



**HAL**  
open science

# Spectrométrie de masse native et pontage chimique pour l'analyse structurale de complexe protéique

Christophe Giorgiutti

## ► To cite this version:

Christophe Giorgiutti. Spectrométrie de masse native et pontage chimique pour l'analyse structurale de complexe protéique. Chimie analytique. Université de Strasbourg, 2020. Français. NNT : 2020STRAF020 . tel-03622476

**HAL Id: tel-03622476**

**<https://theses.hal.science/tel-03622476>**

Submitted on 29 Mar 2022

**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 DES SCIENCES CHIMIQUES**

**UMR 7140 – Chimie de la Matière Complexe**

**THÈSE** présentée par :

**Christophe GIORGIUTTI**

soutenue le : 26 Novembre 2020

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie/Chimie Analytique

**Spectrométrie de masse native et  
pontage chimique pour l'analyse  
structurale de complexe protéique**

**THÈSE dirigée par :**

**Dr. Emmanuelle LEIZE-WAGNER** Directeur de recherche, Université de Strasbourg

**RAPPORTEURS :**

**Pr. Frédéric AUBRIET**

Rapporteur externe, Université de Lorraine

**Dr. Virginie REDEKER**

Rapporteur externe, MIRCen, CEA-Fontenay-aux-Roses

---

**AUTRES MEMBRES DU JURY :**

**Pr. Burkhard BECHINGER**

Examineur, Université de Strasbourg

**Dr. Claudio IACOBUCCI**

Examineur, Martin Luther University Halle-Wittenberg

**Dr. Noelle POTIER**

Co-encadrant, Université de Strasbourg



*A mes grands-parents, Irène et Raymond Leiber,*

*A mes parents,*

*A ma famille*

« La suprême récompense du travail n'est pas ce qu'il vous permet de gagner, mais ce qu'il vous permet de devenir. »

John Ruskin



# Remerciements

---

J'aimerais remercier dans un premier temps mes rapporteurs, le Dr. Virginie Redeker et le Pr. Frédéric Aubriet, ainsi que les autres membres, le Pr. Burkhard Bechinger et le Dr. Claudio Iacobucci pour avoir accepté d'évaluer mes travaux de recherche.

Je tenais également à remercier la Fondation Recherche Chimie (FRC) de m'avoir octroyé un financement afin mener à bien cette thèse et, ce dans les meilleures conditions.

Je souhaiterais dans un premier temps exprimer toute ma gratitude au Dr. Emmanuelle Leize-Wagner, directrice du Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) pour son accueil et son encadrement. Merci de m'avoir accordé ta confiance pour réaliser cette thèse ainsi que ton investissement dans l'ensemble des projets. J'ai énormément appris à tes côtés autant sur le plan scientifique que sur le plan personnel. Je ne pourrais jamais assez te remercier pour tout ce que tu as fait pour moi.

Je souhaiterais également remercier le Dr. Noelle Potier pour son encadrement et son soutien lors de cette thèse. Merci d'avoir cru en moi et de m'avoir accordé ta confiance. Merci de m'avoir encadré tout au long de cette thèse. Merci de m'avoir fait découvrir la spectrométrie de masse, pour tes précieux conseils et ton soutien dans les moments difficiles de la thèse et pour ta gentillesse.

Je tenais également à remercier l'ensemble des personnes qui ont contribué à cette thèse en particulier mes collaborateurs à l'IGBMC : le Dr. Natacha Rochel, le Dr. Judit Osz ainsi que Carole Peluso-Iltis. Pour votre aide et votre patience lors des différents projets que l'on a démarrés ensemble. J'ai énormément apprécié travailler à vos côtés. Merci pour tous vos conseils et votre efficacité dans la préparation des nombreux échantillons que je vous ai demandés. Merci à toi Judit, pour ton aide, ta bonne humeur ainsi que ton soutien durant cette thèse que ce soit pour le projet RAR/RXR ou pour le projet CBP.

Merci également au Pr. Patrick Schultz et à Alexandre Frechard pour leur aide et leur partenariat pour le projet NuA4. Merci de m'avoir accordé votre confiance et pour toutes les explications que vous m'avez apportées sur ce complexe et sur les analyses réalisées en microscopie électronique.

J'aimerais également remercier l'ensemble des collaborateurs d'ISIS : le Dr. Vladimir Torbeev, le Dr. Ana Maria Garcia ainsi que le Dr. Aromal Asokan pour leur confiance et leur aide dans ce beau projet sur l'étude de l'agrégation de la protéine NCBD. Merci pour votre patience et votre confiance. Ce fut un réel plaisir et un grand honneur de travailler avec vous.

Je tiens également à remercier l'ensemble des membres du LSMIS. Merci au Dr. Armelle Charrié et à Nathalie Zorn pour leur aide et leur soutien.

Un grand Merci au Dr. Yannis François pour son soutien, son aide pendant toute cette aventure. Tu as toujours trouvé les mots pour m'aider à reprendre confiance et à relativiser. Tu m'as soutenu dans les moments les plus difficiles, pour cela, je t'en suis infiniment reconnaissant.

La thèse m'a également permis de rencontrer des personnes formidables avec qui j'ai noué des liens forts. Merci à toi Jérémie pour ton aide, ta bonne humeur, ton soutien et pour tous tes jeux de mots... même les plus pourris. Je suis très fier de te compter parmi mes amis ! Merci de m'avoir supporté même pendant les moments les plus difficiles. Rien que pour cela, je te souhaite le meilleur ! C'était vraiment top de bosser avec toi ! Je te souhaite le meilleur pour la suite et en espérant te revoir bientôt. Je tenais également à remercier le Dr. Antony LECHNER. On s'est connu tous les deux au Master 1 et on a terminé nos thèses la même année. Merci pour ton soutien, ton aide et ta bonne humeur ! Merci de m'avoir supporté durant toutes ces années, c'était vraiment top. Avec Jerem on a vraiment été un trio de choc !

Merci d'avoir supporté mes blagues et mes jeux de mots les plus pourris. Je te souhaite le meilleur pour la suite ! Merci les gars pour ces trois ans fantastiques. Je tenais également à remercier Clarisse Gosset-Erard et Mathieu Galmiche qui nous ont rejoints lors de la troisième année. Merci à toi Clarisse pour ta bonne humeur, ton aide, ton soutien, même dans les moments les plus difficiles. Merci de m'avoir aidé à trouver des publications et de m'avoir corrigé et relu ma thèse. Merci pour le temps que tu m'as consacré pour revoir l'ensemble de la thèse. Je te souhaite plein de bonheur pour la suite. Je suis certain que tu feras une thèse formidable... et que tu réussiras à finir les mots fléchés force 2 voire force 3 avant la fin de ta thèse. Merci à toi Mathieu pour ton aide et ton soutien et pour avoir repris la traditionnelle pause-café de 10 h. Merci aussi pour tes jeux de mots pourris et pour les bons moments que l'on a passés ensemble. Je suis certain qu'un jour tu arriveras à battre le record de Yannis aux fléchettes. Je te souhaite bon courage et bonne continuation pour la suite.

Merci également à Tessa Reinert qui nous a rejoint dans les deux derniers mois de thèse. Je te souhaite plein de bonnes choses pour le début de cette aventure !

Merci également à Stéphanie que j'ai eu l'honneur de côtoyer et qui m'a beaucoup apporté. Tu as toujours fait preuve de beaucoup de patience envers Jerem, Anto et moi. Merci pour ton soutien et ton aide. Un jour, on arrivera à faire les mots fléchés force 4 et les sudokus experts !

Merci également à Gaël, Julien et Yu Minh, les doctorants du labo d'à côté. On était tous ensemble dans la même situation et on est restés soudés tout au long de nos thèses. Merci pour les discussions plus ou moins scientifiques de nos fins de journées.

J'aimerais également remercier l'ensemble des étudiants en stage qui ont contribué à cette thèse ou que j'ai pu côtoyer : Merci à Lucas, Lionel, Ludovic, Florian et Geoffrey pour leur aide et leurs implications dans cette thèse. Un grand merci également à Apolline, Thomas, Valentin, Marie, Josiane, Elise et Eminé et tous ceux que j'aurais pu oublier qui nous ont rejoints.

Enfin, mes plus gros remerciements vont à des personnes aussi chaleureuses que compétentes. Merci à l'ensemble du personnel de la plateforme protéomique : Philippe Hammann, Lauriane Kuhn et Johana Chicher. Merci d'avoir répondu à mes nombreuses interrogations ou lors de mes nombreuses demandes en analyse d'échantillons sur l'orbitrap ! Je tenais à vous remercier pour votre aide, votre soutien et votre bonne humeur. Je vous remercie également de m'avoir pris en tant que stagiaire à la plateforme pour me former à l'analyse protéomique et pour votre confiance légendaire.

Merci également aux techniciens de la salle de TP qui m'ont suivi pendant tout le monitorat, j'ai nommé Mme Laura Bom et Mr. Thomas Weissenberger ! Merci de votre soutien, de votre aide et de votre gentillesse. Je vous souhaite à tous les deux bonnes continuations pour la suite. Merci également au Pr. Maurice Millet pour son encadrement lors de ma première année de monitorat.

J'aimerais également remercier les membres du laboratoire de Spectroscopie et de Bioélectrochimie qui m'ont accueilli lors de mes stages en master et qui m'ont donné l'envie de faire cette thèse. En particulier, je voudrais remercier dans un premier temps le Pr. Pétra Hellwig. Merci de m'avoir accueilli au sein de votre laboratoire, de m'avoir fait confiance et de m'avoir donné envie de continuer dans la recherche. Je voudrais dans un second temps également remercier les Dr. Frédéric Melin, Dr. Youssef El Khoury pour leur encouragement et leur soutien depuis le master et jusqu'à la fin de cette thèse. Dans un dernier temps, j'aimerais également remercier le Dr. Zahia Boubegtiten. Merci Zahia de m'avoir accordé ta confiance, de m'avoir formé sur les différentes techniques spectroscopiques, de ton soutien et de ton aide précieuse. Ce fut très agréable de travailler en ta compagnie.

Pour conclure, je voudrais remercier ma famille, Merci d'avoir toujours cru en moi et d'avoir été présent à mes côtés.

J'aimerais également adresser une pensée particulière à mes grands-parents, Raymond et Irène Leiber à qui je dédie ces travaux de thèse.

Bien entendu, cette liste n'est pas exhaustive et je remercie tous celles et ceux qui m'ont encouragé...



# Table des matières

Liste des abréviations .....	1
Liste des figures.....	5
Liste des tableaux.....	13
Introduction générale .....	15
Partie 1 : Introduction Bibliographique.....	19
Chapitre I : La MS et son implication dans l'étude des complexes biologiques .....	21
1. Introduction .....	21
2. Les sources d'ionisation .....	22
3. L'interface.....	28
4. Les analyseurs .....	28
5. Spectrométrie de masse en tandem (MS/MS) .....	34
6. Conclusion .....	37
7. Bibliographie .....	39
Chapitre II : Stratégies d'analyse protéomique par MS .....	43
1. Introduction .....	43
2. Stratégies d'identification de protéines et de peptides par spectrométrie de masse .....	44
3. Bibliographie .....	50
Chapitre III : Caractérisation structurale de complexes biologiques non-covalents .....	53
1. Les différentes techniques structurales.....	54
2. La spectrométrie de masse .....	72
3. Conclusion .....	104
4. Bibliographie .....	106
Partie 2 : Résultats.....	121
Chapitre IV : Etude de l'oligomérisation d'un coactivateur nucléaire par spectrométrie de masse native .....	123
1. Introduction .....	123
2. Etude de l'auto-assemblage de NCBD par MS native.....	125
3. Etude des interactions entre NCBD et TAD par MS native .....	143
4. Conclusions et Perspectives.....	156
5. Bibliographie .....	158
Chapitre V : Développement de la spectrométrie de masse pour la caractérisation structurale de complexe biologique par pontage chimique (XL-MS).....	158
1. Introduction .....	161

2. RAR/RXR comme modèle : mise en place d'un protocole de pontage chimique.....	163
3. Enrichissement en données de peptides pontés : sonder de façon plus précise la structure.....	177
4. Etude des changements conformationnels de RAR-RXR lors de sa liaison avec l'ADN .....	200
5. Mise en évidence d'un réarrangement conformationnel par analyse quantitative.....	215
6. Conclusion et Perspectives .....	227
7. Bibliographie .....	228
Chapitre VI : Vers la caractérisation d'une protéine partiellement désordonnée par XL-MS : la protéine CBP.....	228
1. Introduction .....	233
2. Protocole expérimental.....	237
3. Résultats et Discussion .....	239
4. Conclusions et Perspectives.....	252
5. Bibliographie .....	253
Chapitre VII : XL-MS, une aide à la compréhension des données de cryo-EM : vers la génération d'un modèle structural du complexe NuA4 .....	257
1. Introduction .....	257
2. Matériels et méthode.....	259
3. Conclusions et Perspectives.....	269
4. Bibliographie .....	271
Conclusion générale et perspectives .....	273
Liste des Communications scientifiques.....	277
Annexes.....	279

## Liste des abréviations

---

[ ]	Concentration
°C	Degré Celsius
Å	Angström
Δ	Delta
AA	Acide aminé
ACN	Acétonitrile
AcONH <sub>4</sub>	Acétate d'ammonium
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
BS3	Bis(sulfosuccinimidyl) suberate
BSA	Albumine sérique bovine
XL-MS	Pontage chimique couplé à la spectrométrie de masse (Cross-linking couplé à la spectrométrie de masse)
Da	Dalton
ESI	Ionisation par Electrospray
DESI	Désorption par ionisation électrospray
HEPES	Acide 4-(2-hydroxyéthyl)-1- pipérazine éthane sulfonique
LC	Chromatographie Liquide
M	Molaire (Mole par litre)
MALDI	Désorption-ionisation laser assisté par matrice
μL	Microlitre
μS	Microseconde
MS	Spectrométrie de masse
MSMS	Spectrométrie de masse en tandem
m/z	Rapport masse sur charge
NanoLC	Nano chromatographie liquide
ppm	Partie par million
Q	Quadripôle
QqQ	Triple-quadripôles
RAR	Récepteur de l'acide rétinoïque
RXR	Récepteur X des rétinoïdes
SA	Acide sinapinique
SDS-PAGE	Electrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium
TFA	Acide trifluoroacétique
TOF	Analyseur à temps de vol
IT	Piège à ions
V	Volt
IM-MS	Spectrométrie de masse couplée à la mobilité ionique
HDX-MS	Spectrométrie de masse couplée à l'échange proton/deutérium (HDX-MS)
Native – MS	Spectrométrie de masse native
D <sub>2</sub> O	Eau deutérée
HCCA	α-Cyano-4-hydroxycinnamic acid

SA	Acide sinapique
ICD	Ion conversion dynode
FTICR	Analyseur à résonance ionique cyclotronique à transformée de Fourier
CID	Dissociation induite par collision
CAD	Dissociation activée par collision
HCD	Dissociation induite par collision à haute énergie
pH	Potentiel hydrogène
NaCl	Chlorure de sodium
KCl	Chlorure de potassium
nL/min	nanolitre par minute
EMR	Extended Mass Range
ETD	Dissociation par transfert d'électron
ETHcd	Dissociation par collision à transfert d'énergie et à haute énergie
MS/MS	Spectrométrie de masse en tandem
CE	Electrophorèse Capillaire
HPLC	Chromatographie Liquide à Haute Pression
PTM	Modification Post-traductionnelle
H <sub>2</sub> O	Molécule d'eau
NH <sub>3</sub>	Molécule d'ammoniac
FA	Acide Formique
TFA	Acide Trifluoroacétique
PSM	Peptide Spectrum Match
FDR	False Discovery Rate
ITC	Calorimétrie par titration isotherme
SPR	Résonance Plasmonique de Surface
EM	Microscopie Electronique
Cryo-EM	Cryo-Microscopie
RMN	Résonance Magnétique Nucléaire
DRX	Diffraction Rayon X
FTIR	Infrarouge à transformées de Fourier
ATR	Réflexion totale atténuée
SAXS	Diffraction des rayons X aux petits angles
CD	Dichroïsme Circulaire
HADDOCK	High Ambiguity Driven DOCKing
Na <sub>2</sub> HPO <sub>4</sub>	Hydrogénophosphate de sodium
MCP	Galette de micro-canaux
IRMPD	InfraRed MultiPhoton Dissociation
UVPD	UltraViolet PhotoDissociation
DTT	Dithiothréitol
TCEP	Tris-(2 carboxyéthyl) phosphine
DDA	Data-dependant acquisition
PDB	Protein Data Bank
RCSB PDB	Research Collaboratory for Structural Bioinformatics PDB
COSY	COrrrelation SpectroscopY
NOESY	Nuclear Overhauser Effect SpectroscopY
TROSY	Transverse Relaxation Optimized SpectroscopY
CDD	Charge coupled device
CMC	Concentration micellaire critique

DT	Tube de dérive
TW	Onde progressive
FAIMS	Field-asymmetric ion mobility spectrometry
NHS	N-hydroxysuccinimides
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
PIR	Protein Interaction Reporter
BSA	Bovin Serum Albumin
NCBD	Nuclear Coactivator Binding Domain
TAD	Transactivation Domain
pTAD	Transactivation Domain phosphorylé
CBP	CREB-Binding Protein
ISCID	In-Source Collision-Induced-Dissociation
AD1	Activation Domain 1
AD2	Activation Domain 2
RNs	Récepteurs Nucléaires
SEC	Chromatographie Exclusion Stérique
SCX	Chromatographie Echange de Cation
DBD	Domaine de liaison à l'ADN
LBD	Domaine de liaison au ligand
DR	Direct Repeat
DIA	Data Independent Acquisition
cAMP	Responsive Elements
IDP	intrinsically disordered protein
CRE	cAMP Responsive Eléments
PPAR	Peroxisome Proliferator Activated Receptor
TBP	TATA box Binding Protein - composant de TFIID
SRC-1	Steroid Receptor Coactivator 1
HAT	histone acétyltransférase
BRD	Bromodomaine
CH2	cystéine / histidine 2
CH3	cystéine / histidine 3
HAT	histones acétyltransférases
NuA4	Nucléosome Acetyltransferase de H4
SBAT	1,1'-(suberoyldioxy) bisazabenzotriazole
BS(PEG) <sub>5</sub>	PEGylated bis(sulfosuccinimidyl)suberate
DSSO	Disuccinimidyl sulfoxide
DSBU	Disuccinimidyl dibutyric urea
CDI	N,N'-Carbodiimide
LC-SMCC	Succinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxy-[6-amidocaproate]



# Liste des figures

---

Figure 1 : Schéma représentant les différentes composantes d'un spectromètre de masse.....	21
Figure 2 : Représentation de l'ionisation d'un échantillon en MALDI-MS extrait de [2].....	22
Figure 3 : Spectre de masse de capsid de virus obtenu par MALDI-MS avec un détecteur HM4 de la société CovalX® permettant la détection d'espèce biologique à haut poids moléculaire (jusqu'à 2 MDa). Extrait de la documentation de CovalX® .....	25
Figure 4 : Schéma de la source à ionisation électrospray (ESI) adapté de [27] .....	26
Figure 5 : Schéma représentative de la source DESI extrait de [28] .....	27
Figure 6 : Représentation schématique de la migration des ions dans un MALDI-TOF.....	28
Figure 7 : Représentation d'un quadripôle ainsi que le mouvement des ions à l'intérieur de celui-ci.....	30
Figure 8 : Représentation schématique de l'orbitrap et du déplacement d'ions (Source : figure extraite de [35] ) .....	31
Figure 9 : Représentation schématique du spectromètre de masse ESI-Orbitrap (Orbitrap Qexactive Plus) (source : adapté de ThermoFisher) .....	31
Figure 10 : Schéma représentant les différentes composantes d'un spectromètre de masse en mode MS/MS..	34
Figure 11 : Représentation schématique d'un spectromètre de masse (a) de type Q-TOF (extrait de [48]) et de (b) l'orbitrap Qexactive plus (adapté du manuel d'utilisation de ThermoFisher).....	38
Figure 12 : Représentation schématique des différentes stratégies protéomiques utilisées. ....	43
Figure 13 Représentation schématique de la stratégie protéomique "bottom-up".....	45
Figure 14 : Distribution statistique des acides aminés présents dans les protéines dans la base de données Uniprot. ....	46
Figure 15 : Nomenclature de la fragmentation peptidique.....	48
Figure 16 : les différentes techniques d'analyse structurale. ....	54
Figure 17 : Principe de la RMN.....	55
Figure 18 : Principe de la DRX. Extrait de [26].....	57
Figure 19 : Principe de la CryoEM. Extrait et adapté de [43] .....	59
Figure 20 : Principe de la SAXS. Extrait de [46] .....	61
Figure 21 : Transmission du rayon UV-VIS à travers un échantillon ( $I_0$ est l'intensité lumineuse initiale et $I_t$ est l'intensité transmise).....	63
Figure 22 : Positions des structures secondaires au sein de la bande amide I .....	66
Figure 23 : Principe d'un spectrophotomètre Raman.....	68
Figure 24 : Schéma récapitulant les différentes étapes de la caractérisation structurales des complexes protéiques <sup>[74]</sup> .....	69
Figure 25 : Les différentes techniques complémentaires pour l'analyse structurale de complexe biologique adapté de <sup>[83]</sup> .....	73
Figure 26 : schéma explicatif du fonctionnement de la tension d'accélération. ....	76

Figure 27 : Influence du Prepulse sur les spectres MS. (a) Prepulse = 30 $\mu$ s (b) Prepulse = 40 $\mu$ s. Discrimination de l'ion 1312.96 par augmentation du paramètre de prepulse lors de l'étude de l'interaction entre NCBD et TAD.	77
Figure 28 : Comparaison spectre ESI-MS de masse de la myoglobine en conditions natives (a) et en conditions dénaturantes (b) .....	78
Figure 29 : Spectre de masse à haute résolution de la protéine NCBD sur une Orbitrap Qexactive plus EMR mettant en évidence la superposition des profils isotopiques d'une espèce monomérique et dimérique à m/z 1750.66 ..	80
Figure 30 : Schéma d'une cellule de mobilité ionique. Extrait de [120].....	81
Figure 31 : Schéma des différentes cellules de mobilité ionique. Extrait [122]. .....	82
Figure 32 : Représentation schématique de la caractérisation structurale de complexe biologique par la spectrométrie de masse en couplage avec une stratégie d'échange hydrogène/Deutérium (HDX-MS).....	84
Figure 33 : Représentation Schématique de la réaction de pontage .....	86
Figure 34 : Nombre d'articles scientifiques publiés chaque année sur l'approche XL-MS et répertoriés dans la base PubMed (recherche effectuée en septembre 2020). .....	87
Figure 35 : Représentation schématique de la caractérisation structurale de complexes biologiques par la spectrométrie de masse en couplage avec une stratégie de pontage chimique (XL-MS) .....	88
Figure 36 : Représentation schématique d'un agent pontant.....	89
Figure 37 : structure du BS3 .....	90
Figure 38 : Mécanisme réactionnel de la réaction d'hydrolyse sur un ester NHS. ....	91
Figure 39 : Structure du SMCC, un agent pontant hétérobifonctionnel réagissant sur les amines et les thiols. ...	92
Figure 40 : Réaction de pontage de l'EDC sur les protéines. (Extrait du manuel d'instruction de ThermoFisher <sup>[192]</sup> .....	92
Figure 41 : Mécanisme réactionnel de l'agent pontant photoréactif NHS-Diazirine.( d'après ThermoFisher) <sup>[196]</sup>	93
Figure 42 : Famille des agents pontant se liant sur les amines avec des tailles de bras espaceurs variées .....	94
Figure 43 : Stratégie d'enrichissement par agent pontant clivable : exemple du DSSO.....	97
Figure 44 : Représentation schématique des étapes de XL-MS.....	98
Figure 45 : Représentation schématique de l'algorithme de pLink, adapté de [164].....	100
Figure 46 : Représentation des différents domaines de CBP. Extrait de [18] .....	124
Figure 47 : Représentation structurale de la protéine NCBD. Extrait de [23].....	124
Figure 48 : Séquence des échantillons de NCBD. Nle est la Norleucine incorporée à la place de la méthionine native pour éviter les phénomènes d'oxydation. ....	126
Figure 49 : Spectre MS native montrant l'oligomérisation de L-NCBD à (a) 10 $\mu$ M, (b) 150 $\mu$ M et (c) 300 $\mu$ M. Le zoom montre la région en m/z où les ions trimériques et tétramériques ont été observés. Le profil isotopique des ions à charges multiples ainsi que les valeurs de $\Delta$ m/z mesurées pour les adduits de TFA (marqués avec des étoiles) ont confirmé les états de charge de toutes les espèces et ont ainsi permis d'identifier chaque oligomère. ....	127
Figure 50 : Spectre MS natif montrant l'effet de l'énergie de collision in source (ISCID) sur le L-NCBD à (a) 0eV, (b) 40 eV et (c) 80 eV .....	130

Figure 51 : Comparaison de résolution entre une analyse MS native réalisée sur un microTOF de Bruker (b et d) et une orbitrap Qexactive plus EMR (a et c). A l'aide de la haute résolution, il a été possible de déterminer l'état de charge des ions sans ambiguïté afin de confirmer l'identité de chaque oligomère. ....	142
Figure 52 : Organisation du domaine de p53 et CBP/p300. (A) Domaines de p53. TAD (domaine de transactivation N-terminal), riche en P (riche en proline), DBD (domaine de liaison à l'ADN), TD (domaine de tétramérisation) et REG (domaine de réglementation C-terminal). L'emplacement des motifs AD1 et AD2 est indiqué sur une séquence d'acides aminés partielle ; les sites connus de phosphorylation sont indiqués par des points. (B) Domaines de CBP / p300. Les domaines qui interagissent avec le domaine de transactivation p53, TAZ1 (résidus 340 - 439), KIX (586 - 672), TAZ2 (1764 -1855), NCBD (2059 -2117) sont affichés en noir. Figure extrait de [38] .....	144
Figure 53 : Les méthionines ont été remplacées par des résidus Norleucine, quasi isostérique pour empêcher l'oxydation lors de la synthèse et des analyses ultérieures. ....	145
Figure 54 : Spectre MS native de TAD. Détection majoritaire de sa forme monomérique à différents états de charge. ....	146
Figure 55 : Spectre MS native de TAD obtenu en fonction de l'énergie de collision in source (ISCID) à 40eV et à 80eV. Les étoiles représentent la présence d'adduit de NH <sub>4</sub> <sup>+</sup> . ....	146
Figure 56 : Spectre de masse native des domaines activations de TAD (a) de AD2 à 150μM et (b) de AD1 à 150μM. Les étoiles représentent les adduits de NH <sub>4</sub> <sup>+</sup> . ....	147
Figure 57 : Spectre de masse native de l'interaction entre le domaine TAD de p53 et le domaine NCBD de CBP selon un ratio NCBD/TAD (2/1) et selon un ratio NCBD/TAD (1/2). Dans un souci de visibilité sur le spectre, seules les espèces complexées ne sont ici annotées, les pics vierges de symboles correspondent à TAD ou NCBD libre. ....	148
Figure 58 : zoom de la région m/z illustrant la résolution nécessaire pour discriminer les espèces libres TAD <sub>4</sub> <sup>+</sup> et complexées (NCBD/TAD : 1/1) <sup>7+</sup> . ....	149
Figure 59 : Schéma montrant que plusieurs hypothèses peuvent expliquer les stœchiométries des complexes formés entre TAD et NCBD telles qu'observées sur le spectre de masse.....	150
Figure 60 : Spectre de masse native montrant l'interaction du domaine d'activation AD1 de p53 avec le domaine NCBD de la protéine CBP (a) à un ratio 1/5 (NCBD/AD1) et (b) à un ratio 1/10 (NCBD/AD1).....	151
Figure 61 : Spectre de masse native montrant l'interaction du domaine d'activation 2 de p53 avec le domaine NCBD de la protéine CBP (a) à un ratio 1/5 (NCBD/AD2) et (b) à un ratio 1/10 (NCBD/AD2).....	152
Figure 62 : Spectre MS native du domaine TAD phosphorylé (p-TAD). Détection majoritaire de sa forme monomérique à différents états de charge. Dans l'encadré se trouve un agrandissement de la zone 1905-1925 m/z qui permet de visualiser les adduits d'ammonium (+17Da) ainsi que la co-détection du monomère chargé 4+ et du dimère chargé 8+. ....	154
Figure 63 : Spectre de masse native de l'interaction entre le domaine pTAD de p53 et le domaine NCBD de CBP selon un ratio NCBD/pTAD (1/2) (a) et selon un ratio NCBD/pTAD (2/1).....	155
Figure 64 : Distribution statistique des espèces. ....	156
Figure 65 : Représentation schématique de RAR/RXR. ....	162

Figure 66 : (A) Représentation schématique des différents domaines observés sur RAR et RXR. (B) Séquence des protéines RAR et RXR. L'ensemble des lysines du complexes ont été mises en rouge. Le soulignement correspond au domaine LBD (surligné en bleu) et DBD (surligné en gris) .....	163
Figure 67 : Suivi cinétique de la réaction de pontage par MALDI-MS (a) Spectre MS avant pontage (b) Spectre MS après pontage par le BS3 (ratio protéine/BS3 : 1/50 ; 30min à T°C ambiante). L'échantillon est analysé selon la méthode de la goutte séchée en utilisant l'acide sinapinique (SA) comme matrice concentrée à 10 mg/mL sur un instrument de type Autoflex Speed (Bruker) .....	164
Figure 68 : Schéma montrant les différents types de peptides pontés formés après digestion. Nous nous sommes focalisés sur les peptides pontés intermoléculaires. ....	166
Figure 69 : Exemple de spectre MSMS des peptides pontés montrant la corrélation entre le score et la qualité d'identification. Le spectre du haut correspond au peptide intermoléculaire RXR381-RAR118 .....	168
Figure 70 : Relation non linéaire entre le score attribué par pLink v2.5 et la confiance en l'interprétation. Représentation par pLabel des spectres de fragmentation de peptides pontés suite à une requête dans le logiciel pLink. Tous ces spectres sont issus d'une même analyse LC-MS/MS du complexe RAR/RXR pontés par BS3.....	169
Figure 71 : Exemple de spectres de fragmentation correspondant à des peptides pontés caractérisés par pLink avec des scores de $2 \cdot 10^{-5}$ (a) et de $1,5 \cdot 10^{-5}$ (b). Malgré un bon score d'identification, les peptides $\beta$ (spectre a) et $\alpha$ (spectre b) sont incertains .....	170
Figure 72 : Exemple de peptides pontés identifiés avec des scores pLink similaires ( $10^{-2}$ ) mais une certitude visuelle différente.....	171
Figure 73 : Résultats XL-MS obtenues après analyse de la BSA. (a) Liste des peptides pontés identifiés par pLink 2.7. En bleu sont notés les peptides identifiés après validation manuelle et en blanc les peptides qui ont été rejetés après validation. (b) Détermination des distances entre les zones pontées à partir du modèle cristallographique obtenu sur PDB. (4F5S). (c) Implémentation des données XL-MS sur la structure cristallographique. ....	173
Figure 74 : Image représentant les différentes fonctions de XiView (image extrait du site).....	175
Figure 75 : Exemple de fichiers .csv utilisé comme format d'entrée pour utiliser XiView. ....	175
Figure 76 : Première carte de proximité obtenue sur le complexe RAR-RXR.....	176
Figure 77 : Carte de proximité obtenue sur le complexe RAR-RXR après SCX. En jaune sont représentés les peptides pontés uniquement identifiés après SCX. La proximité entre RAR-DBD et RXR-LBD est maintenant illustrée par de nombreux peptides cross-linkés .....	179
Figure 78 : Schéma montrant la complémentarité des sites de pontages détectés sur un même peptide ponté. En rouge les peptides pontés entre lysine et sérine et en bleu, le peptide ponté identifié à partir des deux lysines. ....	182
Figure 79 : Spectres de fragmentation illustrant différents sites de pontage possibles à partir des mêmes séquences $\alpha$ et $\beta$ trypsiniques connectées. ....	184
Figure 80 : : Mécanisme réactionnel du pontage des protéines avec le BS(PEG) <sub>5</sub> .....	185
Figure 81 : Mécanisme réactionnel du pontage des protéines avec le LC-SMCC. ....	185
Figure 82 : Mécanisme réactionnel du DSBU ainsi que son mécanisme de fragmentation dans les cellules HCD. Adapté de [35].....	186

Figure 83 : Mécanisme réactionnel du CDI ainsi que son mécanisme de fragmentation dans les cellules HCD. Adapté de [35].....	186
Figure 84 : Carte d'interaction entre RAR et RXR. En orange sont présent les peptides pontés par LC-SMCC et en noir ceux pontés par le BS3. ....	187
Figure 85 : Carte de proximité montrant les peptides pontés observés par le BS3 et le BS(PEG) <sub>5</sub> . En vert sont les peptides uniquement observés en présence de BS(PEG) <sub>5</sub> .....	188
Figure 86 : Séquence de RAR-RXR montrant les sites dead end observés pour le CDI (en jaune) .....	189
Figure 87 : Carte de proximité des peptides pontés par le DSBU. En jaune sont représentés les peptides uniquement observés par le DSBU. En noir sont représentés les peptides observés de manière commune entre le DSBU et le BS3. ....	190
Figure 88 : Carte de proximité obtenue en considérant (a) un score arbitraire de 50 (peptides interprétés par au moins un logiciel) et (b) pour un score de 200 (les peptides ont été détectés par les 4 logiciels). ....	196
Figure 89 : Comparaison des paires inter peptidique réticulées uniques identifiées de RAR / RXR par MeroX, pLink, XiSearch et XlinkX.....	198
Figure 90 : Peptide RAR (128) -RXR (181) interprété par XiSearch. +P est la masse de l'ion précurseur.....	198
Figure 91 : représentation schématique illustrant la différence entre site de pontage (position des acides aminés pontés dans la séquence protéique) et zone de proximité (les acides aminés pontés ne sont pas identiques mais situés très proches dans l'espace tridimensionnel de la structure protéique). Les traits rouges entre 2 K représentent les cross-links, les flèches bleues représentent les hydrolyses trypsiniques.....	199
Figure 92 : Représentation schématique des RAREs de type "Direct Repeat" à gauche et de type "Inverted Repeat" à droite. R=G/A et K=G/T .....	201
Figure 93 : Séquence de RAREs utilisés dans cette étude. Les deux sites hexanucléotidique sont notés en rouge. Les nucléotides présents entre les deux demi-sites sont notés en minuscule. ....	202
Figure 94 : Protocole d'analyse XLMS pour l'identification des zones réarrangé en présence d'ADN.....	203
Figure 95 : Carte de proximité des différents complexes étudiés en présence d'ADN DR5, d'ADN DR1 et sans ADN. En jaune sont les peptides uniquement observés en présence d'ADN. ....	205
Figure 96 : Structures du complexe RAR/RXR en présence d'ADN. (a) structure cristallographique de RAR/RXR avec DR1. (b) Structure SAXS de RAR/RXR avec DR5. ....	207
Figure 97 : Structure chimique de l'agent pontant de la famille des esters NHS. (A) R=H, structure du disuccinimidylsuberate (DSS) ; R=SO <sub>3</sub> <sup>-</sup> , structure du bis(sulfosuccinimidyl)suberate (BS3). (B) formation d'une liaison pontée entre deux lysines et l'agent pontant.[16].....	208
Figure 98 : Distribution des distances entre deux acides aminés pontés en fonction de la structure. ....	210
Figure 99 : Déconvolution de la bande Amide I (1700 - 1600 cm <sup>-1</sup> ) obtenu par spectroscopie infrarouge montrant la structure secondaire du complexe RAR / RXR (a) et du complexe RAR / RXR réticulé par BS3 (b) Aucune évolution flagrante de la structure secondaire en présence de BS3. ....	212
Figure 100 : Gel natif 4-15% avec différents dépôts de RAR/RXR sans ADN et avec ADN DR5 et DR0. (a) Gel natif obtenu lors d'une préparation antérieure. (b) gel natif obtenu lors de la préparation de nos échantillons. ....	214
Figure 101 : Protocole expérimental SILAC .....	216

<i>Figure 102 : Exemple d'agent pontant marqué isotopiquement et disponible commercialement. ....</i>	<i>217</i>
<i>Figure 103 : Protocole expérimental de QXL-MS. Extrait de Chen et al 2015. ....</i>	<i>219</i>
<i>Figure 104 : Exemple de marquage isotopique réalisé pour l'analyse quantitative. ....</i>	<i>221</i>
<i>Figure 105 : Schéma explicative sur les principaux modes d'acquisition MSMS utilisé en protéomique. Extrait de Aebersold 2018.[84] .....</i>	<i>222</i>
<i>Figure 106 : Schéma montrant la distribution des valeurs des rapports de signaux pour le marquage inverse montrant la liste des peptides unique détectée pour chaque conformation. ....</i>	<i>226</i>
<i>Figure 107 : Modèle structural proposé par cryo-EM du complexe formé par l'homodimère ER<math>\alpha</math> (constitué des domaines AF1, LBD et DBD et fixé sur l'ADN), le coactivateur SRC-3 et p300 d'après[28] .....</i>	<i>235</i>
<i>Figure 108 : Représentation des différents domaines de CBP. Extrait de [26] .....</i>	<i>235</i>
<i>Figure 109 : Structure cristalline du noyau catalytique de CBP souris. Les domaines BRD, PHD, HAT et ZZ sont représentés respectivement en bleu, vert, rouge et cyan. Les atomes de zinc sont représentés par des sphères orange. [37].....</i>	<i>236</i>
<i>Figure 110 : Représentation schématique de l'ambiguïté de l'identification des peptides pontés intramoléculaire et intermoléculaire. ....</i>	<i>239</i>
<i>Figure 111 : Analyse MALDI-MS de la protéine CBP (a) avant pontage (b) après pontage avec le BS(PEG)<sub>5</sub>. (c) après pontage avec le BS3. ....</i>	<i>240</i>
<i>Figure 112 : Comparaison des cartes de pontage générées par XiView lorsque le cross-link est réalisé par le BS(PEG)<sub>5</sub> (violet) et le BS3 (rouge). La schématisation du haut montre que des peptides uniques différents représentent le pontage BS3 et BS(PEG)<sub>5</sub> mais que les mêmes acides aminés sont accessibles dans les deux expériences. La schématisation du bas montre la complémentarité des peptides pontés entre les acides aminés 1200 et 2442 de CBP. ....</i>	<i>242</i>
<i>Figure 113 : Exemples de spectres MSMS dont l'interprétation n'a été réalisée que sur un logiciel. ....</i>	<i>243</i>
<i>Figure 114 : Cartes de proximité de la protéine CBP pontée par le BS(PEG)<sub>5</sub> en fonction du score arbitraire. (a) Carte de proximité obtenue pour un score de 200 (b) pour un score de 150 (c) pour un score de 100 et (d) pour un score de 50. ....</i>	<i>244</i>
<i>Figure 115 : Cartes de proximité de la protéine CBP pontée par le BS3 observé en fonction du score. (a) Carte de proximité obtenue pour un score de 200 (b) pour un score de 150 (c) pour un score de 100 et (d) pour un score de 50. ....</i>	<i>245</i>
<i>Figure 116 : Superposition de la structure du cœur catalytique de CBP et de P300 de type souris montrant la légère différence de structure du cœur catalytique entre les deux protéines. Extrait de [37] .....</i>	<i>248</i>
<i>Figure 117 : Implémentation des données XL-MS sur la structure cristallographique du cœur catalytique de CBP de type souris (PDB : 4U7G). (a) liste des peptides pontés identifiés sur la structure. (b) Implantation des pontages chimiques sur la structure du cœur catalytique de CBP de type souris. ....</i>	<i>249</i>
<i>Figure 118 : Carte de proximité de la protéine CBP obtenue après digestion par la chymotrypsine. (a) après pontage par le BS(PEG)<sub>5</sub>. (b) après pontage avec le BS3.....</i>	<i>251</i>
<i>Figure 119 : Représentation schématique du complexe NuA4. Extrait de Wang et al 2018.[17] Le sous-domaine Piccolo est représenté en vert. ....</i>	<i>258</i>

Figure 120 : Carte de proximité (xiNET) du complexe NuA4 cross-linké par 5mM BS3 en fonction du score arbitraire. L'épaisseur du trait entre les sous-unités est proportionnelle au nombre de peptides pontés entre les sous-unités. Carte de proximité obtenue pour un score de 50 (a) ; score de 100 (b) ; score de 150 (c) ; score de 200 (d)..... 266

Figure 121 : Visualisation du réseau de pontage entre la K448 de EPL1 et différentes sous-unités de NuA4. ... 267

Figure 122 : Implémentation des cross-links intramoléculaires sur la structure cryo-EM de Tra1. En vert sont les distances atteintes lors du pontage par le BS3 comprises entre 10 et 30Å. En jaune sont les peptides observés avec une distance supérieure de 5Å. En violet sont les distances aberrantes (>30Å). ..... 268

Figure 123 : Mise en évidence d'un comportement conformationnel anormal de NuA4 après pontage par 3mM de Bs3. L'intégralité du complexe est visualisée lors d'un cross-link au glutaraldéhyde par grafix.[30] ..... 268

Figure 124 : zoom de la structure cryo-EM de NuA4 sur l'ancrage de Eaf1 et Swc4 sur Tra1. La K241 de Eaf1 est impliquée dans des liaisons hydrogènes entre les trois sous-unités. .... 269

Figure 125 : (a) Spectres MALDI-MS du complexe avant et après pontage chimique avec le BS3 à différentes conditions réactionnel (1) avant pontage (2) après 30min à Tamb à un ratio protéine/BS3 (1:50) (3) après 1h00 à tamb à un ratio protéine/BS3 (1:50) (4) après 1h00 à tamb à un ratio protéine/BS3 (1:250). La figure b représente l'analyse SDS-PAGE sur un gel d'électrophorèse à 10% acrylamide du complexe RAR-RXR purifié avant et après la réaction de pontage. La stœchiométrie du complexe ainsi que la réaction de pontage ont été cinétiquement suivies par MALDI-MS et SDS-PAGE (figure a). Ainsi, plusieurs ions du complexe ont été détectés avant pontage à  $44509 \pm 70$  Da et  $37300 \pm 70$  Da par gel et par MALDI-MS. Ces ions correspondent respectivement au composé RAR et RXR qui constituent le complexe. Après pontage, un poids moléculaire de  $84994 \pm 200$  Da a été détecté. Aucun dimère ou tétramère du complexe réticulé n'a été trouvé par MALDI-MS. Cette masse correspond à la formation hétérodimérique du complexe RAR/RXR ponté selon une stœchiométrie 1 :1. Le signal correspondant au complexe ponté complet est grand et très intense, montrant ainsi une bonne efficacité de réaction. .... 287

Figure 126 : Spectre obtenu sur un MALDI-MS équipé d'un détecteur CovalX® de RAR/RXR10µM avant pontage (a) et après réaction de pontage réalisé pendant 30min à Tamb (b)..... 288



## Liste des tableaux

---

Tableau 1 : Quelques caractéristiques des principaux analyseurs fournis par les constructeurs [41].	33
Tableau 2 : Position de la bande amide I en fonction de la structure secondaire.	66
Tableau 3 : Tableau récapitulatif listant les intérêts et les limites des différentes techniques structurales pour l'étude de complexe biologique.	71
Tableau 4 : Liste des techniques pouvant être utilisées pour déterminer différents types d'informations structurales.	72
Tableau 5 : Liste des logiciels disponibles publiquement pour l'identification, la quantification, la visualisation et / ou la modélisation des zones pontées.	101
Tableau 6 : Liste de l'ensemble des peptides pontés détectés par pLink. En jaune sont identifiés les peptides intermoléculaires.	167
Tableau 7 : Liste des peptides pontés intermoléculaires détectés par pLink. En bleu sont notés les peptides identifiés après validation manuelle. En blanc sont notés les peptides non validés après validation manuelle.	174
Tableau 8 : Tableau de comparaison des peptides pontés identifié avant et après enrichissement par SCX. Les peptides en jaune représentent les peptides pontés nouvellement identifiés après enrichissement par cartouche SCX.	180
Tableau 9 : Comparaison des méthodes d'acquisition MS appliquées à un échantillon cross-linké de RAR/RXR par du BS3.	181
Tableau 10 : liste des peptides pontés intermoléculaires obtenus par plusieurs agents pontant interprétés par pLink et MeroX après validation manuelle. Les peptides en jaune sont les peptides qui ont été observés avec plusieurs agents pontants.	192
Tableau 11 : Comparaison des peptides cross-linkés identifiés par 4 logiciels. En jaune sont représentés les peptides identifiés par au moins 2 logiciels. Entre parenthèses est noté le nombre de spectres MSMS qui ont conduit à la même identification.	197
Tableau 12 : Comparaison des peptides pontés identifiés pour des échantillons en présence et en absence d'ADN. En jaune sont les peptides pontés observés uniquement sur le complexe en présence d'ADN.	206
Tableau 13 : Tableaux regroupant les distances des peptides pontés identifiés à partir de différents modèles. En vert sont notées les distances compatibles avec à la longueur d'extension maximale du BS3.	209
Tableau 14 : Pourcentages des éléments de la structure secondaire de RAR/RXR avant et après pontage par le BS3 déduits de la figure 99. La précision des pourcentages est estimée à $\pm 3.5\%$ .	212
Tableau 15 : Liste des peptides pontés quantifiés dans l'ensemble des échantillons et observés comme singulet (RAR/RXR+ADN unique et RAR/RXR unique) et comme doublet (RAR/RXR+ADN préférentielle et RAR/RXR préférentielle).	225
Tableau 16 : Liste des peptides pontés identifiés entre les sous-unités de CBP pour l'échantillon ponté par le BS3.	246

<i>Tableau 17 : Liste des peptides pontés identifiés entre les sous-unités de CBP pour l'échantillon ponté par le BS(PEG)<sub>5</sub>.</i>	247
<i>Tableau 18 : Listes des peptides pontés obtenus avec le BS(PEG)<sub>5</sub> après digestion chymotrypsique. En jaune sont les peptides observés par plusieurs logiciels.</i>	250
<i>Tableau 19 : Listes des peptides pontés obtenus par le BS3 après digestion chymotrypsique.</i>	250
<i>Tableau 20 : Liste de l'ensemble des protéines présentes dans l'échantillon de NuA4.</i>	259
<i>Tableau 21 : Nombre de peptides pontés identifiés par pLink (avant validation manuelle) en fonction de la fraction d'élution après SCX.</i>	261
<i>Tableau 22 : Liste de l'ensemble des peptides pontés identifiés sur NuA4 après pontage par du BS3 à 5mM et à 3mM. En orange apparaissent les peptides pontés identifiés par 4 logiciels, en bleu par 3 logiciels et en gris par 2 logiciels.</i>	262

# Introduction générale

Plus la recherche scientifique avance, plus on découvre à quel point notre corps est un système complexe qu'il nous faut appréhender de manière holistique. Cette observation s'applique également à l'échelle moléculaire où la compréhension des mécanismes qui régissent la communication entre les différents acteurs de la cellule passe de plus en plus largement par des approches intégratives. Notre projet se situe dans un contexte d'analyse en biologie structurale. En effet, les protéines sont les véritables acteurs de la cellule, elles se replient dans l'espace et adoptent une structure tridimensionnelle native, qui leur confère leurs propriétés. Elles sont la plupart du temps intégrées dans des édifices macromoléculaires impliqués dans les processus biologiques. C'est pourquoi il est crucial d'obtenir des informations structurales afin de mieux visualiser et comprendre les mécanismes qui génèrent et contrôlent l'activité de ces complexes biologiques non-covalents et ainsi pouvoir émettre des hypothèses diagnostiques et thérapeutiques lors de leur dysfonctionnement. Pendant longtemps, on a cru qu'une séquence d'acides aminés donnait une structure native unique repliée qui conduisait à une ou plusieurs fonctions biologiques. Aujourd'hui on sait que la structure d'une protéine est une donnée dynamique, sensible à de multiples facteurs, pouvant subir d'importants changements conformationnels selon leur environnement. D'autre part, certaines protéines sont intrinsèquement non structurées, mais peuvent être fonctionnelles tandis que d'autres nécessitent l'aide de chaperonnes. Un changement conformationnel, ou encore la dégradation de la protéine ou de l'un de ses partenaires peut présenter la cause ou le catalyseur de certaines maladies.

L'objectif de mon travail de thèse est de développer une méthodologie basée sur l'analyse par Spectrométrie de masse (MS) qui, une fois couplée à des approches mécanistiques et biochimiques, pourrait intégrer le panel des techniques d'analyse afin de nous guider vers la génération d'un modèle structural le plus proche possible de l'état physiologique *in vivo*. La détermination de la structure des macromolécules biologiques seules, ou intégrées dans des complexes hiérarchisés porteurs du message biologique, s'avère aujourd'hui une étape fondamentale, car elle nous permettra de mieux appréhender les mécanismes très subtils qui régulent le fonctionnement des biomolécules.

De nombreuses approches d'analyse en biologie structurale existent, combinant parfois des techniques purement empiriques à des études plus théoriques par modélisation moléculaire. Parmi elles, la diffraction aux rayons-X (DRX), la résonance magnétique nucléaire (RMN) et la microscopie électronique (EM) sont souvent utilisées pour obtenir des informations structurales de meilleure résolution. Les techniques spectroscopiques telles que la Spectroscopie Infrarouge, la spectroscopie Raman, la spectroscopie UV-Vis, le Dichroïsme circulaire (CD), ainsi que la Fluoroscopie apportent des informations à plus faible résolution, mais qui, combinées entre elles moduleront, sublimeront ou réfuteront les hypothèses structurales proposées jusqu'alors. Parmi toutes ces techniques, certaines ont connu des avancées importantes en termes de performances technique et méthodologique, c'est le cas notamment de la EM, de l'IR, mais aussi de la MS.

La MS est une technique analytique qui est apparue au XX<sup>ème</sup> siècle. Sir JJ Thomson fut le fondateur en découvrant l'électron en 1897. Au début du XX<sup>ème</sup> siècle, il construisit le premier spectromètre de masse afin de pouvoir déterminer les ratios de masse sur charge d'ions.<sup>[1]</sup> Les ions étaient générés dans des tubes à décharge et passaient à travers des champs électriques et magnétiques. Ces ions étaient alors détectés sur un écran fluorescent ou sur une plaque photographique. Aujourd'hui, la technique a fortement évolué, permettant d'analyser des macromolécules de plusieurs millions de daltons ! Des avancées technologiques, méthodologiques, informatiques génèrent chaque année des instruments plus précis, rapides, sensibles qui permettent de repousser les limitations intrinsèques et nous ouvrent un champ nouveau d'expérimentation. Alors que des échantillons aussi variés que des poussières stellaires, un antibiotique, un hydrolysate cellulaire ou un virus peuvent être analysés sur des instruments aussi variés qu'un IE-Q-MS (Impact Electronique-Quadripôle), un MALDI-TOF-MS (Matrix Assisted Laser Désorption Ionisation-Time of Flight) ou encore un système nanoLC-ESI-Orbitrap MS (nano Chromatographie Liquide-Electrospray-Orbitrap), deux paramètres nous semblent essentiels et communs pour toute expérience : (i) nous analysons des ions en phase gazeuse, (ii) nous analysons des ions suite à une préparation d'échantillon ciblée. Ces deux observations sont particulièrement importantes dans le contexte de notre projet d'analyse structurale puisqu'il sera nécessaire de trouver des conditions expérimentales et instrumentales compatibles à la fois avec la technique, la problématique et le système biologique d'intérêt. C'est en 1991<sup>[2]</sup>, avec la détection du premier complexe biologique non covalent, que la MS fut considérée comme une approche à fort potentiel pour étudier des édifices non-covalents non déstructurés.

Grâce au couplage avec la source électrospray (ESI), la MS a très vite montré de grandes capacités et reste aujourd'hui encore une technique de choix pour démontrer l'existence d'un complexe non covalent et déterminer sa stœchiométrie, ou encore analyser la dynamique des interactions.<sup>[3-6]</sup> Grâce à la MS, il a ainsi été possible de caractériser des complexes biologiques tels que des virus <sup>[7-9]</sup>, des protéines membranaires<sup>[10]</sup> ou encore des complexes ribosomiques.<sup>[11]</sup>

Cependant, cette approche est affectée par deux limitations qui, jusqu'à aujourd'hui, n'ont pas été dénouées : (i) une préparation de l'échantillon très pointue, notamment le transfert intact des complexes protéiques dans les solvants compatibles avec l'ESI-MS (ii) la déstabilisation des interactions hydrophobe liée à l'analyse des ions en phase gazeuse, ce qui nous incite à développer de nouveaux protocoles d'analyse pour permettre leur étude structurale par MS : marquage chimique<sup>[12]</sup>, échange proton/deutérium<sup>[13]</sup>, pontage chimique<sup>[14]</sup>.

C'est dans ce contexte que débutèrent mes travaux de thèse dont l'objectif est d'approfondir le développement de la MS pour la caractérisation structurale de complexe biologique par pontage chimique (XL) et par MS native (Native-MS).

Le pontage chimique aussi appelé « cross-linking »<sup>[14]</sup> est mis en place en amont de l'analyse par MS (XL-MS). Cette stratégie consiste à lier de façon covalente les sous-unités qui sont en interactions non covalentes ou qui sont proches dans l'espace par l'intermédiaire d'un agent de pontage connu également sous le nom de « cross-linker ». Ce pontage peut être effectué pour relier des protéines, des peptides, des acides nucléiques..., au sein d'une unique protéine ou entre plusieurs partenaires au sein d'un complexe biologique. En identifiant les sites réticulés, on espère positionner les acides aminés proches dans l'espace et accéder ainsi à des informations structurales à basse résolution qui pourront être confrontées aux données obtenues par différentes techniques d'analyse expérimentales (telles que la EM) ou théoriques (telles que la modélisation moléculaire). De nombreuses études structurales de complexes biologiques ont déjà été réalisées grâce à l'utilisation de cette stratégie. Toutefois, ni la mise en place du protocole ni l'interprétation des données ne sont aujourd'hui standardisées et beaucoup de difficultés restent d'actualité, telles qu'un faible rendement de réaction de pontage ou encore un manque d'automatisation dans l'identification des peptides pontés. Dans ce contexte, une optimisation plus poussée est nécessaire afin de mieux définir les limitations ou la résolution de cette approche.

La première partie de ce manuscrit sera consacrée à une approche bibliographique où nous développerons dans un premier chapitre, un état de l'art de la MS et son implication dans l'étude d'échantillons biologiques. Dans un second chapitre, nous traiterons plus en détail les différentes stratégies d'analyses protéomiques basée sur la MS. Enfin, dans un troisième chapitre, nous nous intéresserons à la caractérisation structurale des complexes biologiques non-covalents par MS en montrant plus en détail les différentes approches qui lui sont associées.

Dans une seconde partie, je présenterai les résultats obtenus au cours de ma thèse illustrant le potentiel de la MS dans un contexte de caractérisation structurale de complexes biologiques. Le quatrième chapitre de cette thèse démontrera les possibilités de la MS native pour l'étude de l'agrégation et de l'amyloïdogénicité d'un coactivateur nucléaire. Dans un cinquième chapitre, et ceci a la majeure partie de mon travail de thèse, nous nous intéresserons au développement de la XL-MS. Nous montrerons en particulier le principe de l'approche puis les optimisations qui ont été effectuées sur un complexe modèle faisant partie de la famille des récepteurs nucléaires. Puis, cette approche optimisée a été utilisée afin de caractériser deux autres complexes biologiques. Dans un sixième chapitre, l'approche a été utilisée afin de caractériser une protéine partiellement désordonnée de la famille des coactivateurs nucléaires. Dans un septième et dernier chapitre, nous montrerons comment l'approche XL-MS peut être utilisée en combinaison avec la cryo-EM afin d'aider dans la génération d'un modèle structural à haute résolution.

A l'aide de ces trois exemples, nous tenterons de faire ressortir les avantages et limitations de cette stratégie. Nous détaillerons également comment l'utilisation d'outils bio-informatiques dédiés s'est avérée cruciale pour une interprétation pertinente des données de XL-MS.

