

# THÈSE

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

par  
Guillaume BÉCHADE

## AVANCÉES EN ANALYSE PROTÉOMIQUE PAR SPECTROMÉTRIE DE MASSE : RECHERCHE DE BIOMARQUEURS ET CARACTÉRISATION DE COMPLEXES COVALENTS

soutenue le 21 décembre 2009 devant la commission d'examen :

Dr Alain VAN DORSSELAER

Pr Jérôme LEMOINE

Dr Thierry RABILLOUD

Dr Philippe BULET

Pr Jacques HAIECH

Dr Sarah SANGLIER CIANFERANI

Directeur de thèse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

*À ceux que j'aime,*

*« Le bonheur est une recherche. Il faut y employer l'expérience et son imagination. »*  
Jean Giono, extrait du Voyage en Italie

# REMERCIEMENTS

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

J'adresse ma reconnaissance et mon profond respect à **Alain Van Dorsselaer** qui m'a permis d'évoluer dans l'environnement qu'il construit, exceptionnel sur les plans humain, scientifique et matériel.

Je remercie tout particulièrement **Sarah Sanglier-Cianferani** pour son encadrement, son regard attentif et sa confiance.

Je prie monsieur **Jacques Haiech**, président du jury, messieurs **Jérôme Lemoine** et **Thierry Rabilloud**, rapporteurs, et monsieur **Philippe Bulet**, examinateur, d'agréer mes respectueux remerciements pour avoir accepté d'évaluer ce travail de thèse et pour le temps qu'ils y ont consacré.

Je remercie également l'ensemble de mes collaborateurs pour leur accompagnement dans des thématiques aux défis aussi variés que stimulants. Mes remerciements vont notamment à **Laurent Mauvieux** et **Laurent Miguet**, **Régine Losson** et **Khalid Ouararhni**, **Guy Branlant**, **Sophie Rahuel-Clermont** et **Xavier Roussel**, **Sandrine Boschi-Muller**, **Adeline Gand**, **Laure Selme-Roussel** et **Marourane Libiad**.

Je suis profondément reconnaissant au **Réseau Européen de Formation des Doctorants GRK532** d'avoir soutenu financièrement mon travail. A l'occasion des échanges organisés, j'ai pu rencontrer des personnes de qualité, issues de divers horizons scientifiques et culturels. Je tiens à les saluer.

Merci finalement – chaleureusement – à tous les membres du LSMBO passés, présents et en devenir :

Spécialement à **Christine Carapito** pour son soutien complice et quotidien à tous niveaux ;

A **Agnès Hovasse**, **Cédric Atmanène**, **Nicolas Barthélemy**, **Sébastien Gallien** et **Véronique Delval** pour les relations scientifiques et amicales que nous avons pu nouer ;

A **Alexandre Burel**, **Amandine Boeuf**, **Antoine Zieger**, **Audrey Bednarczyk**, **Bertrand Chaumande**, **Chrystel Husser**, **Cyril Colas**, **Daniel Ayoub**, **Dimitri Heintz**, **Flavie Robert**, **Haiko Herschbach**, **Jean-Michel Saliou**, **Julien Roeser**, **Laëtitia Fouillen**, **Marie-Laure Lortz**, **Thierry Wasselin** et aux valeurs montantes **Tchilabalo (Eli) Alayi** et **Stéphanie Petiot**.

Merci aux piliers du laboratoire, **Christine Schaeffer**, **Danièle Thierse**, **Fabrice Bertile**, **Fabrice Varier**, **François Delalande**, **Hélène Diemer**, **Jean-Marc Strub**, **Jennifer Jund-Riegert**, **Laurence Sabatier**, **Véronique Trimbour** et **René Winkler** pour leur soutien inestimable.

Je n'oublie pas **les Snockeles** qui se cachent toujours ici ou là pour un tour de malice... Je les ai vus changer facétieusement des échantillons en tablettes de chocolat, ou souffler à certains des envies de barbecue ou de poulet rôti à partager ... Ils resteront pour moi comme de petits esprits qui ont contribué et contribuent à lier chacun de nous au quotidien.

# PLAN GÉNÉRAL

<b>Introduction générale</b> .....	1
------------------------------------	---

---

## INTRODUCTION BIBLIOGRAPHIQUE

<b>Introduction</b> .....	7
<b>Définitions</b> .....	8
<b>Chapitre I</b>	
Instruments et technologies au service de l'analyse protéomique .....	9
<b>Chapitre II</b>	
Les stratégies d'identification en analyse protéomique par spectrométrie de masse .....	33
<b>Chapitre III</b>	
Les données générées en analyse protéomique : gestion, validation et partage de l'information .....	49

---

## RESULTATS – 1<sup>RE</sup> PARTIE

### DÉVELOPPEMENTS MÉTHODOLOGIQUES POUR L'ANALYSE PROTÉOMIQUE CLASSIQUE

<b>Chapitre I</b>	
De nouveaux outils et des méthodes optimisées pour l'analyse protéomique par spectrométrie de masse.....	69
<b>Chapitre II</b>	
Recherche de nouveaux biomarqueurs caractéristiques d'hémopathies lymphoïdes chroniques .....	113
<b>Chapitre III</b>	
Étude des partenaires d'interaction du facteur de transcription suppresseur de tumeur Trim24 .....	133

---

## RESULTATS – 2<sup>E</sup> PARTIE

### CARACTÉRISATION FINE DE COMPLEXES PROTÉIQUES COVALENTS À LIAISON CYSTÉINE–CYSTÉINE

<b>Chapitre I</b>	
Techniques et stratégies pour la caractérisation de complexes covalents à liaison cystéine – cytséine par spectrométrie de masse.....	161
<b>Chapitre II</b>	
Mise en évidence de la formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroydée par la sulfirédoxine.....	183
<b>Chapitre III</b>	
Étude de mécanismes de régénération de Methionine Sulfoxyde Réductases .....	201
<b>Conclusion générale</b> .....	223
<b>Bibliographie</b> .....	229
<b>Annexes</b> .....	245

# PLAN DÉTAILLÉ

Introduction générale .....	1
-----------------------------	---

## INTRODUCTION BIBLIOGRAPHIQUE

---

Introduction .....	7
--------------------	---

Définitions .....	8
-------------------	---

### Chapitre I

Instruments et technologies au service de l'analyse protéomique .....	9
-----------------------------------------------------------------------	---

<b>1. La spectrométrie de masse pour la caractérisation des protéines</b> .....	10
---------------------------------------------------------------------------------	----

1.1. Les sources d'ionisation .....	10
-------------------------------------	----

1.1.1. La source Electrospray (ESI) et Nanoelectrospray (nanoESI) .....	10
-------------------------------------------------------------------------	----

1.1.2. La source MALDI .....	12
------------------------------	----

1.2. Les analyseurs .....	13
---------------------------	----

1.2.1. Les analyseurs simples pour la spectrométrie de masse MS .....	13
-----------------------------------------------------------------------	----

1.2.2. Les analyseurs pour la spectrométrie de masse MS/MS .....	17
------------------------------------------------------------------	----

1.2.3. Performances caractéristiques des analyseurs .....	19
-----------------------------------------------------------	----

1.3. La fragmentation peptidique .....	22
----------------------------------------	----

1.3.1. La fragmentation induite par collision (CID) .....	22
-----------------------------------------------------------	----

1.3.2. La dissociation par transfert d'électron (ETD) .....	23
-------------------------------------------------------------	----

<b>2. Les techniques de séparation de protéines et peptides</b> .....	25
-----------------------------------------------------------------------	----

2.1. La séparation de protéines : l'électrophorèse sur gel et ses alternatives .....	25
--------------------------------------------------------------------------------------	----

2.1.1. L'électrophorèse sur gel mono- et bidimensionnelle .....	25
-----------------------------------------------------------------	----

2.1.2. L'électrophorèse Off-Gel .....	27
---------------------------------------	----

2.1.3. Séparation de protéines par chromatographie liquide .....	28
------------------------------------------------------------------	----

2.2. La séparation de peptides par chromatographie liquide .....	28
------------------------------------------------------------------	----

2.2.1. Les systèmes nanoHPLC et nanoUPLC .....	28
------------------------------------------------	----

2.2.2. Les techniques de chromatographie liquide multidimensionnelle .....	29
----------------------------------------------------------------------------	----

2.3. Conclusion .....	30
-----------------------	----

### Chapitre II

Les stratégies d'identification en analyse protéomique par spectrométrie de masse .....	33
--------------------------------------------------------------------------------------------	----

<b>1. Les stratégies d'identification de protéines</b> .....	34
--------------------------------------------------------------	----

1.1. L'identification par empreinte peptidique massique (PMF) .....	34
---------------------------------------------------------------------	----

1.1.1. Principe de l'identification par PMF .....	34
---------------------------------------------------	----

1.1.2. Les limites du PMF .....	35
---------------------------------	----

1.2. Les stratégies nanoLC-MS/MS .....	36
----------------------------------------	----

1.2.1. Identification en mode d'acquisition données-dépendant (DDA).....	36
1.2.2. De nouveaux modes d'acquisition originaux (MS <sup>F</sup> ).....	38
<b>2. Exemples de défis stratégiques en analyse protéomique</b> .....	<b>40</b>
2.1. Les stratégies de quantification.....	40
2.1.1. Le marquage de protéines ou de peptides.....	40
2.1.2. Les approches sans marquage.....	41
2.1.3. Des modes d'acquisition particulièrement adaptés à la quantification.....	41
2.2. L'analyse de partenaires d'interaction protéiques.....	42
2.2.1. La co-immunopurification et la copurification par affinité (TAP-Tag).....	42
2.2.2. Les limites des techniques de copurification par affinité.....	43
2.2.3. La séparation du complexe purifié : SDS-PAGE et BN-PAGE.....	44
2.3. L'analyse de protéines membranaires.....	44
2.3.1. Les difficultés de l'analyse de protéines membranaires par spectrométrie de masse.....	45
2.3.2. Les microparticules : un outil de choix pour l'isolation de protéines membranaires plasmiques.....	45
2.4. Conclusion.....	47

## Chapitre III

Les données générées en analyse protéomique : gestion, validation et partage de l'information.....	49
<b>1. Les banques de données de séquence</b> .....	<b>50</b>
1.1. Les banques de séquences nucléotidiques.....	50
1.2. Les banques de séquences protéiques.....	50
1.2.1. Les banques de séquences « corrigées ».....	50
1.2.2. Les banques généralistes de dépôt (« <i>sequence repositories</i> »).....	51
1.2.3. La croissance exponentielle des banques.....	52
1.3. Les annotations « <i>Gene Ontology</i> ».....	53
<b>2. La validation des résultats d'identification</b> .....	<b>55</b>
2.1. Les différents moteurs de recherche et leurs spécificités.....	55
2.1.1. Les algorithmes de recherche heuristiques.....	55
2.1.2. Les algorithmes de recherche probabilistes.....	55
2.1.3. Les limites des moteurs de recherche et le besoin de valider les résultats.....	56
2.2. La stratégie de vérification manuelle des identifications et ses limites.....	57
2.3. Les stratégies d'évaluation de taux de faux positifs.....	59
2.3.1. Signification des taux de faux positifs.....	59
2.3.2. Les stratégies « <i>Target-Decoy</i> ».....	59
2.3.3. Les approches empiriques de Bayes (PeptideProphet).....	60
2.3.4. Combiner les informations pour améliorer l'évaluation des résultats.....	60
<b>3. Les difficultés du partage de l'information protéomique</b> .....	<b>62</b>
3.1. Le minimum requis pour la publication de données protéomiques (MIAPE).....	62
3.1.1. Le besoin de directives pour la publication de données protéomiques.....	62
3.1.2. Vers l'unification et la standardisation des directives.....	62
3.2. Les projets de partage des données.....	63
3.2.1. L'objet des banques de dépôt protéomiques.....	63
3.2.2. Les principales banques de dépôt protéomiques.....	64
<b>4. Conclusion</b> .....	<b>65</b>

DÉVELOPPEMENTS MÉTHODOLOGIQUES POUR L'ANALYSE PROTÉOMIQUE CLASSIQUE

**Chapitre I**

De nouveaux outils et des méthodes optimisées

pour l'analyse protéomique par spectrométrie de masse.....69

**1. Etude de la répétabilité des analyses nanoLC-MS/MS**

**d'échantillons biologiques complexes**.....70

1.1. Stochastique des analyses nanoLC-MS/MS.....71

1.1.1. Origine des phénomènes de non-répétabilité.....71

1.1.2. Quelques options permettent de limiter la perte d'information.....72

1.1.3. Intérêt de réaliser plusieurs analyses nanoLC-MS/MS d'un même échantillon.....72

1.2. Evaluation de la répétabilité des analyses nanoLC-MS/MS.....73

1.2.1. Objectif de l'étude.....73

1.2.2. Mise en oeuvre expérimentale.....73

1.3. Résultats et discussion.....74

1.3.1. Observations préliminaires.....74

1.3.2. Détermination du nombre d'analyse à répéter.....76

1.4. Proposition et évaluation d'une stratégie visant à améliorer la répétabilité des analyses nanoLC-MS/MS.....80

1.4.1. La concaténation des listes de masses avant la recherche par les moteurs.....80

1.4.2. Evaluation de la stratégie sur un échantillon total.....81

1.5. Valorisation de ces travaux.....83

1.5.1. Communications orales.....83

1.5.2. Communication par affiche.....84

1.6. Conclusions.....84

**2. Des optimisations instrumentales permettent une meilleure prise des données**

**protéomiques**.....85

2.1. La résolution chromatographique : nanoHPLC-Chip et nanoUHPLC.....85

2.1.1. Le système nanoHPLC-Chip Cube (Agilent Technologies).....85

2.1.2. Le système nanoAcquity UPLC (Waters Corporation).....90

2.1.3. Conclusion.....93

2.2. Une exactitude accrue de la mesure de masse grâce au système « *lock-mass* ».....93

2.2.1. Présentation et fonctionnement d'un système lock-mass : le LockSpray™ Waters Corporation.....94

2.2.2. Mise en place et évaluation du bénéfice apporté par le système LockSpray.....94

2.3. L'optimisation de l'acquisition des données en spectrométrie de masse.....97

2.3.1. Les paramètres physiques de détection et sélection des ions parents.....97

2.3.2. Les paramètres temporels des acquisitions MS et MS/MS.....97

2.3.3. Les paramètres de fragmentation des ions parents.....98

2.3.4. Proposition d'une feuille de paramètres optimisés.....98

2.4. Conclusion.....99

**3. De nouveaux outils bio-informatiques améliorent le nombre et la crédibilité des**

**identifications de protéines**.....100

3.1. Les paramètres de traitement des données brutes.....100

3.1.1. La génération des fichiers de « listes de masses » ou « *peak list* ».....100

3.1.2. Optimisation des paramètres d'extraction des données brutes avec ProteinLynxGlobalServer (Waters Corp.).....101

3.1.3. Optimisation des paramètres d'extraction des données brutes avec DataAnalysis (Bruker Dal.).....103

3.1.4. Vers un format libre de fichiers « <i>peak list</i> »	103
3.2. Mise au point d'une stratégie de détermination d'un taux de faux positifs par l'approche Target-Decoy	103
3.2.1. Les limites des stratégies d'identification manuelle	104
3.2.2. La préparation de banques « <i>Target-Decoy</i> »	105
3.2.3. L'estimation et l'ajustement du taux de faux positifs (FDR)	106
3.2.4. Les bienfaits du logiciel Scaffold pour la mise en oeuvre de notre stratégie	108
3.3. L'émergence de stratégies complémentaires	109
3.3.1. L'utilisation combinée de plusieurs moteurs de recherche	109
3.3.2. Les stratégies de recherche « orthogonale » avec Peaks Studio	110
3.4. Conclusion	111
4. Conclusion	112

## Chapitre II

### Recherche de nouveaux biomarqueurs caractéristiques

d'hémopathies lymphoïdes chroniques	113
-------------------------------------	-----

#### 1. La nécessité de mettre en évidence de nouveaux biomarqueurs membranaires spécifiques de leucémies

1.1. La classification des leucémies	114
1.2. Vers un meilleur diagnostic des cas ambigus de leucémies	115
1.2.1. Les techniques de cytométrie en flux	115
1.2.2. La découverte de biomarqueurs comme moteur du progrès en matière de diagnostic	115
1.3. Rechercher de nouveaux marqueurs membranaires spécifiques	116
1.3.1. Les méthodes classiques	116
1.3.2. Approche protéomique par l'analyse des protéines membranaires par spectrométrie de masse	117

#### 2. Analyse comparative du protéome membranaire de plusieurs leucémies

2.1. Choix expérimentaux permettant la recherche avancée de marqueurs membranaires potentiels	119
2.1.1. Préparation des échantillons et enrichissement en protéines membranaires	119
2.1.2. Optimisation des analyses nanoLC-MS/MS	121
2.2. Stratégie visant à mettre en évidence des marqueurs membranaires potentiels	123
2.2.1. Définition de filtres de sélection de candidats biomarqueurs pertinents	123
2.2.2. Application au cas du lymphome de la zone du manteau	123
2.3. La validation des cibles proposées	126
2.3.1. Le « <i>pipeline</i> » de validation de biomarqueurs	126
2.3.2. Validation du biomarqueur candidat CD148 pour le lymphome de la zone du manteau	127
2.4. Publication de ces travaux	128

#### 3. Perspectives : Recherche de biomarqueurs potentiels dans le cas de la leucémie de Waldenström

3.1. La leucémie de Waldenström	129
3.2. Application de la stratégie analytique	129
3.3. Résultats préliminaires	130

#### 4. Conclusion

## Chapitre III

### Etude des partenaires d'interaction du facteur de transcription

suppresseur de tumeur Trim24	133
------------------------------	-----

<b>1. Les mécanismes de régulation de la transcription</b> .....	134
1.1. La transcription de l'ADN .....	134
1.2. Les récepteurs nucléaires et le mode d'action des cofacteurs .....	135
1.2.1. Les récepteurs nucléaires .....	135
1.2.2. Les cofacteurs des récepteurs nucléaires .....	136
1.3. Dysfonctionnements dans les processus de régulation de la transcription .....	136
<b>2. Le rôle du facteur de transcription Trim24 dans l'hépatocarcinogénèse</b> .....	137
2.1. Hépatocarcinogénèse et acide rétinoïque .....	137
2.1.1. Le carcinome hépatocellulaire .....	137
2.1.2. L'acide rétinoïque et l'hépatocarcinogénèse .....	137
2.2. Mise en évidence de Trim24 comme répresseur de tumeur dans le foie .....	138
2.2.1. Le facteur intermédiaire de transcription Trim24 .....	138
2.2.2. Mise en évidence de Trim24 comme suppresseur de tumeur dans le foie .....	138
2.2.3. Le phénotype tumoral des hépatocytes Trim24 <sup>-/-</sup> est sous le contrôle de RAR $\alpha$ .....	138
2.2.4. Conclusions .....	139
<b>3. Identification des mécanismes moléculaires par lesquels Trim24 régule la transcription des gènes oncogéniques dans le foie</b> .....	140
3.1. Mise au point d'une stratégie de purification et d'identification des partenaires spécifiques de Trim24 .....	140
3.1.1. Stratégie générale mise en oeuvre .....	140
3.1.2. Purification des complexes multiprotéiques associés à Trim24 .....	141
3.1.3. Séparation des protéines sur gel 1-D .....	143
3.1.4. Analyse par nanoLC-MS/MS .....	144
3.1.5. Identification des partenaires d'interaction de Trim24 .....	144
3.2. Résultats .....	145
3.2.1. Complémentarité des résultats issus des échantillons préparés à partir de Trim24 FLAG-HA C-ter ou N-ter .....	145
3.2.2. Identification de sous-complexes de Trim24 : Trim 33 un partenaire majeur .....	145
3.2.3. Mise en évidence de l'interaction Trim24 - Trim33 .....	148
3.2.4. Comparaison des partenaires de Trim24 purifiés à partir d'une lignée cellulaire traitée ou non traitée à l'acide rétinoïque .....	150
3.2.5. Perspectives .....	154
3.3. Valorisation de ce travail : communication par affiche .....	156
<b>4. Conclusion</b> .....	157

## RESULTATS – 2<sup>E</sup> PARTIE

---

### CARACTÉRISATION FINE DE COMPLEXES PROTÉIQUES COVALENTS À LIAISON CYSTÉINE-CYSTÉINE

#### Chapitre I

#### Techniques et stratégies pour la caractérisation de complexes covalents à liaison cystéine – cystéine par spectrométrie de masse .....

161

##### 1. La chimie des résidus cystéiques : oxydation, réduction, formation de complexes covalents à liaison cystéine-cystéine .....

162

###### 1.1. Les oxydations impliquant des cystéines .....

162

###### 1.1.1. Cas d'un résidu cystéique isolé .....

162

###### 1.1.2. Oxydation impliquant deux cystéines : les liaisons disulfure, thiosulfinate et thiosulfonate .....

162

###### 1.2. Les fonctions biologiques de l'oxydoréduction des cystéines .....

163

###### 1.2.1. Structuration secondaire des protéines .....

163

1.2.2. Les cystéines oxydées comme intermédiaires catalytiques.....	164
1.3. Les difficultés liées à l'analyse de protéines à liaisons cystéine-cystéine.....	164
<b>2. Analyse de complexes covalents à liaison cystéine-cystéine entiers produits lors de réactions enzymatiques d'oxydoréduction.....</b>	<b>166</b>
2.1. Préparation des échantillons pour l'analyse de complexes covalents à liaison cystéine-cystéine.....	166
2.1.1. Optimisation de la formation des complexes à analyser.....	166
2.1.2. Séparation de complexes covalents à liaison cystéine-cystéine entiers.....	167
2.2. Spectrométrie de masse de complexes covalents à liaison cystéine-cystéine entiers.....	170
2.3. Conclusion.....	171
<b>3. Une combinaison de stratégies complémentaires pour la caractérisation fine de complexes covalents à liaison cystéine-cystéine.....</b>	<b>172</b>
3.1. Préparation et traitement des échantillons.....	173
3.1.1. Préparation des échantillons.....	173
3.1.2. Les traitements réducteurs et alkylants différentiels.....	173
3.1.3. L'utilisation d'enzymes adaptées et complémentaires.....	175
3.2. Analyse et interprétation des données.....	177
3.2.1. Données MALDI-MS et données nanoLC-MS/MS.....	177
3.2.2. L'analyse de dipeptides par spectrométrie de masse en tandem.....	177
3.3. Conclusion.....	180
<b>4. Conclusion.....</b>	<b>181</b>

## Chapitre II

Mise en évidence de la formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroxydée par la sulfirédoxine.....	183
-----------------------------------------------------------------------------------------------------------------------------------------------------------	-----

<b>1. Suroxydation et régénération des peroxyrédoxine.....</b>	<b>184</b>
1.1. Le peroxyde d'hydrogène : un toxique et un messager cellulaire.....	184
1.1.1. L'activité du peroxyde d'hydrogène.....	184
1.1.2. Les peroxyrédoxines.....	184
1.2. Réversibilité de l'oxydation d'une cystéine sous forme acide sulfinique.....	185
1.2.1. Les sulfirédoxines régulent l'activité des peroxyrédoxines.....	185
1.2.2. Les hypothèses formulées pour expliquer le mécanisme enzymatique de Srx pour la réduction des Prx suroxydées.....	186
<b>2. La stratégie analytique mise en oeuvre pour la caractérisation des complexes peroxyrédoxine-sulfirédoxine.....</b>	<b>187</b>
2.1. Objectifs de l'étude.....	187
2.2. Contrôle de la réaction conduisant à la formation des complexes covalents Prx-Srx.....	187
2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre.....	188
<b>3. Résultats.....</b>	<b>189</b>
3.1. Analyse des complexes entiers par ESI-MS Haute Résolution.....	189
3.1.1. Suivi par LC-MS de la réaction conduisant à la formation des complexes Srx-Prx.....	189
3.1.2. Optimisation de la résolution en masse sur le système micrOTOF-Q™ (Bruker Dal.).....	190
3.1.3. Résultats des analyses LC-MS haute résolution.....	192
3.2. Caractérisation des complexes par nanoLC-MS/MS.....	194
3.2.1. Caractérisation indirecte de la liaison cystéine-cystéine entre la peroxyrédoxine C171A et la sulfirédoxine C48A/C106A.....	195
3.2.2. Caractérisation directe des complexes impliqués dans la réaction.....	196
3.3. Valorisation de ces travaux.....	198
3.3.1. Publication.....	198

3.3.2. Communications par affiches .....	198
<b>4. Conclusion .....</b>	<b>199</b>
<b>Chapitre III</b>	
<b>Etude de mécanismes de régénération de Methionine Sulfoxyde Réductases .....</b>	<b>201</b>
<b>1. Etude du mécanisme de réduction des méthionine sulfoxyde réductases par la thiorédoxine .....</b>	<b>202</b>
1.1. Le rôle des enzymes à activité méthionine sulfoxyde réductase .....	202
1.2. Stratégie analytique mise en oeuvre pour l'étude du mécanisme de réduction des méthionine sulfoxyde réductases par les thiorédoxines .....	203
1.2.1. Objectifs de l'étude .....	203
1.2.2. Formation des complexes Msr-Trx .....	203
1.2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre .....	204
1.3. Résultats .....	205
1.3.1. Les résultats acquis préalablement pour MsrA .....	205
1.3.2. Caractérisation des complexes MsrB-Trx selon la stratégie d'alkylation différentielle biotine/iodoacétamide par nanoLC-MS/MS .....	206
1.3.3. Analyses LC-MS des complexes entiers .....	210
1.4. Conclusion .....	212
<b>2. Etude du mécanisme de régénération des activités méthionine sulfoxyde réductase de la protéine périplasmique PilB .....</b>	<b>213</b>
2.1. Le rôle de la protéine PilB .....	213
2.1.1. Une structure à 3 domaines .....	213
2.1.2. Hypothèses de mécanismes de réduction du domaine MsrB .....	213
2.2. Stratégie analytique mise en oeuvre pour l'étude du mécanisme de réduction du domaine MsrB de PilB .....	214
2.2.1. Objectifs de l'étude .....	214
2.2.2. Contrôle de la réaction conduisant à la formation des complexes .....	215
2.2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre .....	215
2.3. Résultats .....	216
2.3.1. Analyses LC-MS de la protéine PilB entière .....	216
2.3.2. Caractérisation des complexes formés selon la stratégie nanoLC-MS/MS .....	218
2.3.3. Conclusions .....	221
2.4. Conclusions .....	222
<b>Conclusion générale .....</b>	<b>223</b>
<b>Bibliographie .....</b>	<b>229</b>
<b>Annexes .....</b>	<b>245</b>

# ABBREVIATIONS PRINCIPALES

<b>1-D</b>	Mono dimensionnel	<b>MRM</b>	<i>Multiple reaction monitoring</i>
<b>2-D</b>	Bidimensionnel	<b>MS<sup>E</sup></b>	<i>Expression mode mass spectrometry</i>
<b>ADN</b>	Acide désoxyribonucléique	<b>MS/MS</b>	Spectrométrie de masse en tandem
<b>ARN</b>	Acide ribonucléique	<b>NCBI</b>	<i>National Center for Biotechnology Information</i>
<b>BLAST</b>	<i>Basic local alignment search tool</i>	<b>OMSSA</b>	<i>Open Mass Spectrometry Search Algorithm</i>
<b>BN</b>	<i>Blue native</i>	<b>PAGE</b>	<i>Polyacrilamide gel electrophoresis</i>
<b>CD</b>	Cluster de différenciation	<b>PFF</b>	<i>Peptide fragment fingerprinting</i>
<b>DDA</b>	<i>Data dependant analysis</i>	<b>PHRC</b>	Programme hospitalier de recherche clinique
<b>DDBJ</b>	<i>DNA Data Bank of Japan</i>	<b>PIR</b>	<i>Protein Information Ressource</i>
<b>EBI</b>	<i>European Bioinformatics Institute</i>	<b>PMF</b>	<i>Peptide mass fingerprinting</i>
<b>EMBL</b>	<i>European Molecular Biology Laboratory</i>	<b>PRIDE</b>	<i>PRoteomics IDentifications database</i>
<b>ESI</b>	<i>Electrospray ionisation</i>	<b>Q</b>	Quadrupole
<b>EST</b>	<i>Expressed sequenced tag</i>	<b>RA</b>	<i>Retinoic acid</i>
<b>ETD</b>	<i>Electron transfer dissociation</i>	<b>RP</b>	<i>Reverse phase</i>
<b>FDR</b>	<i>False discovery rate</i>	<b>RMS</b>	<i>Root Mean Square</i>
<b>FPR</b>	<i>False positive rate</i>	<b>ROC</b>	<i>Receiver Operating Characteristic</i>
<b>FT-ICR</b>	<i>Fourier transform - Ion cyclotron resonance</i>	<b>RT-PCR</b>	<i>Reverse transcription-polymerase chain reaction</i>
<b>FWMH</b>	<i>Full width at half maximum</i>	<b>SCX</b>	<i>Strong cation exchange</i>
<b>GO</b>	<i>Gene ontology</i>	<b>SDS</b>	<i>Sodium dodecyl sulfate</i>
<b>HPLC</b>	<i>High performance liquid chromatography</i>	<b>SIB</b>	<i>Swiss Institute of Bioinformatics</i>
<b>HUPO</b>	<i>Human Proteome Organisation</i>	<b>SILAC</b>	<i>Stable isotope labelling with amino acids in cell culture</i>
<b>ICAT</b>	<i>Isotope-coded affinity tag</i>	<b>TAP</b>	<i>Tandem affinity purification</i>
<b>IT</b>	<i>Ion trap</i>	<b>TOF</b>	<i>Time of flight</i>
<b>iTRAQ</b>	<i>Isobaric tag for relative and absolute quantitation</i>	<b>UHPLC</b>	<i>Ultra high performance liquid chromatography</i>
<b>LC</b>	<i>Liquid chromatography</i>	<b>WB</b>	<i>Westen Blot</i>
<b>MALDI</b>	<i>Matrix assisted laser desorption ionisation</i>		
<b>MIAPE</b>	<i>The minimum information about a proteomics experiment</i>		
<b>MP</b>	Microparticule		
<b>MudPIT</b>	<i>Multidimensional Protein Identification Technology</i>		

# **INTRODUCTION GÉNÉRALE**

---

# Introduction générale

Il y a une vingtaine d'années, l'application des méthodes d'ionisation MALDI et Electrospray aux biomolécules a permis à la spectrométrie de masse de s'imposer rapidement comme la **technique analytique majeure** pour l'étude de protéines. Parallèlement, les projets de **séquençage des génomes** ont apporté les ressources nécessaires à l'identification rapide et robuste de protéines par la corrélation des résultats d'analyses de peptides par spectrométrie de masse avec des banques de séquences théoriques. La spectrométrie de masse a alors pris une place centrale dans l'**analyse protéomique** qui désigne l'analyse systématique des protéines de systèmes biologiques. L'objectif ultime est de produire une description détaillée des structures de ces systèmes pour aboutir à une meilleure compréhension de leurs fonctions et de leurs mécanismes de contrôle.

L'engouement suscité par les nouvelles possibilités offertes par l'analyse protéomique a d'abord masqué des questions essentielles comme **la qualité et la validation des résultats**. Il faut pourtant encore relever **des difficultés majeures** pour identifier toutes les protéines d'un échantillon et en obtenir un recouvrement de séquence complet. Actuellement, l'information recueillie est largement **incomplète et incertaine** : on s'appuie toujours sur des probabilités.

Les progrès à réaliser passent d'abord par des **avancées technologiques en spectrométrie de masse**. Des optimisations doivent être menées pour davantage de sensibilité et de résolution, une meilleure précision de mesure de masse et des vitesses d'acquisition accrues.

Il s'agit surtout d'articuler ces atouts techniques autour de stratégies affinées en accord avec les questions biologiques posées. Ainsi, pour **identifier et caractériser finement le contenu protéique d'échantillons biologiques complexes**, l'amélioration des préparations d'échantillons, des techniques de séparation et des paramètres de mesure est essentielle. Le développement et l'optimisation d'**outils bioinformatiques** pour la génération de données protéomiques, leur manipulation et leur validation le sont tout autant.

## Présentation de ce travail de thèse

J'ai souhaité ouvrir la présentation de mon travail de thèse par une **partie préliminaire** composée de trois chapitres d'**introduction bibliographique**. Chacun représente un des trois piliers soutenant l'essor de l'analyse protéomique par spectrométrie de masse :

- Les développements technologiques et instrumentaux pour l'analyse de peptides et protéines par spectrométrie de masse et pour leur séparation ;
- Les stratégies d'identification et de caractérisation de protéines et les solutions méthodologiques permettant de surmonter certains défis comme la quantification de protéines, l'analyse protéomique fonctionnelle, ou l'étude de protéines membranaires ;

▸ Les outils et les approches de bioinformatique pour l'exploitation, la validation et le partage des données générées.

Dans ce contexte, **la partie « résultats »** de mon travail de thèse s'articule autour d'optimisations et de développements méthodologiques sur deux thématiques traitées dans deux parties distinctes.

La **première partie** est consacrée au développement de nouvelles méthodologies alliant l'utilisation combinée de nouveaux outils et de stratégies pour l'amélioration des analyses protéomiques classiques.

▸ Un premier chapitre rapporte le travail accompli pour **évaluer le nombre d'analyses nanoLC-MS/MS à réaliser pour une comparaison fiable du contenu protéique d'échantillons biologiques complexes**. Parallèlement, des optimisations instrumentales fines ont permis une meilleure prise des données protéomiques et l'émergence de nouveaux outils bioinformatiques a accompagné un gain de confiance dans les informations collectées.

Ces optimisations, innovations instrumentales et stratégiques ont été appliquées dans le cadre de deux collaborations pour répondre à deux problématiques biologiques.

▸ Le deuxième chapitre est consacré à **la recherche de nouveaux biomarqueurs caractéristiques d'hémopathies lymphoïdes chroniques au diagnostic ambigu**. Les recherches se sont notamment focalisées sur l'identification et la validation de marqueurs de surface spécifiques du lymphome du manteau et de la maladie de Waldenström. La phase de découverte a permis la mise en oeuvre des optimisations réalisées pour l'analyse protéomique différentielle d'échantillons enrichis en protéines membranaires grâce à l'utilisation de microparticules de lymphocytes T.

▸ Le troisième chapitre traite de **l'identification de partenaires d'interaction du facteur de transcription Trim24 connu pour être impliqué dans l'hépatocarcinogénèse**. Une approche d'analyse protéomique fonctionnelle a été développée pour l'étude des sous-complexes dans lesquels Trim24 est potentiellement impliqué.

La **deuxième partie** est consacrée au développement de méthodologies pour la caractérisation de complexes protéiques covalents à liaison cystéine-cystéine.

▸ Le premier chapitre détaille **la mise au point de stratégies complémentaires en spectrométrie de masse pour l'analyse d'intermédiaires réactionnels impliqués dans des mécanismes d'oxydoréduction intracellulaires**. Il s'agit de déterminer les espèces mises en jeu et d'établir le type et la localisation de liaisons covalentes entre résidus cystéiques. Notamment la mise au point de stratégies d'analyse LC-MS haute résolution et d'approches nanoLC-MS/MS pour l'analyse de dipeptides à ponts disulfures ont permis l'étude de mécanismes réactionnels d'oxydoréduction dans le cadre de deux collaborations.

▸ Le deuxième chapitre expose le travail réalisé pour **la mise en évidence de la formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroxydée par la sulfirédoxine chez *Saccharomyces cerevisiae***.

▸ Le troisième chapitre détaille **l'étude de mécanismes de régénération de protéines oxydées, par des méthionine sulfoxyde réductases (MrsB, PilB) chez *Neisseria meningitidis***. Des stratégies ont été mises au point pour stabiliser, purifier et identifier les intermédiaires formés. Dans l'objectif de caractériser les complexes et déterminer les cystéines impliquées dans chacun d'eux, des analyses nanoLC-MS/MS différentielles ont notamment pu être mises en oeuvre.

*Ce travail de thèse a été soutenu par le réseau européen de formation des doctorants GRK532 et par le ministère français de la recherche.*

# **INTRODUCTION BIBLIOGRAPHIQUE**

## **Chapitre I**

Instruments et technologies  
au service de l'analyse protéomique

## **Chapitre II**

Les stratégies d'identification en analyse  
protéomique par spectrométrie de masse

## **Chapitre III**

Les données générées en analyse protéomique :  
gestion, validation et partage de l'information

---

# Introduction

Depuis le milieu des années 1990, la spectrométrie de masse a acquis rapidement une maturité lui permettant de devenir un outil remarquable pour l'analyse de biomolécules, tout particulièrement des protéines<sup>1</sup>. Ce développement s'est nourri, dans un champ interdisciplinaire, de l'émergence conjointe de nouvelles stratégies analytiques, d'innovations instrumentales et de nouvelles techniques préparatives et séparatives d'échantillons protéiques. Alors que les projets de séquençage du génome de multiples organismes – notamment celui de l'Homme – ont abouti, l'analyse de protéines par spectrométrie de masse est entrée dans une nouvelle ère : celle de l'analyse protéomique<sup>2</sup>.

Avec l'accumulation de séquences d'ADN dans les banques génomiques, il est apparu qu'il n'était pas suffisant de disposer de séquences complètes de génomes pour comprendre les fonctions biologiques des protéines pour lesquelles ces séquences codent<sup>3</sup>. En effet, l'identification des protéines, la caractérisation de leur séquence en acides aminés et de leurs modifications posttraductionnelles, la découverte de produits de l'épissage alternatif ne sont possibles qu'en étudiant les protéines elles-mêmes. Les mécanismes de régulation et la dynamique de l'expression protéique font également partie des défis à relever pour traduire l'information issue des données génomiques en connaissances biologiques<sup>4</sup>.

Pour y parvenir, la spectrométrie de masse s'est imposée comme l'outil de choix puisqu'aucune autre technologie ne peut aujourd'hui recueillir autant d'informations avec autant de sensibilité et de rapidité. Les développements technologiques et méthodologiques pour l'analyse protéomique par spectrométrie de masse conduisent encore à la mise au point ou à l'amélioration de stratégies toujours plus poussées pour mieux comprendre les fondements moléculaires de processus cellulaires et physiologiques<sup>5</sup>.

Cette introduction bibliographique présente dans un premier chapitre les avancées instrumentales et technologiques qui ont permis à la spectrométrie de masse de devenir une technique dominante en analyse protéomique. Puis, une deuxième partie détaille les principales approches permettant l'identification des protéines et met l'accent sur quelques défis stratégiques en analyse protéomique. Enfin, le troisième chapitre traite de la gestion et de la validation des données générées et souligne les difficultés liées au partage de l'information.

---

<sup>1</sup> Domon, B., et al. (2006). *Science* **312**, 212–7.

<sup>2</sup> Patterson, S. D., et al. (2003). *Nat Genet* **33 Suppl**, 311–23.

<sup>3</sup> Fields, S. (2001). *Science* **291**, 1221–4.

<sup>4</sup> Pandey, A., et al. (2000). *Nature* **405**, 837–46.

<sup>5</sup> Cravatt, B. F., et al. (2007). *Nature* **450**, 991–1000.

## Définitions

Le terme « **protéome** » a été proposé pour la première fois en 1994 par Mark Wilkins<sup>6,7</sup>, un étudiant australien travaillant à l'identification rapide de protéines. Ce mot-valise, construit à partir des substantifs **protéine** et **génome** pour désigner « l'ensemble des protéines produites par un génome, ou dans une cellule ou un tissu », a rapidement séduit la communauté scientifique.

Beaucoup de définitions plus précises ont été proposées avec plus ou moins de succès. A la différence du génome, le protéome n'est pas unique à chaque organisme et il est vain de tenter de figer sa conceptualisation<sup>8</sup>. Les protéines présentes dans une cellule, ainsi que leur fonction, leur localisation, et même leur structure peuvent évoluer selon l'âge, la différenciation ou l'état pathologique de la cellule, les stress environnementaux ou de multiples autres évènements<sup>9</sup>.

Alors, pour désigner l'étude du protéome, le terme « **protéomique** » s'est imposé de lui-même. Il désigne non seulement l'analyse des protéines produites dans une cellule, mais aussi l'ensemble de leurs isoformes et des modifications qu'elles portent, les interactions qu'elles nouent, les organisations structurales et les complexes dans lesquels elles sont engagées<sup>10</sup>.

Par conséquent, l'analyse protéomique vise à l'identification, la caractérisation et la quantification des protéines d'une cellule ou d'un système définis pour établir comment leurs propriétés varient dans l'espace, le temps et les états physiologiques.

---

<sup>6</sup> Cohen, J. (2001). *Technology Review* **104**, 54–60.

<sup>7</sup> Wilkins, M. R., et al. (1996). *Biotechnology (N Y)* **14**, 61–5.

<sup>8</sup> Kenyon, G. L., et al. (2002). *Mol Cell Proteomics* **1**, 763–80.

<sup>9</sup> Huber, L. A. (2003). *Nat Rev Mol Cell Biol* **4**, 74–80.

<sup>10</sup> Tyers, M., et al. (2003). *Nature* **422**, 193–7.

---

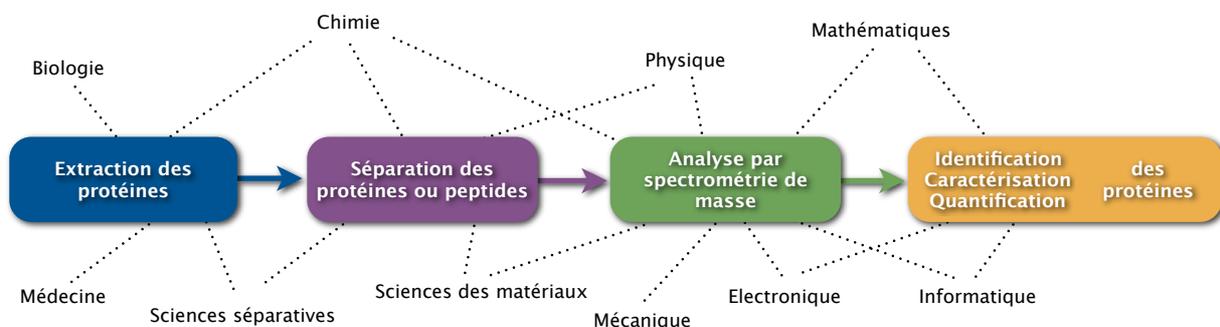
# Chapitre I

## Instruments et technologies au service de l'analyse protéomique

---

L'analyse protéomique par spectrométrie de masse a pris son essor grâce au développement simultané de multiples disciplines académiques (Figure 1). Elles ont notamment contribué à la maturation de trois domaines essentiels en accompagnant :

- le développement de méthodes séparatives de protéines et peptides, comme l'électrophorèse sur gel 1-D ou 2-D ou la chromatographie liquide ;
- des innovations instrumentales majeures en spectrométrie de masse pour l'analyse de biomolécules ;
- l'émergence d'outils bioinformatiques permettant la construction et la recherche dans des banques de données protéiques, ressource de base pour l'identification de protéines par spectrométrie de masse.



**Figure 1** | Disciplines académiques ayant permis l'essor de l'analyse protéomique par spectrométrie de masse.

# 1. La spectrométrie de masse pour la caractérisation des protéines

C'est au début du XXe siècle qu'apparaît la spectrométrie de masse, avec les découvertes de Thompson et Aston<sup>11</sup>. Il faut pourtant attendre 1942 pour voir le premier instrument commercialisé aux Etats-Unis. Dans les 40 années qui suivent, l'analyse de biomolécules par spectrométrie de masse ne reste qu'à ses balbutiements, mais l'apparition de nouveaux principes de séparation des ions en phase gazeuse pose les fondements des futurs développements<sup>12</sup>. Les années 1980 connaissent une rupture technologique avec la découverte de 2 techniques d'ionisation douces qui révolutionnent l'analyse des molécules biologiques : le MALDI (Matrix Assisted Laser Desorption Ionisation) et l'ionisation Electrospray (ESI)<sup>13</sup>. Ces techniques, qui rendent possibles l'ionisation et le transfert de macromolécules intactes en phase gazeuse, ont ouvert la voie à la spectrométrie de masse moderne pour l'analyse protéomique. Leur découverte est récompensée en 2002 par un Prix Nobel de Chimie.

Un spectromètre de masse est classiquement composé de 3 éléments : une source d'ionisation, un analyseur de masse, et un détecteur de courant d'ions (Figure 2). Les molécules, en phase liquide ou solide, doivent être ionisées et transférées en phase gazeuse. Puis les ions sont séparés dans un analyseur en fonction de leur rapport masse-sur-charge ( $m/z$ ). Le mouvement des ions dans l'analyseur est assuré sous l'effet de champs électriques et/ou magnétiques : la vitesse et/ou la trajectoire des ions jusqu'au détecteur dépendent donc de leur  $m/z$  caractéristique. Au bout de leur course, les ions frappent le détecteur qui convertit alors ces événements en signal électrique.

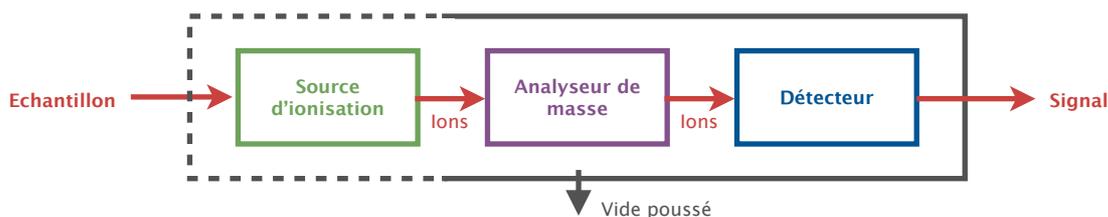


Figure 2 | Principales fonctions des 3 éléments de base d'un spectromètre de masse.

## 1.1. Les sources d'ionisation

Les techniques d'ionisation des molécules biologiques comme les protéines ou les peptides doivent permettre de les vaporiser sans les fragmenter ni les dégrader. Les sources Electrospray et MALDI sont aujourd'hui les 2 sources les plus courantes en analyse protéomique.

### 1.1.1. La source Electrospray (ESI) et Nanoelectrospray (nanoESI)

L'ionisation electrospray (ESI) est réalisée par la nébulisation à pression atmosphérique d'un fin nuage d'ions créé sous l'effet d'un fort champ électrique ( $10^6 \text{ V.m}^{-1}$ ) et de l'assistance pneumatique d'un gaz de nébulisation<sup>14</sup>. En appliquant une différence de potentiel entre le capillaire à travers lequel la solution d'analyte s'écoule et l'entrée du spectromètre de masse, les espèces de charge opposée en solution sont séparées et des gouttelettes multichargées sont produites (Figure 3). Le chauffage de l'enceinte dans laquelle les gouttelettes sont produites favorise l'évaporation du solvant. Le volume des gouttelettes chargées rétrécit

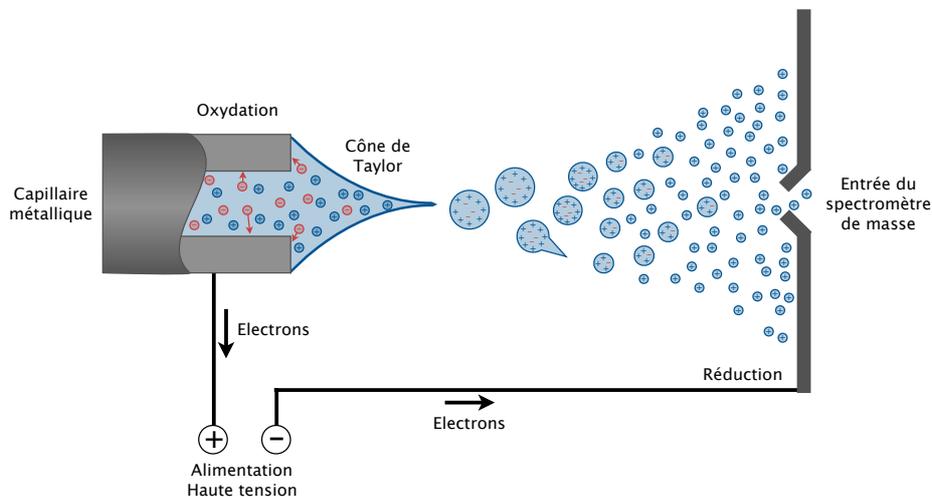
<sup>11</sup> Thompson, J. J. (1913). Rays of positive electricity and their application to chemical analysis., Longmans Green.

<sup>12</sup> Gelpi, E. (2008). J Mass Spectrom **43**, 419–35.

<sup>13</sup> Gelpi, E. (2009). J Mass Spectrom **44**, 1137–61.

<sup>14</sup> Fenn, J. B., et al. (1989). Science **246**, 64–71.

jusqu'à ce que la limite de stabilité de Rayleigh soit atteinte : les forces de répulsions coulombiennes dépassent alors forces de tension de surface et les gouttelettes subissent une succession d'explosions coulombiennes à mesure que le solvant s'évapore, jusqu'à ce que l'ion se trouve nu. Deux modèles sont admis pour expliquer la production d'ions désolvatés en phase gazeuse : Dole<sup>15</sup> décrit le « mécanisme de la charge résiduelle » où les ions sont progressivement désolvatés ; et Iribarne et Thomson<sup>16</sup> proposent le « mécanisme de l'évaporation unique » où des ions peuvent être directement désorbés des gouttelettes.



**Figure 3** | Principe de l'ionisation électrospray (mode d'ionisation positif).

Grâce à l'ionisation électrospray, des molécules en solution sont rapidement transformées sous forme d'ions en phase gazeuse, à condition qu'elles puissent être ionisées. Lorsque plusieurs sites sont ionisables, les ions produits sont multichargés<sup>17</sup>. Dans cette situation les calculs de masse moléculaire sont plus compliqués puisque les grandeurs mesurées sont des rapports  $m/z$  avec  $z \neq 1$ . Cependant, en présence de charges multiples, les valeurs de  $m/z$  mesurées sont plus petites et la gamme de masse réelle mesurée par un spectromètre de masse peut être largement étendue<sup>18</sup>.

Une amélioration majeure des performances de l'ionisation électrospray est née de la possibilité de diminuer le débit de la solution utilisée pour créer le spray. Les premières sources nanospray ( $20 \text{ nL} \cdot \text{min}^{-1}$ ) sont apparues en 1996<sup>19</sup>. A bas débit, le petit diamètre de l'orifice de sortie du capillaire est l'élément clé permettant d'assurer la stabilité du spray<sup>20</sup>. Dans ces montages, la taille des gouttelettes générées est plus petite et aucune assistance pneumatique n'est plus nécessaire pour la formation du spray. La miniaturisation des sources ESI présente l'avantage majeur d'augmenter la concentration des analytes qui éluent d'une colonne de chromatographie liquide à leur entrée dans le spectromètre de masse, et donc d'améliorer la sensibilité de la mesure, qui est dépendante de la concentration en mode d'ionisation ESI. De plus, à plus faible débit et pour un même volume d'échantillon, la durée d'analyse – et donc la précision de mesure – peuvent être augmentées.

Avec le développement des microtechnologies, de nouveaux dispositifs électrospray sur puces microfluidiques ont pu être mis au point<sup>21</sup>. Ils présentent le double avantage de permettre une conception reproductible des pièces (notamment par ablation laser, photolithographie ...) et d'éviter la manipulation délicate des fragiles aiguilles nanoelectrospray grâce à des dispositifs intégrés et motorisés. Ces dispositifs sont généralement associés à des systèmes automatisés de préparation des échantillons (nanoESI Triversa

<sup>15</sup> Dole, M., et al. (1968). *The Journal of Chemical Physics* **49**, 2240–2249.

<sup>16</sup> Iribarne, J. V., et al. (1976). *The Journal of Chemical Physics* **64**, 2287–2294.

<sup>17</sup> Kebarle, P., et al. (2009). *Mass Spectrom Rev* **28**, 898–917.

<sup>18</sup> de Hoffmann, E., et al. (2007). *Mass Spectrometry: Principles and Applications*, Wiley.

<sup>19</sup> Wilm, M., et al. (1996). *Anal Chem* **68**, 1–8.

<sup>20</sup> Gibson, G. T., et al. (2009). *Mass Spectrom Rev* **28**, 918–36.

<sup>21</sup> Koster, S., et al. (2007). *Lab Chip* **7**, 1394–412.

Nanomate™, Advion Biosciences) ou à des systèmes de séparation d'électrophorèse capillaire ou de chromatographie liquide (HPLC-Chip™, Agilent Technologies ; cf. 1<sup>re</sup> partie des résultats, chapitre I, point 2.1.1.).

### 1.1.2. La source MALDI

Les tentatives d'utiliser l'énergie laser pour transférer des molécules en phase gazeuse par désorption thermique ont longtemps été infructueuses. En 1988, Koichi Tanaka a finalement montré que des ions pouvaient être désorbés d'une solution à base de glycérol contenant une poudre métallique ultra fine<sup>22,23</sup>. Plutôt que d'ajuster l'énergie laser aux longueurs d'onde d'absorption des analytes, une matrice qui absorbe l'énergie du laser est utilisée pour assister et promouvoir la désorption thermique. Ces premières expériences ont rapidement conduit au développement d'une technique simple et fiable de production d'ions en phase gazeuse sous vide poussé ( $10^{-7}$  bar) – l'ionisation par désorption laser assistée par matrice (MALDI) – qui utilise la cocrystallisation de l'analyte avec une matrice organique<sup>24</sup>.

Les phénomènes qui se produisent sont encore débattus, mais un modèle en trois étapes semble se dégager<sup>25</sup> (Figure 4) :

- ▶ L'énergie du laser pulsé ( $\lambda_{UV} = 314$  nm) est transférée à la matrice, induisant un rapide échauffement des molécules ;
- ▶ Les analytes sont désorbés sous l'effet d'un abaissement des forces de cohésion des cristaux ;
- ▶ Les molécules sont ionisées en phase gazeuse par un transfert de proton soit lors de collisions, soit lors de réactions acide – base entre les molécules de matrice et des analytes<sup>26</sup>. De fait, les ions produits sont généralement monochargés.

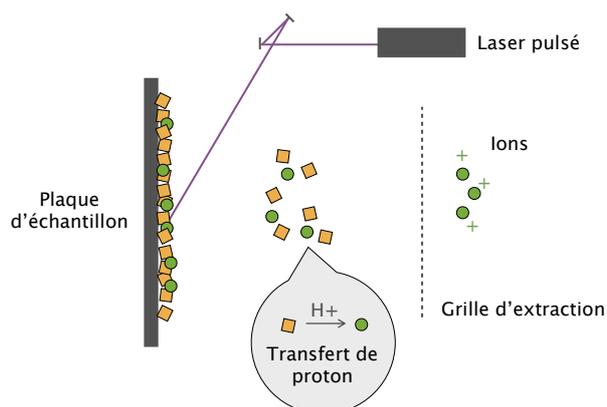


Figure 4 | Principe de l'ionisation MALDI. (Adapté de Aebersold, 2003)

D'un point de vue pratique, plusieurs paramètres entrent en jeu pour une ionisation efficace. Le choix de la matrice ( $\alpha$ -cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid, 2',4',-dihydroxyacetophenone) et le mode de dépôt (couche mince, sandwich ou goutte séchée) sont par exemple essentiels. Ils dépendent grandement de la nature de l'échantillon (protéines ou peptides, tampon de solubilisation, pH ...), sachant que – le dépôt étant solide – la présence de sel n'est pas à proscrire.

<sup>22</sup> Tanaka, K., et al. (1988). *Rapid Commun Mass Spectrom* **2**, 151–153.

<sup>23</sup> Tanaka, K. (2003). *Angew Chem Int Ed Engl* **42**, 3860–70.

<sup>24</sup> Hillenkamp, F., et al. (1991). *Anal Chem* **63**, 1193A–1203A.

<sup>25</sup> Zenobi, R., et al. (1998). *Mass Spectrometry Reviews* **17**, 337–366.

<sup>26</sup> Knochenmuss, R., et al. (2003). *Chem Rev* **103**, 441–52.

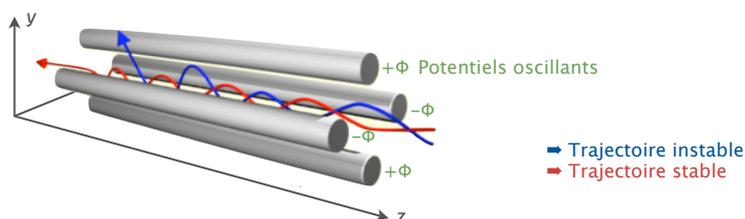
## 1.2. Les analyseurs

L'analyseur de masse est – au sens propre comme au sens figuré – la pièce centrale du spectromètre de masse. Dans le contexte de l'analyse protéomique, les caractéristiques fondamentales d'un analyseur sont sa sensibilité, sa résolution, sa précision de mesure de masse, sa gamme de masse et ses capacités à produire des spectres de fragmentation informatifs<sup>27</sup>. Il existe 4 types d'analyseurs couramment utilisés en analyse protéomique : l'analyseur quadripolaire (Q), l'analyseur à temps de vol (TOF), les trappes d'ions (IT), les analyseurs à transformée de Fourier (FT-ICR, Orbitrap). Ils peuvent fonctionner en tant que tel ou certaines associations peuvent être réalisées pour former des instruments hybrides, tirant parti des avantages de chacun des analyseurs assemblés<sup>28</sup>.

### 1.2.1. Les analyseurs simples pour la spectrométrie de masse MS

#### ▶ L'analyseur quadripolaire (Q)

Le principe du quadripôle<sup>29</sup> est décrit dès les années 1950. Ses capacités d'analyse d'un courant d'ions continu font de la source ESI son partenaire idéal. L'analyseur quadripolaire est constitué de 4 barres métalliques parallèles, opposées 2 à 2 (Figure 5). L'un et l'autre des couples de barres sont portés à un potentiel oscillant opposé (caractérisé par son amplitude, sa fréquence et sa composante continue)<sup>30</sup>. La séparation des ions dans le quadripôle en fonction de leur rapport  $m/z$  est fondée sur la stabilité de leur trajectoire entre les barres. Accélérés selon l'axe longitudinal, les ions subissent des accélérations dans le plan orthogonal en fonction des potentiels appliqués. Les équations de Mathieu (1866) décrivent ces trajectoires et permettent de déterminer les conditions de stabilité pour chaque masse et charge. Un balayage en potentiel permet donc de transmettre successivement chacun des ions selon leur caractéristique  $m/z$ <sup>31</sup>.



**Figure 5** | Trajectoires stable et instable d'ions dans l'analyseur quadripolaire ( $\pm\Phi$  : potentiels oscillants appliqués aux barres parallèles). D'après Elisemariou, CC-BY-SA.

#### ▶ L'analyseur à temps de vol (TOF)

Le principe de l'analyseur à temps de vol (« *time-of-flight* », TOF)<sup>32</sup> a été inventé dans les années 50, mais il faut attendre la fin des années 1980 et l'apparition des sources MALDI<sup>33</sup> pour qu'il soit pleinement reconnu.

L'analyseur TOF permet de séparer des ions et de mesurer leur rapport  $m/z$  en fonction du temps qu'ils mettent à parcourir un tube sous vide poussé et libre de champ, après qu'ils ont été accélérés. En

<sup>27</sup> Aebersold, R., et al. (2003). *Nature* **422**, 198–207.

<sup>28</sup> Domon, B., et al. (2006). *Science* **312**, 212–7.

<sup>29</sup> Paul, W., et al. (1953). *Zeitschrift für Naturforschung A* **8**, 448–450.

<sup>30</sup> Campana, J. E. (1980). *International Journal of Mass Spectrometry and Ion Processes* **33**, 101–117.

<sup>31</sup> Douglas, D. J. (2009). *Mass Spectrom Rev* **28**, 937–60.

<sup>32</sup> Wiley, W. C., et al. (1955). *Review of Scientific Instruments* **26**, 1150–1157.

<sup>33</sup> Karas, M., et al. (1987). *International Journal of Mass Spectrometry and Ion Processes* **78**, 53–68.

appliquant le théorème de l'énergie cinétique, la vitesse  $v$  [ $\text{m}\cdot\text{s}^{-1}$ ] d'un ion de masse  $m$  [kg] et de charge  $z.e$  [C], accéléré par une différence de potentiel  $U$  [V] s'écrit :  $v = (2zeU/m)^{1/2}$ .

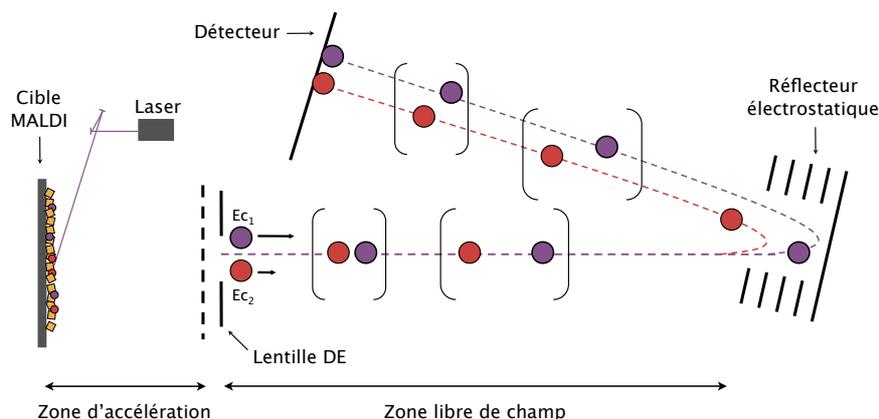
L'ion dans la zone libre de champ conserve son énergie. Le temps de vol  $t$  [s] de cet ion sur une distance  $L$  [m] est donc directement donné par la relation :  $t = L(m/2zeU)^{1/2}$ .

Les premiers analyseurs TOF souffraient d'une résolution médiocre imputable aux variations de distributions temporelles (temps de formation), spatiales (lieu de formation) et cinétiques (énergie de formation) des ions. Deux améliorations majeures ont permis de pallier à ces défauts (Figure 6).

▸ Les dispersions spatiale et temporelle peuvent être limitées en utilisant le principe d'extraction retardée (Delay extraction)<sup>34</sup>. Il a pu être mis en oeuvre en ajoutant une lentille supplémentaire entre une source MALDI et l'analyseur. L'application d'un potentiel à cette lentille, légèrement plus élevé que celui de la cible, permet de freiner les ions formés pendant un court instant après le tir laser. Les ions sont alors focalisés et ont un départ synchronisé pour pénétrer dans le tube de vol.

▸ La dispersion cinétique peut être corrigée grâce à la mise en place d'un réflecteur électrostatique (« reflectron »)<sup>35</sup>. Ce dispositif est composé d'une série d'anneaux ou de grilles sur lesquels sont appliqués des potentiels croissants et qui font office de miroirs électrostatiques. Ils renvoient les ions vers le détecteur, mais en fonction de l'énergie cinétique qu'ils possèdent, les ions pénétreront plus ou moins profondément dans le réflecteur. Des ions de même rapport  $m/z$  se trouvent alors refocalisés et atteindront le détecteur au même moment.

Par ailleurs, pour les instruments les plus récents, les performances de résolution ont encore pu être augmentées grâce à l'amélioration des techniques de refocalisation des ions. Elles ont principalement permis sur une augmentation de la longueur du tube de vol (jusqu'à environ 250 cm, MaXis Bruker Dal.) et la mise au point de doubles réflecteurs électrostatiques (géométrie en W, Synapt HDMS Waters Corp.).



**Figure 6** | Principe de fonctionnement d'un instrument MALDI-TOF avec réflecteur électrostatique et système d'extraction retardée (DE). ● et ● sont des ions de même  $m/z$ .

#### ▸ L'analyseur trappe d'ions (IT)

Dans les années 50, W. Paul et H. Steinwedel décrivent pour la 1<sup>re</sup> fois de nouveaux analyseurs de type trappe ionique. Ce travail a été reconnu par un Prix Nobel<sup>36</sup> décerné en 1989 à Wolfgang Paul et de nombreux autres groupes ont par la suite participé à l'amélioration des performances de l'analyseur notamment en termes de détection, stockage et éjection sélectifs des ions<sup>37</sup>.

La trappe ionique est constituée de 3 électrodes, une électrode annulaire en forme de diabolo et deux électrodes chapeaux à géométrie quasi hyperbolique (Figure 7). L'ensemble forme une cavité d'environ 1

<sup>34</sup> Vestal, M. L., et al. (1995). Rapid Communications in Mass Spectrometry 9, 1044-1050.

<sup>35</sup> Mamyrin, B. A., et al. (1973). Sov. Phys. JETP 37, 45.

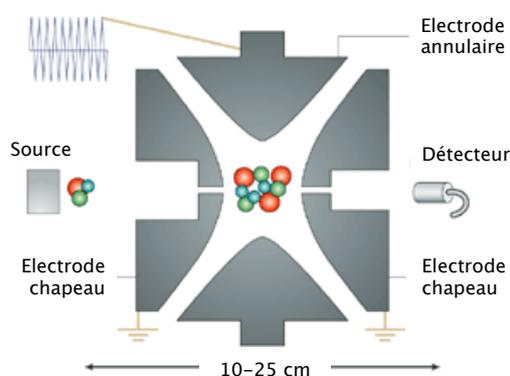
<sup>36</sup> Paul, W. (1990). Angewandte Chemie International Edition in English 29, 739-748.

<sup>37</sup> March, R. E. (1998). Rapid Communications in Mass Spectrometry 12, 1543-1554.

cm<sup>2</sup>. Les ions entrent et sortent de la trappe par des orifices situés dans les électrodes chapeaux. A la différence des analyseurs décrits précédemment, les ions restent confinés à l'intérieur de la trappe sans la traverser directement. Au sein du piège règne une pression résiduelle d'hélium de l'ordre de 5.10<sup>-5</sup> bar permettant d'augmenter l'efficacité de piégeage des ions. Une fois qu'ils ont pénétré dans la trappe, une tension oscillante appliquée à l'électrode annulaire permet de générer un champ quadripolaire qui stabilise la trajectoire des ions en les piégeant dans un puits de potentiel dont la profondeur est proportionnelle à l'amplitude de la tension appliquée<sup>38,39</sup>. Les équations de Mathieu régissent cette trajectoire qui décrit dans l'espace une courbe de Lissajou. L'éjection des ions de la trappe peut être :

- une éjection simple : une diminution de l'amplitude de la tension réduit la profondeur du puits de potentiel et permet aux ions de m/z croissant de s'échapper du piège.

- Une éjection résonante<sup>40</sup> : les ions de m/z croissants, chacun oscillant avec une fréquence qui lui est propre, sont soumis à un champ électrique oscillant supplémentaire appliqué entre les électrodes chapeaux et choisi spécifiquement pour faire entrer les ions de m/z croissant en résonance et donc les éjecter prématurément.



**Figure 7** | Schéma de l'analyseur trappe d'ions. Les tensions oscillantes appliquées au niveau de l'électrode annulaire permettent de piéger les ions sur des trajectoires stables. L'analyse est réalisée en déstabilisant les trajectoires des ions de rapport m/z défini. (D'après Glish, 2003<sup>41</sup>)

Cet analyseur à la fois robuste, sensible et de coût modéré est un instrument de choix pour l'analyse protéomique. Cependant, sa précision de mesure de masse constitue une de ses principales limites. Elle est en effet limitée en partie par la capacité de piégeage d'ions dans la trappe, qui s'ils sont trop nombreux subissent des effets d'espace-charge perturbant les champs appliqués. Les constructeurs font encore des progrès dans la conception de systèmes toujours plus performants : Bruker Daltonics par exemple propose de nouvelles générations de trappes d'ions (HCT Ultra, AmaZon). La focalisation des ions avant leur entrée dans la trappe est améliorée grâce à des systèmes de lentilles électrostatiques (« *funnels* ») en lieu et place des hexapôles classiques. De plus, de nouveaux principes d'éjection plus rapides des ions de la trappe sont mis en oeuvre (avant d'entrer en résonance, les ions à éjecter sont préexcités sur une trajectoire plus large que celle des autres ions).

En parallèle, une génération d'instruments de type trappe d'ions linéaire ou 2-D, présentant des capacités de piégeage beaucoup plus élevées que les trappes 3D, ont été développées. Ils apportent des améliorations tant au niveau de la sensibilité que de la résolution et de la précision de masse<sup>42, 43</sup>.

<sup>38</sup> Dehmelt, H. (1967). *Adv. At. Mol. Phys.* 3, 53.

<sup>39</sup> March, R. E. (2009). *Mass Spectrom Rev* 28, 961-89.

<sup>40</sup> March, R. E. (1997). *Journal of Mass Spectrometry* 32, 351-369.

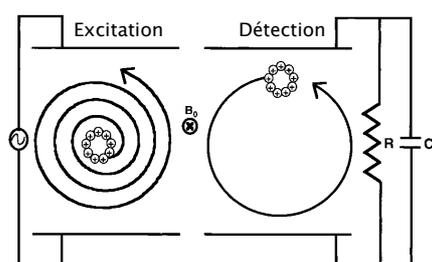
<sup>41</sup> Glish, G. L., et al. (2003). *Nat Rev Drug Discov* 2, 140-50.

<sup>42</sup> Ouyang, Z., et al. (2004). *Anal Chem* 76, 4595-605.

<sup>43</sup> Ouyang, Z., et al. (2007). *Eur J Mass Spectrom (Chichester, Eng)* 13, 13-8.

▸ L'analyseur à résonance cyclotronique des ions à transformée de Fourier (FT-ICR)

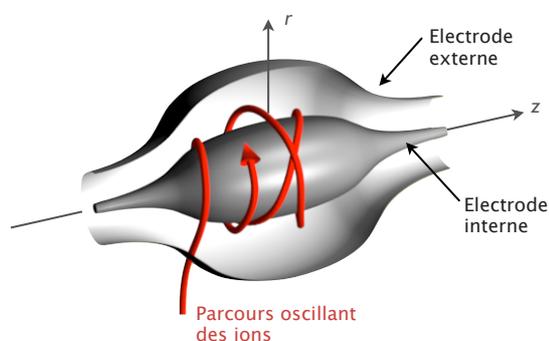
Dans les années 1970, un nouveau type de spectromètre de masse aux capacités résolutive étonnantes a été introduit : l'analyseur à résonance cyclotronique des ions à transformée de Fourier<sup>44</sup>. Les FT-ICR piègent les ions dans une cellule où il règne un champ magnétique uniforme qui impose aux ions une trajectoire cyclotronique (Figure 8). L'application d'un champ électrique balayant une gamme de fréquences identique à celle de rotation des ions rend leur mouvement cohérent. Le mouvement cyclotronique cohérent génère un signal électrique sur deux électrodes opposées qui est mesuré et converti en fréquence par transformée de Fourier. Les valeurs de  $m/z$  des ions, directement reliées à la fréquence de rotation des ions, sont alors déterminées. Un désavantage majeur du FT-ICR est la nécessité de disposer d'aimants supraconducteurs permettant la génération de champs magnétiques très hauts (18 Tesla pour le plus récent), le pouvoir résolutif et la gamme de masse de l'instrument étant directement dépendants de cette valeur<sup>45</sup>.



**Figure 8** | Le mouvement cyclotronique incohérent des ions dans le champ magnétique  $B_0$  est converti en mouvement cohérent (et donc détectable) par l'application d'un champ électrique oscillant à la fréquence de chaque ion de  $m/z$  donné. (D'après Marshall, 1998)

▸ L'analyseur Orbitrap™

Au début des années 2000, l'Orbitrap – un nouveau type d'analyseur – a été introduit sur le marché<sup>46</sup>. C'est la 1<sup>re</sup> fois en 30 ans qu'un analyseur basé sur un nouveau principe physique a émergé. L'instrument présente des caractéristiques similaires au FT-ICR en termes de résolution et de précision de mesure de masse, mais ne nécessite pas l'utilisation d'un aimant supraconducteur cher et très contraignant. En effet, la séparation des ions n'a pas lieu grâce à l'application d'un champ magnétique, mais d'un champ électrique oscillant. L'Orbitrap est composé de deux électrodes imbriquées – une interne et une externe – dont la forme est spécialement conçue pour qu'un champ électrostatique quadralogarithmique soit généré (Figure 9). Les ions tournent autour de l'électrode interne et oscillent de manière harmonique selon l'axe longitudinal avec une fréquence caractéristique de leurs valeurs de  $m/z$ . Une image du courant induit par ces oscillations est convertie en un spectre de fréquence grâce à la transformée de Fourier selon une approche similaire à celle utilisée pour le FT-ICR.



**Figure 9** | Mouvement oscillant des ions autour de l'électrode interne de l'Orbitrap. (D'après Gbdivers, CC-BY-SA)

<sup>44</sup> Marshall, A. G., et al. (1998). *Mass Spectrom Rev* **17**, 1–35.

<sup>45</sup> Marshall, A. G., et al. (1996). *Rapid Commun Mass Spectrom* **10**, 1819–1823.

<sup>46</sup> Makarov, A. (2000). *Anal Chem* **72**, 1156–62.

## 1.2.2. Les analyseurs pour la spectrométrie de masse MS/MS

Grâce aux analyseurs classiques qui viennent d'être présentés, les rapports masse-sur-charge d'ions intacts générés par ionisation MALDI ou ESI peuvent être rapidement et précisément mesurés. Cependant, ces données n'informent pas sur la structure covalente des ions analysés. En analyse protéomique, on cherche notamment à déterminer la séquence d'acides aminés des peptides ou protéines analysés. Pour atteindre ces informations, de nouvelles configurations de spectromètres de masse ont été mises au point pour permettre la sélection d'ions parents, leur fragmentation et la mesure des rapports  $m/z$  de leurs fragments. Ces instruments sont de manière générale appelés spectromètres de masse en tandem.

### ▸ L'analyse MS/MS avec les trappes d'ions

L'analyseur trappe ionique, grâce à ses capacités de piégeage et d'éjection sélectifs des ions, offre la possibilité de réaliser des expériences de spectrométrie de masse multiple (Figure 10) : MS/MS, MS<sup>3</sup>, MS<sup>4</sup>, ..., MS<sup>n</sup>. Le processus d'isolation dans la trappe d'ions de  $m/z$  fixé peut être décomposé en 2 étapes :

- Les ions de  $m/z$  inférieurs au  $m/z$  fixé sont éjectés par un balayage croissant de l'amplitude de la tension appliquée sur l'électrode annulaire jusqu'à une amplitude proche de celle nécessaire à l'éjection de l'ion précurseur fixé.
- Les ions de  $m/z$  supérieurs sont mis en résonance et éjectés par un balayage en fréquence partant d'une fréquence proche de la fréquence de résonance des ions précurseurs à isoler.

La fragmentation est provoquée par l'ajustement de la fréquence de la tension alternative sur l'électrode chapeau de manière à ce qu'elle coïncide avec la fréquence propre de résonance des ions précurseurs : cette amplitude est réglée afin d'éviter l'éjection des ions de la trappe (l'amplitude de fragmentation est plus faible que l'amplitude d'éjection). Ainsi, dopés en énergie cinétique, les ions précurseurs entrent en collision avec les molécules d'hélium. A mesure que des collisions se produisent, leur énergie cinétique est convertie en énergie interne vibrationnelle et les ions finissent par se fragmenter.

L'énergie de fragmentation est exclusivement transmise aux ions précurseurs (à leur fréquence propre de résonance). Ainsi, les ions fils issus d'une fragmentation sont, théoriquement, incapables de se fragmenter à nouveau pour donner naissance à des ions petits-fils.

Un avantage supplémentaire de l'analyseur trappe d'ions est la possibilité de réaliser des expériences de MS<sup>n</sup>. En effet, le processus piégeage-isolation-fragmentation peut être répété plusieurs fois. Cela peut permettre l'élucidation structurale des composés analysés.

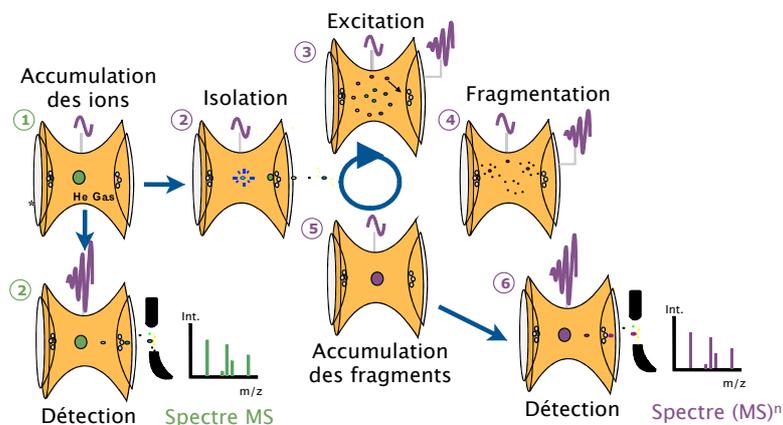


Figure 10 | Principe d'isolation, fragmentation et détection d'ions dans une trappe d'ions pour les expériences de MS et MS multiples.

▸ Les analyseurs hybrides de type Q-TOF

Les instruments dits hybrides sont des spectromètres de masse en tandem constitués de deux analyseurs séparés par une cellule de collision. Les instruments de type triple-quadripôle (Q-Q-Q), temps de vol-temps de vol (TOF-TOF) et quadripôle-temps de vol (Q-TOF) sont les principaux instruments hybrides rencontrés.

Sur un instrument de type Q-TOF, le quadripôle peut fonctionner soit comme guide d'ions, soit comme analyseur<sup>47</sup>.

▸ En mode « guide d'ion » (ou « *RF-only* »), les ions sont analysés par le TOF. Cependant, la transmission des ions dans le quadripôle reste dépendante de l'amplitude de la tension oscillante appliquée sur ses barres. La transmission de différentes gammes de masses peut donc être favorisée (fonction « *MS profile* »).

▸ En mode analyseur, un ion peut être sélectionné dans le quadripôle et transmis vers la cellule de collision où il est fragmenté. Les ions fils générés sont alors analysés dans analyseur TOF.

Dans cette configuration, le Q-TOF est particulièrement adapté à la réalisation d'analyses protéomiques en couplage nanoLC-MS/MS : durant ces analyses, des cycles MS suivis de une ou plusieurs MS/MS sur les peptides sélectionnés permettent d'obtenir des informations sur les peptides tryptiques élués de la colonne chromatographique.

A titre d'exemple, la figure 11 présente le **Synapt HDMS** (« *High Definition Mass Spectrometry* ») conçu par Waters Corporation et acquis par le laboratoire en 2008. Cet instrument de type ESI-Q-TOF intègre des innovations particulièrement intéressantes pour l'analyse protéomique par spectrométrie de masse.

▸ La source d'ionisation nanoESI présente une géométrie « en Z » qui évite le transfert des espèces non chargées dans l'interface. Elle est équipée d'un nébuliseur « *lock-mass* » diffusant une solution de référence et permettant l'acquisition régulière en cours d'analyse de spectres de référence pour la correction de la dérive de l'étalonnage en masse (cf. 1<sup>re</sup> partie des résultats, au chapitre I, point 2.2.).

▸ La technologie Triwave™, installée entre l'analyseur quadripolaire et l'analyseur à temps de vol, permet l'utilisation de l'instrument dans diverses configurations. Elle est constituée de deux cellules de lentilles « *trap* » et « *transfer* », utilisées comme guides d'ions ou comme cellules de collision. La cellule IMS intermédiaire est utilisée pour des expériences de mobilité ionique<sup>48</sup> au cours desquelles les espèces peuvent être séparées selon leur mobilité ionique au sein d'un gaz lorsqu'elles sont soumises à un champ électrique (c.-à-d. selon leur vitesse de déplacement qui dépend de leur charge et leur section efficace). Cette technologie est en cours d'évaluation au laboratoire<sup>49</sup>.

▸ L'analyseur TOF est capable de travailler en mode simple ou double réflecteur électrostatique (mode « *V* » ou « *W* »). Le passage du mode « *V* » au mode « *W* » permet de doubler la distance de vol et d'augmenter de ce fait la résolution (FWMH) de 10 000 à 17 500, mais au détriment de la sensibilité qui est alors environ divisée par 3.

▸ Enfin, couplé à un système de chromatographie ultra-haute performance (UHPLC), le Synapt HDMS permet pour des expériences de MS/MS d'utiliser le mode d'acquisition « *Expression* » (MS<sup>E</sup>, détaillé au chapitre II).

<sup>47</sup> Chernushevich, I. V., et al. (2001). *J Mass Spectrom* **36**, 849-65.

<sup>48</sup> Pringle, S. D., et al. (2007). *International Journal of Mass Spectrometry* **261**, 1-12.

<sup>49</sup> Atmanène, C. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2009).

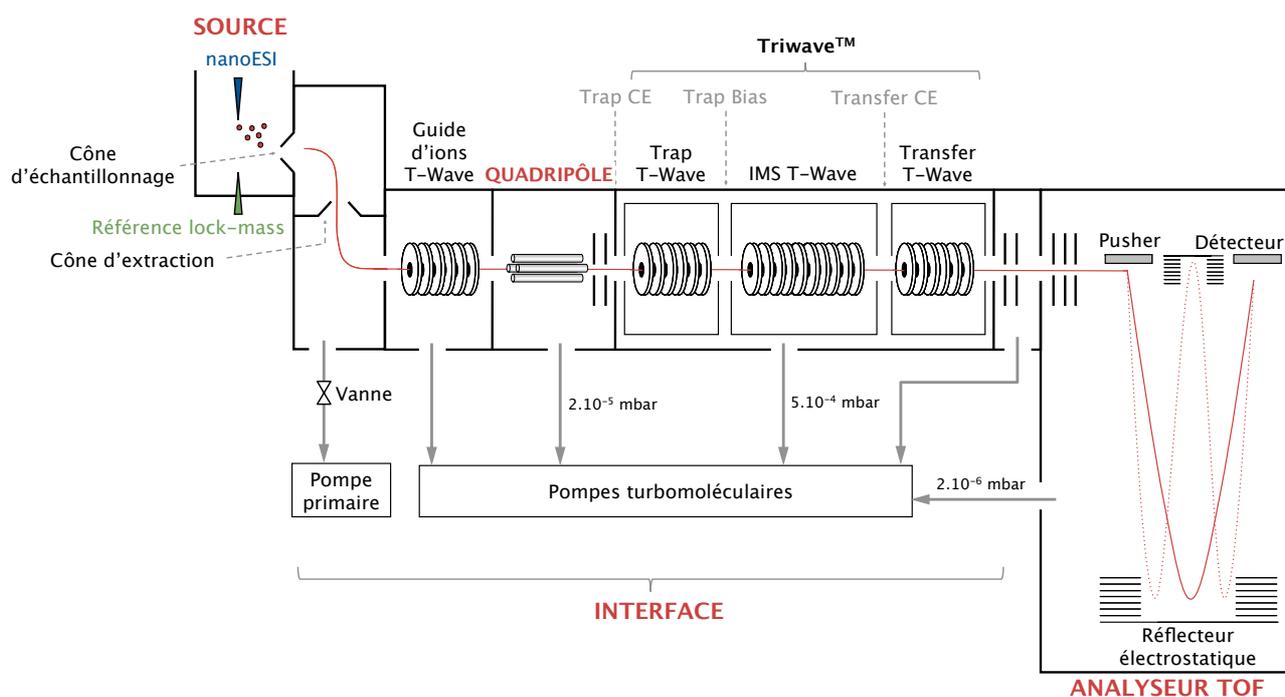


Figure 11 | Schéma général du Synapt HDMS (Waters Corp.) en mode de fonctionnement ESI-Q-TOF. (Adapté de Atmanène, C., 2009.)

#### ▸ Les analyseurs hybrides à ultra-haute résolution

Le développement d'instruments hybrides assemblant une trappe d'ions linéaire (LIT) à un FT-ICR ou à une Orbitrap a rendu possible l'utilisation en routine d'analyseurs à ultra-haute résolution pour l'analyse protéomique<sup>50</sup>. Le principal avantage de ce système est de permettre une mesure MS ultrahaute résolution dans l'analyseur haute résolution (FT-ICR ou Orbitrap) et d'acquérir simultanément des spectres de fragmentation dans la trappe, les ions à fragmenter ayant été choisis lors d'un pré-scan rapide. Ces données MS de haute précision en masse permettent de réaliser des identifications plus sûres<sup>51,52</sup> et peuvent être utiles dans des stratégies de quantification.

### 1.2.3. Performances caractéristiques des analyseurs

#### ▸ Les principaux types d'instruments commerciaux et leurs caractéristiques

Pour résumer, les caractéristiques techniques de différents instruments disponibles sur le marché aujourd'hui sont rassemblées en table 1. Le laboratoire a fait le choix de se doter de technologies complémentaires pour être en mesure de traiter une large gamme de questions analytiques en spectrométrie de masse (protéomique, caractérisation, quantification, analyse structurale de complexes supramoléculaires ...). Outre les instruments acquis ces dernières années et présentés dans la table 1, le laboratoire dispose d'un Ultraflex MALDI-TOF Bruker Dal (résolution : 25 000, précision de mesure : 50 ppm) utilisé dans les approches d'identification par empreinte peptidique massique (cf. chapitre II de cette partie), d'un Q-TOF II Waters Corp. (résolution : 10 000, précision de mesure : 5 ppm), d'un LCT Waters Corp. (résolution : 5000, précision de mesure : 5 ppm) utilisé pour l'analyse de protéines entières (interactions non covalentes)<sup>53</sup> et d'un triple-quadripôle 6400 Agilent Tech. (résolution 5000, précision de mesure 80 ppm) destiné aux expériences de quantification.

<sup>50</sup> Makarov, A., et al. (2006). *Anal Chem* **78**, 2113-20.

<sup>51</sup> Scherl, A., et al. (2008). *Proteomics* **8**, 2791-7.

<sup>52</sup> Scherl, A., et al. (2008). *J Am Soc Mass Spectrom* **19**, 891-901.

<sup>53</sup> Atmanène, C. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2009).

Instrument	HCT-Ultra Bruker Dal.	AmaZon Bruker Dal.	MicrOTOF-Q Bruker Dal.	Synapt HDMS Waters Corp.	LTQ-Orbitrap Thermo Sci.	MaXis Bruker Dal.	Solarix Bruker Dal.
Type	Trappe d'ion (IT)	Trappe d'ion (IT)	Quadripôle- Temps de vol (Q-TOF)	Quadripôle- Temps de vol (Q-TOF)	LTQ-Orbitrap	Quadripôle- Temps de vol (Q-TOF)	FT-ICR
Pouvoir résolutif	3 000 – 5 000	5 000 – 20 000	15 000	17 500	30 000 – 100 000	40 000 – 60 000	50 000 – 1 000 000
Précision de mesure	80 ppm	30 ppm	5 ppm	3 ppm	3 ppm	600 ppb	1 ppm
Gamme de masse	4 000 m/z	3 000 m/z	20 000 m/z	32 000 m/z	4 000 m/z	20 000 m/z	6 000 m/z
Sensibilité	++	+++	++	+++	++	+++	++
Possibilité MS/MS	✓	✓	✓	✓	✓	✓	✓
Vitesse d'acquisition	+++	+++	++	+++	++	+++	+
Coût	€	€€	€€	€€€	€€€	€€€	€€€€
Disponibilité au laboratoire	2006	Prév. 2010	2008	2008		2009	

**Table 1** | Principales caractéristiques techniques et performances moyennes des instruments couramment disponibles sur le marché en 2009. (Sources : Bruker Dal., Waters Corp., Thermo Sci.)

#### ▸ Le pouvoir résolutif et l'importance de la résolution

Deux caractéristiques instrumentales particulièrement importantes pour la mesure des ions par spectrométrie de masse sont le pouvoir résolutif et la précision de mesure en masse. Une bonne résolution spectrale est utile pour déterminer l'état de charge des ions, avoir accès à la distribution isotopique et discriminer des ions isobares. Elle est aussi liée à la précision de mesure de masse, le centroïde d'un pic étant d'autant mieux déterminé que le pic est fin.

Particulièrement, la mesure de masse de protéines en spectrométrie de masse (en réalité, la mesure de rapports  $m/z$ ) suit par nature des règles quantifiées<sup>54</sup> : la charge est induite par un nombre entier de charges élémentaires (par exemple le proton  $H^+$ ), les protéines sont composées d'un assortiment d'acides aminés, de groupes fonctionnels, d'éléments et d'isotopes. Par conséquent, il existe plusieurs niveaux discontinus de résolution spectrale (Figure 12) qui peuvent être représentés sous forme de « marches » (Figure 13). L'accès à chacune de ces marches représente un défi analytique. En augmentant le pouvoir résolutif instrumental, de toutes nouvelles familles de pics deviennent accessibles à mesure que de nouveaux plateaux sont atteints.

Sur la figure 13, l'axe des ordonnées définit la hauteur de chaque marche, c'est-à-dire l'ordre de grandeur de la distance en Dalton séparant deux pics consécutifs d'un spectre. L'axe des abscisses  $1/\Delta m_{50\%}$  (où  $\Delta m_{50\%}$  correspond à la largeur d'un pic à mi-hauteur) équivaut au pouvoir résolutif de l'instrument ( $m/\Delta m_{50\%}$ ) rapporté à la masse  $m$  de l'espèce mesurée. On obtient ainsi une représentation du pouvoir résolutif indépendante de la masse.

<sup>54</sup> Marshall, A. G., et al. (2002). Anal Chem 74, 252A–259A.

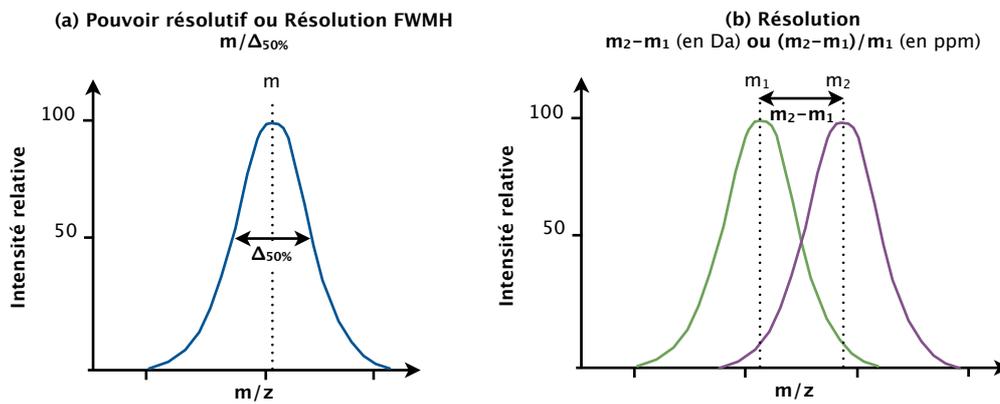


Figure 12 | Pouvoir résolutif (ou résolution FWHM : « full width at half-maximum ») (a) et résolution (b) selon Marshall, A. G. (2002).

L'information accessible augmente avec de meilleurs pouvoirs résolutifs instrumentaux et/ou lorsque des espèces de masse moléculaire plus petite sont analysées :

- ▶ Comme il l'a été décrit dans le paragraphe précédent, le premier plateau de résolution concerne la discrimination des différents états de charge de pics moléculaires d'ions de même masse. La masse moyenne d'une protéine peut en être déduite.

- ▶ Sur un deuxième plateau de résolution, des ions de masse proche à un même état de charge peuvent être mesurés. Il peut par exemple s'agir d'adduits (sodium ...).

- ▶ Le troisième palier correspond à la résolution de la distribution statistique des différents isotopes naturels constituant l'espèce mesurée. Dans ce cas, l'état de charge d'un pic moléculaire peut être évalué en calculant la distance séparant deux pics isotopiques successifs. Également, la masse exacte monoisotopique (déterminée à partir des valeurs exactes des masses des isotopes principaux) peut être mesurée directement.

- ▶ Enfin, la composition isotopique ou atomique d'espèces de même masse peut être évaluée lorsque la résolution atteint le millidalton (pour les petites molécules jusqu'à ~300 Da).

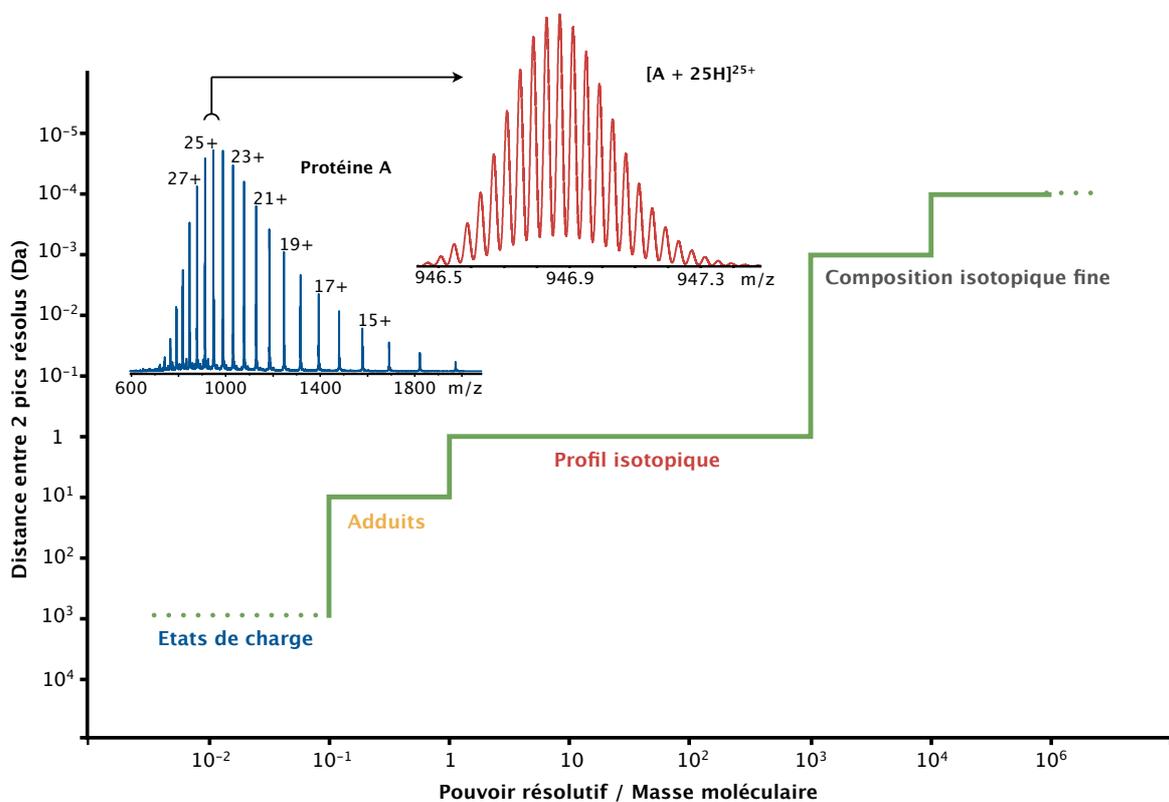


Figure 13 | Résolution spectrale (distance en Da entre 2 pics à résoudre) représentée en fonction du pouvoir résolutif de l'instrument rapporté à la masse moléculaire de l'espèce mesurée. En insert, spectres de masse modélisés d'une protéine (Prx C171A + IAM) de masse moyenne 23 646,59 Da et de masse monoisotopique 23 632,25 Da. (Adapté de Marshall, A. G., 2002.)

Les performances résolutive moyennes des principaux instruments disponibles sur le marché peuvent être trouvées en table 1. Ces résolutions annoncées comme spécification sont généralement calculées à partir d'espèces 2 ou 3 fois chargées autour de 1000 Da. Les instruments de résolution standard (trappes d'ion, quadruples...) sont en mesure de résoudre les différents états de charge sur les spectres ESI-MS de protéines et permettent l'accès à la distribution isotopique sur ceux de petits peptides. Les instruments de moyenne résolution (Q-TOF, TOF) sont en mesure de séparer les isotopes constituant chaque pic moléculaire multichargé de spectres ESI-MS de peptides (jusqu'à ~ 5000 Da, 5 ou 6 fois chargés). Les instruments à haute résolution (comme les FT-ICR ou l'Orbitrap™) permettent d'atteindre une résolution isotopique pour des protéines entières.

### 1.3. La fragmentation peptidique

L'intérêt majeur de la spectrométrie de masse en tandem est de permettre la détermination de tout ou partie de la séquence en acides aminés de peptides. Ces possibilités offrent à l'analyse MS/MS un pouvoir discriminant et une spécificité d'identification considérables en analyse protéomique. Cependant, l'identification des séquences peptidiques à partir de spectres MS/MS est souvent complexe et demande une connaissance des modes et des mécanismes de fragmentation des peptides en phase gazeuse.

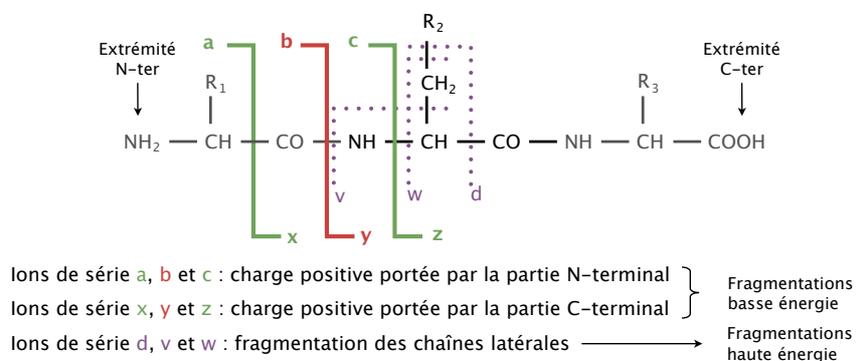
#### 1.3.1. La fragmentation induite par collision (CID)

Le transfert des ions en phase gazeuse par ESI n'est pas un processus hautement énergétique : les ions qui entrent dans le spectromètre de masse ont donc des énergies internes faibles<sup>55</sup>. La fragmentation des ions est généralement réalisée en favorisant leur collision avec des molécules de gaz non chargées (Collision Induced Dissociation, CID ou Collision Activated Dissociation, CAD). On peut distinguer deux types de fragmentations par CID : les fragmentations basse énergie (quelques eV, sur les instruments de type Q-TOF et trappe d'ions) et les fragmentations haute énergie (quelques keV, généralement sur les analyseurs TOF-TOF).

##### ▶ La nomenclature des fragmentations peptidiques

La nomenclature de fragmentation des peptides établie par Biemann dans les années 1990<sup>56</sup> fait toujours référence aujourd'hui (Figure 14). Deux types d'ions sont observés :

- ▶ ceux pour lesquels la charge positive est portée par la partie N-terminale : les séries a, b et c ;
- ▶ ceux pour lesquels la charge positive est portée par la partie C-terminale : les séries x, y et z.



**Figure 14** | Nomenclature des fragmentations peptidiques selon Biemann. R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub> représentent les chaînes latérales de 3 acides aminés.

<sup>55</sup> Gaskell, S. J. (1997). *Journal of Mass Spectrometry* **32**, 677-688.

<sup>56</sup> Biemann, K. (1990). *Methods Enzymol* **193**, 886-7.

---

Les séries b et y, observables en mode de fragmentation à haute ou basse énergie, sont généralement majoritaires. Elles représentent la rupture des liaisons peptidiques, plus labiles que les autres. Les différences de masse entre des ions consécutifs issus d'une même série permettent de déterminer l'identité des acides aminés successifs et donc de reconstruire la séquence du peptide fragmenté. Les acides aminés isobares comme la Leucine et l'Isoleucine ou la Glutamine et la Lysine ont une différence de masse de 0,036 Da et ne peuvent pas être différenciés avec les analyseurs à résolution classique.

▸ Modèle de fragmentation du « proton mobile »

Le transfert d'énergie par activation collisionnelle implique une délocalisation du proton à différents endroits du squelette peptidique. La présence de ce proton initie une fragmentation au niveau des liaisons peptidiques pour former des ions des séries b et y : on dit que ces fragmentations basses énergies sont dirigées par la charge (« *charge-directed fragmentation* ») et suivent le modèle du « proton mobile » décrit en 1996<sup>57</sup>. L'énergie requise pour initier la fragmentation d'un peptide dépend de sa composition en acides aminés (présence ou absence d'acides aminés basiques), de sa séquence et de sa taille.

Pour des peptides monochargés, la présence d'acides aminés basiques (Arginine, Lysine) augmente considérablement l'énergie nécessaire à leur fragmentation : le proton est « séquestré » sur la chaîne latérale de l'acide aminé basique et l'énergie requise pour le délocaliser sur le squelette peptidique est plus élevée.

Pour des peptides tryptiques doublement chargés (espèces majoritaires à 65–80% dans le cas de peptides tryptiques<sup>58</sup>), les deux charges sont généralement portées par l'extrémité N-terminale et par la chaîne latérale de la lysine ou de l'arginine C-terminale : la barrière énergétique pour transférer le proton N-terminal sur le squelette peptidique est plus faible (par rapport au proton séquestré par l'acide aminé basique en C-terminal) ce qui favorise la génération d'ions b et y. Les ions y sont souvent plus intenses : le proton séquestré par le résidu C-terminal augmentant leur stabilité par rapport aux ions b.

Dans les approches mettant en oeuvre des digestions aspécifiques (par la pepsine par exemple), les peptides – non tryptiques – subissent des fragmentations moins régulières du fait de la non-présence systématique d'acides aminés N- et C-terminaux basiques. Les spectres de fragmentation sont dans ce cas bien plus difficiles à interpréter.

De nombreuses équipes travaillent à l'établissement de règles de fragmentations très précises qui pourront ensuite être implémentées dans les algorithmes de recherche en banques de données avec des données MS/MS. L'établissement de telles règles est un outil précieux pour l'amélioration de l'efficacité et de la spécificité de ces algorithmes<sup>59,60</sup>.

### 1.3.2. La dissociation par transfert d'électron (ETD)

En 2004, une nouvelle méthode de fragmentation des peptides fondée sur la dissociation par transfert d'électrons (« *Electron Transfer Dissociation* », ETD) a été développée<sup>61</sup>. Le transfert d'électrons est assuré par l'intermédiaire d'un réactif : un radical anion (le fluoranthène par exemple) qui joue le rôle de transporteur vers le peptide protoné.

La fragmentation engendrée n'ayant pas de coupure préférentielle, le type d'ions fragments générés est aléatoire. Les spectres de fragmentations peuvent être plus riches et plus informatifs que les spectres obtenus par CID avec notamment la présence des séries d'ions c et z. Par ailleurs, des modifications posttraductionnelles (phosphorylation, glycosylation ...) peuvent être observées puisqu'elles sont potentiellement conservées et la fragmentation de peptides de taille élevée est plus efficace qu'en CID.

---

<sup>57</sup> Dongre, A. R., et al. (1996). *J Mass Spectrom* **31**, 339–50.

<sup>58</sup> Savitski, M. M., et al. (2005). *J Proteome Res* **4**, 2348–54.

<sup>59</sup> Paizs, B., et al. (2005). *Mass Spectrom Rev* **24**, 508–48.

<sup>60</sup> Barton, S. J., et al. (2009). *Mass Spectrom Rev* **28**, 177–87.

<sup>61</sup> Syka, J. E., et al. (2004). *Proc Natl Acad Sci U S A* **101**, 9528–33.

Notamment, les spectres de fragmentation ETD de dipeptides à ponts disulfures sont généralement plus informatifs et la liaison cystéine–cystéine peut être rompue.

Néanmoins, l'interprétation de ces spectres de fragmentations très riches peut s'avérer difficile puisque le grand nombre d'ions fragments entraîne une perte de sensibilité importante. Par ailleurs, la formation d'ions fragments multichargés doit être considérée. Pour déterminer les états de charges des ions fragments multichargés, une bonne résolution isotopique est alors indispensable, souvent au détriment de la sensibilité.

---

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

La spectrométrie de masse ne peut, à elle seule, permettre la caractérisation directe à grande échelle de protéines. La raison principale est que les échantillons biologiques contiennent généralement de plusieurs centaines à plusieurs dizaines de milliers de protéines. Ces protéines peuvent être présentes sous de multiples formes modifiées posttraductionnellement et leurs concentrations peuvent s'étendre sur de larges gammes dynamiques, souvent au-delà de celles qui peuvent être appréhendées en spectrométrie de masse.

Une simplification des échantillons avant leur analyse par spectrométrie de masse est alors requise pour que leur complexité soit compatible avec les performances des instruments utilisés. La séparation peut être réalisée au niveau des protéines d'un échantillon biologique, au niveau des peptides (générés lors d'une digestion enzymatique d'un échantillon protéique), ou à l'un et l'autre de ces deux niveaux. Parmi les diverses techniques de séparation de protéines et peptides connues, l'électrophorèse sur gel et la chromatographie liquide sont pratiquement les deux seules utilisées en analyse protéomique par spectrométrie de masse.

### 2.1. La séparation de protéines : l'électrophorèse sur gel et ses alternatives

#### 2.1.1. L'électrophorèse sur gel mono- et bidimensionnelle

Par définition, l'électrophorèse permet la séparation d'analytes sous forme d'ions en les soumettant à un champ électrique qui leur impose de migrer différemment dans un milieu adapté. Malgré son efficacité énergétique, l'électrophorèse capillaire pour la séparation de mélanges protéiques complexes est restée peu utilisée, du fait de sa faible capacité de charge et des effets d'hystérèse inhérents à l'adsorption des protéines sur le capillaire. La séparation de protéines par électrophorèse est donc généralement réalisée sur gel, le plus souvent de polyacrylamide. Les gels offrent l'avantage majeur de permettre une séparation continue, résolutive et compatible avec les traitements d'échantillons qui peuvent être nécessaires avant l'analyse (réduction et alkylation, digestion « *in gel* » par exemple).

##### ▸ Electrophorèse sur gel monodimensionnelle (Gel 1-D ou SDS-PAGE)

L'électrophorèse 1-D de protéines en conditions dénaturantes utilise généralement le Sodium Dodecyl Sulfate (SDS) comme détergent. La technique telle qu'elle est connue aujourd'hui est apparue dans les années 1970 avec les expériences de Laemmli<sup>62</sup>. Le SDS se fixe extrêmement bien aux résidus hydrophobes des protéines et favorise leur dénaturation. Il confère aux protéines une charge nette négative qui masque complètement leur charge native. La densité de charge étant sensiblement constante pour toutes les protéines d'un échantillon (1 molécule de SDS pour 2 acides aminés), la séparation dans un champ électrique est essentiellement conduite par la masse moléculaire des protéines (Figure 15). De plus, la charge négative nette conférée aux protéines implique de multiples répulsions électrostatiques qui favorisent leur solubilisation. De fortes concentrations peuvent alors être atteintes dans un gel de préconcentration où les protéines sont retenues avant la séparation de manière à augmenter sensiblement son pouvoir résolusif.

La taille des protéines étant très variée, il est souvent intéressant d'optimiser la résolution en fonction des besoins expérimentaux<sup>63</sup>. Cela peut être réalisé :

▸ en modifiant la taille des pores du gel (lié à la concentration en acrylamide), la résistance des protéines à la migration étant accentuée dans des gels à hauts pourcentages en acrylamide. Un bon

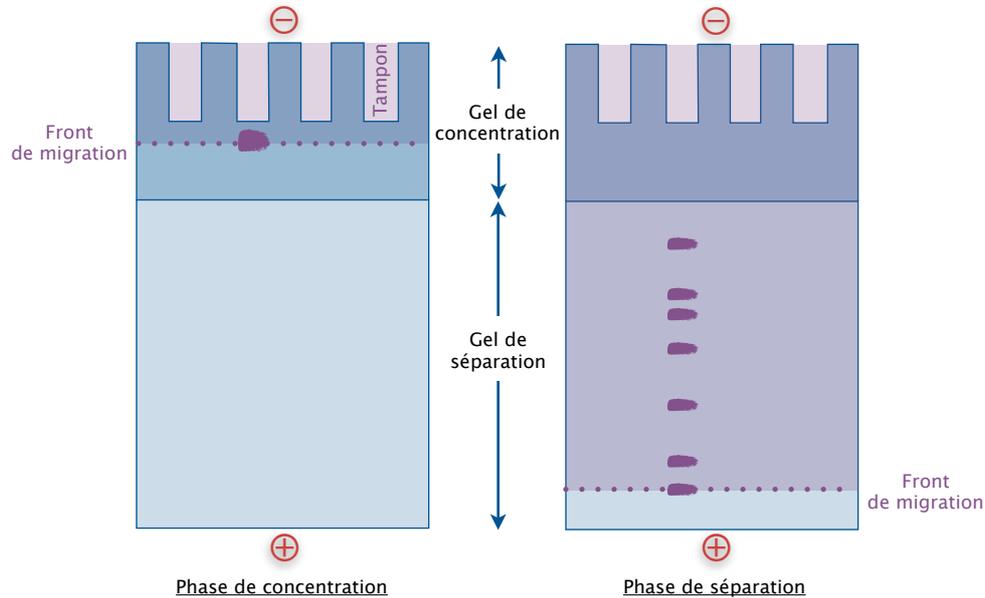
---

<sup>62</sup> Laemmli, U. K. (1970). *Nature* 227, 680-5.

<sup>63</sup> Rabilloud, T., et al. (2009). *Mass Spectrom Rev* 28, 816-43.

compromis est de réaliser des gels présentant un gradient de porosité vertical afin d'optimiser la résolution de séparation des protéines de toutes tailles.

▸ en modifiant les tampons d'électrophorèse utilisés en termes de nature des ions et de pH. Pour des systèmes anioniques (tris-chloride-glycine par exemple), la vitesse de migration des ions augmente avec le pH. Mais une vitesse de migration rapide implique une séparation moins résolutive des protéines de haute masse moléculaire.



**Figure 15** | Principe de l'électrophorèse SDS. Un gel discontinu est fait de 3 composants : un tampon d'électrode (Tris-Glycine), un gel de concentration à pores larges contenant un tampon (Tris-HCl) à pH 6,8, et un gel de séparation à pores resserrés contenant un tampon (Tris-HCl) à pH 8,8. L'échantillon est chargé dans un tampon (Tris-HCl) à pH 6,8. Lorsqu'un champ électrique est appliqué, l'interaction entre le champ électrique, le tampon d'électrode et le tampon de concentration crée un front de migration où les protéines sont concentrées. Lorsque le front atteint le gel de séparation, la vitesse de migration augmente sous l'effet du changement de pH. Les protéines sont alors séparées en bandes étroites, et les ions capables de parcourir le gel à la vitesse du front de migration restent dans la bande concentrée.

L'électrophorèse SDS présente d'excellentes capacités de solubilisation des protéines et peut être adaptée à de multiples situations. Elle est particulièrement utilisée en analyse protéomique fonctionnelle, après que des interactants d'une protéine appât ont été purifiés (cf. chapitre 2 de cette partie, point 2.2.1 et cf. Résultats 1<sup>re</sup> partie, chapitre 3). Toutefois, pour beaucoup d'échantillons plus complexes utilisés en analyse protéomique, les gels 1-D ne sont pas suffisamment résolutifs. Généralement, pour des raisons pratiques, l'électrophorèse SDS est la dernière étape de séparation de protéines entières. Pour ces mêmes raisons, il est préférable de réaliser une 1<sup>re</sup> dimension de séparation directement sur gel : dans ces conditions, la stratégie d'isoélectrofocalisation s'est révélée la plus populaire.

#### ▸ Electrophorèse sur gel bidimensionnelle (Gels 2-D)

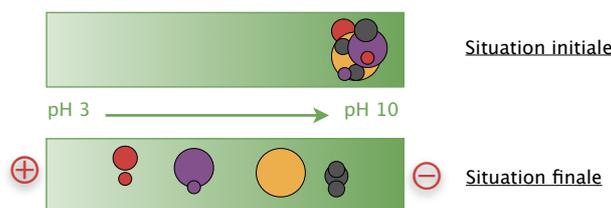
Le couplage de l'isoélectrofocalisation dénaturante à l'électrophorèse SDS est dit orthogonal, car les deux méthodes de séparation utilisent des paramètres complètement indépendants : le point isoélectrique et la masse moléculaire des protéines. Dès 1970, O'Farrel décrit la séparation de 1000 spots sur un gel 2-D. Les capacités résolutives de la technique ont contribué à son utilisation généralisée : jusqu'à 5000 protéines peuvent être séparées et, dans un spot, détectées jusqu'à 1 ng<sup>64</sup>.

L'isoélectrofocalisation est mise en oeuvre dans une bandelette de gel présentant un gradient de pH et traversée par un champ électrique allant de sa borne acide à sa borne basique (Figure 16)<sup>65</sup>. Les protéines chargées sur le gel sont alors séparées en fonction de leur point isoélectrique et s'immobilisent au niveau du gradient où le pH est égal à leur pI, là où par définition leur charge globale est nulle. Les phénomènes

<sup>64</sup> Gorg, A., et al. (2004). *Proteomics* 4, 3665-85.

<sup>65</sup> Righetti, P. G., et al. (1987). *Methods Biochem Anal* 32, 215-78.

opposés de diffusion et de focalisation électrique tendent à concentrer les analytes à leurs pI respectifs. La bandelette est alors saturée en SDS puis chargée en haut d'un gel 1-D et une séparation SDS-PAGE est réalisée comme décrit précédemment.



**Figure 16** | Principe de l'isoélectrofocalisation (IEF). Dans la situation initiale, une bandelette de gel de polyacrylamide présentant un gradient de pH est chargée du mélange protéique à séparer. Lorsque les électrodes sont connectées, les protéines migrent jusqu'à la zone de pH égale à leur pI où elles ont une charge globale nulle. A  $pH < pI$ , une protéine est chargée positivement et migre vers la cathode ; à  $pH > pI$ , une protéine est chargée négativement et migre vers l'anode.

Parmi ses qualités majeures, outre la très bonne résolution et les indications quantitatives accessibles, les gels 2-D permettent la séparation et la visualisation d'un grand nombre de variants posttraductionnels (formes phosphorylées, glycosylées ...)<sup>66</sup>. Toutefois, la technique présente aussi un certain nombre de limites. D'abord, une découpe systématique du gel – souvent réalisée en séparation SDS-PAGE – est impossible de par la taille des gels 2-D : seuls les spots observés sont donc sélectionnés et les protéines les moins abondantes échappent à l'analyse. Une deuxième limite réside dans la sous-représentation de certaines catégories de protéines dans les gels 2-D. C'est notamment le cas des protéines membranaires, hydrophobes<sup>67</sup> ou fortement basiques.

#### ▸ Coloration des gels

La sensibilité de détection des gels dépend essentiellement des colorants utilisés. Trois types de colorations sont couramment utilisées en analyse protéomique :

▸ La coloration au bleu de Coomassie<sup>68</sup> : malgré la faible sensibilité de cette coloration (100 ng), elle présente le grand avantage d'être linéaire et réversible et donc très adaptée pour les études protéomiques comparatives.

▸ La coloration à l'argent : elle présente une limite de détection 100 fois plus faible que la coloration au bleu de Coomassie<sup>69</sup>. Cependant, pour les fortes concentrations, la coloration n'est pas linéaire. Des incompatibilités avec l'analyse par spectrométrie de masse ont également été décrites<sup>70</sup>, mais des protocoles améliorés ont pu être mis au point<sup>71</sup>.

▸ Les colorations fluorescentes<sup>72</sup> : elles peuvent présenter des sensibilités jusqu'à l'attomole, mais requièrent des outils optiques perfectionnés.

### 2.1.2. L'électrophorèse Off-Gel

En 2002, un nouveau type de fractionnement est apparu : « l'isoélectrofocalisation Off-Gel »<sup>73</sup>. Dans cette approche, les protéines sont séparées en fonction de leur point isoélectrique directement en solution, sans tampons ni ampholytes. Les extraits protéiques sont déposés dans des puits installés sur une bande de

<sup>66</sup> Rabilloud, T., et al. (2009). *Mass Spectrom Rev* **28**, 816–43.

<sup>67</sup> Santoni, V., et al. (2000). *Electrophoresis* **21**, 1054–70.

<sup>68</sup> Neuhoff, V., et al. (1988). *Electrophoresis* **9**, 255–62.

<sup>69</sup> Shevchenko, A., et al. (1996). *Analytical Chemistry* **68**, 850–8.

<sup>70</sup> Gevaert, K., et al. (2000). *Electrophoresis* **21**, 1145–54.

<sup>71</sup> Richert, S., et al. (2004). *Proteomics* **4**, 909–16.

<sup>72</sup> Rabilloud, T., et al. (2001). *Proteomics* **1**, 699–704.

<sup>73</sup> Ros, A., et al. (2002). *Proteomics* **2**, 151–156.

gel à gradient de pH immobilisé (IPG). Sous l'effet d'un champ électrique, les protéines migrent d'un puits à un autre via la bande de gel et restent bloquées dans un puits lorsqu'elles atteignent leur  $pI$ <sup>74</sup>.

Ce type de fractionnement présente l'avantage de pouvoir séparer les protéines en fonction de leur  $pI$  au dixième d'unité de pH près, tout en récupérant les échantillons en solution. Ainsi, il est possible en aval de refractionner les extraits par d'autres techniques (gel 1-D par exemple).

### 2.1.3. Séparation de protéines par chromatographie liquide

Une approche de séparation de mélanges simples de protéines entières par chromatographie liquide peut être intéressante dans le cas d'une mesure directe par spectrométrie de masse<sup>75</sup>. L'analyse LC-MS peut être assurée par un couplage via une source ESI ou peut être découplée en collectant des fractions chromatographiques à analyser. La séparation a généralement lieu sur des colonnes analytiques en phase inverse  $C_4$  ou  $C_8$ . Cette stratégie est traitée dans la 2<sup>e</sup> partie des résultats (au chapitre III, point 2.1.2.) dans le cadre de l'analyse de complexes protéiques covalents.

## 2.2. La séparation de peptides par chromatographie liquide

Face à la complexité des mélanges protéiques étudiés et aux difficultés d'identifications de protéines entières par spectrométrie de masse MS ou MS/MS, les analyses ne sont pas réalisées au niveau des protéines entières, mais sur des digests peptidiques de ces protéines, produits lors de réactions de protéolyse contrôlées. Cela a pour conséquence d'augmenter directement la complexité des échantillons, mais de faciliter leur analyse par spectrométrie de masse. Si des techniques de séparation de mélanges peptidiques sur gel ont pu être développés<sup>76</sup>, la chromatographie liquide haute performance (HPLC) s'impose depuis le milieu des années 1980 comme la technique de choix pour la séparation de mélanges peptidiques, grâce à ses capacités résolutes, sa flexibilité et ses possibilités de couplage direct aux spectromètres de masse via une source ESI<sup>77</sup>. Différentes sélectivités peuvent être apportées par les nombreux modes de séparation en HPLC comme la chromatographie d'échange d'ions (IEC), la chromatographie d'exclusion stérique (SEC) et la chromatographie d'interaction hydrophile (HILIC) ; mais c'est la chromatographie en phase inverse (RPLC) qui s'est installée comme technique analytique principale pour les peptides.

### 2.2.1. Les systèmes nanoHPLC et nanoUPLC

Les techniques de séparation par chromatographie liquide ont fait l'objet de nombreuses adaptations pour rendre robuste et sensible le couplage LC-MS via la source ESI (compatible avec l'introduction d'échantillons en solution)<sup>78</sup>. La sensibilité de la mesure en utilisant une source ESI étant dépendante de la concentration des échantillons et le matériel biologique à analyser étant souvent peu concentré, des systèmes de chromatographie liquide ont du être développés pour descendre à des débits de l'ordre de 100 à 300 nL/min (cf. point 1.1.1 de ce chapitre). A ce jour la technique la plus utilisée est la nano-chromatographie liquide haute performance (nanoHPLC) en phase inverse (RP), couplée directement à la spectrométrie de masse via une source nanoESI.

Les échantillons étant souvent dilués au moment de l'injection, ce type de système se compose d'une colonne de chargement et d'une colonne analytique. La colonne de chargement permet de concentrer et de dessaler les échantillons avant leur séparation sur la colonne analytique par un gradient de solvants eau/acétonitrile ou eau/méthanol. C'est en général, le choix de la phase utilisée (taille des particules, taille des

<sup>74</sup> Michel, P. E., et al. (2003). *Electrophoresis* **24**, 3-11.

<sup>75</sup> Aguilar, M. I., et al. (1996). *Methods Enzymol* **270**, 3-26.

<sup>76</sup> Rabilloud, T., et al. (2009). *Mass Spectrom Rev* **28**, 816-43.

<sup>77</sup> Aguilar, M.-I. (2004). *HPLC of Peptides and Proteins: Methods and Protocols*, Humana Press.

<sup>78</sup> Sandra, K., et al. (2008). *J Chromatogr B Analyt Technol Biomed Life Sci* **866**, 48-63.

pores, type de phase stationnaire) qui aura le plus d'impact sur la sélectivité de séparation des protéines et des peptides. De manière générale pour un système nanoHPLC-MS/MS, les colonnes utilisées pour séparer les peptides sont remplies de billes de silice sphériques – dont la taille varie de 5 à 3 µm avec une porosité de 300 Å – modifiées à leur surface par des chaînes alkyles de type C18<sup>79</sup>.

A ce jour, les défis résident dans le développement de nouveaux systèmes de chromatographie liquide permettant de gagner en sensibilité, en résolution, mais également en temps d'analyse. Les industriels du secteur de l'instrumentation analytique développent et commercialisent des solutions toujours plus innovantes :

▸ En 2005, la chromatographie liquide sur puce microfluidique à fait son apparition sur le marché avec le système HPLC-Chip™ proposé par Agilent Technologies<sup>80, 81</sup>. Il devrait bientôt être rejoint par une solution similaire mise au point par Waters Corporation (TRIZAIC™ UPLC System with nanoTile technology).

▸ Depuis 2004, des colonnes chromatographiques avec des tailles de particules toujours plus petites sont proposées<sup>82</sup>. Ce type de colonne permet de travailler à des débits très élevés (pour augmenter la résolution et la sensibilité de la séparation), mais impose une augmentation de pression sur les systèmes. Pour pallier ce problème, des systèmes de chromatographie permettant de travailler à des pressions très élevées sont développés, notamment par Waters Corporation<sup>83</sup> qui propose le système de chromatographie à ultra haute performance UPLC™.

L'utilisation et l'optimisation de ces deux systèmes sont détaillées dans la 1<sup>re</sup> partie des résultats, au chapitre 1, point 2.1.

## 2.2.2. Les techniques de chromatographie liquide multidimensionnelle

D'une manière générale, il est nécessaire de combiner une étape de séparation des protéines et, après que les fractions collectées ont été soumises à digestion enzymatique, une étape de séparation des peptides. Les limites de ces méthodes ont encouragé le développement de stratégies ne passant pas par la séparation des protéines sur gel. C'est ainsi qu'une approche par chromatographie liquide multidimensionnelle a été introduite en 1999<sup>84</sup>.

Cette approche nommée MudPIT pour « *Multi-Dimensional Protein Identification Technology* »<sup>85</sup> a ouvert la voie à ce qu'on appelle la protéomique « *shotgun* ». Les approches de protéomique « *shotgun* » désignent l'analyse directe de mélanges complexes de protéines pour générer rapidement un profil global des protéines présentes dans le mélange<sup>86, 87</sup>. Le mélange complexe de protéines est directement digéré en peptides qui sont séparés lors d'étapes de chromatographies successives, avant d'être analysés par spectrométrie de masse en tandem.

L'approche la plus utilisée consiste à coupler une première étape de séparation des peptides par une chromatographie d'échange de cations (SCX, « *Strong Cation Exchange* ») suivie d'une étape de chromatographie de phase inverse classique. Trois approches ont été décrites dans la littérature pour la réalisation des analyses SCX/RP-MS/MS<sup>88, 89</sup> :

▸ La stratégie MudPIT (Figure 17a) : Les peptides sont chargés sur une colonne microcapillaire biphasique remplie en série par une résine SCX et par une résine de phase inverse. La colonne chargée est intégrée dans le système de chromatographie et les peptides sont élués de la résine SCX par paliers avec un

<sup>79</sup> Neverova, I., et al. (2005). *J Chromatogr B Analyt Technol Biomed Life Sci* **815**, 51–63.

<sup>80</sup> Yin, H., et al. (2005). *Analytical Chemistry* **77**, 527–33.

<sup>81</sup> Yin, H., et al. (2007). *J Sep Sci* **30**, 1427–34.

<sup>82</sup> Liu, H., et al. (2007). *J Chromatogr A* **1147**, 30–6.

<sup>83</sup> Kay, R. G., et al. (2007). *Rapid Commun Mass Spectrom* **21**, 2585–93.

<sup>84</sup> Link, A. J., et al. (1999). *Nat Biotechnol* **17**, 676–82.

<sup>85</sup> Washburn, M. P., et al. (2001). *Nat Biotechnol* **19**, 242–7.

<sup>86</sup> Wolters, D. A., et al. (2001). *Analytical Chemistry* **73**, 5683–90.

<sup>87</sup> Wu, C. C., et al. (2002). *Curr Opin Mol Ther* **4**, 242–50.

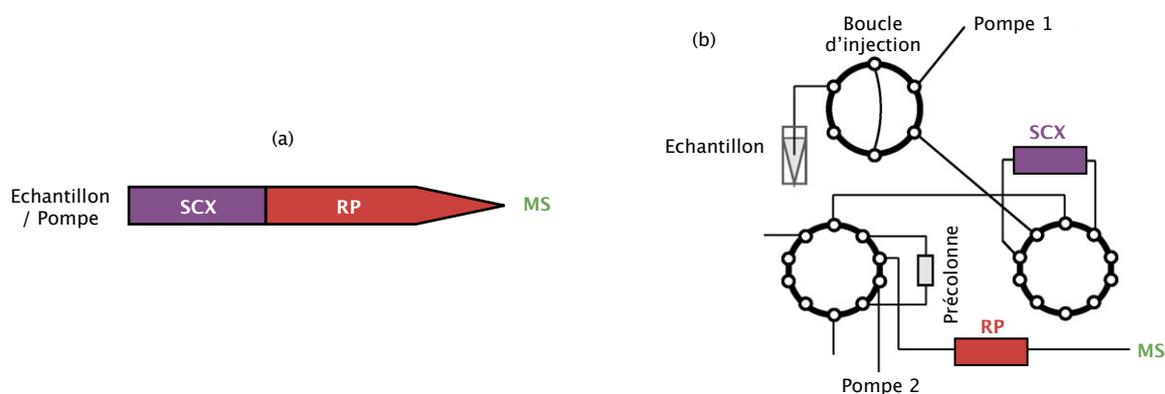
<sup>88</sup> Fournier, M. L., et al. (2007). *Chem Rev* **107**, 3654–86.

<sup>89</sup> Sandra, K., et al. (2009). *J Chromatogr B Analyt Technol Biomed Life Sci* **877**, 1019–39.

gradient en sels et se déposent en tête de la phase RP. A chaque palier, les peptides élués sont séparés dans une deuxième étape en fonction de leur hydrophobicité sur la résine de phase inverse et sont analysés en ligne par spectrométrie de masse en tandem<sup>90</sup>.

▸ La stratégie « *on-line* » (Figure 17b) : Dans cette configuration les deux colonnes SCX et RP sont couplées via un système de valves et une colonne d'enrichissement (précolonne). Les peptides sont élués de la colonne SCX sur cette précolonne d'enrichissement permettant d'éliminer les éluants utilisés en SCX et incompatibles avec la MS (chlorure de sodium et chlorure de potassium). Les peptides sont ensuite élués de la colonne d'enrichissement sur la colonne RP couplée au spectromètre de masse<sup>91</sup>. L'architecture de la valve dans cette configuration est très complexe et rend la mise en oeuvre délicate.

▸ La stratégie « *off-line* » : L'étape de SCX est réalisée de façon indépendante, les fractions sont collectées avant d'être analysées individuellement en nanoLC-MS/MS avec séparation sur colonne de phase inverse classique. N'étant pas directement couplée à la colonne RP, cette approche est très flexible vis-à-vis des solvants utilisés pour la SCX.



**Figure 17** | Exemples de stratégies décrites pour les expériences SCX/RP-MS/MS en configuration MudPIT (a) et en configuration on-line (b). (D'après Sandra, 2009)

Le nombre de protéines pouvant être identifiées par ces approches MudPIT peut largement dépasser le nombre de protéines identifiées par des approches gel 2-D : 1484 protéines uniques ont pu être identifiées par rapport à 300 protéines identifiées après gel 2-D dans un lysat cellulaire total de *Saccharomyces cerevisiae*<sup>92</sup>. Cependant, dans les approches « *shotgun* », les protéines étant digérées en premier lieu, la séparation n'est réalisée qu'au niveau peptidique. La relation entre les peptides et la protéine dont ils sont issus est donc perdue et les identifications sont souvent difficiles et ambiguës. De nombreuses identifications ne sont notamment réalisées qu'avec un unique peptide. Des critères de validation précis doivent alors être choisis pour éviter un taux de faux positifs trop élevé. De plus, la gestion du grand volume de données générées durant ces analyses présente des difficultés (classiquement 10 000 à 100 000 spectres MS/MS doivent être traités)<sup>93</sup>.

## 2.3. Conclusion

Il existe en amont des stratégies d'identification par spectrométrie de masse de nombreuses techniques séparatives dédiées à la séparation de protéines et de peptides. L'électrophorèse sur gel (1-D ou 2-D) pour la séparation d'échantillons protéiques et les systèmes nanoHPLC pour l'analyse de peptides sont des techniques maintenant bien établies (Figure 18).

<sup>90</sup> Link, A. J., et al. (1999). *Nat Biotechnol* **17**, 676–82.

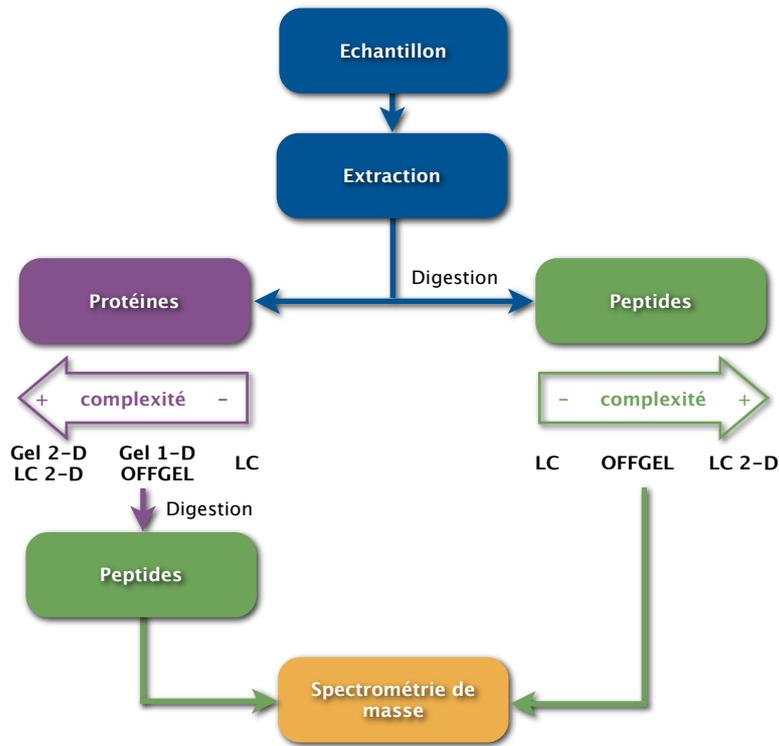
<sup>91</sup> Nagele, E., et al. (2004). *J Biomol Tech* **15**, 134–43.

<sup>92</sup> Wolters, D. A., et al. (2001). *Analytical Chemistry* **73**, 5683–90.

<sup>93</sup> Nesvizhskii, A. I., et al. (2005). *Mol Cell Proteomics* **4**, 1419–40.

Les sciences séparatives ne sont néanmoins pas en déclin. Au contraire, les dernières années ont vu arriver de nouvelles technologies innovantes comme la chromatographie sur puce microfluidique ou la chromatographie ultrahaute performance. Par ailleurs, face à la complexité des échantillons biologiques, de nouvelles approches sont développées comme la chromatographie multidimensionnelle.

