THESE

présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE LOUIS PASTEUR

DE STRASBOURG

Par

Sophie RICHERT

De l'identification à la caractérisation de protéines :

Développement de la spectrométrie de masse dans le cadre de l'approche protéomique

Soutenue le 18 décembre 2003 devant la commission d'examen :

Dr. Alain VAN DORSSELAER Prof. Jacques HAIECH Prof. Hubert HONDERMARCK Prof. Jean-Marie SCHMITTER Dr. Thierry RABILLOUD Dr. Emmanuelle LEIZE-WAGNER Directeur de thèse Rapporteur interne Rapporteur externe Rapporteur externe Examinateur Examinateur

Remerciements

Pour commencer, je tiens à remercier Alain Van Dorsselaer pour m'avoir accueillie dans son laboratoire et surtout m'avoir décidée à commencer une thèse en spectrométrie de masse, même si je n'étais « qu'une biologiste »! Je vous suis particulièrement reconnaissante de m'avoir communiqué votre enthousiasme pour la spectrométrie de masse, plus particulièrement pour la protéomique et pour m'avoir permis de développer des collaborations très enrichissantes. Merci aussi pour les grandes discussions le soir dans le bureau et les nombreux conseils prodigués tout au long de cette thèse au laboratoire mais aussi en congrès, ils m'ont été fort utiles et le seront encore.

J'exprime ma gratitude au Graduiertenkolleg « groupe 532 » pour leur soutien financier au cours de ces trois années de thèse, mais aussi pour m'avoir donné la possibilité de présenter mes travaux lors de leurs nombreuses réunions.

J'adresse mes vifs remerciements à Monsieur le Professeur Jacques Haiech qui a accepté de présider ce jury de thèse. Je remercie également Messieurs les Professeurs Jean-Marie Schmitter et Hubert Hondermarck, ainsi que Monsieur Thierry Rabilloud qui ont bien voulu prendre le temps de juger ce travail de thèse.

Un grand merci à Emmanuelle Leize-Wagner qui m'a encadrée tout au long de ces trois années ! J'ai particulièrement apprécié tes conseils et nos discussions scientifiques mais surtout le fait que tu aies été à l'écoute et disponible. Merci de m'avoir fait confiance pour mener à bien nos différents projets et pour ton coté humain !

Ce travail étant le fruit de nombreuses collaborations, je tiens à remercier l'ensemble des personnes qui y ont participé :

Tout d'abord, je tiens à remercier une nouvelle fois Thierry Rabilloud (cette fois-ci en tant que collaborateur et plus examinateur) du CEA de Grenoble pour m'avoir donné mon tout premier projet (l'étude sur la compatibilité avec la spectrométrie de masse des colorants à l'argent) et surtout de m'avoir laissé le temps de le mener à bien et d'y faire mes premières armes.

J'exprime également ma reconnaissance à Catherine Florentz, Benedicte Sohm et Petra Tryoen de l'Institut de Biologie Moléculaire et Cellulaire de Strasbourg pour notre très enrichissante collaboration sur les mitochondries.

Un grand merci à Jean-Paul Klein, Marie Schöller et Jordane Biarc de l'INSERM U392 d'Illkirch pour notre étude sur *Streptococcus bovis*. Maintenant la route est ouverte, (place aux suivants, n'est ce pas Jordane ?).

Merci aussi à Jean-François Launay et Natacha Turck de l'INSERM U381 d'Hautepierre pour l'étude protéomique des noyaux de cellules Caco2.

Je remercie également Catherine Braun-Breton et Laetitia Vincensini, pour nos études *De Novo* sur *Plasmodium falciparum*, délicates mais passionantes.

Enfin, merci à Christiane Moog et Renaud Burrer de l'institut de virologie de Strasbourg, pour m'avoir permis de continuer mon projet de DEA. (Désolée Renaud, je ne crois pas que je vais vous rejoindre à San Diego...).

Pour finir, je tiens à dire un grand MERCI à l'ensemble des personnes du LSMBO ! Dans le désordre, merci à : Christine 5. pour ses conseils à tous niveaux et pour m'avoir écoutée d'une oreille attentive. Raymond, cycliste comme moi qu'il neige ou pleuve et solidaire jusque dans les crevaisons ! As tu pensé à ta petite trousse de secours anti-crevaison ? Merci aussi pour les conseils concernant les machines, les pompes...dommage qu'on n'y fasse pas toujours suffisamment attention, peut-être faut-il finir sa thèse pour s'en rendre compte.

Fabrice, ingénieur informaticien, toujours disponibles pour les petits soucis informatiques mais qui n'accepte pas les cadeaux (ou les oublie), même lorsqu'il s'agit d'une bouteille de Crémant ! Jean-Marc pour m'avoir appris la protéomique (digestions manuelles...) et ses jolies photos. Noëlle, partenaire de step

Nathalie Z., toute jeune maman, félicitations!

Danièle, collègue de bureau et de gym, experte en gels 2D

Aux ex-post-docs, Christelle, partenaire au squash et traductrice officielle du labo et Virginie, fournisseur des bonnes bouteilles au labo.

Aux filles, Karine (le fauteuil de ton bureau est vraiment confortable !), merci pour nos longues discussions scientifiques et autres ! Bonne chance pour la suite !

Stéphanie, l'espagnole, bonne chance, peut-être que nos routes se recroiseront, elles étaient tellement semblables l'an passé.

Hélène D., ancienne partenaire de basket, pour sa bonne humeur et sa disponibilité (pour les problèmes de robot, MALDI et autres...)!

A Elsa, la prochaine à soutenir, bonne chance pour la suite !

A Rossela, future collègue italienne ? et Stéphanie L.

Au bureau des garçons (désolée Christine C. !), Guido, maintenant tu es le grand, il va falloir t'occuper des autres !, Laurent merci pour ton invitation, mais j'ai de meilleures propositions, Andy, 2° traducteur officiel et Christine C., la relève en protéomique.

A François, qui n'a toujours pas compris que je n'ai jamais mis d'artichauts dans mes tartes mais des courgettes (mais je ne t'en veux pas, tout le monde ne peut pas voir la différence entre les 2). Merci aussi pour nos discussions même si on n'est pas toujours d'accord, mais c'est mieux comme ça !

A Haiko, mon successeur dans le Graduiertenkolleg

A Véro qui gère et s'occupe du coté administratif de nos allers et retours, René qui s'occupe des problèmes techniques (voiture…) et Dominique des petits soucis de tous les jours

Aux anciens, qui m'ont transmis leur savoir, Nathalie C. pour la trappe, Sarah (heureuse maman d'Arthur), toujours souriante et de bons conseils, Nukhet (maintenant parisienne) et Hélène N.

Aux derniers venus et aux stagiaires, Fred 1 l'informaticien bis, Fred 2, qui m'a bien aidé pour mon dernier projet de thèse, Philippe et Mickaël, venus apprendre la spectrométrie de masse.

Enfin, je souhaite remercier ma famille (Maxime pour la mise en page automatique !), Laurent et mes amis les plus proches, pour leur patience et leur compréhension, sans eux je n'aurais pas réussi à aller au bout de cette belle aventure de la même façon. Votre présence m'a profondément aidée, MERCI !

PLAN GENERAL

INTRODUCTI Présentation c	INTRODUCTION GENERALE Présentation du Travail de Thèse7			
PREMIERE PA Evolution de la	RTIE : a spectrométrie de masse en analyse protéomique			
Chapitre I :	La spectrométrie de masse : Principes et description			
Chapitre II :	La protéomique			
SECONDE PAI Résultats : O protéines	RTIE : ptimisation des étapes de l'approche protéomique pour la caractérisation des			
Chapitre I :	Amélioration de la compatibilité des colorants des gels 2D avec l'analyse par spectrométrie de masse			
Chapitre II :	Optimisation des étapes d'extraction et de digestion des protéines <i>in-gel</i> 91			
Chapitre III :	Evaluation des approches MS/MS en analyse protéomique105			
TROISIEME F Résultats : Ap	PARTIE : oplications biologiques : De l'identification à la caractérisation des protéines			
Chapitre I :	Introduction : approches MALDI-MS, nanoLC-MS/MS et séquençage De Novo			
Chapitre II :	Approche MALDI-MS : Identification des protéines			
Chapitre III :	Approche nanoLC-MS/MS			

QUATRIEME PARTIE :

Résultats : La spectrométrie de masse, un outil de choix pour identifier les protéines mais pas seulement...

Chapitre I :	La spectrométrie de r pas seulement	nasse, un outil	de choix pou	ır identifier	les protéines ma 2(is)1
CONCLUSION	N GENERALE				21	7
PARTIE EXPE	RIMENTALE				22	23

Introductio	on et présentation du travail de thèse	7
Partie I :		
Evolution of	le la spectrométrie de masse en analyse protéomique	11
Chapitre 1	:	13
La spectro	nétrie de masse :	13
Principes e	t description	
1 Les so	urces et les interfaces	
1.1 L 1.1.1	a source Electrospray (ESI) Production de gouttelettes chargées	14
1.1.2 1.1.3	L'émission des jouttelettes chargees L'émission des ions en phase gazeuse	
b 1.1.4	Modèle d'Iribarne et Thomson (modèle d'évaporation des ions) Miniaturisation de la source electrospray : le nanospray	
1.1.5	Préparation d'échantillon en ESI et nano-ESI	17
1.2 I 1.2.1 1.2.2 1.2.3	A source MALDI Historique et description Principe d'ionisation Paramètres essentiels	18 18 19 21
a b c	Laser Choix de la matrice Différents dépôts	21 21 22
1.3 L 1.3.1 1.3.2	Le rôle de l'interface La fragmentation induite par collisions dans l'interface	23 23 23 26
2 Les ar	nalyseurs	
2.1 L	e quadripôle	
2.2 L	analyseur à temps de vol (Time Of Flight, TOF)	
2.2.1 a b c 2.2.2	MALDI-TOF en mode linéaire L'extraction retardée (D.E, Delayed Extraction) MALDI-TOF en mode réflecteur Analyse en mode MS/MS Le PSD (Post Source Decay)	
a b c	L'ISD (In Source Decay) Le LIFT	
2.3 L	a trappe ionique Description et fonctionnement	
2.3.2 2.3.3	Analyse en mode MS : piégeage et éjection Analyse en mode MS/MS ou MS ⁿ : isolement et fragmentation	
2.4 L	es analyseurs en tandem	40
2.4.1 2.4.2	Triple quadripôle Q-TOF	

Chapitre 2 :		
La protéomique		
1 Séparation des protéines		
1.1 Gel d'électrophorèse bi-dimensionnel		
1.2 Chromatographie liquide (nano-LC)		
1.3Approches alternatives au gel 2D1.3.1Gel SDS PAGE couplé à la nano-LC1.3.2La chromatographie multidimensionnelle1.3.3Les approches par marquage	45 45 45 47	
2 Identification des protéines		
2.1 Présentation de l'approche protéomique		
2.2 L'empreinte peptidique massique (Fingerprint) : approche MALDI-M	MS 50	
 2.3 Les informations de séquences : approche NanoLC-MS/MS 2.3.1 Les règles de fragmentation peptidique 2.3.2 Identification par de courtes séquences (peptide sequence tag) 2.3.3 Le séquençage <i>De Novo</i> 	51 51 52 53	
3 Recherche dans les banques de données		
3.1 Quelques moteurs de recherche		
3.2 Recherche par alignement de séquences		
4 Quelques logiciels d'archivage		
Bibliographie :		
Partie II :		
Optimisation des étapes de l'approche protéomique pour la caractérisation des pro	otéines. 65	
Chaptere 1 :		
<i>Amenoration ae la compatibilite aes coloranis aes gels 2D avec i analyse par spec</i> <i>de masse</i>	<i>irometrie</i>	
1 Limites des colorants classiques : sensibilité, compatibilité, expertise		
2 Publication		
3 Résultats complémentaires		
3.1 Autres protéines testées		
3.2 Autres colorants testés		
3.3 Analyses avec différentes techniques de spectrométrie de masse		
3.4 Discussion		
Chapitre 2 :	91	
Optimisation des étapes d'extraction et de digestion des protéines in-gel		
1 Tests d'extraction		
2 Tests de digestion		

Chapitre	2 3 :	105
Evaluati	ion des approches MS/MS en analyse protéomique	105
1 Mét	thodologie	106
2 Con MS/MS.	nparaison de recouvrement de séquences des protéines identifiées par nanoLC-	107
3 Con MS/MS.	nparaison de la nature des peptides et de leurs fragments obtenus par nanoLC-	109
4 Etu	des complémentaires, MALDI-TOF/TOF	114
Partie II	И:	119
Applicat	ions biologiques : De l'identification à la caractérisation des protéines	119
Chapitre	e 1 :	121
Introduc	ction : Approches MALDI-MS,	121
nanoLC	-MS/MS et séquençage De Novo	121
Chapitre	2 2 :	125
Approch	e MALDI-MS : Identification des protéines	125
1 Dét	ermination du protéome de noyaux de cellules Caco 2	125
1.1	Intérêt biologique	125
1.2	Approche protéomique descriptive	126
1.3	Publication	126
1.4	Discussion	144
1.5	Approche protéomique différentielle	144
1.6	Perspectives	147
2 Iden	ntification de protéines de mitochondries impliquées dans la maladie MERRF	149
2.1	Contexte biologique	149
2.2	Approche protéomique différentielle	150
2.3	Publication	151
2.4	Discussion	162
2.5 tradu	Résultats complémentaires : Caractérisation de modifications post- ctionnelles	163
2.6	Perspectives : Quantification par marquage des protéines mitochondriales.	166
2.7	Résultats complémentaires concernant la maladie MELAS	167
Chapitre	23:	171
Approch	e NanoLC-MS/MS	171
1 Car responsa	ractérisation par homologie de séquences de protéines de Streptococcus bovis Ibles de l'inflammation du côlon	171
1.1	Contexte biologique	171

1.2	Fractionnement et tests biologiques	172
1.3	Choix du protocole d'analyse : Analyse de la fraction S300	174
1.4	Fraction WEA	178
15	Fraction excrétée	181
1.5	Fraction "Protéine 1" · Utilisation de l'approche Séquencage De Novo	18/
1.0	Fraction Trotemer : Otinsation de l'approche Sequençage De Novo	104
l./ l'inflai	Perspectives : Identification plus fine des proteines responsables de	186
171	Tests biologiques	186
1.7.2	2 Caractérisation par spectrométrie de masse	186
Chapitre	4:	189
Approche	e par séquençage De Novo	189
1 Care	actérisation de protéines parasitaires de Plasmodium falciparum	189
1.1	Contexte biologique	189
1.2	Identifications de sérine hydrolases	191
1.2.1	Préparation des échantillons	191
1.2.2	2 Analyses par nanoLC-MS/MS	192
1.2.3	B Perspectives biologiques	193
1.2.4	Perspectives en spectrométrie de masse	193
1.3	Protéomique différentielle : Erythrocytes sains et parasités	194
1.4	Etude systématique des protéines parasitaires enchassées dans la membr	rane de
l'éryth	rocyte ou à proximité	194
1.4.1	Méthode : Marquage des parasites	194
1.4.2	2 Résultats	195
1.4.3	B Limites de l'approche et perspectives	198
Partie IV	· · · · · · · · · · · · · · · · · · ·	201
La spectr seulemen	ométrie de masse, un outil de choix pour identifier les protéines mais pas	201
1 Oua	ntification de molécules pharmacologianes dans les sérums de patients	201
1 Quu 11	Contorto kiele rique	201
1.1	Contexte biologique	201
1.2	Premières études HPLC: l'hypothèse d'une protéine active réfutée	203
1.3	Dosage de la molécule d'intérêt par spectrométrie de masse	204
1.3.1	Préparation des échantillons (extraction)	204
1.3.2	2. Validation de la méthode	204
1.3.3	B Dosage de la molécule d'intérêt dans les deux sera	206
1.4	Perspectives	207
1.4.1	Utilisation d'un étalon interne	207
1.4.2	2 Analyse d'un mélange de molécules anti-virales	207
Bibliogra	phie :	209
Conclusi	on :	217
Partie ex	périmentale :	223

1	Les s	pectromètres de masse	223	
	1.1	Le MALDI-TOF		
	1.1.1	Les paramètres importants de l'analyse par MALDI-TOF-MS		
	1.1.2	Paramètres importants de l'analyse par MALDI-TOF MS/MS		
	1.1.3	L'étalonnage du spectromètre de masse MALDI-TOF		
	1.1.4	La préparation de l'échantillon		
	1.2	Les spectromètres de masse électrospray		
	1.2.1	Le spectromètre de masse ES-TOF.		
	1.2.2	Le spectromètre de masse Q-TOF		
	а	Introduction des échantillons pour les analyses nano-ESI		
	b Paramètres des analyses MS et MSMS en nanospray			
	c	La préparation de l'échantillon en nanospray		
	1.2.3	La trappe ionique		
	1.2.4	L'étalonnage des spectromètres de masse électrospray		
2	Les s	ystèmes HPLC		
	2.1	HPLC Waters Alliance		
	2.2	Micromass CapLC et nanoLC d'Agilent		
3	Prote	éomique	231	
	3.1	Réduction-alkylation-digestion		
	3.2	L'extraction des peptides		
	3.3	3 La nanoLC-MS-MS232		

Introduction et présentation du travail de thèse

La spectrométrie de masse permet, depuis une dizaine d'années, de répondre à de nombreuses questions biologiques, notamment concernant l'identification des protéines et leur caractérisation (modifications post-traductionnelles, modifications chimiques, mutations d'acides-aminés...)

Ainsi, la spectrométrie de masse, au sein de l'approche protéomique a évolué afin de devenir un outil de choix. En effet, les analyseurs sont de plus en plus résolutifs et sensibles, rendant possible les analyses de quelques femtomoles de produit. Cette évolution s'est faite en parallèle de celle des techniques séparatives (gel d'électrophorèse bi-dimensionnelle, Chromatographie Liquide à nanodébits (NanoLC)), en amont du spectromètre de masse, les rendant mutuellement plus compatibles.

De plus, devant la diversité des problèmes biologiques, une approche protéomique classique (mesure de masses des peptides trypsiques par MALDI-MS (Matrix Assisted Laser Desorption Ionisation- Mass Spectrometry) ne suffit pas toujours. Ainsi, il a été nécessaire de développer des alternatives, possibles grâce aux nombreuses potentialités de la spectrométrie de masse. En effet, la spectrométrie de masse permet d'obtenir des informations sur les masses des peptides intacts, mais aussi des informations de séquences par spectrométrie de masse en tandem (MS/MS), indispensables pour confirmer les identifications litigieuses mais aussi pour caractériser les peptides de protéines n'appartenant pas aux banques de données protéiques.

Dans ce contexte général, l'objectif de mon travail de thèse est d'optimiser les différentes approches de la spectrométrie de masse en analyse protéomique pour passer de l'identification à la caractérisation des protéines. En effet, si l'identification d'une protéine peut se faire à partir de 20% de recouvrement de séquence, sa caractérisation nécessite un recouvrement de séquence maximum, pour déterminer aussi bien les mutations, les modifications post-traductionnelles que les modifications chimiques. Pour atteindre cet objectif, mon travail de thèse a donc porté principalement sur l'optimisation des approches protéomiques, visant à répondre à divers problèmes biologiques, en déterminant quelle était la technique de spectrométrie de masse, la plus adaptée pour répondre aux questions posées. A mon arrivée au laboratoire, nous disposions de deux techniques majeures de spectrométrie de masse : le MALDI-MS et la nanoLC-MS/MS

Ainsi, ce travail de thèse s'articule autour de deux axes principaux :

- L'optimisation des étapes des différentes approches protéomiques pour passer de l'identification globale des protéines à leur caractérisation fine.
- La détermination des approches de spectrométrie de masse (MALDI-MS et/ou nanoLC-MS/MS) les mieux adaptées aux divers problèmes biologiques rencontrés en analyse protéomique. En effet, celles-ci devront être adaptées aux connaissances biologiques dont on dispose : génomes connus ou inconnus, séquences protéiques modifiées ou non, homologie de séquences protéiques avec d'autres organismes.

L'ensemble des développements réalisés dans ce travail l'ont été au travers de problèmes biologiques réels.

Ce manuscrit se compose de 4 parties. La première partie décrit l'évolution de la spectrométrie de masse en analyse protéomique, en résumant les différentes techniques habituellement utilisées en analyse protéomique, à savoir la spectrométrie de masse (ESI-MS (Electrospray), MALDI-MS et MS/MS) ainsi que les différentes techniques de séparation des protéines (gel 1D, gel 2D, nanoLC et 2D-LC).

Ensuite, les différents résultats obtenus seront présentés en 3 parties distinctes.

La première partie, «**Optimisation des étapes de l'approche protéomique pour la caractérisation des protéines**», a été réalisée au travers de 3 études.

Première étude : Amélioration de la compatibilité des colorations des gels 2D avec l'analyse par spectrométrie de masse. Il s'agissait de trouver, en collaboration avec T. Rabilloud, de nouveaux colorants des gels 2D sensibles et compatibles avec la spectrométrie de masse et de mieux comprendre les mécanismes de la coloration au nitrate d'argent.

Seconde étude : Optimisation des étapes de digestion des protéines *in-gel* et d'extraction des peptides du gel. Pour cette étude, différentes conditions d'extraction ont été testées, en modifiant les solvants d'extraction et leurs durées, ainsi que différentes protéases, afin de trouver des conditions optimales nous permettant d'augmenter le nombre de peptides analysés et ainsi le recouvrement de séquence.

Troisième étude : Validation des approches MS/MS en analyse protéomique. Cette étude a pour but de comparer les performances des différentes approches MS/MS (MALDI-TOF/TOF (Time of Flight), NanoLC-Q-TOF et NanoLC-Trappe ionique), par rapport à l'information de séquence protéique obtenue.

La seconde partie de ce travail de thèse, « **Applications biologiques : de l'identification à la caractérisation des protéines** », a consisté à adapter les différentes techniques de spectrométrie de masse pour répondre à différents types de problèmes biologiques rencontrés en analyse protéomique.

Première étude : Détermination du protéome de noyaux de cellules Caco 2 (collaboration avec l'unité INSERM U381 de Strasbourg). Cette étude consiste en l'identification des protéines de noyaux des cellules Caco2 de façon systématique (protéomique descriptive) et ceci à différents stades de différenciation des cellules.

Seconde étude : Identification de protéines mitochondriales impliquées dans la maladie MERRF (myoclonic epilepsy with ragged red fibers) (collaboration avec l'Institut de Biologie Moléculaire et Cellulaire de Strasbourg). Il s'agit d'une analyse protéomique différentielle sur les protéines mitochondriales sur et sous exprimées dans les cellules malades. Cette étude a pour but de mieux comprendre l'impact de la mutation du tRNA lysine mitochondrial, responsable de cette maladie, sur le fonctionnement de la cellule.

Troisième étude : Caractérisation des protéines de *Streptococcus bovis* responsables de l'inflammation du colon et soupçonnées avoir un rôle dans l'aggravation de cancers coliques (collaboration avec l'unité INSERM U392 d'Illkirch). Dans ce cas, la difficulté de l'étude réside dans le fait que seules une trentaine de protéines de *Streptococcus bovis* sont référencées dans les banques. Cependant, de nombreuses bactéries de la même famille (Streptococcus) sont présentes dans les banques, offrant la possibilité d'identifier les protéines après recherche par homologie de séquence.

Quatrième étude: Caractérisation de protéines parasitaires de *Plasmodium falciparum* (collabration avec l'Institut Pasteur à Paris). Il s'agit, d'une part, de caractériser des protéines avec une activité sérine-protéase, après purification par chromatographie d'affinité et séparation sur gel 1D. D'autre part, une étude protéomique différentielle des protéines solubles des fantômes de globules rouges parasités ou non a été réalisée. Enfin, une analyse systématique des protéines parasitaires enchassées ou à proximité de la membrane erythrocytaire a été menée. Pour faciliter cette dernière étude, le mélange protéique a été simplifié en marquant les protéines parasitaires à l'aide de Lysines deutérées. Ici, la difficulté résidait dans le fait que le génome n'était pas entièrement disponible et que seules quelques séquences protéiques étaient connues.

La troisième partie, « La spectrométrie de masse, un outil de choix pour identifier les protéines mais pas seulement... » a porté sur l'utilisation de la micro-LC en couplage avec la spectrométrie de masse pour analyser de petites molécules pharmacologiques présentes dans les fluides biologiques (sérum et plasma) de patients séropositifs résistants au virus VIH (collaboration avec l'unité INSERM U74 de Strasbourg).

Partie I :

Evolution de la spectrométrie de masse en analyse protéomique

- Chapitre 1 : La spectrométrie de masse : Principes et description
- Chapitre 2 : La protéomique

Chapitre 1:

La spectrométrie de masse : Principes et description

1 Les sources et les interfaces

Jusqu'au milieu des années 80, les processus d'ionisation (ionisation par impact électronique ou ionisation chimique) fragmentent les molécules à analyser, soit par des processus très énergétiques (ionisation par impact électronique) soit par instabilité chimique (ionisation chimique). Ces fragmentations permettent d'obtenir de nombreuses informations structurales, mais elles limitent l'application de la spectrométrie de masse à l'analyse de petites molécules (<1000 Da). Cette restriction s'explique par la difficulté à vaporiser et ioniser les macromolécules par bombardement ou par chauffage. En effet, le plus souvent, ces macromolécules étaient pyrolysées avant même d'être ionisées. En 1988, deux nouvelles méthodes d'ionisation dites « douces » ont été développées : l'ionisation par ESI (Electrospray) [Meng et coll., 1988, Fenn et coll., 1989] et l'ionisation par MALDI (Matrix Assisted Laser Desorption/Ionisation) [Karas et Hillenkamp, 1988, Karas et coll., 1989, Tanaka, 1989]. Grâce à ces deux techniques, l'analyse de macromolécules intactes a pu être possible, ouvrant un nouveau champ d'application à la spectrométrie de masse. C'est pourquoi, J. Fenn et K. Tanaka, qui ont développé ces deux techniques ont été récompensés en 2002 par le prix Nobel.

Ainsi, l'ionisation ESI permet sous l'action d'un fort champ électrique de transférer les ions en phase vapeur. L'ionisation MALDI, quant à elle, permet sous l'action d'un faisceau laser de désorber et d'ioniser les molécules cristallisées dans une matrice. L'avènement de ces deux techniques a rendu accessible à la spectrométrie de masse, un nouveau domaine d'étude, celui des bio-molécules. Ainsi,

ces dernières années, l'ionisation MALDI et l'ionisation ESI ont largement participé à l'essor de la spectrométrie de masse dans le cadre des analyses protéomiques, qui correspondent à l'identification et à la caractérisation des protéines. Ces études protéomiques s'inscrivent dans la suite logique des génomiques entreprises dans les années 90.

Ce premier chapitre a pour but de décrire les principes de ces deux méthodes d'ionisation, utilisées à de nombreuses reprises pour mener à bien les différentes études réalisées pendant ce travail de thèse

1.1 La source Electrospray (ESI)

L'ionisation ESI ou encore électronébulisation correspond à l'émission en phase gazeuse d'ions préformés en solution sous l'effet d'un champ électrique intense. Ce processus a lieu à pression atmosphérique contrairement au processus MALDI qui nécessite un vide poussé pour la formation des ions. Par ionisation ESI, l'échantillon est entraîné vers l'analyseur du spectromètre de masse à la fois par des différences de potentiels appliquées à des lentilles électrostatiques et par la différence de pression établie entre la source (pression atmosphérique) et l'analyseur ($10^{-6} - 10^{-7}$ mbar).

Les premiers travaux de Dole en 1968 sur l'utilisation de l'ESI pour générer des ions à partir de polymères en solution n'ont pas été poursuivis, à cette époque, en raison du manque de résultats pertinents obtenus [Dole et coll., 1968]. L'intérêt pour cette technique réapparaît avec les travaux de Fenn qui a pu, par un couplage entre cette source et un analyseur quadripolaire, effectuer les premières mesures de protéines de hauts poids moléculaires. Ce couplage a également permis la mise en évidence de la nature multichargée des protéines durant ce processus, en présentant le premier spectre multichargé d'une protéine de 40kDa [Yamashita et Fenn, 1984a] ; [Yamashita et Fenn, 1984b].

A présent, le mécanisme macroscopique de l'électrospray est bien décrit. Même si quelques divergences persistent sur la désorption des ions de la solution vers la phase gazeuse, le processus d'ionisation-désorption ESI peut être divisé en 3 étapes majeures :

- La production des gouttelettes chargées à partir de l'électrolyte en solution
- La fission des gouttelettes chargées en gouttelettes de plus petites tailles (gouttelettes « filles ») par explosions coulombiennes successives
- L'émission des ions en phase gazeuse

1.1.1 Production de gouttelettes chargées

L'application d'un champ électrique intense (10^6 V/m) à la pointe du capillaire provoque une polarisation du liquide et une séparation des charges positives et négatives. L'accumulation de charges à la pointe du capillaire, déstabilise la surface du liquide qui prend alors la forme d'un cône appelé "cône de Taylor". De ce cône, sont émises des gouttelettes chargées (figure 1). Pour respecter

l'équilibre des charges, la source ESI se comporte comme une cellule électrochimique [Ikonomou et coll, 1991]; [Blade et coll., 1991]. La nébulisation est par ailleurs assistée par un gaz de nébulisation (azote).





1.1.2 Fission des gouttelettes chargées

Dans la source chauffée, l'évaporation du solvant va entraîner la diminution du volume des gouttelettes jusqu'au moment où l'énergie due aux interactions coulombiennes répulsives s'approche du niveau d'énergie des forces de cohésion des gouttelettes et provoque ainsi leur explosion. Les gouttelettes explosent en « gouttelettes filles » avec une densité de charge plus importante. La taille critique de ces gouttelettes est donnée par leur rayon, appelé rayon limite de Rayleigh. La charge de la gouttelette est proportionnelle à son rayon (équation 1). Celui-ci dépend, en plus de la charge de la gouttelette (Q), de la tension superficielle du liquide (γ) et de la permittivité du vide (ϵ_0).

Equation 1 :
$$Q^2 = 64 \pi^2 \epsilon_0 \gamma R_R^3$$

1.1.3 L'émission des ions en phase gazeuse

L'émission des ions à partir des gouttelettes chargées vers la phase gazeuse reste un processus très controversé. Deux modèles sont pourtant admis aujourd'hui, le modèle de la charge résiduelle proposé par Dole [Dole et coll., 1968] et celui de l'évaporation ionique d'Iribarne et de Thomson [Iribarne et Thomson, 1976].

a Modèle de Dole (modèle de la charge résiduelle)

Les étapes successives d'évaporation et d'explosions coulombiennes aboutissent à la formation de gouttelettes filles ultimes ne contenant qu'une seule charge (figure 2).

Figure 2 : Schématisation de l'émission des ions en phase gazeuse lors de l'ionisation/désorption ES selon la théorie de Dole. [Leize, 1994].



b Modèle d'Iribarne et Thomson (modèle d'évaporation des ions)

D'après cette théorie, lorsque les gouttelettes filles ont une densité de charge suffisante correspondant à un rayon précis ($Q = 10^{-17}$ C et R < 10 nm), elles ne subissent plus de fissions supplémentaires. Une fois cette densité de charges atteinte, les ions sont directement émis en phase gazeuse (figure 3).

Figure 3: Schématisation de l'émission des ions en phase gazeuse lors de l'ionisation/désorption ES selon la théorie selon la théorie d'Iribarne et Thomson [Leize, 1994]



Il est désormais admis que le modèle de Dole expliquerait le processus de désorption/ionisation pour les macromolécules (protéines...) et le modèle d'Iribarne et Thomson celui des molécules de petites tailles [Kébarlé, 2000].

1.1.4 Miniaturisation de la source electrospray : le nanospray

En mode d'ionisation électrospray, le spectromètre de masse détecte un signal (courant d'ions) dépendant de la concentration de l'analyte et non du débit avec lequel il est infusé [Ikonomou et coll., 1990]; [Covey et coll., 1990]. Aussi, la diminution du débit permet l'émission de plus petites gouttelettes à la pointe d'un cône de Taylor [Cole, 2000]. La plus grande densité de charge de ces gouttelettes contribue à l'augmentation du rendement d'ionisation/désorption des molécules, ce qui équivaut à une augmentation de la sensibilité. La diminution des débits (environ 50-100nL/min)prolonge le temps d'analyse, puisqu'avec 1 à 2μ L d'analyte, l'analyse pourra durer environ 30 min. Ainsi, pour répondre à ces nouvelles exigences, les sources micro-électrospray, [Emmet et Caprioli, 1994], et en 1996 le nano-électrospray ou nano-spray [Wilm et Mann, 1996] ont été développées.

La source nano-spray est constituée d'un capillaire en quartz étiré de 1 à 3 µm de diamètre interne, recouvert d'une fine couche de métal qui assure sa conductivité. L'efficacité de l'ionisation nano-spray repose sur la faible taille des gouttelettes (environ 200 nm de diamètre), qui leurs confère une densité de charge importante, un rapport surface/volume augmenté permettant une ionisation/désorption plus rapide [Fligge et coll., 1999]. Contrairement à l'ESI, ce montage ne nécessite plus, dans la majorité des cas, de gaz de nébulisation. De plus, la distance avec la contre électrode étant plus courte et pour éviter les décharges de Corona, les tensions appliquées sur le capillaire en quartz sont plus faibles (de l'ordre de 1000V).

Ainsi, la miniaturisation des sources électrospray présente de nombreux avantages pour l'analyse (tolérance aux sels non volatils, augmentation de la sensibilité et de ce fait, augmentation de la précision de la mesure de masse) [Wilm et coll., 1996]. Les diverses applications des sources nano-ESI, notamment dans le domaine de la protéomique, montrent aujourd'hui l'intérêt d'un tel dispositif [Shevchenko et coll. 1996], compatible avec des quantités de produit de l'ordre de la dizaine de femtomole.

1.1.5 Préparation d'échantillon en ESI et nano-ESI

Si la mesure est réalisée en mode d'ionisation positive, la solution est acidifiée pour favoriser la protonation des sites protonables de l'analyte (sites basiques pour les protéines). Pour l'analyse de peptides et de protéines, on considère que ce sont les lysines, arginines ainsi que l'extrémité Nterminale qui seront protonées. Inversement, si on travaille en mode d'ionisation négative, comme pour la détection d'oligonucléotides, de modifications post-traductionelles (phosphorylations), il faudra se placer dans un environnement basique (déprotonation des sites acides) et inverser les polarités appliquées au niveau du spectromètre de masse. Le choix des tampons est une étape cruciale lors des analyses en mode ESI, en raison des contraintes liées à la source et de celles liées à l'échantillon. En effet, l'ESI est une technique particulièrement sensible aux sels non volatils et aux détergents classiquement utilisés en biologie (NaCl, KCl, SDS...). Ceux-ci peuvent venir perturber la formation du spray, empêcher l'étape d'ionisation en entrant en compétition avec les molécules d'analytes ou encore former des adduits et aboutir à l'extinction du signal de la molécule à analyser [Mann et Wilm, 1995]. Il est donc essentiel de porter une attention particulière aux tampons utilisés. C'est pourquoi, une étape de dessalage précédant l'analyse électrospray est souvent réalisée pour éliminer les sels non volatils apportés par les tampons. Ce dessalage peut être réalisé en ligne grâce au couplage chromatographie liquide-spectrométrie de masse détaillé dans le chapitre B 1 ou par l'utilisation de kit de dessalage (Zip TipTM) ou divers autres procédés (dialyse, dessalage par centrifugation...).

1.2 La source MALDI

1.2.1 Historique et description

La spectrométrie de masse MALDI-TOF (Désorption Ionisation Laser Assistée par Matrice-Temps de Vol) permet la mesure de masse de biopolymères pouvant aller de 1 à 500 kDa en pratique, même si, en théorie, sa gamme de masse est illimitée. Les molécules d'analytes sont incorporées à une matrice absorbant fortement à la longueur d'onde du laser. Les ions sont générés par excitation de la matrice par le laser. Cette source est généralement couplée à un analyseur TOF même si de plus en plus de constructeurs développent d'autres couplage (MALDI-Ion Trap, MALDI-Q-TOF, MALDI-Q-Trap...) [Zhang, 2003].

Les principaux atouts de ce mode d'ionisation sont sa sensibilité (détection de quantités subfemtomolaires), sa rapidité d'analyse (quelques secondes) mais également une bonne tolérance aux sels non volatils et aux tampons couramment utilisés en biologie, contrairement à l'ESI. Le spectromètre de masse MALDI-TOF s'adapte donc particulièrement bien à l'étude des biomolécules telles que les peptides, protéines, oligonucléotides, oligosaccharides et polymères synthétiques.

Dès 1976, la désorption laser a été utilisée à partir d'échantillons solides pour produire des peptides intacts en phase gazeuse afin de les analyser par spectrométrie de masse [Posthumus et coll.,1978]. Mais aucun peptide intact de masse supérieure à 1000 Da n'a pu être détecté. Ce n'est qu'à la fin des années 80 que la désorption ionisation laser assistée par matrice a été développée par Karas et Hillenkamp [Karas et coll., 1987] et par une équipe japonaise [Tanaka et coll., 1988]. L'idée a été d'ajouter une matrice en large excès (en général 100 :1 à 10000 :1), absorbant à la longueur d'onde du laser pour protéger l'analyte.

1.2.2 Principe d'ionisation

Les différentes étapes de désorption et d'ionisation MALDI sont résumées ci-dessous [Fenselau, 1997] :

- Absorption par la matrice de l'énergie fournie par le laser UV,
- Ionisation de la matrice
- Dissociation de la matrice
- > Sublimation
- Expansion de la matrice en un jet supersonique
- > Entraînement de l'analyte dans le plasma d'expansion généré par la matrice
- Transfert de charges vers les molécules d'analyte

Figure 3A : Schématisation du processus de désorption/ionisation par MALDI [thèse N. Cavusoglu]



Si le processus d'ionisation MALDI reste peu connu, il est néanmoins bien établi que la protonation, la déprotonation ou encore la cationisation se produisent en phase gazeuse [Glückmann et al., 1999]. De plus, deux grandes phases dans le processus d'ionisation ont été décrites: l'ionisation primaire et l'ionisation secondaire [Zenobi et al., 2003]. Les processus d'ionisation primaire seraient le résultat de l'association entre un mécanisme de photo-ionisation dû aux protons absorbés par la matrice et un mécanisme thermique favorisé par la température élevée (environ 500K) régnant dans le plasma d'expansion. Il faut rappeler que l'ionisation de la matrice dépend de son potentiel d'ionisation, calculé pour les matrices en solution, mais il n'est pas connu pour les matrices sous leurs formes cristallines.

Ainsi, le résultat de cette ionisation primaire est la formation de molécules de matrices excitées.



Le processus d'ionisation secondaire aurait lieu dans le plasma d'expansion par transfert de protons en phase gazeuse [Zigilei et al., 1998]. Deux types de réactions ont lieu pendant ce transfert, d'une part les réactions Matrice-Matrice (1), d'autre part les réactions Matrice-Analyte (2) :

$$M^{+} + M \longrightarrow MH^{+} + (M - H)^{-}$$
 (1)

$$MH^+ + A \longrightarrow M + AH^+$$
(2)

Durant ces processus d'ionisation, l'affinité protonique est un paramètre essentiel qui favorise l'ionisation de l'analyte. Les protéines et les peptides ont une affinité protonique de l'ordre de 240 kcal/mol [Harrison, 1997]. Le tableau 1 résume les affinités protoniques (A.P.) connues et la basicité en phase gazeuse pour un certain nombre de matrices.

La majorité des matrices couramment utilisées ont des valeurs de A.P. inférieures à celles des protéines et peptides à analyser. Dès lors la protonation de ces derniers est thermodynamiquement favorisée. Ainsi l'association d'une matrice à P.A. basse et d'une protéine ou d'un peptide relativement basique, favorise un transfert de proton qui devient exothermique. Cette énergie libérée peut entraîner la fragmentation de l'analyte [Gross et al., 1998], [Jorgensen et al., 1998], ce qui permet d'apporter une explication complémentaire au processus de fragmentation d'ions métastables aussi décrit comme étant la conséquence de l'élévation de la température durant la désorption [Karas et al., 1995]. Au contraire, lorsque l'affinité protonique des co-matrices est supérieure ou équivalente à celle de l'analyte, il y a une baisse significative de la fragmentation [Simmons et al., 1998].

		1	
	Affinités protoniques mesurées,	Basicité en phase	
Matrices	(kcal/mol)	gazeuse mesurées,	Références
		(kcal/mol)	
	223	215	Steenvoorden et al., 1998
4HCCA	183		Burton et al., 1997
	201		Jorgensen et al., 1998
	203 (calculée)		Nelson et al., 1996
	204	197	Burton et al., 1997
2,5 DHB	202.9		Nelson et al., 1996
	204		Steenvoorden et al., 1997
HABA	205		Jorgensen et al., 1998

 Tableau 1 : Valeurs d'affinités protoniques mesurées ou calculées de matrices et d'analytes (glycine, lysine et arginine).

183		Burton et al., 1997
214	206	Steenvoorden et al., 1997
204		Burton et al., 1997
212		Jorgensen et al., 1998
210		Nelson et al., 1996
210.8		Nelson et al., 1996
214		Jorgensen et al., 1998
214.5		Nelson et al., 1996
210.5	202.7	Harrison et al., 1997
235.6	221.8	Harrison et al., 1997
244.8	237	Harrison et al., 1997
	183 214 204 212 210 210.8 214 214.5 210.5 235.6 244.8	183 206 204 206 204 212 210 210 210.8 214 214.5 202.7 235.6 221.8 244.8 237

1.2.3 Paramètres essentiels

a Laser

La longueur d'onde du laser utilisé est un paramètre important pour obtenir une bonne désorption de l'analyte. Son choix dépend de la matrice utilisée et de la nature de l'analyte. Le plus répandu est le laser à azote qui irradie à 337 nm. Le tableau 2 regroupe une liste non exhaustive de différents types de lasers utilisés.

 Tableau 2: Liste de lasers utilisés pour l'ionisation MALDI, longueur d'onde d'émission, énergie et durée de l'impulsion

 [thèse Cavusoglu].

Laser	Longueur d'onde	Energie des photons	Durée de l'impulsion
	nm	eV	ns
Azote	337 (UV)	3,68	< 1
4Nd:YAG x3	355	3,49	5
Nd:YAG x4	266	4,66	5
Er: YAG	2,94 µm (IR)	0,42	85
CO_2	10,6 µm	0,12	100

b Choix de la matrice

La matrice permet la transmission à l'analyte des protons, tout en assurant sa protection. Son choix est dirigé par le type de molécules à analyser et le laser utilisé.

Les caractéristiques des matrices sont les suivantes :

- Absorber à la longueur d'onde du laser (ajustement des longueurs d'onde par rapport à l'analyte devient inutile).
- Protéger l'analyte vis à vis du laser par co-cristallisation de la matrice en excès par rapport à l'analyte.
- Eviter l'agrégation des molécules d'analyte afin d'obtenir des ions moléculaires.
- Désorber l'analyte quel que soit sa taille : processus d'ionisation/désorption indépendant des propriétés intrinsèques de la molécule d'analyte.

Les matrices les plus couramment utilisées sont l'acide α -cyano-4-hydroxycinnamique [Beavis et coll., 1989] (peptides et protéines), l'acide sinapinique (proteines), l'acide 2,5dihydrobenzoïque (oligosaccharides), l'acide 3-hydroxypicolinique (oligonucléotide). Il existe également quelques matrices [Jespersen, 1998] utilisées pour l'étude de complexes non-covalents puisqu'elles permettent de travailler dans des conditions non dénaturantes [Vogl, 1999].

Nom	Structure	λ d'absorption (nm)	Applications
Acide α-cyano-4- hydroxycinnamique (α-HCCA)	HO C COOH	337, 355 et 266 (UV)	Peptides, protéines, oligonucléotides, glycoprotéines
Acide 3,5-diméthoxy-4- hydroxycinnamique (acide sinapinique)		337, 355 et 266 (UV)	protéines
Acide 2,5- dihydrobenzoïque (DHB)	но Соон	337, 355 et 266 (UV) 2,94 et 10,6 μm (IR)	Peptides, oligosaccharides, glycolipides, glycoprotéines
Acide 3- hydroxypicolinique	ОН	337, 355 et 266 (UV)	oligonucléotides

Tableau 3 : Exemples de matrices employées en spectrométrie de masse MALDI

c Différents dépôts

Si le choix de la matrice est primordial, la technique de cristallisation doit aussi être choisie méticuleusement en fonction de l'analyte. En effet, des observations empiriques ont établi que les

différents modes de cristallisation participent à l'optimisation et à la réussite d'une analyse par MALDI-MS.

Voici très brièvement les modes de dépôts les plus usuels :

- Méthode en sandwich
- Méthode de la goutte séchée

La méthode en sandwich consiste à insèrer l'analyte entre 2 couches de matrice. Cette méthode est plutôt employée avec des matrices de la famille des acides cinnamiques. Elle génère des cristaux de petites tailles et un dépôt d'apparence très homogène. Ce dépôt permet des mesures reproductibles d'un tir laser à un autre et il est employé pour les analyses qui requièrent une grande précision de mesure.

La méthode de la goutte séchée (matrice mélangée à l'analyte) génère des cristaux plus grossiers et plus hétérogènes. Cette hétérogénéité se retrouve au niveau des spectres de masse obtenus qui sont d'intensités très variables. Ce mode de dépôt ne permet pas d'obtenir une excellente précision de mesure, étant donnée son hétérogénéité. De plus, les cristaux étant plus résistants aux impulsions laser, ce mode de dépôt possède un grand intérêt pour les analyses par MALDI-MS/MS (spectrométrie de masse en tandem).

De manière générale, il n'est pas possible de donner des règles concernant le mode de dépôt et le choix de la matrice, il est échantillon-dépendant. Ainsi, avant toute analyse, il faudra tester plusieurs matrices, sous différentes conditions expérimentales (évaporation rapide ou lente du solvant, méthode de la goutte séchée ou méthode en sandwich, etc.) [Moniate et coll.,1997]. Cependant, le mode de dépôt goutte séchée est couramment utilisé en analyse protéomique et permet d'obtenir une précision de mesure de l'ordre de la dizaine de ppm (Ultraflex Bruker). Par ailleurs, un des avantages majeurs de l'ionisation MALDI réside dans sa tolérance aux sels non volatils. Cette caractéristique vient essentiellement de la nature solide du dépôt à analyser, ce qui rend possible un « dessalage sur cible » par simple ajout d'une goutte d'eau acidifiée sur le dépôt qui dissoudra les sels et les éliminera.

1.3 Les interfaces

1.3.1 Le rôle de l'interface

Les ions produits par le processus électrospray sont transmis vers l'analyseur par une interface.

L'interface a trois rôles :

- assurer le transfert des ions d'une pression atmosphérique (source) à une pression de l'ordre de 10⁻⁶-10⁻⁷ mbar (analyseur)
- refocaliser le faisceau ionique divergent
- terminer le processus de désolvatation

Les ions sont acheminés vers l'analyseur par un gradient de pression et par un gradient de potentiels appliqués aux différentes lentilles de l'interface.

De nombreuses configurations d'interfaces ont été développées, trois d'entre elles sont schématisées sur les figures 5, 6 et 7. Il s'agit, respectivement, de la configuration d'un appareil EStriple quadripôle (BioQ, Waters), d'un appareil hybride de type ESI-Q-TOF (Q-TOF II, Waters) et d'un instrument ESI-trappe ionique (Esquire 3000+, Bruker).

Les interfaces possèdent aujourd'hui en plus d'un système de lentilles, des multipôles (hexapôles ou octapôles), fonctionnant en mode de radio-fréquence et assurant une meilleure transmission et une refocalisation de tous les ions vers l'analyseur.

Par ailleurs, la géométrie orthogonale et Z-spray des interfaces est une caractéristique de plus en plus répandue. La géométrie de ces sources est particulièrement adaptée à l'utilisation de hauts débits notamment pour les couplages LC-MS, en limitant le passage de molécules neutres dans le spectromètre de masse. En effet contrairement aux molécules ionisées, les molécules neutres ne sont pas influencées par la différence de potentiel établie entre le capillaire de la source ESI et la contreélectrode de l'interface ; elles ne peuvent donc pas être déviées et ne pénètrent pas dans l'interface du spectromètre de masse.

Figure 4: Représentation schématique de l'interface du quattro II (BioQ, Micromass), un seul des trois quadripôles de l'appareil est représenté pour plus de clarté [Rogniaux, thèse 2000].





Figure 5 : Représentation schématique de l'interface de l'ESI-Q-TOF (Q-TOF II, Micromass) [Sanglier, 2002].

Figure 6 : schéma de la trappe ionique (Esquire 3000+, Bruker).



1.3.2 La fragmentation induite par collisions dans l'interface

Les différences de potentiel générées entre la source puis les lentilles successives de l'interface vont accélérer les ions. Ces derniers ayant accumulés une énergie interne plus ou moins élevée en fonction de l'accélération donnée pourront se fragmenter par collisions avec des molécules de gaz résiduelles. Il est donc possible de favoriser les fragmentations en augmentant la différence de potentiel, par augmentation de la tension d'accélération Vc. Inversement ces risques de fragmentation sont fortement diminués à des tensions d'accélération basses. Ainsi, en contrôlant l'énergie communiquée aux ions et la pression de gaz appliquée dans l'interface, il devient possible de préserver les gros édifices non-covalents, d'origine biologique [Potier N. thèse 1996], [Rogniaux H. thèse 2000], [Sanglier S. thèse 2002], ou issus de la chimie de synthèse [Leize E. thèse 1994], [Nierengarten H. thèse 2001].

2 Les analyseurs

Une grande variété d'analyseurs peut être couplée à une source à ionisation ESI ou MALDI. Dans un premier temps, l'ionisation ESI a été couplée à des analyseurs quadripôlaires [Meng et coll. 1988]. Actuellement de nombreuses combinaisons sont possibles : ESI-trappe ionique (ESI-IT) [Van Berkel et coll. 1990, Thèse Carte N. 2001], ESI-temps de vol (ESI-TOF) [Verentchikov et coll. 1994], ESI-résonance cyclotronique (ESI-FTICR) [Henry et coll. 1989], ESI-secteur magnétique ou électrostatique [Gallagher et coll. 1990]. La tendance actuelle est également aux géométries hybrides, de type ESI-Q-TOF [Morris et coll. 1997], développées pour des applications MS-MS. Par ailleurs, les sources MALDI, même si elles sont couplés en majorité à des analyseurs à temps de vol (MALDI-TOF), peuvent aussi être couplées à des analyseurs de types Trappe ionisque ou Q-TOF [Zhang, 2003].

Pour caractériser les performances de ces analyseurs, il est courant de comparer leurs résolutions (tableau 4). Elle peut être définie par la capacité à distinguer une masse M par rapport à M+ Δ M ; elle est donnée par la relation : $R = \frac{M}{\Delta M}$ (figure 7). A titre d'exemple, un appareil avec une résolution de 1000 permettra de distinguer un ion de m/z 1000 de celui à 1001, en fait, la résolution définit la précision de mesure de masse obtenue sur un spectromètre de masse.

Figure 7 : La résolution est le pouvoir séparateur du spectromètre de masse, elle est définie par la relation encadrée ci-dessus.