## Bibliographie

1. Griffiths, J., *A brief history of mass spectrometry*. Analytical Chemistry, 2008. **80**(15): p. 5678-5683.
2. Katta, V. and B.T. Chait, *Observation of the Heme Globin Complex in Native Myoglobin by Electrospray-Ionization Mass-Spectrometry*. Journal of the American Chemical Society, 1991. **113**(22): p. 8534-8535.
3. Hilton, G.R. and J.L.P. Benesch, *Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry*. Journal of the Royal Society Interface, 2012. **9**(70): p. 801-816.
4. Yugandhar, K., S. Gupta, and H. Yu, *Inferring Protein-Protein Interaction Networks From Mass Spectrometry-Based Proteomic Approaches: A Mini-Review*. Computational and structural biotechnology journal, 2019. **17**: p. 805-811.
5. Erba, E.B., L. Signor, and C. Petosa, *Exploring the structure and dynamics of macromolecular complexes by native mass spectrometry*. Journal of Proteomics, 2020. **222**.
6. Ishii, K., M. Zhou, and S. Uchiyama, *Native mass spectrometry for understanding dynamic protein complex*. Biochimica Et Biophysica Acta-General Subjects, 2018. **1862**(2): p. 275-286.
7. Arold, S., et al., *Characterization and molecular basis of the oligomeric structure of HIV-1 Nef protein*. Protein Science, 2000. **9**(6): p. 1137-1148.
8. Guttman, M. and K.K. Lee, *Isotope Labeling of Biomolecules: Structural Analysis of Viruses by HDX-MS*. Methods in enzymology, 2016. **566**: p. 405-426.
9. Snijder, J., et al., *Studying 18 MDa Virus Assemblies with Native Mass Spectrometry*. Angewandte Chemie International Edition, 2013. **52**(14): p. 4020-4023.
10. Bolla, J.R., et al., *Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry*. Annual Review of Biochemistry, 2019. **88**(1): p. 85-111.
11. Rostom, A.A., et al., *Detection and selective dissociation of intact ribosomes in a mass spectrometer*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(10): p. 5185-5190.
12. Huang, T., et al., *Chemical Tagging in Mass Spectrometry for Systems Biology*. Analytical Chemistry, 2019. **91**(1): p. 109-125.
13. Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments*. Nature Methods, 2019. **16**(7): p. 595-602.
14. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes*. Journal of Structural Biology, 2011. **173**: p. 530-540.

# **Partie 1 : Introduction Bibliographique**

Dans cette première partie, nous avons souhaité présenter les méthodologies, basées sur la MS, auxquelles nous avons fait appel au cours de ces travaux de thèses. Il s'agira de fournir une revue bibliographique définissant ainsi les bases théoriques, mais également de dresser un état de l'art à mon arrivée au laboratoire.

Cette partie bibliographique se divise en 3 chapitres :

**Chapitre I** : La MS et son implication dans l'étude des complexes biologiques.

**Chapitre II** : Stratégies d'analyse protéomique par MS

**Chapitre III** : MS et caractérisation structurale de complexes biologiques non-covalents



# Chapitre I : La MS et son implication dans l'étude des complexes biologiques

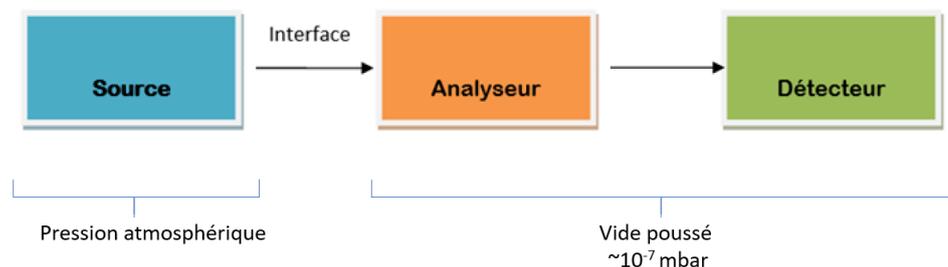
## 1. Introduction

La MS est une technique analytique permettant d'identifier et de quantifier une molécule ou un ensemble de molécules composant un échantillon en déterminant leur rapport masse/charge ( $m/z$ ). Cette technique opère en phase gazeuse où les espèces seront ionisées et séparées selon leur rapport  $m/z$ .

Cette technique a également la possibilité d'analyser une large variété de molécules (échantillons organiques ou inorganiques, depuis des poussières stellaires jusqu'aux macromolécules biologiques). Des spectromètres de masse sont désormais installés dans des aéroports, des hôpitaux, des navettes spatiales, des musées voire en version portable dans des stades olympiques par exemple.

Mais quel que soit le domaine d'application, un spectromètre de masse est divisé en quatre parties :

- La source : qui va permettre d'ioniser et de nébuliser les composés.
- L'interface : qui va assurer la transmission des ions vers les différentes composantes du spectromètre de masse.
- L'analyseur : qui va séparer les composés en fonction de leur  $m/z$ .
- Le détecteur : qui va détecter et compter les ions générant un signal qui, après intégration, aboutira au spectre de masse.



**Figure 1 : Schéma représentant les différentes composantes d'un spectromètre de masse**

Pour chaque partie, il existe différentes sources, analyseurs et détecteurs. Ce schéma général (figure 1) peut ainsi être décliné de beaucoup de manières en fonction de la problématique donnée puisque différents types de sources peuvent être couplés à différents types d'analyseurs. Chacun possède ses propres limitations et ses propres caractéristiques qui devront être choisies en fonction de l'analyte que l'on souhaite caractériser. Dans ce chapitre, nous décrirons plus en détail les sources ESI (Electrospray) et MALDI (Matrix Assisted Laser Desorption Ionisation) ainsi que les analyseurs de type TOF (Time-Of-Flight), Q (Quadripôle) ou Orbitrap qui constituaient les instruments utilisés durant mes travaux de thèse.

Les performances d'un spectromètre de masse sont largement liées aux performances de l'analyseur (gamme de rapport  $m/z$ , résolution, vitesse de balayage, précision de mesure de masse, mode de fragmentation...). Les avantages certains de la MS sont sa sensibilité (jusqu'à quelques attomoles), sa capacité à détecter individuellement toutes les espèces coexistant en solution (contrairement à d'autres techniques analytiques qui rendent un signal global de l'échantillon), ainsi que sa capacité à pouvoir se coupler à des techniques séparatives en amont (chromatographie gazeuse ou liquide, électrophorèse capillaire).

## 2. Les sources d'ionisation

### 2.1 L'ionisation MALDI

La spectrométrie de masse à source d'ionisation par désorption-ionisation laser assistée par matrice (MALDI) a été développée par les Pr. Michael Karas et Frank Hillenkamp d'une part et le Pr. Koichi Tanaka d'autre part en 1984.<sup>[1]</sup>

Cette technique peut être appliquée à une grande variété de molécules et permet de travailler sur une large gamme de masse (0.5kDa à plusieurs centaines de kDa), avec une préparation d'échantillon relativement simple, une grande rapidité et une tolérance vis-à-vis de nombreux sels ou tensioactifs intéressants dans un contexte d'analyse de molécules biologiques.

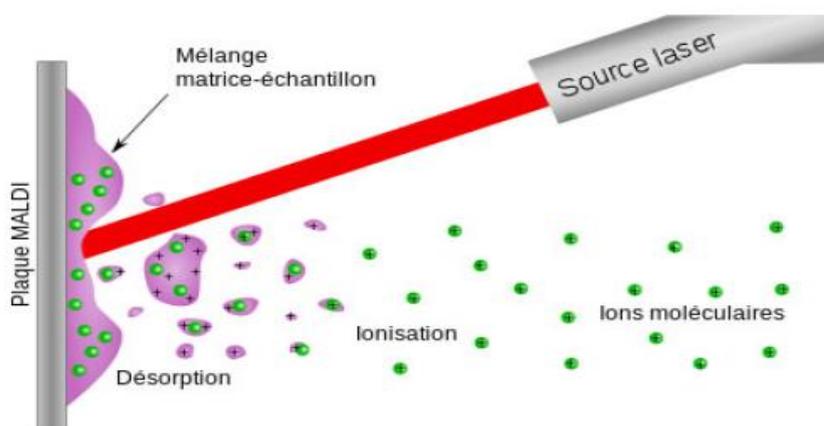


Figure 2 : Représentation de l'ionisation d'un échantillon en MALDI-MS extrait de [2]

D'un point de vue macroscopique, la technique MALDI utilise un faisceau laser UV pulsé (bien que certains instruments utilisent des lasers IR pour augmenter la désorption des molécules<sup>[3, 4]</sup>) pour désorber un échantillon cristallisé dans une solution concentrée de matrice sur une cible en acier inoxydable (matériau conducteur et chimiquement inerte). L'énergie transmise par l'irradiation laser est absorbée par la matrice, ce qui provoque son expansion en phase gazeuse en entraînant les molécules d'analytes co-cristallisées. Le choix de la matrice est très important pour le succès d'une analyse MALDI, tant dans sa nature que dans le ratio matrice/échantillon. Outre son rôle dans l'absorption de l'énergie du faisceau laser, elle participe également à l'ionisation des molécules d'intérêt en induisant des transferts de protons soit avant désorption soit après désorption avec d'autres molécules du plasma.

### **2.1.1 Les différents modes de dépôts**

La qualité des spectres obtenus par MALDI est largement dépendante d'une bonne préparation de la cible d'échantillons. En particulier, la matrice doit posséder certaines propriétés pour être sélectionnée pour une analyse MALDI : elle doit bien sûr absorber à la longueur du laser, mais doit également former des cristaux incluant l'analyte qui résistent au vide poussé de la source et être soluble dans un solvant compatible avec l'analyte<sup>[5]</sup>. Il n'y a pas vraiment de règle permettant de choisir la matrice en fonction de notre application et la matrice la mieux appropriée (seule ou en mélange) est souvent issue d'observations empiriques. Toutefois, on peut noter que certaines matrices sont privilégiées pour certaines applications. Tandis que l'acide  $\alpha$ -Cyano-4-hydroxycinnamic (HCCA) et l'acide sinapinique (SA) sont les matrices les plus souvent choisies pour les peptides et les protéines respectivement, l'acide picolinique (PA) et l'acide 3-hydroxypicolinique (3-HPA) sont plus utilisés pour les acides nucléiques.<sup>[6]</sup>

Actuellement, il existe plusieurs modes de dépôt de l'échantillon sur la plaque métallique<sup>[7, 8]</sup>.

- Des dépôts « gouttes séchées » : l'échantillon et la matrice sont mélangés (soit dans un tube, soit directement sur cible) et déposés sur la plaque métallique. Une fois le dépôt effectué, l'ensemble est laissé évaporé à température ambiante pour être cristallisé. La cristallisation lente permet de générer des cristaux enrichis en analyte.
- Des dépôts « couche mince » : la matrice est tout d'abord déposée sur la cible en un film fin puis, dans un second temps, l'analyte est appliqué par-dessus cette couche de matrice et laissé évaporé à l'air ambiant.
- Des dépôts « sandwich » : l'échantillon est déposé sur une couche mince de matrice puis recouvert d'une seconde couche de matrice.

La différence entre ces types de dépôt réside principalement dans le fait qu'avec les dépôts « couche mince » ou « sandwich » les cristaux formés sont beaucoup plus homogènes que ceux obtenus par « gouttes séchées ». Ainsi, plus le dépôt sera régulier, moins il y aura d'irrégularités de surface modifiant la distance entre la cible et la lentille d'extraction et donc meilleure sera la résolution.<sup>[7]</sup>

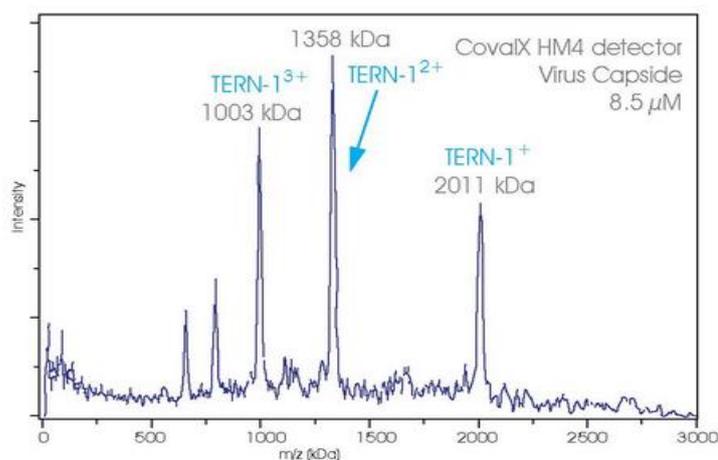
Une fois le dépôt réalisé, une irradiation laser y est réalisée sous un vide poussé (environ  $10^{-7}$  mbar). Cette irradiation permet ainsi de générer des ions en phase gazeuse qui seront ensuite guidés vers l'analyseur et le détecteur grâce à un champ électrique.

Aujourd'hui encore, le mécanisme d'ionisation en spectrométrie MALDI n'est toujours pas très bien compris, particulièrement à cause de différents paramètres physiques et chimiques tels que le pH, la température, le type de matrice utilisée, les solvants... Une des hypothèses introduites par Zenobi et Knochenmuss<sup>[9]</sup> impliquerait deux étapes : une première étape où l'on observerait une ionisation primaire des molécules de matrice se produisant au moment de l'impact laser, générant ainsi des ions primaires par photo-ionisation<sup>[10]</sup> rapidement suivie par la formation d'autres ions cluster de matrice et d'échantillon qui seraient éjectés sous forme de plumeau.<sup>[11-13]</sup> Une fois cette ionisation primaire réalisée, des réactions de transfert de protons et de cations, ou des réactions de transfert d'électrons et d'anions<sup>[14]</sup> induisant la formation d'analytes ionisés. Les analytes ainsi formés seraient donc majoritairement des ions monochargés.

Les principaux analyseurs couplés avec la spectrométrie de masse MALDI sont l'analyseur à résonance ionique cyclotronique à transformée de Fourier<sup>[15]</sup>, l'analyseur piège à ions<sup>[16]</sup> ou l'analyseur à temps de vol,<sup>[17]</sup> ce dernier étant particulièrement bien adapté à la nature pulsée du MALDI ainsi qu'à la production résultante d'ions monochargés de rapports m/z parfois très élevés.

Les avantages certains du MALDI sont :

- Sa rapidité d'analyse : typiquement les lasers de fréquence 200 Hz permettent d'obtenir l'acquisition d'un spectre de masse en quelques secondes.
- La facilité d'interprétation des spectres : les ions détectés sont majoritairement monochargés (même si un petit pourcentage d'ions multichargés est susceptible d'être observé notamment dans le cas de macromolécules).
  - Sa sensibilité : quelques attomoles peuvent suffire pour mesurer la masse.
  - La possibilité de réaliser de l'imagerie moléculaire sur des tissus : ceci permet de suivre la localisation ou les changements de localisation de molécules d'intérêt ciblées avec une résolution pouvant atteindre quelques microns selon la lentille de convergence utilisée.<sup>[18]</sup>
- En théorie, la gamme de masse mesurable n'est pas limitée<sup>[19]</sup>. Dans le cas des analytes de haut poids moléculaire (>100 KDa), la limitation du succès de l'analyse sera également liée à la nature du détecteur en place sur l'instrument. En effet, le détecteur classiquement installé est un détecteur MCP (multi-channel-plate) qui a une faible efficacité de conversion des ions lourds en électrons. La société CovalX AG<sup>®</sup> (Zurich, Suisse) a réussi à développer un système de détection haute masse en se basant sur la technologie du « ion conversion dynode » (ICD). Dans ce détecteur, les ions de haute masse sont dans un premier temps convertis par le biais d'une dynode en ions secondaires avant d'être réaccélérés et détectés par le multiplicateur.<sup>[20]</sup> Ce détecteur peut être installé sur toute configuration de spectromètre de masse MALDI, il s'ajuste entre le réflectron et le détecteur MCP préinstallé ainsi les paramètres de l'analyse sont conservés. Des ions de rapports m/z > 1 million Da ont ainsi pu être observés<sup>[21]</sup>.



**Figure 3 : Spectre de masse de capside de virus obtenu par MALDI-MS avec un détecteur HM4 de la société CovalX® permettant la détection d'espèce biologique à haut poids moléculaire (jusqu'à 2 MDa). Extrait de la documentation de CovalX®**

En revanche, l'ionisation MALDI présente certaines limitations :

- Une reproductibilité très largement dépendante des conditions de cristallisation sur cible.
- La difficulté d'analyser des ions de faibles rapports m/z liée à l'interférence avec les ions de matrice. Une des solutions serait d'utiliser des matrices nanocristallines<sup>[22]</sup> ou de rajouter des surfactants dans la matrice<sup>[23]</sup>

## **2.2 Spectrométrie de masse à source d'ionisation Electrospray (ESI)**

### **2.2.1 Principe**

La spectrométrie de masse à source électrospray (ESI) a été développée par Fenn et ses collaborateurs et décrite en 1985.<sup>[24]</sup> Il s'agit d'une méthode d'ionisation douce, permettant ainsi la désorption de molécules sans fragmentation. Elle est depuis largement utilisée dans des domaines d'application très variés, mais fut une véritable révolution dans le domaine analytique des macromolécules biologiques grâce notamment à sa capacité à générer des ions multichargés.<sup>[25]</sup>

D'un point de vue macroscopique, l'ionisation ESI consiste en la production d'ions en phase gazeuse à partir d'une solution d'analytes en mouvement soumise à un champ électrique. Le processus d'ionisation en ESI-MS s'effectue à pression atmosphérique et comporte plusieurs étapes dont certaines sont encore sujettes à débat <sup>[26]</sup>. Sous l'action du champ électrique, un spray contenant des petites gouttelettes d'analytes et de solvant est généré. Le solvant est alors évaporé sous l'action de la chaleur (utilisation d'un gaz inerte chauffé), mais également sous l'action de l'énergie provenant des nombreuses collisions avec les molécules de gaz résiduelles dans la source et l'interface du spectromètre de masse. Selon la polarité du champ électrique appliqué, des ions positivement chargés (mode positif) ou

négativement chargés (mode négatif) seront extraits de la solution d'injection. Suite à l'évaporation du solvant, la taille des gouttelettes diminue jusqu'à ce que les répulsions électrostatiques excèdent les forces de cohésion du liquide ce qui induit l'explosion des gouttelettes mères en gouttelettes plus petites. Ce phénomène est récurrent jusqu'à l'obtention d'ions totalement désolvatés ou bien, selon les théories développées, jusqu'à l'obtention d'une gouttelette de taille précise (limite de Rayleigh), pour laquelle les répulsions électrostatiques deviennent assez fortes pour provoquer la directe désorption des ions analytes<sup>[26]</sup>.

## 2.2.2 Génération d'ions multichargés

L'ionisation par ESI génère des ions possiblement multichargés. Dans le cas des biomolécules, ces ions sont obtenus par protonation ou déprotonation et sont de type  $(M+zH)^{z+}$  en mode positif ou  $(M-zH)^{z-}$  en mode négatif. Le nombre de charges observé dépend de plusieurs facteurs comme la masse de la molécule, ses propriétés chimiques (notamment de l'affinité protonique de ses différents groupes fonctionnels en phase gazeuse), de l'accessibilité des groupements ionisables (notamment liée à sa conformation), mais aussi du solvant utilisé. Ainsi, un spectre de masse obtenu par ionisation ESI sera représenté par une distribution des différents états de charges des différentes molécules présentes dans notre échantillon.

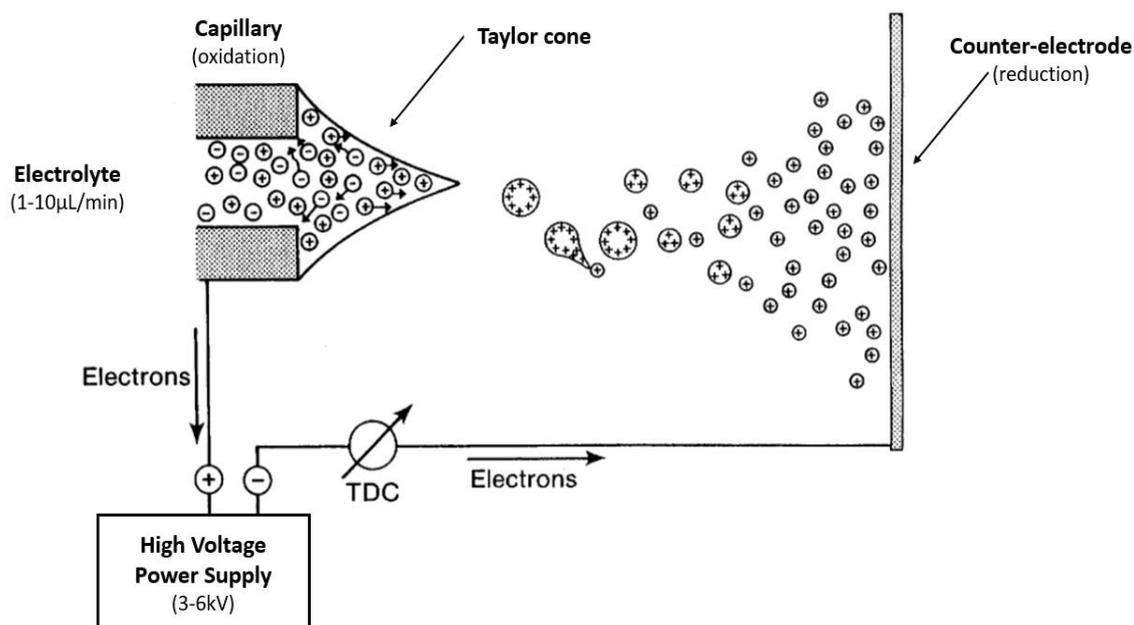


Figure 4 : Schéma de la source à ionisation électrospray (ESI) adapté de [27]

L'ionisation ESI peut être couplée à de nombreux analyseurs grâce à la génération d'ions de rapports  $m/z$  relativement faibles comparés à ceux obtenus par MALDI. Ainsi une grande flexibilité dans la configuration des spectromètres de masse à ionisation ESI est disponible et sera choisie selon notre problématique.

Les avantages certains de l'ESI sont :

- La possibilité d'analyser des molécules de natures variées liées à une grande versatilité dans le choix du solvant.
- L'analyse de molécules polaires de hauts poids moléculaires liée à la génération d'ions multichargés.
- Le couplage facile avec des techniques séparatives telles que la chromatographie liquide haute performance (HPLC) ou l'électrophorèse capillaire (CE).
- L'ouverture à l'analyse d'échantillons solides grâce au développement du DESI (désorption par ionisation électrospray). Cette source à ionisation douce a la particularité d'effectuer l'analyse des composés en phase solide ne nécessitant ainsi aucune préparation en amont de l'analyse. Elle est adaptée à de nombreux types d'échantillons allant des petites molécules à l'analyse de complexes biologiques avec un débit d'analyse rapide.<sup>[28, 29]</sup>

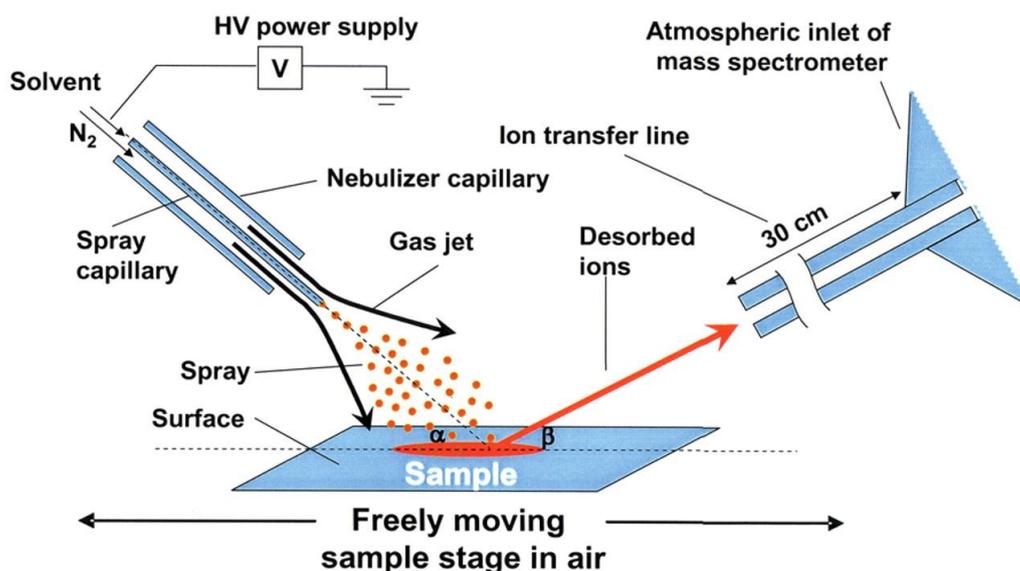


Figure 5 : Schéma représentative de la source DESI extrait de <sup>[28]</sup>.

L'inconvénient majeur de cette méthode d'ionisation est lié à une préparation de l'échantillon souvent lourde puisque qu'elle est très intolérante à la présence de sels (souvent indispensables à une bonne solubilisation des biomolécules) ou d'additifs. Les sels interagissant avec l'analyte vont former des adduits stables qui seront alors visibles sur le spectre de masse. La formation de ces adduits a pour conséquence une hétérogénéité de l'échantillon en masse, rendant ainsi l'interprétation des spectres très difficile. De plus, la présence de sels va diminuer la charge nette des analytes entraînant une diminution du signal lors de l'ionisation par ESI-MS.<sup>[30-32]</sup>

### 3. L'interface

Que ce soit avec une source MALDI ou avec une source ESI, une fois les ions produits, ils devront être guidés vers l'analyseur se trouvant sous un vide poussé. Les interfaces vont donc permettre d'acheminer ces ions vers l'analyseur tout en réduisant la pression petit à petit. Pour ce faire, l'interface est souvent séparée en plusieurs tronçons qui sont constitués de lentilles électrostatiques. Différentes géométries sont proposées par les constructeurs d'instruments, optimisées dans un souci de sensibilité ou de gamme de rapport  $m/z$  notamment. Ce bloc soumis à une moyenne pression ( $10^{-2} - 10^{-3}$  mbar), est crucial pour la refocalisation, tant spatiale qu'énergétique, du faisceau d'ions divergent en sortie de source. Ce point sera détaillé dans le contexte particulier de la caractérisation des complexes non-covalents (cf. Chapitre III).

### 4. Les analyseurs

Tous les analyseurs visent à séparer et mesurer les rapports  $m/z$  des ions en phase gazeuse, mais ils reposent sur des principes physiques différents. La demande en capacités analytiques puissantes est de plus en plus forte ce qui a motivé une évolution rapide et constante de l'instrumentation au cours de ces dernières années. De nouveaux types d'analyseurs plus rapides ou plus résolutifs font leur apparition ainsi que des instruments hybrides, fournissant ainsi de nouvelles opportunités dans l'étude des protéomes complexes. En particulier, des spectromètres de masse permettant des mesures de masse exactes sont très intéressants dans un cadre d'analyses protéomiques, non seulement pour identifier des protéines, mais aussi pour caractériser d'éventuelles modifications post-traductionnelles, ou encore pour mettre en évidence des interactions entre des protéines.

#### 4.1 L'analyseur à temps de vol

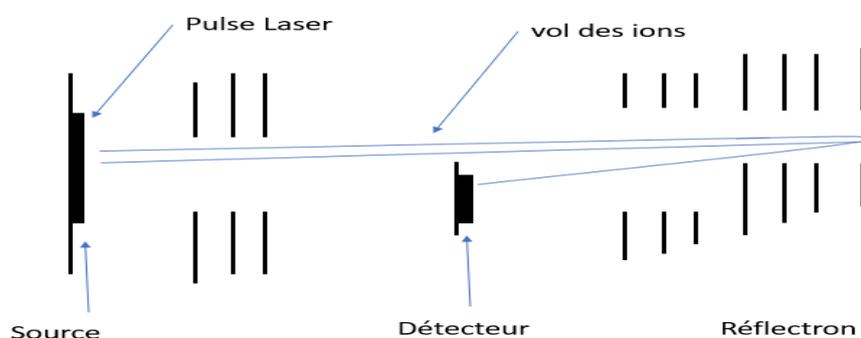


Figure 6 : Représentation schématique de la migration des ions dans un MALDI-TOF

L'analyseur à temps de vol est un analyseur de MS qui sépare les ions en fonction de leur vitesse dans un tube de vol libre de champ. Les ions sont tout d'abord accélérés par un champ électrique à l'entrée du tube de vol (chambre de pulse). Accélérés sur une même distance et par une même force, ceux-ci acquièrent la même énergie cinétique et montreront une distribution de vitesse définie dans le tube de vol (eq 1). Les ions aux plus faibles rapports  $m/z$  vont alors arriver en premier au détecteur tandis que ceux qui ont les  $m/z$  les plus grands vont prendre plus de temps à traverser l'analyseur. Intrinsèquement, la gamme de mesure de masse de l'analyseur est donc en théorie illimitée. Une fois le tube traversé, les ions arrivent au niveau d'un détecteur qui convertit les courants ioniques en courant électrique. Le détecteur le plus couramment utilisé est le détecteur MCP.

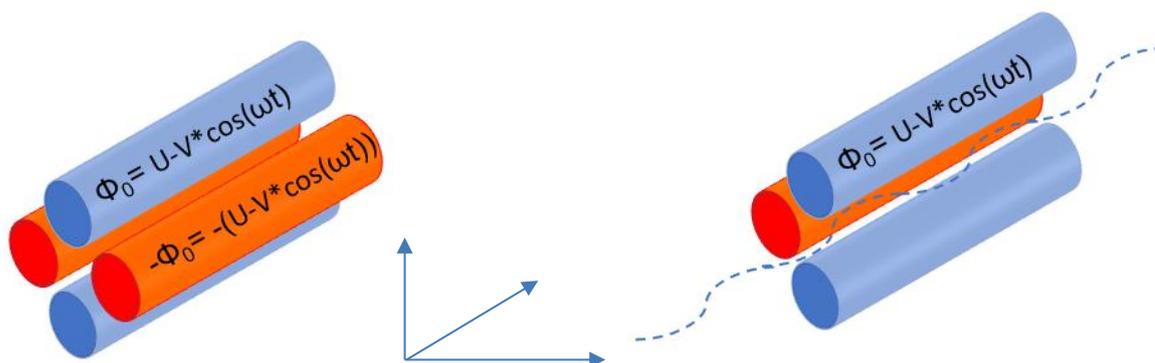
$$\text{Eq 1 : } t_{\text{vol}} \approx \sqrt{m/z}$$

Dans ce type d'analyseur, les ions sont envoyés par paquets dans le tube de vol ce qui en fait un analyseur à haute sensibilité. Son pouvoir de résolution est lié à plusieurs facteurs : la distribution d'énergie initiale, la distance d'accélération ainsi que sa durée, la distance de vol. En effet, une distribution spatiale ou d'énergie cinétique des ions à l'origine entraînera une distribution des temps d'arrivée d'ions identiques et donc une baisse de résolution.

Deux modes d'analyse sont accessibles :

- Le mode **linéaire** : les ions sont expulsés par paquets et traversent le tube de vol de manière linéaire jusqu'au détecteur MCP. Cette configuration d'analyseur permet de mesurer des rapports  $m/z$  théoriquement illimités. Cependant dans ce mode, une forte dispersion spatiale des ions et une différence d'énergie cinétique entre des ions de même  $m/z$  aboutissent souvent à une forte perte de résolution.
- Le mode **réflectron** : les ions sont envoyés par paquets et traversent le tube de vol de manière linéaire puis pénètrent dans un réflectron électrostatique qui impose un champ électrique de direction opposée au champ d'accélération. Les ions les plus énergétiques pénètrent plus profondément dans le réflectron et vont donc mettre plus de temps à en ressortir de sorte qu'ils arrivent en même temps sur le détecteur que les ions moins énergétiques de même rapport  $m/z$ . Ce réflectron permet de corriger les dispersions cinétiques des ions de même rapport  $m/z$  améliorant ainsi la résolution des pics et la précision de masse. Une limitation du mode réflectron est qu'il est optimisé sur une gamme de masse restreinte.

## 4.2 Le quadripôle (Q)



**Figure 7 : Représentation d'un quadripôle ainsi que le mouvement des ions à l'intérieur de celui-ci**

Le quadripôle est un analyseur constitué de quatre électrodes hyperboliques ou cylindriques connectées deux à deux. Les électrodes opposées auront alors le même potentiel tandis que celles qui seront adjacentes auront un potentiel de polarité opposé. Ce potentiel  $\Phi$  est composé d'une tension continue  $U$  et d'une tension alternative de radiofréquence  $V \cos \omega t$ , tel que  $\Phi = \pm (U - V \cos \omega t)$ . Les ions pénétrant dans le quadripôle subissent ce champ électrique oscillant et ont une trajectoire plus ou moins stable, décrite par les équations de Mathieu. Pour  $U/V$  fixé, seuls les ions possédant une valeur de  $m/z$  définie auront une trajectoire stable jusqu'au détecteur tandis que les autres seront déviés. En modifiant les valeurs de tension, l'analyseur va donc effectuer un balayage de l'ensemble de la gamme des  $m/z$  de fonctionnement.

Dans un cadre d'analyse protéomique, le quadripôle est très souvent utilisé en tandem avec un autre analyseur, du fait de sa faible résolution et vitesse de scan. Il est également largement utilisé comme élément de focalisation (« RF only ») en mode MS et comme filtre pour la MS/MS.

## 4.3 La trappe orbitale (Orbitrap)

La trappe orbitale communément appelée orbitrap est un analyseur de type piège à ions qui a été développé à partir de 1923 par Kindon puis repris dans les années 1990 par l'équipe du Dr. Alexander Makarov.<sup>[33]</sup> C'est un analyseur breveté par la société ThermoFisher Scientific. Depuis sa commercialisation en 2005, l'orbitrap est devenu un analyseur de choix pour les analyses protéomiques<sup>[34]</sup>.

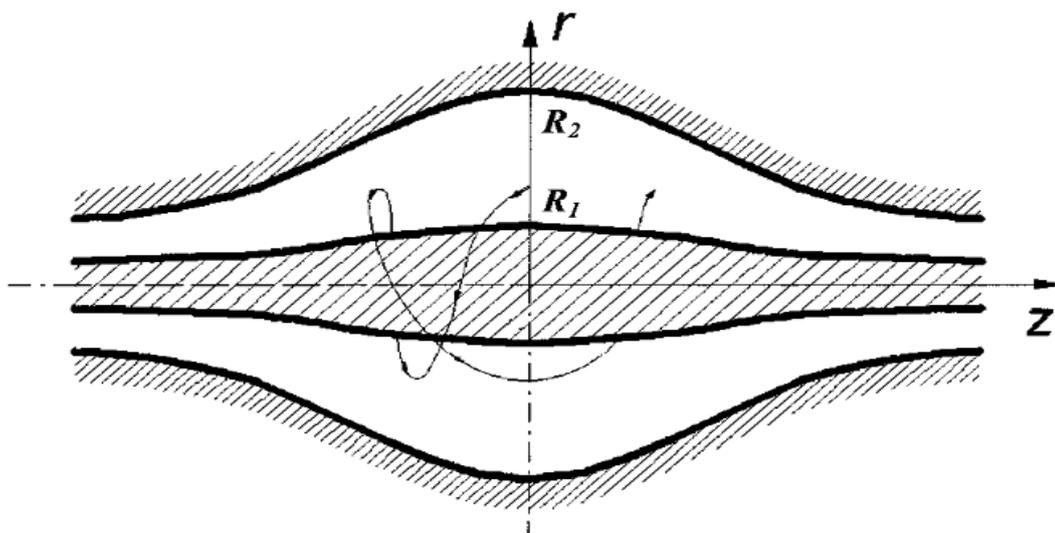


Figure 8 : Représentation schématique de l'orbitrap et du déplacement d'ions (Source : figure extraite de [35] )

Cet analyseur est composé d'une électrode centrale en forme de fuseau et de deux électrodes externes. Les ions sont introduits dans la trappe par paquets tangentiels à l'électrode centrale, suite à leur accumulation en amont dans une trappe linéaire (C-trap).

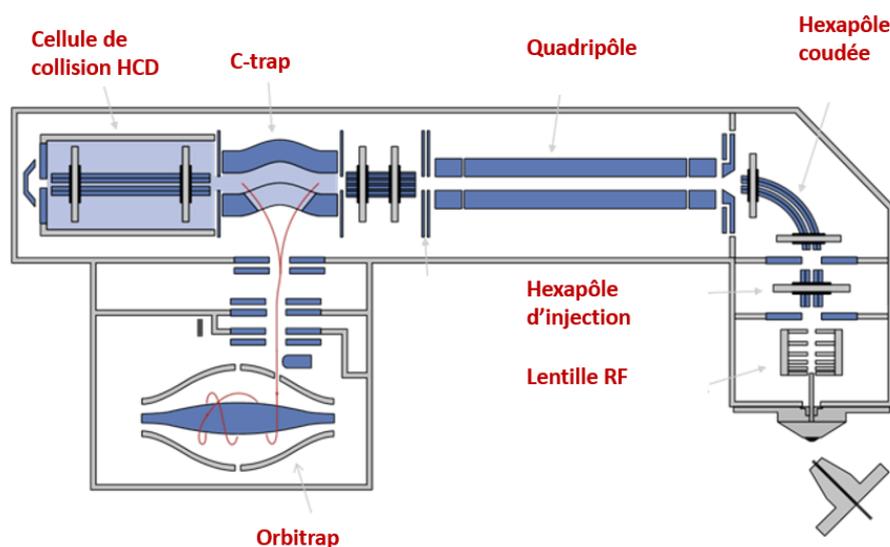


Figure 9 : Représentation schématique du spectromètre de masse ESI-Orbitrap (Orbitrap Qexactive Plus) (source : adapté de ThermoFisher)

Sous l'effet du champ, les ions vont donc entrer en oscillation harmonique autour de l'électrode centrale selon une fréquence qui est inversement proportionnelle à la racine carrée du rapport  $m/z$  de l'ion analysé, et effectuer un mouvement de rotation autour de l'électrode centrale ainsi qu'un mouvement axial le long de cette même électrode. Le mouvement axial des ions génère un courant induit enregistré par les électrodes externes qui sera exploité pour déterminer le rapport  $m/z$  via l'algorithme de la transformée de Fourier.

La stabilité des ions dans le champ électrique de l'analyseur est régie par la solution de l'équation de Mathieu. Le champ électrostatique présent dans l'orbitrap est décrit par l'équation :

$$U(r, z) = \frac{k}{2} \left( z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \ln \left[ \frac{r}{R_m} \right] + C$$

Où  $r$  et  $z$  sont les coordonnées cylindriques,  $R_m$ , le rayon caractéristique,  $C$  une constante,  $k$  le paramètre déterminé par la forme de l'électrode centrale et la tension appliquée.

Le mouvement des ions est par ailleurs décrit par l'équation suivante :

$$z(t) = z_0 \cos(\omega t) + \left( \frac{2E_z}{k} \right)^{\frac{1}{2}} \sin(\omega t)$$

Où  $Z_0$  est l'amplitude initiale et  $E_z$  l'énergie cinétique initiale selon la direction  $Z$  avec  $\omega$  la fréquence d'oscillation.

$$\omega = \left( \frac{kq}{m} \right)^{1/2}$$

La fréquence d'oscillation est régie par  $k$  une constante,  $q$  la charge et  $m$  la masse de l'ion

Cette équation montre ainsi que la fréquence axiale ne dépend donc seulement du rapport masse sur charge de l'ion. Cette fréquence ainsi enregistrée est ensuite convertie par le biais d'une transformée de Fourier en un spectre de masse.

Les analyseurs de type orbitrap ont un pouvoir résolutif excellent ce qui permet d'atteindre des précisions de mesure de masse  $< 1 \text{ ppm}$ <sup>[36]</sup>. Afin de maintenir ces performances tout au long d'expériences à haut débit telles que celles rencontrées en analyse protéomique, un « *lock mass* » dans le spectre est systématiquement introduit. Cette « *lock mass* » correspond à un ion de  $m/z$  connu avec précision qui est analysé en même temps que les ions d'intérêt. Par exemple, en protéomique, l'ion  $m/z$  445,120025 est souvent utilisé comme « *lock mass* », cet ion est formé à partir des molécules siloxanes présentes dans l'atmosphère. L'utilisation de « *lock mass* » permet d'améliorer la précision en masse, sans pour autant affecter la vitesse d'analyse.<sup>[37]</sup> Notamment, une amélioration de la précision en masse de 3,7 ppm à 0,9 ppm a été observée en utilisant un « *lock mass* ».

Depuis sa première commercialisation en 2005 de nouveaux instruments sont régulièrement mis sur le marché présentant des améliorations dans les performances en résolution ou gamme de  $m/z$ . Par exemple, l'orbitrap Fusion Lumos permet d'atteindre des résolutions de 500 000 ou l'orbitrap EMR pouvant atteindre une gamme de masse allant de 15 à 15000  $m/z$ .<sup>[38]</sup>

#### **4.4 Pourquoi a-t-on besoin d'analyseurs toujours plus performants ?**

- *Résolution* : Tout d'abord, un pouvoir de résolution élevé est nécessaire pour atteindre une précision de mesure suffisante pour attribuer une masse la plus proche possible de la masse exacte. D'autre part, un haut pouvoir de résolution permettra de définir de manière précise les massifs isotopiques des ions qui, pour un même rapport m/z, peuvent présenter des abondances relatives très différentes selon les ions jusqu'à constituer une empreinte caractéristique des molécules (notamment dans le cas des molécules organométalliques, mais également dans une moindre mesure pour les peptides) ou encore qui peuvent se chevaucher. Un massif isotopique défini avec précision conduira également à une détermination précise de la charge portée par l'ion.

- *Précision de mesure de masse* : la validation ou l'identification d'une structure moléculaire dépendra largement de la correspondance entre la masse mesurée et la masse théorique de cette molécule. Plus la précision de mesure de la masse sera élevée, meilleure sera la certitude de l'interprétation que ce soit lors d'une analyse par MS ou lors d'une analyse par fragmentation en MS/MS. Dans le cadre d'une analyse protéomique, le nombre de séquences en acides aminés potentielles décroît rapidement avec la précision de mesure de masse. [39, 40] La précision de mesure de masse est également liée au degré avec lequel les mesures fournies par les analyseurs sont reproductibles. Or, les analyseurs ne mesurent pas directement le rapport m/z, mais d'autres paramètres comme le temps de vol ou la fréquence réduite cyclotronique : une procédure d'étalonnage spécifique à chaque instrument est donc nécessaire pour convertir les quantités mesurées en rapport m/z en tenant compte des facteurs susceptibles d'influencer l'exactitude de la mesure telle que par exemple l'effet de charge spatiale ou les effets de bord des champs.

- *Rapidité de balayage* : cette caractéristique prend toute son importance dans le contexte d'analyse protéomique où le nombre de peptides issus de la digestion enzymatique du mélange protéique d'intérêt devient de plus en plus grand.

**Tableau 1 : Quelques caractéristiques des principaux analyseurs fournis par les constructeurs [41].**

Analyseur	Gamme de masse	Résolution	Précision de mesure (ppm)
Quadripôle (Q)	15 – 2 000	1 000 – 6 000	100 ppm
Piège à ions (IT)	0 – 4 000	5 000 – 30 000	< 100 ppm
Temps de vol (TOF)	0 – 25 000	5 000 – 75 000	< 2 ppm
Trappe électrostatique (Orbitrap)	15 – 4 000	15 000 – 150 000	< 1 ppm
Résonance Cyclotronique (ICR)	100 – 10 000	20 000 – 10 <sup>6</sup>	< 1 ppm

#### **4.5 Etalonnage des analyseurs**

Deux types d'étalonnage sont disponibles, externe et interne. L'étalonnage externe est effectué en amont de l'analyse et fixe les paramètres instrumentaux sur toute la durée de l'expérience ou d'une série d'expériences. L'étalonnage interne consiste en la co-injection d'un mélange de standards de rapports  $m/z$  connus qui permettent de compenser les erreurs liées à des variations des paramètres instrumentaux (dérive de la température, fluctuation de la charge spatiale...).

La calibration de l'Orbitrap est réalisée de manière automatisée. L'opérateur règle dans un premier temps les conditions d'analyses en particulier les paramètres de source ou d'acquisition tels que la gamme de masse (généralement comprise entre 150 et 2000  $m/z$ ), la résolution de l'instrument (dans notre cas 120000 en MS et 17500 en MS/MS), le mode d'ionisation. Le nombre de scans réalisés, le nombre d'ions accumulés dans l'orbitrap (AGC target) ainsi que le temps d'injection (environ 30ms) sont également fixés. La calibration interne est réalisée sur une solution calibrante standard constituée de molécule aromatique (caféine), d'une séquence peptidique ainsi que d'un polymère. La calibration va ainsi optimiser les différents paramètres tels que les tensions du quadripôle, les tensions sur les lentilles de focalisation... afin que la détection des molécules présentes dans la solution de calibration soit réalisée à une bonne sensibilité. Celle-ci est effectuée deux fois par mois.

### **5. Spectrométrie de masse en tandem (MS/MS)**

Pour obtenir des informations précises sur des échantillons complexes, la MS en tandem (MS/MS) est primordiale car elle permettra de différencier ou de valider des structures chimiques parmi des ions de rapports  $m/z$  très proches. La MS en tandem s'effectue en trois étapes : sélection de l'ion précurseur, fragmentation de cet ion précurseur et analyse des ions fragments résultants. De nombreux appareils peuvent être ainsi utilisés pour effectuer de la MS/MS. Les trois étapes peuvent s'opérer dans un même espace (trappe d'ions) ou dans des compartiments différents (analyseur 1, cellule de collision et analyseur 2).



**Figure 10 : Schéma représentant les différentes composantes d'un spectromètre de masse en mode MS/MS**

## **5.1 La fragmentation**

Elle vise à rompre des liaisons covalentes sur l'ion précurseur afin de générer des ions fragments dont les rapports m/z guideront à l'identification non ambiguë de sa structure. De nombreuses méthodes ont été développées pour fragmenter les ions mettant en jeu des mécanismes de rupture de liaison différents ce qui permet de caractériser des molécules de structure très variées et d'en retirer des informations complémentaires.

### **5.1.1 La Dissociation induite par collision (CID)**

La dissociation induite par collision (CID) également appelée dissociation activée par collision (CAD), implique la collision d'un ion avec un atome ou une molécule neutre en phase gazeuse et la dissociation ultérieure de l'ion.

Selon l'instrument utilisé, cette cellule de collision peut être :

- De type faisceau : les ions précurseurs sont fragmentés en cours de vol sur des instruments tels que les Q-TOF.
- De type piège à ions : les ions précurseurs sont d'abord piégés, puis fragmentés.

C'est le mode de fragmentation le plus utilisé en analyse protéomique. Il est basé sur le modèle du proton mobile<sup>[41]</sup> et est particulièrement bien adapté aux peptides tryptiques doublement chargés. Les collisions entraînent la délocalisation d'un proton mobile (qui est facilement transférable le long du squelette peptidique) puis la rupture de la liaison amide (C – N qui est la plus faible). Lorsque le squelette peptidique est rompu, la charge peut être retenue par l'extrémité C- ou N-terminale du site de fragmentation. La nomenclature des ions fragments résultants a été décrite par Biemann.<sup>[42]</sup> Ainsi, la fragmentation à basse énergie (en eV) générera principalement des ions b et y, alors qu'une fragmentation à plus haute énergie (en keV) peut engendrer la rupture d'autres liaisons et l'obtention d'ions de type a, x, c, z.

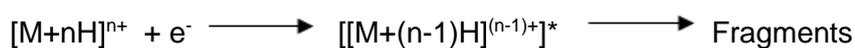
Selon la théorie du proton mobile, la présence d'un résidu fortement basique à l'intérieur de la séquence peptidique peut affecter sensiblement le déplacement du proton mobile et donc l'efficacité de fragmentation<sup>[43]</sup>. La présence de résidus basiques supplémentaires peut ainsi augmenter l'état de charge de l'ion en ESI et générer des ions 3 voire 4 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.

Avec le développement des instruments de type Orbitrap, un autre type de fragmentation basé sur de la dissociation induite par collision a été développé. Cette fragmentation se nomme la fragmentation HCD (Higher Collision Dissociation).<sup>[44]</sup> Le principe de la fragmentation est exactement le même que celui du CID. Le terme higher collision (collision plus élevée) fait ici référence à l'énergie nécessaire pour piéger les ions avant leur entrée dans la cellule de collision.

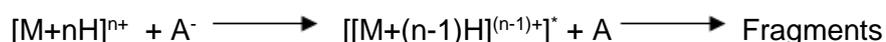
### 5.1.2 Fragmentation par capture et transfert d'électrons

Une autre possibilité pour induire la fragmentation d'une molécule est d'utiliser l'énergie libérée lorsqu'un électron est transféré ou capturé par un ion à charges multiples. Il existe plusieurs techniques pour effectuer de la fragmentation par capture et transfert d'électrons :

La dissociation par capture électronique (ECD) : un électron est ajouté à un ion positif multichargé. La réaction de dissociation est :



La dissociation par transfert d'électron (ETD) : l'ajout d'un électron par une réaction ionique permet la fragmentation d'une molécule par transfert d'électron. Elle induit la fragmentation des cations en leur transférant des électrons.



L'ECD est mise en œuvre sur les spectromètres de masse de type FTICR. Cette technique favorise la fragmentation des peptides le long du squelette peptidique tout en préservant les chaînes latérales des acides aminés <sup>[45]</sup>. Ainsi, les modifications labiles en CID telles que certaines glycosylations ou la phosphorylation sont conservées sur les acides aminés, ce qui facilite le séquençage et la localisation des sites modifiés. L'ETD a été initialement développé sur les trappes ioniques, mais des modules ETD équipent aujourd'hui d'autres instruments commerciaux tels que les orbitraps ou les Q-TOFs. En ETD, des anions radicalaires de fluoranthène ou d'azulène sont utilisés pour transférer un électron sur les ions précurseurs multi-protonés préalablement isolés. 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. Contrairement au mode CID, l'ETD est particulièrement bien adapté pour fragmenter les espèces multi-protonées telles que les peptides ou même les protéines en laissant intact les chaînes latérales. <sup>[45]</sup>. L'ETD est ainsi complémentaire à la CID.

Récemment, un type de fragmentation alternatif a été décrit. Cette fragmentation se nomme la dissociation par collision à transfert d'énergie et à haute énergie (ETHcD) qui est une combinaison de la fragmentation ETD et HCD.<sup>[46]</sup> La molécule d'intérêt est soumise dans un premier temps à une réaction ion/ion dans une trappe linéaire. Dans la deuxième étape, la fragmentation HCD est appliquée à tous les ions formés par ETD. Cette méthode de fragmentation est utilisée pour générer des spectres MS/MS riches en ions et faciliter le séquençage des peptides ainsi que la localisation de modification post-traductionnelles.

### **5.1.3 IRMPD**

Il s'agit d'une méthode de fragmentation par 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 : la dissociation multi-photonique infrarouge (IRMPD : InfraRed MultiPhoton Dissociation) et de la photodissociation par rayonnement ultra-violet (UVPD : UltraViolet PhotoDissociation). Aujourd'hui, seuls des spectromètres de masse de type FTICR équipés de laser IR pour l'IRMPD sont disponibles commercialement.

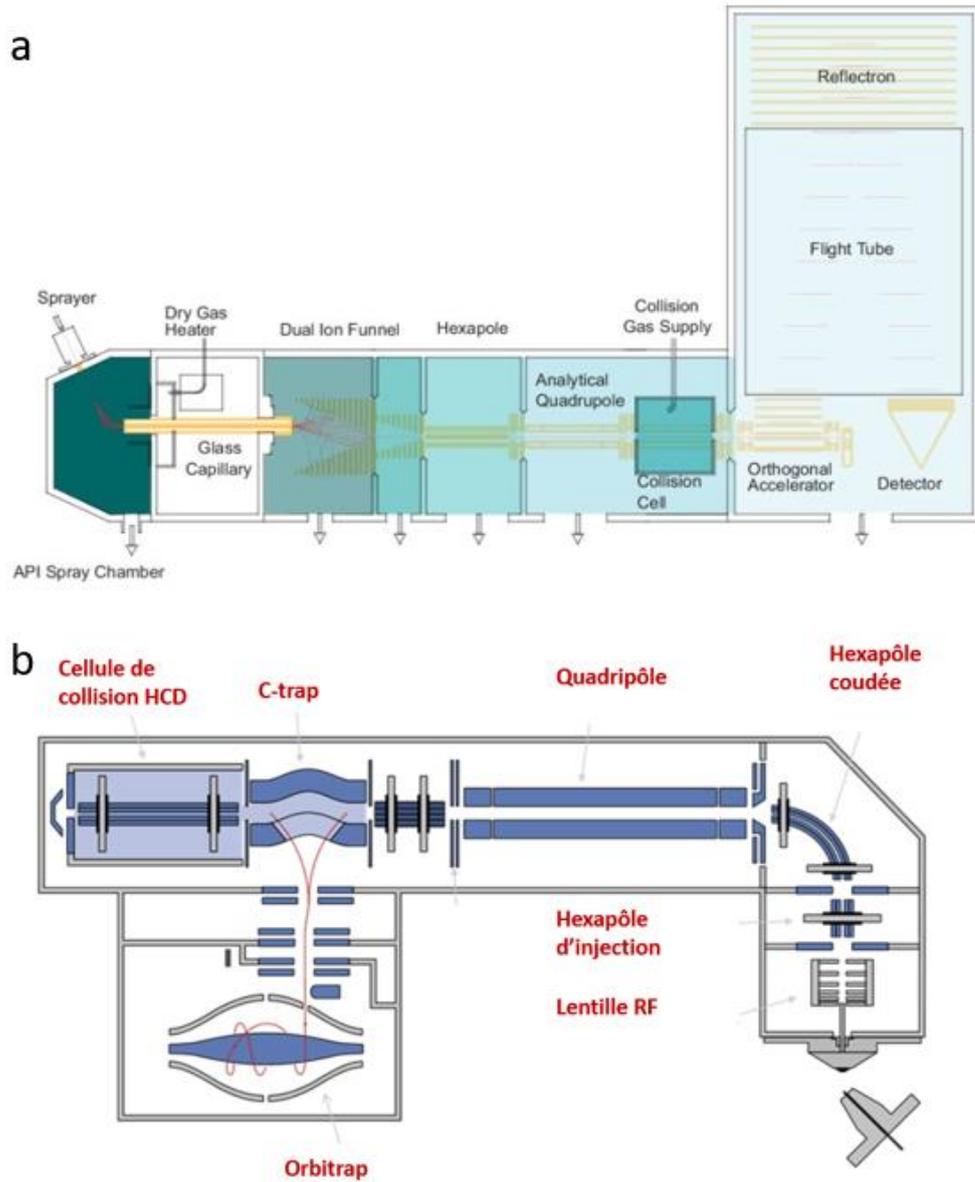
### **5.2 Appareillage permettant la MS/MS**

Il existe de nombreuses configurations instrumentales permettant d'effectuer de la MS/MS. Les quadripôles sont très utilisés, car ils permettent non seulement de servir d'analyseur, mais également servir de chambres de collision. Ainsi, on peut obtenir des spectromètres de masses de types triple-quadripôles (QqQ) et des spectromètres de masses combinant des quadripôles avec un analyseur à temps de vol (Q-TOF).

Certains analyseurs peuvent fonctionner de manières différentes tels que les instruments à piège à ions (trappe linéaire ou orbitrap) où la séparation des ions est réalisée avec des ions piégés au même endroit.

## **6. Conclusion**

Durant ce travail de thèse, nous avons essentiellement travaillé avec des instruments de type ESI-Q-TOF (Micro-Tof QII, Bruker), ESI-Orbitrap (Q-ex plus, ThermoFisher) et MALDI-TOF (Autoflex, Bruker).



**Figure 11 : Représentation schématique d'un spectromètre de masse (a) de type Q-TOF (extrait de [47]) et de (b) l'orbitrap Qexactive plus (adapté du manuel d'utilisation de ThermoFisher)**

## 7. Bibliographie

1. Karas, M., D. Bachmann, and F. Hillenkamp, *Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules*. Analytical Chemistry, 1985. **57**(14): p. 2935-2939.
2. Mème, A., *Développement de la spectrométrie de masse pour l'étude d'édifices non-covalents en chimie et en biologie*. 2010, University of Strasbourg.
3. Leisner, A., et al., *Infrared laser post-ionization of large biomolecules from an IR-MALDI(I) plume*. Journal of the American Society for Mass Spectrometry, 2004. **15**(6): p. 934-941.
4. Woods, A.S., et al., *IR-MALDI-LDI combined with ion mobility orthogonal time-of-flight mass spectrometry*. Journal of proteome research, 2006. **5**(6): p. 1484-1487.
5. Dreisewerd, K., *The desorption process in MALDI*. Chemical Reviews, 2003. **103**(2): p. 395-425.
6. Chen, F., et al., *Applying mass spectrometry to study non-covalent biomolecule complexes*. Mass Spectrometry Reviews, 2016. **35**(1): p. 48-70.
7. Hillenkamp, F., Peter-Katalinic, J., *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications*. Wiley-VCH, 2007, p. 1-362.
8. Toh-Boyo, G.M., S.S. Wulff, and F. Basile, *Comparison of Sample Preparation Methods and Evaluation of Intra- and Intersample Reproducibility in Bacteria MALDI-MS Profiling*. Analytical Chemistry, 2012. **84**(22): p. 9971-9980.
9. Zenobi, R. and R. Knochenmuss, *Ion formation in MALDI mass spectrometry*. Mass Spectrometry Reviews, 1998. **17**(5): p. 337-366.
10. Ehring, H., M. Karas, and F. Hillenkamp, *Role of Photoionization and Photochemistry in Ionization Processes of Organic-Molecules and Relevance for Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry*. Organic Mass Spectrometry, 1992. **27**(4): p. 472-480.
11. Knochenmuss, R. and R. Zenobi, *MALDI ionization: The role of in-plume processes*. Chemical Reviews, 2003. **103**(2): p. 441-452.
12. Karas, M. and R. Kruger, *Ion formation in MALDI: The cluster ionization mechanism*. Chemical Reviews, 2003. **103**(2): p. 427-439.
13. Kruger, R., et al., *Analyte incorporation and ionization in matrix-assisted laser desorption/ionization visualized by pH indicator molecular probes*. Analytical Chemistry, 2001. **73**(24): p. 5812-5821.
14. McCarley, T.D., R.L. McCarley, and P.A. Limbach, *Electron transfer ionization in matrix assisted laser desorption/ionization mass spectrometry*. Analytical Chemistry, 1998. **70**(20): p. 4376-4379.
15. Buck, A., et al., *High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples*. Journal of Pathology, 2015. **237**(1): p. 123-132.
16. Laiko, V.V., S.C. Moyer, and R.J. Cotter, *Atmospheric pressure MALDI/ion trap mass spectrometry*. Analytical Chemistry, 2000. **72**(21): p. 5239-5243.
17. Muller, M.Q., et al., *Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS*. Analytical Chemistry, 2010. **82**(16): p. 6958-6968.
18. Aichler, M. and A. Walch, *MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice*. Laboratory Investigation, 2015. **95**(4): p. 422-431.
19. Mitobe, J., et al., *Superstructure formation by RodZ hexamers of Shigella sonnei maintains the rod shape of bacilli*. PLOS ONE, 2020. **15**(2).
20. Weidmann, S. and R. Zenobi, *High-Mass MALDI-MS Using Ion Conversion Dynode Detectors: Influence of the Conversion Voltage on Sensitivity and Spectral Quality*. Journal of the American Society for Mass Spectrometry, 2014. **25**(6): p. 950-954.