Chacune de ces techniques présente des avantages et des inconvénients et le choix d'utiliser l'une ou une combinaison d'entre elles est souvent dicté par la nature des échantillons biologiques et l'objectif analytique recherché<sup>94</sup>. Une stratégie universelle n'existe pas : la recherche et l'optimisation de nouvelles stratégies doivent être systématiques.



**Figure 18** | Principales techniques séparatives utilisées en analyse protéomique par spectrométrie de masse.

<sup>94</sup> Issaq, H. J. (2003). *Adv Protein Chem* **65**, 249–69.



---

## Chapitre II

# Les stratégies d'identification en analyse protéomique par spectrométrie de masse

---

L'identification de protéines par spectrométrie de masse est généralement réalisée à partir de données générées lors de l'analyse de peptides produits par digestion enzymatique. L'ensemble des stratégies basées sur ces approches est désigné par le terme « *bottom-up proteomics* ». Parallèlement, une approche connue sous le nom de « *top-down proteomics* » est apparue. Elle implique l'analyse directe des protéines entières par spectrométrie de masse en tandem ultra-haute résolution<sup>1</sup>.

Ce chapitre propose la description des principales stratégies d'identification de protéines mises en oeuvre en analyse protéomique « *bottom-up* » qui demeure l'approche la plus utilisée et la plus simple à mettre en oeuvre pour les études à grande échelle<sup>2</sup>. Dans un deuxième temps, quelques exemples de défis stratégiques rencontrés en analyse protéomique sont présentés.

---

<sup>1</sup> McLafferty, F. W., et al. (2007). FEBS J 274, 6256–68.

<sup>2</sup> Han, X., et al. (2008). Curr Opin Chem Biol 12, 483–90.

# 1. Les stratégies d'identification de protéines

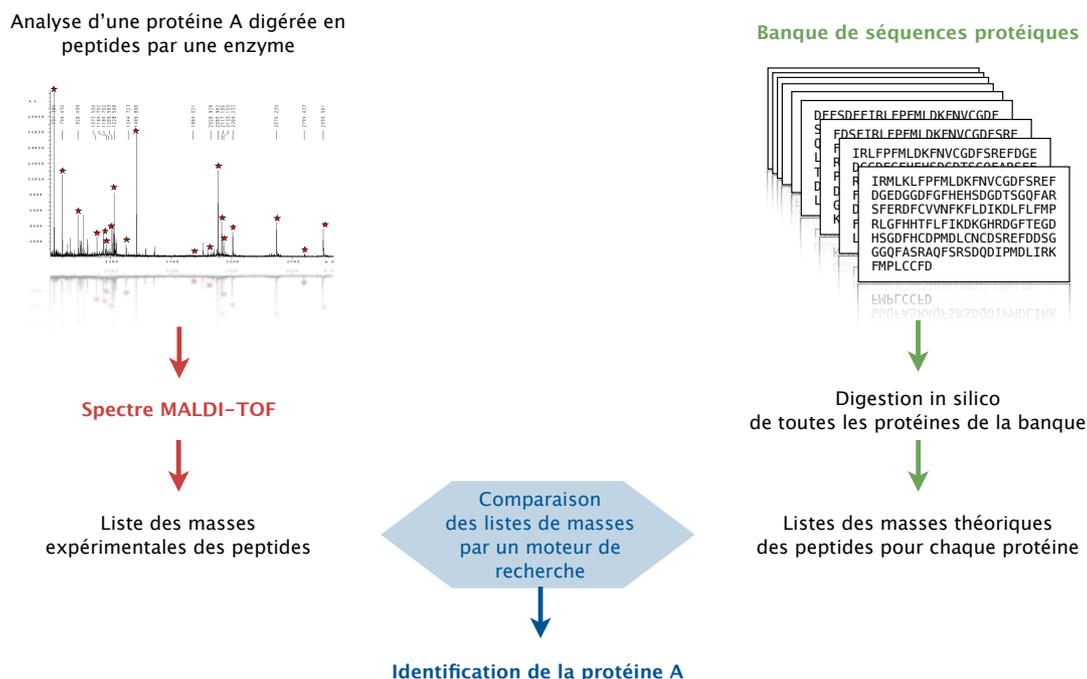
L'apparition de banques de séquences protéiques ainsi que les développements instrumentaux et analytiques pour la spectrométrie de masse des 20 dernières années ont donné naissance à toute une série de stratégies d'identification des protéines.

L'année 1993 marque la naissance du concept d'identification par empreinte peptidique massique (PMF), décrit simultanément par 5 groupes de chercheurs<sup>3,4,5,6,7</sup>. Un an après, des techniques similaires s'appuyant sur des masses de fragments de peptides ont été proposées<sup>8</sup> et les premières recherches utilisant des tags de séquences reconstitués à partir de spectres MS/MS ont paru<sup>9</sup>. A partir de là, de nombreux algorithmes de recherche permettant le traitement de données de spectrométrie de masse en tandem ont été développés. Chacun présente des particularités et participe à l'amélioration des identifications de protéines.

## 1.1. L'identification par empreinte peptidique massique (PMF)

### 1.1.1. Principe de l'identification par PMF

L'identification par empreinte peptidique massique (Figure 1) consiste en l'identification des protéines par la mesure des masses des peptides générés suite à une digestion enzymatique. Ces masses sont ensuite comparées à celles de peptides théoriques, calculées à partir de la digestion *in silico* de toutes les protéines stockées dans les banques de séquences .



**Figure 1** | Stratégie d'identification de protéines par empreinte peptidique massique à partir d'une analyse MALDI-TOF.

<sup>3</sup> Henzel, W. J., et al. (1993). Proc Natl Acad Sci U S A **90**, 5011–5.  
<sup>4</sup> James, P., et al. (1993). Biochem Biophys Res Commun **195**, 58–64.  
<sup>5</sup> Mann, M., et al. (1993). Biol Mass Spectrom **22**, 338–45.  
<sup>6</sup> Pappin, D. J., et al. (1993). Curr Biol **3**, 327–32.  
<sup>7</sup> Yates, J. R., 3rd, et al. (1993). Anal Biochem **214**, 397–408.  
<sup>8</sup> Eng, J. K., et al. (1994). Journal of the American Society for Mass Spectrometry **5**, 976–989.  
<sup>9</sup> Mann, M., et al. (1994). Anal Chem **66**, 4390–9.

---

Généralement, les identifications de protéines par empreinte peptidique massique sont réalisées sur des extraits peptidiques peu complexes (digests du contenu de spots de gels 2D par exemple) analysés par spectrométrie de masse MALDI-TOF<sup>10</sup>. Rapides et facilement automatisables, les analyses par MALDI-TOF sont particulièrement adaptées pour le PMF. Leurs tolérances aux sels et aux détergents, leur bonne sensibilité, leur excellente résolution (de l'ordre de 25 000) et leur bonne précision de masse (50 ppm) sont autant d'atouts permettant une analyse (et donc une identification) rapide et facile du contenu protéique d'extraits simples. Les identifications par PMF peuvent aussi être réalisées à partir de données de LC-MS obtenues sur des instruments permettant de mesurer des masses avec une bonne précision (dernières générations de spectromètres de masse ESI-TOF ou FT-ICR).

La validation des identifications par empreinte peptidique massique répond à quelques critères simples :

- le pourcentage de recouvrement de séquence obtenu ou plus exactement le nombre de peptides identifiés (le pourcentage de recouvrement dépend de la taille de la protéine et peut donc être élevé même si le nombre de peptides identifiés est faible chez les petites protéines) ;
- les erreurs observées sur les mesures de masse des peptides ;
- le nombre de pics non identifiés présents sur le spectre ou appartenant à d'autres protéines ;
- la comparaison entre le pI et le poids moléculaire théorique de la protéine identifiée à sa position sur le gel lorsqu'il s'agit d'analyses de spots de gel 2D.

### 1.1.2. Les limites du PMF

Les critères d'acceptation des identifications par PMF sont inclus dans les directives de publication de données de protéomique décrites au chapitre 3 de cette partie, point 3.1. Toutefois, la publication de résultats protéomiques uniquement issus des approches par empreinte peptidique massique est de plus en plus restreinte. En effet, avec la croissance exponentielle du nombre d'entrées de protéines dans les banques (cf. Chapitre 3 de cette partie, point 1), la spécificité du PMF est de plus en plus limitée<sup>11</sup>. De fait, des peptides ayant la même masse peuvent avoir des séquences en acides aminés différentes. N'avoir accès qu'à la masse de ces peptides ne permet pas de les discriminer et l'identification ne peut pas être validée. Avec le nombre de séquences dans les banques, le nombre de masses théoriques augmente et donc le nombre de masses redondantes.

L'apparition des instruments à ultra-haute résolution de type MALDI-FT-ICR, capables de mesures sub-ppm, a permis d'améliorer considérablement la spécificité des identifications PMF<sup>12</sup>. Toutefois, un certain nombre de limites persiste, au premier rang desquelles on peut placer l'impossibilité d'analyser directement des mélanges complexes qui vont donner lieu à la superposition d'empreintes peptidiques sur le même spectre, rendant l'identification impossible. Par ailleurs, on peut citer les difficultés rencontrées dans l'analyse de protéines de faible masse moléculaire générant peu de peptides et qui vont donc pouvoir correspondre à un nombre important d'entrées dans les banques. Enfin, le signal des peptides faiblement abondants peut se retrouver « écrasé » par les pics les plus intenses présents sur le même spectre. Ces effets de « suppression de signal » impliquent une perte d'information dommageable.

---

<sup>10</sup> Thiede, B., et al. (2005). *Methods* **35**, 237-47.

<sup>11</sup> Gattiker, A., et al. (2002). *Proteomics* **2**, 1435-44.

<sup>12</sup> Horn, D. M., et al. (2004). *International Journal of Mass Spectrometry* **238**, 189-196.

## 1.2. Les stratégies nanoLC–MS/MS

Face aux difficultés rencontrées dans les approches par empreinte peptidique massive, les stratégies qui s'appuient sur les analyses de spectrométrie de masse en tandem se sont rapidement développées. Ces stratégies nanoLC–MS/MS représentent aujourd'hui le coeur de l'analyse protéomique par spectrométrie de masse<sup>13</sup>. En utilisant l'information de séquence des peptides analysés, elles profitent d'une spécificité inégalée pour l'identification de protéines. De plus, l'étape de chromatographie en phase inverse directement couplée à la spectrométrie de masse en tandem permet la séparation préalable des peptides et évite la compétition entre eux au moment de l'ionisation ESI.

Lors d'analyses nanoLC–MS/MS, la sélection d'ions parents à fragmenter peut se faire manuellement, mais la plupart des stratégies développées mettent en oeuvre des solutions automatisées. Le mode d'acquisition données-dépendant (DDA) est le plus couramment adopté. Il permet notamment de reconstituer la séquence en acides aminés du peptide analysé (séquençage *de novo*), mais récemment de nouveaux modes d'acquisition originaux ont été proposés.

### 1.2.1. Identification en mode d'acquisition données-dépendant (DDA)

#### ▸ Le mode d'acquisition données-dépendant

Lors d'une analyse nanoLC–MS/MS, les spectromètres de masse en tandem permettent d'automatiser en temps réel les décisions pour la réalisation de mesures MS/MS. Cela permet à l'instrument de sélectionner lors d'un scan MS un ou plusieurs ions qui seront fragmentés et analysés lors de mesures MS/MS successives. La sélection est réalisée sans intervention systématique de l'utilisateur, mais en fonction de paramètres de sélection physiques (nombre, intensité, charge des parents) et temporels (temps d'acquisition MS et MS/MS maximums) qu'il aura préalablement déterminés. Le fonctionnement des analyses en mode DDA et l'optimisation des paramètres de sélections sont détaillés dans le chapitre I de la 1<sup>re</sup> partie des résultats, aux points 1.1.1 et 2.3, respectivement.

Les données générées par ces analyses sont regroupées dans des fichiers de résultats de type « *Peak List* ». Ils rassemblent l'ensemble des caractéristiques (masse, état de charge, intensité et éventuellement temps de rétention chromatographique) des ions parents sélectionnés et l'ensemble des masses des ions fragments qui leurs correspondent.

#### ▸ Identification par empreinte de fragmentation peptidique (« *Peptide Fragment Fingerprinting* », PFF)

La dénomination « *Peptide Fragment Fingerprinting* » (PFF) a été introduite en 2004<sup>14</sup> par analogie avec la dénomination PMF. Cette approche est basée sur la comparaison des listes de masses des ions parents (mesurées lors des séquences MS) et de leurs fragments associés (mesurés dans chaque spectre MS/MS) avec des listes de masses issues de la digestion théorique des protéines présentes dans des banques de données et de la fragmentation théorique des peptides associés (Figure 2).

De nombreux algorithmes de recherche dans les banques de données fonctionnent sur ces approches par PFF<sup>15, 16</sup>. Chaque programme peut présenter des particularités de paramétrage et des stratégies de traitement des résultats différents : leurs principales différences sont présentées au point 2.1 du chapitre III de cette partie.

<sup>13</sup> Domon, B., et al. (2006). *Science* **312**, 212–7.

<sup>14</sup> Blueggel, M., et al. (2004). *Curr Pharm Biotechnol* **5**, 79–88.

<sup>15</sup> Hernandez, P., et al. (2006). *Mass Spectrom Rev* **25**, 235–54.

<sup>16</sup> Nesvizhskii, A. I., et al. (2007). *Nat Methods* **4**, 787–97.

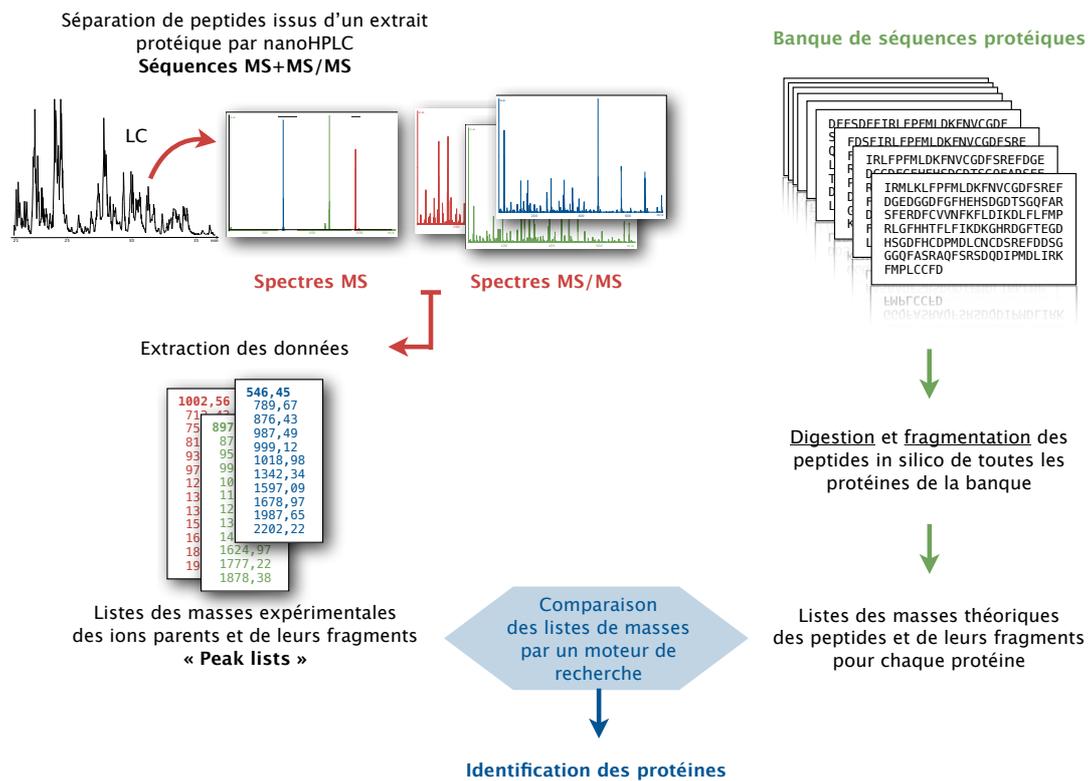


Figure 2 | Stratégie d'identification de protéines par nanoLC-MS/MS

► Les approches par séquençage « de novo »

En mode d'acquisition DDA, une alternative pour l'identification des protéines est le séquençage de novo des peptides analysés. Dans cette approche, les séquences en acides aminés sont directement déduites des différences de masses entre pics successifs sur les spectres MS/MS expérimentaux. Comme détaillé au point 1.2 du chapitre I de cette partie, la différence de masse entre deux fragments de peptide consécutifs issus d'une même série correspond à la masse d'un acide aminé. Ainsi, il est possible de reconstituer toute ou partie de la séquence du peptide fragmenté en considérant les différences de masses entre les membres d'une série de pics successifs (Figure 3).

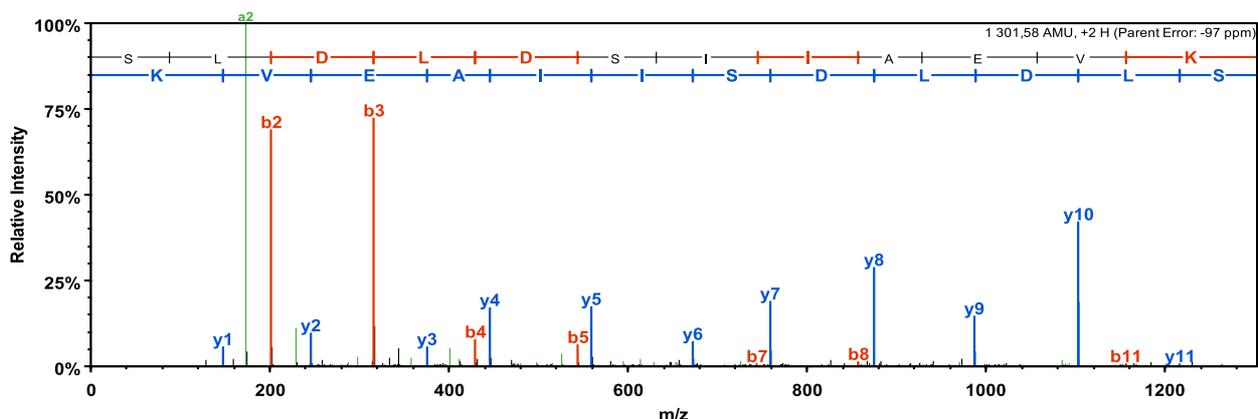


Figure 3 | Exemple de spectre de fragmentation interprété manuellement d'un ion à  $m/z$  651,799, 2+ analysé par nanoLC-MS/MS sur le système CapLC-Q/TOF II Waters Corp. La séquence SLDLDSIAAEVK du peptide peut être directement reconstituée en relevant les différences de masses entre les pics consécutifs d'une même série de fragmentation (ici et généralement, la série y et la série b).

Initialement, l'interprétation des spectres ne pouvait se faire que manuellement et à l'aide des outils informatiques fournis par les constructeurs de spectromètres de masses (DataAnalysis, Bruker Dal. ou

PepSeq, Waters Corp. par exemple). Récemment, de nombreux outils bio-informatiques ont été développés afin d'automatiser le séquençage direct des peptides sur leur spectre de fragmentation<sup>17</sup>.

Le logiciel Peaks Studio (Bioinformatics Solutions Inc.)<sup>18</sup> apparu en 2003 est notamment utilisé au laboratoire. A partir des « *Peak Lists* », fichiers de liste de masses des spectres de fragmentation acquis, le logiciel recherche de manière automatique les séquences peptidiques pouvant être déduites des données MS/MS. Ces recherches peuvent être affinées en fixant des paramètres tels que l'enzyme de digestion utilisée ou les modifications posttraductionnelles éventuellement présentes.

Les séquences proposées à partir de chaque spectre MS/MS, annoté manuellement ou automatiquement, peuvent alors être soumises à des programmes de recherche de similarité de séquences, par exemple MS-BLAST<sup>19</sup> (« *Basic Local Alignment Search Tool* »). Ce type de programme va comparer les séquences proposées par Peaks aux séquences présentes dans les banques (figure 24). Si ce type d'approche nécessite d'avoir des spectres de très bonne qualité et que le temps d'interprétation est considérablement augmenté, elle peut néanmoins se révéler très utile.

▸ D'abord, parce que l'approche « *Peptide Fragment Fingerprint* » peut se révéler vaine notamment dans le cas d'études menées sur des organismes peu ou non séquencés ou encore dans le cas de recherches de modifications posttraductionnelles ou de mutations sur les chaînes d'acides aminés.

▸ Ensuite, parce que l'approche par séquençage direct des peptides a longtemps été considérée comme un argument de poids pour la validation manuelle des identifications ambiguës (cf. chapitre 3 de cette partie).

▸ Enfin, parce que les logiciels d'interprétation des données de fragmentation de peptides comme Peaks Studio offrent depuis peu de nouvelles fonctions prometteuses permettant la recherche directement dans les banques de données à partir du traitement de novo des spectres expérimentaux. Les premiers tests de cette approche sont présentés dans la 1<sup>re</sup> partie des résultats, au chapitre I, point 3.3.2.

### 1.2.2. De nouveaux modes d'acquisition originaux (MS<sup>E</sup>)

En 2006, la société Waters Corp. a proposé un nouveau mode d'acquisition des données en spectrométrie de masse en tandem couplée à la chromatographie liquide ultra-haute performance : le mode « *MS-Expression* » (MS<sup>E</sup>). Dans le mode d'acquisition DDA, des mesures MS sont suivies – de manière séquentielle – de mesures MS/MS sur un nombre donné d'ions sélectionnés : de nombreux ions ne sont donc pas analysés (cf. 1<sup>re</sup> partie des résultats, chapitre I, point 1.1). Dans le mode MS<sup>E</sup>, à l'issue d'une mesure MS, tous les ions sont dans le même temps fragmentés en parallèle ; la perte d'information est donc théoriquement limitée (Figure 4)<sup>20</sup>. Les données MS sont utilisées pour déterminer la masse exacte et l'intensité du signal des peptides parents. Les spectres MS<sup>E</sup>, très complexes, représentent les données MS/MS de tous les peptides visualisés dans le cycle MS.

Puisque chaque ion fragment suit le profil d'élution de son peptide précurseur, chaque fragment peut être associé à son parent par alignement des temps d'élution exacts. Des listes de masses d'ions fragments et de leurs ions précurseurs sont ainsi générées et leur comparaison avec les données issues des banques de données protéiques aboutit à l'identification des protéines contenues dans l'échantillon.

L'avantage apparent de ce mode d'acquisition est qu'il permet théoriquement d'acquérir des données de fragmentation sur tous les peptides élués sans les discriminer. Cependant, il reste encore peu utilisé et sa présence dans la littérature est encore discrète. Deux raisons principales l'expliquent. La première est que la fragmentation sans présélection de tous les peptides d'une fraction chromatographique empêche tout traitement ou lecture manuelle de l'information brute présente sur les spectres de fragmentation. Aucune

<sup>17</sup> Pevtsov, S., et al. (2006). *J Proteome Res* 5, 3018–28.

<sup>18</sup> Ma, B., et al. (2003). *Rapid Commun Mass Spectrom* 17, 2337–42.

<sup>19</sup> Shevchenko, A., et al. (2001). *Anal Chem* 73, 1917–26.

<sup>20</sup> Geromanos, S. J., et al. (2009). *Proteomics* 9, 1683–95.

vérification ou validation ultérieure ne peut donc avoir lieu et une confiance aveugle doit être donnée à l'algorithme de traitement des données MS<sup>E</sup>. La deuxième est justement l'existence d'un unique algorithme commercial : aucun comparatif ou validation croisée des résultats ne peut être réalisé avec un autre logiciel de recherche dans les banques de données.

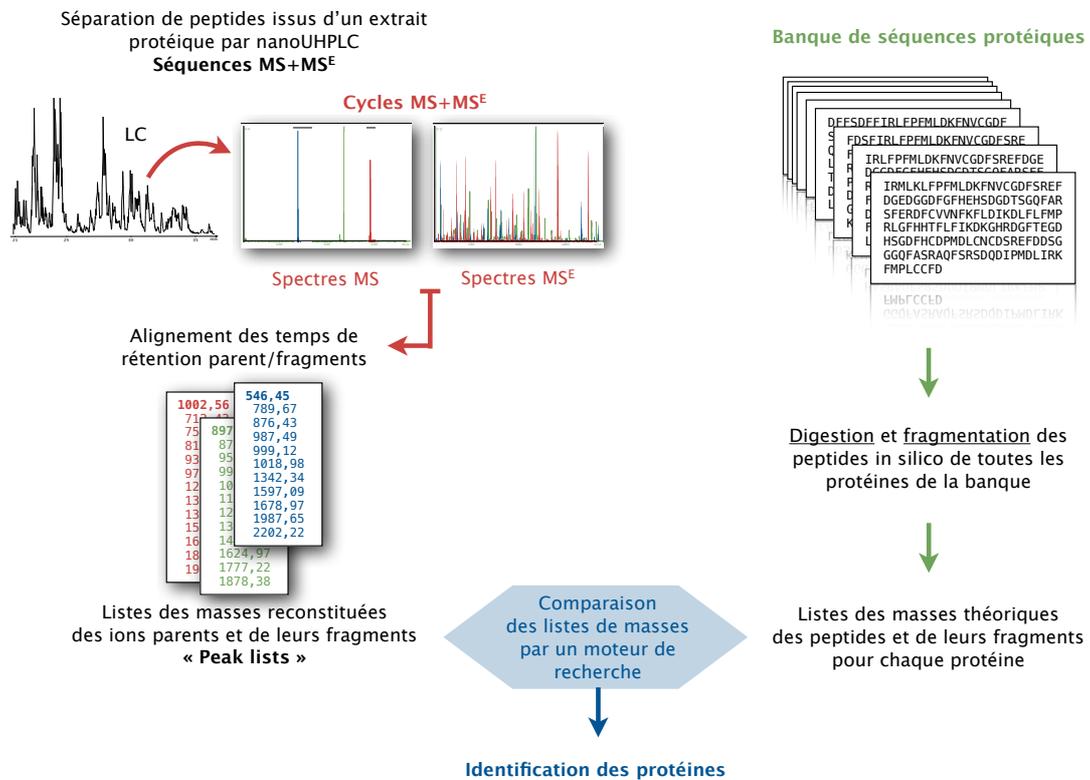


Figure 4 | Stratégie d'identification de protéines par UHPLC-MS<sup>E</sup>

## 2. Exemples de défis stratégiques en analyse protéomique

Comme nous l'énoncions en introduction, la définition de l'analyse « protéomique » a évolué à mesure que les défis et le champ des possibilités offertes par les stratégies d'analyse par spectrométrie de masse s'élargissaient. Le cahier des charges ne se limite plus à la « simple » identification des protéines présentes dans un échantillon : la caractérisation des modifications posttraductionnelles, la quantification relative ou absolue de protéines, l'étude fonctionnelle de complexes protéiques et l'analyse de protéines comme les protéines membranaires, sous-représentées dans les préparations d'échantillons classiques, sont autant de nouvelles questions adressées à l'analyse protéomique par spectrométrie de masse.

La suite de ce chapitre décrit brièvement quelques options stratégiques qui ont été développées ces dernières années pour répondre à ces défis.

### 2.1. Les stratégies de quantification

La quantification de protéines par spectrométrie de masse n'est pas une chose triviale<sup>21</sup>. La relation entre la concentration en analyte dans l'échantillon de départ et l'intensité du signal mesurée avec le spectromètre de masse dépend d'une multitude de facteurs qui sont difficilement contrôlables : l'accessibilité de l'enzyme lors de la digestion, la solubilité des peptides lors de l'extraction des gels, l'efficacité d'ionisation aussi bien en MALDI qu'en ESI<sup>22</sup>.

Des techniques d'analyse protéomique comparative et quantitative ont été développées ces dernières années. La quantification relative de protéines séparées sur gel d'électrophorèse 2-D peut être réalisée grâce à des outils d'analyse d'images, notamment lorsque des techniques de marquage sont utilisées (approches DiGE, « *Differential Gel Electrophoresis* »)<sup>23</sup>. Cependant, ces stratégies sont sujettes aux limites imposées par l'analyse sur gel (cf. chapitre I de cette partie, point 2.1.). De nouvelles techniques basées sur l'utilisation d'isotopes stables ou de standards internes marqués ont alors été développées<sup>24</sup>.

#### 2.1.1. Le marquage de protéines ou de peptides

Ces stratégies profitent du fait que des formes d'une même molécule qui ne diffèrent que par la substitution d'isotopes stables se comportent de manière identique lors d'une analyse MS : seule leur différence de masse permet de les distinguer. Le rapport entre l'intensité des deux pics indique donc précisément les quantités relatives des deux espèces (Figure 5).

Plusieurs protocoles majeurs permettant la quantification en utilisant des isotopes stables (principalement <sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N et <sup>18</sup>O) ont été décrits<sup>25</sup> :

- ▶ La quantification absolue par utilisation de **peptides marqués (AQUA)**<sup>26</sup>. Un peptide d'intérêt est synthétisé en substituant certains isotopes et est ajouté au digest protéique à une concentration fixée. Cette méthode de quantification absolue est très onéreuse (un peptide ad hoc marqué doit être synthétisé pour chaque protéine à quantifier).

- ▶ Le marquage métabolique utilisant des **acides aminés marqués**<sup>27</sup> : stratégie **SILAC** (« *Stable Isotope Labelling with Amino acids in Cell culture* »). Un lot de cellules contrôle sont mises en culture dans un milieu où

<sup>21</sup> Bantscheff, M., et al. (2007). *Anal Bioanal Chem* **389**, 1017–31.

<sup>22</sup> Steen, H., et al. (2004). *Nat Rev Mol Cell Biol* **5**, 699–711.

<sup>23</sup> Unlu, M., et al. (1997). *Electrophoresis* **18**, 2071–7.

<sup>24</sup> Panchaud, A., et al. (2008). *J Proteomics* **71**, 19–33.

<sup>25</sup> Lill, J. (2003). *Mass Spectrom Rev* **22**, 182–94.

<sup>26</sup> Gerber, S. A., et al. (2003). *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940–6945.

<sup>27</sup> Ong, S. E., et al. (2002). *Mol Cell Proteomics* **1**, 376–86.

un ou plusieurs acides aminés ont été remplacés par une forme marquée isotopiquement. Des quantités égales de cellules issues de l'échantillon et du contrôle sont mélangées avant tout traitement analytique.

▸ Le marquage par **réactions chimiques**<sup>28</sup> : stratégie **ICAT** (« *Isotope-Coded Affinity Tag* »). Les deux extraits cellulaires à comparer sont d'abord marqués avec un réactif, l'un avec un réactif « lourd » (2H) et l'autre avec un réactif « léger » (1H). Les extraits sont ensuite mélangés et les protéines digérées. Le marqueur possède à son extrémité un groupement biotine qui permet la capture sélective des peptides marqués suite à un passage sur colonne d'affinité. Seuls les peptides marqués seront donc analysés et quantifiés relativement.

▸ **L'incorporation enzymatique d'<sup>18</sup>O** durant la protéolyse<sup>29</sup>. Un lot de protéines contrôle sont digérées dans H<sub>2</sub><sup>18</sup>O ce qui a pour conséquence d'incorporer un <sup>18</sup>O aux extrémités C-ter des peptides générés. Les peptides marqués sont alors mélangés à des peptides digérés de manière classique.

▸ Le **marquage isobare**<sup>30</sup> pour la quantification relative et absolue : stratégie **iTRAQ** (« *isobaric Tag for Relative and Absolute Quantitation* »). Un réactif iTRAQ est fixé à l'extrémité N-terminale de tous les peptides d'un digest. Jusqu'à 8 étiquettes isobares peuvent être utilisées pour marquer plusieurs lots échantillon ou contrôle avant de les mélanger. Les différentes formes isotopiques n'apparaissent que sur le spectre MS/MS des peptides, l'élément « rapporteur » de l'étiquette étant détaché lors de la fragmentation.

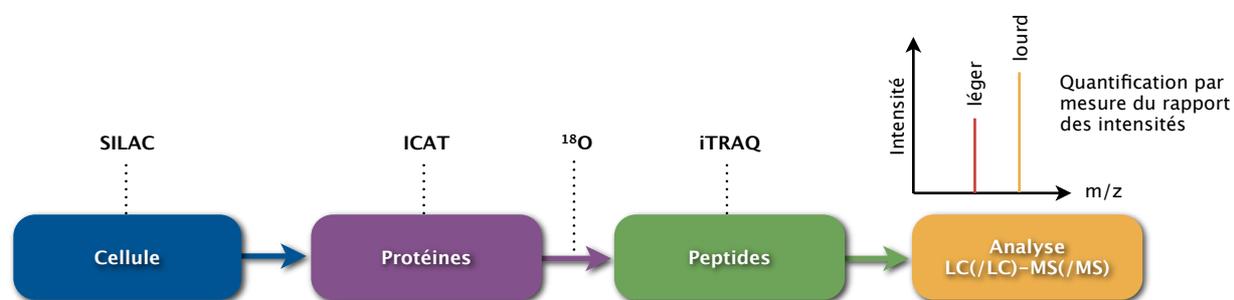


Figure 5 | Stratégies de quantification par marquage couramment utilisées en spectrométrie de masse.

### 2.1.2. Les approches sans marquage

Les approches de quantification sans marquage sont particulièrement intéressantes, car elles n'impliquent pas de modification majeure de l'échantillon<sup>31</sup>. Pour quantifier une protéine sans marquage, deux méthodes distinctes peuvent être mises en œuvre en ionisation electrospray :

▸ La mesure et la comparaison de l'**intensité du signal de ses peptides** (intégration et somme de leurs chromatogrammes d'ions) ;

▸ Le comptage et la comparaison du **nombre de spectres acquis de ses peptides** (empiriquement, le nombre de spectres MS/MS acquis pour les peptides d'une protéine est proportionnel à son abondance dans l'échantillon).

Une normalisation globale sur toute l'analyse (somme des intensités ou du nombre de spectres total) ou l'utilisation de standards internes permettent une quantification relative ou absolue.

### 2.1.3. Des modes d'acquisition particulièrement adaptés à la quantification

Outre l'analyse en mode DDA classique – au cours de laquelle des mesures MS permettent la sélection d'un nombre d'ions limité pour fragmentation et analyses MS/MS successives – deux modes d'acquisition sont décrits comme étant particulièrement intéressants pour les expériences de quantification :

<sup>28</sup> Gygi, S. P., et al. (1999). Nat Biotechnol 17, 994–9.

<sup>29</sup> Bantscheff, M., et al. (2004). Rapid Commun Mass Spectrom 18, 869–76.

<sup>30</sup> Ross, P. L., et al. (2004). Mol Cell Proteomics 3, 1154–69.

<sup>31</sup> Patel, V. J., et al. (2009). J Proteome Res 8, 3752–9.

▸ L'**acquisition ciblée par MRM**<sup>32</sup> (« *Multiple Reaction Monitoring* ») est un mode d'acquisition permettant une détection et une quantification très sensible de molécules ciblées dans des mélanges complexes. Son avantage principal est de permettre la détection d'un ion même lorsque celui-ci est masqué par d'autres présents sur le même spectre. Le mode MRM est généralement mis en oeuvre grâce à l'analyseur triple quadripolaire. Il permet le suivi d'un seul ion spécifique, détecté par filtrage dans le 1<sup>er</sup> quadripôle du précurseur cible parmi tous les autres et, après fragmentation, par la sélection dans le 3<sup>e</sup> quadripôle des fragments spécifiques attendus. Seules les molécules répondant à ces deux critères (précurseur de m/z fixé produisant les fragments attendus) seront détectées. Cette méthode ne s'applique donc que pour les études ciblées de protéines aux espèces dont le profil de fragmentation est établi.

▸ Le **mode MS<sup>E</sup>**, précédemment décrit, utilise l'approche opposée<sup>33</sup> : tous les précurseurs et les fragments de tous les précurseurs sont détectés quasi simultanément. Ce mode est très favorable aux approches de quantification sans marquage, car il n'impose pas de sous échantillonnage comme en mode DDA classique (cf. point 1.2.2. de ce chapitre). Cependant, il présente des désavantages importants : une faible sensibilité sur les spectres de fragmentation des peptides peu abondants est notamment à déplorer<sup>34</sup>.

## 2.2. L'analyse de partenaires d'interaction protéiques

L'analyse de complexes protéiques est un des défis majeurs de l'analyse protéomique et un des domaines où elle a eu un impact important. La plupart des protéines (80 %) exercent leur fonction au sein de complexes multiprotéiques<sup>35</sup>. Notamment des fonctions cellulaires importantes telles que la réplication, la transcription et la traduction de l'ADN requièrent l'intervention coordonnée d'un grand nombre de protéines (cf. 1<sup>re</sup> partie des résultats, chapitre III). Ainsi, une des questions essentielles pour la compréhension des mécanismes cellulaires porte sur l'identité des partenaires potentiels des protéines étudiées et l'évolution des complexes au cours du temps et en fonction de leur environnement<sup>36</sup>. On parle alors d'analyse protéomique fonctionnelle qui vise à l'étude de l'interactome<sup>37</sup>.

### 2.2.1. La co-immunopurification et la co-purification par affinité (TAP-Tag)

Pour traiter ces questions par spectrométrie de masse, des méthodes de purification de complexes protéiques ont été développées : la protéine cible elle-même est utilisée comme un réactif d'affinité pour isoler ses partenaires d'interaction. Comparée aux autres techniques (double-hybride, Co-IP ...)<sup>38</sup>, cette stratégie par spectrométrie de masse présente l'avantage de permettre une utilisation directe de la protéine comme appât, dans son environnement physiologique et son compartiment cellulaire, pour isoler et analyser en une seule opération les complexes multiprotéiques<sup>39</sup>. Pour purifier des complexes endogènes multiprotéiques, sont distinguées :

▸ les méthodes de **co-immunopurification** (Figure 6a) pour lesquelles le complexe protéique est purifié depuis un lysat cellulaire en utilisant un anticorps immobilisé et dirigé contre un des constituants du complexe (la protéine cible). Cette technique simple impose toutefois la disponibilité d'anticorps spécifique et efficace ;

▸ les méthodes de **co-précipitation par affinité** (Figure 6b) qui exploitent les propriétés biochimiques d'une étiquette fixée sur la protéine cible constituant un appât pour les autres membres du

<sup>32</sup> Lange, V., et al. (2008). *Mol Syst Biol* 4, 222.

<sup>33</sup> Silva, J. C., et al. (2006). *Mol Cell Proteomics* 5, 144-56.

<sup>34</sup> Wilm, M. (2009). *Proteomics*.

<sup>35</sup> Berggard, T., et al. (2007). *Proteomics* 7, 2833-42.

<sup>36</sup> Aebersold, R., et al. (2003). *Nature* 422, 198-207.

<sup>37</sup> Kocher, T., et al. (2007). *Nat Methods* 4, 807-15.

<sup>38</sup> Suter, B., et al. (2008). *Biotechniques* 44, 681-91.

<sup>39</sup> Ashman, K., et al. (2001). *Sci STKE* 2001, PE33.

complexe. Les séquences de la protéine cible et de l'étiquette sont fusionnées et le construit est transfecté dans les cellules cibles. Le complexe purifié est obtenu en le lavant sur une colonne d'affinité spécifique pour l'étiquette.

De nombreuses étiquettes sont disponibles, chacune pouvant exercer une influence différente sur le repliement ou la solubilité du construit et le rendement de la purification<sup>40</sup>. Pour tirer parti d'une possible complémentarité des étiquettes, une méthode de **double co-purification par affinité** est apparue en 1999 (Figure 6c) : la méthode TAP-Tag (« *Tandem Affinity Purification* »)<sup>41</sup>. Cette stratégie met en oeuvre 2 étapes de purification et élution spécifique en utilisant deux étiquettes séparées par un lien clivable enzymatiquement. A l'origine, les deux étiquettes utilisées étaient le « *calmodulin-binding peptide* » (CBP) et la protéine A (Staphylococcus Aureus) mais de nombreuses autres possibilités ont été utilisées (cf. notamment la mise en oeuvre de la technique TAP-Tag FLAG-HA dans la 1<sup>re</sup> partie des résultats, au chapitre III).

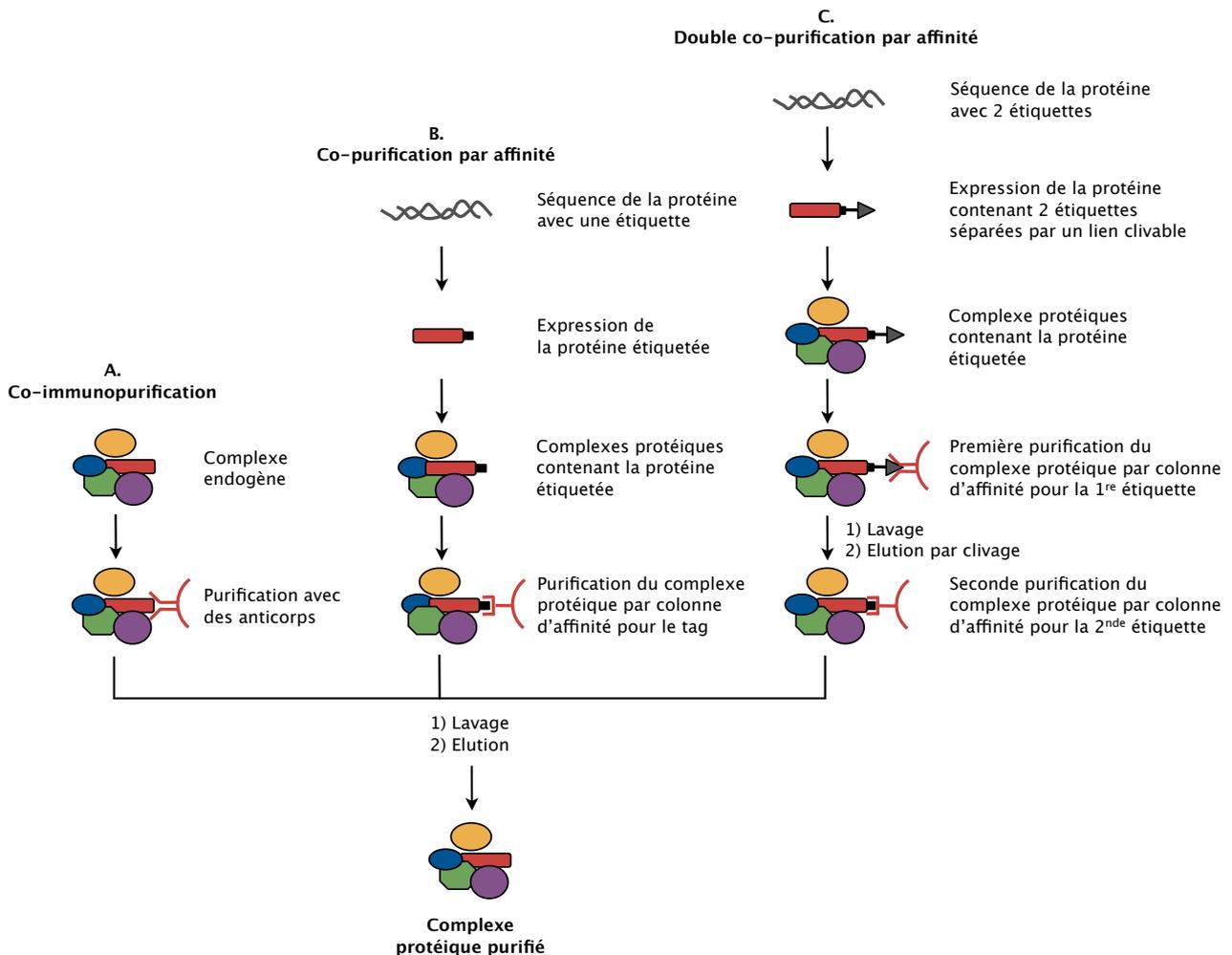


Figure 6 | Stratégies principales pour la purification de complexes protéiques (d'après Köcher, 2007).

## 2.2.2. Les limites des techniques de co-purification par affinité

L'approche TAP-Tag a permis la purification de nombreux complexes protéiques avec une faible contamination, mais généralement au prix d'une perte des interactions transitoires les plus faibles qui ne résistent pas au double lavage<sup>42</sup>. Beaucoup d'interactions protéine-protéine, bien qu'elles puissent être importantes, présentent une faible affinité, sont transitoires ou dépendent fortement de l'environnement cellulaire dans lequel elles se forment. Les méthodes d'identification par spectrométrie de masse de

<sup>40</sup> Waugh, D. S. (2005). Trends Biotechnol 23, 316–20.

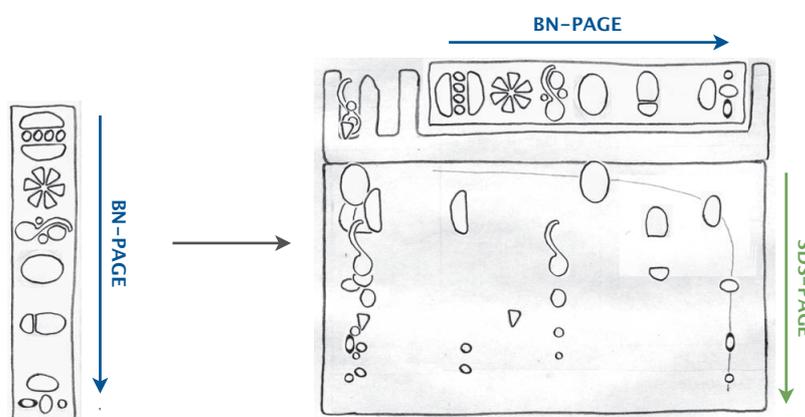
<sup>41</sup> Rigaut, G., et al. (1999). Nat Biotechnol 17, 1030–2.

<sup>42</sup> Puig, O., et al. (2001). Methods 24, 218–29.

complexes directement purifiés ne pourront donc avoir accès qu'à une partie des interactions qui ont effectivement lieu. La corrélation de données obtenues avec d'autres méthodes ou des expériences par spectrométrie de masse en utilisant des réactifs formant des liaisons croisées (« *chemical crosslinking* ») permet de confirmer certaines interactions directes et la topologie des complexes multiprotéiques étudiés.

### 2.2.3. La séparation du complexe purifié : SDS-PAGE et BN-PAGE

Une fois le complexe purifié par une méthode par affinité, ses constituants sont généralement séparés par SDS-PAGE. Toutefois dans ce cas, la technique ne permet pas de déterminer la taille, l'abondance relative et les sous-unités du complexe étudié<sup>43</sup>. Pour répondre à ces questions, une séparation du complexe purifié peut être réalisée sur gradient de sucrose par ultracentrifugation ou sur gel d'électrophorèse natif (« *Blue Native PAGE* »)<sup>44</sup>. Pour cette approche BN-PAGE, potentiellement plus résolutive, le complexe est traité au bleu de Coomassie, qui se fixe de manière aspécifique à toutes les protéines sans dénaturer les sous-complexes et qui est lui-même chargé négativement. Ainsi, seules la forme et la taille des sous complexes déterminent leur mobilité électrophorétique dans le gel de polyacrylamide. Cette dimension de séparation peut être doublée d'une deuxième étape SDS-PAGE classique permettant de séparer les protéines membres de chacun des sous-complexes identifiés<sup>45</sup>.



**Figure 7** | Principe du BN-PAGE. Les sous-complexes sont séparés en conditions natives dans un première dimension BN-PAGE. Pour un gel 2-D, les sous complexes sont ensuite dénaturés par traitement SDS de la bandelette de gel et séparés par SDS-PAGE classique. Les monomères sont élués selon une diagonale « hyperbolique » dans le cas d'un gel BN-PAGE gradient et SDS-PAGE linéaire. (D'après Camacho-Carvajal, 2004).

## 2.3. L'analyse de protéines membranaires

Les membranes plasmiques jouent un rôle crucial dans la structure cellulaire en constituant une barrière physique entre la cellule et son environnement. Bien que la structure de base des membranes plasmiques (et de toute membrane biologique) soit constituée d'une double couche lipidique, la plupart de leurs fonctions spécifiques sont assurées par les protéines qui y sont enchâssées ou associées<sup>46</sup>. L'étude des protéines membranaires ne trouve pas uniquement son intérêt dans la compréhension du fonctionnement des processus biologiques essentiels dans lesquels elles sont engagées (signalisation cellulaire, interactions cellulaires, transport sélectif de matière, processus énergétiques ...), mais aussi dans l'identification de protéines membranaires de surface exprimées spécifiquement par certaines cellules, par exemple pathogènes. Des protéines membranaires plasmiques peuvent alors représenter des marqueurs potentiels

<sup>43</sup> Swamy, M., et al. (2006). *Sci STKE* **2006**, pl4.

<sup>44</sup> Wittig, I., et al. (2008). *Proteomics* **8**, 3974-90.

<sup>45</sup> Camacho-Carvajal, M. M., et al. (2004). *Mol Cell Proteomics* **3**, 176-82.

<sup>46</sup> Tan, S., et al. (2008). *Proteomics* **8**, 3924-32.

---

pouvant être considérés comme cibles diagnostiques et/ou thérapeutiques (cf. 1<sup>re</sup> partie des résultats, chapitre II).

### 2.3.1. Les difficultés de l'analyse de protéines membranaires par spectrométrie de masse

Les protéines membranaires ont la particularité d'être composées de domaines hydrophobes et de domaines hydrophiles. Ce caractère amphiphile leur permet d'être localisées dans la membrane plasmique, mais les rend dans le même temps difficiles à analyser. Elles sont en effet généralement sous-représentées dans les résultats d'analyse protéomique<sup>47</sup>, alors qu'il est estimé que 20 à 30 % du génome humain code pour des protéines membranaires<sup>48</sup>.

Les difficultés attribuées à l'identification de protéines membranaires reposent généralement sur leur hétérogénéité, leur hydrophobicité et leur faible abondance. Afin de dépasser ces limites, de multiples stratégies ont été développées pour améliorer l'enrichissement des échantillons, la solubilisation des protéines, leur séparation et leur digestion<sup>49</sup>.

▸ La **faible abondance** de protéines membranaires limite leur détection dans des mélanges complexes où des protéines sont présentes sur de larges gammes dynamiques de concentration. Des stratégies de fractionnement ont été proposées pour enrichir les échantillons en protéines membranaires. Elles sont généralement réalisées par centrifugation différentielle sur gradient de sucrose ou par purification par affinité après biotinylation, mais d'importants problèmes de contamination par des protéines des membranes d'autres organelles se posent<sup>50</sup>. Récemment, une technique d'enrichissement simple et originale par production de microparticules a été proposée. Elle est détaillée au point suivant.

▸ Les **difficultés de solubilisation** des protéines membranaires, liées à leur grande hydrophobicité, nécessitent l'utilisation de détergents incompatibles avec les techniques classiques de séparation, digestion et analyse. Elles sont notamment sous représentées dans les gels 2-D<sup>51</sup> puisque seuls des détergents non ioniques ou zwitterioniques (CHAPS, thiourée ...) – bien moins efficaces que le SDS par exemple – peuvent être utilisés dans la première dimension de séparation. En outre, les protéines membranaires ont tendance à précipiter à leur pI lors de l'isoélectrofocalisation. Leur pI est d'ailleurs souvent alcalin et le nombre de protéines membranaires accessibles sur gel classique est de fait réduit.

La séparation sur gel 1-D (SDS-PAGE), bien que moins résolutive, reste une solution intéressante pour la séparation des protéines membranaires. D'autres stratégies comme les gels 2-D SDS-PAGE/SDS-PAGE<sup>52</sup> ou BN-PAGE/SDS-PAGE<sup>53</sup> ou les techniques sans gel de chromatographie liquide multidimensionnelle (« *Shotgun proteomics - MudPIT* »)<sup>54</sup> ont également été proposées.

### 2.3.2. Les microparticules : un outil de choix pour l'isolation de protéines membranaires plasmiques

Bien qu'elles aient été découvertes dans les années 1940<sup>55</sup>, les microparticules (MP) – aussi appelées microvésicules ou ectosomes – ont longtemps été considérées comme des « débris cellulaires » impliqués dans la coagulation<sup>56</sup>. Il s'agit de particules de 50 nm à 1 µm de diamètre émises de manière spontanée par

---

<sup>47</sup> Macher, B. A., et al. (2007). *Mol Biosyst* **3**, 705–13.

<sup>48</sup> Wallin, E., et al. (1998). *Protein Sci* **7**, 1029–38.

<sup>49</sup> Josic, D., et al. (2007). *Proteomics* **7**, 3010–29.

<sup>50</sup> Speers, A. E., et al. (2007). *Chem Rev* **107**, 3687–714.

<sup>51</sup> Santoni, V., et al. (2000). *Electrophoresis* **21**, 1054–70.

<sup>52</sup> Rabilloud, T., et al. (2008). *Proteomics* **8**, 3965–73.

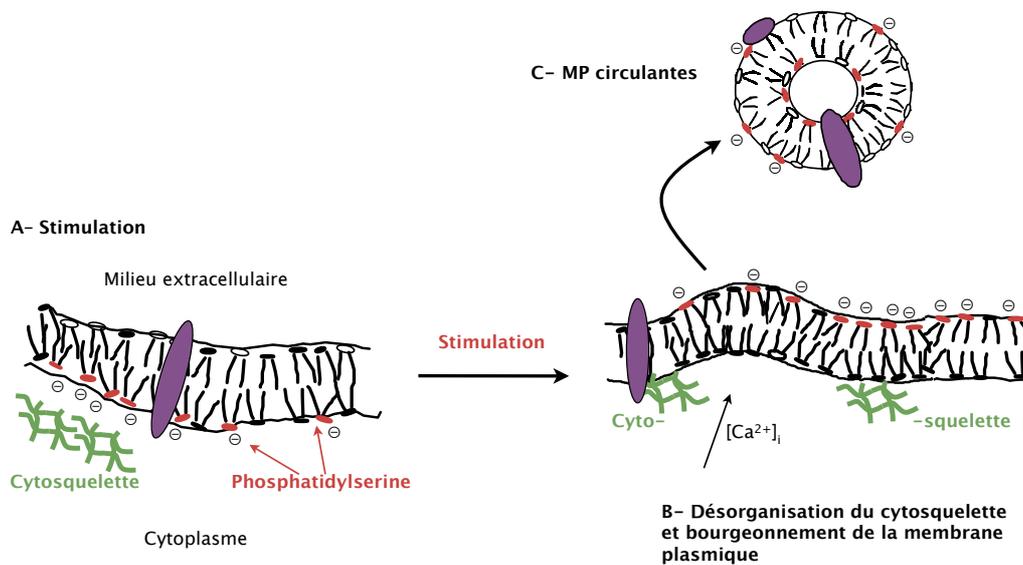
<sup>53</sup> Krause, F. (2006). *Electrophoresis* **27**, 2759–81.

<sup>54</sup> Wu, C. C., et al. (2003). *Nat Biotechnol* **21**, 532–8.

<sup>55</sup> Chargaff, E., et al. (1946). *Journal of Biological Chemistry* **166**, 189–197.

<sup>56</sup> Freyssinet, J. M. (2003). *J Thromb Haemost* **1**, 1655–62.

des cellules stimulées ou en apoptose (mort cellulaire, figure 8). Plus grosses que les exosomes, les microparticules contiennent des protéines issues des cellules dont elles sont issues<sup>57</sup> et – étant dérivées directement de leur membrane plasmique – sont naturellement enrichies en protéines membranaires<sup>58</sup>.



**Figure 8** | Production de microparticules. Un stimulus implique une augmentation du calcium cytosolique qui induit une exposition des phosphatidylsérines à la surface externe de la membrane (A). La membrane plasmique « bourgeonne » sous l'effet du déséquilibre phospholipidique (B). La désorganisation du cytosquelette sous l'effet du calcium permet alors la libération de microparticules (C) contenant des protéines membranaires et cytoplasmiques. (D'après Miguet et al, 2006).

Alors que peut d'études protéomiques traitent des microparticules<sup>59</sup>, le laboratoire a récemment entrepris de tirer parti du fait que les cellules sont capables de produire des microparticules en grande quantité sous l'action de substances chimiques telles que des antibiotiques comme l'actinomycine D1 ou stimulant la division cellulaire comme la phytohémagglutinine (PHA)<sup>60</sup>. La production de microparticules et leur isolation par centrifugation lente (15 000 à 30 000 g) sont simples et reproductibles. De plus, l'analyse de microparticules issues de lymphocytes pathogènes par spectrométrie de masse nanoLC-MS/MS, après séparation de l'extrait sur gel 1-D, a permis l'identification de 390 protéines dont 131 (34 %) appartenaient à la membrane plasmique. Par comparaison, les techniques classiques pour les préparations membranaires (gradient de sucrose, 100 000 g) ont conduit sur la même lignée cellulaire à l'identification de 300 protéines dont 24 % seulement étaient issues de la membrane plasmique.

Ces résultats ont montré la pertinence de l'outil microparticules pour l'étude de protéines membranaires. Notamment, leur utilisation dans l'analyse différentielle de cellules pathologiques ou tumorales permet la recherche de marqueurs membranaires spécifiques. Cette question a été traitée dans le cas de leucémies lymphoïdes chroniques dans la 1<sup>re</sup> partie des résultats au chapitre II.

<sup>57</sup> Morel, O., et al. (2004). *Curr Opin Hematol* **11**, 156–64.

<sup>58</sup> Miguet, L., et al. (2007). *Subcell Biochem* **43**, 21–34.

<sup>59</sup> Josic, D., et al. (2007). *Proteomics* **7**, 3010–29.

<sup>60</sup> Miguet, L., et al. (2006). *Proteomics* **6**, 153–71.

## 2.4. Conclusion

Les stratégies d'analyse protéomique par spectrométrie de masses ne sont pas figées : il n'existe pas de protocole unique permettant de résoudre toutes les questions biologiques posées dans une étude. Au contraire, il est généralement nécessaire de considérer chacune des questions successivement et d'adapter l'approche protéomique à mettre en oeuvre.

S'agit-il d'identifier, de caractériser, de quantifier le contenu protéique d'un échantillon ? Cet échantillon est-il « simple » ou bien issu d'un extrait protéique total ? Vise-t-on à étudier des protéines difficilement accessibles (comme les protéines membranaires) ou des partenaires de cibles déterminées (analyse fonctionnelle) ? De toutes ces questions dépendent les choix d'outils et de méthodes à associer pour tenter de résoudre le problème biologique posé (Figure 9).

Le panel des techniques et des approches protéomiques par spectrométrie de masse s'élargit constamment. Si certaines voies sont bien établies, de nouvelles applications, de nouveaux outils et de nouvelles méthodologies sont toujours mis au point pour dépasser les limites des approches existantes.

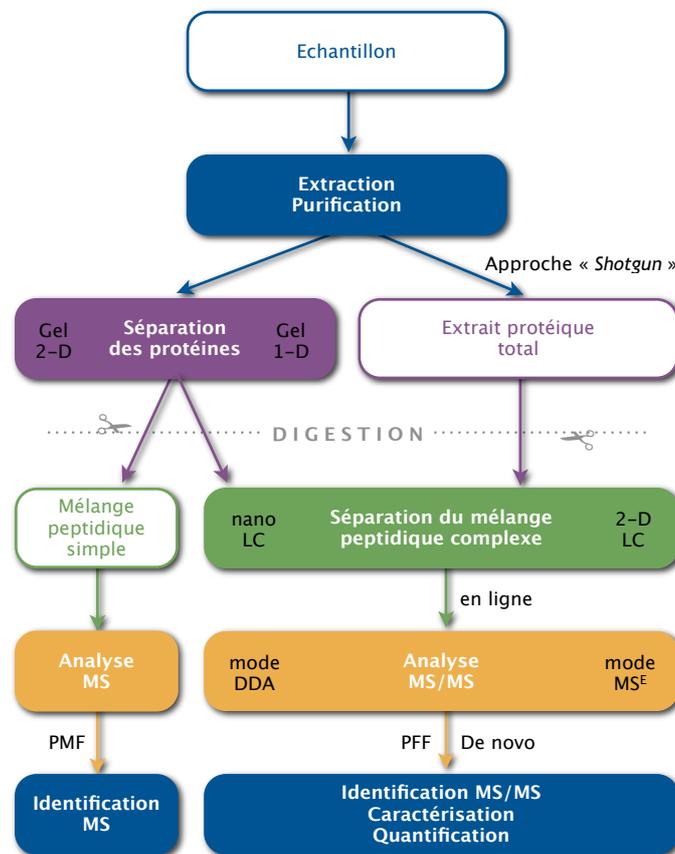


Figure 9 | Stratégies « bottom-up » pour l'identification et la caractérisation de protéines par spectrométrie de masse.



---

## Chapitre III

# Les données générées en analyse protéomique : gestion, validation et partage de l'information

---

En 1997, la « bioinformatique » apparaît comme un nouveau domaine de recherche : elle correspond à la mise au point et à l'utilisation d'outils informatiques permettant de manipuler, gérer, comparer, valider, partager des données biologiques afin de développer de nouvelles connaissances dans le domaine de la biologie elle-même<sup>1</sup>. Aujourd'hui, de nombreux outils bioinformatiques pour la protéomique sont disponibles et recouvrent un large champ d'application<sup>2,3</sup>.

L'identification de protéines en analyse protéomique par spectrométrie de masse repose essentiellement sur la comparaison de données mesurées (par MS ou MS/MS) et de données issues de banques protéiques par des moteurs de recherche adaptés. Les banques de données et les moteurs de recherches jouent donc un rôle essentiel dans le processus d'identification. Ce chapitre expose les différents choix disponibles pour le traitement bioinformatique des résultats d'analyse. Puis, il présente les approches couramment mises en oeuvre pour la validation des identifications. L'accent est notamment mis sur les solutions permettant de répondre aux exigences récentes d'évaluation des résultats. Enfin, la présentation des difficultés liées au partage de l'information protéique est l'occasion faire le point sur les règles de publication en cours de préparation.

---

<sup>1</sup> Andrade, M. A., et al. (1997). *Curr Opin Biotechnol* **8**, 675–83.

<sup>2</sup> Palagi, P. M., et al. (2006). *Proteomics* **6**, 5435–44.

<sup>3</sup> Lisacek, F., et al. (2006). *Proteomics* **6**, 5445–66.

## 1. Les banques de données de séquence

Les identifications de protéines et les interprétations biologiques qui en découlent dépendent directement de la disponibilité de banques de séquences et de la qualité de leurs annotations. Le développement des banques de données protéiques a accompagné l'essor de l'analyse protéomique par spectrométrie de masse. En effet, la disponibilité croissante de banques de séquences nucléotidiques (issues des projets de séquençage des génomes) et de banques de séquences protéiques a permis l'émergence des stratégies d'identification automatique de protéines (décrites au chapitre précédent).

### 1.1. Les banques de séquences nucléotidiques

Au niveau mondial, les données publiques de séquençage des génomes sont disponibles dans 3 banques étroitement liées. Les objectifs de cette collaboration internationale formée en 1988 sont de collecter, gérer et distribuer l'information représentée par toutes les séquences nucléotidiques déterminées expérimentalement. Les données sont rassemblées quelque soit leur origine, quelque soit l'organisme ou l'espèce moléculaire (ADN, ARN ...) dans :

▸ la **banque EMBL**<sup>4</sup> (« *European Molecular Biology Laboratory* »), créée en 1980, est gérée par EBI (« *European Bioinformatics Institute* ») ;

▸ la **banque GenBank**<sup>5</sup>, créée en 1982, est gérée par NCBI (« *National Center for Biotechnology Information* ») ;

▸ la **banque DDBJ**<sup>6</sup> (« *DNA Data Bank of Japan* »), est gérée par l'Institut National de Génétique du Japon qui a rejoint les deux autres organismes en 1986.

La contrepartie de tendre à constituer des archives exhaustives est de référencer des données redondantes et hétérogènes du fait de la diversité des sources de qualité variable.

### 1.2. Les banques de séquences protéiques

De nombreuses banques protéiques sont aujourd'hui disponibles<sup>7</sup>. Chacune est caractérisée par son exhaustivité, son degré de redondance, la qualité de ses annotations. Notamment, certaines banques sont vérifiées, corrigées, implémentées manuellement alors que d'autres ne sont constituées que de séquences déduites informatiquement (Figure 1). Deux types de banques sont ainsi distinguées :

#### 1.2.1. Les banques de séquences « corrigées »

Ces banques possèdent une forte valeur ajoutée en fonction de la qualité du travail de tri des entrées de la banque, de leur documentation et de leur annotation qui sont réalisés par des experts. De nombreuses banques spécialisées pour certaines familles de protéines ou certains organismes en font partie. Mais la plus réputée et la plus utilisée est la **banque UniProtKB** du consortium « *Universal Protein Resource* » : UniProt<sup>8</sup>.

Le consortium UniProt est né en 2003 du rapprochement de SIB (« *Swiss Institute of Bioinformatics* »), EBI (« *European Bioinformatics Institute* ») et PIR (« *Protein Information Resource* »). Il a pour mission de maintenir

<sup>4</sup> Cochrane, G., et al. (2009). *Nucleic Acids Res* 37, D19–25.

<sup>5</sup> Benson, D. A., et al. (2009). *Nucleic Acids Res* 37, D26–31.

<sup>6</sup> Sugawara, H., et al. (2009). *Nucleic Acids Res* 37, D16–8.

<sup>7</sup> Apweiler, R., et al. (2004). *Curr Opin Chem Biol* 8, 76–80.

<sup>8</sup> TheUniProtConsortium (2009). *Nucleic Acids Res*.

---

des banques de données stables, exhaustives, complètement référencées, largement et précisément annotées, et proposant de nombreuses références croisées. UniProt est constitué de 3 composantes majeures, chacune étant destinée à des utilisations précises :

▸ La **banque UniProtKB** (« *UniProt Knowledgebase* ») est la banque de référence, finement corrigée par des experts. Elle est constituée de 2 sous-banques : UniProtKB/SwissProt et UniProtKB/TrEMBL.

La première contient des séquences annotées manuellement avec des informations extraites de la littérature et corrigées à l'aide d'outils d'analyse informatiques. Notamment, un soin tout particulier est pris pour indiquer les fonctions et domaines biologiquement pertinents, les modifications posttraductionnelles et modifications connues, les localisations subcellulaires, la structure, les interactions possibles des protéines. En outre, les isoformes d'une même protéine sont rassemblées sous une séquence référence et sont indexées lorsque cela est possible.

La partie UniProtKB/TrEMBL (« *Translation from EMBL* ») contient les séquences protéiques traduites de toutes les séquences codantes des banques nucléotidiques EMBL/GenBank/DDBJ à l'exception de celles déjà présentes dans UniProtKB/SwissProt. Les séquences redondantes sont par ailleurs combinées. Des scripts de pré-annotation automatique des entrées permettent d'intégrer des informations primitives qui aideront à l'annotation manuelle avant l'indexation des séquences dans UniProtKB/SwissProt. La banque TrEMBL a été créée en complément de SwissProt pour permettre de suivre l'accroissement rapide de la taille des banques nucléotidiques (traduites automatiquement) sans compromettre la qualité et homogénéité de la banque SwissProt à très haute valeur ajoutée, mais plus longue et délicate à enrichir.

▸ La **banque UniParc** (« *Uniprot Archive* ») est une vaste banque exhaustive reflétant l'historique complet de toutes les séquences protéiques. Toutes les séquences ayant été référencées à un moment ou à un autre dans UniProtKB sont archivées dans cette banque. Elle constitue un élément important puisqu'une séquence résultant d'une annotation automatique est susceptible d'être retirée après une mise à jour des algorithmes de prédiction ou lorsque de nouvelles données permettent de préciser la longueur ou la composition de la séquence.

▸ La **banque UniRef** (« *UniProt Reference Clusters* ») fournit une collection de données de séquences non redondantes, avec une certaine « résolution ». C'est-à-dire que les séquences présentant 100 %, 90 % ou 50 % d'identité (y compris entre plusieurs espèces) sont regroupées dans une seule et unique entrée. Ces ressources sont nettement moins volumineuses que les banques entières, ce qui permet d'accélérer les recherches. La perte d'annotation pour chaque cluster est minimisée en leur attribuant les annotations les plus représentatives parmi ses membres.

### 1.2.2. Les banques généralistes de dépôt (« *sequence repositories* »)

Ces banques contiennent principalement des séquences protéiques directement issues de la traduction de gènes prédits *in silico* par des algorithmes. Ces séquences ne sont pas vérifiées et ne possèdent que des annotations très restreintes.

▸ La **banque GenPept** (« *GenBank Gene Products Data Bank* ») est produite par NCBI et regroupe l'ensemble des protéines traduites à partir des banques nucléotidiques GenBank/EMBL/DDBJ, sans annotation et avec de multiples redondances. Elle est donc peu utilisée en analyse protéomique par spectrométrie de masse.

▸ La **banque NCBI nr** (« *NCBI non redondante* ») est essentiellement une banque hybride rassemblant les séquences issues de plusieurs sources (notamment UniProtKB/SwissProt, PIR, PRF, PDB, GenBank, RefSeq ...). Le suffixe « non redondante » signifie que les séquences identiques issues de sources différentes sont fusionnées en une entrée unique. Cependant, les séquences différant en longueur (fragments ...) ou en composition (même de quelques acides aminés) constituent des entrées différentes. Cela peut se révéler utile lorsque des variations mineures sont recherchées, mais la répétition de nombreuses parties de séquences

résulte en la présence d'un grand nombre de peptides (par exemple tryptiques) identiques. La banque NCBI nr est donc très redondante au niveau peptidique et dispose d'annotations de qualité très variable selon la source. Elle peut néanmoins être considérée comme plus exhaustive que les autres banques disponibles.

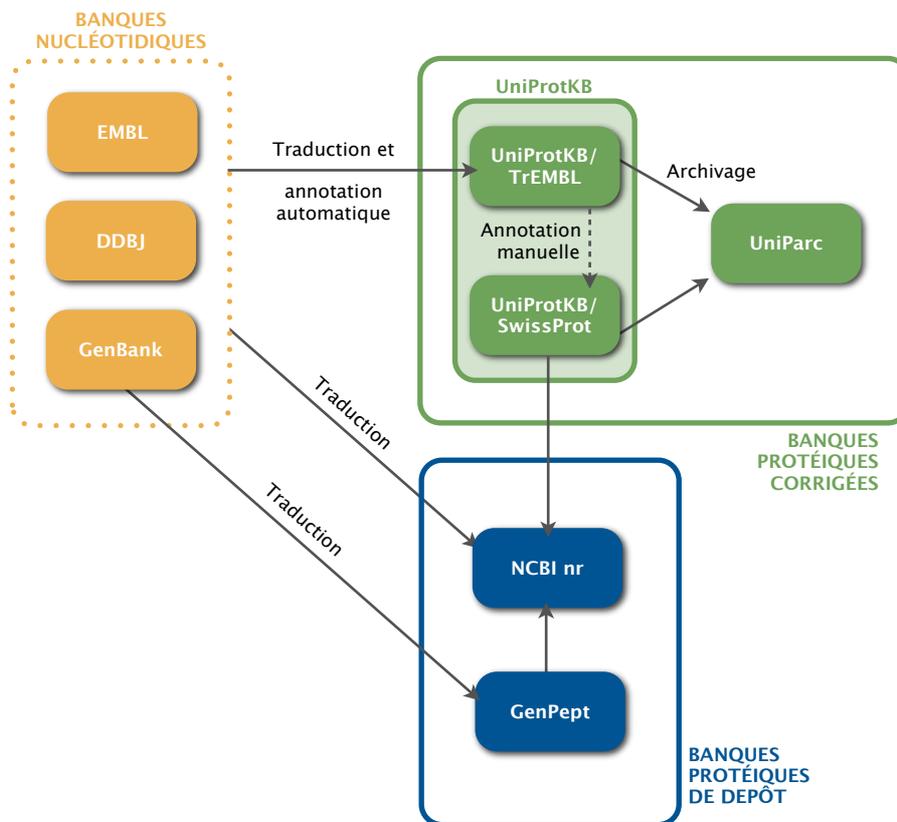


Figure 1 | Schéma de l'organisation des principales banques de données de séquences

### 1.2.3. La croissance exponentielle des banques

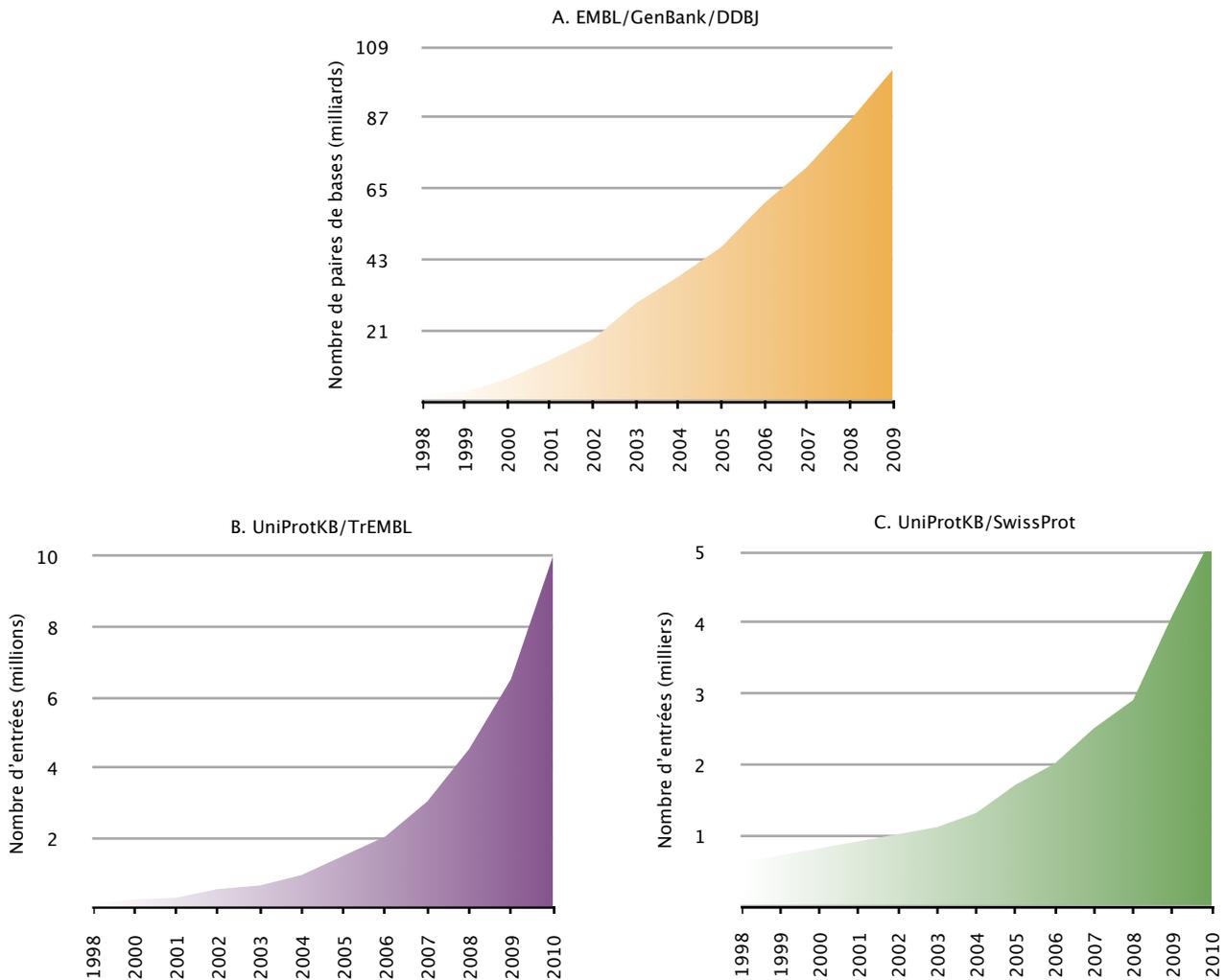
En 2009, la banque de séquences nucléotidiques EMBL/GenBank/DDBJ contient plus de 108 milliards de bases appartenant aux séquences ADN de plus de 300 000 espèces différentes. La banque croît de manière exponentielle et continue sa progression en doublant en 30 mois (Figure 2). Le nombre de génomes complètement séquencés s'élève à 170 pour les organismes microbiens et près de 270 génomes eucaryotes pour lesquels une couverture et des assemblages significatifs sont disponibles<sup>9</sup>.

Avec cette croissance exponentielle, il n'est pas possible de concilier homogénéité, non-redondance, exhaustivité et qualité d'annotation des banques. Outre les ressources informatiques nécessaires, les investissements humain et financier nécessaires au maintien de banques à forte valeur ajoutée sont colossaux. Ainsi, en complément des banques généralistes qui tendent à l'exhaustivité, on voit se développer de plus en plus de petites banques spécialisées.

Pour les mêmes raisons à l'échelle du laboratoire, le maniement et l'utilisation des banques deviennent de plus en plus difficiles. Récemment, le besoin de programmes permettant de créer et de manipuler des sous-banques à taxonomie réduite s'est fait sentir<sup>10</sup> (notamment pour la mise en oeuvre des stratégies de validation « *Target-Decoy* », détaillées au point suivant). Plusieurs outils informatiques sont en cours de réalisation : ils permettront le téléchargement, l'édition et la gestion de banques adaptées à chaque projet de recherche.

<sup>9</sup> Benson, D. A., et al. (2009). *Nucleic Acids Res* 37, D26-31.

<sup>10</sup> Reisinger, F., et al. (2009). *Proteomics* 9, 4421-4.



**Figure 2** | Croissance exponentielle de la taille des banques nucléotidiques EMBL/GenBank/DDBJ (A) et des banques protéiques UniProtKB/TrEMBL (B) et UniProtKB/SwissProt (C). (Sources : <http://www.uniprot.org> et <http://www.ddbj.nig.ac.jp>).

### 1.3. Les annotations « *Gene Ontology* »

La qualité de l'annotation des entrées des banques génomiques et protéiques est essentielle pour valoriser directement les résultats d'analyse. Un défi supplémentaire réside dans l'organisation formelle de ces connaissances biologiques de sorte qu'elles puissent être liées entre elles de manière homogène et exploitées aisément dans les banques de données. L'organisation de données biologiques sous forme d'ontologies (« *Gene Ontology* », GO) est une solution en plein essor<sup>11</sup>. Elle permet de construire des représentations formalisées de données sous forme de graphes acycliques orientés (Figure 3)<sup>12</sup> : les termes clés sont liés entre eux par des règles décrivant leurs relations, chaque terme étant lié à un ou plusieurs parents. Les termes indexés peuvent ensuite être reliés aux entrées des banques de séquences : on parle de « *Gene Ontology annotation* ».

Les ontologies diffèrent des annotations classiques dans la mesure où elles permettent de formaliser les données et de faciliter la recherche d'informations ciblées ou globales grâce à des connecteurs logiques. Dans le cas des « *Gene Ontologies* », trois hiérarchies indépendantes sont utilisées pour décrire les **processus biologiques**, les **fonctions moléculaires** et les **compartiments cellulaires** des protéines. Ainsi, il est par

<sup>11</sup> Ashburner, M., et al. (2000). Nat Genet 25, 25–9.

<sup>12</sup> Bard, J. B., et al. (2004). Nat Rev Genet 5, 213–22.

exemple possible de recherche directement à l'aide de scripts adaptés<sup>13</sup>, quelles sont parmi une liste de protéines identifiées celles appartenant à tel ou tel compartiment cellulaire (Figure 3).

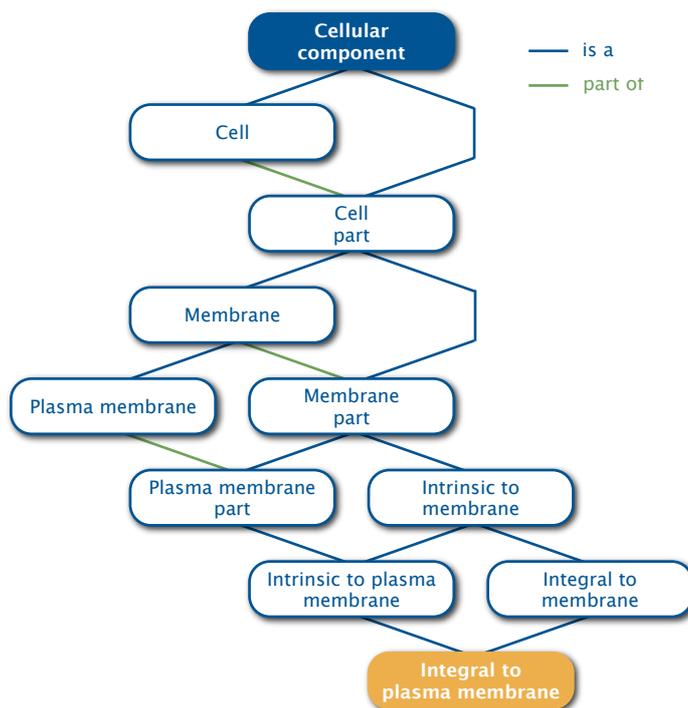


Figure 3 | Exemple du noeud « *Integral to plasma membrane* » dans l'ontologie « *Cellular component* » de la base « *EBI GO Annotation* », représenté sous forme de graphe acyclique orienté.

Les annotations « *Gene Ontology* » représentent un outil qui gagne en intérêt à mesure que la quantité de données indexées croît<sup>14</sup>. Afin de fournir une forte valeur ajoutée aux banques, les experts chargés de l'annotation (notamment de la banque UniProtKB/SwissProt) consultent systématiquement la littérature pour chaque protéine référencée et contribuent directement au travail du consortium GO<sup>15</sup>. En 2009, sur les 20 334 protéines contenues dans la banque UniProtKB/SwissProt *Homo Sapiens*, 18 587 protéines se partagent 165 769 annotations. Parmi ces protéines, 11 890 sont annotées manuellement et les 69 076 annotations correspondantes sont justifiées par 16 886 articles de la littérature (source : EBI GOA Human version 78 du 08.10.09). Ce travail d'annotation colossal représente un des défis majeurs pour la valorisation des résultats issus de l'analyse protéomique permettant une meilleure caractérisation des protéines au niveau moléculaire dans un contexte physiologique.

<sup>13</sup> Dimmer, E. C., et al. (2008). Proteomics.

<sup>14</sup> Rhee, S. Y., et al. (2008). Nat Rev Genet 9, 509–15.

<sup>15</sup> Camon, E., et al. (2003). Genome Res 13, 662–72.

---

## 2. La validation des résultats d'identification

### 2.1. Les différents moteurs de recherche et leurs spécificités

Les algorithmes d'identification de protéines les plus utilisés sont les algorithmes de recherche dans les banques de données. Ils visent à comparer des spectres expérimentaux (sous forme de listes de masse ou « *peak lists* ») à des spectres théoriques générés à partir des séquences protéiques contenues dans les banques. Il existe deux types d'algorithmes principaux capables de traiter les spectres de fragmentation de peptides acquis lors d'analyses MS/MS<sup>16</sup> (approche PFF, cf. chapitre II de cette partie) :

#### 2.1.1. Les algorithmes de recherche heuristiques

Ils mettent en corrélation les spectres de fragmentation expérimentaux et théoriques et calculent un score selon leur degré de similarité. De manière simplifiée, ces algorithmes « comptent » simplement le nombre de pics communs aux deux spectres.

▸ Le moteur **SEQUEST**<sup>17</sup> est le premier algorithme pour l'interprétation directe de spectres MS/MS à avoir été publié. Sous licence de la société Thermo, il représente le moteur de recherche heuristique le plus répandu et a été spécialement conçu pour une bonne sensibilité d'identification (c.-à-d., une capacité à identifier des peptides même à partir de spectre de qualité moyenne).

Lors d'une recherche, SEQUEST utilise un premier algorithme attribuant un score préliminaire (Sp) qui permet de sélectionner les 500 meilleures séquences peptidiques candidates. La corrélation directe de leur spectre théorique avec les données expérimentales permet le calcul de scores de « *Cross-Correlation* » (XCORR). La séquence candidate dont le spectre théorique présente le meilleur score de corrélation est considérée comme la meilleure identification pour le spectre expérimental.

▸ D'autres moteurs comme **Spectrum Mill** (Agilent Tech.) ou **X!Tandem**<sup>18</sup> (libre) fonctionnent sur ce même principe. D'une manière générale, tous les moteurs de recherche heuristiques tendent à intégrer, dans leurs versions les plus récentes, un niveau supplémentaire d'évaluation probabiliste des scores.

#### 2.1.2. Les algorithmes de recherche probabilistes

Ils calculent un score selon la probabilité que le spectre théorique d'un peptide donné corresponde par hasard au spectre expérimental considéré.

▸ Le moteur **MASCOT**<sup>19</sup> (Matrix Sciences) est considéré comme l'algorithme de référence, essentiellement en raison de sa compatibilité avec les données générées sur tous les modèles d'instruments, quel que soit le format des données. En outre, une version en ligne permet d'utiliser le moteur librement pour de petits lots de données.

Le modèle de calcul de score de MASCOT n'est pas publié. Toutefois, on sait qu'il permet de discriminer de manière itérative la meilleure séquence théorique pour chaque spectre de fragmentation en considérant notamment l'intensité des ions fragments expérimentaux attribués, la taille de la séquence attribuée, le nombre de fragments attribués ou non, le nombre d'ions attribués consécutifs d'une même série de fragmentation, et l'erreur de masse sur la masse des fragments attribués. A chaque peptide identifié,

---

<sup>16</sup> Kapp, E., et al. (2007). *Curr Protoc Protein Sci* **Chapter 25**, Unit25 2.

<sup>17</sup> Eng, J. K., et al. (1994). *Journal of the American Society for Mass Spectrometry* **5**, 976–989.

<sup>18</sup> Craig, R., et al. (2004). *Bioinformatics* **20**, 1466–7.

<sup>19</sup> Perkins, D. N., et al. (1999). *Electrophoresis* **20**, 3551–67.

MASCOT attribue un « score d'ion » représentatif de la probabilité que les fragments aient pu être attribués par hasard ; et un « score identité » qui dépend de la taille de l'espace de recherche. Un « score d'ion » supérieur au « score identité » signifie que la probabilité pour que l'identification soit le résultat d'un phénomène aléatoire est inférieure à 5 %. Pour rendre compte de cette probabilité, la différence entre le « score d'ion » et le « score identité » est généralement considérée.

▶ Le moteur **OMSSA**<sup>20</sup> (« *Open Mass Spectrometry Search Algorithm* ») est un algorithme développé initialement par NCBI dans l'objectif de proposer un nouveau moteur de recherche rapide et libre de droits.

Un modèle mathématique explicite est utilisé pour l'attribution de scores aux peptides identifiés. Il s'appuie sur l'hypothèse que pour tout peptide théorique ayant la masse du peptide parent mesuré, le nombre d'ions fragments théoriques correspondants aux ions fragments d'un spectre expérimental varie selon une loi de Poisson en fonction du nombre de fragments considérés. La moyenne de la distribution dépend de la tolérance sur la masse des ions fragments fixée, du nombre de pics théoriques et expérimentaux, et de la masse du peptide parent considéré. OMSSA calcule alors une « *E-value* » représentant la probabilité pour que l'association des spectres expérimental et théorique soit due au hasard. Pour des raisons pratiques, le score utilisé est généralement de la forme  $-\log_{10}(E\text{-value})$ .

▶ Il existe par ailleurs de nombreux autres moteurs de recherche fondés sur l'un ou l'autre des principes (empirique ou statistique) de calcul de score probabilistes présentés ci-dessus. On peut notamment citer : **Phenyx**<sup>21</sup> (Genebio) et **ProteinLynx** (Waters Corp.).

### 2.1.3. Les limites des moteurs de recherche et le besoin de valider les résultats.

#### ▶ Les pistes de développement pour les algorithmes de recherche

Si les moteurs de recherche sont de plus en plus rapides et performants<sup>22</sup>, ils ne sont pas parfaits. Les algorithmes doivent par exemple être constamment améliorés pour affiner les calculs de score et l'attribution des séquences théoriques en tenant compte notamment de la qualité des spectres de fragmentation<sup>23</sup>. Les objectifs de développement visent également à permettre l'identification de nouvelles modifications posttraductionnelles (ponts disulfures)<sup>24</sup>, à accélérer les recherches dans les banques de taille importante ou à partir de données expérimentales toujours plus volumineuses. Enfin, la prise en compte des nouvelles stratégies (MS<sup>F</sup>) et technologies d'analyses (fragmentation ETD) et la combinaison de plusieurs moteurs de recherche<sup>25</sup> pour l'identification sont progressivement disponibles.

#### ▶ Les « faux positifs » : un problème majeur pour l'identification fiable des protéines

Un écueil majeur pour l'identification de protéines via les moteurs de recherche dans les banques de données réside dans la mise en évidence des faux positifs identifiés et dans l'évaluation de leur nombre (Figure 4). L'identification de faux positifs peut être liée à plusieurs phénomènes<sup>26</sup> :

▶ Les moteurs sont incapables de faire la différence entre le spectre d'un peptide réel et celui ne contenant que du bruit de fond ou des contaminants, celui de petites molécules organiques, ou de mélanges de peptides isobares coélus.

<sup>20</sup> Geer, L. Y., et al. (2004). *J Proteome Res* **3**, 958–64.

<sup>21</sup> Colinge, J., et al. (2004). *Proteomics* **4**, 1977–84.

<sup>22</sup> Kapp, E., et al. (2007). *Curr Protoc Protein Sci* **Chapter 25**, Unit25 2.

<sup>23</sup> Salmi, J., et al. (2009). *Proteomics* **9**, 848–60.

<sup>24</sup> Xu, H., et al. (2008). *J Proteome Res* **7**, 138–44.

<sup>25</sup> Alves, G., et al. (2008). *J Proteome Res* **7**, 3102–13.

<sup>26</sup> Sadygov, R. G., et al. (2004). *Nat Methods* **1**, 195–202.

▸ Les peptides sont souvent présents sur de larges gammes de concentration dans un échantillon. Les peptides dont la concentration avoisine la limite de détection peuvent produire des spectres de fragmentation de mauvaise qualité.

▸ La fragmentation des peptides (qui reste dans certains cas incomprise) peut suivre un mécanisme aberrant et conduire à des spectres inattendus.

▸ Les peptides présentant des modifications posttraductionnelles qui ne sont pas prises en compte dans les paramètres de recherche génèrent des spectres de fragmentation que le moteur ne peut attribuer correctement.

▸ Les digestions enzymatiques réalisées avant l'analyse peuvent être incomplètes et les peptides qui en résultent ne sont pas identifiables directement si les paramètres de recherche sont restrictifs (peptides semi-trypsiques...)

▸ Enfin, les algorithmes de recherche ne sont en mesure que d'identifier des peptides dont la séquence est présente dans la banque protéique utilisée. Les spectres de fragmentation de ces peptides représentent alors de potentiels faux positifs.

		Le spectre MS/MS est celui d'un peptide présent dans la banque protéique	
		VRAI	FAUX
La recherche conduit à l'identification correcte de la séquence	VRAI	Vrai positif	Faux positif
	FAUX	Faux négatif	Vrai négatif

**Figure 4** | Définition des termes vrai/faux positif/négatif dans le cas de l'identification de spectres de fragmentation de peptides par un moteur de recherche dans les banques de données protéiques.

▸ Le besoin de valider les résultats d'identification de protéines

Les scores ou les probabilités calculés par les différents moteurs de recherches ne constituent en aucun cas des indicateurs fiables pour considérer la pertinence des identifications (vrais positifs)<sup>27</sup>. Une étape supplémentaire déterminante doit permettre de « vérifier » les identifications afin d'évaluer les faux positifs parmi les résultats.

Les premières approches visant à « valider » les résultats s'appuyaient sur une inspection manuelle des identifications. Mais devant la nécessité de plus en plus pressante d'estimer empiriquement la confiance dans les identifications et face aux projets d'analyse à haut débit mis en oeuvre, des stratégies d'évaluation de taux de faux positifs sont apparues au milieu des années 2000. Le principe de ces approches manuelles et statistiques est présenté dans les 2 points ci-dessous, mais leur évaluation et leur mise au point sont détaillées dans la 1<sup>re</sup> partie des résultats, au chapitre 1, point 3.2.

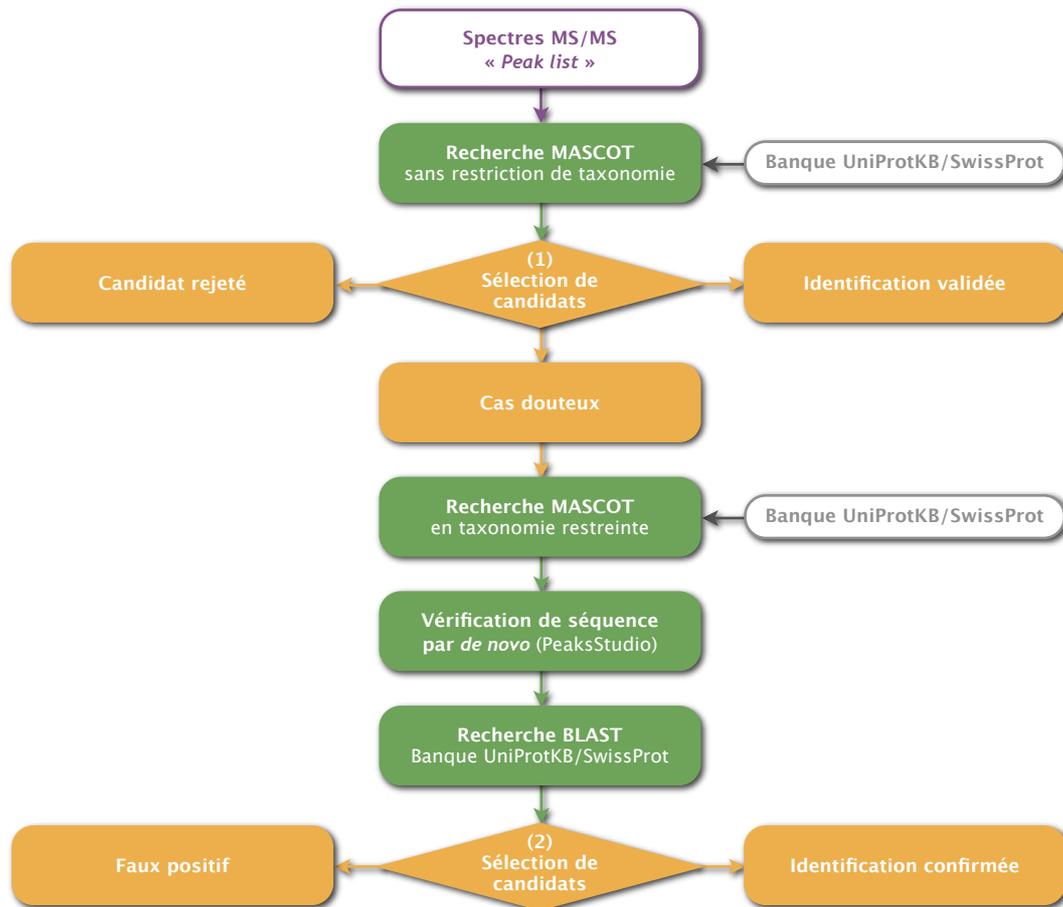
## 2.2. La stratégie de vérification manuelle des identifications et ses limites

Les analyses LC-MS/MS conduisent généralement à un grand nombre de résultats d'identification qui présentent des scores très disparates. Alors qu'une identification réalisée avec plusieurs peptides aux spectres de fragmentation non ambigus est généralement indiscutable, on rencontre également un certain

<sup>27</sup> Nesvizhskii, A. I., et al. (2007). Nat Methods 4, 787-97.

nombre de protéines dont l'identification est plus contestable<sup>28</sup>. Ainsi, si une interprétation exhaustive des résultats est attendue, une inspection plus poussée et au cas par cas des identifications ambiguës est nécessaire.

La figure 5 présente la stratégie de validation manuelle qui était appliquée au laboratoire pour la validation des identifications. De manière similaire, des approches de vérification des cas douteux par l'application d'un deuxième critère empirique ont été décrites dans la littérature<sup>29,30</sup>. La stratégie présentée s'appuie sur l'interprétation *de novo* des spectres pour les identifications « pauvres » (réalisées avec 1 ou 2 peptides uniques). La spécificité pour la protéine des fractions de séquences validées manuellement ou avec l'assistance d'un logiciel comme Peaks Studio est alors vérifiée par une recherche BLAST. En fonction de ces résultats, la décision de valider ou non l'identification peut alors être arrêtée et justifiée.



**Figure 5** | Stratégie de validation « manuelle » des identifications. Les choix réalisés par l'opérateur aux niveaux (1) et (2) se conforment aux critères suivants. Sélection des candidats (1) : (a) Identification validée pour une protéine proposée avec au moins trois peptides dont les spectres ne sont pas ambigus. (b) Cas douteux lorsqu'une protéine est proposée avec un ou deux peptides dont les spectres ne sont pas ambigus. (c) Les autres candidats sont rejetés. Sélection des candidats (2) : (a') Identification confirmée pour une protéine dont la séquence d'au moins un des peptides proposés est vérifiée par séquençage *de novo* et est reconnue comme spécifique de la protéine par une recherche BLAST. (b') Les autres cas sont rejetés et considérés comme « faux positifs ».

De nombreux groupes ont mis au point des logiciels d'aide à la décision pour l'identification des protéines. Ils permettent d'extraire sélectivement l'information des résultats d'identification produits par les moteurs de recherche (c'est le cas notamment d'un script développé en interne au laboratoire) ou sont fondés sur le principe de la comparaison manuelle des spectres de fragmentation douteux à des banques de spectres déjà établis et validés<sup>31</sup>.

<sup>28</sup> Thomas, H., et al. (2008). *Proteomics* 8, 4173–7.

<sup>29</sup> Wielsch, N., et al. (2006). *J Proteome Res* 5, 2448–56.

<sup>30</sup> Hernandez, P., et al. (2006). *Mass Spectrom Rev* 25, 235–54.

<sup>31</sup> Pouillet, P., et al. (2007). *Proteomics* 7, 2553–6.

---

## 2.3. Les stratégies d'évaluation de taux de faux positifs

### 2.3.1. Signification des taux de faux positifs

Les stratégies d'évaluation des résultats en analyse protéomique sont progressivement passées de l'inspection manuelle et de la considération directe des scores à des approches probabilistes permettant la mesure statistique d'indices de confiance et l'estimation de taux d'erreurs. On rencontre dans la littérature de nombreuses confusions dans la terminologie notamment en ce qui concerne les notions de « *False Positive Rate* » (FPR) et de « *False Discovery Rate* » (FDR)<sup>32</sup>.

▸ Le **taux de faux positifs FPR** est la propriété d'un spectre MS/MS définie comme la probabilité qu'un spectre identifié par hasard soit jugé correctement attribué. Il représente le nombre de faux positifs rapporté à la somme des faux positifs et vrais négatifs.

▸ Le **taux de faux positifs FDR** est la propriété d'un ensemble de spectres MS/MS définie comme la proportion d'identifications incorrectes parmi toutes les identifications jugées correctes. Il représente le nombre de faux positifs rapporté à la somme des vrais et faux positifs. La définition est identique qu'on parle d'identification de peptides ou de protéines, mais les deux valeurs de FDR ne sont pas égales pour une même analyse.

Alors que le FPR est difficilement estimable (le nombre de vrais négatifs ne pouvant pas être estimé empiriquement), deux types d'approches ont été présentées pour estimer des FDR : les stratégies « *Target-Decoy* » et les approches empiriques de Bayes (Figure 6).

### 2.3.2. Les stratégies « *Target-Decoy* »

La stratégie « *Target-Decoy* »<sup>33</sup> ou « Cible-Leurre » repose sur la recherche des résultats d'identification par un moteur de recherche directement dans une banque protéique classique « cible / *target* » concaténée avec une banque « leurre / *decoy* » qui contient toutes les séquences inversées (ou mélangées ou randomisées) de la même banque. L'approche suppose que les identifications de peptides dans la banque « leurre » et les identifications de faux positifs dans la banque « cible » originale suivent la même distribution. Dans une deuxième étape, les identifications de peptides sont filtrées en ajustant les seuils de scores du moteur de recherche et le taux de faux positifs (FDR) associé peut être directement estimé en calculant le rapport du nombre de protéines identifiées dans la banque « leurre » ( $N_d$ ) par le nombre total de protéines identifiées ( $N$ ) en appliquant un seuil de score donné ( $FDR = 2 \cdot N_d / N$ ).

L'avantage majeur de la stratégie « *Target-Decoy* » est de permettre une estimation du taux de faux positifs (FDR) de manière directe, avec une mise en oeuvre simple. Elle s'applique dans de nombreuses situations que le volume de données soit très (« *shotgun proteomics* ») ou moyennement important (stratégies nanoLC-MS/MS après séparation sur « gel 1-D »). Par contre, le doublement de la taille des banques implique le doublement du temps de recherche par les algorithmes.

Les questions relatives à la mise en oeuvre de la stratégie « *Target-Decoy* » sont particulièrement détaillées dans la 1<sup>re</sup> partie des résultats, au chapitre 1, point 3.2.2.

---

<sup>32</sup> Nesvizhskii, A. I., et al. (2007). *Nat Methods* 4, 787-97.

<sup>33</sup> Elias, J. E., et al. (2007). *Nat Methods* 4, 207-14.

### 2.3.3. Les approches empiriques de Bayes (PeptideProphet)

Les approches empiriques de Bayes<sup>34,35</sup>, utilisées par le script PeptideProphet<sup>36</sup>, permettent de modéliser la distribution des scores attribués par l'algorithme de recherche à chaque peptide identifié. Le modèle représente le nombre de peptides identifiés en fonction de leur score probabiliste comme une courbe bimodale représentant les identifications correctes et incorrectes (figure 6, à droite). En utilisant les statistiques bayésiennes, PeptideProphet peut alors déduire de la modélisation la probabilité pour qu'une identification soit correcte en fonction du score qui lui est associé. De manière simplifiée, cela revient à calculer la contribution des identifications correctes (courbe bleue) à la distribution globale (somme des deux courbes bleue et rouge), pour chacun des scores. Les probabilités sont ensuite utilisées pour estimer le taux de faux positifs (FDR) associé à chacun des seuils de probabilité minimale considérés.

Dans une deuxième étape, l'évaluation des identifications incorrectes doit être transposée au niveau protéique. Grâce au script ProteinProphet, les probabilités d'identification correcte des peptides participant à l'identification d'une même protéine sont utilisées pour calculer la probabilité que l'identification de la protéine soit correcte. Le calcul prend notamment en compte le nombre de peptides participant à une même identification et réajuste éventuellement les probabilités associées à ces peptides.

Contrairement aux approches « *Target-Decoy* », PeptideProphet ne nécessite pas l'implémentation d'une banque leurre à l'espace de recherche. Cependant, la modélisation de la distribution des identifications ne peut être réalisée que si la quantité de données expérimentales (ie. le nombre de spectres acquis) est très importante. En pratique, cette approche n'est donc pas utilisable directement pour les petits lots de données et se limite aux expériences de « *shotgun proteomics* ».

### 2.3.4. Combiner les informations pour améliorer l'évaluation des résultats

Les deux approches « *Target-Decoy* » et PeptideProphet sont maintenant couramment utilisées en analyse protéomique pour l'évaluation des résultats. Toutefois, la mise en oeuvre de ces stratégies est parfois délicate, notamment pour de petits lots de données expérimentales et lorsque la modélisation n'est pas parfaite<sup>37</sup>. Pour écarter ces difficultés, les deux approches peuvent être combinées : les données sur les peptides issus d'identifications dans une banque leurre apportent une information supplémentaire explicite aux algorithmes PeptideProphet qui peuvent alors affiner le modèle de distribution.

Une autre possibilité pour améliorer la confiance dans les identifications de peptides et protéines via les moteurs de recherche est la prise en compte de données auxiliaires – notamment physico-chimiques – accessibles lors d'une analyse protéomique. C'est probablement un des développements les plus prometteurs, et certains algorithmes de recherche (ProteinLynx) ou de validation (PeptideProphet) utilisent déjà certaines de ces informations comme :

- ▶ la précision de masse mesurée sur les peptides parents lors de l'acquisition MS<sup>38</sup> ;
- ▶ le temps de rétention chromatographique<sup>39,40,41</sup> ;
- ▶ le pI des peptides et protéines identifiés<sup>42,43</sup> ;
- ▶ la présence de sites de coupure enzymatique manqués (« *missed cleavages* »).

<sup>34</sup> Bayes, M., et al. (1763). *Philosophical Transactions* **53**, 370–418.

<sup>35</sup> Storey, J. D., et al. (2003). *Proc Natl Acad Sci U S A* **100**, 9440–5.

<sup>36</sup> Keller, A., et al. (2002). *Anal Chem* **74**, 5383–92.

<sup>37</sup> Nesvizhskii, A. I., et al. (2007). *Nat Methods* **4**, 787–97.

<sup>38</sup> Scherl, A., et al. (2008). *Proteomics* **8**, 2791–7.

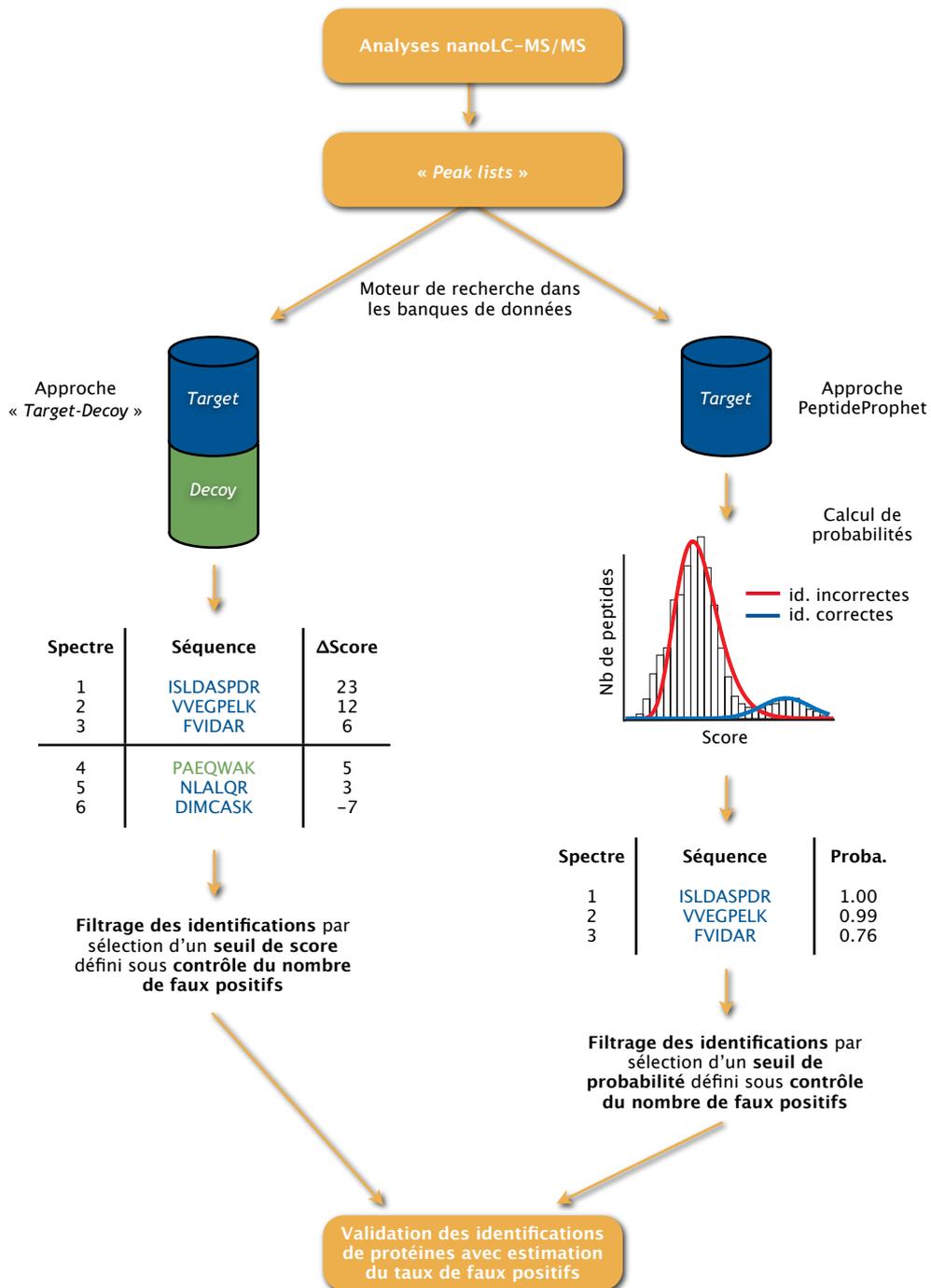
<sup>39</sup> Strittmatter, E. F., et al. (2004). *J Proteome Res* **3**, 760–9.

<sup>40</sup> Pfeifer, N., et al. (2009). *J Proteome Res* **8**, 4109–15.

<sup>41</sup> Baczek, T., et al. (2009). *Proteomics* **9**, 835–47.

<sup>42</sup> Cargile, B. J., et al. (2004). *J Proteome Res* **3**, 112–9.

<sup>43</sup> Malmstrom, J., et al. (2006). *J Proteome Res* **5**, 2241–9.



**Figure 6** | Validation des identifications par approche « Target-Decoy » (branche de gauche) et PeptideProphet (branche de droite). Dans la stratégie « Target-Decoy », les identifications sont réalisées dans une banque composite cible et leurre (de mêmes taille et distribution) et le nombre d'identifications réalisées dans la banque leurre est utilisé pour estimer le taux de faux positifs (FDR) associé à différents seuils de score. Dans une approche probabiliste de Bayse (PeptideProphet), la distribution entre les identifications correctes (courbe bleue) et incorrectes (courbe rouge) est modélisée et ajustée aux données expérimentales (histogramme). Une probabilité est calculée pour chaque peptide analysé et ces données sont utilisées pour estimer le taux de faux positifs (FDR). D'après Nesvizhskii (2007).

### 3. Les difficultés du partage de l'information protéomique

En 2009, dans un éditorial de *Journal of Proteome Research*<sup>44</sup> intitulé « *Ending the 'Publish and Vanish' Culture* », Sandra Orchard de l'EBI affirme que l'analyse protéomique a atteint un tel degré de maturité qu'il n'est maintenant plus concevable de ne pas rendre accessibles et exploitables les données expérimentales de sujets publiés. Sortir de l'habitude de « publier puis oublier » demande un effort sur deux plans *a priori* contradictoires. En effet, si la communauté des protéomistes souhaite bénéficier d'un accès aux données brutes, celle des biologistes – plus grande – est plutôt intéressée par des données déjà interprétées et aisément exploitables<sup>45</sup>. Afin de faciliter la diffusion de l'information protéomique, deux initiatives sont en plein développement :

- la formalisation et la clarification de directives de publication ;
- la construction de banques de dépôt de données protéomiques (« *proteomics MS data repositories* »)

alimentées par l'information protéomique générée par la communauté avec toute l'information nécessaire à son exploitation.

#### 3.1. Le minimum requis pour la publication de données protéomiques (MIAPE)

##### 3.1.1. Le besoin de directives pour la publication de données protéomiques

Les analyses protéomique par spectrométrie de masse génèrent, à partir d'un large panel de techniques et de stratégies originales, une énorme quantité de données brutes et de données interprétées. La publication dans les journaux scientifiques de ces informations soulève des questions cruciales quant à la manière de décrire les approches analytiques et les opérations de traitement des données mises en oeuvre, et quant à la manière de présenter les résultats<sup>46</sup>. Il est évident qu'il n'est ni possible ni pertinent de présenter l'intégralité des données générées. Il n'est plus question aujourd'hui de publier des « annuaires » entiers de protéines identifiées : la plupart sont généralement sans intérêt pour l'objet de l'étude (protéines majoritaires, aspécifiques...) <sup>47</sup>. Des directives de publications sont alors nécessaires pour inciter les auteurs :

- à présenter de manière exhaustive les paramètres expérimentaux et de traitement des données ;
- à présenter une évaluation de la fiabilité des résultats ;
- à présenter les données essentielles soutenant les résultats – notamment celles dont l'interprétation est délicate, de sorte que le lecteur puisse en évaluer lui même la pertinence.

##### 3.1.2. Vers l'unification et la standardisation des directives

En 2005, un rassemblement d'experts en analyse protéomique, de constructeurs d'instruments, de concepteurs de moteurs de recherche et de représentants des éditeurs a abouti à la rédaction d'un premier document : les « Directives de Paris »<sup>48</sup>. Parallèlement, les grands journaux spécialisés comme *Journal of Proteome Research*<sup>49</sup> et surtout *Molecular and Cellular Proteomics*<sup>50</sup> ont précisé leurs directives et ont annoncé la vérification de la conformité des articles avant toute relecture par les comités de critique.

<sup>44</sup> Orchard, S. (2009). *J Proteome Res* **8**, 3219.

<sup>45</sup> Eastman, Q. (2009). *J Proteome Res* **8**, 3220.

<sup>46</sup> Latterich, M. (2006). *Proteome Sci* **4**, 8.

<sup>47</sup> Hamacher, M., et al. (2008). *Proteomics* **8**, 4616–7.

<sup>48</sup> Bradshaw, R. A. (2005). *Mol Cell Proteomics* **4**, 1223–5.

<sup>49</sup> Wilkins, M. R., et al. (2006). *Proteomics* **6**, 4–8.

<sup>50</sup> Bradshaw, R. A., et al. (2006). *Mol Cell Proteomics* **5**, 787–8.

Enfin, dans l'objectif d'unifier les directives, de standardiser les informations recueillies, et de prendre en compte la diversité et la complexité des stratégies et des technologies mises en oeuvre en analyse protéomique, « HUPO – Proteomics Standards Initiative » a entrepris la rédaction de recommandations appelées MIAPE (« *Minimum Information About a Proteomics Experiment* »)<sup>51</sup>. Elles visent à inciter les auteurs à produire les informations essentielles permettant de contextualiser leurs résultats protéomiques (ie. qui doivent figurer dans la section « Matériels et Méthodes » d'un article). Les MIAPE consistent en un système documentaire de plusieurs modules adaptés à chaque technologie utilisée en analyse protéomique. Deux modules concernent particulièrement la spectrométrie de masse (MIAPE-MS)<sup>52</sup> et les outils informatiques qui lui sont associés (MIAPE-MSI)<sup>53</sup>. Les directives en elle mêmes sont divisées en catégories qui rassemblent les informations requises pour la publication de données protéomiques (Table 1).

MIAPE-MS	MIAPE-MSI
<p><b>1. Généralités</b> Informations sur l'instrument, les logiciels de pilotage et les paramètres utilisés.</p>	<p><b>1. Généralités</b> Moteur de recherche, logiciel de traitement des données ...</p>
<p><b>2. Source d'ionisation</b> Par exemple ESI ou MALDI et paramètres appliqués.</p>	<p><b>2. Données initiales et paramètres</b> Format des données, banque de données et paramètres de recherche utilisés.</p>
<p><b>3. Installations post-source</b> Type(s) d'analyseur(s), cellule de fragmentation, détecteur ...</p>	<p><b>3. Informations relatives aux identifications de protéines et peptides présentées</b> Numéro d'accension, description, recouvrement de séquence, masse et charge, nombre de peptides, scores, modifications, erreur de masse, données de quantification ...</p>
<p><b>4. Données générées</b> Information sur les spectres générés, paramètres de création des « <i>peak lists</i> », et éventuellement méthode de quantification utilisée.</p>	<p><b>4. Interprétation et validation</b> Méthode et seuils de validation des identifications, stratégie et statistiques permettant l'évaluation de taux de faux positifs ...</p>

Table 1 | Aperçu du contenu des MIAPE-MS et MIAPE-MSI.

## 3.2. Les projets de partage des données

### 3.2.1. L'objet des banques de dépôt protéomiques

En 2008, les conclusions du sommet d'Amsterdam<sup>54</sup> sur les politiques de diffusion et de partage des données protéomique soulignent qu'un des obstacles majeurs à l'accélération des découvertes en analyse protéomique réside dans le manque d'incitations pour la diffusion rapide de données protéomiques à grande échelle. Cela doit pouvoir être réalisé notamment en tenant compte d'un certain nombre de principes qui ont été énumérés :

- Les données protéomiques doivent être déposées dans les banques le plus rapidement possible. Dès leur validation lorsqu'elles ont été générées dans un contexte collectif et au moment de la publication pour les projets individuels afin de garantir la propriété intellectuelle aux auteurs.
- Le format des données publiées doit permettre de rendre compte de l'ensemble des données acquises de manière exhaustive et documentée (notamment quant à leur qualité).
- Le support des données doit permettre leur large accessibilité. Un format libre et une organisation formalisée sont donc préconisés.
- La procédure de dépôt doit être simple tout en encourageant l'annotation complète des données.

<sup>51</sup> Taylor, C. F., et al. (2007). Nat Biotechnol 25, 887–93.

<sup>52</sup> Taylor, C. F., et al. (2008). Nat Biotechnol 26, 860–1.

<sup>53</sup> Binz, P. A., et al. (2008). Nat Biotechnol 26, 862.

<sup>54</sup> Rodriguez, H., et al. (2009). J Proteome Res 8, 3689–92.

- Des mécanismes d'évaluation de la qualité des données doivent être mis en place pour permettre la comparaison des informations disponibles au sein même de la banque ou entre des banques différentes.
- La communauté scientifique, les agences de financement et les éditeurs partagent la responsabilité de définir des directives qui incitent au partage de l'information.

Le développement des banques de dépôt protéomiques publiques (« *public proteomics MS data repositories* ») doit permettre de répondre à ces exigences. Ces banques de dépôt, accessibles et exploitables librement, sont définies comme des banques de données brutes de spectrométrie de masse et/ou d'identification de protéines et peptides issues de divers laboratoires. Sous l'influence des éditeurs, le partage des données devrait rapidement devenir la norme dans les années à venir. Le journal « *Molecular and Cellular Proteomics* », par exemple, annonce dans une ébauche de ses directives pour 2010 qu'il imposera la soumission des données brutes dans des banques de dépôt avant toute publication<sup>55</sup>.

### 3.2.2. Les principales banques de dépôt protéomiques

Ces dernières années, plusieurs banques de dépôt protéomiques publiques se sont développées. Chacune présente des objectifs différents et offre donc des possibilités complémentaires<sup>56</sup>. Les trois banques principales sont présentées ci-dessous, mais d'autres – plus petites et spécialisées – se développent parallèlement.

- La banque **PRIDE**<sup>57</sup> (« *Proteomics IDentifications database* »), contrairement à ce qu'indique son nom, n'est pas limitée aux seuls résultats d'identification, mais inclut également des données de type « *peak list* ». En lien étroit avec HUPO-PSI, PRIDE s'attache à respecter les directives de publication MIAPE et à privilégier les formats libres et standardisés de données (comme mzML, cf. 1<sup>re</sup> partie des résultats, chapitre I, point 3.1.4.). La banque est notamment utilisée pour héberger des données issues de diverses approches MS/MS et relatives à des articles publiés.

- La banque **PeptideAtlas**<sup>58</sup> est sous le contrôle de « *Institute for Systems Biology* ». A ce titre, elle est conçue pour être une plateforme permettant la sélection et la validation de cibles par spectrométrie de masse. La banque s'organise autour de catégories d'espèces ou de type d'échantillons qui rassemblent les peptides identifiés à partir de toute expérience de protéomique par spectrométrie de masse en tandem. Les données déposées sont soumises à une stratégie d'évaluation globale en utilisant les algorithmes de recherche SEQUEST et X!Tandem dont les scores d'identification sont convertis en probabilités par les approches PeptideProphet et ProteinProphet.

- La banque **Tranche** (ProteomeCommons) a pour objectif direct de permettre le partage de très grosses quantités de données protéomiques non interprétées. Elle peut également avoir pour objet le stockage et la sauvegarde de données brutes en ligne. Comme Tranche n'a pas la vocation de mettre les données qu'elle contient en relation, ses outils de recherche sont assez limités.

<sup>55</sup> Cottingham, K. (2009). *J Proteome Res* **8**, 4887–8.

<sup>56</sup> Mead, J. A., et al. (2009). *Proteomics* **9**, 861–81.

<sup>57</sup> Jones, P., et al. (2008). *Nucleic Acids Res* **36**, D878–83.

<sup>58</sup> Deutsch, E. W., et al. (2008). *EMBO Rep* **9**, 429–34.

---

## 4. Conclusion

Avec la croissance exponentielle des banques protéiques et l'amélioration des techniques d'acquisition de données protéomiques par spectrométrie de masse en tandem, l'identification de protéines est devenue aussi efficace que problématique. Bien que de nombreux algorithmes de recherches dans les banques de données ont montré leurs capacités à identifier de grands nombres de peptides et protéines à partir de lots de données MS/MS importants, ils souffrent tous d'une incapacité à estimer de manière statistique des indices de confiance et d'estimer des taux de faux positifs pour les identifications.

L'amoncellement de données protéomiques publiées, de qualité hétérogène et peu documentées, a conduit la communauté scientifique à rechercher des moyens de les normaliser et les consolider. Trois politiques sont en cours de déploiement : la validation systématique des données par des méthodes d'évaluation probabilistes, la formalisation des informations nécessaires à la compréhension et l'exploitation des données, l'incitation au partage de l'information brute. A mesure que l'analyse protéomique installe de nouveaux standards, les règles de validation, de publication et de diffusion des données expérimentales s'affinent. C'est en gagnant cette maturité que l'analyse protéomique par spectrométrie de masse pourra, dans les années à venir, aller plus loin dans l'exploration de questions biologiques toujours plus pointues et diverses.



# **RÉSULTATS – 1<sup>RE</sup> PARTIE**

## **DÉVELOPPEMENTS MÉTHODOLOGIQUES POUR L'ANALYSE PROTÉOMIQUE CLASSIQUE**

### **Chapitre I**

De nouveaux outils et des méthodes optimisées pour l'analyse protéomique par spectrométrie de masse

### **Chapitre II**

Recherche de nouveaux biomarqueurs caractéristiques d'hémopathies lymphoïdes chroniques

### **Chapitre III**

Étude des partenaires d'interaction du facteur de transcription suppresseur de tumeur Trim24



## Chapitre I

# De nouveaux outils et des méthodes optimisées pour l'analyse protéomique par spectrométrie de masse

---

En 2009, l'Organisation internationale pour le Protéome Humain (HUPO) organisait une vaste étude interlaboratoires afin d'identifier les principaux écueils méthodologiques à l'origine de résultats non reproductibles en analyse protéomique par spectrométrie de masse<sup>1</sup>. Un échantillon contenant 20 protéines en quantités équimolaires a été remis à un groupe de 27 laboratoires de recherche pour analyse. Afin d'évaluer les performances des méthodes protéomiques, leur reproductibilité et la qualité des résultats produits, chaque équipe a mis en oeuvre ses stratégies analytiques de routine sur ses propres instruments. Sur les 27 équipes, seules 7 sont parvenues à identifier la totalité du contenu protéique de l'échantillon. Toutefois, l'évaluation des données brutes par les organisateurs a montré que toutes les protéines auraient pu être identifiées par chacun des laboratoires. Données de qualité médiocre, erreurs systématiques des différentes méthodes, contaminations, caractère aléatoire intrinsèque de l'analyse, problèmes de manipulation et de validation des données sont autant d'obstacles identifiés qui, une fois signalés, ont généralement pu être dépassés par les laboratoires concernés.

Dans un commentaire de cette étude<sup>2</sup>, Aebersold R. souligne l'extrême importance de l'expertise et de l'expérience technique nécessaires au succès des analyses protéomiques par spectrométrie de masse. En effet, à l'échelle d'échantillons plus complexes (extraits bruts, bandes de gel 1-D, ...), les problèmes soulignés sont d'autant plus importants : les mélanges protéiques étudiés peuvent contenir de plusieurs

---

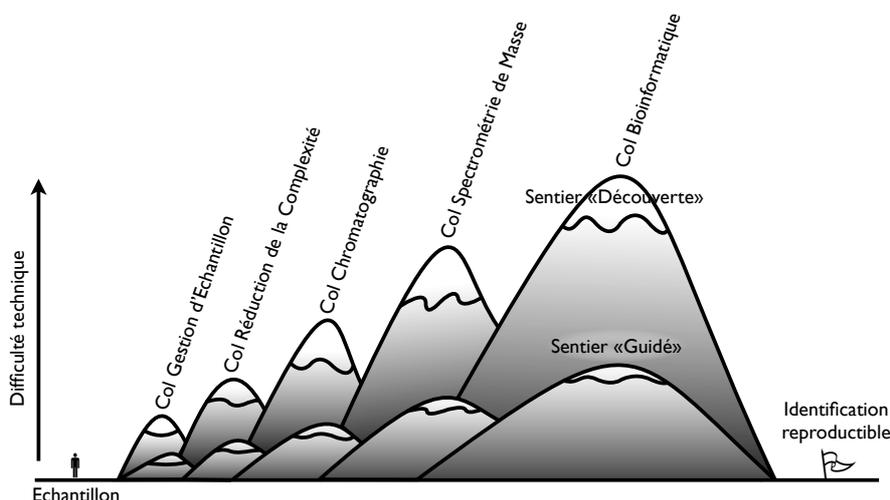
<sup>1</sup> Bell, A. W., et al. (2009). Nat Methods.

<sup>2</sup> Aebersold, R. (2009). Nat Methods.

centaines à plusieurs milliers de protéines, sur une large gamme dynamique. Pour produire une information reproductible à partir de tels échantillons, deux approches peuvent être envisagées (Figure 1) :

- Tenter de séparer le mélange peptidique pour séquencer et identifier chacun de ses constituants ;
- Viser un certain nombre de cibles préalablement identifiées pour les caractériser et les quantifier.

Atteindre une reproductibilité parfaite dans la première approche se révèle souvent être un défi technique qui demande plus d'investissement en temps et échantillon, notamment dans le cadre d'études ambitieuses comme la comparaison du contenu protéique de 2 échantillons ou la recherche de biomarqueurs spécifiques de pathologies.



**Figure 1** | Le sentier escarpé vers l'identification reproductible de protéines par spectrométrie de masse (adapté de B. Wollscheid<sup>2</sup>). L'identification reproductible du contenu protéique total d'un échantillon est un défi technique plus grand que l'étude de quelques cibles connues.

Nous avons beaucoup travaillé ces dernières années pour éviter les écueils conduisant à des résultats incomplets. Ce chapitre décrit, dans un 1<sup>er</sup> point, nos efforts pour comprendre les effets conduisant à des résultats non répétables, pour une même méthode d'analyse nanoLC-MS/MS d'échantillons biologiques de complexité moyenne (comme des bandes de gel 1-D pouvant contenir plusieurs dizaines de protéines). A partir de ces observations, le 2<sup>e</sup> point de ce chapitre détaille les optimisations instrumentales réalisées pour dépasser les obstacles induits par la prise de données expérimentales. Enfin, le 3<sup>e</sup> point met en lumière le travail d'optimisation réalisée pour améliorer le traitement informatique des données et la recherche dans les banques protéiques.

## 1. Etude de la répétabilité des analyses nanoLC-MS/MS d'échantillons biologiques complexes

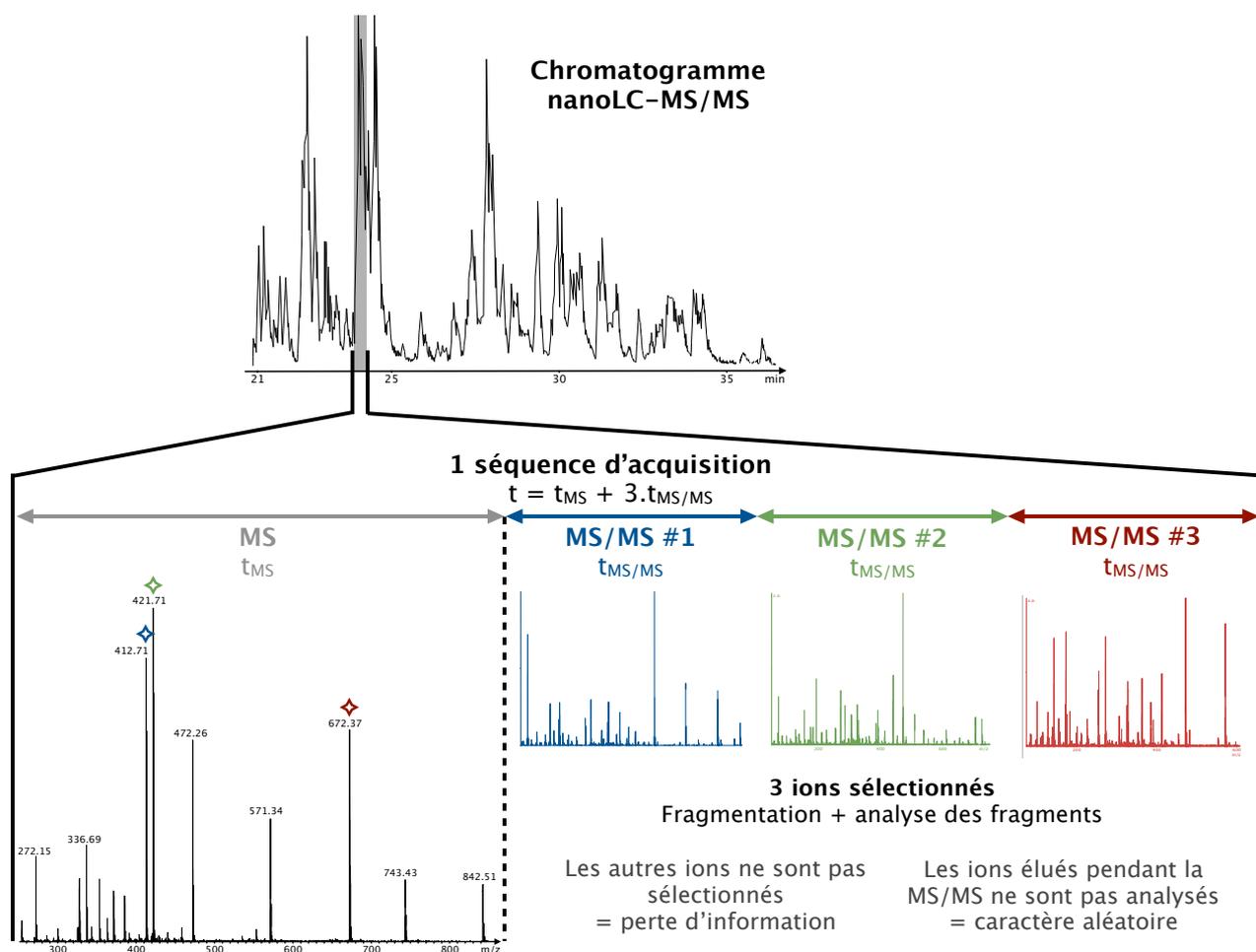
La chromatographie haute performance à nano débit (nanoLC) couplée à la spectrométrie de masse en tandem via une source électrospray (ESI-MS/MS) est largement utilisée dans les analyses protéomiques pour identifier peptides et protéines. La plupart des études protéomiques différentielles nécessitent la comparaison du contenu protéique d'échantillons à différents états (avant et après un traitement par exemple) ou dans le cadre de la recherche de biomarqueurs spécifiques de diverses pathologies. Dans tous ces cas, il est de première importance d'être en mesure d'affirmer sans ambiguïté qu'une protéine donnée est réellement présente dans un échantillon étudié. Il est tout aussi crucial de pouvoir affirmer qu'une protéine identifiée dans un premier échantillon est réellement absente d'un deuxième auquel il est comparé. Cependant malgré les avancées technologiques et attendue la complexité des échantillons étudiés, il est difficile d'identifier directement toutes les protéines présentes dans un échantillon.

