurs représentent des peptides qui sont présent dans deux fractions chromatographiques successives.

On peut remarquer sur ce tableau un certain nombre de masses redondantes entre les différentes analyses issues de fractions chromatographiques successives. Cette redondance rend difficile l'utilisation des données. Il est nécessaire de diminuer cette redondance pour faciliter l'utilisation et la compréhension des données.

Pour cela, en partenariat avec la société Bruker, nous avons développée une approche qui consiste en la création d'un spectre unique non redondant à partir de tous les spectres MALDI générés à partir des différentes fractions LC. Le logiciel développé par Bruker Daltonic (WarpLC™) permet de combiner les différents spectres MALDI. Trois paramètres sont à prendre en compte :

- La précision en masse des différents pics
- Le nombre de fractions successives dans lequel une même masse correspond au même peptide
- Le rapport signal sur bruit

On entend par précision en masse des différents pics l'erreur tolérée (en ppm) pour considérer que deux pics observés sur deux spectres consécutifs proviennent ou non du même peptide. Dans notre étude cette valeur était de 20ppm (Figure 10). Pour considérer qu'un pic n'était pas du bruit de fond, le rapport signal sur bruit devait être supérieur à 10.

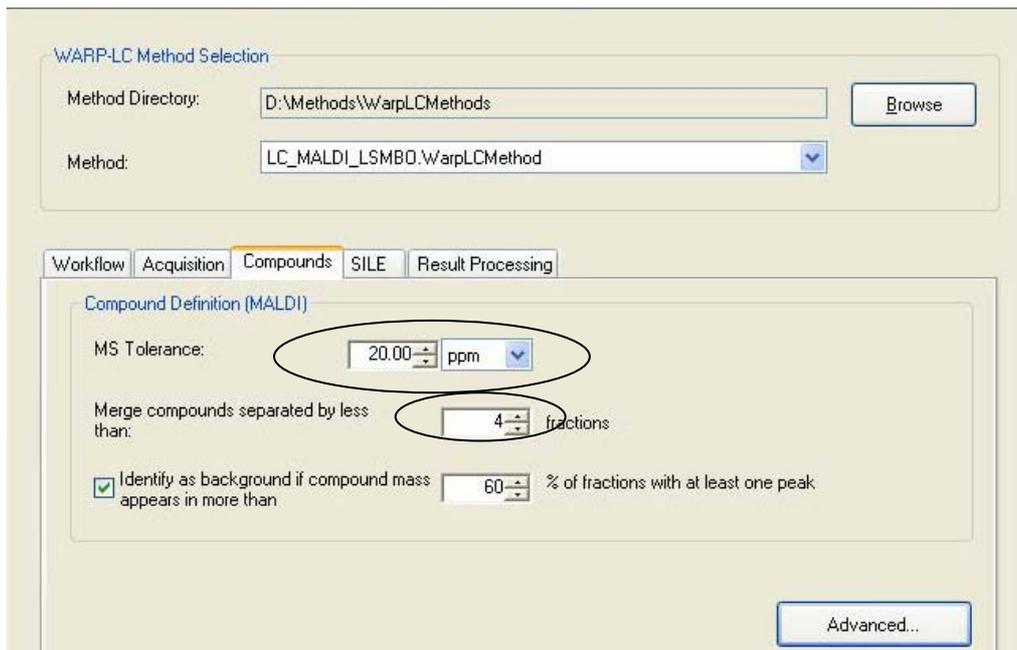


Figure 10 : Fenêtre de sélection des paramètres du logiciel WARP-LC.

Nous avons donc pu reconstituer un spectre unique, et donc une liste de masses observées unique et non redondante (figure 11) :

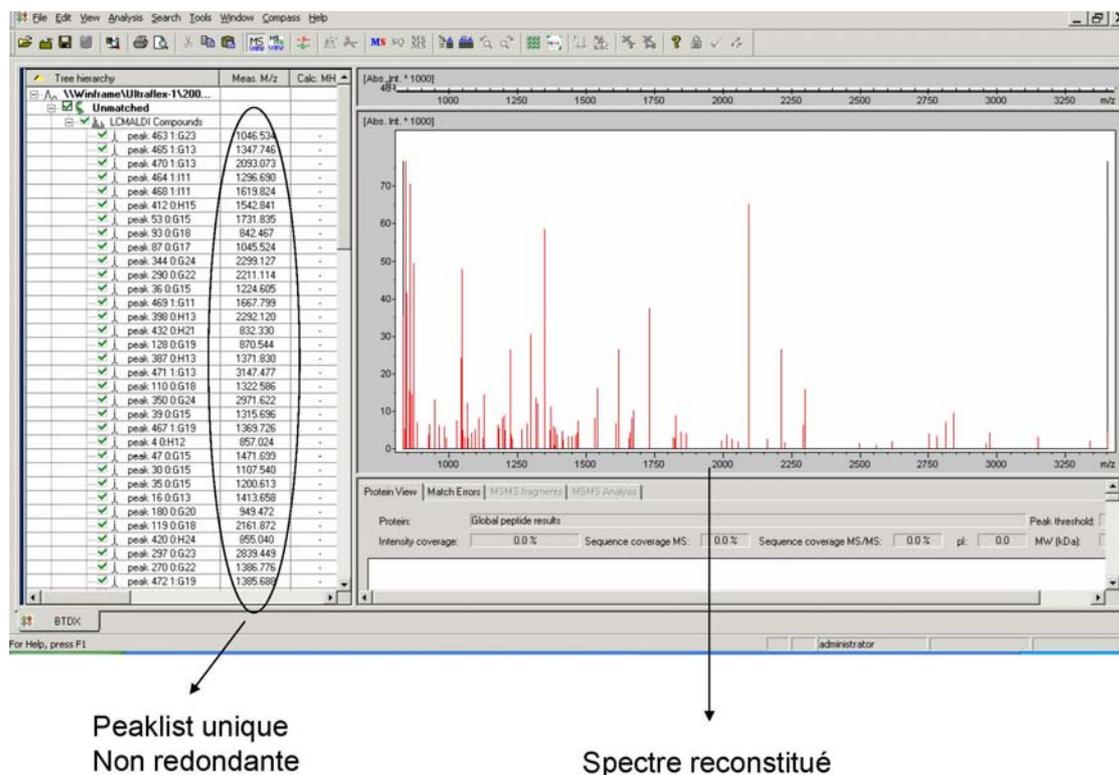


Figure 11 : Reconstitution d'un spectre unique à partir des différentes fractions chromatographiques.

Les différentes masses non redondantes obtenues par cette approche (plusieurs centaines) ne nous ont malgré tout pas permis d'identifier des protéines par empreinte peptidique massique. En effet, un des

critères d'identification des moteurs de recherche par empreinte peptidique massique, dépend du nombre de peptides correspondant à la protéine par rapport au nombre total de peptides soumis lors de l'interrogation. Il en résulte que des listes comprenant plusieurs centaines de masses différentes ne permettent pas d'identifier les différentes protéines du mélange. Par contre nous avons pu générer une liste de masses peptidiques extrêmement riche en informations provenant des différentes protéines de l'échantillon. Ces données peuvent être exploitées pour augmenter le pourcentage de recouvrement des différentes protéines identifiées par nanoLC-MS/MS.

3.5.3) Combinaison des données MALDI et ESI

L'objectif de cette approche n'est pas d'augmenter le nombre de protéines identifiées comme décrit par Bodnar et al et Zhang et al [Bodnar et al. 2003] [Zhang et al. 2004], qui comparent les résultats obtenus par LC-MALDI-MS/MS et par LC-ESI-MS/MS, mais d'augmenter le pourcentage de recouvrement des différentes protéines identifiées par nanoLC-MS/MS.

Pour cela, nous avons mis en place une nouvelle stratégie qui, en combinant les données LC-MALDI et nanoLC-MS/MS, permet de tirer partie de la multitude d'information contenue dans les listes de masses issues du WARP-LC. Les différentes protéines identifiées lors des analyses nanoLC-MS/MS ont donc été digérées *in silico* (<http://www.expasy.org/cgi-bin/peptide-mass>), et les différentes masses peptidiques théoriques ont été comparées à la liste de masse générée par les différentes analyses MALDI-MS. Les masses observées par LC-MALDI ont été considérées comme provenant des peptides lorsqu'ils présentaient une précision en masse meilleure que 20 ppm, et avec un rapport signal sur bruit supérieur à 5.

Cette approche a été appliquée à l'identification des différentes protéines contenues dans différentes bandes de gel 1D de microparticules issues de lymphocytes T. Les identifications des protéines ont été réalisées par nanoLC-MS/MS sur un Q-TOF II (Waters). Les résultats obtenus sur l'analyse d'une bande de gel 1D sont reportés dans le tableau 3.

N° accession	Protéine	Nbr peptides identifiés par LC-MS/MS	% de recouvrement	Nbr de peptides identifiés par LC-MALDI	Combinaison des données MALDI et ESI	
					Nbr de peptides différents	% de recouvrement
P08238	Heat shock protein HSP 90-beta	25	31%	20	30	47%
P07900	Heat shock protein HSP 90-alpha	16	21%	27	33	55%
P46940	Ras GTPase-activating-like protein	7	5%	22	27	22%
Q01813	6-phosphofruktokinase type C	4	5%	7	11	19%
P08195	4F2 cell-surface antigen	4	9%	10	11	30%
P23921	Ribonucleoside-diphosphate reductase	3	4%	13	14	20%
P26639	Threonyl-tRNA synthetase	3	4%	13	15	21%
P02786	Transferrin receptor protein 1 (CD71 antigen)	2	2%	14	16	29%
P16070	CD44 antigen precursor	1	1%	3	3	4%
Q15758	Neutral amino acid transporter B(0)	1	2%	3	4	9%

Tableau 3 : Augmentation des pourcentages de recouvrement des différentes protéines par notre approche LC-MALDI.

On peut remarquer que pour toutes les protéines identifiées par nanoLC-MS/MS, un nombre supplémentaire de peptides correspondant à ces protéines a pu être obtenu par la combinaison des données issues des deux approches. Le pourcentage de recouvrement des protéines a donc été largement augmenté.

Afin de valider cette approche, nous avons tenté de déterminer un taux de faux positifs (peptides injustement attribués à la protéine pour en augmenter le pourcentage de recouvrement). Ces faux positifs sont générés par le fait que plusieurs protéines peuvent produire des peptides tryptiques de masses similaires (tolérance de 20 ppm). Etant donné que cette approche ne permet pas d'avoir d'information de séquence, il nous est donc impossible de discriminer ces différents peptides.

Pour déterminer le taux de faux positifs, nous avons pris trois protéines de masse moléculaire différentes qui n'ont pas été identifiées lors de l'analyse nanoLC-MS/MS (la tubuline à environ 100kDa, l'actine à 40kDa et la myosine à 20kDa) que nous avons digérées *in silico* et soumis à la même comparaison que dans notre exemple. La myosine n'a présenté aucun peptide théorique, l'actine un seul et la tubuline 4. On peut donc estimer un taux de faux positifs moyen entre 1 et 2 peptides par protéines. Dans les différents cas étudiés, ce taux de faux positif est très faible par rapport au nombre moyen de peptides que cette approche nous permet d'identifier.

3.6) Conclusions et perspectives

Nous venons de présenter dans cette partie différentes améliorations du point de vue de l'approche nanoLC-MS/MS en déterminant la reproductibilité des différentes analyses. Nous avons montré qu'il était nécessaire d'injecter 6 fois le même échantillon pour avoir une vue à peu près globale du contenu protéique d'un échantillon complexe issu d'une bande de gel 1D, et que trois injections successives permettent d'obtenir environ 90% des informations contenues dans un échantillon. De plus, nous avons montré au cours de cette étude l'importance de cumuler les données MALDI et les données ESI afin de mieux caractériser les différentes protéines.

Les perspectives sont de réaliser un split après la colonne chromatographique pour pouvoir analyser directement les fractions par ESI-MS/MS et MALDI-MS. De plus, il sera intéressant de réaliser des analyses MALDI-MS/MS afin d'avoir des informations de séquence des différents peptides présents dans l'échantillon.

3.7) Publication

Une publication sur ces différents résultats est en cours de préparation.

Microparticles: a new tool for plasma membrane sub-cellular proteomic.

Authors:

Laurent Miguet¹, Sarah Sanglier¹, Christine Schaeffer¹, Noelle Potier¹, Laurent Mauvieux², Alain Van Dorsselaer^{1}.*

¹ Laboratoire de Spectrométrie de Masse Bio-Organique, ECPM, UMR/CNRS 7178, Institut Pluridisciplinaire Hubert CURIE, Université Louis Pasteur, Strasbourg, France.

² Institut d'Hématologie et d'Immunologie, Faculté de Médecine, Université Louis Pasteur and Laboratoire d'Hématologie, Hôpital Hautepierre, Strasbourg, France

Running title: Microparticles sub-proteome

* To whom correspondence should be addressed:

Dr. Alain Van Dorsselaer Laboratoire de Spectrométrie de Masse Bio-Organique, ECPM, UMR/CNRS 7178, Institut Pluridisciplinaire Hubert CURIE, Université Louis Pasteur, 25 rue Becquerel 67087 Strasbourg Cedex 2 France

E-mail: vandors@chimie.u-strasbg.fr

Fax: +33-390-242781

Keywords: Microparticles, proteomic, plasma membrane proteins, LC-MALDI

Abbreviations:

ActinomycinD: ActD; Microparticles: MPs; cluster of differentiation: CD; phytohemagglutinin: PHA

Table of contents

Chapter 1: Introduction

Chapter 2: Microparticles

Chapter 3: Strategy used for microparticles proteomics analysis

3.1) Presentation of the strategy

3.2) Technical details.

Chapter 4: Validation of the microparticles as a new tool for plasma membrane preparation

Chapter 5: Perspective

Chapter 6: Bibliographie

1) Introduction:

Membranes are critical components of cellular structure. It has been reported that plasma membrane proteins represent about 30% of the all cellular proteins (Wallin and von Heijne 1998). Even if plasma membrane was considered for a long time as a simple biological barrier between the cytosol of the cell and the extra-cellular environment, these membrane proteins have been demonstrated to play a crucial role in the different fundamental biological processes as exchange of component or signal transduction. Also, more than half of all the predicted pharmacological drug targets are predicted to be localized to the plasma membrane (Jang and Hanash 2003). Plasma membrane can than be clearly considered as a sub-cellular compartment of first interest in regard to different diagnosis and/or therapeutic target proteins. Indeed, for each membrane protein there is potentially a specific antibody which can be used for the diagnosis of several pathologies and also for treatment using armed antibodies (Harris 2004). Therefore, the proteomic analysis of plasma membrane proteins is of first importance. Despite the importance of plasma membrane proteins, there is less understanding in this class of proteins due to the difficulty to obtain enriched plasma membrane proteins preparation from eukaryotic cells. Until now many different strategies have been applied but are still laborious and imperfect. In example of these different approaches, biotinylation (Zhao et al. 2004) (Peirce et al. 2004), silica coated (Rahbar and Fenselau 2004), partition phase repartition (Qoronfleh et al. 2003) or partial tryptic surface digestion have been tested, but they still remain unsatisfactory. Although some studies of plasma membrane proteins using 2-DE have been reported (Luche et al. 2003) (Galeva and Altermann 2002), the separation of such hydrophobic proteins have been often poor (Santoni et al. 2000). We describe here a new strategy in order to increase the proportions of plasma membrane proteins identified with the highest possible coverage percentage in the different membrane proteins preparations analyzed by mass spectrometry. We will focus on microparticles as a source of membrane proteins.

2) Microparticles

Each of the two leaflets of the plasma membrane bilayer has a specific lipid composition. Aminophospholipids (phosphatidylserine and phosphatidylethanolamine) are specifically segregated in the inner layer of the membrane, whereas the phosphatidylcholine and the sphingomyelin are enriched in the external leaflet (Beyers et al. 1998). When cells are submitted to various stress conditions as mitogenic activations or apoptosis, the constitutive asymmetry between the inner and the outer leaflet of the plasma membrane is disrupted. The major changes in the plasma membrane constitution will be the delocalization of the phosphatidylserine to the outer leaflet and an augmentation of the concentration in ion Ca^{2+} in the cytoplasm. Such changes are going to disrupt the organisation of the cytoskeleton and drive to a blebbing of the plasma membrane and release microvesicles which are named microparticles (MPs) (figure 2.1)(Hugel et al. 2005) (Miguet et al. 2005).

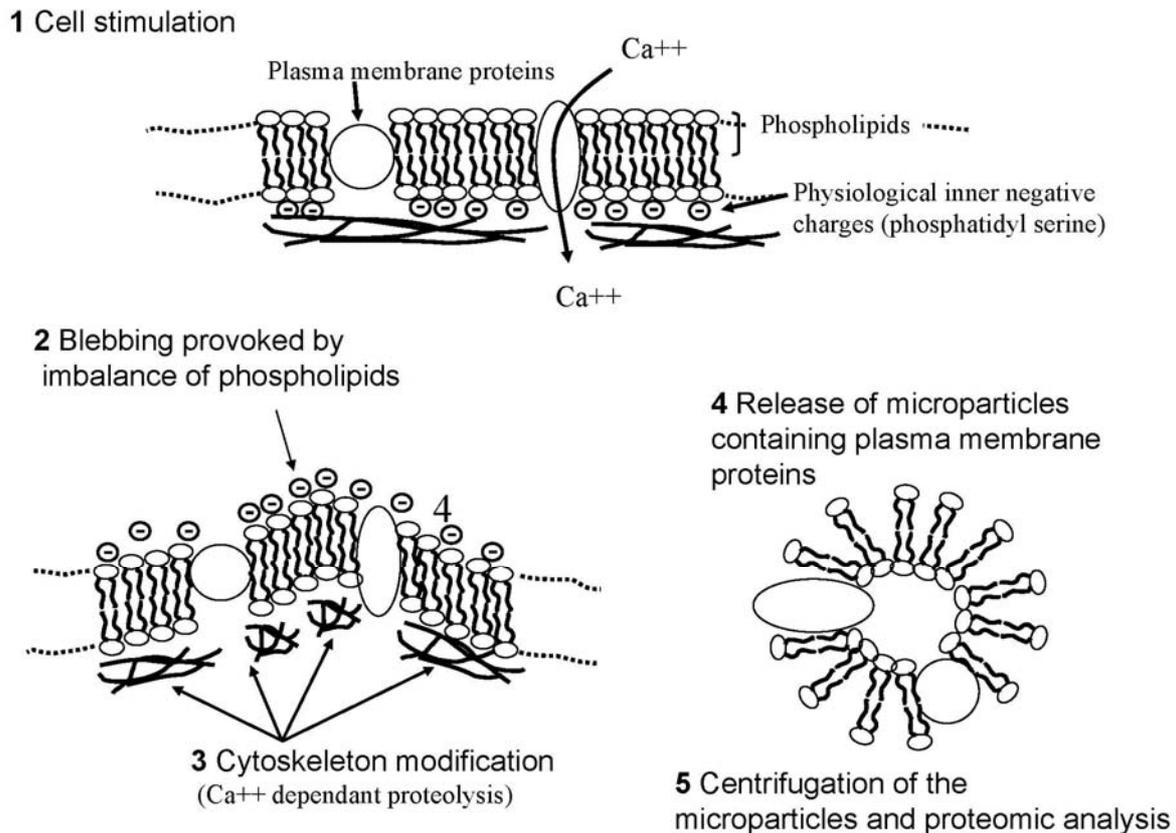


Figure 2.1:

Scheme of the microparticles formation. Cell stimulation (1) induce a remodeling of the phosphatidyl serine repartition) (2) and an increase of cytosolic calcium concentration (3). Such remodeling induces the MPs formation (4) which are easily separated by centrifugation (5).

Such microvesicles have size variable between 50 nm to 1 μ m and differs from other vesicles (like exosomes (30-100 nm)). In general, microparticles are phospholipids vesicles derived from eukaryotic cells as a result of different types of stimulation. Microparticles can also be defined as phospholipids microvesicles containing certain membranes proteins and some proteins originating from the parental cell. Microparticles circulate in the blood and contribute to numerous physiological processes. MPs have been described in various haematopoietic cells as platelets (Heijnen et al. 1999), T-cells (Blanchard et al. 2002), polynuclear neutrophils (Mesri and Altieri 1999) or dendritic cells. After have been considered as cell dust, MPs are now considered to reflect cell activation. Platelet derived microparticles have been the most extensively studied until now. They are now considered to play an important role in the procoagulant process due to the presentation of phospholipids as phosphatidylserine to other platelet which activate them for example (Hugel et al. 2005; Pilzer et al. 2005; Simak and Gelderman 2006).

As microparticles are microvesicles formed directly from plasma membrane, exploration of their protein composition in different pathologies may provide valuable plasma membrane markers specific of the cells they originate from (Pilzer et al. 2005). We then proposed to take the advantage of the physiological production of MPs in order to increase the proportion of plasma membrane proteins and lower the number of irrelevant proteins of the sample.

3) Strategy used for microparticles proteomics analysis

3.1) Presentation of the strategy

Microparticles are naturally enriched in plasma membrane proteins. They correspond themselves to a sub-proteome, and it is tempting to use them as a substitute to perform plasma membrane proteomics. The methodology described bellow has been developed to perform the proteomic analysis of

microparticles. The major problem is linked to the fact that proteins extracted from microparticles are very hydrophobic and therefore cannot be separated on 2-DE (Santoni et al. 2000). Indeed, as a strong ionic detergent must be used for solubilization of the hydrophobic proteins an electrofocalisation step cannot be performed. The protein extract can then only be separated on 1D-SDS-PAGE gel (Galeva and Altermann 2002). The gel lanes obtained were then cut in consecutive pieces of 2 mm. All these pieces (about 100 for a 20 cm gel) were then digested with trypsin. The tryptic peptides were then extracted from each 2 mm piece of the gel and analyzed by MALDI-TOF and ESI-nano-LC-MS-MS. MALDI spectra obtained on these very complex peptide mixtures (several hundred of peptides are expected in a 2 mm slice of 1D gel) did only allowed the identification of the few most abundant proteins using MASCOT (<http://www.matrixscience.com/>), Profound or ProteinProspector (<http://prospector.ucsf.edu/>).

The peptide mixtures are then analyzed by ESI-nano-LC-MS/MS. The mass data recorded during the different analysis were processed and converted into MassLynx.pkl (Q-TOFII, Waters) or Esquire.mgf (IonTrap HCT+, Bruker) format in order to be submitted to a protein database via the different search engines. This step allowed usually the identification of about 10 to 20 different proteins in each 2mm gel slices. This strategy could seem to be time-consuming method (a 20 cm gel generates about 100 slices), but this approach allowed the identification of 400 to 500 proteins for each microparticles preparations. These 400-500 proteins were mainly identified in the 2 mm gel slices regrouped from 20 to 100 kDa (about 50 2mm slices). In each slice an average of 10 proteins were identified. This rather low number is probably due to the fact that the protein extracted from each slice is a sub-proteome. The identification of these proteins using nanoLC-MS/MS was often obtained with a low coverage and few peptides, which was probably due to the fact that only major proteins were detected in each 2mm gel slice. Since sensitivity and chromatographic resolution are probably one of the key elements to improve both the number of proteins identified and the % coverage, we have recently used a LC-Chip chromatographic system (Agilent) coupled to an IonTrap (Bruker) on about 10 slices. In these experiments, the number of protein identified was indeed increased (a factor of 30%). Since MALDI-MS analysis performed on the raw peptide extract was very disappointing from the point of view of protein identification, mainly because of the signal suppression effect, we have developed a new MALDI approach. The suppression effect can only be overcome if the number of peptides present in each fraction is reduced. In order to reduce the complexity of the peptide mixture, a chromatographic fractionation step was performed on the tryptic peptides extract. Figure 3.1 summarizes the strategy and the different analysis performed on each 2mm 1D gel tryptic peptide extract.

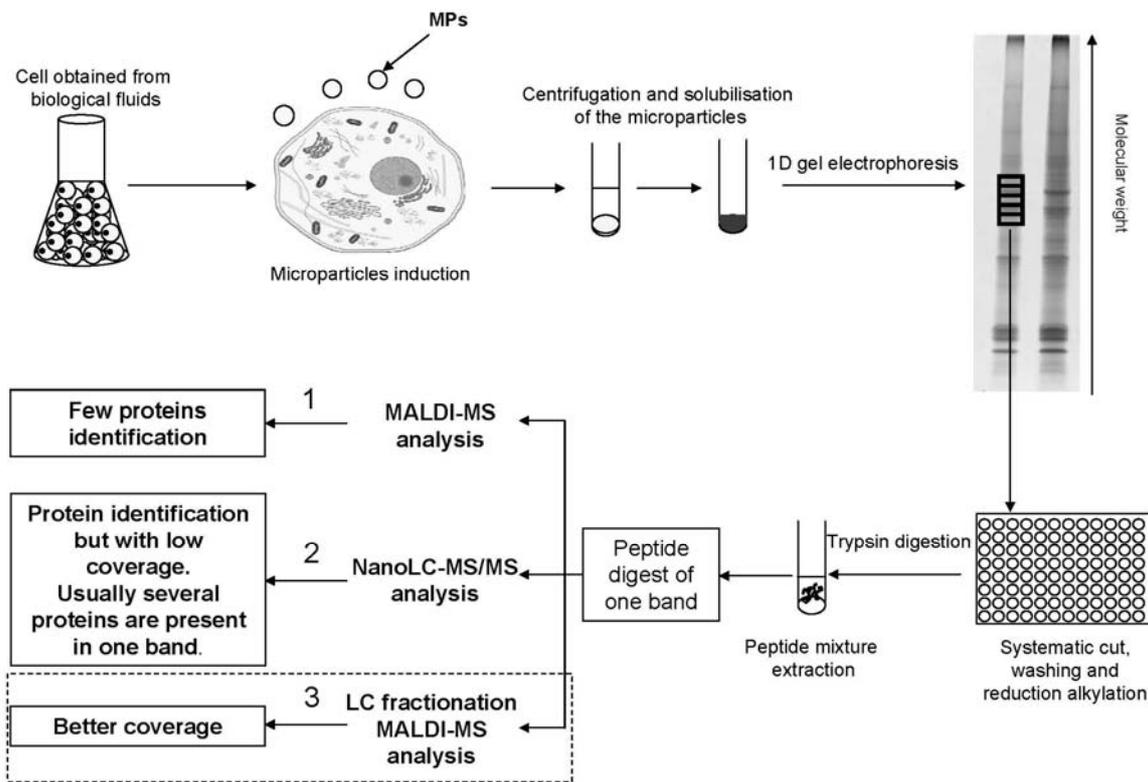


Figure 3.1:

Strategy developed for the microparticles preparation and analysis by mass spectrometry. The step 3 is described in more detail on figure 3.

The column used for this separation was adapted in size to the small quantities of peptides extracted from each slice. It was found that micro-chromatography with 0.3 mm diameter columns yield the best compromise for chromatographic resolution and fractionation feasibility. This approach is therefore based on the possibility to collect efficiently fractions of a few μL from a flow rate of about $4 \mu\text{L}/\text{min}$. In order to speed up the process of data acquisition, a post-column split was used in some cases, allowing to have both ESI-microLC-MS-MS data and fraction collection. Each micro-fraction collected was then submitted to MALDI-MS analysis ($0.5 \mu\text{L}$ used). All MALDI-MS data were then merged to obtain a single peak list showing masses of peptides present in the original extract from one single slice. This single peak list, corresponding to a fictitious MALDI spectrum obtained on the total peptides extract where suppression effect would be absent, allowed to improve very significantly the coverage percentage on the proteins identified by MS-MS data (see below).

The “online” or “offline” LC-ESI analysis also generated a list of peptide masses which were usually complementary to the MALDI peaklist. The few microliters left from the collected fractions after MALDI-MS analysis can be used, when necessary, for complementary analysis (MALDI-MS-MS or nano-ESI-MSn).

The complementarity of all the data obtained in this strategy (LC-MALDI-MS and nanoLC-MS/MS) allowed to increase significantly the coverage percentage for all the gel slices. Figure 3.2 summarize this strategy and technical details are presented below.