Tableau 4 : Performances des analyseurs dans des conditions de fonctionnement standard

Analyseurs	Résolution	Gamme de masse (m/z)
Quadripôle	2000	8000
Magnétique (EB)	20000	20000
Temps de vol	20000	100000
Trappe ionique	5000	6000
Cyclotron à résonance des		
ions	1000000	4000

2.1 Le quadripôle

Le quadripôle est l'analyseur le plus couramment couplé à la source ESI. Il est capable d'analyser un courant d'ion continu et la mesure du rapport m/z de ces ions se fait en fonction de leur trajectoire dans le quadripôle. Son principe de fonctionnement a été décrit en détails par Campana [Campana, 1980]. Nous rappellerons ici son fonctionnement dans les grandes lignes.

Le quadripôle est composé de 4 barres parallèles. Ces barres sont électriquement reliées deux à deux (figure 8). Les barres adjacentes sont à des protentiels opposés (Φ_0 et - Φ_0) et les barres opposées au même protentiel. Le potentiel Φ_0 comporte une composante sinusoïdale Vcosot , où V est l'amplitude de la radio-fréquence, de fréquence ω et une composante continue U.

$$\begin{cases} \Phi_0 = U - V \cos \omega t \\ -\Phi_0 = -U + V \cos \omega t \end{cases}$$

Les ions traversent le quadripôle suivant l'axe z. Le balayage des valeurs U et V en maintenant le rapport U/V constant permet aux ions de valeurs m/z différentes de traverser successivement le quadripôle avec une trajectoire stable. Cette trajectoire est donnée par les équations différentielles de Mathieu. Il existe en réalité deux types de solutions aux équations différentielles : les trajectoires stables qui permettent aux ions de traverser le quadripôle et les trajectoires instables qui

entraînent la déflection des ions sur les barres du quadripôle et leur élimination. La zone de stabilité d'un ion est représentée sur la figure 8A et la résolution d'un quadripôle est donnée dans le tableau 4.



Figure 8A : Zones de stabilité en fonction de U et V des ions de différentes masses (m1<m2<m3)

Figure 8 : Schéma d'un analyseur quadripolaire avec la représentation de la trajectoire des ions selon l'axe z [thèse Bitsch F.]



2.2 L'analyseur à temps de vol (Time Of Flight, TOF)

Le concept de l'analyseur à temps de vol date des années 50. Cependant, l'électronique d'acquisition et de détection n'étaient pas suffisamment perfectionnées pour être efficaces. Il fut repris plus tard [Wiley et coll., 1955], mais son véritable avènement a été possible grâce au développement de la source MALDI en 1987 [Karas et coll., 1987], [Tanaka, 1989]. Le principe de base d'un analyseur TOF consiste à mesurer le temps de vol d'un ion, préalablement accéléré, au travers d'une région libre de champ (le tube de vol), pour atteindre le détecteur (figure 9). Le rapport m/z est directement corrélé au temps de vol ; en effet, les ions de m/z élevé arriveront plus tardivement au détecteur que ceux de faible m/z.

Figure 9 : Principe de l'analyseur à temps de vol



La force appliquée sur un ion dans un champ électrique est :

$$F = zeV = \frac{1}{2}mv^2$$

z : charge de l'ion e : charge de l'électron (1,602.10⁻¹⁹ C) V : différence de potentiel entre la cible et la grille d'extraction m : masse de l'ion (kg) v : vitesse de l'ion (m/s)

On en déduit la vitesse de l'ion :

$$v = \sqrt{\frac{2zeV}{m}} \propto \sqrt{\frac{z}{m}}$$

La vitesse des ions est inversement proportionnelle à la racine carré du rapport m/z et le temps mis par un ion pour traverser le tube de vol de longueur L est de :

$$t = L \sqrt{\frac{m}{2zeV}} \propto \sqrt{\frac{m}{z}}$$

L'analyseur à temps de vol peut être couplé aux sources ESI et MALDI (configuration la plus courante). C'est pourquoi, nous allons décrire les différentes possibilités d'analyse de cet analyseur couplé avec une source MALDI, notamment en analyse protéomique.

2.2.1 Analyse en mode MS

a MALDI-TOF en mode linéaire

L'analyseur à temps de vol linéaire est de conception assez simple (figure 10). Après irradiation laser, les ions sont expulsés de la source par paquets, accélérés par un gradient de potentiels décroissants (20 à 15kV) qui va conférer la même énergie cinétique à tous les ions formés.

Figure 10 : Schéma du MALDI-TOF linéaire, [Cavusoglu, 2002].



L'analyseur en mode linéaire a une bonne sensibilité et une résolution (M/ Δ M) de l'ordre de 1000 [Beavis et coll., 1991 ; Zhou et coll., 1992]. Cette résolution très moyenne est en réalité due à un élargissement des pics dont les causes principales sont les suivantes :

- > La distribution en énergie cinétique initiale non-homogène
- Les petites différences de temps de vol (de l'ordre d'une centaine de nanosecondes pour des ions de petites masses).
- Les collisions, pouvant survenir pendant le processus d'ionisation/désorption qui accentuent les dispersions en temps et énergie.
- Par conséquent, l'influence de ces phénomènes est amplifiée lorsqu'on veut analyser des molécules de plus hautes masses. En effet, travailler à des potentiels d'accélération plus élevés entraîne également l'augmentation des risques de collisions.

b L'extraction retardée (D.E, Delayed Extraction)

Appelée « time lag focusing » par ses concepteurs, [Wiley et Mc Laren, 1955], le principe de l'extraction retardée a été utilisé par Cotter et al., qui ont remarqué que le délai entre l'ionisation et l'extraction entraînait une dispersion spatiale des ions (due à la distribution en énergie cinétique initiale) ce qui a pour conséquence une perte de la résolution. L'extraction retardée est basée sur le concept de la focalisation des ions de vitesses initiales différentes par l'utilisation de potentiels d'accélérations appropriés. L'introduction d'un délai commun à tous les ions formés dans la source, en les refocalisant ainsi dans l'espace va corriger la perte de résolution [Cotter et al., 1989], [Van Breemen et al., 1983], [Spengler et al., 1990]. Ainsi la simple introduction de la lentille d'extraction retardée va faire passer la résolution de 1000 à 5000 en mode linéaire et au delà de 10000 en mode réflecteur [Colby et al., 1994] ; [Brown et Lennon, 1995] ; [Vestal et al., 1995] ; [Witthal et Li, 1995].

c MALDI-TOF en mode réflecteur

La résolution a pu être améliorée par l'ajout à l'extrémité du TOF d'un réflecteur électrostatique. Le réflecteur électrostatique est composé d'une série d'anneaux ou de grilles de potentiels croissants à l'extrémité du tube de vol qui agissent comme un miroir électrostatique (figure 11) [Mamyrin et coll., 1973].



Figure 11 : Schéma du MALDI-TOF avec réflecteur [Cavusoglu, 2002]

En effet, les ions de mêmes masses ayant un petit décalage en énergie vont être ralentis, réfléchis et ré-accélérés plus tardivement par le réflecteur. Ainsi ces ions auront été refocalisés temporellement à leur arrivée au niveau du détecteur, entraînant un gain de résolution. Le MALDI-

TOF avec réflecteur a une résolution comprise entre 15000 et 20000 et l'augmentation de la distance du vol va accroître la précision de la mesure. L'analyse en mode réflecteur ne pourra, cependant s'appliquer qu'à des molécules dont la masse n'excèdera pas 10000 Da.

2.2.2 Analyse en mode MS/MS

a Le PSD (Post Source Decay)

Les ions ayant accumulés trop d'énergie interne pendant l'ionisation peuvent se dissocier dans le tube de vol [Spengler et al., 1991].

Ces ions métastables ne peuvent être vus par un détecteur linéaire puisqu'ils sont générés dans le tube de vol et gardent la même vitesse que l'ion parent [Kaufmann et al., 1993]. Par contre l'ajout du réflecteur, dont le rôle est de refocaliser les ions en énergie, permet de distinguer les ions fragments [Spengler, 1997]; [Chaurand et al., 1999]. Il est nécessaire pour détecter les ions métastables dans toute la gamme de masse d'abaisser progressivement le potentiel du réflecteur, qui agit comme un analyseur d'énergie. Le spectre complet est ensuite reconstitué par concaténations des signaux obtenus dans les différentes gammes de m/z.

b L'ISD (In Source Decay)

A l'opposé du PSD, l'ISD va avoir besoin d'une extraction retardée qui va augmenter la probabilité de collision et donc de fragmentation des ions de la source. Les ions, fragments issus de l'ISD en source vont être refocalisés dans l'espace par le biais des lentilles d'extraction puis d'accélération et auront tous la même énergie cinétique initiale. Les fragments seront donc détectés en mode linéaire tout comme en mode réflecteur [Katta et al.,1998].

c Le LIFT

Si l'analyse en mode PSD (cf a PSD) permet l'obtention d'informations de séquences, elle n'en est pas moins une technique longue, fastidieuse et nécessitant plusieurs picomoles de produit, ce qui n'est pas compatible avec l'analyse protéomique. Pour faciliter ces études, différents constructeurs ont développés des appareils de type MALDI TOF/TOF où le premier TOF permet de sélectionner l'ion parent et le second, après passage dans une cellule de collision (avec ou sans gaz de collision), analyse les ions fragments générés [Yergey et coll. ; 2002]. Ce type d'analyses est rapide et permet d'obtenir des fragments issus de collisions hautes énergies.


Figure12 : schéma MALDI TOF/TOF (Ultraflex, Bruker).

La société Bruker a développé une variante du MALDI TOF/TOF (figure 12), utilisant la technique LIFT (ascenseur à potentiel) [Suckau et coll. ; 2003]. En effet, lors des analyses PSD classiques, le réflecteur peut focaliser uniquement des ions qui différent en énergie cinétique de moins de 30%, d'où un temps d'analyse conséquent (environ 1 heure). Les acquisitions doivent être réalisées à différentes gammes de potentiel sur le réflecteur pour pouvoir refocaliser les ions sur l'ensemble de la gamme de m/z. La technique LIFT consiste à augmenter l'énergie cinétique de tous les ions (+19 keV) à leur sortie de la cellule de collision afin d'avoir des différences d'énergie cinétique inférieures à 30% entre tous les ions fils et l'ion précurseur. Ainsi, le réflecteur pourra refocaliser l'ensemble des ions en une seule étape, d'où l'obtention d'un spectre MS/MS complet en quelques tirs laser sur quelques centaines de femtomoles de produit.

2.3 La trappe ionique

2.3.1 Description et fonctionnement

La trappe ionique (Figure 13) est constituée de deux électrodes quasi hyperboliques, électrodes chapeaux, et d'une électrode annulaire soumises à différents potentiels en fonction de l'étape d'analyse.





Figure 14 : schéma de la trappe ionique (Esquire 3000+, Bruker) couplée à une source ESI.



2.3.2 Analyse en mode MS : piégeage et éjection

Les ions sont injectés et extraits de la trappe ionique par les orifices situés sur les électrodes chapeaux. Lors de l'injection des ions dans la trappe, les électrodes chapeaux sont à la terre tandis que l'électrode annulaire soumise à une radiofréquence ($Vcos\omega_{RF}t$) assure le piégeage des ions en créant un champ quadripolaire dans la trappe ionique. De l'hélium est maintenu à pression constante (environ 5.10⁻⁵ mbar) dans la trappe, pour ralentir les ions et augmenter ainsi l'efficacité du piégeage. Lors du piégeage, les trajectoires des ions épousent le champ quadripolaire en suivant le modèle d'une courbe de Lissajous (figure 15), due aux changements de polarité alternatif de l'électrode annulaire (figure 16).





Figure 16 :Représentation d'un ion piégé par le champ quadripolaire électrique établi dans la trappe ionique. Changement successif de la polarité de l'électrode annulaire qui assure le confinement des ions, les électrodes chapeaux étant mises à la terre (adaptée d'après Walther, H. 1990).



Plus précisément, le champ quadripôlaire appliqué peut être assimilé à un puit de potentiel dirigé suivant l'axe horizontal de la trappe ionique (z) [Dehlmet, HG, 1967]. La profondeur de ce puit est proportionnelle au carré de l'amplitude (V) de la radiofréquence appliquée sur l'électrode annulaire pour les ions de fréquences faibles (fréquence d'oscillation ionique, f ion). La capacité à piéger les ions sera alors d'autant plus forte que l'amplitude (V) de la radiofréquence sera grande (Figure 17) [Marche et Landry, 1995]. Figure 17 :Représentation du puits de potentiel établi dans la trappe. Variation de l'amplitude de la radiofréquence de l'électrode annulaire pour un ion de même rapport m/z.



Par ailleurs, pour une amplitude constante (V) de la radiofréquence, les ions de m/z différents ne subissent pas le même potentiel de piégeage : la profondeur du puit étant inversement proportionnelle au rapport m/z. Ainsi, une grande amplitude (V) de la radiofréquence sera favorable au piégeage des ions de m/z élevés, mais inadaptée au piégeage efficace des ions de faibles rapports m/z(Figure 18).

Figure 18 : Représentation du puit de potentiel établi dans la trappe pour une amplitude de la radiofréquence constante en fonction du rapport masse sur charge (m/z) des ions piégés.



Il existe une limite inférieure du piégeage des ions dans la trappe pour les ions de faible rapport m/z. dont les fréquences d'oscillation seront trop importantes pour qu'ils puissent être piégés. Cette limite inférieure du piégeage ionique est appelée cut-off et dépendra de l'amplitude de la tension alternative de piégeage (V) et de la radiofréquence ω_{RF} .

Les trajectoires ioniques des ions piégés par le champ quadripolaire $(\Phi_{(r,z)})$ de la trappe peuvent être décrites suivant un système d'équations différentielles de Mathieu.

La résolution de ce système aboutit à des trajectoires ioniques stables et instables qui dépendent du rapport masse sur charge des ions, de leurs accélérations définies par les voltages appliqués sur les électrodes mais également de la géométrie de la trappe ionique.



Figure 19 : Diagramme de stabilité des ions A, B, C ($m/z_A > m/z_B > m/z_C$) soumis à un champ quadripolaire ($\Phi_{(r,z)}$).

Les deux zones de stabilité sont délimitées par une frontière où $q_z = 0.908$ (figure 19):

pour des valeurs de $q_z < 0,908$, les ions ont une trajectoire stable et sont piégés dans la trappe par le potentiel ($\Phi_{(r,z)}$).

pour des valeurs de $q_z > 0,908$, les ions sont dans une zone dite instable, ils ne sont plus piégés par le potentiel ($\Phi_{(r,z)}$). Les ions sont déflectés sur les électrodes.

Le facteur de stabilité q_z dont l'expression littérale est indiquée figure 19 est proportionnel à l'amplitude de la tension alternative et inversement proportionnel au rapport *m/z*.

Si l'on augmente l'amplitude de la tension alternative (V), les ions se déplacent vers des valeurs de q_z croissantes (figure 19). Lorsque les valeurs q_z des ions passeront la frontière de stabilité (q_z =0,908), ils seront éjectés de la trappe à commencer par les ions de rapport m/z les plus faibles. Un balayage d'amplitude croissante de la tension alternative sur l'électrode annulaire permet alors d'éjecter de la trappe ionique les ions suivant un ordre croissant des *m/z*.

Cette technique d'éjection sélective est limitée par l'amplitude de la tension alternative maximale applicable (amplitude de 20kV d'après March, R.E. et coll. 1995). Il existe alors des ions « lourds » pour lesquels l'amplitude maximale de la tension alternative ne sera pas suffisante pour les

éjecter de la trappe ionique. Cette éjection simple par balayage croissant de l'amplitude de la tension alternative limite la gamme de masse à une valeur de détection de m/z 650.

D'après l'expression de q_z , la gamme de masse de l'analyseur trappe ionique peut être étendue, en diminuant la taille de la trappe (r_0 , z_0), et/ou en réduisant la fréquence de la tension alternative appliquée à l'électrode annulaire (ω_{RF}) [Kaiser, R.E. et coll. 1995]. Toutefois, si la gamme de m/z est augmentée, c'est au détriment de la résolution et de la sensibilité. La méthode la plus couramment utilisée pour augmenter la gamme de masse d'un analyseur trappe ionique, est l'éjection résonante.

L'éjection par résonance est réalisée en déplaçant sur le diagramme de stabilité la frontière $q_z=0,908$ vers une valeur plus faible. Cette éjection est réalisée en transmettant aux ions de l'énergie cinétique suivant l'axe z de la trappe par l'intermédiaire d'une tension alternative supplémentaire (A $\cos\omega$ 't) appliquée sur les électrodes chapeaux. Les ions soumis au potentiel croissant de l'électrode annulaire rentrent successivement en résonance dans la trappe ionique. L'application d'une tension alternative sur les électrodes chapeaux va augmenter la fréquence axiale des ions déjà résonants, provoquant leur éjection prématurée de la trappe.

Le spectromètre de masse, ESQUIRE 3000+ de Bruker, possède trois gammes de masse étendues avec des points d'éjection q_z spécifiques (Tableau 5).

Tableau 5 : V	aleurs des	points d'	éjection d	les ions	suivant la	gamme c	le masse d	le l'anal	yseur
---------------	------------	-----------	------------	----------	------------	---------	------------	-----------	-------

Gamme de masse étendue à	Valeur de q _z		
<i>m/z</i> 2200 <i>m/z</i> 3400	0,78 0,45		
<i>m/z</i> 6000	0,25		

2.3.3 Analyse en mode MS/MS ou MSⁿ: isolement et fragmentation

Pour les analyses de spectrométrie de masse en tandem (MS/MS), l'isolement d'un ion spécifique s'effectue en éjectant tous les autres ions de la trappe ionique. L'isolement sélectif est suivi d'une étape de fragmentation au cours de laquelle de l'énergie cinétique est communiquée aux ions précurseurs. Les ions précurseurs, dont l'énergie cinétique augmente, rentrent en collision avec les

molécules d'hélium. Si les collisions sont efficaces, elles induisent la dissociation des ions précurseurs en ions fragments appelés ions fils.

Ce processus en deux étapes (isolement et fragmentation) peut se répéter maintes fois (MS/MS/MS.../MS ou MSⁿ) et s'avère utile lors de l'élucidation structurale de composés.

Les différentes étapes du processus d'isolement d'un ion précurseur sont les suivantes :

- Elimination des ions de m/z < m/z des ions précurseurs : balayage croissant de l'amplitude (V) de la tension alternative de l'électrode annulaire jusqu'à une amplitude proche de l'amplitude nécessaire pour l'éjection des ions précurseurs.
- Expulsion de la trappe, des ions de m/z supérieurs aux ions précurseurs : balayage croissant en fréquence de la tension alternative des électrodes chapeaux qui se limite à une fréquence proche de la fréquence de résonance des ions précurseurs (pour éviter de les éjecter). Les ions de m/z plus élevés que les ions précurseurs, ont des fréquences d'oscillation plus faibles que les ions à isoler.
- Isolement plus fin des ions précurseurs. Des balayages successifs en fréquences inférieures puis supérieures à la fréquence de résonance des ions précurseurs, éjectent les ions résiduels proches en m/z des ions à isoler.

Les ions précurseurs piégés dans la trappe ionique seront déstabilisés de leur trajectoire en appliquant une tension alternative sur les "électrodes chapeaux", à la fréquence de résonance des ions précurseurs. Les ions précurseurs dopés en énergie cinétique suivant l'axe horizontal de la trappe ionique, ne sont pas déflectés sur les électrodes chapeaux, puisqu'ils rentrent en collision avec les molécules d'hélium pour se fragmenter. La trappe ionique contient un gaz tampon, de l'hélium à une pression de l'ordre de 5.10⁻⁵ mbar, dont le rôle est de diminuer l'énergie cinétique des ions par collision.

Le choix du gaz est lié au type d'instrument. En effet, dans la trappe ionique, les ions sont piègés, ainsi, ils rentreront plusieurs fois en contact avec des molécules de gaz et accumuleront de l'énergie interne à chaque collision. De ce fait, même si les gaz lourds tels que l'argon et le xénon sont connus pour rendre la MS/MS plus efficace [Vachet R. W., et coll. 1996 et 1998], dans le cas de la trappe ionique, l'utilisationde l'Hélium à une faible pression sera possible.

Au contraire, pour un instrument de type Q-TOF, les ions précurseurs ne traversant qu'une seule fois la cellule de collision, la pression de gaz devra être élevée et les gaz choisis seront des gaz lourds pour rendre la collision efficace. En effet, la probabilité pour un ion précurseur de rencontrer une molécule de gaz, dans la cellule de collision est faible.

Les ions fils, obtenus après fragmentation, sont alors expulsés de la trappe comme décrit pour l'analyse MS (cf 2.3.2)

2.4 Les analyseurs en tandem

2.4.1 Triple quadripôle

Le triple quadripôle (figure 4) est constitué de trois quadripôles en série (Q1, Q2 et Q3), il est souvent utilisé pour étudier la fragmentation des ions par décomposition induite par collision (CID).

Q1 et Q3 agissent dans ce cas comme simples filtres de masse et Q2 comme cellule de collision où les ions entrent en collision avec des molécules de gaz (Ar, He, $N_2...$). L'intérêt de l'utilisation du triple quadripôle vient des propriétés intrinsèques des quadripôles. En effet, il est possible de travailler avec des modes de balayage différents de manière indépendante.

Trois modes de balayage sont couramment utilisés pour les triples quadripôles:

- La détection d'ions fils : Sélection d'un « ion parent » dans Q1, sa fragmentation dans Q2 et analyse des « ions fils » dans Q3 (MS/MS classique).
- La perte de neutre : Détection des ions qui auraient perdu après passage dans Q2 une masse bien définie. Le rôle de Q3 est de permettre l'analyse exclusive des ions ayant perdu la masse définie au départ (par exemple, études de phosphorylation...). Q1 et Q3 balayent simultanément en déphasage correspondant à la masse définie.
- La détection d'ions parents : Détection des « ions parents » qui ont donné par fragmentation dans Q2 des ions fils de masse prédéfinie. Ce mode permet de détecter les ions pouvant témoigner de l'existence de modifications posttraductionnelles dans un mélange complexe de peptides [Huddleston et coll., 1993a et b]. Q1 balaye sur toute la gamme de masse, alors que Q3 est calé sur la masse de l'ion fragment d'intérêt.

2.4.2 Q-TOF

Le Q-TOF est un instrument hybride à source ESI ou MALDI composé d'un premier analyseur quadripôlaire couplé à un analyseur à temps de vol (TOF), orienté perpendiculairement (figure 6). Entre les deux analyseurs se situe la cellule de collision, les ions fragments sont focalisés et envoyés par paquets dans le TOF par un système de lentilles qui permet la thermalisation (ralentissement) ainsi que la refocalisation des ions sortant de la cellule de collision. Le "pusher" situé à l'entrée du TOF donne ensuite l'impulsion nécessaire aux ions pour les ré-accélérer dans une direction perpendiculaire. Les principaux avantages du couplage d'un quadripôle avec un analyseur TOF sont le gain en sensibilité, l'augmentation de résolution et des possibilités de mesures sur une large gamme de m/z, en théorie illimitée. Un des inconvénients est de ne plus pouvoir faire de balayage sur des ions spécifiques puisque le TOF mesure le temps de vol de tous les ions, sur toute la gamme de masse, sans moyen de contrôle.

Chapitre 2 :

La protéomique

Le séquençage de plusieurs génomes, dont le génome humain, donne accès, après traduction, à de nombreuses séquences protéiques codées par les gènes. Mais le fait de disposer de ces gènes ne permet pas de savoir lesquels sont exprimés dans chaque cellule dans les différentes situations physiopathologiques rencontrées. De plus, le faible nombre de gènes eucaryotes laisse à penser qu'une grande partie de la complexité de ces organismes va être représenté par des variations au niveau des protéines exprimées. En effet, si les protéines sont le résultat de la traduction des gènes, l'ajout de modifications post-traductionnelles et les modes d'épissage différentiels contribuent à la production de protéines différentes à partir d'un même gène.

Ainsi, de la même façon que l'ensemble des gènes d'un organisme constitue son génome, l'ensemble des protéines d'un organel, d'une cellule ou d'un tissu constitue son protéome. Mais si le génome est constant dans les cellules d'un organisme, ce n'est pas le cas du protéome de la cellule qui varie en permanence en réponse aux différentes conditions de vie de la cellule [Kenyon et coll., 2002]. Cette définition a été illustrée dans l'article de Lottspeich intitulé, «même génome, différentes protéomes,» lorsqu'il compare la chenille au papillon [Lottspeich, 1999].

L'analyse protéomique consistera à étudier l'expression des protéines au cours de divers processus cellulaires comme la différenciation, la transformation cancéreuse ou l'action d'agents pharmacologiques [Rabilloud, 2001], [Anderson et Seilhamer, 1997], [Gygi et coll., 1999a].

Cette approche étant la combinaison de nombreuses techniques, son développement n'a été possible qu'en raison de nombreuses avancées technologiques, tant au niveau du séquençage des génomes qu'au niveau de l'évolution des gels bi-dimensionnel, de la spectrométrie de masse et des outils informatiques (logiciels d'interprétation et d'archivage des données, capacités de stockage...).

Ainsi, afin de mieux comprendre cette approche, les différentes techniques utilisées seront évoquées selon 3 parties : la séparation des protéines, l'analyse par spectrométrie de masse et l'identification des protéines au travers des recherches dans les banques de données.

1 Séparation des protéines

1.1 Gel d'électrophorèse bi-dimensionnel

L'outil séparatif le plus couramment utilisé en analyse protéomique est basé sur les techniques d'électrophorèse bi-dimensionnelle, possédant une résolution suffisante pour séparer des variants post-traductionnels et une dynamique d'analyse quantitative appréciable [O'Farrel, 1975].

L'échantillon déposé sur gel est séparé suivant une première dimension en fonction du point isoélectrique (pI) des protéines (iso-électrofocalisation, IEF) puis suivant une deuxième dimension en fonction de leurs masses moléculaires (électrophorèse SDS PAGE) [Kenrick et coll., 1970].

L'IEF possède une grande résolution qui permet de distinguer certains variants posttraductionnels contrairement à l'électrophorèse SDS. Mais, le spectre reste limitant, c'est à dire qu'une certaine catégorie de protéines (hydrophobes, de pI extrêmes) ne seront pas visibles. A l'opposé, l'électrophorèse SDS est une technique robuste et éprouvée qui possède un très large spectre [Rabilloud ; 2001].

Les limites liées à la technique sont, malgré tout, assez nombreuses :

La résolution :

Certaines modifications post-traductionnelles ne sont pas détectables, leurs

propriétés physico-chimiques étant peu altérées.

Protéines minoritaires non visibles pour conserver une densité de tâches (spots)

faibles (éviter de surcharger le gel, au risque de ne plus distinguer les spots)

permettant de comparer plusieurs échantillons

La quantification :

Faible détection par rapport à la dynamique d'expression protéique dans la cellule.

Le spectre d'analyse :

Protéines avec des masses moléculaires inférieures à 10 kDa ou supérieures à 300 kDa non détectées.

Protéines avec des pI inférieurs à 4 ou supérieurs à 10 non détectées.

Différentes solutions ont été proposées pour pallier à ces problèmes :

Développement des gradients de pH immobilisés [Bjellqvist et coll., 1982] ;

[Görg et coll., 1988].: - Meilleure détection des protéines basiques

- Amélioration de la résolution

Développement de nouvelles **techniques de détections** plus sensibles et linéaires [Shevchenko, 1996], [Giometti et coll., 1991], [Rabilloud et coll.,2001] : - Amélioration de la résolution - Détection de protéines minoritaires Développement de **nouveaux agents de solubilisation** [Rabilloud 1996], [Rabilloud et coll., 1997], [Rabilloud et coll., 1994] ; [Sanchez et coll., 1997] : - Meilleure détection des protéines peu solubles et de hauts poids moléculaires Réduction du nombre de protéines différentes en travaillant sur des organelles ou encore des fractions membranaires [Rabilloud et coll., 1997]: - Détection des protéines minoritaires

L'introduction des gradients de pH immobilisés ont permis de progresser pour l'analyse des protéines basiques et pour améliorer la résolution via les gradients de pH étroits. De la même manière, les progrès récents de la solubilisation des protéines ont déjà accrus les possibilité d'analyse (spectres d'analyse). Cependant, malgré tous les progrès réalisés, il n'est pas sûr que cette technique réussisse à prendre en compte toute la complexité du protéome d'une cellule (ex : protéines de pH extrêmes, protéines membranaires ..). De plus, cette technique reste difficilement automatisable ce qui pose alors, le problème des analyses à haut débit. C'est pourquoi, d'autres approches ont été développées : approche par chromatographie liquide mono ou bi-dimensionnelle, techniques de marquage, couplage gel SDS/nano-LC [Rabilloud ; 2001].

1.2 Chromatographie liquide (nano-LC)

En protéomique, la chromatographie liquide (cf partie experimentale) a pour objectif de concentrer et de séparer un mélange plus ou moins complexe de peptides. Plus le volume dans lequel est élué la molécule est petit, plus sa concentration sera importante et meilleure sera l'intensité du signal. En effet, la réponse est concentration dépendante et non débit dépendant. [Ikonomou et coll. ; 1990]. Le développement de la nano-LC a 2 objectifs :

- Améliorer la limite de détection afin de pouvoir travailler avec des quantités de plus en plus faibles de matériel

- Rendre les nano-colonnes et la nano-LC compatibles avec les sources nano-ESI.

De ce fait, de grandes avancées techniques ont été réalisées au niveau des nano-colonnes (techniques de remplissage, nouvelles particules plus fines...) et au niveau de la chromatographie (nano-débits). Ainsi, du fait de la bonne performance de ce couplage nanoLC-MS ou nanoLC-MS/MS, cette approche est devenue complémentaire, voir même une alternative à l'approche par gel 2D. Son

avantage majeur reste sa sensibilité. En effet, la diminution du débit (200nl/min) ne modifiant pas la concentration de l'échantillon, le temps d'analyse peut être augmenté, permettant l'obtention d'informations de séquences de bonnes qualités, sur de faibles quantités de matériels (environ 50 femtomoles). De plus, cette approche est assez aisément automatisable. Cependant, elle ne peut pas être utilisée seule, sans autre technique de séparation en amont, si l'échantillon est trop complexe. En effet, un extrait protéique contient plusieurs milliers de protéines d'où des centaines de milliers de peptides trypsiques qui ne pourront pas être séparés efficacement sur nano-colonnes. Cette technique sera donc couplée en amont à une autre technique séparative, généralement le gel 2D. D'autres alternatives seront décrites plus loin, comme la chromatographie multidimensionnelle ou le couplage gel SDS/nano-LC, méthodes développées afin de s'affranchir des limites liées aux gels 2D.

1.3 Approches alternatives au gel 2D

1.3.1 Gel SDS PAGE couplé à la nano-LC

L'électrophorèse bi-dimensionnelle ne permettant pas l'accès à certaines protéines (protéines de pH extrêmes ou hydrophobes), une technique alternative consistant au couplage gel SDS (sodium dodecylsulfate) PAGE (PolyAcrylamide Gel Electrophoresis) suivi d'une séparation par nanoLC-MS/MS a été développée. Le gel SDS permet de séparer les protéines en fonction de leurs masses apparentes quelles que soient leurs caractéristiques physico-chimiques. Ainsi, il est possible de disposer des protéines basiques, hydrophobes auxquelles nous n'avions pas accès avec les gels 2D. Cependant, le pouvoir séparateur de cette technique reste faible, c'est pourquoi, elle est suivie par une séparation par nano-LC sur nano-colonne de phase inverse. Dans cette approche, la séparation est réalisée d'une part sur les protéines (gel SDS) et sur les peptides (nano-LC) alors que dans l'approche gel 2D, les 2 séparations (1^{ère} et 2^{ème} dimensions) sont réalisées sur les protéines [Rabilloud, 2002].

1.3.2 La chromatographie multidimensionnelle

Pour contourner les limitations du gel 2D, telles que la capacité limitée de chargement, la difficulté de faire migrer des protéines hydrophobes ou la détection de protéines minoritaires, une nouvelle technique alternative a été envisagée : la chromatographie multidimensionnelle (figure 20A) [Link et coll., 1999] ; [Yates, 2000], [Washburn et coll., 2000].



Figure 20 A :Schéma de la chromatographie multi-dimensionelle

Les peptides sont séparés sur une colonne échangeuse de cations (les peptides étant majoritairement chargés positivement à pH acide) suivie d'une séparation sur colonne de phase inverse. Deux variantes de cette approche ont été décrites : le couplage de 2 colonnes séparées physiquement [Davis et coll. 2003] et la combinaison des 2 types de phases dans une même colonne (MUDPIT) [Yates, 2000]. Les 2 variantes étant, dans leur principe, identiques, nous ne décrirons que le MUDPIT. Les peptides sont chargés sur une colonne capillaire bi-phasique remplie en série par une résine échangeuse de cations suivie par une résine de phase inverse. Dans un premier temps, les peptides sont élués de la résine échangeuse de cations par paliers avec un gradient en sels pour se déposer en tête de la seconde partie remplie de phase inverse. Ils sont ensuite élués une seconde fois en fonction de leur hydrophobicité croissante vers le spectromètre de masse. Cette approche a été réalisée avec succès sur la sous-unité 80S du ribosome et il a été montré que la 2D LC a permis l'identification de 4 fois plus de protéines par rapport à l'approche LC-MS/MS classique [Link, et coll., 1999].

Il est à noter que le produit de digestion d'un échantillon biologique complet représente des centaines de milliers de peptides. Si le pouvoir séparateur des colonnes reste minimum, seul le couplage avec la MS/MS permettra l'identification de nombreuses protéines puisque le séquençage d'un unique peptide permet d'identifier la présence d'une protéine. Ainsi cette approche a permis d'identifier 1500 protéines de levure d'abondance et de classes fonctionnelles diverses alors que l'approche baséee sur la séparation par gel 2D suivie d'une LC-MS/MS n'en avait identifié que quelques centaines [Rabilloud ; 2002].

En fait, la principale limite de cette approche n'est pas liée à la spectrométrie de masse mais à la chromatographie liquide, délicate à mettre en œuvre en raison des colonnes de faibles diamètres

internes. De plus, il faut noter que ce genre d'analyse peut durer plus de 10 heures selon le nombre de paliers d'élution, et que le monceau de données générées nécessite un appareillage informatique conséquent pour l'archivage (cluster de processeur).

Par ailleurs, si la majorité des couplages décrits allient colonne échangeuse de cations / colonne de phase inverse, il est tout à fait envisageable d'utiliser des colonnes d'affinités ou échangeuses d'anions en fonction des protéines étudiées. Il faudra cependant que les solvants d'élution soient adaptés aux colonnes et au spectromètre de masse.

Toutefois, ces méthodes ne permettent pas une quantification précise des protéines identifiées. Elles permettent de voir l'apparition ou la disparition d'une protéine, le différentiel reste possible mais n'est pas aussi précis car il reste dépendant de nombreux paramètres : propreté de l'appareil, suppression de signal, qualité chromatographique. Seules les approches par marquage permettront d'obtenir une quantification précise des protéines d'un mélange en utilisant l'approche nano-LC MS/MS.

1.3.3 Les approches par marquage

Si les approches décrites précédemment permettent de séparer de façon efficace les protéines, elles ne permettent pas toujours de quantifier les protéines présentes dans les mélanges analysés. En effet, les étapes de digestion enzymatique et d'ionisation en spectrométrie de masse ne sont pas quantitatives. Si les problèmes de quantification par spectrométrie de masse sont évidents, du point de vue biologique, il apparaît de plus en plus que les cellules modifient leur activité physiologique non pas en synthétisant de nouvelles protéines mais plutôt en modulant l'expression de leurs gènes, c'est à dire en augmentant ou en diminuant légèrement la synthèse de certaines protéines. Même si la quantification peut être aisément réalisée par les gels 2D, il reste la limite liée à cette technique qui ne permet pas d'analyser toutes les protéines (protéines hydrophobes, de pI extrêmes...) [Rabilloud, 2001]. Or les méthodes de marquage décrites ici permettent d'éviter les limites rencontrées avec les gels 2D. En effet, si les échantillons sont immédiatement mélangés après le marquage, la digestion est uniforme pour les 2 échantillons. De même, les peptides marqués sont censés co-éluer en HPLC, l'efficacité d'ionisation sera ainsi la même.

Ainsi, différentes approches de marquages d'acides aminés ont été développées. Nous citerons 2 approches : la méthode ICAT (Isotope Coded Affinity Tag) et la méthode ICNAT (Isotope Coded Non Affinity Tag).

L'approche ICAT, d écrite dans la figure 20 permet de simplifier le mélange protéique, en sélectionnant uniquement les peptides à cystéines marquées par chromatographie d'affinité, de les quantifier [Gygi et coll., 1999b] ; [Gygi et coll., 2000] ; [Griffin et coll., 2001] et d'identifier les protéines à partir des séquences peptidiques.





Un des avantages évident de cette méthode est qu'elle s'affranchit de la complexité des extraits puisque ne sont conservés que les peptides comportant au moins une cystéine. De là, il devient possible d'identifier les protéines présentes en faible quantité. Cependant, cette approche ne semble pas compatible avec les échantillons complexes, les peptides ne pourront pas être correctement séparés avant leur entrée dans le spectromètre de masse (pouvoir séparatif des nanocolonnes). De plus, de nombreuses pertes de peptides sont observées (adsorption des peptides sur les différentes colonnes...). De ce fait, l'acide aminé marqué étant rare pour simplifier le mélange, la perte d'un peptide peut signifier la perte de la protéine correspondante. De plus, certaines protéines ne possèdent pas de cystéines et échappent donc totalement à l'analyse. Enfin, lors de la séparation chromatographique, les paires de peptides lourds et légers ne sont pas co-élués, ils ont des temps de rétention différents, empêchant toute analyse quantitative.

Afin d'identifier l'ensemble des protéines présentes dans le mélange tout en les quantifiant, une alternative à cette approche à été développée : l'approche ICNA T avec comme marqueur le N Ethyl Iodo Acetamide (NEAIM) [Guillier ; 2003]. Cette approche permet aussi de quantifier les peptides à cystéines marquées avec du NEAIM lourd et léger mais l'étape de chromatographie d'affinité n'est pas réalisée. Ainsi les peptides non marqués permettent d'identifier l'ensemble des protéines de l'extrait. Pour simplifier l'étude, une première séparation sur gel SDS est réalisée, avec prélèvement systématique de tous les spots. De plus contrairement au marquage ICAT, les paires de peptides sont co-élués, ce qui permet une bonne quantification.

D'autres marquages ont été développés pour simplifier les mélanges complexes et quantifier, par exemple l'utilisation d'acides aminés marqués dans les milieux de culture pour mesurer la néosynthèse d'une protéine dans des conditions de culture particulières [Wagner 2003] ou encore pour distinguer plus aisément les protéines issues d'un organisme en cas de co-cultures. Cette approche a été utilisée lors de ce travail de thèse pour différencier les protéines parasitaires de *Plasmodium falciparum* de celles du globule rouge.

Ces nouvelles stratégies sont utilisées pour des études systématiques, utilisant la nanoLC-MS-MS ; elles ne peuvent fonctionner qu'avec un système de sélection et de fragmentation des ions totalement automatisé. Le spectromètre de masse devra analyser les masses des peptides élués puis les sélectionner une à une selon des critères bien définis afin de les fragmenter avec l'énergie de collision appropriée.

2 Identification des protéines

2.1 Présentation de l'approche protéomique



Figure 21 : Schéma de les différentes approches protéomiques

L'approche protéomique présentée (figure 21) est l'approche classique : gel 2D / Spectrométrie de masse. Après la séparation des protéines par électrophorèse bi-dimensionnelle, les spots sont excisés, digérés (classiquement avec la trypsine mais d'autres endoprotéases peuvent être utilisées, AspN, Chymotrypsine). Ces étapes peuvent être entièrement automatisées à l'aide de robot de digestion. Les peptides trypsiques sont ensuite extraits et analysés par spectrométrie de masse. Ensuite, plusieurs alternatives s'offrent à nous. Si l'organisme étudié est référencé dans les banques de données protéiques, l'approche MALDI (Peptide Mass Fingerprint) sera envisagée ; dans le cas d'un organisme non référencé ou ayant subi des mutations, il faudra se tourner vers l'approche MS/MS permettant d'obtenir des informations de séquences sur les peptides analysés.

2.2 L'empreinte peptidique massique (Fingerprint) : approche MALDI-MS

Cette stratégie fut décrite, dès 1993, par plusieurs équipes [Henzel, et coll., 1993] ; [James et coll., 1993] ; [Mann et coll., 1993] ; [Yates et coll., 1993] comme alternative au séquençage peptidique d'Edman.

L'approche MALDI-MS, illustrée figure 21, permet l'identification rapide des protéines, sans séquençage, directement à partir de leurs cartes peptidiques massiques et par confrontation aux cartes théoriques générées *in silico*.

Le crédit accordé à l'identification dépend du nombre de peptides reconnus, du pourcentage de recouvrement de séquence de la protéine, ainsi que de la précision de mesure. Le spectromètre de masse MALDI-TOF a été choisi, car il est l'instrument le plus adapté à ce type d'analyse :

- > Pour sa précision de mesure [Jensen et coll., 1996].
- > Pour sa tolérance aux sels et aux tampons [Vorm et coll., 1994].
- > Pour sa sensibilité, détection de quantités femtomolaires, voire attomolaires.
- Pour sa possibilité d'analyse à haut-débit grâce à l'automatisation [Shevchenko, et coll. 1996]
- Pour ses spectres directement exploitables. Seuls les ions monochargés sont détectés, donc aucune étape de déconvolution ne sera nécessaire contrairement à l'ESI

Néanmoins il peut arriver que l'identification par cette approche échoue pour plusieurs raisons :

- Pourcentage de recouvrement insuffisant
- Précision de mesure médiocre (supérieure à 50 ppm)
- Suppression de signal en fonction de la composition en acides aminés des peptides [Krause et coll., 1999] ou de la présence de contaminants

- Présence massive de modifications post-traductionnelles ou mutations de la protéine.
- Peptides générés par la digestion trypsique (inférieurs à 15 ou 20 kDa) confondus avec les pics de matrice
- Protéine absente des banques de données

Si la stratégie de l'empreinte peptidique massique ne permet pas d'identifier les protéines, le recours à l'approche MS/MS sera envisagée afin d'obtenir des informations de séquences.

2.3 Les informations de séquences : approche NanoLC-MS/MS

2.3.1 Les règles de fragmentation peptidique

La spectrométrie de masse en tandem est utilisée pour obtenir des informations sur la structure des composés analysés. Les peptides se fragmentent principalement au niveau de la liaison peptidique (ions y et b). Mais d'autres fragments ont également été observés de part et d'autre de cette liaison (ions a, c et x, z). Une première nomenclature a été introduite par Roepstorff et Fohlmann [Roepstorff et Fohlmann, 1984]. Elle a ensuite été complétée et simplifiée par Biemann quelques années plus tard et fait toujours référence dans ce domaine [Biemann et coll., 1990] (figure 22).

Les 2 types d'ions issus de la fragmentation à basse énergie de la liaison peptidique sont les suivants :

Ions a, b, c pour lesquels la charge positive est portée par l'acide aminé N terminal

Ions **b** issus de la fragmentation de la liaison peptidique (CO---NH) Ions **a** issus de la coupure en amont de la liaison peptidique (CH---CO) Ions **c** issus de la coupure en aval de la liaison peptidique (NH---CH)

Ions x, y, z pour lesquels la charge positive est portée par l'acide aminé C terminal

Ions y issus de la fragmentation de la liaison peptidique (CO---NH)

Ions x issus de la coupure en amont de la liaison peptidique (CH---CO)

Ions z issus de la coupure en aval de la liaison peptidique (NH---CH)

D'autres fragments peuvent être observés dans le cas des fragmentations haute énergie :

Ions d pour lesquels la charge positive est portée par l'acide aminé N terminal

Coupure de la chaîne latérale après le carbone β

Ions w et v pour lesquels la charge positive est portée par l'acide aminé C terminal Ions v issus de la coupure de la chaîne latérale après le carbone α Ions w issus de la coupure de la chaîne latérale après le carbone β



Figure 22 : Illustration de la nomenclature des fragmentations d'après Biemann sur un tripeptide, les chaînes latérales sont notées R1, R2 et R3.

D'autres nomenclatures basées sur le même principe ont été développées et proposées pour désigner les fragments observés lors de l'analyse par spectrométrie de masse des oligonucléotides [McLuckey et coll., 1992] et des oligosaccarides [Domon et coll., 1988].

2.3.2 Identification par de courtes séquences (peptide sequence tag)

Lorsque l'approche MALDI-MS ne permet pas d'identifier les protéines, il faut envisager l'approche MS/MS (figure 21) soit seule (nanospray offline), soit couplée à une nanoLC pour obtenir des informations de séquences. Jusqu'à présent, une alternative plus classique pour avoir des informations de séquence était le séquençage par dégradation d'Edman, mais elle nécessitait la séparation des peptides par HPLC, une quantité importante de peptide, et pouvait néanmoins s'avérer inutile si le peptide était bloqué à son extrémité N-terminale. De plus, ce genre d'analyse reste très long et onéreux. Par ailleurs la dégradation d'Edman ne permet pas d'obtenir d'informations sur la nature d'une éventuelle modification post-traductionnelle, alors que la spectrométrie de masse apporte au moins une information sur la masse de la modification.

Ainsi, par MS/MS, il sera possible de séquencer les peptides d'intérêt. La majorité des fragments obtenus sont de types b et y. La différence de masse entre deux fragments consécutifs de même type donne une masse correspondant à l'un des 20 acides aminés. Ainsi l'interprétation des spectres MS/MS est relativement aisée grâce aux outils informatiques disponibles. En effet, de la même manière qu'il existe des logiciels de recherche pour comparer les listes de masses peptidiques théoriques et expérimentales avec la stratégie MALDI-MS, il existe aussi un outil informatique qui

compare la liste des masses des fragments peptidiques obtenus avec une liste de masses théoriques [Yates et coll., 1995], ou encore des moteurs de recherche capables de fonctionner avec une information de séquence partielle, appelée « peptide sequence tag » [Wilm et coll., 1996]. Il suffit donc d'une petite séquence peptidique pour pouvoir identifier une protéine par une recherche dans les banques de données [Mann et coll ; 1994]. Il faut néanmoins que la séquence du peptide soit référencée sans erreurs dans les banques de données, puisque toutes les recherches sont réalisées à partir de listes de masses. Si aucune séquence n'est présente dans les banques ou si celle-ci est mutée, les masses théoriques et mesurées seront différentes, il faudra donc étudier les spectres un à un, afin d'en déduire les séquences, c'est à dire en réalisant un séquençage *De Novo*.

Comme nous l'avons mentionné précédemment, cette stratégie peut être utilisée en nanospray ou en couplage nanoLC-MS/MS. Cependant, si la stratégie nanoESI-MS/MS est très performante pour l'identification de protéines, il n'en reste pas moins qu'elle est très exigeante en terme de temps et très difficilement automatisable (chargement des capillaire, dessalage...). Son automatisation sera uniquement possible au travers du couplage nanoLC MS/MS. Par cette approche, il sera possible de fragmenter les peptides trypsiques de façon automatique et d'obtenir des spectres MS/MS de bonnes qualités qui pourront être soumis directement à différentes recherches dans les banques de données ou interprétés manuellement.

2.3.3 Le séquençage De Novo

Le séquençage De Novo consiste à étudier les spectres MS/MS un à un afin de déterminer à partir de chacun d'entre eux une séquence correspondant aux peptides trypsiques isolés et fragmentés. Cette interprétation manuelle des spectres MS/MS est longue et fastidieuse. Cependant de nouveaux logiciels d'aide au séquençage ont été développés par les constructeurs (Biotools chez Bruker, ProteinLynx chez Waters, Peaks studio Inc.). A partir de chaque spectre, une séquence est déduite et est soumise à des recherches informatiques dans des banques de données en s'affranchissant de toutes masses. Ainsi, si la séquence est légèrement modifiée, l'identification de la protéine pourra néanmoins être possible.

3 Recherche dans les banques de données

3.1 Quelques moteurs de recherche

L'accumulation des données spectrales MALDI-MS ou LC-MS/MS qu'elles soient issues de la fragmentation de peptides ou non ne permet pas, seule, d'identifier les protéines. Il s'agit ensuite de traiter ces données afin d'identifier les protéines d'où les peptides trypsiques sont issus. Pour cela les spectres doivent subir un certain nombre de traitements (déconvolution, lissage, calcul de centroïde, création de carte massique appelée peak-list...) avant d'être utilisés pour entamer des recherches dans les banques de données en accès libre *via* Internet, par exemple au travers de serveur comme ExPASy de Genève (http://www.expasy.ch/ch2d/2d-index.html) ou en utilisant les serveurs locaux si les laboratoires en sont équipés. Plusieurs moteurs de recherche, Mascot Protein Identification System géré par Matrix Science (http://www.matrixscience.com/cgi/index.pl?page=../home.html), Global server de la société Micromass ou Protein Prospector (http://prospector.ucsf.edu) étaient à notre disposition. La comparaison des cartes massiques expérimentales avec les cartes massiques théoriques des banques de données s'effectue à l'aide de logiciels appropriés gérés par les différents moteurs de recherche tels que MS-fit (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm), Peptide Mass (http://www.expasy.ch/tools/-peptidemass.html), ou encore Peptide search (http://www.mann.emblheidelberg.de/services-/peptidesearchintro.html). Aussi plusieurs équipes développent actuellement de nouveaux moteurs de recherche plus performants fonctionnant sur des modes de recherches variés. Ainsi, l'équipe de John Yates a développé la méthode de «l'empreinte de la fragmentation peptidique» [MacCoss, 2002], par analogie avec la méthode de l'empreinte peptidique massique [Eng et coll., 1994]. Ce programme appelé SEQUEST compare des spectres de fragmentations théoriques avec le spectre expérimental et peut cribler ainsi les banques de données protéiques. Par ailleurs le groupe de Mathias Mann a développé PeptideSearch. Cet algorithme requiert la masse du peptide intact, une séquence de 3-4 acides aminés et les masses de part et d'autre de la séquence [Mann et Wilm, 1994] (voir le site de l'EMBL : http://www.mann.emblheildelbreg.de-/Services/peptidesearch-/peptidesearchintro.html).

Pour le futur, les avancées dans le domaine de la protéomique ne pourront se faire sans le développement en parallèle d'outils et de supports informatiques, de mémoires vives importantes, de clusters de PC et de multiprocesseurs, capables d'extraire toute l'information délivrée par les analyses nanoLC-MS-MS.

3.2 Recherche par alignement de séquences

Si la protéine n'est pas référencée dans les banques de données, la séquence peptidique de plus de 5 à 6 acides aminés peut permettre d'identifier la protéine par homologie de séquence grâce aux logiciels de type BLAST ou Fasta3 (facilement disponible sur Expasy). Ces logiciels réalisent des alignements de séquences qui pourront permettre l'identification de la protéine, même si celle-ci est légèrement mutée.

4 Quelques logiciels d'archivage

Si les informations obtenues en protéomique nous permettent de caractériser de plus en plus finement les protéines étudiées, il n'en reste pas moins que la taille des données à traiter est énorme.

En effet, une simple analyse LC-MS/MS génère des fichiers de 80 à 100Mo. Ces fichiers comprennent à la fois les données brutes, mais aussi les données retraitées (déconvolution, élaboration de la peak list...) et les résultats des recherches dans les banques de données. Aux vues de la taille de ces fichiers, les informaticiens ont développés de nouveaux logiciels permettant l'archivage des données. Différents constructeurs se sont intéressés avec plus ou moins de succès au développement de ces nouveaux logiciels (Bruker, Applied Biosystem, Agilent...)

Proteinscape, en fin de développement par la société Bruker, permet d'ores et déjà d'archiver les données MS et très prochainement les données MS/MS.