21. Aksenov, A.A. and M.E. Bier, *The Analysis of Polystyrene and Polystyrene Aggregates into the Mega Dalton Mass Range by Cryodetection MALDI TOF MS*. Journal of the American Society for Mass Spectrometry, 2008. **19**(2): p. 219-230.
22. Lu, M., et al., *Nanomaterials as Assisted Matrix of Laser Desorption/Ionization Time-of-Flight Mass Spectrometry for the Analysis of Small Molecules*. Nanomaterials (Basel, Switzerland), 2017. **7**(4): p. 87-88.
23. Guo, Z., et al., *A Method for the Analysis of Low-Mass Molecules by MALDI-TOF Mass Spectrometry*. Analytical Chemistry, 2002. **74**(7): p. 1637-1641.
24. Whitehouse, C.M., et al., *Electrospray interface for liquid chromatographs and mass spectrometers*. Analytical Chemistry, 1985. **57**(3): p. 675-679.
25. Yamashita, M. and J.B. Fenn, *Electrospray Ion-Source - Another Variation on the Free-Jet Theme*. Journal of Physical Chemistry, 1984. **88**(20): p. 4451-4459.
26. Konermann, L., et al., *Unraveling the Mechanism of Electrospray Ionization*. Analytical Chemistry, 2013. **85**(1): p. 2-9.
27. Kebarle, P., *A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry*. Journal of Mass Spectrometry, 2000. **35**(7): p. 804-817.
28. Takáts, Z., et al., *Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization*. Science, 2004. **306**(5695): p. 471.
29. Takáts, Z., J.M. Wiseman, and R.G. Cooks, *Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology*. Journal of Mass Spectrometry, 2005. **40**(10): p. 1261-1275.
30. Constantopoulos, T.L., G.S. Jackson, and C.G. Enke, *Effects of salt concentration on analyte response using electrospray ionization mass spectrometry*. Journal of the American Society for Mass Spectrometry, 1999. **10**(7): p. 625-634.
31. Kebarle, P. and L. Tang, *Dependence of Ion Intensity in Electrospray Mass-Spectrometry on Concentration in the Solution*. Abstracts of Papers of the American Chemical Society, 1994. **207**: p. 3654-3668.
32. Ganem, B., Y.T. Li, and J.D. Henion, *Observation of Noncovalent Enzyme Substrate and Enzyme Product Complexes by Ion-Spray Mass-Spectrometry*. Journal of the American Chemical Society, 1991. **113**(20): p. 7818-7819.
33. Scigelova, M. and A. Makarov, *Makarov, A.: Orbitrap Mass Analyzer - Overview and Applications in Proteomics. Practical Proteomics 2, 16-21*. Proteomics, 2006. **6** : p. 16-21.
34. Macek, B., et al., *Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer*. Molecular & Cellular Proteomics, 2006. **5**(5): p. 949-958.
35. Makarov, A., *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis*. Analytical Chemistry, 2000. **72**(6): p. 1156-1162.
36. ThermoFisher. *Liquid Chromatography-Mass Spectrometry : Orbitrap LC-MS*. 2020 [cited 2020 12/10/2020].
37. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap*. (1535-9476 ).
38. Maple, H.J., et al., *Application of the Exactive Plus EMR for automated protein-ligand screening by non-covalent mass spectrometry*. Rapid Communications in Mass Spectrometry, 2014. **28**(13): p. 1561-1568.
39. Cardoza, J.D., et al., *Mass spectrometry-based proteomics: qualitative identification to activity-based protein profiling*. Wiley interdisciplinary reviews. Systems biology and medicine, 2012. **4**(2): p. 141-162.
40. Zubarev, R.A., P. Håkansson, and B. Sundqvist, *Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements*. Analytical Chemistry, 1996. **68**(22): p. 4060-4063.

41. Boyd, R. and Á. Somogyi, *The mobile proton hypothesis in fragmentation of protonated peptides: A perspective*. Journal of the American Society for Mass Spectrometry, 2010. **21**(8): p. 1275-1278.
42. Johnson, R.S., S.A. Martin, and K. Biemann, *Collision-induced fragmentation of (M + H)<sup>+</sup> ions of peptides. Side chain specific sequence ions*. International Journal of Mass Spectrometry and Ion Processes, 1988. **86**: p. 137-154.
43. Kinter, M.a.S., N.E., *Collisionally Induced Dissociation of Protonated Peptide Ions and the Interpretation of Product Ion Spectra*, in *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. 2000. p. 64-116.
44. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis*. Nature Methods, 2007. **4**(9): p. 709-712.
45. Zhurov, K.O., et al., *Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis*. Chemical Society Reviews, 2013. **42**(12): p. 5014-5030.
46. Reiding, K.R., et al., *The benefits of hybrid fragmentation methods for glycoproteomics*. TrAC Trends in Analytical Chemistry, 2018. **108**: p. 260-268.
47. Beck, S., et al., *The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics*. Mol Cell Proteomics, 2015. **14** : p. 2014-2029.

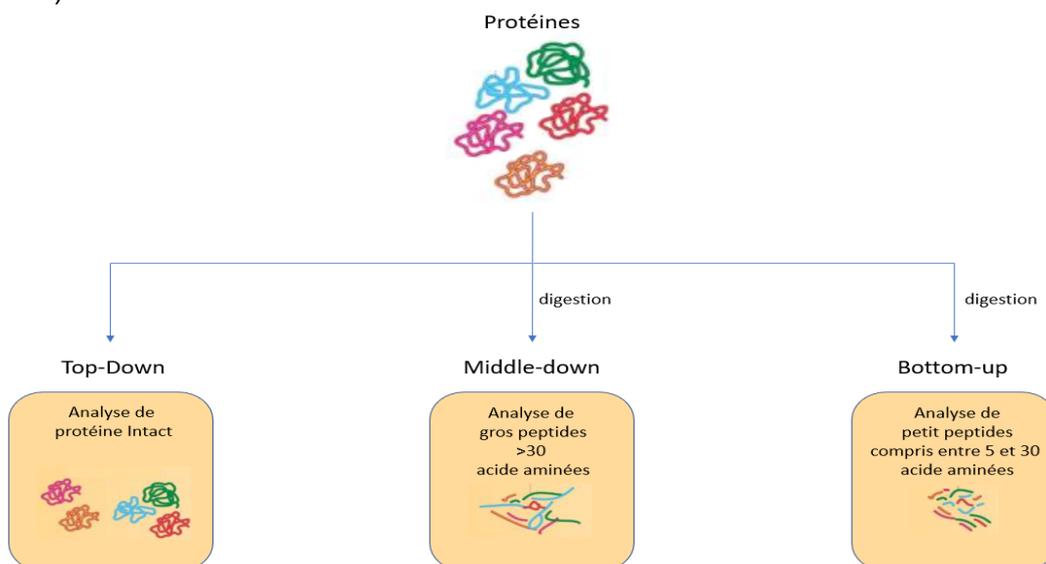


# Chapitre II : Stratégies d'analyse protéomique par MS

## 1. Introduction

Outre les études de protéines intactes ou de complexes biologiques en conditions non dénaturantes, toutes les approches d'analyse structurale basées sur la MS visent à positionner les acides aminés les uns par rapport aux autres afin de construire une structure tridimensionnelle de plus en plus précise. Pour cela, afin de les repérer grâce à une masse moléculaire qui leur sera spécifique, certains acides aminés seront modifiés selon des conditions expérimentales directement liées à la conformation de la protéine d'intérêt puis l'identification de ces acides aminés s'effectuera en utilisant les protocoles classiquement décrits en analyse protéomique.

L'analyse protéomique consiste à identifier, caractériser et quantifier l'ensemble des protéines exprimées dans une cellule ou un tissu à un moment donné, dans des conditions données et avec un historique donné<sup>[1]</sup>. Aujourd'hui des avancées techniques considérables ont rendu cette « nouvelle science » très populaire et ont conduit au développement de nouvelles approches méthodologiques. Ces avancées concernent aussi bien les techniques séparatives en amont de la MS (techniques électrophorétiques.<sup>[2, 3]</sup>) que la MS proprement parlée (développement d'analyseurs à haute résolution comme les Orbitrap<sup>[4]</sup> ou FTICR<sup>[5]</sup>, nouvelles techniques de fragmentation par ETD<sup>[6]</sup>). A ces avancées techniques s'est combiné le développement d'outils bio-informatiques (avec la création des bases de données regroupant toutes les protéines connues pour l'ensemble des espèces ou l'amélioration constante des moteurs de recherche adaptés aux formats de données générés toujours plus complexes).



**Figure 12 : Représentation schématique des différentes stratégies protéomiques utilisées.**

L'approche la plus couramment utilisée en protéomique est l'approche « bottom-up »<sup>[7]</sup>. Elle consiste à protéolyser (ou digérer) à l'aide d'une enzyme les protéines et à analyser les peptides de digestion par chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS). Les peptides sont ensuite identifiés à partir de l'interprétation des spectres MS et MS/MS acquis en les confrontant aux séquences protéiques stockées dans des banques de données (telles que Uniprot<sup>[8]</sup>), ce qui permet ensuite de remonter à l'identification des protéines.

L'approche « middle-down »<sup>[9]</sup>, diffère de l'approche précédente par l'enzyme utilisée, en générant des peptides plus gros, ou polypeptides, dont la masse peut avoisiner les 20kDa.

Enfin, la dernière approche, appelée « top-down »<sup>[10]</sup> consiste à fragmenter les protéines entières. L'avantage essentiel de cette approche est qu'elle permet d'éviter d'éventuelles pertes de modifications qui pourraient être instables pendant la longue étape de protéolyse enzymatique. L'essor de cette approche a été rendu possible grâce à l'intégration des modes de fragmentation ECD-ETD dans des instruments à haute résolution et haute précision de masse (de type FTICR, Orbitrap ou HR-Q-TOF). Les protéines sont identifiées grâce aux spectres de masse MS et MS/MS, qui sont très complexes étant donné le nombre d'états de charge que peuvent prendre les fragments de protéines. C'est pourquoi, outre une bonne résolution ( $\geq 140000$ ), les spectromètres de masse doivent aussi posséder une gamme de masse étendue (20000m/z).

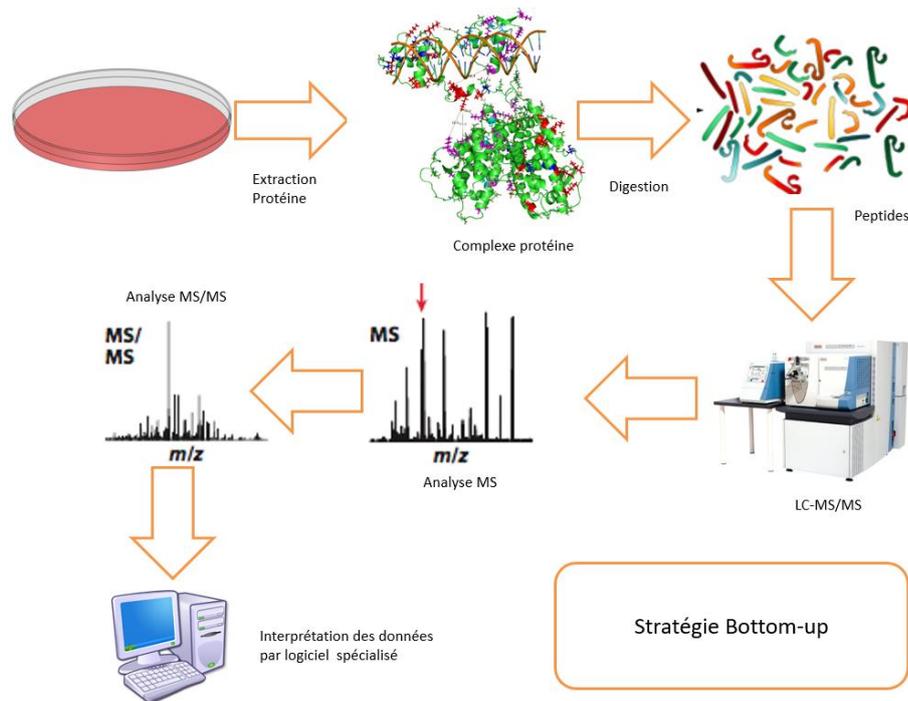
Une des applications majeures avec ce type d'approche est la mise en évidence de la coexistence de différentes isoformes d'une protéine produite à partir des mêmes gènes<sup>[11]</sup> ou encore l'identification de modifications post-traductionnelles (PTMs)<sup>[12]</sup> pouvant être à l'origine de dysfonctionnements physiologiques majeurs. Une limitation serait que les ions fragments correspondent majoritairement aux extrémités terminales des protéines laissant ainsi le cœur de la séquence vide d'informations. Ce phénomène augmente avec la taille de la protéine.

Dans la suite de cette partie seront décrites les différentes étapes de l'approche « bottom-up », qui a été utilisée dans mes travaux de thèse.

## 2. Stratégies d'identification de protéines et de peptides par spectrométrie de masse

### 2.1 La Stratégie Bottom-up

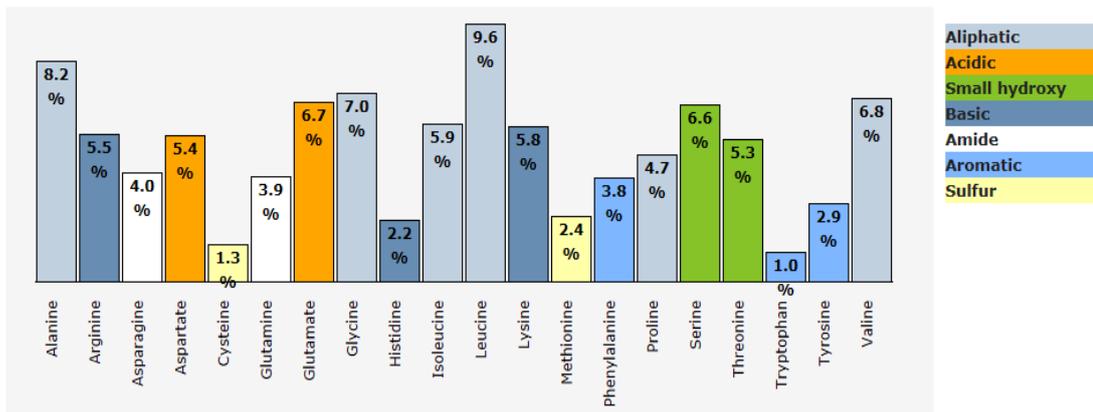
La stratégie Bottom-up est la stratégie d'identification de protéines actuellement la plus utilisée en analyse protéomique.



**Figure 13 Représentation schématique de la stratégie protéomique "bottom-up"**

La première étape consiste à digérer en peptides l'ensemble de protéines présentes dans un échantillon grâce à un agent d'hydrolyse, en général une protéase. Afin d'augmenter l'accessibilité de la protéase aux sites d'hydrolyse, une étape consistant à réduire les ponts disulfures (S-S) par chauffage en présence d'agents réducteurs (tels que le dithiothréitol DTT ou le tris-(2 carboxyéthyl) phosphine TCEP) est quasi systématiquement effectuée. Parfois, le recours à l'utilisation d'agents dénaturants (l'urée ou l'hydrochlorure de guanidine) ou de détergents (SDS) peut également être nécessaire. L'enzyme la plus fréquemment utilisée est la trypsine. Cette enzyme hydrolyse la liaison peptidique du côté C-terminal des lysines et arginines, sauf si ces acides aminés sont suivis par une proline. Elle est particulièrement bien adaptée à une analyse par LC-MS/MS, car (i) la plupart des protéines possèdent en moyenne 10% d'acides aminés de type K ou R répartis régulièrement sur la séquence (Cf figure 14), ce qui permet de générer des peptides de masses moléculaires souvent comprises entre 400 et 4000 Da, (ii) les peptides obtenus par ESI sont majoritairement di-chargés (ce qui est favorable pour une fragmentation classique par CID-MS/MS), (iii) la trypsine est tolérante à de légères variations des conditions expérimentales (présence d'un faible pourcentage de solvant organique ou de détergent).

De très nombreuses protéases sont également compatibles avec l'approche bottom-up et peuvent être soit utilisées individuellement en mode complémentaire, soit combinées dans une expérience unique. De façon intéressante, certaines équipes ont recours à une étape de pré-protéolyse par une autre enzyme lorsque la trypsine ne permet pas à elle seule de cliver tous les peptides d'intérêt, notamment la LysC endopeptidase<sup>[13]</sup>. Contrairement à la trypsine, cette enzyme clive également les lysines suivies d'une proline.



**Figure 14 : Distribution statistique des acides aminés présents dans les protéines dans la base de données Uniprot.**

La deuxième étape consiste en l'analyse proprement dite. Le mélange peptidique sera alors décomplexifié à l'aide de techniques chromatographiques (chromatographie liquide en phase inverse notamment) ou par électrophorèse capillaire (CE)<sup>[14]</sup> puis soumis à de la fragmentation dans le spectromètre de masse afin d'obtenir des informations de séquence en acides aminés. L'instrument est alors paramétré pour sélectionner puis fragmenter automatiquement les ions par alternance de cycles MS et MS/MS. Ce mode d'acquisition s'appelle le mode DDA (data-dependant acquisition). La sélection des ions est basée sur leur intensité : le plus souvent on parle de « top N » c'est-à-dire qu'on sélectionne, les uns après les autres, les N ions les plus intenses du spectre MS correspondant. Cette série d'évènements (1 spectre MS + N spectres MS/MS) constitue un cycle d'analyse. En fonction des performances de l'analyseur du spectromètre de masse considéré, l'utilisateur choisira (i) le nombre d'ions précurseurs à isoler par cycle MS, (ii) le seuil d'intensité ainsi que l'état de charge préférentiel de l'ion pour être sélectionné, (iii) le temps d'exclusion dynamique qui correspond au temps durant lequel un rapport m/z déjà fragmenté ne doit plus être sélectionné, (iv) l'énergie de collision de la cellule de collision.

Dans la dernière étape, la combinaison des données obtenues par MS et MS/MS sera alors exploitée par confrontation avec les banques de données protéiques ou génomiques pour l'identification des protéines. Il existe deux approches principales conduisant à l'identification d'une protéine par LC-MS/MS :

- Une approche par recherche d'homologie de masse : les spectres MS/MS expérimentaux sont comparés aux spectres théoriques issus de la digestion *in silico* des protéines contenues dans la banque de données. Différents paramètres de recherches tels que la taxonomie, la nature de l'enzyme utilisée, les modifications chimiques autorisées, le type de fragmentation MS/MS, les tolérances d'erreur en masse pour les ions précurseurs et les ions fragments sont précisées afin de rendre possible la convergence de l'algorithme de recherche. Plus un peptide est identifié avec un grand nombre de fragments, plus l'identification est fiable.

- Une approche par homologie de séquence grâce à du séquençage *de novo* : des informations de séquences peptidiques sont ici extraites à partir des spectres de fragmentations MS/MS avant d'être soumises à des requêtes d'homologie dans des banques de données protéiques ou génomiques. Cette approche a été proposée par M.Mann<sup>[15]</sup> en 1994. Il rapporte l'utilité de séquences en AA, même partielles, déduites à partir de la différence de masse observée entre les ions fragments de même nature du spectre MS/MS. Des logiciels ont aujourd'hui été conçus pour faciliter la démarche de séquençage *de novo* tels que le logiciel PEAKS<sup>[16]</sup>. Néanmoins, cette approche s'avère longue et délicate, nécessitant souvent encore une étape de validation manuelle de l'interprétation, ce qui limite son application à des problématiques ciblées à bas débit. Elle est en revanche très performante pour identifier et localiser des mutations d'AA.

## **2.2 Fragmentations des peptides**

L'analyse des digests peptidiques est très souvent basée sur l'interprétation des spectres de fragmentation, car, aux vues des échantillons toujours plus complexes et hétérogènes en termes de gamme de concentration, il est impératif de disposer d'une spécificité suffisante pour une identification non ambiguë des protéines. Dans un contexte d'analyse protéomique, le mode de fragmentation le plus couramment utilisé est la CID, car il est particulièrement bien adapté à la fragmentation des peptides tryptiques doublement chargés.

Aujourd'hui, une nomenclature décrivant la nature des ions fragments générés en phase gazeuse est disponible<sup>[17]</sup>. On a ainsi pu différencier 2 types d'ions :

- Les fragments dont la charge est portée sur la partie N-terminale. On va alors parler de série a, b, c.
- Les fragments dont la charge est portée sur la partie C-terminale. On va alors parler de série x, y, z.

Au cours de la fragmentation, il est aussi possible d'observer la perte de molécules neutres telles que les molécules d'eau (H<sub>2</sub>O) et les molécules d'ammoniac (NH<sub>3</sub>).

Sur le spectromètre de masse que nous avons été amenés à utiliser lors de nos études structurales (Orbitrap Qexactive plus), seule la fragmentation par HCD<sup>[18]</sup> a été utilisée. La fragmentation des peptides s'effectue alors essentiellement au niveau des liaisons amides composant la liaison peptidique (Figure 15 fragments b et y). Toutefois, il est également possible d'observer des fragments de type a et c, issus de la rupture de la chaîne latérale des acides aminés, ces ions étant issus de la fragmentation à haute énergie.

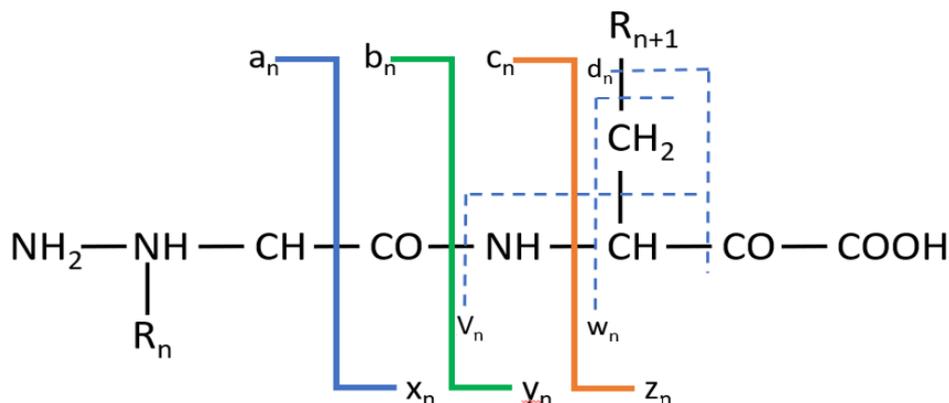


Figure 15 : Nomenclature de la fragmentation peptidique.

### **2.3 Evaluation de l'identification des protéines à l'aide d'outils bio-informatique**

Après l'analyse LC-MS/MS, un fichier de type mgf (mascot generic file) est généré contenant pour chaque spectre le rapport m/z de l'ion précurseur ainsi que la liste des masses des ions fragments correspondants, la charge de l'ion précurseur ainsi que son temps de rétention. Un des outils les plus utilisés pour l'identification des protéines est l'algorithme de recherche Mascot. Bien que Mascot<sup>[19]</sup> soit historiquement l'algorithme le plus populaire, d'autres algorithmes de recherche sont aujourd'hui couramment utilisés tels que Sequest<sup>[20]</sup>, OMSSA<sup>[21]</sup>, MS Amanda<sup>[22]</sup> ou Andromeda<sup>[23]</sup>. Tous ces logiciels sont très similaires sur le principe, néanmoins, chacun utilise un algorithme ainsi qu'un traitement des résultats spécifique, et ainsi des divergences significatives peuvent ressortir lors du rendu des résultats d'identifications des protéines (tant dans le nombre de protéines identifiées que dans la certitude de l'identification reflétée par un score). Ainsi l'utilisation de plusieurs moteurs de recherche peut s'avérer intéressante pour renforcer la validation des identifications.

L'algorithme effectue la recherche selon des informations spécifiées par l'utilisateur telles que la taxonomie, l'enzyme de digestion utilisée, les modifications post-traductionnelles possiblement présentes, la tolérance en masses sur les peptides et leurs fragments ou encore la banque de données. En effet, seules les protéines présentes dans la banque de données sélectionnée ne pourront être identifiées.

Une fois l'interrogation effectuée, afin d'évaluer la qualité de l'assignation d'une séquence à un spectre (Peptide Spectrum Match PSM), l'algorithme de recherche attribue un score sur l'ensemble des spectres qu'il a identifié. Ce score est un indice de confiance attribué aux peptides identifiés. Il reflète globalement la probabilité que l'assignation soit due au hasard. Plus le score est élevé, meilleure sera la confiance en l'assignation. Ce score dépend du nombre de fragments identifiés, de l'intensité des ions fragments attribués et de l'erreur de masse. L'utilisateur peut ainsi accepter ou refuser l'identification des peptides ou des protéines en fixant un score seuil en dessous duquel l'identification n'est pas acceptée.

Un taux de faux positifs important et variable d'une expérience à l'autre a poussé la communauté scientifique à établir un ensemble de directives pour l'analyse et la publication de données protéomiques<sup>[24]</sup>. Une stratégie nommée « target-decoy » permet de calculer un taux de faux positifs (FDR, False Discovery Rate) en effectuant la recherche en parallèle dans la base de données cible et dans une banque leurre dite *decoy*. Cette banque *decoy* contient des séquences inversées ou aléatoires par rapport à la banque cible, dans le but de repérer les spectres qui seront interprétés de manière erronée. Le FDR est ainsi calculé en effectuant le ratio du nombre de correspondances dans la base *decoy* et du nombre d'identifications dans la banque cible. Les seuils de score seront ensuite ajustés afin d'obtenir le taux de FDR souhaité.

De façon surprenante, une fraction relativement modeste des spectres MS/MS enregistrés conduit à une identification peptidique. Ceci est lié à plusieurs facteurs tels que la qualité du spectre MS/MS, la présence de peptides modifiés, une mauvaise attribution de l'état de charge du précurseur, une digestion non spécifique de la protéase, des peptides correspondant à une protéine non présente dans la banque de données.

Au cours de cette thèse, l'approche protéomique bottom-up sera utilisée par défaut lors de la mise en place de l'approche XL-MS.

### 3. Bibliographie

1. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
2. Chakavarti, B. and D. Chakavarti, *Electrophoretic separation of proteins*. Journal of visualized experiments : JoVE, 2008(16): p. 758.
3. Sun, L., et al., *Capillary zone electrophoresis for bottom-up analysis of complex proteomes*. Proteomics, 2016. **16**(2): p. 188-196.
4. Perry, R.H., R.G. Cooks, and R.J. Noll, *Orbitrap mass spectrometry: Instrumentation, ion motion and applications*. Mass Spectrometry Reviews, 2008. **27**(6): p. 661-699.
5. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: A primer*. Mass Spectrometry Reviews, 1998. **17**(1): p. 1-35.
6. Kim, M.S. and A. Pandey, *Electron transfer dissociation mass spectrometry in proteomics*. Proteomics, 2012. **12**(4-5): p. 530-542.
7. Amunugama, R., et al., *Bottom-Up Mass Spectrometry-Based Proteomics as an Investigative Analytical Tool for Discovery and Quantification of Proteins in Biological Samples*. Advances in wound care, 2013. **2**(9): p. 549-557.
8. Bateman, A., et al., *UniProt: a worldwide hub of protein knowledge*. Nucleic Acids Research, 2019. **47**(D1): p. D506-D515.
9. Sidoli, S. and B.A. Garcia, *Middle-down proteomics: a still unexploited resource for chromatin biology*. Expert Review of Proteomics, 2017. **14**(7): p. 617-626.
10. Kellie, J.F., et al., *The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput*. Molecular Biosystems, 2010. **6**(9): p. 1532-1539.
11. Tran, J.C., et al., *Mapping intact protein isoforms in discovery mode using top-down proteomics*. Nature, 2011. **480**(7376): p. 254-U141.
12. Tsur, D., et al., *Identification of post-translational modifications via blind search of mass-spectra*. 2005 IEEE Computational Systems Bioinformatics Conference, Proceedings, 2005: p. 157-166.
13. Saveliev, S., et al., *Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis*. Nature Methods, 2013. **10**(11): p. i-ii.
14. Loo, J.A., H.R. Udseth, and R.D. Smith, *Peptide and Protein-Analysis by Electrospray Ionization Mass-Spectrometry and Capillary Electrophoresis Mass-Spectrometry*. Analytical Biochemistry, 1989. **179**(2): p. 404-412.
15. Mann, M. and M. Wilm, *Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags*. Analytical Chemistry, 1994. **66**(24): p. 4390-4399.
16. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. Rapid Communications in Mass Spectrometry, 2003. **17**(20): p. 2337-2342.
17. Roepstorff, P. and J. Fohlman, *Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides*. Biomedical Mass Spectrometry, 1984. **11**(11): p. 601-601.
18. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis*. Nature Methods, 2007. **4**(9): p. 709-712.
19. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-3567.
20. Tabb, D.L., *The SEQUEST family tree*. Journal of the American Society for Mass Spectrometry, 2015. **26**(11): p. 1814-1819.
21. Geer, L.Y., et al., *Open Mass Spectrometry Search Algorithm*. Journal of Proteome Research, 2004. **3**(5): p. 958-964.
22. Dorfer, V., et al., *MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra*. Journal of Proteome Research, 2014. **13**(8): p. 3679-3684.

23. Cox, J., et al., *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment*. Journal of Proteome Research, 2011. **10**(4): p. 1794-1805.
24. Seymour, S.L., et al., *A standardized framing for reporting protein identifications in mzIdentML 1.2*. Proteomics, 2014. **14**(21-22): p. 2389-2399.



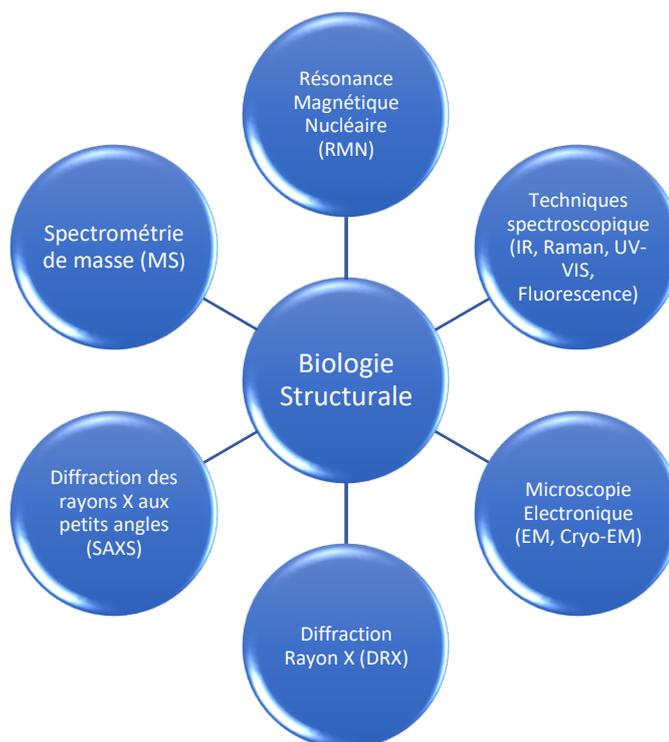
## Chapitre III : Caractérisation structurale de complexes biologiques non-covalents

Adopter une conformation spécifique est l'une des stratégies qui permettent aux protéines de reconnaître leurs partenaires d'interaction leur conférant ainsi leurs propriétés biologiques. Un certain nombre de repliements de protéines sont décrits dans la littérature, mais il est difficile de savoir si ce nombre est limité ou non. Au regard du nombre très élevé de chaînes polypeptidiques existantes dans la nature, caractériser ces repliements ainsi que les motifs structuraux associés et savoir les relier aux propriétés physiologiques des protéines est indispensable pour concevoir des stratégies spécifiques et efficaces de pronostiques, diagnostiques ou de thérapie lors de dysfonctionnements.

Au vu du nombre croissant de structures tridimensionnelles de protéines ou de complexes biologiques, une base de données structurales, la *Protein Data Bank* (PDB) est née. Cette base de données gratuite et accessible à la communauté répertoriait plus de 150000 structures en 2019. Aujourd'hui, *Research Collaboratory for Structural Bioinformatics PDB* (RCSB PDB) offre un accès ouvert aux connaissances accumulées sur la structure, la fonction et l'évolution 3D des macromolécules biologiques.

Pour nous aider à déterminer ces structures ou éléments de structures tridimensionnelles, un panel de techniques analytiques est à notre disposition. Celles-ci seront choisies en fonction de la problématique, mais aussi de la nature de l'échantillon (quantité, taille, homogénéité). Il nous a semblé intéressant d'en décrire quelques-unes dans ce chapitre et de montrer comment la MS y occupe désormais une place de choix. Comprendre comment ses techniques sont utilisées et quelles sont leurs limitations permettra de mieux imaginer comment les méthodologies basées sur la MS s'y intégreront dans une optique d'analyses structurales.

## 1. Les différentes techniques structurales



**Figure 16 : les différentes techniques d'analyse structurale.**

### 1.1 La RMN

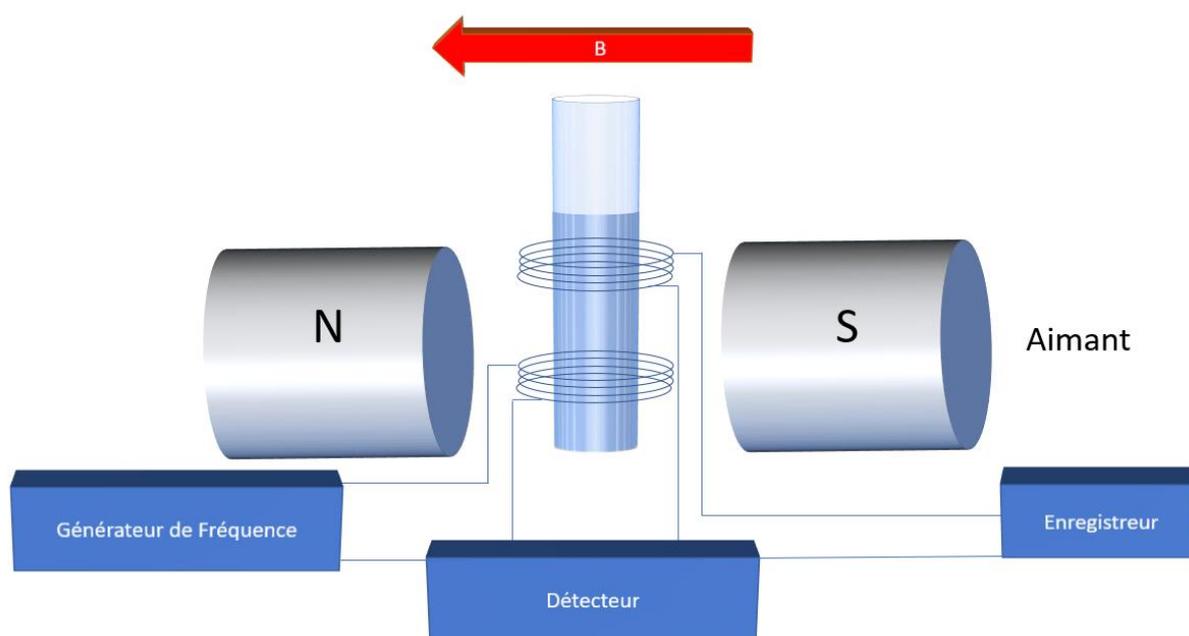
La résonance magnétique nucléaire (RMN) des molécules s'est très vite imposé comme une technique incontournable dans la caractérisation structurale des complexes biologiques.<sup>[1-3]</sup> Tout en étant non-destructive, elle permet d'analyser une molécule en solution. Cette méthode permet d'avoir une vision dynamique des arrangements conformationnels adoptés par les protéines par mesure des temps de relaxation, des temps de corrélation, des vitesses d'échange chimique. Le premier spectre d'une macromolécule, la ribonucléase A, fut enregistré en 1957, soit dix ans seulement après que la méthode eut été créée. Dès le début des années 1960, les scientifiques avaient déjà remarqué à partir des spectres RMN qu'au-delà de la composition chimique, la structure tridimensionnelle d'une macromolécule influence son spectre.<sup>[4]</sup> Toutefois, c'est au début des années 70 que la RMN acquiert sa puissance à la suite de deux percées majeures. La première fut la mise en place de la RMN « impulsionnelle par transformée de Fourier »<sup>[5]</sup> qui a permis de gagner en sensibilité et en rapidité, la seconde, la RMN multidimensionnelle.<sup>[6]</sup> Il a été ainsi possible à l'aide d'expériences de type NOESY de mesurer les distances entre les atomes d'hydrogène au sein d'une biomolécule<sup>[7]</sup>. La technique s'est ensuite étendue à l'étude de petites protéines à partir de 1985<sup>[8]</sup> puis a permis de montrer que les complexes biologiques étaient des objets dynamiques.<sup>[9]</sup> Kurt Wüthrich a reçu le prix Nobel de Chimie en 2002 pour le développement de la RMN pour la caractérisation structurale de protéine en solution.

### 1.1.1 Principe

La RMN est une technique qui utilise les propriétés magnétiques des noyaux des atomes dans le but d'obtenir des informations géométriques sur les molécules. A partir de ces informations, il est ainsi possible de reconstruire par le calcul une structure 3D.

Elle étudie les interactions entre spins dans une molécule. Elle permet ainsi d'observer deux types d'effets :

- Les interactions scalaires qui sont des couplages contrôlés par les électrons de liaison. Ces interactions permettent d'identifier et de reconstruire la topologie d'une molécule. Elles dépendent de la géométrie des liaisons et en particulier des angles entre les atomes.
- Les interactions dipolaires qui permettent d'identifier la distance entre les atomes et sont observables jusqu'à 0.5nm.



**Figure 17 : Principe de la RMN**

Afin de caractériser des complexes biologiques, les spectres contenant des corrélations des interactions scalaires et dipolaires sont enregistrés. A partir de ces spectres, les différents signaux sont traités et attribués individuellement à chaque atome de la molécule. Ensuite, l'ensemble des données géométriques (distance interatomique, angles dièdres...) obtenues sur la molécule est collecté<sup>[10]</sup> et une structure 3D compatible avec ces contraintes sera proposée. Pour cela, les calculs sont réalisés sur une conformation aléatoire de la molécule à laquelle les contraintes sont rajoutées progressivement pour induire le repliement de la molécule. Ce processus calculatoire est ensuite répété sur plusieurs conformations initiales différentes afin de vérifier que les calculs convergent vers une solution unique. Ainsi, en RMN, l'étude des macromolécules biologiques fournit non pas une structure, mais un ensemble de conformations probables.

### **1.1.2 Application à l'étude de composé biologique**

Les récents développements dans le marquage isotopique ainsi que l'amélioration de la technologie des aimants, de l'électronique et de la spectroscopie ont permis de repousser les limites de l'analyse biomoléculaire.

Pour les peptides ou petites protéines, il est possible de réaliser de la RMN à deux dimensions, sans marquage isotopique des molécules (COSY *CORrelation SpectroscopY*, NOESY *Nuclear Overhauser Effect SpectroscopY*). Pour les plus grosses protéines, on effectue des expériences TROSY (*Transverse Relaxation Optimized SpectroscopY*). Des méthodes de marquages isotopiques (échange proton/deutérium<sup>[11]</sup>, marquage carbone 13 et azote 15) sont également possibles. Des superpositions de signaux ainsi qu'un élargissement des bandes dus à une plus grande relaxation dipolaire rendent l'approche méthodologique assez longue et l'interprétation des spectres compliquée, ce qui limite encore la taille des protéines étudiées : environ 100kDa.<sup>[12, 13]</sup> L'utilisation de la RMN du solide<sup>[14]</sup> a également été développée dans le but de caractériser les multi-mères de grandes protéines solubles ou encore les protéines membranaires (« Magic-angle-spinning »<sup>[15, 16]</sup>).

Récemment, la RMN a été également appliquée au niveau cellulaire. La RMN cellulaire est une nouvelle application de la RMN à haute résolution qui permet d'analyser directement les caractéristiques structurales et dynamiques des protéines et d'autres macromolécules dans les cellules vivantes.<sup>[17]</sup>

En combinant les informations géométriques obtenues par cette technique et des calculs de structure obtenus par modélisation moléculaire, il est possible d'établir des modèles 3D à haute résolution.<sup>[18]</sup> Les interactions moléculaires (avec d'autres protéines, cofacteurs ou partenaires oligonucléotidiques) peuvent être cartographiées en utilisant les mêmes paramètres spectraux. Aujourd'hui, la RMN est également un outil majeur dans le domaine des protéines intrinsèquement désordonnées<sup>[19]</sup>, peu susceptibles de cristalliser ou encore pour l'étude de complexes transitoires. La dynamique des protéines sur une large échelle de temps est étudiée à l'aide de RMN en temps réel, de spectroscopie d'échange ou de mesures de relaxation.<sup>[20]</sup>

## **1.2 La diffraction par rayon X**

En 1914, le scientifique allemand Max von Laue postule que les rayons X en passant au travers d'un cristal seraient réfléchis par les centres atomiques du réseau cristallin et interféreraient entre eux pour créer un diagramme de diffraction et il obtient ainsi le prix Nobel de Physique. Depuis, cette nouvelle science appelée cristallographie des rayons X a été appliquée sur des molécules de plus en plus lourdes et complexes, voire sur des virus<sup>[21]</sup>. Avec l'amélioration de la technologie, de nouvelles structures sont décrites chaque année (<https://www.nature.com/subjects/x-ray-crystallography>). Actuellement, c'est la technique d'analyse structurale la plus résolutive avec une résolution pouvant être inférieure à 1Å, et de nouvelles sources de rayons X promettent des images interprétables à partir de cristaux toujours plus petits.

### 1.2.1 Principe

Cette technique est fondée sur la diffraction des rayons X par la matière quand celle-ci est sous forme cristalline.<sup>[22]</sup> L'état cristallin est caractérisé par une architecture périodique et ordonnée à l'échelle atomique, la maille élémentaire. La formation d'un cristal ou "cristallogénèse" est une étape critique et limitante de cette technique, car il faut passer d'un état liquide désordonné à un état solide ordonné, contrôlé par de nombreux paramètres comme : la température, la nature de l'agent précipitant, le temps de croissance et des lois cinétiques complexes. Aujourd'hui, cette étape est automatisable à l'aide de robot pouvant tester beaucoup de conditions de cristallogénèse.

Afin d'avoir une diffraction optimale, il faut avoir un rayonnement du même ordre de grandeur que la taille du réseau.<sup>[23]</sup>

Une autre source de rayonnement possible est le synchrotron. Il s'agit d'un accélérateur de particules à haute énergie élémentaire émettant un rayonnement électromagnétique continu dans une large gamme de longueurs d'onde allant des ultraviolets au rayon X. Cette source permet alors d'analyser les échantillons avec une résolution beaucoup plus fine ou de caractériser des cristaux aux dimensions très faibles.<sup>[24, 25]</sup>

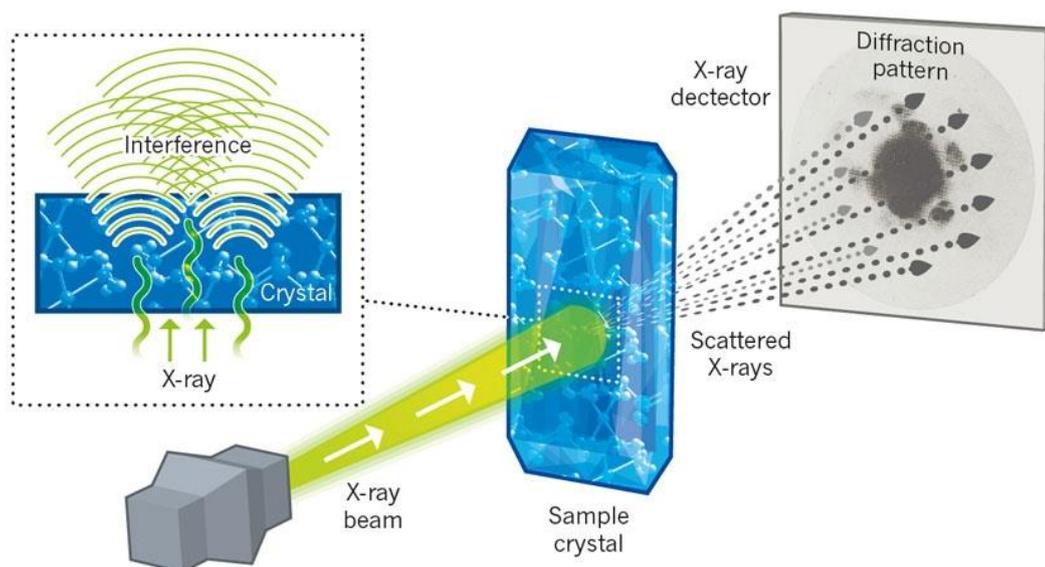


Figure 18 : Principe de la DRX. Extrait de [26]

Les rayons X produits par la source vont ainsi rencontrer un cristal qui va diffracter les rayons et les renvoyer dans des directions spécifiques déterminées par les dimensions et l'orientation du réseau cristallin ainsi que la longueur d'onde des rayons X. En mesurant les angles et l'intensité des rayons diffractés, il est ainsi possible d'obtenir des informations structurales telles que les dimensions de la maille cristalline, les symétries de la structure cristalline ainsi qu'une image de la densité électronique dans la maille. A partir de cette densité électronique, la position moyenne des atomes du cristal ainsi que la nature de ces atomes, leurs liaisons chimiques, leur agitation thermique seront déterminées.

### **1.2.2 Caractérisation structurale des composés biologiques**

La DRX a très vite été appliquée à la caractérisation structurale des macromolécules biologiques.<sup>[27, 28]</sup> En 1953, James Watson et Francis Crick ont ainsi réussi à déduire la structure à double hélice de l'ADN. Suite à cette découverte, des études sur des molécules de plus en plus complexes ont été réalisées. En 1957, il a été ainsi possible de déterminer la première structure cristallographique d'une protéine : la myoglobine.<sup>[29]</sup> L'observation des structures cristallines de la myoglobine et de l'hémoglobine ont permis de montrer qu'il était possible de relier la fonction des protéines à des motifs structuraux. De nombreuses structures d'enzymes dans les années 1970 et au-delà<sup>[30, 31]</sup>, révèlent les conformations des sites actifs permettant d'émettre des hypothèses quant aux mécanismes catalytiques, et constituent la base physique de l'enzymologie moléculaire. Cette technique est aujourd'hui appliquée à des protéines membranaires, des complexes à haut poids moléculaire comme des sous-unités ribosomiques incluant 2833 nucléotides d'ARN et 27 protéines avec une résolution atomique de 2,7 Å.<sup>[32-34]</sup>

Aujourd'hui, les détails atomiques offerts par la DRX permettent d'accéder à une meilleure compréhension mécanistique de la fonction des protéines. La résolution des structures de protéines cibles ainsi obtenues est également à l'origine des approches de conception rationnelle de ligand (« drug design »<sup>[35]</sup>). Toutefois, cette technique comporte quelques limitations telles qu'une consommation importante d'échantillon (plusieurs mg) et la nécessité d'obtenir des cristaux diffractants. De la qualité de ces cristaux dépendra en effet fortement la qualité d'interprétation des résultats<sup>[36]</sup>. De plus, le fait d'avoir des cristaux n'est pas un gage de refléter exactement la structure de la macromolécule en solution. En effet, la forme cristalline ne correspond peut-être pas à la conformation majoritaire !

### **1.3 Le dichroïsme circulaire**

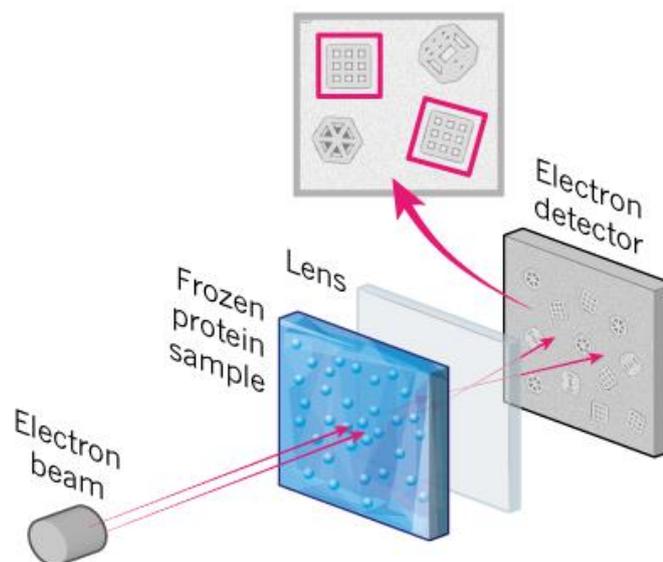
Le dichroïsme circulaire (CD) est une technique analytique qui repose sur l'observation des molécules chromophores qui sont actives optiquement soit par leur structure intrinsèque chirale soit parce qu'elles sont liées d'une manière covalente à un centre chiral d'une molécule<sup>[37]</sup>.

L'échantillon en solution est placé dans un polarimètre qui délivre alternativement une lumière polarisée circulairement droite et gauche. Le spectre CD est obtenu par soustraction de l'absorbance entre ces deux rayonnements en fonction de la longueur d'onde. Les protéines chromophores absorbent d'une part via la liaison peptidique (190-220nm), les acides aminés aromatiques (260 à 320nm) et d'autre part via les ponts disulfures (260nm).<sup>[38]</sup> Le CD permet d'obtenir des informations sur la proportion d'hélice- $\alpha$ , de feuillet- $\beta$ , de  $\beta$ -coude ou encore de structure désorganisée de la protéine (signature spectrale dans l'UV lointain) et sur l'environnement des acides aminés aromatiques (mesures dans l'UV proche).

Les limites du dichroïsme circulaire restent à ce jour sa faible résolution en comparaison avec les techniques citées précédemment. Mais il permet de faire des mesures rapidement (outil de cinétique à temps court), avec peu de matériel, et sur une large gamme de masse. De plus, la préparation de l'échantillon et le traitement des données sont simples et rapides. C'est une technique souvent utilisée pour contrôler la « non-dénaturation » des protéines recombinantes notamment. Il est également facile de déterminer si des changements conformationnels opèrent en faisant varier les conditions expérimentales (solvant, température, pH, salinité) ou sous l'ajout d'un partenaire d'interaction.

#### **1.4 Microscopie et Cryo-Microscopie électronique**

La microscopie électronique (EM) est une technique qui utilise les faisceaux d'électrons pour illuminer un échantillon et en créer une image agrandie. La résolution des microscopes dépend essentiellement de la longueur d'onde du rayonnement utilisé. Elle connaît depuis quelques années de grandes améliorations techniques et méthodologiques qui font d'elle une technologie de choix pour l'analyse structurale de biomolécules<sup>[39]</sup>. Un modèle 3D de la structure est alors obtenu après reconstruction tomographique des images 2D avec une résolution d'environ 5Å<sup>[40, 41]</sup>. Elle permet ainsi d'imager les échantillons biologiques de taille comprise entre la cellule et les protéines de taille supérieure à environ 150kDa. Le prix Nobel de Chimie a été attribué à Jaques Dubochet, Joachim Frank et Richard Henderson en 2017 pour le développement de la cryo-microscopie électronique<sup>[42]</sup>.



**Figure 19 : Principe de la CryoEM. Extrait et adapté de [43]**

### **1.4.1 Préparation des échantillons en EM :**

Il existe ainsi plusieurs stratégies pour préparer les échantillons pour l'analyse par EM. Les deux plus connus sont <sup>[44]</sup>:

- La coloration négative :

Il s'agit de la première méthode qui a été développée. Elle consiste à imprégner de substances (généralement des métaux lourds) les échantillons afin de les rendre plus opaques au faisceau d'électrons. La fixation de ces « colorants » permet ainsi d'améliorer le contraste négatif lors de l'analyse ce qui facilite l'observation des objets moléculaires.

La limitation de cette technique est que l'étape de fixation ou même des réactions ultérieures à la fixation peuvent induire des modifications de la structure de la protéine ciblée et ainsi donner des images correspondant à des structures artéfactuelles.

- La cryo-Microscopie :

La cryo-microscopie est une stratégie qui a été développée au début des années 1980 dans le but de réduire les dommages d'irradiation causés par le faisceau d'électrons sur l'échantillon. C'est une avancée majeure de cette méthode qui permet de s'affranchir de l'utilisation de colorant ou de fixateur chimique. Ici, l'échantillon est analysé à une température cryogénique (dans de l'éthane liquide à environ -195°C). Dans ces conditions de refroidissement rapide, l'eau n'a pas le temps de former des cristaux. Les protéines sont ainsi « figées » dans leur état natif dans une glace amorphe non cristalline. En congelant les échantillons et en les maintenant à basse température pendant toute la durée de l'analyse, les échantillons restent stables et peuvent être analysés beaucoup plus longtemps, car l'impact du faisceau d'électrons sur l'échantillon reste plus faible.

### **1.4.2 Construction du modèle structural :**

Après avoir effectué l'analyse, des milliers d'images de l'échantillon sont obtenues. Ces images sont des projections de l'échantillon et sont réalisées avec un contraste faible. Les images obtenues sous des angles identiques sont dans un premier temps additionnées afin de moyenniser le bruit de fond, améliorer le rapport signal sur bruit. Des images moyennées représentant les différentes vues de l'échantillon sous différents angles sont ainsi générées. Puis, ces images sont combinées à l'aide de modèles mathématiques dans le but de reconstruire la structure tridimensionnelle de l'objet. Le modèle 3D est ainsi obtenu après reconstruction tomographique des images de 2D enregistrées.

Avec l'arrivée des nouveaux détecteurs ainsi que les algorithmes de correction des mouvements, une résolution quasi atomique peut être obtenue<sup>[45]</sup>.

Les besoins en quantité de matériel biologique purifié sont moindres que pour la diffraction des rayons X ou la RMN. Toutefois, l'optimisation de la méthode est longue et difficile car le microscope enregistre les données d'un très grand nombre de particules orientées au hasard. Une limitation majeure de cette technique est qu'il est très difficile d'effectuer l'analyse de petits complexes dont la taille est inférieure à 150kDa.

## 1.5 La diffusion des rayons X aux petits angles (SAXS) :

La diffusion des rayons X aux petits angles, communément appelée SAXS, utilise des sources de radiation de type synchrotron. Cette technique s'appuie sur l'interaction des photons avec les électrons des atomes. Quand ils interagissent avec la macromolécule, les photons sont diffusés en fonction des densités électroniques. Plus l'atome est léger, moins l'interaction est forte. On mesure alors l'intensité des rayons X diffusés par l'échantillon en fonction de son angle de diffusion. Contrairement à la DRX, les mesures réalisées par SAXS sont effectuées sur des petits angles, généralement compris entre 0.1 et 5 degrés. Cette technique s'applique ainsi de la même manière que pour la DRX (la loi de Bragg).