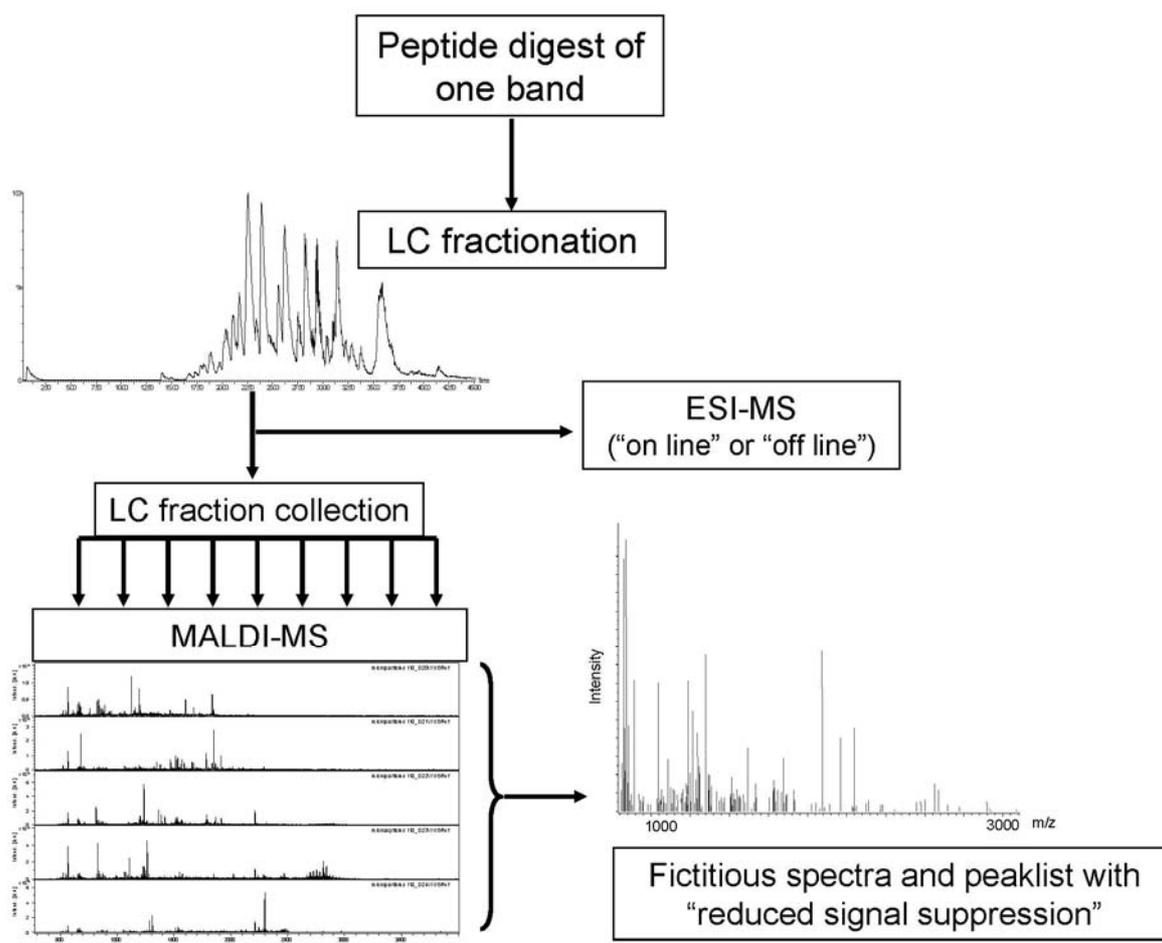


Figure 3.2:
LC-MALDI strategy. Fractionation of the peptide mixture and MALDI-MS analysis.

3.2) Technical details.

The micro-fractionation was realized using a capillary LC system (Agilent 1100 series, Wilmington, DE, USA) equipped with a micro-fraction collector (Agilent 1100 series micro-fraction collector, Wilmington, DE, USA), and all the capillary are optimized for very small column. 15 μl of the samples were injected onto the LC column (Zorbax 300SB-C18, 15 cm x 300 μm , 3.5 μm). Mobile phase solvents were water (A) and acetonitrile (B), containing 0.05% and 0.045% TFA, respectively. Separation was achieved with the following gradient: 5 min at 10% B; from 10% to 80% B in 60 min. The column was maintained at 30°C and the flow rate was set at 4 $\mu\text{l}\cdot\text{min}^{-1}$. Between 3 minutes and 60 minutes, fractions were collected each 2 minutes in a 96 microwell plate using the automatized micro-fraction collector. Such chromatographic steps generate about 15 fractions containing peptides. All the collected fractions are analyzed by MALDI-MS in order to obtain 20 MALDI-MS spectra. This acquisition is realized on ULTRAFLEX (Bruker) using WARP-LC software (Bruker) for acquisition and treatment of the data (figure 3.3).

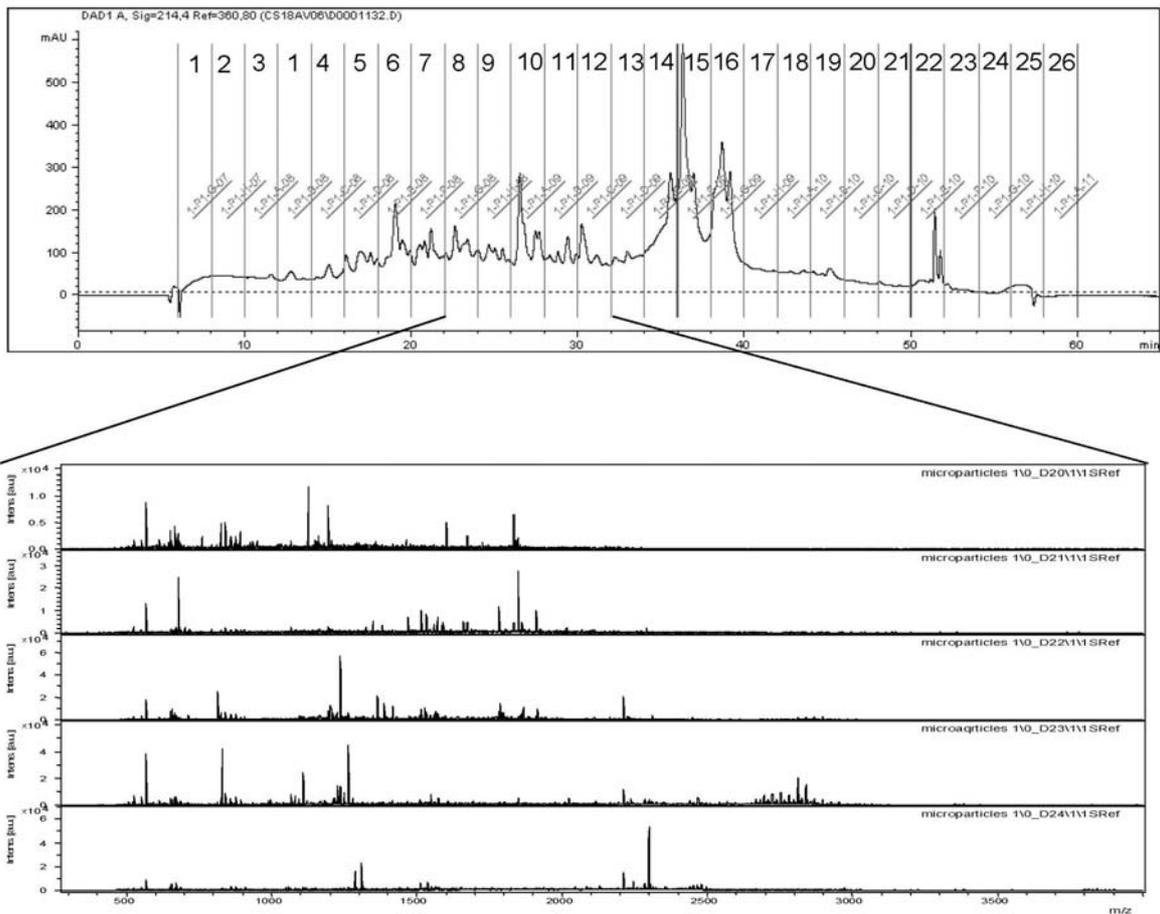


Figure 3.3:

Example of peptide fractionation performed on a 1D gel slice. 26 fractions were collected and analyzed by MALDI-MS. The figure focuses on five of them.

The different spectra are automatically treated in order to generate a single non redundant peaklist. Even if this reconstituted peaklist do not allow new protein identification using peptide mass fingerprint, it combines hundreds different peptides masses. These masses can be used to obtain a better coverage percentage of the different proteins identified in the different spot. Indeed, in the 1DE LC-MS/MS approach, the identification of the different proteins is performed with few peptides generating low sequence coverage. The different proteins identified by nanoLC-MS/MS are digested *in-silico* in order to generate the theoretical peaklist. The theoretical and the experimental peaklists are compared together in order to identify the masses corresponding to the tryptic peptides of the identified proteins. The peaks accepted for this compilation must have a signal/noise greater than 5 and a mass error less than 20 ppm. This strategy allows than a better coverage of the different proteins identified by nanoLC-MS/MS.

The figure 3.4 summarize the all strategy, and an example of this approach obtained from the analysis of a single 1D gel slice is reported in table 3.1.

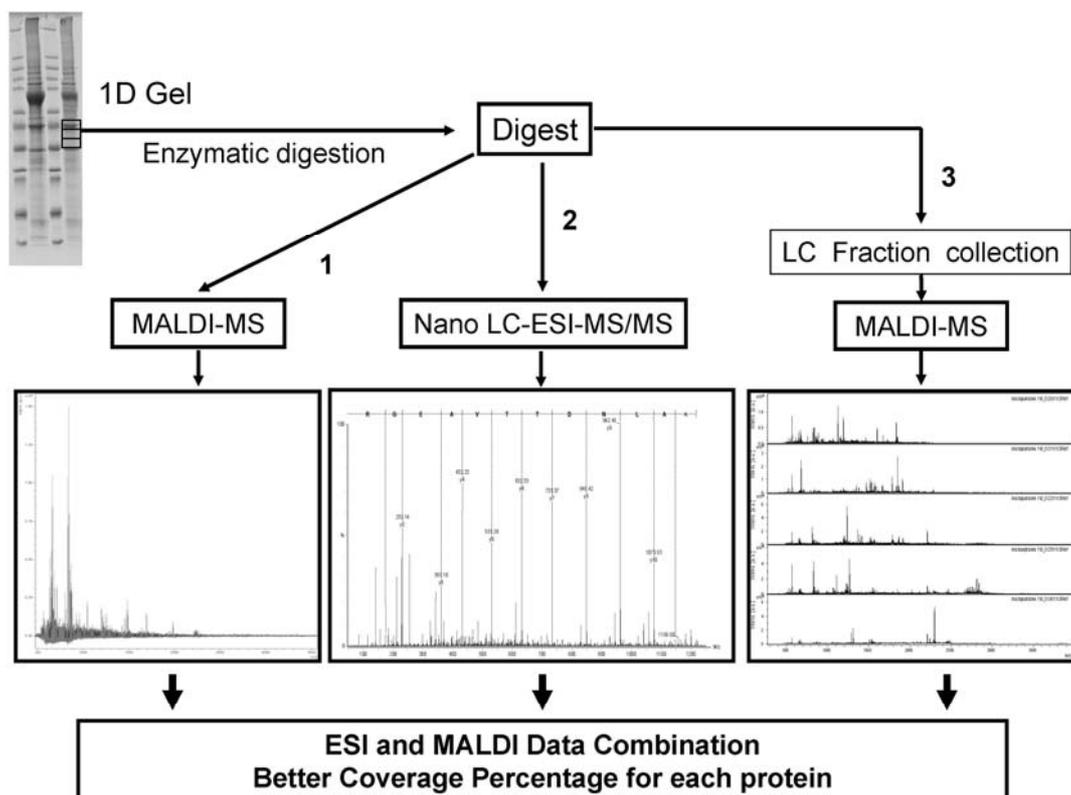


Figure 3.4:

Summarize of the strategy applied for a better coverage of the proteins identified in 1D gel. The first step one is the analysis of the raw peptide mixture by MALDI. The step 2 corresponds to the nanoLC-MS/MS analysis with the peptide digest, and the step 3 corresponds to the “off-line” fractionation of the peptide digest and MALDI-MS analysis. The combination of these approach allow a better coverage of the different proteins identified.

Table 3.1: Summarize of the protein identified by nanoLC-MS/MS from a single 1D gel slice, and the increase of the coverage using LC-MALDI-MS strategy.

accession number	Protein	Peptides identified by nanoLC-MS/MS	Coverage	Peptides identified by LC-MALDI	Combination MALDI/ESI	
					Number of peptides	Final coverage
P08238	Heat shock protein HSP 90-beta	25	31%	20	30	47%
P07900	Heat shock protein HSP 90-alpha	16	21%	27	33	55%
P46940	Ras GTPase-activating-like protein	7	5%	22	27	22%
Q01813	6-phosphofruktokinase type C	4	5%	7	11	19%
P08195	4F2 cell-surface antigen (CD98)	4	9%	10	11	30%
P23921	Ribonucleoside-diphosphate reductase	3	4%	13	14	20%
P26639	Threonyl-tRNA synthetase	3	4%	13	15	21%
P02786	Transferrin receptor (CD71 antigen)	2	2%	14	16	29%
P16070	CD44 antigen precursor	1	1%	3	3	4%
Q15758	Neutral amino acid transporter B(0)	1	2%	3	4	9%

In this table are reported the different proteins identified during the nanoLC-MS/MS analysis of a single 1D gel slice. The peptide obtained from the nanoLC-MS/MS and LC-MALDI-MS experiments are also reported. The combination of both ESI and MALDI experiments shows well the advantage in term of coverage of this double ionization process. Indeed, some peptides are identified specifically from the ESI ionization process and other from the MALDI process, increasing the whole number of identified peptides.

The improvement of sequence coverage of the proteins is very important first to enhance the confidence of the different identifications and second to have a better characterization of the proteins, particularly in term of post-translational modifications and/or mutations.

Moreover, we have determined the false positive rate for this approach. Many tryptic peptides originated from different proteins can be attributed to a single mass (example: HQHPLQCVMEK 1364.63 Da and EADFVNCVIWR 1364.65 Da; $\Delta M < 20\text{ppm}$). Thereby false positive identification may occur. To evaluate the false positive rate, we have selected three common proteins which were not identified during the nanoLC-MS/MS analysis. These proteins (tubulin, actin and myosin) were digested *in-silico*, and the generated mass lists were compared to the LC-MALDI-MS peaklist. A total of only five masses were attributed to the three proteins, which is very few compared to the masses attributed to the different proteins in table 1. This low false positive rate is due to the low tolerance accepted in our strategy ($<20\text{ppm}$).

To conclude, we have developed a strategy which combines LC-MALDI-MS and nanoLC-MS/MS in order to identify proteins originating from 1D gel with a high coverage compatible with plasma membrane study (CD98, CD71, CD44).

4) Validation of the microparticles as a new tool for plasma membrane preparation

In order to validate the hypothesis that microparticles are structures enriched in plasma membranes proteins, we have analysed the proteome composition of different microparticle preparations obtained from a T-Lymphocytic cell line. Such microparticles can be produced and enriched *in-vitro* by mitogenic activation (PHA) or apoptosis induction (Act D1, TNF α) for instance. Microparticles proteins mixture obtained from the two stimulations were separated independently on a 1D gel. The different gels were cut each 2 mm and the different slide were digested and analyzed by nanoLC-MS/MS.

This approach allowed the identification of 390 proteins among which 34% were localized to the plasma membrane. The microparticles obtained from the two different ways of stimulation did not show significant variation in their proteome (Miguet et al. 2005).

These results have been compared to a systematic analysis of a classical plasma membrane preparation by ultracentrifugation (100000g on sucrose). This plasma membrane isolation has allowed the identification of 292 proteins, and only 69 proteins were located to the plasma membrane (24% versus 34% for the microparticles preparation). Moreover, 75 plasma membrane proteins were identified specifically in the microparticles preparation, and only 9 were identified only in the plasma membrane preparation. These results are summarized on figure 4.1.

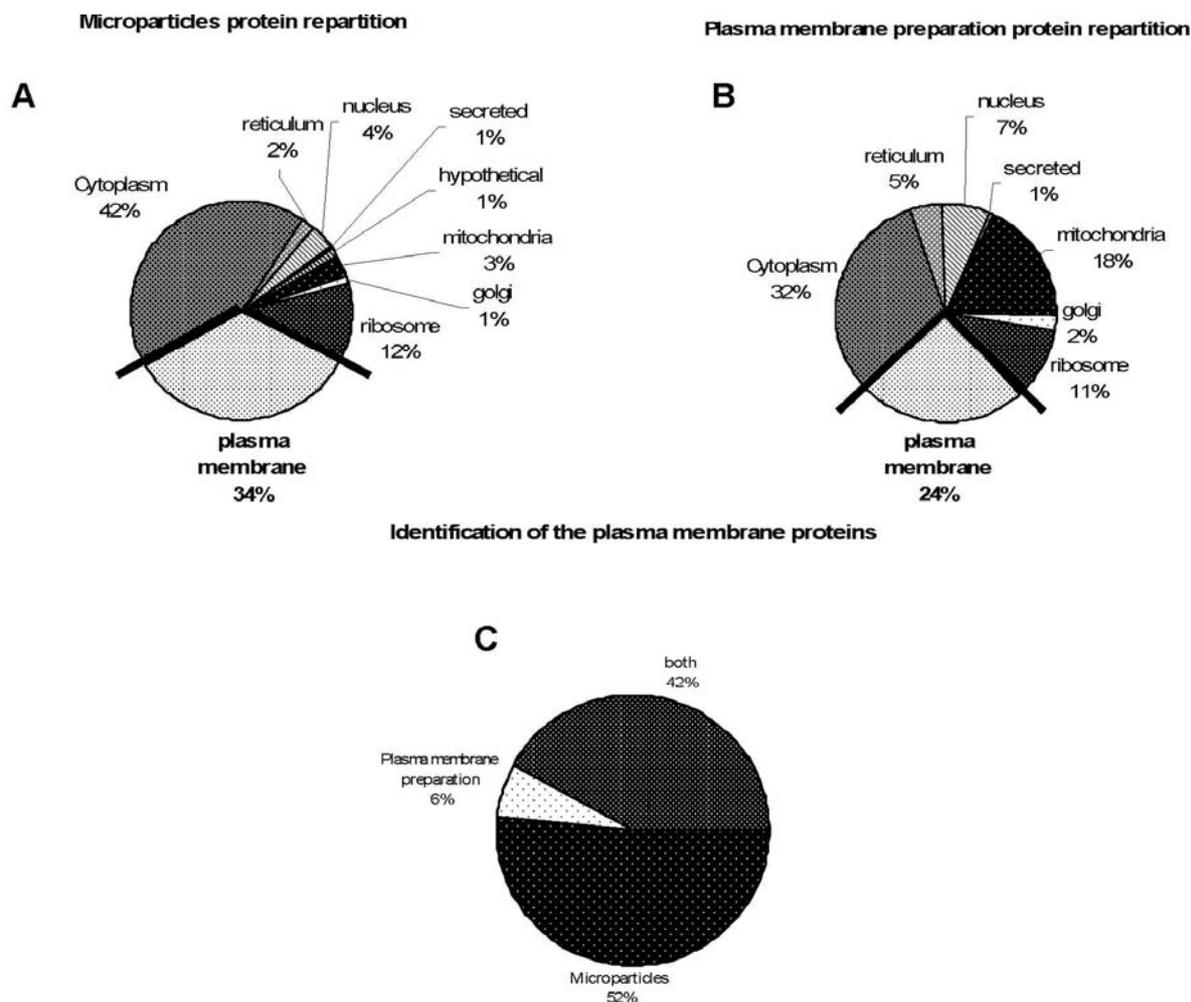


Figure 4.1:

Representation of the protein repartition in function of the preparation. A: repartition of the proteins identified in the microparticles, B: repartition of the proteins identified in the classical plasma membrane preparation; C: comparison of the plasma membrane identified in the microparticles and in the plasma membrane preparation.

On the other hand different groups have started working on the plasma membrane microparticles proteome (Garcia et al. 2005) (Banfi et al. 2005). The results obtained by these groups are very different from the results coming from our lab. This might be due to the difference in the microparticles centrifugation step used for MPs isolation. Indeed these two groups used intense centrifugation (between 100000 and 250000 g) which is very high in comparison with the theoretical microparticles size. This step culotate indeed the microparticles, but also the other vesicles contained in the solution as exosomes which are much smaller. For example, the classical way to prepare exosome preparation is to centrifuge for half an hour between 100000 and 200000 g (They et al. 2001; Wubbolts et al. 2003). We believe that this centrifugation step is crucial in order to generate a pure preparation of microparticles. This point was already underlined, even if not discussed, by Jin et al. (Jin et al. 2005). In this aim a centrifugation of 20000g during 45 minutes, that we used, is appropriate to generate a homogeneous microparticles preparation (Hugel et al. 2005).

5) Perspective

As it is potentially possible to generate microparticles from almost all cell types, it could be a new tool for sub-cellular plasma membrane study and biomarker discovery. Even if quantitation is not assessed,

the comparison of the microparticles sub-proteome, originating from different pathologies, may enlighten differences in protein expression. As our group is interested in the study of the different B-cell chronic lymphopathies, we have decided to take advantage of the microparticles preparations obtained from various pathologies in order realize their proteome. In this aim, fresh cells were obtained following informed consent, and subjected to actinomycinD stimulation. The obtained microparticles were first separated on 1D gel before to be systematically analyzed by nanoLC-MS/MS. We analyzed a lymphoma (small cell lymphoma). This analysis has allowed the identification of more than 400 proteins among which 30% were localized to the plasma membrane or associated to the plasma membrane.

Moreover, in this study, a large number of proteins involved in the signal transduction were identified. Indeed B-cell receptor signaling proteins (numerous membrane receptors, kinase proteins) implied in the signal transduction was detected. This complexome identified in the microparticles is reported on figure 5.1. The proteins in grey were identified in our study.

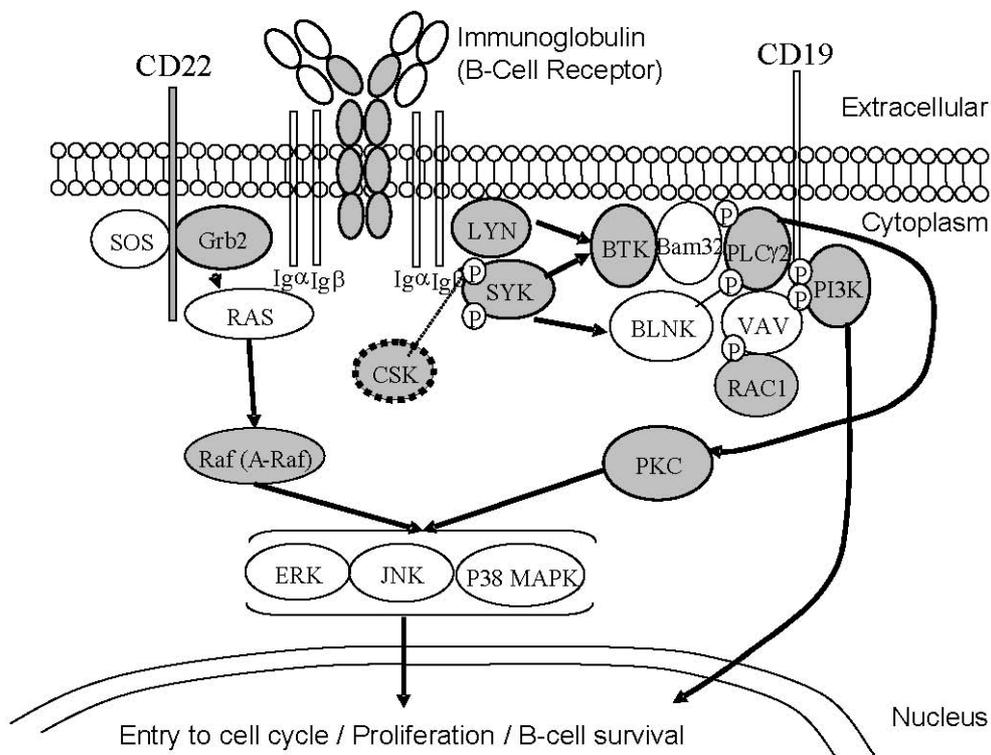


Figure 5.1:

Representation of the B-cell receptor, and the different kinase implied in the signal transduction for the cell proliferation. The proteins in grey were identified in the microparticles obtained from the patient. Most of the proteins implied in the signal transduction were identified using this strategy.

In conclusion, the plasma membrane microparticles represent a sub-proteome reflecting the plasma membrane composition. These structures have the advantage of not being contaminated by membrane proteins originating from other cell components (golgi, nucleus, mitochondria...). As microparticles can be produced by almost all cell lines, they can be useful to discover new potential biomarkers. In this goal we have undertaken the comparison of microparticles originating from different kind of lymphoma. The analytical strategy developed for the characterization of the plasma membrane sub-proteome, based on micro-fractionation, allowed the identification of hundreds proteins with a better coverage percentage than the classical nanoLC-MS/MS approach. The combinations of nanoLC-MS/MS, micro-fractionation and “off-line” MALDI-MS data is now used routinely in general study for the discovery of new specific biomarkers.

6) Bibliographie

- Banfi, C., Brioschi, M., Wait, R., Begum, S., Gianazza, E., Pirillo, A., Mussoni, L. and Tremoli, E. (2005). "Proteome of endothelial cell-derived procoagulant microparticles." *Proteomics* **5**(17): 4443-55.
- Bevers, E. M., Comfurius, P., Dekkers, D. W., Harmsma, M. and Zwaal, R. F. (1998). "Regulatory mechanisms of transmembrane phospholipid distributions and pathophysiological implications of transbilayer lipid scrambling." *Lupus* **7 Suppl 2**(2): S126-31.
- Blanchard, N., Lankar, D., Faure, F., Regnault, A., Dumont, C., Raposo, G. and Hivroz, C. (2002). "TCR activation of human T cells induces the production of exosomes bearing the TCR/CD3/zeta complex." *J Immunol* **168**(7): 3235-41.
- Galeva, N. and Altermann, M. (2002). "Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins." *Proteomics* **2**(6): 713-22.
- Garcia, B. A., Smalley, D. M., Cho, H., Shabanowitz, J., Ley, K. and Hunt, D. F. (2005). "The platelet microparticle proteome." *J Proteome Res* **4**(5): 1516-21.
- Harris, M. (2004). "Monoclonal antibodies as therapeutic agents for cancer." *Lancet Oncol* **5**(5): 292-302.
- Heijnen, H. F., Schiel, A. E., Fijnheer, R., Geuze, H. J. and Sixma, J. J. (1999). "Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules." *Blood* **94**(11): 3791-9.
- Hugel, B., Martinez, M. C., Kunzelmann, C. and Freyssinet, J. M. (2005). "Membrane microparticles: two sides of the coin." *Physiology (Bethesda)* **20**: 22-7.
- Jang, J. H. and Hanash, S. (2003). "Profiling of the cell surface proteome." *Proteomics* **3**(10): 1947-54.
- Jin, M., Drwal, G., Bourgeois, T., Saltz, J. and Wu, H. M. (2005). "Distinct proteome features of plasma microparticles." *Proteomics* **5**(7): 1940-52.
- Luche, S., Santoni, V. and Rabilloud, T. (2003). "Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis." *Proteomics* **3**(3): 249-53.
- Mesri, M. and Altieri, D. C. (1999). "Leukocyte microparticles stimulate endothelial cell cytokine release and tissue factor induction in a JNK1 signaling pathway." *J Biol Chem* **274**(33): 23111-8.
- Miguet, L., Pacaud, K., Felden, C., Hugel, B., Martinez, M. C., Freyssinet, J. M., Herbrecht, R., Potier, N., van Dorsselaer, A. and Mauvieux, L. (2005). "Proteomic analysis of

- malignant lymphocyte membrane microparticles using double ionization coverage optimization." *Proteomics* **6**(1): 153-171.
- Peirce, M. J., Wait, R., Begum, S., Saklatvala, J. and Cope, A. P. (2004). "Expression profiling of lymphocyte plasma membrane proteins." *Mol Cell Proteomics* **3**(1): 56-65. Epub 2003 Oct 21.
- Pilzer, D., Gasser, O., Moskovich, O., Schifferli, J. A. and Fishelson, Z. (2005). "Emission of membrane vesicles: roles in complement resistance, immunity and cancer." *Springer Semin Immunopathol* **27**(3): 375-87.
- Qoronfleh, M. W., Benton, B., Ignacio, R. and Kaboord, B. (2003). "Selective Enrichment of Membrane Proteins by Partition Phase Separation for Proteomic Studies." *J Biomed Biotechnol* **2003**(4): 249-255.
- Rahbar, A. M. and Fenselau, C. (2004). "Integration of Jacobson's pellicle method into proteomic strategies for plasma membrane proteins." *J Proteome Res* **3**(6): 1267-77.
- Santoni, V., Molloy, M. and Rabilloud, T. (2000). "Membrane proteins and proteomics: un amour impossible?" *Electrophoresis* **21**(6): 1054-70.
- Simak, J. and Gelderman, M. P. (2006). "Cell membrane microparticles in blood and blood products: potentially pathogenic agents and diagnostic markers." *Transfus Med Rev* **20**(1): 1-26.
- They, C., Boussac, M., Veron, P., Ricciardi-Castagnoli, P., Raposo, G., Garin, J. and Amigorena, S. (2001). "Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles." *J Immunol* **166**(12): 7309-18.
- Wallin, E. and von Heijne, G. (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." *Protein Sci* **7**(4): 1029-38.
- Wubbolts, R., Leckie, R. S., Veenhuizen, P. T., Schwarzmann, G., Mobius, W., Hoernschemeyer, J., Slot, J. W., Geuze, H. J. and Stoorvogel, W. (2003). "Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation." *J Biol Chem* **278**(13): 10963-72. Epub 2003 Jan 7.
- Zhao, Y., Zhang, W. and Kho, Y. (2004). "Proteomic analysis of integral plasma membrane proteins." *Anal Chem* **76**(7): 1817-23.