## 1.1. Stochastique des analyses nanoLC-MS/MS

### 1.1.1. Origine des phénomènes de non-répétabilité

Les analyses nanoLC-MS/MS ne sont ni directes ni triviales : elles ont des limites intrinsèques liées à la cascade d'évènements qui s'enchaînent lors de chaque cycle MS/MS. De manière classique, les stratégies protéomiques mises en oeuvre en nanoLC-MS/MS sont complètement automatisées et utilisent un mode d'analyse données-dépendant (DDA, « *Data Dependant Analysis* »). Cela signifie que pendant une acquisition, l'instrument est programmé pour sélectionner les ions les plus intenses dans un premier balayage MS pour les fragmenter et analyser ces fragments (MS/MS). Cette opération peut prendre plus ou moins de temps en fonction de la vitesse d'acquisition de l'instrument. Pour les spectromètres de masse classiques comme de pointe, les cycles d'acquisition LC-MS/MS ne sont pas capables d'analyser tous les peptides élués dans une fraction chromatographique donnée (Figure 2).

Par conséquent, tous les peptides coélus ayant des intensités plus faibles que ceux sélectionnés pour la MS/MS ne pourront pas être fragmentés. Beaucoup d'ions ne sont donc pas analysés. De plus, les ions élués pendant la fragmentation et l'acquisition MS/MS ne sont pas analysés non plus. Ces phénomènes impliquent un biais récurrent et intrinsèque aux spectromètres de masse en tandem, à l'origine d'une perte d'information pendant une analyse en mode données-dépendant (DDA). Ce sous-échantillonnage confère à l'analyse nanoLC-MS/MS un caractère aléatoire (ou stochastique) difficilement maîtrisable.



**Figure 2** | Origine de la perte d'information lors d'une analyse nanoLC-MS/MS en mode DDA. Dans chaque fraction chromatographique, seuls les 3 ions les plus intenses sont sélectionnés pour fragmentation et analyse. Les autres ions, ainsi que ceux élués pendant la fragmentation et l'acquisition MS/MS, ne sont donc pas analysés.

### 1.1.2. Quelques options permettent de limiter la perte d'information

Différentes stratégies peuvent être mises en oeuvre pour dépasser les limites intrinsèques à l'instrument et se rapprocher d'une analyse complète.

▸ Les innovations techniques en chromatographie liquide contribuent de façon certaine à diminuer les effets stochastiques de la nanoLC-MS/MS. Citons par exemple la miniaturisation des systèmes de chromatographie liquide ou le développement des systèmes de chromatographie liquide à ultra-haute performance (UHPLC). Les systèmes nanoLC sur puce microfluidique (HPLC-Chip) ont de meilleures résolutions chromatographiques grâce à une réduction importante des volumes morts. Dans le cas de l'UHPLC, la phase stationnaire est constituée de billes d'un diamètre inférieur aux systèmes classiques (1,7 µm contre 3 à 5 µm), ce qui impose une pression de travail plus importante (8000–15000 psi contre 2500–5000 psi). Cela induit une plus grande sensibilité et une meilleure séparation des peptides que sur les systèmes conventionnels. Couplés aux spectromètres de masses adaptés en termes de vitesse d'acquisition et de sensibilité, des systèmes chromatographiques de ce type permettent de sélectionner plus d'ions dans le temps d'une analyse.

▸ Les progrès instrumentaux en termes de vitesse de balayage et d'acquisition permettent également de minimiser la perte d'information. Effectivement, accroître la vitesse d'acquisition (comme sur les instruments de type trappe d'ions et grâce aux progrès de l'électronique pour les instruments récents) permet de réaliser des cycles MS+MS/MS plus rapides ou de sélectionner plus d'ions dans un cycle MS+MS/MS. Des ions d'intensité plus faible sont ainsi susceptibles d'être analysés (Table 1).

▸ Par ailleurs, une amélioration généralement utilisée pour éviter que des ions coélus ne soient pas sélectionnés pour la MS/MS consiste à exclure dynamiquement les ions après qu'ils sont analysés. Cela veut dire qu'une fois qu'un ion avec une intensité raisonnable a été sélectionné et analysé, ses caractéristiques (m/z, état de charge, temps de rétention) sont ajoutées à une liste d'exclusion qui interdit sa sélection pendant une durée donnée après son analyse (par exemple 2 min). Aujourd'hui, la plupart des instruments sont implémentés d'une option d'exclusion dynamique des ions.

Spectromètre de masse	Durée de balayage en mode MS	Durée d'acquisition de 3 spectres MS/MS	Durée totale d'un cycle MS + 3 MS/MS	Système chromatographique	Largeur approximative d'un pic chromatographique à mi-hauteur
Q-TOF II Waters Corp.	2,1 s	18,1 s	20,2 s	CapLC Waters Corp.	≈ 15 s
Trappe d'ions HCTUltra Bruker Dal.	1,5 s	8 s	9,5 s	HPLC Chip Agilent Tech.	≈ 10 s
Synapt HDMS Waters Corp.	0,5 s	4,2 s	4,7 s	nanoAcquity UPLC Waters Corp.	≈ 5 s

**Table 1** | Durées d'acquisition d'un spectre MS et de 3 spectres MS/MS sur les principaux instruments utilisés au laboratoire pour les analyses protéomiques. En regard, largeur approximative d'un pic chromatographique sur le système associé.

### 1.1.3. Intérêt de réaliser plusieurs analyses nanoLC-MS/MS d'un même échantillon

Depuis quelque temps, l'idée de réaliser plusieurs analyses nanoLC-MS/MS d'un même échantillon est proposée par quelques équipes pour dépasser les limites imposées par une analyse unique<sup>3</sup>. Alors que peu d'études traitent explicitement de la question dans la littérature, la stratégie est mentionnée de plus en

<sup>3</sup> Elias, J. E., et al. (2005). Nat Methods 2, 667–75.

plus fréquemment, notamment dans le cas d'analyses d'échantillons complexes. Certaines équipes font état d'un gain de confiance en l'identification de peptides et protéines grâce à la répétition des analyses<sup>4</sup>, d'autres vont même jusqu'à émettre une hypothèse sur le lien entre l'abondance d'une protéine et la répétabilité de son identification<sup>5</sup>. Dans une étude publiée par Durr et al.<sup>6</sup>, la répétabilité est définie comme « la confiance en une analyse complète » du contenu protéique d'un échantillon ; autrement dit le nombre d'analyses individuelles et indépendantes nécessaires pour obtenir une image aussi complète que possible des protéines présentes dans un échantillon. Cette définition de la répétabilité nous a semblé pertinente et nous avons proposé de la transposer à une étude de nos systèmes d'analyse protéomique.

## 1.2. Evaluation de la répétabilité des analyses nanoLC-MS/MS

### 1.2.1. Objectif de l'étude

Dans ce travail, nous avons voulu précisément évaluer le nombre d'analyses à réaliser sur un même échantillon afin d'obtenir des informations de séquence sur un maximum de peptides et donc une image la plus complète possible du protéome. Il s'agit :

- de déterminer la quantité d'information non collectée lorsqu'on se limite à un nombre donné d'analyses (en général les quantités de matériel biologique disponibles ne permettent qu'une seule analyse nanoLC-MS/MS) ;
- de s'assurer qu'une protéine non identifiée est indéniablement absente de l'échantillon analysé. A l'inverse, la spécificité d'une protéine identifiée dans un seul échantillon peut alors être validée.

On peut ainsi déterminer un critère à partir duquel l'analyse est considérée comme représentative de l'échantillon : ce critère permettra par la suite de définir la répétabilité d'une analyse nanoLC-MS/MS et de valider par exemple des stratégies de comparaison d'échantillons.

Pour beaucoup de projets traités au laboratoire, la quantité d'échantillon disponible est très limitée. Comprendre le comportement de nos systèmes nanoLC-MS/MS et évaluer la répétabilité des identifications de protéines sur ces systèmes est donc un sujet de première importance.

### 1.2.2. Mise en oeuvre expérimentale

L'étude de la répétabilité a nécessité la mise au point d'une stratégie expérimentale originale. Nous avons choisi de travailler à partir d'échantillons modèles utilisés pour nos études protéomiques différentielles : des bandes de gel SDS-PAGE de microparticules de lymphocytes enrichies en protéines membranaires<sup>7</sup>. Chaque bande contenant potentiellement quelques dizaines de protéines, ces échantillons de complexité moyenne correspondent parfaitement à l'objet de l'étude. Au regard du temps et de la quantité d'échantillon nécessaires, nous avons choisi de sélectionner pour chaque système 3 bandes de gel 1-D à différents niveaux de masse moléculaire ( $\approx$  200 kDa, 100 kDa et 50 kDa). Le contenu protéique de chaque bande a été soumis au traitement classique de digestion à la trypsine et analysé sur nos instruments (Table 2) avec leurs paramètres de routine respectifs. Les identifications ont été réalisées selon notre méthode de recherche via le moteur Mascot<sup>®</sup> en mettant en oeuvre une stratégie « *Target-Decoy* » pour l'évaluation des critères de sélection. Un seuil limite de 1 % de faux positifs a été fixé pour l'ensemble des échantillons analysés sur chaque système (cf. point 3.2.3 de ce chapitre).

<sup>4</sup> Resing, K. A., et al. (2004). *Anal Chem* **76**, 3556-68.

<sup>5</sup> Liu, H., et al. (2004). *Anal Chem* **76**, 4193-201.

<sup>6</sup> Durr, E., et al. (2004). *Nat Biotechnol* **22**, 985-92.

<sup>7</sup> Miguet, L., et al. (2007). *Subcell Biochem* **43**, 21-34.

Système	Chromatographie liquide	Spectromètre de masse
1	CapLC, Waters Corp. LC Packings C18, 75 µm, 150 mm	QTOF II Waters Corp.
2	HPLC-Chip, Agilent Tech. Zorbax 300SB-C18, 75 µm, 150 mm	Trappe d'ions HCT Ultra Bruker Dal.
3	nanoAcquity UPLC, Waters Corp. Acquity UPLC BEH130 C18, 75 µm, 200 mm	Synapt HDMS Waters Corp.

**Table 2** | Systèmes utilisés pour l'étude de la répétabilité des analyses nanoLC-MS/MS. Les instruments ont été testés dans leur configuration de routine.

Six analyses nanoLC-MS/MS successives de chaque échantillon ont été réalisées afin de déterminer la fraction (le pourcentage) de protéines et peptides identifiés en fonction du nombre d'analyses nanoLC-MS/MS effectuées. Puis, nous avons évalué le nombre d'analyses nanoLC-MS/MS à répéter pour obtenir un bon aperçu du contenu protéique global de chaque échantillon. Les interprétations ont été réalisées au niveau des protéines puis des peptides, car l'augmentation du nombre de peptides identifiés n'est pas nécessairement répercutée directement sur le nombre de protéines validées.

L'étude de chaque échantillon s'articule en 3 phases :

- ▶ L'identification des protéines analysées pour chaque injection ;
- ▶ La comparaison pour chaque analyse des protéines identifiées et des peptides permettant une identification ;
- ▶ La mise en évidence du pourcentage de protéines nouvellement identifiées lors de chaque injection par rapport à toutes les précédentes.

Un soin particulier a été pris pour s'affranchir des « effets d'ordre » entre les 6 essais. En effet, la réalisation successive d'analyses d'un même échantillon et la comparaison des résultats imposent un ordre qui biaise le caractère aléatoire de l'expérience : les résultats de la deuxième analyse par exemple doivent, dans la modélisation, avoir la même probabilité d'être mesurés lors de chacune des 5 autres injections. Pour obtenir des données statistiques sensées, nous avons calculé la moyenne des résultats sur les 6! permutations possibles, soit 720 ordres de comparaison. Chaque analyse réalisée contribue ainsi au résultat final pour chaque injection. L'écriture d'une macro VisualBasic® a permis de réaliser ces lourds calculs qui ont pu prendre plusieurs jours pour aboutir.

## 1.3. Résultats et discussion

### 1.3.1. Observations préliminaires

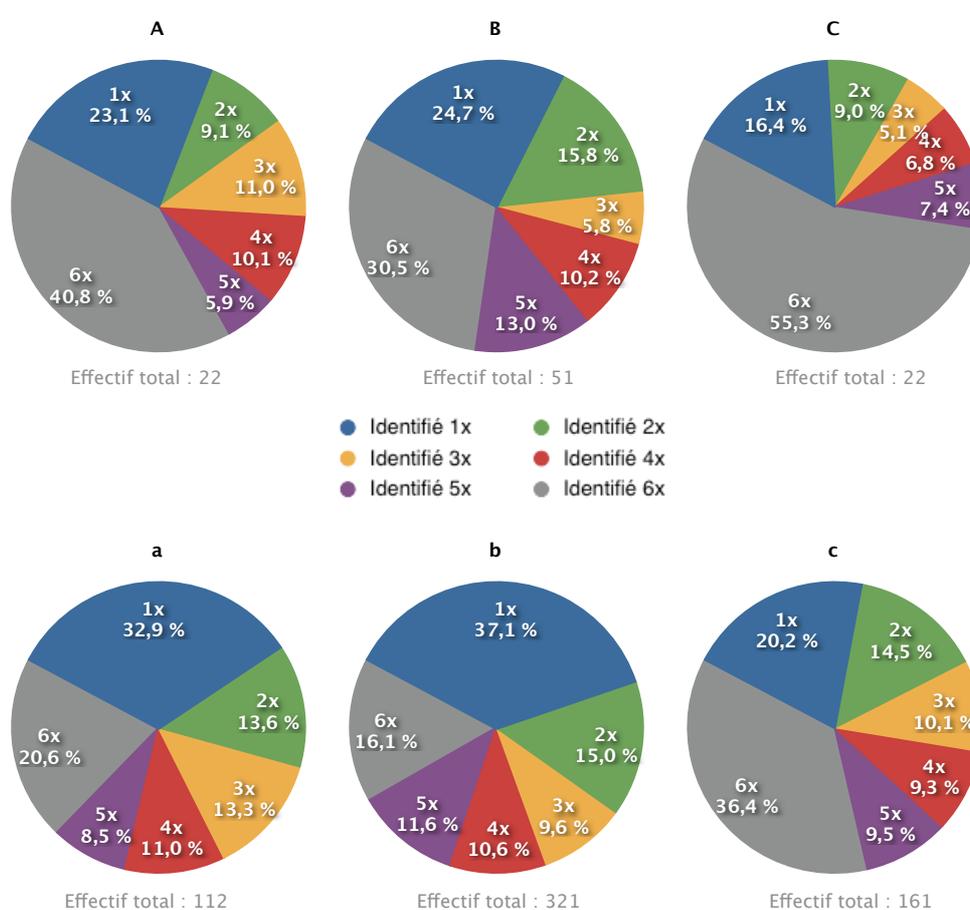
- ▶ Absence d'effets mémoires

Avant de travailler sur les questions de répétabilité et d'analyser les résultats fournis par des analyses successives d'un même échantillon, il est important de vérifier qu'il n'existe pas d'effet mémoire à l'origine de phénomènes d'hystérèses. En effet, la rétention sur la colonne chromatographique de peptides d'une analyse à l'autre pourrait conduire à la contamination de l'analyse suivante et fausser les résultats de répétabilité. Afin d'étudier ces éventuels effets accumulatifs, une série d'analyses alternant échantillon et « blanc » a été réalisée.

Aucune protéine n'a pu être significativement identifiée dans les « blancs ». La quantité de matériel retenue sur la colonne chromatographique entre deux analyses consécutives est donc négligeable et ne perturbe nullement les résultats des injections ultérieures.

► Mise en évidence de la non-répétabilité des identifications de protéines et peptides associés

Les résultats suivants présentent le bilan sur 6 analyses consécutives des identifications de protéines et de peptides, pour chacun des instruments testés (Figure 3). Certaines protéines (et certains peptides) ont été identifiées dans chacune des 6 répliques alors que d'autres n'ont pu être identifiées que dans une seule des analyses. En moyenne sur le système CapLC-Q/TOF II, 22 protéines uniques ont été identifiées. 41 % l'ont été dans les 6 analyses et 23 % dans une seule. Sur un total moyen de 112 peptides, seuls 21 % apparaissent dans l'ensemble des résultats et près de 33 % figurent dans une unique analyse. Les résultats obtenus sur le couplage HPLC-Chip-HCT Ultra sont similaires avec un effectif moyen de 51 protéines et 321 peptides. Sur le système nanoAcquity UPLC-Synapt HDMS, on observe par contre que plus de 55 % d'un total moyen de 22 protéines ont été identifiés dans les 6 injections consécutives et seuls 16 % dans une unique analyse. En termes de peptides, 36 % des 161 peptides identifiés en moyenne sont visibles dans toutes les analyses et à peine 21 % dans une seule.



**Figure 3** | Répétabilité des identifications de protéines (A, B, C) et de peptides (a, b, c) par nanoLC-MS/MS sur les systèmes CapLC-Q/TOF II (A et a), HPLC-Chip-HCT Ultra (B et b) et nanoAcquity UPLC-Synapt HDMS (C et c). Les données présentées sont des moyennes calculées sur 3 échantillons pour chaque instrument.

Comme exposé précédemment, le caractère aléatoire intrinsèque des analyses nanoLC-MS/MS implique clairement une non-répétabilité des identifications de protéines et de peptides. Ce phénomène est directement amplifié par les effets de suppression de signal imposés lors de la coélution de peptides et par de possibles petites variations dans le profil de séparation chromatographique impliquant des mesures MS d'ions parents à des temps d'éluion différents. Par ailleurs, la gamme dynamique de concentration des protéines (et donc des peptides) dans l'échantillon peut être importante. Des peptides très abondants vont donc être identifiés de manière répétée, alors que des peptides présents en faible concentration ne seront pas systématiquement détectés.

Le système CapLC-Q/TOF II, le plus ancien de nos équipements, est aussi le moins sensible et présente les temps d'acquisition les plus longs (bien supérieurs à la largeur d'un pic chromatographique). La répétabilité des analyses nanoLC-MS/MS y demeure donc médiocre. La trappe d'ions couplée à la chromatographie sur puce microfluidique (HPLC-Chip-HCT Ultra) est plus sensible : pour un même échantillon, jusqu'à 50 % de protéines supplémentaires peuvent être identifiées par rapport au système CapLC-Q/TOF II (données non présentées). La vitesse d'acquisition en mode « *ultra-scan* » semble par ailleurs adaptée au système de chromatographie sur puce microfluidique puisqu'ils présentent des échelles de temps similaires. Malgré cela, la répétabilité des identifications n'est pas sensiblement améliorée : près de 50 % des protéines étant identifiées dans moins de 3 injections sur 6. C'est finalement le couplage nanoAcquity UPLC-Synapt HDMS, le système le plus récemment acquis par le laboratoire, qui présente le cas le plus favorable : il bénéficie des innovations électroniques lui permettant d'atteindre des temps de balayages 5 fois inférieurs aux instruments de première génération et compatibles avec la chromatographie ultra-haute performance.

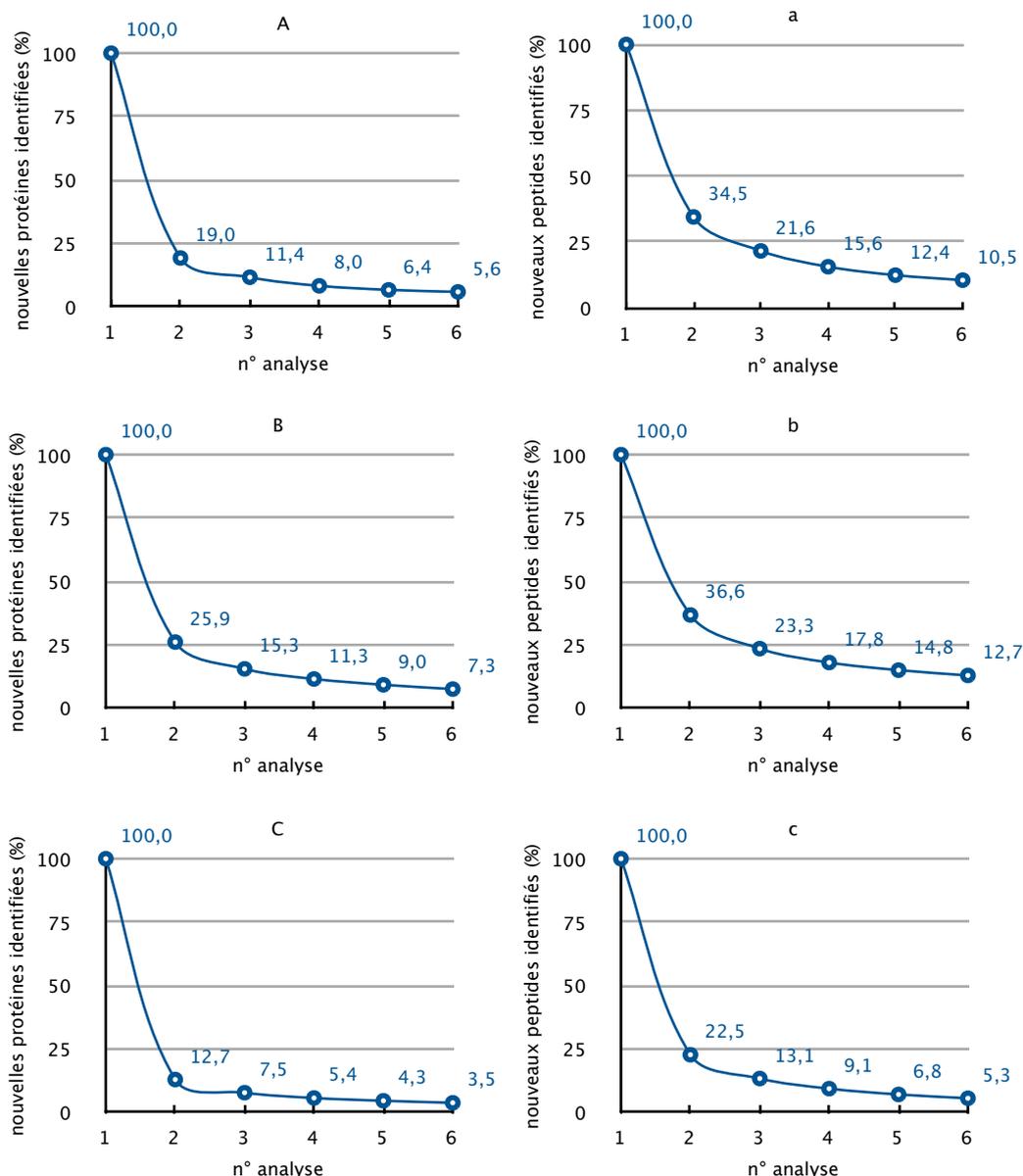
Ces données nous confortent surtout dans l'idée que la répétition d'analyses d'un même échantillon permet d'accroître significativement le nombre de peptides identifiés et donc, le nombre et/ou la qualité des identifications de protéines. Il semble par conséquent qu'en dehors des avancées technologiques, une optimisation méthodologique permettrait d'améliorer la « confiance en une analyse complète » du contenu protéique d'un échantillon. Nous avons pour cela proposé d'évaluer un critère de répétabilité, au-delà duquel on estime que l'information recueillie par toute analyse supplémentaire est négligeable.

### 1.3.2. Détermination du nombre d'analyses à répéter

#### ▸ Evaluation et définition du contenu protéique global d'un échantillon

Afin d'estimer le nombre d'analyses à répéter pour obtenir une image significative du contenu protéique identifiable d'un échantillon, le pourcentage moyen de protéines et peptides nouvellement identifiés dans chacune des 6 analyses successives par rapport aux précédentes a été calculé selon la méthode décrite plus haut (Figure 4). Il est ainsi établi qu'entre 13 et 26 % des protéines (22 et 37 % des peptides) identifiées lors de la deuxième analyse ne l'avaient pas été dans la précédente. De la même façon, entre 7 et 15 % des protéines (13 et 23 % des peptides) identifiées lors d'une troisième injection n'avaient été découverts dans aucune des 2 analyses précédentes.

Au-delà de la 6<sup>e</sup> analyse, il peut être estimé en extrapolant que toute analyse supplémentaire ne permettrait de découvrir que peu d'information : moins de 5 % des protéines et moins de 10 % des peptides analysés seraient nouvellement découverts, ce qui n'améliorerait la connaissance du contenu protéique total que de façon négligeable. Par conséquent, l'approximation peut être posée que pour un échantillon de complexité moyenne issu d'une bande de gel 1-D, la totalité des protéines et peptides identifiables est connue en réalisant 6 analyses.

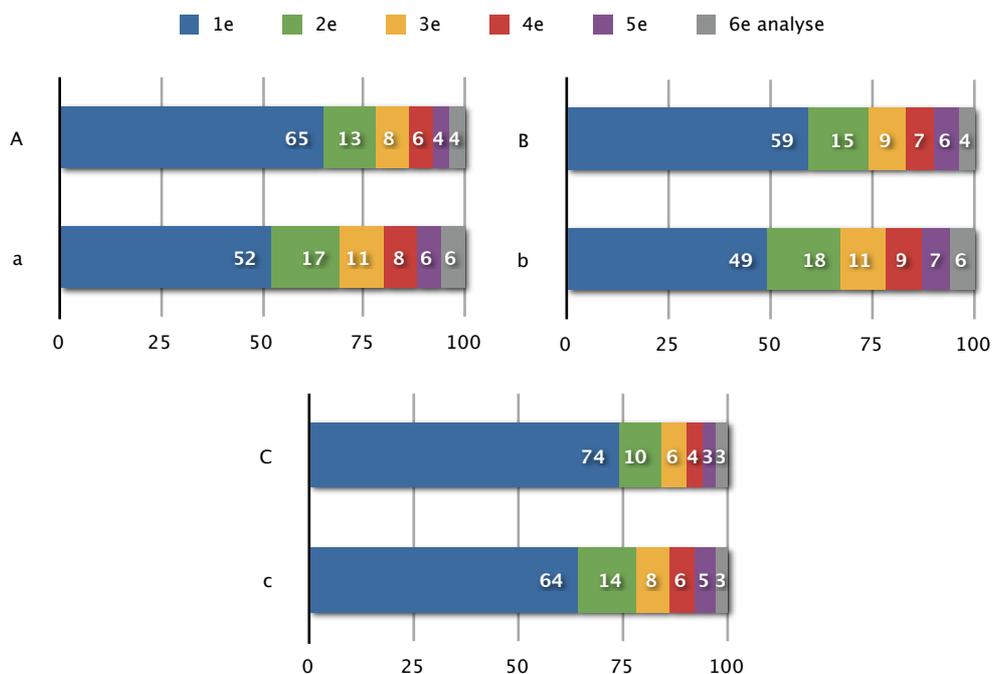


**Figure 4** | Pourcentage de nouvelles protéines (A, B, C) et de nouveaux peptides (a, b, c) découverts dans la dernière analyse nanoLC-MS/MS par rapport aux précédentes, sur les systèmes CapLC-Q/TOF II (A et a), HPLC-Chip-HCT Ultra (B et b) et nanoAcquity UPLC-Synapt HDMS (C et c). Les données présentées sont des moyennes des 720 permutations possibles calculées pour 3 échantillons.

#### ► Détermination d'un critère de répétabilité

Cette hypothèse sur l'information accessible avec la méthode et les instruments utilisés permet d'exprimer les résultats en termes de pourcentage du contenu protéique découvert lors de chaque analyse. En effet, considérant que 100 % des protéines sont identifiées après 6 analyses, la part contributive de chaque injection peut être évaluée (Figure 5).

Il est montré que sur le système CapLC-Q/TOF II, 65 % des protéines et 52 % des peptides accessibles sont identifiés lors de la première analyse. Plus de 86 % du contenu protéique (et 80 % du contenu peptidique) est découvert après 3 analyses. Au-delà, les injections supplémentaires ne participent à la découverte de nouvelles protéines ou peptides qu'à hauteur de 5 % chacune. De fait, lors d'un travail sur du matériel de complexité moyenne et disponible en quantité limitée, il n'est pas pertinent de réaliser plus de 3 analyses nanoLC-MS/MS pour des raisons évidentes d'économies de temps, d'argent et d'échantillon.



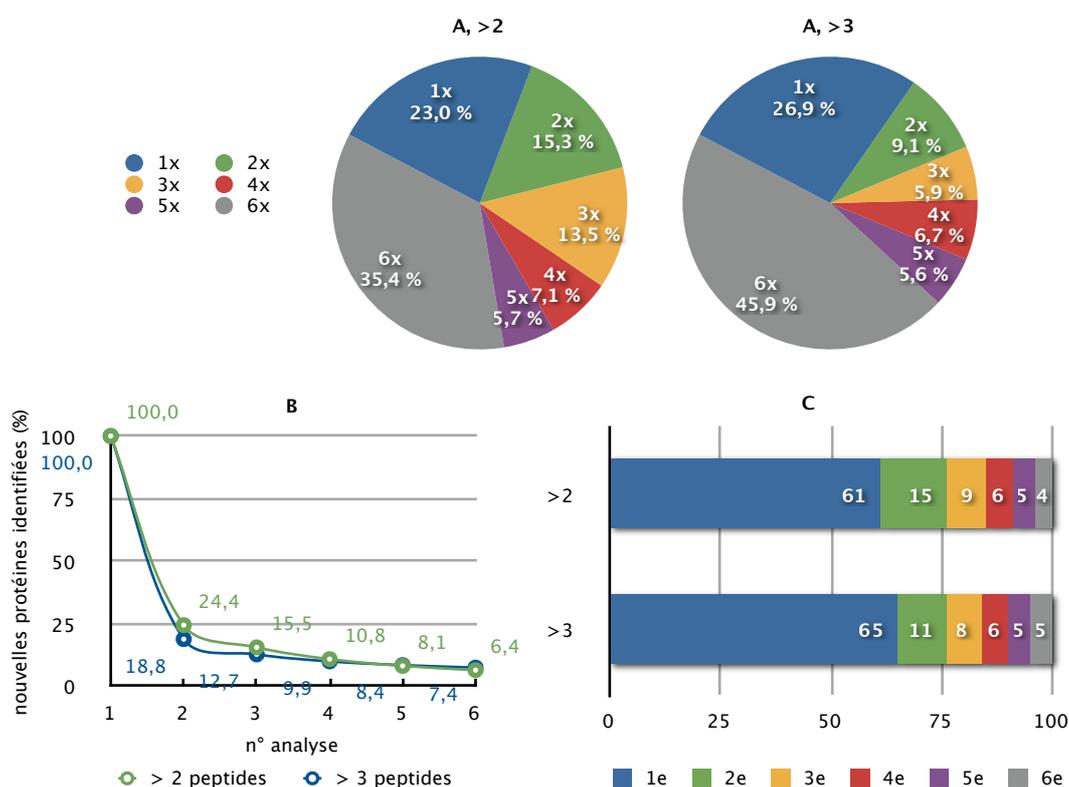
**Figure 5** | Pourcentage du contenu protéique global (A, B, C) et peptidique global (a, b, c) — identifiables en 6 analyses — en fonction du nombre d’analyses nanoLC-MS/MS effectuées, sur les systèmes CapLC-Q/TOF II (A et a), HPLC-Chip-HCT Ultra (B et b) et nanoAcquity UPLC-Synapt HDMS (C et c). Les données présentées sont des moyennes des 720 permutations possibles calculées pour 3 échantillons.

Pour le couplage HPLC-Chip-HCT Ultra, les résultats sont équivalents. Mais on retiendra qu’en terme d’effectif, la trappe — plus sensible — permet théoriquement d’identifier plus de peptides et de protéines pour un même échantillon.

Par contre, le système nanoAcquity UPLC-Synapt HDMS a la capacité d’identifier près de 85 % du contenu protéique global en seulement 2 analyses, et 90 % en 3 analyses (78 et 86 % du contenu peptidique en 2 et 3 analyses, respectivement). Si ces résultats sont plus satisfaisants, en majeure partie grâce aux améliorations technologiques, plusieurs analyses du même échantillon sont toujours nécessaires pour obtenir une image aussi complète que possible du contenu protéique. Dans le cas du couplage nanoAcquity UPLC-Synapt HDMS, réaliser un duplicata de chaque analyse semble être un bon compromis. En outre en réalisant plusieurs analyses du même échantillon, les protéines déjà identifiées lors d’une première mesure pourront bénéficier d’un meilleur recouvrement de séquence et donc d’une confiance en l’identification supérieure grâce à la découverte de nouveaux peptides appartenant potentiellement à ces protéines.

► Influence du seuil d’identification choisi

Pour des projets nécessitant la comparaison de contenus protéiques de plusieurs échantillons, comme la recherche de candidats biomarqueurs, l’identification des protéines minoritaires est requise. Dans cet objectif, il est raisonnable de considérer qu’un unique peptide permet d’identifier une protéine dans une stratégie « *Target-Decoy* » fixant un taux de faux positifs inférieur à 1 %. Nos résultats précédents s’inscrivaient dans cette stratégie de validation. Dans des cas plus généraux — où l’identification de protéines majoritaires est suffisante —, les critères appliqués sont plus sélectifs. Certaines équipes ajoutent un filtre supplémentaire en ne retenant que les protéines identifiées avec plus de 2 ou 3 peptides. Nous avons souhaité estimer la répétabilité des identifications par nanoLC-MS/MS dans ces conditions et évaluer l’influence des critères fixant un nombre de peptides minimum pour une identification. Les outils précédemment décrits ont été mis en oeuvre pour cette étude dont les résultats sont présentés ci-dessous (Figure 6).



**Figure 6** | Résultats des analyses nanoLC-MS/MS sur le système HPLC-Chip-HCT Ultra avec un critère de validation sélectif : identification avec plus de 2 peptides (>2) ou plus de 3 peptides (>3). (A) Répétabilité des identifications de protéines. (B) Pourcentage de nouvelles protéines identifiées dans la dernière analyse nanoLC-MS/MS par rapport aux précédentes. (C) Pourcentage du contenu protéique global — identifiable en 6 analyses — en fonction du nombre d'analyses nanoLC-MS/MS effectuées. Les données présentées sont des moyennes calculées sur 3 échantillons.

Des critères de validation plus exigeants en termes de nombre de peptides réduisent considérablement le contenu protéique identifiable : 51 protéines sont identifiées en moyenne avec un critère à 1 peptide, 40 et 26 pour des critères à 2 et 3 peptides, respectivement. Pourtant, il est clairement établi que pour des seuils à 2 ou 3 peptides, la répétabilité de nos systèmes nanoLC-MS/MS n'est toujours pas idéale. En fixant des seuils d'identification stricts, on abandonne généralement les protéines minoritaires, identifiées avec peu de peptides et qui participent fortement à la non-répétabilité des analyses. Mais on impose aussi un nouveau biais en faisant disparaître dans quelques répliques, des identifications de protéines qui sont potentiellement identifiées dans toutes les analyses, parfois certes avec seulement 1 ou 2 peptides. Ces phénomènes disparaissent peu à peu en augmentant la rigueur des filtres.

Cela montre qu'il faut manier ces filtres sur le nombre de peptides avec précaution, en dépit de l'idée que les analyses sont plus fiables lorsque les identifications à peptide unique ne sont pas considérées<sup>8</sup>. Pour améliorer la confiance dans les identifications, on optimisera plutôt les critères de scores évalués par les méthodes « *Target-Decoy* ». Ils sont moins liés aux limites intrinsèques des couplages nanoLC-MS/MS qu'à la qualité des données générées. En revanche, si seules les protéines majoritaires intéressent l'objet de l'étude, alors un critère fort sur le nombre de peptides pourra être choisi sans risque (plus de 5 peptides par exemple).

<sup>8</sup> Gupta, N., et al. (2009). J Proteome Res.

## 1.4. Proposition et évaluation d'une stratégie visant à améliorer la répétabilité des analyses nanoLC-MS/MS

### 1.4.1. La concaténation des listes de masses avant la recherche par les moteurs

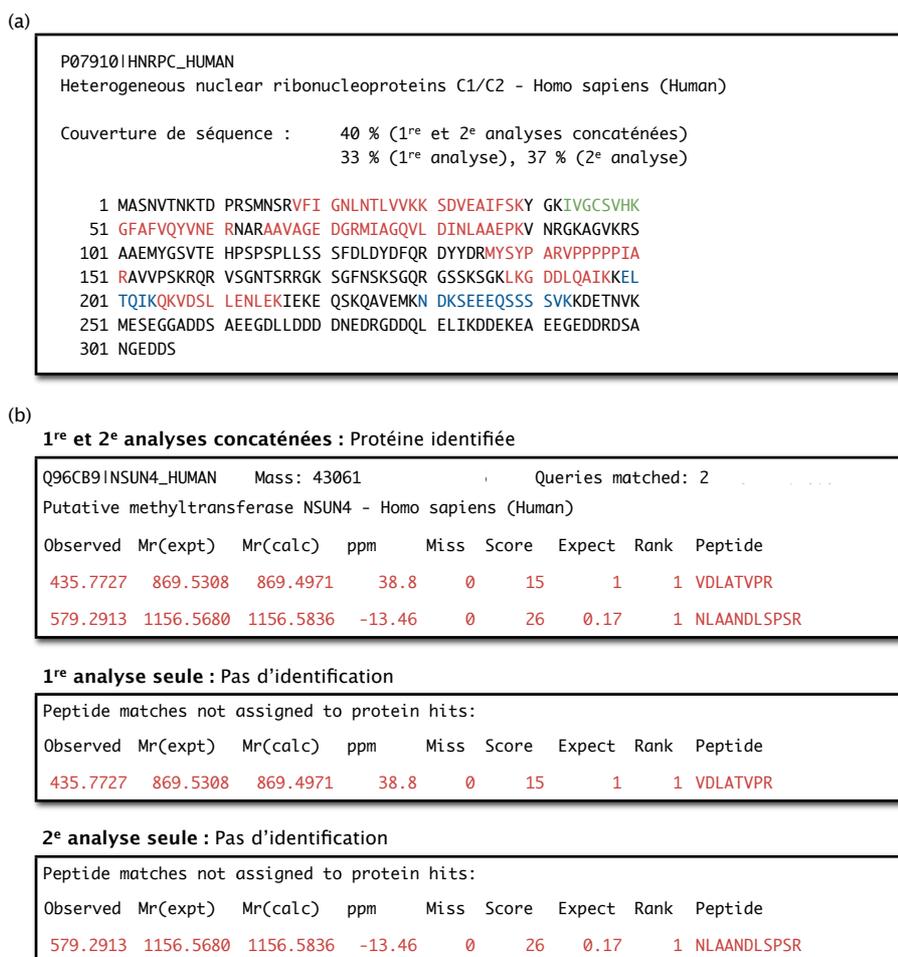
Tirant parti des informations précieuses que révèle cette étude de la répétabilité des analyses nanoLC-MS/MS sur nos systèmes utilisés pour l'analyse protéomique d'échantillons de complexité moyenne, une stratégie affinée peut être proposée pour augmenter le nombre de protéines identifiées et améliorer la fiabilité des identifications. En effet, il a été montré que sur un système récent comme le couplage nanoAcquity UPLC-Synapt HDMS, 85 % du contenu protéique identifiable d'un échantillon pouvaient être découverts en dupliquant les analyses ; de nouvelles protéines et de nouveaux peptides étant détectés lors de la deuxième analyse (13 et 23 % de nouvelles identifications, respectivement).

Prises individuellement, ces 2 analyses participent chacune indépendamment à la liste finale des protéines identifiées puisque 2 recherches indépendantes sont menées à partir de chacun des fichiers de résultats (« listes de masses » ou « *peak list* »). Afin de cumuler l'information recueillie lors des mesures successives, nous avons proposé de concaténer *in silico* les fichiers « *peak list* » issus de chaque doublet d'analyses avant la soumission des résultats au moteur de recherche. L'unique fichier rassemblant deux à deux le contenu des fichiers de résultats mis bout à bout se comporte comme s'il résultait d'une unique analyse au cours de laquelle toute l'information des duplicatas aurait été mesurée.

Cette opération permet en théorie d'exploiter l'information complémentaire recueillie lors de chacune des 2 analyses. En premier lieu, de nouvelles identifications sont susceptibles d'apparaître pour des raisons intrinsèques aux algorithmes des moteurs de recherche puisque de nouvelles informations leur sont soumises. De manière directe, cela a aussi un effet potentiel sur 2 niveaux (Figure 7).

- D'abord, pour une protéine identifiée avec plus de 2 peptides dans une première analyse, la découverte de nouveaux peptides appartenant à cette protéine dans une deuxième analyse contribuera directement à l'augmentation du recouvrement de séquence de la protéine considérée.

- Ensuite, à un peptide unique qui, dans une première analyse, ne passe pas le seuil de validation pour l'identification de la protéine à laquelle il appartient, il peut s'ajouter la découverte dans une deuxième analyse d'un peptide différent et au score individuel toujours insuffisant. Puisque pour 2 peptides le seuil de validation est moins strict que pour une identification à un peptide, alors une nouvelle protéine peut éventuellement être identifiée grâce à la combinaison de l'information issue des 2 analyses (cf. point 3.4. de ce chapitre : Mise au point d'une stratégie de détermination d'un taux de faux positifs par l'approche « *Target-Decoy* »).



**Figure 7** | La concaténation des fichiers « *peak list* » issus des résultats de 2 analyses nanoLC-MS/MS d'un même échantillon permet : (a) d'améliorer le recouvrement de séquence des protéines identifiées. Exemple de l'identification de HNRPC\_HUMAN (en rouge, information commune aux 2 analyses ; en vert, information spécifique à la 1<sup>re</sup> analyse ; en bleu, information spécifique à la 2<sup>e</sup> analyse). (b) d'identifier des protéines supplémentaires grâce à la combinaison de 2 peptides uniques séquencés séparément dans chacune des 2 analyses. Exemple de l'identification de NSUN4\_HUMAN (rapport Mascot®). Pour cette analyse, les scores d'ions seuils pour l'identification ont été fixés à 15 pour des identifications avec plus de 2 peptides et à 35 pour des identifications avec un peptide unique.

### 1.4.2. Evaluation de la stratégie sur un échantillon total

Le bénéfice apporté par la duplication des analyses nanoLC-MS/MS d'un échantillon et par la concaténation des fichiers « *peak list* » issus de ces mesures a été évalué pour des échantillons complets (pistes de gel 1-D). Ces éléments permettent de confirmer l'intérêt de notre stratégie pour améliorer le nombre et la qualité des identifications à l'échelle d'une étude protéomique entière.

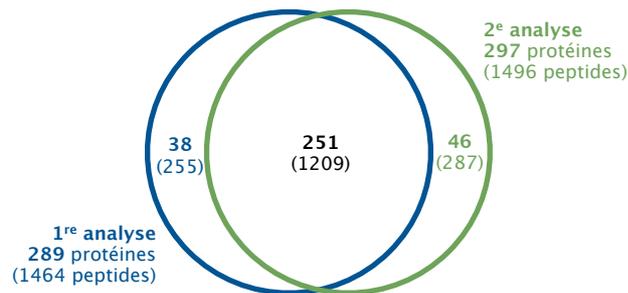
#### ► Mise en oeuvre expérimentale

Dans le cadre de l'étude des partenaires d'interaction de Trim23 développée au chapitre III de cette partie, des échantillons de complexité moyenne ont été déposés sur gel 1-D et des bandes ont été découpées de manière systématique ( $\approx 2$  mm). Quatre échantillons différents ont été étudiés. Les analyses nanoLC-MS/MS ont été réalisées en duplicata dans les conditions standards d'utilisation sur le couplage nanoAcquity UPLC-Synapt HDMS (cf. point 2.1.2 de ce chapitre). Les fichiers de résultats — les listes de masses ou « *peak list* » (.pkl) — ont été concaténés grâce à un script ad hoc fait maison. Les identifications ont été réalisées selon notre méthode de recherche par le moteur Mascot®, soit à partir des fichiers « *peak list* » individuels, soit à partir des fichiers concaténés. La stratégie « *Target-Decoy* » a permis de fixer un critère de sélection correspondant à un taux de faux positifs inférieur à 1 %. Finalement entre chaque cas, le nombre et la qualité des identifications de protéines et de peptides ont été comparés afin de valider l'intérêt

de dupliquer les analyses nanoLC-MS/MS et d'évaluer le bénéfice apporté par la concaténation des fichiers « *peak list* » des analyses dupliquées.

▸ Résultats et discussion

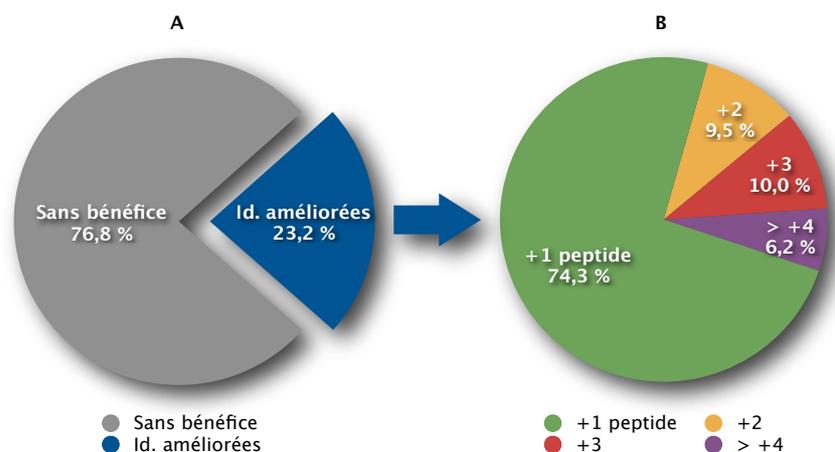
Une liste totale des protéines et des peptides identifiés a été établie pour chaque groupe de duplicatas : une liste rassemblant les résultats de toutes les « premières analyses » d'une part, et une autre de toutes les « deuxièmes analyses » d'autre part. Les identifications de chaque liste ont été comparées (comme le montre l'exemple présenté Figure 8). En moyenne pour les 4 échantillons étudiés, en considérant les 2 permutations possibles, 17 % de protéines et 18 % de peptides supplémentaires ont été rendus accessibles grâce à la duplication des analyses par rapport à une analyse unique.



**Figure 8** | Exemple du gain en nombre d'identifications réalisé en dupliquant les analyses nanoLC-MS/MS d'une piste totale de gel 1-D découpée de manière systématique. Les données présentées sont issues de l'analyse du contenu protéique de 37 bandes sur le couplage nanoAcquity UPLC-Synapt HDMS.

Ces résultats confortent notre stratégie visant à améliorer la confiance en une analyse complète du contenu protéique de l'échantillon analysé. La réalisation de deux analyses de chaque échantillon n'est pas suffisante pour obtenir une image exhaustive, mais est un bon compromis en termes d'impératifs matériels (temps, échantillon, coût) et d'une réalisation pertinente de l'étude protéomique.

Finalement, l'intérêt de concaténer deux à deux les fichiers « *peak list* » résultant de chaque duplicata a été confirmé. Les identifications de protéines (et de tous leurs peptides) ont été comparées pour une recherche dans les banques soit à partir des fichiers concaténés, soit à partir des fichiers individuels de chaque réplique.



**Figure 9** | Amélioration de la qualité des identifications de protéines grâce à la duplication des analyses nanoLC-MS/MS et à la concaténation des paires de fichiers « *peak list* » avant la recherche. (A) Pourcentage de protéines qui ont été identifiées avec plus de peptides que dans le cas d'un simple cumul des résultats considérés individuellement. (B) Peptides gagnés par la part de protéines dont l'identification est améliorée. Les données présentées sont des moyennes calculées sur le total des identifications de 4 pistes de gel 1-D découpées de manière systématique et analysées sur le système nanoAcquity UPLC-Synapt HDMS.

Contrairement à ce qui pouvait être espéré lorsque les duplicatas sont concaténés, seules quelques nouvelles protéines apparaissent ( $\approx 1\%$  en moyenne, données non présentées). On observe cependant une amélioration de la qualité des identifications dans plus de 23 % des cas (Figure 9). En effet, grâce à la fusion des fichiers « *peak list* », de nouveaux peptides inaccessibles dans les identifications individuelles viennent compléter le recouvrement de nombreuses protéines. Les 3/4 de ces protéines reçoivent un seul peptide séquencé supplémentaire, mais on a pu observer jusqu'à 9 peptides de plus pour une même identification.

Ces arguments renforcent encore l'intérêt et le soin qu'il faut prendre pour la réalisation d'analyses protéomiques par nanoLC-MS/MS et le traitement des données issues de l'expérience. Dans ce cas, il est montré que de dupliquer les analyses de chaque bande de gel 1-D et de fusionner les paires de fichiers « *peak list* » issus de chaque réplique avant la recherche dans les banques de données, permet d'améliorer le nombre et la qualité des identifications accessibles de manière simple et pertinente.

Dans la même idée, certaines équipes proposent de fusionner les fichiers résultats de toutes les analyses nanoLC-MS/MS réalisées sur une piste entière de gel 1-D. Cela est pertinent dans le cas d'une analyse LC/LC-MS/MS de type MudPit où un digest de protéines total est soumis à séparation chromatographique à 2 dimensions. Mais cette opération semble bien moins appropriée lorsque, dans une première dimension, les protéines sont séparées sur gel d'électrophorèse. En effet, la concaténation des analyses de toutes les bandes découpées a pour conséquence directe de rendre possible l'appariement de peptides analysés dans deux bandes très éloignées au profit de l'identification d'une protéine. De fait, l'apparition d'identifications erronées devient fortement probable. Il est alors nécessaire de fixer des critères de sélection beaucoup plus stricts pour conserver un taux de faux positifs acceptable, au détriment de l'identification de nombreuses protéines minoritaires.

## 1.5. Valorisation de ces travaux

### 1.5.1. Communications orales

Béchade, G., Mauvieux, L., Miguet, L., Sanglier, S., Van Dorsselaer, A.

Étude de la répétabilité des analyses nanoLC-MS/MS en protéomique : Vers une image complète du contenu protéique d'un échantillon biologique complexe.

12<sup>e</sup> Journées du Club Jeunes de la Société Française de Spectrométrie de Masse.  
Saint-Dié des Vosges (France). 19-23 mars 2007.

Béchade, G., Mauvieux, L., Miguet, L., Sanglier, S., Van Dorsselaer, A.

Repeatability of nanoLC-MS/MS analysis in proteomic analysis: Towards comprehensive determination of protein content in complex biological samples.

International Research Training Group GRK532, 3rd Period, Seminar C  
Strasbourg (France). 30 mai-01 juin 2007.

### 1.5.2. Communication par affiche

.....  
: Béchade, G., Mauvieux, L., Miguet, L., Sanglier, S. and Van Dorsselaer, A.  
:

: Étude de la répétabilité des analyses nanoLC-MS/MS en protéomique : Vers une image complète du contenu  
: protéique d'un échantillon biologique complexe.  
:

: 24<sup>e</sup> JFSM, Spectrométrie de Masse et Analyse protéomique.  
:

: Pau (France). 17-20 septembre 2007.  
:.....

### 1.6. Conclusions

Dans le cadre de nos projets de recherche, il est souvent question de comparer le contenu protéique d'échantillons relativement complexes (déposés sur gel 1-D). Ces analyses différentielles ne sont pertinentes que si le contenu protéique total des échantillons peut être découvert de manière fiable.

En tentant de comprendre les phénomènes entravant la répétabilité des analyses nanoLC-MS/MS, l'intérêt de répéter les analyses pour augmenter la quantité et la qualité des identifications a été évalué. Une stratégie d'analyse est finalement proposée afin d'accroître la crédibilité des études protéomiques différentielles.

▶ Les analyses nanoLC-MS/MS présentent un caractère aléatoire intrinsèque responsable de la non-répétabilité des identifications de protéines et de peptides ;

▶ Réaliser plusieurs analyses nanoLC-MS/MS d'un même échantillon permet d'améliorer la « confiance en une découverte complète » du contenu protéique d'un échantillon ;

▶ Jusqu'à 26 % de protéines et 37 % de peptides nouveaux sont découverts parmi les identifications d'une analyse dupliquée ;

▶ Sur nos couplages, de 2 à 3 analyses peuvent être nécessaires pour recouvrir plus de 80 % du contenu protéique et peptidique d'un échantillon ;

▶ Concaténer les résultats issus de répliques d'analyses, avant de les soumettre aux moteurs de recherche, permet d'améliorer la qualité des identifications et, dans une moindre mesure, leur nombre.

## 2. Des optimisations instrumentales permettent une meilleure prise des données protéomiques

Malgré le grand engouement que connaît l'analyse protéomique par spectrométrie de masse nanoLC-MS/MS, l'information recueillie demeure grandement incertaine et incomplète. Nous venons d'établir que même pour des stratégies optimisées pour améliorer la répétabilité et l'exhaustivité des données générées, des difficultés majeures restent à relever pour identifier toutes les protéines d'un échantillon et en obtenir un recouvrement de séquence complet. Les progrès à réaliser passent d'abord par des avancées technologiques en spectrométrie de masse. Il ne faut cependant pas confondre expertise et maîtrise de la pratique analytique avec nouvelles technologies. Un échantillon bien préparé, des paramètres instrumentaux correctement affinés et contrôlés permettent généralement d'augmenter directement les performances. La suite de ce chapitre détaille le travail d'optimisation qui a été mené afin d'améliorer la prise de données protéomiques.

### 2.1. La résolution chromatographique : nanoHPLC-Chip et nanoUHPLC

Alors que les couplages de type nanoLC-ESI-MS se sont imposés comme technologie de choix pour l'analyse protéomique de mélanges peptidiques complexes, de nombreux défis techniques ont commencé à être relevés ces dernières années. La technique souffrait en effet de difficultés pratiques causées par l'utilisation de nombreux capillaires, tubulures et connexions qui pouvaient se boucher ou fuir et généraient d'importants volumes morts. Par ailleurs, les systèmes nanoélectrospray à aiguille nécessitaient une manipulation délicate et une maintenance contraignante.

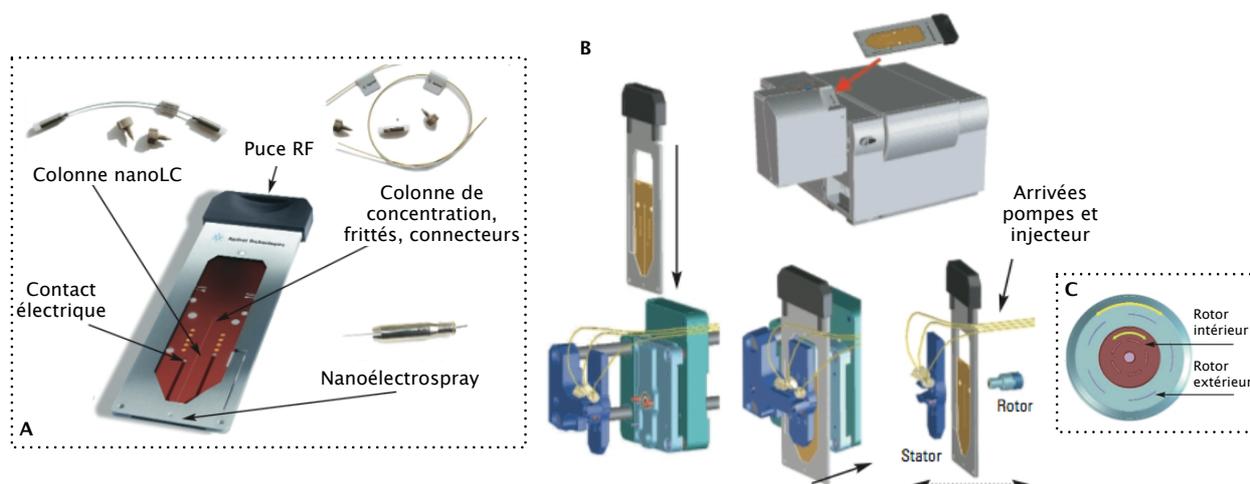
Le développement de la chromatographie sur puce microfluidique et de la chromatographie à ultra-haute performance a permis d'atteindre de meilleures résolution, sensibilité et débit d'analyse, ainsi qu'une facilité d'utilisation et une fiabilité inégalées.

#### 2.1.1. Le système nanoHPLC-Chip Cube (Agilent Technologies)

##### ▸ Présentation

Le système nanoHPLC-Chip Cube commercialisé par la société Agilent Technologies depuis 2005 représente une innovation technologique majeure. Rassemblant sur une même puce microfluidique les colonnes d'enrichissement et de séparation et la pointe de nébulisation nanoélectrospray, il permet d'éliminer 50 % des connexions et raccords indispensables aux systèmes classiques (Figure 10). Il en résulte une diminution draconienne des volumes morts et risques de fuites, et une amélioration significative de la résolution chromatographique et de la sensibilité. La conception de ces puces par ablation laser de films de polyimide biocompatibles permet d'en réaliser pour de multiples applications.

Les puces sont mises en place de manière automatisée grâce au « Cube » garantissant l'interface entre le système chromatographique et le spectromètre de masse et permettant un positionnement optimal de la pointe de nébulisation. La puce y est installée fermement entre un stator et un rotor délivrant directement les fluides dans les canaux appropriés sur le même modèle de fonctionnement qu'une vanne traditionnelle.



**Figure 10** | Schémas du système nanoHPLC-Chip Agilent intégré permettant l'élimination des connexions complexes et délicates pour une diminution radicale des volumes morts (A). Disposée de manière automatisée en position orthogonale face à l'électrode d'entrée du spectromètre de masse (B), la puce est fermement enserrée entre un stator et un rotor composé d'une double microvanne permettant la distribution des fluides sans connexions (C).

Au laboratoire, la nanoHPLC-Chip a été couplée avec un spectromètre de masse HCT Ultra de type trappe d'ions (Bruker Daltonics). Une étude du couplage, réalisée lors de son installation, a déjà été détaillée dans un précédent travail de thèse<sup>9</sup> et une évaluation des performances dans le cadre d'analyses protéomiques d'échantillons peu complexes (gel 2-D) a déjà paru<sup>10</sup>. Leurs conclusions confirment la simplicité d'utilisation du système, une meilleure résolution et une meilleure séparation chromatographique que sur les systèmes classiques. Cela semble induire une qualité de spectres MS/MS supérieure et des temps d'analyse raccourcis. Par ailleurs, la sensibilité accrue du système favorise l'analyse de protéines minoritaires. Associés à une bonne répétabilité chromatographique, ces atouts présentent des perspectives prometteuses pour les projets de recherche de biomarqueurs, par exemple.

Nous avons choisi d'évaluer l'utilisation de ce système pour la comparaison du contenu protéique d'échantillons issus de gel 1-D ou pour la recherche de biomarqueurs. Une optimisation des paramètres chromatographiques a été réalisée sur des échantillons modèles. Il s'agissait :

- d'augmenter le contenu protéique total accessible lors d'une analyse nanoLC-MS/MS sur le système ;
- d'améliorer la qualité des identifications réalisées grâce à l'acquisition de plus d'information de fragmentations et de spectres de qualité.

▸ Le choix de la puce

Désignation	Colonne d'enrichissement	Colonne de séparation	Phase stationnaire
Protein ID chip #1	40 nL	75 µm x 43 mm	5 µm, C-18 SB-ZORBAX, 300A
Protein ID chip #2	40 nL	75 µm x 150 mm	5 µm, C-18 SB-ZORBAX, 300A

**Table 3** | Pucés microfluidiques HPLC-Chip Agilent Technologies disponibles au laboratoire pour la séparation de mélanges peptidiques.

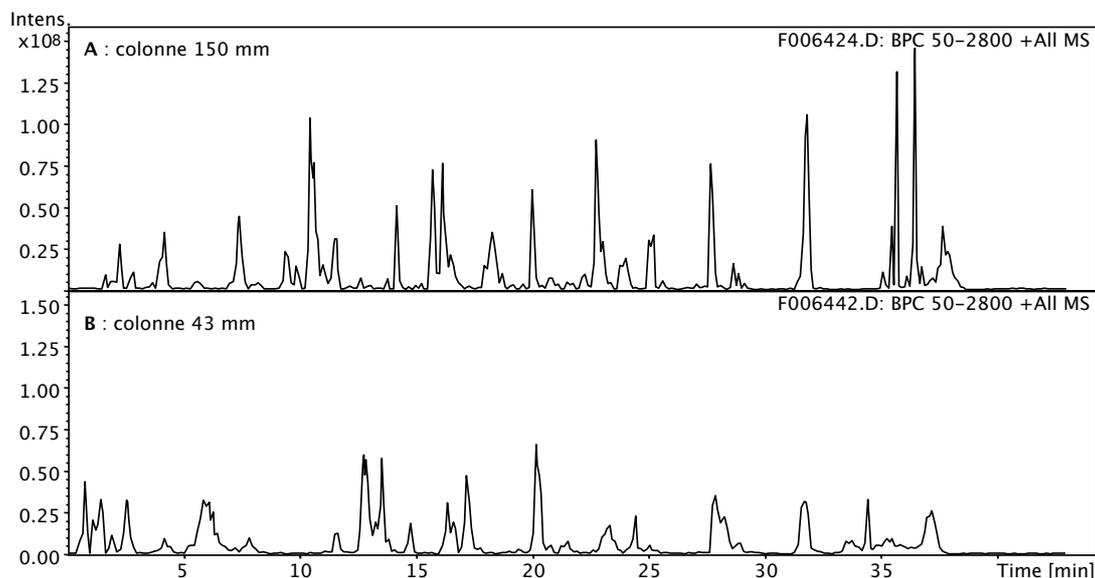
Plusieurs familles de pucés sont disponibles pour la séparation classique de mélanges peptidiques, l'enrichissement en phosphopeptides, la séparation de glycanes. D'autres pucés, permettant la LC 2-D par exemple, sont en cours de développement. Pour nos applications protéomiques, deux pucés de type phase inverse C-18 ont été testées (Table 3). La première, disposant d'une colonne de séparation d'une longueur de

<sup>9</sup> Carapito, C. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2006).

<sup>10</sup> Hardouin, J., et al. (2006). Rapid Commun Mass Spectrom **20**, 3236-44.

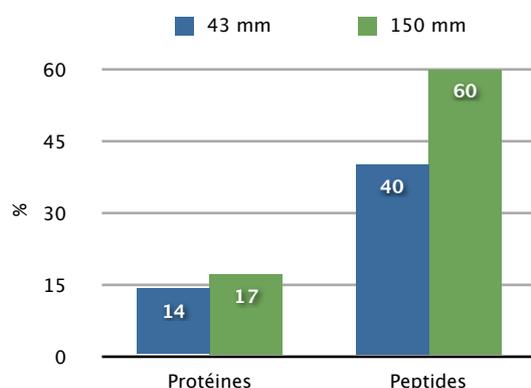
43 mm, est plutôt destinée à la séparation de mélanges peptidiques peu ou moyennement complexes. La seconde, de 150 mm, est indiquée pour la séparation de mélanges peptidiques moyennement ou très complexes.

Les performances chromatographiques des deux puces ont été évaluées à l'aide d'échantillons modèles : digests peptidiques de microparticules de lymphocytes déposées sur gel 1-D. En utilisant un même gradient (de 8 à 40 % d'acétonitrile en 30 min, soit une pente de  $1 \text{ \%} \cdot \text{min}^{-1}$ ), plusieurs échantillons ont été analysés sur chacune des puces et le nombre d'identifications de protéines et de peptides réalisées a été comparé.



**Figure 11** | Chromatogrammes de l'analyse d'une bande de gel 1-D sur le système HPLC-Chip-HCT Ultra avec une puce équipée d'une colonne de 150 mm (A) ou 43 mm (B).

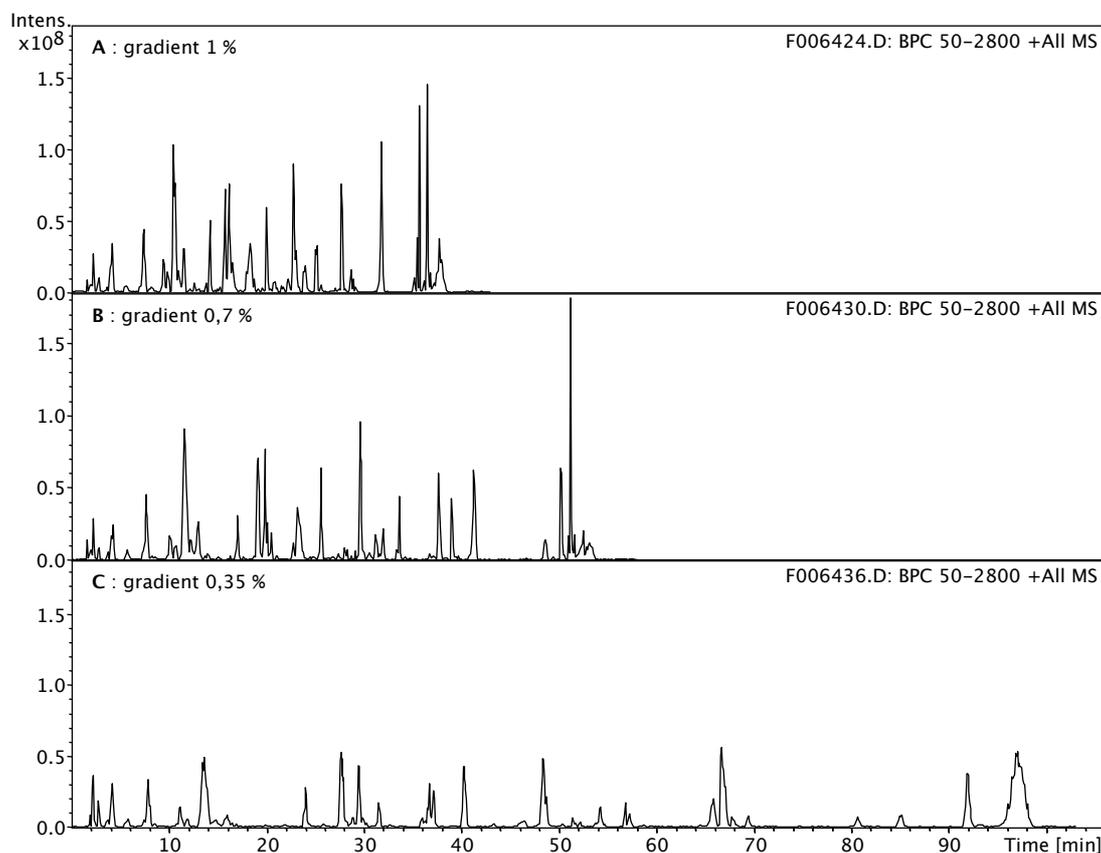
D'une manière générale, comme le montre le chromatogramme proposé en exemple (Figure 11), une colonne de 150 mm présente un profil chromatographique meilleur : les pics sont plus fins et mieux résolus que pour une puce de 43 mm. De fait, le système gagne en sensibilité lorsqu'une colonne longue est utilisée et les fractions chromatographiques contiennent potentiellement moins de peptides coélus. Ces observations sont vérifiées en comparant le nombre d'identifications de protéines et de peptides dans chaque cas (Figure 12). En utilisant une colonne de 150 mm, plus de protéines sont identifiées. Surtout, le nombre total de peptides sélectionnés est bien plus grand (50 % de plus). Il traduit une meilleure qualité des identifications grâce à une couverture de séquence plus importante. Des résultats équivalents ont été obtenus avec tous les échantillons testés. Une plus grande part du contenu protéique total d'une bande de gel 1-D peut donc être identifiée en utilisant une puce microfluidique équipée d'une colonne de 150 mm.



**Figure 12** | Nombre d'identifications de protéines et de peptides pour une bande de gel 1-D analysée sur le système HPLC-Chip-HCT Ultra avec une puce équipée d'une colonne de 150 ou 43 mm.

▸ L'optimisation des gradients

La deuxième étape de l'optimisation des performances chromatographiques pour l'analyse de mélanges peptidiques de complexité moyenne a consisté en l'adaptation des gradients d'éluion. En utilisant une puce équipée d'une colonne de 150 mm, nos échantillons modèles ont été analysés plusieurs fois avec des gradients entre 8 et 40 % d'acétonitrile avec des pentes de 1 %.min<sup>-1</sup>, 0,7 %.min<sup>-1</sup> et 0,35 %.min<sup>-1</sup> (Figure 13).

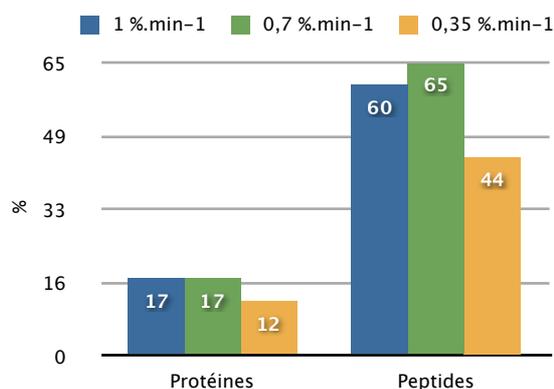


**Figure 13** | Chromatogrammes de l'analyse d'un digest du contenu protéique d'une bande de gel 1-D sur le système HPLC–Chip–HCT Ultra avec une puce équipée d'une colonne de 150 mm pour un gradient appliqué de pente 1 %.min<sup>-1</sup> (A), 0,7 %.min<sup>-1</sup> (B), 0,35 %.min<sup>-1</sup> (C).

Théoriquement, appliquer une pente faible permet de séparer plus de peptides et d'obtenir une meilleure résolution chromatographique. Mais lorsque la pente du gradient est trop ralentie, les pics chromatographiques finissent par s'élargir entraînant une perte en sensibilité et un effet de dilution. La comparaison du nombre de protéines et peptides identifiés pour un même échantillon confirme que moins d'identifications sont validées pour un gradient trop long (0,35 %.min<sup>-1</sup>, Figure 14). En revanche, si les gradients de 1 %.min<sup>-1</sup> et 0,7 %.min<sup>-1</sup> permettent d'identifier autant de protéines, le nombre de peptides associé est sensiblement plus important dans le deuxième cas. Des résultats similaires ont été obtenus pour tous les échantillons testés. En conséquence, des échantillons issus de bandes de gel 1-D pourront être analysés de façon optimale avec le gradient proposé ci-dessous (Table 4).

Temps (min)	0	45	46	49	50	58
% ACN	8	40	70	70	8	8

**Table 4** | Gradient d'acétonitrile optimisé pour la séparation d'un mélange peptidique complexe sur un système HPLC–Chip–HCT Ultra utilisant une puce équipée d'une colonne de 150 mm.



**Figure 14** | Nombre d'identifications de protéines et de peptides pour une bande de gel 1-D analysée sur le système HPLC-Chip-HCT Ultra avec une puce équipée d'une colonne de 150 mm pour un gradient appliqué de pente 1 %.min<sup>-1</sup>, 0,7 %.min<sup>-1</sup>, 0,35 %.min<sup>-1</sup>.

#### ▸ Le problème de pénurie d'acétonitrile et ses solutions

Au début de l'année 2009, l'acétonitrile est devenu une ressource rare et onéreuse : les prix ont été multipliés par 15 et une pénurie mondiale a touché tous les utilisateurs. Il n'existe aucun site dédié à la production d'acétonitrile qui est obtenue comme produit secondaire d'acrylonitrile, très utilisé par l'industrie automobile. Sur les systèmes HPLC Agilent, qui travaillent à des débits primaires élevés (200 à 500  $\mu\text{L}\cdot\text{min}^{-1}$ ), l'économie d'acétonitrile ne pouvait pas être une solution. Pour pallier ces manques, nous avons évalué une solution alternative en utilisant le méthanol comme solvant organique pour nos séparations chromatographiques. Si le méthanol est souvent déconseillé lors de l'utilisation de détecteurs UV (sa transparence étant moindre que celle de l'acétonitrile), il est parfaitement compatible pour un couplage HPLC-MS. Des études font même état d'un gain de sensibilité en masse et de gradients raccourcis pour l'analyse de peptides par nanoLC-MS/MS<sup>11,12</sup>.

Des tests semblables à ceux présentés précédemment ont été réalisés pour aboutir à un gradient optimisé, 2 fois plus court que celui utilisant l'acétonitrile, mais nécessitant un temps de rééquilibrage plus long. La sensibilité MS a été accrue d'un facteur 5 à 10, permettant l'analyse d'ions minoritaires avec des spectres de fragmentation de qualité. Cependant, les premiers éléments à disposition montrent une dégradation plus rapide des performances des puces microfluidiques – probablement due à la pression de travail supérieure liée à la moindre compressibilité du méthanol – et une instabilité handicapante de l'électrospray.

Temps (min)	0	5	20	21	26	27	42
% MeOH	10	30	60	85	85	10	10

**Table 5** | Gradient de méthanol optimisé pour la séparation d'un mélange peptidique complexe sur un système HPLC-Chip-HCT Ultra utilisant une puce équipée d'une colonne de 150 mm.

#### ▸ Les défauts de jeunesse

Le système HPLC-Chip Cube est une technologie récente qui a subi de nombreuses améliorations depuis son arrivée au laboratoire, bénéficiant de notre retour d'expérience notamment pour une utilisation permanente avec un débit d'analyse important.

▸ La technique et la composition de la métallisation de la pointe de nébulisation ont d'abord été revues pour améliorer l'application de la haute tension sur la pointe, qui influe directement sur la stabilité du spray. Le platine utilisé pour la métallisation des chips de première génération a été remplacé par l'or pour leur conférer plus de robustesse.

<sup>11</sup> Giorgianni, F., et al. (2004). Anal Chem **76**, 7028–38.

<sup>12</sup> Fritz, R., et al. (2009). Rapid Commun Mass Spectrom **23**, 2139–45.

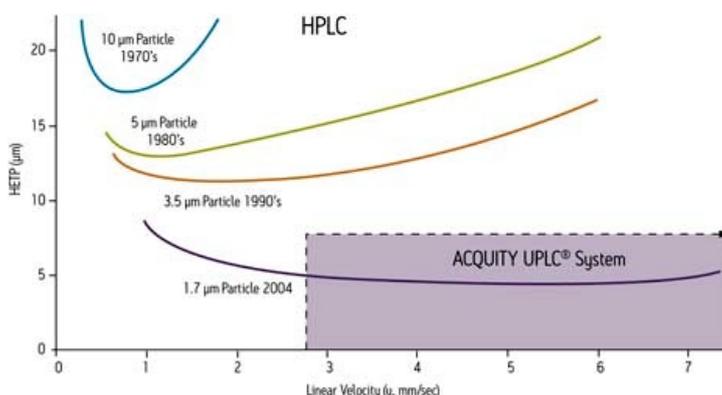
▸ Le point faible du système qui a reçu le plus de modifications est le dispositif rotor–stator enserrant la puce en opération. Le frottement du dispositif sur le revêtement polymère causait des problèmes de fuites, aggravées par l’oxydation des parties mécaniques qui se trouvaient en contact avec les solvants acides. Le revêtement des puces a été amélioré pour renforcer la zone de contact avec le rotor et le stator. Puis, le matériau métallique du bras permettant la rotation de la microvanne a été changé pour un matériau céramique insensible aux acides. Enfin, des évacuations pour d’éventuelles fuites ont été percées pour prévenir l’accumulation de liquide dans le dispositif.

▸ La troisième amélioration a porté sur le renforcement et le design émoussé de pièces constituant le rotor–stator qui se brisaient régulièrement, car soumises à des forces trop répétées.

## 2.1.2. Le système nanoAcquity UPLC (Waters Corporation)

### ▸ Présentation

Au début des années 2000, des études indiquant de nouvelles pistes pour améliorer les performances chromatographiques ont été publiées<sup>13, 14</sup>. Elles proposent d’utiliser de bien plus petites particules de phase stationnaire afin de diminuer la Hauteur Equivalente de Plateaux Théoriques (HEPT). Cela implique une meilleure efficacité chromatographique notamment en terme de résolution. Les courbes de Van Deemter (Figure 15) montrent en outre que cette efficacité peut être atteinte sur une gamme de débits beaucoup plus large qu’avec des particules de taille supérieure. Il en résulte que la vitesse de l’analyse peut être optimisée sans sacrifier la résolution chromatographique et la sensibilité.



**Figure 15** | Courbes de Van Deemter présentant l’amélioration des performances chromatographiques grâce à la diminution de la taille des particules. (Figure Waters Corporation)

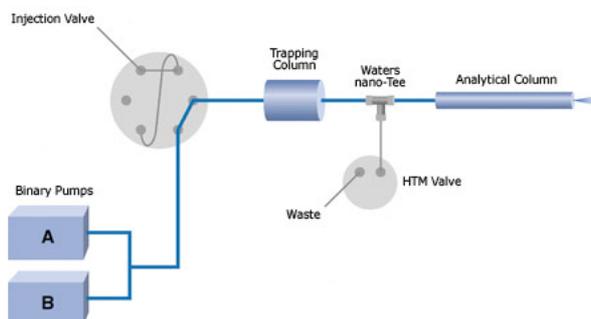
Waters Corporation commercialise depuis 2004 des colonnes chromatographiques à particules de 1,7 µm de diamètre et des systèmes capables de délivrer et supporter des solvants aux pressions de travail requises en UHPLC. Alors que les systèmes HPLC classiques travaillent à des pressions moyennes comprises entre 100 et 250 bar, les pressions en UHPLC doivent être de l’ordre de 500 à 800 bar pour atteindre une sensibilité, une résolution et un débit optimaux. Des efforts importants ont été réalisés dans la conception de pompes capables de compenser la compressibilité des solvants sur de larges variations de pression. Les solvants sont délivrés directement au débit requis, sans vanne de fuite, ce qui permet une économie de solvant de plus de 95 % par rapport aux systèmes classiques. Des vannes, colonnes et capillaires supportant des pressions élevées (jusqu’à 1000 bar) ont également été conçus avec le souci de réduire au minimum les volumes morts : le volume du système entier n’exède pas 120 µL.

Notre installation étant dévolue à l’analyse de mélange peptidiques, nous l’avons équipé d’une colonne analytique Acquity UPLC BEH130 C18 à particules de 1,7 µm, de 75 µm de diamètre interne et d’une longueur de 200 mm. Une colonne de préconcentration Symmetry C18 à particules de 5 µm, de 180 µm de

<sup>13</sup> Jerkovich, A. D., et al. (2003). *Lc Gc Europe* **16**, 20–23.

<sup>14</sup> Wu, N., et al. (2001). *J Chromatogr A* **911**, 1–12.

diamètre interne et d'une longueur de 20 mm est installée en amont (Figure 16). Nous avons choisi d'utiliser un nébuliseur électrospray à jonction liquide, plus robuste et fiable que les nano-aiguilles fragiles et onéreuses (maximum 24 h de durée de vie) ; la différence de volume mort n'étant pas dramatique (20 nL).



**Figure 16** | Schéma du module de préconcentration et de séparation nanoAcquity UPLC. Les solvants sont délivrés directement au débit d'analyse (400 nL/min) régulé par des contrôleurs et sans vanne de fuite. La colonne analytique est en ligne avec la colonne de préconcentration et de dessalage, le chargement et le dessalage de l'échantillon étant permis grâce à l'ouverture d'une vanne d'évacuation entre les deux colonnes. (Figure Waters Corporation)

#### ▸ L'optimisation des gradients

La configuration de ce système permet de combiner résolution, vitesse et sensibilité même pour des mélanges peptidiques complexes.

▸ La réduction considérable des volumes morts et des volumes de capillaire réduit la dispersion des pics chromatographiques de manière remarquable. Mais c'est surtout de la taille des particules (1,7  $\mu\text{m}$ ) que le système tire le plein bénéfice pour un gain d'efficacité. Ces particules offrent théoriquement une résolution améliorée de 70 % par rapport à des particules de 5  $\mu\text{m}$  (selon Van Deemter, la résolution étant proportionnelle à la racine carrée de l'inverse de la taille des particules). Ainsi, avec l'UHPLC, la largeur des pics chromatographiques a pu être réduite à 5 s par rapport aux pics de 15 à 30 s obtenus avec les systèmes nanoHPLC classiques.

▸ Avec des particules inférieures à 2  $\mu\text{m}$ , le débit de solvant peut être multiplié par 3 sans perte d'efficacité ou de résolution. Nous avons testé plusieurs débits et comparé les résolutions chromatographiques obtenues lors de l'analyse de mélange peptidiques complexes. Une finesse des pics optimale et une séparation idéale entre deux pics consécutifs sont obtenues pour un débit de 400 nL/min. Des débits plus rapides permettraient d'améliorer encore la finesse des pics, mais un compromis doit être trouvé pour que la durée d'éluion d'une fraction chromatographique soit compatible avec la durée d'un cycle MS/MS.

▸ Un gradient optimal pour la séparation de mélanges peptidiques complexes a été mis au point pour tirer profit des avantages du système tout en s'assurant obtenir une analyse la plus complète possible de mélanges peptidiques complexes pour l'identification d'un maximum de protéines et peptides (Table 4).

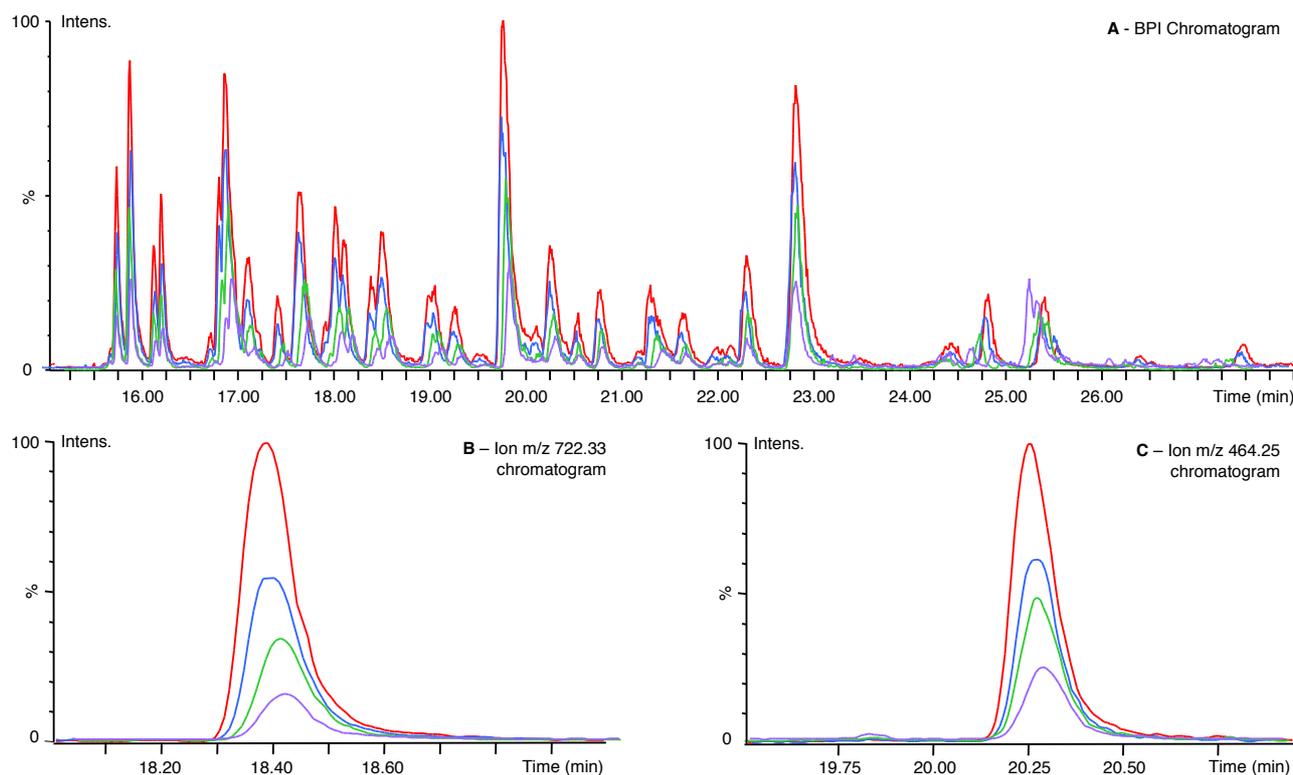
Temps (min)	0	35	36	37	50
% ACN	1	50	90	1	1

**Table 4** | Gradient d'acétonitrile optimisé pour la séparation d'un mélange peptidique complexe sur un système nanoAcquity UPLC-Synapt HDMS.

#### ▸ Evaluation de la sensibilité du système

La sensibilité du système nanoAcquity UPLC a été évaluée en injectant un mélange de peptides tryptiques issus d'un digest de BSA (Bovine Serum Albumin) en quantité croissante de 10 à 40 fmol (Figure 17). 36 peptides tryptiques appartenant à la BSA ont pu être identifiés sans difficulté suite à l'injection de 10 fmol de digest. Ces tests de sensibilité ont été réalisés en mode MS, mais aussi en mode MS/MS afin de

juger du recouvrement de séquence et de la qualité des spectres MS/MS obtenus sur le digest de BSA (Table 6, données issues de recherches Mascot® à partir des fichiers « peak list » générés).



**Figure 17** | (A) Evaluation de la sensibilité du système nanoAcquity UPLC-Synapt HDMS avec une pré-colonne d'enrichissement (4mm, 40 nl, Zorbax 300SB-C18, 5µm) et une colonne analytique de 75 µm de diamètre interne (43 mm, Zorbax 300SB-C18, 5µm). Des quantités croissantes – de 10 fmol à 40 fmol – de peptides tryptiques de BSA ont été injectées. (B) Zoom sur deux chromatogrammes d'ions extraits pour les peptides de m/z = 922,33 et de m/z = 464,25. Légende : — 10 fmol, — 20 fmol, — 30 fmol, — 40 fmol

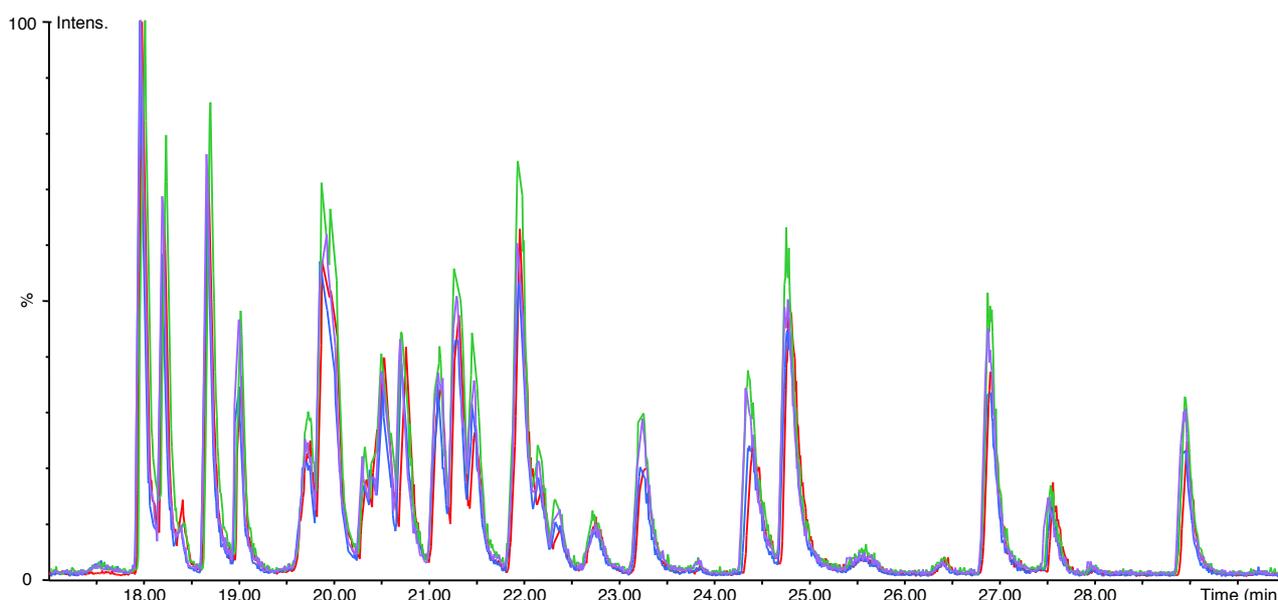
Quantité de BSA injectée (fmol)	10	20	30	40	50
% de recouvrement de séquence	54	54	52	52	61
Nombre de peptides uniques identifiés	36	36	34	34	40

**Table 6** | Pourcentage de recouvrement de séquence obtenu en MS/MS et nombre de peptides identifiés pour des quantités croissantes de digest de BSA injectées.

► Evaluation de la répétabilité chromatographique

La répétabilité du système a été évaluée par des injections successives d'un mélange de peptides tryptiques issus de la digestion de BSA. La figure 18 présente la superposition des chromatogrammes obtenus pour 4 injections successives de 50 fmol de digest.

La répétabilité, même sur des gradients relativement longs comme celui que nous avons mis au point, est excellente. Cependant, après un arrêt du système jusqu'à 5 cycles peuvent être nécessaires pour retrouver une équilibration satisfaisante et un profil chromatographique répétable. Cette inertie du système impose une vérification méticuleuse de ses performances avant toute série d'analyses.



**Figure 18** | Superposition de 4 chromatogrammes de 4 injections successives de 50 fmol d'un digest de BSA acquis après un grand nombre de cycles d'analyse.

### 2.1.3. Conclusion

Les efforts technologiques consentis par les constructeurs et le développement de méthodes adaptées aux analyses protéomiques par nanoLC-MS/MS de mélanges peptidiques complexes ont abouti ces dernières années à une amélioration spectaculaire des performances chromatographiques. Le laboratoire a acquis l'HPLC-Chip Cube Agilent Technologies en 2006 et la nanoAcquity UPLC en 2008, ouvrant de nouvelles perspectives tant la résolution et la sensibilité atteintes par ces systèmes sont prometteuses pour l'identification de protéines minoritaires dans des mélanges complexes. Il faut toutefois se satisfaire d'un compromis : une résolution chromatographique trop fine impliquant une perte de données si la rapidité des cycles MS/MS n'est pas adaptée.

## 2.2. Une exactitude accrue de la mesure de masse grâce au système « *lock-mass* »

Les spectromètres de masses de type Q-TOF sont particulièrement utilisés en analyse protéomique pour leurs performances résolutive et leur précision en masse. Toutefois, à l'inverse des trappes d'ions, ils sont très enclins à une dérive d'étalonnage. Elle est en grande partie due aux variations de température de l'environnement d'analyse qui implique une dilatation du tube de vol de l'analyseur TOF<sup>15</sup>. Un étalonnage avec une solution de référence sur toute la gamme d'analyse en  $m/z$  – avant et après une série d'acquisitions – est nécessaire. A haut débit, cet étalonnage n'est cependant pas suffisant pour conserver une exactitude de masse en MS et MS/MS tout au long de la série d'analyse, ce qui peut empêcher l'identification de nombreux peptides et protéines et favoriser l'apparition de faux-positifs. Pour obtenir des mesures de masse d'une bonne exactitude lors d'analyses de routine, il est nécessaire d'introduire une référence dans le spectromètre de masse afin de corriger la dérive de l'étalonnage en masse<sup>16</sup>. Le laboratoire a récemment acquis une nouvelle source nanoélectrospray développée par Waters Corporation qui permet, grâce à un

<sup>15</sup> Chernushevich, I. V., et al. (2001). *J Mass Spectrom* **36**, 849–65.

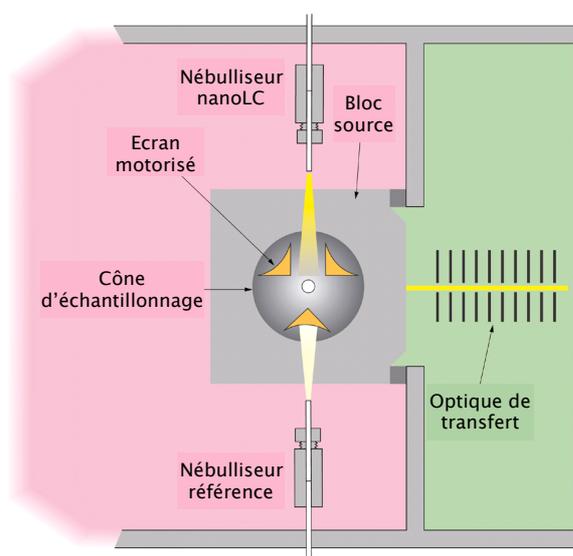
<sup>16</sup> Charles, L. (2003). *Rapid Commun Mass Spectrom* **17**, 1383–8.

système « *lock-mass* », de réaliser des mesures exactes de masses. Elle est installée sur notre couplage nanoAcquity UPLC–Synapt HDMS.

### 2.2.1. Présentation et fonctionnement d'un système *lock-mass* : le *LockSpray*<sup>™</sup> Waters Corporation

La source nanoélectrospray équipée du système *LockSpray* est conçue spécialement pour les sources en Z dans lesquelles les ions sont défectés deux fois avant d'entrer dans le spectromètre de masse. Le système *LockSpray* est composé de 2 nébulisateurs nanoélectrospray : l'un est couplé à la nano-chromatographie liquide, l'autre permet d'introduire directement une solution de référence infusée en continu. Chacun des deux nébulisateurs est maintenu séparé à tout moment grâce à un écran motorisé, piloté par le logiciel (Figure 19).

L'avantage majeur de ce dispositif est la conservation de l'intégrité des données tout au long de l'analyse ; l'information produite par chacun des nébuliseurs étant séparée de l'autre. Au contraire, les systèmes qui introduisent la référence dans le flux post-colonne imposent des volumes morts supplémentaires et causent une perte de résolution chromatographique dommageable. Ils peuvent être par ailleurs à l'origine d'effets de suppression de signal et d'interférences puisque référence et analyse se retrouvent confondues sur les spectres acquis<sup>17</sup>.



**Figure 19** | Schéma du dispositif *LockSpray*. Au cours d'une analyse, l'écran motorisé bascule à des intervalles de temps prédéterminés pour permettre l'acquisition de spectres de références qui seront utilisés *a posteriori* pour corriger l'étalonnage des données mesurées.

### 2.2.2. Mise en place et évaluation du bénéfice apporté par le système *LockSpray*

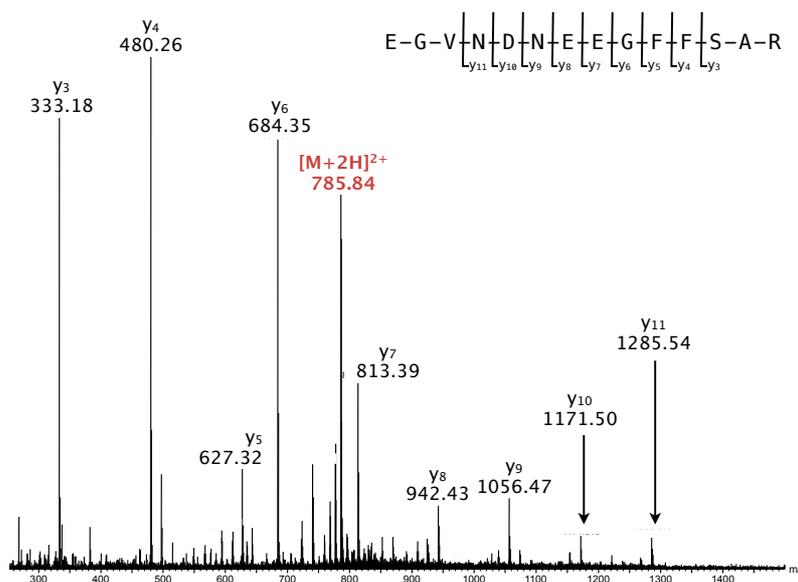
#### ▸ Optimisation du système

Le système est simple à mettre en place et est complètement automatisé. Cependant, un positionnement fin des nébuliseurs et de l'écran motorisé est nécessaire pour éviter la formation d'arcs électriques entre ces différents éléments confinés dans un volume restreint.

Sur le plan expérimental, le travail d'optimisation à réaliser réside essentiellement dans la détermination de l'échantillonnage idéal de la référence. Il doit être suffisant pour qu'une mesure correcte puisse être réalisée, mais sans perturber trop fortement les acquisitions nanoLC–MS/MS. Les paramètres du système ont été ajustés en fonction des observations qui ont pu être faites sur les premières séries d'analyse réalisées sur l'instrument.

<sup>17</sup> Eckers, C., et al. (2000). *Anal Chem* 72, 3683–8.

La référence utilisée est une solution de [Glu1]-fibrinopeptide B (GFP) à 200 fmol/ $\mu\text{L}$ . Toutes les 30 s, l'écran motorisé bascule devant le nébullisateur nanoLC pour permettre l'acquisition d'un spectre de référence en 0,5 s. La mesure est réalisée en mode de fragmentation de sorte que le peptide  $[M+H]^{2+}$  à 785,844 m/z et ses fragments (entre 72 et 1285 m/z) puissent être utilisés pour le réétalonnage des spectres MS comme des spectres MS/MS (Figure 20). Le logiciel de pilotage du système acquiert les données issues de l'analyse et de la référence dans deux canaux différents, afin que les informations soient toujours séparées et puissent être utilisées indépendamment. A l'issue de l'analyse, lors de la création des fichiers « *peak list* » à partir des données brutes, une correction de l'étalonnage de chaque spectre MS et MS/MS est réalisée en fonction des références mesurées avant et après le spectre considéré. Ainsi, grâce à la source *LockSpray*, l'information envoyée aux moteurs de recherche contiendra des mesures de masses exactes à moins de 10 ppm près, ce qui doit permettre d'accroître le nombre et la confiance des identifications de peptides et protéines analysés.

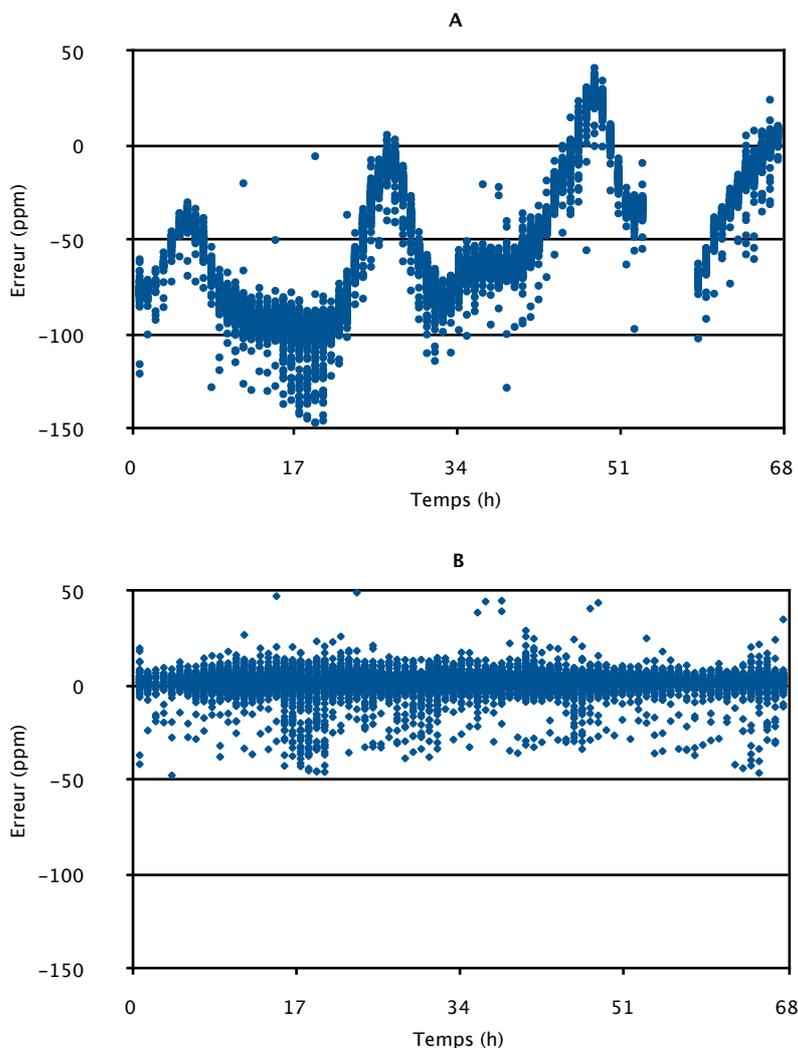


**Figure 20** | Spectre de référence acquis en mode de fragmentation de l'ion  $[M+2H]^{2+}$  du GFP, toutes les 30 s au cours d'une analyse nanoLC-MS/MS. La correction de l'étalonnage des spectres d'analyse est réalisée à partir des mesures références acquises avant et après chaque spectre corrigé.

#### ► Evaluation du bénéfice apporté par l'utilisation du système *LockSpray*

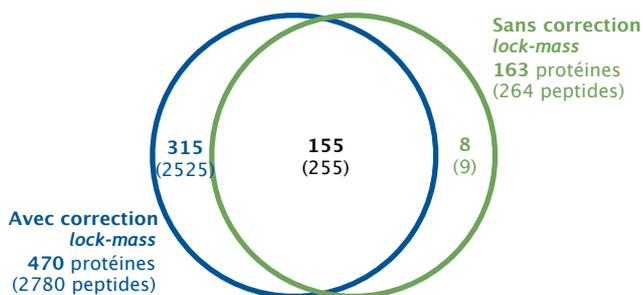
Premier système de ce type disponible au laboratoire sur nos systèmes Q-TOF et permettant la correction permanente de l'étalonnage, la source *LockSpray* a immédiatement montré son intérêt. Même sur les couplages récents, bien que l'enceinte soit calorifugée, on note généralement une forte fluctuation de l'étalonnage en fonction de la température de la pièce de travail, même si celle-ci est climatisée. La température de l'électronique embarquée semble également influencer fortement sur l'exactitude des mesures.

Nous avons mis en évidence l'utilité du système pour des mesures de masses exactes sur une longue série d'analyses. La figure 21 montre l'erreur mesurée sur les peptides identifiés au cours de l'analyse de bandes de gel 1-D issues d'une piste complète. Sur une durée totale de 68 heures, on note clairement que sans utiliser la correction, la fluctuation de l'erreur correspond à l'évolution de la température dans la pièce. L'analyse ayant commencé à 23h, on suit sur le graphe l'aggravation de l'erreur lorsque l'atmosphère se rafraîchit en milieu de nuit et une évolution inverse au cours de la journée. A ces effets de température, s'ajoute une dérive constante de l'étalonnage de l'instrument. En corrigeant les mêmes données avec les informations acquises sur le canal référence, les mesures de masses sont exactes sur toute la durée de l'analyse. La moyenne quadratique des erreurs mesurées (RMS) est proche de 64 ppm pour l'analyse non corrigée et d'à peine plus de 1 ppm lorsque la *lock-mass* est utilisée.



**Figure 21** | Erreurs sur les mesures de masse en MS des peptides identifiés au cours d'une longue série d'analyses. (A) Sans réétalonnage *LockSpray* ; RMS = 64,68 ppm. (B) Avec réétalonnage *LockSpray* ; RMS = 1,19 ppm.

Les moteurs de recherche imposent de fixer une fenêtre d'erreur maximale sur les peptides parents et fragments. Une tolérance trop importante implique des identifications moins fiables et un taux de faux positifs potentiellement plus grand. Au contraire, une exigence trop forte sur l'exactitude des mesures exclut du résultat final des identifications de valeur (faux négatifs). Il est par conséquent nécessaire de bien connaître les performances de l'instrument en terme d'exactitude de mesure de masse. Pour l'illustrer, les données utilisées précédemment, corrigées et non corrigées par les informations de la *lock-mass*, ont été soumises avec une tolérance de 50 ppm sur les ions parents et de 0,1 Da sur les fragments (Figure 22). Un mauvais étalonnage ou une fenêtre d'erreur mal réglée met en péril la qualité et le nombre d'identifications.



**Figure 22** | Nécessité d'une mesure de masse exacte en MS et MS/MS pour l'identification correcte de protéines et peptides. Echantillons issus d'une piste de gel 1-D découpée de manière systématique et analysés 2 fois par nanoLC-MS/MS sur le couplage nanoAcquity UPLC-Synapt HDMS.

## 2.3. L'optimisation de l'acquisition des données en spectrométrie de masse

Comme il l'a été décrit au point 1.1.1., les acquisitions nanoLC-MS/MS utilisent de manière générale le mode d'analyse données-dépendant (DDA) grâce auquel la sélection des ions parents pour la fragmentation est réalisée de manière automatique en fonction de critères optimisés par l'utilisateur.

Selon le couplage utilisé, les pics chromatographiques ont une largeur comprise entre 5 et 15 s. Or le nombre d'ions précurseurs sélectionnés conditionne la durée d'un cycle MS+MS/MS. Celui-là doit être adapté à la largeur des pics chromatographiques afin d'obtenir un maximum d'informations de fragmentation. Pour optimiser la compatibilité de la durée des cycles d'acquisition avec les performances chromatographiques, un travail de mise au point des paramètres d'acquisition influant directement sur cette durée a été effectué sur notre nouveau système nanoAcquity UPLC-Synapt HDMS.

### 2.3.1. Les paramètres physiques de détection et sélection des ions parents

#### ▸ Le type de peptides sélectionnés

Les peptides issus de la digestion trypsique sont généralement doublement (espèces majoritaires à 65–80 %<sup>18</sup>) ou triplement chargés lorsqu'ils sont analysés en mode d'ionisation ESI positif : ils présentent au moins deux sites de protonation, un sur l'amine N-terminale et l'autre sur la chaîne latérale de la lysine ou de l'arginine du côté C-terminal. Pour favoriser la fragmentation des ions doublement et triplement chargés – qui fragmentent mieux que les ions monochargés<sup>19</sup> – une option permettant de privilégier leur sélection, même en présence d'ions monochargés, peut être activée. Des ions monochargés seront alors sélectionnés uniquement si aucun ion doublement ou triplement chargé ne dépasse le seuil de sélection.

#### ▸ Le seuil de détection des ions parents

Le basculement du mode MS au mode MS/MS s'opère de manière automatique dès que l'intensité d'un ion est supérieure à une valeur seuil. Ce seuil doit être finement optimisé : sa valeur conditionne le moment de la sélection des ions parents. Lorsqu'il est fixé trop bas, les ions parents sont sélectionnés avant leur maximum d'intensité, la fragmentation ne produisant alors pas l'information la plus informative. Lorsque le seuil est fixé trop haut, l'information sur des ions minoritaire risque d'être perdue.

### 2.3.2. Les paramètres temporels des acquisitions MS et MS/MS

#### ▸ Temps de balayage

Parmi les paramètres d'acquisition MS+MS/MS, ceux fixant les temps de balayage sont les plus critiques. En effet, un balayage trop rapide en MS ou MS/MS génère des spectres trop peu intenses, mal résolus et difficilement interprétables. Au contraire, un balayage lent produit des spectres de bonne qualité, au détriment du nombre d'ions analysés.

Par ailleurs, la durée totale d'un cycle MS/MS est fonction du nombre d'ions sélectionnés à l'issue de la MS. Ce nombre est fixé (entre 1 et 8) en cohérence avec la durée de balayage MS/MS choisie. La règle générale est qu'un compromis doit être trouvé pour adapter la durée des cycles MS+MS/MS à la largeur des pics chromatographiques et éviter ainsi de perdre l'information.

#### ▸ Exclusion dynamique

Pour éviter la sélection répétée d'un ion qui serait largement majoritaire dans une analyse et qui empêcherait ainsi la sélection d'ions minoritaires, l'option d'exclusion dynamique doit être paramétrée. Elle

<sup>18</sup> Savitski, M. M., et al. (2005). *Mol Cell Proteomics* **4**, 1180–8.

<sup>19</sup> Dongre, A. R., et al. (1996). *J Mass Spectrom* **31**, 339–50.

permet d'ajouter à des listes d'exclusion (temporaires ou définitives) les masses des ions parents après qu'ils ont été analysés. Le paramétrage de la fenêtre d'exclusion doit être prudent, car plusieurs peptides tryptiques de séquence différente peuvent être isobares. Il faut donc éviter de perdre des informations sur ces peptides en fixant des fenêtres d'exclusion trop larges et il est plus pertinent d'utiliser l'exclusion de façon temporaire en interdisant la sélection pendant un temps donné puis de l'autoriser à nouveau. Il est en effet plus rare que des peptides aient la même masse et le même temps de rétention chromatographique.

### 2.3.3. Les paramètres de fragmentation des ions parents

Les options de fragmentation des ions parents sélectionnés pour la MS/MS doivent également être finement réglées. Plusieurs possibilités s'offrent à l'utilisateur qui a généralement le choix entre :

- une fragmentation identique quelque soit l'ion sélectionné par l'application d'un gradient de tension dans la chambre de fragmentation
- ou une fragmentation différente selon la masse et/ou la charge de l'ion parent par l'application de tensions successives adaptées.

Cette deuxième option a montré de meilleurs résultats pour la qualité des spectres de fragmentation : des tensions adaptées à la taille du peptide isolé permettent de lui communiquer l'énergie adéquate pour une fragmentation optimale. La qualité des spectres peut de surcroît être améliorée en utilisant successivement deux tensions différentes : une « faible » qui favorise la formation de gros fragments (dans une gamme de m/z haute) et une plus « forte » qui favorise les fragments plus petits (dans une gamme de m/z basse).

### 2.3.4. Proposition d'une feuille de paramètres optimisés

En conclusion, en prenant en compte tous les paramètres interdépendants qui viennent d'être détaillés, un compromis doit être trouvé pour perdre le moins d'informations sur des ions minoritaires et obtenir le plus d'informations sur les ions sélectionnés. Notre nouveau couplage nanoAcquity UPLC–Synapt HDMS produit des pics chromatographiques très résolus et dispose d'une capacité d'acquisition rapide. Ainsi, il est possible de paramétrer des temps de balayages relativement courts, de sélectionner jusqu'à 4 ions et d'analyser leurs fragments produits grâce à 2 énergies complémentaires (Table 7). La durée d'un cycle MS +MS/MS demeure alors compatible avec la largeur d'un pic chromatographique (environ 5 s à mi-hauteur) :

$$\text{Durée MS} + \text{Durée MS/MS} \times 4 \text{ ions} \times 2 \text{ énergies} = \text{Durée 1 cycle MS+MS/MS}$$

$$0,5 + 0,6 \times 4 \times 2 = 5,3 \text{ s}$$

<b>Sélection des ions parents</b>	Nombre d'ions sélectionnés par cycle Intensité seuil pour activer le mode MS/MS Intensité limite pour revenir en mode MS	4 40 coups/s 15 coups/s
<b>Paramètres temporels</b>	Balayage MS Balayage MS/MS Durée maximale d'une mesure MS/MS  Durée de l'exclusion dynamique Fenêtre d'exclusion	0,5 s 0,6 s 1,25 s  15 s ± 2 Da
<b>Fragmentation des ions parents</b>	<b>m/z</b> 300–500 501–600 601–700 701–800 801–900 901–1000 1001–1200 1201–1500	<b>V<sub>1</sub> (V)</b> 14 19 24 28 32 39 45 55 <b>V<sub>2</sub> (V)</b> 18 24 28 32 39 45 55 60

**Table 7** | Paramètres d'acquisition optimisés pour une analyse nanoLC–MS/MS d'un échantillon protéique complexe sur le système nanoAcquity UPLC–Synapt HDMS (Waters Corp.).

## 2.4. Conclusion

A travers les dernières innovations instrumentales pour l'analyse protéomique par nanoLC-MS/MS, les performances chromatographiques et l'acquisition des données se sont grandement améliorées. Toutefois, les possibilités techniques ne sont rien sans le nécessaire travail d'optimisation et de paramétrage. Des pistes d'optimisation pour trouver le meilleur compromis entre qualité et quantité de résultats ont été proposées. Mais elles ne doivent pas rester figées : il est toujours nécessaire de remettre en cause les facteurs critiques qui doivent être adaptés en fonction de la nature, de la complexité des échantillons à analyser et des objectifs à relever.

- ▶ Ces dernières années, des innovations technologiques en chromatographie liquide ont permis d'atteindre des performances inégalées :
  - La chromatographie liquide sur puce microfluidique, en éliminant les volumes morts importants, permet d'atteindre de meilleures résolution et sensibilité,
  - La chromatographie liquide ultra-haute performance (UHPLC), en mettant en oeuvre des tailles de particules inférieures à 2  $\mu\text{m}$ , travaille à des débits supérieurs à 300 nL/min et gagne en résolution et en sensibilité ;
- ▶ L'utilisation d'un système de correction permanente de l'étalonnage pour les analyseurs TOF permet d'atteindre des précisions de masse en MS inférieures à 2 ppm ;
- ▶ Les paramètres d'acquisition en spectrométrie de masse doivent être finement adaptés en fonction de l'échantillon, de système et des objectifs analytiques afin de trouver le meilleur compromis entre qualité de l'information acquise et quantité de résultats générés.

### 3. De nouveaux outils bio-informatiques améliorent le nombre et la crédibilité des identifications de protéines

Les outils bioinformatiques pour l'analyse protéomique représentent aujourd'hui une large gamme d'applications allant de simples outils de digestion *in silico* de protéines à des logiciels sophistiqués pour les études *de novo* ou la comparaison du contenu protéique d'échantillons<sup>20,21</sup>. Ces outils interviennent à tous les niveaux de l'analyse protéomique : de la préparation des analyses, à l'interprétation et la validation des identifications, en passant par le traitement des données brutes avant leur exploitation.

La suite de ce chapitre décrit le paramétrage et l'utilisation des outils utilisés dans nos stratégies d'identification et de validation. Il détaille également l'évaluation de nouvelles approches en devenir.

#### 3.1. Les paramètres de traitement des données brutes

Une fois une analyse nanoLC-MS/MS réalisée, une étape d'extraction des données brutes est indispensable pour soumettre l'information aux moteurs de recherche en vue de l'identification du contenu protéique de l'échantillon. Cette extraction est une étape critique, car elle affecte tout le traitement des données et donc la qualité, la quantité et la fidélité des résultats. Alors que beaucoup de méthodes ont été proposées dans la littérature, peu de solutions existent en terme d'offre logicielle<sup>22</sup>.

Les constructeurs fournissent généralement un algorithme intégré à leur plateforme de gestion des données. Il est alors indispensable d'ajuster les paramètres pour obtenir l'information la plus exploitable et offrant les meilleurs résultats. C'est ce qui est proposé ici pour nos algorithmes commerciaux inclus dans ProteinLynxGlocalServer™ (Waters Corp.) et DataAnalysis™ (Bruker Dal.).

##### 3.1.1. La génération des fichiers de « listes de masses » ou « *peak list* »

Les fichiers bruts contiennent toute l'information acquise lors d'une analyse nanoLC-MS/MS sous forme binaire (chromatogrammes et spectres MS et MS/MS), exploitable uniquement par des logiciels propriétaires. Pour utiliser l'information utile pour la recherche dans les banques de données via les moteurs de recherche, il est nécessaire d'extraire la masse et l'intensité des peptides parents, des fragments associés, et éventuellement les temps de rétention chromatographique correspondants. Cette information est rassemblée dans un fichier « *peak list* » qui peut être soumis aux moteurs de recherche. Ces opérations nécessitent plusieurs étapes de traitement du signal contenu dans les fichiers bruts<sup>23</sup> :

- Une étape de correction de l'étalonnage de chaque spectre (éventuellement)

Pour les instruments équipés d'un système de correction de l'étalonnage, il est nécessaire d'indiquer le nombre de spectres de références à utiliser (3 par exemple) autour de chaque spectre considéré pour calculer sa correction. Une fenêtre de tolérance en  $m/z$ , dans laquelle l'ion référence est recherché, doit aussi être fixée. Si cette fenêtre est trop étroite et que la dérive de l'étalonnage est trop importante, la correction n'a pas lieu puisque l'ion référence n'est pas trouvé par l'algorithme. Une fenêtre de quelques daltons doit donc être choisie pour assurer une bonne correction de l'étalonnage.

<sup>20</sup> Palagi, P. M., et al. (2006). *Proteomics* 6, 5435–44.

<sup>21</sup> Lisacek, F., et al. (2006). *Proteomics* 6, 5445–66.

<sup>22</sup> Matthiesen, R., et al. (2008). *Methods Mol Biol* 453, 105–22.

<sup>23</sup> Matthiesen, R. (2007). *Proteomics* 7, 2815–32.

▸ Une étape d'élimination du bruit de fond et de lissage des spectres

Le bruit de fond correspond à une information aléatoire, généralement d'intensité faible et variant peu. L'algorithme de suppression du bruit calcule une fonction, dont l'ordre et le seuil de suppression sont fixés par l'utilisateur. Le lissage itératif permet d'adoucir les variations rapides d'intensité et d'améliorer la détection des pics. On dispose généralement de deux options : un lissage « moyen » ou « de Stavitsky-Golay ». Ce dernier est préféré pour conserver la largeur à la base des pics.

▸ Une étape de déconvolution et de désisotopage

La déconvolution correspond à la détermination de l'état de charge de chaque ion : la résolution isotopique étant généralement observée pour des peptides ou fragments faiblement chargés (1 à 3 fois), la mesure de l'état de charge est directe en calculant la différence de  $m/z$  entre deux isotopes consécutifs. Le désisotopage remplace ensuite chaque cluster d'ions par une unique mesure de masse<sup>24</sup>. Les isotopes sont effacés et des spectres d'ions monochargés et monoisotopiques sont obtenus.

▸ Une étape de calcul du centroïde des pics

La dernière étape de traitement des spectres avant l'export de l'information est la transformation des pics sous forme de « bâtons », centrés sur la valeur de  $m/z$  déterminée à partir de leur apex à 80 %. Cette limite peut être modifiée par l'utilisateur.

### 3.1.2. Optimisation des paramètres d'extraction des données brutes avec ProteinLynxGlobalServer (Waters Corp.)

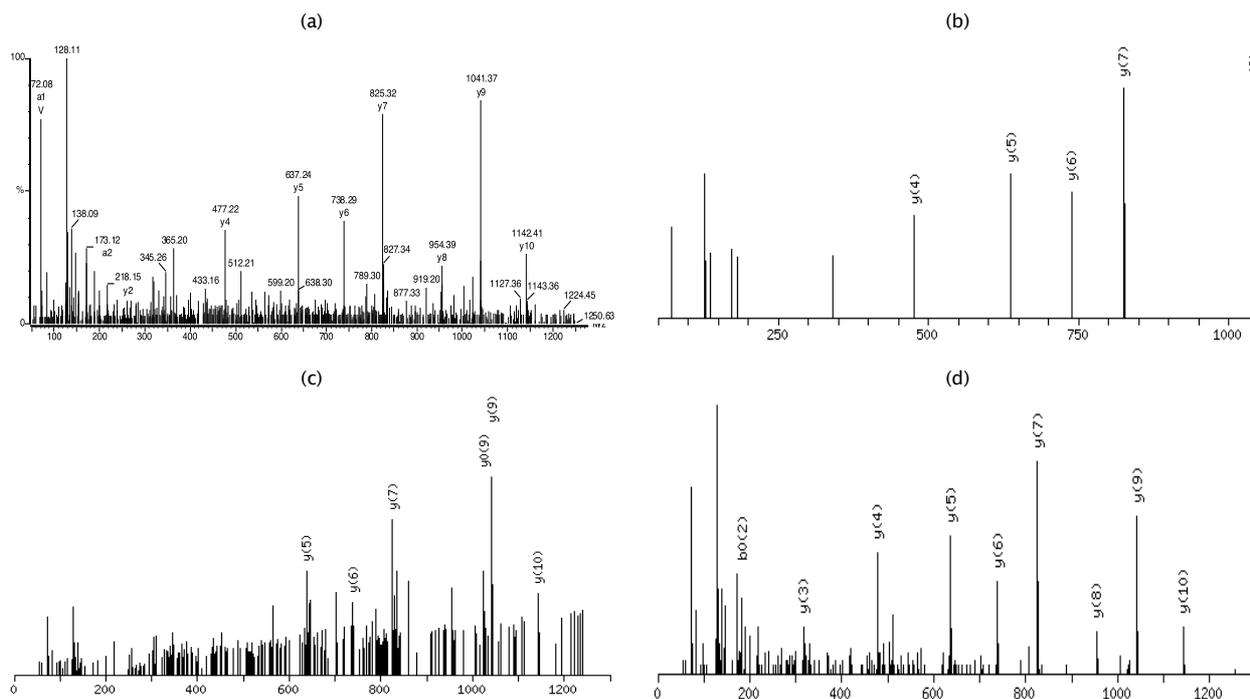
Les instruments Waters Corp. – tel notre système Synapt HDMS – nécessitent l'utilisation du logiciel PLGS (ProteinLynxGlobalServer) pour générer des fichiers de « listes de masses » de type « *peak list* » (.pkl), exploitables par les moteurs de recherche.

Après avoir remarqué que des spectres de mauvaise qualité étaient exportés avec la méthode standard, nous avons souhaité affiner les paramètres d'extraction des données brutes en réalisant de multiples essais de paramétrages (Table 8). Les résultats montrent l'énorme influence de ces options sur la qualité de l'information soumise aux moteurs de recherche (Figure 23). Par ailleurs, les résultats d'identification sont fortement compromis lorsque les paramètres d'extraction des données sont inadaptés.

		Méthode	Stand.	#2	#3	#4	#5	#6
P A R A M	Deisotopage		✓	✓	✗	✗	✗	✗
	Suppression du bruit de fond Seuil		Auto	35 %	35 %	5 %	35 %	5 %
	Lissage		✗	✗	✓	✓	✗	✗
R E S U L T A T S	Nombre de protéines identifiées		2	3	5	6	3	4
	Trypsine Score Mascot		258	337	695	803	664	660
	Nombre de peptides uniques identifiés		13	14	14	19	16	14
	Protéine modèle 1 (Fsh alpha) Score Mascot		147	67	472	521	469	409
	Nombre de peptides uniques identifiés		12	11	14	15	14	14
	% de recouvrement de séquence		58 %	57 %	58 %	58 %	58 %	58 %
Taille du fichier .pkl		453 Kb	1608 Kb	1178 Kb	4370 Kb	1817 Kb	928 Kb	
Nombre de composés exportés		447	497	354	354	354	354	

**Table 8** | Paramètres testés et résultats évalués pour la mise au point d'une extraction des données brutes idéale. Les essais ont été réalisés à partir de l'analyse d'un digest trypsique de protéines modèles.

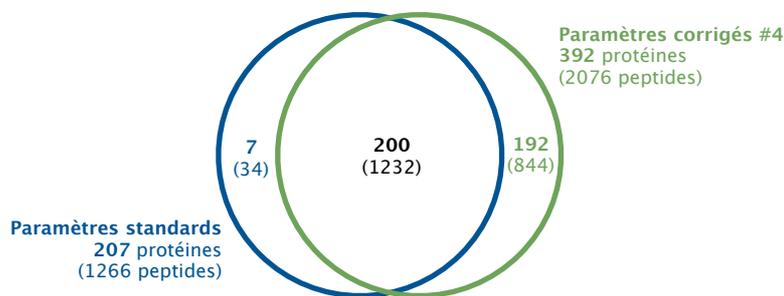
<sup>24</sup> Wehofskey, M., et al. (2002). J Mass Spectrom 37, 223–9.



**Figure 23** | Spectres de fragmentation du peptide VTSESTCIAMCIAMVAK ( $[M+2H]^{2+} = 621$  m/z). Spectre brut déconvolué (a) et extraits des spectres reconstitués par le moteur Mascot après recherche dans les banques à partir des fichiers « peak list » générés par les lots de paramètres standard (b), #2 (c) et #4 (d).

Cette étude montre que les automatisations du déisotopage et de la détermination de seuil de suppression du bruit de fond sont loin d'être optimales. En effet, leur action sur la simplification des données est efficace (les fichiers générés ne sont pas trop gros) mais elles semblent réaliser des opérations non désirées et incontrôlables (les rapports signal/bruit sont modifiés, des ions d'intérêt disparaissent notamment dans les bas m/z). Il semblerait que l'opération de déisotopage supprime tout ce qui n'est pas reconnu comme étant un massif isotopique attendu pour des peptides. Une nouvelle page de paramètres a donc été proposée sur la base du modèle #4 de notre étude.

Pour aller plus loin, nous avons évalué l'effet de ces paramètres sur le nombre d'identifications pour des échantillons entiers, issus de pistes de gel 1-D (Figure 24). En utilisant des options optimales d'extraction des données brutes pour la génération des fichiers « peak list », il a été possible d'identifier jusqu'à 93 % de protéines supplémentaires par rapport à un lot de paramètres standard.



**Figure 24** | Comparaison du nombre de protéines et de peptides identifiés, après soumission au moteur de recherche de fichiers « peak list », créés grâce aux paramètres standards et aux paramètres optimisés. Echantillons issus d'une piste de gel 1-D découpée de manière systématique et analysés 2 fois par nanoLC-MS/MS sur le couplage nanoAcquity UPLC-Synapt HDMS. Un critère de validation des identifications correspondant à un taux de faux positifs (FDR) inférieur à 1 % a été utilisé.

### 3.1.3. Optimisation des paramètres d'extraction des données brutes avec DataAnalysis (Bruker Dal.)

Les instruments Bruker Dal. – tel notre système HCT Ultra – nécessitent l'utilisation d'un script et de paramètres intégrés au logiciel DataAnalysis pour générer des fichiers « *peak list* » de type « *Mascot Generic Format* » (.mgf), exploitables par les moteurs de recherche. Jusqu'en 2009, il était nécessaire de traiter chaque analyse individuellement pour générer ces fichiers. Récemment, la société a édité un programme permettant de traiter des lots d'analyses, ce qui représente un gain de temps considérable.

Les paramètres à affiner sont similaires à ceux décrits précédemment. Notre expérience d'utilisateurs de ce logiciel montre cependant :

- qu'il est préférable de fixer un seuil d'intensité bas (autour de 1000 u.a.) pour la suppression du bruit de fond ;
- que le nombre de composés à exporter doit être relevé au-delà de 2000 pour l'analyse d'échantillons complexes de type bande de gel 1-D ;
- qu'il est préférable de désactiver l'option de déconvolution, car la résolution isotopique est parfois insuffisante sur un instrument de type trappe d'ions. La mauvaise attribution d'un état de charge sur un peptide rendrait alors son information incorrecte pour l'exploitation par le moteur de recherche.

### 3.1.4. Vers un format libre de fichiers « *peak list* »

Ces deux exemples d'extraction des données et de création des fichiers « *peak list* » montrent l'impérieuse nécessité d'optimiser les paramètres en fonction des instruments utilisés et des échantillons analysés. Il s'agit de s'affranchir des automatismes proposés par les constructeurs qui peuvent produire une information incomplète, voire erronée.

Pour une exploitation facilitée des informations, un problème réside également dans le format propriétaire des fichiers « *peak list* » générés. Des initiatives de développement de formats standardisés et libres ont été entreprises entre 2002 et 2004 : les formats mzXML<sup>25</sup> et mzData ont fait leur apparition. Leur intérêt premier est d'être potentiellement utilisable par toutes les suites logicielles d'exploitation de données MS. Cependant, l'existence de deux formats ayant le même objectif peut être source de confusion et demande un travail supplémentaire de développement. Le groupe « *Proteomics Standards Initiative* » (PSI) de l'Organisation pour le Protéome Humain (HUPO) a alors entrepris, autour des constructeurs, des développeurs et d'utilisateurs, de mettre au point un unique format tirant parti des avantages des deux formats existants (flexibilité de l'information reportée pour l'un, organisation clairement standardisée pour l'autre)<sup>26</sup>. Ce format nommé mzML est en cours de test et d'implémentation dans la plupart des programmes. Son succès dépendra largement de sa parfaite compatibilité entre toutes les solutions informatiques pour l'analyse protéomique.

## 3.2. Mise au point d'une stratégie de détermination d'un taux de faux positifs par l'approche « *Target-Decoy* »

Quelles que soient la stratégie expérimentale et la méthode d'identification de protéines mises en oeuvre, il est scientifiquement essentiel d'être en mesure de valider les résultats produits ; ou du moins qu'une estimation de l'erreur puisse être réalisée. Il s'agit de confirmer le crédit et la confiance donnés à l'information, générée par une technologie aussi avancée que l'analyse protéomique par spectrométrie de masse. Or, ces derniers temps, le doute a grandi quant à la fiabilité de résultats publiés dans la littérature, notamment en ce qui concerne les études à grande échelle. En 2008, certains éditeurs ont renforcé les

<sup>25</sup> Pedrioli, P. G., et al. (2004). *Nat Biotechnol* **22**, 1459–66.

<sup>26</sup> Orchard, S., et al. (2008). *Brief Bioinform* **9**, 166–73.

exigences en demandant que l'incertitude relative aux résultats d'identifications soit évaluée pour toute étude publiée (cf. Introduction Bibliographique, Chapitre III, point 3.1 – Le minimum requis pour la publication de données protéomiques (MIAPE)).

Jusqu'alors, le manque de ressources informatiques et de logiciels adaptés nous avait contraints à évaluer manuellement les résultats d'identification ambigus. Face aux nouvelles directives et devant les limites de nos méthodes manuelles – de plus en plus handicapantes –, le laboratoire a décidé d'évaluer de nouvelles approches d'évaluation des résultats pour les études à moyenne ou grande échelle.

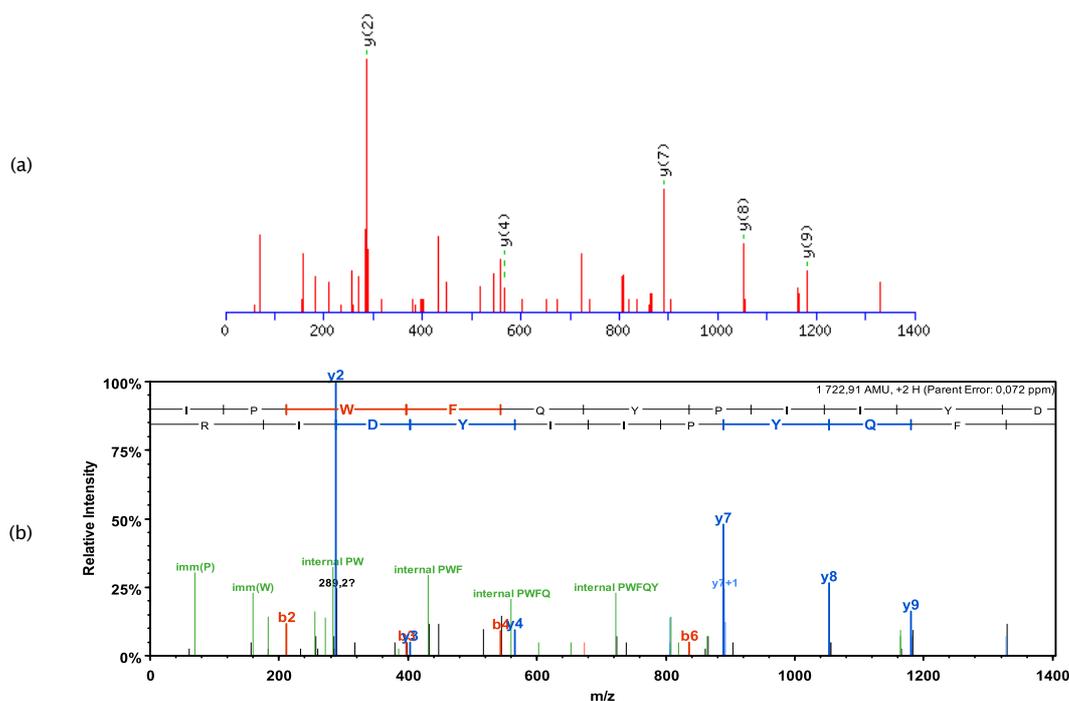
### 3.2.1. Les limites des stratégies d'identification manuelle

Notre stratégie d'identification et de vérification manuelles est largement décrite dans l'Introduction Bibliographique, au chapitre III, point 2.2. Sont détaillées ci-dessous les limites humaines et stratégiques qui, sous l'impulsion des directives éditoriales, nous ont conduites à rechercher des stratégies plus adéquates pour les projets de grande échelle.

#### ▸ Les limites humaines

Les projets traités au laboratoire ont récemment gagné en complexité : les stratégies développées impliquent de plus en plus d'analyses d'échantillons complexes (gels 1-D ou extraits cellulaires séparés ou non par 2D-LC) et la comparaison de multiples échantillons traités différemment. Alors qu'une évaluation manuelle de la qualité des identifications est possible et encore tolérée pour des cas simples, elle devient très fastidieuse et nécessite beaucoup de temps pour des études à grande échelle.

