



# Spectrométrie de masse des modifications induites ou post-traductionnelles de protéines : méthodologie et application à des protéines d'intérêt thérapeutique

Guillaume Gabant

## ► To cite this version:

Guillaume Gabant. Spectrométrie de masse des modifications induites ou post-traductionnelles de protéines : méthodologie et application à des protéines d'intérêt thérapeutique. Biochimie, Biologie Moléculaire. Université d'Orléans, 2014. Français. <NNT : 2014ORLE2061>. <tel-01246054>

**HAL Id: tel-01246054**

**<https://tel.archives-ouvertes.fr/tel-01246054>**

Submitted on 18 Dec 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**ÉCOLE DOCTORALE SANTÉ, SCIENCES BIOLOGIQUES  
ET CHIMIE DU VIVANT**

CENTRE DE BIOPHYSIQUE MOLECULAIRE

**THÈSE** présentée par :  
**Guillaume GABANT**

soutenue le : **17 décembre 2014**

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline / Spécialité : Spectrométrie de Masse

**Spectrométrie de masse des modifications  
induites ou post-traductionnelles de protéines :  
méthodologie et application à des protéines  
d'intérêt thérapeutique**

**THÈSE dirigée par :**  
**Martine CADÈNE**

Directeur de Recherche, CNRS Orléans

**RAPPORTEURS :**  
**Catherine GUETTE**  
**Jean-Michel CAMADRO**

Chargée de Recherche, ICO Paul Papin Angers  
Directeur de Recherche, CNRS Paris 5

**JURY :**

**Marc BOUDVILLAIN**

Directeur de Recherche, CNRS Orléans  
Président du jury

**Catherine GUETTE**

Chargée de Recherche, ICO Paul Papin Angers

**Christine KELLENBERGER**

Chargée de Recherche, CNRS Marseille

**Jean-Michel CAMADRO**

Directeur de Recherche, CNRS Paris 5

**Martine CADÈNE**

Directeur de Recherche, CNRS Orléans



# Remerciements

*Ce travail a été réalisé au sein du Centre de Biophysique Moléculaire, UPR4301 du CNRS, situé à Orléans.*

*Mes premiers remerciements vont à Martine Cadène qui m'a permis de faire cette thèse en parallèle à mon activité de responsable technique de plateforme. Je tiens à lui exprimer ma profonde reconnaissance pour m'avoir fait confiance, pour son encadrement et sa disponibilité pendant ces cinq années, pour sa rigueur scientifique et ses nombreux conseils.*

*Je remercie Catherine Guette et Jean-Michel Camadro pour avoir accepté d'être rapporteurs de cette thèse.*

*Je remercie également les autres membres de mon jury, Christine Kellenberger et Marc Boudvillain d'avoir accepté de juger ce travail.*

*Je remercie les personnes avec qui nous avons travaillé en collaboration : Corinne Augé-Gouillou pour la collaboration sur MOS1, Francis Gauthier Baris Korkmaz et Christophe Epinette pour la collaboration sur la gréglène, ainsi que Bertrand Castaing, Françoise Culard et Stéphane Goffinont pour la collaboration sur Abf2. Petit clin d'œil à Stéphane avec qui j'ai travaillé plus particulièrement.*

*Mes remerciements s'adressent également à tous les participants du projet METASUPP, notamment Françoise Schoentgen et Ludivine Errien pour la production de la protéine hPEBP1 et Vincent Aucagne pour les discussions utiles sur la locostatine.*

*Je remercie Alain Boyer et Yannick Berteaux pour la programmation de l'approche segmentation ainsi que Stéphane Bourg et Norbert Garnier pour la modélisation moléculaire.*

*Je tiens à remercier tous les membres actuels de mon équipe, Martine Beaufour, Cyril Colas, David Ginguené mais aussi les anciens, Lucie Jaquillard, Emmanuelle Mebold et Kevin Mark, pour nos multiples discussions et leur sympathie.*

*J'adresse mes plus vifs remerciements à Justo Torres et Martial Bruneau pour leur sympathie et leurs multiples dépannages. Grâce à vous, j'ai pu passer plus de temps sur ma thèse ! Justo, encore merci pour m'avoir trouvé un détecteur de fuite en plines vacances d'été ; sans toi, je serais encore en train de démonter tout le spectro !*

*Je tiens à remercier particulièrement,*

*David Gosset, pour nos discussions quotidiennes toujours agréables, pour son amitié, sa disponibilité et son authenticité, une denrée rare de nos jours...*

*Mes coéquipiers de foot du BRGM, notamment Marc Oltra, Guillaume Lointier, Stéphane Gaboreau, Pascal Perrier et Stéphane Guillemoto, pour leur sympathie et pour faire vivre ce club !*

*Lucie Jaquillard et Hélène Adihou, pour toutes nos grandes discussions et les sorties en dehors du labo.*

*Je souhaite remercier chaleureusement mes parents et Noëlle pour leur soutien et leur encouragement, ainsi que mon ami Julien, mon frère Yohan et Claire, toujours là pour partager des moments de vie, sincèrement merci !*

*Enfin un grand merci plein d'amour à ma chère et tendre épouse Christine et mes enfants Clément et Martin, pour avoir été auprès de moi tout au long de ce périple à la fois enrichissant et sinueux qu'est la thèse.*

# Liste des abréviations

4-HCCA	Acide $\alpha$ -cyano-4-hydroxycinnamique
ABC	Bicarbonate d'ammonium
Abf2	Autonomously replicating sequence (ARS)-binding factor 2
ACTH	Acétylcholine
ADN	Acide DésoxyriboNucléique
AMP	Adénosine MonoPhosphate
AMPc	AMP cyclique
AP	site apurinique/apyrimidique
ARN	Acide RiboNucléique
ATP	Adenosine TriPhosphate
CAD	Collision Activated Dissociation
CE	Cap Exit
CID	Collision Induced Dissociation
Da	Dalton
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
DMSO	DiMéthylSulfOxyde
DTT	DiThioThreitol
ECD	Electron Capture Dissociation
ESI	ElectroSpray Ionisation
ETD	Electron Transfer Dissociation
Fpg	Formamidopyrimidine-DNA glycosylase
FMN	Flavine MonoNucléotide
FWHM	Full Width at Half Maximum
FWI	Formic acid/Water/Isopropanol (3/1/2)
GSH	Glutathion libre
GTP	Guanosine TriPhosphate
HEPES	2-(4-(2-HydroxyEthyl)-1-Piperazinyl)EthaneSulfonic Acid
HMGB	High Mobility Group Box
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
IMAC	Immobilized Metal Affinity Chromatography
isCID	In source CID
IT	Ion Trap
$k_{off}$	constante de vitesse de dissociation
$k_{on}$	constante de vitesse d'association
$K_D$	constante de dissociation à l'équilibre
MALDI	Matrix Assisted Laser Desorption Ionisation
MAP	Mitogen-activated protein
MS	Mass Spectrometry
m/z	rapport masse sur charge
MRM	Multiple Reaction Monitoring
NF- $\kappa$ B	Nuclear Factor kappa B
PEBP	PhosphatidylEthanolamine Binding Protein
p/p	poids/poids
PTM	Post-Translational Modification

Q	Quadrupole
RKIP	Raf Kinase Inhibitory Protein
RMN	Résonance Magnétique Nucléaire
ROS	Reactive Oxygen Species
SRM	Selected Reaction Monitoring
TCEP	Tris(2-CarboxyEthyl)Phosphine
TFA	TriFluoroacetic Acid
TRIS	Tris(hydroxyméthyl)aminométhane
TOF	Time Of Flight
TWA	TFA 0.1%, Water 66.6%, Acetonitrile 33.3% (v/v)
WFA	Water 50%, Acetonitrile 49.8%, Formic Acid 0.2% (v/v)

# Table des matières

<b>Introduction .....</b>	<b>11</b>
<b>Section 1. Les modifications de protéines en médecine.....</b>	<b>12</b>
1. Les relations structure-fonction des protéines .....	12
2. Les modifications de protéines .....	12
2.1 Définition .....	12
2.2 Rôles et conséquences des modifications de protéines .....	13
2.3 Les types de modifications de protéines .....	13
2.4 Les modifications post-traductionnelles.....	14
2.4.1 Généralités .....	14
2.4.2 La phosphorylation .....	19
2.4.2.1 Formation, destruction et effets moléculaires et cellulaires .....	19
2.4.2.2 Les méthodes de détection .....	20
2.4.3 La N-acétylation.....	20
2.4.4 La protéolyse.....	22
2.5 Les modifications induites .....	23
2.5.1 Modification induite <i>in vivo</i> , exemple de la fixation covalente d'un ligand.....	24
2.5.2 Modification induite <i>in vitro</i> , exemple de l'oxydation des protéines.....	27
2.5.2.1 Les espèces réactives de l'oxygène (ROS).....	27
2.5.2.2 L'oxydation des acides aminés.....	28
2.5.2.3 Recommandations pour éviter l'oxydation accidentelle des protéines .....	30
2.5.2.4 Les autres modifications induites accidentelles.....	30
3. Intérêt des médicaments covalents en utilisation thérapeutique .....	34
3.1 Sécurité sanitaire .....	36
3.2 Sélectivité .....	38
3.3 Efficacité.....	38
3.4 Conception et optimisation de nouveaux inhibiteurs covalents.....	39
<b>Section 2. Les bases de la caractérisation de modification de protéines par spectrométrie de masse.....</b>	<b>42</b>
1. Identification de modifications et localisation de sites modifiés .....	43
1.1 Détermination du type et du nombre de modifications sur protéine entière.....	43
1.2 Identification et localisation de modifications par approche bottom-up .....	44
1.2.1 La protéolyse : génération de peptides.....	44
1.2.2 Détection des peptides modifiés .....	47
1.2.2.1 Empreinte peptidique massique.....	47
1.2.2.2 Détection en MS/MS : Modes d'analyse en spectrométrie de masse avec fragmentation.....	48
1.2.3 Fractionnement et enrichissement des peptides modifiés .....	49
1.2.3.1 Difficultés associées à la détection des peptides modifiés en mélange.....	49
1.2.3.2 Fractionnement par HPLC .....	49
1.2.3.3 Enrichissement d'espèces modifiées .....	51
1.2.4 Localisation de résidus modifiés .....	53
1.2.4.1 Décomposition dans la source (ISD) .....	55
1.2.4.2 Décomposition après la source (PSD) .....	55
1.2.4.3 Dissociation induite par collision (CID) .....	55
1.2.4.4 Dissociation par transfert d'électrons (ETD) .....	57
1.2.4.5 Photodissociation .....	57
1.3 Identification et localisation de modifications par approche top-down.....	58
1.3.1 Les difficultés des approches top-down.....	59
1.3.1.1 Séparation LC .....	59
1.3.1.2 Sensibilité .....	59
1.3.1.3 Fragmentation .....	59
1.3.1.4 Complexité du mélange de protéines.....	60
1.3.2 Les pistes de résolution de problèmes .....	60
1.3.2.1 Les nouveaux supports chromatographiques.....	60
1.3.2.2 La photodissociation.....	62



1.4	Traitement et Interprétation des spectres de masses (Bioinformatique).....	62
2.	De la quantification de protéines à la quantification de leurs modifications .....	64
2.1	Bottom-up vs top-down .....	64
2.2	Quantification de protéine.....	65
2.2.1	La question biologique .....	67
2.2.1.1	Approche globale .....	67
2.2.1.2	Approche ciblée .....	67
2.2.2	Les différentes méthodes de quantification par spectrométrie de masse.....	67
2.2.2.1	Les méthodes pour une quantification relative.....	68
2.2.2.2	Les méthodes pour une quantification absolue.....	72
2.2.3	Modes d'acquisition .....	76
2.2.3.1	Modes d'acquisition DDA et DIA.....	76
2.2.3.2	Modes d'acquisition SRM-MRM.....	77
2.2.3.3	Considérations expérimentales.....	78
2.3	Quantification de modification.....	82
2.3.1	Les objectifs .....	82
2.3.1.1	Quantification de chaque protéoforme d'une protéine.....	84
2.3.1.2	Quantification de la fraction globale de protéine modifiée .....	84
2.3.1.3	Quantification de la stœchiométrie totale de modification de la protéine entière.....	85
2.3.1.4	Quantification de la stœchiométrie de modification d'un ou plusieurs site(s).....	85
2.3.2	Méthodes disponibles pour la quantification de modifications .....	86
2.3.2.1	Quantification de protéoformes.....	86
2.3.2.2	Quantification de sites modifiés .....	86
2.3.3	Ce qu'il manque en approche bottom-up pour quantifier les modifications de résidus de manière robuste et exhaustive.....	91

## **Présentation du travail de thèse.....93**

1.	Contexte.....	94
2.	Objectifs du travail de thèse .....	95
2.1	Corriger la séquence de la gréglie et localiser ses PTMs .....	95
2.2	Caractériser le complexe MOS1-inhibiteur .....	95
2.3	Localiser le site actif d'Abf2 .....	95
2.4	Caractériser l'interaction entre hPEBP1 et la locostatine afin de fournir des éléments pour la synthèse raisonnée d'activateurs des propriétés antimétastatiques naturelles de hPEBP1 .....	95
2.5	Développer des outils pour la caractérisation de modifications de protéines .....	96

## **Résultats et Discussion.....97**

### **Chapitre 1 Une nouvelle modification de la gréglie, un inhibiteur de protéases à sérine.....99**

1.	Contexte.....	101
2.	Publication .....	102
3.	Supplementary informations .....	117
4.	Annexes à la publication .....	129
4.1	Analyse des cristaux du complexe gréglie-subtilisine par spectrométrie de masse.....	129
4.2	Détermination de la formule brute de la modification inconnue de la gréglie.....	130
4.2.1	Interprétation du spectre pseudoMS3 de la modification inconnue .....	130
4.2.2	Recherche de la formule brute de la modification inconnue .....	132
4.3	Recherche dans les bases de données .....	140
5.	Perspectives.....	140

### **Chapitre 2 Etude du complexe entre la transposase MOS1 et des inhibiteurs.....141**

1.	Contexte.....	143
2.	Objectifs.....	145
3.	Etude de la protéine MBP-MOS1 produite dans <i>Escherichia coli</i> .....	146
3.1	Analyse de la pureté de la protéine.....	146
3.2	Analyse de la pureté des inhibiteurs .....	148
3.3	Analyse du complexe entre la protéine MBP-MOS1 <sub>E.coli</sub> et l'inhibiteur MCV41 .....	150
3.4	Analyse de la pureté de la protéine MBP-MOS1 produite en cellules d'insecte .....	151
4.	Publication .....	154
5.	Conclusions.....	167
6.	Perspectives.....	167

<b>Chapitre 3</b>	<b>Vers la localisation du site actif d'Abf2, une protéine impliquée dans la réparation de l'ADN mitochondrial.</b>	<b>169</b>
1.	Contexte.....	171
1.1	La protéine Abf2 .....	171
1.2	Analyse d'hétéroconjugés protéine-ADN ou peptide-oligonucléotide par spectrométrie de masse....	173
2.	Objectif du projet.....	173
3.	Contrôle qualité de la protéine Abf2 et de son substrat ADN .....	174
3.1	Analyse de la pureté d' Abf2.....	174
3.2	Analyse de la pureté des oligonucléotides .....	175
4.	Génération de duplex ADN contenant un site AP ou 5' dRP, substrat d' Abf2 .....	177
5.	Stratégie expérimentale 1.....	178
5.1	Formation d'un complexe Abf2-ADN covalent et irréversible .....	180
5.2	Enrichissement du complexe Abf2-ADN.....	180
5.3	Dégradation enzymatique de la protéine et de l'oligonucléotide.....	180
6.	Les résultats de localisation de la stratégie 1.....	181
6.1	Identification des hétéroconjugés peptide-acides nucléiques .....	181
6.2	Localisation du site actif présumé .....	184
6.3	Vérification des résultats par mutagenèse dirigée.....	186
7.	Stratégie expérimentale 2.....	189
8.	Les résultats de localisation de la stratégie 2.....	191
8.1	Détermination de la stœchiométrie du complexe Abf2-ADN .....	191
8.2	Identification des hétéroconjugés peptide-acides nucléiques .....	192
8.3	Aspects méthodologiques .....	197
8.3.1	L'apport du deuxième enrichissement monoQ.....	197
8.3.2	Le cas du peptide trypsique [1-16] de l'étiquette 6 Histidine.....	197
8.3.3	Mise en évidence d'isomères par fragmentation ETD.....	198
9.	Conclusions.....	200
10.	Perspectives .....	201
<b>Chapitre 4</b>	<b>Caractérisation de l'interaction entre la locostatine et la protéine PEBP1/RKIP humaine, un suppresseur de métastases naturel.....</b>	<b>203</b>
1.	Contexte.....	205
1.1	La protéine PEBP1/RKIP.....	205
1.2	Structure de PEBP1 humaine .....	205
1.3	Les fonctions de PEBP1.....	207
1.3.1	PEBP1 et voies de signalisation .....	207
1.3.1.1	La voie Ras/Raf/MEK/ERK.....	208
1.3.1.2	La voie Nf-κB .....	209
1.3.1.3	La voie des récepteurs couplés aux protéines G.....	210
1.3.2	PEBP et cancer.....	211
1.3.2.1	Cancer et métastase .....	211
1.3.2.2	Activité suppresseur de métastases de hPEBP1 .....	212
1.3.2.3	Activité pro-apoptotique de hPEBP1.....	213
1.4	La locostatine .....	215
1.4.1	Activité anti-migratoire.....	215
1.4.2	Interaction avec PEBP1 .....	217
2.	Objectifs.....	218
3.	Analyse de la pureté de hPEBP1 et de la locostatine.....	218
3.1	Analyse de la pureté de la locostatine.....	218
3.2	Analyse de la pureté de hPEBP1 .....	220
3.2.1	Le stock initial de hPEBP1 est pur mais contient des modifications induites non désirées .....	220
3.2.2	Le deuxième stock de hPEBP1 est pur et adapté pour l'étude du complexe avec la locostatine....	222
4.	Dissection du mécanisme de liaison .....	223
4.1	Nature de la liaison.....	223
4.2	Stabilité du complexe .....	224
4.3	Cinétique d'association.....	229
4.4	Stœchiométrie du complexe.....	231
5.	Optimisation des conditions de réaction et de protéolyse pour la localisation du site spécifique.....	232
5.1	Evaluation de la réactivité de la locostatine .....	232
5.1.1	Réaction de la locostatine avec des agents réducteurs.....	233
5.1.2	Réaction de la locostatine avec des peptides.....	233
5.2	Solutions pour limiter les réactions non spécifiques.....	235
5.2.1	Au niveau de la protéine .....	235

5.2.2	Au niveau des peptides .....	235
5.2.2.1	Neutralisation de la locostatine en excès .....	235
5.2.2.2	Elimination de la locostatine en excès par TCEP immobilisé sur bille.....	239
5.2.2.3	Elimination de la locostatine en excès par micro gel filtration.....	239
5.3	Optimisation des conditions de protéolyse.....	241
5.3.1	Efficacité de protéolyse .....	241
5.3.2	Cartographie peptidique.....	244
6.	Essai de localisation du site spécifique - Approche qualitative .....	247
6.1	Analyse MS des peptides .....	248
6.2	Analyse MS/MS des peptides.....	251
6.3	Limites de l'approche qualitative.....	256
7.	Modélisation mathématique pour discriminer le site spécifique des sites non spécifiques .....	257
7.1	Objectif du modèle .....	257
7.2	Description mathématique du modèle.....	258
7.3	Représentations graphiques du modèle.....	263
7.3.1	Cinétiques comparées d'un site spécifique avec un site non-spécifique.....	263
7.3.2	Contribution du site spécifique à la liaison globale .....	264
7.3.2.1	Représentation en fonction du temps.....	264
7.3.2.2	Représentation en fonction de fbound site total .....	264
7.3.3	Influence de n.....	266
7.3.4	Influence du rapport $k_{onS}/k_{onNS}$ .....	267
7.3.5	Facteur de discrimination.....	268
8.	Essai de localisation du site spécifique – Approche quantitative.....	271
8.1	Développement d'une nouvelle méthode pour la quantification exhaustive et parallèle du degré de modification à chaque site .....	271
8.2	Compilation des résultats des approches qualitative et quantitative .....	319
9.	Modélisation moléculaire de l'interaction hPEBP1-locostatine .....	319
10.	Conclusions .....	321
11.	Perspectives .....	322

**Conclusion générale et perspectives .....325**

**Matériel et méthodes.....331**

1.	Produits biologiques et chimiques .....	332
2.	Purification d'hétéroconjugués Abf2-ADN.....	333
2.1	Formation du duplex ADN contenant un site AP ou 5'dRP.....	333
2.2	Formation d'un complexe covalent Abf2-ADN .....	333
2.3	Purification du complexe covalent Abf2-ADN en deux étapes (stratégie expérimentale 1) .....	333
2.3.1	Elimination des protéines libres par chromatographie d'affinité avidine-biotine .....	333
2.3.2	Elimination des oligonucléotides libres.....	334
2.4	Purification du complexe covalent Abf2-ADN par chromatographie échangeuse d'anions (monoQ) (stratégie expérimentale 2).....	334
2.5	Dénaturation et protéolyse sur billes.....	334
2.6	Enrichissement d'hétéroconjugués peptide-oligonucléotide .....	335
2.7	Hydrolyse de l'oligonucléotide .....	335
3.	Dessalage/Concentration.....	335
4.	Spectrométrie de masse.....	336
4.1	Spectrométrie de masse MALDI.....	336
4.2	Spectrométrie de masse à source électrospray.....	336
4.2.1	nanoUltraHPLC-MS .....	336
4.2.2	Trappe ionique HCT Ultra PTM Discovery .....	337
4.2.2.1	Couplage nanoLC-MS .....	337
4.2.2.2	Infusion directe .....	337
4.2.3	Maxis UHR-Q-TOF 4GHz.....	338
4.2.3.1	Couplage nanoLC-MS .....	338
4.2.3.2	Infusion directe .....	338
4.3	Analyse bioinformatique des données .....	339

**Bibliographie.....341**

# Introduction

# Section 1. Les modifications de protéines en médecine

Les protéines sont des composants essentiels de tous les organismes vivants. Impliquées dans la plupart des processus biologiques, elles peuvent être des cibles thérapeutiques ou des médicaments potentiels si leur fonction est connue.

## 1. Les relations structure-fonction des protéines

Au sein de la cellule, les protéines agissent en formant des complexes (permanents ou transitoires) avec différents partenaires (protéines, acides nucléiques, lipides, glycanes, ou petites molécules). Dans leur état natif, la plupart des protéines présentent un repliement spatial très organisé. Cette structure tridimensionnelle donne l'agencement des sites actifs, des domaines d'interaction et des lieux de fixation des ligands et donc renseigne sur leur mode de fonctionnement. L'établissement de la relation structure/fonction des protéines permet la compréhension de leur rôle dans les processus biologiques. C'est pourquoi toute information structurale (au niveau tertiaire ou quaternaire) peut apporter des précisions sur la fonction des protéines au sein de la cellule. Notons également que certaines protéines, appelées protéines intrinsèquement désordonnées, adoptent des conformations aléatoires et n'ont pas de structure tridimensionnelle stable dans des conditions physiologiques *in vitro*. Dans la cellule, cette plasticité accrue leur permet notamment d'interagir avec un grand nombre de protéines et d'être impliquées dans des fonctions biologiques essentielles telles que la reconnaissance et l'assemblage moléculaire (Uversky et al., 2008). De nombreuses protéines possèdent à la fois des domaines repliés mais également des régions intrinsèquement désordonnées qui sont modifiables et importantes pour leur fonction.

## 2. Les modifications de protéines

### 2.1 Définition

Une modification de protéine correspond à un changement de sa structure chimique, y compris de sa structure primaire, par addition, soustraction ou remplacement d'au moins une liaison covalente. Le produit de la modification doit encore pouvoir être considéré comme une protéine ou au minimum comme un peptide bioactif.

La protéine est généralement modifiée après synthèse de la chaîne peptidique, ce qui donne le terme de modification post-traductionnelle. Ce terme est également couramment employé pour décrire des modifications qui ont lieu en cours de synthèse, et qui sont donc, au sens strict, co-traductionnelles.

## 2.2 Rôles et conséquences des modifications de protéines

Les modifications de protéines changent les propriétés physico-chimiques des protéines par le biais de réactions chimiques et biochimiques qui modulent la fonction des protéines. Ainsi, les modifications peuvent activer ou inhiber les protéines, les stabiliser, entraîner leur dégradation, changer leurs partenaires d'interactions ou encore leur localisation dans la cellule. Le plus souvent, une modification provoque un changement de la structure de la protéine en modifiant la charge et/ou l'hydrophobie du résidu cible. Dans le cas d'addition d'un groupement ionisé sur un résidu apolaire, les résidus d'acides aminés chargés négativement ou positivement, voisins du résidu modifié, seront repoussés ou attirés pour aboutir à la formation d'une région hydrophile et polaire. Des changements conformationnels peuvent ainsi être induits et modifier l'agencement de sites actifs, de domaines d'interaction ou de sites de fixation des ligands.

## 2.3 Les types de modifications de protéines

Il existe deux grands types de modifications de protéines selon qu'elles proviennent d'un fonctionnement endogène de la cellule ou qu'elles sont la conséquence d'une exposition à des rayonnements ou à des agents chimiques ou xénobiotiques, d'origine accidentelle ou due à une intervention humaine. Ainsi, on distinguera les modifications post-traductionnelles et les modifications induites.

Les modifications post-traductionnelles (PTMs) ont lieu dans un contexte biologique. Elles sont nécessaires au fonctionnement même de la cellule, comme par exemple pour les voies de signalisation, ou pour son métabolisme. Les PTM sont finement régulées et font intervenir des molécules naturelles.

Les modifications induites sont en général la conséquence d'une intervention humaine, intentionnelle ou non, ou d'une exposition à des agents environnementaux externes tels que les rayonnements, les agents infectieux, la pollution environnementale, le mode de vie (tabac, alcool), le stress, l'alimentation et les médicaments. Les modifications induites peuvent être générées *in vivo* ou *in vitro*. *In vivo*, elles ne sont pas nécessaires au fonctionnement d'une cellule. *In vitro*, elles sont utilisées dans le cadre de projets de recherche pour mieux comprendre la fonction des protéines *in vivo*.

## 2.4 Les modifications post-traductionnelles

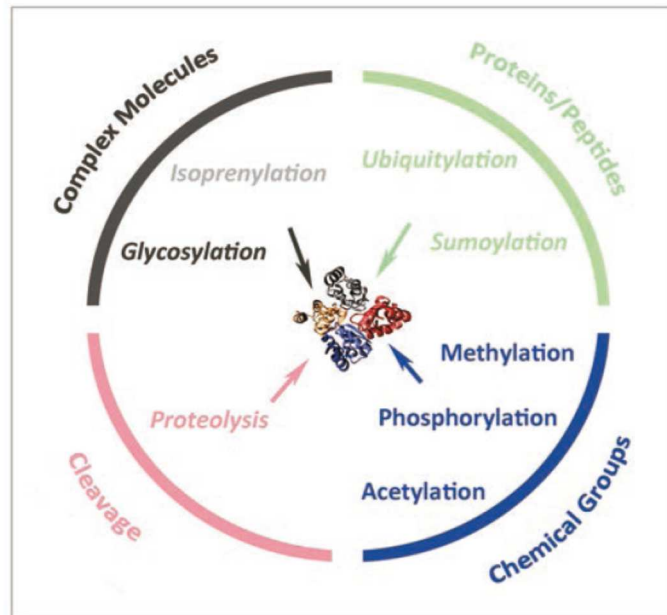
Depuis le séquençage du génome humain en 2000, il apparaît que l'Homme possède un nombre de gènes assez modeste (20687 d'après Pennisi, 2012), alors que le nombre total de protéines est estimé à plus d'un million (Nørregaard Jensen, 2004). Cette différence montre qu'un gène code pour de multiples protéines. Récemment, le terme « protéoforme » a été proposé pour définir les différentes formes moléculaires d'une protéine codée par un seul gène. L'ensemble de ces formes résulte de variations génétiques (variations alléliques, recombinaison génomique), transcriptionnelles (utilisation de promoteurs alternatifs, épissage alternatif de l'ARN transcrit), de sites d'initiation de la traduction alternatifs ou de modifications post-traductionnelles des protéines (Smith *et al.*, 2013).

### 2.4.1 Généralités

Les modifications post-traductionnelles (PTMs) sont des mécanismes essentiels utilisés par les cellules eucaryotes afin de diversifier les fonctions de leurs protéines, de leur donner leur forme mature active et de coordonner dynamiquement leurs réseaux de signalisation. Des défauts de PTM sont liés à de nombreux troubles du développement et à des maladies humaines, soulignant l'importance des PTMs dans le maintien d'un état cellulaire normal.

Les PTMs modifient la structure et les propriétés des protéines par le biais de réactions biochimiques. Ces modifications peuvent être divisées en trois classes principales: la première est l'addition covalente d'une ou plusieurs molécules sur une protéine, la seconde est le clivage protéolytique d'une ou plusieurs liaisons peptidiques (Mann and Jensen, 2003) et la troisième est la formation de liaisons intra- ou inter-moléculaires. Dans la première catégorie, des groupements tels que les phosphoryles ( $-\text{PO}_3^{2-}$ ), acétyles ou glycosyles sont attachés covalamment aux chaînes latérales des acides aminés *via* l'activité catalytique d'enzymes spécifiques. Ces groupements sont de complexité variable ; on distinguera ainsi les petits groupements chimiques (par exemple la phosphorylation et l'acétylation), les molécules complexes (par exemple la glycosylation) et les protéines entières (par exemple l'ubiquitination) (Figure 1). Dans la deuxième catégorie, des enzymes appelées protéases catalysent la coupure des liaisons peptidiques.

Figure 1. Représentation des PTMs les plus étudiées (Wang *et al.*, 2013)



Enfin, la formation de liaisons covalente intramoléculaire la plus fréquente est l'oxydation de cystéines en cystines avec formation de ponts disulfure, mais il est également possible de ponter des chaînes latérales avec l'extrémité N- ou C-terminale d'une protéine ou d'un peptide pour former des structures en lasso (Bayro *et al.*, 2003). Des liaisons intermoléculaires comme le pontage de fibrilles de collagène *via* des hydroxyprolines existent également (Bailey *et al.*, 1974).

Plus de 300 modifications de protéines, générées par addition d'un groupement sur un ou plusieurs types d'acides aminés, par clivage protéolytique, ou par liaison intra- ou inter-moléculaire, sont décrites à ce jour (Nørregaard Jensen, 2004). Ces modifications sont répertoriées dans des bases de données telles que Unimod [a]. La base de données dbPTM [b] recense des données statistiques sur le nombre de sites expérimentaux et putatifs des principales PTMs (Lu *et al.*, 2013). Parmi, les PTMs les plus communes, on peut citer l'élimination de la méthionine en position N-terminale des protéines par une méthionine aminopeptidase, la formation des ponts disulfures, la glycosylation, la phosphorylation, l'ubiquitination ou la lipidation (Tableau 1).



**Tableau 1. Modifications post-traductionnelles les plus communes et les plus importantes<sup>#</sup>.**

Type de modification post-traductionnelle	Formule chimique ajoutée et perdue sur protéine	Δ Masse monoisotopique (Da)	Stabilité <sup>a</sup>	Fonction et commentaires
<b>Phosphorylation</b>				
pTyr	+H <sub>3</sub> PO <sub>4</sub> -H <sub>2</sub> O	+79.9663	+++	Transduction du signal, activation/inactivation de l'activité des enzymes, modulation des interactions protéine-protéine et protéine-ligand.
pSer, pThr	+H <sub>3</sub> PO <sub>4</sub> -H <sub>2</sub> O	+79.9663	+/>++	
<b>Acétylation</b> (résidu N-terminal, Lys)	+CH <sub>3</sub> CO-H	+42.0106	+++	Stabilité protéique, protection de l'extrémité N-terminale. Régulation des interactions protéine-ADN (histones).
<b>Méthylation</b> (Lys, Arg)	+CH <sub>3</sub> -H	+14.0157	+++	Régulation de l'expression des gènes, des interactions protéine-protéine et protéine-acide nucléique.
<b>Acylation</b> , modification par des acides gras				Localisation cellulaire, signalisation, ancrage membranaire, interactions protéine-protéine.
Farnésyl (Cys)	+C <sub>15</sub> H <sub>24</sub>	+204.1878	+++	
Myristoyl (Cys, Lys, Gly)	+C <sub>14</sub> H <sub>26</sub> O	+210.1984	+++	
Palmitoyl (Cys, Lys, Ser, Thr, N-ter) etc.	+C <sub>16</sub> H <sub>30</sub> O	+238.2297	+/>++	
<b>Glycosylation</b>				Sécrétion des protéines, reconnaissance et signalisation cellulaire, stabilité des protéines.
N-glycosylation (Asn)		>800	+/>++	
O-glycosylation (Ser, Thr)	+C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N	+203.0793, >800	+/>++	
<b>Ancre GPI</b> (glycosylated phosphatidylinositol)		>1000	++	Ancrage membranaire pour des enzymes et récepteurs extérieurs à la cellule.
<b>Hydroxyproline (Pro)</b>	+OH-H	+15.9949	+++	Stabilité des protéines et interactions protéine-protéine.
<b>Sulfonation (Tyr)</b>	+SO <sub>3</sub>	+79.9568	+	Modulation des interactions protéine-protéine et récepteur-ligand.
<b>Pont disulfure (Cys)</b>	-H <sub>2</sub>	-2.0157	++	Réticulation intra et intermoléculaire, stabilité des protéines.
<b>Déamidation</b> (Asn, Gln)	-NH <sub>2</sub> +OH	+0.9909	+++	Régulation des interactions protéine-ligand et protéine-protéine. Associée au vieillissement.
<b>Acide pyroglutamique</b>	-NH <sub>3</sub>	-17.0265	+++	Stabilité des protéines, blocage de l'extrémité N-terminale.
<b>Oxydation</b> (Met)	+O	+15.9949	+++	Signalisation. Activité des protéines.
(Trp)	+O-C,	+3.9949,	+++	
	+O,	+15.9949,	+++	
	+O <sub>2</sub>	+31.9898	+++	
<b>Ubiquitination (Lys)</b>		>1000	+/>++	Signal de destruction. Après digestion trypsique, le site d'ubiquitination est observé avec un incrément de masse de +114 Da (Gly-Gly).
<b>Nitration des tyrosines</b>	+NO <sub>2</sub> -H	+44.9851	+/>++	Domages oxydatifs au cours de l'inflammation.

<sup>#</sup>d'après Mann et Jensen (2003)

<sup>a</sup>Stabilité : + labile en spectrométrie de masse, ++ modérément stable, +++ stable.

Une modification post-traductionnelle peut se produire à n'importe quelle étape du «cycle de vie» d'une protéine. Par exemple, de nombreuses protéines sont modifiées peu après la traduction afin de leur permettre d'être correctement repliées ou d'être dirigées vers d'autres compartiments cellulaires tels que le noyau ou la membrane plasmique. D'autres modifications se produisent une fois que le repliement et la localisation sont terminés, pour activer ou inactiver l'activité catalytique de la protéine.

Les PTMs peuvent être réversibles ou irréversibles. Des modifications permanentes, telles que le clivage protéolytique, la formation de pont disulfure ou la glycosylation, sont essentielles pour une maturation structurelle et fonctionnelle appropriée des protéines nouvellement synthétisées. D'autres modifications covalentes sont transitoires et jouent des rôles importants dans les voies de signalisation et la régulation des protéines (Yang, 2005).

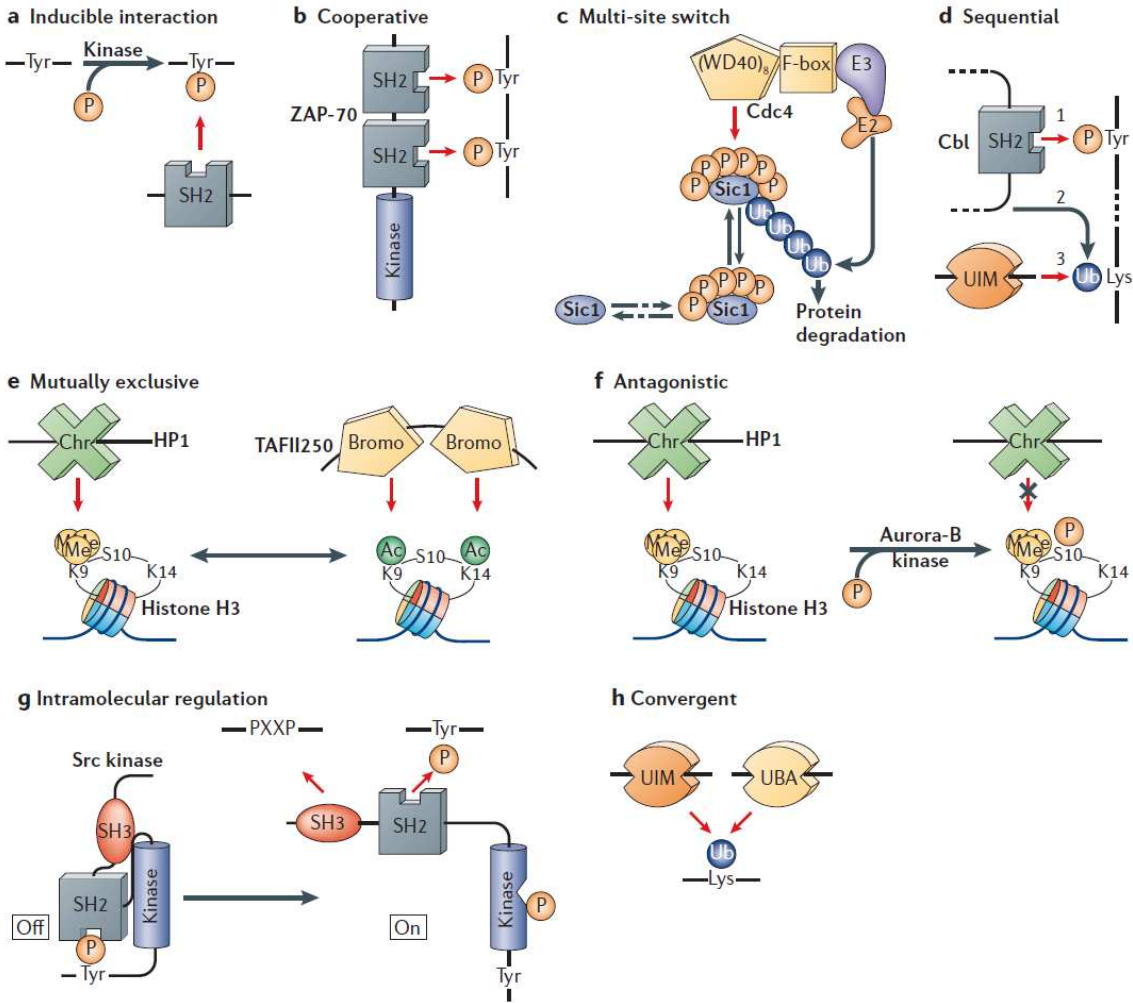
Une même protéine peut être modifiée sur plusieurs sites et par différentes PTMs. Cela permet d'obtenir un grand nombre de combinaisons et d'accroître la diversité du protéome.

Le niveau d'abondance des PTMs est finement ajusté par des enzymes qui ajoutent ou suppriment les modifications. En agissant de manière coordonnée, les PTMs jouent un rôle clé dans la fonction des protéines, car elles régulent l'activité enzymatique, les interactions avec d'autres molécules (protéines, acides nucléiques, lipides, cofacteurs) et la localisation dans la cellule (Beltrao *et al.*, 2013). Pour moduler la fonction des protéines, les PTMs peuvent agir de plusieurs manières. L'addition d'une molécule sur une protéine peut :

- provoquer un changement de la structure tridimensionnelle de la protéine et ainsi modifier l'agencement de son site actif, de domaines d'interaction et de sites de fixation des ligands.
- créer un site de liaison pour le recrutement d'effecteurs contenant des domaines d'interaction qui reconnaissent le résidu modifié (Seet *et al.*, 2006; Deribe *et al.*, 2010). Plusieurs mécanismes d'interaction dépendants des PTMs sont décrits (Figure 2).

Les développements récents en spectrométrie de masse et en préparation d'échantillon, tels que l'enrichissement, permettent d'identifier les PTMs à haut débit (Choudhary and Mann, 2010). Les PTMs les plus étudiées sont la phosphorylation, l'acétylation, la méthylation, l'ubiquitination et la glycosylation. Par souci de brièveté, seule une partie de ces PTMs sera détaillée ici.

**Figure 2. Exemples de mécanismes de régulation par interaction de PTMs avec des domaines de reconnaissance (d'après (Seet *et al.*, 2006)).**



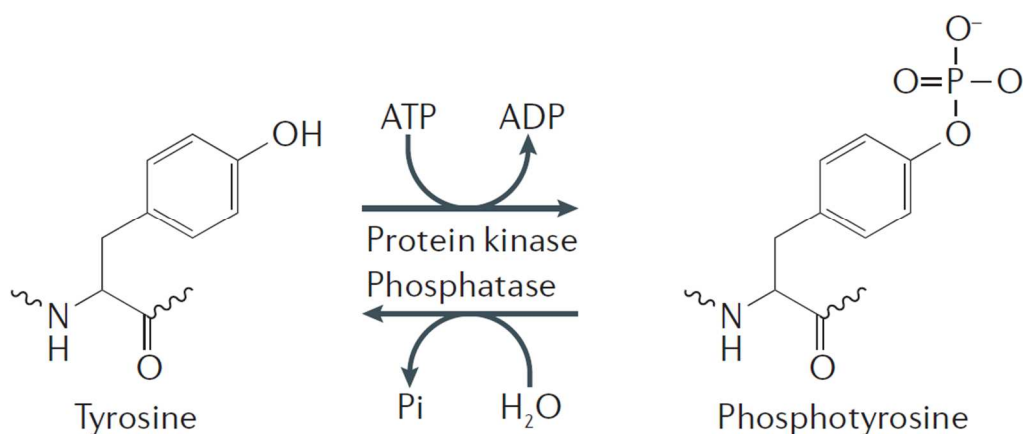
## 2.4.2 La phosphorylation

La phosphorylation des protéines est l'une des modifications post-traductionnelles les plus importantes et les plus étudiées. Elle joue un rôle essentiel dans la régulation de nombreux processus cellulaires, tels que le cycle cellulaire, la croissance, l'apoptose et les voies de transduction du signal.

### 2.4.2.1 Formation, destruction et effets moléculaires et cellulaires

Chez les eucaryotes, la phosphorylation cible la chaîne latérale des acides aminés sérine, thréonine et tyrosine. C'est une modification réversible médiée par des kinases et des phosphatases, qui phosphorylent et déphosphorylent les résidus cibles, respectivement (Figure 3). Ces deux familles d'enzymes contrôlent précisément l'état de phosphorylation des protéines. On estime qu'environ 1000 kinases et 500 phosphatases sont codées par le génome humain, soit environ 2% du génome (Venter *et al.*, 2001), et que presque un tiers de toutes les protéines dans la cellule est phosphorylé à un moment donné (Cohen, 2000).

**Figure 3. Réaction de phosphorylation des protéines**



Dans la plupart des cas, l'addition d'un groupement phosphoryle provoque des changements conformationnels dans la protéine phosphorylée. Ces changements peuvent affecter l'activité enzymatique et/ou les interactions protéine-protéine (Johnson and Lewis, 2001). Ainsi, une protéine peut être activée (Leonard *et al.*, 2013) ou inactivée (Burke *et al.*, 2010) par phosphorylation, ou même passer d'une voie de signalisation à une autre. En plus de modifier les propriétés intrinsèques des protéines par régulation allostérique, un rôle important de la phosphorylation est de créer des sites pour l'interaction avec d'autres protéines (Figure 2). Ces dernières contiennent des domaines qui reconnaissent les phospho-motifs et s'y lient. Ces domaines présentent une spécificité pour les

différents acides aminés phosphorylés. Par exemple, les domaines SH2 (Src Homology 2) et PTB (PhosphoTyrosine Binding) montrent une spécificité pour la phosphotyrosine (pY) (Yaffe, 2002) tandis que la phosphothréonine (pT) est reconnue par les domaines FHA (ForkHead-Associated) (Li *et al.*, 2000). La capacité des phosphoprotéines à recruter d'autres protéines est capitale pour la transduction du signal.

Le caractère réversible de la phosphorylation des protéines rend ce type de modification idéal pour la transduction de signal. En effet, elle permet aux cellules de répondre rapidement à des stimuli extracellulaires ou intracellulaires. Les cascades de transduction de signal sont caractérisées par une ou plusieurs protéine(s) senseur(s) qui détecte(nt) physiquement le signal soit par liaison à un ligand, un clivage ou une autre réponse, et qui le relaie(nt) à des seconds messagers et des enzymes de signalisation. Dans le cas de la phosphorylation, ces récepteurs activent des kinases en aval, qui ensuite phosphorylent et activent leurs substrats, y compris d'autres kinases, jusqu'à ce que la réponse spécifique soit atteinte.

#### 2.4.2.2 Les méthodes de détection

A cause du rôle central de la phosphorylation dans la régulation des processus cellulaires chez les eucaryotes, les efforts de développement de méthodes pour étudier les phosphorylations des protéines ont été nombreux. Les approches actuelles pour caractériser les phosphorylations comprennent l'immunodétection, le test d'activité kinase et la localisation des phosphorylations par analyse de phosphoprotéines ou phosphopeptides en spectrométrie de masse. L'objectif à terme vise à caractériser entièrement le réseau de régulation hautement sophistiqué dans lequel est impliquée la phosphorylation. Cela implique l'identification de toutes les phosphoprotéines ainsi que leur(s) site(s) modifié(s) et la quantification de cette modification. L'ensemble de ces travaux constitue la phosphoprotéomique et devrait accroître notre connaissance de la régulation de différents processus biologiques.

#### 2.4.3 La N-acétylation

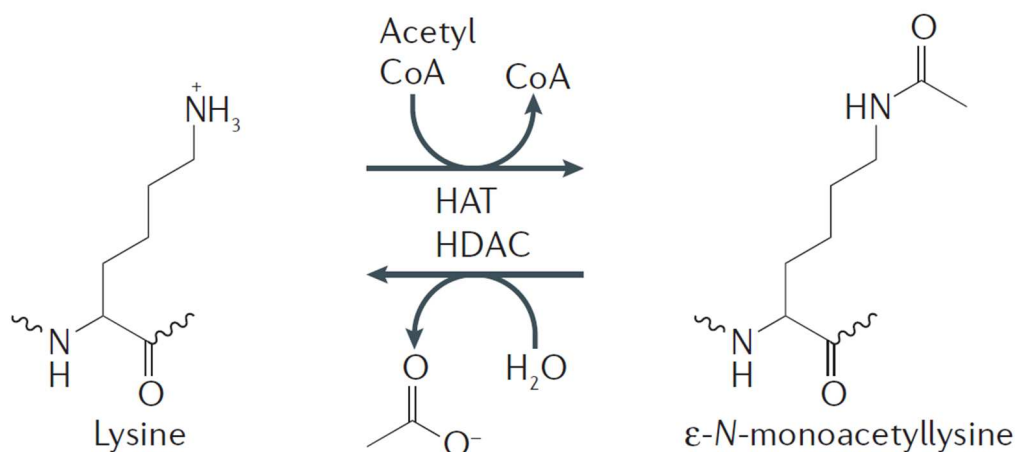
La N-acétylation consiste à transférer un groupement acétyle sur l'azote N-terminal (acétylation N-terminale) ou sur l'azote d'une lysine (acétylation de lysine).

L'acétylation N-terminale nécessite le clivage préalable de la méthionine N-terminale par la méthionine aminopeptidase (MAP). Ensuite, l'enzyme N-acétyltransférase (NAT) ajoute un groupement acétyle en utilisant l'acétyl-CoA. Ce type d'acétylation est irréversible et se produit au cours de la synthèse de la protéine. Bien que 80 à 90 % de protéines eucaryotes soient acétylées de

cette manière, la signification biologique exacte n'est pas encore claire. Des données récentes suggèrent un rôle dans la stabilité des protéines et les voies de sécrétions (Arnesen, 2011).

L'acétylation de la fonction  $\epsilon$  - NH<sub>2</sub> de la lysine (appelée acétylation de la lysine) sur les extrémités N-terminales des histones est une voie courante de régulation de la transcription des gènes. L'acétylation des histones est un événement réversible qui réduit la condensation chromosomique pour promouvoir la transcription. L'acétylation de ces lysines est régulée par des facteurs de transcription qui contiennent une activité histone acétyl transférase (HAT) et des histones déacétylases (HDAC) (Figure 4). En combinaison avec des méthylations, ces acétylations constituent le code histone à la base des phénomènes épigénétiques. Certains auteurs ont proposé la mesure systématique des modifications d'histones par spectrométrie de masse dans l'optique du "déchiffrement" du code histone (Young *et al.*, 2009).

**Figure 4. Réaction d'acétylation des protéines**



L'acétylation de protéines peut être détectée par immunoprécipitation en utilisant des anticorps spécifiques des lysines acétylées ou par spectrométrie de masse en recherchant des incréments de masse de +42 Da par acétylation. Avec 3600 sites d'acétylation identifiés sur 1750 protéines, Choudhary *et al.* ont ainsi montré que l'acétylation des lysines est une modification répandue impliquée dans la régulation de plusieurs processus cellulaires tels que le cycle cellulaire ou le transport nucléaire (Choudhary *et al.*, 2009). L'acétylation des lysines semble également avoir un rôle dans la régulation des voies métaboliques (Zhao *et al.*, 2010).

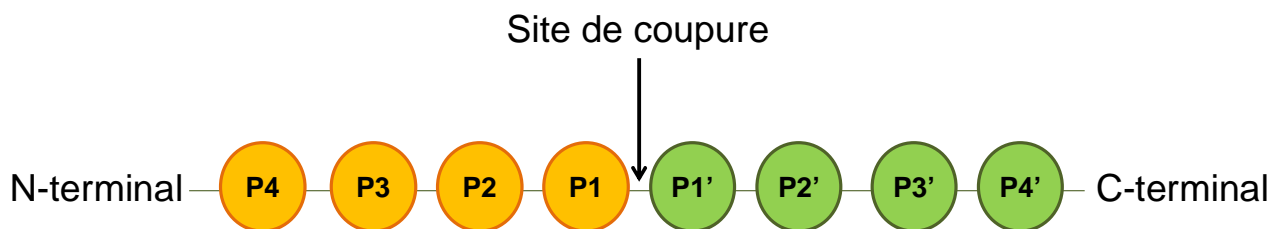
#### 2.4.4 La protéolyse

Les protéinases (appelées couramment protéases) sont les enzymes qui catalysent l'hydrolyse des liaisons peptidiques de protéines cibles et peuvent ainsi modifier leur activité. La protéolyse est essentielle pour éliminer les sous-unités de protéines non assemblées, les protéines mal repliées et pour maintenir les concentrations en protéine à des concentrations homéostatiques. Les protéases jouent également un rôle biosynthétique tel que le clivage du peptide signal de protéines et l'activation des zymogènes, ou sont associées à un processus de dégradation (protéasome). Notons que d'après la définition d'une modification de protéine décrite au paragraphe 2.1, les clivages enzymatiques impliqués dans la dégradation des protéines ne sont pas considérés comme des modifications post-traductionnelles de protéines dans la mesure où le produit de la modification n'est plus une protéine. La protéolyse joue de nombreux rôles dans les processus cellulaires, tels que la formation d'antigènes, la mort cellulaire programmée, la sécrétion de protéines à la surface cellulaire, les voies de signalisation ou encore les processus de développement comme l'action protéolytique des enzymes de spermatozoïdes pour pénétrer dans un ovocyte (Thornberry and Lazebnik, 1998; Ye and Fortini, 2000). Au niveau extracellulaire, une cascade de protéinases est responsable des phénomènes de coagulation.

La protéolyse est une réaction thermodynamiquement favorable et irréversible. Par conséquent, afin d'éviter une protéolyse non contrôlée, l'activité des protéases est strictement régulée par des mécanismes de contrôle temporel et/ou spatial (Bai and Pfaff, 2011), incluant la compartimentation (par exemple, les granules de sécrétion, les lysosomes) (Ciechanover, 2005). Un défaut de régulation de l'activité protéolytique par les inhibiteurs naturels de protéinases conduit à la mort cellulaire par apoptose au niveau intracellulaire et à une destruction des tissus au niveau extracellulaire.

Certaines protéases montrent une grande sélectivité avec une spécificité leur permettant de cibler une protéine unique, alors que d'autres sont de spécificité large et peuvent dégrader plusieurs substrats. Les protéases peuvent être classées en fonction de leur site d'action avec un clivage de l'extrémité amino ou carboxy-terminale du substrat pour les exoprotéases, ou un clivage à l'intérieur de la séquence pour les endoprotéases. On les classe également par leur mécanisme catalytique en fonction de l'identité du résidu catalytique (par exemple, protéinase à sérine). Il existe à ce jour neuf catégories catalytiques répertoriées dans la base de données MEROPS [c] (Rawlings *et al.*, 2011). Cette base de données propose notamment des matrices de spécificité qui correspondent aux résidus P4 à P4' de la poche de liaison du substrat (Figure 5) et qui peuvent être analysées en détail lorsque l'on s'intéresse à la spécificité des protéases.

Figure 5. Nomenclature des résidus de la poche de liaison du substrat (peptide, protéine) de part et d'autre du site de coupure par une protéinase. D'après Schechter and Berger (1967).



A ce jour, plus de 90% des protéases connues peuvent être classées dans une des quatre catégories catalytiques suivante (Puente *et al.*, 2003):

- Les protéases à sérines
- Les protéases à cystéines
- Les protéases à acide aspartique
- Les métalloprotéases à zinc

## 2.5 Les modifications induites

Les propriétés physico-chimiques des protéines peuvent également être modulées non pas lors de processus physiologiques, mais par des réactions chimiques induites par une exposition à des agents environnementaux externes ou à l'action de l'Homme. Ces modifications peuvent avoir lieu *in vivo* ou *in vitro*.

*In vivo*, elles sont la conséquence de ce que nous mangeons, de l'air que nous respirons, des médicaments que nous utilisons ou encore de nos choix de vie, comme le tabagisme et l'exercice physique. Ces modifications induites *in vivo* se traduisent souvent par la fixation covalente d'une molécule sur les protéines.

*In vitro*, les modifications induites peuvent être intentionnelles ou accidentelles. Dans le cadre de projets de recherche, le scientifique peut être amené à modifier les protéines *in vitro* afin de les étudier et mieux comprendre leur fonction *in vivo*. Ces modifications peuvent être des protéolyses volontaires, des PTMs telles que la phosphorylation ou la glycosylation, l'action d'agents de réticulation chimiques dans le cadre d'études structurales ou encore des ligands dans le cadre d'études de complexes protéine-ligand. Toutefois, au cours de ces expériences ou même pendant le stockage d'un échantillon, des modifications accidentelles telles que l'oxydation, la protéolyse involontaire ou la déamidation peuvent également avoir lieu. Ces artéfacts, selon leur ampleur et l'objectif de l'expérience, peuvent devenir un vrai problème et doivent être évités.



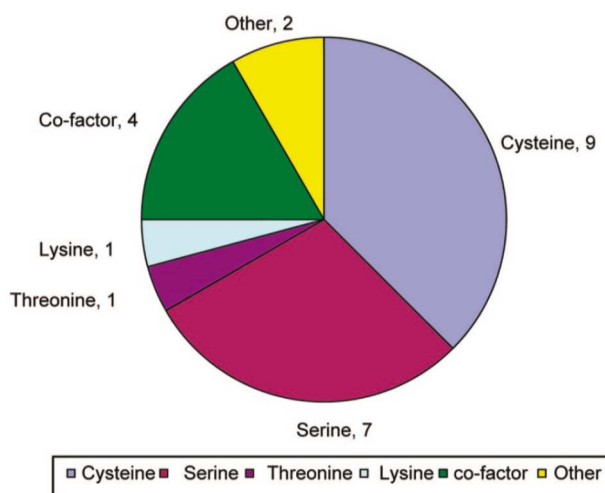
## 2.5.1 Modification induite *in vivo*, exemple de la fixation covalente d'un ligand

Par définition, en biochimie, un ligand est une molécule capable de se lier à une protéine. Pour éviter toute confusion, nous ne parlerons pas ici de ligand au sens d'un métal capable de former des liaisons de coordination avec un groupement prosthétique. Classiquement, le terme de ligand se réfère à une petite molécule capable de se placer au niveau d'un site spécifique d'une protéine cible afin de moduler son activité et/ou sa conformation tridimensionnelle. Ce site peut être très structuré et former une poche (modèle clé-serrure) ou être plutôt plan et correspondre à une zone d'interaction protéine-protéine.

La grande majorité des ligands endogènes interagissent avec les protéines de manière non-covalente et réversible, mettant en jeu des liaisons telles que les liaisons hydrogènes, les forces de van der Waals et les liaisons ioniques. La liaison covalente d'un ligand sur une protéine est plus rare dans les conditions physiologiques et concerne plutôt les ligands exogènes. Ces derniers, tels que les médicaments et d'autres xénobiotiques, peuvent être d'origine naturelle (Gersch *et al.*, 2012), synthétiques ou semi-synthétiques.

Une liaison covalente est une liaison chimique qui résulte de la mise en commun d'une ou plusieurs paires d'électrons entre deux atomes. La fixation covalente d'un ligand sur une protéine implique le plus souvent une réaction entre un ligand électrophile et un résidu nucléophile de la protéine (Rubino *et al.*, 2009). Les résidus les plus nucléophiles sont la cystéine, la sérine, la thréonine et la lysine (Figure 6) mais d'autres résidus sont également capables de réagir avec un électrophile, comme la tyrosine, l'histidine, l'asparagine, l'arginine et les acides aspartiques et glutamiques (Rubino *et al.*, 2009; Gersch *et al.*, 2012).

**Figure 6. Distribution des résidus modifiés de manière covalente par des médicaments (Potashman and Duggan, 2009).**



Pour former une liaison covalente entre deux molécules en milieu biologique, plusieurs mécanismes réactionnels sont possibles, notamment l'addition de Michael. L'addition de Michael implique l'addition d'un nucléophile (composé riche en électrons), également appelé donneur de Michael, sur un électrophile (composé déficitaire en électrons), l'accepteur de Michael. Le plus souvent, le groupement électrophile est un système conjugué de type carbonyle  $\alpha,\beta$ -insaturé (cétone, aldéhyde) ou ester  $\alpha,\beta$ -insaturé. Cette réaction permet la création de liaisons carbone-carbone, carbone-soufre, carbone-oxygène, carbone-azote ou encore carbone-phosphore (Mather *et al.*, 2006). Le mécanisme de réaction est le suivant : dans le cas d'une cétone  $\alpha,\beta$ -insaturée, la délocalisation des électrons  $\pi$  fait apparaître une charge partielle négative sur l'atome d'oxygène du carbonyle et une charge partielle positive sur le carbone  $\beta$  (Figure 7). Ce carbone  $\beta$  est électrophile et peut subir l'attaque d'un nucléophile (Figure 8).

Figure 7. Formation d'un carbone  $\beta$  électrophile d'une cétone  $\alpha,\beta$ -insaturée.

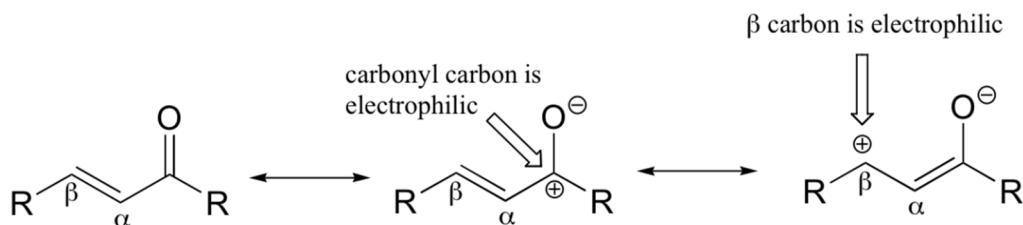
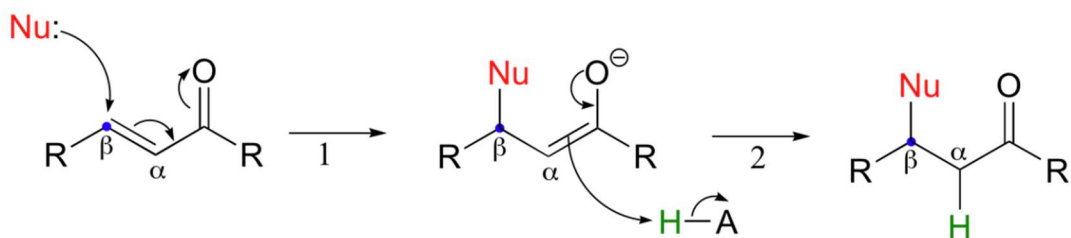
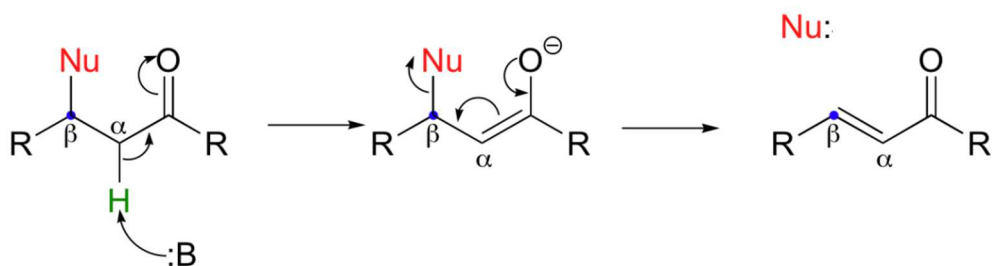


Figure 8. Addition de Michael sur une cétone  $\alpha,\beta$ -insaturée.



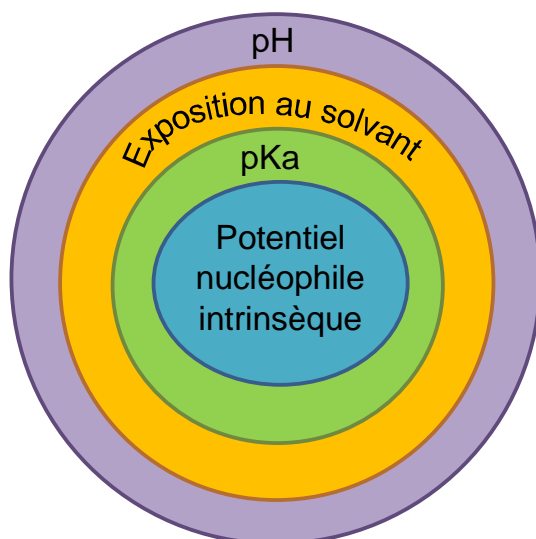
L'addition de Michael peut être réversible (Figure 9), c'est ce que l'on appelle la réaction rétro-Michael ou encore la bêta-élimination. Cela concerne notamment les cystéines. La réversibilité de la liaison dépend de la nature électronique du donneur et de l'accepteur, du pH, de la température et de l'environnement local du résidu dans la protéine (Wilson *et al.*, 2013).

Figure 9. Réaction de rétro-Michael



La réactivité du groupement électrophile peut varier selon son environnement. De même, d'une protéine à l'autre ou bien à l'intérieur d'une même protéine, la réactivité d'un résidu nucléophile peut varier. La réactivité d'un résidu nucléophile est une composante de quatre paramètres (Figure 10) qui sont le potentiel nucléophile intrinsèque du résidu, le pKa du résidu dans la protéine, l'exposition au solvant (résidu exposé ou enfouit), et le pH local.

Figure 10. Les paramètres influençant la réactivité d'un résidu nucléophile



Cette Figure 10 illustre bien que le potentiel nucléophile d'un résidu ne pourra s'exprimer que si les conditions de pKa, d'accessibilité au solvant et de pH sont réunies. En d'autres termes, selon l'état de protonation d'un résidu nucléophile, ce dernier sera plus ou moins réactif. En effet, d'après la relation d'Henderson-Hasselbalch, on voit que des variations de pKa et/ou de pH vont influencer la proportion d'espèces protonées.

$$pH = pKa + \log\left(\frac{[A^-]}{[AH]}\right) \quad (1) \text{ relation d'Henderson-Hasselbalch}$$

avec  $[A^-]$  et  $[AH]$ , les concentrations respectives en base conjuguée de l'acide et en acide.

Ainsi, une diminution de pKa et/ou une augmentation de pH vont accroître la proportion des espèces déprotonées qui sont plus réactives que les espèces protonées du point de vue de la nucléophilicité. Dans le cas des cystéines par exemple, le pKa détermine si le résidu est plus susceptible d'exister sous forme d'un thiol (SH) ou un anion thiolate (S<sup>-</sup>) qui est plus réactif. Par conséquent, les résidus cystéines avec un faible pKa ont tendance à être plus réactifs que ceux qui ont une valeur de pKa supérieure au pH environnant et qui restent protonés et extrêmement moins réactifs. Les variations de valeurs de pKa sont liées à l'environnement structural local du résidu. Dans le cas de la cystéine, la présence au voisinage de résidus aromatiques ou chargés positivement induit une diminution du pKa en stabilisant l'anion thiolate *via* des interactions polaires ou électrostatiques (Ferrer-Sueta *et al.*, 2011). Cette stabilisation du thiolate permet ainsi d'augmenter la proportion de l'espèce réactive de la cystéine.

## 2.5.2 Modification induite *in vitro*, exemple de l'oxydation des protéines

L'oxydation est une modification particulière car elle est à la fois une PTM, impliquée dans le stress oxydatif et sa détection par les mécanismes cellulaires de réparation, et une modification induite générée *in vitro*. L'oxydation de protéines *in vitro* peut être intentionnelle ou accidentelle. Dans le cadre de projets de recherche, *via* des expériences d'empreinte protéique (footprinting), l'oxydation des protéines est utilisée pour étudier la structure des protéines (les zones exposées au solvant ou enfouies) (Wang and Chance, 2011) ou les surfaces d'interactions avec un partenaire ou dans un dimère (Charvatova *et al.*, 2008). Toutefois, l'oxydation de protéines *in vitro* peut également être accidentelle et être générée par le protocole expérimental utilisé. Bien trop souvent négligé, cet artefact expérimental augmente l'hétérogénéité de l'échantillon, contribue à la diminution du signal d'intérêt et peut même diminuer la stabilité et l'activité biologique des protéines, comme dans le cas des anticorps thérapeutiques par exemple (Pan *et al.*, 2009).

### 2.5.2.1 Les espèces réactives de l'oxygène (ROS)

Bien que l'oxygène et l'eau soient essentiels à la vie, ils peuvent devenir hautement réactifs et potentiellement néfastes lorsqu'ils sont convertis en radicaux libres, généralement connus sous le nom d' « espèces réactives de l'oxygène » (ROS). *In vitro*, les ROS peuvent être générés par radiolyse de l'eau, photoactivation (lumière UV ou visible) ou *via* des métaux (réaction de Fenton, voir ci-après). Les ROS sont composés d'espèces radicalaires, telles que l'anion superoxyde (O<sub>2</sub><sup>•-</sup>) ou le radical hydroxyle (OH<sup>•</sup>) et d'espèces non radicalaires, telles que le peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>). Le ROS le plus réactif est OH<sup>•</sup>. Son caractère très électrophile conduit à des dommages non

spécifiques immédiats au niveau des molécules environnantes riches en électrons. De plus, le peroxyde d'hydrogène H<sub>2</sub>O<sub>2</sub>, bien que moins réactif, peut facilement être réduit *via* des réactions de Fenton ou Haber-Weiss et constituer ainsi une source de OH<sup>•</sup> (Höhn *et al.*, 2013). La réaction de Fenton peut également être catalysée par des métaux autres que Fe<sup>2+</sup> tels que Cu<sup>2+</sup>, Mn<sup>2+</sup>, Ni<sup>2+</sup> ou Zn<sup>2+</sup>, bien que la réactivité de Zn<sup>2+</sup> soit si faible qu'il est généralement considéré comme biologiquement inerte.



### 2.5.2.2 L'oxydation des acides aminés

L'oxydation des protéines concerne des acides aminés bien déterminés. Les acides aminés soufrés sont les plus facilement oxydés mais les aromatiques ainsi que l'histidine peuvent l'être également (Tableau 2). L'ordre de réactivité des acides aminés est le suivant : Cys, Trp, Tyr, Met > Phe > His, Arg, cystine, Ile, Leu, Val, Pro. Les autres acides aminés sont nettement moins réactifs.

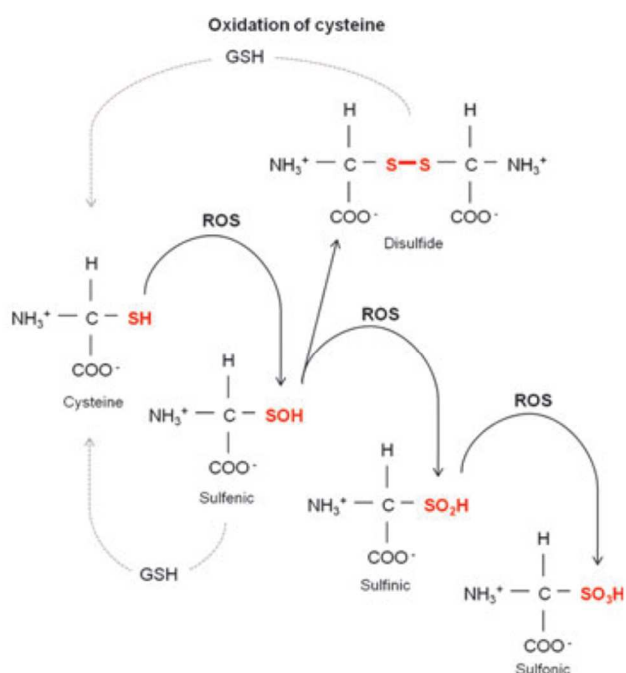
**Tableau 2. Delta masse des produits d'oxydation les plus courants générés par les radicaux hydroxyles.** D'après (Wang and Chance, 2011).

Acides aminés	Changement dans la formule brute	Changement de la masse monoisotopique (Da)	Type de dommage
Arg	+ O, - C, - 3N, - 5H	- 43.0534	déguanidination
Glu, Asp	- C, - O, - 2H	- 30.0106	Décarboxylation oxydative : -CH <sub>2</sub> -COO <sup>-</sup> → -C=O
Met	+ O, - S, - C, - 4H	- 32.0085	
His	+ O, - 2C, - N, - H	- 23.0160	
His	+ 2O, - 2C, - 2N, - 2H	- 22.0320	
Cys	+ O, - S	- 15.9772	
His	+ 2O, - C, - 2N, - 2H	- 10.0320	
Trp	+ O, - C	+ 3.9949	Formation de kynurénine
Lys, Arg, Leu, Ile, Val, Ala Pro, Glu, Gln,	+ O, - 2H	+ 13.9793	« carbonylation » : -CH <sub>2</sub> → -C=O
Cys, Met, Trp, Tyr, Phe, His, Ser, Thr, Leu, Ile, Val, Asp, Asn, Glu, Gln, Arg, Lys	+ O	+ 15.9949	Addition d'oxygène sous forme hydroxy (-OH) ou oxo (=O)
Met, Cys, Trp, Tyr, Phe	+ 2O	+ 31.9898	Formation de méthionine sulfone, Formation d'acide sulfinique, Formation de N-formylkynurénine
Cys	+ 3O	+ 47.9847	

La chaîne latérale de la cystéine subit une oxydation conduisant à la formation réversible d'un acide sulfénique intermédiaire (SOH), qui peut former un pont disulfure (S-S) avec une autre cystéine

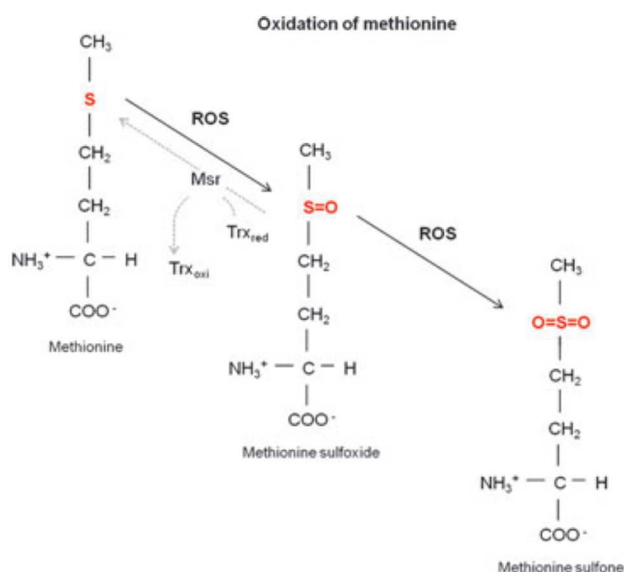
physiquement proche. Dans des conditions d'oxydations extrêmes, l'oxydation se poursuit et conduit à la formation irréversible d'acide sulfinique (SO<sub>2</sub>H) et d'acide sulfonique (SO<sub>3</sub>H) (Figure 11).

**Figure 11. Produits d'oxydation des cystéines (Corcoran and Cotter, 2013)**



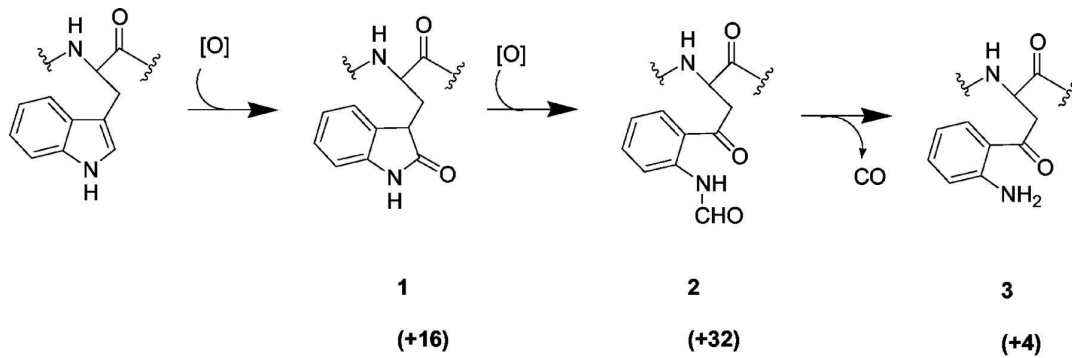
L'oxydation de la méthionine (Figure 12) conduit à la méthionine sulfoxyde (+ 16 Da). Cette oxydation est réversible *in vivo* via l'action de la méthionine sulfoxyde réductase (Msr) et de la thiorédoxine (Lee and Gladyshev, 2011) mais pas *in vitro*. Dans les conditions d'oxydation extrêmes, les ROS peuvent convertir la méthionine sulfoxyde en une forme hyper-oxydée : la méthionine sulfone.

**Figure 12. Produits d'oxydation des méthionines (Corcoran and Cotter, 2013)**



Les résidus aromatiques peuvent également être oxydés. L'oxydation de la tyrosine génère, entre autres composés, la di-tyrosine qui est considérée comme un marqueur de l'oxydation dans les cellules (Heinecke *et al.*, 1993). L'oxydation du tryptophane conduit successivement à l'hydroxytryptophane (+ 16 Da), à la *N*-formylkynurénine (+ 32 Da) qui est un produit de double oxydation et à la kynurénine (+ 4 Da) (Figure 13).

**Figure 13. Produits d'oxydation du tryptophane (Taylor *et al.*, 2003).**



Dans le cadre d'étude sur l'oxydation des protéines en tant que PTM, il est primordial de maîtriser toutes les étapes du protocole expérimental afin de ne pas introduire d'oxydation artificielle.

#### 2.5.2.3 Recommandations pour éviter l'oxydation accidentelle des protéines

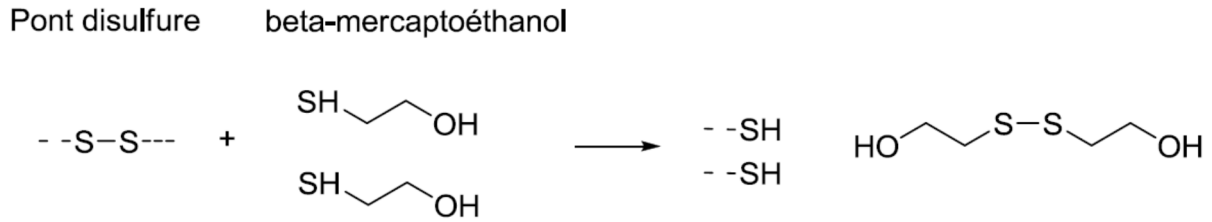
Les solutions pour éviter l'oxydation accidentelle des protéines consistent 1) à éliminer l'oxygène dissout par bullage des tampons à l'aide d'un gaz inerte de solubilité supérieure à celle de l'oxygène, tel que l'Argon; 2) à ajouter un antioxydant dans la solution de conservation de la protéine et/ou dans le milieu réactionnel; 3) à privilégier les temps courts de stockage et de réaction et 4) à vérifier que chaque étape du protocole expérimental n'induit pas d'oxydation, comme par exemple le gel SDS-PAGE qui peut provoquer l'oxydation du tryptophane (Perdivara *et al.*, 2010).

#### 2.5.2.4 Les autres modifications induites accidentelles

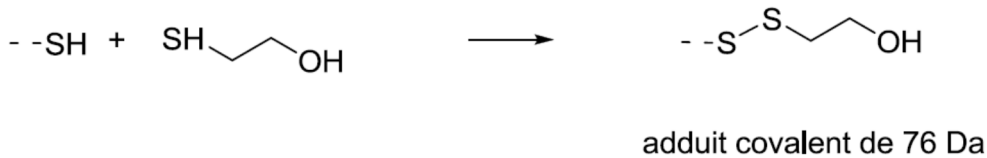
##### a) Addition de $\beta$ -mercaptoéthanol

Le  $\beta$ -mercaptoéthanol est un agent réducteur très utilisé en biochimie pour éviter la formation de ponts disulfures intra- ou inter-protéines. Le  $\beta$ -mercaptoéthanol est efficace mais présente le désavantage, lorsqu'il est en excès, de former un pont disulfure persistant avec les thiols de la protéine et ainsi de former des adduits de +76 Da sur la protéine (Figure 14).

**Figure 14. Mécanismes de réduction des ponts disulfures et d'adduction des cystéines libres par le  $\beta$ -mercaptoéthanol**



Si le  $\beta$ -mercaptoéthanol est en excès:



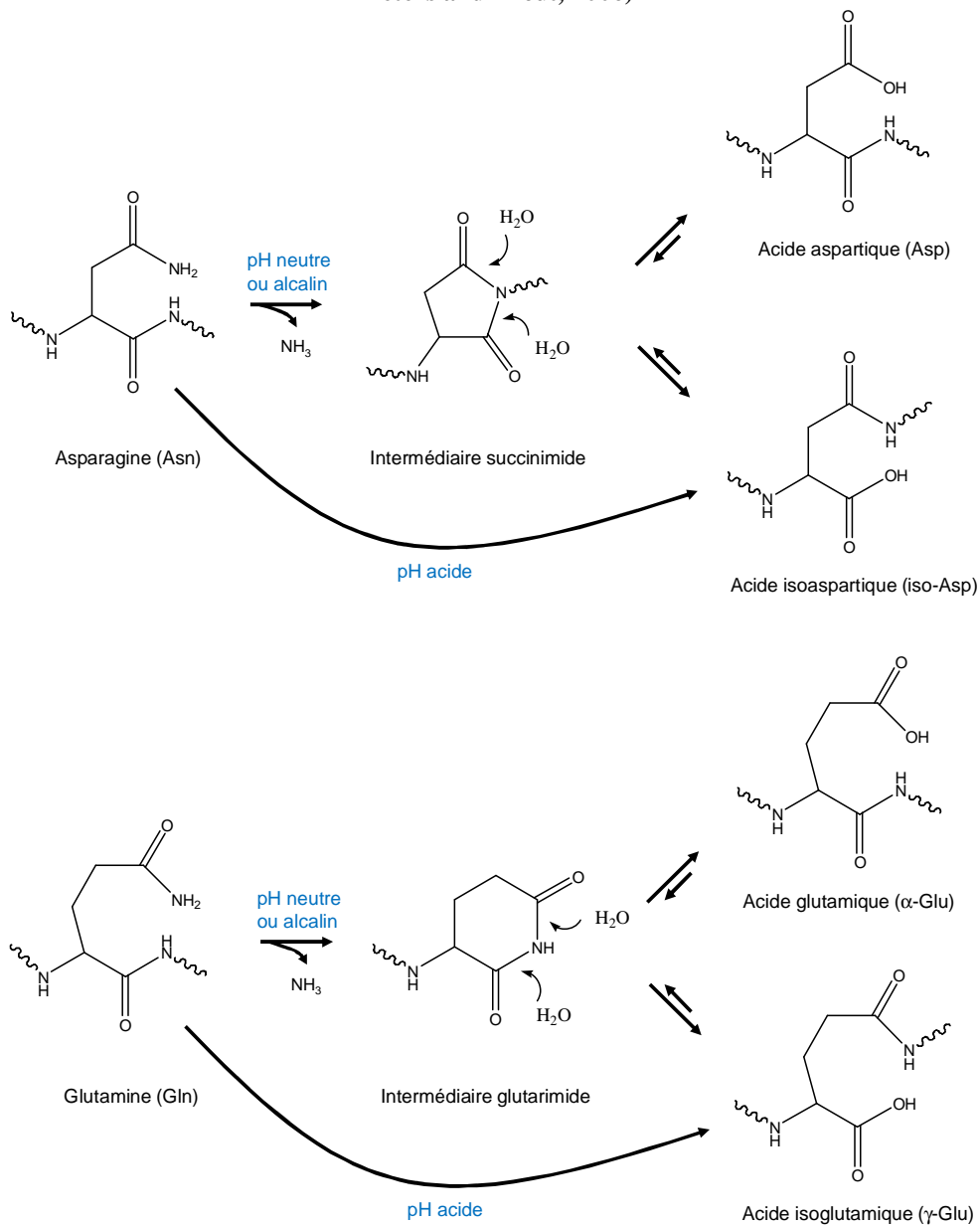
Ces adduits peuvent perturber les interactions protéine-protéine ou protéine-ligand. Il faudra privilégier d'autres antioxydants comme le dithiothréitol (DTT) ou encore la tris(2-carboxyethyl)phosphine (TCEP). Ce dernier présente l'avantage d'être stable dans le temps, à des températures élevées, de fonctionner dans une large gamme de pH et d'être non odorant.

#### b) Déamidation

La déamidation est une PTM impliquée dans la régulation des interactions protéine-ligand et protéine-protéine. Elle correspond à la conversion des asparagines (Asn) et glutamines (Gln) en acides aspartiques (Asp) et acides glutamiques (Glu), respectivement et entraîne un incrément de masse de +1 Da par déamidation. La vitesse de déamidation d'une glutamine est beaucoup plus lente que celle d'une asparagine, exceptée lorsqu'une glutamine est située en N-terminal. Dans ces conditions, la glutamine subit une déamidation plus rapide que l'asparagine pour former un acide pyroglutamique (Li *et al.*, 2010). *In vitro*, la déamidation peut se produire accidentellement et selon le pH, deux voies réactionnelles sont décrites. A pH acide (pH < 3), la déamidation de l'asparagine et de la glutamine se produit par hydrolyse directe de la fonction amide et conduit à la formation d'acide aspartique et d'acide glutamique, respectivement. A pH physiologique et alcalin, l'azote de la liaison peptidique de l'acide aminé N + 1 attaque le carbone du carbonyle de la chaîne latérale de l'asparagine ou de la glutamine pour former un intermédiaire cyclique succinimide ou glutarimide, respectivement. Cet intermédiaire est ensuite hydrolysé et génère deux espèces : l'acide aspartique et l'acide isoaspartique, pour la déamidation de l'asparagine; l'acide glutamique et l'acide isoglutamique pour la déamidation de la glutamine (Figure 15) (Peters and Trout, 2006; Li *et al.*, 2010).



**Figure 15. Les voies de la déamidation de l'asparagine et de la glutamine selon le pH. D'après (Li *et al.*, 2010 ; Peters and Trout, 2006)**



La déamidation introduit une charge négative dans les protéines, ce qui peut perturber leur structure tridimensionnelle et/ou modifier leurs sites d'interaction et donc leur activité.

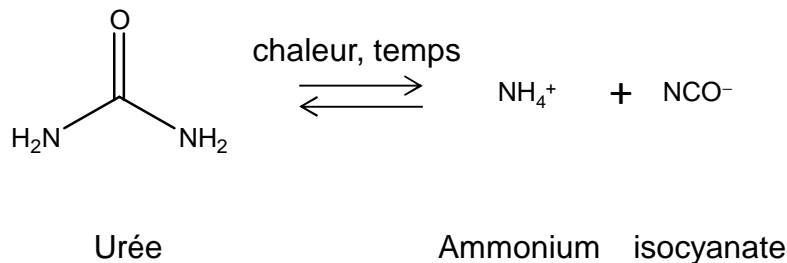
### c) Carbamoylation

La carbamoylation ou carbamylation est une PTM liée au vieillissement et concerne les amines primaires (extrémité N-terminale des protéines et chaîne latérale des lysines). *In vitro* elle peut être générée en présence d'urée, molécule classiquement utilisée pour dénaturer les protéines. Sous l'effet de la chaleur et/ou du temps, l'urée se dégrade en ammonium et isocyanate. L'isocyanate est réactif

et peut réagir covalamment avec les amines primaires, entraînant un incrément de masse de +43 Da par carbamoylation (Figure 16).

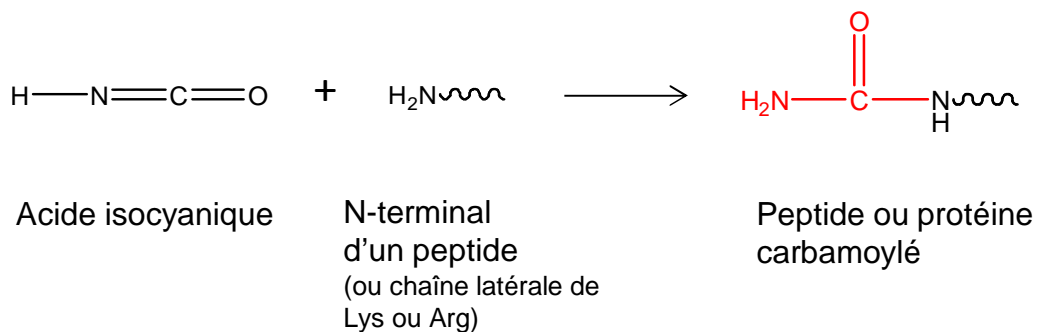
Figure 16. Carbamoylation de peptides ou de protéines par l'urée [d]

## Décomposition de l'urée



## Carbamoylation des protéines

(l'extrémité N-terminale d'un peptide est utilisée comme exemple)



La réaction est favorisée en augmentant la température (Kollipara and Zahedi, 2013) et inhibée par ajout d'ammonium dans le tampon (Sun *et al.*, 2014).

### d) Alkylation à l'iodoacétamide

L'iodoacétamide (IAM) est utilisé dans les approches protéomiques « bottom-up » pour alkyler les cystéines afin de les bloquer et qu'elles ne puissent plus former de ponts disulfures. Les cystéines ainsi modifiées sont carbamidométhylées et entraînent un incrément de masse de +57 Da par carbamidométhyle. Cette modification induite peut créer des artefacts de plusieurs manières. D'une part, elle peut réagir sur d'autres nucléophiles que la cystéine tels que l'extrémité N-terminale, l'histidine, la lysine (Boja and Fales, 2001) ou encore la méthionine (Lapko *et al.*, 2000). D'autre

part, elle peut être confondue avec 1) une mutation par insertion d'une glycine ou 2) la cicatrice d'une PTM d'ubiquitination (incrément de masse de +114 Da après protéolyse trypsique) lorsque les peptides sont deux fois carbamidométhylés (Nielsen *et al.*, 2008).

Ainsi, d'une manière générale, il est important de limiter voire d'éviter l'apparition de ces artéfacts 1) lors de la production et de la conservation de protéines afin qu'elles conservent leur structure et leur activité, et 2) lors de protocoles expérimentaux afin d'obtenir des résultats fiables et cohérents.

### 3. Intérêt des médicaments covalents en utilisation thérapeutique

Bien que près de 30% des médicaments disponibles sur le marché agissent *via* un mécanisme d'action covalent, la majorité d'entre eux ont été découverts lors de criblages par des tests biologiques et le mécanisme covalent n'a été découvert que plus tard. En outre, dans les années 1970, des études ont démontré l'effet hépatotoxique de molécules telles que le bromobenzène et l'acétaminophène (DCI: paracétamol) *via* une liaison covalente (Singh *et al.*, 2011). Ces observations ont été le point de départ d'une perte de confiance vis-à-vis des médicaments covalents et ont conduit à leur désaffection au profit des médicaments non-covalents, réputés moins dangereux. Par conséquent, en dépit de nombreux exemples de médicaments efficaces qui fonctionnent grâce à un mécanisme covalent (par exemple, l'acide acétylsalicylique (aspirine), la pénicilline, l'ésoméprazole et le clopidogrel), l'industrie pharmaceutique a été réticente à utiliser des molécules au mode d'action covalent dans les programmes de recherche de médicaments (Singh *et al.*, 2011).

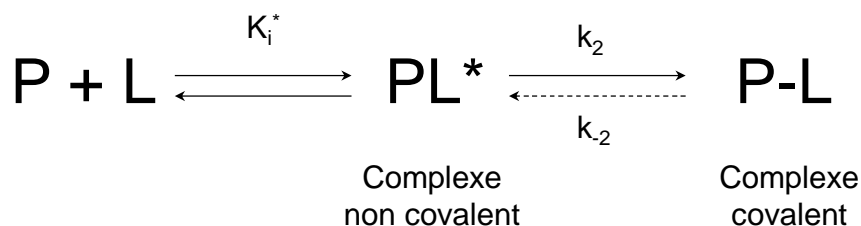
Depuis quelques années, beaucoup d'efforts sont déployés pour cibler les kinases qui, lorsqu'elles sont dérégulées, surexprimées ou mutées, peuvent se retrouver constitutivement activées et entraîner des maladies immunologiques, cardiovasculaires, neurologique (Pandya *et al.*, 2005; Bennett, 2006; Huang *et al.*, 2008) mais également l'apparition de cancers (Santarius *et al.*, 2010). Les premières stratégies visaient à inhiber ces kinases en utilisant une molécule non-covalente et réversible destinée à prendre la place de l'ATP dans le site actif de la kinase cible, et ainsi empêcher l'activité catalytique. C'est ce que l'on appelle l'inhibition par compétition de l'ATP. Le premier succès a été l'imatinib, mis sur le marché en 2001 aux Etats-Unis, qui est très efficace contre la leucémie myéloïde chronique (CML) et les tumeurs du tube digestif (GIST, pour gastro intestinal stromal tumors) et possède peu d'effets secondaires. Ce succès a stimulé l'industrie pharmaceutique qui s'est alors intéressée de plus près à la thérapeutique ciblée *via* l'inhibition de protéines kinases anormalement activées dans les maladies. Ainsi, en 2011, il y avait 11 médicaments ciblant les kinases présents sur le marché. Cet essor est étroitement lié à l'augmentation des structures

tridimensionnelles de kinases disponibles dans les bases de données. En effet, ces informations structurales, combinées à la modélisation moléculaire, ont permis une meilleure compréhension du site actif et ont guidés les chercheurs pour dessiner et concevoir des molécules nouvelles, ce que l'on appelle le drug design.

Toutefois, ces molécules de première génération rencontrent deux difficultés majeures. D'une part, elles ont un problème de sélectivité du fait que le site actif des kinases est très conservé au cours de l'évolution. Ainsi, un inhibiteur développé pour une kinase en particulier peut inhiber d'autres kinases non ciblées initialement (Karaman *et al.*, 2008). D'autre part, ces molécules peuvent être confrontées à des phénomènes de résistance développés par leur cible kinase, entraînant une inefficacité du médicament (Azam *et al.*, 2008). Ces phénomènes de résistance correspondent à des mutations qui peuvent apparaître dans le gène de la kinase et entraîner divers effets selon l'endroit où elle se trouve. Une mutation peut concerner un résidu du site actif et augmenter l'affinité de l'ATP (Yun *et al.*, 2008) ou un résidu en dehors du site actif et dans ce cas modifier la structure tridimensionnelle de la protéine ou bien empêcher l'accès de l'inhibiteur au site actif (Mian *et al.*, 2009).

Face à ces difficultés, une nouvelle génération d'inhibiteurs de kinases plus sélectifs et capables de rester efficaces face aux mutations est actuellement développée. Ces nouveaux médicaments combinent à la fois les modes d'action non-covalent et covalent et ont été reconnus comme très prometteurs. Le mode d'action est un mécanisme en deux étapes : tout d'abord, la molécule médicament (ligand) interagit spécifiquement avec sa cible protéique pour former un complexe non-covalent. Cette interaction permet d'agencer la fonction électrophile du ligand à proximité du résidu nucléophile cible dans la protéine, ce qui entraîne la formation d'un complexe covalent (Figure 17). Comme nous le verrons plus en détail ci-après, les molécules covalentes de deuxième génération sont plus sélectives, ce qui représente un atout majeur dans les aspects de sécurité sanitaire d'un médicament.

**Figure 17. Schéma réactionnel en deux étapes de la formation d'une liaison covalente entre une protéine P et un ligand L.** Le ligand forme tout d'abord un complexe non-covalent avec la protéine cible, selon la constante d'équilibre de dissociation  $K_i^*$ . Cette interaction permet de placer le groupement électrophile à proximité du résidu nucléophile spécifique dans la protéine, résultant ensuite à la formation du complexe covalent, selon la constante de vitesse de réaction  $k_2$ . Dans le cas des liaisons irréversibles,  $k_{-2}$  est égale à zéro.



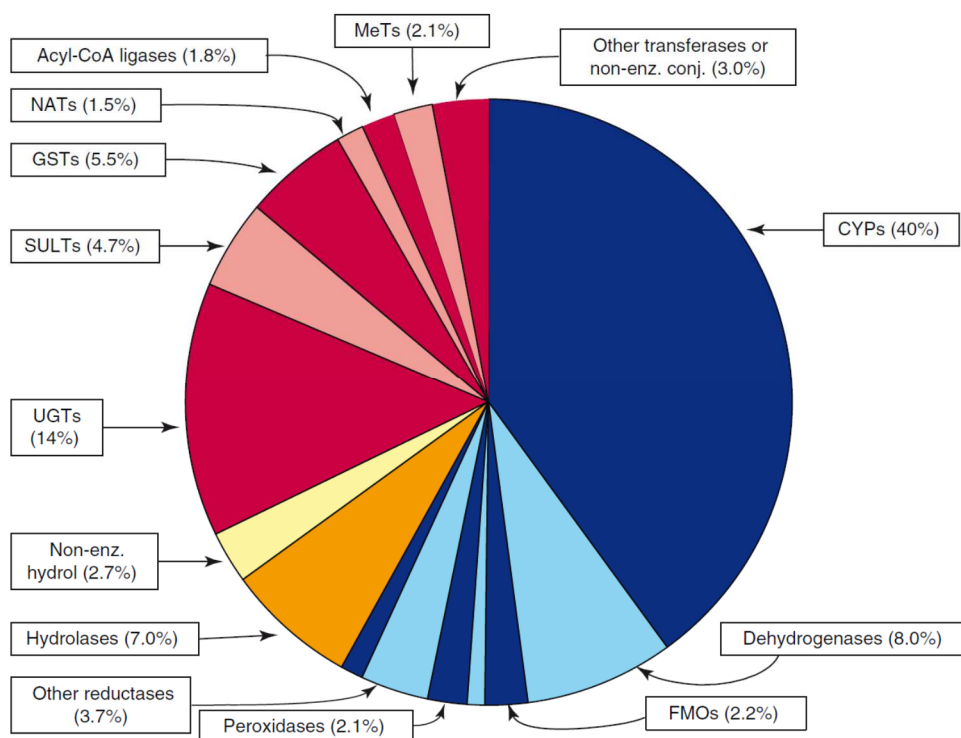
### 3.1 Sécurité sanitaire

Une des raisons de la faveur accordée aux médicaments non-covalents est qu'ils peuvent être éliminés par dialyse dans certains cas de surdosage, alors que les médicaments covalents nécessitent la conception d'un antidote, tel que la N-acétylcystéine pour les réactifs des thiols.

La mauvaise réputation historique des médicaments covalents provient en grande partie des réactions non spécifiques sur des protéines non cibles. La liaison aléatoire d'un médicament covalent peut aboutir à une lésion tissulaire aigüe ou à l'activation du système immunitaire par hapténisation de protéines. L'hapténisation peut conduire à la génération d'anticorps qui reconnaissent des épitopes des protéines modifiées par le médicament et qui sont dès lors considérées comme étrangères à l'organisme. Ces réactions indésirables sont étroitement liées à la réactivité chimique et à la sélectivité de la molécule. Afin de limiter ces réactions, des groupements électrophiles moins réactifs ont été incorporés aux nouveaux médicaments covalents et leur sélectivité a été améliorée en ciblant des résidus peu conservés dans la famille de protéine. Ceci tend à réduire les réactions non spécifiques et à favoriser la réaction covalente au niveau du site actif de la protéine-cible. Le fait de gagner en sélectivité permet de réduire les doses de médicaments administrées, ce qui représente un atout pour la sécurité sanitaire.

Les agents thérapeutiques covalents peuvent être classés en deux catégories : les médicaments qui contiennent un groupement réactif électrophile dans la molécule initiale (par exemple la pénicilline) et ceux qui subissent une biotransformation *in vivo* générant des métabolites électrophiles chimiquement réactifs (par exemple le paracétamol). Ces activations métaboliques concernent principalement les oxydations catalysées par les cytochromes P450 et les glucuronidations catalysées par les UDP-glucuronosyl transférases, mais de nombreuses autres réactions sont possibles (Figure 18). Aujourd'hui, afin d'évaluer les réactions non spécifiques sur les protéines non cibles et ainsi mieux anticiper d'éventuels problèmes de toxicité, des efforts considérables sont investis dans la prédiction *in silico* de la biotransformation et de la toxicité d'une nouvelle molécule (Gleeson *et al.*, 2012). En parallèle, sa réactivité peut être testée sur des thiols non spécifiques tels que le glutathion (Gan *et al.*, 2005) ou encore sur un plasma déplété en albumine. Après incubation, une analyse protéomique permet de déterminer si la molécule testée a modifié de manière covalente les protéines non cibles du plasma (Mah *et al.*, 2014).

**Figure 18. Répartition des métabolites selon les (super) familles ou catégories d'enzymes modifiées (mesure qualitative) (Testa *et al.*, 2012).** Les pourcentages indiqués se réfèrent à 6967 enzymes = 100 % et ont été arrondi à deux chiffres significatifs. Le code de couleur est comme suit : les réactions d'oxydoréduction sont en bleu; les hydrolyses en jaune; les conjugaisons en rouge. L'alternance des champs sombres et clairs est utilisée pour plus de clarté graphique. À partir du quadrant supérieur droit, les boîtes montrent (i) les réactions catalysées par les cytochromes P450, (ii) les réactions catalysées par les déshydrogénases, (iii) les réactions catalysées par les monooxygénases contenant flavine, (iv) les réactions catalysées par la xanthine oxydoréductase et l'aldéhyde oxydase (non indiquée), (v) les réactions catalysées par les peroxydases, (vi) les réactions catalysées par d'autres réductases, (vii) les réactions catalysées par d'autres oxydoréductases ou des réactions d'auto-oxydation (non indiquée), (viii) les réactions catalysées par des hydrolases, (ix) les hydrolyses non-enzymatiques, (x) les réactions catalysées par les UDP glucuronosyl transférases, (xi) les réactions catalysées par les sulphotransférases, (xii) les réactions catalysées par des glutathion-S- transférases et des enzymes associées, (xiii) les réactions catalysées par les N- acétyltransférases, (xiv) les réactions catalysées par des acyl-coenzyme A ligases 1 et enzymes associées, (xv) les réactions catalysées par les méthyltransférases, (xvi) les réactions catalysées par d'autres transférases ou des réactions non enzymatiques de conjugaison. Abréviations: CYP, cytochrome P450 ; FMO, monooxygénase à flavine ; GST, glutathion-S-transférase ; NATs, N -acétyltransférase ; SULT, sulphotransférase ; UGT, UDP –glucuronosyl transférases



Malgré toutes ces précautions, des drames humains arrivent encore aujourd'hui. C'est le cas de la méthyl-bardoxolone qui en activant la voie Nrf2, impliquée dans la réponse au stress oxydatif, a un effet anti-inflammatoire et anti-oxydant. En agissant sur cette voie, elle a un effet inhibiteur sur NF- $\kappa$ B. Les résultats cliniques étaient très encourageants mais les essais ont été interrompus en 2013, en phase III de l'étude, à cause de graves effets indésirables allant jusqu'à la létalité (Wilson *et al.*, 2013). Ces événements nous montrent qu'il faut redoubler d'efforts pour encore mieux comprendre et appréhender les réactions secondaires d'un nouveau médicament.

## 3.2 Sélectivité

Pour gagner en sélectivité, un médicament ne doit pas réagir avec n'importe quel nucléophile. Contrairement aux médicaments covalents de première génération qui étaient très réactifs et qui ciblaient un résidu nucléophile catalytique d'un site actif d'une enzyme (inhibiteur suicide), les nouveaux médicaments covalents, appelés inhibiteurs covalents ciblés (TCI pour Targeted Covalent Inhibitor), sont moins réactifs et ciblent un résidu nucléophile non catalytique et faiblement conservé au sein de la famille de protéines (Singh *et al.*, 2011). Ces composés sont ainsi plus sélectifs que les inhibiteurs réversibles ou les substrats suicides et ne sont pas cantonnés à cibler des enzymes.

Pour gagner en sélectivité, les médicaments covalents présentent un avantage car il est possible d'agir à deux niveaux : la formation du complexe non covalent intermédiaire,  $K_i^*$ , et l'étape de réaction chimique pour former la liaison covalente,  $k_2$  (Figure 17). Ainsi, l'affinité du médicament doit être suffisamment élevée, et donc la valeur de  $K_i^*$  suffisamment basse, pour 1) assurer une interaction sélective avec la cible désirée et 2) avoir un temps de résidence suffisant pour que la réaction covalente ( $k_2$ ) ait lieu. Aujourd'hui, étant donné que les électrophiles hautement réactifs sont proscrits, la réaction covalente est rendu possible en positionnant de manière favorable et optimale le groupement électrophile par rapport au résidu nucléophile de la protéine. De plus, si le résidu ciblé est rare au sein de la famille de la protéine cible, la sélectivité en est encore améliorée.

## 3.3 Efficacité

Les inhibiteurs covalents présentent plusieurs avantages :

- ils restent efficaces face aux mutations du site actif. En effet, bien que ces mutations puissent affecter le  $K_i^*$  de l'interaction protéine-inhibiteur, même les mutants les plus lents, sous réserve d'une exposition suffisante entre l'électrophile et le nucléophile, seront pleinement inhibés. Ainsi, c'est la vitesse à laquelle les complexes covalents sont formés qui va être altérée par ces mutations. Toutefois, si la mutation concerne le résidu ciblé pour la réaction covalente, il n'y aura pas de complexe covalent et l'efficacité du médicament sera modifiée. De même, une équipe de l'entreprise Pfizer a montré très récemment qu'une modification chimique, telle qu'une oxydation de la cystéine 797 de l'EGFR, diminuait de 11 à 1100 fois l'efficacité de certains inhibiteurs (Schwartz *et al.*, 2014).

- si le mécanisme est irréversible, un inhibiteur compétitif reste efficace même si la molécule avec laquelle il est en compétition est présente à forte concentration dans la cellule, ce qui serait par exemple le cas de l'ATP (Yun *et al.*, 2008). Dans ce cas, la concentration en compétiteur agit comme

un "facteur de freinage" qui modifie le  $K_i^*$  apparent selon l'équation:  $K_{i\text{ (app)}}^* = K_i^* (1 + S_o/K_m)$ , où  $S_o$  est la concentration en substrat.

- le temps de résidence dans le site actif est amélioré par rapport au mécanisme non-covalent réversible, ce qui assure une bonne efficacité de l'inhibiteur covalent. Dans le cas d'une liaison covalente irréversible, le temps de résidence est maximum car la protéine cible restera inactivée jusqu'à ce qu'elle soit dégradée (Spicer and Rudman, 2010). Le seul risque de régénérer de la protéine active provient alors d'une éventuelle néo-synthèse par la cellule. Par ailleurs, ces temps de résidence élevés contribuent à améliorer la sélectivité en limitant les temps d'interaction possible avec d'autres cibles (Copeland *et al.*, 2006).

### 3.4 Conception et optimisation de nouveaux inhibiteurs covalents

La conception de médicaments covalents nécessite à la fois une optimisation de l'affinité du complexe non-covalent initial et de la réactivité du groupement électrophile. Au début d'un programme de conception d'un inhibiteur covalent ciblé, l'analyse bioinformatique est utilisée afin d'identifier un résidu nucléophile approprié, c'est-à-dire non catalytique, faiblement conservé, accessible et convenablement orienté, à proximité du site de liaison de la protéine-cible. Ensuite, quelques inhibiteurs sont testés pour leur capacité à modifier la protéine cible. Les composés les plus actifs sont ensuite caractérisés pour déterminer  $K_i^*$  et  $k_2$  (Figure 17), permettant ainsi d'optimiser leur activité de manière raisonnée et selon les données observées. A cette étape du processus, le caractère irréversible ou réversible de la liaison doit être considéré. Dans les deux cas, les ligands ont des temps de résidence élevés mais la liaison covalente réversible présente l'avantage de ne pas former d'adduits persistants sur les protéines non cibles et donc de potentiellement réduire les risques de toxicité indésirable tels que les réactions immunologiques ou idiosyncrasiques<sup>1</sup>.

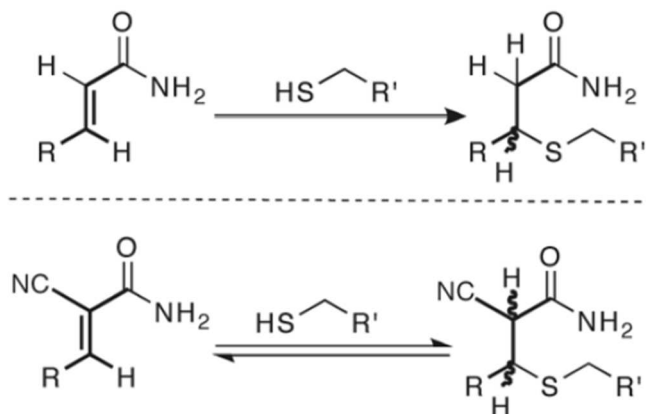
Idéalement, la fonction électrophile doit peu réagir avec les thiols non spécifiques et réagir efficacement avec le nucléophile cible (souvent une cystéine) lorsqu'ils sont maintenus à proximité. Si une liaison irréversible est souhaitée, l'acrylamide est l'électrophile de choix (Garuti *et al.*, 2011) alors que le cyanoacrylamide sera utilisé pour avoir une liaison réversible (Miller *et al.*, 2013) (Figure 19).

---

<sup>1</sup> Une réaction idiosyncrasique à un médicament est une réaction immunologique rare et difficile à prévoir. Elle est généralement initiée par la fixation covalente de molécules, médicaments ou métabolites associés, à des protéines endogènes. Ces protéines endogènes modifiées sont dès lors considérées comme étrangère à l'organisme, déclenchant ainsi une réaction auto-immune très problématique pour le patient.

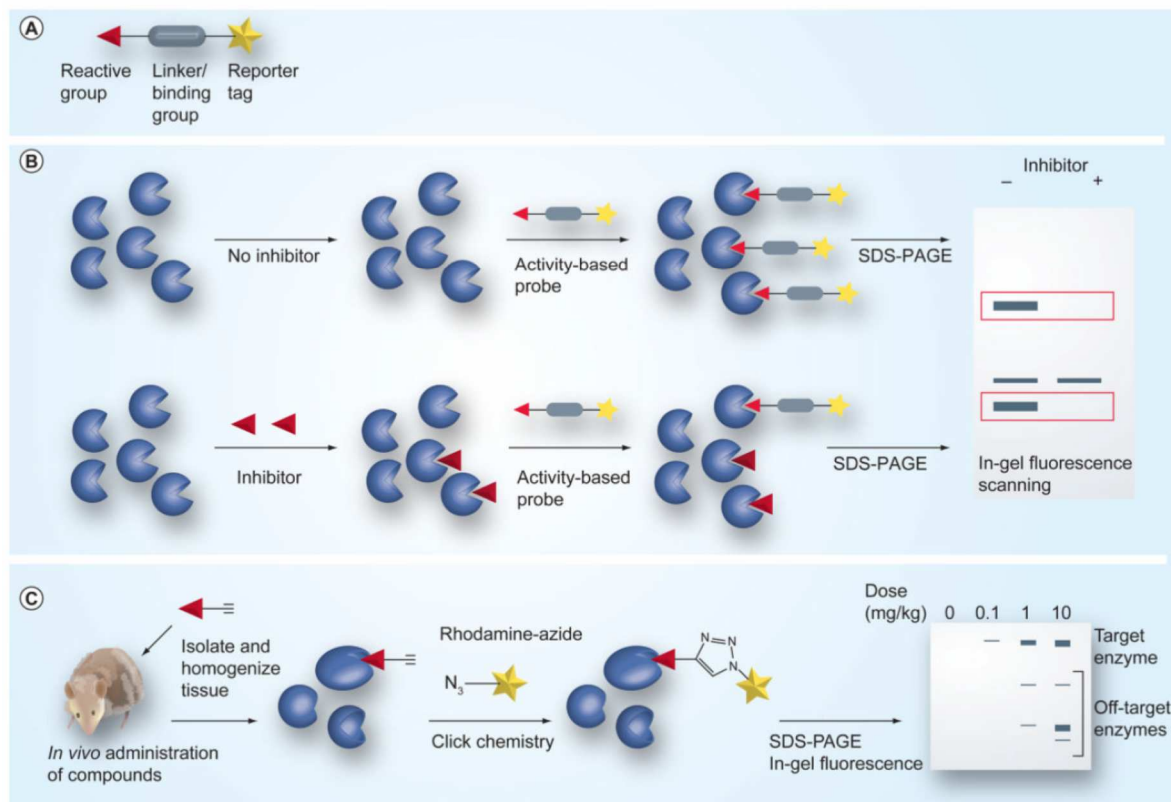


**Figure 19 Addition de Michael irréversible ou réversible (Miller *et al.*, 2013).** L'addition d'un thiol sur un acrylamide conduit à une liaison irréversible alors que l'addition sur un cyanoacrylamide forme une liaison réversible



Comme évoqué précédemment, au cours de la conception d'un inhibiteur, sa réactivité peut être évalué *in silico* et testé *in vitro* sur des thiols non spécifiques. Après ces tests préliminaires, la sélectivité d'un inhibiteur est évaluée par des approches de protéomique fonctionnelle qui visent à profiler les protéines en utilisant des sondes qui ciblent le site actif des protéines (technique ABPP pour activity-based protein profiling). Dans les systèmes vivants, cette stratégie peut être combinée à la « click chemistry », qui utilise la cycloaddition entre un azoture et un alcyne, pour identifier les protéines cibles d'un inhibiteur covalent (Speers and Cravatt, 2004) (Figure 20).

**Figure 20. Evaluation de la sélectivité globale des inhibiteurs covalents par l'approche ABPP (activity-based protein profiling) (Johnson *et al.*, 2010).** (A) Structure représentative de la sonde ciblant le site actif de protéines. Cette sonde contient un groupement réactif, un domaine de liaison et une étiquette comportant un groupement rapporteur. (B) Approche ABPP par compétition afin de déterminer la sélectivité d'un inhibiteur. Les protéines ciblées par la sonde sont analysées par SDS-PAGE (révélation par fluorescence) en absence et en présence d'inhibiteur. Les protéines dont la fluorescence diminue en présence d'inhibiteur correspondent à des protéines ciblées par l'inhibiteur. (C) Pour caractériser la sélectivité d'un inhibiteur *in vivo*, l'approche ABPP peut être combinée à la cycloaddition entre un azoture et un alcynes catalysée par le cuivre (I) (CuAAC) ou « click chemistry ». L'inhibiteur covalent est converti en sonde par incorporation d'un alcynes. Cette sonde est ensuite administrée aux systèmes vivants (cellules ou animaux). Les protéines ciblées par la sonde sont ensuite conjuguées à l'azoture-rhodamine en utilisant la « click chemistry » puis analysées par SDS-PAGE (révélation par fluorescence).



Dans cette approche, la sonde contient une fonction azoture qui pourra ensuite réagir sélectivement avec un fluorophore contenant la fonction alcynes. Les protéines modifiées par la sonde sont ensuite détectées sur gel SDS-PAGE et identifiées par spectrométrie de masse. Bien qu'une quantification puisse être réalisée sur gel SDS-PAGE, par fluorescence le plus souvent, cette dernière est peu précise et peu résolutive. En effet, plusieurs protéines de masses voisines peuvent être contenues dans une même bande de gel. Pour pallier aux limitations de cette méthode, une caractérisation des protéines modifiées incluant leur identification est nécessaire.

De plus, la quantification des sites de réaction des médicaments covalents à l'échelle des acides aminés des protéines est nécessaire pour ces études d'impact de médicaments covalents. Enfin, il est important de pouvoir intégrer très tôt dans le développement du médicament la présence potentielle de ces adduits aux protéines endogènes, et de quantifier ces adduits. Les pré-requis pour une méthode quantitative fiable seront détaillés dans la partie "objectifs de thèse".

## Section 2. Les bases de la caractérisation de modification de protéines par spectrométrie de masse

La spectrométrie de masse (MS) est une technique d'analyse qui permet de mesurer précisément le rapport masse sur charge ( $m/z$ ) de molécules ionisées en phase gazeuse, et d'en déduire la masse  $M$  des molécules analysées. L'analyse MS comprend trois étapes. La première étape consiste à convertir les molécules en ions en phase gazeuse, ce qui représente un défi considérable pour les molécules qui sont en phase solide ou liquide. Ensuite, les ions sont séparés selon leur rapport  $m/z$ , *via* des champs magnétiques ou électriques, dans un analyseur. Enfin, les ions séparés sont détectés. Depuis la fin des années 1980 et l'invention des méthodes d'ionisation douces MALDI (Karas and Hillenkamp, 1988) et ESI (Fenn *et al.*, 1989), l'analyse de macromolécules biologiques par MS est possible. En 25 ans, grâce à une sensibilité, une sélectivité et une vitesse d'analyse plus élevées que la plupart des techniques biochimiques, la MS est devenue incontournable pour l'étude des protéines et de leurs modifications. Ainsi tout incrément ou défaut de masse généré suite à une modification peut en principe être détecté par MS. La caractérisation de modifications de protéines par MS comprend deux volets : 1) l'identification de modifications et la localisation de sites modifiés et 2) la quantification de ces modifications.

Des enjeux et défis analytiques demeurent cependant dans ce domaine, notamment au niveau de :

- la détection et localisation exhaustive des modifications d'une protéine,
- la caractérisation des protéoformes dues aux combinaisons de modifications sur des sites multiples, en particulier à partir de milieux où ces protéoformes sont en mélange,
- la quantification en parallèle des modifications en termes de degré de modification site par site et/ou la quantification de chaque protéoforme relative aux autres protéoformes.

# 1. Identification de modifications et localisation de sites modifiés

## 1.1 Détermination du type et du nombre de modifications sur protéine entière

Pour étudier les modifications de protéines par spectrométrie de masse, il est pertinent, lorsque cela est possible, d'analyser la protéine entière, afin d'estimer le type et le nombre global de modifications et ainsi recenser les différentes protéoformes d'une protéine. De telles mesures, utilisées en parallèle avec le calcul de la masse moléculaire théorique de la protéine non modifiée, permettent la détermination du nombre moyen de modifications par protéine pour un type de modification donné. En supposant que la résolution MS soit suffisamment élevée et dans les cas où il y a plusieurs modifications sur la protéine, il est également possible de déterminer la distribution du nombre de modifications. Bien que ces informations soient précieuses, elles ne permettent pas d'identifier le ou les résidu(s) ciblé(s) par une modification. Pour améliorer la résolution spatiale et localiser les sites modifiés, il existe deux stratégies qui seront développées dans les paragraphes 1.2 et 1.3 suivants :

- l'approche « bottom-up » qui consiste à analyser des peptides issus d'une protéolyse enzymatique de la protéine en MS et MS/MS
- l'approche « top-down » qui consiste à fragmenter la protéine modifiée entière directement dans le spectromètre de masse.

La détermination du type et du nombre de modifications sur protéine entière va dépendre de la complexité de l'échantillon à analyser. Ce type de mesure s'adresse plutôt à des échantillons peu complexes tels qu'une protéine purifiée et est particulièrement utile dans des expériences de cinétique d'association d'un ligand sur une protéine ou lorsque l'on souhaite comparer la réactivité de plusieurs ligands sur une même protéine.