Il est organisé suivant 6 niveaux :

- Description du projet général
- Description de l'échantillon
- Photo du gel
- Description du spot
- Données MS et liste de masses associées
- Identification des protéines et résultat de la recherche Mascot

De tels logiciels présentent l'avantage majeur de donner la possibilité à l'utilisateur retraiter et de lancer facilement de nouvelles recherches sur d'anciennes analyses. En effet, les banques de données évoluent quotidiennement. Ainsi des analyses restées sans résultat pourront aboutir ultérieurement à une identification. De plus, ces logiciels facilitent la comparaison entre différentes analyses, ce qui donne la possibilités de mettre en relation les identifications de protéines réalisées sur des sujets différents qui n'ont à priori aucun rapport.

Bibliographie :

Anderson L. and Seilhamer J.

A comparison of selected mRNA and protein abundances in human liver, *Electrophoresis*, **1997**, 18, 533-7.

Beavis R. C. and Chait B. T.

Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins, *Rapid Commun Mass Spectrom*, **1989**, 3, 432-5.

Beavis R. C. and Chait B. T.

Velocity distributions of intact hight mass polypeptide molecule ions produced by matrix assisted laser desorption., *Chem. Phys. Lett.*, **1991**, 181, 479-484.

Biemann K.

Appendix 5. Nomenclature for peptide fragment ions (positive ions), *Methods Enzymol*, 1990, 193, 886-7.

Bitsch F.,

Analyses structurales de protéines et de complexes de coordination par spectrométrie de masse, Thèse de l'Université Louis Pasteur de Strasbourg **1991**.

Bjellqvist B., Ek K., Righetti P. G., Gianazza E., Gorg A., Westermeier R. and Postel W.

Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications, J Biochem Biophys Methods, **1982**, 6, 317-39.

Bjellqvist B., Pasquali C., Ravier F., Sanchez J. C. and Hochstrasser D.

A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale, *Electrophoresis*, **1993b**, 14, 1357-65.

Bjellqvist B., Sanchez J. C., Pasquali C., Ravier F., Paquet N., Frutiger S., Hughes G. J. and Hochstrasser D.

Micropreparative two-dimensional electrophoresis allowing the separation of samples containing milligram amounts of proteins, *Electrophoresis*, **1993a**, 14, 1375-8.

Blades A. T., Ikonomou M. G. and Kebarle P.

Mechanism of electrospray mass spectrometry. Electrospray as an electrolysis cell, *Anal Chem*, **1991**, 63, 2109 - 2114.

Brown, R. S., Lennon, J.

Sequence-Specific Fragmentation of Matrix-Assisted Laser-Desorbed Protein/Peptide Ions J.Anal. Chem. 1995a. 67, 3990-3999.

Brown, R. S., Lennon, J.

Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer, J. Anal. Chem. **1995b**. 67, 1998-2003.

Campana J. E.

Elementary theory of the quadrupole mas filter., *Int. J. Mass Spectrom. Ion Processes*, **1980**, 33, 101-117.

Carte N.

La trappe ionique et l'ionisation electrospray:un nouveau potentiel pour la caractérisation de biomolécules, *Thèse de l'Université Louis Pasteur de Strasbourg.*, **2001**,

Cavusoglu N.

Apport de la spectrométrie de masse à l'étude de modifications post-traductionnelles et à la protéomique., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **2002**,

Colby S. M., King T. B. and Reilly J. P.

Improving the resolution of matrix assisted laser desorption/ionization mass spectrometry by exploiting the correlation between ion position and velocity., *Rapid Commun. Mass Spectrom.*, **1994**, 8, 865-868.

Cole R. B.

Some tenets pertaining to electrospray ionization mass spectrometry, *J Mass Spectrom*, **2000**, 35, 763-72.

Cooks, R.G., .Glish, G.L., McLuckey, S.A., Kaiser, R.E.,

on trap mass spectrometry. Chemical & Engineering News, 1991, Mars 25, 26-41.

Cotter R. J.

Time-of-flight mass spectrometry: an increasing role in the life sciences, *Biomed Environ Mass Spectrom*, **1989**, 18, 513-32.

Covey T., Shushan B. I. and Bonner R.

Practical advantages of Ionspray over Electrospray, Sciex, 1990, Note 15789.

Dehmelt, H.G.

Radiofrequency spectroscopy of stored ions. I :Storage Adv. At. Mol. Phys., 1967, 3, 53-72.

Dole M., Mack L. L., Himes R. L., Mobley R. C., Ferguson L. D. and Alicce M.B.

Molecular beams of macroions., J. Chem. Phys., 1968, 49, 2240-2249.

Domon B. and Costello C. E.

Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry, *Biochemistry*, **1988**, 27, 1534-43.

Emmet M. R. and Caprioli R.M.

Micro-electrospray : ultra-high-sensitivity- analysis of peptides and proteins., J. Am. Soc. Mass Spectrom., 1994, 5, 605-613.

Eng J. K., McCormack A. L. and Yates J. R.

An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database., J. Am. Soc. Mass Spectrom., **1994**, 5, 976-989.

Fligge T. A., Kast J., Bruns K. and Przybylski M.

Direct monitoring of protein-chemical reactions utilising nanoelectrospray mass spectrometry, J. Am. Soc. Mass Spectrom., **1999**, 10, 112-8.

Gallagher R. T., Chapman J. R. and Mann M.

Design and performance of an electrospray ionization source for a doubly-focusing magnetic sector mass spectrometer., *Rapid Commun Mass Spectrom*, **1990**, 4, 369-372.

Giometti C. S., Gemmell M. A., Tollaksen S. L. and Taylor J.

Quantitation of human leukocyte proteins after silver staining: a study with two-dimensional electrophoresis, *Electrophoresis*, **1991**, 12, 536-43.

Glückmann, M. and Karas, M.

The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization, J. Mass Spectrom. 1999. 34, 467.

Gorg A., Postel W., Domscheit A. and Gunther S.

Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects, *Electrophoresis*, **1988**, 9, 681-92.

Griffin T. J., Gygi S. P., Rist B. and Aebersold R.

Quantitative proteomic analysis using a MALDI quadrupole time of flight mass spectrometer, *Anal. Biochem.*, **2001**, 73, 978-86.

Gross J., Leisner A., Hillenkamp F., Hahner S., Karas M., Schäfer J., Lützenkirchen F. and Nordhoff E.

Investigations of the metastable decay of DNA under ultraviolet matrix-assisted laser desorption/ionization conditions with post-source-decay analysis and hydrogen/deuterium exchange, J Am Soc. Mass Spectrom. **1998**. 9, 866.

Gygi S. P., Rist B. and Aebersold R.

Measuring gene expression by quantitative proteome analysis, *Curr. Opin. in Biotech.*, **2000**, 11, 396-401.

Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H. and Aebersold R.

Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.*, **1999b**, 17, 994-9.

Gygi S. P., Rochon Y., Franza B. R. and Aebersold R.

Correlation between protein and mRNA abundance in yeast, Mol Cell Biol, 1999a, 19, 1720-30.

Hanash S. M., Strahler J. R., Neel J. V., Hailat N., Melhem R., Keim D., Zhu X. X., Wagner D., Gage D. A. and Watson J. T.

Highly resolving two-dimensional gels for protein sequencing, *Proc Natl Acad Sci U S A*, **1991**, 88, 5709-13.

Harrison, A. G.

The gas-phase basicities and proton affinities of amino acids and peptides, *Mass Spectrom. Rev.* 1997. 16, 201.

Henry K. D., Williams E. R., Wang B. H., McLafferty F. W., Shabanowitz J. and Hunt D. F.

Fourier-transform mass spectrometry of large molecules by electrospray ionization, *Proc Natl Acad Sci USA*, **1989**, 86, 9075-8.

Henzel W. J., Billeci T. M., Stults J. T., Wong S. C., Grimley C. and Watanabe C.

Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases, *Proc Natl Acad Sci U S A*, **1993**, 90, 5011-5.

Huddleston M. J., Bean M. F. and Carr S. A.

Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests, *Anal Chem*, **1993**, 65, 877-84.

Huddleston M. J., Anan R. S. and Bean M. F.

Selective detection of phosphopeptides in complex mixtures by electrospray liquid chromatography/mass spectrometry., J. Am. Soc. Mass Spectrom., **1993b**, 4, 710-717.

Ikonomou M., Blades A. T. and Kebarle P.

Investigations of the Electrospray Interface for Liquid ChromatographyIMass Spectrometry, *Anal Chem*, **1990**, 62, 957-967.

Ikonomou M. G., Blades A. T. and Kebarle P.

Electrospray-Ion spray: A comparaison of mechanisms and performance, *Anal. Chem.*, **1991**, 63, 1989-98.

Iribarne J. V. and Thomson B. A.

On the evaporation of small ions from charged droplets., J. Chem. Phys., 1976, 64, 2287-2294.

James P., Quadroni M., Carafoli E. and Gonnet G.

Protein identification by mass profile fingerprinting, *Biochem Biophys Res Commun*, **1993**, 195, 58-64.

Jensen O. N., Vorm O. and Mann M.

Sequence patterns produced by incomplete enzymatic digestion or one-step Edman degradation of peptide mixtures as probes for protein database searches, *Electrophoresis*, **1996**, 17, 938-44.

Jorgensen, T.J.D., Heck, A.J.R., Roepstorff, P.

Direct determination of solution binding constants for noncovalent complexes between bacterial cell wall peptide analogues and vancomycin group antibiotics by electrospray ionization mass spectrometry., *Anal. Chem.*, **1998**, 70, 4427-4432.

Jespersen S., Niessen W. M., Tjaden U. R. and van der Greef J.

Basic matrices in the analysis of non-covalent complexes by matrix-assisted laser desorption/ionization mass spectrometry, *J Mass Spectrom*, **1998**, 33, 1088-93.

Kaiser, R.E., Cooks, R.G., Stafford, G.C., Syka, J.E.P., Hemberger, P.H.

Operation of a quadrupole ion trap mass spectrometer to achieve high mass/charge ratios. Int. J. Mass Spectrom. Ion Proc., **1991**, 106, 79-115.

Karas M., Bachmann D., Bahr U. and Hillenkamp F.

Matrix-assisted ultraviolet laser desorption of non-volatile compounds., Int. J. Mass Spectrom. Ion Processes, 1987, 78, 53-68.

Karas, M., Bahr, U., Ingendoh, A., Hillenkamp, F.

Laser desorption ionization mass spectrometry of proteins of mass 100000 to 250000 Daltons. Angew. *Chem. Int. Ed. Engl.*, **1989**, 28, 760-761.

Kebarle P. and Tang L.

From ions in solution to ions in the gas phase, The mechanism of electrospray mass spectrometry, *Anal. Chem.*, **1993**, 65, 972-86.

Kebarle P.

A brief overview of the present status of the mechanism involved in electrospray mass spectrometry., *J. Mass Spectrom.*, **2000**, 35, 804-817.

Kenrick K. G. and Margolis J.

Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique, *Anal Biochem*, **1970**, 33, 204-7.

Krause E., Wenschuh H. and Jungblut P. R.

The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins, *Anal Chem*, **1999**, 71, 4160-5.

Knochenmuss R. and Zenobi R.

MALDI ionization: the role of in-plume processes, Chem Rev, 2003, 103, 441-52.

Leize E.

Caractèrisation d'édifices supramoléculaires par spectrométrie de masse avec ionisaton électropsray., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **1994**,

Lemaitre-Guillier C., Sanvito R., Guillier F., Varrier F., Rabilloud T., Leize-Wagner E. and Van Dorsselaer A.

N-ethyl iodoacetamide alkylating agent use in a non affinity tag approach for proteomic quantification in mass spectrometry (submitted)

Liao P. C., Leykam J., Andrews P. C., Gage D. A. and Allison J.

An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry, *Anal Biochem*, **1994**, 219, 9-20.

Link A. J., Eng J., Schieltz D. M., Carmack E., Mize G. J., Morris D. R., Garvik B. M. and Yates J. R.

Direct analysis of protein complexes using mass spectrometry, *Nature Biotechnology*, **1999**, 17, 676-82.

Lottspeich F.

Proteome analysis: A pathway to the functionnal analysis of protein, *Angew. Chem. Int. Ed.*, **1999**, 38, 2476-92.

MacCoss M. J., Wu C. C. and Yates J. R. 3rd

Probability-based validation of protein identifications using a modified SEQUEST algorithm, *Anal Chem*, **2002**, 74, 5593-9.

Mamyrin B. A., Karataev V. I., Shmikk D.V. and Zagulin V. A.

Mass reflectron. New nonmagnetic time-of-flight high-resolution mass spectrometer., *Zh. Eksp. Theor. Fiz.*, **1973**, 64, 82-89.

Mann M. and Wilm M.

Electrospray mass spectrometry for protein characterization, Trends Biochem Sci, 1995, 20, 219-24.

March, R.E., Londry, F.A.

Fundamentals of ion trap mass spectrometry, 1995, March, R.E., Todd, J.F.J. Eds, CRC Press, Inc, Boca Raton, 1, 25.

McLuckey S. A., Van Berkel G. J. and Glish G. L.

Tandem Mass Spectrometry of Small, Multiply charged Oligonucleotides., J. Am. Soc. Mass Spectrom., 1992, 3, 60-70.

Medzihradszky K. F., Campbell J. M., Baldwin M. A., Falick A. M., Juhasz P., Vestal M. L. and Burlingame A. L.

The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer., *Anal. Chem.*, **2000**, 72, 552-58.

Meng C.K., Mann M. and Fenn J.B.

Of protons or proteins., Z. Phys. D., 1988, 10, 361-368.

Moniatte M.,

Caractérisation structurale de biomolécules par spectrométrie de masse à désorption/ionisation laser assistée par matrice, Thèse de l'Université Louis Pasteur de Strasbourg **1997**.

Moniatte M., Lesieur C., Vécsey-Semjen B., Buckley J. T., Pattus F., van der Goot F. G. and van Dorsselaer A.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in the subunit stoichiometry study of high-mass non-covalent complexes., *Int. J. Mass Spectrom. Ion Processes*, **1997**, 169, 179-199.

Morris H. R., Paxton T., Panico M., McDowell R. and Dell A.

A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing, *J Protein Chem*, **1997**, 16, 469-79.

Nierengarten H.

Nouvelles approches pour l'étude de supramolécules synthétiques par spectrométrie de masse., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **2001**,

Oda Y., Huang K., Cross F. R., Cowburn D. and Chait B. T.

Accurate quantitation of protein expression and site-specific phosphorylation, *Proc Natl Acad Sci U S A*, **1999**, 96, 6591-6.

O'Farrrell

High resolution two-dimentional electrophoresis of proteins., J. Biol. Chem., 1975, 250, 4007-4021.

Posthumus M. A., Kistemaker P. G., Meuzelaar H. L. and Ten Noever de Brauw M. C.

Laser desorption-mass spectrometry of polar nonvolatile bio-organic molecules., *Anal Chem*, **1978**, 50, 985-991.

Potier N.

Caractérisaton de complexes non-covalents par spectrométrie de masse avec ionisation électrospray., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **1996**,

Rabilloud T.

Two-dimensional electrophoresis of basic proteins with equilibrium isoelectric focusing in carrier ampholyte-pH gradients, *Electrophoresis*, **1994**, 15, 278-82.

Rabilloud T.

Solubilization of proteins for electrophoretic analyses, *Electrophoresis*, 1996, 17, 813-29.

Rabilloud T., Adessi C., Giraudel A. and Lunardi J.

Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients, *Electrophoresis*, **1997**, 18, 307-16.

Rabilloud T., Strub J. M., Luche S., van Dorsselaer A. and Lunardi J.

A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels, *Proteomics*, **2001**, 1, 699-704.

Roepstorff P. and Fohlman J.

Proposal for a common nomenclature for sequence ions in mass spectra of peptides, *Biomed Mass Spectrom*, **1984**, 11, 601.

Rogniaux H.

Etude d'interactions faibles entra- et inter-moléculaires de biomolécules par spectrométrie de masse., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **2000**,

Rudd P. M. and Dwek R. A.

Glycosylation: heterogeneity and the 3D structure of proteins, *Crit Rev Biochem Mol Biol*, **1997**, 32, 1-100.

Sanglier S.

La spectrométrie de masse : un nouvel outil pour l'étude des interactions faibles en biologie., *Thèse de l'Université Louis Pasteur de Strasbourg.*, **2002**,

Sanchez J. C., Rouge V., Pisteur M., Ravier F., Tonella L., Moosmayer M., Wilkins M. R. and Hochstrasser D. F.

Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients, *Electrophoresis*, **1997**, 18, 324-7.

Shevchenko A., Wilm M., Vorm O. and Mann M.

Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal Chem*, **1996**, 68, 850-8.

Shevchenko A., Loboda A., Shevchenko A., Ens W. and Standing K. G.

MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research., *Anal. Chem.*, **2000**, 72, 2132-41.

Simmons, T.A., Limbach, P.A.,

Influence of Co-matrix Proton Affinity on Oligonucleotide Ion Stability in Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry, *J Am Soc. Mass Spectrom.* **1998**. 9, 668.

Spengler B. and Cotter R. J.

Ultraviolet laser desorption/ionization mass spectrometry of proteins above 100,000 daltons by pulsed ion extraction time-of-flight analysis, *Anal Chem*, **1990**, 62, 793-6.

Suckau D., Resemann A., Schuerenberg M., Hufnagel P., Franzen J. and Holle A.

A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics, *Anal Bioanal Chem*, **2003**, 376, 952-65.

Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y. and Yoshida T.

Protein and polymer analysis up to m/z 100.000 by laser ionization time-of-flight mass spectrometry., *Rapid Commun. Mass Spectrom.*, **1988**, 2, 151-152.

Vachet, R.W., Glish, G.L.

Effects of heavy gases on the tandem mass spectra of peptide ions in the quadrupole ion trap. J. Am. Soc. Mass Spectrom., 1996, 7, 1194-1202.

Vachet, R.W., Glish, G.L.

Boundary activated dissociation of peptide ions in a quadrupole ion trap. Anal. Chem., 1998, 70, 340-346.

Van Berkel G. J., Glish G. L. and McLuckey S. A.

Electrospray ionization combined with ion trap mass spectrometry., Anal Chem, 1990, 62, 1284-1295.

Verentchikov A. N., Ens W. and Standing K. G.

Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction, *Anal Chem*, **1994**, 66, 126-33.

Vestal M. L., Juhasz P. and Martin S. A.

Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry., *Rapid Commun. Mass Spectrom.*, **1995**, 9, 1044-1050.

Vogl T., Roth J., Sorg C., Hillenkamp F. and Strupat K.

Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 detected by ultraviolet matrixassisted laser desorption/ionization mass spectrometry, *J Am Soc Mass Spectrom*, **1999**, 10, 1124-30.

Whittal R. M. and Li L.

High-Resolution Matrix-Assisted Laser Desorption/Ionization in a Linear Time-of-Flight Mass Spectrometer, *Anal Chem*, **1995**, 67, 1950-1954.

Wiley W. C. and McLaren I. H.

Time-of-flight mass spectrometry with improved resolution., Rev. Sci. Instrum., 1955, 26, 1150-1157.

Wiley W.C. and McLaren I.H.

Time-of-ight mass spectrometer with improved resolution, Rev. Sci. Instrum., 1955, 26, 1150-1157.

Wilm M. and Mann M.

Electrospray and taylor-cone theory, Dole's beam of macromolecules at last., *Int. J. Mass Spectrom. Ion Processes*, **1994**, 136, 167-180.

Wilm M. and Mann M.

Analytical properties of the nanoelectrospray ion source, Anal Chem, 1996, 68, 1-8.

Yamashita M. and Fenn J.

Electrospray ion source. Another variation on the Free-jet theme., J. Phys. Chem., 1984a, 88, 4451-4460.

Yamashita M. and Fenn J.

Negative Ion Production with the electrospray Ion Source, J. Phys. Chem., 1984b, 88, 4671-4675.

Yates J. R. 3rd

Mass spectrometry. From genomics to proteomics, Trends Genet, 2000, 16, 5-8.

Yates J. R. 3rd, Eng J. K., McCormack A. L. and Schieltz D.

Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database, *Anal Chem*, **1995**, 67, 1426-36.

Yates J. R. 3rd, Speicher S., Griffin P. R. and Hunkapiller T.

Peptide mass maps: a highly informative approach to protein identification, *Anal Biochem*, **1993**, 214, 397-408.

Yergey A. L., Coorssen J. R., Backlund P. S., Blank P. S., Humphrey G. A., Zimmerberg J., Campbell J. M. and Vestal M. L.

De novo sequencing of peptides using MALDI/TOF-TOF, J Am Soc Mass Spectrom, 2002, 13, 784-91.

Zigilei, L. V., and Garrison, B. J.

Velocity distributions of analyte molecules in matrix-assisted laser desorption from computer simulations, *Rapid Comm Mass Spectrom.* **1998**, 12, 1273.

Zhang W., Krutchinsky A. N. and Chait B. T.

"De novo" peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: a preliminary study, *J Am Soc Mass Spectrom*, **2003**, 14, 1012-21.

Zhou J., Ens W., Standing K. G. and Verentchikov A.

Kinetic energy measurements of molecular ions ejected into an electric field by matrix-assisted laser desorption, *Rapid Commun Mass Spectrom*, **1992**, 6, 671-8.

Partie II :

Optimisation des étapes de l'approche protéomique pour la caractérisation des protéines

Chapitre 1 : Amélioration de la compatibilité des colorants des gels 2D avec l'analyse par spectrométrie de masse

Chapitre 2 : Optimisation des étapes d'extraction et de digestion des protéines *in-gel*

Chapitre 3 : Evaluation des approches MS/MS en analyse protéomique

Chapitre 1:

Amélioration de la compatibilité des colorants des gels 2D avec l'analyse par spectrométrie de masse

Le but de cette 1^{ère} étude était de mettre au point, en collaboration avec T. Rabilloud (CEA Grenoble) de nouveaux colorants des gels d'électrophorèse bi-dimensionnels à la fois sensibles et compatibles avec la spectrométrie de masse (MALDI-MS et ESI-MS).

1 Limites des colorants classiques : sensibilité, compatibilité, expertise

Si l'électrophorèse bi-dimensionnelle est une technique de choix pour séparer la majorité des protéines d'un échantillon, la visualisation de ces protéines reste une étape délicate. En effet, la coloration doit permettre dans certains cas, de quantifier la protéine présente dans le gel et/ou d'être compatible avec l'analyse par spectrométrie de masse. Trois colorations sont couramment utilisées : les colorations à l'argent, au bleu de coomassie colloïdal et fluorescentes.

La première, coloration au nitrate d'argent, permet une bonne quantification des protéines [Rabilloud, 2000]. Par contre, bien qu'étant la plus sensible des colorations non radioactives [Rabilloud, 2000], elle reste délicate à utiliser et est peu compatible avec la spectrométrie de masse MALDI et ESI [Gevaert, 2000]. En effet, les ions argent sont complexés par les acides aminés des protéines [Rabilloud, 1990], [Shevchenko, 1996], [Rabilloud, 2000] et la force de complexation dépend de l'affinité entre les ions et les acides aminés. De ce fait, si les ions argent sont trop fortement complexés, certaines protéines pourront être perdues pour l'analyse par spectrométrie de masse. Il sera donc important de colorer le gel superficiellement pour laisser un maximum de peptides disponibles

pour l'analyse. De ce fait, elle demande une grande expertise pour être compatible avec la spectrométrie de masse.

Le second colorant est le bleu de coomassie colloïdal, plus facile à utiliser, compatible avec la spectrométrie de masse, mais moins sensible, ce qui le restreint à l'analyse des protéines majoritaires (tableau 1) [Rabilloud, 2000]. Or, le manque de sensibilité est un inconvénient majeur, puisque les analyses protéomiques ont pour but de caractériser, entre autre, des protéines en très faibles quantités.

Enfin, une nouvelle technique de détection fluorescente a été élaborée, à l'aide de sondes organiques (Sypro Orange, Sypro Red) ou de ruthénium (Sypro Ruby, Sypro Rose et Complexes RUBPS) [Rabilloud, 2000]. Ces colorants possèdent une sensibilité intermédiaire entre les colorations au bleu de coomassie colloïdal et au nitrate d'argent classique [Lauber, 2001], ainsi qu'une bonne compatibilité avec la spectrométrie de masse [Rabilloud, 2001]. Cependant, ils nécessitent un appareillage onéreux (détection de fluorescence).

D'autres techniques de détection, moins classiques, ont aussi été développées, parmi lesquelles, les colorations négatives par précipitation de métaux lourds, qui sont sensibles, douces mais non linéaires, la plus connue étant la coloration au Zinc–Imidazole [Ortiz, 1992]. Cette coloration offre une bonne compatibilité avec les protocoles de digestion enzymatique et d'extraction des peptides du gel et ne présente pas d'interférence avec la spectrométrie de masse, les protéines étant ni fixées ni colorées directement. Cependant, certaines classes de protéines, de faibles masses moléculaires et les glycoprotéines [Courchesnes, 1997] ne sont pas détectées. De plus, il est difficile de distinguer les taches sur le gel d'où une excision délicate.

Colorants	Sensibilité	Reproductibilité	Linéarité	Compatibilité MS
Bleu colloïdal	10ng	oui	oui	++
fluorescents	5ng	oui	oui	++
argent	0.1ng	faible	+/-	+/-

Tableau 1: Caractéristiques des colorants classiques

L'idéal serait donc une nouvelle coloration au moins aussi sensible que la coloration au nitrate d'argent, avec la même simplicité de visualisation, la même stabilité dans le temps et complètement compatible avec la spectrométrie de masse. Dans cette optique, plusieurs protocoles ont déjà permis d'améliorer la compatibilité des colorations à l'argent avec la spectrométrie de masse en y ajoutant une étape de décoloration soit à l'aide de ferricyanide [Gharahdaghi, 1999] soit de peroxyde
d'hydrogène [Sumner, 2002]. Ces améliorations ne permettent cependant pas, après analyse par spectrométrie de masse, d'obtenir les mêmes résultats pour les échantillons issus de gels colorés avec des colorations fluorescentes ou au bleu de coomassie colloïdal. Ainsi, le but de notre étude était de comprendre pour quelles raisons les colorations classiques à l'argent sont peu compatibles avec la spectrométrie de masse et d'établir de nouveaux protocoles de colorations avec une meilleure compatibilité, tout en conservant une bonne sensibilité comparable à celle des colorants classiques. La coloration élaborée est en fait une coloration au nitrate d'argent pour laquelle les solvants utilisés lors de l'étape de fixation ont été modifiés. Ainsi, le formaldéhyde, couramment utilisé lors de cette étape, a été remplacé par du carbohydrazide. Des tests de sensibilités ont été réalisés en parallèle sur les différentes colorations et les analyses par spectrométrie de masse ont été effectuées aussi bien en ESI qu'en MALDI pour vérifier la compatibilité de cette coloration avec les différents modes d'ionisation. Les différents résultats de cette étude ont fait l'objet d'une publication dans le journal *Proteomics*.

2 Publication

About mechanism of interference of silver staining with peptide mass spectrometry.

Sophie Richert, Sylvie Luche, Mireille Chevallet, Alain Van Dorsselaer, Emmanuelle Leize-Wagner, Thierry Rabilloud

(Acceptée dans Proteomics, Août 2003)

ABOUT THE MECHANISM OF INTERFERENCE OF SILVER STAINING WITH PEPTIDE MASS SPECTROMETRY

Sophie Richert 1, Sylvie Luche 2, Mireille Chevallet 2, Alain Van Dorsselaer 1, Emmanuelle Leize-Wagner 1 and Thierry Rabilloud 2*

1: Laboratoire de Spectrométrie de Masse Bio-Organique, UMR CNRS 7509, ECPM, 25 rue Becquerel, 67008 Strasbourg Cedex

 CEA- Laboratoire de Bioénergétique Cellulaire et Pathologique, EA 2943, DRDC/BECP CEA-Grenoble, 17 rue des martyrs, F-38054 GRENOBLE CEDEX 9, France

(Running title): silver staining interference with mass spectrometry

*: to whom correspondence should be addressed

Correspondence : Thierry Rabilloud, DRDC/BECP CEA-Grenoble, 17 rue des martyrs, F-38054 GRENOBLE CEDEX 9 Tel (33)-438-78-32-12 Fax (33)-438-78-51-87 e-mail: Thierry.Rabilloud@cea.fr

Abbreviations: ESI: electrospray ionisation; MALDI: Matrix-Assisted Laser Desorption and Ionization; MS: mass spectrometry; RuBPS: ruthenium II tris (bathophenanthroline disulfonate).

Abstract

The mechanism by which silver staining of proteins in polyacrylamide gels interferes with mass spectrometry of peptides produced by proteolysis has been investigated. It was demonstrated that this interference increases with time between silver staining and gel processing, although the silver image is constant. This suggested an important role of the formaldehyde used in silver staining development in this interference process. Consequently, a formaldehyde-free staining protocol has been devised, using carbohydrazide as the developing agent. This protocol showed much increased peptide coverage and retained the sensitivity of silver staining. These results were however obtained at the expense of an increased background in the stained gels and of a reduced staining homogeneity.

1 Introduction

Proteomics analyses impart special constraints on the detection techniques used after twodimensional gel electrophoresis. For example, in addition to the standard sensitivity issues, proteomics put special emphasis on the interference with the microcharacterization techniques used afterwards, i.e. in most cases mass spectrometry analysis by MALDI-MS or nanoESI-MS/MS. For these reasons, silver staining, which is still the most sensitive non-radioactive detection technique, is not optimal, because losses in peptide masses or problems in nanoESI-MS/MS have been documented [1, 2].

Apart from silver staining, the most widely used detection technique relies on colloidal Coomassie Blue [3]. While this technique provides much better results in terms of linearity, homogeneity and interference with mass spectrometry, its sensitivity is much too low. Thus, either high loads of proteins must be used, at the risk of protein losses by precipitation, or the analysis is restricted to major proteins only.

Heavy salt precipitation has also been used as a mild and sensitive detection technique. The most sensitive technique in this family is represented by the zinc-imidazole protocol [4]. This technique performs extremely well for mass spectrometry, as there is no fixation of the proteins and therefore maximal accessibility of the cleavage sites to digestion and maximal extractability of the resulting peptides. Despite these very strong points, this technique suffers from some drawbacks that have limited its widespread use. Among those are its poor ability to detect some classes of proteins, such as low molecular weight proteins and glycoproteins [5] and its poor contrast that makes precise band or spot excision difficult.

Another detection technique which appeared recently is based on fluorescence, either of purely organic probes (e.g. Sypro Orange and Sypro Red) or of ruthenium or europium complexes (Sypro Ruby, Sypro Rose and RUBPS complex). Organic probes show a sensitivity slightly better than colloidal Coomassie blue, while ruthenium complexes have a sensitivity slightly inferior to that of silver staining but a much improved compatibility with mass spectrometry [6, 7]. These methods, however, require expensive hardware to detect the fluorescent signal. Moreover, the uniform illumination required for spot excision leads to some fading of fluorescence and thus losses of performance. This effect is much more pronounced for purely organic fluorophores than for metal-organic fluorophores. In addition, UV tables used sometimes for this illumination are rather hazardous. Last but not least, many fluorescent detection protocols are not steady-state, so that the stained gel cannot be kept for future use if convenient.

Thus, the best situation among the methods available to date would be to combine the sensitivity of silver staining, its simplicity of visualization for spot excision and its stability over time, but to improve its compatibility with peptide mass analysis. Interesting methods in this trend have been devised by destaining protocols performed after staining, either with ferricyanide [8] or

with hydrogen peroxide [9]. While these methods allow an important improvement in peptide recovery over undestained gels, the peptide coverage obtained are still inferior to those obtained with Coomassie Blue, zinc imidazole or fluorescent probes. We therefore decided to investigate the chemical mechanism of the interference of silver staining with peptide analysis with mass spectrometry techniques, in order to devise improved silver staining or post-staining processing protocols.

2. Materials and methods

2.1. Gel electrophoresis

Proteins were separated by SDS-PAGE, either in the standard Tris-glycine system, or in the Tris-taurine [10]. Molecular weight markers (Bio Rad, Broad range) were diluted 200 and 2000 fold in SDS sample buffer to reach a concentration range of 10 and 1 ng/ μ 1 for each band. the required volumes were loaded on top of a 10% gel to give the adequate concentrations, ranging from 5 to 400 ng per band.

2.3. Detection of proteins after electrophoresis

Detection of proteins by colloidal Coomassie blue or fluorescent complexes was carried out according to published protocols [3, 6]. For silver staining, fast and long methods with silver nitrate were used [11, 12]. For comparison, ammoniacal silver methods were also used [13]. The formaldehyde-free staining protocol was based on the fast silver staining methods, but used carbohydrazide instead of formaldehyde as the developing agent. Other chemicals, i.e. orthoand para-phenylenediamine, phenidone, hydroxylamine, semithiocarbazide and semicarbazide were also tested. The silver staining protocol is as follows:

1. Fix the gels (1 h + overnight) in 5% acetic acid/30% ethanol (v/v). If a shorter time is preferred, gels can be fixed in 10% acetic acid/30% ethanol (v/v) for 3 x 30 min.

2. Rinse in water for 4 x 10 min.

3. To sensitize, soak gels for 1 min (1 gel at a time) in 0.8 mM sodium thiosulfate

4. Rinse 2x 1 min in water .

5. Impregnate for 30-60 min in 12 mM silver nitrate (0.2g/l). The gels may become yellowish at this stage.

6. Rinse in water for 5-15 s

7. Develop image (1-2 min) in 3% potassium carbonate containing 300-500 μ M carbohydrazide and 125 μ 1 10% sodium thiosulfate per liter.

 Stop development (30-60 min) in a solution containing 40 g of Tris and 20 ml of acetic acid per liter.

9. Rinse with water (several changes) prior to drying or densitometry.

Spot destaining was performed according to previously-described methods 8,9 either immediately after silver staining (including the stop and wash steps) or 2 days after silver staining. In control experiments, 3mM formaldehyde was used in place of carbohydrazide for image development.

2.4. Mass spectrometry analysis

Stained proteins spots or bands were excised (on a UV table for fluorescent detection), and shrunk in 1 ml of 50% ethanol for 2 hours.

In gel digestion :

Each gel slice was cut into small pieces with a scalpel, washed with 100 μ l of 25 mM NH4HCO3 and dehydrated with 100 μ 1 of acetonitrile. This operation was repeated twice. Reduction was achieved by 1 hour treatment with 10mM DTT at 57°C. Alkylation reaction was performed by 25mM Iodoacetamide for 45 min at room temperature, protected from light. Finally, gel spots were washed 3 times for 5 minutes again alternately with 25mM ammonium carbonate and acetonitrile. Gel pieces were completely dried with a Speed Vac before tryptic digestion. The dried gel volume was evaluated and three volumes of trypsin (Promega, V5111), 12.5 ng/µl, in 25 mM NH4HCO3 (freshly diluted) were added. The digestion was performed at 35°C overnight with 5 to 10 μ l of buffer. The gel pieces were centrifuged and 5 μ l of 25% H2O/70% Acetonitrile/5% HCOOH were added to extract peptides. The mixture was sonicated for 5 min. and centrifuged. The supernatant was recovered and the operation was repeated once. For MALDI-MS analysis, the supernatant volume was reduced under nitrogen flow to 4 μ l, 1 μ l of H2O/5% HCOOH were added and 0.5 µl of the mix were used for the analysis. For nano LC-MS/MS, the supernatant solvent was completely evaporated in order to remove all acetonitrile from the sample. Then, 10 µl of H2O/5% HCOOH were added and injected in the nano HPLC system.

MALDI-MS : For MALDI mass spectrometry, mass measurements were carried out on a Bruker BIFLEX tion was dried under vacuum. The sample was washed one to three times by applying 1 μ 1 of aqueous HCOOH (5 %) solution on the target and then flushed after a few seconds. In positive mode, internal calibration is performed with tryptic peptides coming from autodigestion of trypsin, with respectively monoisotopic masses at m/z = 842.51; m/z = 1045.564; m/z = 2211.105. Monoisotopic peptide masses were assigned and used for databases searches.

These files were then fed into the search engine MASCOT (Matrix Science, London, UK). The data were searched against NCBI non-redundant protein sequence database with trypsin plus potentially two missed cleavages. All proteins present in NCBI were taken into account without any pI and MW restrictions. Some variable modifications are taken into account, like methionine oxidation, cysteine carbamidomethylation. The peptide mass error was limited to 50 ppm.

NanoLC-MS/MS : Nanoscale capillary liquid chromatography-tandem mass spectrometric (LC-MS-MS) analysis of the digested proteins were performed using a CapLC capillary LC system (Micromass, Manchester, UK) coupled to a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Q-TOF II, Micromass). The LC-MS union was made with a PicoTip (New Objective, Woburn, MA) fitted on a ZSPRAY (Micromass) interface.

Chromatographic separations were conducted on a reversed-phase (RP) capillary column (Pepmap C18, 75µm i.d., 15 cm lenght, LC Packings) with a 200 nL/min flow. The gradient profile used consisted of a linear gradient from 95% A (H2O / 0.05% HCOOH) to 45% B (acetonitrile / 0.05% HCOOH) in 35 min. followed by a linear gradient to 95% B in 1 min. Mass data acquisitions were piloted by MassLynx software (Micromass, Manchester, UK) using automatic switching between MS and MS/MS modes. The internal parameters of Q-TOF II were set as follows. The electrospray capillary voltage was set to 3.0 kV, the cone voltage set to 30 V, and the source temperature set to 80°C. The MS survey scan was m/z 300-1500 with a scan time of 1 s and a interscan time of 0.1s. When the intensity of a peak rose above a threshold of 8 counts, tandem mass spectra were acquired. Normalized collision energies for peptide fragmentation was set using the charge-state recognition files for +1, +2 and +3peptides ions. The scan range for MS/MS acquisition was from m/z 50 to 1500 with a scan time of 1 s and a interscan time of 0.1s. Fragmentation was performed using argon as the collision gas and with a collision energy profile optimized for various mass ranges of precursor ions. Mass data collected during a nanoLC-MS/MS analysis were processed and converted into a .PKL file to be submitted to the search software MASCOT (Matrix Science, London, UK). Searches were done with a tolerance on mass measurement of 0.25 Da in MS mode and 0.5 Da in MS/MS mode.

Results

3.1. Sensitivity evaluation

This test was carried out by staining serial dilutions of protein markers separated by SDS PAGE, as shown in Figure 1. The formaldehyde-free protocol resulted in extremely fast development (1-2 minutes), but also gave a dark background that can be partly filtered at the scanning step. In addition, the staining was located at the very surface of the gel and was almost undetectable by scanning the gel in its thickness, while the formaldehyde silver staining gave a much thicker silver deposit (data not shown). The sensitivity was roughly equivalent for the two staining methods, and 10 ng were detectable with both methods. Rather similar results were obtained with semicarbazide and semithiocarbazide. However, these chemicals were less easy to use and gave more erratic results. We could not obtain any adequate staining with phenidone, hydroxylamine or phenylene diamine.

3.2. Mass spectrometry

The peptide mass fingerprinting by MALDI-MS and nanoLC ESI-MS/MS experiments were carried out at least in triplicate to get an average value of sequence coverage. Typical results are shown in figures 2 and 3 and in Table 1. From these data, several important trends emerged:

The classical silver staining with formaldehyde in the developer gives much lower sequence coverage, even after destaining, than the more benign but less sensitive methods (colloidal coomassie blue, fluorescent ruthenium complexes)

The silver staining protocol with carbohydrazide gives much improved sequence coverage over the classical formaldehyde silver stain. The sequence coverage is increased if the stained bands are destained by the ferricyanide method [8]. The resulting coverage is then very close to the one obtained for example with Coomassie Blue.

It must be mentioned that the nanoLC ESI-MS/MS experiments were carried out on both an ion trap (Esquire; Bruker) and a Q-TOF II instrument (Micromass). Comparable results were obtained in both instruments, thereby showing the versatility of the methods described here.

In order to further investigate the mechanisms at play in the silver staining interference process, additional experiments were carried out by varying the destaining protocol. Briefly, the gels were either destained the same day as staining or left for another 48 hours in a water bath before destaining. While the sequence coverage was hardly effected by this additional period of time between staining and destaining in the case of the carbohydrazide-silver stain, the coverage decreased dramatically when the gels were left for 48 hours in water after formaldehyde-silver staining and before band excision and destaining. As a matter of fact, no protein could be identified from a 200 ng band after this 48 hours bath, while this amount of protein gave quite adequate identification when destaining is carried out on the same day as staining (see Table 1). In another experiment, the silver metal-containing surface layer was scraped from the gel plugs, which were then used without destaining. The sequence coverage was slightly better than the one observed on undestained gels, but clearly inferior to the one reached with additional destaining. Destaining was therefore adopted as a routine procedure prior to MS analysis.

To determine the useful threshold for the different detection methods, sensitivity experiments were carried out on serial dilutions of proteins separated on gels and then detected by different techniques. The results are shown on Table 2. They clearly demonstrate that the ferricyanide destaining protocol is clearly superior to the one based on hydrogen peroxide. More importantly, these data also show that the useful sensitivity is close to 0.5 picomoles, although much lower levels of proteins can be detected. This limit does not seem to be related to the silver staining per se, as the same limit is also encountered for the ruthenium-stained gels. It must be reminded, however, that this limit is obtained in SDS PAGE with rather wide lanes (ca 8 mm). Due to the focusing effect that allows a much smaller gel plug to be analyzed, the limits seems to be lower for two-dimensional gels, close to 100-200 fmol.

Finally, to evaluate the usefulness of the carbohydrazide method in a real proteomics experiment, this method was tested on a complex cell extract separated by 2D gel electrophoresis. The results are shown on figure 4. Apart from the high background, which has been already mentionned and can be corrected, a second drawback appears, which is a limited staining homogeneity, in the sense that some proteins which are easily detected with a classical silver staining are hardly detected with the carbohydrazide protocol (compare panel 4A vs. 4B). However, the carbohydrazide staining method is still fairly more sensitive than colloidal Coomassie Blue (compare panel 4B vs. 4C), and delivered results in less than 24 hours, instead of the 48 hours requested for optimal colloidal Commassie Blue staining

4. Discussion

The results described above provided some precise insights about the chemical mechanisms involved in the interference phenomenon induced by silver staining. While the exact role of the silver metal image itself is not very clear yet, previous blotting experiments from undestained gels [14], in which whole proteins are able to cross the silver image, strongly suggest that the metal image itself is not the major interfering entity. The difference between the classical and formaldehyde-free protocols strongly suggests that formaldehyde is a major interfering compound. The mechanism of this interference is probably some crosslinking between aminoacids reactive side chains (lysine, cysteine, and to a lesser extent serine and threonine), leading to protein reticulation by methylene bridges, in a way quite similar to that observed during histochemical fixation [15]. It is very likely that this crosslinking occurs in a two-step mechanism. The first step involves grafting of a hydroxymethyl group on a reactive side chain, and is followed by a second substitution leading to crosslinking. This two-step mechanism would explain the slow worsening in interference induced upon simple storage in water. However, the improvement shown by destaining in formaldehyde-free, silver stained gels also shows that formaldehyde is not responsible alone for interference, and that remaining silver ions also provide some interference. This is to be linked to the observation that zinc-stained proteins, although transparent, must be freed from zinc ion (so-called mobilization in [4]) before any further analysis by blotting or mass spectrometry can take place. Thus, heavy metal ions protein salts do not lead themselves easily to further analysis.

The conclusions of this work for future improvement is two-fold.

The first track for improvement lies in the availability of formaldehyde-free silver staining protocols for improved sequence coverage at high sensitivity and with all the simplicity of silver staining (especially in hardware) over fluorescent staining. Carbohydrazide was selected over other chemicals such as semicarbazide, thiosemicarbazide on the basis of more regular performances. Hydroxylamine also gave a positive staining, but it was much less sensitive. All of these chemicals did not lead to any peptide modification. However, we were not able with these chemicals to devise a high-contrast and wide scope silver staining protocol of the quality reached with formaldehyde. Decreasing the reducer concentration in the developer decreased the background but also dramatically decreased the sensitivity (data not shown). We also tried reducers used in photography. However, most of them are said to be "tanning", i.e. to induce strong protein crosslinking (e.g. hydroquinone, aminophenols, pyrogallol etc...), and were therefore not tested. Only a few molecules are known to be non -tanning, i.e. not to induce crosslinks. Among those are phenylene diamines and phenidone. However, we could not devise a practical silver staining protocol with any of the latter chemicals.

The second track for improvement is the development of improved destaining protocols. We now know that the main action of destaining protocols is not only the removal of the silver image but the oxidation of remaining formaldehyde and the complexation of silver ion. In the destaining protocols described to date, silver ion complexation is achieved by thiosulfate [8] or by ammonia [9], while silver and formaldehyde oxidation is achieved by hydrogen peroxide [9] or by ferricyanide [8], which is the active component of the classical Benedict's test solution for aldehydes.

Now that the chemical species responsible for interference have been identified, post-silver staining processing protocols with increased efficiency can probably be devised. They would allow even better compatibility between the high sensitivity and high contrast classical silver stain and the downstream peptide mass analysis.

On a more practical point of view, the formaldehyde-free silver staining described here combines the MS compatibility of colloidal Coomassie or fluorescent stains with the sensitivity of silver staining. It allows to reach the same performances in terms of identification than fluorescent probes, but with the permanence of the silver image and the simplicity of spot cutting from such an image. These are obvious practical advantages which combine to the much lower price of silver staining compared to commercial fluorescent probes (e.g. Sypro Ruby) to make this procedure an attractive choice. In addition, this procedure allows to perform a complete study with silver staining, i.e. both the analytical gels and the preparative work, thereby minimizing the confusing variations in spot intensities arising sometimes from changes in detection protocols. However, the poor staining homogeneity, when compared to either classical silver staining or colloidal Coomasie Blue or fluorescent probes, reduces the interest of this protocol for 2D gel-based proteomics. It could be successfully used for fast, sensitive and easy lane visualization in SDS-PAGE-MS/MS-based proteomics.

References

(1) Gevaert, K.; Vandekerckhove, J. Electrophoresis 2000, 21, 1145-1154.

(2) Scheler, C.; Lamer, S.; Pan, Z.; Li, X.P.; Salnikow, J.; Jungblut, P. Electrophoresis 1998, 19, 918-927.

(3) Neuhoff, V.; Arold, N.; Taube, D.; Ehrhardt, W. Electrophoresis 1988, 9, 255-262

(4) Ortiz, M.L.; Calero, M.; Fernandez-Patron, C.; Patron, C.F.; Castellanos, L.; Mendez, E. FEBS Lett. 1992, 296(3), 300-304.

(5) Courchesne, P.L.; Luethy, R.; Patterson, S.D. Electrophoresis. 1997, 18, 369-381

(6) Rabilloud, T.; Strub, J.M.; Luche, S.; van Dorsselaer, A.; Lunardi, J. Proteomics. 2001, 1, 699-704.

(7) Berggren, K.; Chernokalskaya, E.; Steinberg, T.H.; Kemper, C.; Lopez, M.F.; Diwu, Z.; Haugland, R.P.; Patton, W.F. Electrophoresis. 2000, 21, 2509-2521.

(8) Gharahdaghi, F.; Weinberg, C.R.; Meagher, D.A.; Imai, B.S.; Mische, S.M. Electrophoresis. 1999, 20, 601-605.

(9) Sumner, L.W.; Wolf-Sumner, B.; White, S.P.; Asirvatham, V.S. Rapid Commun. Mass Spectrom. 2002, 16, 160-168.

(10) Rabilloud, T.; Valette, C.; Lawrence, J.J. Electrophoresis. 1994, 15, 1552-1558.

(11) Rabilloud, T. Electrophoresis 1992, 13, 429-439.

(12) Sinha, P.; Poland, J.; Schnolzer, M.; Rabilloud T. Proteomics. 2001, 1, 835-840.

(13) Rabilloud, T.; Kieffer, S.; Procaccio, V.; Louwagie, M.; Courchesne, P.L.; Patterson, S.D.; Martinez, P.; Garin, J.; Lunardi, J. Electrophoresis 1998, 19, 1006-1014

(14) Wise, G.E.; Lin, F. J Biochem. Biophys. Methods. 1991, 22, 223-231.

(15) Pearse A.G.E. Histochemistry, theoretical and applied. Churchill Livingstone Edinburgh, 1980, chapter 5

Legends to figures

Figure 1. Sensitivity evaluation of the silver staining

Molecular weight markers (BioRad, broad range) were diluted serially in SDS buffer and separated by SDS electrophoresis. The corresponding gel was stained with standard (formaldehyde) silver staining (panel A), or with the carbohydrazide silver stain (panels B and C). The difference between panels B and C is a background substraction operated at the scanning stage. Protein separated: MYO: myosin (205kDa), GAL: beta galactosidase (116kDa), PHO: glycogen phosphorylase (97kDa) BSA: bovine serum albumin (67kDa), OVA: ovalbumin (46 kDa), CAR: carbonic anhydrase (30 kDa), STI: soybean trypsin inhibitor (21 kDa), LYS: lysozyme (14.5 kDa). Protein content per lane in each panel, from left to right: 200 ng/protein, 100 ng, 50 ng, 20 ng, 10 ng, 5 ng, 2 ng

Figure. 2: Comparison of MALDI-MS Spectra after silver staining.

The Beta galactosidase band (ca. 2 picomoles) was excised from a silver-stained gel either stained with a formaldehyde developer and immediately destained with ferricyanide-thiosulfate (spectrum A) or stained with the carbohydrazide developer and not destained (spectrum B). The MALDI-MS spectra obtained after trypsin digestion of the corresponding bands are shown on the figure. The peaks matching with calculated masses for trypsin digestion of E. coli beta-galactosidase are marked with a solid circle

Figure. 3: Comparison of nanoLC ESI-MS Spectra after silver staining.

The glycogen phosphorylase band (2 picomoles) was excised from a silver-stained gel either stained with the carbohydrazide developer and immediately destained with ferricyanide-thiosulfate (spectrum A) or stained with a formaldehyde developer and destained (spectrum B). Spectrum C come from a ruthenium-stained gel. The combined nanoLC ESI-MS spectra (from retention time 5 to 40 min.) obtained on a Q-TOF II instrument (Micromass) after trypsin digestion of the corresponding bands are shown on the figure. Spectra are normalized in fonction of intensity. The peaks matching with calculated masses for trypsin digestion of glycogen phosphorylase are marked with a solid circle.

Figure 4: comparison of formaldehyde and carbohydrazide silver staining A complex total cell extract from heLa cells was separated by 2D gel electrophoresis. 0.1 mg of proteins were loaded by in gel rehydration. IEF: immobilized pH gradient, pH 4 to 8, linear. SDS-PAGE: 10%T continuous gel.