En outre, l'intervention humaine à ce stade du traitement des données implique des erreurs aléatoires indécélables à posteriori. La subjectivité de l'opérateur est également à prendre en compte : inconstante, elle dépend d'un grand nombre de facteurs humains et de la qualité de l'assistance informatique à disposition. A titre d'exemple, la figure 25 présente le même spectre de masse MS/MS annoté par 2 logiciels différents. L'identification risque d'être validée par l'utilisateur dans le deuxième cas uniquement, sous l'influence d'une annotation plus informative (fragments internes, fragments de type a ou c, ...) et d'une présentation graphique plus aboutie.



**Figure 25** | Exemple de spectre de fragmentation du peptide IPWFQYPIIDIR ( $[M+2H]^{2+} = 862,5$  m/z) dont la validation manuelle est très influencée par l'outil informatique utilisé. (a) Spectre MS/MS annoté par le moteur de recherche Mascot. (b) Le même spectre annoté par le logiciel Scaffold.

▸ Les limites stratégiques

Un écueil majeur de la vérification manuelle des identifications réside également dans l'impossibilité d'évaluer numériquement un taux de faux positifs. En effet, le seuil de sélection étant subjectif, il ne peut être quantifié et justifié par des critères chiffrés. Par ailleurs, Il est impossible de moduler simplement des critères subjectifs pour imposer une sélection plus ou moins exigeante, par exemple. Enfin, l'exportation des informations validées est très laborieuse : sans assistance logicielle, les pourcentages de recouvrement doivent être recalculés manuellement, les informations exportées pas à pas. Des initiatives ont permis de mettre au point des programmes permettant de faciliter de telles opérations<sup>27</sup> – dont quelques unes au sein du laboratoire – ; mais si ces outils facilitent la création d'un rapport final, ils ne comblent pas les défauts cruciaux des stratégies de validation manuelles.

▸ Vers une évaluation mesurée des résultats

Dans l'objectif de pallier les faiblesses importantes des stratégies de validation manuelle des identifications, et pour répondre aux conseils des éditeurs, le laboratoire a décidé d'investir dans l'évaluation des approches « *Target-Decoy* » (cf. Introduction Bibliographique, Chapitre III, point 2.3. – Les stratégies d'évaluation de faux positifs). Elles nous ont semblé être les plus adaptées à nos projets et les plus faciles à mettre en oeuvre. Surtout, elles permettent de justifier simplement et de manière empirique les seuils de scores choisis pour l'identification. Voyons quels choix ont été faits et quels sont les outils que nous avons mis en oeuvre.

### 3.2.2. La préparation de banques « *Target-Decoy* »

La première étape consiste à préparer la banque « leurre / *decoy* » à partir de la banque « cible / *target* ». Plusieurs options sont possibles et des règles importantes doivent être posées. De nombreux scripts sont disponibles pour l'extraction des séquences et la création des banques « *Target-Decoy* » ; mais nous avons développé notre propre programme qui répond au mieux aux choix expérimentaux décrits ci-dessous.

▸ Banques inversées ou banques aléatoires

D'une manière générale, il s'agit de construire une banque de données contenant des entrées de séquences protéiques inexistantes, mais semblables à celles de réelles protéines. Pour cela, la banque « leurre » peut être préparée soit en inversant toutes les séquences de la banque « cible », soit en créant des séquences aléatoires. L'inversion est une approche simple et raisonnable, car elle permet de préserver les entrées aux compositions très riches en certains acides aminés (collagènes et kératines par exemple) et de conserver le degré de redondance de la banque (c'est-à-dire le nombre de séquences peptidiques uniques). A l'inverse, la fabrication d'une banque aléatoire nécessite plus de précautions pour garantir que la taille et la composition des peptides soient réalistes.

▸ Equilibre des espaces de recherche

Une autre considération importante concerne la taille des espaces de recherche. En effet, la banque « leurre » doit avoir la même taille que la banque « cible » pour que le taux de faux positifs ne soit pas surévalué. Dès lors, deux points doivent être vérifiés :

(1) Les banques « *Target-Decoy* » doivent être conçues pour des taxonomies restreintes, correspondant à celles de l'échantillon. Dans le cas contraire, les séquences protéiques appartenant à des

<sup>27</sup> Pouillet, P., et al. (2007). *Proteomics* 7, 2553–6.

espèces non visées peuvent être considérées comme appartenant à la banque « leurre », ce qui augmente sa taille de manière disproportionnée.

(2) Les séquences des contaminants connus (kératines humaines, trypsine, ...) doivent être ajoutées aux banques pour que les peptides analysés issus de ces contaminants puissent être correctement identifiés et ne constituent pas de potentiels faux positifs.

▸ Banques concaténées ou banques séparées

Enfin, les banques « cible » et « leurre » peuvent être concaténées ou considérées séparément. Une recherche indépendante peut conduire à un taux de faux positif plus important et nous semble moins pertinente : en concaténant les deux banques, un peptide réellement identifié dans la banque « cible » ne peut pas simultanément l'être dans la banque « leurre ».

### 3.2.3. L'estimation et l'ajustement du taux de faux positifs (FDR)

A l'issue d'une recherche réalisée dans une banque « *Target-Decoy* », le taux de faux positif (FDR) pour l'identification des protéines peut être directement évalué comme étant égal à  $2.N_d/N$ , où  $N$  est le nombre total de protéines identifiées et  $N_d$  est le nombre de protéines identifiées dans la banque « leurre ».

En modulant un score seuil pour la sélection des peptides, il est possible d'ajuster le taux de faux positifs au niveau souhaité. Quelques précautions doivent cependant être observées :

▸ Le choix du score seuil est justifié par le taux de faux positifs qui lui est associé

Avant le développement des approches « *Target-Decoy* », les modèles de validation qui faisaient uniquement intervenir des seuils de scores au niveau peptidique<sup>28, 29</sup> pouvaient être largement critiqués. En effet, ils ne s'appuyaient que sur les approches probabilistes propres à chaque moteur de recherche, sans qu'il ne soit possible de déterminer un pourcentage d'erreur ou d'évaluer empiriquement la pertinence des seuils de scores choisis. Il pouvait être aussi nécessaire d'utiliser des seuils différents selon l'état de charge des peptides, la complexité de l'échantillon, la taille de la banque, l'instrument utilisé et bien sûr, le moteur de recherche employé. Une unification simple et claire des résultats issus de tous ces cas – et a fortiori leur comparaison – était impossible.

Les stratégies permettant la détermination d'un taux de faux positifs ont permis de résoudre ces problèmes. En effet, il est possible de fixer un taux de faux positifs maximal à ne pas dépasser et d'ajuster le score seuil de manière à tendre vers cette valeur. Ainsi, le taux de faux positifs justifie de manière simple et expérimentale le score seuil choisi pour l'identification.

▸ Moduler le score seuil d'identification des peptides pour estimer un taux de faux positifs au niveau protéique (FDR)

Une subtilité réside dans le fait d'évaluer un taux de faux positifs pour l'identification de protéines en ajustant un score seuil au niveau peptidique. Cela est justifié par deux raisons<sup>30</sup> :

(1) Il n'est pas pertinent de considérer des scores seuils pour l'identification de protéines. Cela revient en effet à valider l'identification de protéines à la condition que la somme des scores de ses peptides dépasse le seuil fixé. Des protéines identifiées avec un peptide unique au score important sont alors privilégiées au détriment de protéines identifiées avec plusieurs peptides aux scores plus faibles.

(2) Il est plus pertinent de suivre le taux de faux positifs (FDR) au niveau protéique, car l'optimisation des identifications de peptides n'implique pas directement l'optimisation des identifications de protéines. En

<sup>28</sup> Chepanoske, C. L., et al. (2005). *Rapid Commun Mass Spectrom* **19**, 9–14.

<sup>29</sup> Wielsch, N., et al. (2006). *J Proteome Res* **5**, 2448–56.

<sup>30</sup> Gupta, N., et al. (2009). *J Proteome Res*.

réalité, lorsqu'on cherche à maximiser le nombre de peptides identifiés dans la banque « cible », beaucoup des peptides supplémentaires appartiennent à des protéines déjà validées et n'augmentent donc pas le nombre d'identifications. Au contraire, l'augmentation simultanée des peptides identifiés dans la banque « leurre » a un impact direct sur le nombre de protéines « leures » identifiées – et donc sur le taux de faux positifs.

- ▶ Calculer des scores seuils avec le moteur Mascot

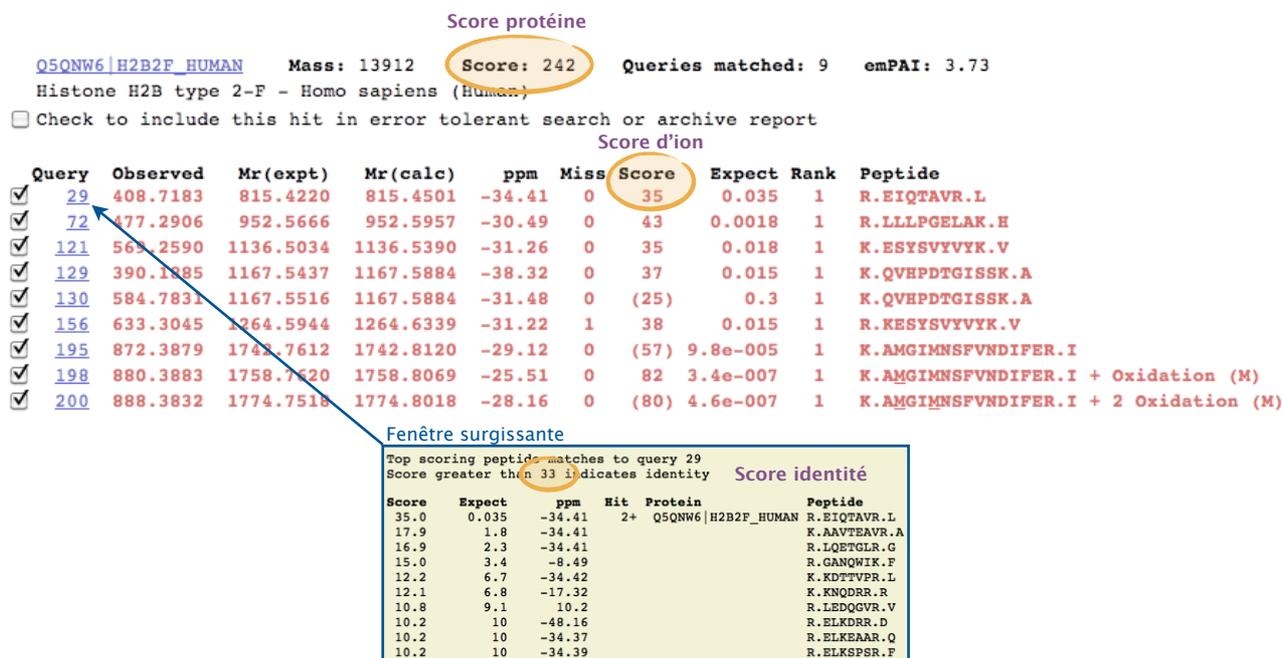


Figure 26 | Les différents scores calculés par le moteur Mascot (exemple d'un rapport d'identification nanoLC-MS/MS pour une protéine) : le score d'ion, le score identité et le score protéine (somme pondérée des scores d'ions qui limite l'importance des peptides au score faible).

Le moteur de recherche Mascot calcule un score pour chaque peptide identifié en fonction de la probabilité pour que leur identification soit un phénomène aléatoire ; c'est-à-dire la probabilité pour que la correspondance observée entre les données expérimentales et les masses calculées à partir du peptide candidat dans la banque de données soit due au hasard. Ce score est appelé « score d'ion » (Figure 26).

Pour chaque peptide, le moteur Mascot calcule également un « score identité », un niveau de confiance au-delà duquel l'identification n'est pas considérée comme le résultat d'un phénomène aléatoire.

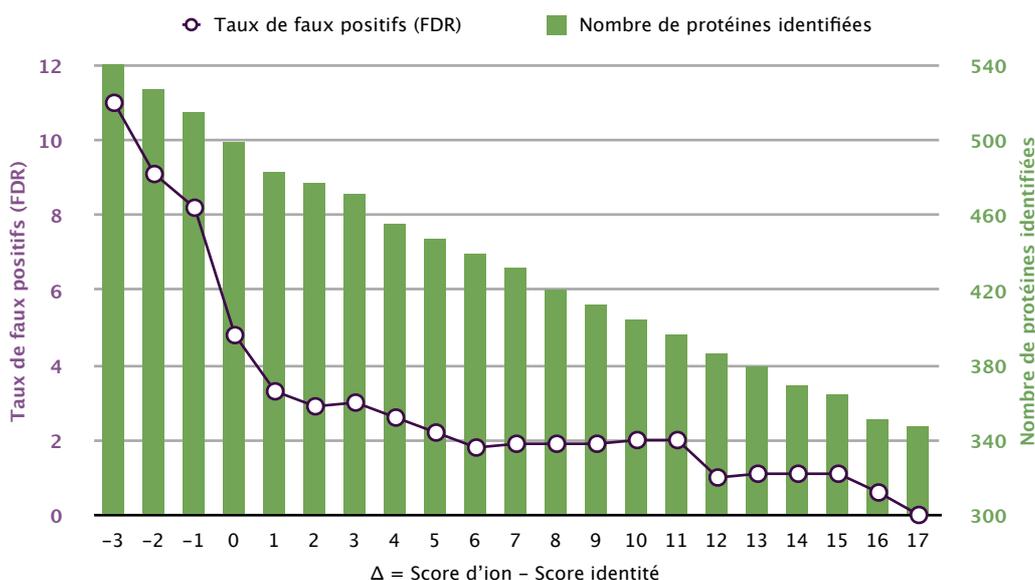
Nous avons choisi d'utiliser la différence entre le « score d'ion » et le « score identité » comme seuil d'identification de peptides. Cette différence  $\Delta$  évalue en quelque sorte la confiance avec laquelle chaque peptide est identifié.

Comme le « score identité » calculé théoriquement par Mascot est extrêmement conservateur, il est généralement possible de trouver des identifications tout à fait acceptables au dessous de cette limite, pour des valeurs de  $\Delta$  négatives.

- ▶ Trouver un compromis entre sensibilité et confiance pour les identifications

Pour chaque série d'analyse, le taux de faux positifs pour l'identification de protéines doit être ajusté en filtrant les peptides identifiés selon la valeur du seuil  $\Delta$ , différence « score d'ion – score identité ». Plus le seuil est strict, plus le taux de faux positifs est faible. Mais pour des seuils trop hauts, le nombre de protéines identifiées chute fortement. La figure 27 illustre ce phénomène : un compromis doit être trouvé entre un taux de faux positifs acceptable et une bonne sensibilité de la méthode. Généralement, ajuster le

seuil de score  $\Delta$  pour obtenir un taux de faux positifs au niveau protéique (FDR) inférieur à 2 % garantit une bonne sensibilité.



**Figure 27** | Evaluation de la sensibilité de la méthode et du taux de faux positifs en fonction du score seuil  $\Delta$  déterminé sur l'ensemble des analyses d'échantillons issus d'une piste de gel 1-D découpée de manière systématique et analysés 2 fois par nanoLC-MS/MS sur le couplage nanoAcquity UPLC-Synapt HDMS.

Afin de conserver un taux de faux positifs (FDR) convenable tout en maintenant une sensibilité intéressante, nous avons proposé de traiter différemment les identifications réalisées avec 1 peptide unique et celles réalisées avec plus de 2 peptides. Il a en effet été montré qu'à taux de faux positifs identique, pour tout lot de protéines identifiées avec plus de 2 peptides (en utilisant un seuil  $\Delta_2$ ), il existait un lot plus grand de protéines, identifiées avec 1 peptide ou plus, en utilisant un seuil  $\Delta_1$  (où  $\Delta_1 > \Delta_2$ )<sup>31</sup>.

Ainsi, pour obtenir une sensibilité maximale associée à un taux de faux positifs (FDR) inférieur à 2 %, les protéines identifiées avec plus de 2 peptides sont d'abord filtrées en déterminant un premier seuil  $\Delta_2$ . Puis les protéines identifiées avec 1 peptide unique sont retenues grâce à un seuil  $\Delta_1$  supérieur, ajusté pour garantir un taux de faux positifs inférieur à 2 %.

### 3.2.4. Les bienfaits du logiciel Scaffold pour la mise en oeuvre de notre stratégie

Jusqu'en 2008, la mise en oeuvre de la stratégie d'évaluation des résultats par approche « *Target-Decoy* » était rendue compliquée par l'absence d'outils informatiques adaptés. L'émergence du logiciel Scaffold (Proteome Software) a permis de relever un grand nombre des difficultés qui étaient rencontrées jusqu'alors dans la gestion des données et des résultats, leur vérification et leur validation, et le partage de l'information avec nos collaborateurs. Nous avons intégré pleinement les fonctions de ce logiciel à notre stratégie d'évaluation des résultats par l'approche « *Target-Decoy* ». Le logiciel permet notamment :

- ▶ L'extraction des données d'identification depuis les fichiers de résultats des moteurs de recherche ;
- ▶ La comparaison et combinaison de résultats d'identification de plusieurs analyses ou échantillons ;
- ▶ Le filtrage facilité des identifications selon des critères textuels, statistiques ou biologiques ;
- ▶ La comparaison, combinaison et validation croisée de résultats de multiples moteurs de recherche ;
- ▶ L'export exhaustif de l'information validée, sous forme de tableaux.

<sup>31</sup> Gupta, N., et al. (2009). J Proteome Res.

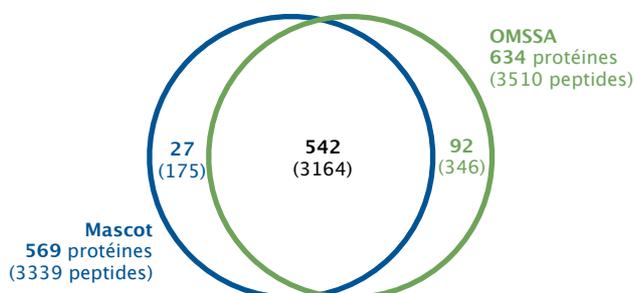
### 3.3. L'émergence de stratégies complémentaires

Pour devancer les attentes des éditeurs et poursuivre notre travail vers une identification la plus confiante et la plus exhaustive possible du contenu protéique de nos échantillons complexes, le laboratoire oriente depuis quelque temps ses investigations vers l'utilisation d'outils complémentaires pour l'identification de protéines. La suite de ce chapitre propose deux exemples en cours d'évaluation : la recherche dans les banques via des moteurs complémentaires, et la stratégie combinant l'approche *de novo* à l'identification dans les banques de données.

#### 3.3.1. L'utilisation combinée de plusieurs moteurs de recherche

Les moteurs de recherche dans les banques de données n'identifient généralement qu'une partie des spectres acquis lors d'une analyse LC-MS/MS. Les nuances entre les multiples algorithmes de recherche propres à chaque moteur sont à l'origine de résultats d'identification différents<sup>32</sup>. Pour recouvrer cette information potentielle, le laboratoire s'est doté d'un moteur de recherche complémentaire à Mascot, traditionnellement utilisé. Le moteur OMSSA (Open Mass Spectrometry Search Algorithm) est un programme libre possédant un algorithme de calcul de scores différent de celui de Mascot. Nous avons voulu évaluer une méthode simple pour comparer et cumuler les résultats d'identification fournis par les deux moteurs.

En soumettant les analyses d'un échantillon issu d'une découpe systématique d'une piste de gel 1-D à Mascot et OMSSA, réalisées sur notre système HPLC-Chip – HCT Ultra, les identifications réalisées par chacun des deux moteurs ont pu être comparées pour un même taux de faux positifs (Figure 28). L'information apportée par un deuxième moteur de recherche apparait complémentaire : près de 20 % de protéines supplémentaires peuvent être identifiées et les identifications déjà établies peuvent bénéficier d'arguments supplémentaires et d'un meilleur recouvrement grâce à de nouveaux peptides identifiés.

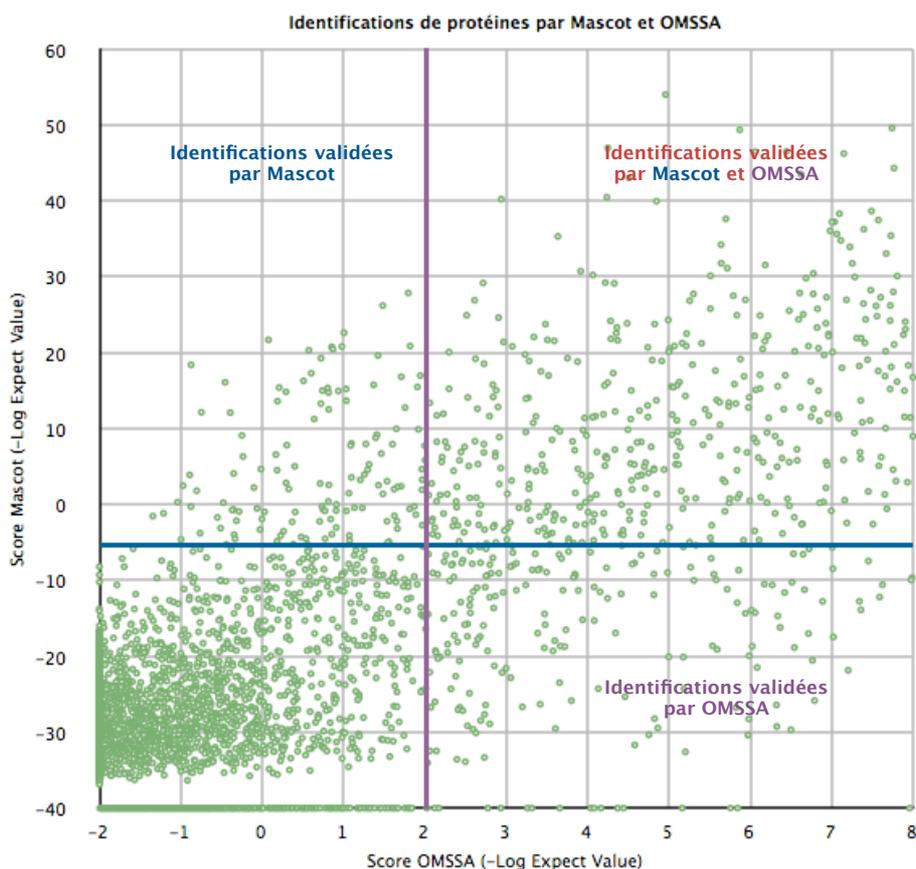


**Figure 28** | Comparatif des identifications réalisées grâce aux moteurs Mascot et OMSSA à partir des analyses d'échantillons issus d'une piste de gel 1-D découpée de manière systématique et analysés 2 fois par nanoLC-MS/MS sur le couplage HPLC-Chip – HCT Ultra.

Afin de tirer parti de cette complémentarité, les logiciels comme Scaffold sont en mesure de combiner les résultats de plusieurs moteurs de recherche pour renforcer le nombre et/ou la qualité des identifications de protéines et peptides (Figure 29). Pour cela, des algorithmes de normalisation des systèmes de score propres à chaque moteur sont mis en oeuvre. La combinaison des identifications issues des deux moteurs permet d'améliorer :

- soit la sensibilité des identifications (en sélectionnant les identifications issues de l'un et l'autre des moteurs à un taux de faux positifs inférieur à 2 %),
- soit la confiance dans les résultats (en sélectionnant les identifications communes aux deux moteurs, à un taux de faux positifs inférieur à 2 %).

<sup>32</sup> Jones, A. R., et al. (2009). *Proteomics* 9, 1220–9.



**Figure 29** | Variations des scores entre Mascot et OMSSA lors de l'analyse d'un échantillon issu d'une bande de gel 1-D, par nanoLC-MS/MS sur le couplage HPLC-Chip - HCT Ultra. Les seuils fixés pour chaque moteur, correspondant à un taux de faux positifs de 2 %, sont représentés par les deux droites parallèles aux axes.

### 3.3.2. Les stratégies de recherche « orthogonale » avec Peaks Studio

Utiliser le bon outil – ou la bonne combinaison d'outils – bioinformatiques pour l'identification de protéines et peptides et la validation des résultats est une question clé en analyse protéomique. Certaines situations peuvent compromettre l'utilisation de moteurs de recherche dans les banques protéiques. En effet, pour des organismes dont le génome n'est pas séquencé ou lorsque des recherches doivent être menées en autorisant des différences entre données théoriques et expérimentales, l'utilisation des stratégies d'identification *de novo* peut être une solution<sup>33</sup>.

Le logiciel Peaks Studio (Bioinformatics Solutions Inc.) utilise comme première approche un séquençage automatique des spectres MS/MS pour générer une liste de peptides candidats. Puis, ces séquences partielles ou complètes sont utilisées pour la recherche dans les banques de données protéiques. Le logiciel produit une liste de protéines candidates en calculant un score en fonction de la concordance des informations mesurées et théoriques. Il peut en outre extraire des données de recherche des moteurs de type Mascot ou OMSSA pour valider les identifications et augmenter la sensibilité de la recherche.

Malgré l'apparent intérêt de telles approches, elles sont encore peu utilisées, car le logiciel souffre de quelques problèmes pour des analyses à grandes échelles. Notre évaluation du système a montré qu'un de ses points forts résidait dans l'interprétation *de novo* automatique des spectres MS/MS. Cependant, les temps de traitement sur des séries d'analyses entières sont particulièrement longs. Dans le cadre d'une comparaison des résultats issus de différents moteurs, il n'est par ailleurs pas possible d'ajuster les seuils individuels pour chaque algorithme et l'utilisation du score Peaks – au calcul assez flou – est imposée. Enfin, l'export des résultats sous forme de tableau est encore laborieux et peu flexible.

<sup>33</sup> Bern, M., et al. (2007). *Anal Chem* **79**, 1393–400.

Peaks Studio trouve donc son principal intérêt dans l'identification de spectres non attribués par les méthodes de recherche des moteurs classiques : grâce au séquençage *de novo*, le logiciel permet d'identifier des peptides absents des banques ou présentant des modifications. Cela permet d'augmenter le recouvrement de séquence de protéines déjà identifiées ou d'être en mesure d'analyser des échantillons issus d'organismes dont le génome est peu ou pas séquencé.

Les possibilités offertes par ces nouvelles approches sont en cours d'évaluation et les possibilités techniques sont étudiées actuellement. Face aux résultats que l'utilisation d'OMSSA semble promettre, et devant la demande de la communauté de plus en plus pressante de produire des identifications validées avec plusieurs moteurs de recherche, ces stratégies pourront devenir des outils essentiels de l'analyse protéomique par spectrométrie de masse.

### 3.4. Conclusion

Si l'analyse protéomique par spectrométrie de masse a atteint un haut niveau de maturité et de maîtrise dans l'acquisition des données, de nombreux défis subsistent notamment en ce qui concerne l'analyse et le traitement de l'information générée. Il peut être escompté que des développements de méthodes et de logiciels comme ceux qui viennent d'être décrits seront en mesure d'améliorer la qualité des données protéomiques. Pour cela, le test et la mise en oeuvre des nouveaux outils à disposition doivent être systématiques. Les éditeurs et la communauté scientifique ont reconnu cette nécessité et ont commencé à imposer des règles de publication pour la validation et le partage des informations.

▸ Les outils bio-informatiques sont aujourd'hui partie prenante de l'analyse protéomique par spectrométrie de masse. Ils interviennent à tous les niveaux de l'expérience, influant sur la qualité des données générées comme sur leur validation.

▸ L'extraction et la préparation des informations à soumettre aux moteurs de recherche sont des points clés du processus. Les paramètres d'export des données vers les fichiers « *peak list* » doivent être finement ajustés en fonction de l'instrument, de l'échantillon et de la stratégie mise en oeuvre.

▸ La validation des identifications de protéines par les stratégies LC-MS/MS est maintenant incontournable. Nous avons évalué et mis au point une approche de vérification et d'estimation du taux de faux positifs (FDR) fondée sur les stratégies « *Target-Decoy* ».

▸ L'évaluation et l'installation de logiciels d'aide au traitement des données comme Scaffold ont permis de mettre en oeuvre ces stratégies et de faciliter le partage des informations.

▸ Enfin, de nouvelles approches sont en cours d'évaluation, notamment en ce qui concerne l'utilisation de moteurs de recherches complémentaires. Les premiers essais avec OMSSA semblent promettre des identifications plus fiables et plus sensibles.

## 4. Conclusion

Nous commençons ce chapitre en soulignant les difficultés majeures rencontrées en analyse protéomique pour obtenir des données reproductibles. L'étude de la répétabilité de nos propres analyses nanoLC-MS/MS a montré quelles pouvaient être les origines des difficultés conduisant à des résultats incomplets. Le développement de stratégies s'appuyant sur des analyses répétées du même échantillon a permis d'améliorer la part du contenu protéique total accessible et la qualité des identifications. Par ailleurs, les innovations technologiques qui ont pu être mises en oeuvre durant ce travail de thèse, tant du point de vue de la chromatographie liquide que de la spectrométrie de masse, ont ouvert de nouvelles possibilités pour une meilleure prise des données protéomiques. Enfin, il a été montré que les outils bioinformatiques étaient déterminants pour l'exploitation, la validation et le partage des résultats produits par l'analyse protéomique.

Comme l'automobile s'est popularisée en étant soutenue par d'autres technologies spécialement développées comme les autoroutes ou les stations essences, l'analyse protéomique a bénéficié d'avancées majeures dans les techniques de séparation, la haute résolution et la vitesse d'acquisition en spectrométrie de masse. Le chaos des premières années de l'automobile a rapidement laissé place à un monde plus ordonné de feux tricolores, limites de vitesse et parkings aménagés. Les règles de publications en cours d'élaboration pour la publication de données en spectrométrie de masse lui conféreront plus de lisibilité et de fiabilité. La production de masse a fait chuter les prix des autos et les a démocratisées. La spectrométrie de masse n'est qu'au début de sa conquête du marché qui lui est promis. Les travaux visant à préparer des stratégies d'analyse et d'identification validées pavent la route vers ce futur prometteur.

## Chapitre II

# Recherche de nouveaux biomarqueurs caractéristiques d'hémopathies lymphoïdes chroniques

---

Ce projet s'inscrit dans le cadre d'une collaboration avec l'équipe du Professeur Laurent Mauvieux, du pôle d'Onco-Hématologie des Hopitaux Universitaires de Strasbourg. Laurent Miguet, qui a rejoint l'équipe de L. Mauvieux, a posé les fondements de ce projet dans le travail de thèse qu'il a réalisé au laboratoire<sup>1</sup>. Son travail a notamment porté sur le développement et l'évaluation de la préparation d'échantillons de microparticules enrichis en protéines membranaires. Les premiers résultats de l'analyse protéomique différentielle de ces échantillons, produits pour plusieurs pathologies au diagnostic ambigu, ont ouvert la voie à une application de la stratégie à la recherche de nouveaux marqueurs membranaires spécifiques.

Le propos de ce chapitre rappelle les points forts de l'approche analytique, développe la stratégie de validation des identifications réalisées et détaille les résultats validés pour le lymphome du manteau (MCL). Grâce aux avancées stratégiques et techniques décrites au chapitre précédent, de nouveaux cas atypiques ou rares ont pu être analysés.

---

<sup>1</sup> Miguet, L. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2006).

# 1. La nécessité de mettre en évidence de nouveaux biomarqueurs membranaires spécifiques de leucémies

Les leucémies désignent une prolifération maligne de cellules hématopoïétiques. Les cellules cancéreuses prolifèrent anormalement dans les organes impliqués dans la production et/ou la maturation des cellules sanguines avant de s'accumuler dans le sang. Alors que les leucémies aiguës, d'évolution rapide, touchent les cellules immatures, les leucémies chroniques touchent les cellules matures.

Les hémopathies lymphoïdes chroniques représentent un ensemble de pathologies cancéreuses touchant les lymphocytes B matures de l'adulte. Avec une augmentation annuelle de 5 à 10 % dans les pays occidentaux, ces pathologies représentent chaque année 4 % des nouveaux cas de cancer. Il s'agit de syndromes accumulatifs qui peuvent évoluer de manière très différente et permettre une survie allant de 4 à 25 ans.

## 1.1. La classification des leucémies

Le diagnostic, la classification, le suivi et le contrôle des néoplasmes hématologiques reposent sur quelques éléments cliniques et biologiques, parmi lesquels les études morphologiques, génotypiques et immunophénotypiques des lymphocytes circulants tiennent une place majeure<sup>2</sup>. Cependant, pour un certain nombre de pathologies, le diagnostic est plus délicat et peut être associé à un pronostic plus réservé. La classification des tumeurs des tissus lymphoïdes et hématopoïétiques élaborée par l'Organisation Mondiale de la Santé (O.M.S.) a lentement évolué ces 10 dernières années et quelques critères utilisés pour distinguer les différentes pathologies ont été affinés (Figure 1)<sup>3</sup>. Cette classification détaille les approches multi-paramétriques permettant de caractériser chaque hémopathie ; mais elle ne lève pas le voile sur les cas atypiques, ambigus, difficiles à diagnostiquer et dont la réponse aux traitements est incertaine. Malgré les efforts de recherche, la classification est toujours délicate dans certains cas.

### Hémopathies Lymphoïdes Malignes B

Leucémies Aiguës Lymphoblastiques (touchent le lymphocyte immature)	Hémopathies Lymphoïdes B Matures Chroniques (touchent le lymphocyte mature)	
<b>Leucémie Aiguë Lymphoblastique (LAL-B)</b>	<b>Syndromes lympho-prolifératifs B chroniques (lymphocytes dans le sang)</b>	<b>Lymphomes Non-Hodgkiniens (lymphocytes dans les ganglions, la rate...)</b>
	Leucémie lymphoïde chronique B (CLL) Leucémie prolymphocytaire (B-PLL) Leucémie à tricholeucocytes (HCL) Lymphome splénique à lymphocytes villeux	Lymphome de la zone marginale (MZL) Lymphome folliculaire (FL) Lymphome à petites cellules (SLL) Lymphome du manteau (MCL)

Figure 1 | Classification simplifiée des hémopathies lymphoïdes malignes B d'après l'Organisation Mondiale pour la Santé.

<sup>2</sup> Craig, F. E., et al. (2008). Blood **111**, 3941-67.

<sup>3</sup> Swerdlow, S. H., Campo, E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Vardiman, J.W (2008). IARC Press, Lyon.

## 1.2. Vers un meilleur diagnostic des cas ambigus de leucémies

### 1.2.1. Les techniques de cytométrie en flux

Parmi les techniques de diagnostic, l'approche d'immunophénotypage par cytométrie en flux s'est imposée comme essentielle. Elle permet d'évaluer la présence ou l'absence et le niveau d'expression d'antigènes spécifiques sur des cellules en suspension. D'abord utilisée pour caractériser de larges populations de cellules anormales, la technique est maintenant mature pour évaluer en routine de petites populations de cellules présentant de subtiles modifications dans leur expression antigénique. Les améliorations instrumentales et la disponibilité d'un nombre croissant d'anticorps et de fluorochromes ont permis de décrire le phénotype des cellules avec plus de précision, et notamment d'améliorer l'identification de populations anormales<sup>4</sup>.

Dans le cas des hémopathies lymphoïdes B matures, les études d'immunophénotypage par cytométrie en flux sont indispensables pour identifier les lymphocytes B au phénotype anormal et pour tenter de reconnaître les caractéristiques phénotypiques de la pathologie concernée. En outre, la cytométrie en flux peut être utilisée pour identifier des cibles thérapeutiques (des antigènes spécifiques) et pour apporter une information complémentaire sur le pronostic grâce à l'étude de protéines de surface ciblées<sup>5</sup>.

Le succès des techniques de diagnostic par cytométrie en flux et la disponibilité d'anticorps thérapeutiques dépendent directement de l'identification de cibles adaptées et pertinentes. Les antigènes de la membrane plasmique sont les plus à propos. Parmi eux, les clusters de différenciation (CD) représentent une catégorie de protéines dont les niveaux d'expressions sont particulièrement contrastés selon les cas et qui peuvent être impliquées dans le processus de développement des pathologies.

### 1.2.2. La découverte de biomarqueurs comme moteur du progrès en matière de diagnostic

Les experts sont maintenant en mesure de recommander des approches orientées d'abord sur des données de cytométrie en flux pour discriminer entre les pathologies plutôt que sur des considérations s'appuyant sur la description de pathologies déterminées.

Dans la stratégie d'évaluation des néoplasmes lymphoïdes B matures, il est généralement utile de considérer 4 larges groupes de pathologies déterminés en fonction du niveau d'expression de CD5 et CD10 (Table 1). Pour chaque groupe, des données complémentaires de cytométrie en flux combinées à des observations morphologiques peuvent aider à resserrer le champ des diagnostics possibles et orienter les études auxiliaires à réaliser.

Si la classification et les possibilités d'un diagnostic précis s'affinent, d'importants progrès restent à accomplir. En effet, face à la complexité des cas, à la diversité des situations rencontrées, et malgré le cocktail de biomarqueurs de surface connu, il est parfois difficile de discriminer entre plusieurs pathologies. Par exemple, le phénotype du lymphome du manteau est plus variable que celui des leucémies lymphoïdes chroniques (CLL) et des lymphomes à petites cellules (SLL), mais ils partagent beaucoup de caractéristiques communes. Il est alors nécessaire de recourir à des analyses complémentaires, longues, coûteuses et pas toujours parfaitement spécifiques.

Par conséquent, l'effort de recherche de nouveaux biomarqueurs pour des pathologies au diagnostic toujours ambigu ou inconnu doit être poursuivi.

<sup>4</sup> Wood, B. (2006). Arch Pathol Lab Med **130**, 680–90.

<sup>5</sup> Scallan, B. J., et al. (2006). J Immunother **29**, 351–64.

Pathologie	Élément phénotypique discriminant	Autre information pour le diagnostic
<b>CD5+ CD10-</b>		
Leucémie lymphoïde chronique (CLL) et lymphome à petites cellules (SLL)	Phénotype typique : CD20 <sup>f</sup> , CD22 <sup>f</sup> , slg <sup>f</sup> , CD23 <sup>+</sup> , FMC-7 <sup>-</sup>	Morphologie caractéristique
Lymphome du manteau (MCL)	Phénotype variable atypique pour CLL; souvent CD20 <sup>m</sup> , slg <sup>m</sup> , CD23 <sup>+/-</sup> , FMC-7 <sup>+/-</sup>	Cyclin-D1 IHC, réarrangement t(11;14)/CCND
Leucémie prolymphocytaire (B-PLL)	Phénotype variable, possiblement similaire à CLL et MCL, CD20 <sup>+m</sup> , slg <sup>+m</sup> , FMC-7 <sup>+/-</sup> , CD5 <sup>+/-</sup>	Grosses cellules à noyau proéminent
Lymphome de la zone marginale (MZL)	Phénotype variable atypique pour CLL : CD23 <sup>-</sup> , souvent CD11c <sup>+/-</sup> , CD103 <sup>+/-</sup> , parfois clg uniquement	Croissance autour et dans les follicules, réarrangement t(11;18), t(1;14), t(14;18)/MALT-1
Lymphome diffus à grandes cellules B (DLBCL)	Phénotype variable	Grosses cellules, croissance diffuse
Lymphome lymphoplasmacytique (LPL)	Phénotype atypique pour CLL, souvent CD23 <sup>-/f</sup> , parfois slg <sup>-</sup> mais clg <sup>+</sup>	Petites cellules, différenciation plasmatique
<b>CD5- CD10+</b>		
Lymphome folliculaire (FL)	Habituellement bcl-2 <sup>+</sup> , CD43 <sup>+/-</sup>	Croissance folliculaire, réarrangement t(14;18)/BCL-2
Lymphome diffus à grandes cellules B (DLBCL)	Phénotype variable, bcl-2 <sup>+/-</sup> , CD43 <sup>+/-</sup>	Grosses cellules quelque peu pleomorphiques, croissance diffuse
Lymphome de Burkitt (BL)	Habituellement bcl-2 <sup>-</sup> , CD10 <sup>+e</sup> , CD43 <sup>+</sup>	Cellule de taille uniforme intermédiaire
Lymphome à tricholeucocytes (HCL)	Phénotype typique : CD20 <sup>e</sup> , CD22 <sup>e</sup> , CD11c <sup>e</sup> , CD25 <sup>+</sup> , CD103 <sup>+</sup> , slg <sup>m</sup> , CD123 <sup>+</sup>	Morphologie caractéristique, Annexin-A1 <sup>+</sup>
<b>CD5+ CD10+</b>		
Lymphome folliculaire (FL)	Habituellement bcl-2 <sup>+</sup> , CD43 <sup>-</sup>	Croissance folliculaire possible, réarrangement t(14;18)/BCL-2
Lymphome diffus à grandes cellules B (DLBCL)	Phénotype variable, bcl-2 <sup>+/-</sup> , CD43 <sup>+/-</sup>	Grosses cellules, croissance diffuse
Lymphome du manteau (MCL)	Phénotype variable atypique pour CLL; souvent CD20 <sup>m</sup> , slg <sup>m</sup> , CD23 <sup>+/-</sup> , FMC-7 <sup>+/-</sup>	Cyclin-D1 IHC, réarrangement t(11;14)/CCND
Lymphome de Burkitt (BL)	Habituellement bcl-2 <sup>-</sup> , CD10 <sup>b</sup> , CD43 <sup>+</sup>	Cellule de taille uniforme intermédiaire
<b>CD5- CD10-</b>		
Lymphome à tricholeucocytes (HCL)	Phénotype typique : CD20 <sup>e</sup> , CD22 <sup>e</sup> , CD11c <sup>e</sup> , CD25 <sup>+</sup> , CD103 <sup>+</sup> , slg <sup>m</sup>	Morphologie caractéristique
Lymphome de la zone marginale (MZL)	Souvent CD11c <sup>+/-</sup> , CD103 <sup>+/-</sup> mais atypique pour HCL, parfois slg <sup>-</sup> mais clg <sup>+</sup>	Croissance autour et dans les follicules, réarrangement t(11;18), t(1;14), t(14;18)/MALT-1
Lymphome diffus à grandes cellules B (DLBCL)	Phénotype variable	Grosses cellules, croissance diffuse
Lymphome folliculaire (FL)	Phénotype variable	Croissance folliculaire possible, réarrangement t(14;18)/BCL-2
Lymphome du manteau (MCL)	Phénotype variable	Cyclin-D1 IHC, réarrangement t(11;14)/CCND

**Table 1** | Approche de diagnostic et de classification par cytométrie en flux des hémopathies lymphoïdes B chroniques.

+ : habituellement positif, - : habituellement négatif, +/- : peut être positif ou négatif, f : intensité faible, m : intensité moyenne, e : intensité élevée, slg : immunoglobuline de surface, clg : immunoglobuline cytoplasmique.

### 1.3. Rechercher de nouveaux marqueurs membranaires spécifiques

#### 1.3.1. Les méthodes classiques

Les avancées dans l'analyse de l'expression génétique ont permis l'identification d'antigènes spécifiques de tumeurs grâce au profilage de gènes à grande échelle. Ces techniques incluent l'analyse sérielle de l'expression génique (SAGE), les études différentielles de réaction en chaîne par polymérase après transcription inverse (RT-PCR), l'hybridation soustractive, le séquençage de tags de séquences exprimés (EST), et surtout les analyses sur puce à ADN (DNA-Microarray) permettant de détecter les gènes surexprimés par les cellules tumorales de manière spécifique. Les observations au niveau génomique et protéique peuvent

être très différentes : pour les protéines membranaires par exemple, de nombreux facteurs posttraductionnels peuvent affecter la fonction et la localisation des protéines dans la membrane plasmique<sup>6</sup>. Le nombre très important de cibles généralement identifiées doit donc être validé au niveau des protéines<sup>7</sup>.

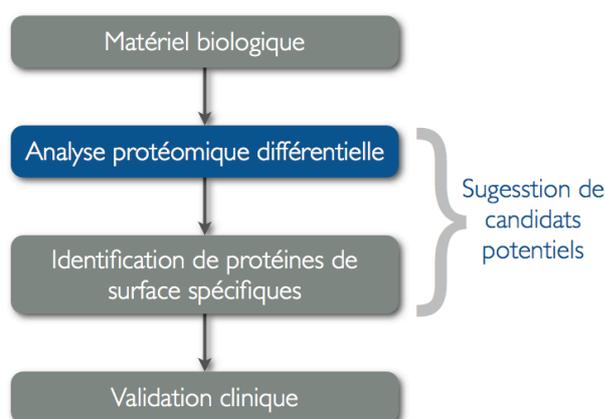
Ces études peuvent alors donner lieu à la proposition de banques d'anticorps monoclonaux à tester de manière systématique contre les cellules tumorales issues de différentes pathologies afin d'identifier de potentiels marqueurs membranaires. Si cette approche nécessite la disponibilité d'une large gamme d'anticorps, elle est encore la plus utilisée. On peut citer par exemple de récents travaux qui ont pu mettre en évidence des différences au niveau protéique entre des lymphocytes B normaux et pathologiques<sup>8</sup>. L'étude a nécessité le test systématique de plus de 80 anticorps afin de réaliser des profils protéiques. Quatre antigènes semblant spécifiques du lymphome du manteau ont été mis en évidence.

### 1.3.2. Approche protéomique par l'analyse des protéines membranaires par spectrométrie de masse

Face aux difficultés des méthodes classiques, les approches d'analyse protéomique par spectrométrie de masse présentent un intérêt certain. En effet, elles ne sont pas limitées par la taille et la composition de banques d'anticorps et peuvent identifier non seulement des cibles connues mais aussi de nouveaux candidats. Les avantages analytiques inhérents à la spectrométrie de masse incluent sa sensibilité, sa résolution, sa vitesse et son débit d'analyse, combinés aux méthodes poussées d'interprétation des données grâce aux outils bioinformatiques. Cette approche permet l'identification rapide et systématique de centaines de protéines.

Au-delà de la simple identification, les stratégies mises au point doivent offrir la possibilité de comparer des échantillons et de proposer des éléments discriminants. Par exemple, l'analyse comparative du protéome membranaire de lymphocytes B issus de patients présentant des pathologies différentes peut produire une information sur les modifications induites par la maladie directement au niveau protéique.

Il est largement entendu que ces approches d'analyse protéomique par spectrométrie de masse ne sont pas encore prêtes pour satisfaire les contraintes des analyses de routine et à grande échelle au plan clinique<sup>9</sup>. La procédure permettant de découvrir de nouveaux biomarqueurs diagnostiques ou des cibles thérapeutiques suit un cheminement long et complexe allant de la collecte et la préparation d'échantillons prélevés sur des patients, à la validation clinique des résultats, en passant par l'analyse protéomique différentielle permettant de proposer des candidats biomarqueurs potentiels (Figure 2). Le défi reste qu'à chaque étape, de nombreux éléments peuvent affecter la reproductibilité et des biais peuvent influencer sur le résultat final.



**Figure 2** | La découverte de biomarqueurs par l'approche protéomique par spectrométrie de masse est un processus long et délicat.

<sup>6</sup> Boyd, R. S., et al. (2009). *Mol Cell Proteomics* **8**, 1501-15.

<sup>7</sup> Ek, S., et al. (2007). *Curr Genomics* **8**, 171-9.

<sup>8</sup> Ek, S., et al. (2006). *Mol Cell Proteomics* **5**, 1072-81.

<sup>9</sup> Beretta, L. (2007). *Nat Methods* **4**, 785-6.

Il faut notamment insister sur le fait que l'analyse protéomique par spectrométrie de masse pour la découverte de biomarqueurs se heurte à l'impossibilité de juger de la pertinence clinique des candidats. L'approche ne permet que la proposition de cibles qu'il faut valider par des essais cliniques sur de grandes populations, afin de déterminer si l'expression des marqueurs discriminants identifiés est cliniquement significative et utile.

Nous avons choisi d'appliquer cette approche à l'identification de nouveaux marqueurs membranaires spécifiques de différents cas de leucémies lymphoïdes chroniques au diagnostic ambigu. Dans ce contexte, une large part du travail a d'abord consisté en la mise au point de conditions et d'outils expérimentaux pour une analyse ciblée et la plus complète possible des protéines membranaires de lymphocytes B cancéreux. Notre stratégie a ensuite été appliquée au cas du lymphome du manteau pour lequel un marqueur spécifique a été identifié et validé cliniquement. De nouvelles pathologies rares, comme la maladie de Waldenström, sont par ailleurs en cours d'étude.

## 2. Analyse comparative du protéome membranaire de plusieurs leucémies

L'analyse protéomique différentielle s'appuie sur la comparaison du contenu protéique de plusieurs échantillons (Figure 3). Elle doit permettre de mettre en évidence des protéines exprimées spécifiquement par un type de cancer par rapport aux autres et de proposer ainsi de nouveaux biomarqueurs candidats, spécifiques d'une pathologie. C'est la comparaison de la composition protéique des membranes issues de cellules de pathologies différentes qui permet d'identifier des protéines exprimées de manière spécifique par certaines d'entre elles.

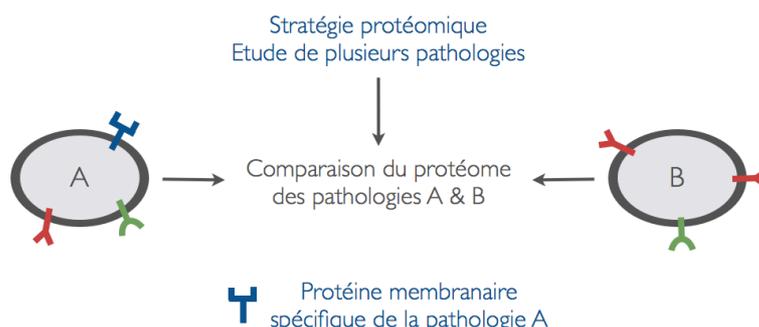


Figure 3 | Approche protéomique différentielle pour la recherche de marqueurs membranaires spécifiques.

### 2.1. Choix expérimentaux permettant la recherche avancée de marqueurs membranaires potentiels

#### 2.1.1. Préparation des échantillons et enrichissement en protéines membranaires

- L'impossibilité de disposer d'échantillons témoins de patients sains

Les cellules d'un individu sain ne peuvent être utilisées comme échantillon témoin : en l'absence de cancer, les lymphocytes sont présents en faible concentration dans le sang et des volumes trop importants seraient nécessaires à la préparation d'échantillons témoins. De ce fait, seule la comparaison de pathologies entre elles est possible : la prolifération et l'accumulation des cellules malades sont suffisantes pour la production d'échantillons analysables ( $200 \times 10^6$  cellules nécessaires).

- Les microparticules : des échantillons naturellement enrichis en protéines membranaires

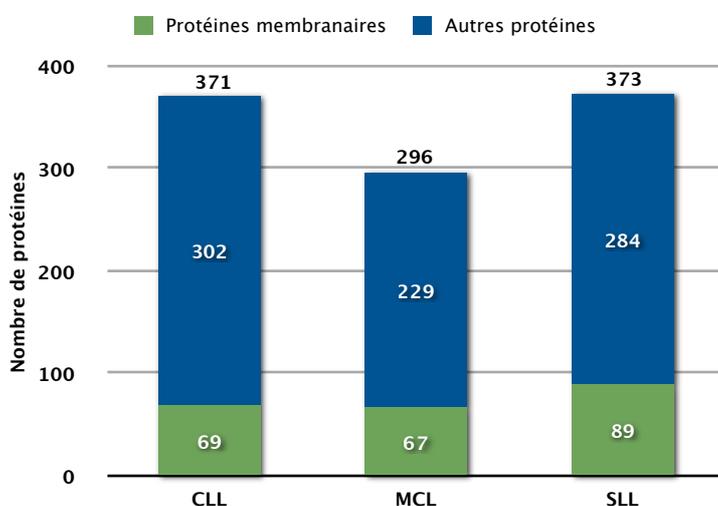
L'utilisation de cellules entières pour la recherche de protéines membranaires n'est pas envisageable. Les échantillons seraient en effet trop hétérogènes et contaminés par de nombreuses protéines cytosoliques ou issues des autres compartiments cellulaires. On s'intéresse alors à l'étude du protéome de la membrane plasmique (sous-protéome de la cellule). La production d'échantillons de préparation membranaire équivaut à une véritable première dimension de séparation analytique. Elle permet une analyse plus aisée et plus spécifique des protéines de surface.

La préparation d'échantillons naturellement enrichis en protéines membranaires a été préalablement mise au point au laboratoire : le développement et la validation de la génération de microparticules à partir de lymphocytes cancéreux a permis de disposer d'une technique d'enrichissement en protéines

membranaires bien plus performante et simple que les préparations classiques<sup>10</sup>. Comme il l'a déjà été décrit (cf. Introduction Bibliographique, Chapitre II, point 2.3 – L'analyse de protéines membranaires), les microparticules sont formées sous l'effet d'un traitement conduisant à l'apoptose des cellules. Elles sont ensuite débarrassées des débris cellulaires par centrifugation différentielle. L'échantillon au final est potentiellement composé de plus de 30 % de protéines membranaires.

La quantité de production de microparticules est difficilement maîtrisable : partant d'un même nombre de cellules initial, il est possible que pour certaines pathologies – en fonction de leur agressivité – des microparticules soient produites de manière plus spontanée.

Pour notre première application (dont la publication figure au point 2.4 de ce chapitre), 3 pathologies ont été comparées : un cas de leucémie lymphoïde chronique (CLL), un cas de lymphome du manteau (MCL) et un cas de lymphome à petites cellules (SLL). Des microparticules ont été générées pour chacun des échantillons et le nombre de protéines membranaires identifiées a été évalué après analyse. L'enrichissement s'élève en moyenne à 23 % pour chacune des pathologies. Le nombre total de protéines identifiées est équivalent dans les 3 cas (Figure 4).



**Figure 4** | Nombre de protéines identifiées et fraction de protéines membranaires dans les trois échantillons de microparticules analysés issus des pathologies CLL, MCL et SLL étudiées.

▸ Les difficultés de doser la quantité de protéines dans les échantillons de microparticules

Dans l'objectif de comparer le contenu protéique de différents échantillons, il est nécessaire de normaliser la quantité de protéines analysées. Cela est généralement réalisé par différents types de dosages colorimétriques, le plus courant étant le dosage de Bradford<sup>11</sup>. Du fait de leur solubilisation difficile, et de l'incompatibilité de ces dosages avec l'utilisation de détergents classiques (de type SDS), le dosage des échantillons de microparticules n'est pas aisé et conduit souvent à des résultats aberrants. En outre, même avec les instruments de dosages colorimétriques récents qui ne nécessitent qu'une simple goutte d'échantillon, la quantité de matériel nécessaire pour le dosage est trop importante par rapport à ce qui est requis pour une analyse protéomique dans de bonnes conditions (plusieurs centaines de microgrammes).

Pour nos études, l'hypothèse est donc posée que partant d'un même nombre de cellules, la production et la purification de microparticules sont reproductibles pour chaque échantillon.

▸ Analyse de protéines membranaires par spectrométrie de masse

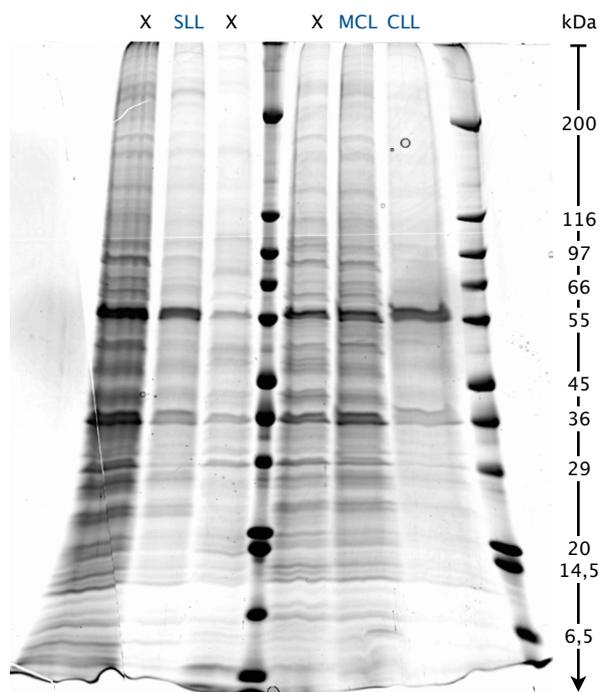
Les protéines membranaires restent une catégorie de protéines plus compliquée à analyser du fait de leurs difficultés de solubilisation. Des essais de séparation sur gel d'électrophorèse bidimensionnel ont été

<sup>10</sup> Miguet, L., et al. (2007). *Subcell Biochem* 43, 21–34.

<sup>11</sup> Bradford

décrits, mais la technique n'est pas très résolutive pour les protéines membranaires et les tampons de solubilisation sont incompatibles avec la première dimension de séparation. Des approches « *shotgun* » mettant en jeu la double séparation par chromatographie liquide de digests peptidiques totaux sont également utilisées, la technique demande cependant beaucoup de travail de mise au point<sup>12</sup>. Nous nous sommes donc dirigés naturellement vers l'électrophorèse monodimensionnelle SDS-PAGE, qui autorise la solubilisation des protéines membranaires, mais ne permet pas d'atteindre une résolution aussi fine qu'avec les autres approches. Cependant robuste et reproductible, la technique est bien maîtrisée au laboratoire.

La figure 5 présente le gel réalisé pour l'analyse des pathologies CLL, MCL et SLL. Les échantillons de microparticules ont été solubilisés dans un tampon de charge à haute concentration en SDS et déposés sur un gel de gradient 8-15 %, permettant une séparation optimale sur une gamme de masse s'étalant de 205 à 5 kDa.



**Figure 5** | Gel d'électrophorèse SDS-PAGE, gradient 8-15 %, coloré au bleu de Coomassie, présentant les 3 pathologies étudiées. Les autres pistes (notées X) correspondent à des pathologies hors du contexte de cette étude.

Chaque bande, découpée de manière systématique tous les 2 mm, contient potentiellement plusieurs dizaines de protéines. Le contenu protéique est soumis aux traitements de réduction et d'alkylation classiques permettant de réduire les ponts disulfures existants entre certaines cystéines. Après digestion trypsique, chaque échantillon est constitué d'un mélange peptidique complexe qui doit être analysé par nanoLC-MS/MS.

### 2.1.2. Optimisation des analyses nanoLC-MS/MS

#### ▸ Valorisation de l'étude de la répétabilité des analyses nanoLC-MS/MS

Les limites et l'optimisation de l'analyse de mélanges peptidiques complexes ont été longuement discutées précédemment (cf. Chapitre I de cette partie, point 1.1 - Etude de la répétabilité des analyses nanoLC-MS/MS d'échantillons biologiques complexes). Notamment pour des approches protéomiques différentielles, la stratégie développée doit permettre une identification maximale du contenu protéique total de chaque échantillon. D'une part, il faut que l'identification des protéines minoritaires soit possible. D'autre

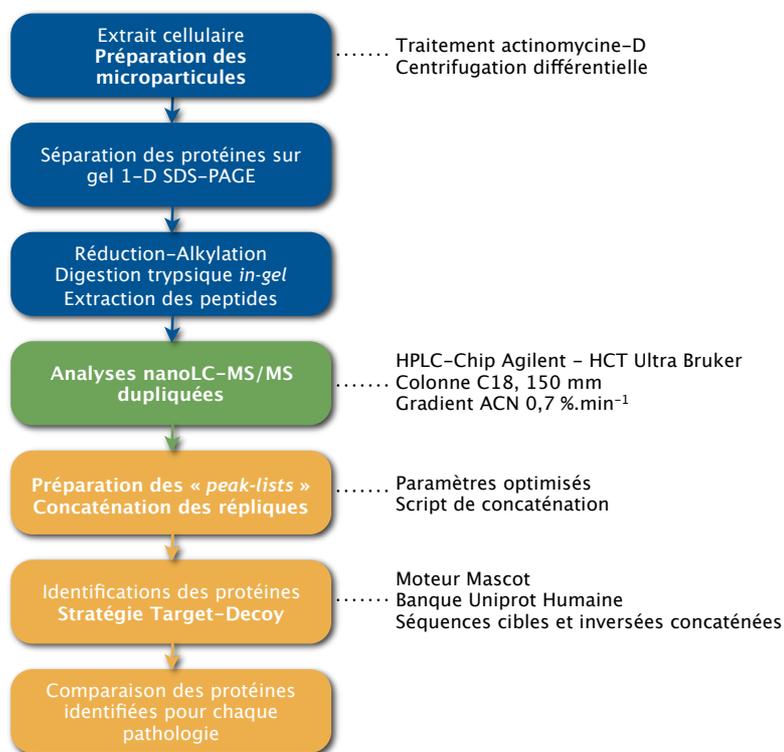
<sup>12</sup> Rabilloud, T. (2003). Nat Biotechnol 21, 508-10.

part, l'hypothèse qu'une protéine non identifiée est réellement absente de l'échantillon analysé doit pouvoir être confortée.

Ces conditions sont nécessaires afin de conclure quant à la spécificité d'une protéine qui serait identifiée dans un seul des échantillons comparés. La stratégie mise au point pour cette étude s'appuie alors sur le travail d'optimisation des méthodes et des outils utilisés pour l'analyse nanoLC-MS/MS décrits dans le chapitre précédent.

▸ La stratégie d'analyse optimisée

La plupart des étapes de la stratégie proposée ont bénéficié d'optimisations spécifiques (Figure 6). Elles ont permis de favoriser l'analyse de protéines membranaires (préparation de microparticules), de mettre au point des conditions d'analyse favorables (chromatographie sur puce microfluidique, cycles d'acquisition optimisés), d'améliorer le recouvrement du contenu protéique total et la répétabilité des mesures (duplication des analyses, concaténation des fichiers de résultats avant la recherche), et de renforcer la confiance en les identifications réalisées (approche « *Target-Decoy* » pour le choix des seuils de sélection et l'évaluation d'un taux de faux positifs).



**Figure 6** | Stratégie analytique optimale permettant l'analyse des protéines membranaires de lymphocytes B. Figurent en **gras**, les étapes qui présentent une approche originale ou qui ont bénéficié d'optimisations particulières.

▸ Application de la stratégie analytique

Les instruments de nouvelle génération (HPLC-Chip Agilent, HCT Ultra Bruker) n'étant pas disponibles au laboratoire au début du projet et l'étude de la répétabilité des analyses nanoLC-MS/MS n'ayant débuté qu'au même moment, la première étude de recherche de biomarqueurs spécifiques du lymphome du manteau n'a pas pu disposer de toutes les innovations. Les analyses ont été réalisées une seule fois sur un couplage CapLC - Q-TOF II (Waters Corp.). Cependant, les fichiers d'analyse ont pu être traités par l'approche « *Target-Decoy* » pour valider et optimiser les identifications.

Le deuxième projet, autour de la pathologie de Waldenström, a bénéficié de tous les outils mis au point pour une analyse optimale.

## 2.2. Stratégie visant à mettre en évidence des marqueurs membranaires potentiels

### 2.2.1. Définition de filtres de sélection de candidats biomarqueurs pertinents

Les listes de protéines résultant de l'analyse de chaque pathologie sont comparées afin de mettre en évidence des biomarqueurs potentiels. Toutes les protéines identifiées de manière spécifique ne sont pas biologiquement pertinentes, ni validées avec la même confiance. Nous avons établi une série de filtres représentant des critères biologiques et des critères d'identification MS qui permettent de sélectionner les protéines les plus intéressantes parmi celles identifiées.

Un panel de différents filtres peut être utilisé :

- Des filtres d'identification des protéines fondés sur l'approche « *Target-Decoy* ». Ils permettent, en ajustant directement le score seuil de validation des identifications, d'augmenter la confiance ou d'affiner la sensibilité. Ces filtres sont directement appliqués aux résultats bruts des moteurs de recherche pour générer la liste des protéines identifiées.

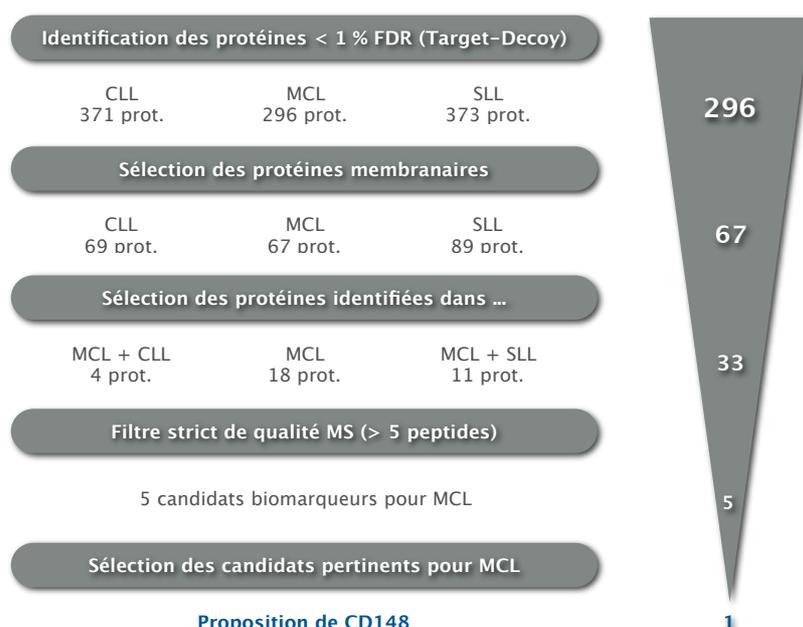
- Des filtres de sélection selon la fonction ou le compartiment cellulaire des protéines. Dans notre cas par exemple, les protéines membranaires étant ciblées, elles ont été sélectionnées à partir de données « *Gene Ontology* » qui rassemblent des informations sur la fonction et la localisation cellulaire des protéines<sup>13</sup>.

- Des filtres comparatifs entre échantillons. Les protéines identifiées dans une ou quelques pathologies peuvent être sélectionnées afin de faire émerger les protéines spécifiques à certains cas.

- Des filtres reflétant des différences dans le niveau d'expression des protéines entre les pathologies. Il a été montré que le nombre de peptides identifiés pour une protéine était lié à sa concentration dans l'échantillon<sup>14</sup>. Une différence importante dans le nombre de peptides validés peut donc révéler une sur- ou sous-expression de la protéine considérée

- Des filtres de pertinence biologique choisis selon les informations rapportées dans la littérature.

### 2.2.2. Application au cas du lymphome du manteau



**Figure 7** | Filtres de pertinence biologique et de qualité MS conduisant à la proposition d'un marqueur membranaire potentiel spécifique du lymphome du manteau.

<sup>13</sup> Rhee, S. Y., et al. (2008). *Nat Rev Genet* **9**, 509–15.

<sup>14</sup> Liu, H., et al. (2004). *Anal Chem* **76**, 4193–201.

La figure 7 présente la combinaison de filtres que nous avons appliqués à pour la recherche de biomarqueurs potentiels du lymphome du manteau. Ces filtres ont permis la sélection de 5 protéines (Table 2) dont la pertinence en tant que biomarqueur potentiel de MCL a été évaluée au regard de leur fonction biologique et de la qualité de leur identification.

Protéine	Numéro d'accension	Masse moléculaire (Da)	Nombre de peptides identifiés par pathologie		
			LLC	LNH	MCL
Chloride intracellular channel protein 1 – Homo sapiens (Human)	O00299 CLIC1_HUMAN	26 905,3		5	6
HLA class I histocompatibility antigen, A-24 alpha chain precursor – Homo sapiens (Human)	P05534 1A24_HUMAN	40 670,1			6
Plexin-A1 precursor – Homo sapiens (Human)	Q9UIW2 PLXA1_HUMAN	211 071,9			9
Receptor-type tyrosine-protein phosphatase eta precursor – Homo sapiens (Human)	Q12913 PTPRJ_HUMAN	145 924,0		4	10
Transgelin-2 – Homo sapiens (Human)	P37802 TAGL2_HUMAN	22 373,9			6

**Table 2** | Liste des 5 protéines passant chacun des filtres successifs pour l'identification de candidats biomarqueurs pour MCL.

Parmi ces 5 protéines, HLA (P05534) n'a pas été retenue car, de la classe des antigènes d'histocompatibilité (impliqués dans la reconnaissance du soi), elle ne représente pas un candidat biomarqueur pertinent. La protéine CLIC1 (O00299) qui sert dans le transport des chlorures et la transduction du signal, n'a pas été retenue puisqu'elle n'a pas été identifiée de manière différentielle entre LNH et MCL. Plexin-A1 (Q9UIW2) et Transgelin-2 (P37802) sont impliquées dans le remodelage du cytosquelette. Ce processus fait partie des mécanismes conduisant à la formation de microparticules. Les deux protéines n'ont donc pas été retenues comme candidat de premier choix. Finalement, seul le Cluster de Différenciation CD148 (Q12913, aussi nommé Receptor-type tyrosine-protein phosphatase eta precursor) a été identifié comme le candidat biomarqueur le plus crédible.

L'identification de CD148 dans l'échantillon de MCL a été validée avec minutie. Son occurrence est validée dans 3 bandes de gel 1-D successives avec un total de 10 peptides répartis sur toute la séquence protéique, conduisant à un recouvrement de plus de 8 % (Figure 8).

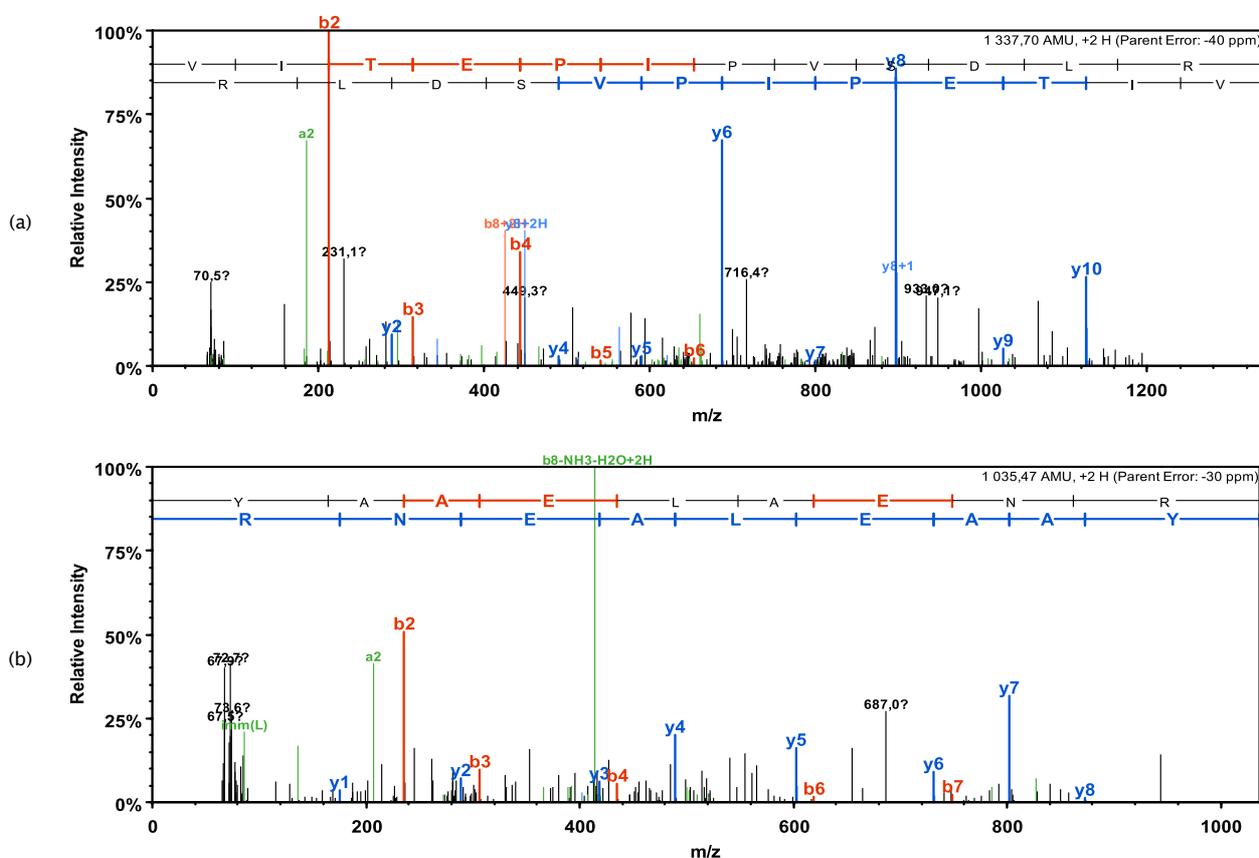
CD148 – Homo sapiens (Human) (ou Receptor-type tyrosine-protein phosphatase eta precursor)										Q12913 PTPRJ_HUMAN										
1	MKPAA	REARL	PPRSP	GLRWA	LPLLL	LLLRL	GQILC	AGGTP	SPIPD	PSVAT										
51	VATGE	NGITQ	ISSTA	ESFHK	QNGTG	TPQVE	TNTSE	DGESS	GANDS	LRTPE										
101	QGSNG	TDGAS	<b>QKTPS</b>	<b>STGPS</b>	<b>PVFDI</b>	<b>KAVSI</b>	SPTNV	ILTWK	SNDTA	ASEYK										
151	YVVKH	KMENE	KTITV	VHQPV	CNITG	LRPAT	SYVFS	ITPGI	GNETW	GDPKV										
201	<b>IKVIT</b>	<b>EPIPV</b>	<b>SDLRV</b>	ALTGV	RKAAL	SWSNG	NGTAS	CRVLL	ESIGS	HEELT										
251	QDSRL	QVNIS	GLKPG	VQYNI	NPYLL	QSNKT	KGDPL	GTEGG	LDASN	TERSR										
301	AGSPT	APVHD	ESLVG	PVDPV	SGQQS	<b>RDTEV</b>	<b>LLVGL</b>	<b>EPGTR</b>	YNATV	YSQAA										
351	NGTEG	QPQAI	EFRTN	AIQVF	DVTAV	NISAT	SLTLI	WKVSD	NESSS	NYTYK										
401	IHVAG	ETDSS	NLNVS	EPRAV	IPGLR	SSTFY	NITVC	PVLGD	IEGTP	GFLQV										
451	HTPPV	PVSDF	RVTVV	STTEI	GLAWS	SHDAE	SFQMH	ITQEG	AGNSR	VEITT										
501	NQSII	IGGLF	PGTKY	<b>CFEIV</b>	<b>PKGPN</b>	GTEGA	SRTVC	NRTVP	SAVFD	IHVVY										
551	VTTTE	MWLDW	KSPDG	ASEYV	YHLVI	ESKHG	SNHTS	TYDKA	ITLQG	LIPGT										
601	LYNIT	ISPEV	DHVWG	DPNST	AQYTR	PSNVS	NIDVS	TNTTA	ATLSW	QNFDD										
651	ASPTY	SYCLL	IEKAG	NSSNA	TQVVT	DIGIT	DATVT	ELIPG	SSYTV	EIFAQ										
701	VGDGI	KSLEP	GRKSF	CTDPA	SMASF	DCEVV	PKEPA	LVLKW	TCPPG	ANAGF										
751	ELEVS	SGAWN	NATHL	ESCSS	ENGTE	YRTEV	TYLNF	STSYN	ISITT	VSCGK										
801	MAAPT	RNTCT	TGITD	PPPPD	GSPNI	TSVSH	NSVKV	KFSGF	EASHG	PIKAY										
851	AVILT	TGEAG	HPSAD	VLKYT	YDDFK	KGASD	TYVTY	LIRTE	EKGRS	<b>QSLSE</b>										
901	<b>VLKYE</b>	IDVGN	ESTTL	GYNG	<b>KLEPL</b>	<b>GSYRA</b>	CVAGF	TNITF	HPQNK	<b>GLIDG</b>										
951	<b>AESYV</b>	<b>SFSRY</b>	SDAVS	LPQDP	GVICG	AVFGC	IFGAL	VIVTV	GGFIF	WRKKR										
1001	KDAKN	NEVSF	SQIKP	KKSKL	IRVEN	<b>FEAYF</b>	<b>KKQQA</b>	DSNCG	FAEYV	EDLKL										
1051	VGISQ	<b>PKYAA</b>	<b>ELAEN</b>	RGKNR	YNNVL	PYDIS	RVKLS	VQTHS	TDDYI	NANYM										
1101	PGYHS	KKDFI	ATQGP	LNPTL	KDFWR	MVWEK	NVYAI	IMLTK	CVEQG	RTKCE										
1151	EYWPS	KQAQD	YGDIT	VAMTS	EIVLP	EWTIR	DFTVK	<b>NIQTS</b>	<b>ESHPL</b>	<b>RQFHF</b>										
1201	TSWPD	HGVDP	TTDLL	INFRY	LVRDY	MKQSP	PESPI	LVHCS	AGVGR	TGTFI										
1251	AIDRL	IYQIE	NENTV	DVYGI	VYDLR	MHRPL	MVQTE	DQYVF	LNQCV	LDIVR										
1301	SQKDS	KVDLI	YQNTT	AMTIY	ENLAP	VTTFG	KTNGY	IA												

**Figure 8** | Séquence de CD148 et peptides identifiés (en rouge) par nanoLC-MS/MS pour l'échantillon MCL. Le pourcentage de recouvrement global obtenu s'élève à 8 %.

La table 3 présente le détail des identifications des 10 peptides de CD148 et la figure 9 présente les spectres de fragmentation de 2 d'entre eux. Ces résultats établissent indiscutablement l'identification de CD148 pour MCL. De plus, le différentiel important en terme de nombre de peptides identifiés pour CD148 entre MCL (10 peptides) et LNH (4 peptides) / CLL (pas d'identification) suggère une surexpression potentielle du candidat biomarqueur pour MCL.

Séquence peptidique	Score d'ion Mascot	Score Identité Mascot	m/z mesuré	Masse du peptide calculée (Da)	Nombre de charges	Erreur sur la mesure de masse de l'ion parent (ppm)
YAAELAENR	18,9	34,2	518,74	1 035,46	2	-30,34
YcFEIVPK	10	34,2	528,25	1 054,49	2	-22,88
GLIDGAESYVFSR	12,7	33,4	750,85	1 499,69	2	-21,37
LEPLGSYR	18,4	34,6	467,74	933,47	2	-19,09
SQSLSEVLK	9,08	34,6	495,76	989,51	2	-27,79
VENFEAYFK	2,43	34,2	573,76	1 145,51	2	-25,01
NIQTSESHPLR	21,2	33,5	641,30	1 280,59	2	-38,99
VITEPIVSDLR	13	33,7	669,85	1 337,70	2	-40,08
TPSSTGSPVFDIK	6,2	33,5	716,84	1 431,68	2	-30,97
DTEVLLVGLPEGTR	53,6	33,1	749,88	1 497,75	2	-34,28

**Table 3** | Liste des 10 peptides de CD148 (Q12913) identifiés pour l'échantillon MCL.



**Figure 9** | (a) Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à m/z 669,85, correspondant au peptide VITEPIVSDLR. (b) Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à m/z 518,74, correspondant au peptide YAAELAENR.

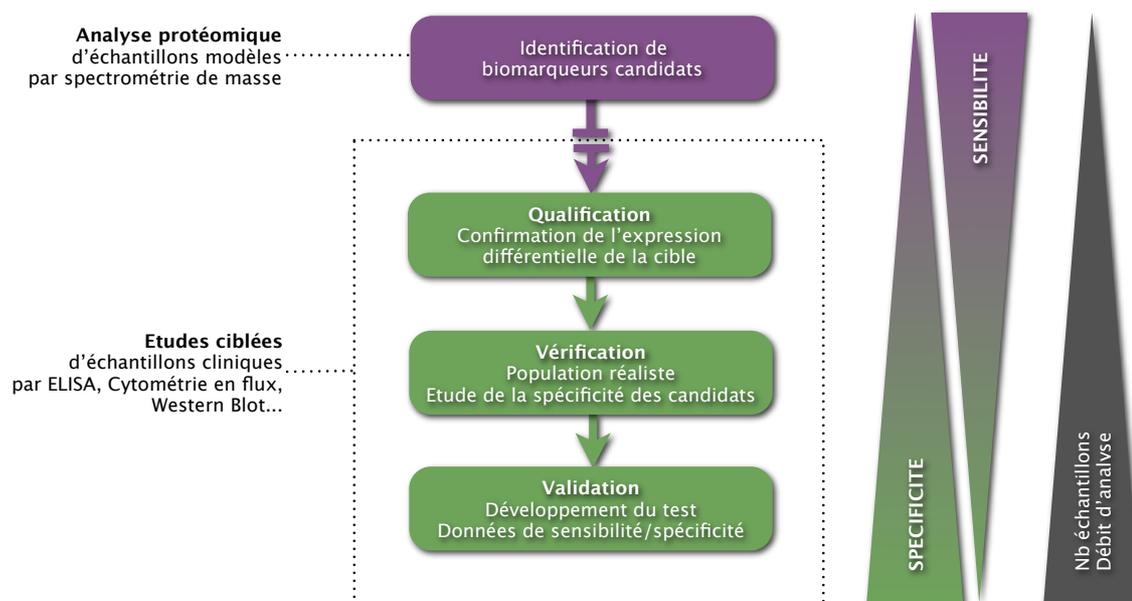
## 2.3. La validation des cibles proposées

A ce jour, toutes pathologies confondues, on ne rencontre pas de succès indiscutable où l'analyse protéomique a été à l'origine de l'application commerciale d'un biomarqueur qu'elle suggérait. Ce faible retour sur investissement, imputable à la difficulté de franchir le guet entre les phases d'identification et de validation, a soulevé de sérieux doutes quant à de possibles futurs succès<sup>15</sup>.

La collaboration entretenue avec l'équipe de médecins sur nos projets a permis de faire déboucher la phase initiale d'identification de cibles potentielles sur un début de processus de validation clinique des candidats proposés.

### 2.3.1. Le « pipeline » de validation de biomarqueurs

La validation clinique de candidats est un travail nécessaire et indispensable pour proposer un biomarqueur utilisable en routine en milieu hospitalier. Des processus cohérents et complets conduisant à la validation et à l'utilisation clinique de biomarqueurs ont été décrits<sup>16</sup>. Ils distinguent 6 phases essentielles : l'identification, la qualification, la vérification, l'optimisation et la validation des tests cliniques et finalement, la commercialisation (Figure 10).



**Figure 10** | La « distillation » de biomarqueurs pertinents d'un lot de candidats est le plus grand défi pour l'émergence de biomarqueurs cliniquement utiles (d'après Rifai, N. *et al.* (2006) *Nat Biotechnol* 24, 971–83).

Les phases de qualification et de vérification permettent généralement d'examiner la pertinence des candidats en mesurant leurs caractéristiques de performance. S'appuyant sur des expériences quantitatives sur plusieurs dizaines ou centaines de patients, elles permettent de vérifier que la concentration des protéines candidates est significativement différente entre la pathologie considérée et les autres cas. C'est également l'occasion de démontrer la corrélation entre l'observation réalisée lors de l'identification des candidats sur des échantillons modèles et les mesures prises sur des échantillons réels, peu ou non simplifiés. Ces deux phases initiales permettent de déterminer les cibles qui méritent d'être retenues pour des études sur des cohortes d'essais plus larges. Ces phases avancées requièrent des investissements en temps et en argent importants afin de développer des tests cliniques et d'optimiser leur robustesse.

<sup>15</sup> Lescuyer, P., *et al.* (2007). *J Proteome Res* 6, 3371–6.

<sup>16</sup> Rifai, N., *et al.* (2006). *Nat Biotechnol* 24, 971–83.

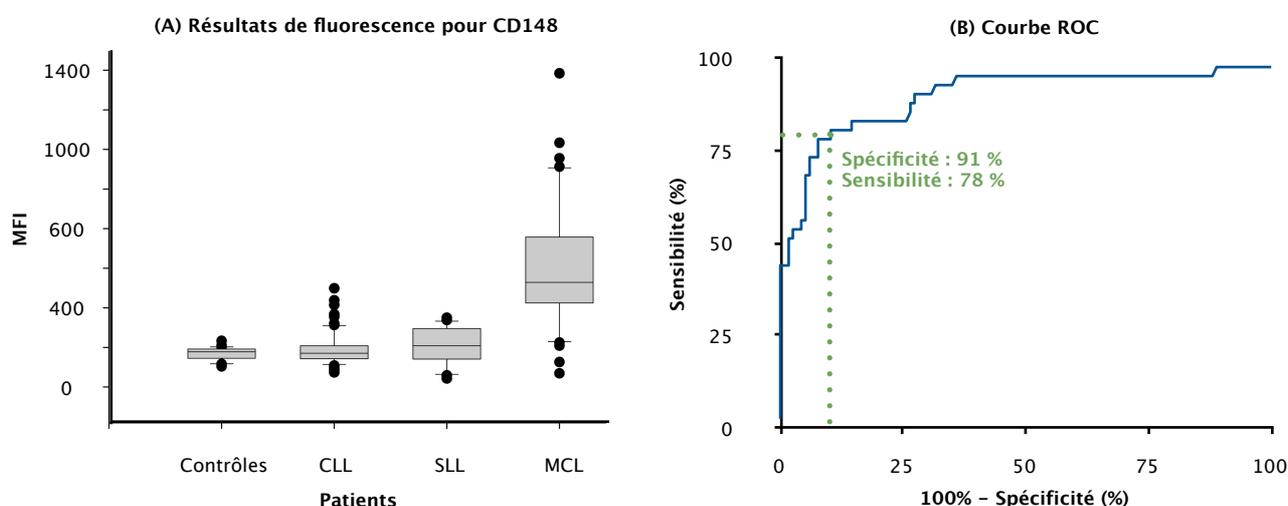
### 2.3.2. Validation du biomarqueur candidat CD148 pour le lymphome du manteau

Nous avons pu faire quelques pas sur le chemin de la validation du biomarqueur candidat CD148 que nous avons identifié pour le lymphome du manteau. L'anticorps commercial spécifique à la protéine membranaire ciblée existe, il a donc pu être utilisé pour une recherche sur 158 patients atteints de l'une des trois pathologies comparées (41 MCL, 24 SLL, 93 CLL) et sur 30 cas contrôles. Des mesures de cytométrie en flux ont pu être réalisées, confirmant la surexpression de CD148 dans le lymphome du manteau (Figure 11A).

Des données de **sensibilité** et de **spécificité** ont par ailleurs été proposées. La sensibilité d'un biomarqueur pour une pathologie est la probabilité que le biomarqueur soit surexprimé si le sujet est atteint de la pathologie considérée. Un test diagnostique est donc d'autant plus sensible que les sujets atteints de la pathologie présentent plus souvent une surexpression du biomarqueur.

La **spécificité** d'un biomarqueur pour une pathologie est la probabilité que le biomarqueur soit exprimé de manière normale si le sujet n'est pas atteint de la pathologie. Un test diagnostique est donc d'autant plus spécifique que les sujets sains présentent plus souvent le biomarqueur exprimé normalement. Pour un test « parfait », c'est-à-dire n'autorisant aucune erreur, les valeurs de la sensibilité et de la spécificité sont égales à 1.

Le meilleur seuil entre les valeurs pathologiques et les valeurs normales doit être défini de manière à obtenir les plus fortes valeurs pour les paramètres de sensibilité et de spécificité, sachant qu'ils varient en sens inverse. On s'aide pour ce choix de courbes ROC (Receiver Operating Characteristics) permettant de représenter la fonction  $f(1-\text{spécificité}) = \text{sensibilité}$ . La figure 11B présente la courbe ROC pour un niveau d'expression de CD148 2 fois supérieur au niveau normal. Elle montre que le diagnostic de MCL peut être établi en utilisant ce seuil avec 91 % de spécificité et 78 % de sensibilité.



**Figure 11** | (A) Comparaison de l'intensité de fluorescence moyenne pour CD148 mesurée par cytométrie en flux, pour les échantillons contrôle et les cas CLL, MCL et SLL. (B) Courbe ROC (Receiver Operator Characteristic) indiquant que pour un niveau d'expression de CD148 2 fois supérieur aux cellules normales, le diagnostic de MCL peut être validé avec 91 % de spécificité et 78 % de sensibilité.

Ces résultats encourageants doivent maintenant permettre au biomarqueur potentiel CD148 de franchir des étapes de validation plus avancées pour qu'il s'impose comme élément discriminant pour le diagnostic du lymphome du manteau vis-à-vis du lymphome à petites cellules et de la leucémie lymphoïde chronique. Une étude multisite dans le cadre d'un programme hospitalier de recherche clinique (PHRC) est en cours de préparation. Si cette étude confirme les résultats obtenus sur les quelques centaines de cas testés, alors le CD148 pourra être considéré comme un *vrai nouveau biomarqueur*.

En parallèle, le laboratoire s'apprête à développer une stratégie de quantification par spectrométrie de masse par MRM (cf. introduction bibliographique, chapitre 2, point 2.1.3). Cette technique de dosage des

protéines différentielles est plus spécifique qu'une méthode immunologique (détection des isoformes, des modifications posttraductionnelles éventuelles...) et ne nécessite pas d'anticorps commerciaux *ad hoc*.

## 2.4. Publication de ces travaux

.....  
: Miguet, L.\*, Bechade, G.\*, Fornecker, L., Zink, E., Felden, C., Gervais, C., Herbrecht, R., van Dorsselaer, A.,  
: Mauvieux, L. et Sanglier-Cianferani, S.  
:  
: Proteomic Analysis of Malignant B-Cell Derived Microparticles Reveals CD148 as a Potentially Useful Antigenic  
: Biomarker for Mantle Cell Lymphoma Diagnosis.  
:  
: J Proteome Res (2009) **8**, 3346–3354.  
:.....  
\* These authors equally contributed to this work

Les données supplémentaires associées à cette publication sont fournies en annexe (A1).

## Proteomic Analysis of Malignant B-Cell Derived Microparticles Reveals CD148 as a Potentially Useful Antigenic Biomarker for Mantle Cell Lymphoma Diagnosis

Laurent Miguet,<sup>†,‡,§,#</sup> Guillaume Béchade,<sup>†,#</sup> Luc Fornecker,<sup>‡,§</sup> Estelle Zink,<sup>‡,§</sup> Claire Felden,<sup>‡</sup> Carine Gervais,<sup>§</sup> Raoul Herbrecht,<sup>§</sup> Alain van Dorselaer,<sup>\*,†</sup> Laurent Mauvieux,<sup>‡,§</sup> and Sarah Sanglier-Cianferani<sup>\*,†</sup>

*Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DNA, UDS, CNRS, UMR7178, ECPM 25 rue Becquerel, 67087 Strasbourg, France, Laboratoire d'Hématologie cellulaire EA 3948, Institut d'Hématologie, Strasbourg, France, and Pôle d'Onco-Hématologie - Hôpitaux Universitaires de Strasbourg, France*

Received December 22, 2008

The diagnosis of mature B-cell neoplasms (MBCN) remains difficult in a number of cases, especially leukemic phases of non-Hodgkin lymphoma, for which discriminating criteria or biomarker are often lacking. To identify new surface biomarkers, we developed an original proteomic approach based on mass spectrometry analysis of plasma membrane microparticles derived from chronic B-cell lymphoproliferations of single patients: chronic lymphocytic leukemia (CLL), small cell lymphoma (SLL) and mantle cell lymphoma (MCL). A straightforward selection process for proteomic-based candidate biomarker identification was further constructed in order to propose potentially useful and relevant biomarkers. Comparison of the lists of the proteins identified in each pathology combined to highly stringent MS validation criteria for protein identification allowed to propose CD148, a membrane receptor with phosphatase activity, as a discriminating biomarker candidate. Flow cytometry analyses, performed on 158 patients and 30 controls, showed that an anti-CD148 antibody stained significantly higher MCL than CLL and SLL circulating cells ( $p < 0.0001$ ), which validates CD148 overexpression in MCL. Our results indicate that a medium or high CD148 expression level may exclude the diagnosis of CLL and high CD148 expression levels (CD148 MFI equal or superior to 2 times the value obtained with CLL/SLL) allows MCL diagnosis to be suspected with 91% specificity (versus CLL and SLL) and 78% sensitivity. This study is one of the first where proteomic strategies allowed to identify a potentially useful biomarker.

**Keywords:** mass spectrometry • proteomics • plasma membrane microparticles • chronic B-cell malignancies • CD148 • mantle cell lymphoma • biomarker

### Introduction

Mature B-cell neoplasms (MBCN) include several entities delimited within the WHO classification.<sup>1</sup> Facing a B-cell hyperlymphocytosis, morphological examination is the first step for diagnosis, which allows to distinguish well-defined entities such as chronic lymphocytic leukemia (CLL), small cell lymphoma (SLL) from hairy cell leukemia or splenic lymphoma with villous lymphocytes. However, due to their marked heterogeneity, many cases remain cytologically ambiguous or unclassified in overlapping entities. Immunophenotype classification is of great help for MBCN diagnosis<sup>2</sup> and investigation

of minimal residual disease,<sup>3</sup> but remains nondiscriminating for some leukemic phases of non-Hodgkin lymphoma (NHL).

Cytogenetic abnormalities can be specific for some MBCN or define prognosis factors, mainly for CLL. New prognosis biomarkers, including CD38, Zap70 expression or mutation status of immunoglobulin heavy locus<sup>4</sup> are extensively studied in this pathology, but are limited to CLL. However other lymphoma still raise routinely complicated diagnostic problems, for instance, marginal zone lymphoma, lymphoplasmacytic lymphoma, and mantle cell lymphoma (MCL).<sup>5–8</sup> MCL, the most aggressive form of MBCN, can be particularly difficult to diagnose with a very variable cytology and ambiguous phenotypes. Moreover, if the translocation t(11;14) leading to a cyclin D1 hyperexpression is a hallmark of the disease, it is not MCL specific and is observed in other MBCN.<sup>8,9</sup>

Despite the ongoing diagnostic advances, the management of NHL patients remains complex, with overlapping diseases and markedly disparate presentations and clinical outcomes. For all these reasons, there is a need for new biomarkers that

\* Corresponding authors: Alain van Dorselaer and Sarah Sanglier-Cianferani, Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DNA, UDS, CNRS, UMR7178, ECPM - 25, rue Becquerel, 67087 Strasbourg - France, email: vandors@chimie.u-strasbg.fr and sarah.sanglier@chimie.u-strasbg.fr, Tel: 00 33 (0)3 90 24 27 83 Fax: 00 33 (0)3 90 24 27 81.

<sup>†</sup> Laboratoire de Spectrométrie de Masse Bio-Organique.

<sup>‡</sup> Institut d'Hématologie.

<sup>§</sup> Pôle d'Onco-Hématologie - Hôpitaux Universitaires de Strasbourg.

<sup>#</sup> These authors equally contributed to this work.

could allow a more precise diagnostic/classification, and that could also represent therapeutic targets. The most valuable biomarkers would be surface antigens. So far, the search for new biomarkers has mostly been driven by the availability of monoclonal antibodies.<sup>10</sup> Development of proteomics and mass spectrometry techniques has allowed the identification and characterization of hundreds of proteins constituting in a single sample.<sup>11–16</sup> These proteomic techniques would therefore be of great interest for proliferative lymphoid cell antigen discovery. As surface biomarkers would be the most appropriate, plasma membrane proteomics is attractive. However, isolation of plasma membrane proteins is difficult because of a laborious, often imperfect purification. Moreover, the classical two-dimensional gel analysis approach was shown to be inappropriate for separation of hydrophobic proteins.<sup>17,18</sup> To circumvent these drawbacks, we previously developed an original method to optimize the analysis of membrane proteins, using secreted plasma membrane microparticles (MPs) and 1D-gel separation.<sup>19</sup> MPs are physiologically secreted in various situations and can be induced by actinomycin D or phytohemagglutinin. This specific methodology allowed the identification of several hundred proteins in the lymphocytic CEM cell line, even if samples were available in relatively small amounts. These proteins included numerous plasma membrane or plasma membrane associated proteins, CD antigens, annexins and signal transduction associated proteins.<sup>19,20</sup>

These preliminary results prompted us to look for surface antigens as potential new biomarkers differentially expressed in various MBCN. In the present work, MPs from hyperlymphocytic patients suffering from CLL, SLL and MCL were generated and their protein contents were investigated with the previously developed proteomic approach.

A straightforward selection process based on membrane localization of the biomarker and highly stringent MS validation criteria for protein identifications<sup>21</sup> was developed in order to propose potentially useful candidate biomarkers. Finally, a limited set of “realistic” candidate biomarkers for MCL was delivered and further submitted to quantitative immunological validation. Among them, we focus in the present study on CD148, a protein differentially identified in MCL by the productive but time-consuming proteomic approach. A differential expression of this protein was confirmed in a limited series of well-documented patients by flow cytometry analyses, suggesting that this two-step strategy is adapted and can be extended to different hematological malignancies.

## Experimental Section

**Patients.** Fresh cells were isolated from blood samples from patients with chronic lymphocytic leukemia and small cell lymphoma (CLL  $n = 93$ , SLL,  $n = 24$ ), mantle cell lymphomas (MCL,  $n = 41$ ) and normal controls ( $n = 30$ ). Initially, patients were included in a study approved by local ethic committee (CCPPRB Alsace, #0279, “Programme de recherche hospitalier des Hôpitaux Universitaires de Strasbourg” #2674). All patients were studied using morphology, immunophenotype and cytogenetics in the clinical hematology laboratory of Hautepierre Hospital (Strasbourg, France). Cytological data were reviewed by at least two different biologists, as well as cytogenetics and molecular data. Diagnosis were assigned following World Health Organization (WHO) classification.<sup>1</sup> All CLL and SLL samples were cytologically univocal, with a Matutes score equal or superior to 4. The diagnosis of MCL was based on cell morphology, immunophenotype as well as the demonstration

of cyclin D1 expression and/or t(11;14) translocation, following WHO recommendations.<sup>1</sup> No follicular lymphoma nor prolymphocytic leukemia were included in this group. Thirty samples used for normal controls were randomly and anonymously selected within remaining blood samples after routine analyses for nonhematological disease.

**Isolation of Membrane MPs from MBCN.** An average of  $200 \times 10^6$  cells were isolated from 3 mL of peripheral blood samples of well-documented hyperlymphocytic patients: one CLL, one SLL and one MCL. Ficoll isolated cells were treated with Actinomycin D (0.5  $\mu\text{g}/\text{mL}$ , Sigma) for 18 h in serum-free X-VIVO 15 medium (Cambrex, East Rutherford, NJ). A supernatant was obtained by centrifugation at 750g for 15 min and then at 1500g for 5 min to remove cells and large fragments, respectively. The latter supernatant was recentrifuged for 45 min at 15 000g. The pelleted MPs were resuspended in 15 mL of Hanks buffered saline solution (HBSS) and centrifuged again for 45 min at 15 000g. The pellet was recovered in 1 mL of HBSS and stored at  $-80^\circ\text{C}$  until proteomic analysis.

**Proteomic Studies. 1. 1D-SDS-PAGE.** Crude MPs pellet were resuspended in 80  $\mu\text{L}$  of SDS-PAGE buffer, before being loaded onto a 8–15% acrylamide SDS-PAGE gel ( $20 \times 20$  cm). After 5 h of migration at 30 mA, the gel was stained with R250 coomassie blue (Sigma). Each electrophoretic lane was cut systematically at 2 mm intervals and the gel bands were transferred into a 96 well plate.

**2. In-Gel Digestion.** In-gel digestion was performed with an automated protein digestion system, a MassPrep Station (Waters, Manchester, U.K.). The gel plugs were washed twice with 50  $\mu\text{L}$  of 25 mM ammonium hydrogen carbonate ( $\text{NH}_4\text{HCO}_3$ ) and 50  $\mu\text{L}$  of acetonitrile. The cysteine residues were reduced by addition of 50  $\mu\text{L}$  of 10 mM dithiothreitol at  $57^\circ\text{C}$  and alkylated by addition of 50  $\mu\text{L}$  of 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins were cleaved in-gel with a 12.5  $\text{ng}/\mu\text{L}$  solution of modified porcine trypsin (Promega, Madison, WI) in 25 mM  $\text{NH}_4\text{HCO}_3$  ( $\sim 30 \mu\text{L}$ ). The digestion was performed overnight at room temperature. Tryptic peptides were extracted twice: the first time with 60% acetonitrile in 5% formic acid for 1 h, and the second time with a 100% acetonitrile solution until the gel pieces were dehydrated. The collected extracts were pooled to a final volume of 60  $\mu\text{L}$ . Excess of acetonitrile was evaporated at  $37^\circ\text{C}$  before analysis.

**3. NanoLC-MS/MS Analysis.** Nanocapillary liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) analyses were performed using a CapLC system (Waters, Manchester, U.K.), coupled to a hybrid Electrospray Quadrupole Time-Of-Flight mass spectrometer (Q-TOF II, Waters, Manchester, U.K.). A total of 6.4  $\mu\text{L}$  of each sample was loaded onto a precolumn, before chromatographic separation on a  $\text{C}_{18}$  column (LC packings  $\text{C}_{18}$ , 75  $\mu\text{m}$  i.d., 150 mm length). Separation was performed at 200 nL/min flow rate using a 35 min gradient from 5 to 45% B (eluent A, 0.1% (v/v) formic acid in water and eluent B, 0.1% (v/v) formic acid in acetonitrile). The ESI capillary voltage was set to 3500 V and cone voltage to 35 V. Mass calibration of the TOF was achieved using phosphoric acid ( $\text{H}_3\text{PO}_4$ ) on the  $m/z$  50–2000 range.<sup>22</sup> The MS survey scan was acquired on the range  $m/z$  300–1500 with a scan time of 1 s and an interscan time of 0.1 s. Automatic switching between MS and MS/MS modes was used. The 3 most intense ions (intensity threshold of 10 counts) were selected on each MS spectrum for further isolation and CID fragmentation. Fragmentation was performed using argon as collision gas and with

a collision energy profile optimized for various mass ranges of precursor ions. The scan range for MS/MS acquisition was from  $m/z$  50 to 2000 with a scan time of 1 s and an interscan time of 0.1 s.

Data acquisition was piloted by MassLynx software V4.0 SP1. Mass data collected during a nanoLC-MS/MS analysis were automatically processed and converted into a .pkl using MassLynx software V4.0 SP1. Normal background subtraction type was used for both MS and MS/MS with 5% threshold, a polynomial correction of order 5 and deisotoping. Peak list files were submitted to Mascot search engine (Matrix Science, London, U.K.) installed in its 2.2.0 version on a local server. Searches were performed against a composite target-decoy database including both forward (target) and reversed (decoy) Swiss-Prot protein sequences of Human (9 July 2008, 39 862 total entries). Target-decoy database was generated using in-house Perl-CGI scripts. Searches were performed with a tolerance on mass measurement of 0.25 Da for both precursor and fragment ions. Carbamidomethylation of cysteine residues, oxidation of methionine residues and acetylation of protein N-term residues were searched as variable modifications. Up to one missed cleavage was allowed.

Scaffold V2.00.03 (Proteome Science, Portland, OR) was used for identification validation and false positive rate estimation for protein identification. For this, specific Mascot score cutoffs were empirically determined for each sample analysis to maximize the number of proteins identified, keeping the false positive rate <1% (number of decoy database assignments divided by the total number of protein identifications in the target-decoy data set). For each sample, peptides were filtered out according to the cutoff set for proteins hits with 2 or more peptides (Mascot Ion Score - Mascot Identity Score > -20, -32, and -33 for CLL, SSL and MCL samples, respectively). In case of single peptide hit, a different and more stringent criterion was used (Mascot Ion Score - Mascot Identity Score > -6, +9 and +3 for CLL, SSL and MCL samples, respectively). Further data manipulations were performed with Microsoft Excel to finally obtain a catalogue of proteins for each sample with an estimated false positive rate for protein identification below 1%. Subcellular localization and classification of the proteins was assigned using the information extractor Gofetcher (Mississippi Computational Biology Consortium, Hattiesburg, MS, <http://mcbc.usm.edu/gofetcher>) searching the GO EBI database (<http://ebi.ac.uk>, Cambridge, U.K.).

**Flow Cytometry Immunophenotypic Studies.** All studies were performed using a FACScalibur flow cytometer (Becton Dickinson Bioscience, Franklin Lakes, NJ). Calibrite (Becton Dickinson Bioscience) beads were used to monitor the instrument settings for both the intensity and color compensation. Acquisition and analysis were performed using CellQuest software program (Becton Dickinson Bioscience).

**1. Quantitative Indirect Immunofluorescence CD148 Antigen Binding Assay.** Quantitative analysis of indirect immunofluorescence staining was performed by using the QIFIKIT purchased from DakoCytomation. QIFIKIT allows expression of fluorescence intensity in antibody-binding capacity (ABC) which represents the number of primary monoclonal antibody per cell. Briefly,  $5 \times 10^5$  cells were incubated for 30 min at 4 °C with CD148 nonconjugated primary mouse monoclonal antibody (clone 143-41, purchased from R&D Systems). After washing two times with phosphate buffer saline (PBS)/20% fetal calf serum (FCS), cells and beads (setup and calibration beads) were incubated for 45 min at 4 °C with FITC-conjugated anti-

mouse secondary antibody provided with the QIFIKIT. After washing 3 times with PBS/20% FCS, cells and beads were resuspended into 500  $\mu$ L of PBS. The lymphocyte populations were gated using side and forward scatter dot plots. Calibration beads were used for the construction of calibration curve (MFI versus ABC). For cell sample, ABC was determined by interpolation of its MFI on the calibration curve. Results were expressed into specific antibody-binding capacity (SABC), which was obtained after subtracting background antibody equivalent (BAE) due to background fluorescence (SABC = ABC - BAE).

Indirect quantitative assay was compared to the direct immunofluorescence staining that was realized as follows.

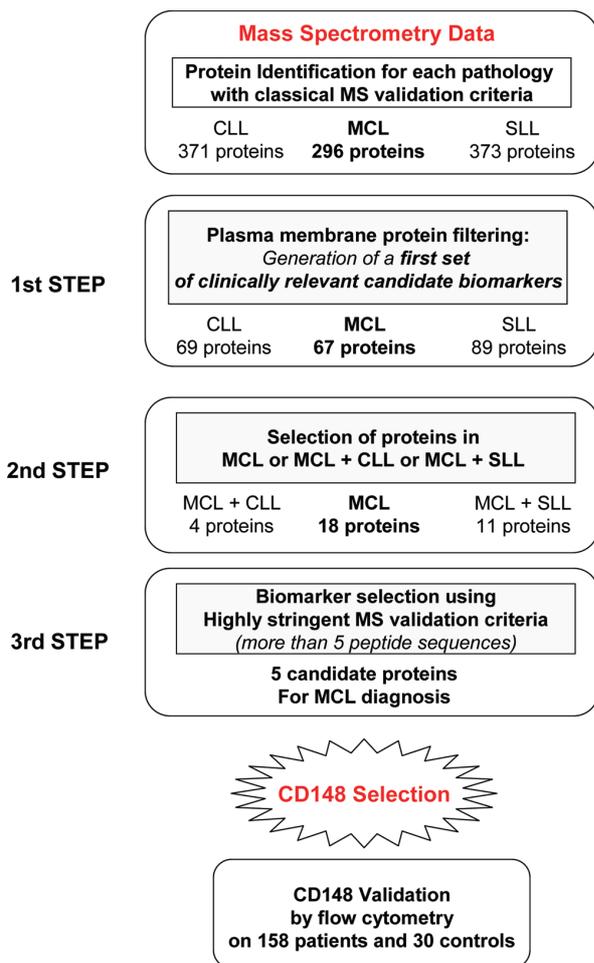
**2. Direct Immunofluorescence Staining.** Whole blood was stained with the following four-color combination of monoclonal antibodies CD45/CD19/CD3/CD148 directly conjugated with peridinin chlorophyll protein (PerCP), allophycocyanin (APC), fluorescein isothiocyanate (FITC) and phycoerythrin (PE), respectively. Appropriate directly conjugated isotypic antibodies were used as nonspecific immunostaining controls in all experiments. CD45 (clone 2D1), CD19 (clone SJ25C1), CD3 (clone SK7) and APC-conjugated isotypic control (clone X40) were purchased from Becton Dickinson Bioscience, FITC-conjugated isotypic control (clone 679.1Mc7) from Beckman Coulter, CD148 (clone 143-41) from Santa Cruz Biotechnology and R&D Systems and PE-conjugated isotypic control from Monosan. A total of  $5 \times 10^5$  cells were used per test and the appropriate saturating amounts of labeled monoclonal antibodies were added. Cell suspensions were incubated for 30 min at room temperature and were subsequently lysed. After washing with PBS, the obtained leukocytes were resuspended in PBS. The lymphocyte area was gated in a CD45 versus side scatter dot plot and 5000 CD19+ cells were acquired for each sample. The values of mean fluorescence intensity (MFI) were obtained after subtracting the autofluorescence values.

As a high correlation ( $r^2 > 0.992$ ) between quantitative indirect SABC measurements and direct fluorescence MFI was observed, direct MFI was used for CD148 expression analysis using flow cytometry.

## Results

**Proteomic-Based Approach to Propose Potentially Relevant MCL Candidate Biomarkers.** The previously published strategy for MP proteomic analysis<sup>19,20</sup> was applied to three different lymphocytic malignancies: one CLL, one SLL and one MCL patient for which diagnosis was assessed as described in the Experimental Section. Briefly, for each pathology, MP proteins were first separated on a 1D-SDS-PAGE gel followed by systematic band cutting and nanoLC-MS/MS analysis.

A straightforward selection process for proteomic-based biomarker identification was further developed in order to propose potentially useful and relevant protein candidates (Figure 1). The first step of our biomarker identification workflow consists of the generation of lists of proteins detected in each sample (i.e., for each pathology) with classical MS validation criteria (see Experimental Section). With these criteria, between 300 and 400 unique proteins were identified in each pathology (296 for MCL, 371 for CLL and 373 for SLL), leading to 579 nonredundant protein identifications when combining results obtained for the three pathologies. Among the 579 unique proteins identified with classical MS validation criteria, roughly 30% were common to all pathologies (300 were detected in at least two of the three types of malignancies).



**Figure 1.** Description of the proteomic based strategy for the proposition of CD148 as a potentially useful candidate biomarker.

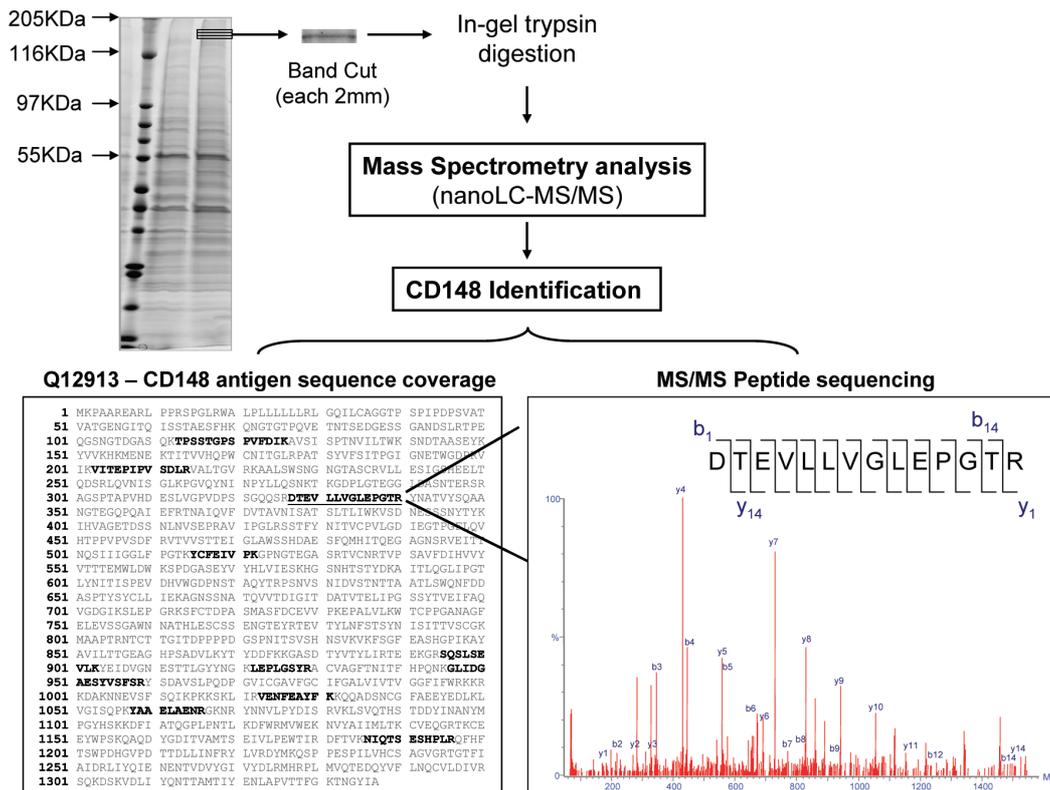
To reduce the number of proteins to those candidates that might be potentially useful, we first used a selection criterion based on the biological features of potential candidate proteins. We focused on plasma membrane proteins presenting extracellular domains which would be most appropriate for further flow cytometry analysis. This first selection step enabled to reduce the initial set of proteins to 67 for MCL, 69 for CLL and 89 for SLL (see Supporting Information S1 and S2). The next filter applied to search for relevant MCL biomarkers consisted of selecting proteins either uniquely identified in MCL or in MCL and another pathology (either CLL or SLL): all membrane proteins concomitantly present in the three pathologies were thus eliminated. This second step in the biomarker selection process allowed us to reduce the number of protein candidates from 67 to 33 for MCL (see Supporting Information S3). To provide the final limited set of “realistic” candidate biomarkers to be further validated in immunological studies, we applied highly stringent MS validation criteria in order to focus on proteins identified with a very high level of confidence in MCL, that is, with a large number of peptide sequences (more than 5 peptides) and high sequence coverage (see Supporting Information S3, proteins in bold boxes). Indeed, it has been shown that the amount of peptides used for the identification

of a protein is directly linked to the concentration of this protein in the sample.<sup>23,24</sup> Among the 33 surface plasma membrane proteins, most of them, including CD20 (P11836), CD37 (P11049) or CD82 (P27701), were differentially identified but with less than 5 peptide sequences and did thus not fulfill our stringent requirements. This final selection step evidenced 5 candidate proteins (see Supporting Information S3, proteins in bold boxes). Among those, HLA (P05534) was not selected due to its lack of biological relevance. Chloride channel CLIC1 (O00299), which serves in signal transduction and chloride transport, was not retained because of its nondifferential expression in SLL versus MCL (5 peptides in SLL versus 6 in MCL). As cytoskeleton remodelling partly induces MP generation, Transgelin-2 (P37802) and Plexin-A1 (Q9UIW2), two proteins involved in the latter process, were also not considered as first-line candidate biomarkers. Thus, CD148 (Q12913), a plasma membrane protein receptor with tyrosine phosphatase activity, was identified as the best “realistic” candidate biomarker that fulfills the stringent MS validation criteria mentioned above. Indeed, this protein was unambiguously identified in the MCL sample in 3 consecutive 1D gel bands with 10 different peptides (Figure 2 and Supporting Information S4), while in SLL only 4 unique peptides were observed (no CD148 tryptic peptide was detected in the CLL sample). The strong confidence in CD148 identification and its differential detection in MCL versus SLL/CLL suggest that CD148 is more expressed in MCL compared to SLL and CLL lymphoma samples.

**Validation of CD148 in 158 Patients and 30 Controls.** We decided to look at CD148 expression in controls and patients with circulating pathologic B-cells of CLL, SLL and MCL by flow cytometry (see Experimental Section). Representative experiments are shown in Figure 3. CD148 MFI of 30 control cells was very homogeneous (mean = 168, SD = 31, Table 1 and Figure 4). In a total of 93 CLL (mean = 189, SD = 78) and 24 SLL cases (mean = 209, SD = 92), MFI was considered not significantly different from controls using nonparametric tests (controls versus CLL,  $p = 0.59$ ; controls versus SLL,  $p = 0.07$ ; and CLL versus SLL  $p = 0.20$  Mann–Whitney Test). In these three groups (CLL, SLL and controls), it can thus be considered that CD148 is expressed at comparable levels. The results were strikingly different for the MCL group ( $n = 41$ ), in which CD148 MFI was 613 (SD = 329) which is significantly different from both CLL and SLL groups ( $p < 0.0001$ , Mann–Whitney Test), and from controls ( $p < 0.0001$ , Mann–Whitney Test) (Figure 4). These results clearly support the differential expression of CD148, as suggested from the proteomic study. CD148 expression appeared to be highly elevated in MCL while not different from controls in CLL or SLL. Moreover, the realization of a Receiver Operator Characteristic test (ROC curve) between MCL and CLL/SLL shows that with a CD148 MFI equal or superior to 2 times the mean value obtained with CLL/SLL, MCL diagnosis can be suspected with 91% of specificity (versus CLL and SLL) and a sensitivity of 78% (Figure 5 and Supporting Information S5). This observation can be a rationale for its use for MCL diagnosis.

## Discussion

In this work, we describe the selection of CD148 as a potential candidate biomarker for clinical use proposed from a proteomic approach, and a preliminary validation of its expression in different B-cell lymphoproliferations using classical immunological tools.



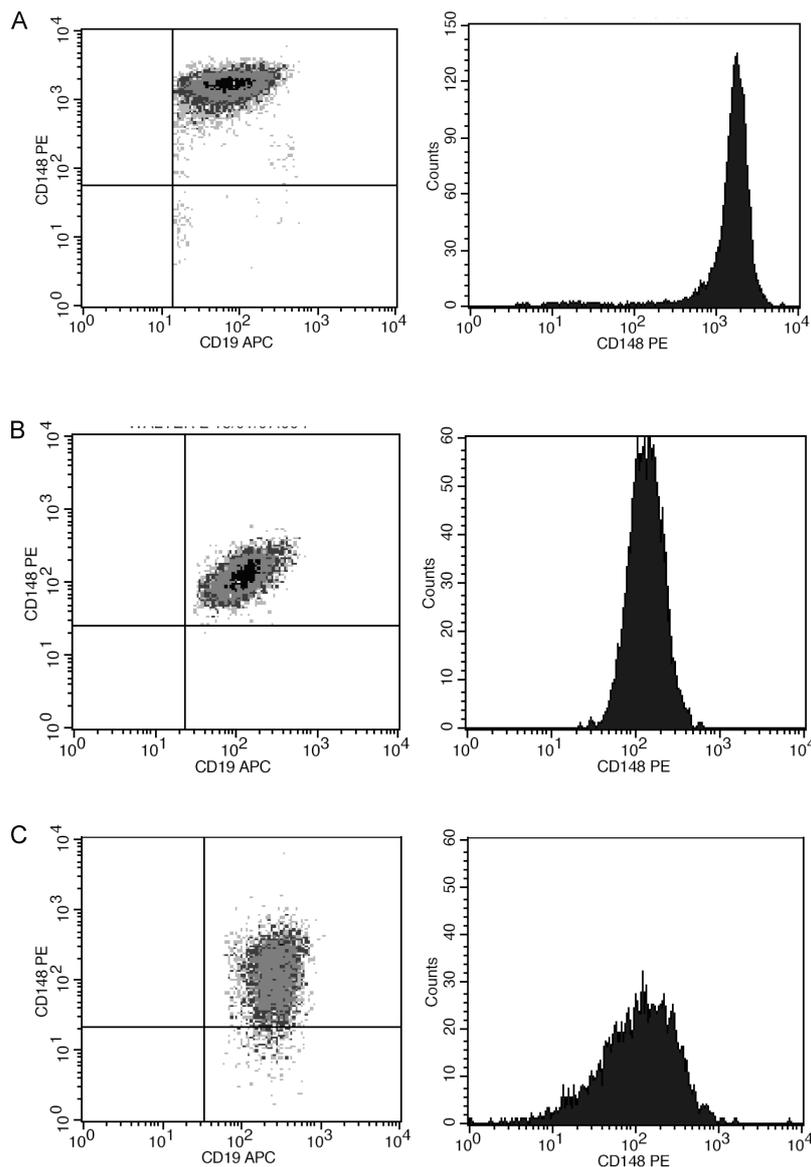
**Figure 2.** Identification of CD148 by mass spectrometry. Plasma membrane MPs are generated from pure cell populations. Proteins are extracted and fractionated using 1D SDS-PAGE electrophoresis. Gels bands are systematically cut into 2 mm pieces, followed by trypsin digestion and nanoLC-MS/MS analysis. In bold are shown the 10 different peptides sequenced by MS/MS during mass spectrometry analyses and used for CD148 identification. The fragmentation spectrum of the underlined peptide (measured monoisotopic mass of 1497.8 Da) is presented in the right panel of the figure. The fragmentation spectrum shows predominantly y and b ions (as indicated on the spectrum) allowing the determination of the DTEVLLVGLLEPGTR amino acid sequence, permitting a clear identification of the CD148.

In the race for disease-specific clinically useful biomarker discovery, the use of proteomics has huge potential, although, to date, there are only a limited number of successful exploitations of proteomic experiments.<sup>25</sup> Recently, Lescuyer et al. have addressed this point of heated debate in a review paper.<sup>26</sup> These authors explain that several reasons can account for the relative failure of proteomic approaches to deliver new potential biomarkers. These include: no optimal definition of the clinical question, complexity of the different biological fluids chosen, inadequate proteomic strategies developed and underestimation of the necessity to validate proteomic-proposed biomarkers. However, a careful design of the proteomic experiment in the discovery phase can lead to successful identification of biomarkers for clinical purposes. Another crucial point is the fact that the proteomic-based discovery phase needs always to be followed by a separate clinical validation phase involving quantitative methodologies, often by immunological methods (mostly ELISA assays). Altogether, these points can explain the relative low yield of successful proteomic studies in the biomarker discovery field.

The study presented here relies on a well-defined question: how can we discover helpful biomarkers in MBCN, and in our case MCL, that could be easily detected using routine analysis in hospital, that is, membrane antigens by flow cytometry? To achieve this goal it was decided to work on MPs, which were shown to be an interesting starting material, affording the

combination of low sample consumption and enabling the identification of hundreds of plasma surface membrane proteins.<sup>19,20</sup> At the proteomic level, the classical approach used for biomarker discovery is to compare protein/peptide profiles of normal versus diseased subjects.<sup>27,28</sup> This approach was not applicable in our case, as sufficient amounts of MPs cannot be easily generated from healthy subjects (about  $200 \times 10^6$  B-cells required). However, it can be applied to a wide range of lymphoma showing an accumulation of lymphocytes in the peripheral blood. Even if the limitation of our approach is to have enough circulating MPs in the concerned malignancy, other studies have already shown platelet MPs<sup>11</sup> and small vesicles<sup>29</sup> proteomic analyses in various biological fluids.<sup>30</sup> Moreover, which population would be explored: memory B-cells, naive B-cells? Also MBCN diagnosis issue is not to distinguish ill versus healthy patients but to precisely classify MBCN subtypes (CLL versus SLL versus MCL for instance). We therefore based our work on a differential proteomic study between CLL, SLL and MCL pathologies.

A proteomic experiment based on 1D-gel separation of proteins, nanoLC-MS/MS analyses of tryptic peptides combined with highly stringent MS validation criteria for protein identifications was subsequently designed. Although absolute quantification is not possible with this approach, comparison of MPs subproteomes originating from different pathologies indicate significant differences in protein expression levels.<sup>31,32</sup>



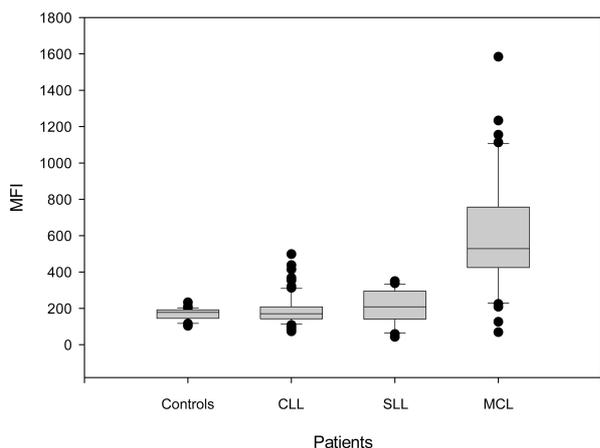
**Figure 3.** Anti-CD148 staining of three different representative samples: (A) mantle cell lymphoma (MCL), (B) chronic lymphocytic leukemia (CLL), and (C) control samples, showing a high CD148 expression in MCL samples. Left panel: CD19/CD148 dot blot (anti-CD19 allophycocyanin conjugated (CD19 APC) in X-axis, CD148 phycoerythrin conjugated (CD148 PE) in Y-axis). Right panel: Anti-CD148 phycoerythrin staining histogram.

**Table 1.** Anti-CD148 Staining Results and Cyclin D1 Expression of Chronic Lymphocytic Leukemias (CLL), Small Cell Lymphomas (SLL), Mantle Cell Lymphomas (MCL), and Control Groups

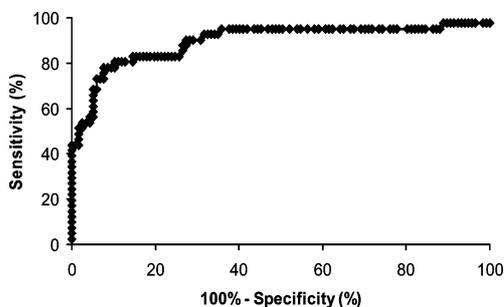
	controls	CLL	SLL	MCL
<i>n</i>	30	93	24	41
Cyclin D 1	Negative	Negative	Negative	Positive
CD148 MFI				
Mean	168	189	209	613
Min	104	80	74	70
Max	233	498	350	1584
SD	31	78	93	329

Cell surface CD148 quantification, which was achieved in our case by immunological methods, is thus necessary. The pro-

teomic selection process we developed, based on high numbers of peptides and high protein sequence coverages, allowed to focus not only on proteins differentially identified with a high confidence, but also proteins showing significant differences in their expression level.<sup>23,33</sup> Even if the choice of CD148 can appear to be arbitrary as there are no quantitative criteria in the selection process, our proteomic data show that CD148 is identified with 10 different peptides in MCL, 4 in SLL and none in CLL, which strongly suggests that CD148 is overexpressed in MCL compared to the two other lymphomas. CD148 seemed thus to be the best protein candidate to be proposed for clinical validation. The presented strategy shows the ability of proteomics to propose strong protein candidates to be further confirmed as potential biomarkers. The quantification, the



**Figure 4.** Comparison of CD148 median fluorescence intensity in controls ( $n = 30$ ), chronic lymphocytic leukemias ( $n = 93$ ), small cell lymphoma ( $n = 24$ ) and mantle cell lymphomas (MCL,  $n = 41$ ).



**Figure 5.** ROC Curve of MCL versus CLL/SLL group. A high CD148 (2-fold mean expression of normal B-cells) strongly indicates MCL diagnosis with 91% of specificity and a sensitivity of 78%.

biological and statistical relevance of CD148 as potential MCL biomarker, are then given by flow cytometry analyses on a larger set of patients.

CD148 is a plasma membrane receptor with phosphatase activity (type 3) also named DEP-1, PTPRJ or HPTPeta. This protein is composed of an extracellular domain containing 8 fibronectin type II-like domains, a transmembrane region and a single intracellular phosphatase domain. It was identified in human F-36 erythroleukemic<sup>34</sup> and HeLa cell lines, in which its expression was enhanced with increasing cell density.<sup>35</sup> It is expressed in various tissues, and particularly in hematopoietic cells. In peripheral blood cells, granulocytes display the highest intensity of anti-CD148 staining using flow cytometry, while lymphocytes are more weakly stained, CD19+ cells being slightly more stained than T cells.<sup>36</sup> The unique study published to date about CD148 expression in MBCN<sup>37</sup> described its expression in almost all the different types of the pathologies studied. On the contrary, our flow cytometry experiments revealed that CD148 is overexpressed in the large majority of MCL cases compared to CLL or SLL. Discrepancies between these results can be explained by several reasons: (i) Dong et al. use immunochemistry, a considerably less sensitive and nonquantitative technique compared to flow cytometry, and probably not helpful in the context of B-cell malignancies diagnosis described here; (ii) only flow cytometry techniques are able to discriminate variations of CD148 in the range of

what was detected in the present study. Thus systematic CD148 flow cytometry detection in MBCN could be very helpful in order to better diagnose MCL that harbor frequently misleading morphologic and immunophenotypic presentations. As a consequence, “gold standard” karyotype or cyclin D1 exploration, that are not systematically realized, may then be performed more rapidly when a high CD148 expression is detected by flow cytometry analysis. On the contrary, 30% of MCL patients display a CD148 below 380 counts. In these cases, CD148 does not help for MCL diagnosis. Other biomarkers are still needed and the current diagnosis relies on available biological tools. From our results, a high CD148 expression (superior to 2-fold the mean of normal B-cells) may be strongly in favor of MCL diagnosis with 91% of specificity (versus CLL and SLL) and a sensitivity of 78%. Interestingly, in the present study, for two cases where the diagnosis of MCL lymphoma was not evoked using conventional criteria (that do not include karyotype or cyclin D1), a high anti-CD148 staining led to CCND1 expression analysis. Both cases were positive and the diagnosis was reconsidered into MCL.

We have no clear explanation for the high level of expression of CD148 in MCL. *In vitro* activation of T cells induces an increase in CD148 expression, indicating a potential role in signal transduction on lymphocytes confirmed by other studies.<sup>36,38</sup> However, for the moment, no CD148 ligand has been clearly identified even if a potential ligand may be present in the extracellular matrix.<sup>39</sup> Replacement of CD148 phosphatase domain with GFP was not viable in mice due to vascular abnormalities, thus, preventing any analysis in mature B-lymphocytes.<sup>40</sup> Mice with constitutively deleted transmembrane domain of CD148 were also generated.<sup>38</sup> In this model, no obvious defect in B-cell development was observed in bone marrow, but a partial block in peripheral B-cell development was detected. In CD45 (a crucial plasma membrane protein with phosphatase activity expressed in hematopoietic cells) and CD148 double KO mice,<sup>38</sup> a severe block in early B-cell development and impaired B-cell antigen receptor (BCR) mediated signaling in relation with impaired tyrosine phosphorylation was observed. Very recently, it was also shown that CD148 may interact with and dephosphorylate p85 and modulate the magnitude of PI3K activity.<sup>41</sup> It is thus likely that CD148 participates in the regulation of signaling in B-cells.

In conclusion, the strategy described in this paper is one of the few successful proteomic studies that have led to the identification of a potentially useful biomarker. Indeed, CD148 may be a useful first-line biomarker for the diagnosis of MBCN using flow cytometry: (i) medium or high CD148 expression (defined by CD148 MFI > MFI of normal B-cells + 2 SD) may exclude the diagnosis for CLL or SLL samples; (ii) a high CD148 (2-fold mean expression of normal B-cells) may strongly indicate MCL diagnosis with 91% specificity and 78% sensitivity. As MCL assessment can still be ambiguous in several cases, an earlier, more precise and accurate diagnosis, using a combination of criteria (such as cyclinD1 activation and CD148 MFI, for example), would imply more aggressive chemo- or immuno-therapies. Further studies are now required to assess the clinical and prognosis strengths of this biomarker. In addition, the approach outlined here may be useful for the identification of potential biomarkers for other pathologies.

**Abbreviations:** MPs, microparticles; CLL, chronic lymphocytic leukemia; MCL, mantle cell lymphoma; SLL, small B-cell

lymphoma; CCND1, cyclin D1; NHL, non-Hodgkin lymphoma; MBCN, Mature B-cell neoplasms; MFI, mean fluorescence intensity; nanoLC-MS/MS, nanocapillary liquid chromatography-tandem mass spectrometry.

**Acknowledgment.** L. Miguet gratefully acknowledges the ARC foundation for financial support. G. Béchade was supported by the French Research Ministry and the European Postgraduate School (GRK 532). This study was also supported part by the CNRS, the Medical School of Strasbourg, and grants from the “Programme Hospitalier de Recherche Clinique des Hôpitaux de Strasbourg” # 2674, the ARECOH/ATGC association and the Institut National du Cancer. The FRM (Fondation pour la Recherche Médicale) is acknowledged for financing a mass spectrometer.