Dans le cadre d'analyses d'échantillons plus complexes tels qu'un protéome entier ou un fluide biologique, la mesure de la masse des protéines entières est très difficile et nécessite plusieurs étapes de fractionnement avant l'analyse MS. Aujourd'hui, seuls les laboratoires experts en approche top-down se sont lancés dans ce type de séparations (cf paragraphe 1.3 ci-après). Dans la très grande majorité des études protéomiques bottom-up, la détermination du type et du nombre de modifications sur protéine entière n'est pas réalisée car bien qu'utile, cette information n'est pas indispensable pour l'identification et la localisation de modifications.

## 1.2 Identification et localisation de modifications par approche bottom-up

### 1.2.1 La protéolyse : génération de peptides

Dans leur conformation native, les protéines sont repliées. Elles possèdent des structures secondaires et tertiaires qui sont maintenues par des liaisons non covalentes (liaisons hydrogène, interactions ioniques, forces de van der Waals, effet hydrophobe) et parfois covalentes (ponts disulfures). Le processus de clivage par une protéase nécessite que les sites de clivages se situent dans des régions accessibles et, pour certaines protéases, flexibles.

On peut distinguer cinq types de protéolyse selon qu'il reste ou non de la protéine intacte et des peptides contenant des sites clivables non clivés (= missed cleavage) : la protéolyse totale, complète, incomplète, ménagée ou la protéolyse pour la détermination de domaines (Tableau 3).

**Tableau 3. Les différents types de protéolyses.**

Type de protéolyse	Protéolyse totale	Protéolyse complète	Protéolyse incomplète	Protéolyse ménagée	Protéolyse pour la détermination de domaines
Couverture de séquence théorique					
Protéine intacte	absence	absence	présence	présence	absence
MC sur peptides <sup>a</sup>	absence	présence	présence	présence (1 site de coupure <u>différent</u> par chaîne)	présence (1 site de coupure <u>identique</u> par chaîne)
Dénaturation préalable de la protéine	oui	oui	oui	non	non

<sup>a</sup> MC : Missed Cleavage

- Historiquement, la définition de la protéolyse totale correspond à une coupure à chaque acide aminé produisant uniquement des acides aminés (pour l'analyse quantitative d'acides aminés par exemple). Cette définition est abandonnée pour la suite de ce glossaire. Dans la définition actuelle, elle correspond à une coupure à chaque résidu clivable par une protéase spécifique. En fin de réaction, la protéine intacte n'est pas observée et il n'y a pas de MC. Sous réserve de l'absence de problèmes de

solubilité ou d'adsorption, la concentration de chaque peptide est égale à la concentration de la protéine intacte de départ. La protéolyse totale est la protéolyse idéale pour la caractérisation de modifications de protéine. Pour certaines applications, une protéase non-spécifique peut être utilisée. A l'extrême, une protéase qui génère une coupure à chaque acide aminé produirait uniquement des acides aminés (aucun peptide), donc aucune information de séquence.

Pour atteindre la protéolyse totale d'une protéine, il convient de la déplier et de rendre accessible tous ses sites de clivage. Ce dépliement est réalisé par une étape de dénaturation qui consiste à rompre les liaisons non covalentes et éventuellement par une étape de réduction de liaisons disulfure covalentes. Les liaisons non covalentes peuvent être rompues par une dénaturation thermique et/ou chimique par ajout d'agents chaotropiques (urée ou chlorure de guanidinium), de détergents ioniques (dodécyle sulfate de sodium (SDS) et/ou de solvants organiques (acétonitrile). Les ponts disulfures sont classiquement rompus par une étape de réduction au DTT ou au TCEP, puis les fonctions thiols libérées sont bloquées par alkylation avec un agent alkylant tel que l'iodoacétamide (IAM) afin d'éviter la formation de nouveaux ponts disulfures. La protéase est ensuite ajoutée dans des rapports enzyme:substrat (E:S) variant de [1:5] à [1:100] (p/p) avec des temps d'incubation variant de 2 à 18 h selon le rapport utilisé.

Cependant, en pratique, à cause de conditions de protéolyse non optimisées, de conformations tridimensionnelles de protéine modifiée ou intacte résistantes à la dénaturation, ou de la présence de MCs liés à la séquence ou à la modification, la protéolyse totale est difficilement obtenue et c'est plutôt la protéolyse complète voire incomplète qui est atteinte (Tableau 3).

- La protéolyse complète définit un clivage à la plupart des résidus clivables mais pas tous. Classiquement, en fin de réaction, chaque chaîne de protéine est coupée au moins une fois et par conséquent la protéine intacte n'est pas observée, mais il subsiste des MCs au niveau des peptides. Ces MCs sont essentiellement dus à la présence de certains résidus à proximité du site de coupure qui perturbent l'activité de la protéase. C'est le cas par exemple des résidus P, K, R, E ou D en P1' du site de coupure (Figure 5) qui perturbent l'activité de la trypsine et génère des MCs (Siepen *et al.*, 2007).

Dans des conditions de protéolyse complète, seuls les peptides sans MC ont une concentration égale à celle de la protéine intacte de départ, sous réserve de l'absence de problèmes de solubilité ou d'adsorption.

- La protéolyse incomplète correspond à l'observation de protéine intacte. Cette protéolyse incomplète peut-être due à des conditions de réactions insuffisamment poussées (temps trop court,

rapport E:S trop bas) ou à la présence de conformations résistantes à la protéolyse, éventuellement due à une résistance à la dénaturation si celle-ci a été appliquée.

Cas particuliers :

- La protéolyse ménagée est utilisée en biologie structurale pour explorer la structure 3D (probing de structure, sites accessibles, footprinting, variations entre homologues). Elle implique qu'il y ait toujours 1) de la protéine intacte afin d'éviter de dépasser le "single-hit statistics" et 2) un seul site de coupure par chaîne. C'est un cas particulier de la protéolyse incomplète.

- La protéolyse pour la détermination de domaines est également utilisée en biologie structurale. Elle échappe au pré-requis du "single-hit statistics" puisque l'on cherche à déterminer les domaines résistants à la protéolyse indépendamment de la relation de stabilisation entre domaines. On compte ici sur la stabilité intrinsèque du domaine clivé. C'est un cas particulier de la protéolyse complète où l'on tire partie de la présence de zones conformationnelles résistantes à la protéolyse, et où l'on s'abstient donc d'appliquer une dénaturation.

Pour la caractérisation de modification de protéine, il faut se placer dans des conditions de protéolyse totale ou complète afin de s'assurer 1) que les formes modifiée et non modifiée de la protéine sont entièrement protéolysées et ainsi qu'il n'y a pas de différence de protéolyse entre les deux formes et 2) que les peptides modifiés et non modifiés sont correctement générés.

Pour cela, il faut

- 1) vérifier qu'il ne reste plus de protéine intacte après protéolyse,
- 2) analyser la cartographie peptidique en termes de couverture de séquence qui indique si chaque résidu de la séquence est recouvert par au moins un peptide en identifiant les zones non observées,
- 3) analyser la cartographie peptidique en termes de degré de protéolyse en détectant la présence et le nombre de MC dans chaque peptide de clivage.

## 1.2.2 Détection des peptides modifiés

### 1.2.2.1 Empreinte peptidique massique

La méthode classique d'identification d'une protéine repose sur la génération en spectrométrie de masse d'un signal corrélé à sa structure primaire. Dans cette méthode, les peptides issus d'une protéolyse enzymatique spécifique (par exemple par protéolyse à la trypsine) sont analysés en spectrométrie de masse. On obtient ainsi une liste de masses  $M$  issues des  $m/z$  spectraux des peptides, que l'on appelle l'empreinte peptidique massique de la protéine. En anglais, cette méthode s'appelle PMF pour Peptide Mass Fingerprinting. L'empreinte peptidique obtenue est ensuite comparée à celle prédite pour toutes les protéines connues et répertoriées au sein d'une banque de données génomique (avec traduction *in silico*) ou de protéines (SwissProt par exemple). Or, les masses de peptides peuvent correspondre à plusieurs produits de protéines, les masses de peptide unique à une protéine représentant une très faible minorité, même pour des masses mesurées à haute résolution. De plus, des formes isomères de protéines peuvent avoir des séquences en commun, ce qui implique de détecter des peptides en-dehors de ces séquences pour lever les ambiguïtés d'identification. Par conséquent, plus le nombre de peptides identifiés et la couverture de séquence de la protéine sont élevés, plus la probabilité d'identifier la protéine sera élevée. Une méthode plus récente et plus fiable à des fins d'identification consiste à analyser les peptides en MS/MS et non plus seulement en MS.

Sous réserve de connaître l'incrément de masse généré par une modification, la méthode PMF originellement destinée à l'identification peut être détournée pour identifier des peptides modifiés. Cette approche est relativement simple à mettre en œuvre, mais elle ne permet pas de détecter ou d'annoter tous les peptides modifiés. En effet, certains peptides modifiés ne seront pas détectés car trop faibles en intensité (voir le paragraphe 1.2.3.1 ci-après) et d'autres ne seront pas annotés car des différences de masses inattendues, générées par une mutation ou une spécificité de clivage non suspectée (par exemple une contamination de la trypsine par la chymotrypsine), n'auront pas été considérées pour la recherche dans les bases de données. Dans le cas des modifications labiles en spectrométrie de masse, il est possible de s'affranchir de ces problèmes d'annotation en recherchant des ions diagnostics de la modification. Cette approche est présentée dans le paragraphe suivant.



### 1.2.2.2 Détection en MS/MS : Modes d'analyse en spectrométrie de masse avec fragmentation

Les peptides modifiés par un groupement labile peuvent être détectés spécifiquement dans un mélange de peptides en utilisant la perte de la modification (ou d'une partie seulement) lors de la dissociation activée ou induite par collision (CAD : collisionally-activated dissociation, CID : collision-induced dissociation).

#### *a) Balayage d'ions précurseurs (PIS)*

Le balayage d'ions précurseurs (PIS : precursor ion scan), également appelé balayage d'ions parents, utilise la détection de fragments spécifiques de la modification générés par CID ou CAD afin de signaler la présence de peptide modifié. Ce type d'analyse implique typiquement les spectromètres de masses à triple quadripôle qui sont des analyseurs à balayage. Classiquement, le premier quadripôle (Q1) balaye constamment toute la gamme de masse de l'instrument, la fragmentation CID étant induite dans le second quadripôle (Q2). Le troisième quadripôle (Q3) est réglé pour laisser passer sélectivement les ions d'un certain  $m/z$  (ions marqueurs ou ions diagnostics). Ainsi, seuls les peptides modifiés qui se fragmentent pour générer les ions marqueurs choisis sont présents dans le spectre final. Cette approche a été utilisée avec succès à des protéines phosphorylées (Huddleston *et al.*, 1993) ou glycosylées (Ritchie *et al.*, 2002).

Dans le cas des phosphorylations, les ions diagnostics recherchés sont les  $m/z$  97 ( $\text{H}_2\text{PO}_4^-$ ), 79 ( $\text{PO}_3^-$ ) et 63 ( $\text{PO}_2^-$ ) pour le mode négatif et 216.043 (ion immonium de la phosphotyrosine) pour le mode positif. Dans le cas des glycosylations, l'ion diagnostic recherché est le  $m/z$  204 correspondant au  $[\text{M}+\text{H}]^+$  d'un hexose N-acétylé (HexNAc).

#### *b) Balayage en perte de neutre (NLS)*

Le balayage en perte de neutre (NLS : neutral loss scan) utilise également la spectrométrie de masse en tandem (avec deux analyseurs quadripôlares) pour détecter, en mode positif, la perte de neutre de la modification après CID. Cette approche n'est pas adaptée aux modifications de masse variable telle que la glycosylation car la perte de neutre n'est pas constante. Dans le cas des phosphorylations, la fragmentation CID des peptides contenant des phosphosérines et phosphothréonines conduisent souvent à une perte de neutre  $\text{H}_3\text{PO}_4$  *via* une  $\beta$ -élimination en phase gazeuse. Les pics correspondant à cette perte (-98, -49 et -32.67 Da à partir de précurseurs monochargés, dichargés et trichargés, respectivement) sont souvent les ions les plus abondants dans les spectres CID. Les trappes ioniques peuvent même détecter ces pertes de neutre et déclencher automatiquement une fragmentation MS/MS/MS ( $\text{MS}^3$ ) pour localiser le site phosphorylé (Jin *et al.*, 2005) (voir paragraphe 1.2.4.3).

## 1.2.3 Fractionnement et enrichissement des peptides modifiés

### 1.2.3.1 Difficultés associées à la détection des peptides modifiés en mélange

Idéalement, tous les peptides modifiés d'une protéine devraient être détectés. Malheureusement, l'analyse MS de peptides issus d'une protéolyse fournit rarement 100% de couverture de la séquence protéique (même pour des protéines non modifiées). En effet, dans un mélange peptidique, il y a compétition pour l'ionisation, ce qui entraîne la suppression du signal des peptides faiblement ionisés. L'ionisation d'un peptide va dépendre de sa séquence, du type d'ionisation utilisé (électrospray ou MALDI), de la polarité utilisée pour l'analyse (positive ou négative), de la complexité du mélange et de la présence de molécules interférentes telles que des sels, tampons ou détergents. Ce phénomène de suppression est accentué pour les peptides modifiés car, la modification étant généralement présente en proportions sous-stœchiométriques, ils sont sous-représentés par rapport aux peptides non modifiés. De plus, suivant les propriétés physico-chimiques de la modification, l'efficacité d'ionisation du peptide modifié peut être sévèrement affectée. C'est le cas par exemple de la phosphorylation qui apporte une charge négative additionnelle qui perturbe l'observation des phosphopeptides en mode positif (Liao *et al.*, 1994). Cette charge négative peut également gêner la protéolyse par la trypsine qui est spécifique des résidus basiques lysine et arginine qui sont chargés positivement dans les conditions de protéolyse standard par la trypsine (McLachlin and Chait, 2001).

Pour l'étude des modifications de protéines et de leurs peptides, la préparation des échantillons est une étape cruciale dont dépendent fortement les méthodes de détection telles que la spectrométrie de masse. En effet, les échantillons biologiques sont trop complexes pour être analysés dans leur forme brute et des étapes de fractionnement, d'enrichissement, de concentration ou de dessalage sont absolument nécessaires avant de réaliser une quelconque analyse en spectrométrie de masse. Ces techniques sont essentiellement développées pour les peptides mais elles peuvent être transposées aux protéines la plupart du temps. La méthode classique pour réduire la complexité d'un mélange peptidique est de le fractionner par HPLC.

### 1.2.3.2 Fractionnement par HPLC

De façon courante, le fractionnement par HPLC consiste à séparer les peptides sur phase inverse (phase stationnaire apolaire composée de chaînes aliphatiques de dix-huit carbones : C18) en fonction de leur hydrophobie. L'HPLC présente également l'avantage de concentrer les échantillons dilués et d'éliminer les sels qui interfèrent avec le processus d'ionisation. C'est une technique reproductible,

sensible et automatisable. La séparation HPLC est très utilisée pour sa facilité de connexion en ligne avec l'ESI-MS. Pour traiter les faibles quantités de peptides généralement disponibles, il est souvent nécessaire de travailler en nanodébits (de 50 à 500 nL/min). On parle alors de nano-HPLC ou nanoLC.

Depuis une dizaine d'année, les développements de la phase stationnaire (réduction de la taille des particules à moins de 2 µm, meilleure fonctionnalisation) et de l'instrumentation (augmentation des débits et des pressions jusqu'à 1000 bars) ont permis l'essor de chromatographies liquides ultra performantes appelées UHPLC. Ces dernières permettent de diminuer le temps d'analyses à quelques minutes (au lieu de plus d'une heure) et d'augmenter la résolution pour atteindre des largeurs de pic à mi-hauteur de trois à quatre seconde. La nanoLC a également bénéficié de ces avancées: on peut parler alors de nano-UHPLC. En aval des chaînes UHPLC, il faut des spectromètres de masse qui travaillent à haute fréquence (~10 Hz) pour assurer des cycles MS - MS/MS suffisamment courts pendant l'élution des peptides. Les spectromètres de masse modernes tels que les QTOF et les orbitrap sont capables d'effectuer 10 MS/MS en une seconde, ce qui permet à la fois d'identifier plus de protéines et d'améliorer leur couverture de séquence

L'HPLC permet notamment de séparer des peptides de séquences différentes mais qui aurait la même masse (isobares), en tirant partie de la différence de conformation résiduelle entre ces peptides.

Pour l'analyse de mélange très complexe, des séparations LC bi-dimensionnelles sont également utilisées. En utilisant une propriété de séparation différente de l'hydrophobie (charge, hydrophilie), cette deuxième dimension orthogonale permet i) de détecter plus de peptides et donc d'identifier plus de protéines et ii) de diminuer les interférences liées à la matrice et ainsi d'améliorer la précision des mesures en quantification. Les couplages LC-LC décrits sont les suivant :

- Echange de cations – phase inverse (SCX-RP : Strong Cation eXchange – Reversed Phase) (Wolters *et al.*, 2001)
- Echange d'anions – phase inverse (SAX-RP : Strong Anion eXchange – Reversed Phase) (Zhou *et al.*, 2011)
- Phase inverse pH basique – phase inverse pH acide (High pH-RP) (Delmotte *et al.*, 2007)
- Interactions hydrophiles – phase inverse (HILIC-RP : Hydrophilic Interaction Liquid Chromatography-Reversed Phase) (Di Palma *et al.*, 2011)

### 1.2.3.3 Enrichissement d'espèces modifiées

L'enrichissement de peptides modifiés devient nécessaire lorsque le pourcentage de modification est faible ou que les mélanges sont complexes, c'est-à-dire lors de protéolyses effectuées sur des protéines de taille élevée ou des mélange de protéines. Les approches les plus sélectives sont l'immunoprécipitation et la chromatographie d'affinité.

#### a) *Enrichissement par immunoprécipitation*

L'immunoprécipitation (IP) utilise un anticorps, généralement attaché de façon covalente à une résine, pour fixer spécifiquement les protéines et/ou les peptides à partir de mélanges complexes tels que des lysats cellulaires. Plusieurs lavages sont nécessaires pour éliminer les interactions non spécifiques. Les peptides et/ou les protéines d'intérêt sont ensuite élués pour analyse.

Dans le cas de la phosphorylation, les anticorps spécifiques des résidus phosphotyrosines sont les plus utilisés car ils présentent une spécificité élevée et une affinité assez forte pour permettre une immunoprécipitation efficace. Les anticorps anti-phosphosérines et anti-phosphothréonines ne sont pas très spécifiques et dépendent souvent de séquences consensus. Bien que l'immunoprécipitation soit le plus souvent employée pour enrichir des protéines phosphorylées à partir de lysats cellulaires (Salomon *et al.*, 2003), elle a également été utilisée avec succès pour enrichir des phosphopeptides (De Corte *et al.*, 1999).

Dans le cas de l'acétylation, les protéines acétylées peuvent être enrichies en utilisant des anticorps spécifiques des lysines acétylées.

#### b) *Enrichissement par chromatographie d'affinité sur métal immobilisé*

La chromatographie d'affinité sur métal immobilisé (IMAC : immobilized metal affinity chromatography) a été décrite pour la première fois par Porath *et al.* (1975). Une des applications les plus connues est la purification de protéines recombinantes porteuses d'une séquence de six histidines. L'IMAC est très utilisée aujourd'hui pour l'enrichissement des phosphopeptides. Dans cette technique, un groupement chimique tridentate (acide iminodiacétique : IDA) ou tétradentate (acide nitrilotriacétique : NTA; carboxyméthyl aspartate: CM-Asp ou TALON<sup>TM</sup>) capable de chélater un métal multi-valent est tout d'abord lié de manière covalente sur un support chromatographique (agarose, silice, sépharose, polystyrène-divinylbenzène). Un ion métallique chargé positivement ( $\text{Fe}^{3+}$ ,  $\text{Ga}^{3+}$ ,  $\text{Al}^{3+}$ ,  $\text{Zr}^{4+}$ ) est ensuite lié à ce support et permet une interaction électrostatique avec le groupement phosphate chargé négativement. Parmi les différents ions métalliques testés, il apparaît

que le plus efficace pour retenir puis pour éluer les phosphopeptides est l'ion ferrique  $Fe^{3+}$  (Reinders and Sickmann, 2005).

Classiquement, les phosphopeptides sont fixés sur la colonne dans des conditions acides puis élués à pH alcalin. Toutefois, l'IMAC présente plusieurs limitations : la perte possible de phosphopeptides à cause de leur incapacité à se fixer sur la colonne, l'enrichissement préférentiel des peptides multiphosphorylés (Mann *et al.*, 2002), l'élution difficile de certains peptides multiphosphorylés (Posewitz and Tempst, 1999) et le bruit de fond généré par les peptides non phosphorylés contenant des résidus acides (acides aspartique et glutamique) ou donneurs d'électrons (histidine) qui ont également une affinité pour les ions métalliques. La sélectivité des colonnes IMAC et l'élution des espèces phosphorylées apparaissent ainsi être largement dépendantes du type d'ion métallique, du matériel de la colonne, des conditions de chargement de l'échantillon et des conditions d'élution utilisés (Mann *et al.*, 2002). Par ailleurs, les colonnes sont souvent miniaturisées pour récupérer de faibles volumes d'élution (Posewitz and Tempst, 1999; Stensballe *et al.*, 2001).

Le pH de l'échantillon peut affecter la sélectivité des colonnes IMAC. En diminuant le pH, Posewitz et Tempst (1999) ont observé une amélioration de la sélectivité (moins de peptides non phosphorylés acides enrichis). Le pH a une influence sur l'état de dissociation du phosphate et des groupements carboxyles des résidus Asp et Glu. Le phosphate, une fois fixé sur un résidu Ser, Thr ou Tyr, est un diacide capable de céder deux protons et dispose donc de deux constantes d'acidité  $pK_{a1}$  (2.15) et  $pK_{a2}$  (7.20). Le  $pK_{a1}$  du phosphate (2.15) est plus faible que celui du groupe carboxyle de Asp (3.65) et de Glu (4.25) (Kokubu *et al.*, 2005). La meilleure sélectivité a été obtenue pour des pH de 2.0-2.5. Des valeurs plus élevées de pH entraînent la rétention de peptides acides non phosphorylés tandis que des valeurs de pH inférieures à 2.0 diminuent la rétention des phosphopeptides. De plus, les interactions hydrophobes (fixation non spécifique sur les billes) peuvent être diminuées en utilisant un solvant organique tel que l'acétonitrile (Kokubu *et al.*, 2005).

L'élution des phosphopeptides est réalisée classiquement par compétition à pH basique avec l'hydroxyde d'ammonium mais l'élution à pH acide avec l'acide phosphorique a également été décrite (Stensballe and Jensen, 2004). Ces conditions acides et non salines permettent de s'affranchir de l'étape de dessalage nécessaire dans le cas de l'élution à pH basique.

Le problème de fixation non spécifique des peptides non phosphorylés contenant des résidus acides a été résolu par Ficarro *et al.* (2002) en convertissant les groupes d'acides carboxyliques en esters de méthyle. Cette approche leur a permis d'identifier plus de 300 phosphoprotéines et plus de 200 sites de phosphorylation chez *Saccharomyces cerevisiae*. Toutefois, les conditions de cette réaction d'estérification doivent être bien établies afin d'éviter une estérification incomplète et une réaction

secondaire avec les asparagines et les glutamines, ce qui augmenterait la complexité de l'échantillon et interférerait avec l'analyse en spectrométrie de masse (He *et al.*, 2004).

Malgré ces limitations, l'IMAC est très employée. Elle peut être couplée en ligne avec l'HPLC ou l'électrophorèse capillaire (Watts *et al.*, 1994; Wu *et al.*, 2000). Des chromatographies bidimensionnelles d'échange d'anions-IMAC (Nühse *et al.*, 2003) et IMAC-échange de cations (Gruhler *et al.*, 2005) ont été également utilisées avec succès. Ficarro *et al.* (2005) ont mis au point une méthode automatisée par IMAC/nanoLC-ESI-MS pour l'enrichissement et l'analyse des phosphopeptides.

#### *c) Enrichissement par chromatographie d'affinité d'oxyde de métal*

Une autre possibilité pour enrichir les espèces phosphorylées est la chromatographie d'affinité d'oxyde de métal (MOAC) utilisant des colonnes de dioxyde de titane (TiO<sub>2</sub>) (Pinkse *et al.*, 2004) ou de dioxyde de zirconium (ZrO<sub>2</sub>) (Kweon and Håkansson, 2006). Ces colonnes sont constituées de microsphères qui présentent une stabilité mécanique, chimique et thermique élevée, ce qui en fait un support alternatif à la silice. En 2004, Pinkse et ses collaborateurs ont utilisés des pré-colonnes de TiO<sub>2</sub> dans un système nanoLC-ESI-MS/MS pour l'enrichissement sélectif de phosphopeptides.

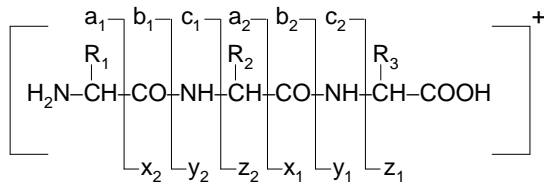
#### *d) Enrichissement par extraction sur support solide avec marquage chimique de site modifié*

Le marquage chimique des résidus modifiés peut être utilisé si la chimie de la modification le permet. Cette approche consiste à introduire, au niveau de chaque site de modification, une étiquette stable et utilisable pour un enrichissement. Dans le cas des phosphorylations, l'introduction d'une étiquette stable peut être réalisée par une réaction de  $\beta$ -élimination suivi d'une addition de Michael de l'éthanedithiol (EDT) ou du dithiothréitol (DTT) sur la déhydro-alanine et d'une fixation à un support solide réactif des thiols (McLachlin and Chait, 2003). Comme c'est généralement le cas avec les approches de dérivation, il faut des quantités importantes d'échantillon et les réactions chimiques doivent être bien contrôlées afin qu'elles soient aussi complètes que possible et qu'il n'y ait pas de réactions secondaires. Des réactions partielles ou non désirées engendrent une complexité accrue des échantillons.

### 1.2.4 Localisation de résidus modifiés

Une fois qu'un peptide a été détecté comme étant modifiés, par PMF ou détection d'ions diagnostics (cf paragraphe 1.2.1), il est analysé par spectrométrie de masse en tandem (MS/MS) afin d'obtenir des informations de séquence et/ou localiser précisément le ou les résidu(s) modifié(s). La

fragmentation d'un peptide (ou d'une protéine) concerne les trois types de liaisons du squelette peptidique, à savoir la liaison peptidique (amide) ou les liaisons C-C et N-C adjacentes. Néanmoins, la dissociation de la liaison peptidique entre le carbone du carbonyle et l'azote de l'amide est la plus fréquemment rencontrée et génère des ions b et/ou y (voir nomenclature Figure 21). L'écart de masse entre deux ions successifs d'une même série (b ou y) correspond à la masse d'un résidu +/- la masse de la modification suivant que le résidu est modifié ou non.



**Figure 21. Nomenclature des ions fragments (d'après Roepstorff and Fohlman, 1984 ; complétée par Biemann, 1988).** Le type d'ions fragments observés dans un spectre MS/MS dépend de plusieurs facteurs incluant la séquence primaire du peptide, la quantité d'énergie interne, la manière dont cette énergie est apportée ou encore l'état de charge.

D'une manière générale, la localisation d'un site modifié est plus difficile que la détection et l'identification d'un peptide modifié. Cela est dû à la nécessité d'avoir des séries d'ions fragments diagnostiques encadrant l'acide aminé modifié afin de l'identifier sans ambiguïté. La faible abondance et l'efficacité d'ionisation parfois défavorable des peptides modifiés contribuent à des faibles rapports signal/bruit ce qui nuit à leur sélection pour la fragmentation MS/MS et conduit des ions fragments de faible intensité.

Lors d'une expérience de fragmentation, l'énergie interne des ions est à considérer. En effet, les ions formés lors du processus d'ionisation, électrospray ou MALDI, sont classés en fonction de leur stabilité au cours de l'analyse. Ainsi, on distingue les ions stables, instables et métastables. Les ions stables (durée de vie  $> 10^{-4}$  s) restent intacts et ne se fragmentent pas d'eux même car ils ont reçu une quantité d'énergie interne insuffisante lors de l'ionisation pour conduire à une quelconque réaction de fragmentation. A l'inverse, les ions instables (durée de vie  $< 10^{-6}$  s) ont un excès d'énergie interne et se fragmentent dans la source pour générer des ions fragments plus stables qui seront observés dans le spectre de masse. Les ions métastables (durée de vie entre  $10^{-6}$  et  $10^{-4}$  s) ont une quantité intermédiaire d'énergie interne et se fragmentent de la source jusqu'au détecteur (Kinter and Sherman, 2000).

Les caractéristiques importantes pour induire la dissociation sont la fenêtre de temps dans laquelle les réactions de dissociation se produisent, la distribution de l'énergie interne des ions activés, et le mécanisme par lequel l'énergie est transférée aux ions. Ainsi, plusieurs techniques de fragmentation sont disponibles suivant les spectromètres de masse utilisés et le caractère labile de la modification étudiée.

#### 1.2.4.1 Décomposition dans la source (ISD)

Le mode *In-Source Decay* ou ISD consiste à analyser les fragments formés dans la source, pendant ou après l'ionisation, par dissociation spontanée des ions instables (prompt fragmentation) et dissociation spontanée ou induite des ions métastables. Dans les sources ESI, l'efficacité de fragmentation peut être modulée en faisant varier 1) la tension d'accélération  $V_C$  qui est la différence entre la tension de sortie du capillaire de transfert  $V_{CE}$  et la tension du skimmer  $V_{SK}$  ou 2) la tension au niveau des funnels. Dans le cas des sources MALDI, on parle également de décomposition induite par laser (LID) (Katta *et al.*, 1998).

#### 1.2.4.2 Décomposition après la source (PSD)

Le mode PSD permet l'analyse d'ions fragments résultant de la décomposition d'ions métastables. Dans les spectromètres de masse MALDI-TOF, la décomposition métastable des peptides est un phénomène courant et spontané. Les premières acquisitions en MALDI-TOF-PSD étaient longues et consommatrice en échantillon car elles consistaient à diminuer par étape les tensions au niveau du réflectron afin de refocaliser les différents ions fragments générés. Depuis 2003, la technologie LIFT permet de gagner en sensibilité, résolution et temps d'acquisition en redonnant de l'énergie cinétique aux ions métastables (Suckau *et al.*, 2003). Pour ce type d'analyse, c'est la matrice HCCA qui est classiquement utilisée car sa faible affinité protonique permet de transférer plus d'énergie aux peptides, ce qui favorise les fragmentations métastables.

#### 1.2.4.3 Dissociation induite par collision (CID)

La dissociation induite par collision (CID) est la méthode de fragmentation la plus courante pour identifier les sites modifiés. Une fois sélectionné, le peptide à fragmenter arrive dans une cellule de collision et entre en collision avec des molécules d'un gaz inerte tel que l'hélium, l'azote ou encore l'argon. L'énergie cinétique de ces collisions est convertie en énergie vibrationnelle dans le peptide, ce qui entraîne des réactions de fragmentation dirigées par le site de protonation dans la séquence. Ainsi, la migration du proton, provenant de l'amine N-terminale, le long des différentes liaisons amides du peptide est capitale pour générer différents fragments et obtenir une information de séquence. Selon la théorie du proton mobile, la présence d'un résidu fortement basique à l'intérieur de la séquence peptidique peut, suivant sa localisation, affecter sensiblement le déplacement du proton mobile N-terminal et donc l'efficacité de fragmentation (Kinter and Sherman, 2000). L'ordre de basicité en phase gazeuse est Lysine  $\approx$  Histidine < Arginine. La présence de résidus basiques supplémentaires dans la séquence du peptide peut potentiellement augmenter l'état de charge de l'ion



en électrospray et générer des ions trois voire quatre fois chargés. Comme ces résidus basiques ont tendance à repousser électrostatiquement le proton mobile, l'augmentation de l'état de charge tend à diminuer l'efficacité de fragmentation en CID.

L'efficacité de fragmentation CID va dépendre de l'énergie de collision, de la nature du gaz et de la pression de gaz dans la cellule de collision. En termes d'énergie de collision, on distinguera les CID de haute énergie (keV/charge) disponibles sur les TOF/TOF et les CID de faible énergie (moins de 100-200 eV/charge) disponibles sur les trappes, les triples quadripôles ou les Q-TOF. Dans la catégorie des CID de faible énergie, on peut encore subdiviser les spectromètres de masse selon le temps d'activation. Ainsi, on aura des vitesses de dissociation rapide de l'ordre  $10^4 \text{ s}^{-1}$  pour les triples quadripôles et les Q-TOF et des vitesses plus lente de l'ordre de 1 à  $100 \text{ s}^{-1}$  pour les trappes ioniques (Xia *et al.*, 2006).

Bien que très populaire, le mode CID présente quelques limitations :

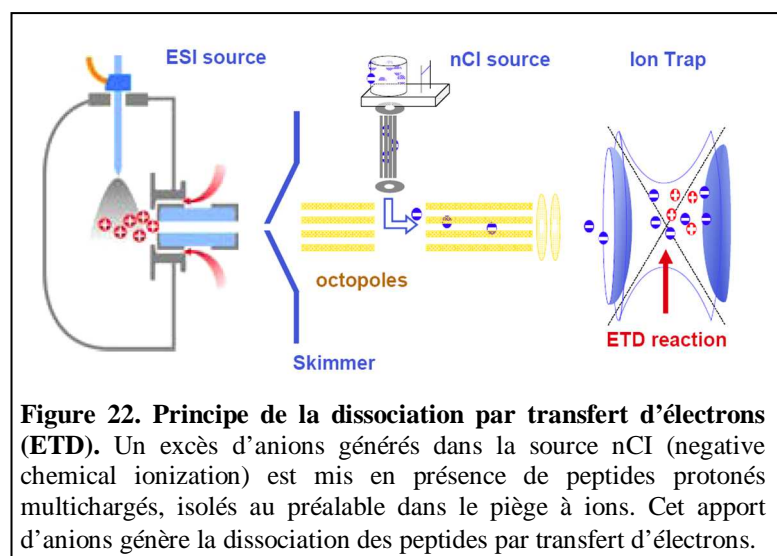
- le mode CID est peu efficace pour localiser les modifications labiles. Comme évoqué précédemment, dans le cas des phosphorylations, la perte d'un groupement phospho  $-\text{HPO}_3$  ou  $-\text{H}_3\text{PO}_4$  est un évènement de fragmentation favorisé. Cette perte de neutre domine souvent le spectre au détriment des clivages du squelette peptidique qui sont pourtant nécessaires à la détermination de la séquence. Afin d'avoir des informations sur la séquence du phosphopeptide dans le but de localiser le site de phosphorylation, une expérience de MS/MS/MS ou  $\text{MS}^3$  peut être réalisée pour l'analyse des phosphorylations sur les sérines et les thréonines. Cette expérience consiste à fragmenter l'ion produit généré par la perte de  $\text{H}_3\text{PO}_4$  en mode MS/MS. Dans le spectre  $\text{MS}^3$ , l'identification d'une déhydroalanine ( $M = 69 \text{ Da}$ ) ou d'un acide déhydroaminobutyrique ( $M = 83 \text{ Da}$ ), issus de la  $\beta$ -élimination en phase gazeuse de la phosphosérine ou de la phosphothréonine, respectivement, permet de localiser un site phosphorylé (Gruhler *et al.*, 2005).

- le mode CID est sensible à la séquence, ce qui se traduit par des abondances relatives très variables des ions fragments selon la séquence.

- les informations de séquences sont faibles dans le cas 1) de peptides non tryptiques tels que ceux issus d'une protéolyse AspN ou GluC, 2) de peptides multichargés ( $z > 2$ ).

#### 1.2.4.4 Dissociation par transfert d'électrons (ETD)

La dissociation par transfert d'électrons (ETD) est un mode de fragmentation analogue à la dissociation par capture d'électrons (ECD : electron capture dissociation) introduite par Zubarev *et al.* en 1998. L'ECD est mise en œuvre sur les spectromètres de masse FTICR. Cette technique favorise la fragmentation des peptides le long du squelette peptidique tout en préservant les chaînes



**Figure 22. Principe de la dissociation par transfert d'électrons (ETD).** Un excès d'anions générés dans la source nCI (negative chemical ionization) est mis en présence de peptides protonés multichargés, isolés au préalable dans le piège à ions. Cet apport d'anions génère la dissociation des peptides par transfert d'électrons.

latérales des acides aminés. Ainsi, les modifications labiles telles que la O-GlcNAc ou la phosphorylation sont conservées sur les résidus, ce qui facilite le séquençage et la localisation des sites de modifiés (Stensballe *et al.*, 2000). L'ETD a été initialement développé sur les trappes ioniques, qui sont des instruments plus répandus et moins onéreux que les FT-ICR, par Syka *et al.* 2004. L'ETD et l'ECD

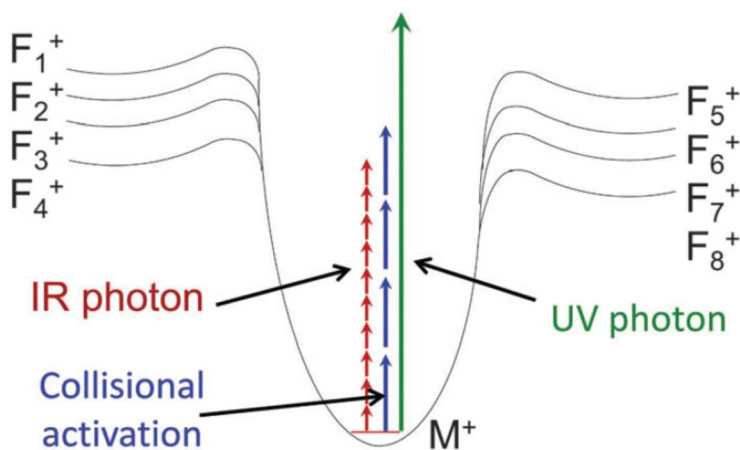
reposent sur un mécanisme « non ergodique », c'est-à-dire qui n'implique pas une redistribution de l'énergie vibrationnelle interne du peptide. En ETD, des anions radicalaires de fluoranthène ou d'azulène sont formés au niveau de la source d'ionisation chimique, un processus appelé nCI : negative chemical ionization. Ces "réactants" anioniques sont utilisés pour transférer un électron sur des peptides multiprotonés préalablement isolés dans la trappe ionique (Figure 22). Cette réaction, très exothermique, réduit la charge des peptides d'une unité et entraîne la dissociation de la liaison N-C $\alpha$  pour générer des ions fragments de type c et z. Aujourd'hui, des instruments commerciaux équipés d'un module ETD existent et sont de type trappe ionique, Orbitrap et Q-TOF. Contrairement au mode CID, l'ETD est particulièrement bien adaptée pour fragmenter les espèces multiprotonées telles que les peptides longs ou même les protéines entières. L'ETD est ainsi complémentaire à la CID.

#### 1.2.4.5 Photodissociation

Lors de la photodissociation, les ions accumulent de l'énergie *via* l'absorption d'un ou plusieurs photons, conduisant ainsi à leur fragmentation. Deux types de photodissociation sont principalement décrits. Il s'agit de la dissociation multiphotonique infrarouge (IRMPD : InfraRed MultiPhoton Dissociation) et de la photodissociation par rayonnement ultra-violet (UVPD : UltraViolet PhotoDissociation). Les photons sont émis par un laser dont la longueur d'onde se situe entre

l'infrarouge (10.6  $\mu\text{m}$ ) et l'ultra-violet (157 nm). Plus la longueur d'onde est faible et plus le photon contient de l'énergie, on parle de photon de faible ou de haute énergie. L'énergie apportée par photon est de l'ordre de 0.1 eV à 10.6  $\mu\text{m}$  (IR) et de 6.4 eV à 193 nm (UV). Ainsi, pour être fragmenté, un ion devra absorber des dizaines de photons IR alors qu'en général un seul photon UV suffit pour dissocier un ion (Figure 23).

**Figure 23. Diagramme d'énergie illustrant le dépôt d'énergie par activation collisionnelle ou par absorption de photons IR ou UV.**  $M^+$  représente un ion précurseur sélectionné et  $F_n^+$  représente les ions fragments avec différentes énergies d'activation. D'après (Brodbelt, 2014).



C'est en 1979 que la photodissociation a été couplée pour la première fois à un spectromètre de masse (Woodin *et al.* 1979). Depuis, elle a été implémentée sur de nombreux prototypes instrumentaux incluant les pièges à ions (trappe 3D, trappe linéaire), l'orbitrap, les FTICR et les temps de vol, mais aujourd'hui seuls des FTICR équipés de laser IR pour l'IRMPD sont disponibles commercialement (Brodbelt, 2014)

### 1.3 Identification et localisation de modifications par approche top-down

Une alternative à la protéolyse enzymatique suivie du séquençage MS/MS est l'approche top-down (Kelleher *et al.*, 1999), qui consiste à fragmenter la protéine entière directement dans le spectromètre de masse. L'avantage essentiel de l'approche top-down est la préservation du lien entre la masse de la protéine entière, issue du spectre MS, qui renseigne sur l'état de la protéine (intacte, tronquée, mutée, portant une PTM) et l'information structurale provenant du spectre MS/MS qui permet d'identifier la protéine et de localiser une modification éventuelle. En outre, l'approche top-down permet d'éviter d'éventuelles pertes de modifications qui pourraient être instables pendant la longue étape de protéolyse enzymatique. L'essor de cette approche depuis ces dernières années a été rendu possible

grâce à l'intégration des modes de fragmentation ECD-ETD dans les instruments haute résolution – haute précision de masse de type FT-ICR, Orbitrap ou HR-Q-TOF. La haute résolution ( $R = 50000$ ) permet d'observer la distribution isotopique de chaque ion fragment et ainsi d'attribuer les états de charges nécessaires pour l'annotation. La précision de masse élevée ( $< 5$  ppm) permet de limiter les ambiguïtés dans les annotations de séquence.

Cependant, bien que le séquençage top-down soit très prometteur, cette approche souffre de plusieurs limitations.

### 1.3.1 Les difficultés des approches top-down

Les approches top-down en protéomique se heurtent à plusieurs difficultés inhérentes à la séparation en chromatographie liquide d'une part et à l'analyse en spectrométrie de masse d'autre part.

#### 1.3.1.1 Séparation LC

En ce qui concerne la LC, l'efficacité de séparation reste encore insuffisante pour pouvoir séparer différentes protéoformes et encore moins des isomères, c'est-à-dire des protéines portant le même type et le même nombre de modification mais sur des sites différents. De plus, les protéines peuvent interagir de manière irréversible avec la phase stationnaire, ce qui provoque de faibles performances chromatographique en termes de résolution et de récupération de matériel. Ce phénomène est notamment observé pour les protéines de plus de 30 kDa (Li *et al.*, 2014).

#### 1.3.1.2 Sensibilité

La difficulté d'obtenir un bon rapport signal sur bruit pour les protéines entières, notamment celles de masse élevée, est l'un des principaux défis des stratégies top-down. En effet, les différentes protéoformes d'une protéine associées aux nombreux états de charge pour chaque espèce et à de multiples isotopes conduisent à disséminer le signal et ainsi diminuer le rapport signal sur bruit.

#### 1.3.1.3 Fragmentation

Le grand nombre d'ions fragments qui résultent de la dissociation de la protéine complique l'interprétation spectrale. En effet, d'une manière générale, les spectres obtenus sont très fournis en pics et le phénomène s'accroît lorsque la taille de la protéine augmente. Ainsi, il est très fréquent d'observer des chevauchements de distributions isotopiques entre les ions fragments. D'une manière

générale, cela a peu d'incidence sur l'identification de la protéine mais peut être une source d'erreur lors de la localisation de modification.

Par ailleurs, les ions fragments observés en CID ou ECD/ETD correspondent majoritairement aux extrémités terminales laissant ainsi le cœur de la séquence vide d'informations. Le phénomène augmente avec la taille de la protéine.

#### 1.3.1.4 Complexité du mélange de protéines

D'autres problèmes sont liés à la complexité du protéome, à la fois en termes de nombre de protéines et de gamme dynamique (par exemple 10 à 12 ordres de grandeur pour les protéines sériques).

L'ensemble de ces difficultés fait que la taille des protéines séquencées par top-down n'excède pas 30-40 kDa ce qui ne correspond qu'à une partie du protéome.

L'équipe de Kelleher a employé un système de séparation de protéines en 4 étapes : 1) isoélectrofocalisation, 2) gel d'électrophorèse SDS-PAGE avec récupération de fractions liquides (système GELFREE : Gel-Eluted Liquid FRaction Entrapment Electrophoresis), 3) nanoHPLC et 4) spectrométrie de masse. Grâce à ce système, ils ont identifié 645 phosphorylations, 538 acétylations de lysines et 158 méthylations à partir d'un protéome de cellules humaine HeLa (Tran *et al.*, 2011). Comparé aux approches bottom-up, qui sont capables aujourd'hui d'identifier et même de quantifier plus de 20000 phosphorylations, 15000 ubiquitinations et 3000 acétylations (Mertins *et al.*, 2013), le nombre de modification identifiée par top-down reste très faible.

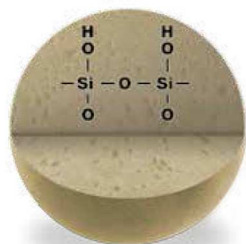
### 1.3.2 Les pistes de résolution de problèmes

#### 1.3.2.1 Les nouveaux supports chromatographiques

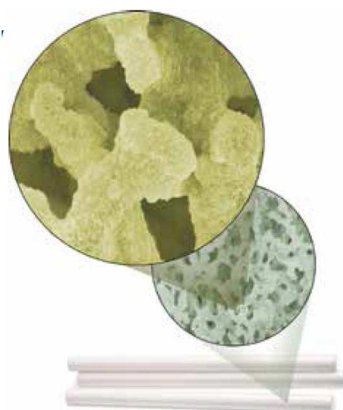
Les développements récents des colonnes HPLC, et plus particulièrement UHPLC, permettent une augmentation spectaculaire de l'efficacité et de la résolution ainsi qu'une diminution du temps d'analyse. Trois types de support sont disponibles commercialement. Il s'agit de particules totalement poreuses (support traditionnel), de particules superficiellement poreuses (SPPs, cores-shell, porous shell, fused-core) et de silice monolithique (Figure 24).

Figure 24. Les différents types de supports chromatographiques

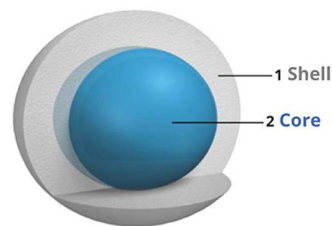
a) Silice totalement poreuse



b) Silice monolithique



c) Particules superficiellement poreuses (SPP, core-shell, porous shell, fused-core)



Aujourd'hui, certains fournisseurs proposent des matériaux entièrement poreux, de diamètre inférieur à  $2\ \mu\text{m}$  et à larges pores ( $300\ \text{\AA}$ ) pour la séparation de protéines en phase inverse. Ces colonnes sont disponibles avec des greffages C3, C4, C8, C18 et diphényle, offrant une large gamme de sélectivité.

Les particules SPPs sont constituées d'un noyau de silice solide, non poreux, entouré d'une fine enveloppe de silice poreuse dans laquelle peuvent diffuser les analytes. Cette enveloppe a des propriétés similaires à celles des particules totalement poreuses, utilisées classiquement en HPLC. Les particules SPPs permettent de réduire le chemin de diffusion et permettent un remplissage des colonnes plus uniforme et régulier, ce qui permet d'améliorer l'efficacité par rapport aux particules totalement poreuses traditionnelles ainsi que les temps d'analyses. Aujourd'hui, les colonnes contenant la dernière génération de particules SPPs de diamètre inférieur à  $3\ \mu\text{m}$  rivalisent avec des colonnes contenant des particules totalement poreuses de diamètre inférieur à  $2\ \mu\text{m}$  mais avec une pression de fonctionnement deux fois plus faible. Pour améliorer l'efficacité de la séparation, des colonnes SPPs de diamètre inférieur à  $2\ \mu\text{m}$  ( $1.3$  à  $1.7\ \mu\text{m}$ ) ont récemment été commercialisées (Fekete and Guillaume, 2014).

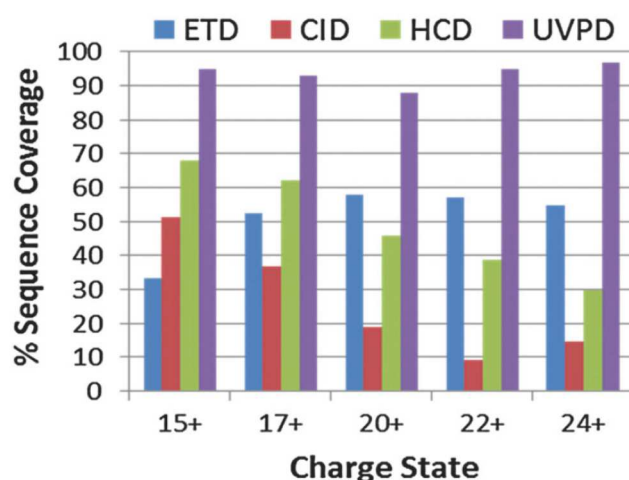
Les colonnes monolithiques contiennent des macropores et des mésopores qui génèrent une contre-pression très faible. Ces colonnes sont bien adaptées pour l'injection directe d'échantillons complexes tels que les extraits de plasma.

En plus des développements pour la chromatographie en phase inverse, des colonnes contenant des particules de diamètre inférieur à  $2\ \mu\text{m}$  peuvent également être utilisées avec d'autres modes de chromatographie, comme la chromatographie liquide d'interaction hydrophile (HILIC : Hydrophilic Interaction LIquid Chromatography), la chromatographie d'exclusion stérique (SEC : Size-Exclusion Chromatography) ou la chromatographie d'échange d'ions (IEX : Ion-EXchange chromatography) (Fekete and Guillaume, 2014).

### 1.3.2.2 La photodissociation

Comparé à la CID, l'ETD ou l'HCD, la photodissociation par rayonnement ultra-violet permet d'obtenir un plus grand nombre d'ions fragments et une meilleure couverture de séquence sur l'ensemble de la protéine (Figure 25). Ce mode de fragmentation semble très prometteur pour la fragmentation de protéines entières mais reste cantonné, à ce jour, à des instruments non commerciaux.

Figure 25. Comparaison de la couverture de séquence obtenue pour cinq états de charge de la myoglobine fragmentés par ETD, CID, HCD et UVPD dans un Orbitrap Elite (Brodbeck, 2014).



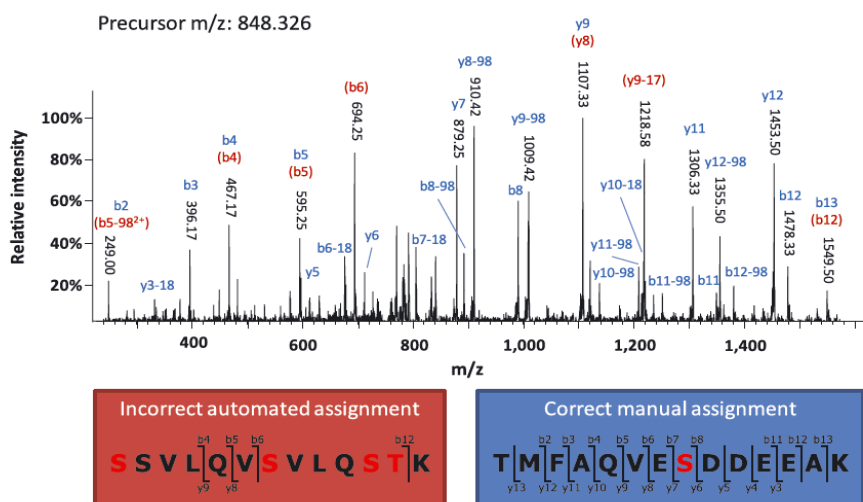
L'objectif annoncé par le groupe de Kelleher d'identifier toutes les protéoformes de protéines par top-down n'a pas été atteint pour l'instant. En effet, le fractionnement préalable à la spectrométrie de masse est encore insuffisant pour pouvoir séparer des protéoformes et encore moins des isomères. Des isomères ne peuvent pas non plus être séparés selon leur rapport  $m/z$  puisqu'il est identique. L'implémentation de la mobilité ionique dans les spectromètres de masse, qui permet de distinguer des isomères de petites molécules, pourrait être envisagée mais n'est pas encore assez résolutive pour séparer des isomères protéiques. Lors de la dissociation, ces isomères, confondus dans le même  $m/z$ , génèrent des spectres de fragmentation encore plus compliqués et les logiciels de retraitement de données ont parfois des difficultés à fournir un résultat fiable et cohérent.

## 1.4 Traitement et Interprétation des spectres de masses (Bioinformatique)

Les logiciels sont omniprésents en spectrométrie de masse. Ils permettent d'abord d'acquérir un spectre de masse mais surtout de le retraiter, de l'interpréter et d'extraire l'information d'intérêt.

Avec les approches à haut-débit qui génèrent des milliers de spectres, il est devenu impossible de vérifier les spectres manuellement. Par conséquent, en dépit de l'utilisation d'algorithmes de plus en plus performants, des erreurs d'annotations peuvent se produire (Figure 26) (Engholm-Keller and Larsen, 2013).

**Figure 26. Exemple d'erreur d'annotation automatique de spectre MS/MS d'un peptide phosphorylé (Engholm-Keller and Larsen, 2013).** Le spectre a été annoté par le logiciel à la séquence pSSVLQVpSVLQpSpTK de la 2',5'-phosphodiesterase 12 (IP|00174390.3) avec seulement quatre ions b et deux ions y (en rouge). L'annotation manuelle du spectre permet d'annoter correctement la séquence à TMoxFAQVEpSDDEEAK avec 10 ions b et 11 ions y (en bleu). Les lysines sont 13C6-Lys car le peptide sélectionné pour la fragmentation est issu d'un marquage SILAC (13C6-Lys) d'un lysat de cellule HeLa.



L'identification des peptides modifiés est plus difficile que celle des peptides non modifiés parce que le moteur de recherche, au lieu de simplement comparer un ensemble expérimental de fragments de peptide avec un ensemble de fragments produit *in silico*, doit introduire un jeu d'incrément de masse possibles correspondant à différentes modifications potentielles. Lorsque plusieurs types et/ou sites de modifications concomitantes sont autorisés, cela conduit à une explosion combinatoire de l'espace de recherche qui peut amener à diminuer la confiance statistique dans les résultats de recherche. En outre, localiser le résidu qui porte la modification nécessite d'observer dans le spectre MS/MS des ions fragments diagnostics qui encadrent le résidu modifié. Dans le cas des isomères (peptides portant la même modification mais sur des sites différents) qui sont souvent co-élués et donc co-fragmentés, les logiciels ont encore plus de difficultés. Bien que plusieurs algorithmes dédiés à la localisation de sites modifiés, tels que Ascore (Beausoleil *et al.*, 2006) ou PTM score (Olsen *et al.*, 2006) pour la phosphorylation, aient été développés pour fournir une probabilité pour chaque site modifiable, la vérification manuelle des spectres de peptides importants est très conseillé. La vérification ultime est de synthétiser le peptide modifié au niveau de la ou les position(s) identifiée(s) et de comparer le profil de fragmentation.



## 2. De la quantification de protéines à la quantification de leurs modifications

Cataloguer les différentes modifications post-traductionnelles en termes d'identification et de localisation est une première étape indispensable pour étudier un processus biologique. Les informations obtenues permettent notamment de faire une mutation ponctuelle en lieu et place de la modification découverte et de confirmer ou non l'importance du résidu cible pour l'activité de la protéine étudiée. Toutefois, savoir qu'une modification est responsable d'une fonction biologique, normale ou aberrante, ne suffit pas à déchiffrer le mécanisme d'un processus biologique car dans un système vivant, l'abondance des protéines et de leurs modifications post-traductionnelles subit des changements dynamiques au cours du cycle cellulaire, de la différenciation cellulaire et en réponse à des stimuli intra- et extracellulaires. Ainsi, la quantification de protéines et de leurs modifications présente un grand intérêt dans i) la compréhension des processus biologiques mais également dans ii) la recherche de biomarqueurs pour identifier et quantifier une protéine dans un biofluide (urine, plasma/sérum) et iii) la découverte et la mise au point de médicaments pour identifier et quantifier les cibles secondaires ainsi que les résidus modifiés. Dans cette optique, depuis une dizaine d'année, de nombreux efforts ont été engagés dans le développement de stratégies pour la quantification de protéine et de leurs modifications. Les techniques existantes comprennent le dosage immunologique ELISA, l'électrophorèse bidimensionnelle avec colorants spécifiques, le Western blot et la spectrométrie de masse. La quantification par spectrométrie de masse est de plus en plus utilisée en raison de sa grande sensibilité et spécificité ainsi que pour sa capacité à quantifier plusieurs protéines simultanément (dosage multiplexé).

### 2.1 Bottom-up vs top-down

Comme pour leur identification, les protéines et/ou leurs modifications peuvent être quantifiées par bottom-up ou top-down. La quantification bottom-up consiste à quantifier un ou plusieurs peptides issu(s) d'une étape de protéolyse enzymatique. La quantité des peptides mesurée est ensuite directement reliée à la quantité de protéine. La quantification top-down consiste à quantifier directement la protéine entière. Du fait de la complexité accrue des échantillons biologiques (tant en termes du nombre de protéines que de la gamme dynamique de l'expression des protéines), la quasi-totalité des méthodes de quantification développées utilisent le couplage LC-MS. Malgré quelques exemples récents de quantification top-down (Hung and Tholey, 2012; Ntai *et al.*, 2014), depuis une dizaine d'années, la grande majorité des développements méthodologiques ont été réalisés pour la quantification bottom-up parce que les peptides sont plus faciles à séparer et à analyser par LC-MS

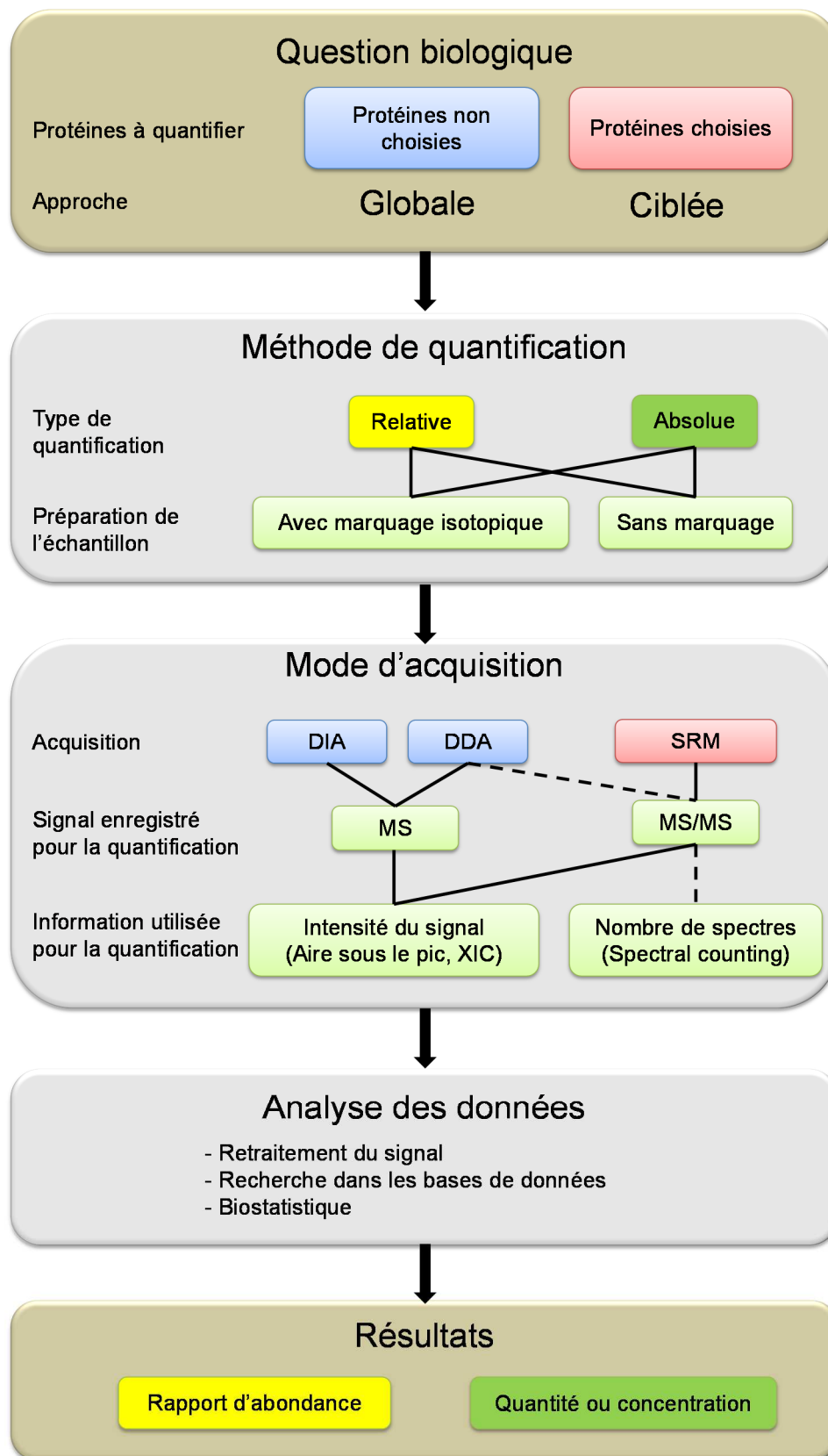
que les protéines et que la MS est beaucoup plus sensible pour les ions de basse masse. Ainsi, les stratégies de quantifications présentées ci-après concernent essentiellement la quantification bottom-up.

## 2.2 Quantification de protéine

Les stratégies de quantification bottom-up sont multiples et varient selon la question biologique posée, le nombre d'échantillons à quantifier, le type de quantification désirée (absolue ou relative), l'utilisation ou non d'un marquage isotopique stable ou encore le mode d'acquisition du spectromètre de masse (Figure 27).

Aujourd'hui les deux principales approches pour la quantification à partir de données LC-MS sont 1) l'extraction des intensités LC-MS des ions précurseurs lors de l'acquisition classique en mode DDA au cours de laquelle les ions précurseurs les plus intenses sont fragmentés lorsqu'ils sont élués de la colonne chromatographique et 2) l'extraction des intensités LC-MS/MS des ions fragments lors d'acquisitions en mode SRM dans laquelle une liste prédéfinie d'ions précurseurs et d'ions fragments sont spécifiquement analysés. Ces modes DDA et SRM seront détaillés au paragraphe 2.2.3.

**Figure 27. Stratégie globale pour la quantification de protéine par spectrométrie de masse.** Sont représentées les approches les plus répandues avec classiquement une analyse globale pour les protéines non choisies et une approche ciblée pour les protéines choisies mais ces dernières peuvent également être analysées en approche globale. Toutes les possibilités ne sont pas illustrées, notamment la combinaison « approche globale-MS/MS-Intensité du signal » utilisée dans les méthodes de quantification avec marquage isotopique chimique de type iTRAQ ou TMT.



## 2.2.1 La question biologique

La question biologique est centrale en quantification car c'est elle qui va orienter l'analyse vers telle ou telle stratégie de quantification. Selon que les protéines à quantifier sont choisies ou non, deux stratégies sont possibles. Il s'agit de l'approche globale et de l'approche ciblée. Le choix entre ces deux approches va dépendre de l'état de connaissance sur le processus biologique, le biomarqueur ou le médicament étudié.

### 2.2.1.1 Approche globale

L'approche globale a été la méthode la plus utilisée et a généré la grande majorité des données protéomiques disponibles actuellement. Elle permet la quantification de toutes les protéines détectables sans sélection préalable. La quantification globale est principalement utilisée pour identifier les protéines dont les quantités varient entre deux ou trois conditions expérimentales. Le plus souvent, elle correspond aux approches de protéomique « shotgun » qui présentent l'intérêt d'identifier et de quantifier un très grand nombre de protéines simultanément et avec peu de développement méthodologique. La quantification globale permet d'avoir un premier aperçu des variations relatives des protéines ou de leurs modifications mais c'est une technique de quantification peu reproductible, notamment dans les échantillons complexes, du fait du mode d'acquisition semi-aléatoire des spectres MS/MS (DDA : data-dependent acquisition).

### 2.2.1.2 Approche ciblée

L'approche ciblée est très à la mode aujourd'hui. C'est la méthode de choix pour une quantification sensible en milieu complexe de protéines choisies *a priori* (souvent à partir de données issues de l'approche globale). Dans la quantification ciblée, les protéines à quantifier sont connues, on ne peut donc pas détecter de nouvelles protéines avec cette approche. C'est une approche appropriée pour tester des hypothèses de mécanisme biologique, pour générer de manière reproductible des ensembles complets de données quantitatives pour la biologie des systèmes ou valider des biomarqueurs en mesurant des changements d'abondance dans les échantillons cliniques. Le nombre d'échantillon peut être élevé mais le nombre de protéines ou de modifications quantifiées est limité.

## 2.2.2 Les différentes méthodes de quantification par spectrométrie de masse

La quantification de protéines par spectrométrie de masse peut être relative ou absolue. La quantification relative permet de comparer l'abondance d'une protéine entre plusieurs échantillons et

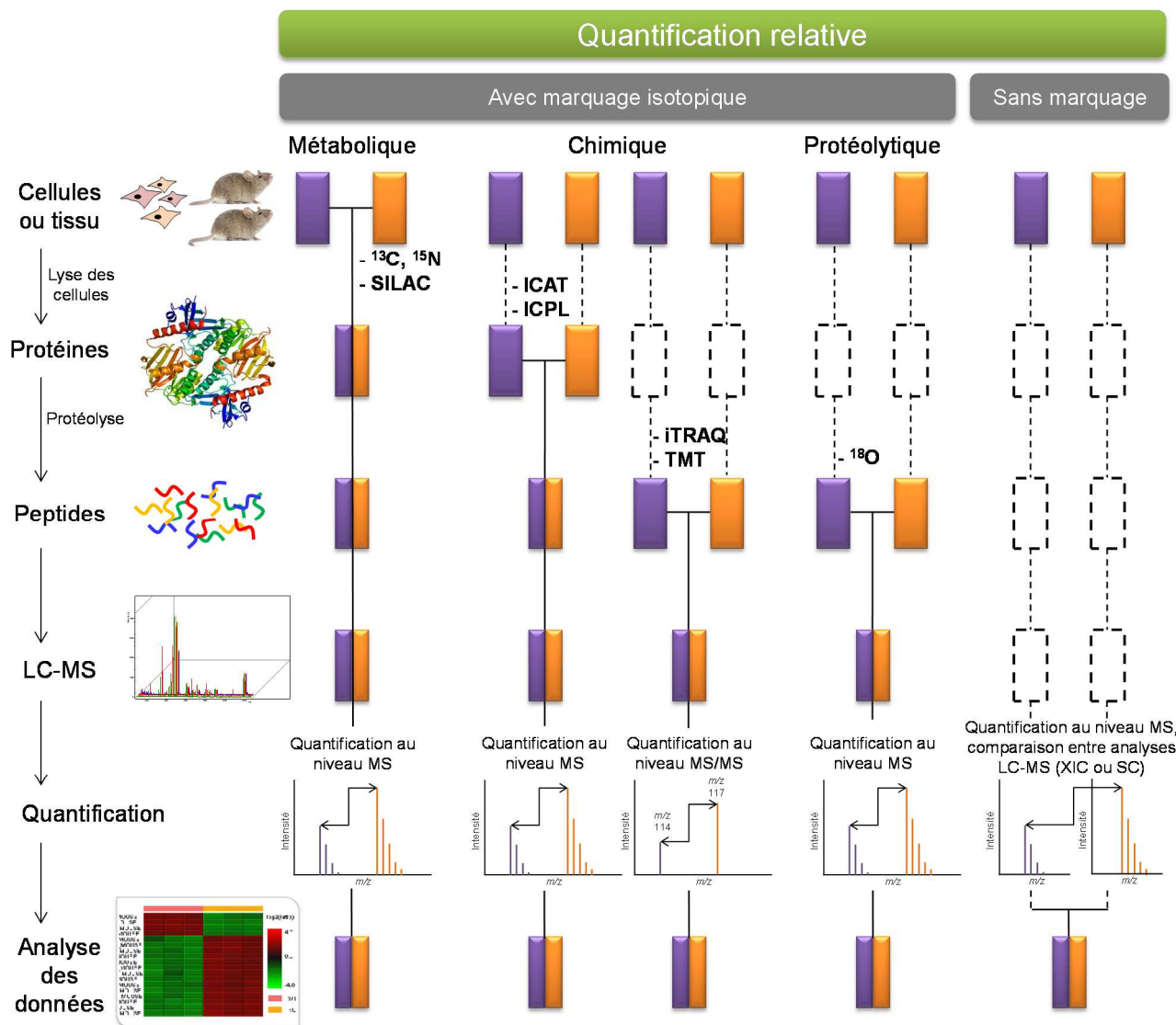
d'en déduire des variations du niveau d'expression et/ou de dégradation. La quantification absolue permet, grâce à l'utilisation d'un standard de concentration connue, de déterminer la quantité et la concentration d'une protéine dans un échantillon. Les méthodes associées à ces deux types de quantification sont présentées ci-après.

#### 2.2.2.1 Les méthodes pour une quantification relative

La quantification relative d'une protéine consiste à mesurer sa variation entre deux ou plusieurs conditions ; sa concentration exacte n'est pas connue. La variation est exprimée sous forme de rapport, par exemple 1:2:4 pour trois conditions A:B:C, indiquant que la protéine est deux fois plus concentrée dans la condition B que dans la condition A et qu'elle est quatre fois plus concentrée dans la condition C. La quantification relative est la plus utilisée car tout en répondant à la question biologique initiale elle est moins contraignante et souvent moins coûteuse que la quantification absolue.

Il y a deux approches principales pour effectuer une quantification relative en spectrométrie de masse : les méthodes basées sur le marquage isotopique stable et les méthodes sans marquage (label-free) (Figure 28).

**Figure 28. Les différentes méthodes de quantification relative par spectrométrie de masse. (Adapté de Bantscheff et al., 2012).** Les rectangles violet et orange représentent deux conditions expérimentales. Les lignes horizontales indiquent le mélange des échantillons. Les lignes en pointillées indiquent les étapes au cours desquelles il peut y avoir des variations expérimentales et donc des erreurs de quantification. Dans le marquage métabolique comme SILAC, les cellules sont cultivées en présence d'isotopes stables, ce qui permet de mélanger les échantillons à un stade très précoce dans le protocole et donc de réduire les biais dues aux variations liées aux manipulations des échantillons. Le marquage chimique tel que l'iTRAQ est généralement effectué sur des peptides après protéolyse enzymatique. Contrairement aux autres méthodes quantitatives, la quantification iTRAQ est effectuée au niveau MS/MS par des ions rapporteurs de faibles masses. Les stratégies sans marquage utilisent la comparaison des abondances peptidiques entre analyses LC-MS. Cette abondance est traduite soit sous forme d'intensité du signal avec la génération d'un chromatogramme d'ion extrait (XIC) soit sous forme de décompte de spectres MS/MS (« spectral counting »).



### a) Le marquage isotopique

Le marquage isotopique consiste à incorporer des isotopes stables dans les protéines ou les peptides avant l'analyse en spectrométrie de masse. Classiquement, deux échantillons sont marqués avec des isotopes différents (isotopes légers :  $^{12}\text{C}$ ,  $^{14}\text{N}$  ou isotopes lourds  $^{13}\text{C}$ ,  $^{15}\text{N}$ ), puis ils sont mélangés et analysés simultanément. Chaque peptide est alors observé sous forme de deux pics sur le spectre et les abondances relatives des pics reflètent la quantité des peptides dans chaque échantillon (Figure 28).

### Le marquage métabolique

Les isotopes stables peuvent être incorporés *in vivo* par marquage métabolique des protéines dans une culture cellulaire en utilisant des isotopes  $^{12}\text{C}/^{13}\text{C}$ ,  $^{14}\text{N}/^{15}\text{N}$  (Oda *et al.*, 1999) ou des acides aminés isotopiques tels que la lysine ( $^{13}\text{C}_6$ ,  $^{15}\text{N}_2$ ) et l'arginine ( $^{13}\text{C}_6$ ,  $^{15}\text{N}_4$ ) (SILAC : stable isotope labeling by amino acids in cell culture) (Ong *et al.*, 2002).

Le marquage  $^{15}\text{N}/^{13}\text{C}$  permet un marquage uniforme de chaque acide aminé contrairement au marquage SILAC où seules les lysines, arginines, leucine et proline peuvent être marquées. Il confère ainsi une liberté dans le choix de la stratégie de protéolyse et assure que chaque peptide ait une différence de masse suffisante par rapport à son homologue non marqué. Il est préférable que cette différence de masse entre le marqué et le non marqué soit supérieure à la largeur de la distribution isotopique afin d'éviter la superposition des deux distributions et faciliter l'interprétation des résultats.

D'une manière générale, le marquage métabolique est la méthode de quantification la plus précise car les échantillons marqués et non marqués sont mélangés très tôt dans le protocole de quantification, ce qui permet de corriger la majorité des pertes et des variations au cours du protocole expérimental. Cependant, cette méthode est réservée aux organismes et tissus dont la culture cellulaire est possible. Pour les échantillons cliniques (par exemple les fluides biologiques), des méthodes de marquage *in vitro* sont disponibles.

### Le marquage chimique

Les isotopes stables peuvent être incorporés *in vitro* par dérivation chimique de groupes fonctionnels tels que les cystéines (ICAT : isotope-coded affinity tags) (Gygi *et al.*, 1999), les amines primaires (iTRAQ : isobaric tags for relative and absolute quantitation) (Ross *et al.*, 2004), (TMT : tandem mass tag) (Thompson *et al.*, 2003), (ICPL : Isotope-coded protein label) (Schmidt *et al.*, 2005) ou l'extrémité C-terminale (He *et al.*, 2004). Le marquage chimique peut être réalisé au niveau des protéines (ICPL, ICAT) ou des peptides (iTRAQ, TMT). L'ajout d'une étape de réaction chimique, outre le surcoût associé, conduit à un allongement du temps de préparation des échantillons, introduit un facteur de variabilité du résultat et peut conduire à des réactions secondaires imprévues.

### Le marquage protéolytique

Les isotopes stables peuvent également être incorporés *in vitro* en utilisant de l'eau lourde  $\text{H}_2^{18}\text{O}$  lors de la protéolyse par la trypsine (Mirgorodskaya *et al.*, 2000). Cette approche conduit à l'incorporation de deux atomes  $^{18}\text{O}$  par peptide (excepté le peptide C-terminal) soit un incrément de masse de 4 Da. Cette différence de masse de 4 Da est souvent insuffisante pour séparer les

distributions isotopiques des peptides marqués ( $^{18}\text{O}$ ) et non marqués ( $^{16}\text{O}$ ). Ces distributions partiellement chevauchantes nécessitent l'utilisation d'équation et de logiciel adaptés pour attribuer le signal respectif au peptide  $^{16}\text{O}$  et au peptide  $^{18}\text{O}$  avant la quantification (Yao *et al.*, 2001).

#### *b) La quantification sans marquage (« label free »)*

Les approches sans marquage sont souvent utilisées quand le marquage isotopique est impossible (par exemple dans les études sur les animaux) ou que le coût de l'expérience est prohibitif (par exemple dans les études sur les biomarqueurs dans lesquelles un grand nombre d'échantillons doivent être analysés).

Dans l'approche sans marquage, les différents échantillons sont analysés séparément. Il est donc essentiel que chaque étape du protocole soit rigoureusement optimisée pour être reproductible. Afin de diminuer les effets de ces variations, des peptides références sont souvent utilisés pour corriger les variations du volume d'injection en LC et de sensibilité de l'instrument.

La quantification sans marquage repose sur le système LC-MS et peut être réalisée de deux manières : le calcul du nombre d'évènements de fragmentation par peptide (spectral counting) ou la mesure de l'intensité MS des peptides.

#### Le « spectral counting »

La quantification par « spectral counting » est une méthode simple qui consiste à utiliser le nombre d'évènements de fragmentation d'un peptide pour estimer la quantité d'une protéine. Ainsi pour un peptide abondant dans la condition A, le nombre de spectre MS/MS sera plus élevé que dans la condition B où le même peptide sera en plus faible quantité. Le spectral-counting est typiquement utilisé pour la quantification relative et sur une petite gamme dynamique (Schilling *et al.*, 2012). Il n'est pas adapté aux variations d'abondance faible. Le spectral-counting nécessite quelques paramétrages lors de l'acquisition, notamment de ne pas utiliser de liste d'inclusion (liste prédéfinie de  $m/z$  qui seront fragmentés) ou d'exclusion dynamique (critère selon lequel un peptide, qui a été fragmenté une fois par le spectromètre de masse, ne sera pas fragmenté de nouveau). Avec l'avènement de la HRMS, l'approche par mesure de l'intensité MS est devenue nettement plus précise et est donc souvent préférée à l'approche par « spectral counting ».

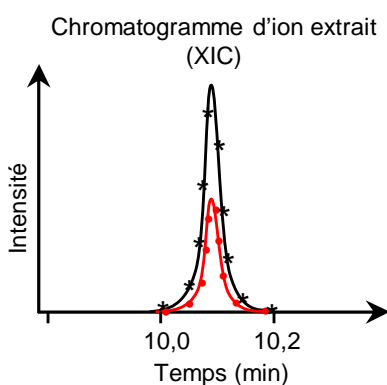
#### La mesure de l'intensité MS

La quantification par mesure de l'intensité MS est analogue à la quantification UV en HPLC basée sur la mesure de l'aire d'un pic. Lorsqu'un peptide est élué de la colonne, il va être ionisé et détecté par le spectromètre de masse. Le profil d'éluion de ce peptide peut être illustré par le courant ionique



mesuré en fonction du temps (Figure 29). L'aire sous la courbe correspond au chromatogramme de l'ion extrait (XIC pour eXtracted Ion Chromatogram) et est proportionnelle à la quantité de peptide. Plusieurs milliers de XICs correspondant à des peptides différents peuvent être générés pour chaque analyse. Ainsi, en comparant les XICs de deux échantillons on peut identifier les peptides et donc les protéines dont la quantité varie. En utilisant une courbe étalon, une quantité absolue peut être calculée.

La mesure de l'intensité MS nécessite une reproductibilité irréprochable en LC et une procédure de normalisation des variations entre les analyses par ajout d'un standard interne (Steen *et al.*, 2005). Classiquement il faut au minimum 10 points pour dessiner correctement un pic chromatographique.



**Figure 29. Quantification sans marquage.** Lorsqu'un peptide A est élué de la colonne, son signal est mesuré plusieurs fois, formant ainsi la courbe noire. L'aire sous la courbe noire est le chromatogramme de l'ion extrait (XIC). La courbe rouge correspond au même peptide A d'un autre échantillon. La comparaison de l'aire des courbes noire et rouge indique un rapport de 50% entre les deux échantillons.

La quantification "label free" apparaît comme la méthode de quantification la plus simple, permettant la comparaison directe de signaux MS entre plusieurs échantillons, applicable à plusieurs sources de matériel biologique et peu onéreuse.

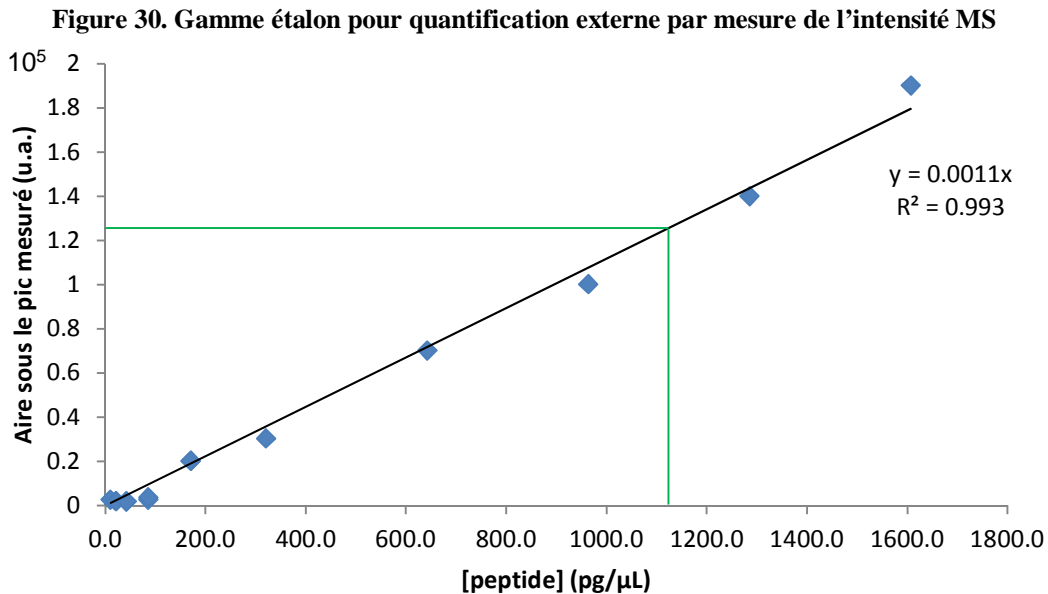
#### 2.2.2.2 Les méthodes pour une quantification absolue

La quantification absolue est capable de i) déterminer le nombre de copies des protéines dans une cellule, ce qui constitue une information précieuse pour la modélisation des processus biologiques (Bennett *et al.*, 2010), ii) mesurer précisément la quantité d'un biomarqueur ou d'une toxine dans un fluide biologique, l'eau ou la nourriture (Dupuis *et al.*, 2008), ou encore de iii) déterminer la stœchiométrie de complexes protéiques (Schmidt *et al.*, 2010).

La quantité absolue d'une protéine dans un échantillon peut être déterminée soit par des approches basées sur le décompte de spectres telle que emPAI (Ishihama, 2005) ou APEX (Lu *et al.*, 2007), soit par la comparaison de l'intensité de son signal à celui d'un standard externe ou interne. Nous ne parlerons ici que des approches utilisant un standard externe ou interne.

a) *Étalonnage externe*

La quantification absolue par étalonnage externe consiste à réaliser une courbe d'étalonnage représentant le signal mesuré d'un standard en fonction de sa concentration (Figure 30).



L'échantillon inconnu est ensuite analysé et sa concentration est déterminée en utilisant l'équation de la courbe d'étalonnage. Dans cet exemple :

$$\text{concentration de l'analyte} = \frac{\text{Aire mesuré}}{0.0011}$$

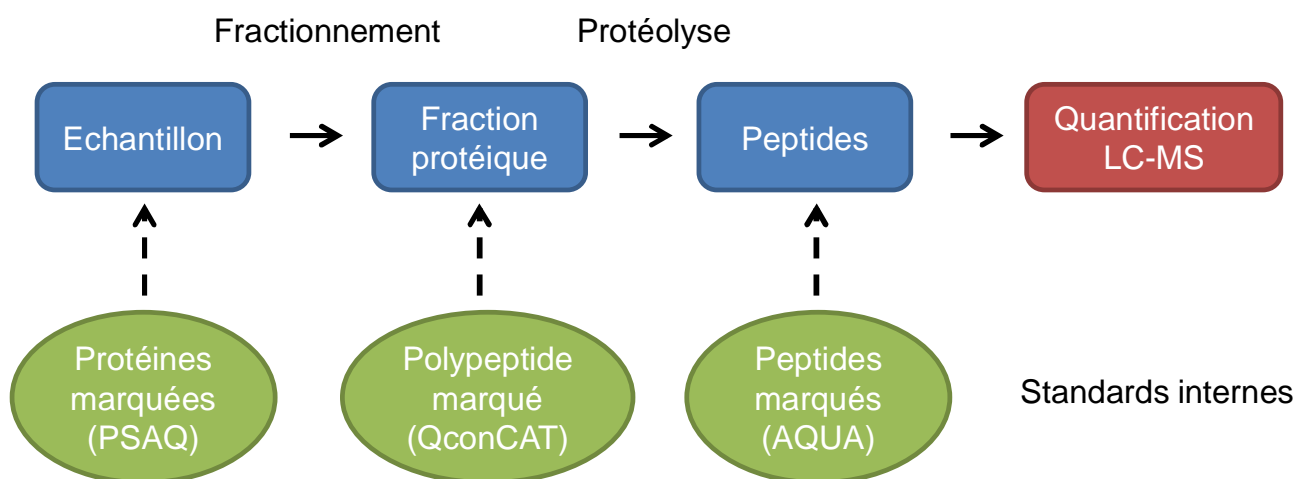
b) *Étalonnage interne*

La détermination de la quantité absolue d'une protéine *via* l'étalonnage interne consiste à comparer son signal à celui d'un standard interne ajouté en quantité connue dans l'échantillon. Afin de corriger les sources de variation des différentes étapes du protocole, le standard interne doit à la fois avoir des propriétés physico-chimiques très proches de l'analyte à quantifier (même temps de rétention chromatographique, même efficacité d'ionisation, même profil de fragmentation MS/MS) et être distinguable en spectrométrie de masse, c'est-à-dire avoir une masse différente de l'analyte. Le meilleur standard interne pour la quantification en spectrométrie de masse est l'analyte marqué avec des isotopes stables <sup>13</sup>C et/ou <sup>15</sup>N. Le marquage au deutérium a également été utilisé mais des décalages dans les temps de rétention LC ont été observés, ce qui n'est pas recherché (Zhang *et al.*, 2001). L'équation générale pour déterminer la concentration absolue de l'analyte est la suivante:

$$\begin{aligned} \text{concentration de l'analyte} \\ = \text{concentration du standard interne} \times \frac{\text{signal de l'analyte}}{\text{signal du standard interne}} \end{aligned}$$

Trois méthodes de référence sont utilisées aujourd'hui suivant que le standard interne est introduit avant une étape de fractionnement, avant la protéolyse ou après la protéolyse : PSAQ, QconCAT et AQUA, respectivement (Figure 31).

**Figure 31. Stratégies utilisées pour la quantification absolue d'après (Domon and Aebersold, 2010).** Trois types de standard interne peuvent être utilisés. Les protéines marquées (PSAQ : Protein Standard for Absolute Quantification) sont ajoutées avant le fractionnement de l'échantillon ; les polypeptides artificiels marqués correspondant à la concaténation de différents peptides marqués à quantifier (QconCAT) sont introduits avant la protéolyse ; les peptides marqués (AQUA) sont ajoutés après la protéolyse.



### AQUA

La stratégie de quantification absolue AQUA (Absolute Quantitation) utilise des peptides synthétisés chimiquement et marqués avec des isotopes stables (Gerber *et al.*, 2003). Ces peptides AQUA, également appelés peptides SIL (stable isotope labeled), sont ajoutés en quantité connue dans les échantillons après l'étape de protéolyse. Classiquement, le marquage du peptide est réalisé par l'incorporation d'un acide aminé marqué tel que la leucine, la lysine ou l'arginine.

Les peptides AQUA sont classiquement générés par synthèse chimique, purifiés par HPLC et quantifiés par une analyse des acides aminés. Les peptides AQUA sont généralement fournis sous forme lyophilisée ce qui nécessite une étape de solubilisation avant utilisation. Il faut noter que cette étape peut affecter les résultats expérimentaux car il n'est pas facile de déterminer si le peptide contenu dans le flacon a été entièrement solubilisé.

Bien que cette méthode AQUA soit très utilisée, notamment grâce à la disponibilité de nombreux peptides AQUA, le standard interne est ajouté assez tardivement dans le protocole de quantification. Ceci peut entraîner des biais de quantification si les variations liées au fractionnement de l'échantillon, souvent nécessaire lors de la manipulation d'échantillon biologique, et à l'étape de

protéolyse ne sont pas évaluées et prises en compte. Par ailleurs, le coût des peptides AQUA n'est pas négligeable (~160€ les 100 pmol), notamment lorsque l'on souhaite quantifier plusieurs protéines.

### QconCAT

Dans l'optique de quantifier plusieurs protéines en même temps (quantification multiplexée) et de minimiser les coûts, l'approche QconCAT a été développée (Beynon *et al.*, 2005). Cette approche consiste à produire un polypeptide artificiel marqué contenant les différents peptides tryptiques standards marqués de chaque protéine à quantifier. Le polypeptide est introduit en quantité connue à l'étape de protéolyse tryptique et les différents peptides sont libérés au cours de la protéolyse. Bien qu'il soit présent dans le même tube que l'analyte lors de la protéolyse, le standard QconCAT ne permet pas de corriger de manière fiable un rendement de protéolyse inférieur à 100% car le rendement de protéolyse dépend de la structure tridimensionnelle d'une protéine et que le standard interne QconCAT et les protéines à quantifier n'ont pas la même conformation. L'utilisation d'une étape de dénaturation en amont de la protéolyse tend à réduire ses différences mais le problème des absences de coupure (MC : missed cleavage) liées à des sites de coupures trop proches subsiste.

### PSAQ

Afin de surmonter les limitations des approches AQUA et QconCAT, l'équipe de Jérôme Garin a développé la stratégie PSAQ (Protein Standard Absolute Quantification) (Brun *et al.*, 2007). Cette approche est basée sur l'utilisation d'une protéine entière marquée uniformément avec l'isotope  $^{15}\text{N}$  ou avec des acides aminés isotopiques stables. Cette protéine marquée est identique à la protéine à quantifier en termes de séquence et de conformation (sous réserve de la présence/absence de PTM sur la protéine non marquée). Après purification, le taux d'incorporation est contrôlé, la concentration est déterminée précisément par analyse d'acides aminés et la protéine est ajoutée en quantité connue à l'échantillon avant un fractionnement éventuel. Le fait que ce standard interne soit introduit tôt dans le protocole de quantification permet de corriger les variations liées aux différentes étapes en amont de l'analyse MS : pertes, rendement d'enrichissement et rendement de protéolyse.

Sous réserve de pouvoir produire une protéine soluble et suffisamment pure La stratégie PSAQ constitue actuellement le meilleur standard interne pour une quantification absolue d'une protéine. Les limitations de cette approche sont qu'elle n'est pas très adaptée à la quantification multiplexée qui nécessiterait la production de plusieurs protéines marquées très coûteuses et à l'obtention de protéine modifiée.

Le Tableau 4 rassemble les avantages et limitations des trois standards internes AQUA, QconQAT et PSAQ.

**Tableau 4. Avantages et limitations des standards internes utilisés en quantification absolue. (Adapté de Boja and Rodriguez, 2012)**

	Avantages	Limitations
AQUA	<ul style="list-style-type: none"> <li>- De nombreux peptides commercialement disponibles.</li> <li>- Correction des variations LC-MS.</li> <li>- Capacité de multiplexage mais pour un nombre limité de protéines</li> <li>- Permet l'introduction de certaine modification dans la séquence peptidique telle que la phosphorylation, l'ubiquitination, la méthylation et l'acétylation</li> </ul>	<ul style="list-style-type: none"> <li>- Les peptides synthétiques sont couteux</li> <li>- Selon la séquence, la synthèse peut prendre du temps voire ne pas aboutir</li> <li>- La solubilisation du peptide lyophilisé n'est pas toujours de 100%</li> <li>- Les peptides ne sont pas compatibles avec un fractionnement de l'échantillon au niveau protéique</li> <li>- Pas de correction des rendements de protéolyses, ce qui affecte la précision de la quantification</li> </ul>
QconQAT	<ul style="list-style-type: none"> <li>- Permet de quantifier plusieurs protéines simultanément (quantification multiplexée) à moindre coût et de manière plus simple que les peptides AQUA.</li> <li>- Correction des variations LC-MS voire même de la protéolyse si les rendements de protéolyses sont identiques entre les protéines à quantifier et le standard QconQAT</li> </ul>	<ul style="list-style-type: none"> <li>- Plusieurs constructions QconQAT peuvent être nécessaires pour avoir une production d'un polypeptide artificiel soluble</li> <li>- Il est difficile de produire des polypeptides contenant plus que 30-40 peptides.</li> <li>- La génération d'un mélange équimolaire de peptides nécessite une protéolyse trypsique complète, ce qui nécessite une optimisation des conditions de protéolyse</li> <li>- L'introduction de modification (phosphorylation ou glycosylation) est possible mais nécessite une étape enzymatique additionnelle</li> </ul>
PSAQ	<ul style="list-style-type: none"> <li>- Correction des variations lors des étapes de fractionnement, protéolyse et LC-MS</li> <li>- Permet d'utiliser facilement et de manière fiable plusieurs peptides pour quantifier une protéine.</li> <li>- Standard de référence pour les approches top-down</li> </ul>	<ul style="list-style-type: none"> <li>- Une protéine PSAQ est onéreuse</li> <li>- Peu compatible avec une quantification multiplexée</li> <li>- Problème de solubilité de certaine protéine</li> <li>- Pureté élevée parfois difficile à atteindre</li> <li>- L'introduction de modification (phosphorylation ou glycosylation) est possible mais nécessite une étape enzymatique additionnelle</li> </ul>

## 2.2.3 Modes d'acquisition

### 2.2.3.1 Modes d'acquisition DDA et DIA

L'acquisition dépendante des données (DDA : data dependent acquisition) est le mode d'acquisition conventionnel dans lequel chaque spectre MS (permettant la quantification) est suivi de plusieurs spectres MS/MS (permettant l'identification des peptides) déclenchés automatiquement par le logiciel d'acquisition. Il en résulte que les peptides les plus intenses sont fragmentés plus fréquemment que les peptides de plus faible intensité (c'est d'ailleurs sur ce principe qu'est basé le « spectral counting », voir paragraphe 2.2.2.1.b). En outre, certains peptides de faible abondance

seront échantillonnés dans certaines analyses, mais pas dans d'autres, créant ainsi un problème de "données manquantes" qui complique les analyses statistiques.

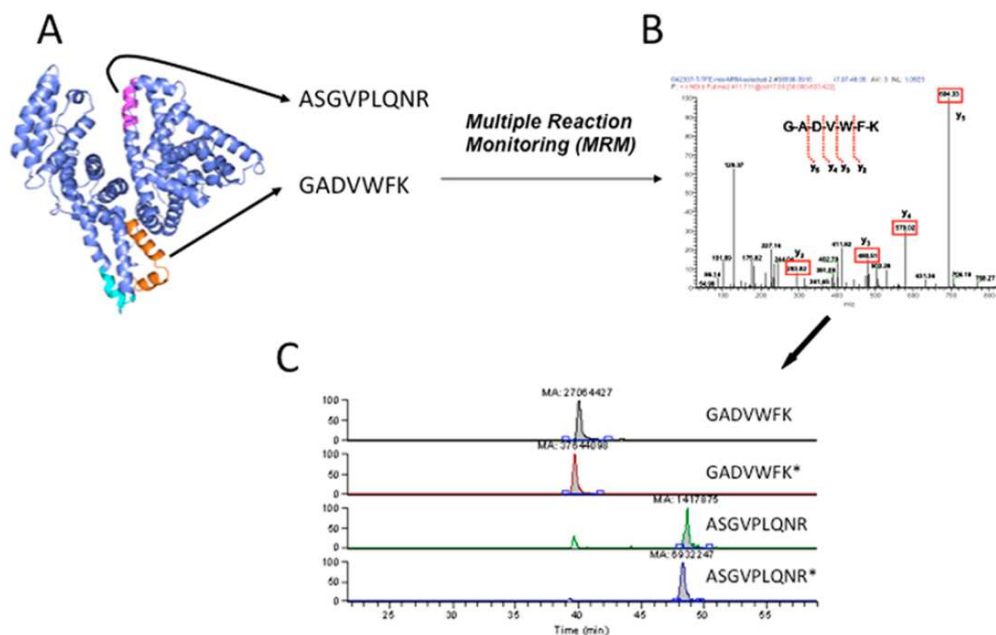
Récemment, de nouvelles stratégies d'acquisition indépendante des données (DIA : data independent acquisition) ont été développées et tentent de collecter simultanément l'information de fragmentation de plusieurs peptides, en élargissant la fenêtre d'isolation. Ce mode DIA permet donc d'acquérir les données sans que le logiciel ait à choisir les pics à fragmenter, ce qui améliore la reproductibilité des mesures (Gillet *et al.*, 2012). Il faut cependant noter que ce mode d'acquisition DIA annihile toute tentative d'exploration manuelle du spectre MS/MS. Pour l'annotation de ces spectres très complexes, les logiciels utilisent des bibliothèques spectrales d'ions fragments.

### 2.2.3.2 Modes d'acquisition SRM-MRM

Les modes SRM (selected reaction monitoring) ou MRM (multiple reaction monitoring) mesurent des peptides spécifiques d'intérêts (Figure 32). Ces modes consistent à sélectionner l'ion parent d'un peptide pour la MS/MS, puis à enregistrer spécifiquement un seul ion du spectre MS/MS. L'enchaînement 1) isolation, fragmentation d'un ion parent et 2) enregistrement sélectif d'un seul événement de fragmentation dans le spectre MS/MS s'appelle une transition. Ce type d'expérience permet de quantifier une transition par SRM ou plusieurs transitions correspondant à un ou plusieurs peptides par MRM pendant l'analyse. Chaque transition doit être optimisée en termes de choix du peptide représentant la protéine à quantifier (peptide protéotypique), d'énergie de fragmentation et de choix d'ions fragments. Les critères de choix d'un peptide protéotypique seront discutés au paragraphe 2.2.3.3.c.

La SRM-MRM, en utilisant des ions fragments pour la quantification, fournit une sélectivité accrue car en plus d'exiger que la masse de l'ion précurseur soit proche de sa masse prédite, les masses des ions fragments doivent également être correctes. Ce gain en sélectivité fait de la quantification ciblée la stratégie la plus sensible, précise, reproductible et la moins affectée par les effets d'interférence lors de l'analyse des échantillons complexes. La SRM, utilisée depuis des décennies en industrie pharmaceutique pour quantifier des petites molécules telles que les métabolites de médicaments (Kostiainen *et al.*, 2003) ou des hormones (Tai *et al.*, 2004), est aujourd'hui la technique de choix pour quantifier des peptides dans un mélange complexe.

**Figure 32. Quantification ciblée d'une protéine par MRM.** (A) Sélection des peptides spécifiques de la protéine cible. (B) Spectre de fragmentation avec les transitions encadrées en rouge. (C) Signal correspondant aux transitions. L'aire sous un pic permet des comparaisons quantitatives entre les peptides cibles et les peptides références marqués par un isotope lourd. D'après (Liebler and Zimmerman, 2013).



### 2.2.3.3 Considérations expérimentales

#### a) Une histoire de compromis

Il n'existe actuellement aucune méthode unique capable d'identifier et de quantifier systématiquement tous les composants d'un protéome. Chaque méthode est donc un compromis qui maximise une performance à certain niveau tout en altérant une autre performance à un autre niveau. Par exemple, dans une expérience SRM, le rapport signal sur bruit enregistré est lié au temps de séjour (« dwell time ») de l'analyte dans le spectromètre de masse (c'est-à-dire, le temps que prend le spectromètre pour enregistrer le signal d'une transition donnée). Ainsi, une limite de détection très faible obtenue par un « dwell time » plus longs affecte négativement le nombre de transitions et donc le nombre de peptides pouvant être mesurés durant un temps donné. De même, une augmentation de la résolution d'un analyseur de masse quadripolaire réduit sa sensibilité.

Concernant le mode DDA, la difficulté récurrente est de trouver un équilibre entre le nombre de spectres MS et MS/MS. D'une part, la précision de quantification est atteinte en réalisant une dizaine de MS au cours de l'élution d'un pic chromatographique pour la reconstruction d'un chromatogramme d'ion extrait (XIC) et d'autre part, le nombre de protéine identifiée sera d'autant plus élevé que le nombre de MS/MS sera grand. Pour trouver le bon compromis permettant d'obtenir une quantification précise tout en maximisant l'identification des protéines, il faut considérer la

résolution du système LC et la vitesse d'acquisition du spectromètre de masse. Par exemple, les spectromètres de masse modernes tels que les QTOF ou orbitrap peuvent aisément générer 10 spectres par secondes (fréquence d'acquisition = 10 Hz) et les largeurs de pics UHPLC sont classiquement de 10 sec à la base (soit 3-4 sec à mi-hauteur). Ainsi, un nombre de 10 MS/MS par MS devrait être un bon réglage. Une alternative possible consiste à faire deux analyses séparées, une pour la MS et l'autre pour la MS/MS.

En outre, beaucoup de performances instrumentales dépendent de la source de l'échantillon et sa complexité. Par exemple, les méthodes DDA, où l'ion précurseur doit être détecté de manière explicite avant la sélection pour la fragmentation, sont plus fortement affectées par le bruit de fond que les méthodes SRM où l'ion précurseur n'a pas besoin d'être détecté de manière explicite pour être fragmenté.

#### *b) Conception d'une expérience SRM*

La conception d'une expérience SRM comprend les étapes suivantes : 1) la sélection des peptides protéotypiques, 2) l'identification des ions fragments les plus appropriés pour chaque peptide protéotypique, 3) l'optimisation des paramètres instrumentaux pour chaque transition, 4) la validation de la spécificité du dosage dans la matrice étudiée (urine, plasma, eau, aliment...), 5) la détermination de la limite de détection et de la limite de quantification dans cette même matrice, et 6) la compilation de tous les paramètres d'analyse :  $m/z$  du précurseur,  $m/z$  du ou des ion(s) fragment(s), temps de rétention, énergie de collision et temps de séjour dans l'instrument (dwell time), pour tous les peptides d'intérêt.

#### *c) Critère de choix d'un peptide protéotypique*

La base de la quantification bottom-up et ciblée réside dans le choix du ou des peptide(s) idéal(s) utilisé(s) pour quantifier la protéine étudiée. On parle de peptide protéotypique. Pour assurer une quantification fiable et robuste, un peptide protéotypique doit avoir les propriétés suivantes :

- être soluble et facilement ionisable. Pour cela, on choisira des peptides entre 5 et 25 acides aminés. Plus un peptide aura une efficacité d'ionisation élevée et plus la limite de quantification (LOQ : limit of quantitation) sera faible, garantissant ainsi un dosage sensible.
- ne pas contenir de séquences susceptibles d'être partiellement clivées par la protéase (missed cleavage).
- ne pas contenir d'acides aminés fortement susceptibles d'être modifiés par une modification post-traductionnelle ou par une modification induite. Ainsi les peptides contenant Ser, Thr, Tyr (résidus



phosphorylables), Cys, Met, Trp (sensible à l'oxydation), Asn, Gln (sensible à la déamidation) ou encore Gln en N-terminal (sensible à la cyclisation pour former un pyro-Glu) sont à éviter. On ignore dans ce raisonnement les modifications potentielles sur les autres types d'acides aminés, en particulier ceux qui sont sensibles à la protéolyse.

- ne pas contenir des séquences instables Asp-Pro et Asp-Gly

Pour la quantification de modification, les peptides contenant la ou les modification(s) (phosphorylation, glycosylation, méthylation, acétylation...) peuvent être utilisés mais le peptide protéotypique idéal est parfois difficile à obtenir car on est contraint par le site modifié à quantifier et ses séquences adjacentes. Le seul élément sur lequel il est possible de jouer est le choix de la protéase qui fera plus ou moins varier la longueur du peptide portant la modification.

Aujourd'hui, plusieurs bases de données regroupant les peptides déjà utilisés pour la quantification sont disponibles: PeptideAtlas, Global Proteome Machine (GPM), PRIDE ou SRMAtlas.

#### *d) Les spectromètres de masse utilisés pour la quantification*

La plupart des configurations instrumentales dédiées à la quantification reposent sur l'ionisation électrospray couplée à une chromatographie liquide (LC-MS). Nous avons vu que les quantifications globale et ciblée diffèrent dans le nombre de protéine quantifiée, la conception expérimentale et le système analytique utilisé. Alors que les études globales traitent souvent un nombre limité d'échantillons (généralement pas plus d'une douzaine), les études ciblées peuvent inclure une centaine d'échantillons. Ainsi, les approches ciblées nécessitent un système à haut débit robuste pour générer des données reproductibles. Elles sont couramment effectuées sur des instruments de type triple quadripôles ou Q-orbitrap. Concernant les approches globales, elles peuvent être en principe réalisées sur n'importe quel spectromètre de masse (Q-TOF, orbitrap, trappe ionique,).

D'autres méthodes, alternatives ou complémentaires, utilisant l'ionisation MALDI ont également été proposées (Pan *et al.*, 2008). Les approches MALDI offre plusieurs caractéristiques uniques telles que 1) le découplage des opérations MS et MS/MS qui permet de sélectionner uniquement les MS/MS d'ions précurseurs dont l'abondance varie, 2) la possibilité de réaliser des analyses répétées du même échantillon et 3) la génération d'ions monochargés qui simplifie les spectres MS/MS. Toutefois, ce couplage LC-MALDI est également associés à des contraintes: 1) temps d'analyse plus long car la séparation LC et l'analyse MS sont dissociées, 2) nécessité d'investir dans un instrument supplémentaire pour réaliser les dépôts en sortie de colonne LC (« spotter »).

*e) Les paramètres LC et MS importants pour la quantification*

Ces dernières années, l'amélioration des performances à la fois en LC et en MS ont permis l'essor de la quantification par spectrométrie de masse.

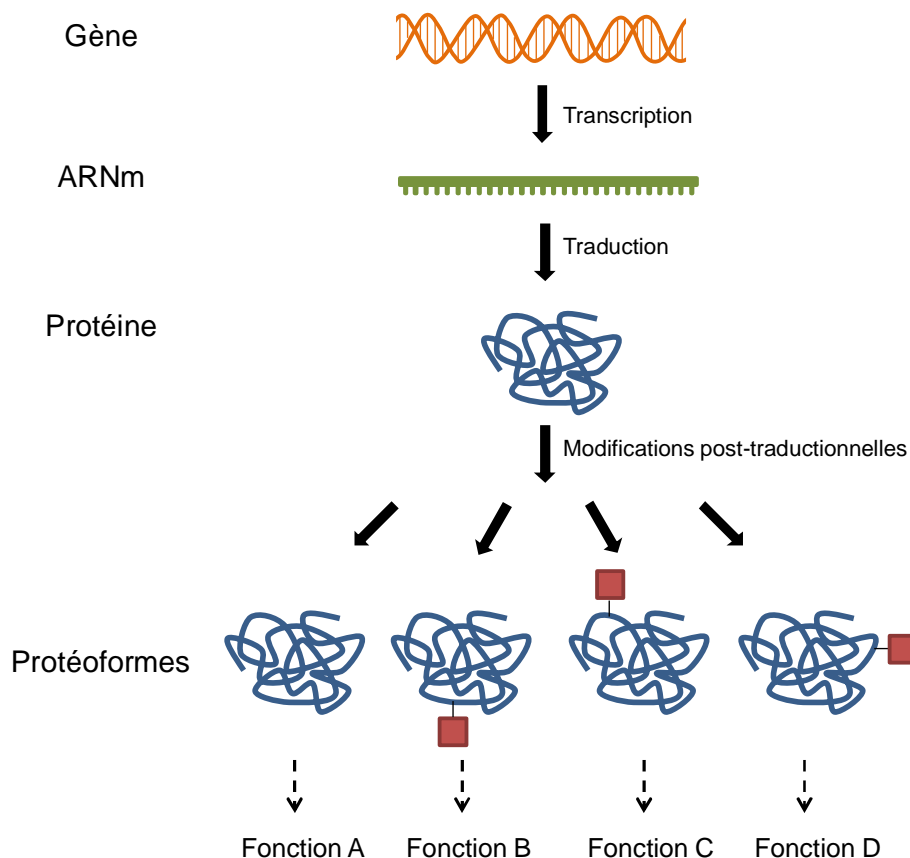
Au niveau LC, l'arrivée de l'UHPLC avec des phases stationnaires de quelques  $\mu\text{m}$  de diamètre et des colonnes de 25 cm de long ont permis d'améliorer la capacité de pics et la sensibilité tout en réduisant les temps d'analyse d'un facteur 5. En quantification, un des paramètres LC important est la robustesse. Celle-ci se traduit par une stabilité des temps de rétention d'une injection à l'autre (de l'ordre de 1%) et permet de réduire les problèmes d'alignement d'un pic chromatographique, de limiter les erreurs d'annotation et d'avoir une quantification plus précise.

Au niveau MS, les performances en termes de sélectivité, sensibilité, vitesse d'acquisitions, précision de masse et résolution sont en progression constante depuis ces dernières années et sont des critères indispensables pour la protéomique quantitative. Parmi eux, la sélectivité est très certainement le plus important. D'une manière générale, la sélectivité d'un système analytique est sa capacité à distinguer et quantifier un analyte particulier dans un mélange (matrice), sans interférences d'autres composés. Concernant le spectromètre de masse, une sélectivité accrue est obtenue 1) avec des analyseurs haute résolution ( $R > 20000$ ) et haute précision de mesure (HR-AM : High Resolution-Accurate Mass) tels que l'ICR, l'orbitrap ou encore les TOF de dernière génération, qui séparent les espèces ioniques quasi-isobares et déterminent leur masse exacte respective, ou 2) en utilisant un deuxième niveau de sélection MS/MS, avec des analyseurs de type quadripôles qui permettent d'isoler un ion précurseur dans une fenêtre d'isolation très restreinte (approches SRM). Dans le 1<sup>er</sup> cas, le bruit de fond est filtré lors du retraitement des données en utilisant une tolérance très stricte, inférieure ou égale à 5 ppm (soit 0.003 Da au  $m/z$  600), pour générer les chromatogrammes des ions extraits (XIC) d'intérêt et dans le 2<sup>ème</sup> cas le bruit de fond est filtré directement à l'acquisition.

## 2.3 Quantification de modification

Comme évoqué dans la section 1, il existe de nombreuses modifications post-traductionnelles de protéine, plus de 300 connues. Ces modifications sont des régulateurs dynamiques de la fonction des protéines, et jouent des rôles importants dans les maladies telles que le cancer. Ainsi, quantifier les PTMs, au cours du cycle cellulaire par exemple ou entre différents échantillons sains et malades, permet d'obtenir des informations plus fines sur les mécanismes moléculaires qui sous-tendent les processus biologiques physiologiques ou pathologiques. Une protéine peut contenir une multitude de PTM et chaque combinaison constitue une protéoforme avec une fonction particulière dans la cellule (Figure 33). Différents objectifs peuvent être ciblés lorsqu'on s'intéresse à la quantification de modification de protéine.

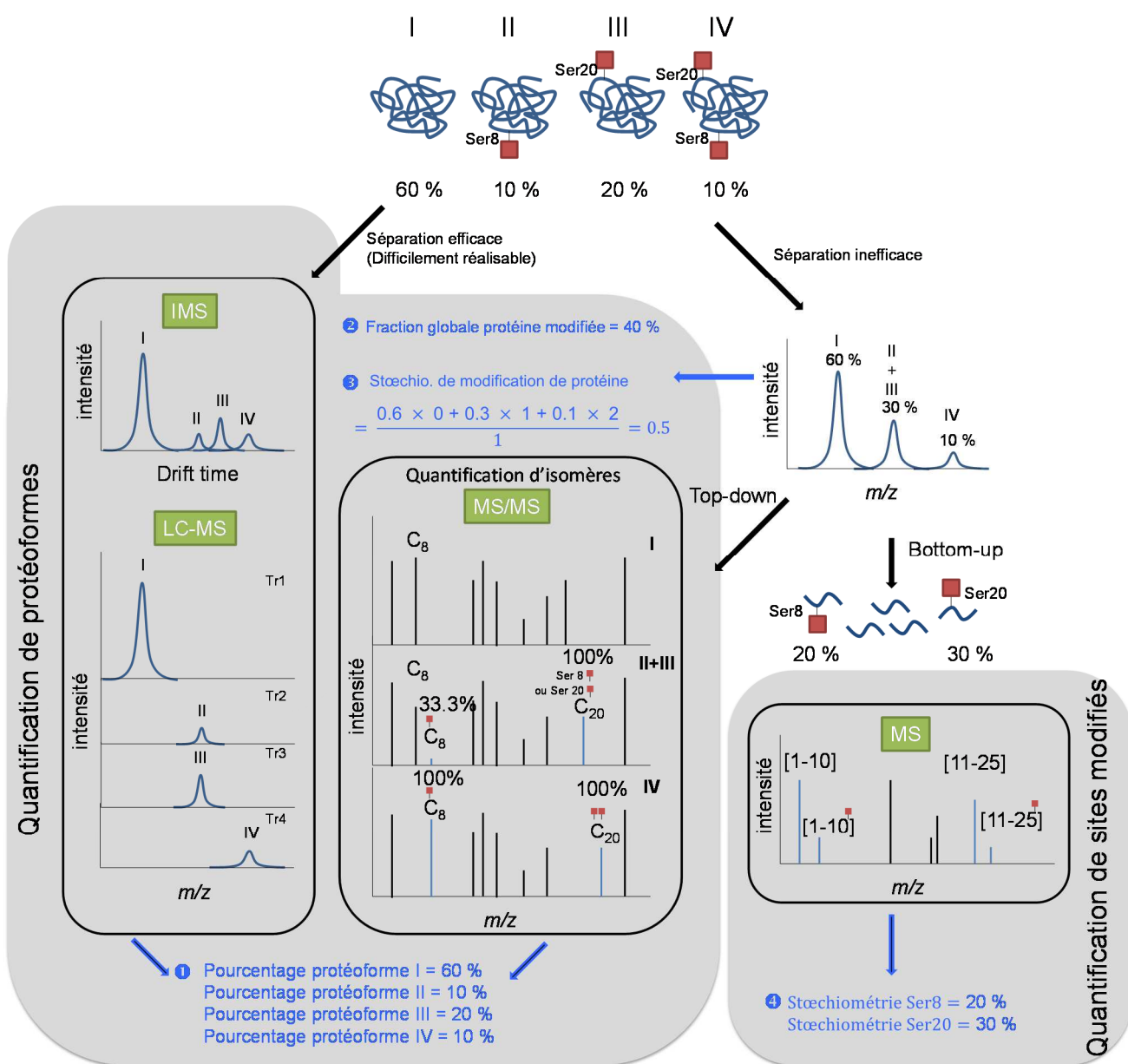
**Figure 33. Du gène aux différentes protéoformes d'une protéine.**



### 2.3.1 Les objectifs

Quatre objectifs principaux peuvent être considérés pour la quantification de modifications de protéine. Ils visent soit à quantifier les modifications à l'échelle de la protéine par quantification de protéoformes (approche top-down) soit à quantifier les modifications au niveau des sites par quantification de peptides modifiés (approche bottom-up) (Figure 34).

**Figure 34. Quantification de protéoformes – quantification de sites modifiés.** Illustration des différentes possibilités de quantification d'un échantillon contenant quatre protéoformes I (non modifiée), II (modifiée en Ser 8), III (modifiée en Ser 20) et IV (modifiée en Ser 8 et Ser 20), d'abondances 60%, 10%, 20% et 10%, respectivement. La quantification de protéoformes nécessite une séparation efficace de chaque protéoforme par LC-MS et/ou mobilité ionique mais cela est encore difficilement réalisable avec les performances actuelles de ces techniques. Lorsque ces dernières permettront de séparer les isomères d'une protéine, la quantification de chaque protéoforme (❶) sera facilement accessible. Dans la plupart des cas, les protéoformes d'une protéine ne sont pas séparées efficacement en LC et sont analysées en mélange en MS. Dans notre exemple, les protéoformes II et III sont confondues sous le même pic. Ces analyses MS de protéines entières permettent de déterminer la fraction globale de protéine modifiée (❷) et la stœchiométrie de modification de la protéine entière (❸). Ensuite, ces protéoformes en mélange peuvent être analysées par approche bottom-up ou top-down afin d'identifier et quantifier les sites modifiés. L'approche bottom-up permet de déterminer la stœchiométrie de modification des sites (❹) et l'approche top-down, en étant capable de détecter et de quantifier les isomères dans le spectre MS/MS, permet d'accéder à la quantification de chaque protéoforme (❶). Le succès de cette dernière dépend de la taille de la protéine (plus la protéine sera de petite taille et plus simple et complet seront les spectres de fragmentation) et de la proximité des résidus modifiés dans la séquence (plus deux résidus seront proche et moins il y aura d'ion fragments discriminants pour les distinguer).



### 2.3.1.1 Quantification de chaque protéoforme d'une protéine

L'enjeu de demain pour la quantification de modifications de protéines est de quantifier, chaque protéoforme de toutes les protéines afin de déterminer la fonction de chaque protéoforme et de comprendre finement et dans leur ensemble les différents mécanismes qui gouvernent un état cellulaire normal ou aberrant. Le problème fondamental avec l'analyse quantitative de ces protéoformes est qu'il est presque impossible de les séparer, notamment les isomères, alors que cela est nécessaire pour estimer leur quantité respective par rapport à la quantité totale de la protéine considérée.