A; detection with formaldehyde silver staining. B: detection with carbihydrazide silver staining. C: detection with colloidal Coomassie Blue

	Carbonic anhydrase	Ovalbumin	Bovine serum albumin	β Galactosidase	Glycogen phosphorylase
MALDI	Number of peptides/ %coverage				
Colloidal blue	6/ 32%	12 /29% .	16 /30%	24 /27%	36 /42%
Sypro ruby	4 /23%	8 /24%	11 /19%	20 /21%	17 /19%
Ruthenium	6 /32%	6 /22%	7 /12%	10 /8%	46 /47%
Silver staining with HCHO, destained	8 /43%	Not identified	5 /9%	13 /10%	Not identified
Silver staining with hydrazide, destained	7 /32%	12 /35%	17 /32%	16 /20%	17 /22%
Silver staining with hydrazide, no destaining	6 /32%	5 /16%	8 /14%	22 /23%	Not identified
LC/ESI-MS					
Colloidal blue	11 /38%	9 /27% .	33 /46%	25 /27%	32 /35%
Sypro ruby	11 /29%	5 /15%	33 /47%	14 /16%	43 /48%
Ruthenium	9 /26%	8 /17%	34 /57%	17 /19%	28 /28%
Silver staining with HCHO, destained	6 /22%	3 /5%	24 /40%	N/A	14 /14%
Silver staining with hydrazide, destained	6 /20%	Not identified	40 /55%	20 /19%	28 /34%
Silver staining with hydrazide, no destaining	6 /20%	4 /13%	26 /42%	12 /11%	20 /23%

Table 1 : Comparison of the coverage efficiency by mass spectrometry after different detection methods

The marker proteins described in the table are separated on a SDS gel, and detected by the method indicated. The 200 ng band is excised for each protein and then submitted to trypsin digestion. The peptide mixture is then analyzed either with MALDI/MS or with LC/ESI-MS, as described in the methods section.



Figure 1. Sensitivity evaluation of the silver staining

Molecular weight markers (BioRad, broad range) were diluted serially in SDS buffer and separated by SDS electrophoresis. The corresponding gel was stained with standard (formaldehyde) silver staining (panel A), or with the carbohydrazide silver stain (panels B and C). The difference between panels B and C is a background substraction operated at the scanning stage. Protein separated: MYO: myosin (205kDa), GAL: beta galactosidase (116kDa), PHO: glycogen phosphorylase (97kDa) BSA: bovine serum albumin (67kDa), OVA: ovalbumin (46 kDa), CAR: carbonic anhydrase (30 kDa), STI: soybean trypsin inhibitor (21 kDa), LYS: lysozyme (14.5 kDa). Protein content per lane in each panel, from left to right: 200 ng/protein, 100 ng, 50 ng, 20 ng, 10 ng, 5 ng, 2 ng



Figure. 2: Comparison of MALDI-MS Spectra after silver staining.

The Beta galactosidase band (ca. 2 picomoles) was excised from a silver-stained gel either stained with a formaldehyde developer and immediately destained with ferricyanide-thiosulfate (spectrum A) or stained with the carbohydrazide developer and not destained (spectrum B). The MALDI-MS spectra obtained after trypsin digestion of the corresponding bands are shown on the figure. The peaks matching with calculated masses for trypsin digestion of E. coli beta-galactosidase are marked with a solid circle



Figure. 3: Comparison of nanoLC ESI-MS Spectra after silver staining. The glycogen phosphorylase band (2 picomoles) was excised from a silver-stained gel either stained with the carbohydrazide developer and immediately destained with ferricyanidethiosulfate (spectrum A) or stained with a formaldehyde developer and destained (spectrum B). Spectrum C come from a ruthenium-stained gel. The combined nanoLC ESI-MS spectra (from retention time 5 to 40 min.) obtained on a Q-TOF II instrument (Micromass) after trypsin digestion of the corresponding bands are shown on the figure. Spectra are normalized in fonction of intensity. The peaks matching with calculated masses for trypsin digestion of glycogen phosphorylase are marked with a solid circle.



Figure 4: comparison of formaldehyde and carbohydrazide silver staining A complex total cell extract from heLa cells was separated by 2D gel electrophoresis. 0.1 mg of proteins were loaded by in gel rehydration. IEF: immobilized pH gradient, pH 4 to 8, linear. SDS-PAGE: 10%T continuous gel.

A; detection with formaldehyde silver staining. B: detection with carbihydrazide silver staining. C: detection with colloidal Coomassie Blue

3 Résultats complémentaires

3.1 Autres protéines testées

Afin de compléter cette étude, réalisée sur des protéines modèles de caractéristiques biochimiques différentes (tableau 1 publication), l'ensemble des colorants a aussi été testé sur des protéines de cellules Hela, dans le cadre d'une analyse protéomique classique (faible quantité de matériel, diversité des protéines, préparation d'échantillon délicate). Les résultats indiqués pour l' α énolase dans le tableau 2 et illustrés par la figure 1, sont représentatifs de ceux obtenus pour les autres protéines issues de ces cellules (Peroxirédoxine 2, Nucléophosmine, GTP binding protéine, Endoplasmine precurseur, Glucosidase II precurseur). En effet, de bon taux d'identification ont été obtenus, à savoir des pourcentages de recouvrement supérieures à 20% pour des protéines de masses moléculaires de 50kD, pour les protéines issues de gels colorés à l'argent avec le protocole sans formaldéhyde. Ces résultats sont aussi conformes à ceux obtenus pour les protéines modèles (publication).

Tableau 2: Pourcentages de recouvrement de séquence pour l' α énolase et du nombre de peptides séquencés en fonction des différentes colorations testées (19/21 signifie que sur 21 peptides correspondants à la protéine isolés par spectrométrie de masse, 19 ont pu être séquencés par MS/MS)

coloration	% recouvrement	peptides
Bleu de Coomassie C8	39%	13/13
Ruthénium R8	47%	19/21
Argent hydrazide CH8	50%	20/20
Argent formaldéhyde F8	44%	18/21
Argent formaldéhyde décoloré FD8	47%	18/19
Zinc imidazole, décoloré EDTA Z1H	41%	17/18
Zinc imidazole, décoloré acide citrique Z2H	58%	21/22

Figure 1 : Histogramme des pourcentages de recouvrement de séquence pour l' α énolase pour les différentes colorations



3.2 Autres colorants testés

Bien que le but de cette étude ait été focalisé sur la compréhension des mécanismes de colorations à l'argent, afin de les améliorer, d'autres colorants ont aussi été testés dont une coloration négative au Zinc-Imidazole et une coloration fluorescente au Ruthénium.

La détection au Zinc-Imidazole, habituellement décolorée à l'EDTA [Rabilloud, Proteome research], a ici été décolorée à l'acide citrique. Les résultats indiqués dans le tableau 2 montrent un bon pourcentage de recouvrement de séquence dans le cas de l' α énolase (représentatif des autres protéines de cellules Hela). Mais les problèmes de prélèvements des spots, de détection de certaines protéines (protéines de faibles poids moléculaires, glycoprotéines [Courchesnes, 1997]) et de non-linéarité demeurent.

Concernant la coloration au Ruthénium, les tests avaient pour but de vérifier sa compatibilité avec la spectrométrie de masse en mode d'ionisation Electrospray et de la comparer avec le colorant fluorescent classique (Sypro Ruby). Il avait déjà été démontré dans une précédente étude que ces deux colorants possédaient les mêmes caractéristiques en MALDI-MS [Rabilloud, 2001]. Ainsi, la comparaison des spectres MS de la glycogène phosphorylase colorée au Ruthénium et au Sypro Ruby, obtenus par ESI-MS sur une trappe ionique après séparation par nanoLC (figure 2), montre que le nombre de peptides caractéristiques détectés ainsi que leurs intensités sont comparables. Cependant, si ces deux colorants fluorescents possèdent les mêmes caractéristiques de sensibilité et de compatibilité avec la spectrométrie de masse, l'avantage du Ruthénium réside dans le fait qu'il est beaucoup moins onéreux que le Sypro Ruby.



Figure 2: Analyse nanoLC-MS (trappe ionique) de la glycogène phosphorylase. Les peptides reconnus de la protéine sont indiqués par le signe $\mathbf{0}$

3.3 Analyses avec différentes techniques de spectrométrie de masse

Enfin, afin de déterminer si cette nouvelle coloration à l'argent était compatible avec les différentes techniques de spectrométrie de masse couramment utilisées en protéomique (MALDI et ESI), les échantillons ont été analysés par MALDI-MS et nanoLC-MS.

En MALDI-MS, le nombre de peptides spécifiques de la protéine est plus important dans le cas de la coloration à l'argent avec carbohydrazide (sans formaldéhyde), comparé à la coloration à l'argent classique (avec formaldéhyde). Ces résultats sont illustrés par les spectres MS en figure 3 obtenus pour la β galactosidase et résumés pour l'ensemble des protéines modèles dans le tableau 3.





Tableau 3 : Pourcentages de recouvrement de séquences pour l'analyse MALDI-MS

	Anhydrase carbonique	Ovalbumine	BSA	β Galactosidase	Glycogène phosphorylase
Bleu colloïdal	6 peptides	12 peptides	16 peptides	24 peptides	36 peptides
	32%	29% .	30%	27%	42%
Sypro Ruby	4 peptides	8 peptides	11 peptides	20 peptides	17 peptides
	23%	24%	19%	21%	19%
Ruthenium	6 peptides	6 peptides	7 peptides	10 peptides	46 peptides
	32%	22%	12%	8%	47%
Ag formaldéhyde	8 peptides 43%	Pas d'identification	5 peptides 9%	13 peptides 10%	Pas d'identification
Ag carbohydrazide	7 peptides	12 peptides	17 peptides	16 peptides	17 peptides
décoloré	32%	35%	32%	20%	22%
Ag carbohydrazide	6 peptides	5 peptides	8 peptides	22 peptides	Pas
	32%	16%	14%	23%	d'identification

Concernant l'approche ESI-MS, si on compare les spectres obtenus en nanoLC Q-TOF pour la BSA colorée à l'argent classique (avec formaldéhyde) avec décoloration et à l'argent (avec carbohydrazide) avec et sans décoloration (figure 4), un plus grand nombre de peptides caractéristiques pour la nouvelle coloration est détecté qu'il y ait ou non décoloration après l'étape de coloration (7 peptides supplémentaires détectés).



Figure 4: NanoLC-MS (Q-TOF): Bovin Serum Albumin, peptides caractéristiques indiqués par le signe o

En plus d'analyses sur des protéines modèles (BSA, β galactosidase, ovalbumine...), des échantillons issus de protéines de cellules Hela ont aussi été analysés sur deux types d'instruments ESI (trappe ionique et Q-TOF). Les résultats sont indiqués figure 5A et B. Malgré quelques résultats aberrants, liés à des problèmes de stabilité de spray et de ce fait, à l'obtention de spectres de mauvaises qualités, les identifications sont globalement du même ordre quelque soit l'analyseur.

Figure 5A: Résultats de nanoLC MS/MS (Q-TOF, Trappe ionique) de protéines de cellules Hela

Q-TOF

nucleophosmin (5)		
spot	% recouvrement	peptides
bleu de coomassie C5	5%	2/2
Ruthénium 5R	22%	7/8
Argent hydrazide CH5	34%	8/9
Argent formol F5	22%	7/7
Argent formol décoloré FD5	38%	7/9
Zn imidazole décoloré EDTA Z1E	24%	5/5
Zn imidazole décoloré acide citrique Z2E	37%	6/8

Trappe ionique

nucleophosmin (5)		
spot	% recouvrement	peptides
C5	53%	9/21
5R	51%	7/16
CH5	59%	8/20
F5	62%	9/18
FD5	29%	6/13
Z1E	43%	5/19
Z2E	0%	

peroxiredoxine 2 (6)

bleu de coomassie C6	33%	8/9
Ruthénium 6R	40%	9/9
Argent hydrazide CH6	40%	11/11
Argent formol F6	58%	11/11
Argent formol décoloré FD6	40%	11/11
Zn imidazole décoloré EDTA Z1F	27%	7/7
Zn imidazole décoloré acide citriqueZ2F	42%	12/12

peroxiredoxine 2 (6)

C6	47%	12/13
6R	62%	11/13
CH6	64%	12/15
F6	42%	9/10
FD6	45%	9/10
Z1F	35%	5/10
Z2F	pas analysé	

GTP binding prot RAN (7)

bleu de coomassie C7	0%	/
Ruthénium 7R	35%	8/9
Argent hydrazide CH7	35%	8/9
Argent formol F7	47%	11/12
Argent formol décoloré FD7	36%	6/7
Zn imidazole décoloré EDTA Z1G	35%	8/9
Zn imidazole décoloré acide citrique Z2G	40%	9/11

α enolase (8)

bleu de coomassie C8	39%	13/13
Ruthénium 8R	47%	19/21
Argent hydrazide CH8	50%	20/20
Argent formol F8	44%	18/21
Argent formol décoloré FD8	47%	18/19
Zn imidazole décoloré EDTA Z1H	41%	17/18
Zn imidazole décoloré acide citrique Z2H	58%	21/22

GTP binding prot RAN (7)

C7	50%	9/15
7R	45%	7/10
CH7	28%	5/8
F7	31%	5/9
FD7	0%	
Z1G	35%	6/10
Z2G	55%	6/15

α enolase (8)

C8	0%	
8R	64%	15/29
CH8	64%	16/22
F8	0%	
FD8	69%	19/35
Z1H	51%	18/29
Z2H	48%	11/24

Figure 5B: Histogramme des pourcentages de recouvrement de séquences pour les résultats de nanoLC MS/MS (Q-TOF, Trappe ionique) de protéines de cellules Hela



Trappe ionique







Grâce à cette étude, il a été possible de déterminer l'étape responsable des problèmes de compatibilité entre la spectrométrie de masse et les colorations à l'argent. Il semble que des phénomènes d'interférences soient générés par les colorations à l'argent au cours de l'étape de

révélation et que l'agent responsable en soit le formaldéhyde. Il a donc été remplacé par du carbohydrazide, choisi parmi d'autres solvants de révélation : le semicarbozide, le thiosemicarbozide ou l'hydroxylamine.

Les interférences liées au formaldéhyde pourraient s'expliquer par la formation de liaisons entre les chaînes latérales de certains acides aminés (lysine, cystéine et dans une moindre mesure sérine et thréonine), conduisant à une agrégation des protéines entre elles par des ponts méthylène. Un groupement hydroxy-méthyl se fixerait dans une première étape sur les chaînes latérales des acides aminés qui initierait une réaction de substitution, liant les acides aminés (figure 6A). Ainsi, plus le formaldéhyde reste en contact avec le gel, plus les liaisons ont le temps de se former, expliquant la diminution des résultats dans le cas de l'utilisation de formaldéhyde couplée à une décoloration tardive (48 ou 72h) par rapport à une décoloration immédiate.



Figure 6A: Interaction acide aminé/formaldéhyde

Cependant, il ne faut pas négliger le rôle des ions argent. En effet, comme dans le cas des ions zinc (coloration Zinc-Imidazole), pour avoir une bonne réponse en spectrométrie de masse, il est nécessaire d'éliminer tous les ions métalliques avant l'analyse.

En plus d'établir un nouveau protocole de coloration à l'argent, cette étude a permis de tester de nouveaux colorants ainsi que différentes méthodes de décoloration et de mieux évaluer l'intérêt de cette dernière étape. En effet, son rôle permet de supprimer les ions argent, d'éviter une trop forte complexation mais aussi d'éliminer le formaldéhyde qui s'oxyde. Les ions argent sont éliminés par du thiosulfate [Gharahdaghi, 1999] ou de l'ammoniaque [Sumner, 2002] alors que le formaldéhyde est éliminé grâce à l'ajout de peroxyde d'hydrogène [Sumner, 2002] ou de ferrycyanide [Gharahdaghi, 1999]. Ainsi, ce nouveau protocole de coloration à l'argent (sans formaldéhyde) conserve la sensibilité des colorants à l'argent classique tout en associant une bonne compatibilité avec

l'analyse par spectrométrie de masse (ESI et MALDI) et en gardant ses avantages face aux colorants fluorescents (appareillage onéreux).

Conclusions générales

Cette étude a permis de mieux comprendre les raisons pour lesquelles la coloration à l'argent est plus difficilement compatible avec la spectrométrie de masse. Ces raisons seraient liées à la présence de formaldéhyde dans les solutions de révélation des protéines et à la présence des ions argent. De ce fait, le nouveau protocole de coloration à l'argent élaboré sans formaldéhyde donne d'aussi bons résultats que les colorants classiques (ruthénium et bleu colloïdal), particulièrement concernant sa compatibilité avec la spectrométrie de masse.

Chapitre 2:

Optimisation des étapes d'extraction et de digestion des protéines *in-gel*

Les protocoles utilisés en analyse protéomique ont été optimisés afin d'obtenir les meilleures identifications pour *différentes catégories de protéines*. Cependant, que ce soit au niveau des procédés d'extraction des peptides des gels qu'au niveau des protocoles de digestion, il peut être important d'adapter les méthodes employées en fonction des protéines étudiées, par exemple en tenant compte, lorsque cela est possible, de leurs propriétés physico-chimiques. L'étude des protéines membranaires, protéines hydrophobes, illustre bien ce problème. En effet, ces protéines nécessitent une préparation d'échantillon particulière (centrifugations, détergents spécifiques) [Santoni, 2000], [Wu, 2003], mais aussi des solvants d'extraction différents de ceux couramment utilisés [Van Montfort, 2002]. Pour d'autres protéines, possédant peu ou pas de lysines et d'arginines, la digestion trypsique, couramment utilisée en analyse protéomique, ne sera d'aucun intérêt.

Il était donc intéressant de vérifier que les conditions d'extraction généralement employées étaient les plus adéquates pour l'analyse protéomique mais aussi de tester de nouvelles enzymes de digestion seules ou en combinaison.

Pour ces différentes études, le protocole a été le suivant :

- Séparation des protéines par gel SDS PAGE
- Digestion enzymatique
- Analyse MALDI-MS

1 Tests d'extraction

Différentes conditions ont été testées en modifiant les solvants et les durées d'extraction. Ces tests ont été réalisés sur des protéines modèles de caractéristiques biochimiques différentes afin d'avoir un panel représentatif de celui rencontré lors des études protéomiques (tableau 4).

Protéine	Masse (Da)	Nombre d'acides aminés
BSA (Bovin Serum Albumin)	69248	607
ADH (Alcohol Deshydrogenase)	36669	347
Lysozyme C (blanc d'oeuf de poule)	16228	147
Porine bactérienne	37639	350

Tableau 4 : Protéines modèles étudiées

Si les trois premières protéines sont peu hydrophobes, la 4^{ème} protéine, la porine, est une protéine à 7 domaines trans-membranaires de séquences très hydrophobes (diagrammes d'hydrophobicité, figure 6)

Figure 6: Comparaison des diagrammes d'hydrophobicité de la porine et de l'ADH



Pour cette étude, plusieurs paramètres ont été testés :

- Temps d'extraction : 1 h, 4 h, 1 nuit
- Solvant d'extraction : concentration en acétonitrile (ACN)

H₂O/ACN/1% acide formique 0%,60%, 100% (ACN)

Chaque expérience a été réalisée 3 fois sur chacune des 4 protéines.

Concernant la concentration en ACN, les résultats varient fortement d'une expérience à l'autre, ne permettant pas d'établir de conclusions, seules quelques hypothèses ont pu être émises.

Au bout d'1 heure d'extraction, les meilleurs résultats (pourcentages de recouvrement de séquences) sont obtenus pour 100% d'ACN suivis de 60% d'ACN (Figure 7).

Par contre, si les temps d'extraction sont augmentés, différents cas de figures sont observés :

- Une chute ou une stabilisation des pourcentages de recouvrement de séquence pour la BSA, le Lysozyme C et l'ADH, ce qui peut être dû à une adsorption aspécifique des peptides au cours du temps sur les parois en polypropylène des puits. Pour vérifier cette hypothèse, de nouvelles expériences devront être réalisées en utilisant d'autres supports, en ajoutant une solution plus concentrée en acide formique ou en soumettant la plaque aux ultra-sons pour tenter de décrocher les peptides.
- Une augmentation des pourcentages de recouvrement de séquences à 0% d'ACN pour la BSA (graphique, figure 7) ou encore à 100% d'ACN pour l'ADH au bout d'une nuit d'extraction. Ces résultats peuvent s'expliquer par une meilleure diffusion des peptides au cours du temps mais reste sans explication face à la diminution ou la stabilisation observée au bout de 4 heures d'extraction. Les expériences devront être à nouveau réalisées mais sur un plus grand nombre d'échantillons et de façon répétée afin d'éliminer ce type de variabilité.

Figure 7 : Résultats des tests d'extraction pour la BSA, l'ADH et le Lysozyme C ; en fonction des temps d'extraction et de la composition des solvants

Pourcentage de recouvrement de séquence de la BSA (%) Pour 1 5 pmol								
% ACN	Temps Exp1 Exp2 Exp3 Movenne Ecart type							
	1h	58	51	60	56	4,7		
	4h	44	39	53	45	7,1		
0	1 nuit	60	49	\ge	54	7,8		
	1h	69	57	59	61	6,4		
	4h	\ge	61	\ge	61	0		
60	1 nuit	64	30	27	40	20,6		
	1h	73	57	66	65	8		
	4h	63	62	60	61	1,5		
100	1 nuit	56	43	45	48	7		



Pourcentage de recouvrement de séquence de l'ADH (%)						
			pour 1	,5 pmo		
% ACN	temps	Exp1	Exp2	Exp3	Moyenne	Ecart type
	1h	52	52	33	46	11
	4h	47	36	39	40	5,7
0	1 nuit	36	27	54	39	13,7
	1h	51	36	53	47	9,3
	4h	41	55	\ge	48	9,9
60	1 nuit	43	42	55	47	7,2
	1h	43	41	55	46	7,6
	4h	57	31	52	47	13,8
100	1 nuit	65	52	58	58	6,5





Les tests d'extraction des porines donnent les mêmes résultats que ceux obtenus pour les autres protéines (figure 8A et B), à savoir de bons pourcentages de recouvrement de séquence pour 100% d'ACN au bout d'1 heure.

	Pourcen	tage de re	couvre	ement	des p	orines	
% ACN	Expériences	Temps	1	2	3	moyenne	Ecart type
		1h	42	44	42	42	1,2
0	Sans détergent	4h	44	42	44	43	1,2
		1nuit	44	42	44	43	1,2
		1h	36	44	44	41	4,6
	Avec détergent	4h	39	44	44	42	1,2 1,2 4,6 2,9 1,2 4 5,1 3,8 0 0,6 2
		1nuit	42	44	44	43	1,2
		1h	51	51	44	48	4
	Sans détergent	4h	51	41	44	45	5,1
		1nuit	51	44	50	48	3,8
60	Avec détergent	1h	51	51	51	51	0
		4h	43	44	44	43	0,6
		1nuit	51	49	47	49	2
		1h	50	50	51	50	0,6
	Sans détergent	4h	51	51	50	50	0,6
		1nuit	48	42	50	46	4,2
100		1h	51	51	51	51	0
	Avec détergent	4h	51	44	44	46	4
		1nuit	54	49	46	49	4

Figure 8A : Résultats des tests d'extraction sur les Porines



Figure 8B : Graphique illustrant les résultats des tests d'extraction sur les Porines

Par contre, l'ajout de détergent (octyl-β-glucopyranoside [Van Montfort, 2002]), décrit pour favoriser l'extraction des peptides hydrophobes, n'a pas permis d'améliorer le pourcentage de recouvrement de séquence protéique. En effet, les peptides trypsiques hydrophobes issus des porines (figure 9) possèdent des masses variant entre de 3300 à 4500 Da. De ce fait, ce sont de grands peptides difficiles à extraire et à analyser par MALDI-MS (bonne résolution sur une gamme de masse jusqu'à 3000m/z en mode réflecteur). Le choix de cette protéine n'a donc pas été judicieux, il faudra, pour tester l'intérêt de ce détergent lors de l'extraction, analyser d'autres protéines hydrophobes générant des peptides trypsiques inférieurs à 3000m/z ou utiliser d'autres enzymes de digestion, par exemple la chymotrypsine.

Figure 9 : Séquence de la Porine, peptides hydrophobes indiqués en vert et en rouge les peptides identifiés par MALDI-MS

¹ MKLKNTLGVV ¹¹ IGSLVAASAM	²¹ NAFAQGQNSV ³¹ EIEAFGKRYF ⁴¹ TDSVRNMKNA
⁵¹ DLYGGSIGYF ⁶¹ LTDDVELALS	⁷¹ YGEYHDVRGT ⁸¹ YETGNKKVHG ⁹¹ NLTSLDAIYH
¹⁰¹ FGTPGVGLRP ¹¹¹ YVSAGLAHQN	¹²¹ ITNINSDSQG ¹³¹ RQQMTMANIG ¹⁴¹ AGLKYYFTEN
¹⁵¹ FFAKASLDGQ ¹⁶¹ YGLEKRDNGH	¹⁷¹ QGEWMAGLGV ¹⁸¹ GFNFGGSKAA ¹⁹¹ PAPEPVADVC
²⁰¹ SDSDNDGVCD ²¹¹ NVDKCPDTPA	²²¹ NVTVDANGCP ²³¹ AVAEVVRVQL ²⁴¹ DVKFDFDKSK
²⁵¹ VKENSYADIK ²⁶¹ NLADFMKQYP	²⁷¹ STSTTVEGHT ²⁸¹ DSVGTDAYNQ ²⁹¹ KLSERRANAV
³⁰¹ RDVLVNEYGV ³¹¹ EGGRVNAVGY	³²¹ GESRPVADNA ³³¹ TAEGRAINRR ³⁴¹ VEAEVEAEAK

2 Tests de digestion

Si la trypsine est l'enzyme la plus couramment utilisée, elle s'avère dans certains cas, inadaptée, notamment pour les protéines pauvres en lysine et arginine. Les peptides générés seront trop grands et de ce fait, non détectés par spectrométrie de masse (MALDI et ESI). Afin d'optimiser cette étape et ainsi améliorer le pourcentage de recouvrement de séquences, d'autres protéases ont été testées (chymotrypsine, AspN) (tableau 5) seules ou en combinaison avec la trypsine (figure 10). Les expériences ont été répétées de 4 (AspN) à 6 fois.

Figure 10 : Stratégie d'analyse pour les tests de digestions. Digestions en parallèle avec une enzyme, digestions combinées avec 2 enzymes et résultats cumulés (cumul des résultats obtenus pour 2 enzymes seules).



Tableau 5 : Caractéristiques des protéases testées

Enzymes	Coupures	Exceptions
Trynsine	niveau C ter	Si K ou R suivis de P (C ter)
Trypsine	de K et R	
Chymatrynging	niveau C ter	Si F, L,M, W ou Y suivis de P (C ter)
Chymou ypsme	de F, L, M, W et Y	Si Y précédé de P (N ter)
Asp-N	niveau N ter de D	

Ces 3 protéases ont des spécificités différentes avec, pour la chymotrypsine un plus large spectre de coupure. Donc, comme l'enzyme coupe après 5 acides aminés, il y a plus de chances *à priori* de générer des peptides de tailles compatibles avec l'analyse par spectrométrie de masse, mais aussi des spectres plus difficiles à interpréter.

Concernant les résultats des tests avec une enzyme seule :

- Les meilleurs pourcentages de recouvrement de séquences (figure 12) sont obtenus avec la trypsine, exception faite du Lysozyme C pour lequel les résultats sont légèrement meilleurs avec la chymotrypsine (73%).
- Dans tous les cas, l'AspN donne de moins bons résultats bien que l'acide aspartique soit bien représenté dans les 3 séquences protéiques. L'efficacité de protéolyse de cette enzyme pourrait être en cause. Pour vérifier cette hypothèse, des expériences complémentaires devront être effectuées en augmentant la concentration d'enzyme.





Expériences	1	2	3	4	5	6	Moyenne	Ecart type
BSA TRY	57	66	55	59	53	56	58	4,5
BSA CHYMO	39	\boxtimes	49	50	23	48	42	11,4
BSA ASPN	16	21	Х	16			18	2,9
BSA TRY+CHYMO	55	50	29	Х	45	Х	45	11,3
BSA TRY+ASPN	52	49	42	41			46	5,4
BSA CHYMO+ASPN	47	60	52	55			53	5,4
ADH TRY	60	58	58	62	53	56	58	3,1
ADH CHYMO	39	6	Х	44	50	66	52	11,3
ADH ASPN	18	17	27	Х			21	5,5
ADH TRY+CHYMO	51	64	36	57	56	34	50	12,1
ADH TRY+ASPN	53	40	63	48			51	9,6
ADH CHYMO+ASPN	60	67	51	57			59	6,7
LYSO TRY	55	76	76	55	76	55	65	11,5
LYSO CHYMO	69	70	72	77	79	70	73	4,2
LYSO ASPN	43	\ge	Х	46			44	2,1
LYSO TRY+CHYMO	66	58	66	51	53	67	60	7,1
LYSO TRY+ASPN	44	\succ	75	43			54	18,2
LYSO CHYMO+ASPN	73	85	73	80			78	5,9

Figure 12 : Pourcentages de recouvrement de séquences en fonction des différentes enzymes de digestion (%)

En combinant *2 enzymes*, les meilleurs pourcentages de recouvrement de séquences sont obtenus lorsque l'AspN est combinée à la chymotrypsine.

Diverses hypothèses peuvent expliquer les bons résultats obtenus pour l'association de la chymotrypsine et de l'AspN :

- Forte spécificité de l'AspN qui génére de grands peptides, peu compatibles avec la spectrométrie de masse, susceptibles d'être coupés par la chymotrypsine
- Faible efficacité de protéolyse de l'AspN (AspN seule, figure 11) qui n'affecte pas la chymotrypsine, qui conserve donc toute son efficacité de protéolyse.
- Protéolyse de l'AspN par la chymotrypsine : la séquence de cette enzyme (AspN) n'étant pas publiée, aucune conclusion ne peut être tirée. Nous pouvons cependant envisager le fait que même si une petite quantité d'enzyme (AspN) est protéolysée, il en restera toujours suffisamment pour couper la protéine d'intérêt, et en combinant cette protéolyse à celle de la chymotrypsine qui semble conserver toute son efficacité, les résultats sont améliorés par rapport à une enzyme seule.

Par ailleurs, de façon inattendue, la trypsine en combinaison avec une autre enzyme donne de moins bons résultats que lorsqu'elle est seule. Plusieurs hypothèses peuvent être émises concernant ce résultat :

> Grande efficacité de protéolyse de la trypsine qui doit aussi couper l'enzyme qui lui est associée en plus de la protéine d'intérêt, d'où une dispersion de son activité et de moins bons résultats.

- Protéolyse efficace des 2 enzymes, génèrant des peptides de trop petites tailles pour être analysés par spectrométrie de masse MALDI
- Protéolyse de la trypsine par l'enzyme associée d'où une diminution de son activité.

Pour remédier à ce problème, peut-être faudrait-il ajouter l'enzyme en plus forte concentration de sorte de limiter l'impact des protéolyses parallèles (enzyme-enzyme). Afin de mieux visualiser les résultats obtenus pour les 3 protéines en fonction des enzymes utilisées, les peptides détectés sont soulignés le long des séquences protéiques (figure 13).

Figure 13 : Recouvrement de séquence de la BSA, de l'ADH et du lysozyme C. La trypsine, l'AspN et la chymotrypsine sont utilisées en parallèle ou combinées 2 à 2.

BSA (digestions en ¹ MKWVTFISLL	parallèle) ¹¹ LLFSSAYSRG	²¹ VFRRD <u>THKSE</u>	³¹ IAHRFKDLGE	⁴¹ EHFKGLVLIA
⁵¹ FSQYLQQCPF	⁶¹ DEHVKLVNEL	⁷¹ TEFAKTCVAD	⁸⁰ ESHAGCEKSL	⁹¹ HTLFGDELCK
¹⁰¹ VASLRETYGD	¹¹¹ MADCCEKQEP	¹²¹ ERNECFLSHK	¹³¹ DDSPDLPKLK	¹⁴¹ PDPNTLCDEF
¹⁵¹ KADEKKFWGF	K ¹⁶¹ YLYEIARRHP	¹⁷¹ YFYAPELLYY	¹⁸¹ ANKYNGVFQE	¹⁹¹ CCQAEDKGAC
²⁰¹ LLPKIETMRE	²¹¹ KVLASSARQR	²²¹ LRCASIQKFG	²³¹ ERALKAWSVA	²⁴¹ RLSQKFPKAE
²⁵¹ <u>FVEVTK</u> LVTD	²⁶¹ LTKVHKECCH	²⁷¹ GDLLECADDR	²⁸¹ ADLAKYICDN	291QDTISSKLKE
³⁰¹ CCDKPLLEKS	³¹¹ HCIAEVEKDA	³²¹ IPENLPPLTA	³³¹ DFAEDKDVCK	³⁴¹ NYQEAKDAFL
³⁵¹ GSFLYEYSRR	³⁶¹ HPEYAVSVLL	³⁷¹ RLAKEYEATL	³⁸¹ EECCAKDDPH	³⁹¹ ACYSTVFDKL
401 KHLVDEPQNL	⁴¹¹ IKQNCDQFEK	421LGEYGFQNAL	⁴³¹ IVRYTRKVPQ	441VSTPTLVEVS
⁴⁵¹ RSLGKVGTRC	⁴⁶¹ CTKPESERMP	471 CTEDYLSLIL	⁴⁸¹ NRLCVLHEKT	⁴⁹¹ PVSEKVTKCC
501 TESLVNRRPC	⁵¹¹ FSALTPDETY	⁵²¹ VPKAFDEKLF	531TFHADICTLP	⁵⁴¹ DTEKQIKKQT
⁵⁵¹ ALVELLKHKP	⁵⁶¹ KATEEQLKTV	⁵⁷¹ MENFVAFVDK	⁵⁸¹ CCAADDKEAC	C ⁵⁹¹ FAVEGPKLVV
601STQTALA				

trypsine ou chymotrypsine ou Asp-N

BSA (digestions combinées)

	MKWVTFISLL	¹¹ LLFSSAYSRG	²¹ VFRRDTHKSE	³¹ IAHRFKDLGE	⁴¹ EHFKGLVLIA
5	FSQYLQQCPF	⁶¹ DEHVKLVNEL	⁷¹ TEFAKTCVAD	⁸⁰ ESHAGCEKSL	⁹¹ HTLFGDELCK
10	¹ VASLRETYGD	¹¹¹ MADCCEKQEP	¹²¹ ERNECFLSHK	¹³¹ DDSPDLPKLK	¹⁴¹ PDPNTLCDEF
15	¹ KADEKKFWGK	X ¹⁶¹ YLYEIARRHP	¹⁷¹ YFYAPELLYY	¹⁸¹ ANKYNGVFQE	¹⁹¹ CCQAEDKGAC
20	¹ LLPKIETMRE	²¹¹ KVLASSARQR	²²¹ LRCASIQKFG	²³¹ ERALKAWSVA	²⁴¹ RLSQKFPKAE
25	¹ FVEVTKLVTD	²⁶¹ LTKVHKECCH	²⁷¹ GDLLECADDR	²⁸¹ ADLAKYICDN	²⁹¹ QDTISSKLKE
30	¹ CCDKPLLEKS	³¹¹ HCIAEVEKDA	³²¹ IPENLPPLTA	³³¹ DFAEDKDVCK	³⁴¹ NYQEAKDAFL
35	¹ GSFLYEYSRR	³⁶¹ HPEYAVSVLL	³⁷¹ RLAKEYEATL	³⁸¹ EECCAKDDPH	³⁹¹ ACYSTVFDKL
40	¹ KHLVDEPQNL	⁴¹¹ IKQNCDQFEK	⁴²¹ LGEYGFQNAL	⁴³¹ IVRYTRKVPQ	⁴⁴¹ VSTPTLVEVS
45	¹ RSLGKVGTRC	⁴⁶¹ CTKPESERMP	471CTEDYLSLIL	⁴⁸¹ NRLCVLHEKT	⁴⁹¹ PVSEKVTKCC
50	¹ TESLVNRRPC	⁵¹¹ FSALTPDETY	⁵²¹ VPKAFDEKLF	⁵³¹ TFHADICTLP	⁵⁴¹ DTEKQIKKQT
55	¹ ALVELLKHKP	⁵⁶¹ KATEEQLKTV	⁵⁷¹ MENFVAFVDK	581CCAADDKEAC	⁵⁹¹ FAVEGPKLVV
60	¹ STQTALA				

trypsine+chymotrypsine trypsine+Asp-N chymotrypsine+Asp-N

ADH (digestions en parallèle)

¹SIPETQKGVI ¹¹FYESHGKLEH ²¹KDIPVPKPKA ³¹NELLINVKYS ⁴¹GVCHTDLHAW ⁵¹HGDWPLPVKL ⁶¹PLVGGHEGAG ⁷¹VVVGMGENVK ⁸¹GWKIGDYAGI ⁹¹KWLNGSCMAC ¹⁰¹EYCELGNESN ¹¹¹CPHADLSGYT ¹²¹HDGSFQQYAT ¹³¹ADAVQAAHIP ¹⁴¹QGTDLAQVAP ¹⁵¹ILCAGITVYK ¹⁶¹ALKSANLMAG ¹⁷¹HWVAISGAAG ¹⁸¹GLGSLAVQYA ¹⁹¹KAMGYRVLGI ²⁰¹DGGEGKEELF ²¹¹RSIGGEVFID ²²¹FTKEKDIVGA ²³¹VLKATDGGAH ²⁴¹GVINVSVSEA ²⁵¹AIEASTRYVR ²⁶¹ANGTTVLVGM ²⁷¹PAGAKCCSD ²⁸¹FNQVVKSISI ²⁹¹VGSYVGNRAD ³⁰¹TREALDFFAR ³¹¹GLVKSPIKVV ³²¹GLSTLPEIYE ³³¹KMEKGQIVGR ³⁴¹YVVDTSK

trypsine ou chymotrypsine ou Asp-N

ADH (digestions combinées)

¹ SIPETQKGVI ¹¹ FYESHGKLEH ²¹ KDIPVPKPKA ³¹ NELLINVKYS ⁴¹ GVCHTDLHAW
⁵¹ HGDWPLPVKL ⁶¹ PLVGGHEGAG ⁷¹ VVVGMGENVK ⁸¹ GWKIGDYAGI ⁹¹ KWLNGSCMAC
¹⁰¹ EYCELGNESN ¹¹¹ CPHADLSGYT ¹²¹ HDGSFQQYAT ¹³¹ ADAVQAAHIP ¹⁴¹ QGTDLAQVAP
¹⁵¹ ILCAGITVYK ¹⁶¹ ALKSANLMAG ¹⁷¹ HWVAISGAAG ¹⁸¹ GLGSLAVQYA ¹⁹¹ KAMGYRVLGI
²⁰¹ DGGEGKEELF ²¹¹ RSIGGEVFID ²²¹ FTKEKDIVGA ²³¹ VLKATDGGAH ²⁴¹ GVINVSVSEA
²⁵¹ <u>AIEASTRYVR</u> ²⁶¹ ANGTTVLVGM ²⁷¹ PAGAKCCSD ²⁸¹ FNQVVKSISI ²⁹¹ VGSYVGNRAD
³⁰¹ TREALDFFAR ³¹¹ GLVKSPIKVV ³²¹ GLSTLPEIYE ³³¹ KMEKGQIVGR ³⁴¹ YVVDTSK
trypsine+chymotrypsine trypsine+Asp-N chymotrypsine+Asp-N

Lysozyme C (digestions en parallèle)

¹ MRSLLILVLC	¹¹ FLPLAALGKV	²¹ FGRCELAAAM	³¹ KRHGLDNYRG	⁴¹ YSLGNWVCAA
⁵¹ KFESNFNTQA	⁶¹ TNRNTDGSTD	⁷¹ YGILQINSRW	⁸¹ WCNDGRTPGS	⁹¹ RNLCNIPCSA
¹⁰¹ LLSSDITASV	¹¹¹ NCAKKIVSDG	¹²¹ NGMNAWVAW	^{IR¹³¹NR<u>CKGTDVQA</u>}	¹⁴¹ WIRGCRL
trypsine ou chy	motrypsine ou As	p-N		
Lysozyme C (dige	estions combinées)			

¹ MRSLLILVLC	¹¹ FLPLAALGKV	²¹ FGRCELAAAM	³¹ KRHGLDNYRG	⁴¹ YSLGNWVCAA
⁵¹ KFESNFNTQA	⁶¹ TNRNTDGSTD	⁷¹ YGILQINSRW	⁸¹ WCNDGRTPGS	91RNLCNIPCSA
¹⁰¹ LLSSDITASV	¹¹¹ NCAKKIVSDG	¹²¹ NGMNAWVAWR	R ¹³¹ NR <u>CKGTDVQA</u>	¹⁴¹ WIRGCRL

trypsine+chymotrypsine trypsine+Asp-N chymotrypsine+Asp-N

Enfin, pour savoir s'il existe un intérêt à digérer successivement un échantillon avec 2 enzymes différentes, les taux de recouvrement de séquence obtenus pour la trypsine seule ont été cumulés à ceux obtenus pour une autre enzyme seule (AspN ou chymotrypsine) (figure10). Les résultats sont dans tous les cas améliorés, lors du cumul de 2 enzymes (Tableau 6). Par contre, le cumul des 3 enzymes ne permet plus d'améliorer significativement les pourcentages de recouvrement de séquence. Il y a logiquement des peptides différents qui sont générés en fonction des enzymes utilisées. Mais de façon plus surprenante, les pourcentages de recouvrement de séquence des résultats cumulés sont meilleurs que ceux obtenus lors d'une digestion combinée (réalisée avec les 2 enzymes simultanément).

Les raisons expliquant les moins bons résultats obtenus pour la digestion par 2 enzymes combinées sont multiples :

- diminution de l'efficacité de protéolyse (protéolyse parallèle enzyme-enzyme),
- genèse de peptides de trop petite taille qui ne seront plus visibles en MALDI-MS.

Figure 14 : Comparaison des résultats de la BSA pour les digestions en parallèle, combinées ou cumulées.



Tableau 6 : Comparaison des résultats de la BSA pour les digestions en parallèle ou combinées.

Enzymes	BSA	ADH	Lysozyme C
Trypsine + Chymotrypsine (combiné)	50%	57%	58%
Trypsine + Chymotrypsine (cumulé)	72%	69%	95%
Trypsine + Asp-N	49%	56%	44%
Trypsine + Asp-N (cumulé)	62%	68%	84%
Chymotrypsine + Asp-N	55%	60%	73%
Chymotrypsine + Asp-N (cumulé)	59%	63%	92%

Il pourra donc être intéressant, dans certains cas, de réaliser en parallèle 2 expériences avec 2 enzymes différentes plutôt que d'ajouter simultanément les 2 enzymes dans le même échantillon. Cette approche ne sera possible uniquement si la quantité d'échantillons est suffisante et si le but est de caractériser la protéine avec le plus grand pourcentage de recouvrement de séquence.

Conclusions générales

Les tests d'extraction des protéines ont permis de déterminer les conditions optimales pour extraire les peptides trypsiques des gels, à savoir, 100% d'acétonitrile pendant une heure. Par ailleurs, l'utilisation de plusieurs enzymes de protéolyse (trypsine, chymotrypsine et AspN) permet d'obtenir des résultats complémentaires aboutissant à une amélioration du pourcentage de recouvrement de séquence protéique, en particulier, si les résultats de 2 digestions réalisées en parallèle sont cumulées (trypsine + chymotrypsine). En effet, dans ce cas, les problèmes de protéolyse d'une enzyme par l'autre sont évités et la taille des peptides générés reste compatible avec l'analyse par spectrométrie de masse (<3000Da). Cette approche (utilisation de plusieurs enzymes) pourra être envisagée si le but de l'analyse protéomique est de caractériser l'ensemble d'une séquence protéique et si l'on dispose surtout de suffisamment de matériel.
Chapitre 3:

Evaluation des approches MS/MS en analyse protéomique

La spectrométrie de masse en tandem (MS/MS) est une étape clé de l'analyse protéomique. En effet, elle permet de passer d'une identification des protéines à leur caractérisation, en donnant des informations de séquences, des indications sur les modifications post-traductionnelles ou chimiques ou des informations sur les mutations de la séquence peptidique. Dans ce contexte, différents instruments MS/MS ont été optimisés pour répondre aux problèmes de caractérisation des protéines.

Les trois instruments rencontrés majoritairement pour l'analyse protéomique sont:

- ➢ le Q-TOF (couplé à une nanoLC)
- la trappe ionique (couplée à une nanoLC)
- ▶ le MALDI-TOF/TOF.

Ainsi, ces trois approches MS/MS ont été testées, afin de déterminer les avantages et les inconvénients de chacune d'elles dans le cadre de l'analyse protéomique.

Ces tests ont été réalisés :

- sur des protéines modèles (BSA, Ovalbumine..)
- sur des protéines de cellules Hela (cf. chapitre 1, tests de colorations)
- sur des protéines issues de parois bactériennes de *Streptococcus bovis* (cf. partie 3 applications biologiques).

1 Méthodologie

Pour ces études, chaque échantillon est divisé en trois après l'étape d'extraction. Chacune des fractions est analysée en parallèle par nanoLC Q-TOF et nanoLC Ion Trap avec les mêmes conditions expérimentales (conditions HPLC, paramètres d'acquisition et de retraitement MS), ainsi que par MALDI-TOF/TOF. Les résultats obtenus sont ensuite comparés. Tout d'abord, on vérifie que la nature de la protéine (identification) est correcte. Ensuite, on compare les peptides sélectionnés et séquencés en fonction des instruments ainsi que les fragments observés (y, b, d et w).

Figure 15 : Protocole d'évaluation des différentes techniques MS/MS en analyse protéomique



2 Comparaison de recouvrement de séquences des protéines identifiées par nanoLC-MS/MS

Concernant les protéines de cellules Hela, les taux de recouvrement de séquences vont de 17% à 55%, avec 5 à 20 peptides séquencés par MS/MS. La Nucléophosmine a par exemple été identifiée par les deux techniques nanoLC-MS/MS avec un pourcentage de recouvrement de séquences de 30% et 8 peptides entièrement ou partiellement séquencés, quelque soit l'instrument utilisé (Tableau 7). Ainsi, concernant l'identification de protéines, les résultats sont comparables pour les 2 instruments.

Tableau 7 : Comparaison des couvertures de séquences obtenues par nanoLC-MS/MS sur Q-TOF et Trappe

 ionique pour 4 protéines différentes

protéine	instrument	% sequence coverage	peptides
Nucleophosmin			
silver staining	Q TOF	30%	8
	Trap	30%	8
ruthenium staining	Q TOF	20%	7
	Trap	23%	7
Peroxiredoxin 2			
silver staining	Q TOF	40%	10
	Trap	42%	10
ruthenium staining	Q TOF	40%	9
	Trap	55%	9
RAN Protein			
silver staining	Q TOF	26%	8
	Trap	22%	5
ruthenium staining	Q TOF	26%	8
	Trap	17%	5
α enolase			
silver staining	Q TOF	44%	17
	Trap	47%	17
ruthenium staining	Q TOF	48%	20
	Trap	43%	18

De la même façon, les résultats obtenus lors d'une analyse protéomique (*Streptococcus bovis*, chapitre 3) sont pour la majeure partie comparables pour les 2 instruments (Tableau 8). Les faibles pourcentages de recouvrement de séquences observés sont attribués à la nature de l'échantillon bactérien étudié (chapitre 3) qui contient de nombreux fragments protéiques et non des protéines entières.

Tableau 8 : Comparaison des couvertures de séquences obtenues par nanoLC-MS/MS sur Q-TOF (indiqué par*) et Trappe ionique pour des protéines de *Streptococcus bovis*

spot	identification	% recouvrement	peptides
S300 30*	putative NADP specific glutamate dehydrogenase	29%	11
S300 30	putative NADP specific glutamate dehydrogenase	32%	13
S300 31*	putative NADP specific glutamate dehydrogenase	34%	14
S300 31	putative NADP specific glutamate dehydrogenase	7%	3
S300 42*	putative aminopeptidase C (fgt int)	4%	2
S300 42	aminopeptidase C	3%	2
	fructose bi phosphate aldolase	3%	3
S300 50*	aminopeptidase C (Ct)	3%	1
S300 50	putative NADP specific glutamate dehydrogenase (Nt)	9%	4
S300 51*	putative NADP specific glutamate dehydrogenase (Nt)	26%	11
S300 51	putative NADP specific glutamate dehydrogenase (Nt)	17%	8
S300 53*	aucun		
S300 53	aucun		
S300 55*	groEL (Ct)	8%	4
S300 55	groEL (Ct)	3%	2
	Dpr	6%	1
S300 56*	peroxide resistance protein Dpr	6%	1
S300 56	peroxide resistance protein Dpr	6%	1
S300 57*	peroxide resistance protein Dpr	6%	1
S300 57	peroxide resistance protein Dpr	6%	1
S300 58*	Glucose 6 P isomerase (Ct)	4%	2
S300 58	Glucose 6 P isomerase (fgt int)	2%	1
S300 60*	groEL (Ct)	7%	4
	class II aldolase (Ct)	9%	2
S300 60	groEL (Ct)	8%	3
	Dpr (Nt)	6%	1
S300 61*	enolase (Nt)	6%	2
	class II aldolase (Ct)	4%	1
	Glucose 6 P isomerase (fgt int)	4%	2
	Dpr (Nt)	8%	2
	L lactate dehydrogenase (Ct)	8%	2
S300 61	Dpr (Nt)	6%	1
	L lactate dehydrogenase (Ct)	3%	1
S300 66*	class II aldolase (Ct)	15%	4
	enolase (fgt int)	3%	1
	glyceraldehyde 3 P dehydrogenase (Nt)	13%	4
S300 66	glyceraldehyde 3 P dehydrogenase (Nt)	5%	2
S300 68*	putative NADP specific glutamate dehydrogenase (Nt)	19%	7
S300 68	putative NADP specific glutamate dehydrogenase (Nt)	19%	7
S300 69*	glyceraldehyde 3 P dehydrogenase (Nt)	19%	7
	class II aldolase (Ct)??	12%	4
S300 69	glyceraldehyde 3 P dehydrogenase (Nt)	8%	3
S300 70*	glyceraldehyde 3 P dehydrogenase (Nt)	18%	7
	pyruvate kinase (Nt)	5%	3
S300 70	glyceraldehyde 3 P dehydrogenase (Nt)	19%	7

Les résultats ci-dessus peuvent être illustrés par l'histogramme suivant (figure 16) qui reflète la quasisimilarité des résultats obtenus par les deux techniques **Figure 16**: Histogramme illustrant la comparaison des pourcentages de recouvrement de séquences obtenus par NanoLC MS/MS sur Q-TOF et trappe ionique pour les protéines de *Streptococcus bovis*



Q-TOF / ION TRAP

3 Comparaison de la nature des peptides et de leurs fragments obtenus par nanoLC-MS/MS

Bien que les pourcentages de recouvrement de séquences soient très proches pour les deux techniques de nanoLC-MS/MS (Q-TOF et trappe ionique), la liste des peptides séquencés n'est pas toujours identique. De ce fait, si on combine les résultats obtenus avec les 2 instruments, le pourcentage de recouvrement de séquence peut, dans certains cas, être amélioré. Par exemple, pour l' α Enolase, il passe de 43% (Q-TOF) et 48% (Trappe ionique) à 61% (Q TOF + Trappe ionique) (figure 17).