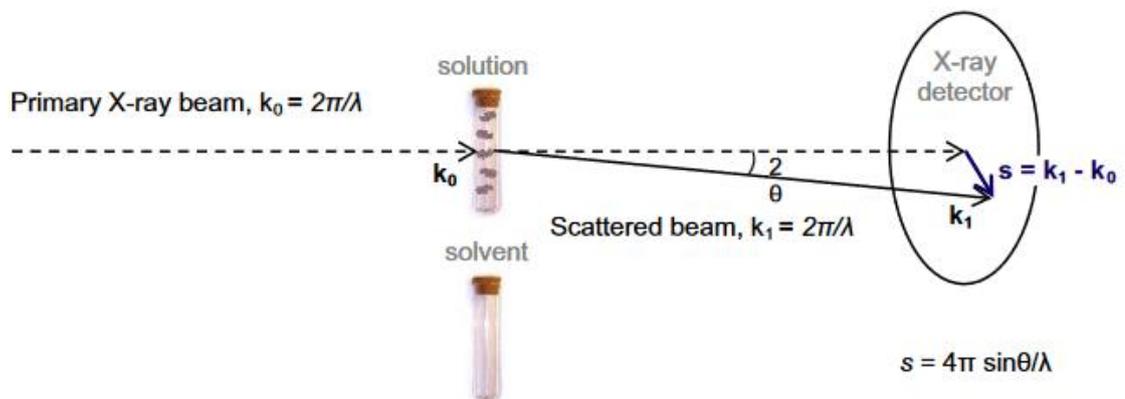


Figure 20 : Principe de la SAXS. Extrait de [46]

### 1.5.1 Principe

L'échantillon sous forme liquide est ainsi placé dans un capillaire en quartz. Les rayons X produits par la source vont ainsi rencontrer l'échantillon présent en solution et les diffracter en les renvoyant dans des directions spécifiques. L'interprétation de la courbe de diffusion, notamment la mesure des angles et l'intensité des rayons diffractés, permet d'extraire des informations structurales sur le complexe telles que sa masse, le rayon de giration ou le diamètre maximal. A partir de ces données, des informations décrivant la forme globale des protéines en solution ou l'enveloppe conformationnelle globale des complexes biologiques sont obtenues et des approches de modélisation moléculaire sont ensuite appliquées pour aboutir à la construction de quelques modèles structuraux les plus probables à basse résolution ( $10 \text{ \AA}$ )<sup>[46]</sup>. Des mesures sont réalisées en parallèle sur un échantillon témoin contenant uniquement le tampon d'analyse afin d'observer le motif de diffusion du solvant.

## **1.5.2 Application à l'étude structurale de protéines**

Tout comme la DRX la gamme de masse est de quelques kDa à plusieurs GDa, mais ici, la SAXS permet l'analyse des échantillons en solution.<sup>[46, 47]</sup> Toutefois, une grande quantité d'échantillons est requise (1 à 2 mg) et cette méthode est dépendante de la concentration : plus la concentration est élevée plus le ratio signal/bruit est élevé<sup>[46]</sup>.

Des informations précieuses sur l'état de dépliement/repliement des complexes biologiques, l'homogénéité et l'état d'oligomérisation des espèces en solution seront accessibles.<sup>[48, 49]</sup> Une limitation de la technique est qu'elle n'est pas adaptée pour l'étude de complexes impliquant l'interaction avec de petites molécules.

## **1.6 Les techniques spectroscopiques vibrationnelles**

La spectroscopie vibrationnelle a pendant longtemps été utilisée exclusivement pour la caractérisation de composés organiques. Aujourd'hui, les méthodes de spectroscopie vibrationnelle sont devenues des outils incontournables dans le domaine de la biologie.<sup>[50]</sup> Elles doivent en particulier leur succès aux avancées techniques qu'il y a eu dans les années 1970 avec le développement d'interféromètres pour la spectroscopie IR et la commercialisation de lasers pour la spectroscopie Raman.<sup>[51]</sup> Ces développements ont permis d'améliorer fortement la sensibilité.

La spectroscopie est une technique basée sur l'interaction de la matière avec une onde électromagnétique. Ainsi, en fonction de l'énergie des ondes électromagnétiques, différents phénomènes peuvent être observés (la diffusion, l'émission ou l'absorption de la lumière) et des méthodologies comme la spectroscopie UV-Vis<sup>[52]</sup>, la spectroscopie IR<sup>[53]</sup> et la spectroscopie Raman<sup>[54]</sup> en ont résulté.

### **1.6.1 La spectroscopie UV-VIS**

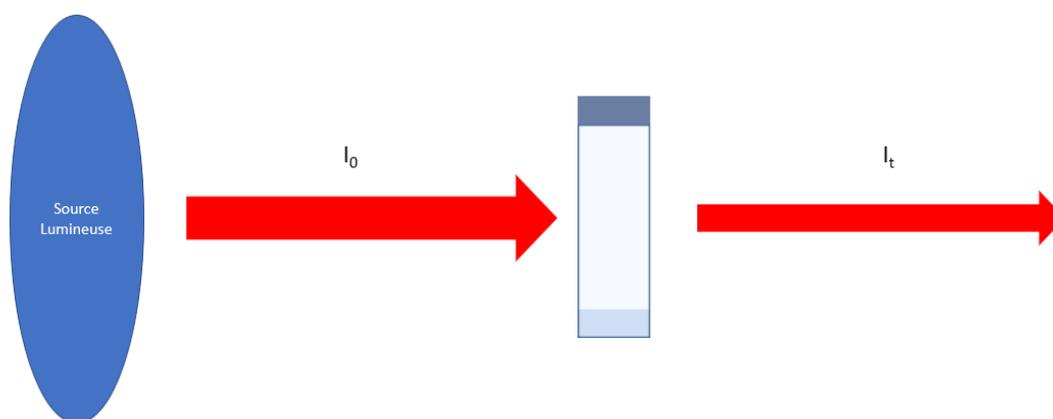
Dans un contexte biologique, la spectroscopie UV-Vis est surtout connue pour deux applications : la détection des composés élués en HPLC<sup>[55]</sup> et la détermination rapide de la concentration des protéines dans un échantillon. Le rayonnement UV émet entre les longueurs d'onde 200 à 400nm alors que la lumière visible émet à des longueurs d'onde entre 400 et 800nm.

#### **1.6.1.1 Principe**

Lorsqu'une molécule absorbe le rayonnement UV-Vis, l'énergie absorbée excite un électron dans une orbitale vide d'énergie supérieure. En spectroscopie UV-Vis, l'absorption d'énergie provoque des transitions qui se produisent entre les niveaux d'énergie électronique des électrons de valence, c'est-à-dire que les orbitales de plus basse énergie sont excitées vers des orbitales de plus hautes énergies. Les spectres UV-Vis sont alors tracés en enregistrant l'absorbance de l'énergie au sein de l'échantillon en fonction de la longueur d'onde de la lumière émise.

### 1.6.1.2 Applications de la spectroscopie UV-Vis à la caractérisation de complexes biologiques

Bien qu'elle remplisse un rôle de détection des composés organiques, cette technique peut s'avérer intéressante pour suivre des changements structuraux de protéines. En effet, la plupart des composés organiques qui absorbent en UV-Vis contiennent des liaisons  $\pi$  conjuguées. La forme du ou des pics et la longueur d'onde de l'absorbance maximale ( $\lambda_{\max}$ ) fournissent des informations sur la structure des composés et sur les propriétés de son environnement. Ainsi, les protéines contenant des acides aminés aromatiques tels que le tryptophane (Trp), la tyrosine (Tyr) et la phénylalanine (Phe) peuvent être caractérisées par la spectroscopie UV-Vis.<sup>[56, 57]</sup> Il s'agit également d'une technique analytique permettant la détermination de l'état d'oxydation d'un centre métallique d'un cofacteur comme par exemple un hème.<sup>[58, 59]</sup>



**Figure 21 : Transmission du rayon UV-VIS à travers un échantillon ( $I_0$  est l'intensité lumineuse initiale et  $I_t$  est l'intensité transmise)**

Des études réalisées en spectroscopie UV-Vis ont par ailleurs permis de montrer un changement conformationnel lors de la formation d'un complexe protéine-ligand entre l'actéoside, une molécule très utilisée dans les médicaments, et la trypsine.<sup>[60]</sup> La formation de ce complexe protéine-drogue a conduit à un relâchement conformationnel et au dépliement du squelette de la trypsine.

### **1.6.2 La spectroscopie infrarouge à transformée de Fourier (FTIR)**

Il s'agit d'une technique que nous avons été amenés à utiliser ponctuellement durant cette thèse, c'est pourquoi nous avons choisi de la présenter plus en détail.

La spectroscopie Infrarouge est une technique analytique classiquement utilisée pour la caractérisation structurale des petites molécules. Comme toute spectroscopie d'absorption, la FTIR va mesurer la quantité de lumière absorbée par l'échantillon en fonction de la longueur d'onde (dans le cas de l'IR allant de 0.8 à 1000 $\mu\text{m}$ , soit de 12500 à 10  $\text{cm}^{-1}$ ). Contrairement à la spectroscopie UV-Vis, un faisceau de lumière à large bande contenant l'ensemble du spectre des longueurs d'onde à mesurer est utilisé, ce qui nécessite en fin d'analyse un traitement des données par transformée de Fourier afin de convertir les données brutes en spectres exploitables. Un spectromètre FTIR peut également être utilisé pour obtenir des spectres d'émission, de photoconductivité ou de diffusion Raman.

Depuis les années 1990 <sup>[61]</sup>, la FTIR s'est fortement développée dans le domaine de la biologie.<sup>[62, 63]</sup> et constitue aujourd'hui un outil précieux pour caractériser la structure des protéines ou encore pour suivre les processus de réarrangement conformationnels, mettre en évidence un dépliement ou mauvais repliement des protéines.<sup>[53]</sup>

Un des exemples les plus marquants a été l'identification de certaines souches bactériennes.<sup>[64]</sup> A l'aide de cette technique, il a été possible de les différencier et de les classer.

### 1.6.2.1 Principe

Cette méthode consiste à envoyer sur l'échantillon un rayonnement électromagnétique dans la gamme de longueurs d'onde de l'infrarouge allant de 0.8 à 1000 $\mu\text{m}$ , soit de 12500 à 10  $\text{cm}^{-1}$ . Trois zones d'infrarouge peuvent ainsi être distinguées : l'infrarouge proche (12500-4000 $\text{cm}^{-1}$ ), l'infrarouge moyen (4000-400 $\text{cm}^{-1}$ ) et l'infrarouge lointain (400-10 $\text{cm}^{-1}$ ). Le champ électrique induit par l'onde va alors interagir avec le moment dipolaire d'un constituant moléculaire présent au sein de l'échantillon. L'interaction entre l'onde électromagnétique et l'échantillon va entraîner la vibration de certaines liaisons chimiques et l'absorption de l'énergie de l'onde excitatrice correspondante.

Un spectre infrarouge représente ainsi l'absorbance de l'énergie en fonction de la longueur d'onde  $\lambda$  exprimée généralement en nombre d'onde  $\nu$ . La longueur d'onde et le nombre d'onde sont reliés par l'équation suivante :

$$\nu = \frac{10^4}{\lambda} \text{ Où } \nu \text{ en } \text{cm}^{-1} \text{ et } \lambda \text{ en } \mu\text{m}.$$

De plus, selon la loi de Beer-Lambert, l'absorbance est directement liée à la composition de l'échantillon selon l'équation

$$A = \sum \varepsilon_i \cdot l \cdot c_i$$

Avec  $l$ , la longueur du trajet optique en cm,  $C_i$ , la concentration de l'espèce  $i$  exprimée en mol.  $\text{L}^{-1}$ ,  $\varepsilon_i$  le coefficient d'absorption molaire  $\text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$

Depuis les années 1970, la majorité des appareils utilisés en IR sont équipés d'interféromètres.<sup>[65]</sup> L'ajout d'interféromètres a permis d'améliorer la sensibilité de la méthode, mais également de réduire le temps d'analyse de quelques heures à quelques secondes. Il est désormais possible d'observer de manière simultanée l'ensemble des fréquences d'une zone spectrale définie contrairement aux anciens appareils qui utilisaient un monochromateur (l'absorbance était alors enregistrée nombre d'onde par nombre d'onde et ce, jusqu'à avoir balayé complètement la zone spectrale désirée). Le temps de balayage est ainsi beaucoup plus court pour une résolution donnée ce qui fait de la FTIR une technique d'analyse à haut débit. Avec le développement des interféromètres, il a également été possible d'augmenter le nombre de scans effectués afin d'améliorer le rapport signal sur bruit et ainsi la qualité des spectres. L'interféromètre le plus utilisé est l'interféromètre de Michelson<sup>[66]</sup> qui est constitué d'une lame séparatrice et de deux miroirs (un fixe et un mobile). Le signal de l'interféromètre sera ensuite converti en signal vibrationnel par transformée de Fourier.

### 1.6.2.2 Techniques d'échantillonnage

Selon la nature de l'échantillon à analyser, il existe différentes techniques d'échantillonnage.<sup>[67]</sup> Les techniques les plus couramment utilisées sont la transmission et la réflexion totale atténuée.

#### a. La transmission

Il s'agit du mode le plus rapide et le plus simple. Le principe consiste à mesurer l'énergie du faisceau IR après sa traversée de l'échantillon. Ainsi, la transmittance est le rapport entre l'intensité du rayon incident ( $I_0$ ) par rapport à l'intensité du rayon transmis ( $I$ ).

$$T = \frac{I}{I_0}$$

Et l'absorbance sera :

$$A = -\log T = \log\left(\frac{I_0}{I}\right)$$

Une des difficultés de ce mode réside dans le fait qu'il est nécessaire d'utiliser des supports transparents au rayonnement infrarouge.

#### b. La réflexion totale atténuée (ATR)

En réflexion, lorsque le faisceau lumineux arrive à l'interface d'un milieu dont son indice de réfraction est plus grand, l'onde subit une réflexion qui peut être soit totale, soit atténuée (une partie a été absorbée par le milieu). Pour les échantillons très absorbants tels que les molécules présentes dans des tampons aqueux, l'analyse des échantillons est exclusivement réalisée par réflexion totale atténuée (ATR).

Le principe de l'ATR est de faire en sorte que le rayon IR soit réfléchi plusieurs fois à l'interface entre l'échantillon et un cristal dont l'indice de réfraction est plus élevé que celui de l'échantillon. Le faisceau IR va traverser le cristal et subir une réflexion totale à l'interface entre le cristal et l'échantillon. Ce phénomène est généralement perturbé par la présence d'une onde

progressive appelée onde évanescente. Cette onde pénètre de quelques micromètres ( $1 \mu\text{m}$  à  $1650 \text{ cm}^{-1}$ ) dans l'échantillon qui est en contact avec le cristal et est donc absorbé. Une partie de l'énergie de l'onde est retenue et la réflexion totale de l'onde sera alors atténuée.

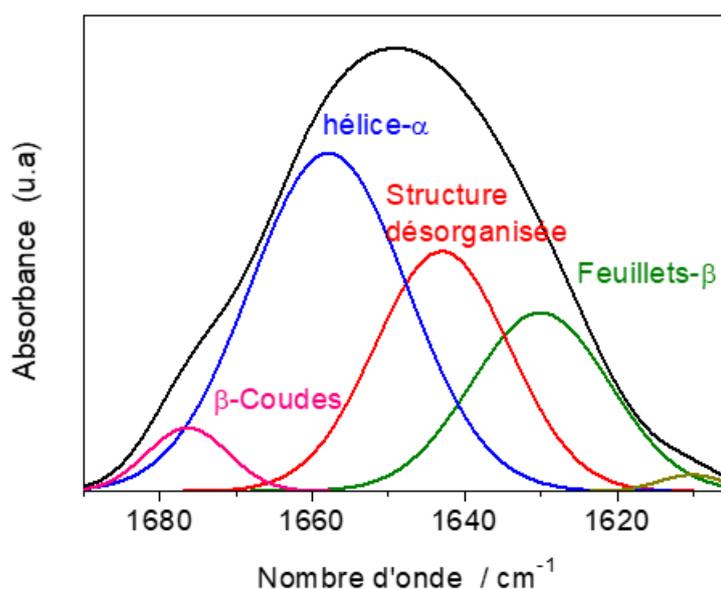
c. La spectroscopie IR appliquée à l'analyse structurale des complexes biologiques

Le spectre IR est constitué de 9 bandes dénommées les bandes amides. Celles-ci représentent les vibrations du squelette peptidique des protéines.

En 1990, il a été montré que les bandes amide I et amide II sont sensibles à la structure secondaire des protéines.<sup>[68]</sup> La bande amide I est retrouvée dans la région  $1700\text{-}1600 \text{ cm}^{-1}$  et correspond principalement aux vibrations d'élongation des liaisons C=O.<sup>[69]</sup> En fonction de la structure secondaire des protéines dans laquelle les groupements C=O sont impliqués, les propriétés spectrales de la bande amide I seront différentes.<sup>[53]</sup> Ainsi, chaque structure secondaire aura une position de bande différente. Par déconvolution de la bande amide I, il sera possible d'identifier et de caractériser chacune des structures secondaires constituant les protéines.

**Tableau 2 : Position de la bande amide I en fonction de la structure secondaire.**

Structure secondaire	Position de la bande dans H <sub>2</sub> O (cm <sup>-1</sup> )	Position de la bande dans D <sub>2</sub> O (cm <sup>-1</sup> )
Hélices $\alpha$	1657 - 1648	1660 - 1642
Feuillets $\beta$	1641 - 1623	1638 - 1615
	1695 - 1674	1694 - 1672
$\beta$ - Coudes	1686 - 1662	1691 - 1653
Structures désorganisées	1657 - 1642	1654 - 1639



**Figure 22 : Positions des structures secondaires au sein de la bande amide I**

Alors que la structure secondaire des protéines est visible en IR moyen, la structure tertiaire est visible en IR lointain qui décrit les mouvements collectifs de la protéine, dont les liaisons hydrogène. Généralement, les modes de vibrations des protéines et peptides visibles en IR lointain sont les amides IV, V, VI (800 à 500  $\text{cm}^{-1}$ ) ainsi que l'amide VII (vers les 200  $\text{cm}^{-1}$ ), et les liaisons hydrogène.<sup>[70]</sup>

Les avantages de l'IR sont une gamme d'applications allant de petites protéines solubles aux protéines membranaires, une résolution temporelle jusqu'à 1  $\mu\text{s}$  avec un effort modéré, souvent un temps de mesure court, une faible quantité d'échantillons requis (généralement 10 à 100  $\mu\text{g}$ ) et les coûts relativement faibles (spectromètres de première classe pour 40000 € ou \$). Ces avantages pour la recherche sur les protéines sont largement reconnus dans le monde académique, mais, de façon surprenante, la percée de la spectroscopie IR dans l'analyse commerciale des protéines reste à venir. Une des limitations de cette technique est qu'il faut travailler dans le  $\text{D}_2\text{O}$  pour pouvoir arriver à identifier efficacement les structures secondaires.

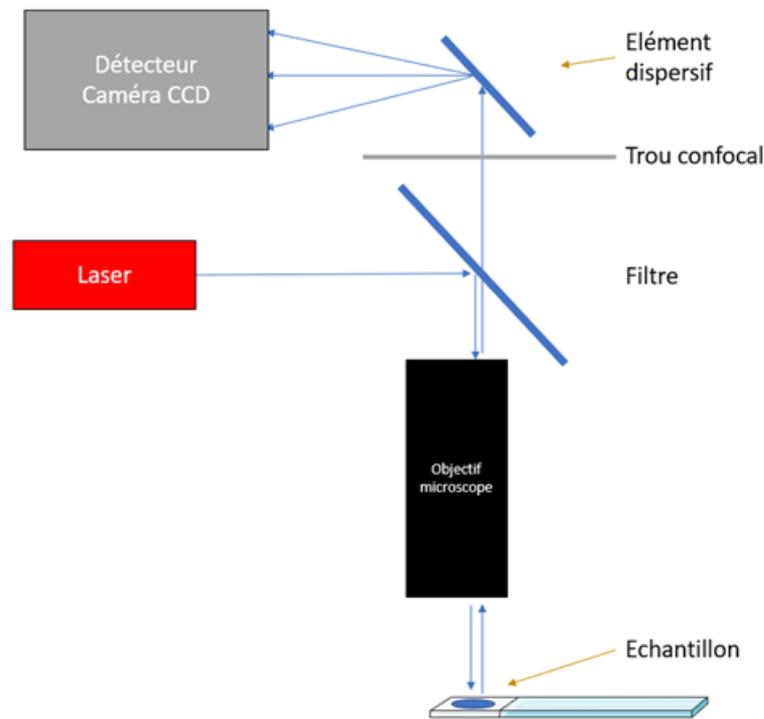
### **1.6.3 La spectroscopie Raman**

La spectroscopie Raman est une technique spectroscopique qui repose sur le principe de la diffusion de la lumière. Il s'agit d'une méthode non destructive et non invasive permettant l'analyse d'échantillon présent dans des milieux divers. Tout comme la spectroscopie IR, il s'agit d'une spectroscopie vibrationnelle. Ainsi, en collectant la lumière diffusée par un échantillon, on obtient un spectre Raman qui renseigne sur la structure des molécules.

Il s'agit également d'une technique qui a évolué grâce à l'arrivée des lasers dans les années 1970. Les lasers utilisés en Raman ont généralement des longueurs d'onde se situant entre l'ultraviolet et le proche infrarouge.

#### **1.6.3.1 Principe**

Les spectromètres Raman sont généralement couplés à un microscope. L'objectif du microscope est de focaliser le faisceau sur l'échantillon et ensuite de le diriger vers le spectrophotomètre. Afin de séparer les différents signaux et de ne détecter que le signal Raman, un filtre holographique est utilisé. Les photons diffusés sont ensuite envoyés vers un réseau de diffraction par un trou confocal afin qu'ils soient séparés en fonction de leur longueur d'onde. Le signal est par la suite détecté par une caméra CDD (charge coupled device) qui mesurera simultanément l'intensité des photons selon différentes longueurs d'onde.



**Figure 23 : Principe d'un spectrophotomètre Raman**

### 1.6.3.2 .La spectroscopie Raman appliquée à l'analyse des protéines

Comme pour la spectroscopie IR, l'analyse des protéines s'effectue par l'observation des vibrations du squelette peptidique qui seront observées sur les spectres Raman des protéines. <sup>[71]</sup> L'analyse de la structure secondaire des protéines s'effectue sur l'analyse des bandes amide I et amide III. Une des particularités de la spectroscopie Raman est qu'elle peut également détecter les vibrations provenant des chaînes latérales des résidus d'acides aminés permettant ainsi d'obtenir des informations sur l'état de protonation ou la coordination des acides aminés.<sup>[72]</sup>

Alors que l'analyse effectuée par spectroscopie IR repose sur l'existence de moments dipolaires à l'intérieur de la molécule, l'analyse effectuée par spectroscopie Raman s'effectue sur la polarisabilité des liaisons et sur la variabilité de celle-ci au cours des vibrations.

En conclusion, il existe énormément de techniques spectroscopiques visant à la caractérisation structurale de composés biologiques en solution. Bien que ces méthodes spectroscopiques permettent de mettre en évidence les interactions moléculaires et en particulier les modifications physiques des molécules, cette technique concède quelques défauts. Par exemple, pour utiliser la spectroscopie à fluorescence, il faut que les échantillons contiennent un chromophore dans au moins une des sous-unités du complexe.

## 1.7 La modélisation moléculaire

La modélisation moléculaire est un ensemble de techniques ayant pour but de modéliser ou de simuler le comportement des molécules. C'est une approche complémentaire des techniques physiques précédemment décrites. Cette démarche consiste à générer des milliers de repliements d'acides aminés probables (modèles structuraux) et d'en extraire les plus stables en prenant en compte les contraintes stériques, de solvation ... Les algorithmes de conception de ces modèles peuvent s'appuyer ou non sur des structures ou éléments de structures connus. L'objectif de cette approche est soit de pouvoir proposer des hypothèses de modèles structuraux ab initio (les visualiser, les manipuler et en calculer les paramètres géométriques ou énergétiques), soit de pouvoir comparer différents modèles structuraux (homologie de structures, simuler des changements conformationnels liés à un processus biologique particulier).

La méthode de modélisation la plus couramment utilisée en biologie structurale est la modélisation intégrative également appelée « docking » moléculaire.<sup>[73, 74]</sup> Elle repose sur le concept d'utilisation de données expérimentales (ou prédites) pour guider le processus de modélisation dans l'espoir d'échantillonner ou de sélectionner uniquement la partie significative de la zone d'interaction conformationnelle du complexe. On espère ainsi éviter d'effectuer des analyses structurales de manière exhaustive sur l'échantillon. Le docking moléculaire est notamment très utilisé dans la conception de médicament.

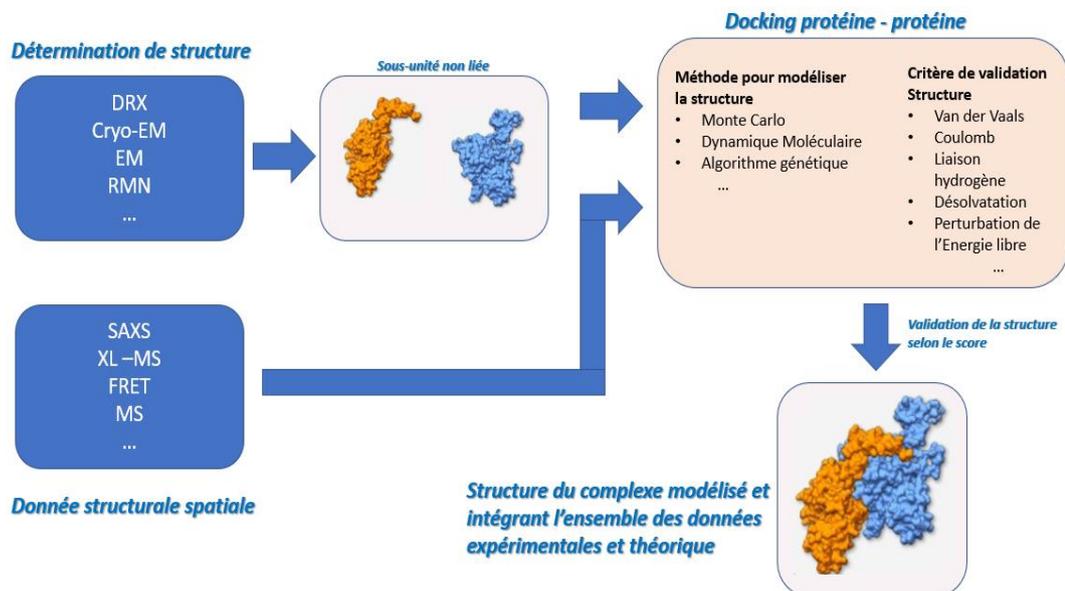


Figure 24 : Schéma récapitulant les différentes étapes de la caractérisation structurales des complexes protéiques<sup>[74]</sup>

La modélisation intégrative combine les données de plusieurs sources d'informations comme par exemple la carte de densité obtenue par Cryo-EM ou les contraintes de distances obtenues par XL-MS pour générer un modèle 3D du complexe. A partir des données expérimentales et des calculs de simulation, il est ainsi possible de générer des modèles 3D dans un temps très court. De nombreuses conformations seront ainsi proposées et la stabilité des modèles sera évaluée par des fonctions de scores<sup>[75]</sup>. Régulièrement, de nouvelles méthodes de calcul sont développées afin de sélectionner le « meilleur » conformère au sein d'un vaste ensemble de conformations potentielles.

Pour effectuer cette modélisation intégrative, il existe différentes méthodes. La plupart des méthodes partent des données expérimentales telles que le logiciel HADDOCK (High Ambiguity Driven DOCKing)<sup>[76]</sup>. Depuis, d'autres logiciels ont été développés et largement utilisés tels que les logiciels ATTRACT<sup>[77]</sup> et ROSETTA.<sup>[78]</sup> Le projet ROSETTA s'appuie sur l'hypothèse que l'information structurale est contenue localement et qu'ainsi, il est possible de construire un modèle structural en assemblant des petits morceaux de séquences protéiques ce qui réduit considérablement les degrés de liberté conformationnelle de la biomolécule<sup>[79]</sup>. Il est possible d'estimer la structure avec une précision pouvant aller jusqu'à quelques Å.<sup>[78]</sup>

Le champ d'application de ces approches théoriques est immense puisqu'à partir de la modélisation de la structure tridimensionnelle des molécules il sera possible d'accéder à diverses interprétations structurales : comparer les macromolécules entre elles selon certains critères, simuler des changements conformationnels ou des phénomènes biologiques, simuler des structures inconnues à partir de structures connues, jusqu'à peut-être concevoir des protéines « artificielles » aux propriétés thérapeutiques originales ...

## **1.8 Conclusion**

La détermination d'une structure tridimensionnelle fiable est aujourd'hui un enjeu majeur et la communauté scientifique se tourne vers une approche intégrative combinant des informations structurales obtenues par différentes techniques. Chaque technique rapporte son lot d'information et vise à compenser les limitations des autres. Par exemple, la DRX est une technique très puissante pour obtenir une structure à haute résolution, mais elle requiert une préparation parfois fastidieuse pour l'obtention de cristaux diffractants qui peuvent ne pas refléter l'image des molécules en solution. La RMN ou la FTIR étudient les biomolécules en solution et proposent un outil permettant d'accéder à la dynamique des arrangements conformationnels en solution, mais des contraintes de taille de molécules ou de quantité de matériel peuvent en limiter les applications. La cryo-EM est une technique en pleine expansion qui a gagné en résolution et permet parfois l'observation de plusieurs conformations d'une même biomolécule. Néanmoins, l'approche est longue et les traitements de données délicats.

C'est dans ce contexte que des méthodologies alternatives visant à compléter ces techniques et pallier leurs limitations sont développées. La spectrométrie de masse fait partie de ces techniques alternatives et grâce aux développements techniques et informatiques, plusieurs approches basées sur la détection des espèces par MS sont aujourd'hui appliquées à la caractérisation structurale de protéines ou de complexes biologiques<sup>[80]</sup>

**Tableau 3 : Tableau récapitulatif listant les intérêts et les limites des différentes techniques structurales pour l'étude de complexe biologique.**

	Méthodes	Préparation Échantillon	Intérêt	Limitation
<b>Echelle atomique</b>	RMN	Solution Solide	<ul style="list-style-type: none"> <li>- Étude de la dynamique</li> <li>- informations sur la structure (distance angles dièdres)</li> <li>- Identification de changement conformationnel global</li> </ul>	<ul style="list-style-type: none"> <li>- Taille des complexes biologiques étudiée (uniquement les complexes &lt;100kDa)</li> <li>- Concentration (entre 0,1 et 3mM)</li> <li>- Utilisation de marquage</li> </ul>
	DRX	Cristaux	<ul style="list-style-type: none"> <li>- Position des atomes</li> <li>- Résolution de l'ordre de l'angström</li> </ul>	<ul style="list-style-type: none"> <li>- Cristallogenèse</li> <li>- Sélection de la conformation cristallisée</li> <li>- Ne permet pas de visualiser les domaines non structurés</li> <li>- nécessite souvent de travailler avec des protéines recombinantes pour accéder aux quantités structurales</li> </ul>
	EM, Cryo-EM	Glace amorphe	<ul style="list-style-type: none"> <li>- Image structure à une résolution de 5 Angströms</li> <li>- Changement conformationnelle</li> </ul>	<ul style="list-style-type: none"> <li>- Difficultés à analyser les petits complexes &lt;150kDa</li> <li>- Ne permet pas de visualiser les domaines non structurés</li> <li>- traitement des données long</li> </ul>
<b>Echelle Moléculaire</b>	SAXS	Solution	<ul style="list-style-type: none"> <li>- enveloppe conformationnelle globale des complexes biologiques</li> <li>- Étude du dépliement/repliement des complexes biologiques</li> <li>- Étude de l'homogénéité et de l'état d'oligomérisation des espèces en solution</li> </ul>	<ul style="list-style-type: none"> <li>- Technique à basse résolution (10 Angström),</li> <li>- Grande quantité d'échantillons nécessaire (1 à 2mg)</li> <li>- Impossible d'étudier l'interaction des protéines avec les petites molécules</li> </ul>
	Spectroscopie IR/Raman	Solution Solide	<ul style="list-style-type: none"> <li>- Étude de la structure secondaire et tertiaire des protéines</li> <li>- Changement conformationnel</li> </ul>	<ul style="list-style-type: none"> <li>- Intolérante aux sels</li> <li>- Information sur un changement conformationnel global</li> <li>- Concentration (environ 1mM)</li> </ul>
	CD	Solution	<ul style="list-style-type: none"> <li>- Accès aux motifs structuraux</li> <li>- Étude des changements conformationnels</li> </ul>	<ul style="list-style-type: none"> <li>- Faible résolution</li> <li>- Intolérante à certains tampons (Tris, acétate...)</li> <li>- Pas d'accès à la structure tertiaire</li> </ul>
	MS (HDX-MS, XL-MS, IM-MS)	Solution Gaz	<ul style="list-style-type: none"> <li>- Dynamique conformationnelle (HDX-MS, IM-MS)</li> <li>- Stœchiométrie</li> <li>- Accès aux distances entre les acides aminés (XL-MS)</li> <li>- Détection simultanée de l'ensemble des espèces en solution et en mélange</li> </ul>	<ul style="list-style-type: none"> <li>- Intolérante aux sels</li> <li>- Pas d'accès direct à la structure</li> <li>- Analyse en phase gazeuse</li> </ul>

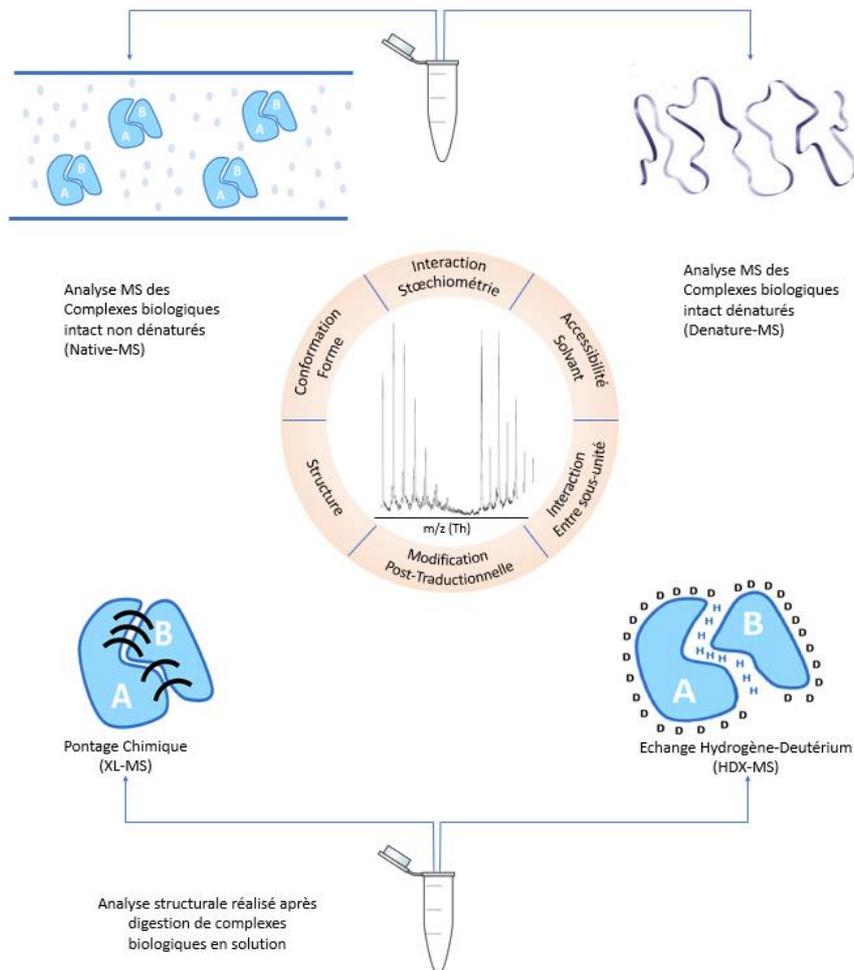
**Tableau 4 : Liste des techniques pouvant être utilisées pour déterminer différents types d'informations structurales.**

<b>Stœchiométrie</b>	MS, SAXS
<b>Dynamique conformationnelle</b>	RMN, SAXS, IR, Raman, CD, HDX-MS, IM-MS
<b>Structure 3D</b>	DRX (Ångström), RMN (Ångström), EM (10 Ångström)
<b>Contrainte de distance</b>	XL-MS

## 2. La spectrométrie de masse

Longtemps cantonnée au contrôle de l'intégrité de séquence des protéines, la MS démontre aujourd'hui un fort potentiel pour l'étude des structures secondaire, tertiaire et quaternaire<sup>[80]</sup>. Depuis quelques années, les interactions entre la MS et les autres techniques d'analyse structurales se sont renforcées, conduisant au développement de méthodologies intégratives. Les performances techniques de la MS en termes de sensibilité, gamme de masse accessible, rapidité ont largement contribué à ouvrir la MS à l'étude de systèmes biologiques natifs de plus en plus complexes et lourds<sup>[81]</sup> et plus récemment à l'étude des protéines membranaires<sup>[82]</sup>. Diverses méthodologies sont aujourd'hui susceptibles de nous aider dans un contexte d'analyse structurale : des approches visant l'analyse directe de protéines ou de complexes biologiques natifs comme la MS native, la mobilité ionique (IM-MS) ou des approches combinées à la méthodologie d'analyse protéomique comme le marquage chimique, la protéolyse ménagée, les échanges Hydrogène/Deutérium (HDX-MS) ou encore la stratégie du pontage chimique aussi appelée « cross-linking – MS » (XL-MS)<sup>[80]</sup> (figure 25).

Dans ce chapitre, nous allons décrire quelques-unes de ces approches basées sur la MS en nous focalisant plus précisément sur les approches par MS non dénaturante et par couplage au pontage chimique que nous avons été amenés à utiliser lors de ce travail de thèse.



**Figure 25 : Les différentes techniques complémentaires pour l'analyse structurale de complexe biologique adapté de <sup>[83]</sup>**

## **2.1 Spectrométrie de masse Native**

### **2.1.1 Introduction**

C'est en 1991 que le premier spectre de masse correspondant à un complexe biologique non-covalent a été observé, il s'agissait du complexe formé entre la myoglobine et son hème<sup>[84]</sup>. De façon surprenante, mais extrêmement intéressante, des espèces maintenues par interactions non-covalentes pouvaient être transférées intactes depuis la solution vers la phase gazeuse du spectromètre de masse. Ce résultat a révolutionné notre vision du champ d'application de la MS et a ouvert la voie à une nouvelle science dans un contexte de biologie structurale appelée la MS native<sup>[85]</sup> ou non dénaturante. Aujourd'hui, la MS non dénaturante a été appliquée à des complexes non-covalents de nature variée (protéine/protéine<sup>[86, 87]</sup>, protéine/ADN<sup>[88]</sup>, protéine/métal<sup>[89]</sup>, protéine/ligand<sup>[90]</sup>, de plus en plus lourds (détection de ribosome<sup>[91]</sup>, d'ions de virus<sup>[92]</sup>), de solubilité délicate (protéines membranaires<sup>[93]</sup>). C'est une

technique très performante pour la mise en évidence d'une affinité entre biomolécules ainsi que pour la détermination de la stœchiométrie d'interaction notamment.

### **2.1.2 La MS native requiert une préparation spécifique de l'échantillon :**

La MS native est une approche basée sur l'ionisation par électrospray. L'échantillon étant introduit en solution dans la source ESI, il est alors possible de choisir un solvant non dénaturant permettant de préserver la conformation des biomolécules en solution (pH et la force ionique contrôlée). Néanmoins, l'ESI-MS présente une faible tolérance aux sels non volatils tels que ceux souvent utilisés lors de la purification des complexes biologiques (NaCl, Na<sub>2</sub>HPO<sub>4</sub>, TRIS, HEPES). En effet, ces sels inorganiques interagissent de façon non spécifique avec les protéines menant à la formation d'adduits sur le spectre de masse. En résulte alors une grande hétérogénéité en masses moléculaires qui se traduit par un élargissement des pics sur le spectre de masse pouvant mener à une interprétation délicate. Une préparation d'échantillon spécifique comprenant une étape de dessalage des échantillons dans un tampon volatil est donc nécessaire en amont de l'analyse.

Pour le choix du tampon volatil, les sels d'ammonium se sont avérés un bon compromis car ils minimisent la formation d'adduits de molécules tampons qui nuiraient à une mesure de masse précise. Les plus couramment utilisés sont l'acétate d'ammonium (AcNH<sub>4</sub>, pH 6.8) ou le bicarbonate d'ammonium (NH<sub>4</sub>HCO<sub>3</sub>, pH 7.5) à des concentrations variant de quelques millimolaires jusqu'à 1 molaire dans certains cas<sup>[94]</sup>. Cette étape de dessalage est capitale pour le succès de l'analyse. Selon les complexes d'intérêt des dispositifs tels que la dialyse, l'ultrafiltration (système Vivaspin® de Millipore) ou encore des cartouches de phase d'exclusion stérique (exemple colonnes Biospin® de Biorad) seront choisies. Ces dernières sont plus rapides et limitent les phénomènes de dilution/concentration successifs susceptibles de déstabiliser certains complexes, mais elles sont moins adaptées pour les complexes de type protéine/ligand qui présentent des constantes cinétiques faibles. D'autre part, un compromis réside toujours entre un échange de tampon suffisant qui implique plusieurs cycles d'échange et une perte potentiellement non négligeable en matériel biologique. La technique d'échange de tampon est donc à adapter suivant les cas. Pour les systèmes biologiques dont la présence de sels inorganiques est impérative au maintien des interactions non-covalentes et qui ne peuvent pas être transférés intacts dans un tampon ammonium, l'analyse par MS native est impossible.

Pour l'analyse de complexes membranaires, il est également possible de rajouter un détergent en faible concentration micellaire critique (CMC).<sup>[95-97]</sup>

### **2.1.3 La MS native requiert un paramétrage instrumental spécifique :**

Pour pouvoir être observés par MS native, les complexes non-covalents doivent survivre à deux processus : i) le transfert des espèces en solution vers la phase gazeuse, ii) la traversée de la zone d'interface entre la source et l'analyseur.

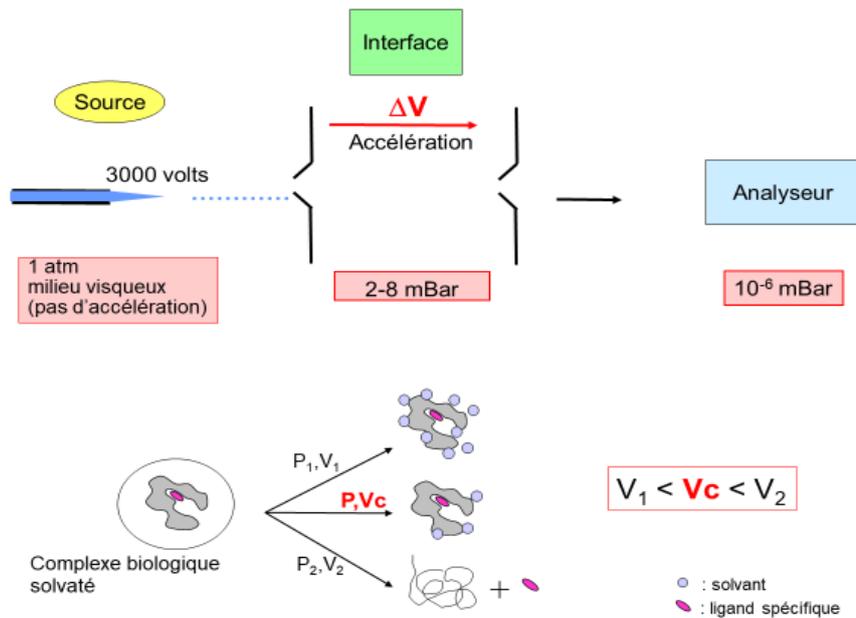
Plusieurs travaux ont montré que les conformations protéiques pouvaient être transférées intactes en phase gazeuse<sup>[98-100]</sup>. Néanmoins, des réarrangements de structure sont susceptibles de survenir lors du passage de la phase condensée à la phase gazeuse. Au cours du processus de désolvatation des ions, des molécules d'eau sont éliminées, ce qui induit un renforcement des interactions électrostatiques et les liaisons hydrogène, tandis que les interactions hydrophobes sont affaiblies et les liaisons de Van der Waals restent inchangées. Par conséquent, si les interactions hydrophobes jouent un rôle prépondérant dans le maintien des contacts intra- ou intermoléculaires, il sera probablement difficile d'observer les complexes intacts lors de l'analyse en ESI-MS<sup>[101, 102]</sup>.

D'autre part, lors du transfert de la source à l'analyseur, l'énergie interne communiquée aux ions peut être suffisamment importante pour induire la dissociation de notre complexe<sup>[103]</sup>. Il conviendra ainsi de contrôler précisément :

- *La température du gaz de séchage* : elle sera adaptée en fonction du solvant utilisé. Pour les échantillons biologiques hydrosolubles, cette température est de l'ordre de 140°C - 180°C et joue peu d'influence sur le maintien de l'édifice, sauf dans le cas des duplexes d'ADN où elle doit être réduite pour ne pas les dissocier<sup>[104, 105]</sup>.

- *L'énergie de collision* avec les molécules de gaz résiduelles dans l'interface (zone de moyenne pression 2-3 mbar) : Les collisions contre les molécules de gaz sont utiles car elles favorisent la désolvatation des ions en les dissociant de leurs adduits de solvant ou de sels (c'est particulièrement important dans le cas de nos complexes biologiques non-covalents qui sont solubilisés dans des tampons aqueux). La tension d'accélération (In-Source Dissociation sur le microTOF Q-II que nous avons utilisé) est la différence entre le potentiel modulable à la sortie du capillaire de transmission et le potentiel du premier skimmer. Elle induit un champ électrique qui accélère les ions dans l'interface. L'énergie cinétique résultante est ensuite convertie en énergie de changement d'état ou chaleur latente lors des collisions avec les molécules de gaz environnantes, provoquant l'évaporation du solvant. Si cette tension d'accélération n'est pas suffisante, l'évaporation ne sera que partielle, et ainsi l'intensité du signal considérablement diminuée. Néanmoins, une intensité trop élevée peut aboutir à des collisions trop énergétiques qui peuvent induire la dissociation des interactions non-covalentes. Ainsi, un bon compromis sera à trouver pour chaque système non-covalent afin d'obtenir une désolvatation efficace sans pour autant provoquer la dissociation des complexes (Cf figure 26).

- *La pression dans l'interface* : lorsque la pression augmente, le libre parcours moyen des ions diminue. Ainsi ils vont subir davantage de collisions, mais, leur accélération étant limitée, ces collisions resteront de faibles énergies et donc peu efficaces, ce qui réduit le risque de dissociation<sup>[106]</sup>. Ce phénomène, appelé processus de thermalisation par collisions « douces », entraîne une dissipation progressive de l'énergie cinétique excluant un échauffement brutal des ions macromoléculaires. Il a été exploité par certains constructeurs de spectromètres de masse (notamment des instruments Waters), qui ont introduit au niveau de l'interface une vanne permettant d'augmenter la pression (5-8 mbar) pour favoriser la transmission des ions de haut rapport  $m/z$ .<sup>[107, 108]</sup> Ceci ne s'est pas é véré nécessaire avec les configurations de source à capillaire de transfert telle que celle installée sur notre Micro-TOFQII.

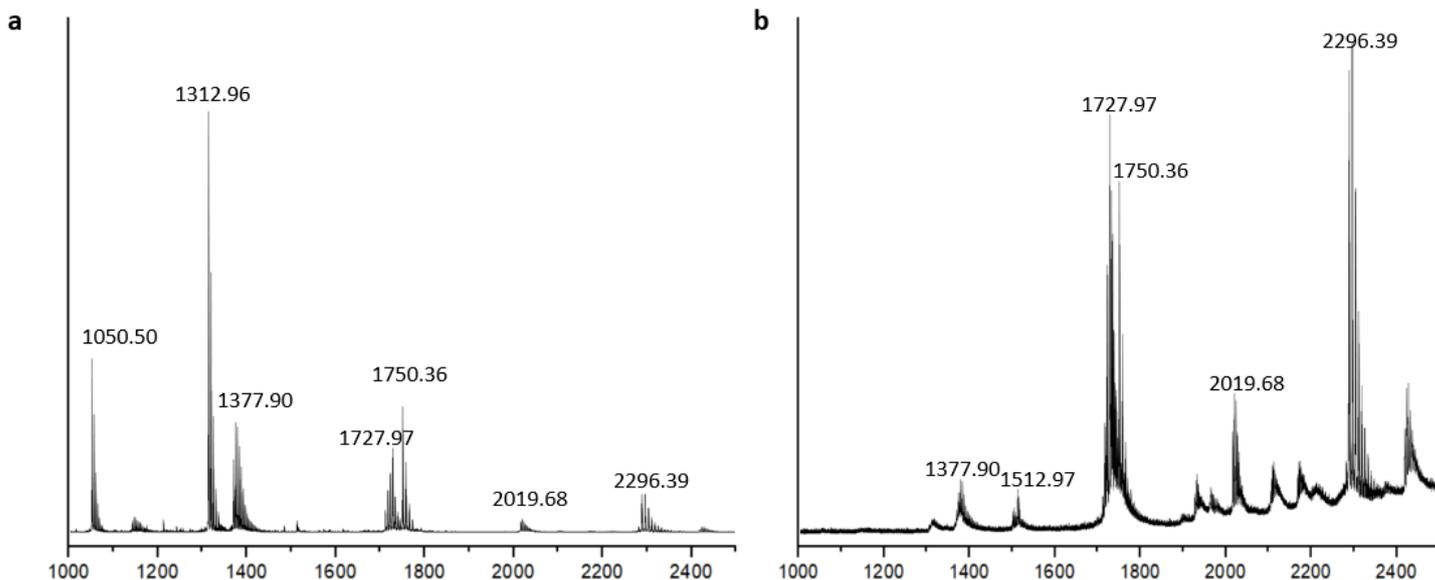


**Figure 26 : schéma explicatif du fonctionnement de la tension d'accélération.**

Par conséquent, il est important d'optimiser ces paramètres pour chaque complexe biologique étudié afin de trouver un bon compromis entre une bonne désolvatation des ions, et un maintien des interactions non-covalentes lors de l'analyse en MS. Cette optimisation dépendra de la nature et du nombre d'interactions non covalentes qui maintiennent l'intégrité de la conformation de la protéine ou du complexe biologique.

La possibilité qu'a l'opérateur de pouvoir moduler l'énergie interne des ions dans l'interface peut présenter un atout du point de vue structural. En effet, la dissociation en phase gazeuse peut être provoquée intentionnellement dans le but d'obtenir des informations sur la nature et la force des interactions mises en jeu au sein du complexe. Ainsi, à l'instar des IC50 (50% inhibition concentration), la valeur d'énergie induisant 50% de dissociation peut être un outil pour comparer la stabilité en phase gazeuse de plusieurs complexes de type protéine-ligand.<sup>[109, 110]</sup>

Enfin, pour pouvoir observer des complexes biologiques de grande masse, il faut également optimiser des paramètres de stockage (« prepulse storage ») et de transmission (« transfer time ») ainsi que les amplitudes de voltage appliquées sur les guides d'ions des multipôles dans l'interface. En effet, les ions ayant des rapports  $m/z$  très éloignés vont épouser des trajectoires différentes dans le spectromètre de masse. Les complexes protéiques dans leur état natif sont moins chargés que leurs homologues dénaturés, nécessitant des tensions de fonctionnement plus élevées pour une convergence efficace<sup>[111]</sup>. Ainsi, en présence d'un mélange d'espèces de rapports  $m/z$  très différents, il conviendra de faire varier ces paramètres afin de ne pas discriminer certains ions au profit d'autres. Il est important de détecter l'existence de toutes les espèces coexistant et de trouver un bon compromis pour la meilleure représentativité de l'échantillon en solution.

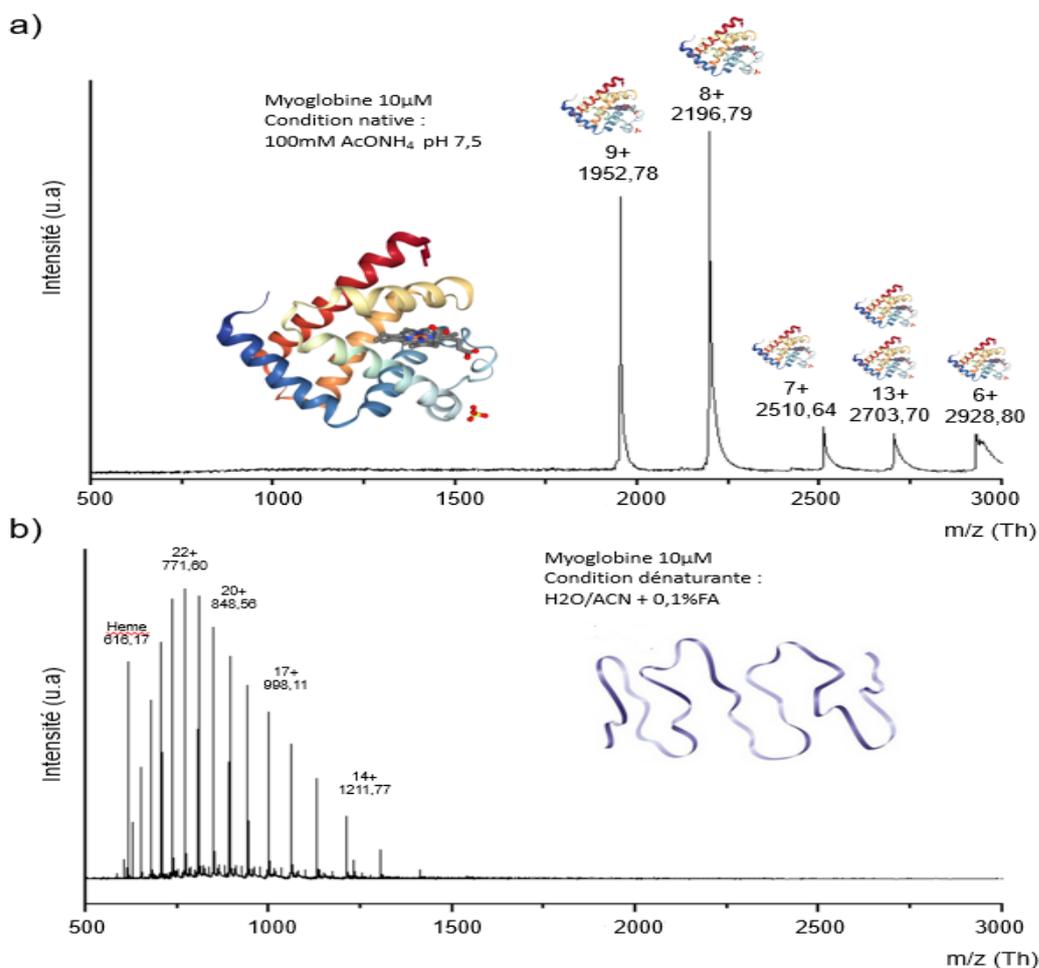


**Figure 27 : Influence du Prepulse sur les spectres MS. (a) Prepulse = 30  $\mu$ s (b) Prepulse = 40  $\mu$ s. Discrimination de l'ion 1312.96 par augmentation du paramètre de prepulse lors de l'étude de l'interaction entre NCBD et TAD.**

La figure 27 montre un exemple de cette observation préférentielle en fonction du paramétrage instrumental. Avec un temps de transmission plus faible, on favorise la détection des espèces à bas  $m/z$ , tandis qu'avec un temps de transmission plus élevé, on observe préférentiellement des espèces à haut  $m/z$ .

L'optimisation de ces paramètres est pointue et il n'est pas toujours évident de les définir afin de refléter au mieux les espèces présentes en solution.

### 2.1.4 Une analyse en MS native possède une signature spectrale :



**Figure 28 : Comparaison spectre ESI-MS de masse de la myoglobine en conditions natives (a) et en conditions dénaturantes (b)**

La figure 28 représente les spectres de masse d'un même échantillon (ici un complexe non-covalent entre la myoglobine et son hème) analysé par MS native (a) et par MS dénaturante (b). Il en ressort une nette différence dans la distribution des états de charge, cette dernière étant plus étroite et décalée vers les rapports m/z élevés en MS native. Ceci est lié à la structure des ions en phase gazeuse : (i) une structure repliée limite le nombre de sites accessibles à la protonation et donc réduit le nombre de charges global porté par l'ion (m/z plus élevé), (ii) la spécificité et l'homogénéité de la conformation du complexe non-covalent limite le nombre de structures en phase gazeuse et donc le nombre d'états de charge (distribution étroite). Ces observations sont récurrentes quels que soient les systèmes biologiques considérés et constituent une empreinte spectrale en MS native.