**Supporting Information Available:** S1, table of plasma membrane proteins identified per pathology with information on the total number of peptides involved in identification; S2, information relative to the identifications of the plasma membrane proteins; S3, table of proteins identified in MCL or in MCL and one of the two other pathologies and Biomarker candidate selection using highly stringent MS-validation criteria (more than 5 peptide sequences); S4, information on CD148 identification and MS/MS spectra of peptides involved in the identification; S5, raw data of ROC Curves. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Jaffe, E. S.; L., H. N.; Stein, H.; Vardiman, J. W. , *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*; IARC Press: Lyon, France, 2001; Vol. 1.
- Matutes, E.; Owusu-Ankomah, K.; Morilla, R.; Garcia Marco, J.; et al. The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia* **1994**, *8*, 1640–1645.
- Rawstron, A. C.; Villamor, N.; Ritgen, M.; Bottcher, S.; et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. *Leukemia* **2007**, *21*, 956–964.
- Montserrat, E. New prognostic markers in CLL. *Hematology* **2006**, 279–284.
- Ruchlemer, R.; Parry-Jones, N.; Brito-Babapulle, V.; Attolico, I.; et al. B-prolymphocytic leukaemia with t(11;14) revisited: a splenomegalic form of mantle cell lymphoma evolving with leukaemia. *Br. J. Haematol.* **2004**, *125*, 330–336.
- Nelson, B. P.; Variakojis, D.; Peterson, L. C. Leukemic phase of B-cell lymphomas mimicking chronic lymphocytic leukemia and variants at presentation. *Mod. Pathol.* **2002**, *15*, 1111–1120.
- Avet-Loiseau, H.; Garand, R.; Gaillard, F.; Daviet, A.; et al. Detection of t(11;14) using interphase molecular cytogenetics in mantle cell lymphoma and atypical chronic lymphocytic leukemia. *Genes Chromosomes Cancer* **1998**, *23*, 175–182.
- Troussard, X.; Mauvieux, L.; Radford-Weiss, I.; Rack, K.; et al. Genetic analysis of splenic lymphoma with villous lymphocytes: a Groupe Français d’Hématologie Cellulaire (GFHC) study. *Br. J. Haematol.* **1998**, *101*, 712–721.
- Resnitzky, P.; Matutes, E.; Hedges, M.; Morilla, R.; et al. The ultrastructure of mantle cell lymphoma and other B-cell disorders with translocation t(11;14)(q13;q32). *Br. J. Haematol.* **1996**, *94*, 352–361.
- Ek, S.; Hogerkerp, C. M.; Dictor, M.; Ehinger, M.; Borrebaeck, C. A. Mantle cell lymphomas express a distinct genetic signature affecting lymphocyte trafficking and growth regulation as compared with subpopulations of normal human B cells. *Cancer Res.* **2002**, *62*, 4398–4405.
- García, B. A.; Smalley, D. M.; Cho, H.; Shabanowitz, J.; et al. The platelet microparticle proteome. *J. Proteome Res.* **2005**, *4*, 1516–1521.
- Pasini, E. M.; Kirkegaard, M.; Mortensen, P.; Lutz, H. U.; et al. In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **2006**, *108*, 791–801.
- Jeong, H. C.; Lee, S. Y.; Lee, E. J.; Jung, K. H.; et al. Proteomic analysis of peripheral T-lymphocytes in patients with asthma. *Chest* **2007**, *132*, 489–496.
- Bartee, E.; McCormack, A.; Fruh, K. Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. *PLoS Pathog.* **2006**, *2*, e107.
- Salonen, J. M.; Valmu, L.; Ronnholm, G.; Kalkkinen, N.; Vihinen, M. Proteome analysis of B-cell maturation. *Proteomics* **2006**, *6*, 5152–5168.
- Barderas, M. G.; Darde, V. M.; Duran, M. C.; Egido, J.; Vivanco, F. Characterization of circulating human monocytes by proteomic analysis. *Methods Mol. Biol.* **2007**, *357*, 319–328.
- Santoni, V.; Kieffer, S.; Desclaux, D.; Masson, F.; Rabilloud, T. Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* **2000**, *21*, 3329–3344.
- Galeva, N.; Altermann, M. Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins. *Proteomics* **2002**, *2*, 713–722.
- Miguet, L.; Pacaud, K.; Felden, C.; Hugel, B.; et al. Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* **2006**, *6*, 153–171.
- Miguet, L.; Sanglier, S.; Schaeffer, C.; Potier, N.; et al. Microparticles: a new tool for plasma membrane sub-cellular proteomic. *Subcell Biochem.* **2007**, *43*, 21–34.
- Taylor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P. A.; et al. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **2007**, *25*, 887–893.
- De Vos, R. C.; Moco, S.; Lommen, A.; Keurentjes, J. J.; et al. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2007**, *2*, 778–791.
- Paoletti, A. C.; Parmely, T. J.; Tomomori-Sato, C.; Sato, S.; et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18928–18933.
- Zybailov, B.; Mosley, A. L.; Sardi, M. E.; Coleman, M. K.; et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5*, 2339–2347.
- Hsich, G.; Kenney, K.; Gibbs, C. J.; Lee, K. H.; Harrington, M. G. The 14-3-3 brain protein in cerebrospinal fluid as a marker for transmissible spongiform encephalopathies. *N. Engl. J. Med.* **1996**, *335*, 924–930.
- Lescuyer, P.; Hochstrasser, D.; Rabilloud, T. How shall we use the proteomics toolbox for biomarker discovery. *J. Proteome Res.* **2007**, *6*, 3371–3376.
- Miguet, L.; Bogumil, R.; Decloquement, P.; Herbrecht, R.; et al. Discovery and identification of potential biomarkers in a prospective study of chronic lymphoid malignancies using SELDI-TOF-MS. *J. Proteome Res.* **2006**, *5*, 2258–2269.
- Sanchez, J. C.; Lescuyer, P.; Hochstrasser, D.; Allard, L. Detection of biomarkers of stroke using SELDI-TOF. *Methods Mol. Biol.* **2007**, *357*, 343–350.
- Gonzalez-Begne, M.; Lu, B.; Han, X.; Hagen, F. K.; et al. Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). *J. Proteome Res.* **2009**, *8*, 1304–1314.
- Gonzales, P. A.; Pisitkun, T.; Hoffert, J. D.; Tchapyjnikov, D.; et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *J. Am. Soc. Nephrol.* **2009**, *20*, 363–379.
- Liu, H.; Sadygov, R. G.; Yates, J. R. 3rd A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193–4201.
- Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265–1272.
- de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008**, *455*, 1251–1254.
- Honda, H.; Inazawa, J.; Nishida, J.; Yazaki, Y.; Hirai, H. Molecular cloning, characterization, and chromosomal localization of a novel protein-tyrosine phosphatase, HPTP. *Blood* **1994**, *84*, 4186–4194.

## research articles

- (35) Ostman, A.; Yang, Q.; Tonks, N. K. Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced with increasing cell density. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9680–9684.
- (36) de la Fuente-Garcia, M. A.; Nicolas, J. M.; Freed, J. H.; Palou, E.; et al. CD148 is a membrane protein tyrosine phosphatase present in all hematopoietic lineages and is involved in signal transduction on lymphocytes. *Blood* **1998**, *91*, 2800–2809.
- (37) Dong, H. Y.; Shahsafaei, A.; Dorfman, D. M. CD148 and CD27 are expressed in B cell lymphomas derived from both memory and naive B cells. *Leuk. Lymphoma* **2002**, *43*, 1855–1858.
- (38) Zhu, J. W.; Brdicka, T.; Katsumoto, T. R.; Lin, J.; Weiss, A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signaling. *Immunity* **2008**, *28*, 183–196.
- (39) Sorby, M.; Sandstrom, J.; Ostman, A. An extracellular ligand increases the specific activity of the receptor-like protein tyrosine phosphatase DEP-1. *Oncogene* **2001**, *20*, 5219–5224.
- (40) Takahashi, T.; Takahashi, K.; St John, P. L.; Fleming, P. A.; et al. A mutant receptor tyrosine phosphatase, CD148, causes defects in vascular development. *Mol. Cell. Biol.* **2003**, *23*, 1817–1831.
- (41) Tsuboi, N.; Utsunomiya, T.; Roberts, R. L.; Ito, H.; et al. The tyrosine phosphatase CD148 interacts with p85 regulatory subunit of PI 3-kinase. *Biochem. J.* **2008**, *413*, 193–200.

Miguet et al.

PR801102C



## 3. Perspectives : Recherche de biomarqueurs potentiels dans le cas de la leucémie de Waldenström

### 3.1. La leucémie de Waldenström

La maladie de Waldenström (Waldenström Macroglobulinemia, WM) est une pathologie lymphoproliférative touchant les lymphocytes B. C'est une leucémie rare affectant généralement les personnes de plus de 60 ans. Elle représente 1 à 2 % des cas de leucémies. Environ 25 % des patients ne présentent pas de symptôme et sont diagnostiqués par hasard. Il n'existe aucun traitement curatif et la survie moyenne ne dépasse pas 5 ans.

Des variations phénotypiques existent, mais la plupart des patients présentent le nouveau profil caractéristique proposé : IgM<sup>+</sup>, CD5<sup>-</sup>, CD10<sup>-</sup>, CD19<sup>+</sup>, CD20<sup>+</sup> et CD23<sup>-</sup>. Ici encore, un besoin de nouveaux biomarqueurs permettant un diagnostic plus précis et suggérant de probables cibles thérapeutiques se fait sentir.

Forts du succès de l'identification et du début de validation d'une protéine membranaire spécifique pour le lymphome du manteau, nous avons poursuivi les efforts de développement et d'application de notre stratégie d'analyse protéomique différentielle par spectrométrie de masse à destination de nouveaux résultats pour la leucémie de Waldenström.

### 3.2. Application de la stratégie analytique

Cette étude a pu bénéficier de toutes nos innovations stratégiques et techniques développées pour l'analyse protéomique différentielle par spectrométrie de masse (revoir la figure 6 au point 2.1.2) :

- Les analyses ont été réalisées sur le système HPLC Chip (Agilent Tech.) – HCT Ultra (Bruker Dal.). L'instrument, qui était alors le plus récemment acquis par le laboratoire, dispose d'une sensibilité et d'une vitesse d'acquisition accrues.
- L'utilisation de puces microfluidiques équipées d'une colonne de 150 mm et de méthanol en lieu et place de l'acétonitrile, a permis d'allier performances résolutives, sensibilité et débit d'analyse.
- Une optimisation fine des gradients et des paramètres de cycles d'acquisition MS+MS/MS a été réalisée.
- Les analyses ont été réalisées en duplicata afin d'augmenter la part du contenu protéique total identifié et d'améliorer la qualité des identifications.
- Les paramètres de génération des fichiers de résultats « *peak-list* » ont été optimisés pour exporter un maximum de données et des informations de qualité.
- Les fichiers « *peak-list* » de chaque paire d'analyses ont été concaténés pour cumuler l'information complémentaire recueillie.

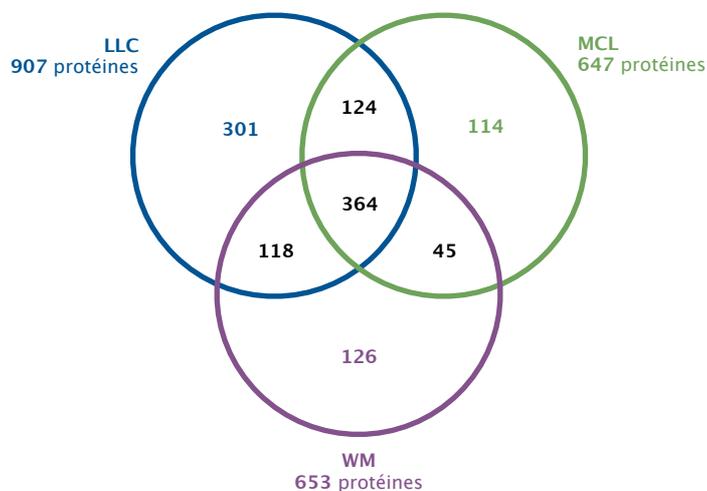
Nous avons pu disposer de 3 échantillons incluant des patients atteints de pathologies qui présentent de fortes similitudes phénotypiques avec la leucémie de Waldenström (WM) : une leucémie lymphoïde chronique (LLC) à l'évolution incertaine et un lymphome du manteau (MCL).

### 3.3. Résultats préliminaires

▸ Sensibilité de notre stratégie optimisée

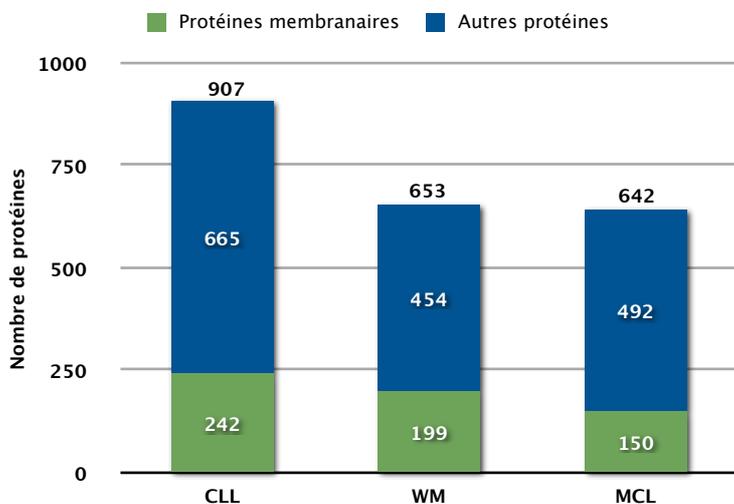
En cumulant les résultats pour les 3 pathologies, 1192 protéines non redondantes ont pu être identifiées. Entre 650 et 900 protéines ont été analysées dans chaque cas (Figure 12).

Ces premiers résultats bruts montrent la sensibilité accrue du couplage HPLC–Chip (Agilent Tech.) – HCT Ultra (Bruker Dal.) et le bénéfice apporté par une optimisation fine de nos systèmes. Associées à la stratégie de validation des identifications que nous avons mis au point, ces améliorations ont permis d’identifier 2 fois plus de protéines que ce qui avait été présenté dans l’application précédente. En effet, pour des échantillons similaires (microparticules de lymphocytes B), l’analyse nanoLC–MS/MS des pathologies LLC, MCL et SLL sur le couplage CapLC – Q–TOF II (Waters Corp.) avait conduit à l’identification de 579 protéines.



**Figure 12** | Nombre de protéines identifiées pour les échantillons de leucémies lymphoïdes chroniques (LLC), de lymphome du manteau (MCL) et de leucémie de Waldenström (WM) grâce à notre stratégie analytique optimisée.

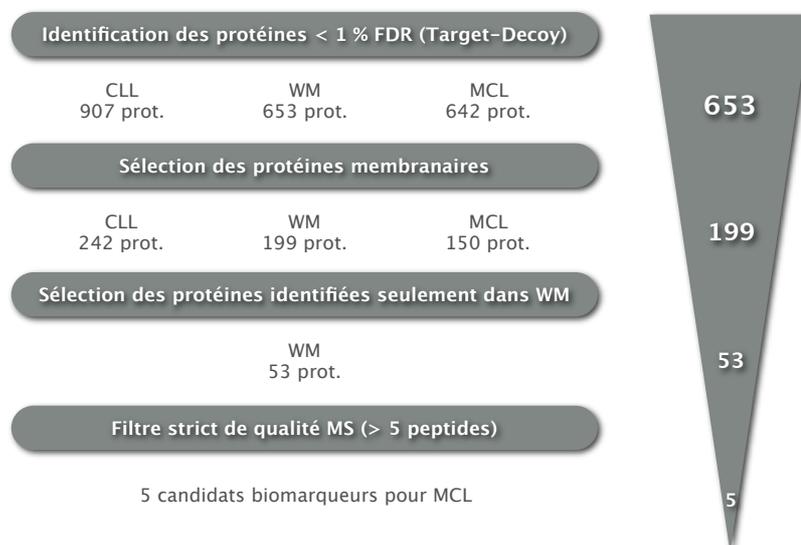
Par ailleurs, la technique de préparation de microparticules s’est révélée une nouvelle fois très performante pour l’enrichissement des échantillons en protéines membranaires : elles représentent en moyenne 36 % des protéines identifiées dans nos échantillons (Figure 13).



**Figure 13** | Nombre de protéines identifiées et fraction de protéines membranaires dans les trois échantillons de microparticules analysés issus des pathologies CLL, WM et MCL étudiées.

▸ Une nouvelle liste de candidats potentiels

A partir de la liste des protéines identifiées, une série de filtres similaires à ceux utilisés dans l'étude précédente a permis de sélectionner quelques candidats biomarqueurs pertinents pour la pathologie de Waldenström (Figure 14).



**Figure 14** | Filtres de pertinence biologique et de qualité MS conduisant à la proposition d'un marqueur membranaire potentiel spécifique du lymphome du manteau.

Parmi 199 protéines membranaires identifiées pour la pathologie de Waldenström, seules 54 lui étaient spécifiques. La table 4 présente les 5 protéines identifiées avec plus de 5 peptides : elle constitue la liste des protéines les plus pertinentes à considérer.

Protéine	Numéro d'accession	Masse moléculaire (Da)	Nombre de peptides identifiés
Hexokinase-3 – Homo sapiens (Human)	P52790 HXK3_HUMAN	98 899,8	17
HLA class I histocompatibility antigen, A-25 alpha chain precursor – Homo sapiens (Human)	P18462 1A25_HUMAN	41 063,4	12
ADP-ribosyl cyclase 2 precursor – Homo sapiens (Human)	Q10588 BST1_HUMAN	35 707,3	5
B-cell receptor CD22 precursor – Homo sapiens (Human)	P20273 CD22_HUMAN	95 329,2	5
HLA class I histocompatibility antigen, B-41 alpha chain precursor – Homo sapiens (Human)	P30479 1B41_HUMAN	40 520,7	5

**Table 4** | Liste des 5 protéines passant chacun des filtres successifs pour l'identification de candidats biomarqueurs pour WM.

Les protéines de la classe HLA ne sont pas des candidats à retenir, car leur expression différentielle est intrinsèque à leur rôle de marqueur du soi. La pertinence des autres protéines candidates est en cours de discussion avec nos collaborateurs. Des analyses complémentaires pourront être nécessaires pour identifier de nouveaux candidats ou des filtres de sélection plus souples pourront faire apparaître de nouvelles cibles potentielles.

## 4. Conclusion

Afin d'améliorer le diagnostic délicat d'hémopathies lymphoïdes ambiguës et éventuellement de proposer de nouvelles cibles thérapeutiques, nous sommes partis à la recherche de marqueurs membranaires spécifiques.

Nous sommes parvenus à identifier le cluster de différenciation CD148 comme biomarqueur potentiel du lymphome du manteau et à vérifier cliniquement sa surexpression pour cette pathologie.

L'objectif était osé : trouver une aiguille dans une botte de foin.

▸ D'abord parce qu'il est nécessaire de mettre au point un protocole de préparation d'échantillons modèles reflétant l'état *in vivo* tout en étant enrichis en protéines membranaires, plus difficiles à analyser.

Le développement des microparticules de membrane plasmique nous a permis de disposer d'un échantillon simple à préparer et présentant une part importante de protéines membranaires.

▸ Ensuite parce que notre stratégie d'analyse différentielle du contenu protéique d'échantillons par spectrométrie de masse pour la recherche de biomarqueurs présente des limites intrinsèques à première vue difficilement surmontables. Sans approche de quantification des protéines identifiées, il n'est envisageable que d'évaluer la présence ou l'absence probable de protéines dans des échantillons issus de patients uniques.

Nous avons finement optimisé cette phase de « découverte » afin de rendre notre méthode plus sensible, plus reproductible et plus fiable. Les innovations stratégiques et techniques décrites au chapitre précédent ont été appliquées ici, démontrant le bénéfice qu'elles apportent. En identifiant une plus grande fraction du contenu protéique total de chaque échantillon (et donc les constituants présents en plus faible concentration), notre approche permet de diminuer le nombre de candidats faux-positifs proposés et de focaliser le choix des cibles à valider cliniquement sur les protéines les plus pertinentes.

▸ Enfin parce que jusqu'à présent, la recherche de biomarqueurs par analyse protéomique n'a pas connu de succès retentissant. Cela est essentiellement dû au manque de passerelles et aux différences culturelles entre les mondes de la « découverte » et de la « validation » clinique des biomarqueurs candidats.

Notre collaboration rapprochée avec l'équipe de médecins a favorisé l'échange des attentes de part et d'autre et a conduit au succès de l'identification d'un biomarqueur spécifique du lymphome du manteau et à la vérification clinique de sa pertinence. L'effort de validation doit maintenant être poursuivi.

Les premiers résultats sur l'étude de la maladie de Waldenström sont encourageants. Ils ont permis de confirmer l'intérêt et les performances de notre stratégie analytique qui, dans le cadre de collaborations efficaces, pourra assurer de nouveaux succès mesurés et ciblés.

On retiendra que dans la course aux biomarqueurs, la préparation des objectifs et l'optimisation de stratégies adaptées prévalent sur des analyses inconsidérées et tous azimuts.

## Chapitre III

# Etude des partenaires d'interaction du facteur de transcription suppresseur de tumeur Trim24

---

Ce projet s'inscrit dans le cadre d'une collaboration avec l'équipe du Docteur Régine Losson, du Département de Génomique Fonctionnelle de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) d'Illkirch. Khalid Ouararhni a poursuivi le travail de thèse de Konstantin Khetchoumian : il a été établi que l'inactivation du cofacteur Trim24, qui est connu pour jouer un rôle crucial dans le contrôle de l'activité transcriptionnelle des récepteurs nucléaires, conduit à une activation aberrante de la voie de signalisation de l'acide rétinoïque dans les hépatocytes, provoquant le développement d'hépatocarcinomes<sup>1</sup>.

A la suite d'une brève introduction sur les objets biologiques étudiés, ce chapitre détaille les stratégies d'analyse protéomique développées pour l'étude des partenaires d'interaction de Trim24 par spectrométrie de masse. Il présente par la suite les résultats qui ont permis d'apporter des connaissances déterminantes pour la compréhension des mécanismes par lesquels Trim24 régule l'expression de ses gènes cibles.

---

<sup>1</sup> Khetchoumian, K. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2005).

# 1. Les mécanismes de régulation de la transcription

Chez les mammifères, la différenciation cellulaire, à partir du même patrimoine génétique, est assurée par une stricte régulation des gènes permettant l'adaptation de chaque cellule individuelle aux conditions externes. Lorsque ces processus de contrôle sont perturbés, des pathologies à l'origine de certains cancers peuvent apparaître<sup>2</sup>.

Une partie majeure de cette régulation s'exerce au niveau de l'initiation de la transcription, première étape – souvent limitante – de l'expression des gènes (Figure 1). C'est pourquoi la compréhension des fondements moléculaires de ces phénomènes et leurs impacts biologiques est devenue une des priorités de la recherche dans les sciences du vivant.

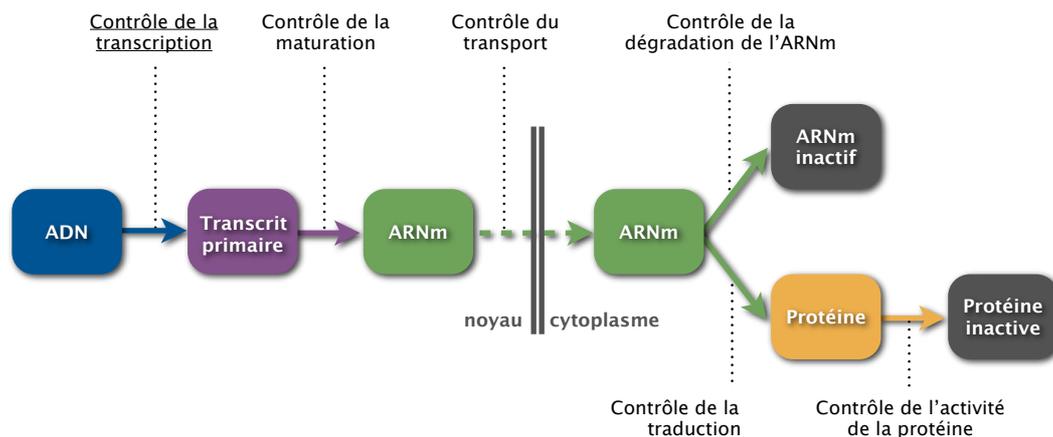


Figure 1 | Les 6 niveaux de contrôle de l'expression d'un gène eucaryote. Parmi eux, le contrôle de la transcription est prééminent.

## 1.1. La transcription de l'ADN

La transcription de l'ADN consiste en la copie des séquences codantes de l'ADN en ARNm. Trois catégories de facteurs participent à ce processus (Figure 2) :

- ▶ La machinerie transcriptionnelle de base

Constituée d'une enzyme, l'ARN polymérase, et de facteurs généraux de transcription, cette machinerie agit en aval des voies de signalisation qui aboutissent au démarrage de la transcription.

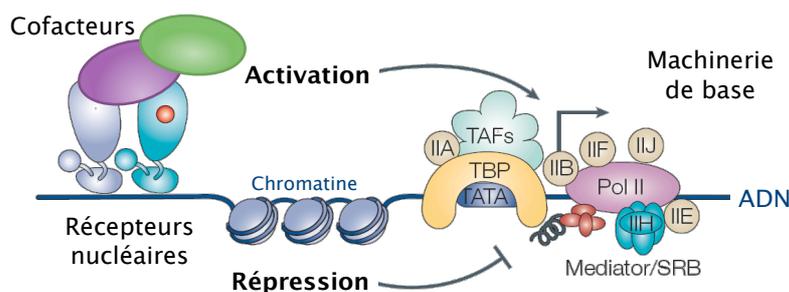
- ▶ Les facteurs de transcription séquence-spécifiques

Ces facteurs ciblent les gènes dont l'expression est à induire ou réprimer en reconnaissant des séquences d'ADN spécifiques situées au niveau des promoteurs des gènes ou à une distance variable du site d'initiation. Cette étude s'intéresse aux récepteurs nucléaires, qui forment une famille de facteurs de transcription séquence-spécifiques et ligand-dépendants.

- ▶ Les facteurs intermédiaires de transcription

Aussi appelés cofacteurs ou médiateurs, ils sont capables de réguler l'accès de la machinerie transcriptionnelle de base. Ils sont recrutés par les récepteurs nucléaires, sur l'effet régulateur desquels ils peuvent avoir un effet de stimulation ou d'inhibition. On parle alors de coactivateurs ou de corépresseurs.

<sup>2</sup> Santos-Rosa, H., et al. (2005). Eur J Cancer **41**, 2381-402.



**Figure 2** | Les 3 catégories de facteurs participant au contrôle de la transcription. La régulation de la transcription par les récepteurs nucléaires requiert un grand nombre de complexes de corégulation qui ont des fonctions diverses et présentent une activité enzymatique. (D'après Perissi et Rosenfeld, 2005)

## 1.2. Les récepteurs nucléaires et le mode d'action des cofacteurs

### 1.2.1. Les récepteurs nucléaires

#### ▸ Les deux voies de signalisation cellulaire

Au cours du développement chez l'adulte, les cellules sont soumises en permanence à une combinaison de signaux extracellulaires (hormones, neurotransmetteurs, facteurs de croissance et de différenciation). Ces signaux agissent par l'intermédiaire de 2 grandes voies de transmission :

▸ La voie de signalisation passant par les récepteurs membranaires. Ils répondent aux signaux en déclenchant une cascade d'évènements intracellulaires pour les transmettre jusqu'au noyau. Là, l'activité de nombreux facteurs de transcription et donc l'expression de certains gènes pourront être modifiées.

▸ La voie de signalisation utilisant les récepteurs nucléaires. Situés dans le noyau de la cellule, ils agissent directement en tant que facteurs de transcription en réponse à des molécules capables de diffuser librement à travers les membranes plasmique et nucléaire des cellules. Parmi ces composés liposolubles, appelés ligands, figurent par exemple les hormones stéroïdiennes, thyroïdiennes, la vitamine D ou les rétinoïdes (comme l'acide rétinoïque), dérivés actifs de la vitamine A.

#### ▸ La régulation de la transcription par les récepteurs nucléaires

Les récepteurs nucléaires sont capables de reconnaître et de se fixer à des séquences spécifiques d'ADN – dites « éléments de réponse » – sous forme de monomères, dimères ou hétérodimères. Alors qu'ils sont similaires dans leur structure, les récepteurs nucléaires ont développé des différences dans leur façon de réguler la transcription<sup>3</sup> :

▸ Les gènes cibles peuvent être activés en présence du ligand. Dans certains cas, le ligand permet le transport du récepteur nucléaire jusqu'à la séquence spécifique où il se fixe. Dans d'autres, les récepteurs nucléaires non liés au ligand se fixent sur leurs éléments de réponse et en association avec des complexes corépresseurs répriment le niveau d'expression basal des gènes cibles. La fixation du ligand va entraîner la dissociation des complexes corépresseurs et permettre le recrutement des complexes coactivateurs.

▸ Les gènes cibles peuvent être réprimés en présence du ligand. Des « éléments de réponse négatifs » peuvent être responsables d'une répression ligand-dépendante des gènes placés sous leur contrôle.

▸ Certains récepteurs nucléaires peuvent moduler l'expression des gènes indépendamment de leur capacité à fixer l'ADN selon des interférences positives ou négatives avec d'autres facteurs de transcription.

<sup>3</sup> Perissi, V., et al. (2005). Nat Rev Mol Cell Biol 6, 542-54.

### 1.2.2. Les cofacteurs des récepteurs nucléaires

#### ▸ Coactivateurs et corépresseurs

Par définition, les cofacteurs sont des protéines incapables de se fixer à l'ADN, qui sont recrutées par des facteurs de transcription séquence-spécifiques tels que les récepteurs nucléaires dont ils contrôlent les fonctions d'activation et de répression. Selon les cas, la présence ou l'absence de ligand est nécessaire. Le plus souvent, ces facteurs fonctionnent eux même en association avec d'autres partenaires au sein de complexes multiprotéiques, leur permettant d'activer ou d'inhiber la transcription selon qu'il s'agisse de coactivateurs ou de corépresseurs.

#### ▸ Le mode d'action des cofacteurs

L'ADN est engagé dans une structure nucléoprotéique compacte, appelée chromatine et formée d'une succession de nucléosomes : des octamères d'histones autour duquel s'enroulent 146 paires de bases d'ADN. Ces nucléosomes permettent l'empaquetage de l'ADN et limitent également l'accès de la machinerie de transcription au niveau des sites d'initiation. L'ADN chromatinien ne peut donc pas être transcrit en absence de facteurs de transcription<sup>4</sup>.

Les complexes de cofacteurs sont en mesure d'agir :

- soit par des interactions directes au niveau de la machinerie transcriptionnelle de base ;
- soit par des modifications de la structure de la chromatine qui augmentent ou diminuent l'accessibilité des sites d'initiation.

## 1.3. Dysfonctionnements dans les processus de régulation de la transcription

Les récepteurs nucléaires jouent un rôle central dans la prolifération et la différenciation cellulaires. Il n'est donc pas surprenant que les dysfonctionnements des voies de signalisation des récepteurs nucléaires soient à l'origine de nombreuses pathologies, telles que des maladies métaboliques ou des cancers. Ils représentent des cibles thérapeutiques de choix pour ces pathologies. Mais considérant la grande variété des effets médiés par les récepteurs nucléaires, il est important de tenir compte du contexte cellulaire ou tissulaire qui peut influencer de façon considérable la réponse aux traitements.

Par ailleurs, des erreurs dans le programme de l'expression des gènes peuvent favoriser le développement de cancers. Régulateur et corégulateurs transcriptionnels sont par conséquent souvent affectés dans des tumeurs. Les modèles actuels postulent que le cancer se développe par une série d'altérations génétiques (mutations) et épigénétiques. Des modèles de souris génétiquement modifiées ont apporté les preuves définitives de l'implication des cofacteurs dans le processus de cancérogenèse *in vivo*. Ces données récentes ouvrent de nouvelles voies pour des thérapies anticancéreuses, en définissant les cofacteurs comme une nouvelle catégorie de cibles potentielles pour des agents thérapeutiques.

---

<sup>4</sup> Santos-Rosa, H., et al. (2005). Eur J Cancer **41**, 2381-402.

## 2. Le rôle du facteur de transcription Trim24 dans l'hépatocarcinogénèse

Le carcinome hépatocellulaire (cancer du foie) est relativement répandu au niveau mondial et peut être induit par de multiples stimuli. Nos collaborateurs ont montré que chez la souris, le facteur de transcription « Tripartate motif-containing 24 (Trim24) » semblait réprimer spécifiquement le développement du cancer via une régulation de la voie de signalisation de l'acide rétinoïque. En ce sens, Trim24 pourrait représenter un suppresseur de tumeur, sa suppression ou sa sous-régulation conduisant au développement de la maladie.

### 2.1. Hépatocarcinogénèse et acide rétinoïque

#### 2.1.1. Le carcinome hépatocellulaire

Le carcinome hépatocellulaire (CHC) représente 80 à 90 % des cancers primitifs du foie. Il se situe au 8<sup>e</sup> rang, en terme d'incidence, chez l'homme et c'est le 3<sup>e</sup> cancer le plus meurtrier. En effet, les hépatocarcinomes sont le plus souvent diagnostiqués à des stades avancés, d'où une mortalité de 95 % dans les 5 ans qui suivent le diagnostic<sup>5</sup>. Bien que son étiologie soit assez bien connue (Table 1), les mécanismes moléculaires en cause restent encore mal caractérisés et rares sont les gènes suppresseurs de tumeurs ou protooncogènes connus pour être couramment impliqués dans le développement d'un CHC.

Facteurs à risque	Conséquences
<b>Facteurs majeurs</b>	
Virus de l'hépatite B (HBV)	Virus à ADN induisant l'hépatite chronique et la cirrhose
Virus de l'hépatite C (HCV)	Virus à ARN induisant l'hépatite chronique et la cirrhose
Aflatoxines	Agent génotoxique et cytotoxique
Alcool	Induit la cirrhose alcoolique
<b>Facteurs mineurs</b>	
Tabac	Cause des dommages de l'ADN et induit la prolifération
Eau polluée	Contient des toxines des algues qui stimulent la prolifération
Contraceptiques oraux	Stimulent la prolifération
Stéroïdes androgènes anabolisants	Stimulent la prolifération

**Table 1** | Les principaux facteurs favorisant le développement du carcinome hépatocellulaire chez l'homme.

#### 2.1.2. L'acide rétinoïque et l'hépatocarcinogénèse

L'acide rétinoïque (RA) est un dérivé de la vitamine A jouant un rôle important dans le développement embryonnaire, le contrôle de la prolifération et de la différenciation cellulaire. Il agit par l'intermédiaire de récepteurs nucléaires spécifiques : les récepteurs de l'acide rétinoïque RAR $\alpha$ ,  $\beta$ ,  $\gamma$ .

Bien que largement connu pour son activité antitumorale dans une grande variété de modèles cellulaires et animaux<sup>6</sup>, l'acide rétinoïque est également capable d'exercer des effets potentiellement oncogéniques. En particulier, il a été décrit que l'acide rétinoïque est un puissant mitogène pour les

<sup>5</sup> Bialecki, E. S., et al. (2005). Eur J Gastroenterol Hepatol 17, 485-9.

<sup>6</sup> Altucci, L., et al. (2001). Nat Rev Cancer 1, 181-93.

hépatocytes normaux des rongeurs. De plus, une étude<sup>7</sup> a montré, dans un modèle murin d'hépatocarcinogenèse induit chimiquement, que l'administration d'acide rétinoïque augmente de façon significative le développement des tumeurs. Cependant, le rôle joué par l'acide rétinoïque dans l'oncogenèse hépatique, tant au niveau cellulaire que moléculaire, reste incompris.

## 2.2. Mise en évidence de Trim24 comme répresseur de tumeur dans le foie

### 2.2.1. Le facteur intermédiaire de transcription Trim24

Trim24 est l'un des premiers cofacteurs de transcription identifiés<sup>8</sup> : il a été montré que la protéine interagissait directement avec la famille des récepteurs nucléaires, notamment des récepteurs de l'acide rétinoïque RAR. Ces interactions dépendent entièrement de la présence du ligand et se traduisent par une inhibition de l'activité transcriptionnelle des récepteurs nucléaires.

### 2.2.2. Mise en évidence de Trim24 comme suppresseur de tumeur dans le foie

Les hépatocarcinomes se développent chez l'homme comme chez la souris en présentant des caractéristiques morphologiques similaires même si les facteurs à risques peuvent différer. Dans le but d'élucider les fonctions physiologiques de Trim24, l'équipe de R. Losson a généré par recombinaison homologue une lignée de souris inactivées (KO) pour Trim24 (souris Trim24<sup>-/-</sup>)<sup>9</sup>. Ces souris Trim24<sup>-/-</sup> sont viables, mais développent des tumeurs du foie avec une pénétrance quasi complète à partir du 10<sup>e</sup> mois. A travers une étude histologique comparative du foie d'une large cohorte (> 100), l'action inhibitrice de Trim24 dans l'oncogénèse hépatique chez la souris a pu être démontrée (Figures 3 et 4a).

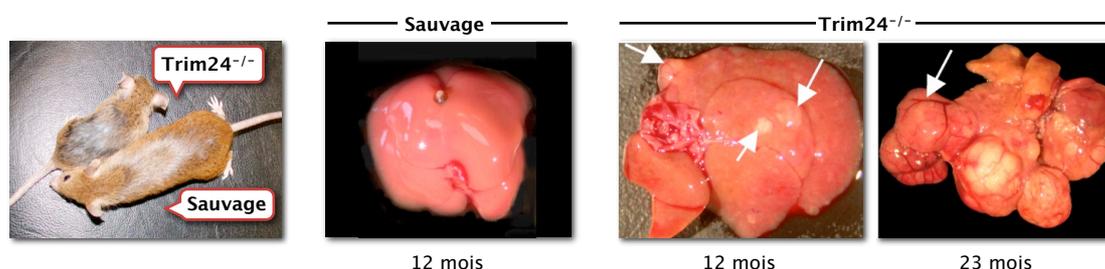


Figure 3 | Formation de tumeurs hépatiques spontanées chez les souris Trim24<sup>-/-</sup>.

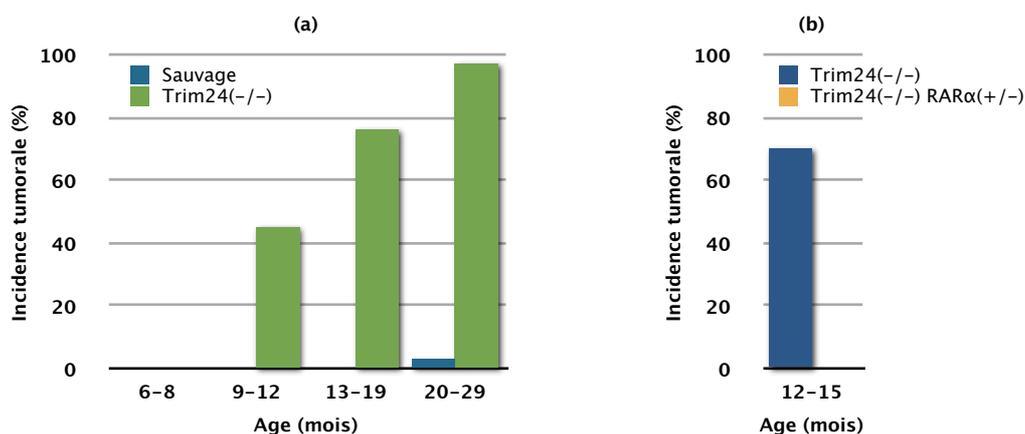
### 2.2.3. Le phénotype tumoral des hépatocytes Trim24<sup>-/-</sup> est sous le contrôle de RAR $\alpha$

Il a enfin été montré que chez les souris Trim24<sup>-/-</sup>, une invalidation d'un allèle du gène RAR $\alpha$  supprime entièrement la transformation tumorale des hépatocytes (Figure 4b). Cette étude montre que RAR $\alpha$  est un facteur indispensable pour le développement de tumeurs en absence de Trim24 et révèle les caractéristiques oncogéniques du récepteur RAR $\alpha$ .

<sup>7</sup> Di Bisceglie, A. M., et al. (2005). J Lab Clin Med 145, 134–8.

<sup>8</sup> Le Douarin, B., et al. (1995). EMBO J 14, 2020–33.

<sup>9</sup> Khetchoumian, K., et al. (2007). Nat Genet 39, 1500–6.

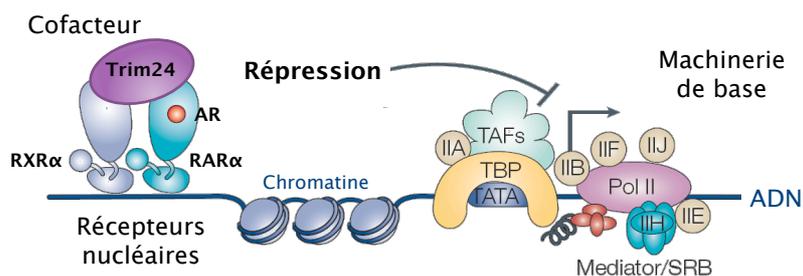


**Figure 4** | (a) L'inactivation de Trim24 implique une bien plus grande susceptibilité au développement de tumeurs dans le foie. (b) L'haploinsuffisance de RAR $\alpha$ <sup>+/-</sup>, dans le fond génétique Trim24<sup>-/-</sup>, supprime entièrement la transformation tumorale des hépatocytes.

## 2.2.4. Conclusions

En résumé, les études préliminaires de nos collaborateurs ont montré que (Figure 5) :

- Trim24 représente un gène suppresseur de tumeur spécifique dans le foie, puisque les souris inactivées pour Trim24 sont des modèles d'hépatocarcinogenèse spontanée.
- Le cofacteur Trim24 et le récepteur nucléaire RAR $\alpha$  sont en interaction fonctionnelle et Trim24 atténue l'expression des gènes oncogéniques activés par RAR $\alpha$ .



**Figure 5** | Trim24 inhibe l'activité transcriptionnelle des récepteurs nucléaires RAR $\alpha$  en présence du ligand acide rétinoïque.

### 3. Identification des mécanismes moléculaires par lesquels Trim24 régule la transcription des gènes oncogéniques dans le foie

Nous avons mis au point avec nos collaborateurs une stratégie d'analyse protéomique fonctionnelle par spectrométrie de masse, visant à l'étude des partenaires d'interaction du cofacteur Trim24. La suite de ce chapitre détaille le développement expérimental, met l'accent sur les résultats obtenus et leur validation.

#### 3.1. Mise au point d'une stratégie de purification et d'identification des partenaires spécifiques de Trim24

##### 3.1.1. Stratégie générale mise en oeuvre

L'étape de définition de la stratégie protéomique est une phase importante au cours de laquelle les questions biologiques, les contraintes et objectifs analytiques doivent s'accorder. La discussion avec nos collaborateurs a permis de mettre au point une approche complète allant de la préparation des échantillons permettant la purification des complexes multiprotéiques associés à Trim24 à l'analyse par spectrométrie de masse nanoLC-MS/MS (Figure 6). Beaucoup des optimisations réalisées ont été décrites au chapitre I de cette partie. Sont résumés ici, les principaux choix expérimentaux et les conditions de leur réalisation.

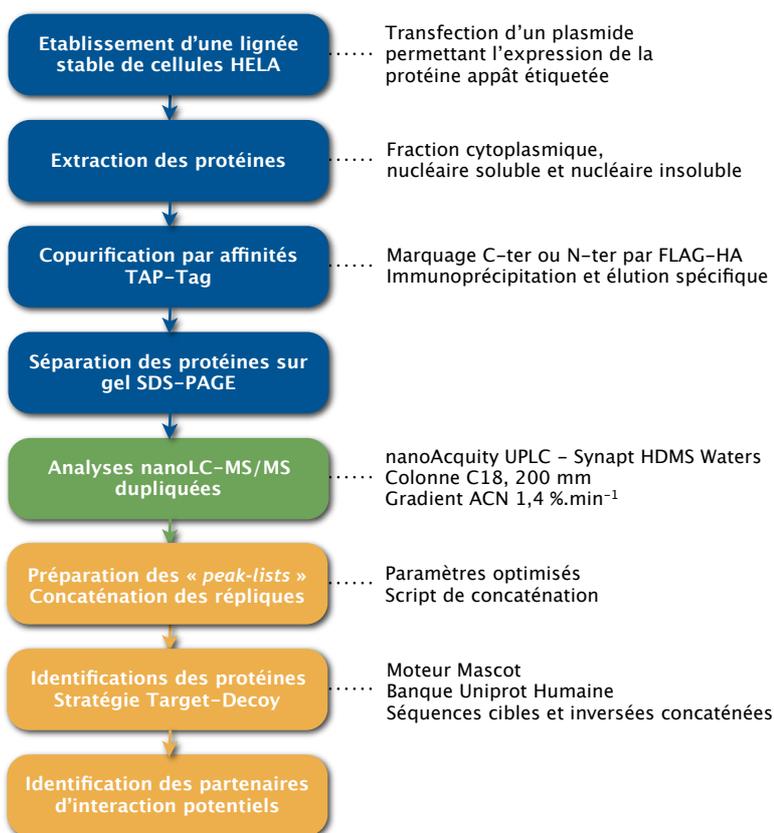


Figure 6 | Stratégie analytique optimale permettant l'analyse des complexes multiprotéiques associés au cofacteur Trim24.

### 3.1.2. Purification des complexes multiprotéiques associés à Trim24

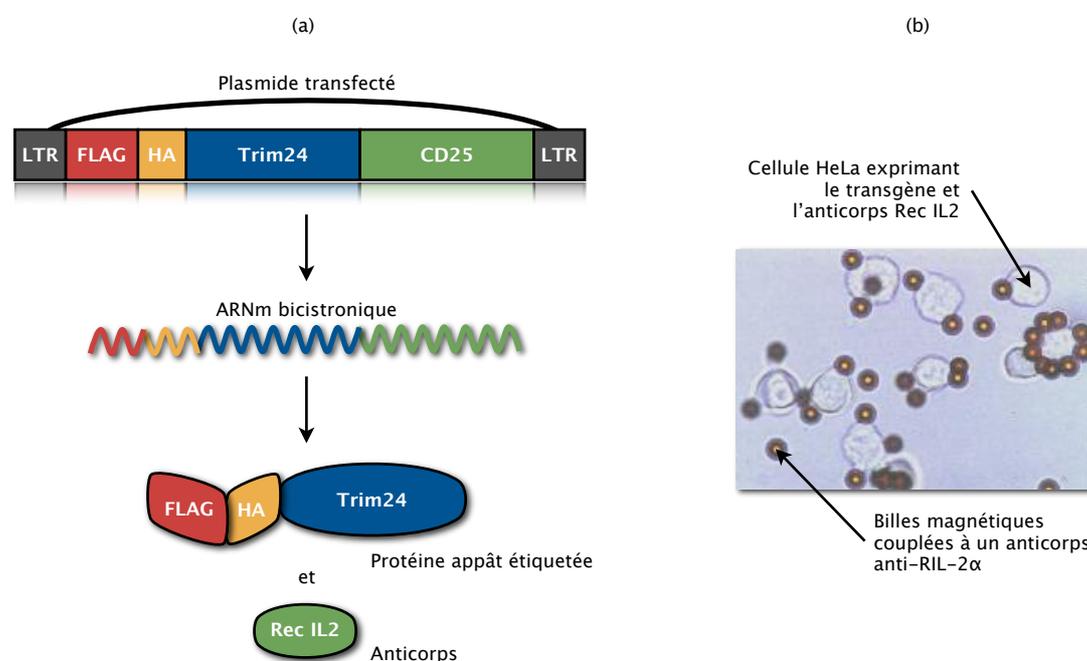
#### ▸ L'approche TAP Tag

Les techniques de purification de complexes multiprotéiques par co-immunoprécipitation exploitent les propriétés biochimiques d'étiquettes (ou « tags ») attachées à la protéine appât de manière à purifier les composés des complexes protéiques endogènes. Une des techniques les plus performantes est mise en oeuvre dans les approches de copurification par affinités (Tandem Affinity Purification, TAP) qui utilise des étapes d'enrichissement successives<sup>10</sup>.

Les approches TAP existantes<sup>11, 12</sup> (notamment les combinaisons d'étiquettes GST, calmoduline, straptavidine et protéine A) ont montré leurs limites : des interférences peuvent avoir lieu entre les ensembles de complexes endogènes et les étiquettes relativement grosses (> 20 kDa), les contaminations par des protéines non ciblées sont importantes et la nécessité de recourir à un traitement enzymatique pour éluer les complexes en séparant la cible de son étiquette implique des contaminations supplémentaires.

Le système de double étiquette FLAG® HA (Sigma) permet d'éliminer ou amoindrir ces problèmes. En effet, les deux étiquettes FLAG (DYKDDDDK) et HA (YPYDVPDYA) sont de petits épitopes et non des dérivés de protéines. Ces qualités permettent de limiter les interférences avec les protéines non impliquées dans les complexes et impliquent une plus grande spécificité. Le caractère hydrophile du peptide FLAG HA augmente les chances pour que le double épitope ne soit pas enfermé au coeur de la protéine de fusion, mais situé en surface, là où il est accessible pour établir des liaisons antigène-anticorps. Par ailleurs, les étiquettes peuvent être fusionnées en position C-terminale ou N-terminale de la protéine appât : la fusion sur l'une des extrémités uniquement risque de perturber les interactions de la protéine avec ses partenaires, il est donc préférable de tester chacune des deux possibilités.

#### ▸ Etablissement d'une lignée stable de cellule HeLa exprimant Trim24 étiquetée par FLAG et HA



**Figure 7** | (a) Construction et transfection du plasmide dans la lignée de cellules HeLa. (b) Etablissement d'une lignée stable exprimant le transgène par la sélection des cellules sur billes magnétiques couplées à l'anti-RIL-2α. (LTR : séquence terminale longue répétée)

<sup>10</sup> Kocher, T., et al. (2007). Nat Methods 4, 807-15.

<sup>11</sup> Rigaut, G., et al. (1999). Nat Biotechnol 17, 1030-2.

<sup>12</sup> Puig, O., et al. (2001). Methods 24, 218-29.

La première étape nécessite le clonage et l’expression dans un modèle cellulaire des séquences fusionnées de la protéine cible Trim24 et des étiquettes FLAG HA. Nos collaborateurs ont choisi le modèle de cellules HeLa (une lignée immortelle de cellules cancéreuses) où ils ont transfecté un plasmide contenant la séquence de Trim24 étiquetée FLAG HA en position C-ter ou N-ter, ainsi que la séquence d’un antigène de surface (Rec IL2) (Figure 7). Cet antigène permet de sélectionner les cellules exprimant le transgène en les triant sur billes magnétiques couplées à l’anti-RIL-2 $\alpha$ <sup>13</sup>.

► Copurification par affinités (TAP) des complexes de Trim24

Les protéines de plus de 2.10<sup>9</sup> cellules HeLa exprimant le transgène ont alors été extraites selon le modèle de la figure 8. Les extraits nucléaires solubles et insolubles (ou chromatinien) ont été séparés des extraits cytoplasmiques. En considérant les 2 extraits nucléaires préparés à partir de chacun des 2 construits C-ter et N-ter Trim24 FLAG-HA, on obtient 4 échantillons (Table 2).

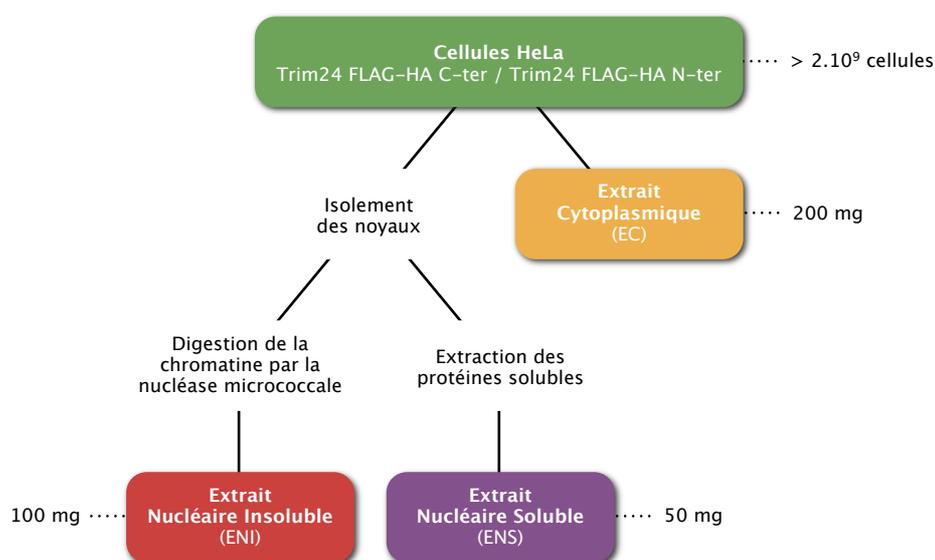


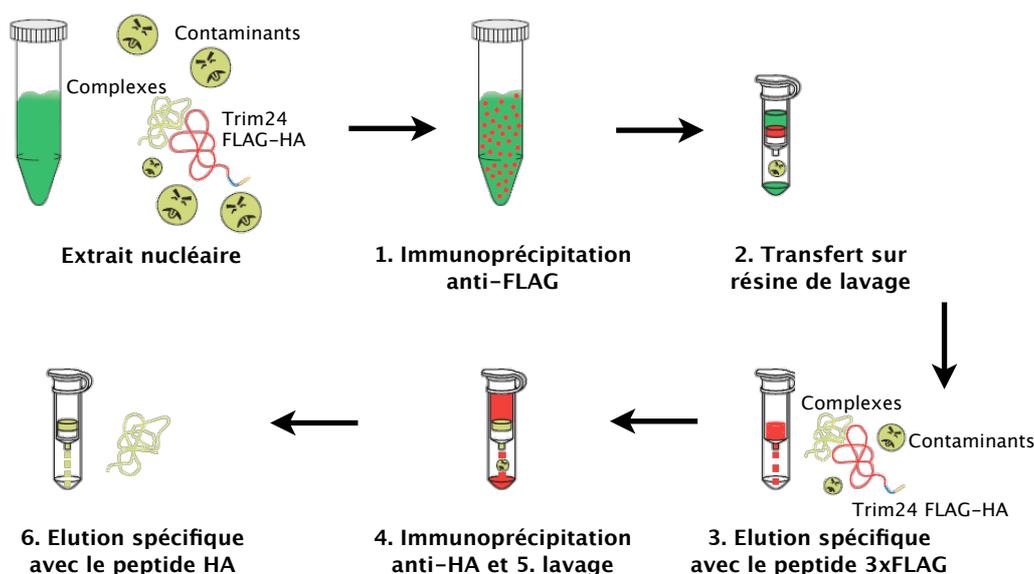
Figure 8 | Extraction des fractions nucléaires solubles et insolubles des protéines des cellules HeLa exprimant le transgène.

Trim24			
C-ter FLAG HA		N-ter FLAG HA	
Extrait nucléaire insoluble (ENI)	Extrait nucléaire soluble (ENS)	Extrait nucléaire insoluble (ENI)	Extrait nucléaire soluble (ENS)

Table 2 | Les 4 types d’échantillons préparés représentent les extraits nucléaires solubles et insolubles de cellules HeLa exprimant Trim24 étiquetée FLAG HA du côté C-ter ou N-ter.

Les complexes impliquant Trim24 dans chaque échantillon ont par la suite été purifiés selon la technique de copurification par affinités TAP (Figure 9). Elle permet de retenir séquentiellement les complexes protéiques endogènes impliquant l’appât Trim24 étiqueté FLAG HA. La procédure implique 6 étapes : (1) la résine anti-FLAG est ajoutée directement aux extraits nucléaires, (2) la résine est transférée sur une colonne de lavage, (3) la 1<sup>re</sup> élution est conduite avec le peptide 3xFLAG, elle est efficace et douce et ne détruit pas les complexes, (4) le 1<sup>er</sup> éluat est transféré directement sur une colonne contenant la résine anti-HA, (5) les complexes sont lavés une seconde fois, (6) la 2<sup>e</sup> élution est conduite avec le peptide HA.

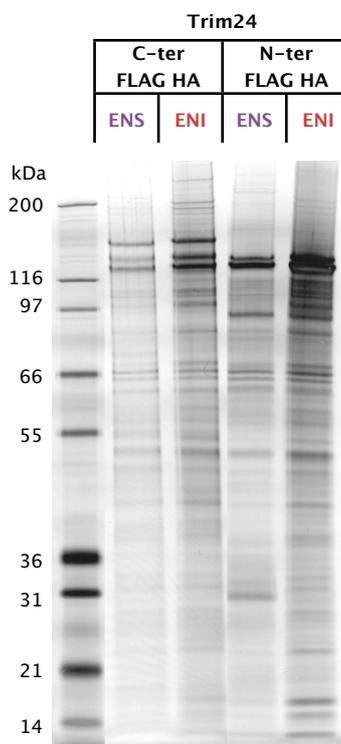
<sup>13</sup> Padmanabhan, R., et al. (1988). Anal Biochem 170, 341-8.



**Figure 9** | Représentation schématique de la stratégie de copurification par affinités (TAP) utilisant le système d'étiquettes FLAG HA. (D'après Zenser, N. – Poster Sigma Aldrich)

### 3.1.3. Séparation des protéines sur gel 1-D

Les protéines partenaires de Trim24 purifiées ont ensuite été séparées sur gel monodimensionnel de gradient 4–12 %. La figure 10 présente le gel réalisé pour notre première série d'analyse des 4 échantillons préparés.



**Figure 10** | Gel 1-D, gradient 4–12 %, coloré au bleu de Coomassie, présentant les partenaires d'interaction de Trim24 purifiés par la technique TAP, issus des extraits nucléaires solubles et insolubles de cellules HeLa exprimant Trim24 étiquetée.

Des contrôles négatifs ont par ailleurs été déposés sur un autre gel qui, une fois révélé, ne présente aucune bande colorée : la spécificité de la technique de purification pour Trim24 étiquetée FLAG-HA est donc excellente, et le gel contrôle n'a pas été analysé.

Chacune des 4 pistes du gel échantillon a été découpée de manière systématique (tous les 1 mm). Le contenu protéique des bandes obtenues a été soumis au traitement classique de réduction et d'alkylation et a été digéré par la trypsine. Le mélange peptidique complexe obtenu a alors été analysé par nanoLC-MS/MS.

### 3.1.4. Analyse par nanoLC-MS/MS

#### ▸ Valorisation des optimisations techniques et méthodologiques pour une meilleure prise de données nanoLC-MS/MS

Il a été décrit au chapitre I de cette partie l'ensemble des efforts d'optimisation et d'amélioration de la prise de données pour l'analyse nanoLC-MS/MS. Dans les approches d'analyse protéomique fonctionnelle, la sensibilité de la stratégie mise en oeuvre doit être maximale. C'est dans ces conditions que les partenaires d'interaction de faible abondance ou transitoires – qui ne sont pas forcément les moins importants – peuvent être accessibles.

Par ailleurs, dans l'objectif de comparer les protéines présentes dans chaque échantillon, un différentiel ne peut être décrit avec pertinence que dans l'hypothèse où le contenu protéique maximal est identifié. Notre étude de la répétabilité des analyses nanoLC-MS/MS et ses retombées en termes d'ajustements stratégiques ont permis de mettre au point une approche optimisée, valorisée dans cette application.

#### ▸ La stratégie d'analyse optimisée

Les améliorations suivantes ont pu être mises en oeuvre pour cette étude (revoir la figure 6) :

▸ Les analyses ont été réalisées sur le système nanoAcquity UPLC – Synapt HDMS (Waters). L'instrument, le plus récemment acquis par le laboratoire, dispose de la meilleure sensibilité et d'une vitesse d'acquisition accrue et adaptée à l'excellente résolution chromatographique du système de chromatographie liquide à ultra-haute performance.

▸ Une optimisation fine des gradients et des paramètres de cycles d'acquisition MS+MS/MS a été réalisée.

▸ Le système de correction permanente de l'étalonnage a été utilisé et paramétré pour atteindre des précisions de mesure inférieures en moyenne à 2 ppm sur une longue série d'analyse.

▸ Les analyses ont été réalisées en duplicata afin d'augmenter la part du contenu protéique total identifié et d'améliorer la qualité des identifications.

▸ Les paramètres de génération des fichiers de résultats « *peak-list* » ont été optimisés pour exporter un maximum de données et des informations de qualité.

▸ Les fichiers « *peak-list* » de chaque paire d'analyses ont été concaténés pour cumuler l'information complémentaire recueillie.

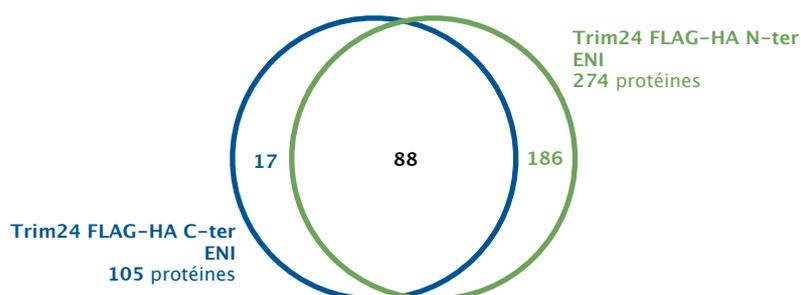
### 3.1.5. Identification des partenaires d'interaction de Trim24

Les identifications de protéines ont été réalisées selon la stratégie « *Target-Decoy* » décrite précédemment. Les données expérimentales ont été soumises au moteur de recherche Mascot contre la banque de protéines humaines UniProt, concaténée avec la banque leurre inversée. Une estimation du taux de faux positifs (FDR) et le filtrage des identifications validées ont été menés via le logiciel Scaffold (Proteome Software). Les seuils d'identification ont été ajustés pour préparer des listes de protéines contenant moins de 1 % de faux positifs.

## 3.2. Résultats

### 3.2.1. Complémentarité des résultats issus des échantillons préparés à partir de Trim24 FLAG-HA C-ter ou N-ter

Contrairement à l'ensemble des résultats qui suivent qui ont été obtenus sur le couplage nanoAcquity UPLC – Synapt HDMS (Waters) nouvellement disponible au laboratoire, des essais préliminaires ont été réalisés sur le système HPLC–Chip (Agilent) – HCT Ultra (Bruker) pour évaluer notre stratégie. Des échantillons d'extraits nucléaires insolubles, issus de purification des partenaires de Trim24 étiquetée du coté C-ter ou N-ter, ont été analysés.



**Figure 11** | Comparatif des identifications d'interactants de Trim24 purifiés par TAP FLAG-HA C-ter ou N-ter, pour les échantillons d'extraits nucléaires insolubles. Ces analyses préliminaires ont été réalisées sur le système HPLC–Chip – HCT Ultra.

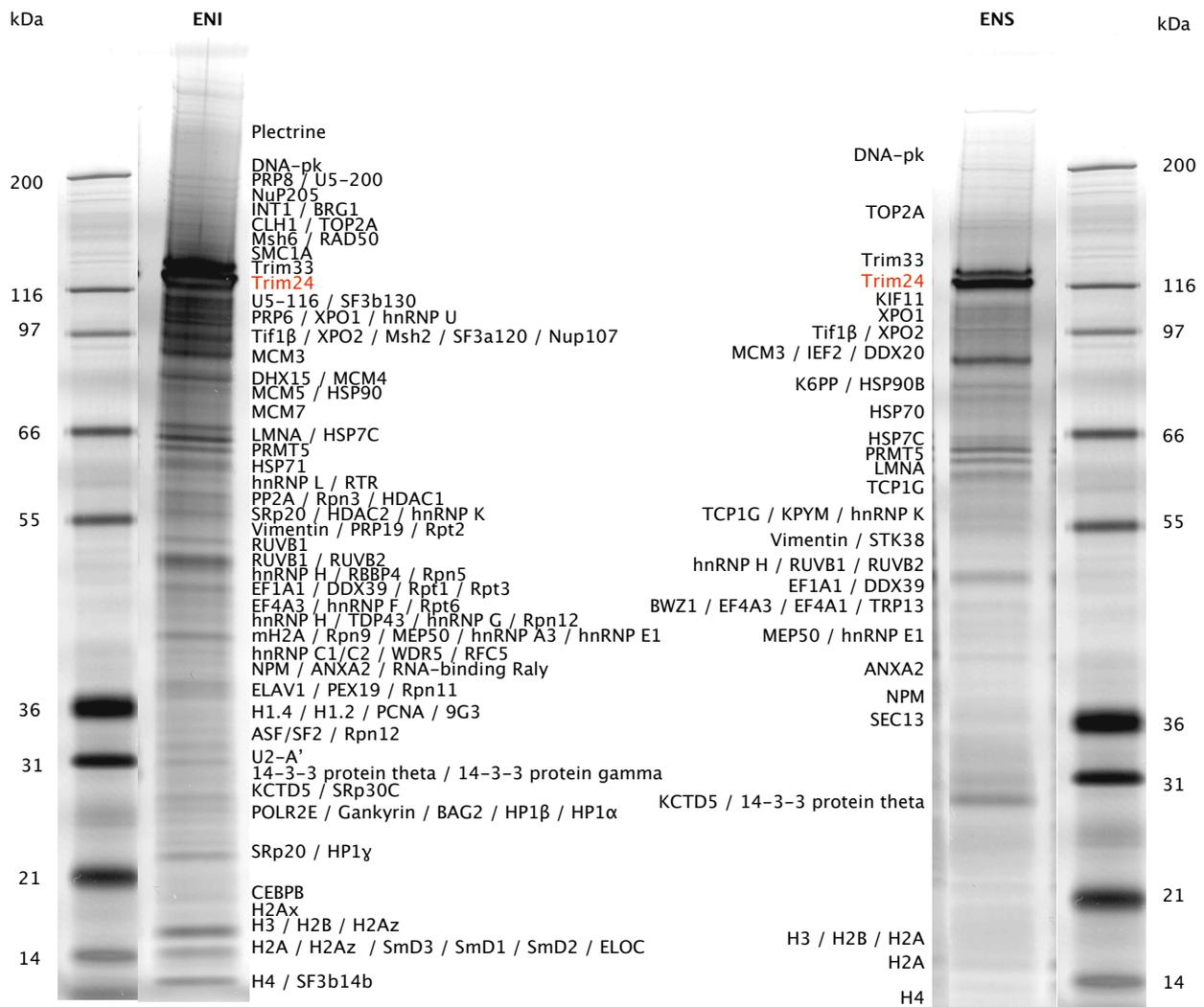
Il est connu que dans la mise en oeuvre de la stratégie TAP, l'étiquette est susceptible de modifier les interactions entre la protéine appât et ses partenaires selon qu'elle est placée du côté C-ter ou N-ter. La comparaison des protéines identifiées dans le cas de purification des partenaires de Trim24 FLAG-HA C-ter et N-ter a montré que moins de 6 % des partenaires identifiés au total étaient accessibles uniquement dans le cas de la purification C-ter (Figure 11). Face aux contraintes matérielles et au temps nécessaire à l'analyse systématique de ces deux types d'échantillons, il a été décidé d'orienter l'étude sur les partenaires préparés uniquement à partir des purifications réalisées avec Trim24 FLAG-HA N-ter.

### 3.2.2. Identification de sous-complexes de Trim24 : Trim 33 un partenaire majeur

L'analyse des partenaires d'interaction de Trim24 purifiés par la technique TAP FLAG-HA a conduit à l'identification de 342 protéines pour l'extrait nucléaire insoluble et 74 pour l'extrait nucléaire soluble. Ces identifications ont pu être obtenues en ajustant les seuils de score du moteur de recherche Mascot pour atteindre un taux de faux positifs (FDR) inférieur à 1% (Table 3). Le détail des identifications figure en annexe (A2) et les protéines particulièrement intéressantes ont été reportées sur le gel 1-D présenté en figure 12.

Echantillon	Seuil pour les identifications à plus de 2 peptides uniques	Seuil pour les identifications à 1 seul peptide unique
N-ter FLAG HA Trim24 ENI	$\Delta > -16$	$\Delta > +6$
N-ter FLAG HA Trim24 ENS	$\Delta > -27$	$\Delta > +8$

**Table 3** | Seuils de score  $\Delta$  (score d'ion – score identité) permettant d'atteindre un taux de faux positifs (FDR) inférieur à 1 % pour les échantillons N-ter FLAG HA Trim24 ENI et ENS.



**Figure 12** | Gel 1-D 4-12 %, coloré au bleu de coumassie, présentant la séparation des partenaires d'interaction de Trim24, purifiés par la technique TAP FLAG-HA. Le gel est annoté avec les protéines d'intérêt qui permettent d'évaluer quels sont les complexes impliquant Trim24.

► Identification de Trim24 et Trim33

En premier lieu, l'appât Trim24 a pu être identifié dans de nombreuses bandes : 57 peptides uniques ont pu être séquencés, conduisant à un recouvrement de séquence de 69 %. Cette identification d'excellente qualité confirme avec force l'efficacité de la purification de Trim24 et ses partenaires.

Sur le gel 1-D, une bande très intense peut être remarquée autour de 116 kDa. L'identification de son contenu protéique révèle qu'il s'agit de la protéine E3 ubiquitin-protein ligase TRIM33 (Q9UPN9), inconnue à ce jour pour interagir avec Trim24. Le recouvrement de séquence atteint 56 % grâce à la caractérisation de 45 peptides (Figure 13). Cette identification de qualité et l'intensité de la bande de gel correspondante laissent supposer que Trim33 est un partenaire majeur de Trim24. Cette interaction fait l'objet d'une étude particulière dans le point suivant.

E3 ubiquitin-protein ligase TRIM33										Q9UPN9
1	GGGEA	ESGGG	GSGSA	PVTAG	AAGPA	AQEAE	PPLTA	VLVEE	EEEEG	GRAGA
51	EGGAA	GPDDG	GVAAG	SSGSA	QAASS	PAASV	GTGVA	GGAVS	TPAPA	PASAP
101	APGPS	AGPPP	GPPAS	LLDTC	AVCQQ	SLQSR	REAEF	KLLPC	LHSFC	LRCLP
151	EPERQ	LSVPI	PGGSN	GDIQQ	VG VIR	CPVCR	QECRQ	IDLVD	NYFVK	DTSEA
201	PSSSD	EKSEQ	VCTSC	EDNAS	AVGFC	VECGE	WLCKT	CIEAH	QRVKF	TKDHL
251	IRKKE	DVSES	VGASG	QRPVF	CPVHK	QEQLK	LFCET	CDRLT	CRDCQ	LLEHK
310	EHRYQ	FLEEA	FQNQK	GAIEN	LLAKL	LEKKN	YVHFA	ATQVQ	NRIKE	VNETN
351	KRVEQ	EIKVA	IFTLI	NEINK	KGKSL	LQQLL	NVTKE	RQMKL	LQQQN	DITGL
401	SRQVK	HVMNF	TNWAI	ASGSS	TALLY	SKRLI	TFQLR	HILKA	RCDPV	PAANG
451	AIRFH	CDPTF	WAKNV	VNLGN	LVIES	KPAPG	YTPNV	VVGQV	PPGTN	HISKT
501	PGQIN	LAQLR	LQHMQ	QQVYA	QKHQQ	LQQMR	MQQPP	APVPT	TTTTT	QQHPR
551	QAAPQ	MLQQQ	PPRLI	SVQTM	QRGNM	NCGAF	QAHQM	RLAQN	AARIP	GIPRH
601	SGPQY	SMMQP	HLQRQ	HSNPG	HAGPF	PVVSF	HNTTI	NPTSP	TTATM	ANANR
651	GPTSP	SVTAI	ELIPS	VTNPE	NLPSL	PDIPP	IQLED	AGSSS	LDNLL	SRYIS
701	GSHLP	PQPTS	TMNPS	PGPSA	LSPGS	SGLSN	SHTPV	RPPST	SSTGS	RGSCG
751	SSGRT	AEKTS	LSFKS	DQVKV	KQEPG	TEDEI	CSFSG	GVKQE	KTEDG	RRSAC
801	MLSSP	ESSLT	PPLST	NLHLE	SELDA	LASLE	NHVKT	EPADM	NESCK	QSGLS
851	SLVNG	KSPIR	SLMHR	SARIG	GDGNN	KDDDP	NEDWC	AVCQN	GGDLL	CCEKC
901	PKVFH	LTCHV	PTLLS	FPSGD	WICTF	CRDIG	KPEVE	YDCDN	LQHSK	KGKTA
951	QGLSP	VDQRK	CERLL	LYLYC	HELSE	EFQEP	VPASI	PNYYK	IIKKP	MDLST
1001	VKKKL	QKKHS	QHYQI	PDDVF	ADVRL	IFKNC	ERFNE	MMKVV	QVYAD	TQEIN
1051	LKADS	EVAQA	GKAVA	LYFED	KLTEI	YSDRT	FAPLP	EFEQE	EDDGE	VTEDS
1101	DEDFI	QPRRK	RLKSD	ERPVI	IK					

Figure 13 | Séquence de Trim33 identifiées (en rouge) par nanoLC-MS/MS pour l'échantillon Trim24 TAP FLAG-HA. Le pourcentage de recouvrement global obtenu s'élève à 56 % grâce à l'identification de 45 peptides uniques.

► Fonction des sous-complexes de Trim24 identifiés

Les protéines d'intérêt identifiées ont été classées sur la figure 14 selon leur fonction connue. Ont pu être relevés parmi ces interactants potentiels de Trim24, des partenaires connus pour jouer des rôles déterminants dans les mécanismes de transcription et pour être impliqués dans des complexes appartenant à la machinerie d'élongation de l'ARNm (Spliceosome C complexe), dans des complexes contrôlant la réplication, la réparation ou la recombinaison de l'ADN ou dans des complexes impliqués dans la modification de la chromatine et le contrôle de la transcription.

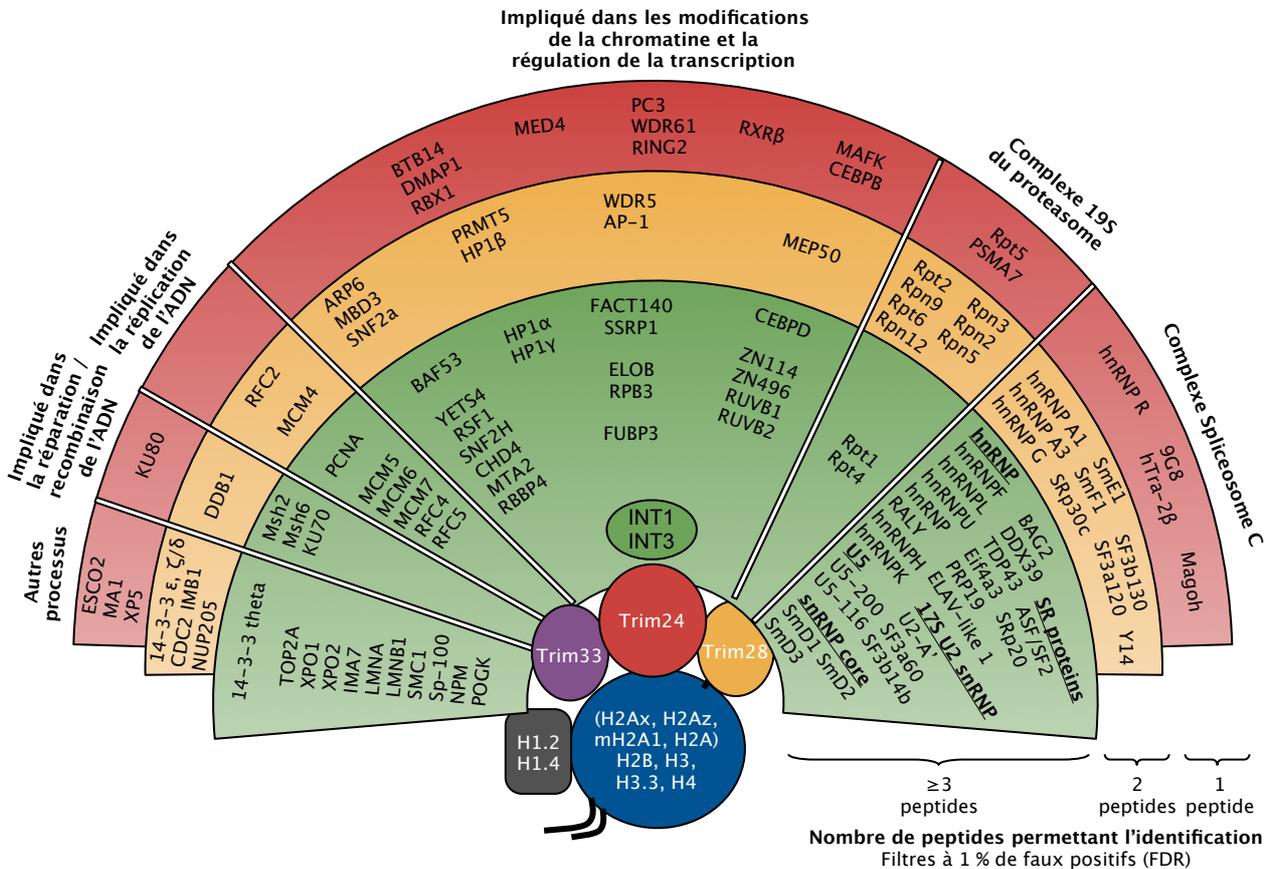


Figure 14 | Caractérisation des complexes contenant Trim24 par analyse protéomique fonctionnelle par spectrométrie de masse nanoLC-MS/MS. Les protéines identifiées sont classées selon leur fonction cellulaire autour de Trim24 et selon le nombre de peptides ayant permis l'identification.

▸ Conclusions

Ces premiers résultats soulèvent des questions importantes sur le plan analytique et biologique :

▸ Il a d’abord été mis en évidence que Trim24 semblait être fortement associée à Trim33. Cela suggère que Trim33 soit également susceptible d’affecter la transcription des gènes cibles de l’acide rétinoïque. Une étude des interactants de Trim33 selon la stratégie qui vient d’être exposée peut être réalisée pour confirmer l’interaction et évaluer les partenaires de Trim33.

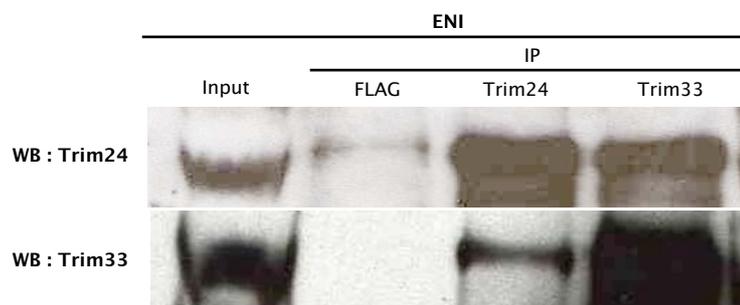
▸ Le complexe chromatinien (fraction nucléaire insoluble) de Trim24 semble contenir plusieurs partenaires impliqués dans l’élongation de la transcription. Trim24 est donc susceptible de jouer un rôle dans ce mécanisme.

▸ Trim24 est impliquée dans plusieurs sous-unités du protéasome. La régulation de la voie de l’acide rétinoïque pourrait donc s’établir par un processus d’ubiquitinylation/protéasome.

**3.2.3. Mise en évidence de l’interaction Trim24 – Trim33**

▸ Expériences de Western-Blotting

Pour évaluer l’interaction entre Trim24 et Trim33, une étude complémentaire a été menée à 2 niveaux. D’abord, l’interaction entre les deux protéines endogènes a été validée par des expériences de Western Blotting (Figure 15). Avec un modèle de cellules MEF (fibroblastes embryonnaires de souris), des immunoprécipitations de Trim24 et Trim33 ont été réalisées et les complexes retenus ont été déposés sur gel. La révélation par les anticorps anti-Trim24 et anti-Trim33 a permis de montrer que la majorité de Trim24 est associée à Trim33, alors que la réciproque n’est pas vérifiée.



**Figure 15** | Etude de l’interaction entre Trim24 et Trim33 endogène par immunoprécipitation (IP) et révélation par Western Blotting (WB). La majorité de Trim24 est associée à Trim33, mais ce n’est pas réciproque.

▸ Purification TAP FLAG-HA des complexes de Trim33

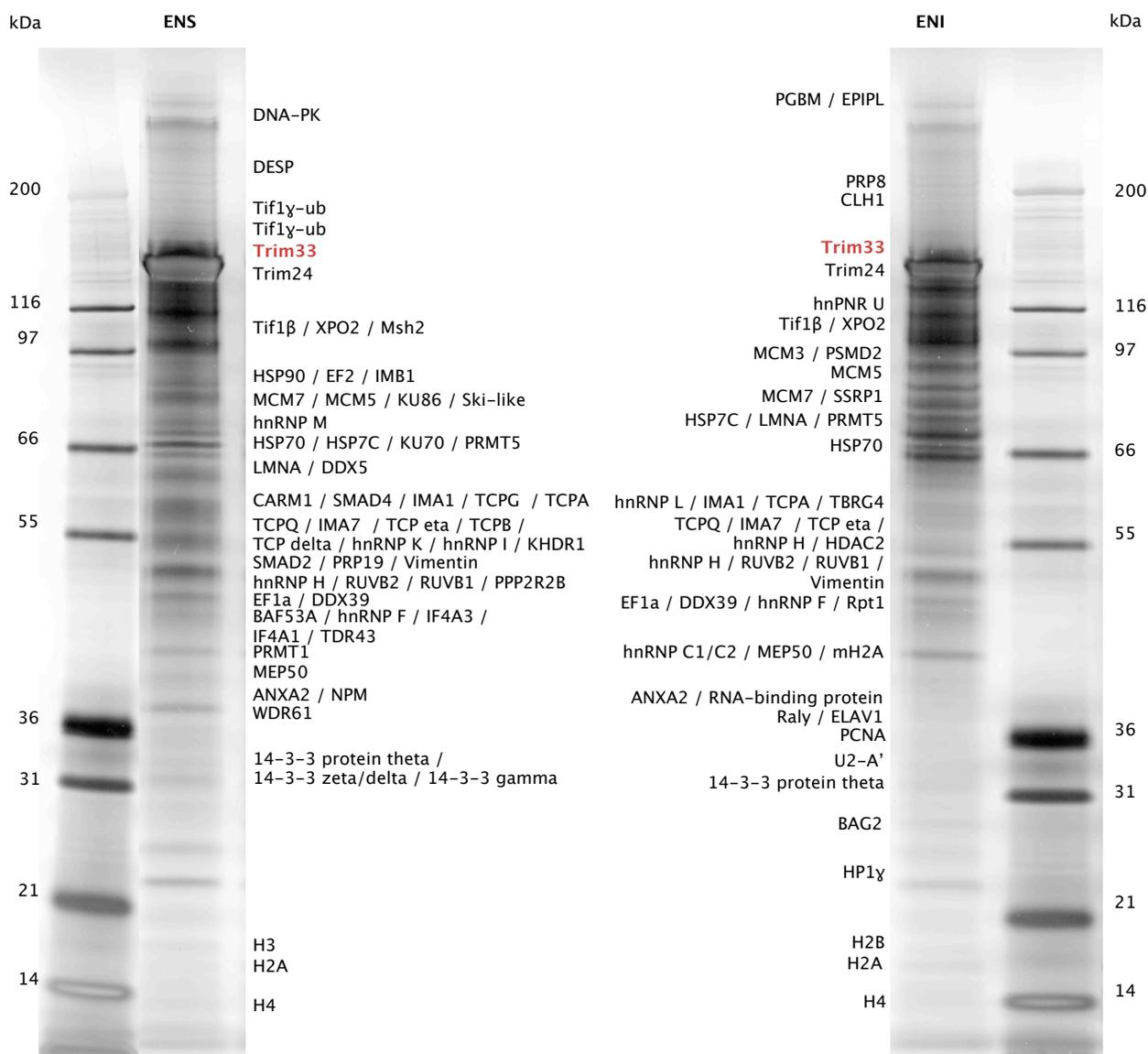
Pour aller plus loin, une lignée stable de cellules HeLa exprimant Trim33 étiquetée FLAG-HA a été préparée selon le protocole décrit précédemment et les partenaires impliqués dans les complexes de Trim33 ont été purifiés. Les échantillons ont été analysés par nanoLC-MS/MS en suivant notre approche optimisée pour l’analyse protéomique fonctionnelle. Notamment, des seuils de score conférant une meilleure confiance dans les identifications (taux maximal de faux positifs (FDR) de 1%) ont été fixés selon les valeurs données en table 4.

Echantillon	Seuil pour les identifications à plus de 2 peptides uniques	Seuil pour les identifications à 1 seul peptide unique
N-ter FLAG HA Trim33 ENI	$\Delta > -24$	$\Delta > +5$
N-ter FLAG HA Trim33 ENS	$\Delta > -26$	$\Delta > +11$

**Table 4** | Seuils de score  $\Delta$  (score d’ion – score identité) permettant d’atteindre un taux de faux positifs (FDR) inférieur à 1 % pour les échantillons N-ter FLAG HA Trim33 ENI et ENS.

Pour la fraction nucléaire insoluble, 337 protéines ont été identifiées et pour la fraction nucléaire soluble, 176 protéines ont été rapportées. Le détail des identifications figure en annexe (A3) et le gel 1-D analysé a été annoté en figure 16 avec les interactants jugés potentiellement intéressants pour une étude ultérieure plus poussée. On retrouve notamment les identifications de Trim33 et Trim24 avec le plus grand nombre de peptides. 55 % de l'information de séquence de Trim33 a été recouverte grâce à la caractérisation de 47 peptides uniques. Pour Trim24, les 38 peptides uniques identifiés représentent 53 % de la couverture de séquence.

A partir de ces résultats qui valident l'interaction Trim24-Trim33, des études permettant d'évaluer sa fonction sont en cours. Les premiers résultats montrent que Trim24 et Trim33 semblent travailler en synergie dans le contrôle des gènes cibles de l'acide rétinolique.



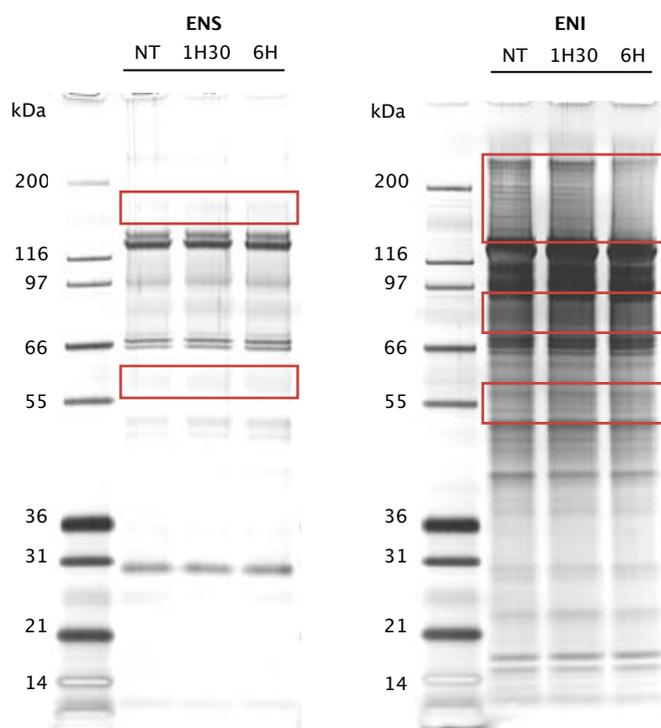
**Figure 16** | Gel 1-D 4-12 %, coloré au bleu de coumassie, présentant la séparation des partenaires d'interaction de Trim33, purifiés par la technique TAP FLAG-HA. Le gel est annoté avec les protéines d'intérêt qui permettent d'évaluer quels sont les complexes impliquant Trim33.

### 3.2.4. Comparaison des partenaires de Trim24 purifiés à partir d'une lignée cellulaire traitée ou non traitée à l'acide rétinoïque

- Un traitement à l'acide rétinoïque semble modifier les complexes impliquant Trim24

Il a été montré qu'activer la voie de signalisation de l'acide rétinoïque dans le foie était oncogénique. Pour évaluer si l'acide rétinoïque est en mesure de modifier les partenaires des complexes chromatinien et solubles de Trim24, une étude des interactants de Trim24 purifiés à partir de lignées cellulaires traitées ou non traitées à l'acide rétinoïque a été menée.

Une lignée de cellules HeLa a été traitée par un large excès d'acide rétinoïque (1  $\mu$ M) pendant plusieurs heures avant l'extraction des protéines nucléaires solubles et insolubles. Le dépôt sur gel 1-D (Figure 17) montre des différences flagrantes entre les échantillons traités et non traités. Notamment, des bandes de haut poids moléculaire disparaissent dans le cas ENI traité. Cela peut témoigner d'une modification importante de la structure du complexe chromatinien.



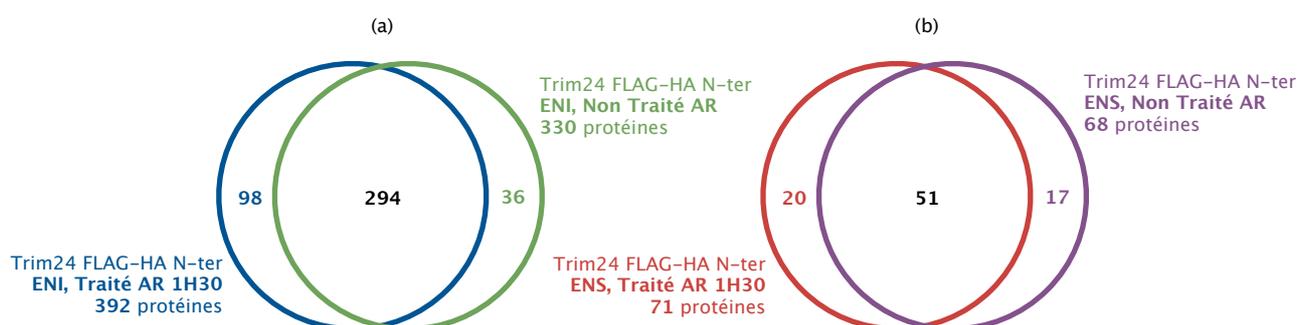
**Figure 17** | Gels 1-D 4-12 %, coloré au bleu de coumassie, présentant des différences importantes (cadres rouges) pour les complexes chromatinien et nucléaire soluble de Trim24 purifiés par TAP FLAG-HA N-ter, entre les cas traités (pendant 1h30 et 6H) et non traités (NT) par un excès d'acide rétinoïque.

- Analyse des partenaires d'interaction de Trim24 de lignées traitées ou non à l'acide rétinoïque

Les pistes de gel 1-D d'échantillons traités et non traités ont été découpées systématiquement et analysées par nanoLC-MS/MS. Une découpe systématique a été préférée à la sélection de quelques bandes pour ne pas risquer de passer à côté de protéines présentes différemment, mais qui pourraient être masquées par les protéines majoritaires. Grâce à notre stratégie d'analyse fonctionnelle et différentielle, dont la sensibilité et la répétabilité ont été optimisées, un total de 428 protéines différentes ont pu être identifiées pour les extraits chromatinien et de 88 pour les extraits nucléaires solubles (Figure 18). Ces résultats ont été obtenus avec une taux de faux positifs évalué à 1 % pour les critères d'identification indiqués en table 5. Le détail des identifications et le tableau comparatif des protéines identifiées sont disponibles en annexe (A4).

Echantillon	Seuil pour les identifications à plus de 2 peptides uniques	Seuil pour les identifications à 1 seul peptide unique
N-ter FLAG HA Trim24 ENI, Traité AR	$\Delta > -15$	$\Delta > +5$
N-ter FLAG HA Trim24 ENI, Non traité	$\Delta > -15$	$\Delta > +9$
N-ter FLAG HA Trim24 ENS, Traité AR	$\Delta > -23$	$\Delta > +3$
N-ter FLAG HA Trim24 ENS, Non traité	$\Delta > -23$	$\Delta > +15$

**Table 5** | Seuils de score  $\Delta$  (score d'ion - score identité) permettant d'atteindre un taux de faux positifs (FDR) inférieur à 1 % pour les échantillons N-ter FLAG HA Trim24, ENI et ENS, traités à l'acide rétinoïque et non traités.



**Figure 18** | Comparatif des identifications d'interactants de Trim24 purifiés par TAP FLAG-HA N-ter, pour les échantillons d'extraits nucléaires insolubles (a) et solubles (b).

Dans le cas des échantillons traités, plus d'un quart des partenaires potentiels de Trim24 semblent recrutés spécifiquement en présence d'acide rétinoïque. En effet, 98 protéines sur les 392 identifiées pour l'extrait nucléaire insoluble traité sont spécifiques à cet échantillon. Pour l'extrait soluble traité, ce sont 20 protéines sur les 71 identifiées qui lui sont propres. Au-delà de ce simple comparatif ne faisant apparaître que les identifications spécifiques à l'un ou l'autre des échantillons, il peut être intéressant de se pencher sur d'éventuelles protéines identifiées dans le cas non traité et potentiellement surexprimées dans le cas traité. Le nombre de peptides permettant l'identification de telles protéines est généralement bien supérieur en cas de surexpression. Étonnamment, peu de protéines identifiées dans les échantillons non traités et dans les échantillons traités, mais avec un nombre de peptides bien plus grand ont pu être mises en évidence.

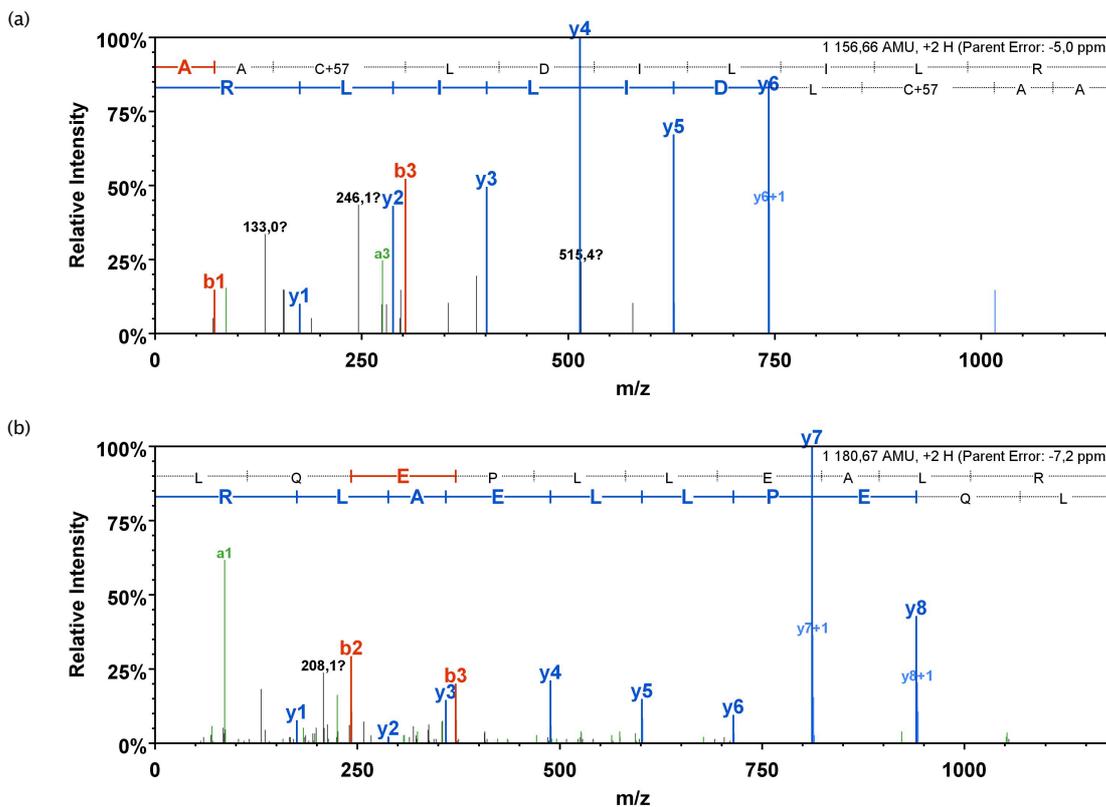
#### ▸ Identification des récepteurs de l'acide rétinoïque

Parmi les protéines identifiées spécifiquement dans le complexe chromatinien des échantillons traités à l'acide rétinoïque, on relève tout particulièrement la présence de RAR $\alpha$  (ou RAR $\beta$ ), RAR $\gamma$ , RXR $\alpha$ , RXR $\beta$  et LXR $\beta$  (Table 6).

Les spectres des peptides attribués aux protéines RAR ont été vérifiés manuellement (Figure 19). Si RAR $\gamma$  a pu être validé sans ambiguïté, il n'a pas été possible de discriminer entre RAR $\alpha$  et RAR $\beta$  car le peptide permettant l'identification appartient à une partie de séquence commune aux deux protéines. Cependant, des expériences de Wester Blotting ont permis de vérifier la présence spécifique de RAR $\alpha$  dans le cas d'échantillons traités à l'acide rétinoïque.

Protéine	Numéro d'accèsion	Nombre de peptides uniques	Recouvrement de séquence
RAR $\alpha$ (ou RAR $\beta$ )	P10276 (P10826)	1	2 %
RAR $\gamma$	P13631	1	2 %
RXR $\alpha$	P19793	3	16 %
RXR $\beta$	P28702	11	24 %
LXR $\beta$	P55055	7	18 %

**Table 6** | Identification de protéines RAR, RXR et LXR comme partenaires d'interaction de Trim24 dans les échantillons traités à l'acide rétinoïque.



**Figure 19** | Validation manuelle par séquençage de novo des peptides uniques permettant l'identification de (a) RAR $\alpha$  ou RAR $\beta$  (peptide AACIAMLDILILR, 579,34 m/z, 2+) ; et (b) RAR $\gamma$  (peptide LQEPLLEALR, 591,34 m/z, 2+).

L'identification de ces protéines, et notamment celle de RAR rarement observée en analyse protéomique du fait du caractère transitoire des complexes qu'elle forme, révèle la participation de Trim24 à la voie de signalisation de l'acide rétinoïque et conforte l'investissement engagé dans la recherche des interactants de Trim24.

▸ Caractérisation des complexes contenant Trim24 sous traitement à l'acide rétinoïque

Parmi toutes les protéines identifiées dans les expériences présentées, on en relève beaucoup qui sont connues pour être impliquées dans de multiples complexes jouant un rôle dans le contrôle de l'activité transcriptionnelle. On relève des complexes agissant comme facteur de remodelage pour ouvrir ou fermer la chromatine (BAF53, RSF, ...), des éléments du protéasome (sous unités 19S) ou participant au mécanisme de la réplication (Cul4A). Dans la plupart des cas, la majorité des protéines formant ces complexes ont été identifiées. Elles ont été rapportées en table 7.

<b>Le Complexe NuRD</b> HDAC1, HDAC2 RbAp46, RbAp48 Mi-2 $\alpha$ , Mi-2 $\beta$ MTA1, MTA2 p66	<b>Le Complexe FACT</b> Fact140 SSRP1	<b>Le Complexe MEDIATEUR</b> MED1, MED4, MED6, MED7, MED8, MED9, MED10, MED11, MED12, MED13, MED13L, MED14, MED15, MED16, MED17, MED18, MED19, MED20, MED21, MED22, MED23, MED24, MED25, MED26, MED27, MED29, MED30, MED31, CCNC, CDK8 et CDC2L6/CDK11.	<b>19S subunits</b> S7 (Rpt1) S4 (Rpt2) S6b (Rpt3) S10b (Rpt4) S6a (Rpt5) S8 (Rpt6) S2 (Rpn1) S1 (Rpn2) S3 (Rpn3) p55 (Rpn5) S9 (Rpn6) S10a (Rpn7) S12 (Rpn8) S11 (Rpn9) S5a (Rpn10) S13 (Rpn11) S14 (Rpn12) ADRM1 (hRpn13) PAAF1 (Rpn14)
<b>Le Complexe Sin3</b> HDAC1, HDAC2 RbAp46, RbAp48 Sin3 SAP18 SAP30	<b>Le Complexe RSF</b> RSF SNF2h	<b>Le Complexe NuA4</b> HTATIP/TIP60, EP400, TRRAP/PAF400, BRD8/ SMAP, EPC1, DMAP1/ DNMAP1, RUVBL1/TIP49, RUVBL2, ING3, actin, ACTL6A/BAF53A, MORF4L1/MRG15, MORF4L2/MRGX, MRGBP, YEATS4/GAS41 et VPS72/ YL1.	<b>Le Complexe Elongin BC</b> ELoC EloB Cul2-Rbx1 ou Cul5-Rbx2
<b>Le Complexe MLL</b> MLL2-MLL3 or MLL3- MLL4 ASH2L, RBBP5, DPY30, NCOA6, WDR5, MEN1, UTX PAXIP1/PTIP	<b>Le Complexe WICH</b> SNF2h BAZ1B	<b>Le Complexe BAF</b> actin (ACTB), ARID1A, ARID1B/BAF250, SMARCA2, SMARCA4/ BRG1, ACTL6A/BAF53, ACTL6B/BAF53B, SMARCE1/BAF57 SMARCC1/BAF155, SMARCC2/BAF170, SMARCB1/SNF5/INI1, et un ou plus parmi : SMARCD1/BAF60A, SMARCD2/BAF60B, ou SMARCD3/BAF60C	<b>Le Complexe Cul4A</b> DDB1 Cul4A RBX1 CAND1
<b>Le Complexe SET</b> (SETD1A or SETD1B) WDR5 WDR82 RBBP5 ASH2/ASH2L CXXC1/CFP1	<b>Le Complexe NORC</b> SNF2h BAZ2A	<b>Le Complexe INO80</b> ACTL6A, ACTR5, ACTR8, RVBL1, RVBL2, INO80, INO80B, INO80C, INO80D et INO80E.	
	<b>Le Complexe BAF53</b> BAF53A, RUVBL1, SMARCA4/BRG1, TRRAP	<b>Le complexe BASC</b> BRCA1, MSH2, MSH6, MLH1, ATM, BLM, RAD50, MRE11A et NBN	

**Table 7** | Liste des complexes reconnus comme impliquant Trim24 et jouant un rôle dans la régulation de l'activité transcriptionnelle. Les protéines identifiées sont indiquées en rouge. Certaines – en gras – ont été analysées par Western Blotting après fractionnement des complexes sur gradient de glycérol.

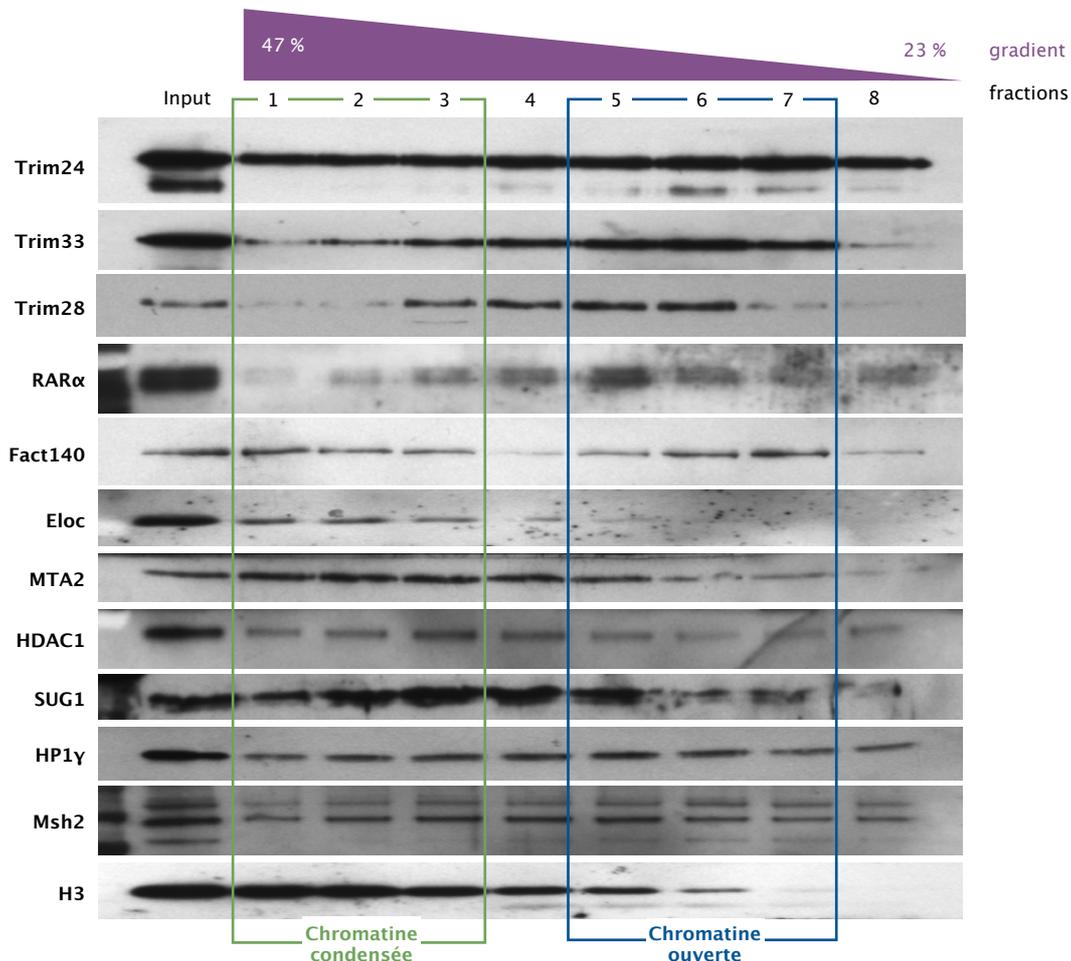
Plus particulièrement, certains interactants potentiels de Trim24 impliqués dans ces complexes, et dont l'identification est présentée en table 8, ont retenu notre attention pour être soumis à validation.

Protéine	Accession (Abbréviation)	Nombre de peptides unique permettant la meilleure identification			
		ENI Non traité	ENI Traité AR	ENS Non traité	ENS Traité AR
Transcription intermediary factor 1–alpha	O15164 (Trim24)	48	47	25	28
E3 ubiquitin–protein ligase TRIM33	Q9UPN9 (Trim33)	35	38	23	23
Transcription intermediary factor 1–beta	Q13263 (Trim28)	39	38	24	17
Retinoic acid receptor alpha	P10276, P10826 (RAR $\alpha$ )		1		
FACT complex subunit SPT16	Q9Y5B9 (Fact140)	6	7		
Transcription elongation factor B polypeptide 1	Q15369 (Eloc)		1		
Metastasis–associated protein MTA2	O94776 (MTA2)	5	5		
Histone deacetylase 1	Q13547 (HDAC1)	6	6		
26S protease regulatory subunit 8	P62195 (SUG1)		2		
Chromobox protein homolog 3	Q13185 (HP1 $\gamma$ )	10	10		
DNA mismatch repair protein Msh2	P43246 (Msh2)	3	4		
Histone H3	P68431, P84243 Q16695, Q71DI3 (H3)	6	9		6

**Table 8** | Identification de partenaires potentiels de Trim24 impliqués dans des complexes de régulation de la transcription.

► Analyse du complexe Trim24 traité à l'acide rétinoïque par gradient de glycérol

Les protéines particulièrement remarquées lors de leur identification par spectrométrie de masse ont été analysées par Western-Blotting. Pour différencier les complexes de Trim24 impliqués dans une chromatine fermée ou ouverte, les complexes ont été fractionnés sur gradient de glycérol 23-47 % par ultracentrifugation. Les fractions ont été analysées par Western Blotting pour les protéines d'intérêt (Figure 20).



**Figure 20** | Fractionnement du complexe non traité sur gradient de glycérol 23-47 % et analyse des fractions pour différentes protéines par Western Blotting. Les protéines les plus retenues (fraction 1-3) sont associées au complexe chromatinien condensé et les protéines moins retenues (fractions 5-7) sont associées à la chromatine ouverte.

Le complexe Trim24 associé à la chromatine condensée contient le complexe NuRD (MTA2) qui est un complexe de répression de la transcription, alors que les complexes associés à une chromatine ouverte contiennent certains candidats de l'élongation de la transcription comme Fact140 et HP1 $\gamma$ .

Il semble par conséquent que Trim24 soit impliquée dans plusieurs sous complexes, distincts selon qu'ils sont associés à une chromatine ouverte ou fermée.

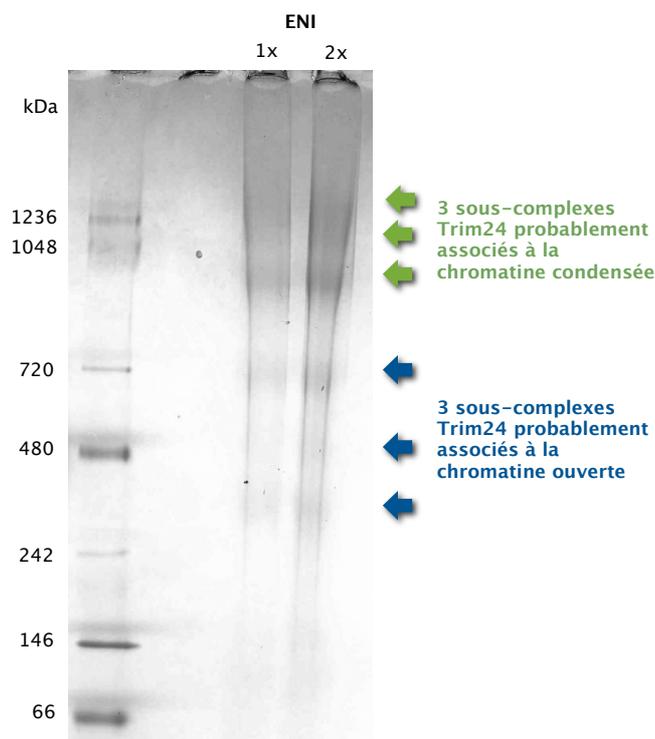
### 3.2.5. Perspectives

Si ces premiers résultats apportent des connaissances précieuses sur l'implication de Trim24 dans des complexes de régulation de la transcription, de nouvelles approches doivent permettre d'aller plus loin dans la compréhension de ces phénomènes moléculaires. La collaboration avec l'équipe de biologistes doit se poursuivre selon les 3 priorités exposées ci-dessous.

▸ Analyse des complexes après séparation sur gel natif

Pour affiner leur identification, les différents complexes mis en évidence peuvent être analysés après séparation sur gel d'électrophorèse natif (Blue Native (BN)-Page). Il s'agit d'une méthode de séparation de complexes multiprotéiques sur gel d'électrophorèse pour laquelle la migration des complexes dépend de la charge négative conférée par le bleu de Coomassie qui s'y lie, de leur taille et de leur forme<sup>14</sup>. Le bleu de Coomassie n'agit pas comme un détergent et préserve la structure des complexes analysés. La résolution du BN-PAGE est bien supérieure aux méthodes de gel-filtration ou de séparation sur gradient de sucrose et peut permettre de déterminer la taille, l'abondance relative et la composition des sous-unités d'un complexe<sup>15</sup>.

Les analyses sont en cours, mais les premiers essais montrent la formation de 3 sous-complexes de Trim24 associés à la chromatine condensée et de 3 sous-complexes de Trim24 associés à la chromatine ouverte (Figure 21). La mise au point des conditions de séparation est délicate et nécessite une quantité de matériel importante.



**Figure 21** | Gel natif (BN-PAGE) 4–16 % réalisé pour l'analyse des complexes de Trim24 purifiés par la technique TAP FLAG-HA pour l'extrait nucléaire insoluble.

▸ Analyse des complexes après purification par double co-immunoprécipitation

L'établissement de nouvelles lignées cellulaires stables exprimant deux protéines étiquetées différemment permettra l'analyse plus spécifique de sous-complexes impliquant les 2 protéines visées. Grâce aux analyses précédentes, les appâts peuvent être choisis de manière à orienter les études vers des sous-complexes particuliers. En réalisant une co-immunoprécipitation des complexes de Trim24 puis des sous-complexes de Trim24 associés à une autre protéine, l'analyse par spectrométrie de masse nanoLC-MS/MS des protéines purifiées et séparées sur gel 1-D permettra de déterminer les partenaires impliqués.

Nos collaborateurs travaillent à la préparation des lignées cellulaires stables présentées en table 9.

<sup>14</sup> Wittig, I., et al. (2008). *Proteomics* **8**, 3974–90.

<sup>15</sup> Swamy, M., et al. (2006). *Sci STKE* **2006**, pl4.

Lignée cellulaire	Protéine 1	Protéine 2
HeLa	Trim24-Nt-FLAG-HA	Trim33-Nt-Strep-V5
HeLa	Trim24-Nt-FLAG-HA	RAR $\alpha$ -Nt-Strep-V5
HeLa	Trim24-Nt-FLAG-HA	MTA2-Nt-Strep-V5
HeLa	Trim24-Nt-FLAG-HA	POLR2C-Nt-Strep-V5
HeLa	Trim24-Nt-FLAG-HA	Sug1-Nt-Strep-V5

**Table 9** | Lignées cellulaires stables en préparation qui permettront l'analyse de sous-complexes de Trim24 impliquant diverses protéines sélectionnées parmi celles qui ont été identifiées dans les analyses préliminaires.

► Analyse des complexes endogènes grâce à l'établissement d'une lignée de souris Trim24<sup>FLAG-HA</sup>

A plus long terme, il sera possible d'analyser les complexes endogènes de Trim24 grâce à la lignée de souris hétérozygote Trim24-FLAG-HA qui vient d'être mise au point. Les complexes de Trim24 pourront donc être purifiés directement à partir de foies de souris.

### 3.3. Valorisation de ce travail : communication par affiche

Béchade, G., Ouararhni, K., Van Dorselaer, A., Losson, R., Sanglier-Cianferani, S.  
 Recherche de partenaires d'interaction du facteur de transcription Trim24.  
 Congrès Spectrométrie de Masse et Analyse Protéomique (SMAP)  
 Dijon (France). 14-17 septembre 2009.

## 4. Conclusion

Partant des connaissances acquises par nos collaborateurs sur le rôle du facteur de transcription Trim24, reconnu comme étant un suppresseur de tumeurs dans le foie et un régulateur de la voie de signalisation de l'acide rétinoïque, nous avons développé une stratégie d'analyse des partenaires d'interaction du cosuppresseur.

Grâce à la mise au point d'un protocole permettant l'étude fonctionnelle par nanoLC-MS/MS des interactants de Trim24 purifiés par co-immunoprécipitation, nous sommes parvenus à identifier des protéines connues pour être impliquées dans des mécanismes d'élongation ou de répression de la transcription. Le travail en synergie de Trim24 et Trim33 pour le contrôle des gènes cibles de l'acide rétinoïque a également pu être établi.

Par ailleurs, l'étude différentielle d'échantillons soumis ou non à un excès d'acide rétinoïque a permis de mettre en évidence des protéines spécifiquement impliquées dans des sous-complexes tels que le protéasome, des facteurs de remodelage ou de réplication de l'ADN.

Puis, des tentatives de séparation des sous-complexes associés à Trim24, combinées à des expériences de Western Blotting sur des protéines identifiées au préalable, ont mis en évidence des différences importantes entre les complexes associées à une chromatine ouverte ou fermée.

Ces résultats, qui permettent d'aller plus loin dans la compréhension des mécanismes moléculaires régulateurs de l'état de quiescence des hépatocytes qui sont contrôlés par Trim24, n'ont pu être obtenus que grâce à la combinaison de stratégies analytiques finement optimisées pour l'analyse protéomique fonctionnelle et différentielle.

En effet, un important travail d'optimisation pour la préparation des échantillons a été réalisé par nos collaborateurs, notamment pour la purification des partenaires de Trim24 par la technique TAP et l'établissement des lignées cellulaires modèles. Les approches d'analyse par nanoLC-MS/MS ont parallèlement été finement préparées pour satisfaire aux exigences de sensibilité et de répétabilité indispensables à l'étude. Les mises au point technique et stratégique décrites au chapitre I ont apporté un bénéfice essentiel : citons notamment, l'optimisation des paramètres du système nanoUPLC - Synapt HDMS, les astuces permettant d'augmenter le contenu protéique identifié (comme la duplication des analyses et la concaténation des résultats) et la validation des identifications par la stratégie Target-Decoy.

Les premiers succès de cette collaboration sont en cours de mise en forme pour publication.

Dans le même temps, la spectrométrie de masse associée aux techniques de séparation des sous-complexes de Trim24 (BN-Page, double co-immunoprécipitation...) permettra d'aller plus loin dans l'identification et la compréhension des mécanismes de régulation de la transcription par Trim24. On peut alors espérer identifier de nouveaux gènes cibles de l'acide rétinoïque dont le niveau d'expression est inhibé et obtenir une meilleure compréhension des modulations épigénétiques de leur expression.



# Recherche de partenaires d'interaction du facteur de transcription Trim24

Guillaume Béchade<sup>1</sup>; Khalid Ouararhni<sup>2</sup>; Régine Losson<sup>2</sup>; Alain Van Dorsselaer<sup>1</sup>; Sarah Sanglier-Cianferani<sup>1</sup>

<sup>1</sup>Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC, UDS, CNRS, UMR 7178  
25 rue Becquerel 67087 Strasbourg Cedex 2, France

<sup>2</sup>Département de Génomique Fonctionnelle, IGBMC, UMR 7104  
BP 10142 67404 Illkirch Cedex

Contact  
gbechade@unistra.fr

## Trim24 : un suppresseur de tumeur dans le foie

«Tripartite motif-containing 24»  
 ► un cofacteur de transcription de l'ADN  
 ► interagit directement avec RAR et RXR  
 ► inhibe leur activité transcriptionnelle en présence d'acide rétinolique (AR)

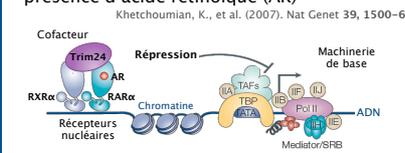


Fig. 1 | Répression de la transcription des gènes cibles de la voie de signalisation de l'acide rétinolique

Souris nulles pour Trim24 (Trim24<sup>-/-</sup>)  
 ► développement des tumeurs spontanées du foie

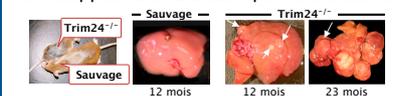


Fig. 2 | Formation de tumeurs hépatiques chez les souris Trim24<sup>-/-</sup>

## Objectifs

► Etude des partenaires d'interaction du cosuppresseur Trim24  
 ► Comprendre le rôle exact joué par Trim24 dans le contrôle de la transcription de la voie de signalisation de l'AR

## Stratégie analytique

- Établissement d'une lignée stable de cellules HELa
- Traitement ou non à l'AR
- Co-purification par affinités TAP-Tag
- Séparation des protéines sur gel SDS-PAGE
- Analyses nanoLC-MS/MS dupliquées
- Préparation des « peak-lists » Concaténation des répliques
- Identifications des protéines Stratégie Target-Decoy
- Identification de partenaires d'interaction potentiels

Fig. 3 | Analyse protéomique fonctionnelle par spectrométrie de masse pour l'étude des complexes multiprotéiques associés à Trim24

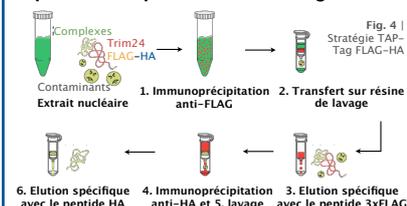
## Purification du complexe Trim24

Mise au point d'une lignée stable HeLa exprimant Trim24 fusionnée avec FLAG et HA (petits épitopes) en Cter ou Nter

► extraction des fractions nucléaire soluble et nucléaire insoluble (chromatinienne)

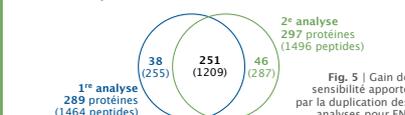
Echantillons de Trim24 analysés	
Trim24 C-ter FLAG HA	Trim24 N-ter FLAG HA
Extrait nucléaire insoluble (ENI)	Extrait nucléaire soluble (ENS)
Extrait nucléaire insoluble (ENI)	Extrait nucléaire soluble (ENS)

## Copurification par affinité TAP-Tag



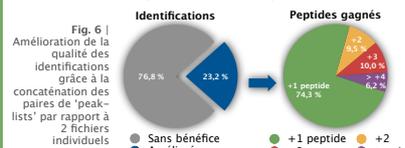
## Identification de partenaires d'interaction de Trim24 - Un interactant important : Trim33

Analyses nanoLC-MS/MS  
 ► ENI chromatinien : 330 protéines identifiées  
 ► ENS : 68 protéines identifiées



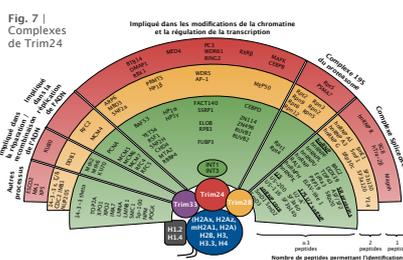
## Dupliquer les analyses : + 15 % identifications

► Fusionner les 'peak-lists' de duplicats : de nouveaux peptides améliorent le recouvrement pour 23 % des protéines



## Fusionner les 'peak-lists' : + recouvrements

## Caractérisation du complexe Trim24



## Identification de Trim24 et d'histones

## Identification de complexes impliqués dans

- le contrôle de la transcription
- le machinerie d'élongation de l'ARNm
- la préparation de la réplication
- la réparation ou recombinaison de l'ADN
- la modification de la chromatine

## Interaction entre Trim24 et Trim33

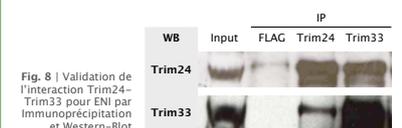
particulièrement mise en évidence

## ► Identification de Trim33

Protéine	Accession	Nb Peptides	% Recouvrement
E3 ubiquitin-protein ligase Trim33	Q9UPN9	45	56 %

► Interaction confirmée par l'identification de Trim24 dans les interactants de Trim33 purifiés par TAP-Tag FLAG-HA

► Validation de l'interaction par Wester-Blot



Conclusion  
 La majorité de Trim24 est associé à Trim33  
 La réciproque n'est pas vérifiée

## Complexes de Trim24 sous traitement à l'Acide Rétinolique

### Gel 1-D : cas traité et non traité à l'AR

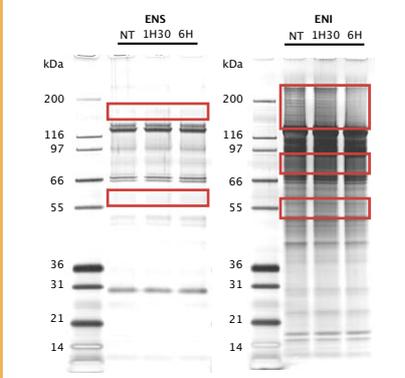


Fig. 9 | Gel 1-D 4-12 % d'échantillons non traités (NT) ou traités 1h30 ou 6h par un excès d'acide rétinolique. Encadrés : différences importantes.

Conclusion  
 Modifications importantes du complexe chromatinien (ENI) traité à l'AR

## Caractérisation du complexe Trim24 traité

### ► Récepteurs nucléaires de l'acide rétinolique

Protéine	Accession	Nb Peptides	% Recouvrement
RARα (ou RARβ)	P10276 (P10826)	1	2 %
RARγ	P13631	1	2 %
RXRα	P19793	3	16 %
RXRβ	P28702	11	24 %
LXRβ	P55055	7	18 %

Conclusion  
 Trim24 est en interaction avec des récepteurs de l'acide rétinolique (RAR, RXR et LXR)

### ► Complexes de régulation de la transcription

Le Complexe NuRD	Le Complexe WICH	Le Complexe MEDIATEUR	195 subunits
HDAC1, HDAC2	SNF2h, BAZ1B	MED4, MED24	S7 (Rpt1)
RbAp46, RbAp48	Le Complexe NORC		S4 (Rpt2)
Mi-2a, Mi-2β	SNF2h, BAZ2A		S6b (Rpt3)
MTA1, MTA2			S10b (Rpt4)
Le Complexe Sin3	Le Complexe BAF53	Elongin BC	S6a (Rpt5)
HDAC1, HDAC2	BAF53A, RUVBL1, SMARCA4/BRG1	EloC, EloB, Rbx1	S8 (Rpt6)
RbAp46, RbAp48			S2 (Rpn1)
Sin3	Le Complexe INO80	Le Complexe Cui4A	S1 (Rpn2)
Le Complexe MLL	ACTL6A, RUVBL1, RUVBL2	DDB1, Cui4A, Rbx1, CAND1	S3 (Rpn3)
RBBP5, DPY30, WDR5			P55 (Rpn5)
Le Complexe SET	Le complexe BASC		S12 (Rpn8)
WDR5, RBBP5	RAD50, MRE11A		S11 (Rpn9)
Le Complexe FACT	Le Complexe BAF		S13 (Rpn11)
Fact140, SSRP1	SMARCA2, SMARCA4/BRG1, ACTL6A/BAF53, SMARCC2/BAF170		S14 (Rpn12)
Le Complexe RSF			
RSF, SNF2h			

Conclusions  
 Chromatine condensée : Trim24 impliquée dans la répression de la transcription (NuRD - HDAC1)  
 Chromatine ouverte : Trim24 impliquée dans l'élongation de la transcription (Fact140, HP1γ)

## Fractionnement du complexe sur gradient de glycérol et analyse par Western Blot de protéines cibles pré-identifiées

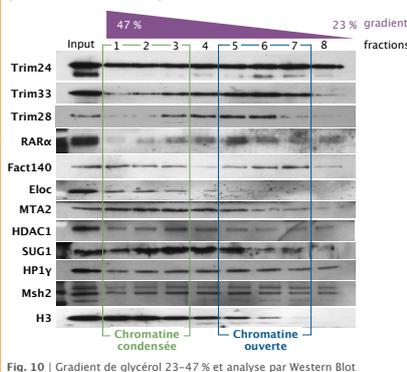


Fig. 10 | Gradient de glycérol 23-47 % et analyse par Western Blot



# **RÉSULTATS – 2<sup>E</sup> PARTIE**

## **CARACTÉRISATION FINE DE COMPLEXES PROTÉIQUES COVALENTS À LIAISON CYSTÉINE–CYSTÉINE**

### **Chapitre I**

Techniques et stratégies pour la caractérisation de complexes covalents par spectrométrie de masse

### **Chapitre II**

Mise en évidence de la formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroxydée par la sulfirédoxine

### **Chapitre III**

Étude de mécanismes de régénération de Methionine Sulfoxyde Réductases



## Chapitre I

# Techniques et stratégies pour la caractérisation de complexes covalents à liaison cystéine – cystéine par spectrométrie de masse

---

Pour les protéines, la liaison peptidique est pratiquement la seule liaison covalente possible entre les acides aminés qui les constituent. Il existe une exception courante à cette règle : la liaison disulfure. Il s'agit d'une liaison chimique soufre-soufre qui résulte d'un processus d'oxydation de cystéines impliquées dans des protéines. Les protéines formant des liaisons cystéine-cystéine peuvent être différenciées selon deux catégories :

- ▶ Les protéines pour lesquelles un pont disulfure entre cystéines est une partie stable de leur structure secondaire finale, qui détermine le repliement de leurs chaînes peptidiques et leur stabilité à l'état natif (non dénaturé) ;
- ▶ Les protéines pour lesquelles les cystéines alternent entre état oxydé et état réduit. Ce cycle d'oxydoréduction peut constituer un élément essentiel pour l'activité de protéines comme les enzymes ou peut impliquer l'activation ou la désactivation de protéines.

Les liaisons entre cystéines peuvent s'établir de manière intra- ou inter-protéique, formant dans ce dernier cas des complexes protéiques covalents.

Ce 1<sup>er</sup> chapitre passe en revue les techniques et stratégies mises au point au laboratoire pour l'analyse et la caractérisation de protéines et complexes protéiques covalents à liaison cystéine-cystéine. Dans le deuxième chapitre de cette partie, ces approches analytiques sont appliquées à l'étude de mécanismes d'oxydoréduction et à la caractérisation d'intermédiaires réactionnels particuliers.

# 1. La chimie des résidus cystéiques : oxydation, réduction, formation de complexes covalents à liaison cystéine-cystéine

Parmi les 20 acides aminés, la cystéine joue un rôle particulier du fait de sa chaîne latérale constituée d'un thiol. C'est ce thiol – ou groupement sulfhydryle (–SH) – qui est souvent impliqué dans des réactions enzymatiques en tant que nucléophile et qui est susceptible de s'oxyder pour former par exemple un dérivé disulfure (ou cystine) qui joue un rôle structural important pour beaucoup de protéines.

## 1.1. Les oxydations impliquant des cystéines

### 1.1.1. Cas d'un résidu cystéique isolé

Une cystéine peut se trouver sous trois degrés d'oxydation différents (Table 1, en haut). L'état d'oxydation 0 du soufre correspond à la forme acide sulfénique. Le résidu, très réactif, est peu stable. Il peut subir une réduction pour former avec un autre thiol un pont disulfure ; ou au contraire subir des oxydations supplémentaires jusqu'aux degrés d'oxydation du soufre +II (acide cystéine sulfinique) et +IV (acide cystéine sulfonique). Ces espèces sulféniques et sulfoniques sont généralement qualifiées de formes « suroxydées »

*In vivo*, la réduction enzymatique de l'acide cystéine sulfinique a récemment été découverte<sup>1</sup>. L'acide cystéine sulfonique par contre est considéré comme correspondant à un état d'oxydation irréductible.



Espèces à une cystéine				
Formule	CysSH	CysSOH	CysSO <sub>2</sub> H	CysSO <sub>3</sub> H
Degrés d'oxydation	-II	0	+II	+IV
Nomenclature	Thiol	Acide sulfénique	Acide sulfinique	Acide sulfonique

Espèces à deux cystéines			
Formule	CysS-SCys	$  \begin{array}{c}  \text{O} \\     \\  \text{CysS}-\text{SCys}  \end{array}  $	$  \begin{array}{c}  \text{O} \\     \\  \text{CysS}-\text{SCys} \\     \\  \text{O}  \end{array}  $
Degrés d'oxydation	-I/-I	+I/-I	+III/-I
Nomenclature	Pont disulfure (ou cystine)	Pont thiosulfinate	Pont thiosulfonate

Table 1 | Les différentes formes oxydées de la cystéine et le degré d'oxydation de son atome de soufre.

### 1.1.2. Oxydation impliquant deux cystéines : les liaisons disulfure, thiosulfinate et thiosulfonate

#### ▸ Liaisons disulfures

Lorsque plusieurs cystéines sont en jeu, l'espèce réactive qu'est l'acide cystéine sulfénique s'engage rapidement dans des réactions biochimiques. En premier lieu, elle réagit avec un thiolate (RS<sup>-</sup>) pour former un

<sup>1</sup> Woo, H. A., et al. (2003). Science 300, 653-6.

pont disulfure entre cystéines (Table 1, en bas). La réduction d'un pont disulfure est possible via la réaction avec un autre thiolate par échange de pont disulfure.

▸ **Liaisons thiosulfonates**

Les peroxydes ( $R-O-O-R'$ ), comme le peroxyde d'hydrogène ( $H_2O_2$ ) qu'on trouve *in vivo*, sont des oxydants très forts capables d'oxyder les ponts disulfures en disulfure S-monoxyde ou pont thiosulfinate (Figure 1). Un thiosulfinate peut également être formé à partir de 2 acides sulféniques. Le soufre qui porte l'atome d'oxygène (soufre sulfinyl) s'y trouve sous un degré d'oxydation de +I, l'autre (soufre sulfényl) sous un degré d'oxydation de -I.