Figure 17: Résultats combinés de l'étude nanoLC-MS/MS (Q-TOF + trappe ionique) pour l'α Enolase, la nucléophosmine, la GTP binding protéine et la peroxyrédoxine 2

Γ

Alpha enolas	Alpha enolase (coloration CH)														
Trap, QTOF, Trap+QTOF															
SILKIHAREI	FDSRGNPTVE	VDLFTSKGLF	RAAVPSGAST	GIYEALELRD	NDK TRYMGKG	VSKAVEHINK	TIAPALVSKK								
LNVTEQEKID	KLMIEMDGTE	NK SKFGANAI	LGVSLAVCKA	GAVEKGVPLY	RHIADLAGNS	EVILPVPAFN	VINGGSHAGN								
KLAMQEFMIL	PVGAANFR EA	MRIGAEVYHN	LKNVIKEKYG	KDATNVGDEG	GFAPNILENK	EGLELLK TAI	GKAGYTDK <mark>VV</mark>								
IGMDVAASEF	FRSGKYDLDF	K SPDDPSR YI	SPDQLADLYK	SFIKDYPVVS	IEDPFDQDDW	GAWQKFTASA	GIQVVGDDLT								
VTNPKRIAKA	VNEKSCNCLL	LKVNQIGSVT	ESLQACKLAQ	ANGWGVMVSH	RS GETEDTFI	ADLVVGLCTG									
IKTGAPCRS	ERLAKYNQLL	RIEEELGSKA	KFAGRNFRNP	LAK											
61% reco. Global (trap 44% +QTOF 47%)															

```
Nucleophosmin (CH)

Trap, QTOF, Trap+QTOF

MEDSMDMDMS PLRPQNYLFG CELKADKDYH FKVDNDENEH QLSLRTVSLG AGAKDELHIV EAEAMNYEGS PIKVTLATLK

MSVQPTVSLG GFEITPPVVL RLKCGSGPVH ISGQHLVAVE EDAESEDEEE EDVKLLSISG KRSAPGGGSK VPQKKVKLAA

DEDDDDDDDEE DDDEDDDDDD FDDEEAEEKA PVKKSIRDTP AKNAQKSNQN GKDSKPSSTP RSKGQESFKK QEKTPKTPKG

PSSVEDIKAK MQASIEKGGS LPKVEAKFIN YVKNCFRMTD QEAIQDLWQW RKSL
```

30% reco. Global (trap 30% +QTOF 30%)

<u>,</u>
?

AFQYTDEHGE VCPAGWKPGS DTIKPNVDDS KEYFSKHN

43% reco. Global (trap 42% +QTOF 40%)

Les peptides préférentiellement isolés et fragmentés par la trappe ionique ou le Q-TOF n'ont pas de caractéristiques physico-chimiques particulières (hydrophobicité, acidité..).

Par contre, il est possible, que les différences observées concernant les peptides sélectionnés soient dues à la fragmentation en source des peptides qui est différente d'un appareil à l'autre. En effet, ces 2 instruments (Q-TOF et Trappe ionique) ont des sources, mais surtout des interfaces différentes qui favoriseront ou non la fragmentation de certains peptides. Or, si un peptide fragmente en source, il sera moins intense. Par conséquent, les 3 ions les plus intenses étant sélectionnés pour chaque spectre MS, un ion pourra être préférentiellement sélectionné avec un instrument et peut-être ne pas l'être avec un autre, si sa fragmentation en source a été plus importante, d'où des différences d'information en fonction des instruments. Par ailleurs, les processus de désolvatation des ions sont également différents entre les 2 appareils. Ainsi, si un ion est mieux désolvaté, il gagnera en intensité et donc aura plus de chance d'être sélectionné pour être analysé par MS/MS.

De plus, dans notre cas, seuls les ions 2+ et 3+ étaient pris en compte comme ions précurseurs pour être fragmentés. Mais, les ions 3+, isolés et fragmentés dans la trappe ionique, donnent des spectres difficilement interprétables en raison d'une moins bonne résolution de cet analyseur par rapport à un TOF. Pour améliorer la résolution [Heller, 2003], la rampe de tensions appliquée sur les éléctrodes de la trappe devrait être plus lente. Mais, à l'opposé, si le balayage est ralenti, le nombre d'ions précurseurs sélectionnés pour un même laps de temps sera plus faible. Si les informations de séquences sont de meilleures qualités, elles seront moins nombreuses et certains peptides pourront ne pas être sélectionnés. De ce fait, certaines protéines pourront ne pas être détectées. Il est donc important de trouver un compromis entre une bonne résolution et l'acquisition d'un maximum d'informations de séquences. Ce point est sans cesse en amélioration par les constructeurs. Ainsi, la société BRUKER a développé une nouvelle trappe ionique qui possède une meilleure résolution, permettant de distinguer efficacement des ions 3 +. Mais dans notre cas, les spectres MS/MS obtenus sur la trappe ionique ont été moins facilement interprétables pour les ions 3 +, impliquant une différence entre les instruments en ce qui concerne les ions précurseurs génèrant des information de séquences interprétables.

Par ailleurs, le Q-TOF et la trappe ionique étant équipés de 2 analyseurs fondamentalement différents, la comparaison des fragments générés (y et b) dans le cadre de ces approches, était intéressante. Les résultats présentés dans le tableau 9 montrent que les fragments générés sont comparables quelque soient les instruments (Q-TOF et trappe ionique)

Tableau 9: Fragments y et b observés en fonction des analyseurs.pour différents peptides des protéines :
peroxiredoxine 2 (pdx2), GTP binding protein (RAN) et nucleophosmine

6 PDX2																		
1211,7	QITVNDLPVGR	y1	y2	y3	y4	y5	y6	у7	y8	y9	y10	y11						
	Fichier 9849 QTOF		Х		Х	Х	Х	Х	Х	Х								
	9857 QTOF		Х		Х	Χ	Х	Х	Х	Х								
	7549 TRAP				X	Χ	X	X		X								
	7553 TRAP		X		X	X	X	X	X	X								
	QITVNDLPVGR	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11						
	9849 QTOF		Х	Х														
	9857 QTOF		Х	Х														
	7549 TRAP			X														
	7553 TRAP				X													
1735	EGGLGPLNIPLLADVTR	y1	y2	y3	y4	y5	y6	у7	y8	y9	y10	y11	y12	y13	y14	y15	y16 y17	
	9857 QTOF			Х	Х				Х	Х	Х		Х	Х	Х	Х		
	7549 TRAP			X	X	X	X	X	X	X	X		X	X				
	7553 TRAP			X	X	X	X		X	X	X	X	X	X				
	EGGLGPLNIPLLADVTR	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15	b16	
	9857 QTOF					Х												
	7549 TRAP							X		X		X	X		X	X	X	
	7553 TRAP									X		X	X	X	X	X	X	
971,5	IGKPAPDFK	y1	y2	y3	y4	y5	y6	у7	y8	y9								
	9849 QTOF	Х	Х		Х	Х	Х	Х	Х									
	7549 TRAP		Χ		X		X	X	X									
	7553 TRAP		Χ		X	X	X	X										
	IGKPAPDFK	b1	b2	b3	b4	b5	b6	b7	b8	b9								
	9849 QTOF			Х				Х										
	7549 TRAP				X	X		X	X									
	7553 TRAP					X		X										

978,6	ATAVVDGAFK	y1	y2	y3	y4	y5	y6	y7	y8	y9 y10
	9849 QTOF				Х	Χ	Х	Х	Χ	Х
	9857 QTOF					Х	Х	Х	Х	
	9861 QTOF					Х	Х	Х	Х	
	7549 TRAP				X	X	X	X	X	
	7553 TRAP		X		X	X	X	X		
	ATAVVDGAFK	b1	b2	b3	b4	b5	b6	b7	b8	b9 b10
	9849 QTOF		Х	Х	Х				Χ	
	9857 QTOF		Х	Х	Х				Х	
	9861 QTOF		Х	Х	Х				Х	
	7549 TRAP					X			X	X
	7553 TRAP								Χ	X

7 RAN																			
2052,1	YVATLGVEVHPLVFHTNR	y1	y2	y3	y4	y5	y6	у7	y8	y9	y10	y11	y12	y13	y14	y15	y16	y17	y18
	ABSENT																		
	7554 TRAP				X			X	X	X	X	X		X					
	YVATLGVEVHPLVFHTNR	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15	b16	b17	b18
	ABSENT																		
	7554 TRAP										X			X	Х			Х	
1015,4	LVLVGDGGTGK	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11							
	9854 QTOF					Χ	Х	Χ	Х	Х									
	9858 QTOF	Х			Х	Х	Х	Χ	Х	Х	Χ								
	7547 TRAP				X	X	Χ	X	X	X									
	7554 TRAP					X		X	X	X									
	7559 TRAP				X	X		X	X	X	X								
	LVLVGDGGTGK	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11							
	9854 QTOF		Х	Х	Х	Х													
	9858 QTOF		Х	Х	Х	Х													
	7547 TRAP																		
	7554 TRAP			X					X	X									
	7559 TRAP									X	X								
960,5	HLTGEFEK	y1	y2	y3	y4	y5	y6	y7	y8										
	9854 QTOF	Х		Х		Х	Х	Х											
	9862 QTOF	Х		Х		Х	Х	Х											
	7554 TRAP			X		X	X	X											
	7559 TRAP			X		X	X	X											
	HLTGEFEK	b1	b2	b3	b4	b5	b6	b7	b8										
	9854 QTOF		Х	Х		Х													
	9862 QTOF		Х	Х		Х													
	7554 TRAP					X	X	X											
	7559 TRAP			Х		Х	Х	Χ											

549,2	FGGLR	y1 y2 y3 y4 y5
	ABSENT	
	7547 TRAP	X X
	7554 TRAP	ХХ
	7559 TRAP	X
	FGGLR	b1 b2 b3 b4 b5
	ABSENT	
	7547 TRAP	ххх
	7554 TRAP	ххх
	7559 TRAP	ххх

nucleophosmine															
717,4	LLSISGK	y1	y2	y3	y4	y5	y6	y7							
	9856 QTOF	Х	Х	Χ	Х	Х	Χ								
	7557 TRAP			X	X	X	X								
	LLSISGK	b1	b2	b3	b4	b5	b6	b7							
	9856 QTOF		Χ	Χ											
	7557 TRAP														
745,46	VTLATLK	y1	y2	уЗ	y4	у5	y6	у7							
	9856 QTOF	Х	Χ	Χ	Χ	Χ	Χ								
	7557 TRAP			X	X	X									
	VTLATLK	b1	b2	b3	b4	b5	b6	b7							
	9856 QTOF		Χ												
	7557 TRAP				X		X								
931,4	GPSSVEDIK	y1	y2	уЗ	y4	у5	y6	у7	y8	y9					
	9856 QTOF	Х	Χ	Х	Х	Χ	Х	Х	Х						
	7557 TRAP					X	X	X							
	GPSSVEDIK	b1	b2	b3	b4	b5	b6	b7	b8	b9					
	9856 QTOF		Х	Х	Х			Х							
	7557 TRAP						X	X	X						
1129,5	GPSSVEDIKAK	y1	y2	уЗ	y4	у5	y6	у7	y8	у9	y10	y11			
	9856 QTOF	Х		Х	Х	Х	Х	Х	Х	Х					
	7557 TRAP				X	X			X	X	X				
	GPSSVEDIKAK	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11			
	9856 QTOF		Х	Х	Х	Х	Х	Х	Х						
	7557 TRAP					X	X	X	X	X					
1567,8	VDNDENEHQLSLR	y1	y2	уЗ	y4	у5	y6	у7	y8	у9	y10	y11	y12	y13	
	9856 QTOF			Х								Х			
	7557 TRAP			X	X		X		X	X					
	VDNDENEHQLSLR	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	
	9856 QTOF														
	7557 TRAP									X	X				

Ces résultats étant comparables quelque soit l'analyseur, ils peuvent s'expliquer par le fait que, les fragmentations générées dans la trappe ionique et le Q-TOF sont de basses énergies. Ainsi, les ions « y et b » sont majoritairement observés puisque l'énergie de fragmentation est faible et insuffisante pour casser les liaisons entre le carbone α et le carbonyl. De plus, même si le processus de fragmentation est complexe et pas encore entièrement compris [Tabb, 2003], les interactions entre les chaînes latérales des acides aminés et la séquence peptidique favorise les fragmentations, comme la localisation des charges le long de la chaîne peptidique (en modifiant le caractère électrophile et nucléophile des sites) qui détermine la partie du peptide qui sera détectée par spectrométrie de masse (charge du coté C-terminal, détection des fragments y, charge du coté N terminal, détection des fragments b). Ainsi, la fragmentation des peptides dépend majoritairement de la localisation des charges qui est fonction de la nature des peptides et pas de l'instrument utilisé, ce qui peut expliquer la présence de fragments (y et b) identiques malgré les différents analyseurs.

Il faut cependant noter qu'aucune différence n'est observée concernant la fragmentation des peptides entre ces 2 instruments car nous nous sommes uniquement intéressés à l'étude de peptides trypsiques. En effet, le mode de fragmentation dans une trappe ionique est beaucoup plus doux que dans un Q-TOF. De ce fait, les spectres de fragmentation obtenus par Q-TOF donnent des informations sur toute la gamme de masse, indiquant que la MS² par Q-TOF ne génère pas uniquement des ions fragments issus de l'ion parent. Les spectres obtenus seront donc plus difficiles à interpréter. Au contraire, les fragments obtenus par trappe ionique permettent d'attribuer une filiation directe des ces ions. En effet, à la différence du Q-TOF où une différence de potentiel est appliquée à l'ion parent à l'entrée de la cellule de collision, dans la trappe ionique, l'énergie de fragmentation est délivrée à la fréquence de résonance de l'ion parent, impliquant une fragmentation spécifique. De plus, les ions fragments restants plus longuement dans la trappe ionique, des réarrangement chimiques sont possibles, générant d'autres espèces ce qui compliquera les spectres MS/MS [thèse N. Carte].

4 Etudes complémentaires, MALDI-TOF/TOF

Pour compléter l'étude des fragmentations réalisées par nanoLC-MS/MS, les mêmes analyses ont été effectuées, à titre de comparaison, par MALDI-MS, les résultats concernant les 4 protéines modèles sont présentés dans le tableau 10.

protéine	instrument	% recouvrement de séquence	peptides	erreur ppm
Nucleophosmine				
coloration argent	MALDI-MS	22%	6	11
Peroxiredoxine 2				
coloration argent	MALDI-MS	35%	7	21
RAN Protein				
coloration argent	MALDI-MS	56%	14	24
α enolase				
coloration argent	MALDI-MS	71%	33	19

Tableau 10 : Résultats MALDI-MS pour les protéines de cellules Hela

Pour la Peroxiredoxine 2 (figure 18), en combinant les résultats obtenus par nanoLC-MS et par MALDI-MS, on passe d'un pourcentage de recouvrement de séquence de 40-42 % (nanoLC-MS) et de 35% (MALDI-MS) à 55% de recouvrement de séquence (nanoLC-MS + MALDI-MS)

Figure 18: Résultats combinés pour la Peroxiredoxine 2

```
      Peroxiredoxine 2 (CH)

      Trap, QTOF, Trap+QTOF, MALDI

      MASGNARIGK PAPDFKATAV VDGAFKEVKL SDYKGKYVVL FFYPLDFTFV CPTEIIAFSN RAEDFRKLGC EVLGVSVDSQ

      FTHLAWINTP RKEGGLGPLN IPLLADVTRR LSEDYGVLKT DEGIAYRGLF IIDGKGVLRQ ITVNDLPVG SVDEALRLVQ

      AFQYTDEHGE VCPAGWKPGS DTIKPNVDDS KEYFSKHN

      55% recouvrement Global (trap 42% +QTOF 40%+MALDI 35%)
```

Ces résultats confirment la complémentarité des 2 modes d'ionisation (MALDI et ESI) dans le cadre des études protéomiques. Ainsi, même si les pourcentages de recouvrement de séquence sont généralement du même ordre, quelque soit la technique utilisée, les peptides observés ne sont pas toujours identiques.

Mais, le fait que l'on obtienne des informations complémentaires en fonction des différents instruments utilisés (ions précurseurs différents) est particulièrement intéressant dans le cadre de la caractérisation de protéines puisqu'un plus grand nombre de peptides sera séquencé par MS/MS avec les trois instruments rendant possible la détection de modifications post-traductionnelles ou chimiques ainsi que des mutations d'acides aminés.

Par ailleurs, les tests de fragmentation MALDI-TOF/TOF ont été réalisés sur les mêmes protéines qu'en nanoLC-MS/MS, avec et sans gaz de collision (Argon) dans la cellule de collision. Les résultats ont été comparés à ceux obtenus par nanoLC-MS/MS. Pour l'ion 1211.7 de la Peroxiredoxine 2, les fragments observés par MALDI-TOF/TOF sont comparables à ceux obtenues par nanoLC-MS/MS (tableau 11). L'ajout de gaz dans la cellule de collision permet en plus de lever les ambiguités liées aux acides aminés isobares (figure 19), par exemple pour distinguer la leucine de l'isoleucine dans la séquence QITVND L/I PVGR. En effet, l'analyse par MALDI-TOF/TOF permet de générer en plus des fragments basses énergies b et y, des fragments hautes énergies de type d et w par perte de la chaîne latérales des acides aminés.

Tableau 11 : Fragments y et b observés pour les 3 instruments de l'ion 1211,7 m/z (peroxiredoxine 2)

1211,7	QITVNDLPVGR	y1	y2	уЗ	y4	y5	y6	y7	y8	y9	y10	y11	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11
	QTOF		X		Х	Х	Х	X	Х	X				Х	X								
	TRAP		X		X	X	X	X	X	X						X							
	MALDI TOF-TOF	Х	X	Χ	X	Χ	Χ							Х	X								



Figure 19: Spectre MS/MS MALDI TOF/TOF ion 1211m/z (peroxiredoxine 2)

Lors de cette étude, seules des séquences peptidiques partielles ont été obtenues par MALDI-MS/MS. En effet, les fragments de haut poids moléculaires sont rarement observés quelques soient les voltages appliqués sur les lentilles et la puissance du laser. En fait, l'appareil (cellule LIFT) a été optimisé afin de visualiser le mieux possible les ions de basses masses, principalement les ions immonium au détriment des ions de haut poids moléculaires.

Ainsi, même si l'approche MALDI-TOF/TOF est une méthode de choix pour l'analyse protéomique, certains peptides de faibles intensités ne peuvent pas être isolés et fragmentés ou donnent des spectres difficilement interprétables mais le pendant existe aussi vis à vis de l'approche nanoLC- MS/MS. En effet, lors des études MS/MS par MALDI, des pertes de sensibilité et de résolution sont observées. La technologie LIFT ne peut atteindre une résolution supérieure à 5000 en mode MS/MS, en raison d'une dispersion en énergie cinétique des ions qui est trop importante. Comme l'énergie cinétique transmise aux ions est moins importante lors d'une étude MS/MS (8kV contre 19kV en MS)pour permettre leur refocalisation, moins d'ions fragments seront observés, c'est pour cela que l'irradiance laser est augmentée afin de générer plus de fragments en source. Cependant, il sera nécessaire de réaliser des dépôts épais avec des quantités d'échantillons plus importantes pour avoir une meilleure sensibilité. De plus, la présence d'acides aminés aromatiques, l'hydrophobicité des peptides, leur taille et leur possibilité à former des structures II stables ont déjà été décrites comme ayant une influence sur l'intensité des ions en MALDI [Krause, 1999]. Des effets de suppression du

signal peuvent aussi être observés pour les mélanges de peptides. De ce fait, la séparation par nanoLC utilisée en amont des instruments à source ESI, limite ces effets et pourrait être envisagée en amont de l'analyse MALDI. Cependant, la rapidité d'analyse, qui est un des avantages majeurs de l'analyse par MALDI serait perdue.

De plus, l'approche MALDI-MS/MS, bien que donnant les mêmes résultats MS/MS (fragments y et b) que l'approche nanoLC MS/MS, est plus délicate à mettre en œuvre. En effet, la sélection des ions précurseurs n'est pas encore automatisée et les effets de suppression limitent l'analyse de certains peptides. Par contre, cette approche permet de différencier les acides aminés isobares (fragments haute énergie, w et d) et de caractériser les protéines. Par conséquent, ces 3 techniques donnent des informations de séquence de bonne qualité qui sont complémentaires en fonction des instruments utilisés.

Conclusions générales

Les tests réalisés sur les différentes approches MS/MS utilisées en analyse protéomique ont permis d'établir la complémentarité de ces trois approches : nanoLC-Q-TOF, nanoLC-Trappe ionique et MALDI-TOF/TOF. En effet, si les protéines sont identifiées avec les mêmes recouvrements de séquence, les peptides séquencés par nanoLC-MS/MS sont parfois différents en fonction de l'instrument utilisé :Trappe ionique ou Q-TOF. Ceci est d'autant plus intéressant que dans le contexte de l'approche MS/MS, Q-TOF et Trappe ionique apparaissent comme complémentaires et permettent en combinant les résultats d'améliorer les pourcentages de recouvrement de séquence.

De plus, si les fragments majeurs (y et b) sont observés de la même façon pour les trois instruments, il est également intéressant d'observer les fragments haute énergie (w et d) par MALDI-TOF/TOF. Cette approche permet de différencier les acides aminés isobares et donc d'améliorer les recouvrements de séquence. Ainsi, en fonction des échantillons et des questions posées (caractérisation de peptides ou identification globale de protéines), il pourra être intéressant de choisir l'une ou l'autre technique. En effet, l'approche MALDI-MS/MS, n'étant pas encore automatisée et nécessitant une plus grande quantité de matériel, sera préférée pour des études ponctuelles de caractérisation fine des protéines (distinction d'acides aminés isobares). Au contraire, l'approche nanoLC-MS/MS, automatisable, sera préférée pour les études par homologie de séquences et les études de séquençage globale (identification de protéines peu ou pas référencées dans les banques de données).

Partie III:

Applications biologiques : De l'identification à la caractérisation des protéines

Chapitre 1 : Introduction : Approches MALDI-MS, nanoLC-MS/MS et séquençage *De Novo*

Chapitre 2 : Approche MALDI-MS : Identification des protéines

Chapitre 3 Approche nanoLC-MS/MS

Chapitre 4 Approche séquençage De Novo

Chapitre 1:

Introduction : Approches MALDI-MS, nanoLC-MS/MS et séquençage *De Novo*

La seconde partie de ce travail de thèse concerne diverses applications biologiques pour lesquelles il a été nécessaire d'utiliser les différentes voies de l'approche protéomique (figure 20), qui peuvent être séparées en deux parties :

Si les protéines étudiées sont issues de génomes connus, 2 approches pourront être choisies :

- L'approche MALDI-MS pour les études de protéomique descriptive (nécessitant une identification rapide des protéines)
- L'approche nanoLC-MS/MS pour les études de protéomique fonctionnelle (caractérisations fines des protéines, déterminations de modifications post-traductionnelles...)

Si les protéines sont issues de génomes inconnus ou partiellement connus :

- L'approche nanoLC-MS/MS, pour les études nécessitant des identifications par homologies de séquences
- L'approche par Séquençage De Novo, pour pouvoir identifier les protéines par alignement de séquences

Figure 20 : Stratégies d'identification de protéines.

stratégie MALDI-MS : comparaison de listes de masses de peptides trypsiques, stratégie nanoLC-MS/MS: comparaison de listes de masses des fragments de peptides trypsiques stratégie séquençage De Novo : alignement de séquences (BLAST)



L'identification des protéines par spectrométrie de masse lors des premières études protéomiques, consistera de façon générale à comparer une liste de masses de peptides trypsiques avec les masses présentes dans les banques de données, en utilisant l'approche MALDI-MS (étude du protéome nucléaire des cellules Caco2, étude protéomique différentielle de mitochondries dans le cadre de la maladie MERRF). Dans les études suivantes, l'approche nanoLC-MS/MS, qui consiste à comparer les masses des fragments de ces peptides et l'approche Séquençage *De Novo* seront utilisées, car elles permettent une caractérisation plus fine des protéines (modifications post-traductionnelles ou chimiques, mutations d'acides aminés...) et peuvent être employées pour l'étude de génomes partiellement connus ou inconnus.

Ces 2 approches ont été utilisées lors de deux études protéomiques (*Streptococcus bovis, Plasmodium falciparum*), car leurs génomes ne sont pas entièrement séquencés (*Streptococcus bovis*) ou ne l'étaient pas au moment de l'étude (*Plasmodium falciparum*).

Pour l'identification de protéines issues d'espèces non séquencées, la stratégie qui consiste à comparer les masses des peptides trypsiques de la protéine avec celles de peptides d'autres organismes présents dans les banques de données (empreinte peptidique massique par homologie), peut être envisagée. Si les organismes, phylogénétiquement proches, présentent une grande proportion de séquences génomiques identiques, et que par conséquent leurs protéines ont un haut degré d'homologie, l'identification par homologie sera donc possible. Par contre, si de nombreuses substitutions nucléotidiques au sein des gènes sont présentes, elles entraînent de nombreuses variations d'acides aminés dans les séquences protéiques. Dans ce cas, l'identification par homologie cèdera la place à l'identification par alignement de séquences.

Trois cas sont donc rencontrés (figure 20):

- Si le degré d'homologie des protéines inter-espèces est important, alors l'empreinte peptidique massique (MALDI-MS) est suffisante pour permettre l'identification. Wilkins suggère que 80% d'homologies entre les protéines sont suffisant pour permettre l'identification [Wilkins et coll., 1997]. Par contre, les peptides présentant une substitution d'un acide aminé, ayant donc des masses peptidiques modifiées, ne participeront pas à l'identification de la protéine.
- Pour les protéines qui ont un faible pourcentage d'homologie avec celles présentes dans les banques de données, l'empreinte peptidique massique échoue bien souvent. Le recours à la MS/MS sur quelques peptides trypsiques, afin d'avoir une liste de masses plus informative (masse des peptides et des fragments peptidiques) permet parfois l'identification inter-espèces [Fenyo, 2000], même si le pourcentage de recouvrement de séquences est faible.
- Lorsque les protéines ont un très faible pourcentage d'identité par rapport à celles présentes dans les banques de données, la comparaison des masses des peptides trypsiques et de leurs fragments ne permet plus d'identifier les protéines. Seules la détermination des séquences des peptides à partir de l'analyse MS/MS (séquençage *De Novo*) et la soumission de ces séquences à une recherche (partie I) par alignement de séquences (BLAST) pourra aboutir à l'identification de la protéine.

Chapitre 2 :

Approche MALDI-MS : Identification des protéines

1 Détermination du protéome de noyaux de cellules Caco 2

Le but de cette étude, réalisée en collaboration avec J.F. Launay de l'Unité INSERM U 381 de Strasbourg, était d'établir la cartographie (protéome) des protéines de noyaux des cellules Caco2/TC7 (lignées cellulaires issues de cellules cancéreuses coliques) au stade prolifératif et de déterminer par une analyse protéomique différentielle les protéines sous et sur-exprimées pour ces cellules en voie de différenciation et différenciées.

1.1 Intérêt biologique

Les lignées intestinales Caco2/ TC7 représentent un bon modèle [Chantret, 1994] pour l'étude *in vitro* de la séquence prolifération/différenciation cellulaire le long de l'axe crypto villositaire intestinal [Zweibaum, 1991] (figure 21).

Ces processus sont influencés par des interactions entre les cellules intestinales (entérocytes) et la lame basale (figure 2), notamment par l'intermédiaire de protéines, les laminines. Jusqu'à présent, seules 4 laminines (1, 2, 5, 10) ont été décrites dans l'intestin [Colognato, 2000]. L'influence des laminines (2, 5 et 10) sur la prolifération cellulaire et de la laminine 1 sur la différenciation [De Arcangelis, 1996] ont déjà été démontrées. Ces processus de prolifération et de différenciation s'accompagnent de nombreux changements cellulaires particulièrement au niveau des noyaux, en le réorganisant et en modulant l'expression de certaines protéines nucléaires. Le but de ce travail a été l'analyse du protéome nucléaire des cellules Caco2/TC7, tout d'abord en phase proliférative

(3 jours de culture) puis en voie de différenciation (j7) et enfin à l'état différencié (j14). Une cartographie complète des protéines de noyaux a été réalisée après 3 jours de culture, les études après 7 et 14 jours sont des analyses de protéomique différentielle par rapport à j3. Pour ces études (j7 et j14), seules les protéines dont l'expression semblait être modulée ont été analysées.

Figure 21 : Axe crypto-villositaire intestinal



1.2 Approche protéomique descriptive

Cette approche consiste à analyser systématiquement tous les spots détectés sur gel 2D par MALDI-MS. Dans le cadre de l'étude des noyaux de cellules Caco2/TC7, le génome était connu et référencé dans les banques de données, donc nous avons pu utiliser cette approche.

Cette étude a fait l'objet d'une publication dans le journal Proteomics.

1.3 Publication

Proteomic analysis of nuclear proteins from proliferative and differentiated human colonic intestinal epithelial cells.

Natacha Turck, <u>Sophie Richert</u>, Patrick Gendry, Jeanne Stutzmann, Michèle Kedinger, Patricia Simon-Assmann, Emmanuelle Leize-Wagner, Alain Van Dorsselaer, Jean-François Launay (Acceptée dans Proteomics, Avril 2003)

Proteomic analysis of nuclear proteins from proliferative and differentiated human colonic intestinal epithelial cells

Natacha Turck¹, Sophie Richert², Patrick Gendry¹, Jeanne Stutzmann¹, Michèle Kedinger¹, Patricia Simon-Assmann¹, Emmanuelle Leize², Alain Van Dorsselaer² and Jean-François Launay¹ ¹INSERM U381, 3, avenue Molière, 67200 Strasbourg

² CNRS-EPCM, UMR7509, LSMBO, 25, rue Becquerel, 67087 Strasbourg

Keywords : Nuclear proteome / Human colonic cancer cells / High resolution two-dimensional gel electrophoresis / MALDI-TOF mass spectrometry

Correspondence : Dr. Jean-François Launay, INSERM U.381, 3 avenue Molière, 67200 Strasbourg, France **E-mail** : <u>Jean-Francois.Launay@inserm.u-strasbg.fr</u>

Fax : 33 (0)3 88 26 35 38

Running title : Nuclear proteome and intestinal cell differentiation

SUMMARY

Self-renewing tissues such as the intestine contain progenitor proliferating cells which subsequently differentiate. Cell proliferation and differentiation of cells involve gene regulation, which takes place in the nucleus. A human intestinal epithelial cell line model (Caco2/TC7), reproducing these dynamic processes, has been used to perform proteomic studies on nuclear proteins. Nuclei from the Caco2/TC7 cells at proliferative and differentiated stages were purified by subcellular fractionation. After two-dimensional gel electrophoresis separation, ruthenium staining and peptide mass fingerprinting by MALDI-MS; among 400 spots detected, 87 corresponding to 61 different proteins were identified in nuclei from proliferative cells. Comparison of nuclear proteomes from proliferative or differentiated cells by differential display highlighted specific protein spots. First, our results revealed that the expression of nucleolin, hnRNPs (A2/B1 and A1) decreased in differentiated cells. Second, Western blotting and immunocytochemistry confirmed that the expression of these nuclear proteins was restricted to the proliferative state of cells. This study of the nuclear proteome as a function of proliferation to differentiation stages, represents the first step towards the establishment of a database which will be a valuable resource in future studies on the differential expression of nuclear proteins

1 Introduction

The completion of the first draft of the human genome will allow genes to be identified directly and their protein products predicted. Therefore, the genome projects provide a basis for many proteomic studies in which the expressed protein profiles are the focus. Proteomics objectives range from describing all proteins localized in a specific cell type or organelle to describing the proteins present and their post-translational modifications during specific developmental and stages, or those found physiological in pathologocal tissues. Proteomics uses а combination of techniques including twodimensional electrophoresis (2-DE). mass spectrometry (MS) and informatic tools [1]The macromolecular architecture of eukaryotic cells can be taken as an advantage to study subcellular proteomes. The nucleus contains the genetic information and is the place of gene expression. The mammalian cell nucleus is a complex but highly organized organelle which contains numerous subcompartments[1], [2]. Thus. determining the nuclear localization of proteins is important for the understanding of genome regulation and function, as well as providing important clues as to the molecular function of novel proteins. Regarding the studies of human nuclear proteomes using a proteomic approach, Jung et al. established a 2D-PAGE preliminary reference map of liver nuclei [3]. However, an analysis of the protein content of nuclei during specific stages of cell proliferation and differentiation has never been investigated. In the present work we have focused on nuclear proteome changes during intestinal epithelial cell proliferation/differentiation. The proliferation and differentiation processes in the intestinal epithelium involve different proteins, e.g., extracellular and signalling molecules, receptors, transcription factors [4], [5]. In vivo, the renewal of intestinal epithelium implies two distinct phases : proliferation, which occurs within crypts, and differentiation which happens during the migration of cells from the crypt to the villus tip. Among the human epithelial cell lines, the colon carcinoma Caco-2 cells [6] are the only ones which differentiate spontaneously as enterocytes when grown in standard culture conditions. Caco-2 cell functional differentiation can be assessed by the targeting of digestive enzymes like sucrase-isomaltase or dipeptidylpeptidase IV (DPP IV) at the apical membrane of cells [7], [8]. Recently, the changes in the transcriptome from undifferentiated to differentiated Caco-2 cells have been investigated; it was found that a high number of genes are down-regulated in the differentiated Caco-2 cells, compensated by few up-regulated genes [9]. Our study was performed using a clonal cell population (TC-7) isolated from late passages of the Caco-2 cell line. This clone essentially exhibits a more rapid growth and a higher differentiation potential, assessed by the expression of sucrase-isomaltase compared to the parental cell line [10]. The present study was designed to explore the

proteome from proliferative nuclear or differentiated intestinal epithelial cells and to initiate the construction of 2-DE maps of human intestinal epithelial cells. The identification of nuclear proteins and 2-DE database construction are necessary prerequisites for the assignment of differentiation-associated proliferation and proteins. that will allows subsequent developmental and physio-pathological studies.

Experimental Procedures

Antibodies

The following monoclonal antibodies were used : anti-nucleolin(C23) (Santa Cruz), antiactin(C4) (Chemicon), anti- α -tubulin, antivinculin and anti-Golgi 58K (Sigma), anti-lamin B2 (Novo), anti-mitochondria Ab-1 (clone 113-1) (Neomarkers Interchim). Affinity-purified polyclonal anti-GAPDH [11] and anticytokeratins 8-18 [12] have been previously described. Polyclonal anti-histone-H2b and -H3 were generously given by Dr.S.Muller (IBMC, Strasbourg). Rabbit polyclonal anti-hnRNP A2/B1 and B1 were generous gifts from Dr. E. Sueoka (Saga Medical School, Japan), rabbit polyclonal anti-hnRNP A1 came from Dr. L. Abel (University of Pennsylvania, USA) and mouse monoclonal anti-sucrase-isomaltase from Dr. H.P. Hauri (Basel, Switzerland). Alexa Fluor 488 or 568 Goat anti-mouse IgG, Alexa Fluor 488 or 568 Goat anti-rabbit IgG were purchased from Molecular Probes.

Cell culture

At confluency, Caco2/TC7 [6] cells are trypsinized and seeded at 17,500 cells per cm² on culture dishes or on glass coverslips. Cells were cultured in Dulbecco's modified Eagle medium (DMEM) from In Vitrogen containing 20% fetal calf serum at 37°C in 5% $CO_2/95\%$ air. After 3, 7 or 14 days of culture, cells were rinsed twice with 0.15M NaCl and once with 10mM

Phosphate Buffer pH 7.2, 0.15M NaCl (PBS) and detached by scraping in PBS.

Nuclei purification

Caco2 cell suspensions were centrifuged at 1000xg for 10 min at 4°C. After discarding the supernatant, the cell pellet was resuspended with a lysis buffer containing 0.5% Igepal (e.g. Nonidet P-40, Sigma), 5mM MgCl₂, 10mM NaCl, 10mM Tris-HCl pH 7.5 and placed on ice 5 min just before centrifugation for 5 min at 500xg at 4°C. This step was repeated twice. The nuclei pellet was resuspended in 0.25M sucrose, 10mM MgCl₂ 20mM Tris pH 7.5 with acocktail of protease inhibitors (Sigma) and with 1mM DTE. The nuclei were layered on a top of a 2M sucrose cushion in 1mM MgCl₂, 20mM Tris pH 7.5 and then centrifuged for 25 min at 30000xg at 4°C. The nuclei pellet was resuspended in a suitable buffer for appropriate analysis. The sucrose concentration is critical and 2.0M was found to be optimal for the isolation of nuclei from Caco-2/TC-7 cells whatever the stage of the culture.

To examine nuclei integrity, nuclei or cells were deposited on glass slides with a cytobucketsystem (Jouan) and centrifuged for 10 min at 100xg. The samples were fixed with 4% paraformaldeyde (Merck) for 20 min and washed with PBS. The cells were permeabilized with 0.5% Triton X-100 for 15 min and washed with PBS. Nuclei or cells were stained for 20 min with 1% hematoxylin and washed extensively with water. Afterwards 1% sodium acetate was added for 10 min and finally samples were stained with 1% eosin. Nuclei preparations were also stained with Hoechst 33258 (Sigma).

Two-dimensional gel electrophoresis (2-DE) To optimize solubilization of proteins, the pellet of purified nuclei was resuspended in 2D-buffer containing 7M urea, 2M thiourea, 0.24% Triton X-100, 4% CHAPS and 20mM DTT. To precipitate DNA, 10mM spermine was added for 1h at room temperature. Centrifugation 30 min at 200,000xg and 4°C allowed to recover nuclear proteins in the supernatant. ReadyStrip IPG stips pH 3-10 (7 and 17 cm, Bio-Rad) were rehydrated overnight in a reswelling tray (Amersham Biosciences) in 350µl sample containing 90 µg or 350ug of nuclear proteins in 2D-buffer and 0.5% IPG buffer (Amersham Biosciences). First non-denaturing dimension. isoelectrofocalisation(IEF), was carried out on Protean IEF Cell (Bio-Rad) at 19°C with a maximum current setting of 50µA/strip. Focusing was performed for a total of 30,000 Vh or 80,000 Vh for respectively 7 cm or 17 cm strips.

Before carrying out the second dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the strips were subjected to a two-step equilibration. The first one was with an equilibration buffer consisting in 6M urea, 30% glycerol, 2% SDS, 50mM Tris-HCl pH 8.8 and 1% w/v DTT during 20 mn at 20°C. The second step was with the same equilibration buffer without DTT but with 2.5% (w/v) iodoacetamide during 20 mn at 20°C. Strips were then transferred onto the seconddimensional 1 mm thick SDS-PAGE gel and sealed in place with 1% low melting point agarose (Sigma). SDS-PAGE was performed on a 7.5-15% acrylamide/piperazine di-acrylamide (PDA) gel gradient at 50V for 30 min, then at 200V / 28mA overnight at 15°C using the Protean II XL cell (Bio-Rad). The gel was run in the following electrode buffer: 25 mM Tris, 192mM glycine, 0.1% SDS, supplemented with 0.02% sodium thiosulphate when silver staining was performed. 2-D SDS-PAGE standards (Bio-Rad) were used for gel calibration .Silver staining of the gel was performed using the Silver Staining Kit from Amersham Biosciences, according to the manufacturer's instructions. The gel was scanned with the GS 700 Imaging Densitometer (Bio Rad). For differential display study, the gel was coloured using the SIS software (Olympus).After synthesis, ruthenium II tris (bathophenantroline disulfonate) was used for protein visualization by fluorescence staining as described [13]. Fluorography scanning was performed with the Molecular FX (Bio Rad) at 488 nm. After scanning, the gel was placed on an UV transilluminator (310 nm) and protein spots were manually excised for mass spectrometry analysis.

Western blotting

The proteins were either separated on a small gel (8x7 cm) by one dimensional SDS-PAGE as previously described [14] or by 2-DE technique as described above using a Protean Mini II chamber (Bio-Rad). Proteins were transferred onto nitrocellulose sheets [15]. The membranes were blocked in 5% non-fat dried milk in PBS/0.05% Tween 20 and probed at room temperature successively with primary (for 4h) and secondary (for 1h) antibodies. Proteins were visualized using ECL reagents (Amersham Biosciences).

Protein identification by MALDI Mass spectrometry analysis

In-gel digestion

Each gel slice was cut into small pieces with a scalpel, washed with 100µl of 25mM NH₄HCO₃ and dehydrated with 100µl of acetonitrile. This operation was repeated twice. Reduction was achieved by 1 hr treatment with 10mM DTT at 57°C. Alkylation reaction was performed by 25mM iodoacetamide for 45 min at room temperature, protected from light. Finally, gel spots were washed 3 times for 5 min again alternately with 25mM ammonium carbonate and acetonitrile. Gel pieces were completely dried with a Speed Vac before tryptic digestion. The dried gel volume was evaluated and three volumes of trypsin (Promega V5111), 12.5ng/µl, in 25mM NH₄HCO₃ (freshly diluted) werre added. The digestion was performed at 35°C overnight. The gel pieces were centrifuged and 5µl of 25% H₂O / 70% acetonitrile/ 5% HCOOH were added to extract peptides. The mixture was sonicated for 5 min and centrifuged. The supernatant was recovered and the operation was repeated once.

MALDI-MS analysis

For MALDI-TOF-Mass Spectrometry analysis, the supernatant volume was reduced under nitrogen flow to 4µl, 1µl of H₂O/HCOOH were added and 0.5µl of the mix were used for the analysis. Mass mesurements were carried out on a Brucker ULTRAFLEXTM TOF/TOF mass spectrometer equipped with the SCOUTTM High Resolution Optics with X-Y mutlisample probe and gridless reflector. This instrument was used at a maximum accelerating potential of 20kV (in positive mode) and was operated in reflector mode. A saturated solution of α -cyano-4hydroxycinnamic acid in acetone was used as a matrix. A first layer of fine matrix crystals was obtained by spreading and fast evaporation of 0.5ul of matrix solution. On this fine layer crystals, a droplet of 0.5µl af aqueous HCOOH (5%) solution was deposited. Afterwards, 0.5µl of sample solution was added and a second 0.3µl droplet of saturated matrix solution (in 50% H₂O/50% acetonitrile) was added. The preparation was dried under vacuum. The sample was washed one to three times by applying 1µl of aquenous HCOOH (5%) solution on the target and then flushed after a few seconds. In positive mode, internal calibration is performed with tryptic peptides coming from autodigestion of trypsin, with respectively monoisotopic masses at m/z= 842.51; m/z=1045.564; m/z= 2211.105. Monoisotopic peptide masses were assigned and used for databases searches.

These files were then fed into the search engine MASCOT (Matrix Science, London, UK). The data were searched against NCBI nonredundant protein sequence database with trypsin plus potentially one missed cleavage. All proteins present in NCBI were taken into account without any pI and Mr restrictions. The peptides mass error was limited to 50ppm.

Immunocytochemistry

Cells seeded on glass coverslips were rinsed in PBS and fixed in 2.5% (w/v) paraformaldhehyde in PBS for 20 min at room temperature. The fixed cells were treated with 0.5% Triton X-100 (15 min) to permeabilize the cells and then with 0.1% sodium borohydride in PBS (15 min) to block the unreacted aldehyde residues. Non-specific binding sites were blocked in 5% normal goat serum for 40 min at room temperature and incubated overnight at 4°C in PBS containing primary antibodies. After 3 washes in PBS (10 min each), cells were incubated for 1 hour in Alexa fluor 488 or Alexa fluor 568 labeled secondary antibodies depending on the primary antibody. Finally, the cells were washed extensively two times (10 min each) with PBS, incubated with Hoechst 33342 (0.5µg/ml PBS) for 10 min to visualize nuclei and rinsed again with PBS. Preparations were mounted in 1% (w/v) paraphenylenediamine dissolved in PBS/glycerol.

All the samples were examined with a Olympus AX 60 fluorescence microscope. Pictures were taken with a Olympus digital camera and a SIS (soft imaging system) software. For negative controls, primary antibodies were omitted.

Results

To analyse the nuclear proteome of intestinal epithelial cells, during the proliferative phase, at the onset and at an advanced stage of cell differentiation, we used the human colon adenocarcinoma cell line Caco2/TC7 [6]. This cell line was studied at three stages of the culture according to the growth curve pubished previously [10], Fig.7: during the exponential phase of growth (day 3), at confluency (day 7), and one week after confluency (day 14). Caco2 cell line represent the best in vitro model mimicking the processes of intestinal proliferation-differentiation occurring in vivo.

As seen on Fig1, the well described differentiation marker, sucrase-isomaltase (SI) localized at the apical surface of the intestinal epithelial cells increases from the stage of growth arrest (day 7) to later culture stages (day14).

The different steps of our work comprised first, the establishment of a nuclear proteome map of Caco2/TC7 cells during the exponential phase of growth (day 3), and, second, the comparison of the nuclear proteomic pattern at two stages of Caco2 cell differentiation (early, day 7, and late day 14) to that of the proliferative cells.

Characterisation of the nuclear proteome of proliferative Caco2/TC7 cells

Quality control of the integrity and purity of the nuclei fraction

It was essential to work on highly purified preparations of nuclei as a source material for nuclear proteome analysis. Light microscopy and Hoechst staining allowed to show that nuclei after purification on sucrose cushion, appear intact and well preserved (Fig.2 B and C). Furthermore, no visible contamination by cytoplasm was observed (Fig.2 B versus A).

Biochemical markers were used to evaluate the protein content of nuclei (Fig.3). Histones H2b, H3 and lamin B2, well-known nuclear proteins, were present among the nuclei proteins (Fig.3A). The presence of lamin B2, a protein specific of the nuclear envelope confirmed the nuclei integrity observed in Fig.2 (B-C). To check the purity of the nuclei fraction, we analysed other cellular organelles markers of (mitochondria, Golgi apparatus). As shown in Fig.3C, the nuclei preparation was devoid of mitochondria and Golgi proteins as tested with organelle-specific antibodies. Membrane proteins such as sucrase-isomaltase (SI), cytoplasmic proteins as housekeeping proteins (GAPDH); focal-adhesion plaque proteins (vinculin), were not found in the nuclei preparation (fig.3B). However, cytokeratins and actin, cytoskeletal-constitutive proteins were present in the nuclear fraction while tubulin was faintly detected. The presence of these cytoskeletal proteins in nuclei will be discussed in Section 4.

3.1.2 Nuclear proteome map from proliferative cells

The establishment of a 2-DE nuclear protein map from proliferative cells is a prerequisite for the subsequent comparison of the nuclear proteome profiles as a function of Caco2/TC7 cell differentiation. To construct a 2-DE map, it is essential to have a representative analytical 2-DE gel, this has been assessed by the comparison of gels obtained from independent five experiments. A representative silver-stained gel of a nuclear extract obtained from 3 days Caco2/TC7 cells is illustrated in Fig.4. We obtained well-resolved and reproducible gels where the majority of nuclear proteins have pI comprised between 5-7. Based on the results of the image analysis with PDQUEST software (Bio-Rad), approximately 1000 spots were detected on a 2-DE gel performed with 90µg of nuclear proteins obtained from proliferative Caco2/TC7 cells. Silver stain is a very sensitive but not quantitative staining tool, moreover, it can also stain nucleic acids and lipids. Therefore, for protein quantification, we used preparative 2-DE gels where nuclear proteins (350µg) were separated as described above but stained with a fluorescent protein stain, synthesized as previously described [13]. Ruthenium II staining displays less protein variability than silver and better performance for mass spectrometry. Using this technique, 400 spots were detected from 350µg of nuclear proteins obtained from proliferative cells (Fig.5). Ruthenium-stained spots were excised for mass spectrometry analysis, among 100 spots excised from the gel shown in Fig.5, 95 of them presented MALDI-MS spectra with good intensity and resolution. Databases searches with peptide masses allowed to identify a total of 87 of these spots corresponding to 61 different protein species; several spots being isoforms of a same protein (Table 1). In this table, identified nuclear proteins are listed with their protein name, assigned with their SWISS-PROT accession number and their relative mass and pI values (theoretical and experimental). The porcentage of proteic sequence coverage obtained with MALD-MS data are also indicated in table I. The proteins belonging to the different functional or structural families involved in the nuclear organization (nuclear matrix) or functions (splicing, ribosome biogenesis) are pooled in Table 2.

Changes in the proteomic pattern as a function of differentiation

Cellular differentiation is generally accompanied by changes in nuclear organization [16], [17], [18]. Therefore, the putative modifications of the nuclear proteome were studied at three stages of cell culture, on proliferative (3days), pre-(7days) or well-differentiated (14 days) Caco2/TC7cells. This study was performed by three different approaches : differential analysis of 2-D gels, MALDI-MS, Western blotting and immunolocalisation.

Differential analysis

The differential display strategy allows analysis of superimposed images of nuclear proteomes from proliferative versus pre-differentiated or proliferative/differentiated Caco2/TC7 cells (Fig.6). The comparative analysis of the expression pattern of nuclear proteins has been performed with the SIS analysis software. Silverstained images of nuclear proteins from the proliferative Caco2/TC7 cells are colorized in red and those from confluent and differentiated cells in green. The superposition of spots allow to obtain yellow spots that that correspond to proteins found to be expressed at the two stages of culture studied (3 versus 7 days or 3 versus 14 days of culture). Red spots will correspond to specifically proteins expressed at the proliferative stage (3 days), while green spots are representative of proteins found at confluent (7 days) or differentiated (14 days) stage. In this study, the comparative analysis was restricted to 3 interesting areas from 2-D gels (Fig.6). As shown on Fig. 6, A and A', a protein spot localised in the acidic region of the 2-D gel and with physicochemical parameters (98 kDa, pI 4.8), was present among nuclear proteins from proliferative (3 days) and confluent (7 days) but absent from differentiated cells. According to the data summarized in Table1, this spot identified MALDI-MS printing corresponds bv to nucleolin. Comparison of other nuclear proteins migrating in a more basic region (Fig.6, B, B', C, C') revealed that among three spots (35 kDa, pI 7.3 ; 34.5 kDa, pI 7.3 ; 35 kDa, pI 8.4), spots of pIs 7.3 were mainly expressed at the proliferative stage (B, B', arrows) while this of pI 8.4 was present at the three stages (3, 7 and 14 days). These spots correspond to hnRNP A2/B1 Interestingly, a spot (33 kDa; pI 8.4) only present in proliferative cells (C, C', arrows) has been attributed to hnRNP A1.

3.2.2. Western blotting

The second approach was undertaken to study by Western blotting, the expression of the nuclear proteins described above and their putative posttranslational modifications (PTMs).

Nucleolin, is a nucleolar protein often described as a 100-110 kDa protein, cloning of its cDNA predicted a molecular mass of 77 kDa (for a review, see [19]). In our Western blot experiments (Fig.7), nucleolin migrated at around 100 kDa, was found in two well-defined spots migrating with a pI of 4.8 and 6.1 in 3 days proliferative Caco2/TC7 cells. At 7 days, nucleolin isotype (pI 6.1) and at 14 days no nucleolin isotypes were detected.

In nucleus, heterogeneous nuclear RNA is complexed with to a set of nuclear proteins called hnRNPs; hnRNP A2/B1 and hnRNP A1 are among the most important and abondant nuclear proteins [20]. The hnRNP A2/B1 gene is a single copy gene, an alternative splicing produced hnRNP A2 protein and less abundant hnRNP B1, the sequence of this latter differs from hnRNP A2 only by a 12 amino acid insertion which corresponds to a variation in molecular mass of 2 kDa. In the Caco2/TC7 cell line, the hnRNP A2/B1 antibody recognize 6 isotypes in the proliferative cells : this putative postcould reflect complexity translational modifications such as phosphorylation of the different transcripts. The expression of all of the isotypes decreased in a specific manner when Caco2/TC7 cells differentiate. The ubiquitously expressed hnRNP A1 is a well-characterized hnRNP involved in pre-mRNA and mRNA metabolism [21]. Immunoblotting with specific hnRNP A1 antibody revealed one protein, only expressed in proliferative and to a lesser extent in confluent cells, but undetected at the differentiated stage (14 days) of Caco2/TC7cells.