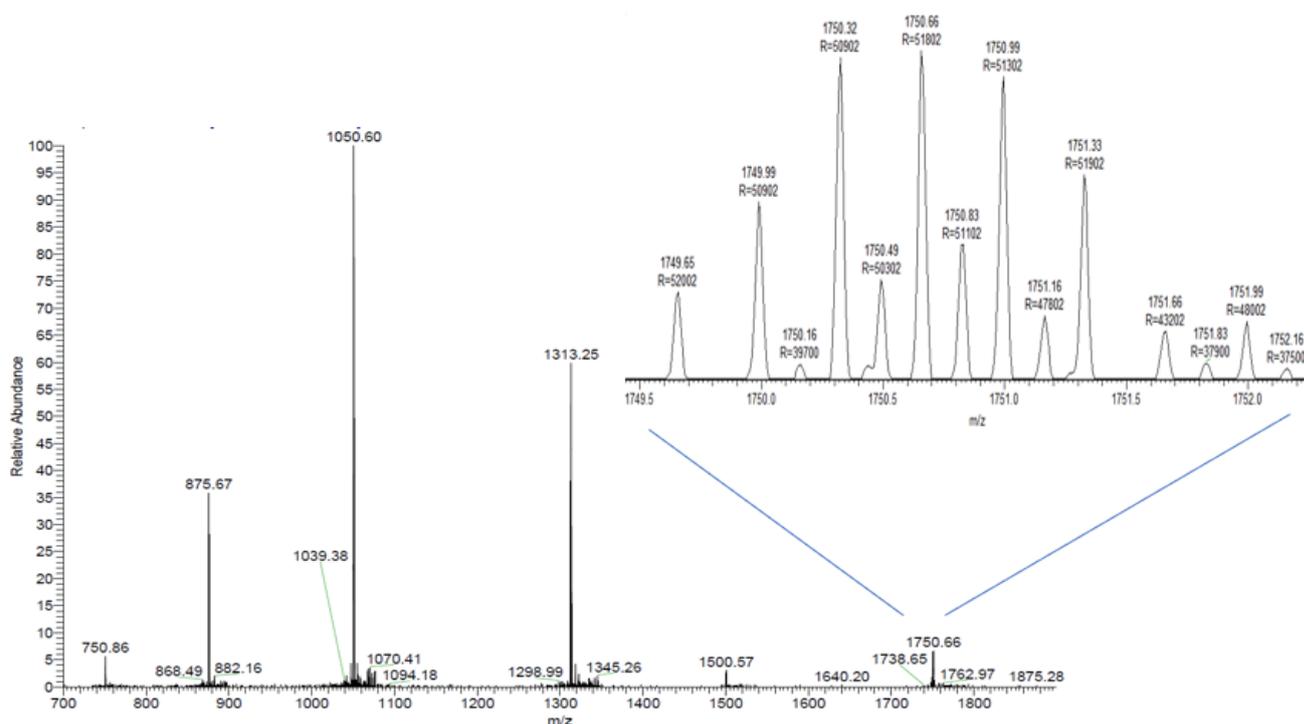
### **2.1.5 La MS native requiert-elle des instruments à haute résolution ?**

La précision dans la détermination de la masse moléculaire est liée à la résolution du spectromètre de masse<sup>[112]</sup>. Cette dernière est optimisée afin d'être capable de séparer 2 espèces de rapports m/z toujours plus « proches ».

En 1<sup>ière</sup> intention, la MS native est utilisée pour mettre en évidence l'existence d'un complexe non-covalent, donc l'affinité d'une biomolécule pour une autre molécule, ainsi que la stœchiométrie d'interaction. Dans de nombreux cas, la différence de masse entre les partenaires d'interaction est suffisante pour discriminer sans ambiguïté les ions correspondants aux espèces libres de ceux correspondants aux espèces complexées, et ceci, sur des instruments que l'on peut qualifier de standard en termes de résolution (ESI-Q ou ESI-TOF). Néanmoins, dès lors que l'on travaille sur des systèmes où la complexation moléculaire induit une faible augmentation de la masse (cela peut être le cas des complexes de type protéine/métal ou lors de l'interaction entre un petit ligand et un complexe multiprotéique) ou encore lorsqu'il s'agira de séparer les différents états de charge d'un complexe multiprotéique de grande taille<sup>[92, 113]</sup>, la résolution instrumentale peut s'avérer très limitante pour le succès de l'analyse.

Intrinsèquement, grâce à des logiciels de simulation de profils isotopiques, il est possible de connaître la résolution nécessaire à l'observation non ambiguë des différentes espèces co-existantes dans notre échantillon. En pratique, il faut tenir compte de l'élargissement des signaux résultant d'une désolvatation incomplète des ions. Des algorithmes améliorant l'assignation des états de charges pour les complexes de très grande taille sont également décrits dans la littérature.<sup>[114]</sup>

Aujourd'hui, des instruments à haute résolution sont développés dans un contexte d'analyse par MS native comme le SynaptG2 (Waters) ou l'orbitrap Exactive plus EMR (ThermoFisher) qui permettent d'atteindre une résolution de 40000 à 200m/z et 140000 à 200m/z respectivement.



**Figure 29 : Spectre de masse à haute résolution de la protéine NCBP sur une Orbitrap Qexactive plus EMR mettant en évidence la superposition des profils isotopiques d'une espèce monomérique et dimérique à m/z 1750.66**

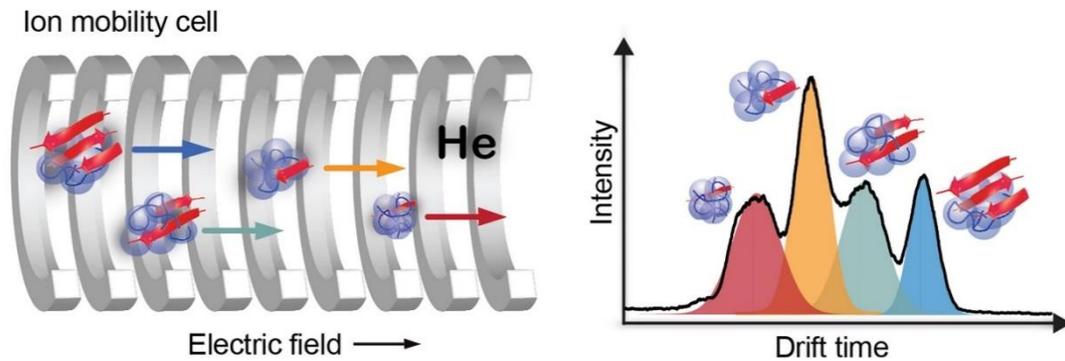
Ainsi, disposer d'instruments à haute résolution en MS native n'est une nécessité que dans certains cas.

## **2.2 La mobilité ionique**

Le couplage de la mobilité ionique (IM-MS) à la MS ouvre le champ d'application de la MS native à l'étude de la conformation des édifices non-covalents. L'IM-MS permet d'accéder aux valeurs de sections efficaces des ions non-covalents issus de la source ESI. Ces CCS correspondent à l'aire projetée en 2 dimensions des ions en phase gazeuse<sup>[115]</sup> et nous aiderons à mettre en évidence une éventuelle hétérogénéité conformationnelle dans l'échantillon<sup>[116]</sup> (coexistence de plusieurs conformères), ou encore des changements conformationnels qui pourraient s'opérer lors d'une modification de l'environnement (changement du pH, ajout d'un partenaire d'interaction notamment)<sup>[101]</sup> Étant donné que les valeurs théoriques du CCS peuvent être calculées à partir de modèles structuraux de biomolécules, la CCS déterminée expérimentalement peut également être utilisée en retour pour aider à éliminer les modèles les moins probables<sup>[117]</sup>.

Cette méthode consiste à mesurer le temps nécessaire mis par les ions pour traverser un gradient électrique faible dans une chambre remplie de gaz inerte. Ce temps se nomme « le temps de dérive » et dépend non seulement de la masse et de la charge du composé, mais également de sa forme. Ainsi les ions de structures dépliées entrent en collision plus fréquemment avec le gaz neutre, ce qui entrave leur progression et augmente ainsi leur temps de dérive par rapport à des ions plus compacts.<sup>[118]</sup> Comme le nombre de collisions est

proportionnel à la surface de la protéine, la mesure du temps de dérive peut être utilisée pour déterminer la valeur moyenne de rotation de la CCS, indiquant la forme tridimensionnelle de la protéine.<sup>[119]</sup>



**Figure 30 : Schéma d'une cellule de mobilité ionique. Extrait de [120]**

Différentes cellules de mobilité peuvent être couplées à la MS :

Les plus connues sont le tube de dérive (DT) et l'onde progressive (TW). En DT, un champ électrique linéaire homogène est utilisé et les valeurs de CCS sont déterminées directement à partir du temps de dérive mesuré et des conditions expérimentales appliquées. Avec la méthode d'onde progressive, des ondes de potentiel se propagent continuellement à travers le tube de dérive. La manipulation de la fréquence et la hauteur des ondes progressives permettent la séparation des ions. Bien que cette méthode améliore le rapport cyclique, les valeurs absolues de CCS ne peuvent pas être déterminées et une procédure d'étalonnage indirect est nécessaire. À ce jour, les cellules DT et TW IM-MS couplés à des instruments de type QTOF sont disponibles dans le commerce, et les applications de la technique aux études structurales se développent rapidement.

Il est à noter que les cellules de type FAIMS (field-asymmetric ion mobility spectrometry) dont le principe repose sur l'application d'un champ électrique intense ne peuvent pas être utilisées dans un cadre structural, mais uniquement comme dimension séparative supplémentaire (notamment lors d'analyses protéomiques d'échantillons très complexes).<sup>[121]</sup>

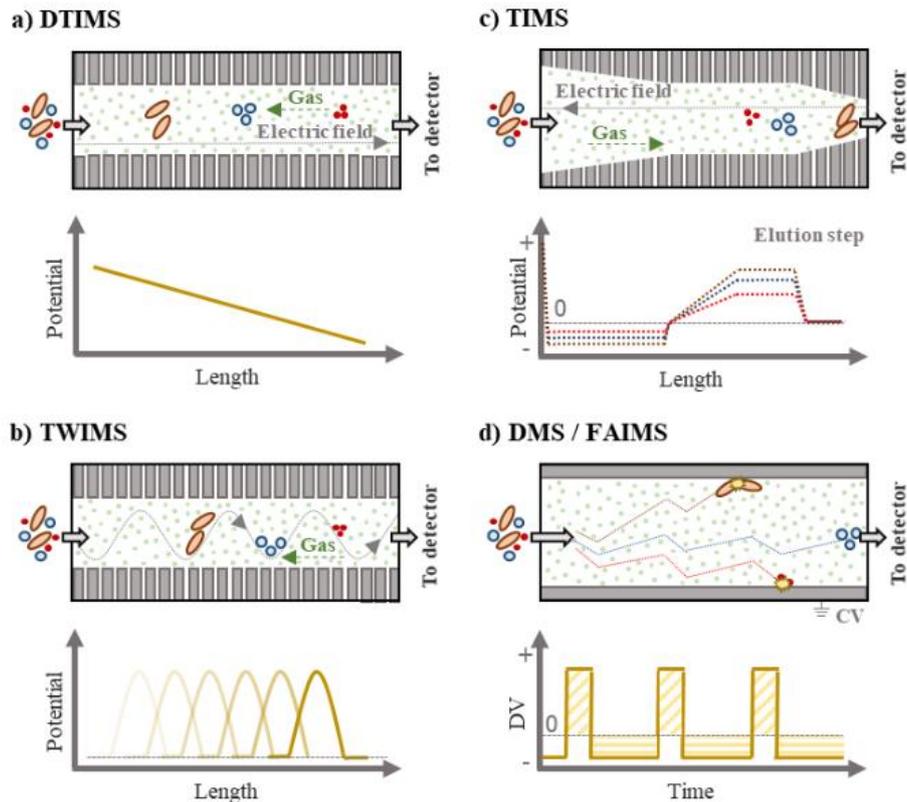


Figure 31 : Schéma des différentes cellules de mobilité ionique. Extrait [122].

#### Applications de l'IM-MS en MS structurale :

De nombreuses revues ont montré le potentiel de l'IM-MS à fournir des informations structurales des édifices non-covalentes.<sup>[123, 124]</sup> De plus, l'IM-MS a également permis la séparation de molécule isomérique dans leur état intact.<sup>[125]</sup> Une autre application de l'IM-MS se concentre sur la réduction de la complexité des spectres de masse, en particulier dans les situations où les ions de fond sont en abondance relative élevée par rapport à l'ion d'intérêt.<sup>[126]</sup>

L'IM-MS est particulièrement intéressante pour les systèmes où les sous-unités génèrent différents multimères qui ne peuvent être séparés uniquement par le rapport  $m/z$ , comme pour la dynamique d'oligomérisation de l'amyloïde  $\beta$ <sup>[127]</sup> ou encore celle de la  $\beta$ 2-microglobuline.<sup>[128]</sup>

D'autre part, d'un point de vue fondamental, l'IM-MS est également un moyen de montrer une certaine rétention de structure tertiaire des ions en phase gazeuse, ce qui a longtemps été et reste un sujet de débat aujourd'hui.<sup>[129]</sup>

L'IM-MS présente néanmoins quelques limitations. Il a été observé que des complexes de protéines possédant une cavité centrale peuvent s'effondrer pendant l'analyse en IM-MS menant à des ions de structures plus compactes donc possédant des CCS inférieures à la conformation native<sup>[130]</sup>. De plus, les informations structurales sont déduites indirectement des valeurs de CCS par des simulations de dynamique moléculaire qui peuvent être délicates dans le cas de grands édifices moléculaires<sup>[131]</sup>. Enfin, l'utilisation de cette technique implique l'achat d'un appareil dédié.

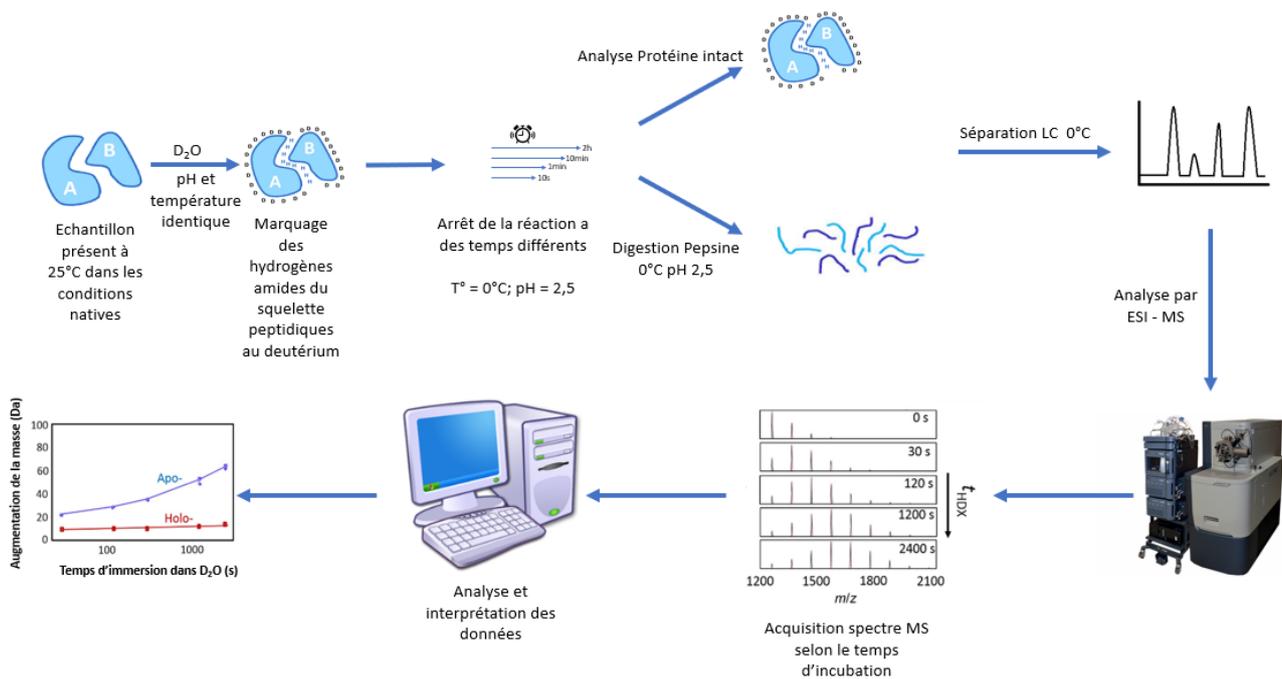
### **2.3 Spectrométrie de masse couplée à l'échange proton/deutérium (HDX-MS)**

Cette approche permet de sonder l'accessibilité des hydrogènes des liaisons amides du squelette peptidique à un solvant deutéré. Le profil d'incorporation des deutériums sera interprété pour fournir des informations structurales en solution, en particulier sur les interfaces en interaction ou sur les éventuels changements conformationnels <sup>[132]</sup>. Les premières mesures de protéines effectuées par HDX remontent aux années 1950 <sup>[133]</sup>. Celle-ci a été surtout utilisée dans les années 1970-1980 en couplage avec la spectroscopie RMN comme méthode de détection et de caractérisation. <sup>[134, 135]</sup>

Couplée à la MS<sup>[136, 137]</sup> et en utilisant les outils de l'analyse protéomique, il est possible de comparer les profils d'incorporation en deutérium de chaque peptide en fonction du temps de marquage et des conditions expérimentales (suivi des profils de dénaturation, impact de l'ajout d'un partenaire d'interaction). La HDX-MS est aujourd'hui appliquée à des complexes biologiques relativement lourds et avec des quantités d'échantillons beaucoup plus faibles (de l'ordre du  $\mu\text{M}$ )<sup>[138-141]</sup>. Il s'agit d'une technique également utilisée dans les industries car elle excelle dans la cartographie des épitopes et dans la caractérisation des produits. <sup>[142]</sup>

#### **Le taux d'incorporation en deutérium dépend de la structure tridimensionnelle :**

Les hydrogènes exposés à la surface accessible au solvant et qui ne sont pas impliqués dans des liaisons hydrogène peuvent s'échanger avec les deutériums du solvant de manière rapide. Ce processus entraîne alors un incrément de masse au niveau de la protéine qui pourra être déterminée par MS et par suivi au cours du temps. <sup>[143]</sup> En revanche, les hydrogènes impliqués dans des liaisons hydrogène intra- ou intermoléculaires et les hydrogènes enfouis au cœur de la structure des édifices macromoléculaires vont s'échanger plus lentement. De même, les zones stables et en interaction ont une vitesse d'échange plus faible que les régions flexibles et libres. Ainsi les vitesses d'échange isotopique reflètent la structure tridimensionnelle, les changements conformationnels et les interfaces de liaison des molécules biologiques et leurs complexes.



**Figure 32 : Représentation schématique de la caractérisation structurale de complexe biologique par la spectrométrie de masse en couplage avec une stratégie d'échange hydrogène/Deutérium (HDX-MS)**

La HDX-MS requiert un protocole expérimental pointu :

Afin d'obtenir des informations structurales pertinentes, le protocole expérimental d'échange se doit d'être parfaitement reproductible<sup>[144]</sup> et aujourd'hui, c'est aussi par le développement d'une formule automatisée de préparation d'échantillon que la technique a gagné en crédibilité (<https://www.leaptec.com/products/hdx-workflow-solution><sup>[145]</sup>). Deux techniques de marquage peuvent être utilisées. La première méthode est le marquage continu. Il s'agit de la méthode la plus utilisée et consiste à diluer la protéine native dans un tampon deutéré<sup>[146]</sup>. Plusieurs temps de marquage sont effectués afin d'identifier les conditions pour lesquelles le taux d'incorporation sera informatif. La seconde méthode est le marquage pulsé. Il s'agit d'une technique qui consiste à marquer l'échantillon dans des conditions non natives pour une durée définie. Une modification du pH et/ou de température, ou encore l'ajout d'agents dénaturants sont réalisés afin de faciliter le marquage, et ce, pendant des temps variables allant de quelques secondes à quelques heures. Puis, l'incorporation de deutérium est enregistrée en diluant la protéine dans un tampon deutéré. Cette approche permet ainsi d'identifier plus rapidement certaines zones protéiques qui ont été dépliées dans les conditions dénaturantes. Ce type de marquage permet d'appréhender les mécanismes de dépliement/repliement des biomolécules.<sup>[147]</sup>

Une fois l'échange abouti vient l'étape importante de « quenching » pour conserver des taux de deutération les plus proches possible de ceux obtenus à la fin du marquage. Etant donné que les vitesses d'échange dépendent du pH et de la température, de manière générale le « quenching » est réalisé en diminuant le pH jusqu'aux alentours de 2,5 et la température jusqu'à 0°C puis une approche d'analyse protéomique « Bottom-up » permettra d'accéder au taux d'incorporation en deutérium de chaque peptide voire de chaque AA accessible<sup>[148]</sup>. Il est impératif de maintenir l'échantillon en condition acide et à faible température jusqu'à l'introduction dans le spectromètre de masse. Ceci requiert un dispositif d'analyse dédié et l'adaptation du protocole standard de protéolyse. C'est pourquoi les enzymes les plus couramment utilisées en protéomique telles que la chymotrypsine ou la trypsine sont ici inadaptées. L'enzyme la plus couramment utilisée en HDX-MS est la pepsine.<sup>[149]</sup> Il s'agit d'une enzyme peu spécifique, mais qui travaille à pH acide et qui donne des résultats très reproductibles.

D'autre part, la séparation des peptides de digestion par LC doit également être optimisée (température et temps d'analyse rapide (environ une quinzaine de minutes). Un analyseur à haute résolution<sup>[150]</sup> ou encore le couplage à la mobilité ionique sont également des atouts pour déterminer avec précision le taux de deutération du peptide analysé.<sup>[151]</sup>

La HDX-MS est ainsi une approche permettant la caractérisation rapide et efficace de zones d'interaction entre les protéines d'un même complexe.<sup>[152]</sup> Cette technique permet également de monitorer le dépliement/repliement de certains des complexes protéiques.

Une des difficultés de cette méthode reste la mise en place d'un protocole robuste limitant le risque de réversibilité de l'échange hydrogène-deutérium.

#### **2.4 Le marquage chimique couplé à la MS**

En plus de l'échange proton/deutérium, d'autres isotopes peuvent être utilisés afin d'étudier la topologie de surface des complexes biologiques. Cette approche consiste à utiliser des agents de marquage qui réagissent sur des acides aminés cibles afin de marquer la protéine d'intérêt. Cette approche permet ainsi d'étudier les changements conformationnels liés à la formation du complexe qui va modifier l'accessibilité et la réactivité des sites. Une analyse différentielle en MS après protéolyse permettra alors d'identifier efficacement les résidus impliqués dans l'interaction.

Le marquage chimique ne doit pas affecter la structure de la protéine ni ses interactions avec ses partenaires. C'est pourquoi le choix de l'agent marquant ainsi que les conditions expérimentales doivent être au préalable optimisés.<sup>[153]</sup>

#### **2.5 La protéolyse ménagée**

La protéolyse ménagée consiste à digérer partiellement l'échantillon dans des conditions non dénaturantes pour mettre en évidence des zones résistantes à l'hydrolyse et en déduire des informations structurales.

Le principe part du postulat suivant : En présence d'une protéase, les régions de la protéine qui sont enfouies, qui font partie d'une structure rigide, ou qui sont impliquées dans les interactions avec des partenaires sont protégées contre la protéolyse. Les régions non structurées, flexibles ou qui sont exposées au solvant sont moins protégées et plus susceptibles d'être digérées.

Les régions résistantes aux protéases sont identifiées grâce aux outils de l'analyse protéomique par nanoLC-MS/MS.

Cette approche a été utilisée avec succès afin d'obtenir des informations sur la flexibilité d'une protéine en solution, sur des interfaces d'interactions entre des sous-unités de complexes de type protéine-protéine<sup>[154]</sup>, protéine-ADN<sup>[155]</sup> ou encore protéine-ligand.<sup>[156]</sup> ou encore pour mettre en évidence des changements conformationnels ayant lieu lors des interactions avec divers partenaires.<sup>[155, 156]</sup> Elle peut être réalisée avec différentes protéases telles que la trypsine, la thermolysine, la Glu-C, la Lys-C, Asp-N, la chymotrypsine...le choix de la protéase est très souvent dépendant des édifices étudiés. La cinétique de protéolyse est généralement suivie à différents temps d'incubation et peut être étudiée soit par MALDI-MS ou par ESI-MS<sup>[155]</sup>

## **2.6 Spectrométrie de masse couplée à la technique de pontage chimique (XL-MS)**

Le pontage chimique (ou cross-linking ou XL), consiste à lier de façon covalente deux résidus d'acides aminés appartenant à des séquences peptidiques intramoléculaires ou à des sous-unités qui sont proches dans l'espace par l'intermédiaire d'un agent de pontage (ou agent réticulant, « cross-linker ») afin de stabiliser la structure tridimensionnelle de notre système biologique à l'état natif<sup>[157]</sup>. La réaction de pontage peut être réalisée sur des protéines, des peptides, mais également sur des acides nucléiques.<sup>[158, 159]</sup> L'identification des sites pontés s'effectuera par MS en utilisant les outils de l'analyse protéomique (XL-MS).

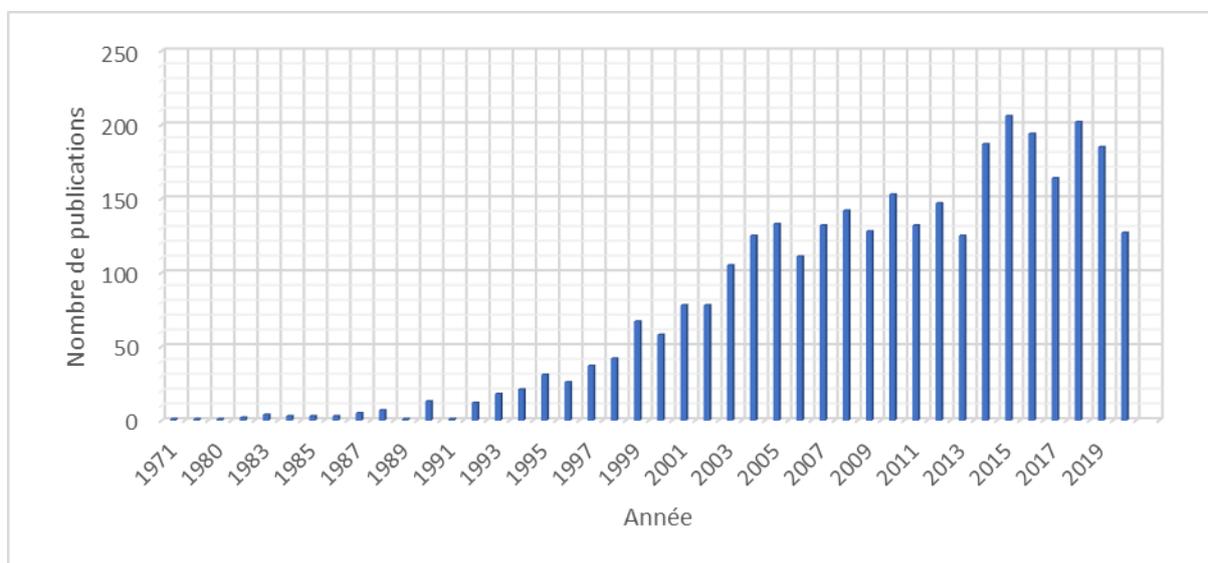
Ainsi cette méthodologie aura deux applications majeures : l'immobilisation d'un état conformationnel lié aux conditions environnementales, ainsi que l'étude de la topologie des interfaces entre les différentes sous-unités constituant le complexe d'intérêt. La détermination de la topologie des interfaces, que l'on pourrait définir comme la détermination des points de contact entre les différentes sous-unités sera possible grâce à l'utilisation d'agent de pontage possédant un bras espaceur de longueur courte et fixe. Ainsi seuls deux acides aminés très proches (quelques Angströms) pourront être liés par une molécule de pontage.



**Figure 33 : Représentation Schématique de la réaction de pontage**

Après réaction avec le cross-linker, il sera possible de localiser ces points de contact grâce à une approche protéomique Bottom-up, après digestion enzymatique. De ces données seront établies des cartes de distances intermoléculaires qui permettront de rapidement mettre en évidence des « nœuds » d'interactions, c'est-à-dire des zones où les partenaires protéiques sont en proche contact, mais aussi l'apparition ou la disparition de nouvelles zones de contact selon les conditions expérimentales. En donnant l'accès à des informations telles que la topologie, la proximité spatiale ou encore les contraintes de distance des groupes fonctionnels, le couplage XL-MS s'intègre dans l'éventail des approches d'analyse structurale.

Depuis une dizaine d'années, grâce au développement instrumental avec l'apparition des spectromètres de masse de plus en plus performants, au développement de nouveaux agents de pontage, et au développement bio-informatique, l'essor de cette approche s'illustre par la publication de revues toujours plus nombreuses. En septembre 2020, une simple recherche sur la base de documentation PubMed avec les mots-clés « cross-linking mass spectrometry » (« pontage couplé à la spectrométrie de masse ») montre une croissance exponentielle du nombre d'études scientifiques utilisant le XL-MS au fil des années.



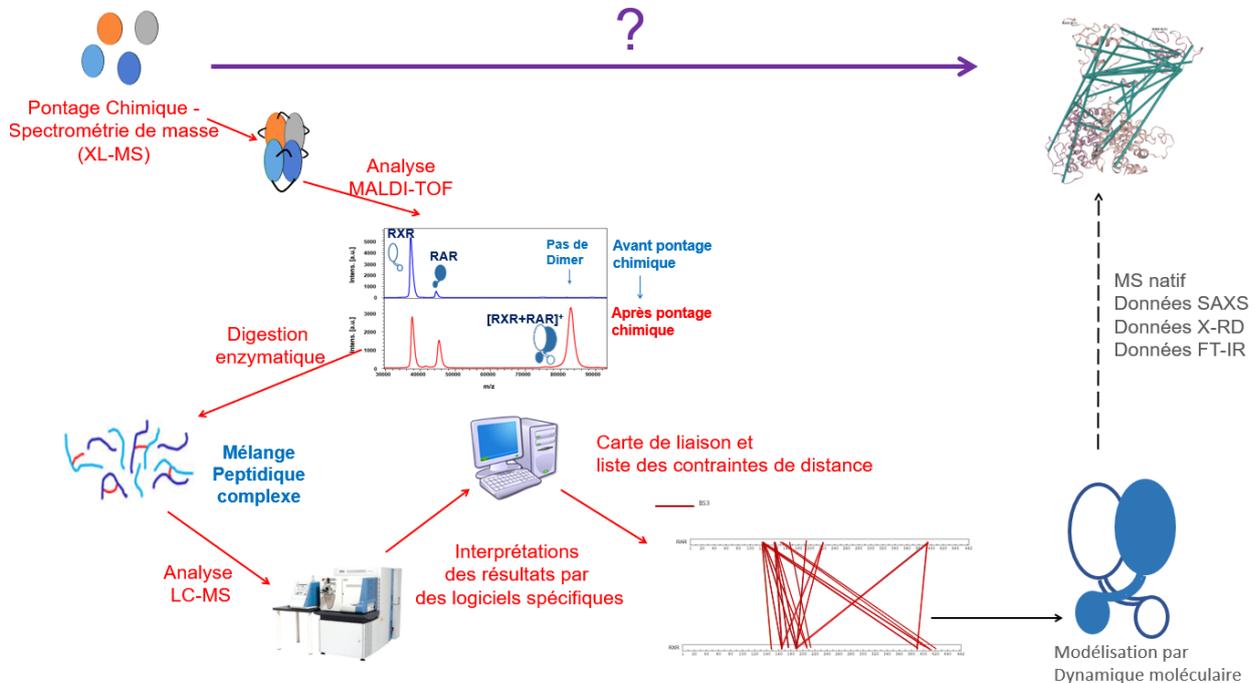
**Figure 34 : Nombre d'articles scientifiques publiés chaque année sur l'approche XL-MS et répertoriés dans la base PubMed (recherche effectuée en septembre 2020).**

Parmi celles-ci, nous pouvons citer quelques exemples d'application XL-MS sur de larges édifices macromoléculaires<sup>[160]</sup> visant à obtenir de nouvelles perspectives sur l'interactome.<sup>[161-163]</sup> ou encore sur les interactions des protéines *in vivo*.<sup>[164]</sup>

Cette approche, à basse résolution, est souvent utilisée en combinaison avec d'autres techniques structurales dans le but d'obtenir une structure à meilleure résolution des protéines. La XL-MS a notamment été utilisée en combinaison avec la cryo-EM afin de caractériser des complexes très gros tels que les complexes ribosomiques, les ARN polymérases...<sup>[165]</sup>

L'approche XL-MS s'effectue selon 3 étapes majeures :

- La réaction de pontage
- La production de peptides pontés analysés par MS
- L'identification des peptides pontés.



**Figure 35 : Représentation schématique de la caractérisation structurale de complexes biologiques par la spectrométrie de masse en couplage avec une stratégie de pontage chimique (XL-MS)**

A mon arrivée au laboratoire, la définition du protocole expérimental à partir des conditions décrites dans la littérature tant concernant le choix de l'agent de pontage, que du choix du logiciel de traitement des données [166-168], suggérait autant d'espoir que d'interrogations. La mise en place d'un protocole robuste et fiable constitue encore aujourd'hui une source de débat et un défi analytique.

Challenge analytique :

- Lié à la réaction de pontage : celle-ci devra être optimisée systématiquement, pour chaque système biologique considéré. Quel agent de pontage ? Pendant combien de temps ? A quelle concentration ? Plusieurs agents de pontage à réactivité orthogonale (chimie, taille du bras espaceur) pourront être évalués afin d'identifier celui/ceux qui permettront au mieux de topographier les protéines ou éventuellement les mouvements protéiques.

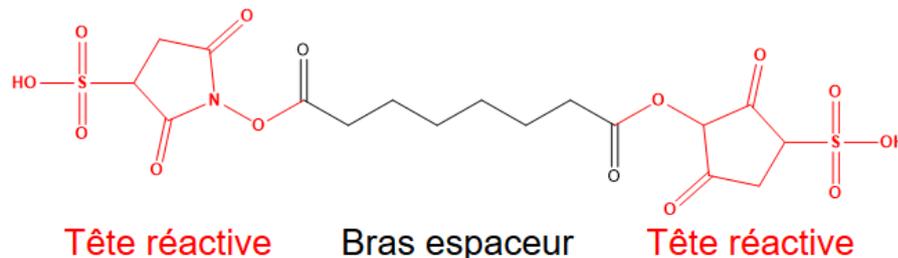
- *Lié à l'hétérogénéité du mélange peptidique obtenu* : bien que la XL-MS repose sur la même technologie que la protéomique classique, l'identification non ambiguë de nos peptides pontés demeure délicate en raison de leur présence à faible concentration en solution et en raison d'interférences liées aux modifications de surface (« dead-end ») ou aux pontages intramoléculaires. Pour faire face à la complexité de l'échantillon, des stratégies d'enrichissement peuvent être envisagées ainsi que l'utilisation d'instruments plus résolutifs.

- *Lié aux outils informatiques* : Cette complexité ainsi que la non-linéarité des ions cross-linkés imposent également l'utilisation de logiciels spécifiques et la qualité des résultats dépendra largement des outils bio-informatiques utilisés, ce dernier point étant à l'heure actuelle encore un frein à l'expansion de la technique. Considérant une moyenne de 22000 spectres de fragmentations par analyse, l'interprétation manuelle des données n'est pas envisageable. Un système de score doit donc être appliqué. Quel poids doit-on attribuer à l'intensité (des ions précurseurs/fragments), à la redondance, au recouvrement de séquence (contenant ou non l'agent de pontage), au taux de faux positifs ?

La mise au point du protocole de pontage, l'optimisation des conditions de spectrométrie de masse associées, mais surtout l'interprétation des données, ont représenté 80% de mes activités de recherches.

### **2.6.1 Les différents types d'agents pontants**

Les agents de pontage ou « cross-linker » sont des molécules chimiques constituées de deux têtes réactives et d'un bras espaceur dont la longueur est connue. Si les fonctions réactives sont identiques, on parlera d'agent de pontage homobifonctionnel, si les fonctions réactives sont différentes, alors on parlera d'agent hétérobifonctionnel. Dans ce cas, les fonctions réactives iront se fixer sur des groupements fonctionnels de nature différente.



**Figure 36 : Représentation schématique d'un agent pontant**

Pour choisir un agent pontant adéquat<sup>[169]</sup> il faut veiller à :

- Sa réactivité : l'agent doit être suffisamment réactif pour se fixer avec un rendement compatible avec l'analyse
- Sa sélectivité : l'agent possède-t-il une spécificité chimique qui dirigera le pontage vers un acide aminé défini ou bien le mécanisme réactionnel de pontage lui permettra-t-il de ponter indifféremment plusieurs types d'acides aminés ?
- Ses conditions de réaction : les conditions réactionnelles ne doivent pas interférer avec la préservation de la conformation de la protéine et entraîner des mouvements artéfactuels d'acides aminés.

Il existe plus d'une centaine d'agents pontants décrits dans la littérature, soit commercialement disponibles (<https://www.thermofisher.com/fr/fr/home/life-science/protein-biology/protein-labeling-crosslinking/protein-crosslinking/crosslinker-selection-tool.html>) soit synthétisés dans des laboratoires de recherche.<sup>[170]</sup> Ceux-ci seront choisis principalement en fonction de la fréquence des acides aminés cibles présents dans la protéine d'intérêt, mais également en fonction de la longueur de leur bras espaceur. La réaction de pontage est donc protéine-dépendante et il faudra peut-être effectuer plusieurs essais afin de trouver l'agent pontant adéquat.

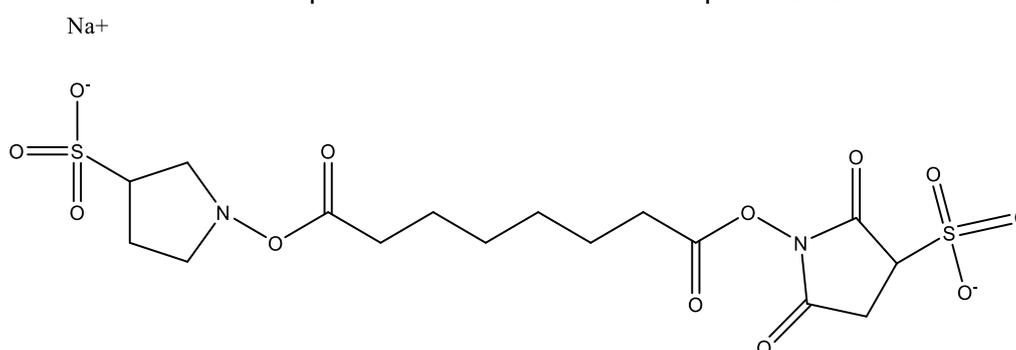
### 2.6.1.1 Les agents pontant réagissant sur les amines

Les agents pontants qui réagissent avec les amines primaires (en particulier au niveau du groupe  $\epsilon$ -amino de la chaîne latérale des lysines et des extrémités N-terminales)<sup>[169]</sup> sont actuellement les plus communément utilisés car beaucoup de protéines hydrosolubles possèdent des lysines régulièrement réparties sur leur séquence et la réaction de substitution résultante est plutôt spécifique.

Pendant très longtemps, le formaldéhyde et le glutaraldéhyde ont été largement appliqués en biologie comme agent pontant.<sup>[171-175]</sup> Malgré leurs avantages en termes de solubilité, perméabilité cellulaire et réactivité très élevée, leur application en XL-MS est limitée car l'identification des sites de pontage à partir de ces aldéhydes reste difficile.<sup>[176, 177]</sup>

En revanche, les N-hydroxysuccinimides (NHS) largement commercialisés constituent une part importante des agents homo-bifonctionnels amino-réactifs.

Le chef de file des agents amino-réactifs est le BS3. Disponible commercialement en version hydro- ou liposoluble, et possédant un bras espaceur de longueur intermédiaire (11.4 Å), il est décrit dans la littérature comme présentant un bon rendement réactionnel<sup>[178-182]</sup>. Il s'est déjà montré efficace au laboratoire pour la caractérisation du complexe SAGA-HAT.<sup>[183]</sup>

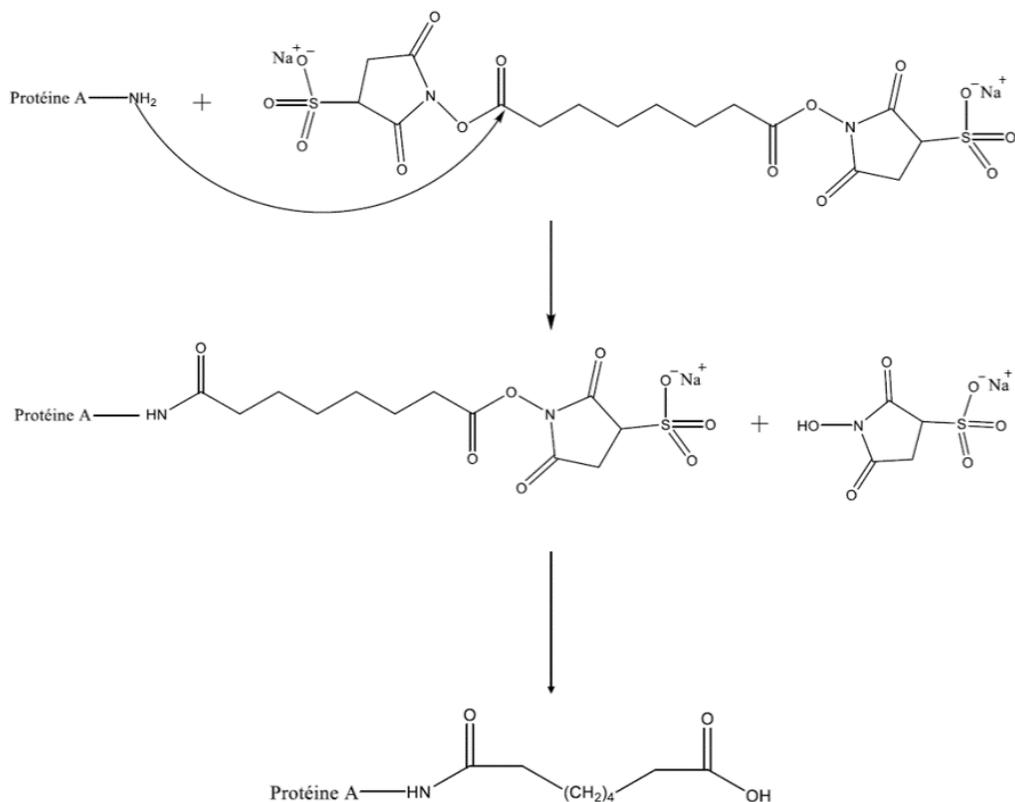


**Figure 37 : structure du BS3**

Malgré leur spécificité vis-à-vis des amines, les esters NHS peuvent aussi réagir sur d'autres groupements nucléophiles. L'équipe d'A. Sinz a signalé qu'environ 12,5 % de l'ensemble des peptides cross-linkés découlait d'un marquage sur la serine, 4,3 % sur la tyrosine et 3 % sur la thréonine lorsque le BS3 et le BS2G étaient utilisés comme réticulant.<sup>[184]</sup> Cependant, la réactivité des serines est beaucoup plus lente comparée à celle des amines. La réaction avec les thiols des cystéines libres (non impliquées dans un pont disulfure) est possible, mais les produits obtenus (les thioesters) sont instables et peuvent réagir avec les amines pour former une amide.<sup>[185]</sup> La sélectivité de la réactivité peut également être guidée

selon la gamme de pH définie. En effet, une étude a pu déterminer que la réaction des esters NHS était cinétiquement plus réactive à pH acide sur les tyrosines alors que celle-ci est cinétiquement plus réactive sur les lysines lorsque le pH est basique.<sup>[186, 187]</sup>

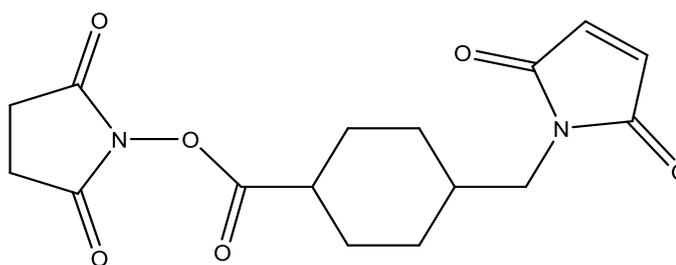
Une des difficultés lors de l'utilisation des esters NHS est qu'ils sont très sensibles à l'hydrolyse, celle-ci augmentant fortement avec le pH. Cette hydrolyse les rend inactifs et il a pu être montré que leur temps de demi-vie était de l'ordre de l'heure dans un pH physiologique.<sup>[177, 188, 189]</sup> C'est pourquoi de grandes précautions sont à respecter lors de leur conservation en minimisant les contacts avec l'humidité environnante et lors de leur utilisation (en les utilisant tout de suite après dilution dans un solvant aqueux).



**Figure 38 : Mécanisme réactionnel de la réaction d'hydrolyse sur un ester NHS.**

Une autre famille amino-réactive permettant de fixer les protéines sont les imidates<sup>[190]</sup>. L'augmentation de la distribution d'état de charge résultante est séduisante en vue de leur analyse par MS/MS. Toutefois, des conditions de réaction de pontage drastiques pour limiter les réactions secondaires (notamment à un pH très alcalin) ainsi qu'un temps de demi-vie en solution aqueuse inférieur à 30 min<sup>[191]</sup> limitent leur champ d'application<sup>[189]</sup>.

### 2.6.1.2 Les agents pontants réagissant sur les thiols

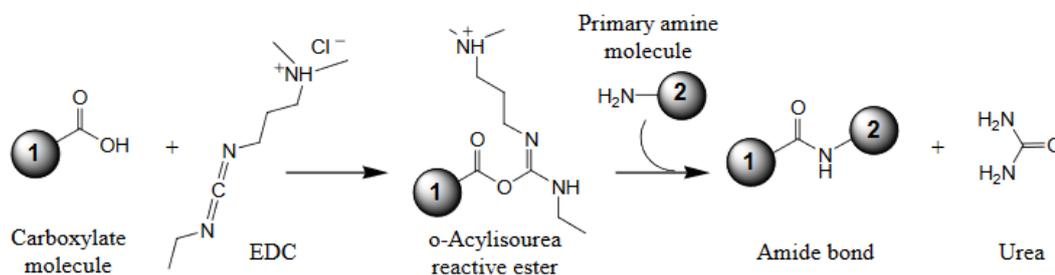


**Figure 39 : Structure du SMCC, un agent pontant hétérobifonctionnel réagissant sur les amines et les thiols.**

Les agents pontants réagissant sur les thiols vont essentiellement se fixer sur les cystéines lorsque celles-ci ne sont pas impliquées dans un pont disulfure. Dans le cas contraire, une étape de réduction des ponts disulfures en sulfhydryles est souvent indispensable afin de les rendre disponibles en amont de la réaction de pontage. Ils constituent le second groupe d'agents pontant le plus utilisé. Ils réagissent le plus souvent en effectuant une alkylation ou un échange dans les ponts disulfures. Les maléimides sont des agents qui sont spécifiques aux thiols dans la gamme de pH de 6,5 à 7,5.<sup>[169]</sup> Ils peuvent également réagir sur les amines, mais cette réaction est cinétiquement beaucoup plus lente que la réaction sur les thiols à pH physiologique.

### 2.6.1.3 Les agents pontants pouvant se fixer sur des acides

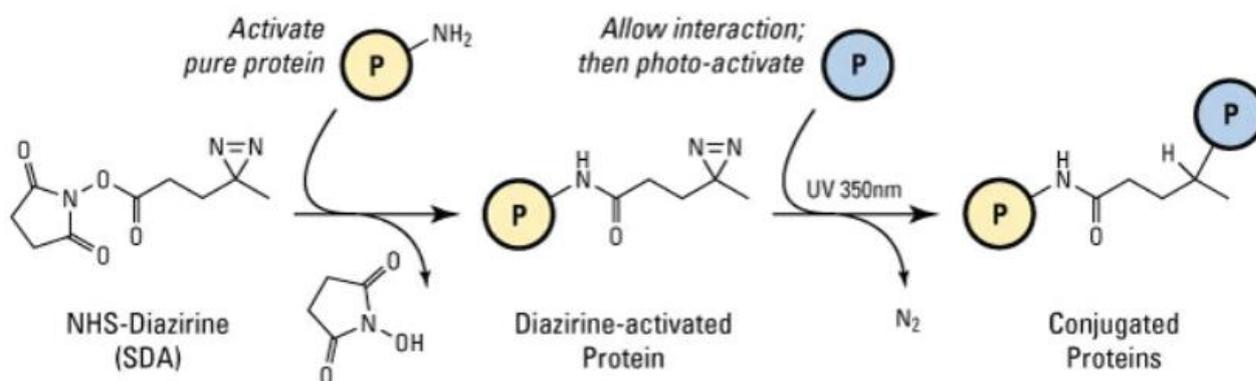
Quelques cross-linkers ont été développés pour aller se fixer sur les chaînes latérales des acides aspartiques et glutamiques. Les premiers agents pontants pouvant se fixer sur des résidus acides sont les dihydrazides.<sup>[169]</sup> Une des problématiques de l'utilisation de ces composés est qu'il faut utiliser un grand excès d'agents pontants pour conduire la réaction limitant ainsi leur utilité pour certaines applications. Un de ses agents est l'EDC. Il s'agit d'un agent hétérobifonctionnel réagissant sur les acides mais également sur les amines.



**Figure 40 : Réaction de pontage de l'EDC sur les protéines. (Extrait du manuel d'instruction de ThermoFisher<sup>[192]</sup>)**

#### 2.6.1.4 Les agents pontants photoréactifs

C'est une famille d'agents pontants non spécifiques. Ils réagissent sur les groupements fonctionnels lors de l'exposition à la lumière UV. <sup>[193]</sup> Avant l'activation, ils sont relativement non-actifs et peuvent donc être utilisés sous des conditions très contrôlées. Ils sont communément utilisés pour ponter les acides nucléiques et leurs complexes <sup>[194]</sup>. Les agents les plus utilisés sont les arylazides <sup>[158, 195]</sup>



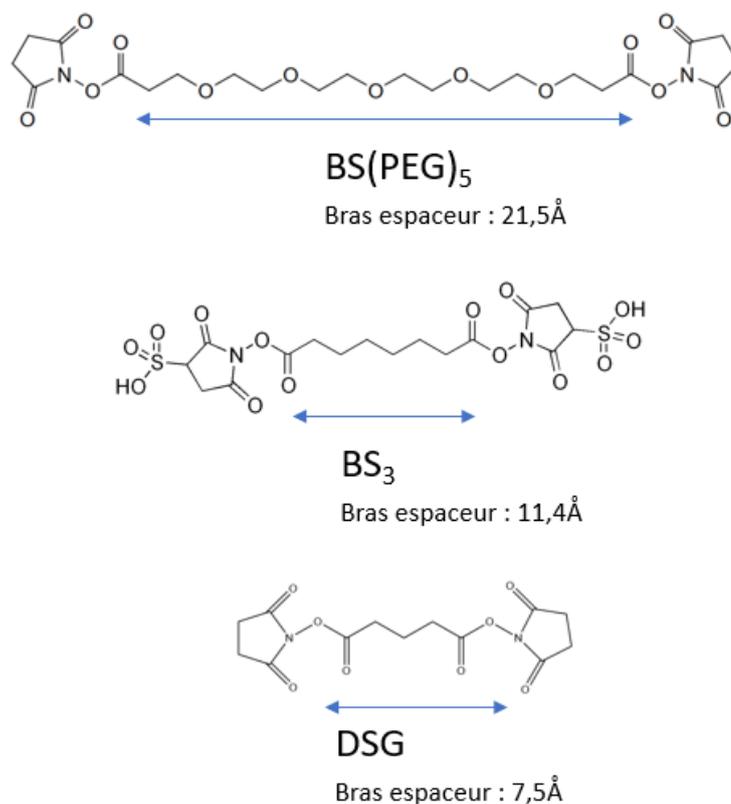
**Figure 41 : Mécanisme réactionnel de l'agent pontant photoréactif NHS-Diazirine. ( d'après ThermoFisher)<sup>[196]</sup>**

L'irradiation des phényl-azides mène à la formation de nitrènes qui peuvent se fixer de manière non spécifique sur tous les résidus de la protéine, en particulier au niveau de doubles liaisons <sup>[169]</sup>. Les diazirines présentent une meilleure photo stabilité et un meilleur rendement d'activation sous UV (330-3470nm). Afin de faciliter l'interprétation des spectres de fragmentation des peptides pontés, ces réactifs sont souvent utilisés sous forme d'agents de pontage hétérobifonctionnels comme le NHS-Diazirine SDA <sup>[197]</sup>.

La photoréaction peut également être utilisée *in vivo* par l'incorporation, directement dans nos protéines d'intérêt, d'analogues de la L-Leucine et de la L-Méthionine pendant la synthèse, puis activés par la lumière pour permettre une réticulation covalente des protéines dans les cellules. <sup>[198-200]</sup>

Ces agents photo-activables hétérobifonctionnels présentent certains avantages par rapport aux réactifs strictement thermo-chimiques. Outre leur réactivité sur des acides aminés de nature variée, ils permettent d'ajouter le réactif de manière précoce dans l'échantillon puis, d'initier le pontage par exposition à la lumière UV lors d'une étape ultérieure.

## 2.6.2 Les bras espaceurs



**Figure 42 : Famille des agents pontant se liant sur les amines avec des tailles de bras espaceurs variées**

Le bras espaceur du cross-linker est généralement une chaîne carbonée reliant les deux têtes réactives. En plus d'accorder une certaine flexibilité aux groupements fonctionnels, la longueur du bras espaceur imposera une limite supérieure à la distance entre deux acides aminés pontés. Cette longueur peut varier de « zéro » à des dizaines d'angströms. <sup>[201]</sup> Les agents pontants « zero-length » offrent les contraintes de distances les plus sévères (généralement <10 angströms) car ils relient de manière covalente deux chaînes latérales sans ajouter de bras espaceur<sup>[202-204]</sup>. Toutefois, les peptides pontés résultants sont souvent plus difficiles à identifier. En effet, l'absence d'un bras espaceur ne permet pas l'incorporation de marqueurs isotopiques, la majorité des systèmes étudiés ayant moins de 100 kDa de séquence protéique unique.<sup>[205, 206]</sup>

Tout comme la réactivité chimique du cross-linker doit être optimisée empiriquement, rien ne permet de prédire la longueur du bras espaceur adéquate en l'absence de données structurales. De façon surprenante, plusieurs études reportent des incohérences entre la longueur du bras espaceur et la distance entre deux AA pontés extraits d'une structure à haute résolution. Par exemple, l'équipe de J. Bruce a identifié des pontages avec un cross-linker PIR (bras espaceur de 14Å) qui ont également été observés avec des cross-linkers beaucoup plus petits comme le DSS (bras espaceur de 11 Å).<sup>[207]</sup>

Ces résultats montrent qu'en solution, des cross-links de distances plus grandes, mais aussi plus courtes que la longueur théorique de pontage liés au bras espaceur peuvent se produire. Ceci est en accord avec l'étude réalisée par Green qui a simulé les longueurs moyennes réalistes de 32 agents de pontage en montrant qu'elles s'avéraient souvent inférieures aux longueurs théoriques.<sup>[208]</sup> Ceci peut être lié à la flexibilité du cross-linker en solution ou du système biologique. C'est pourquoi le choix du cross-linker ne doit pas être abordé de la même façon si la finalité de l'étude est l'identification des séquences protéiques en interaction ou bien l'établissement d'une carte de contraintes de distance.

### **2.6.3 Nouvelles classes d'agent pontant modifiées aux propriétés orthogonales**

Afin d'améliorer l'identification des peptides pontés, deux stratégies sont disponibles<sup>[209]</sup> : (i) une réduction de la complexité de l'échantillon (méthode d'enrichissement en amont de l'analyse XL-MS) ou (ii) une réduction de la complexité des données (méthode d'enrichissement en spectres MSMS d'ions cross-linkés en aval de l'analyse XL-MS). Dans tous les cas, les agents pontants sont modifiés au niveau du bras espaceur.

#### **2.6.3.1 Réduction de la complexité de l'échantillon : développement d'agent pontant trifonctionnel**

Dans ce cas, une étiquette d'affinité est introduite au niveau du bas espaceur du cross-linker. L'étiquette la plus communément utilisée est une biotine. Après la réaction de pontage et protéolyse du complexe cross-linké, les peptides pontés sont enrichis grâce à des billes marquées à l'avidine<sup>[210, 211]</sup>. Une autre méthode d'introduction du tag d'affinité par chimie click a été proposée par Chowdhury<sup>[212]</sup>. Ainsi le groupement d'affinité biotinylé est incorporé au cross-linker après la réaction de pontage, ce qui évite de travailler en présence d'un agent de pontage à fort encombrement stérique. Cette approche a été utilisée pour l'étude de l' $\alpha$  synucléine, une protéine qui est connue pour s'agréger et pour jouer un rôle important dans la maladie de Parkinson<sup>[213]</sup>.