Sous réserve de pouvoir isoler une protéoforme donnée, dans les mélanges peu complexes par exemple, la quantification de cette protéoforme peut s'exprimer en termes

1) d'abondance relative (= pourcentage) de cette protéoforme par rapport à l'ensemble des protéoformes d'une protéine :

$$\text{① Pourcentage d'une protéoforme} = \frac{\Sigma \text{ du signal d'une protéoforme}}{\text{signal protéine totale}} \times 100$$

2) de variations d'abondance en comparant son signal entre plusieurs échantillons (quantification relative)

3) d'abondance absolue en comparant son signal à un standard (quantification absolue), sous réserve de pouvoir le produire et le purifier.

### 2.3.1.2 Quantification de la fraction globale de protéine modifiée

A défaut de pouvoir séparer et quantifier chaque protéoforme, on peut exprimer la quantité de protéoformes en termes de fraction globale de protéine modifiée qui correspond à :

$$\text{② Fraction globale de protéine modifiée} = \frac{\Sigma \text{ du signal des protéoformes modifiées}}{\text{signal protéine totale}}$$

La différence majeure avec la quantification d'une protéoforme est que l'on quantifie les modifications dans leur ensemble. Cette mesure est obtenue simplement en analysant le mélange de protéoformes d'une protéine entière en MS. Une étape de séparation optionnelle peut être réalisée en amont si les échantillons sont complexes mais dans le cadre d'études structurales, réalisées sur des protéines purifiées, cette séparation n'est souvent pas nécessaire.

### 2.3.1.3 Quantification de la stœchiométrie totale de modification de la protéine entière

La quantification de la stœchiométrie de modification d'une protéine correspond à la somme des stœchiométries des sites modifiés individuels (④). Cette mesure se distingue de la fraction globale en ce sens que l'on fait intervenir le nombre de modification :

$$\textcircled{3} \text{ Stœchio. Tot. Modif. Prot.} = \frac{\sum (\text{nombre de modification d'une protéoforme } n \times \text{signal d'une protéoforme } n)}{\text{signal protéine totale}}$$

Lorsque chaque site est modifié à 100% (= saturation), la mesure tend vers le nombre total de modifications fixées sur la protéine.

Dans ce type de quantification et d'une manière générale en quantification de protéine entière, on fait l'approximation que les différentes protéoformes ont la même efficacité d'ionisation car en général l'effet néfaste ou bénéfique d'une modification sur l'efficacité d'ionisation est largement dilué par la quantité accrue de sites de protonation de la protéine.

### 2.3.1.4 Quantification de la stœchiométrie de modification d'un ou plusieurs site(s)

La quantification de la stœchiométrie de modification d'un ou plusieurs site(s) (④) correspond à l'approche historique pour quantifier les modifications de protéines. Elle est classiquement réalisée par la stratégie bottom-up « shotgun ». Dans cette approche, la modification peut affecter l'efficacité d'ionisation d'un peptide, car les sites de protonation sont bien moins nombreux que sur protéine entière. Ainsi, le calcul intuitif qui consiste à mesurer le rapport entre le signal du peptide modifié et la somme des signaux des peptides modifié et non modifié, est souvent entaché d'erreur. Les astuces pour contourner cette difficulté sont présentées au paragraphe 2.3.2.2.b ci-après.

Par ailleurs, cette stœchiométrie de modification d'un site ne doit pas être confondue avec la fraction globale modifiée d'une protéine qui représente la modification globale, et non locale, d'une protéine. Le seul cas où la stœchiométrie de modification d'un site correspond à la fraction globale modifiée c'est lorsqu'il n'y a qu'un seul et unique site modifié. Par ailleurs, étant donné que plusieurs protéoformes peuvent avoir un site modifié commun, la quantification de la stœchiométrie de modification d'un site peut correspondre à la somme des signaux de ce site modifié présent sur plusieurs protéoformes. On voit ainsi que l'approche « shotgun » entraîne la perte de la filiation entre une protéoforme donnée et les modifications qui lui sont associées. Cela constitue la limitation majeure de cette approche.

En parallèle à la stœchiométrie de modification d'un site, la quantification des variations d'abondance d'un site modifié entre plusieurs conditions biologiques peut être réalisée. Cette quantification relative permet d'identifier et quantifier des peptides modifiés et donc des protéines modifiées impliquées dans le processus biologique étudié. Enfin, une quantification absolue de site modifié est également possible en comparant le signal du peptide modifié avec un standard de concentration connue, sous réserve que ce dernier puisse être synthétisé et purifié.

## 2.3.2 Méthodes disponibles pour la quantification de modifications

### 2.3.2.1 Quantification de protéoformes

Des approches de quantification « label-free » top-down combinant des méthodes électrophorétiques et chromatographiques avec la MS à haute résolution commencent à voir le jour mais cela nécessite beaucoup d'optimisations et ne concerne que les protéoformes les plus abondants et de taille inférieure à 40 kDa (Ntai *et al.*, 2014). Malgré quelques exemples fructueux, notamment sur la quantification en MS/MS d'isomères acétylés de l'histone H4 (Pesavento *et al.*, 2006), la localisation de sites modifiés et la quantification de protéoformes restent difficiles par top-down. Dans la pratique, la quantification de modification est le plus souvent réalisée à l'échelle du peptide par des approches bottom-up avec pour objectif de quantifier les sites modifiés.

### 2.3.2.2 Quantification de sites modifiés

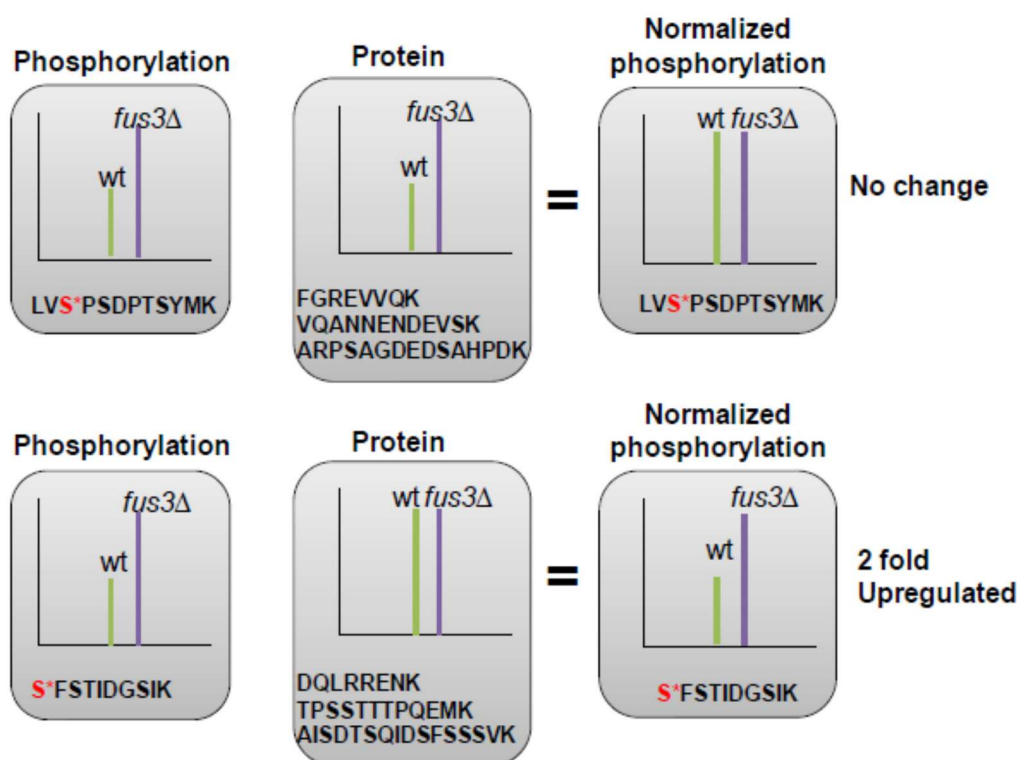
Tout comme pour la localisation de sites modifiés, les méthodes utilisées pour quantifier les modifications ont été dans l'ensemble développées à partir de méthodes bottom-up destinées initialement à la quantification de protéines non-modifiées. Cependant, la quantification de modification est plus difficile car, contrairement à la quantification de protéine où plusieurs peptides sont généralement disponibles pour quantifier chaque protéine, elle est basée sur des peptides uniques et qui sont souvent présents en très faible abondance.

#### a) *Quantification relative d'un site modifié*

Les approches SILAC (Kettenbach *et al.*, 2011), iTRAQ (Boja *et al.*, 2009) ou encore les approches sans marquage (Montoya *et al.*, 2011) sont classiquement utilisées pour comparer les profils de modification (essentiellement la phosphorylation) dans plusieurs états cellulaires. Toutefois, bien que ces études aient permis de mieux comprendre la dynamique de la phosphorylation, les données quantitatives doivent être interprétées avec prudence, car l'abondance différentielle des phosphopeptides reflète non seulement les changements de l'état de phosphorylation mais aussi les

changements d'expression de protéines. La solution simple consiste à quantifier simultanément l'expression des protéines et l'abondance de modification afin de normaliser les variations et ainsi distinguer l'abondance différentielle réelle de la modification et les changements d'expression de la protéine correspondante (Figure 35) (Wu *et al.*, 2011b).

**Figure 35. L'interprétation des variations de peptides modifiés nécessite une normalisation par les changements de l'expression des protéines.** (B) Exemples de l'effet de la normalisation du niveau d'expression de protéines sur le rapport observé de deux phosphopeptides issus d'une souche de levure sauvage et mutante (*fus3Δ*). D'après Wu 2011b.



Toutefois, plusieurs difficultés techniques font que cette approche est difficile à appliquer pour les modifications minoritaires. Tout d'abord, les peptides modifiés requièrent souvent une étape d'enrichissement en raison de leur faible abondance et de leurs propriétés chimiques parfois défavorables pour l'ionisation en MS. Au cours de cette étape, les peptides non modifiés sont perdus, empêchant alors la normalisation des variations observées. Ainsi, classiquement, deux analyses LC-MS sont réalisées : une pour l'expression globale de la protéine avant enrichissement et l'autre pour la phosphorylation après enrichissement. Un avantage de l'enrichissement de peptides modifiés est qu'il permet d'identifier des protéines très faiblement exprimées qui seraient autrement indétectables, même avec un fractionnement poussé. Toutefois, ces protéines ultra minoritaires, invisibles sans enrichissement, ne sont donc pas observées lors de l'analyse destinée à quantifier l'expression globale, empêchant alors toute normalisation des variations de modifications mesurées. Enfin, l'assemblage de ces données protéomiques, associant telle abondance de site modifié avec tel niveau d'expression de protéine, représente un défi majeur, notamment dans le cas d'études à grande échelle.



En plus de la quantification relative du site de modification d'intérêt, il est très utile de déterminer de la stœchiométrie de modification d'un site modifié afin d'évaluer son importance fonctionnelle. En effet, prenons l'exemple d'une modification dont l'abondance relative augmente d'un facteur deux entre deux conditions biologique. Cette augmentation peut en fait résulter d'un changement de la stœchiométrie de modification de 1% à 2% ou de 50% à 100%, ce qui représente des stratégies cellulaires fondamentalement différentes.

*b) Détermination de la stœchiométrie de modification d'un site – Quantification relative*

Pour déduire la stœchiométrie de modification d'un site, plusieurs études ont utilisées les intensités relatives des signaux des peptides modifiés et du peptide non modifié correspondant (Ruse *et al.*, 2002; Haydon, 2003). Toutefois, il a été démontré qu'une modification, notamment la phosphorylation qui apporte une charge négative, pouvait perturber le processus d'ionisation et à quantité de peptide égale, changer l'intensité du signal d'ions en spectrométrie de masse (Steen *et al.*, 2006; Gropengiesser *et al.*, 2009). Ces résultats indiquent que, d'une manière générale, la stœchiométrie de modification ne peut être déterminée par une simple comparaison des signaux d'un peptide modifié *versus* peptide non modifié.

Dans la plupart des études du phosphoprotéome, et la stœchiométrie de phosphorylation est déterminée en quantifiant les variations de signal du peptide non modifié ; l'hypothèse centrale étant qu'une diminution de signal associée à un peptide non modifié est due à une augmentation de phosphorylation et *vice versa*. Deux approches principales sont décrites et toutes les deux utilisent un marquage isotopique stable.

Dans la première approche, la stœchiométrie de phosphorylation est déterminée directement à partir des données (Olsen *et al.*, 2010). Elle consiste à établir trois rapports correspondants aux variations de protéine, phosphopeptide et peptide non modifié. Cette approche utilise le marquage SILAC et nécessite de faire deux analyses : avec (variation des phosphopeptides) et sans (variation de protéine et de peptides non modifiés) enrichissement de phosphopeptides. Les limitations de cette stratégie sont 1) qu'il est nécessaire d'avoir des changements importants dans la quantité de phosphopeptides pour qu'une stœchiométrie puisse être déterminée 2) qu'elle est sensible au caractère labile de certaines modifications lorsqu'elles sont analysées en MS et 3) qu'elle implique beaucoup d'opérations mathématiques. Ainsi, même une faible erreur de mesure peut conduire à une erreur importante sur le calcul de la stœchiométrie de modification d'un site (Wu *et al.*, 2011a).

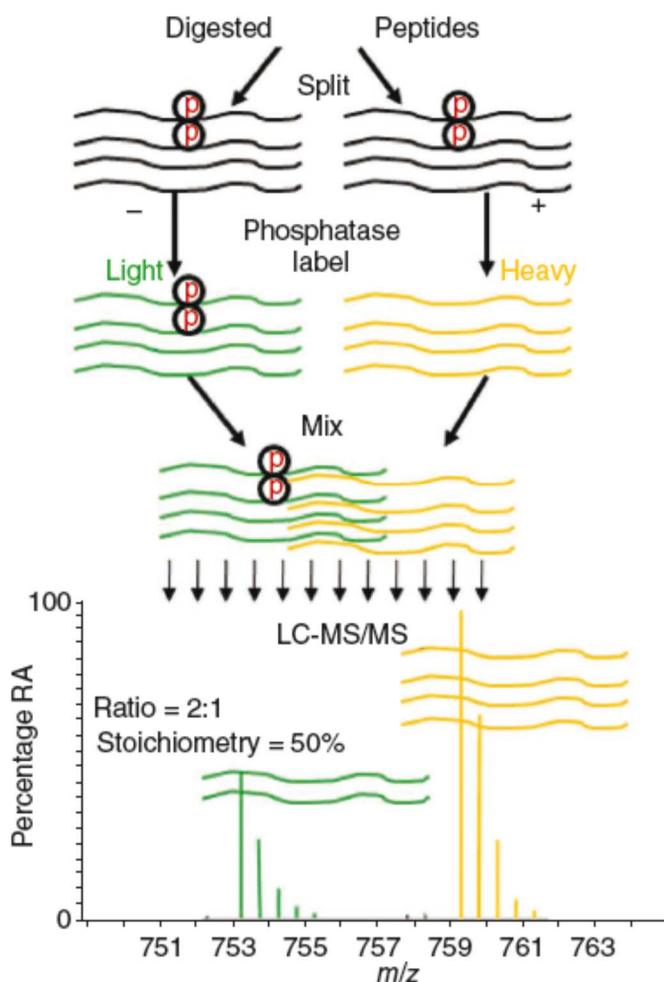
La deuxième approche utilise une combinaison de marquage chimique et d'un traitement à la phosphatase alcaline (Hegeman *et al.*, 2004; Wu *et al.*, 2011a). Dans cette approche, la stœchiométrie

de phosphorylation est mesurée en divisant un échantillon en deux, en réalisant un traitement à la phosphatase sur une des deux fractions, en marquant chimiquement chaque fraction avec des isotopes stables (une fraction marquée avec la forme légère et l'autre avec la forme lourde de l'étiquette), en mélangeant les deux fractions et analysant l'échantillon par MS (Figure 36). Le traitement à la phosphatase entraîne une augmentation du signal du peptide non modifié (version « lourde » dans notre exemple). La stœchiométrie de phosphorylation du site est obtenue par la relation suivante :

$$\frac{\text{signal peptide non modifié version lourde} - \text{signal peptide non modifié version légère}}{\text{signal peptide non modifié version lourde}}$$

Les limitations de cette approche sont 1) qu'elle nécessite à la fois une déphosphorylation par la phosphatase et une dérivation chimique complètes et 2) qu'elle n'est pas applicable aux modifications irréversibles.

**Figure 36. Détermination de la stœchiométrie de phosphorylation d'un site par une combinaison de marquage isotopique chimique et de traitement à la phosphatase.** L'échantillon de départ est divisé en deux aliquotes. L'une d'entre elles subit un traitement à la phosphatase puis les deux aliquotes sont marquées chimiquement avec des isotopes stables différents avant d'être mélangées. La stœchiométrie de phosphorylation du site est obtenue avec la relation suivante :  $\frac{\text{signal peptide non modifié version lourde} - \text{signal peptide non modifié version légère}}{\text{signal peptide non modifié version lourde}}$ . P, site de phosphorylation. (Wu *et al.*, 2011a)



c) Détermination de la stœchiométrie de modification d'un site – Quantification absolue

La quantification absolue de peptides modifiés et de leur peptide non modifié correspondant, en s'affranchissant des différences d'ionisation éventuelles entre peptides modifiés et peptides non modifiés, permet de déterminer simplement la stœchiométrie de modification d'un site avec la formule suivante :

Stœchiométrie de modification d'un site

$$= \frac{\frac{\text{signal peptide modifié}}{\text{signal SI modifié}} \times [\text{SI modifié}]}{\left( \frac{\text{signal peptide non modifié}}{\text{signal SI non modifié}} \times [\text{SI non modifié}] + \frac{\text{signal peptide modifié}}{\text{signal SI modifié}} \times [\text{SI modifié}] \right)}$$

En 2003 l'équipe de Gygi a rapporté la première étude AQUA pour la quantification de phosphorylation (Gerber *et al.*, 2003). Dans ces travaux, deux peptides de référence (standards internes) ont été synthétisés : la forme non modifié et la forme monophosphorylée en Ser1126, pour déterminer quantitativement les changements dynamiques de la phosphorylation en Ser1126 de la séparase humaine au cours du cycle cellulaire. La même stratégie AQUA a été utilisée pour la détermination de la stœchiométrie d'autres PTMs telles que la méthylation (Hart-Smith *et al.*, 2014) et l'acétylation (Weinert *et al.*, 2014). Le principal défi de ces approches est de réussir à synthétiser le résidu modifié à introduire dans la séquence du standard interne. C'est le problème de la glycosylation par exemple.

La glycosylation est une PTM importante, notamment pour les protéines excrétées et les protéines membranaires. Cependant, du fait de l'hétérogénéité et de la complexité structurale des glycanes, notamment des N-glycanes, la synthèse de peptides AQUA glycosylés n'est jamais envisagée. Pour contourner cette difficulté, une astuce consiste à réaliser une étape de déglycosylation par la PNGaseF et de quantifier le peptide déglycosylé d'intérêt. La déglycosylation entraîne la conversion de l'asparagine initialement glycosylée en acide aspartique, générant un delta-masse de +0.984 Da. En synthétisant un peptide contenant un acide aspartique en lieu et place d'une asparagine glycosylée *in vivo*, la quantification de plusieurs protéines glycosylées a été réalisée dans le sérum (Pan *et al.*, 2005) et le plasma (Stahl-Zeng *et al.*, 2007).

Enfin, deux exemples utilisant un standard de type PSAQ phosphorylé (Ciccimaro *et al.*, 2009) ou PSAQ ubiquitinylé (Kaiser *et al.*, 2011) ont également été décrits. Dans ces stratégies, la modification est introduite spécifiquement au niveau d'un site par voie enzymatique *in vitro*.

### 2.3.3 Ce qu'il manque en approche bottom-up pour quantifier les modifications de résidus de manière robuste et exhaustive

Aujourd'hui la méthode de quantification robuste, applicable à l'ensemble des modifications, et exhaustive, permettant de quantifier tous les sites modifiables, ne semble pas exister. Le cahier des charges pour une telle méthode serait qu'elle soit à la fois :

- universelle, en ce sens qu'elle peut être appliquée à toutes les modifications de protéines,
- insensible au caractère labile de certaine modification lors de l'analyse MS,
- insensible à l'éventuelle différence d'efficacité d'ionisation entre une forme modifiée et non modifiée d'une protéine,
- globale à l'échelle de la protéine, c'est-à-dire une méthode capable de quantifier les modifications sans à priori et en parallèle,
- sélective à l'échelle du résidu modifié, c'est-à-dire une méthode capable de distinguer la contribution de chaque site dans les peptides contenant plusieurs PTM,
- exhaustive (= quantification de tous les sites modifiables). En effet, dans une approche bottom-up, même dans les meilleures conditions possibles, les sites de modifiés peuvent être manqués par MS en raison de plusieurs facteurs propres à la purification et / ou la détection de peptide. Ainsi, il convient de garder à l'esprit que la carte des sites modifiés observée peut ne pas être complète et que le fait de ne pas identifier un site modifié est souvent insuffisant pour écarter sa non-existence.



# Présentation du travail de thèse

# 1. Contexte

Les modifications post-traductionnelles (PTMs) sont essentielles au bon fonctionnement des cellules. Par exemple, le cycle cellulaire, les voies de signalisation ou encore le métabolisme s'appuient sur des PTMs qui sont finement régulées et qui impliquent des molécules endogènes. Les modifications induites sont généralement le résultat d'une exposition à des agents externes tels que les médicaments, le stress, les agents infectieux, l'alimentation, le mode de vie (par exemple le tabac ou l'alcool) ou encore les polluants environnementaux et les rayonnements. Un certain nombre de pathologies humaines sont dues à un déséquilibre de PTM ou de modifications induites. Il est fondamental de caractériser ces modifications afin de comprendre leur rôle et leur dynamique, et de pouvoir ainsi concevoir un jour un médicament contre les pathologies qu'elles occasionnent. Par ailleurs, les médicaments capables de se lier aux protéines de manière covalente représentent une classe ancienne dont l'intérêt a été renouvelé ces dernières années. La spectrométrie de masse (MS) est une technique de choix pour détecter, localiser et quantifier les modifications de protéines. Les modifications de quatre protéines d'intérêt thérapeutique en cancérologie, virologie ou dans les maladies inflammatoires sont étudiées par MS.

La grégline est une protéine capable d'inhiber très efficacement l'élastase de neutrophile humaine (HNE), protéase impliquée dans l'invasion tumorale et dans les maladies pulmonaires telles que la maladie pulmonaire obstructive chronique et la mucoviscidose. Les propriétés physico-chimiques de la grégline suggèrent qu'elle peut être utilisée en tant que modèle pour concevoir une nouvelle génération d'inhibiteurs de HNE très résistants à la dégradation. Pour ce faire, une caractérisation de la structure primaire et des modifications de la grégline est nécessaire pour comprendre leur importance sur son activité inhibitrice de protéases.

La protéine MOS1 est une transposase de l'élément mariner-Mos1 qui présente des caractéristiques communes avec l'intégrase du VIH-1. Plusieurs inhibiteurs de MOS1 ont été identifiés et pourraient être utilisés à des fins thérapeutiques pour inhiber les intégrases rétrovirales. Une meilleure compréhension du mode d'action de ces inhibiteurs serait très utile pour concevoir de meilleurs médicaments.

La protéine Autonomously replicating sequence (ARS)-binding factor 2 (Abf2), bien connue pour son rôle essentiel dans le maintien de l'ADN mitochondrial, a récemment été décrite comme étant une enzyme impliquée dans la réparation de l'ADN mitochondrial de *Saccharomyces cerevisiae*. Afin de mieux comprendre cette nouvelle activité d'Abf2 et ainsi d'appréhender son potentiel éventuel comme cible thérapeutique, une identification du site actif est nécessaire.

La protéine humaine Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1), est un suppresseur de métastases dans différents types de cancer. La locostatine est le seul composé connu qui empêche la migration des cellules en se liant à hPEBP1. Aujourd'hui, la plupart des approches utilisées pour prévenir les métastases sont basées sur l'inhibition des facteurs pro-métastatiques. La particularité de la protéine hPEBP1 est d'être au contraire un facteur anti-métastatique. Les travaux présentés ici s'inscrivent dans un projet collaboratif dont l'objectif est d'exploiter le potentiel anti-métastatique naturel de PEBP1 en synthétisant, de façon rationnelle, des analogues de locostatine ayant un effet agoniste (ou activateur) de la protéine pour en faire de futurs médicaments capables d'empêcher le développement des métastases cancéreuses.

## 2. Objectifs du travail de thèse

### 2.1 Corriger la séquence de la grégline et localiser ses PTMs

La grégline a des propriétés remarquables de résistance à la protéolyse et à la dénaturation thermique, acide ou basique. L'objectif est d'explorer la séquence de la grégline et d'inventorier puis localiser ses PTMs, et d'utiliser ces données en combinaison avec des données de structure cristallographique pour comprendre l'origine de cette résistance.

### 2.2 Caractériser le complexe MOS1-inhibiteur

Des inhibiteurs de l'activité de transposition de MOS1 ont été découverts. Afin de décrypter le mécanisme de liaison et de localiser le ou les site(s) cible(s) des inhibiteurs, il est nécessaire de caractériser l'interaction de ces inhibiteurs avec MOS1.

### 2.3 Localiser le site actif d'Abf2

Récemment, une activité lyase d'Abf2 a été découverte dans notre institut au CBM. Pour contribuer à la compréhension du mécanisme catalytique, il faut localiser le site actif d'Abf2.

### 2.4 Caractériser l'interaction entre hPEBP1 et la locostatine afin de fournir des éléments pour la synthèse raisonnée d'activateurs des propriétés antimétastatiques naturelles de hPEBP1

Dans le but de fournir des informations précieuses pour la synthèse raisonnée d'analogues de locostatine, l'interaction entre hPEBP1 et la locostatine doit être caractérisée au niveau cinétique et



structural. Cette caractérisation comprend la dissection du mécanisme de liaison entre les deux partenaires et la localisation du site spécifique lié de manière covalente à la locostatine.

## 2.5 Développer des outils pour la caractérisation de modifications de protéines

Plusieurs des protéines cibles étudiées ici nécessitent des outils fiables pour l'identification, la localisation et la quantification de modifications, qu'elles soient post traductionnelles ou induites dans le cadre de complexes protéine-ligand. Pour cela, plusieurs méthodes plus ou moins générales et combinant des outils biochimiques avec la MS vont devoir être développées.

L'étude bibliographique de l'état de l'art a mis en évidence qu'il n'existe pas de méthode qui soit à la fois universelle, exhaustive, et globale pour quantifier de manière fiable le degré de modification de tous les sites d'une protéine modifiée. Un des objectifs de ce travail est donc de développer une nouvelle méthode de quantification répondant à des critères précis. En effet, elle doit être insensible au caractère labile ou dégradable de certaines modifications, précise à l'échelle du résidu et insensible aux différences d'efficacité d'ionisation entre les formes modifiées et non modifiées.

# Résultats et Discussion

- Chapitre 1 Une nouvelle modification de la grégline, un inhibiteur de protéases à sérine
- Chapitre 2 Etude du complexe entre la transposase MOS1 et des inhibiteurs
- Chapitre 3 Vers la localisation du site actif d'Abf2, une protéine impliquée dans la réparation de l'ADN mitochondrial
- Chapitre 4 Caractérisation de l'interaction entre la locostatine et la protéine PEBP1/RKIP humaine, un suppresseur de métastases naturel

# Chapitre 1 Une nouvelle modification de la grégligne, un inhibiteur de protéases à sérine

*Résumé du chapitre 1 :*

*La grégline est une protéine capable d'inhiber très efficacement l'élastase de neutrophile humaine (HNE), protéase majeure de l'inflammation, également impliquée dans l'invasion tumorale et dans les maladies pulmonaires. La grégline a des propriétés remarquables de résistance à la protéolyse et à la dénaturation thermique, acide ou basique. Ces propriétés physico-chimiques suggèrent que la grégline peut être utilisée en tant que modèle pour concevoir une nouvelle génération d'inhibiteurs de HNE très résistants à la dégradation. Pour comprendre l'origine de cette résistance, la spectrométrie de masse haute résolution combinée avec plusieurs outils biochimiques a été utilisée pour obtenir la séquence correcte de la grégline et déterminer ses modifications post-traductionnelles. Ces travaux ont fait l'objet d'une publication dans FEBS Journal. Par ailleurs, une modification nouvelle a été détectée et une formule brute est proposée.*

# 1. Contexte

Les protéases à sérine sont impliquées dans de nombreux processus biologiques tels que la réponse immunitaire ou la coagulation sanguine. Dans un organisme sain, l'activité protéolytique de ces enzymes est régulée par des inhibiteurs endogènes. Un déséquilibre entre les protéases et les inhibiteurs peut conduire à des dysfonctionnements graves. Ainsi, l'élastase de neutrophile humaine (HNE) est considérée comme la protéase majeure impliquée dans la destruction des tissus lors de l'inflammation. En agissant directement sur les protéines de la matrice extracellulaire, cette sérine-protéase est impliquée dans l'invasion tumorale et dans les maladies pulmonaires telles que la maladie pulmonaire obstructive chronique (COPD : chronic obstructive pulmonary disease) et la mucoviscidose (cystic fibrosis). Le développement de nouvelles thérapies anti-HNE pour traiter des maladies inflammatoires semble prometteur. La grégline est un inhibiteur puissant de HNE. C'est une protéine purifiée à partir d'ovaires de criquet *Schistocerca gregaria*. Elle est résistante à la protéolyse et à la dénaturation (thermique, acide ou basique) et ses propriétés physico-chimiques suggèrent qu'elle peut être utilisée en tant que modèle pour concevoir une nouvelle génération d'inhibiteurs de HNE très résistants à la dégradation.

La grégline est une protéine de 83 acides aminés qui contient trois sites de phosphorylations putatifs S8, S11 et S15 (Brillard-Bourdet *et al.*, 2006). Ces résidus sont suspectés être phosphorylés car 1) le séquençage d'Edman n'a pas permis d'identifier d'acide aminé standard (non modifié) à ces positions, 2) un quadruplet de pic séparés de 80 Da a été observé en spectrométrie de masse sur la grégline entière et 3) le programme de prédiction de site de phosphorylation NetPhos propose ces positions comme étant phosphorylées (Brillard-Bourdet *et al.*, 2006). Les alignements de séquence n'ont révélé aucune homologie avec d'autres inhibiteurs connus. Pour découvrir les mécanismes qui sous-tendent ses propriétés biologiques, une caractérisation de sa structure primaire et tertiaire a été effectuée par spectrométrie de masse et cristallographie aux rayons X.

En combinant plusieurs outils biochimiques avec la spectrométrie de masse haute résolution nous avons déterminé la séquence correcte de la grégline ainsi que ses modifications post-traductionnelles (PTMs). La structure cristallographique indique que la grégline représente un nouveau membre de la famille des inhibiteurs Kazal non classiques, avec la particularité de contenir une extrémité C-terminale additionnelle qui pourrait être impliquée dans la sélectivité d'inhibition de la grégline.

## 2. Publication

**Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with subtilisin**

## Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with subtilisin

Chrystelle Derache<sup>1,\*</sup>, Christophe Epinette<sup>2</sup>, Alain Roussel<sup>3</sup>, Guillaume Gabant<sup>1</sup>, Martine Cadene<sup>1</sup>, Brice Korkmaz<sup>2</sup>, Francis Gauthier<sup>2</sup> and Christine Kellenberger<sup>3</sup>

<sup>1</sup> Centre de Biophysique Moléculaire, UPR 4301 CNRS conventionnée avec l'Université d'Orléans, Orléans Cedex 2, France

<sup>2</sup> Pathologies Respiratoires, Protéolyse et Aérosolthérapie, INSERM U1100 Faculté de Médecine, Université François Rabelais, Tours, France

<sup>3</sup> Architecture et Fonction des Macromolécules Biologiques, CNRS UMR7257 and Aix-Marseille Université, Marseille Cedex, France

### Keywords

disulfide bond; Kazal; protease-inhibitor complex; specificity; thermostability

### Correspondence

C. Kellenberger, Architecture et Fonction des Macromolécules Biologiques, CNRS UMR7257 and Aix-Marseille Université, 13288 Marseille Cedex, France  
Fax: +33 4 91 26 67 20  
Tel: +33 4 91 82 55 83  
E-mail: christine.kellenberger@afmb.univ-mrs.fr

### \*Present address

Medical Microbiology, Department of Laboratory Medicine Malmö, Lund University, Skåne University Hospital, Malmö, S-20520, Sweden

(Received 29 August 2012, revised 5 October 2012, accepted 9 October 2012)

doi:10.1111/febs.12033

Greglin is an 83-residue serine protease inhibitor purified from the ovaries of the locust *Schistocerca gregaria*. Greglin is a strong inhibitor of subtilisin and human neutrophil elastase, acting at sub-nanomolar and nanomolar concentrations, respectively; it also inhibits neutrophil cathepsin G,  $\alpha$ -chymotrypsin and porcine pancreatic elastase, but to a lesser extent. In the present study, we show that greglin resists denaturation at high temperature (95 °C) and after exposure to acetonitrile and acidic or basic pH. Greglin is composed of two domains consisting of residues 1–20 and 21–83. Mass spectrometry indicates that the N-terminal domain (1–20) is post-translationally modified by phosphorylations at three sites and probably contains a glycosylation site. The crystal structure of the region of greglin comprising residues 21–78 in complex with subtilisin was determined at 1.75 Å resolution. Greglin represents a novel member of the non-classical Kazal inhibitors, as it has a unique additional C-terminal region (70–83) connected to the core of the molecule via a supplementary disulfide bond. The stability of greglin was compared with that of an ovomucoid inhibitor. The thermostability and inhibitory specificity of greglin are discussed in light of its structure. In particular, we propose that the C-terminal region is responsible for non-favourable interactions with the autolysis loop (140-loop) of serine proteases of the chymotrypsin family, and thus governs specificity.

### Database

The atomic coordinates and structure factors for the greglin–subtilisin complex have been deposited with the RCSB Protein Data Bank under accession number [4GI3](#).

### Structured digital abstract

- [Greglin and Subtilisin Carlsberg bind by X-ray crystallography \(View interaction\)](#)

## Introduction

Serine proteases are involved in many biological processes. In addition to their role in digestion, they have highly regulated functions in embryonic development, the immune response and blood coagulation [1]. In a healthy organism, the proteolytic activity of these

enzymes is regulated by natural endogenous inhibitors. An imbalance between proteases and inhibitors may lead to severe dysfunctions. Serine protease inhibitors are classified as canonical inhibitors, non-canonical inhibitors and serpins based on their mechanism of

### Abbreviations

AEI, *Anemonia* elastase inhibitor; CID, collision-induced dissociation; HRMS, high-resolution mass spectrometry; OMTKY3, third domain of turkey ovomucoid; RSL, reactive site loop.



### 3. Supplementary informations





## **Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with subtilisin**

Chrystelle Derache, Christophe Epinette, Alain Roussel, Guillaume Gabant, Martine Cadene, Brice Korkmaz, Francis Gauthier and Christine Kellenberger

DOI: 10.1111/febs.12033

## 4. Annexes à la publication

### 4.1 Analyse des cristaux du complexe grégligne-subtilisine par spectrométrie de masse.

Comme évoqué dans la publication, aucune densité électronique n'a été observée pour les résidus 1-20 et 79-83. Pour rappel, la Figure 37 montre la séquence corrigée de la grégligne.

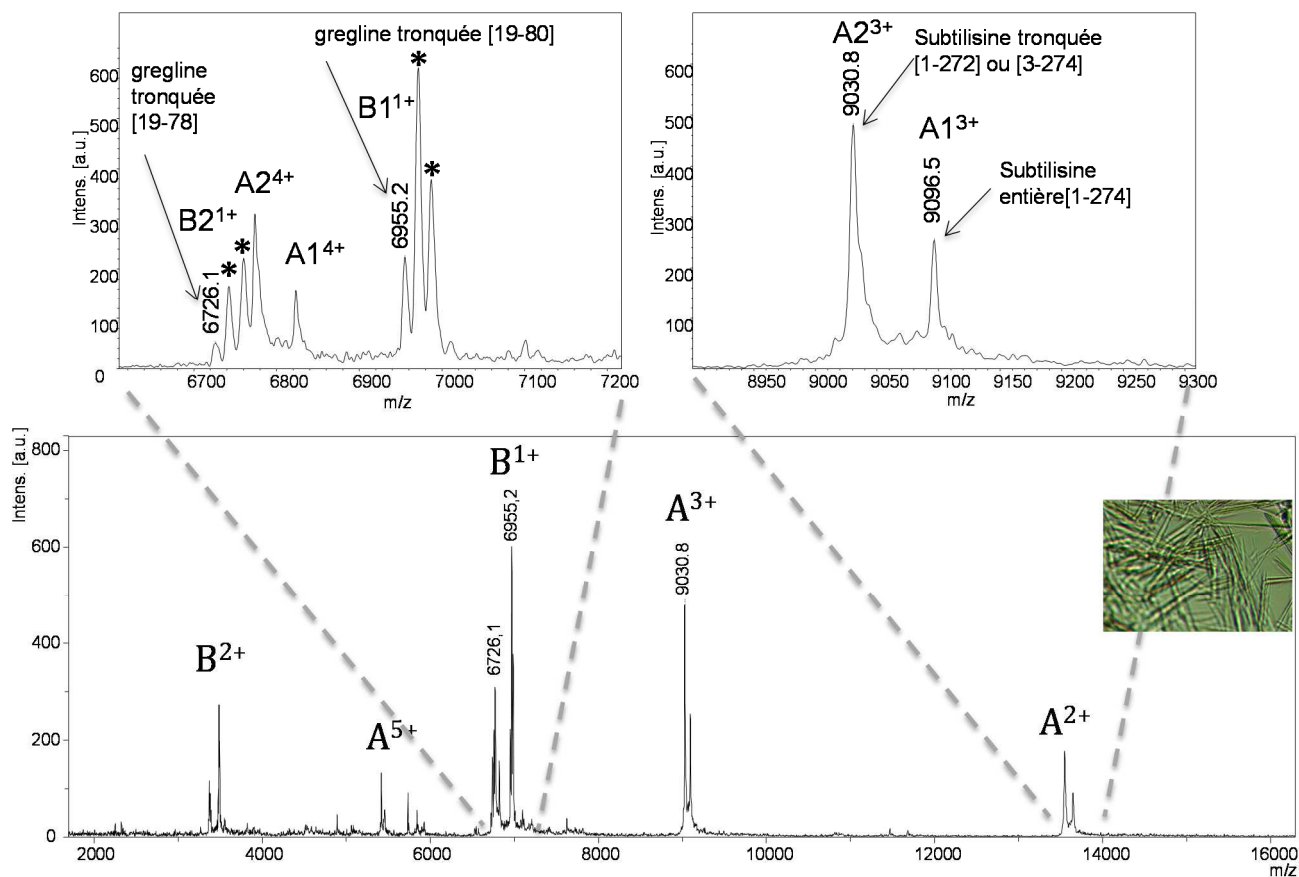
**Figure 37. Séquence de la grégligne issue des bases de données (DB) et corrigée par spectrométrie de masse (MS).** Les résidus corrigés sont en rouge souligné.

	10	20	30	40	50	60	70	80	83
<i>DB</i>	SEDDGSVSPÉ	SQEMSYLELP	LPSISPLIYA	PVCVEDSNSD	FYLFVNECEV	RKCGCEAGFV	YTFVPREMCK	ATTSLCPMQT	KSS
<i>MS</i>	SEDDGS <u>AS</u> PE	SQEMSY <u>I</u> ELP	<u>C</u> PSI <u>C</u> PLIYA	PVCVEDSN <u>Q</u>	FYLFVNECEV	RKCGCEAGFV	YTFVPREMCK	ATTSLCPMQT	KSS

+ 472.146 Da

Afin de savoir si les zones 1-20 et 79-83 sont désordonnées ou simplement absentes, les cristaux du complexe grégligne-subtilisine ont été analysés par spectrométrie de masse MALDI-TOF (Figure 38). Les résultats indiquent que la grégligne est tronquée en N-terminal et qu'il manque 3 à 5 résidus en C-terminal : observation des  $m/z$  6726.1 et 6955.2 annotés à la grégligne [19-78] et [19-80], respectivement. Les formes une fois et deux fois oxydées de ces espèces ont également été observées. La subtilisine est observée intacte ( $m/z$  9096.5) et tronquée [1-272] ou [3-274] ( $m/z$  9030.8).

**Figure 38. Analyse MALDI-TOF des cristaux du complexe grégligne-subtilisine.** Les cristaux ont été lavés avec du tampon de cristallisation puis avec de l'eau. Ils ont ensuite été solubilisés dans 2  $\mu\text{L}$  d'eau avant d'être dilué au  $1/10^{\text{ème}}$  dans une solution de matrice HCCA/FWI. Le mélange matrice-échantillon a été déposé selon la méthode de la couche ultra-mince (Cadene and Chait, 2000; Gabant and Cadene, 2008) et les dépôts ont été lavés avec du TFA 0.1% avant acquisition. Les acquisitions ont été réalisées en mode linéaire positif. Les masses moyennes observées sont A1 = 27287 Da, A2 = 27089 Da, B1 = 6954.2 Da et B2 = 6725.1 Da correspondant respectivement à la subtilisine entière [1-274] ( $M_{\text{moyenne théorique}} = 27288$  Da), subtilisine tronquée [1-272] ou [3-274] ( $M_{\text{moyenne théorique}} = 27089$  Da), grégligne tronquée [19-80] ( $M_{\text{moyenne théorique}} = 6954.1$  Da) et grégligne tronquée [19-78] ( $M_{\text{moyenne théorique}} = 6724.9$  Da). Les oxydations sont représentées par un astérisque.



## 4.2 Détermination de la formule brute de la modification inconnue de la grégligne

### 4.2.1 Interprétation du spectre pseudoMS3 de la modification inconnue

La Figure S3 de l'article indique que la modification contient un hexose (perte de neutre de 162.0528 Da correspondant à  $\text{C}_6\text{H}_{10}\text{O}_5$ ) et des groupements alcool ou acide (pertes de neutres de 18.0106 Da correspondant à  $\text{H}_2\text{O}$ ).

De plus, nous observons un changement de parité parmi les ions fragments ( $m/z$  150, 168, 186 vs 215, 233, 257, 275, 293, 311) ce qui est informatif sur la présence d'azote dans la molécule. En effet,

d'après la « règle de l'azote<sup>2</sup>», le fragment de  $m/z$  473 correspondant au  $[M+H]^+$  de la modification indique une masse M paire de 472 Da, ce qui signifie la présence d'un nombre pair d'azote : 0, 2, 4... L'observation d'un changement de parité parmi les ions fragments nous amène à deux interprétations possibles :

- 1) si tous les ions fragment sont de forme  $[M+H]^+$ , le changement de parité des ions fragments est lié à une variation du nombre d'atomes d'azote. Les M impaires ( $m/z$  150, 168, 186) indiqueraient ainsi la présence d'un nombre impair d'azotes, et donc d'au moins un atome d'azote. Comme l'ion parent possède un nombre pair d'azotes, si l'on exclu la possibilité que ce nombre soit nul, alors le nombre d'azotes dans l'ion parent doit être supérieur ou égal à deux.
- 2) le changement de parité des ions fragments est lié à la présence d'ions radicalaires. Les  $m/z$  paires 150, 168 et 186 correspondraient alors à des  $M^{\bullet+}$  (M paire) et le nombre d'azotes dans l'ion parent pourrait être égal à zéro, deux, quatre...

Par ailleurs, nous avons également examiné le défaut de masse<sup>3</sup> de la modification inconnue qui est de +0.1535 ( $473.1535 - 473 = 0.1535$ ). Ce dernier est plutôt faible pour cette masse sachant que le défaut de masse moyen pour une molécule constituée uniquement de C et H est de + 0.1 par 100 Da. En considérant les défauts de masse des éléments les plus courants dans les biomolécules (Tableau 5), le défaut de masse de +0.1535 au  $m/z$  473 est cohérent avec la présence accrue d'oxygène qui a un défaut de masse négatif de -0.0051 par atome d'oxygène. La présence d'autres hétéroatomes à défaut de masse négatif tels que le soufre (défaut de masse de -0.0279) ou le phosphore (défaut de masse de -0.0262) est également possible. Enfin, la présence de plusieurs insaturations peut également contribuer à ce faible défaut de masse de +0.1535.

Afin d'approfondir la caractérisation de cette modification, la recherche de sa formule brute a été entreprise.

---

<sup>2</sup> La règle de l'azote permet de déterminer si une molécule organique contient un nombre pair ou impair d'atome d'azote. Elle repose sur le fait que la plupart des éléments présents dans les molécules, excepté l'azote, ont une valence et une masse nominale de même parité: ainsi le carbone 12 a une valence de 4, l'hydrogène 1 a une valence de 1, l'oxygène 16 a une valence de 2, alors que l'azote 14 a une valence de 3. Ainsi, une molécule organique de masse M paire aura un nombre pair d'atome d'azote (0, 2, 4,...) et une molécule de masse impaire aura un nombre impair d'atome d'azote (1, 3, 5,...). Cette règle de l'azote de fonctionne pas avec une molécule inorganique.

<sup>3</sup> Le défaut de masse, en spectrométrie de masse, est la différence entre la masse exacte (ou masse monoisotopique) d'un élément, d'une molécule ou d'un ion et sa masse nominale. La masse nominale correspond à la somme du nombre de masse A des isotopes les plus abondants de chaque atome c'est-à-dire la somme du nombre de neutron et de proton. Pour exemple, l'hydrogène a une masse nominale de 1 et contient un proton. Le carbone a une masse nominale de 12 et contient six protons et six neutrons. Ainsi une molécule comme le glucose, de formule brute  $C_6H_{12}O_6$ , a une masse monoisotopique calculée de 180.06339 Da, une masse nominale de 180 et un défaut de masse de +0.06339.

Pour information, la définition du défaut de masse en physique est différente et correspond à la différence entre la masse d'un atome et la somme des masses des protons, neutrons et électrons qui le constituent. Ce défaut de masse, qui varie pour chaque élément, représente l'énergie nécessaire pour la formation et la stabilisation du noyau, c'est l'énergie de liaison d'un noyau (Thurman and Ferrer, 2010).

**Tableau 5. Défauts de masse des éléments les plus courants dans les biomolécules**

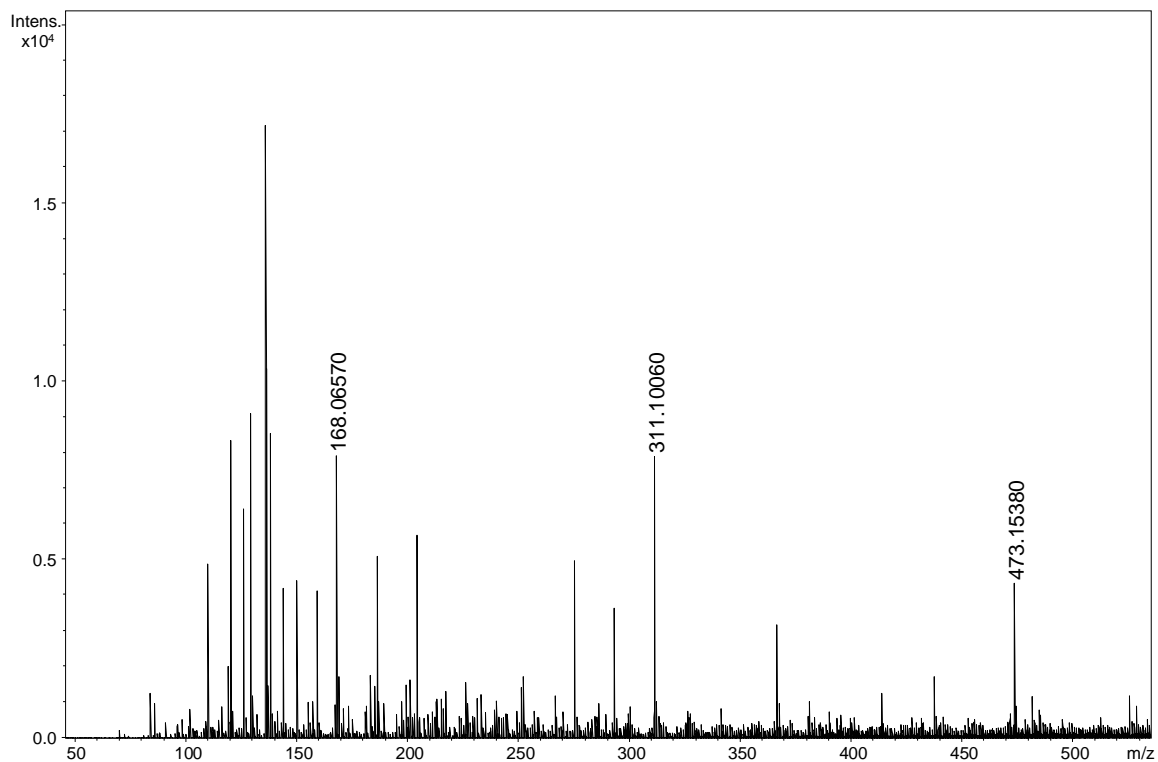
Elément	Symbole	Masse Nominale	Masse Exacte	Abondance	Défaut de masse	
Oxygène	<sup>16</sup> O	16	15.99491	0.99757	-0.00509	défaut de masse < 0
	<sup>17</sup> O	17	16.99913	0.00038	-0.00087	
	<sup>18</sup> O	18	17.99916	0.00205	-0.00084	
Phosphore	<sup>31</sup> P	31	30.97376	1.00000	-0.02624	
Soufre	<sup>32</sup> S	32	31.97207	0.94930	-0.02793	
	<sup>33</sup> S	33	32.97146	0.00760	-0.02854	
	<sup>34</sup> S	34	33.96787	0.04290	-0.03213	
Carbone	<sup>12</sup> C	12	12.00000	0.98930	0.00000	défaut de masse = 0
	<sup>13</sup> C	13	13.00335	0.01070	0.00335	défaut de masse > 0
Hydrogène	H	1	1.00783	0.99989	0.00783	
	<sup>2</sup> H ou D	2	2.01410	0.00012	0.01410	
Azote	<sup>14</sup> N	14	14.00307	0.99632	0.00307	
	<sup>15</sup> N	15	15.00011	0.00368	0.00011	

#### 4.2.2 Recherche de la formule brute de la modification inconnue

Les spectres relatifs à la modification, présentés dans la publication, sont des spectres MS/MS qui sont moins adaptés que les spectres MS à la génération de formule brute du fait d'une précision de mesure insuffisante : < 2 ppm en MS et pouvant aller jusqu'à 0.01 Da en MS/MS soit une erreur de 21 ppm au  $m/z$  473. En effet, afin de réduire le risque de proposer une formule brute fautive, il est primordial d'avoir une précision de masse élevée, c'est-à-dire une valeur de ppm faible. La précision de mesure dégradée en MS/MS est liée à trois facteurs : 1) l'absence de calibrant interne (« lock-mass ») dans les spectres MS/MS, 2) la faible intensité de certains pics et 3) le peu d'accumulation du signal (< 200 msec) car les spectres ont été obtenus en nanoLC-HRMS.

Une deuxième expérience a ainsi été conduite. Elle a été effectuée en infusion afin d'accumuler du signal et d'optimiser les paramètres d'acquisition (ionisation, transmission, fragmentation) pour le  $m/z$  473. La stratégie pseudo-MS3 (ISD + CID) présentée dans la publication a cette fois-ci été utilisée pour l'analyse directe de la gréglie entière. Pour éviter une influence éventuelle de la structure repliée de la gréglie sur la dissociation de la modification dans la source, les ponts disulfures ont été préalablement réduits par une incubation dans 2 mM de TCEP pendant une heure à 50°C. La modification inconnue a ainsi été générée dans la source (ISD) puis son  $m/z$  a été mesuré précisément en MS grâce à la présence d'un calibrant interne (« lock-mass »). Deux autres  $m/z$ , 168 et 311, correspondant à des fragments de la modification ont également été détectés (Figure 39).

**Figure 39. Spectre ISD-HRMS de la gréline réduite au TCEP et dessalée-concentrée par ZipTip C18.** Le spectre correspond à la moyenne des spectres acquis pendant 2.5 min. Trois ions caractéristiques de la modification inconnue sont observés en MS.



Le Tableau 6 regroupe les différentes formules brutes obtenues pour ces trois ions en utilisant le logiciel Smart Formula (Bruker Daltonics) et en considérant les éléments C, H, O, N, S et P.



**Tableau 6. Formules brutes générées à partir des  $m/z$  168.0657, 311.1006 et 473.1538 relatifs à la modification inconnue en utilisant le logiciel SmartFormula (Bruker).** Les éléments C, H, O, N, S et P ont été considérés. La tolérance a été fixée à 2 ppm. Le score mSigma rend compte de la probabilité que la distribution isotopique observée corresponde à la distribution théorique d'une formule brute donnée. Plus le score mSigma est faible et plus cette probabilité est grande. Le nombre d'insaturation pour chaque formule est affiché : rdb pour « ring and double bond ». Les configurations électroniques paire (« even ») et impaire (« odd ») ont été examinées. La règle de l'azote (N-rule) a été vérifiée pour chaque formule. Les formules surlignées en bleu sont les formules retenues.

Meas. $m/z$	#	Formula	$m/z$	err [ppm]	mSigma	rdb	e <sup>-</sup> Conf	N-Rule
168.0657	1	C 8 H 10 N O 3	168.06552	-1.1	12.1	4.5	even	ok
311.1006	1	C 6 H 11 N 14 S	311.10063	0.09	6.4	8.5	even	ok
	2	C 10 H 20 N 2 O 7 P	311.10026	-1.09	13.2	2.5	even	ok
	3	C 7 H 17 N 7 O 5 S	311.10064	0.11	15.6	3	odd	ok
	4	C 9 H 14 N 9 O 2 P	311.10026	-1.11	15.9	8	odd	ok
473.1538	1	C 13 H 27 N 7 O 10 S	473.15346	-0.7	16.4	4	odd	ok
	2	C 16 H 30 N 2 O 12 P	473.15309	-1.5	16.4	3.5	even	ok
	3	C 17 H 26 N 6 O 8 P	473.15442	1.33	17.2	8.5	even	ok
	4	C 15 H 24 N 9 O 7 P	473.15308	-1.51	20.5	9	odd	ok
	5	C 21 H 23 N 5 O 8	473.15411	0.67	22.8	13	odd	ok
	6	C 16 H 20 N 13 O 3 P	473.15442	1.32	23.6	14	odd	ok
	7	C 14 H 18 N 16 O 2 P	473.15308	-1.52	24.7	14.5	even	ok
	8	C 23 H 28 N 3 O 4 P S	473.15327	-1.12	33.9	12	odd	ok
	9	C 20 H 17 N 12 O 3	473.15411	0.66	35.9	18.5	even	ok
	10	C 25 H 30 O 5 P S	473.15461	1.72	39.6	11.5	even	ok
	11	C 24 H 24 N 7 P S	473.1546	1.71	45.8	17	odd	ok
	12	C 27 H 25 N 2 O 4 S	473.15295	-1.78	54.6	16.5	even	ok
	13	C 28 H 21 N 6 S	473.15429	1.05	66.7	21.5	even	ok
	14	C 31 H 24 N O 2 P	473.15392	0.26	67	21	odd	ok
	15	C 35 H 21 O 2	473.15361	-0.4	86.8	25.5	even	ok

Pour le  $m/z$  168, il n'y a qu'une seule formule générée  $C_8H_{10}NO_3$  et ce jusqu'à une tolérance de 9 ppm ce qui consolide l'annotation. On notera que la parité électronique est paire (« even ») ce qui correspond à l'interprétation 1) présentée au paragraphe 4.2.1.

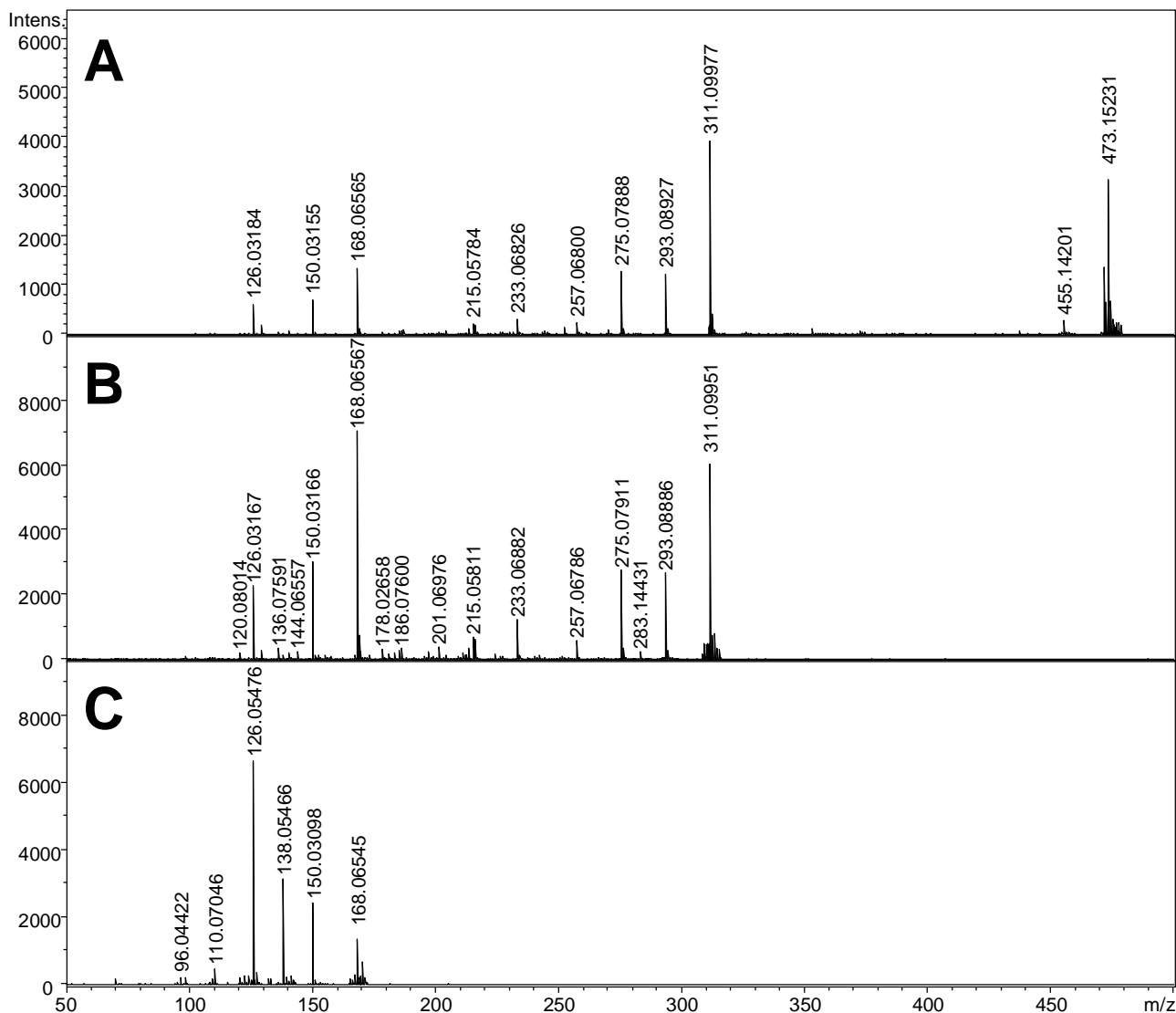
Pour le  $m/z$  311, quatre formules brutes sont proposées. Sur la base des trois pertes de neutre d' $H_2O$  observées pour le  $m/z$  311 (Figure S3 de la publication), qui suggère la présence d'un minimum de trois atomes d'oxygène, les formules annotées N°2 et N°3 sont possibles. En restant cohérent en termes de configuration électronique, c'est la formule brute  $C_{10}H_{20}N_2O_7P$ , correspondant à la configuration paire, qui est retenue.

Pour le  $m/z$  473, quatorze formules brutes sont proposées. En considérant un nombre minimum de huit atomes d'oxygène, correspondant à la somme des trois pertes d' $H_2O$  du  $m/z$  311 et de la perte de l'hexose ( $C_6H_{10}O_5$ ), les formules annotées N°1, 2, 3 et 5 sont possibles. En examinant le nombre d'insaturation (colonne « rdb »), les formules N°3 et 5, avec un nombre d'insaturation élevé de 9 (rdb

= 8.5) et 13 (rdb = 13) respectivement, sont peu probables. En prenant en compte la présence de phosphore dans le  $m/z$  311 et la configuration électronique paire, c'est la formule brute  $C_{16}H_{30}N_2O_{12}P$  qui est retenue.

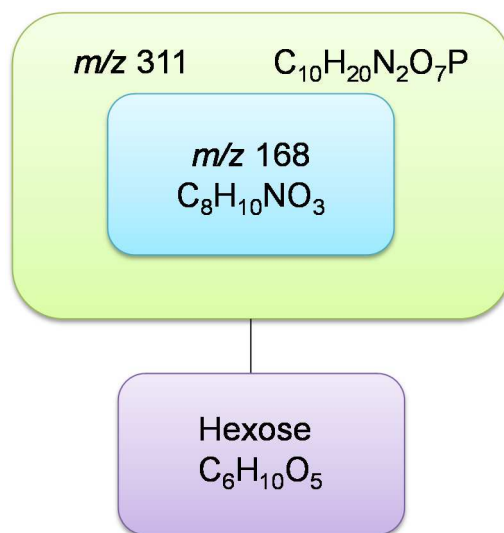
La modification intacte et les ions fragments 311 et 168 ont également été fragmentés en pseudo MS3 afin d'apporter le maximum d'information structurale (Figure 40).

**Figure 40. Spectres pseudo MS3 (ISD + CID) des  $m/z$  473 (modification intacte), 311 et 168 (fragments de la modification).** A, spectre pseudo MS3 de la modification intacte. B, spectre pseudo MS3 du fragment de  $m/z$  311. C, spectre pseudo MS3 du fragment de  $m/z$  168.



Ces spectres indiquent une fragmentation en « poupée russe » avec 473 qui donne 311 (perte d'un hexose) qui donne 168 (Figure 41).

Figure 41. Représentation graphique de la modification de  $m/z$  473.1538.



L'analyse de ces données de fragmentation avec le logiciel SmartFormula3D (Bruker) est cohérente avec les formules brutes proposées pour les  $m/z$  168, 311 et 473 (Tableau 7).

**Tableau 7. Formules brutes générées pour les ions fragments des  $m/z$  168, 311 et 473 en utilisant le logiciel SmartFormula3D (Bruker).** Les formules brutes sont dans l'encadré rouge.

Ion précurseur	Formules brutes pour les ions fragments																																																																																																																													
168	<p><b>Precursor Spectrum</b> +MS2(168.06000), 9.75-10.76min #(375-395)</p> <table border="1"> <thead> <tr> <th>m/z</th> <th>Intensity</th> <th>FWHM</th> <th>S/N</th> </tr> </thead> <tbody> <tr><td>165.07038</td><td>168.3</td><td>0.01169</td><td>319.8</td></tr> <tr><td>166.09489</td><td>129.3</td><td>0.01412</td><td>246.1</td></tr> <tr><td>167.08087</td><td>312.2</td><td>0.01928</td><td>594.8</td></tr> <tr><td>167.11699</td><td>234.9</td><td>0.00921</td><td>447.7</td></tr> <tr><td>168.04346</td><td>292.9</td><td>0.01305</td><td>558.7</td></tr> <tr><td><b>168.06545</b></td><td><b>1362.6</b></td><td><b>0.00992</b></td><td><b>2599.9</b></td></tr> <tr><td>168.10122</td><td>190.7</td><td>0.01843</td><td>363.9</td></tr> <tr><td><b>169.07240</b></td><td><b>196.7</b></td><td><b>0.02161</b></td><td><b>375.9</b></td></tr> <tr><td>169.09690</td><td>137.7</td><td>0.01428</td><td>263.2</td></tr> <tr><td>169.13281</td><td>274.4</td><td>0.00984</td><td>524.3</td></tr> <tr><td>170.05922</td><td>666.8</td><td>0.00872</td><td>1276.1</td></tr> <tr><td>170.09188</td><td>125.0</td><td>0.02341</td><td>239.2</td></tr> <tr><td>171.08041</td><td>110.9</td><td>0.02972</td><td>212.5</td></tr> <tr><td>171.11178</td><td>189.4</td><td>0.01579</td><td>362.9</td></tr> </tbody> </table>	m/z	Intensity	FWHM	S/N	165.07038	168.3	0.01169	319.8	166.09489	129.3	0.01412	246.1	167.08087	312.2	0.01928	594.8	167.11699	234.9	0.00921	447.7	168.04346	292.9	0.01305	558.7	<b>168.06545</b>	<b>1362.6</b>	<b>0.00992</b>	<b>2599.9</b>	168.10122	190.7	0.01843	363.9	<b>169.07240</b>	<b>196.7</b>	<b>0.02161</b>	<b>375.9</b>	169.09690	137.7	0.01428	263.2	169.13281	274.4	0.00984	524.3	170.05922	666.8	0.00872	1276.1	170.09188	125.0	0.02341	239.2	171.08041	110.9	0.02972	212.5	171.11178	189.4	0.01579	362.9	<p><b>Fragment Spectrum</b> +MS2(168.06000), 9.75-10.76min #(375-395)</p> <table border="1"> <thead> <tr> <th>m/z</th> <th>Intensity</th> <th>FWHM</th> <th>S/N</th> </tr> </thead> <tbody> <tr><td>70.06502</td><td>166.7</td><td>0.00566</td><td>285.6</td></tr> <tr><td>96.04422</td><td>207.6</td><td>0.00683</td><td>358.4</td></tr> <tr><td>98.05988</td><td>194.8</td><td>0.00793</td><td>337.2</td></tr> <tr><td>109.00414</td><td>158.5</td><td>0.00566</td><td>278.5</td></tr> <tr><td>110.07046</td><td>458.5</td><td>0.00646</td><td>806.6</td></tr> <tr><td>120.08020</td><td>193.6</td><td>0.00743</td><td>345.2</td></tr> <tr><td>122.06058</td><td>285.8</td><td>0.00735</td><td>511.0</td></tr> <tr><td>124.03854</td><td>256.0</td><td>0.00785</td><td>459.0</td></tr> <tr><td>125.07009</td><td>136.1</td><td>0.00827</td><td>244.4</td></tr> <tr><td><b>126.05476</b></td><td><b>6652.0</b></td><td><b>0.00775</b></td><td><b>11959.0</b></td></tr> <tr><td>126.08887</td><td>125.6</td><td>0.00563</td><td>225.9</td></tr> <tr><td><b>127.05710</b></td><td><b>359.7</b></td><td><b>0.00980</b></td><td><b>647.6</b></td></tr> <tr><td>132.02012</td><td>170.0</td><td>0.00863</td><td>308.1</td></tr> <tr><td>133.00448</td><td>155.6</td><td>0.00738</td><td>282.4</td></tr> <tr><td><b>138.05466</b></td><td><b>3141.5</b></td><td><b>0.00723</b></td><td><b>5742.5</b></td></tr> </tbody> </table>	m/z	Intensity	FWHM	S/N	70.06502	166.7	0.00566	285.6	96.04422	207.6	0.00683	358.4	98.05988	194.8	0.00793	337.2	109.00414	158.5	0.00566	278.5	110.07046	458.5	0.00646	806.6	120.08020	193.6	0.00743	345.2	122.06058	285.8	0.00735	511.0	124.03854	256.0	0.00785	459.0	125.07009	136.1	0.00827	244.4	<b>126.05476</b>	<b>6652.0</b>	<b>0.00775</b>	<b>11959.0</b>	126.08887	125.6	0.00563	225.9	<b>127.05710</b>	<b>359.7</b>	<b>0.00980</b>	<b>647.6</b>	132.02012	170.0	0.00863	308.1	133.00448	155.6	0.00738	282.4	<b>138.05466</b>	<b>3141.5</b>	<b>0.00723</b>	<b>5742.5</b>
	m/z	Intensity	FWHM	S/N																																																																																																																										
	165.07038	168.3	0.01169	319.8																																																																																																																										
	166.09489	129.3	0.01412	246.1																																																																																																																										
167.08087	312.2	0.01928	594.8																																																																																																																											
167.11699	234.9	0.00921	447.7																																																																																																																											
168.04346	292.9	0.01305	558.7																																																																																																																											
<b>168.06545</b>	<b>1362.6</b>	<b>0.00992</b>	<b>2599.9</b>																																																																																																																											
168.10122	190.7	0.01843	363.9																																																																																																																											
<b>169.07240</b>	<b>196.7</b>	<b>0.02161</b>	<b>375.9</b>																																																																																																																											
169.09690	137.7	0.01428	263.2																																																																																																																											
169.13281	274.4	0.00984	524.3																																																																																																																											
170.05922	666.8	0.00872	1276.1																																																																																																																											
170.09188	125.0	0.02341	239.2																																																																																																																											
171.08041	110.9	0.02972	212.5																																																																																																																											
171.11178	189.4	0.01579	362.9																																																																																																																											
m/z	Intensity	FWHM	S/N																																																																																																																											
70.06502	166.7	0.00566	285.6																																																																																																																											
96.04422	207.6	0.00683	358.4																																																																																																																											
98.05988	194.8	0.00793	337.2																																																																																																																											
109.00414	158.5	0.00566	278.5																																																																																																																											
110.07046	458.5	0.00646	806.6																																																																																																																											
120.08020	193.6	0.00743	345.2																																																																																																																											
122.06058	285.8	0.00735	511.0																																																																																																																											
124.03854	256.0	0.00785	459.0																																																																																																																											
125.07009	136.1	0.00827	244.4																																																																																																																											
<b>126.05476</b>	<b>6652.0</b>	<b>0.00775</b>	<b>11959.0</b>																																																																																																																											
126.08887	125.6	0.00563	225.9																																																																																																																											
<b>127.05710</b>	<b>359.7</b>	<b>0.00980</b>	<b>647.6</b>																																																																																																																											
132.02012	170.0	0.00863	308.1																																																																																																																											
133.00448	155.6	0.00738	282.4																																																																																																																											
<b>138.05466</b>	<b>3141.5</b>	<b>0.00723</b>	<b>5742.5</b>																																																																																																																											
<table border="1"> <thead> <tr> <th>SumFormula</th> <th>m/z calc</th> <th>err[mDa]</th> <th>err[ppm]</th> <th>mSigma</th> <th>eConf</th> <th>Comb...</th> <th>Co...</th> <th>Comb....</th> </tr> </thead> <tbody> <tr> <td><input checked="" type="checkbox"/> C<sub>8</sub>H<sub>10</sub>N<sub>0</sub>O<sub>3</sub></td> <td>168.06552</td> <td>0.07</td> <td>0.40</td> <td>31.5</td> <td>even</td> <td>22.6</td> <td>0.23</td> <td>0.16</td> </tr> </tbody> </table>	SumFormula	m/z calc	err[mDa]	err[ppm]	mSigma	eConf	Comb...	Co...	Comb....	<input checked="" type="checkbox"/> C <sub>8</sub> H <sub>10</sub> N <sub>0</sub> O <sub>3</sub>	168.06552	0.07	0.40	31.5	even	22.6	0.23	0.16	<table border="1"> <thead> <tr> <th>SumFormula</th> <th>SumFormula Loss</th> <th>m/z Loss</th> <th>err[mDa] ...</th> <th>D...</th> <th>Intensity</th> <th>m/z calc</th> <th>err[mDa]</th> <th>m</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/> C<sub>8</sub>H<sub>6</sub>O<sub>3</sub></td> <td>H<sub>4</sub>N</td> <td>18.03447</td> <td>-0.10</td> <td></td> <td>2438</td> <td>150.03115</td> <td>0.16</td> <td></td> </tr> <tr> <td><input type="checkbox"/> C<sub>7</sub>H<sub>8</sub>N<sub>0</sub>O<sub>2</sub></td> <td>CH<sub>2</sub>O</td> <td>30.01079</td> <td>-0.22</td> <td></td> <td>3141</td> <td>138.05495</td> <td>0.29</td> <td></td> </tr> <tr> <td><input type="checkbox"/> C<sub>6</sub>H<sub>8</sub>N<sub>0</sub>O<sub>2</sub></td> <td>C<sub>2</sub>H<sub>2</sub>O</td> <td>42.01069</td> <td>-0.13</td> <td></td> <td>6652</td> <td>126.05495</td> <td>0.20</td> <td></td> </tr> </tbody> </table>	SumFormula	SumFormula Loss	m/z Loss	err[mDa] ...	D...	Intensity	m/z calc	err[mDa]	m	<input type="checkbox"/> C <sub>8</sub> H <sub>6</sub> O <sub>3</sub>	H <sub>4</sub> N	18.03447	-0.10		2438	150.03115	0.16		<input type="checkbox"/> C <sub>7</sub> H <sub>8</sub> N <sub>0</sub> O <sub>2</sub>	CH <sub>2</sub> O	30.01079	-0.22		3141	138.05495	0.29		<input type="checkbox"/> C <sub>6</sub> H <sub>8</sub> N <sub>0</sub> O <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> O	42.01069	-0.13		6652	126.05495	0.20																																																																								
SumFormula	m/z calc	err[mDa]	err[ppm]	mSigma	eConf	Comb...	Co...	Comb....																																																																																																																						
<input checked="" type="checkbox"/> C <sub>8</sub> H <sub>10</sub> N <sub>0</sub> O <sub>3</sub>	168.06552	0.07	0.40	31.5	even	22.6	0.23	0.16																																																																																																																						
SumFormula	SumFormula Loss	m/z Loss	err[mDa] ...	D...	Intensity	m/z calc	err[mDa]	m																																																																																																																						
<input type="checkbox"/> C <sub>8</sub> H <sub>6</sub> O <sub>3</sub>	H <sub>4</sub> N	18.03447	-0.10		2438	150.03115	0.16																																																																																																																							
<input type="checkbox"/> C <sub>7</sub> H <sub>8</sub> N <sub>0</sub> O <sub>2</sub>	CH <sub>2</sub> O	30.01079	-0.22		3141	138.05495	0.29																																																																																																																							
<input type="checkbox"/> C <sub>6</sub> H <sub>8</sub> N <sub>0</sub> O <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> O	42.01069	-0.13		6652	126.05495	0.20																																																																																																																							

Ion précurseur

### Formules brutes pour les ions fragments

311

**Precursor Spectrum**

+MS2(311.10001), 7.78-8.26min #(341-350)

m/z	Intensity	FWHM	S/N
308.16195	164.6	0.03259	95.7
309.08585	164.9	0.02977	97.4
309.13068	391.7	0.02650	231.6
309.16241	486.5	0.03573	287.8
309.19917	283.1	0.02599	167.6
310.10874	237.5	0.02944	142.9
310.16167	484.6	0.02587	291.9
310.20086	218.3	0.02864	131.6
311.02843	214.4	0.03481	131.2
311.09951	6038.3	0.01520	3699.8
311.13586	580.4	0.02944	355.8
311.17539	461.8	0.03666	283.3
312.10454	730.2	0.02637	455.8
312.15796	417.3	0.02916	260.8
313.05870	125.9	0.02619	80.0

**Fragment Spectrum**

+MS2(311.10001), 7.78-8.26min #(341-350)

m/z	Intensity	FWHM	S/N
234.05341	127.5	0.01074	47.8
240.13415	115.9	0.01266	43.1
242.11517	128.4	0.01333	47.6
251.07835	105.1	0.01556	38.4
257.06786	580.4	0.01144	210.2
275.07911	2780.0	0.01303	1023.1
276.08013	357.7	0.01309	133.1
283.14431	252.9	0.01255	102.2
293.08886	2679.0	0.01402	1231.9
294.09294	271.0	0.01570	126.4
308.16195	164.6	0.03259	95.7
309.08585	164.9	0.02977	97.4
309.13068	391.7	0.02650	231.6
309.16241	486.5	0.03573	287.8
309.19917	283.1	0.02599	167.6

SumFormula	m/z calc	err[mDa]	err[ppm]	mSigma	eConf	Comb...	Co...	Comb...
<input checked="" type="checkbox"/> C 10 H 20 N 2 O 7 P	311.10026	0.76	2.43	20.5	even	32.9	0.50	0.99
<input checked="" type="checkbox"/> C 12 H 15 N 4 O 6	311.09861	-0.90	-2.88	23.7	even	47.9	0.60	0.54
<input checked="" type="checkbox"/> C 8 H 18 N 5 O 6 P	311.09892	-0.59	-1.89	25.7	odd	33.2	0.60	0.46
<input checked="" type="checkbox"/> C 9 H 14 N 9 O 2 P	311.10026	0.75	2.41	26.0	odd	43.9	0.74	0.64
<input checked="" type="checkbox"/> C 14 H 17 N O 7	311.09995	0.45	1.43	26.2	odd	46.1	0.45	0.77
<input checked="" type="checkbox"/> C 7 H 12 N 12 O P	311.09892	-0.59	-1.90	32.0	even	41.5	0.60	0.64
<input checked="" type="checkbox"/> C 6 H 17 N 7 O 6 Si	311.10041	0.90	2.90	32.3	odd	53.1	0.63	0.90
<input checked="" type="checkbox"/> C 11 H 9 N 11 O	311.09861	-0.90	-2.90	33.9	odd	60.7	0.86	0.64
<input checked="" type="checkbox"/> C 7 H 17 N 7 O 5 S	311.10064	1.13	3.64	34.3	odd	44.8	0.81	1.00
<input checked="" type="checkbox"/> C 13 H 11 N 8 O 2	311.09995	0.44	1.42	35.9	even	61.1	0.51	0.58

SumFormula	SumFormula Loss	m/z Loss	err[mDa] ...	D...	Intensity	m/z calc	err[mDa]	m
<input type="checkbox"/> C 10 H 18 N 2 O 6 P	H 2 O	18.01065	-0.08		2679	293.08970	0.84	
<input type="checkbox"/> C 10 H 16 N 2 O 5 P	H 4 O 2	36.02040	0.73		2780	275.07913	0.03	
<input type="checkbox"/> C 8 H 14 N 2 O 4 P	C 2 H 6 O 3	78.03068	1.01		1217	233.06857	-0.25	
<input type="checkbox"/> C 8 H 10 N O 3	C 2 H 10 N O 4 P	143.03384	0.91		7058	168.06552	-0.15	
<input type="checkbox"/> C 7 H 9 N 2 O P	C 3 H 11 O 6	143.05515	0.41		1096	168.04470	0.34	
<input type="checkbox"/> C 4 H 9 N O 3 P	C 6 H 11 N O 4	161.06784	0.96		3020	150.03146	-0.21	
<input type="checkbox"/> C 8 H 6 O 3	C 2 H 14 N 2 O 4 P	161.06784	1.27		3020	150.03115	-0.52	
<input type="checkbox"/> C 6 H 8 N O 2	C 4 H 12 N O 5 P	185.04414	1.17		793	126.05495	-0.41	
<input type="checkbox"/> C 6 H 6 O 3	C 4 H 14 N 2 O 4 P	185.06783	1.28		2290	126.03115	-0.53	

Ion  
précurseur

### Formules brutes pour les ions fragments

473

Precursor Spectrum										Fragment Spectrum									
+MS2(473.60001), 4.95-6.31min #(285-315)										+MS2(473.60001), 4.95-6.31min #(285-315)									
m/z	Intensity	FWHM	S/N							m/z	Intensity	FWHM	S/N						
471.17900	223.9	0.07304	129.6							126.03184	602.4	0.00701	874.3						
471.22140	303.5	0.04525	175.8							126.05436	120.1	0.00789	174.3						
471.25814	340.7	0.05413	197.5							129.10290	184.1	0.00793	258.0						
471.30694	1361.1	0.02103	789.4							150.03155	698.3	0.00798	792.9						
472.18381	211.4	0.16248	124.0							168.04293	462.6	0.01051	451.5						
472.24582	362.8	0.06090	212.9							168.06565	1354.2	0.00981	1321.5						
472.30937	647.5	0.02178	380.3							169.06831	132.4	0.01063	128.2						
473.15231	3140.4	0.01953	1864.7							213.08697	114.5	0.01116	87.0						
473.22599	469.5	0.05829	279.1							215.05784	214.9	0.01080	161.8						
473.30498	165.7	0.03215	98.6							216.04246	188.7	0.00978	141.5						
474.15672	666.0	0.02567	400.7							233.06826	310.9	0.01227	216.8						
474.23039	339.7	0.06714	204.5							252.06319	142.6	0.01473	92.2						
475.19458	301.3	0.07291	183.7							257.06800	233.1	0.01185	147.9						
475.23122	283.4	0.04088	172.9							275.07888	1262.9	0.01275	765.9						
475.33302	114.2	0.03001	69.8							276.08176	122.4	0.01359	74.4						

SumFormula	m/z calc	err[mDa]	err[ppm]	mSigma	eConf	Comb...	Co...	Comb...		SumFormula	SumFormula Loss	m/z Loss	err[mDa] ...	D...	Intensity	m/z calc	err[mDa]	m
<input checked="" type="checkbox"/> C 14 H 11 N 21	473.15008	-2.23	-4.72	16.2	odd	49.4	1.47	1.9		<input type="checkbox"/> C 10 H 20 N 2 O 7 P	C 6 H 10 O 5	162.05255	0.28		3915	311.10026	0.50	
<input checked="" type="checkbox"/> C 16 H 23 N 7 O 10	473.15009	-2.22	-4.70	20.6	odd	40.3	1.13	2.5		<input type="checkbox"/> C 14 H 17 N O 7	C 2 H 13 N O 5 P	162.05255	0.59		3915	311.09995	0.19	
<input checked="" type="checkbox"/> C 15 H 24 N 9 O 7 P	473.15308	0.77	1.63	20.8	odd	32.5	0.74	1.0		<input type="checkbox"/> C 10 H 18 N 2 O 6 P	C 6 H 12 O 6	180.06305	0.34		1226	293.08970	0.43	
<input checked="" type="checkbox"/> C 17 H 26 N 6 O 8 P	473.15442	2.11	4.46	21.3	even	33.3	0.95	2.2		<input type="checkbox"/> C 14 H 15 N O 6	C 2 H 15 N O 6 P	180.06305	0.65		1226	293.08939	0.12	
<input checked="" type="checkbox"/> C 18 H 25 N 4 O 11	473.15143	-0.88	-1.86	22.3	even	38.1	0.69	0.7		<input type="checkbox"/> C 10 H 16 N 2 O 5 P	C 6 H 14 O 7	198.07343	0.52		1263	275.07913	0.25	
<input checked="" type="checkbox"/> C 13 H 22 N 12 O 6 P	473.15174	-0.57	-1.21	22.4	even	29.3	0.72	0.6		<input type="checkbox"/> C 14 H 13 N O 5	C 2 H 17 N O 7 P	198.07343	0.83		1263	275.07882	-0.06	
<input checked="" type="checkbox"/> C 16 H 30 N 2 O 12 P	473.15309	0.77	1.64	24.6	even	29.0	0.39	0.7		<input type="checkbox"/> C 8 H 10 N O 3	C 8 H 20 N O 9 P	305.08666	0.90		1354	168.06552	-0.13	
<input checked="" type="checkbox"/> C 20 H 27 N O 12	473.15278	0.46	0.98	25.3	odd	35.7	0.30	0.5		<input type="checkbox"/> C 4 H 9 N O 3 P	C 12 H 21 N O 9	323.12076	0.87		698	150.03146	-0.10	
<input checked="" type="checkbox"/> C 17 H 29 O 15	473.15010	-2.22	-4.68	26.6	even	56.7	2.25	1.8		<input type="checkbox"/> C 8 H 6 O 3	C 8 H 24 N 2 O 9 P	323.12076	1.18		698	150.03115	-0.41	
<input checked="" type="checkbox"/> C 21 H 23 N 5 O 8	473.15411	1.80	3.81	30.5	odd	40.8	1.06	1.8										
<input checked="" type="checkbox"/> C 14 H 28 N 5 O 11 P	473.15174	-0.57	-1.20	30.5	odd	30.9	0.65	0.6										
<input checked="" type="checkbox"/> C 19 H 21 N 8 O 7	473.15277	0.46	0.97	34.7	even	40.7	0.82	0.8										
<input checked="" type="checkbox"/> C 15 H 27 N 10 O 2 P 2 S	473.15089	-1.42	-3.01	35.1	even	36.8	0.75	1.8										
<input checked="" type="checkbox"/> C 17 H 19 N 11 O 6	473.15143	-0.88	-1.87	35.5	odd	40.2	0.92	1.2										

### 4.3 Recherche dans les bases de données

Afin d'identifier d'éventuelles molécules déjà décrites ayant la même formule brute que la modification, une recherche dans des bases de données collectant des milliers de molécules et/ou de spectres de masse de molécules a été réalisée. Trois bases de données ont été explorées : Chemspider, METLIN et MassBank mais aucune ne contient la formule brute recherchée ce qui suggère que la modification découverte sur la gréglina est nouvelle.

Sur la base de la présence de phosphore et d'un hexose, l'hypothèse d'une structure analogue à une ancre glycosylphosphatidylinositol (GPI) est considérée mais reste à confirmer à ce jour. L'autre hypothèse majeure est que la molécule contient un atome "exotique" peu souvent rencontré en biologie.

## 5. Perspectives

Afin de localiser le résidu modifié par cette nouvelle modification, une expérience de fragmentation ETD pourrait être effectuée. Cela n'a pas encore été réalisé car il n'y a plus de protéine disponible, celle-ci étant produite par purification après une très longue étape de dissection manuelle de centaines d'ovaires de criquet.

D'autres techniques orthogonales, telles que la RMN, l'analyse élémentaire ou la spectroscopie infrarouge, pourraient apporter des informations précieuses sur la structure de cette nouvelle modification, le pourcentage en C, H, O, N, S, P ou d'autres éléments non considérés jusqu'à présent et l'existence de groupement fonctionnel particulier (N-H, O-H, C=O, C=N, C=C...), respectivement.

Enfin le rôle biologique des phosphorylations et de la nouvelle modification reste à élucider. Elles ne semblent pas être directement liées à l'activité anti-protéase de la gréglina dans la mesure où Gauthier et ses collaborateurs ont montré que la gréglina clivée [27-81] conservait ses propriétés inhibitrices (Brillard-Bourdet *et al.*, 2006). Ces modifications post-traductionnelles pourraient par contre être impliquées dans la localisation et/ou la stabilisation de la gréglina, ou dans le recrutement d'autres protéines pour former des complexes.

## Chapitre 2 Etude du complexe entre la transposase MOS1 et des inhibiteurs



## *Résumé du chapitre 2 :*

*La transposition d'éléments génétiques est un mécanisme essentiel pour la variabilité génétique et l'évolution des génomes, mais il peut également être génotoxique lorsqu'un élément transposable s'intègre dans un gène vital pour la cellule hôte. La protéine MOS1 est une transposase de l'élément mariner-Mos1 qui présente des caractéristiques communes avec l'intégrase du VIH-1. Plusieurs inhibiteurs de MOS1 ont été identifiés et pourraient être utilisés à des fins thérapeutiques pour inhiber les intégrases rétrovirales. En préalable au criblage d'inhibiteurs, les modifications post-traductionnelles de MOS1 ont été détectées et identifiées. Nous montrons que MOS1 est à la fois acétylée et phosphorylée, et que la phosphorylation a un effet direct sur l'activité de la transposase. Ces travaux ont donné lieu à une publication dans Nucleic Acids Research.*

# 1. Contexte

La transposition d'éléments génétiques correspond au déplacement d'une séquence d'ADN, appelée élément transposable, d'une position à une autre dans le génome. Ce mécanisme est une source de variabilité génétique et constitue un outil essentiel pour l'évolution des génomes, mais il peut également être génotoxique lorsqu'un élément transposable s'intègre dans un gène vital pour la cellule hôte. Le séquençage des génomes a révélé que les éléments transposables (actifs et inactifs) représentent une proportion élevée des génomes eucaryotes avec 45% du génome humain, 40% de celui de la souris, 60% de celui du maïs et jusqu'à 77% de celui de la grenouille (Biémont and Vieira, 2006). Initialement considérés comme des parasites génétiques ou de l'ADN inutile (« junk DNA »), il est aujourd'hui admis que certains transposons ont été domestiqués par leur hôte pour fournir de nouvelles fonctions cellulaires. Par exemple, le système de recombinaison V(D)J, qui génère la diversité des anticorps, a très certainement évolué à partir d'un transposon eucaryote (Kapitonov and Jurka, 2005), tout comme la méthyltransférase SETMAR, également appelée metnase, qui est une protéine fusion entre un domaine SET et une transposase *mariner* et qui est impliquée dans la réparation de l'ADN (Beck *et al.*, 2008). Les éléments transposables peuvent également jouer un rôle important dans la régulation épigénétique (Slotkin and Martienssen, 2007). La capacité des transposons à s'intégrer dans le génome est exploitée pour développer des vecteurs non viraux pour la transgénèse (Frokjaer-Jensen *et al.*, 2012) ou la thérapie génique (Hong *et al.*, 2014).

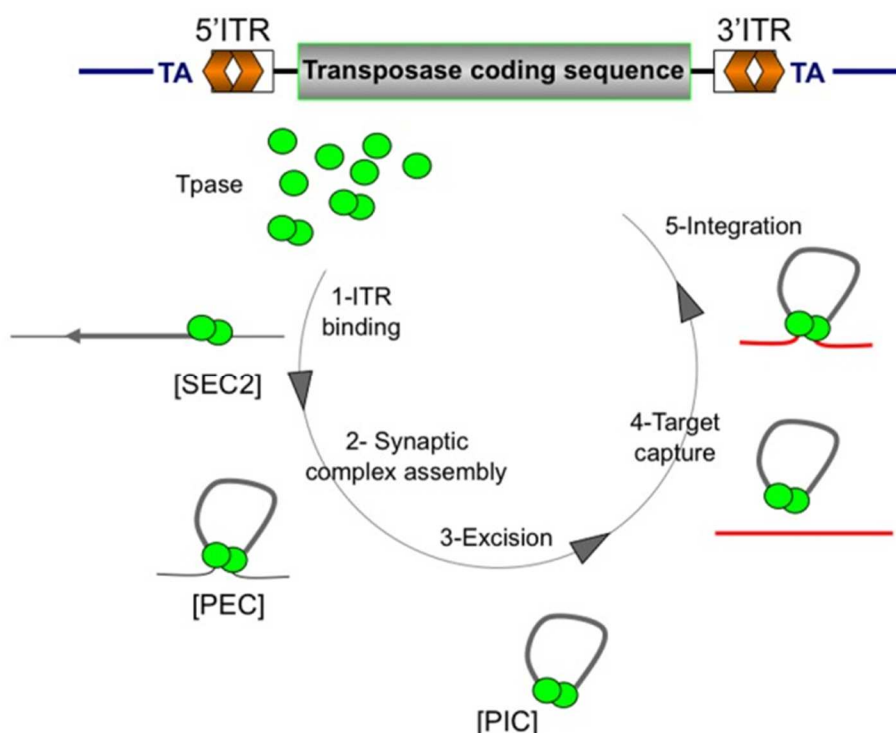
Les éléments transposables présentent une grande diversité chez les organismes vivants. Ils sont classés en deux catégories : les éléments de classe I, appelés rétrotransposons, qui utilisent un intermédiaire ARN et les éléments de classe II qui se transposent directement par un intermédiaire ADN. La transposition des éléments de classe II utilise une enzyme spécifique, appelée transposase, codée par le transposon lui-même, qui est capable d'exciser le transposon de son locus et de l'intégrer ailleurs dans le génome.

La protéine MOS1 est une transposase de l'élément *mariner-Mos1* qui appartient à la superfamille des éléments transposables *Tc1-mariner*, transposons les plus répandus chez les eucaryotes (Miskey *et al.*, 2005). Les éléments *mariner-Mos1* transposent *via* un mécanisme « couper-coller » qui se décline en 5 étapes (Jaillet *et al.*, 2012) (Figure 42) :

- (1) dimérisation des protéines MOS1 au niveau d'une répétition terminale inversée (ITR pour Inverted Terminal Repeat),
- (2) formation du complexe synaptique,
- (3) excision du transposon,
- (4) capture de l'ADN cible,

- (5) intégration de l'élément au niveau d'un site cible.

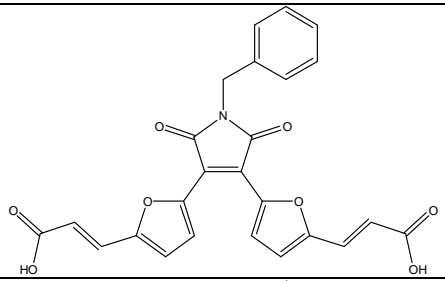
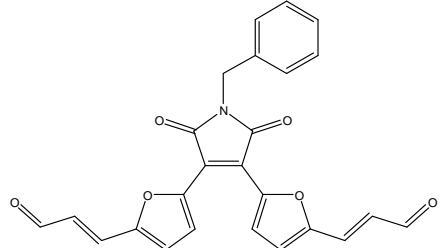
**Figure 42. Cycle de transposition de l'élément transposable mariner (d'après Jaillet *et al.*, 2012).** Une représentation schématique d'un transposon mariner est illustrée en haut de la figure avec les principaux composants que sont la séquence codant pour la transposase (en gris), les répétitions terminales inversées (ITR pour Inverted Terminal Repeat) (en orange) et le dinucléotide TA flanquant l'élément (point de repère pour la transposition). La transposition de Mos1 comprend cinq étapes : (1) dimérisation des protéines MOS1 (en vert) au niveau d'une ITR pour former le complexe SEC2 (Single-end complex 2). (2) L'assemblage du complexe synaptique est obtenu par l'addition du second ITR à SEC2 pour former le complexe PEC (Paired-end complex). (3) Les brins d'ADN sont ensuite clivés par la transposase, c'est l'excision du transposon. Une fois que le complexe PIC (Pre-integration complex) est généré, la capture de l'ADN cible se produit (4), suivie de l'intégration de l'élément au niveau d'un dinucléotide TA (5).



Ce mécanisme est similaire au mécanisme d'intégration des rétrovirus tels que le virus de l'immunodéficience humaine 1 (VIH-1), ce qui fait de MOS1 un bon modèle pour l'étude du mécanisme des intégrases rétrovirales.

L'équipe de C. Augé-Gouillou a identifié plusieurs inhibiteurs de la transposase MOS1 qui peuvent agir sur une étape en particulier (excision par exemple) ou plusieurs étapes (excision + intégration) (Bouchet *et al.*, 2009). Les inhibiteurs étudiés dans ce chapitre sont des dérivés bis(furanyl)*N*-maléimides (Tableau 8).

**Tableau 8. Inhibiteurs de MOS1 étudiés dans ce chapitre.** La réaction suspectée est une addition de Michael au niveau de l'insaturation située entre le noyau furane et la fonction acide carboxylique pour MCV22 et entre le noyau furanyl et la fonction aldéhyde pour MCV41.

Nom	Structure des inhibiteurs	Formule brute	Masse moléculaire (Da)	Formule chimique de l'inhibiteur ajouté	Incrément de masse ajouté	
					Masse théorique monoisotopique (Da)	Masse théorique moyenne (Da)
MCV22		C <sub>25</sub> H <sub>17</sub> NO <sub>8</sub>	459.41	C <sub>25</sub> H <sub>17</sub> NO <sub>8</sub>	459.0954	459.4053
MCV41		C <sub>25</sub> H <sub>17</sub> NO <sub>6</sub>	427.41	C <sub>25</sub> H <sub>17</sub> NO <sub>6</sub>	427.1056	427.4066

Ces inhibiteurs de MOS1 présentent deux intérêts majeurs : ils peuvent être utilisés 1) en recherche fondamentale pour étudier en détail le mécanisme moléculaire de la transposition en bloquant certaines étapes du processus ou encore 2) à des fins thérapeutiques pour inhiber les intégrases rétrovirales telles que l'intégrase du VIH-1 qui présente des caractéristiques communes avec MOS1. Récemment, Wolkowicz *et al.* (2014) ont montré que le médicament antirétroviral Raltegravir, initialement développé contre l'intégrase du VIH-1, inhibe les étapes d'excision et d'intégration du transposon *mariner-Mos1* (Wolkowicz *et al.*, 2014). Les auteurs ont obtenu la structure cristallographique du complexe MOS1-Raltegravir, ce qui permet de mieux comprendre le mode d'action de cet inhibiteur et ainsi de concevoir de meilleurs médicaments. Les inhibiteurs de MOS1 ont une fonction accepteur de Michael potentielle, ce qui a suscité notre intérêt dans notre démarche d'identification de sites de modification par des accepteurs de Michael sur des protéines d'intérêt thérapeutique.

## 2. Objectifs

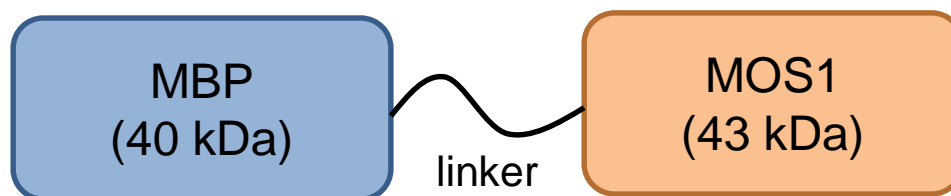
Ce projet collaboratif visait à caractériser par spectrométrie de masse le complexe entre MOS1 et les inhibiteurs MCV22 et MCV41. Plus précisément, il faut 1) confirmer que les inhibiteurs se fixent de manière covalente à MOS1 en complexe ou non avec son ADN cible au vu de leur fonction accepteur

de Michael, 2) déterminer la stœchiométrie des complexes inhibiteur-MOS1 et inhibiteur-MOS1-ADN et 3) localiser le ou les résidu(s) ciblé(s).

### 3. Etude de la protéine MBP-MOS1 produite dans *Escherichia coli*

La transposase a été exprimée en fusion avec la protéine MBP (« Maltose Binding Protein ») pour surmonter des problèmes de solubilité de MOS1 lorsqu'elle est exprimée seule (Figure 43). Cette protéine MBP-MOS1 a été produite, dans un premier temps, dans *Escherichia coli* et sera dénommée MBP-MOS1<sub>E.coli</sub> par la suite.

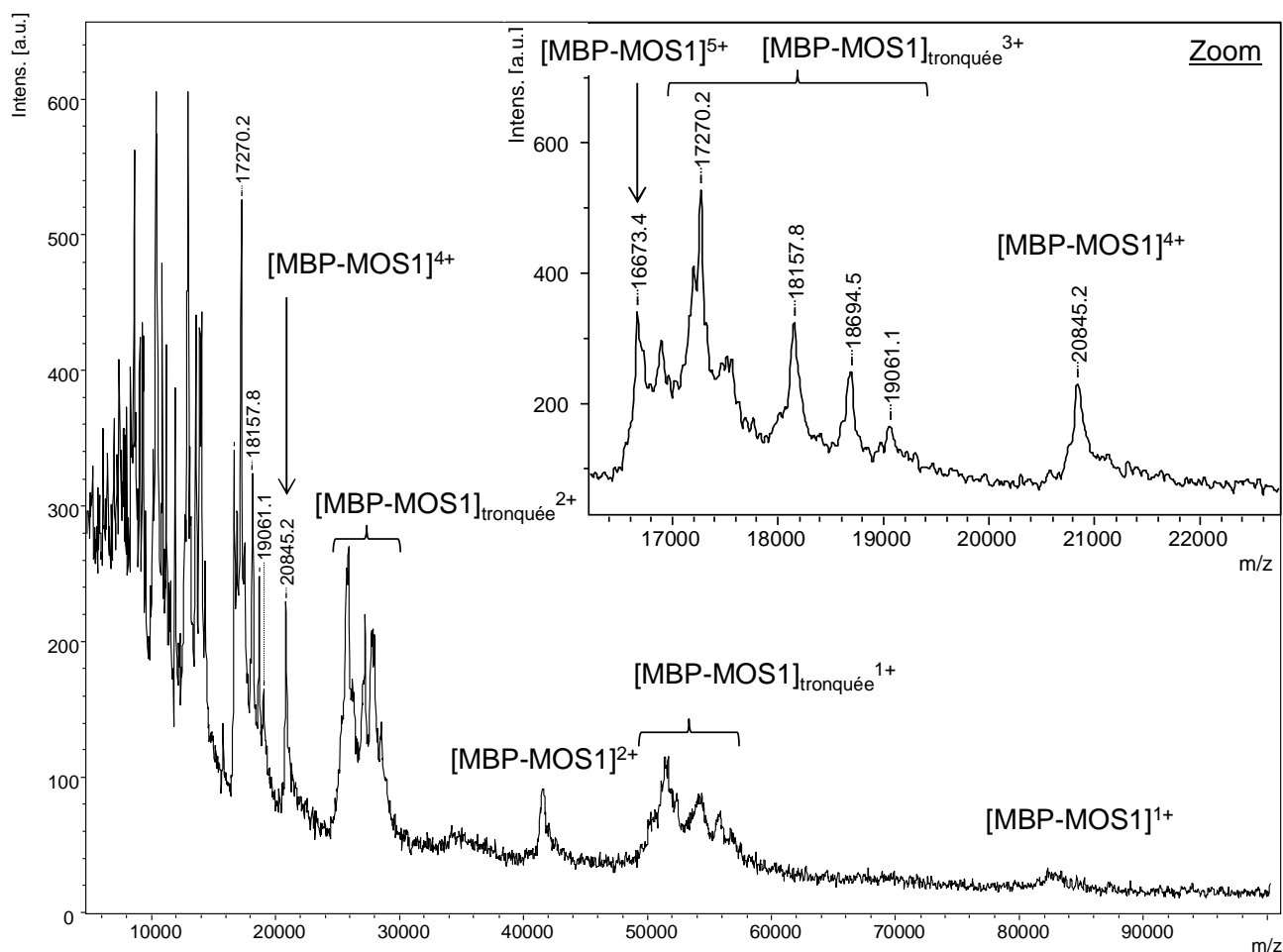
Figure 43. Représentation schématique de la protéine fusion MBP-MOS1.



#### 3.1 Analyse de la pureté de la protéine

Dans un premier temps, l'intégrité et l'homogénéité de la protéine MBP-MOS1<sub>E.coli</sub> ont été vérifiées par spectrométrie de masse MALDI-TOF en mesurant la masse de la protéine entière. Ce type d'analyse est nécessaire avant de commencer le travail minutieux de la caractérisation de complexes protéine-ligand. En effet, l'existence d'une dégradation protéolytique de la protéine et/ou de modifications induites ou post-traductionnelles impacte négativement ce type d'étude par spectrométrie de masse, en augmentant la micro-hétérogénéité de la protéine ou en créant des ambiguïtés d'attribution de séquence aux masses observées. La Figure 44 montre la présence de nombreux pics qui correspondent à la protéine intacte MBP-MOS1<sub>E.coli</sub> mais également à des produits de dégradation protéolytique de la protéine.

**Figure 44. Analyse MALDI-TOF de la protéine MBP-MOS1<sub>E.coli</sub>.** Une solution à 10 µM de protéine MBP-MOS1<sub>E.coli</sub> stockée dans un tampon 10 mM Tris-HCl pH 9, 100 mM NaCl, 1 mM DTT, a été dessalée par micropurification ZipTip C4. L'éluat a ensuite été concentré par évaporation speedvac puis dilué au 1/20<sup>ème</sup> dans la matrice HCCA/FWI. Le mélange échantillon-matrice a été déposé sur cible MALDI par la méthode en couche ultra-mince (Cadene and Chait, 2000; Gabant and Cadene, 2008). Les acquisitions ont été réalisées en mode linéaire positif.



La masse moyenne observée pour MBP-MOS1<sub>E.coli</sub> est de 83377 Da soit un écart de masse de +117 Da par rapport à la masse moyenne théorique de 83260 Da (sans méthionine initiale). Cet écart de masse, trop faible pour correspondre à la méthionine initiale (131 Da), pourrait correspondre soit à une modification, soit à une mutation. Si l'on considère des modifications induites ou post-traductionnelles, le delta masse de +117 Da ne semble pas correspondre à une modification communément décrite. De plus, MBP-MOS1<sub>E.coli</sub> intacte est observée sous la forme d'un seul pic homogène (Figure 44) impliquant que la protéine serait modifiée à 100%. Or, les protéines recombinantes surexprimées dans un système bactérien, sont généralement peu ou pas modifiées, ce qui rend l'hypothèse d'une modification peu crédible (Sahdev *et al.*, 2008). Dans le cas d'une mutation, l'écart de masse observé pourrait être attribué à une mutation par substitution d'une alanine

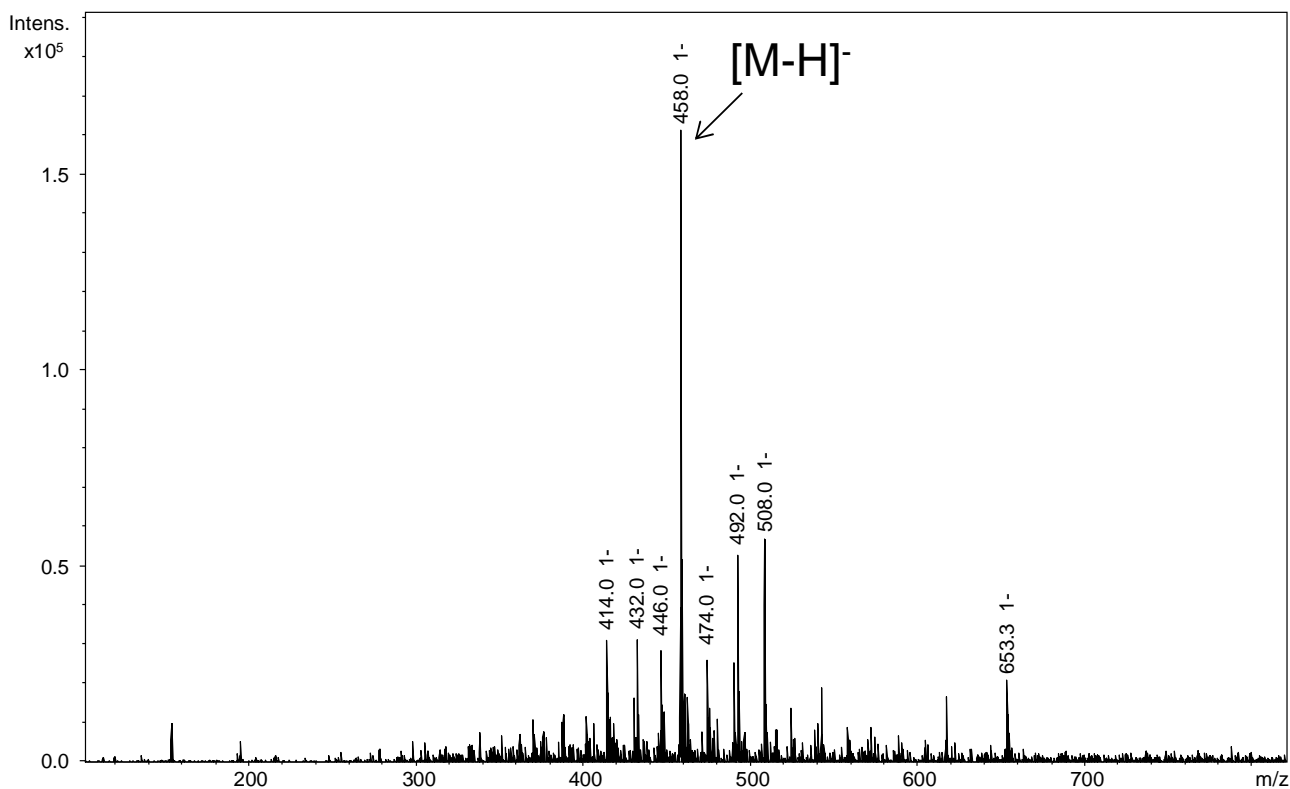
en tryptophane (+115 Da) ou encore à une insertion d'une asparagine (+114 Da) au niveau du linker dont la séquence contient dix asparagine successives : NSSSNNNNNNNNNNLGIEGRT.

Afin de localiser la mutation suspectée, une expérience bottom-up (protéolyse de MBP-MOS1<sub>E.coli</sub> suivie d'une analyse LC-MS/MS des peptides générés) ou top-down (fragmentation de la protéine entière dans le spectromètre de masse) pourrait être réalisée. Etant donné que la protéine contient des produits de dégradation protéolytique, l'approche top-down semble plus adaptée car, en isolant spécifiquement le pic d'intérêt pour la fragmentation, elle permettrait de générer des informations de séquence correspondant uniquement à MBP-MOS1<sub>E.coli</sub> intacte. Cette analyse n'a pas été réalisée pour le moment car la protéine étant active, il semble que la mutation suspectée ne concerne pas la séquence de MOS1 dans la protéine de fusion avec MBP, ou qu'elle affecte une zone de séquence non-indispensable à son activité.

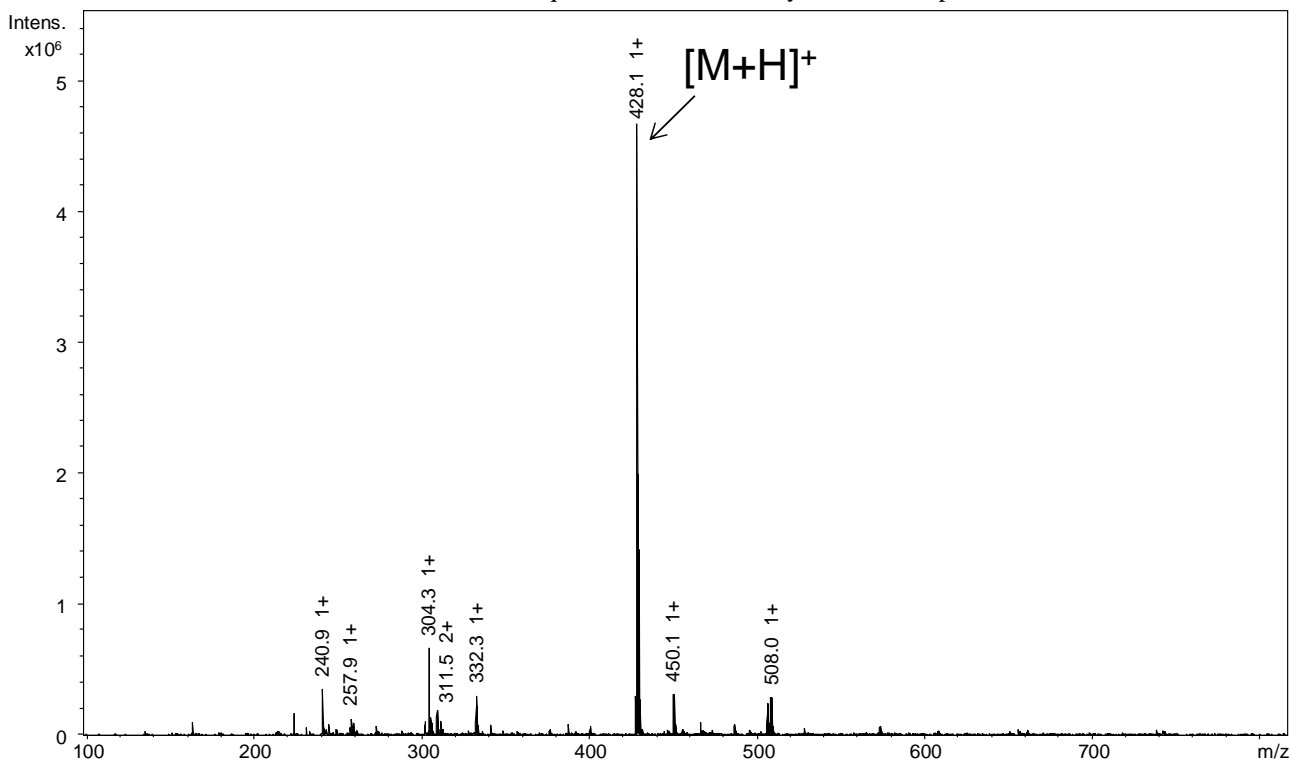
### 3.2 Analyse de la pureté des inhibiteurs

La pureté des inhibiteurs MCV22 et MCV41 a été évaluée par spectrométrie de masse ESI-IT. En raison de la présence de deux fonctions acide dans sa structure, MCV22 a été analysé en mode négatif (Figure 45) et est observé sous forme  $[M-H]^-$ . Plusieurs pics additionnels sont détectés indiquant que MCV22 n'est pas pur à 100%. MCV41 a été analysé en mode positif (Figure 46) et est observé sous forme  $[M+H]^+$ . De la même manière que pour MCV22, des impuretés sont mises en évidence dans l'échantillon MCV41. La présence de ces impuretés peut être préjudiciable lors de la réaction sur la protéine car si elles sont capables de réagir de manière non spécifique avec des résidus de la protéine, elles vont créer de l'hétérogénéité et disperser le signal de la protéine modifiée.

**Figure 45. Analyse ESI-IT de MCV22.** La solution stock de MCV22 100 mM dans le DMSO a été diluée à 10  $\mu$ M dans la solution d'ionisation acétonitrile/eau 50/50 et analysée en mode négatif.



**Figure 46. Analyse ESI-IT de MCV41.** La solution stock de MCV41 100 mM dans le DMSO a été diluée à 1  $\mu$ M dans la solution d'ionisation acétonitrile/eau/acide formique 50/49.8/0.2 et analysée en mode positif.

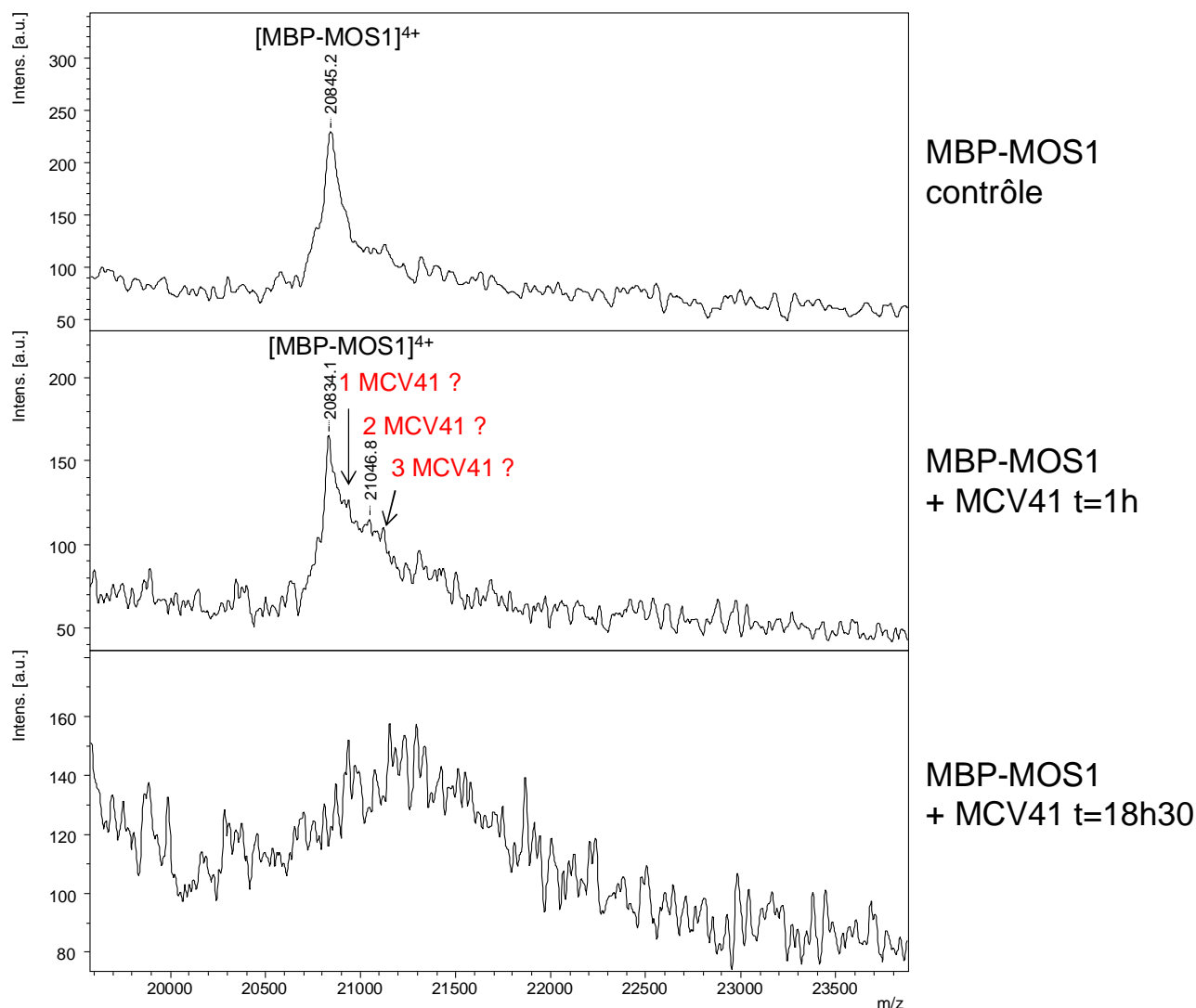




### 3.3 Analyse du complexe entre la protéine MBP-MOS1<sub>E.coli</sub> et l'inhibiteur MCV41

Bien que MBP-MOS1<sub>E.coli</sub> soit partiellement dégradée, une première réaction avec l'inhibiteur MCV41 a été effectuée afin d'avoir un résultat préliminaire. Deux temps d'incubation ont été analysés par MALDI-TOF : 1h et 18h30. En général, les conditions d'ionisation en MALDI classique : conditions dénaturantes (pH < 2, [solvant organique] = 30-50 %) et dépôts solides, sont très défavorables à la survie de complexes non-covalents, ce qui fait du MALDI-TOF un bon outil pour vérifier l'existence d'un complexe covalent entre deux molécules. La Figure 47 montre qu'après une heure de réaction, un épaulement est observé, suggérant l'existence d'un complexe MCV41-MBP-MOS1<sub>E.coli</sub> covalent. Cet épaulement correspondrait à la fixation de 1 à 3 molécules MCV41 sur MBP-MOS1. A t=18h30, le pic MBP-MOS1<sub>E.coli</sub> a disparu en faveur d'un massif non résolu de masse plus élevée, suggérant un nombre accru de molécules MCV41 sur MBP-MOS1<sub>E.coli</sub>. Le fait d'observer un massif non résolu pour le complexe MCV41-MBP-MOS1<sub>E.coli</sub> peut être expliqué par l'hétérogénéité au niveau de l'inhibiteur (Figure 46).