Les liaisons thiosulfonates sont peu stables et très réactives vis-à-vis des espèces nucléophiles<sup>2</sup>.

▸ L'attaque par un thiolate se fait sur le soufre sulfényl de la liaison thiosulfinate et conduit à la libération d'un acide sulfénique et d'une espèce pont disulfure mixte.

▸ Par contre, l'hydrolyse du thiosulfinate a lieu par attaque d'un résidu  $HO^-$  sur le soufre sulfinyl conduisant à la libération d'un thiol et d'un acide sulfonique.

*In vivo*, la formation d'un intermédiaire thiosulfinate résultant de la réaction entre un thiolate et un acide sulfénique activé a été proposée comme hypothèse dans le mécanisme de réduction des acides sulfoniques<sup>3</sup>. L'étude de ce mécanisme et la caractérisation de l'intermédiaire thiosulfinate suggéré sont traitées dans le 2<sup>e</sup> chapitre de cette partie.

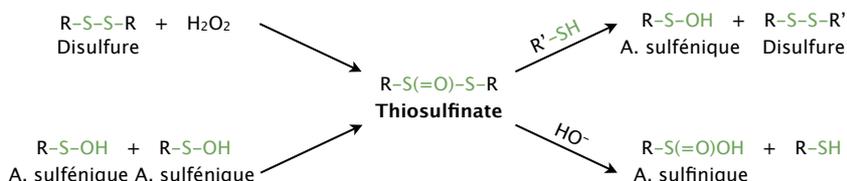


Figure 1 | Réactions de formation et de réduction des espèces thiosulfonates.

▸ **Liaisons thiosulfonates**

Les liaisons thiosulfonate sont issues de l'oxydation d'un intermédiaire thiosulfinate ou de la réaction d'un acide sulfénique et d'un acide sulfonique. Les nucléophiles attaquent uniquement au niveau du soufre sulfényl (Figure 2). L'attaque par thiolate conduit à la formation d'un acide sulfonique et d'un pont disulfure, alors que l'hydrolyse de l'espèce produit un acide sulfonique et un acide sulfénique.

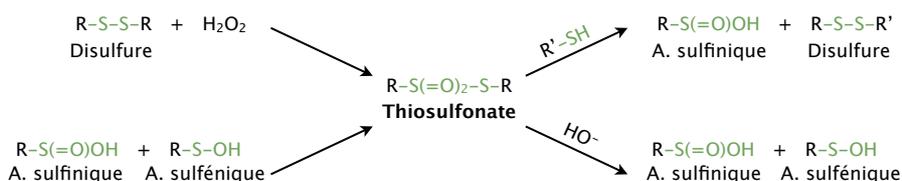


Figure 2 | Réactions de formation et de réduction des espèces thiosulfonates.

## 1.2. Les fonctions biologiques de l'oxydoréduction des cystéines

### 1.2.1. Structuration secondaire des protéines

Les ponts disulfures entre cystéines jouent un rôle important dans le repliement et la stabilité de nombreuses protéines, notamment celles sécrétées dans le milieu extracellulaire. Cette stabilisation peut être liée à plusieurs phénomènes :

<sup>2</sup> Nagy, P., et al. (2007). *J Org Chem* 72, 8838-46.

<sup>3</sup> Biteau, B., et al. (2003). *Nature* 425, 980-4.

- Un pont disulfure peut maintenir ensemble deux parties d'une chaîne peptidique et induire une topologie repliée de la protéine. En d'autres termes, il diminue l'entropie de la forme dépliée de la protéine ;
- Un pont disulfure peut constituer le noyau d'un coeur hydrophobe pour une protéine repliée : des résidus hydrophobes proches peuvent se rassembler autour du pont disulfure et autour les uns des autres par des interactions hydrophobes :
- En relation avec les deux premiers points, un pont disulfure lie deux segments d'une chaîne peptidique, augmente la concentration locale de résidus protéiques et diminue la concentration locale en molécules d'eau. Puisque l'eau est susceptible d'attaquer les liaisons hydrogènes entre les groupements amides participant à la structuration secondaire d'une protéine, cette dernière peut être stabilisée dans l'environnement proche d'un pont disulfure.

### 1.2.2. Les cystéines oxydées comme intermédiaires catalytiques

Outre le rôle joué par les résidus cystéiques dans la structuration de certaines protéines, les systèmes mettant en jeu des réactions d'oxydoréduction de thiols sont impliqués dans la régulation de divers processus biologiques comme la réponse au stress oxydant ou la transduction de signaux. Ces contrôles sont assurés par une famille d'enzymes – les thiol oxidoréductases – qui engagent leurs résidus cystéiques actifs dans des mécanismes catalytiques d'oxydoréduction<sup>4</sup>. Parmi ces enzymes par exemple, les peroxyrédoxines (Prx) et les méthionine sulfoxyde réductases (Msr) permettent la réduction des espèces réactives de l'oxygène (comme H<sub>2</sub>O<sub>2</sub>)<sup>5</sup> et des méthionines oxydées, respectivement. Des mécanismes impliquant ces deux enzymes sont étudiés dans les chapitres II et III de cette partie.

## 1.3. Les difficultés liées à l'analyse de protéines à liaisons cystéine–cystéine

D'une manière générale, la spectrométrie de masse est une méthode de choix pour déterminer la nature et la position de liaisons cystéine–cystéine inter- ou intra-protéiques.

En analyse protéomique classique – lorsque l'objectif est d'identifier le contenu protéique total d'un échantillon – les protéines sont soumises à un traitement de réduction et d'alkylation permettant de bloquer les résidus cystéiques. Cette « normalisation » permet : (i) de faciliter l'accès de la protéase lors de la digestion en évitant le repliement des protéines pontées à l'état natif ; (ii) d'améliorer le rendement d'extraction du gel 1-D les dipeptides à cystéine en les réduisant (les peptides pontés – plus gros – étant plus difficiles à extraire) ; (iii) de faciliter l'identification par les moteurs de recherche des peptides à cystéines en les indiquant comme potentiellement alkylés (les moteurs étant incapables de traiter d'autre information que les séquences protéiques linéaires présentes dans les banques).

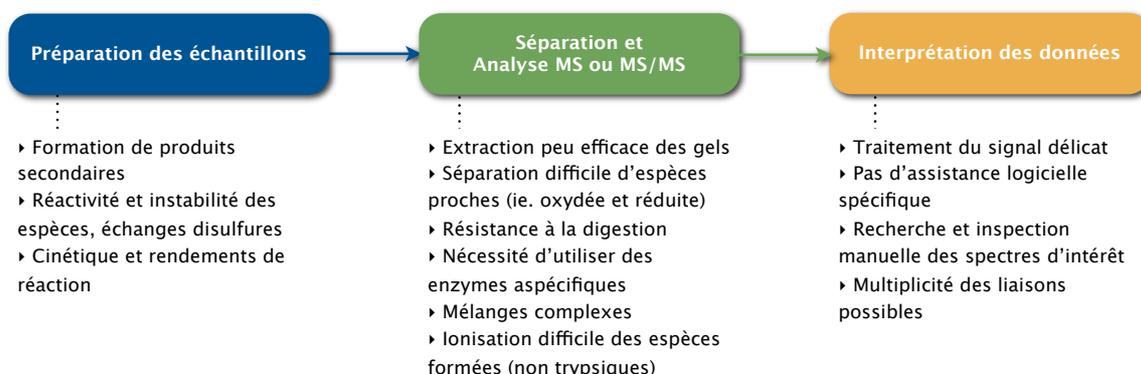
Ce traitement détruit toute chance de recueillir une quelconque information sur les liaisons cystéine–cystéine ou les protéines liées. Au contraire, lorsque l'objectif est de caractériser les ponts inter- ou intra-protéiques, le plus grand soin doit être pris pour maintenir les liaisons intactes et rendre l'information accessible. Un nombre important d'écueils doit alors être contourné (Figure 3) :

- La préparation des échantillons doit être spécifiquement adaptée pour chaque étude. Notamment dans le cas de l'étude de réactions enzymatiques, la réactivité des espèces sauvages impose des stratégies originales pour isoler et stabiliser les intermédiaires réactionnels à étudier.
- Des conditions expérimentales particulières doivent être mises au point pour maintenir les espèces pontées intactes, en évitant leur réarrangement pendant la mesure, pour assurer leur séparation et faciliter leur analyse en permettant par exemple la discrimination entre des espèces proches (réduite vs oxydée).

<sup>4</sup> Fomenko, D. E., et al. (2008). *Mol Cells* **26**, 228–35.

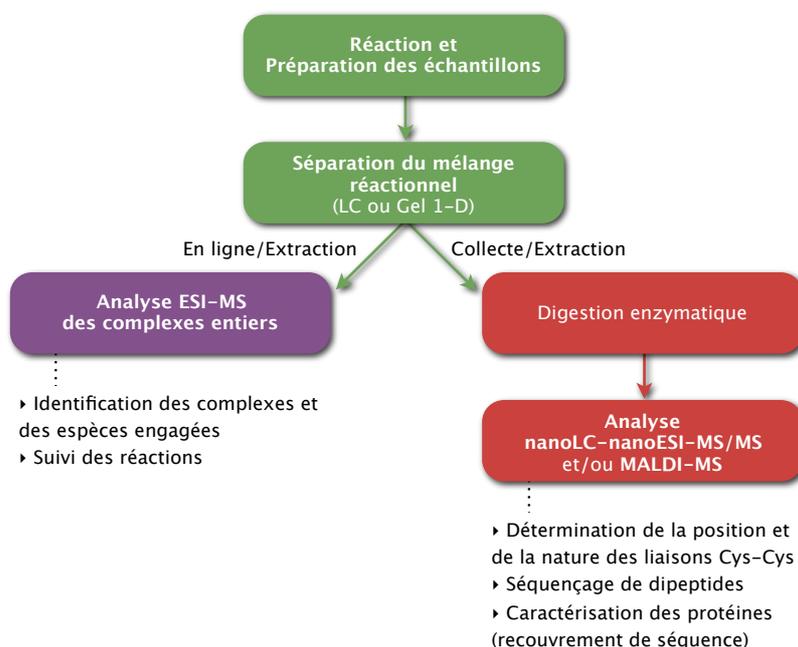
<sup>5</sup> Wood, Z. A., et al. (2003). *Science* **300**, 650–3.

▸ Loin d'un travail de routine, l'exploitation des données d'analyse par spectrométrie de masse (MS ou MS/MS) doit généralement être réalisée manuellement, car l'offre d'outils bioinformatiques pour l'étude des liaisons entre cystéines est encore très limitée. Le traitement du signal (lissage, exportation) doit par ailleurs être finement ajusté pour conserver l'intégrité des informations.



**Figure 3** | Difficultés rencontrées lors de la caractérisation de complexes covalents par spectrométrie de masse.

La suite de ce chapitre propose une série d'outils et de stratégies expérimentales pour l'analyse de complexes covalents dans le cadre de l'étude de mécanismes enzymatiques d'oxydoréduction (Figure 4). En première approche, la séparation et l'analyse de complexes entiers sont proposées. Elle permet d'identifier la nature des complexes et les espèces en jeu dans une réaction catalytique. Ensuite, pour déterminer la nature et la position des liaisons cystéine–cystéine, la caractérisation des complexes peut être réalisée en mettant en oeuvre des stratégies alliant la digestion enzymatique des complexes à l'analyse LC–MS/MS et MALDI–MS des mélanges peptidiques obtenus.



**Figure 4** | Stratégie de caractérisation de complexes covalents dans le cadre de l'étude de réactions enzymatiques d'oxydoréduction.

## 2. Analyse de complexes covalents à liaison cystéine–cystéine entiers produits lors de réactions enzymatiques d'oxydoréduction

### 2.1. Préparation des échantillons pour l'analyse de complexes covalents à liaison cystéine–cystéine

Lors du suivi d'une réaction enzymatique d'oxydoréduction, toutes les espèces formées ne sont pas forcément connues. Certaines peuvent correspondre à des réactions secondaires ou annexes, d'autres peuvent résulter de l'hydrolyse d'espèces instables. Dans tous les cas, le mélange réactionnel n'est jamais trivial et son analyse directe est impossible : une étape de mise au point des conditions réactionnelles pour favoriser la formation et l'accumulation des complexes à caractériser est nécessaire ; et la séparation des réactifs et des différents produits de la réaction doit être optimisée.

#### 2.1.1. Optimisation de la formation des complexes à analyser

- Évaluer la cinétique de formation des intermédiaires réactionnels

La première étape consiste à déterminer les conditions idéales de formation des complexes à analyser. La température, la durée, le pH de la réaction et la concentration des réactifs doivent être adaptés.

Dans le cas d'intermédiaires réactionnels instables et rapidement détruits, cette étape est déterminante. Pour l'étude du mécanisme de réduction des peroxyrédoxines suroxydées via un intermédiaire thiosulfinate (cf. chapitre 2 de cette partie), les premières analyses d'échantillons préparés quelques jours avant ont montré que la cinétique d'hydrolyse du complexe thiosulfinate était rapide : la réaction a alors dû être réalisée fraîchement sous contrôle de temps et de température, arrêtée par acidification du milieu et analysée immédiatement.

Par ailleurs, les conditions doivent être définies pour que la réaction puisse conduire à une formation maximale de complexe en évitant la subsistance d'excès de réactifs et la formation d'espèces secondaires.

- Bloquer l'intermédiaire à analyser, éviter les réactions secondaires

Dans un mécanisme d'oxydoréduction, les intermédiaires réactionnels qu'on cherche à caractériser peuvent être difficiles à isoler. Ceux-là sont souvent rapidement détruits par l'attaque d'autres résidus cystéiques impliqués dans les étapes de réactions suivantes. Par ailleurs, des réactions parasites avec des résidus cystéiques libres et ne jouant pas de rôle dans la réaction principale peuvent perturber ou rendre moins efficace la formation de l'intermédiaire à caractériser.

Pour pallier ces perturbations, la réaction enzymatique peut être bloquée par ajout d'un agent alkylant (iodoacétamide, biotine) qui permet d'inactiver les cystéines non engagées dans des liaisons inter- ou intra-protéiques. Toutefois, l'ajout d'agent alkylant peut produire des adduits aspécifiques et augmente le nombre d'espèces potentiellement en présence. Pour éviter une complexité inopportune, une autre option peut être préférée. Il s'agit du blocage des réactions par ajout d'acide (formique ou trifluoroacétique) pour interdire l'hydrolyse des complexes et le réarrangement des liaisons cystéine–cystéine qui peuvent avoir lieu à pH neutre ou basique.

L'utilisation de mutants pour lesquels certaines cystéines ont été substituées en alanine est également une possibilité stratégique intéressante. En effet, l'engagement de différents mutants en lieu et place des protéines sauvages permet éventuellement :

- d'éviter l'attaque du complexe par des cystéines réduites non impliquées dans le cycle catalytique en les mutant ;

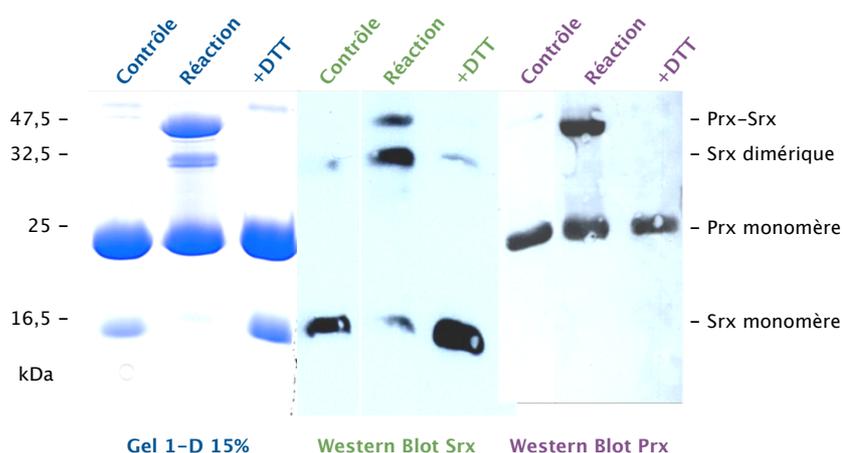
- d'accumuler l'intermédiaire réactionnel à caractériser en substituant les cystéines connues ou supposées être impliquées dans sa réduction dans la suite du cycle catalytique ;
- d'étudier les produits formés lorsque certaines cystéines sont mutées pour évaluer leur rôle dans le cycle catalytique.

### 2.1.2. Séparation de complexes covalents à liaison cystéine–cystéine entiers

Une fois les conditions de formation et d'isolation des complexes mises au point, les mélanges réactionnels sont susceptibles de contenir outre le reste d'excès de réactifs mis en jeu, plusieurs produits, sous-produits ou intermédiaires de réaction. Leur caractérisation impose leur séparation préalable.

#### ▸ Séparation sur gel 1-D et élution passive

La séparation de ces différentes espèces peut être réalisée sur gel 1-D. Couplé à une révélation par Western Blot, le gel 1-D permet généralement d'obtenir des informations sur les espèces formées et de suivre une réaction sous différentes conditions expérimentales. L'exemple en figure 5 présente la réaction de la sulfirédoxine C48A/C106A et de la peroxyrédoxine C171A suroxydée qui est étudiée au chapitre 2 de cette partie. En présence d'ATP-MgCl<sub>2</sub>, on observe la formation d'un hétérodimère entre la sulfirédoxine et la peroxyrédoxine et d'un homodimère de sulfirédoxine. Ils sont tous deux réduits en présence de DTT.



**Figure 5** | Gel 1-D 15 %, coloré au bleu de Coomassie, analysée par Western Blot des espèces formées lors de la réaction de la sulfirédoxine C48A/C106A et de la peroxyrédoxine C171A suroxydée en présence d'ATP-MgCl<sub>2</sub> pendant 2 min.

Si une séparation sur gel 1-D permet de suivre la réaction et d'évaluer la quantité de produit formé, elle ne renseigne pas sur le type de complexes en jeu, ni sur l'état d'oxydation des espèces suivies. Pour obtenir ces informations, les bandes d'intérêt peuvent être excisées et leur contenu analysé par spectrométrie de masse.

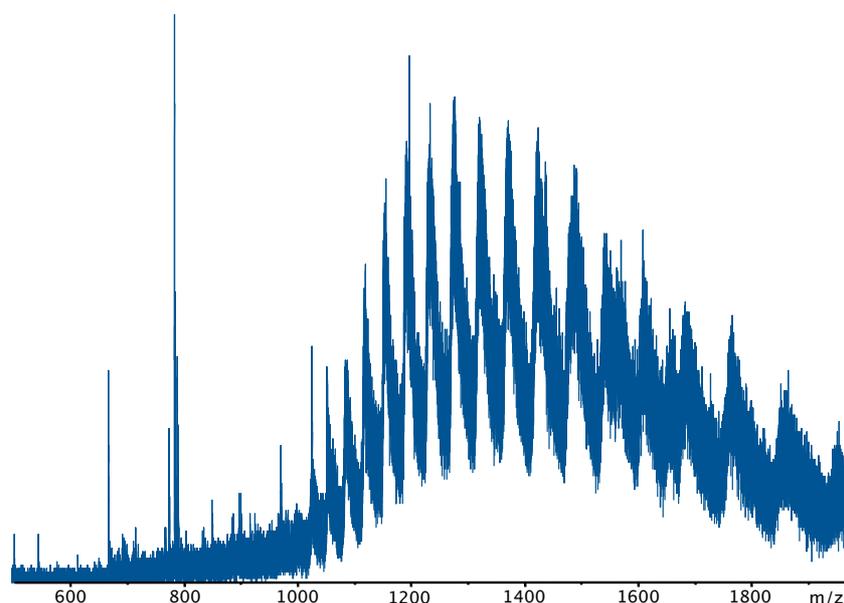
Si l'on vise à analyser les protéines entières, il a été montré que l'élution passive des protéines peut être réalisée directement (après une étape de décoloration) avec une solution d'acide formique/eau/2-propanol (1:3:2 v/v/v) sous agitation vigoureuse pendant 4 à 8 heures<sup>6</sup>. Théoriquement, pour des bandes suffisamment intenses (> 25 pmol), la réponse en spectrométrie de masse ESI-MS ou MALDI-MS est suffisante pour une mesure de qualité. Nous avons exploré cette voie. Cependant, les essais que nous avons réalisés sur nos échantillons n'ont pas été concluants pour plusieurs raisons.

- D'abord, parce qu'avec cette méthode la quantité de protéine extraite est souvent très limitée : le rapport signal sur bruit des analyses réalisées est trop important pour une mesure correcte.
- Deuxièmement, la technique SDS-PAGE n'est pas suffisamment résolutive pour séparer des espèces proches (comme une forme oxydée et une forme réduite par exemple). Les spectres ESI-MS obtenus par

<sup>6</sup> Cohen, S. L., et al. (1997). *Anal Biochem* 247, 257-67.

infusion directe des extraits présentent plusieurs espèces simultanément, ce qui cause des effets de suppression de signal et des chevauchements de spectres interdisant toute mesure fiable.

► Enfin, la forte concentration en acide formique de la solution d'extraction génère de multiples adduits formyl (+28 Da) sur les protéines, rendant la mesure difficile et les spectres de mauvaise qualité (Figure 6).



**Figure 6** | Spectre de masse ESI-MS obtenu après élution passive du contenu d'une bande de gel 1-D correspondant au complexe Srx-Prx. La mesure de masse est impossible en raison du rapport signal/bruit trop faible, de la cohabitation sur le même spectre d'espèces de masse proche éventuellement formylées.

Face à ces difficultés de mise en oeuvre et aux résultats médiocres obtenus pour des mesures de masse entière, nous avons exploré une autre voie : la stratégie d'analyse LC-MS.

► Séparation par chromatographie liquide analytique

Une séparation par chromatographie liquide présente le premier avantage de permettre une analyse MS en ligne grâce à la source d'ionisation électrospray. Parmi les colonnes aux dimensions adaptées à la séparation de protéines (Table 2), une colonne « *Narrowbore* » de 2,1 mm possédant une phase inverse C8 (colonne Vydac 208TP52 C8, 2,1 x 250 mm, 5 µm), permet la séparation de mélanges protéiques simples avec une bonne résolution chromatographique pour une charge maximale de 50 µg. Les débits de solvants sont par ailleurs compatibles avec une analyse MS en ligne : la plupart des sources électrospray requièrent un débit d'entrée entre 250 et 400 µL.min<sup>-1</sup>.

Colonne	Diamètre interne	Débit de solvant	Charge maximale
« <i>Microbore</i> »	1,0 mm	10 – 50 µL.min <sup>-1</sup>	10 µg
« <i>Narrowbore</i> »	2,1 mm	50 – 250 µL.min <sup>-1</sup>	50 µg
	3,2 mm	0,24 – 0,75 mL.min <sup>-1</sup>	100 µg
Analytique	4,6 mm	0,5 – 1,5 mL.min <sup>-1</sup>	200 µg

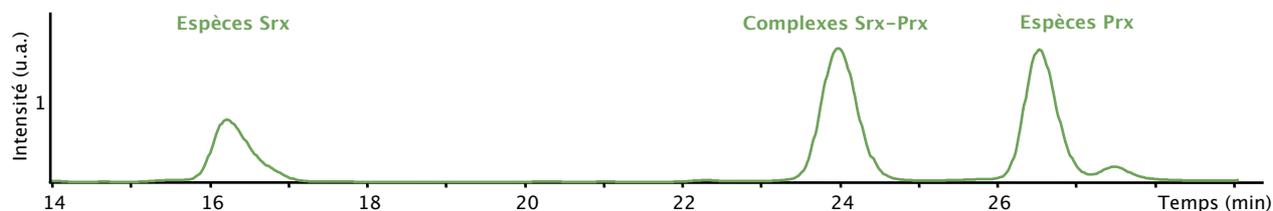
**Table 2** | Types de colonnes disponibles pour la séparation de mélanges protéiques.

Le laboratoire dispose d'une chaîne analytique de chromatographie liquide de la gamme HPLC 1100 Agilent Technologies, couplée à un détecteur UV et pouvant être mise en ligne avec nos spectromètres de masse. Nous avons mis au point des conditions de séparation de mélanges protéiques pour l'analyse directe de mélanges réactionnels enzymatiques. Le gradient (H<sub>2</sub>O + 0,1 % TFA / ACN + 0,08 % TFA) optimisé est

proposé en table 3. La figure 7 reprend l'exemple de l'analyse de la réaction de la sulfirédoxine C48A/C106A et de la peroxyrédoxine C171A suroxydée en présence d'ATP-MgCl<sub>2</sub>.

Temps (min)	0	2	32	33	38	39	54
% ACN	25	25	60	80	80	25	25

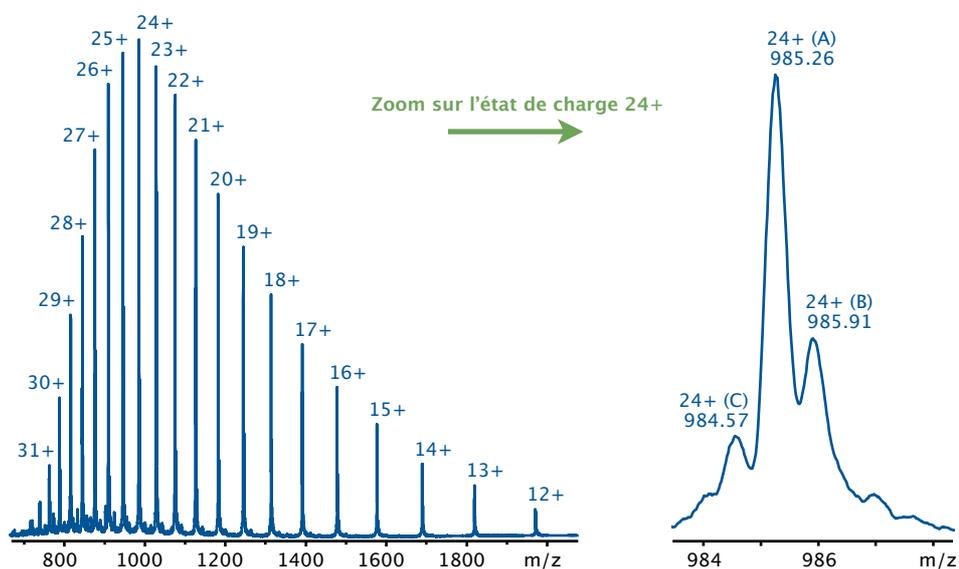
**Table 3** | Gradient d'acétonitrile optimisé pour la séparation d'un mélange protéique sur un système HPLC 1100 Agilent équipé d'une colonne Vydac C8, 2,1 x 250 mm, 5 µm.



**Figure 7** | Séparation par chromatographie liquide en phase inverse C8 des réactifs et espèces formées lors de la réaction de la sulfirédoxine C48A/C106A et de la peroxyrédoxine C171A suroxydée en présence d'ATP-MgCl<sub>2</sub> pendant 2 min.

Il est remarquable en premier lieu qu'alors qu'ils sont constitués des espèces Srx et Prx, les complexes produits sont élués à un temps intermédiaire. Le repliement des complexes imposé par les liaisons inter-protéiques rend probablement les interactions hydrophobes avec la phase stationnaire moins nombreuses ; ce qui implique une diminution du temps de rétention.

Les différentes familles d'espèces sont bien séparées. Mais en réalité chaque pic représente potentiellement plusieurs formes de la même protéine : les espèces oxydées et réduites d'une même protéine ne pouvant pas ou peu être séparées en phase inverse (Figure 8). Il faut donc être en mesure de discriminer des espèces coéluees sur un même spectre de masse.



	Masse moyenne théorique	Masse moyenne calculée
PrxSO	23 605,5	(A) 23 604,7 ± 0,6
PrxSO <sub>2</sub>	23 621,5	(B) 23 621,9 ± 0,6
PrxSO <sub>3</sub>	23 637,5	(C) 23 637,6 ± 0,3

**Figure 8** | Exemple de spectre de masse présentant trois espèces coéluees. Il s'agit des trois formes oxydées de la peroxyrédoxine C171A, dont la masse peut être déterminée aisément grâce à une résolution en masse classique (résolution spectrale théorique nécessaire < 1500 à 23 kDa).

## 2.2. Spectrométrie de masse de complexes covalents à liaison cystéine-cystéine entiers

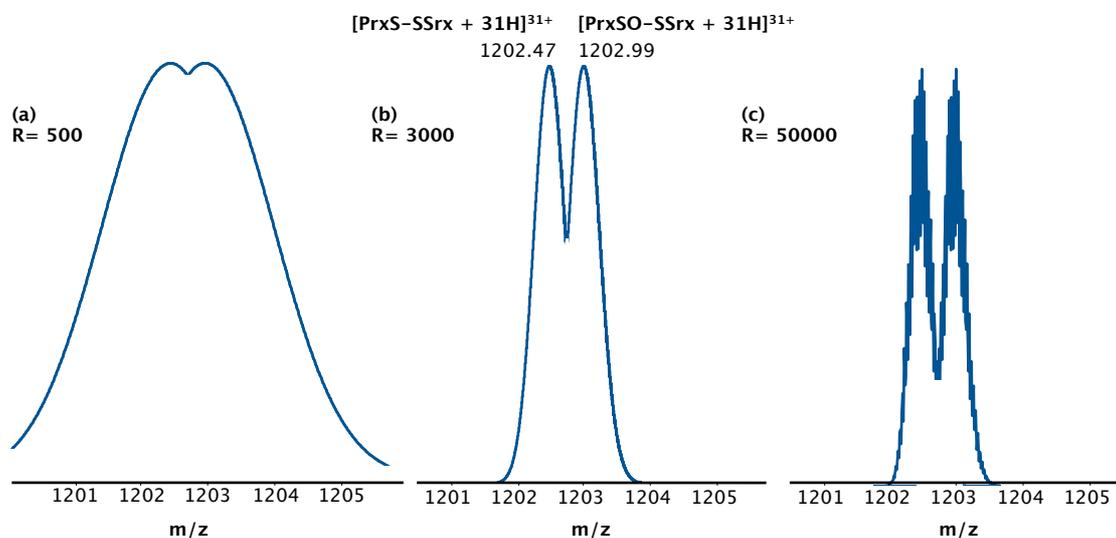
Une très haute résolution n'est pas exigée pour la mesure de masses moyennes de protéines entières (cf. Introduction Bibliographique, Chapitre I, point 1.2.3.). En effet, pour accéder à la masse moyenne calculée à partir des pics moléculaires multichargés d'une protéine, une résolution de 5000 est largement suffisante. En revanche, si plusieurs espèces sont présentes dans l'échantillon, leur discrimination peut devenir plus difficile.

Cela est particulièrement vrai dans le cas de protéines à cystéines réduites et/ou oxydées. Ces formes n'étant pas séparées lors d'une étape de chromatographie liquide en phase inverse, elles se trouvent coéluees et doivent être analysées simultanément. Pour qu'un résultat exploitable soit produit, la résolution doit être suffisante pour permettre de mesurer avec précision les masses des 2 composés sur chacun de leurs pics moléculaires multichargés du spectre ESI-MS. Par exemple, l'étude du mécanisme de réduction des peroxyrédoxines (Prx) par la sulfirédoxine (Srx) via un intermédiaire thiosulfinate (cf. chapitre II de cette partie) a requis la possibilité de discriminer les espèces présentées en table 4.

Espèces coéluees	Masse moyenne	Résolution théorique requise
<b>Prx sauvage réduite</b> (2 cystéines sous forme thiol)	13 722,7 Da	2 Da soit R > 6 900 à 15 kDa
<b>Prx sauvage oxydée</b> (2 cystéines engagées dans une liaison disulfure intra)	13 720,7 Da	
<b>Complexe disulfure PrxS-SSrx</b> Prx C171A - Srx C48A/C106A	37 246,1 Da	16 Da soit R > 2 350 à 37 kDa
<b>Complexe thiosulfinate PrxSO-SSrx</b> Prx C171A - Srx C48A/C106A	37 262,1 Da	

**Table 4** | Résolutions théoriques requises pour la mesure de masse de complexes coéluees impliqués dans le mécanisme de réduction des peroxyrédoxines (Prx) par la sulfirédoxine (Srx) via un intermédiaire thiosulfinate.

La figure 9 présente des simulations de résolutions FWHM à 500, 3000 et 50000 pour le mélange des complexes disulfure et thiosulfinate formés entre la peroxyrédoxine et la sulfirédoxine. On remarque qu'une résolution de 500 ne permet pas de discriminer les deux espèces coéluees (Figure 9a). Dans ce cas, la mesure de masse n'est pas possible, car les pics moléculaires multichargés du spectre ESI-MS obtenu sont confondus et le rapport m/z relevé pour ce pseudo-pic conduit à une mesure de masse inexacte.



**Figure 9** | Simulations de résolution FWHM à 500 (a), 3000 (b) et 50000 (c) pour l'état de charge 31+ d'un spectre ESI-MS des complexes disulfure PrxS-SSrx (masse moyenne théorique : 37 246,1 Da) et thiosulfinate PrxSO-SSrx (Masse moyenne théorique : 37 262,1 Da).

Une résolution de 3000 permet de séparer les pics moléculaires des deux espèces. Chacune des deux masses moyennes peut alors être mesurée de manière précise (Figure 9b). Pour de très hautes résolutions (au-delà de 50000, figure 9c), le massif isotopique de chaque enveloppe moléculaire apparaît. Dans une telle situation cependant, outre le fait de donner accès sans calcul à l'état de charge des pics moléculaires, l'information isotopique n'est pas valorisable. En effet, la mesure de la masse monoisotopique des deux espèces n'est pas possible puisque les deux massifs s'interpénètrent. Les spectres doivent alors être lissés pour faire apparaître les pics moléculaires à partir desquels la masse moyenne peut être calculée.

**Ces simulations montrent que rien ne sert de courir après toujours plus de résolution : il faut plutôt choisir un instrument adapté au problème traité et finement le paramétrer pour obtenir les performances appropriées.**

## 2.3. Conclusion

Dans l'objectif d'identifier et de caractériser les espèces formées lors de réactions d'oxydoréduction enzymatique, nous avons proposé en première approche analytique l'analyse des espèces entières par spectrométrie de masse ESI-MS. Les biologistes, malgré leurs hypothèses, ne savent pas toujours quelles sont les espèces potentiellement formées lors de ces réactions. Les stratégies et les optimisations techniques que nous venons de détailler sont une base essentielle pour préparer, purifier et identifier les intermédiaires en jeu. Dans un deuxième temps, il s'agit de tenter de les caractériser finement.

▸ Une préparation adaptée de l'échantillon est nécessaire pour bloquer les réactions et isoler les produits d'intérêt. Cela peut être réalisé :

- en acidifiant le milieu ou en ajoutant des agents alkylants pour bloquer les thiols libres ;
- en utilisant des mutants pour lesquels des cystéines spécifiques sont substituées.

▸ La séparation et la purification des espèces composant le mélange réactionnel peu être réalisée :

- sur gel 1-D, leur analyse par Western Blot permettant une évaluation rapide des espèces en jeu. Elles peuvent être extraites du gel par élution passive, mais la mise en oeuvre est délicate et la technique peu sensible.
- par chromatographie liquide. Reproductible, elle permet une analyse ESI-MS en ligne. Toutefois, la LC n'est pas toujours en mesure de séparer les formes proches d'une même espèce (les états oxydés et réduits d'une protéine par exemple). Elles se retrouvent donc sur le même spectre de masse où elles doivent être discriminées.

▸ La résolution d'un spectromètre de masse est une de ses caractéristiques essentielles. Mais les instruments les plus résolus – et les plus chers – ne sont pas forcément les plus adaptés. Etre en mesure d'accéder à l'information isotopique est inutile lorsque les massifs moléculaires de protéines sont enchâssés. Reste que pour certaines applications, un effort d'optimisation des paramètres instrumentaux est indispensable pour obtenir une résolution permettant de discriminer des espèces coélues.

### 3. Une combinaison de stratégies complémentaires pour la caractérisation fine de complexes covalents à liaison cystéine-cystéine.

Dans le cadre de l'étude de réactions enzymatiques d'oxydoréduction, la stratégie d'analyse de complexes protéiques covalents à liaison cystéine-cystéine entiers par ESI-MS permet d'identifier les espèces engagées et le type de liaison formée. Cependant, elle est incapable de fournir une quelconque information sur la position des ponts inter-cystéiques. Un effort de caractérisation fine des complexes doit donc être réalisé pour indiscutablement valider les hypothèses émises grâce à la première approche.

Cette caractérisation implique une étape de digestion des complexes isolés et purifiés, suivie d'une analyse des peptides générés par spectrométrie de masse MALDI-MS ou nanoLC-MS/MS. Deux objectifs simultanés doivent être atteints :

- Obtenir un recouvrement de séquence maximal des protéines analysées pour caractériser finement l'enchaînement d'acides aminés et vérifier la présence d'éventuelles modifications.
- Vérifier l'existence des dipeptides inter- ou intra-protéiques et caractériser la position et la nature des liaisons cystéine-cystéine formées.

Devant la diversité et la complexité des cas, il n'existe pas de stratégie directe et universelle pour la caractérisation des complexes covalents. Mais nous avons évalué tout un panel d'approches complémentaires dont chacune doit être adaptée à l'objet de l'étude (Figure 10). La fin de ce chapitre propose de détailler les quelques options stratégiques mises au point au laboratoire.

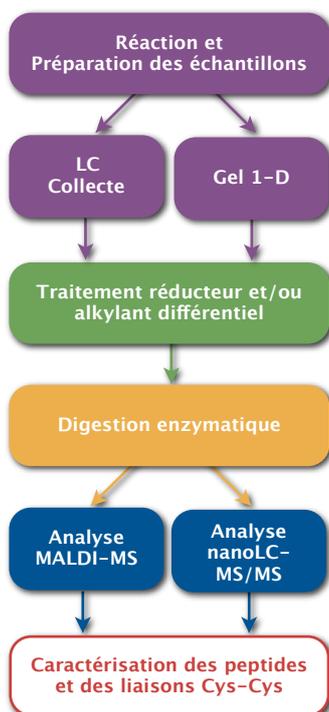


Figure 10 | Options stratégiques permettant la caractérisation fine de complexes protéiques covalents.

## 3.1. Préparation et traitement des échantillons

### 3.1.1. Préparation des échantillons

Pour l'analyse d'intermédiaires réactionnels rédox, les complexes d'intérêt dont la formation a été optimisée peuvent être séparés sur gel 1-D ou par chromatographie liquide comme décrit au point 2.1. Toutefois, la caractérisation impliquant une étape de digestion, les deux techniques ne présentent pas les mêmes limites que lors de l'analyse de complexes entiers. De nouvelles contraintes sont imposées à ces étapes de séparation préalables.

#### ▸ Séparation sur gel 1-D

Les éventuels traitements de réduction-alkylation et la digestion enzymatique peuvent être réalisés directement *in gel*. L'extraction des peptides générés constitue une étape limitante, notamment en ce qui concerne les dipeptides : plus gros et ramifiés, leur extraction est moins efficace et leur concentration finale dans l'échantillon peut être faible. Pour faciliter l'extraction des gros peptides du gel, quelques règles simples peuvent être observées :

▸ Les bandes de gel 1-D excisées sont finement découpées pour augmenter les surfaces d'échange gel/liquide.

▸ Une double extraction des peptides est plus efficace : la première est réalisée pendant une heure sous agitation vigoureuse avec une solution ACN/H<sub>2</sub>O (60:40) + 0,1 % acide formique ; la seconde avec une solution d'ACN pur jusqu'à déshydratation du gel.

#### ▸ Séparation par chromatographie liquide analytique

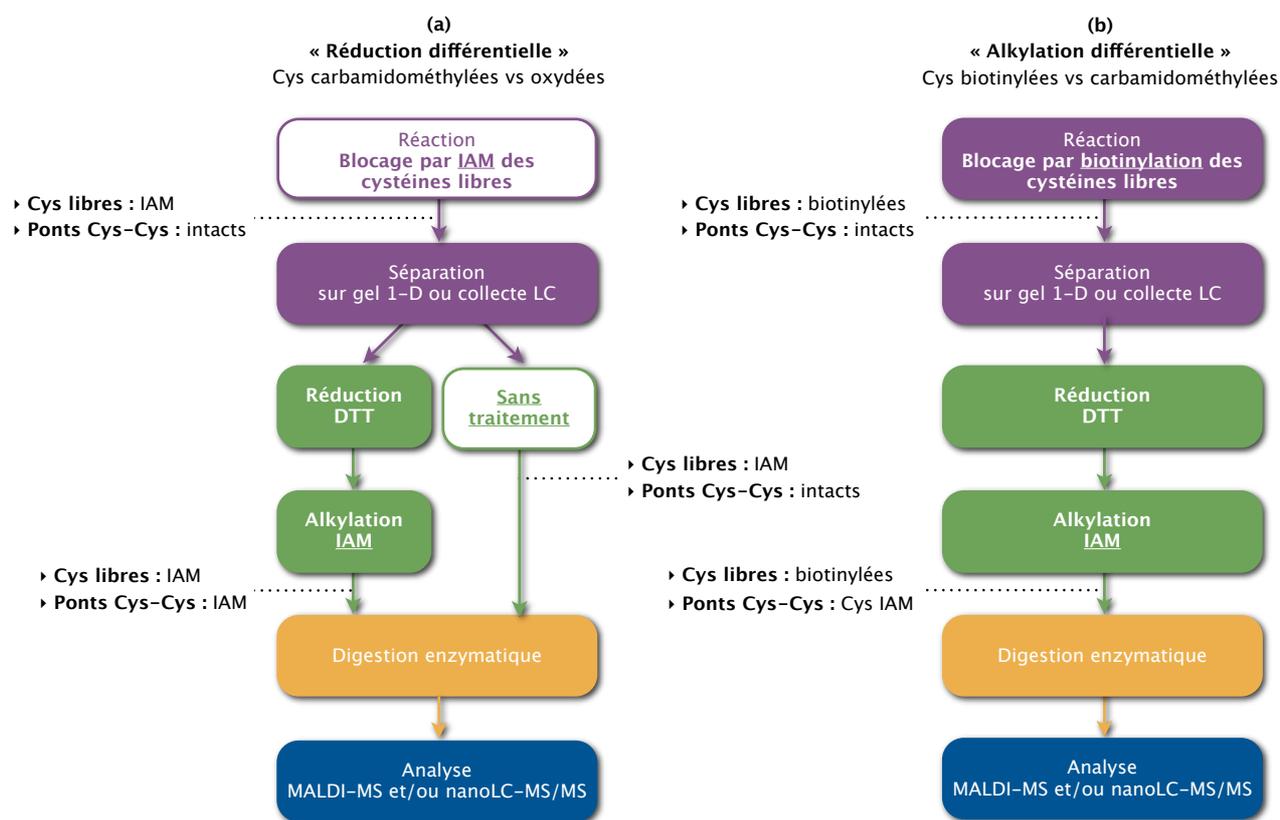
Couplée à un détecteur UV, elle permet la collecte sélective des pics élués : chaque espèce peut donc être récupérée successivement, diluée dans l'éluant de chromatographie (un mélange H<sub>2</sub>O/ACN + TFA). La plupart des protéases généralement utilisées pour la digestion sont intolérantes aux solvants organiques. L'excès d'acétonitrile doit donc être évaporé et le pH des échantillons doit être ajusté dans la gamme où l'activité de l'enzyme de digestion choisie est maximale (généralement à pH neutre). Ce traitement de pré-digestion est susceptible de modifier l'agencement des liaisons cystéine-cystéine<sup>7</sup> et peut conduire à la dégradation des complexes covalents peu stables (cf. chapitre 2 de cette partie).

### 3.1.2. Les traitements réducteurs et alkylants différentiels

Pour les protéines comportant plusieurs cystéines, les réarrangements de ponts disulfures ou les échanges thiol-disulfure qui peuvent avoir lieu sont susceptibles de conduire à une mauvaise assignation des liaisons cystéine-cystéine. Pour lever le doute qui peut subsister, des approches de réduction ou d'alkylation différentielle sont communément admises pour la détermination de la nature et de la position des liaisons cystéine-cystéine<sup>8</sup>. Elles s'appuient toutes deux sur la comparaison d'un même échantillon traité de manière différentielle.

<sup>7</sup> Gorman, J. J., et al. (2002). *Mass Spectrom Rev* **21**, 183–216.

<sup>8</sup> Seiwert, B., et al. (2008). *J Am Soc Mass Spectrom* **19**, 1–7.



**Figure 11** | Stratégie de « réduction différentielle » et « d'alkylation différentielle » (exemple avec la biotine et l'IAM) permettant la caractérisation de liaisons intercystéiques.

▸ Stratégie de « réduction différentielle »

Dans la première approche (Figure 11a), la réaction de formation des complexes est bloquée par ajout d'iodoacétamide (IAM) qui permet de carbamidométyler toutes les cystéines libres du mélange réactionnel. Après purification, la fraction est séparée en deux lots. L'un est soumis, avant l'étape de digestion enzymatique, à un traitement de réduction et d'alkylation à l'iodoacétamide ; alors que l'autre ne subit aucun traitement. Dans le premier lot, toutes les cystéines sont donc potentiellement carbamidométhylées alors que dans le deuxième, les liaisons cystéine-cystéine demeurent intactes. Ainsi, la comparaison des peptides identifiés pour chaque lot permet de déterminer les peptides qui ont subi une modification suite au traitement de réduction-alkylation. Ces peptides sont alors potentiellement impliqués dans un dipeptide à liaison cystéine-cystéine, qui peut être directement caractérisé dans le lot non traité.

Cette stratégie présente l'avantage majeur de permettre une caractérisation directe des dipeptides. Dans le cas d'une analyse nanoLC-MS/MS, la position et la nature de la liaison peuvent être prouvées. Toutefois, si la présence de multiples liaisons cystéine-cystéine est possible, des précautions doivent être prises pour limiter leur possible réarrangement lors des étapes de pré-traitement et l'analyse doit être réalisée rapidement.

▸ Stratégie « d'alkylation différentielle »

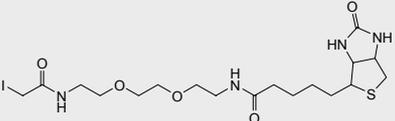
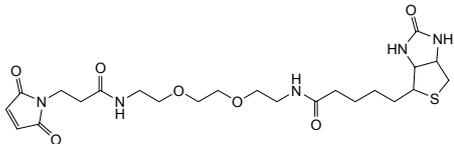
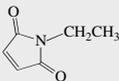
L'approche mettant en oeuvre une alkylation différentielle des cystéines impliquées ou non dans une liaison permet d'exclure tout risque d'échanges disulfure et thiol-disulfure (Figure 11b). Directement après la réaction, les protéines sont soumises à un traitement non réductif d'alkylation par N-éthylmaléimide (NEM)<sup>9</sup>, N- ou I-biotine<sup>10</sup>. Il s'agit d'agents alkylants spécifiques des cystéines (Table 5) qui permettent à cette étape

<sup>9</sup> Yen, T. Y., et al. (2002). J Mass Spectrom 37, 15-30.

<sup>10</sup> Yen, T. Y., et al. (2000). J Mass Spectrom 35, 990-1002.

de bloquer toutes celles qui ne sont pas impliquées dans une liaison inter- ou intra-protéique. Après séparation des produits de réaction, chaque lot (bande de gel 1-D ou fraction chromatographique collectée) est soumis à un traitement de réduction et d'alkylation à l'iodoacétamide. Nous avons vérifié que la réduction et la carbamidométhylation n'entraient pas en compétition avec l'alkylation préalablement réalisée et qu'elles permettaient la modification effective des cystéines engagées dans une liaison. Après digestion enzymatique, l'analyse de ces échantillons permet alors d'identifier les peptides qui présentaient des ponts cystéine-cystéine (on les retrouve carbamidométhylés) et les peptides qui contenaient des cystéines libres (retrouvées biotinylées ou avec un adduit NEM).

Cette approche présente 2 atouts : le premier est qu'elle limite les réarrangements des liaisons cystéine-cystéine grâce au blocage de tous les résidus avant la digestion ; le second est qu'une recherche automatique tenant compte des deux modifications possibles permet une identification directe des cystéines libres ou liées. L'alkylation différentielle reste cependant une méthode de caractérisation indirecte moins informative que l'inspection d'un échantillon non réduit : elle ne permet pas de déterminer les couples de cystéines lorsque de multiples ponts existent, et elle est incapable de renseigner sur la nature de la liaison formée.

Nom	Abréviation	Masse nette de l'adduit	pH de réaction	Structure
Iodoacetyl-LC-Biotin	<b>I-Biotin</b>	382,53 Da	8,3	
Maleimide-PEO-Biotin	<b>M-Biotin</b>	525,62 Da	6,5	
N-ethylmaleimide	<b>NEM</b>	125,05 Da	6,5	
Iodoacétamide	<b>IAM</b>	57,05 Da	7,5	

**Table 5** | Principaux agents alkylants des résidus cystéiques.

### 3.1.3. L'utilisation d'enzymes adaptées et complémentaires

#### ▸ Un problème de taille pour l'analyse de dipeptides tryptiques

Un des problèmes majeurs pour la caractérisation directe des liaisons cystéine-cystéine en conditions non réductrices est la taille importante des dipeptides formés lors d'une digestion classique par la trypsine. Intrinsèquement, un dipeptide – qui lie deux chaînes entre elles – est plus gros qu'un peptide classique. En outre, les liaisons inter-cystéiques peuvent perturber la digestion en gênant l'accès et en empêchant la coupure des sites de protéolyse qui leur sont proches. Dans les cas les plus défavorables, des dipeptides de plus de 20 résidus et de masse supérieure à 3000 Da peuvent être formés. Deux problèmes peuvent alors se poser :

▸ L'analyse de gros dipeptides peut être compliquée : ils sont difficilement extraits des gels 1-D et leur concentration peut être trop faible pour qu'un spectre de fragmentation de qualité soit acquis en nanoLC-MS/MS.

▸ Les gros dipeptides contiennent au moins et généralement plus de deux acides aminés basiques. Lors de l'ionisation électrospray, les dipeptides acquièrent des états de charges plus hauts et plus nombreux que les peptides classiques : plusieurs espèces de 3 à 7 fois chargées sont généralement observées. Cette dispersion en états de charge conduit inexorablement à une baisse de sensibilité pour l'analyse d'un dipeptide, le signal qui lui correspond étant partagé entre chacune de ses formes multichargées. Par ailleurs, en mode MS/MS, la présence de multiples résidus basiques interdit la protonation aléatoire du squelette peptidique et implique une fragmentation très pauvre du dipeptide<sup>11</sup>. L'information utile résultant de l'analyse peut alors être très limitée.

Pour pallier les difficultés rencontrées lors de l'analyse de dipeptides de taille importante, l'utilisation d'autres protéases que la trypsine a pu être évaluée (Table 6). Selon les cas, elles sont susceptibles de produire des dipeptides de taille plus adaptée à l'analyse nanoLC-MS/MS. Lorsque la séquence des protéines étudiées est connue, il est utile d'évaluer quelle est l'enzyme la plus adéquate. Dans certains cas, des digestions plus agressives combinant plusieurs protéases spécifiques ou non peuvent être mises en oeuvre.

L'utilisation d'autres enzymes que la trypsine est toutefois plus délicate. Dans le cas de cocktails enzymatiques ou d'enzymes aspécifiques, des mélanges peptidiques très complexes sont générés. Ils peuvent être significativement plus difficiles à interpréter. Par ailleurs, en mode MS/MS, les peptides non tryptiques sont généralement plus difficiles à fragmenter ou conduisent à des fragmentations plus atypiques, la répartition de leurs sites basiques étant plus aléatoire que chez les peptides tryptiques qui possèdent nécessairement une lysine ou une arginine du côté C-ter.

Enzyme	pH optimal	Ratio optimal enzyme/protéine (en masse)	Spécificité : Coupures préférentielles / secondaires
Trypsine	7,5 - 9	1/100 - 1/5	K, R (C-ter)
Chymotrypsine	7 - 9	1/200 - 1/20	Y, F, W / L, M, A, D, E (C-ter)
Asp-N	6 - 8,5	1/200 - 1/20	D, C (N-ter)
Glu-C (ou V8)	4 ou 7,8	1/100 - 1/20	E / D (C-ter)
Lys-C	7 - 9	1/200 - 1/50	K (C-ter)
Pepsine	1,8 - 2,2	1/50 - 1/100	F, L, W, Y, M, ... (C-ter)

Table 6 | Principales protéases utilisées pour l'analyse de protéines par spectrométrie de masse

▸ Un problème de digestion : le pH et l'hydrolyse.

La plupart des enzymes utilisées pour la digestion de protéines fonctionnent à un pH compris entre 6 et 9. Or, il est connu que le réarrangement et l'hydrolyse des liaisons cystéine-cystéine peuvent être favorisés à pH neutre<sup>12</sup>. Pour éviter ces phénomènes, les possibilités d'une digestion enzymatique à pH acide ont été évaluées. La pepsine est l'une des rares enzymes actives à un pH inférieur à 4. Malheureusement, sa large spécificité de substrat en fait une enzyme difficile à contrôler : les conditions de digestion doivent être préférentiellement évaluées et adaptées pour chaque cas. Et on retrouve avec la pepsine les mêmes difficultés d'analyse des peptides non tryptiques.

<sup>11</sup> Mikesch, L. M., et al. (2006). *Biochim Biophys Acta* **1764**, 1811-22.

<sup>12</sup> Gorman, J. J., et al. (2002). *Mass Spectrom Rev* **21**, 183-216.

## 3.2. Analyse et interprétation des données

### 3.2.1. Données MALDI-MS et données nanoLC-MS/MS

Dans l'objectif de caractériser finement les protéines impliquées dans des liaisons cystéine-cystéine, une analyse par empreinte peptidique massique MALDI-MS peut constituer une première approche pour :

- suivre et contrôler l'avancement de la digestion lorsqu'une enzyme originale est utilisée ;
- identifier les peptides liés présents dans le mélange peptidique et pour lesquels une analyse par spectrométrie de masse en tandem pourra être privilégiée ;
- comparer de manière directe les peptides présents ou absents dans le cas d'échantillons ayant subi ou non un traitement de réduction-alkylation.

L'analyse MALDI-MS de digests peut apporter une information précieuse pour l'identification des peptides présents dans l'échantillon, elle n'est généralement pas suffisante pour une caractérisation totale des complexes analysés. Tous les peptides ne sont d'ailleurs pas accessibles par MALDI-MS du fait des effets de suppression de signal que subissent les espèces les moins abondantes face aux ions les plus intenses.

L'approche par nanoLC-MS/MS met en oeuvre une dimension de séparation qui permet de décomplexifier l'échantillon. Surtout, le séquençage par spectrométrie de masse en tandem apporte une information supplémentaire qui permet de caractériser les peptides analysés, l'enchaînement de leurs acides aminés et les modifications éventuelles que ces derniers peuvent porter (alkylation, oxydation...). Cela facilite en outre l'identification des peptides lorsqu'une enzyme aspécifique est utilisée.

Pour obtenir un recouvrement de séquence optimal, il peut être bénéfique de combiner les données obtenues par MALDI-MS et nanoLC-MS/MS pour tirer parti de ces deux techniques d'ionisation différentes et qui peuvent apporter des informations complémentaires.

### 3.2.2. L'analyse de dipeptides par spectrométrie de masse en tandem

Pour indiscutablement caractériser les dipeptides et les peptides modifiés, la stratégie d'analyse nanoLC-MS/MS demeure incontournable. Elle n'en reste pas moins dans ce cas un défi analytique.

#### ▸ Analyse avec liste d'inclusion

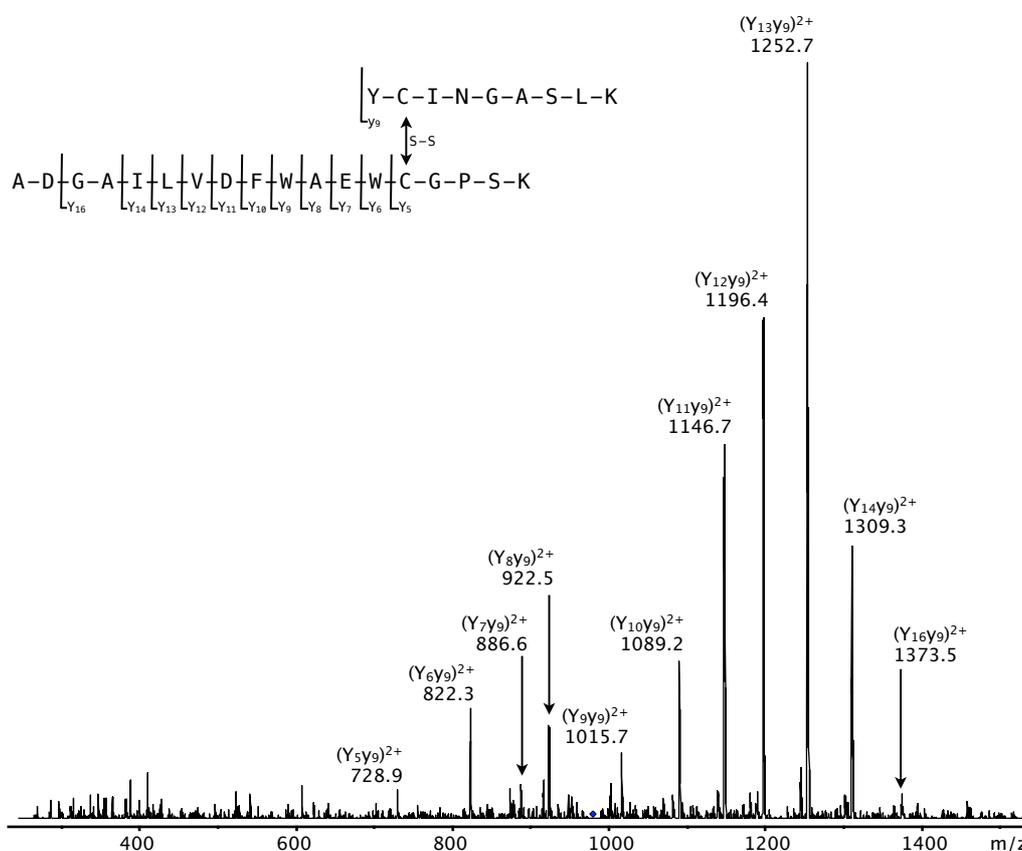
Comme il a été décrit plus haut, la grande taille des dipeptides implique généralement des états de charges plus diversifiés et plus élevés que les peptides classiques en ionisation électrospray. Ce phénomène a pour conséquence de diminuer la sensibilité de l'analyse vis-à-vis des dipeptides. Lors des cycles MS+MS/MS automatiques, les dipeptides sont généralement trop peu intenses par rapport aux autres peptides présents dans la fraction chromatographique analysée et ne sont pas sélectionnés pour la MS/MS. Pour y pallier, les logiciels de pilotage des spectromètres de masse proposent des options permettant de forcer la sélection d'ions ciblés en indiquant leurs caractéristiques (rapports  $m/z$  et éventuellement charges et temps de rétention) :

- Des listes de peptides préférentiels peuvent être proposées. Dans ce mode, les ions indiqués sont sélectionnés pour la MS/MS en premier lieu s'ils sont présents. Les autres ions sont sélectionnés sans priorité selon les règles classiques.
- Des listes d'inclusion de peptides peuvent être imposées. Cette option permet de sélectionner pour la MS/MS uniquement les ions indiqués. Les autres ions sont totalement ignorés.

Lors de l'utilisation de l'une ou l'autre de ces possibilités, notre expérience a montré qu'il était préférable de désactiver l'option d'exclusion dynamique des ions qui interdit, pendant une durée définie, la sélection des ions déjà analysés.

▸ Les spectres de fragmentation de dipeptides

Pour les raisons évoquées précédemment (taille et charge importante des dipeptides), les spectres de fragmentation de dipeptides sont généralement peu informatifs en mode CID classique (« *Collision Induced Dissociation* »). Toutefois, en mode positif, la fragmentation de dipeptides de taille importante conserve généralement la liaison cystéine-cystéine intacte, ce qui en permet la caractérisation<sup>13</sup>. Sur la plupart des spectres MS/MS de dipeptides, une seule des chaînes d'acides aminés est complètement séquencée, la fragmentation s'arrêtant au niveau de la cystéine impliquée dans la liaison. Restant intacts, les ions contenant le pont et les autres branches peptidiques sont généralement di- ou tri-chargés sur le spectre de fragmentation. C'est une différence majeure avec les spectres MS/MS de peptides classiques qui ne présentent généralement que des séries d'ions monochargés. A titre d'illustration, la figure 12 présente le spectre de fragmentation d'un dipeptide à pont disulfure étudié au chapitre 3. Le pont disulfure est maintenu dans la fragmentation et les ions fragments obtenus sont dichargés.



**Figure 12** | Spectre de fragmentation de l'ion  $[M+3H]^{3+}$  à 978,2 m/z, correspondant au dipeptide YCINGASLK – ADGAILVDFWAEWCGPSK à pont disulfure. Les fragments majeurs correspondent à la série  $(Y_n Y_9)^{2+}$  ( $n=5-16$ ).

Des modes de fragmentation et d'analyse alternatifs, comme la fragmentation ETD (« *Electron Transfer Dissociation* »)<sup>14</sup> ou le mode d'analyse négatif en fragmentation CID<sup>15</sup>, permettent d'observer la rupture de la liaison cystéine-cystéine et peuvent apporter des informations complémentaires. Ces stratégies n'ont toutefois pas été mise en oeuvre dans ce travail de thèse.

<sup>13</sup> Bilusich, D., et al. (2007). *Rapid Commun Mass Spectrom* **21**, 619–28.

<sup>14</sup> Mikesch, L. M., et al. (2006). *Biochim Biophys Acta* **1764**, 1811–22.

▸ Recherche de dipeptides dans les données nanoLC-MS/MS

Du fait de leurs particularités, les spectres de fragmentation de dipeptides requièrent une interprétation manuelle : les moteurs de recherche classiques ne sont pas capables d'identifier des peptides liés et d'exploiter l'information des spectres MS/MS qui peuvent présenter des fragments issus des 4 branches d'un dipeptide.

L'exploitation des données peut donc être difficile et demander des efforts importants, même lorsque la séquence des protéines en jeu et les différentes liaisons cystéine-cystéine possibles sont connues. La recherche s'effectue en 4 étapes :

▸ Digérer virtuellement la séquence des protéines selon les règles de coupure de l'enzyme utilisée et relever les peptides possédant une cystéine potentiellement impliquée dans une liaison. La digestion pouvant être aspécifique ou rendue difficile par la présence d'un pont cystéine-cystéine, il est préférable de considérer également les peptides présentant un site de coupure manqué d'une part et d'autre de la cystéine (« *missed cleavage* »). Pour chaque cystéine, 4 peptides différents sont donc à envisager.

▸ Etablir la liste de tous les dipeptides potentiellement contenus dans l'échantillon digéré. Le nombre de façons de former p liaisons cystéine-cystéine entre n résidus cystéiques est donné par la formule :

$$n! / [p! (n-2p)! 2^p]$$

▸ Calculer les rapports m/z des dipeptides à rechercher en considérant les états de charges couramment observés pour ces ions (de 3 à 7).

▸ Rechercher manuellement chacun des ions possibles et séquencer le spectre MS/MS de tous ceux qui figurent effectivement parmi les ions analysés.

Par exemple, pour déterminer quel pont disulfure est susceptible de se former entre 3 cystéines, il faudra rechercher parmi 240 ions potentiellement présents dans les données MS et MS/MS :

$$\begin{array}{l} 3 \text{ ponts possibles} \\ \times (4 \times 4) \text{ dipeptides à rechercher pour chaque pont} \\ \times 5 \text{ états de charge probables pour chaque dipeptide} \\ \hline = 240 \text{ ions} \end{array}$$

Des outils informatisés commencent à apparaître pour faciliter la recherche de liaisons cystéine-cystéine dans les données d'analyses nanoLC-MS/MS<sup>15</sup>. Mais encore à l'état de test et peu utilisés, leur fiabilité reste à démontrer.

<sup>15</sup> Xu, H., et al. (2008). J Proteome Res 7, 138-44.

### 3.3. Conclusion

Une méthode de caractérisation directe et complète de complexes protéiques covalents n'existe pas. Pour parvenir à détailler finement leur structure primaire en acides aminés, établir l'existence de liaisons inter-cystéiques, prouver leur position et leur nature, une combinaison de stratégies complémentaires doit être mise en oeuvre. Nous avons détaillé dans ce chapitre quelques options analytiques qui ont été évaluées au laboratoire.

- Des stratégies de réduction et d'alkylation différentielle sont mises au point pour déterminer facilement le nombre et la position des liaisons inter-cystéines d'un complexe. Elles permettent une interprétation simple des données, mais ne produisent que des preuves indirectes de l'existence de ponts entre protéines.
- L'utilisation et la combinaison d'enzymes adaptées à la génération de petits dipeptides favorisent leur analyse et améliorent le recouvrement de séquence du complexe caractérisé.
- L'analyse des digests produits par MALDI-MS permet une comparaison rapide des profils peptidiques d'échantillons ayant subi ou non un traitement de réduction-alkylation. Elle vient par ailleurs en complément de l'analyse nanoLC-MS/MS pour cumuler les informations d'identification.
- Le séquençage de dipeptides par spectrométrie de masse en tandem est considéré comme la méthode la plus informative : la position et la nature des dipeptides peuvent être prouvées directement. Les spectres de fragmentation de peptides à liaisons cystéine-cystéine demeurent cependant très difficiles à interpréter. La recherche et l'exploitation des informations doivent être réalisées manuellement.

## 4. Conclusion

La présence de résidus cystéiques dans une protéine est le fondement de la construction de liaisons inter- ou intra-protéiques. Celles-là ont une influence directe sur la structure tridimensionnelle des protéines et sont essentielles à leur fonction. C'est dans l'objectif de caractériser finement les complexes covalents protéiques qui se forment lors de réactions enzymatiques d'oxydoréduction que nous avons mis au point et évalué les possibilités offertes par deux stratégies complémentaires.

La première consiste en l'analyse de complexes covalents entiers par ESI-MS. Les performances résolutive de nos spectromètres de masse ont été optimisées pour satisfaire aux applications les plus pointues nécessitant la séparation sur le même spectre de masse d'espèces de masses proches.

La seconde implique l'identification de peptides à pont cystéine-cystéine dans les digests des complexes qu'on cherche à caractériser finement. Cette stratégie permet de prouver la position et la nature des liaisons en jeu, mais demeure plus difficile à mettre en oeuvre. Pour faciliter ces analyses, nous avons pu évaluer une série d'approches complémentaires qui doivent toujours être adaptées à chaque étude.

Nous proposons dans les deux chapitres qui suivent de mettre en valeur ces stratégies à travers deux applications. Elles s'intéressent à des mécanismes d'oxydoréduction enzymatiques qu'il s'agit d'expliquer grâce à la caractérisation d'intermédiaires réactionnels.



## Chapitre II

# Mise en évidence de la formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroxydée par la sulfirédoxine

---

Ce projet s'inscrit dans le cadre d'une collaboration avec l'équipe du Professeur Guy Branlant, du laboratoire « RN-RNP, Enzymologie Moléculaire et Structurale » (AREMS) de l'Université Henri Poincaré de Nancy. Sophie Rahuel-Clermont et Xavier Roussel ont étudié le mécanisme catalytique de la sulfirédoxine chez *S. cerevisiae*, impliquée dans la régénération des peroxyrédoxines suroxydées. Nous avons notamment montré par spectrométrie de masse que la réduction de la fonction acide sulfénique activée de la peroxyrédoxine, par attaque d'une cystéine catalytique de la sulfirédoxine, conduisait à un intermédiaire covalent thiosulfinate.

Ce chapitre propose une brève présentation du contexte biologique de l'étude. Puis, s'appuyant sur les développements méthodologiques pour la caractérisation des complexes covalents cystéine-cystéine, dont la mise au point a été détaillée au chapitre précédent, il met en lumière les possibilités offertes par une combinaison de techniques en spectrométrie de masse pour la caractérisation des espèces formées entre la sulfirédoxine et la peroxyrédoxine suroxydée.

# 1. Suroxydation et régénération des peroxyrédoxines

## 1.1. Le peroxyde d'hydrogène : un toxique et un messager cellulaire

### 1.1.1. L'activité du peroxyde d'hydrogène

Le peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>) a longtemps été considéré comme une espèce réactive de l'oxygène impliquée, directement ou non, dans la modification des macromolécules biologiques. Ce peut être en effet une molécule « non désirée » du métabolisme cellulaire, produite de manière non spécifique. Dans une cellule, H<sub>2</sub>O<sub>2</sub> peut réagir directement sur les biomolécules (ADN, lipides, protéines) ou être à l'origine d'autres espèces réactives (radical hydroxyle HO•, acide hypochloreux HOCl) et doit donc être éliminé efficacement pour éviter l'altération des fonctions cellulaires.

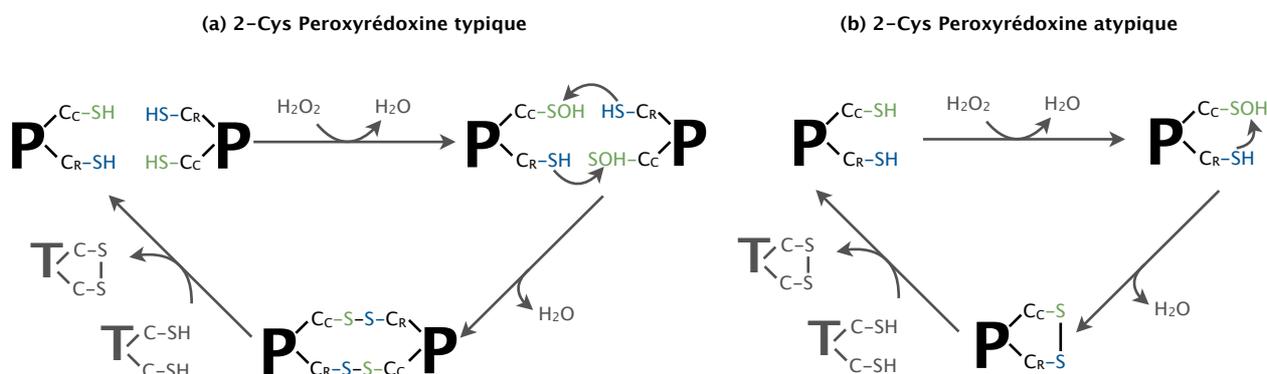
De récentes études ont montré que H<sub>2</sub>O<sub>2</sub> peut être aussi produit en faible concentration, de manière spécifique et contribue au bon fonctionnement cellulaire. Pour certaines voies de signalisation, sa production spécifique et la régulation de voies d'élimination en font un véritable messager cellulaire<sup>1</sup>.

### 1.1.2. Les peroxyrédoxines

Les peroxyrédoxines font partie de la famille des peroxydases, des enzymes impliquées dans l'élimination de H<sub>2</sub>O<sub>2</sub>. Cette famille est divisée en 3 classes selon leur mécanisme de réduction des peroxydes et en particulier selon leur mode de recyclage. On distingue :

- les peroxyrédoxines à 2 cystéines typiques et atypiques (2-Cys Prx), qui contiennent une cystéine activée dans leur site actif (dite cystéine catalytique C<sub>C</sub>), dont le recyclage est assuré par une cystéine de régénération (C<sub>R</sub>) ;
- les peroxyrédoxines à 1 cystéine, chez qui le recyclage est assuré par un système réducteur externe.

Dans le mécanisme de réduction de H<sub>2</sub>O<sub>2</sub> par les peroxyrédoxines à 2 cystéines, la cystéine activée est oxydée en acide cystéine sulfénique lorsqu'un peroxyde est réduit. Puis elle engage une liaison disulfure avec une cystéine de régénération de manière intra- ou inter-protéique (Figure 1). Afin de compléter le cycle enzymatique catalytique, les peroxyrédoxines sont réduites dans leur forme thiol active d'origine grâce par exemple à l'intervention d'enzymes thiorédoxines capables de réduire les ponts disulfures.

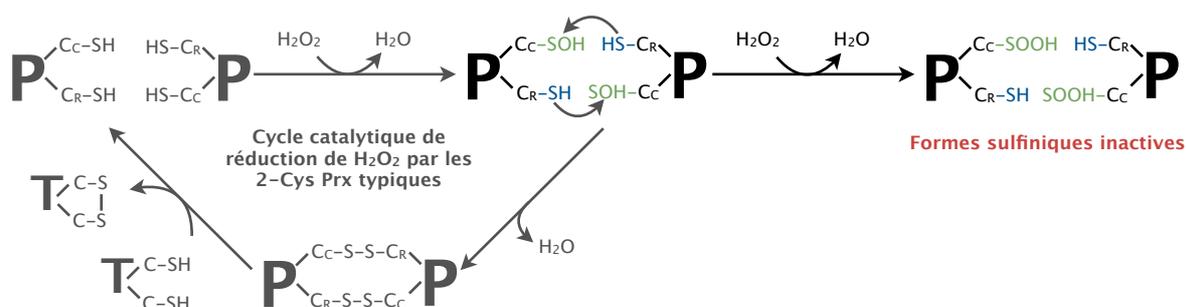


**Figure 1 | Mécanismes catalytiques simplifiés de la réduction du peroxyde d'hydrogène par les peroxyrédoxines à 2 cystéines typiques (a, mécanisme dimérique) et atypiques (b, mécanisme monomérique). (P : Peroxyrédoxine, T : Thiorédoxine.)**

<sup>1</sup> Wood, Z. A., et al. (2003). Science 300, 650-3.

### ▸ Suroxydation des peroxyrédoxines

Pour l'ensemble des 2-Cys Prx typiques des mammifères, de plantes et de levures, il a été montré que la cystéine catalytique pouvait être suroxydée sous forme d'acide sulfénique voir sulfonique<sup>2, 3</sup> et que l'enzyme était alors inactivée. Cette suroxydation a lieu lors de l'activité peroxydase des 2-Cys Prx eucaryotes. La réduction de l'acide sulfénique de la cystéine catalytique par la cystéine de recyclage nécessite des réarrangements structuraux plus ou moins lents au court desquels la cystéine catalytique peut être oxydée à de plus hauts degrés d'oxydation sous l'action de H<sub>2</sub>O<sub>2</sub> (Figure 2).



**Figure 2** | Inactivation des peroxyrédoxines à 2 cystéines typiques. La cystéine catalytique est suroxydée sous forme sulfénique par H<sub>2</sub>O<sub>2</sub> en conditions de stress oxydant important.

## 1.2. Réversibilité de l'oxydation d'une cystéine sous forme acide sulfénique

### 1.2.1. Les sulfirédoxines régulent l'activité des peroxyrédoxines

Alors qu'on pensait que la suroxydation des peroxyrédoxines à 2 cystéines sous forme sulfénique était irréversible *in vivo*, en 2003, des enzymes à activité sulfinyl réductase ont été découvertes chez *S. cerevisiae* : les sulfirédoxines (Srx)<sup>4</sup>. Les premières études ont montré :

- l'implication des sulfirédoxines dans le métabolisme cellulaire de H<sub>2</sub>O<sub>2</sub>, leur expression étant fortement induite par H<sub>2</sub>O<sub>2</sub> et les levures étant intolérantes à H<sub>2</sub>O<sub>2</sub> en l'absence de Srx.
- que les sulfirédoxines pouvaient être liées covalamment à certaines peroxyrédoxines à 2 cystéines.
- que les levures sauvages peuvent réduire les peroxyrédoxines à 2 cystéines suroxydées après exposition à H<sub>2</sub>O<sub>2</sub> ; alors que celles invalidées pour Srx n'en sont pas capables.
- que la complémentation en sulfirédoxine recombinante des levures invalidées pour Srx implique une restauration de la réduction des peroxyrédoxines suroxydées.
- la réduction de peroxyrédoxines suroxydées par la sulfirédoxine nécessite la présence d'ATP, de Mg<sup>2+</sup> ou Mn<sup>2+</sup>, et d'un système réducteur de type thiorédoxine.

Les sulfirédoxines contiennent une cystéine conservée chez de nombreux organismes eucaryotes, en position 84 chez *S. cerevisiae*. Son remplacement par une alanine abolit l'activité sulfinyl réductase, alors que la substitution des autres cystéines non conservées n'a pas (C106S) ou n'a que peu d'effet (C48S) sur la réduction des peroxyrédoxines à 2 cystéines suroxydées *in vivo*. Ces résultats suggèrent que la cystéine 48 possède une fonction essentielle dans le mécanisme catalytique de réduction.

<sup>2</sup> Rabilloud, T., et al. (2002). J Biol Chem 277, 19396–401.

<sup>3</sup> Wagner, E., et al. (2002). Biochem J 366, 777–85.

<sup>4</sup> Biteau, B., et al. (2003). Nature 425, 980–4.

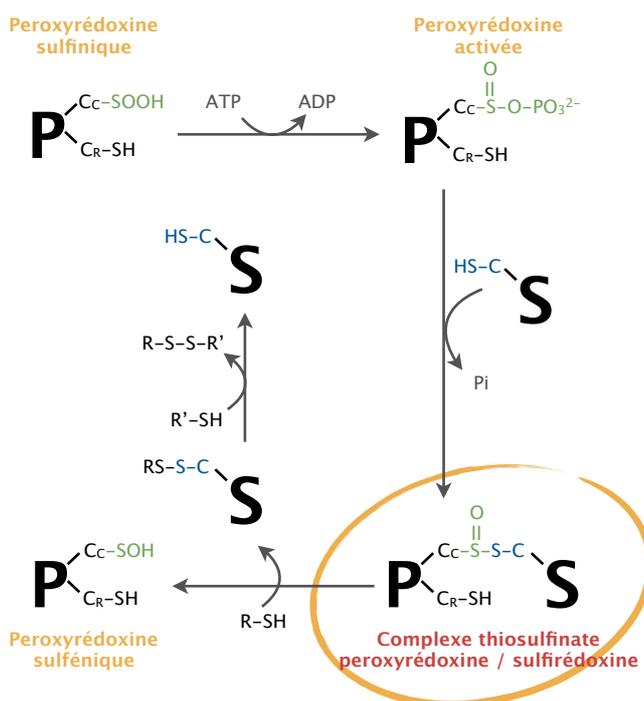
### 1.2.2. Les hypothèses formulées pour expliquer le mécanisme enzymatique de Srx pour la réduction des Prx suroxydées

▸ Le mécanisme de Biteau *et al.*

Les observations qui ont accompagné la découverte des sulfirédoxines ont permis d'émettre des hypothèses quant au mécanisme catalytique de réduction des peroxyrédoxines à 2 cystéines. Biteau *et al.* a proposé le mécanisme présenté en figure 3.

Comme c'est le cas pour un acide carboxylique, la réduction *in vivo* d'un acide sulfinique nécessite une activation chimique. La présence nécessaire d'ATP et de Mg<sup>2+</sup> a suggéré que l'activation était assurée par le transfert d'un groupement phosphate de l'ATP sur la fonction acide sulfinique. Cette activation aboutit à la formation d'un intermédiaire anhydride phosphoryle sulfinique qui subit l'attaque nucléophile de la cystéine catalytique de la sulfirédoxine, celle-là libérant un ion phosphate inorganique. On aboutit à un intermédiaire covalent entre la peroxyrédoxine et la sulfirédoxine, liées par un pont thiosulfinate. Comme il l'a été décrit au chapitre I, les liaisons thiosulfinate sont peu stables et sensibles à l'attaque nucléophile des thiols au niveau du soufre sulfénique de la liaison (ici celui de la cystéine catalytique de la sulfirédoxine). Cette attaque permet la libération de la peroxyrédoxine sous forme acide sulfénique et de la sulfirédoxine lié par un pont disulfure avec le réducteur. La sulfirédoxine est alors recyclée par la réduction de ce pont disulfure par un autre équivalent de réducteur (la thiorédoxine par exemple).

Dans l'hypothèse où se mécanisme serait vérifié, Srx présenterait double activité phosphotransférase et réductase.



**Figure 3** | Hypothèse de mécanisme catalytique de réduction des 2-Cys peroxyrédoxines suroxydée par la sulfirédoxine, proposée par Biteau *et al.*, 2003. (P : peroxyrédoxine, S : sulfirédoxine, R : réducteur à thiol, Pi : phosphate inorganique)

▸ Le mécanisme de Rhee *et al.*

En 2006, l'équipe de Rhee *et al.* proposa un autre mécanisme catalytique pour Srx<sup>5</sup> dans lequel elles possèdent une activité de phosphotransférase, mais où l'activité réductase est assurée par un autre réducteur à thiol. Dans cette hypothèse, ce réducteur engagerait lui même une liaison thiosulfinate avec la Prx.

<sup>5</sup> Rhee, S. G. (2006). Science **312**, 1882-3.

## 2. La stratégie analytique mise en oeuvre pour la caractérisation des complexes peroxyrédoxine–sulfirédoxine

### 2.1. Objectifs de l'étude

Ce projet vise à expliquer le mécanisme catalytique de la sulfirédoxine chez *S. cerevisiae* en validant l'hypothèse de Biteau *et al.* En particulier, il s'agit de démontrer l'activité réductase des sulfirédoxines. Pour cela, nous proposons la caractérisation la plus complète possible des espèces formées entre la sulfirédoxine et la peroxyrédoxine suroxydée grâce à une combinaison de techniques en spectrométrie de masse.

Trois étapes sont nécessaires pour atteindre l'objectif :

- ▶ Déterminer les conditions idéales pour observer l'intermédiaire thiosulfinate supposé ;
- ▶ Optimiser des conditions d'analyse LC-MS afin de caractériser la structure des multiples produits de réaction entre les espèces Srx et Prx ;
- ▶ Déterminer précisément la position du pont et de l'oxydation sur l'intermédiaire en optimisant la collecte des espèces, leur digestion et leur analyse par nanoLC-MS/MS.

Le travail de développement qui a été décrit au chapitre I a permis la mise au point de stratégies analytiques par spectrométrie de masse visant à caractériser des complexes covalents à liaison cystéine-cystéine. La suite de ce chapitre décrit l'application de ces approches analytiques à l'étude du mécanisme catalytique de réduction de la peroxyrédoxine suroxydée par la sulfirédoxine. Les résultats de ce travail ont été valorisés par la publication d'un article qui figure en fin de chapitre.

### 2.2. Contrôle de la réaction conduisant à la formation des complexes covalents Prx–Srx

Dans l'hypothèse de Biteau *et al.*, le complexe peroxyrédoxine–sulfirédoxine à pont thiosulfinate, s'il existe, est un intermédiaire réactionnel. Il s'agit de déterminer les conditions de réaction optimales pour favoriser la formation et l'accumulation du complexe en vue de l'isoler et de le purifier. Pour cela, les 2 aspects suivants ont été pris en compte :

- ▶ L'élimination des réactions parasites ou secondaires en bloquant les cystéines non réactives ou potentiellement impliquées dans la réduction de l'intermédiaire thiosulfinate.

Des mutants de la peroxyrédoxine (PrxC171A) et de la sulfirédoxine (SrxC48A/C106A) ne contenant que leur cystéine active ont pu être préparés. Les autres mutants possibles et les protéines sauvages ont également pu être étudiés (cf. publication au point 3.3.1. de ce chapitre). De plus, la réaction a été bloquée par ajout d'iodoacétamide (IAM) ou d'acide trifluoroacétique (TFA) pour éviter l'attaque des thiols libres des réactifs en excès.

- ▶ La mise au point de conditions de réaction favorisant la formation de l'espèce thiosulfinate.

Il est apparu qu'une température de plus de 30 °C ou une durée de réaction supérieure à 30 s favorisaient l'hydrolyse rapide du complexe thiosulfinate. La réaction a donc été systématiquement menée sous contrôle précis de température, puis bloquée (par acidification du milieu ou ajout d'un agent alkylant) et le mélange analysé immédiatement.

## 2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre

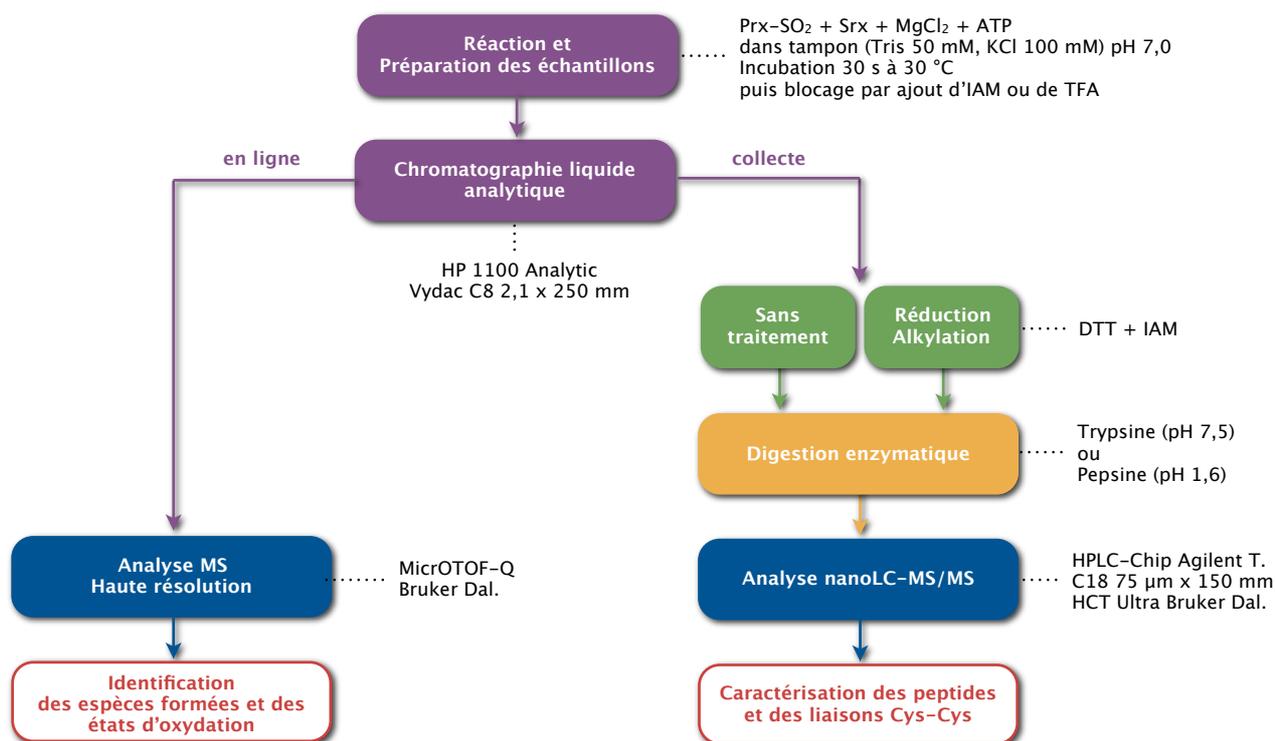
Au commencement du projet, sans recul sur la stabilité du complexe thiosulfinate potentiellement formé lors de la réaction, nous pensions pouvoir isoler l'intermédiaire purifié seul. Des approches d'analyse directe par ESI-MS après séparation des produits de réaction sur gel 1-D et élution passive ont été infructueuses (pour les raisons détaillées au chapitre I de cette partie). De même, les tentatives de caractérisation par nanoLC-MS/MS après digestion enzymatique *in gel* n'ont pas abouti : aucun dipeptide n'a pu être caractérisé.

Nous nous sommes alors tournés vers l'approche mettant en oeuvre une séparation par chromatographie liquide analytique du mélange réactionnel et deux approches complémentaires d'analyse par spectrométrie de masse ont été développées pour la caractérisation de l'intermédiaire thiosulfinate.

- La première vise à analyser, par chromatographie liquide analytique couplée à la spectrométrie de masse ESI-MS, les espèces formées (complexes, produits secondaires) lors de la réduction de la peroxyrédoxine suroxydée par la sulfirédoxine.

- La deuxième vise à caractériser finement chaque complexe pour valider la nature et la position des liaisons. Une stratégie par digestion enzymatique et analyse par nanoLC-MS/MS a été utilisée.

La figure 4 présente les chemins stratégiques empruntés et détaille les principales variables expérimentales utilisées.



**Figure 4** | Stratégies mises en oeuvre pour la caractérisation des intermédiaires de réaction lors de la réduction des peroxyrédoxines suroxydées par les sulfirédoxines.

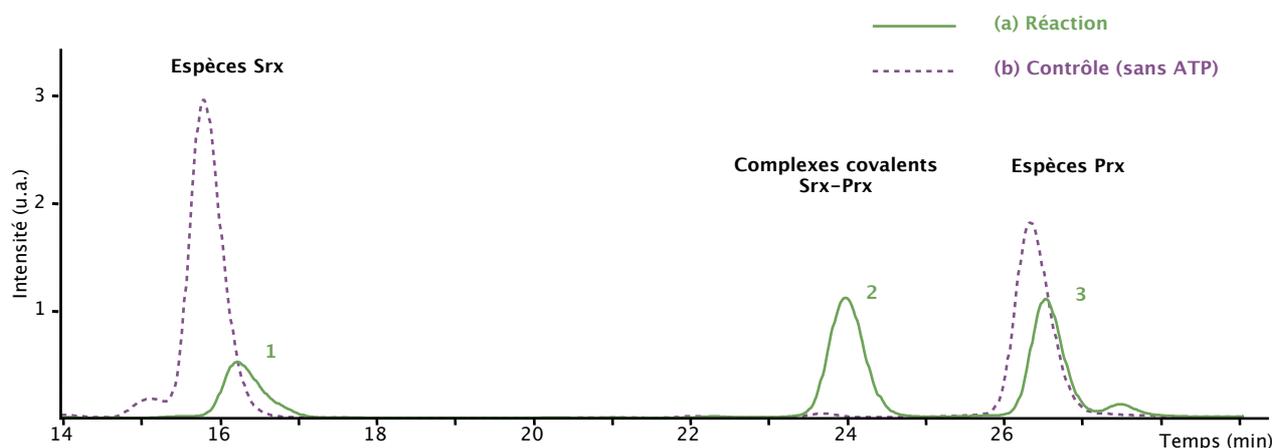
### 3. Résultats

Cette étude a fait l'objet d'une publication dans *The Journal of Biological Chemistry*, reproduite dans son intégralité à la fin de ce chapitre. Les points suivants détaillent les principaux résultats obtenus pour la caractérisation par spectrométrie de masse des complexes formés entre Srx C48A/C106A et Prx C171A.

#### 3.1. Analyse des complexes entiers par ESI-MS Haute Résolution

##### 3.1.1. Suivi par LC-MS de la réaction conduisant à la formation des complexes Srx-Prx

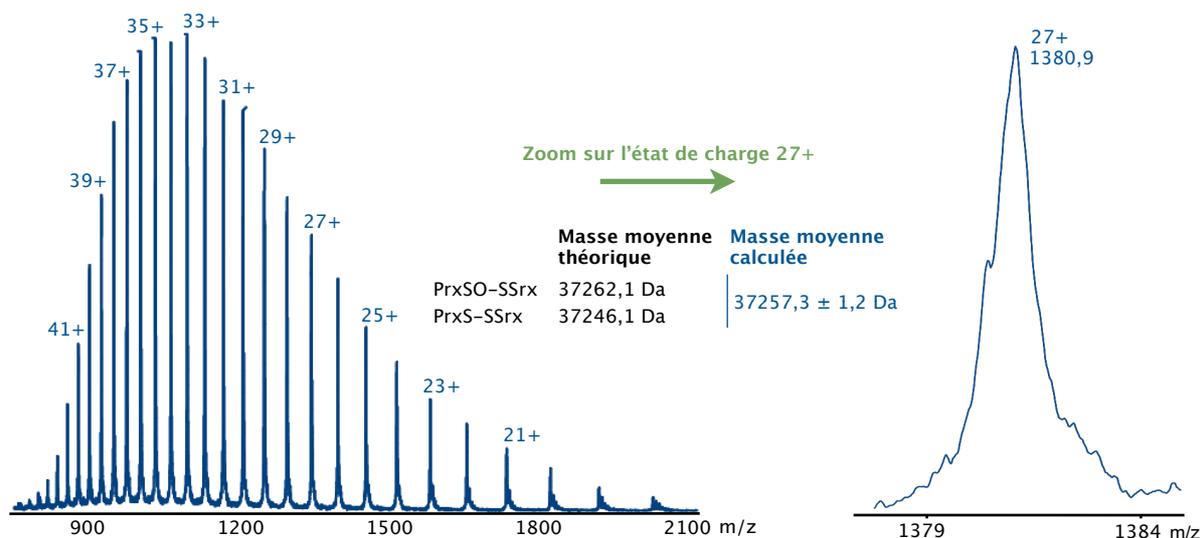
Le suivi par LC-MS de la réaction de la sulfirédoxine C48A/C106A avec la peroxyrédoxine C171A suroxydée en présence d'ATP a permis d'identifier les produits de la réaction. Des analyses de contrôle (sans ATP) ont conduit à l'identification des espèces en jeu (Figure 5) : alors que seuls des pics chromatographiques correspondant aux réactifs sont présents dans l'analyse de l'échantillon sans ATP, un troisième pic apparaît lorsque la réaction est réalisée.



**Figure 5** | Séparations par chromatographie liquide en phase inverse C8 : (a) des réactifs et espèces formées lors de la réaction de la SrxC48A/C106A et de la PrxC171A suroxydée en présence d'ATP-MgCl<sub>2</sub> pendant 2 min, (b) de la réaction de contrôle sans ATP.

Les premières analyses des complexes formés entre la peroxyrédoxine et la sulfirédoxine ont été réalisées sur notre instrument LCT (Waters Corp.) de type TOF. Les résultats obtenus avec cet instrument ont permis de soulever l'hypothèse que l'intermédiaire thiosulfinate à caractériser s'hydrolysait rapidement en complexe disulfure, les deux espèces étant vraisemblablement coéluées en chromatographie liquide analytique.

L'instrument, bien qu'ayant une résolution spécifiée de 5000, n'est pas capable de résoudre les deux espèces (Figure 6). La masse moyenne mesurée est une masse « intermédiaire » et les pics sont anormalement larges. En outre, les pics multichargés de plus haut m/z (correspondant aux états de charge de 19+ à 30+) présentent un pic naissant dans leur enveloppe. C'est grâce à ces observations qu'il a été possible de supposer la présence sur le même spectre de deux complexes de masse proche, sans qu'il n'ait été possible dans un premier temps de les séparer.



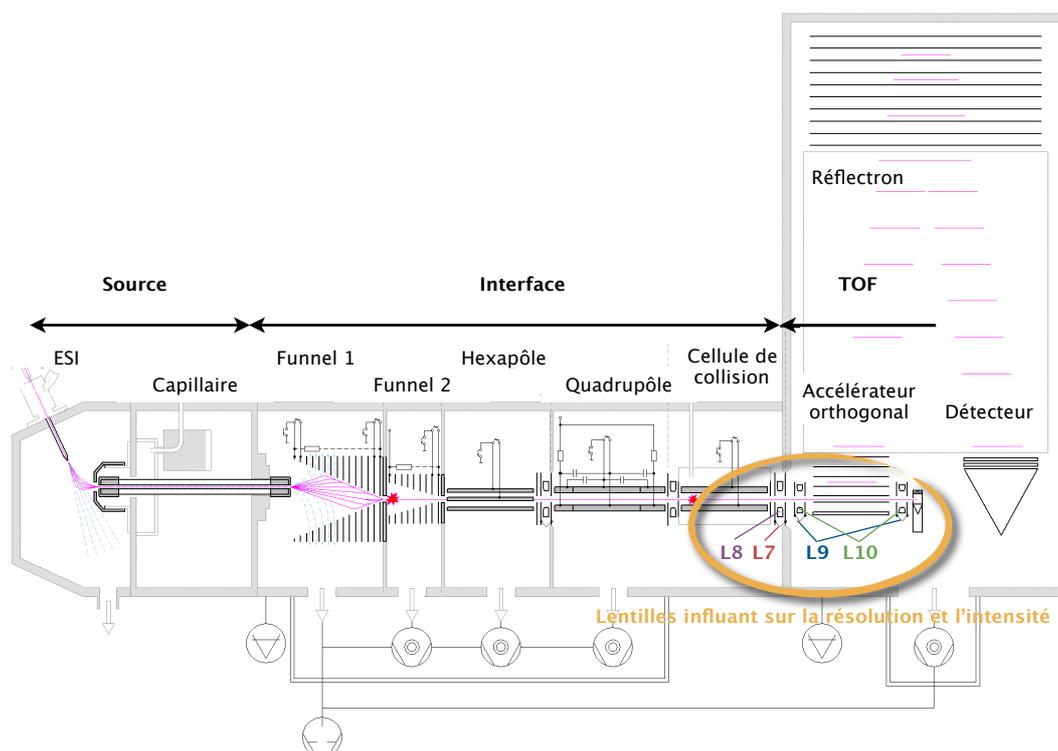
**Figure 6** | Spectre de masse ESI-MS, acquis sur LCT Waters Corp., des complexes disulfure PrxS-SSrx et thiosulfinate PrxSO-SSrx. La résolution de l'instrument ne permet pas de séparer les deux espèces présentes sur le même spectre : on devine un pic caché dans l'enveloppe des pics multichargés.

### 3.1.2. Optimisation de la résolution en masse sur le système microTOF-Q™ (Bruker Dal.)

En matière de pouvoir résolutif en spectrométrie de masse, les résultats expérimentaux ne sont souvent pas exactement conformes aux simulations (cf. chapitre I de cette partie, point 2.2.). La résolution nécessaire pour séparer deux espèces de masse proche est souvent sous-estimée. D'abord parce que les complexes coélués ne sont généralement pas présents dans les mêmes proportions. Si les deux pics n'ont pas la même intensité, une résolution supérieure est nécessaire pour qu'ils demeurent bien séparés : le pic le plus faible se trouvant en partie masqué par l'enveloppe plus large du pic le plus intense. Par ailleurs, selon la qualité de la désolvatation et de l'ionisation, les pics moléculaires de protéines entières ont tendance à s'élargir.

En 2008, le laboratoire s'est doté d'un nouveau spectromètre de masse à temps de vol de la gamme microTOF-Q Bruker Dal., dont la résolution spécifiée s'élève à 15 000. Pour permettre une mesure précise, sur le même spectre ESI-MS, de la masse des deux complexes supposés, un travail d'optimisation de la résolution spectrale a été nécessaire. Les paramètres d'analyse ont été finement ajustés afin d'atteindre une résolution maximale pour nos analyses.

A l'aide d'une solution de référence (TuneMix, Agilent Technologies) composée de molécules organiques stables, les valeurs de tension de certaines lentilles de focalisation des ions à l'entrée du tube de vol (au niveau de l'accélérateur orthogonal, figure 7) ont été optimisées. Le module d'optimisation des paramètres instrumentaux inclus dans le logiciel de pilotage de l'instrument a été utilisé. Il permet de suivre l'évolution de la résolution et de l'intensité pour une espèce déterminée en fonction de gammes de tensions appliquées aux lentilles de focalisation. Ce travail a été réalisé sur un standard monochargé de m/z 1221,99 pour les lentilles 7 à 10, indiquées comme ayant une influence importante sur la résolution et l'intensité du signal. Les résultats sont présentés en table 1.



**Figure 7** | Parcours des ions dans le spectromètre de masse microTOF-Q Bruker Dal. et désignation des lentilles influant sur la résolution et l'intensité du signal. Schéma en coupe extrait du manuel utilisateur.

Lentille	Influence majeure	Gamme de tension applicable (V)	Valeur optimale expérimentale (V)
L7	Intensité + résolution	[+8 ; +9]	+8
L8	Intensité	[-25 ; -15]	-16
L9	Résolution	[-2 ; +2]	+1
L10	Résolution	[-30 ; -20]	-27

**Table 1** | Optimisation de la résolution et de l'intensité du signal sur un standard de  $m/z$  1221,99 sur microTOF-Q Bruker Dal., grâce à l'ajustement des tensions appliquées sur les lentilles de focalisation 7 à 10.

Cette optimisation visant à obtenir les performances maximales de l'instrument a permis d'atteindre un **pouvoir résolitif (FWMH) de 17457** pour le composé standard monochargé à  $m/z$  1221,99. Des mesures réalisées sur des échantillons de référence comme la myoglobine bovine ont montré le gain de performances obtenu (Figure 8) : le massif isotopique de chaque pic multichargé se détachant des enveloppes moléculaires a pu être visualisé. Mais dans cette situation, les performances résolutes ne peuvent être évaluées précisément puisque la vallée entre chaque pic n'est pas suffisante pour mesurer une résolution FWHM. Cela illustre le caractère discontinu de l'information accessible en spectrométrie de masse (cf. chapitre I de l'introduction bibliographique, point 1.2.3.).

Toutefois, les mesures expérimentales réalisées sur nos échantillons biologiques ont montré qu'il était possible d'atteindre un pouvoir résolitif de près de 18000 pour un composé de 15 kDa et un pouvoir résolitif de près de 2800 pour un composé de 37 kDa.

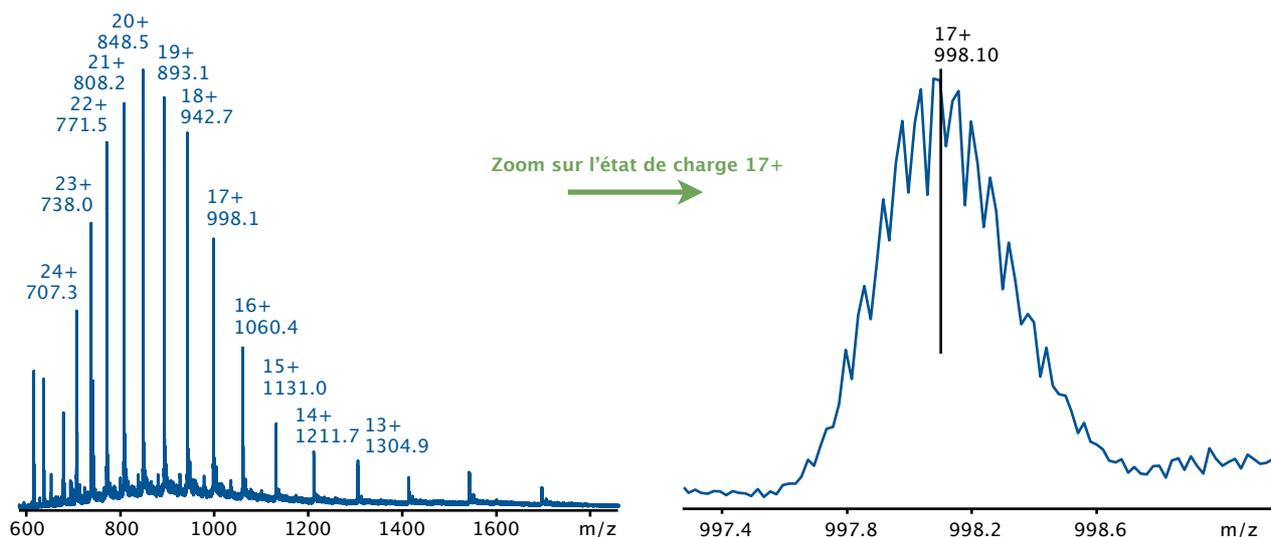


Figure 8 | Spectre de masse ESI-MS, acquis sur microTOF-Q Bruker Dal., de la myoglobine bovine ( $M_{th} = 16\,951,5\text{ Da}$  ;  $M_{mesurée} = \pm\text{ Da}$ )

### 3.1.3. Résultats des analyses LC-MS haute résolution

► Mesures de masses et hypothèses sur les espèces formées lors de la réaction

Grâce à l'optimisation des performances résolutes de notre instrument, les espèces formées lors de la réduction des peroxyrédoxines par la sulfirédoxine ont pu être discriminées et une mesure précise de la masse moyenne de ces composés a pu être réalisée. La figure 9 présente le spectre résolu des complexes disulfure et thiosulfinate formés entre la peroxyrédoxine et la sulfirédoxine, obtenu sur microTOF-Q Bruker Dal. avec nos paramètres optimisés.

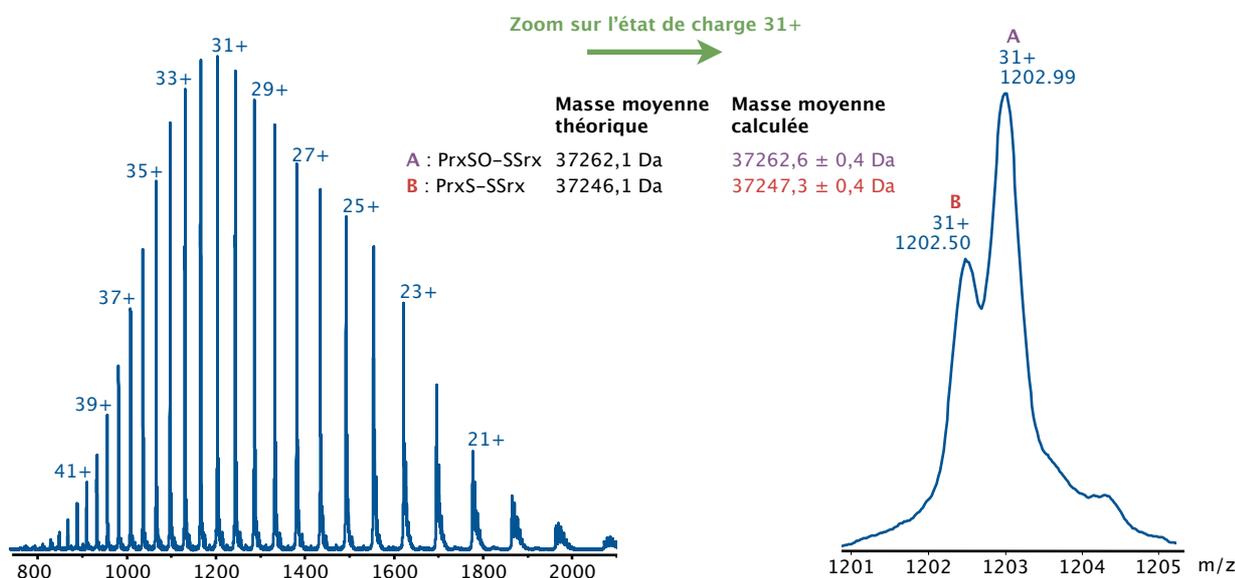


Figure 9 | Spectre de masse ESI-MS, acquis sur microTOF-Q Bruker Dal., des complexes disulfure PrxS-SSrx et thiosulfinate PrxSO-SSrx. Les performances résolutes optimisées de l'instrument permettent la séparation et la mesure de la masse moyenne des deux espèces présentes sur le même spectre.

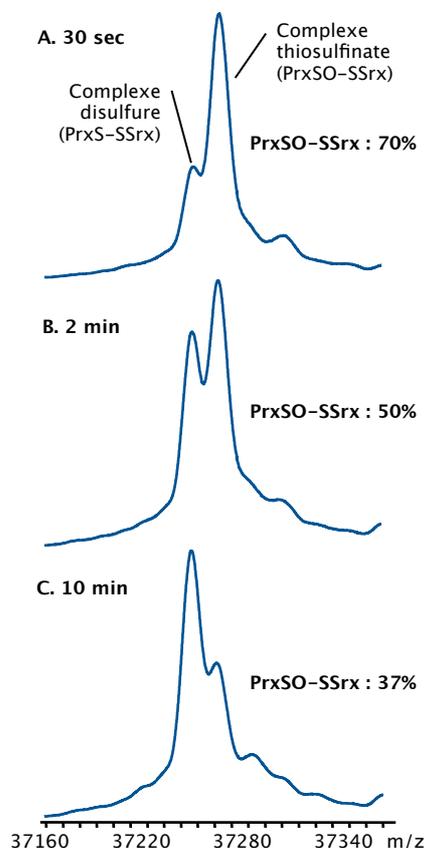
La table 2 rassemble l'ensemble des masses mesurées et leur assignation, pour chaque fraction chromatographique de l'échantillon et du contrôle. Les masses des réactifs (sulfirédoxine alkylée et peroxyrédoxine suroxydée) sont mesurées dans chacun des deux échantillons. Par ailleurs, la réaction semble conduire à la formation d'espèces dimériques de sulfirédoxine et de peroxyrédoxine réduite.

Temps de rétention	Espèces	Formes	Masses moyennes observées (Da)	Masses moyennes attendues (Da)
15-18 min	Srx	-	13 658,6 ± 0,1	13 658,6
		+IAM	13 715,5 ± 0,1°	13 715,6
	Dimère de Srx	SrxS-SSrx	27 316,6 ± 0,7	27 315,2
+IAM		27 430,6 ± 0,5	27 429,3	
+2IAM		27 543,6 ± 0,7	27 543,4	
23-25 min	Complexes Prx-Srx	Pont disulfure	37 247,3 ± 0,4	37 246,1
		Pont thiosulfinate	37 262,6 ± 0,4	37 262,1
26-29 min	Prx oxydée	PrxSO <sub>2</sub>	23 621,9 ± 0,4°	23 621,5
		PrxSO <sub>3</sub>	23 637,6 ± 0,3°	23 637,5
	Prx réduite	-	23 589,9 ± 0,4	23 589,5

**Table 2** | Masses mesurées lors de l'analyse LC-MS du mélange réactionnel SrxC48A/C106A - Prx C171A. ° : Masses observées également dans l'expérience de contrôle sans ATP.

▸ Suivi de l'évolution cinétique du ratio complexe thiosulfinate / disulfure par LC-MS

A ce stade, il n'est pas possible d'affirmer sans ambiguïté que la masse mesurée de 37 262,6 Da correspond effectivement à un complexe Prx-Srx à pont thiosulfinate. En effet, un complexe Prx-Srx à pont disulfure dont un résidu non cystéique est oxydé a la même masse théorique. Pour lever l'ambiguïté, le suivi de l'évolution cinétique du ratio entre les hypothétiques complexes thiosulfinate et disulfure a été réalisé. Le mélange réactionnel incubé pendant des durées croissantes (30 sec, 2 min et 10 min) a été analysé directement par LC-MS selon la méthode détaillée précédemment. La figure 10 montre qu'en augmentant le temps d'incubation, l'intensité du pic correspondant à la masse du complexe thiosulfinate diminue au profit de celle du pic correspondant à la masse du complexe disulfure.



**Figure 10** | Suivi de l'évolution cinétique du rapport entre les complexes Prx-Srx thiosulfinate et disulfure. Spectres de masse déconvolués des espèces covalentes formées à partir de PrxSO<sub>2</sub> C171A et Srx C48A/C106A en présence d'ATP après 30 s (A), 2 min (B) et 10 min (C) d'incubation.

Cette expérience permet d'invalider l'hypothèse de l'oxydation d'un résidu non cystéique (une méthionine oxydée par exemple ne pourrait pas être réduite dans les conditions de l'expérience). Elle conforte cependant l'idée que le complexe thiosulfinate est instable et s'hydrolyse rapidement.

▸ Formation d'espèces secondaires lors de l'hydrolyse du complexe Prx-Srx à pont thiosulfinate

A partir de ces observations et de l'observation, lors du suivi de la réaction par LC-MS, de masses correspondantes à un dimère de sulfirédoxines et à la peroxyrédoxine réduite, un mécanisme expliquant l'hydrolyse du complexe Prx-Srx à pont thiosulfinate et la formation d'espèces secondaires a pu être proposé (Figure 11).

▸ Un dimère de sulfirédoxine à pont disulfure peut résulter de l'hydrolyse du complexe peroxyrédoxine-sulfirédoxine à pont thiosulfinate qui libère une sulfirédoxine libre capable d'agir en tant que réducteur.

▸ Un complexe peroxyrédoxine-sulfirédoxine à pont disulfure peut résulter de l'attaque du complexe peroxyrédoxine-sulfirédoxine à pont thiosulfinate par une sulfirédoxine libre. La peroxyrédoxine sulfénique libérée peut alors elle-même être attaquée par une nouvelle sulfirédoxine libre pour former l'hétérodimère à pont disulfure.

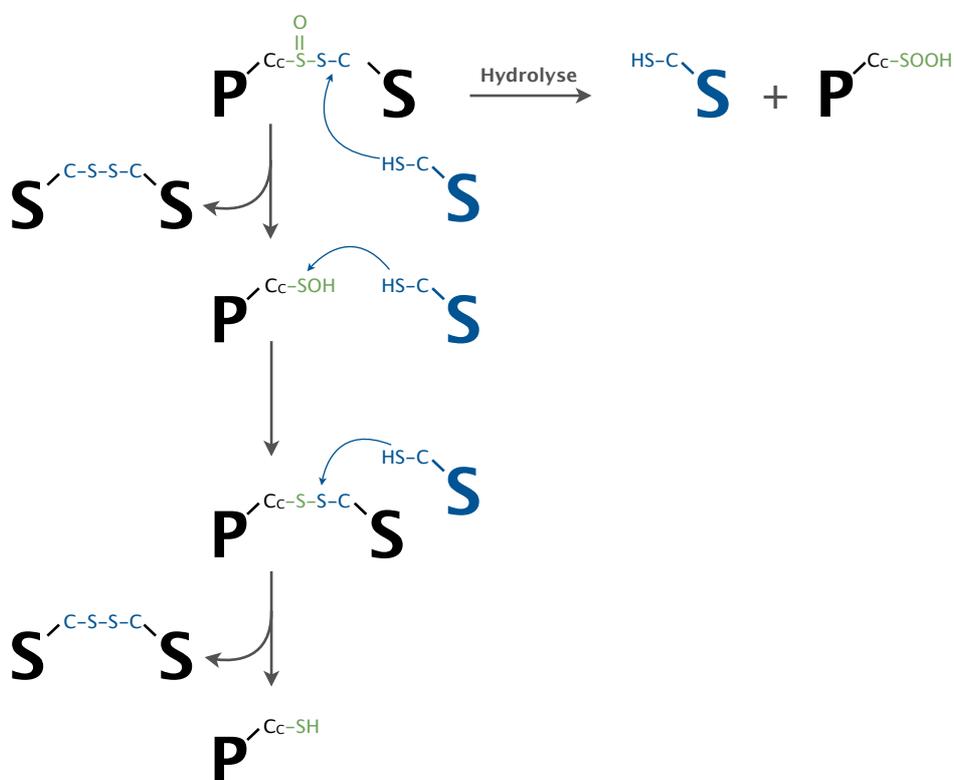


Figure 11 | Cascade de réactions possibles, consécutive à l'hydrolyse du complexe Prx C171A - Srx C48A/C106A à pont thiosulfinate.

### 3.2. Caractérisation des complexes par nanoLC-MS/MS

Afin de déterminer précisément la position des ponts et de caractériser finement les complexes entre la peroxyrédoxine et la sulfirédoxine, nous avons souhaité aller plus loin et tenter d'analyser les complexes suivant la stratégie nanoLC-MS/MS après digestion enzymatique.

La fraction chromatographique contenant les complexes Prx-Srx thiosulfinate et disulfure a été collectée et soumise à un traitement de digestion enzymatique. Comme il l'a été décrit au chapitre I de cette partie, la stabilité des liaisons cystéine-cystéine peut être remise en cause à pH neutre ou légèrement alcalin.

Pour ces raisons, la digestion du mélange de complexes a été réalisée par la trypsine d'une part (permettant des coupures spécifiques, mais fonctionnant à pH neutre) et par la pepsine d'autre part (largement aspécifique, mais active en milieu acide). Les échantillons ont par ailleurs été soumis ou non à un traitement de réduction et d'alkylation à l'iodoacétamide. Les analyses nanoLC-MS/MS ont été réalisées sur le système HPLC-Chip (Agilent Tech.) – HCT Ultra (Bruker Dal.).

### 3.2.1. Caractérisation indirecte de la liaison cystéine-cystéine entre la peroxyrédoxine C171A et la sulfirédoxine C48A/C106A

L'information de séquence obtenue après analyse nanoLC-MS/MS des échantillons non réduits, digérés par la trypsine ou la pepsine, recouvre pratiquement tous les peptides de chacune des deux protéines, à l'exception de ceux supposés être impliqués dans une liaison cystéine-cystéine (Figure 12).

Sulfirédoxine C48A/C106A										
	<u>SLQS</u>	<u>NSVKP</u>	<u>TEIPL</u>	<u>SEIRR</u>	<u>PLAPV</u>	<u>LDPOK</u>	<u>IDAMV</u>	<u>ATMKG</u>	<u>IPTAS</u>	<u>KTASL</u>
51	<u>EQAEA</u>	<u>AASAG</u>	<u>ELPPV</u>	<u>DVLGV</u>	<u>RVKGQ</u>	<u>TLYYA</u>	<u>FGGCH</u>	<u>RLQAY</u>	<u>DRRAR</u>	<u>ETQNA</u>
101	<u>AFPVR</u>	<u>ARVLP</u>	<u>ATPRQ</u>	<u>IRMYL</u>	<u>GSSLD</u>	<u>IE</u>				
Peroxyrédoxine C171A										
	<u>GSSH</u>	<u>HHHHH</u>	<u>SSGLV</u>	<u>PRGSH</u>	<u>MVAQV</u>	<u>QKQAP</u>	<u>TEKKT</u>	<u>AVVDG</u>	<u>VFDEV</u>	<u>SLDKY</u>
51	<u>KGKYV</u>	<u>VLAFI</u>	<u>PLAFT</u>	<u>FVCPT</u>	<u>EIIAF</u>	<u>SEAAK</u>	<u>KFEEQ</u>	<u>GAQVL</u>	<u>FASTD</u>	<u>SEYSL</u>
101	<u>LAWTN</u>	<u>IPRKE</u>	<u>GGLGP</u>	<u>INIPL</u>	<u>LADTN</u>	<u>HSLSR</u>	<u>DYGVV</u>	<u>IEEEG</u>	<u>VALRG</u>	<u>LFIID</u>
151	<u>PKGVI</u>	<u>RHITI</u>	<u>NDLPV</u>	<u>GRNVD</u>	<u>EALRL</u>	<u>VEAFQ</u>	<u>WTDKN</u>	<u>GTVLP</u>	<u>ANWTP</u>	<u>GAATI</u>
201	<u>KPTVE</u>	<u>DSKEY</u>	<u>FEAAN</u>	<u>K</u>						

Figure 12 | Recouvrement de séquence obtenu par l'analyse nanoLC-MS/MS. En couleur : digestion à la trypsine, Souligné : digestion à la pepsine.

Par ailleurs, après traitement de réduction et d'alkylation, le peptide tryptique Srx 74-87 a été séquencé avec sa cystéine carbamidométhylée et le peptide pepsique Prx 66-74 a été séquencé avec sa cystéine carbamidométhylée (Figures 13 et 14).

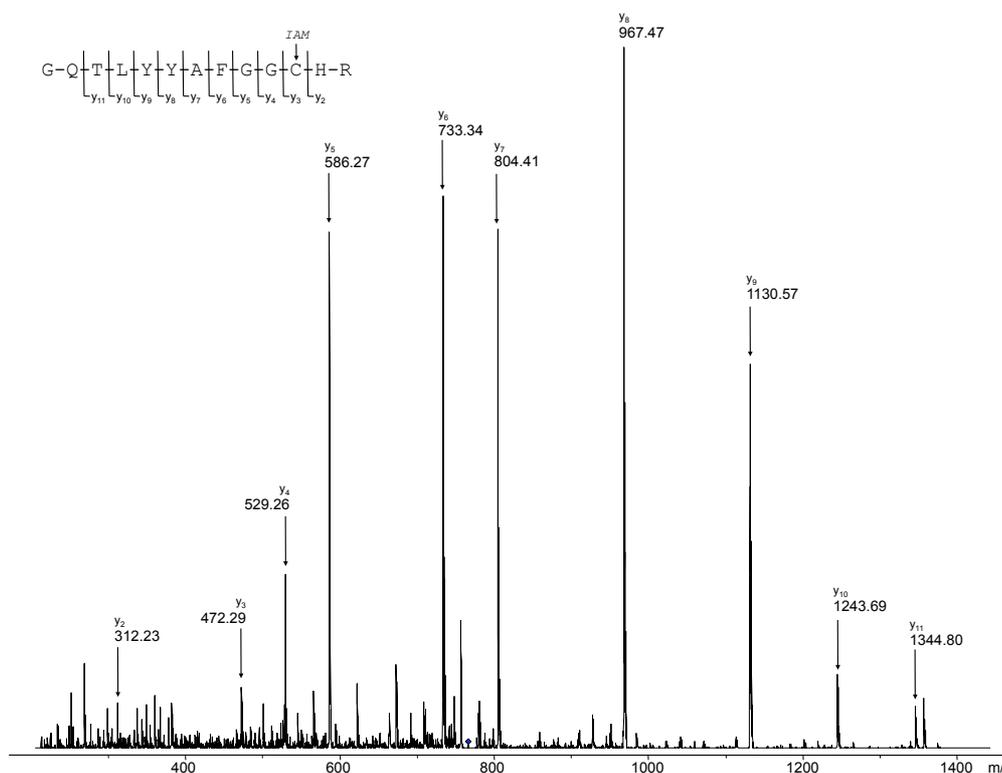
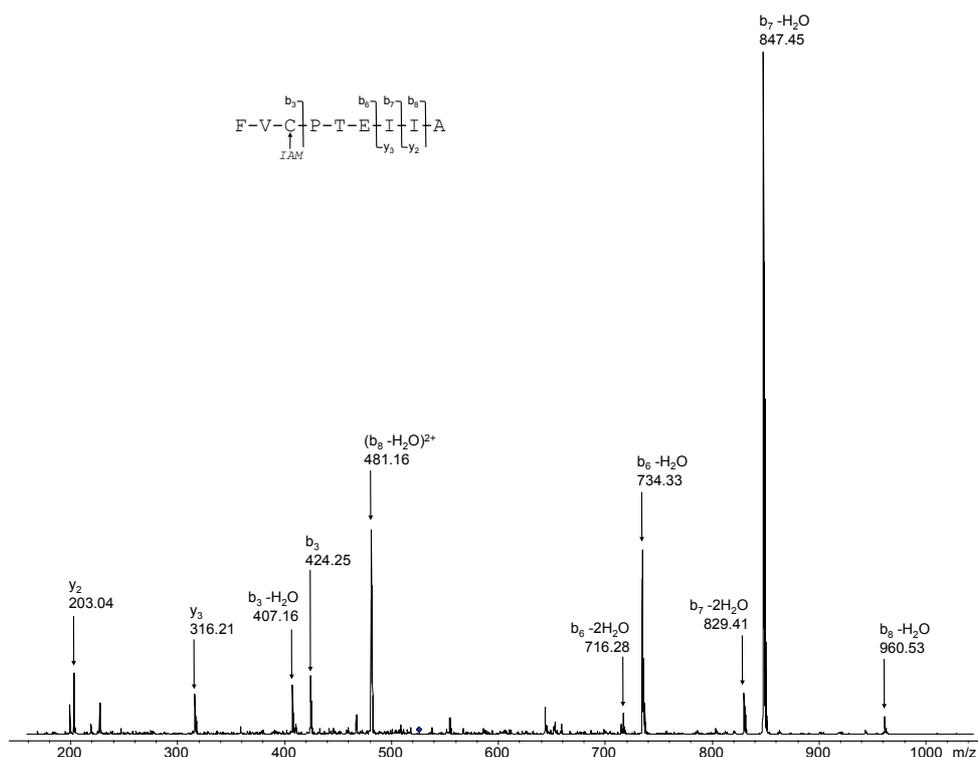


Figure 13 | Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à 765,92 m/z, correspondant au peptide Srx 74-87 carbamidométhylé, obtenu après traitement de réduction-alkylation et digestion tryptique. Les fragments majeurs correspondent à la série  $y_n$  ( $n=2-11$ ).



**Figure 14** | Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à 525,58 m/z, correspondant au peptide Prx 66–74 carbamidométhylé, obtenu après traitement de réduction-alkylation et digestion à la pepsine. Les fragments majeurs correspondent à la série  $b_n-H_2O$  ( $n=3-8$ ).

Deux conclusions peuvent être émises :

- Lorsque le complexe est maintenu en conditions non réductrices, l'identification de la quasi-totalité des peptides à l'exception des deux peptides portant une cystéine indique que ces peptides sont engagés dans une liaison réciproque.

- Lorsque le complexe est réduit et alkylé avant analyse, ces deux peptides sont identifiés et clairement séquencés avec leur cystéine alkylée : le pont qui les liait est donc réductible.

**Ces observations permettent de confirmer indirectement la présence d'une liaison cystéine-cystéine entre les deux protéines et de valider sa position entre les cystéines 84 de Srx et 68 de Prx.**

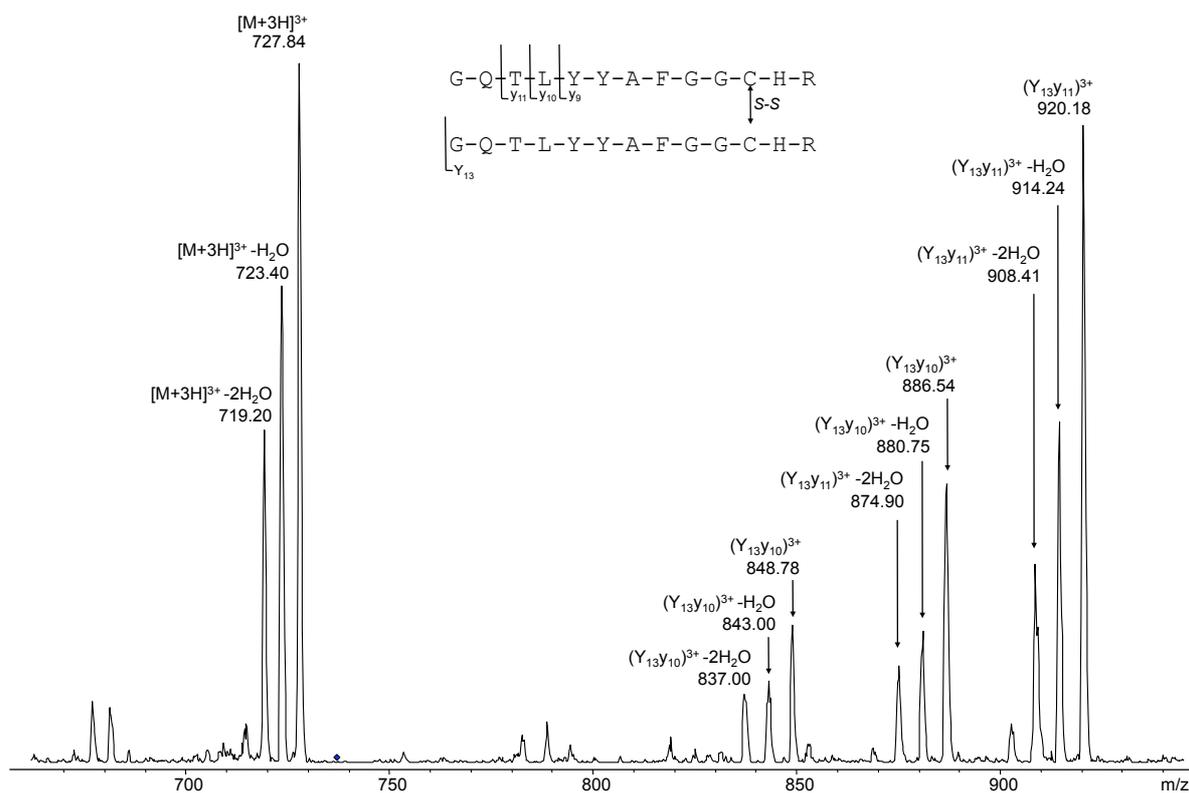
Afin de caractériser directement la nature de la liaison cystéine-cystéine entre la peroxyrédoxine et la sufirédoxine, le séquençage *de novo* du dipeptide Prx-Srx à pont thiosulfinate doit être réalisé.

### 3.2.2. Caractérisation directe des complexes impliqués dans la réaction

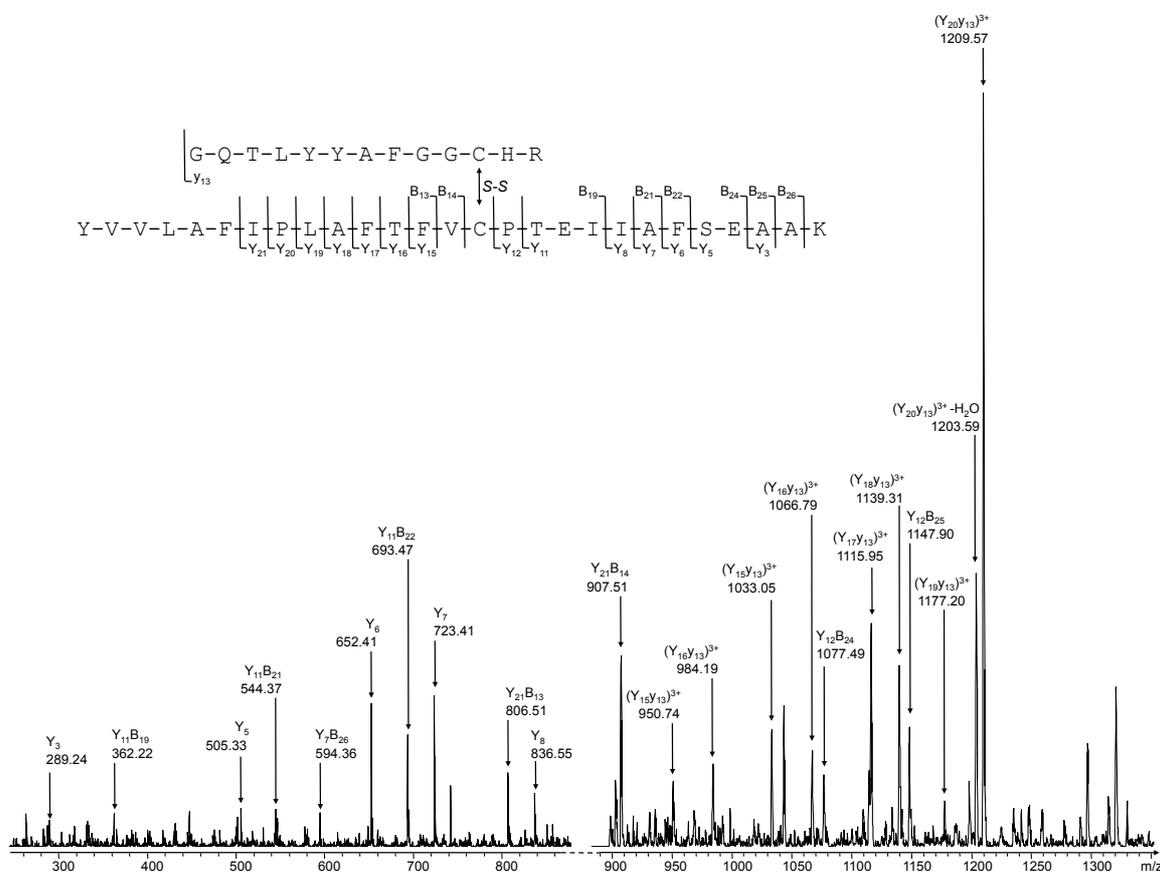
Malheureusement, aucun dipeptide à pont thiosulfinate n'a pu être directement séquencé en conditions non réductrices que ce soit après digestion trypsique ou pepsique. Il est probable que la liaison thiosulfinate instable soit rapidement hydrolysée dans les conditions de digestion trypsique (pH 7,5 sur la nuit, à 37 °C). La digestion par la pepsine – réalisée à pH 1,6 – évite la dégradation de la liaison thiosulfinate, mais génère de nombreux peptides de manière très peu spécifique. Devant cette complexité, et les difficultés d'analyse de peptides de ce type (expliquées au chapitre I), aucun dipeptide thiosulfinate n'a pu être identifié.

Des dipeptides à pont disulfure ont toutefois pu être identifiés et clairement séquencés dans le cas d'une digestion à la trypsine. La caractérisation des dipeptides Srx84-S-S-Srx84 (Figure 15) et Prx68-S-S-Srx84 (Figure 16) valide les hypothèses de formation d'un dimère de sulfirédoxine et d'un complexe peroxyrédoxine-sufirédoxine à pont disulfure. Ces produits secondaires ne peuvent être expliqués que par l'hydrolyse du complexe à liaison thiosulfinate et les réactions successives qui s'ensuivent (qui ont été proposées précédemment, figure 11).