Les agents pontants avec des marqueurs d'affinité sont particulièrement intéressants lors du pontage de cellules ou de tissus intacts car l'agent de pontage doit pénétrer la membrane cellulaire et les organites subcellulaires.<sup>[214]</sup> En effet, les premières études *in vivo* étaient réalisées avec le formaldéhyde. Toutefois, cet agent pontant génère une multitude de pontages différents qui sont encombrants, voire impossibles à identifier par MS. C'est pourquoi un enrichissement en peptides pontés est obligatoire avant d'effectuer des analyses LC-MS sur ces systèmes.<sup>[214]</sup> Par ailleurs, l'équipe de Julia Chamot-Rooke a développé un nouvel agent pontant trifonctionnel constitué de deux fonctions carbamate NHS et une troisième fonction, un groupe azido "cliquable" afin de permettre un enrichissement supplémentaire des peptides réticulés.<sup>[215]</sup>

### 2.6.3.2 Réduction de la complexité des données : développement de cross-linkers marqué isotopiquement

Les agents pontants marqués par des isotopes ont été introduits il y a déjà quelques années et sont largement utilisés afin de faciliter la détection des peptides pontés. Cette approche consiste à substituer plusieurs atomes du bras espaceur de l'agent de pontage par leurs isotopes lourds stables. Très souvent, un mélange équimolaire d'agents de pontage lourds et légers est utilisé. Les groupements fonctionnels des acides aminés ayant autant de chance de réagir avec un agent de pontage léger que son homologue marqué, les signaux des peptides pontés apparaissent sur le spectre de masse sous forme de doublets avec un ratio 1 :1. Les peptides non modifiés par le cross-linker seront uniquement observés sous forme de singulet. Ces agents pontants sont très utilisés pour les expériences réalisées en analyse quantitative afin d'identifier les changements conformationnels des complexes protéiques.<sup>[181, 216, 217]</sup>

### 2.6.3.3 Réduction de la complexité des données : développement d'agent pontant MS-clivable

L'objectif ici est de diminuer l'espace de recherche en créant sur le spectre de fragmentation une empreinte de l'agent pontant<sup>[218]</sup> et pouvoir ainsi extraire les spectres de peptides pontés au sein du mélange peptidique extrêmement complexe. Les agents de pontages MS-clivables sont ainsi fonctionnalisés afin de posséder un groupement qui sera dissocié soit physiquement en amont de l'analyse MS (traitement par agents réducteurs notamment<sup>[219]</sup>) soit fragmenté directement pendant l'analyse MS.<sup>[220]</sup> Cette nouvelle stratégie permet de distinguer les peptides modifiés par le cross-linker des peptides linéaires grâce à des profils de fragmentation caractéristiques.

Un premier cross-linker clivable appelé Protein Interaction Reporter (PIR) fut développé en 2005 par l'équipe de J. Bruce<sup>[221]</sup>. Lors de la fragmentation CID, des ions « reporter » caractéristiques du cross-linker sont générés permettant de mieux distinguer les signaux correspondant aux peptides pontés. Plus récemment, l'incorporation de liaisons CID-labiles a révolutionné l'approche XL-MS.

Après activation des ions par CID, les deux peptides initialement réticulés sont libérés ce qui permet d'en déterminer la masse puis de les soumettre individuellement à de la MS/MS pour l'identification des deux séquences<sup>[218]</sup>. On contourne ainsi la complexité combinatoire qui survient avec la fragmentation des peptides pontés non clivables, et permet une affectation non ambiguë des séquences peptidiques réticulées.<sup>[218, 220, 222]</sup>

Trois liaisons CID-clivables couramment utilisées sont :

- Les liaisons C-S adjacentes au sulfoxyde, les ions sulfonium tels que le DSSO...<sup>[223]</sup>
- Les Liaison D-P (Asp-Pro) tels que le SUDP<sup>[224]</sup>
- Les liaisons C-N associées aux structures, l'urée ou les diamines quaternaires telles que le DSBU.<sup>[225]</sup>

En fonction de la force des liaisons clivables, ces réactifs peuvent être mis en œuvre pour des analyses effectuées en MS<sub>n</sub> et/ou en MS<sub>2</sub> pour faciliter l'identification des peptides pontés.<sup>[226]</sup>

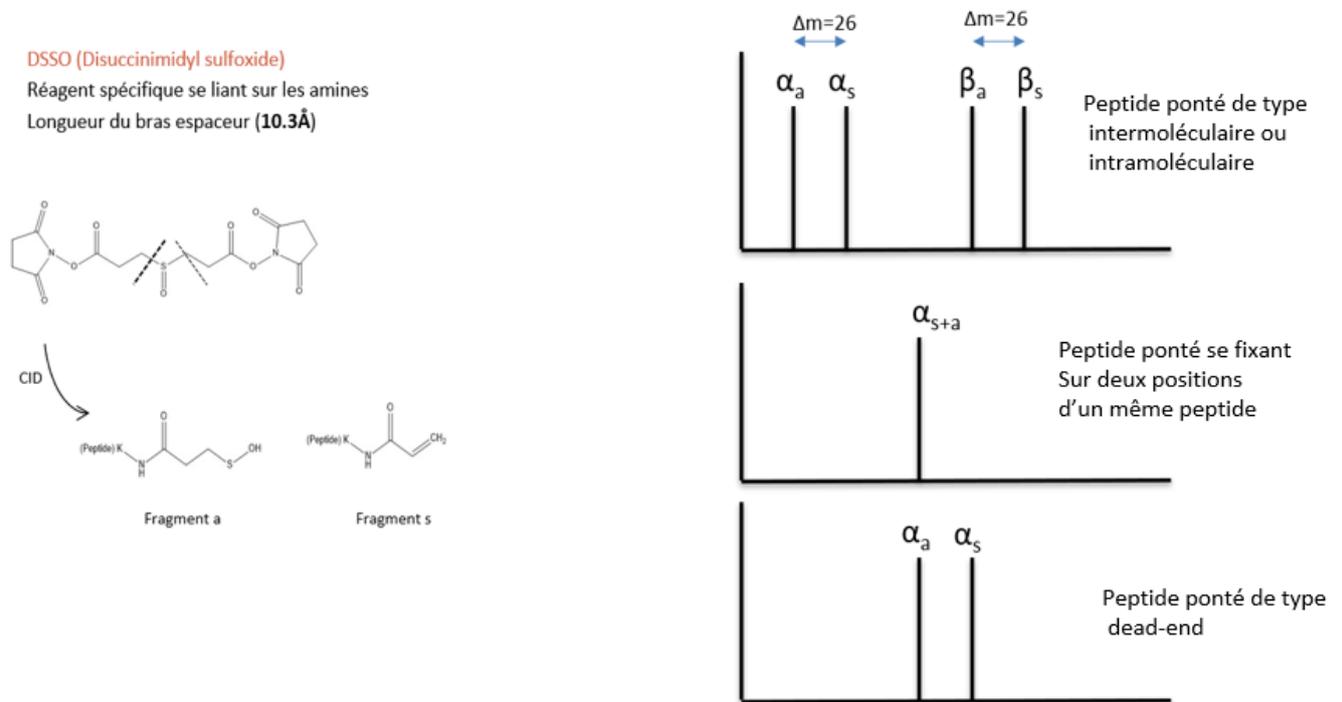
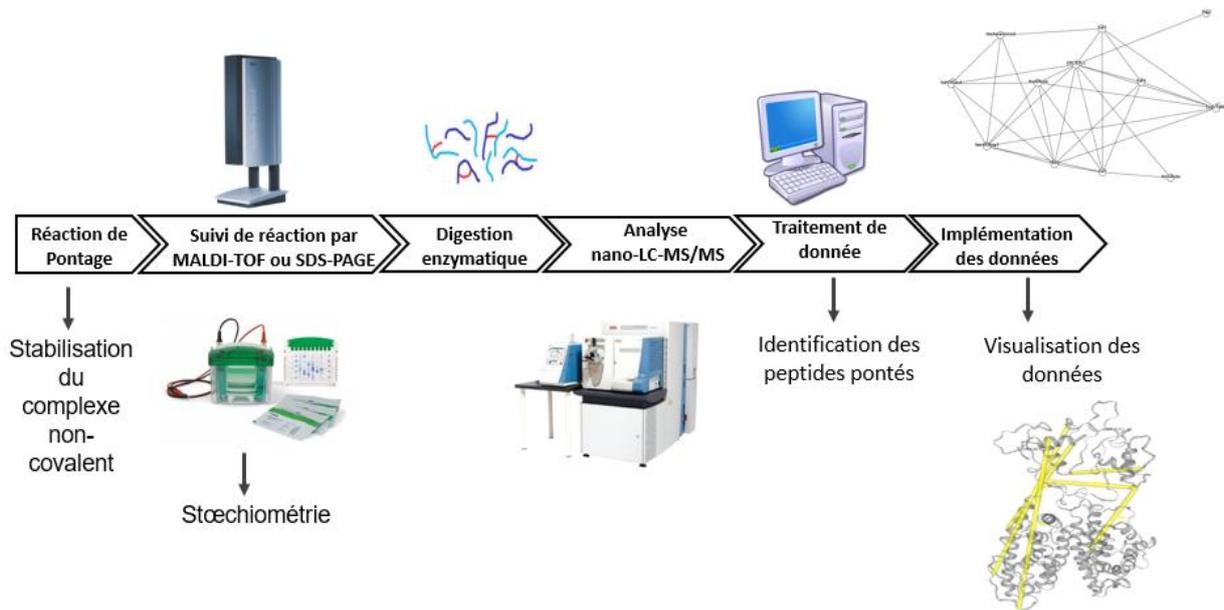


Figure 43 : Stratégie d'enrichissement par agent pontant clivable : exemple du DSSO

#### 2.6.4 Protocole expérimental

Une analyse XL-MS est constituée de quatre grandes étapes. Nous avons choisi de les décrire ici afin de mieux se représenter les éléments de contrôle à imposer lors d'une expérience XL-MS pour éviter l'introduction de biais :

- La réaction de pontage qui s'effectue en solution dans des conditions expérimentales choisies pour que le complexe d'intérêt soit maintenu dans sa forme native.
- La protéolyse du complexe ponté par digestion enzymatique
- L'analyse par MS
- L'identification des peptides pontés par l'utilisation d'outils bio-informatiques dédiés.



**Figure 44 : Représentation schématique des étapes de XL-MS**

La réaction de pontage : Celle-ci doit s'opérer dans les conditions proches des conditions physiologiques afin de stabiliser grâce au pontage les biomolécules dans une conformation la plus proche possible de la conformation *in vivo*. Le rendement de réaction doit être suffisamment poussé pour pouvoir générer des peptides pontés en concentration suffisante pour être détectée, sélectionnée et fragmentée par MS, mais un pontage excessif pourrait entraîner des réarrangements conformationnels artéfactuels qu'il sera par la suite impossible à différencier avec les informations structurales pertinentes.<sup>[227, 228]</sup>

Outre un tampon de solubilisation et un pH adapté, il faudra optimiser les paramètres cinétiques de la réaction comme la concentration en cross-linker, le temps de réaction ainsi que la température. Tous ces paramètres seront à définir en fonction du système biologique, mais aussi de la nature du cross-linker (réactivité chimique ou photo-activable notamment). En fonction de la taille du complexe biologique, l'état d'avancement de la réaction de pontage peut être suivi par gel d'électrophorèse SDS-PAGE. A partir de gels standards de polyacrylamide, il est envisageable de détecter des complexes de masses moléculaires inférieures à 250 kDa<sup>[229]</sup>.

Pour pallier la limite en taille des gels SDS-PAGE, le suivi de l'état d'avancement de la réaction de pontage peut également s'effectuer par MALDI-MS. Sa rapidité, sensibilité et tolérance raisonnable aux sels non volatils présents dans les tampons en font une technique appropriée pour analyser un échantillon cross-linké. Il est alors possible de prélever 0.2-0.5  $\mu$ l d'échantillon à intervalle de temps régulier et mesurer la masse moléculaire des différentes espèces présentes en solution. La technique MALDI étant destructrice pour l'étude d'interactions non covalentes (sauf dans des conditions très spécifiques)<sup>[230]</sup>, les masses mesurées refléteront les modifications covalentes subies par le complexe d'origine et en particulier le pontage entre sous-unités. Pour cela, un détecteur adapté à la détection d'ions de hauts rapports  $m/z$  doit être utilisé car les détecteurs standards de type MCP montrent rapidement une faible efficacité de conversion ion-électrons comme nous l'avons exposé dans

le chapitre I de la partie bibliographie. Le détecteur de haute masse HM4 basé sur la technologie « ion conversion dynode » (ICD) et développé par l'entreprise CovalX AG (Zürich, Suisse) est parfaitement bien adapté à ce type d'application<sup>[183]</sup>. Entre nos mains, il a démontré un gain en sensibilité dès  $m/z > 100\text{kDa}$ .

Génération de peptides pontés par approche Bottom-up : La protéolyse peut être effectuée soit sur le complexe ainsi ponté en solution soit à partir des bandes du gel SDS-PAGE. L'équipe d'Andrea Sinz a remarqué que la digestion en solution est beaucoup plus efficace que la digestion en gel.<sup>[189]</sup> Dans les deux cas, la stratégie est la même et repose sur le choix de l'enzyme de protéolyse. Les enzymes de faible spécificité permettent l'obtention de peptides de plus petite taille préférables pour l'analyse en MS, mais identifier les enchaînements d'acides aminés connectés à partir d'ions de petite taille n'est pas forcément un atout. D'autre part, l'absence de spécificité va considérablement augmenter l'espace de recherche ce qui complique encore plus le nombre de combinaisons potentielles de séquences peptidiques et donc l'identification des peptides pontés.

L'enzyme la plus fréquemment utilisée reste, comme dans le cas d'analyses protéomiques standards, la trypsine.<sup>[227]</sup> Toutefois, d'autres enzymes peuvent être utilisées, individuellement ou couplées sous forme de doubles digestions peuvent être réalisées afin d'améliorer l'identification des peptides pontés.<sup>[231, 232]</sup>

Les types de peptides pontés : Après digestion, on obtient un mélange peptidique très hétérogène car il est constitué de peptides linéaires (des peptides issus de zones qui n'ont pas été pontées) ainsi que des peptides pontés. Il est possible d'observer plusieurs types de peptides pontés, une nomenclature a été proposée par Schilling<sup>[233]</sup> :

- Des peptides dits « dead-end » où seulement une tête réactive du cross-linker a été fixée sur le complexe et l'autre tête a été hydrolysée (pontage de type 0).
- Des peptides dits « loop-link » où les deux extrémités fonctionnelles de l'agent de pontage ont réagi sur deux résidus d'acides aminés d'un même peptide (pontage de type 1).
- Des peptides dits « inter-link » où les deux extrémités fonctionnelles de l'agent de pontage ont réagi sur deux résidus d'acides aminés de deux peptides différents (pontage de type 2).

Il y a également des peptides à pontages multiple<sup>[157]</sup>, mais vu leur complexité et la difficulté à les interpréter, nous ne les prendrons pas en compte dans le cadre de ce travail.

Analyse par MS : Etant donné que les paires de peptides pontés sont généralement de faible abondance par rapport aux peptides non pontés, des stratégies d'enrichissement en peptides pontés ont été développées en amont de l'analyse par MS. Dans ce cadre, les techniques séparatives telles que la chromatographie par échange de cation (SCX) ou la chromatographie d'exclusion stérique (SEC) peuvent être utilisées.<sup>[234, 235]</sup> Certaines équipes préconisent par ailleurs de combiner l'utilisation d'agent pontant clivable.<sup>[236, 237]</sup>

Les échantillons seront analysés par nanoLC-MS/MS sur des spectromètres de masse disposant d'une précision de masse et d'une résolution adéquate pour faciliter l'assignation efficace des peptides pontés. Les méthodes d'acquisition utilisées sont similaires aux méthodes utilisées pour les analyses protéomiques classiques. De par leur haute résolution et vitesse d'échantillonnage, les instruments de type Orbitrap sont parfaitement adaptés et très largement utilisés<sup>[178]</sup>.

### 2.6.5 La stratégie XL-MS requiert l'utilisation d'outils bio-informatiques dédiés

L'identification des peptides pontés est très complexe et reste une difficulté importante de la technique. Ceci est dû à la complexité du mélange peptidique obtenu après pontage, à la grande quantité de spectres MS/MS à traiter ainsi qu'à la multitude de combinaisons théoriques possibles de peptides pontés. La fragmentation des peptides pontés est généralement complexe et imprévisible car elle s'effectue sur deux peptides de manière simultanée rendant l'interprétation des spectres MS/MS plus délicate que pour leurs homologues linéaires.<sup>[238]</sup>

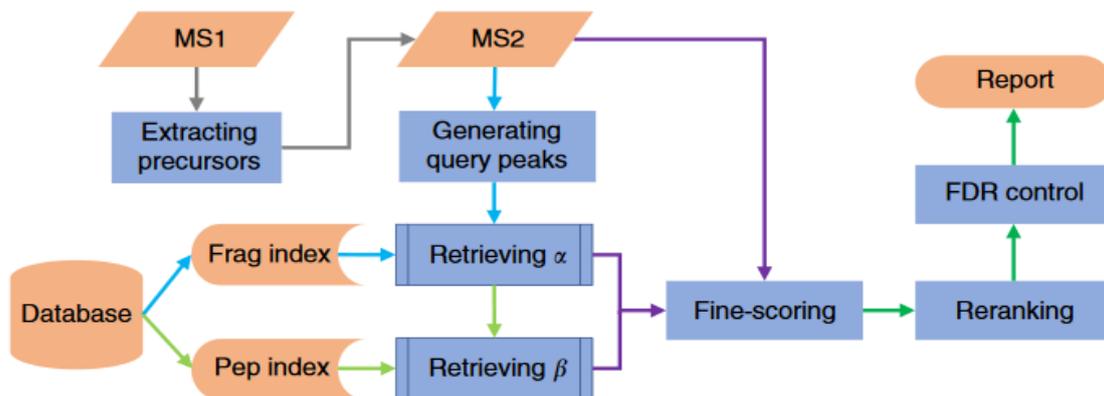


Figure 45 : Représentation schématique de l'algorithme de pLink, adapté de [164]

Bien que les logiciels d'identification classiques tels que Mascot ou Sequest sont très performants pour l'identification des peptides « linéaires » et permettent l'introduction de modifications complexes, aucun ne permet d'identifier les séquences des deux peptides liés par un pontage. De ce fait des outils informatiques spécialisés ont dû être développés<sup>[239]</sup>. Ceci représente ainsi un véritable défi d'un point de vue bio-informatique car la majorité des plateformes de recherche dans les bases de données fonctionnent en comparant les spectres MS/MS expérimentaux à une banque de données constituée de spectres MS/MS théoriques.<sup>[240]</sup>

Pour les peptides pontés, l'espace de recherche augmente de façon exponentielle, car toutes les paires de combinaisons de peptides possibles (d'ordre  $\frac{n^2+n}{2}$ )<sup>[240, 241]</sup> doivent être prises en compte lors de la construction d'une bibliothèque. Ainsi, plus l'échantillon est complexe, plus la puissance de calcul et le temps requis pour identifier les peptides pontés augmentent.

Les différents logiciels d'identification de peptides pontés :

De nombreux algorithmes d'analyses dédiés au pontage sont proposés dans la littérature. Tous ont la capacité d'interpréter des peptides intramoléculaires (pour accéder à l'étude structurale d'une protéine d'intérêt) ou des peptides intermoléculaires (pour accéder aux interfaces de proximité) au sein d'un mélange peptidique très complexe, mais peu d'entre eux se sont avérés robustes et stables entre nos mains...

**Tableau 5 : Liste des logiciels disponibles publiquement pour l'identification, la quantification, la visualisation et / ou la modélisation des zones pontées**

Nom du logiciel	Agent pontant		Identification	FDR	Quantification	Visualisation	Modélisation	Reference
	Non-Clivable	Clivable						
pLink2	X		X	X				[242]
StavroX	X		X	X				[243]
MeroX		X	X	X				[237]
XlinkX	X	X	X	X				[244]
XiSEARCH	X	X	X	X				[239]
xQuest/xProphet	X		X	X				[245]
MetamorpheusXL	X	X	X	X				[246]
Protein Prospector	X		X	X				[209]
Kojak	X		X	X				[247]
SIM-XL	X		X	X		X		[248]
Xilmass	X		X	X				[249]
MaxQuant	X		X	X	X			[250]
CLMSVault	X		X	X	X	X		[251]
XiView						X		[252]
ProXL						X		[253]
XlinkDB 2.0						X	X	[254]
Xwalk						X		[255]
Xlink Analyzer						X	X	[256]
DynaXL							X	[257]
ROSETTA							X	[78]
HADDOCK							X	[76]

Une stratégie courante consiste à utiliser des algorithmes spécialisés pour réduire l'espace de recherche, allégeant ainsi la charge de calcul. Par exemple, xQuest / xProphet<sup>[245, 258]</sup> a été conçu pour permettre des recherches à grande échelle dans les bases de données en utilisant un préfiltrage des peptides candidats basés sur le marquage isotopique du cross-linker et minimiser ainsi le nombre de permutations qui doivent être prises en compte. StavroX<sup>[243]</sup> utilise un 1<sup>er</sup> filtre sur la masse des ions précurseurs en la comparant à une liste précalculée de peptides pontés obtenus de manière théorique. Des logiciels tels que ProteinProspector<sup>[209]</sup> et pLink<sup>[242]</sup> traitent les peptides pontés comme des peptides uniques portant de grandes modifications inconnues, identifiant les peptides individuels puis recombinaient les résultats provenant des spectres communs en peptides pontés. Un autre logiciel couramment utilisé est XiSearch<sup>[259]</sup> développé par J. Rappsilber. Ce logiciel applique la même approche que pLink. Il dissocie par le calcul les peptides pontés et contourne ainsi le problème de taille des banques de données. Quel que soit le logiciel considéré, une fonction de score détermine ensuite la similarité entre les spectres théoriques et expérimentaux et les meilleurs peptides candidats sont extraits.<sup>[239]</sup>

Des logiciels adaptés à l'identification des peptides pontés par des agents de pontages clivables ont également été spécifiquement développés. À ce jour, seuls quelques logiciels exploitant les propriétés des agents clivables sont disponibles tels que MeroX<sup>[237]</sup> et XlinkX<sup>[163]</sup>. Ils permettent non seulement d'identifier les peptides pontés clivés par MS en identifiant les ions fragments caractéristiques des agents pontants dans les cellules de collision CID, HCD et analysés en MS3, mais également les peptides pontés clivés par ETD.

La grande majorité des logiciels bio-informatiques dédiés à l'identification des peptides pontés sont des logiciels open source développés par des laboratoires de recherche publics et nécessitant un développement constant et des mises à jour. Malgré les progrès continus, ces logiciels souffrent toujours d'un manque d'automatisation et de compatibilité entre différents formats de données. A mon arrivée au laboratoire, il était considéré qu'aucun logiciel ne satisfaisait tous les critères requis pour analyser une large variété de types de données obtenues par XL-MS. Après plusieurs tests sur notre système modèle, le Récepteur Nucléaire de l'Acide Rétinoïque, nous avons été amenés à travailler avec 4 d'entre eux : pLink, XlinkX, MeroX et XiSearch.

## **2.6.6 Validation des identifications de peptides pontés**

Une fois la recherche réalisée, comme toute méthode dont le résultat repose sur une fonction de score, une étape de validation est indispensable afin d'éliminer les erreurs d'interprétation (faux positifs). Une méthode populaire en protéomique pour évaluer l'erreur attendue parmi les résultats rapportés est le taux de fausses découvertes (False Discovery Rate FDR) par l'approche cible-leurre (target-decoy) comme nous l'avons décrite dans le chapitre 2. Pour l'analyse effectuée en XL-MS, l'estimation du FDR est délicate du fait que chaque correspondance est un composite de deux peptides, chacun avec sa propre probabilité d'être faux.<sup>[166, 167]</sup> Rappsilber a montré que la validation des peptides pontés était réalisée de la même manière que ce soit lors de l'utilisation d'un agent pontant hétérobifonctionnel ou

homobifonctionnel et que celle-ci est importante afin de limiter les fausses identifications réalisées par les algorithmes de recherche. [166, 167]

Une des difficultés rencontrées lors de l'estimation du FDR est l'influence de la taille des bases de données utilisées. En effet, travailler avec une petite base de données (<10 protéines) pourrait être considéré comme un atout en XL-MS car cela permet de réduire le nombre de permutations et donc réduit le nombre de cross-links théoriques, mais en revanche, très peu de peptides leurres sont générés. Il est alors délicat d'effectuer une stratégie decoy pertinente, ce qui entraîne une réduction de l'estimation du FDR. Pour contourner ce problème, Claudio Iacobucci et Andrea Sinz conseillent dans le protocole d'utilisation de MeroX d'ajouter des séquences protéiques supplémentaires à la base de données utilisée afin d'améliorer l'estimation du FDR. [237, 238]

Dans MeroX, il est possible d'ajouter la base de données cRAP qui contient des séquences de 116 contaminants fréquemment observés dans les analyses de protéines MS. Toutefois, l'augmentation de la taille des bases de données augmentera le nombre de peptides pontés leurre et peut entraîner une diminution du nombre de peptides identifiés. [237] D'autres groupes ont également tenté d'améliorer la validation des peptides pontés en mettant en place des algorithmes supplémentaires. [260]

La dispersion et l'absence de « normalisation » des logiciels dédiés à l'identification des peptides pontés ainsi que le calcul du FDR pour identifier les faux positifs reste sujet à débat au sein de la communauté scientifique. C'est pourquoi une validation manuelle rigoureuse des peptides pontés identifiés est encore souvent indispensable. Ces critères de validation manuelle seront détaillés et discutés dans la seconde partie de la thèse.

### **2.6.7 Les contraintes de la méthodologie XL-MS**

L'approche XL-MS est devenue ces dernières années un outil structural accessible et puissant pour cartographier les interactions protéine-protéine. Son application générale et universelle reste entravée par un manque de robustesse du protocole général. Une source d'erreur peut être introduite à presque chaque étape de la méthodologie et des expériences de contrôles tant au niveau expérimental que traitement informatique sont nécessaires.

Une des contraintes méthodologiques les plus importantes de la XL-MS est qu'aucune standardisation de la méthode n'a été pour l'instant réalisée. Une étude interlaboratoire réalisée sur la BSA a ainsi montré une grande hétérogénéité de la stratégie d'un point de vue expérimental, analytique et interprétation des données [178]. En effet, chaque équipe utilise sa propre méthodologie afin d'identifier les peptides pontés selon ses propres critères.

Néanmoins, il n'est pas évident d'imaginer un protocole expérimental défini, ou en tout cas généralisable, comme en protéomique classique, principalement parce que la réaction de pontage est protéine-dépendante. La complexité et l'inhomogénéité du mélange peptidique qui en résulte (tant en nombre, qu'en taille, en nature linéaire versus branchée, en gamme de concentration,) est également une étape à choix multiple quant à la méthode d'enrichissement qu'il faudra choisir en amont de l'analyse MS et à la méthode d'acquisition MS elle-même.

Lors d'analyse en MS/MS des peptides cross-linkés, l'énergie interne est distribuée à travers un grand nombre de liaisons, il y a donc moins d'énergie par liaison ce qui peut mener à une fragmentation incomplète.<sup>[227]</sup> D'autre part, la fenêtre  $m/z$  utilisée est limitée par un bon compromis entre le nombre de cycles d'analyses (donc nombre de peptides sélectionnés pour un évènement de fragmentation), et la précision de mesure de masse que l'on souhaite obtenir. La faible abondance des ions pontés par rapport aux peptides non cross-linkés linéaires est également une difficulté car en acquisition DDA les ions les plus abondants sont sélectionnés en premier pour la fragmentation<sup>[228]</sup>. Toutes ces raisons rendent la standardisation d'un protocole compliquée.

A tout ceci, s'ajoute une source de variabilité supplémentaire des résultats liée à la reproductibilité des expériences. En effet, la forte complexité du mélange peptidique peut être à l'origine d'une baisse dans la reproductibilité technique, les mêmes ions n'étant pas systématiquement sélectionnés par MS/MS d'une analyse à l'autre. D'autre part, la reproductibilité biologique peut être fortement impactée par plusieurs facteurs (stabilité de l'échantillon lors de la préparation *ex vivo*, flexibilité de la structure biologique en solution, influence de l'efficacité de la digestion...). Ce point sera illustré lors de la caractérisation du complexe de récepteur nucléaire RAR/RXR.

### 3. Conclusion

Ainsi, dans de ce chapitre, nous avons montré que de nombreuses techniques et méthodologies sont à notre disposition pour appréhender une étude structurale de protéine ou de complexes biologiques. Au sein de ces techniques, la MS se fraye une place de choix. Alors qu'elle n'est pas encore capable de générer des données structurales à haute résolution spatiale, elle offre des informations de stœchiométrie, d'accessibilité au solvant et de contraintes de distances qui s'avèrent capitales pour compléter, valider ou infirmer les hypothèses émises par d'autres approches. Aujourd'hui, la communauté scientifique se dirige vers une approche intégrative où plusieurs méthodologies d'analyses structurales sont interprétées en parallèle et appréhendées de façon complémentaire. En particulier, la stratégie XL-MS entre en interaction forte avec de nombreuses techniques telles que la DRX, cryo-EM, SAXS ou encore la modélisation moléculaire. Une des difficultés persistantes est l'intégration des données venant de plusieurs techniques. C'est pourquoi il est nécessaire, pour une interprétation pertinente des résultats, de réussir à établir une standardisation des protocoles et d'améliorer les plateformes intégratives permettant de mieux comparer les résultats obtenus par les différentes techniques.

Dans la partie résultat de ce manuscrit, nous présenterons au travers de quelques exemples le potentiel de la MS dans cette quête de données structurales. Pour cela, nous montrerons comment la MS native a permis de mieux appréhender le mécanisme d'oligomérisation d'un peptide issu de la protéine CREB-Binding Protein. Puis, nous exposerons l'optimisation d'un protocole de cross-link au laboratoire en considérant trois problématiques d'étude : la mise en évidence d'un changement conformationnel sur un

système biologique simple, l'étude d'un système biologique nommée NuA4 formé de 12 sous-unités et l'étude d'une protéine partiellement déstructurée.

## 4. Bibliographie

1. Zuiderweg, E.R.P., *Mapping protein-protein interactions in solution by NMR Spectroscopy*. Biochemistry, 2002. **41**(1): p. 1-7.
2. Delsuc, M.A., M. Vitorino, and B. Kieffer, *Determination of Protein Structure and Dynamics by NMR*, in *Structural Biology in Drug Discovery*, J.P. Renaud, Editor. 2020. p. 295-323.
3. Markwick, P.R.L., T. Malliavin, and M. Nilges, *Structural Biology by NMR: Structure, Dynamics, and Interactions*. Plos Computational Biology, 2008. **4**(9).
4. Saunders, M., A. Wishnia, and J.G. Kirkwood, *The Nuclear Magnetic Resonance Spectrum of Ribonuclease1*. Journal of the American Chemical Society, 1957. **79**(12): p. 3289-3290.
5. Ernst, R.R., *Nuclear magnetic resonance Fourier transform spectroscopy*. Bioscience Reports, 1992. **12**(3): p. 143-187.
6. Wuthrich, K., et al., *Sequential Resonance Assignments as a Basis for Determination of Spatial Protein Structures by High-Resolution Proton Nuclear Magnetic-Resonance*. Journal of Molecular Biology, 1982. **155**(3): p. 311-319.
7. Agris, P.F. and P.G. Schmidt, *Structure of transfer RNA by carbon NMR: resolution of single carbon resonances from <sup>13</sup>C-enriched, purified species*. Nucleic acids research, 1980. **8**(9): p. 2085-2091.
8. Williamson, M.P., T.F. Havel, and K. Wüthrich, *Solution conformation of proteinase inhibitor IIA from bull seminal plasma by <sup>1</sup>H nuclear magnetic resonance and distance geometry*. Journal of Molecular Biology, 1985. **182**(2): p. 295-315.
9. Ishima, R. and D.A. Torchia, *Protein dynamics from NMR*. Nature Structural Biology, 2000. **7**(9): p. 740-743.
10. Thompson, J.M., et al., *Accurate protein structure modeling using sparse NMR data and homologous structure information*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(25): p. 9875-9880.
11. Kay, L.E., *NMR methods for the study of protein structure and dynamics*. Biochemistry and Cell Biology, 1997. **75**(1): p. 1-15.
12. Göbl, C., et al., *NMR approaches for structural analysis of multidomain proteins and complexes in solution*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2014. **80**: p. 26-63.
13. Kang, C., *Applications of In-Cell NMR in Structural Biology and Drug Discovery*. International journal of molecular sciences, 2019. **20**(1): p. 139.
14. Bechinger, B., *DNP Solid-State NMR of Biological Membranes*. eMagRes, 2018: p. 25-34.
15. Andronesi, O.C., et al., *Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy*. Journal of the American Chemical Society, 2005. **127**(37): p. 12965-12974.
16. Andronesi, O.C., et al., *Probing membrane protein orientation and structure using fast magic-angle-spinning solid-state NMR*. Journal of Biomolecular NMR, 2004. **30**(3): p. 253-265.
17. Luchinat, E. and L. Banci, *In-cell NMR: a topical review*. Iucrj, 2017. **4**: p. 108-118.
18. van Ingen, H. and A.M.J.J. Bonvin, *Information-driven modeling of large macromolecular assemblies using NMR data*. Journal of Magnetic Resonance, 2014. **241**: p. 103-114.
19. Gibbs, E.B., E.C. Cook, and S.A. Showalter, *Application of NMR to studies of intrinsically disordered proteins*. Archives of Biochemistry and Biophysics, 2017. **628**: p. 57-70.
20. Case, D.A., *Molecular dynamics and NMR spin relaxation in proteins*. Accounts of Chemical Research, 2002. **35**(6): p. 325-331.
21. Finch, J.T., *Chapter 3 X-ray diffraction from viruses*, in *Perspectives in Medical Virology*, M.V. Nermut and A.C. Steven, Editors. 1987, Elsevier. **3** : p. 75-88.

22. Bunaciu, A.A., E.G. Udristioiu, and H.Y. Aboul-Enein, *X-Ray Diffraction: Instrumentation and Applications*. Critical Reviews in Analytical Chemistry, 2015. **45**(4): p. 289-299.
23. Ryland, A.L., *X-ray diffraction*. Journal of Chemical Education, 1958. **35**(2): p. 80.
24. Koch, M.H.J. and J. Bordas, *X-Ray-Diffraction and Scattering on Disordered-Systems Using Synchrotron Radiation*. Nuclear Instruments & Methods in Physics Research, 1983. **208**(1-3): p. 461-469.
25. Sakabe, N., *X-Ray-Diffraction Data-Collection System for Modern Protein Crystallography with a Weissenberg Camera and an Imaging Plate Using Synchrotron Radiation*. Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 1991. **303**(3): p. 448-463.
26. Jones, N., *Crystallography: Atomic secrets*. 2014(1476-4687 (Electronic)).
27. Egli, M., *Diffraction Techniques in Structural Biology*. Current protocols in nucleic acid chemistry, 2016. **65**: p. 7.13.1-7.13.41.
28. Shi, Y.G., *A Glimpse of Structural Biology through X-Ray Crystallography*. Cell, 2014. **159**(5): p. 995-1014.
29. de Chadarevian, S., *John Kendrew and myoglobin: Protein structure determination in the 1950s*. Protein science : a publication of the Protein Society, 2018. **27**(6): p. 1136-1143.
30. Yano, J.K., et al., *The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-angstrom resolution*. Journal of Biological Chemistry, 2004. **279**(37): p. 38091-38094.
31. Eisenberg, D., *1 X-Ray Crystallography and Enzyme Structure*, in *The Enzymes*, P.D. Boyer, Editor. 1970, Academic Press. p. 1-89.
32. Cate, J.H., et al., *X-ray crystal structures of 70S ribosome functional complexes*. Science, 1999. **285**(5436): p. 2095-2104.
33. Zubay, G. and M.H.F. Wilkins, *X-ray diffraction studies of the structure of ribosomes from Escherichia coli*. Journal of Molecular Biology, 1960. **2**(2): p. 105-IN4.
34. Hunter, M.S., et al., *X-ray Diffraction from Membrane Protein Nanocrystals*. Biophysical Journal, 2011. **100**(1): p. 198-206.
35. Zheng, H., et al., *The future of crystallography in drug discovery*. Expert opinion on drug discovery, 2014. **9**(2): p. 125-137.
36. Davis, A.M., S.J. Teague, and G.J. Kleywegt, *Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design*. Angewandte Chemie International Edition, 2003. **42**(24): p. 2718-2736.
37. Keiderling, T.A., *Protein and peptide secondary structure and conformational determination with vibrational circular dichroism*. Current Opinion in Chemical Biology, 2002. **6**(5): p. 682-688.
38. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. Nature protocols, 2006. **1**(6): p. 2876-2890.
39. Nogales, E., *The development of cryo-EM into a mainstream structural biology technique*. Nature methods, 2016. **13**(1): p. 24-27.
40. De Rosier, D.J. and A. Klug, *Reconstruction of Three Dimensional Structures from Electron Micrographs*. Nature, 1968. **217**(5124): p. 130-134.
41. Harris, J.R., *Transmission electron microscopy in molecular structural biology: A historical survey*. Archives of biochemistry and biophysics, 2015. **581**: p. 3-18.
42. Henderson, R., et al., *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy*. 1990. **213**(4) : p. 899 - 929.
43. Cressey, D. and E. Callaway, *Cryo-electron microscopy wins chemistry Nobel*. Nature, 2017. **550**(7675): p. 167-167.

44. Altenhoff, M., et al., *An optimized evaluation strategy for a comprehensive morphological soot nanoparticle aggregate characterization by electron microscopy*. Journal of Aerosol Science, 2020. **139**: p. 130 - 139.
45. Bai, X.-c., G. McMullan, and S.H.W. Scheres, *How cryo-EM is revolutionizing structural biology*. Trends in Biochemical Sciences, 2015. **40**(1): p. 49-57.
46. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins*. FEBS Letters, 2015. **589**(19, Part A): p. 2570-2577.
47. Brosey, C.A. and J.A. Tainer, *Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology*. Current Opinion in Structural Biology, 2019. **58**: p. 197-213.
48. Chen, P.-C. and J. Hennig, *The role of small-angle scattering in structure-based screening applications*. Biophysical reviews, 2018. **10**(5): p. 1295-1310.
49. Hauseman, Z.J., et al., *Homogeneous Oligomers of Pro-apoptotic BAX Reveal Structural Determinants of Mitochondrial Membrane Permeabilization*. Molecular Cell, 2020. **79**(1): p. 68-83.
50. Singh, B., et al., *Application of vibrational microspectroscopy to biology and medicine*. Current Science, 2012. **102**(2): p. 232-244.
51. Capuano, E. and S. Ruth, *Infrared Spectroscopy: Applications*. Encyclopedia of Food and Health, 2016 p. 424-431.
52. Schmid, F.-X., *Biological Macromolecules: UV-visible Spectrophotometry*. eLS, 2001.
53. Barth, A., *Infrared spectroscopy of proteins*. Biochimica Et Biophysica Acta-Bioenergetics, 2007. **1767**(9): p. 1073-1101.
54. Butler, H.J., et al., *Using Raman spectroscopy to characterize biological materials*. Nature Protocols, 2016. **11**(4): p. 664-687.
55. Stradi, R., E. Pini, and G. Celentano, *The chemical structure of the pigments in Ara macao plumage*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 2001. **130**(1): p. 57-63.
56. Rathore, N., W. Qi, and W.C. Ji, *Cleaning characterization of protein drug products using UV-vis spectroscopy*. Biotechnology Progress, 2008. **24**(3): p. 684-690.
57. Proshlyakov, D.A., *UV optical absorption by protein radicals in cytochrome c oxidase*. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2004. **1655**: p. 282-289.
58. Beitlich, T., et al., *Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-Vis spectroscopy and X-ray crystallography*. Journal of Synchrotron Radiation, 2007. **14**: p. 11-23.
59. Pattison, D.I., A.S. Rahmanto, and M.J. Davies, *Photo-oxidation of proteins*. Photochemical & Photobiological Sciences, 2012. **11**(1): p. 38-53.
60. Antosiewicz, J.M. and D. Shugar, *UV-Vis spectroscopy of tyrosine side-groups in studies of protein structure. Part 2: selected applications*. Biophysical reviews, 2016. **8**(2): p. 163-177.
61. Goormaghtigh, E., V. Raussens, and J.-M. Ruysschaert, *Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes*. Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes, 1999. **1422**(2): p. 105-185.
62. Vigano, C., et al., *Hydrogen-deuterium exchange in membrane proteins monitored by IR spectroscopy: A new tool to resolve protein structure and dynamics*. Biopolymers, 2004. **74**(1-2): p. 19-26.
63. Grimard, V., et al., *Phosphorylation-induced conformational changes of cystic fibrosis transmembrane conductance regulator monitored by attenuated total reflection-fourier transform IR spectroscopy and fluorescence spectroscopy*. Journal of Biological Chemistry, 2004. **279**(7): p. 5528-5536.
64. Maity, J.P., et al., *Identification and discrimination of bacteria using Fourier transform infrared spectroscopy*. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2013. **116**: p. 478-484.

65. Chalmers, J.M. and G. Dent, *Vibrational Spectroscopic Methods in Pharmaceutical Solid-state Characterization*. Polymorphism, 2006: p. 95-138.
66. Genzel, L. and J. Kuhl, *A new version of a Michelson interferometer for Fourier transform infrared spectroscopy\**. Infrared Physics, 1978. **18**(2): p. 113-120.
67. Stuart, B., *Infrared Spectroscopy*. Kirk-Othmer Encyclopedia of Chemical Technology, 2005.
68. Dousseau, F. and M. Pezolet, *Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods*. Biochemistry, 1990. **29**(37): p. 8771-8779.
69. Dong, A., P. Huang, and W.S. Caughey, *Protein secondary structures in water from second-derivative amide I infrared spectra*. Biochemistry, 1990. **29**(13): p. 3303-3308.
70. Fabian, H. and W. Mäntele, *Infrared Spectroscopy of Proteins*. Handbook of Vibrational Spectroscopy, 2001.
71. Pelton, J.T. and L.R. McLean, *Spectroscopic Methods for Analysis of Protein Secondary Structure*. Analytical Biochemistry, 2000. **277**(2): p. 167-176.
72. Mikhonin, A.V., et al., *Peptide Secondary Structure Folding Reaction Coordinate: Correlation between UV Raman Amide III Frequency,  $\Psi$  Ramachandran Angle, and Hydrogen Bonding*. The Journal of Physical Chemistry B, 2006. **110**(4): p. 1928-1943.
73. Braitbard, M., D. Schneidman-Duhovny, and N. Kalisman, *Integrative Structure Modeling: Overview and Assessment*. Annual Review of Biochemistry, 2019. **88** : p.113 - 135.
74. Soni, N. and M.S. Madhusudhan, *Computational modeling of protein assemblies*. Current Opinion in Structural Biology, 2017. **44**: p. 179-189.
75. Meng, X.-Y., et al., *Molecular docking: a powerful approach for structure-based drug discovery*. Current computer-aided drug design, 2011. **7**(2): p. 146-157.
76. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin, *HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information*. Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.
77. May, A. and M. Zacharias, *Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility*. Proteins: Structure, Function, and Bioinformatics, 2007. **69**(4): p. 774-780.
78. Rohl, C.A., et al., *Protein Structure Prediction Using Rosetta*, in *Methods in Enzymology*. 2004, Academic Press. **384** : p. 66-93.
79. Koga, N., et al., *Principles for designing ideal protein structures*. Nature, 2012. **491**(7423): p. 222-227.
80. Faini, M., F. Stengel, and R. Aebersold, *The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology*. Journal of The American Society for Mass Spectrometry, 2016. **27**(6): p. 966-974.
81. Snijder, J., et al., *Studying 18 MDa Virus Assemblies with Native Mass Spectrometry*. Angewandte Chemie International Edition, 2013. **52**(14): p. 4020-4023.
82. Calabrese, A.N. and S.E. Radford, *Mass spectrometry-enabled structural biology of membrane proteins*. Methods, 2018. **147**: p. 187-205.
83. Lössl, P., M. van de Waterbeemd, and A. Heck, Jr., *The diverse and expanding role of mass spectrometry in structural and molecular biology*. The EMBO journal, 2016. **35**(24): p. 2634-2657.
84. Katta, V. and B.T. Chait, *Observation of the Heme Globin Complex in Native Myoglobin by Electrospray-Ionization Mass-Spectrometry*. Journal of the American Chemical Society, 1991. **113**(22): p. 8534-8535.
85. Leney, A.C. and A.J.R. Heck, *Native Mass Spectrometry: What is in the Name?* Journal of the American Society for Mass Spectrometry, 2017. **28**(1): p. 5-13.
86. Thompson, N.J., S. Rosati, and A.J.R. Heck, *Performing native mass spectrometry analysis on therapeutic antibodies*. Methods, 2014. **65**(1): p. 11-17.

87. Smits, A.H. and M. Vermeulen, *Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities*. Trends in Biotechnology, 2016. **34**(10): p. 825-834.
88. Shiu-Hin Chan, D., et al., *Structural insights into the EthR–DNA interaction using native mass spectrometry*. Chemical Communications, 2017. **53**(25): p. 3527-3530.
89. Guo, C., M. Cheng, and M.L. Gross, *Protein-Metal-Ion Interactions Studied by Mass Spectrometry-Based Footprinting with Isotope-Encoded Benzhydrazide*. Analytical chemistry, 2019. **91**(2): p. 1416-1423.
90. Ishii, K., M. Noda, and S. Uchiyama, *Mass spectrometric analysis of protein-ligand interactions*. Biophysics and physcobiology, 2016. **13**: p. 87-95.
91. Rostom, A.A., et al., *Detection and selective dissociation of intact ribosomes in a mass spectrometer*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(10): p. 5185-5190.
92. Uetrecht, C., et al., *High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(27): p. 9216-9220.
93. van Dyck, J.F., A. Konijnenberg, and F. Sobott, *Native Mass Spectrometry for the Characterization of Structure and Interactions of Membrane Proteins*, in *Membrane Protein Structure and Function Characterization: Methods and Protocols*, J.-J. Lacapere, Editor. 2017, Springer New York: New York, NY. **1635** : p. 205-232.
94. Sterling, H.J., et al., *Effects of Buffer Loading for Electrospray Ionization Mass Spectrometry of a Noncovalent Protein Complex that Requires High Concentrations of Essential Salts*. Journal of the American Society for Mass Spectrometry, 2010. **21**(6): p. 1045-1049.
95. Xia, Z.J., J.B. Degrandchamp, and E.R. Williams, *Native mass spectrometry beyond ammonium acetate: effects of nonvolatile salts on protein stability and structure*. Analyst, 2019. **144**(8): p. 2565-2573.
96. Zhang, H., et al., *Native Mass Spectrometry Characterizes the Photosynthetic Reaction Center Complex from the Purple Bacterium Rhodospirillum rubrum*. Journal of the American Society for Mass Spectrometry, 2017. **28**(1): p. 87-95.
97. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes*. Nature protocols, 2013. **8**(4): p. 639-651.
98. Hunter, C.L., A.G. Mauk, and D.J. Douglas, *Dissociation of Heme from Myoglobin and Cytochrome b5: Comparison of Behavior in Solution and the Gas Phase*. Biochemistry, 1997. **36**(5): p. 1018-1025.
99. Heck, A.J.R. and R.H.H. van den Heuvel, *Investigation of intact protein complexes by mass spectrometry*. Mass Spectrometry Reviews, 2004. **23**(5): p. 368-389.
100. Frey, B.L., et al., *Controlling gas-phase reactions for efficient charge reduction electrospray mass spectrometry of intact proteins*. Journal of the American Society for Mass Spectrometry, 2005. **16**(11): p. 1876-1887.
101. Robinson, C.V., et al., *Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly*. Journal of the American Chemical Society, 1996. **118**(36): p. 8646-8653.
102. Chen, F., et al., *Applying mass spectrometry to study non-covalent biomolecule complexes*. Mass Spectrometry Reviews, 2016. **35**(1): p. 48-70.
103. Gabelica, V. and E.D. Pauw, *Internal energy and fragmentation of ions produced in electrospray sources*. Mass Spectrometry Reviews, 2005. **24**(4): p. 566-587.
104. Wan, K.X., M.L. Gross, and T. Shibusawa, *Gas-phase stability of double-stranded oligodeoxynucleotides and their noncovalent complexes with DNA-binding drugs as revealed by collisional activation in an ion trap*. Journal of the American Society for Mass Spectrometry, 2000. **11**(5): p. 450-457.

105. Gupta, R., et al., *Comparison of the binding stoichiometries of positively charged DNA-binding drugs using positive and negative ion electrospray ionization mass spectrometry*. Journal of the American Society for Mass Spectrometry, 2004. **15**(10): p. 1382-1391.
106. Schmidt, A., U. Bahr, and M. Karas, *Influence of Pressure in the First Pumping Stage on Analyte Desolvation and Fragmentation in Nano-ESI MS*. Analytical Chemistry, 2001. **73**(24): p. 6040-6046.
107. Tahallah, N., et al., *The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument*. Rapid Communications in Mass Spectrometry, 2001. **15**(8): p. 596-601.
108. Lorenzen, K., et al., *Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases*. International Journal of Mass Spectrometry, 2007. **268**(2): p. 198-206.
109. Cubrilovic, D., et al., *Direct monitoring of protein–protein inhibition using nano electrospray ionization mass spectrometry*. Chemical Science, 2014. **5**(7): p. 2794-2803.
110. Cubrilovic, D., et al., *Quantifying protein-ligand binding constants using electrospray ionization mass spectrometry: a systematic binding affinity study of a series of hydrophobically modified trypsin inhibitors*. Journal of the American Society for Mass Spectrometry, 2012. **10** : p. 1768-1777.
111. van Duijn, E., *Current Limitations in Native Mass Spectrometry Based Structural Biology*. Journal of the American Society for Mass Spectrometry, 2010. **21**(6): p. 971-978.
112. Lössl, P., J. Snijder, and A.J.R. Heck, *Boundaries of Mass Resolution in Native Mass Spectrometry*. Journal of the American Society for Mass Spectrometry, 2014. **25**(6): p. 906-917.
113. Winkler, R., *ESIprot: a universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data*. Rapid Communications in Mass Spectrometry, 2010. **24**(3): p. 285-294.
114. Tseng, Y.-H., et al., *Interpreting the Charge State Assignment in Electrospray Mass Spectra of Bioparticles*. Analytical Chemistry, 2011. **83**(6): p. 1960-1968.
115. Guntner, A.S., et al., *Collision cross sections obtained with ion mobility mass spectrometry as new descriptor to predict blood-brain barrier permeation by drugs*. Scientific Reports, 2019. **9**(1): p. 19182.
116. Baird, M.A., et al., *Differential Ion Mobility Separations/Mass Spectrometry with High Resolution in Both Dimensions*. Analytical Chemistry, 2019. **91**(2): p. 1479-1485.
117. Laphorn, C., F. Pullen, and B.Z. Chowdhry, *Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: Separating and assigning structures to ions*. Mass Spectrometry Reviews, 2013. **32**(1): p. 43-71.
118. Kanu, A.B., et al., *Ion mobility-mass spectrometry*. Journal of Mass Spectrometry, 2008. **43**(1): p. 1-22.
119. Ben-Nissan, G. and M. Sharon, *The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes*. Current opinion in chemical biology, 2018. **42**: p. 25-33.
120. Hoffmann, W., G. von Helden, and K. Pagel, *Ion mobility-mass spectrometry and orthogonal gas-phase techniques to study amyloid formation and inhibition*. Current Opinion in Structural Biology, 2017. **46**: p. 7-15.
121. Meher, A.K. and Y.-C. Chen, *Polarization induced electrospray ionization mass spectrometry for the analysis of liquid, viscous and solid samples*. Journal of Mass Spectrometry, 2015. **50**(3): p. 444-450.
122. Hernández-Mesa, M., et al. *Ion Mobility Spectrometry in Food Analysis: Principles, Current Applications and Future Trends*. Molecules (Basel, Switzerland), 2019. **24**, DOI: 10.3390/molecules24152706.

123. Maurer, M.M., G.C. Donohoe, and S.J. Valentine, *Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity*. *Analyst*, 2015. **140**(20): p. 6782-6798.
124. Jurneczko, E. and P.E. Barran, *How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase*. *Analyst*, 2011. **136**(1): p. 20-28.
125. Bowman, A.P., R.R. Abzalimov, and A.A. Shvartsburg, *Broad Separation of Isomeric Lipids by High-Resolution Differential Ion Mobility Spectrometry with Tandem Mass Spectrometry*. *Journal of The American Society for Mass Spectrometry*, 2017. **28**(8): p. 1552-1561.
126. Levin, D.S., et al., *Using a nanoelectrospray-differential mobility spectrometer-mass spectrometer system for the analysis of oligosaccharides with solvent selected control over ESI aggregate ion formation*. *Journal of the American Society for Mass Spectrometry*, 2007. **18**(3): p. 502-511.
127. Bleiholder, C., et al., *Ion Mobility Spectrometry Reveals the Mechanism of Amyloid Formation of A $\beta$ (25–35) and Its Modulation by Inhibitors at the Molecular Level: Epigallocatechin Gallate and Scyllo-inositol*. *Journal of the American Chemical Society*, 2013. **135**(45): p. 16926-16937.
128. Marcinko, T.M., et al., *Structural Heterogeneity in the Preamyloid Oligomers of  $\beta$ -2-Microglobulin*. *Journal of Molecular Biology*, 2020. **432**(2): p. 396-409.
129. van Duijn, E., et al., *Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa*. *Molecular & cellular proteomics : MCP*, 2012. **11**(11): p. 1430-1441.
130. Hall, Z., et al., *Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics*. *Journal of the American Chemical Society*, 2012. **134**(7): p. 3429-3438.
131. Lanucara, F., et al., *The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics*. *Nature Chemistry*, 2014. **6**(4): p. 281-294.
132. Hodge, E.A., M.A. Benhaim, and K.K. Lee, *Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry*. *Protein Science*, 2020. **29**(4): p. 843-855.
133. Englander, S.W., et al., *Hydrogen exchange: the modern legacy of Linderstrøm-Lang*. *Protein science : a publication of the Protein Society*, 1997. **6**(5): p. 1101-1109.
134. Englander, J.J., N.R. Kallenbach, and S.W. Englander, *Hydrogen exchange study of some polynucleotides and transfer RNA*. *Journal of Molecular Biology*, 1972. **63**(1): p. 153-169.
135. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics*. *Biochimica et biophysica acta*, 2011. **1814**(8): p. 942-968.
136. Zhang, Z. and D.L. Smith, *Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation*. *Protein Science*, 1993. **2**(4): p. 522-531.
137. Katta, V. and B.T. Chait, *Conformational-Changes in Proteins Probed by Hydrogen-Exchange Electrospray-Ionization Mass-Spectrometry*. *Rapid Communications in Mass Spectrometry*, 1991. **5**(4): p. 214-217.
138. Baños-Mateos, S., et al., *High-fidelity DNA replication in Mycobacterium tuberculosis relies on a trinuclear zinc center*. *Nature Communications*, 2017. **8**(1): p. 855.
139. Shukla, A.K., et al., *Visualization of arrestin recruitment by a G-protein-coupled receptor*. *Nature*, 2014. **512**(7513): p. 218-222.
140. Lee, S., et al., *Allosteric inhibition of antiapoptotic MCL-1*. *Nature Structural & Molecular Biology*, 2016. **23**(6): p. 600-607.
141. Gallagher, E.S. and J.W. Hudgens, *Mapping Protein-Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry*. *Methods in Enzymology*, 2016. **566** : p. 357 - 404.

142. Puchades, C., et al., *Epitope mapping of diverse influenza Hemagglutinin drug candidates using HDX-MS*. Scientific Reports, 2019. **9**(1): p. 1-10.
143. Konermann, L., Y.-H. Pan J Fau - Liu, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics*. Chemical Society Review, 2011, **40** : p.1224-1234.
144. Moroco, J.A. and J.R. Engen, *Replication in bioanalytical studies with HDX MS: aim as high as possible*. Bioanalysis, 2015. **7**(9): p. 1065-1067.
145. Anderson, K.A.-O., E.A.-O. Gallagher, and J.A.-O. Hudgens, *Automated Removal of Phospholipids from Membrane Proteins for H/D Exchange Mass Spectrometry Workflows*. Analytical Chemistry, 2018. **90**(11): p.6409-6412.
146. Guttman, M. and K.K. Lee, *Isotope Labeling of Biomolecules: Structural Analysis of Viruses by HDX-MS*. Methods in enzymology, 2016. **566**: p. 405-426.
147. Konermann, L. and D.A. Simmons, *Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry*. Mass Spectrometry Reviews, 2003. **22**(1): p. 1-26.
148. Pan, J., S. Zhang, and C.H. Borchers, *Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry*. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2016. **1864**(12): p. 1801-1808.
149. Ahn, J., et al., *Assessing the reproducibility and specificity of pepsin and other aspartic proteases*. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(6): p. 1222-1229.
150. Zheng, J., et al., *High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5*. Nucleic Acids Research, 2014. **43**(2): p. 1216-1230.
151. Sørensen, L.A.-O. and R. Salbo, *Optimized Workflow for Selecting Peptides for HDX-MS Data Analyses*. Journal of the American Society of Mass Spectrometry, 2018. **29**(11): p. 2278-2281.
152. Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments*. Nature Methods, 2019. **16**(7): p. 595-602.
153. Mendoza, V.L. and R.W. Vachet, *Probing protein structure by amino acid-specific covalent labeling and mass spectrometry*. Mass Spectrometry Review, 2009. **28**(5): 785-815.
154. Gervasoni, P., et al., *Identification of the Binding Surface on  $\beta$ -Lactamase for GroEL by Limited Proteolysis and MALDI-Mass Spectrometry*. Biochemistry, 1998. **37**(33): p. 11660-11669.
155. Cohen, S.L., et al., *Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry*. Protein science : a publication of the Protein Society, 1995. **4**(6): p. 1088-1099.
156. Shields, S.J., et al., *Mass spectrometry and non-covalent protein-ligand complexes: Confirmation of binding sites and changes in tertiary structure*. Journal of the American Society for Mass Spectrometry, 2003. **14**(5): p. 460-470.
157. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes*. Journal of Structural Biology, 2011. **173**: p. 530-540.
158. Kramer, K., et al., *Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins*. Nature methods, 2014. **11**(10): p. 1064-1070.
159. Zhang, Q., et al., *Toward Building a Database of Bifunctional Probes for the MS3D Investigation of Nucleic Acids Structures*. Journal of the American Society for Mass Spectrometry, 2006. **17**(11): p. 1570-1581.
160. Merkley, E.D., et al., *Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances*. Protein science : a publication of the Protein Society, 2014. **23**(6): p. 747-759.
161. Kiosze-Becker, K., et al., *Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry*. Nature communications, 2016. **7**: p. 13248-13248.