4) Conclusions générales.

Lors de ce chapitre principalement axé sur le développement de nouvelles méthodologies en spectrométrie de masse, nous avons tout d'abord développé un nouvel outil pour l'étude des protéines membranaires en protéomique. En effet, nous avons montré que les microparticules sont des structures naturellement enrichies en protéines issues de la membrane plasmique, et que de ce fait les microparticules peuvent servir de préparation initiales pour l'étude des protéines membranaires.

Dans un deuxième temps, nous avons déterminé la reproductibilité des analyses nanoLC-MS/MS. Cette approche a permis de déterminer le nombre d'analyses à réaliser pour avoir une vision globale du contenu protéique d'un échantillon complexe issu de gel 1D. Nous avons effectivement pu montrer qu'il faut réaliser six injections successives du même échantillon pour ne plus identifier de nouvelles protéines et donc avoir une vision globale du mélange protéique. Ces données sont extrêmement importantes lorsque l'on veut comparer différents protéomes (listes de protéines) entre eux. En effet, il faut pouvoir affirmer qu'une protéine est indéniablement identifiée dans un échantillon, et inversement qu'une protéine qui n'a pas été identifiées est effectivement absente de l'échantillon, ou si elle n'a pu être identifiée du à différents biais induits par la nature même de l'analyse. Cette stratégie pourra donc être appliquée à la recherche de nouveaux biomarqueurs, en comparant les protéomes issus des différentes pathologies.

Enfin nous avons montré la complémentarité des approches nanoLC-MS/MS et LC-MALDI-MS afin d'augmenter le pourcentage de recouvrement des différentes protéines issues d'échantillons complexes. Cette technique a pour but d'aller toujours plus loin dans la caractérisation des différentes protéines identifiées, et de confirmer les protéines identifiées avec peu de peptides.

Recherche de nouveaux biomarqueurs dans le cadre des hémopathies lymphoïdes chroniques

Conclusions majeures :

- Approche protéomique : nous avons identifié plusieurs centaines de protéines à partir de préparations de microparticules issues de patients atteints de différentes formes d'hémopathies lymphoïdes chroniques. Nous avons proposé deux nouveaux marqueurs potentiels spécifiques du lymphome du manteau : le CD148 et la sorcin. Ces deux biomarqueurs potentiels sont en cours de validation (RT-PCR et cytométrie en flux).
- Approche peptidomique : nous avons identifié trois peptides qui sont surexprimés dans les sérums de patients atteints d'hémopathie lymphoïde chronique (la forme sulfitée de la transthyréline, des peptides de dégradation du complément C3 et C4).

Cette étude s'inscrit dans le cadre d'une collaboration entre notre laboratoire de spectrométrie de masse, le laboratoire d'Hématologie cellulaire de l'Institut d'Hématologie et d'Immunologie de la Faculté de Médecine de Strasbourg, le laboratoire d'hématologie (Dr Laurent Mauvieux) et le département d'onco-hématologie des hôpitaux universitaires de Strasbourg (Pr Raoul Herbrecht).

Cette collaboration a pour but de mettre en évidence des biomarqueurs spécifiques des différentes hémopathies lymphoïdes chroniques. Pour cela, deux nouvelles stratégies ont été développées au laboratoire afin de répondre à cette problématique.

La première, que l'on pourrait qualifier d'approche protéomique, consiste à rechercher directement par spectrométrie de masse et analyse protéomique des biomarqueurs exprimés spécifiquement au niveau de la membrane plasmique des lymphocytes pathologiques.

La deuxième, que l'on pourrait qualifier d'approche peptidomique, consiste à comparer le profil peptidique de sérums d'individus sains à ceux de patients présentant une hémopathie lymphoïde chronique afin de déterminer des masses discriminantes et d'identifier les peptides correspondant à ces masses par spectrométrie de masse.

1) Contexte biologique :

1.1) Définitions :

- **Hémopathie** : Maladie du sang (maligne ou non).
- **Lymphome** : Désigne la prolifération maligne touchant les cellules lymphoïdes matures d'organes donnés (ganglions lymphatiques, rate...). Les phases dites leucémiques des lymphomes, correspondent à l'envahissement du sang par les cellules pathologiques.
- **Leucémie** : (sang blanc) Prolifération maligne qui peut toucher les différentes lignées de cellules hématopoïétiques. Les cellules cancéreuses prolifèrent anormalement dans les différents organes impliqués dans la production et/ou la maturation des cellules sanguines, et vont s'accumuler dans le sang. Les leucémies aiguës, d'évolution rapide, touchent les cellules immatures, et les leucémies chroniques touchent des cellules matures (lymphoïdes ou myéloïdes).
- **Leucémie lympho-proliférative chronique** : Pathologie durant laquelle les lymphocytes s'accumulent dans le sang.

1.2) Problématique

Les hémopathies lymphoïdes chroniques représentent un ensemble de pathologies cancéreuses touchant les lymphocytes B matures de l'adulte. Ces pathologies sont en constante augmentation dans tous les pays occidentaux, et représentent 4% des nouveaux cancers chaque année [Harris 2004]. Ces pathologies sont des syndromes accumulatifs qui peuvent évoluer de manière totalement différente (taux de survie pouvant aller de 4 ans à 25 ans).

Le diagnostic clinique de ces différentes pathologies repose principalement sur l'analyse morphologique et immuno-phénotypique des différentes cellules pathologiques circulantes en utilisant le score de Matutes [Matutes et al. 1994]. Cependant, dans un certain nombre de cas, cette méthode n'est pas suffisante pour établir un diagnostic précis. De plus, dans la majorité des cas, l'évolution de ces différentes pathologies reste imprévisible au moment du diagnostic. Ces difficultés aussi bien du point de vue diagnostique que pronostique rendent difficile le choix du traitement adéquat (chimiothérapie intensive ou non, anticorps armés,...) [Harris 2004].

Bien que certaines hémopathies ne peuvent pas être clairement classifiées, un certain nombre de biomarqueurs ont néanmoins été identifiés. Toutefois, il n'existe pas de biomarqueur spécifique de chacune de ces pathologies. Le fait de ne pas avoir de biomarqueurs spécifiques de ces pathologies

permettant une classification précise est aujourd'hui un obstacle majeur à la lutte contre ces leucémies [Matutes 1995].

L'objectif de cette étude était de développer différentes techniques de spectrométrie de masse afin de trouver et d'identifier de nouveaux biomarqueurs de ces leucémies. Deux nouvelles approches complémentaires ont été développées en parallèle afin de mieux comprendre ces différentes pathologies :

- Une approche protéomique, qui cherche à étudier les protéines membranaires de surface en utilisant les microparticules.
- Une approche peptidomique, qui consiste à comparer les profils peptidiques issus de sérums sains ou de patients leucémiques.

2) Recherche et identification de biomarqueurs de leucémies : méthodes classiques.

La manière dont les biomarqueurs sont ou ont été trouvés jusqu'alors repose sur des critères de disponibilité d'anticorps spécifiques. Historiquement, les différents travaux de recherche portaient sur l'étude de la différenciation des lymphocytes immatures en lymphocytes matures, et la comparaison de l'expression protéique entre ces différents états. Ces études ont donc donné lieu à la génération de banques d'anticorps monoclonaux contre les protéines de lymphocytes. Ces différents anticorps ont été testés de manière systématique sur les cellules issues des différentes pathologies. Cette approche a permis d'identifier les différents marqueurs utilisés à ce jour.

Par exemple, CD5 (T-cell surface glycoprotein CD5 precursor ; P06127), était à l'origine une protéine connue pour être exprimée spécifiquement par les lymphocytes T, et pour ne pas être exprimée par les lymphocytes B matures. Or des études de 1984 et 1985 ont pu montrer que cette protéine de surface était exprimée par des lymphocytes B pathologiques ainsi que par les lymphocytes B fœtaux [Flug et al. 1985] [Antin et al. 1986]. Aujourd'hui CD5 est un biomarqueur permettant la classification des différentes pathologies comme les lymphomes du manteau, les leucémies lymphoïdes chroniques, et les lymphomes prolymphocytaires par exemple.

Aujourd'hui encore, c'est cette approche par anticorps qui est la plus employée. On peut citer pour exemple les récents travaux de Ek et al [Ek et al. 2006], intitulés « From gene expression analysis to tissue microarrays - a rational approach to identify therapeutic and diagnostic targets in lymphoid malignancies ». Dans cette étude, une centaine d'anticorps, présélectionnés via des banques d'EST (Expressed Sequence Tag), sont testés sur des lymphocytes B normaux et pathologiques afin de mettre en évidence des différences au niveau protéique. L'étude de ces dizaines d'anticorps a permis de réaliser des profils protéiques et de mettre en évidence 4 anticorps qui sembleraient spécifiques du lymphome du manteau.

2.1 Les différents biomarqueurs connus

Il existe à l'heure actuelle un cocktail de différents biomarqueurs de surface utilisés pour réaliser l'immuno-phénotypage des différentes cellules pathologiques (Tableau 1).

Hémopathies	Marqueurs Immunophénotypiques
Leucémie Lymphoïde Chronique	CD5 ⁺ , CD10 ⁻ , CD11c ^{-/+} , CD19 ⁺ , CD20 ⁺ , CD 22 ^{+/-} , CD23 ⁺ , CD43 ⁺ , CD79a ⁺
Lymphome du manteau	CD5 ⁺ , CD10 ^{-/+} , CD11c ⁻ , CD19 ⁺ , CD20 ⁺ , CD23 ⁻ , CD43 ⁺ , CD79a ⁺
Lymphome de la zone marginale	CD5 ⁻ , CD10 ⁻ , CD11c ^{+/-} , CD19 ⁺ , CD20 ⁺ , CD23 ⁻ , CD43 ^{+/-} , CD79a ⁺
Lymphome de la zone marginale ganglionnaire	CD5 ⁻ , CD10 ⁻ , CD23 ⁻ , CD43 ^{+/-} , CD103 ⁻
Leucémie à lymphocytes villeux	CD5 ⁻ , CD10 ⁻ , CD11c ⁺ , CD19 ⁺ , CD20 ⁺ , CD23 ⁻ , CD25 ⁺ , CD79a ⁺ , CD103 ⁺ FMC7 ⁺
leucémie lymphoplasmocytaire	CD5 ⁻ , CD10 ⁻ , CD19 ⁺ , CD20 ⁺ , CD22 ⁺ , CD23 ⁻ , CD43 ^{+/-} , CD79a ⁺ , CD103 ⁻
Myélome	CD5 ⁻ , CD10 ^{+/-} , CD19 ⁻ , CD20 ⁻ , CD23 ⁻ , CD39 ⁺ , CD43 ^{+/-} , CD103 ⁻ , CD79a ⁺
Lymphome prolymphocytaire	CD19 ⁺ , CD20 ⁺ , CD22 ⁺ , CD79a ⁺ , CD5 ^{+/-} , CD23 ⁻
Lymphome folliculaire	CD5 ⁻ , CD10 ⁺ , CD19 ⁺ , CD20 ⁺ , CD22 ⁺ , CD43 ⁻ , CD79a ⁺

Tableau 1 : Biomarqueurs utilisés pour la classification des différentes hémopathies

On remarque qu'il est nécessaire d'analyser une dizaine de marqueurs différents pour pouvoir classer et différencier les différentes pathologies. Parmi ces différents marqueurs, on trouve les marqueurs spécifiques des lymphocytes B (comme CD20, CD22...) et des marqueurs aberrants comme CD5 par exemple. Il est aussi à souligner que dans certains cas, ces différents marqueurs ne sont pas suffisamment spécifiques pour obtenir une bonne classification des pathologies, et que par exemple les lymphomes de la zone marginale et les lymphomes du manteau présentent une expression de marqueurs immunophénotypiques très proches.

2.2) Classification actuelle des différentes leucémies

Aujourd'hui, la classification des leucémies communément acceptée est celle de l'OMS. Une représentation extrêmement schématique de cette classification est représentée en figure 1.

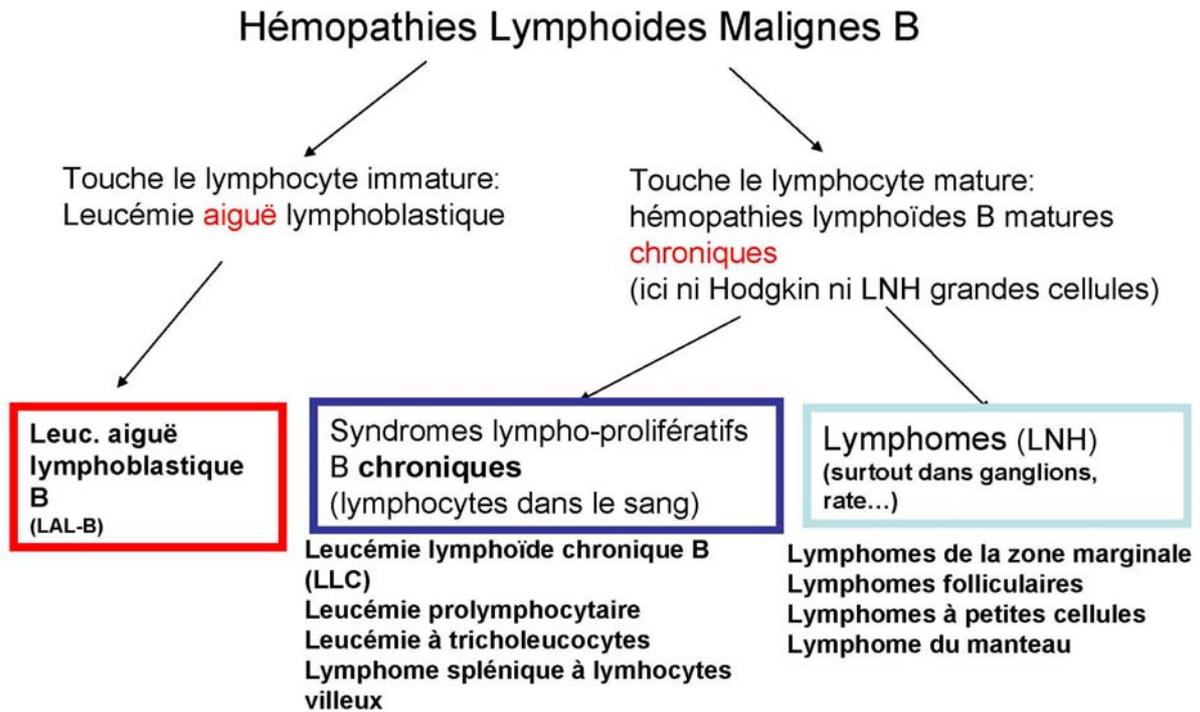


Figure 1 : Tableau de classifications des hémopathies lymphoïdes malignes B.

Il est à souligner que cette classification date des années 90 [Harris et al. 1994; Matutes et al. 1994], et qu'elle est toujours d'actualité. Très peu de progrès du point de vue de la classification ont donc été réalisés grâce à l'outil immunologique. De plus, il reste un grand nombre de cas litigieux, ou le diagnostic n'est pas possible. On peut prendre comme exemple le cas des LLC dites atypiques qui représentent une catégorie « fourre tout » dans laquelle sont rangés tous les syndromes qui ne répondent pas aux critères de classification. Il est donc nécessaire de développer de nouvelles approches afin de pouvoir mieux classer ces différentes pathologies.

3) Recherche et identification de biomarqueurs de leucémies : approche protéomique par analyse des protéines membranaires.

Comme nous l'avons montré dans le chapitre 2 de la partie résultats, les microparticules sont naturellement enrichies en protéines issues de la membrane plasmique, et sont donc un outil d'étude efficace pour identifier de nouveaux marqueurs membranaires. Nous avons donc décidé de comparer entre eux les protéomes de microparticules issues de différents lymphocytes leucémiques.

Dans le cadre de cette étude, nous n'avons pas pu comparer les différents protéomes obtenus à partir de cellules leucémiques à un protéome de référence issu de cellules saines (échantillon contrôle). En effet, il n'a pas été possible d'obtenir suffisamment de lymphocytes B sains pour générer une quantité suffisante de microparticules (nécessité de plusieurs ml de sang). Cette approche doit donc nous permettre d'observer des protéines exprimées de façon différentielle entre les différents états pathologiques étudiés, et ainsi de proposer de nouveaux candidats spécifiques d'une leucémie par rapport aux autres (protéine identifiée spécifiquement dans une pathologie par rapport à une autre).

Néanmoins cette approche pourrait paraître discutable. En effet, nous ne disposons d'aucun moyen de quantification des différentes protéines identifiées. Par cette approche nous ne pouvons pas déterminer si une protéine est sur ou sous exprimée dans une pathologie par rapport à une autre, et ce n'est pas parce que une protéine est identifiée dans un seul échantillon qu'elle n'est pas présente dans les autres (problème de suppression de signal, protéine masquée par une protéine abondante...). Cependant, conscients de ces limitations, nous avons tout de même pris le risque de mettre en place cette approche qui est la seule qui permette d'identifier un grand nombre de protéines membranaires, et de donner une image globale du protéome membranaire des cellules pathologiques. De plus, nous pensons que cette approche protéomique peut apporter des informations importantes quant à une meilleure compréhension des hémopathies lymphoïdes chroniques en proposant de nouveaux candidats, qui pourront être confirmés par d'autres techniques biochimiques (ELISA, RT-PCR...), spécifiques d'une leucémie. Cette stratégie est résumée dans la figure 2.

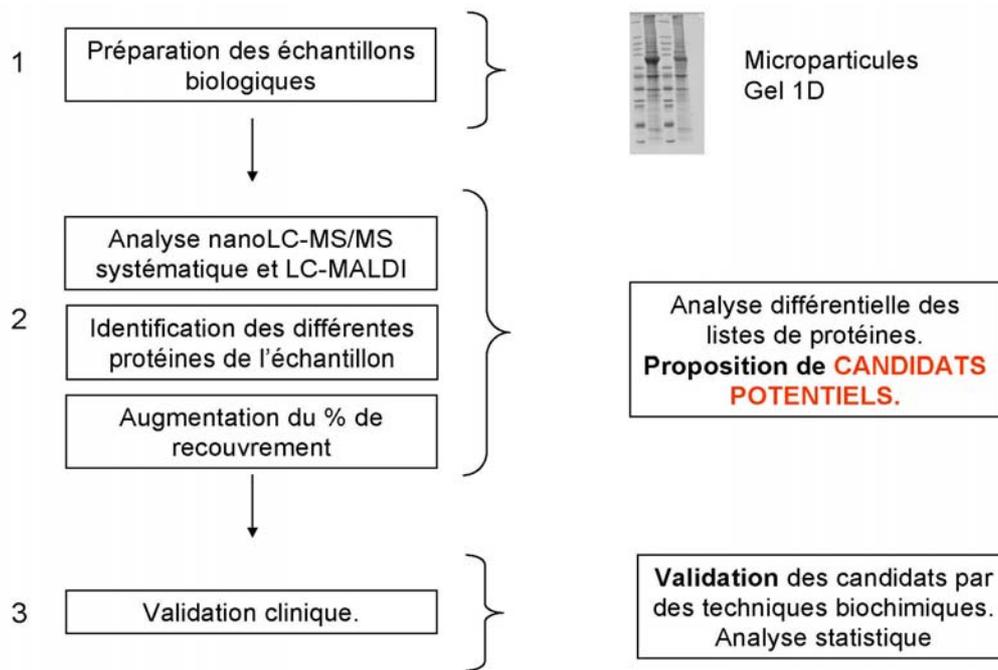


Figure 2 : La stratégie consiste en l'identification des protéines par spectrométrie de masse et comparaison des listes de protéines entre elles. Les candidats potentielles seront alors validés par des techniques biochimiques et des analyses statistiques.

3.1) Production des microparticules et préparation de l'échantillon

Les microparticules ont été produites à partir de lymphocytes B de patients atteints d'hémopathies lymphoïdes chroniques.

Cette étude a été réalisée à partir de trois échantillons de leucémies chroniques :

- Lymphome du manteau (MCL)
- Leucémie lymphoïde chronique (LLC)
- Lymphome à petites cellules (LNH)

et d'une leucémie aigue (contrôle):

- Leucémie aigue lymphoblastique (LAL)

Les lymphocytes B contenus dans les différents prélèvements sanguins ont été séparés par cytométrie en flux. Pour chacune des pathologies, environ 150 millions de cellules ont été soumises à l'action de l'actinomycineD pendant 18h afin de stimuler la production des microparticules. Les microparticules ont alors été culottées par centrifugation à 15000g pendant 45 minutes. Les différentes protéines

contenues dans les microparticules ont été solubilisées dans un tampon de charge d'électrophorèse (riche en SDS), et déposées sur un gel à gradient 8-15% (figure 3).

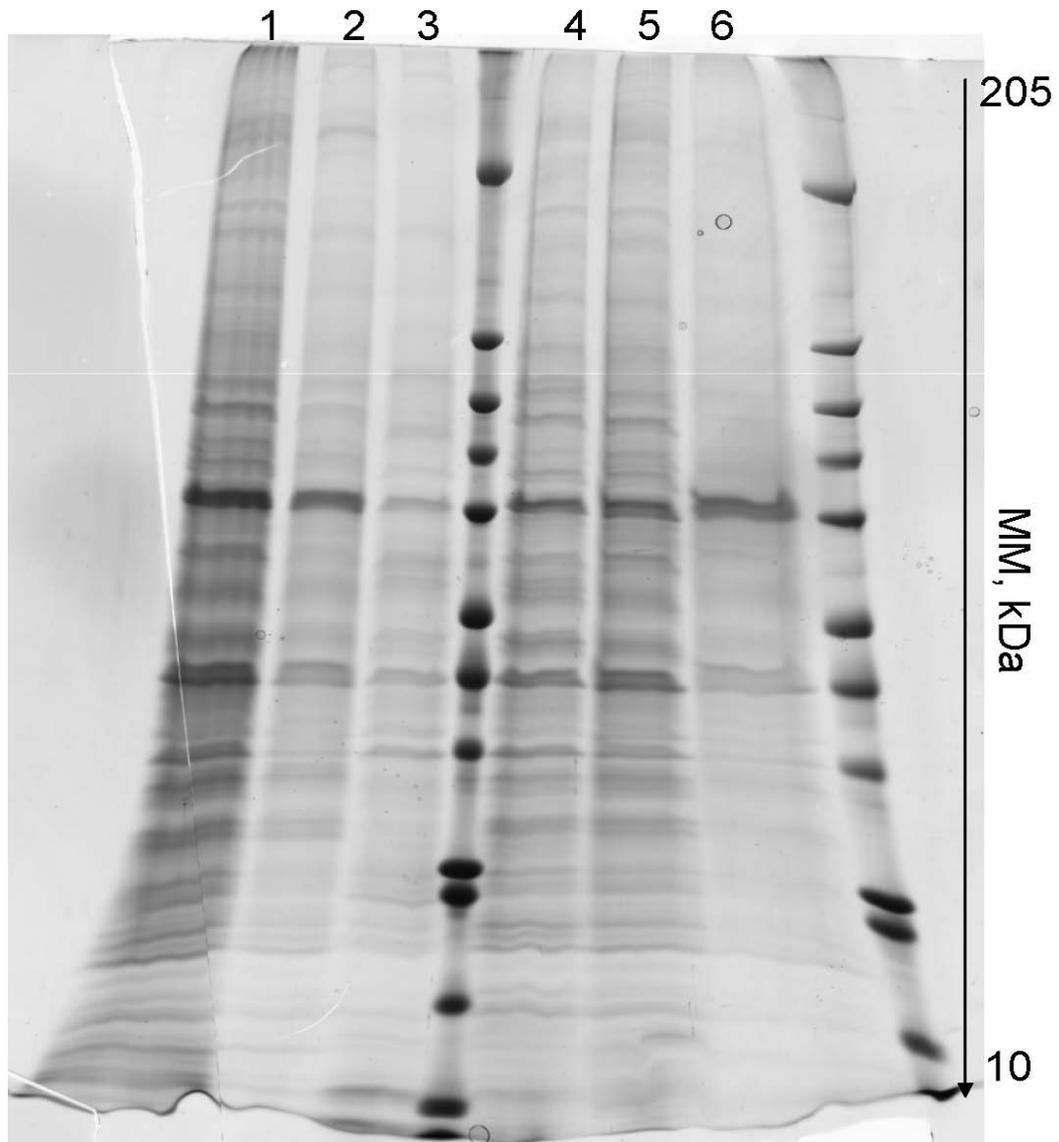


Figure 3 : Gel 1D gradient 8-15% coloré au bleu de Coomassie des différentes pathologies étudiées. Piste 1et 5 MCL ; piste 2 LNH ; piste 3 lymphome polynucléaires (pas analysé) ; piste 4 LAL ; piste 6 LLC.

Après coloration du gel, les différentes pistes du gel ont été découpées systématiquement tous les 2 millimètres.

3.2) Analyse par spectrométrie de masse

Les protéines contenues dans les différentes bandes ont été réduites et alkylées avant d'être digérées à la trypsine. Les peptides ont été extraits et analysés par spectrométrie de masse nanoLC-MS/MS. Les analyses ont été réalisées sur un système de chromatographie capillaire (CapLC Waters) couplé à un spectromètre de masse en tandem Q-TOF (Waters). Les différents extraits peptidiques ont

été analysés sur une colonne C18 (LC Packings). Le débit a été fixé à 200nL/min avec un gradient d'élution de 5% à 45% d'acétonitrile dont la vitesse d'application était de 1% par minute pour toutes les analyses. Les listes de masses correspondant aux peptides détectés et analysés en MS et MS/MS ont été générées automatiquement en fichiers .pkl par le logiciel Masslynx™.

3.3) Résultats et comparaison entre les différentes pathologies

Etant donné le peu de matériel biologique dont nous disposions lors de cette étude, nous n'avons pu réaliser qu'une analyse pour chacun des échantillons. A partir des données de spectrométrie de masse, les recherches par homologie de masse ont été réalisées par interrogation de la banque de données SwissProt. Cette étude a permis d'identifier environ 2000 protéines sur les quatre pathologies dont 800 différentes (la liste de protéines est fournie en annexe). La proportion en protéines membranaires était sensiblement identique à celle déterminée lors du chapitre précédent (Chapitre II § 2.2.1) (autour de 30%). Ainsi nous avons identifié environ 240 protéines membranaires ou associées à la membrane, parmi lesquelles on retrouve un grand nombre de clusters de différenciation. Les différents clusters de différenciation en fonction des pathologies sont reportés dans le tableau 2.