**Ces preuves indirectes confirment l'hypothèse de la formation d'un intermédiaire thiosulfinate entre la sulfirédoxine et la peroxyrédoxine suroxydée lors de sa réduction catalytique.**



**Figure 15** | Spectre de fragmentation de l'ion  $[M+4H]^{4+}$  à  $736,97$   $m/z$ , correspondant à l'homodipeptide  $Srx$  74-87 à pont disulfure ( $Srx$  C84 -  $Srx$  C84), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $(Y_{13}Y_n)^{3+}$  ( $n=11-10$ ).



**Figure 16** | Spectre de fragmentation de l'ion  $[M+5H]^{5+}$  à  $887,65$   $m/z$ , correspondant à l'hétérodipeptide  $Prx$  54-80 -  $Srx$  74-87 à pont disulfure ( $Prx$  C68 -  $Srx$  C84), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $(Y_nY_{13})^{3+}$  ( $n=15-20$ ).

### 3.3. Valorisation de ces travaux

#### 3.3.1. Publication

Xavier Roussel, Guillaume Béchade, Alexandre Kriznik, Alain Van Dorsselaer, Sarah Sanglier-Cianferani, Guy Branlant and Sophie Rahuel-Clermont

Evidence for the Formation of a Covalent Thiosulfinate Intermediate with Peroxiredoxin in the Catalytic Mechanism of Sulfiredoxin.

The Journal of Biological Chemistry (2008) **33**, 22371–22382.

#### 3.3.2. Communications par affiches

Béchade, G., Roussel, X., Rahuel-Clermont, S., Branlant, G., Sanglier, S., Van Dorsselaer, A.

A combination of mass spectrometry techniques to assess the formation of a covalent thiosulfinate catalytic intermediate between overoxidized peroxiredoxin and sulfiredoxin.

56th ASMS Conference on Mass Spectrometry and Allied Topics  
Denver (Colorado, Etats-Unis). 01–05 juin 2008.

Béchade, G., Roussel, X., Rahuel-Clermont, S., Branlant, G., Van Dorsselaer, A., Sanglier, S.

Mise en évidence par spectrométrie de masse de la formation d'un intermédiaire thiosulfinate entre la peroxyrédoxine et la sulfirédoxine.

25e Journées Françaises de Spectrométrie de Masse  
Grenoble (France). 08–11 juin 2008.

## 4. Conclusion

Depuis la découverte des sulfirédoxines – enzymes capables de réduire les peroxyrédoxines à 2 cystéines impliquées dans la régulation du stress oxydant – deux mécanismes ont été proposés. L'enjeu de notre étude consistait à caractériser l'activité réductase de Srx et en particulier l'intermédiaire impliqué dans le processus catalytique de réduction de la fonction sulfinique : un complexe thiosulfinate entre la peroxyrédoxine et la sulfirédoxine.

Nous avons pu relever ce défi analytique en travaillant à la mise au point d'une stratégie à 2 niveaux. Le premier, s'appuyant sur la mesure de masse de complexes entiers par LC-MS, a notamment nécessité la fine optimisation des performances résolutive de notre spectromètre de masse MicrOTOF-Q Bruker Dal. Cela a permis la discrimination de complexes Srx-Prx thiosulfinate et disulfure, coélués en chromatographie liquide analytique. Le deuxième niveau a nécessité le développement de méthodes permettant la caractérisation des complexes collectés. Il a été montré, grâce à l'identification de produits d'hydrolyse du complexe thiosulfinate, que l'espèce ne pouvait pas être caractérisée directement par une approche nanoLC-MS/MS. Néanmoins, des preuves indirectes ont pu être rassemblées pour conclure à l'activité réductase des Sulfirédoxines.

Peu après leur publication, nos travaux ont été confirmés par une étude de Jönsson *et al.*<sup>6</sup>, dans laquelle une approche de cinétique rapide couplée en ligne à la spectrométrie de masse a permis de caractériser les mêmes espèces que nous avons observées.

<sup>6</sup> Jönsson, T. J., et al. (2008). J Biol Chem **283**, 22890–4.



# Evidence for the Formation of a Covalent Thiosulfinate Intermediate with Peroxiredoxin in the Catalytic Mechanism of Sulfiredoxin<sup>\*S</sup>

Received for publication, January 18, 2008, and in revised form, May 28, 2008. Published, JBC Papers in Press, June 14, 2008, DOI 10.1074/jbc.M800493200

Xavier Roussel<sup>†1</sup>, Guillaume Béchade<sup>§1</sup>, Alexandre Kriznik<sup>‡</sup>, Alain Van Dorsselaer<sup>§</sup>, Sarah Sanglier-Cianferani<sup>§</sup>, Guy Branlant<sup>‡2</sup>, and Sophie Rahuel-Clermont<sup>‡3</sup>

From the <sup>†</sup>Unité Mixte de Recherche, CNRS-UHP 7567, Maturation des Acide Ribonucléique et Enzymologie Moléculaire, Nancy Université, Faculté des Sciences, Bld. des Aiguillettes, BP 239, Vandoeuvre-lès-Nancy 54506, France and the <sup>§</sup>Unité Mixte de Recherche, CNRS-ULP 7178, Laboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien, Département Sciences Analytiques, Ecole de Chimie, Polymères et Matériaux, 25, rue Becquerel, Strasbourg 67087, France

The typical 2-Cys peroxiredoxins are thiol-peroxidases involved in the physiology of hydrogen peroxide not only as a toxic but also as a signaling molecule. Coordination of these functions depends on the sulfinylation of the catalytic Cys, a modification reversed by ATP-dependent sulfiredoxin, which specifically reduces the sulfinic acid group of overoxidized 2-Cys peroxiredoxins into a sulfenic acid. Sulfiredoxin was originally proposed to operate by covalent catalysis, with formation of a peroxiredoxin-sulfiredoxin intermediate linked by a thiosulfinate bond between the catalytic Cys of both partners, a hypothesis rejected by a study of the human enzyme. To settle the argument, we investigated the catalytic mechanism of *Saccharomyces cerevisiae* sulfiredoxin, by the characterization of the nature and kinetics of formation of the protein species formed between sulfiredoxin and its substrate in the presence of ATP, using mutants of the non-essential Cys residues of both proteins. We observed the formation of a dithiothreitol-reducible peroxiredoxin-sulfiredoxin species using SDS-PAGE and Western blot analysis, and its mass was shown to correspond to a thiosulfinate complex by high resolution mass spectrometry coupled to liquid chromatography. We next measured indirectly and directly a rate constant of formation of the thiosulfinate species of  $\sim 2 \text{ min}^{-1}$ , for both wild-type and mutant sulfiredoxins, at least equal to the steady-state rate constant of the reaction, with a stoichiometry of 1:1 relative to peroxiredoxin. Taken altogether, our results strongly argue in favor of the formation of a covalent thiosulfinate peroxiredoxin-sulfiredoxin species as an intermediate on the catalytic pathway.

Reactive oxygen species, including hydrogen peroxide originating both from exogen and endogen sources, are compounds

deeply involved in cellular physiology. They can result in either toxic or beneficial effects through chemical modification of cellular components. Sulfinylation of a protein cysteine thiol was considered to be a biologically irreversible event *in vivo* until the discovery of new enzymatic sulfinyl reductase activities referred to as sulfiredoxin (Srx)<sup>4</sup> (1) and sestrins (2), capable of reducing a sulfinic acid group -SO<sub>2</sub>H (sulfur oxidation number +II) into a sulfenic acid -SOH (oxidation number 0). The strong interest toward the elucidation of the molecular and structural enzymology of these enzymes lies in their potential regulatory functions affecting the bipolar nature of H<sub>2</sub>O<sub>2</sub> as a toxic and a carcinogen, and as a signaling molecule (3–5).

In these mechanisms specific to eukaryotic organisms, a family of thiol-peroxidases susceptible to Cys sulfinylation, the peroxiredoxins (Prx), plays a central role and constitutes the specific substrate of Srx and sestrins (2, 6). Typical 2-Cys Prx (2-Cys-Prx) represents a class of peroxidase distinct from catalase and glutathione peroxidase, the functional unit of which is structured as a symmetrical homodimer. Typical 2-Cys-Prxs reduce H<sub>2</sub>O<sub>2</sub> by oxidation of the N-terminal catalytic Cys (C<sub>p</sub>) to a sulfenic acid intermediate, then recycled by disulfide formation with a regenerating Cys (C<sub>R</sub>) located on the symmetrical subunit, finally coupled to reduction by thioredoxin. Furthermore, Prxs have been proposed to participate in the transduction and regulation of H<sub>2</sub>O<sub>2</sub>-mediated cell signaling (7).

The modulation of Prx activity by post-translational modification adds another layer to the potential regulations of H<sub>2</sub>O<sub>2</sub> effects in eukaryotic cells. The Prx peroxidase activity is inhibited by Thr phosphorylation (8). In addition, eukaryotic 2-Cys-Prxs have been described to be susceptible to the oxidation of the catalytic Cys to a higher level, the sulfinic acid state, during the catalytic cycle, by a mechanism of escape of the reactive sulfenic acid intermediate (9). This overoxidation of 2-Cys-Prx, also observed *in vivo*, not only abolishes peroxidase activity but

<sup>\*</sup> This work was supported in part by the CNRS, the University of Nancy I, the Institut Fédératif de Recherche 111 Bioingénierie, and by the French Agence Nationale de la Recherche (Program ANR-06-BLAN-0369). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>S</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental text and Figs. S1–S5.

<sup>†</sup> Supported by the French Research Ministry.

<sup>2</sup> To whom correspondence may be addressed. Tel.: 33-3-83-68-43-04; Fax: 33-3-83-68-43-07; E-mail: [guy.branlant@maem.uhp-nancy.fr](mailto:guy.branlant@maem.uhp-nancy.fr).

<sup>3</sup> To whom correspondence may be addressed. Tel.: 33-3-83-68-43-12; Fax: 33-3-83-68-43-07; E-mail: [sophie.rahuel@maem.uhp-nancy.fr](mailto:sophie.rahuel@maem.uhp-nancy.fr).

<sup>4</sup> The abbreviations used are: Srx, sulfiredoxin; Prx, peroxiredoxin; 2-Cys-Prx, two-cysteine peroxiredoxin; C<sub>p</sub>, 2-Cys-Prx catalytic Cys; C<sub>R</sub>, 2-Cys-Prx regeneration Cys; DTT, 1,4-dithiothreitol; IAM, iodoacetamide; LC-MS, reversed-phase liquid chromatography coupled to mass spectrometry; <sup>m</sup>e<sup>7</sup>Guo, methyl-7-guanosine; MES, 2-(N-morpholino)ethanesulfonic acid; NTR, *E. coli* NADPH thioredoxin reductase; PNP, His-tagged purine nucleoside phosphorylase from *E. coli*; PrxSO<sub>2</sub>, overoxidized *S. cerevisiae* His-tagged peroxiredoxin Tsa1; Trx, *E. coli* thioredoxin 1; FWHM, full width at half-maximum.

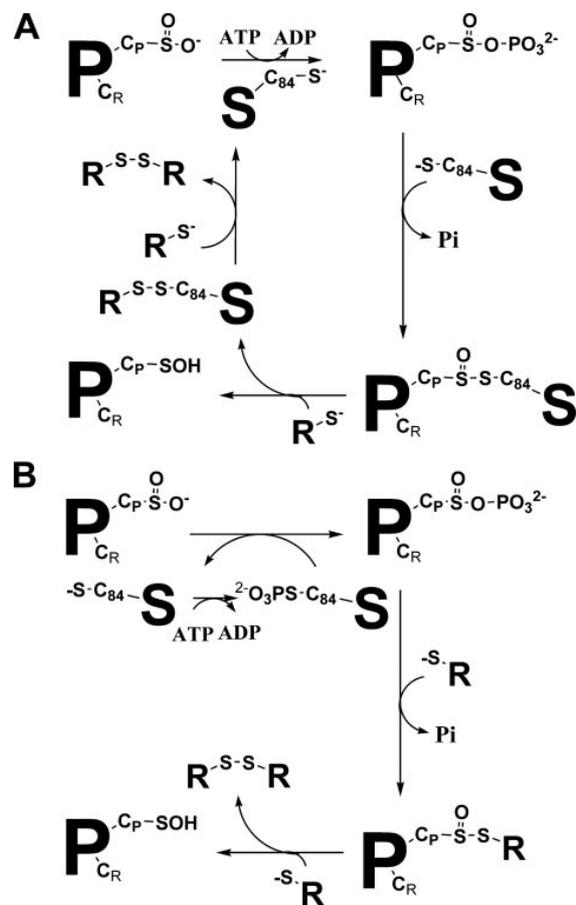
## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

also stabilizes high molecular weight forms of Prx, which show chaperone-like activities. Therefore, sulfanyl reductase activities can potentially regulate several aspects of H<sub>2</sub>O<sub>2</sub> physiology involving Prx in eukaryotes, by restoring the native state of 2-Cys-Prx: 1) in the “floodgate” theory (9), sulfenylation/sulfanyl reduction would be associated with the requirement of temporary and/or local inactivation of Prx peroxidase activity to allow the transmission of H<sub>2</sub>O<sub>2</sub>-mediated signal; 2) as proven in *Schizosaccharomyces pombe*, Prx overoxidation and its reversion by Srx constitute a redox switch between two signal transduction paths dependent on H<sub>2</sub>O<sub>2</sub> concentration (10); and 3) in view of the recently reported chaperone activity induced by overoxidation, sulfanyl reductase would also regulate a switch between oxidative stress conditions, with overoxidized 2-Cys-Prx acting as a cellular protector, and H<sub>2</sub>O<sub>2</sub> signaling conditions, with native reduced 2-Cys-Prx acting as redox sensor for example (11). In addition, a new regulator function for human Srx has been proposed, by catalysis of the deglutathionylation of several distinct proteins in response to oxidative stress (12). In support of the role of Srx as part of cellular redox regulation and signaling systems, several studies have shown that expression of Srx is induced under oxidative stress conditions, including yeasts, plants, and mammals (1, 13–15).

Srxs are small proteins of 13–18 kDa generally with a basic character, identified in the genome of some eukaryotes (in plant and algae, metazoans, and fungi) and some cyanobacteria species. They catalyze the ATP-dependent reduction of the overoxidized catalytic Cys C<sub>p</sub> (as sulfinic acid) of typical 2-Cys-Prx, with substrate specificity exclusive of other Prx classes or of other proteins possessing a Cys sensitive to overoxidation (e.g. glyceraldehyde 3-phosphate dehydrogenase) (6). The Srx activities described so far for mammalian enzymes are slow, with turnover numbers of 0.2–0.5 min<sup>-1</sup> (16, 17).

The Srx activity was initially described by Biteau *et al.* (1), on the *Saccharomyces cerevisiae* enzyme, who proposed a catalytic mechanism in which Srx acted first as an ATP-dependent phosphotransferase to activate the sulfinic group of the Prx substrate as a phosphoryl sulfenic intermediate, leading to the release of ADP. In a second step, Srx would act as a reductase through the formation of a PrxSO·SSrx thiosulfinate intermediate involving Srx essential Cys (Cys-99 in mammalian enzymes, Cys-84 in *S. cerevisiae* enzyme), a phosphate group being thus released in this step (Fig. 1A). This intermediate would be reduced by thioredoxin or by another thiol reductant, leading to the release of the Prx reduced in the sulfenic acid state and of a mixed disulfide thioredoxin-Srx, followed by recycling of Srx to the reduced form and release of oxidized thioredoxin. In 2006, an alternative mechanism based on a kinetic study of the human enzyme reaction and infirming the reductase role of Srx was supported. In this mechanism, the Prx phosphoryl sulfenic intermediate is directly reduced by a thiol reductant distinct from Srx (17) (Fig. 1B).

In the present report, we examined the catalytic mechanism of *S. cerevisiae* Srx. First, we characterized the protein species formed between Srx and its overoxidized 2-Cys-Prx substrate in the absence of added reductant by the use of SDS-PAGE, Western blot, and mass spectrometry coupled to liquid chromatography. Second, to validate the formation of these species as part



**FIGURE 1. Srx hypothetical catalytic mechanisms.** In hypothesis A, Srx catalyzes both the phosphotransferase and the reductase steps by reductive attack of the catalytic Cys on the phosphorylated substrate, leading to a covalent thiosulfinate intermediate with Srx. Subsequent reduction is achieved by an external reductant. In hypothesis B, Srx only catalyzes the reductase step by a phosphorylated Srx intermediate. All subsequent reduction steps depend on an external reductant. No covalent intermediate is formed between Srx and the Prx substrate. P, Prx; S, Srx; C<sub>p</sub>, peroxidatic Cys; C<sub>r</sub>, regeneration Cys; C<sub>84</sub>, Srx catalytic Cys; R, thiol reductant distinct from Srx.

of the catalytic mechanism, we then designed kinetic approaches aimed at monitoring the kinetics and stoichiometry of the putative steps. To avoid secondary reactions, we used mutants of the non-essential Cys residues of both Srx and the overoxidized Prx. Altogether, our results favor the formation of a covalent thiosulfinate Prx-Srx species as an intermediate of Srx catalytic pathway.

## EXPERIMENTAL PROCEDURES

**Materials**—Tris was from VWR (West Chester, PA). KCl and MgCl<sub>2</sub> were from Merck (Darmstadt, Germany), NADPH was obtained from Roche (Basel, Switzerland), and dithiothreitol (DTT) and ammonium sulfate were from Euromedex (Souffelweyersheim, France). ATP, MES, iodoacetamide (IAM), ammonium hydrogen carbonate, and horse heart myoglobin were from Sigma-Aldrich, and phosphoric acid and H<sub>2</sub>O<sub>2</sub> were from Acros Organics (Geel, Belgium). Rabbit antiserum specific for purified *S. cerevisiae* Srx was produced by Centre Lago (Vonnas, France). Mouse antiserum specific for His tag was

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

from Qiagen (Hilden, Germany), and secondary antisera were obtained from Pierce. Trifluoroacetic, formic acids, and <sup>m7</sup>guanosine (<sup>m7</sup>Guo) were from Fluka. Acetonitrile was obtained from Carlo Erba Reactifs SDS (Val de Reuil, France).

Thioredoxin 1 (Trx) and NADPH thioredoxin reductase (NTR) from *Escherichia coli* were prepared following experimental procedures described previously (18, 19). Purine nucleoside phosphorylase from *E. coli* was expressed from a plasmid derived from pET28b(+) encoding the N-terminal His tag fusion protein of the enzyme (referred to as PNP), obtained by cloning the *deoD* open reading frame amplified by PCR using *E. coli* K12 genomic DNA as template between the NdeI and SacI sites (oligonucleotides not shown). The PNP was produced and purified according to the same procedures as the His-tagged *S. cerevisiae* Prx (see below), and stored as a lyophilizate at -20 °C.

**Plasmid Constructions and Preparation of Recombinant *S. cerevisiae* Srx and Prx Proteins**—The plasmid pET20bSrx $\Sigma$  encoding the *S. cerevisiae* Srx protein was obtained by subcloning the synthetic *srx $\Sigma$  open reading frame optimized for expression in *E. coli* (GeneArt AG, Regensburg, Germany), into the pET20b(+) plasmid (Novagen, Merck Chemicals, Darmstadt, Germany) between the NdeI and SacI sites. The plasmid pET28bHTTsa1 encoding the N-terminal His tag fusion protein of *S. cerevisiae* 2-Cys-Prx Tsa1 (referred to as Prx in the following text) was obtained by cloning the *tsa1* open reading frame amplified by PCR (GC-rich system, Roche Applied Science, Basel, Switzerland) using *S. cerevisiae* W303 genomic DNA as template, into the pET28b(+) plasmid between the NdeI and SacI sites. The forward primer contained an NdeI restriction site, and the reverse primer contained an SacI restriction site (sequences of oligonucleotides not shown). The mutant proteins C48A-C106A Srx and C171A Prx were generated by standard PCR site-directed mutagenesis. *E. coli* C41(DE3) [F<sup>-</sup> ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm (DE3)] (20) transformants containing the wild-type or mutant pET20bSrx $\Sigma$  were grown at 37 °C for 24 h and 30 h, respectively, in the autoinducible media ZYM-5052 or N-5052 (21), respectively, supplemented with ampicillin (200 mg l<sup>-1</sup>). Wild-type and mutant Prxs were overexpressed in the same strain transformed by the pET28bHTTsa1 plasmid, by overnight culture at 37 °C in the ZYM-5052 medium supplemented with kanamycin (200 mg l<sup>-1</sup>).*

For Srx purification, cells were harvested by centrifugation, resuspended in a minimal volume of buffer A (100 mM MES pH 6.1) containing 20 mM DTT, and disrupted by sonication. Srx contained in the soluble fraction of the cellular extract was then precipitated by ammonium sulfate at 55% saturation, followed by cation exchange chromatography on a sulfopropyl-Sepharose column connected to a fast protein liquid chromatography system (Amersham Biosciences) equilibrated with buffer A. Srx was eluted by a 0–1 M KCl linear gradient with a step at 350 mM. Final purification of Srx was achieved by hydrophobic chromatography on a phenyl-Sepharose column (Amersham Biosciences) equilibrated with buffer A plus 1 M ammonium sulfate, eluted with a linear 1 to 0 M ammonium sulfate gradient. For Prx purification, cells were harvested by centrifugation, resuspended in a minimal volume of buffer B (20 mM K<sub>2</sub>HPO<sub>4</sub>,

1 M KCl, pH 7.5) containing 20 mM DTT and disrupted by sonication. Prx contained in the soluble fraction was purified on nickel-Sepharose column equilibrated with buffer B plus 50 mM imidazole, connected to a fast protein liquid chromatography system, and eluted by a 0.5 M imidazole step.

At this stage, wild-type and mutant proteins were pure as checked by electrophoresis on 15% SDS-polyacrylamide gel followed by Coomassie Brilliant Blue R-250 staining and by electrospray mass spectrometry analyses. Purified Srx and Prx were stored at -20 °C in the presence of 20 mM DTT, and 70% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 15% glycerol, respectively, and were stable for several weeks under these conditions. Immediately before use, both proteins were incubated with 50 mM DTT at 4 °C, followed by desalting or oxidation reaction. Protein molar concentrations were determined spectrophotometrically, using extinction coefficients at 280 nm determined experimentally, of 7280 M<sup>-1</sup> cm<sup>-1</sup> for wild-type and mutant Srxs, and of 29500 M<sup>-1</sup> cm<sup>-1</sup> for wild-type and mutant Prxs. Prx enzymatic activity was measured by a coupled assay using the Trx recycling system (0.5 μM NTR, 200 μM NADPH, 50 μM Trx), with 100 μM H<sub>2</sub>O<sub>2</sub> as substrate. Initial rate measurements were carried out at 30 °C in buffer TK (50 mM Tris-HCl, 100 mM KCl, pH 7.0) on a UV mc2 spectrophotometer (Safas, Monaco) by following the decrease of the absorbance at 340 nm due to the oxidation of NADPH. A steady-state rate constant of 2 s<sup>-1</sup> was measured in these conditions.

**Preparation of Overoxidized Prx**—Oxidation of Cys C<sub>p</sub> of both wild-type and C171A Prxs was achieved in TK buffer at 30 °C, by incubation of 500 μM Prx with 5 mM H<sub>2</sub>O<sub>2</sub> in the presence of 100 mM DTT, followed by two additions of 5 mM H<sub>2</sub>O<sub>2</sub> at 10-min intervals. The oxidized protein (PrxSO<sub>2</sub>) was then desalted through an Econo-Pac 10 DG column (Bio-Rad) equilibrated with buffer TK. The sulfinic oxidation state of Cys C<sub>p</sub> was checked by titration of the protein with 5,5'-dithiobis(2-nitro)benzoate under denaturing conditions (1% SDS), by following the loss of Prx enzymatic activity and by electrospray mass spectrometry analysis. PrxSO<sub>2</sub> molar concentration was determined spectrophotometrically, using extinction coefficients at 280 nm determined experimentally of 27500 M<sup>-1</sup> cm<sup>-1</sup> for the wild-type and mutant proteins.

**Kinetics of Srx Reaction under Steady-state Conditions**—The reaction was followed in the steady state using Trx as reducer and the Trx recycling system as a coupled assay (50 μM Trx, 0.5 μM NTR, and 200 μM NADPH), in the presence of 1 mM ATP, 1 mM MgCl<sub>2</sub>, 100 μM PrxSO<sub>2</sub>, and variable concentrations of Srx. Initial rate measurements were carried out at 30 °C in buffer TK on a Safas UV mc2 spectrophotometer by following the decrease of the absorbance at 340 nm due to the oxidation of NADPH. A blank measurement recorded in the absence of Srx was systematically deduced from the assay to account for the nonspecific oxidation of NTR. When wild-type PrxSO<sub>2</sub> was used as substrate, a stoichiometry of 2 mol of NADPH per mol of PrxSO<sub>2</sub> was used in rate calculations to account for the oxidation of 1 mol of Trx per mol PrxSO<sub>2</sub> in Prx own catalytic cycle, following the reduction of Cys C<sub>p</sub> to the sulfenic acid state.

**SDS-PAGE and Immunoblot Analysis**—Reaction mixtures containing 30 μM Srx and 30 μM PrxSO<sub>2</sub>, 1 mM ATP, 1 mM MgCl<sub>2</sub>, were incubated in buffer TK for 30 s, 1 min, and 2 min at

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

30 °C. DTT was added to 50 mM in an additional sample incubated for 2 min. All reactions were then treated by 250 mM IAM, and the proteins were immediately loaded onto 15% SDS-PAGE gel in non-reducing conditions, followed by Coomassie Blue staining. The proteins separated on two additional identical gels run in parallel were transferred on nitrocellulose membranes for revelation with antibodies against Srx or His tag. Immune complexes were detected with horseradish peroxidase-conjugated secondary antibodies and enhanced chemiluminescence reagents (Pierce).

**PNP-coupled Assay**—The kinetics of  $P_i$  release were followed by the decrease of the emission fluorescence intensity associated with the phosphorolysis of  $^{m7}Guo$  catalyzed by PNP as a coupled assay (22). Time courses were recorded at 30 °C on an SX18MV-R stopped-flow apparatus (Applied PhotoPhysics, Leatherhead, UK) fitted for fluorescence measurements, with the excitation wavelength set at 305 nm, and the emitted light was collected above 455 nm using a cutoff filter. One syringe contained 1 mM ATP, 1 mM  $MgCl_2$ , 300  $\mu M$   $^{m7}Guo$ , 5  $\mu M$  PNP in buffer TK, and the other contained 100  $\mu M$  PrxSO<sub>2</sub>, 20  $\mu M$  PNP, variable concentration of Srx in the presence or absence of 300  $\mu M$  Trx (final concentrations after mixing). For each condition, a blank time course was recorded in the absence of Srx and deduced from the corresponding experience in the presence of Srx. An average of three runs was recorded for each set of conditions. Under single turnover conditions (in the absence of Trx), the rate constant  $k_{obs}$  was obtained by fitting fluorescence traces after blank deduction against Equation 1, in which  $c$  represents the end point and  $a$  is the amplitude of the fluorescence decrease.

$$y = ae^{-k_{obs}t} + c \quad (\text{Eq. 1})$$

The steady-state rate constant was obtained by measurement of initial rate in the presence of Trx. The fluorescence signal was calibrated against  $P_i$  concentration to allow determination of the stoichiometry of  $P_i$  released during Srx reaction.

**Kinetics of Formation of the Prx-Srx Complexes followed by Reverse Phase Chromatography**—Reaction mixtures containing 20  $\mu M$  C48A-C106A Srx and 100  $\mu M$  C171A PrxSO<sub>2</sub>, 1 mM ATP, 1 mM  $MgCl_2$  were incubated in the TK buffer at 30 °C. Aliquots were quenched by using trifluoroacetic acid, 0.1%, and analyzed by reversed-phase liquid chromatography on an Aquapore RP-300 (C8) column, 4.6 × 100 mm, 7  $\mu m$  (PerkinElmer Life Sciences), coupled to the ÄKTAexplorer system (Amersham Biosciences), using the same gradient as described below for LC-MS.

**Liquid Chromatography Coupled to Mass Spectrometry: Sample Preparation**—Reaction mixtures containing 30  $\mu M$  Srx (C48A-C106A double mutant/wild-type Srx) and 30  $\mu M$  PrxSO<sub>2</sub> (C171A/wild-type PrxSO<sub>2</sub>), 1 mM ATP, 1 mM  $MgCl_2$  were incubated in buffer TK for 30 s, 2 min, and 10 min at 30 °C. Reactions were then treated by 25 mM IAM. Finally, samples were diluted 2.5 times and were analyzed by reversed-phase LC-MS immediately to cope with thiosulfinate instability and to prevent secondary reactions.

**LC-MS Analyses**—MicroLC-MS analyses were performed on a Hewlett Packard 1100 Series Analytic LC system coupled with a MicrOTOF-Q Mass spectrometer (Bruker Daltonics, Bre-

men, Germany). An aliquot of 50  $\mu l$  of each sample was loaded immediately after preparation onto a Vydac 208TP52 (C8) column, 2.1 × 250 mm, 5  $\mu m$ . Proteins were eluted from the column using a linear gradient from 30 to 60% of B in 30 min and from 60 to 80% of solvent B in 1 min with A:H<sub>2</sub>O plus 0.1% trifluoroacetic acid (v/v) and B:acetonitrile plus 0.08% trifluoroacetic acid (v/v) at a flow rate of 250  $\mu l$  min<sup>-1</sup>. The column was then flushed with 80% of solvent B during 5 min.

The capillary and end plate voltage of the mass spectrometer were set to -4500 V and -500 V, respectively. The complete system was fully controlled by HyStar and micrOTOFControl software (Bruker Daltonics). For the measurements, external calibration was performed with a horse heart myoglobin solution at 0.2 pmol  $\mu l^{-1}$  in a water/acetonitrile mixture (1:1, v/v) acidified with 1% (v/v) formic acid. The MS acquisition range was set between  $m/z$  500 and 3000, and three scans were averaged to obtain MS spectra. Collected data were treated using DataAnalysis software (Bruker Daltonics). For all spectra, the baseline was subtracted before they were smoothed once using the Savitsky-Golay algorithm with a smoothing width of 0.2  $m/z$ . Deconvoluted mass spectra were performed using the Maximum Entropy Charge Deconvolution module (Bruker Daltonics). Ratios of thiosulfinate and disulfide intermediates were estimated from peak heights of the deconvoluted mass spectra.

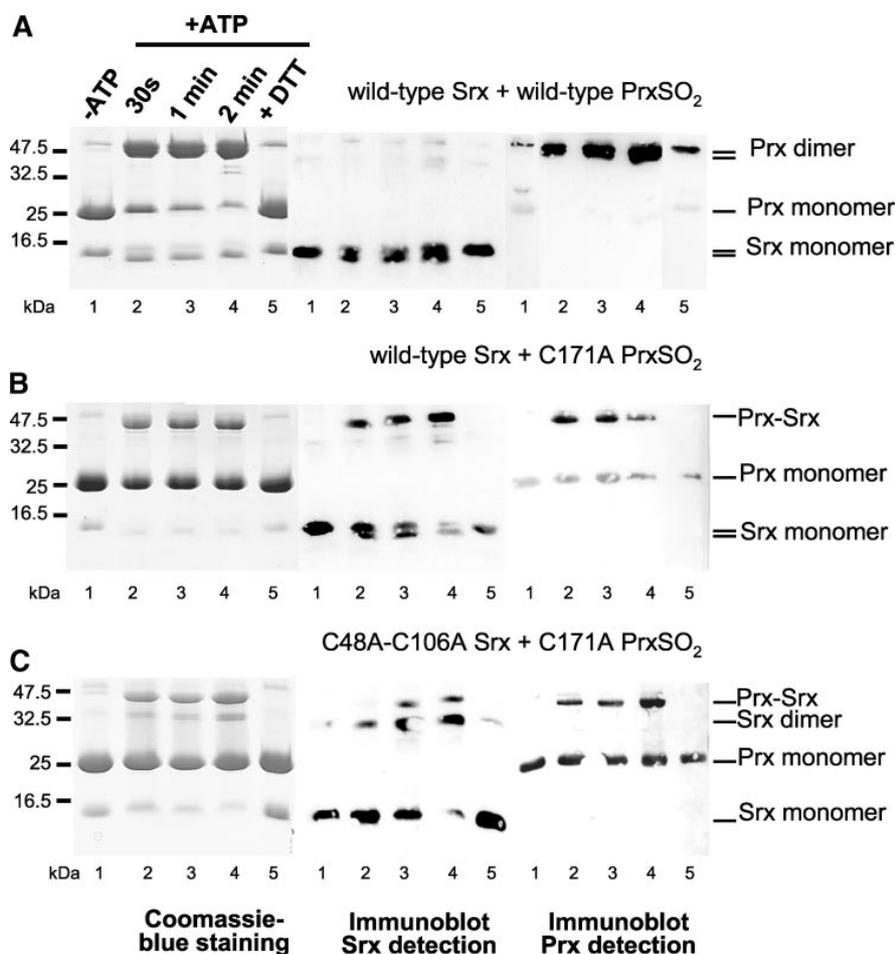
**High Resolution MS Analysis**—To achieve the required resolution to discriminate co-eluting Prx-Srx thiosulfinate and disulfide complexes (mass difference of 16 Da), a fine tuning of the mass spectrometer was performed to get a minimal theoretical mean mass resolution (full width at half-maximum (FWHM)) of 2,350 at 37 kDa. Preliminary experiments performed in the "focus mode" with tuning mix solution (Agilent Technologies, Santa Clara, CA), aiming at maximizing resolution around  $m/z$  1000, allowed to reach a mean resolution (FWHM) of 18,000 for a 15-kDa protein. This optimization subsequently led to a resolution (FWHM) of 2,800 for a 37-kDa protein complex.

**Control Experiments**—Prior to each reaction monitoring, a control experiment, in the same conditions as those described above but without ATP addition, was performed (see Fig. 3). Mass assignments were performed after LC-MS analysis. The following molecular masses were measured: (i) peak 1': wild-type Srx plus 2 IAM (13,836.9 ± 0.1 Da); (ii) peak 2': wild-type PrxSO<sub>2</sub> plus IAM (23,711.4 ± 0.1 Da), wild-type PrxSOH plus IAM (23,694.6 ± 0.6 Da), wild-type PrxSO<sub>2</sub> (23,654.3 ± 0.3 Da) and wild-type PrxSO<sub>3</sub> (23,726.4 ± Da); (iii) peak 3': wild-type Srx plus 2 IAM (13,835.8 ± 0.1 Da) and wild-type Srx plus IAM (13,778.8 ± 0.2 Da); (iv) peak 5': C171A PrxSO<sub>2</sub> (23,620.1 ± 0.3), C171A PrxSOH (23,603.1 ± 0.5 Da) and C171A PrxSO<sub>3</sub> (23,635.5 ± 0.5 Da); (v) peak 7': C48A/C106A Srx plus IAM (13,715.17 ± 0.4 Da); and (vi) peak 9': C171A PrxSO<sub>2</sub> (23,621.0 ± 0.6 Da), C171A PrxSOH (23,604.1 ± 0.5 Da), and C171A PrxSO<sub>3</sub> (23,636.6 ± 0.6 Da).

## RESULTS AND DISCUSSION

**Identification of Prx-Srx Species Formed in the Absence of Added Reducer**—Mechanistic studies on Srxs have been reported for *S. cerevisiae* and mammalian enzymes, including searching for covalent intermediates between overoxidized Prx

### Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism



**FIGURE 2. Non-reducing SDS-PAGE and immunoblot analysis of the species formed for the reaction catalyzed by Srx in the absence of added reducer.** Equimolar concentrations (30  $\mu$ M) of Srx and PrxSO<sub>2</sub> were mixed in the absence (lane 1) or presence of 1 mM ATP-MgCl<sub>2</sub> for the indicated times (lanes 2–4), and for 2 min followed by the addition of 50 mM DTT (lane 5), immediately followed by addition of 25 mM IAM and non-reducing SDS-PAGE fractionation. Reactions were carried out in TK buffer at 30 °C. Three identical SDS-PAGE gels were run in parallel for Coomassie Blue staining (left panel) or immunoblot revelation against Srx (middle panel) and His tag (right panel) specific antisera. A, wild-type PrxSO<sub>2</sub> and wild-type Srx; bands corresponding to monomeric Prx in the immunoblot are very faint due to the relative amounts and sensitivity to His tag antisera of both species; B, C171A PrxSO<sub>2</sub> and wild-type Srx; C, C171A PrxSO<sub>2</sub> and C48A-C106A Srx.

substrate and Srx enzyme. Studies on *S. cerevisiae* cells subjected to H<sub>2</sub>O<sub>2</sub> stress identified a Cys-84-dependent (Prx)<sub>2</sub>-Srx species with DTT-sensitive linkage in cellular extracts which, although not fully chemically characterized, led to the suggestion that Srx catalytic Cys could be involved in the formation of a Prx-Srx thiosulfinate intermediate species (1). On the contrary, no thiosulfinate species could be detected after *in vitro* reaction of human overoxidized Prx1 and human Srx in the presence of ATP, followed by Western blot and mass spectrometry analyses (17). We therefore tackled the *in vitro* study of the reaction products of *S. cerevisiae* Srx with the His-tagged overoxidized peroxiredoxin Tsa1 (referred to as PrxSO<sub>2</sub>), shown to interact with Srx *in vivo* (1), in non-reducing conditions to allow the accumulation of intermediate species.

The potential reactivity and instability of the thiosulfinate bond have already been studied on model compounds. In particular, it is known that thiosulfonates react readily with thiols,

which are described to attack preferentially on the sulfenic sulfur (-S-) rather than on the sulfinic sulfur (-SO-) of the thiosulfinate bond, giving off a disulfide product and a sulfenic acid product (23, 24). Therefore, we set up experimental conditions aiming at reducing secondary reactions on this putative intermediate. First, the potential secondary reactions of free thiol groups of Srx and Prx with intermediate species in the absence of reducer were avoided during the subsequent analysis steps by carbamidomethylation with IAM, which alkylates the remaining free thiol groups, immediately followed by non-reducing SDS-PAGE or LC-MS. Second, because free thiols of Srx and PrxSO<sub>2</sub> could lead to secondary reactions also during the time course of experiments, assays were carried out both on wild-type and C48A-C106A double mutant Srx and both on wild-type and C171A PrxSO<sub>2</sub>. Indeed, *S. cerevisiae* Srx contains two non-conserved Cys in positions 48 and 106 in addition to the catalytic Cys-84, for which substitutions were described to have minor effects on the reaction *in vivo* (1). Using the C171A mutant PrxSO<sub>2</sub>, which lacks the regeneration Cys<sub>R</sub>, as a substrate for Srx, gave a similar steady-state rate constant as wild-type PrxSO<sub>2</sub> (1.7 ± 0.4 min<sup>-1</sup> versus 1.9 ± 0.4 min<sup>-1</sup>) in the presence of Trx as reducer by the Trx/NTR assay. This suggests that this residue was not

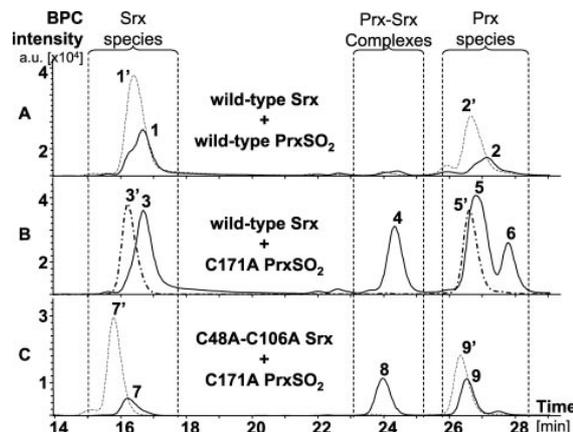
involved in Srx catalytic mechanism, as observed for the human enzyme (17).

The products of the reaction between wild-type Srx and wild-type or C171A PrxSO<sub>2</sub> and between C48A-C106A Srx and C171A PrxSO<sub>2</sub> in non-reducing conditions were analyzed by SDS-PAGE and Western blots against Srx and His tag antisera, after 30-s, 1-min, and 2-min incubations (Fig. 2). For the wild-type proteins, in the presence of ATP/MgCl<sub>2</sub>, two major species migrating at ~47 kDa were formed at the expense of the one corresponding to monomeric Prx (23.6 kDa) and were sensitive to DTT (Fig. 2A). Immunoblots revealed that these species comprised only the Prx protein, suggesting that dimeric, DTT-reducible Prx species were formed. Srx was found almost exclusively as a monomeric species (13.7 kDa), although it migrated slightly higher than expected with regard to its molecular weight, probably due to its basic character (theoretical pI of 9.2). In addition, two bands with migration similar to mono-

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

meric Srx were observed in the presence of ATP, suggesting that a significant fraction of the Srx population migrated slightly faster, resulting in the apparent doubling of this band into two bands. Addition of DTT after 2-min incubation restored the population of Srx migrating as observed before ATP addition, suggesting that the band migrating slightly faster corresponded to an oxidized Srx monomer with an intramolecular disulfide bond. When C171A PrxSO<sub>2</sub> was used as the substrate with wild-type Srx (Fig. 2B), a major DTT-sensitive species migrating at ~45 kDa appeared in the presence of ATP, corresponding to a Prx-Srx species, as revealed by immunoblotting. Again, likely due to Srx basic character, this species migrated significantly higher than expected for a molecular mass of 37.6 kDa. Division of the Srx band into two was also suggested by Coomassie staining and confirmed by immunoblotting against Srx antisera. Similar to the experiment with wild-type PrxSO<sub>2</sub>, this phenomenon was reversed by DTT treatment. However, no apparent doubling of the monomeric Srx band occurred for the C48A-C106A mutant of Srx incubated with the C171A substrate (Fig. 2C), thus supporting the formation of an oxidized Srx monomer with an intramolecular disulfide bond in the case of wild-type Srx. Instead, a new species migrating at ~32 kDa appeared that comprised only Srx, thus indicating the formation of dimeric Srx, whereas in the same conditions, a Prx-Srx species was produced. As observed in previous conditions, both species dissociated in the presence of DTT, indicating covalent linkages involving oxidized Cys residues.

Given that they were produced specifically in the presence of ATP, several hypotheses on the origins of these species can be raised based on the catalytic mechanisms proposed for Srx. First, the formation of oxidized forms of Srx implies chemical activation of Cys-84. This residue could be activated either as a phosphorylated intermediate according to Jeong *et al.* (17), or as a thiosulfinate intermediate (Fig. 1). In the case of wild-type Srx, intramolecular attack of either Cys-48 or Cys-106 would prevail, resulting in a monomeric oxidized species, whereas in the case of C48A-C106A mutant Srx, intermolecular attack of Cys-84 from native Srx would occur concurrently to the catalytic reactions, leading to a dimeric species. In the hypothesis of formation of a PrxSO-SSrx thiosulfinate intermediate, an important point to note is that this reaction would result in release of the reactive sulfenate form of the peroxidatic Cys C<sub>P</sub> of Prx. Second, formation of the dimeric Prx species is associated with the presence of Cys-171 (C<sub>R</sub>), suggesting the reaction of Cys C<sub>R</sub> on the activated Cys C<sub>P</sub> of Prx, likewise the regeneration step in the catalytic cycle of any typical 2-Cys-Prx. Species that could play the role of activated Prx are either the phosphorylated, the thiosulfinate intermediates or the oxidized form of Prx with a sulfenate Cys C<sub>P</sub>. Two forms, including one or two symmetrical bonds, would be expected, as observed in Fig. 2A (lanes 2–4) and in a previous study (25). Finally, the Prx-Srx species observed with the C171A mutant of PrxSO<sub>2</sub> (Fig. 2, B and C) can only be explained by the hypothesis involving a PrxSO-SSrx thiosulfinate intermediate: either it corresponds to the intermediate itself, or to a disulfide issued from the reaction of a free thiol of Srx with the sulfenic Prx-C<sub>P</sub>SOH released from the thiosulfinate intermediate, as suggested above.



**FIGURE 3. LC-MS analysis of the reaction catalyzed by Srx in the absence of added reducer.** Reactions were carried out for 2 min in the same conditions as in Fig. 2, in the absence (*dashed lines*) and in the presence of ATP (*solid lines*), immediately followed by addition of 25 mM IAM and by LC-MS analysis. Base peak chromatograms of wild-type PrxSO<sub>2</sub> and wild-type Srx (A), C171A PrxSO<sub>2</sub> and wild-type Srx (B), and C171A PrxSO<sub>2</sub> and C48A-C106A Srx (C) are presented. Protein species are eluted in three clusters of peaks corresponding to Srx species, covalent Prx-Srx complexes and Prx species. Peaks are numbered for reference in the text and in Table 1.

*Identification of a Thiosulfinate Intermediate by LC-MS and NanoLC-MS/MS Analyses*—To test the hypothesis of the formation of a thiosulfinate complex, we focused on the structural characterization of the various products of the reactions between Prx (wild type/C171A mutant) and Srx (wild type/C48A-C106A double mutant) species by LC-MS. To distinguish between possible disulfide and thiosulfinate complexes between Srx and Prx, which are differentiated by a mass increase of 16 Da, a minimal theoretical mean mass resolution (FWHM) of 2,350 at 37 kDa is needed, which implies the use of a mass spectrometer finely tuned for high resolution (see “Experimental Procedures” for details).

Fig. 3 presents base peak chromatograms obtained for control experiments performed in the absence of ATP and for enzymatic reactions run for 2 min in the presence of ATP, in conditions similar as for Western blot analyses. For control experiments (*dashed lines*), two main peaks were detected. MS analysis showed that species eluting at retention times between 15 and 18 min and between 26 and 28 min corresponded to monomeric Srx and to monomeric PrxSO<sub>2</sub>, respectively, labeled with carbamidomethyl groups inherent to IAM treatment. Molecular masses measured for control experiments (peak numbers 1', 2', 3', 5', 7', and 9') are given under “Experimental Procedures.”

In the presence of ATP, one additional peak eluting between 23 and 25 min was observed for reactions involving the C171A PrxSO<sub>2</sub> mutant (Fig. 3, B and C, *solid lines*), which corresponds to Prx-Srx complexes (peaks 4 and 8). Mass assignments after LC-MS analyses in the presence of ATP are reported in Table 1. By contrast, for the reaction of wild-type Srx with wild-type PrxSO<sub>2</sub>, no additional Prx-Srx complex peak was observed on the chromatogram (Fig. 3A), which is in agreement with the results of Western blots. Peak 2 (at retention time ~27 min) revealed two major species, with masses that may be attributed to dimeric Prx linked with one (PrxS-SPrxSO<sub>2</sub>) or two (PrxS-S/

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

**TABLE 1**

**Mass assignments obtained after LC-MS analysis for the reaction catalyzed by Srx in the absence of added reducer**

Reactions were carried out for 2 min in the same conditions as in Fig. 2, immediately followed by addition of 25 mM IAM and by LC-MS analysis. Minor species (<10%) are not reported, and major ones are in bold. Peak numbers refer to Fig. 3. Data are expressed in daltons.

RT	Species	Wild-type Srx and Prx		Wild-type Srx and C171A Prx		C48A-C106A Srx and C171A Prx	
		Observed molecular mass	Expected molecular mass	Observed molecular mass	Expected molecular mass	Observed molecular mass	Expected molecular mass
<i>min</i>							
15–18	Srx	<b>Peak 1</b>		<b>Peak 3</b>		<b>Peak 7</b>	
	+IAM	13,779.6 ± 0.3	13,722.7	13,779.9 ± 0.3	13,722.7	<b>13,658.6 ± 0.1</b>	13,658.6
	+2 IAM	13,836.8 ± 0.4	13,836.8	13,836.4 ± 0.7	13,836.8	<b>13,715.5 ± 0.1</b>	13,715.6
	SS Bridge	<b>13,721.0 ± 0.2</b>	13,720.7	<b>13,721.1 ± 0.1</b>	13,720.7		13,772.6
	Dimer of Srx						
	SrxS-SSrx		27,443.4		27,443.4	27,316.6 ± 0.7	27,315.2
	+2 IAM		27,557.5		27,525.5	<b>27,430.6 ± 0.5</b>	27,429.3
	+4 IAM		27,671.6		27,639.6	27,543.6 ± 0.7	27,543.4
23–25	Disulfide complex			<b>Peak 4</b>		<b>Peak 8</b>	
	PrxS-SSrx		37,342.3		37,310.3	<b>37,247.3 ± 0.4</b>	37,246.1
	+IAM		37,399.4	<b>37,368.7 ± 0.2</b>	37,367.3		
	Thiosulfinate complex		37,358.3		37,326.3	<b>37,262.6 ± 0.4</b>	37,262.1
26–29	Oxidized Prx	<b>Peak 2</b>		<b>Peak 5</b>		<b>Peak 9</b>	
	PrxSO <sub>2</sub>	23,654.6 ± 0.5	23,653.6	<b>23,622.2 ± 0.4</b>	23,621.5	<b>23,621.9 ± 0.4</b>	23,621.5
	PrxSO <sub>3</sub>		23,669.6	23,638.3 ± 0.1	23,637.5	23,637.6 ± 0.3	23,637.5
	Reduced Prx			<b>Peak 6</b>		<b>Peak 10</b>	
			23,621.6	<b>23,590.2 ± 0.2</b>	23,589.5	23,589.9 ± 0.4	23,589.5
	Dimer of Prx				47,177.1		47,177.1
	PrxS-S/S-Prx	<b>47,241.5 ± 0.3</b>	47,239.2				
	PrxS-SPrxSO <sub>2</sub>	<b>47,277.8 ± 0.7</b>	47,273.2				
	+IAM	<b>47,338.2 ± 0.6</b>	47,330.2				

S-SPrx) disulfide bonds (Table 1), likely between the Cys C<sub>P</sub> and C<sub>R</sub> from symmetrical subunits. However, the resolution of the MS instrument at 47 kDa did not allow us to discriminate between these two species. Such a result provides the first evidence in favor of the formation of a thiosulfinate intermediate between Srx and Prx, because it implies the reaction of the Cys C<sub>R</sub> on a released sulfenic Prx-C<sub>P</sub>SOH species. Therefore, reduction of the sulfinic PrxSO<sub>2</sub> is possible by a Srx thiol group in the absence of added reducer. For the reaction of wild-type Srx and C171A PrxSO<sub>2</sub>, *peak 4* (Fig. 3B) corresponds to one single ion series with a measured molecular mass of 37,368.7 ± 0.2 Da (Table 1), which could be attributed to a disulfide PrxS-SSrx complex with one IAM adduct (expected molecular mass of 37,367.3 Da), which is in agreement with Western blot analyses.

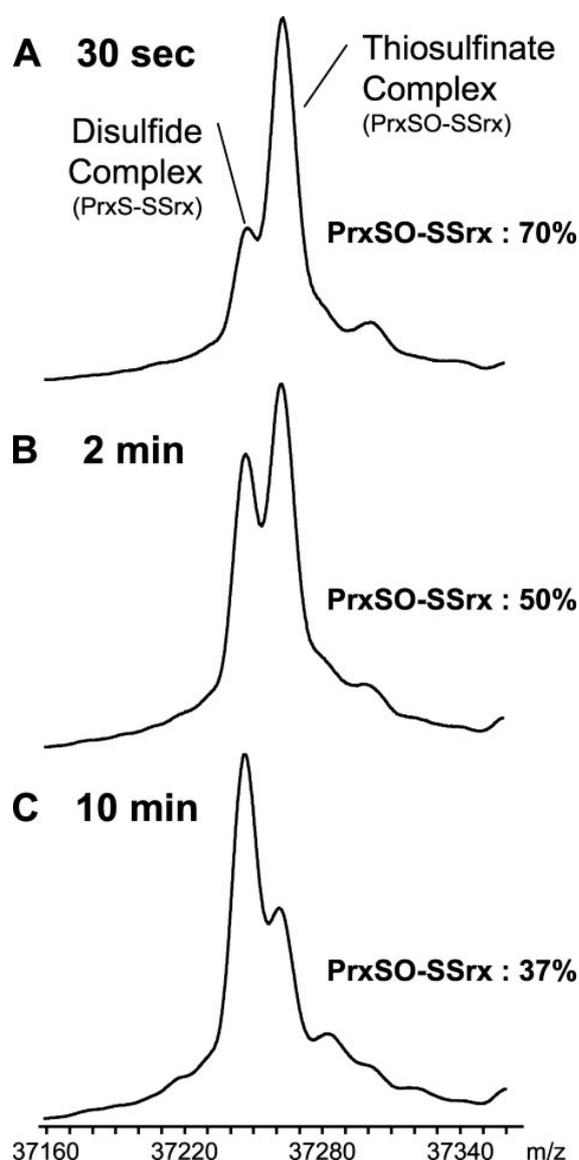
Finally, direct evidence in favor of the formation of a thiosulfinate-type PrxSO-SSrx species was obtained by LC-MS analysis of the reaction of the C48A-C106A double mutant Srx and C171A PrxSO<sub>2</sub>. Indeed, MS analysis of *peak 8* complexes showed a mixture of two species (Fig. 3C, Table 1): i) a compound with a mass of 37,262.6 ± 0.4 Da which could correspond to the expected PrxSO-SSrx thiosulfinate complex (expected MW: 37,262.1 Da, see Table 1) and ii) a species with a mass of 37,247.3 ± 0.4 Da which could be attributed to the PrxS-SSrx disulfide complex (expected MW: 37,246.1 Da, see Table 1). Since Srx Cys-84 and Prx C<sub>P</sub> are the only Cys residues available for the reaction, this result strongly argues in favor of the accumulation of a PrxSO-SSrx thiosulfinate species involving Srx Cys-84 and Prx C<sub>P</sub> in these conditions.

Even if the most probable explanation for the 37,262-Da species detected by LC-MS is the formation of the thiosulfinate intermediate, another hypothesis cannot be ruled out at this stage. Indeed, the 37,262-Da species could also correspond to a PrxS-SSrx disulfide compound with an additional oxidation of a non-Cys residue. However, LC-MS-monitored kinetic evolution of the ratio of putative thiosulfinate *versus* disulfide species

contained in *peak 8* invalidates this hypothesis. Indeed, the fraction of the 37,262-Da complex, which constitutes the major Prx-Srx species after 30-s reaction, decreased with time while the fraction of disulfide PrxS-SSrx species increased (Fig. 4). If the 37,262-Da species was a disulfide, the additional oxidation located for example on a Met residue, could not be reduced in these *in vitro* conditions to produce the PrxS-SSrx species. On the contrary, the evolution observed in Fig. 4 could be explained by the release of PrxSOH from a thiosulfinate species, followed by the attack of a reduced Cys thiol from Srx, which would lead to the increase of the disulfide species. In addition, these reactions would produce a disulfide SrxS-SSrx species. Indeed, such a complex was observed by immunoblotting (Fig. 2C) and by LC-MS (*peak 7*, Fig. 3C, Table 1). Although no free Cys is available on the SrxS-SSrx dimer, it was observed with two and four carbamidomethyl adducts, which can probably be attributed to increased His and Met reactivity toward IAM for this mutant Srx (26, 27).

To further establish the chemical nature of the bridging bond, *peak 8* was collected after LC-MS analysis and submitted to enzymatic digestions, followed by subsequent tandem mass spectrometry (MS/MS) analysis (see supplemental materials). First, these conditions allowed us to unambiguously confirm the existence of the peptide corresponding to the 74–86 (GQT-LYYAFGGCHR)<sub>2</sub> disulfide dipeptide of the SrxS-SSrx dimer (see supplemental Fig. S1). Because the digestions were performed on the isolated complexes from *peak 8*, this species can only be explained by the release of free reduced Srx, and it thus implies the hydrolysis of a thiosulfinate species. Indeed, model thiosulfinate compounds have been shown to be sensitive to hydrolysis at neutral or basic pH. Although the mechanism of thiosulfinate hydrolysis appears more complex than nucleophilic attack by thiols, it would significantly involve release of reduced Srx and PrxSO<sub>2</sub> as an initial step (28, 29). This result again supports the thiosulfinate nature of the 37,262-Da species. In addition, the dipeptide resulting from the PrxS-SSrx

### Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism



**FIGURE 4. Kinetic evolution of the ratio of thiosulfinate versus disulfide complexes monitored by MS.** Deconvoluted mass spectra of the covalent species formed between C171A PrxSO<sub>2</sub> and C48A-C106A Srx in the presence of ATP (*peak 8*, Fig. 3) obtained after 30 s (A), 2 min (B), and 10 min (C) reaction in identical conditions as in Fig. 2C.

species was also detected, showing the cross-linking of Prx Cys C<sub>p</sub> and Srx Cys-84. Indeed, the ion at  $m/z$  887.65 corresponds to the 5+ charge state of the 54–80 YVLAFLAFTFCPTETIAFSEAAK Prx peptide cross-linked by a disulfide bond to the 74–86 GQTLYYAFGGCHR Srx peptide (see supplemental Fig. S2). Due to the instability of thiosulfinate bonds at near neutral pH, no tryptic peptide containing a thiosulfinate cross-link was detected by nanoLC-MS/MS.

To attempt to directly characterize the putative thiosulfinate intermediate, pepsin digestions were performed at acidic pH, which is expected to stabilize the thiosulfinate bond (28). Despite accurate mass measurement of pepsic peptides, no peptide containing either the disulfide or the thiosulfinate bond

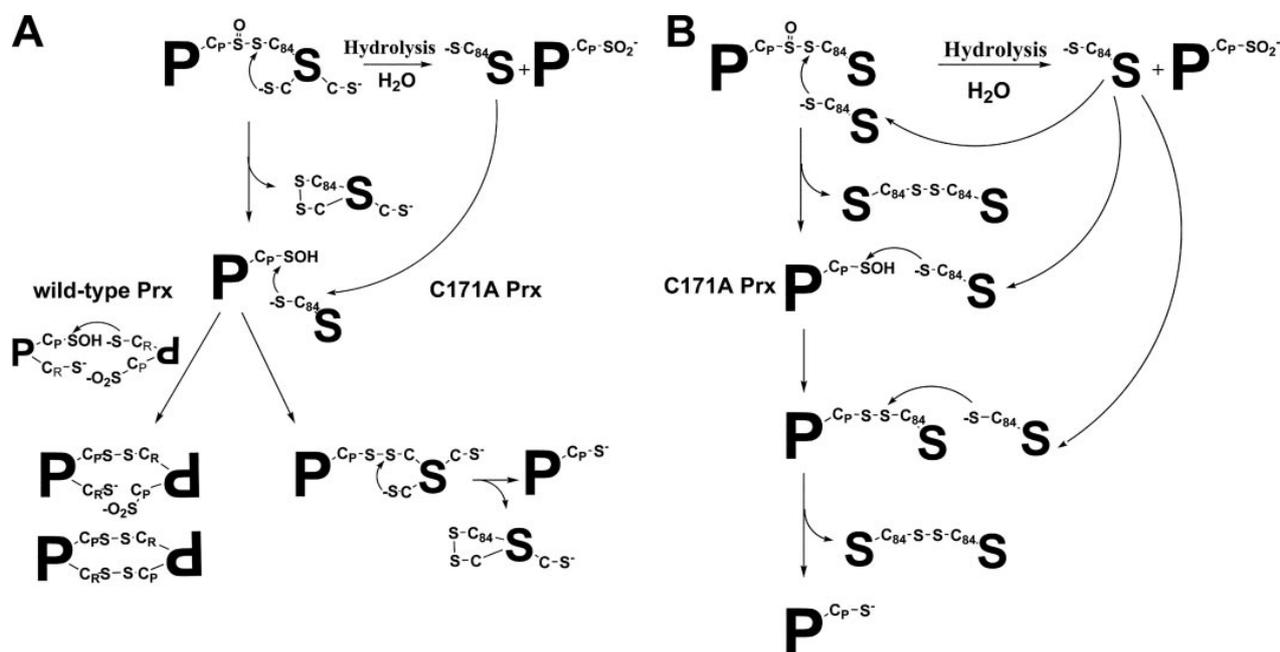
could be unambiguously sequenced even if molecular masses that could correspond to such peptides were measured with a high accuracy. Two reasons can account for this failure: (i) pepsin is a highly aspecific enzyme, and thus numerous missed cleavages can be considered resulting in a high number of potential theoretical proteolytic peptides, and (ii) MS/MS data for pepsic peptides were of low quality, which made precise sequence determination impossible.

However, analysis of the enzymatic digestions of the complexes by nanoLC-MS/MS allowed us to unambiguously localize the putative thiosulfinate bond on the amino acid scaffold. Indeed, the only peptides that could not be detected after enzymatic digestion (trypsin or pepsin) correspond to peptides containing either Prx Cys C<sub>p</sub> or Srx Cys-84 (see supplemental Fig. S3). When reduction (DTT) and alkylation (IAM) treatment was performed before tryptic digestion, the 74–86 GQTLYYAFGGC<sub>Cam</sub>HR peptide corresponding to the Srx Cys-84-containing peptide was clearly detected and sequenced by nanoLC-MS/MS analysis (supplemental Fig. S4). Even if the corresponding Prx Cys C<sub>p</sub>-containing peptide could not be observed after trypsin digestion, it could be sequenced in acidic conditions after pepsin digestion (supplemental Fig. S5).

Based on the preceding results, we propose the following scenario to explain the evolution of the reaction products (Fig. 5): formation of a thiosulfinate intermediate between Srx Cys-84 and Prx Cys C<sub>p</sub> occurs first after activation of Prx by phosphorylation. All secondary products derive from the high reactivity of this species, in particular toward thiols. For the reaction of C48A-C106A double mutant Srx with C171A PrxSO<sub>2</sub> (Fig. 5B), reduced Srx, resulting from the hydrolysis of the thiosulfinate intermediate, would attack the sulfenyl sulfur of the thiosulfinate intermediate, producing dimeric oxidized Srx (*peak 7*, Figs. 3C and 5B) and PrxC<sub>p</sub>SOH. PrxC<sub>p</sub>SOH would then react rapidly with reduced Srx, generating the PrxS-SSrx species observed in *peak 8* (Figs. 3C and 5B). Finally, a second attack of reduced Srx on the PrxS-SSrx disulfide complex would then result in the formation of a small amount of reduced Prx observed (*peak 10*, Fig. 3C, Table 1, and Fig. 5B). Reduced Prx was no longer observed after 10-min incubation (data not shown), likely due to the formation of a mixed disulfide with reduced Srx.

For wild-type Srx (Fig. 5A), the hypothesis of an intramolecular attack of a non-catalytic Cys thiol on Cys-84 sulfur atom is supported both by the electrophoretic shift of the monomeric Srx band (Fig. 2) and by LC-MS analysis. Indeed, a species at  $13,721.0 \pm 0.2$  Da, corresponding to a decrease of 2 Da compared with the expected molecular mass for Srx (13,722.7 Da, see Table 1) was observed in *peaks 1* and *3* when wild-type Srx reacts with wild-type and C171A PrxSO<sub>2</sub>, respectively (Fig. 3, A and B). Such an intramolecular disulfide bond was not detected for the reaction of C48A-C106A Srx double mutant with C171A PrxSO<sub>2</sub>. Alternatively, the hypothesis of an intramolecular attack of a non-catalytic Cys on the putative cysteinyl-phosphate intermediate (Fig. 1B) cannot be definitely ruled out. The highly reactive sulfenic species PrxC<sub>p</sub>SOH then rapidly evolved (within 30-s incubation, data not shown) into dimeric disulfide Prx species (*peak 2*, Fig. 3A) in the presence of the regeneration Cys C<sub>R</sub> (Fig. 5A). Indeed, this reaction is part of

*Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism*



**FIGURE 5. Proposed scenario for the evolution of the thiosulfinate species.** Secondary reactions are deduced from SDS-PAGE, immunoblot, and LC-MS analysis of the reaction of wild-type Srx with wild type and C171A PrxSO<sub>2</sub> (A) and of C48A-C106A mutant Srx with C171A mutant PrxSO<sub>2</sub> (B), run in the conditions of Fig. 2.

the catalytic cycle of the peroxidase, at rates at least 50 times higher than Srx (rate constants of  $2 \text{ s}^{-1}$  versus  $1.8 \text{ min}^{-1}$ ). When the regeneration Cys of Prx is mutated, the residual free thiols are carried only by Srx, which explains the formation of intermolecular mixed disulfide PrxS-SSrx as secondary species (peak 4, Fig. 3B). If hydrolysis of the thiosulfinate is significant during the time scale of the reaction, release of reduced Srx could provide the free thiol groups (Fig. 5A). Two arguments support the later proposition: first, the disulfide PrxS-SSrx species is labeled by IAM and thus cannot arise from reaction of the remaining Cys of oxidized Srx released, and second, there is formation of a significant amount of reduced Prx (Fig. 3B, peak 6, and Table 1) in less than 30 s (data not shown). This probably results from an intramolecular attack of a thiol of Srx on the PrxS-SSrx disulfide linkage, similar to the reaction taking place within the thiosulfinate intermediate (Fig. 5A). Therefore, no significant reaction would occur between the released Prx sulfenate and the remaining thiol of oxidized Srx, which suggests that recognition between both partners is needed for efficient attack of Srx catalytic Cys-84 on Prx Cys C<sub>p</sub>. In addition, a significant hydrolysis of the thiosulfinate complex would also explain why a substantial fraction of PrxSO<sub>2</sub> is left non-reacted, as visible both on SDS-PAGE (Fig. 2, B and C) and by LC-MS (Fig. 3, peaks 2, 5, and 9, and Table 1).

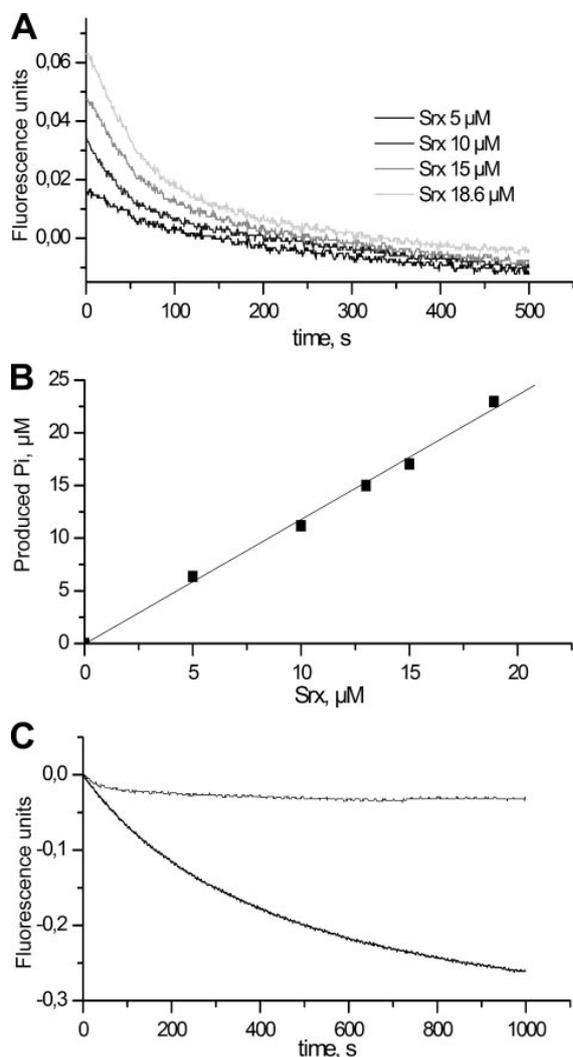
**Kinetic Analysis of the Formation of the Prx-Srx Species**—Because all products of the reactions between the wild-type and mutant forms of both Srx and PrxSO<sub>2</sub> can be explained by the formation of a thiosulfinate PrxSO-SSrx complex, the question arises of the relevance of this species as an intermediate in the catalytic pathway. We addressed this question by a kinetic approach for real-time monitoring of the formation of this species by the production of P<sub>i</sub>. For this purpose, we used the PNP

coupled assay, which monitors the decrease of fluorescence upon phosphorolysis of <sup>m7</sup>Guo. The assay was validated in the steady-state conditions by comparison with the Trx/NTR coupled assay in the presence of Trx as the reducing agent, which gave a rate constant of  $1.7 \pm 0.2 \text{ min}^{-1}$  for wild-type PrxSO<sub>2</sub>, similar to  $1.9 \text{ min}^{-1}$  for the Trx/NTR assay using saturating/optimal concentrations of PrxSO<sub>2</sub>, ATP, MgCl<sub>2</sub>, and Trx. This value was also confirmed by a method based on the titration of the product of the Srx reaction, *i.e.* reduced Prx, by following the regeneration of its peroxidase activity in the presence of 10 mM DTT as electron donor, which gave a slightly slower steady-state rate constant of  $1.5 \pm 0.1 \text{ min}^{-1}$ .

Each of the reactions analyzed previously by Western blot and LC-MS was followed by the PNP assay. Single turnover conditions were used, *i.e.* absence of reducer and excess PrxSO<sub>2</sub> relative to Srx, to allow the determination of the rate constant and stoichiometry of formation of the intermediate relative to Srx. If the formation of the intermediate is part of the catalytic pathway, the rate constant of the reaction is expected to be at least equal to the steady-state rate constant, and the amplitude of the reaction should be equal to the limiting concentration, *i.e.* Srx. Indeed, depending on the relative rates of the secondary reactions following the formation of thiosulfinate, the species that accumulates will be either the intermediate itself, or an oxidized form of Srx (Fig. 5). Furthermore, one advantage of this method is to “filter off” secondary reactions such as formation of the disulfide species Prx-Srx, which are not monitored by the PNP assay.

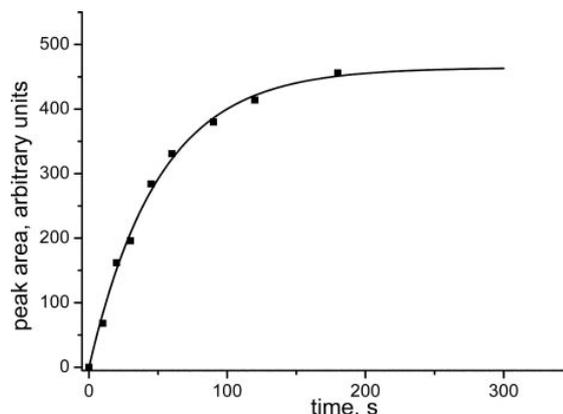
The kinetics recorded for the reaction of wild-type Srx with wild-type and C171A PrxSO<sub>2</sub> gave very similar results. In the absence of Srx, a nonspecific linear phosphate release was measured, at a rate constant of  $\sim 0.015 \text{ min}^{-1}$ . After subtrac-

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism



**FIGURE 6. Kinetics of the reaction catalyzed by wild-type Srx with wild-type PrxSO<sub>2</sub> monitored by P<sub>i</sub> release.** *A*, the reaction of 100 μM wild-type PrxSO<sub>2</sub> with variable concentrations of wild-type Srx (5, 10, 15, and 18.6 μM) in the presence of 1 mM ATP/MgCl<sub>2</sub> was monitored on a rapid kinetics spectrofluorometer by the PNP coupled assay (see text). Reactions were carried out in TK buffer at 30 °C. The progress curves are corrected from blank traces collected in the absence of Srx. *B*, stoichiometry of P<sub>i</sub> produced as a function of Srx concentration deduced from the amplitude of the time courses from *panel A* analyzed as monoexponential processes, and from a calibration curve of the fluorescence signal against P<sub>i</sub> established in identical conditions. *C*, comparison of the kinetics of P<sub>i</sub> release for the reaction of wild-type Srx (5 μM) with 100 μM wild-type PrxSO<sub>2</sub> in the presence of 1 mM ATP/MgCl<sub>2</sub>, in the presence (*thick line*) or absence (*thin line*) of 300 μM Trx.

tion from the traces collected in the same conditions in the presence of increasing concentrations of Srx, the resulting progress curves were best described by a monoexponential equation, corresponding to a first order process, characterized by a rate constant of  $1.8 \pm 0.7 \text{ min}^{-1}$  independent of Srx concentration (Fig. 6*A*). This rate constant is equal to the steady-state rate constant of the Srx-catalyzed reaction measured in the presence of a thiol reducer ( $1.7 \pm 0.2 \text{ min}^{-1}$ ). Furthermore, the fluorescence response of <sup>m7</sup>Guo phosphorylation was calibrated against phosphate (see "Experimental Procedures"), which



**FIGURE 7. Kinetics of the reaction catalyzed by C48A-C106A Srx with C171A PrxSO<sub>2</sub> monitored by reversed-phase chromatography.** The reaction of 100 μM C171A PrxSO<sub>2</sub> with 20 μM C48A-C106A Srx in the presence of 1 mM ATP and MgCl<sub>2</sub> at 30 °C was followed by analysis of aliquots of the reaction mixture quenched by acidification in trifluoroacetic acid 0.1%, on reversed-phase chromatography with the same gradient as in Fig. 3, monitored by absorption spectrophotometry at 215 nm. The evolution of the area of the peak corresponding to the Prx-Srx complexes (see Fig. 3*B*, peak 8) is plotted as a function of time (*dots*) and analyzed by a monoexponential kinetic model (*solid line*) with a rate constant of  $1.2 \pm 0.1 \text{ min}^{-1}$ .

allowed the analysis of the amplitude of the process as a function of Srx concentration. As shown on Fig. 6*B*, the reaction occurs with a stoichiometry of  $0.96 \pm 0.16 \text{ mol}$  of P<sub>i</sub> released per mol of Srx. These results support the accumulation of an intermediate species, the stoichiometry of which is limited by Srx concentration, at a rate compatible with catalysis. On the contrary, the amount of P<sub>i</sub> released in the presence of an excess of Trx was compatible with the initial concentration of PrxSO<sub>2</sub> in the assay (Fig. 6*C*). Therefore, formation of the thiosulfinate species is strongly supported as a specific step on the catalytic pathway. In addition, the similarity between the rate constant measured in single turnover conditions and in steady-state conditions indicates that the rate-limiting step of the overall reaction occurs before or at this step. The kinetics of the reaction of C48A-C106A Srx with C171A PrxSO<sub>2</sub> revealed a similar behavior in the absence of reducer, with a rate constant of P<sub>i</sub> release of  $2.0 \pm 0.1 \text{ min}^{-1}$  independent of Srx concentration (data not shown). A slightly higher stoichiometry of P<sub>i</sub> release of 1.3 mol of P<sub>i</sub> released per mol of Srx was observed. This supports hydrolysis of the thiosulfinate intermediate, as hypothesized above, which would result in 30% additional catalytic cycles in addition to the secondary reactions on the thiosulfinate intermediate already described (Fig. 5*B*).

To confirm the catalytic intermediacy of the thiosulfinate species, we followed the kinetics of evolution of the area of the chromatographic peak corresponding to peak 8 (Fig. 3*C*), after reaction of C48A-C106A Srx with C171A PrxSO<sub>2</sub> in the presence of ATP, monitored by UV spectrophotometry. Indeed, we showed that this peak contains two species, the putative thiosulfinate PrxSO-SSrx and the disulfide PrxS-SSrx which, assuming the scenario proposed in Fig. 5*B*, should derive from the secondary reactions of the thiosulfinate species. Therefore, the kinetics of evolution of the area of this peak should approximate the kinetics of production of the PrxSO-SSrx species. As shown in Fig. 7, this follows monoexponential kinetics with a

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

rate constant of  $1.2 \pm 0.1 \text{ min}^{-1}$  close to the rate of  $P_i$  release. Such results again support the role of the Prx-Srx species with a mass corresponding to a thiosulfinate as a catalytic intermediate. The slightly lower rate constant measured by chromatography can probably be attributed to the evolution of this reactive species by the side reactions described Fig. 5B.

This interpretation raises the problem of the compatibility of the results obtained on *S. cerevisiae* (this study) and mammalian Srx (17). In their study on human Srx, Jeong *et al.* did not detect an Srx-Prx thiosulfinate species, but a disulfide species between the human Srx catalytic Cys and the regeneration Cys  $C_R$  of the Prx. The Prx protein exists in solution as an obligate dimer, or decamer. Because experiments were conducted on the wild-type substrate PrxSO<sub>2</sub>, this species could result from the attack of Cys  $C_R$  from the symmetrical Prx monomer on the sulfenyl sulfur of the thiosulfinate bond, a reaction favored by the existence of the complex. After a 10-min reaction, the greater part of the thiosulfinate intermediate would have reacted to give the disulfide species observed. In addition, it would release Prx with a Cys  $C_P$  as a sulfenate, which can react with the catalytic Cys-99 of another human Srx molecule to eventually produce the Prx-(Srx)<sub>2</sub> species, the second most abundant form in the experiment shown in Jeong *et al.*, study (17). This interpretation was supported by a control experiment performed on the equivalent reaction for the *S. cerevisiae* proteins, using C48A-C106A Srx and wild-type PrxSO<sub>2</sub> for 5 min in the presence of ATP/MgCl<sub>2</sub>. LC-MS analyses revealed a complex mix, including three chromatographic peaks in addition to the peaks of Srx and Prx, corresponding to mixed disulfide species formed between Prx and Srx (1:1 and Srx 1:2 complexes) and PrxSO<sub>2</sub> and Srx (1:1 complex) (data not shown). The authors also reported that these species were formed nonspecifically between Srx and reduced Prx, a result that we also observed in the yeast system (data not shown). However, for the reaction between the substrate PrxSO<sub>2</sub> and *S. cerevisiae* Srx, no Prx-Srx species was produced in the absence of ATP (Fig. 2), clearly supporting that the thiosulfinate species observed in this study is specifically formed along the catalytic mechanism.

Kinetics of  $P_i$  release for the reaction of human Srx and sulfinic Prx in the absence of reducer were characterized by a stoichiometry largely higher than Srx concentration (17). They were described as a linear phase interpreted as a futile cycle of hydrolysis of the phosphorylated substrate, followed by inactivation of Srx from the nonspecific reaction between the catalytic Cys and Cys  $C_R$  of the substrate. However, as proposed above, thiosulfinate hydrolysis would also lead to a futile cycle of recycling of the sulfinic acid substrate and Srx. The kinetics of  $P_i$  release for human Srx in the absence of reducer could correspond to hydrolysis. As shown for the yeast system, secondary reactions of hydrolysis and reaction with thiols could compete with the slow formation of the thiosulfinate intermediate (the rate constant for the reaction catalyzed by mammalian Srx is reported to be of  $0.2\text{--}0.5 \text{ min}^{-1}$  (16, 17)). Therefore, for human Srx, either the rate of formation of the thiosulfinate intermediate would be too slow to be resolved from the process of thiosulfinate hydrolysis, or the technique of  $P_i$  detection on the hour time scale employed by Jeong *et al.* could not resolve

the kinetics of formation of the thiosulfinate (on the minute time scale for yeast Srx) and of secondary reactions.

### CONCLUSION

In the present study, the role of Srx as a reductase has been proven by three types of arguments from the analysis of the Srx reaction in the absence of added reducers: first, the observation of the transient formation of a DTT-reducible species with a mass compatible with the thiosulfinate intermediate postulated in Fig. 1A, second, the analysis of the different species formed in secondary reactions, which implies obligatory transit through such an intermediate, and finally, the kinetics and stoichiometry of the reaction indirectly monitored in real-time by the release of  $P_i$ , or directly by the evolution of the Prx-Srx complex chromatographic peak, which authenticated this species as a catalytic intermediate. Such an interpretation can probably be extrapolated to the catalytic mechanism of Srx from mammals for which no such results were obtained in a previous study by Jeong *et al.* (17), likely due to secondary reactions of the regeneration Cys of the Prx substrate on the thiosulfinate intermediate.

This interpretation implies the attack of Srx catalytic Cys on the phosphorylated Prx substrate, which therefore requires precise recognition and positioning of the reacting Cys from both partners at distances compatible with the chemical steps. Observations from this study, and substrate specificity of Srx for the class of the typical 2-Cys-Prx (6), are in agreement with these conclusions. Furthermore, when this report was under submission, the structure of an artificial human PrxS-SSrx disulfide complex was published (30). This structure strongly supports the formation of a covalent Prx-Srx intermediate involving Prx Cys  $C_P$  and Srx catalytic Cys, through the local unfolding of Prx active site triggered by Srx binding. The complex is also stabilized by an interaction with the unfolded C-terminal tail of Prx acting as an embrace onto the Srx molecule.

In addition, catalysis of the ATP-dependent activation of PrxSO<sub>2</sub> as a phosphorylated intermediate remains to be elucidated for *S. cerevisiae* Srx. In this regard, the hypothesis of the existence of a phosphorylated Srx intermediate (Fig. 1B) is not supported by the structural data on the PrxS-SSrx complex (30). The importance of an Asn residue in position 186 from the Prx was suggested for the human enzyme (31). The role of the corresponding Thr residue in Prx will require further investigation at the chemical level. Finally, the results obtained with wild-type Srx from *S. cerevisiae*, showing the release of oxidized Srx with intramolecular disulfide bond, raise the question of the potential role of the non-catalytic Cys residues of Srx Cys-48 and Cys-106, which are not conserved in human Srx.

*Acknowledgments*—We thank J. Ugolini and S. Alaoui for excellent technical assistance and Prof. S. Boschi-Muller, Dr. A. Gruetz, Dr. H. Mazon, and Dr. C. Schaeffer-Reiss for helpful discussions.

### REFERENCES

1. Biteau, B., Labarre, J., and Toledano, M. B. (2003) *Nature* **425**, 980–984
2. Budanov, A. V., Sablina, A. A., Feinstein, E., Koonin, E. V., and Chumakov, P. M. (2004) *Science* **304**, 596–600
3. Rhee, S. G., Kang, S. W., Jeong, W., Chang, T. S., Yang, K. S., and Woo,

## Prx-Srx Thiosulfinate Intermediate in the Srx Mechanism

- H. A. (2005) *Curr. Opin. Cell Biol.* **17**, 183–189
- Sundaresan, M., Yu, Z. X., Ferrans, V. J., Irani, K., and Finkel, T. (1995) *Science* **270**, 296–299
  - Rhee, S. G. (2006) *Science* **312**, 1882–1883
  - Woo, H. A., Jeong, W., Chang, T. S., Park, K. J., Park, S. J., Yang, J. S., and Rhee, S. G. (2005) *J. Biol. Chem.* **280**, 3125–3128
  - Kang, S. W., Rhee, S. G., Chang, T. S., Jeong, W., and Choi, M. H. (2005) *Trends Mol. Med.* **11**, 571–578
  - Chang, T. S., Jeong, W., Choi, S. Y., Yu, S., Kang, S. W., and Rhee, S. G. (2002) *J. Biol. Chem.* **277**, 25370–25376
  - Wood, Z. A., Poole, L. B., and Karplus, P. A. (2003) *Science* **300**, 650–653
  - Vivancos, A. P., Castillo, E. A., Biteau, B., Nicot, C., Ayte, J., Toledano, M. B., and Hidalgo, E. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8875–8880
  - Jang, H. H., Lee, K. O., Chi, Y. H., Jung, B. G., Park, S. K., Park, J. H., Lee, J. R., Lee, S. S., Moon, J. C., Yun, J. W., Choi, Y. O., Kim, W. Y., Kang, J. S., Cheong, G. W., Yun, D. J., Rhee, S. G., Cho, M. J., and Lee, S. Y. (2004) *Cell* **117**, 625–635
  - Findlay, V. J., Townsend, D. M., Morris, T. E., Fraser, J. P., He, L., and Tew, K. D. (2006) *Cancer Res.* **66**, 6800–6806
  - Liu, X. P., Liu, X. Y., Zhang, J., Xia, Z. L., Liu, X., Qin, H. J., and Wang, D. W. (2006) *Cell Res.* **16**, 287–296
  - Glauser, D. A., Brun, T., Gauthier, B. R., and Schlegel, W. (2007) *BMC Mol. Biol.* **8**, 54–67
  - Diet, A., Abbas, K., Bouton, C., Guillon, B., Tomasello, F., Fourquet, S., Toledano, M. B., and Drapier, J.-C. (2007) *J. Biol. Chem.* **282**, 36199–36205
  - Chang, T. S., Jeong, W., Woo, H. A., Lee, S. M., Park, S., and Rhee, S. G. (2004) *J. Biol. Chem.* **279**, 50994–51001
  - Jeong, W., Park, S. J., Chang, T. S., Lee, D. Y., and Rhee, S. G. (2006) *J. Biol. Chem.* **281**, 14400–14407
  - Mulrooney, S. B. (1997) *Protein Expr. Purif.* **9**, 372–378
  - Mossner, E., Huber-Wunderlich, M., and Glockshuber, R. (1998) *Protein Sci.* **7**, 1233–1244
  - Miroux, B., and Walker, J. E. (1996) *J. Mol. Biol.* **260**, 289–298
  - Studier, F. W. (2005) *Protein Expr. Purif.* **41**, 207–234
  - Banik, U., and Roy, S. (1990) *Biochem. J.* **266**, 611–614
  - Kice, J. L., and Liu, C.-C. A. (1979) *J. Org. Chem.* **44**, 1918–1923
  - Kice, J. L., and Rogers, T. E. (1974) *J. Am. Chem. Soc.* **96**, 8015–8019
  - Chae, H., Uhm, T., and Rhee, S. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7022–7026
  - Veniamin, N., Lapko, D. L. S., and Smith, J. B. (2000) *J. Mass Spectrom.* **35**, 572–575
  - Lundell, N., and Schreitmuller, T. (1999) *Anal. Biochem.* **266**, 31–47
  - Nagy, P., and Ashby, M. T. (2007) *Chem. Res. Toxicol.* **20**, 1364–1372
  - Kice, J. L., and Rogers, T. E. (1974) *J. Am. Chem. Soc. U. S. A.* **96**, 8009–8015
  - Jonsson, T. J., Johnson, L. C., and Lowther, W. T. (2008) *Nature* **451**, 98–101
  - Lee, D. Y., Park, S. J., Jeong, W., Sung, H. J., Oho, T., Wu, X., Rhee, S. G., and Gruschus, J. M. (2006) *Biochemistry* **45**, 15301–15309

## Supplemental data for

# EVIDENCE FOR THE FORMATION OF A COVALENT THIOSULFINATE INTERMEDIATE WITH PEROXIREDOXIN IN THE CATALYTIC MECHANISM OF SULFIREDOXIN

by

Xavier Roussel, Guillaume Béchade, Alexandre Kriznik, Alain Van Dorsselaer, Sarah Sanglier-Cianferani, Guy Branlant and Sophie Rahuel-Clermont

### Methods for supplemental data:

#### *NanoLC-MS/MS experiments*

*Samples preparation*- Peaks containing complexes of interest were collected after LC-MS (see above). Acetonitrile was removed from these samples using Savant Speedvac Plus (ThermoQuest, San Jose, CA, USA). For reduction and alkylation treatment, the pH was first adjusted to 7.5 by addition of 250  $\mu$ l of ammonium hydrogen carbonate (25 mM) followed by addition of 100  $\mu$ l of 10 mM dithiothreitol (1 h at 60°C) and 55 mM IAM (1 h at room temperature).

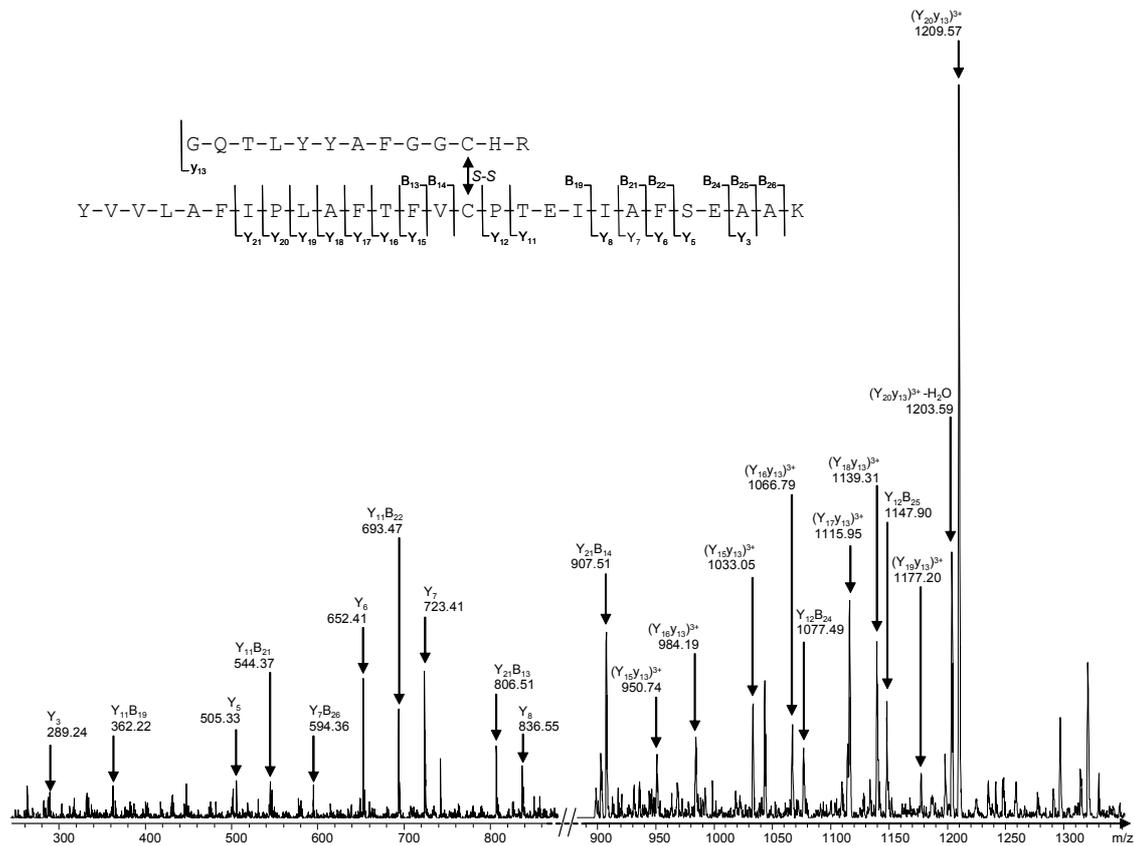
*Digestion step*- Trypsin digestion (Promega, Madison, WI, USA) was performed at pH 7.5 in 25 mM ammonium hydrogen carbonate, over 4 h at 37°C. Digestion was stopped by addition of formic acid. Pepsin digestion (Roche Diagnostics, Mannheim, Germany) was conducted at pH 1.6 in 0.11 M phosphoric acid over 1 h at 4°C. NanoLC-MS/MS analyses were performed immediately after pepsin digestion.

*NanoLC-MS/MS analyses*- NanoLC-MS/MS analyses were performed on an Agilent 1100 Series capillary LC system (Agilent Technologies, Santa Clara, CA, USA) coupled with a HCT Ultra Ion Trap (Bruker Daltonics, Bremen, Germany). Coupling was ensured by HPLC-Chip Cube interface (Agilent Technologies, Santa Clara, CA, USA) loaded with a HPLC-Chip. The chip comprised a 40 nl enrichment column and a 75  $\mu$ m x 150 mm separation column packed with Zorbax 300SB-C18 5  $\mu$ m material.

Aliquots of 4  $\mu$ l of each sample were loaded on the precolumn, the gradient was set at a 300  $\text{nl min}^{-1}$  flow rate and with a linear profile from solvent A (92%  $\text{H}_2\text{O}/1\%$  formic acid, v/v) to solvent B (40% acetonitrile/1% formic acid, v/v) over 45 min. A second ramp up to 70% of B followed over 4 min. The capillary and cone voltage were set to 1850 V and -500 V respectively. Automatic switching between MS and MS/MS modes was used. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares.

NanoLC-MS/MS accurate mass measurements were performed on a nanoAquity HPCL system coupled to a Synapt HDMS mass spectrometer (Waters, Manchester, UK) and equipped with lock-spray. Aliquots of 4  $\mu$ l of each sample were loaded on a 5  $\mu$ m Symmetry C18 180  $\mu$ m x 20 mm precolumn coupled to a 1.7  $\mu$ m BEH130 C18 180  $\mu$ m x 20 mm separation column. The gradient was set at a 400  $\text{nl min}^{-1}$  flow rate and with a linear profile from solvent A (99%  $\text{H}_2\text{O}/0.1\%$  formic acid, v/v) to solvent B (50% acetonitrile/0.1% formic acid, v/v) over 35 min. A second ramp up to 90% of B followed over 4 min. The capillary and cone voltage were set to 3.3 kV and 35 V respectively. The complete system was fully controlled by MassLynx and Synapt HDMS software (Waters).





**Suppl. Figure 2:** Tandem mass spectrum (nanoLC-Chip-IT) of the  $[M+5H]^{5+}$  precursor ion at  $m/z$  887.65 of the disulfide-bonded (Prx Cys C<sub>P</sub> - Srx C84) dipeptide Prx 54-80 and Srx 74-86 obtained after trypsin digestion without prior reduction/alkylation treatment. The major fragment ion  $(Y_{20}Y_{13})^{3+}$  and the subsequent amino acid tags unambiguously attest the existence of the PrxS-SSrx species.

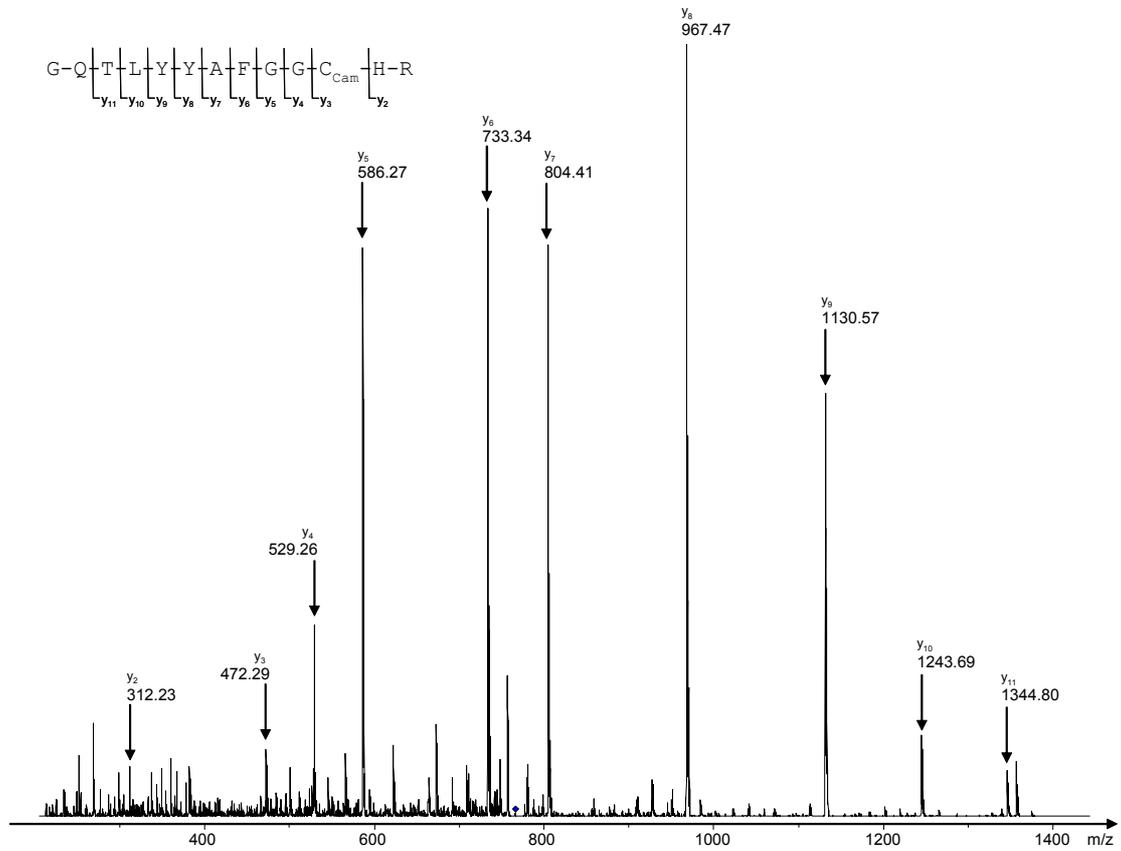
**C48A-C106A Srx**

SLQS NSVKP TEIPL SEIRR PLAPV LDPOK IDAMV ATMKG IPTAS KTASL EQAEA  
AASAG ELPPV DVLGV RVKGQ TLYYA FGGCH RLQAY DRRAR ETQNA AFPVR ARVLP  
ATPRQ IRMYL GSSLD IE

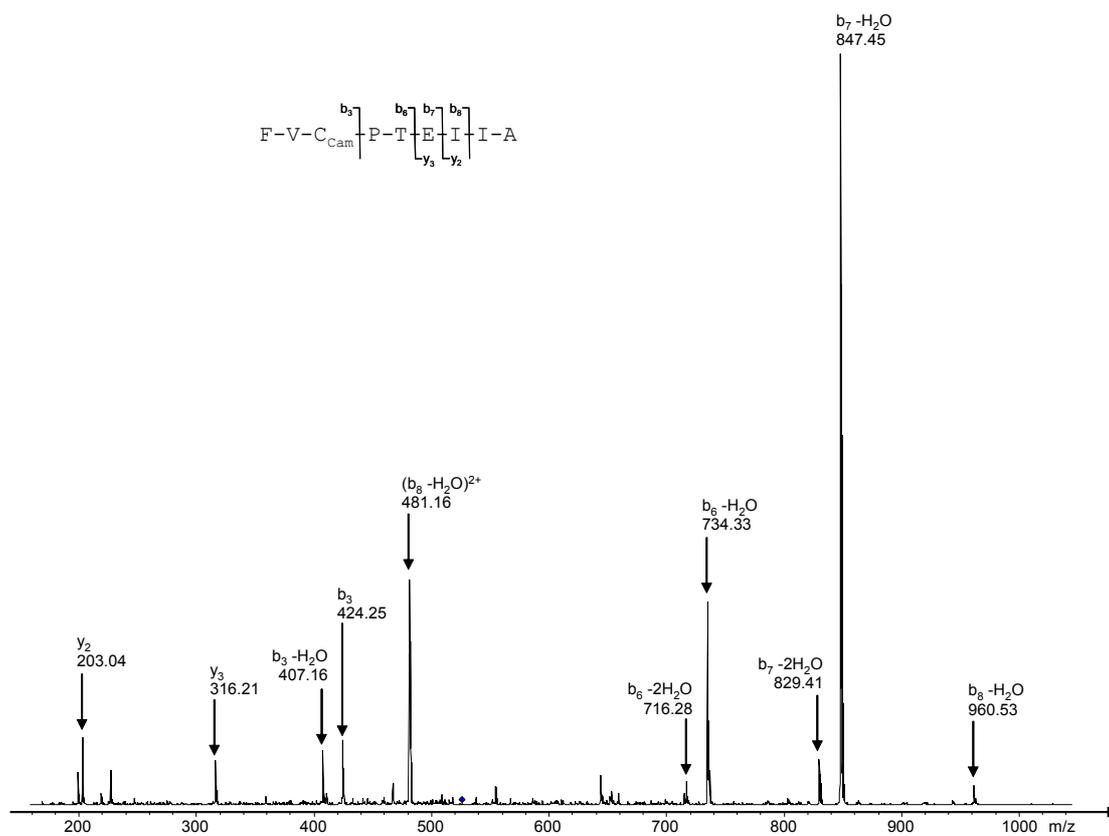
**C171A Prx**

GSSH HHHHH SSGLV PRGSH MVAQV OKQAP TFKKT AVVDG VFDEV SLDKY KGKYV  
VLAFI PLAFT FVCP EIIAF SEAAK KFEEQ GAQVL FASTD SEYSL LAWTN IPRKE  
GGLGP INIPL LADTN HLSLR DYGVL IEEEG VALRG LFIID PKGVI RHITI NDLPV  
GRNVD EALRL VEAFQ WTDKN GTVLP ANWTP GAATI KPTVE DSKEY FEAAN K

**Suppl. Figure 3:** Sequence coverage obtained by nanoLC-MS/MS analyses after trypsin and pepsin digestions without prior reduction/alkylation treatment. Bold and underlined tags correspond to peptides recovered in trypsin and pepsin digested samples, respectively. A fraction of 88% of the total Prx-Srx complex sequence has been recovered. Peptides including cysteins of interest (labelled with \*) have not been detected in any case, inferring a link between these two Cys.



**Suppl. Figure 4:** Tandem mass spectrum (nanoLC-Chip-IT) of the  $[M+2H]^{2+}$  precursor ion at  $m/z$  765.92 of the carbamidomethylated peptide SrX 74-87 obtained after trypsin digestion with reduction-alkylation treatment. Major fragment ions correspond to  $y_n$  ( $n=2-11$ ) sequence tag.



**Suppl. Figure 5:** Tandem mass spectrum (nanoLC-Chip-IT) of the  $[M+2H]^{2+}$  precursor ion at  $m/z$  525.58 of the carbamidomethylated peptide Prx 66-74 obtained after pepsin digestion with prior reduction-alkylation treatment. Major fragment ions correspond to  $b_n - H_2O$  ( $n=3-8$ ) sequence tag.



# Mise en évidence par spectrométrie de masse de la formation d'un intermédiaire thiosulfinate entre la peroxyredoxine et la sulfiredoxine.

Guillaume Béchade<sup>1</sup>; Xavier Roussel<sup>2</sup>; Sophie Rahuel-Clermont<sup>2</sup>; Guy Branlant<sup>2</sup>; Alain Van Dorselaer<sup>1</sup>; Sarah Sanglier<sup>1</sup>

<sup>1</sup>Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC, ULP, CNRS, UMR 7178  
25 rue Bequerel 67087 Strasbourg Cedex 2, France

<sup>2</sup>Maturation des ARN et Enzymologie Moléculaire, Faculté des Sciences, UHP, CNRS, UMR 7567  
BP 239 Vandœuvre-les-Nancy Cedex, France

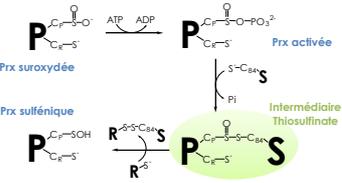
Contact  
gbechade@chimie.u-strasbg.fr

## Suroxydation des peroxyredoxines



Wood et al., 2003, Science, 300

## Réduction de Prx Hypothèse de mécanisme



P. Prx, S: Srx, R: réducteur

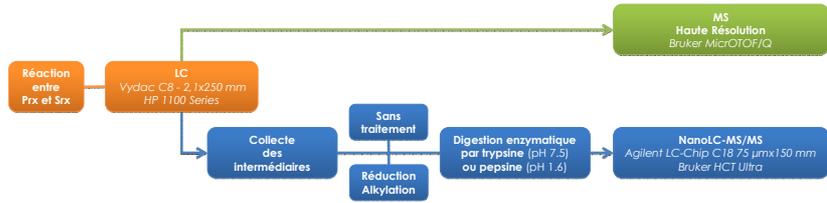
Biteau et al., Nature 2003, 425

## Objectifs

Caractériser totalement les espèces formées entre la sulfiredoxine et la peroxyredoxine suroxydée grâce à une combinaison de techniques en Spectrométrie de Masse :

- Déterminer les conditions idéales pour observer l'intermédiaire thiosulfinate supposé ;
- Optimiser les conditions d'analyse LC-MS afin de caractériser la structure des multiples produits de réaction entre les espèces Srx et Prx ;
- Déterminer précisément la position du pont et de l'oxydation sur l'intermédiaire en optimisant la collecte des espèces, leur digestion et leur analyse MS.

## Stratégies de caractérisation



## Détection des complexes covalents Prx-Srx

Conditions expérimentales :

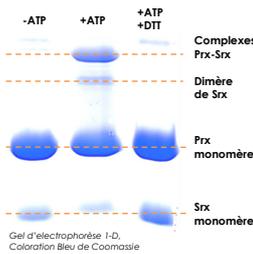
Prx-SO<sub>2</sub> + Srx + MgCl<sub>2</sub> dans tampon (Tris 50 mM, KCl 100 mM) pH 7.0 incubés 30 secondes à 30°C

Afin d'éliminer les réactions parasites :

- Les Cys non réactives ont été mutées en Ala, donnant les mutants PrxC171A et SrxC48A/C106A, contenant chacun une Cys unique ;
- Les réactions ont été bloquées par ajout soit d'iodoacétamide, soit d'acide trifluoroacétique.

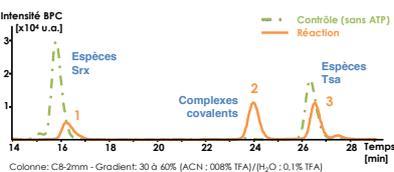
## Gel d'électrophorèse 1-D

Gel d'électrophorèse SDS-Page non réducteur : les espèces formées par la réaction catalytique entre PrxSO<sub>2</sub> et Srx ont été identifiées grâce à des analyses immunoblot.



- Les complexes covalents formés en présence d'ATP sont réduits en présence de DTT.
- La formation d'un dimère de Srx en absence de réducteur indique l'existence probable de réactions secondaires.

## Analyses LC-MS des complexes



Mesures de masses obtenues par analyse LC-MS du milieu réactionnel

Pic #	Espèces	MM (Da) observées	MM (Da) attendues
1	Srx	- +IAM 13 658.6 ± 0.1 13 715.5 ± 0.1*	13 658.6 13 715.6
	Dimère de Srx	Srx-S-Srx +IAM +2IAM 27 316.6 ± 0.7 27 430.6 ± 0.5 27 543.6 ± 0.7	27 315.2 27 429.3 27 543.4
2	Complexes Prx-Srx	Pont disulfure Pont thiosulfinate 37 246.3 ± 0.4 37 261.7 ± 0.4	37 246.1 37 262.1
	Prx oxydée	PrxSO <sub>2</sub> PrxSO <sub>3</sub> 23 621.9 ± 0.4* 23 637.6 ± 0.3*	23 621.5 23 637.5
3	Prx réduite	- 23 589.9 ± 0.4	23 589.5

\* Observées également dans les expériences de contrôle (sans ATP)

- Hypothèse sur les produits de réactions :  
Coélution des deux complexes Prx-Srx à pont thiosulfinate et disulfure
- Spectrométrie de masse à haute résolution requise

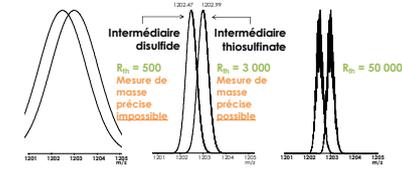
## MS Haute Résolution sur les espèces Prx-Srx

Un pouvoir résolutif théorique minimal de 2300 (FWHM) est requis afin de distinguer entre les complexes à pont thiosulfinate et disulfure :



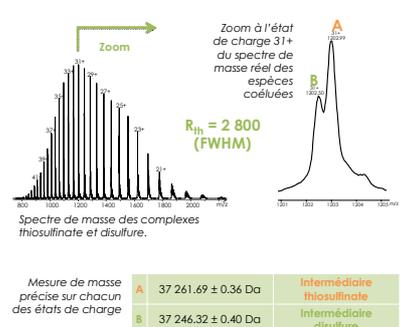
## Optimisation instrumentale

- Avec une résolution trop basse, les 2 espèces coéluees ne peuvent pas être distinguées.
- Simulations des spectres des complexes thiosulfinate et disulfure à l'état de charge 31+, pour différents pouvoirs résolutifs (FWHM) :



- Notre spectromètre Bruker MicroTOF-Q a été finement paramétré pour atteindre à 1 200 m/z un pouvoir résolutif de :  
  - R = 17 000 (FWHM) pour une molécule monochargée de 1 222 Da ;
  - R = 2 800 (FWHM) pour un complexe protéique de 37 kDa.

## Analyse des complexes par LC-MS Haute Résolution



- Un spectre de masse correctement résolu des complexes thiosulfinate et disulfure a pu être obtenu ;
- Une mesure précise des deux espèces a pu être réalisée.

## Caractérisation des complexes

Afin de déterminer précisément le position des ponts et de caractériser les complexes Prx-Srx, les intermédiaires collectés ont été soumis à une digestion par la trypsine ou la pepsine.

La fragmentation de dipeptides thiosulfinate n'est pas connue dans la littérature et aucun spectre de référence n'est disponible.

Trypsine	Pepsine
<ul style="list-style-type: none"> <li>pH 7.5: pont thiosulfinate instable</li> <li>Clivages spécifiques</li> <li>Fragmentation MS/MS hasardeuse</li> </ul>	<ul style="list-style-type: none"> <li>pH 1.6: pont thiosulfinate plus stable</li> <li>Clivages spécifiques</li> <li>Fragmentation MS/MS efficace</li> </ul>

## Caractérisation indirecte du complexe Prx-Srx thiosulfinate

Aucun dipeptide thiosulfinate n'a été directement séquencé par nanoLC-MS/MS.

Information de séquence obtenue par nanoLC-MS/MS après digestion par la trypsine et la pepsine sans réduction/alkylation : tous les peptides, sauf ceux supposés être impliqués dans le pont, ont été détectés.

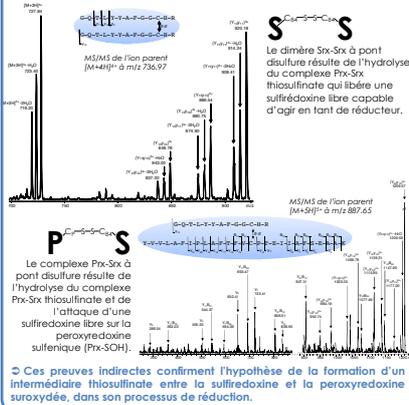
Trypsine	Pepsine
<ul style="list-style-type: none"> <li>pH 7.5: pont thiosulfinate instable</li> <li>Clivages spécifiques</li> <li>Fragmentation MS/MS hasardeuse</li> </ul>	<ul style="list-style-type: none"> <li>pH 1.6: pont thiosulfinate plus stable</li> <li>Clivages spécifiques</li> <li>Fragmentation MS/MS efficace</li> </ul>

- Traitement de réduction & d'alkylation :  
  - Peptide Srx 74-87 séquencé avec IAM-Cys (digestion / trypsine) ;
  - Peptide Prx 66-74 séquencé avec IAM-Cys (digestion / pepsine).

Ces indices confirment la position du pont entre les 2 protéines :  
La Cys68 de Prx est liée à la Cys84 de Srx.

## Séquençage de dipeptides

Des dipeptides à pont disulfure ont été clairement séquencés dans le cas d'une digestion par la trypsine. Ces produits secondaires ne peuvent être expliqués que par l'hydrolyse du complexe thiosulfinate.



- Ces preuves indirectes confirment l'hypothèse de la formation d'un intermédiaire thiosulfinate entre la sulfiredoxine et la peroxyredoxine suroxydée, dans son processus de réduction.

## Chapitre III

# Etude de mécanismes de régénération de Méthionine Sulfoxyde Réductases

---

Ce projet s'inscrit dans le cadre d'une collaboration avec l'équipe du Professeur Guy Branlant, du laboratoire « ARN-RNP, Enzymologie Moléculaire et Structurale » (AREMS) de l'Université Henri Poincaré de Nancy. Sandrine Boschi-Muller, Adeline Gand et Laure Selme-Roussel, travaillent sur le mécanisme de régénération de méthionine sulfoxyde réductases. Nous avons étudié, par une combinaison d'approches en spectrométrie de masse, les intermédiaires covalents à pont disulfure qui sont susceptibles de se former lors de ces réactions.

Ce chapitre complète les éléments déjà introduits précédemment en proposant une brève présentation des objets biologiques utilisés et détaille les résultats obtenus pour l'étude des mécanismes de réduction :

- des méthionine sulfoxyde réductases cytoplasmiques, par les thiorédoxines ;
- de la protéine PilB périplasmique, qui possède un domaine N-ter à l'activité disulfure oxydoréductase et un domaine central et C-ter à l'activité méthionine sulfoxyde réductase A et B, respectivement.

# 1. Etude du mécanisme de réduction des méthionine sulfoxyde réductases par la thiorédoxine

## 1.1. Le rôle des enzymes à activité méthionine sulfoxyde réductase

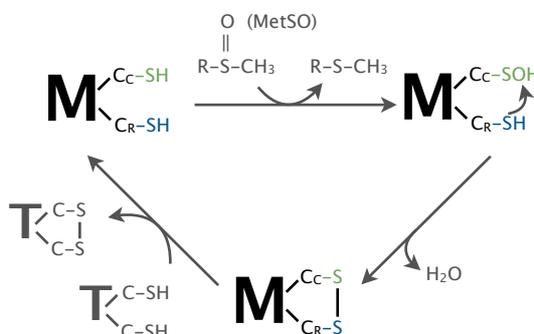
Les méthionines (Met) sont, avec les cystéines, les seuls acides aminés présentant un atome de soufre sur leur chaîne latérale. Les méthionines peuvent subir 2 degrés d'oxydation (Table 1). L'oxydation des méthionine (Met) en méthionine sulfoxyde (MetSO), forme oxydée majoritaire *in vivo*, représente un mécanisme de modification posttraductionnel non enzymatique des protéines. Cette modification est décrite comme état impliquée dans l'étape de reconnaissance des protéines non fonctionnelles à dégrader<sup>1</sup> ou dans des mécanismes de régulation de la fonction de certaines protéines<sup>2</sup>. Elle peut avoir lieu également lors d'un stress oxydant imposé par des espèces réactives de l'oxygène comme H<sub>2</sub>O<sub>2</sub> (cf. point 1.1.1 au chapitre II).

<b>Formule</b>	R-(CH <sub>2</sub> ) <sub>2</sub> -S-CH <sub>3</sub>	R-(CH <sub>2</sub> ) <sub>2</sub> -SO-CH <sub>3</sub>	R-(CH <sub>2</sub> ) <sub>2</sub> -SO <sub>2</sub> -CH <sub>3</sub>
<b>Degrés d'oxydation</b>	-II	0	+II
<b>Nomenclature</b>	Méthionine	Méthionine sulfoxyde	Méthionine sulfone

**Table 1** | Les différentes formes oxydées de la méthionine et le degré d'oxydation de son atome de soufre.

Les méthionines sulfoxydes sont, contrairement aux méthionines sulfones, connues pour être réduites par des enzymes présentant une activité méthionine sulfoxyde réductase comme les Msr<sup>3</sup>. Dans cette classe, les MsrA et MsrB sont connues pour réduire spécifiquement l'isomère S- ou R-MetSO, respectivement. Les membres de ces deux classes ont des structures différentes, mais opèrent selon un même mécanisme catalytique fondé sur la chimie des résidus cystéiques, qui se décompose en au moins en 3 étapes (Figure 1) :

- Les groupements méthionine sulfoxyde sont réduits par les Msr sur la cystéine dite « catalytique » desquelles se forme un acide sulfénique ;
- La fonction thiol d'une cystéine de « recyclage » attaque l'acide sulfénique pour conduire à la formation d'une liaison disulfure entre cystéines ;
- Le pont disulfure est réduit par l'intervention d'une thiorédoxine (enzyme ubiquitaire à activité réductase, figure 2).



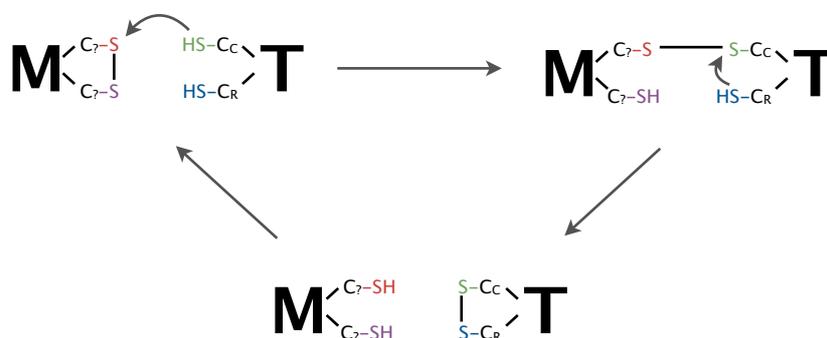
**Figure 1** | Mécanisme catalytique simplifié de la réduction des méthionine sulfoxydes par les méthionine sulfoxyde réductases. (M : méthionine sulfoxyde réductase, T : thiorédoxine.)

<sup>1</sup> Hokenson, M. J., et al. (2004). *Biochemistry* **43**, 4621-33.

<sup>2</sup> Ciorba, M. A., et al. (1999). *FEBS Lett* **442**, 48-52.

<sup>3</sup> Brot, N., et al. (1981). *Proc Natl Acad Sci U S A* **78**, 2155-8.

Le mécanisme de réduction des méthionine sulfoxyde réductases par la thiorédoxine demeure mal connu. La réaction nécessite une étape de reconnaissance entre les 2 protéines ; mais bien que la cystéine catalytique de la thiorédoxine ait été identifiée, on ne sait pas laquelle des deux cystéines de la méthionine sulfoxyde réductase est attaquée en premier (Figure 2).



**Figure 2** | Mécanisme général de réduction des méthionine sulfoxyde réductases par la thiorédoxine. (M : méthionine sulfoxyde réductase, T : thiorédoxine, C<sub>C</sub> : cystéine catalytique, C<sub>R</sub> : cystéine de recyclage)

## 1.2. Stratégie analytique mise en oeuvre pour l'étude du mécanisme de réduction des méthionine sulfoxyde réductases par les thiorédoxines

### 1.2.1. Objectifs de l'étude

L'objectif de cette étude est de déterminer, dans le cas de la méthionine sulfoxyde réductase A d'une part et B d'autre part, quelle est la cystéine impliquée dans le pont disulfure avec la thiorédoxine. Pour cela, il s'agit de :

- mettre au point les conditions idéales favorisant l'accumulation des complexes à caractériser ;
- déterminer une stratégie d'analyse par spectrométrie de masse permettant de discriminer les différents complexes pouvant être formés (ie. présentant une liaison disulfure entre l'une ou l'autre des cystéines des Msr) ;
- déterminer précisément la position du ou des ponts formés en caractérisant chacun des complexes MsrA-Trx et MsrB-Trx.

Le travail de développement qui a été décrit au chapitre I a permis la mise au point de stratégies analytiques par spectrométrie de masse visant à caractériser des complexes covalents à liaison cystéine-cystéine. La suite de ce chapitre décrit l'application de ces approches analytiques à l'étude du mécanisme catalytique de réduction des méthionine sulfoxyde réductases par la thiorédoxine.

### 1.2.2. Formation des complexes Msr-Trx

#### ▸ Conventions de notation

La figure 3 présente les séquences des protéines utilisées pour cette étude et indique les conventions de notation qui seront utilisées par la suite.

- La thiorédoxine possède deux cystéines proches situées en position 32 et 35.
- La méthionine sulfoxyde réductase A possède deux cystéines situées en position 12 et 154.
- La méthionine sulfoxyde réductase B possède deux cystéines situées en position 65 et 120.

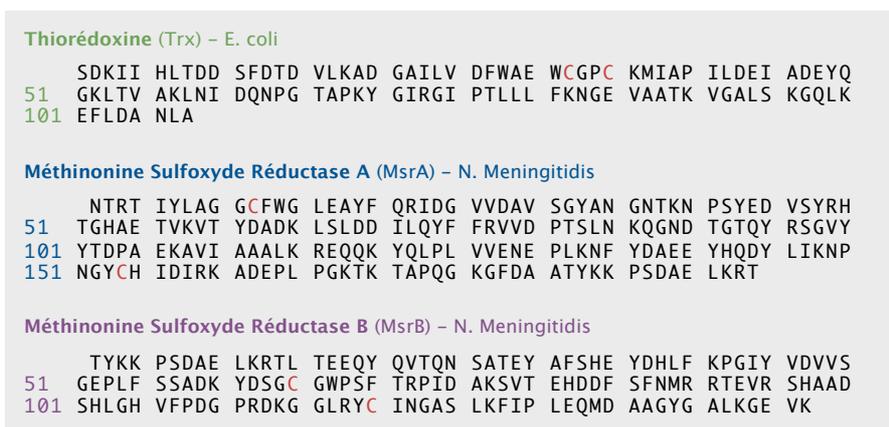


Figure 3 | Séquences de la thiorédoxine (Trx) de *E. coli* et des Méthionine Sulfoxyde Réductases A (MsrA) et B (MsrB) de *N. meningitidis*.

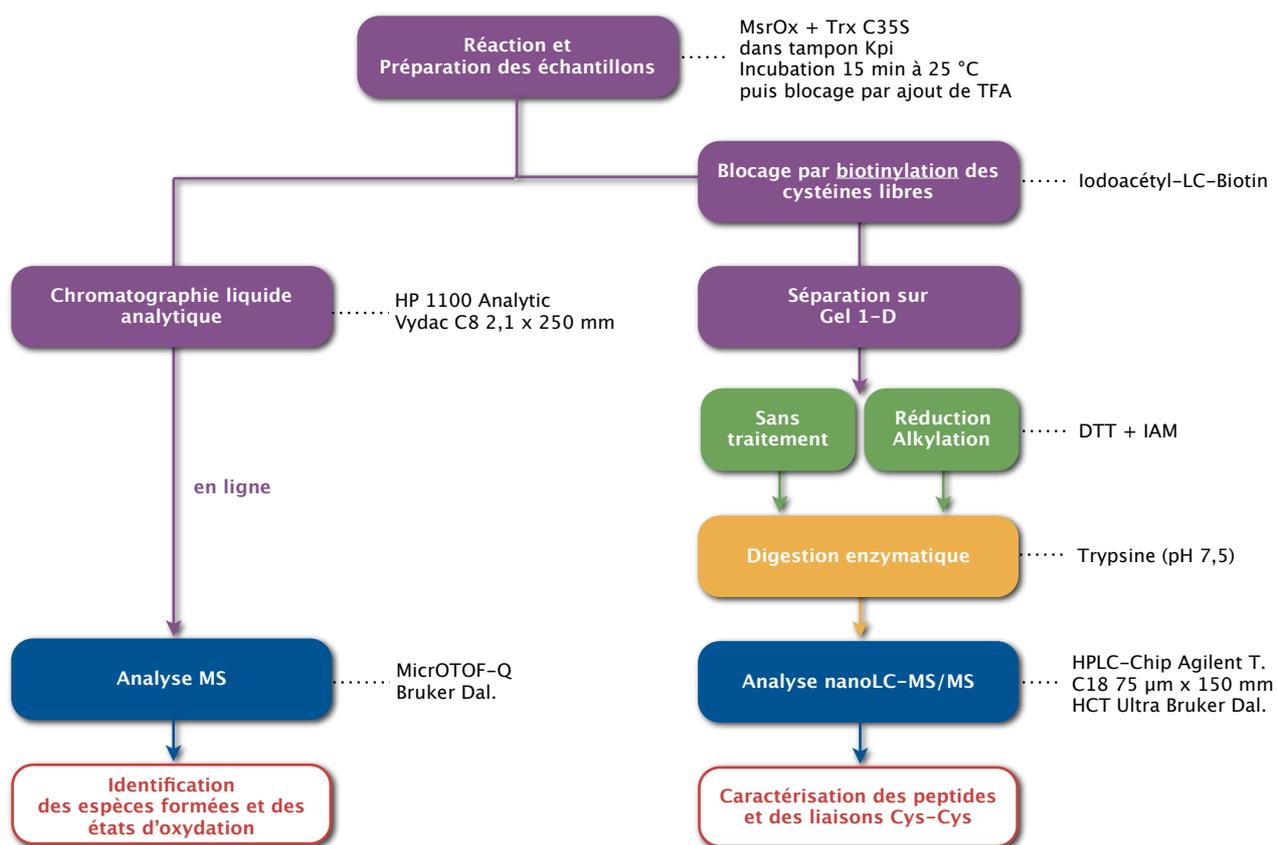
▸ Contrôle de la réaction conduisant à la formation des complexes

Comme le montre la figure 2, les complexes Msr-Trx sont des espèces transitoires dans le mécanisme de réduction des méthionine sulfoxyde réductases par la thiorédoxine. Ces complexes sont en effet attaqués par la cystéine de recyclage C35 de la thiorédoxine. Pour éviter cette attaque et permettre aux complexes à caractériser de s'accumuler dans le milieu réactionnel, la cystéine de recyclage de la thiorédoxine peut être substituée en sérine et la réaction réalisée avec ce mutant Trx C35S, incapable de réduire la liaison cystéine-cystéine formée avec les méthionine sulfoxyde réductases.

L'oxydation des méthionine sulfoxyde réductases a été réalisée en mélangeant 100 µM de MsrA ou B avec 100 mM de racémique MetSO dans le tampon (Tris-HCl, 2 mM EDTA, pH 8) pendant 10 min à température ambiante. Puis, 30 µM de Msr oxydée ont été ajoutés à 100 µM de Trx C35S dans un tampon Kpi (qsp 100 µL) pendant 15 min à 25 °C. La réaction est bloquée par ajout d'acide trifluoroacétique pour éviter la formation d'espèces non spécifiques ou les réarrangements.

### 1.2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre

Dans ce cas où il s'agit de discriminer 2 complexes à liaison disulfure, la stratégie d'analyse LC-MS des protéines entières n'apporte a priori pas d'information directe. Par conséquent, nous nous sommes en premier lieu dirigés vers une stratégie d'analyse nanoLC-MS/MS après digestion et alkylation différentielle (biotine / iodoacétamide). La figure 4 présente les approches et les conditions expérimentales mises au point.



**Figure 4** | Stratégies mises en oeuvre pour la caractérisation des intermédiaires de réaction formés lors de la réduction des méthionine sulfoxyde réductases par les thiorédoxines.

## 1.3. Résultats

### 1.3.1. Les résultats acquis préalablement pour MsrA

Une étude préalable à ce travail<sup>4</sup> avait été menée dans le même objectif : caractériser les complexes disulfures Msr A (et B) – Trx formés lors de la réaction de réduction. La résistance de la thiorédoxine à la digestion par l'enzyme AspN avait été exploitée : l'analyse ESI-MS directe des complexes digérés en milieu liquide avait alors permis l'identification d'une unique liaison disulfure MsrA C154 – Trx C32. Par contre, le doute subsistait toujours quant à la cystéine de MsrB impliquée dans le pont disulfure avec Trx. De plus, un complément de résultat avait été demandé avant publication : la séparation des espèces formées lors de la réaction avait été requise pour certifier que les complexes caractérisés n'étaient pas le reflet de réactions secondaires entre les réactifs.

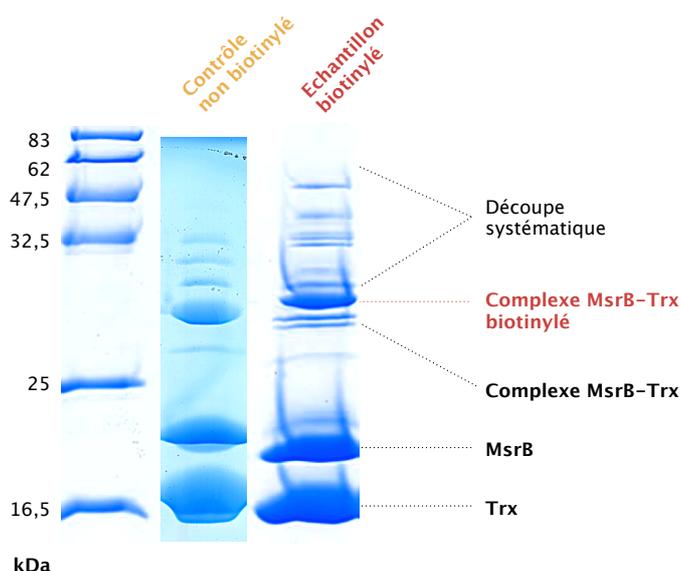
Les efforts se sont donc concentrés sur la caractérisation des complexes MsrB-Trx suivant la stratégie d'analyse nanoLC-MS/MS exposée précédemment.

<sup>4</sup> Chevreux, G. Thèse de doctorat. Université Louis Pasteur. Strasbourg (2005).

### 1.3.2. Caractérisation des complexes MsrB-Trx selon la stratégie d'alkylation différentielle biotine/iodoacétamide par nanoLC-MS/MS

► Gel 1-D

La figure 5 présente le gel 1-D réalisé par nos collaborateurs à partir du mélange réactionnel après blocage des cystéines libres par l'iodoacétyl-LC-Biotin ainsi qu'un contrôle sans biotinylation. On remarque en premier lieu que le rendement de réaction est faible, les bandes correspondant aux complexes MsrB-Trx (biotinylé ou non) étant de faible intensité par rapport à celles des réactifs (MsrB et Trx). Par ailleurs, d'autres bandes de très faible intensité sont visibles dans les hautes masses moléculaires. Pour que leur contenu soit identifié, elles ont été découpées systématiquement, soumises à digestion trypsique et analysées par nanoLC-MS/MS. Les résultats de recherche dans les banques de données ont montré qu'il s'agissait de contaminants (essentiellement des kératines, cf. table 2).



**Figure 5** | Gel 1-D 15 % coloré au bleu de Coomassie réalisé à partir du mélange réactionnel de MsrB oxydée avec Trx C35S traité ou non à l'iodoacétyl-LC-Biotin.

Numéro d'accèsion	Protéine	Masse moléculaire (Da)	Nombre maximum de peptides uniques
P04264	Keratin, type II cytosqueletal 1 - Homo Sapiens	65847	17
P02538	Keratin, type II cytosqueletal 6A - Homo Sapiens	58877	17
P08779	Keratin, type I cytosqueletal 16 - Homo Sapiens	51105	16
P13645	Keratin, type I cytosqueletal 10 - Homo Sapiens	59483	14
P02533	Keratin, type I cytosqueletal 14 - Homo Sapiens	51458	12
P35527	Keratin, type I cytosqueletal 9 - Homo Sapiens	62092	10
Q04695	Keratin, type I cytosqueletal 17 - Homo Sapiens	47945	5
P81605	Dermcidin precursor - Homo Sapiens	11277	4
P0AEX9	Maltose-binding periplasmic protein precursor - E. Coli	43360	4

**Table 2** | Principaux contaminants identifiés sur le gel 1-D présenté en figure 5 par analyse nanoLC-MS/MS après découpe systématique et digestion trypsique.

▸ **Analyses nanoLC-MS/MS du complexe MsrB-Trx C35S biotinylé**

La bande de gel 1-D correspondant au complexe MsrB-Trx biotinylé a été découpée et séparée en 2. La première fraction n'a reçu aucun traitement autre qu'une digestion par la trypsine, cette enzyme ayant été choisie pour la petite taille de dipeptides qu'elle était susceptible de générer. La deuxième fraction a été soumise à un traitement de réduction et d'alkylation à l'iodoacétamide précédant la digestion.

Globalement, les analyses nanoLC-MS/MS de chaque fraction ont permis de recouvrir 67 % des séquences protéiques dans le cas de l'échantillon non traité et 54 % dans le cas de l'échantillon réduit-alkylé (Figure 6, indiqué en couleur). Ce recouvrement est satisfaisant au regard du long peptide MsrB 1-60, qui sans site de coupure intermédiaire, est difficile à analyser. Des essais de digestion par la chymotrypsine, moins spécifique que la trypsine, et des analyses MALDI-MS n'ont pas apporté d'informations supplémentaires.