162. Liu, C.H., et al., *Combining Proximity Labeling and Cross-Linking Mass Spectrometry for Proteomic Dissection of Nuclear Envelope Interactome*. Journal of Proteome Research, **19**(3): p.1109-1118.
163. Liu, F., et al., *Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry (vol 12, pg 1179, 2015)*. Nature Methods, 2017. **14**(5): p. 539-539.
164. Chavez, J.D., et al., *Systems structural biology measurements by in vivo cross-linking with mass spectrometry*. Nature Protocols, 2019. **14**(8): p. 2318-2343.
165. Schmidt, C. and H. Urlaub, *Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies*. Current Opinion in Structural Biology, 2017. **46**: p. 157-168.
166. Fischer, L. and J. Rappsilber, *Quirks of Error Estimation in Cross-Linking/Mass Spectrometry*. Analytical chemistry, 2017. **89**(7): p. 3829-3833.
167. Fischer, L. and J. Rappsilber, *False discovery rate estimation and heterobifunctional cross-linkers*. PloS one, 2018. **13**(5): p. e0196672-e0196672.
168. O'Reilly, F.J. and J. Rappsilber, *Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology*. Nature Structural & Molecular Biology, 2018. **25**(11): p. 1000-1008.
169. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes*. Journal of Mass Spectrometry, 2003. **38**(12): p. 1225-1237.
170. Tang, X. and J.E. Bruce, *A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies*. Molecular bioSystems, 2010. **6**(6): p. 939-947.
171. Gerrard, J.A., P.K. Brown, and S.E. Fayle, *Maillard crosslinking of food proteins I: the reaction of glutaraldehyde, formaldehyde and glyceraldehyde with ribonuclease*. Food Chemistry, 2002. **79**(3): p. 343-349.
172. Groehler, A., A. Degner, and N.Y. Tretyakova, *Mass Spectrometry-Based Tools to Characterize DNA-Protein Cross-Linking by Bis-Electrophiles*. Basic & Clinical Pharmacology & Toxicology, 2017. **121**: p. 63-77.
173. Marquié, C., *Chemical Reactions in Cottonseed Protein Cross-Linking by Formaldehyde, Glutaraldehyde, and Glyoxal for the Formation of Protein Films with Enhanced Mechanical Properties*. Journal of Agricultural and Food Chemistry, 2001. **49**(10): p. 4676-4681.
174. Tayri-Wilk, T., et al., *Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins*. bioRxiv, 2019: p. 820779.
175. Toews, J., et al., *Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions*. Analytica Chimica Acta, 2007 p(1873-4324).
176. Sutherland, B.W., J. Toews, and J. Kast, *Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions*. Journal of Mass Spectrometry, 2008. **43**(6): p. 699-715.
177. Hermanson, G.T., *Chapter 5 - Homobifunctional Crosslinkers*, in *Bioconjugate Techniques (Third Edition)*, G.T. Hermanson, Editor. 2013, Academic Press: Boston. p. 275-298.
178. Iacobucci, C., et al., *First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study*. Analytical chemistry, 2019. **91**(11): p. 6953-6961.
179. Haupt, C., et al., *Combining Chemical Cross-linking and Mass Spectrometry of Intact Protein Complexes to Study the Architecture of Multi-subunit Protein Assemblies*. Journal of visualized experiments : JoVE, 2017. **129**: p. 56747.
180. Linden, A., et al., *A cross-linking mass spectrometry approach defines protein interactions in yeast mitochondria*. Molecular & Cellular Proteomics, 2020: p. mcp.RA120.002028.

181. Chen, Z., et al., *Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes*. Wellcome open research, 2016. **1**: p. 5-5.
182. Fürsch, J., et al., *Proteome-Wide Structural Probing of Low-Abundant Protein Interactions by Cross-Linking Mass Spectrometry*. Analytical Chemistry, 2020. **92**(5): p. 4016-4022.
183. Nguyen-Huynh, N.-T., et al., *Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex*. Protein science : a publication of the Protein Society, 2015. **24**(8): p. 1232-1246.
184. Kalkhof, S. and A. Sinz, *Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters*. Analytical and Bioanalytical Chemistry, 2008. **392**(1): p. 305-312.
185. Paramelle, D., et al., *Chemical cross-linkers for protein structure studies by mass spectrometry*. Proteomics, 2013. **13**: p. 438-456.
186. Leavell, M.D., et al., *Strategy for selective chemical cross-linking of tyrosine and lysine residues*. Journal of the American Society for Mass Spectrometry, 2004. **15**(11): p. 1604-1611.
187. Swaim, C.L., J.B. Smith, and D.L. Smith, *Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides*. Journal of the American Society for Mass Spectrometry, 2004. **15**(5): p. 736-749.
188. Besselink, G.A.J., T. Beugeling, and A. Bantjes, *N-Hydroxysuccinimide-activated glycine-sepharose*. Applied Biochemistry and Biotechnology, 1993. **43**(3): p. 227.
189. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes*. Journal of Mass spectrometry, 2003. **38**: p(1076-5174).
190. Koolen, H.H., et al., *Imidate-based cross-linkers for structural proteomics: increased charge of protein and peptide ions and CID and ECD fragmentation studies*. Journal of the American Society for Mass Spectrometry, 2014. **25**(7): p. 1181-1191.
191. Hunter, M.J. and M.L. Ludwig, *The Reaction of Imidoesters with Proteins and Related Small Molecules*. Journal of the American Chemical Society, 1962. **84**(18): p. 3491-3504.
192. ThermoFisher. *Manuel d'utilisation de l'agent pontant EDC*. 2020 [cited 2020 12/10/20]; Available from: <https://www.thermofisher.com/order/catalog/product/77149#/77149>
193. Krieg Uc Fau - Walter, P., A.E. Walter P Fau - Johnson, and A.E. Johnson, *Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle*. (0027-8424).
194. Geyer, H., R. Geyer, and V. Pingoud, *A novel strategy for the identification of protein-DNA contacts by photocrosslinking and mass spectrometry*. Nucleic Acids Research, 2004. **32**(16): p. e132-e132.
195. Müller, F., A. Graziadei, and J. Rappsilber, *Quantitative Photo-crosslinking Mass Spectrometry Revealing Protein Structure Response to Environmental Changes*. Analytical Chemistry, 2019. **91**(14): p. 9041-9048.
196. ThermoFisher. *Manuel d'utilisation de l'agent pontant photoréactif NHS-Diazarine*. 2020 [cited 2020 12102020]; Available from: <https://www.thermofisher.com/order/catalog/product/26167#/26167>.
197. Belsom, A., et al., *Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology*. Molecular & Cellular Proteomics, 2016. **15**(3): p. 1105.
198. Sinz, A., et al., *Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology*. Protein Science, 2015. **24**(8): p. 1193-1209.
199. Shiota, T., T. Nishikawa S Fau - Endo, and T. Endo, *Analyses of protein-protein interactions by in vivo photocrosslinking in budding yeast*. Methods in molecular Biology, 2013. **1033**: p207-217.

200. Lössl, P. and A. Sinz, *Combining Amine-Reactive Cross-Linkers and Photo-Reactive Amino Acids for 3D-Structure Analysis of Proteins and Protein Complexes*. *Methods in molecular Biology*, 2016. **1384**: p109-127.
201. Huang, B.X., H.-Y. Kim, and C. Dass, *Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry*. *Journal of the American Society for Mass Spectrometry*, 2004. **15**(8): p. 1237-1247.
202. Hage, C., et al., *The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis*. *Angew Chem Int Ed Engl*, 2017. **56**(46): p. 14551-14555.
203. Rivera-Santiago, R.F., et al., *Probing structures of large protein complexes using zero-length cross-linking*. *Methods (San Diego, Calif.)*, 2015. **89**: p. 99-111.
204. Hwang, Y.-J., J. Granelli, and J. Lyubovitsky, *Effects of Zero-Length and Non-Zero-Length Cross-Linking Reagents on the Optical Spectral Properties and Structures of Collagen Hydrogels*. *ACS Applied Materials & Interfaces*, 2012. **4**(1): p. 261-267.
205. Yamashiro, S., et al., *Mammalian tropomodulins nucleate actin polymerization via their actin monomer binding and filament pointed end-capping activities*. *Journal of Biological Chemistry*, 2010. **285** (43), (1083-351X).
206. Zhao, C., et al., *Cross-linking mass spectrometry and mutagenesis confirm the functional importance of surface interactions between CYP3A4 and holo/apo cytochrome b(5)*. *Biochemistry*, 2012. **51**(47): p. 9488-9500.
207. Tang, X.T. and J.E. Bruce, *A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies*. *Molecular Biosystems*, 2010. **6**(6): p. 939-947.
208. Green, N.S., E. Reisler, and K.N. Houk, *Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers*. *Protein Science*, 2001. **10**(7): p. 1293-1304.
209. Trnka, M.J., et al., *Matching cross-linked peptide spectra: only as good as the worse identification*. *Molecular & cellular proteomics : MCP*, 2014. **13**(2): p. 420-434.
210. Sinz, A., S. Kalkhof, and C. Ihling, *Mapping protein interfaces by a trifunctional cross-linker combined with MALDI-TOF and ESI-FTICR mass spectrometry*. *Journal of the American Society for Mass Spectrometry*, 2005. **16**(12): p. 1921-1931.
211. Kaake, R.M., et al., *A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells*. *Molecular & cellular proteomics : MCP*, 2014. **13**(12): p. 3533-3543.
212. Chowdhury, S.M., et al., *Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry*. *Analytical Chemistry*, 2009. **81**(13): p.5524-5532.
213. Rey, M., et al., *eXL-MS: An Enhanced Cross-Linking Mass Spectrometry Workflow To Study Protein Complexes*. *Analytical Chemistry*, 2018. **90**(18): p.10707-10714.
214. Sinz, A., *Crosslinking Mass Spectrometry Goes In-Tissue*. *Cell Systems*, 2018. **6**(1): p. 10-12.
215. Nury, C., et al., *A novel bio-orthogonal cross-linker for improved protein/protein interaction analysis*. *Analytical Chemistry*, 2015. **87**(3): p.1853-1860.
216. Nielsen, T., et al., *Determination of protein conformation by isotopically labelled cross-linking and dedicated software: Application to the chaperone, calreticulin*. *International Journal of Mass Spectrometry*, 2007. **268**(2): p. 217-226.
217. Chen, Z.A., et al., *Quantitative Cross-linking/Mass Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant*. *Molecular & Cellular Proteomics*, 2016. **15**(8): p. 2769.
218. Muller, M.Q., et al., *Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS*. *Analytical Chemistry*, 2010. **82**(16): p. 6958-6968.
219. Bennett, K.L., et al., *Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular*

- protein contacts*. Protein science : a publication of the Protein Society, 2000. **9**(8): p. 1503-1518.
220. Sinz, A., *Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions*. Analytical and Bioanalytical Chemistry, 2017. **409**(1): p. 33-44.
  221. Tang, X., et al., *Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein-Protein Interactions*. Analytical Chemistry, 2005. **77**: p. 311-318.
  222. Arlt, C., et al., *Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker*. Analytical Chemistry, 2016. **88**(16): p. 7930-7937.
  223. Gutierrez, C.B., et al., *Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions*. Analytical Chemistry, 2018. **90**(12): p.7600-7607.
  224. Soderblom, E.J. and M.B. Goshe, *Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis*. Analytical Chemistry, 2006.**78**(23): p.8059-8068.
  225. Pan, D., et al., *Simplified Protocol for Cross-linking Mass Spectrometry Using the MS-Cleavable Cross-linker DSBU with Efficient Cross-link Identification*. Analytical Chemistry, 2018. **90**(18): p. 10990-10999.
  226. Gotze, M., et al., *Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis*. Journal of the American Society for Mass Spectrometry, 2015. **26**(1): p. 83-97.
  227. Petrotchenko, E.V. and C.H. Borchers, *Crosslinking combined with mass spectrometry for structural proteomics*. Mass Spectrometry Reviews, 2010. **29**(6): p. 862-876.
  228. Merkley, E.D., J.R. Cort, and J.N. Adkins, *Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze*. Journal of structural and functional genomics, 2013. **14**(3): p. 77-90.
  229. Nowakowski, A.B., W.J. Wobig, and D.H. Petering, *Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions*. Metallomics : integrated biometal science, 2014. **6**(5): p. 1068-1078.
  230. Wolstenholme, R., et al., *Study of latent fingerprints by matrix-assisted laser desorption/ionisation mass spectrometry imaging of endogenous lipids*. Rapid Communication in mass spectrometry, 2009. **23**(19): p.3031-3039.
  231. Dau, T., et al., *Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry*. Analytical Chemistry, 2019. **91**(7): p. 4472-4478.
  232. Belsom, A., et al., *Blind testing cross-linking/mass spectrometry under the auspices of the 11<sup>th</sup> CASP11*. Wellcome Open Research, 2016: p. 053173.
  233. Schilling, B., et al., *MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides*. Journal of the American Society for Mass Spectrometry, 2003. **14**(8): p. 834-850.
  234. Fritzsche, R., et al., *Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis*. Rapid Communications in Mass Spectrometry, 2012. **26**(6): p. 653-658.
  235. Leitner, A., et al., *Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography*. Molecular & cellular proteomics : MCP, 2012. **11**(3): p. M111.014126-M111.014126.
  236. Yu, C. and L. Huang, *Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology*. Analytical chemistry, 2018. **90**(1): p. 144-165.
  237. Iacobucci, C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions*. Nature Protocols, 2018. **13**(12): p. 2864-2889.

238. Iacobucci, C. and A. Sinz, *To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry*. Analytical Chemistry, 2017. **89**(15): p. 7832-7835.
239. Mendes, M.L., et al., *An integrated workflow for crosslinking mass spectrometry*. Molecular systems biology, 2019. **15**(9): p. e8994-e8994.
240. Yang, B., et al., *Identification of cross-linked peptides from complex samples*. Nature Methods, 2012. **9**(9): p. 904-906.
241. Stengel, F., R. Aebersold, and C.V. Robinson, *Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes*. Molecular & cellular proteomics : MCP, 2012. **11**(3): p. R111.014027-R111.014027.
242. Chen, Z.L., et al., *A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides*. Nature Communications, 2019. **10**, p1-11.
243. Götze, M., et al., *StavroX--a software for analyzing crosslinked products in protein interaction studies*. Journal of the American Society of Mass spectrometry, 2012. **23**(1): p.76-87.
244. Liu, F., et al., *Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification*. Nature Communications, 2017. **8**(1): p. 15473.
245. Leitner, A., T. Walzthoeni, and R. Aebersold, *Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline*. Nature Protocols, 2014. **9**(1): p. 120-137.
246. Lu, L., et al., *Identification of MS-Cleavable and Noncleavable Chemically Cross-Linked Peptides with MetaMorpheus*. Journal of proteome research, 2018. **17**(7): p. 2370-2376.
247. Hoopmann, M.R., et al., *Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes*. Journal of Proteome Research, 2015. **14**(5): p. 2190-2198.
248. Lima, D.B., et al., *SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis*. Journal of Proteomics, 2015. **129**: p. 51-55.
249. Yilmaz, Ş., et al., *Xilmass: A New Approach toward the Identification of Cross-Linked Peptides*. Analytical Chemistry, 2016. **88**(20): p. 9949-9957.
250. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nature Biotechnology, 2008. **26**(12): p. 1367-1372.
251. Courcelles, M., et al., *CLMSVault: A Software Suite for Protein Cross-Linking Mass-Spectrometry Data Analysis and Visualization*. Journal of proteome research, 2017. **16**(7): p. 2645-2652.
252. Graham, M.J., et al., *xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data*. bioRxiv, 2019: p. 561829.
253. Riffle, M., et al., *ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data*. Journal of Proteome Research, 2016. **15**(8): p. 2863-2870.
254. Zheng, C., et al., *XLink-DB: database and software tools for storing and visualizing protein interaction topology data*. Journal of proteome research, 2013. **12**(4): p. 1989-1995.
255. Kahraman, A., L. Malmström, and R. Aebersold, *Xwalk: computing and visualizing distances in cross-linking experiments*. Bioinformatics (Oxford, England), 2011. **27**(15): p. 2163-2164.
256. Kosinski, J., et al., *Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures*. (1095-8657 ).
257. Gong, Z., et al., *Protocol for analyzing protein ensemble structures from chemical cross-links using DynaXL*. Biophysics reports, 2017. **3**(4): p. 100-108.
258. Leitner, A., et al., *Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines*. Trends in Biochemical Sciences, 2016. **41**: p. 20-32.
259. Giese, S.H., L. Fischer, and J. Rappsilber, *A Study into the Collision-induced Dissociation (CID) Behavior of Cross-Linked Peptides*. Molecular & Cellular Proteomics, 2016. **15**(3): p. 1094-1104.

260. Walzthoeni, T., et al., *False discovery rate estimation for cross-linked peptides identified by mass spectrometry*. *Nature Methods*, 2012. **9**(9): p. 901-903.



## **Partie 2 : Résultats**

Dans cette seconde partie, nous avons souhaité présenter les résultats obtenus au cours de cette thèse. Dans un premier temps, nous avons étudié l'oligomérisation d'un coactivateur nucléaire par MS native. Puis, dans un second temps, nous avons optimisé au laboratoire la stratégie XL-MS sur trois édifice structuraux différents.

Cette partie se divise ainsi en 4 chapitres :

**Chapitre IV** : Etude de l'oligomérisation d'un coactivateur nucléaire par spectrométrie de masse native

**Chapitre V** : Développement de la spectrométrie de masse pour la caractérisation structurale de complexe biologique par pontage chimique (XL-MS)

**Chapitre VI** : Vers la caractérisation d'une protéine partiellement désordonnée par XL-MS : la protéine CBP

**Chapitre VII** : XL-MS, une aide à la compréhension des données de cryo-EM : vers la génération d'un modèle structural du complexe NuA4



## Chapitre IV : Etude de l'oligomérisation d'un coactivateur nucléaire par spectrométrie de masse native

Dans ce chapitre, nous illustrerons le potentiel de la MS native pour l'étude du comportement d'agrégation d'un fragment protéique le NCBD (Nuclear Coactivator Binding Domain) de la protéine CBP (CREB- Binding Protein). Après avoir optimisé les paramètres instrumentaux spécifiques au maintien des interactions non covalentes lors de l'analyse par ESI-MS, nous tenterons d'obtenir des informations telles que la stœchiométrie d'interaction ainsi que le suivi du processus d'auto-assemblage en fonction des conditions d'analyse.

Dans un second temps, nous étudierons la complexation entre NCBD et le domaine N-terminal de transactivation de la protéine P53 (TAD : Transactivation Domain) afin d'apporter des informations importantes sur la stœchiométrie d'interaction en complément aux hypothèses émises par nos collaborateurs à partir des résultats obtenus avec d'autres techniques analytiques.

Ce travail a été réalisé dans le cadre d'une collaboration avec l'équipe du Dr. Vladimir Torbeev du laboratoire de Chimie des Protéines au sein de l'Institut de Science et de l'ingénierie Supramoléculaires (ISIS – UMR7006). Le projet a démarré il y a un an et est toujours en cours.

### 1. Introduction

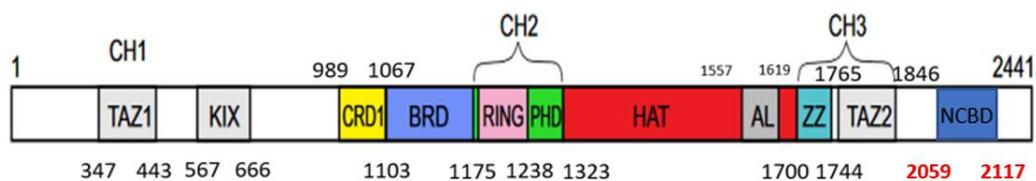
Dans un contexte pathologique, l'agrégation protéique peut conduire à la formation de fibres insolubles telles que celles impliquées dans la maladie d'Alzheimer<sup>[1]</sup> ou de Parkinson<sup>[2]</sup>. Comprendre comment s'effectue la transition structurale entre la protéine monomérique soluble et les fibres multimériques souvent riches en feuillet  $\beta$  est capital pour développer des stratégies diagnostiques et thérapeutiques pour le traitement de ces maladies<sup>[3]</sup>.

Dans notre cas, nous nous sommes intéressés à l'oligomérisation d'un polypeptide correspondant au domaine de liaison au Coactivateur de récepteur nucléaire (NCBD) de la protéine régulatrice CREB-Binding Protein (CBP) ainsi qu'à son interaction avec le domaine N-terminal de transactivation (TAD) de la protéine P53 par spectrométrie de masse.<sup>[4]</sup> Après avoir optimisé des conditions expérimentales appropriées à l'étude de complexes non covalents par MS, nous avons déterminé l'état d'oligomérisation de NCBD seul ou en présence de son partenaire TAD.

Peu de publications décrivent l'étude de l'oligomérisation des protéines par MS native en comparaison aux complexes hétéro-multimériques.<sup>[5-8]</sup> Ceci peut être lié à la difficulté de distinguer par la simple mesure de la masse moléculaire les complexes spécifiques des potentiels agrégats artéfactuels, ces derniers pouvant résulter de l'effet de concentration lors du processus d'ionisation/désolvatation en ESI.<sup>[9, 10]</sup> Des études ont notamment été réalisées sur la protéine amyloïde

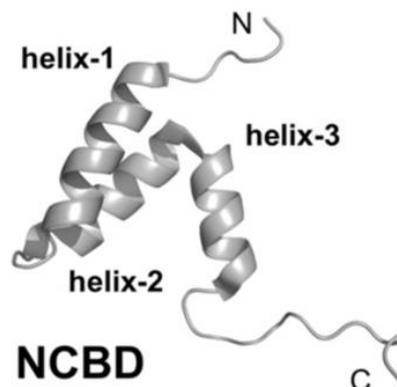
béta (A $\beta$ ) afin de comprendre son mécanisme d'agrégation et la relation possible avec des déficiences dans les tissus cérébraux.

La protéine de liaison CREB (CBP) est une protéine coactivatrice qui régule l'expression des gènes dans une pléthore de processus cellulaires.<sup>[11-14]</sup> en particulier lors de l'activation ou la répression de la transcription cellulaire régulée par les récepteurs nucléaires.<sup>[15]</sup> Cette protéine est constituée de 10 sous-domaines et de plus de 2400 acides aminés. Des études effectuées par DRX ont permis d'établir une première structure du noyau catalytique de CBP englobant les domaines BRD, CH2 (constitué des sous-domaines PHD et RING) et du domaine HAT. (cf. figure 46) <sup>[16]</sup> Alors que beaucoup d'efforts furent déployés sur le domaine HAT et sur le domaine BRD très peu d'études ont été réalisées sur le domaine CH3 présent sur le domaine C-terminal de la protéine CBP.<sup>[17]</sup> Il est difficile de caractériser la protéine entière car elle est très grande et structuralement désordonnée, ce qui rend délicate tant la production recombinante que l'analyse structurale.



**Figure 46 : Représentation des différents domaines de CBP. Extrait de [18]**

Le domaine NCBD est un des domaines du coactivateur CBP. Il est présent au niveau C-terminal de la protéine et présente une structure majoritaire hélicoïdale. Lors de l'interaction entre CBP et d'autres partenaires, la structure de NCBD est modifiée en formant une structure de 3 hélices qui diffèrent dans leurs arrangements topologiques en fonction des liaisons réalisées avec d'autres protéines <sup>[19]</sup>. NCBD est connu dans la littérature pour interagir avec d'autres coactivateurs nucléaires tels que la protéine p160 et la protéine coactivatrice pour la thyroïde et les récepteurs nucléaires (ACTR) qui sont respectivement des coactivateurs des récepteurs thyroïdiens et rétinoïdes. En particulier, une étude par RMN a mis en valeur que la structure secondaire du domaine en solution restait structurée lors des interactions avec la protéine ATCR. NCBD peut interagir avec SRC-1, une autre protéine coactivatrice ainsi qu'avec les différents domaines d'activations de p53.<sup>[20-22]</sup>



**Figure 47 : Représentation structurale de la protéine NCBD. Extrait de [23]**

Ce chapitre sera ainsi divisé en deux parties. Une première partie traitera essentiellement de l'auto-assemblage de NCBD alors qu'une seconde étudiera son interaction avec son partenaire TAD.

## 2. Etude de l'auto-assemblage de NCBD par MS native

### 2.1 Etat de l'art

L'auto-assemblage fibreux de protéine est impliqué dans une variété de fonctions biologiques incluant le transport moléculaire intracellulaire, la contraction musculaire, l'échafaudage et la connexion des tissus, dans la formation de caillots sanguins ... Outre les rôles bénéfiques, l'auto-assemblage des protéines peut être responsable d'une perte de fonction. Par exemple, un pliage aberrant des protéines peut entraîner l'agrégation de protéines en fibres amyloïdes insolubles structurellement riches en feuillet  $\beta$  et impliqués dans un certain nombre de maladies humaines, y compris les troubles neurodégénératifs.<sup>[24]</sup>

L'agrégation des protéines est un processus très complexe. Ce phénomène peut être initié à partir d'une protéine complètement repliée, par exemple, la  $\beta$ 2-microglobuline<sup>[25]</sup> ou la transthyréline<sup>[26]</sup>, ou à partir de problèmes intrinsèques des séquences tels que la  $\alpha$  synucléine<sup>[27]</sup> ou le peptide amyloïde  $\beta$  (A $\beta$ )<sup>[28]</sup>. Lors de l'agrégation des protéines, divers intermédiaires transitoires et espèces oligomériques précèdent l'assemblage des protéines en fibres amyloïdiques. De plus, il a été démontré dans des études antérieures que dans les maladies neurodégénératives les oligomères solubles possédaient une toxicité plus élevée que les fibres amyloïdes matures.<sup>[29]</sup> C'est pourquoi différentes approches analytiques ont dû être mises en place afin de mieux comprendre ce phénomène d'agrégation en identifiant les différents états conformationnels adoptés par ces protéines.<sup>[30]</sup>

Une des approches qui peut aider à explorer l'espace conformationnel des propriétés d'agrégation distinctes ou la prédisposition à s'assembler est de modifier la chiralité des chaînes polypeptidiques.<sup>[31]</sup> En effet, il a été démontré que l'incorporation spécifique d'acides aminés de conformation D dans les peptides A $\beta$  responsables de la maladie d'Alzheimer peut considérablement modifier les propriétés d'agrégation et permettre ainsi d'identifier les zones critiques responsables de l'amyloïdogénicité.<sup>[32]</sup> De plus, l'ajout d'énantiomères de forme D au sein des formes énantiomériques natives L des peptides amyloïdogéniques pourrait faciliter leur agrégation et la formation des fibres amyloïdes. Dans ce cas particulier, il a été démontré que les mélanges racémiques des peptides amyloïdes détruiraient les formes oligomériques les plus toxiques.<sup>[33]</sup> Vraisemblablement, l'accélération de l'auto-assemblage est entraînée par la formation de structure sous forme de feuillets  $\beta$  « ondulés » où les formes polypeptidiques L et D adoptent des conformations sous forme de feuillet.<sup>[34]</sup> Il a également été montré que les formes racémiques L/D des peptides sont thermodynamiquement plus stables que les formes énantiomériques purs.<sup>[35]</sup>

Dans ce projet, nous avons voulu utiliser la capacité de la MS native à déterminer les stœchiométries des différentes espèces présentes en solution pour suivre la formation des oligomères de NCBD en fonction des conditions expérimentales :

- Quelles sont les stœchiométries d'oligomérisation du L- et du D-NCBD natif ? Peut-on mettre en évidence la formation d'une structure mixte L/D-NCBD ?
- Les 3 hélices constituant NCBD sont-elles similairement impliquées dans l'oligomérisation ?

Les résultats présentés dans cette partie ont fait l'objet d'une publication dans le journal scientifique *Chemistry - A European Journal*.

Dans les paragraphes suivants, nous exposerons la mise en œuvre de ces expériences, en particulier un réglage instrumental contrôlé permettant de relier le spectre de masse natif à l'image des espèces présentes en solution.

## **2.2 La spectrométrie de masse native dans la caractérisation de protéine oligomérique : Optimisation des conditions expérimentales**

### **2.2.1 Mise au point de l'étude de NCBD par Native MS**

#### 2.2.1.1 Préparation des échantillons

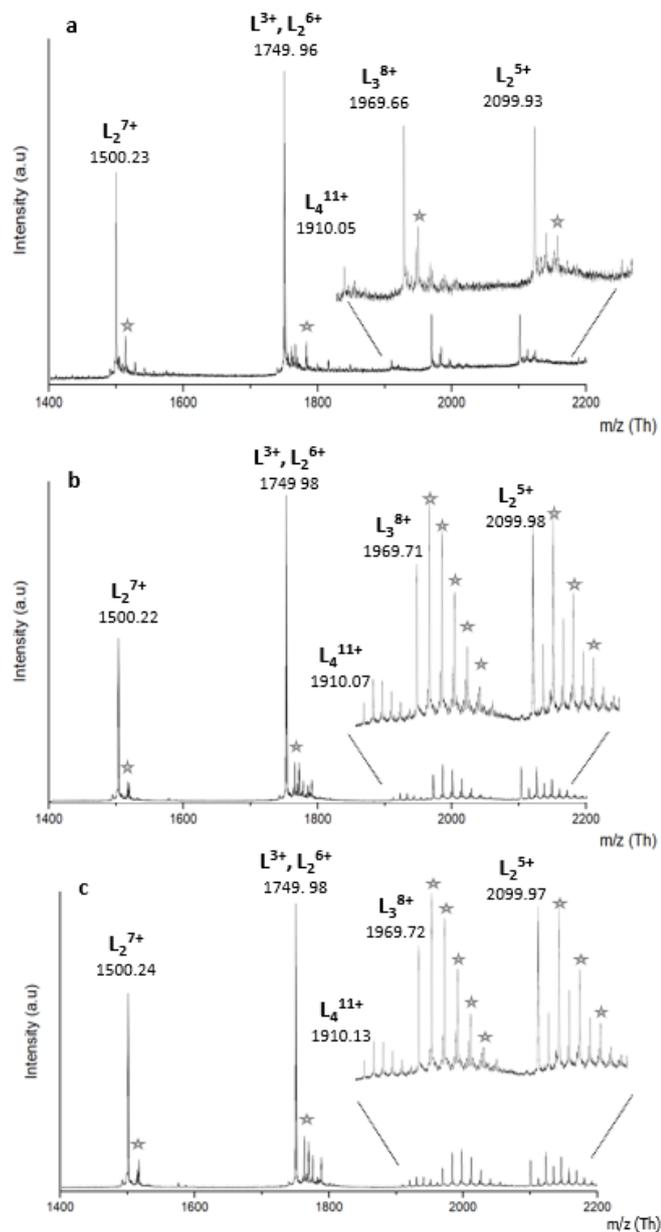
Comme nous l'avons vu dans la partie bibliographique, la 1<sup>ère</sup> étape consiste à transférer notre échantillon dans un tampon compatible à la fois avec le maintien de la structure de notre protéine en solution et avec la technique d'ESI. Dans notre cas, après synthèse, NCBD a directement été resolubilisé dans un tampon acétate d'ammonium (AcONH<sub>4</sub>, 100 mM pH 6,8). La séquence des protéines ou peptides sur lesquels nous avons travaillé est donnée sur la figure 48.

Peptide	Séquence <sup>†</sup>
L-NCBD	(H <sub>2</sub> N)-SALQDLLRTLKSPSSPQQQQVVLNILKSNPQLNleAAFIKQRTAKYVAN-(OH)
D-NCBD	(H <sub>2</sub> N)-SALQDLLRTLKSPSSPQQQQVVLNILKSNPQLNleAAFIKQRTAKYVAN-amide
L*-NCBD	(H <sub>2</sub> N)-SALQDLLRTLKSPSSPQQQQVVLNILKSNPQLNleAAFIKQRTAKYVANQP-amide
Helix-1	(H <sub>2</sub> N)-SALQDLLRTLKS-(OH)
Helix-2	(H <sub>2</sub> N)-SPQQQQVVLNILKS-(OH)
Helix-3	(H <sub>2</sub> N)-NPQLNleAAFIKQRTAKYVAN-(OH)

**Figure 48 : Séquence des échantillons de NCBD. Nle est la Norleucine incorporée à la place de la méthionine native pour éviter les phénomènes d'oxydation.**

La concentration de l'échantillon est un paramètre important puisqu'elle doit être en accord avec les constantes d'association des édifices en solution et permettre la détection des espèces multimériques avec une bonne intensité, sans pour autant induire la formation d'artéfacts d'associations lors du processus de désolvatation par ESI. Pour cela, des tests à des concentrations différentes de

protéine (10 $\mu$ M, 150 $\mu$ M et 300 $\mu$ M) ont été réalisés sur un instrument de type Q-TOF (Micro-TOF-QII, Bruker) (Cf figure 49). Avant analyse, les protéines ont été incubées pendant 5 min à 37 ° C et ont été infusées à un débit de 4  $\mu$ L/min à l'aide d'une pompe à seringue KDS100 (KD Scientific, Holliston, MA, États-Unis).



**Figure 49 : Spectre MS native montrant l'oligomérisation de L-NCBD à (a) 10 $\mu$ M, (b) 150 $\mu$ M et (c) 300 $\mu$ M. Le zoom montre la région en m/z où les ions trimériques et tétramériques ont été observés. Le profil isotopique des ions à charges multiples ainsi que les valeurs de  $\Delta m/z$  mesurées pour les adduits de TFA (marqués avec des étoiles) ont confirmé les états de charge de toutes les espèces et ont ainsi permis d'identifier chaque oligomère.**

Des ions reflétant la coexistence de monomères et de multi-mères (dimère L2, trimère L3, tétramère L4 et jusqu'au pentamère L5) ont été observés. Dans ces conditions non dénaturantes, le nombre d'états de charge par espèce est restreint. D'autre part, nous devons être attentifs au recouvrement des signaux identiques entre les monomères et oligomères multichargés (dimère 6-fois chargés et monomère 3-fois chargés par exemple). La résolution de notre instrument est suffisante pour assigner sans ambiguïté la charge des ions observés et en déduire la masse moléculaire de l'édifice correspondant. Les espèces monomériques sont détectées comme des ions triplement chargés à un  $m/z$  de 1749,98. Des états de charge impairs confirment la présence d'espèces dimériques à un  $m/z$  1500,22 et 2099,98. Les ions à  $m/z$  1969.71 sont observés comme un état chargé 8 + correspondant aux espèces trimériques, tandis que les ions chargés 11 + à  $m/z$  1910.07 donnent une masse moléculaire correspondant aux espèces tétramériques. Le profil isotopique a pu être exploité jusqu'aux états de charge 7-fois chargés. Les ions notés avec des étoiles (\*) correspondent à des adduits d'acide trifluoroacétique (TFA) pour lesquels les paires d'ions formées sont stables en phase gazeuse dans nos conditions douces d'analyse. Une augmentation de l'énergie interne des ions dans l'interface peut provoquer leur dissociation mais elle induit la dissociation des oligomères en priorité (Cf paragraphe suivant « optimisation des paramètres instrumentaux »). Bien que compliquant les spectres multichargés, la présence de ces adduits de TFA nous a en réalité aidé à attribuer les états de charge supérieurs à 7+ pour lesquels la résolution isotopique était trop faible.

Ces résultats montrent que l'oligomérisation de L-NCBD s'opère déjà à la concentration réduite de 10  $\mu\text{M}$ . Alors que peu d'ions correspondants aux trimères et tétramères sont observés à cette concentration de 10  $\mu\text{M}$  les spectres obtenus à 150 et 300 $\mu\text{M}$  sont très similaires. Afin de ne pas trop saturer le spectromètre en analyte, nous avons préféré travailler avec une concentration de 150 $\mu\text{M}$ .

#### 2.2.1.2 Optimisation des paramètres instrumentaux

Comme expliqué dans le chapitre précédent (Chapitre III.2), les paramètres MS doivent être optimisés afin de maintenir les protéines dans leur état natif en particulier lors du passage des échantillons en phase gazeuse et lors de leur transmission au sein de l'interface du spectromètre. Ainsi, la pression du gaz de nébulisation et du débit du gaz de séchage furent optimisés afin d'avoir un spray stable et de participer à la désolvatation des ions. Dans notre étude, le débit du gaz de séchage a été fixé à 3.5 L/min et la pression du gaz de nébulisation à 1 bar. Dans un second temps, les paramètres gérant le transfert des ions de la source vers l'analyseur ont été ajustés, en particulier la tension d'accélération (Capillary Exit), les paramètres de stockage (le « pre-pulse storage time ») et également les temps de transfert (« transfer time »).

La tension d'accélération doit être correctement optimisée afin de ne pas causer la rupture des liaisons non covalentes et ainsi dissocier les complexes. Dans notre cas, ce paramètre a été fixé à 340V. Le pre-pulse storage time ainsi que le transfer time sont des paramètres importants qui vont nous permettre de contrôler le transfert des ions selon leur rapport  $m/z$  (Cf figure 27) Pour la suite de nos expériences, le pre-pulse a été fixé à 36 $\mu\text{s}$  et le transfer time à 80 $\mu\text{s}$ .

Un autre paramètre crucial pour l'étude de complexes non-covalents est l'énergie de collision dans la source (In-Source Collision-Induced-Dissociation, ISCID). Ce paramètre permet d'améliorer la désolvatation des échantillons au sein de la source.

La figure 50 présente les spectres de masse enregistrés en fonction de la valeur de l'ISCID c'est-à-dire pour des énergies allant de 0 à 80 eV. Ils montrent comment une valeur de ISCID trop élevée induit la dissociation des trimères et tétramères en phase gazeuse, notamment à partir de 40eV. A 80eV, seuls NCBD monomérique et en moindre abondance dimérique ne sont détectés. En revanche, la distribution des espèces ne varie pas entre 0 et 40eV, ce qui indique qu'aucune dissociation en phase gazeuse ne se produit et que les stœchiométries observées reflètent bien les espèces présentes en solution. Cette valeur sera ajustée pour chaque nouvel échantillon, car les interactions non-covalentes impliquées d'un complexe à l'autre peuvent varier.

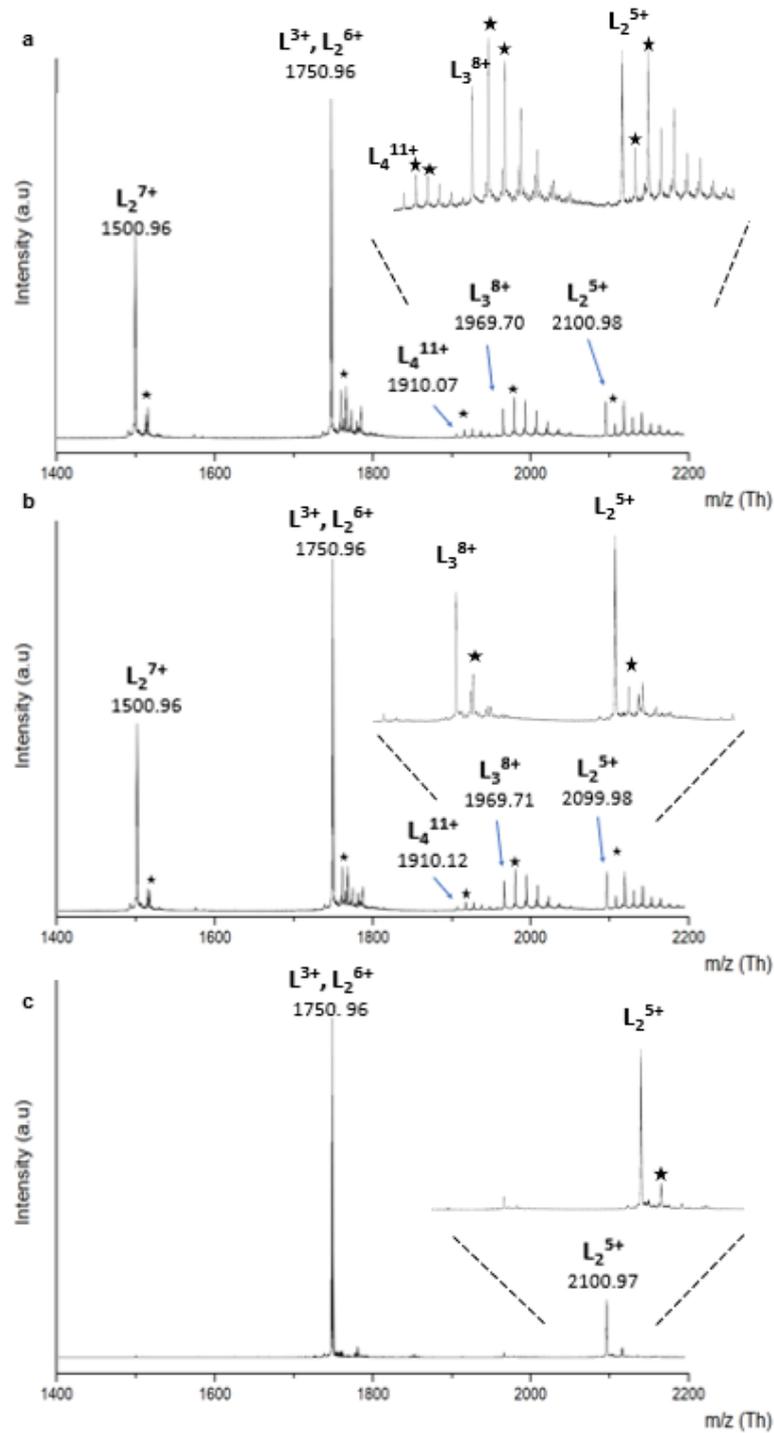


Figure 50 : Spectre MS natif montrant l'effet de l'énergie de collision in source (ISCID) sur le L-NCBD à (a) 0eV, (b) 40 eV et (c) 80 eV

### 2.2.1.3 Spécificité des espèces détectées par MS native

Afin de retirer de ces expériences des informations pertinentes, il est crucial de vérifier que les oligomères observés sur le spectre de masse résultent effectivement d'interactions préformées en solution et non d'une agrégation artéfactuelle liée au processus ESI. Pour cela, nous avons modifié les conditions expérimentales et contrôlé si ces variations se sont reflétées sur le spectre de masse. En 1<sup>er</sup> lieu, nous avons changé le solvant d'analyse. Ainsi, dans des conditions dénaturantes (H<sub>2</sub>O/CH<sub>3</sub>CN (50/50) avec 1% d'acide formique), les interactions non-covalentes sont rompues en solution et les oligomères doivent être dissociés. Dans ces conditions, seul le monomère n'est effectivement détecté (masse mesurée de 5246,9 Da ± 0,5 Da) ce qui suggère que les espèces multimériques observées ne sont pas l'œuvre d'un effet de concentration des ions en phase gazeuse.

Une seconde expérience de contrôle a consisté à modifier la séquence de fragments de NCBD et vérifier que l'oligomérisation résultante telle que mesurée sur le spectre de masse est affectée. Pour cela, nous avons injecté dans des conditions strictement identiques des peptides correspondants aux séquences N-terminale (Hélice I), C-terminale (Hélice III) et intermédiaire (Hélice II). Les spectres de masse révèlent un comportement très différent des trois hélices vis-à-vis de l'oligomérisation (voir publication ci-dessous) : l'hélice I montre un taux d'oligomérisation élevé (jusqu'à 10-mers) alors que l'hélice II est principalement observée sous forme de monomère et l'hélice III présente une agrégation modérée (jusqu'à 5-mers). Ainsi, les complexes non-covalents observés ne sont pas dus à un simple effet de condensation en phase gazeuse car la capacité à former des interactions est dépendante de la séquence peptidique. Les paramètres mis en place en MS native sont donc adaptés pour suivre de façon pertinente l'agrégation de la protéine NCBD.

### **2.3 Publication: Aggregation and amyloidogenicity of nuclear coactivator binding domain of CREB-binding protein dans Chemistry - A European Journal.**

Les conditions expérimentales validées, nous avons obtenus les résultats suivants :

- L-NCBD montre une tendance à s'oligomériser même à basse concentration
- Les différents oligomères peuvent être suivis individuellement par MS native
- Pas de différence significative dans le profil d'oligomérisation entre L- et D-NCBD
- Des oligomères mixtes L/D-NCBD ont été observés par MS native
- L'hélice I joue un rôle moteur dans l'oligomérisation de NCBD



# Aggregation and amyloidogenicity of nuclear coactivator binding domain of CREB-binding protein

Ana Maria Garcia,<sup>[a]</sup> Christophe Giorgiutti,<sup>[b]</sup> Youssef El Khoury,<sup>[c]</sup> Valentin Bauer,<sup>[a]</sup> Coralie Spiegelhalter,<sup>[d]</sup> Emmanuelle Leize-Wagner,<sup>[b]</sup> Petra Hellwig,<sup>[c,e]</sup> Noelle Potier<sup>[b]</sup> and Vladimir Torbeev<sup>\*[a]</sup>

[a] Dr. A. M. Garcia, V. Bauer, Dr. V. Torbeev  
ISIS (Institut de Science et d'Ingénierie Supramoléculaires) and icFRC (International Center for Frontier Research in Chemistry), University of Strasbourg and CNRS - UMR 7006  
8 allée Gaspard Monge, 67083 Strasbourg (France)  
Email: torbeev@unistra.fr

[b] C. Giorgiutti, Dr. E. Leize-Wagner, Dr. N. Potier  
Laboratory of Mass-Spectrometry of Interactions and Systems, University of Strasbourg and CNRS - UMR 7140  
1 rue Blaise Pascal, 67070 Strasbourg (France)

[c] Dr. Y. El Khoury, Prof. P. Hellwig  
Laboratory of Bioelectrochemistry and Spectroscopy, University of Strasbourg and CNRS - UMR 7140  
1 rue Blaise Pascal, 67070 Strasbourg (France)

[d] C. Spiegelhalter  
Imaging Center, IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire)  
INSERM-U964, University of Strasbourg and CNRS - UMR 7104  
1 rue Laurent Fries, 67404 Illkirch (France)

[e] Prof. P. Hellwig  
USIAS (University of Strasbourg, Institute for Advanced Study)  
5 allée du Général Rouvillois, 67083 Strasbourg (France)

Supporting information for this article is given via a link at the end of the document.

**Abstract:** Nuclear coactivator binding domain (NCBD) of transcriptional co-regulator CREB-binding protein (CBP) is an example of conformationally malleable protein that can bind to structurally unrelated protein targets and adopt distinct folds in the respective protein complexes. Here, we show that the folding landscape of NCBD contains an alternative pathway, which results in protein aggregation and self-assembly into amyloid fibers. The initial steps of such protein misfolding are driven by intermolecular interactions of its N-terminal  $\alpha$ -helix bringing multiple NCBD molecules in contact. Then, these oligomers undergo slow but progressive interconversion into  $\beta$ -sheet containing aggregates. To reveal the concealed aggregation potential of NCBD we used chemically synthesized mirror-image *D*-NCBD form. The addition of *D*-NCBD promoted self-assembly into amyloid precipitates presumably due to formation of thermodynamically more stable racemic  $\beta$ -sheet structures. The unexpected aggregation of NCBD needs to be taken into consideration given multitude of protein-protein interactions and resulting biological functions mediated by CBP.

## Introduction

Protein fibrous self-assembly is involved in a variety of biological functions including intracellular molecular transport, muscle contraction, scaffolding and connection of tissues, blood clot formation and many others.<sup>[1,2]</sup> Apart from beneficial roles, protein self-assembly can be responsible for a loss of function. For instance, an aberrant protein folding (or *protein misfolding*) can result in protein aggregation into insoluble amyloid fibers structurally rich in  $\beta$ -sheet content and implicated in a number of human diseases, including neurodegenerative disorders.<sup>[3]</sup> A detailed understanding of how soluble monomeric proteins

misfold and self-assemble into amyloid fibers is important for developing new strategies for the treatment of these diseases.<sup>[4-6]</sup>

Protein misfolding is a highly complex process.<sup>[7,8]</sup> The misfolding and aggregation into amyloids can be initiated from a fully folded protein, for example,  $\beta$ 2-microglobulin<sup>[9]</sup> or transthyretin,<sup>[10]</sup> or starting from intrinsically disordered sequences such as  $\alpha$ -synuclein<sup>[11]</sup> or amyloid- $\beta$  (A $\beta$ ) peptide.<sup>[12]</sup> Typically, conditions that lead to destabilization of fully folded forms are required to trigger misfolding and aggregation.<sup>[13,14]</sup> Various transient intermediates and oligomeric species precede the assembly into amyloid fibers.<sup>[15,16]</sup> In neurodegenerative diseases, the soluble oligomers were shown to possess higher toxicity than mature amyloid fibers.<sup>[16,17]</sup>

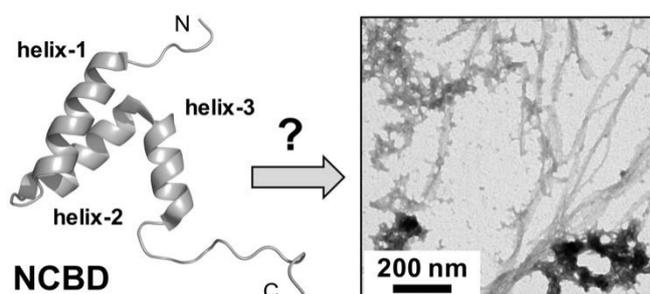
Previously, it was suggested that amyloid formation is not restricted to only a particular group of proteins but instead it can be a general feature of almost any polypeptide under the appropriate conditions.<sup>[18,19]</sup> This is supported by a high diversity of sequences and structural properties of proteins known to form amyloids, and suggests that the formation of amyloid fibers is a self-assembly phenomenon inherently dependent on a polypeptide backbone and to lesser extent on side-chains.<sup>[20-22]</sup> Various conditions are associated with protein misfolding and aggregation into amyloids, such as denaturing conditions (acidic pH, presence of chaotrops and surfactants), presence of transition metals (e.g., Cu<sup>2+</sup>) or organic solvents.<sup>[9,23,24]</sup> These factors may induce conformations in folded or intrinsically disordered proteins that are predisposed for aggregation.

Moreover, to aid experimental work several computational approaches are available to estimate the amyloid propensity of a given protein sequence. Short amyloidogenic sequence fragments (~4-6 residues long) were proposed to act as

'aggregation hotspots' defining the conversion into amyloids of an entire protein or a protein domain.<sup>[25,26]</sup> Based on this premise, such amyloid-prone sequences can be predicted by either empirical bioinformatic algorithms evaluating the physicochemical properties of amino acids and their possible contribution to amyloid formation or with the help of structure-based methods that use atomistic description of the polypeptide chains to provide a structural interpretation of aggregation into  $\beta$ -sheets.<sup>[25,26]</sup> While these approaches work well for short polypeptides, their accuracy is generally not sufficient to make precise prediction for any arbitrary protein. The mechanisms of amyloid self-assembly are at least as complex as for protein folding and experimental studies are needed to validate the propensities of a protein to aggregate or not.<sup>[27]</sup>

A complementary approach which may help exploring conformational space and identifying conformational states that have distinct aggregation properties and predisposition to self-assemble into amyloids is to modify the chirality of polypeptide.<sup>[28]</sup> It was shown that the site specific incorporation of *D*-amino acids into Alzheimer's disease-related A $\beta$ -peptide can drastically alter its aggregation properties allowing to identify locations in the sequence that are critical for amyloidogenicity.<sup>[29,30]</sup> Moreover, in many examples, adding mirror-image analogues (*D*-enantiomers) of amyloidogenic peptides or proteins to its native *L*-counterpart was reported to facilitate formation of amyloid fibers.<sup>[31-35]</sup> Probably the most striking case was reported for A $\beta$ (1-42) peptide, where the addition of *D*-enantiomer to the corresponding *L*-peptide accelerated the formation of amyloids and, moreover, attenuated the cytotoxicity commonly observed for A $\beta$ (1-42).<sup>[35]</sup> In this particular case, it was shown that racemic amyloids acted as a sink for soluble and most toxic oligomers. Presumably, the accelerated self-assembly is driven by the formation of the 'rippled'  $\beta$ -sheet structures where *L*- and *D*-polypeptides adopting extended  $\beta$ -strand conformations alternate.<sup>[36,37]</sup> Provided that such *L/D*-amyloids can be more thermodynamically stable than enantiomerically pure *L*- or *D*-amyloids,<sup>[32,33]</sup> the experiments in the presence of *D*-enantiomer may serve as an additional test for amyloidogenicity of a given peptide/protein.

In our previous work, we investigated the mechanism of complex formation between activation domain 1 (AD1) of transcriptional co-activator ACTR (also known as NCoA3 or SRC-3 from p160 family) and nuclear coactivator binding domain (NCBD) of CREB-binding protein (CBP or p300) which were prepared by chemical protein synthesis.<sup>[38,39]</sup> When performing chemical synthesis of NCBD by native chemical ligation of two peptide segments, even in the presence of 6 M guanidine hydrochloride, we frequently observed precipitate which we suspected can be composed of amyloids (Figure 1). Provided that CBP/p300 are highly important proteins regulating gene transcription<sup>[40]</sup> and, on the other hand, the diverse roles played by amyloid structures in different biological processes,<sup>[41]</sup> we investigated the aggregation of NCBD in detail including in the presence of mirror-image *D*-NCBD. Here we present the experimental data confirming the self-assembly of NCBD into amyloid-like fibers under different experimental conditions. Using various tools, we derived a possible molecular mechanism of earlier steps of NCBD aggregation into oligomers and their conversion into amyloids.



**Figure 1.** Largely  $\alpha$ -helical molten globule NCBD (PDB ID: 2KKJ) aggregates and transforms into amyloids. NCBD corresponds to [2066-2112]-fragment of 270 kDa transcriptional co-regulator CREB-binding protein (CBP).<sup>[40]</sup> CBP and its paralog p300 protein participate in multiple protein-protein interactions that fine-tune transcription of various genes. The objective of our study is to delineate the molecular details of NCBD aggregation mechanism. Three  $\alpha$ -helical fragments as well as N- and C-termini are labeled.

## Results

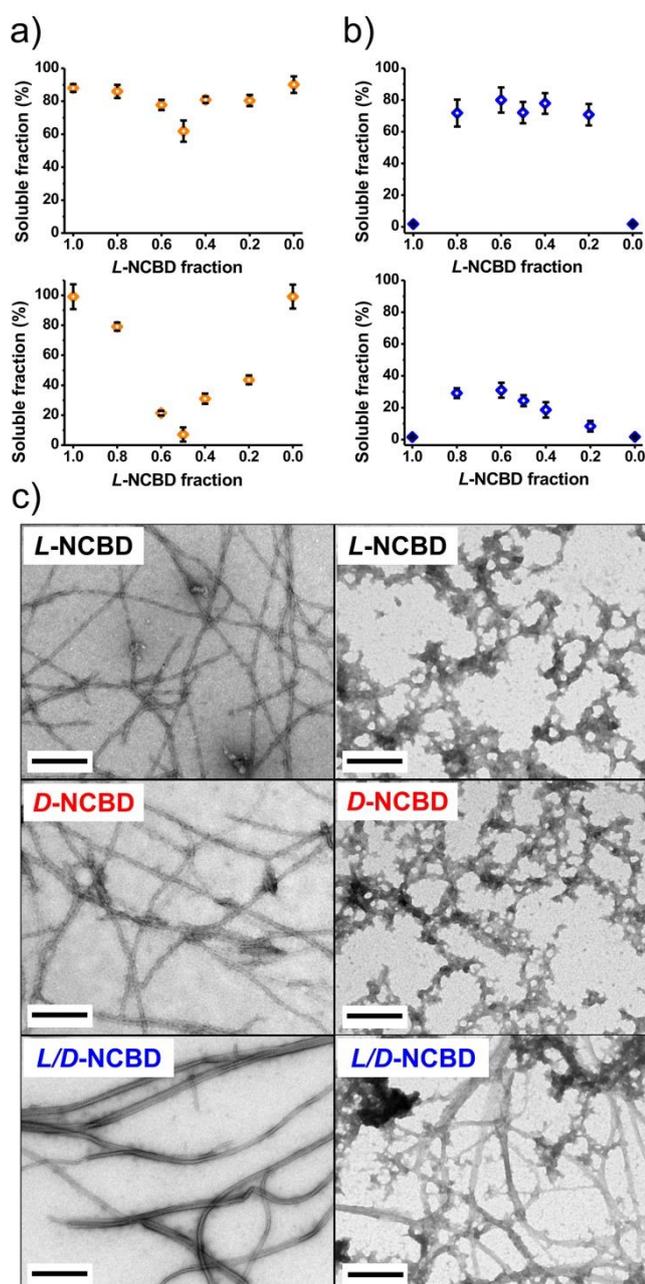
### Solubility and aggregation of NCBD

NCBD was previously shown to exhibit properties of molten globule possessing elements of secondary structure but lacking a well-defined tertiary structure typical for folded proteins.<sup>[42]</sup> In its isolated form, NCBD presents a compact structure with a high content of  $\alpha$ -helical conformation, while showing structural adaptability by forming distinct well-defined structures in complexes with other specific protein binding partners. This relatively small fragment of CBP is responsible for interactions with other relevant proteins such as transcriptional co-activators from p160 family<sup>[43]</sup> or interferon regulatory proteins,<sup>[44]</sup> through the formation of stable and well-defined tertiary structures at the interaction interface.