CD	LNH	LLC	LAL	MCL
CD11a		X	X	
CD11b			X	
CD18	X	X	X	X
CD20	X	X		X
CD22				X
CD29		X	X	
CD31	X			
CD37	X			X
CD43	X			
CD44		X	X	X
CD45	X	X	X	X
CD47		X		X
CD49c		X		
CD49d			X	
CD50	X	X		X
CD53	X			
CD71				X
CD81				X
CD82				X
CD140			X	
CD148				X
CD205	X			
CD222	X			

Tableau 2 : Clusters de différenciation identifiés en fonction des pathologies.

Par cette approche, nous avons pu identifier 23 clusters de différenciation différents dont des tetraspanines, comme CD81 qui est une protéine associée aux récepteurs B ou CD53 et qui est connue pour interagir avec des glycoprotéines de surface qui se complexent à différentes intégrines [Levy et Shoham 2005].

Parmi ces différents clusters de différenciation, on en trouve un grand nombre qui sont spécifiques des lymphocytes B comme CD22 ou CD11. Par contre d'autres marqueurs de ces pathologies comme CD5 n'ont pu être identifiés. Ceci peut être expliqué entre autre par la faible abondance absolue de ce marqueur (environ 3000 copies/cellules cancéreuses) [Lavabre-Bertrand et al. 1994], alors qu'il est présent entre 15 et 20000 copies sur les lymphocytes T. On peut aussi remarquer que la protéine CD20 qui sert à la classification de différentes pathologies (marqueur protéique spécifique des lymphocytes B) a été identifiée dans les trois pathologies chroniques, mais pas dans la pathologie aigue. Ceci est logique, étant donné que CD20 est un marqueur de cellules matures, et qu'il est donc peu/ou pas exprimé dans les cellules immatures. Cet élément a été confirmé par cytométrie en flux.

De plus, l'identification des protéines détectées a permis de mettre en évidence différentes voies de transduction du signal dont l'une d'elles, correspondant à l'activation de la division cellulaire via le récepteur des cellules B (immunoglobuline) est représentée sur la figure 4. Les protéines grisées ont été identifiées lors de l'analyse du lymphome à petites cellules.

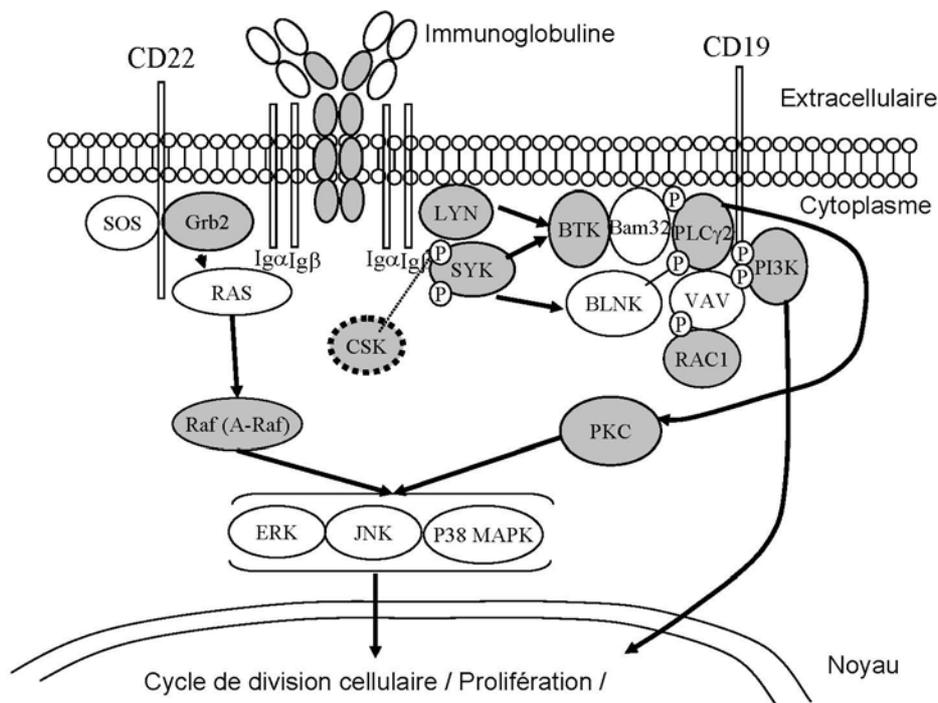


Figure 4 : Schéma du récepteur des cellules B (Immunoglobuline) et des différentes protéines impliquées dans la transduction du signal pour la division cellulaire. Les protéines grisées ont été identifiées dans l'étude du lymphome à petites cellules.

On peut remarquer sur cette figure que nous avons pu identifier une immunoglobuline membranaire impliquée dans la division cellulaire, ainsi que la majeure partie des protéines associées à la membrane impliquées dans la cascade de kinases impliquée dans la transduction du signal.

De plus, nous avons identifié 7 annexines différentes (tableau 3)

Annexin	LNH	LLC	MCL	LAL
1		X	X	X
2	X	X	X	X
4	X			X
5	X	X	X	X
6	X	X	X	X
7			X	X
11	X	X	X	

Tableau 3 : Annexines identifiées en fonction des pathologies.

Parmi les différentes annexines identifiées, nous avons pu identifier l'annexine 1 qui joue un rôle important pour la résistance des cellules cancéreuses aux différentes drogues [Wang et al. 2004], ou l'annexine 7 qui joue un rôle important dans les différents transports ioniques et en particulier celui du calcium. De plus, l'annexine 7 est connue pour interagir avec une autre protéine : la sorcin [Verzili et al. 2000] [Salzer et al. 2002]. Or, nous avons identifié la sorcin uniquement dans le lymphome du manteau, et la sorcin est connue pour jouer un rôle important dans la résistance aux drogues (couplé avec des protéines MDR (multidrug resistance protein))[Zhou et al. 2006].

3.4) Proposition de nouveaux candidats

En étudiant de façon différentielle les protéomes obtenus pour chacune de ces pathologies au cours de cette étude, nous avons pu mettre en évidence des protéines membranaires ou associées à la membrane uniquement identifiées dans une des pathologies. Parmi ces différentes protéines, nous nous sommes intéressés à deux grandes familles :

- Les clusters de différenciation (CD)
- Les annexines

En effet, les CD sont des molécules de surface faciles d'accès pour des techniques biochimiques qui travaillent sur cellules entières (il existe des anticorps spécifiques de ces protéines). De plus, les différents marqueurs déjà utilisés pour la classification des différentes leucémies sont des clusters de différenciation (cf tableau 1). Le deuxième groupe de protéines sont les annexines qui, d'après de récentes études, semblent jouer un rôle important dans ces pathologies [Holder et al. 2006] [Falini et al. 2004].

Sélection des candidats :

Nous avons donc sélectionné deux candidats identifiés uniquement dans le cadre du lymphome du manteau afin des réaliser des analyses complémentaires pour valider la spécificité de ces protéines.

La première protéine sélectionnée est la sorcin (Numéro Swiss Prot : P30626, Sorcin (22 kDa protein)). En effet, comme nous l'avons déjà mentionné elle semble jouer un rôle très important dans la résistance aux différentes drogues

La deuxième protéine que nous avons sélectionnée lors de cette étude est la protéine CD148 (Numéro Swiss Prot : Q12913, Protein-tyrosine phosphatase eta). Cette protéine a été identifiée avec 6 peptides différents et un pourcentage de recouvrement de 6%, ce qui est assez important pour une protéine de 150 kDa. De plus, aucun peptide correspondant à cette protéine n'a été identifié dans les autres pathologies.

Bien que ce choix puisse paraître arbitraire, il nous semble justifié d'une part par l'intérêt biologique de ces deux candidats, et d'autre part par la qualité des identifications de ces protéines. En effet, ces protéines ont été identifiées par des analyses nanoLC-MS/MS avec un nombre conséquent de peptides. Ceci laisse à penser qu'elles ne sont que peu voir pas exprimées dans les autres pathologies.

3.5) Validation de ces nouveaux candidats

La validation des différents candidats consiste à déterminer et à mettre en évidence les variations des niveaux d'expression de ces différentes protéines en fonction des pathologies.

Validation de la Sorcin :

La validation de ce candidat est réalisée par analyse RT-PCR (équipe de L. Mauvieux). Ces analyses sont actuellement en cours. Des résultats préliminaires sur une dizaine de patients choisis aléatoirement à partir des différentes leucémies montrent déjà des taux d'expression allant de 0 (patient 14) à 10 (patient 11), et donc une très grande hétérogénéité entre les différentes pathologies (figure 5). Le contrôle représente une lignée cellulaire T.

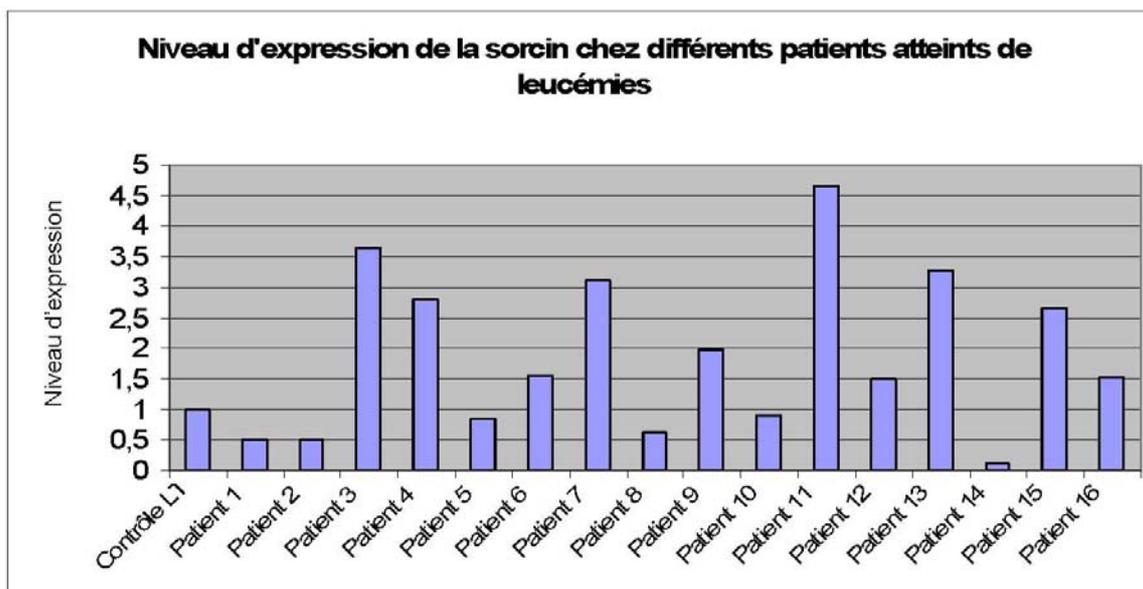


Figure 5 : résultats préliminaires du niveau d'expression de la sorcin par RT-PCR.

Une étude complémentaire est en cours pour corrélér ces résultats de RT-PCR avec les différentes pathologies.

Validation de CD148 :

La validation consiste en un marquage par anticorps anti-CD148 couplé à un fluorochrome et une détection par cytométrie en flux. Cette validation est elle aussi en cours. Les résultats préliminaires semblent indiquer que les lymphocytes B normaux expriment fortement cette protéine. Or nous n'avons identifié le CD148 que dans une seule pathologie (lymphome du manteau). Donc nous nous attendons à trouver des différences d'expression entre les différentes pathologies (diminution de l'expression de CD148 dans les LAL, les LLC et les LNH par rapport aux MCL et aux lymphocytes B normaux).

3.6) Conclusions et perspectives

Lors de cette étude, nous avons comparé le protéome de préparation de microparticules issues de 4 pathologies distinctes. Bien que cette approche fut osée, puisqu'aucune notion de quantification n'a été envisagée, elle a permis d'obtenir des informations quant aux protéines « présentes et absentes » de l'échantillon (protéines identifiées / protéines pas identifiées). Ainsi nous avons identifié des protéines seulement dans certains de ces échantillons. Ces protéines en fonction de leur relevance biologique seront analysées par d'autres techniques (RT-PCR, Immuno-détection...) afin d'être confirmées ou infirmées comme différemment exprimées entre les pathologies et ainsi servir comme nouveaux biomarqueurs spécifiques d'un type de leucémie.

Nous avons pu par cette approche identifier plusieurs centaines de protéines dont un grand nombre issues de la membrane plasmique. Parmi ces protéines, nous avons pu entre autre identifier spécifiquement dans une pathologie précise (lymphome du manteau) la Sorcin qui est une protéine associée à la membrane et impliquée dans la résistance aux drogues [Zhou et al. 2006], et la protéine tyrosine phosphatase eta (CD148). Cette protéine semble être exprimée, d'après une étude de Dong et al. [Dong et al. 2002], uniquement dans certaines leucémies. Mais à l'heure actuelle, il n'existe que cette publication dans la littérature.

Les perspectives de ce travail sont tout d'abord de finir la validation des différentes protéines proposées comme biomarqueurs potentiels (en cours de réalisation par l'équipe du Dr Laurent Mauvieux), et de recommencer ces analyses sur plus de matériel biologique afin de pouvoir réaliser trois analyses successives. En effet, comme nous l'avons montré dans le chapitre II de ce travail de thèse, si nous sommes capables de réaliser trois analyses successives, nous pourrons alors affirmer que nous avons identifié au moins 90% des protéines contenues dans l'échantillon. A ce moment, si nous identifions une protéine spécifiquement dans une pathologie, et pas dans les autres, nous pourrons conclure avec une forte probabilité que la protéine est spécifique de l'échantillon, ou tout au moins supposer que la variation d'expression de cette protéine entre les différentes pathologies sera grande. Les candidats qui seront alors proposés en vue d'une validation biologique seront déjà « pré-validés ». De plus, l'identification de ces nouveaux biomarqueurs permettra de développer des tests biologiques de type ELISA qui pourront servir au diagnostic des différentes pathologies.

4) Utilisation de l'approche peptidomique par SELDI-TOF-MS sur des sérums de patients

Toujours dans le but d'identifier des biomarqueurs d'hémopathies lymphoïdes chroniques, nous avons décidé d'utiliser une deuxième approche que l'on pourrait qualifier d'approche peptidomique. Cette approche consiste à réaliser des profils peptidiques (inférieurs à 20kDa) à partir de fluides biologiques complexes. Dans cette étude, nous nous sommes axés sur l'étude de sérums issus de patients atteints de différentes hémopathies lymphoïdes chroniques.

4.1) Description de l'approche

L'approche SELDI-TOF-MS consiste à simplifier le mélange peptidique/protéique à analyser en utilisant diverses surfaces chromatographiques qui sont greffées sur des barrettes. Les différentes protéines vont interagir spécifiquement avec les différentes surfaces chromatographiques en fonction de leurs propriétés physico-chimiques. Les échantillons sont ensuite lavés dans des conditions plus ou moins rigoureuses, ce qui permet d'éliminer les protéines interagissant de façon aspécifique. Les analytes qui sont spécifiquement fixés sur les surfaces chromatographiques sont analysés par SELDI-TOF-MS (figure 6).

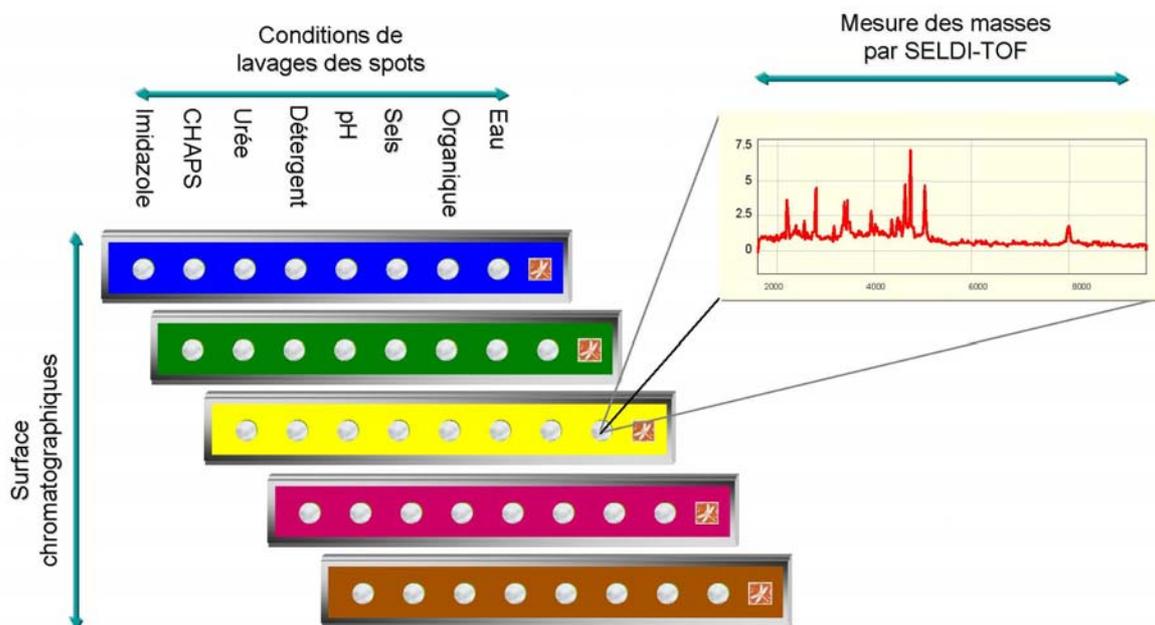


Figure 6 : Schéma de la stratégie SELDI.

Les analyses ont été réalisées sur les différentes surfaces chromatographiques proposées par CIPHERGEN®. Nous avons ainsi généré des profils spécifiques en fonction des différentes surfaces chromatographiques. Les spectres obtenus sont reportés sur la figure 7.

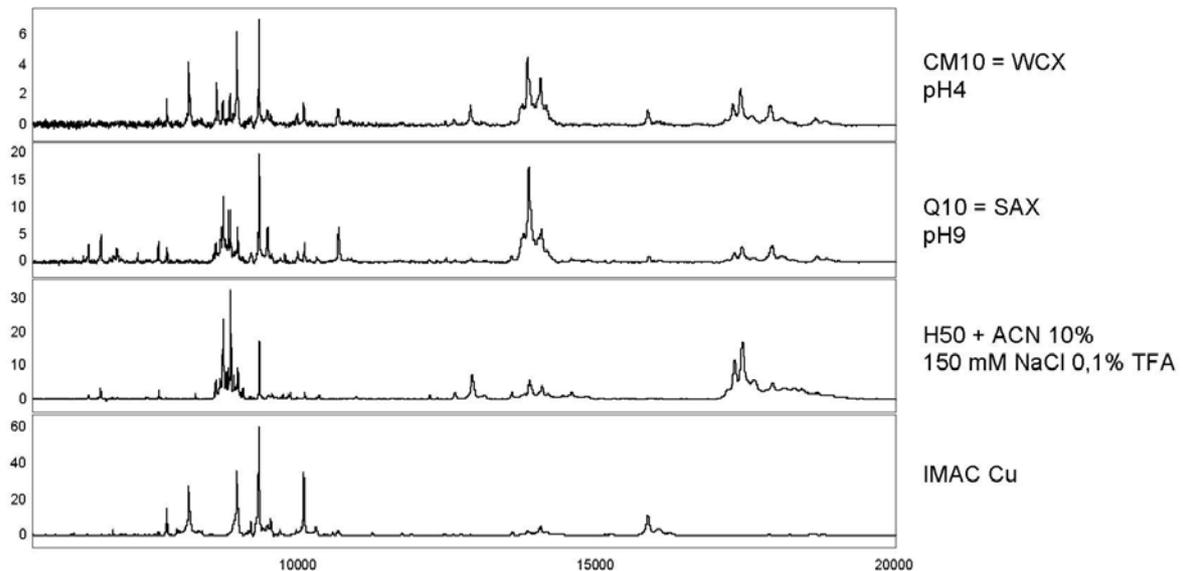


Figure 7 : Spectres de masse représentant la complémentarité des différentes surfaces chromatographiques (WCX : weak cation exchange ; SAX : Strong anion exchange ; IMAC : immobilized metal affinity chromatography ; H50 : surface hydrophobe en C8) obtenus sur un SELDI PBS II.

On peut observer sur les différents spectres la complémentarité qui existe entre les différentes surfaces chromatographiques. De plus, les conditions de lavage des spots permettent elles aussi d'obtenir des informations complémentaires (Figure 8)

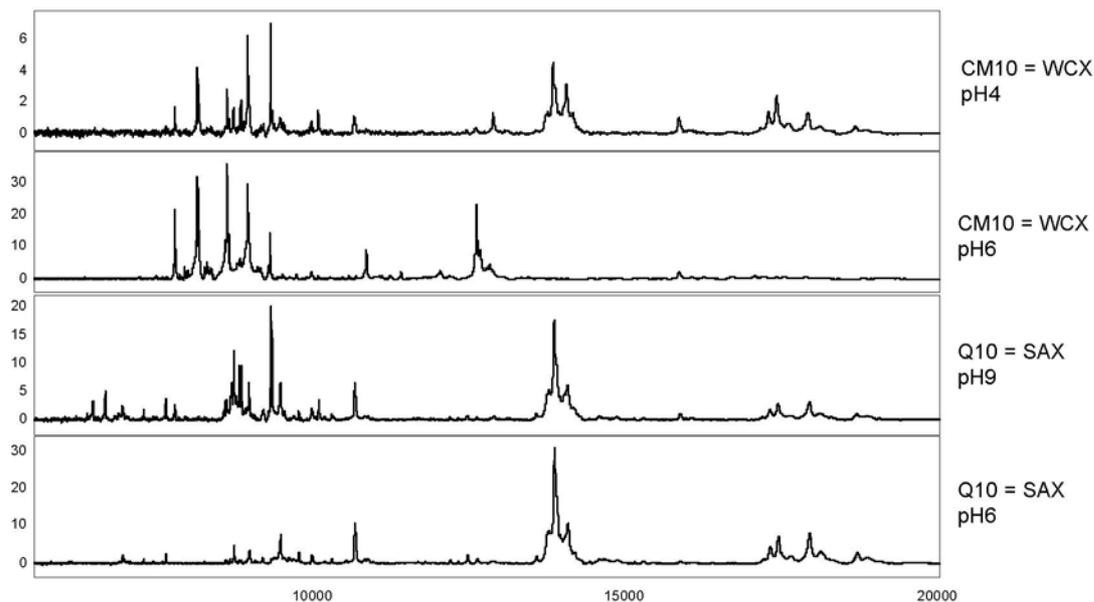


Figure 8 : Spectres illustrant la complémentarité des différents lavages.

On obtient ainsi en fonction des surfaces chromatographiques et des différentes conditions de lavage des profils peptidiques complémentaires. Ces différents profils pourront être comparés entre eux afin de déterminer la présence de pics discriminants entre le groupe de contrôles et le groupe de patients.

4.2) Résultats : Profiling et analyse statistique

Nous avons appliqué cette approche sur 39 sérums issus de patients atteints d'hémopathies lymphoïdes chroniques au diagnostic contre 20 contrôles.

Les analyses SELDI-TOF ont permis de générer des spectres présentant trois groupes de peptides qui variaient entre les patients et les contrôles. L'analyse statistique de ces variations a permis de confirmer que les différents massifs peptidiques étaient spécifiques (figure 9).

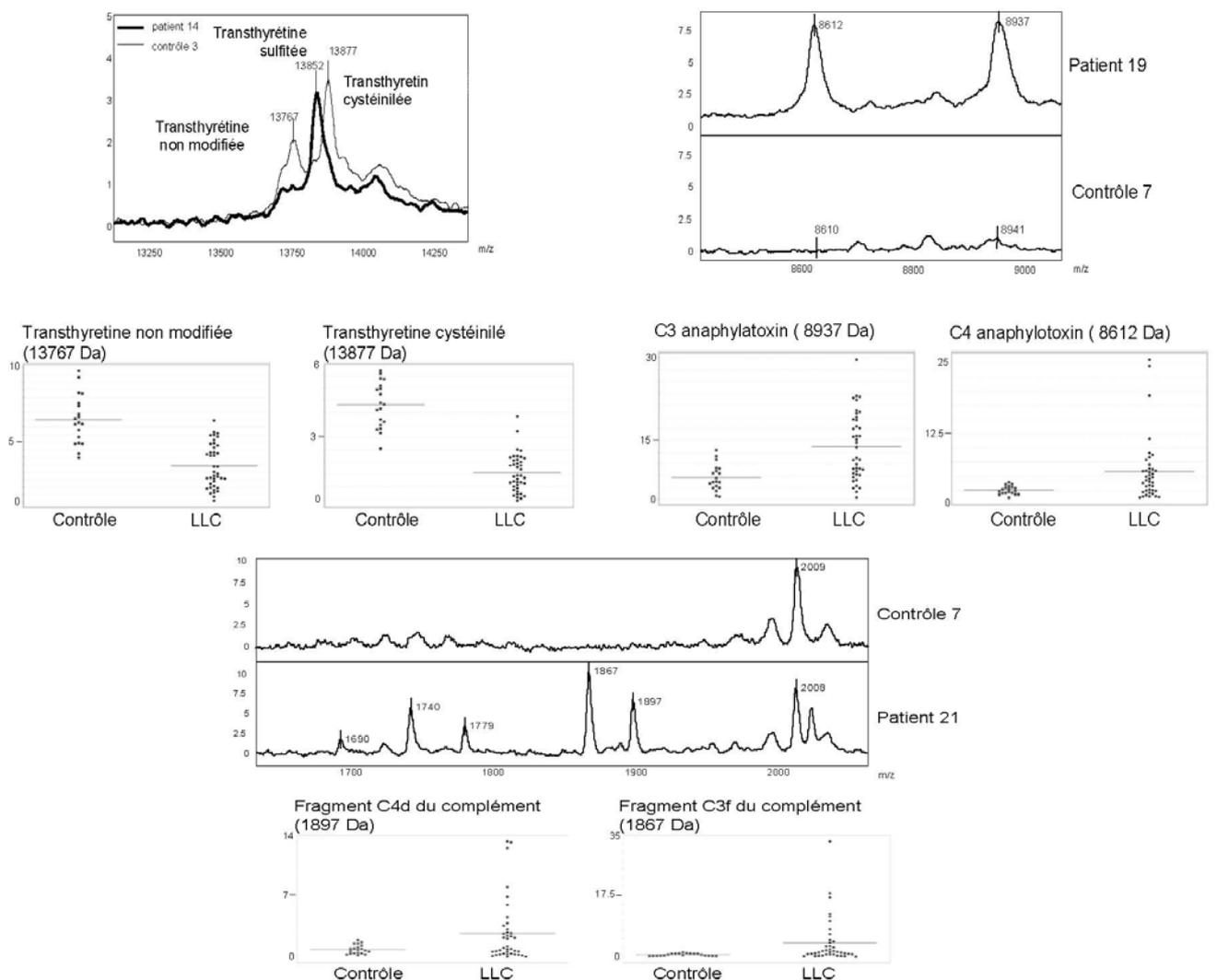


Figure 9 : Représentation des différents candidats peptidiques analysés par SELDI-TOF-MS.

4.3) Purification des peptides d'intérêts et identification par spectrométrie de masse

Les différents peptides d'intérêts ont été purifiés par différentes techniques (HPLC, gel d'électrophorèse, colonnes échangeuse d'ions) avant d'être analysés et identifiés par spectrométrie de masse. Les stratégies de purifications et d'identifications sont reportées dans la figure 10.