**Figure 47. Analyse MALDI-TOF de la réaction entre la protéine MBP-MOS1<sub>E.coli</sub> et l'inhibiteur MCV41.** La protéine MBP-MOS1<sub>E.coli</sub> a été incubée à une concentration de 10 µM à température ambiante avec l'inhibiteur MCV41 (2 mM) dans un tampon suivant Tris-HCl 20 mM pH=9, NaCl 100 mM, DTT 1 mM et DMSO 2.15 % (apporté par MCV41). La réaction a ensuite été dessalée par micropurification ZipTipC4. L'éluat du ZipTipC4 a été concentré par évaporation au speedvac avant d'être dilué au 1/20<sup>ème</sup> dans la solution de matrice HCCA/FWI puis déposé en couche ultra-mince. Le spectre contrôle sans MCV41 et les spectres en présence de MCV41 à t=1h et t=18h30 sont présentés avec un zoom sur l'état de charge 4+.



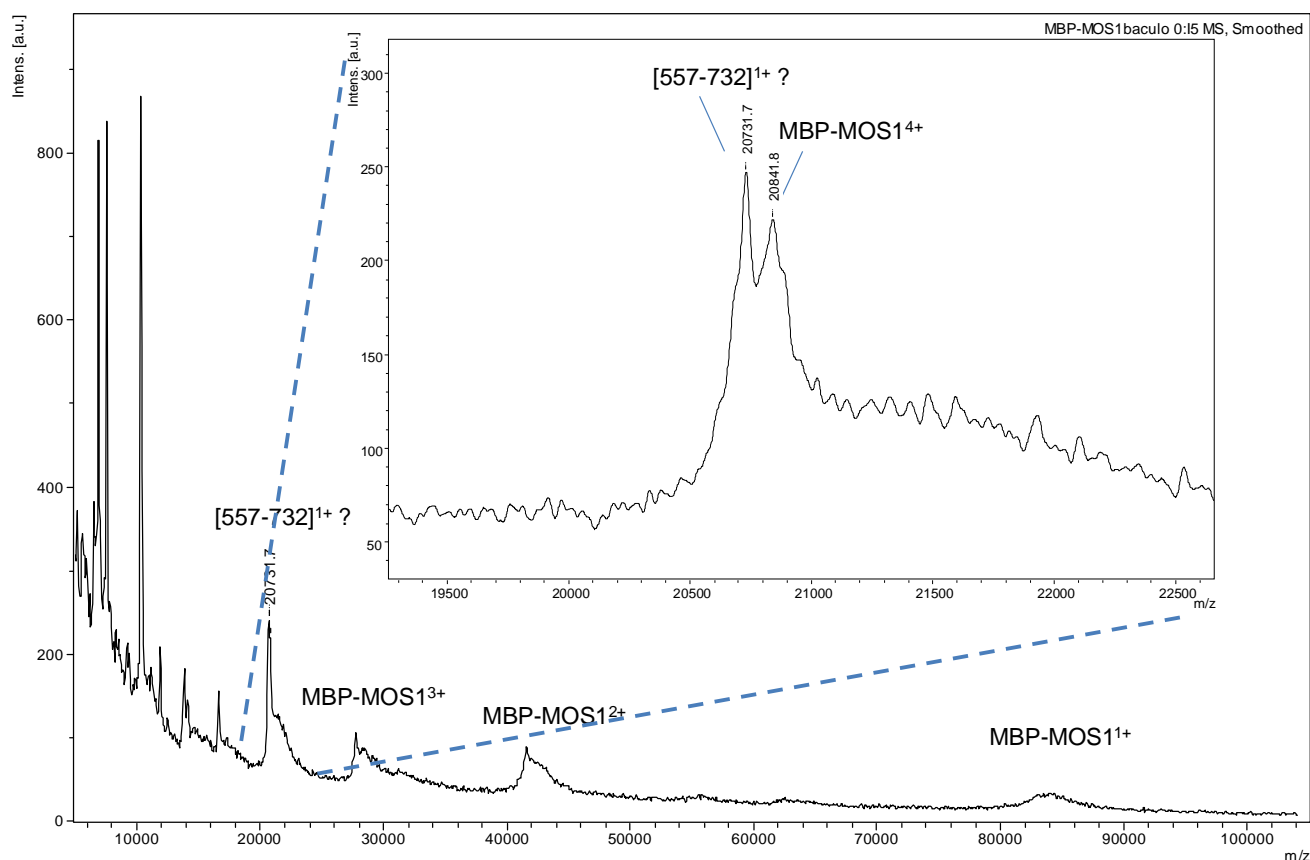
Avant de poursuivre la caractérisation, il était nécessaire de travailler sur un autre stock non dégradé de MBP-MOS1. Ainsi, une nouvelle production, en cellule d'insecte cette fois-ci, a été réalisée par nos collaborateurs.

### 3.4 Analyse de la pureté de la protéine MBP-MOS1 produite en cellules d'insecte

La protéine de fusion MBP-MOS1 exprimée dans *E. coli* contenant des produits de dégradation protéolytique, une autre alternative permettant d'obtenir une préparation de protéine pure et résistante aux protéolyses accidentelles a été considérée. La protéine MBP-MOS1 a donc été produite en

cellules d'insecte (MBP-MOS1<sub>insecte</sub>) par l'équipe de C. Augé-Gouillou. MBP-MOS1<sub>insecte</sub> a été analysée en MALDI-TOF. Le spectre obtenu est présenté Figure 48 et montre à nouveau une hétérogénéité dans l'échantillon. MBP-MOS1<sub>insecte</sub> est observée entière à une masse moyenne de 83363 Da et un pic additionnel de masse 20731 Da est également détecté. Les pics de MBP-MOS1<sub>insecte</sub> sont larges, ce qui suggère une microhétérogénéité apportée par une modifications induite et/ou post-traductionnelle. Concernant le pic de masse 20731 Da, il pourrait correspondre à un contaminant ou à de la dégradation protéolytique de MBP-MOS1<sub>insecte</sub> : fragment [557-732] de masse moyenne théorique 20735.3 Da. Toutefois, ce fragment ne contient pas l'étiquette de purification, et le fragment complémentaire [1-556] de masse moyenne théorique 62542.8 Da n'est pas observé, ce qui rend l'hypothèse d'une dégradation peu plausible. Même si cette préparation de protéine contient encore un produit de protéolyse, elle est de ce point de vue beaucoup plus homogène que la protéine exprimée dans *E. coli*.

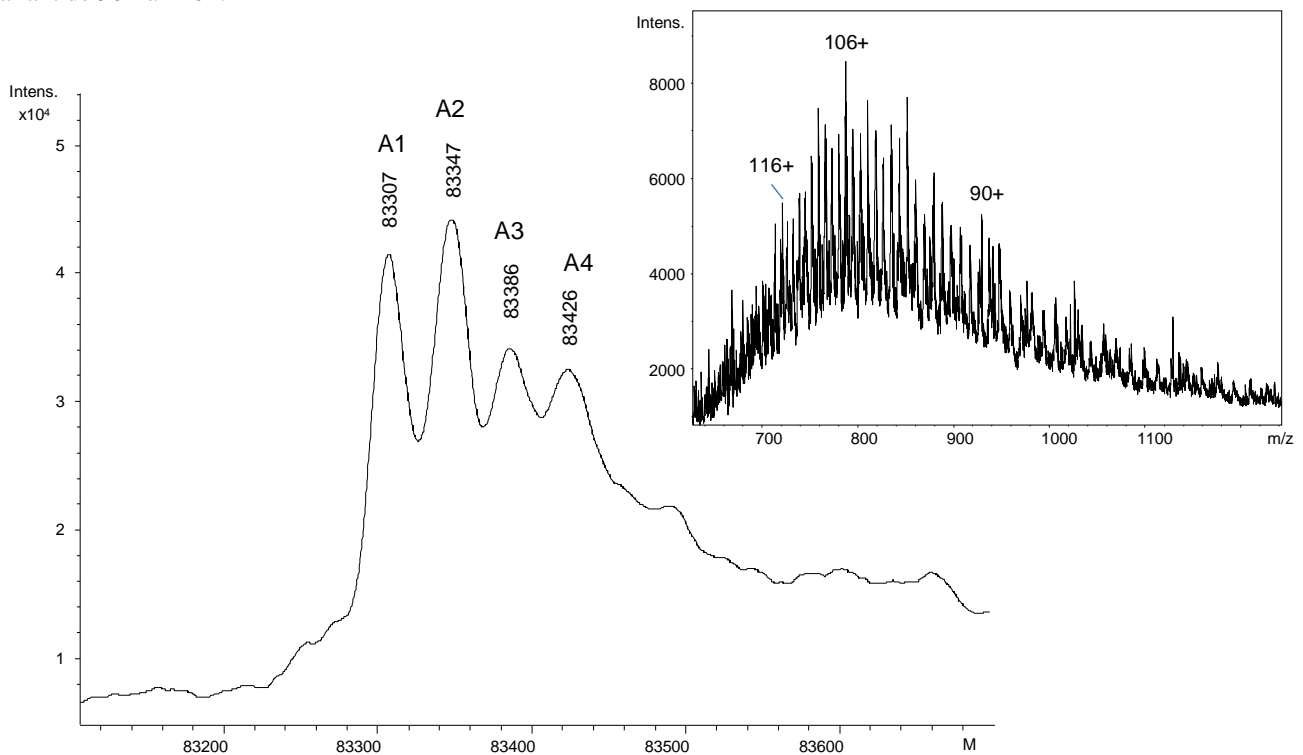
**Figure 48. Analyse MALDI-TOF de la protéine MBP-MOS1<sub>insecte</sub>.** Une solution de protéine MBP-MOS1<sub>insecte</sub> (10 μM) stockée dans un tampon Tris-HCl pH 9, 100 mM NaCl a été diluée au 1/20<sup>ème</sup> dans la matrice HCCA/TWA. Le mélange échantillon-matrice a été déposé sur cible MALDI par la méthode en couche ultra-mince et les dépôts ont été lavés avec du TFA 0.1% avant acquisition en mode linéaire positif.



Afin de vérifier la présence d'une microhétérogénéité de MBP-MOS1<sub>insecte</sub> et de la caractériser, une analyse haute résolution en ESI-UHR-Q-TOF a été effectuée. Le spectre obtenu est présenté

Figure 49. Quatre pics de masses 83307 (A1), 83347 (A2), 83386 (A3) et 83426 Da (A4) sont observés, ce qui confirme l'existence d'une microhétérogénéité de MBP-MOS1<sub>insecte</sub>. Des écarts de masse de +42 Da et +80 Da sont mesurés entre les espèces A2-A1 et A3-A1, suggérant la présence d'acétylation (+42 Da) et de phosphorylation (+80 Da) sur la protéine. La masse observée en MALDI-TOF de 83363 Da correspond en réalité à la masse moyenne des quatre pics observés en HRMS qui est égale à 83362 Da. On voit ici l'apport de la spectrométrie de masse haute résolution pour la mise en évidence de modifications sur des protéines de haut poids moléculaire (83 kDa).

**Figure 49. Analyse ESI-UHR-Q-TOF déconvolué de MBP-MOS1<sub>insecte</sub>.** La protéine MBP-MOS1<sub>insecte</sub> (10 µL) a été dessalée par micropurification ZipTip C4 puis analysée en infusion dans une solution d'acétonitrile/eau/acide formique 50/49.8/0.2. Le spectre déconvolué est présenté. L'insert correspond au spectre brut avec une distribution d'état de charge allant de 90+ à 116+.



Avant d'aller plus loin dans l'étude des complexes entre MBP-MOS1 et les inhibiteurs, nous avons décidé d'entreprendre la caractérisation des modifications de MBP-MOS1<sub>insecte</sub>. Ces travaux sont présentés dans la publication ci-après.

#### 4. Publication

cAMP protein kinase phosphorylates the Mos1 transposase and regulates its activity: evidences from mass spectrometry and biochemical analyses

# cAMP protein kinase phosphorylates the Mos1 transposase and regulates its activity: evidences from mass spectrometry and biochemical analyses

Nicolas Bouchet<sup>1</sup>, Jérôme Jaillet<sup>1</sup>, Guillaume Gabant<sup>2</sup>, Benjamin Brillet<sup>3</sup>, Luis Briseño-Roa<sup>4</sup>, Martine Cadene<sup>2</sup> and Corinne Augé-Gouillou<sup>1,\*</sup>

<sup>1</sup>Innovation Moléculaire Thérapeutique, EA 6306, UFR Sciences Pharmaceutiques, Parc Grandmont, Université François Rabelais, 37200 Tours, France, <sup>2</sup>Centre de Biophysique Moléculaire, UPR 4301 CNRS, Rue Charles Sadron, 45071 Orléans, France, <sup>3</sup>Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, IUT de Quimper, Université de Bretagne Occidentale, 6 rue de l'Université, 29000 Quimper, France and <sup>4</sup>Biologie Cellulaire de la Synapse, INSERM U789, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France

Received June 19, 2013; Revised September 5, 2013; Accepted September 6, 2013

## ABSTRACT

Genomic plasticity mediated by transposable elements can have a dramatic impact on genome integrity. To minimize its genotoxic effects, it is tightly regulated either by intrinsic mechanisms (linked to the element itself) or by host-mediated mechanisms. Using mass spectrometry, we show here for the first time that MOS1, the transposase driving the mobility of the *mariner* Mos1 element, is phosphorylated. We also show that the transposition activity of MOS1 is downregulated by protein kinase AMP cyclic-dependent phosphorylation at S170, which renders the transposase unable to promote *Mos1* transposition. One step in the transposition cycle, the assembly of the paired-end complex, is specifically inhibited. At the cellular level, we provide evidence that phosphorylation at S170 prevents the active transport of the transposase into the nucleus. Our data suggest that protein kinase AMP cyclic-dependent phosphorylation may play a double role in the early stages of genome invasion by *mariner* elements.

## INTRODUCTION

The transposition of genetic elements is potentially genotoxic for the host cell despite being an essential tool for perpetuating genome plasticity, and thus for its evolutionary potential. This means that tight regulation of transposition is essential to maintain the balance

between maintaining active transposons in host genomes and preventing the damage they could cause by potentially lethal DNA rearrangements. Studies of the relationships between transposable elements and the host genome have revealed diverse examples of the regulations that can be achieved through intrinsic, emergent or host-mediated mechanisms. Intrinsic regulation can be due to topological constraints (1), to the poisoning of transposition reactions by the overproduction of transposase (2,3) or to strong negative complementation between active transposases and the product of mutated alleles (2). Transposon silencing, in which transposition is restricted by various epigenetic pathways (4,5), is probably the best known of the regulatory mechanisms that have recently emerged. Little is known about host-mediated regulatory mechanisms, which are triggered when a transposon invades a naive genome that may be nonspecialized mechanisms widely used by eukaryotic cells. They could therefore consist of posttranslational modifications (PTMs), which are often found to drive the regulation of protein activity. Such modifications could alter the cellular localization of a protein, trapping the modified protein in a cellular compartment that is inappropriate for its activity. Alternatively, PTM could modify the stability of the protein, shortening its residence time in the cell, and thereby limiting its activity. Finally, PTM could directly alter protein activity, making it unable to promote any biochemical process.

Over 200 types of PTMs have been identified in eukaryotes so far. The most common ones include (i) phosphorylation, a key reversible modification used as a regulatory mechanism in virtually every process in eukaryotic

\*To whom correspondence should be addressed. Tel: +33 2 47 36 74 72; Fax: +33 2 47 36 70 42; Email: auge@univ-tours.fr

Present address:

Nicolas Bouchet, Centre National de Recherche et de Formation sur le Paludisme, 01 BP 2208 Ouagadougou 01, Burkina Faso.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## 5. Conclusions

Nous avons montré que les deux stocks de protéines MBP-MOS1 présentaient une hétérogénéité : MBP-MOS1<sub>E.coli</sub> contient des produits de protéolyse et MBP-MOS1<sub>insecte</sub> est à la fois acétylée et phosphorylée. Une première expérience exploratoire a été effectuée entre MCV41 et MBP-MOS1<sub>E.coli</sub> et suggère l'existence d'un complexe covalent. Cependant, ce complexe est observé sous forme d'un massif non résolu, ce qui est peu informatif et manque de robustesse pour valider définitivement l'existence d'un complexe covalent en prévision d'études de localisation plus poussées.

## 6. Perspectives

Pour les prochaines expériences, il serait pertinent de diminuer les hétérogénéités, tant au niveau de la protéine que de l'inhibiteur, car elles ajoutent de la complexité non désirée et ne sont de ce fait pas propices à l'étude de complexes protéine-ligand. Ainsi, il faudrait obtenir des inhibiteurs de degré de pureté proche de 100% et une protéine MOS1 non dégradée et peu ou pas modifiée. Alternativement, la protéine exprimée en cellules d'insectes pourrait être déphosphorylée par la phosphatase après réaction avec un inhibiteur et avant l'analyse par une combinaison de protéolyse et de LC-MS/MS.

Une fois ces nouveaux stocks à disposition, les inhibiteurs seront incubés avec la protéine en présence et en absence d'ADN cible puis les complexes formés seront analysés par MALDI-TOF. Une fois avoir validé l'existence de complexes covalents, la stœchiométrie des différents complexes sera déterminée et des expériences visant à localiser le ou les résidu(s) modifiés seront conduites.





## Chapitre 3 Vers la localisation du site actif d'Abf2, une protéine impliquée dans la réparation de l'ADN mitochondrial.

### Résumé du chapitre 3 :

Lors de la respiration cellulaire, des espèces réactives de l'oxygène (ROS) endogènes sont naturellement générées au cours de la chaîne de transport d'électrons. L'ADN mitochondrial, de par sa proximité avec la chaîne respiratoire, est très exposé aux ROS. Ces dernières créent de multiples lésions de l'ADN qui sont potentiellement mutagènes et cytotoxiques. Pour réparer ces dommages, la cellule est dotée d'une batterie d'enzymes de réparation qui détectent et corrigent les anomalies afin de garantir l'intégrité de la cellule. La voie de réparation la plus connue est la voie BER (Base Excision Repair). La protéine *Autonomously replicating sequence (ARS)-binding factor 2 (Abf2)* est une protéine initialement décrite pour son rôle essentiel dans le maintien de l'ADN mitochondrial de *Saccharomyces cerevisiae*. Récemment, Françoise Culard de l'équipe de Bertrand Castaing au CBM, a montré que la protéine Abf2 présente une activité lyase de sites abasiques. Afin de mieux comprendre cette nouvelle activité d'Abf2 et ainsi d'appréhender son potentiel éventuel comme cible thérapeutique, la localisation du site actif par spectrométrie de masse a été entreprise. L'activité de coupure de site abasique d'Abf2 met en jeu un intermédiaire réactionnel formant une base de Schiff entre Abf2 et l'ADN. Ce complexe Abf2-ADN étant transitoire, il a tout d'abord été nécessaire de le stabiliser (réduction au borohydrure de sodium) pour former un complexe covalent. Une stratégie basée sur une purification du complexe covalent Abf2-ADN suivie d'une protéolyse sur billes Ni-NTA, d'une hydrolyse de l'oligonucléotide puis d'une analyse MS a été développée. Deux versions ont été conçues. La première approche met en œuvre une purification du complexe Abf2-ADN en deux étapes : chromatographie d'affinité avidine-biotine et chromatographie d'affinité sur métal immobilisé. La deuxième méthode, plus sensible, combine une purification du complexe Abf2-ADN et un enrichissement en hétéroconjugués peptide-oligonucléotide par chromatographie échangeuse d'anions (monoQ). Cette stratégie expérimentale a permis de détecter efficacement les hétéroconjugués peptide-oligonucléotide et de permettre leur fragmentation informative en ETD. Plusieurs zones modifiées ont été observées, notamment l'étiquette 6 His, ce qui témoigne d'une réaction non spécifique lors de l'étape de piégeage de l'intermédiaire réactionnel. Sur la base des résultats de MS, plusieurs mutants ont été produits et purifiés. Un seul mutant montre une diminution de l'activité lyase, mais celle-ci est très réduite et indissociable d'une diminution de l'affinité pour l'ADN. Les autres mutants n'ont pas montré de baisse d'activité lyase. Afin de localiser le site actif d'Abf2 pour l'activité de clivage de sites abasiques, il faudra s'affranchir de ces réactions non spécifiques.

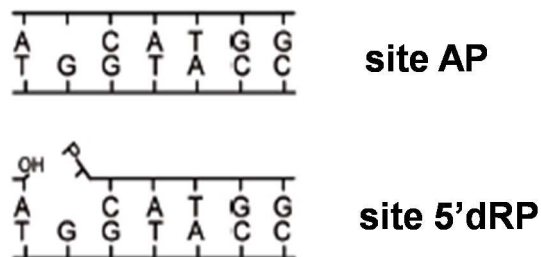
# 1. Contexte

## 1.1 La protéine Abf2

Les cellules procaryotes et eucaryotes sont constamment exposées à des agents chimiques endogènes ou environnementaux qui oxydent les protéines mais également l'ADN. Les génomes mitochondriaux sont plus particulièrement endommagés par des espèces réactives de l'oxygène (ROS) endogènes générées au cours de la chaîne de transport d'électrons de la respiration cellulaire (Richter *et al.*, 1988). Ces lésions de l'ADN, telles que les bases oxydées ou les sites abasiques (AP), sont potentiellement mutagènes et cytotoxiques et la plupart sont corrigées grâce à la voie de réparation par excision de base (BER Base excision repair) (Dizdaroglu, 2005). Le mécanisme de réparation BER se décompose en plusieurs étapes comprenant une reconnaissance et une élimination de la base endommagée, une coupure du brin, une resynthèse et une ligation.

Un site abasique, ou site apurinique/apirimidique (AP), correspond à un emplacement de l'ADN où la base azotée (purine ou pyrimidine) est manquante (Figure 50). Un site AP résulte de l'hydrolyse de la liaison N-glycosydique par les ROS, ou est le produit final des ADN glycosylases qui éliminent les bases endommagées. Un site AP se décline en site 5'desoxyribose phosphate (5'dRP) lorsqu'il y a eu clivage en 5' du site AP (Figure 50).

**Figure 50. Illustration d'un site abasique (AP) et d'un site 5'desoxyribose phosphate (5'dRP).**

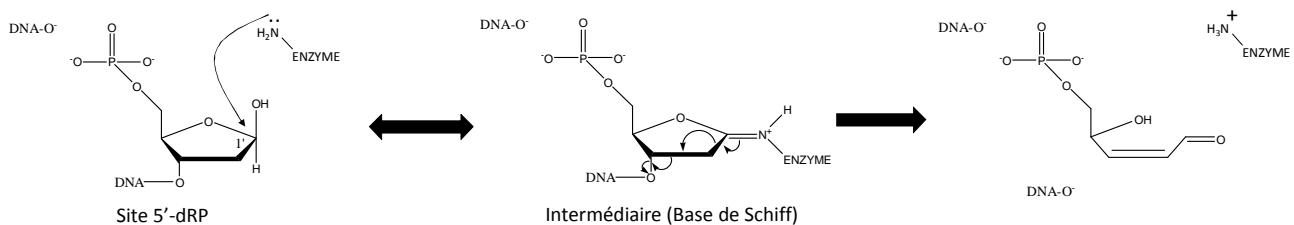


Les sites AP peuvent se former spontanément ou *via* l'action de glycosylases monofonctionnelles, telles que l'uracile DNA glycosylase (UDG), qui excisent les bases oxydées ou inappropriées. Chez les eucaryotes, les sites AP peuvent être pris en charge par certaines glycosylases bifonctionnelles, telles que la 8-oxoguanine DNA glycosylase (OGG1), qui, par leur activité lyase, clivent l'ADN en 3' ou en 3' et 5' des sites AP, mais ils sont essentiellement traités par les AP endonucléases qui clivent l'ADN en 5' des sites AP formant ainsi un site 5'dRP. Ces extrémités 5'dRP seront prise en charge par différentes protéines telles que les glycosylases bifonctionnelles, des polymérases (Pol4 dans la levure) ou encore des enzymes spécifiques comme Rad27 (Boiteux and Guillet, 2004).

La protéine Autonomously replicating sequence (ARS)-binding factor 2 (Abf2) est essentielle pour le maintien de l'ADN mitochondrial de *Saccharomyces cerevisiae*. Elle appartient à la famille des protéines HMGB (High Mobility Group Box) et est composée de deux domaines HMG, A et B, séparés par une zone de liaison de quatre résidus (Diffley and Stillman, 1991). Aucune structure d'Abf2, en complexe ou non avec l'ADN, n'est disponible à ce jour. La protéine HMGB1 est impliquée dans le BER en tant que cofacteur de régulation (Prasad *et al.*, 2007). L'orthologue chez l'humain, le facteur de transcription mitochondrial A (Tfam), pourrait jouer un rôle dans la réparation de l'ADN *via* la reconnaissance de l'ADN oxydé (Yoshida *et al.*, 2002) et est impliqué dans la rétinopathie diabétique (Santos *et al.*, 2014).

Françoise Culard (équipe de Bertrand Castaing au CBM) a récemment montré que la protéine Abf2 présente une activité lyase (activité de clivage du désoxyribose) des sites AP et 5'dRP (résultats non publiés). L'hypothèse de travail est que le mécanisme d'excision du désoxyribose d'un site abasique par Abf2 implique une amine primaire. Cette amine nucléophile attaquerait l'aldéhyde réactif en C1' du site AP ou 5'dRP pour former un complexe intermédiaire covalent de type base de Schiff. Cet intermédiaire transitoire subirait ensuite une réaction de  $\beta$ -élimination conduisant au clivage du brin d'ADN en C3' du site abasique (Figure 51).

**Figure 51. Activité lyase : mécanisme d'excision du désoxyribose d'un site 5'dRP.**



Ce mécanisme a déjà été décrit chez d'autres enzymes à activité lyase telles que Fpg d'*E. coli* ou Ogg1 de *S. cerevisiae* qui ont pour sites actifs respectifs l'amine secondaire de la proline N-terminale (Zharkov *et al.*, 1997) et l'amine primaire de la lysine 241 (Girard *et al.*, 1997).

Cette nouvelle activité lyase d'Abf2 suggère qu'en plus de son rôle structural, elle pourrait être impliquée dans la réparation de l'ADN.

## 1.2 Analyse d'hétéroconjugués protéine-ADN ou peptide-oligonucléotide par spectrométrie de masse.

La caractérisation MS d'hétéroconjugués protéine-ADN ou peptide-oligonucléotide est compliquée par les propriétés physico-chimiques très contrastées des acides aminés et des nucléotides. En effet, ces différences se traduisent par des conditions d'analyse divergentes en spectrométrie de masse :

- En MALDI, les peptides et les protéines sont analysés avec la matrice HCCA alors que les oligonucléotides sont observés avec des matrices d'affinité protonique supérieure : HPA, THAP.
- En ESI, bien que les peptides et les protéines puissent être observés en mode négatif, ils sont classiquement analysés en mode positif qui est plus sensible et qui permet des fragmentations plus efficaces. A l'inverse, les oligonucléotides, du fait des nombreuses charges négatives apportées par les phosphates des liaisons phosphodiester, sont exclusivement analysés en mode négatif.
- En MS/MS par CID, une labilité accrue de l'oligonucléotide par rapport au peptide par perte d'alkyl phosphates et de carbohydrates est observée.

Ainsi, la partie oligonucléotide d'un hétéroconjugué peptide-oligonucléotide doit être suffisamment petite, c'est-à-dire deux à trois nucléotides maximum, pour que l'hétéroconjugué présente des propriétés analogues aux peptides. Ceci permet d'analyser l'hétéroconjugué dans des conditions adaptées aux peptides et de limiter l'effet de suppression de l'ionisation de l'oligonucléotide (Steen and Jensen, 2002)

## 2. Objectif du projet

Ce projet collaboratif vise à localiser le ou les résidu(s) catalytique(s) de l'activité lyase d'Abf2 par spectrométrie de masse. Pour atteindre cet objectif, il a tout d'abord été nécessaire de développer une stratégie expérimentale permettant de :

- 1) stabiliser le complexe intermédiaire Abf2-ADN,
- 2) enrichir le complexe protéine-ADN
- 3) observer et fragmenter efficacement les hétéroconjugués peptide-oligonucléotide en spectrométrie de masse.

### 3. Contrôle qualité de la protéine Abf2 et de son substrat ADN

#### 3.1 Analyse de la pureté d'Abf2

La protéine Abf2 a été exprimée en fusion avec une étiquette six histidines N-terminale. La séquence de la protéine est présentée Figure 52.

**Figure 52. Structure primaire d'Abf2 avec une étiquette 6 histidines en N-terminal (157 acides aminés).** Les domaines A et B sont surlignés en vert et bleu, respectivement. Les résidus soulignés constituent l'étiquette 6 histidines N-terminale. Au total, 25 lysines sont comptabilisées.

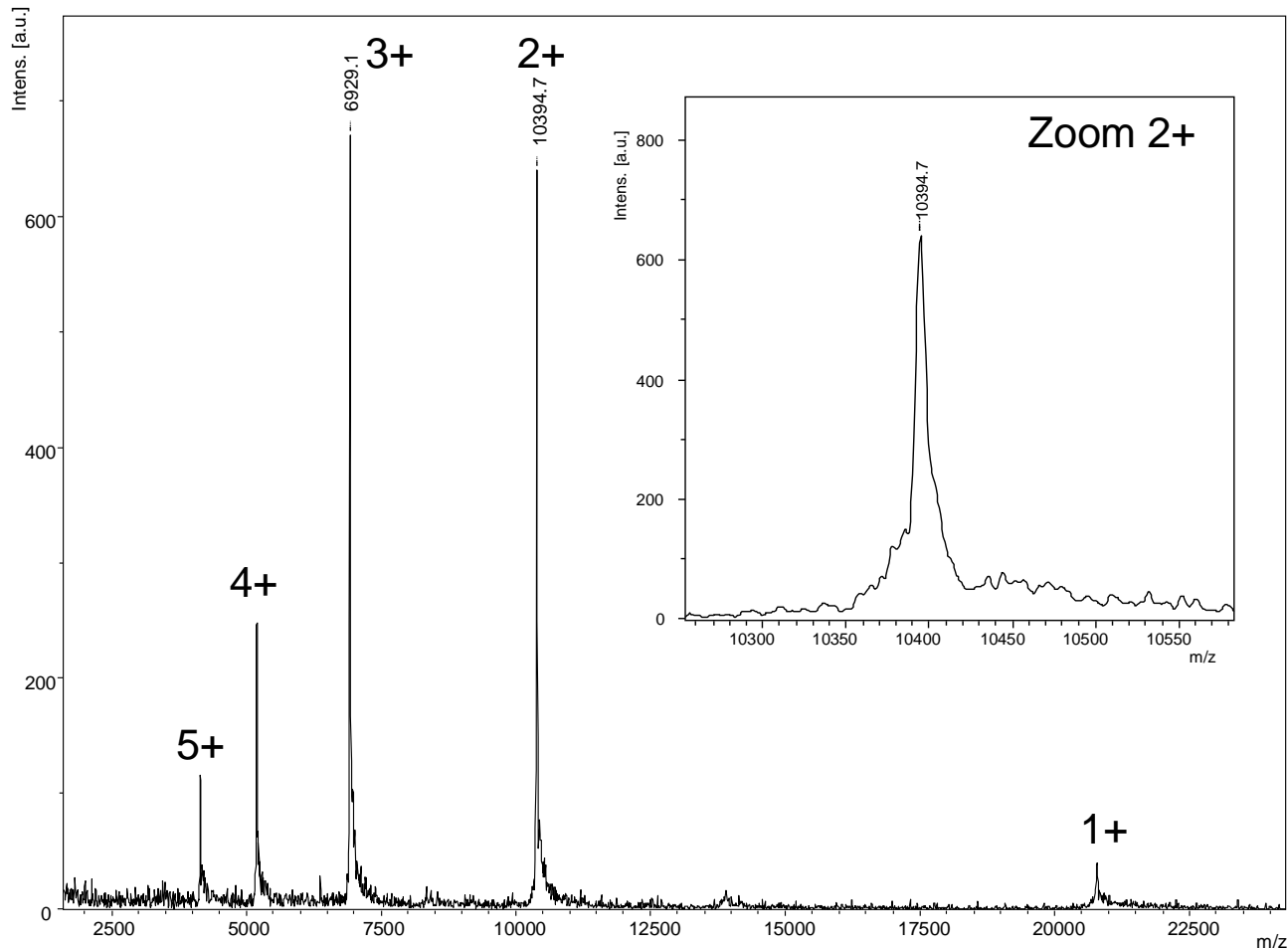
```

GSSH1HHHHS SGLVPRGSHM 10KASKRTQLRN 20ELIKQGPKRP 30TSAYFLYLQD 40HRSQFVKENP
50TLRPAEISKI 60AGEKWQNLEA 70DIKEYISER 80KKLYSEYQKA 90KKEFDEKLPP 100KKPAGPFIKY
110ANEVRSQVFA 120QHPDKSQLDL 130MKIIGDKWQS 140LDQSIKDKYI 150QEYKKAIQEY 157NARYPLN

```

Après purification, l'intégrité et l'homogénéité de la protéine ont été vérifiées par spectrométrie de masse MALDI-TOF (Figure 53). La protéine est observée à une masse moyenne de 20785.9 Da (masse calculée à partir des états de charges 2<sup>+</sup> et 3<sup>+</sup>) correspondant à la masse attendue (masse moyenne théorique = 20785.6 Da). La protéine est très pure, non dégradée et homogène (absence de modification induite).

**Figure 53. Analyse MALDI-TOF de la protéine Abf2.** Une solution de protéine Abf2 à 52  $\mu\text{M}$  dans une solution de Tris 2 mM NaCl 20 mM, pH7.3 a été diluée au 1/20<sup>ème</sup> et un aliquot a été déposé par la méthode en couche ultra-mince (Cadene and Chait, 2000; Gabant and Cadene, 2008). Le dépôt a été lavé avec du TFA 0.1%.



### 3.2 Analyse de la pureté des oligonucléotides

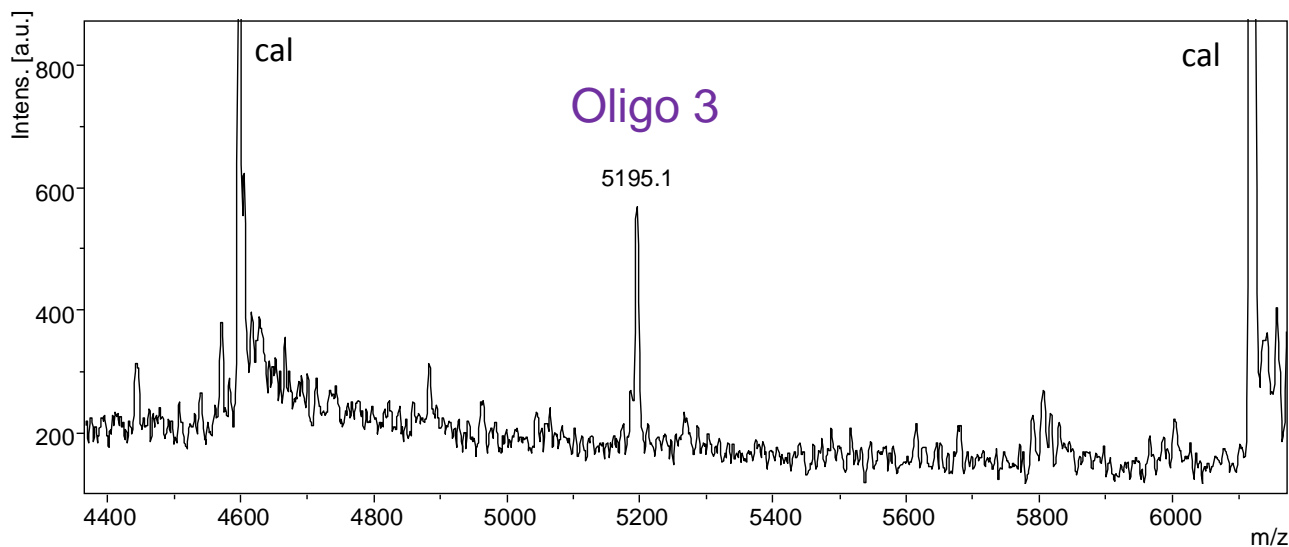
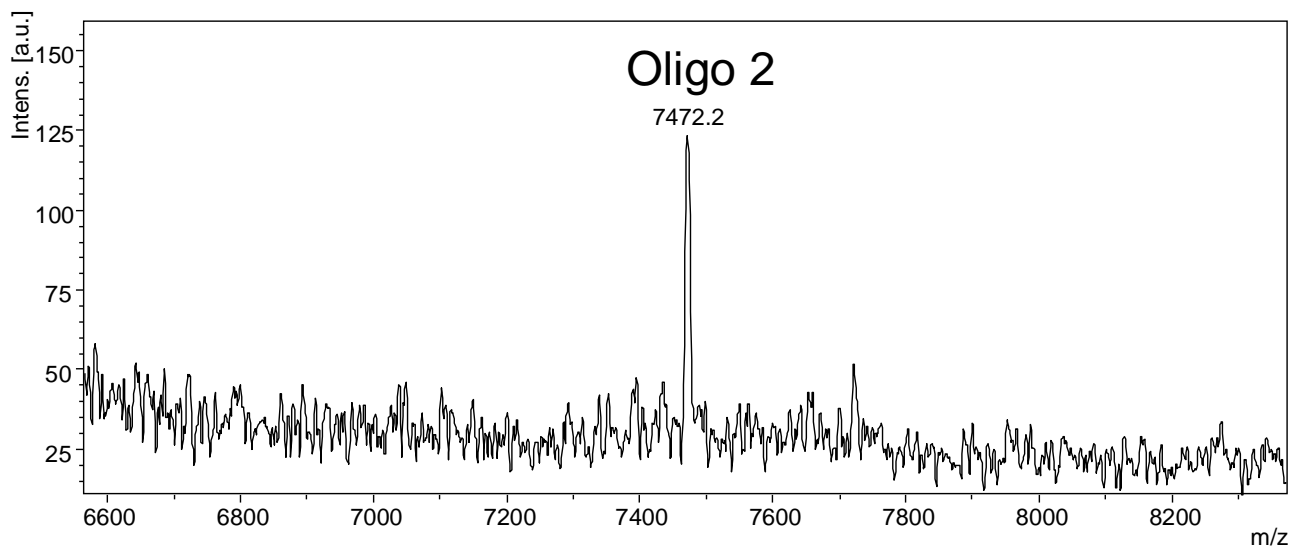
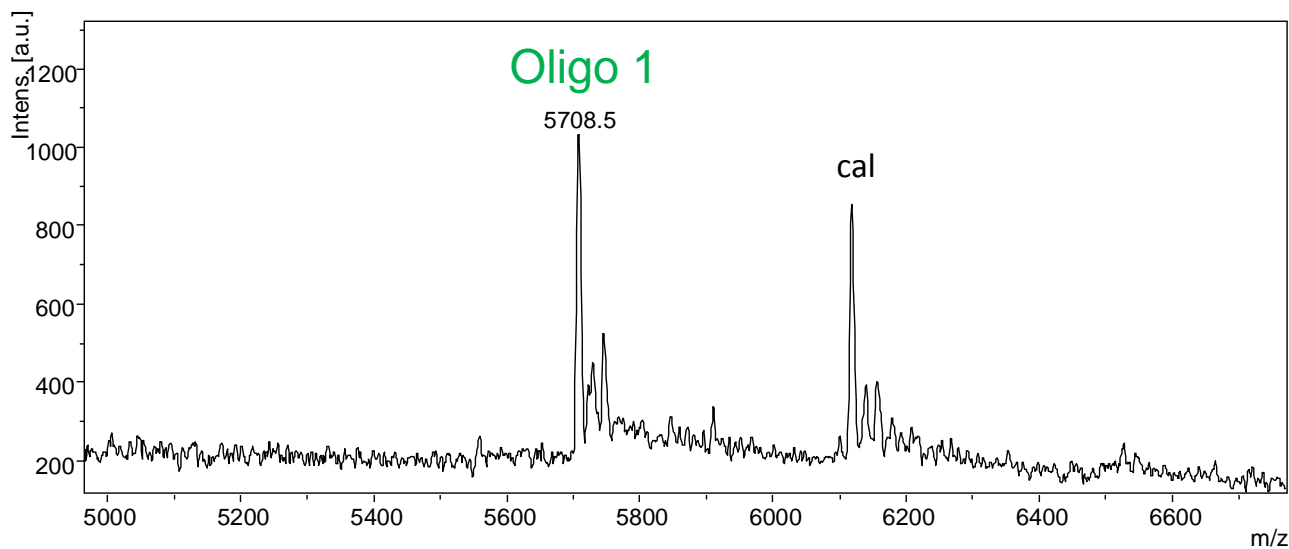
Plusieurs oligonucléotides, de séquence et de longueur différentes, ont été utilisés pour générer un duplex ADN contenant un site AP ou 5'dRP. L'ensemble de ces oligonucléotides utilisés est répertorié dans le Tableau 9.

**Tableau 9. Séquence et masse théorique des oligonucléotides utilisés dans ce projet.**

Nom	Séquence 5' → 3'	Longueur	Masse théorique moyenne 5'P (Da)	Masse théorique moyenne 5'OH (Da)
Oligo 1	<u>U</u> TTCTCGCCGGGTGCAG-biotine	17 mers biotinylé	5703	5623
Oligo 2	CCCGGCGAGAAACAAAGAGGCGCC	24 mers	7471	7391
Oligo 3	GTACGGGCGCCTCTTTG	17 mers	5273	5193
Oligo 4	GGCGCTCTT <u>U</u> TTTCTCGCCGGG	24 mers	7200	7280

La pureté de ces oligonucléotides a été évaluée par MALDI-TOF MS. Les résultats sont présentés Figure 54 pour les oligonucléotides 1 (5'P), 2 (5'P) et 3 (5'OH). Les oligonucléotides sont observés à leur masse attendue et présentent une bonne pureté.

**Figure 54. Analyse MALDI-TOF des oligonucléotides 1 (5'P), 2 (5'P) et 3 (5'OH) utilisés pour la formation d'un duplex contenant un site 5'dRP.** Chaque oligonucléotide a été préparé à une concentration d'environ 5  $\mu$ M dans de l'eau milliQ. La solution de matrice est composée d'acide 3-hydroxypicolinique (HPA) à 10 mg/mL, de citrate d'ammonium dibasique à 1 mg/mL et d'acétonitrile 10%. Les dépôts ont été réalisés sur cible Anchorchip selon la méthode en goutte séchée : 1  $\mu$ L de solution de matrice a été tout d'abord déposé puis 1  $\mu$ L d'échantillon a été ajouté avant que la matrice ne sèche. Les acquisitions ont été réalisées en mode linéaire positif. Une calibration interne a été effectuée en utilisant un mélange standard Bruker. Les pics de calibrant sont indiqués par « cal ».





#### 4. Génération de duplex ADN contenant un site AP ou 5'dRP, substrat d'Abf2

La génération de duplex ADN contenant un site AP ou 5'dRP a été réalisée par nos collaborateurs Françoise Culard et Stéphane Goffinont. Classiquement, la première étape consiste à hybrider deux brins d'ADN complémentaires dont un contient un uracile dans la séquence. Ensuite, l'uracile est éliminé par un traitement à l'Uracile DNA glycosylase (UDG) générant ainsi un site AP. A ce stade, un site 5'dRP peut être obtenu, en lieu et place du site AP, par un traitement à l'apurique/apyrimidique endonucléase 1 (APE1) qui clive la liaison phosphodiester en 5' du site abasique (Figure 55). Pour générer un site 5'dRP, une variante est possible en utilisant trois brins d'ADN dont un contient l'uracile à l'extrémité 5'. Le site 5'dRP est directement obtenu après hybridation et traitement à l'UDG (Figure 56).

Figure 55. Génération de duplex ADN 24mers contenant un site AP ou 5'dRP.

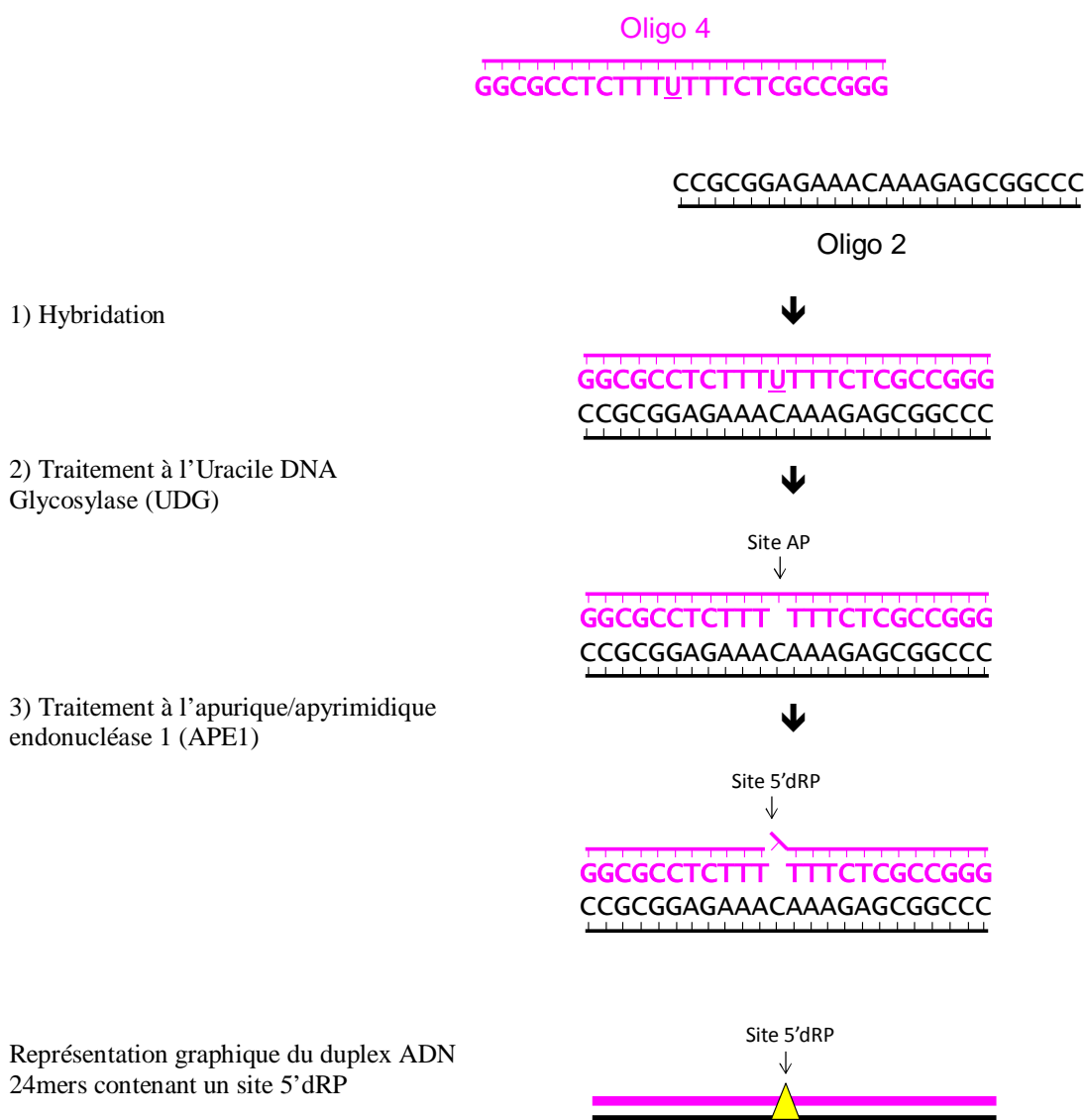
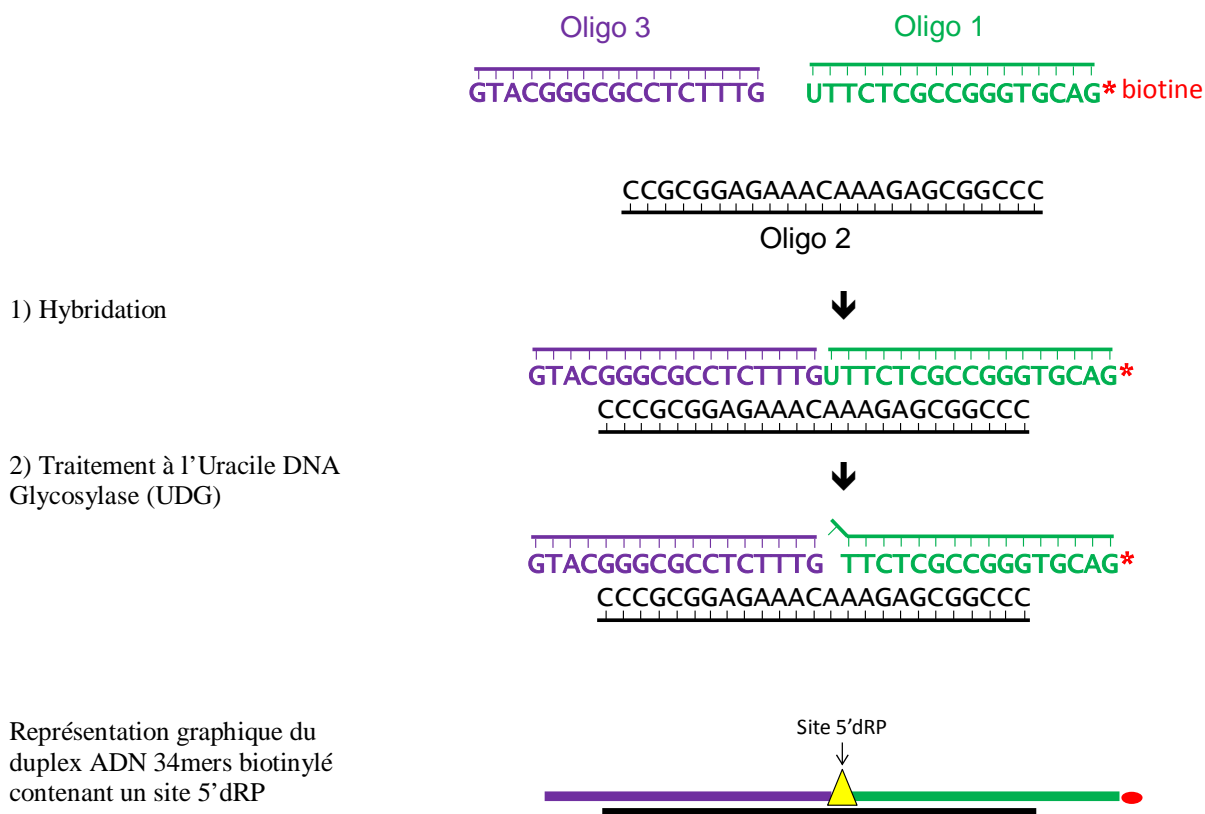


Figure 56. Génération de duplex ADN 34mers biotinylé contenant un site 5'dRP à partir de trois brins d'ADN.

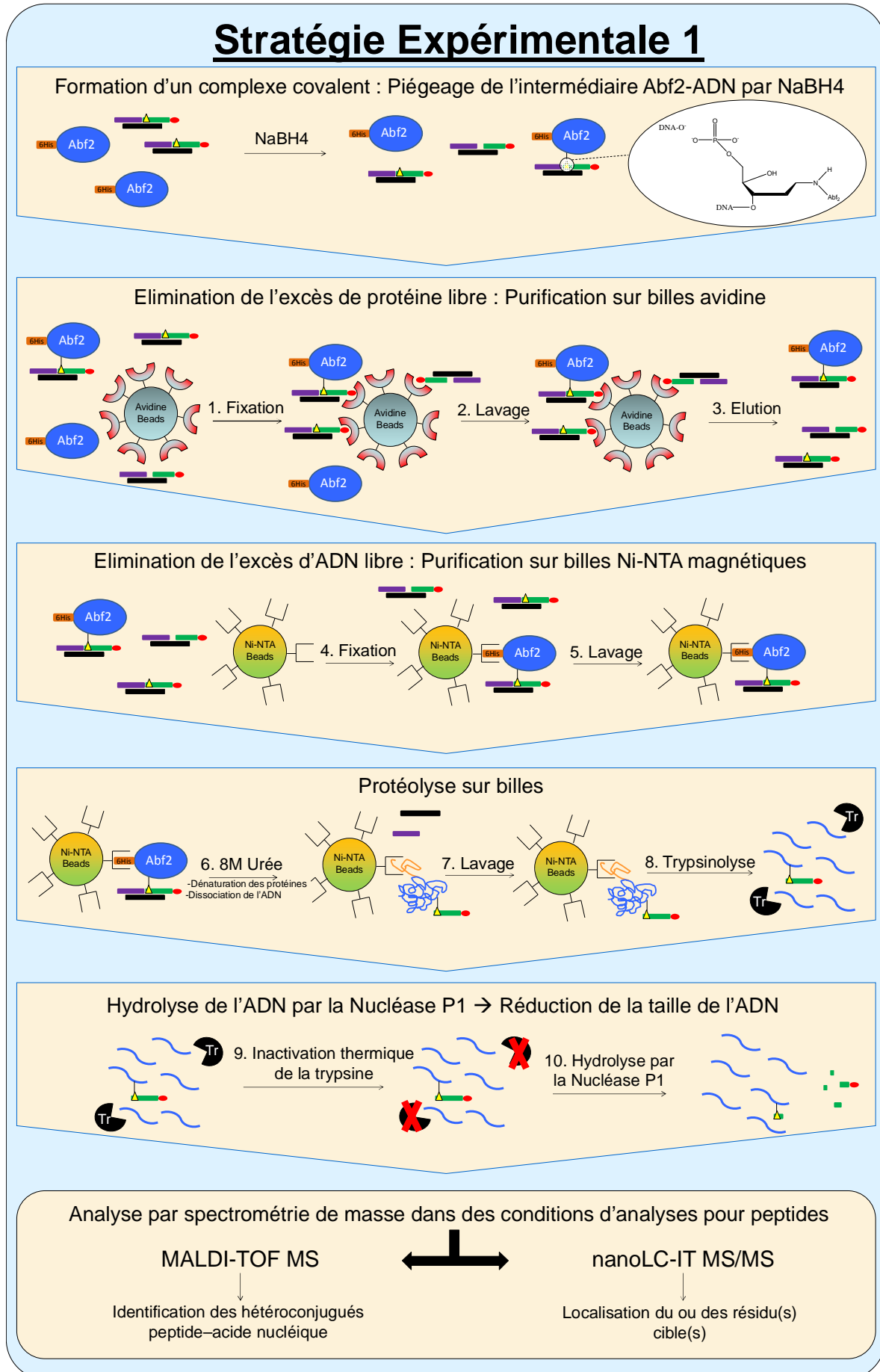


## 5. Stratégie expérimentale 1

Pour les premiers essais, nous avons utilisé le duplex 34-mers biotinylé contenant un site 5'dRP (Oligo 1). Ce duplex ADN, formé à partir des Oligo 1, 2 et 3, a été choisi pour deux raisons : 1) Abf2 présente une meilleure activité lyase pour les sites 5'dRP que pour les sites AP et 2) la biotine a été utilisée pour l'enrichissement du complexe Abf2-ADN.

La stratégie expérimentale 1 est présentée dans son ensemble dans la Figure 57 puis détaillée dans les paragraphes suivants.

Figure 57. Stratégie expérimentale 1 pour la purification du complexe covalent Abf2-ADN et l'analyse d'hétéroconjugués peptide-acide-nucléique en spectrométrie de masse



## 5.1 Formation d'un complexe Abf2-ADN covalent et irréversible

Afin d'identifier le ou les site(s) catalytique(s) de l'activité lyase d'Abf2, il a tout d'abord été nécessaire de stabiliser le produit intermédiaire de la réaction enzymatique. Cet intermédiaire, qui est une base de Schiff entre l'enzyme et l'ADN, peut être converti en complexe covalent irréversible par une réduction au borohydrure de sodium ( $\text{NaBH}_4$ ) (Figure 57, panneau du haut). Cette réduction au borohydrure de sodium a déjà été utilisée pour piéger avec succès des intermédiaires Fpg-ADN (Castaing *et al.*, 1999) ou Ogg1-ADN (Girard *et al.*, 1997) et est bien maîtrisée par l'équipe de B. Castaing. Le rendement de piégeage estimé sur gel SDS-PAGE est inférieur à 10%. Ce rendement est faible et a nécessité la mise en œuvre d'une méthode de purification du complexe d'intérêt Abf2-DNA en amont de l'analyse par spectrométrie de masse.

## 5.2 Enrichissement du complexe Abf2-ADN

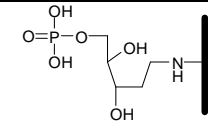
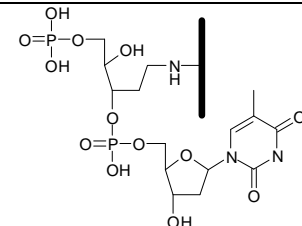
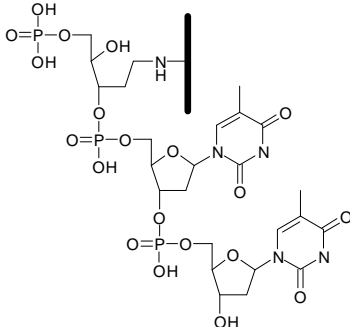
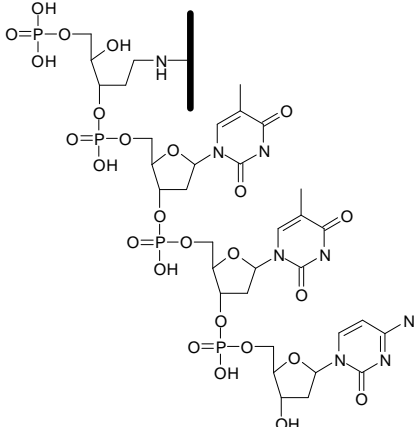
L'enrichissement du complexe Abf2-ADN a consisté en une approche à deux étapes : 1) chromatographie d'affinité avidine-biotine et 2) chromatographie d'affinité sur métal immobilisé. Cette approche permet d'éliminer successivement l'excès de protéine libre puis l'excès d'ADN libre. En fin d'étape de purification, le complexe Abf2-ADN d'intérêt se trouve fixé sur les billes Ni-NTA.

## 5.3 Dégradation enzymatique de la protéine et de l'oligonucléotide

La stratégie retenue pour la localisation du site actif d'Abf2 est une approche bottom-up. Ainsi, le complexe encore fixé sur les billes a été dénaturé afin de déplier la protéine puis protéolysé par la trypsine, générant ainsi, en solution, un mélange de peptides non modifiés et modifiés (hétéroconjugué peptide-oligonucléotide). Le fait de réaliser la protéolyse de la protéine alors qu'elle est encore fixée sur les billes permet d'utiliser des concentrations élevées en agent dénaturant et donc d'assurer une dénaturation efficace. Ce dénaturant en concentration élevé est ensuite facilement éliminé par lavage immédiatement avant la protéolyse.

Enfin, comme évoqué dans le paragraphe 1.2 de ce chapitre, afin de limiter l'effet suppression d'ionisation dû à l'oligonucléotide, la partie oligonucléotide des hétéroconjugués peptide-oligonucléotide a été réduite à deux ou trois nucléotides par un traitement à la nucléase P1. Les incréments de masses théoriques de la base de Schiff stabilisée avec zéro à trois nucléotides sont présentés dans le Tableau 10.

**Tableau 10. Incréments de masses théoriques de la base de Schiff stabilisée avec zéro à trois nucléotides.** La base de Schiff stabilisée est représentée par un triangle. Les lettres T et C correspondent à la thymine et la cytosine, respectivement.

Modifications	Structure des modifications	Formule chimique de la modification ajoutée	Incrément de masse ajouté	
			Masse théorique monoisotopique (Da)	Masse théorique moyenne (Da)
$\Delta$		$C_5H_{11}O_6P_1$	198.0293	198.1112
$\Delta^T$		$C_{15}H_{24}N_2O_{13}P_2$	502.0754	502.3049
$\Delta^{TT}$		$C_{25}H_{37}N_4O_{20}P_3$	806.1214	806.4986
$\Delta^{TTC}$		$C_{34}H_{49}N_7O_{26}P_4$	1095.1678	1095.6809

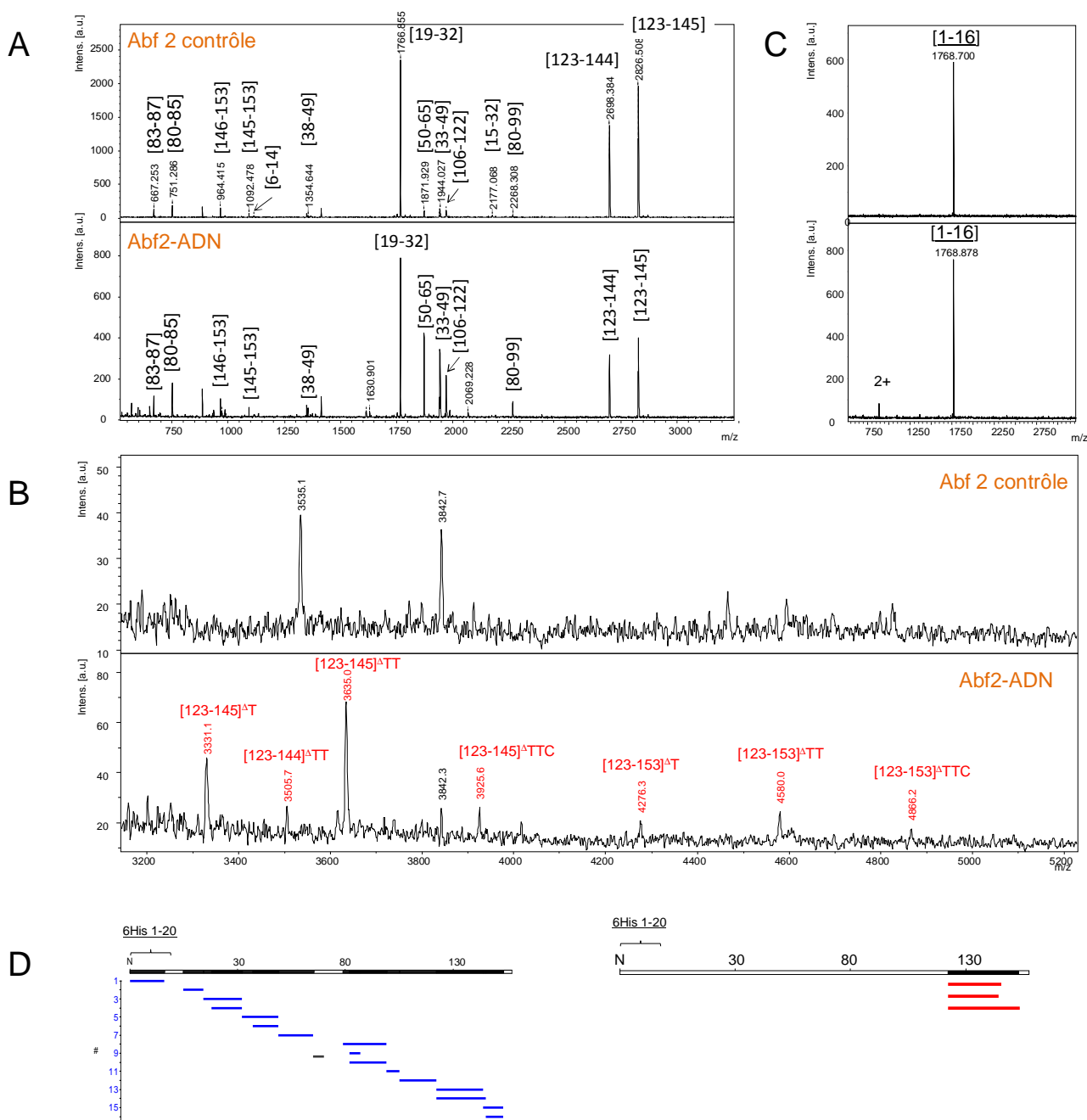
## 6. Les résultats de localisation de la stratégie 1

### 6.1 Identification des hétéroconjugués peptide-acides nucléiques

Le complexe covalent Abf2-ADN a été formé en présence de 75  $\mu$ M d'Abf2, de 50  $\mu$ M de duplex ADN et de 5 mM de NaBH<sub>4</sub>, soit un rapport protéine:substrat de 1.5:1. A ces concentrations, 100% de la protéine est fixé à l'ADN. Le complexe covalent Abf2-ADN a ensuite été purifié et protéolysé

selon le protocole expérimental 1 décrit précédemment. Un contrôle, correspondant à Abf2 seule, a été introduit au niveau de l'étape de protéolyse sur billes Ni-NTA. Les deux échantillons protéolysés par la trypsine ont ensuite été analysés par MALDI-TOF MS (Figure 58).

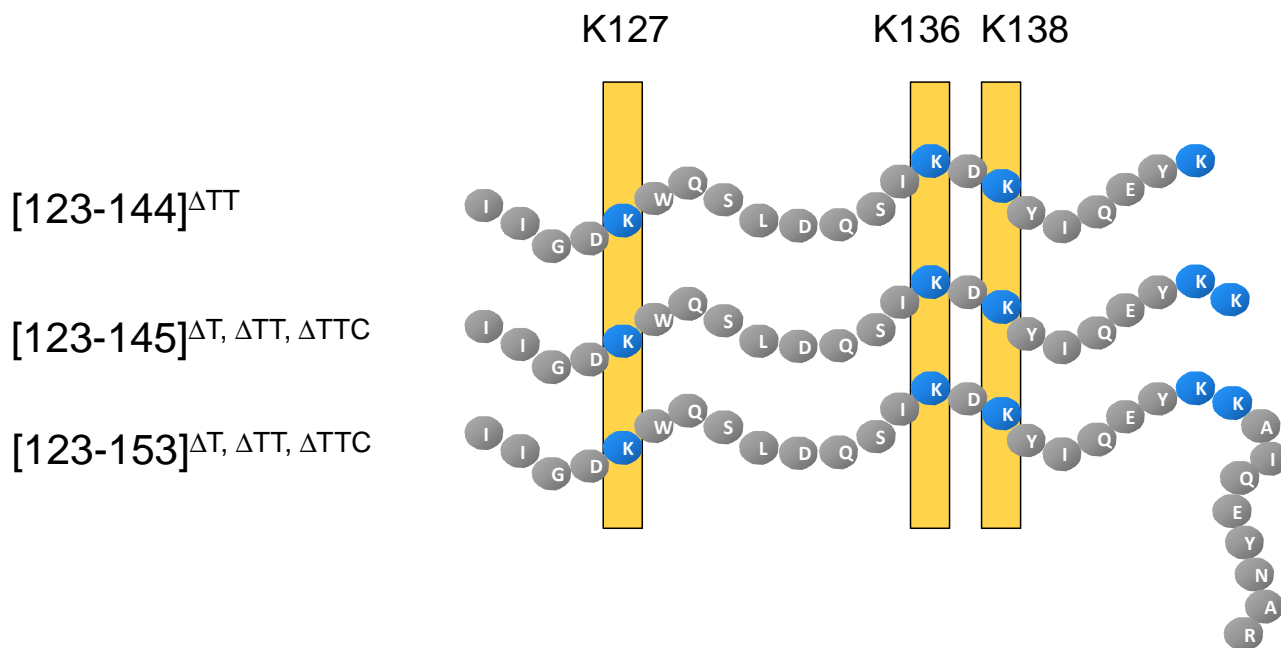
**Figure 58. Analyse MALDI-TOF des hétéroconjugués peptides-acides nucléiques provenant du complexe Abf2-ADN à l'issue du protocole expérimental 1.** Les échantillons ont été dilués au  $1/20^{\text{ème}}$  dans HCCA/TWA et un aliquot a été déposé selon la méthode en couche ultra-mince (Cadene and Chait, 2000; Gabant and Cadene, 2008). A. Analyse du mélange de peptides tryptiques en mode réflectron positif (gamme  $m/z$  500-3250, masses monoisotopiques). B. Analyse du mélange de peptides tryptiques en mode linéaire positif (gamme  $m/z$  3200-5200, masses moyennes). C. Acquisition en mode réflectron positif après incubation des billes Ni-NTA dans l'imidazole 250 mM. D. Cartographie peptidique des peptides non modifiés (en bleu) et des hétéroconjugués (en rouge). La base de Schiff réduite au  $\text{NaBH}_4$  est représentée par un triangle. Les lettres T et C correspondent à la Thymine et Cytosine, respectivement.



Une très bonne couverture de séquence a été observée dans cette protéolyse en phase hétérogène après dénaturation (Figure 58-D, couverture de séquence bleue). Notons que le peptide N-terminal [1-16] n'est pas détecté dans le surnageant de protéolysat (Figure 58-A) et qu'il est observé uniquement après élution à l'imidazole (Figure 58-C), ce qui est conforme nos attentes. Trois peptides [123-144], [123-145] et [123-153], sont observés avec la base de Schiff stabilisée et de un à trois nucléotides (Figure 58-B). Ces résultats indiquent une hétérogénéité à la fois au niveau de la protéolyse (clivages manqués) et de l'hydrolyse par la nucléase P1. Cette hétérogénéité complique les spectres et disperse le signal, ce qui n'est pas optimal. L'existence de ces clivages manqués (ou missed cleavage en anglais) pourrait être expliquée par un encombrement stérique de l'oligonucléotide 1 lors de la protéolyse. En effet, étant donné sa masse élevée d'environ 5 kDa, on peut supposer qu'une fois attaché de manière covalente à la protéine, l'oligonucléotide 1 empêche l'accès de la protéase aux résidus situés à proximité du site modifié. Par ailleurs, il est à noter que les peptides [123-144] et [123-145] non-modifiés sont observés de manière intense. Cette observation suggère soit qu'il reste de la protéine libre non pontée après purification, soit que des sites sont modifiés ailleurs dans la séquence. Concernant le peptide [123-153], il n'est pas observé non-modifié, ce qui indique qu'il résulte d'un missed cleavage dû à une modification de lysine en amont, c'est-à-dire des lysines 127, 136, 138, 144 ou 145.

Notre hypothèse initiale est qu'il y a un seul site actif par domaine HMG A et B. Les peptides observés sont chevauchants et recouvrent la même zone de séquence [123-144] qui se trouve dans le domaine B. Ainsi nous avons considéré que tous les hétéroconjugués observés concernaient le site actif du domaine B et qu'ils étaient donc tous modifiés sur le même résidu. En supposant que la trypsine ne peut pas cliver une lysine modifiée, et qu'une seule lysine est modifiée, dans la séquence commune à tous les produits observés, la Lys144 doit être exclue puisqu'elle est clivée. Il reste trois lysines: Lys127, Lys136 et Lys138, qui peuvent être considérées comme site actif potentiel (Figure 59).

**Figure 59. Illustration des lysines communes entre les trois hétéroconjugués [123-144], [123-145] et [123-153].** Les lysines communes K127, K136 et K138 sont représentées par un rectangle jaune.



## 6.2 Localisation du site actif présumé

Afin de localiser le site actif, les échantillons ont ensuite été analysés en nanoLC-IT CID en utilisant une liste d'inclusion correspondant aux hétéroconjugués peptide-acides nucléiques identifiés en MALDI-TOF MS. Cependant, comme illustré Figure 60, malgré la petite taille de l'oligonucléotide, les spectres de fragmentation CID étaient dominés par des pertes de neutres de phosphate, d'eau et d'acide nucléique qui appauvrissent l'information structurale. Les masses théoriques de ces différentes pertes de neutres sont présentées dans le Tableau 11. Les quelques fragments pertinents  $b_4$ ,  $b_6$  et  $b_{10}$  observés sans modification suggèrent que Lys127 n'est pas modifiée, mais qu'il subsiste une ambiguïté entre Lys136 et Lys138. Aucun ion y n'est observé dans ce spectre MS/MS.



Figure 60. Spectre de fragmentation nanoLC-IT CID de l'hétéroconjugué [123-145]<sup>ΔT</sup> m/z 1110.2 [M+3H]<sup>3+</sup>.

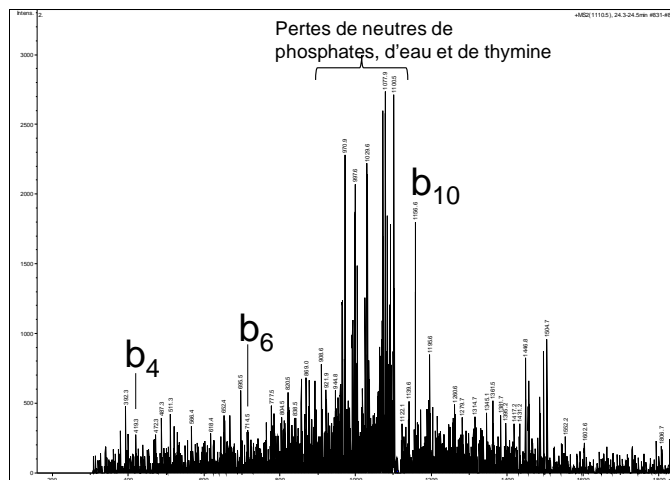
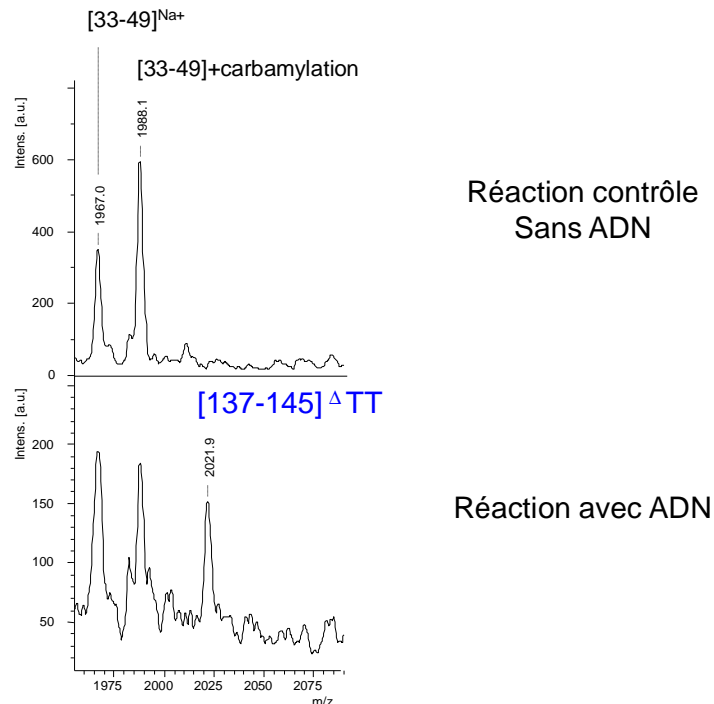


Tableau 11. Masses des pertes de neutres issues de la fragmentation CID de la modification ΔT

Pertes de neutres	Structure de la modification ΔT après les pertes de neutres	Formule chimique des pertes de neutres	Masse théorique monoisotopique (Da)	Masse théorique moyenne (Da)
<sup>ΔT</sup> – Pi (phosphate inorganique)		H <sub>3</sub> PO <sub>4</sub>	97.9769	97.9952
<sup>ΔT</sup> – nucléoside (base azotée + ribose)		C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O <sub>4</sub>	224.0797	224.2138
<sup>ΔT</sup> – nucléotide (nucléoside + phosphate)		C <sub>10</sub> H <sub>15</sub> N <sub>2</sub> O <sub>8</sub> P	322.0566	322.2090
<sup>ΔT</sup> – Pi – nucléotide		C <sub>10</sub> H <sub>18</sub> N <sub>2</sub> O <sub>12</sub> P <sub>2</sub>	420.0335	420.2042

Afin de lever l'ambiguïté entre Lys136 et Lys138, une protéolyse supplémentaire des peptides trypsiques a été réalisée avec l'AspN. Le spectre MALDI-TOF indique la présence de l'hétéroconjugué  $[137-145]^{\Delta}TT$  issu du clivage au niveau de Asp137 (Figure 61). Ce résultat suggère que Lys138 est le résidu modifié.

**Figure 61. Analyse MALDI-TOF des peptides hétéroconjugués après clivage AspN**



### 6.3 Vérification des résultats par mutagenèse dirigée

Afin de vérifier les résultats obtenus, le résidu Lys138 a été remplacé par Gln138 par mutagenèse dirigée. Lorsqu'une protéine est mutée, notamment lorsqu'on remplace un résidu chargé en résidu non chargé, comme la lysine en glutamine, la structure peut être perturbée. Ces changements de conformations peuvent affecter l'affinité de la protéine mutée pour son partenaire et donc au final altérer l'activité de la protéine. Ainsi, afin de pouvoir conclure si une diminution d'activité lyase d'un mutant d'Abf2 est due à l'absence du site actif ou à un changement de conformation, il est important de vérifier son affinité pour l'ADN.

Le simple mutant Abf2 K138Q a donc été produit, purifié et son affinité à l'ADN ainsi que son activité lyase ont été mesurées. Les résultats ont montré qu'il n'y a pas de différences majeures entre K138Q et Abf2 sauvage en termes d'affinité à l'ADN, d'activité lyase et d'efficacité de piégeage.

Ce résultat inattendu nous a conduit à considérer 1) l'existence potentielle d'un phénomène de compensation entre le domaine HMG A non muté et le domaine HMG B muté, ce qui biaiserait les tests d'affinité à l'ADN et d'activité lyase et 2) l'éventualité que Lys138 ne soit pas la lysine

catalytique mais plutôt une lysine non compétente pour l'activité lyase, c'est-à-dire une lysine capable de former une base de Schiff avec un site abasique mais qui ne peut pas aller jusqu'au bout de la réaction d'hydrolyse. Ceci impliquerait l'existence d'autres résidus modifiés dans la zone [123-153].

Concernant l'existence d'un phénomène de compensation entre les domaines HMG A et B, un alignement de séquence entre les domaines A et B a été effectué afin d'identifier d'éventuelles lysines conservées entre les deux domaines. Les résultats sont présentés Figure 62 et indiquent qu'il y a sept lysines participant à la zone homologue des deux domaines, dont Lys138 du domaine B qui est homologue avec Lys65 du domaine A. Les quatre autres lysines du peptide [123-145] ont également un homologue dans le domaine A.

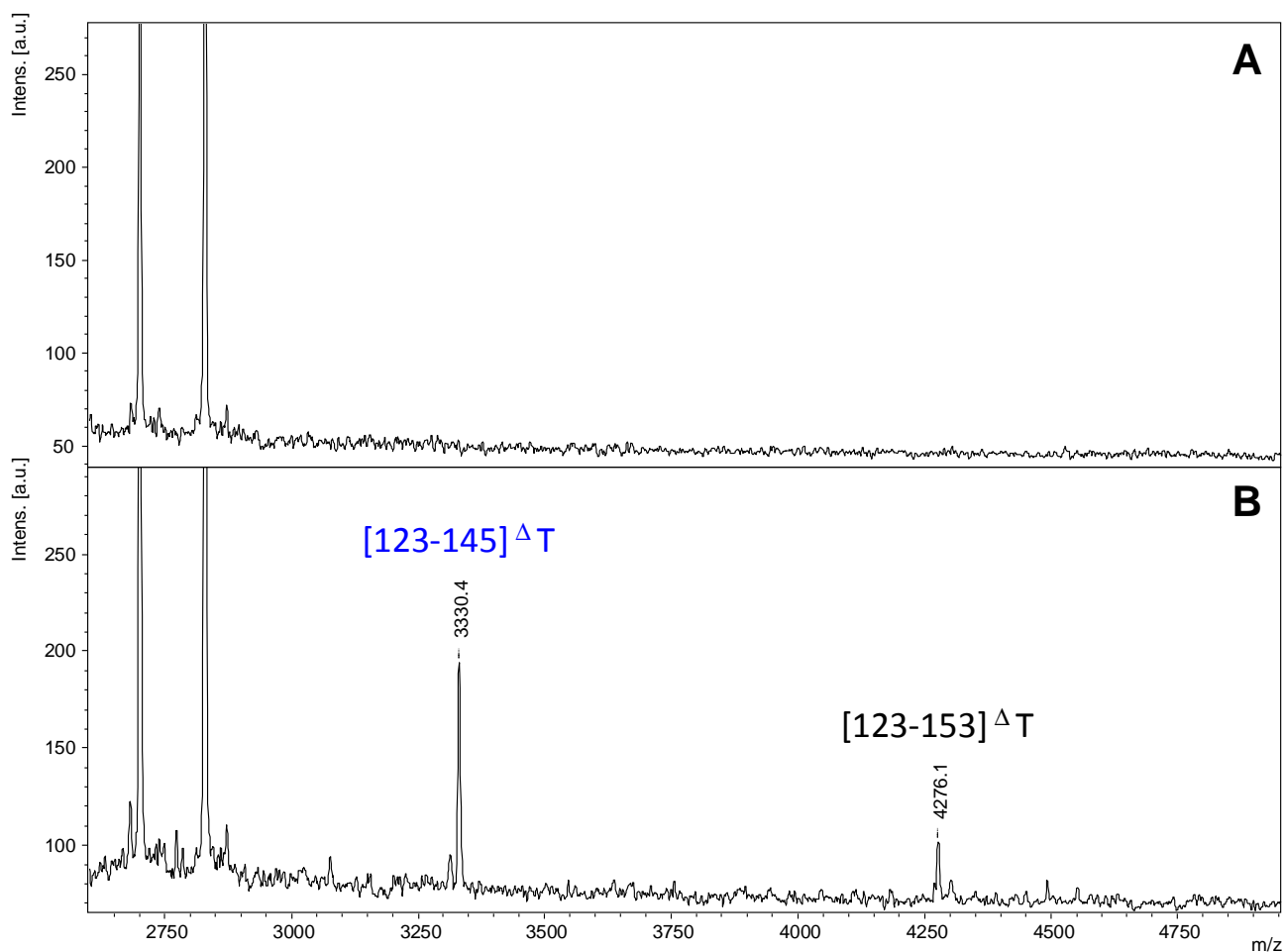
**Figure 62. Alignement de séquence entre les domaines HMG A et B d'Abf2.** Sept lysines sont alignées dans la zone d'homologie entre les deux domaines.

		18		49	54	63 65	71 72	
Domaine A	1	KASKRTQLRNELIKQGPKRPTSAYFLYLQDHRSQFVKENPTLRPAEISKIAGEKWQNL	LEAD	IK	EY	SERKKLYSEYQKAKKEFD	85	
		PK+P	+ Y	+ RSQ	++P	++ KI	G+KWQ+L+	IK+KYI E KK EY
Domaine B	90	PKKPAGPFIKYANEVRSQVFAQHPDKS	QLDLMKII	GDKWQSLDQ	SIKDKYIQEYK	KAIQEYNARYPLN	157	
		91		122	127	136 138	144 145	

Le double mutant K65Q, K138Q pourrait ainsi être produit, purifié et testé en termes d'affinité à l'ADN et d'activité lyase. Ceci n'a pas encore été réalisé pour le moment. Par ailleurs, d'après cette redondance fonctionnelle, un hétéroconjugué contenant Lys65 aurait dû être observé. Toutefois, si ce peptide [50-72] équivalent du [123-145] existe, le fait de ne pas le détecter pourrait être expliqué par le rapport signal sur bruit réduit des peptides modifiés. De plus, le peptide [50-65] ( $m/z$  1871.9) est clairement observé alors que le peptide [123-138] n'est pas visible dans le contrôle. Afin de s'affranchir d'un éventuel phénomène de compensation et ainsi mieux évaluer l'importance de Lys138 sur l'activité lyase, le domaine HMG B seul muté en K138Q pourrait être produit. Toutefois, la production du domaine HMG B seul se heurte à des problèmes de stabilité de la protéine, ce qui empêche son étude.

Pour tester l'hypothèse de l'existence d'autres résidus modifiés dans la zone [123-153], le double mutant Abf2 K136Q, K138Q a été produit, purifié et analysé avec le protocole expérimental 1. Le mélange de peptides obtenu a été analysé par MALDI-TOF. La partie haute du spectre, contenant le signal d'intérêt, est présentée Figure 63.

**Figure 63. Analyses MALDI-TOF des hétéroconjugués peptides-acides nucléiques issus du complexe Abf2 K136Q, K138Q-ADN après le protocole expérimental 1.** A. Abf2 K136Q, K138Q contrôle. B. Abf2 K136Q, K138Q-ADN. Les solutions de matrice et d'échantillon ont été déposées selon la méthode en couche ultra-mince puis analysées en mode linéaire positif. La base de Schiff réduite au NaBH<sub>4</sub> est représentée par un triangle. La lettre T correspond à la Thymine.



L'hétéroconjugué [123-144], préalablement observé avec Abf2 sauvage, n'est plus détecté avec le double mutant Abf2 K136Q, K138Q, ce qui corrobore l'annotation initiale que Lys138 est modifiée. Toutefois, les deux hétéroconjugués  $[123-145]^{\Delta T}$  et  $[123-153]^{\Delta T}$  sont encore observés, ce qui démontre qu'ils ne sont modifiés ni sur Lys138 ni sur Lys136 et donc que d'autres résidus sont concernés. Contrairement à notre postulat initial, au moins deux résidus sont donc modifiés dans la zone [123-153], ce qui complique fortement la découverte du site actif dans le domaine B. En plus de Lys138 (et potentiellement Lys136), Lys144 et Lys145 sont désormais à considérer. Il est intéressant de noter que la Lys144, si elle est modifiée, ne peut produire que les peptides [123-145] et [123-153] dès lors que l'on supprime les Lys136 et 138. Ce comportement est cohérent avec le fait que le peptide [123-144] soit dû à une modification à bas bruit du résidu 138 (et éventuellement 136).

La localisation de ces résidus pourrait être envisagée par fragmentation ETD qui a la faculté de préserver les modifications labiles. Dans cette optique, nous avons développé une seconde stratégie expérimentale afin d'augmenter le rapport signal sur bruit des hétéroconjugués peptide-acides nucléiques, et ainsi espérer obtenir des spectres ETD exploitables et fiables.

## 7. Stratégie expérimentale 2

La seconde stratégie, illustrée Figure 64, a consisté à faire un double enrichissement par chromatographie d'échange d'anions (MonoQ) :

- un premier enrichissement au niveau protéique pour purifier le complexe Abf2-ADN, et
- un deuxième enrichissement au niveau peptidique après l'étape de protéolyse.

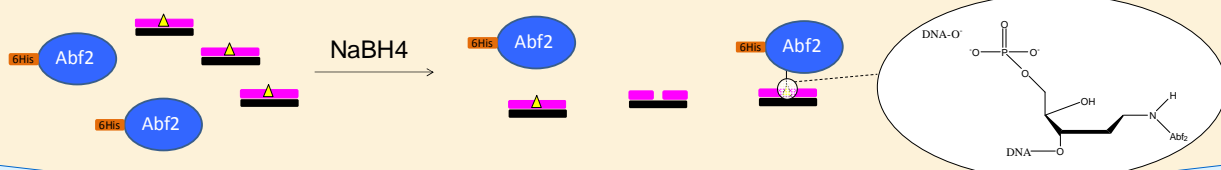
Enfin, la partie oligonucléotide des hétéroconjugués enrichis a été hydrolysée par la nucléase P1 avant micropurification ZipTipC18 et analyses par spectrométrie de masse MALDI-TOF et ESI-IT-ETD en infusion. La fragmentation ETD est connue pour préserver les modifications labiles sur les peptides et ainsi promouvoir la fragmentation du squelette peptidique. L'analyse de l'échantillon en infusion a été préférée à l'HPLC car elle permet d'optimiser les paramètres de fragmentation et d'accumuler du signal.

Concernant l'enrichissement par chromatographie d'échange d'anions, il repose essentiellement sur la charge négative des oligonucléotides et ne requiert donc pas la présence de la biotine. Ainsi pour la stratégie 2, ce sont les oligonucléotides 2 et 4 (Tableau 9) qui ont été utilisés pour former le duplex ADN contenant un site 5'dRP.

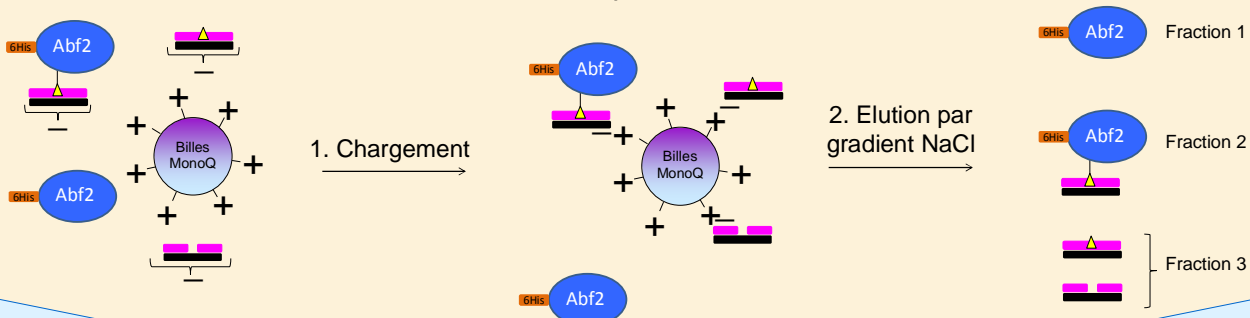
Figure 64. Stratégie expérimentale 2 pour la purification du complexe covalent Abf2-ADN et l'analyse d'hétéroconjugués peptide-acide-nucléique en spectrométrie de masse.

## Stratégie Expérimentale 2

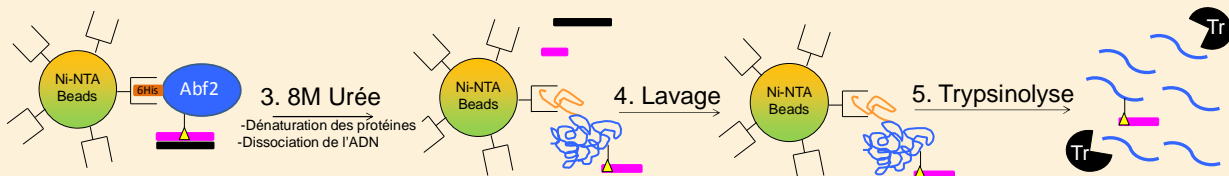
Formation d'un complexe covalent : Piégeage de l'intermédiaire Abf2-ADN par NaBH4



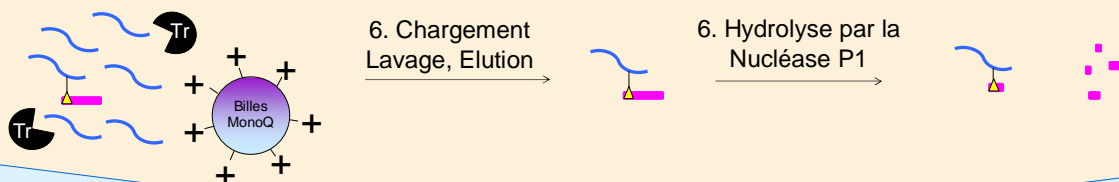
Purification du complexe covalent par chromatographie d'échange d'ions Mono Q :  
Élimination de l'excès de protéine libre et d'ADN libre



Fixation sur billes Ni-NTA de la fraction 2 et Protéolyse sur billes



Enrichissement sur billes MonoQ et Hydrolyse de l'ADN par la Nucléase P1



Analyse par spectrométrie de masse dans des conditions d'analyses pour peptides

MALDI-TOF MS

Identification des hétéroconjugués  
peptide-acide nucléique



ESI IT-ETD en infusion

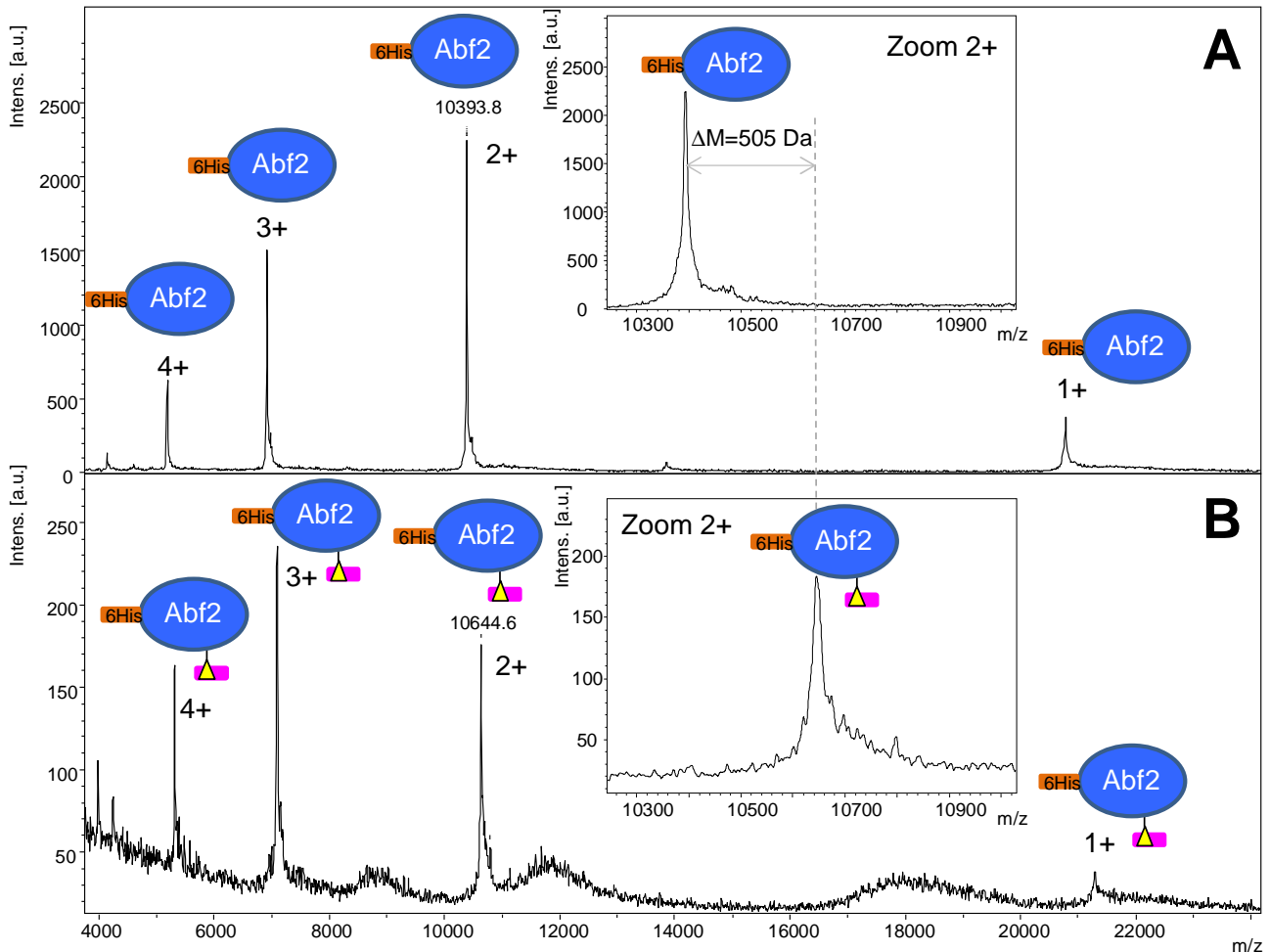
Localisation du ou des résidu(s)  
cible(s)

## 8. Les résultats de localisation de la stratégie 2

### 8.1 Détermination de la stœchiométrie du complexe Abf2-ADN

Etant donné qu'Abf2 possède deux domaines HMG A et B, nous supposons qu'ils ont tous les deux une activité lyase. A ce stade, nous nous sommes demandés si les problèmes rencontrés ne pourraient pas être dus à la fixation d'un ADN par domaine, ce qui correspondrait à une stœchiométrie protéine:substrat 1:2. Après purification, un traitement à la nucléase P1 a été réalisé et la stœchiométrie du complexe a été déterminée par spectrométrie de masse MALDI-TOF. Le spectre obtenu est présenté Figure 65. Une première distribution avec des pics intenses et résolus est observée. L'incrément de masse mesuré par rapport à Abf2 seule est de 505 Da, ce qui correspond à la modification  $\Delta T$  illustrée dans le Tableau 10.

**Figure 65. Analyse MALDI-TOF du complexe Abf2-ADN après purification et hydrolyse par la Nucléase P1. A.** Abf2 seule. **B.** Complexe Abf2-ADN après purification monoQ et hydrolyse par la Nucléase P1.

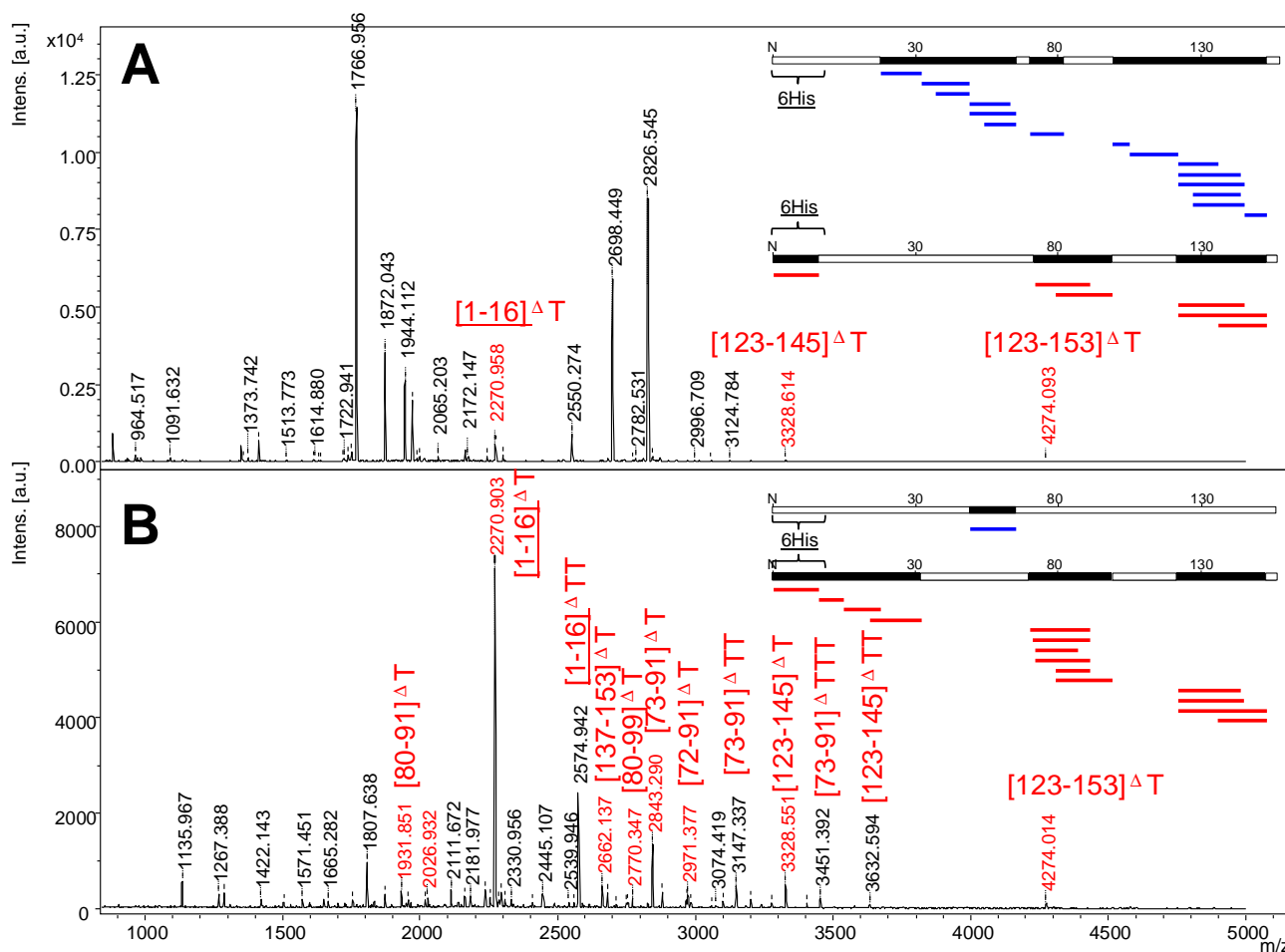


Une seule modification est donc observée sur Abf2, ce qui traduit une stœchiométrie 1:1. Ce résultat est conforme à nos attentes du fait de l'encombrement stérique très probable de deux duplex ADN de 34mers pour une protéine Abf2 de 21 kDa. Une deuxième distribution de faible intensité avec des massifs larges et non résolus est détectée et correspond au complexe Abf2-ADN non hydrolysé par la nucléase P1. Par ailleurs, le fait de ne pas observer Abf2 seule dans le spectre B indique que la purification monoQ est très efficace.

## 8.2 Identification des hétéroconjugués peptide-acides nucléiques

Le complexe covalent Abf2-ADN a été formé en présence de 66  $\mu\text{M}$  d'Abf2 et de duplex ADN et de 5 mM de  $\text{NaBH}_4$ , soit un rapport de protéine:substrat de 1:1. A la suite du protocole expérimental 2, le mélange de peptides a été analysé par MALDI-TOF MS (Figure 66-B).

**Figure 66. Analyse MALDI-TOF des hétéroconjugués peptides-acides nucléiques provenant du complexe Abf2-ADN à l'issue du protocole expérimentale 2.** Les échantillons ont été dilués au  $1/20^{\text{ème}}$  dans HCCA/TWA et un aliquot a été déposé selon la méthode en couche ultra-mince. Les spectres A et B correspondent à l'analyse du mélange de peptides tryptiques en mode réflectron positif avant et après enrichissement sur la deuxième monoQ, respectivement.

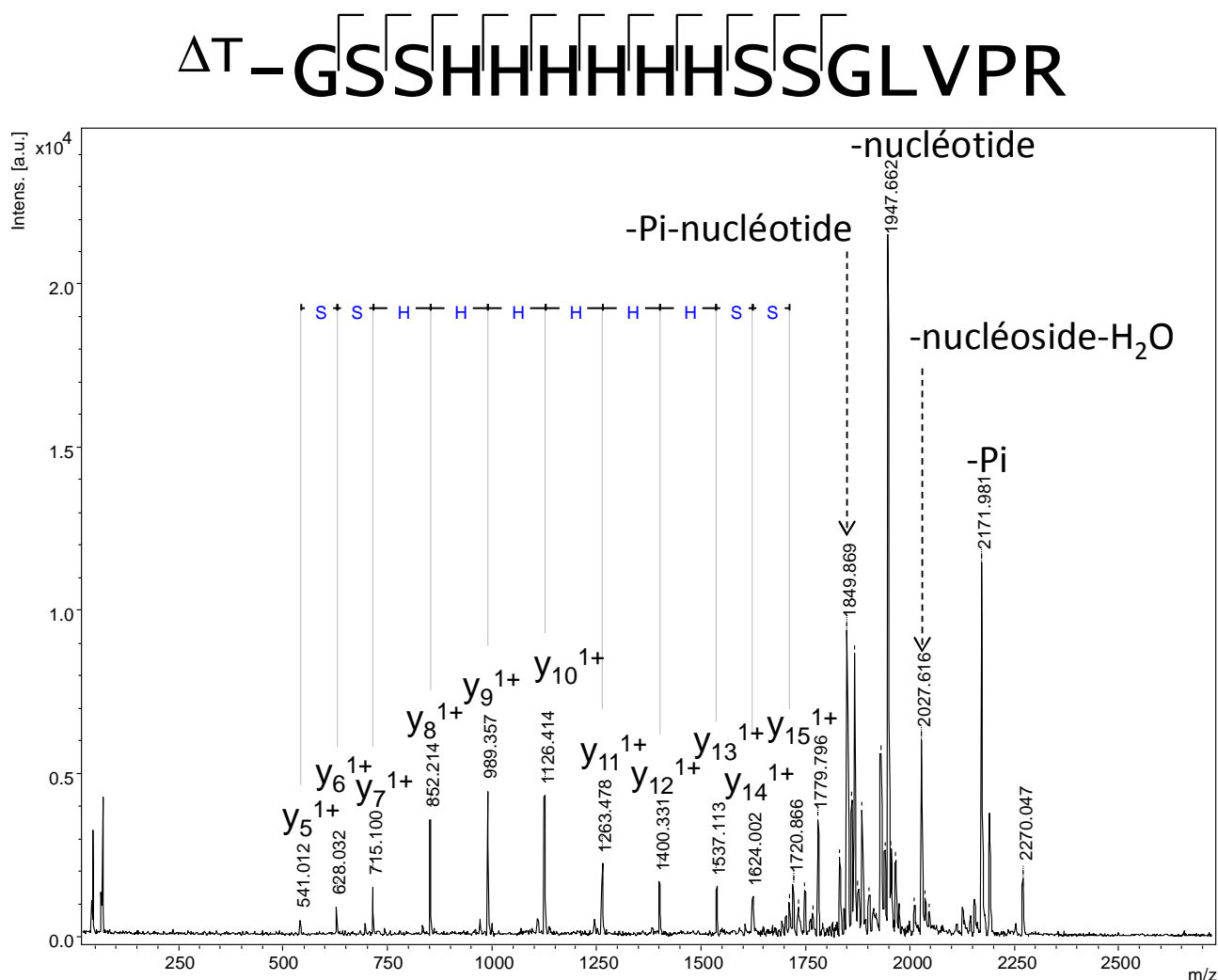


De nombreux peptides modifiés sont détectés. En plus des hétéroconjugués [123-144], [123-145] et [123-153] déjà observés avec la stratégie 1, dix autres hétéroconjugués répartis en N-terminal et au



niveau de la zone 72-99 sont mis en évidence. Concernant la zone N-terminale, l'observation de l'hétéroconjugué [1-16] de l'étiquette 6 His est inattendue et révèle l'existence d'une réaction non spécifique, très certainement au niveau de l'amine primaire N-terminale. Cette annotation est confirmée par une fragmentation en mode LIFT (Figure 67).

Figure 67. Analyse MALDI-TOF/TOF LIFT du  $m/z$  2270.0



Qu'en est-il alors des autres hétéroconjugués observés ? Correspondent-ils tous à une réduction non spécifique par NaBH<sub>4</sub> ou sommes-nous en présence d'un mélange d'hétéroconjugués contenant le ou les site(s) catalytique(s) et d'hétéroconjugués contenant des sites non spécifiques ? A ce stade, la spectrométrie de masse ne peut répondre à ces questions et vu le nombre de lysines concernées par les hétéroconjugués observés, la mutagenèse dirigée n'est pas envisageable, car elle serait trop chronophage et fastidieuse. Il faut donc optimiser les conditions de piégeage du complexe transitoire Abf2-ADN par NaBH<sub>4</sub> afin d'éviter des réactions non spécifiques.

D'autres conditions de pontage ont été testés sur Abf2 entière afin de tenter de gagner en spécificité: diminution des concentrations à 0.25  $\mu$ M en ADN avec un rapport Abf2:ADN 1:1 (Figure 68) ou 4:1

(Figure 69) afin de se placer dans des conditions de « single turnover », c'est à dire dans des conditions où chaque molécule d'enzyme ne peut faire qu'une seule réaction d'hydrolyse du substrat ADN. Cependant, à chaque tentative, plusieurs peptides modifiés pour chaque domaine d'Abf2 sont identifiés.

**Figure 68. Analyse MALDI-TOF d'une réaction Abf2-ADN dans des conditions diluées et avec un rapport E:S 1:1.**

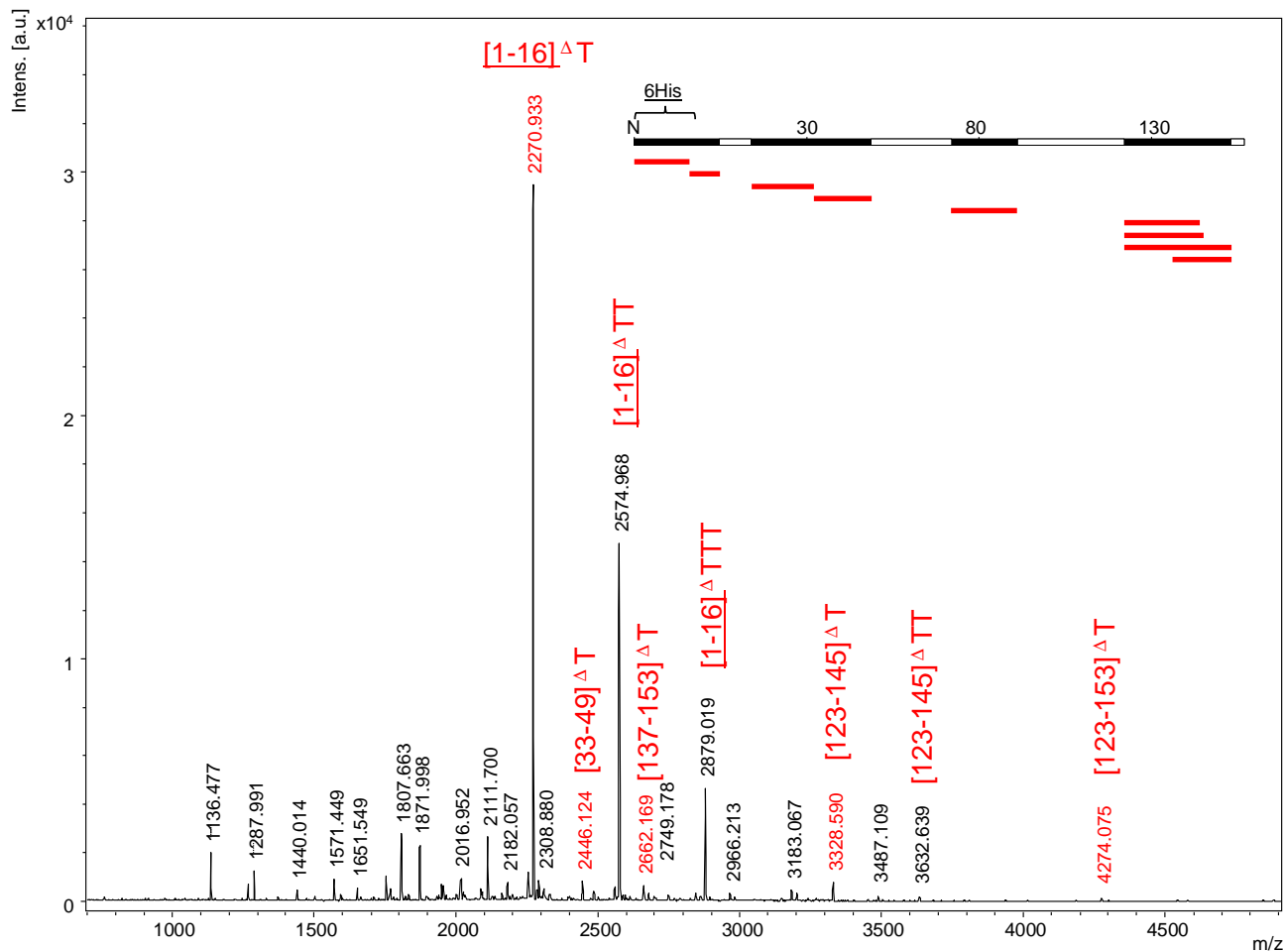
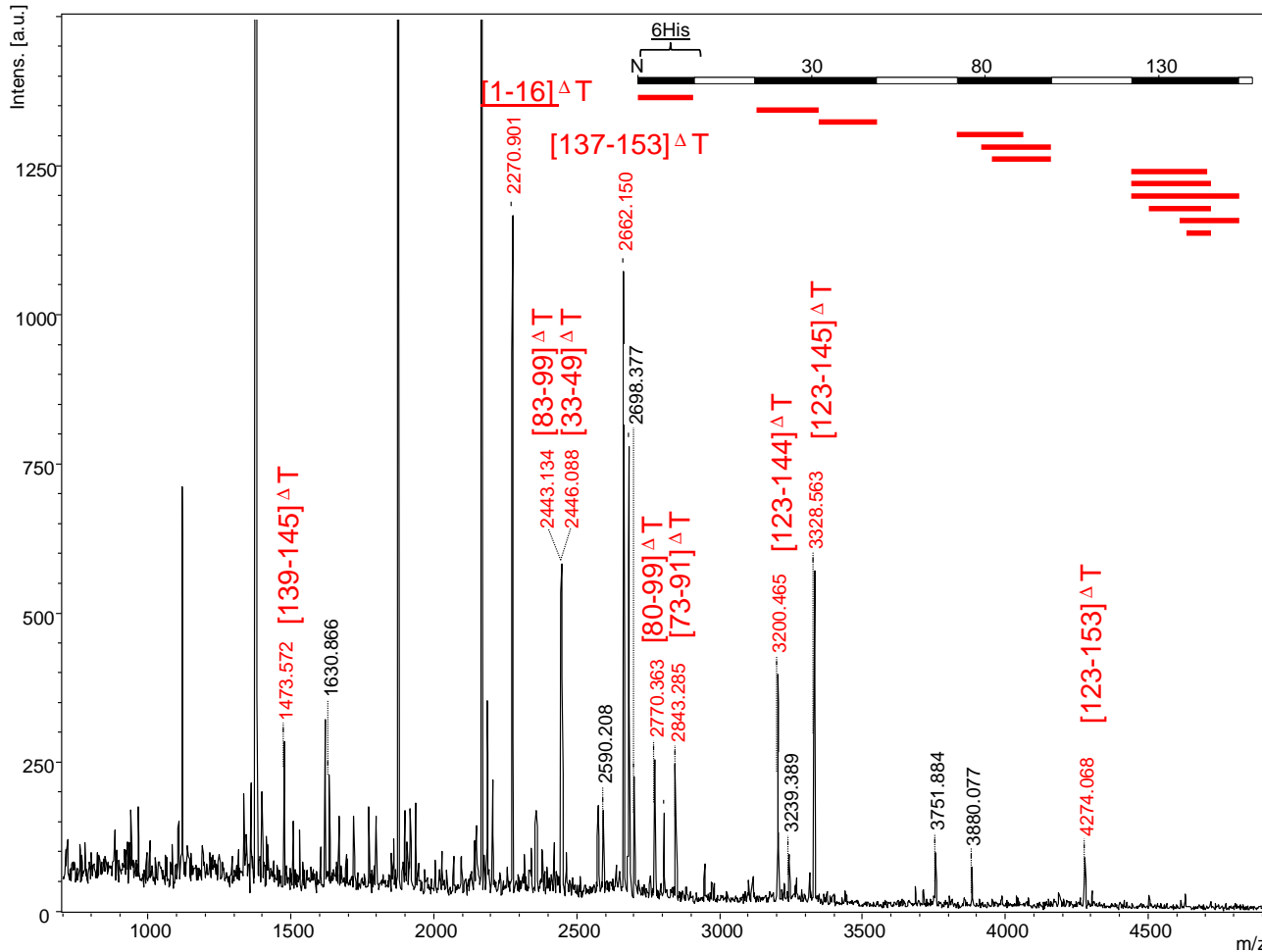
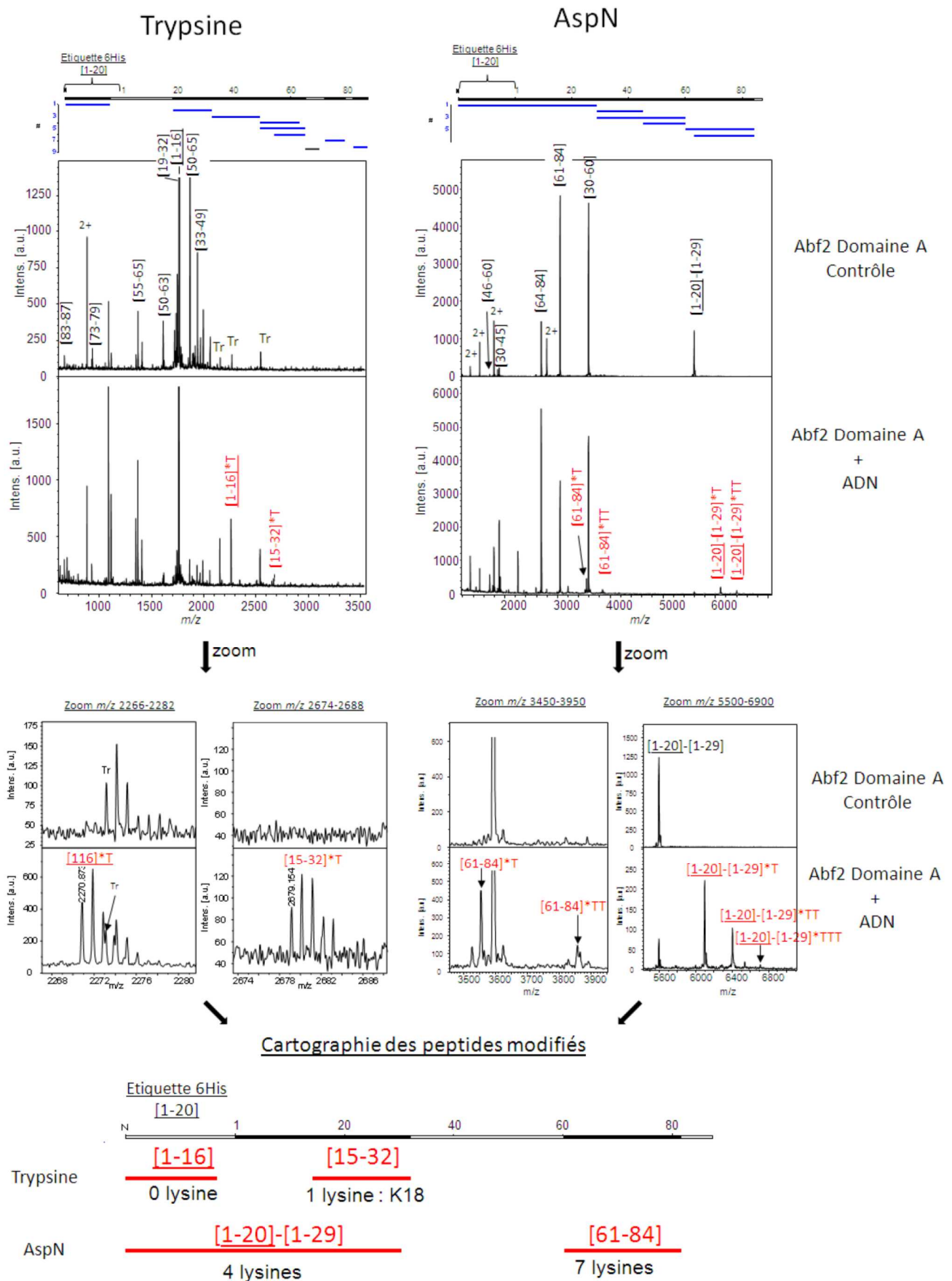


Figure 69. Analyse MALDI-TOF d'une réaction Abf2-ADN dans des conditions diluées et avec un rapport E:S 4:1.