• Immunolocalisation

Immunocytochemistry was performed to study the distribution of the three nuclear proteins (nucleolin, hnRNP A2/B1, hnRNP A1) during the growth phase, at confluency and at a differentiated stage of Caco2/TC7 cells. At 3 and 7 days of culture, an intense immunofluorescent staining of nucleolin was observed in the nucleus, predominantly distributed as intensively stained nucleoli-like bodies (Fig. 8, A-B). This staining was strongly diminished in the welldifferentiated cells cultured for 14 days. (Fig. 8 C). The hnRNP A2/B1 immunostaining was distributed all over the nucleus excluding less intense territories, with a complementary pattern of expression compared to that of nucleolin (Fig. 8, D-E). A slight decrease of staining was observed in the nucleus of differentiated cells (Fig. 8 F). The hnRNP A1 immunostaining, broadly distributed throughout the nucleus, was intense in proliferative and confluent cells (Fig.8 G and H), but decreased in differentiated cells (Fig.8 I). As for hnRNP A2/B1, the hnRNP A1 immunostaining seemed to be excluded from

nucleoli-like foci (compare Fig. 8 H and E compared to B).

Discussion

Apart from the development of a membrane protein database of another human carcinoma cell line, LIM 1215, [22], this study represents the first analysis of a nuclear proteome during proliferation and differentiation of human intestinal epithelial intestinal cells. As a model, we used a clonal cell population (TC-7) isolated from late passages of the Caco-2 cell line [10]. Caco2 cell line, although tumoral [6], has been widely accepted as a useful in vitro model to study the proliferation-differentiation processes which takes place in the intestine along the crypt-villus axis [8]. Caco2/TC7 cells were studied at three stages of the culture and allowed us to identify (i) the proteomic pattern of the cell nuclei during the proliferative phase, (ii) the changes occurring during the differentiation process, and (iii) the differential expression and distribution of three proteins involved in nuclear functions (transcription, pre-mRNA splicing, ribosome assembly). Our results are in agreement with several reports which indicate that the nuclear organisation changes with cellular functions, such as proliferation and differentiation, [18], [16,17].

Our study consisted in developing firstly the optimal preparation of the nuclear content and secondly the differential analysis of the nuclear proteome, comparing proliferative and differentiated cells, which allowed us to validate the cellular model as well as the techniques used.

Besides morphological observations showing the good integrity of nuclei obtained from Caco2/TC7 cells, we checked the quality of the preparations showing on one hand the presence of specific nuclear proteins like histones H2b, H3 and lamin B2 and on the other hand the absence of various cellular markers. It was surprising to observe that cytokeratins were highly represented in the nuclear preparation. As the nucleus is connected to cytokeratin filaments, several alternative protocols have been performed to prevent a potential contamination. First, nuclear proteins were extracted in presence of Triton X-100, but in absence of urea which solubilizes cytokeratins [12]; second, the nuclei were treated as previously described [23] with vanadyl ribonucleoside, a compound which affects physical properties of cytofilaments. In both conditions, cytokeratins were still detected in the nuclei fraction (unpublished observations), interestingly, these nuclear cytokeratins have a higher molecular mass (Fig.3), this difference could be due to phosphorylation as it has been described in human epithelial cells during stress, apoptosis and mitosis [24]. The presence of cytokeratins has also been previously described in the nuclear proteome from human liver [3] and even in the nucleolus proteome [25]. The presence among the nuclear proteins of other cytoskeletal proteins, like actin was not surprising in the light of published data. Indeed, previous works have described actin as a nuclear protein besides its location in the cytoplasm [26] or as a transcription initiation factor [27]. Thus, some of these structural proteins (cytokeratins, actin) could be considered as integral nuclear proteins.

Among the 400 spots detected by ruthenium staining of 2D gel and obtained from proliferative Caco2/TC7 cells, MALDI-MS allow us to identify 87 spots corresponding to 61 different proteins, some of them beeing protein isotypes. By the differential analysis allowing the direct comparison of spots corresponding to nuclear proteins obtained from Caco2/TC7 cells at different stages of cell culture, we focussed our interest on 3 sets of nuclear proteins (nucleolin, hnRNP A2/B1, hnRNP A1).

As far as nucleolin is concerned, its expression was found to be related to the proliferation state of the Caco2/TC7 cells, this observation is confirmed by previous works which established that nucleolin plays a fundamental role in cell proliferation [19,28]. It has been also described that nucleolin levels are highest in tumors or other rapidly dividing cells [29]. Nucleolin is used in studies of different cancer cell lines as a useful marker of cell proliferation [30]. in cell growth may be Nucleolin function regulated by phosphorylation [31]. Nucleolin phosphorylation must cooperate with its dephosphorylation for the regulation of its function in cell growth [28]. Dephosphorylation of proteins by protein phosphatases (PPs) is a common mechanism. It has been recently described that nucleolin interacted with a protein phosphatase (PP1 δ) in human osteoblastic cells [32], these results suggest that nucleolin could be a possible candidate substrate for PP18. The presence of another isoform of protein phosphatase (PP1y) among the identified nuclear proteins needs further studies on the colocalisation of nucleolin and PP1y in Caco2/TC7 cells during the different stages of cell culture. In differentiated Caco2/TC7 cells,

nucleolin was no more found in the nuclear fraction of differentiated cells as attested by immunoblotting and immunofluorescence. The loss of this protein could result from the proteolysis or autodegradation of nucleolin as it has been previously described [33], [34]. Another class of molecules, the hnRNPs A2/B1 was also found to be highly expressed at the proliferative state of Caco2/TC7 cells. The hnRNPs A2/B1 were also up-regulated during mammalian lung development [35] or in pancreatic tumor cells [36]. Similarly, hnRNP A1 is up-regulated in proliferating Cao2/TC7 This well-characterized hnRNP cells. is proliferating overexpressed in and/or transformed cells compared to differentiated tissues [21]. In nucleus, transcription factors are classified under proteins with a diffuse nucleoplasmic distribution, and represented a low percentage of nuclear proteins. In spite of this limitation, the transcription factor Prox-1 was identified among nuclear proteins obtained from proliferative Caco2/TC7 cells, this is the first evidence of its presence in intestinal epithelial cells. The pattern of the expression of Prox-1 in function of differentiation will be investigated . Until now, other transcription factors were not found from the excised spots; this could be due to their low level of expression and/or to the sensitivity threshold of the ruthenium staining. A further purification step, like DNA-affinity chromatography may be needed to purify such nuclear proteins.

These data highlight the dynamic nature of the nuclear proteome in function of cell proliferation/differentiation and show that proteins can associate transiently with specific regions of nucleus, i.e. nucleoli.

Thus this baseline nuclear proteomic study of a widely used intestinal epithelial cell model will be a reference tool for further changes occurring in response to physiological, pharmacological and pathological modulations, in vitro but also in transgenic versus wild type mice or in pathological samples.

Acknowledgements

Natacha Turck is recipient of a fellowship from the French Ministry of Research and Education.

We would like to acknowledge financial supports from the French Ministry of Research and Education (ACI Developmental Biology and Physiopathology n°172), the Ligue Contre le

Cancer (Comités Départementaux du Bas-Rhin et du Haut-Rhin, France).

We thank Drs. M. Rousset and A. Zweibaum (INSERM U.505, Paris, France) for providing Caco2/TC7 cells. We are grateful to Dr. E. Sueoka (Saitama, Japan) for giving us hnRNP A2/B1 antibodies. We also thank Dr. HP. Hauri (Basel, Switzerland) for anti-sucrase-isomaltase antibody and S. Muller (Strasbourg, France for anti-histones antibodies. We would like to thank L. Mathern for illustrations. We especially thank Dr. O. Lefebvre for the use of the SIS software.

LEGENDS OF FIGURES



Figure 1. Caco2/TC7 cell differentiation as a function of culture time period. Indirect immunofluorescence staining of sucrase-isomaltase (SI) at day 3 (A), day 7 (B) and day 14 (C) of culture. The staining becomes visible at the apical pole of the cells at day 7; its intensity increases from day 7 to day 14 of culture attesting cell differentiation. Corresponding Hoechst staining (D-F) allows to visualize nuclei. Note the size variation of nuclei in function of cell culture (D-F). This reflects the fact that the cells which form small clones of large cells at day 3 are smaller at day 14 due to polarization. Bar, 25µm.



Figure 2. A, C, D: Hematoxylin-eosin staining of Caco2/TC7 cells (A) or of purified nuclei (C,D). Hematoxylin-eosin stains the cytoplasm in pink and nuclei in blue. Note the absence of cytoplasm in (D). B: Hoechst staining shows the integrity of purified nuclei. N: nuclei, C: cytoplasm. Bars, 5μ m (A), 10μ m (B,D), 20μ m (C).

E, **F**: Purity control of nuclei obtained from Caco2/TC7 cells. (E): Biochemical protein markers specific of nuclei: histone H2b (15 kDa), histone H3 (15 kDa) and lamin B2 (68.5 kDa). (F): Markers of cytoplasmic components - mitochondria (Mito: 60 kDa), Golgi (58 kDa) - cytoskeleton: cytokeratins (CK 8-18: 52 kDa), actin (43 kDa), tubulin (55 kDa) - adhesion-plaque (Vinc, vinculin: 130 kDa) - housekeeping protein (glyceraldehyde-3-phosphate-dehydrogenase: 35kDa) -membrane proteins (SI, sucrase-isomaltase: 212 kDa) have been analyzed by western blotting to test the cross-contamination of purified nuclei. SI was only detected in cells at day 7 or 14 of culture. Note that cytokeratins (CK) and actin were present among nuclear proteins. Neither other organelles (mitochondria or Golgi) nor cytoplasmic proteins were observed. N: nuclear fraction (30 μ g), C: total cell extract (30 μ g).



Figure 3. Representative 2-DE silver-stained gel of the nuclear proteome from day 3 proliferative Caco2/TC7 cells. 90 μ g of nuclear proteins were focused on a pH 3-10 ReadyStrip IPG strip of 17cm before being separated by a 7.5-15% SDS-PAGE.



Figure 4. Two-dimensional map of nuclear proteins from proliferative (day 3) Caco2/TC7 cells. 350 µg of proteins were focused on a pH 3-10 ReadyStrip IPG strip before being separated by a 7.5-15% SDS-PAGE. Protein spots were visualized by RuBPS staining. Proteins identified by MALDI-MS analysis are indicated by their SWISS-PROT accession number. The identity of the proteins is mentioned in Table 1.



Figure 5. A-C: Differential analysis on defined areas of 2-DE silver stained gels of nuclear proteins in function of cell differentiation. Each nuclear extracts (90 μ g) from either day 3, day 7 or day 14 Caco2/TC7 cells were separated by 2-DE. 2-DE silver-stained gels were scanned and the digitalized images were colorized by the SIS software in red for day 3 and in green for day 7 or day 14. Spots were overlaid using the software. Yellow spots represent proteins commonly expressed between day 3 and day 7 or between day 3 and day 14.

A'-C': Western blot analysis of expression of nucleolin (A'), hnRNP A2/B1 (B'), hnRNP A1 (C') in function of cell differentiation. Nuclear proteins ($40\mu g$) prepared from proliferative (day 3), confluent (day 7) and differentiated (day 14) Caco2/TC7 cells were separated by 2-DE. Nucleolin, hnRNP A2/B1 and hnRNP A1 are down-regulated during Caco2/TC7 cell differentiation. Only parts of Western blots are illustrated.



Figure 6. Intracellular distribution of nucleolin, hnRNP A2/B1 and hnRNP A1 in function of cell differentiation. Immunofluorescent staining of nucleolin (A-C), hnRNP A2/B1 (D-F) and hnRNP A1 (G-I). The three proteins are found located in the nuclei. Nucleolin was distributed as intense spots in day 3 and day 7 cells while hnRNP A2/B1 and hnRNP A1 were distributed all over the nucleus except in some less intense territories. Decreased staining was obvious for the three proteins at day 14 (C, F, I). Bar, 25µm.

	SWISS-PROT		% segu.	Theoretical values		Experimental values		number of	error in
	accession number Protein name		coverage	Da	DI.	Da	υI	peptides	ppm
1	P08206	Calpactin I light chain	55	11072	7.30	10500	8.05	6	24
2	P98179	Putative RNA binding protein 3	44	17170	8.86	18000	8	80	22
3	O06830	Peroxiredoxin	34	22110	8.27	22000	9.70	5	32
4	P82979	Nuclear protein Hcc1	40	23671	6.10	30500	6.15	8	25
5	P17080	GTP binding nuclear protein RAN	26	24423	7.01	26000	6.05	5	5
6	P19388	DNA directed RNA polymerase II 23kD polymentide	42	24612	5.53	26000	5.75	9	28
7	P24534	Elongation factor 1 ^β	35	24632	4.50	28500	4.75	6	27
8	075934	Putative spliceosome associated protein	33	26131	5.48	26000	5.45	7	18
9	P56537	Eukaryotic initiation factor 6	33	26599	4.56	28000	4.75	5	29
10	Q99714	3- hydroxyacyl-CoA dehydrogenase type II	31	26973	7.65	21000	8.70	5	18
11	Q92786	Homeobox prox	59	28068	6.19	26000	5.95	10	28
12	P09661	u2 small nuclear ribonucleoprotein A'	22	28443	8.73	24500	9.75	6	38
13	075822	Eukaryotic initiation factor 3A	15	28990	4.74	28000	4.75	5	29
14	P12750	40s Ribosomal protein S4X	33	29466	10.16	29000	10	10	33
15	P35232	Prohibitin	58	29804	5.57	29000	5.45	13	27
16	P35232	Prohibitin	24	29804	5.57	29000	5.55	6	23
17	P35232	Prohibitin	61	29804	5.57	36000	5.45	13	50
18	P35232	Prohibitin	41	29804	5.57	36000	5.50	7	28
19	P21796	Voltage dependent anion selective channel protein-1 (VDAC)	46	30641	8.63	32000	9.30	11	44
20	P09012	ul small nuclear ribonucleoprotein A	45	31279	9.83	31500	5.25	10	18
21	O60812	Heterogeneous nuclear ribonucleoprotein C	25	32142	4.93	40000	5.80	9	50
22	P06748	Nucleophosmin	34	32575	4.64	35000	4.95	11	50
23	P06748	Nucleophosmin	34	32575	4.64	36000	5	10	27
24	P09493	Tropomyosin a	41	32708	4.69	30500	4.95	18	20
25	P09493	Tropomyosin a	41	32708	4.69	33500	5	15	19
26	P08865	40s ribosomal protein SA	30	32854	4.79	41000	5	9	50
27	P32322	Pyrroline-5 carboxylate reductase	40	33374	7.18	29000	9.30	10	17
28	P07910	hnRNP C1/C2	37	33688	4.95	38000	5.10	10	43
29	P07910	hnRNP C1/C2	31	33688	4.95	39500	5.10	9	50
30	P05388	60s acidic ribosomal protein P0	18	34273	5.75	34500	5.45	4	12,5
31	P25388	Gaznine nucleotide binding protein β subunit like 12,3	47	35077	7.60	31000	5.95	11	20
32	P25388	Guanine nucleotide binding protein β subunit like 12.3	43	35077	7.60	31000	7.20	10	23
33	P31942	hnRNP H3	46	36926	6.37	35500	6.10	12	17
34	P31942	hnRNP H3	43	36926	6.37	35500	6.45	10	9
35	P36873	Serine threanine protein phosphatase PP17	27	36983	6.12	35500	5.95	7	23
36	P36873	Serine threanine protein phosphatase PP17	26	36983	6.12	35500	6.10	8	21
37	P22626	hnRNP A2/B1	43	37429	8.97	34000	7.25	13	17
38	P22626	hnRNP A2/B1	37	37429	8.97	34500	7.25	18	30
39	P22626	hnRNP A2/B1	34	37429	8.97	34000	8.30	18	39
40	Q14103	hnRNP D0	23	38434	7.61	40000	6	6	39
41	P07365	Annexin 2	61	38473	7.56	35500	7.65	20	43
42	P09651	hnRNPA1	24	38700	9.30	33000	8.40	7	7
43	P51991	hnRNP A3	15	39686	8.74	37000	8.10	5	35
44	Q15019	Septin 2	50	41487	6.15	41000	5.95	14	50
45	Q15019	Septin 2	37	41487	6.15	41000	6	10	15
46	P02570	Actine	39	41737	5.29	42000	5.25	13	48
47	P38159	hnRNP G	19	42404	10.02	45000	10	8	27
48	Q9UQ80	Proliferated associated protein 2G4	22	43787	6.13	47000	5.95	7	24
49	P08727	Cytokeratin 19	74	44106	5.05	41000	5.15	27	50
50	015427	Splicing factor 3B 4subunit	20	44385	8.55	47500	8.50	5	3.5
51	P04765	Eukaryotic imitiation factor 4A-I	48	46154	5.32	46000	5.95	20	15
52	P04765	Eukaryotic initiation factor 4A-I	29	46154	5.32	46000	6.15	13	4
53	Q09028	Chromatin accembly factor p48 subunit	24	47656	4.74	50000	4.90	10	28
34	P05783	Cytokeratin 18	49	47926	5.34	40000	5.45	81	18
- 55	P05785	Cytoseratin 18	62	47926	5.34	54000	5.60	28	50

Table 1: Nuclear proteins from the 2-DE map of proliferative Caco2/TC7 cells identified by peptide mass fingerprinting

Table 1 - continued

	SWISS-PROT	ROT % sequ. Theoretical value		al values	alnes Experimental values number of err			error in	
	accession number	Protein name	coverage	Da	nI	Da	N	peptides	DDID
- 56	P26641	Elongation factor 17	31	50118	6.25	46000	5.95	15	50
57	P26641	Elongation factor 17	40	50118	6.25	46000	6.15	14	25
58	P02551	Tubuline 🌣	27	50135	4.94	51500	5.05	12	42
59	P05787	Cytokeratin S	35	53543	5.52	51500	5.65	16	49
60	P14866	hnRNP L	24	60187	6.65	61500	6.55	12	25
61	P20700	Lamin B1	44	66277	5.11	67000	5.20	23	30
62	Q03252	Lamin B2	52	68539	5.90	64000	5.45	41	50
63	P15311	Ezrin	55	69267	5.95	77000	6.05	10	22
64	P12956	ATP dependent DNA helicase 2	29	69712	6.23	69500	6.10	11	12
65	P12956	ATP dependent DNA helicase 2	21	69712	6.23	69500	6.15	8	13
66	P08107	Heat shock protein 70	26	70052	5.48	67000	4.90	11	26
67	P11142	Heat shock protein 71	48	70898	5.38	68000	4.85	29	40
68	P11021	Glucose regulated protein precursor 78kD	24	72333	5.03	73500	4.6	10	14
69	O00571	DEAD box protein 3	25	73112	6.73	74500	6.45	13	9
70	P02545	Lamin A and C	25	74139	6.57	63000	6	14	18
71	P02545	Lamin A and C	36	74139	6.57	63000	6.10	20	30
72	P02545	Lamin A and C	28	74139	6.57	63000	6.20	17	17
73	P02545	Lamin A and C	43	74139	6.57	63000	6.30	23	11
74	P02545	Lamin A and C	20	74139	6.57	74500	6.35	10	10
75	P23246	Splicing factor proline glutamine riche	41	76149	9.45	97500	10	28	50
76	P19338	Nucleolin	29	76200	4.59	96000	4.90	19	22
77	P19338	Nucleolin	16	76200	4.59	96000	6.10	8	16
78	P52272	Heterogenous ribonucleoprotein M	36	77470	8.94	72500	7.70	22	50
79	P52272	Heterogenous ribonucleoprotein M	43	77470	8.94	72500	8.80	33	21
80	Q16891	Motor protein	28	\$3678	6.08	\$3500	5.70	21	1
81	Q16891	Motor protein	33	\$3678	6.08	\$3500	5.75	18	23
82	Q16891	Motor protein	29	\$3678	6.08	\$2500	5.80	10	19
83	Q16891	Motor protein	26	\$3678	6.08	\$5000	5.75	14	50
84	P43243	Matrine 3	31	94623	5.87	103000	5.60	27	42
85	P11586	C1 tetrahydrofolate syntase	23	101428	6.94	100000	6.50	17	19

Table 2: Classification of the identified nuclear proteins according to their known function

* Nuclear matrix proteins

P09493	Tropomyosin a
P02570	Actine β
P05787	Cytokeratin 8
P08727	Cytokeratin 19
P05783	Cytokeratin 18
P02551	Tubuline α
P02545	Lamin A and C
P20700	Lamin B1
Q03252	Lamin B2
P15311	Ezrin
P43243	Matrine 3

* Ribosomal proteins

P08865	40s ribosomal protein SA	
P12750	40s ribosomal protein S4X	
P05388	60s acidic ribosomal protein	P0

* Elongation factors

P04765	Eukaryotic initiation factor 4A-I
P56537	Eukaryotic initiation factor 6
075822	Eukaryotic initiation factor 3A
P24534	Elongation factor 1β
P26641	Elongation factor 1y

* Transduction signal

P25388	Guanine nucleotide binding protein
	β subunit like 12,3 (RACK-1)
P08206	Calpactin I light chain
P07365	Annexin 2

* Enzymes

C1 tetrahydrofolate syntase
3 hydroxyacyl coA dehydrogenase
Pyrroline-5 carboxylate reductase
Serine threonine protein phosphatase PP1y
Peroxiredoxin

* Cellular cycle

P17080	GTP binding nuclear protein RAN
P35232	Prohibitin
Q15019	Septin 2
Q9UQ80	Proliferated associated protein 2G4

* Stress proteins

P08107	Heat shock protein 70
P11142	Heat shock protein 71
P11021	Glucose regulated protein precursor 78kD
* hnRNP	s / splicing factors
P09651	hnRNPA1
P22626	hnRNP A2/B1
P51991	hnRNP A3
O60812	hnRNP C
P07910	hnRNP C1/C2
Q14103	hnRNP D0
P31942	hnRNP H3
P38159	hnRNP G
P14866	hnRNP L
P52272	hnRNP M
P23246	Splicing factor proline glutamine riche
Q15427	Splicing factor 3B 4subunit
075934	Putative spliceosome associated protein
P09012	ul small nuclear ribonucleoprotein A

P09661 u2 small nuclear ribonucleoprotein A'

* DNA / RNA binding proteins

P98179	Putative RNA binding protein 3
Q09028	Chromatin assembly factor p48 subunit
P12956	ATP dependent DNA helicase 2
P19388	DNA directed RNA polymerase II
P82979	Nuclear protein Hccl
P06748	Nucleophosmin
000571	DEAD box protein 3
P19338	Nucleolin

* Transcription factors

Q92786	Homeobox prox

* Unclassified proteins

P21796	VDAC-1
Q16891	Motor protein
References

- 1. Misteli, T. (2000). Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. J.Cell Sci. 113 (Pt 11), 1841-1849.
- Dundr, M. and Misteli, T. (2001). Functional architecture in the cell nucleus. Biochem.J. 356, 297-310.
- Jung, E., Hoogland, C., Chiappe, D., Sanchez, J. C., and Hochstrasser, D. F. (2000). The establishment of a human liver nuclei two-dimensional electrophoresis reference map. Electrophoresis 21, 3483-3487.
- 4. Kedinger, M., Freund, J. N., Launay, J. F., and Simon-Assmann, P. (2000). Cell interactions through the basement membrane in intestinal development and differentiation. In "Development of the Gastrointestinal Tract" (I. R. Sanderson and W. A. Walker, Eds.), B.C.Decker Inc., London.
- 5. Brittan, M. and Wright, N. A. (2002). Gastrointestinal stem cells. J.Pathol. 197, 492-509.
- Fogh, J. and Trempe, G. (1975). Human tumor cells in vitro. (J. Fogh, Ed.), Plenum Publishing Corp., New York.
- Pinto, M., Robine-Leon, S., APPAY, M. D., Kedinger, M., Triadou, N., Dussaulx, E., LACROIX, B., SIMON-ASSMANN, P., Haffen, K., Fogh, J., and Zweibaum, A. (1983). ENTEROCYTE-LIKE DIFFERENTIATION AND POLARIZATION OF THE HUMAN COLON CARCINOMA CELL LINE CACO-2 IN CULTURE. BIOL.CELL 47, 323-330.
- 8. Zweibaum, A., LABURTHE, M., GRASSET, E., and LOUVARD, D. (1991). THE USE OF CULTURED CELL LINES IN STUDIES OF INTESTINAL AND CELL DIFFERENTIATION FUNCTION. **IN: "HANDBOOK** OF PHYSIOLOGY.INTESTINAL TRANSPORT THE OF SYSTEM." GASTROINTESTINAL FIELD M., FRIZZELL R.A., EDS,

AMERICAN PHYSIOLOGICAL SOCIETY.

- Tadjali, M., Seidelin, J. B., Olsen, J., and Troelsen, J. T. (2002). Transcriptome changes during intestinal cell differentiation. Biochim.Biophys.Acta 1589, 160-167.
- Chantret, I., Rodolosse, A., Barbat, A., Dussaulx, E., Brot-Laroche, E., Zweibaum, A., and Rousset, M. (1994). Differential expression of sucraseisomaltase in clones isolated from early and late passages of the cell line Caco-2: evidence for glucose-dependent negative regulation. J.Cell Sci. 107 (Pt 1), 213-225.
- Launay, J. F., Jellali, A., and Vanier, M. T. (1989). Glyceraldehyde-3-phosphate dehydrogenase is a microtubule binding protein in a human colon tumor cell line. Biochim.Biophys.Acta 996, 103-109.
- Fontao, L., Dirrig, S., Owaribe, K., Kedinger, M., and Launay, J. F. (1997). Polarized expression of HD1: relationship with the cytoskeleton in cultured human colonic carcinoma cells. Exp.Cell Res. 231, 319-327.
- Rabilloud, T., Strub, J. M., Luche, S., van Dorsselaer, A., and Lunardi, J. (2001). A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. Proteomics. 1, 699-704.
- 14. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc.Natl.Acad.Sci.U.S.A 76, 4350-4354.
- Bartova, E., Kozubek, S., Kozubek, M., Jirsova, P., Lukasova, E., Skalnikova, M., and Buchnickova, K. (2000). The influence of the cell cycle, differentiation and irradiation on the nuclear location of the abl, bcr and c-myc genes in human leukemic cells. Leuk.Res. 24, 233-241.

- Bartova, E., Kozubek, S., Jirsova, P., Kozubek, M., Gajova, H., Lukasova, E., Skalnikova, M., Ganova, A., Koutna, I., and Hausmann, M. (2002). Nuclear structure and gene activity in human differentiated cells. J.Struct.Biol. 139, 76-89.
- Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat.Rev.Genet. 2, 292-301.
- 19. Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999). Structure and functions of nucleolin. J.Cell Sci. 112 (Pt 6), 761-772.
- 20. Krecic, A. M. and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. Curr.Opin.Cell Biol. 11, 363-371.
- Biamonti, G., Bassi, M. T., Cartegni, L., Mechta, F., Buvoli, M., Cobianchi, F., and Riva, S. (1993). Human hnRNP protein A1 gene expression. Structural and functional characterization of the promoter. J.Mol.Biol. 230, 77-89.
- Simpson, R. J., Connolly, L. M., Eddes, J. S., Pereira, J. J., Moritz, R. L., and Reid, G. E. (2000). Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. Electrophoresis 21, 1707-1732.
- Gerner, C., Holzmann, K., Grimm, R., and Sauermann, G. (1998). Similarity between nuclear matrix proteins of various cells revealed by an improved isolation method. J.Cell Biochem. 71, 363-374.
- Toivola, D. M., Zhou, Q., English, L. S., and Omary, M. B. (2002). Type II keratins are phosphorylated on a unique motif during stress and mitosis in tissues and cultured cells. Mol.Biol.Cell 13, 1857-1870.
- 25. Dundr, M. and Misteli, T. (2002). Nucleolomics: an inventory of the nucleolus. Mol.Cell 9, 5-7.
- Rando, O. J., Zhao, K., and Crabtree, G. R. (2000). Searching for a function for nuclear actin. Trends Cell Biol. 10, 92-97.

- 27. Egly, J. M., Miyamoto, N. G., Moncollin, V., and Chambon, P. (1984). Is actin a transcription initiation factor for RNA polymerase B? EMBO J. 3, 2363-2371.
- Srivastava, M. and Pollard, H. B. (1999). Molecular dissection of nucleolin's role in growth and cell proliferation: new insights. FASEB J. 13, 1911-1922.
- Derenzini, M., Sirri, V., Trere, D., and Ochs, R. L. (1995). The quantity of nucleolar proteins nucleolin and protein B23 is related to cell doubling time in human cancer cells. Lab Invest 73, 497-502.
- de Verdugo, U. R., Selinka, H. C., Huber, M., Kramer, B., Kellermann, J., Hofschneider, P. H., and Kandolf, R. (1995). Characterization of a 100kilodalton binding protein for the six serotypes of coxsackie B viruses. J.Virol. 69, 6751-6757.
- Peter, M., Nakagawa, J., Doree, M., Labbe, J. C., and Nigg, E. A. (1990). Identification of major nucleolar proteins as candidate mitotic substrates of cdc2 kinase. Cell 60, 791-801.
- Morimoto, H., Okamura, H., and Haneji, T. (2002). Interaction of protein phosphatase 1 delta with nucleolin in human osteoblastic cells. J.Histochem.Cytochem. 50, 1187-1193.
- Bouche, G., Caizergues-Ferrer, M., Bugler, B., and Amalric, F. (1984). Interrelations between the maturation of a 100 kDa nucleolar protein and pre rRNA synthesis in CHO cells. Nucleic Acids Res. 12, 3025-3035.
- Fang, S. H. and Yeh, N. H. (1993). The self-cleaving activity of nucleolin determines its molecular dynamics in relation to cell proliferation. Exp.Cell Res. 208, 48-53.
- Montuenga, L. M., Zhou, J., Avis, I., Vos, M., Martinez, A., Cuttitta, F., Treston, A. M., Sunday, M., and Mulshine, J. L. (1998). Expression of heterogeneous nuclear ribonucleoprotein A2/B1 changes with critical stages of mammalian lung development. Am.J.Respir.Cell Mol.Biol. 19, 554-562.

36. Yan-Sanders, Y., Hammons, G. J., and Lyn-Cook, B. D. (2002). Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. Cancer Lett. 183, 215-220.

1.4 Discussion

Cette étude a permis d'identifier 87 protéines exprimées dans les noyaux de cellules Caco2/TC7 au stade prolifératif, avec seulement 61 protéines différentes indiquant la présence de nombreuses isoformes. Même si ce modèle est un modèle *in vitro*, il a permis de donner de nombreuses informations quant aux protéines présentes au stade prolifératif. Cette étude peut être considérée comme une première étape pour une meilleure connaissance des protéines nucléaires à ce stade *in vivo*. La cartographie des protéines de noyaux de ces cellules servira de base pour une étude protéomique différentielle en fonction des stades de différenciation des cellules et des conditions de culture.

Par ailleurs, cette étude démontre le rôle primordial d'une bonne préparation des échantillons. En effet, pour cette étude, les noyaux ont été purifiés sur gradient de sucrose et divers détergents ont été utilisés, ce qui a pour but d'éviter les contaminations par des protéines cytoplasmiques qui auraient faussé l'étude. En effet, les cytokératines qui entourent le noyau sont solubles dans certains détergents (urée), il fallait donc choisir des détergents permettant d'éliminer les protéines cytoplasmiques (Triton). L'intégrité des noyaux a été vérifiée par microscopie électronique et la présence d'éventuels contaminants contrôlée par Western Blot contre des marqueurs d'autres compartiments cellulaires (mitochondrie, appareil de Golgi). La bonne préparation de l'échantillon, sans contaminant, a permis de mettre en évidence la présence de cytokératines dans le noyau ainsi que d'autres composants du cytosquelette [Rando, 2000], ce qui avait déjà été décrit pour les noyaux de cellules de foie humain [Jung, 2000], mais jamais pour ces cellules.

De plus, lors de ces études aucun facteur de transcription n'a été identifié. Or, leur présence a déjà été décrite dans les noyaux de ces cellules après des études par Western blot [Freund, 1998]. Le fait que ces protéines n'aient pas été détectées est certainement du à des problèmes de sensibilité au niveau des gels 2D (bleu colloidal, utilisé pour l'étape de révélation) ou au niveau de la spectrométrie de masse (suppression de signal). De plus, la présence de protéines très intenses masque les protéines minoritaires. Il serait donc intéressant de sous-fractionner l'échantillon avant de réaliser l'analyse par gel 2D ou alors en réalisant sur l'ensemble des protéines de noyaux une nanoLC-LC MS/MS qui ajouterait un fractionnement supplémentaire.

1.5 Approche protéomique différentielle

Après l'étude systématique du protéome de noyaux de cellules Caco2/TC7 à l'état prolifératif, l'analyse différentielle des protéines en fonction de l'état de différenciation des cellules a été entreprise. En effet, il est intéressant d'identifier les protéines sur-ou sous-exprimées par rapport au protéome nucléaire des cellules en voie de prolifération, puisque la différenciation cellulaire est souvent accompagnée de changements dans l'organisation du noyau [Bartova, 2000], [Bartova, 2002], [Cremer, 2001]. Les premiers résultats de l'étude protéomique différentielle sur les noyaux de cellules Caco2/TC7 différenciées ou en cours de différenciation, ont permis d'identifier 3 protéines principalement exprimées lors de la phase proliférative des cellules : Nucléoline, hnRNP A2/B1 et hnRNP A1 [Montuenga, 1998], [Yan-Sanders, 2002]. La Nucléoline, déjà décrite comme marqueur de prolifération cellulaire [Ginisty, 1999], [Srivastava, 1999] serait active en fonction de son état de phosphorylation. Cet état serait modulé par la protéine phosphatase (PP18) dans les cellules osteoblastiques humaines [Morimoto, 2002]. Or, si la PP18 n'a pas été identifiée dans les noyaux des cellules Caco2/TC7, une autre isoforme de cette protéine a été identifiée, PP1 γ . Des études complémentaires pour vérifier la co-localisation de PP1 γ et de la nucléoline aux différents stades de différenciation devront être réalisées pour vérifier l'hypothèse d'une phosphorylation de la nucléoline par PP1 γ . Des études MS/MS seront également menées pour caractériser l'état de phosphorylation de la nucléoline. Par ailleurs, l'analyse différentielle globale a donné les résultats présentés dans les tableaux suivants (tableaux 12 et 13).

 Tableau 12 : Analyse différentielle des protéines nucléaires de cellules Caco2 entre 2 stades de prolifération cellulaire J7 et J3

Spot	Swiss Prot	Protéines	% recouvrement
1	P02304	Histone H4	52%
2	P02278	Histone H2b	60%
3	P23152	Splicing factor, arginine/serine-rich 3	43%
4	P32969	60s ribosomal protein S7	56%
5	P23821	40s ribosomal protein S7	35%
6	P46782	40s ribosomal protein S5	41%
7	P49798	Regulator of G protein signaling 4	21%
8	Q01130	Splicing factor, arginine/serine- rich 2	26%
9	P17080	GTP binding nuclear protein RAN	43%
10	P56537	Eukaryotic translation initiation factor 6	
		(eIF-6)	52%
11	Q16629	Splicing factor, arginine/serine- rich 7	36%
12	Q07955	Splicing factor, arginine/serine- rich 1	57%
13	P49241	40s ribosomal protein S3A	53%
14	P07910	Heterogeneous nuclear ribonucleoprotein C1/C2	59%
15	P06748	Nucleophosmin	43%
16	P09493	Tropomyosin a	38%
17	P08865	40s ribosomal protein SA	57%
18	P05388	60s acidic ribosomal protein P0	56%
19	015144	Arp 2/3 complex 34kDa subunit	45%
20	P25388	Guanine nucleotide binding protein	
-		b subunit like 12.3 (RACK-1)	77%
21	O9Y4J5	hnRNP 2H9	51%
22	P22626	hnRNP A2/B1	49%
23	P08129	Serine/threonine protein phosphatase PP1a	48%
24	015365	hnRNP E1	52%
25	096DD5	Annexine 2	63%
26	P51991	hnRNP A3	25%
27	P78406	mRNA binding protein mrnp 41	26%
28	P02570	Actine b	48%
29	P38159	hnRNP G	40%
30	P04765	Eukarvotic translation initiation factor 4A-I	63%
31	075955	Flotillin-1	56%
32	P31943	hnRNP H	56%
33	O9P0I1	Nucleoporin p54 protein	37%
34	O9UMS4	Nuclear matrix nmp200	39%
35	09Y6M1	Hepatocellular carcinoma autoantigen	26%
36	P46060	RAN GTPase activating protein 1	23%
37	P20700	Lamin B1	48%
38	003252	Lamin B2	41%
39	P12956	ATP-dependent DNA helicase II	29%
40	P08107	Heat shock protein 70	55%
41	P11142	Heat shock protein 71	47%
42	000571	DEAD box protein 3	35%
43	P02545	Lamin A/C	52%
44	P52272	hnRNP M	30%
45	016891	Mitofilin	40%
46	043143	Pre-mRNA splicing RNA helicase	17%
47	P09327	Villin I	29%
48	P43243	Matrin 3	35%
49	012906	Interleukin enhancer binding factor 3	17%
50	Q13435	Splicing factor 3B subunit 2	12%

spot	Swiss prot	protéine	% recouvrement
1	P02304	Histone H4	53%
2	P16106	Histone H3	38%
3	P07910	Heterogeneous nuclear ribonucleoprotein C1/C2	33%
4	P06748	Nucleophosmin	25%
5	P05388	60s acidic ribosomal protein P0	47%
6	P25388	Guanine nucleotide binding protein	
		b subunit like 12,3 (RACK-1)	26%
7	Q9Y4J5	hnRNP 2H9	27%
8	P22626	hnRNP A2/B1	42%
9	Q96DD5	Annexine 2	41%
10	P38159	hnRNP G	31%
11	P04765	Eukaryotic translation initiation factor 4A-I	35%
12	Q9NZI8	mRNA binding protein CRDBP	15%
13	P20700	Lamin B1	38%
14	Q03252	Lamin B2	38%
15	P08107	Heat shock protein 70	41%
16	P11021	Glucose regulated protein precursor 78kD	22%
17	P02545	Lamin A/C	32%
18	Q9UHB6	Epithelial protein lost in neoplasm	12%

 Tableau 13 : Analyse différentielle des protéines nucléaires de cellules Caco2 entre 2 stades de prolifération cellulaire J14 et J3

Les protéines identifiées, lors de cette étude protéomique différentielle étaient toutes globalement sous exprimées aux différents états de différenciation cellulaires (j7 ou j14). Comme précédemment, lors de l'étude de protéomique descriptive, aucun facteur de transcription n'a été détecté.

1.6 Perspectives

L'analyse du protéome des noyaux de cellules Caco2/TC7 a mis en évidence la présence de plusieurs isoformes de certaines protéines à j3, sur 87 protéines identifiées seules 61 étaient différentes. Ces isoformes diffèrent probablement par la présence de modifications post-traductionnelles (masses moléculaires cohérentes sur le gel 2D, mais pI différent de celui attendu, chapelet de spots..) ou de protéines tronquées (masses moléculaires inférieures à celles attendues) qu'il serait intéressant de caractériser par MS/MS pour compléter l'étude.

Par ailleurs, concernant l'étude protéomique différentielle, il pourra être envisagé de réaliser une étude par marquage des protéines exprimées à j7 ou j14 par rapport à celles exprimées à j3. Il sera ainsi possible de quantifier de façon exacte les protéines sur-et sous-exprimées en fonction des stades de différenciation. De plus, par cette approche et par une séparation par nanoLC-LC-MS/MS, il sera peut-être possible d'identifier les protéines minoritaires, par exemple les facteurs de transcriptions.

De plus, les cellules Caco2/TC7 étant un modèle d'étude très fréquent, il serait intéressant d'en réaliser une cartographie complète, en commençant par l'étude d'autres organels comme les mitochondries. Le fractionnement de l'échantillon en compartiments cellulaires permet d'avoir accès à un plus grand nombre de protéines qui seraient masquées par des protéines majoritaires lors d'une étude cellulaire globale. En effet, après la réalisation de cette cartographie, il serait possible de l'utiliser, comme référence, dans le cadre d'autres études, par exemple pour tester *in vitro* l'action de certains agents pharmacologiques ou bactériens (cf. étude *Streptococcus bovis*) sur ces cellules. Il serait ainsi possible de savoir quelles protéines sont sur ou sous exprimées en fonction des conditions de culture. Par ailleurs, ces cellules étant des lignées de cellules cancéreuses coliques, les résultats pourraient être comparés aux protéomes de cellules cancéreuses issues de patients.

2 Identification de protéines de mitochondries impliquées dans la maladie MERRF

Le but de cette étude, réalisée en collaboration avec Catherine Florentz de l'Institut de Biologie Moléculaire et Cellulaire de Strasbourg, était d'identifier les protéines mitochondriales sur et sous-exprimées dans le cadre de la maladie MERRF (Myoclonic Epilepsy and Ragged Red Fibers).

2.1 Contexte biologique

Les mitochondries (figure 22) sont le lieu de nombreuses fonctions métaboliques d'une importance primordiale pour la cellule, dont la principale est celle de la synthèse d'ATP par l'intermédiaire des enzymes des chaînes respiratoires et de l'ATP synthase, enzymes qui sont toutes localisées au niveau de la membrane interne de la mitochondrie. En plus de la synthèse d'énergie, les mitochondries sont aussi le lieu du cycle du catabolisme de l'acide citrique, d'oxydation des acides gras et d'une partie du cycle de l'urée [Scheffler, 1999], ainsi que de celui de la biosynthèse de l'hème, de l'ubiquinol et des cardiolipines [Scheffler, 1999].





Pour assurer son rôle, la mitochondrie possède son propre génome, codant pour 13 protéines, les autres protéines étant importées après avoir été synthétisées à partir du génome nucléaire (400 à 2000 protéines). Pour assurer la traduction de ces 13 protéines, la mitochondrie dispose de 22 tRNA et de 2 rRNA. Malheureusement, plusieurs mutations du génome mitochondrial ont été décrites [Kogelnik, 1998], [Ingman, 2000], [Servidei, 2001] et ont été corrélées à de nombreuses maladies comme des myopathies, cardiopathies, diabètes...[Schon, 1997], [Wallace, 1999]. En particulier, nombreuses sont les mutations situées au niveau des tRNA, souvent responsables de graves maladies,

en raison de leur rôle central dans la traduction des protéines mitochondriales. La mutation du tRNA Lysine est par exemple associée à une maladie neuromusculaire, la maladie MERRF [Hanna, 1999].

Dans ce contexte, une étude protéomique comparative sur des mitochondries mutées au niveau du gène A8344G codant pour le tRNA Lysine a été menée, afin de mieux comprendre les implications de cette mutation au niveau protéique.

Cette étude a fait l'objet d'une publication dans The Journal of Biological Chemistry.

2.2 Approche protéomique différentielle

Cette approche consiste à comparer 2 gels 2D sur lesquels ont été déposées, des protéines issues de cellules saines sur l'un et des protéines issues de cellules malades sur l'autre.

Pour avoir une représentation statistique des résultats et éviter toute variation liée à l'expérience, les gels sont réalisés en plusieurs exemplaires. Dans le cas de l'étude de la maladie MERRF, 11 gels de chaque ont été réalisés, et analysés informatiquement à l'aide de logiciels (PDQuest[®]) afin de visualiser le plus finement possible les variations d'expression d'un gel à l'autre (figure 23A et B). Ensuite, l'analyse par spectrométrie de masse a permis d'identifier les protéines suret sous-exprimées dans le cadre de la maladie.

Figure 23A : Gels 2D de protéines issues de mitochondries saines et malades, étoiles bleues correspondant aux protéines sous exprimées et rouges aux protéines sur exprimées.







2.3 Publication

Petra Tryoen, <u>Sophie Richert</u>, Bénédicte Sohm, Manuele Mine, Cécile Marsac, Alain Van Dorsselaer, Emmanuelle Leize-Wagner and Catherine Florentz

Proteomic exploration of a mitochondrial tRNA disorder monitors wide range effects on nuclear encoded proteins

Journal of Biological Chemistry, 2003, Jul 4; 278(27), 24314-24323



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

Proteomic Consequences of a Human Mitochondrial tRNA Mutation beyond the Frame of Mitochondrial Translation

Petra Tryoen-Tóth, **Sophie Richert**, Bénédicte Sohm, Manuele Mine, Cécile Marsac, Alain Van Dorsselaer, Emmanuelle Leize, and Catherine Florentz

The Journal of Biological Chemistry, 2003, Vol. 278, N°27, Pages 24314 - 24323

<u>Pages 24314 – 24323 :</u>

- 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://www.jbc.org/cgi/content/full/278/27/24314

• 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: <u>peb.sciences@scd-ulp.u-strasbg.fr</u>

2.4 Discussion

Si lors d'une étude précédente [Rabilloud, 2001], deux protéines avaient déjà été identifiées comme sous-exprimées dans la maladie MERRF, avec cette nouvelle étude protéomique comparative, 38 protéines ont été identifiées comme sur-ou sous-exprimées, après analyse des gels. Parmi ces protéines, la spectrométrie de masse a permis d'en identifier 16 (figure 23B), les autres protéines n'ont pas pu être identifiées, étant en trop faible quantité pour être visibles à la coloration au bleu de coomassie colloïdal (les analyses quantitatives ayant été réalisées sur des gels colorés au nitrate d'argent). Néanmoins, parmi les 16 protéines identifiées, plusieurs protéines sont des composants de la chaîne respiratoire mitochondriale, d'autres, des enzymes du métabolisme mitochondrial (pyruvate deshydrogénase, ornithine amino transférase) et une autre, une protéine intervenant dans la traduction protéique (EF1y). Par ailleurs, de nombreuses protéines cytosoliques ont aussi été identifiés (peroxysomal enoyl CoA hydratase) et ceci malgré les nombreuses étapes de purification des mitochondries, elles sont toujours co-purifiées avec la fraction mitochondriale, ce qui tend à indiquer que ces protéines peuvent avoir un lien avec la pathologie, puisqu'elles semblent se lier aux mitochondries. De plus, la variation d'expression des protéines cytosoliques peut être liée directement à la mutation du tRNA lui-même. Par exemple, la sur-expression du facteur de traduction $EF1\gamma$ cytosolique, peut être le reflet d'une augmentation globale de la traduction cytosolique avec pour but de compenser les troubles au niveau mitochondrial, processus déjà décrit dans de nombreux cancers [Lew, 1992], [Ender, 1993], [Mimori, 1995], [Mathur, 1998] ou pour les cellules soumises à un stress oxydatif [Chacko, 1998]. Ainsi, l'information majeure de cette étude réside dans le fait que 6 protéines de la chaîne respiratoire mitochondriale sont sous-exprimées et que les protéines cytoplasmiques subissent aussi les conséquences de la mutation du tRNA Lysine. En effet, il semblerait que la traduction au niveau cytoplasmique soit autorégulée en fonction des traductions réalisées dans la mitochondrie (figure 24). Ainsi, les protéines d'un même complexe, bien que constitué à la fois de protéines issues du génome mitochondrial et du génome nucléaire, sont affectées de la même façon (sur-ou sous-exprimées).



Figure 24: Implication des protéines identifiées dans le fonctionnement mitochondrial

2.5 Résultats complémentaires : Caractérisation de modifications posttraductionnelles

Dans cette étude, certaines protéines ont été identifiées plusieurs fois (spots adjacents, figure 25), comme la Pyruvate deshydrogénase E1.

Figure 25: Isoformes de la Pyruvate déshydrogénase E1 α (spots 20, 28, 31)



Il est probable que ces isoformes correspondent à des modifications post-traductionnelles. En effet, ces modifications, comme les phosphorylations, les acétylations ou encore les méthylations modifient les pI des protéines et ont déjà été décrites pour de nombreuses maladies telles que la cataracte congénitale, la maladie de Parkinson ou d'autres maladies métaboliques [MacCoss, 2002], [Przedborski, 2001], [Mills, 2001]. Afin de répondre à cette nouvelle question, il sera intéressant d'analyser systématiquement tous les spots correspondants à ces isoformes et de les caractériser par MS/MS afin de trouver les acides aminés portant des modifications. Une première vérification a été entreprise en regardant attentivement les spectres MALDI-MS et en tentant de rechercher les pics qui auraient pu correspondre à un peptide trypsique avec une modification post-traductionnelle.

Certaines masses ont été identifiées comme correspondant à celles d'un peptide trypsique avec une ou plusieurs modifications post-traductionnelles (phosphorylations) (figure 26), il s'agit à présent de les analyser par MS/MS pour confirmer ces hypothèses.

Figure 26: Etude MALDI-MS des phosphorylations de la pyruvate deshydrogénase E1 α

Pyruvate deshydrogenase E1 ss unité α , mitochondriale				
Spot 28 :3 peptides, 8 phosphorylations suspectées 2249.16 m/z 4-22 : S(9), S(13), S(16), S(21), (4 phosphorylations) 1768.86 m/z 275-288 : S(275), T(286), Y(287) (3 phosphorylations) 1283.72 m/z 314-323 : S(314) (1 phosphorylation)				
Spot 20: 3 peptides, 3	B phosphorylations suspectées			
1033.4 m/z	67-73 : T(70) (1 phosphorylation)			
1334.6 m/z	74-83 : Y(82) (1 phosphorylation)			
1153.53 m/z	268-277 : Y(273) ou S(275) (1 phosphorylation)			
Spot 31: 3 peptides, 0	Spot 31: 3 peptides, 6 phosphorylations suspectées			
1477.82 m/z	1477.82 m/z 29-40 : T(35) (<i>1 phosphorylation</i>)			
1688.73 m/z	1688.73 m/z 120-132 : T(124), T(126), S(130) (<i>3 phosphorylations</i>)			
1673.97 m/z	1673.97 m/z 128-141 : S(130), T(139) (<i>2 phosphorylations</i>)			

Une première étude nanoLC-MS/MS sur le spot 28 (figure 27) n'a pas permis de séquencer les peptides susceptibles d'être modifiés. Pour cette analyse, la sélection des ions qui devaient être fragmentés par spectrométrie de masse (MS/MS) se faisait suivant un programme standard, qui sélectionnait et fragmentait seulement les 3 ions les plus intenses. Ainsi, d'une part, aucun des peptides suspecté être phosphorylé après l'analyse MALDI-MS n'a été observé sur les spectres nanoLC-MS, ce qui peut être dû à l'hydrophobicité de ces peptides mais surtout aux charges négatives apportées par les phosphates, rendant plus délicate la détection de ces peptides par nanoLC-MS/MS (problèmes d'ionisation). D'autre part, les peptides portant les sites préférentiels de

phosphorylation n'avaient pas été observés par MALDI-MS (peptides minoritaires, problèmes de suppression de signal).



Ainsi, les analyses tentées par nanoLC-MS/MS n'ont pas permis de déterminer si les peptides du site actif étaient ou non phosphorylés et ceci malgré un mode d'ionisation différent du MALDI. Plusieurs hypothèses peuvent être émises quant à ces premiers résultats, les peptides d'intérêt sont hydrophobes, trop grands ou en faible quantité. Mais, dans notre cas, si l'hydrophobicité des peptides peut gêner à leur extraction du gel ou générer une forte rétention sur la colonne de phase inverse, les problèmes liés à la détection de ces peptides sont certainement dûs aux charges négatives apportées par les phosphates, rendant leur ionisation difficile. Pour vérifier la présence de ces peptides phosphorylés, l'approche décrite par B. Chait pourra être envisagée, isoler les peptides d'intérêts en "aveugle" [Kalkum, 2003]. Il faudra donc isoler préférentiellement ces ions à l'aide d'un programme d'inclusion des ions. Un tel programme permettra de sélectionner en priorité certains ions de m/z définis, quelque soit leurs intensités. Le choix des masses à intégrer dans ce programme pourra être orienté par le fait que les Sérines 232, 293 et 300 sont connues comme ayant un rôle dans la régulation de l'activité enzymatique. Cette activité serait modulée au niveau de 3 sites de phosphorylations préférentiels [Arden, 1996].