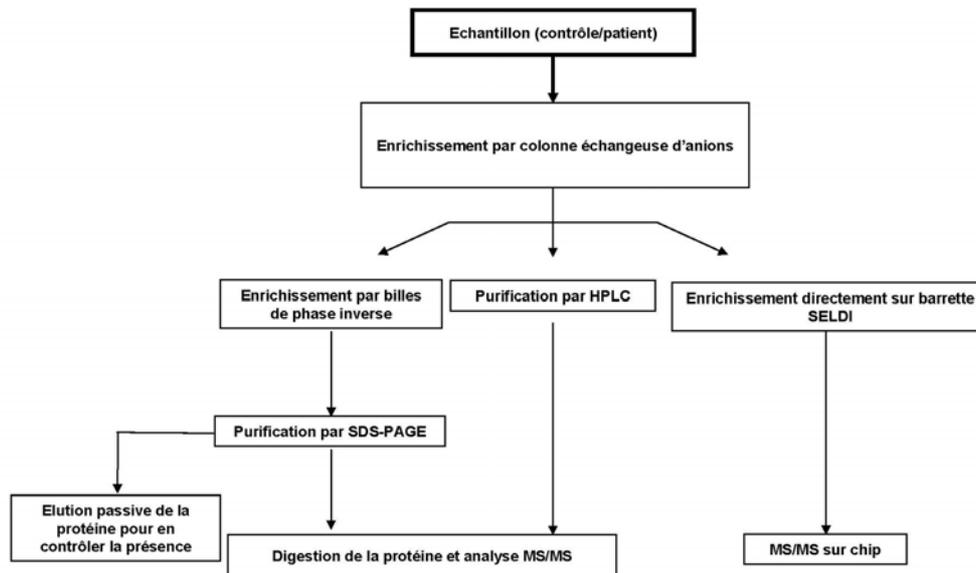


Figure 10 : Schéma de la stratégie de purification et d'identification des différents peptides d'intérêts.

Cette stratégie a permis d'identifier trois groupes de peptides qui sont les suivants:

- La transthyréline (TTR)
- Fragments du complément C3
- Fragments du complément C4

De plus, nous avons pu mettre en évidence par spectrométrie de masse que la TTR était présente sous quatre formes différentes dans nos échantillons :

- La forme non modifiée (13767 Da)
- La forme sulfitée (13852 Da)
- La forme cystéinillée (13877 Da)
- La forme glutathionillée (14060 Da)

Nous avons aussi montré que la forme sulfitée de la TTR était très fortement augmentée dans les LLC, alors que la forme non modifiée et la forme cystéinillées étaient largement diminuées.

4.4) Publication Journal of proteom research

Ces travaux ont fait l'objet d'une publication dans le Journal of Proteome Research.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Discovery and Identification of Potential Biomarkers in a Prospective Study of Chronic Lymphoid Malignancies Using SELDI-TOF-MS

Laurent Miguet, Ralf Bogumil, Philippe Decloquement, Raoul Herbrecht, Noelle Potier, Laurent Mauvieux, and Alain Van Dorsselaer

Journal of Proteome Research, 2006, Vol. 5, Pages 2258-2269

Pages 2258 à 2269 :

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

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <http://pubs.acs.org/cgi-bin/article.cgi/jprobs/2006/5/i09/pdf/pr060058y.pdf>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

4.5) Conclusions et perspectives

Lors de cette étude, nous avons analysé 39 sérums de patients atteints de LLC. Nous avons pu mettre en évidence trois groupes peptidiques exprimées de façon différentielle entre les patients et les contrôles. Ces différents peptides ont pu être purifiés et analysés par spectrométrie de masse. Nous avons pu montrer que la forme sulfitée de la transthyréline était augmentée dans les LLC alors que la forme cystéinillée et la forme normale étaient diminuées. Nous avons pu aussi identifier différents peptides de dégradation des molécules du complément C3 et C4. Ces différents marqueurs nous permettent de réaliser une classification avec une spécificité de 100%, et une sensibilité de 100%. Un arbre de classification est montré sur la figure 11.

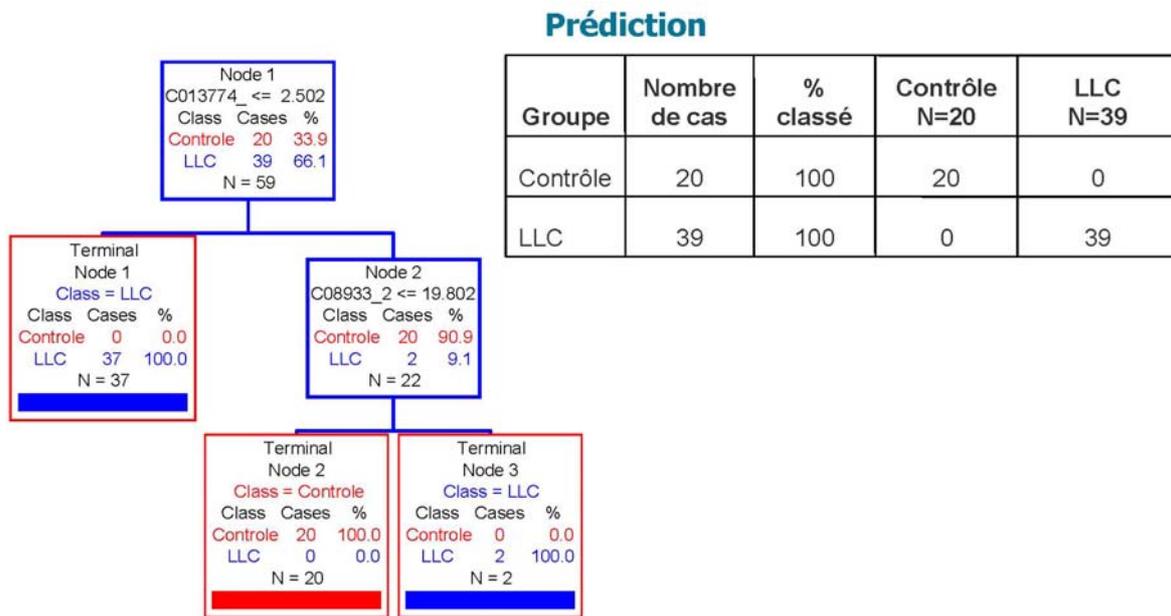


Figure 11 : Arbre de classification réalisé à partir des biomarqueurs mis en évidence lors de cette étude.

Les perspectives de ce travail sont maintenant de valider ce profil sur un plus grand nombre de patients, et de rechercher d'autres protéines qui présenteraient la même modification post-traductionnelle que celle observée sur la TTR (sulfitation de résidus cystéines).

5) Conclusions générales.

Lors de cette étude sur les hémopathies lymphoïdes chroniques, nous avons développé deux approches distinctes. L'approche protéomique nous a permis d'identifier un grand nombre de protéines différentes dont un nombre important provenait de la membrane plasmique. Parmi ces protéines, nous avons pu identifier la sorcin qui semble impliquée dans différents processus de résistances aux drogues. Cette protéine est à l'heure actuelle la cible d'analyses complémentaires pour valider notre hypothèse comme quoi la sorcin est exprimée de façon différentielle entre les différentes formes d'hémopathies lymphoïdes chroniques. Nous avons aussi mis en évidence le CD148, qui n'a été identifié que dans le lymphome du manteau, et qui est aussi en cours de validation. Cette approche est à élargir à d'autres formes d'hémopathies lymphoïdes chroniques et à des quantités de protéines légèrement plus importantes pour pouvoir réaliser trois injections successives afin d'avoir une vision plus globale du contenu protéique des différents échantillons, et donc de proposer des biomarqueurs potentiels d'une façon plus rationnelle.

Enfin, l'approche peptidomique a permis de réaliser un profil peptidique spécifique des hémopathies lymphoïdes chroniques. En effet, nous avons montré que trois peptides étaient augmentés de façon notable dans les hémopathies lymphoïdes chroniques. L'expression de ces trois protéines nous permet de réaliser une classification entre les LLC et les contrôles avec une spécificité et une sensibilité de 100%. Les perspectives de cette approche sont de réaliser des expériences complémentaires sur un plus grand nombre de sujets pour vérifier la sensibilité du classement proposé, et de comparer ces résultats avec d'autres pathologies afin de déterminer la spécificité des marqueurs proposés. Enfin, nous pouvons aussi essayer de déterminer des marqueurs plus spécifiques des différentes leucémies lympho-prolifératives chroniques.

Références bibliographiques

A

- Ahram, M. and Springer, D. L. (2004). "Large-scale proteomic analysis of membrane proteins." *Expert Rev Proteomics* **1**(3): 293-302.
- Aleksandrov, L., Mengos, A., Chang, X., Aleksandrov, A. and Riordan, J. R. (2001). "Differential interactions of nucleotides at the two nucleotide binding domains of the cystic fibrosis transmembrane conductance regulator." *J Biol Chem* **276**(16): 12918-23.
- Antin, J. H., Emerson, S. G., Martin, P., Gadol, N. and Ault, K. A. (1986). "Leu-1+ (CD5+) B cells. A major lymphoid subpopulation in human fetal spleen: phenotypic and functional studies." *J Immunol* **136**(2): 505-10.

B

- Banfi, C., Brioschi, M., Wait, R., Begum, S., Gianazza, E., Pirillo, A., Mussoni, L. and Tremoli, E. (2005). "Proteome of endothelial cell-derived procoagulant microparticles." *Proteomics* **5**(17): 4443-55.
- Bevers, E. M., Comfurius, P., Dekkers, D. W., Harmsma, M. and Zwaal, R. F. (1998). "Regulatory mechanisms of transmembrane phospholipid distributions and pathophysiological implications of transbilayer lipid scrambling." *Lupus* **7 Suppl 2**(2): S126-31.
- Blanchard, N., Lankar, D., Faure, F., Regnault, A., Dumont, C., Raposo, G. and Hivroz, C. (2002). "TCR activation of human T cells induces the production of exosomes bearing the TCR/CD3/zeta complex." *J Immunol* **168**(7): 3235-41.
- Bodnar, W. M., Blackburn, R. K., Krise, J. M. and Moseley, M. A. (2003). "Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage." *J Am Soc Mass Spectrom* **14**(9): 971-9.

C

- Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T. (1998). "New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis" *Electrophoresis* **19**(11): 1901-1909.

D

- Delalande, F., Carapito, C., Brizard, J. P., Brugidou, C. and Van Dorsselaer, A. (2005). "Multigenic families and proteomics: extended protein characterization as a tool for paralog gene identification." *Proteomics* **5**(2): 450-60.
- Dong, H. Y., Shahsafaei, A. and Dorfman, D. M. (2002). "CD148 and CD27 are expressed in B cell lymphomas derived from both memory and naive B cells." *Leuk Lymphoma* **43**(9): 1855-8.
- Durr, E., Yu, J., Krasinska, K. M., Carver, L. A., Yates, J. R., Testa, J. E., Oh, P. and Schnitzer, J. E. (2004). "Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture." *Nat Biotechnol* **22**(8): 985-92.

E

- Ek, S., Andreasson, U., Hober, S., Kampf, C., Ponten, F., Uhlen, M., Merz, H. and Borrebaeck, C. A. (2006). "From gene expression analysis to tissue microarrays: a rational approach to identify therapeutic and diagnostic targets in lymphoid malignancies." *Mol Cell Proteomics* **5**(6): 1072-81.

F

- Falini, B., Tiacci, E., Liso, A., Basso, K., Sabbatini, E., Pacini, R., Foa, R., Pulsoni, A., Dalla Favera, R. and Pileri, S. (2004). "Simple diagnostic assay for hairy cell leukaemia by immunocytochemical detection of annexin A1 (ANXA1)." *Lancet* **363**(9424): 1869-70.
- Flug, F., Dodson, L., Wolff, J., Guarini, L., Rausen, A., Wang, C. Y. and Knowles, D. M., 2nd (1985). "B-lymphocyte associated differentiation antigen expression by 'non-B, non-T' acute lymphoblastic leukemia." *Leuk Res* **9**(8): 1051-8.

G

- Galeva, N. and Altermann, M. (2002). "Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins." *Proteomics* **2**(6): 713-22.
- Garcia, B. A., Smalley, D. M., Cho, H., Shabanowitz, J., Ley, K. and Hunt, D. F. (2005). "The platelet microparticle proteome." *J Proteome Res* **4**(5): 1516-21.
- Guo, L., Eisenman, J. R., Mahimkar, R. M., Peschon, J. J., Paxton, R. J., Black, R. A. and Johnson, R. S. (2002). "A proteomic approach for the identification of cell-surface proteins shed by metalloproteases." *Mol Cell Proteomics* **1**(1): 30-6.

H

- Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A. and Burlingame, A. L. (2003). "Mass spectrometric analysis of protein mixtures at low levels using cleavable ¹³C-isotope-coded affinity tag and multidimensional chromatography." *Mol Cell Proteomics* **2**(5): 299-314.
- Harris, M. (2004). "Monoclonal antibodies as therapeutic agents for cancer." *Lancet Oncol* **5**(5): 292-302.
- Harris, N. L., Jaffe, E. S., Stein, H., Banks, P. M., Chan, J. K., Cleary, M. L., Delsol, G., De Wolf-Peeters, C., Falini, B., Gatter, K. C. and et al. (1994). "A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group." *Blood* **84**(5): 1361-92.
- Hassett, D. J., Schweizer, H. P. and Ohman, D. E. (1995). "Pseudomonas aeruginosa sodA and sodB mutants defective in manganese- and iron-cofactored superoxide dismutase activity demonstrate the importance of the iron-cofactored form in aerobic metabolism." *J Bacteriol* **177**(22): 6330-7.
- Heijnen, H. F., Schiel, A. E., Fijnheer, R., Geuze, H. J. and Sixma, J. J. (1999). "Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules." *Blood* **94**(11): 3791-9.
- Holder, M. J., Barnes, N. M., Gregory, C. D. and Gordon, J. (2006). "Lymphoma cells protected from apoptosis by dysregulated bcl-2 continue to bind annexin V in response to B-cell receptor engagement: a cautionary tale." *Leuk Res* **30**(1): 77-80.
- Hugel, B., Martinez, M. C., Kunzelmann, C. and Freyssinet, J. M. (2005). "Membrane microparticles: two sides of the coin." *Physiology (Bethesda)* **20**: 22-7.

J

- James, H. E., Beare, P. A., Martin, L. W. and Lamont, I. L. (2005). "Mutational analysis of a bifunctional ferrisiderophore receptor and signal-transducing protein from Pseudomonas aeruginosa." *J Bacteriol* **187**(13): 4514-20.
- Jang, J. H. and Hanash, S. (2003). "Profiling of the cell surface proteome." *Proteomics* **3**(10): 1947-54.
- Jensen, B. C., Wang, Q., Kifer, C. T. and Parsons, M. (2003). "The NOG1 GTP-binding protein is required for biogenesis of the 60 S ribosomal subunit." *J Biol Chem* **278**(34): 32204-11.

Jin, M., Drwal, G., Bourgeois, T., Saltz, J. and Wu, H. M. (2005). "Distinct proteome features of plasma microparticles." *Proteomics* **5**(7): 1940-52.

K

Kallstrom, G., Hedges, J. and Johnson, A. (2003). "The putative GTPases Nog1p and Lsg1p are required for 60S ribosomal subunit biogenesis and are localized to the nucleus and cytoplasm, respectively." *Mol Cell Biol* **23**(12): 4344-55.

Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F. and Oesterhelt, D. (2005). "The membrane proteome of *Halobacterium salinarum*." *Proteomics* **5**(1): 180-97.

Kratzer, R., Eckerskorn, C., Karas, M. and Lottspeich, F. (1998). "Suppression effects in enzymatic peptide ladder sequencing using ultraviolet - matrix assisted laser desorption/ionization - mass spectrometry." *Electrophoresis* **19**(11): 1910-9.

L

Lavabre-Bertrand, T., Janossy, G., Exbrayat, C., Bourquard, P., Duperray, C. and Navarro, M. (1994). "Leukemia-associated changes identified by quantitative flow cytometry. II. CD5 over-expression and monitoring in B-CLL." *Leukemia* **8**(9): 1557-63.

Levy, S. and Shoham, T. (2005). "The tetraspanin web modulates immune-signalling complexes." *Nat Rev Immunol* **5**(2): 136-48.

Lill, J. (2003). "Proteomic tools for quantitation by mass spectrometry." *Mass Spectrom Rev* **22**(3): 182-94.

Luche, S., Santoni, V. and Rabilloud, T. (2003). "Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis." *Proteomics* **3**(3): 249-53.

M

Matutes, E. (1995). "Contribution of immunophenotype in the diagnosis and classification of haemopoietic malignancies." *J Clin Pathol* **48**(3): 194-7.

Matutes, E., Owusu-Ankomah, K., Morilla, R., Garcia Marco, J. and Houlihan, A., Que, T. H., Catovsky, D. (1994). "The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL." *Leukemia* **8**(10): 1640-5.

Mesri, M. and Altieri, D. C. (1999). "Leukocyte microparticles stimulate endothelial cell cytokine release and tissue factor induction in a JNK1 signaling pathway." *J Biol Chem* **274**(33): 23111-8.

Miguet, L., Pacaud, K., Felden, C., Hugel, B., Martinez, M. C., Freyssinet, J. M., Herbrecht, R., Potier, N., van Dorsselaer, A. and Mauvieux, L. (2005). "Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization." *Proteomics* **6**(1): 153-171.

Miller, D. A. and Mahenthiralingam, E. (2003). "Sequencing of the *Pseudomonas aeruginosa* and *Burkholderia cepacia* genomes and their applications in relation to cystic fibrosis." *J R Soc Med* **96 Suppl 43**: 57-65.

O

Obritsch, M. D., Fish, D. N., MacLaren, R. and Jung, R. (2005). "Nosocomial infections due to multidrug-resistant *Pseudomonas aeruginosa*: epidemiology and treatment options." *Pharmacotherapy* **25**(10): 1353-64.

P

Peirce, M. J., Wait, R., Begum, S., Saklatvala, J. and Cope, A. P. (2004). "Expression profiling of lymphocyte plasma membrane proteins." *Mol Cell Proteomics* **3**(1): 56-65. Epub 2003 Oct 21.

Pilzer, D., Gasser, O., Moskovich, O., Schifferli, J. A. and Fishelson, Z. (2005). "Emission of membrane vesicles: roles in complement resistance, immunity and cancer." *Springer Semin Immunopathol* **27**(3): 375-87.

Q

Qoronfleh, M. W., Benton, B., Ignacio, R. and Kaboord, B. (2003). "Selective Enrichment of Membrane Proteins by Partition Phase Separation for Proteomic Studies." *J Biomed Biotechnol* **2003**(4): 249-255.

R

Rahbar, A. M. and Fenselau, C. (2004). "Integration of Jacobson's pellicle method into proteomic strategies for plasma membrane proteins." *J Proteome Res* **3**(6): 1267-77.

S

Salzer, U., Hinterdorfer, P., Hunger, U., Borken, C. and Prohaska, R. (2002). "Ca⁺⁺-dependent vesicle release from erythrocytes involves stomatin-specific lipid rafts, synexin (annexin VII), and sorcin." *Blood* **99**(7): 2569-77.

- Santoni, V., Molloy, M. and Rabilloud, T. (2000). "Membrane proteins and proteomics: un amour impossible?" *Electrophoresis* **21**(6): 1054-70.
- Sauer, M., Hantke, K. and Braun, V. (1987). "Ferric-coprogen receptor FhuE of *Escherichia coli*: processing and sequence common to all TonB-dependent outer membrane receptor proteins." *J Bacteriol* **169**(5): 2044-9.
- Schalk, I. J., Abdallah, M. A. and Pattus, F. (2002). "A new mechanism for membrane iron transport in *Pseudomonas aeruginosa*." *Biochem Soc Trans* **30**(4): 702-5.
- Schalk, I. J., Kyslik, P., Prome, D., van Dorsselaer, A., Poole, K., Abdallah, M. A. and Pattus, F. (1999). "Copurification of the FpvA ferric pyoverdine receptor of *Pseudomonas aeruginosa* with its iron-free ligand: implications for siderophore-mediated iron transport." *Biochemistry* **38**(29): 9357-65.
- Simak, J. and Gelderman, M. P. (2006). "Cell membrane microparticles in blood and blood products: potentially pathogenic agents and diagnostic markers." *Transfus Med Rev* **20**(1): 1-26.
- Siroy, A., Molle, V., Lemaitre-Guillier, C., Vallenet, D., Pestel-Caron, M., Cozzone, A. J., Jouenne, T. and De, E. (2005). "Channel formation by CarO, the carbapenem resistance-associated outer membrane protein of *Acinetobacter baumannii*." *Antimicrob Agents Chemother* **49**(12): 4876-83.
- Stewart, II, Zhao, L., Le Bihan, T., Larsen, B., Scozzaro, S., Figeys, D., Mao, G. D., Ornatsky, O., Dharsee, M., Orsi, C., Ewing, R. and Goh, T. (2004). "The reproducible acquisition of comparative liquid chromatography/tandem mass spectrometry data from complex biological samples." *Rapid Commun Mass Spectrom* **18**(15): 1697-710.

T

- Thery, C., Boussac, M., Veron, P., Ricciardi-Castagnoli, P., Raposo, G., Garin, J. and Amigorena, S. (2001). "Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles." *J Immunol* **166**(12): 7309-18.
- Tseng, C. F., Burger, A., Mislin, G. L., Schalk, I. J., Yu, S. S., Chan, S. I. and Abdallah, M. A. (2006). "Bacterial siderophores: the solution stoichiometry and coordination of the Fe(III) complexes of pyochelin and related compounds." *J Biol Inorg Chem* **11**(4): 419-32.

V

- Verzili, D., Zamparelli, C., Mattei, B., Noegel, A. A. and Chiancone, E. (2000). "The sorcin-annexin VII calcium-dependent interaction requires the sorcin N-terminal domain." *FEBS Lett* **471**(2-3): 197-200.

W

- Wallin, E. and von Heijne, G. (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." *Protein Sci* **7**(4): 1029-38.
- Wang, Y., Serfass, L., Roy, M. O., Wong, J., Bonneau, A. M. and Georges, E. (2004). "Annexin-I expression modulates drug resistance in tumor cells." *Biochem Biophys Res Commun* **314**(2): 565-70.
- Wubbolts, R., Leckie, R. S., Veenhuizen, P. T., Schwarzmann, G., Mobius, W., Hoernschemeyer, J., Slot, J. W., Geuze, H. J. and Stoorvogel, W. (2003). "Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation." *J Biol Chem* **278**(13): 10963-72. Epub 2003 Jan 7.

Z

- Zhang, N., Li, N. and Li, L. (2004). "Liquid chromatography MALDI MS/MS for membrane proteome analysis." *J Proteome Res* **3**(4): 719-27.
- Zhang, R., Sioma, C. S., Thompson, R. A., Xiong, L. and Regnier, F. E. (2002). "Controlling deuterium isotope effects in comparative proteomics." *Anal Chem* **74**(15): 3662-9.
- Zhao, Y., Zhang, W. and Kho, Y. (2004). "Proteomic analysis of integral plasma membrane proteins." *Anal Chem* **76**(7): 1817-23.
- Zhou, Y., Xu, Y., Tan, Y., Qi, J., Xiao, Y., Yang, C., Zhu, Z. and Xiong, D. (2006). "Sorcin, an important gene associated with multidrug-resistance in human leukemia cells." *Leuk Res* **30**(4): 469-76.

Conclusions générales

Conclusions générales

Lors de la première partie de ce manuscrit, nous avons brièvement rappelé les différents objectifs de l'approche protéomique et présenté les différents spectromètres de masse utilisés lors de nos études. Enfin, nous avons fait un récapitulatif des différentes techniques utilisées pour la préparation et la simplification des échantillons protéiques. Lors de cette partie introductive, nous avons aussi essayé de souligner les différents problèmes posés par l'étude des protéines membranaires (obtention de préparations membranaires contaminées par différentes protéines, incompatibilité avec les gels 2D...)

Cette étude bibliographique nous a permis de mettre en évidence les différents objectifs de notre travail de thèse :

- Le développement de nouvelles stratégies de quantification, applicable aux protéines membranaires, et qui ne présentent pas d'étapes de chromatographie d'affinité.
- Le développement de nouvelles méthodologies pour l'étude de protéines membranaires par nanoLC-MS/MS et LC-MALDI-MS.
- La recherche de nouveaux biomarqueurs de surface spécifiques de différentes pathologies.

La seconde partie de ce manuscrit est consacrée à la présentation des résultats obtenus lors de ce travail de thèse. Cette partie se présentait en trois chapitres.

- Le premier point abordé concerne la quantification de mélanges protéiques complexes issus des membranes externes de bactéries, ainsi que le développement et la validation d'un nouveau marqueur isotopique. Nous avons montré lors de cette étude que le marqueur développé au laboratoire pouvait être utilisé pour les analyses protéomiques, et que bien que peu compatible avec l'étude des protéines membranaires, il est tout à fait approprié pour l'étude de mélanges protéiques complexes

cytosolubles. Nous avons pu quantifier un grand nombre de protéines à partir d'un mélange protéique issu de cellules Hela.

- Le deuxième point de ce travail de thèse a été la mise en place un nouvel outil biologique pour l'étude des protéines issues de la membrane plasmique. Pour cela, nous avons réalisé le protéome de microparticules issues de la membrane plasmique de lymphocytes T. Nous avons pu montrer que le contenu protéique de ces structures est fixe et dépend du type cellulaire. En effet, nous avons identifié un grand nombre de clusters de différenciation des cellules spécifiques du type cellulaire étudié. De plus, nous avons comparé le protéome des microparticules au protéome obtenu lors de l'analyse d'une préparation membranaire plasmique classique. Nous avons pu montrer lors de cette comparaison que les microparticules sont des structures naturellement enrichies en protéines issues de la membrane plasmique et sont donc beaucoup plus faciles d'accès que les différentes préparations membranaires classiquement utilisées (centrifugation sur gradient de sucrose, biotinylation de surface...).

De plus, lors de cette étude, nous avons développé une nouvelle stratégie qui vise à combiner les données issues des analyses MALDI et ESI afin d'augmenter le pourcentage de recouvrement des différentes protéines identifiées. Cette stratégie se base sur la réalisation d'une chromatographie liquide couplée avec un microcolecteur. Les différentes fractions chromatographiques peuvent alors être analysées soit par MALDI-MS ou par ESI-MS.

- Le troisième point abordé lors de ce travail de thèse est la recherche et la mise en évidence de biomarqueurs spécifiques des différentes formes d'hémopathies lymphoïdes chroniques. Pour cela, deux approches ont été envisagées.

La première concerne l'approche SELDI-TOF qui nous a permis de mettre en évidence des protéines exprimées de façon différentielle entre les sérums de patients par rapport à des sérums contrôles. Nous avons pu montrer que la forme sulfitée de la transthyréline est augmentée dans les hémopathies lymphoïdes chroniques, ainsi que différents peptides de dégradation du complément C3 et C4. Bien que les protéines identifiées comme exprimées de façon différentielle ne soient pas très pertinentes d'un point de vue biologique, cette étude nous a tout de même permis de réaliser un arbre de classification qui présente une spécificité et une sensibilité de 100%.

La deuxième approche que nous avons utilisée vise à mettre en évidence des marqueurs de surface spécifiques des différentes hémopathies lymphoïdes chroniques. Pour cela, nous avons décidé de comparer les protéomes obtenus par l'analyse de préparation de microparticules issues de différents patients atteints de différentes formes d'hémopathies lymphoïdes chroniques. Bien que cette approche puisse paraître osée, car ne permettant pas de réaliser la quantification des différentes protéines identifiées, nous avons pu déterminer une liste de protéines identifiées spécifiquement dans une pathologie par rapport aux autres. Parmi ces protéines, nous avons sélectionné deux candidats qui ont été identifiés uniquement dans une pathologie et qui présentent un intérêt biologique important : la

sorcin, une molécule associée à la membrane et qui joue un rôle important dans les mécanismes de résistance aux drogues, et le CD148, un récepteur à activité tyrosine phosphatase qui présente un domaine transmembranaire. Nous réalisons actuellement des expériences complémentaires (RT-PCR et analyses par cytométrie en flux) afin de confirmer les différences d'expression et la spécificité de ces protéines en fonction des pathologies.

Ce travail de thèse permet aussi d'ouvrir de nouvelles perspectives dans le cadre de la recherche de biomarqueurs spécifiques des différentes formes d'hémopathies lymphoïdes chroniques.

Il s'agit tout d'abord d'augmenter le nombre d'échantillons à analyser pour réaliser le protéome des microparticules de chacune des pathologies. De plus comme nous l'avons montré lors du chapitre deux des résultats, il est possible d'obtenir une image globale du contenu protéique d'un échantillon en réalisant six analyses consécutives. En regroupant différentes préparations de microparticules produites à partir d'une même pathologie, il sera possible de réaliser plusieurs analyses successives et d'avoir une vision globale du protéome de ces microparticules. Les protéines qui seront alors identifiées spécifiquement dans une pathologie présenteront forcément de grandes variations d'expression entre les différentes pathologies.