D'autres essais ont été effectués en utilisant le domaine A d'Abf2 seul et un duplex ADN contenant un site AP. Plusieurs peptides recouvrant différentes zones du domaine A ont été observés modifiés après protéolyse par la trypsine ou par l'Asp-N : [1-16] de l'étiquette 6 histidines N-terminale, [15-32], [1-20]-[1-29] (peptide contenant l'étiquette 6 histidines et le début de séquence d'Abf2) et [61-84] (correspondant à la séquence 134-157 du domaine B) (Figure 70).

**Figure 70. Analyse MALDI-TOF des hétéroconjugués peptides-acides nucléiques provenant du complexe entre le domaine A d'Abf2 et l'ADN à l'issue du protocole expérimental 2 modifié.** La protéolyse par la trypsine a été réalisée en solution et il n'y a pas eu de 2ème étape d'enrichissement des peptides-oligonucléotides. Les échantillons ont été dilués au 1/20<sup>ème</sup> dans HCCA/TWA et un aliquot a été déposé selon la méthode en couche ultra-mince.



Ce résultat montre encore une fois la présence de plusieurs résidus modifiés. Parmi eux, Lys18 est suspectée modifiée puisqu'il y a une seule lysine dans le peptide [15-32]. Des données MS/MS CID de ce peptide corrobore cette localisation. Le mutant Abf2-domaine A-K18Q a été produit et purifié. Comme pour Lys138, le peptide [15-32] n'est plus observé avec ce mutant confirmant la modification de Lys18. L'affinité à l'ADN et l'activité lyase de ce mutant Abf2-domaine A-K18Q ont été mesurées. Les résultats montrent que l'activité lyase mais également l'affinité à l'ADN sont diminuées d'un facteur 3 et 3.5, respectivement, par rapport à celles d'Abf2-domaine A-sauvage. Ces résultats suggèrent que la structure du mutant Abf2-domaine A-K18Q est altérée, ce qui affecte son activité. La Lys18 ne semble donc pas être une lysine catalytique. .

## 8.3 Aspects méthodologiques

### 8.3.1 L'apport du deuxième enrichissement monoQ

Afin d'évaluer l'influence du deuxième enrichissement monoQ sur le nombre et le rapport signal sur bruit des hétéroconjugués, une analyse MALDI-TOF MS a été réalisée avant cette étape d'enrichissement (Figure 66-A). On peut observer que l'enrichissement permet d'augmenter le nombre d'hétéroconjugués détectés de 6 à 14 et que leur rapport signal/bruit est amélioré d'un facteur d'environ 5 à 8 fois (augmentation d'un facteur 4.9 pour le  $m/z$  3328.6 et 7.5 pour le  $m/z$  2270.9). On note que seul un peptide non modifié est annoté après enrichissement : [50-65]. Ces résultats indiquent que l'enrichissement monoQ est efficace et bénéfique pour l'observation des hétéroconjugués. Toutefois, la présence, après enrichissement, d'hétéroconjugués portant jusqu'à trois acides nucléiques indique une hydrolyse incomplète par la nucléase P1 qui provient très certainement de la présence accrue de NaCl (650 mM) nécessaire pour l'élution de l'hétéroconjugué peptide-oligonucléotide. Or, la dispersion du signal sur plusieurs pics nuit à la sensibilité de détection. L'élimination du NaCl par un dessalage ZipTipC18 serait probablement bénéfique pour l'activité de la nucléase P1 et donc pour l'amélioration du rapport signal sur bruit.

### 8.3.2 Le cas du peptide tryptique [1-16] de l'étiquette 6 Histidine

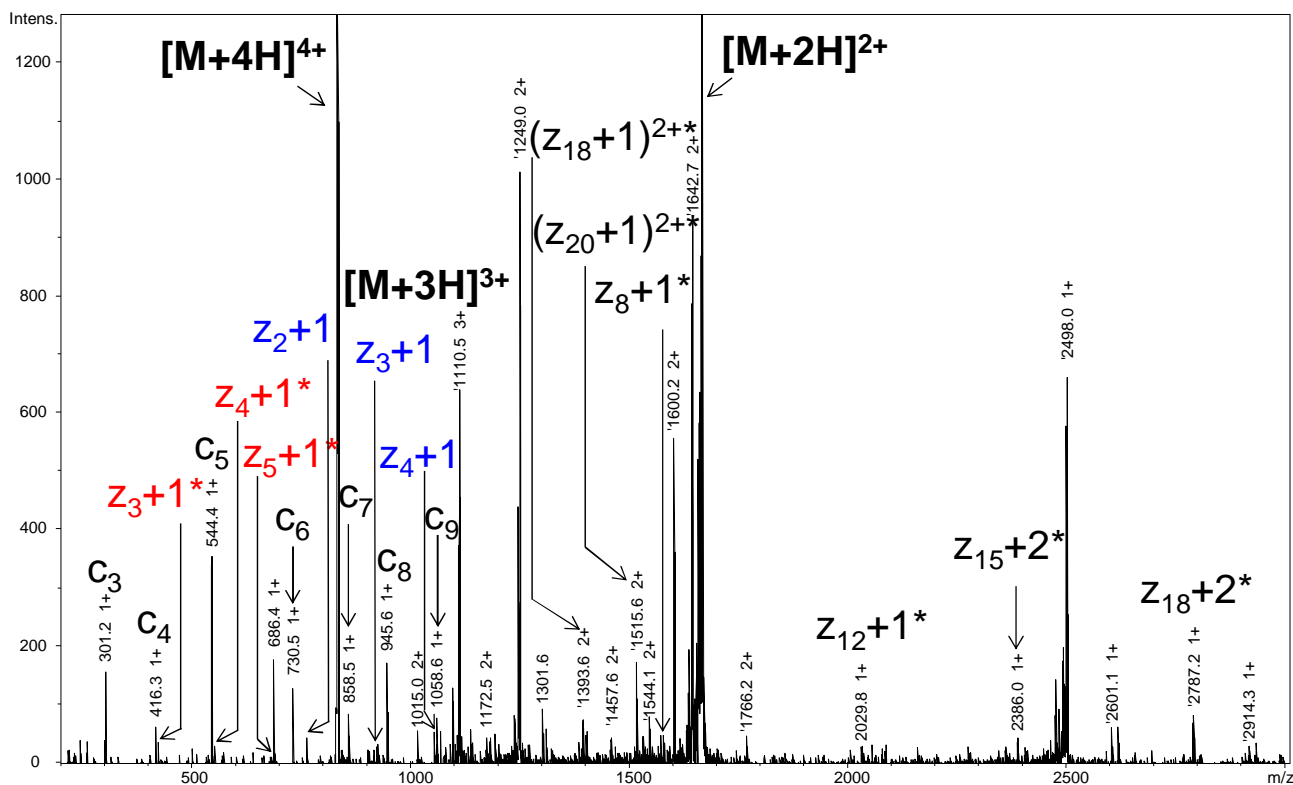
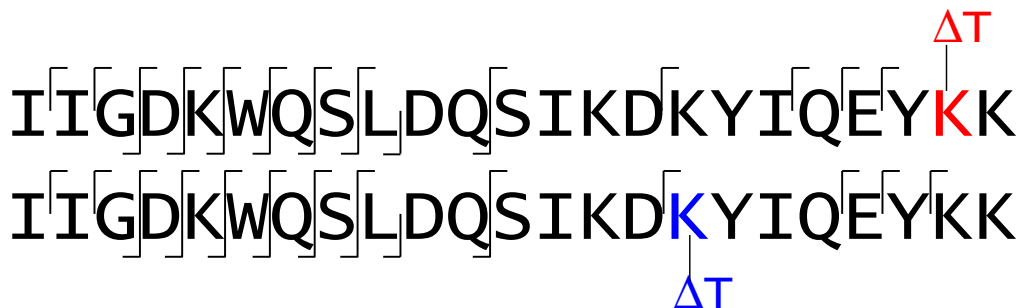
Au-delà de la détection d'une réaction non spécifique sur l'étiquette 6 Histidines, il est surprenant d'observer le peptide [1-16] en tant que tel, dans la mesure où la protéolyse a été réalisée sur le complexe Abf2-ADN fixé sur les billes Ni-NTA. En effet, dans le cas où l'oligonucléotide est fixé sur un site en dehors de l'étiquette 6 His, cette dernière est accessible et disponible pour l'interaction avec les billes Ni-NTA. Après protéolyse de la protéine encore fixée sur les billes, les peptides sont en solution mais l'étiquette reste sur les billes. C'est pourquoi, logiquement, l'étiquette n'est pas

détectée, comme illustré au niveau de la cartographie peptidique bleue (Figure 66-A). Lorsque la protéolyse est réalisée en solution (Figure 70), l'étiquette est détectée *via* le peptide [1-16] ( $m/z$  monoisotopique théorique  $[M+H]^+ = 1767.849$ ). On notera que la distribution isotopique de ce peptide se superpose partiellement avec la celle du peptide [19-32] ( $m/z$  monoisotopique théorique  $[M+H]^+ = 1765.890$ ), ce qui peut affecter la détection du peptide [1-16]. Dans le cas où l'oligonucléotide est fixé au niveau de l'étiquette 6 His, deux cas de figures sont possibles : soit l'oligonucléotide ne perturbe pas l'interaction de l'étiquette 6 His avec les billes Ni-NTA et l'on se trouve alors dans le cas précédent; soit l'oligonucléotide, *via* un encombrement stérique, empêche l'interaction avec les billes Ni-NTA, ce qui laisse l'hétéroconjugué dans le surnageant. Ensuite, l'hétéroconjugué est censé être éliminé au cours de l'étape de lavage avant protéolyse (cf. étape 4 de la stratégie 2, Figure 64). Ainsi, que l'étiquette soit modifiée ou non, l'approche expérimentale développée devait en théorie nous permettre de ne pas l'observer. Néanmoins, le peptide [1-16] est observé dans les spectres sous une forme modifiée uniquement. La présence de ce peptide pourrait être expliquée par un phénomène d'adsorption sur les billes. Une autre possibilité serait que la protéine peut se dimériser et que seule l'étiquette du deuxième monomère, à distance du support, est modifiée. Le peptide [1-16] modifié n'a pas été observé du tout au cours de la stratégie 1. Ceci peut s'expliquer par le fait que l'étiquette 6 His est utilisée pour la purification et que toute la population d'Abf2 contenant l'étiquette modifiée perd très probablement son affinité pour les billes Ni-NTA. Dans ce cas, l'hypothèse d'une dimérisation n'est pas nécessaire pour interpréter les données.

### 8.3.3 Mise en évidence d'isomères par fragmentation ETD

Nous avons déjà constaté que la fragmentation des hétéroconjugués en CID générait peu de fragments informatifs, produisant essentiellement des pertes de neutres. Dans l'objectif d'obtenir des spectres de fragmentation exploitables pour localiser le ou les résidus modifiés, une fragmentation ETD, décrite pour préserver les modifications labile, a été réalisée. La Figure 71 montre la fragmentation du peptide [123-145]<sup>ΔT</sup> déjà observé lors de l'expérience initiale avec la stratégie 1.

**Figure 71. Analyse ESI-IT ETD en infusion de l'hétéroconjugué [123-145]<sup>ΔT</sup>.** L'ion précurseur est de *m/z* 832.9 et d'état de charge 4+. Les ions fragments modifiés de couleur rouge et non modifiés de couleur bleue démontrent l'existence d'une modification de Lys138 et Lys144, respectivement.



Ce spectre montre la présence d'ions fragments  $y_4$ ,  $y_5$  et  $y_6$  modifiés, indiquant l'existence d'une modification sur Lys144, mais il montre également la présence d'ions fragments  $y_2$ ,  $y_3$  et  $y_4$  non modifiés, démontrant que la modification est ailleurs, à savoir sur la Lys138. Ce spectre indique donc la présence d'isomères, c'est-à-dire de peptides qui portent le même type et le même nombre de modifications mais sur des résidus différents. Ces peptides sont par définition isobares et co-sélectionnés en fragmentation.

## 9. Conclusions

Afin de localiser le ou les site(s) catalytique(s) de l'activité lyase d'Abf2, deux stratégies basées sur une purification du complexe covalent Abf2-ADN suivi d'une protéolyse sur billes par la trypsine, d'une hydrolyse de l'oligonucléotide par la Nucléase P1 puis d'une analyse MS ont été développées. La première approche met en œuvre une purification du complexe Abf2-ADN en deux étapes : chromatographie d'affinité avidine-biotine et chromatographie d'affinité sur métal immobilisé. Une combinaison de protéolyse par la trypsine et l'AspN suivies d'analyses ESI-IT CID MS/MS et/ou MALDI-TOF MS a conduit à l'identification de la lysine 138 comme étant le résidu modifié. Sur la base de cette identification, la protéine mutée Abf2 K138Q a été produite et purifiée et ne présente aucune différence par rapport à Abf2 sauvage en termes d'affinité à l'ADN, d'activité lyase et d'efficacité de piégeage. Ceci suggère soit l'existence d'un phénomène de compensation entre le domaine HMG A non muté et le domaine HMG B muté, soit que la Lys138 est une lysine non compétente pour l'activité lyase. Un cas où les réactions non-spécifiques dominent l'étape de piégeage a déjà été montré dans la littérature (Haracska *et al.*, 2003). Une réaction avec le double mutant K136Q K138Q a permis d'identifier au moins deux lysines dans le domaine HMG B comme étant un site catalytique potentiel : Lys138, ainsi que Lys144 ou Lys145. La Lys136 ne peut pas non plus être exclue de la liste des lysines candidates. Afin de localiser les résidus potentiellement modifiés, c'est-à-dire Lys65, homologue de Lys138 dans le domaine HMG A, ou Lys144, Lys145 ou d'autres, une deuxième approche plus sensible a été développée. Cette dernière combine une purification du complexe Abf2-ADN et un enrichissement en hétéroconjugués peptide-oligonucléotide. Cette stratégie repose sur la chromatographie d'échanges d'anions (monoQ) et a permis d'améliorer les rendements de purification de la protéine modifiée et de ses peptides modifiés et d'augmenter le rapport signal sur bruit des hétéroconjugués en MS. Elle a également permis d'obtenir des fragments informatifs en ETD. Cette méthode révèle qu'en plus de la zone [123-153], cinq autres zones sont modifiées dont l'extrémité N-terminal de l'étiquette 6 histidines. L'observation d'une modification de l'étiquette 6 histidines témoigne d'une réaction non spécifique lors de la première étape du protocole : le piégeage au NaBH<sub>4</sub>. La présence de réactions non spécifiques sur d'autres résidus d'Abf2 ne peut alors être exclue. Afin de tenter de gagner en spécificité, les concentrations en enzyme et en substrat ont été diminuées à 0.25-1 μM dans des conditions de « single turnover », mais sans succès. Des essais sur le domaine A exprimé seul ont été effectués et montrent également plusieurs zones modifiées dont l'étiquette 6 histidines. Le mutant Abf2-domaine A-K18Q a été produit et purifié mais son affinité à l'ADN diminue par rapport à Abf2-domaine A sauvage, ce qui suggère un changement de conformation rendant inexploitable les résultats du test d'activité lyase.



## 10. Perspectives

Les conditions de piégeage de l'intermédiaire Abf2-ADN sont à optimiser afin d'augmenter la spécificité envers la lysine catalytique de chaque domaine. Plusieurs paramètres pourraient être testés tels que le type de site abasique (dRP/AP), la force ionique (NaCl) qui en étant augmentée pourrait réduire les duplex Abf2-ADN non spécifiques ou encore la température et le temps de la réaction. Un autre agent réducteur moins puissant pourrait également être testé. La longueur de l'ADN ainsi que la position du site abasique serait une voie intéressante à explorer. Si, malgré ces tests, plusieurs sites étaient encore modifiés, la quantification de tous les sites modifiés permettrait d'identifier le site modifié le plus abondant pour chaque domaine d'Abf2. Si les deux sites actifs sont les sites majoritairement piégés par NaBH<sub>4</sub>, ils pourraient ainsi être mis en évidence. A l'inverse, si l'on fait l'hypothèse que le site catalytique à un  $k_{obs}$  plus élevé que celui des lysines non compétentes pour l'activité lyase, l'intermédiaire réactionnel spécifique sera peut-être plus difficile à piéger par NaBH<sub>4</sub> que les non-spécifiques. Dans ces conditions, les sites actifs seraient les sites minoritaires et seraient donc très difficile à localiser avec cette approche de réduction de la base de Schiff. Une autre stratégie de capture de l'intermédiaire réactionnel, telle que l'utilisation d'un substrat suicide, pourrait alors être envisagée.

Malgré les conditions de piégeage qui restent à optimiser, nous avons développé une méthode efficace pour la purification d'un complexe covalent protéine-ADN ainsi que pour la détection d'hétéroconjugués peptide-acides nucléiques et la localisation des résidus modifiés par spectrométrie de masse. Ces outils pourraient être utilisés d'une manière plus générale pour disséquer le mécanisme d'autres enzymes de réparation de l'ADN par excision.



## Chapitre 4 Caractérisation de l'interaction entre la locostatine et la protéine PEBP1/RKIP humaine, un suppresseur de métastases naturel

#### Résumé du chapitre 4 :

A ce jour, très peu de cibles cellulaires ayant une activité suppresseur de métastases ont été découvertes et les molécules thérapeutiques visant cette nouvelle classe de cible sont plus rares encore. Plusieurs études récentes ont montré que la protéine humaine Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1), est un suppresseur de métastases dans différents types de cancer. La locostatine est le seul composé connu ayant une activité démontrée envers PEBP1. Au niveau cellulaire, elle empêche la migration des cellules en se liant à PEBP1.

Mes travaux de thèse portent sur la caractérisation fine du complexes hPEBP1-locostatine par spectrométrie de masse. Cette caractérisation aidera à la conception et la synthèse de ligands de hPEBP1 à visée anti-métastases

Le mécanisme de liaison a été disséqué et montre que la locostatine se fixe de manière covalente à la protéine puis s'hydrolyse en groupement butyrate. Des expériences de cinétiques suggèrent l'existence d'un intermédiaire réactionnel non covalent et réversible ( $P + L \rightleftharpoons PL^* \rightarrow PL$ ). Afin d'identifier le résidu ciblé par la locostatine, les conditions de réaction et de protéolyse en solution ont été optimisées. L'approche qualitative révèle la présence de plusieurs sites de réaction de la locostatine sur la protéine, indiquant de probables réactions non spécifiques. Ces résultats ont conduit au développement 1) d'un modèle mathématique pour déterminer la fraction de liaison optimale permettant de discriminer le site spécifique des sites non-spécifiques, et 2) d'une méthode pour la quantification parallèle et exhaustive du degré de modification de tous les sites modifiés de la protéine. Selon nos critères, les résultats de quantification de ces sites/séquence modifiés ne permettent pas de proposer un site majoritaire. Ce résultat indique que la vitesse de réaction spécifique n'est pas visiblement plus rapide que celle des réactions non spécifiques, ce qui complique fortement l'identification du site spécifique de la locostatine sur hPEBP1.

Au-delà de l'application à hPEBP1, la quantification de sites modifiés est un outil précieux pour corrélérer la liaison moléculaire avec l'activité cellulaire d'agents thérapeutiques covalents potentiels.

# 1. Contexte

## 1.1 La protéine PEBP1/RKIP

La protéine Phosphatidylethanolamine Binding Protein 1 (PEBP1), également connue chez les mammifères sous le nom de Raf Kinase Inhibitory Protein (RKIP), fait partie de la famille des protéines PEBPs qui est un groupe très conservée au cours de l'évolution et qui compte plus de 400 membres. PEBP1 est une protéine ubiquitaire, retrouvée dans de nombreux tissus (cerveau, foie, estomac, rate, muscle...) et de nombreux organismes (mammifères, insectes, plantes, levures, bactéries). Elle fût initialement identifiée en 1984 dans le cerveau de boeuf (Bernier and Jollès, 1984) puis caractérisée comme une protéine liant la phosphatidyléthanolamine (PEBP) (Bernier *et al.*, 1986). En 1999, à l'issue d'un criblage de molécules par double hybride chez le rat (Yeung *et al.*, 1999), rPEBP1 fût nommée RKIP pour refléter sa capacité à se lier et à inhiber la kinase Raf. Ce nom a été ensuite étendu à la PEBP1 humaine (hPEBP1). La localisation principale de PEBP1 est cytoplasmique, associée à la membrane plasmique ou à celle du réticulum endoplasmique granuleux (Roussel *et al.*, 1988), mais elle est également retrouvée dans le milieu extracellulaire. PEBP1 est fortement exprimée dans le système nerveux, les glandes surrénales et la thyroïde ; modérément exprimée dans le foie, les testicules et la prostate, et est faiblement exprimée dans les globules blancs, le cœur et d'autres tissus (Al-Mulla *et al.*, 2013).

## 1.2 Structure de PEBP1 humaine

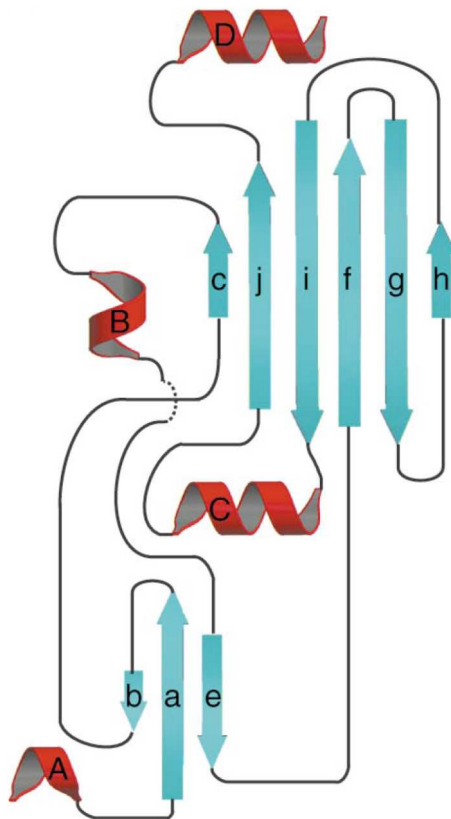
La structure primaire de PEBP1 humaine est constituée de 187 acides aminés en comptant la méthionine initiale. Cette dernière est clivée lors de la maturation de la protéine. Les masses moléculaires monoisotopique et moyenne de hPEBP1 sont respectivement de 20 912.6255 Da et 20 925.3744 Da. La séquence consensus qui définit l'appartenance à la famille des PEBPs se trouve entre les résidus 70 et 119 (Banfield *et al.*, 1998) et sont représentés en rouge dans la Figure 72. La protéine contient deux cystéines en position 133 et 168.

**Figure 72. Séquence de hPEBP1.** La numérotation des résidus correspond à la protéine avec la méthionine initiale. Les résidus en rouge représentent la séquence consensus de la famille PEBP. D'après Banfield *et al.* (1998)

```
 2 PVDLSKWSGP LSLQEVDEQP QHPLHVITYAG AAVDELGKVL TPTQVKNRPT
52 SISWDGLDSG KLYTLVLTDP DAPSRKDPKY REWHHFLVVN MKGNDISSGT
102 VLSDYVGS GP PKGTGLHRYV WLVEYQDRPL KCDEPILSNR SGDRHGKFKV
152 ASFRKKEYELR APVAGTCYQA EWDDYVPKLY EQLSGK
```

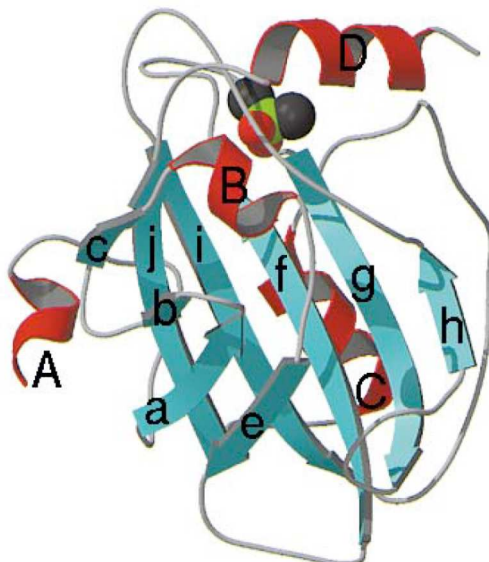
La structure cristallographique de la PEBP1 bovine a été réalisée ici au CBM en 1998 (Serre *et al.*, 1998). Simultanément et dans le même journal, la structure cristallographique de la PEBP1 humaine était également publiée (Banfield *et al.*, 1998). Les deux structures sont très similaires avec deux 2 feuillets  $\beta$  antiparallèles et un élément C-terminal  $\alpha\beta\alpha$  (Figure 73).

**Figure 73. Structure secondaire de hPEBP1.** La protéine est constituée de 9 feuillets  $\beta$  et 4 hélices  $\alpha$  (Banfield *et al.*, 1998).



En outre, ces structures montrent une protéine globulaire avec un cœur en feuillets  $\beta$  et une poche proéminente exposée au solvant, comme illustré avec la structure de hPEBP1 Figure 74. Cette poche est un élément structural très conservé au sein des PEBPs. Elle est composée de seize résidus (D70, A73, P74, Y81, W84, H86, V107, G108, G110, P111, P112, H118, Y120, L180, Y181, et L184) (Tavel *et al.*, 2012) et peut accueillir une variété de ligands anioniques tels que des phospholipides (Vallée *et al.*, 2001) ou encore le GTP, le FMN (Tavel *et al.*, 2012) ou encore. Cette poche est couramment appelée poche à anions. La Figure 74 indique par ailleurs que la protéine a été co-cristallisée avec une molécule de cacodylate provenant du tampon de cristallisation. Ce cacodylate est logé dans la poche à anions. Cette structure montre également que les deux cystéines sont sous forme réduite.

Figure 74: Structure tertiaire de la hPEBP1 avec un cacodylate dans la poche à anions (Banfield *et al.*, 1998)



### 1.3 Les fonctions de PEBP1

La PEBP1 agit dans un grand nombre de processus physiologiques tels que la croissance, la différenciation, la prolifération, la migration, la motilité, le cycle cellulaire, la stabilité génomique, l'apoptose et la résistance aux médicaments (Al-Mulla *et al.*, 2013). Son rôle exact dans la cellule est encore mal défini. Elle est impliquée dans différentes maladies comme la maladie d'Alzheimer, le diabète, les cancers de la prostate, de l'estomac, du sein, de l'ovaire, du poumon, le mélanome (cancer de la peau), le cancer colorectal, les tumeurs stromales gastro-intestinales, le carcinome hépatocellulaire, le carcinome du nasopharynx, et les gliomes (tumeurs cérébrales) (Al-Mulla *et al.*, 2013). Depuis 2003, elle a été surtout identifiée comme le produit d'un des rares gènes suppresseurs de métastases, dans un nombre croissant de cancers majeurs.

#### 1.3.1 PEBP1 et voies de signalisation

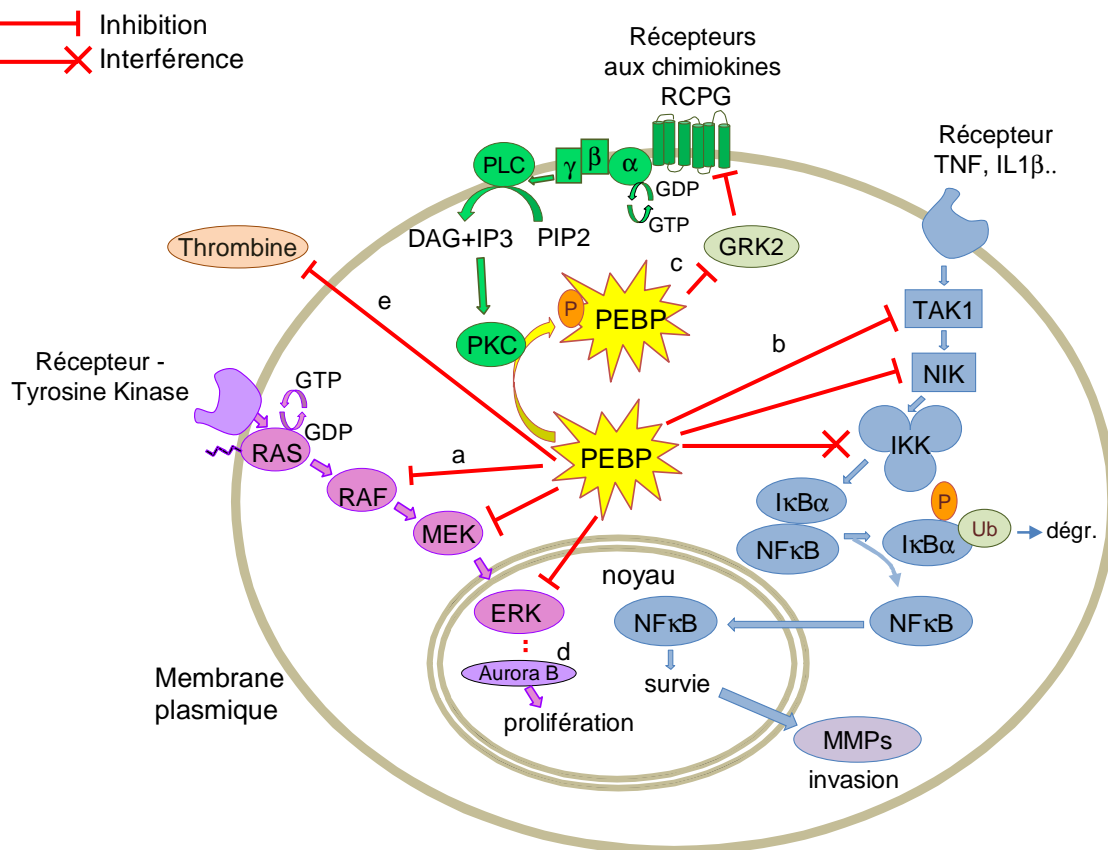
Les voies de signalisation ont un rôle vital pour la cellule. Elles assurent la transmission d'un signal extracellulaire, capté par un récepteur transmembranaire, vers différents effecteurs intracellulaires conduisant à la génération d'une réponse cellulaire appropriée. Cette réponse peut se traduire par l'activation ou l'inhibition de protéines et/ou à l'induction ou la répression de l'expression de gènes. Depuis plusieurs années, PEBP1 apparaît de plus en plus comme une protéine pivot, au carrefour des principales voies de signalisation impliquées dans la survie de la cellule telles que la voie MAP kinases Ras/Raf/MEK/ERK, la voie NF- $\kappa$ B et la voie des récepteurs couplés aux protéines G. Son homologue chez la levure a été caractérisé au CBM comme un pont possible entre signalisation cellulaire et métabolisme (Beaufour *et al.*, 2012).

### 1.3.1.1 La voie Ras/Raf/MEK/ERK

La voie Ras/Raf/MEK/ERK est essentielle pour la croissance cellulaire. Cela comprend la prolifération, la transformation, la différenciation et l'apoptose (Kolch, 2000). Une grande variété d'hormones, de facteurs de croissance et de facteurs de différenciation, ainsi que des substances favorisant les tumeurs, emploient cette voie. Ces molécules, lorsqu'elles se lient aux récepteurs tyrosine kinase TRK, activent la protéine GTPase Ras qui déclenche alors la cascade de phosphorylation de la voie de signalisation MAPK : phosphorylation de Raf par Ras, puis de MEK par Raf, puis de ERK par MEK (Figure 75 en rose). Raf, MEK et ERK sont des sérine/thréonine kinases, et sont de type MAPKKK, MAPKK et MAPK, respectivement. Une fois activée dans le noyau, ERK phosphoryle ensuite les facteurs de transcription nucléaires qui influencent de nombreuses fonctions cellulaires (Zeng *et al.*, 2008). En 1999, Yeung *et al.* ont démontré que rPEBP1 est capable d'inhiber la voie Ras/Raf/MEK/ERK en se liant à Raf-1. Cette interaction PEBP1-Raf-1 empêche l'activation de Raf-1 en bloquant sa phosphorylation en S338 et Y340/341 (Trakul *et al.*, 2005). Il a également été constaté que rPEBP1 peut dissocier le complexe Raf-1-MEK, empêchant ainsi la phosphorylation et l'activation de MEK par Raf-1 (Yeung *et al.*, 2000). La voie Ras/Raf/MEK/ERK est régulée par la phosphorylation de rPEBP1 en Ser153 (Corbit *et al.*, 2003). Ainsi la forme non phosphorylée de PEBP1 est liée à Raf1 et inhibe la voie alors que la forme phosphorylée de PEBP1 est dissociée de Raf1 et lève l'inhibition de la voie.



**Figure 75. Les activités cellulaires de PEBP1.** En rose, inhibition de la voie Ras/Raf/MEK/ERK. En bleu, inhibition de la voie NF-κB. En vert, inhibition de la voie des récepteurs couplés aux protéines G. a : Yeung et al., 1999, b : Yeung et al., 2001, c : Lorenz et al., 2003, d : Eves et al., 2006, e : Hengst et al., 2001.



### 1.3.1.2 La voie Nf-κB

PEBP1 contrôle également l'activation du facteur nucléaire κB (NF-κB). Dans sa forme inactive, NF-κB est localisée dans le cytoplasme, où il est lié à sa protéine inhibitrice IκBα (Inhibitory κBα) (Dyson and Komives, 2012). En réponse à différents stimuli, tels que des infections microbiennes et virales ou encore des cytokines inflammatoires, les kinases TAK1 (transforming growth factor beta-activated kinase 1), NIK (Nf-κB inducing kinase) et IKK (IκB kinase) sont activées résultant à la phosphorylation d'IκBα liée à Nf-κB (Figure 75 en bleu). Il s'ensuit une ubiquitination et une dégradation d'IκBα, puis une libération de Nf-κB. Une fois libéré, Nf-κB transloque dans le noyau où il régule la transcription de gènes impliqués dans l'immunité, l'inflammation, la prolifération cellulaire, l'apoptose, et la migration cellulaire (Karin *et al.*, 2002). rPEBP1 est capable de bloquer la voie NF-κB en inhibant les kinases en amont TAK1, NIK, et IKK (Yeung *et al.*, 2001). Cette inhibition entraîne l'inhibition en aval des produits des gènes anti-apoptotiques et sensibilise la cellule à l'apoptose induite par la chimiothérapie (Chatterjee *et al.*, 2004).

### 1.3.1.3 La voie des récepteurs couplés aux protéines G

PEBP1 est également impliquée dans le contrôle de la cascade de signalisation des récepteurs couplés aux protéines G (RCPG, GPCR en anglais) (Lorenz *et al.*, 2003). Les RCPG constituent une grande famille de récepteurs membranaires et sont des déterminants physiologiques essentiels de la neurotransmission, de l'inflammation, et de la régulation de la pression artérielle (Ribas *et al.*, 2007). Lorsqu'elle est inactive, la protéine G est constituée de trois sous-unités  $G\alpha$ ,  $G\beta$  et  $G\gamma$  et la sous-unité  $G\alpha$  lie une molécule de GDP (Figure 75 en vert). Lorsqu'un signal chimique ou physique stimule le récepteur RCPG, ce dernier devient actif et change de conformation. Ce changement conformationnel favorise l'interaction entre RCPG et la protéine G. Il en résulte une diminution de l'affinité de la sous-unité  $G\alpha$  pour le GDP au profit du GTP. Cet échange GDP/GTP conduit à une dissociation du complexe hétérotrimérique :  $G\alpha$ -GTP et le complexe  $G\beta\gamma$  (Neer, 1995). Les sous-unités dissociées activent alors différents effecteurs spécifiques : adénylate cyclase (AC), phospholipase C (PLC)... (Neves *et al.*, 2002). La PLC, une fois activée par la sous-unité  $G\alpha$ , clive le phosphatidyl-inositol-4,5-diphosphate (PIP2) en inositol-triphosphate (IP3) et en diacylglycérol (DAG). Ces deux second-messagers agissent à leur tour comme des effecteurs, le DAG active notamment la protéine kinase C (PKC).

La voie RCPG est retro-contrôlée négativement par la kinase GRK2. Cette dernière phosphoryle les RCPGs actifs et initie leur internalisation ce qui conduit au final à une inhibition de la voie RCPG (Krupnick and Benovic, 1998). hPEBP1 a été identifiée comme un inhibiteur physiologique de GRK-2 (Lorenz *et al.*, 2003), ce qui veut dire qu'elle lève l'inhibition par GRK2. Cependant, sa fonction inhibitrice est dépendante d'une phosphorylation. Quand une cellule est stimulée par un facteur de croissance, la kinase PKC est activée et phosphoryle rPEBP1 au niveau de Ser153 (Corbit *et al.*, 2003), ce qui provoque la dissociation de rPEBP1 et de Raf-1. La protéine phosphorylée rP-PEBP1 se lie alors à GRK-2 et bloque son activité inhibitrice. Récemment, Deiss et ses collaborateurs ont montré que cette phosphorylation en S153 déclenche la dimérisation de hP-PEBP1 et que ce processus est nécessaire pour que hPEBP1 permute de Raf-1 à GRK2 (Deiss *et al.*, 2012).

La forme phosphorylée hP-Ser153PEBP1, en levant son contrôle négatif sur la kinase Raf et en inhibant GRK2, semble favoriser la survie des cellules cancéreuses et l'invasion (Escara-Wilke *et al.*, 2012). Par exemple, l'expression de P-Ser153PEBP1 est un indicateur prédictif de la survie du cancer du poumon (Huerta-Yeppez *et al.*, 2011).

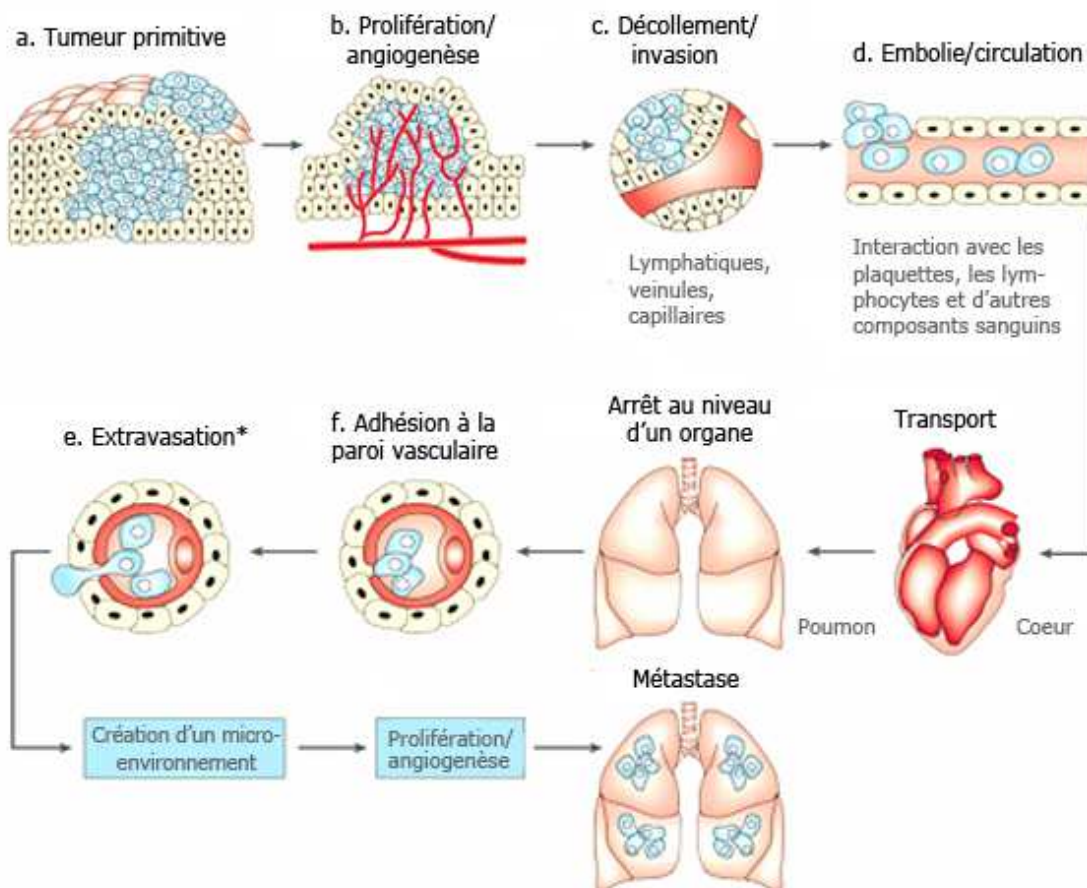
## 1.3.2 PEBP et cancer

### 1.3.2.1 Cancer et métastase

Le cancer est une maladie causée par une croissance et une prolifération cellulaire incontrôlées formant ce que l'on appelle une tumeur. Lorsqu'elle est localisée, la tumeur est dite primaire et peut en général être traitée efficacement par tout un arsenal thérapeutique anti-cancer constitué de la chirurgie, la chimiothérapie, la radiothérapie, la thérapie hormonale, l'immunothérapie, et des combinaisons de celles-ci (Bonavida *et al.*, 2013). Actuellement, le pronostic vital du patient est surtout engagé lors du développement de métastases.

Une métastase correspond à la formation d'une tumeur secondaire au niveau d'un site éloigné de celui de la tumeur primaire. Ce processus est complexe et comprend plusieurs étapes. Tout d'abord, les cellules tumorales acquièrent un phénotype métastatique, c'est-à-dire qu'elles vont perdre leurs caractéristiques épithéliales pour acquérir des caractéristiques mésenchymateuses et des propriétés invasives, un processus connu sous le nom de transition épithélio-mésenchymateuse (EMT pour Epithelial to Mesenchymal Transition) (Baritaki *et al.*, 2009). Ceci a pour effet de rompre les contacts intercellulaires et de favoriser la motilité cellulaire (Figure 76).

Figure 76. Principales étapes de la formation des métastases [e].



S'ensuit une étape d'angiogenèse pour vasculariser et alimenter les cellules au cœur de la tumeur, d'invasion du tissu environnant, d'intravasation (passage des cellules du tissu vers les vaisseaux sanguins), de survie dans la circulation sanguine, d'extravasation (passage des cellules depuis les vaisseaux sanguins vers un tissu environnant) et une colonisation d'un nouveau site, distant de la tumeur primaire (Escara-Wilke *et al.*, 2012).

### 1.3.2.2 Activité suppresseur de métastases de hPEBP1

Les premiers résultats suggérant l'activité suppresseur de métastases de hPEBP1 ont été décrits en 2003 par Fu *et al.* Ces auteurs ont montré que l'expression de hPEBP1 est plus faible dans les cellules cancéreuses métastatiques de la prostate que dans les cellules cancéreuses non métastatiques et que cette expression est inversement corrélée avec le potentiel invasif des cellules du cancer de la prostate *in vitro*. D'autres études ont également montré cette corrélation inverse entre le niveau d'expression de hPEBP1 et le potentiel métastatique de plusieurs cancers chez l'Homme tels que le mélanome (Schuierer *et al.*, 2004), le cancer du sein (Hagan *et al.*, 2005), le cancer colorectal (Al-Mulla *et al.*, 2006), le cancer de l'ovaire (Li *et al.*, 2008) et le cancer du nasopharynx (Chen *et al.*, 2008). De plus, la restauration de l'expression de hPEBP1 dans des cellules du cancer de la prostate réduit l'angiogenèse tumorale et les métastases dans un modèle murin (Fu *et al.*, 2003). De la même manière, la surexpression de hPEBP1 se traduit par une diminution des métastases de cellules cancéreuses du sein (Dangi-Garimella *et al.*, 2009) et de l'ovaire (Li *et al.*, 2008). En outre, il a été montré que le niveau sérique de hPEBP1 a une valeur pronostique chez les patients atteints du cancer de la prostate, et un dosage sérique de hPEBP1 a été proposé comme marqueur pronostique en soutien aux décisions thérapeutiques (Fu *et al.*, 2006). En plus du niveau d'expression totale de hPEBP1, il a été suggéré que la quantité absolue de forme phosphorylée de hPEBP1 est un indicateur pronostique pour le cancer du poumon (Huerta-Yepez *et al.*, 2011). Considérant l'ensemble de ces données, on peut conclure que hPEBP1 a un rôle central sur la croissance tumorale et les métastases et est considérée comme une nouvelle cible thérapeutique et diagnostique contre le cancer.

Néanmoins, les mécanismes précis qui favorisent les métastases suite à la perte ou à la réduction de hPEBP1 restent à élucider. Est-ce la capacité de hPEBP1 à moduler la migration cellulaire, l'invasion, la croissance ou encore la stabilité génomique qui est responsable de ses fonctions suppresseur de métastases? La réponse à cette question est difficile du fait des fonctions multimodales de cette protéine et de la complexité du processus métastatique.

L'influence du niveau d'expression de PEBP1 humaine, canine ou de rat sur la migration cellulaire a été explorée mais les résultats sont controversés. Une corrélation négative entre le niveau

d'expression de hPEBP1 et la migration a été observée dans les cellules de carcinome hépatocellulaire humain (HHC) (Lee *et al.*, 2006). A l'inverse, une corrélation positive a été mise en évidence avec hPEBP1 canine dans les cellules épithéliales de chien (MDCK) (Zhu *et al.*, 2005). De même, l'expression de rPEBP1 est positivement corrélée avec la migration des cellules étoilées hépatiques de rat (HSC) (Ma *et al.*, 2009) et des cellules du cancer du colon humain (LOVO) (Beshir *et al.*, 2010). Enfin, en utilisant d'autres lignées cellulaires du cancer du sein (T47D, BT-20), du colon (HCT-116) et de la peau (MDA-MB-435), Beshir *et al.* (2010) n'ont observé aucune influence du niveau d'expression de rPEBP1 sur la migration. Les contradictions entre l'expression de PEBP1 et la migration des cellules peuvent être attribuées à l'utilisation de différentes lignées cellulaires et d'espèces de PEBP1 différentes. Ce constat suggère que le rôle de PEBP1 sur la migration est dépendant du type cellulaire.

Concernant l'invasion, Beshir *et al.* (2010) proposent que la fonction de rPEBP1 dans les métastases soit attribuée à sa capacité à réguler à la baisse l'expression de certaines métalloprotéinases de la matrice extracellulaire (MMP), à savoir la MMP-1 et MMP-2. Cette répression des MMPs se ferait par la voie Nf-kB (Beshir *et al.*, 2010). Cette hypothèse n'a pas été testée au niveau expérimental par ces auteurs.

Par ailleurs, la PEBP1 canine a également une influence sur l'adhérence cellule-substrat et cellule-cellule dans les cellules MDCK (Zhu *et al.*, 2005; Mc Henry *et al.*, 2008). D'autres procédés pouvant affecter la formation de métastases ont été décrits comme étant régulés par PEBP1 humaine ou de rat ; à savoir, la progression du cycle cellulaire (Eves *et al.*, 2006), l'angiogenèse (Fu *et al.*, 2003) et l'apoptose (Chatterjee *et al.*, 2004).

### 1.3.2.3 Activité pro-apoptotique de hPEBP1

Les différentes thérapies anti-cancer telles que la chimiothérapie, la radiothérapie, la thérapie hormonale ou l'immunothérapie ont un seul et même objectif : induire la mort des cellules cancéreuse. Cette mort cellulaire est le plus souvent induite par la voie de l'apoptose, également appelée mort cellulaire programmée. Cependant, il peut subsister une sous-population de cellules qui est très résistante aux diverses thérapies cytotoxiques. En outre, ces cellules résistantes sont généralement la cause d'une récurrence du cancer, très souvent dévastatrice pour le patient car les traitements actuellement disponibles sont inefficaces (Bonavida 2013). Les cellules résistantes sont en fait équipées de mécanismes moléculaires qui leur permettent d'échapper à l'activation des voies apoptotiques induite par le médicament. La résistance à l'apoptose de ces cellules tumorales résulte,

dans une large mesure, de l'hyperactivation constitutive de plusieurs voies de survie / anti-apoptotiques telles que les voies NF- $\kappa$ B, Ras/Raf/MEK/ERK et PI3K-Akt (Bonavida 2013).

Le rôle déterminant de hPEBP1 dans l'induction de l'apoptose a été mis en évidence par Chatterjee *et al.* (2004) en montrant que l'expression de hPEBP1 sensibilise les cellules du cancer du sein et de la prostate aux agents chimiothérapeutiques. De même, l'expression de hPEBP1 est induite en réponse au Rituximab, médicament utilisé en chimiothérapie, qui inhibe la voie Raf-MEK-ERK et augmente l'apoptose dans les cellules de lymphome non hodgkinien B (Jazirehi *et al.*, 2004). A l'inverse, la perte de hPEBP1 diminue l'apoptose induite par radiothérapie et chimiothérapie, protégeant ainsi les cellules cancéreuses contre la mort.

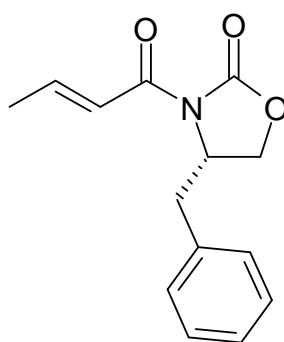
Dans la littérature, deux mécanismes de résistance chimiothérapeutique des cellules cancéreuses sont décrits :

- 1) Le circuit NF- $\kappa$ B / Snail / hPEBP1 proposé par Baritaki *et al.* (2009). Snail est un répresseur transcriptionnel de hPEBP1 (Beach *et al.*, 2008) qui joue un rôle important dans l'invasion et les propriétés migratoires des cellules cancéreuses au cours de la progression tumorale en étant un médiateur de la transition épithélio-mésenchymateuse (Nieto, 2002). La transcription de Snail est régulée, en partie, par NF- $\kappa$ B (Julien *et al.*, 2007) et par lui-même en réprimant son propre promoteur (Peiró *et al.*, 2006). Il a été montré que la surexpression de PEBP1 dans les cellules tumorales humaines de la prostate et de mélanome était capable d'inverser la résistance aux médicaments et d'induire l'apoptose médiée par TRAIL (Baritaki *et al.*, 2007). En outre la sous expression de hPEBP1 confère une résistance aux médicaments anti-cancéreux dans les cellules cancéreuses intrinsèquement sensibles (Chatterjee *et al.*, 2004).
- 2) La voie KEAP1 / NRF2 proposée par Al-Mulla *et al.* (2012). Ces auteurs ont constaté que la déplétion de hPEBP1 induisait un stress oxydatif intense dans les cellules HEK293 (Al-Mulla *et al.*, 2011). Il est bien établi que le stress oxydatif induit l'activation de Nrf2 (Nuclear factor (erythroid-derived 2)-like 2), une molécule qui régule à la hausse les antioxydants intracellulaires et qui augmente la survie cellulaire (Lee *et al.*, 2003). Dans les cellules normales et en présence de hPEBP1, NRF2 est séquestrée dans le cytoplasme *via* son interaction avec KEAP1 (Kelch-like ECH-associated protein 1). En cas de perte ou de réduction de hPEBP1, KEAP1 est ubiquitinylée ce qui provoque la libération de Nrf2 et sa translocation au noyau afin d'activer la transcription de gènes antioxydants.

## 1.4 La locostatine

Comme évoqué précédemment, PEBP1 régule des processus physiologiques essentiels. Ainsi, l'identification de molécules qui se lient à PEBP1 et influencent son aptitude à inhiber des protéines cibles telles que Raf-1, représentent un grand intérêt thérapeutique pour le futur. En 2002, une petite molécule, la locostatine, a été identifiée dans un crible de molécules pro-cicatrisantes comme étant capable d'inhiber la migration de cellules épithéliales de rein de chien Madin-Darby (MDCK) (Mc Henry *et al.*, 2002), puis en 2005, par une approche dite de chimie génomique, il a été montré par Zhu *et al.* qu'elle est capable de se lier à rPEBP1 (Zhu *et al.*, 2005). La structure de la locostatine est présentée Figure 77.

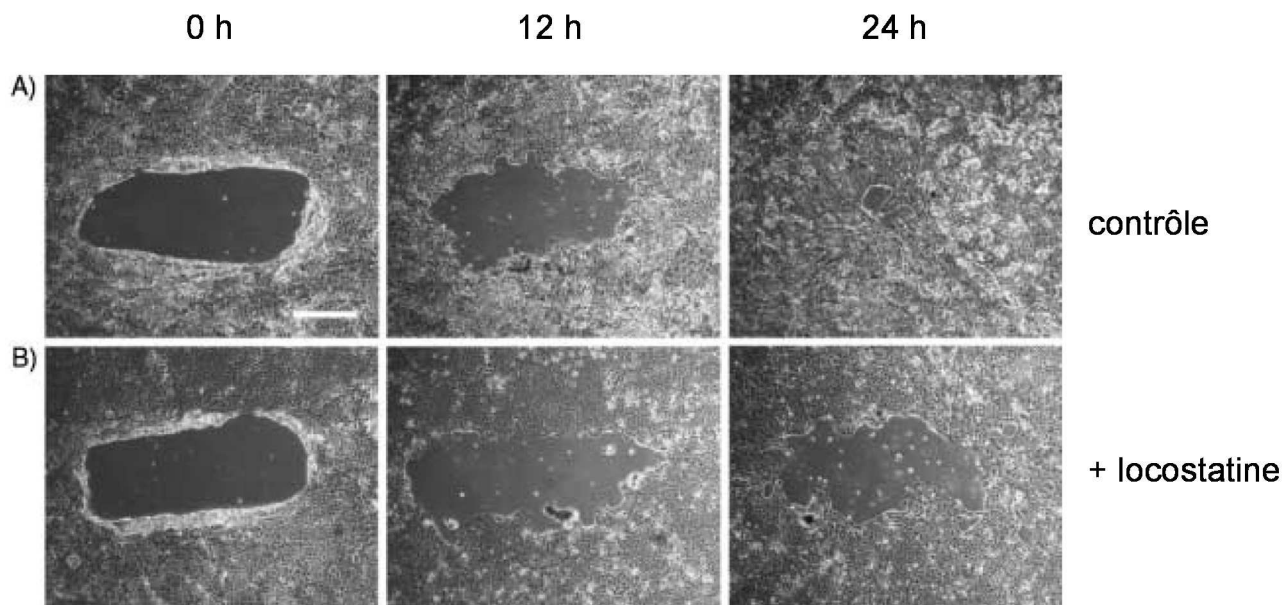
Figure 77. Structure chimique de la locostatine



### 1.4.1 Activité anti-migratoire

Initialement, la locostatine, également appelée (4S)-3-[(E)-But-2-enoyl]-4-benzyl-2-oxazolidinone, a été découverte en 2002 lors d'un criblage d'agents cicatrisants sur des cellules MDCK (Mc Henry *et al.*, 2002). La (4S)-3-[(E)-But-2-enoyl]-4-benzyl-2-oxazolidinone a montré des propriétés inattendues d'inhibition de la cicatrisation, plus précisément de la migration cellulaire, ce qui lui a valu son nom de locostatine (Figure 78).

**Figure 78: Activité antimigratoire de la locostatine sur des cellules MDCK** (Mc Henry *et al.*, 2002). Images de microscopie à contraste de phase de cellules monocouches MDCK lors d'un test de « wound healing » qui consiste à mesurer la capacité des cellules à combler une lésion introduite expérimentalement : A) Contrôle avec 0.1% de DMSO 0, 12 et 24 h après la blessure B) 50  $\mu$ M de locostatine, 0, 12 et 24 h après la blessure



Zhu *et al.* (2005) ont montré par ailleurs que la déplétion, par interférence ARN, de PEBP1 canine inhibe également la migration des cellules épithéliales MDCK et que la surexpression de PEBP1 canine dans ces cellules annule l'effet anti-migratoire de la locostatine. En accord avec un rôle dans la voie de signalisation de Raf, la locostatine empêche la co-immunoprécipitation de rPEBP1 avec Raf-1 *in vitro*, démontrant ainsi que la locostatine lève l'inhibition de rPEBP1 sur Raf-1 (Zhu *et al.*, 2005). Plus récemment, il a été observé que la locostatine altère également la migration de cellules cancéreuses humaines d'adénocarcinome du sein (MCF7) (Mc Henry *et al.*, 2008). Ces résultats suggèrent que locostatine régule la migration cellulaire *via* un mécanisme dépendant de PEBP1 et médié par la voie Raf/MEK/ERK. Toutefois d'autres études montrent que la locostatine est capable d'inhiber la migration de cellules qui n'expriment plus PEBP1. C'est le cas des cellules non-tumorales de souris (MEF) (Shemon *et al.*, 2009) et de la bactérie *Mycobacterium smegmatis* (Eulenburg *et al.*, 2013). Ces données suggèrent que la locostatine pourrait avoir d'autres cibles cellulaires encore non identifiées. Zhu *et al.* (2005) ont montré sur gel SDS-PAGE, en employant de la locostatine tritiée, qu'en plus de rPEBP1, la locostatine se lie à trois enzymes : la glutathione S-transférase omega 1-1 (GSTO1-1), l'aldéhyde déshydrogénase 1A1 (ALDH1A1) et la prolyle oligopeptidase (POP). L'inhibition de chacune de ces enzymes par un inhibiteur spécifique n'a cependant pas révélé d'effet pro-migratoire ni anti-migratoire, ce qui les exclut donc comme cibles médiatrices d'un effet anti-migratoire.

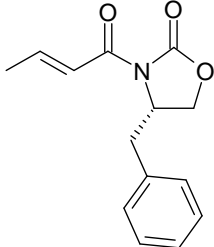
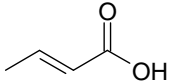
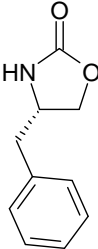


## 1.4.2 Interaction avec PEBP1

Depuis la découverte que PEBP1 est une cible cellulaire de la locostatine (Zhu *et al.*, 2005), plusieurs équipes se sont intéressées à l'interaction entre ces deux partenaires afin de mieux comprendre le rôle de PEBP1 dans les propriétés anti-migratoires de la locostatine, ainsi que d'ébaucher une stratégie pour exploiter les propriétés anti-métastases naturelles de PEBP1.

Le mécanisme réactionnel supposé est une addition de Michael entre un résidu nucléophile de PEBP1 et l'acyl insaturé de la locostatine qui constitue un centre électrophile (Tableau 12).

**Tableau 12. Locostatine et molécules dérivées. NC : liaison non covalente**

Nom	Structure	Formule brute	Masse moléculaire (Da)	Formule chimique ajoutée sur protéine	Incrément de masse ajouté	
					Masse théorique monoisotopique (Da)	Masse théorique moyenne (Da)
Locostatine		C <sub>14</sub> H <sub>15</sub> NO <sub>3</sub>	245.27	C <sub>14</sub> H <sub>15</sub> NO <sub>3</sub>	245.1052	245.2744
Acide crotonique		C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	86.09	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	86.0368	86.0894
(S)-4-benzyloxazolidin-2-one		C <sub>10</sub> H <sub>11</sub> NO <sub>2</sub>	177.20	C <sub>10</sub> H <sub>11</sub> NO <sub>2</sub> (NC)	177.0789 (NC)	177.2003 (NC)

Sans cette liaison insaturée, la molécule ne possède plus de propriétés anti-migratoires (Mc Henry *et al.*, 2002; Zhu *et al.*, 2005). Par ailleurs, ce groupe a pu déterminer une constante de vitesse d'association  $k_{on}$  égale à  $13 \text{ M}^{-1} \text{ s}^{-1}$  entre rPEBP1 et la locostatine.

Le site ciblé par la locostatine a fait l'objet de plusieurs recherches, que ce soit en RMN ou en spectrométrie de masse, mais sans résultats clairs à ce jour. Shemon *et al.* (2009) ont tenté d'étudier le complexe entre rPEBP1 et la locostatine par RMN mais l'analyse n'a pu être réalisée à cause d'une précipitation induite par l'ajout de locostatine à de fortes concentrations de rPEBP1. Afin de contourner la difficulté, ces auteurs ont utilisé un dérivé non réactif de la locostatine, qui ne contient pas la partie crotonyl, la (S)-4-benzyloxazolidin-2-one (Tableau 12), et ont montré que cette molécule interagit au niveau de la poche à anions (Shemon *et al.*, 2009). Tavel *et al.* (2012) ont obtenu le

même résultat avec la hPEBP. Une autre étude, utilisant la spectrométrie de masse, initialement publiée sous forme de proceedings d'un workshop autoédité, prétend avoir identifié l'His86 comme étant le site cible de la locostatine (Beshir *et al.*, 2011). Cependant, les données MS/MS sont critiquables dans la mesure où 1) les ions fragments ont tous la même intensité, ce qui suggère que le signal est au niveau du bruit de fond et 2) les ions fragments  $b_5$  et  $y_8$ , seuls ions permettant de discriminer entre une modification de His85 et His86, ne sont pas observés. De plus, de manière surprenante, ces résultats ont été obtenus avec le complexe hPEBP1-locostatine précipité, avec une destruction profonde du repliement de la protéine et en présence d'un excès de locostatine, puis resolubilisé, ce qui semble peu compatible avec une réalité biologique. Les auteurs indiquent l'existence d'au moins deux sites modifiés sur la protéine entière, mais choisissent d'accorder au résidu His86 une crédibilité particulière du fait qu'il s'agit d'un résidu conservé de la poche anionique. Il est à noter que la locostatine n'est pas un anion et qu'elle n'a pas montré d'effet de compétition avec les nucléotides GTP et FMN (Jaquillard, 2012) qui se lient avec des affinités de l'ordre de  $10^{-5}$  M à hPEBP1 (Tavel *et al.*, 2012)

## 2. Objectifs

Aujourd'hui, la plupart des approches utilisées pour prévenir les métastases sont basées sur l'inhibition des facteurs pro-métastatiques. La particularité de hPEBP1 est d'être au contraire un facteur anti-métastatique. Les travaux présentés ici s'inscrivent dans un projet collaboratif dont l'objectif est d'exploiter le potentiel anti-métastatique naturel de PEBP1 en synthétisant, de façon rationnelle, des analogues de locostatine ayant un effet agoniste (ou activateur) de la protéine pour en faire de futurs médicaments capables d'empêcher le développement des métastases cancéreuses. Dans le but de fournir des informations précieuses pour la synthèse raisonnée de ces analogues, les travaux présentés ici visent à caractériser, au niveau cinétique et structural, l'interaction entre hPEBP1 et la locostatine. Cette caractérisation comprend la dissection du mécanisme de liaison entre les deux partenaires et la localisation du site spécifique lié de manière covalente à la locostatine.

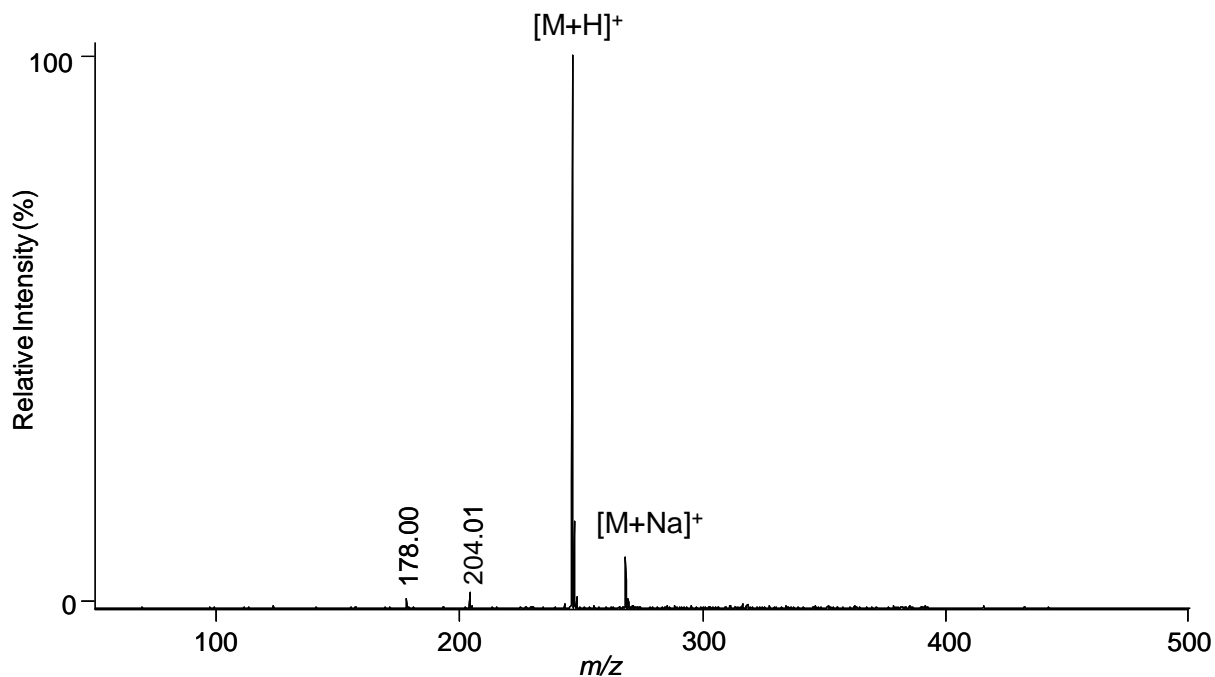
## 3. Analyse de la pureté de hPEBP1 et de la locostatine

### 3.1 Analyse de la pureté de la locostatine

La pureté de la locostatine a été évaluée par ESI-IT. La Figure 79 montre que la locostatine présente une pureté élevée. Elle est observée sous les formes  $[M+H]^+$  et  $[M+Na]^+$ . Deux composés additionnels, donnant des pics de faible intensité, sont détectés à  $m/z$  178 et 204. Le pic de  $m/z$  178

peut être annoté à la forme  $[M+H]^+$  de la (S)-4-benzyloxazolidin-2-one. Le pic à  $m/z$  204 n'est pas annoté pour le moment et pourrait être une impureté.

**Figure 79. Analyse ESI-IT de la locostatine.** Une solution de locostatine commerciale à 100 mM dans l'acétonitrile pur a été préparée. Cette solution stock a été diluée au 1/2000<sup>ème</sup> dans une solution d'ionisation acétonitrile:eau:acide formique 45:54.8:0.2 contenant du bicarbonate d'ammonium (ABC) 10 mM puis analysée en mode positif. La locostatine est observée sous les formes  $[M+H]^+$  et  $[M+Na]^+$ .



Cette pureté élevée de la locostatine est adaptée pour l'étude de son interaction avec hPEBP1.

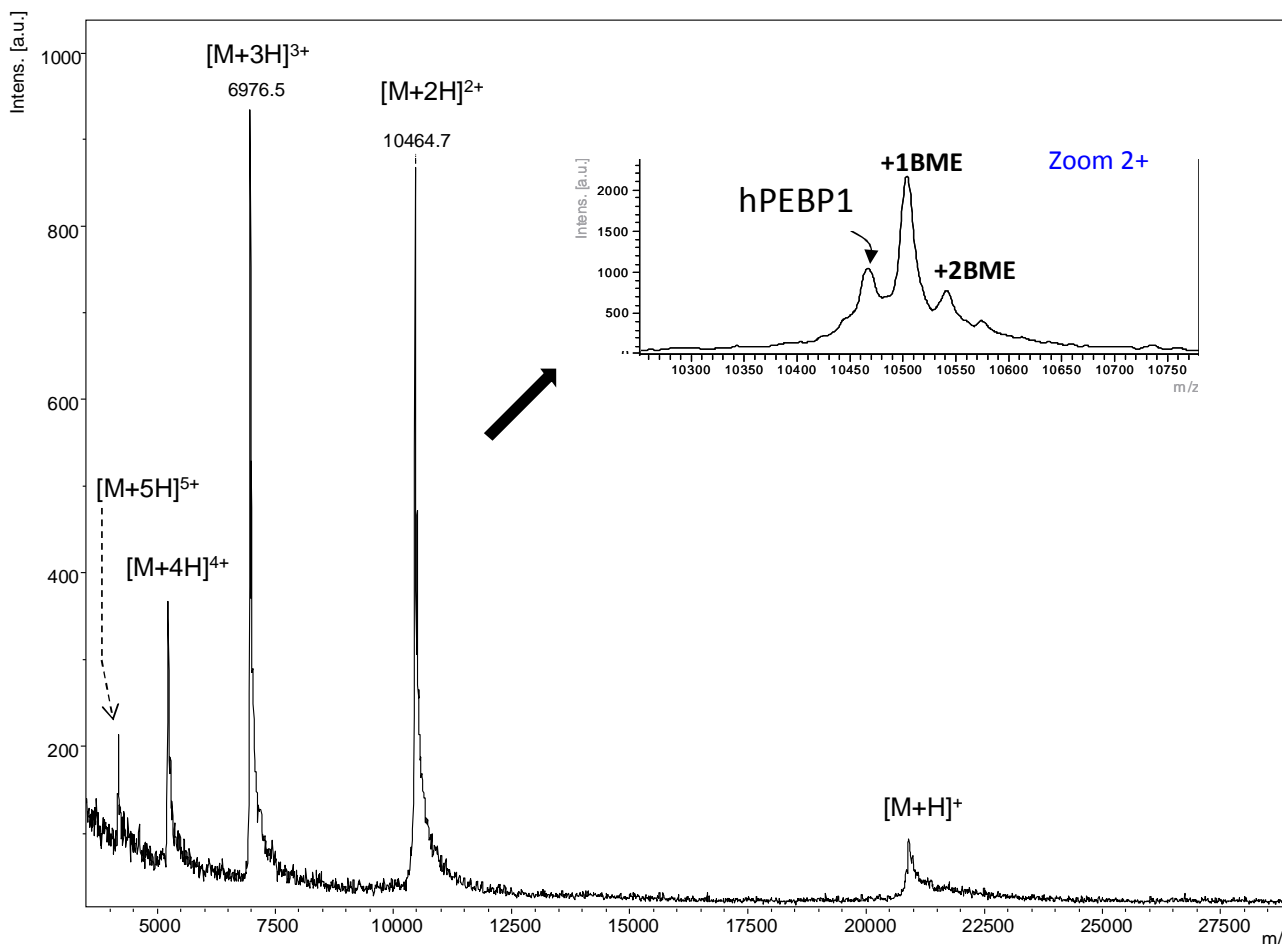
## 3.2 Analyse de la pureté de hPEBP1

Deux stocks de hPEBP1 ont été successivement utilisés pour ce projet.

### 3.2.1 Le stock initial de hPEBP1 est pur mais contient des modifications induites non désirées

Le degré de pureté de hPEBP1 a été vérifié par MALDI-TOF MS. Le spectre obtenu est présenté Figure 80.

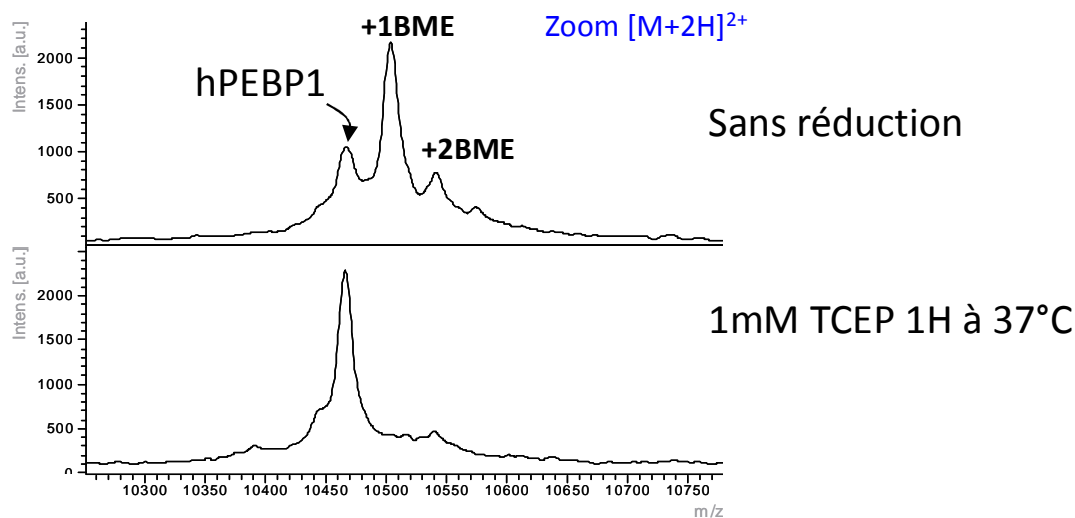
**Figure 80. Analyse MALDI-TOF du stock initial de la protéine hPEBP1.** Une solution de protéine hPEBP1 à 20  $\mu\text{M}$  dans une solution d'ABC 20mM, pH 8.5 a été diluée au  $1/20^{\text{ème}}$  dans HCCA/TWA et un aliquot a été déposé par la méthode en couche ultra-mince (Cadene and Chait, 2000; Gabant and Cadene, 2008).



La protéine est très pure et est observée à une masse moyenne de 20927.0 Da (masse calculée à partir des états de charges 2<sup>+</sup> et 3<sup>+</sup>) correspondant à hPEBP1 sans méthionine initiale (masse moyenne théorique = 20925.6 Da). Toutefois, le spectre montre également une microhétérogénéité traduite par la présence de deux pics additionnels. L'écart de masse mesuré entre les pics est de 74.5 Da ce qui peut correspondre à l'addition de  $\beta$ -mercaptoéthanol (+76 Da) sur les deux cystéines de la protéine. La présence de  $\beta$ -mercaptoéthanol dans le tampon de conservation de la protéine et l'observation

d'une stœchiométrie d'addition compatible avec le nombre de cystéines renforcent cette hypothèse. En effet, comme évoqué dans le paragraphe 2.4.2.4.a de la section 1, le  $\beta$ -mercaptoéthanol peut former un pont disulfure persistant avec les thiols de la protéine. Ces adduits de  $\beta$ -mercaptoéthanol 1. introduisent une micro-hétérogénéité nuisible à la qualité du signal, et 2. ne sont pas favorables à l'observation du complexe hPEBP1-locostatine. Ils doivent donc être éliminés. Ceci peut être réalisé par une incubation avec un autre agent réducteur tel que le DTT ou le TCEP. La Figure 81 montre la disparition des adduits de  $\beta$ -mercaptoéthanol après une réduction au TCEP 1 mM pendant une heure à 37°C. Le signal homogène ainsi obtenu est approprié pour la détection du complexe hPEBP1-locostatine. Toutefois, comme il sera discuté dans le paragraphe 4.1.1 ci-après, la locostatine est capable de réagir avec les agents réducteurs, ce qui a nécessité d'ajuster la concentration en TCEP afin qu'il ne consomme pas toute la locostatine. Cette découverte nous a conduit à faire une autre production de hPEBP1, purifiée cette fois ci dans un tampon compatible avec la MS et ne contenant pas de  $\beta$ -mercaptoéthanol.

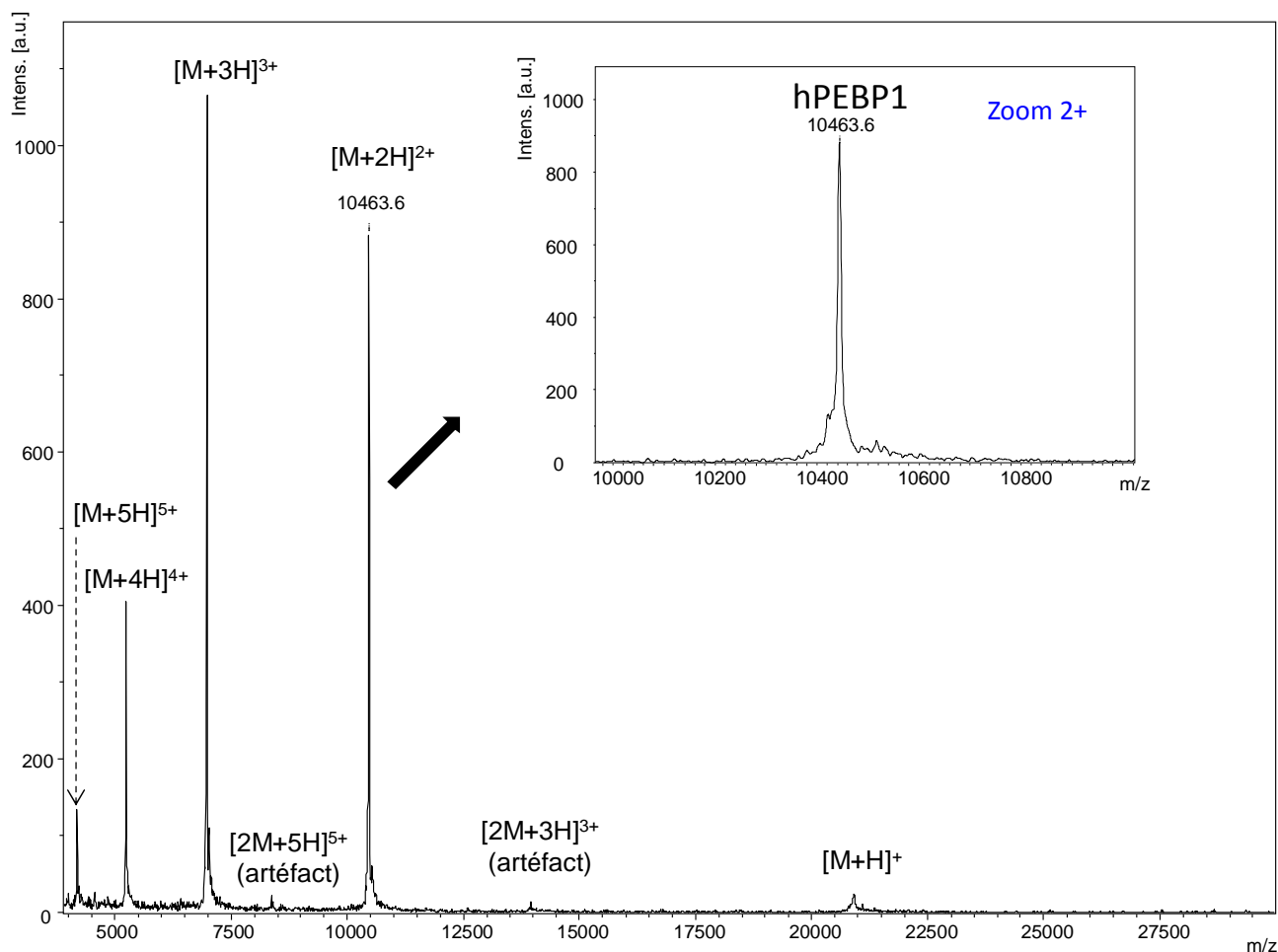
**Figure 81. Elimination des adduits de  $\beta$ -mercaptoéthanol (BME) par une réduction au TCEP.**



### 3.2.2 Le deuxième stock de hPEBP1 est pur et adapté pour l'étude du complexe avec la locostatine

Le deuxième stock de hPEBP1 a été préparé à 380  $\mu\text{M}$  dans un tampon composé d'ABC 20 mM pH 8.0 et de DTT 0.5 mM. La Figure 82 montre que la protéine est pure et qu'elle ne contient pas de micro-hétérogénéité. La masse M calculée à partir des états de charge 2+ et 3+ est de 20925.3 Da.

**Figure 82. Analyse MALDI-TOF du deuxième stock de la protéine hPEBP1.** Pour l'analyse, la protéine hPEBP1, préparée à 6.7  $\mu\text{M}$  dans une solution d'ABC 20mM, pH 8.5 contenant 5% d'acétonitrile, a été diluée au 1/20<sup>ème</sup> dans HCCA/FWI et un aliquot a été déposé par la méthode en couche ultra-mince.



Bien que de nombreuses expériences aient été réalisées avec le stock initial de hPEBP1, l'ensemble des résultats présentés ci-après a été généré avec le deuxième stock de hPEBP1.

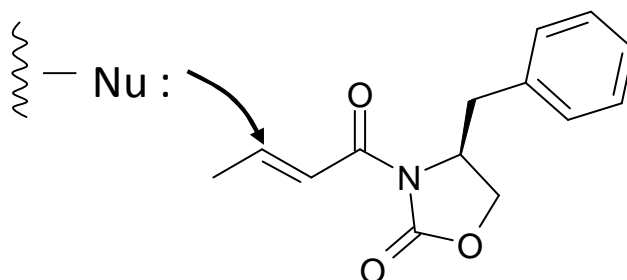
## 4. Dissection du mécanisme de liaison

La dissection du mécanisme de liaison de la locostatine à hPEBP1 a pour objectif d'établir une base solide pour la conception rationnelle d'analogues de locostatine comme agents anti-métastases potentiels. Pour cela, la nature de la liaison, la stabilité du complexe ou encore la cinétique d'association ont été déterminées.

### 4.1 Nature de la liaison

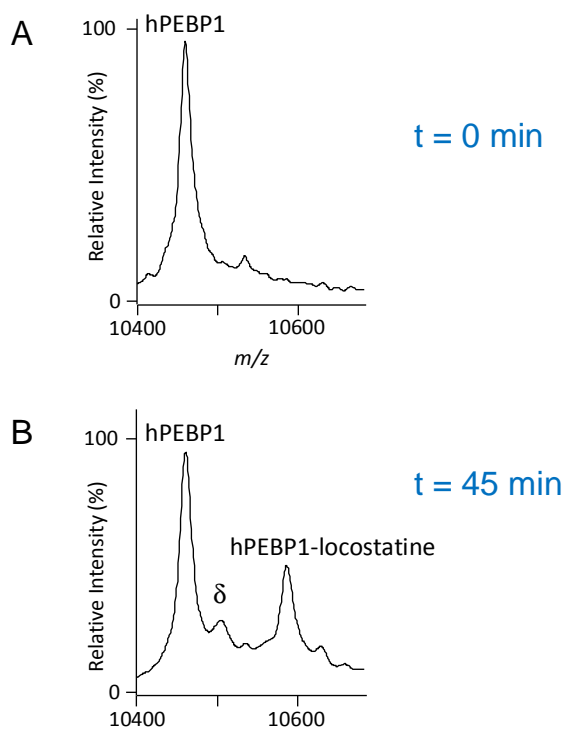
Dès la découverte de la locostatine en 2002, Mc Henry *et al.* avaient fait l'hypothèse qu'elle réagissait de manière covalente avec ses cibles cellulaires *via* une addition de Michael d'un résidu nucléophile de la protéine cible sur la double liaison de la locostatine (Figure 83).

Figure 83. Réaction d'un résidu nucléophile de hPEBP1 sur la locostatine.



Cependant, avant nos travaux, l'interaction covalente entre la locostatine et hPEBP1 n'avait jamais été vérifiée. Les deux partenaires ont donc été incubés ensemble puis analysés par MALDI-TOF MS. Comme déjà évoqué dans le paragraphe 3.3 du chapitre 2, l'analyse MALDI-TOF est un bon outil pour vérifier la nature covalente d'une interaction entre deux molécules. En effet, les conditions d'ionisation en MALDI classique : conditions dénaturantes (pH < 2, [solvant organique] = 30-50 %) et dépôts solides, sont très défavorables à la survie de complexes non-covalents et seuls les complexes covalents sont observés. La Figure 84 montre, après 45 min d'incubation dans un tampon bicarbonate d'ammonium pH 8.0, l'apparition d'un pic à +245.3 Da correspondant au complexe hPEBP1-locostatine (incrément de masse théorique = 245.27 Da). Ces résultats confirment que l'interaction est covalente et qu'elle est très certainement formée par une addition 1,4 d'un résidu nucléophile de la protéine (thiol, alcool ou amine) sur la double liaison électrophile de la locostatine, comme initialement proposé par Mc Henry *et al.*

**Figure 84. Suivi en MALDI-TOF MS de la réaction de hPEBP1 avec la locostatine à t = 0 min et t = 45 min, pH 8.0 37°C. Les espèces dichargées sont illustrées.**



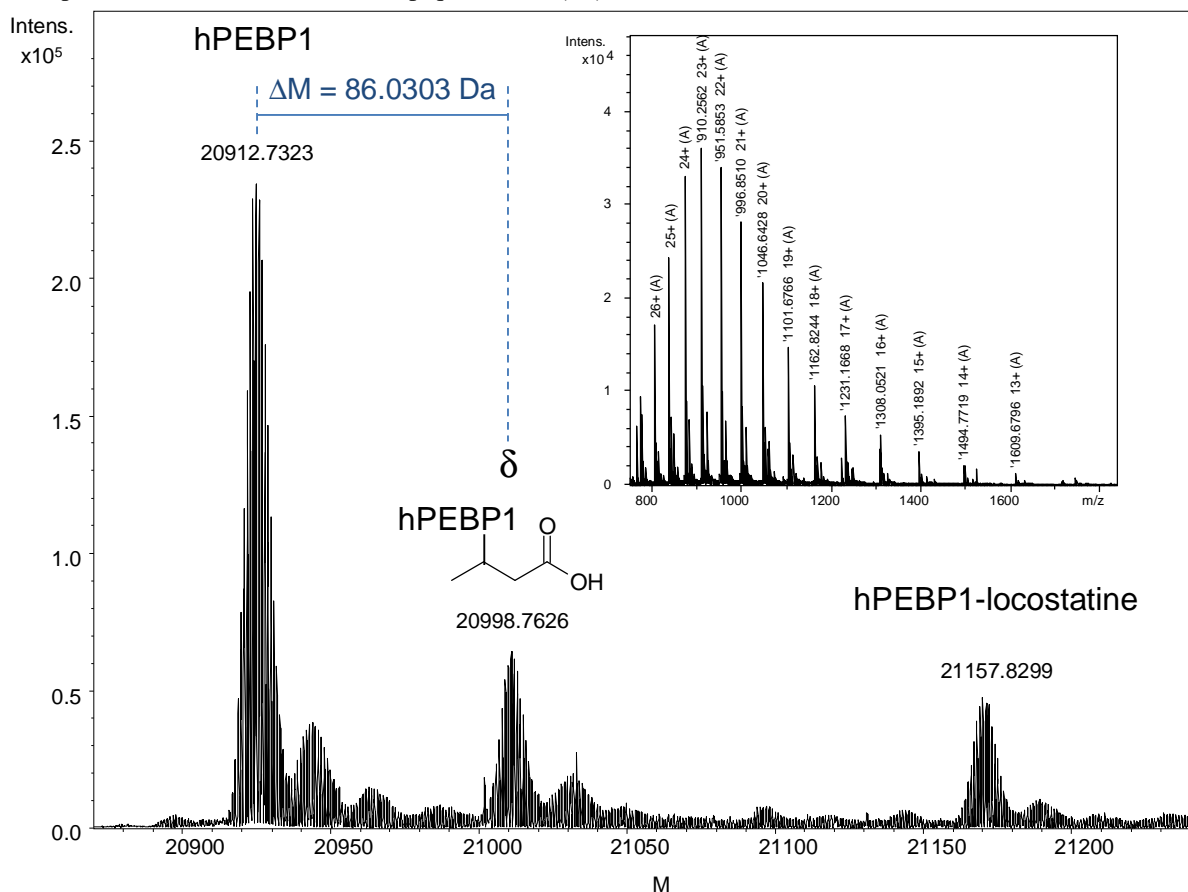
Par ailleurs, le spectre de la Figure 84-B laisse entrevoir un pic additionnel, noté  $\delta$ . Nous avons fait l'hypothèse que ce pic  $\delta$  correspond à un produit de dégradation de la locostatine. La résolution et l'intensité de ce pic sont insuffisantes pour mesurer un delta-masse précis, nécessaire à l'identification de ce composé.

## 4.2 Stabilité du complexe

Afin d'apporter des informations plus précises sur le pic  $\delta$  détecté en MALDI-TOF, le complexe hPEBP1-locostatine a été analysé en ESI-QTOF-HRMS qui présente l'avantage d'avoir une meilleure résolution et une meilleure précision de masse. Le spectre HRMS présenté dans la Figure 85 confirme l'existence de ce pic  $\delta$  et permet d'obtenir un écart de masse monoisotopique expérimentale précis de 86.0303 Da.



**Figure 85. Analyse ESI-Q-TOF HRMS des produits de réaction de hPEBP1 avec la locostatine.** Les masses correspondent à des masses monoisotopiques.  $M_{\text{mono}}(\text{théo}) = 20912.6255 \text{ Da}$



Les formules brutes compatibles dans une fourchette de 10 mDa sont présentées dans le Tableau 13.

**Tableau 13. Formules brutes compatibles avec la masse monoisotopique 86.0303 Da.** La recherche a été réalisée avec l'outil en ligne CHEMCALC ([http://www.chemcalc.org/mf\\_finder/mfFinder\\_em\\_new](http://www.chemcalc.org/mf_finder/mfFinder_em_new)) en autorisant les atomes C, H, O et N. Seuls les résultats inférieurs à 10 mDa sont présentés.

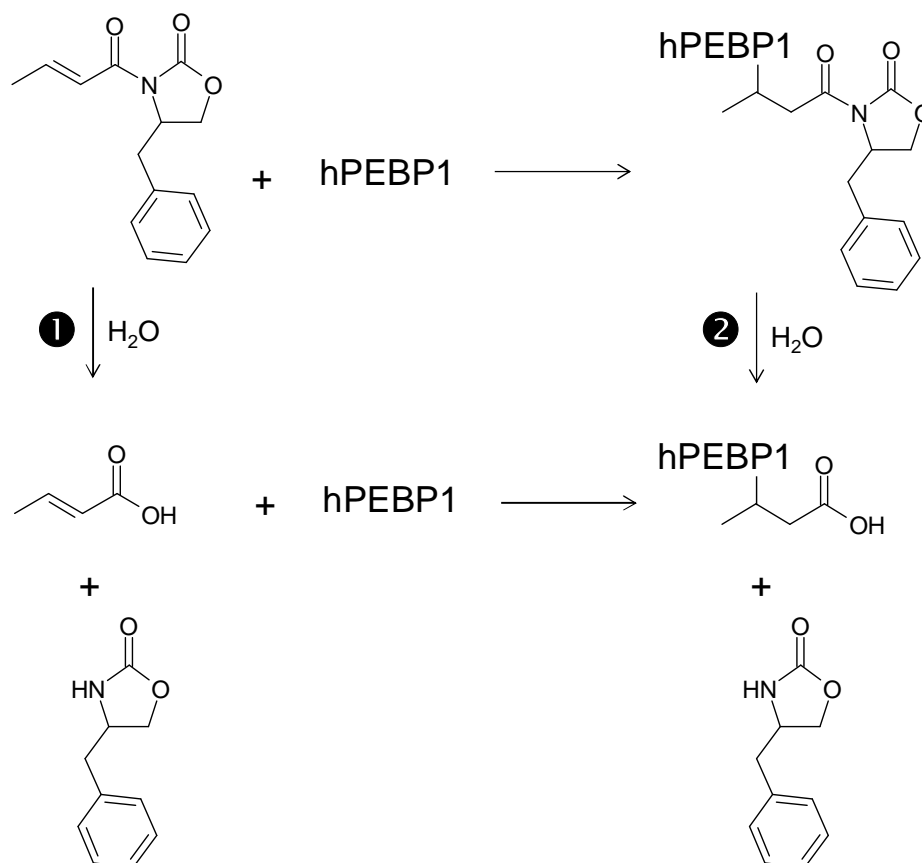
	Formule brute	Masse Monoisotopique	ppm	mDa	insaturation
1	H <sub>2</sub> N <sub>6</sub>	86.0341	59.705	5.137	3
2	C <sub>2</sub> H <sub>4</sub> N <sub>3</sub> O	86.0354	70.871	-6.097	2.5
3	C <sub>3</sub> H <sub>4</sub> NO <sub>2</sub>	86.0242	75.31	6.479	2.5
4	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	86.0368	86.48	-7.439	2
5	CH <sub>2</sub> N <sub>4</sub> O	86.0229	59.705	5.137	3

Selon notre hypothèse d'une dégradation de la locostatine, cette masse pourrait correspondre à l'acide butyrique fixé sur la protéine : formule brute C<sub>4</sub>H<sub>6</sub>O<sub>2</sub> et masse théorique monoisotopique / moyenne = 86.0368 / 86.0894 Da.

Deux voies peuvent conduire à l'observation de ce produit de dégradation (Figure 86) :

- 1 avant fixation sur hPEBP1, dégradation de la locostatine en acide crotonique et (S)-4-benzyloxazolidin-2-one, puis réaction de l'acide crotonique sur hPEBP1
- 2 fixation de la locostatine sur hPEBP1 puis dégradation de la locostatine fixée en acide butyrique et (S)-4-benzyloxazolidin-2-one

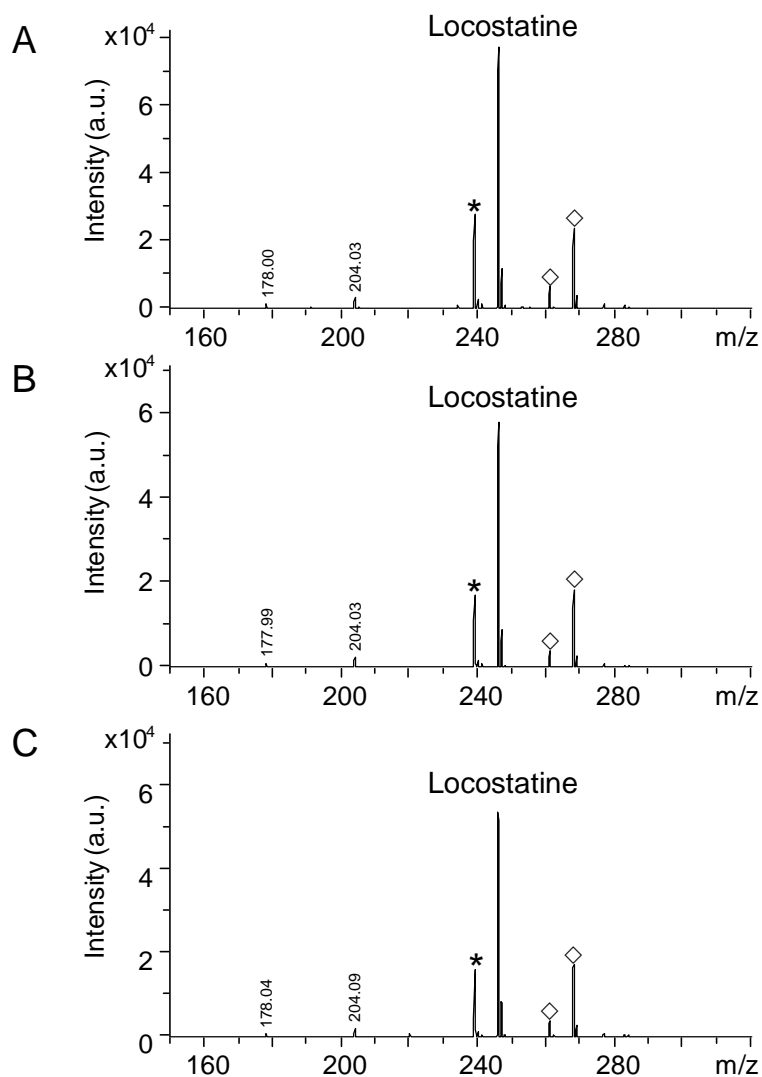
Figure 86. Les voies de dégradation possibles de la locostatine.



Afin de savoir si cette dégradation se produit avant ou après fixation de la locostatine sur la protéine, 1) la stabilité de la locostatine seule a été évaluée et 2) la protéine a été incubée avec l'acide crotonique.

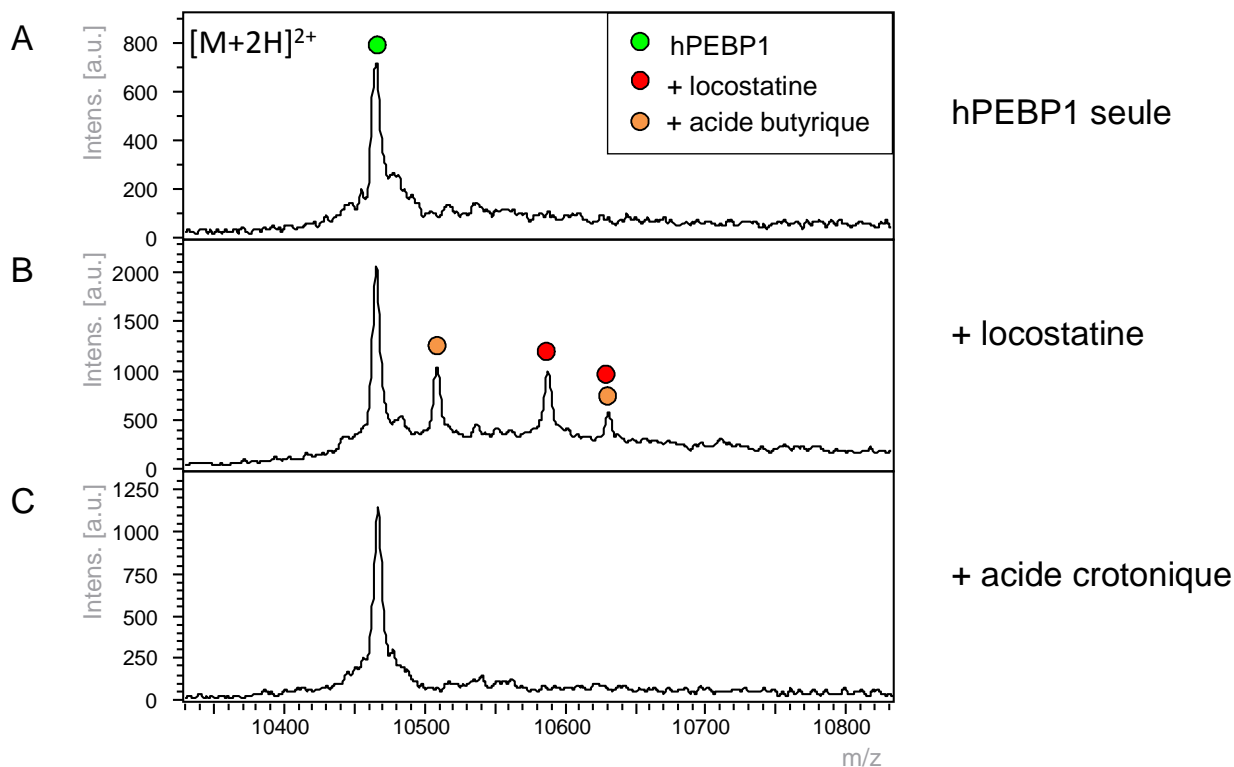
La locostatine a été incubée seule dans les mêmes conditions que celles utilisées en présence de hPEBP1 puis a été analysée en ESI-IT à  $t = 0, 1$  et  $6h$ . Les spectres sont présentés Figure 87 et ne montrent aucune diminution du pic de locostatine ( $m/z$  246) en faveur du pic de (S)-4-benzyloxazolidin-2-one ( $m/z$  178). Ce résultat indique que la locostatine est stable pendant au moins  $6h$ , ce qui va dans le sens de la voie 2 de dégradation.

**Figure 87. Stabilité de la locostatine à pH 7.6 et à 37°C.** Une solution de locostatine 1 mM dans un tampon HEPES-NaOH 2 mM, pH 7.6 contenant 5% d'acétonitrile a été incubée pendant 0 min (A), 1 h (B) et 6 h (C). Des aliquots prélevés aux temps indiqués ont été dilués au 1/20<sup>ème</sup> dans une solution d'acétonitrile:eau:acide formique 45:54.8:0.2 contenant 10 mM d'ABC, puis ils ont été analysés en ESI-IT mode positif. \* ion HEPES,  $\diamond$  adduit de sodium.



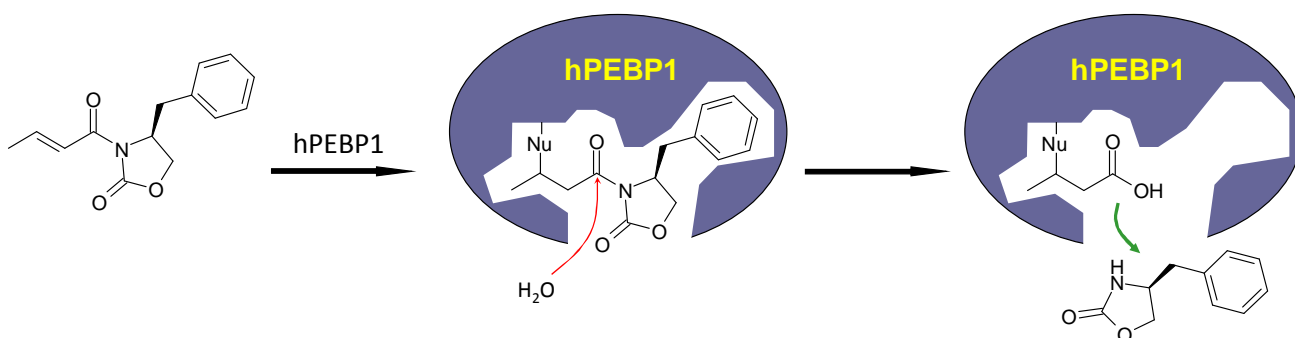
Afin de valider définitivement la voie ②, de l'acide crotonique a été incubé avec hPEBP1 dans des conditions où la locostatine est réactive. Les spectres obtenus sont présentés Figure 88. Aucun pic de complexe hPEBP1-acide butyrique n'est observé après 1h de réaction (Figure 88-C), ce qui indique que l'acide crotonique ne se fixe pas à hPEBP1. Lorsque la locostatine est incubée avec hPEBP1 dans les mêmes conditions de réaction, un pic correspondant à hPEBP1-acide butyrique est observé (Figure 88-B).

**Figure 88. Suivi par MALDI-TOF MS de la réaction de l'acide crotonique avec hPEBP1 à pH 8.5 et 37°C.** Une solution de protéine hPEBP1 à 6.7  $\mu\text{M}$  a été incubée en présence d'acide crotonique ou de locostatine à 1 mM dans un tampon d'ABC 20 mM pendant 1 h puis analysée par MALDI-TOF MS. Pour cela, les réactions ont été diluées au 1/20<sup>ème</sup> dans une solution de matrice HCC/FWI et le mélange matrice-échantillon a été déposé sur cible MALDI selon la méthode en couche ultra-mince.



Ces résultats indiquent que la locostatine s'hydrolyse en acide butyrique après fixation sur hPEBP1. C'est donc la locostatine entière qui est active sur hPEBP1 (Figure 89). L'instabilité de la locostatine fixée peut être liée à la présence de centres électrophiles secondaires.

**Figure 89. Hydrolyse de la locostatine après fixation sur hPEBP1.**



### 4.3 Cinétique d'association

Une cinétique d'association entre hPEBP1 et locostatine a été réalisée afin d'obtenir la constante de vitesse  $k_{on}$  (Figure 90). Pour faciliter l'interprétation des données de liaison, l'expérience a été réalisée en conditions de pseudo-premier ordre, c'est-à-dire  $[ligand] \gg [protéine]$ , ce qui permet de simplifier l'équation cinétique en négligeant la consommation de ligand. La fraction de liaison  $f$  est donnée par l'équation 1:

$$f = \frac{[protéine-ligand]}{[protein]_{totale}} \quad (4)$$

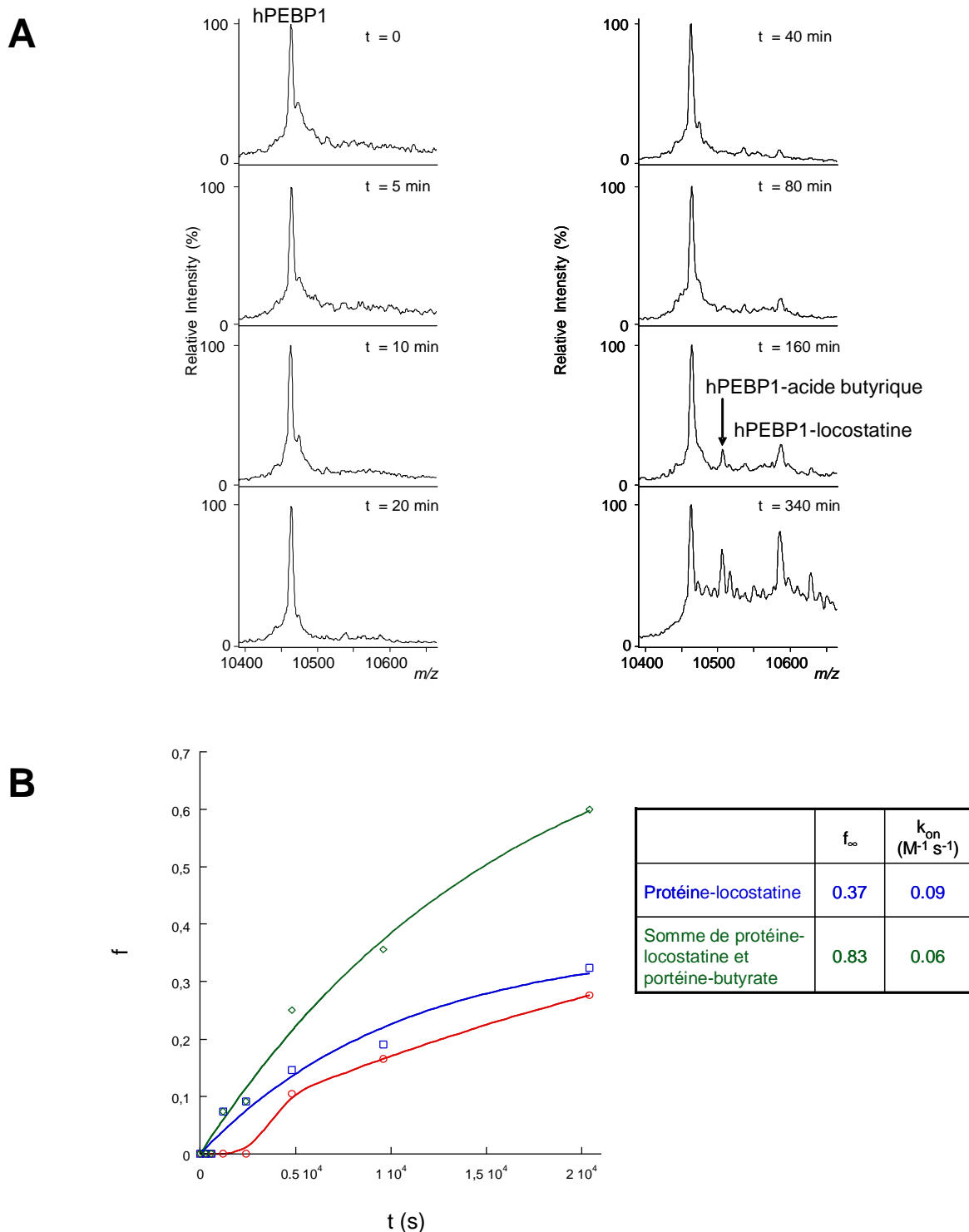
Si l'on pose  $f_{\infty}$ , la fraction de liaison à saturation, une association irréversible suit la loi décrite par l'équation 2:

$$f = f_{\infty} (1 - e^{(-k_{obs}.t)}) \quad (5)$$

La constante de cinétique  $k_{on}$  est ensuite calculée en utilisant l'équation 3 :

$$k_{on} = k_{obs} / [locostatine] \quad (6)$$

**Figure 90. Cinétique d'association entre hPEBP1 et locostatine à pH 7.7 et 37°C.** A, Spectres MALDI-TOF MS de suivi des complexe hPEBP1-locostatine: hPEBP1 (6.6  $\mu$ M) a été incubée en présence de locostatine (1 mM) dans un tampon HEPES 38.5 mM, pH 7.7, TCEP 13  $\mu$ M et acétonitrile 5%. Au cours du temps, des aliquots de 0.5  $\mu$ L de réaction ont été prélevés et dilués au 1/20<sup>ème</sup> dans une solution de matrice HCCA/FWI pour mesurer la formation du complexe hPEBP1-locostatine par MALDI-TOF-MS. La fraction de liaison  $f$  correspondant à  $[\text{protéine-ligand}]/[\text{protéine}]_{\text{totale}}$  a été calculée à partir du rapport signal sur bruit des ions dichargés. B, Courbes de liaison à hPEBP1. Ronds bleu: données du complexe locostatine intacte-hPEBP1; ronds rouges: données  $f$  du complexe acide butyrique-hPEBP1; diamants verts: sommes de  $f$  locostatine intacte-hPEBP1 et acide butyrique-hPEBP1. Les lignes correspondant respectivement à la courbe calculée par régression non linéaire des données pour la fraction hPEBP1-locostatine intacte (équation 4), à l'interpolation des données pour l'apparition de l'espèce acide butyrique-hPEBP1, et à la courbe calculée par régression non linéaire des données pour la somme des fractions locostatine-hPEBP1 et acide butyrique-hPEBP1 (équation 4). Les calculs de régression ont été réalisés en utilisant le logiciel Kaleidagraph 3.6 (Synergy Software).

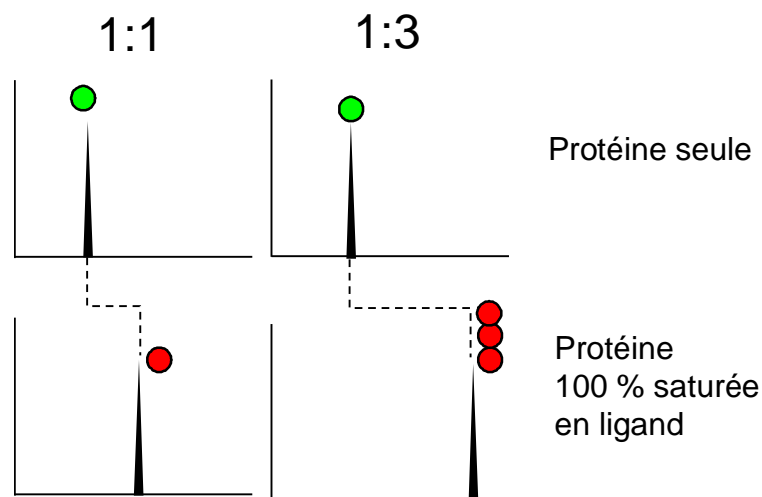


La constante de vitesse  $k_{on}$  obtenue est de  $0.06 \text{ M}^{-1} \text{ s}^{-1}$ , ce qui indique une réaction très lente entre la locostatine et hPEBP1. Cette vitesse de réaction très basse suggère l'existence d'un intermédiaire réactionnel non-covalent et réversible ( $P + L \rightleftharpoons PL^* \rightarrow PL$ ). L'analyse de la réaction en conditions natives (adaptées à l'observation des complexes non covalents), dans des temps trop courts pour la formation de la liaison covalente, n'a pas permis de mettre en évidence le complexe non-covalent. Ceci peut être dû à 1) un  $k_{on}$  très lent et un  $k_{off}$  très rapide, conduisant à une proportion de complexe intermédiaire trop mineure pour être observée; ou 2) à la nature du complexe non covalent, comme par exemple une implication majeure d'interactions hydrophobes qui sont difficiles à préserver en spectrométrie de masse native. La prise en compte de la fraction de liaison hPEBP1-acide butyrique permet également d'atteindre une saturation plus proche de 1.

#### 4.4 Stœchiométrie du complexe

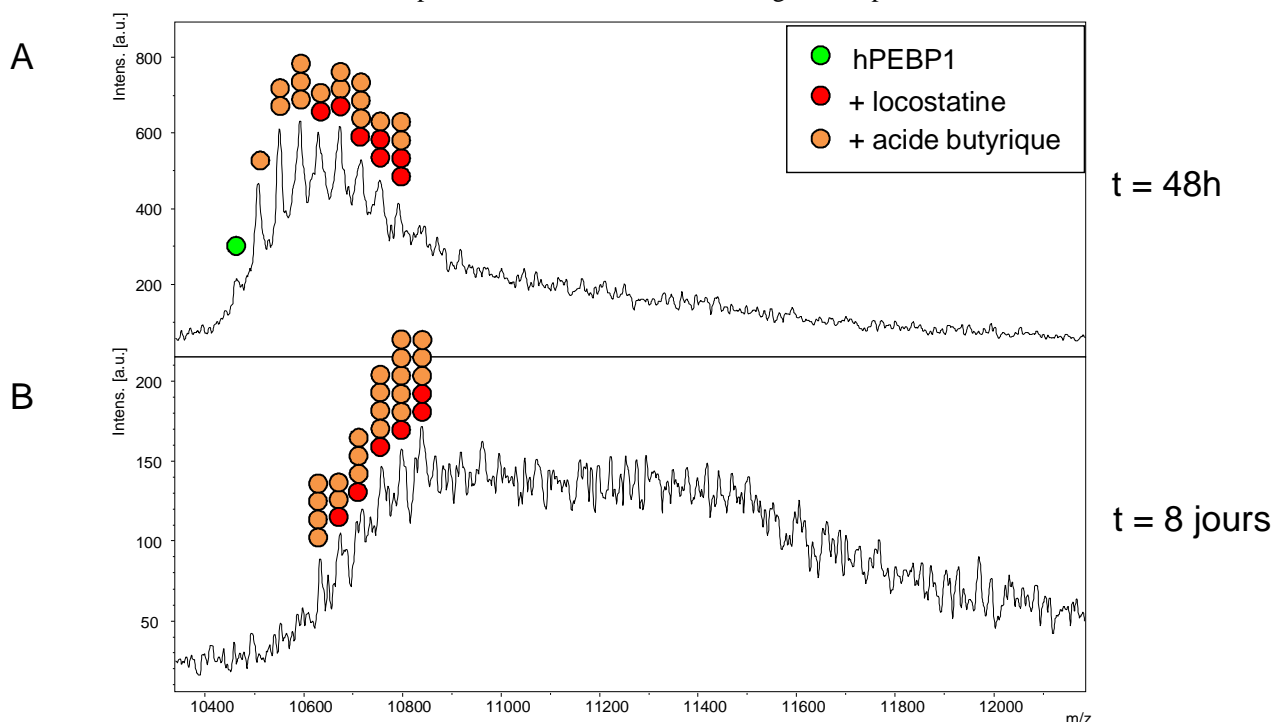
La stœchiométrie du complexe hPEBP1-locostatine a été évaluée par une expérience de saturation de la protéine en ligand. Etant donné la taille modeste de hPEBP1 (~21 kDa) et sa structure en un seul domaine, nous supposons une stœchiométrie 1:1. A titre d'illustration, la Figure 91 montre les spectres théoriques pour des complexes saturés de stœchiométries 1:1 et 1:3.

**Figure 91. Spectres théoriques pour des complexes saturés de stœchiométries 1:1 et 1:3.**



Les spectres expérimentaux obtenus à  $t = 48 \text{ h}$  et  $t = 8 \text{ jours}$  sont présentés Figure 92.

**Figure 92. Essai de saturation du complexe hPEBP1-locostatine à pH 7.7 et 37°C.** Une solution de protéine hPEBP1 à 7.3  $\mu\text{M}$  a été incubée avec 1 mM de locostatine dans un tampon 100 mM HEPES pH 7.7 contenant 1% d'acétonitrile. Afin de prévenir tout risque d'évaporation totale du milieu réactionnel, le volume de la réaction était de 100  $\mu\text{L}$  et des centrifugations toutes les 12 h ont été réalisées. La réaction a été analysée au cours du temps par MALDI-TOF MS après dilution au 1/20<sup>ème</sup> dans HCCA/FWI, dépôt en couche ultra-mince et lavage des dépôts avec du TFA 0.1%.