(a) Fraction non traitée – Recouvrement : 67 %	
<b>Thiorédoxine C35S (Trx) – E. Coli</b>	
51	SDKII HLTDD SFDTD VLKAD GAILV DFVAE WCGPS KMIAP ILDEI ADEYQ
101	GKLTV AKLNI DQNPQ TAPKY GIRGI PTLLE FKNGE VAATK VGALS KGQLK EFLDA NLA
<b>Méthionine Sulfoxyde Réductase B (MsrB) – N. meningitidis</b>	
51	TYKK PSDAE LKRTL TEEQY QVTQN SATEY AFSHE YDHLF KPGIY VDVVS
101	GEPLF SSADK YDSGC GWPSF TRPID AKSVT EHDDF SFNMR RTEVR SHAAD SHLGH VFPDG PRDKG GLRYC INGAS LKFIP LEQMD AAGYG ALKGE VK
(b) Fraction réduite et alkylée IAM – Recouvrement : 54 %	
<b>Thiorédoxine C35S (Trx) – E. Coli</b>	
51	SDKII HLTDD SFDTD VLKAD GAILV DFVAE WCGPS KMIAP ILDEI ADEYQ
101	GKLTV AKLNI DQNPQ TAPKY GIRGI PTLLE FKNGE VAATK VGALS KGQLK EFLDA NLA
<b>Méthionine Sulfoxyde Réductase B (MsrB) – N. meningitidis</b>	
51	TYKK PSDAE LKRTL TEEQY QVTQN SATEY AFSHE YDHLF KPGIY VDVVS
101	GEPLF SSADK YDSGC GWPSF TRPID AKSVT EHDDF SFNMR RTEVR SHAAD SHLGH VFPDG PRDKG GLRYC INGAS LKFIP LEQMD AAGYG ALKGE VK

**Figure 6** | Peptides tryptiques de la thiorédoxine (en vert) et de la méthionine sulfoxyde réductase B (en violet) identifiés par nanoLC-MS/MS pour la fraction non traitée (a) et réduite-alkylée (b).

Par ailleurs, en réalisant des analyses en paramétrant une liste d'inclusion des peptides cibles pour forcer leur sélection, les espèces présentées en table 3 ont pu être identifiées.

Peptide ou dipeptide	Masse théorique (Da)	Echantillon non traité	Echantillon réduit-alkylé IAM
<b>MsrB 61-77 – Trx 19-36</b> YDSGCGWPSFTRPIDAK ADGAILVDFWAEWCGPSK	3860,8	<b>Pont MsrB C65 – Trx C32</b> (962,2 m/z, 4+)	n.i.
<b>MsrB 119-127 – Trx 19-36</b> YCINGASLK ADGAILVDFWAEWCGPSK	2929,4	<b>Pont MsrB C120 – Trx C32</b> (978,2 m/z, 3+)	n.i.
<b>Trx 19-36</b> ADGAILVDFWAEWCGPSK	1963,9	n.i.	n.i.
<b>MsrB 61-77</b> YDSGCGWPSFTRPIDAK	1898,9	<b>C65-IBiotine</b> (1141,0 m/z, 2+)	<b>C65-IBiotine</b> (1141,0 m/z, 2+) <b>C65-IAM</b> (978,4 m/z, 2+)
<b>MsrB 119-127</b> YCINGASLK	967,5	<b>C120-IBiotine</b> (675,8 m/z, 2+)	<b>C120-IBiotine</b> (675,8 m/z, 2+) <b>C120-IAM</b> (513,2 m/z, 2+)

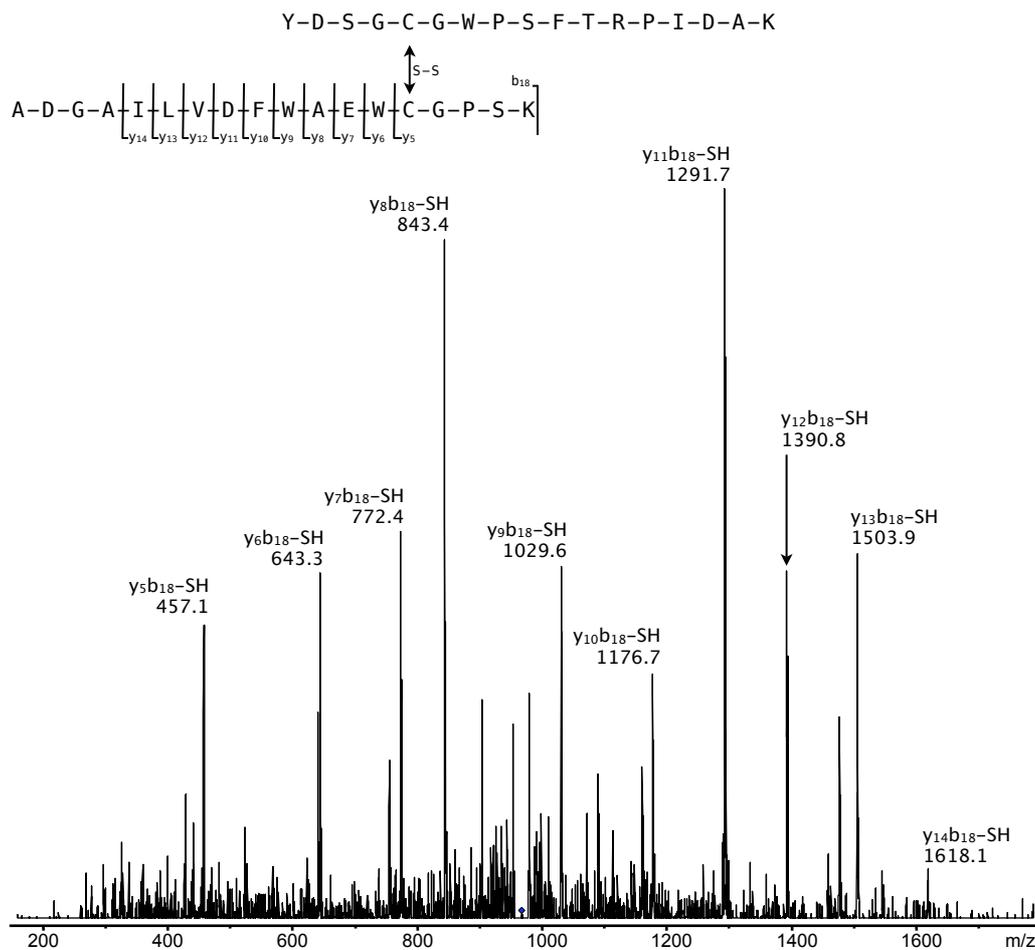
**Table 3** | Peptides caractérisés lors des analyses nanoLC-MS/MS des échantillons non traité et réduit-alkylé. (n.i. : non identifié, masse nette de l'adduit I-Biotine : +382,5 Da, masse nette de l'adduit IAM : +57,0 Da)

Les figures 7 à 10 présentent les spectres de fragmentation ayant permis la caractérisation des dipeptides et des peptides biotinylés dans le cas de l'échantillon non traité.

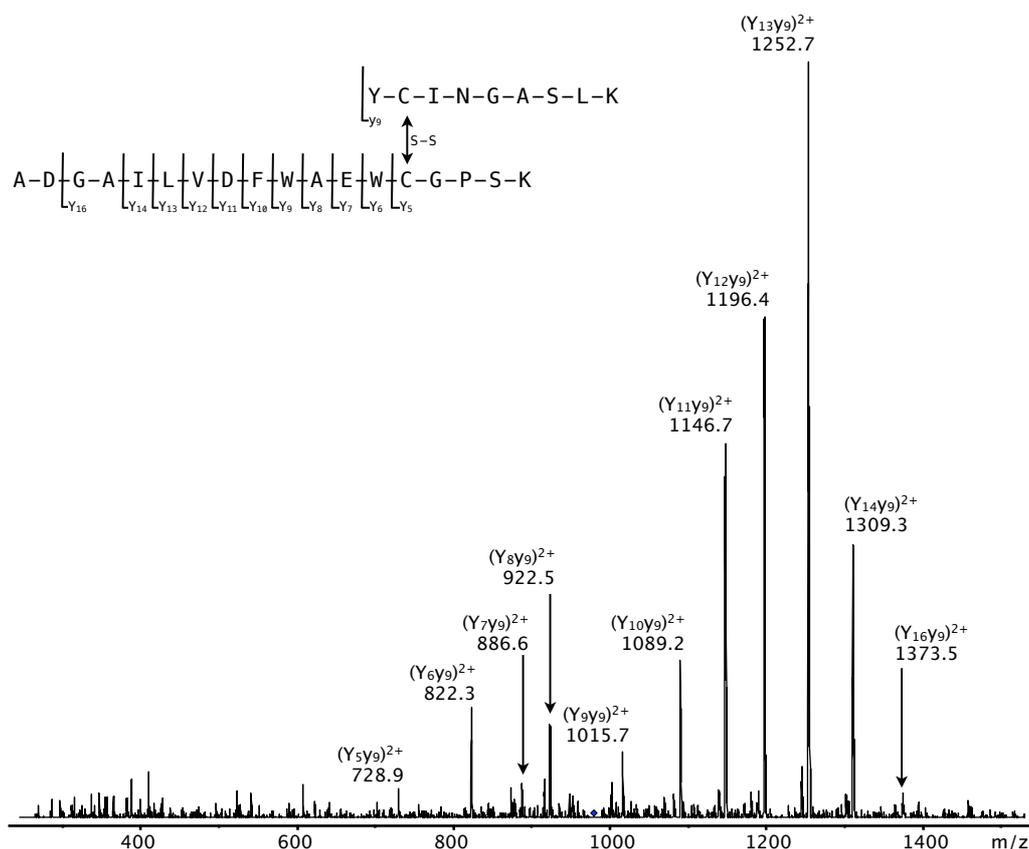
► Conclusions

Les différents indices rassemblés dans cette étude suggèrent que les 2 complexes disulfures MsrB C65 – Trx C32 et MsrB C120 – Trx C32 sont susceptibles de se former lors de la réduction de MsrB oxydée par Trx. En effet, l'analyse de l'échantillon non traité montre que les deux complexes co-existent alors que toutes les cystéines en mesure de réagir ont été biotinyllées et sont donc inactivées. Par ailleurs, en soumettant l'échantillon à un traitement de réduction et d'alkylation par l'iodoacétamide, les liaisons cystéine-cystéine sont bien rompues et on retrouve un mélange des cystéines de Msr biotinyllées et carbamidométhylées. Le peptide Trx 19–36 n'est pas retrouvé dans l'échantillon traité, mais du fait de la présence de multiples acides aminés acides (2 acides glutamiques D, 1 acide aspartique E) il est probable qu'il ait des difficultés à se charger positivement. Ce peptide ne peut donc pas être analysé.

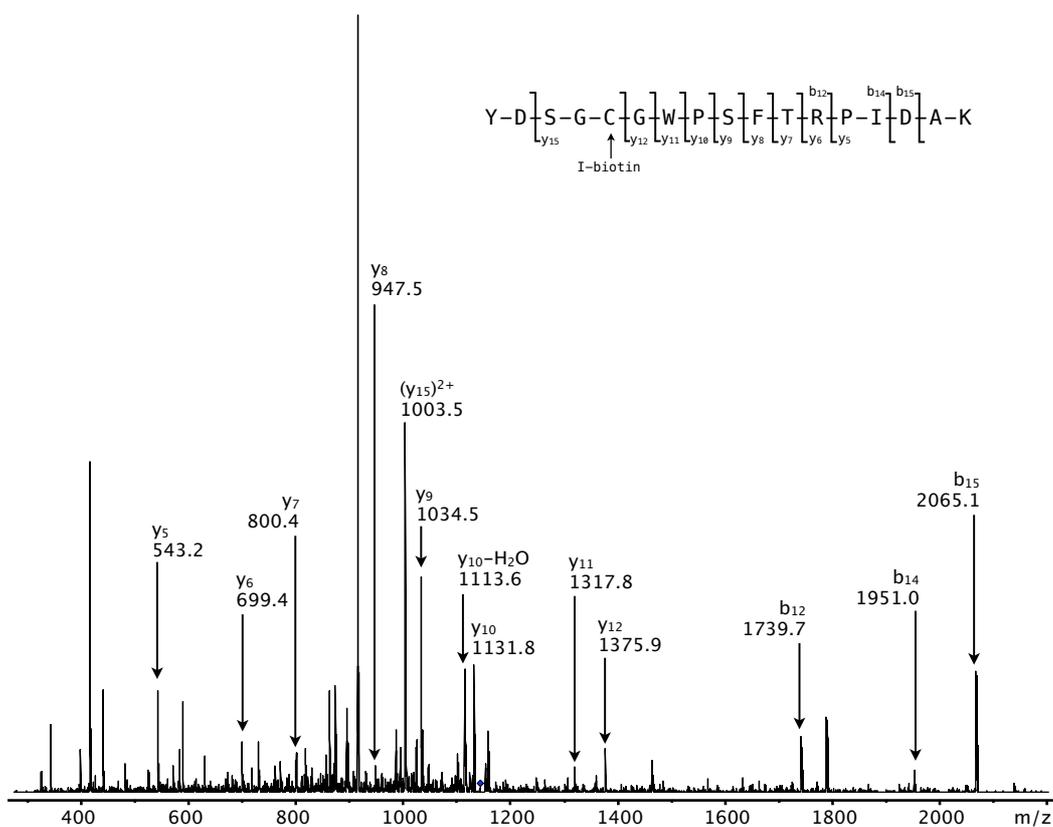
Toutefois, il est peu vraisemblable que les deux cystéines de la méthionine sulfoxyde réductase B puissent être attaquées indifféremment par la cystéine de recyclage C32 de la thiorédoxine. A l'échelle de temps de l'analyse, il n'est pas possible d'observer un équilibre cinétique court (l'attaque préférentielle d'une des deux cystéines du pont intra-protéique de Msr oxydée). Il se peut que notre stratégie permette uniquement d'observer le produit thermodynamique de la réaction (issu de réarrangements entre les liaisons cystéine-cystéine).



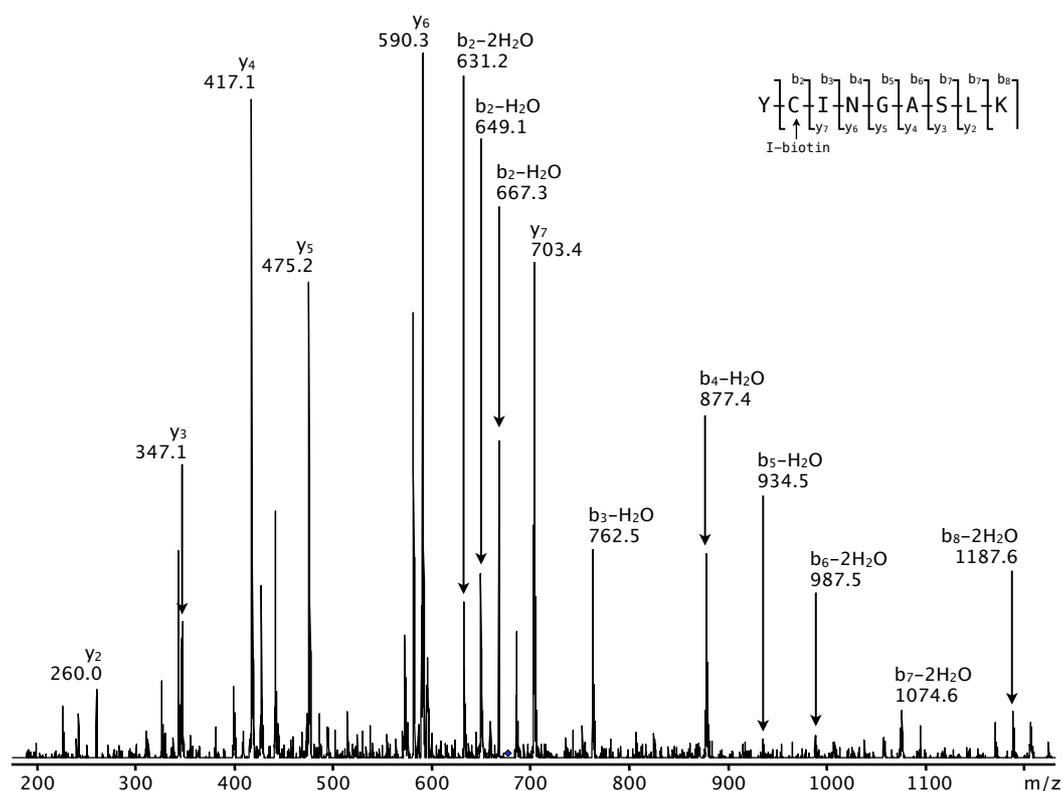
**Figure 7** | Spectre de fragmentation de l'ion  $[M+4H]^{4+}$  à 966,2 m/z, correspondant au dipeptide MsrB 61-77 – Trx 19-36 à pont disulfure (MsrB C65 – Trx C32), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $y_n b_{18} - SH$  ( $n=5-14$ ).



**Figure 8** | Spectre de fragmentation de l'ion  $[M+3H]^{3+}$  à 978,2 m/z, correspondant au dipeptide MsrB 119-127 - Trx 19-36 à pont disulfure (MsrB C120 - Trx C32), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $(Y_nY_9)^{2+}$  ( $n=5-16$ ).



**Figure 9** | Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à 1141,03 m/z, correspondant au peptide MsrB 61-77 avec C65 biotinylée, obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $y_n$  ( $n=5-12$ ).

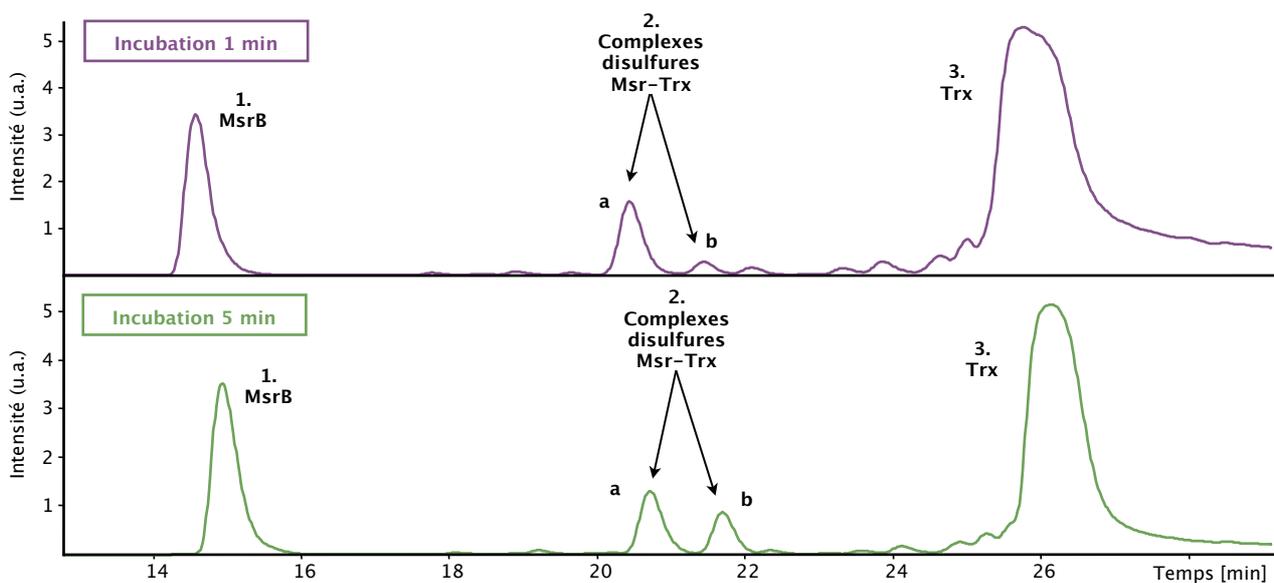


**Figure 10** | Spectre de fragmentation de l'ion  $[M+2H]^{2+}$  à 675,8 m/z, correspondant au peptide MsrB 119-127 avec C120 biotinyllée, obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $y_n$  ( $n=2-7$ ).

### 1.3.3. Analyses LC-MS des complexes entiers

De façon inattendue, l'approche analytique mettant en oeuvre une séparation du mélange réactionnel en ligne avec une mesure de masse ESI-MS des protéines entières apporte des informations supplémentaires sur la possible évolution du complexe préférentiellement formé en un mélange des 2 formes disulfures.

La figure 11 présente les chromatogrammes obtenus lors de l'analyse des complexes MsrB-Trx formés avec des temps d'incubation de 1 et 5 min. Les masses mesurées des espèces contenues dans chaque fraction chromatographique sont rassemblées en table 4.



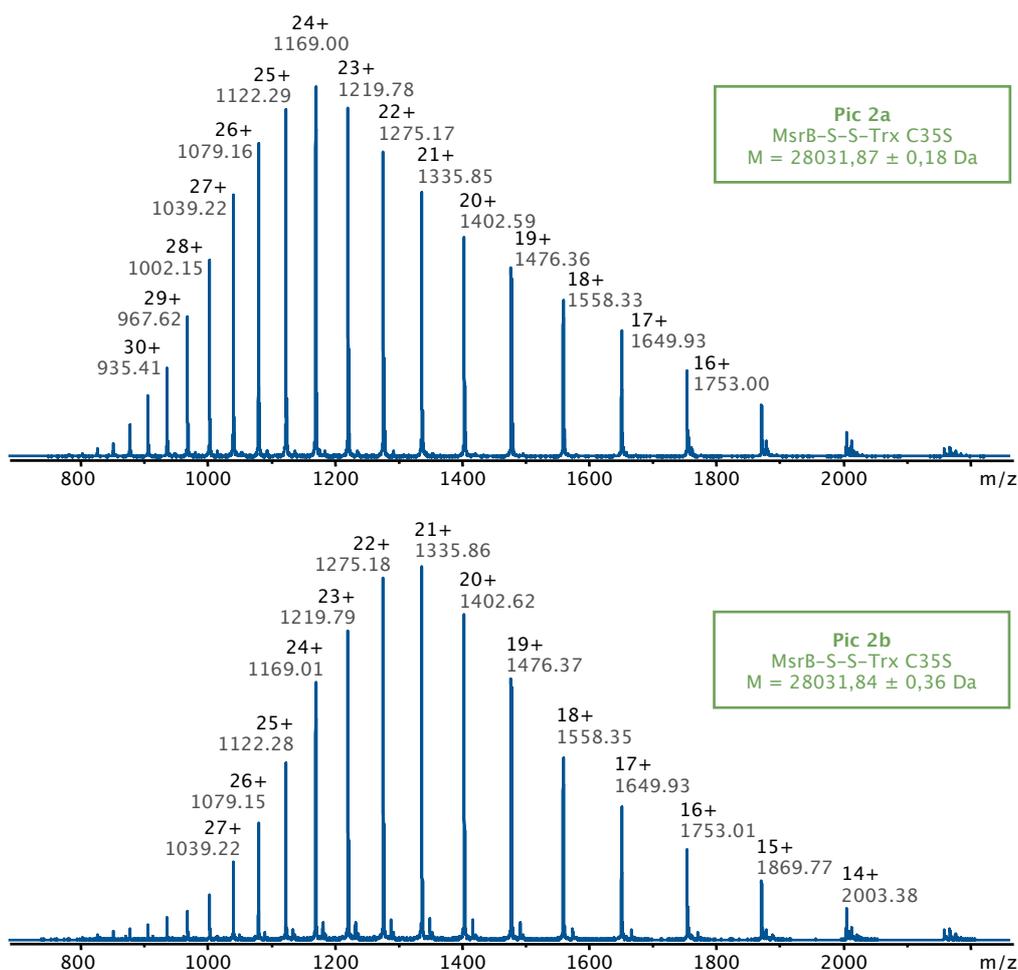
**Figure 11** | Séparations par chromatographie liquide en phase inverse C8 du mélange réactionnel MsrB - Trx C35S incubé : (a) 1 min à 25 °C et (b) 5 min à 25 °C.

Pic #	Espèces	Formes	Masses moyennes observées (Da)		Masses moyennes attendues (Da)
			Incubation 1min	Incubation 5 min	
1	MsrB	Oxydée	16 372,05 ± 0,23	16 372,08 ± 0,20	16 372,00
2	Complexe MsrB-Trx C35S	Pont disulfure	2a : 28 031,81 ± 0,26 2b : 28 031,79 ± 0,28	2a : 28 031,87 ± 0,18 2b : 28 031,84 ± 0,36	28 031,25
3	Trx C35S	-	11 658,70 ± 0,45	11 659,07 ± 0,23	11 659,25

**Table 4** | Masses mesurées lors de l'analyse LC-MS du mélangé réactionnel MsrB - Trx C35S incubé 1 ou 5 min.

Les masses des produits de réaction sont identiques quelque soit le temps d'incubation. Les complexes sont élués entre 20 et 22 min et il est possible de distinguer deux pics complètement séparés qui contiennent tous deux des espèces de même masse moyenne. Cette masse correspond à celle d'un complexe MsrB-Trx C35S à pont disulfure.

Les chromatogrammes montrent clairement que l'intensité du deuxième pic de complexe (2b) augmente pour un temps d'incubation plus long. Par ailleurs, on peut s'intéresser à la distribution des états de charge sur les spectres de chaque espèce (Figure 12). Il apparaît que la distribution est centrée sur le pic représentant le complexe 24 fois chargé pour la fraction chromatographique 2a, alors que pour la fraction 2b la distribution est centrée sur l'état de charge 21+.



**Figure 12** | Spectres de masse ESI-MS, acquis sur MicroTOF-Q Bruker, de complexes disulfures MsrB-Trx C35S, issus des fractions chromatographiques 2a et 2b. Echantillon incubé 5 min.

Le fait que ces deux espèces, correspondant à la masse du complexe disulfure MsrB-Trx, ne soient pas coéluees et présentent des distributions en états de charge décalées indique qu'il s'agit probablement des deux complexes à pont disulfure MsrB C65 – Trx C32 et MsrB C120 – Trx C32. En effet, si l'on suppose que les 2 complexes présentent une conformation différente à cause de leur liaison cystéine-cystéine respective, une modification des temps de rétention chromatographique peut être expliquée. De la même manière, une conformation différente peut modifier l'efficacité d'ionisation des deux molécules en exposant plus ou moins d'acides aminés basiques protonables.

Ces arguments étayaient par conséquent l'hypothèse qu'un des deux complexes MsrB-Trx se forme préférentiellement sous contrôle cinétique et que pour des temps d'incubation plus longs l'autre complexe se forme pour atteindre l'équilibre thermodynamique.

Malheureusement, il n'a pas été possible de déterminer quelle était l'espèce qui se formait en premier malgré nos tentatives de collecte sélective de chacun des deux pics suivie d'une analyse nanoLC-MS/MS après digestion enzymatique. De nouvelles stratégies doivent être développées pour discriminer les deux espèces et tenter de définir laquelle des deux cystéines de la méthionine sulfoxyde réductase B est attaquée préférentiellement par la thiorédoxine dans le mécanisme de réduction.

## 1.4. Conclusion

Alors qu'il a précédemment été possible de montrer que la cystéine 154 de la méthionine sulfoxyde réductase A était attaquée par la cystéine de recyclage C32 de la thiorédoxine, les études concernant la méthionine sulfoxyde réductase B présentées ici n'ont pas pu aboutir à un résultat tranché. En effet, il est apparu que la cinétique de la réaction était trop rapide et que l'équilibre thermodynamique conduisant à la formation des deux complexes possible était atteint rapidement.

Cette différence entre MsrA et MsrB peut s'expliquer par l'éloignement plus grand des 2 cystéines chez MsrA que MsrB. Les 2 cystéines de MsrB étant proches l'une de l'autre, les ponts disulfures MsrB-Trx sont potentiellement sujets à des réarrangements plus rapides.

De nouvelles expériences, réalisées en utilisant des mutants inactivant sélectivement l'une ou l'autre des cystéines de MsrB, peuvent apporter des informations supplémentaires sur le complexe formé préférentiellement. On pourra également réfléchir à des approches permettant un suivi cinétique des espèces formées. Par exemple, de récentes études proposent le marquage isotopique différentiel des cystéines<sup>5,6</sup> : un premier traitement alkylant par un groupement acide iodoacétique marqué <sup>12</sup>C est suivi d'une réduction des ponts disulfures et d'une alkylation par un groupement acide iodoacétique marqué <sup>13</sup>C. Après digestion et analyse MS, la fraction de peptides initialement pontés ou non peut alors être directement estimée et calculant le rapport d'intensité des peptides marqués <sup>12</sup>C et <sup>13</sup>C.

<sup>5</sup> Xiang, T., et al. (2009). *Anal Chem* **81**, 8101–8.

<sup>6</sup> Chumsae, C., et al. (2009). *Anal Chem*.

## 2. Etude du mécanisme de régénération des activités méthionine sulfoxyde réductase de la protéine périplasmique PilB

### 2.1. Le rôle de la protéine PilB

#### 2.1.1. Une structure à 3 domaines

PilB est une protéine périplasmique trouvée chez *N. meningitidis* qui comporte 3 domaines distincts (Figure 13) :

- Le domaine N-terminal (N-ter) présente une activité disulfure oxydoréductase
- Le domaine central et le domaine C-terminal présentent une activité méthionine sulfoxyde réductase A (MsrA) et B (MsrB), respectivement.

PilB est donc impliquée dans la défense contre le stress oxydant<sup>7</sup> touchant particulièrement les méthionines et dans les mécanismes de survie contre l'attaque des espèces réactives de l'oxygène comme H<sub>2</sub>O<sub>2</sub>. Ces questions ont déjà été présentées aux points 1.1 du chapitre III et 1.1.1 du chapitre II.

La production du domaine N-ter en domaine isolé et soluble a récemment permis de montrer son activité de recyclage des formes oxydées des domaines Msr de PilB. Toutefois, le mécanisme de réduction n'est réellement efficace qu'envers le domaine MsrB.

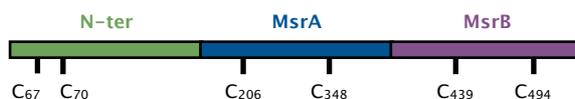


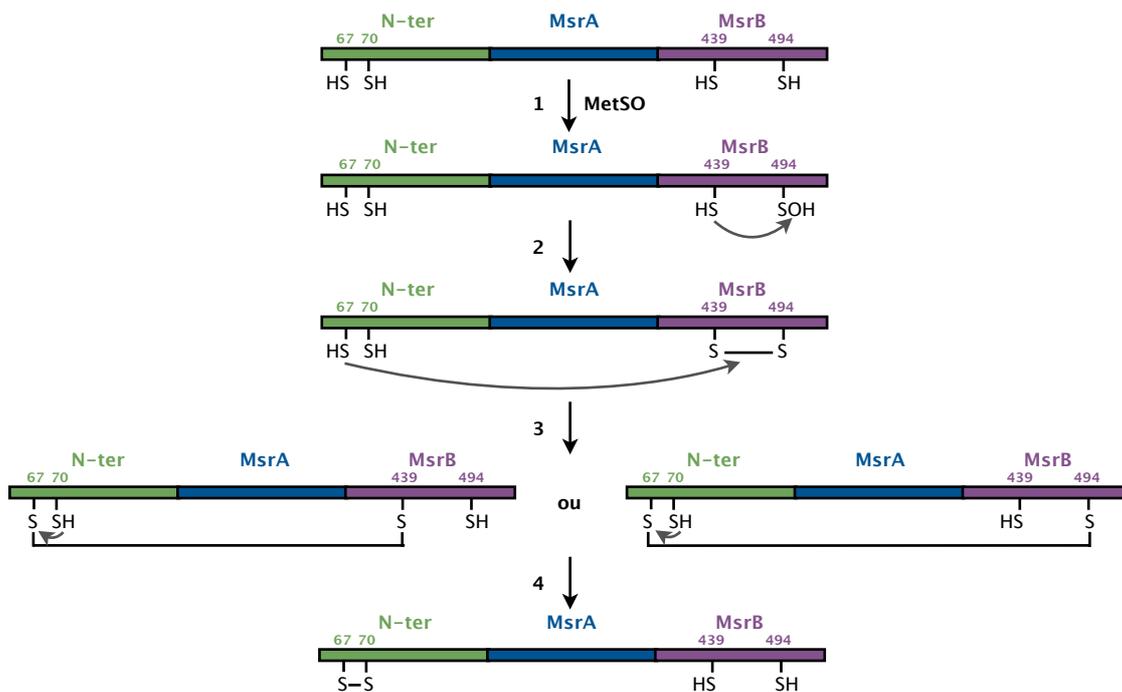
Figure 13 | Représentation schématique de la protéine PilB de *N. meningitidis* et position des cystéines dans sa séquence protéique.

#### 2.1.2. Hypothèses de mécanismes de réduction du domaine MsrB

Chacun des domaines de PilB possède deux cystéines. Des études cinétiques, la préparation de mutants et les analogies avec la réduction des méthionine sulfoxyde réductases par les thiorédoxines ont permis à nos collaborateurs d'identifier les cystéines catalytiques (C67, C206, C494) et de proposer une hypothèse de mécanisme intramoléculaire pour le recyclage du domaine MsrB oxydé par le domaine N-ter (Figure 14).

Après que le domaine MsrB a réduit une méthionine R sulfoxyde (étape 1), la cystéine C439 attaque l'acide sulfénique formé sur la cystéine C494 (étape 2). Le pont disulfure C439-C494 formé est alors réduit par la cystéine catalytique C67 du domaine N-ter selon un mécanisme – c'est une hypothèse – intramoléculaire (étape 3). La cystéine du domaine MsrB qui est attaquée préférentiellement reste à déterminer. Finalement, au cours d'une 4<sup>e</sup> étape, l'attaque de la cystéine C67 par la cystéine de recyclage C70 permet de réduire totalement le domaine MsrB.

<sup>7</sup> Skaar, E. P., et al. (2002). Proc Natl Acad Sci U S A **99**, 10108–13.



**Figure 14** | Etapes 1 et 2 : mécanisme général de réduction des MetSO par le domaine MsrB. Etapes 3 et 4 : hypothèse de mécanisme intramoléculeire de régénération du domaine MsrB par le domaine N-ter.

## 2.2. Stratégie analytique mise en oeuvre pour l'étude du mécanisme de réduction du domaine MsrB de PilB

### 2.2.1. Objectifs de l'étude

Cette étude vise 2 objectifs. Le premier est de valider l'hypothèse selon laquelle la réduction du domaine MsrB oxydé est réalisée par le domaine N-ter suivant un mécanisme intramoléculeire. Le second objectif est de déterminer quelle est la cystéine du domaine MsrB qui est attaquée lors de sa réduction par la cystéine catalytique C67 du domaine N-ter.

Pour cela, il s'agit comme précédemment de :

- mettre au point les conditions idéales favorisant l'accumulation des complexes à caractériser ;
- déterminer une stratégie d'analyse par spectrométrie de masse permettant de discriminer les différents complexes pouvant être formés (ie. complexe intra- ou interprotéique, présentant une liaison disulfure entre la cystéine N-ter C67 et l'une ou l'autre des cystéines du domaine MsrB) ;
- déterminer précisément la position du ou des ponts formés en caractérisant chacun des complexes identifiés.

La suite de ce chapitre s'appuie à nouveau sur le travail de développement méthodologique décrit au chapitre I de cette partie et détaille l'application d'une combinaison de ces stratégies à l'étude de la réduction du domaine MsrB de PilB.

## 2.2.2. Contrôle de la réaction conduisant à la formation des complexes

Dans l'hypothèse de mécanisme décrite à la figure 14, les formes à ponts disulfures intramoléculaires C67–C439 et C67–C494 de PilB sont des espèces transitoires. Ils sont en effet attaqués par la cystéine de recyclage C70 du domaine N-ter. Pour éviter toute attaque conduisant à la rupture des complexes à caractériser et pour leur permettre de s'accumuler dans le milieu réactionnel, **la cystéine de recyclage C70 peut être substituée en sérine**. De plus, en inactivant le domaine MsrB, il a été montré que le domaine MsrA n'était pas réduit par le domaine N-ter. Les cystéines de MsrA peuvent donc être également substituées pour inactiver le domaine. La réaction peut alors être réalisée avec le mutant PilB C70S/C206S/C348S, théoriquement incapable de réduire les liaisons cystéine–cystéine formées entre C67 et C439 et/ou C494.

La figure 15 présente la séquence de PilB utilisée pour notre étude. L'indexation de la séquence utilisée dans le texte correspond à la position réelle des acides aminés dans la séquence de PilB sauvage chez *N. meningitidis*.

- Le domaine N-ter possède une cystéine située en position 67.
- Le domaine MsrA ne possède plus aucune cystéine active.
- Le domaine MsrB possède deux cystéines en position 439 et 494.

PilB	Domaine N-terminal (N-ter)										
	Domaine Méthionine Sulfoxyde Réductase A (MsrA)						Domaine Méthionine Sulfoxyde Réductase B (MsrB)				
32	MVPHT	LSTLK	TADNR	PASVY	LKKDK	PTLIK	FWASW	CPLSL	SELGQ	TEKWA	
82	QDAKF	SSANL	ITVAS	PGFLH	EKKDG	DFQKW	YAGLN	YPKLP	VVTDN	GGTIA	
132	QSLNI	SVYPS	WALIG	KDGDV	QRIVK	GSINE	AQALA	LIRDP	NADLG	SLKHS	
182	FYKPD	TQKKD	SKIAN	TRTIY	LAGGS	FWGLE	AYFQR	IDGVV	DAVSG	YANGN	
232	TKNPS	YEDVS	YRHTG	HAETV	KVTYD	ADKLS	LDDIL	QYFFR	VVDPT	SLNKQ	
282	GNDTG	TQYRS	GVYYT	DPAEK	AVIAA	ALKRE	QQKYQ	LPLVV	ENEPL	KNFYD	
332	AEEYH	QDYLI	KNPNG	YSHID	IRKAD	EPLPG	KTKTA	PQGGK	FDAAT	YKKPS	
382	DAELK	RTLTE	EQYQV	TQNSA	TEYAF	SHEYD	HLFKP	GIYVD	VVSGE	PLFSS	
432	ADKYD	SGCGW	PSFTR	PIDAK	SVTEH	DDFSY	NMRRT	EVRSH	AADSH	LGHVF	
482	PDGPR	DKGGL	RYCIN	GASLK	FIPLE	QMDAA	GYGAL	KSKVK			

Figure 15 | Séquence de PilB *N. meningitidis* produite et indexation utilisée dans le texte.

La réaction d'oxydation et de réduction du domaine MsrB de PilB a été réalisée en mélangeant 15  $\mu\text{M}$  de PilB avec 15 mM de racémique MetSO dans le tampon potassium phosphate (Kpi, qsp 100  $\mu\text{L}$ ) pendant 15 min à 25°C. La réaction est bloquée par ajout d'acide trifluoroacétique pour éviter la formation d'espèces non spécifiques ou les réarrangements.

## 2.2.3. Les stratégies de caractérisation par spectrométrie de masse mises en oeuvre

Dans l'objectif de déterminer si la réduction du domaine MsrB oxydé par le domaine N-ter de PilB a lieu selon un mécanisme intra- ou inter-moléculaire, nous avons proposé de mesurer la masse entière du mutant C70S/C206S/C348S en présence ou non de méthionine sulfoxyde après que les cystéines libres ont été bloquées par le N-éthylmaléimide (NEM). Une approche de purification par LC analytique et d'analyse en ligne ESI-MS a d'abord été mise en oeuvre. Puis, pour identifier les cystéines impliquées dans l'intermédiaire réactionnel domaine N-ter – domaine MsrB, une stratégie d'analyse nanoLC-MS/MS précédée ou non d'une étape de digestion et d'alkylation par l'iodoacétamide a été développée. La figure 16 présente les approches et conditions expérimentales mises au point.

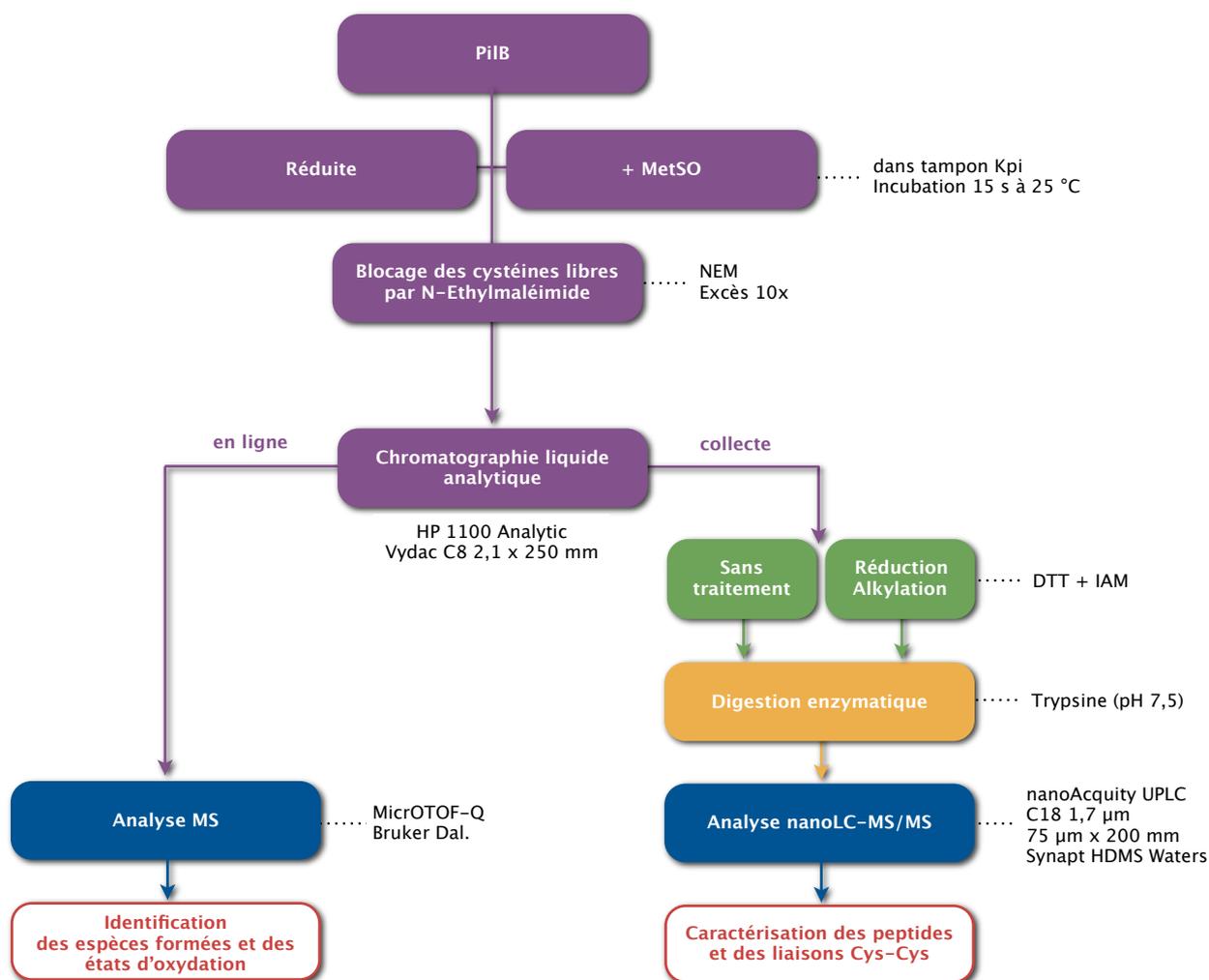


Figure 16 | Stratégies mises en oeuvre pour la caractérisation des intermédiaires de réaction formés lors de la réduction du domaine MsrB oxydé de PiIB.

## 2.3. Résultats

### 2.3.1. Analyses LC-MS de la protéine PiIB entière

Le chromatogramme obtenu lors de l'analyse des échantillons de PiIB (non présenté) présente un unique pic à un temps de rétention de 21 min (en travaillant avec le gradient mis au point au chapitre I) et contenant les complexes formés lors de la réaction. Les masses mesurées lors des analyses LC-MS des échantillons de la protéine PiIB réduite ou oxydée et soumise à un traitement alkylant par le N-Ethylmaléimide, sont présentées en table 5.

Il apparaît d'abord que pour l'échantillon de contrôle « PiIB réduite », l'alkylation par le NEM est efficace, mais peut ne pas être totale : une forme minoritaire de PiIB ne présente que 2 cystéines alkylées sur les 3 libres que la protéine contient (Figure 17).

Dans le cas de l'échantillon mis en présence de méthionine sulfoxyde, une unique masse est relevée : elle peut correspondre à une forme de PiIB possédant un pont disulfure et une cystéine modifiée par le NEM (Figure 18). La caractérisation complète de la ou des formes oxydées de PiIB doit alors être réalisée.

Echantillon	Espèces	Formes	Masses moyennes observées (Da)	Masses moyennes attendues (Da)
PiIB réduite + NEM	PiIB	+ 3 NEM + 2 NEM (minoritaire)	55 002,59 ± 0,62 54 877,43 ± 0,55	55 002,54 54 877,41
PiIB + MetSO + NEM	PiIB	Pont disulfure + 1 NEM	54 750,54 ± 0,57	54 750,26

Table 5 | Masses moyennes et complexes correspondants mesurés par LC-MS à partir des échantillons réduit et oxydé de PiIB.

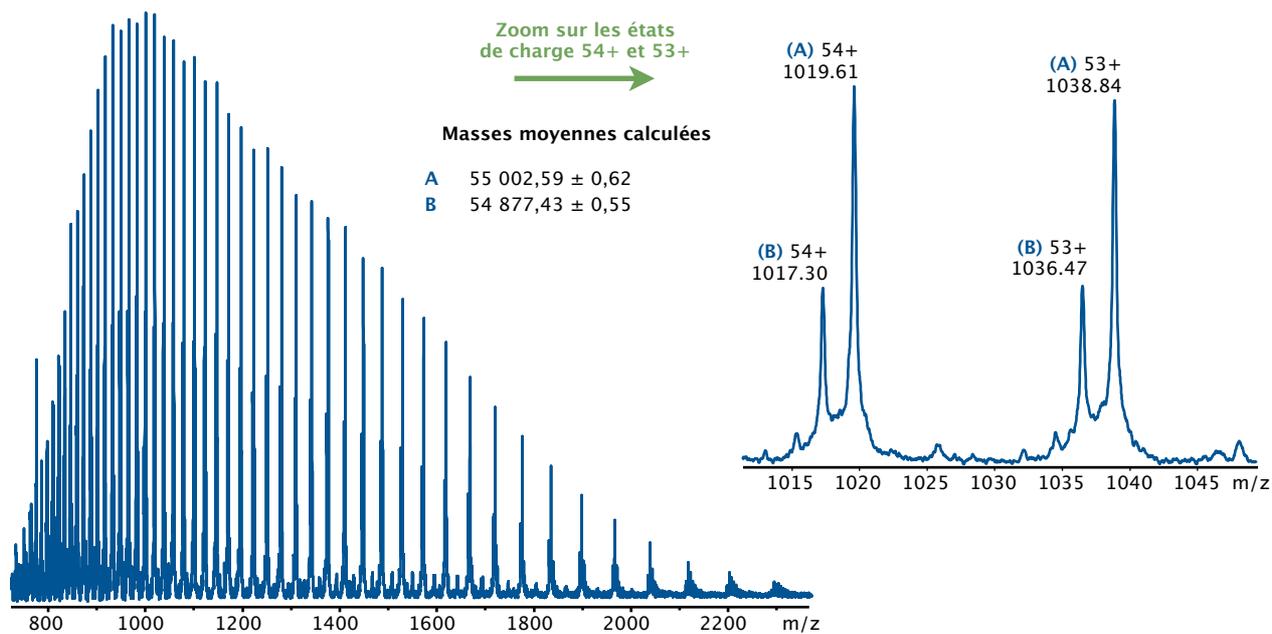


Figure 17 | Spectre de masse ESI-MS acquis sur MicroTOF-Q Bruker, de PiIB réduite avec traitement NEM. Echantillon incubé 15 s.

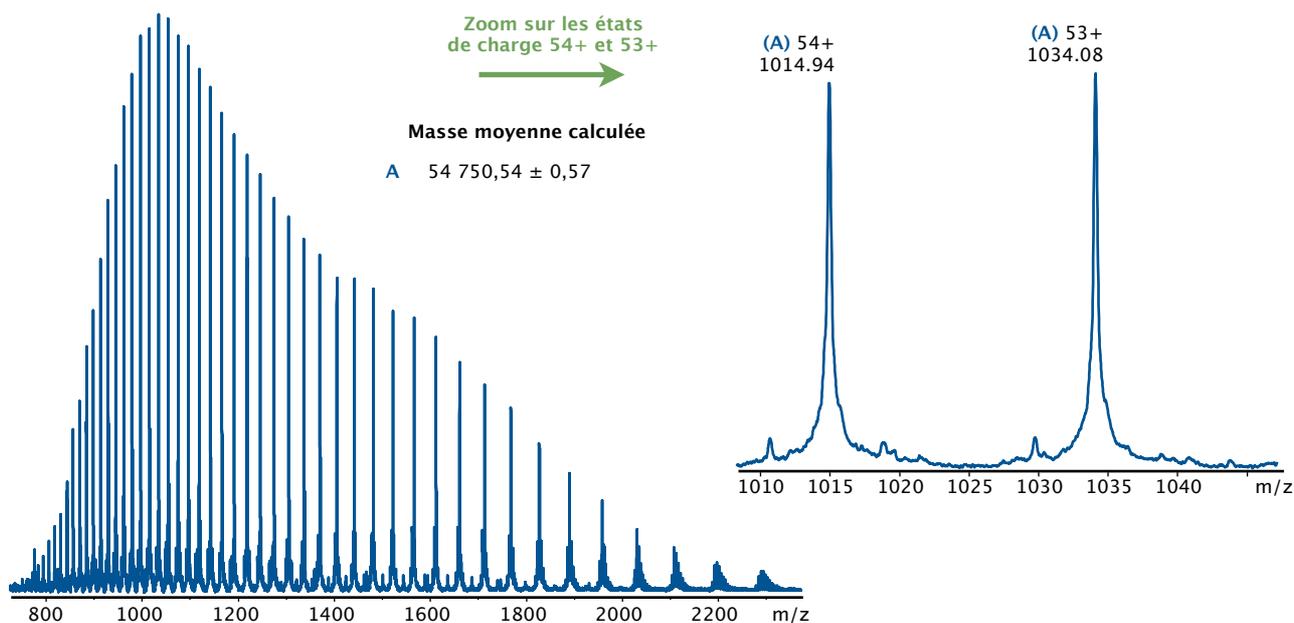


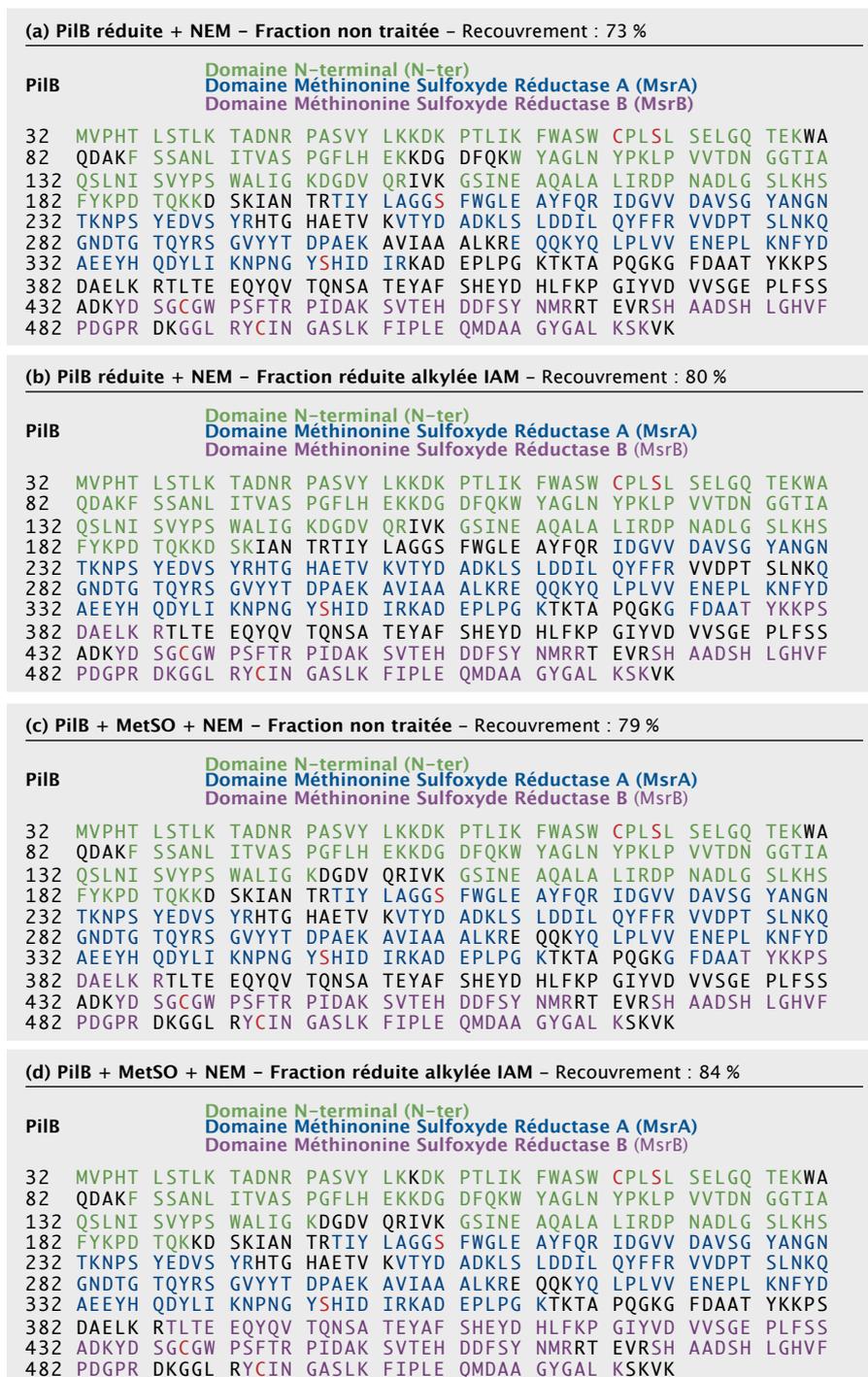
Figure 18 | Spectre de masse ESI-MS acquis sur MicroTOF-Q Bruker, de PiIB oxydée avec traitement NEM. Echantillon incubé 15 s.

### 2.3.2. Caractérisation des complexes formés selon la stratégie nanoLC-MS/MS

▸ Recouvrements de séquences

Pour aller plus loin et caractériser la position exacte de la ou des formes disulfures identifiées précédemment, les fractions chromatographiques des échantillons de PiIB ont été collectées et soumises à digestion trypsique après un éventuel traitement réductif et alkylant par l'iodoacétamide. La trypsine a été choisie, car elle permet la formation de peptides de taille adaptée à une analyse nanoLC-MS/MS optimale.

La figure 19 montre l'information de séquence recouverte lors de l'analyse nanoLC-MS/MS des 4 types d'échantillon de PiIB.



**Figure 19** | Peptides tryptiques de la protéine PiIB identifiés (en couleur) par nanoLC-MS/MS pour les échantillons réduits non traité (a) et réduit alkylé (b) et les échantillons oxydés non traité (c) et réduit alkylé (d).

Globalement, les analyses nanoLC-MS/MS ont permis de recouvrir entre 73 et 84 % de la séquence de PiIB. D'autres essais de digestion avec des enzymes moins spécifiques telles que la chymotrypsine et l'analyse par MALDI-MS des échantillons n'ont pas apporté d'informations supplémentaires. L'information de séquence recouverte est toutefois pratiquement complète : seuls manquent quelques peptides plus difficiles à analyser du fait de leur taille.

Les trois peptides contenant une cystéine sont systématiquement analysés. Dans le cas particulier des échantillons non soumis à un traitement de réduction préalable, cela semble indiquer que chacune des cystéines est susceptible de ne pas être impliquée dans un pont disulfure.

▸ Caractérisation des peptides d'intérêt

Les peptides cibles pour la caractérisation de l'état des cystéines et des liaisons disulfures qui peuvent se former ont été inclus dans les options d'analyse pour forcer leur sélection et fragmentation en nanoLC-MS/MS. Les espèces identifiées sont rassemblées en table 6. La caractérisation des dipeptides a nécessité une recherche systématique et une interprétation manuelle des spectres sans assistance automatisée.

(a) Echantillon PiIB réduite

Peptide ou dipeptide	Masse théorique (Da)	Echantillon non traité	Echantillon réduit-alkylé IAM
<b>PiIB 62-79 - PiIB 435-451</b> FWASWCPLSLSSELGQTEK YDSGCGWPSFTRPIDAK	3977,8	n.i.	n.i.
<b>PiIB 62-79 - PiIB 493-501</b> FWASWCPLSLSSELGQTEK YICINGASLK	3046,5	n.i.	n.i.
<b>PiIB 435-451 - PiIB 493-501</b> YDSGCGWPSFTRPIDAK YICINGASLK	2864,34	n.i.	n.i.
<b>PiIB 62-79</b> FWASWCPLSLSSELGQTEK	2081,0	C67-NEM (1104,0 m/z, 2+)	C67-NEM (1104,0 m/z, 2+) C67-IAM (1070,0 m/z, 2+)
<b>PiIB 435-451</b> YDSGCGWPSFTRPIDAK	1898,9	C439-NEM (675,6 m/z, 3+)	C439-NEM (675,6 m/z, 3+) C439-IAM (653,0 m/z, 3+)
<b>PiIB 493-501</b> YICINGASLK	967,5	C494-NEM (547,3 m/z, 2+) C494-réduite (484,7 m/z, 2+)	C494-NEM (547,3 m/z, 2+) C494-IAM (513,2 m/z, 2+)

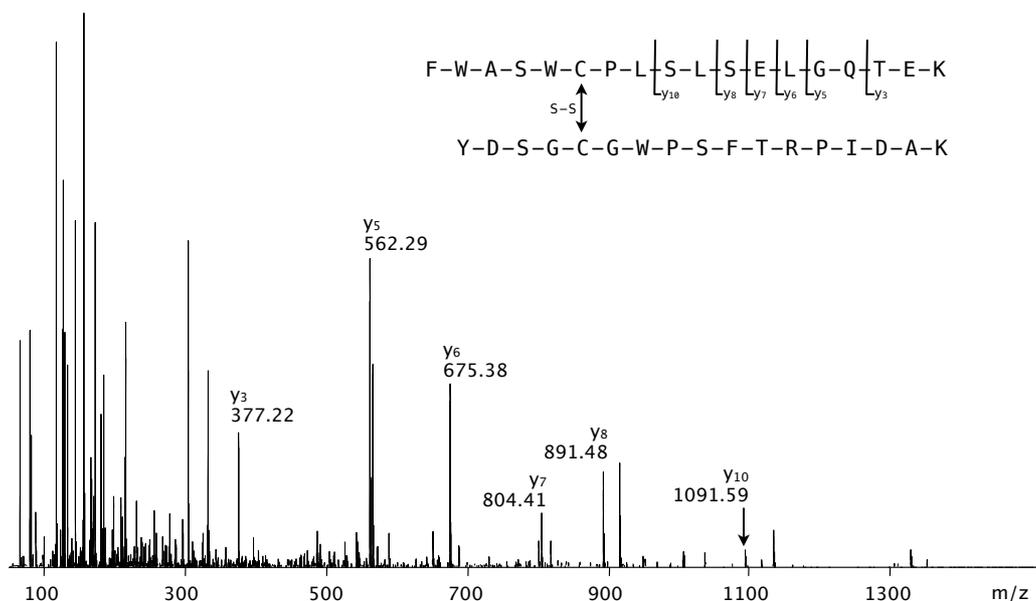
(b) Echantillon PiIB + MetSO

Peptide ou dipeptide	Masse théorique (Da)	Echantillon non traité	Echantillon réduit-alkylé IAM
<b>PiIB 62-79 - PiIB 435-451</b> FWASWCPLSLSSELGQTEK YDSGCGWPSFTRPIDAK	3977,8	Pont C67 - C439 (797,0 m/z, 5+)	n.i.
<b>PiIB 62-79 - PiIB 493-501</b> FWASWCPLSLSSELGQTEK YICINGASLK	3046,5	Pont C67 - C494 (762,9 m/z, 4+)	n.i.
<b>PiIB 435-451 - PiIB 493-501</b> YDSGCGWPSFTRPIDAK YICINGASLK	2864,34	Pont C439 - C494 (717,3 m/z, 4+)	n.i.
<b>PiIB 62-79</b> FWASWCPLSLSSELGQTEK	2081,0	C67-NEM (1104,0 m/z, 2+)	C67-NEM (1104,0 m/z, 2+) C67-IAM (1070,0 m/z, 2+)
<b>PiIB 435-451</b> YDSGCGWPSFTRPIDAK	1898,9	C439-NEM (675,6 m/z, 3+)	C439-NEM (675,6 m/z, 3+) C439-IAM (653,0 m/z, 3+)
<b>PiIB 493-501</b> YICINGASLK	967,5	C494-NEM (547,3 m/z, 2+) C494-réduite (484,7 m/z, 2+)	C494-NEM (547,3 m/z, 2+) C494-IAM (513,2 m/z, 2+)

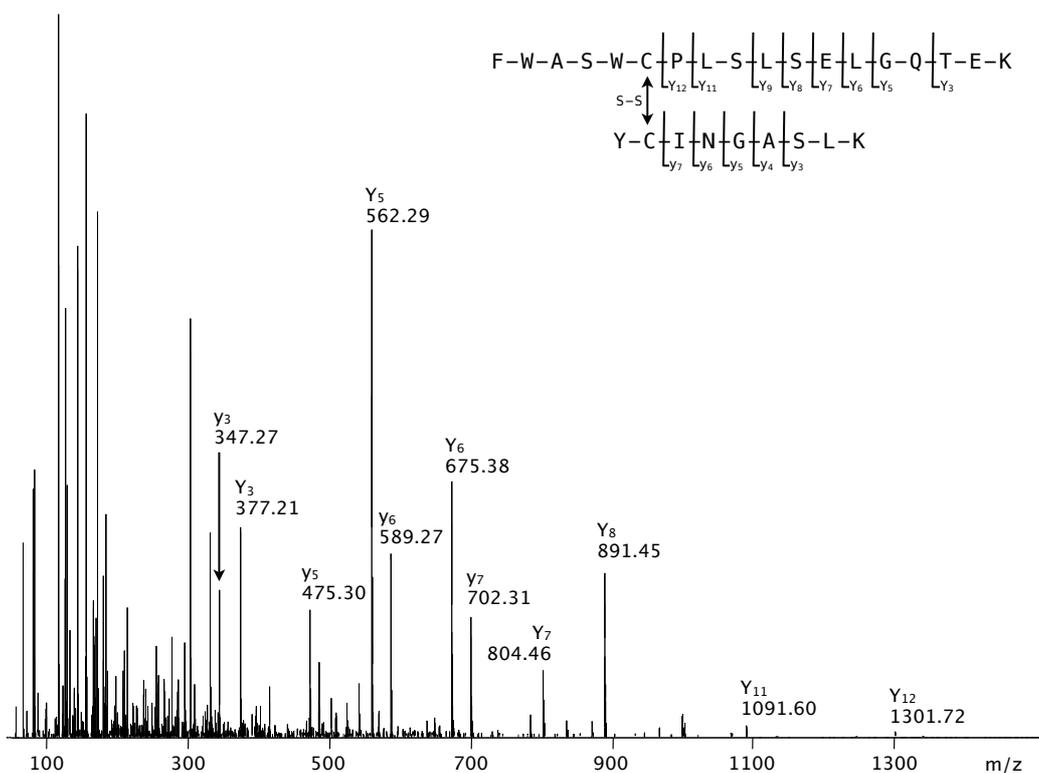
**Table 6** | Peptides caractérisés lors des analyses nanoLC-MS/MS des échantillons de PiIB réduite (a) et PiIB + MetSO (b), non traités et réduits-alkylés. (n.i. : non identifié, masse nette de l'adduit NEM : +125,0 Da, masse nette de l'adduit IAM : +57,0 Da)

▸ En absence de MetSO (Table 6a), aucun pont disulfure n'a pu être caractérisé. Par contre, chacun des peptides possédant une cystéine a été identifié : avec un adduit NEM uniquement, pour l'échantillon non traité et avec un adduit NEM ou un adduit IAM pour l'échantillon réduit-alkylé.

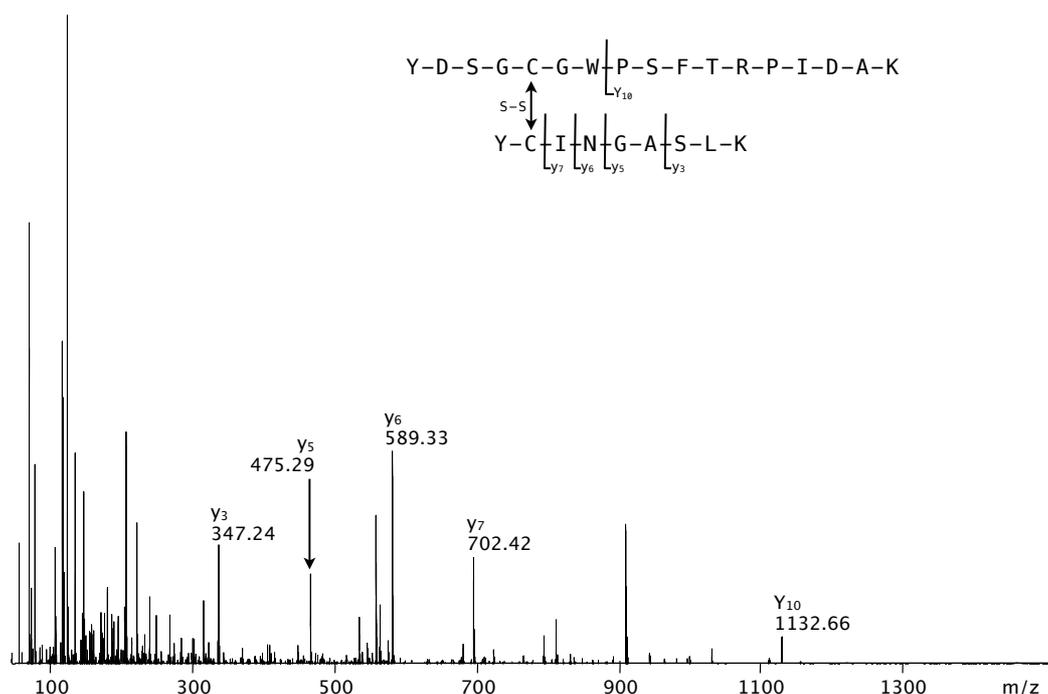
▸ Dans le cas de l'échantillon oxydé par la MetSO (Table 6b), chacun des trois ponts disulfures possibles a été caractérisé. Leurs spectres de fragmentation respectifs sont présentés en figure 20 à 22. Ces ponts sont détruits par un traitement réducteur et ne sont donc pas identifiés après réduction et alkylation de l'échantillon. Par ailleurs, on retrouve des peptides portant une cystéine avec un adduit NEM pour l'échantillon non traité ; et un adduit NEM ou un adduit IAM pour l'échantillon réduit-alkylé.



**Figure 20** | Spectre de fragmentation de l'ion  $[M+5H]^{5+}$  à 797,0 m/z, correspondant au dipeptide PiB 62-79 - PiB 435-451 à pont disulfure (C67 - C439), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $y_n$  ( $n=3-10$ ).



**Figure 21** | Spectre de fragmentation de l'ion  $[M+4H]^{4+}$  à 762,9 m/z, correspondant au dipeptide PiB 62-79 - PiB 493-501 à pont disulfure (C67 - C494), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $Y_n$  ( $n=3-12$ ).



**Figure 22** | Spectre de fragmentation de l'ion  $[M+4H]^{4+}$  à 717,3 m/z, correspondant au dipeptide PiIB 435-451 - PiIB 493-501 à pont disulfure (C439 - C494), obtenu après digestion trypsique sans traitement réducteur préalable. Les fragments majeurs correspondent à la série  $y_n$  ( $n=3-7$ ).

### 2.3.3. Conclusions

► **Le domaine MsrB oxydé est réduit par le domaine N-ter selon un mécanisme intramoléculaire.**

En effet, la mesure de masse entière de la protéine PiIB oxydée révèle la formation d'un pont disulfure intramoléculaire. Cette observation est confortée non seulement par la caractérisation du pont disulfure C439-C494 qui révèle l'oxydation de PiIB par la MetSO, mais aussi par les deux autres formes disulfures possibles entre le domaine N-ter et le domaine MsrB (liaisons C67-C439 et C67-C494). L'existence de ces liaisons disulfures entre les deux domaines prouve que la réduction du domaine MsrB par le domaine N-ter est réalisée selon un mécanisme intramoléculaire qui n'implique d'une molécule de PiIB.

► Tout indique donc que **les trois liaisons disulfures possibles PiIB C67 - C439, PiIB C67 - C494 et PiIB C439 - 494 sont susceptibles de se former** lors de la réduction d'une méthionine sulfoxyde par le domaine MsrB, suivi de la régénération de ce domaine par le domaine N-ter de la même molécule. En effet, l'analyse de l'échantillon oxydé par la MetSO et non traité au DTT montre que les trois complexes co-existent alors que toutes les cystéines en mesure de réagir ont été alkylées (et donc inactivées) par le N-Ethylmaléimide. La présence du peptide C439-C494 montre que PiIB a bien été oxydé. Par ailleurs, la caractérisation des deux autres peptides semble indiquer que l'attaque de la cystéine 67 du domaine N-ter peut avoir lieu sur l'une ou l'autre des cystéines C439 ou C494 du domaine MsrB oxydé.

Le contrôle négatif, représenté par l'analyse des échantillons non soumis à la MetSO, montre qu'aucune cystéine n'est en mesure d'être oxydée en absence de MetSO. Par ailleurs, la présence des modifications NEM et IAM sur chacune d'elles dans le cas réduit-alkylé, souligne qu'une compétition entre les deux réactifs est possible ou que l'alkylation initiale par le NEM n'est pas homogène. Cette deuxième hypothèse est renforcée par les résultats LC-MS détaillés précédemment, où il apparaît que la protéine PiIB réduite n'est pas toujours totalement alkylée après un traitement au NEM.

Enfin, en soumettant les échantillons oxydés à un traitement de réduction et d'alkylation par l'iodoacétamide, les liaisons cystéine-cystéine sont bien rompues et on retrouve le mélange des peptides de PiIB dont la cystéine présente un adduit NEM ou IAM.

Toutefois, comme dans le cas de l'étude précédente du mécanisme de réduction de MsrB par Trx, il est peu vraisemblable que les cystéines C439 et C494 de PilB puissent être attaquées indifféremment par la cystéine de recyclage C67 du domaine N-ter. Nos résultats soulignent une nouvelle fois qu'il n'est probablement pas possible d'observer le produit cinétique d'une réaction rapide. Ils reflètent vraisemblablement la formation des produits thermodynamiques, qui est consécutive aux réarrangements disulfures pouvant avoir lieu.

## 2.4. Conclusions

Cette étude a permis de montrer pour la protéine PilB que la réaction de réduction du domaine MsrB par le domaine N-ter suivait un mécanisme intramoléculaire.

Par ailleurs, les résultats obtenus grâce à notre stratégie nanoLC-MS/MS ont confirmé l'oxydation du domaine MsrB et la formation d'un pont disulfure entre les 2 cystéines qu'il contient. Cette liaison est par la suite réduite par la cystéine C67 du domaine N-ter. Si cette attaque a pu être démontrée, il n'a pas été possible de trancher sur la question de la cystéine de MsrB prioritairement attaquée. La question cinétique est à nouveau probablement la cause de ce résultat incertain.

Des études mettant en oeuvre l'analyse de mutants de PilB inactivant l'une ou l'autre des cystéines du domaine de MsrB sont susceptibles d'éclairer davantage sur la question. On peut ici aussi réfléchir également à de nouvelles expériences visant à mettre en évidence l'évolution cinétique des complexes formés grâce au développement de stratégies utilisant un marquage isotopique.

Une publication présentant ces résultats de spectrométrie de masse et ceux de nos collaborateurs obtenus par des méthodes en enzymologie est en cours de rédaction.

# **CONCLUSION GÉNÉRALE**



---

## Conclusion générale

Les objectifs de mon travail de thèse se sont articulés sur deux thématiques :

- Le développement de nouvelles méthodologies alliant l'utilisation combinée de nouveaux outils et stratégies pour l'amélioration d'analyses protéomiques classiques.
- La mise au point d'approches pour la caractérisation par spectrométrie de masse de complexes protéiques covalents à liaison cystéine–cystéine.

En réponse à ces objectifs, les optimisations techniques et stratégiques réalisées durant ce travail de thèse ont pu être appliquées à une série de problématiques biologiques.

### Développements méthodologiques pour l'analyse protéomique classique

Au-delà de la simple identification de quelques protéines contenues dans un échantillon, les approches visant à évaluer l'expression protéique différentielle au sein de plusieurs échantillons biologiques demeurent plus délicates. La première partie de mon travail de thèse a consisté à optimiser la prise de données protéomiques pour améliorer le nombre et la qualité des identifications.

▸ Les analyses protéomiques par nanoLC–MS/MS présentent un caractère aléatoire intrinsèque qui implique de fait une perte d'information importante. La question du nombre d'analyses à réaliser afin d'obtenir l'image la plus complète possible du contenu protéique identifiable a été posée grâce à l'**étude des phénomènes de répétabilité des analyses nanoLC–MS/MS**. En fonction de l'instrument utilisé, ce travail a montré que 2 à 3 analyses consécutives peuvent être nécessaires pour identifier 80% des protéines accessibles. Près de 30 % de protéines supplémentaires sont accessibles en dupliquant les analyses.

▸ Les **nouveaux outils et instruments** à disposition ont ouvert de nouvelles perspectives. L'optimisation des gradients d'élution de peptides et le choix de colonnes adaptées sur les dispositifs de chromatographie sur puce microfluidique ou à ultrahaute performance (UHPLC) ont permis d'améliorer sensibilité et résolution chromatographique. La définition des paramètres optimaux pour l'utilisation d'un système « *lock mass* » de calibration en continu sur notre spectromètre de masse de type Q–TOF a offert des précisions de mesure de masse inférieure au ppm, sans dérive d'étalonnage au cours de l'analyse. Enfin, les spectromètres de masse de nouvelle génération embarquent une électronique toujours plus rapide qui permet d'améliorer la vitesse de balayage en masse et donc la rapidité des cycles d'analyse MS/MS. La mise au point des paramètres d'acquisition (seuils de détections des ions à fragmenter, temps de scans, paramètres de fragmentions...) a permis d'améliorer la qualité et les performances de l'analyse.

▸ En aval de l'analyse à proprement parler, le traitement des données grâce à des **outils bioinformatiques** prend de plus en plus d'importance pour l'automatisation de l'analyse protéomique. A ce titre, l'optimisation des algorithmes permettant de convertir les données brutes en fichiers de listes de masses acceptés par les moteurs de recherche dans les banques protéiques a par exemple permis

d'améliorer de manière drastique la qualité et le nombre des identifications. Par ailleurs, le développement de méthodes de validation est aujourd'hui incontournable. L'évaluation de taux de faux positifs parmi les identifications a été rendue possible par la mise au point de stratégies *Target-Decoy*.

Ces paramètres, stratégies et outils pour l'analyse protéomique ont été utilisés et adaptés dans le cadre de deux projets :

▸ **La recherche de marqueurs membranaires spécifiques de leucémies au diagnostic ambigu.** Le diagnostic de certaines hémopathies lymphoïdes chroniques reste difficile dans un certain nombre de cas, comme les pathologies rares ou notamment les lymphomes non hodgkiniens pour lesquels des critères morphologiques discriminants ou des biomarqueurs sont souvent inconnus. Avec l'objectif d'identifier de nouveaux marqueurs de surface, une approche protéomique a été développée pour l'analyse de microparticules issues de la membrane plasmique de lymphocytes B de patients uniques. Nos développements méthodologiques ont permis une comparaison optimale du contenu protéique de chaque échantillon. Puis des critères de sélection et de validation rigoureux, déterminés grâce aux outils bioinformatiques mis au point, ont conduit à la proposition de candidats biomarqueurs. Le cluster de différenciation CD148 (Q12913), sélectionné pour sa pertinence biologique, a été validé par cytométrie en flux sur un grand nombre de patients. Pour la première fois, une stratégie protéomique a été en mesure de proposer un biomarqueur potentiel intéressant.

▸ **La recherche de partenaires d'interaction d'un facteur de transcription Trim24.** L'inactivation de Trim24, connue pour jouer un rôle crucial dans le contrôle de l'activité transcriptionnelle des récepteurs nucléaires, conduit à une activation aberrante de la voie de signalisation de l'acide rétinoïque (AR) dans les hépatocytes, provoquant le développement d'hépatocarcinomes. Pour déterminer quels sont les mécanismes régulateurs qui sont contrôlés par Trim24, une étude des partenaires d'interaction du cosuppresseur a été menée. A partir d'un modèle de cellules HELA, les partenaires de Trim24 ont été purifiés puis analysés selon une stratégie protéomique optimisée. Nous avons notamment pu suggérer l'existence d'une interaction entre Trim24 et Trim33. Plusieurs protéines intervenant dans des complexes impliqués dans les mécanismes d'élongation de la transcription et dans des sous unités du protéasome ont été mises en évidence. Enfin, en comparant des échantillons traités et non traités à l'acide rétinoïque, nous avons évalué les modifications qu'il pouvait induire sur les complexes impliquant Trim24.

## Développement de méthodes pour la caractérisation de complexes protéiques covalents à liaison cystéine–cystéine

Parmi les 20 acides aminés naturels, les cystéines jouent un rôle tout particulier. Elles sont en effet impliquées dans de nombreux mécanismes biologiques d'oxydation et de réduction où elles peuvent se trouver sous forme de sulfhydryles ou de ponts disulfures. Ces modifications jouent un rôle clé dans la structure des protéines et dans leurs fonctions biologiques. Ainsi, pour décrire et comprendre les mécanismes d'oxydoréduction catalytiques, il est primordial de pouvoir caractériser les protéines impliquées et déterminer l'état d'oxydation de leurs cystéines. Cela reste cependant un défi dans de nombreux cas.

Une série de stratégies mettant en œuvre différents outils de spectrométrie de masse pour traiter les cas concrets apportés par nos collaborateurs enzymologistes a été développée. Dans un premier temps, pour l'analyse de complexes covalents entiers, les conditions expérimentales de formation des complexes ont été finement travaillées pour favoriser la formation des espèces d'intérêt tout en restant compatibles avec l'analyse par spectrométrie de masse. La stabilisation des intermédiaires réactionnels a par ailleurs fait l'objet de nombreuses mises au point : mutation des cystéines réactives et blocage des réactions parasites par ajout de réactifs alkylants ou par traitement acide, par exemple. Finalement, outre le choix de conditions de séparation chromatographique adaptées, l'analyse en elle-même a nécessité un réglage très fin de nos spectromètres de masse de type ESI-Q/TOF pour obtenir les meilleures performances résolutes.

---

Afin de localiser précisément les cystéines en interaction et la nature des liaisons en jeu, il a été développé dans un deuxième temps une stratégie passant par la purification des intermédiaires réactionnels, leur digestion enzymatique et l'analyse des peptides générés (et particulièrement les dipeptides) par spectrométrie de masse en tandem. Le choix de stratégies de réduction et d'alkylation différentielles et la sélection d'enzymes adaptées et complémentaires par exemple ont été explorés en détail.

Ces méthodes de caractérisation d'intermédiaires réactionnels mis en jeu dans des cycles catalytiques ont été développées et appliquées à des problèmes biologiques traitant de la défense cellulaire contre le stress oxydant.

▸ **La formation d'un intermédiaire thiosulfinate dans le mécanisme de réduction de la peroxyrédoxine suroxydée (Prx) par la sulfirédoxine (Srx) chez *Saccharomyces cerevisiae*** a été étudiée. Il a notamment été montré que l'intermédiaire thiosulfinate était rapidement hydrolysé sous forme de disulfure. La coexistence des deux espèces, non séparées par chromatographie liquide, a requis les capacités résolutive maximales de nos instruments pour qu'elles puissent être discriminées. En outre, l'instabilité de la forme thiosulfinate a rendu difficile la caractérisation directe du dipeptide par nanoLC-MS/MS. Cependant, les produits de l'hydrolyse du complexe (dimère de Srx, complexe disulfure...) ont pu être clairement séquencés, prouvant l'existence de l'intermédiaire thiosulfinate.

▸ La deuxième étude porte sur les **mécanismes de régénération de protéines oxydées par des méthionine sulfoxyde réductases chez *Neisseria meningitidis*** (MrsB et PilB). Les Msr, étroitement liées au stress oxydant, sont des protéines capables de catalyser la réaction de réduction de protéines oxydées. Elles interviennent selon un mécanisme encore mal connu qui implique une cystéine catalytique sur laquelle se forme un intermédiaire acide sulfénique puis un pont disulfure, lui-même réduit par la thiorédoxine (ubiquitaire). Nos stratégies mises au point pour stabiliser, identifier et purifier les intermédiaires formés ont été utilisées pour finalement caractériser les complexes et déterminer les cystéines impliquées dans chacun d'eux grâce à des analyses différentielles nanoLC-MS/MS.

## Perspectives

Dans ses premières années, le domaine de l'analyse de protéines par spectrométrie de masse s'est élargi rapidement en intégrant de nombreuses innovations technologiques. Mais beaucoup d'avancées méthodologiques, instrumentales, bioinformatiques sont encore nécessaires pour être en mesure de caractériser toutes les protéines contenues dans des échantillons complexes.

Mon travail de thèse a permis d'améliorer la préparation d'échantillons, les stratégies et paramétrages analytiques et la validation des données générées pour une meilleure confiance dans l'information recueillie. Appliquées à des problématiques biologiques, ces avancées ont permis une utilisation plus fine et plus fiable de la spectrométrie de masse en protéomique. Cela est particulièrement vrai en ce qui concerne les approches de comparaison du contenu protéique d'échantillons biologiques complexes ou la caractérisation d'intermédiaires réactionnels catalytiques.

D'une manière générale, il me semble important de garder à l'esprit que l'analyse protéomique par spectrométrie de masse est une discipline relativement jeune et en plein essor. J'ai tenté, dans l'introduction bibliographique et tout au long de la présentation des résultats de ce travail de thèse, de mettre en lumière les tendances actuelles dans l'évolution de l'analyse protéomique par spectrométrie de masse. Trois axes de développement me semblent particulièrement importants :

▸ De nouveaux instruments, toujours plus sensibles, résolutifs et aux vitesses d'acquisition accrues sont constamment développés. Les techniques de séparation de peptides et de protéines gagnent en résolution, en robustesse et sont capables de séparer des mélanges toujours plus complexes.

▸ Parallèlement, les stratégies d'identification, de caractérisation et de quantification des protéines gagnent en maturité. Tout particulièrement ces dernières années, la prise de conscience de la nécessité

d'évaluer systématiquement la confiance des résultats d'identification de protéines en analyse protéomique s'est installée.

▸ Les outils bioinformatiques de validation, d'évaluation et partage des informations jouent alors un rôle de plus en plus déterminant dans la formalisation et la régulation des résultats produits en analyse protéomique par spectrométrie de masse.

De nombreux défis restent toutefois en suspens en ce qui concerne la caractérisation qualitative et quantitative de protéomes. La multitude et la diversité des modifications posttraductionnelles possibles ajoutent un degré de complexité difficile à maîtriser et nécessitent des technologies et stratégies analytiques encore meilleures. Il reste qu'un des plus gros défis de l'analyse protéomique par spectrométrie de masse réside dans l'utilisation de l'information générée : son maniement, sa validation et son partage ne sont toujours pas parfaitement établis.

# **BIBLIOGRAPHIE**



---

**A**

- Aebersold, R.  
A stress test for mass spectrometry-based proteomics.  
*Nat Methods* (2009).
- Aebersold, R. and Mann, M.  
Mass spectrometry-based proteomics.  
*Nature* (2003) **422**, 198–207.
- Aguilar, M.-I.  
HPLC of Peptides and Proteins: Methods and Protocols.  
Humana Press, Totowa (New Jersey), (2004).
- Aguilar, M. I. and Hearn, M. T.  
High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins.  
*Methods Enzymol* (1996) **270**, 3–26.
- Altucci, L. and Gronemeyer, H.  
The promise of retinoids to fight against cancer.  
*Nat Rev Cancer* (2001) **1**, 181–93.
- Alves, G., Wu, W. W., Wang, G., Shen, R. F. and Yu, Y. K.  
Enhancing peptide identification confidence by combining search methods.  
*J Proteome Res* (2008) **7**, 3102–13.
- Andrade, M. A. and Sander, C.  
Bioinformatics: from genome data to biological knowledge.  
*Curr Opin Biotechnol* (1997) **8**, 675–83.
- Apweiler, R., Bairoch, A. and Wu, C. H.  
Protein sequence databases.  
*Curr Opin Chem Biol* (2004) **8**, 76–80.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G.  
Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.  
*Nat Genet* (2000) **25**, 25–9.
- Ashman, K., Moran, M. F., Sicheri, F., Pawson, T. and Tyers, M.  
Cell signalling – the proteomics of it all.  
*Sci STKE* (2001) **2001**, PE33.
- Atmanène, C.  
Nouvelles perspectives pour l'analyse and la caractérisation structurale de complexes biologiques par spectrométrie de masse supramoléculaire.  
Thèse de doctorat. Université Louis Pasteur. Strasbourg (2009).

**B**

- Baczek, T. and Kaliszczan, R.  
Predictions of peptides' retention times in reversed-phase liquid chromatography as a new supportive tool to improve protein identification in proteomics.  
*Proteomics* (2009) **9**, 835–47.
- Bantscheff, M., Dumpelfeld, B. and Kuster, B.  
Femtomol sensitivity post-digest (18)O labeling for relative quantification of differential protein complex composition.  
*Rapid Commun Mass Spectrom* (2004) **18**, 869–76.
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B.  
Quantitative mass spectrometry in proteomics: a critical review.  
*Anal Bioanal Chem* (2007) **389**, 1017–31.
- Bard, J. B. and Rhee, S. Y.  
Ontologies in biology: design, applications and future challenges.  
*Nat Rev Genet* (2004) **5**, 213–22.
- Barton, S. J. and Whittaker, J. C.  
Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors.  
*Mass Spectrom Rev* (2009) **28**, 177–87.
- Bayes, M. and Price, M.  
An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.  
*Philosophical Transactions* (1763) **53**, 370–418.

Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., Beavis, R., Sechi, S., Nilsson, T., Bergeron, J. J., Beardslee, T. A., Chappell, T., Meredith, G., Sheffield, P., Gray, P., Hajivandi, M., Pope, M., Predki, P., Kullolli, M., Hincapie, M., Hancock, W. S., Jia, W., Song, L., Li, L., Wei, J., Yang, B., Wang, J., Ying, W., Zhang, Y., Cai, Y., Qian, X., He, F., Meyer, H. E., Stephan, C., Eisenacher, M., Marcus, K., Langenfeld, E., May, C., Carr, S. A., Ahmad, R., Zhu, W., Smith, J. W., Hanash, S. M., Struthers, J. J., Wang, H., Zhang, Q., An, Y., Goldman, R., Carlsohn, E., van der Post, S., Hung, K. E., Sarracino, D. A., Parker, K., Krastins, B., Kucherlapati, R., Bourassa, S., Poirier, G. G., Kapp, E., Patsiouras, H., Moritz, R., Simpson, R., Houle, B., Laboissiere, S., Metalnikov, P., Nguyen, V., Pawson, T., Wong, C. C., Cociorva, D., Yates Iii, J. R., Ellison, M. J., Lopez-Campistrous, A., Semchuk, P., Wang, Y., Ping, P., Elia, G., Dunn, M. J., Wynne, K., Walker, A. K., Strahler, J. R., Andrews, P. C., Hood, B. L., Bigbee, W. L., Conrads, T. P., Smith, D., Borchers, C. H., Lajoie, G. A., Bendall, S. C., Speicher, K. D., Speicher, D. W., Fujimoto, M., Nakamura, K., Paik, Y. K., Cho, S. Y., Kwon, M. S., Lee, H. J., Jeong, S. K., Chung, A. S., Miller, C. A., Grimm, R., Williams, K., Dorschel, C., Falkner, J. A., Martens, L. and Vizcaino, J. A.

A HUPO test sample study reveals common problems in mass spectrometry-based proteomics.  
*Nat Methods* (2009).

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W.  
GenBank.

*Nucleic Acids Res* (2009) **37**, D26–31.

Beretta, L.

Proteomics from the clinical perspective: many hopes and much debate.  
*Nat Methods* (2007) **4**, 785–6.

Berggard, T., Linse, S. and James, P.

Methods for the detection and analysis of protein-protein interactions.  
*Proteomics* (2007) **7**, 2833–42.

Bern, M., Cai, Y. and Goldberg, D.

Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry.  
*Anal Chem* (2007) **79**, 1393–400.

Bialecki, E. S. and Di Bisceglie, A. M.

Clinical presentation and natural course of hepatocellular carcinoma.  
*Eur J Gastroenterol Hepatol* (2005) **17**, 485–9.

Biemann, K.

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

Bilusich, D. and Bowie, J. H.

Identification of intermolecular disulfide linkages in underivatized peptides using negative ion electrospray mass spectrometry. A joint experimental and theoretical study.  
*Rapid Commun Mass Spectrom* (2007) **21**, 619–28.

Binz, P. A., Barkovich, R., Beavis, R. C., Creasy, D., Horn, D. M., Julian, R. K., Jr., Seymour, S. L., Taylor, C. F. and Vandenbrouck, Y.

Guidelines for reporting the use of mass spectrometry informatics in proteomics.  
*Nat Biotechnol* (2008) **26**, 862.

Biteau, B., Labarre, J. and Toledano, M. B.

ATP-dependent reduction of cysteine-sulphinic acid by *S. cerevisiae* sulphiredoxin.  
*Nature* (2003) **425**, 980–4.

Blueggel, M., Chamrad, D. and Meyer, H. E.

Bioinformatics in proteomics.  
*Curr Pharm Biotechnol* (2004) **5**, 79–88.

Boyd, R. S., Jukes-Jones, R., Walewska, R., Brown, D., Dyer, M. J. and Cain, K.

Protein profiling of plasma membranes defines aberrant signaling pathways in mantle cell lymphoma.  
*Mol Cell Proteomics* (2009) **8**, 1501–15.

Bradshaw, R. A.

Revised draft guidelines for proteomic data publication.  
*Mol Cell Proteomics* (2005) **4**, 1223–5.

Bradshaw, R. A., Burlingame, A. L., Carr, S. and Aebersold, R.

Reporting protein identification data: the next generation of guidelines.  
*Mol Cell Proteomics* (2006) **5**, 787–8.

Brot, N., Weissbach, L., Werth, J. and Weissbach, H.

Enzymatic reduction of protein-bound methionine sulfoxide.  
*Proc Natl Acad Sci U S A* (1981) **78**, 2155–8.

C

Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V. and Schamel, W. W.

Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach.  
*Mol Cell Proteomics* (2004) **3**, 176–82.

Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., Mulder, N., Oinn, T., Maslen, J., Cox, A. and Apweiler, R.

The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro.  
*Genome Res* (2003) **13**, 662–72.

Campana, J. E.

Elementary Theory of the Quadrupole Mass Filter.  
*International Journal of Mass Spectrometry and Ion Processes* (1980) **33**, 101–117.

Carapito, C.

Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique.  
Thèse de doctorat. Université Louis Pasteur. Strasbourg (2006).

---

Cargile, B. J., Bundy, J. L., Freeman, T. W. and Stephenson, J. L., Jr.  
Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification.  
*J Proteome Res* (2004) **3**, 112–9.

Chargaff, E. and West, R.  
The biological significance of the thromboplastic protein of blood.  
*Journal of Biological Chemistry* (1946) **166**, 189–197.

Charles, L.  
Flow injection of the lock mass standard for accurate mass measurement in electrospray ionization time-of-flight mass spectrometry coupled with liquid chromatography.  
*Rapid Commun Mass Spectrom* (2003) **17**, 1383–8.

Chepanoske, C. L., Richardson, B. E., von Rechenberg, M. and Peltier, J. M.  
Average peptide score: a useful parameter for identification of proteins derived from database searches of liquid chromatography/tandem mass spectrometry data.  
*Rapid Commun Mass Spectrom* (2005) **19**, 9–14.

Chernushevich, I. V., Loboda, A. V. and Thomson, B. A.  
An introduction to quadrupole-time-of-flight mass spectrometry.  
*J Mass Spectrom* (2001) **36**, 849–65.

Chevreux, G.  
Etude des interactions moléculaires spécifiques par spectrométrie de masse : application à la chimie du vivant.  
Thèse de doctorat. Université Louis Pasteur. Strasbourg (2005).

Chumsae, C., Gaza-Bulsecu, G. and Liu, H.  
Identification and Localization of Unpaired Cysteine Residues in Monoclonal Antibodies by Fluorescence Labeling and Mass Spectrometry.  
*Anal Chem* (2009).

Ciorba, M. A., Heinemann, S. H., Weissbach, H., Brot, N. and Hoshi, T.  
Regulation of voltage-dependent K<sup>+</sup> channels by methionine oxidation: effect of nitric oxide and vitamin C.  
*FEBS Lett* (1999) **442**, 48–52.

Cochrane, G., Akhtar, R., Bonfield, J., Bower, L., Demiralp, F., Faruque, N., Gibson, R., Hoad, G., Hubbard, T., Hunter, C., Jang, M., Juhos, S., Leinonen, R., Leonard, S., Lin, Q., Lopez, R., Lorenc, D., McWilliam, H., Mukherjee, G., Plaister, S., Radhakrishnan, R., Robinson, S., Sobhany, S., Hoopen, P. T., Vaughan, R., Zalunin, V. and Birney, E.  
Petabyte-scale innovations at the European Nucleotide Archive.  
*Nucleic Acids Res* (2009) **37**, D19–25.

Cohen, J.  
Beyond the genome – the proteomics payoff.  
*Technology Review* (2001) **104**, 54–60.

Cohen, S. L. and Chait, B. T.  
Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels.  
*Anal Biochem* (1997) **247**, 257–67.

Colinge, J., Masselot, A., Cusin, I., Mahe, E., Niknejad, A., Argoud-Puy, G., Reffas, S., Bederr, N., Gleizes, A., Rey, P. A. and Bougueleret, L.  
High-performance peptide identification by tandem mass spectrometry allows reliable automatic data processing in proteomics.  
*Proteomics* (2004) **4**, 1977–84.

Cottingham, K.  
MCP ups the ante by mandating raw-data deposition.  
*J Proteome Res* (2009) **8**, 4887–8.

Craig, F. E. and Foon, K. A.  
Flow cytometric immunophenotyping for hematologic neoplasms.  
*Blood* (2008) **111**, 3941–67.

Craig, R. and Beavis, R. C.  
TANDEM: matching proteins with tandem mass spectra.  
*Bioinformatics* (2004) **20**, 1466–7.

Cravatt, B. F., Simon, G. M. and Yates, J. R., 3rd  
The biological impact of mass-spectrometry-based proteomics.  
*Nature* (2007) **450**, 991–1000.

D

de Hoffmann, E. and Stroobant, V.  
Mass Spectrometry: Principles and Applications.  
Wiley, Chichester, (2007).

Dehmelt, H.  
Radiofrequency Spectroscopy of Stored Ions.  
*Adv. At. Mol. Phys.* (1967) **3**, 53.

Deutsch, E. W., Lam, H. and Aebersold, R.  
PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows.  
*EMBO Rep* (2008) **9**, 429–34.

Di Bisceglie, A. M., Osmack, P. and Brunt, E. M.  
Chemoprevention of hepatocellular carcinoma: use of tamoxifen in an animal model of hepatocarcinogenesis.  
*J Lab Clin Med* (2005) **145**, 134–8.

Dimmer, E. C., Huntley, R. P., Barrell, D. G., Binns, D., Draghici, S., Camon, E. B., Hubank, M., Talmud, P. J., Apweiler, R. and Lovering, R. C.  
The Gene Ontology – Providing a Functional Role in Proteomic Studies.  
*Proteomics* (2008).

Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D. and Alice, M. B.  
Molecular Beams of Macroions.  
The Journal of Chemical Physics (1968) **49**, 2240–2249.

Domon, B. and Aebersold, R.  
Mass spectrometry and protein analysis.  
Science (2006) **312**, 212–7.

Dongre, A. R., Somogyi, A. and Wysocki, V. H.  
Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides.  
J Mass Spectrom (1996) **31**, 339–50.

Douglas, D. J.  
Linear quadrupoles in mass spectrometry.  
Mass Spectrom Rev (2009) **28**, 937–60.

Durr, E., Yu, J., Krasinska, K. M., Carver, L. A., Yates, J. R., Testa, J. E., Oh, P. and Schnitzer, J. E.  
Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture.  
Nat Biotechnol (2004) **22**, 985–92.

E

Eastman, Q.  
Proteomics researchers solidifying principles for data sharing.  
J Proteome Res (2009) **8**, 3220.

Eckers, C., Wolff, J. C., Haskins, N. J., Sage, A. B., Giles, K. and Bateman, R.  
Accurate mass liquid chromatography/mass spectrometry on orthogonal acceleration time-of-flight mass analyzers using switching between separate sample and reference sprays. 1. Proof of concept.  
Anal Chem (2000) **72**, 3683–8.

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

Ek, S. and Borrebaeck, C. A.  
Parallel gene expression profiling of mantle cell lymphoma – how do we transform 'omics data into clinical practice.  
Curr Genomics (2007) **8**, 171–9.

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

Elias, J. E., Haas, W., Faherty, B. K. and Gygi, S. P.  
Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations.  
Nat Methods (2005) **2**, 667–75.

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.  
Journal of the American Society for Mass Spectrometry (1994) **5**, 976–989.

F

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

Fields, S.  
Proteomics. Proteomics in genomeland.  
Science (2001) **291**, 1221–4.

Fomenko, D. E., Marino, S. M. and Gladyshev, V. N.  
Functional diversity of cysteine residues in proteins and unique features of catalytic redox-active cysteines in thiol oxidoreductases.  
Mol Cells (2008) **26**, 228–35.

Fournier, M. L., Gilmore, J. M., Martin-Brown, S. A. and Washburn, M. P.  
Multidimensional separations-based shotgun proteomics.  
Chem Rev (2007) **107**, 3654–86.

Freyssinet, J. M.  
Cellular microparticles: what are they bad or good for?  
J Thromb Haemost (2003) **1**, 1655–62.

Fritz, R., Ruth, W. and Kragl, U.  
Assessment of acetone as an alternative to acetonitrile in peptide analysis by liquid chromatography/mass spectrometry.  
Rapid Commun Mass Spectrom (2009) **23**, 2139–45.

G

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

- 
- Gattiker, A., Bienvenut, W. V., Bairoch, A. and Gasteiger, E.  
FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification.  
*Proteomics* (2002) **2**, 1435–44.
- Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. and Bryant, S. H.  
Open mass spectrometry search algorithm.  
*J Proteome Res* (2004) **3**, 958–64.
- Gelpi, E.  
From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part I  
1965–1984.  
*J Mass Spectrom* (2008) **43**, 419–35.
- Gelpi, E.  
From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II  
1985–2000.  
*J Mass Spectrom* (2009) **44**, 1137–61.
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. and Gygi, S. P.  
Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.  
*Proceedings of the National Academy of Sciences of the United States of America* (2003) **100**, 6940–6945.
- Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H. and Langridge, J. I.  
The detection, correlation, and comparison of peptide precursor and product ions from data independent LC–MS with data dependant LC–MS/MS.  
*Proteomics* (2009) **9**, 1683–95.
- Gevaert, K. and Vandekerckhove, J.  
Protein identification methods in proteomics.  
*Electrophoresis* (2000) **21**, 1145–54.
- Gibson, G. T., Mugo, S. M. and Oleschuk, R. D.  
Nanoelectrospray emitters: trends and perspective.  
*Mass Spectrom Rev* (2009) **28**, 918–36.
- Giorgianni, F., Cappiello, A., Beranova-Giorgianni, S., Palma, P., Truffelli, H. and Desiderio, D. M.  
LC–MS/MS analysis of peptides with methanol as organic modifier: improved limits of detection.  
*Anal Chem* (2004) **76**, 7028–38.
- Glish, G. L. and Vachet, R. W.  
The basics of mass spectrometry in the twenty-first century.  
*Nat Rev Drug Discov* (2003) **2**, 140–50.
- Gorg, A., Weiss, W. and Dunn, M. J.  
Current two-dimensional electrophoresis technology for proteomics.  
*Proteomics* (2004) **4**, 3665–85.
- Gorman, J. J., Wallis, T. P. and Pitt, J. J.  
Protein disulfide bond determination by mass spectrometry.  
*Mass Spectrom Rev* (2002) **21**, 183–216.
- Gupta, N. and Pevzner, P.  
False discovery rates of protein identifications: a strike against the two-peptide rule.  
*J Proteome Res* (2009).
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R.  
Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.  
*Nat Biotechnol* (1999) **17**, 994–9.
- Hamacher, M., Eisenacher, M., Meyer, H. E. and Stephan, C.  
Proteomics today: bioinformatics at its best. *Proteomics and Bioinformatics -- an inseparable couple.*  
*Proteomics* (2008) **8**, 4616–7.
- Han, X., Aslanian, A. and Yates, J. R., 3rd  
Mass spectrometry for proteomics.  
*Curr Opin Chem Biol* (2008) **12**, 483–90.
- Hardouin, J., Duchateau, M., Joubert-Caron, R. and Caron, M.  
Usefulness of an integrated microfluidic device (HPLC–Chip–MS) to enhance confidence in protein identification by proteomics.  
*Rapid Commun Mass Spectrom* (2006) **20**, 3236–44.
- 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.
- Hernandez, P., Muller, M. and Appel, R. D.  
Automated protein identification by tandem mass spectrometry: issues and strategies.  
*Mass Spectrom Rev* (2006) **25**, 235–54.
- Hillenkamp, F., Karas, M., Beavis, R. C. and Chait, B. T.  
Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.  
*Anal Chem* (1991) **63**, 1193A–1203A.
- Hokenson, M. J., Uversky, V. N., Goers, J., Yamin, G., Munishkina, L. A. and Fink, A. L.  
Role of individual methionines in the fibrillation of methionine-oxidized alpha-synuclein.  
*Biochemistry* (2004) **43**, 4621–33.

Horn, D. M., Peters, E. C., Klock, H., Meyers, A. and Brock, A.  
Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting.  
*International Journal of Mass Spectrometry* (2004) **238**, 189–196.

Huber, L. A.  
Is proteomics heading in the wrong direction?  
*Nat Rev Mol Cell Biol* (2003) **4**, 74–80.

Iribarne, J. V. and Thomson, B. A.  
On the evaporation of small ions from charged droplets.  
*The Journal of Chemical Physics* (1976) **64**, 2287–2294.

Issaq, H. J.  
Application of separation technologies to proteomics research.  
*Adv Protein Chem* (2003) **65**, 249–69.

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

Jerkovich, A. D., Mellors, J. S. and Jorgenson, J. W.  
The use of micron-sized particles in ultrahigh-pressure liquid chromatography.  
*Lc Gc Europe* (2003) **16**, 20–23.

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

Jones, P., Cote, R. G., Cho, S. Y., Klie, S., Martens, L., Quinn, A. F., Thorneycroft, D. and Hermjakob, H.  
PRIDE: new developments and new datasets.  
*Nucleic Acids Res* (2008) **36**, D878–83.

Jonsson, T. J., Tsang, A. W., Lowther, W. T. and Furdai, C. M.  
Identification of intact protein thiosulfinate intermediate in the reduction of cysteine sulfinic acid in peroxiredoxin by human sulfiredoxin.  
*J Biol Chem* (2008) **283**, 22890–4.

Josic, D. and Clifton, J. G.  
Mammalian plasma membrane proteomics.  
*Proteomics* (2007) **7**, 3010–29.

Kapp, E. and Schutz, F.  
Overview of tandem mass spectrometry (MS/MS) database search algorithms.  
*Curr Protoc Protein Sci* (2007) **Chapter 25**, Unit25 2.

Karas, M., Bachmann, D., Bahr, U. and Hillenkamp, F.  
Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds.  
*International Journal of Mass Spectrometry and Ion Processes* (1987) **78**, 53–68.

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

Kebarle, P. and Verkerk, U. H.  
Electrospray: from ions in solution to ions in the gas phase, what we know now.  
*Mass Spectrom Rev* (2009) **28**, 898–917.

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

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

Khetchoumian, K.  
Caractérisation fonctionnelle des facteurs intermédiaires de transcription Tif1 $\alpha$  and Tif1 $\delta$  chez la souris.  
Thèse de doctorat. Université Louis Pasteur. Strasbourg (2005).

Khetchoumian, K., Teletin, M., Tisserand, J., Mark, M., Herquel, B., Ignat, M., Zucman-Rossi, J., Cammas, F., Lerouge, T., Thibault, C., Metzger, D., Chambon, P. and Losson, R.  
Loss of Trim24 (Tif1 $\alpha$ ) gene function confers oncogenic activity to retinoic acid receptor  $\alpha$ .  
*Nat Genet* (2007) **39**, 1500–6.

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

---

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

Koster, S. and Verpoorte, E.  
A decade of microfluidic analysis coupled with electrospray mass spectrometry: an overview.  
*Lab Chip* (2007) **7**, 1394–412.

Krause, F.  
Detection and analysis of protein–protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes.  
*Electrophoresis* (2006) **27**, 2759–81.

L

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

Lange, V., Picotti, P., Domon, B. and Aebersold, R.  
Selected reaction monitoring for quantitative proteomics: a tutorial.  
*Mol Syst Biol* (2008) **4**, 222.

Latterich, M.  
Publishing proteomic data.  
*Proteome Sci* (2006) **4**, 8.

Le Douarin, B., Zechel, C., Garnier, J. M., Lutz, Y., Tora, L., Pierrat, P., Heery, D., Gronemeyer, H., Chambon, P. and Losson, R.  
The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18.  
*EMBO J* (1995) **14**, 2020–33.

Lescuyer, P., Hochstrasser, D. and Rabilloud, T.  
How shall we use the proteomics toolbox for biomarker discovery?  
*J Proteome Res* (2007) **6**, 3371–6.

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

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

Lisacek, F., Cohen-Boulakia, S. and Appel, R. D.  
Proteome informatics II: bioinformatics for comparative proteomics.  
*Proteomics* (2006) **6**, 5445–66.

Liu, H., Finch, J. W., Lavalley, M. J., Collamati, R. A., Benevides, C. C. and Gebler, J. C.  
Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations.  
*J Chromatogr A* (2007) **1147**, 30–6.

Liu, H., Sadygov, R. G. and Yates, J. R., 3rd  
A model for random sampling and estimation of relative protein abundance in shotgun proteomics.  
*Anal Chem* (2004) **76**, 4193–201.

M

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

Macher, B. A. and Yen, T. Y.  
Proteins at membrane surfaces—a review of approaches.  
*Mol Biosyst* (2007) **3**, 705–13.

Makarov, A.  
Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.  
*Anal Chem* (2000) **72**, 1156–62.

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K. and Horning, S.  
Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.  
*Anal Chem* (2006) **78**, 2113–20.

Malmstrom, J., Lee, H., Nesvizhskii, A. I., Shteynberg, D., Mohanty, S., Brunner, E., Ye, M., Weber, G., Eckerskorn, C. and Aebersold, R.  
Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis.  
*J Proteome Res* (2006) **5**, 2241–9.

Mamyrin, B. A., Karataev, V. I., Shmikk, D. V. and Zagulin, V. A.  
The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution.  
*Sov. Phys. JETP* (1973) **37**, 45.

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

- Mann, M. and Wilm, M.  
Error-tolerant identification of peptides in sequence databases by peptide sequence tags.  
*Anal Chem* (1994) **66**, 4390–9.
- March, R. E.  
An introduction to quadrupole ion trap mass spectrometry.  
*Journal of Mass Spectrometry* (1997) **32**, 351–369.
- March, R. E.  
Quadrupole ion trap mass spectrometry: Theory, simulation, recent developments and applications.  
*Rapid Communications in Mass Spectrometry* (1998) **12**, 1543–1554.
- March, R. E.  
Quadrupole ion traps.  
*Mass Spectrom Rev* (2009) **28**, 961–89.
- Marshall, A. G. and Guan, S. H.  
Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry.  
*Rapid Commun Mass Spectrom* (1996) **10**, 1819–1823.
- Marshall, A. G., Hendrickson, C. L. and Jackson, G. S.  
Fourier transform ion cyclotron resonance mass spectrometry: a primer.  
*Mass Spectrom Rev* (1998) **17**, 1–35.
- Marshall, A. G., Hendrickson, C. L. and Shi, S. D.  
Scaling MS plateaus with high-resolution FT-ICRMS.  
*Anal Chem* (2002) **74**, 252A–259A.
- Matthiesen, R.  
Methods, algorithms and tools in computational proteomics: a practical point of view.  
*Proteomics* (2007) **7**, 2815–32.
- Matthiesen, R. and Jensen, O. N.  
Analysis of mass spectrometry data in proteomics.  
*Methods Mol Biol* (2008) **453**, 105–22.
- McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X. and Begley, T. P.  
Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics.  
*FEBS J* (2007) **274**, 6256–68.
- Mead, J. A., Bianco, L. and Bessant, C.  
Recent developments in public proteomic MS repositories and pipelines.  
*Proteomics* (2009) **9**, 861–81.
- Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J., Girault, H. H. and Rossier, J. S.  
Protein fractionation in a multicompartment device using Off-Gel isoelectric focusing.  
*Electrophoresis* (2003) **24**, 3–11.
- Miguet, L.  
Nouvelles méthodologies dans l'analyse protéomique par spectrométrie de masse. Application à la recherche de biomarqueurs dans le cadre des leucémies.  
Thèse de doctorat. Université Louis Pasteur. Strasbourg (2006).
- Miguet, L., Pacaud, K., Felden, C., Hugel, B., Martinez, M. C., Freyssinet, J. M., Herbrecht, R., Potier, N., van Dorsselaer, A. and Mauvieux, L.  
Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization.  
*Proteomics* (2006) **6**, 153–71.
- Miguet, L., Sanglier, S., Schaeffer, C., Potier, N., Mauvieux, L. and Van Dorsselaer, A.  
Microparticles: a new tool for plasma membrane sub-cellular proteomic.  
*Subcell Biochem* (2007) **43**, 21–34.
- Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J. and Hunt, D. F.  
The utility of ETD mass spectrometry in proteomic analysis.  
*Biochim Biophys Acta* (2006) **1764**, 1811–22.
- Morel, O., Toti, F., Hugel, B. and Freyssinet, J. M.  
Cellular microparticles: a disseminated storage pool of bioactive vascular effectors.  
*Curr Opin Hematol* (2004) **11**, 156–64.
- Nagele, E., Vollmer, M. and Horth, P.  
Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome.  
*J Biomol Tech* (2004) **15**, 134–43.
- Nagy, P., Lemma, K. and Ashby, M. T.  
Reactive sulfur species: kinetics and mechanisms of the reaction of cysteine thiosulfinate ester with cysteine to give cysteine sulfenic acid.  
*J Org Chem* (2007) **72**, 8838–46.
- Nesvizhskii, A. I. and Aebersold, R.  
Interpretation of shotgun proteomic data: the protein inference problem.  
*Mol Cell Proteomics* (2005) **4**, 1419–40.
- Nesvizhskii, A. I., Vitek, O. and Aebersold, R.  
Analysis and validation of proteomic data generated by tandem mass spectrometry.  
*Nat Methods* (2007) **4**, 787–97.

---

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

Neverova, I. and Van Eyk, J. E.  
Role of chromatographic techniques in proteomic analysis.  
*J Chromatogr B Analyt Technol Biomed Life Sci* (2005) **815**, 51-63.

O

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M.  
Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.  
*Mol Cell Proteomics* (2002) **1**, 376-86.

Orchard, S.  
Ending the "publish and vanish" culture: how the data standardization process will assist in data harvesting.  
*J Proteome Res* (2009) **8**, 3219.

Orchard, S. and Hermjakob, H.  
The HUPO proteomics standards initiative--easing communication and minimizing data loss in a changing world.  
*Brief Bioinform* (2008) **9**, 166-73.

Ouyang, Z., Gao, L., Fico, M., Chappell, W. J., Noll, R. J. and Cooks, R. G.  
Quadrupole ion traps and trap arrays: geometry, material, scale, performance.  
*Eur J Mass Spectrom* (Chichester, Eng) (2007) **13**, 13-8.

Ouyang, Z., Wu, G., Song, Y., Li, H., Plass, W. R. and Cooks, R. G.  
Rectilinear ion trap: concepts, calculations, and analytical performance of a new mass analyzer.  
*Anal Chem* (2004) **76**, 4595-605.

P

Padmanabhan, R., Corsico, C. D., Howard, T. H., Holter, W., Fordis, C. M., Willingham, M. and Howard, B. H.  
Purification of transiently transfected cells by magnetic affinity cell sorting.  
*Anal Biochem* (1988) **170**, 341-8.

Paizs, B. and Suhai, S.  
Fragmentation pathways of protonated peptides.  
*Mass Spectrom Rev* (2005) **24**, 508-48.

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

Panchaud, A., Affolter, M., Moreillon, P. and Kussmann, M.  
Experimental and computational approaches to quantitative proteomics: status quo and outlook.  
*J Proteomics* (2008) **71**, 19-33.

Pandey, A. and Mann, M.  
Proteomics to study genes and genomes.  
*Nature* (2000) **405**, 837-46.

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

Patel, V. J., Thalassinou, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. and Scrivens, J. H.  
A comparison of labeling and label-free mass spectrometry-based proteomics approaches.  
*J Proteome Res* (2009) **8**, 3752-9.

Patterson, S. D. and Aebersold, R. H.  
Proteomics: the first decade and beyond.  
*Nat Genet* (2003) **33 Suppl**, 311-23.

Paul, W.  
Electromagnetic Traps for Charged and Neutral Particles (Nobel Lecture).  
*Angewandte Chemie International Edition in English* (1990) **29**, 739-748.

Paul, W. and Steinwedel, H.  
Ein neues Massenspektrometer ohne Magnetfeld.  
*Zeitschrift für Naturforschung A* (1953) **8**, 448-450.

Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W. and Aebersold, R.  
A common open representation of mass spectrometry data and its application to proteomics research.  
*Nat Biotechnol* (2004) **22**, 1459-66.

Perissi, V. and Rosenfeld, M. G.  
Controlling nuclear receptors: the circular logic of cofactor cycles.  
*Nat Rev Mol Cell Biol* (2005) **6**, 542-54.

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

- Pevtsov, S., Fedulova, I., Mirzaei, H., Buck, C. and Zhang, X.  
Performance evaluation of existing de novo sequencing algorithms.  
*J Proteome Res* (2006) **5**, 3018–28.
- Pfeifer, N., Leinenbach, A., Huber, C. G. and Kohlbacher, O.  
Improving peptide identification in proteome analysis by a two-dimensional retention time filtering approach.  
*J Proteome Res* (2009) **8**, 4109–15.
- Poullet, P., Carpentier, S. and Barillot, E.  
myProMS, a web server for management and validation of mass spectrometry-based proteomic data.  
*Proteomics* (2007) **7**, 2553–6.
- Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinou, K., Bateman, R. H., Bowers, M. T. and Scrivens, J. H.  
An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument.  
*International Journal of Mass Spectrometry* (2007) **261**, 1–12.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Seraphin, B.  
The tandem affinity purification (TAP) method: a general procedure of protein complex purification.  
*Methods* (2001) **24**, 218–29.
- Rabilloud, T.  
Membrane proteins ride shotgun.  
*Nat Biotechnol* (2003) **21**, 508–10.
- Rabilloud, T., Chevallet, M., Luche, S. and Lelong, C.  
Fully denaturing two-dimensional electrophoresis of membrane proteins: a critical update.  
*Proteomics* (2008) **8**, 3965–73.
- Rabilloud, T., Heller, M., Gasnier, F., Luche, S., Rey, C., Aebersold, R., Benahmed, M., Louisot, P. and Lunardi, J.  
Proteomics analysis of cellular response to oxidative stress. Evidence for in vivo overoxidation of peroxiredoxins at their active site.  
*J Biol Chem* (2002) **277**, 19396–401.
- 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.
- Rabilloud, T., Vaezzadeh, A. R., Potier, N., Lelong, C., Leize-Wagner, E. and Chevallet, M.  
Power and limitations of electrophoretic separations in proteomics strategies.  
*Mass Spectrom Rev* (2009) **28**, 816–43.
- Reisinger, F. and Martens, L.  
Database on Demand – an online tool for the custom generation of FASTA-formatted sequence databases.  
*Proteomics* (2009) **9**, 4421–4.
- Resing, K. A., Meyer-Arendt, K., Mendoza, A. M., Aveline-Wolf, L. D., Jonscher, K. R., Pierce, K. G., Old, W. M., Cheung, H. T., Russell, S., Wattawa, J. L., Goehle, G. R., Knight, R. D. and Ahn, N. G.  
Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics.  
*Anal Chem* (2004) **76**, 3556–68.
- Rhee, S. G.  
Cell signaling. H<sub>2</sub>O<sub>2</sub>, a necessary evil for cell signaling.  
*Science* (2006) **312**, 1882–3.
- Rhee, S. Y., Wood, V., Dolinski, K. and Draghici, S.  
Use and misuse of the gene ontology annotations.  
*Nat Rev Genet* (2008) **9**, 509–15.
- Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E. and Rabilloud, T.  
About the mechanism of interference of silver staining with peptide mass spectrometry.  
*Proteomics* (2004) **4**, 909–16.
- Rifai, N., Gillette, M. A. and Carr, S. A.  
Protein biomarker discovery and validation: the long and uncertain path to clinical utility.  
*Nat Biotechnol* (2006) **24**, 971–83.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B.  
A generic protein purification method for protein complex characterization and proteome exploration.  
*Nat Biotechnol* (1999) **17**, 1030–2.
- Righetti, P. G. and Gianazza, E.  
Isoelectric focusing in immobilized pH gradients: theory and newer methodology.  
*Methods Biochem Anal* (1987) **32**, 215–78.
- Rodriguez, H., Snyder, M., Uhlen, M., Andrews, P., Beavis, R., Borchers, C., Chalkley, R. J., Cho, S. Y., Cottingham, K., Dunn, M., Dylag, T., Edgar, R., Hare, P., Heck, A. J., Hirsch, R. F., Kennedy, K., Kolar, P., Kraus, H. J., Mallick, P., Nesvizhskii, A., Ping, P., Ponten, F., Yang, L., Yates, J. R., Stein, S. E., Hermjakob, H., Kinsinger, C. R. and Apweiler, R.  
Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: the Amsterdam principles.  
*J Proteome Res* (2009) **8**, 3689–92.
- Ros, A., Faupel, M., Mees, H., van Oostrum, J., Ferrigno, R., Reymond, F., Michel, P., Rossier, J. S. and Girault, H. H.  
Protein purification by Off-Gel electrophoresis.  
*Proteomics* (2002) **2**, 151–156.

- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A. and Pappin, D. J.  
Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents.  
*Mol Cell Proteomics* (2004) **3**, 1154–69.
- Sadygov, R. G., Cociorva, D. and Yates, J. R., 3rd  
Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book.  
*Nat Methods* (2004) **1**, 195–202.
- Salmi, J., Nyman, T. A., Nevalainen, O. S. and Aittokallio, T.  
Filtering strategies for improving protein identification in high-throughput MS/MS studies.  
*Proteomics* (2009) **9**, 848–60.
- Sandra, K., Moshir, M., D'Hondt, F., Tuytten, R., Verleysen, K., Kas, K., Francois, I. and Sandra, P.  
Highly efficient peptide separations in proteomics. Part 2: bi- and multidimensional liquid-based separation techniques.  
*J Chromatogr B Analyt Technol Biomed Life Sci* (2009) **877**, 1019–39.
- Sandra, K., Moshir, M., D'Hondt, F., Verleysen, K., Kas, K. and Sandra, P.  
Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography.  
*J Chromatogr B Analyt Technol Biomed Life Sci* (2008) **866**, 48–63.
- Santoni, V., Molloy, M. and Rabilloud, T.  
Membrane proteins and proteomics: un amour impossible?  
*Electrophoresis* (2000) **21**, 1054–70.
- Santos-Rosa, H. and Caldas, C.  
Chromatin modifier enzymes, the histone code and cancer.  
*Eur J Cancer* (2005) **41**, 2381–402.
- Savitski, M. M., Nielsen, M. L., Kjeldsen, F. and Zubarev, R. A.  
Proteomics-grade de novo sequencing approach.  
*J Proteome Res* (2005) **4**, 2348–54.
- Savitski, M. M., Nielsen, M. L. and Zubarev, R. A.  
New data base-independent, sequence tag-based scoring of peptide MS/MS data validates Mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques.  
*Mol Cell Proteomics* (2005) **4**, 1180–8.
- Scallon, B. J., Snyder, L. A., Anderson, G. M., Chen, Q., Yan, L., Weiner, L. M. and Nakada, M. T.  
A review of antibody therapeutics and antibody-related technologies for oncology.  
*J Immunother* (2006) **29**, 351–64.
- Scherl, A., Shaffer, S. A., Taylor, G. K., Hernandez, P., Appel, R. D., Binz, P. A. and Goodlett, D. R.  
On the benefits of acquiring peptide fragment ions at high measured mass accuracy.  
*J Am Soc Mass Spectrom* (2008) **19**, 891–901.
- Scherl, A., Tsai, Y. S., Shaffer, S. A. and Goodlett, D. R.  
Increasing information from shotgun proteomic data by accounting for misassigned precursor ion masses.  
*Proteomics* (2008) **8**, 2791–7.
- Seiwert, B., Hayen, H. and Karst, U.  
Differential labeling of free and disulfide-bound thiol functions in proteins.  
*J Am Soc Mass Spectrom* (2008) **19**, 1–7.
- Shevchenko, A., Sunyaev, S., Loboda, A., Bork, P., Ens, W. and Standing, K. G.  
Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching.  
*Anal Chem* (2001) **73**, 1917–26.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M.  
Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.  
*Analytical Chemistry* (1996) **68**, 850–8.
- Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. and Geromanos, S. J.  
Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.  
*Mol Cell Proteomics* (2006) **5**, 144–56.
- Skaar, E. P., Tobiason, D. M., Quick, J., Judd, R. C., Weissbach, H., Etienne, F., Brot, N. and Seifert, H. S.  
The outer membrane localization of the *Neisseria gonorrhoeae* MsrA/B is involved in survival against reactive oxygen species.  
*Proc Natl Acad Sci U S A* (2002) **99**, 10108–13.
- Speers, A. E. and Wu, C. C.  
Proteomics of integral membrane proteins—theory and application.  
*Chem Rev* (2007) **107**, 3687–714.
- Steen, H. and Mann, M.  
The ABC's (and XYZ's) of peptide sequencing.  
*Nat Rev Mol Cell Biol* (2004) **5**, 699–711.
- Storey, J. D. and Tibshirani, R.  
Statistical significance for genomewide studies.  
*Proc Natl Acad Sci U S A* (2003) **100**, 9440–5.
- Strittmatter, E. F., Kangas, L. J., Petritis, K., Mottaz, H. M., Anderson, G. A., Shen, Y., Jacobs, J. M., Camp, D. G., 2nd and Smith, R. D.  
Application of peptide LC retention time information in a discriminant function for peptide identification by tandem mass spectrometry.  
*J Proteome Res* (2004) **3**, 760–9.

Sugawara, H., Ikeo, K., Fukuchi, S., Gojobori, T. and Tateno, Y.  
DDBJ dealing with mass data produced by the second generation sequencer.  
*Nucleic Acids Res* (2009) **37**, D16–8.

Suter, B., Kittanakom, S. and Stagljar, I.  
Interactive proteomics: what lies ahead?  
*Biotechniques* (2008) **44**, 681–91.

Swamy, M., Siegers, G. M., Minguet, S., Wollscheid, B. and Schamel, W. W.  
Blue native polyacrylamide gel electrophoresis (BN–PAGE) for the identification and analysis of multiprotein complexes.  
*Sci STKE* (2006) **2006**, pl4.

Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F.  
Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.  
*Proc Natl Acad Sci U S A* (2004) **101**, 9528–33.

T

Tan, S., Tan, H. T. and Chung, M. C.  
Membrane proteins and membrane proteomics.  
*Proteomics* (2008) **8**, 3924–32.

Tanaka, K.  
The origin of macromolecule ionization by laser irradiation (Nobel lecture).  
*Angew Chem Int Ed Engl* (2003) **42**, 3860–70.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. and Matsuo, T.  
Protein and polymer analyses up to  $m/z$  100 000 by laser ionization time-of-flight mass spectrometry.  
*Rapid Commun Mass Spectrom* (1988) **2**, 151–153.

Taylor, C. F., Binz, P. A., Aebersold, R., Affolter, M., Barkovich, R., Deutsch, E. W., Horn, D. M., Huhmer, A., Kussmann, M., Lilley, K., Macht, M., Mann, M., Muller, D., Neubert, T. A., Nickson, J., Patterson, S. D., Raso, R., Resing, K., Seymour, S. L., Tsugita, A., Xenarios, I., Zeng, R. and Julian, R. K., Jr.  
Guidelines for reporting the use of mass spectrometry in proteomics.  
*Nat Biotechnol* (2008) **26**, 860–1.

Taylor, C. F., Paton, N. W., Lilley, K. S., Binz, P. A., Julian, R. K., Jr., Jones, A. R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E. W., Dunn, M. J., Heck, A. J., Leitner, A., Macht, M., Mann, M., Martens, L., Neubert, T. A., Patterson, S. D., Ping, P., Seymour, S. L., Souda, P., Tsugita, A., Vandekerckhove, J., Vondriska, T. M., Whitelegge, J. P., Wilkins, M. R., Xenarios, I., Yates, J. R., 3rd and Hermjakob, H.  
The minimum information about a proteomics experiment (MIAPE).  
*Nat Biotechnol* (2007) **25**, 887–93.

TheUniProtConsortium  
The Universal Protein Resource (UniProt) in 2010.  
*Nucleic Acids Res* (2009).

Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F. and Jungblut, P. R.  
Peptide mass fingerprinting.  
*Methods* (2005) **35**, 237–47.

Thomas, H. and Shevchenko, A.  
Simplified validation of borderline hits of database searches.  
*Proteomics* (2008) **8**, 4173–7.

Thompson, J. J.  
[Rays of positive electricity and their application to chemical analysis.](#)  
Longmans Green, London, (1913).

Tyers, M. and Mann, M.  
From genomics to proteomics.  
*Nature* (2003) **422**, 193–7.

U

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

V

Vestal, M. L., Juhasz, P. and Martin, S. A.  
Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry.  
*Rapid Communications in Mass Spectrometry* (1995) **9**, 1044–1050.

W

Wagner, E., Luche, S., Penna, L., Chevallet, M., Van Dorsseleer, A., Leize-Wagner, E. and Rabilloud, T.  
A method for detection of overoxidation of cysteines: peroxiredoxins are oxidized in vivo at the active-site cysteine during oxidative stress.  
*Biochem J* (2002) **366**, 777–85.

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

- 
- Washburn, M. P., Wolters, D. and Yates, J. R., 3rd  
Large-scale analysis of the yeast proteome by multidimensional protein identification technology.  
*Nat Biotechnol* (2001) **19**, 242–7.
- Waugh, D. S.  
Making the most of affinity tags.  
*Trends Biotechnol* (2005) **23**, 316–20.
- Wehofsky, M. and Hoffmann, R.  
Automated deconvolution and deisotoping of electrospray mass spectra.  
*J Mass Spectrom* (2002) **37**, 223–9.
- Wielsch, N., Thomas, H., Surendranath, V., Waridel, P., Frank, A., Pevzner, P. and Shevchenko, A.  
Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches.  
*J Proteome Res* (2006) **5**, 2448–56.
- Wiley, W. C. and McLaren, I. H.  
Time-of-Flight Mass Spectrometer with Improved Resolution.  
*Review of Scientific Instruments* (1955) **26**, 1150–1157.
- Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W. and Dunn, M. J.  
Guidelines for the next 10 years of proteomics.  
*Proteomics* (2006) **6**, 4–8.
- Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L. and Hochstrasser, D. F.  
From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.  
*Biotechnology (N Y)* (1996) **14**, 61–5.
- Wilm, M.  
Quantitative proteomics in biological research.  
*Proteomics* (2009).
- Wilm, M. and Mann, M.  
Analytical properties of the nanoelectrospray ion source.  
*Anal Chem* (1996) **68**, 1–8.
- Wittig, I. and Schagger, H.  
Features and applications of blue-native and clear-native electrophoresis.  
*Proteomics* (2008) **8**, 3974–90.
- Wolters, D. A., Washburn, M. P. and Yates, J. R., 3rd  
An automated multidimensional protein identification technology for shotgun proteomics.  
*Analytical Chemistry* (2001) **73**, 5683–90.
- Woo, H. A., Chae, H. Z., Hwang, S. C., Yang, K. S., Kang, S. W., Kim, K. and Rhee, S. G.  
Reversing the inactivation of peroxiredoxins caused by cysteine sulfinic acid formation.  
*Science* (2003) **300**, 653–6.
- Wood, B.  
9-color and 10-color flow cytometry in the clinical laboratory.  
*Arch Pathol Lab Med* (2006) **130**, 680–90.
- Wood, Z. A., Poole, L. B. and Karplus, P. A.  
Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling.  
*Science* (2003) **300**, 650–3.
- Wu, C. C. and MacCoss, M. J.  
Shotgun proteomics: tools for the analysis of complex biological systems.  
*Curr Opin Mol Ther* (2002) **4**, 242–50.
- Wu, C. C., MacCoss, M. J., Howell, K. E. and Yates, J. R., 3rd  
A method for the comprehensive proteomic analysis of membrane proteins.  
*Nat Biotechnol* (2003) **21**, 532–8.
- Wu, N., Lippert, J. A. and Lee, M. L.  
Practical aspects of ultrahigh pressure capillary liquid chromatography.  
*J Chromatogr A* (2001) **911**, 1–12.
- X
- Xiang, T., Chumsae, C. and Liu, H.  
Localization and quantitation of free sulfhydryl in recombinant monoclonal antibodies by differential labeling with 12C and 13C iodoacetic acid and LC-MS analysis.  
*Anal Chem* (2009) **81**, 8101–8.
- Xu, H., Zhang, L. and Freitas, M. A.  
Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine.  
*J Proteome Res* (2008) **7**, 138–44.
- Y
- 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.

Yen, T. Y., Joshi, R. K., Yan, H., Seto, N. O., Palcic, M. M. and Macher, B. A.  
Characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography/electrospray ionization tandem mass spectrometry.  
J Mass Spectrom (2000) **35**, 990–1002.

Yen, T. Y., Yan, H. and Macher, B. A.  
Characterizing closely spaced, complex disulfide bond patterns in peptides and proteins by liquid chromatography/electrospray ionization tandem mass spectrometry.  
J Mass Spectrom (2002) **37**, 15–30.

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

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

Z

Zenobi, R. and Knochenmuss, R.  
Ion formation in MALDI mass spectrometry.  
Mass Spectrometry Reviews (1998) **17**, 337–366.

# ANNEXES



**ANNEXE A1**

*relative à : Résultats - 1<sup>re</sup> partie, chapitre II, point 2.4.*

Données supplémentaires de la publication :

Miguet, L. et Béchade, G. (2009). *J Proteome Res* **8**, 3346–3354.

**ANNEXE A2**

*relative à : Résultats - 1<sup>re</sup> partie, chapitre III, point 3.2.2.*

Protéines partenaires potentiels de Trim24, identifiées dans les extraits ENI et ENS après purification TAP Flag-HA

**ANNEXE A3**

*relative à : Résultats - 1<sup>re</sup> partie, chapitre III, point 3.2.3.*

Protéines partenaires potentiels de Trim33, identifiées dans les extraits ENI et ENS après purification TAP Flag-HA

**ANNEXE A3**

*relative à : Résultats - 1<sup>re</sup> partie, chapitre III, point 3.2.4.*

Protéines partenaires potentiels de Trim24, identifiées dans les extraits ENI et ENS après purification TAP Flag-HA, avec ou sans traitement préalable à l'acide rétinoïque.



Institut Pluridisciplinaire Hubert Curien

Service de reprographie  
Michel Schlée

23, rue du Loess  
67037 Strasbourg Cedex 2