Partie expérimentale

Partie expérimentale

1) Synthèse du marqueur NEIAM

La synthèse du NEIAM se réalise très facilement par l'addition lente d'anhydride iodoacétique (1,2 équivalent) sur une solution d'éthylamine (D0 ou D5) (0,9 équivalent) dans l'acétonitrile à 0°C, contenant 1,2 équivalent de diisopropyléthylamine. La solution est lentement portée à température ambiante et devient incolore. Au bout de 5 heures le solvant est évaporé. Le résidu est redissous dans du dichlorométhane et lavé avec une solution d'hydrogène sulfate d'ammonium à 1M. La solution aqueuse est ré extraite (volume à volume) par du dichlorométhane. La phase organique est alors séchée sur sulfate de magnésium, filtrée et évaporée à sec. Le résidu est purifié dans une solution d'éther diéthylique et d'héxane, offrant un cristal jaune pale.

2) Préparation des gels d'électrophorèse

Les gels d'électrophorèses sont des gels SDS-PAGE à gradient 8-15% en acrylamide de 20 cm.

GEL DE SEPARATION

	5%	6%	8%	10%	12%	15%	20%
H2O	5,7 ml	5,4 ml	4,7 ml	4,0 ml	3,4 ml	2,4 ml	0,7 ml
Tris 1,5 M pH 8,8	2,5 ml						
Acryl/Bis 30 %	1,7 ml	2,0 ml	2,7 ml	3,4 ml	4 ml	5 ml	6,7 ml
SDS 10 %	100 ul						
Temed	10 ul						
Persulfate NH4 10 %	25 ul						

GEL DE CONCENTRATION

H2O	1 ml	2,1 ml	3,5 ml	7 ml	10,5 ml
Tris 1 M pH 6,8	0,2 ml	0,4 ml	0,6 ml	1,2 ml	1,8 ml
Acryl/Bis 30 %	0,25 ml	0,5 ml	0,9 ml	1,8 ml	2,7 ml
SDS 10 %	15 ul	30 ul	50 ul	100 ul	150 ul
Temed	3 ul	6 ul	10 ul	20 ul	30 ul
Persulfate NH4 10 %	15 ul	30 ul	50 ul	100 ul	150 ul

La migration des gels se réalise pendant 5 heures avec un ampérage de 30mA. La coloration des gels est réalisée au bleu de coomassie R250 (sigma).

3) Protéomique

3.1) Réduction-alkylation-digestion

Pour l'étude protéomique des protéines séparées sur gels mono et bi-dimensionnels, nous avons mis au point un protocole de réduction-alkylation-digestion in gel. L'acquisition d'un robot MassPREP Station (Micromass, Manchester, UK), nous a permis d'automatiser cette étape contraignante et longue si elle est faite manuellement. Le protocole utilisé est basé sur les différentes méthodes utilisées avant l'achat de cet instrument. Les deux avantages principaux de ce robot sont d'une part le gain de temps considérable (5h00 pour une plaque de 96 spots contre 2 jours à la main, hors temps de digestion) et d'autre part une contamination des spots par la kératine totalement éradiquée. En effet, toutes les étapes étant automatisées, le risque de pollution lié à l'expérimentateur est minime, néanmoins cette contamination peut se produire lors de la préparation des diverses solutions et de la manipulation du gel.

Les étapes de lavages

Une fois séparées sur le gel d'électrophorèse, les protéines d'intérêts sont excisées du gel et placées dans une plaque à 96 puits. Les spots sont lavés par 100 µL avec un mélange 50/50 v/v ACN/NH₄HCO₃ 25mM pendant 10 min. Cette opération est répétée 2 fois. Les spots sont ensuite déshydratés par 50 µL d'ACN pendant 5 min.

La réduction et l'alkylation

La solution réductrice est une solution de dithiothreitol 10 mM dans NH₄HCO₃ 25mM. 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 55mM dans NH₄HCO₃ 25mM, 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 la trypsine

La digestion trypsique

Les protéines sont digérées à l'aide d'une solution de trypsine à 12,5 ng/µL dans NH₄HCO₃ 25mM. Le volume ajouté est fonction de la taille des spots (5-10µL). La réaction de digestion in gel se fait à température ambiante durant la nuit.

3.2) L'extraction des peptides.

L'étape d'extraction doit permettre la diffusion passive des peptides du gel vers la solution d'extraction. Cette étape est primordiale pour avoir un signal le plus intense possible lors de l'analyse par MS. Dans nos différentes analyses nous avons réalisé une extraction double :

Extraction des peptides pendant une heure à température ambiante avec une solution eau/ACN/acide formique (35%/60%/5%).

Extraction des peptides avec une solution ACN/acide Formique (95%/5%).

4) La chromatographie liquide

4.1) HPLC Waters Alliance (purification des protéines dans le cadre de l'étude SELDI)

Les molécules à analyser sont chromatographiées sur un système HPLC couplé à une pompe Waters 2690 de la série Alliance et à détecteur à barrette de diode Waters 996.

Colonne: Macherey Nagel Nucleosil 300-5C4 (1mm x 125mm; taille des particules 5 µm)

Gradient:

Temps en min	Solvant A (Eau TFA 0,1%)	Solvant B (ACN TFA 0,1%)
0	98	2
5	98	2
15	70	30
90	10	90
91	98	2
105	98	2

Ce système HPLC est couplé au spectromètre de masse LC-TOF (waters) décrit plus loin.

4.2) HPLC Agilent 1100 et collecteur de fraction Agilent 1100 (fractionnement des peptides pour les expériences LC-MALDI-MS)

L'HPLC Agilent 1100 comporte un programmeur de gradient piloté par le logiciel ChemStation (Agilent). 15µl des différents extraits peptidiques obtenus pour chacune des bandes de gel 1D sont injectées sur une colonne de 300µ (Zorbax 300SB-C18, 15 cm x 300 µm, 3.5µm).

Gradient :

Temps en min	Solvant A (Eau TFA 0,05%)	Solvant B (ACN TFA 0,045%)
0	90	10
5	90	10
45	50	50
46	80	20
50	80	20
51	10	90
65	10	90

Le microcollecteur permet de collecter les différentes fractions d'élutions avant d'être analysées par MALDI-MS.

4.3) CapLC Micromass

Le système HPLC utilisé est une CapLC (Waters) 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 µm de diamètre interne remplie d'une phase stationnaire de type PepMap C18 de granulométrie 5 µm et de 100 Å de porosité (LC-Packings) à un débit de 30 µl/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" (Waters) 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 15 cm long et 75 µm de diamètre interne remplie d'une phase stationnaire de type PepMap C18 de granulométrie 3µm et de porosité 100 Å (LC-Packings) (figure 1).

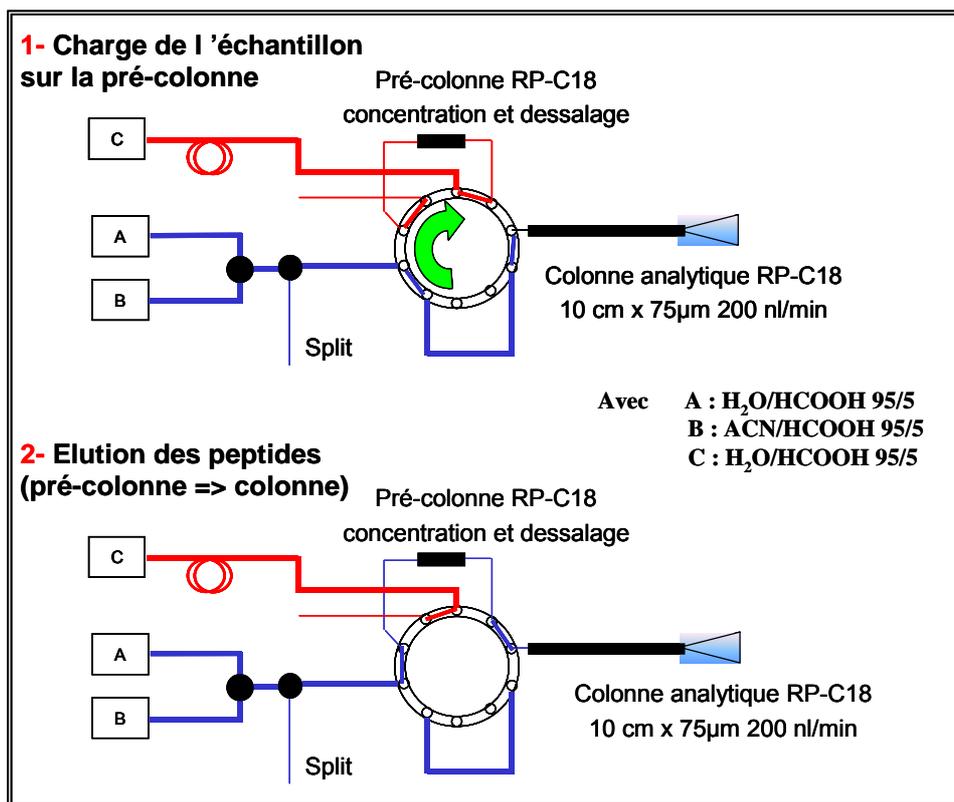


Figure 1 : Schéma représentant les deux étapes de la LC-MS-MS. 1 : chargement de l'échantillon sur la pré-colonne et lavage, 2 : élution des peptides retenus sur la pré-colonne et séparation de ces peptides sur la colonne chromatographique.

Gradient :

Temps en min	Solvant A (Eau HCOOH 0,1%)	Solvant B (ACN HCOOH 0,1%)
0	90	10
50	40	60
51	95	5
55	95	5
56	10	90
70	10	90

5) Les spectromètres de masse

5.1) Le MALDI-TOF

Toutes les mesures de masses par MALDI-TOF ont été réalisées sur un Ultraflex (Bruker). Le logiciel d'acquisition des spectres est Flexcontrol 2.2. Le logiciel d'exploitation des données est FlexAnalysis 2.2. L'Ultraflex est équipé d'une source SCOUT. La désorption/ionisation est réalisée par irradiation laser, avec un laser à azote de longueur d'onde 337 nm et de fréquence 20 Hz.

L'énergie (fluence) du laser peut varier de 150 à 200 μJ . La puissance est modulable grâce à un atténuateur. Une caméra permet de visualiser le dépôt et les impacts des tirs lasers. Un spectre de masse est généré à chaque tir laser, l'accumulation de plus de 100 spectres étant parfois nécessaire pour avoir un bon rapport signal sur bruit.

Les paramètres importants de l'analyse par MALDI-TOF-MS :

Pour la précision et la résolution (analyses protéomiques, peptides <5kDa) : l'acquisition se fait en mode réflecteur. Les tensions appliquées aux lentilles d'accélération et de focalisation sont : IS1 : 20kV ; IS2 : 17,5kV ; lens : 5kV et reflecteur : 21kV.

Le délai d'extraction peut être augmenté ou diminué en fonction de l'analyse. Il est en général «court» pour l'analyse des peptides et «long» pour l'analyse des protéines (en centaines de ns). L'augmentation de ce délai permet de gagner en résolution pour la mesure de masses moléculaires de protéines supérieures à 20 kDa. En deçà de 20 kDa, on observe plutôt un élargissement des pics à mi-hauteur.

L'étalonnage du spectromètre de masse MALDI-TOF :

Pour les peptides de 700 à 2500 Da : nous avons utilisé les fragments d'autolyse de la trypsine à 842,5 et à 2211,10 m/z mono-isotopiques pour les peptides issus de digestion trypsique, dans les autres cas nous avons utilisé un mélange peptidique : 712,38 m/z (leu-enképhaline) ; 1046,54 m/z (angiotensine) ; 1348,72 m/z (substance P) ; 1620,81 m/z (bombésine) ; 2004,95 m/z (Druuna).

L'étalonnage interne est utilisé pour un maximum de précision de mesure. L'échantillon à analyser est incorporé au mélange étalon. Le tout est co-cristallisé avec la matrice. L'assignation des masses des composés standards permet de calculer les constantes de calibration. Ces constantes sont ensuite utilisées pour le calcul des masses des composés inconnus. Le principal inconvénient de cette méthode est le risque de suppression des ions analytes par une trop forte concentration du mélange étalon. Lors de la mesure de masse de digest trypsique la calibration interne est de fait automatique.

La préparation de l'échantillon

La matrice la plus couramment utilisée pour nos analyses par MALDI-MS a été l'acide α -cyano-4-hydroxynamique (HCCA). Le mode de dépôt utilisé est le dépôt en sandwich (figure 2) ou le dépôt goutte séchée.

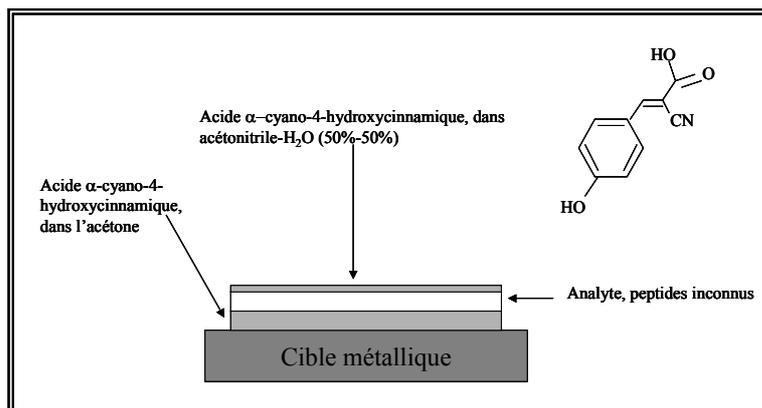


Figure 2A : Modes de dépôt en sandwich

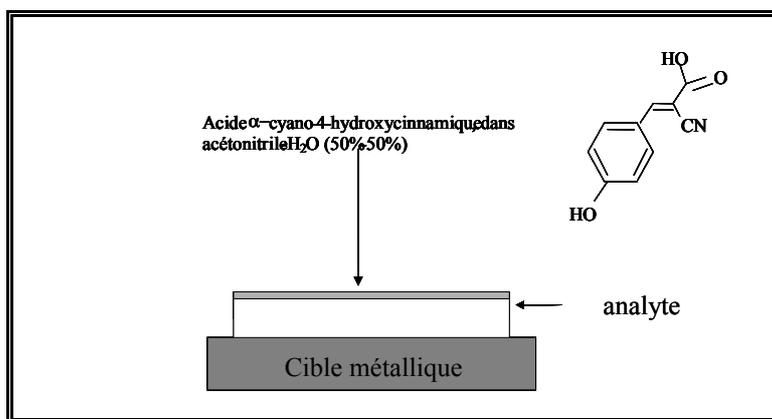


Figure 2B : Modes de dépôt en goutte séchée

En comparant différentes matrices avec différents modes de dépôts, nous avons opté pour une préparation en sandwich avec la matrice HCCA pour les raisons suivantes :

L'HCCA est la matrice indiquée pour les analyses de peptides sur une assez large gamme de masse, et permet d'effectuer des mesures de masse sans discriminer les petits peptides des plus gros.

La préparation en sandwich comme en goutte séchée se prête bien au lavage de la cible, cette étape de dessalage améliore également la qualité de l'analyse. De plus, la double couche de matrice du dépôt sandwich permet d'obtenir un dépôt plus homogène.

Lorsque le signal est faible, la seule façon d'obtenir in fine une intensité suffisante susceptible de donner une mesure de masse fiable est d'accumuler les signaux. Cette accumulation n'est possible qu'avec des pics qui se superposent parfaitement. Par conséquent, la qualité du dépôt sur cible est un point important pour la réalisation d'une mesure de masse précise. L'homogénéité du dépôt permet des mesures reproductibles donc la possibilité et l'avantage de pouvoir accumuler les signaux lorsque ceux-ci sont de faible intensité. Plus la protéine est présente en faible quantité plus la qualité de préparation de l'échantillon prend de l'importance. Il s'agit de trouver un compromis entre la matrice idéale et le mode de dépôt pour analyser un maximum de peptides.

5.2) Les spectromètres de masse électrospray

Trois spectromètres de masse électrospray différents ont été utilisés : un temps de vol, un hybride quadripôle-temps de vol et une trappe ionique.

5.2.1) Le spectromètre de masse LC-TOF.

Le dosage de molécules pharmacologiques issues de fluides biologiques a été effectué sur le ES-TOF (LCT, Waters, UK). L'acquisition des spectres de masse ainsi que le traitement des données est réalisé par le logiciel MassLynx, version 4.0 (Waters). Le TOF mesure environ 80 cm aller-retour. La tension de cône est réglée en fonction de l'étude et oscille entre 20 et 50V. La tension appliquée sur la canne d'introduction ESI est de l'ordre de 3000-3500V.

5.2.2) Le spectromètre de masse Q-TOF

Sur le Q-TOF II, trois sortes de source sont adaptables : la source électrospray (ESI), la source nanospray et la source LC-nanospray pour le couplage nanoLC-MS.

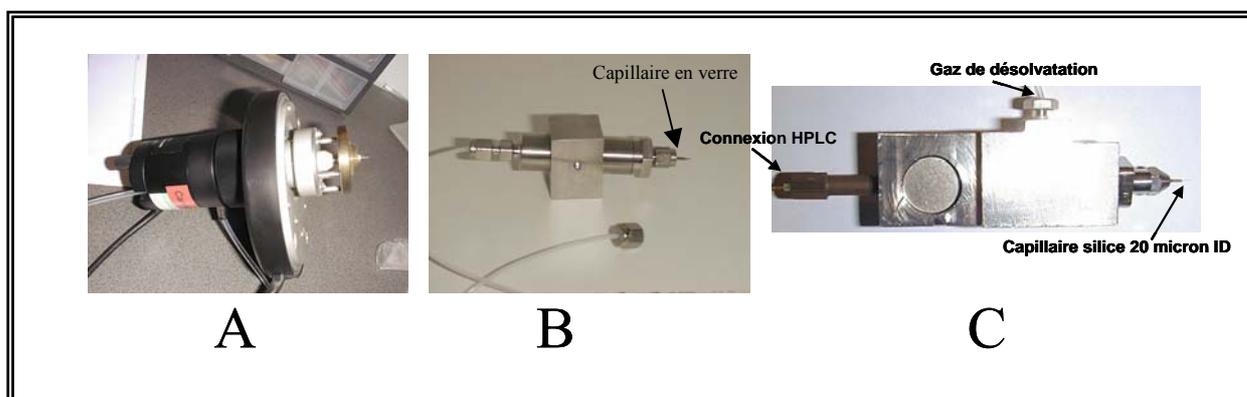


Figure 3: Représentation des sources adaptables sur le Q-TOF II Micromass.

Source	Débit	Tension du capillaire
Electrospray	4 à 50 μ l/min	3500 V
Nanospray (aiguille)	200 nl/min	1100 V
LC-nanospray	200 nl/min	3500 V

Tableau 1 : Tableau des débits et des tensions utilisés.

L'acquisition des spectres de masse ainsi que le traitement des données sont réalisés par le logiciel MassLynx, version 4.0 (Waters). Le Q-TOF II est un instrument hybride et, comme nous l'avons vu dans les chapitres précédents, il allie les avantages des analyseurs quadripolaires (possibilité de

sélectionner très précisément un ion) et ceux des analyseurs TOF (sensibilité, excellente résolution grâce à un TOF de 1.80 m aller-retour (2x90 cm)). (cf. schémas des appareils Partie Introduction, Chapitre II)

Introduction des échantillons pour les analyses nano-ESI.

La canne d'introduction de l'échantillon utilisée pour les analyses ESI-MS, est remplacée par un montage portant le capillaire nano-ESI en verre (Protana, Odensee, Danemark) et contenant l'analyte (1-5µl) (figure 3B). Il est placé perpendiculairement à l'entrée du spectromètre de masse, juste devant le cône d'échantillonnage (source Z-spray, nano-ESI). Un tube téflon amène le gaz de nébulisation au capillaire. Un manchon conducteur permet d'appliquer la haute-tension au capillaire métallisé. Lorsque le capillaire est bouché, l'augmentation de la pression d'azote permet de réamorcer le spray. Il est possible de travailler pendant environ 1 heure sur un échantillon de 2 µl grâce au nano-débit.

Paramètres des analyses MS et MSMS en nanospray :

La tension appliquée au capillaire nano-spray est de l'ordre de 1000 V. La différence de potentiel générée entre le capillaire et le cône d'échantillonnage (30V), entraîne l'échantillon ionisé dans le premier analyseur. Lorsqu'il est utilisé en mode MS simple, le quadripôle joue le rôle de filtre de masse (RF only), et laisse passer tous les ions vers le deuxième analyseur (TOF). En mode MSMS, un courant continu (DC) est appliqué, permettant l'isolement d'un ion précurseur. Juste avant la cellule de collision, où règne une pression d'argon de 0.1 mbar, une énergie est appliquée à l'ion (entre 10 et 80 eV) et entraîne sa fragmentation dans la cellule de collision. Les fragments générés sont analysés par le TOF.

5.3) Le SELDI

Les analyses SELDI ont été réalisées sur un PBS IIC. Les différentes surfaces chromatographiques utilisées sont les échangeuses anioniques et cationiques (Q10 et CM10), les surfaces hydrophobes (H50), les surfaces d'affinité métallique (IMAC) et les surfaces hydrophiles (NP20).

5.3.1) Préparation des échantillons

Les différents sérums ont été dénaturés dans un tampon CHAPS 1% 9M urée avant d'être dilués au 1/100 dans les différents tampons de liaison :

- CM10 : 100mM acétate de sodium à pH 4 ou pH 6
- Q10 : 100mM tampon Tris pH 9 ou pH 7

- H50 : 10% CH₃CN 0,1% TFA 150mM NaCl
- IMAC Cu : 100mM phosphate de sodium 0,5M NaCl pH 7

Les échantillons sont déposés sur les barrettes chromatographiques pendant une heure avant d'être lavés en utilisant le même tampon de liaison, avant d'être rincés à l'eau MilliQ. Les protéines sont désorbées par l'addition de matrice (acide sinapinic) avec 0,5% TFA (figure3).

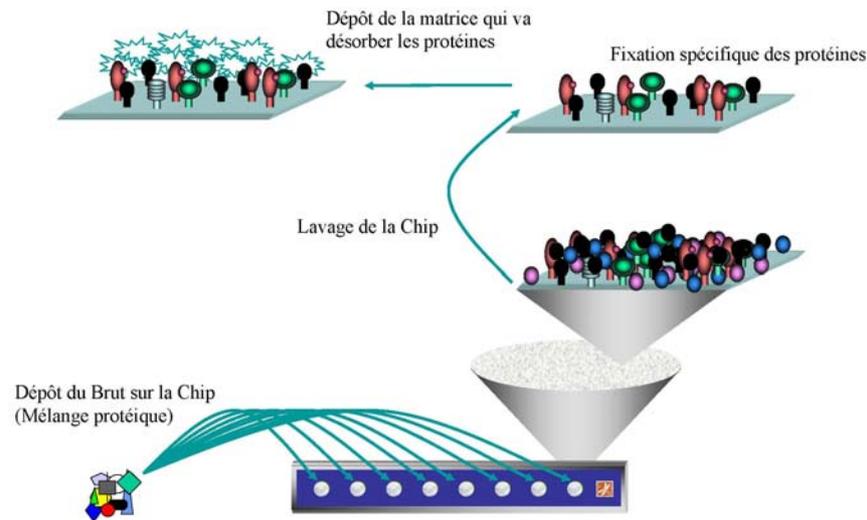


Figure 3 : Représentation schématique de la préparation de l'échantillon pour l'analyse SELDI_TOF

5.3.2) Acquisition des données SELDI

Pour l'acquisition des spectres SELDI, trois paramétrages ont été utilisés :

- Petites masses : Intensité du laser à 160, focus à 2000
- Masses moyennes : Intensité du laser à 180, focus à 8000
- Hautes masses : Intensité du laser à 200, focus à 13000

Les acquisitions sont pilotées par le logiciel ProteinChip Software 3.2.1

5.3.3) traitement statistique des données

Les données ont été traitées soit par le logiciel ProteinChip software soit par CIPHERGENExpress 2.1. Les clusters de pics ont été générés à partir de pics présentant un rapport signal/bruit supérieur à 5 et une tolérance de masse de 0,2% du pic. Les calculs de p-value ont été réalisés sur les spectres normalisés en utilisant un test de Mann-Whitney pour des données non paramétriques.

6) Recherche dans les banques de données en utilisant l'algorithme MASCOT

MASCOT MS/MS Ions Search

Your name	<input type="text"/>	Email	<input type="text"/>
Search title	<input type="text"/>		
Database	Sprot		
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)	Variable modifications	N-Acetyl (Protein) N-Formyl (Protein) NIPCAM (C) O18 (C-term) Oxidation (M)
Protein mass	<input type="text"/> kDa	ICAT	<input type="checkbox"/>
Peptide tol.	± 0.25 Da	MS/MS tol.	± 0.25 Da
Peptide charge	1+, 2+ and 3+	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Data file	<input type="text"/>	Parcourir...	
Data format	Micromass (.PKL)	Precursor	<input type="text"/> m/z
Instrument	ESI-QUAD-TOF		
Overview	<input type="checkbox"/>	Report top	50 hits
<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>	

Figure 4 : Vue de la fenêtre MASCOT. Les critères de recherche important sont entourés.

Les paramètres importants sont :

- Le choix de la banque protéique. Pour nos études, nous avons majoritairement travaillé avec la banque SwissProt qui est non redondante et assez complète par rapport aux protéines humaines.
- Le nombre de sites de coupure manqués par la trypsine (missing cleavage). De façon générale, ce paramètre est porté à 1.
- La taxonomie. Toutes nos recherches sont faites sans restriction de taxonomie.
- Les modifications : N-acétylation de la protéine, oxydation des méthionines, carbamidométhylation des cystéines.
- La tolérance sur la masse des peptides parents et des fragments. Dans nos études avec des données Q-TOF, nous acceptons une erreur de 0,25 Da sur l'ion parent et sur les fragments.

Annexes

Annexes

1) Tableau des protéines identifiées en fonction des pathologies.