Si l'approche nanoLC-MS/MS automatique ne parvient pas à isoler ce peptide, l'approche nanospray offline pourrait permettre de les isoler puisque dans ce cas les paramètres du spectromètre de masse sont optimisés pour le peptide d'intérêt, qui pourra dans certains cas aussi être isolé en "aveugle", approche aussi envisageable par MALDI MS/MS.

La stratégie IMAC [Ueda 2003], chromatographie d'affinité dirigée contre les peptides phosphorylés en amont de l'analyse MS/MS, pourrait aussi permettre d'isoler l'ensemble des peptides phosphorylés. Cependant, cette approche nécessite des échantillons concentrés, ce qui n'est pas le cas ici, puisque l'échantillon déposé sur gel, 500µg de mitochondries, demandait plus d'un mois de culture cellulaire. De ce fait, il sera donc difficile de tester toutes ces approches afin de déterminer laquelle est la plus adaptée à l'étude des peptides phophorylés. Neanmoins, l'approche nanoLC-MS/MS semble la plus adaptée.

2.6 Perspectives : Quantification par marquage des protéines mitochondriales

L'étude présentée est basée sur une quantification informatique à partir des gels 2D. Cependant, bien que cette approche soit très performante, elle possède un biais lié aux propriétés intrinsèques des gels 2D. En effet, les protéines hydrophobes, de pI extrêmes, de très faibles ou très haut poids moléculaires ne sont pas ou difficilement détectées par cette approche. Or, un grand nombre de protéines essentielles pour le métabolisme de la mitochondrie est enchassé dans les membranes et est par conséquent très hydrophobe. Pour contourner ces limites, tout en gardant la possibilité de réaliser une quantification, l'approche ICNAT (figure 28) est envisagée. Ainsi, grâce au marquage des protéines mitochondriales, leur quantification sera possible. L'étape de séparation par gel 2D est éliminée, par conséquent, cette approche donne accès à toute une série de protéines supplémentaires, ce qui permettra de compléter l'étude. Le nombre de peptides à analyser étant très important, il faudra ajouter une étape de séparation supplémentaire, soit avec un gel 1D, soit avec une seconde séparation chromatographique, chromatographie échangeuse de cations par exemple avant la chromatographie en phase inverse.



Figure 28: Approche protéomique par marquage ICNAT

2.7 Résultats complémentaires concernant la maladie MELAS

Les résultats de protéomique différentielle obtenus pour la maladie MERRF ayant permis de valider cette approche, la même étude a été envisagée pour une autre maladie mitochondriale, la maladie MELAS (Mitochondrial Encephalopathy with Lactic Acidosis and Strocke-like Episodes). Cette maladie est due, comme la maladie MERRF, à la mutation d'un tRNA, ici, le tRNA Leucine au niveau A3243G.

Si cette étude semble identique à la précédente, la réalisation des gels 2D est beaucoup plus délicate, car cette mutation affecte beaucoup plus les mitochondries, ce qui rend la préparation des échantillons très peu homogène et l'obtention de gels 2D peu reproductibles. En fait, ces cellules mutées produisent de l'acide lactique en grande quantité, qui dégraderait les protéines dans les cellules mais aussi au moment de l'extraction des mitochondries. Malgré tout, deux zones semblaient peu varier d'une expérience à l'autre, ce qui a permis d'en étudier les protéines sur et sous exprimées (tableau 14).



Tableau 14: protéines mitochondriales sur et sous exprimées dans le cadre de la maladie MELAS au niveau de deux zones du gel 2D



Les protéines sur et sous exprimées dans le cadre de la maladie MELAS sont pour certaines identiques (ornithine amino-transférase) à celles identifiées dans le cadre de l'étude de la maladie MERRF. Il s'agira par la suite d'établir, comme pour la maladie MERRF un schéma général permettant de comprendre l'influence au niveau protéique de la mutation du tRNA mitochondrial.

Conclusions générales

L'approche MALDI-MS a d'une part, permis de cartographier 87 protéines de noyaux des cellules Caco2/TC7 ainsi que 68 protéines exprimées de façon différentielle en fonction des états de différenciation des cellules. Ce protéome pourra servir de référence pour les nombreuses études in vitro réalisées sur ces cellules. Cependant, cette étude n'a pas permis d'identifier les protéines minoritaires (facteurs de transcriptions) détectées par d'autres tests biologiques (Western Blot). Afin d'avoir accès à ces protéines, il s'agira de fractionner l'échantillon ou d'ajouter des étapes de séparations supplémentaires (nanoLC-MS/MS) pour diminuer la complexité du mélange.

D'autre part, il a également été possible, grâce à cette approche MALDI-MS, d'identifier 16 protéines sur ou sous exprimées dans les mitochondries mutées responsables de la maladie MERRF. L'identification de ces protéines a permis de mieux comprendre les incidences de la mutation du tRNA Lysine mitochondrial au niveau cellulaire. Par ailleurs, les études protéomiques différentielles pourront être complétées par des études par marquage ICNAT, donnant accès aux protéines de pI extrêmes et hydrophobes. De plus, il serait intéressant de caractériser les modifications posttraductionnelles (phosphorylations) de certaines isoformes protéiques décrites comme ayant un rôle biologique important dans l'activité mitochondriale (pyruvate deshydrogénase E1 α)

Chapitre 3 :

Approche NanoLC-MS/MS

1 Caractérisation par homologie de séquences de protéines de *Streptococcus bovis* responsables de l'inflammation du côlon

Cette étude, réalisée en collaboration avec J.P. Klein et M. Schöller de l'unité INSERM U392, Illkirch, avait pour but d'identifier d'une part, les différentes protéines de *Streptococcus bovis* exposées à la surface ou sécrétées par la bactérie et d'autre part de déterminer celles qui sont responsables de l'inflammation du côlon pour connaître leurs rôles dans la cancérogénèse colique.

1.1 Contexte biologique



Figure 29 : Streptococcus bovis

Streptococcus bovis est une bactérie commensale (figure 29), Gram+, non invasive présente dans l'environnement, y compris dans l'alimentation (laitages). Elle devient un constituant de la flore intestinale après ingestion. Cependant, cette bactérie peut devenir pathogène suite à des lésions et conduire à des bactériémies se manifestant plus particulièrement chez les personnes atteintes d'immunodéficience. Il apparaît également, selon diverses études, que 25% des cas de bactériémies et d'endocardites à *Streptococcus bovis* sont associées à un adénocarcinome du côlon [Beebe, 1995]. De plus, la proportion de *Streptococcus bovis* présente dans les selles de patients présentant un adénocarcinome du côlon est augmentée en comparaison avec les autres bactéries [Klein, 1977].

Les premières études *in vitro* sur les conséquences d'une co-culture entre ces bactéries et diverses lignées cellulaires humaines (épithéliales coliques (Caco2) et buccales (KB) ou monocytaires (THP-1)) ont permis de montrer que *Streptococcus bovis* entraîne la libération de cytokines proinflammatoires, l'Interleukine 8 (IL-8) par les cellules epithéliales et de TNF- α (tumor necrosis factor) par les monocytes [Ellmerich, 1999]. D'autres expériences préliminaires *in vivo* sur des rats traités avec l'azoxyméthane (carcinogène chimique spécifique de l'intestin) ont montré que si la bactérie n'est pas par elle même un facteur déclenchant du processus cancéreux, sa présence dans le côlon stimule l'expression de marqueurs de la prolifération cellulaire (PCNA, proliferative cell nuclear antigen et polyamines) et accélère le développement de lésions précancéreuses préexistantes. Elle peut également entraîner la libération d'IL-8 dans la muqueuse [Ellmerich, 2000]. Il a également été montré que des extraits protéiques de paroi bactérienne ou Wall Extracted Antigens (WEA) produisaient les mêmes effets proinflammatoires et procancérogènes [Ellmerich, 1999], [Ellmerich, 2000] et qu'une inflammation chronique pouvait favoriser le développement d'un cancer [Coussens et Werb, 2001].

Il était donc intéressant d'identifier par spectrométrie de masse les protéines conduisant à la libération d'IL-8 pour pouvoir ensuite les tester dans un modèle de cancérogénèse. Mais l'approche protéomique MALDI-MS ne pouvait pas être utilisée en raison de la faible représentativité de *Streptococcus bovis* dans les banques de données, ce qui mène souvent l'identification par comparaison de masses peptidiques à l'échec. En effet, actuellement il n'existe que 6 entrées dans Swiss Prot et 31 entrées dans TrEMBL spécifiques de cette espèce (contre respectivement 498 et 4848 entrées pour *Streptococcus pneumoniae*)

1.2 Fractionnement et tests biologiques

Plusieurs fractions bactériennes ont été testées puis analysées (figure 30). Tout d'abord, la fraction WEA, extraite par agitation du culot bactérien en présence de billes de verre [Scholler M] est ensuite séparée par chromatographie anionique (MonoQ). Les fractions induisant une libération d'IL-8 par les cellules Caco-2 sont rassemblées en une fraction MonoQ qui sera elle-même séparée par chromatographie d'exclusion (Sephacryl S300) donnant la fraction active S300. A chaque étape, des gels mono-dimensionnels ont été réalisés. Des anticorps polyclonaux ont aussi été produits chez le

lapin contre une protéine de très haut poids moléculaire présente dans la fraction WEA, fraction des protéines excrétées, la fraction MonoQ et la fraction S300 et ont été utilisés pour réaliser une colonne d'immuno-affinité. La protéine1 est purifiée par séparation du WEA sur cette colonne d'immuno-affinité.

Après incubation de S*treptococcus bovis* dans un tampon PBS pendant 4h à 37°C puis après une centrifugation des bactéries, le surnageant contenant les protéines excrétées est récupéré.





Au cours de ces différentes étapes de purification, des tests *in vitro* sont réalisés afin de déterminer quelles sont les fractions actives, c'est à dire ayant une activité proinflammatoire, mesurée par leur capacité à induire une libération d'IL-8 par les cellules Caco2. Si ces premiers tests s'avèrent positifs, des tests *in vivo* sont réalisés (figure 31). Ces tests consistent tout d'abord, à induire un cancer du côlon chimiquement chez des rats par l'azoxyméthane puis à administrer par gavage les différentes fractions de protéines bactériennes et enfin à vérifier leurs effets sur le nombre de cryptes anormales et l'apparition de tumeurs au niveau de la muqueuse colique. Les rats témoins sont traités à l'azoxyméthane seul. Après ces deux types de tests, les fractions actives *in vitro* et *in vivo* sont analysées par spectrométrie de masse pour tenter d'identifier les protéines présentes dans ces fractions et à terme déterminer laquelle ou lesquelles sont responsables de cette activité.





1.3 Choix du protocole d'analyse : Analyse de la fraction S300

L'étude a débuté par l'analyse de gels 1D, car seules quelques bandes étaient visibles, ce qui pouvait signifier qu'il y avait peu de protéines présentes dans ces échantillons (figure 32A). Après analyse par spectrométrie de masse nanoLC-MS/MS seules quelques protéines ont pu être identifiées par homologie de séquence par rapport au genre *Streptococcus*. L'identification des protéines n'était possible qu'en utilisant l'approche nanoLC-MS/MS avec laquelle il est possible d'obtenir des informations de séquences (si nécessaire par un séquençage *De Novo*). En effet, même si l'ensemble des différentes souches de Streptococques est phylogénétiquement proche, il est néanmoins possible que certaines séquences aient subi des mutations qui ne participeront pas à l'identification si elle n'est basée que sur les masses des peptides trypsiques et de leurs ions fragments, d'où l'intérêt d'un séquençage *De Novo*.

Figure 32A : Gel 1D de la fraction bactérienne S300



Par ailleurs, une autre approche a été envisagée, l'approche gel 2D pour ajouter une seconde séparation des protéines en fonction de leurs pI et vérifier si certaines protéines avaient pu ne pas être détectées par spectrométrie de masse. Par cette approche, il a été possible de visualiser un nombre de spots plus important (72 spots analysés par spectrométrie de masse contre 6 pour le gel 1D) pour la même fraction S300 (figure33A et B), tout en sachant que les protéines hydrophobes et de pI extrêmes étaient perdues pour l'analyse par spectrométrie de masse.

Figure 33A : Gel 2D de la fraction bactérienne S300



Figure 33B : identification des protéines de la fraction S300 par homologie de séquence après séparation sur gel 2D et nanoLC-MS/MS. ("agrégat" correspond à des protéines de faibles masses moléculaires qui ont été identifiées en haut du gel, présumant qu'elles sont restées agrégées)

spot	identification	% recouvrement	peptides
S300 1	dipeptidase (agrégats?)	6%	5/5
S300 2	Dpr (agrégats?)	15%	1/2
S300 3	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 4	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 5	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 6	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 7	Dpr (agrégats?)	15%	1/2
S300 8	Dpr (agrégats?)	14%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 9	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 10	putative NADP specific glutamate dehydrogenase	27%	8/11
S300 11	Dpr (agrégats?)	15%	1/2
	putative peroxide resistance protein (agrégats?)	15%	2/2
S300 12	putative dipeptidase	10%	4/6
S300 13	dipeptidase	13%	4/6
	putative dipeptidase	8%	5/6
S300 14	dipeptidase	12%	5/7
	putative dipeptidase	13%	4/9
S300 15	aucune		
S300 16	aucune		
S300 17	putative NADP specific glutamate dehydrogenase	20%	8/9
S300 18	putative NADP specific glutamate dehydrogenase	12%	5/5
S300 19	putative NADP specific glutamate dehydrogenase	13%	6/6
	glutamate dehydrogenase	15%	3/7
S300 20	putative dipeptidase	8%	4/5
	dipeptidase	11%	3/4
	Glucose 6 P isomerase	9%	2/4
S300 21	putative NADP specific glutamate dehydrogenase	4%	2/2
	Glucose 6 P isomerase	7%	2/4
S300 22	putative NADP specific glutamate dehydrogenase	14%	4/7
	Glucose 6 P isomerase	14%	3/4
S300 23	putative NADP specific glutamate dehydrogenase	13%	3/5
	Glucose 6 P isomerase	29%	8/10
S300 24	putative NADP specific glutamate dehydrogenase	31%	10/12
S300 25	putative NADP specific glutamate dehydrogenase	26%	9/11
S300 26	putative NADP specific glutamate dehydrogenase	32%	13/13

S300 27	putative NADP specific glutamate dehydrogenase	30%	11/12
S300 28	putative NADP specific glutamate dehydrogenase	33%	12/13
S300 29	putative NADP specific glutamate dehydrogenase	33%	13/14
S300 30	putative NADP specific glutamate dehydrogenase	32%	11/12
S300 31	putative NADP specific glutamate dehydrogenase	36%	14/15
S300 32	putative NADP specific glutamate dehydrogenase	34%	13/13
S300 33	putative NADP specific glutamate dehydrogenase	27%	9/10
S300 34	putative NADP specific glutamate dehydrogenase	15%	4/5
S300 35	putative NADP specific glutamate dehydrogenase	9%	13/14
S300 36	NADP specific glutamate dehydrogenase	12%	1/4
S300 37	putative NADP specific glutamate dehydrogenase	18%	6/7
	glutamyl amino peptidase	12%	4/5
S300 38	putative NADP specific glutamate dehydrogenase	15%	6/6
	glutamyl amino peptidase	12%	4/5
S300 39	Glucose 6 P isomerase (Ct)	6%	2/3
	putative dipeptidase Ct	6%	2/3
S300 40	Glucose 6 P isomerase (Ct)	12%	2/4
	dipeptidase Ct	12%	1/5
S300 41	putative cysteine aminopeptidase C (fgt)	4%	2/2
S300 42	class II aldolase	26%	6/7
	putative aminopeptidase C (fgt int)	4%	2/2
	Dpr	6%	1/1
	putative NADP specific glutamate dehydrogenase		
S300 43	(Nt)	14%	4/5
	class II aldolase (Ct)	10%	1/3
	Glucose 6 P isomerase (Ct)	7%	2/3
S300 44	aucun		
S300 45	groEL (Ct)	3%	2/2
	Glucose 6 P isomerase (Ct)	3%	2/2
S300 46	Glucose 6 P isomerase (Ct)	6%	3/3
	triose P isomerase	8%	1/2
\$300.47	putative NADP specific glutamate denydrogenase (Ct)	10%	2/3
S300 48	superoxide dismutase	31%	Δ/Δ
0000 40	aroEL (Ct)	12%	2/5
\$300.49	nutative cysteine aminopentidase C (Nt)	6%	2/3
0000 40	Glucose 6 P isomerase (Ct)	10%	3/5
	putative NADP specific glutamate dehydrogenase	1070	0/0
S300 50	(Nt)	15%	5/5
	aminopeptidase C (Ct)	5%	1/2
	putative NADP specific glutamate dehydrogenase		
S300 51	(Nt)	26%	11/11
S300 52	aucun		
S300 53	aucun		
S300 54	dipeptidase (Nt)	11%	3/6
S300 55	groEL (Ct)	14%	4/6
S300 56	peroxide resistance protein Dpr	6%	1/1
S300 57	peroxide resistance protein Dpr	6%	1/1
S300 58	Glucose 6 P isomerase (Ct)	6%	3/3

S300 59	enolase (Nt)	10%	2/3
	groEL (Ct)	6%	2/3
S300 60	groEL (Ct)	9%	4/5
	class II aldolase (Ct)	9%	2/2
S300 61	enolase (Nt)	6%	2/2
	class II aldolase (Ct)	7%	1/2
	Glucose 6 P isomerase (fgt int)	4%	2/2
	Dpr (Nt)	8%	2/2
	L lactate dehydrogenase (Ct)	8%	2/3
S300 62	enolase (fgt int)	10%	1/3
	putative peroxide resistance protein	15%	2/2
S300 63	glyceraldehyde 3 P dehydrogenase (Nt)	26%	4/8
S300 64	L lactate dehydrogenase (Ct)	10%	2/4
S300 65	class II aldolase (Ct)	12%	2/3
	enolase (Nt)	7%	1/2
S300 66	class II aldolase (Ct)	21%	5/6
	enolase (fgt int)	14%	1/5
	glyceraldehyde 3 P dehydrogenase (Nt)	19%	4/5
S300 67	Dpr (Nt)	6%	1/1
	putative NADP specific glutamate dehydrogenase		
S300 68	(Nt)	22%	7/9
S300 69	glyceraldehyde 3 P dehydrogenase (Nt)	19%	7/7
	class II aldolase (Ct)??	19%	
S300 70	glyceraldehyde 3 P dehydrogenase (Nt)	25%	7/8
	pyruvate kinase (Nt)	5%	3/3
S300 71	glyceraldehyde 3 P dehydrogenase (Nt)	15%	4/5
	class II aldolase (Ct)??	11%	2/4
S300 72	class II aldolase	17%	5/6
	Dpr	17%	1/2
	NADP specific glutamate dehydrogenase (fgt int)	18%	3/7

Par conséquent, l'approche gel 2D a été choisie pour les études suivantes, puisqu'elle nous permet d'avoir accès à un plus grand nombre d'informations. Certaines protéines de faibles masses moléculaires et qui ont été identifiées en haut du gel sont notées agrégat, présumant du fait qu'elles sont restées agrégées au cours de la préparation de l'échantillon.

1.4 Fraction WEA

46 spots issus de la fraction WEA ont été analysés par nanoLC-MS/MS (figure 34A et B). La plupart des identifications a été possible, grâce aux homologies de séquences avec les autres souches de Streptocoques. Il n'est cependant pas exclu que ces protéines n'aient pas exactement les mêmes séquences, la souche ayant permis l'identification protéique par homologie de séquence pouvant être légèrement différente, ce qui expliquerait parfois les faibles recouvrements de séquences obtenus. De plus, de nombreuses isoformes ont été détectées, ce qui peut s'expliquer par des modifications post-

traductionnelles. Mais, si des glycosylations ont déjà été décrites au niveau de la paroi des Streptocoques, dans notre cas la plupart des isoformes rencontrées avaient des masses moléculaires inférieures à celles attendues. L'hypothèse avancée pour expliquer la présence de ces nombreux fragments protéiques est la préparation de l'échantillon. En effet, à chaque étape, il peut se produire des fragmentations de protéines (extraction mécanique par des billes de verre, succession de lyophilisation, de congélation, passage dans des tampons acides lors de l'élution de la protéine 1).

De plus, de façon surprenante, la majorité des protéines identifiées sont des protéines du métabolisme cellulaire (Glucose 6 phosphate Isomérase..) que l'on pourrait imaginer localisées dans le cytoplasme. La présence de telles protéines à la surface des bactéries a déjà été décrite [Hughes, 2002]. Ces protéines, déjà identifiées pour une activité enzymatique donnée, semblent posséder d'autres activités. Certaines de ces enzymes du métabolisme seraient impliquées dans l'adhésion (Ornithine carbamoyltransferrase: [Hussain M, 1999]), ou agiraient comme une neuroleukine (Glucose 6 Phosphate Isomérase avec l'activité de l'autocrine motility factor)) ou encore comme une transporteuse de fer (glycéraldéhyde-3P dehydrogenase) [Modun, 1999].





Figure 34B : Protéines de la fraction WEA identifiées par homologie de séquence après nanoLC-MS/MS. (les protéines dites hybrides ont été identifiées à l'aide des séquences de 2 souches de streptocoques différentes, la séquence de *Streptococcus bovis* pour certaines protéines semblant être un hybride de 2 séquences de Streptocoques)

spot	identification	%	peptides
		recouvreme	
		nt	
WEA 1	aucune		
WEA 2	aucune		
WEA 3	lysyl amino peptidase	3%	2/3
WEA 4	amino peptidase N (hybride?)	8%	4/6
	amino peptidase N (hybride?)	7%	3/6
	lysyl amino peptidase (hybride? 5 pep différents)	6%	3/6
WEA 5	ATP dpt CLP protease proteolytic sub unit	26%	10/16
WEA 6	hsp 70 (DNA K)	33%	19/21
WEA 7	hsp 60 (GroEL, Cpn60)	48%	24/26
WEA 8	GroEl	39%	17/17
	dipeptidase	8%	3/3
WEA 9	Pyruvate kinase	28%	11/11
	GMP synthase	17%	2/6
WEA 10	Pyruvate kinase	32%	10/12
	NADP dpt glyceraldehyde 3 P dehydrogenase	10%	3/5
WEA 11	Proton translocating ATPase α sub-unit	32%	13/15
	Pyruvate kinase	12%	4/5
	NADP dpt glyceraldehyde 3 P dehydrogenase	11%	2/5
	NADP dpt glyceraldehyde 3 P dehydrogenase	14%	2/6
WEA 12	Glutamine synthetase type 1	22%	6/9
	Glucose 6 P isomerase	16%	3/5
	enolase (qq différence) (hybride?)	20%	6/6
	2 phosphoglycerate dehydratase (enolase) (hybride?)	18%	6/6
	Proton translocating ATPase β sub-unit	8%	3/3
WEA 13	Glucose 6 P isomerase	25%	7/8
	putative aminopeptidase C	9%	3/3
	Enolase	44%	13/15
WEA 14	Enolase	52%	18/18
WEA 15	Glyceraldehyde 3 P dehydrogenase (hybride?)	37%	12/12
	Glyceraldehyde 3 P dehydrogenase (2 pep diff,) (hybride?)	37%	13/13
	Enolase?? (contamination)	47%	14/16
WEA 16	Glyceraldehyde 3 P dehydrogenase	38%	14/15
	Glyceraldehyde 3 P dehydrogenase (3 pep diff,) (hybride?)	48%	14/15
WEA 17	Glyceraldehyde 3 P dehydrogenase	48%	13/15
	Phosphoglycerate kinase	26%	8/9
WEA 18	Glyceraldehyde 3 P dehydrogenase	50%	12/15
	Phosphoglycerate kinase	33%	11/12
WEA 19	Phosphoglycerate kinase	57%	18/20
	Glyceraldehyde 3 P dehydrogenase	36%	9/12
WEA 20	Phosphoglycerate kinase (contamination)	39%	11/11
	Proline dipeptidase (fgt?)	9%	2/2
	Pyruvate kinase (Nt)	20%	4/7
	aspartate aminotransferase (Ct?)	22%	4/4
WEA 21	L lactate dehydrogenase	30%	9/11
	Glyceraldehyde 3 P dehydrogenase (fgt)	20%	4/5

WEA 22	L lactate dehydrogenase	28%	10/12
	Glyceraldehyde 3 P dehydrogenase (fgt)	24%	5/6
WEA 23	glyceraldehyde 3 P dehydrogenase (fgt)	33%	7/9
	6 Phosphofructokinase (Nt)	30%	5/10
WEA 24	Glucokinase? (peu de signal)	10%	1/3
WEA 25	class II aldolase (hybride)	40%	13/13
	class II aldolase (hybride) (6 pep diff,)	25%	7/7
WEA 26	class II aldolase (hybride)	29%	5/8
	class II aldolase (hybride) (6 pep diff,)	29%	3/8
	glucose 1P thymidyltransferase	22%	4/6
WEA 27	Glyceraldehyde 3Pdehydrogenase (Nt)	22%	4/6
WEA 28	2,3 biphosphoglycerate dpt phosphorglycertae mutase	32%	7/8
	ribosomal prot L3 (incohérent,pl)	29%	6/6
	ribosomal prot L2 (incohérent, pl) (fgt int?)	25%	3/4
WEA 29	class II aldolase (hybride) (fgt Ct?)	21%	4/5
	ribosomal prot L3	15%	2/2
WEA 30	aucune		
WEA 31	Triosephosphate isomerase	11%	2/3
	superoxide dismutase (peu de signal)	6%	1/1
WEA 32	Alkyl hydroperoxidase	13%	3/3
WEA 33	Alkyl hydroperoxidase (faible signal)	29%	1/3
WEA 35	Polypeptide deformylase	14%	2/3
WEA 36	superoxide dismutase (fgt?)	6%	1/1
WEA 37	DTDP 4 keto L rhamnose reductase (fgt int)	22%	4/5
	orotate phosphoribosyl transferase (Nt)	13%	3/4
	DTDP 4 keto 6 deoxyglucose 3,5 epimerase (fgt int)	12%	2/2
WEA 38	class II aldolase (Ct)	19%	3/5
	ribosomal prot L4	19%	2/3
WEA 39	ribosomal prot L4	11%	2/2
	ATP dpt C1P protease proteolytic sub unit (int)	33%	3/3
WEA 40	aucune		
WEA 42	Glyceraldehyde 3 P dehydrogenase (hybride?)	13%	2/4
	Glyceraldehyde 3 P dehydrogenase (hybride?)	12%	2/4
WEA 43	aucune identification cohérente		
WEA 45	UDP N acetylmuramoylalanyl D glutamate??(faible signal)	25%	2/9
WEA 46	50S ribosomal L2	20%	3/4
	L lactate dehydrogenase (Nt)	20%	4/5
	Glyceraldehyde 3 P dehydrogenase (fgt)	13%	3/4

1.5 Fraction excrétée

De la même façon, les protéines bactériennes excrétées ont été analysées, puisqu'elles provoquent aussi une réponse *in vitro*. Les résultats sont présentés dans la figure 35 (A et B). La majorité des protéines identifiées sont identiques à celles identifiées dans le cas de la fraction WEA.



Figure 35A : Gel 2D et identification des protéines bactériennes excrétées

Figure 35B : identification par homologie de séquence des protéines bactériennes excrétées après nanoLC-MS/MS

spot	identification	% recouvrement	peptides
1	Dpr (agrégats??)	16%	2/3
2	Dpr (agrégats??)	16%	2/3
4	Dpr (agrégats??)	15%	1/2
5	aucune		
6	hsp 70 DNA K	24%	17/18
7	ATP dpt protease ATP binding sub unit	23%	7/13
8	hsp 70 DNA K	6%	2/4
10	GroEL	16%	6/7
11	GroEL	23%	11/11
	enolase	36%	11/13
12	aucune		
13	GroEL	18%	1/8!!!!
14	Zn binding protein adcA precursor	18%	4/9
16	GroEL	10%	6/6
	EfTu	7%	2/2
	putative dipeptidase	5%	2/2
17	peptidyl prolyl isomerase RopA	14%	5/7
18	peptidyl prolyl isomerase RopA	17%	5/8

			1
20	glucose 6 P isomerase	10%	3/4
21	aucune		
22	aucune		
23	glyceraldehyde 3P dehydrogenase	45%	14/14
24	phosphoglycerate kinase	34%	12/13
25	phosphoglycerate kinase	20%	5/6
	Ef Ts	11%	4/4
26	phosphoglycerate kinase	29%	7/10
27	phosphoglycerate kinase	21%	5/7
28	aucune		
29	D ala D ala ligase	13%	2/4
30	enolase	17%	5/5
31	aucune		
32	Zn binding protein adcA precursor (Nt)	7%	3/3
33	GroEL (Ct)	7%	3/4
34	Zn binding protein adcA precursor (Nt)	15%	3/7
35	L lactate dehydrogenase	27%	7/7
36	6 phosphofructokinase	13%	3/4
37	aucune	1070	0,1
38	Manganese ABC transporter	15%	2/5
30	Manganese ABC transporter	22%	2/3
40		10%	2/1
40		16%	2/5
41		210/	4/3
42		3170	7/7
43		10%	1/1
44		31%	10/12
45		4%	1/1
46	fructose 1,6 biphosphate aldolase	20%	1/5
47	aucune		
48	aucune		
49	pas d'identification cohérente		
50	hsp70 (tgt??)		
51	class II aldolase	16%	3/4
52	aucune		
53	aucune		
54	aucune		
55	aucune		
56	pas d'identification cohérente		
57	aucune		
58	superoxide dismutase	28%	3/3
59	class II aldolase	24%	5/8
60	class II aldolase	10%	2/3
61	aucune		
62	peptide deformylase	6%	2/2
63	polypeptide deformylase	14%	2/3
64	ribosome recycling factor	28%	5/7
65	ribosome recycling factor	17%	3/4
66	superoxide dismutase	28%	3/3
67	superoxide dismutase	71%	9/9
68	superoxide dismutase	39%	5/5
69	chromosome IV reading frame ORF!!	35%	6/7
70	50S ribosomal prot L6	18%	3/4
71	50S ribosomal prot L6	10%	2/2
		- · · •	· ·

	superoxide dismutase	37%	3/4
	peroxide resistance prot Dpr	6%	1/1
72	class II aldolase (Ct)	12%	2/3
	putative peroxide resistance prot	21%	2/3
73	peroxide resistance prot Dpr	15%	2/2
	Dpr ??	15%	1/2
74	Dpr	6%	1/1!!
75	superoxide dismutase	15%	1/1
76	DNA binding prot HU	35%	2/3
77	30S ribosomal prot S6	42%	4/4
78	30S ribosomal prot S6	23%	2/2
79	phosphocarrier prot HPr	13%	2/2
80	50S ribosomal prot L9 (Ct)	14%	2/2
	50S ribosomal prot L13 (Ct) (6%	1/1
	HPr	11%	1/1
81	aucune		
82	30S ribosomal prot S8	71%	6/9
	HPr	11%	1/1
83	HPr	13%	2/2
	ribosomal protein L22	34%	3/3
84	HPr	11%	1/1

1.6 Fraction "Protéine 1" : Utilisation de l'approche Séquençage De Novo

La fraction « protéine1 » (figure 36), obtenue après séparation du WEA sur la colonne d'immuno-affinité, décrite précédemment, a été séparée sur gel 1D. Son profil permet d'identifier une bande majoritaire à 48kDa (la protéine1) et d'autres bandes minoritaires qui ont été analysées par nanoLC-MS/MS (figure 38). Les recherches par homologies de séquences n'ont rien donné. Par conséquent, l'approche par séquençage *De Novo* (figure 37) a été réalisée suivie d'une recherche par alignement de séquences par BLAST. Différentes protéines ont pu être identifiées, principalement des protéines membranaires (figure 38).





Figure 37 : Séquence obtenue après analyse d'un spectre MS/MS de la fraction « protéine 1 ». Peptide à 478.2 m/z du spot 3. L'interprétation a été réalisée par PEAKS®



Figure 38 : Protéines de la fraction « protéine 1 » après analyse par BLAST

spot	identification	numéro accession
1	Dpr	Q99YU7
	Ton B dpt receptor	Q8PGQ6
	putative glutamine ABC transporter	Q8DVC6
2	Ton B dpt receptor	Q8P6U8
3	Ton B dpt receptor	Q8P6U8
4	Ton B dpt receptor	Q8P6U8
5	outer membrane protein	Q8PNP2
6	outer membrane protein	Q83QU7
7	outer membrane protein	Q8PNP2
	extra cellular protease precursor	P23314
8	aucun résultat	

Lors de cette étude, les recherches par alignement de séquences obtenues par séquençage *De Novo* ont été difficiles. Il s'agira donc, pour valider l'identification de la protéine TonB
de séquencer une partie de la protéine présente dans cette fraction par séquençage d'Edman, si la fraction est suffisamment pure. Par ailleurs, afin de s'affranchir de la séparation par gel, il est envisageable de réaliser directement après la chromatographie d'affinité une séparation par colonne échangeuse d'ions (nanoLC-LC-MS/MS). Cette approche permettrait peut-être de mieux visualiser les protéines minoritaires.

1.7 Perspectives : Identification plus fine des protéines responsables de l'inflammation

1.7.1 Tests biologiques

Tout d'abord, bien que la fraction S300 montre une activité proinflammatoire et procancérogène, il reste à déterminer s'il existe ou non une corrélation entre ces deux activités. Pour cela, chaque protéine identifiée devra être clonée puis testée individuellement dans une expérience de cancérogenèse chez le rat. Certaines protéines semblent plus intéressantes, par exemple la gluose-6-phosphate isomérase qui a déjà montré une activité importante sur la mobilité de certaines cellules, activité qui semble importante dans la formation des métastases [Chia-Cheng Chou, 2000]. On peut encore citer la protéine dpr (Dps [DNA binding protein from starved cells]-like peroxide resistance gene) qui fait partie d'une famille de dps, dans laquelle on retrouve la protéine HP-NAP de *helicobacter pylori* qui possède une forte activité dans le recrutement de neutrophiles [Satin B, 2000]. Or ces neutrophiles sont des sources de radicaux libres connus pour induire des mutations au niveau de l'ADN. Cette protéine induit également la production de radicaux libres de manière directe en stimulant la NADPH oxydase membranaire de cellules humaines.

En ce qui concerne la protéine1, une identification plus fine est nécessaire. La localisation du gène dans *Streptococcus bovis* et son clonage permettra son séquençage complet. Elle est actuellement testée dans une expérience *in vivo*.

Enfin, l'utilisation d'inhibiteurs reste envisagée mais on ne peut exclure que ces inhibiteurs agissent sur l'activité enzymatique (connue au niveau du métabolisme) et non sur l'activité potentielle (proinflammatoire).

1.7.2 Caractérisation par spectrométrie de masse

Une fois que les protéines d'intérêt seront identifiées, il sera intéressant de vérifier la séquence de la protéine par une étude de séquençage *De Novo*. En effet, si les protéines peuvent être isolées spécifiquement par chromatographie d'affinité, la quantité de matériel sera peut-être suffisante pour réaliser une étude complète de séquençage pour vérifier que la séquence est homologue avec celle de

l'espèce Streptococcus ayant permis l'identification. D'autre part, cette analyse pourra aussi servir à déterminer les modifications post-traductionnelles de ces protéines.

Par la suite, il serait intéressant d'identifier les cibles de ces protéines sur les cellules humaines, particulièrement sur les cellules Caco2 et d'en déterminer précisément les partenaires d'interaction. Les complexes d'interaction entre les protéines bactériennes et les cellules humaines pourraient être analysés par spectrométrie de masse.

Enfin, la cartographie des noyaux des cellules Caco2 ayant été réalisée lors d'une autre étude (cf chapitre 1, partie III), il serait intéressant de réaliser une étude protéomique différentielle sur les noyaux de ces cellules en présence et en absence de la fraction bactérienne. Cette étude permettrait de déterminer les protéines nucléaires sur- et sous- exprimées dans le cadre de l'exposition des cellules à la fraction bactérienne. Si par la suite, la cartographie complète des cellules Caco2 est réalisée, la même étude pourra être envisagée.

Chapitre 4:

Approche par séquençage De Novo

1 Caractérisation de protéines parasitaires de Plasmodium falciparum

Cette étude a été réalisée en collaboration avec Catherine Braun-Breton de l'Institut Pasteur (Biologie des Interactions Hôte-Parasite, Département de Parasitologie) à Paris et Thierry Rabilloud du CEA à Grenoble. Elle avait pour but de mieux comprendre le fonctionnement global de *Plasmodium falciparum*, en utilisant les différentes techniques de l'analyse protéomique.

1.1 Contexte biologique

Plasmodium falciparum est un parasite protozoaire, responsable de la malaria, qui représente un problème de santé publique mondial (figure 39), surtout en raison de l'apparition de moustiques résistants aux insecticides classiques et de parasites de plus en plus insensibles aux médicaments. Il devient donc urgent de découvrir de nouvelles stratégies pour lutter et comprendre les voies majeures du cycle parasitaire qui aideraient à la découverte de bonnes cibles pour les drogues futures. L'objectif de l'équipe de l'Institut Pasteur est de comprendre le fonctionnement biologique global du parasite dans les cellules parasitées dans l'optique de découvrir à terme de nouvelles cibles thérapeutiques pour des drogues aux modes d'action inhabituels. Le cycle parasitaire se déroule majoritairement dans les globules rouges et les cellules hépatiques, lieux de l'activité morbide des parasites. Il peut être schématisé par la figure 40.





Figure 40 : Cycle de multiplication de Plasmodium falciparum



En fait, de petits organels sécrétoires sont impliqués dans les étapes essentielles du cycle parasitaire. L'invasion des hépatocytes et des globules rouges (Erythrocytes) dépend de sécrétions enzymatiques à l'extrémité apicale de ces organels parasitaires, les Rhoptries, Micronèmes et Granules denses. De plus, le développement intra-érythrocytaire du parasite, la libération de parasites infectieux dépend de la libération de protéines parasitaires au niveau de la paroi érythrocytaire via les structures de Maurer. Une technique de préparation de ces compartiments a été développée par l'équipe de T. Rabilloud par lyse osmotique ménagée des globules rouges parasités. De ce fait, il a été possible d'envisager l'analyse protéomique globale de ces structures selon différentes approches, dans le but de mieux comprendre leur fonctionnement, avec un intérêt tout particulier pour les protéines avec une activité Sérine hydrolase. En effet, ces protéines sont spécifiques du parasite et il serait intéressant de les identifier afin de pouvoir, à terme, élaborer des drogues spécifiquement dirigées contre ces protéines.

- La première approche consiste à utiliser un inhibiteur spécifique des sérines hydrolases couplé à la biotine (FP-biotine) [Kidd, 2001], pour isoler spécifiquement ces protéines et les identifier par spectrométrie de masse.
- La seconde approche consiste en une analyse protéomique différentielle entre les protéines solubles des fantômes de globules rouges parasités et sains.
- La troisième approche consiste en une analyse protéomique globale des protéines parasitaires présentes dans les fantômes de globules rouges parasités. Afin de simplifier le mélange, les parasites ont été marqués au préalable avec de la Lysine deutérée.

1.2 Identifications de sérine hydrolases

1.2.1 Préparation des échantillons

Le but de cette première approche vise à caractériser l'ensemble des Sérine hydrolases parasitaires d'intérêt pour la libération des mérozoïtes (figure 41), leur maturation ou leur entrée dans l'érythrocyte. Ces protéines sont isolées par chromatographie d'affinité spécifique dirigée contre les Sérine hydrolases [Kidd, 2001]. Après avoir isolé les protéines d'intérêt, elles sont séparées par gel 1D et analysées par spectrométrie de masse. Dans notre cas, l'approche nanoLC-MS/MS a été envisagée car le génome de *Plasmodium falciparum* n'était pas complètement accessible au moment de cette première étude.



Figure 41 : Chromatographie d'affinité dirigée contre les protéines à activité Sérine hydrolase

1.2.2 Analyses par nanoLC-MS/MS

Trois enzymes ont plus particulièrement été analysées. L'approche nanoLC-MS/MS était nécessaire pour obtenir des informations de séquences avec des spectres de bonnes qualités pour pouvoir analyser les spectres un à un manuellement et en déduire des informations de séquence. Après une recherche par alignement de séquence par BLAST, PFI 1445w a été clairement identifiée (figures 42 et 43) ainsi que PF 13-0116 et PF08-0036. Ces protéines n'ont pas encore de fonction véritablement attitrée. L'identification de ces protéines n'a donc été possible qu'en réalisant une étude *De Novo*.



Figure 42 : Spectres MS/MS brut (Q-TOF) et interprété pour l'ion 573.7 m/z

m/z	charge	séquence	BLAST
401,1	2+	LA(FG)HEK	
459,2	2+	TTEYYLK	PFI 1445w
477,2	2+	QYVTTLTK	PFI 1445w
487,2	2+	LSLAPD(MA)R	PFI 1445w
493,7	2+	GENPAELEK	
495,2	2+	LSLAPD(M(ox)A)R	PFI 1445w
512,19	2+	?	
520,69	2+	LPLEDYYK	PFI 1445w
538,2	2+	YNFLDLYK	PFI 1445w
573,7	2+	QLDDEELER	PFI 1445w
575,1	2+	VSNLTND(EM)K	PFI 1445w
579,7	2+	WSDLTY M/F VK	PFI 1445w
592,7	2+	?	
597,7	2+	(YL)LDNSNPMK	PFI 1445w
610,7	2+	VFNGLPA M/F LDK	PFI 1445w
716,2	2+	_FDEY M/F M/F ASNK	PFI 1445w
718,25	2+	(VQ)LPE(GC)FG	

Figure 43 : Résultats des alignements de séquence (BLAST) pour la protéine PFI 1445w

1.2.3 Perspectives biologiques

L'identification protéique terminée, il s'agit à présent de vérifier la localisation de ces protéines au sein des fantômes d'érythrocytes parasités. Pour cela, des anticorps spécifiques des protéines ont été synthétisés. De plus, il s'agira aussi de confirmer l'activité de sérine hydrolases de ces protéines afin de valider, une nouvelle fois, la méthode de chromatographie d'affinité et de confirmer l'activité des protéines identifiées, activité qui n'est pas encore clairement définie.

1.2.4 Perspectives en spectrométrie de masse

Dans cette étude, le choix du programme de MS/MS a été primordial, puisqu'il fallait disposer de données spectrales de bonnes qualités facilement interprétables manuellement. En effet, différents tests ont été réalisés avant cette étude en utilisant plusieurs énergies de fragmentation (une rampe d'énergie) [thèse Delalande F.], pour obtenir des spectres MS/MS sur lesquels la majorité des fragments (faibles et hauts poids moléculaires) seraient présents.

Par la suite, il pourra être intéressant de compléter l'étude en augmentant le nombre de séquences obtenues pour les 3 protéines. En effet, si la protéine PFI 1445w a été identifiée avec plus d'une dizaine de séquences, les autres l'ont été avec seulement quelques séquences. Il s'agira donc d'utiliser des conditions d'analyse (temps de fragmentation plus longs, concentration de l'échantillon) favorisant l'obtention de spectres de très bonne qualité, qui pourront être analysés par séquençage *De Novo*.

1.3 Protéomique différentielle : Erythrocytes sains et parasités

L'approche protéomique différentielle décrite précédemment (partie II, chapitre2) a été utilisée. Après séparation sur gels 2D, de protéines issues de fantômes de globules rouges sains et parasités, les protéines parasitaires visibles ont été analysées par spectrométrie de masse.

Lors de l'analyse d'image, quatre spots ont été observés uniquement dans l'échantillon de fantômes de globules rouges parasités. Trois d'entre eux se sont avérés être des protéines de la famille des chaperons moléculaires, le dernier a permis d'identifier la protéine Disulfide isomérase (PDI) parasitaire.

Pour cette étude, les identifications ont pu être réalisées par recherche directe dans les banques de données protéiques avec les liste de masses des fragments, sans avoir recours aux alignements de séquences, les protéines étant, dans ce cas, référencées.

Figure 44 : Pourcentages de recouvrement de séquences des protéines issues de l'analyse différentielle entre globules rouges sains et parasités

spot	identification	numéro SP	% recouvrement
1	hsp 70	Q07615	34%
2	hsp 70	Q07615	19%
3	hsp 70	Q07615	24%
4	disulfide isomerase precurseur	Q9GRI2	40%

Comme pour l'étude précédente, il s'agira de vérifier la localisation de la protéine PDI (Disulfide protein isomerase) dans les structures de Maurer à l'aide d'anticorps spécifiques.

1.4 Etude systématique des protéines parasitaires enchassées dans la membrane de l'érythrocyte ou à proximité

1.4.1 Méthode : Marquage des parasites

L'identification globale des protéines parasitaires à proximité ou présentes dans les membranes de globules rouges a pour but de caractériser les structures de Maurer, pour avoir à terme une meilleure compréhension de son rôle dans le développement du parasite au sein de la cellule hôte.

Après un marquage des parasites et l'infection des globules rouges, les protéines de fantômes de globules rouges parasités sont séparés sur gel 1D. Le gel est coupé de façon systématique (64 bandes) et analysé par spectrométrie de masse MS/MS (figure 45). Seules les protéines parasitaires

nous intéressaient pour cette étude, c'est pourquoi elles ont été marquées pour simplifier le mélange au moment de l'interprétation des résultats. Ainsi, les recherches dans les banques de données sont réalisées automatiquement en ajoutant une modification fixe dans les logiciels de recherche.



Figure 45 : protocole de préparation des échantillons

1.4.2 Résultats

Les protéines identifiées sont présentées dans le tableau 15.

 Tableau 15: Résultats de l'analyse après marquage des protéines parasitaires enchassées ou à proximité de la membrane des globules rouges

spot	identification	% recouvrement	peptides
2	contamination Hb		
	60S acidic ribosomal prot P2	22%	1/1
3	contamination Hb		
	Sexual stage spe prot precurseur	33%	2/2
	60S acidic ribosomal prot P2	21%	1/1
	40S ribosomal subunit prot S14	41%	4/5
	40S ribosomal subunit prot S16	36%	5/6
4	aucune		
5	60S ribosomal prot L11a	26%	4/5
	40S ribosomal prot S15	20%	2/2
	40S ribosomal prot S12	16%	2/2
6	40S ribosomal prot S12	41%	4/6
	60S ribosomal prot L11a	32%	4/6
	60S ribosomal protein L12	35%	4/5
	ribosomal protein L14	29%	3/5
7	ribosomal protein S19S	58%	6/9
	hypothetical protein	34%	5/9
8	hypothetical protein	25%	5/6
	40S ribosomal prot S11	30%	3/6

	ribosomal protein S19	22%	3/3
	40S ribosomal prot S5	15%	3/3
9	60S ribosomal prot L18	46%	5/10
· ·	40S ribosomal prot S9	36%	5/7
	hypothetical protein	49%	4/11
10	40S ribosomal prot S7	38%	5/7
-	malaria prot Exp 1 prec,	29%	5/6
11	ribosomal protein L13	38%	5/11
12	aucune		-
13	ribosomal prot S3	31%	5/7
	ribosomal prot L10	42%	5/10
	hypothetical protein	33%	2/8
14	ribosomal prot S3	54%	9/12
	ribosomal prot L10	46%	4/8
15	hypothetical protein	49%	8/10
	60S ribosomal protein L6e	56%	6/12
	GTP binding prot RAN/TC4	43%	5/11
	purine nucleoside phosphorylase	27%	3/5
	ribosomal prot S19s	41%	5/6
	hypothetical protein	42%	4/9
16	HGXPRT	22%	6/6
	triose phosphate isomérase	8%	2/2
17	RESA like prot	48%	12/19
	ribosomal prot L10	49%	4/11
18	RESA like prot	48%	8/20
	Adenylate kinase	26%	5/5
	calcyclin binding prot	12%	3/3
	ADP/ATP transporter/ adenylate translocase	14%	4/4
	60S ribosomal prot L7	15%	3/3
	ribosomal prot S4	10%	2/2
19	ribosomal prot S4	10%	2/2
	Phosphoethanolamine N methyl transferase	15%	3/3
	merozoite surface protein 7	14%	3/3
20	L lactate dehydrogenase	53%	7/13
	prot homologue	23%	4/4
	ribosomal prot S8e	28%	4/4
~ ·	60S ribosomal prot L8	12%	3/3
21	L lactate dehydrogenase	58%	9/10
	40S ribosomai prot S3A	54%	7/19
	14-3-3 prot nomologue	54%	0/10
22	Liaciate denyorogenase	じ <u>ン</u> % 120/	0/12 2/2
		10%	5/5 2/2
00	evported pret	200/	3/3
23	ribosomal prot L 5	J∠70 160/	2/16
	ribosomal protein L73	40%	5/10
	alveeraldebyde 3 D debydrogenase	+5 % 25%	2/2
	hypothetical protein	20%	Δ/7
24	divceraldehvde 3 P dehvdrogenase	59%	9/16
27	guanine nucleotide binding prot	25%	5/9
	exported prot 2	26%	6/7
	ribosomal prot L 5	44%	5/12
	putative prot	15%	2/5
	hypothetical protein	33%	3/8
25			
	•	•	•

26			
20	40S ribosomal prot S6	28%	4/7
27	40S ribosomal prot S6	20%	7/7
21	ribo phosphoprotein P0	49%	7/11
	60S prot P0	26%	3/7
	alvcerol 3 P dehvdrogenase	2070	3/13
28	ribosomal Phospho protein P0	34%	7/8
20	alvcerol 3 P dehydrogenase	21%	2/6
	40S ribosomal prot	36%	2/0 4/7
	hypothetical protein	8%	2/2
29	hsn	12%	<u> </u>
30	hsp	22%	4/9
31	fructose biP aldolase	38%	6/9
0.	rhontry associated prot	22%	4/8
32	rhoptry associated prot	21%	6/6
02	nhosnhoribosyl pyronhosnhate synthetase	8%	3/3
33		070	0/0
34			
35			
36	hypothetical protein	24%	5/8
27		24 /0	5/0 6/10
37	elioiase skeleton binding prot	20%	3/5
	bypothetical protein	27 /0	2/6
20	alongation factor 1a	20/	2/0
30	digulfide inomerane pres	0 70	2/2
39	skoloton binding prot	10%	2/2
40	digulfide inomerane pres	10 /0	Z/Z 5/5
40		70/	0/0 1/0
11		1 70	1/2
41	aucurie bypothetical protein	100/	2/0
42		19%	3/0
43			
44			
45	aucune	070/	E M C
46	nsp 70	27%	5/16
47	v type H+ translocating pyrophosphatase	14%	4/10
47	aucune	000/	5/40
48	nsp 70	22%	5/16
40	GRP78	21%	4/12
49	nypotnetical protein	19%	3/10
50	aucune erythrocyte mb	400/	0/45
51	nsp	16%	3/15
50	endoplasmin	15%	2/12
52	aucune CD233 hu	000/	0/4.4
53	knob associated his rich protein	20%	3/11
5.4	GBP 130	8%	3/5
54	knob associated his rich protein	20%	2/11
	GBP 130	10%	4/10
55	knob associated his rich protein	14%	2/8
56	serine repeat Ag prot	21%	3/19
	nypotnetical protein	12%	4/7
	KNOD ASSOCIATED NIS FICH PROTEIN	14%	3/9
		24%	2/18
5/	knob associated his rich protein	22%	4/13
	nypothetical protein	28%	4/32

	SERA Ag	32%	13/29
58	SERA Ag	26%	10/22
	hypothetical protein	28%	3/32
	cystein protease	22%	2/17
59	serine repeat Ag prot	26%	4/21
	high molecular weight rhoptry prot 2	3%	4/4
60	high molecular weight rhoptry prot 2	27%	2/36
61	high molecular weight rhoptry prot 2	24%	1/28
62	major merozoite surface Ag	21%	2/33
63	aucune spectrin hu		
64	hypothetical protein	14%	2/19

Il faut à présent vérifier les localisations de ces différentes protéines. D'ores et déjà, il est intéressant de constater que ces protéines peuvent être classées suivant quatre catégories :

- Protéines solubles sans signal de sécrétion ni signal d'ancrage membranaire
- Protéines avec signal de sécrétion, sans signal d'ancrage membranaire : HRP1
- Protéines sans signal de sécrétion mais avec signal d'ancrage : PfSBP1
- Protéines avec signal d'ancrage et de sécrétion : 3 protéases de la famille SERA

1.4.3 Limites de l'approche et perspectives

Les différentes études menées sur *Plasmodium falciparum* n'auraient pas pu être réalisées en utilisant l'approche MALDI-MS. En effet, les premières études nécessitaient des informations de séquences pour pouvoir réaliser l'identification des protéines après un séquençage *De Novo* (figure 47). D'autre part, quelques tests MALDI-MS (figure 46) ont été tentés pour l'analyse globale des protéines parasitaires marquées, mais la complexité de l'échantillon n'a pas permis d'identifier les protéines parasitaires puisqu'elles étaient masquées par les protéines de globules rouges.