Ces spectres indiquent que la saturation n'est pas atteinte même au temps très long, ce qui empêche de déterminer la stœchiométrie du système. Par ailleurs, jusqu'à six ligand sont fixés sur hPEBP1, suggérant l'existence de réactions non spécifiques. La protéine hPEBP1 elle-même tend à devenir instable après quelques jours, ce qui rend l'analyse à des temps très long délicate. Ceci nous a conduits à favoriser des temps courts de réaction et à évaluer la réactivité de la locostatine avec des nucléophiles libres en solution.

## 5. Optimisation des conditions de réaction et de protéolyse pour la localisation du site spécifique

### 5.1 Evaluation de la réactivité de la locostatine

La « sur-modification » de hPEBP1 par la locostatine observée pour les temps longs de réaction semble indiquer que la locostatine est capable de réagir avec plusieurs résidus, suggérant que l'exposition au solvant d'un résidu nucléophile peut être suffisante pour conduire à une addition sur la

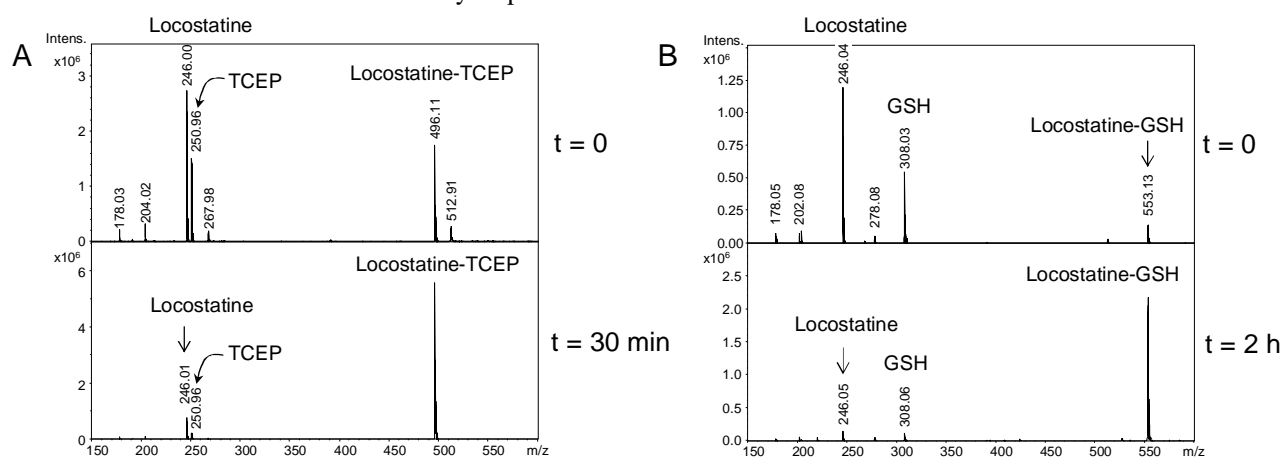


locostatine. Afin de vérifier cette hypothèse, la réactivité de la locostatine a été évaluée sur des agents réducteurs et des peptides.

### 5.1.1 Réaction de la locostatine avec des agents réducteurs

La locostatine a été incubée avec deux agents réducteurs : le TCEP ou le glutathion ( $\gamma$ -Glu-Cys-Gly, GSH), puis la réaction a été analysée en ESI-IT. Les spectres présentés Figure 93 montrent l'apparition de pics de  $m/z$  496 et 553 correspondant respectivement au produit de réaction locostatine-TCEP et locostatine-GSH. Ces résultats indiquent que la locostatine possède un potentiel électrophile suffisant pour réagir avec des nucléophiles libres en solution et nous ont conduits à limiter l'ajout d'agent réducteur afin que la locostatine soit disponible pour réagir avec hPEBP1

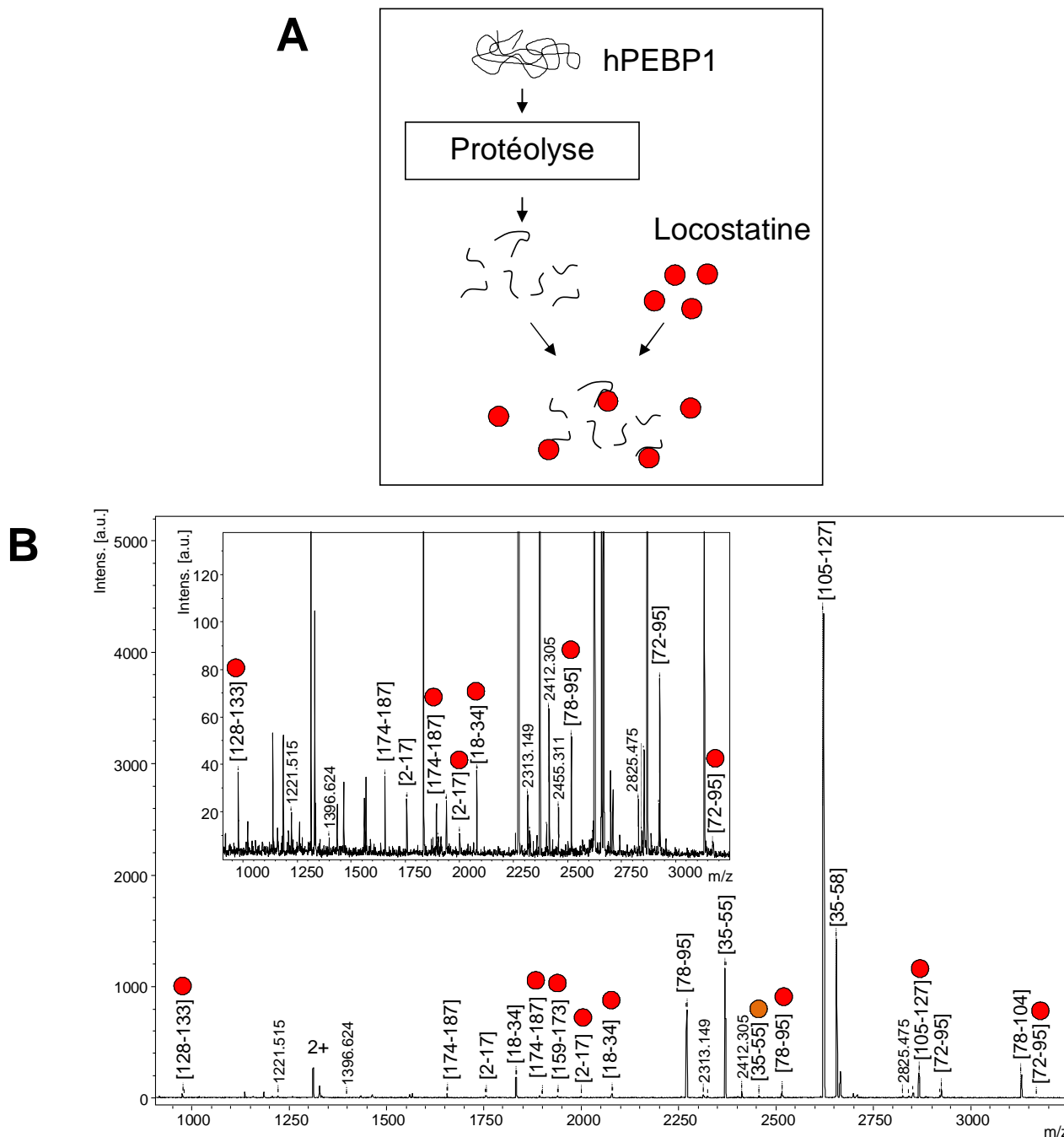
**Figure 93. Réaction de la locostatine avec des agents réducteurs à pH 8.5 et 21°C.** Les réactions ont été réalisées dans une solution d'ABC 20 mM. A, une solution de locostatine à 1 mM a été incubée en présence de TCEP à 1 mM. Des aliquots du mélange réactionnel ont été prélevés à  $t = 0$  et  $t = 30$  min et dilués au  $1/50^{\text{ème}}$  dans une solution d'ionisation acétonitrile:eau:acide formique 50:49.8:0.2 puis analysés par ESI-IT MS. B, une solution de locostatine à 0.5 mM a été incubée en présence de GSH à 0.5 mM. Des aliquots du mélange réactionnel ont été prélevés à  $t = 0$  et  $t = 2$  h, dilués au  $1/10^{\text{ème}}$  dans la solution d'ionisation et analysés par ESI-IT MS.



### 5.1.2 Réaction de la locostatine avec des peptides

La réactivité de la locostatine a été testée avec des peptides issus de la protéolyse de hPEBP1 par la protéinase Asp-N. Le suivi de la réaction par analyse MALDI-TOF MS en mode réflectron au temps  $t = 4$  h (Figure 94) permet de mettre en évidence neuf peptide modifiés par la locostatine. Parmi eux, le peptide [35-55] est observé avec la modification hydrolysée en groupement butyrate.

**Figure 94. Suivi par analyse MALDI-TOF MS de la réactivité de la locostatine sur des peptides de hPEBP1 à pH 7.7 et à 37°C.** A, Schéma simplifié de l'expérience : la protéine hPEBP1 à 5  $\mu$ M a été dénaturée par un traitement thermique de 10 min à 60°C en présence d'urée 1 M et d'acétonitrile 20%. La protéolyse par Asp-N a ensuite été effectuée à un rapport enzyme:substrat [1/20] (p/p) pendant 2 h à 37°C dans un tampon HEPES à 50 mM pH 7.7. Après élimination de l'acétonitrile par évaporation speedvac de 10 min, le volume des protéolysats a été complété par de l'eau q.s.p. 20  $\mu$ L puis les peptides de clivage de hPEBP1 ont été mis en contact avec 2 mM de locostatine. B, Analyse MALDI-TOF MS de la réaction de la locostatine sur des peptides de clivage de hPEBP1 après 4 h d'incubation. Le milieu réactionnel a été dilué au 1/10<sup>ème</sup> dans HCCA/TWA, déposé en couche ultra-mince sur cible MALDI puis les dépôts ont été lavés avec du TFA 0.1%. Les peptides modifiés par la locostatine ou l'acide butyrique sont annotés d'un disque rouge ou orange, respectivement.



Les résidus nucléophiles pouvant réagir avec la locostatine sont : Cys, Lys, Ser, Thr, Tyr, ou His. Ce résultat montre qu'il est absolument nécessaire de neutraliser ou d'éliminer la locostatine n'ayant pas réagi sur hPEBP1 avant l'étape de protéolyse, afin d'éviter des réactions non spécifiques de la

locostatine sur des peptides. Ceci nous a conduits à mettre en œuvre plusieurs stratégies pour neutraliser ou éliminer la locostatine.

## 5.2 Solutions pour limiter les réactions non spécifiques

### 5.2.1 Au niveau de la protéine

Nous avons vu que les temps longs de réaction entraînent des stœchiométries de modification de hPEBP1 supérieures à 1, correspondant très certainement à des réactions non spécifiques. Si l'on fait l'hypothèse qu'il existe un site spécifique, on peut ajouter le corolaire que ce site est majoritaire par rapport aux sites non-spécifique. Afin de limiter ces réactions parasites, le temps de réaction devra donc être ajusté pour s'assurer :

- qu'après la fixation de la locostatine sur le site spécifique de hPEBP1, il n'y ait pas de fixation secondaire non spécifique;
- que la réaction respecte si possible des conditions de "single-hit statistics", c'est-à-dire qu'en moyenne, une seule réaction par chaîne protéique a lieu. Cette dernière condition permet également d'éviter des effets allostériques d'un site sur un autre.

Ces deux conditions nous conduisent à privilégier la stœchiométrie 1:1. Les arguments en faveur de cette stœchiométrie seront détaillés au paragraphe 7.

### 5.2.2 Au niveau des peptides

Deux voies ont été explorées pour limiter les réactions non spécifiques de la locostatine sur les peptides de hPEBP1 : la neutralisation et l'élimination de l'excès de locostatine.

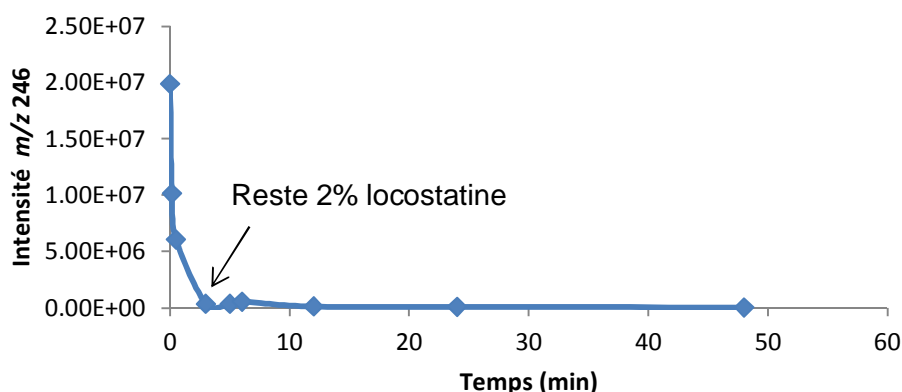
#### 5.2.2.1 Neutralisation de la locostatine en excès

Ayant démontré que la locostatine était capable de réagir avec des agents réducteurs, la neutralisation de la locostatine a été réalisée par ajout de TCEP. Une concentration élevée de 10 mM, équivalent à un rapport TCEP:locostatine de 10:1, a été utilisé afin de garantir une neutralisation efficace de la locostatine dans un minimum de temps. La Figure 95 présente une cinétique de disparition de la locostatine en présence de TCEP et en absence de protéine. Cette cinétique montre que le TCEP permet d'éliminer 98% de la locostatine en 3 min et plus de 99.5% en 30 min, ce que nous

considérons comme satisfaisant. Les conditions retenues pour la neutralisation de réaction hPEBP1-locostatine sont 10 mM TCEP pendant 30 min à température ambiante.

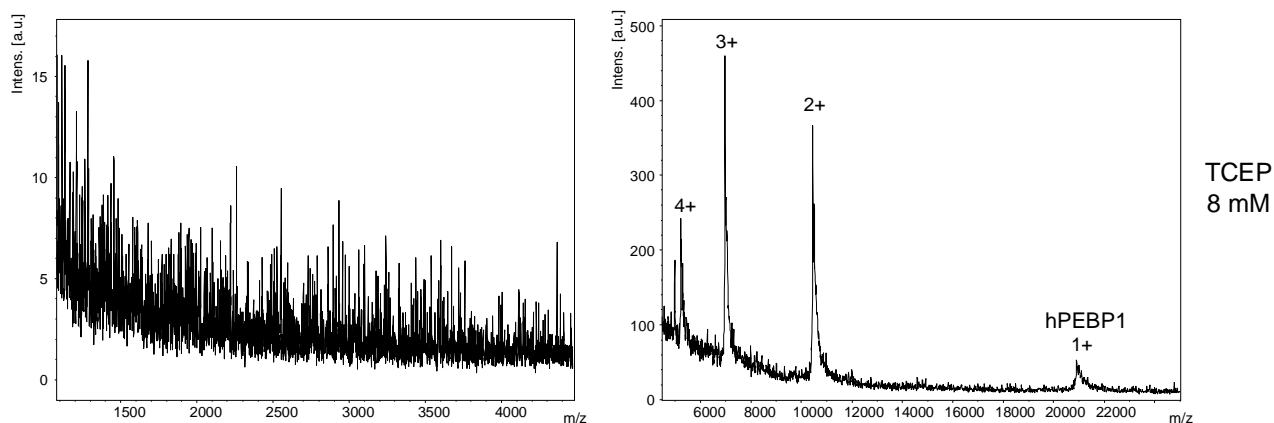
**Figure 95. Cinétique de réaction entre la locostatine et le TCEP à pH 8.3 et 21°C.** Une solution de locostatine 1 mM a été incubée en présence de TCEP à 10 mM dans un tampon ABC 100 mM, pH 8.3 contenant 5% d'acétonitrile. La cinétique de réaction de la locostatine avec le TCEP immobilisé sur billes a été évaluée en mesurant la décroissance de la locostatine libre ( $m/z$  246) au cours du temps par ESI-IT MS. Pour cela, un aliquot de 0.5  $\mu$ L de milieu réactionnel a été prélevé, dilué au 1/100<sup>ème</sup> dans une solution d'ionisation acétonitrile:eau:acide formique 50:49.8:0.2 puis analysé en infusion. Afin d'observer le signal minoritaire de la locostatine ( $m/z$  246), son intensité a été mesurée non pas en MS simple mais après isolation du  $m/z$  246. L'instrument fonctionnait avec un temps d'accumulation fixé à 0.30 ms.

### Cinétique de réaction entre la locostatine et le TCEP



Toutefois, la concentration élevée en agent réducteur n'est pas adaptée pour l'étape suivante du protocole : la protéolyse. En effet, dans ces conditions hautement réductrices (8 mM final lors de la protéolyse), l'activité des protéases telles que la trypsine est perturbée, ce qui se traduit par la présence accrue de protéine intacte après protéolyse (Figure 96).

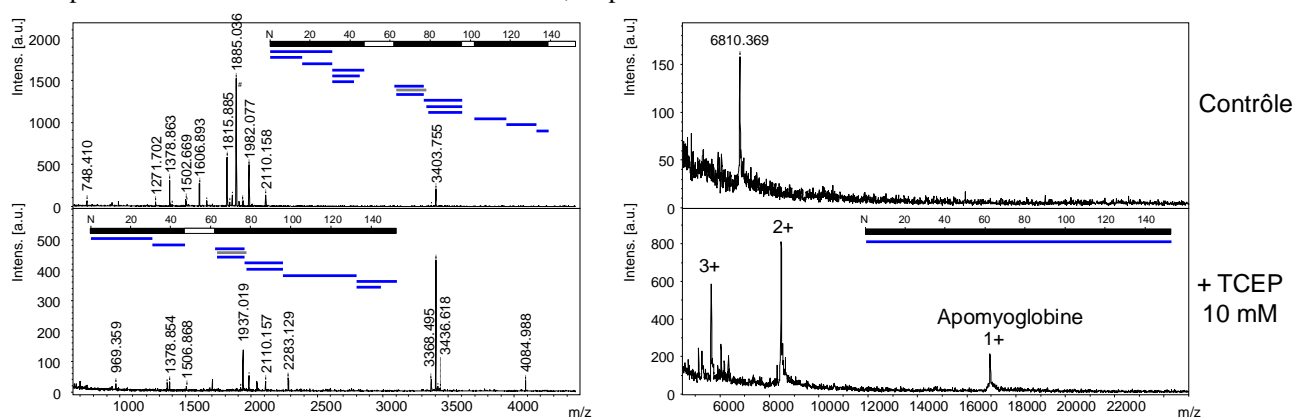
**Figure 96. Analyse MALDI-TOF de hPEBP1 protéolysée par la trypsine en présence de 8 mM de TCEP.** Une solution de protéine hPEBP1 à 7  $\mu$ M a été incubée avec 1 mM de locostatine pendant 1 h à 37°C dans 20 mM d'ABC, pH 8.5 contenant 10% d'acétonitrile. La réaction a été arrêtée par ajout de TCEP 10 mM final pendant 30 min à 21°C suivie d'une protéolyse par la trypsine ajoutée à un rapport enzyme:substrat de [1:10] (p/p) pendant 4 h à 37°C dans 100 mM d'ABC, pH 8.5. La concentration finale de TCEP lors de la protéolyse est de 8 mM. L'analyse MALDI-TOF a été réalisée après dilution des échantillons au 1/10<sup>ème</sup> dans HCCA/TWA, dépôt en couche ultra-mince et lavage des dépôts avec du TFA 0.1%. Deux gammes d'acquisition sont présentées :  $m/z$  1000-4500 et 4500-25000.



Ce problème de protéolyse a été confirmé avec l'apomyoglobine, utilisée classiquement comme contrôle de protéolyse dans notre laboratoire (Figure 97). L'étude de l'effet du TCEP sur l'efficacité de protéolyse de l'apomyoglobine par la trypsine montre que :

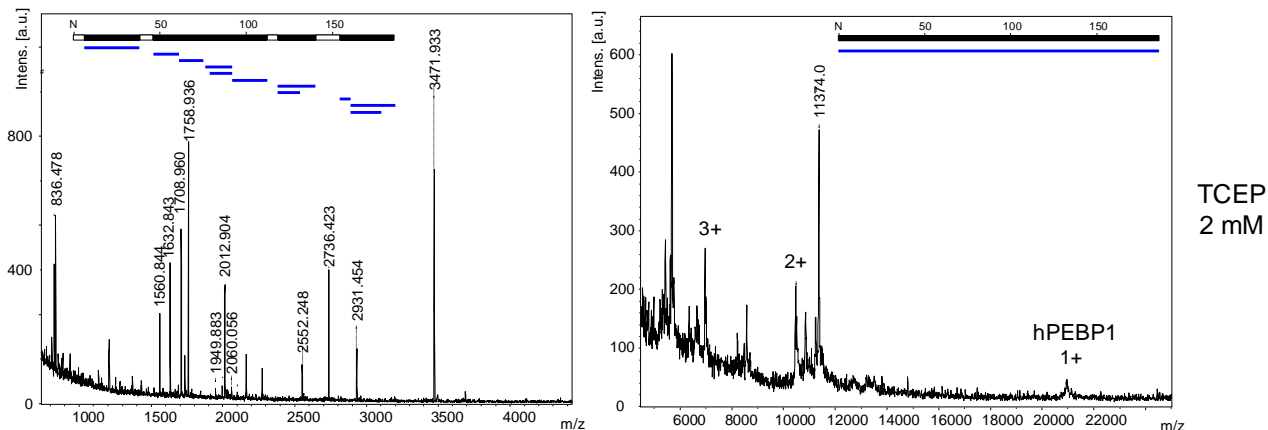
- globalement, la couverture de séquence peptidique n'est pas impactée, mais que
- la progression de la protéolyse est ralentie, ce qui mène à davantage de peptides longs et de protéine intacte.

**Figure 97. Effet du TCEP à 10 mM sur la protéolyse de l'apomyoglobine par la trypsine.** L'apomyoglobine (5  $\mu$ M) a été protéolysée par la trypsine à un rapport E:S de [1:10] (p/p) pendant 4 h à 37°C, en présence ou absence de TCEP à 10 mM, dans une solution d'ABC à 100 mM pH 8.5. Les échantillons ont été dilués au 1/10 dans HCCA/TWA, déposés en couche ultra-mince sur la cible MALDI et les dépôts ont été lavés avec du TFA 0.1%. Deux gammes d'acquisition sont présentées :  $m/z$  700-4500 et 4500-25000. Le graphique en haut des spectres  $m/z$  700-4500 représente la couverture de séquence obtenue. Les peptides observés sont représentés par des rectangles bleus. Les zones noires et blanches correspondent aux zones observées et non observées, respectivement.

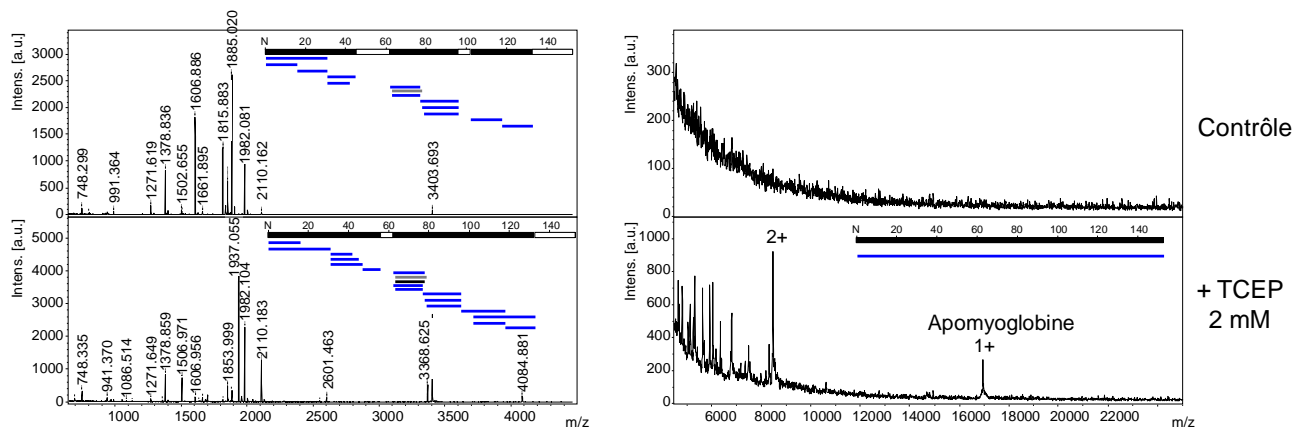


Des essais ont été réalisés en diminuant la concentration de TCEP à 2 mM lors de la protéolyse (Figures 98 et 99).

**Figure 98. Analyse MALDI-TOF de hPEBP1 protéolysée par la trypsine en présence de 2 mM de TCEP.** Une solution de protéine hPEBP1 à 7  $\mu$ M a été incubée avec 1 mM de locostatine pendant 1 h à 37°C dans 20 mM d'ABC, pH 8.5 contenant 10% d'acétonitrile. La réaction a été arrêtée par ajout de TCEP à une concentration finale de 5 mM pendant 30 min à 21°C suivie d'une protéolyse par la trypsine ajoutée à un rapport enzyme:substrat de [1:10] (p/p) pendant 14h30 à 37°C dans 100 mM d'ABC, pH 8.5. La concentration finale de TCEP lors de la protéolyse est de 2 mM. L'analyse MALDI-TOF a été réalisée après dilution des échantillons au 1/10<sup>ème</sup> dans HCCA/TWA, dépôt en couche ultra-mince et lavage des dépôts avec du TFA 0.1%. Deux gammes d'acquisition sont présentées :  $m/z$  700-4500 et 4500-25000. Le graphique en haut à gauche du spectre  $m/z$  700-4500 représente la couverture de séquence obtenue. Les peptides observés sont représentés par des rectangles bleus. Les zones noires et blanches correspondent aux zones observées et non observées, respectivement.



**Figure 99. Effet du TCEP à 2 mM sur la protéolyse de l'apomyoglobine par la trypsine.** L'apomyoglobine (5  $\mu$ M) a été protéolysée par la trypsine à un rapport E:S de [1:10] (p/p) pendant 4 h à 37°C, en présence ou absence de TCEP à 2 mM, dans une solution d'ABC à 100 mM pH 8.5. Les échantillons ont été dilués au 1/10<sup>ème</sup> dans HCCA/TWA, déposés en couche ultra-mince sur la cible MALDI et les dépôts ont été lavés avec du TFA 0.1%. Deux gammes d'acquisition sont présentées :  $m/z$  700-4500 et 4500-25000. Le graphique en haut des spectres  $m/z$  700-4500 représente la couverture de séquence obtenue. Les peptides observés sont représentés par des rectangles bleus. Les zones noires et blanches correspondent aux zones observées et non observées, respectivement.



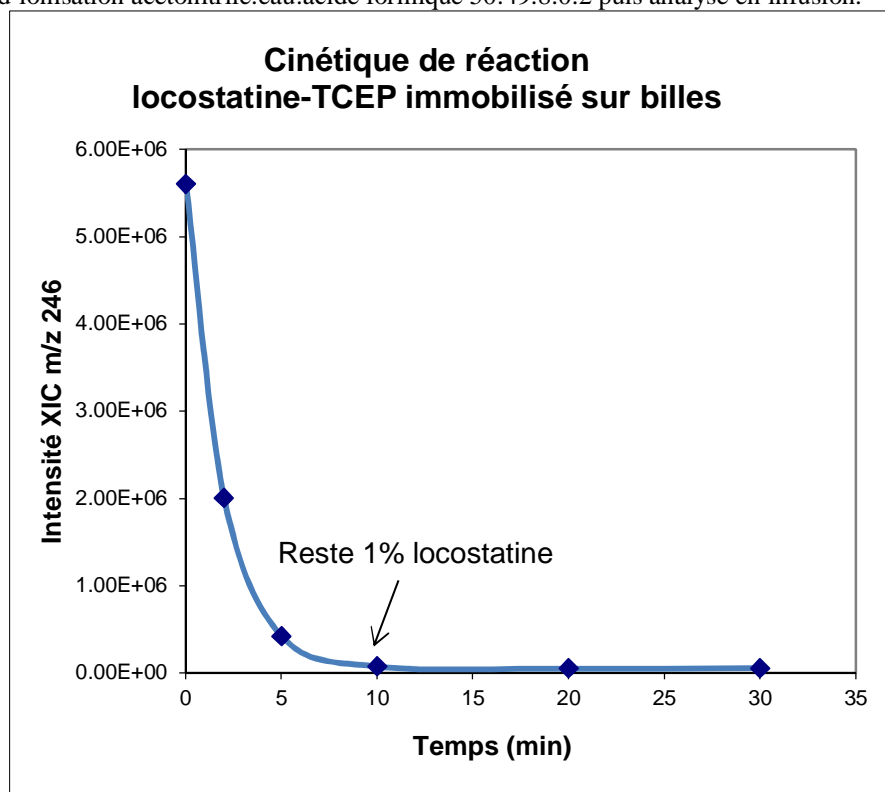
Ces conditions moins réductrices permettent d'améliorer l'activité de la trypsine, avec notamment l'observation de peptides de clivage qui n'étaient pas visibles en présence de 8 mM de TCEP, mais il reste encore de la protéine intacte. Une dilution plus importante de la réaction hPEBP1-locostatine avant protéolyse permettrait de diminuer encore la concentration de TCEP et donc d'améliorer les performances de la protéase mais serait au détriment d'un bon rapport signal/bruit.

Etant donné que pour la caractérisation de modification, il est essentiel d'avoir une protéolyse complète (cf paragraphe 1.2.1 de la section 2), la neutralisation de la locostatine en excès par le TCEP n'a pas été retenue.

### 5.2.2.2 Elimination de la locostatine en excès par TCEP immobilisé sur bille

L'utilisation de TCEP immobilisé sur bille est inspirée de la neutralisation en solution de la locostatine, avec l'avantage de ne pas laisser d'agent réducteur résiduel dans le milieu réactionnel. L'efficacité d'élimination de la locostatine par le TCEP immobilisé sur bille a d'abord été évaluée en absence de protéine en mesurant la cinétique de disparition de la locostatine libre au cours du temps (Figure 100).

**Figure 100. Cinétique de réaction entre la locostatine et le TCEP immobilisé sur billes à pH 8.5 et 21°C.** Une solution de locostatine à 1 mM (soit 0.245 mg/mL) a été incubée avec 30 µL de TCEP immobilisé sur billes préalablement équilibré avec deux lavages de 600 µL d'ABC 20 mM, dans l'ABC 20 mM contenant 5% d'acétonitrile. D'après le fournisseur, le TCEP est fonctionnalisé à hauteur de 8 nmol/µL de gel soit une quantité de 240 nmol pour 30 µL. Cette quantité de TCEP est 12 fois plus élevée que la quantité de locostatine (20 nmol). Au cours de l'incubation, la solution a été agitée afin de maintenir les billes en suspension. La cinétique de réaction de la locostatine avec le TCEP immobilisé sur billes a été évaluée en mesurant la décroissance de la locostatine libre ( $m/z$  246) au cours du temps par ESI-IT MS. Pour cela, un aliquot de 0.5 µL de milieu réactionnel a été prélevé (après centrifugation), dilué au 1/100<sup>ème</sup> dans une solution d'ionisation acétonitrile:eau:acide formique 50:49.8:0.2 puis analysé en infusion.



Cette cinétique indique que le TCEP immobilisé sur billes permet d'éliminer 99% de la locostatine en 10 min, ce que nous considérons comme efficace.

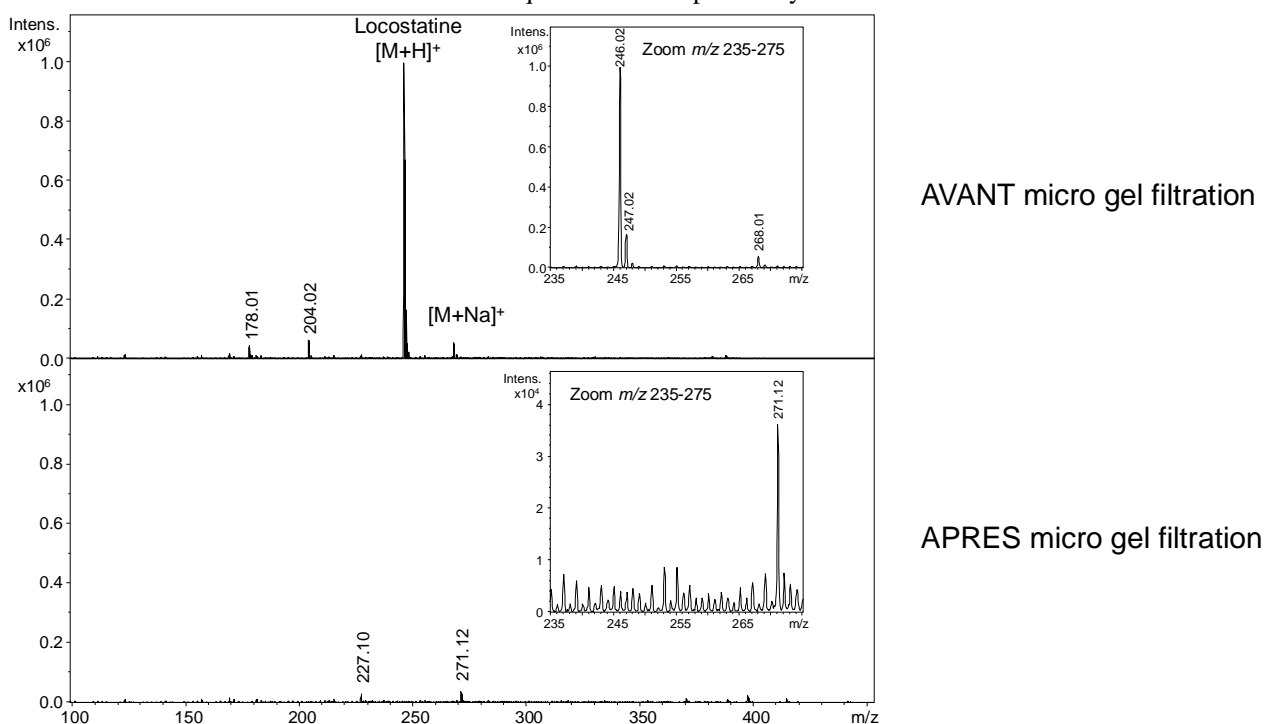
### 5.2.2.3 Elimination de la locostatine en excès par micro gel filtration

La micro gel filtration consiste à séparer physiquement les molécules de "tailles" différentes. C'est une méthode classiquement utilisée pour dessaler les protéines. Le principe repose sur la chromatographie d'exclusion stérique (SEC : Size Exclusion Chromatography) dont la phase

stationnaire poreuse sert de tamis moléculaire. Dans notre cas, seules les molécules de masse  $< 7000$  Da vont pouvoir entrer dans les pores de la phase stationnaire et y séjourner. Les molécules de masse  $> 7000$  Da, sont exclues du gel. Après une simple étape de centrifugation, hPEBP1 est ainsi récupérée dans la fraction non-retenue, la locostatine étant retenue dans la phase stationnaire.

L'efficacité d'élimination de la locostatine par micro gel filtration a été évaluée en mesurant le signal MS de la locostatine avant et après cette étape. Aucun signal correspondant à la locostatine n'est détecté après micro gel filtration (Figure 101) ce qui traduit une très bonne efficacité de séparation.

**Figure 101. Elimination de la locostatine libre par micro gel filtration.** Une solution de locostatine à 2 mM préparée dans l'ABC 20 mM pH 8.5 contenant 5% d'acétonitrile, a été passée sur micro gel filtration 0.5 mL préalablement conditionnée par trois lavages de 300  $\mu$ L d'ABC 20 mM. L'élimination de la locostatine libre a été évaluée par ESI-IT MS. Pour cela, la solution de locostatine, avant et après passage sur micro gel filtration, a été diluée au 1/50<sup>ème</sup> dans une solution d'ionisation acétonitrile:eau:acide formique 50:49.8:0.2 puis analysée en infusion.



Par rapport au TCEP immobilisé sur billes, la micro gel filtration présente l'avantage d'être plus rapide (30 sec vs 30 min), ce qui limite les risques d'évolution de la réaction ou les pertes par adsorption.

Les deux méthodes sont efficaces pour éliminer la locostatine en excès et sont compatibles avec une protéolyse ultérieure.



## 5.3 Optimisation des conditions de protéolyse

L'optimisation des conditions de protéolyse est essentielle pour la caractérisation de modification par approche bottom-up. Elle a pour objectif d'obtenir 1) une protéolyse totale ou complète de hPEBP1 dans un temps court d'incubation et 2) une couverture de séquence de 100%.

### 5.3.1 Efficacité de protéolyse

Pour les protéines modifiées, il est préférable de se placer dans des conditions de protéolyse totale ou la plus complète possible afin d'avoir une efficacité de protéolyse proche de 100%. Ces conditions garantissent qu'il n'y ait plus de protéine intacte après protéolyse et que les peptides ne contiennent que peu voire pas de MC. En effet, si la forme modifiée est plus résistante que la protéine non-modifiée dans les conditions adoptées pour la protéolyse, le degré de complétion de la protéolyse sera différent entre les deux formes, impactant directement la détection et la quantification des peptides modifiés. Dans les cas où une modification génère une nouvelle conformation tridimensionnelle, ces deux critères permettent d'assurer que les conditions de protéolyse produisent une efficacité de protéolyse identique entre les formes modifiée et non modifiée d'une protéine.

Par ailleurs, effectuer une protéolyse rapide présente plusieurs avantages : 1) limiter l'hydrolyse de modification dégradable, telle que la locostatine, 2) limiter l'apparition de modifications induites accidentelles, telles que l'oxydation et 3) limiter les pertes par adsorption. Il y a donc un grand intérêt d'obtenir une protéolyse complète dans un temps court d'incubation.

Notre méthode pour évaluer l'efficacité de protéolyse consiste à déterminer la limite de détection (LOD : Limit Of Detection) de hPEBP1 intacte, puis à analyser le protéolysat de hPEBP1 dans lequel la concentration de départ en protéine est connue. L'absence de signal correspondant à la protéine entière dans le protéolysat indique que la concentration de hPEBP1 intacte est en dessous de la LOD. Cette absence de signal nous permet d'estimer l'efficacité de protéolyse avec l'équation suivante :

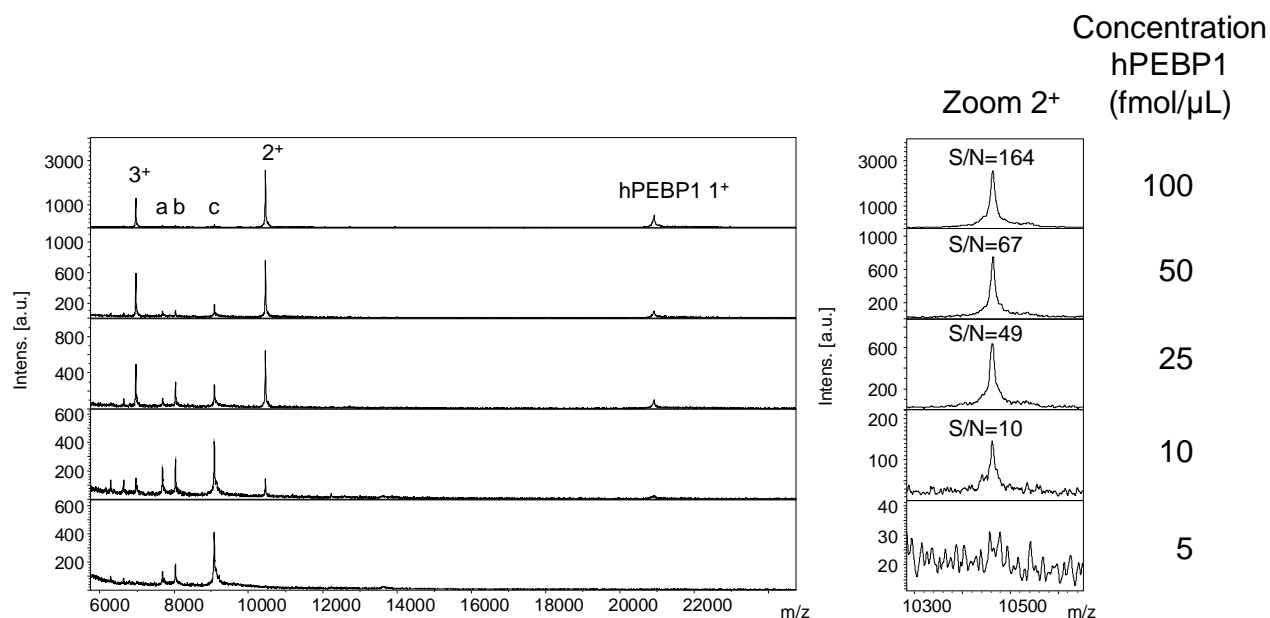
$$\text{Efficacité de protéolyse} = \left(1 - \frac{\text{LOD hPEBP1}}{[\text{hPEBP1}]_{\text{protéolysat}}}\right) \times 100$$

Ainsi, l'analyse d'un protéolysat cent fois plus concentré que la LOD permettrait de déterminer dans le meilleur des cas (absence de protéine intacte détectée) une efficacité de protéolyse supérieure ou égale à 99%.

La LOD de hPEBP1 a été déterminée en présence d'un mélange de peptides pour avoir une compétition lors de l'ionisation et ainsi se rapprocher au mieux de l'analyse de hPEBP1 protéolysée.

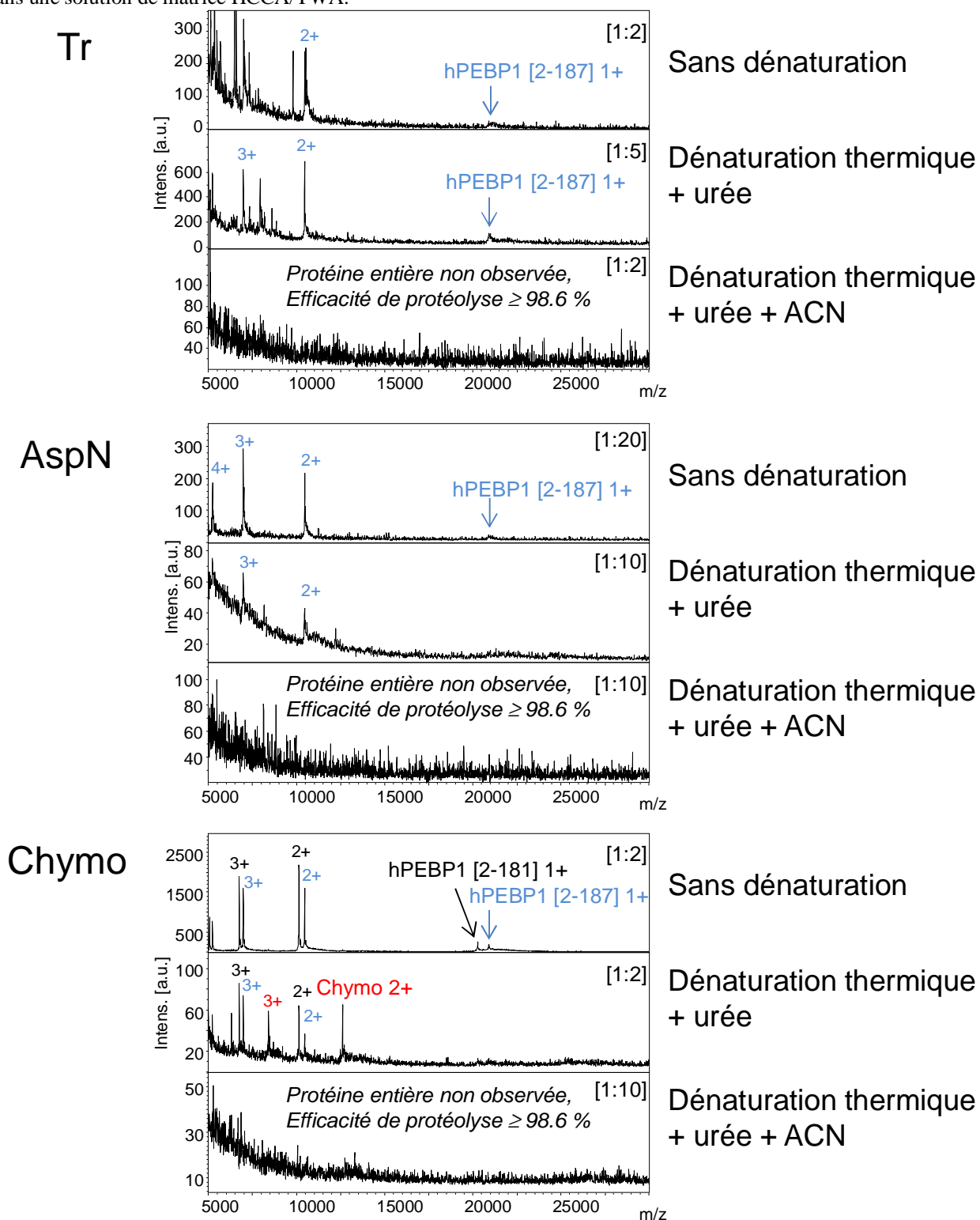
La LOD de hPEBP1 obtenue se situe autour de 5 fmol/μL en MALDI-TOF avec la méthode en couche ultra-mince (Figure 102).

**Figure 102. Limite de détection de hPEBP1 en MALDI-TOF avec la méthode en couche ultra-mince.** Un mélange de peptides contenant les peptides de calibration pepmix 1 (Bruker), l'ACTH 1-39 et l'insuline B oxydée, a été utilisé pour avoir une compétition entre les peptides et la protéine lors de l'ionisation et ainsi mimer l'analyse de hPEBP1 après protéolyse. Pour cela ce mélange de peptides a été dilué à 100 fmol/μL dans la matrice HCCA/TWA et cette solution matrice-peptides a été utilisée pour préparer les différentes concentrations de hPEBP1. Les concentrations indiquées correspondent aux concentrations de hPEBP1 dans le mélange matrice-peptides. Le rapport signal sur bruit de l'état de charge 2+ de hPEBP1 est indiqué. Les pics annotés a, b et c correspondent à des dimères d'ACTH [1-39]-somatostatine, ACTH [1-39]-Insuline B oxydée et ACTH [1-39]-ACTH [1-39], respectivement.



La Figure 103 présente les efficacités de protéolyses obtenues après 2 h de protéolyse de hPEBP1 par les protéases trypsine (Tr), Asp-N et chymotrypsine (Chymo).

**Figure 103. Efficacités de protéolyse de hPEBP1 par les protéases trypsine, Asp-N et chymotrypsine.** La dénaturation de hPEBP1 (7.3  $\mu$ M) a été réalisée par un traitement thermique de 10 min à 60°C en présence d'urée 1 M ou d'urée 1 M et d'acétonitrile 20%. Les protéolyses ont été effectuées dans une solution d'ABC 100 mM, pH 8.3 pour la trypsine et la chymotrypsine et dans une solution d'acétate d'ammonium 100 mM, pH 7.1 pour l'Asp-N. Pour la chymotrypsine du CaCl<sub>2</sub> 10 mM a également été ajouté. L'incubation a été de 2h à 37°C pour la trypsine et l'Asp-N et 2h à 25°C pour la chymotrypsine. Les rapports enzyme:substrat (p/p) sont indiqués en haut à droite de chaque spectre. Les échantillons ont été analysés par MALDI-TOF par la méthode en couche ultra-mince après avoir été dilués au 1/20<sup>ème</sup> dans une solution de matrice HCCA/TWA.



Trois conditions ont été testées : 1) sans dénaturation, 2) avec une dénaturation thermique de 10 min à 60°C en présence d'urée 1 M et 3) avec une dénaturation thermique de 10 min à 60°C en présence d'urée 1 M et d'acétonitrile 20%. Une efficacité de protéolyse supérieure ou égale à 98.6 % ( $(1 - \frac{5}{7300}) \times 100$ ) a été obtenue avec la condition 3) et en utilisant un rapport enzyme:substrat élevé de [1:2] pour la trypsine et [1:10] pour l'Asp-N et la chymotrypsine. Nous estimons que cette efficacité de protéolyse est satisfaisante. En parallèle, le faible nombre de MC doit être vérifié avec la couverture de séquence.

### 5.3.2 Cartographie peptidique

La protéolyse complète d'une protéine génère un mélange de peptides. Généralement, la majorité de ces peptides a une masse comprise entre 500 et 3000 Da, ce qui correspond à la gamme classique de mesure de la plupart des spectromètres de masse. Toutefois, comme nous l'avons vu précédemment (cf § 1.2.3.1 de la section 2), tous ces peptides ne seront pas ionisés et donc détectés par spectrométrie de masse. Ainsi, lors des analyses classiques d'identification de protéines par empreintes peptidiques massiques ou LC-MS/MS, les protéines sont identifiées à partir de quelques peptides qui correspondent, dans le meilleur des cas, à environ 50 % de la séquence totale de la protéine. Pour l'identification de protéines, ce résultat peut être suffisant, mais il devient limitant pour caractériser de manière exhaustive les modifications d'une protéine par spectrométrie de masse. En effet, pour localiser et quantifier tous les sites modifiés d'une protéine, il apparaît judicieux d'avoir une couverture de séquence la plus proche possible de 100%.

La finalité de la cartographie peptidique d'une protéine est 1) d'analyser la couverture de séquence pour identifier d'éventuelles zones de la protéine qui ne sont jamais observées en spectrométrie de masse et pour lesquelles on sait qu'aucune information ne pourra être obtenue et 2) de mettre en évidence des MCs. Cette analyse concerne essentiellement les protéines purifiées et n'est généralement pas abordée pour des études globales sur des mélanges complexes.

Afin d'obtenir la couverture de séquence la plus complète possible nous avons utilisé trois endoprotéases : l'Asp-N, la trypsine et la chymotrypsine. Le choix de la protéase dépend de la séquence de chaque protéine et doit être établi après protéolyse théorique de la protéine étudiée. Le critère déterminant dans ce choix est la taille des peptides générés lors de la protéolyse théorique. Cette taille dépend directement du nombre et de la répartition le long de la séquence protéique des résidus cibles de la protéase utilisée. Un autre critère à considérer est le nombre de  $m/z$  proches qui génèrent des ambiguïtés d'annotation, qui ont lieu lorsque plusieurs  $m/z$  théoriques peuvent correspondre à un  $m/z$  expérimental mesuré. La Figure 104 présente la séquence de la protéine

hPEBP1. Les résidus en bleu correspondent aux sites de clivage par la trypsine, endoprotéase la plus utilisée en protéomique de par son faible coût, son efficacité et sa bonne stabilité. L'analyse de la séquence de hPEBP1 montre que les résidus lysine et arginine sont répartis régulièrement, ce qui permet de générer des peptides de taille adaptée à l'analyse MS (< 5000 Da). Le Tableau 14 présente le nombre de peptides et les ambiguïtés d'annotation obtenues après protéolyse théorique de hPEBP1 par différentes protéases. Très peu de risques d'ambiguïtés d'annotation sont observés : 0 pour la trypsine et l'Asp-N et seulement trois pour la chymotrypsine. Un exemple d'ambiguïté potentielle est visible pour les peptides [26-58] et [30-63] qui ont un  $m/z$  monoisotopique théorique de 3523.854 et 3523.875, respectivement, correspondant à une différence de 6 ppm. La précision de masse de nos instruments MALDI-TOF et ESI-IT ne permet pas de distinguer l'une ou l'autre des possibilités. L'utilisation de notre QTOF haute résolution-haute précision de masse (< 5 ppm), ou encore du couplage LC-MS, permet de circonvenir ces problèmes d'ambiguïtés d'annotation. D'une manière générale, ces ambiguïtés sont statistiquement plus fréquentes lorsque la taille de la protéine augmente et varient en fonction de la protéase utilisée. L'étape de réduction/alkylation des cystéines, en créant des valeurs supplémentaires de  $m/z$  théoriques, augmente également la fréquence de ces ambiguïtés.

**Figure 104. Séquence de hPEBP1.** La numérotation des résidus correspond à la protéine avec la méthionine initiale. Les résidus spécifiques d'une protéolyse par la trypsine sont indiqués en bleu.

2 PVDLS**K**WSPG LSLQEVDEQP QHPLHVITYAG AAVDELG**K**V**L** TPTQV**K****N**RPT  
52 SISWDGLDSG **K**LYTLVLTDP DAP**S**R**K**DP**K****Y** REWHHFLVV**N** **M**K**G**NDISSGT  
102 VLSDYVGS**G**P **P****K**GTGLH**R**YV WLVEYQD**R**PL **K**CDEPIL**S****N**R SG**D**H**R**G**K****F****K**V  
152 AS**F****R****K****K**YEL**R** APVAGTCYQA EWDDYVP**K**LY EQLS**G****K**

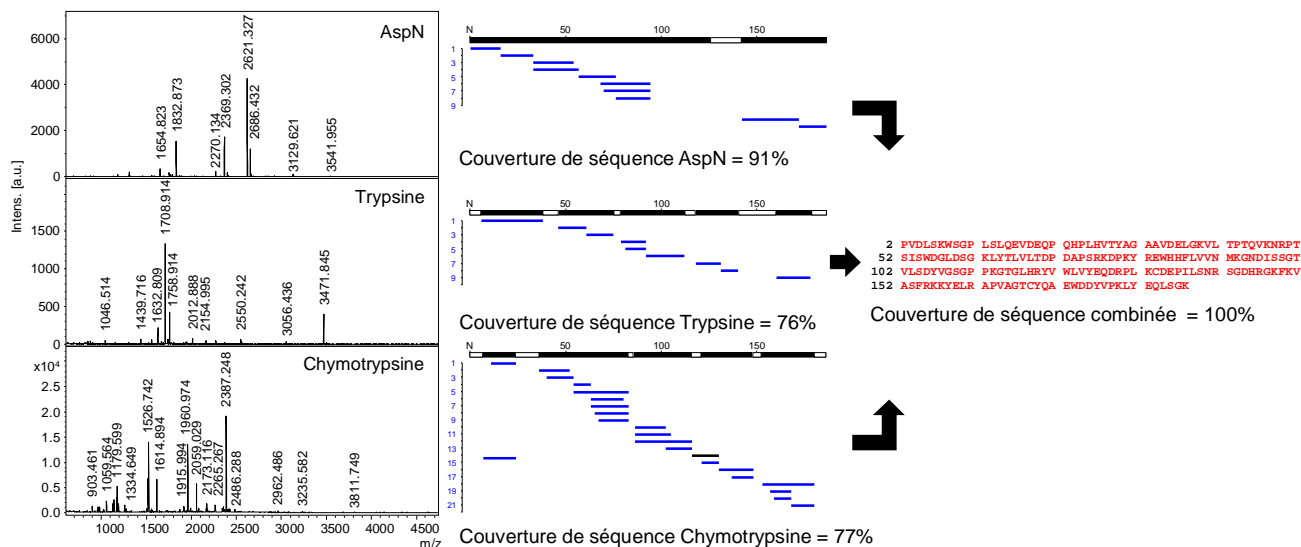
**Tableau 14. Nombre d'ambiguïtés d'annotation sur l'ensemble des peptides générés lors de protéolyses théoriques de hPEBP1#.**

Endoprotéases	Nbre d'ambiguïtés / nbre peptides théoriques
Asp-N (Nter D)	0/27 (673.362 – 4610.347)
Trypsine (Cter K et R)	0/33 (571.258 – 4338.240)
Chymotrypsine (Cter F, Y, W, I, L)	3/202 (519.256 – 4539.326)

#L'ambiguïté a été considérée lorsque deux  $m/z$  théoriques étaient différents de moins de 50 ppm. Un « missed cleavage » a été considéré pour la trypsine et l'Asp-N et cinq pour la chymotrypsine. Les peptides de  $m/z$  inférieurs à 500 n'ont pas été comptabilisés. Les chiffres entre parenthèses correspondent aux extrémités de la gamme de  $m/z$  des peptides théoriques. Les résidus cibles de chaque endoprotéase sont également illustrés.

Les protéolyses ont été réalisées en solution puis analysées par MALDI-TOF MS et nanoLC-MS. Les couvertures de séquences obtenues en MALDI-TOF sont de 91 % pour l'Asp-N, 76 % pour la trypsine et 77 % pour la chymotrypsine (Figure 105).

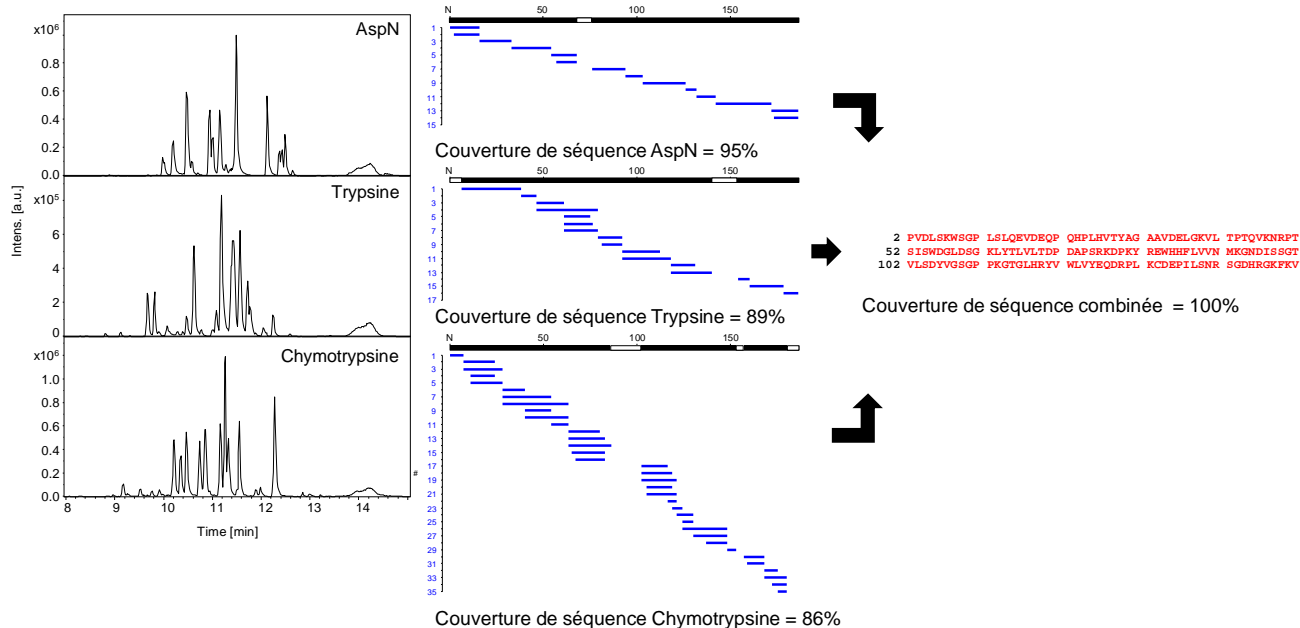
**Figure 105. Cartographie peptidique et couverture de séquence obtenues après protéolyse de hPEBP1 par l'Asp-N, la trypsine et la chymotrypsine et analyse par MALDI-TOF MS.** Pour chaque analyse, environ 200 fmol d'échantillon ont été analysés en mode réflectron positif en utilisant la matrice HCCA/TWA et la méthode en couche ultra-mince. Les pics annotés dans les spectres de gauche correspondent à des peptides représentés en bleu. Ces derniers permettent d'établir la couverture de séquence de la protéine, visualisée par les zones noires. Les zones blanches représentent les zones non observées. L'utilisation de trois protéases différentes permet d'obtenir 100% de couverture de séquence. Les recherches ont été effectuées avec une tolérance de 50 ppm et un nombre de missed cleavage de 2 pour l'Asp-N et la trypsine et 5 pour la chymotrypsine.



Nous pouvons observer que pour chaque protéolyse la distribution des peptides le long de la séquence est homogène, suggérant une absence de domaine récalcitrant à la protéolyse. Le cumul de ces trois couvertures de séquence permet d'atteindre une couverture totale de 100 %, ce qui est idéal pour une caractérisation complète de la protéine. En termes de MC, les protéolyses Asp-N et trypsine présentent des résultats satisfaisants, avec un faible nombre de MC observés. Il n'en va pas de même avec la chymotrypsine où quasiment chaque zone de la séquence est concernée par un MC. Cette différence peut s'expliquer par l'existence de sites secondaires pour la chymotrypsine (I, L) qui sont clivés moins fréquemment que les sites primaires (F, W, Y), et par la forte proximité des sites de clivage entre eux, générant ainsi une multitude de MCs.

Des résultats similaires ont été obtenus en nanoLC-HRMS avec des couvertures de séquence de 95 % pour l'Asp-N, 89 % pour la trypsine et 86 % pour la chymotrypsine (Figure 106).

**Figure 106. Cartographie peptidique et couverture de séquence obtenues après protéolyse de hPEBP1 par l'Asp-N, la trypsine et la chymotrypsine et analyse en nanoLC HRMS.** Pour chaque analyse, environ 200 fmol d'échantillon ont été analysés. Les pics du chromatogramme de gauche correspondent à des peptides représentés en bleu. Ces derniers permettent d'établir la couverture de séquence de la protéine, visualisée par les zones noires. Les zones blanches représentent les zones non observées. L'utilisation de trois protéases différentes permet d'obtenir 100% de couverture de séquence. Les recherches ont été effectuées avec une tolérance de 5 ppm et un nombre de missed cleavage de 2 pour l'Asp-N et la trypsine et 5 pour la chymotrypsine.



Pour chaque protéase, la couverture de séquence obtenue en nanoLC-ESI-HRMS est supérieure à celle obtenue en MALDI-TOF MS (par exemple, pour la chymotrypsine, 77% en MALDI-TOF MS vs 86% en nanoLC-HRMS). Ceci pourrait être expliqué par l'étape de fractionnement de l'échantillon en nanoLC qui permet d'analyser successivement les peptides, limitant ainsi les phénomènes d'interférence nuisibles à leur détection.

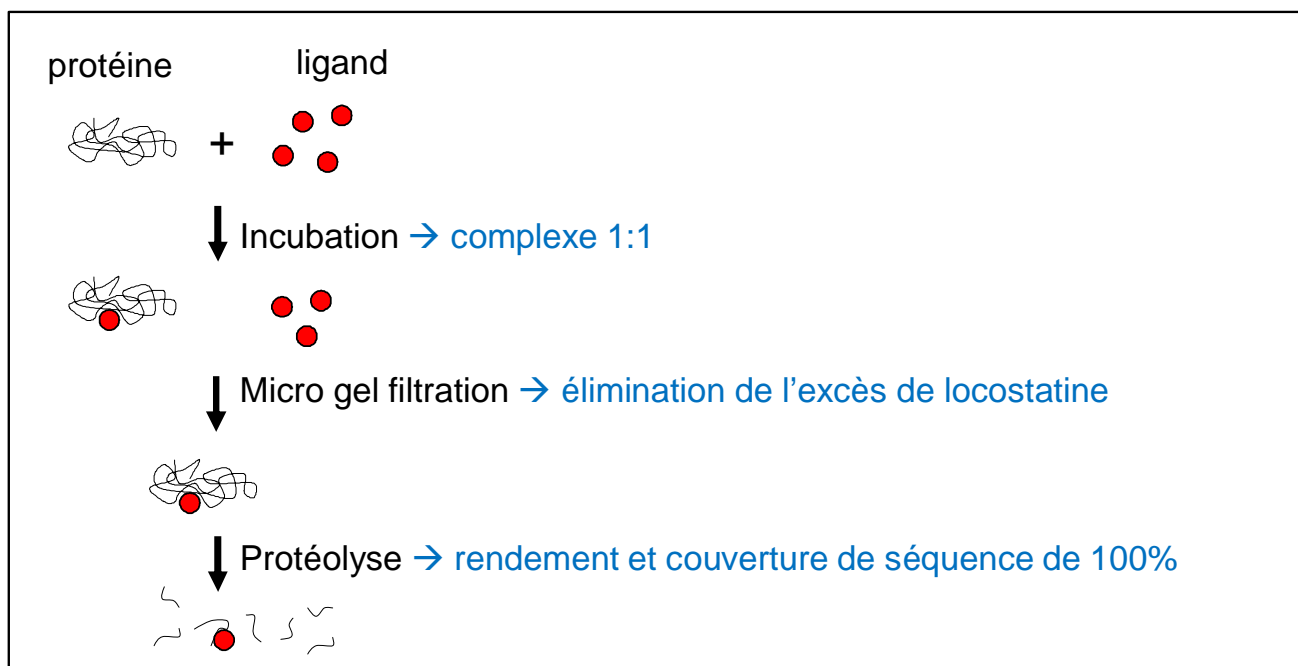
## 6. Essai de localisation du site spécifique - Approche qualitative

Dans l'objectif de localiser le site actif de hPEBP1 ciblé par la locostatine un protocole a été élaboré en respectant le cahier des charges suivant :

- 1) Contrôler la réaction hPEBP1-locostatine afin d'obtenir un complexe 1:1
- 2) Eliminer l'excès de ligand en fin de réaction et avant l'étape de protéolyse
- 3) Protéolyser le complexe dans des conditions permettant de se rapprocher d'une efficacité de protéolyse et d'une couverture de séquence de 100 % ; le tout en 2 h d'incubation afin de limiter la dégradation de la locostatine.

Le protocole rassemblant l'ensemble de ces critères est illustré Figure 107.

Figure 107. Protocole optimisé pour la localisation du site spécifique de hPEBP1 ciblé par la locostatine.



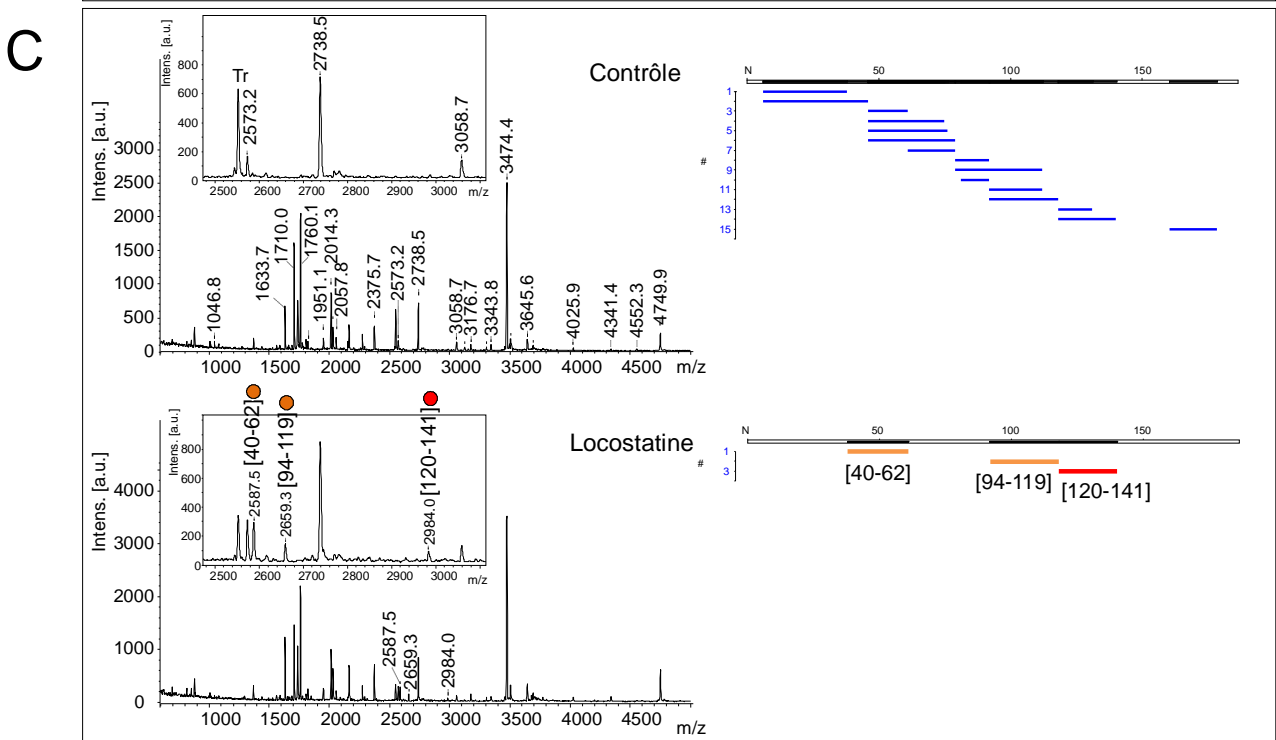
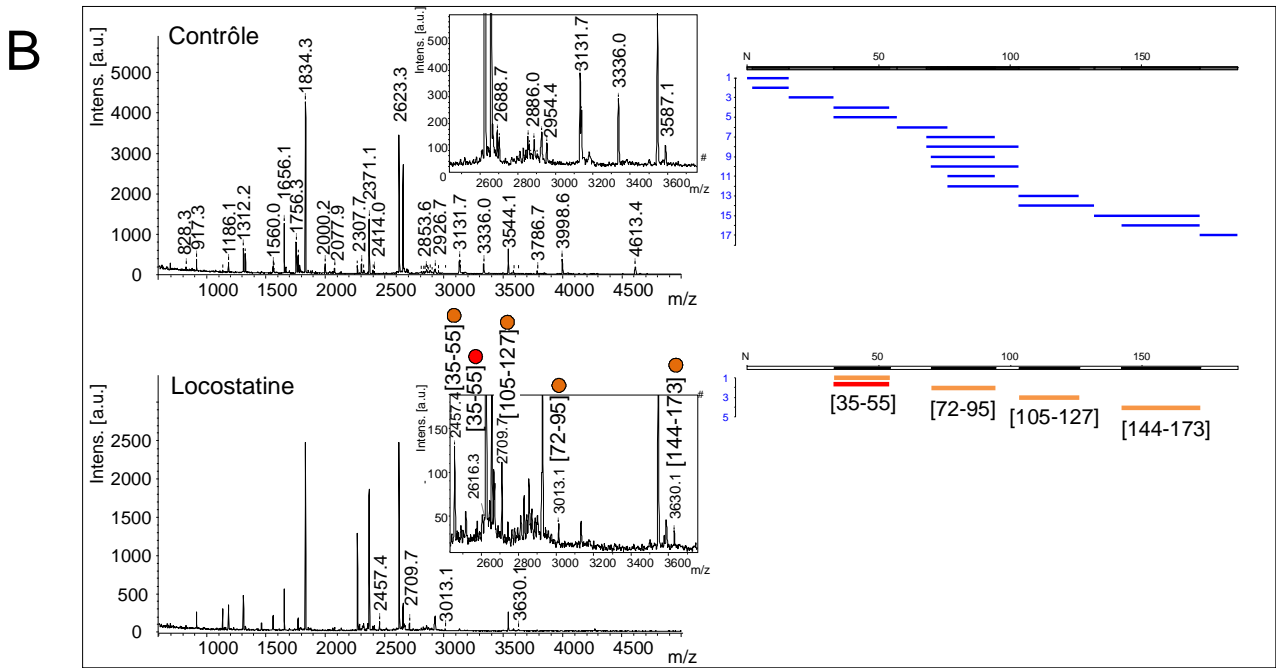
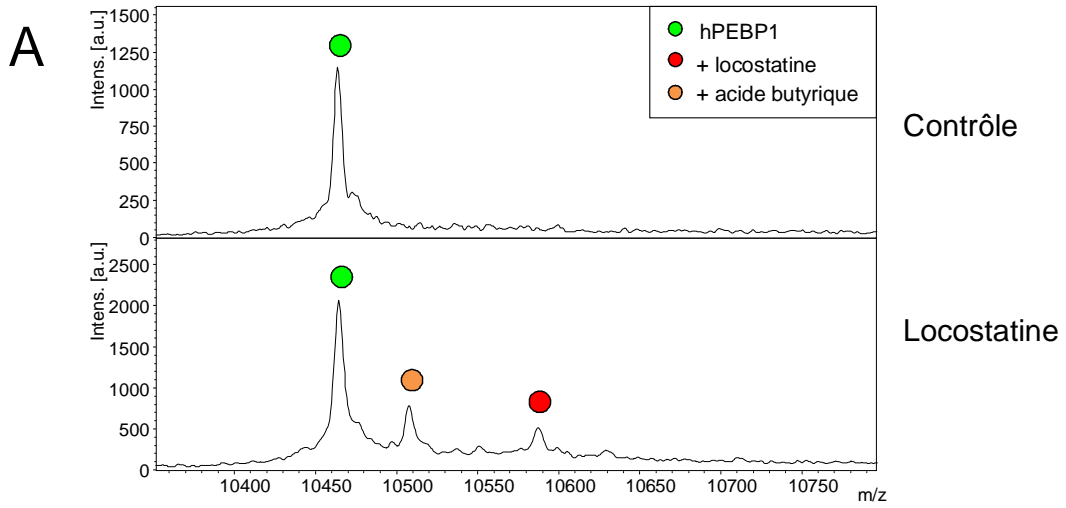
Un complexe hPEBP1-locostatine 1:1 a été obtenu en 4h30 d'incubation dans un tampon HEPES 35 mM, pH 7.8 (Figure 108-A). Après élimination de l'excès de locostatine, deux protéolyses ont été effectuées : par l'Asp-N et la trypsine.

## 6.1 Analyse MS des peptides

Les peptides issus des protéolyses par l'Asp-N et la trypsine ont été analysés en MALDI-TOF MS (Figures 108-B et C).

**Figure 108. Peptides modifiés observés après protéolyse du complexe hPEBP1-locostatine par l'Asp-N et la trypsine.** A, Une solution de protéine hPEBP1 à 6.8  $\mu\text{M}$  a été incubée avec 1 mM de locostatine pendant 4h30 à 37°C dans un tampon HEPES 35 mM, pH 7.8 contenant 5% d'acétonitrile. La réaction a été arrêtée par micro gel filtration puis analysée en MALDI-TOF MS après une dilution de l'échantillon au 1/10<sup>ème</sup> dans HCCA/FWI, dépôt en couche ultra-mince et lavage des dépôts avec du TFA 0.1%. B, Analyses MALDI-TOF en mode linéaire positif du protéolysat Asp-N de la réaction après dilution au 1/10<sup>ème</sup> dans HCCA/TWA, dépôt sur cible MALDI par la méthode en couche ultra-mince et lavage des dépôts avec du TFA 0.1%. Les peptides non modifiés sont en bleu et les peptides modifiés par la locostatine ou l'acide butyrique sont indiqués en rouge ou orange, respectivement. C, Analyses MALDI-TOF en mode linéaire positif du protéolysat trypsine de la réaction après dilution au 1/10<sup>ème</sup> dans HCCA/TWA, dépôt sur cible MALDI par la méthode en couche ultra-mince et lavage des dépôts avec du TFA 0.1%. Les peptides non modifiés sont en bleu et les peptides modifiés par la locostatine ou l'acide butyrique sont indiqués en rouge ou orange, respectivement.





Les résultats indiquent que plusieurs peptides sont modifiés. Ces derniers sont listés dans le Tableau 15.

**Tableau 15. Liste des peptides modifiés observés après protéolyse du complexe hPEBP1-locostatine par l'Asp-N et la trypsine.**

Séquence	Position [début-fin]	Modification	Nombre de missed cleavage <sup>a</sup>
Asp-N			
DELGKVLTPQTQVKNRPTSISW	[35-55]	locostatine	0
DELGKVLTPQTQVKNRPTSISW	[35-55]	acide butyrique	0
DAPSRKDPKYREWHHFLVVMKGN	[72-95]	acide butyrique	1
DYVGSPPKGTGLHRYVWLVYEQ	[105-127]	acide butyrique	0
DHRGKFKVASFRKKYELRAPVAGTCYQAEW	[144-173]	acide butyrique	0
Trypsine			
VLTPQTQVKNRPTSISWDGLDSGK	[40-62]	acide butyrique	1
GNDISSGTVLSDYVGSPPKGTGLHR	[94-119]	acide butyrique	1
YVWLVYEQDRPLKCDEPILSNR	[120-141]	locostatine	1

<sup>a</sup> Un missed cleavage généré par la présence d'une proline en P1' de K ou R est un phénomène bien décrit et n'a pas été considéré

En considérant les peptides chevauchants entre les deux protéases, c'est-à-dire les peptides qui partagent une séquence commune, cinq zones modifiées peuvent être définies :

- la zone [40-55] correspondant au chevauchement des peptides [35-55] et [40-62] observés modifiés par l'Asp-N et la trypsine respectivement,
- le peptide [72-95],
- la zone [105-119] correspondant aux peptides chevauchant [105-127] et [94-119] observés modifiés par l'Asp-N et la trypsine respectivement,
- le peptide [120-141], et
- le peptide [144-173].

L'observation de cinq zones modifiées indique qu'il y a, au minimum, cinq résidus modifiés (1 par zone), sous réserve de l'absence d'isobares. En considérant les résidus modifiables C, K, H, Y, S et T, et en faisant l'hypothèse qu'un résidu modifié ne peut être clivé par une protéase, les missed-cleavages associés aux peptides tryptique [40-62], [94-119] et [120-141] nous font suspecter les résidus Lys47, Lys113 et Lys132, respectivement comme étant modifiés. Ceci devra être confirmé par une analyse MS/MS des peptides. La zone [72-95] contient également un missed-cleavage mais c'est un résidu Asp, donc non-modifiable.

## 6.2 Analyse MS/MS des peptides

Afin de localiser définitivement et sans ambiguïté le résidu modifié, les peptides modifiés ont été fragmentés en MS/MS. Toutefois, leur rapport signal sur bruit étant trop faible pour une analyse en MALDI-TOF/TOF LIFT, ils ont été fragmentés lors d'une analyse nanoLC-IT MS/MS CID avec liste d'inclusion. Le protéolysat trypsine a été utilisé dans un premier temps car, d'une manière générale, les peptides tryptiques, du fait de leur résidu protonable Lys ou Arg en C-terminal, ont une meilleure efficacité de fragmentation que les peptides Asp-N ou chymotrypsine qui, soit sont monochargés, soit contiennent un ou plusieurs résidus protonables à l'intérieur de la séquence du peptide, ce qui est néfaste pour la fragmentation CID. La protéolyse a été effectuée sur des échantillons de protéine peu modifiée, donc dans des conditions de "single-hit" (une modification par chaîne de protéine). S'il existe malgré tout des peptides isobares, l'étape de nano-ultraHPLC en amont de la MS doit permettre de bien les séparer dans ce mélange peu complexe. En absence d'ions de fragmentation à la fois modifiés et non-modifiés indicatifs de la présence d'isobares, nous avons donc fait l'hypothèse que tous les spectres interprétés correspondent à des espèces uniques.

La Figure 109 présente les spectres MS/MS CID des peptides [40-62]-acide butyrique (Figure 109-A), [94-119]-acide butyrique (Figure 109-B) et [120-141]-locostatine (Figure 109-C).

Le peptide [40-62]-acide butyrique contient sept sites modifiables : Thr42, Thr44, Lys47, Thr51, Ser52, Ser54 et Ser60 (la Lys62 n'est pas modifiée puisqu'elle a été clivée par la trypsine) (Figure 109-A). La présence des ions  $y_3$  à  $y_{15}$  et  $b_3$  non modifiés permet d'exclure la modification des résidus Thr42, Thr51, Ser52, Ser54 et Ser60. L'observation des ions  $y_{16}$  à  $y_{18}$  portant un groupement butyrate indique que c'est la Lys47 qui est modifiée. L'ion  $b_9$  modifié détecté conforte cette annotation.

Le peptide [94-119]-acide butyrique contient neuf résidus modifiables : Ser98, Ser99, Thr101, Ser104, Tyr106, Ser109, Lys113, Thr115 et His118 (Figure 109-B). La détection des ions  $b_3$  à  $b_{18}$  non modifiés montre que les résidus Ser98, Ser99, Thr101, Ser104, Tyr106, Ser109 ne portent pas la modification. De la même manière, les ions  $y_3$  et  $y_6$  non modifiés indiquent que Thr115 et His118 ne sont pas modifiés. La modification de Lys113 est supportée par la présence des ions  $b_{21}$  et  $y_7$  modifiés par l'acide butyrique.

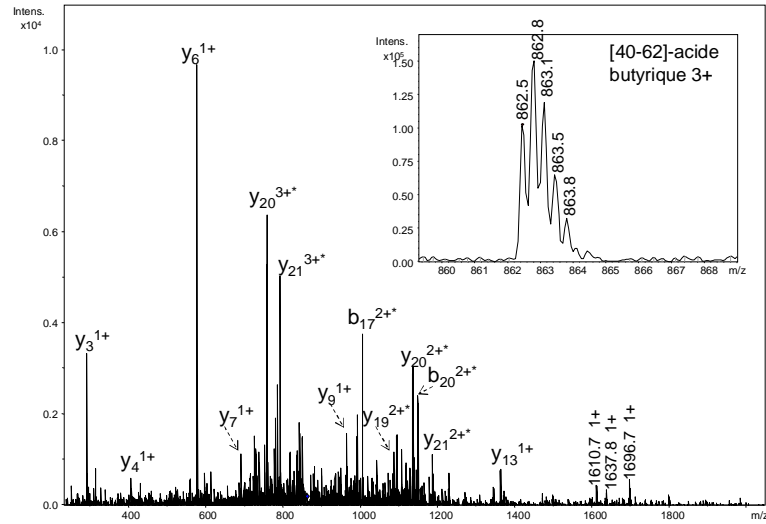
Le peptide [120-141]-locostatine contient cinq résidus modifiables : Tyr120, Tyr125, Lys132, Cys133 et Ser139 (Figure 109-C). Une forte fragmentation se produit entre les résidus D134-E135 et E135-P136, avec production d'ions internes détectés comme étant modifiés en N-terminal de cette zone. L'observation des ions  $y_3$  à  $y_7$  et  $b_6$  à  $b_{12}$  non modifiés montre l'absence de modification au niveau des résidus Tyr120, Tyr125 et Ser139. Les ions  $y_{10}$  à  $y_{16}$  et  $b_{15}$  modifiés par la locostatine

permettent de proposer Lys132 ou Cys133 comme étant modifiée. La présence de l'ion  $y_9$  modifié par la locostatine permet de lever l'ambiguïté et révèle que la locostatine est localisée au niveau de Cys133. Ceci signifie qu'une modification en  $P_1'$  est suffisante pour prévenir la coupure par la trypsine. Un réexamen des données précédentes tenant compte de cette information n'a pas invalidé nos conclusions. Si l'on fait l'hypothèse que, réciproquement, la modification en  $P_1$  empêche la coupure par l'Asp-N, on peut suggérer la Lys77 comme résidu modifié en position proximale d'un site de missed-cleavage (Asp78) au vu des données MS.

**Figure 109. Analyse nanoLC-IT MS/MS CID du protéolysat trypsique.** Une liste d'inclusion contenant les  $m/z$  d'intérêt à fragmenter a été utilisée A, Spectre CID du peptide [40-62]-acide butyrique. L'ion précurseur de  $m/z$  862.5 est trichargé (encadré). B, Spectre CID du peptide [94-119]-acide butyrique. L'ion précurseur de  $m/z$  886.4 est trichargé (encadré). C, Spectre CID du peptide [120-141]-locostatine. L'ion précurseur de  $m/z$  994.5 est trichargé (encadré). Plusieurs ions internes sont annotés. Les ions fragments modifiés par l'acide butyrique ou la locostatine sont représentés par une étoile.

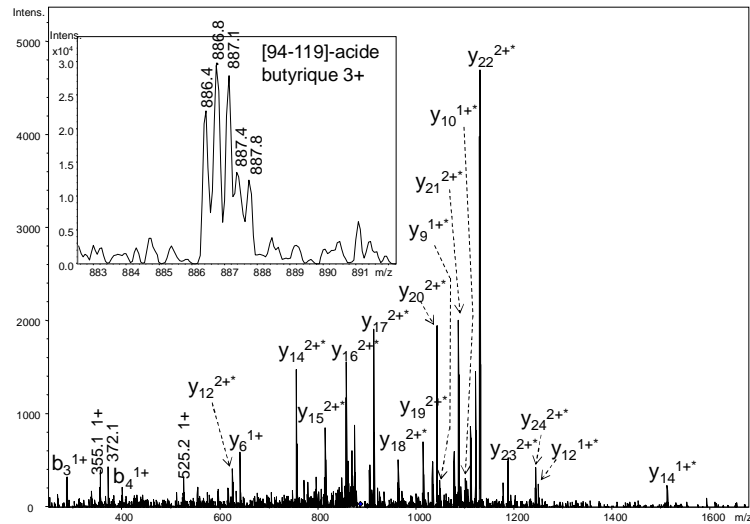
VLTPPTQVK\*NRPTSISWDGLDSGK

A



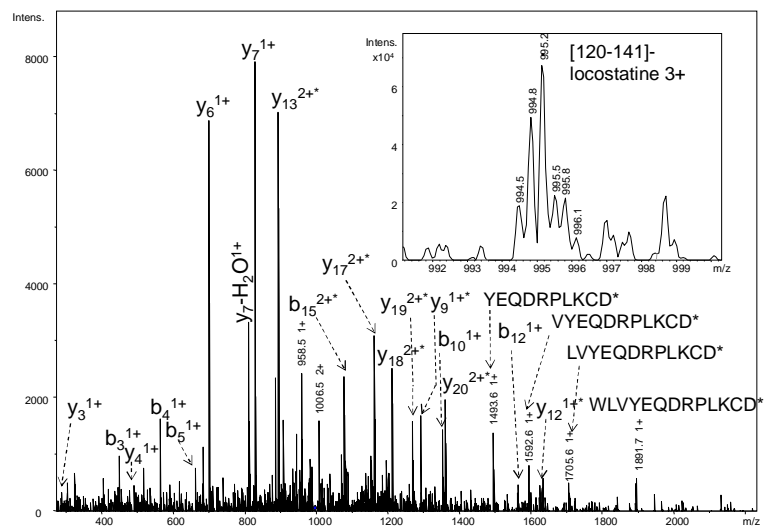
GNDISSGTVLSDYVGSPPK\*GTGLHR

B



YVWLVEQDRPLKC\*DEPILSNR

C



En résumé, l'approche qualitative permet d'identifier les résidus cibles suivants:

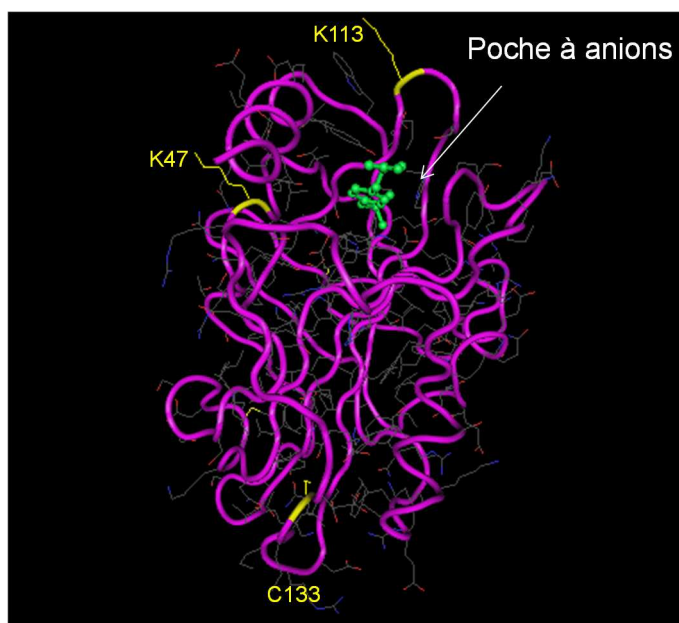
Lys47, Lys113 et Cys133. Les peptides observés modifiés en MALDI-TOF MS: [72-95] et [144-173] n'ont pas été analysés en MS/MS pour le moment.

L'observation de plusieurs sites modifiés issus de la protéolyse d'un complexe hPEBP1-locostatine 1:1 peut être expliquée par trois hypothèses :

- 1) Présence d'une "poche" spécifique de liaison de la locostatine avec attaque de plusieurs résidus nucléophiles qui sont proches dans la structure tridimensionnelle de hPEBP1. Ce type d'attaque traduirait une certaine flexibilité de conformation de la locostatine dans la poche de liaison ("wobble")
- 2) Présence de plusieurs sites spécifiques
- 3) Présence de n sites non spécifiques sur la protéine entière, en plus du site spécifique

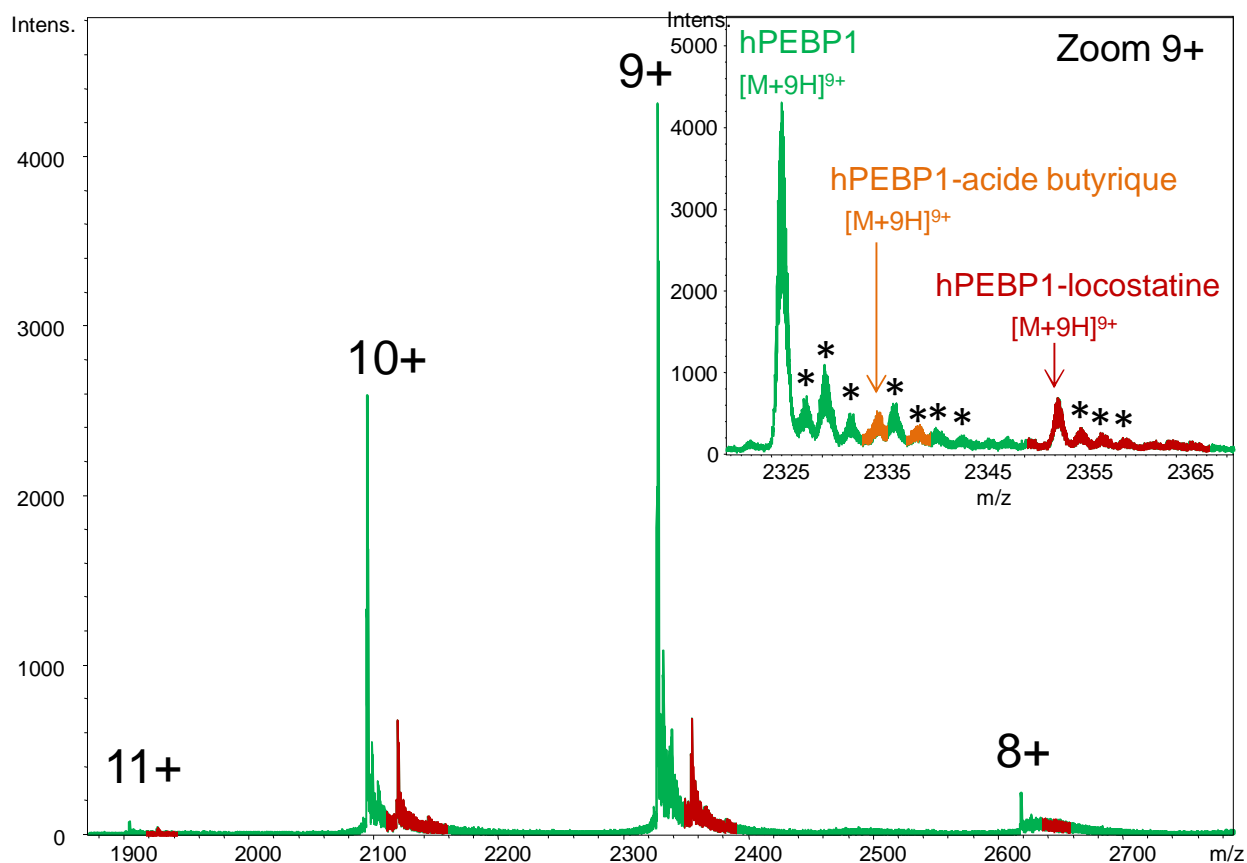
D'après la Figure 110, les sites modifiés suspectés Lys47, Lys113 et Cys133 sont éloignés voire à l'opposé dans la structure, ce qui n'est pas en faveur de l'hypothèse 1). Bien que la structure illustrée corresponde au complexe de hPEBP1 avec la phosphotyrosine, pour que les résidus modifiés des zones [72-95], [144-173] et les trois résidus Lys47, Lys113 et Cys133 se retrouvent au niveau d'une zone commune dans la structure, il faudrait que la fixation de la locostatine sur hPEBP1 en une zone unique entraîne une réorganisation structurale très importante de hPEBP1.

**Figure 110. Localisation spatiale des résidus Lys47, Lys113 et Cys133 de hPEBP1.** Le modèle structural 2QYQ de la PDB (Protein Data Bank) représentant hPEBP1 en complexe avec la O-phosphotyrosine est illustré.



Cette hypothèse est considérée comme très peu probable, notamment à la vue des distributions d'états de charges qui sont très similaires entre hPEBP1 seule et le complexe hPEBP1-locostatine covalent en conditions natives (Figure 111).

**Figure 111. Analyse HRMS en conditions natives du complexe hPEBP1-locostatine covalent.** Une solution de protéine hPEBP1 à 12.2  $\mu\text{M}$  a été incubée avec 1.2 mM de locostatine pendant 1h à 37°C dans un tampon acétate d'ammonium-ABC 100 mM, pH 7.7 contenant 5% d'acétonitrile. La réaction a été arrêtée par micro gel filtration puis analysée en HRMS dans une solution d'acétate d'ammonium-ABC 100 mM, pH 7.7 en infusion. Les spectres vert, orange et rouge représentent hPEBP1 non modifiée, le complexe hPEBP1-acide butyrique et le complexe hPEBP1-locostatine, respectivement. Les astérisques indiquent des adduits de  $\text{Na}^+$  et  $\text{K}^+$ .



L'hypothèse 2) force, dans le cas présent, à envisager plus de deux sites spécifiques. Elle semble donc difficile à appliquer à une protéine de ~21 kDa. L'hypothèse 3) a donc été privilégiée et nous avons considéré qu'en plus du site spécifique, nous étions en présence de réactions non spécifiques sur la protéine entière, et ce malgré l'arrêt de la réaction à une stœchiométrie globale de 1:1. L'approche qualitative ne permet cependant pas de classer les sites les uns par rapport aux autres.

### 6.3 Limites de l'approche qualitative

L'interaction *in vitro* entre hPEBP1 et la locostatine révèle la présence de plusieurs sites modifiés correspondant très certainement à un mélange du site spécifique avec plusieurs sites non spécifiques. Pour le moment, nous ne savons pas si ce phénomène se produit également *in vivo*.

La discrimination d'un site spécifique parmi plusieurs sites non spécifiques n'est pas triviale. Tout d'abord, rappelons qu'un site spécifique est un site corrélé à l'activité biologique. Ici, la modification de hPEBP1 par la locostatine inhibe la migration cellulaire en renforçant un effet naturel de hPEBP1. Pour démontrer qu'un site est associé à une activité biologique, l'approche ultime consiste à muter le résidu en question par mutagenèse dirigée et à observer la perte d'activité. Cependant, en présence de multiples sites potentiels, comme c'est le cas pour le système hPEBP1-locostatine, la mutation site par site n'est pas pertinente car trop fastidieuse et coûteuse en temps. Avant de se lancer dans la mutagenèse dirigée de hPEBP1, il apparaît ainsi judicieux de développer une approche permettant de discriminer *in vitro* le spécifique du non spécifique.

L'approche qualitative utilisée pour mettre en évidence tous les sites modifiés de hPEBP1 par la locostatine ne fait qu'inventorier tous les sites mais ne peut distinguer un site spécifique parmi plusieurs sites non spécifiques. Dans l'hypothèse rationnelle où le site spécifique est le site majoritaire, seule une approche quantitative capable de quantifier tous les sites modifiés permettrait d'identifier le site le majoritaire et donc le site cible de la locostatine. Une telle approche nécessite cependant de planifier les conditions expérimentales afin de pouvoir distinguer le site spécifique des sites non-spécifique.



## 7. Modélisation mathématique pour discriminer le site spécifique des sites non spécifiques

### 7.1 Objectif du modèle

L'objectif de la modélisation de l'association covalente d'un ligand sur une protéine est d'évaluer la contribution des différents sites, spécifique et non spécifique, à la fraction modifiée globale. Ceci permet d'identifier les paramètres qui influencent cette contribution et de déterminer la fraction modifiée globale optimale. Celle-ci correspond au meilleur compromis entre

1. une contribution élevée du site spécifique, et
2. un rapport signal sur bruit (défini comme le signal lié à la modification au niveau de la protéine entière ou des peptides) suffisant pour que l'espèce modifiée d'intérêt soit détectée en spectrométrie de masse.

L'association covalente d'un ligand avec une protéine possédant  $1 + n$  sites de liaison indépendants (1 site spécifique et  $n$  sites non spécifiques) dépend des paramètres décrivant le système d'un point de vue cinétique (constante(s) de vitesse d'association  $k_{on}$ ), et stœchiométrique (nombre de sites non spécifiques  $n$ ). Pour quantifier les sites spécifiques et non spécifiques, il est nécessaire de déterminer 1) les conditions expérimentales, concentrations initiales en ligand  $[L]_0$  et en protéine  $[P]_0$ , ratio  $[L]_0/[P]_0$ , temps de réaction ou fraction modifiée globale optimal(e) permettant de discriminer au mieux le site spécifique (majoritaire) des sites non spécifiques (minoritaires), et 2) un critère objectif utilisé pour dire qu'un site est majoritaire.

## 7.2 Description mathématique du modèle

Le modèle présenté ici décrit l'association d'un ligand à l'échelle des sites de la protéine et non pas à l'échelle de la protéine prise comme une entité simple, c'est-à-dire que l'on va quantifier la fraction stœchiométrique globale de modification de la protéine et non pas la fraction globale de protéine modifiée (cf paragraphe 2.3.1 de la section 2).

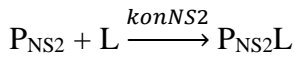
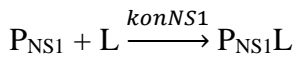
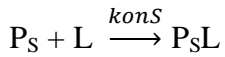
- Description du système:

- Réaction covalente en une étape
- Sites indépendants
- Excès de ligand par rapport à la protéine
- 1 site spécifique → site majoritaire
- n sites non spécifiques → sites minoritaires

### Définitions

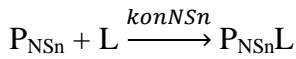
- $P_S$  : site spécifique libre
- $P_{NS1}$  : site non spécifique 1 libre
- $P_{NS2}$  : site non spécifique 2 libre
- $P_{NSn}$  : site non spécifique n libre
- L : ligand libre
- $k_{onS}$  : constante de vitesse d'association spécifique de 2<sup>ème</sup> ordre ( $M^{-1} s^{-1}$ )
- $k_{onNS1}$  : constante de vitesse d'association non spécifique 1 de 2<sup>ème</sup> ordre ( $M^{-1} s^{-1}$ )
- $k_{onNS2}$  : constante de vitesse d'association non spécifique 2 de 2<sup>ème</sup> ordre ( $M^{-1} s^{-1}$ )
- $k_{onNSn}$  : constante de vitesse d'association non spécifique n de 2<sup>ème</sup> ordre ( $M^{-1} s^{-1}$ )
- $P_S L$ : complexe ligand-site spécifique de la protéine
- $P_{NS1} L$  : complexe ligand-site non-spécifique 1 de la protéine
- $P_{NS2} L$  : complexe ligand-site non-spécifique 2 de la protéine
- $P_{NSn} L$  : complexe ligand-site non-spécifique n de la protéine

- L'association du ligand en excès sur 1 + n sites indépendants de la protéine est représentée par:



.

.



On définit les concentrations molaires:

$[P_S]_t$  : concentration du site spécifique libre à l'instant t

$[P_{NS1}]_t$  : concentration du site non spécifique 1 libre à l'instant t

$[P_{NS2}]_t$  : concentration du site non spécifique 2 libre à l'instant t

$[P_{NSn}]_t$  : concentration du site non spécifique n libre à l'instant t

$[L]_t$  : concentration de ligand libre à l'instant t

On peut alors calculer la vitesse de production instantanée en complexe spécifique  $P_S L$  et en complexes non spécifique  $P_{NS1} L$ ,  $P_{NS2} L$ , ...,  $P_{NSn} L$  suivant les équations 1 à 4:

$$V_S = \frac{d[P_S L]_t}{dt} = - \frac{d[P_S]_t}{dt} = k_S \cdot [P_S]_t \cdot [L]_t \quad (7)$$

$$V_{NS1} = \frac{d[P_{NS1} L]_t}{dt} = - \frac{d[P_{NS1}]_t}{dt} = k_{NS1} \cdot [P_{NS1}]_t \cdot [L]_t \quad (8)$$

$$V_{NS2} = \frac{d[P_{NS2} L]_t}{dt} = - \frac{d[P_{NS2}]_t}{dt} = k_{NS2} \cdot [P_{NS2}]_t \cdot [L]_t \quad (9)$$

.

.

$$V_{NSn} = \frac{d[P_{NSn} L]_t}{dt} = - \frac{d[P_{NSn}]_t}{dt} = k_{NSn} \cdot [P_{NSn}]_t \cdot [L]_t \quad (10)$$

Sachant qu'au temps  $t = 0$ , on est aux conditions initiales avec

$[L]_0$  : concentration initiale (= totale) de ligand

$[P_S]_0$  : concentration initiale en site spécifique

$[P_{NS1}]_0$  : concentration initiale en site non spécifique 1

$[P_{NS2}]_0$  : concentration initiale en site non spécifique 2

$[P_{NSn}]_0$  : concentration initiale en site non spécifique n,

si l'on se place dans des conditions où  $[L]_0/[P_S]_0 > 10$ ,  $[L]_0/[P_{NS1}]_0 > 10$ ,  $[L]_0/[P_{NS2}]_0 > 10$ ,  $[L]_0/[P_{NSn}]_0 > 10$ , en s'assurant que  $[L]_0/\Sigma[P_S + P_{NS1} + P_{NS2} + \dots + P_{NSn}]_0 > 10$ , alors on a  $[L]_t \sim [L]_0$  (11)

ce qui signifie que l'on est en condition de pseudo 1<sup>er</sup> ordre.

Après intégration, on obtient l'équation de calcul de  $[P_S]_t$ ,  $[P_{NS1}]_t$ ,  $[P_{NS2}]_t$  et  $[P_{NSn}]_t$ , c'est-à-dire les concentrations consommées à l'instant t en site spécifique libre  $P_S$ , et en sites non spécifiques libres de 1 à n,  $P_{NS1}$  à  $P_{NSn}$ :

$$[P_S]_t = [P_S]_0 \cdot e^{-k_{obs(S)} \cdot t} \quad (12)$$

$$\text{avec } k_{obs(S)} = k_{onS} \cdot [L]_0 \quad (13)$$

$$[P_{NS1}]_t = [P_{NS1}]_0 \cdot e^{-k_{obs(NS1)} \cdot t} \quad (14)$$

$$\text{avec } k_{obs(NS1)} = k_{onNS1} \cdot [L]_0 \quad (15)$$

·  
·

$$[P_{NSn}]_t = [P_{NSn}]_0 \cdot e^{-k_{obs(NSn)} \cdot t} \quad (16)$$

$$\text{avec } k_{obs(NSn)} = k_{onNSn} \cdot [L]_0 \quad (17)$$

En considérant les équations de conservation de la matière

$$[P_S]_0 = [P_S]_t + [P_S L]_t \quad (18)$$

$$[P_{NS1}]_0 = [P_{NS1}]_t + [P_{NS1} L]_t \quad (19)$$

·  
·

$$[P_{NSn}]_0 = [P_{NSn}]_t + [P_{NSn} L]_t \quad (20)$$

avec

$[P_S L]_t$  : concentration en complexe ligand-site spécifique à l'instant t (M)

$[P_{NS1} L]_t$  : concentration en complexe ligand-site non spécifique 1 à l'instant t (M)

·  
·

$[P_{NSn} L]_t$  : concentration en complexe ligand-site non spécifique n à l'instant t (M)

on obtient l'équation de calcul de  $[P_S L]_t$ ,  $[P_{NS1} L]_t$ , à  $[P_{NSn} L]_t$  les concentrations en complexe  $P_S L$ ,  $P_{NS1} L$ ,  $P_{NS2} L$  et  $P_{NSn} L$  accumulées à l'instant t:

$$[P_S L]_t = [P_S]_0 (1 - e^{-k_{obs(S)} \cdot t}) \quad (21)$$

$$[P_{NS1} L]_t = [P_{NS1}]_0 (1 - e^{-k_{obs(NS1)} \cdot t}) \quad (22)$$

.

.

$$[P_{NSn} L]_t = [P_{NSn}]_0 (1 - e^{-k_{obs(NSn)} \cdot t}) \quad (23)$$

Les fractions de liaisons  $f_{bound\ S}$  et  $f_{bound\ NS1, \dots, NSn}$  s'écrivent:

$$\frac{[P_S L]_t}{[P_S]_0} = f_{bound\ S} = 1 - e^{-k_{onS} \cdot [L]_0 \cdot t} \quad (24)$$

$$\frac{[P_{NS1} L]_t}{[P_{NS1}]_0} = f_{bound\ NS1} = 1 - e^{-k_{onNS1} \cdot [L]_0 \cdot t} \quad (25)$$

$$\frac{[P_{NSn} L]_t}{[P_{NSn}]_0} = f_{bound\ NSn} = 1 - e^{-k_{onNSn} \cdot [L]_0 \cdot t} \quad (26)$$

- Du point de vue de la stœchiométrie de liaison à la protéine

Le nombre de sites de liaison visibles au ligand (c'est-à-dire la stœchiométrie globale de modification) découle de :

$$[sites\ P_{liés}]_{\infty} = [P_S L]_{\infty} + [P_{NS1} L]_{\infty} + [P_{NS2} L]_{\infty} + \dots + [P_{NSn} L]_{\infty} \quad (27)$$

stœchiométrie : nombre de termes à droite

ici stœchiométrie = n+1.

- Du point de vue de la cinétique sur la protéine entière

La production de sites occupés par le ligand dans un système à n+1 sites est donc :

$$[sites\ P_{liés}]_t = [P_S L]_t + [P_{NS1} L]_t + [P_{NS2} L]_t + \dots + [P_{NSn} L]_t \quad (28)$$

$$[sites\ P_{liés}]_t = [P_S]_0 (1 - e^{-k_{obs(S)} \cdot t}) + [P_{NS1}]_0 (1 - e^{-k_{obs(NS1)} \cdot t}) + [P_{NS2}]_0 (1 - e^{-k_{obs(NS2)} \cdot t}) + \dots + [P_{NSn}]_0 (1 - e^{-k_{obs(NSn)} \cdot t}) \quad (29)$$

Si on pose  $[P]_0$ , la concentration initiale en protéine,

et que l'on raisonne en fraction de liaison totale ( $f_{bound\ sites\ total}$ ) par rapport à  $[P]_0$ , (division par  $[P]_0$  des n termes),

alors on a :

$$(f_{bound\ sites\ total}) = \frac{[P_S]_0}{[P]_0} \cdot (1 - e^{-k_{obs(S)} \cdot t}) + \frac{[P_{NS1}]_0}{[P]_0} (1 - e^{-k_{obs(NS1)} \cdot t}) + \frac{[P_{NS2}]_0}{[P]_0} (1 - e^{-k_{obs(NS2)} \cdot t}) + \dots + \frac{[P_{NSn}]_0}{[P]_0} (1 - e^{-k_{obs(NSn)} \cdot t}) \quad (30)$$

En considérant que  $[P]_0 = [P_S]_0 = [P_{NS1}]_0 = [P_{NS2}]_0 = [P_{NSn}]_0$  on a :

$$f_{\text{bound sites total}} = (1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t}) + (1 - e^{-k_{\text{onNS1}} \cdot [L]_0 \cdot t}) + (1 - e^{-k_{\text{onNS2}} \cdot [L]_0 \cdot t}) + \dots + (1 - e^{-k_{\text{onNSn}} \cdot [L]_0 \cdot t}) \quad (31)$$

A  $t = \infty$ ,

$$(f_{\text{bound sites total}}) = \frac{[P_S]_0}{[P]_0} + \frac{[P_{NS1}]_0}{[P]_0} + \frac{[P_{NS2}]_0}{[P]_0} + \dots + \frac{[P_{NSn}]_0}{[P]_0} = \frac{[P_S L]_{\infty} + [P_{NS1} L]_{\infty} + [P_{NS2} L]_{\infty} + \dots + [P_{NSn} L]_{\infty}}{[P]_0} \quad (32)$$

En d'autres termes, la fraction de liaison totale ( $f_{\text{bound sites total}}$ ) tend vers la fraction stœchiométrique globale de modification.

La contribution du site spécifique à la fraction de modification globale correspond à :

$$f_{\text{bound S}} / f_{\text{bound sites total}} = \frac{(1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t})}{(1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t}) + (1 - e^{-k_{\text{onNS1}} \cdot [L]_0 \cdot t}) + (1 - e^{-k_{\text{onNS2}} \cdot [L]_0 \cdot t}) + \dots + (1 - e^{-k_{\text{onNSn}} \cdot [L]_0 \cdot t})} \quad (33)$$

Dans le cas particulier où les constantes de vitesses non spécifiques sont identiques,  $k_{\text{onNS1}} = k_{\text{onNS2}} = k_{\text{onNSn}}$ , on a:

$$f_{\text{bound sites total}} = (1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t}) + n (1 - e^{-k_{\text{onNS}} \cdot [L]_0 \cdot t}) \quad (34)$$

$$f_{\text{bound S}} / f_{\text{bound sites total}} = \frac{(1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t})}{(1 - e^{-k_{\text{onS}} \cdot [L]_0 \cdot t}) + n (1 - e^{-k_{\text{onNS}} \cdot [L]_0 \cdot t})} \quad (35)$$

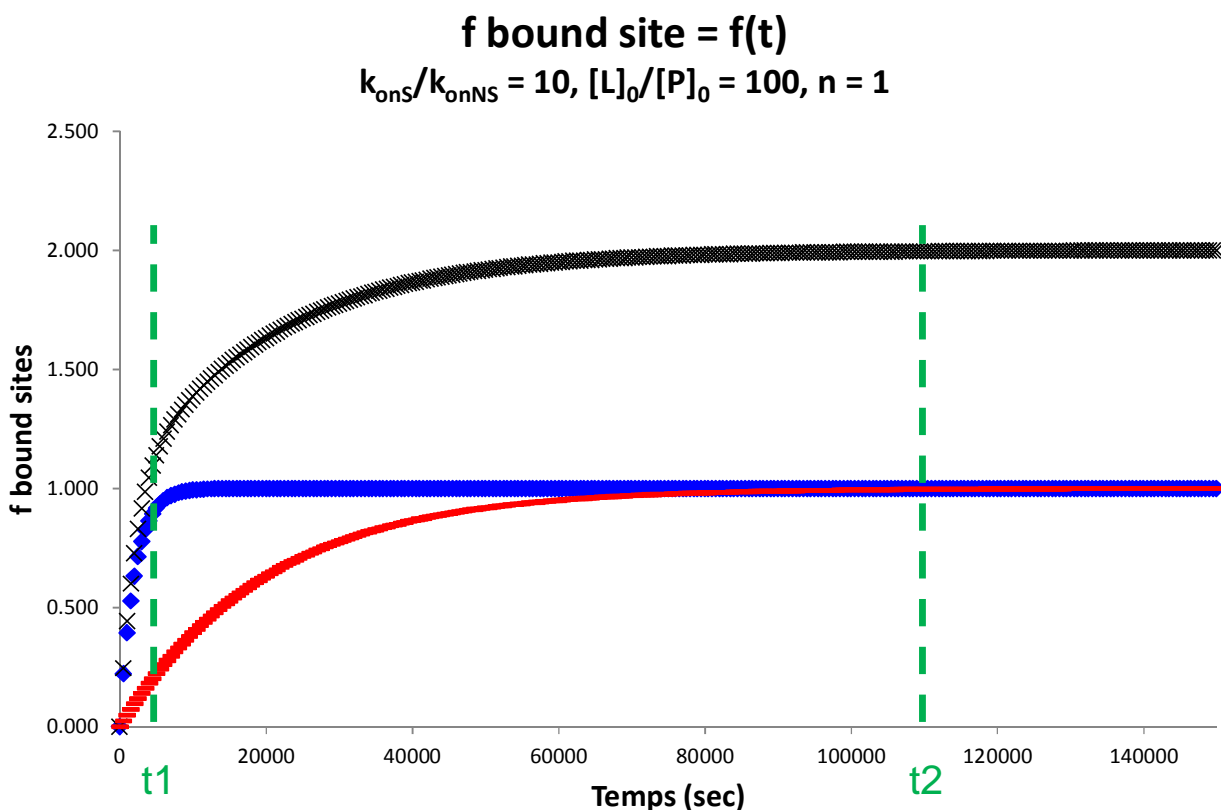
Cette équation simplifiée (35) montre que la contribution du site spécifique dépend de  $n$ ,  $k_{\text{onS}}$ ,  $k_{\text{onNS}}$ ,  $[L]_0$  et  $t$ . Les paramètres que l'on peut contrôler sont  $[L]_0$  et  $t$  alors que  $n$ ,  $k_{\text{onS}}$  et  $k_{\text{onNS}}$  vont varier en fonction du système étudié. Afin d'évaluer graphiquement l'influence de ces paramètres sur la contribution du site spécifique, plusieurs représentations ont été réalisées.

## 7.3 Représentations graphiques du modèle

### 7.3.1 Cinétiques comparées d'un site spécifique avec un site non-spécifique

L'association d'un ligand sur une protéine est illustrée Figure 112 pour un seul site non-spécifique ( $n = 1$ ) et  $k_{\text{onS}}/k_{\text{onNS}} = 10$  pour  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$  et  $k_{\text{onNS}} = 0.01 \text{ M}^{-1} \text{ s}^{-1}$ .

**Figure 112. Simulation de courbe d'association covalente d'un ligand sur une protéine à 1 site spécifique et 1 site non spécifique ( $n = 1$ ).** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-5} \text{ M}$ ,  $[L]_0/[P]_0 = 100$ ,  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{onS}}/k_{\text{onNS}} = 10$ . La fraction de liaison du site spécifique et du site non spécifique est représentée en bleu et rouge, respectivement. La fraction de liaison totale est indiquée en noire.



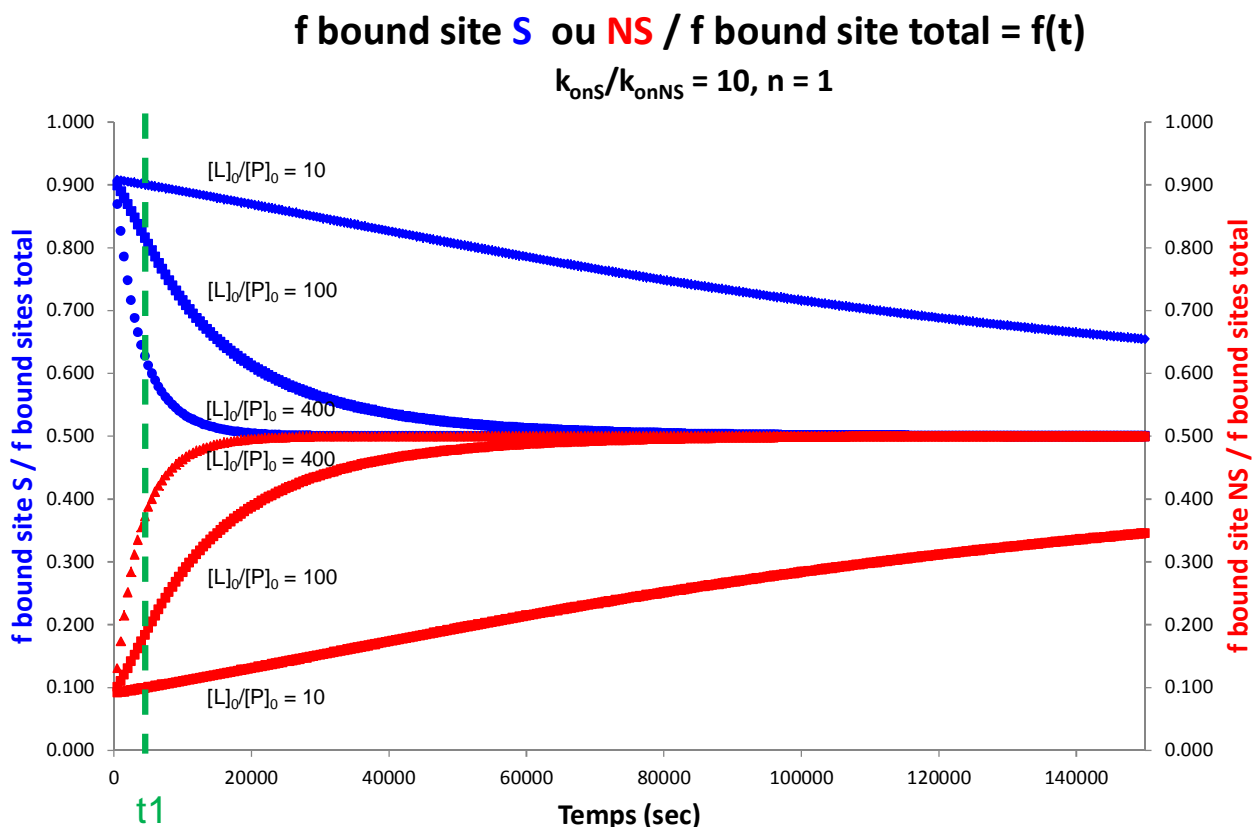
Les courbes bleue et rouge représentent la fraction de liaison des sites spécifique et non spécifique, respectivement. La fraction de liaison totale est indiquée en noir. Pour les temps longs ( $t_2$ ), chaque site est saturé avec un  $f_{\text{bound}}$  égal à 1 et un  $f_{\text{bound}}$  total égal à  $n+1$ , soit 2 dans cet exemple. Ce graphique montre que suivant le temps auquel est arrêtée la réaction,  $t_1$  ou  $t_2$ , les sites spécifique et non spécifique pourront ou non être distingués quantitativement. Au temps  $t_1$ , le site spécifique est majoritaire et donc distinguable du site non spécifique, mais au temps  $t_2$  les deux sites sont modifiés au même degré. Par ailleurs, comme le montre l'équation (35), les paramètres  $k_{\text{on}}$ ,  $[L]_0$  et  $t$  se compensent. Ainsi une augmentation de  $k_{\text{on}}$  ou de  $[L]_0$  d'un facteur cinq permet d'atteindre la saturation cinq fois plus rapidement et *vice versa*.