Nom de la protéine	Numéro d'accèsion	LAL	LLC	MCL	LNH
10 kDa heat shock protein, mitochondrial (Hsp10) (10 kDa chaperonin)	P61604			X	
130 kDa leucine-rich protein (LRP 130) (GP130)	P42704				X
14-3-3 protein beta/alpha (Protein kinase C inhibitor protein 1)	P31946	X	X	X	X
14-3-3 protein epsilon (14-3-3E)	P62258	X	X	X	
14-3-3 protein eta (Protein AS1)	Q04917		X		
14-3-3 protein gamma (Protein kinase C inhibitor protein 1) (KCIP-1)	P61981	X	X	X	
14-3-3 protein sigma (Stratifin) (Epithelial cell marker protein 1)	P31947		X		
14-3-3 protein tau (14-3-3 protein theta) (14-3-3 protein T-cell) (HS1 protein)	P27348				X
14-3-3 protein theta (14-3-3 protein tau) (14-3-3 protein T-cell) (HS1 protein)	P27348	X	X	X	
14-3-3 protein zeta/delta (Protein kinase C inhibitor protein 1) (KCIP-1)	P63104	X	X	X	X
1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma 2	P16885	X	X	X	X
2,4-dienoyl-CoA reductase, mitochondrial precursor	Q16698		X	X	X
26S protease regulatory subunit 6B (MIP224)	P43686		X		
26S proteasome non-ATPase regulatory subunit 1	Q99460		X		
26S proteasome non-ATPase regulatory subunit 13	Q9UNM6	X			
26S proteasome non-ATPase regulatory subunit 2	Q13200	X	X	X	X
26S proteasome non-ATPase regulatory subunit 5	Q16401	X	X		
26S proteasome non-ATPase regulatory subunit 6	Q15008		X		
26S proteasome non-ATPase regulatory subunit 7	P51665		X		X
3'(2'),5'-bisphosphate nucleotidase 1	O95861		X		
3,2-trans-enoyl-CoA isomerase, mitochondrial precursor	P42126			X	
3-ketoacyl-CoA thiolase, mitochondrial	P42765			X	
40S ribosomal protein S10	P46783	X	X	X	X
40S ribosomal protein S11	P62280		X		
40S ribosomal protein S13	P62277	X	X	X	X
40S ribosomal protein S14	P62263	X	X		
40S ribosomal protein S15 (RIG protein)	P62841		X	X	
40S ribosomal protein S15a	P62244	X	X	X	X
40S ribosomal protein S16	P62249	X	X	X	
40S ribosomal protein S17	P08708	X	X		
40S ribosomal protein S18 (Ke-3) (Ke3)	P62269	X	X	X	
40S ribosomal protein S19	P39019	X	X		
40S ribosomal protein S2 (S4) (LLRep3 protein)	P15880	X	X	X	X
40S ribosomal protein S20	P60866		X		
40S ribosomal protein S23	P62266	X		X	
40S ribosomal protein S24	P62847	X		X	
40S ribosomal protein S24 (S19)	P16632				X
40S ribosomal protein S25	P62851	X	X	X	
40S ribosomal protein S26	P62854	X	X	X	
40S ribosomal protein S28	P62857			X	
40S ribosomal protein S3	P23396		X	X	X
40S ribosomal protein S3a	P61247			X	X
40S ribosomal protein S4, X isoform	P62701	X	X	X	X
40S ribosomal protein S5	P46782		X	X	X
40S ribosomal protein S6 (Phosphoprotein NP33)	P62753		X	X	X
40S ribosomal protein S7	P62081	X	X	X	
40S ribosomal protein S7 (S8)	P23821				X

40S ribosomal protein S8	P62241	X	X	X	X
40S ribosomal protein S9	P46781				X
40S ribosomal protein SA (p40) (34/67 kDa laminin receptor)	P08865	X	X		
40S ribosomal protein SA (p40) (C10 protein)	P26452		X		
4F2 cell-surface antigen heavy chain (4F2hc)	P08195	X			
4-trimethylaminobutyraldehyde dehydrogenase	P49189		X	X	
5'(3')-deoxyribonucleotidase, cytosolic type	Q8TCD5		X		
60 kDa heat shock protein, mitochondrial precursor	P10809		X	X	X
60S acidic ribosomal protein P0 (L10E)	P05388	X	X	X	X
60S acidic ribosomal protein P1	P05386			X	
60S ribosomal protein L10a (CSA-19)	P62906		X		
60S ribosomal protein L10-like	Q96L21		X		
60S ribosomal protein L11 (CLL-associated antigen KW-12)	P62913	X	X	X	
60S ribosomal protein L12	P30050		X		
60S ribosomal protein L13 (Breast basic conserved protein 1)	P26373		X		X
60S ribosomal protein L13a (23 kDa highly basic protein)	P40429		X		X
60S ribosomal protein L15	P61313	X			X
60S ribosomal protein L17 (L23)	P18621	X	X	X	X
60S ribosomal protein L18	Q07020	X	X		X
60S ribosomal protein L18a	Q02543	X	X	X	X
60S ribosomal protein L19	P84098	X			X
60S ribosomal protein L21	P46778	X	X		X
60S ribosomal protein L22	P35268	X			
60S ribosomal protein L23 (Ribosomal protein L17)	P62829		X	X	
60S ribosomal protein L23a	P62750	X	X		
60S ribosomal protein L24 (L30)	P83731				X
60S ribosomal protein L26	P61254	X			X
60S ribosomal protein L26-like 1	Q9UNX3		X	X	
60S ribosomal protein L27	P08526				X
60S ribosomal protein L27a	P46776	X	X	X	X
60S ribosomal protein L28	P46779				X
60S ribosomal protein L3	P39023		X		
60S ribosomal protein L30	P62888	X	X		
60S ribosomal protein L31	P62899	X		X	
60S ribosomal protein L32	P02433				X
60S ribosomal protein L36	Q9Y3U8				X
60S ribosomal protein L38	P63173			X	
60S ribosomal protein L4 (L1)	P36578		X		
60S ribosomal protein L6	Q02878			X	X
60S ribosomal protein L7	P18124	X	X	X	X
60S ribosomal protein L7a	P62424		X		
60S ribosomal protein L8	P62917		X		X
60S ribosomal protein L9	P32969	X	X		X
6-phosphofructokinase type C	Q01813		X		
6-phosphofructokinase, liver type	P17858	X	X	X	X
6-phosphofructokinase, type C	Q01813			X	X
6-phosphogluconate dehydrogenase	P52209	X	X	X	
6-phosphogluconolactonase (EC 3.1.1.31) (6PGL)	O95336	X	X		
70 kDa WD-repeat tumor rejection antigen homolog	P57737				X
Acetyl-CoA acetyltransferase, mitochondrial precursor	P24752		X		X
Acid ceramidase precursor	Q13510		X		
Acidic leucine-rich nuclear phosphoprotein 32 family member A	P39687	X	X	X	
Aconitate hydratase, mitochondrial precursor	Q99798		X	X	X
Actin, aortic smooth muscle (Alpha-actin 2)	P03996			X	X
Actin, cytoplasmic 1 (Beta-actin)	P60709	X	X	X	X
Actin, cytoplasmic 2 (Gamma-actin)	P63261	X	X	X	X
Actin-like protein 2 (Actin-related protein 2)	P61160	X	X	X	X
Actin-like protein 3 (Actin-related protein 3)	P61158	X	X	X	
Actin-like protein 3 (Actin-related protein 3) (Actin-2)	P61157		X		X

Actin-related protein 2/3 complex subunit 1B	O15143	X	X	X	
Actin-related protein 2/3 complex subunit 2	O15144	X	X	X	
Actin-related protein 2/3 complex subunit 3	O15145	X	X	X	
Actin-related protein 2/3 complex subunit 4	P59998	X	X	X	
Actin-related protein 2/3 complex subunit 5	O15511		X		
Acylamino-acid-releasing enzyme (EC 3.4.19.1)	P13798	X	X	X	
Acyl-CoA dehydrogenase, medium-chain specific	P11310				X
Adapter-related protein complex 1 beta 1 subunit	Q10567				X
Adapter-related protein complex 1 gamma 1 subunit	O43747		X		X
Adaptor-related protein complex 1	Q9BXS5		X		
Adenine phosphoribosyltransferase	P07741	X	X	X	X
Adenosylhomocysteinase (EC 3.3.1.1)	P23526	X	X	X	X
Adenylate kinase isoenzyme 2, mitochondrial	P54819	X	X	X	X
Adenylosuccinate lyase (EC 4.3.2.2)	P30566	X			
Adenylyl cyclase-associated protein 1	Q01518	X	X	X	X
ADP/ATP translocase 3	P12236		X		
ADP-ribosylation factor 1	P84080	X	X	X	X
ADP-ribosylation factor 3	P61204		X		
ADP-ribosylation factor 4	P18085			X	X
ADP-ribosylation factor 5	P84085			X	X
ADP-ribosylation factor 6	P62330	X	X		X
ADP-sugar pyrophosphatase	Q9UKK9		X		
Alanyl-tRNA synthetase	P49588				X
Alcohol dehydrogenase	P14550		X		
Alpha enolase (EC 4.2.1.11)	P06733			X	X
Alpha-actinin 1	P12814		X	X	X
Alpha-actinin 4	O43707	X	X	X	X
Alpha-centractin (Centractin)	P42024				X
Alpha-enolase (EC 4.2.1.11)	P06733	X	X		
Amidophosphoribosyltransferase precursor (EC 2.4.2.14)	Q06203		X		
Aminopeptidase B (EC 3.4.11.6) (Ap-B)	Q9H4A4				X
Angiomotin	Q4VCS5	X			
Annexin A1	P04083	X	X	X	
Annexin A11	P50995		X	X	X
Annexin A2	P07355	X	X	X	X
Annexin A4	P50994	X	X		X
Annexin A5	P08758	X	X	X	X
Annexin A6	P08133	X	X	X	X
Annexin A7 (Annexin VII) (Synexin)	P20073	X		X	
Antibacterial protein FALL-39 precursor	P49913	X			
Apoptosis-associated speck-like protein	Q9ULZ3	X	X		
Arachidonate 5-lipoxygenase	P09917		X	X	X
A-Raf proto-oncogene serine/threonine-protein kinase	P10398				X
ARF GTPase-activating protein GIT2	Q14161	X			
Arginyl-tRNA synthetase	P54136		X		
ARP2/3 complex 20 kDa subunit	P59998				X
ARP2/3 complex 21 kDa subunit	O15145				X
ARP2/3 complex 34 kDa subunit	O15144				X
ARP2/3 complex 41 kDa subunit	O15143				X
Asparaginyl-tRNA synthetase	O43776			X	
Aspartate aminotransferase	P17174	X			
Aspartate aminotransferase	P00505		X	X	X
Aspartyl aminopeptidase	Q9ULA0	X	X	X	
Aspartyl-tRNA synthetase	P14868		X	X	X
ATP synthase alpha chain	P25705	X	X		X
ATP synthase beta chain	P06576	X	X	X	X
ATPase family AAA domain containing	Q5T9A4	X			
ATP-citrate synthase	P53396	X	X	X	X
ATP-dependent DNA helicase II, 70 kDa subunit	P12956		X	X	X

ATP-dependent DNA helicase II, 80 kDa subunit	P13010		X	X	X
ATP-dependent helicase DDX1	Q92499	X	X	X	X
ATP-dependent helicase DDX39	O00148	X	X		
ATP-dependent RNA helicase A	Q08211		X	X	X
B cell-stimulating factor 3 precursor	Q9UBD9	X			
B-cell receptor CD22 precursor	P20273			X	
Beta platelet-derived growth factor receptor precursor	P09619	X			
Beta-2-microglobulin precursor	P61769	X		X	
Bifunctional aminoacyl-tRNA synthetase	P07814			X	X
Bifunctional purine biosynthesis protein PURH	P31939	X	X	X	X
B-lymphocyte antigen CD20	P11836		X	X	X
Brain acid soluble protein 1	P80723			X	
Brain-specific angiogenesis inhibitor 1 precursor	O14514	X			
Bullous pemphigoid antigen 1, isoforms 6/9/10	O94833	X			
C-1-tetrahydrofolate synthase, cytoplasmic	P11586	X		X	X
CAD protein	P27708				X
Calcium/calmodulin-dependent protein kinase type II	Q13557	X		X	
Calcium-binding protein 39	Q9Y376		X		
Calcium-regulated heat stable protein 1	Q9Y2V2			X	
Calgranulin A	P05109	X			
Calgranulin B	P06702	X			
Calmodulin (CaM)	P62158		X	X	
Calpain 1, large	P07384				X
Calpain 2, large	P17655				X
Calpain small subunit 1	P04632	X	X		
Calpain-1 catalytic subunit	P07384	X		X	
cAMP-dependent protein kinase type I-	P10644	X			
cAMP-dependent protein kinase type II-alpha	P13861		X		
cAMP-dependent protein kinase	P17612	X			X
<i>Carboxypeptidase D precursor</i>	O75976		X		X
Casein kinase II, alpha'	P19784				X
Catalase (EC 1.11.1.6)	P04040	X	X	X	
Cathepsin G precursor (EC 3.4.21.20) (CG)	P08311	X			
Cation-independent mannose-6-phosphate receptor precursor	P11717		X		X
CCG1-interacting factor B	Q96IU4		X		X
CD2-associated protein	Q9Y5K6	X			
CD44 antigen precursor	P16070	X	X	X	
CD81 antigen	P60033			X	
CD82 antigen	P27701			X	
CD97 antigen precursor (Leukocyte antigen CD97)	P48960	X			
Cell division control protein 42 homolog	Q90694		X		
Chloride intracellular channel protein 1	O00299	X	X	X	X
Chromobox protein homolog 3	Q13185			X	
Ciliary dynein heavy chain 11	Q96DT5		X		
Citrate synthase	O75390		X	X	X
Clathrin heavy chain	P49951	X	X		
Clathrin heavy chain 1 (CLH-17)	Q00610	X	X	X	X
Clathrin heavy chain 2 (CLH-22)	P53675				X
c-Myc-responsive protein Rcl	O43598		X		
Coactosin-like protein	Q14019		X	X	X
Coagulation factor XIII A chain precursor	P00488	X			
Coatomer alpha subunit	P53621		X	X	X
Coatomer beta subunit	P53618		X	X	X
Coatomer beta' subunit	P35606			X	
Coatomer delta subunit	P48444		X		
Coatomer gamma subunit	Q9Y678		X	X	X
Coatomer gamma-2 subunit	Q9UBF2		X		
Coatomer zeta-1 subunit (Zeta-1 coat protein) (Zeta-1 COP)	P61923	X			
Cofilin, non-muscle isoform (18 kDa phosphoprotein) (P18)	P23528				X

Cofilin-1	P23528	X	X	X	
<i>Complement C5 precursor</i>	<i>P06684</i>				X
Copine III	O75131				X
Copine-1 (Copine I)	Q99829		X	X	
Copine-2 (Copine II)	Q96FN4	X	X		
Copine-3 (Copine III)	O75131	X		X	
Coronin 1A (Coronin-like protein p57)	P31146	X	X	X	X
CTP synthase (EC 6.3.4.2)	P17812	X			
Cullin homolog 4A (CUL-4A)	Q13619				X
Cullin-2 (CUL-2)	Q13617		X		
Cullin-3 (CUL-3)	Q13618		X		
Cullin-4A (CUL-4A)	Q13619		X	X	
Cullin-associated NEDD8-dissociated protein 1	Q86VP6	X	X	X	
Cyclin A1	P78396		X		
Cystatin B	P04080	X			
CysteinyI-tRNA synthetase	P49589				X
Cytochrome c	P99999			X	
Cytoplasmic antipeptidase 3	P50453	X	X	X	
Cytosol aminopeptidase	P28838	X	X		
Cytosolic nonspecific dipeptidase	Q96KP4	X	X	X	X
D-dopachrome tautomerase	P30046	X			
Dedicator of cytokinesis protein 10	Q96BY6				X
Dedicator of cytokinesis protein 2	Q92608		X		
Dedicator of cytokinesis protein 8 (Fragment)	Q8NF50		X		X
Desmin	P17661				X
Diablo homolog, mitochondrial precursor	Q9NR28		X	X	X
Diacylglycerol kinase, alpha	P23743		X		X
Diaphanous protein homolog 1	O60610		X	X	X
Dihydrolipoyl dehydrogenase, mitochondrial precursor	P09622		X		
Dihydropyrimidinase related protein-2	Q16555	X		X	X
Dipeptidyl-peptidase III	Q9NY33	X	X		X
DJ-1 protein (Oncogene DJ1)	Q99497	X	X	X	
DNA damage binding protein 1	Q16531	X		X	
DNA repair protein RAD50 (EC 3.6.-.-) (hRAD50)	Q92878			X	
DNA replication licensing factor MCM2	P49736			X	
DNA replication licensing factor MCM7	P33993			X	
DNA-(apurinic or apyrimidinic site) lyase	P27695		X	X	
DNA-dependent protein kinase catalytic subunit	P78527				X
DNA-directed RNA polymerase II 140 kDa polypeptide	P30876				X
DnaJ homolog subfamily A member 3	Q96EY1				X
DnaJ homolog subfamily C member 13	O75165				X
Dynactin 1	Q14203				X
Dynamamin 2 (EC 3.6.5.5)	P50570		X		X
Dynamamin-like 120 kDa protein	O60313		X	X	
Dynein heavy chain, cytosolic	Q14204		X	X	X
Early endosome antigen 1	Q15075		X		
EH-domain containing protein 1 (Testilin) (hPAST1)	Q9H4M9		X	X	
EH-domain containing protein 3	Q9NZN3				X
Electron transfer flavoprotein alpha-subunit, mitochondrial precursor	P13804				X
Elongation factor 1-alpha	P68104	X	X	X	X
Elongation factor 1-alpha 2	Q05639			X	
Elongation factor 1-alpha, somatic form (EF-1-alpha-S)	P13549			X	
Elongation factor 1-beta (EF-1-beta)	P24534		X	X	
Elongation factor 1-delta	P29692	X	X	X	X
Elongation factor 1-gamma	P26641	X	X	X	X
Elongation factor 2 (EF-2)	P13639	X	X	X	X
Elongation factor Tu, mitochondrial precursor	P49411				X
Endoplasmin precursor	P14625	X			X
Enoyl-CoA hydratase, mitochondrial precursor	P30084				X

Erythrocyte band 7 integral membrane proteine	P27105	X			
ES1 protein homolog, mitochondrial precursor	P30042		X		
Esterase D (EC 3.1.1.1)	P10768	X	X	X	
Estradiol 17 beta-dehydrogenase 2	P37059				X
Eukaryotic initiation factor 4A-I (eIF4A-I) (eIF-4A-I)	P60842		X	X	X
Eukaryotic translation initiation factor 3 subunit 10	Q14152		X	X	X
Eukaryotic translation initiation factor 3 subunit 2	Q13347	X	X		X
Eukaryotic translation initiation factor 3 subunit 3	O15372			X	X
Eukaryotic translation initiation factor 3 subunit 5	O00303		X		
Eukaryotic translation initiation factor 3 subunit 6	Q9Y262		X	X	
Eukaryotic translation initiation factor 3 subunit 8	Q99613		X	X	X
Eukaryotic translation initiation factor 3 subunit 9	P55884		X		X
Eukaryotic translation initiation factor 5A	P63241			X	
Eukaryotic translation initiation factor 5B	O60841			X	
Exocyst complex component Sec8	Q96A65				X
Exportin-1	O14980	X	X	X	
Exportin-7 (Ran-binding protein 16)	Q9UIA9		X	X	
Ezrin (p81) (Cytovillin) (Villin 2)	P15311	X	X	X	X
F-actin capping protein alpha-1 subunit	P52907	X	X	X	X
F-actin capping protein alpha-2 subunit	P47755		X		
F-actin capping protein beta subunit	P79136	X	X	X	X
Far upstream element binding protein 1	Q96AE4		X		
Farnesyl pyrophosphate synthetase	P14324			X	
Fatty acid-binding protein, epidermal	Q01469			X	
Ferritin heavy chain	P02794				X
Ferritin light chain	P02792	X		X	
Filamin A	P21333		X	X	X
Filamin B	O75369		X		X
Filamin C	Q14315				X
Filamin-A	Q8BTM8	X			
Filamin-B	O75369	X			
FK506-binding protein 1A	P62942	X		X	
FK506-binding protein 4	Q02790			X	X
Flightless-I protein homolog	Q13045		X	X	X
Flotillin-1	O75955	X	X	X	X
Flotillin-2 (Epidermal surface antigen) (ESA)	Q14254		X	X	X
Formin-like protein (Protein C17orf1)	O95466				X
Fructose-bisphosphate aldolase A	P04075	X	X	X	X
Fumarate hydratase	P07954			X	
Galectin-1	P09382		X	X	
Galectin-7	P47929		X		
Gamma-enolase	P09104		X		
Gelsolin precursor	P06396	X			
General vesicular transport factor p115	O60763		X	X	X
Glia maturation factor gamma (GMF-gamma)	O60234	X			
Glucose-6-phosphate 1-dehydrogenase	P11413	X	X		X
Glucose-6-phosphate isomerase	P06744	X	X	X	X
Glutamate dehydrogenase 1, mitochondrial precursor	P00367				X
Glutamate dehydrogenase 2, mitochondrial precursor	P49448				X
Glutamate--cysteine ligase catalytic subunit	P48506				X
Glutaminyl-tRNA synthetase	P47897				X
Glutathione peroxidase 1	P07203	X			
Glutathione reductase	P00390				X
Glutathione S-transferase P	P09211	X	X	X	X
Glutathione synthetase	P48637	X			
Glutathione transferase omega 1	P78417		X	X	
Glyceraldehyde 3-phosphate dehydrogenase	P04406	X	X	X	X
Glyceraldehyde 3-phosphate dehydrogenase	P00354	X		X	X
Glycogen debranching enzyme	P35573		X	X	X

Glycogen phosphorylase, brain form	P11216		X	X	X
Glycyl-tRNA synthetase	P41250		X		X
Glyoxylate reductase/hydroxypyruvate reductase	Q9UBQ7		X	X	
Growth factor receptor-bound protein 2	P62993	X	X		X
GTPase Nras	P01111			X	
GTP-binding nuclear protein Ran	P62826	X	X	X	
GTP-binding nuclear protein RAN	P17080				X
GTP-binding protein ARD-1	P36406		X	X	X
GTP-binding protein SAR1a	Q9NR31		X		
Guanine nucleotide-binding protein beta subunit 2-like 1	P63244		X	X	
Guanine nucleotide-binding protein beta subunit 4	Q9HAV0	X		X	
Guanine nucleotide-binding protein beta subunit-like protein 12.3	P25388				X
Guanine nucleotide-binding protein G(i), alpha-1 subunit	P63096		X	X	
Guanine nucleotide-binding protein G(i), alpha-2 subunit	P04899	X	X	X	X
Guanine nucleotide-binding protein G(l)/G(s)/G(t)	P62873	X	X	X	X
Guanine nucleotide-binding protein G(l)/G(s)/G(t)	P62879		X	X	
Guanine nucleotide-binding protein G(k)	P08754	X		X	X
Guanine nucleotide-binding protein G(olf)	P38405				X
Guanine nucleotide-binding protein G(s)	P63092		X	X	X
Guanine nucleotide-binding protein, alpha-13 subunit (G alpha 13)	Q14344	X	X	X	X
Haptoglobin precursor	P00738				X
Haptoglobin-related protein precursor	P00739		X		X
Heat shock 70 kDa protein 1	P08107				X
Heat shock 70 kDa protein 1L	P34931		X		
Heat shock 70 kDa protein 4	P34932		X	X	X
Heat shock 70 kDa protein 6	P17066				X
Heat shock cognate 71 kDa protein	P11142	X	X	X	X
Heat shock protein HSP 90-alpha (HSP 86)	P07900	X	X	X	X
Heat shock protein HSP 90-beta (HSP 84) (HSP 90)	P08238	X	X	X	X
Hematopoietic lineage cell-specific protein	P14317		X		
Hemoglobin beta chain	P68871			X	
Hemopexin precursor (Beta-1B-glycoprotein)	P02790				X
Heterogeneous nuclear ribonucleoprotein A/B	Q99729		X	X	
Heterogeneous nuclear ribonucleoprotein A1	P09651	X	X	X	X
Heterogeneous nuclear ribonucleoprotein D0	Q14103	X	X	X	X
Heterogeneous nuclear ribonucleoprotein F	P52597		X	X	X
Heterogeneous nuclear ribonucleoprotein G	P38159			X	
Heterogeneous nuclear ribonucleoprotein H	P31943		X	X	X
Heterogeneous nuclear ribonucleoprotein K	Q07244	X	X	X	X
Heterogeneous nuclear ribonucleoprotein L	P14866			X	
Heterogeneous nuclear ribonucleoprotein M	P52272				X
Heterogeneous nuclear ribonucleoprotein R	O43390			X	X
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	X	X	X	X
Heterogeneous nuclear ribonucleoproteins C1/C2	P07910		X	X	X
Heterogeneous nuclear ribonucleoprotein U	Q00839		X	X	X
High mobility group protein 1	P09429			X	
High mobility group protein 1-like 10	Q9UGV6		X		
Hippocalcin-like protein 1	P37235	X			
Histidyl-tRNA synthetase	P12081			X	
Histidyl-tRNA synthetase homolog	P49590		X		
Histone H2A.a (H2A/a) (H2A.2)	P28001	X	X	X	X
Histone H2A.e (H2A/e)	Q99878	X			X
Histone H2A.o (H2A/o) (H2A.2) (H2a-615)	P20670				X
Histone H2B.a/g/h/k/l (H2B.1 A)	P62807	X	X	X	
Histone H2B.a/g/k (H2B.1 A)	P02278				X
Histone H2B.b (H2B/b) (H2B.1 B)	P58876	X			
Histone H2B.f (H2B/f) (H2B.1)	P33778	X			
Histone H3.3 (H3.A) (H3.B) (H3.3Q)	P06351				X
Histone H3/b	Q93081	X		X	X

Histone H4	P62805	X	X	X	X
HLA class I histocompatibility antigen, A-1	P30443	X	X	X	X
HLA class I histocompatibility antigen, B-15	P30464	X	X	X	X
HLA class I histocompatibility antigen, Cw-12	P30508			X	
HLA class I histocompatibility antigen, E	P13747				X
HLA class II histocompatibility antigen, DP	P20036	X	X	X	X
HLA class II histocompatibility antigen, DQ	P01919	X	X	X	X
HLA class II histocompatibility antigen, DR	P01903	X	X	X	X
HLA class II histocompatibility antigen, DRB1-1	P04229	X	X	X	X
HLA class II histocompatibility antigen, DW2.2	P01911	X	X	X	X
HLA class II histocompatibility antigen,	P01916			X	
Hornerin	Q86YZ3				X
Hsc70-interacting protein (Hip)	P50502		X	X	X
Huntingtin	P42858				X
Huntingtin interacting protein 1	O00291			X	
Hypoxanthine-guanine phosphoribosyltransferase	P00492	X			
Ig mu chain C region	P01871		X	X	X
Importin beta-1 subunit	Q14974	X	X	X	X
Importin beta-3	O00410	X	X	X	X
Importin-7	O95373		X		X
Importin-alpha re-exporter	P55060		X	X	X
Inorganic pyrophosphatase 2	Q9H2U2			X	
Inositol-1(or 4)-monophosphatase	P29218				X
Integrin alpha-3 precursor	P26006		X		
Integrin alpha-4 precursor	P13612	X			
Integrin alpha-L precursor	P20701	X	X	X	
Integrin alpha-M precursor	P11215	X			
Integrin beta-1 precursor	P05556	X	X		
Integrin beta-2 precursor	P05107	X	X		X
Intercellular adhesion molecule 3 precursor	P32942		X	X	X
Interferon-induced transmembrane protein 1	P13164	X			
Interferon-regulated resistance GTP-binding protein MxA	P20591				X
Interferon-stimulated gene 20 kDa protein	Q96AZ6			X	
Isocitrate dehydrogenase	P48735			X	X
Isoleucyl-tRNA synthetase,	P41252		X	X	X
Isovaleryl-CoA dehydrogenase	P26440				X
Keratin, type I cytoskeletal 10	P13645	X	X	X	X
Keratin, type I cytoskeletal 14	P02533		X		X
Keratin, type I cytoskeletal 9	P35527	X	X	X	X
Keratin, type II cytoskeletal 1	P04264	X	X	X	X
Keratin, type II cytoskeletal 2 epidermal	P35908	X	X	X	X
Keratin, type II cytoskeletal 4	P19013			X	
Keratin, type II cytoskeletal 5	P13647	X	X		X
Keratin, type II cytoskeletal 6A	P02538	X			X
Kinectin (Kinesin receptor)	Q86UP2			X	X
Kinesin heavy chain	P33176				X
Lactoylglutathione lyase	Q04760				X
Lamin B1	P20700	X	X	X	X
Lamin B2	Q03252		X		
LanC-like protein 1	O43813			X	
Leucyl-cystinyl aminopeptidase	Q9UIQ6		X	X	X
Leukocyte antigen CD37	P11049			X	X
Leukocyte common antigen precursor	P08575	X	X	X	X
Leukocyte elastase inhibitor	P30740	X	X	X	X
Leukocyte surface antigen CD47 precursor	Q08722		X	X	
Leukocyte surface antigen CD53	P19397				X
Leukosialin precursor	P16150				X
Leukotriene A-4 hydrolase	P09960	X	X	X	X
L-lactate dehydrogenase A chain	P00338	X	X	X	X