Figure 46 Spectre MALDI-MS et identification du spot 56





Figure 47 Spectre nanoLC-MS du spot 17, agrandi de l'ion à 404.23 (2+) m/z avec et sans marquage de la lysine

Ainsi sans séparation LC au préalable, les protéines humaines masquent intégralement le signal des protéines d'intérêt. Pour pallier à cette limite, le couplage LC-MALDI-MS aurait pu être une solution bien que l'approche MALDI-MS/MS n'étant pas encore automatisée de façon efficace, cette étude aurait été très fastidieuse. Par contre, elle nous aurait permis de disposer d'informations supplémentaires sur les fragments « w et d» afin de lever les ambiguités liées aux acides aminés isobares.

Par ailleurs, il serait intéressant d'éliminer les protéines de globules rouges qui masquent très souvent les protéines parasitaires. De ce fait, il serait peut-être possible d'identifier des protéines plus minoritaires. Ainsi, afin de compléter l'étude protéomique globale, un marquage des protéines parasitaires avec cette fois-ci un marqueur permettant une séparation par chromatographie d'affinité (ex :ICAT) en aval de l'analyse par spectrométrie de masse permettrait d'éliminer une majorité de protéines humaines. Une telle approche, comporte évidemment de nombreuses limites liées à la sélectivité du marquage qui, dans le cas d'un marquage ICAT ne sera pas réalisé sur toutes les protéines mais uniquement sur les protéines à Cystéines (cf limites de l'approche ICAT).

L'approche par marquage des Lysines est donc très intéressant, puisqu'il s'agit d'un acide aminé essentiel qui sera incorporé dans toutes les protéines. Par contre, il n'est pas possible de réaliser une chromatographie d'affinité dirigée contre un acide aminé deutéré. Donc pour l'instant il n'y a pas de possibilité de simplifier encore plus le mélange, si ce n'est en réalisant une nanoLC-LC-MS/MS.

Une autre possibilité serait de modifier les moteurs de recherches. En effet, s'il est possible de réaliser des recherches dans les banques de données en ajoutant des modifications sur les acides aminés, cette modification n'est pas considérée comme fixe, c'est à dire que les résultats tiendront compte de l'acide aminé avec et sans modification, dans notre cas, les recherches seront réalisées avec les Lysines marquées et non marquées. Par conséquent, nous retrouverons aussi dans la liste des résultats, les protéines non marquées.

Conclusions générales

Grâce à l'approche nanoLC-MS/MS et aux recherches dans les banques de données par homologie de séquence, il a été possible d'identifier 72 protéines exposées à la surface de Streptococcus bovis et 84 protéines excrétées par la bactérie. Grâce aux fractionnement des échantillons, il a été possible de réduire le nombre des protéines candidates responsables de l'activité anti-inflammatoire. A présent, il s'agira par des études biologiques de déterminer la ou les protéines responsables de l'inflammation du côlon pour mieux comprendre leurs rôles dans la cancérogénèse colique.

Parallèlement, dans le but de mieux comprendre le fonctionnement global de Plasmodium falciparum (parasite responsable de la Malaria), cette approche suivie d'un séquençage De Novo et de recherches par alignement de séquences a permis d'identifier :

- 3 protéines avec une activité enzymatique de Sérine hydrolase, qui n'étaient pas encore décrites.

- 4 protéines exprimées intensément dans les globules rouges parasités par rapport aux globules rouges sains.

- 64 protéines parasitaires présentes à la surface ou enchassées dans la membrane du globule rouge Même si le génome n'était pas entièrement disponible lors de cette étude, le fractionnement des échantillons et l'approche par séquençage De Novo a permis de caractériser plusieurs protéines dont les fonctions biologiques n'ont pas encore été décrites avec certitude. Par ailleurs, il serait intéressant pour identifier un plus grand nombre de protéines ainsi que des protéines minoritaires, dans le cadre de l'analyse protéomique différentielle, de fractionner un peu plus l'échantillon, afin d'éliminer les protéines chaperonnes.

Ainsi, grâce à l'identification de ces protéines, il sera possible de mieux comprendre le fonctionnement global du parasite pour à terme, élaborer de nouvelles drogues spécifiques.

Partie IV :

La spectrométrie de masse, un outil de choix pour identifier les protéines mais pas seulement...

1 Quantification de molécules pharmacologiques dans les sérums de patients

Cette étude a été réalisée en collaboration avec C. Moog de l'Institut de Virologie (INSERM U 74) de Strasbourg, dans le but d'identifier de nouveaux facteurs antiviraux.

1.1 Contexte biologique

Le VIH (Virus de l'Immunodéficience Humaine) responsable du SIDA (Syndrome d'Immunodéficience Acquise) est une préoccupation majeure de l'organisation mondiale de la santé. En effet, certains pays d'Afrique voient leurs populations infectées à 35%. Il s'agit donc d'un véritable problème de santé publique. Si le cycle de développement global du virus est de mieux en mieux connu (figure 48), le développement de drogues contre le virus reste difficile. En effet, l'enveloppe de ce virus est en perpétuelle mutation pour éviter les mécanismes de défense immunitaires.

Figure 48 Cycle du développement viral



Si des antiviraux associés en multi-thérapie sont utilisés comme traitement pour diminuer la charge virale, ils ne permettent pas d'éradiquer le virus. De plus les coûts de ces traitements rendent impossible leurs utilisations dans les pays les plus touchés, à savoir les pays en voie de développement. Ainsi, les recherches continuent pour tenter d'élaborer un vaccin en stimulant la réponse immunitaire, par exemple en aidant à la synthèse d'anticorps dirigés contre le virus. Dans ce contexte, l'Institut de Virologie tente de comprendre les mécanismes de défense des individus contre le virus VIH en étudiant dans les sera et plasmas de patients infectés et en recherchant ceux qui résistent au virus. Les détections de sera et plasmas résistants sont réalisées à l'aide de tests immunologiques [Burrer, 1999]. Au cours de ces études, si la plupart des sera et plasmas possédaient une activité antivirale humorale (anticorps actifs contre le virus), deux d'entre eux possédaient une activité qui n'était pas due aux anticorps, mais certainement à des facteurs solubles.

Certains facteurs avaient déjà été décrits comme ayant un rôle dans l'inhibition de la multiplication virale : les β chimiokines, RANTES, MIP 1 α , MIP 1 β , a chimiokines et CAF [Barker, 1999], [Amara, 1997]. De plus, une nouvelle catégorie de protéines, avec une activité antivirale, ont récemment été identifiées, les Défensines [Zhang, 2002].

Le but de cette étude était donc d'identifier les facteurs actifs présents dans ces deux sera. Les études préliminaires de ces deux sera (dialyses, chromatographie d'affinité et tests d'activités) ont permis d'établir que leur activité était due à des facteurs solubles de masses moléculaires comprises entre 10 et 50kd [Burrer, 1999]. Les facteurs solubles connus ont été dosés sans montrer de taux exceptionnellement élevés.

1.2 Premières études HPLC: l'hypothèse d'une protéine active réfutée

Pour identifier ce facteur, une approche par LC-MS a été envisagée. Après purification (élimination des grosses particules) du sérum sur colonne de phase inverse (Sep Pak®, C18), l'éluat (100% d'ACN) a été séparé par HPLC (colonne de phase inverse C18, 1mm Diamètre Interne (D.I.)). Chaque pic, observé par le détecteur UV (214nm) a été récolté et testé *in vitro* sur des cellules humaines infectées par le virus, afin de détecter les fractions actives contre le virus. Le but de cette purification était de simplifier le mélange pour trouver la fraction active. Une fois le pic contenant la fraction active déterminée, il était ainsi possible de la séparer par gel 2D et d'analyser toutes les protéines présentes sur le gel par spectrométrie de masse. A terme, la ou les protéines responsables de cette activité antivirale aurait pu être identifiée.

Malheureusement, l'hypothèse d'un facteur protéique responsable de cette activité a été abandonnée pour plusieurs raisons :

- Elution des agents actifs à 50% d'acétonitrile (caractère hydrophobe peu cohérente avec l'hypothèse d'un facteur soluble)
- Conservation de l'activité même après un passage dans des solvants dénaturants (ACN, acide formique)
- Patients traités par des anti-viraux, ce qui n'était pas connu au moment des études préliminaires (particulièrement un inhibiteur non nucléosidique de la reverse transcriptase (Efavirenz®))

Après avoir éliminé l'intervention d'une protéine anti-virale spécifique dans les deux sera étudiés, il nous a semblé intéressant de doser la molécule (Efavirenz®) active chez ces patients par LC-MS.

Figure 49 : Molécule d'Efavirenz®



1.3 Dosage de la molécule d'intérêt par spectrométrie de masse

1.3.1 Préparation des échantillons (extraction)

Avant d'analyser cette molécule, il était important de l'extraire des sera et plasmas pour être capable de la doser. De plus, il était indispensable de développer une méthode (figure 50) de préparation de l'échantillon qui permette l'élimination de contaminants susceptibles de posséder une activité antivirale (par exemple, anticorps dirigés contre le virus)

Après centrifugation de l'échantillon avec de l'éthanol, le surnageant est lyophilisé et repris dans de l'eau. Bien que cette préparation aboutisse à un échantillon propre, un filtre ainsi qu'une précolonne ont été installés en amont de la colonne HPLC (1mm D.I.), afin d'éviter toute altération du matériel.

1.3.2 Validation de la méthode

Afin d'améliorer la sensibilité de la méthode, différents diamètres de colonnes HPLC ont été testées : 2mm DI (diamètre interne) et 1mm DI. Les résultats obtenus étaient de meilleures qualités (gain de sensibilité) avec les colonnes de 1mm de DI. De plus, différents gradients ont été testés afin d'avoir une bonne séparation et un temps d'élution le plus court possible.

Après ces premiers tests, une courbe de calibration externe (figures 51 et 52) a été élaborée avec un produit de référence, la molécule (Efavirenz®) pure, afin d'estimer la quantité de produit actif dans les échantillons. Cet étalon a subi la même préparation que le plasma ou le sérum.

L'analyse a été effectuée comme suit :

- Ajouts dosés d'Efavirenz® dans 10µl d'H2O
- Analyse LC-MS
- Mesure de l'intensité du courant d'ions générés par la molécule
- Report de cette intensité sur un graphe Intensité = f(Concentration)





Figure 51: Chromatogramme d'ions et spectre MS (LC-TOF) à 316.2 m/z







1.3.3 Dosage de la molécule d'intérêt dans les deux sera

Une fois la gamme étalon élaborée, l'analyse LC-MS des sera a permis de doser la molécule d'intérêt. Les résultats sont indiqués dans la figure53.

Figure 53: Courbe de calibration externe et dosage des sera de patients



Cette étude préliminaire nous a permis de constater que notre approche était suffisamment sensibles pour les deux dosages effectués.

1.4 Perspectives

1.4.1 Utilisation d'un étalon interne

Le dosage par LC-MS pourrait être amélioré en utilisant un étalon interne, par exemple la molécule d'intérêt deutérée. Cette molécule possèderait le même temps de rétention que la molécule à doser et un processus d'ionisation identique [Shi, 2003]. De plus, la précision du dosage serait améliorée et les temps d'analyses raccourcis (calibrations externes inutiles). Ainsi, il serait possible de réaliser une lecture directe de la concentration de l'analyte par comparaison des intensités entre la molécule deutérée (étalon) et la molécule hydrogénée (molécule à doser). Il s'agit donc de l'étalon idéal qui devrait être utilisé en priorité [Xu, 2002], [Mastuoka, 2003]. Une étude comparative de quantification avec un étalon interne deutéré qu'avec un étalon interne de structure analogue à la molécule à quantifier [Morrison, 2002]. Cependant, dans notre cas, l'étude n'a pu être réalisée avec un étalon interne deutéré en raison du coût trop élevé de cette molécule.

1.4.2 Analyse d'un mélange de molécules anti-virales

La méthode mise au point dans cette étude, pourra servir à doser un cocktail d'anti-viraux chez les patients, ce qui pourra être utile dans le cas du suivi des malades. En effet, l'intérêt des dosages réside dans le fait que les variations de concentrations, par exemple d'Efavirenz® dans le sang peuvent avoir des conséquences dramatiques pour les patients (contrôle de la virémie, effets secondaires importants...). Il est donc important de vérifier que les prises de ces molécules se font de manière correcte.

Par ailleurs, en fonction de la complexité des échantillons, il serait intéressant de développer une méthode par MS/MS, en dosant les ions fils générés par fragmentation des molécules d'intérêt. Ceci permettrait un gain de spécificité de la méthode. Par contre, le mode MS/MS donne des signaux de plus faible intensité que le mode MS. Malgré tout, en mode MS/MS, en plus du gain de spécificité, le bruit de fond observé est atténué, le rapport signal/bruit est augmenté aboutissant à un gain de sensibilité. Il faudra cependant s'assurer que les molécules à doser peuvent être fragmentées à basse énergie par MS/MS.

Conclusions générales

Grâce au couplage de la microLC et de la spectrométrie de masse, il a été possible d'élaborer une méthode de dosage d'une molécule pharmacologique (Efavirenz®) dans les sera de patients séropositifs résistants au VIH. Cette méthode de dosage pourra servir à contrôler, rapidement, la prise de médicaments chez ces patients.

Bibliographie :

A. Amara, S. L. Gall, O. Schwartz, J. Salamero, M. Montes, P. Loetscher, M. Baggiolini, J. L. Virelizier and F. Arenzana-Seisdedos

HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication, *J Exp Med*, **1997**, 186, 139-46.

S. D. Arden, B. O. Roep, P. I. Neophytou, E. F. Usac, G. Duinkerken, R. R. de Vries and J. C. Hutton

Imogen 38: a novel 38-kD islet mitochondrial autoantigen recognized by T cells from a newly diagnosed type 1 diabetic patient, *J Clin Invest*, **1996**, 97, 551-61.

E. Barker

CD8+ cell-derived anti-human immunodeficiency virus inhibitory factor, *J Infect Dis*, **1999**, 179 Suppl 3, S485-8.

E. Bartova, S. Kozubek, P. Jirsova, M. Kozubek, H. Gajova, E. Lukasova, M. Skalnikova, A. Ganova, I. Koutna and M. Hausmann

Nuclear structure and gene activity in human differentiated cells, *J Struct Biol*, **2002**, 139, 76-89.

E. Bartova, S. Kozubek, M. Kozubek, P. Jirsova, E. Lukasova, M. Skalnikova and K. Buchnickova

The influence of the cell cycle, differentiation and irradiation on the nuclear location of the abl, bcr and c-myc genes in human leukemic cells, *Leuk Res*, **2000**, 24, 233-41.

J. L. Beebe and E. W. Koneman

Recovery of uncommon bacteria from blood: association with neoplastic disease, *Clin Microbiol Rev*, **1995**, 8, 336-56.

R. Burrer, D. Salmon-Ceron, S. Richert, G. Pancino, G. Spiridon, S. Haessig, V. Roques, F. Barre-Sinoussi, A. M. Aubertin and C. Moog

Immunoglobulin G (IgG) and IgA, but also nonantibody factors, account for in vitro neutralization of human immunodeficiency virus (HIV) type 1 primary isolates by serum and plasma of HIV-infected patients, *J Virol*, **2001**, 75, 5421-4.

G. Chacko, Q. Ling and K. A. Hajjar

Induction of acute translational response genes by homocysteine. Elongation factors-1alpha, - beta, and -delta, *J Biol Chem*, **1998**, 273, 19840-6.

Chantret, A. Rodolosse, A. Barbat, E. Dussaulx, E. Brot Laroche, A. Zweibaum and M.Rousset,

Differential expression of sucrase-isomaltase in clones isolated from early and late passages of the cell line Caco-2: evidence for glucose-dependent negative regulation. *J.Cell Sci.* 1994. 107 (Pt 1), 213 225.

M. Chevallet, E. Wagner, S. Luche, A. van Dorsselaer, E. Leize-Wagner and T. Rabilloud

Regeneration of peroxiredoxins during recovery after oxidative stress: only some overoxidized peroxiredoxins can be reduced during recovery after oxidative stress, *J Biol Chem*, **2003**, 278, 37146-53.

H. Colognato and P. D. Yurchenco

Form and function: the laminin family of heterotrimers, Dev Dyn, 2000, 218, 213-34.

P. L. Courchesne, R. Luethy and S. D. Patterson

Comparison of in-gel and on-membrane digestion methods at low to sub-pmol level for subsequent peptide and fragment-ion mass analysis using matrix-assisted laser-desorption/ionization mass spectrometry, *Electrophoresis*, **1997**, 18, 369-81.

L. M. Coussens and Z. Werb

Inflammatory cells and cancer: think different!, J Exp Med, 2001, 193, F23-6.

T. Cremer and C. Cremer

Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nat Rev Genet*, **2001**, 2, 292-301.

A. De Arcangelis, O. Lefebvre, A. Mechine-Neuville, C. Arnold, A. Klein, L. Remy, M. Kedinger and P. Simon-Assmann

Overexpression of laminin alpha1 chain in colonic cancer cells induces an increase in tumor growth, *Int J Cancer*, **2001**, 94, 44-53.

F. Delalande,

Application du couplage chromatographie liquide-spectrométrie de masse à l'étude de la biodisponibilité de peptides issus de produits laitiers et à la protéomique, Thèse de l'Université Louis Pasteur de Strasbourg **2003**.

S. Ellmerich, N. Djouder, M. Scholler and J. P. Klein

Production of cytokines by monocytes, epithelial and endothelial cells activated by Streptococcus bovis, *Cytokine*, **2000**, 12, 26-31.

S. Ellmerich, M. Scholler, B. Duranton, F. Gosse, M. Galluser, J. P. Klein and F. Raul

Promotion of intestinal carcinogenesis by Streptococcus bovis, *Carcinogenesis*, **2000**, 21, 753-6.

D.Fenyo

Identifying the proteome: software tools, Curr Opin Biotechnol, 2000, 11, 391-5.

J. N. Freund, C. Domon-Dell, M. Kedinger and I. Duluc

The Cdx-1 and Cdx-2 homeobox genes in the intestine, Biochem Cell Biol, 1998, 76, 957-69.

F. Gharahdaghi, C. R. Weinberg, D. A. Meagher, B. S. Imai and S. M. Mische

Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity, *Electrophoresis*, **1999**, 20, 601-5.

H. Ginisty, H. Sicard, B. Roger and P. Bouvet

Structure and functions of nucleolin, J Cell Sci, 1999, 112, 761-72.

M. G. Hanna, I. P. Nelson, J. A. Morgan-Hughes and A. E. Harding

Impaired mitochondrial translation in human myoblasts harbouring the mitochondrial DNA tRNA lysine 8344 A-->G (MERRF) mutation: relationship to proportion of mutant mitochondrial DNA, *J Neurol Sci*, **1995**, 130, 154-60.

M. Heller, H. Mattou, C. Menzel and X. Yao

Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers, *J Am Soc Mass Spectrom*, **2003**, 14, 704-18.

M. J. Hughes, J. C. Moore, J. D. Lane, R. Wilson, P. K. Pribul, Z. N. Younes, R. J. Dobson, P. Everest, A. J. Reason, J. M. Redfern, F. M. Greer, T. Paxton, M. Panico, H. R. Morris, R. G. Feldman and J. D. Santangelo

Identification of major outer surface proteins of Streptococcus agalactiae, *Infect Immun*, **2002**, 70, 1254-9.

M. Hussain, G. Peters, G. S. Chhatwal and M. Herrmann

A lithium chloride-extracted, broad-spectrum-adhesive 42-kilodalton protein of Staphylococcus epidermidis is ornithine carbamoyltransferase, *Infect Immun*, **1999**, 67, 6688-90.

M. Ingman, H. Kaessmann, S. Paabo and U. Gyllensten

Mitochondrial genome variation and the origin of modern humans, *Nature*, **2000**, 408, 708-13.

E. Jung, C. Hoogland, D. Chiappe, J. C. Sanchez and D. F. Hochstrasser

The establishment of a human liver nuclei two-dimensional electrophoresis reference map, *Electrophoresis*, **2000**, 21, 3483-7.

M. Kalkum, G. J. Lyon and B. T. Chait

Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry, *Proc Natl Acad Sci U S A*, **2003**, 100, 2795-800.

D. Kidd, Y. Liu and B. F. Cravatt

Profiling serine hydrolase activities in complex proteomes, *Biochemistry*, 2001, 40, 4005-15.

R. S. Klein, R. A. Recco, M. T. Catalano, S. C. Edberg, J. I. Casey and N. H. Steigbigel Association of Streptococcus bovis with carcinoma of the colon, *N Engl J Med*, **1977**, 297, 800-2.

A. M. Kogelnik, M. T. Lott, M. D. Brown, S. B. Navathe and D. C. Wallace

MITOMAP: a human mitochondrial genome database--1998 update, *Nucleic Acids Res*, **1998**, 26, 112-5.

E. Krause, H. Wenschuh and P. R. Jungblut

The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins, *Anal Chem*, **1999**, 71, 4160-5.

W. M. Lauber, J. A. Carroll, D. R. Dufield, J. R. Kiesel, M. R. Radabaugh and J. P. Malone

Mass spectrometry compatibility of two-dimensional gel protein stains, *Electrophoresis*, **2001**, 22, 906-18.

Y. Liu, M. P. Patricelli and B. F. Cravatt

Activity-based protein profiling: the serine hydrolases, *Proc Natl Acad Sci U S A*, **1999**, 96, 14694-9.

M. J. MacCoss, W. H. McDonald, A. Saraf, R. Sadygov, J. M. Clark, J. J. Tasto, K. L. Gould, D. Wolters, M. Washburn, A. Weiss, J. I. Clark and J. R. Yates, 3rd

Shotgun identification of protein modifications from protein complexes and lens tissue, *Proc Natl Acad Sci U S A*, **2002**, 99, 7900-5.

S. Mathur, K. R. Cleary, N. Inamdar, Y. H. Kim, P. Steck and M. L. Frazier

Overexpression of elongation factor-1gamma protein in colorectal carcinoma, *Cancer*, **1998**, 82, 816-21.

K. Mills, P. B. Mills, P. T. Clayton, A. W. Johnson, D. B. Whitehouse and B. G. Winchester

Identification of alpha(1)-antitrypsin variants in plasma with the use of proteomic technology, *Clin Chem*, **2001**, 47, 2012-22.

K. Mimori, M. Mori, S. Tanaka, T. Akiyoshi and K. Sugimachi

The overexpression of elongation factor 1 gamma mRNA in gastric carcinoma, *Cancer*, **1995**, 75, 1446-9.

B. Modun and P. Williams

The staphylococcal transferrin-binding protein is a cell wall glyceraldehyde-3-phosphate dehydrogenase, *Infect Immun*, **1999**, 67, 1086-92.

L. M. Montuenga, J. Zhou, I. Avis, M. Vos, A. Martinez, F. Cuttitta, A. M. Treston, M. Sunday and J. L. Mulshine

Expression of heterogeneous nuclear ribonucleoprotein A2/B1 changes with critical stages of mammalian lung development, *Am J Respir Cell Mol Biol*, **1998**, 19, 554-62.

H. Morimoto, H. Okamura and T. Haneji

Interaction of protein phosphatase 1 delta with nucleolin in human osteoblastic cells, *J Histochem Cytochem*, **2002**, 50, 1187-93.

M. L. Ortiz, M. Calero, C. Fernandez Patron, C. F. Patron, L. Castellanos and E. Mendez

Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins, *FEBS Lett*, **1992**, 296, 300-4.

S. Przedborski, Q. Chen, M. Vila, B. I. Giasson, R. Djaldatti, S. Vukosavic, J. M. Souza, V. Jackson-Lewis, V. M. Lee and H. Ischiropoulos

Oxidative post-translational modifications of alpha-synuclein in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) mouse model of Parkinson's disease, *J Neurochem*, **2001**, 76, 637-40.

T. Rabilloud

Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis, *Electrophoresis*, **1990**, 11, 785-94.

T. Rabilloud

Detecting proteins separated by 2-D gel electrophoresis, Anal Chem, 2000, 72, 48A-55A.

T. Rabilloud, J. M. Strub, N. Carte, S. Luche, A. Van Dorsselaer, J. Lunardi, R. Giege and C. Florentz

Comparative proteomics as a new tool for exploring human mitochondrial tRNA disorders, *Biochemistry*, **2002**, 41, 144-50.

T. Rabilloud, J. M. Strub, S. Luche, A. van Dorsselaer and J. Lunardi

A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels, *Proteomics*, **2001**, 1, 699-704.

O. J. Rando, K. Zhao and G. R. Crabtree

Searching for a function for nuclear actin, Trends Cell Biol, 2000, 10, 92-7.

V. Santoni, M. Molloy and T. Rabilloud

Membrane proteins and proteomics: un amour impossible?, *Electrophoresis*, **2000**, 21, 1054-70.

B. Satin, G. Del Giudice, V. Della Bianca, S. Dusi, C. Laudanna, F. Tonello, D. Kelleher, R. Rappuoli, C. Montecucco and F. Rossi

The neutrophil-activating protein (HP-NAP) of Helicobacter pylori is a protective antigen and a major virulence factor, *J Exp Med*, **2000**, 191, 1467-76.

I. E. Scheffler

Mitochondria make a come back, Adv Drug Deliv Rev, 2001, 49, 3-26.

M. Scholler, J. P. Klein and R. M. Frank

Common antigens of streptococcal and non-streptococcal oral bacteria: immunochemical studies of extracellular and cell-wall-associated antigens from Streptococcus sanguis, Streptococcus mutans, Lactobacillus salivarius, and Actinomyces viscosus, *Infect Immun*, **1981**, 31, 52-60.

E. A. Schon, E. Bonilla and S. DiMauro

Mitochondrial DNA mutations and pathogenesis, J Bioenerg Biomembr, 1997, 29, 131-49.

S. Servidei

Mitochondrial encephalomyopathies: gene mutation, Neuromuscul Disord, 2001, 11, 774-9.

A. Shevchenko, M. Wilm, O. Vorm and M. Mann

Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal Chem*, **1996**, 68, 850-8.

G. Shi, J. T. Wu, Y. Li, R. Geleziunas, K. Gallagher, T. Emm, T. Olah and S. Unger

Novel direct detection method for quantitative determination of intracellular nucleoside triphosphates using weak anion exchange liquid chromatography/tandem mass spectrometry, *Rapid Commun Mass Spectrom*, **2002**, 16, 1092-9.

M. Srivastava and H. B. Pollard

Molecular dissection of nucleolin's role in growth and cell proliferation: new insights, *Faseb J*, **1999**, 13, 1911-22.

L. W. Sumner, B. Wolf-Sumner, S. P. White and V. S. Asirvatham

Silver stain removal using H2O2 for enhanced peptide mass mapping by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun Mass Spectrom*, **2002**, 16, 160-8.

Y. J. Sun, C. C. Chou, W. S. Chen, R. T. Wu, M. Meng and C. D. Hsiao

The crystal structure of a multifunctional protein: phosphoglucose isomerase/autocrine motility factor/neuroleukin, *Proc Natl Acad Sci USA*, **1999**, 96, 5412-7.

D. L. Tabb, L. L. Smith, L. A. Breci, V. H. Wysocki, D. Lin and J. R. Yates, 3rd

Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides, *Anal Chem*, **2003**, 75, 1155-63.

E. K. Ueda, P. W. Gout and L. Morganti

Current and prospective applications of metal ion-protein binding, *J Chromatogr A*, **2003**, 988, 1-23.

B. A. van Montfort, B. Canas, R. Duurkens, J. Godovac-Zimmermann and G. T. Robillard

Improved in-gel approaches to generate peptide maps of integral membrane proteins with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *J Mass Spectrom*, **2002**, 37, 322-30.

D. C. Wallace

Mitochondrial diseases in man and mouse, Science, 1999, 283, 1482-8.

M. R. Wilkins and K. L. Williams

Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation, *J Theor Biol*, **1997**, 186, 7-15.

C. C. Wu, M. J. MacCoss, K. E. Howell and J. R. Yates

A method for the comprehensive proteomic analysis of membrane proteins, *Nat Biotechnol*, **2003**, 21, 532-8.

Y. Yan-Sanders, G. J. Hammons and B. D. Lyn-Cook

Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells, *Cancer Lett*, **2002**, 183, 215-20.

L. Zhang, W. Yu, T. He, J. Yu, R. E. Caffrey, E. A. Dalmasso, S. Fu, T. Pham, J. Mei, J. J. Ho, W. Zhang, P. Lopez and D. D. Ho

Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor, *Science*, **2002**, 298, 995-1000.

A. Zweibaum, M. Laburthe, E. Grasset and D. Louvard

The use of cultured cell lines in studies of intestinal cell differentiation and function in Handbool of physiology intestinal transport of the gastrointestinal sytem FIELD M., FRIZZELL R.A., EDS, AMERICAN PHYSIOLOGICAL SOCIETY, 1991.

Conclusion :

Au cours de ce travail de thèse, nous avons optimisé les approches protéomiques visant à répondre à différents problèmes biologiques, en déterminant quelle était la technique de spectrométrie de masse, la plus adaptée pour répondre aux questions posées. Pour cela, nous avons développé trois approches protéomiques : approche MALDI-MS, approche nanoLC-MS/MS, approche par séquençage *De Novo* (figure 1).

Le choix de ces approches est déterminé notamment par les connaissances biologiques dont nous disposions au moment de l'étude. En effet, si les protéines étudiées sont issues de *génomes connus*, 2 approches pourront être choisies :

- L'approche MALDI-MS pour les études de protéomique descriptive (nécessitant une identification rapide des protéines)
- L'approche nanoLC-MS/MS pour les études de protéomique fonctionnelle (caractérisations fines des protéines, déterminations de modifications post-traductionnelles...)

Si les protéines sont issues de génomes inconnus ou partiellement connus :

- L'approche nanoLC-MS/MS, pour les études nécessitant des identifications par homologies de séquences
- L'approche par Séquençage *De Novo*, pour pouvoir identifier les protéines par alignement de séquences

Figure 20 : Stratégies d'identification de protéines.

stratégie MALDI-MS : comparaison de listes de masses de peptides trypsiques, stratégie nanoLC-MS/MS: comparaison de listes de masses des fragments de peptides trypsiques stratégie séquençage De Novo : alignement de séquences (BLAST)



Mais dans toutes ces approches, il a été nécessaire, auparavant, d'optimiser les étapes de coloration des gels, de digestion *in-gel* des protéines, d'extraction des peptides du gel et de MS/MS.

Optimisation de l'étape de coloration des gels 2D

Cette étude a permis de mieux comprendre les raisons pour lesquelles la coloration à l'argent est plus difficilement compatible avec la spectrométrie de masse. Ces raisons seraient liées à la présence de formaldéhyde dans les solutions de révélation des protéines et à la présence des ions argent. De ce fait, le nouveau protocole de coloration à l'argent élaboré sans formaldéhyde donne d'aussi bons résultats que les colorants classiques (ruthénium et bleu colloïdal) tant en sensibilité qu'en compatibilité avec la spectrométrie de masse.

Etapes de digestion et d'extraction des peptides du gel

Les tests d'extraction des protéines ont permis de déterminer les conditions optimales pour extraire les peptides trypsiques des gels, à savoir, 100% d'acétonitrile pendant une heure. Par ailleurs, l'utilisation de plusieurs enzymes de protéolyse (trypsine, chymotrypsine et AspN) permet d'obtenir des résultats complémentaires aboutissant à une amélioration du pourcentage de recouvrement de séquence protéique, en particulier, si les résultats de 2 digestions réalisées en parallèle sont cumulées (trypsine + chymotrypsine). En effet, dans ce cas, les problèmes de protéolyse d'une enzyme par l'autre sont évités et la taille des peptides générés restent compatibles avec l'analyse par spectrométrie de masse (<3000Da). Cette approche (utilisation de plusieurs enzymes) pourra être envisagée si le but de l'analyse protéomique est de caractériser l'ensemble d'une séquence protéique et si l'on dispose surtout de suffisamment de matériel.

Optimisation des approches MS/MS

Les tests réalisés sur les différentes approches MS/MS utilisées en analyse protéomique ont permis d'établir la complémentarité de ces trois approches : nanoLC-Q-TOF, nanoLC-Trappe ionique et MALDI-TOF/TOF. En effet,si les protéines sont identifiées avec les mêmes recouvrements de séquence, les peptides séquencés par nanoLC-MS/MS sont parfois différents en fonction de l'instrument utilisé :Trappe ionique ou Q-TOF. Ceci est d'autant plus intéressant que dans le contexte de l'approche MS/MS, Q-TOF et Trappe ionique apparaissent comme complémentaires et permettent en combinant les résultats d'améliorer les pourcentages de recouvrement de séquence.

De plus, si les fragments majeurs (y et b) sont observés de la même façon pour les trois instruments, il est également intéressant d'observer les fragments haute énergie (w et d) par MALDI-TOF/TOF. Cette approche permet de différencier les acides aminés isobares et donc d'améliorer les recouvrements de séquence. Ainsi, en fonction des échantillons et des questions posées (caractérisation fine des protéines ou identification globale de protéines), il pourra être intéressant de choisir l'une ou l'autre technique. En effet, l'approche MALDI-MS/MS, n'étant pas encore automatisée et nécessitant une plus grande quantité de matériel, sera préférée pour des études ponctuelles de caractérisation fine des protéines (distinction d'acides aminés isobares). Au contraire, l'approche nanoLC-MS/MS, automatisable, sera préférée pour les études par homologie de séquences et les études de séquençage globale (identification de protéines peu ou pas référencées dans les banques de données).

L'optimisation de ces différentes étapes, ainsi que le choix adapté des approches protéomiques ont permis d'obtenir des éléments de réponse à différents problèmes biologiques.

Approches protéomiques MALDI-MS

L'approche MALDI-MS a d'une part, permis de cartographier 87 protéines de noyaux des cellules Caco2/TC7 ainsi que 68 protéines exprimées de façon différentielle en fonction des états de différenciation des cellules. Ce protéome pourra servir de référence pour les nombreuses études in vitro réalisées sur ces cellules. Cependant , cette étude n'a pas permis d'identifier les protéines minoritaires (facteurs de transcriptions) détectées par d'autres tests biologiques (Western Blot) Afin

d'avoir accès à ces protéines, il s'agira de fractionner l'échantillon ou d'ajouter des étapes de séparations supplémentaires (nanoLC-MS/MS) pour diminuer la complexité du mélange.

D'autre part, il a également été possible, grâce à cette approche MALDI-MS, d'identifier 16 protéines sur ou sous exprimées dans les mitochondries mutées responsables de la maladie MERRF. L'identification de ces protéines a permis de mieux comprendre les incidences de la mutation du tRNA Lysine mitochondrial au niveau cellulaire. Par ailleurs, les études protéomiques différentielles pourront être complétées par des études par marquage ICNAT, donnant accès aux protéines de pI extrêmes et hydrophobes. De plus, il serait intéressant de caractériser les modifications post-traductionnelles (phosphorylations) de certaines isoformes protéiques décrites comme ayant un rôle biologique important dans l'activité mitochondriale (pyruvate deshydrogénase E1 α)

Approches protéomiques par analyse nanoLC-MS/MS

Grâce à l'approche nanoLC-MS/MS et aux recherches dans les banques de données par homologie de séquence, il a été possible d'identifier 72 protéines exposées à la surface de *Streptococcus bovis* et 84 protéines excrétées par la bactérie. Grâce aux fractionnement des échantillons, il a été possible de réduire le nombre des protéines candidates responsables de l'activité anti-inflammatoire. A présent, il s'agira par des études biologiques de déterminer la ou les protéines responsables de l'inflammation du côlon pour mieux comprendre leurs rôles dans la cancérogénèse colique.

Approches protéomiques par séquençage De Novo

Dans le but de mieux comprendre le fonctionnement global de Plasmodium

falciparum (parasite responsable de la Malaria), l'approche nanoLC-MS/MS suivie d'un séquençage De Novo et de recherches par alignement de séquences a permis d'identifier :

- 3 protéines avec une activité enzymatique de Sérine hydrolase, qui n'étaient pas encore décrites.

- 4 protéines exprimées intensément dans les globules rouges parasités par rapport aux globules rouges sains.

- 64 protéines parasitaires présentes à la surface ou enchassées dans la membrane du globule rouge.

Même si le génome n'était pas entièrement disponible lors de cette étude, le fractionnement des échantillons et l'approche par séquençage De Novo a permis de caractériser plusieurs protéines dont les fonctions biologiques n'ont pas encore été décrites avec certitude. Par ailleurs, il serait intéressant pour identifier un plus grand nombre de protéines ainsi que des protéines minoritaires, dans le cadre de l'analyse protéomique différentielle, de fractionner un peu plus l'échantillon, afin d'éliminer les protéines chaperonnes.

Ainsi, grâce à l'identification de ces protéines, il sera possible de mieux comprendre le fonctionnement global du parasite pour à terme, élaborer de nouvelles drogues spécifiques.

Pour conclure, notre travail a permis d'optimiser les différentes étapes de l'analyse protéomique, des gels 2D aux analyses par spectrométrie de masse. Ces différentes optimisations ainsi que le choix adapté des approches protéomiques nous ont ensuite permis de répondre à divers problèmes biologiques.

Cependant, afin d'améliorer ces différentes approches, il reste à développer ou à optimiser des logiciels de traitement informatiques, particulièrement dans les études par séquençage *De Novo*. Nous avons également montré l'intérêt de bien adapter les techniques d'analyse en fonction des problèmes à résoudre. Pour ce faire, une connaissance approfondie des différents protocoles de préparation des échantillons et des paramètres instrumentaux (chromatographie liquide, spectrométrie de masse) est requise. Cette connaissance des techniques garantit la qualité des informations sans laquelle les résultats obtenus ne pourraient aboutir à une bonne identification des protéines voire à leur caractérisation.

Enfin, ces études ont montré que la préparation de l'échantillon était primordiale pour avoir une bonne caractérisation des protéines. En effet, lors de l'analyse du protéome nucléaire de cellules Caco2, il n'a pas été possible d'identifier des protéines minoritaires. Au contraire, l'analyse des fractions bactériennes de *Streptococcus bovis* a permis d'identifier différentes protéines candidates pour l'inflammation du colon. Par conséquent, l'accès aux protéines minoritaires sera plus facile après fractionnement des échantillons, d'où la grande importance de la préparation des échantillons.
Partie expérimentale :

1 Les spectromètres de masse

1.1 Le MALDI-TOF

Toutes les mesures de masses par MALDI-TOF ont été réalisées sur des instruments de la gamme Bruker : le Biflex, le Biflex III et l'Ultraflex (Bruker-Franzen Analytic GmbH, Bremen, Germany). L'Ultraflex, plus récent, a permis d'effectuer des analyses avec un seuil de sensibilité plus bas et une meilleure résolution grâce à un tube de vol plus long (1,5 m pour l'Ultraflex). De plus, cet instrument nous a permis de réaliser des analyses MS/MS sur les ions métastables. Le logiciel d'acquisition des spectres est Flexcontrol 1.2. Le logiciel d'exploitation des données est XMASS, version 5.0 (Bruker-Franzen Analytic GmbH, Bremen, Germany).

Les trois instruments sont équipés 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.

1.1.1 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 pour l'Ultraflex.

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 mihauteur.

1.1.2 Paramètres importants de l'analyse par MALDI-TOF MS/MS

L'analyse MS/MS par la technologie LIFT correspond à l'analyse des ions métastables. Après formation des ions par tir laser avec une différence de potentiel de 8kV, l'ion parent et ses ions fils (ions métastables) sont réaccélérés avec une tension de 19kV dans la cellule LIFT. L'ion parent est ensuite éliminé pour favoriser la détection des ions métastables. Les ions fragments sont enfin analysés dans le TOF en mode réflecteur. L'ajout de gaz dans la cellule de collision (Argon à une pression de 6 10⁻⁶ Torr) est possible pour accentuer la présence des fragments haute énergie. L'interprétation des données MS/MS est réalisée à l'aide du logiciel Biotools 2.2.

1.1.3 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 isus 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).

Deux types d'étalonnages sont possibles :

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.

<u>L'étalonnage externe</u> est beaucoup moins précis. Le dépôt de l'étalon se fait en un endroit distinct de celui de l'échantillon sur la cible. La première étape consiste en l'acquisition d'un spectre de masse du mélange standard et de l'étalonnage proprement dit. Dans un deuxième temps, les ions analytes sont analysés dans les mêmes conditions que les étalons. Cette méthode présente l'avantage de pouvoir travailler sur de très faibles quantités d'échantillons sans risque de suppression du signal. Mais l'étalonnage externe n'est pas assez précis pour les analyses protéomiques.

Un compromis entre ces deux modes d'étalonnage peut être réalisé par un dépôt particulier, schématisé en figure 1.

Figure 1 : Dépôt d'échantillon et de l'étalon pour l'analyse de masses précises.



Ce dépôt a permis de contourner le problème de suppression lié à la co-cristallisation de l'analyte avec le mélange étalon, en préservant une zone sans mélange étalon. Ainsi en se déplaçant de quelques mm sur la cible, il est possible d'obtenir des signaux dans la zone mixte, puis de passer dans la zone ne contenant que l'analyte pour accumuler suffisament de signal.

1.1.4 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 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.

1.2 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.

1.2.1 Le spectromètre de masse ES-TOF.

Le dosage de molécules pharmacologiques issues de fluides biologiques a été effectué sur le ES-TOF (LCT, Micromass Ldt, Altricham, UK). L'acquisition des spectres de masse ainsi que le traitement des données est réalisé par le logiciel MassLynx, version 3.5 (Micromass Ltd., Altricham, UK). 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.

1.2.2 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.



Tableau 1 : Tableau des débits et des tensions utilisés.

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

L'acquisition des spectres de masse ainsi que le traitement des données sont réalisés par le logiciel MassLynx, version 3.5 (Micromass Ltd., Altricham, UK). Le Q-TOF II est un instrument hybride et, comme nous l'avons vu dans les chapitres précédents, il allie les avantages des analyseurs quadripôlaires (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)

a 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 (1-5 μ l) (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 μ l grâce au nanodébit.

b 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.

c La préparation de l'échantillon en nanospray

Le dessalage des échantillons est une étape incontournable de l'analyse par nano-ESI-MS. Nous avons mis au point un protocole de dessalage sur ZipTip $_{C18}$ (Millipore, USA) compatible avec nos analyses :

- Le ZipTip est un cône de capacité 10-20μl contenant 0,6 μl de phase C18 type silice sphérique (15μm, 200 Å). La quantité de matériel retenue sur le ZipTip est de l'ordre de 2 μg. Le ZipTip peut être utilisé dans une gamme de pH de 2 à 13.
- L'étape clé est l'activation de la colonne : l'activation se fait dans une solution ACN/H2O (1/1) sans acide. Les ZipTips sont rincés avec cette solution. Il s'agit de chasser l'air contenu dans la phase (il faut veiller ensuite à ne plus introduire d'air).
- Le conditionnement du ZipTip : l'acétonitrile est chassé par des rinçages successifs avec une solution d'eau acidifiée, H₂O/HCOOH 0,1%, pour permettre la fixation des peptides.
- La fixation des peptides se fait par des mouvements de «va et vient» par pipetage de la solution contenant les peptides.
- Le lavage ou dessalage : Une solution H₂O/HCOOH 0,1% est utilisée pour laver les peptides retenus.
- L'élution : Les peptides sont décrochés de la phase par une solution d'ACN dont la concentration varie selon l'hydrophobicité des peptides (ACN/H2O (1/1) acidifié avec 0,1% HCOOH).

1.2.3 La trappe ionique

Les spectres de masse ES-Trappe Ionique sont obtenus sur un spectromètre de masse ESQUIRE 3000+ (Bruker-Franzen Analytic GmbH, Bremen, Germany), équipé soit d'une source électrospray orthogonale Hewlett Packard soit d'une source nanoélectrospray linéaire (Bruker-Franzen Analytic GmbH) soit d'une nanoélectrospray on line (Bruker-Franzen Analytic GmbH).

L'acquisition des spectres de masse est réalisée par le logiciel Bruker ESQUIRE control (Version NT 3.0, Bruker–Franzen Analytic GmbH). Le traitement des données s'effectue à l'aide du logiciel Bruker DataAnalysis (Version 3.2, Bruker–Franzen Analytic GmbH). Les spectres sont enregistrés en additionnant et en moyennant de 10 à 100 balayages de la gamme m/z. Ce nombre de balayages augmente significativement lorsque le degré de la spectrométrie de masse multiple croit.

1.2.4 L'étalonnage des spectromètres de masse électrospray.

Plusieurs étalons sont utilisés selon la nature des peptides ou molécules à analyser et selon l'instrument utilisé :

<u>LC-TOF et Q-TOF II</u> : Pour les protéines, l'étalonnage est réalisé sur une gamme de masse de 200 à 2000 m/z, par l'infusion dans le spectromètre de masse d'une solution à 2picomoles/µl de myoglobine de cheval (Sigma, M-1882). L'étalonnage se fait à partir des ions multi-chargés.

<u>Q-TOF II</u> : pour l'analyse de peptides par nanospray ou LC-nanospray, l'étalonnage est réalisé par l'infusion dans le spectromètre d'une solution micromolaire de poly-alanine (Sigma P-9003) dans ACN/H2O 50/50 acidifiée avec 0,1% HCOOH. Dernièrement, une nouvelle méthode de calibration a été mise au point en utilisant le Glu-Fibrinopeptide (Sigma, F3261). Cette calibration se fait sur les fragments générés par la fragmentation de l'ion parent dans la cellule de collision. L'énergie de collision optimale est de 19eV. L'avantage de cette calibration est l'absence d'adsorption sur les capillaires en silice contrairement à la poly-alanine. Cette adsorption est parfois responsable d'une pollution pour les analyses suivant la calibration.

<u>Trappe ionique</u> : pour l'analyse de peptides par nanospray ou LC-nanospray, l'étalonnage est réalisé par l'infusion dans le spectromètre d'une solution d'un mélange peptidique : 712,38 m/z (leuenképhaline) ; 1046,54 m/z (angiotensine) ; 1348,72 m/z (substance P) ; 1620,81 m/z (bombésine) ; 2004,95 m/z (Druuna). La calibration de cet appareil étant très stable, il n'est pas nécessaire de la réaliser avant chaque analyse.

2 Les systèmes HPLC

Lors de cette thèse, nous avons utilisées quatre pompes HPLC différentes : HP 1100 (Hewlett Packard), Micromass CapLC (Micromass), HPLC Waters Alliance et HPLC Waters 625 (Waters).

2.1 HPLC Waters Alliance

Les molécules à analyser sont chromatographiées sur un système HPLC comportant un programmeur de gradient piloté par le logiciel Millennium (Waters) couplé à une pompe Waters 2690 de la série Alliance et à détecteur à barrette de diode Waters 996.

2.2 Micromass CapLC et nanoLC d'Agilent

Le système HPLC utilisé est une CapLC (Micromass Ltd., Manchester, UK) et une nanoLC (Agilent..) qui permet l'injection automatique des échantillons et leur concentration sur une précolonne. L'échantillon injecté est concentré et dessalé sur une pré-colonne de 1mm de long et 300 μ m de diamètre interne remplie d'une phase stationnaire de type PepMap C18 de granulométrie 5 μ m et de 100 Å de porosité (LC-Pakings) à un débit de 30 μ l/min durant 3 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

La pré-colonne est ensuite mise automatiquement à l'aide du "Stream Select" (Micromass Ltd., Manchester, UK) 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 4).

La phase mobile A est une solution d'eau acidifiée par 0,1 % d'acide formique, la phase mobile B est une solution d'acétonitrile acidifiée par 0,1 % d'acide formique.

L'élution des peptides est réalisée à un débit de 200 nl/min par un gradient de 5 à 45 % de solvant B durant 35 minutes, puis 95 % de solvant B durant 5 minutes , puis la colonne est équilibrée 20 minutes par une solution d'eau acidifiée par 0,1 % d'acide formique.

Figure 4 : Schéma représentant les deux étapes de la LC-MS-MS. 1 : chargement de l'échantillon sur la précolonne et lavage, 2 :élution des peptides retenus sur la pré-colonne et séparation de ces peptides sur la colonne chromatographique.



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 μ L 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 μ L d'ACN pendant 5 min.

La réduction et l'alkylation

La solution réductrice est une solution de dithiothreitol 10 mM dans NH_4HCO_3 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 NH_4HCO_3 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/ μ L dans NH₄HCO₃ 25mM. Le volume ajouté est fonction de la taille des spots (5-10 μ L). La réaction de digestion *in gel* se fait à température ambiante et *overnight*.

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. La qualité de l'extraction dépend principalement de deux paramètres : le temps d'extraction et le solvant d'extraction. Lors des études protéomiques présentées deux protocoles d'extraction sont utilisés selon le type de spectromètre de masse utilisé pour l'analyse :

- MALDI-MS : Dans chaque puits, 5 µl de 60/35/5 (Acétonitrile/eau/acide formique) sont ajoutés. Le temps d'extraction est de 60 minutes.
- LC-MS-MS : Ajout de 30 µl de 60/35/5 (Acétonitrile/eau/acide formique) dans chaque puits. Ensuite la plaque est placée 60 minutes dans un sonicateur, avant de récupérer le solvant d'extraction (environ 20 à 25 µl). Ce volume est réduit à 8 µl par évaporation de l'acétonitrile à température ambiante.

3.3 La nanoLC-MS-MS

Les analyses nanoLC-MS-MS ont été effectuées sur le couplage CapLC-Q-TOFII. Le gradient et les débits des pompes A, B et C sont détaillés ci-dessous. La colonne utilisée est une colonne de 75µm de diamètre interne (LC Packings).

Temps (min)	%A	%B
0.01	95	5
3	95	5
35	55	45
36	5	95
48	5	95
48.1	95	5
70	95	5

Avec \mathbf{A} : H₂O/ 0.1% HCOOH

et

B: Acétonitrile/ 0.1% HCOOH

Débit constant avant split 5.5 µl/min

<u>Débit</u> pompe auxiliaire

Gradient CapLC

(pompe A et B)

	Débit
Temps (min)	Pompe C
0.01	30 μl/min
3	30 μl/min
3.1	2 μl/min

Avec C: H2O/ 0.1% HCOOH