## 7.3.2 Contribution du site spécifique à la liaison globale

### 7.3.2.1 Représentation en fonction du temps

La contribution du site spécifique à la liaison globale, c'est-à-dire  $f_{\text{bound S}} / f_{\text{bound site total}}$ , est présentée Figure 113 en bleu.

**Figure 113. Contribution du site spécifique et du site non spécifique à la fraction modifiée globale en fonction du temps, pour différents rapports  $[L]_0/[P]_0$ .** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-05} \text{ M}$ ,  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{onS}}/k_{\text{onNS}} = 10$ . Les contributions du site spécifique et du site non spécifique sont représentées en bleu et rouge, respectivement. Les rapports  $[L]_0/[P]_0$  sont indiqués pour chaque courbe.



D'une manière générale, plus le temps augmente et plus la contribution du site spécifique diminue. Cette décroissance s'accroît d'autant plus vite que le rapport  $[L]_0/[P]_0$  est élevé : au temps  $t_1$ , la contribution du site spécifique est de 0.9, 0.8 et 0.6 pour les rapports  $[L]_0/[P]_0$  de 10, 100 et 400, respectivement. Des résultats similaires sont obtenus lorsque l'on fait varier  $k_{\text{onS}}/k_{\text{onNS}}$ .

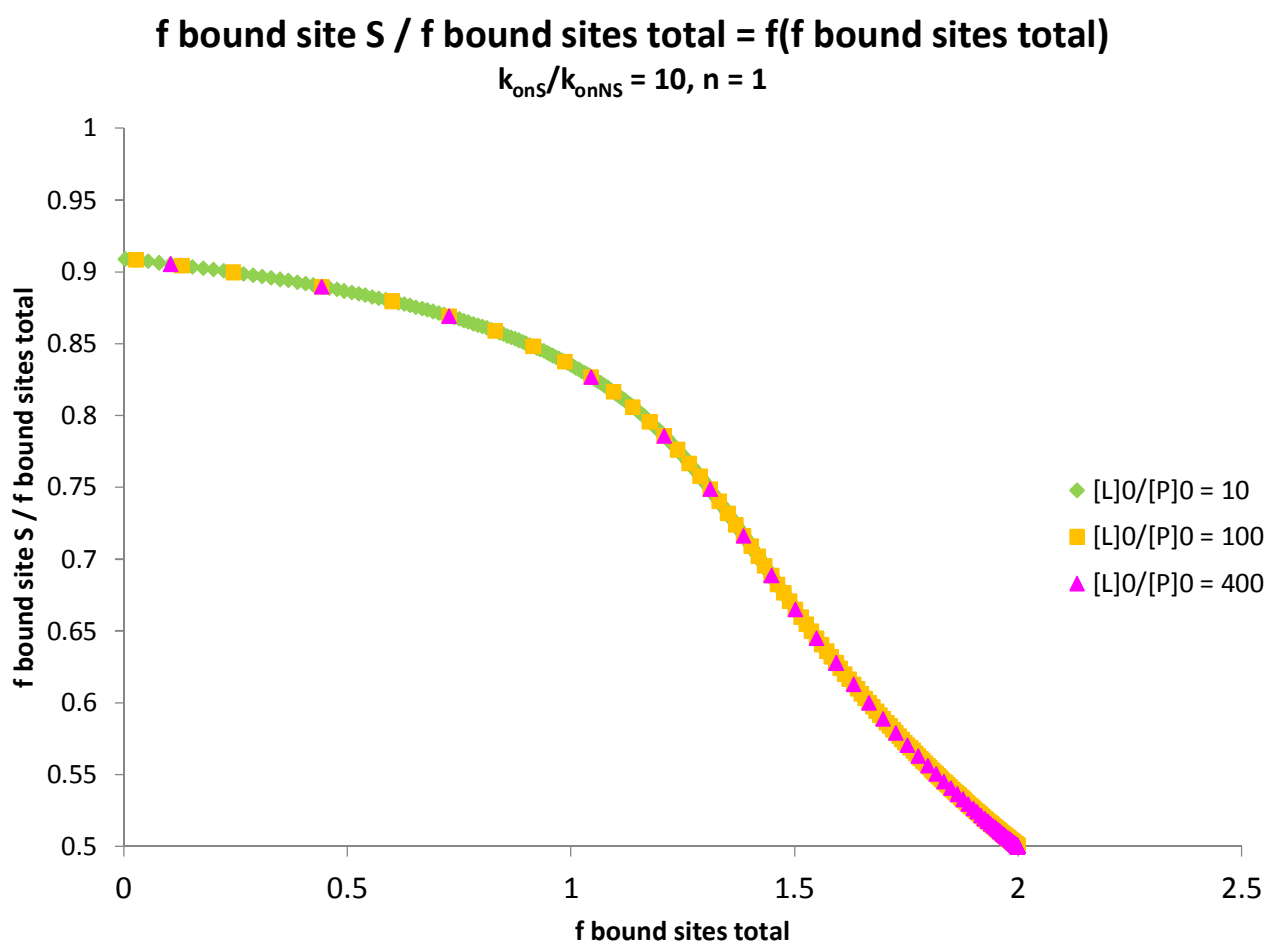
### 7.3.2.2 Représentation en fonction de fbound site total

Dans la représentation de la contribution du site spécifique en fonction du temps, la pente des courbes varie en fonction de  $[L]_0/[P]_0$ . Pour simplifier et avoir une vision globale, la contribution du site spécifique a été tracée en fonction de la fraction de liaison totale, ce qui permet de s'affranchir des



variations du rapport  $[L]_0/[P]_0$  (Figure 114). En effet, les courbes obtenues avec les trois valeurs de  $[L]_0/[P]_0$  : 10, 100 et 400 ont un profil identique et superposable. Cette représentation permet de montrer que pour une valeur de  $f_{\text{bound sites total}}$  donnée, le rapport  $[L]_0/[P]_0$  n'a pas d'influence sur la contribution du site spécifique et va uniquement impacter la cinétique de réaction. Par contre, il faut être conscient que la valeur de  $f_{\text{bound sites total}}$  (la fraction stœchiométrique globale de modification) en abscisse, ne sera en pratique observable qu'à un temps  $t$  bien précis, et que ce temps sera variable en fonction de  $[L]_0/[P]_0$  (Figure 113).

**Figure 114. Contribution du site spécifique à la fraction modifiée globale en fonction de la fraction modifiée globale.** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-05} \text{ M}$ ,  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{onS}}/k_{\text{onNS}} = 10$  et  $n = 1$ . Les rapports  $[L]_0/[P]_0$  sont indiqués pour chaque courbe.

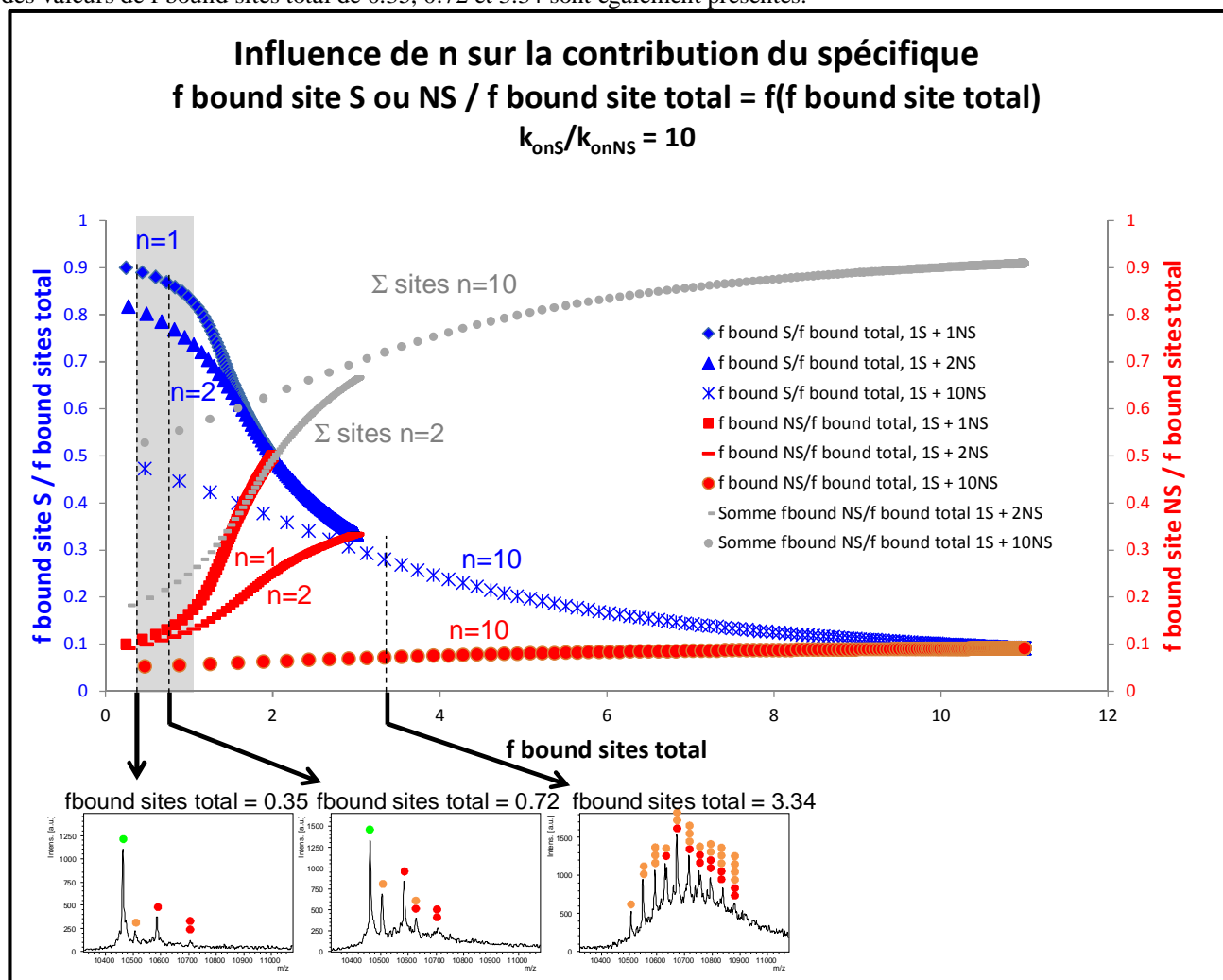


Dans l'exemple de la Figure 114, le site spécifique est largement majoritaire au moins jusqu'à une fraction stœchiométrique globale de 1. D'autres considérations peuvent influencer le choix d'une valeur expérimentale de  $f_{\text{bound sites total}}$ .

### 7.3.3 Influence de n

L'influence de n, nombre de sites non spécifiques, sur la contribution du site spécifique a été évaluée. Ce nombre dépend de chaque système protéine-ligand étudié et est souvent inconnu. La Figure 115 présente les résultats obtenus pour n = 1, 2 et 10.

**Figure 115. Influence du nombre de sites non spécifique n sur la contribution du spécifique.** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-05} \text{ M}$ ,  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{onS}}/k_{\text{onNS}} = 10$ . La contribution du site spécifique, d'un site non spécifique et de tous les sites non spécifiques est représentée en bleu, rouge et gris, respectivement. Les valeurs de n sont indiquées pour chaque courbe. Les spectres MALDI-TOF MS de l'état de charge 2+ du complexe hPEBP1-locostatine correspondant à des valeurs de f bound sites total de 0.35, 0.72 et 3.34 sont également présentés.

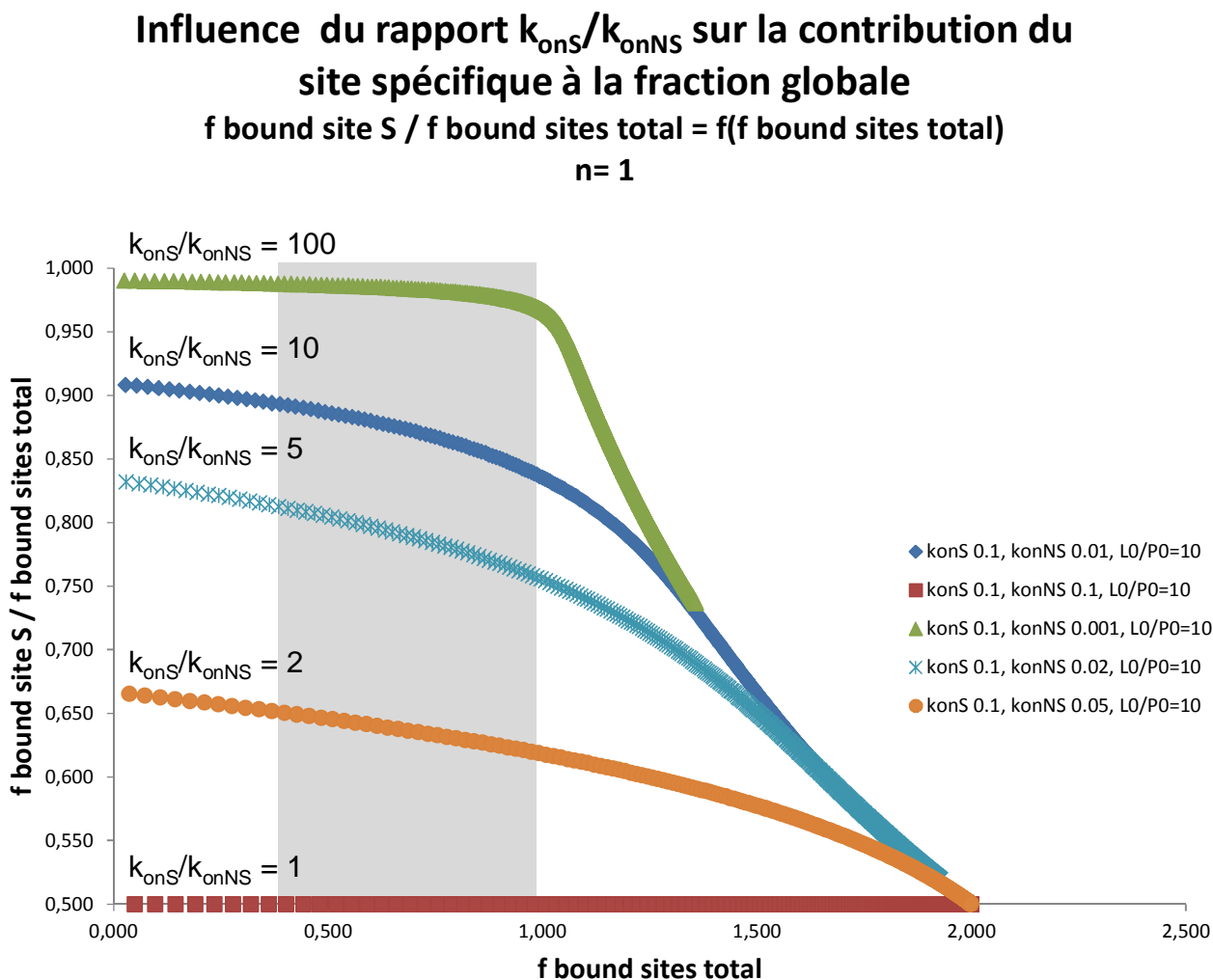


Lorsque n augmente, la contribution du spécifique diminue. Pour chaque cas, lorsque la saturation est atteinte, c'est-à-dire  $f_{\text{bound sites total}} = 1+n$ , la contribution du site spécifique est égale à  $1 / (1+n)$ . Les spectres du complexe hPEBP1-locostatine correspondant à différentes valeurs de f bound sites total de 0.35, 0.72 et 3.34 sont également présentés. Les valeurs de f bound sites total comprises entre 0.4 et 1, représentées par un rectangle gris, correspondent à un bon compromis entre une contribution maximale du site spécifique et un rapport S/N suffisant en spectrométrie de masse.

### 7.3.4 Influence du rapport $k_{onS}/k_{onNS}$

L'influence du rapport  $k_{onS}/k_{onNS}$  sur la contribution du site spécifique a également été évaluée. De la même manière que  $n$ , les valeurs de  $k_{onS}$  et  $k_{onNS}$  dépendent de chaque système protéine-ligand étudié. Pour des systèmes simples à un seul site, les expériences de cinétiques permettent d'obtenir  $k_{onS}$ . La Figure 116 présente les résultats obtenus pour  $k_{onS}/k_{onNS} = 100, 10, 5, 2$  et  $1$ .

**Figure 116. Influence du rapport  $k_{onS}/k_{onNS}$  sur la contribution du spécifique.** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-5}$  M,  $k_{onS} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $n = 1$ . Les valeurs de  $k_{onS}/k_{onNS}$  sont indiquées pour chaque courbe.



Ces courbes montrent que la contribution de liaison spécifique est d'autant plus basse que le rapport des vitesses est bas, ce qui rend la liaison spécifique d'autant plus difficile à distinguer de la liaison non-spécifique. Encore une fois, les valeurs de  $f_{\text{bound sites total}}$  comprises entre 0.4 et 1, représentées par un rectangle gris, correspondent au meilleur compromis entre une contribution maximale du spécifique et un rapport S/N suffisant en spectrométrie de masse.

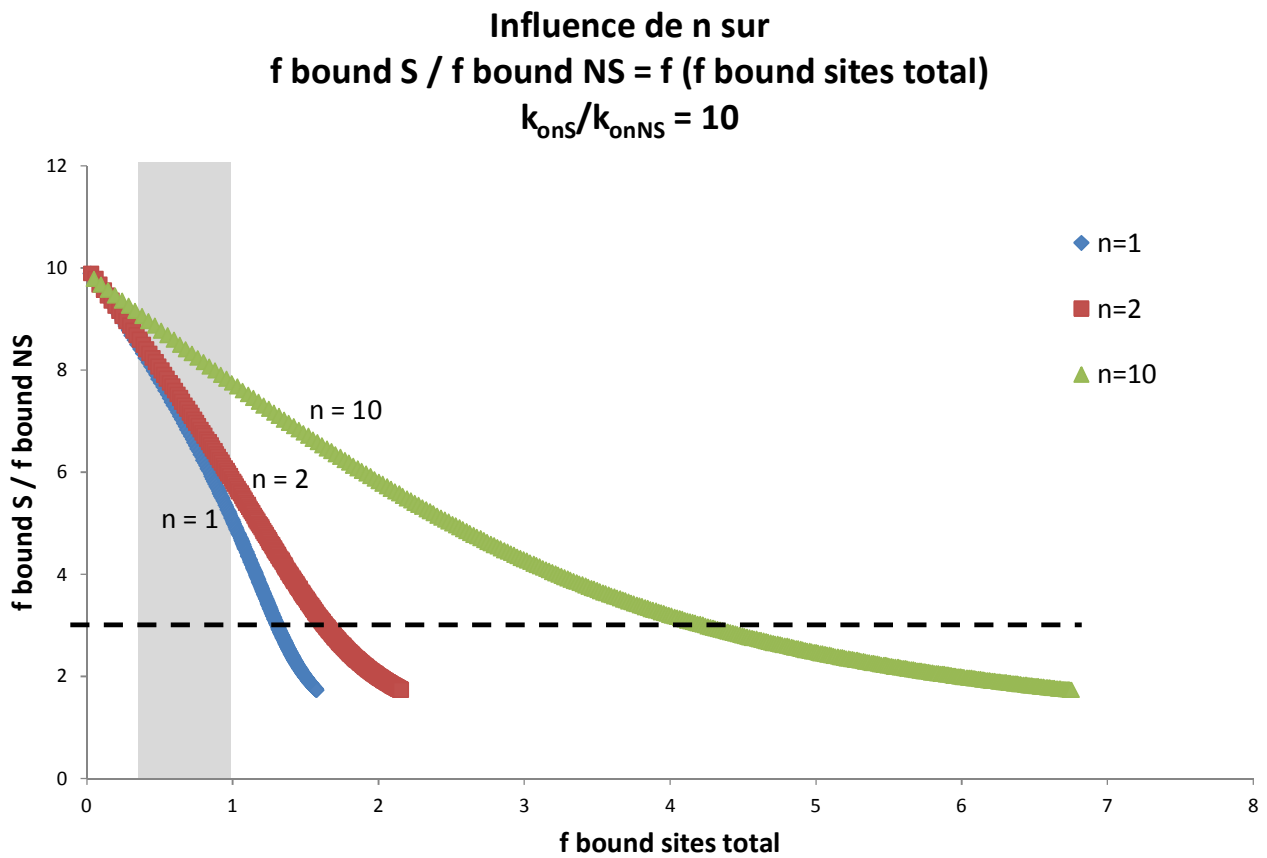
### 7.3.5 Facteur de discrimination

Pour discriminer le site majoritaire des sites minoritaires, il est nécessaire de définir un critère objectif de discrimination. Ce critère peut être exprimé comme un seuil à partir duquel un site peut être considéré comme majoritaire. Les graphiques précédents indiquent qu'il n'est pas judicieux de fixer une valeur absolue pour la contribution du site spécifique car cette dernière varie en fonction de  $n$  et du rapport  $k_{\text{onS}}/k_{\text{onNS}}$ . En d'autres termes, on ne peut pas se baser sur la proportion de liaison spécifique dans l'ensemble des liaisons pour définir un seuil.

Par contre, on peut intuitivement concevoir un critère qui compare les liaisons deux à deux et considère que le spécifique doit se distinguer quantitativement de toute liaison non-spécifique. Nous avons donc considéré le rapport  $f_{\text{bound S}} / f_{\text{bound NS}}$ . Plus ce rapport est élevé, meilleure sera la discrimination entre le site spécifique et un site non spécifique. On peut ensuite fixer une valeur seuil de  $f_{\text{bound S}} / f_{\text{bound NS}}$ , à partir de laquelle la liaison est considérée comme spécifique. Cette valeur seuil est définie comme le facteur de discrimination (FD). Par analogie avec les principes de détection du signal, qui imposent de ne considérer comme du signal que les valeurs supérieures ou égales à trois fois la valeur du bruit ( $S/N = 3$ ), on peut proposer le principe que la liaison spécifique doit être supérieure ou égale à trois fois la valeur de la liaison non-spécifique la plus abondante. Ceci correspondrait à  $FD = 3$ .

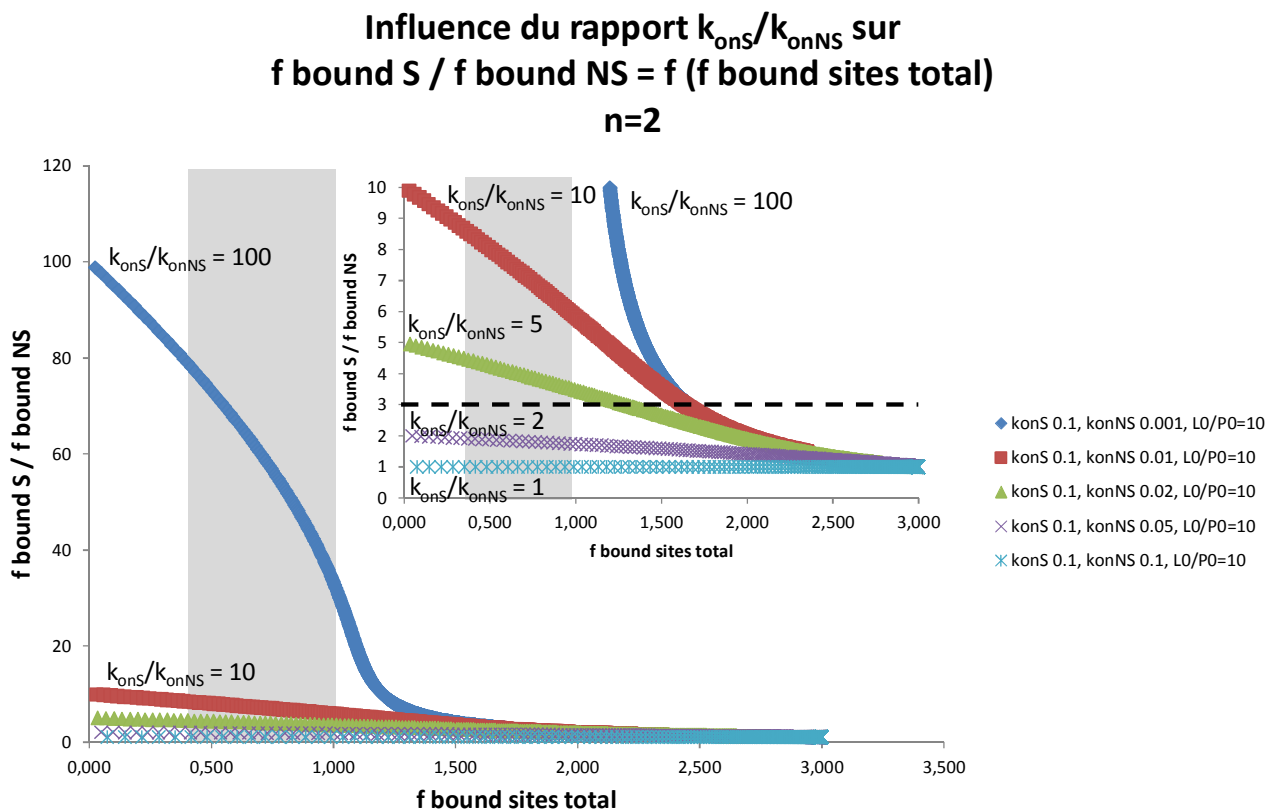
L'influence de  $n$  et de  $k_{\text{onS}}/k_{\text{onNS}}$  sur le rapport  $f_{\text{bound S}} / f_{\text{bound NS}}$  a été évaluée (Figures 117 et 118).

**Figure 117. Influence de n sur le rapport f bound S / f bound NS.** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{-05} \text{ M}$ ,  $k_{\text{onS}} = 0.1 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{onS}}/k_{\text{onNS}} = 10$ . Les valeurs de n sont indiquées pour chaque courbe. La ligne en pointillée noire représente un f bound S / f bound NS = 3.



La Figure 117 montre que n a peu d'influence sur le rapport f bound S / f bound NS, notamment pour des faibles valeurs de f bound site total : f bound S / f bound NS est égal à environ 8.5 pour n=1 et n=2 et environ 9 pour n=10 lorsque f bound site total est égal 0.4.

**Figure 118. Influence du rapport  $k_{onS}/k_{onNS}$  sur le rapport  $f_{bound S} / f_{bound NS}$ .** Les paramètres sont :  $[P]_0 = 5 \cdot 10^{05} M$ ,  $k_{onS} = 0.1 M^{-1} s^{-1}$ ,  $n = 2$ . Les valeurs de  $k_{onS}/k_{onNS}$  sont indiquées pour chaque courbe. La ligne en pointillée noire représente un  $f_{bound S} / f_{bound NS} = 3$ . Insert: zoom sur les rapports  $f_{bound S} / f_{bound NS} < 10$ .



La Figure 118 indique que plus  $k_{onS}/k_{onNS}$  diminue et plus le rapport  $f_{bound S} / f_{bound NS}$  diminue. Ce graphe montre qu'en fixant une valeur seuil de  $f_{bound S} / f_{bound NS}$  égale à 3, seuls les systèmes protéine-ligand avec un  $k_{onS}/k_{onNS}$  très faible  $< 5$  ne pourront être discriminés, ce que nous considérons comme acceptable. Des résultats similaires sont obtenus pour  $n=1$  et 10.

## 8. Essai de localisation du site spécifique – Approche quantitative

Afin d'identifier tous les sites de hPEBP1 modifiés par la locostatine, une approche quantitative originale a été développée.

### 8.1 Développement d'une nouvelle méthode pour la quantification exhaustive et parallèle du degré de modification à chaque site

Draft : Exhaustive parallel inventory and quantification of the extent of modification at each site of a protein: a subtractive approach.





# Exhaustive parallel inventory and quantification of the extent of modification at each site of a protein: a subtractive approach

*Guillaume Gabant, Alain Boyer, Martine Cadene\**

Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique, UPR 4301, affiliated with the University of Orléans and to INSERM, rue Charles Sadron, 45071 Orléans cedex 2

## **RECEIVED DATE**

\* To whom correspondence should be addressed. E-mail: [martine.cadene@cnrs-orleans.fr](mailto:martine.cadene@cnrs-orleans.fr).

Tel: +33-238-25-5624, Fax: +33-238-63-1517.

**Keywords:** Covalent ligand, protein modification, PTM, quantification, mass spectrometry

## **Abstract**

## Introduction

Protein modifications play a crucial role in living organisms. By changing the physicochemical properties of proteins through chemical and biochemical reactions, modifications modulate the function of proteins. Protein modifications can be the result of an endogenous cell process in the form of post-translational modifications. Alternatively, they can be induced, intentionally or not, by exposure to any number of external environmental or man-made agents. Broad cellular functions including signaling and metabolism rely on PTM that are finely regulated and involve endogenous molecules. Induced modifications on the other hand are usually the result exposure to drugs, diets, stress, infectious agents, lifestyle-derived xenobiotics (e.g. tobacco, alcohol), environmental pollutants and radiations, (Rappaport and Smith, 2010). Disorders of PTMs and/or induced chemical modifications may lead to developmental dysfunctions and human diseases (Rubino et al., 2009).

The quantification of protein modifications is of great interest for the comprehension of biological processes and for clinical research investigations. The determination of the extent of *in vivo* modification gives a measure of the effect of enzyme inhibitors (e.g. anti-kinase drugs) on modification (in this case, phosphorylation) at a given site. Covalent drugs represent an ancient class of medicine whose interest has been renewed in recent years (Singh et al., 2011). The quantification of the extent of a residue modification by these drugs gives access to the addition kinetics. The quantification of *in vitro* induced modification is of great interest for recombinant therapeutic proteins such as antibodies, as oxidation could occur during the manufacturing process and storage, leading to the inactivation of the protein (Torosantucci et al., 2014). Today, a concerted effort has been made to evaluate the impact of modifications, whether induced by a covalent drug or undergone by a biodrug, on efficacy and human safety (Johnson et al., 2010; Torosantucci et al., 2014).

Mass spectrometry (MS) is a powerful technique for the detection, localization and quantification of modifications (Olsen and Mann, 2013). Although it is hoped that top-down methods can solve the proteoform riddle, the complete separation of isoforms they rely on are elusive at this point. Focusing on methods based on proteolysis of the target protein prior to MS analysis, i.e. bottom-up methods, several approaches have been described to determine the degree of modification of proteins. A number of studies have used the relative signal intensities of the modified peptide and its corresponding nonmodified form to infer a degree of modification. However, since a modification such as phosphorylation can alter the ionization efficiency of a peptide to an unpredictable degree (Gropengiesser et al., 2009), the extent of modification of a peptide cannot be determined by simple comparison of these signal intensities.

In an alternative label-free approach, the ion count of modified and nonmodified peptides was titrated over the course of an experiment, thereby allowing for the calculation of normalized signals according to the saturation point (Steen et al., 2005). A direct ratio of normalized signals at each time point was then used to quantify the extent of phosphorylation. This approach cannot compensate for any sample-to-sample variation during sample handling and depends on the analysis of a series of samples spanning the gamut to saturation.

Other approaches use isotopically labeled internal standards to carry out absolute quantification of modified and nonmodified protein in the form of AQUA peptides (Gerber et al., 2003) or PSAQ proteins (Ciccimaro et al., 2009). The degree of modification can then be deduced by calculation from the ratio of absolute quantities. These approaches require the costly synthesis and purification of numerous stable isotope-labeled internal standard peptides or proteins in their modified and nonmodified forms. Furthermore, they require prior knowledge of the type of modification and the nature of modified residue(s).

Elucidation of the fraction of modification has also been reported in the case of phosphorylation using a minimum of three different ratios representing protein, phosphopeptide and unmodified peptide changes based on stable-isotope labeling with amino acids in cell culture (SILAC) (Olsen et al., 2010). However, because of the complex math involved, even slight errors in measurement can cause wildly changing stoichiometry calculation results.

Finally, another elegant strategy used a combination of deuterioformaldehyde/formaldehyde stable isotope chemical labeling and alkaline phosphatase treatment (Wu et al., 2011a). The sample of interest was divided into two aliquots: one is treated with phosphatase while the other is not. Following differential chemical labeling of free amines with stable isotopes, both aliquots are recombined. Mass spectrometric analysis of the recombined mixture reveals the degree of phosphorylation by measuring the signal increase from the dephosphorylated peptide of the corresponding phosphopeptide (Wu et al., 2011a). Nonetheless, this strategy is compromised if the derivatization step or the enzymatic removal of the phosphate group are incomplete. Moreover it cannot be applied to irreversible modifications.

In an attempt to circumvent the limitations of existing approaches, we set out to develop a method for a robust, reliable and comprehensive quantification of modification. The ideal method should be insensitive to changes in the solubility and the ionization efficiency of peptides containing the modified residues, able to detect and deal with missed cleavages caused by the modification, capable of quantifying modifications in parallel and without prior knowledge of the type of target residues. It should be universal in that it can be applied to all types of protein modification and exhaustive, i.e. capable of quantifying all modifiable sites. This last criterion is crucial because, in a bottom-up approach, even under the best possible conditions, several peptide purification or detection factors can affect the accessibility of

residues to the analysis, so that some modified sites may be missed by MS. Thus, the map of observed modified sites may not be complete and it should be borne in mind that the failure to identify a modified site is insufficient to claim its non-existence. Finally, the developed tools should allow for the straightforward visualization of the result.

In this study, we present a subtractive method for the parallel quantification of the extent of modification at each site of a purified protein. In this strategy, we use isotopically labeled unmodified protein as an internal standard. This approach is a powerful tool to list all the detectable modified sites on the protein, to clearly show the areas where no information could be drawn, if any, and to demonstrate the absence of modification on other residues.

As a concrete example, this method was applied to the interaction of human Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1) with a covalent ligand, locostatin. The hPEBP1 protein, named according to the current classification of the PEBP family of proteins (Zeng et al., 2008) - also named Raf Kinase Inhibitory Protein (RKIP) in mammals - is a metastasis suppressor gene product in different types of cancers (Lee et al., 2006; Klysik et al., 2008). Locostatin, (S)-4-benzyl-3-crotonyl-oxazolidin-2-one, is the only known compound with demonstrated activity towards PEBP1. At a cellular level, it prevents cell migration by binding to PEBP1, as demonstrated by a chemical genetics approach (Mc Henry et al., 2002; Zhu et al., 2005). The unique observation of the *in vitro* anti-migration effect of locostatin, together with the anti-metastasis effect of its target protein, is the basis for the study of the hPEBP1-locostatin complex *in vitro* with the aim to obtain valuable information for the design of new molecules as potential anti-metastatic leads. To date, there is no known X-ray or NMR structure of the hPEBP1-locostatin complex, and the determination of the covalent site of addition of locostatin on hPEBP1 has proved a vexing analytical challenge to several

teams. Thus, we have undertaken the quantification of the degree of modification of all modifiable sites of hPEBP1 by locostatin. In the present study, all the sites modified by locostatin were detected and quantified in a controlled reaction, i.e. in conditions of single-hit statistics with elimination of the excess of locostatin before complete proteolysis. The method we developed could be generally applied to target-drug or protein-reagent complexes of a covalent nature, as well as to PTMs where a removal of the modification is available.

## Results and Discussion

### Strategy and general workflow for parallel quantification of the degree of site modification

The quantification strategy presented here was developed in the context of an induced modification, i.e. a protein-ligand covalent complex. However, it can be applied to *de facto* modifications such as PTMs, provided that the modification can be removed. The strength and originality of the subtractive method is that it relies exclusively on signals derived from nonmodified peptides. For the sake of clarity, the logic behind the strategy can be explained in the case of an induced modification of the protein, the reasoning being the same for removable PTMs. The method is based upon the fact that the modification of a given site in a protein automatically leads to a decrease in the pool of protein that is non-modified at this site. Consequently, the modification produces a proportional decrease of the signal associated with the peptide that includes this site and is left nonmodified in the sample. This phenomenon is even cumulative, as the subtractive effects of modification at two or more sites would add up to a total sum of decrease in the peptide's signal. Thus, by quantifying the decrease of the nonmodified peptide, we can indirectly quantify the increase in modified peptide, and thereby calculate the extent of modification.

Depending on whether a *de facto* or an induced modification is measured, different front-ends are required to produce the samples and their controls for quantification (Figure 1-A). In the case of *de facto* modification, the  $^{14}\text{N}$  protein control is generated by removal of the modification by chemical or enzymatic means, and the experimental sample is left untouched. In the case of induced modifications, the  $^{14}\text{N}$  protein in the absence of ligand or reagent is the control. The method then consists in performing three incubations in parallel:  $^{14}\text{N}$  protein without ligand (or after modification removal) as control,  $^{14}\text{N}$  protein with ligand (or *de facto*

modification) as experiment, and  $^{15}\text{N}$  protein without ligand (or modification) as internal standard. The internal standard is used in the early steps of the workflow to compensate for losses during sample processing, as well as for incomplete proteolysis and variations in the chromatographic injection volume. After incubation, the same volume of internal standard solution is added into the tubes containing the control and experiment solutions. This step is the only source of random variability for quantification and must be done carefully. For induced modifications, there is an extra step where the excess ligand/reagent is removed by a size exclusion spin column to prevent non-specific modification of peptides. At this point both workflows merge (Figure 1-B). The next two steps aim at achieving the most complete proteolysis to prevent bias in the measurement of the extent of modification and to have optimal peptide signal to noise ratios. They also ensure that the modified and non-modified forms of the target protein undergo proteolysis with the same efficiencies despite possible differences in their tridimensional structure. This is achieved by the combined use of a denaturation step followed by a proteolysis in the presence of denaturing agent at a high enzyme:substrate ratio. Multiple proteolyses are carried out in parallel in order to have maximum coverage of sequence by combining the coverage obtained for different proteases. Ideally, each and every residue of the protein sequence should be covered by at least one proteolytic peptide, so that no residue can escape the quantitative measurement of its modification. The proteolysis efficiency is assessed through detection of the intact protein by MALDI-TOF MS, and counting of missed cleavages in product peptides. The mixtures of  $^{14}\text{N}$  and  $^{15}\text{N}$  peptide solutions are analyzed by nanoUltraHPLC-UHR\_QTOF MS in triplicate. Data are then processed using a quantification software such as Skyline to identify peptides, to extract related nonmodified  $^{14}\text{N}$  and  $^{15}\text{N}$  ions to construct the extracted ion chromatogram (XIC), and to calculate the area under the curve (AUC) of each  $^{14}\text{N}$  and  $^{15}\text{N}$  XIC peak. By dividing the  $^{14}\text{N}$  area with the  $^{15}\text{N}$  area, one obtains a ratio of decrease in the  $^{14}\text{N}$  sample



versus the  $^{15}\text{N}$  standard. In the control sample, the expected area ratio is 1, while in the experiment, the area lost due to the modification finds itself subtracted from the  $^{14}\text{N}$  area, so that the ratio drops below 1 upon modification. The bound fraction is then calculated using Equation 1:

$$f \text{ bound} = 1 - (\text{Area } ^{14}\text{N} / \text{Area } ^{15}\text{N}) \quad (1)$$

In the control sample, as the protein is not modified,  $f$  bound values are expected to be equal to zero for all peptides. Even if some sites suffer from non-optimal cleavage efficiency, the effect will be compensated by the similar behavior of the internal standard. The observed deviation from zero corresponds to a combination of systematic and random error. The systematic error, which we call  $e$ , is calculated from triplicate data and corresponds to an average of bound fraction ( $f$  bound av) (Equation 2).

$$e = f \text{ bound av. ctrl} \quad (2)$$

The  $e$  value calculated in the control sample is then used to correct the bound fraction of each charge state and peptide in the experiment sample (Equation 3).

$$f \text{ bound corr.} = f \text{ bound av. exp.} - e \quad (3)$$

The corrected bound fractions  $f$  bound corr. for different charge states of the same peptide are then averaged to obtain the value of the averaged corrected bound fraction per peptide ( $f$  bound corr. av.). Finally, a graphical representation of the degree of modification of each peptide was designed using Equation (4) to provide a comprehensive view of the modified, nonmodified and nonobserved sequences of the protein.

$$f \text{ bound corr. av.} = f(\text{sequence}) \quad (4)$$

In the remainder of the paper, experimental values are shown for the hPEBP1-locostatin complex.

### **Evaluation of $^{15}\text{N}$ metabolic labeling and experimental considerations**

Isotopic metabolic labeling efficiency has to be measured at the beginning of the workflow to validate the use of the labeled protein as internal standard. Here, metabolic labeling with the stable  $^{15}\text{N}$  nitrogen isotope was performed. Metabolic incorporation was evaluated by MALDI-TOF MS analysis by measuring the delta mass between the labeled and nonlabeled protein peaks (Figure 2). The observed masses of  $^{15}\text{N}$  hPEBP1 and  $^{14}\text{N}$  hPEBP1 are 21,160 Da and 20924 Da, respectively. The mass increase caused by  $^{15}\text{N}$  labeling is 236 Da. Considering that there are 256 nitrogen atoms in the protein, the incorporation efficiency is thus 92.2%. This isotope incorporation efficiency is rather low. As a consequence, the isotopic peak distribution of  $^{15}\text{N}$ -labeled peptides will be broad. In many quantification strategies, only the first three isotopic peaks of the monoisotopic distribution, a0, a1, and a2 are taken into account to calculate the area under the curve (AUC). When the incorporation is complete (= 100%), no bias on the  $^{14}\text{N}/^{15}\text{N}$  ratio is expected with this approach. When the isotope incorporation is incomplete as is the case here, the contribution of the first three isotopic peaks to the distribution will depend on the molecular composition. To evaluate the effect of less-than-100% isotope incorporation efficiency, we compared the  $^{14}\text{N}/^{15}\text{N}$  ratios calculated using the first three isotopic peaks *versus* all isotopic peaks in the distribution. For this test, Asp-N peptides from a 1:1 mixture of  $^{14}\text{N}$  /  $^{15}\text{N}$  hPEBP1 protein solutions were analyzed in decuplicate by nanoUltraHPLC-UHR-QTOF MS (MaXis, Bruker), and the  $^{14}\text{N}/^{15}\text{N}$  ratios for each peptide were plotted as a function of  $m/z$  (Figure 3-A). When the first three peaks of the monoisotopic distribution are considered, the  $^{14}\text{N}/^{15}\text{N}$  ratio increases as a

function of  $m/z$ , demonstrating the influence of molecular composition on their contribution. Furthermore, the ratio values range from 1.5 to 2.5, which is far from the expected value of 1. When the whole distribution is used, the regression curve is practically a horizontal line with a y-intercept at 1.13, i.e. much closer to the expected value of 1. This remaining error of 0.13 can be explained by errors in protein concentration combined with pipetting errors. These results were confirmed by plotting the theoretical relative isotopic abundances using either three or all isotopic peaks (Figure 3-B), thus highlighting the importance of considering the whole isotopic distribution. As the incorporation is rarely 100% complete, we propose to systematically use the whole isotopic distribution to ensure a reliable and accurate quantification.

### **Graphical representation of parallel peptides quantification**

The hPEBP1 protein was incubated for 5 hours with or without locostatin and cleaved with Asp-N, chymotrypsin and trypsin. The proteolysis efficiency for each protease was assessed by checking for the absence of intact protein after proteolysis (data shown). After nanoUltraHPLC-UHR-QTOF MS analysis, the extent of sequence coverage and the number of missed cleavages were checked. The quantification software Skyline was used to derive  $^{14}\text{N} / ^{15}\text{N}$  ratios, and the raw f bound for each charge state of each peptide was calculated. In the control, as the protein is not modified, the expected f bound is zero for all peptides. However, a combination of systematic and random error results in deviation from zero for each data point, as shown for Asp-N peptides in Figure 4-A. An error calculation based on a global average of control f bounds would seem appropriate to correct for systematic error. However, correcting all of the control data with the global average error generates a dataset with a wide dispersion, suggesting a weakness in this approach to correction (data not shown).

To help trace the origin of the dispersion, we decided to plot the control f bound against each charge state of each peptide (Figure 4-A). Contrary to expectations, strong local effects are observed, i.e. the spread and trend of the error seems to depend on the peptide identity and on its charge state. Thus, a local epsilon  $\epsilon$  corresponding to an average bound fraction for each charge state of each peptide was calculated from triplicate data (Equation 5).

$$\epsilon = f \text{ bound av. ctrl} = \Sigma (f \text{ bound}_{\text{charge state}}) / n_{\text{charge states}} \quad (5)$$

Each local epsilon was then used to correct the f bound of each corresponding peptide and charge state (Equation 6) (Figure 4-B), leaving only dispersion due to random error.

$$f \text{ bound corr.} = f \text{ bound av. ct.} - \epsilon \quad (6)$$

The dispersion of the corrected f bound can be measured for each protease. A standard deviation of 0.0164, 0.0238 and 0.0232 was obtained for Asp-N, chymotrypsin and trypsin datasets, respectively.

Each local epsilon calculated in the control sample was also used to correct the bound fraction of the corresponding charge state and peptide in the experiment sample (equation 7).

$$f \text{ bound corr.} = f \text{ bound av. exp.} - \epsilon \quad (7)$$

The corrected bound fractions f bound corr. for different charge states of the same peptide were then averaged to obtain the value of the averaged corrected bound fraction per peptide (f bound corr. av.).

The corrected bound fractions f bound corr. av. obtained in presence of locostatin for each protease Asp-N, chymotrypsin and trypsin are presented Figure 5. The standard deviations calculated from the control datasets were used to assess the likelihood that a f bound differs

from zero. At twice the standard deviation, there is a risk of 5% to consider a  $f$  bound as different from zero while it is not. The threshold values are 0.033, 0.048 and 0.046 for Asp-N, chymotrypsin and trypsin, respectively. A band corresponding to  $f$  bound corr. av. values within the threshold boundaries is represented for each protease in Figure 5. The  $f$  bound outside this band are considered as significantly positive or negative.

At this stage, all  $f$  bound values different from zero were expected to correspond to modified peptides, and thus to have positive values. However, Figure 5 shows that some  $f$  bound are significantly negative. Indeed, this is the case for the [83-93] and [133-141] trypsin peptides which are observed at  $f$  bound -0.290 and -0.141, respectively. The significance of this observation will be apparent in the next section.

### **Detection of missed and shifted cleavages caused by the modification**

If a cleavable residue of a proteinase is modified, we assume that it cannot be cleaved, thus leading to a missed cleavage caused by the modification (MCm). This MCm must be differentiated from the classical MC caused by incomplete proteolysis due to close proximity of sites, or due to an effect of the substrate-binding sequence specificity of the protease (e.g. for trypsin: presence of a proline, lysine or arginine residue in the P1' position, or of a negatively charged residue such as glutamate and aspartate residues in the vicinity of the cleavage site). A MCm leads to a decrease of the amount of both of the nonmodified peptides flanking the modified residue. In terms of quantification, the MCm gives the peptide adjacent to the modification an artificially high  $f$  bound equal to the  $f$  bound of the peptide which is really modified. The specific pattern of peptide  $f$  bounds generated by this behavior is illustrated by the theoretical scheme in Figure 6. This pattern was observed experimentally in the trypsin graph, specifically with peptides [40-47] and [48-62] (Figure 5) which are detected

as two contiguous peptides with near identical fractions of modification ( $f_{\text{bound}} = 0.05$ ). This interpretation is corroborated by the fact that the residue in position 47 is a lysine, a modifiable and cleavable residue. On the two adjacent peptides, the peptide to eliminate is the one that does not contain the cleavable residue, i.e. [48-62] peptide in this example. From a qualitative point of view, since the protease cannot cleave a modified residue, a MCm constitutes a clue that the residue where it occurs is indeed modified. Lys47 can thus be proposed to be modified by locostatin.

Although the presence of MCm was expected, the negative  $f_{\text{bound}}$  values such as those observed for the [83-93] and [133-141] trypsin peptides were not. We thus set out to find the causal factor for this unexpected behavior. We first checked for the presence of an interference in the isotopic distributions, which was ruled out as described (Zhang et al., 2010). If the observed negative  $f_{\text{bound}}$  is due to a change in the  $^{14}\text{N} / ^{15}\text{N}$  area ratio, it means that the  $^{14}\text{N} / ^{15}\text{N}$  area ratio in the experiment increases relative to the ratio in the control, which is constant by design.

To explain a relative increase of the experiment  $^{14}\text{N} / ^{15}\text{N}$  area ratio, we hypothesized the existence of a shifted cleavage (SC). The definition of a SC between experiment and control is a modification-induced difference in the proteolysis efficiency at a site adjacent to the modification. In other words, the SC involves the presence a cleavable residue in the immediate vicinity of a modified cleavable residue. In the control, the close proximity of the two cleavable residues leads to classical MC, so that cleavage at site #1 is much more efficient than at site #2 (see theoretical scheme Figure 7). In the experiment, based on the rational hypothesis that a cleavable residue which is modified cannot be cleaved, the modification of the cleavable residue prevents cleavage, producing a MCm at site #1, and more importantly, promotes cleavage at site #2, thus increasing the  $^{14}\text{N} / ^{15}\text{N}$  area ratio of the corresponding peptide. A recognizable pattern of three peptides is produced, two of which

have adjacent positive  $f$  bounds due to the MCm, while the third with negative  $f$  bound is shorter than the second (Figure 7). There is symmetry between the positive and negative  $f$  bounds if and only if the cleavage is spread 50-50% between adjacent sites.

This SC pattern is observed for the [63-80], [81-93] and [83-93] peptides. A MC at Lys82 may be caused by the neighboring cleavable Lys80. In this case, if Lys80 is modified, it can no longer be cleaved, and a MCm is generated. The cleavage of Lys82 will thus be promoted, thereby increasing the area of the  $^{14}\text{N}$  [83-93] peptide in the experiment and leading to the observation of a negative  $f$  bound. The existence of a SC at Lys82 is corroborated by the relative increase of the  $^{14}\text{N}$  [83-93] area compared to its  $^{15}\text{N}$  counterpart in the experiment, while both  $^{14}\text{N}$  and  $^{15}\text{N}$  [83-93] areas are predictably small in the control.

The [133-141] peptide also bears a negative  $f$  bound. No interference in its isotopic distribution was detected. The shifted cleavage phenomenon was thus considered for this peptide. However, in the absence of a cleavable site in the immediate neighborhood of residue Lys 32, this seems unlikely. In Asp N and chymotrypsin cleavages, the corresponding sequence was not associated with negative  $f$  bounds, suggesting that an artefact was produced in the trypsin incubation alone. It should be noted that this peptide contains an exposed cysteine which is normally free, so that some component that is exclusive to the trypsin incubation may affect the corresponding peptide differently than in the other proteolysis. While the source of this odd behavior is not clear yet, this  $f$  bound cannot be taken at face value and was not used for subsequent data processing.

At this stage, there is a lot of overlapping peptides both within and between proteolyses, preventing a straight interpretation of the data.

## **Narrowing down the modified positions**

Usually, the quantification of modifications is performed at the peptide level, whereas the quantification of interest for the biologist is at the amino acid site level. The quantitative measurement from a peptide can be narrowed down to a residue only if it is the only modifiable residue in the peptide. If there are several modifiable residues in a peptide, qualitative MS/MS analysis is classically carried out to operate this narrowing down of candidate residues. However, the localization of modification sites is a more difficult task than the detection and identification of modified peptides. This is because MS/MS localization requires a good quality spectrum composed of relatively complete series of fragment ions to unambiguously identify the specific modified amino acid. The low ionization efficiency and/or low abundance of some modified peptides lead to low signal to noise ratios which affect their selection for fragmentation and yield low fragment ion intensities. In some cases, current interpretation algorithms may actually lead to mislocalization and thus misquantification of the modified site, especially in the case of modification at different positions of co-eluting peptides isomers, i.e. isobaric modified peptides. There is thus great advantage in exploiting MS data for this purpose. The specifics of the experiment presented here led us to propose an original angle to address the problem.

## **The segmentation approach**

Figure 6 shows that, in spite of our efforts to get complete cleavage with maximum sequence coverage, MC still occurred at a significant level, especially in the case of chymotrypsin due to the close proximity of sites. This leads to local ladders which differ by a few residues, sometimes as little as 1, 2, or 3. Modified residues could be contained in these short stretches of sequence difference. This means that there is precious, unexploited information contained



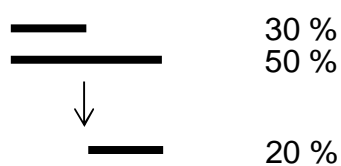
in the dataset. To extract f bound data from groups of ladder peptides, we developed a powerful tool, which we call the segmentation approach, to refine at the MS level the quantification of the extent of modification, with the ultimate goal of narrowing down the measurement to a single modifiable amino-acid.

Peptides couples can be classified into four categories as shown in Scheme I.



Scheme I

Only class A pairs, corresponding to fully overlapping peptides with a common end, can be used for segmentation. The principle of segmentation is that the difference in f bound of two peptides with a common end is attributable to modification of the sequence that differs between them. It is based on the identification and use of class A peptides, within a given proteolysis and/or from different proteolyses, to deduce a f bound value of the complementary sequence as shown in Scheme 2.



Scheme II

If segmentation generates more than one f bound per segment of sequence, these fbound values are averaged and assigned to this segment for the next cycle. The segmentation process

can be iterated by reusing the complementary sequence for segmentation. The segmentation process stops when only B, C and D class peptides remain.

To prevent the generation of artificially high f bounds, all artefacts caused by MCm and SC are removed manually from the dataset before launching the segmentation routine.

Figure 8 shows the segmentation results for each proteinase. The graph shows segments that are significantly above threshold and can thus be considered as modified, segments with f bounds that are not significantly different from zero, and nonobserved sequences of the protein. Propagation of random error is expected to occur as a result of the math used in the segmentation process. The propagated error will depend on the number of applied segmentation cycles that generate subtractions and sometimes averages, and thus vary along the sequence. An estimation of propagated error was calculated based on the location and number of observed segmentation cycles for each protease. Applying  $i$  iterations of the segmentation cycle generates a  $\sqrt{(1 + i)}$  error multiplier, while applying  $j$  iterations of the average calculation generates a  $\sqrt{\frac{1}{2^j}}$  error multiplier. The combined error propagation given in equation 8 was applied to each proteolysis dataset, using a conservatively low number of averages as these tend to reduce the error (Table 1).

$$\sigma (i,j) = \sqrt{(1 + i)} \cdot \sqrt{\frac{1}{2^j}} \cdot \sigma_0 \quad (8)$$

Table 1. Calculated  $\sigma$  (i,j) errors propagated through the segmentation process (Equation 8)

proteinase	$\sigma_0$	segmentation cycles i	averaging operations j	$\sigma$ (i,j)
Asp-N	0.0164	1	0	0.0232
chymotrypsin	0.0238	3	1	0.0337
trypsin	0.0232	1	0	0.0328

The threshold of significant deviation from null f bound was set at twice the propagated error, and the resulting numbers of +/- 0.046, 0.067 and 0.066 for Asp-N, chymotrypsin and trypsin, respectively, were reported as 95% confidence bands on the graphs.

For Asp-N, four sequences appear as modified: [35-55], [78-95], [128-133] and [144-173]. No information is present for [2-3] and [70-77].

For chymotrypsin, two segments have a clearly positive f bound: [69-84] and [132-149]. For the [69-84] sequence, segmentation consolidates the information which was initially distributed over three peptides: [65-84], [67-84] and [69-84] (Figure 5). Nonobserved sequences are [85-87], [121-131] and [150-154].

For trypsin, [78-80] is the only segment that comes above the threshold as modified. The [78-80] sequence clearly shows the benefit of the segmentation approach: the initial information distributed over three peptides: [63-76], [63-77] and [63-80] was refined down to three residues. Nonobserved sequences are [2-7], [48-62] and [142-156].

The data in Figure 8 was combined in Figure 9 to provide a synthetic view of the results. Figure 9 shows a number of B and C class overlapping modified pairs which cannot be easily interpreted. Before applying a final processing step to attempt to improve resolution in the

sequence dimension however, the  $f$  bounds obtained after segmentation can be used to test for coherence in the whole dataset.

### **Determination of the total modified stoichiometric fraction**

Determination of the total modified stoichiometric fraction is performed at the protein level. It corresponds to the sum of the extent of modification at each modified sites. Mass spectrometry is one of the few methods that can simultaneously measure the amplitude of a proteoform signal and calculate the modification stoichiometry in terms of number of sites per protein chain for a given modification type. Contrary to the commonly used total modified fraction, which corresponds to the proportion of modified chains in the total pool of protein chains, the total modified stoichiometric fraction (TMSF) takes into account the stoichiometry of modification (Equation 9):

Total modified stoichiometric fraction =

$$\frac{\sum (\text{number of modification of a proteoform } n \times \text{signal of a proteoform } n)}{\text{signal of total protein}} \quad (9)$$

If each site is 100% modified, i.e. if saturation is attained at all sites, the TMSF value tends to the total integer stoichiometry of the modification fixed on the protein. For example, if a protein has 3 modification sites capable of reaching saturation, TFMS tends to 3, whereas the total modified fraction tends to 1

In most cases, the detrimental or beneficial effect of a modification on the ionization efficiency of a whole protein is greatly diluted by the large number of protonation sites on the protein. When using TMSF as a direct quantitative measurement of protein modification, the approximation that different proteoforms have the same ionization efficiency can thus be made.

The point of measuring TMSF is that it can then be compared to the sum of individual f bound measured by parallel quantification in a bottom-up approach such as the subtractive method. The comparison gives a measure of the coherence of the extent of modification found on the whole protein with the data obtained at the peptide level.

The hPEBP1 and locostatin partners were incubated for 5 h at 37°C before removing the excess of locostatin by micro gel filtration. The covalent complex was then analyzed by MALDI-TOF MS (Figure 10). Figure 10 shows a peak at +243.8 Da corresponding to the hPEBP1-locostatine complex which bears a theoretical mass increment of 245.27 Da compared to locostatin. An additional peak at +87.6 Da is detected and corresponds to butyric acid, a hydrolysis product of locostatin we previously identified as butyric acid (theoretical mass increment of 86.09 Da) and which is also covalently attached to hPEBP1 (Gabant et al.). Taking into account the intensity of the hPEBP1-locostatin and hPEBP1-butyric acid complexes, a total modified stoichiometric fraction of 0.50 is found. In this experiment, the <sup>15</sup>N labeled internal standard was not added because the corresponding peak partially overlaps with the peak of <sup>14</sup>N hPEBP1-locostatin complex, thus preventing its detection and quantification.

By comparing the extent of modification on the whole protein measured as TMSF with the sum of individual f bounds for segments for each proteinase after segmentation, a percent recovery of modification at the peptide level is obtained (Table 2)

Table 2. Comparison of the extent of modification on the whole protein measured as TMSF with the sum of individual f bounds for segments for each proteinase after segmentation.

Proteinase	$\Sigma f_{\text{bounds}}$	TMSF	% recovery
AspN proteolysis	0.37	0.50	74
Chymotrypsin proteolysis	0.22	0.50	44
Trypsin proteolysis	0.12	0.50	24

The subtractive method is not sensitive to the hydrolysis of locostatin since it only measures the nonmodified peptides. Thus  $\Sigma f_{\text{bounds}}$  is expected to be equal to TMSF. However, only 24 to 74% of TMSF is recovered at the peptide level after segmentation. The remainder may correspond to modification that is widely distributed over slightly modified segments and/or nonobserved sequences. The origin of the difference of percent recovery between proteinases is not clearly identified yet, but it may be due, at least in part, to the difference in average length of the proteolytic peptides resulting in generally higher f bounds per segment.

### Local Minimum

As mentioned above, Figure 9 still shows a number of class B and C pairs of overlapping segments between proteolyses, which lead to residual ambiguity. The reason for this is the presence of nonoverlapping sequences on both sides of at least one of the peptides. Contrary to class A pairs, any f bound difference between peptides cannot be unambiguously assigned to a modification location on either side of the sequence in common. While f bound difference calculations are meaningless for class B and C pairs, one piece of information can be derived from the comparison of f bounds in the common sequence, and used to further refine the data as shown hereafter.

The principle of the local minimum is that in the sequence that is common to two or more partially overlapping class B or C pairs or groups, the lowest f bound, i.e. the local minimum (LocMin), gives the highest possible value of the extent of modification in that region. If one of the segments has a higher value, logic dictates that it can only be due to a modification in a nonoverlapping region. As stated above however, the exact f bound of this modification cannot be calculated from the difference since these are not a class A pair. The Local Minimum concept is thus simply based on the selection, for each residue, of the minimum value of f bound found for all segments. Results with LocMin treatment applied on the 3 combined proteolyses data are shown in Figure 11. A f bound histogram as a function of sequence is shown in Figure 12. The propagated  $\sigma$  calculated for Figure 8 was used for each protease to assign an error bar to the LocMin of each segment in the figure. Table 3 summarizes the observed segments positively identified as modified, along with modifiable amino acids they contain and the observation of a MCm or SC in the segment.

Table 3. Final segments identified as modified after LocMin narrowing of modified sequences. The italicized sites in the table correspond to f bounds that are higher than  $\sigma$  yet within  $2\sigma$  of zero. The non-italicized sites are significantly different from zero.

Sequence segment	f bound after LocMin	Modifiable sites	Site leading to MCm or SC due to a modification	Unambiguously identified site
[2-3]	0.047	<i>N-terminal</i>		<i>N-terminal</i>
[78-80]	0.084	Lys80	Lys80	Lys80
[81-87]	0.039	<i>Tyr81, His85, His86</i>		
[107-113]	0.034	<i>Ser109, Lys113</i>		
[144-149]	0.111	His145, Lys148		
[150-154]	0.128	Lys150, Ser153		
[155-156]	0.059	<i>Lys156</i>		<i>Lys156</i>

Care should be taken in the interpretation of LocMin f bounds, as class B pairs generate a falsely high f bound on at least one of the nonoverlapping sides. This is represented by a small linking line between the residual sequences in Figure 11. In this dataset, an example of such

potential problem areas are [4-8] and [13-17] with  $f$  bounds around zero. LocMin  $f$  bounds generated by class C groups can also be overestimated, because modifications leading to these  $f$  bounds can be located outside the common sequence for all segments in the group. There may even be cases where the common sequence does not bear any modification. So, while LocMin minimizes errors by representing the upper boundary for the true  $f$  bound in the common sequence, this minimization may still not be enough, preventing unambiguous assignment of the measured  $f$  bound to a modifiable residue in that sequence. Examples of LocMin of class C groups leading to potentially overestimated  $f$  bounds concern sequences [81-87] and [144-149] in this dataset.

The most striking aspect of the parallel quantification of modifications on hPEBP1 is that no single residue appears with an  $f$  bound value that is clearly superior to any other. This result explains why, in spite of years of the best efforts of several teams (Shemon et al., 2009; Beshir et al., 2011), including ours, the locostatin binding site was never clearly identified.

Table 3 shown seven narrowed segments. Three segments contain only one modifiable residue: [2-3], [78-80] and [155-156]. The  $f$  bounds observed in these areas can thus be attributed to these sites, i.e. the N-terminal site, Lys80 and Lys156, with some level of confidence. Lys 80 is the only residue of the three with a  $f$  bound that is significantly above zero. For the remaining segments [81-87], [107-113], [144-149] and [150-154], up to two or three residues may be modified, thus preventing a quantitative measurement at the amino acid level.

Lys47 was identified as modified in the [42-55] segment based on the generation of a MCm with trypsin cleavage. As the segmentation process slightly increases the global error, the final threshold of significance is increased. After the segmentation and LocMin processes, the



f bound associated with this segment indicates a low level of modification at this site, below the threshold of significance.

One of the two modifiable residues in the [150-154] segments is Ser153, a residue which is pivotal to signalization since its phosphorylation by PKC switches PEBP1 from the Raf1-MAP kinase pathway to GRK2 (Deiss et al., 2012). Modification by locostatin followed by its hydrolysis into butyric acid would introduce a negative charge in this area. If located on Ser153, the negative charge could mimic phosphorylated Ser153. Downstream signaling of this "always-on phosphorylation" switch may then favor anti-migratory effects. On the other, if the negative charge of butyrate was located on another residue near Ser153, it could prevent GRK2 kinase binding to Ser153, preventing the switch away from Raf1, thus re-enforcing the inhibition of the Raf1 MAP kinase pathway. This could indirectly trigger Aurora with positive anti-tumoral and anti-metastasis effects. It could also act through hPEBP1's ability to modulate the  $\text{nF-}\kappa\text{B}$  pathway (Yeung et al., 2001). Interestingly, the [81-87] segment contains Tyr81 and His86, two residues that are part of the evolutionarily conserved ligand-binding pocket of PEBP1 (Tavel et al., 2012). Beshir et al., in the proceedings of a workshop (2011), claimed to have identified His86 as the primary target site of locostatin, while suggesting that other residues were alkylated. His86 was singled out because it is a highly conserved residue at the bottom of the anion pocket. However, the authors used a high concentration of hPEBP1, leading to precipitation in the presence of excess locostatin. The MS results were obtained after resolubilization of the insoluble fraction of hRKIP/hPEBP1-locostatin complex. The precipitation of the protein obviously indicates a profound destruction of the protein structure, which casts serious doubt about the biological relevance of the result. Despite the lack of credible evidence and the fact that locostatin is not an anion, these results were then used in support of a computational simulation for locostatin binding in the anion pocket (Rudnitskaya

et al., 2012). Our results show that the [81-87] segment modification cannot be distinguished from the background.

In light of the present quantification study of locostatin modifications on hPEBP1, both functional and structural studies could be undertaken, with the aim to better characterize the binding site responsible for the anti-metastatic activity. The rational design of activators of hPEBP1 natural antimetastasis effect would greatly benefit from these developments.

The present parallel quantification method can measure  $f$  bounds with less than 0.07 standard deviation. The reliability of the method thus depends on modifications generating sufficiently high  $f$  bounds. In the present study, the individual  $f$  bound values are relatively low due to the number of parallel sites that were modified, while TMSF was kept at a low 0.5. For systems with TMSF over 1 and only 2 or 3 sites with no background, the relative error should be comparatively low. The method is also essentially modular: segmentation could be omitted if none of the proteolyses generate MCs. Proteolysis results can be combined before or after segmentation, and LocMin can be applied or not depending on the presence of class B and C groups.

## **Conclusions**

The subtractive method developed here is well adapted to address some of the common issues associated with the parallel quantitation of modifications. Because it deals only with nonmodified peptides, it circumvents problems linked to ionization efficiencies and the progressive loss of integrity of the modification after covalent addition, as was the case for locostatin, and poor solubility of modified peptides becomes a non-issue. In the case of induced modifications, it can also work without prior knowledge of the types of modified

residue, provided that a sufficient set of proteases is used to cover the whole sequence and that cleavages are achieved at least once between each modified residue. With a slightly different front end, it could be applied more generally to PTM quantification with the appropriate enzymatic removal tools.

## Materials and Methods

**Reagents.** Unless otherwise stated, all chemicals used in this study were obtained from Sigma (St. Louis, MO). Locostatin was purchased from Acros Organics. Ammonium acetate and calcium chloride ( $\text{CaCl}_2$ ) were procured from Merck (Darmstadt, Germany). Glycine was purchased from Eurobio (Courtaboeuf, France). Endoproteinases Asp-N, chymotrypsin and trypsin (sequencing grade) were from Roche. Alpha-cyano-4-hydroxycinnamic acid was obtained from Bruker Daltonics (Bremen, Germany). Acetonitrile and isopropanol were procured from Carlo Erba (Milan, Italy). Formic acid 90% (FA) and trifluoroacetic acid (TFA) were purchased from Fisher (Loughborough, UK). All solvents and buffers were prepared using 18 M $\Omega$  purified water (MilliQ reagent grade system, Millipore).

Purification of recombinant hPEBP1. Recombinant human Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1) was cloned and overexpressed in BL21DE3 *E.coli* as previously described by Schoentgen and coworkers (Atmanene et al., 2009). The protein was purified without tag using the protocol originally developed by Bernier and Jolles (Bernier and Jollès, 1984) consisting of QAE Sephadex A-50 chromatography, isoelectrofocusing and Blue-Sepharose chromatographic steps, and concentrated as previously described (Jaquillard et al., 2012) (Jaquillard et al 2012). The purified protein was stored in milliQ water with 2 equivalents of TCEP at -80°C.

**Interaction of hPEBP1 with locostatin.** A solution of hPEBP1 protein at 7.3  $\mu\text{M}$  was incubated with or without 1 mM locostatin for 4 h 30 at 37°C in a buffer consisting of 100 mM HEPES, pH 7.7, and 2% acetonitrile. The control and experiment samples were then spiked with an  $^{15}\text{N}$  labeled unmodified hPEBP1 solution at 5  $\mu\text{M}$ . After internal standard

addition, the excess locostatin was removed by micro gel filtration as follows. Spin columns with a bed volume of 75  $\mu$ L were used. Efficient elimination of the conditioning buffer from the gel filtration phase required 5 cycles of equilibration with 50  $\mu$ L volumes of incubation buffer. The sample was loaded onto the spin column and the collection performed immediately with a 30 s centrifugation at 1000 g.

**Denaturation and Enzymatic Proteolysis.** At this stage the sample was split into three tubes in order to perform parallel proteolyses. Proteins were thermally denatured for 10 min at 60°C in a solution of 1 M urea, 100 mM glycine, and 20% acetonitrile. The glycine was added during this step to prevent urea-induced carbamylation of Lysines, Arginines, Cysteines, and the N-terminal amine. Denatured proteins were then proteolyzed in parallel for 2 h a high enzyme:substrate (E:S) ratio using optimized conditions for Asp-N, chymotrypsin and trypsin. For Asp-N, the reaction was carried out at 37°C in 100 mM ammonium acetate, pH 7.0, with an E:S ratio of 1:10. For trypsin, the cleavage was carried out at 37°C in a solution of 100 mM ammonium acetate-ammonium bicarbonate, pH 7.5, with an E:S ratio of 1:2. For chymotrypsin, proteolysis was carried out at 25°C in the same buffer with 10 mM CaCl<sub>2</sub>, with an E:S ratio of 1:10. Proteolysis reactions were stopped by addition of TFA at a final concentration of 0.1% and immediate desalting/concentration on a C18 ZipTip from Millipore (Billerica, MA, USA). The volume of TFA solution was adjusted so as to dilute acetonitrile as well, to a final concentration of 2.5%. This low concentration of organic solvent ensures efficient binding of the peptides on the reversed phase of the ZipTip.

**Mass spectrometry.** Whole proteins were analyzed by MALDI-TOF MS. The matrix solution consisted of a saturated solution of 4-hydroxy- $\alpha$ -cyano-cinnamic acid (4HCCA) in 3:1:2 formic acid:water:isopropanol. Proteins in the micromolar range were prepared by

twenty-fold dilution into the matrix solution. The analyte-matrix samples were spotted onto a gold-plated sample probe using the ultra-thin layer method as described (Cadene and Chait, 2000; Gabant and Cadene, 2008). Spots were washed with 0.1% TFA before acquisition. Analyses were performed using an Ultraflex I mass spectrometer (Bruker Daltonics, Germany) equipped with a 337 nm nitrogen laser and a gridless delayed extraction ion source. An accelerating voltage of 25 kV was used and delay optimized at 250 ns to achieve a mass resolution greater than 1000 over the mass range of interest (10000–20000 Da). A deflection of matrix ions up to 800 Da was applied to prevent detector saturation. Spectra were acquired in the linear positive ion mode by accumulation of 1000-1200 laser shots. Calibration was performed externally using a neighboring spot with a solution of apomyoglobin and cytochrome c. The instrument was controlled using Bruker FlexControl 3.3 software and MALDI-TOF-MS spectra were processed using FlexAnalysis 3.3 software from Bruker Daltonics.

Proteolytic peptides were analyzed by nanoUltraHPLC-nanoESI UHR-QTOF MS. Experiments were performed using an UltiMate<sup>TM</sup> 3000 NanoRSLC System (Dionex, Sunnyvale, CA) connected to a Bruker MaXis UHR-QTOF 4 GHz mass spectrometer equipped with an online nano-ESI ion source. The LC-MS setup was controlled by Bruker Hystar<sup>TM</sup> software version 3.2. Peptides were pre-concentrated online on a Dionex Acclaim PepMap100 C18 reverse-phase pre-column (inner diameter 100  $\mu\text{m}$ , length 2 cm, particle size 5  $\mu\text{m}$ , pore size 100  $\text{\AA}$ ), and separated on a nanoscale Acclaim Pepmap100 C18 column (inner diameter 75  $\mu\text{m}$ , length 25 cm, particle size 2  $\mu\text{m}$ , pore size 100  $\text{\AA}$ ) at a flow rate of 450 nL/min using a 2–35% gradient of acetonitrile in 0.1% formic acid. Chromatographic peaks were about 3-4 s at Full Width at Half Maximum (FWHM), corresponding to a peak width at the base around 8-10 s. Mass spectrometer scans were set at a frequency of 1 Hz in MS mode

only. These settings ensure that there is around 10 data points per extracted ion chromatogram (XIC) allowing for an accurate determination of the area under the curve (AUC). Mass spectra were acquired in positive ion mode from  $m/z$  50–3000. Lock mass calibration was performed on  $m/z$  622 (Hexakis(2,2-Difluoroethoxy)Phosphazene; CAS NO: 186817-57-2) and 1222 (Hexakis(1H, 1H, 4H-hexafluorobutyloxy)Phosphazine; CAS NO: 186406-47-3).

**Skyline MS1 Quantification.** MS data from nanoUltraHPLC-UHR-QTOF MS were processed using the open source software Skyline 2.1 (Schilling et al., 2012) to calculate the area under the curve for each peak from XIC. XICs were constructed for all isotopic ion peaks of each peptide with a base peak threshold of 1% and a  $m/z$  tolerance of 5 ppm. Each XIC was manually checked in terms of retention time alignment between light and heavy peptides as well as across acquisition replicates. The coefficient of correlation between the observed and expected isotopic distributions was below 0.95 and 0.89 for light and heavy peptides, respectively. The area for each charge state of each peptide was exported to .csv format and used as an input for our in-house software dedicated to the quantification of modification at each site.

## Bibliography

- Atmanene, C.; Laux, A.; Glattard, E.; Muller, A.; Schoentgen, F.; Metz-Boutigue, M.-H.; Aunis, D.; Van Dorselaer, A.; Stefano, G. B.; Sanglier-Cianfèrani, S.; et al. Characterization of Human and Bovine Phosphatidylethanolamine-Binding Protein (PEBP/RKIP) Interactions with Morphine and Morphine-Glucuronides Determined by Noncovalent Mass Spectrometry. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2009**, *15*, BR178–BR187.
- Bernier, I.; Jollès, P. Purification and Characterization of a Basic 23 kDa Cytosolic Protein from Bovine Brain. *Biochim. Biophys. Acta* **1984**, *790*, 174–181.
- Beshir, A. B.; Argueta, C. E.; Menikarachchi, L. C.; Gascón, J. A.; Fenteany, G. Locostatin Disrupts Association of Raf Kinase Inhibitor Protein With Binding Proteins by Modifying a Conserved Histidine Residue in the Ligand-Binding Pocket. *Forum Immunopathol. Dis. Ther.* **2011**, *2*, 47–58.
- Cadene, M.; Chait, B. T. A Robust, Detergent-Friendly Method for Mass Spectrometric Analysis of Integral Membrane Proteins. *Anal. Chem.* **2000**, *72*, 5655–5658.
- Ciccimaro, E.; Hanks, S. K.; Yu, K. H.; Blair, I. A. Absolute Quantification of Phosphorylation on the Kinase Activation Loop of Cellular Focal Adhesion Kinase by Stable Isotope Dilution Liquid Chromatography/mass Spectrometry. *Anal. Chem.* **2009**, *81*, 3304–3313.
- Deiss, K.; Kisker, C.; Lohse, M. J.; Lorenz, K. Raf Kinase Inhibitor Protein (RKIP) Dimer Formation Controls Its Target Switch from Raf1 to G Protein-Coupled Receptor Kinase (GRK) 2. *J. Biol. Chem.* **2012**, *287*, 23407–23417.
- Gabant, G.; Cadene, M. Mass Spectrometry of Full-Length Integral Membrane Proteins to Define Functionally Relevant Structural Features. *Methods San Diego Calif* **2008**, *46*, 54–61.
- Gabant, G.; Beaufour, M.; Schoentgen, F.; Cadene, M. A Detailed Characterization of the Interaction between the PEBP/RKIP Protein and Locostatin, a Potential Antimetastatic Lead. *56th ASMS Conference on Mass Spectrometry and allied topics, Denver, USA, 1-5 June 2008*.
- Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6940–6945.
- Gropengiesser, J.; Varadarajan, B. T.; Stephanowitz, H.; Krause, E. The Relative Influence of Phosphorylation and Methylation on Responsiveness of Peptides to MALDI and ESI Mass Spectrometry. *J. Mass Spectrom. JMS* **2009**, *44*, 821–831.
- Jaquillard, L.; Saab, F.; Schoentgen, F.; Cadene, M. Improved Accuracy of Low Affinity Protein-Ligand Equilibrium Dissociation Constants Directly Determined by Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 908–922.
- Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem.* **2010**, *2*, 949–964.
- Klysik, J.; Theroux, S. J.; Sedivy, J. M.; Moffit, J. S.; Boekelheide, K. Signaling Crossroads: The Function of Raf Kinase Inhibitory Protein in Cancer, the Central Nervous System and Reproduction. *Cell. Signal.* **2008**, *20*, 1–9.
- Lee, H. C.; Tian, B.; Sedivy, J. M.; Wands, J. R.; Kim, M. Loss of Raf Kinase Inhibitor Protein Promotes Cell Proliferation and Migration of Human Hepatoma Cells. *Gastroenterology* **2006**, *131*, 1208–1217.
- Mc Henry, K. T.; Ankala, S. V.; Ghosh, A. K.; Fenteany, G. A Non-Antibacterial Oxazolidinone Derivative That Inhibits Epithelial Cell Sheet Migration. *Chembiochem Eur. J. Chem. Biol.* **2002**, *3*, 1105–1111.



- Olsen, J. V.; Mann, M. Status of Large-Scale Analysis of Post-Translational Modifications by Mass Spectrometry. *Mol. Cell. Proteomics MCP* **2013**, *12*, 3444–3452.
- Olsen, J. V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M. L.; Jensen, L. J.; Gnad, F.; Cox, J.; Jensen, T. S.; Nigg, E. A.; et al. Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy during Mitosis. *Sci. Signal.* **2010**, *3*, ra3.
- Rappaport, S. M.; Smith, M. T. Environment and Disease Risks. *Science* **2010**, *330*, 460–461.
- Rubino, F. M.; Pitton, M.; Di Fabio, D.; Colombi, A. Toward an “omic” Physiopathology of Reactive Chemicals: Thirty Years of Mass Spectrometric Study of the Protein Adducts with Endogenous and Xenobiotic Compounds. *Mass Spectrom. Rev.* **2009**, *28*, 725–784.
- Rudnitskaya, A. N.; Eddy, N. A.; Fenteany, G.; Gascón, J. A. Recognition and Reactivity in the Binding between Raf Kinase Inhibitor Protein and Its Small-Molecule Inhibitor Locostatin. *J. Phys. Chem. B* **2012**, *116*, 10176–10181.
- Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; et al. Platform-Independent and Label-Free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline: APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION. *Mol. Cell. Proteomics* **2012**, *11*, 202–214.
- Shemon, A. N.; Eves, E. M.; Clark, M. C.; Heil, G.; Granovsky, A.; Zeng, L.; Imamoto, A.; Koide, S.; Rosner, M. R. Raf Kinase Inhibitory Protein Protects Cells against Locostatin-Mediated Inhibition of Migration. *PloS One* **2009**, *4*, e6028.
- Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.
- Steen, H.; Jebanathirajah, J. A.; Springer, M.; Kirschner, M. W. Stable Isotope-Free Relative and Absolute Quantitation of Protein Phosphorylation Stoichiometry by MS. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3948–3953.
- Tavel, L.; Jaquillard, L.; Karsisiotis, A. I.; Saab, F.; Jouvensal, L.; Brans, A.; Delmas, A. F.; Schoentgen, F.; Cadene, M.; Damblon, C. Ligand Binding Study of Human PEBP1/RKIP: Interaction with Nucleotides and Raf-1 Peptides Evidenced by NMR and Mass Spectrometry. *PloS One* **2012**, *7*, e36187.
- Torosantucci, R.; Brinks, V.; Kijanka, G.; Halim, L. A.; Sauerborn, M.; Schellekens, H.; Jiskoot, W. Development of a Transgenic Mouse Model to Study the Immunogenicity of Recombinant Human Insulin. *J. Pharm. Sci.* **2014**, *103*, 1367–1374.
- Wu, R.; Haas, W.; Dephoure, N.; Huttlin, E. L.; Zhai, B.; Sowa, M. E.; Gygi, S. P. A Large-Scale Method to Measure Absolute Protein Phosphorylation Stoichiometries. *Nat. Methods* **2011**, *8*, 677–683.
- Yeung, K. C.; Rose, D. W.; Dhillon, A. S.; Yaros, D.; Gustafsson, M.; Chatterjee, D.; McFerran, B.; Wyche, J.; Kolch, W.; Sedivy, J. M. Raf Kinase Inhibitor Protein Interacts with NF-kappaB-Inducing Kinase and TAK1 and Inhibits NF-kappaB Activation. *Mol. Cell. Biol.* **2001**, *21*, 7207–7217.
- Zeng, L.; Imamoto, A.; Rosner, M. R. Raf Kinase Inhibitory Protein (RKIP): A Physiological Regulator and Future Therapeutic Target. *Expert Opin. Ther. Targets* **2008**, *12*, 1275–1287.
- Zhang, G.; Ueberheide, B. M.; Waldemarson, S.; Myung, S.; Molloy, K.; Eriksson, J.; Chait, B. T.; Neubert, T. A.; Fenyo, D. Protein Quantitation Using Mass Spectrometry. *Methods Mol. Biol. Clifton NJ* **2010**, *673*, 211–222.
- Zhu, S.; Mc Henry, K. T.; Lane, W. S.; Fenteany, G. A Chemical Inhibitor Reveals the Role of Raf Kinase Inhibitor Protein in Cell Migration. *Chem. Biol.* **2005**, *12*, 981–991.

# Figures

**Figure 1. Quantification workflow.** A. Preparation of the samples and their controls for quantification of *de facto* or induced modifications. B. Production of peptides and nanoLC-HRMS quantification of the unmodified peptides.

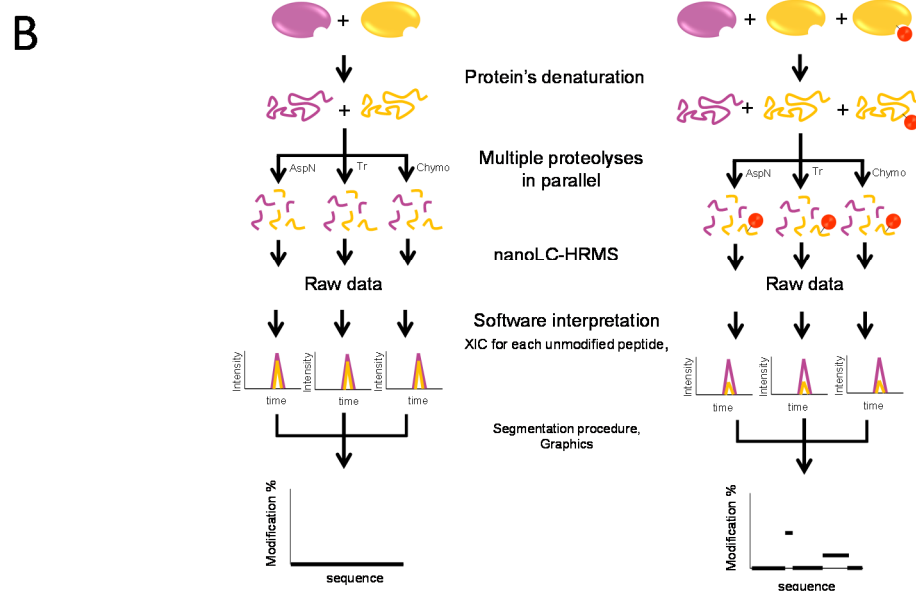
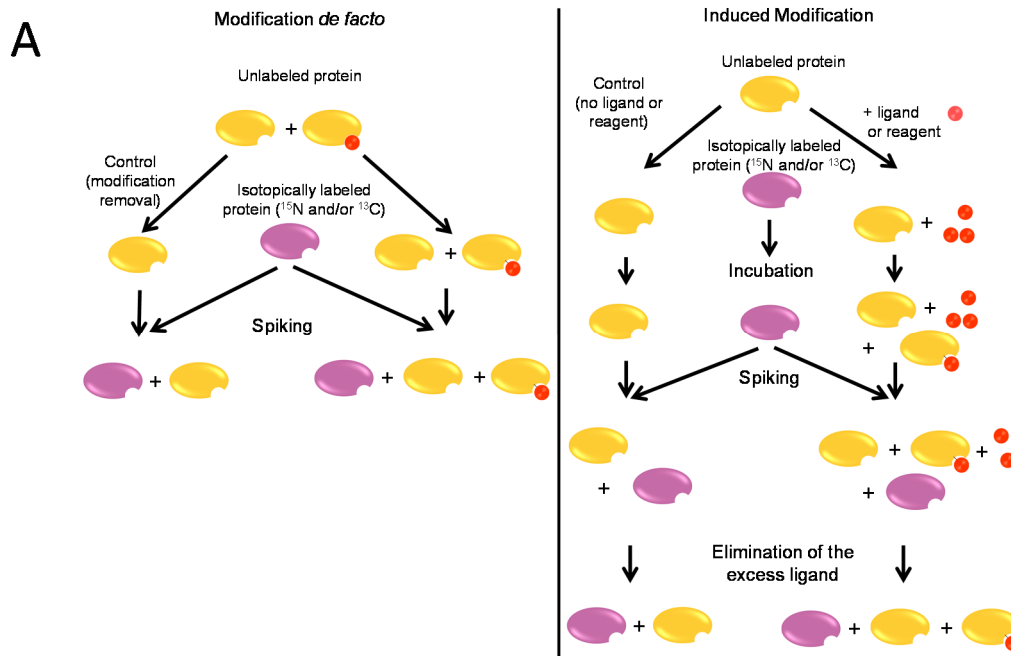


Figure 2. **Metabolic incorporation measurement on hPEBP1.** A 1:1 mixture of  $^{14}\text{N}/^{15}\text{N}$  hPEBP1 protein solutions was analyzed by MALDI-TOF MS.

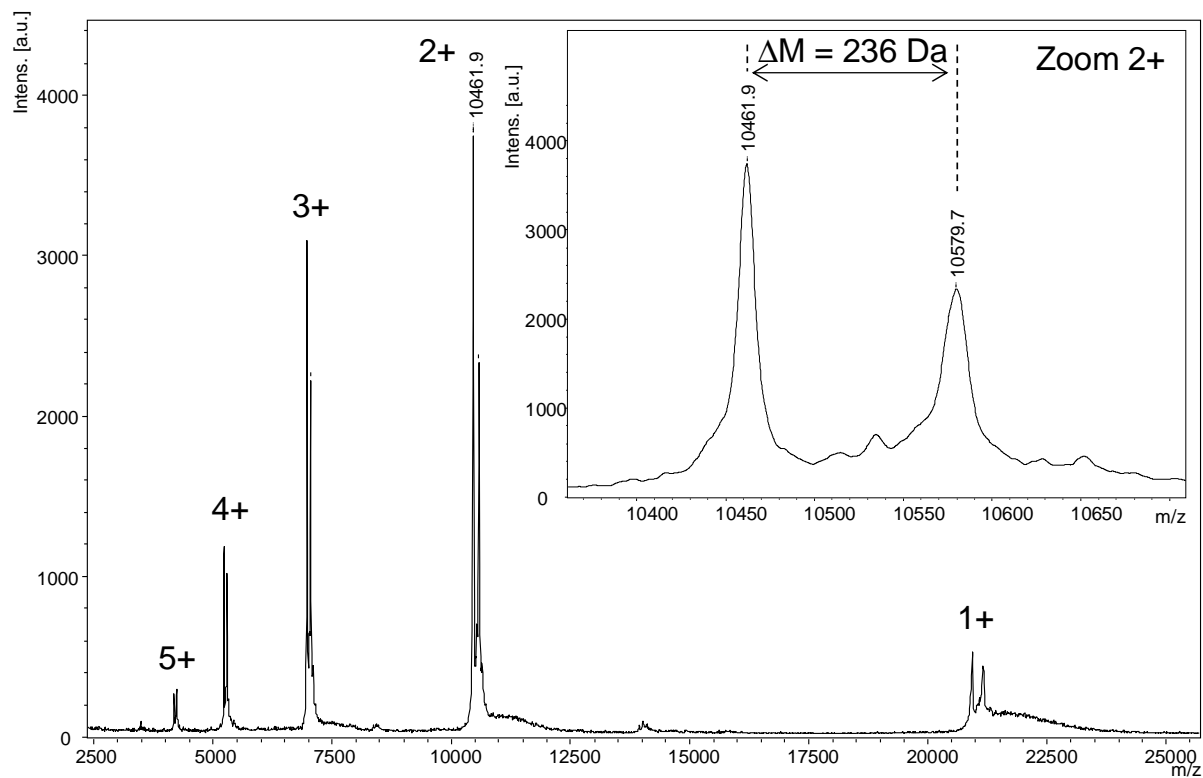


Figure 3. **Influence of the number of isotopic distribution peaks used for area calculation on the quantification result.** The  $^{14}\text{N}/^{15}\text{N}$  ratios for each peptide were calculated using Skyline software and plotted as a function of  $m/z$ . A. Asp-N peptides from a 1:1 mixture of  $^{14}\text{N}/^{15}\text{N}$  hPEBP1 protein solutions were analyzed in decaplicate by nanoUltraHPLC-UHR-QTOF MS. B. Theoretical relative isotopic abundances of the Asp-N peptides of hPEBP1. In blue : the first three peaks of the monoisotopic distribution are considered. In red : The whole monoisotopic distribution (isotope > 1 %) is considered.

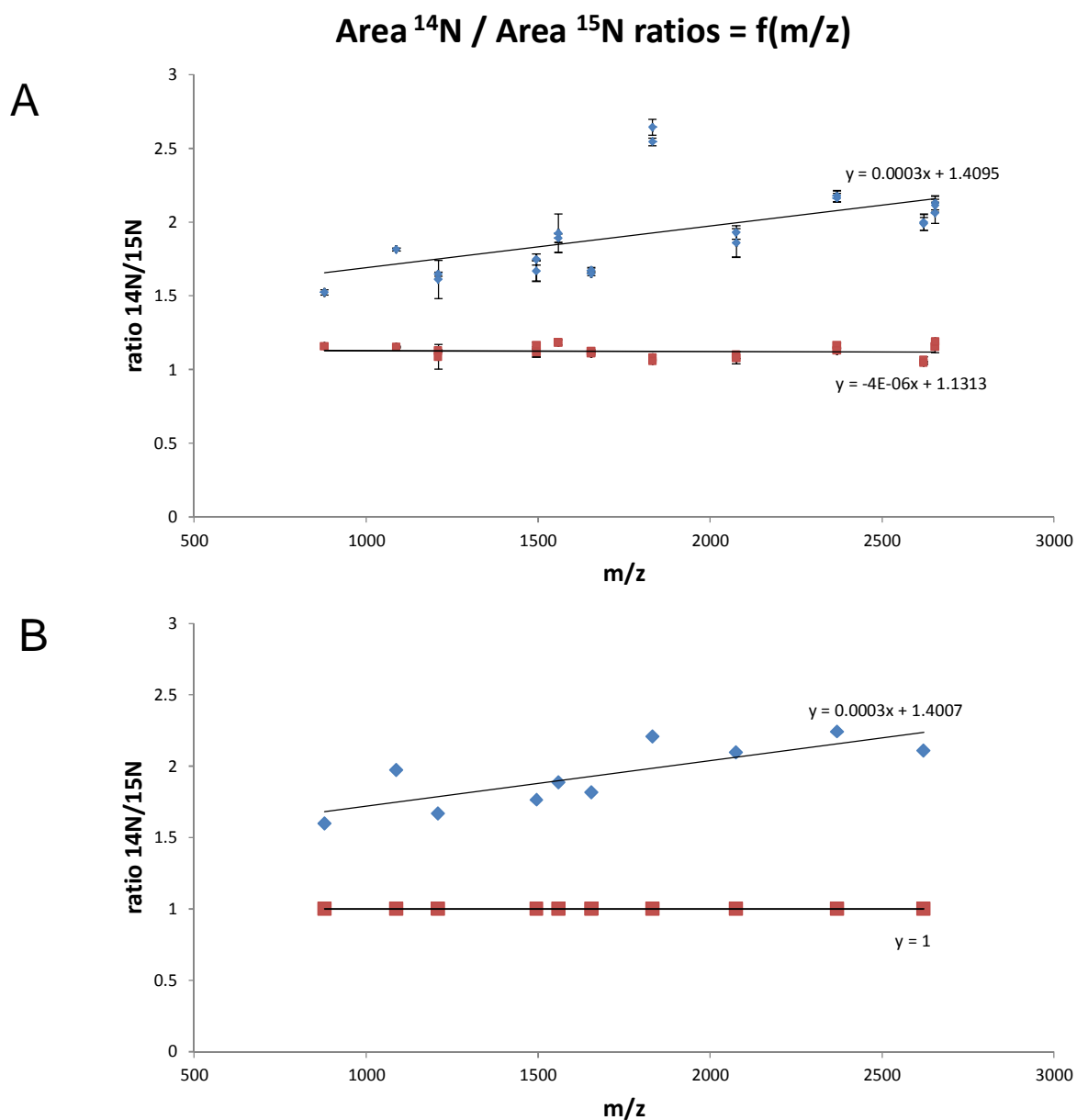


Figure 4. Local epsilon correction of the f bound of hPEBP1 Asp-N peptides in the control. A. Dispersion of the f bound before correction. B. Dispersion of the f bound after local epsilon correction.

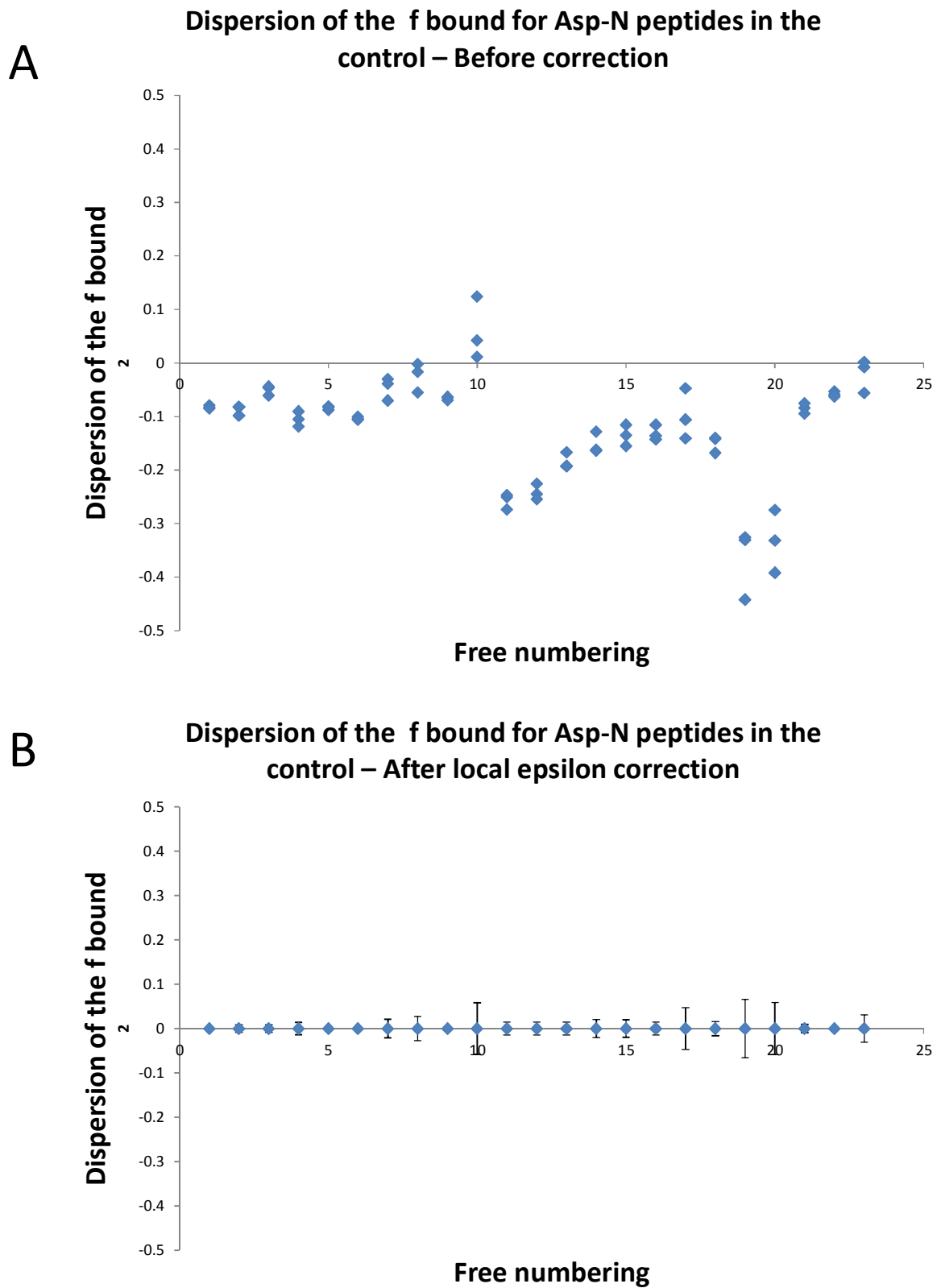


Figure 5. Graphical representation of parallel quantification of peptide modifications for the hPEBP1-locostatin t=5h reaction time.

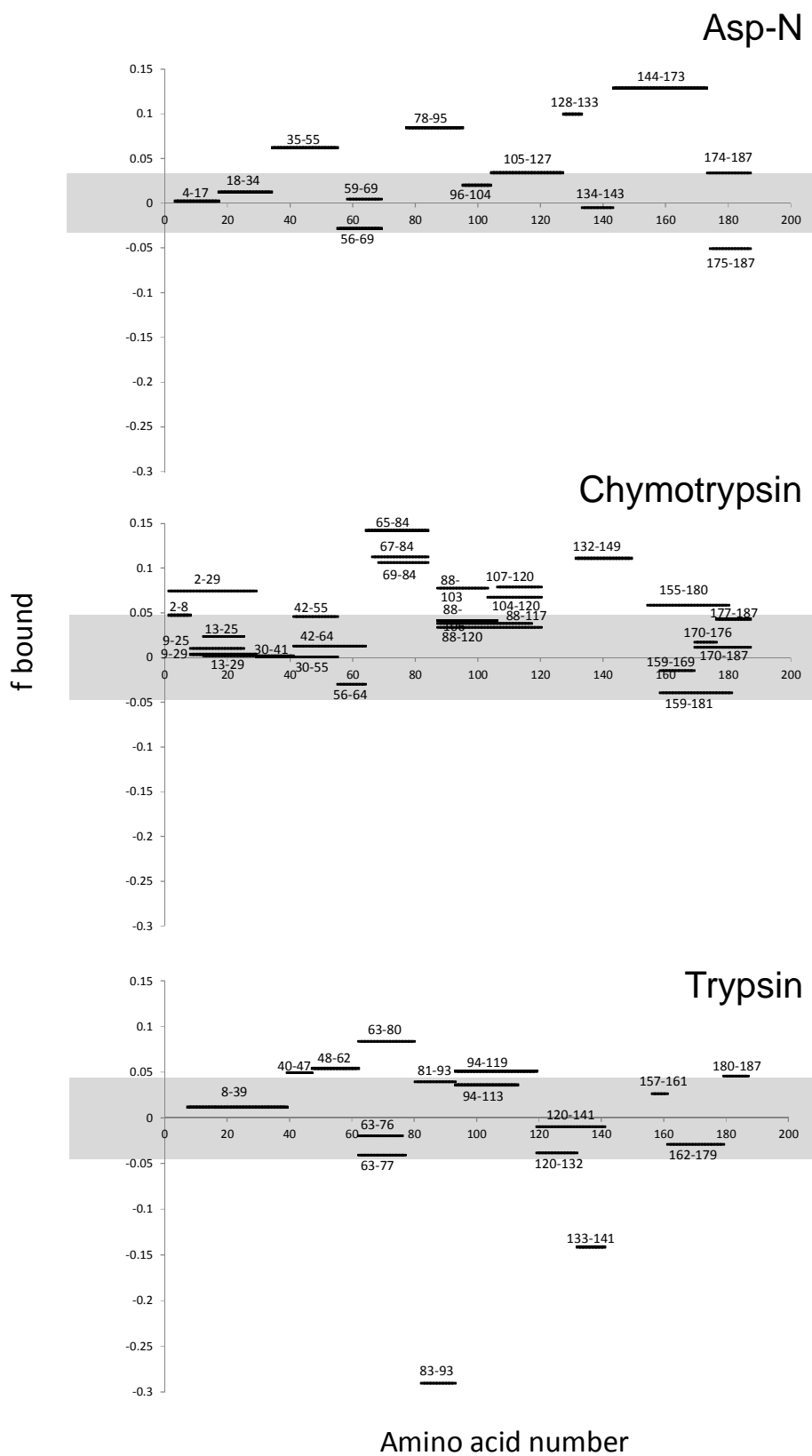


Figure 6. **Bound fraction pattern caused by a missed cleavage caused by the modification.** This scheme is valid for both N- and C-endoproteinases. Protein and peptides are represented in color with associated abundances. The corresponding  $f$  bounds are calculated using non modified peptides and graphed at the bottom.

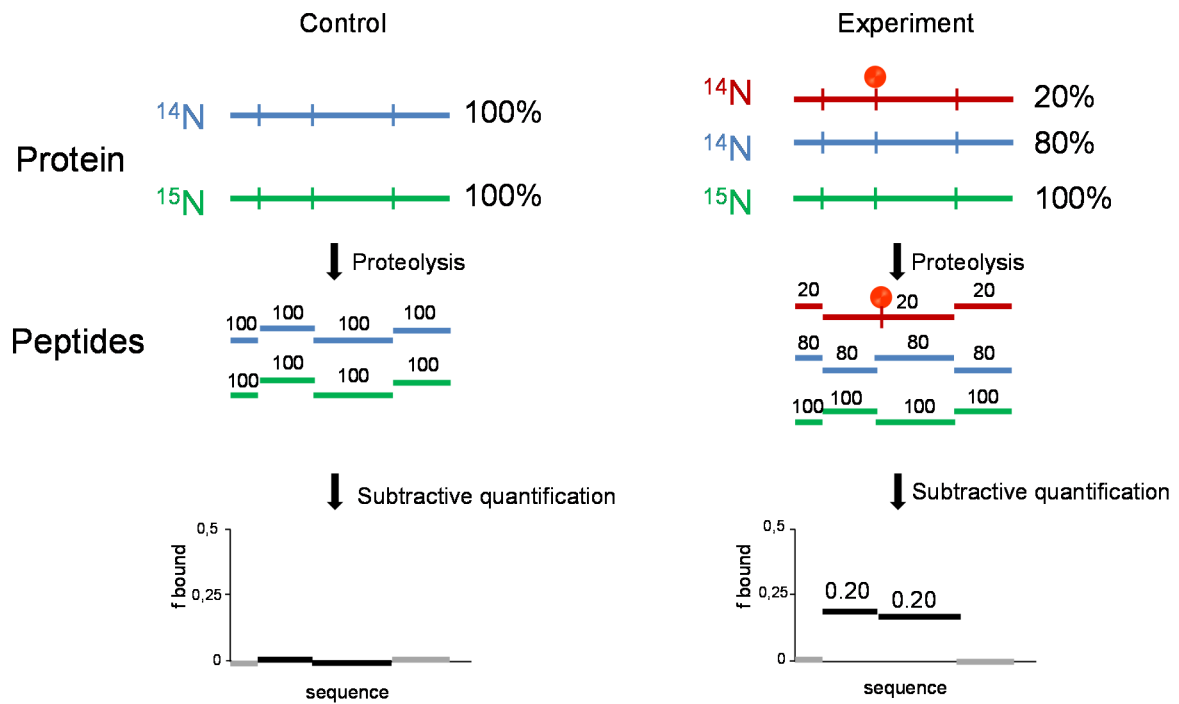


Figure 7. **Bound fraction pattern caused by a shifted cleavage.** This scheme is valid for both N- and C-endoproteinases. Protein and peptides are represented in color with associated abundances. The corresponding  $f$  bounds are calculated using non modified peptides and graphed at the bottom. Two cleavable and modifiable sites separated by two residues are shown: site 1 and site 2. In the control, site 1 is completely cleaved while a missed cleavage of 70% is considered at site 2. In the experiment, 20% of  $^{14}\text{N}$  protein is modified at site 1, so that there is a MCm at site 1 and there is no more MC at site 2 in this modified population of protein. The di-peptide generated after cleavages at sites 1 and 2 is not represented in the bottom graphs because of its short length which impairs its detection by MS.

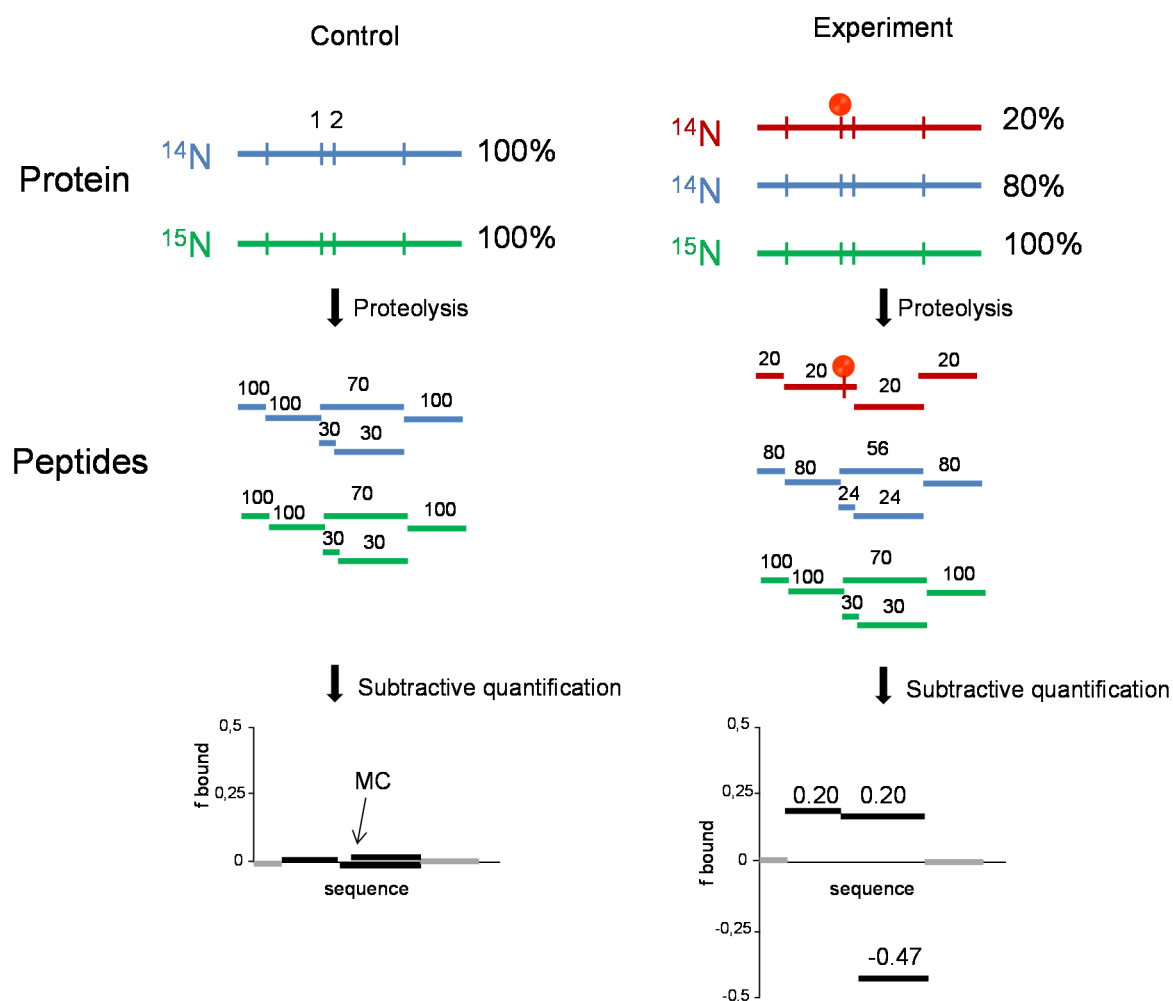




Figure 8. Graphical representation of parallel quantification of sites modifications for the hPEBP1-locostatin t=5h reaction time. Segmentation was performed on the dataset.

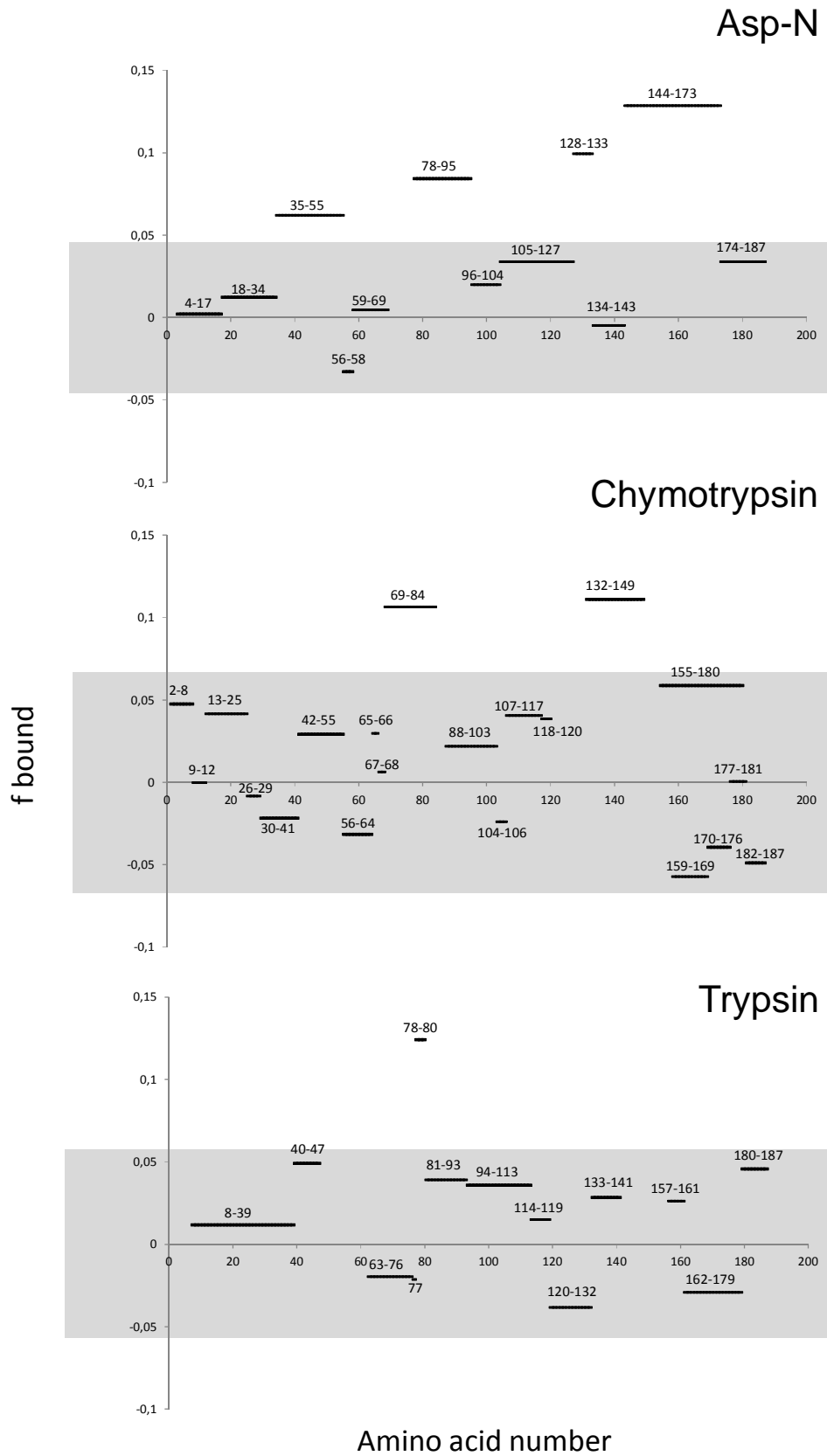


Figure 9. **Graphical representation of parallel sites quantification for the hPEBP1-locostatin t=5h reaction time.** Proteolytic peptides results from the three proteinases were combined.

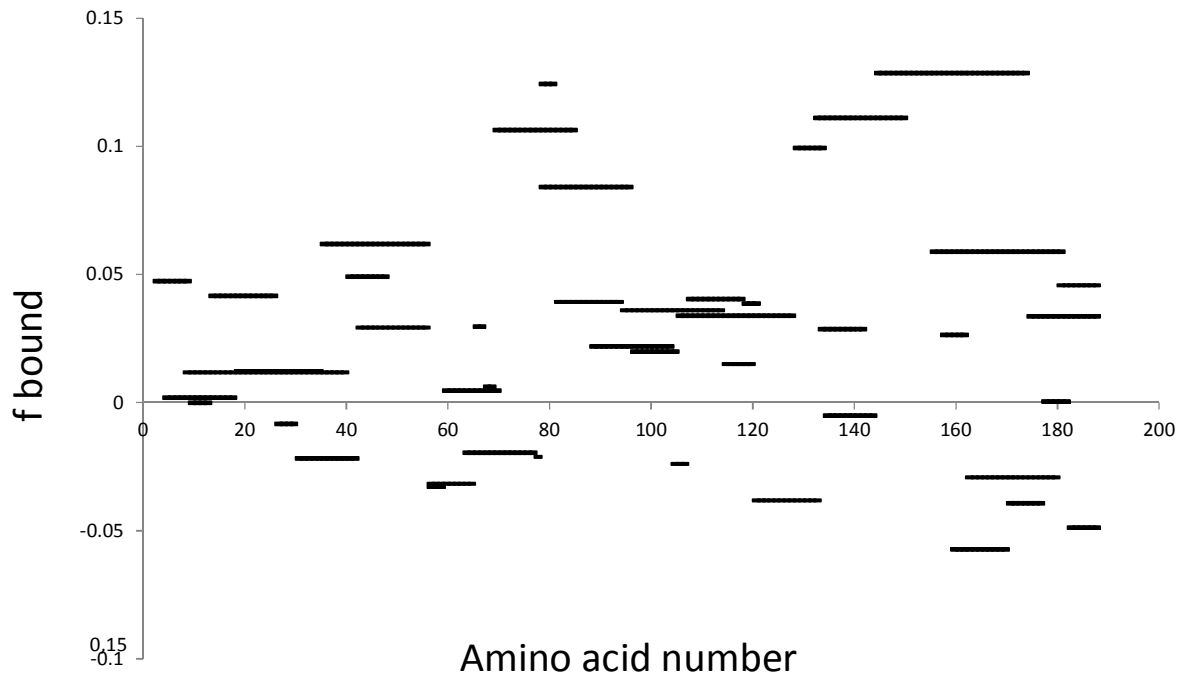


Figure 10. **MALDI-TOF MS analysis of the hPEBP1/RKIP-locostatin complex.** The hPEBP1 protein and locostatin were incubated for 5 h at 37°C before removing the excess of locostatin by micro gel filtration. The covalent complex was then analyzed by MALDI-TOF MS.