L-lactate dehydrogenase B chain	P07195	X	X	X	X
L-plastin	P13796	X	X	X	X
Lymphocyte antigen 75 precursor	O60449				X
Macrophage capping protein	P40121		X	X	X
Macrophage migration inhibitory factor	P14174	X		X	
Mago nashi protein homolog	P61326			X	
Major vault protein (MVP)	Q14764	X	X		X
Malate dehydrogenase, cytoplasmic	P40925	X	X	X	X
Malate dehydrogenase, mitochondrial precursor	P40926		X	X	X
Matrin-3	P43243		X		
Membrane transport protein XK	P51811		X		
Membrane-associated protein HEM-1	P55160		X	X	X
Methionyl-tRNA synthetase	P56192				X
Mitochondrial processing peptidase alpha subunit, mitochondrial precursor	Q10713			X	
Mitogen-activated protein kinase kinase 1 interacting protein	Q9UHA4			X	
Mitogen-activated protein kinase kinase kinase 2	Q12851				X
Moesin	P26038	X	X	X	X
Mu-crystallin homolog	Q14894			X	
Multifunctional protein ADE2	P22234		X		
Multisynthetase complex auxiliary component p38	Q13155		X		
Myeloperoxidase precursor	P05164	X			
Myosin heavy chain, nonmuscle type A	P35579				X
Myosin heavy chain, smooth muscle isoform	P35749				X
Myosin Ie (Myosin Ic)	Q12965				X
Myosin light polypeptide 6	P60660	X	X	X	
Myosin regulatory light chain 2, nonsarcomeric	P19105	X	X		
Myosin XVIIIa	Q92614				X
Myosin-9 (Myosin heavy chain, nonmuscle IIa)	P35579	X	X	X	
Myotrophin	P58546	X		X	
Myristoylated alanine-rich C-kinase substrate	P29966			X	
N-acetylglucosamine kinase (EC 2.7.1.59)	Q9UJ70		X		X
Nanos homolog 1 (NOS-1) (EC_Rep1a)	Q8WY41	X			
Neddylin (Ubiquitin-like protein NEDD8)	Q15843			X	
Neuroblast differentiation associated protein AHNAK	Q09666				X
Neuron specific calcium-binding protein hippocalcin	P84074			X	
Neutral amino acid transporter B(0) (ATB(0))	Q15758	X			
Neutrophil cytosol factor 1	P14598				X
Neutrophil cytosol factor 4 (NCF-4)	Q15080	X			
Niban-like protein (Meg-3)	Q96TA1	X			
NipSnap1 protein	Q9BPW8				X
NipSnap3A protein (NipSnap4)	Q9UFN0				X
Nuclear factor NF-kappa-B p105 subunit	P19838		X		
Nuclear mitotic apparatus protein 1	Q14980		X	X	
Nuclear protein Hcc-1	P82979		X		
<i>Nuclear receptor ROR-gamma</i>	<i>P51450</i>				X
Nucleolin (Protein C23)	P19338			X	
Nucleophosmin (NPM)	P06748			X	X
Nucleoside diphosphate kinase A	P15531	X	X	X	
Nucleoside diphosphate kinase B	P22392	X			
Nucleosome assembly protein 1-like 1	P55209		X	X	
Nucleosome assembly protein 1-like 4	Q99733			X	
Osteoclast stimulating factor 1	Q92882				X
Peptidyl-prolyl cis-trans isomerase A	P62937	X	X	X	X
Peptidyl-prolyl cis-trans isomerase B precursor	P23284	X			
Perforin 1 precursor (P1)	P14222				X
Peroxiredoxin 1	Q06830	X	X	X	X
Peroxiredoxin 2	P32119	X		X	X
Peroxiredoxin 5, mitochondrial precursor	P30044	X	X	X	
Peroxiredoxin 6	P30041	X	X		X

Peroxisomal multifunctional enzyme type 2	P51659				X
Phenylalanyl-tRNA synthetase alpha chain	Q9Y285		X		X
Phosphate carrier protein, mitochondrial precursor	Q00325		X		
Phosphatidylethanolamine-binding protein	P30086	X	X	X	X
Phosphatidylinositol 3-kinase regulatory alpha subunit	P27986		X		X
Phosphatidylinositol 4-kinase alpha	P42356		X		X
Phosphoenolpyruvate carboxykinase, mitochondrial precursor	Q16822				X
Phosphoglucomutase 1	P36871		X	X	
Phosphoglycerate kinase 1	P00558	X	X	X	X
Phosphoglycerate mutase 1	P18669	X	X	X	X
Phosphoribosyl pyrophosphate synthetase-associated protein 2	O60256			X	
Phosphoribosylformylglycinamide synthase	O15067		X		
Placental ribonuclease inhibitor	P13489			X	X
Placental thrombin inhibitor	P35237	X	X		
Plasma membrane calcium-transporting ATPase 1	P20020		X		
Platelet endothelial cell adhesion molecule precursor	P16284			X	X
Platelet-activating factor acetylhydrolase IB beta subunit	P68402	X	X	X	
Plectin 1 (PLTN) (PCN)	Q15149		X	X	X
Plexin B2 precursor (MM1)	O15031	X			
Poly	P09874		X	X	X
Poly(rC)-binding protein 1	Q15365		X		
Polypyrimidine tract-binding protein 1	P26599	X	X	X	X
Potassium voltage-gated channel subfamily KQT member 1	P51787	X			
Pre-mRNA processing splicing factor 8	Q6P2Q9			X	
Probable ATP-dependent helicase DDX17	Q92841		X	X	
Probable cation-transporting ATPase 13A1	Q9HD20		X		
Probable RNA-dependent helicase p72	Q92841			X	
Probable ubiquitin carboxyl-terminal hydrolase FAF-X	Q93008				X
Profilin I	P07737	X	X	X	X
Programmed cell death 6-interacting protein	Q8WUM4		X	X	X
Programmed cell death protein 6	O75340	X		X	
Prohibitin	P35232		X		
Proliferation-associated protein 2G4	Q9UQ80	X	X	X	
Proteasome activator complex subunit 1	Q06323	X	X		
Proteasome activator complex subunit 2	Q9UL46	X	X	X	X
Proteasome subunit alpha type 1	P25786		X	X	
Proteasome subunit alpha type 2	P25787	X	X		X
Proteasome subunit alpha type 3	P25788	X			
Proteasome subunit alpha type 4	P25789	X	X	X	X
Proteasome subunit alpha type 5	P28066	X	X	X	
Proteasome subunit alpha type 6	P60900	X	X	X	X
Proteasome subunit alpha type 7	O14818	X	X	X	X
Proteasome subunit beta type 1	P20618	X	X	X	X
Proteasome subunit beta type 10 precursor	P40306	X			
Proteasome subunit beta type 2	P49721	X	X	X	X
Proteasome subunit beta type 4 precursor	P28070				X
Proteasome subunit beta type 6 precursor	P28072				X
Proteasome subunit beta type 8 precursor	P28062	X	X	X	
Proteasome subunit beta type 9 precursor	P28065	X		X	X
Protein 4.1 (Band 4.1) (P4.1) (EPB4.1) (4.1R)	P11171				X
Protein arginine N-methyltransferase 1	Q99873			X	X
Protein C19orf10 precursor	Q969H8	X			
Protein disulfide-isomerase A3 precursor	P30101	X			
Protein disulfide-isomerase A6 precursor	Q15084	X		X	
Protein FAM40A	Q5VSL9	X			
Protein FAM49B (L1)	Q9NUQ9	X		X	
Protein KIAA0196	Q12768		X		
Protein kinase C, beta type (EC 2.7.1.37) (PKC-beta) (PKC-B)	P05126		X	X	X
Protein TMED8	Q6PL24			X	

Protein transport protein Sec23A (SEC23-related protein A)	Q15436			X	
Protein transport protein Sec23B (SEC23-related protein B)	Q15437				X
Protein transport protein Sec24C (SEC24-related protein C)	P53992		X		X
Protein tyrosine phosphatase receptor type C-associated protein	Q14761			X	
Protein tyrosine phosphatase type IVA protein 1	Q93096			X	
Protein-tyrosine phosphatase eta precursor	Q12913				X
Protein-tyrosine phosphatase, non-receptor type 6	P29350				X
Protocadherin alpha 11 precursor	Q9Y511				X
Proto-oncogene tyrosine-protein kinase FGR	P09769		X		X
Proto-oncogene tyrosine-protein kinase LCK	P06239		X	X	X
PRP19/PSO4 homolog	Q9UMS4		X	X	
Purine nucleoside phosphorylase	P00491	X	X		X
Puromycin-sensitive aminopeptidase	P55786	X	X	X	X
Putative deoxyribose-phosphate aldolase	Q9Y315		X		
Putative GTP-binding protein PTD004	Q9NTK5	X			X
Putative pre-mRNA splicing factor RNA helicase	O43143		X	X	X
Pyruvate kinase, isozymes M1/M2	P14618	X	X	X	X
Quinone oxidoreductase	Q08257		X	X	
Rab GDP dissociation inhibitor alpha	P31150	X	X	X	
Rab GDP dissociation inhibitor beta (Rab GDI beta) (GDI-2)	P50395	X	X	X	X
Ras GTPase-activating-like protein IQGAP1 (p195)	P46940	X	X	X	X
Ras GTPase-activating-like protein IQGAP2	Q13576	X		X	
Ras-related C3 botulinum toxin substrate 1	P63000	X	X		X
Ras-related C3 botulinum toxin substrate 2	P15153	X	X	X	X
Ras-related C3 botulinum toxin substrate homolog DJ20J23.1	O95916				X
Ras-related protein Rab-10	P61026			X	X
Ras-related protein Rab-11A (Rab-11) (YL8)	P62491	X	X		X
Ras-related protein Rab-11B (GTP-binding protein YPT3)	Q15907			X	X
Ras-related protein Rab-14	P61106	X		X	X
Ras-related protein Rab-18	Q9NP72				X
Ras-related protein Rab-1A (YPT1-related protein)	P62820	X	X	X	X
Ras-related protein Rab-1B	Q9H0U4	X	X	X	X
Ras-related protein Rab-21	Q9UL25			X	X
Ras-related protein Rab-26	Q9ULW5			X	
Ras-related protein Rab-2A	P61019	X	X	X	X
Ras-related protein Rab-35 (Rab-1C)	Q15286			X	X
Ras-related protein Rab-4A	P20338				X
Ras-related protein Rab-5B	P61020	X	X		X
Ras-related protein Rab-5C (RAB5L) (L1880)	P51148	X	X	X	X
Ras-related protein Rab-7	P51149	X	X	X	X
Ras-related protein Rab-8A (Oncogene c-mel)	P61006	X	X	X	X
Ras-related protein Rab-8B	Q92930			X	X
Ras-related protein Ral-A	P11233		X	X	X
Ras-related protein Ral-B	P11234		X		
Ras-related protein Rap-1A	P62834	X	X	X	X
Ras-related protein Rap-1b (GTP-binding protein smg p21B)	P61224	X		X	X
Ras-related protein Rap-2b	P61225			X	X
Ras-related protein R-Ras (p23)	P10301	X			
Ras-related protein R-Ras2	P17082				X
RCC2 protein	Q9P258			X	
Receptor-type tyrosine-protein phosphatase eta precursor	Q12913			X	
Retinal dehydrogenase 1	P00352	X			
Rho GDP-dissociation inhibitor 1	P52565	X	X	X	X
Rho GDP-dissociation inhibitor 2	P52566	X	X	X	X
Rho-associated protein kinase 1	Q13464			X	
Rho-GTPase-activating protein 1	Q07960	X	X		
Rho-GTPase-activating protein 4	P98171				X
Rho-related GTP-binding protein RhoC	P08134	X			
Ribonuclease inhibitor	P13489		X		

Ribose-phosphate pyrophosphokinase I	P60891		X	X	X
RNA-binding protein 8A	Q9Y5S9			X	
RuvB-like 1 (EC 3.6.1.-)	Q9Y265	X		X	X
Secernin-1	Q12765		X	X	
Semaphorin-3B precursor	Q13214	X			
Septin 1 (LARP)	Q8WYJ6				X
Septin 2 (NEDD5 protein homolog)	Q15019	X	X	X	X
Septin 6	Q14141		X	X	X
Septin 7 (CDC10 protein homolog)	Q16181		X	X	X
Septin-9	Q9UHD8		X	X	
Serine hydroxymethyltransferase	P34897			X	X
Serine/threonine protein phosphatase 2A, 55 kDa	P63151		X	X	
Serine/threonine protein phosphatase 2A, 56 kDa	Q14738				X
Serine/threonine protein phosphatase 2A, 65 kDa regulatory subunit A	P30153	X	X	X	
Serine/threonine protein phosphatase 2A, catalytic subunit, alpha isoform	P67775		X		
Serine/threonine protein phosphatase PP1-alpha catalytic subunit	P62136	X		X	
Serine/threonine-protein kinase 24 (EC 2.7.1.37)	Q9Y6E0		X	X	
Serine/threonine-protein kinase MST4	Q9P289			X	
Serine/threonine-protein kinase/endoribonuclease IRE2 precursor	Q76MJ5	X			
Serine-threonine kinase receptor-associated protein	Q9Y3F4		X		
Serotransferrin precursor	P02787		X		X
Serpin A12 precursor	Q8IW75	X	X		
Serum albumin precursor	P02768	X	X	X	X
Seryl-tRNA synthetase (EC 6.1.1.11)	P49591		X	X	X
SET protein (Phosphatase 2A inhibitor I2PP2A) (I-2PP2A)	Q01105	X	X	X	
SH3 domain-binding glutamic acid-rich-like protein	O75368	X	X	X	
Short chain 3-hydroxyacyl-CoA dehydrogenase	Q16836			X	
Signal recognition particle 14 kDa protein	P37108	X			
Single-stranded DNA-binding protein	Q04837			X	
Small nuclear ribonucleoprotein E	P62304	X		X	
Small nuclear ribonucleoprotein G	P62308	X			
Small nuclear ribonucleoprotein Sm D2	P62316	X			
Small nuclear ribonucleoprotein Sm D3	P62318			X	
Small ubiquitin-related modifier 2 precursor	P61956	X	X	X	
Sodium/potassium-transporting ATPase alpha-1 chain precursor	P05023	X	X	X	X
Sodium/potassium-transporting ATPase beta-3 chain	P54709		X		
Sorcin (22 kDa protein) (CP-22) (V19)	P30626	X		X	
Sorting nexin-3 (SDP3 protein)	O60493			X	
Spectrin alpha chain, brain	Q13813	X	X	X	X
Spectrin beta chain, brain 1	Q01082	X	X	X	X
Spectrin beta chain, brain 2	O15020		X	X	
Spermatogenesis-associated protein 7	Q9P0W8	X			
Spermidine synthase (EC 2.5.1.16)	P19623				X
Spliceosome RNA helicase BAT1	Q13838		X	X	
Splicing factor 3A subunit 3	Q12874			X	
Splicing factor 3B subunit 1	O75533			X	
Splicing factor U2AF 35 kDa subunit	Q01081				X
Splicing factor, arginine/serine-rich 3	P84103			X	
Splicing factor, proline-and glutamine-rich	P23246			X	X
Staphylococcal nuclease domain containing protein	Q7KZF4		X	X	
Stathmin (Phosphoprotein p19) (pp19)	P16949	X		X	
Stress-70 protein, mitochondrial precursor	P38646		X	X	X
Stress-induced-phosphoprotein 1	P31948	X	X	X	
Succinyl-CoA ligase	Q96199				X
Superoxide dismutase	P04179		X	X	
Suppressor of tumorigenicity 14	Q9Y5Y6		X		
Synaptic vesicle membrane protein VAT-1 homolog	Q99536		X	X	
Synaptophysin-like protein (Pantophysin)	Q16563	X			
Syntaxin 4	Q12846				X

Syntaxin-binding protein 3 (UNC-18 homolog 3)	O00186		X	X	
Talin 1	Q9Y490	X	X	X	X
T-cell leukemia/lymphoma protein 1A	P56279	X		X	
T-complex protein 1, alpha subunit	P17987	X	X	X	X
T-complex protein 1, beta subunit	P78371		X	X	X
T-complex protein 1, delta subunit	P50991	X	X	X	X
T-complex protein 1, epsilon subunit	P48643	X	X	X	X
T-complex protein 1, eta subunit	Q99832	X	X	X	X
T-complex protein 1, gamma subunit	P49368	X	X	X	X
T-complex protein 1, theta subunit	P50990	X	X	X	X
T-complex protein 1, zeta subunit	P40227	X	X	X	X
Telomerase-binding protein p23	Q15185	X		X	
Tetraspanin-14 (Tspan-14)	Q8NG11	X			
Thimet oligopeptidase	P52888		X	X	
Thioredoxin (ATL-derived factor) (ADF)	P10599			X	
Thioredoxin-like protein 2	O76003				X
Threonyl-tRNA synthetase, cytoplasmic	P26639	X	X	X	X
Transaldolase (EC 2.2.1.2)	P37837	X	X	X	
Transcription intermediary factor 1-beta	Q13263			X	
Transferrin receptor protein 1	P02786			X	
Transforming protein RhoA (H12)	P61586	X	X	X	X
Transforming protein RhoC (H9)	P08134				X
Transgelin-2 (SM22-alpha homolog)	P37802	X		X	
Transitional endoplasmic reticulum ATPase	P55072	X	X	X	X
Transketolase (EC 2.2.1.1) (TK)	P29401	X	X	X	
Translationally-controlled tumor protein	P13693	X			
Translin	Q15631	X	X		X
Transmembrane protein 24	O14523	X			
Trifunctional enzyme alpha subunit, mitochondrial precursor	P40939				X
Trifunctional purine biosynthetic protein adenosine-3	P22102				X
Triosephosphate isomerase	P60174	X	X	X	X
Tripeptidyl-peptidase 2	P29144	X	X		
tRNA-splicing endonuclease subunit Sen54	Q7Z6J9	X			
Tropomyosin 1 alpha chain	P09493	X			
Tropomyosin alpha 4 chain	P67936		X		
Tropomyosin beta chain	P07951	X			
Tryptophanyl-tRNA synthetase	P23381		X	X	X
Tubulin alpha-1 chain (Alpha-tubulin 1)	P68366	X	X	X	X
Tubulin alpha-3 chain (Alpha-tubulin 3) (Tubulin B-alpha-1)	Q71U36			X	
Tubulin alpha-4 chain (Alpha-tubulin 4)	P05215				X
Tubulin alpha-6 chain (Alpha-tubulin 6)	Q9BQE3			X	X
Tubulin alpha-ubiquitous chain	P68363	X	X	X	
Tubulin beta-? chain (Tubulin beta-2 chain)	P68371	X	X	X	
Tubulin beta-1 chain	P07437				X
Tubulin beta-2 chain	P07437	X	X	X	X
Tubulin beta-5 chain	P05218				X
Tyrosine-protein kinase BTK	Q06187				X
Tyrosine-protein kinase CSK	P41240				X
Tyrosine-protein kinase JAK1	P23458				X
Tyrosine-protein kinase Lyn	P07948		X	X	X
Tyrosine-protein kinase SYK	P43405				X
Tyrosine-protein phosphatase, non-receptor type 6	P29350	X	X	X	
U5 small nuclear ribonucleoprotein 200 kDa helicase	O75643			X	
UBA/UBX 33.3 kDa protein	Q04323				X
Ubiquitin	P62988	X		X	
Ubiquitin carboxyl-terminal hydrolase 14	P54578			X	X
Ubiquitin carboxyl-terminal hydrolase 5	P45974				X
Ubiquitin carboxyl-terminal hydrolase 7	Q93009			X	
Ubiquitin carboxyl-terminal hydrolase isozyme L1i	P09936			X	

Ubiquitin thiolesterase protein OTUB1	Q96FW1	X		X	
Ubiquitin-activating enzyme E1	P22314	X	X	X	X
Ubiquitin-conjugating enzyme E2 I	P63279	X	X		
Ubiquitin-conjugating enzyme E2 N	P61088	X	X	X	
Ubiquitin-like protein 7	Q96S82	X			
Unc-112 related protein 2	Q86UX7		X		X
UTP--glucose-1-phosphate uridylyltransferase 1	Q07131	X	X		X
Vacuolar ATP synthase catalytic subunit A	P38606			X	X
Vacuolar protein sorting 26	O75436		X		X
Vacuolar protein sorting 35	Q96QK1	X	X	X	X
Valyl-tRNA synthetase	P26640	X	X	X	X
Vasodilator-stimulated phosphoprotein (VASP)	P50552				X
Vesicle-fusing ATPase (EC 3.6.4.6)	P46459		X		X
Vimentin	P08670	X	X	X	X
Vinculin (Metavinculin)	P18206	X	X		X
Voltage-dependent anion-selective channel protein 1	P21796				X
WD-repeat protein 1	O75083	X	X	X	X
Xaa-Pro aminopeptidase 1 (EC 3.4.11.9)	Q9NQW7	X			
XRP2 protein	O75695			X	
Zinc finger protein 197 (ZnF20)	O14709				X

Tableau 1 : Liste des protéines identifiées à partir de préparation de microparticules en fonction des différentes pathologies.

LAL : Leucémie aigue lymphoblastique ; LLC : Leucémie lymphoïde chronique ; MCL : lymphome du manteau ; LNH : Lymphome à petites cellules.

2) Détermination par spectrométrie de masse non-covalente de la fixation d'un atome de ZN

Cette étude visait à déterminer s'il existait des interactions entre la protéine (Ataxin-7), et des atomes de Zn, et si oui, combien d'atomes de Zn étaient impliqués. En étudiant la protéine en conditions natives par spectrométrie de masse, par rapport à sa forme dénaturée, nous avons pu mettre en évidence la fixation d'un unique atome de Zn sur la protéine. Ces travaux ont été publiés dans :
Human Molecular Genetics, 2004, Vol. 13, No. 12 1257–1265

Ataxin-7 is a subunit of GCN5 histone cetyltransferase-containing complexes

Dominique Helmlinger¹, Sara Hardy², Souphatta Sasorith³, Fabrice Klein¹, Flavie Robert², Chantal Weber¹, Laurent Miguet⁴, Noelle Potier⁴, Alain Van-Dorsselaer⁴, Jean-Marie Wurtz³, Jean-Louis Mandel¹, La`szlo` Tora² and Didier Devys¹

Spinocerebellar ataxia type 7 (SCA7) is a neurodegenerative disorder caused by a CAG repeat expansion in the SCA7 gene leading to elongation of a polyglutamine tract in ataxin-7, a protein of unknown function. A putative ataxin-7 yeast orthologue (SGF73) has been identified recently as a new component of the SAGA (Spt/Ada/Gcn5 acetylase) multisubunit complex, a coactivator required for transcription of a subset of RNA polymerase II-dependent genes. We show here that ataxin-7 is an integral component of the mammalian SAGA-like complexes, the TATA-binding protein-free TAF-containing complex (TFTC) and the SPT3/TAF9/GCN5 acetyltransferase complex (STAGA). In agreement, immunoprecipitation of ataxin-7 retained a histone acetyltransferase activity, characteristic for TFTC-like complexes. We further identified a minimal domain in ataxin-7 that is required for interaction with TFTC/STAGA subunits and is conserved highly through evolution, allowing the identification of a SCA7 gene family. We showed that this domain contains a conserved Cys(3)His motif that binds zinc, forming a new zinc-binding domain. Finally, polyglutamine expansion in ataxin-7 did not affect its incorporation into TFTC/STAGA complexes purified from SCA7 patient cells. We demonstrate here that ataxin-7 is the human orthologue of the yeast SAGA SGF73 subunit and is a bona fide subunit of the human TFTC-like transcriptional complexes.

3) Détermination par spectrométrie de masse non-covalente du nombre d'atomes de Zn fixés sur la partie c-terminale de la protéine E6

Lors de cette étude, nous avons montré, par spectrométrie de masse en comparant la masse de la protéine en conditions natives par rapport aux conditions dénaturantes qu'il n'y avait qu'un seul atome de Zn qui se fixait sur la protéine E6. Ce travail a été publié dans :

Journal of Biomolecular NMR (2005) 31: 129–141

¹H and ¹⁵N resonance assignment, secondary structure and dynamic behaviour of the C-terminal domain of human papillomavirus oncoprotein E6

Yves Nominé¹, Sebastian Charbonnier¹, Laurent Miguet², Noelle Potier², Alain Van Dorsselaer², R. Andrew Atkinson³, Gilles Travé¹ and Bruno Kieffer³

E6 is a viral oncoprotein implicated in cervical cancers, produced by human papillomaviruses (HPVs). E6 contains two putative zinc-binding domains of about 75 residues each. The difficulty in producing recombinant E6 has long hindered the obtention of structural data. Recently, we described the expression and purification of E6-C 4C/4S, a stable, folded mutant of the C-terminal domain of HPV16 E6. Here, we have produced ¹⁵N-labelled samples of E6-C 4C/4S for structural studies by NMR. We have assigned most ¹H and ¹⁵N resonances and identified the elements of secondary structure of the domain. The domain displays an original alpha/beta topology with roughly equal proportions of alpha-helix and beta-sheet. The PDZ-binding region of E6, located at the extreme C-terminus of the domain, is in a random conformation. Mass spectrometry demonstrated the presence of one zinc ion per protein molecule. Kinetics of replacement of zinc by cadmium followed by ¹H,¹⁵N-HSQC experiments revealed specific frequency changes for the zinc-binding cysteines and their immediate neighbours. NMR spectra were affected by severe line-broadening effects which seriously hindered the assignment work. Investigation of these effects by ¹⁵N relaxation experiments showed that they are due to heterogeneous dynamic behaviour with mus-ms time scale motions occurring in localised regions of the monomeric domain.

4) Détermination de l'incorporation d'isotopes dans les protéines lors de la croissance de bactéries.

Lors de cette étude, nous devons déterminer la quantité d'isotopes stables (C_{13} , N_{15} et H_2) incorporés dans la protéine E6. L'analyse par spectrométrie de masse de la protéine en fonction de son milieu de culture a permis de manière efficace de déterminer le taux d'incorporation des différents isotopes. Ces travaux ont fait l'objet d'une publication :

Biotechniques. 2005 Sep;39(3):405-11

Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120.

Desplancq D, Bernard C, Sibling AP, Kieffer B, Miguet L, Potier N, Van Dorselaer A, Weiss E.

The difficulty and expense of preparing protein samples highly enriched in stable isotopes is a bottleneck for structural studies by nuclear magnetic resonance (NMR) spectroscopy. We have developed a new regulatable expression/labeling vector system in the cyanobacterium *Anabaena* sp. PCC 7120 using the endogenous promoter of the nitrate assimilation *nir* operon. Standard proteins were overexpressed upon induction with $NaNO_3$, yielding up to 250 mg/L of culture. When the cyanobacteria were grown in the presence of inexpensive ^{15}N -, ^{13}C -labeled mineral salts and $2H_2O$, the expressed polypeptides were highly (>90%) enriched in stable isotopes. Furthermore, the tight repression of the *nir* promoter upon induction allowed the production of the toxic oncoprotein E6. In addition, under these conditions, the malE31 protein, while insoluble in *Escherichia coli*, was found to be soluble in *Anabaena*. Together, these properties render the described system especially suitable for the production and/or triple labeling of recombinant protein samples. It represents an interesting alternative to conventional protein expression systems used in structural genomics.

5) Détermination des protéines interagissant avec la protéine NIR par purification Tap-TAG du complexe et analyse protéomique

Lors de cette étude nous devons déterminer les différentes protéines qui interagissent avec la protéine NIR. Les analyses par spectrométrie de masse MALDI-MS ont permis de mettre en évidence la protéine p53 impliquée dans la suppression des tumeurs comme un partenaire spécifique de la protéine NIR. Ces travaux ont fait l'objet d'une publication :

Genes & Dev. 2005 19: 2912-2924

NIR is a novel INHAT repressor that modulates the transcriptional activity of p53

Philip Hublitz, Natalia Kunowska, Ulrich P. Mayer, Judith M. Müller, Kristina Heyne, Na Yin, Claudia Fritzsche, Cecilia Poli, Laurent Miguet, Ingo W. Schupp, Leo A. van Grunsven, Noëlle Potiers, Alain van Dorsseleer, Eric Metzger, Klaus Roemer and Roland Schüle

Most transcriptional repression pathways depend on the targeted deacetylation of histone tails. In this report, we characterize NIR, a novel transcriptional corepressor with inhibitor of histone acetyltransferase (INHAT) activity. NIR (Novel INHAT Repressor) is ubiquitously expressed throughout embryonic development and adulthood. NIR is a potent transcriptional corepressor that is not blocked by histone deacetylase inhibitors and is capable of silencing both basal and activator-driven transcription. NIR directly binds to nucleosomes and core histones and prevents acetylation by histone acetyltransferases, thus acting as a bona fide INHAT. Using a tandem affinity purification approach, we identified the tumor suppressor p53 as a NIR-interacting partner. Association of p53 and NIR was verified *in vitro* and *in vivo*. Upon recruitment by p53, NIR represses transcription of both p53-dependent reporters and endogenous target genes. Knock-down of NIR by RNA interference significantly enhances histone acetylation at p53-regulated promoters. Moreover, p53-dependent apoptosis is robustly increased upon depletion of NIR. In summary, our findings describe NIR as a novel INHAT that plays an important role in the control of p53 function.