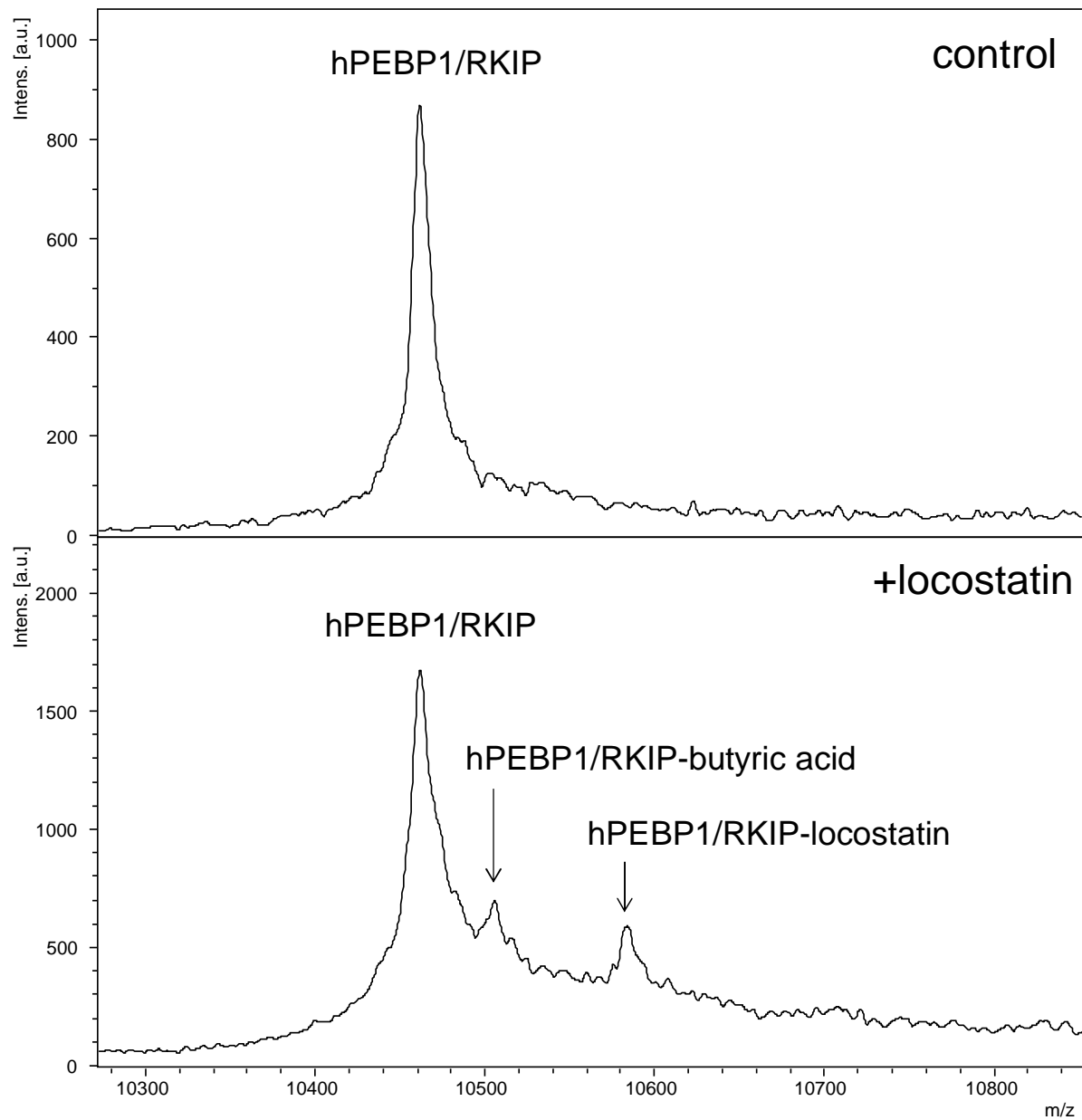


Figure 11. **Graphical representation of parallel sites quantification for the hPEBP1-locostatin t=5h reaction time.** LocMin calculation was applied to the dataset from Figure 8.

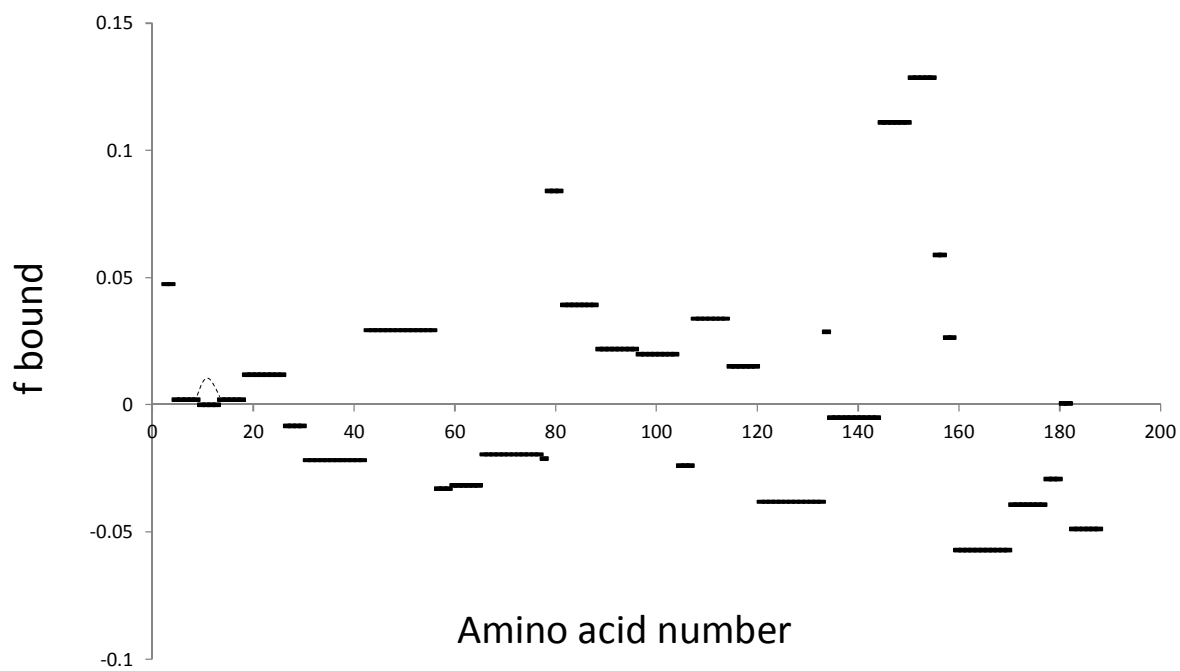
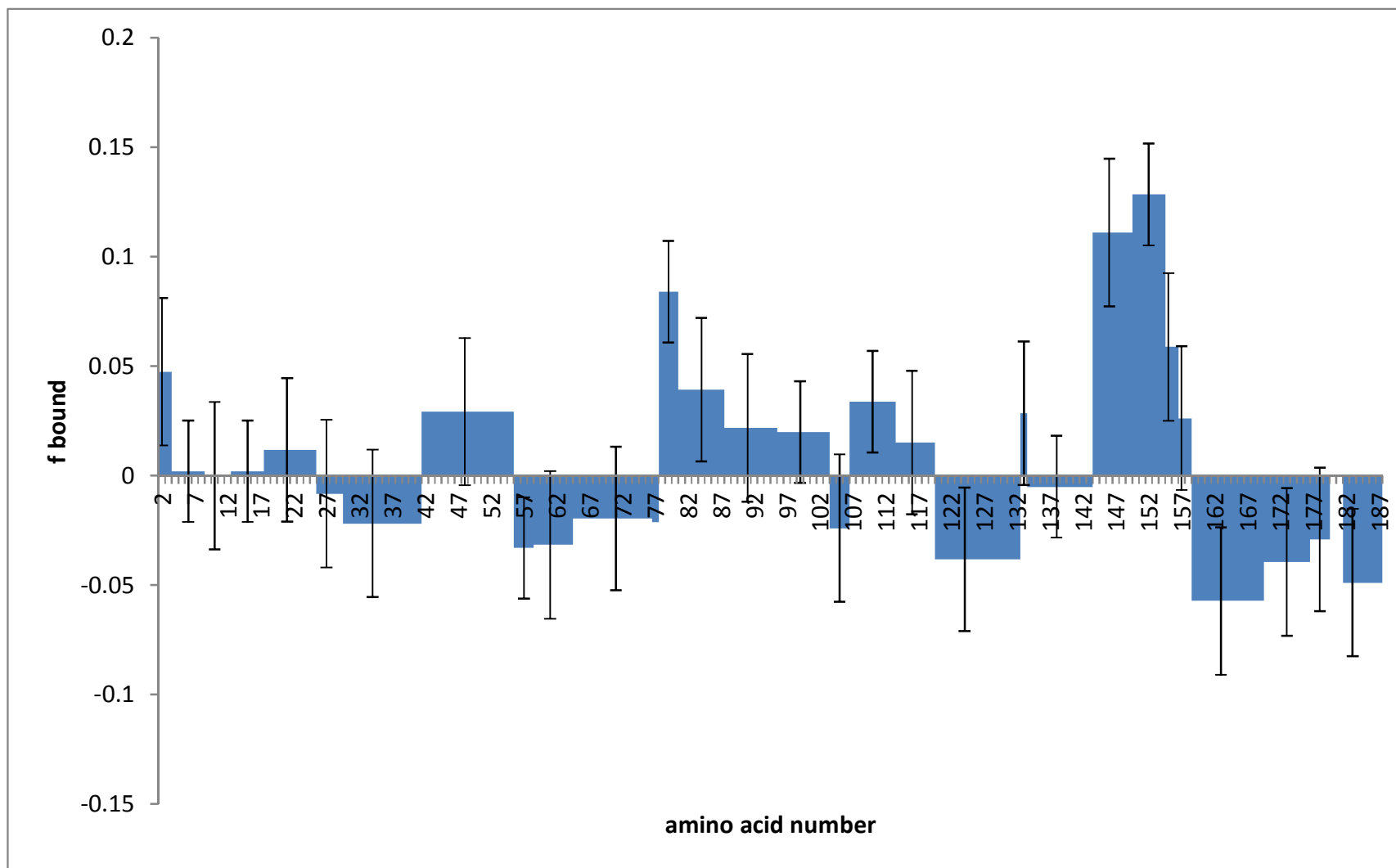


Figure 12. Histogram representation of parallel sites quantification for the hPEBP1-locostatin t=5h reaction time.





## 8.2 Compilation des résultats des approches qualitative et quantitative

L'approche quantitative révèle que Lys80 et le segment [150-154] sont significativement modifiés. Ce résultat précieux permet de localiser le résidu modifié du peptide Asp-N [72-95] et de restreindre la zone du peptide Asp-N [144-173], deux peptides observés avec l'approche qualitative mais qui n'avait pas encore été fragmentés. Par ailleurs, les résidus Lys47, Lys113 et Cys133 identifiés comme étant modifiés par l'approche qualitative ont une fraction de liaison ( $f_{\text{bound}}$ ) en dessous du seuil significatif. L'analyse quantitative révèle donc qu'ils sont peu modifiés. Les résultats obtenus par les approches qualitative et quantitative sont compilés dans le Tableau 16.

**Tableau 16. Résultats compilés des approches qualitative et quantitative.** La stœchiométrie de modification de la protéine entière est de 0.50. NSM : non significativement modifié (seuil  $\sim 0.05$ ).

Séquence ou site modifié	Fraction de liaison ( $f_{\text{bound}}$ )
Lys47	NSM
Lys80	0.084
Lys113	NSM
Cys133	NSM
[150-154]	0.128

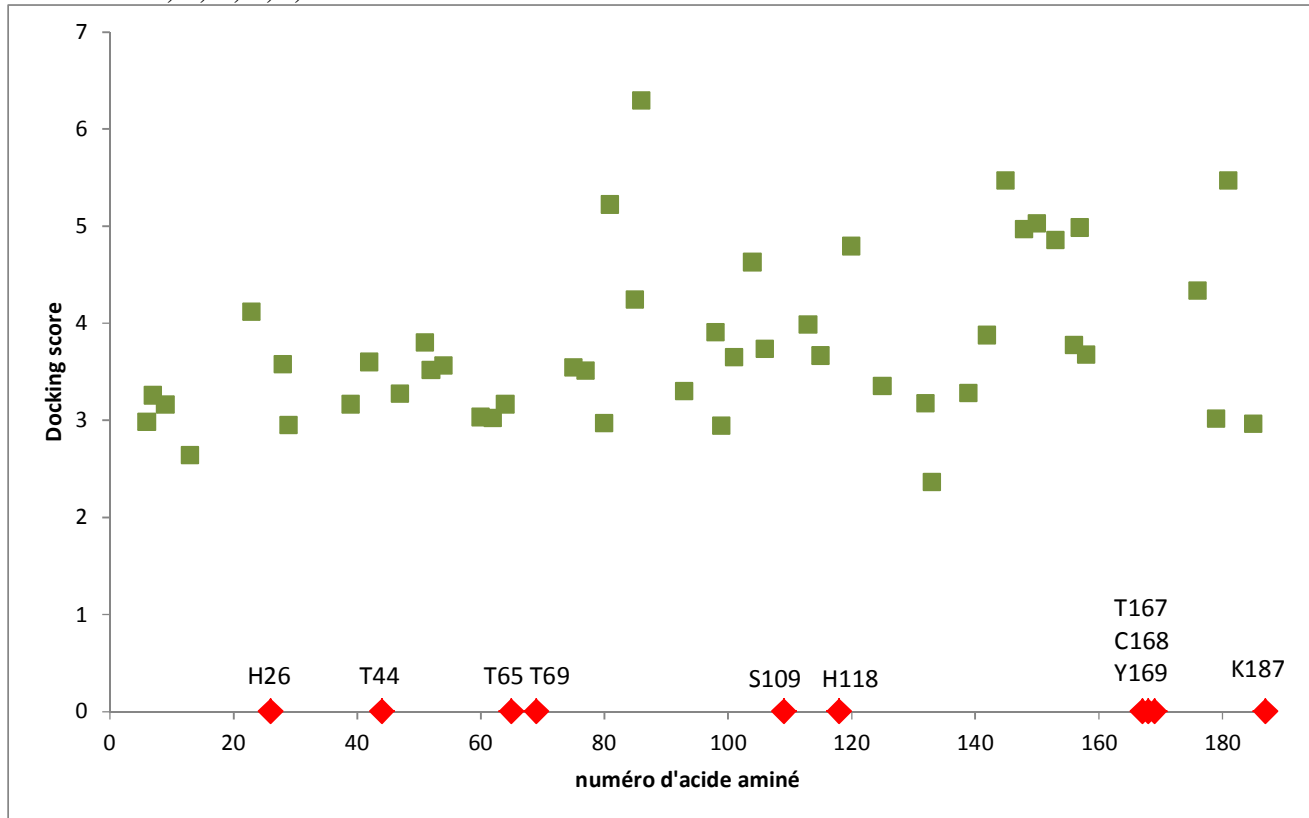
En faisant le rapport entre les deux fractions de liaison les plus élevées on obtient une valeur de 1.5 ( $0.128/0.084$ ). Cette valeur est nettement inférieure au facteur de discrimination (FD) de 3 que nous avons fixé suite aux simulations réalisées avec le modèle mathématique présenté paragraphe 7. Ce résultat révèle que, parmi les sites modifiés par la locostatine, il n'y a pas de site majoritaire. D'après la modélisation mathématique, ceci signifie que le rapport  $k_{\text{onS}}/k_{\text{onNS}}$  est inférieur à 5 (Figure 118), autrement dit, que la vitesse de réaction spécifique n'est pas visiblement plus rapide que celle des réactions non spécifiques.

## 9. Modélisation moléculaire de l'interaction hPEBP1-locostatine

La modélisation moléculaire a été entreprise en parallèle à l'approche expérimentale pour tenter de lever d'éventuelles ambiguïtés sur les segments contenant plusieurs résidus modifiables. Un calcul de minimisation de conformation de la locostatine liées de manière covalente sur tous les résidus modifiables de hPEBP1 : C, K, S, T, Y et H a été effectué en collaboration avec Stéphane Bourg et Norbert Garnier au CBM. Le logiciel utilisé est Glide. Le principe a été d'explorer le volume de balayage du ligand (= volume d'accessibilité, échantillonnage conformationnel du ligand) et de trouver la meilleure conformation liée. L'approche se déroule en trois étapes : 1) fixation du ligand

sur un résidu modifiable, 2) exploration de l'espace conformationnel du ligand, 3) mesure de l'énergie d'interaction. Les vingt meilleures poses du ligand ont été retenues, ce qui ne correspond pas à une exploration exhaustive de toutes les poses possibles de la locostatine sur un résidu. La meilleure pose de chaque résidu parmi les vingt retenues est présentée Figure 119.

**Figure 119. Meilleur score conformationnel sur 20 poses pour chaque addition de locostatine sur les résidus modifiables K, C, H, Y, S, T de hPEBP1.**



Ces scores nuls indiquent un problème d'encombrement stérique se traduisant par un volume disponible insuffisant à proximité du résidu cible. Un résultat d'"absence de score" est précieux et fiable, car il permet de déterminer avec certitude qu'un résidu ne peut être modifié. Dix résidus, His26, Thr44, Thr65, Thr69, Ser109, His118, Thr167, Cys168, Tyr169 et Lys186, n'ont donné aucun score. Ce type de résultat a été utilisé pour lever l'ambiguïté de segments modifiés contenant plusieurs résidus modifiables. Seul le segment [107-133] observé avec l'approche quantitative est concerné. Ce segment contient Ser109 et Lys113 comme résidu modifiable. Etant donné que Ser109 ne peut pas être modifiée d'après la modélisation moléculaire, Lys 113 peut être proposée comme un site modifiée par la locostatine. Ce résultat est cohérent avec les données MS/MS du peptide [94-119]-acide butyrique détecté avec l'approche qualitative (Figure 109).

Malheureusement, cette approche ne permet pas lever l'ambiguïté entre Lys150 et Ser153 du segment [150-154] observé modifié par quantification.



On remarque que l'ensemble des scores non nuls est dans la même gamme de valeurs, ce qui ne permet pas de dégager un résidu parmi les autres.

D'une manière générale, les résultats de modélisation sont à considérer avec précaution car on ne sait pas si la locostatine induit ou non un changement conformationnel de hPEBP1. En effet, hPEBP1 n'a pu être co-cristallisée avec la locostatine pour le moment, et la structure du complexe en RMN n'est pas atteignable du fait de son insolubilité à concentration élevée. Les seules données de RMN disponibles concernent le produit d'hydrolyse (S)-4-benzyloxazolidin-2-one (Shemon *et al.*, 2009) et sa pertinence biologique n'est pas établie.

## 10. Conclusions

L'interaction entre la protéine hPEBP1 et la locostatine a été finement caractérisée au niveau structural et cinétique par spectrométrie de masse. Les résultats montrent que la locostatine se fixe de manière covalente à la protéine puis s'hydrolyse en groupement butyrate. Des expériences de cinétiques suggèrent l'existence d'un intermédiaire réactionnel non covalent et réversible ( $P + L \rightleftharpoons PL^* \rightarrow PL$ ). L'analyse de la réaction en conditions natives (adaptées à l'observation des complexes non covalents) n'a pas permis de mettre en évidence le complexe non-covalent, très certainement dû à un  $k_{on}$  très bas et un  $k_{off}$  très élevé. La stœchiométrie du complexe a été évaluée par une expérience de saturation de la protéine en ligand mais n'a pu être obtenue du fait de l'absence de saturation même après huit jours de réaction. Par ailleurs, ces longues incubations ont mis en évidence la présence de réactions non spécifiques de la locostatine sur hPEBP1 en comptabilisant jusqu'à six ligands fixés sur la protéine. Ces réactions non spécifiques ont été corroborées en démontrant que la locostatine a la capacité de réagir avec des nucléophiles libres en solution tels que le GSH, le TCEP ou les chaînes latérales d'acides aminés C, H, Y, S, T et K. Dans l'objectif de localiser le site spécifique ciblé par la locostatine, les conditions de réaction et de protéolyse en solution ont été optimisées. Un temps court de réaction a été favorisé et l'excès de locostatine libre a été éliminé après réaction. Concernant les conditions de protéolyse, une protéolyse complète de hPEBP a été obtenue en présence d'urée, d'acétonitrile et d'un ratio Enzyme:Substrat élevé. Malgré ces précautions, l'approche qualitative révèle la présence de plusieurs peptides modifiés, ce qui indique la présence de sites de réaction non spécifiques de la locostatine sur la protéine même pour un temps court de réaction (stœchiométrie 1:1). Ces résultats ont conduit au développement 1) d'un modèle mathématique pour déterminer la fraction de liaison optimale permettant de discriminer le site spécifique des sites non-spécifiques, et 2) d'une méthode pour la quantification parallèle et exhaustive du degré de modification de tous les sites modifiés de la protéine. Un site modifié, Lys80, est quantifié ainsi qu'une séquence [150-154] contenant Lys150 et Ser153 modifiables. Selon nos

critères, la quantification de ces sites/séquence modifiés ne permet pas de proposer un site majoritaire. Afin de tenter de lever l'ambiguïté sur les séquences contenant plusieurs résidus, dont [150-154], une modélisation moléculaire a été entreprise mais n'a pas été fructueuse.

## 11. Perspectives

L'ensemble de ces informations met en avant la nécessité de concevoir des analogues de locostatine qui soient plus stables et plus spécifiques. Sur la base des résultats obtenus, plusieurs analogues ont déjà été synthétisés par les équipes de Luigi Agrofoglio et de Sylvain Routier. Ces analogues ont été testés au niveau moléculaire pour leur capacité à se lier à hPEBP1 et au niveau cellulaire pour leur capacité à inhiber la migration et l'invasion de cellules tumorales mammaires et pulmonaires. Les résultats indiquent que certains analogues sont moins actifs que la locostatine et d'autres apparaissent plus actifs. Une corrélation entre l'activité cellulaire et la quantification de tous les sites modifiés de chaque analogue de locostatine pourrait peut-être permettre d'identifier le site spécifique.

Par ailleurs dans l'optique de concevoir des analogues plus efficaces que la locostatine, il serait pertinent de synthétiser des analogues qui soient plus affins pour hPEBP1 et peut-être moins électrophile que la locostatine. En effet, l'augmentation de l'affinité permettrait de maintenir plus longtemps le groupement électrophile de l'analogue de locostatine à proximité du résidu nucléophile spécifique de la protéine, favorisant ainsi la réaction covalente. Dans ces conditions, la cinétique de fixation pour le site spécifique devrait être plus rapide que les réactions non spécifiques.

Une autre possibilité pouvant expliquer la difficulté d'observer un site majoritaire est que hPEBP1 n'est pas dans une conformation optimale pour interagir avec la locostatine. En effet, *in vivo*, on ne peut exclure que lorsque la locostatine se fixe sur hPEBP1, hPEBP1 soit en complexe avec une ou plusieurs protéine(s) et que la structure tridimensionnelle de hPEBP1 dans ce complexe protéique soit différente de celle de hPEBP1 seule. Si cette hypothèse est vraie, tant que les partenaires protéiques en question ne sont pas identifiés, le complexe ne pourra être réalisé *in vitro*. Une solution serait alors d'identifier le site spécifique sur le complexe hPEBP1-locostatine généré *in vivo*. Pour cela, un analogue stable de la locostatine contenant une sonde, du type fonction azoture par exemple, devra être conçu. Ensuite, ce complexe pourra réagir sélectivement avec un support solide fonctionnalisé avec un groupement alcyne.

Sur le plan biologique, le mécanisme par lequel la locostatine inhibe la migration cellulaire et renforce le potentiel antimétastatique de PEBP1 n'est encore élucidé. Lorsque PEBP1 est associée à Raf1 et TAK1, elle inhibe les voies ERK et NFκB impliquées dans la prolifération et la survie cellulaire. Si l'on fait l'hypothèse que l'inhibition de la migration cellulaire, médiée par le complexe

PEBP1-locostatine, passe par une inhibition de ces voies, on peut supposer que la locostatine renforce l'interaction de PEBP1 avec Raf1 ou TAK1 ou plutôt empêche la dissociation de ces complexes. Sachant que la phosphorylation en Ser153 entraîne la dissociation de ces complexes, nous faisons l'hypothèse que la locostatine pourrait empêcher la kinase PKC de phosphoryler la Ser153 de PEBP1. L'observation de cette zone modifiée est une information excitante.

Sur le plan technique, la méthode de quantification développée doit être éprouvée avec une protéine modifiée de référence. Le choix de cette protéine est à l'étude. Pour compléter la méthode, il reste à déterminer sa gamme de linéarité. Lors du développement de cette méthode, un test avec un rapport 0.33 :1 a été réalisé et les résultats ont été conformes à nos attentes.

Au-delà de l'application à hPEBP1, la quantification de sites modifiés est un outil précieux pour corréler la liaison moléculaire avec l'activité cellulaire d'agents thérapeutiques covalents potentiels.



# Conclusion générale et perspectives

Les modifications de protéines jouent un rôle crucial dans les organismes vivants. En modifiant les propriétés physico-chimiques des protéines par des réactions chimiques et biochimiques, les modifications modulent la fonction des protéines. Ces modifications peuvent être post-traductionnelles (PTMs), c'est-à-dire qu'elles sont le résultat d'un processus cellulaire endogène, ou alors elles peuvent être induites suite à une exposition, intentionnelle ou non, à un des agents environnementaux externes ou liés à une intervention humaine. Un certain nombre de pathologies humaines sont dues à un déséquilibre de PTMs ou de modifications induites. Il est fondamental de caractériser ces modifications afin de comprendre leur rôle et leur dynamique, et de pouvoir ainsi concevoir un jour un médicament contre les pathologies qu'elles occasionnent. Par ailleurs, ces dernières années, les médicaments capables de se lier aux protéines de manière covalente suscitent un regain d'intérêt. La spectrométrie de masse (MS) est une technique de choix pour l'étude de modifications de protéine. Elle présente l'avantage d'être sélective, sensible, rapide et peu consommatrice d'échantillon.

L'objectif de cette thèse est de caractériser au niveau structural et cinétique des complexes protéine-ligand et des PTMs, dans le but de disséquer leur mécanisme et de fournir des informations précieuses pour la conception de médicaments covalents contre des protéines liant l'ADN, des protéinases, ou des kinases. Selon le cas, il nous a fallu aller plus ou moins loin dans le développement d'approches de biochimie combinées à des méthodes MS en résolution standard ou à haute résolution et de MS/MS CID et ETD pour tenter d'atteindre l'objectif.

La grégline est une protéine capable d'inhiber très efficacement l'élastase de neutrophile humaine (HNE), protéase impliquée dans l'invasion tumorale et dans les maladies pulmonaires. Elle possède des propriétés remarquables de résistance à la protéolyse et à la dénaturation thermique, acide ou basique. Pour comprendre l'origine de cette résistance, nous avons caractérisé les PTMs de la grégline et déterminé sa séquence exacte. Pour cela, en complément des études cristallographiques, nous avons utilisé la spectrométrie de masse haute résolution-haute précision combinée à des outils biochimiques tels que la réduction de ponts disulfure, la déphosphorylation par la  $\lambda$ -phosphatase et l'enrichissement de phosphopeptides. Les résultats, associés aux données de cristallographie aux rayons X de la grégline clivée [19-80], montrent que la grégline possède un pont disulfure additionnel par rapport autres membres Kazal. Ce quatrième pont disulfure permet de connecter l'extrémité C-terminale à l'hélice  $\alpha$  et explique très certainement l'origine des propriétés de résistance accrue de la grégline. Trois phosphorylations ont été identifiées et une modification nouvelle a été détectée dans la région N-terminale [1-27]. Ces modifications ne semblent pas être directement liées à l'activité anti-protéase de la grégline dans la mesure où Gauthier et ses collaborateurs ont montré que la grégline clivée [27-81] conservait ses propriétés inhibitrices

(Brillard-Bourdet *et al.*, 2006). Elles pourraient par contre être impliquées dans la localisation subcellulaire et/ou la stabilisation de la gréghine, ou dans le recrutement d'autres protéines pour former des complexes. Dans ce projet, l'apport de la spectrométrie de masse haute résolution – haute précision a été décisif. En effet, cette technique de pointe a permis 1) de déterminer de manière fiable le nombre de ponts disulfure, modification délicate à étudier car elle induit une différence de masse faible de 2 Da et 2) de proposer une formule brute pour une nouvelle modification post-traductionnelle.

La protéine MOS1 est une transposase de l'élément mariner-Mos1 qui présente des caractéristiques communes avec l'intégrase du VIH-1. Plusieurs inhibiteurs de MOS1 ont été identifiés et pourraient être utilisés à des fins thérapeutiques pour inhiber les intégrases rétrovirales. Afin de décrypter le mécanisme de liaison et de localiser le ou les site(s) cible(s) des inhibiteurs, la caractérisation MS de leur interaction avec MOS1 a été entreprise. Les essais préliminaires en MALDI-TOF suggèrent l'existence d'un complexe covalent, mais une hétérogénéité détectée sur la protéine empêche la caractérisation de la liaison de MOS1 aux inhibiteurs. Avant de poursuivre cette caractérisation, l'hétérogénéité de MOS1 a été étudiée. L'utilisation de la HRMS et de la fragmentation ETD combinées à une déphosphorylation par la  $\lambda$ -phosphatase a permis de montrer que la transposase MOS1 est à la fois acétylée et phosphorylée. Etant donné la taille conséquente de la protéine MOS1 recombinante (~80 kDa), il paraît judicieux de réduire cette hétérogénéité car elle ajoute de la complexité non désirée pour l'étude de complexes protéine-ligand. Par ailleurs, nos collaborateurs ont montré que la phosphorylation en Ser170 inhibe l'activité de transposition de MOS1. Il serait ainsi très pertinent de trouver un inhibiteur ciblant cette sérine. Ces deux objectifs pourraient être conciliés en déphosphorylant MOS1 ou mieux encore, en produisant une forme de MOS1 non modifiée.

La protéine Abf2 a récemment été décrite comme étant une enzyme impliquée dans la réparation de l'ADN *via* une activité lyase. Pour contribuer à la compréhension du mécanisme catalytique et ainsi appréhender le potentiel éventuel d'Abf2 comme cible thérapeutique, son site actif a été recherché. Une stratégie originale combinant le piégeage du complexe transitoire Abf2-ADN, la purification du complexe covalent Abf2-ADN, suivie d'une protéolyse sur billes et d'une hydrolyse de l'ADN, avec l'analyse MS, a été développée. Plusieurs zones modifiées ont été observées, notamment l'étiquette 6 His, ce qui témoigne d'une réaction non spécifique lors de l'étape de piégeage de l'intermédiaire réactionnel. Sur la base des résultats de MS, plusieurs mutants ont été produits et purifiés, mais aucun mutant ne montre une baisse d'activité indépendante d'une diminution de l'affinité pour l'ADN. Afin de localiser le site actif d'Abf2 pour l'activité de clivage de sites abasiques, il faudra s'affranchir de ces réactions non spécifiques. Sur le plan technique, des outils performants et fiables de purification

de complexes protéine-ADN, de détection d'hétéroconjugués peptide-oligonucléotide et de localisation de résidus modifiés ont été développés.

La protéine humaine Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1), est un suppresseur de métastases dans différents types de cancer. La locostatine est le seul composé connu qui empêche la migration des cellules en se liant à hPEBP1. Les travaux présentés ici s'inscrivent dans un projet collaboratif dont l'objectif est d'exploiter le potentiel anti-métastatique naturel de PEBP1 en synthétisant, de façon rationnelle, des analogues de locostatine ayant un effet agoniste (ou activateur) de la protéine pour en faire de futurs médicaments capables d'empêcher le développement des métastases cancéreuses. Dans le but de fournir des informations précieuses pour la synthèse raisonnée d'analogues de locostatine, l'interaction entre hPEBP1 et la locostatine a été finement caractérisée au niveau cinétique et structural. Le mécanisme de liaison a été disséqué et montre que la locostatine se fixe de manière covalente à la protéine puis s'hydrolyse en groupement butyrate. Nous montrons également que la cinétique d'association est lente et que la locostatine est capable de réagir de manière non spécifique sur la protéine. Ces résultats ont conduit au développement 1) d'un modèle mathématique permettant de déterminer la fraction de liaison optimale pour discriminer le site spécifique des sites non-spécifiques, et 2) d'une méthode pour la quantification parallèle et exhaustive du degré de modification de tous les sites modifiés de la protéine. Selon nos critères, les résultats de quantification de ces sites/séquence modifiés ne permettent pas de proposer un site majoritaire. Ce résultat indique que la vitesse de réaction du site spécifique n'est pas visiblement plus rapide que celle du non spécifique, ce qui complique fortement l'identification du site spécifique de la locostatine sur hPEBP1.

L'ensemble de ces informations met en avant la nécessité de concevoir des analogues de locostatine qui soient plus stables et plus spécifiques. Pour cela, il serait pertinent de synthétiser des analogues qui soient plus affins pour hPEBP1 et peut-être moins électrophile que la locostatine. En effet, l'augmentation de l'affinité permettrait de maintenir plus longtemps le groupement électrophile de l'analogue de locostatine à proximité du résidu nucléophile spécifique de la protéine, favorisant ainsi la réaction covalente. Dans ces conditions, la cinétique de fixation pour le site spécifique devrait être plus rapide que les réactions non spécifiques.

Mes travaux montrent que la caractérisation de complexes protéine-ligand covalents et notamment la localisation du ou des résidu(s) cible(s) (ou résidu modifié dans le cas de modifications post-traductionnelles) peuvent se heurter à des difficultés. Le niveau de difficulté rencontré dépend de l'écart de masse associé à la modification, du degré de modification de la protéine, de la stabilité de



la modification au cours des analyses MS et MS/MS, de l'effet de la modification sur l'efficacité d'ionisation de la protéine et/ou du peptide modifié, de la présence d'isomères de position. De plus, dans plusieurs des cas étudiés ici, un recensement exhaustif de sites de modification sur une protéine montre un niveau non-négligeable de réactions protéine-ligand non spécifiques. En conséquence, la localisation de résidus modifiés post-traductionnellement ou par une molécule covalente, médicament potentiel, n'est pas toujours triviale et nécessite le plus souvent des développements méthodologiques originaux permettant de surmonter les différentes difficultés rencontrées. La quantification a un rôle important à jouer pour pouvoir évaluer de manière rationnelle si le niveau de modification non-spécifique est "à bas bruit" et pour le comparer à la modification spécifique.

Les méthodes développées au cours de ces travaux pourront être utilisées pour l'étude structurale et fonctionnelle d'autres PTMs ou d'autres complexes protéine-ligand covalent. Le traitement d'échantillons plus complexes tel qu'un protéome issu d'un biofluide nécessitera la mise en œuvre de techniques séparatives pour réduire la complexité de l'échantillon. Actuellement, le domaine de la séparation HPLC de protéines est en pleine effervescence et permet de séparer de plus en plus de protéines par des approches bidimensionnelles. Ces avancées technologiques, associées au développement de spectromètres de masses plus sensibles, plus précis et avec une meilleure gamme dynamique, seront très bénéfiques pour caractériser et quantifier les PTMs, pour approfondir les études d'impact de médicaments covalents, ainsi que pour caractériser finement les modifications induites au cours de la purification ou du stockage d'un biomédicament.



# Matériel et méthodes

# 1. Produits biologiques et chimiques

La protéine gréglina a été isolée à partir d'ovaires de criquet *Schistocerca gregaria* comme décrit dans la section Matériel et Méthodes de la publication « Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with subtilisin ».

La protéine MOS1 a été exprimée en fusion avec la Maltose Binding Protein (MBP) dans *Escherichia coli* et baculovirus et purifiée comme décrit dans la section Matériel et Méthodes de la publication « cAMP protein kinase phosphorylates the Mos1 transposase and regulates its activity: evidences from mass spectrometry and biochemical analyses ».

La protéine Abf2 a été exprimée avec une étiquette 6 histidines (6 His) dans la souche BL-21(DE3) d'*E. coli* à partir d'une construction obtenue dans le plasmide pET28a. Après lyse des cellules par sonication puis centrifugation, la fraction soluble a été récupérée. Cette fraction a ensuite été passée sur une chromatographie d'affinité Ni-NTA en utilisant un système FPLC (Fast Protein Liquid Chromatography). Après élution avec 250 mM imidazole, la fraction éluée a été purifiée par chromatographie échangeuse de cations (colonne SP pour Sulfo-Propyl) qui utilise les propriétés basique d'Abf2. La protéine a ensuite été éluée par un gradient NaCl 0.1-1 M. Enfin une dernière étape de gel filtration ACA 54 a été réalisée.

La protéine PEBP1 humaine (hPEBP1) a été produite et purifiée sans étiquette à partir d'*E. coli* comme décrit dans la section Matériel et Méthodes de la publication en cours de soumission « Exhaustive parallel inventory and quantification of the extent of modification at each site of a protein: a subtractive approach ».

L'Uracile DNA Glycosylase (UDG) et l'apurique/apyrimidique endonucléase 1 (Ape 1) ont été achetées chez New England Biolabs. Les oligonucléotides ont été obtenus chez Eurogentec (Hampshire, UK). Les endoprotéases utilisées, trypsine, AspN et chymotrypsine, proviennent de chez Roche Diagnostics GmbH (Mannheim, Allemagne). Les matrices acide 4-hydroxy- $\alpha$ -cyano-cinnamique (4-HCCA) et acide 3-hydroxy-picolinique (3-HPA) ainsi que le mélange de peptides standards de calibration pour l'analyse par spectrométrie de masse MALDI-TOF proviennent de chez Bruker Daltonics (Leipzig, Allemagne). Les standards de calibration pour les analyses ESI trappe ionique ou UHR-QTOF ont été achetés chez Agilent Technologies (Santa Clara, CA, USA). Les solvants organiques acétonitrile, isopropanol ont été obtenus chez Carlo Erba (Milan, Italie). L'eau ultra-pure de qualité 18 M $\Omega$  utilisée pour toutes les expériences a été purifiée en utilisant le système MilliQ reagent grade system de chez Millipore (Bedford, MA, USA). Tous les autres réactifs et solvants proviennent de chez Sigma-Aldrich Chemical Company (St Louis, MO, USA).

## 2. Purification d'hétéroconjugués Abf2-ADN

### 2.1 Formation du duplex ADN contenant un site AP ou 5'dRP

Les duplex ADN ont été préparés par nos collaborateurs comme indiqués ci-après. Les oligonucléotides ont préalablement été purifiés par électrophorèse dénaturante sur gel acrylamide-urée. L'hybridation des oligonucléotides a été réalisée dans des concentrations équimolaires de 50  $\mu\text{M}$  dans un tampon 10 mM Tris-HCl, pH 8.3, 1 mM EDTA et 100 mM NaCl par un cycle de dénaturation à 85°C pendant 5 min, puis refroidissement à température ambiante pendant 30 min. Le traitement à l'Uracile DNA Glycosylase (UDG) a été effectué pendant 15 min à 37°C dans le tampon d'hybridation avec une concentration en UDG de 1 U/400 pmol de duplex. Lorsqu'il a été réalisé, le traitement à l'apurique/apyrimidique endonucléase 1 (Ape 1) a consisté en une incubation de 15 min à 37°C de l'Ape1 à 1 nM dans le tampon d'hybridation contenant 10 mM MgCl<sub>2</sub>.

### 2.2 Formation d'un complexe covalent Abf2-ADN

La protéine Abf2 et le duplex ADN contenant un site abasique (AP ou 5'dRP) ont tout d'abord été incubés par nos collaborateurs pendant 2 min à température ambiante pour former le complexe Abf2-ADN. Ensuite 5 mM NaBH<sub>4</sub> ont été ajoutés pour réduire la base Schiff et ainsi piéger de manière covalente et irréversible l'intermédiaire réactionnel. L'incubation en présence de NaBH<sub>4</sub> a été de 1 h dans la glace. Les concentrations et rapports enzyme:substrat suivants ont été utilisés : rapport E:S 1.5:1 avec 75  $\mu\text{M}$  d'Abf2, 1:1 avec 66  $\mu\text{M}$  d'Abf2, 4:1 avec 1  $\mu\text{M}$  d'Abf2 et 1:1 avec 0.25  $\mu\text{M}$  d'Abf2. Le rendement de piégeage est estimé à environ 15 % d'après l'intensité des bandes sur SDS-PAGE.

### 2.3 Purification du complexe covalent Abf2-ADN en deux étapes (stratégie expérimentale 1)

#### 2.3.1 Élimination des protéines libres par chromatographie d'affinité avidine-biotine

La purification par chromatographie d'affinité avidine-biotine a été réalisée par nos collaborateurs avec de l'avidine monomérique immobilisée sur résine (Pierce, ref 20227). La mini-colonne d'avidine (1.2 mL) a été équilibrée avec 1 M NaCl, 100 mM NaH<sub>2</sub>PO<sub>4</sub> (PBS) puis l'échantillon a été chargé. La résine a ensuite été lavée dans le même tampon PBS afin d'éliminer Abf2 libre. L'élution

a été réalisée avec 2 mL de D-biotin à 2 mM dans le PBS. Le rendement de cette étape est estimé à 20%.

### 2.3.2 Elimination des oligonucléotides libres

La purification par chromatographie d'affinité sur métal immobilisé a été réalisée par nos collaborateurs avec des billes magnétiques Ni-NTA de chez Qiagen. Le mélange complexe Abf2-ADN et ADN libre (2 mL) a été chargé en quatre passages de 500  $\mu$ L sur les billes magnétiques (800  $\mu$ L) en présence de 10 mM imidazole. L'ADN libre a ensuite été éliminé par lavage des billes avec 100 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl et 20 mM imidazole.

## 2.4 Purification du complexe covalent Abf2-ADN par chromatographie échangeuse d'anions (monoQ) (stratégie expérimentale 2)

Le complexe covalent Abf2-ADN a été purifié par nos collaborateurs par chromatographie échangeuse d'anions (MonoQ) en utilisant un système FPLC et un tampon 10 mM Hepes pH7.5, 1 mM DTT. Le volume de colonne utilisé a été de 1 mL. L'élution a été réalisée par un gradient de NaCl 0.1-1 M. Ces conditions permettent de séparer efficacement les trois populations de molécules : la protéine seule n'est pas retenue sur la colonne et le complexe covalent Abf2-ADN puis le duplex ADN sont élués successivement au cours du gradient. Les fractions correspondantes au complexe covalent Abf2-ADN ont été collectées puis groupées dans un volume total de 2 mL. Ensuite les billes magnétiques Ni-NTA ont été ajoutées en présence de tween20 0.005% suivi d'une incubation de 30 min à 4°C sous agitation douce.

## 2.5 Dénaturation et protéolyse sur billes

La dénaturation a été réalisée sur le complexe Abf2-ADN fixé sur les billes Ni-NTA. Un traitement thermique de 10 min à 60°C dans 8 M urée a été effectué pour déplier la protéine et dissocier le duplex ADN. Le surnageant a ensuite été éliminé puis trois lavages de 100  $\mu$ L de bicarbonate d'ammonium pH 8.0 (ABC) ont été réalisés. Rapidement, une protéolyse par la trypsine et/ou l'AspN a été réalisée dans 25  $\mu$ L d'une solution 100 mM ABC, 20 % acétonitrile pendant 3h à 37°C sous agitation 1400 rpm au thermomixer. Le rapport E:S estimé est de 1:5 et 1:20 pour la trypsine et l'AspN, respectivement. La protéolyse a été arrêtée par un traitement thermique de 5 min à 95°C suivi d'un stockage immédiat dans la glace.

## 2.6 Enrichissement d'hétéroconjugués peptide-oligonucléotide

L'échantillon est repassé aux mains de nos collaborateurs pour cette étape. Le surnageant de la protéolyse (20-50µL) a été mis en contact avec 10 µL de résine monoQ préalablement équilibrée dans 50 mM ABC, 20 % acétonitrile. Après une courte étape de vortex (~10 sec), l'échantillon a été centrifugé à 300g puis le surnageant a été éliminé. La résine a ensuite été lavée avec 2 × 50 µL 50 mM ABC, 20 % acétonitrile puis 1 × 50 µL 50 mM ABC. Les hétéroconjugués peptide-oligonucléotide ont été élués avec deux fois 2 × 5 µL 50 mM ABC, 650 mM NaCl. J'ai réalisé toutes les étapes restantes du protocole.

## 2.7 Hydrolyse de l'oligonucléotide

Les échantillons (10 µL) ont tout d'abord été acidifiés à un pH de 5.2 par ajout de 0.15 µL d'acide acétique 50 % puis 1.1 µL d'acétate d'ammonium 1 M. L'hydrolyse de l'oligonucléotide a été réalisée selon les indications du fabricant par un traitement à la Nucléase P1 (Sigma, ref N8630) à 0.1 U/10 µL de volume réactionnel pendant 30 min à 70°C.

## 3. Dessalage/Concentration

L'étape de dessalage/concentration a été effectuée avec le ZipTipC18 (Millipore) pour les peptides et les protéines de masse moléculaire inférieure à 50 kDa et avec le ZipTipC4 (Millipore) pour les protéines de masse moléculaire supérieure à 50 kDa. Les ZipTip utilisés contenaient 0.6 µL de phase stationnaire capable de retenir 5 µg de matériel. Après activation de la colonne par une solution 1 constituée d'acétonitrile/acide 70/30 (v/v), puis équilibration avec une solution 2 d'acétonitrile/acide 2.5/97.5 (v/v), les peptides ou les protéines ont été fixés sur la colonne par une série d'étapes d'aspiration et de refoulement. Les sels ont ensuite été éliminés par lavage avec la solution 2 et les peptides ou les protéines ont été élués avec 10 µL de solution 1.

Pour les échantillons destinés à l'analyse nanoLC ou MALDI, l'acide utilisé était l'acide trifluoroacétique à une concentration finale de 0.1%. Pour les échantillons analysés en infusion électrospray, l'acide formique à une concentration finale de 1% a été utilisé. Lorsque les échantillons dessalés/concentrés étaient destinés à l'analyse nanoLC-MS, une étape d'évaporation au speedvac puis de reprise dans une solution d'acétonitrile/TFA 0.1% 2.5/97.5 (v/v) a été effectuée.

## 4. Spectrométrie de masse

### 4.1 Spectrométrie de masse MALDI

Les échantillons (protéines entières ou mélange peptidique) ont été dilués à une concentration finale située entre 0.5 et 5  $\mu\text{M}$  dans une solution d'acide 4-hydroxy- $\alpha$ -cyano-cinnamique (4-HCCA) saturée dans un mélange acétonitrile:eau:acide trifluoroacétique 33.3:66.6:0.1 (TWA) ou acide formique:eau:isopropanol 3:1:2 (FWI). Le mélange échantillon-matrice a ensuite été déposé sur une cible plaquée or en utilisant la méthode en couche ultramince (Cadene and Chait, 2000; Gabant and Cadene, 2008). Les dépôts contenant des molécules pouvant nuire au processus d'ionisation telles que le NaCl, le KCl, le Tris, l'Hepes, l'urée ou la glycine ont été lavés par du TFA 0.1% avant l'acquisition. Les analyses ont été réalisées sur un MALDI-TOF/TOF LIFT UltraFlex I (Bruker Daltonics) équipé d'un laser à azote ( $\lambda = 337 \text{ nm}$ ) et d'une extraction retardée. Les spectres ont été acquis à une fréquence de 5 à 25 Hz en mode réflectron positif (gamme  $m/z$  500-5000) ou en mode linéaire positif (gamme  $m/z$  1000-100000) et représentent la somme cumulée de 1000 tirs. Une tension d'accélération de 25 kV a été utilisée et le délai a été optimisé à 50 ns pour le mode réflectron ou 250 ns pour le mode linéaire afin d'atteindre une résolution supérieure à 18000 au  $m/z$  2465 pour le mode réflectron ou 1000 entre la gamme 10000-20000 de  $m/z$  pour le mode linéaire. Une déflexion des ions de matrice jusqu'à 450 Da pour le mode réflectron ou 800 Da pour le mode linéaire a été appliquée pour empêcher la saturation du détecteur. Une calibration externe a été réalisée en utilisant des solutions de calibration standard, déposées à proximité de l'analyte après mélange avec la solution de matrice: Pepmix I (Bruker Daltonics) pour le mode réflectron et un mélange de cytochrome c et d'apomyoglobine pour le mode linéaire. L'acquisition des spectres a été réalisée *via* le logiciel Flexcontrol 3.3 (Bruker Daltonics) et l'exploitation des données a été réalisée *via* le logiciel Flexanalysis 3.3 (Bruker Daltonics).

### 4.2 Spectrométrie de masse à source électrospray

#### 4.2.1 nanoUltraHPLC-MS

Le matériel utilisé est un système UltiMate<sup>TM</sup> 3000 nanoUltraHPLC (Dionex, Sunnyvale, CA) connecté soit à une trappe ionique HCT Ultra PTM Discovery (Bruker), soit à un MaXis UHR-Q-TOF 4 GHz (Bruker), tous deux équipés d'une source électrospray. La configuration LC-MS était pilotée par le logiciel Hystar<sup>TM</sup> version 3.2 (Bruker). Les peptides ont été pré-concentrés en ligne



sur une précolonne en phase inverse Thermo Acclaim PepMap100 C18 (diamètre interne 100  $\mu\text{m}$ , longueur 2 cm, taille des particules 5  $\mu\text{m}$ , taille des pores 100  $\text{\AA}$ ), puis séparés sur une colonne analytique Thermo Acclaim Pepmap100 C18 (diamètre interne 75  $\mu\text{m}$ , longueur 25 cm, taille des particules 2  $\mu\text{m}$ , taille des pores 100  $\text{\AA}$ ) à l'aide d'un gradient de 2 à 35% d'acétonitrile dans 0.1% d'acide formique réalisé à un débit de 450 nL/min. La colonne était thermostatée à 40°C. La sortie de la nano-colonne a été reliée directement à la source nanoESI en ligne en utilisant un NanoFlow ESI Sprayer (nanosprayer) (Bruker).

## 4.2.2 Trappe ionique HCT Ultra PTM Discovery

### 4.2.2.1 Couplage nanoLC-MS

Une tension de - 4000 V et une pression d'azote de 11 psi ont été appliquées au niveau de l'aiguille de nébulisation du nanosprayer. Le gaz de séchage ( $\text{N}_2$ ) était à une température de 300°C avec un débit de 5 L/min. Les spectres ont été enregistrés en mode positif entre 300 et 1500 de  $m/z$  en MS et entre 100 et 2800 de  $m/z$  en MS/MS. Les acquisitions ont été réalisées avec une résolution de balayage améliorée (8100  $m/z$ .s-1). La valeur cible d'accumulation des ions (ICC pour Ion Charge Control) a été fixée à 90 000 et le temps maximum pour atteindre cette valeur d'ICC a été fixé à 200 ms. La fragmentation CID a été réalisée avec une énergie de collision de 1.20 V et de l'hélium (gaz de collision), et d'une manière préférentielle sur les ions dichargés. La fragmentation ETD a été réalisée en accumulant du fluoranthène jusqu'à atteindre une ICC de 400 000 en 200 ms. Le temps de la réaction de transfert d'électrons a été fixé à 100 ms. Les spectres MS/MS ont été acquis en utilisant une liste d'inclusion ou en mode « data dependent acquisition » (DDA). Dans le mode DDA, un maximum de trois ions précurseurs par cycle a été fragmenté en MS/MS en utilisant une exclusion dynamique de 12 s. Les fragmentations CID et ETD ont été effectuées dans des analyses nanoLC-IT-MS/MS séparées. Une calibration externe a été réalisée en utilisant le calibrant ESI-T Tuning Mix (Agilent Technologies).

### 4.2.2.2 Infusion directe

Pour les expériences en infusion directe, les échantillons ont été analysés dans une solution d'acétonitrile:eau:acide formique 50/49.8/0.2 (WAFa, pour Water Acetonitrile Formic Acid). Les échantillons ont été infusés à un débit de 180  $\mu\text{L}/\text{h}$  à l'aide d'un pousse seringue et d'une aiguille de nébulisation ESI (Agilent Technologies). Les paramètres d'ionisation, de désolvatation et d'analyse MS et MS/MS sont identiques à ceux utilisés pour le couplage nanoRSLC à l'exception de l'exclusion dynamique qui est inactivée. L'analyse en infusion permet d'accumuler du signal et

d'optimiser certains paramètres tels que la fenêtre d'isolation ou l'énergie de fragmentation. L'acquisition des spectres de masse été réalisé *via* le logiciel Esquire Control 6.2, et le traitement des données *via* Data Analysis 4.0 (Bruker Daltonics).

### 4.2.3 Maxis UHR-Q-TOF 4GHz

Le Maxis UHR-Q-TOF 4GHz est un instrument à haute résolution-haute précision de masse capable d'atteindre une résolution supérieure à 50000 et une précision de masse inférieure à 2 ppm.

#### 4.2.3.1 Couplage nanoLC-MS

Pour le couplage avec la nanoUltraHPLC, le Maxis 4 GHz a fonctionné uniquement en mode MS. Au niveau de l'aiguille de nébulisation du nanosprayer, une tension de - 4000 V et une pression d'azote de 1.1 bar ont été appliquées. Un offset de - 500 V a été fixé à l'extrémité du capillaire de transfert (« end plate »). La désolvatation a été réalisée avec un gaz séchant (N<sub>2</sub>) de débit 4 L/min et de température 200°C. Les spectres ont été enregistrés en mode positif entre 300 et 3000 de  $m/z$  en MS. La fréquence d'acquisition a été réglée à 1 Hz correspondant à une sommation de 5494 spectres. Une calibration interne (« lock mass ») a été réalisée en utilisant le  $m/z$  622 (Hexakis(2,2-Difluoroethoxy)Phosphazene; Apollo Scientific) et le  $m/z$  1222 (Hexakis(1H, 1H, 4H-hexafluorobutyloxy)Phosphazine; Agilent Technologies).

#### 4.2.3.2 Infusion directe

Pour les expériences en infusion directe, les échantillons ont été analysés dans une solution d'acétonitrile:eau:acide formique 50/49.8/0.2 (WAFA). Les échantillons ont été infusés à un débit de 180  $\mu$ L/h à l'aide d'un pousse seringue et d'une aiguille de nébulisation ESI (Bruker). Les paramètres d'ionisation, de désolvatation et d'analyse MS sont identiques à ceux utilisés pour le couplage nanoUltraHPLC. Pour les analyses MS/MS, la fragmentation a été obtenue en mode CID en utilisant de l'azote et en ajustant l'énergie de collision entre 8 et 60 eV. Le paramètre iSCID (in source collision induced dissociation) a été ajusté entre 0 et 200 eV pour éliminer des adduits de types TFA ou pour induire une fragmentation dans la source. Une calibration externe a été réalisée en utilisant le calibrant ESI-L Low Concentration Tuning Mix (Agilent Technologies). L'acquisition des spectres de masse a été réalisée grâce au logiciel micrOTOFcontrol 3.2 (Bruker Daltonics) et le logiciel Data Analysis 4.0 a été utilisé pour retraiter les données (Bruker Daltonics).

Pour la déconvolution des spectres de protéines entières, l'algorithme MaxEnt a été utilisé.

### 4.3 Analyse bioinformatique des données

Pour l'analyse de peptides, les données de spectrométrie de masse ont été confrontées à une banque de données particulière, composée uniquement des séquences de la gregline, MBP-MOS1, Abf2 et hPEBP1 avec leur différentes étiquettes, à l'aide d'une copie locale du logiciel Mascot 2.3.02 (Matrix Science). L'utilisation de cette « mini » base de données permet de faire la recherche sur des protéines fusions et de diminuer les temps lors de la recherche. Toutefois, il est parfois nécessaire de faire la recherche dans une base de données standard telle que SwissProt pour pouvoir annoter des pics contaminants, issus par exemple de kératine. Pour la recherche, les différentes modifications recherchées ont été considérées comme variables. En MALDI-TOF, la tolérance a été fixée à 50 ppm en MS. En ESI-IT, la tolérance a été fixée à 0.5 Da en MS et MS/MS. En ESI-UHR-QTOF, la tolérance a été fixée à 5 ppm en MS et 0.02 Da en MS/MS. Le nombre de « missed cleavage » a été fixé à 2 pour la trypsine ou l'Asp-N et 5 pour la chymotrypsine. L'annotation des résidus modifiés a été validée manuellement à l'aide du logiciel Biotools 3.2 (Bruker Daltonics). En parallèle à Mascot, le logiciel PAWS version 8.5.0.3 (ProteoMetrics, New York, NY) a également été utilisé pour l'annotation MS des peptides. Pour la recherche de formules brutes, les logiciels SmartFormula et Smart Formula 3D (Bruker) ont été utilisés pour les données MS et MS/MS, respectivement.



# Bibliographie

- Arnesen, T. Towards a Functional Understanding of Protein N-Terminal Acetylation. *PLoS Biol* **2011**, *9*, e1001074.
- Azam, M.; Seeliger, M. A.; Gray, N. S.; Kuriyan, J.; Daley, G. Q. Activation of Tyrosine Kinases by Mutation of the Gatekeeper Threonine. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1109–1118.
- Bai, G.; Pfaff, S. L. Protease Regulation: The Yin and Yang of Neural Development and Disease. *Neuron* **2011**, *72*, 9–21.
- Bailey, A. J.; Robins, S. P.; Balian, G. Biological Significance of the Intermolecular Crosslinks of Collagen. *Nature* **1974**, *251*, 105–109.
- Banfield, M. J.; Barker, J. J.; Perry, A. C.; Brady, R. L. Function from Structure? The Crystal Structure of Human Phosphatidylethanolamine-Binding Protein Suggests a Role in Membrane Signal Transduction. *Struct. Lond. Engl.* **1993** **1998**, *6*, 1245–1254.
- Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present. *Anal. Bioanal. Chem.* **2012**, *404*, 939–965.
- Baritaki, S.; Katsman, A.; Chatterjee, D.; Yeung, K. C.; Spandidos, D. A.; Bonavida, B. Regulation of Tumor Cell Sensitivity to TRAIL-Induced Apoptosis by the Metastatic Suppressor Raf Kinase Inhibitor Protein via Yin Yang 1 Inhibition and Death Receptor 5 up-Regulation. *J. Immunol. Baltim. Md 1950* **2007**, *179*, 5441–5453.
- Baritaki, S.; Chapman, A.; Yeung, K.; Spandidos, D. A.; Palladino, M.; Bonavida, B. Inhibition of Epithelial to Mesenchymal Transition in Metastatic Prostate Cancer Cells by the Novel Proteasome Inhibitor, NPI-0052: Pivotal Roles of Snail Repression and RKIP Induction. *Oncogene* **2009**, *28*, 3573–3585.
- Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V. T.; Huang, J. Y.; Ma, L.-C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H. Structure of Antibacterial Peptide Microcin J25: A 21-Residue Lariat Protoknot. *J. Am. Chem. Soc.* **2003**, *125*, 12382–12383.
- Beach, S.; Tang, H.; Park, S.; Dhillon, A. S.; Keller, E. T.; Kolch, W.; Yeung, K. C. Snail Is a Repressor of RKIP Transcription in Metastatic Prostate Cancer Cells. *Oncogene* **2008**, *27*, 2243–2248.
- Beaufour, M.; Godin, F.; Vallée, B.; Cadene, M.; Bénédicti, H. Interaction Proteomics Suggests a New Role for the tfs1 Protein in Yeast. *J. Proteome Res.* **2012**, *11*, 3211–3218.
- Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A Probability-Based Approach for High-Throughput Protein Phosphorylation Analysis and Site Localization. *Nat. Biotechnol.* **2006**, *24*, 1285–1292.
- Beck, B. D.; Park, S.-J.; Lee, Y.-J.; Roman, Y.; Hromas, R. A.; Lee, S.-H. Human Pso4 Is a Metnase (SETMAR)-Binding Partner That Regulates Metnase Function in DNA Repair. *J. Biol. Chem.* **2008**, *283*, 9023–9030.
- Beltrao, P.; Bork, P.; Krogan, N. J.; van Noort, V. Evolution and Functional Cross-Talk of Protein Post-Translational Modifications. *Mol. Syst. Biol.* **2013**, *9*, n/a – n/a.
- Bennett, B. L. C-Jun N-Terminal Kinase-Dependent Mechanisms in Respiratory Disease. *Eur. Respir. J.* **2006**, *28*, 651–661.
- Bennett, E. J.; Rush, J.; Gygi, S. P.; Harper, J. W. Dynamics of Cullin-RING Ubiquitin Ligase Network Revealed by Systematic Quantitative Proteomics. *Cell* **2010**, *143*, 951–965.
- Bernier, I.; Jollès, P. Purification and Characterization of a Basic 23 kDa Cytosolic Protein from Bovine Brain. *Biochim. Biophys. Acta* **1984**, *790*, 174–181.
- Bernier, I.; Tresca, J. P.; Jollès, P. Ligand-Binding Studies with a 23 kDa Protein Purified from Bovine Brain Cytosol. *Biochim. Biophys. Acta* **1986**, *871*, 19–23.
- Beshir, A. B.; Ren, G.; Magpusao, A. N.; Barone, L. M.; Yeung, K. C.; Fenteany, G. Raf Kinase Inhibitor Protein Suppresses Nuclear Factor- $\kappa$ B-Dependent Cancer Cell Invasion through Negative Regulation of Matrix Metalloproteinase Expression. *Cancer Lett.* **2010**, *299*, 137–149.
- Beshir, A. B.; Argueta, C. E.; Menikarachchi, L. C.; Gascón, J. A.; Fenteany, G. Locostatin Disrupts Association of Raf Kinase Inhibitor Protein With Binding Proteins by Modifying a Conserved Histidine Residue in the Ligand-Binding Pocket. *Forum Immunopathol. Dis. Ther.* **2011**, *2*, 47–58.
- Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J. Multiplexed Absolute Quantification in Proteomics Using Artificial QCAT Proteins of Concatenated Signature Peptides. *Nat. Methods* **2005**, *2*, 587–589.
- Biemann K., Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **1988**; *16*; 99
- Biémont, C.; Vieira, C. Genetics: Junk DNA as an Evolutionary Force. *Nature* **2006**, *443*, 521–524.
- Boiteux, S.; Guillet, M. Abasic Sites in DNA: Repair and Biological Consequences in *Saccharomyces Cerevisiae*. *DNA Repair* **2004**, *3*, 1–12.
- Boja, E. S.; Fales, H. M. Overalkylation of a Protein Digest with Iodoacetamide. *Anal. Chem.* **2001**, *73*, 3576–3582.
- Boja, E. S.; Rodriguez, H. Mass Spectrometry-Based Targeted Quantitative Proteomics: Achieving Sensitive and Reproducible Detection of Proteins. *PROTEOMICS* **2012**, *12*, 1093–1110.
- Boja, E. S.; Phillips, D.; French, S. A.; Harris, R. A.; Balaban, R. S. Quantitative Mitochondrial Phosphoproteomics Using iTRAQ on an LTQ-Orbitrap with High Energy Collision Dissociation. *J. Proteome Res.* **2009**, *8*, 4665–4675.
- Bonavida, B.; Jazirehi, A.; Vega, M. I.; Huerta-Yepez, S.; Baritaki, S. Roles Each of Snail, Yin Yang 1 and RKIP in the Regulation of Tumor Cells Chemo-Immuno-Resistance to Apoptosis. *Forum Immunopathol. Dis. Ther.* **2013**, *4*.
- Bouchet, N.; Bischerour, J.; Germon, S.; Guillard, J.; Dubernet, M.; Viaud-Massuard, M. C.; Delelis, O.; Ryabinin, V.; Bigot, Y.; Augé-Gouillou, C. First Mariner Mos1 Transposase Inhibitors. *Mini Rev. Med. Chem.* **2009**, *9*, 431–439.

- Brillard-Bourdet, M.; Hamdaoui, A.; Hajjar, E.; Boudier, C.; Reuter, N.; Ehret-Sabatier, L.; Bieth, J. G.; Gauthier, F. A Novel Locust (*Schistocerca Gregaria*) Serine Protease Inhibitor with a High Affinity for Neutrophil Elastase. *Biochem. J.* **2006**, *400*, 467–476.
- Broadbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.
- Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J. Isotope-Labeled Protein Standards: Toward Absolute Quantitative Proteomics. *Mol. Cell. Proteomics MCP* **2007**, *6*, 2139–2149.
- Burke, J. R.; Deshong, A. J.; Pelton, J. G.; Rubin, S. M. Phosphorylation-Induced Conformational Changes in the Retinoblastoma Protein Inhibit E2F Transactivation Domain Binding. *J. Biol. Chem.* **2010**, *285*, 16286–16293.
- Cadene, M.; Chait, B. T. A Robust, Detergent-Friendly Method for Mass Spectrometric Analysis of Integral Membrane Proteins. *Anal. Chem.* **2000**, *72*, 5655–5658.
- Castaing, B.; Fourrey, J. L.; Hervouet, N.; Thomas, M.; Boiteux, S.; Zelwer, C. AP Site Structural Determinants for Fpg Specific Recognition. *Nucleic Acids Res.* **1999**, *27*, 608–615.
- Charvatova, O.; Foley, B. L.; Bern, M. W.; Sharp, J. S.; Orlando, R.; Woods, R. J. Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1692–1705.
- Chatterjee, D.; Bai, Y.; Wang, Z.; Beach, S.; Mott, S.; Roy, R.; Braastad, C.; Sun, Y.; Mukhopadhyay, A.; Aggarwal, B. B.; *et al.* RKIP Sensitizes Prostate and Breast Cancer Cells to Drug-Induced Apoptosis. *J. Biol. Chem.* **2004**, *279*, 17515–17523.
- Chen, Y.; Ouyang, G.-L.; Yi, H.; Li, M.-Y.; Zhang, P.-F.; Li, C.; Li, J.-L.; Liu, Y.-F.; Chen, Z.-C.; Xiao, Z.-Q. Identification of RKIP as an Invasion Suppressor Protein in Nasopharyngeal Carcinoma by Proteomic Analysis. *J. Proteome Res.* **2008**, *7*, 5254–5262.
- Choudhary, C.; Mann, M. Decoding Signalling Networks by Mass Spectrometry-Based Proteomics. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 427–439.
- Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325*, 834–840.
- Ciccimaro, E.; Hanks, S. K.; Yu, K. H.; Blair, I. A. Absolute Quantification of Phosphorylation on the Kinase Activation Loop of Cellular Focal Adhesion Kinase by Stable Isotope Dilution Liquid Chromatography/mass Spectrometry. *Anal. Chem.* **2009**, *81*, 3304–3313.
- Ciechanover, A. Proteolysis: From the Lysosome to Ubiquitin and the Proteasome. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 79–87.
- Cohen, P. The Regulation of Protein Function by Multisite Phosphorylation – a 25 Year Update. *Trends Biochem. Sci.* **2000**, *25*, 596–601.
- Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discov.* **2006**, *5*, 730–739.
- Corbit, K. C.; Trakul, N.; Eves, E. M.; Diaz, B.; Marshall, M.; Rosner, M. R. Activation of Raf-1 Signaling by Protein Kinase C through a Mechanism Involving Raf Kinase Inhibitory Protein. *J. Biol. Chem.* **2003**, *278*, 13061–13068.
- Corcoran, A.; Cotter, T. G. Redox Regulation of Protein Kinases. *FEBS J.* **2013**, *280*, 1944–1965.
- De Corte, V.; Demol, H.; Goethals, M.; Van Damme, J.; Gettemans, J.; Vandekerckhove, J. Identification of Tyr438 as the Major *in Vitro* c-Src Phosphorylation Site in Human Gelsolin: A Mass Spectrometric Approach. *Protein Sci. Publ. Protein Soc.* **1999**, *8*, 234–241.
- Dangi-Garimella, S.; Yun, J.; Eves, E. M.; Newman, M.; Erkeland, S. J.; Hammond, S. M.; Minn, A. J.; Rosner, M. R. Raf Kinase Inhibitory Protein Suppresses a Metastasis Signalling Cascade Involving LIN28 and Let-7. *EMBO J.* **2009**, *28*, 347–358.
- Deiss, K.; Kisker, C.; Lohse, M. J.; Lorenz, K. Raf Kinase Inhibitor Protein (RKIP) Dimer Formation Controls Its Target Switch from Raf1 to G Protein-Coupled Receptor Kinase (GRK) 2. *J. Biol. Chem.* **2012**, *287*, 23407–23417.
- Delmotte, N.; Lasasoa, M.; Tholey, A.; Heinzle, E.; Huber, C. G. Two-Dimensional Reversed-Phase X Ion-Pair Reversed-Phase HPLC: An Alternative Approach to High-Resolution Peptide Separation for Shotgun Proteome Analysis. *J. Proteome Res.* **2007**, *6*, 4363–4373.
- Deribe, Y. L.; Pawson, T.; Dikic, I. Post-Translational Modifications in Signal Integration. *Nat. Struct. Mol. Biol.* **2010**, *17*, 666–672.
- Diffley, J. F.; Stillman, B. A Close Relative of the Nuclear, Chromosomal High-Mobility Group Protein HMG1 in Yeast Mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 7864–7868.
- Dizdaroglu, M. Base-Excision Repair of Oxidative DNA Damage by DNA Glycosylases. *Mutat. Res.* **2005**, *591*, 45–59.
- Domon, B.; Aebersold, R. Options and Considerations When Selecting a Quantitative Proteomics Strategy. *Nat. Biotechnol.* **2010**, *28*, 710–721.
- Dupuis, A.; Hennekinne, J.-A.; Garin, J.; Brun, V. Protein Standard Absolute Quantification (PSAQ) for Improved Investigation of Staphylococcal Food Poisoning Outbreaks. *Proteomics* **2008**, *8*, 4633–4636.
- Dyson, H. J.; Komives, E. A. Role of Disorder in I $\kappa$ B-NF $\kappa$ B Interaction. *IUBMB Life* **2012**, *64*, 499–505.
- Engholm-Keller, K.; Larsen, M. R. Technologies and Challenges in Large-Scale Phosphoproteomics. *Proteomics* **2013**, *13*, 910–931.
- Escara-Wilke, J.; Yeung, K.; Keller, E. T. Raf Kinase Inhibitor Protein (RKIP) in Cancer. *Cancer Metastasis Rev.* **2012**, *31*, 615–620.

- Eulenburg, G.; Higman, V. A.; Diehl, A.; Wilmanns, M.; Holton, S. J. Structural and Biochemical Characterization of Rv2140c, a Phosphatidylethanolamine-Binding Protein from Mycobacterium Tuberculosis. *FEBS Lett.* **2013**, *587*, 2936–2942.
- Eves, E. M.; Shapiro, P.; Naik, K.; Klein, U. R.; Trakul, N.; Rosner, M. R. Raf Kinase Inhibitory Protein Regulates Aurora B Kinase and the Spindle Checkpoint. *Mol. Cell* **2006**, *23*, 561–574.
- Fekete, S.; Guillarme, D. Ultra-High-Performance Liquid Chromatography for the Characterization of Therapeutic Proteins. *TrAC Trends Anal. Chem.* **2014**, *63*, 76–84.
- Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64–71.
- Ferrer-Sueta, G.; Manta, B.; Botti, H.; Radi, R.; Trujillo, M.; Denicola, A. Factors Affecting Protein Thiol Reactivity and Specificity in Peroxide Reduction. *Chem. Res. Toxicol.* **2011**, *24*, 434–450.
- Ficarro, S. B.; McClelland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome Analysis by Mass Spectrometry and Its Application to *Saccharomyces Cerevisiae*. *Nat. Biotechnol.* **2002**, *20*, 301–305.
- Ficarro, S. B.; Salomon, A. R.; Brill, L. M.; Mason, D. E.; Stettler-Gill, M.; Brock, A.; Peters, E. C. Automated Immobilized Metal Affinity Chromatography/nano-Liquid Chromatography/electrospray Ionization Mass Spectrometry Platform for Profiling Protein Phosphorylation Sites. *Rapid Commun. Mass Spectrom. RCM* **2005**, *19*, 57–71.
- Frokjaer-Jensen, C.; Davis, M. W.; Ailion, M.; Jorgensen, E. M. Improved Mos1-Mediated Transgenesis in *C. Elegans*. *Nat. Methods* **2012**, *9*, 117–118.
- Fu, Z.; Smith, P. C.; Zhang, L.; Rubin, M. A.; Dunn, R. L.; Yao, Z.; Keller, E. T. Effects of Raf Kinase Inhibitor Protein Expression on Suppression of Prostate Cancer Metastasis. *J. Natl. Cancer Inst.* **2003**, *95*, 878–889.
- Fu, Z.; Kitagawa, Y.; Shen, R.; Shah, R.; Mehra, R.; Rhodes, D.; Keller, P. J.; Mizokami, A.; Dunn, R.; Chinnaiyan, A. M.; *et al.* Metastasis Suppressor Gene Raf Kinase Inhibitor Protein (RKIP) Is a Novel Prognostic Marker in Prostate Cancer. *The Prostate* **2006**, *66*, 248–256.
- Gabant, G.; Cadene, M. Mass Spectrometry of Full-Length Integral Membrane Proteins to Define Functionally Relevant Structural Features. *Methods San Diego Calif* **2008**, *46*, 54–61.
- Gallien, S.; Duriez, E.; Domon, B. Selected Reaction Monitoring Applied to Proteomics. *J. Mass Spectrom.* **2011**, *46*, 298–312.
- Gan, J.; Harper, T. W.; Hsueh, M.-M.; Qu, Q.; Humphreys, W. G. Dansyl Glutathione as a Trapping Agent for the Quantitative Estimation and Identification of Reactive Metabolites. *Chem. Res. Toxicol.* **2005**, *18*, 896–903.
- Garuti, L.; Roberti, M.; Bottegoni, G. Irreversible Protein Kinase Inhibitors. *Curr. Med. Chem.* **2011**, *18*, 2981–2994.
- Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6940–6945.
- Gersch, M.; Kreuzer, J.; Sieber, S. A. Electrophilic Natural Products and Their Biological Targets. *Nat. Prod. Rep.* **2012**, *29*, 659–682.
- Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics MCP* **2012**, *11*, O111.016717.
- Girard, P. M.; Guibourt, N.; Boiteux, S. The Ogg1 Protein of *Saccharomyces Cerevisiae*: A 7,8-Dihydro-8-Oxoguanine DNA glycosylase/AP Lyase Whose Lysine 241 Is a Critical Residue for Catalytic Activity. *Nucleic Acids Res.* **1997**, *25*, 3204–3211.
- Gleeson, M. P.; Modi, S.; Bender, A.; Robinson, R. L. M.; Kirchmair, J.; Promkatkaew, M.; Hannongbua, S.; Glen, R. C. The Challenges Involved in Modeling Toxicity Data in Silico: A Review. *Curr. Pharm. Des.* **2012**, *18*, 1266–1291.
- Gropengiesser, J.; Varadarajan, B. T.; Stephanowitz, H.; Krause, E. The Relative Influence of Phosphorylation and Methylation on Responsiveness of Peptides to MALDI and ESI Mass Spectrometry. *J. Mass Spectrom. JMS* **2009**, *44*, 821–831.
- Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. Quantitative Phosphoproteomics Applied to the Yeast Pheromone Signaling Pathway. *Mol. Cell. Proteomics MCP* **2005**, *4*, 310–327.
- Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.
- Hagan, S.; Al-Mulla, F.; Mallon, E.; Oien, K.; Ferrier, R.; Gusterson, B.; García, J. J. C.; Kolch, W. Reduction of Raf-1 Kinase Inhibitor Protein Expression Correlates with Breast Cancer Metastasis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2005**, *11*, 7392–7397.
- Haracska, L.; Prakash, L.; Prakash, S. A Mechanism for the Exclusion of Low-Fidelity Human Y-Family DNA Polymerases from Base Excision Repair. *Genes Dev.* **2003**, *17*, 2777–2785.
- Hart-Smith, G.; Chia, S. Z.; Low, J. K. K.; McKay, M. J.; Molloy, M. P.; Wilkins, M. R. Stoichiometry of *Saccharomyces Cerevisiae* Lysine Methylation: Insights into Non-Histone Protein Lysine Methyltransferase Activity. *J. Proteome Res.* **2014**, *13*, 1744–1756.



- Haydon, C. E. Identification of Novel Phosphorylation Sites on *Xenopus Laevis* Aurora A and Analysis of Phosphopeptide Enrichment by Immobilized Metal-Affinity Chromatography. *Mol. Cell. Proteomics* **2003**, *2*, 1055–1067.
- Hegeman, A. D.; Harms, A. C.; Sussman, M. R.; Bunner, A. E.; Harper, J. F. An Isotope Labeling Strategy for Quantifying the Degree of Phosphorylation at Multiple Sites in Proteins. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 647–653.
- Heinecke, J. W.; Li, W.; Daehnke, H. L., 3rd; Goldstein, J. A. Dityrosine, a Specific Marker of Oxidation, Is Synthesized by the Myeloperoxidase-Hydrogen Peroxide System of Human Neutrophils and Macrophages. *J. Biol. Chem.* **1993**, *268*, 4069–4077.
- He, T.; Alving, K.; Feild, B.; Norton, J.; Joseloff, E. G.; Patterson, S. D.; Domon, B. Quantitation of Phosphopeptides Using Affinity Chromatography and Stable Isotope Labeling†. *J. Am. Soc. Mass Spectrom.* **2004**.
- Höhn, A.; König, J.; Grune, T. Protein Oxidation in Aging and the Removal of Oxidized Proteins. *J. Proteomics* **2013**, *92*, 132–159.
- Hong, I.-S.; Lee, H.-Y.; Kim, H.-P. Novel Therapeutic Approaches for Various Cancer Types Using a Modified Sleeping Beauty-Based Gene Delivery System. *PLoS One* **2014**, *9*, e86324.
- Huang, L.; He, Z.; Guo, L.; Wang, H. Improvement of Cognitive Deficit and Neuronal Damage in Rats with Chronic Cerebral Ischemia via Relative Long-Term Inhibition of Rho-Kinase. *Cell. Mol. Neurobiol.* **2008**, *28*, 757–768.
- Huddleston, M. J.; Annan, R. S.; Bean, M. F.; Carr, S. A. Selective Detection of Phosphopeptides in Complex Mixtures by Electrospray Liquid Chromatography/mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 710–717.
- Huerta-Yepe, S.; Yoon, N. K.; Hernandez-Cueto, A.; Mah, V.; Rivera-Pazos, C. M.; Chatterjee, D.; Vega, M. I.; Maresh, E. L.; Horvath, S.; Chia, D.; *et al.* Expression of Phosphorylated Raf Kinase Inhibitor Protein (pRKIP) Is a Predictor of Lung Cancer Survival. *BMC Cancer* **2011**, *11*, 259.
- Hung, C.-W.; Tholey, A. Tandem Mass Tag Protein Labeling for Top-down Identification and Quantification. *Anal. Chem.* **2012**, *84*, 161–170.
- Ishihama, Y. 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.
- Jaillet, J.; Genty, M.; Cambefort, J.; Rouault, J.-D.; Augé-Gouillou, C. Regulation of Mariner Transposition: The Peculiar Case of Mos1. *PLoS One* **2012**, *7*, e43365.
- Jaquillard, L. Spectrométrie de Masse Supramoléculaire: Caractérisation de L'interaction Non-Covalente Entre PEBP1/RKIP Humaine et Des Analogues de Nucléotides. *Thèse de doctorat*, 2012.
- Jazirehi, A. R.; Vega, M. I.; Chatterjee, D.; Goodglick, L.; Bonavida, B. Inhibition of the Raf-MEK1/2-ERK1/2 Signaling Pathway, Bcl-xL down-Regulation, and Chemosensitization of Non-Hodgkin's Lymphoma B Cells by Rituximab. *Cancer Res.* **2004**, *64*, 7117–7126.
- Jin, M.; Bateup, H.; Padovan, J. C.; Greengard, P.; Nairn, A. C.; Chait, B. T. Quantitative Analysis of Protein Phosphorylation in Mouse Brain by Hypothesis-Driven Multistage Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 7845–7851.
- Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem.* **2010**, *2*, 949–964.
- Johnson, L. N.; Lewis, R. J. Structural Basis for Control by Phosphorylation. *Chem. Rev.* **2001**, *101*, 2209–2242.
- Julien, S.; Puig, I.; Caretti, E.; Bonaventure, J.; Nelles, L.; van Roy, F.; Dargemont, C.; de Herreros, A. G.; Bellacosa, A.; Larue, L. Activation of NF-kappaB by Akt Upregulates Snail Expression and Induces Epithelium Mesenchyme Transition. *Oncogene* **2007**, *26*, 7445–7456.
- Kaiser, S. E.; Riley, B. E.; Shaler, T. A.; Trevino, R. S.; Becker, C. H.; Schulman, H.; Kopito, R. R. Protein Standard Absolute Quantification (PSAQ) Method for the Measurement of Cellular Ubiquitin Pools. *Nat. Methods* **2011**, *8*, 691–696.
- Kapitonov, V. V.; Jurka, J. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biol.* **2005**, *3*, e181.
- Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; *et al.* A Quantitative Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.
- Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60*, 2299–2301.
- Karin, M.; Cao, Y.; Greten, F. R.; Li, Z.-W. NF-kappaB in Cancer: From Innocent Bystander to Major Culprit. *Nat. Rev. Cancer* **2002**, *2*, 301–310.
- Katta, V.; Chow, D. T.; Rohde, M. F. Applications of in-Source Fragmentation of Protein Ions for Direct Sequence Analysis by Delayed Extraction MALDI-TOF Mass Spectrometry. *Anal. Chem.* **1998**, *70*, 4410–4416.
- Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.
- Kettenbach, A. N.; Rush, J.; Gerber, S. A. Absolute Quantification of Protein and Post-Translational Modification Abundance with Stable Isotope-Labeled Synthetic Peptides. *Nat. Protoc.* **2011**, *6*, 175–186.
- Kinter, M.; Sherman, N. E. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*; John Wiley & Sons, Ed.; John Wiley: New York, 2000.

- Kokubu, M.; Ishihama, Y.; Sato, T.; Nagasu, T.; Oda, Y. Specificity of Immobilized Metal Affinity-Based IMAC/C18 Tip Enrichment of Phosphopeptides for Protein Phosphorylation Analysis. *Anal. Chem.* **2005**, *77*, 5144–5154.
- Kolch, W. Meaningful Relationships: The Regulation of the Ras/Raf/MEK/ERK Pathway by Protein Interactions. *Biochem. J.* **2000**, *351 Pt 2*, 289–305.
- Kollipara, L.; Zahedi, R. P. Protein Carbamylation: In Vivo Modification or in Vitro Artefact? *PROTEOMICS* **2013**, *13*, 941–944.
- Kostiainen, R.; Kotiaho, T.; Kuuranne, T.; Auriola, S. Liquid Chromatography/atmospheric Pressure Ionization-Mass Spectrometry in Drug Metabolism Studies. *J. Mass Spectrom. JMS* **2003**, *38*, 357–372.
- Krupnick, J. G.; Benovic, J. L. The Role of Receptor Kinases and Arrestins in G Protein-Coupled Receptor Regulation. *Annu. Rev. Pharmacol. Toxicol.* **1998**, *38*, 289–319.
- Kweon, H. K.; Håkansson, K. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. *Anal. Chem.* **2006**, *78*, 1743–1749.
- Lapko, V. N.; Smith, D. L.; Smith, J. B. Identification of an Artifact in the Mass Spectrometry of Proteins Derivatized with Iodoacetamide. *J. Mass Spectrom. JMS* **2000**, *35*, 572–575.
- Lee, B. C.; Gladyshev, V. N. The Biological Significance of Methionine Sulfoxide Stereochemistry. *Free Radic. Biol. Med.* **2011**, *50*, 221–227.
- Lee, H. C.; Tian, B.; Sedivy, J. M.; Wands, J. R.; Kim, M. Loss of Raf Kinase Inhibitor Protein Promotes Cell Proliferation and Migration of Human Hepatoma Cells. *Gastroenterology* **2006**, *131*, 1208–1217.
- Lee, J.-M.; Calkins, M. J.; Chan, K.; Kan, Y. W.; Johnson, J. A. Identification of the NF-E2-Related Factor-2-Dependent Genes Conferring Protection against Oxidative Stress in Primary Cortical Astrocytes Using Oligonucleotide Microarray Analysis. *J. Biol. Chem.* **2003**, *278*, 12029–12038.
- Leonard, P. G.; Golemi-Kotra, D.; Stock, A. M. Phosphorylation-Dependent Conformational Changes and Domain Rearrangements in Staphylococcus Aureus VraR Activation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8525–8530.
- Liao, P. C.; Leykam, J.; Andrews, P. C.; Gage, D. A.; Allison, J. An Approach to Locate Phosphorylation Sites in a Phosphoprotein: Mass Mapping by Combining Specific Enzymatic Degradation with Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Anal. Biochem.* **1994**, *219*, 9–20.
- Liebler, D. C.; Zimmerman, L. J. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry (Mosc.)* **2013**, *52*, 3797–3806.
- Li, H. Z.; Wang, Y.; Gao, Y.; Shao, J.; Zhao, X. L.; Deng, W. M.; Liu, Y. X.; Yang, J.; Yao, Z. Effects of Raf Kinase Inhibitor Protein Expression on Metastasis and Progression of Human Epithelial Ovarian Cancer. *Mol. Cancer Res. MCR* **2008**, *6*, 917–928.
- Li, J.; Lee, G. I.; Van Doren, S. R.; Walker, J. C. The FHA Domain Mediates Phosphoprotein Interactions. *J. Cell Sci.* **2000**, *113 Pt 23*, 4143–4149.
- Li, X.; Lin, C.; O'Connor, P. B. Glutamine Deamidation: Differentiation of Glutamic Acid and Gamma-Glutamic Acid in Peptides by Electron Capture Dissociation. *Anal. Chem.* **2010**, *82*, 3606–3615.
- Li, Y.; Compton, P. D.; Tran, J. C.; Ntai, I.; Kelleher, N. L. Optimizing Capillary Electrophoresis for Top-down Proteomics of 30–80 kDa Proteins. *PROTEOMICS* **2014**, *14*, 1158–1164.
- Lorenz, K.; Lohse, M. J.; Qwitterer, U. Protein Kinase C Switches the Raf Kinase Inhibitor from Raf-1 to GRK-2. *Nature* **2003**, *426*, 574–579.
- Lu, C.-T.; Huang, K.-Y.; Su, M.-G.; Lee, T.-Y.; Bretaña, N. A.; Chang, W.-C.; Chen, Y.-J.; Chen, Y.-J.; Huang, H.-D. DbPTM 3.0: An Informative Resource for Investigating Substrate Site Specificity and Functional Association of Protein Post-Translational Modifications. *Nucleic Acids Res.* **2013**, *41* (Database issue), D295–D305.
- Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. Absolute Protein Expression Profiling Estimates the Relative Contributions of Transcriptional and Translational Regulation. *Nat. Biotechnol.* **2007**, *25*, 117–124.
- Mah, R.; Thomas, J. R.; Shafer, C. M. Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 33–39.
- Ma, J.; Li, F.; Liu, L.; Cui, D.; Wu, X.; Jiang, X.; Jiang, H. Raf Kinase Inhibitor Protein Inhibits Cell Proliferation but Promotes Cell Migration in Rat Hepatic Stellate Cells. *Liver Int. Off. J. Int. Assoc. Study Liver* **2009**, *29*, 567–574.
- Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261.
- Mann, M.; Ong, S. E.; Grønborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. Analysis of Protein Phosphorylation Using Mass Spectrometry: Deciphering the Phosphoproteome. *Trends Biotechnol.* **2002**, *20*, 261–268.
- Mather, B. D.; Viswanathan, K.; Miller, K. M.; Long, T. E. Michael Addition Reactions in Macromolecular Design for Emerging Technologies. *Prog. Polym. Sci.* **2006**, *31*, 487–531.
- Mc Henry, K. T.; Ankala, S. V.; Ghosh, A. K.; Fenteany, G. A Non-Antibacterial Oxazolidinone Derivative That Inhibits Epithelial Cell Sheet Migration. *Chembiochem Eur. J. Chem. Biol.* **2002**, *3*, 1105–1111.
- Mc Henry, K. T.; Montesano, R.; Zhu, S.; Beshir, A. B.; Tang, H.-H.; Yeung, K. C.; Fenteany, G. Raf Kinase Inhibitor Protein Positively Regulates Cell-Substratum Adhesion While Negatively Regulating Cell-Cell Adhesion. *J. Cell. Biochem.* **2008**, *103*, 972–985.
- McLachlin, D. T.; Chait, B. T. Analysis of Phosphorylated Proteins and Peptides by Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2001**, *5*, 591–602.
- McLachlin, D. T.; Chait, B. T. Improved B-Elimination-Based Affinity Purification Strategy for Enrichment of Phosphopeptides. *Anal. Chem.* **2003**, *75*, 6826–6836.

- Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. Integrated Proteomic Analysis of Post-Translational Modifications by Serial Enrichment. *Nat. Methods* **2013**, *10*, 634–637.
- Mian, A. A.; Schüll, M.; Zhao, Z.; Oancea, C.; Hundertmark, A.; Beissert, T.; Ottmann, O. G.; Ruthardt, M. The Gatekeeper Mutation T315I Confers Resistance against Small Molecules by Increasing or Restoring the ABL-Kinase Activity Accompanied by Aberrant Transphosphorylation of Endogenous BCR, Even in Loss-of-Function Mutants of BCR/ABL. *Leukemia* **2009**, *23*, 1614–1621.
- Miller, R. M.; Paavilainen, V. O.; Krishnan, S.; Serafimova, I. M.; Taunton, J. Electrophilic Fragment-Based Design of Reversible Covalent Kinase Inhibitors. *J. Am. Chem. Soc.* **2013**, *135*, 5298–5301.
- Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Körner, R.; Sönksen, C. P.; Roepstorff, P. Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/ionization Mass Spectrometry Using (18)O-Labeled Internal Standards. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14*, 1226–1232.
- Miskey, C.; Izsvák, Z.; Kawakami, K.; Ivics, Z. DNA Transposons in Vertebrate Functional Genomics. *Cell. Mol. Life Sci. CMLS* **2005**, *62*, 629–641.
- Montoya, A.; Beltran, L.; Casado, P.; Rodríguez-Prados, J.-C.; Cutillas, P. R. Characterization of a TiO<sub>2</sub> Enrichment Method for Label-Free Quantitative Phosphoproteomics. *Methods* **2011**, *54*, 370–378.
- Al-Mulla, F.; Hagan, S.; Behbehani, A. I.; Bitar, M. S.; George, S. S.; Going, J. J.; García, J. J. C.; Scott, L.; Fyfe, N.; Murray, G. I.; *et al.* Raf Kinase Inhibitor Protein Expression in a Survival Analysis of Colorectal Cancer Patients. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2006**, *24*, 5672–5679.
- Al-Mulla, F.; Bitar, M. S.; Al-Maghrebi, M.; Behbehani, A. I.; Al-Ali, W.; Rath, O.; Doyle, B.; Tan, K. Y.; Pitt, A.; Kolch, W. Raf Kinase Inhibitor Protein RKIP Enhances Signaling by Glycogen Synthase Kinase-3 $\beta$ . *Cancer Res.* **2011**, *71*, 1334–1343.
- Al-Mulla, F.; Bitar, M. S.; Feng, J.; Park, S.; Yeung, K. C. A New Model for Raf Kinase Inhibitory Protein Induced Chemotherapeutic Resistance. *PLoS One* **2012**, *7*, e29532.
- Al-Mulla, F.; Bitar, M. S.; Taqi, Z.; Yeung, K. C. RKIP: Much More than Raf Kinase Inhibitory Protein. *J. Cell. Physiol.* **2013**, *228*, 1688–1702.
- Neer, E. J. Heterotrimeric G Proteins: Organizers of Transmembrane Signals. *Cell* **1995**, *80*, 249–257.
- Neves, S. R.; Ram, P. T.; Iyengar, R. G Protein Pathways. *Science* **2002**, *296*, 1636–1639.
- Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. Iodoacetamide-Induced Artifact Mimics Ubiquitination in Mass Spectrometry. *Nat. Methods* **2008**, *5*, 459–460.
- Nieto, M. A. The Snail Superfamily of Zinc-Finger Transcription Factors. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 155–166.
- Nørregaard Jensen, O. Modification-Specific Proteomics: Characterization of Post-Translational Modifications by Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2004**, *8*, 33–41.
- Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. Applying Label-Free Quantitation to Top down Proteomics. *Anal. Chem.* **2014**, *86*, 4961–4968.
- Nühse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. Large-Scale Analysis of in Vivo Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry. *Mol. Cell. Proteomics* **2003**, *2*, 1234–1243.
- Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate Quantitation of Protein Expression and Site-Specific Phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591–6596.
- Olsen, J. V.; Blagoev, B.; Gnäd, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **2006**, *127*, 635–648.
- Olsen, J. V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M. L.; Jensen, L. J.; Gnäd, F.; Cox, J.; Jensen, T. S.; Nigg, E. A.; *et al.* Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy during Mitosis. *Sci. Signal.* **2010**, *3*, ra3.
- Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics MCP* **2002**, *1*, 376–386.
- Di Palma, S.; Boersema, P. J.; Heck, A. J. R.; Mohammed, S. Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-chILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. *Anal. Chem.* **2011**, *83*, 3440–3447.
- Pandya, N.; Santani, D.; Jain, S. Role of Mitogen-Activated Protein (MAP) Kinases in Cardiovascular Diseases. *Cardiovasc. Drug Rev.* **2005**, *23*, 247–254.
- Pan, H.; Chen, K.; Chu, L.; Kinderman, F.; Apostol, I.; Huang, G. Methionine Oxidation in Human IgG2 Fc Decreases Binding Affinities to Protein A and FcRn. *Protein Sci. Publ. Protein Soc.* **2009**, *18*, 424–433.
- Pan, S.; Zhang, H.; Rush, J.; Eng, J.; Zhang, N.; Patterson, D.; Comb, M. J.; Aebersold, R. High Throughput Proteome Screening for Biomarker Detection. *Mol. Cell. Proteomics MCP* **2005**, *4*, 182–190.
- Pan, S.; Rush, J.; Peskind, E. R.; Galasko, D.; Chung, K.; Quinn, J.; Jankovic, J.; Leverenz, J. B.; Zabetian, C.; Pan, C.; *et al.* Application of Targeted Quantitative Proteomics Analysis in Human Cerebrospinal Fluid Using a Liquid Chromatography Matrix-Assisted Laser Desorption/ionization Time-of-Flight Tandem Mass Spectrometer (LC MALDI TOF/TOF) Platform. *J. Proteome Res.* **2008**, *7*, 720–730.

- Peiró, S.; Escrivà, M.; Puig, I.; Barberà, M. J.; Dave, N.; Herranz, N.; Larriba, M. J.; Takkunen, M.; Francí, C.; Muñoz, A.; *et al.* Snail1 Transcriptional Repressor Binds to Its Own Promoter and Controls Its Expression. *Nucleic Acids Res.* **2006**, *34*, 2077–2084.
- Pennisi, E. Genomics. ENCODE Project Writes Eulogy for Junk DNA. *Science* **2012**, *337*, 1159, 1161.
- Perdivara, I.; Deterding, L. J.; Przybylski, M.; Tomer, K. B. Mass Spectrometric Identification of Oxidative Modifications of Tryptophan Residues in Proteins: Chemical Artifact or Post-Translational Modification? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1114–1117.
- Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Anal. Chem.* **2006**, *78*, 4271–4280.
- Peters, B.; Trout, B. L. Asparagine Deamidation: pH-Dependent Mechanism from Density Functional Theory. *Biochemistry (Mosc.)* **2006**, *45*, 5384–5392.
- Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **2004**, *76*, 3935–3943.
- Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionation. *Nature* **1975**, *258*, 598–599.
- Posewitz, M. C.; Tempst, P. Immobilized gallium(III) Affinity Chromatography of Phosphopeptides. *Anal. Chem.* **1999**, *71*, 2883–2892.
- Potashman, M. H.; Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J. Med. Chem.* **2009**, *52*, 1231–1246.
- Prasad, R.; Liu, Y.; Deterding, L. J.; Poltoratsky, V. P.; Kedar, P. S.; Horton, J. K.; Kanno, S.-I.; Asagoshi, K.; Hou, E. W.; Khodyreva, S. N.; *et al.* HMGB1 Is a Cofactor in Mammalian Base Excision Repair. *Mol. Cell* **2007**, *27*, 829–841.
- Puente, X. S.; Sánchez, L. M.; Overall, C. M.; López-Otín, C. Human and Mouse Proteases: A Comparative Genomic Approach. *Nat. Rev. Genet.* **2003**, *4*, 544–558.
- Rawlings, N. D.; Barrett, A. J.; Bateman, A. MEROPS: The Database of Proteolytic Enzymes, Their Substrates and Inhibitors. *Nucleic Acids Res.* **2011**, *40*, D343–D350.
- Reinders, J.; Sickmann, A. State-of-the-Art in Phosphoproteomics. *Proteomics* **2005**, *5*, 4052–4061.
- Ribas, C.; Penela, P.; Murga, C.; Salcedo, A.; García-Hoz, C.; Jurado-Pueyo, M.; Aymerich, I.; Mayor, F. The G Protein-Coupled Receptor Kinase (GRK) Interactome: Role of GRKs in GPCR Regulation and Signaling. *Biochim. Biophys. Acta* **2007**, *1768*, 913–922.
- Richter, C.; Park, J. W.; Ames, B. N. Normal Oxidative Damage to Mitochondrial and Nuclear DNA Is Extensive. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6465–6467.
- Ritchie, M. A.; Gill, A. C.; Deery, M. J.; Lilley, K. Precursor Ion Scanning for Detection and Structural Characterization of Heterogeneous Glycopeptide Mixtures. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1065–1077.
- Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *J. Biomed. Mass Spectrom.* **1984**; *11* ; 601
- Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; *et al.* Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics MCP* **2004**, *3*, 1154–1169.
- Roussel, G.; Nussbaum, F.; Schoentgen, F.; Jollès, P.; Nussbaum, J. L. Immunological Investigation of a 21-Kilodalton Cytosolic Basic Protein in Rat Brain. *Dev. Neurosci.* **1988**, *10*, 65–74.
- Rubino, F. M.; Pitton, M.; Di Fabio, D.; Colombi, A. Toward an “omic” Physiopathology of Reactive Chemicals: Thirty Years of Mass Spectrometric Study of the Protein Adducts with Endogenous and Xenobiotic Compounds. *Mass Spectrom. Rev.* **2009**, *28*, 725–784.
- Ruse, C. I.; Willard, B.; Jin, J. P.; Haas, T.; Kinter, M.; Bond, M. Quantitative Dynamics of Site-Specific Protein Phosphorylation Determined Using Liquid Chromatography Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 1658–1664.
- Sahdev, S.; Khattar, S. K.; Saini, K. S. Production of Active Eukaryotic Proteins through Bacterial Expression Systems: A Review of the Existing Biotechnology Strategies. *Mol. Cell. Biochem.* **2008**, *307*, 249–264.
- Salomon, A. R.; Ficarro, S. B.; Brill, L. M.; Brinker, A.; Phung, Q. T.; Ericson, C.; Sauer, K.; Brock, A.; Horn, D. M.; Schultz, P. G.; *et al.* Profiling of Tyrosine Phosphorylation Pathways in Human Cells Using Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 443–448.
- Santarius, T.; Shipley, J.; Brewer, D.; Stratton, M. R.; Cooper, C. S. A Census of Amplified and Overexpressed Human Cancer Genes. *Nat. Rev. Cancer* **2010**, *10*, 59–64.
- Santos, J. M.; Mishra, M.; Kowluru, R. A. Posttranslational Modification of Mitochondrial Transcription Factor A in Impaired Mitochondria Biogenesis: Implications in Diabetic Retinopathy and Metabolic Memory Phenomenon. *Exp. Eye Res.* **2014**.
- Schechter, I and Berger, A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun.* 1967 Apr 20;27(2):157-62.
- Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; *et al.* Platform-Independent and Label-Free Quantitation of Proteomic Data Using MS1

Extracted Ion Chromatograms in Skyline: APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION. *Mol. Cell. Proteomics* **2012**, *11*, 202–214.

- Schmidt, A.; Kellermann, J.; Lottspeich, F. A Novel Strategy for Quantitative Proteomics Using Isotope-Coded Protein Labels. *Proteomics* **2005**, *5*, 4–15.
- Schmidt, C.; Lenz, C.; Grote, M.; Lührmann, R.; Urlaub, H. Determination of Protein Stoichiometry within Protein Complexes Using Absolute Quantification and Multiple Reaction Monitoring. *Anal. Chem.* **2010**, *82*, 2784–2796.
- Schuijser, M. M.; Bataille, F.; Hagan, S.; Kolch, W.; Bosserhoff, A.-K. Reduction in Raf Kinase Inhibitor Protein Expression Is Associated with Increased Ras-Extracellular Signal-Regulated Kinase Signaling in Melanoma Cell Lines. *Cancer Res.* **2004**, *64*, 5186–5192.
- Schwartz, P. A.; Kuzmic, P.; Solowiej, J.; Bergqvist, S.; Bolanos, B.; Almaden, C.; Nagata, A.; Ryan, K.; Feng, J.; Dalvie, D.; *et al.* Covalent EGFR Inhibitor Analysis Reveals Importance of Reversible Interactions to Potency and Mechanisms of Drug Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 173–178.
- Seet, B. T.; Dikic, I.; Zhou, M.-M.; Pawson, T. Reading Protein Modifications with Interaction Domains. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 473–483.
- Serre, L.; Vallée, B.; Bureaud, N.; Schoentgen, F.; Zelwer, C. Crystal Structure of the Phosphatidylethanolamine-Binding Protein from Bovine Brain: A Novel Structural Class of Phospholipid-Binding Proteins. *Struct. Lond. Engl.* **1993**, *6*, 1255–1265.
- Shemon, A. N.; Eves, E. M.; Clark, M. C.; Heil, G.; Granovsky, A.; Zeng, L.; Imamoto, A.; Koide, S.; Rosner, M. R. Raf Kinase Inhibitory Protein Protects Cells against Locostatin-Mediated Inhibition of Migration. *PLoS One* **2009**, *4*, e6028.
- Siepen, J. A.; Keevil, E.-J.; Knight, D.; Hubbard, S. J. Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics. *J. Proteome Res.* **2007**, *6*, 399–408.
- Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.
- Slotkin, R. K.; Martienssen, R. Transposable Elements and the Epigenetic Regulation of the Genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285.
- Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla\*, L.; Kruppa, G.; Zubarev, R.; *et al.* Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10*, 186–187.
- Speers, A. E.; Cravatt, B. F. Profiling Enzyme Activities in Vivo Using Click Chemistry Methods. *Chem. Biol.* **2004**, *11*, 535–546.
- Spicer, J. F.; Rudman, S. M. EGFR Inhibitors in Non-Small Cell Lung Cancer (NSCLC): The Emerging Role of the Dual Irreversible EGFR/HER2 Inhibitor BIBW 2992. *Target. Oncol.* **2010**, *5*, 245–255.
- Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites. *Mol. Cell. Proteomics MCP* **2007**, *6*, 1809–1817.
- Steen, H.; Jensen, O. N. Analysis of Protein-Nucleic Acid Interactions by Photochemical Cross-Linking and Mass Spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 163–182.
- Steen, H.; Jebanathirajah, J. A.; Springer, M.; Kirschner, M. W. Stable Isotope-Free Relative and Absolute Quantitation of Protein Phosphorylation Stoichiometry by MS. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3948–3953.
- Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W. Phosphorylation Analysis by Mass Spectrometry: Myths, Facts, and the Consequences for Qualitative and Quantitative Measurements. *Mol. Cell. Proteomics MCP* **2006**, *5*, 172–181.
- Stensballe, A.; Jensen, O. N. Phosphoric Acid Enhances the Performance of Fe(III) Affinity Chromatography and Matrix-Assisted Laser Desorption/ionization Tandem Mass Spectrometry for Recovery, Detection and Sequencing of Phosphopeptides. *Rapid Commun. Mass Spectrom. RCM* **2004**, *18*, 1721–1730.
- Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14*, 1793–1800.
- Stensballe, A.; Andersen, S.; Jensen, O. N. Characterization of Phosphoproteins from Electrophoretic Gels by Nanoscale Fe(III) Affinity Chromatography with off-Line Mass Spectrometry Analysis. *Proteomics* **2001**, *1*, 207–222.
- Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. A Novel MALDI LIFT-TOF/TOF Mass Spectrometer for Proteomics. *Anal. Bioanal. Chem.* **2003**, *376*, 952–965.
- Sun, S.; Zhou, J.-Y.; Yang, W.; Zhang, H. Inhibition of Protein Carbamylation in Urea Solution Using Ammonium-Containing Buffers. *Anal. Biochem.* **2014**, *446*, 76–81.
- Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528–9533.
- Tai, S. S.-C.; Bunk, D. M.; White, E.; Welch, M. J. Development and Evaluation of a Reference Measurement Procedure for the Determination of Total 3,3',5-Triiodothyronine in Human Serum Using Isotope-Dilution Liquid Chromatography-Tandem Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 5092–5096.
- Tavel, L.; Jaquillard, L.; Karsisiotis, A. I.; Saab, F.; Jouvensal, L.; Brans, A.; Delmas, A. F.; Schoentgen, F.; Cadene, M.; Damblon, C. Ligand Binding Study of Human PEBP1/RKIP: Interaction with Nucleotides and Raf-1 Peptides Evidenced by NMR and Mass Spectrometry. *PLoS One* **2012**, *7*, e36187.

- Taylor, S. W.; Fahy, E.; Murray, J.; Capaldi, R. A.; Ghosh, S. S. Oxidative Post-Translational Modification of Tryptophan Residues in Cardiac Mitochondrial Proteins. *J. Biol. Chem.* **2003**, *278*, 19587–19590.
- Testa, B.; Pedretti, A.; Vistoli, G. Reactions and Enzymes in the Metabolism of Drugs and Other Xenobiotics. *Drug Discov. Today* **2012**, *17*, 549–560.
- Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904.
- Thornberry, N. A.; Lazebnik, Y. Caspases: Enemies within. *Science* **1998**, *281*, 1312–1316.
- Thurman, E. M.; Ferrer, I. The Isotopic Mass Defect: A Tool for Limiting Molecular Formulas by Accurate Mass. *Anal. Bioanal. Chem.* **2010**, *397*, 2807–2816.
- Trakul, N.; Menard, R. E.; Schade, G. R.; Qian, Z.; Rosner, M. R. Raf Kinase Inhibitory Protein Regulates Raf-1 but Not B-Raf Kinase Activation. *J. Biol. Chem.* **2005**, *280*, 24931–24940.
- Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; *et al.* Mapping Intact Protein Isoforms in Discovery Mode Using Top-down Proteomics. *Nature* **2011**, *480*, 254–258.
- Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246.
- Vallée, B. S.; Tauc, P.; Brochon, J. C.; Maget-Dana, R.; Lelièvre, D.; Metz-Boutigue, M. H.; Bureaud, N.; Schoentgen, F. Behaviour of Bovine Phosphatidylethanolamine-Binding Protein with Model Membranes. Evidence of Affinity for Negatively Charged Membranes. *Eur. J. Biochem. FEBS* **2001**, *268*, 5831–5841.
- Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; *et al.* The Sequence of the Human Genome. *Science* **2001**, *291*, 1304–1351.
- Wang, L.; Chance, M. R. Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem.* **2011**, *83*, 7234–7241.
- Wang, Y.-C.; Peterson, S. E.; Loring, J. F. Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Res.* **2013**.
- Watts, J. D.; Affolter, M.; Krebs, D. L.; Wange, R. L.; Samelson, L. E.; Aebersold, R. Identification by Electrospray Ionization Mass Spectrometry of the Sites of Tyrosine Phosphorylation Induced in Activated Jurkat T Cells on the Protein Tyrosine Kinase ZAP-70. *J. Biol. Chem.* **1994**, *269*, 29520–29529.
- Weinert, B. T.; Iesmantavicius, V.; Moustafa, T.; Scholz, C.; Wagner, S. A.; Magnes, C.; Zechner, R.; Choudhary, C. Acetylation Dynamics and Stoichiometry in *Saccharomyces Cerevisiae*. *Mol. Syst. Biol.* **2014**, *10*, 716.
- Wilson, A. J.; Kerns, J. K.; Callahan, J. F.; Moody, C. J. Keap Calm, and Carry on Covalently. *J. Med. Chem.* **2013**, *56*, 7463–7476.
- Wolkowicz, U. M.; Morris, E. R.; Robson, M.; Trubitsyna, M.; Richardson, J. M. Structural Basis of Mos1 Transposase Inhibition by the Anti-Retroviral Drug Raltegravir. *ACS Chem. Biol.* **2014**, *9*, 743–751.
- Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73*, 5683–5690.
- Woodin, R. R.; Bomse D. S.; Beauchamp J. L., in *Chemical and Biochemical Applications of Lasers*, ed. C. B. Moore, Academic Press, New York, **1979**, vol. IV, pp. 355–388.
- Wu, R.; Haas, W.; Dephoure, N.; Huttlin, E. L.; Zhai, B.; Sowa, M. E.; Gygi, S. P. A Large-Scale Method to Measure Absolute Protein Phosphorylation Stoichiometries. *Nat. Methods* **2011a**, *8*, 677–683.
- Wu, R.; Dephoure, N.; Haas, W.; Huttlin, E. L.; Zhai, B.; Sowa, M. E.; Gygi, S. P. Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Mol. Cell. Proteomics* **2011b**, *10*, M111.009654.
- Wu, X.; Ranganathan, V.; Weisman, D. S.; Heine, W. F.; Ciccone, D. N.; O'Neill, T. B.; Crick, K. E.; Pierce, K. A.; Lane, W. S.; Rathbun, G.; *et al.* ATM Phosphorylation of Nijmegen Breakage Syndrome Protein Is Required in a DNA Damage Response. *Nature* **2000**, *405*, 477–482.
- Xia, Y.; Liang, X.; McLuckey, S. A. Ion Trap versus Low-Energy Beam-Type Collision-Induced Dissociation of Protonated Ubiquitin Ions. *Anal. Chem.* **2006**, *78*, 1218–1227.
- Yaffe, M. B. Phosphotyrosine-Binding Domains in Signal Transduction. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 177–186.
- Yang, X.-J. Multisite Protein Modification and Intramolecular Signaling. *Oncogene* **2005**, *24*, 1653–1662.
- Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic <sup>18</sup>O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001**, *73*, 2836–2842.
- Yeung, K.; Seitz, T.; Li, S.; Janosch, P.; McFerran, B.; Kaiser, C.; Fee, F.; Katsanakis, K. D.; Rose, D. W.; Mischak, H.; *et al.* Suppression of Raf-1 Kinase Activity and MAP Kinase Signalling by RKIP. *Nature* **1999**, *401*, 173–177.
- Yeung, K.; Janosch, P.; McFerran, B.; Rose, D. W.; Mischak, H.; Sedivy, J. M.; Kolch, W. Mechanism of Suppression of the Raf/MEK/extracellular Signal-Regulated Kinase Pathway by the Raf Kinase Inhibitor Protein. *Mol. Cell. Biol.* **2000**, *20*, 3079–3085.
- Yeung, K. C.; Rose, D. W.; Dhillon, A. S.; Yaros, D.; Gustafsson, M.; Chatterjee, D.; McFerran, B.; Wyche, J.; Kolch, W.; Sedivy, J. M. Raf Kinase Inhibitor Protein Interacts with NF-kappaB-Inducing Kinase and TAK1 and Inhibits NF-kappaB Activation. *Mol. Cell. Biol.* **2001**, *21*, 7207–7217.
- Ye, Y.; Fortini, M. E. Proteolysis and Developmental Signal Transduction. *Semin. Cell Dev. Biol.* **2000**, *11*, 211–221.

- Yoshida, Y.; Izumi, H.; Ise, T.; Uramoto, H.; Torigoe, T.; Ishiguchi, H.; Murakami, T.; Tanabe, M.; Nakayama, Y.; Itoh, H.; *et al.* Human Mitochondrial Transcription Factor A Binds Preferentially to Oxidatively Damaged DNA. *Biochem. Biophys. Res. Commun.* **2002**, *295*, 945–951.
- Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. *Mol. Cell. Proteomics MCP* **2009**, *8*, 2266–2284.
- Yun, C.-H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.-K.; Meyerson, M.; Eck, M. J. The T790M Mutation in EGFR Kinase Causes Drug Resistance by Increasing the Affinity for ATP. *Proc. Natl. Acad. Sci.* **2008**, *105*, 2070–2075.
- Zeng, L.; Imamoto, A.; Rosner, M. R. Raf Kinase Inhibitory Protein (RKIP): A Physiological Regulator and Future Therapeutic Target. *Expert Opin. Ther. Targets* **2008**, *12*, 1275–1287.
- Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Anal. Chem.* **2001**, *73*, 5142–5149.
- Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; *et al.* Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science* **2010**, *327*, 1000–1004.
- Zharkov, D. O.; Rieger, R. A.; Iden, C. R.; Grollman, A. P. NH<sub>2</sub>-Terminal Proline Acts as a Nucleophile in the glycosylase/AP-Lyase Reaction Catalyzed by Escherichia Coli Formamidopyrimidine-DNA Glycosylase (Fpg) Protein. *J. Biol. Chem.* **1997**, *272*, 5335–5341.
- Zhou, F.; Sikorski, T. W.; Ficarro, S. B.; Webber, J. T.; Marto, J. A. An Online Nanoflow RP-SAX-RP LC-MS/MS Platform for Efficient and In-Depth Proteome Sequence Analysis of Complex Organisms. *Anal. Chem.* **2011**, *83*, 6996–7005.
- Zhu, S.; Mc Henry, K. T.; Lane, W. S.; Fenteany, G. A Chemical Inhibitor Reveals the Role of Raf Kinase Inhibitor Protein in Cell Migration. *Chem. Biol.* **2005**, *12*, 981–991.
- Zubarev RA, Kelleher NL and McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* *120*, 3265-3266

# Webographie

[a] <http://www.unimod.org>

[b] <http://dbptm.mbc.nctu.edu.tw/>

[c] <http://merops.sanger.ac.uk/>

[d] <http://www.ionsource.com/Card/carbam/carbam.htm>

[e] [http://reflexions.ulg.ac.be/cms/c\\_15009/fr/siblings-les-armes-multifonctionnelles-du-cancer?portal=j\\_55&printView=true](http://reflexions.ulg.ac.be/cms/c_15009/fr/siblings-les-armes-multifonctionnelles-du-cancer?portal=j_55&printView=true) (accessed Sep 16, 2014).



## Guillaume GABANT

### Spectrométrie de masse des modifications induites ou post-traductionnelles de protéines : méthodologie et application à des protéines d'intérêt thérapeutique

#### Résumé :

Les modifications de protéines, qu'elles soient post-traductionnelles (PTMs) ou induites chimiquement, ont une influence considérable sur l'activité des protéines. Des méthodes de spectrométrie de masse (MS) HRMS, MS/MS CID et ETD, et de biochimie ont été développées pour la caractérisation structurale et cinétique de complexes protéine-ligand et de PTMs, dans le but de disséquer leur mécanisme et de concevoir des médicaments covalents contre des protéines liant des protéases, des kinases, ou l'ADN.

La MS combinée avec des outils biochimiques a permis de séquencer l'inhibiteur de protéases grégliné, et de détecter une PTM originale. De même, la transposase MOS1, modèle de l'intégrase du VIH pour la conception d'inhibiteurs, s'avère être à la fois acétylée et phosphorylée. Pour la lyase Abf2, une stratégie de piégeage, purification, protéolyse et hydrolyse ADN du complexe covalent Abf2-ADN, couplée à l'analyse MS, a été développée. Enfin, l'interaction entre le surpresseur de métastase hPEBP1 et la locostatine a été disséquée sur la protéine entière et par approche bottom-up. La locostatine s'hydrolyse en butyrate après fixation. Afin d'identifier le site ciblé par la locostatine, les conditions de réaction et de protéolyse ont été optimisées. La présence de réactions non spécifiques a conduit au développement 1) d'un modèle mathématique permettant de déterminer la fraction de liaison optimale pour discriminer le site spécifique des sites non-spécifiques, et 2) d'une méthode pour la quantification parallèle et exhaustive du degré de modification de tous les sites modifiés d'une protéine. Ces outils sont applicables aux ligands covalents de protéines et/ou à leurs PTMs.

Mots clés : modification de protéine, PTM, ligand covalent, quantification, spectrométrie de masse.

### Mass spectrometry for induced or post-translational modifications: methodology and application to proteins of therapeutic interest

#### Abstract :

Protein modifications, whether post-translational (PTMs) or chemically induced, play a crucial role on the activity of proteins. Mass spectrometry (MS) techniques such as HRMS, CID/ETD MS/MS, and biochemistry-based methods for structural and kinetic characterization of protein-ligand complexes and PTMs have been developed.

MS combined with several biochemical tools has been used to sequence the proteinase inhibitor grégliné and to detect a novel PTM. A similar approach shows that the transposase MOS1, a model for the design of HIV integrase inhibitors, is both phosphorylated and acetylated. For the lyase Abf2, a strategy of trapping, purification, proteolysis, and DNA hydrolysis of the Abf2-DNA covalent complex, coupled to MS analysis, has been developed. Finally, the interaction between the metastasis suppressor hPEBP1 and locostatin was dissected. Upon binding to hPEBP1, locostatin undergoes hydrolysis. To identify the site targeted by locostatin, the conditions of reaction and proteolysis were optimized. The qualitative approach reveals the presence of non-specific reactions, leading to the development of 1) a mathematical model to determine the optimum bound fraction for discriminating the specific site from non-specific sites, and 2) a method for the parallel and exhaustive quantification of the degree of modification of all modified sites of a protein. These tools are widely applicable to covalent protein ligands and/or PTMs.

Keywords : protein modification, PTM, covalent ligand, quantification, mass spectrometry



CBM CNRS,  
Rue Charles Sadron, CS 80054  
45071 Orléans Cedex 2