The *D*-NCBD enantiomer was prepared by chemical synthesis analogously to *L*-NCBD,<sup>[39]</sup> using commercially available *D*-amino acid building blocks (see Experimental Section). *D*-NCBD folds into mirror-image molten globule inferred through reciprocal circular dichroism (CD) spectra (Figure S1, Supporting Information). Our motivation to use *D*-NCBD along with *L*-NCBD was to bias the self-assembly of *L*-NCBD, so that if *L*-NCBD is predisposed to aggregate and self-assemble into amyloids, the addition of *D*-enantiomer will facilitate this process and help explore the aggregation landscape of *L*-NCBD at various conditions.

The samples containing *L*- and *D*-NCBD or mixtures thereof at different ratios and at two different pH values (acidic at pH 2.5 and neutral at pH 7.0) were incubated at 37 °C. The soluble fraction of NCBD in each sample was analyzed by HPLC as a function of the *L/D* ratio and time of incubation after ultrafiltration using the centrifugal devices with membrane of 0.2  $\mu$ m porosity (Figure 2 a,b). The solubility as a function of the *L/D* ratio suggests a distinct aggregation behavior for the two different conditions tested. At pH 2.5, *L/D* mixtures showed lesser quantities of soluble protein that passes through 0.2  $\mu$ m membrane than the single enantiomers; while at neutral pH, single enantiomers form aggregates that are trapped by membrane more readily than the different *L/D*-NCBD mixtures. These results confirm presence of chiral recognition effects between *L*- and *D*-NCBD. The differences in aggregation for mixtures of *L*- and *D*-NCBD can be caused by distinct conformations of polypeptides at different pH

values and thus distinct modes of interaction. The solubility diagrams in Figure 2 a,b resemble the solubility diagrams observed for crystals formed by enantiomers of small molecules, where the situation at pH 2.5 corresponds to solubility behavior of racemates that behave as racemic compounds whereas at neutral pH solubility curves remind those of conglomerates.<sup>[45-47]</sup> In the former case heterochiral interactions are more favorable and result in mixed racemic crystals from the mixture of enantiomers, whereas for the latter case the preference for homochiral assemblies enables the formation of enantiopure crystals from *L/D*-mixture.



**Figure 2.** (a,b) Solubility curves of NCBD enantiomers incubated at 37 °C at different *L*-NCBD/*D*-NCBD ratios in buffers at pH 2.5 (in a) and pH 7.0 (in b). Top panels correspond to one week incubation and bottom to two weeks of incubation. The plotted data are the average of two independent experiments

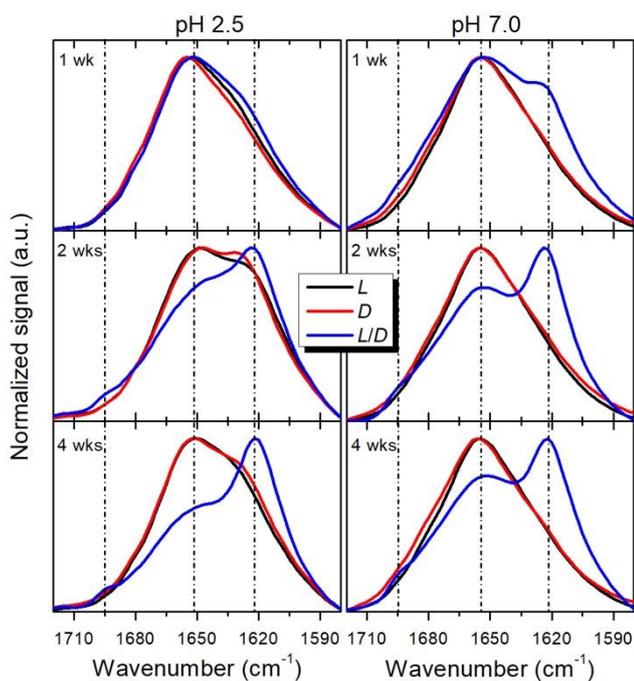
(*n*=2) where samples have been analyzed in triplicate. Standard deviations are represented by black bars. (c) TEM micrographs of NCBD samples after two weeks of incubation at pH 2.5 (left column) and at pH 7.0 (right column). Scale bars = 200 nm.

A pH-dependent behavior is also reflected in the morphology of the aggregates in the samples, as shown by TEM (transmission electron microscopy) (Figure 2 c and Figures S2-S9, Supporting Information). At acidic pH, the *L*- and *D*-NCBD single enantiomers assembled to form fibers with widths of  $12.1 \pm 1.3$  nm, a typical width reported for other amyloids.<sup>[48]</sup> When combinations of *L*- and *D*-enantiomers were incubated, fibers with larger widths of  $25.9 \pm 3.0$  nm and composed of bundles of several thinner fibers were frequently observed. Aggregation at neutral pH yielded initially amorphous precipitate, but upon two weeks of incubation fibrous structures could also be detected. Morphologically, these fibers are thinner in comparison to longer and thicker fibers observed at pH 2.5. Overall, fibrous self-assembly is promoted by the addition of the non-natural *D*-NCBD counterpart and, importantly, aggregation of NCBD generates morphologically different species depending on the pH of the media.

#### Monitoring structural changes during NCBD aggregation by FT-IR spectroscopy

In order to follow the secondary structure adopted by NCBD aggregates over the time, attenuated total reflection Fourier transform infrared (ATR-FT-IR) spectra were recorded for NCBD samples upon incubation at 37 °C at different timepoints (one, two and four weeks) (Figure 3). This spectroscopic technique is very reliable to observe the structural changes which accompany protein aggregation into  $\beta$ -sheets characterized by amide I stretching frequencies at around  $1630\text{ cm}^{-1}$ .<sup>[49]</sup> Moreover, it is suitable for the analysis of heterogeneous samples and requires little amount of material.<sup>[50]</sup> After one week of incubation, samples containing single NCBD enantiomers (*L*- or *D*-) showed a relatively sharp maximum in the amide I region ( $\sim 1650\text{--}1654\text{ cm}^{-1}$ ) at both pH 2.5 and 7.0, typically assigned to  $\alpha$ -helical structures, which is not surprising given high helical content observed for both *L*- and *D*-NCBD molten globule.<sup>[42]</sup> Notably, the single enantiomers at acidic pH showed a wider peak shifted towards the  $\beta$ -sheets region ( $\sim 1635\text{ cm}^{-1}$ ) in comparison with pH 7.0. The spectra of the racemic mixtures at both pHs were dominated by the  $\alpha$ -helix characteristic peak with a shoulder in the  $\beta$ -sheet region ( $\sim 1626\text{ cm}^{-1}$ ), more pronounced at neutral pH.

Over time, the spectra of samples at distinct pH values evolved differently. At pH 2.5, a band in the  $\beta$ -sheet region ( $\sim 1630\text{ cm}^{-1}$ ) in the spectra of all three samples (*L*-, *D*- or *L/D*-NCBD) increased in intensity, being more intense than the  $\alpha$ -helical signal for *L/D*-mixture after four weeks of incubation, which reflects a progressive transition from  $\alpha$ -helical to  $\beta$ -sheet structures. At neutral pH, the spectra of pure enantiomers were dominated by the  $\alpha$ -helical signal ( $\sim 1655\text{ cm}^{-1}$ ) even after four weeks, whereas racemic mixture showed a major peak centered at  $\sim 1623\text{ cm}^{-1}$  corresponding to  $\beta$ -sheet structure concomitant with the presence of the  $\alpha$ -helical structure. These data confirm the presence of  $\beta$ -sheet secondary structure of fibrous aggregates typical of amyloids and again point to the chiral effects of NCBD aggregation. The *D*-NCBD enantiomer acts as a promoter of  $\beta$ -sheet self-assembly and is essential for facile formation of amyloid fibers especially at neutral pH.



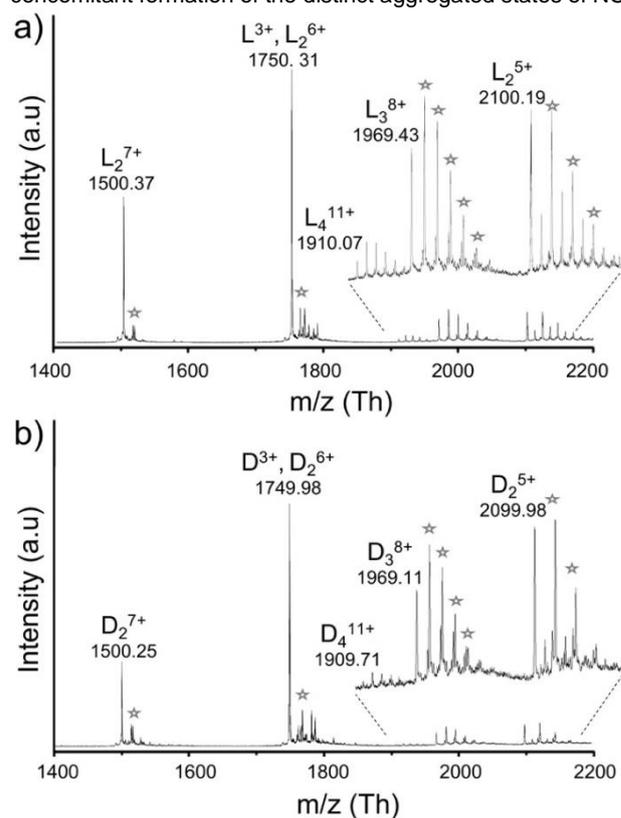
**Figure 3.** ATR-FT-IR spectra of *L*-NCBD, *D*-NCBD and *L/D*-NCBD racemic mixture in phosphate buffer at pH 2.5 (left) and pH 7.0 (right) after incubation at 37 °C for one, two and four weeks. The vertical dotted lines highlight the frequencies of the maxima at four weeks.

### Insights into the early steps of NCBD aggregation by native mass-spectrometry

To better understand the initial stages of self-association and aggregation of NCBD we used electrospray ionization mass spectrometry (ESI-MS) under native-like conditions. This technique enables to monitor association/dissociation of intact proteins and to get structural insights into the corresponding protein complexes by determining their stoichiometry, topology and dynamics.<sup>[51-55]</sup> In our case, *L*- and *D*-NCBD enantiomers as well as the racemic mixture were separately analyzed at different concentrations (10, 150 and 300  $\mu$ M). The native ESI mass spectra provide a charge state distribution of the species within a mass-over-charge ( $m/z$ ) range. By deconvolution of the charge states, the molecular masses of the complexes were determined, and their resulting stoichiometry deduced. The results showed the formation of clusters by NCBD (monomers, dimers, trimers and even tetramers) at any concentration (Figure 4, Figure S10). Determination of stoichiometry was achieved thanks to the resolution of isotopic profiles of all ions which led to unambiguous assignment of the charge states. All samples present the same pattern of aggregation, with a slight growth in the intensity of multimeric species upon increasing concentration but with no significant changes of coexisting stoichiometries. No multimeric interactions could be detected under denaturing conditions (50 % acetonitrile, 0.1 % formic acid), which strongly supports that the spectra reflect the different populations of NCBD oligomers present in solution rather than non-specific aggregation in the gas phase.

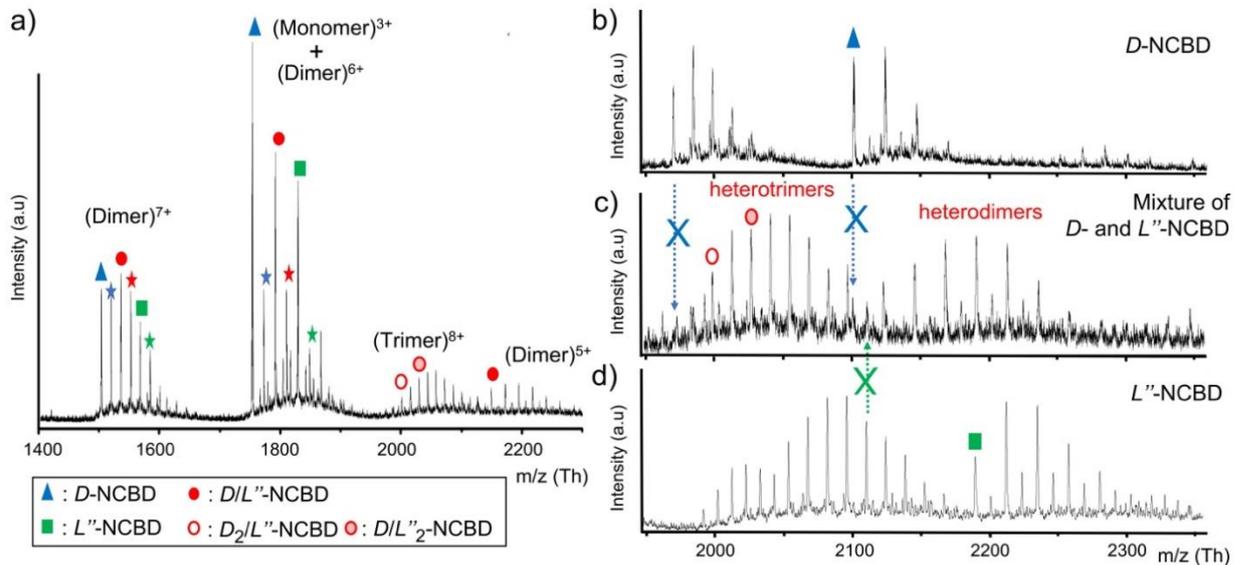
Due to the same molecular weight of *L*- and *D*-enantiomers of NCBD, native MS is not capable to distinguish between homo- and heterochiral interactions within the *L/D*-NCBD racemic

mixture and cannot give information about chiral composition of the aggregates. We envisaged that the addition of two more amino acids at the C-terminus of *L*-NCBD sequence, i.e., non-charged Gln-Pro native residues from the corresponding full-length CBP sequence will increase the molecular weight of *L*-NCBD by  $\sim$ 200 Da and, therefore, can be used as a tag to distinguish *L*-NCBD from *D*-NCBD by mass. In this way, native MS can be used to identify differences in the aggregation patterns between pure enantiomers and quasi-racemic mixture. We chemically synthesized this new *L*'-NCBD analogue (abbreviated here as *L*'-NCBD) (see Experimental Section). By native MS of an equimolar mixture of *L*'- and *D*-NCBD in addition to the expected *L*'- and *D*-NCBD dimers a third species at  $10725 \pm 1$  Da was observed and unambiguously confirmed the presence of a *L*'/*D* heterocomplex, thus demonstrating that heterochiral interactions occur in solution. No species revealing the presence of homochiral NCBD trimeric or tetrameric complexes were furthermore observed suggesting that heterochiral oligomerization is favored compared to homochiral multimerization at these conditions. (Figure 5). Because each species present in solution displays a different molecular mass, all oligomeric states could be followed individually. Native MS appeared to be a powerful technique for monitoring the concomitant formation of the distinct aggregated states of NCBD.



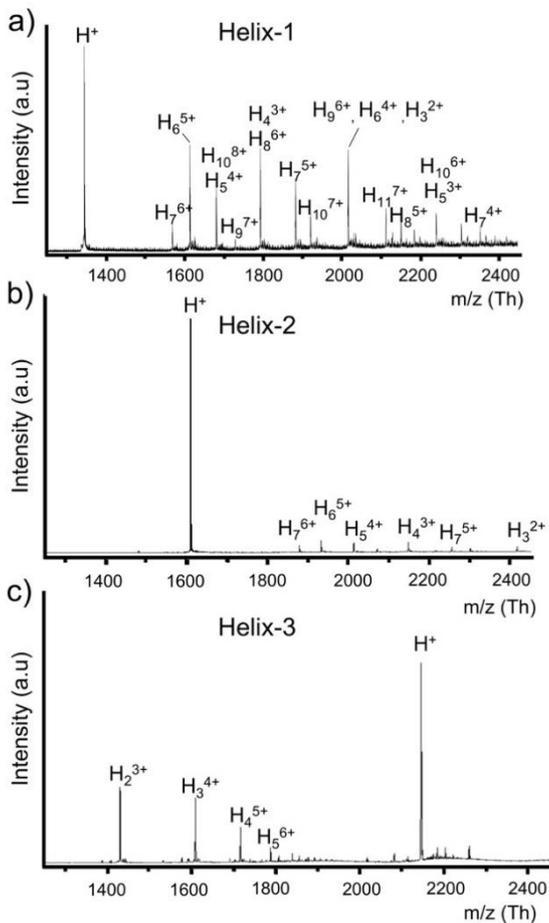
**Figure 4.** Native ESI mass-spectra of *L*-NCBD (a) and *D*-NCBD (b) in 100 mM AcONH<sub>4</sub>, pH 6.8. Resolution of the isotopic profiles allows to determine the charge states of all multiply-charged ions and thus to unambiguously assign the molecular masses of all species. Monomeric species (*L* and *D*) are detected as triply-charged ions at  $m/z$  1750.31 and 1749.98 respectively. Uneven charged states confirm the presence of dimeric species at  $m/z$  1500.37 and 2100.19 for *L*<sub>2</sub> and  $m/z$  1500.25 and 2099.98 for *D*<sub>2</sub>. Ions at  $m/z$  1969.43 and 1969.11 are observed as 8<sup>+</sup>-charged state corresponding to trimeric species (*L*<sub>3</sub> and *D*<sub>3</sub>), while 11<sup>+</sup>-charged ions at  $m/z$  1909.07 and 1909.71 yield a molecular mass

corresponding to tetrameric species ( $L_4$  and  $D_4$ ). The star symbols indicate trifluoroacetic acetate (TFA) adducts of proteins.



**Figure 5.** Evidence for self-assembly of heterochiral  $L/D$ -NCBD multimers by native ESI mass-spectrometry. (a) Mass-spectrum of an equimolar mixture of  $D$ - and two-extra amino acids modified  $L''$ -NCBD ( $L''$ ). (b-d) Zoom of the 2000-2350  $m/z$  region of the spectra and comparison between  $D$ -NCBD sample (in b),  $D/L''$ -NCBD mixture (in c), and  $L''$ -NCBD (in d). Homochiral ions at  $m/z$  1969.11 ( $D$ -NCBD homotrimer),  $m/z$  2099.98 ( $D$ -NCBD homodimer) or  $m/z$  2110.87 ( $L''$ -NCBD homotrimer) are not observed in the mass-spectrum of  $D/L''$ -NCBD mixture (in c). The star symbols indicate TFA adducts of proteins.

**Figure 6.** Characterization of the aggregation behavior of (a) **helix-1**, (b) **helix-2**, (c) **helix-3** by native ESI mass spectrometry in 100 mM AcONH<sub>4</sub>. Ions showing coexisting stoichiometries from monomeric (H) up to 11-mer ( $H_{11}$ ) are detected for **helix-1** while less pronounced aggregation is observed with **helix-2** and **helix-3**.



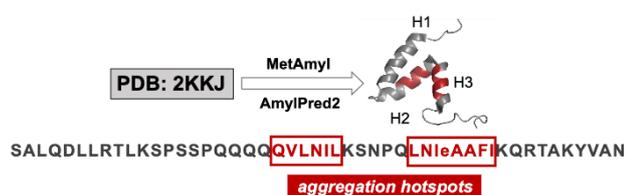
### Identifying the aggregation-prone segment in NCBD sequence

The structure of NCBD molten globule contains 3 helical fragments (Figure 1).<sup>[42]</sup> To identify which sequence fragment of NCBD is responsible for triggering aggregation, we prepared the three polypeptides corresponding to the three different helices (**helix-1**, **helix-2** and **helix-3**) observed in the structure of NCBD (PDB ID: 2KKJ).<sup>[42]</sup> When each helix was analyzed separately by native MS, only **helix-1** showed the pronounced aggregation revealing clusters ranging from dimers to undecamers, while **helix-2** was mostly observed as a monomer and **helix-3** demonstrated moderate aggregation propensity (Figure 6). Due to the instrument effects, the relative abundance of free peptide versus peptide clusters could not be directly extracted from the MS spectra. Indeed, all species display distinct  $m/z$  ratios and adopt different conformations, two parameters that are known to affect the ion intensities. Nevertheless, qualitatively the MS spectra of the three helical segments acquired under the same ESI-MS parameters illustrate the distinct aggregation properties of these peptides. The inspection of samples of the corresponding peptides incubated at pH 2.5 and 7.0 with TEM revealed similar trend of aggregation propensities, i.e., **helix-1** segment is particularly prone to aggregate (Figure S11).

The mixtures of full length NCBD with every helix were also studied (Figures S12-S14). According to native MS, **helix-1** again revealed its 'sticky' behavior by forming a variety of mixed clusters in high abundance under equimolar incubation conditions with NCBD, whereas this was not the case for **helix-2** and **helix-3**,

where major observed species were individual NCBD and **helix-2** or **helix-3**, respectively. The association pattern of *L*-NCBD or *D*-NCBD with the three helices showed no strong chiral preferences as the spectra for two enantiomers and the corresponding helices were similar. The measured molecular masses allowed to assign the stoichiometries of the various complexes coexisting in solution and it appeared that clusters formed by the interaction of one NCBD with several units of helices ( $L/H_n$ ) were favored compared to clusters implying dimerization of NCBD ( $L_2/H_n$ ). No higher stoichiometries of aggregated NCBD (trimers or tetramers) were detected anymore in the presence of **helix-1** which clearly shows that homologous NCBD multimers were displaced in favor of heterologous NCBD/**helix-1** multimers. (Figure S12). These data suggest that **helix-1** plays a critical role in the aggregation of the NCBD sequence, promoting the self-assembly between several NCBD units.

However, the computational predictors point out distinct parts in the sequence of NCBD that are aggregation-prone (Figure 7). Two applied consensus methods, *MetAmyl*<sup>[56]</sup> and *AmylPred2*,<sup>[57]</sup> yielded consistent results showing the two probable ‘aggregation hot spots’ localized in **helix-2** and **helix-3** of the NCBD sequence (*QVLN*IL and *LMAAF*I). Indeed, inspecting samples of **helix-2** and **helix-3** after four weeks incubation revealed amyloid-like fibers which supports computational prediction (see Figure S11). The apparent discrepancy of the aggregation-prone fragments in NCBD based on the experimental native MS measurements and the computational approaches can be explained by the fact that these methods probe different aspects of the aggregation. Native MS provides insights into the initial stages of NCBD multimerization via helix-helix interactions whereas computational predictors suggest which sequences are likely to be found in the core of the amyloid fiber after  $\alpha$ -helix-to- $\beta$ -sheet transition. Thus, the two methods are complementary and allow to propose a mechanistic model for sequential steps in NCBD aggregation where the initial aggregation is driven by promiscuous helix-helix contacts of N-terminal helix of NCBD and followed by the structural conversion involving segments located in different parts of the sequence.



**Figure 7.** Results of computational prediction of amyloidogenicity of NCBD using consensus methods *MetAmyl* and *AmylPred2*. Two aggregation hotspots are highlighted in red and located in **helix-2** (H2) and **helix-3** (H3) fragments. Note: Nle residue in the sequence is a nearly isosteric replacement of Met introduced to avoid oxidation of thioether moiety.

## Discussion

As discussed previously, amyloid formation is not restricted to a particular group of proteins but can occur for highly diverse sequences under proper conditions.<sup>[18-22]</sup> In this article, for the first

time we characterized the misfolding and aggregation of NCBD into amyloid-like structures. We tested different pH values and used *D*-NCBD to modulate and explore protein aggregation process. Indeed, the solubility and morphology of the aggregates is affected by the conditions of incubation (e.g., pH and fraction of *D*-NCBD in the mixture) (Figure 2, Figure S2-S9). According to FT-IR data, NCBD experiences  $\alpha$ -helix-to- $\beta$ -sheet conversion over time particularly at acidic pH (Figure 3). Notably, the *L/D*-racemic mixture presents a marked preference to adopt a  $\beta$ -sheet structure both at acidic and neutral pH.

Native MS data suggest that NCBD has a high propensity to form aggregates even at low concentrations (10  $\mu$ M) (Figure 4, Figure S10). The results obtained for the quasi-racemate of *L*'- and *D*-NCBD as an approach to distinguish *L*- and *D*-species in the mass-spectra suggest that enantiomeric proteins behave similarly and that the initial association of homo- and heterochiral multimers occurs rapidly in solution (Figure 5). In addition, the N-terminal helix in NCBD was identified being a trigger of the early stages of protein aggregation (Figure 6). This helix contains the leucine-rich *LXXLLXXL* sequence motif known to participate in various hydrophobic interactions mediating the formation of relevant biological complexes.<sup>[58,59]</sup>

Taking into account all the results, we hypothesize that the aggregation of NCBD starts from its self-assembly into  $\alpha$ -helical oligomers mediated through hydrophobic interactions of the ‘sticky’ N-terminal helical segment, as suggested by a high tendency of amphiphilic **helix-1** to multimerize. Modelling using CCBUILDER 2.0 software<sup>[60]</sup> showed that the formation of 2-, 3- and 4-helix bundles by **helix-1** segment can easily be achieved (Figure S15) and this is likely an explanation of the results of native MS measurements involving **helix-1** peptide.

On the other hand, NCBD is known to adopt different structures upon binding to different binding partners. The inspection of the available structures allowed to deduce a feasible molecular mechanism of NCBD oligomerization (Figure 8). We compared two conformational states of NCBD extracted from crystal structures of complexes with ACTR (PDB ID: 6SCQ)<sup>[39]</sup> and IRF3 (PDB ID: 1ZOQ).<sup>[44]</sup> In these two structures the biggest difference is the orientation of N-terminal helix (**helix-1**): in 6SCQ it forms tight contacts with the middle helical fragment (**helix-2**), whereas in 1ZOQ structure it is located in the groove and interacts with both middle (**helix-2**) and C-terminal (**helix-3**) helices. Apparently, the position of the middle helix in 1ZOQ also coincides with fragments of other proteins that interact with NCBD, such as helical fragment of ACTR (as in 6SCQ structure) or fragment of transactivation domain of p53 (PDB ID: 2I14). Notably, the structures of the latter two fragments are very different.

NCBD being a molten globule in free state implies conformational isomerism and heterogeneity. Thus, we envisaged that a minimal change of NCBD-6SCQ conformation such as  $\sim 180^\circ$  rotation around the axis of **helix-1** to expose its hydrophobic side to the solvent will make it predisposed to bind to another molecule of NCBD via interactions of solvent exposed **helix-1** and a promiscuous binding groove of a second NCBD molecule formed by **helix-2-helix3** (Figure 8). Furthermore, this conformational modification and docking can be repeated multiple times resulting in high order NCBD oligomers.

Over time, these  $\alpha$ -helical oligomers undergo isomerization into more stable  $\beta$ -sheet structures and convert into amyloid

fibers. Indeed, amyloid formation through  $\alpha$ -helical intermediates is described to happen for many proteins, including tau protein<sup>[61]</sup> and  $\alpha$ -synuclein.<sup>[62]</sup> Thus, the results of our work can be qualitatively described by *association-conformational conversion* model.<sup>[63]</sup> According to this model, amyloidogenic proteins first self-associate into soluble oligomers via hydrophobic interactions and then undergo structural conversion into  $\beta$ -sheet-rich amyloids. Mechanistically, the initial helical assemblies increase the local concentration necessary for the subsequent formation of intermolecular  $\beta$ -sheets.<sup>[64]</sup> The importance of high local concentration of proteins can be further illustrated by “droplets” formed due to liquid-liquid phase separation of various disordered proteins.<sup>[65]</sup> Conformational rearrangements within these compartments often lead to the formation of  $\beta$ -sheet “lumps” that propagate to ultimately yield more stable fibrillar aggregates with  $\beta$ -sheet structure.

## Conclusion

To summarize, we discovered that NCBD sequence is aggregation-prone and amyloidogenic. The predisposition of N-terminal helix of NCBD molten globule protein to mediate intermolecular self-assembly results in helical aggregates. They evolve towards more stable  $\beta$ -sheet structures which adopt fibrous amyloid-like morphologies particularly at low pH. At physiologically relevant neutral pH this effect can be highlighted in the presence of *D*-NCBD enantiomer. Thus, using *D*-NCBD helped to reveal the presence of an alternative pathway on the folding landscape of NCBD that is not obviously detected for enantiopure *L*-NCBD. These results are important given multiple studies reported previously for this protein domain that served as a classical example of molten globule that forms distinct folded structures upon complexation with different protein binders.<sup>[66]</sup> Further research is needed to explore the potential role of NCBD misfolding and aggregation in biological processes mediated by CBP.

## Experimental Section

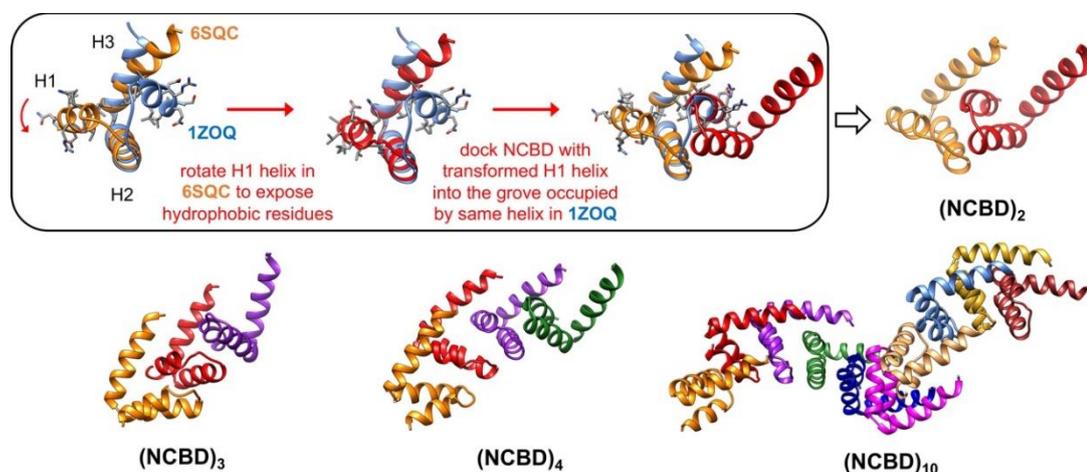
**Materials:** All solvents, chemicals, and reagents were purchased from commercial sources and used without further purification.

Fmoc-amino acids and resins for solid-phase peptide synthesis, the coupling reagents ***N,N'*-diisopropylcarbodiimide** (DIC) and **ethyl cyano(hydroxyimino)acetate** (Oxyrna pure), diisopropylethylamine, piperidine, triisopropylsilane (TIPS) and 1,4-dithiothreitol (DTT) were purchased from Iris Biotech (Marktredwitz, Germany) or Sigma Aldrich (St. Louis, Missouri, USA). Trifluoroacetic acid (Biograde) (TFA) was purchased from Halocarbon (Peachtree Corners, Georgia, USA). All other chemicals were purchased from Sigma-Aldrich. Nanosep MF centrifugal devices with Bio-Inert membrane of 0.2  $\mu$ m porosity were provided by Pall Corporation (Ann Arbor, Michigan, USA).

**Peptides synthesis:** The *L*-NCBD and *D*-NCBD peptides fragments (corresponding to the [2066-2112]-fragment of CBP (CREB-binding protein, UniProtKB - P45481 (CBP\_MOUSE)) were synthesized using a Liberty Blue microwave-assisted automatic synthesizer (CEM Corporation, USA), following a Fmoc/tBu-protecting group strategy, on either 0.1 or 0.25 mmol, as previously described.<sup>[39]</sup> For the synthesis of *L*-NCBD and the different *L*-NCBD fragments (helices 1-3), the preloaded Fmoc-Asn(Trt) and Fmoc-Ser(tBu) Wang resins (0.7 mmol/g and 0.65 mmol/g loading, respectively) were used, while the *L'*-NCBD and *D*-NCBD polypeptides were synthesized using Rink-Amide ChemMatrix resin (0.45 mmol/g loading). Peptides were cleaved and fully deprotected by treatment with a mixture of TFA/TIPS/H<sub>2</sub>O/DTT (92.5/2.5/2.5/2.5, v/v) for 2 hours (1 mL of mixture for 0.1 g of resin). Afterwards, the mixture was added onto cold diethyl ether (-20 °C) for precipitation, centrifuged and the precipitates were dissolved in H<sub>2</sub>O/CH<sub>3</sub>CN (1:1, v/v) containing 0.1% TFA and lyophilized. Pure peptide fractions (> 98 %) were obtained after purification by HPLC (Tables S1,S2 and Figures S16,S17).

**Measurement of protein concentrations:** After synthesis and purification proteins were obtained as TFA salts. In order to calculate concentrations of the samples used in various experiments (CD, TEM, solubility curves), the following procedure was applied: the nitrogen content of a sample of *L*-NCBD was determined by elemental analysis. This allowed to exactly calculate the mass portion of protein in the purified product (~80%). This sample was used to calibrate the analytical HPLC system. The derived calibration was used to determine the concentrations of other NCBD samples.

**Preparation of aggregated samples:** *L*-NCBD and *D*-NCBD samples were dissolved at peptide concentrations of 0.5 mM (exact concentration was determined by analytical HPLC). The prepared samples were incubated at 37 °C, at different *L*-NCBD/*D*-NCBD ratios and under two different conditions: 200 mM sodium citrate buffer pH 2.5 (or 200 mM sodium phosphate buffer pH 2.5) and 200 mM sodium phosphate buffer pH 7.0 (with NaN<sub>3</sub>, 0.05 %, w/v). The samples were prepared at the following *L*-NCBD/*D*-NCBD ratios: 100/0; 80/20; 60/40; 50/50; 40/60; 20/80; 0/100.



**Figure 8.** Model of the oligomerization of NCBD. Two distinct conformations of NCBD were extracted from crystal structures of NCBD with ACTR (in orange, PDB ID: 6SQC) and IRF3 (in blue, PDB ID:1Z0Q). Structural alignment was performed using region H2-H3 of 6SQC structure as a reference with MatchMaker tool of USCF-Chimera software.<sup>[67]</sup> Then, helix H1 was rotated to expose the hydrophobic side to the solvent making it aggregation prone. This new conformation was overlaid onto H1 helix of NCBD from 1Z0Q structure that is in the groove formed by H2 and H3 helices. Such a transformation effectively results in (NCBD)<sub>2</sub> dimer. This operation can be repeated again by docking a copy of aggregation-prone NCBD conformation to the free binding groove exposed in (NCBD)<sub>2</sub> leading to trimer and subsequently to tetramer and higher order oligomers (the largest shown is decamer).

**Analysis of the soluble peptide by HPLC:** The soluble fraction of NCBD for the samples prepared at different *L*-NCBD/*D*-NCBD ratios was analyzed by HPLC after incubation for one and two weeks. To this end, incubated samples were centrifuged at 10,000 rpm for 10 min using Nanosep MF centrifugal devices with membrane of 0.2 μm porosity (Pall Corporation). Then, 30 μL of the samples were diluted with 120 μL of milliQ water (final volume = 150 μL) and analyzed by analytical HPLC (in triplicate). The values for the soluble fraction of NCBD for each sample were obtained by integration of the HPLC traces using a calibration curve at OD 220 nm. Incubation and analysis were repeated twice. The percentages of soluble fraction for each sample were calculated based on the precise concentration of NCBD obtained by HPLC before incubation was started.

**Computational prediction of amyloid propensity:** NCBD sequence (PDB file: 2KKJ)<sup>[42]</sup> was screened with two different consensus methods available online, *MetAmyl*<sup>[66]</sup> and *AmylPred2*.<sup>[57]</sup> The first program computes the propensity of amyloid fibrillogenesis combining the following individual predictors: Pafig, SALSA, Waltz and FoldAmyloid. The score,  $p(x)$ , represents the probability for a fragment to form amyloid fiber, and it is calculated using a logistic regression model. *AmylPred2* gives as primary output the consensus prediction that represents the overlap of at least  $n/2$  of the different  $n$  methods selected by the user. The tool offers a maximum of 11 algorithms available: Aggrescan, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conformational Energy, NetCSSP, Pafig, SecStr, Tango and Waltz.

**Mass-spectrometry:** Liquid chromatography - mass spectrometry (LC-MS) spectra of peptides were recorded on a Thermo Scientific Accela UHPLC (Hypersil II GOLD column, 50 × 2.1 mm<sup>2</sup>, 1.9 μm) integrated with a Thermo Scientific LCQ XL linear ion trap mass-spectrometer. The mass spectra were recorded in the positive ionization mode. Obtained raw data were deconvoluted using MagTran (ver. 1.03) software.

**High performance liquid chromatography (HPLC):** Analytical reversed-phase HPLC of all peptides was performed on a Dionex Ultimate 3000 (Thermo Fisher) equipped with a UV detector, a column heater set to 40 °C and an autosampler. Analyzes were performed on a Reprosil-XR 120 C18 column (Dr. Maisch GmbH, particle size 3 μm, pore size 120 Å, dimensions 50 × 2.1 mm<sup>2</sup>) or a Kinetex XB C18 column (Phenomenex, particle size 2.6 μm, pore size 100 Å, dimensions 50 × 2.1 mm<sup>2</sup>) at a flow rate of 1 mL/min and a gradient of 2-50% of eluent B (0.08% TFA in CH<sub>3</sub>CN) in eluent A (0.1% TFA in H<sub>2</sub>O) within 4 min. Preparative reversed-phase HPLC was performed on a Shimadzu instrument equipped with a CBM-20A communication module, a SPD-M20A UV detector, a SIL-10AP autosampler and a FRC-10A fraction collector. *L*-NCBD, *D*-NCBD and *L*'-NCBD peptides were purified on a Kinetex C4 column (Phenomenex, particle size 5 μm, pore size 100 Å, dimensions 100 × 21.2 mm<sup>2</sup>) at a flow rate of 10 mL/min and a gradient of 20-40% of eluent B (0.08%TFA in CH<sub>3</sub>CN) in eluent A (0.1% TFA in H<sub>2</sub>O) within 90 min. After purification, pure fractions were identified by analytical HPLC and LC/MS, combined and lyophilized. *L*-NCBD fragments (**helices 1, 2 and 3**) were purified on a C18 column (Phenomenex, particle size 10 μm, pore size 300 Å, dimensions 250 × 21.00 mm<sup>2</sup>) at a flow rate of 10 mL/min and a gradient of 5-30 % of eluent B (0.08% TFA in CH<sub>3</sub>CN) in eluent A (0.1% TFA in H<sub>2</sub>O) within 40 min. After purification, pure fractions were identified by analytical HPLC and LC/MS and lyophilized.

**Circular dichroism (CD):** A 1 mm quartz cell was used on a Jasco J-810 spectropolarimeter. All samples were measured in quartz cuvettes with a thickness of 1 mm and a volume of 350 μL. The following

parameters were set for all scans performed: scanning speed 100 nm/min over a range of wavelengths from 185 to 280 nm, with 1 s integrations, bandwidth of 1 nm, data pitch 0.1 nm, temperature 25 °C. NCBD samples prepared in the two buffers used were incubated at 0.5 mM for two weeks. For measurements, they were diluted to 50 μM in the corresponding buffer. Every CD curve was obtained by averaging 5 scans and subtracting the background signal. The obtained ellipticities were transformed to mean residue molar ellipticities.

**TEM analysis:** TEM micrographs were acquired on a Philips CM12 electron microscope, operating at 80 keV. TEM grids (carbon-coated 300 mesh copper EM grids) were first exposed to the UV-ozone cleaner for 5 minutes. Then, grids were placed coated-side-down for 60 s onto sample drops (15 μL). The grids were then retrieved, washed with deionized water and stained with 2% (w/v) uranyl acetate for 20 s. Grids were air-dried and kept under vacuum before analysis. Average thickness of fibers was determined by considering at least 50 individual measurements using ImageJ software.

**FT-IR spectroscopy:** ATR-FT-IR experiments were performed on a Bruker Vertex 70 FT-IR spectrometer equipped with a Harrick-Diamond ATR. Residual TFA was removed from the samples by 10 mM HCl treatment and lyophilization (3 cycles). Each sample was left to dry under air. For each sample, ten spectra of 128 scans each (4000 to 700 cm<sup>-1</sup>, resolution: 4 cm<sup>-1</sup>) were recorded and averaged. All spectra were plotted using OriginPro 8.5 software.

**Native mass-spectrometry:** All proteins were solubilized in 100 mM ammonium acetate (AcONH<sub>4</sub>) pH 6.8 and analyzed at 150 μM, unless specified. MS data were acquired in the positive ionization mode on an electrospray time-of-flight mass spectrometer (MicroTOF-QII, BrukerDaltonic, Germany). Denaturing conditions measurements were conducted by injecting protein separately at a concentration of 10 μM in H<sub>2</sub>O/CH<sub>3</sub>CN (50/50) with 1% formic acid to check their sequence integrity and homogeneity. *L*- and *D*-NCBD were detected at 5248.4 Da ± 0.5 Da and 5248.9 ± 0.5 Da respectively, which is in good agreement with the theoretical molecular mass of the monomer (5248.8 Da). Native mass spectrometry analyses of the noncovalent complexes were carried out in 100 mM AcONH<sub>4</sub> pH 6.8. Proteins or complexes (150 μM) were incubated for 5 min at 37 °C and were infused at 4 μL/min through a KDS100 syringe pump (KD Scientific, Holliston, MA, USA). Electrospray source and transfer parameters were set to optimize the desolvation process while preventing dissociation of the noncovalent species. In particular, the pre-pulse time which controls a delay between transfer time and TOF pulser on for injection into the TOF analyzer was set at 36 μs. The collision energy in source (ISCID) was also carefully controlled and mass spectra were recorded as a function of the ISCID value from 0 to 100 eV. All spectra are shown at 40 eV where it was checked that no gas phase dissociation occurred. Finally, a manual interpretation of the MS spectra was then carried out in order to assign all charged states of the ions and infer the molecular masses of coexisting populations.

## Acknowledgements

C. G. would like to acknowledge the "Fondation Recherche Chimie" (FRC foundation) for his PhD thesis funding. This work has been funded by the European Research Council (ERC-2016-StG, grant number 715062-HiChemSynPro) and the LabEx CSC (ANR-10-LABX-0026\_CSC). Electron microscopy imaging was supported by the Imaging Center of IGBMC (ici.igbmc.fr).

**Keywords:** self-assembly • protein folding • mirror-image protein  
• amyloids • transcription coactivator

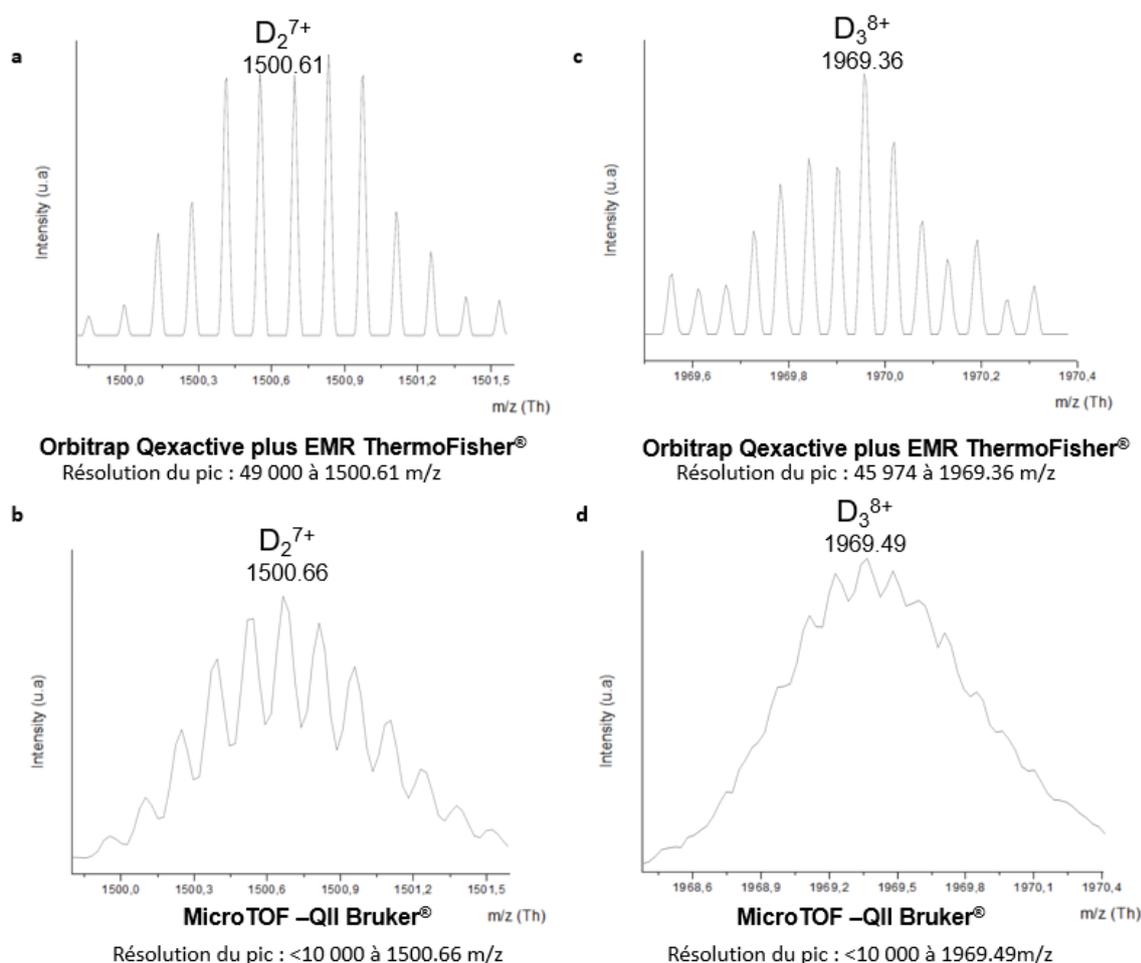
- [1] T. Scheibel, *Curr. Opin. Biotech.* **2005**, *16*, 427-433.
- [2] K. Gelse, E. Pöschl, T. P. Aigner, K.T. Hughes, *Adv. Drug Deliv. Rev.* **2003**, *55*, 1531-1546.
- [3] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2017**, *20*, 27-68.
- [4] D. S. Ong, J. W. Kelly, *Curr. Opin. Cell Biol.* **2011**, *23*, 231-238.
- [5] M. Jucker, L. C. Walker, *Nat. Rev.* **2013**, *501*, 45-51.
- [6] B. Cheng, H. Gong, H. Xiao, R. B. Petersen, L. Zheng, K. Huang, *Biochim. Biophys. Acta* **2013**, *1830*, 4860-4871.
- [7] R. Gallardo, N. A. Ranson, S. E. Radford, *Curr. Opin. Struct. Biol.* **2019**, *60*, 7-16.
- [8] R. Riek, D. S. Eisenberg, *Nature* **2016**, *539*, 227-235.
- [9] T. Eichner, S. E. Radford, *FEBS J.* **2011**, *278*, 3868-3883.
- [10] S. M. Johnson, R. L. Wiseman, Y. Sekijima, N. S. Green, S. L. Adamski-Werner, J. W. Kelly, *Acc. Chem. Res.* **2005**, *38*, 911-921.
- [11] H. A. Lashuel, C. R. Overk, A. Oueslati, E. Masliah, *Nat. Rev. Neurosci.* **2013**, *14*, 38-48.
- [12] D. M. Walsh, D. B. Teplow, *Prog. Mol. Biol. Transl. Sci.* **2012**, *107*, 101-124.
- [13] T. P. J. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, M. E. Welland, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10016-10021.
- [14] E. Serebryany, J. A. King, *Prog. Biophys. Mol. Biol.* **2014**, *115*, 32-41.
- [15] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, L. E. Kay, *Science* **2012**, *336*, 362-366.
- [16] P. Alam, L. Bousset, R. Melki, D. E. Otzen, *J. Neurochem.* **2019**, *150*, 522-534.
- [17] S. Li, D. J. Selkoe, *J. Neurochem.* **2020**, doi: 10.1111/jnc.15007.
- [18] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3590-3594.
- [19] F. Hane, *FEBS Letters* **2013**, *587*, 3617-3619.
- [20] M. Fändrich, C. M. Dobson, *EMBO J.* **2002**, *21*, 5682-5690.
- [21] M. T. Pastor, A. Esteras-Chopo, L. Serrano, *Prion* **2007**, *1*, 9-14.
- [22] N. S. de Groot, R. Sabate, S. Ventura, *Trends Biochem. Sci.* **2009**, *34*, 408-416.
- [23] G. Reddy, A. N. Muttathukattil, B. Mondal, *Curr. Opin. Struct. Biol.* **2020**, *60*, 101-109.
- [24] M. Calamai, F. Chiti, C. M. Dobson, *Biophys. J.* **2005**, *89*, 4201-4210.
- [25] A. B. Ahmed, A. V. Kajava, *FEBS Lett.* **2013**, *587*, 1089-1095.
- [26] G. Meric, A. S. Robinson, C. J. Roberts, *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8*, 139-59.
- [27] J. S. Ebo, N. Guthertz, S. E. Radford, D. J. Brockwell, *Curr. Opin. Struct. Biol.* **2020**, *60*, 157-166.
- [28] J. A. Raskatov, D. B. Teplow, *Sci Rep.* **2017**, *7*, 12433.
- [29] E. Y. Hayden, K. K. Hoi, J. Lopez, M. Inayathullah, M. M. Condron, D. B. Teplow, *Sci Rep.* **2017**, *7*, 12434.
- [30] C. J. Warner, S. Dutta, A. R. Foley, E. Chen, D. S. Kliger, J. A. Raskatov, *Chirality* **2017**, *29*, 5-9.
- [31] J. H. Fuhrhop, M. Krull, G. Büldt, *Angew. Chem. Int. Ed.* **1987**, *26*, 699-700.
- [32] K. Nagy-Smith, P. J. Beltramo, E. Moore, R. Tycko, E. M. Furst, J. P. Schneider, *ACS Cent Sci.* **2017**, *3*, 586-597.
- [33] J. M. Urban, J. Ho, G. Piester, R. Fu, B. L. Nilsson, *Molecules* **2019**, *24*, 1983.
- [34] K. Kar, I. Arduini, K. W. Drombosky, P. C. A. van der Wel, R. Wetzler, *J. Mol. Biol.* **2014**, *426*, 816-829.
- [35] S. Dutta, A. R. Foley, C. J. A. Warner, X. Zhang, M. Rolandi, B. Abrams, J. A. Raskatov, *Angew. Chem. Int. Ed.* **2017**, *56*, 11506-11510.
- [36] L. Pauling, R. B. Corey, *Proc. Nat. Acad. Sci. USA* **1951**, *39*, 253-256.
- [37] V. Torbeev, M. Grogg, J. Ruiz, R. Boehringer, A. Schirer, P. Hellwig, G. Jeschke, D. Hilvert, *J. Pept. Sci.* **2016**, *22*, 290-304.
- [38] B. Schmidtgall, O. Chaloin, V. Bauer, M. Sumyk, C. Birck, V. Torbeev, *Chem. Commun.* **2017**, *53*, 7369-7372.
- [39] V. Bauer, B. Schmidtgall, G. Gögl, J. Dolenc, J. Osz, Y. Nominé, C. Kostmann, A. Cousido-Siah, A. Mitschler, N. Rochel, G. Travé, B. Kieffer, V. Torbeev, ChemRxiv, doi: 10.26434/chemrxiv.10113128.v1.
- [40] N. Vo, R. H. Goodman, *J. Biol. Chem.* **2001**, *276*, 13505-13508.
- [41] D. Otzen, R. Riek, *Cold Spring Harb. Perspect. Biol.* **2019**, *11*, a033860.
- [42] M. Kjaergaard, K. Teilum, F. M. Poulsen, *Proc. Nat. Acad. Sci. USA* **2010**, *107*, 12535-12540.
- [43] S. J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H. J. Dyson, R. M. Evans, P. E. Wright, *Nature* **2002**, *415*, 549-553.
- [44] R. T. Lin, C. Heylbroeck, P. M. Pitha, J. Hiscott, *Mol. Cell Biol.* **1998**, *18*, 2986-2996.
- [45] J. Jacques, A. Collet, S. H. Wilen, Enantiomers, Racemates and Resolutions. Krieger Publishing Company, Malabar (Florida), 1994.
- [46] S. Srisanga, J. H. ter Horst, *Cryst. Growth Des.* **2010**, *10*, 1808-1812.
- [47] G. Coquerel, *J. Pharm. Pharmacol.* **2015**, *67*, 869-878.
- [48] F. T. Senguen, N. R. Lee, X. Gu, D. M. Ryan, T. M. Doran, E. A. Anderson, B. L. Nilsson, *Mol. BioSyst.* **2011**, *7*, 486-496.
- [49] A. Barth, C. Zscherp, *Q. Rev. Biophys.* **2002**, *35*, 369-430.
- [50] E. Goormaghtigh, J.-M. Ruyschaert, V. Raussens, *Biophys. J.* **2006**, *90*, 2946-2957.
- [51] A. J. R. Heck, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5-13.
- [52] S. K. Chaturvedi, M. K. Siddiqi, P. A. Rizwan, H. Khan, *Process Biochem.* **2016**, *51*, 1183-1192.
- [53] M. Marcinko, C. Liang, S. Savinov, J. Chen, R. W. Vachet, *J. Mol. Biol.* **2020**, *432*, 396-409.
- [54] W. Tong, G. Wang, *Methods* **2018**, *144*, 3-13.
- [55] S. Mehmood, T. M. Allison, C. V. Robinson, *Annu. Rev. Phys. Chem.* **2015**, *66*, 453-74.
- [56] M. Emily, A. Talvas, C. M. Delamarche, *PLoS ONE* **2013**, *8*, e79722.
- [57] A. C. Tsois, N. C. Papandreou, V. A. Iconomidou, S. J. Hamodrakas, *PLoS ONE* **2013**, *8*, e54175.
- [58] E. M. McInerney, D. W. Rose, S. E. Flynn, S. Westin, T. M. Mullen, A. Kronos, J. Inostroza, J. Torchia, R. T. Nolte, N. Assa-Munt, M. V. Milburn, C. K. Glass, M. G. Rosenfeld, *Genes Dev.* **1998**, *12*, 3357-3368.
- [59] J. M. Voegel, M. J. Heine, M. Tini, V. Vivat, P. Chambon, H. Gronemeyer, *EMBO J.* **1998**, *17*, 507-519.
- [60] C. W. Wood, D. N. Woolfson, *Protein Sci.* **2018**, *27*, 103-111.
- [61] R. Kunjithapatham, F. Y. Oliva, U. Doshi, M. Pérez, J. Avila, V. Muñoz, *Biochemistry* **2005**, *44*, 149-156.
- [62] D. Ghosh, P. K. Singh, S. Sahay, N. N. Jha, R. S. Jacob, S. Sen, A. Kumar, R. Riek, S. K. Maji, *Sci. Rep.* **2015**, *5*, 9228.
- [63] R. H. Walters, R. M. Murphy, *J. Mol. Biol.* **2011**, *412*, 505-519.
- [64] A. Abedini, D. P. Raleigh, *Protein Eng. Des. Sel.* **2009**, *22*, 453-459.
- [65] C. P. Brangwynne, P. Tompa, R. V. Pappu, *Nat. Phys.* **2015**, *11*, 899-904.
- [66] H. J. Dyson, P. E. Wright, *J. Biol. Chem.* **2016**, *291*, 6714-6722.
- [67] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

## 2.4 Difficultés de l'identification des ions multichargés :

Afin de valider nos interprétations sur la définition des états de charge (et confirmer ainsi l'identité de chaque oligomère) mais aussi afin de mieux distinguer d'éventuelles superpositions de profils isotopiques, nous avons réalisé quelques expériences sur un spectromètre de masse à haute résolution de type Orbitrap Qex+ EMR (ThermoFisher Scientific). Interrompues suite à la crise sanitaire Covid, ces analyses sont prometteuses et démontrent l'intérêt à disposer de haute résolution, même en présence de complexes de petites tailles.

La figure 51 montre un zoom sur le profil isotopique des deux mêmes ions obtenus à partir du même échantillon analysé sur un instrument de type Orbitrap EMR (fig. a et c) et sur un ESI-QTOF (fig. b et d). Il s'agit des ions caractéristiques du dimère (chargé 7+ à m/z 1500,66) et du trimère (chargé 8+ à m/z 1969,49). Alors que la charge ne peut être définie pour l'ion chargé 8+ sur le spectre QTOF, elle est lisible sans ambiguïté à partir des données Orbitrap.

Une telle résolution s'avérerait utile pour distinguer les complexes hétérotrimériques de type D<sub>2</sub>/ L'' -NCBD et D/L<sub>2</sub>'' -NCBD notamment.



**Figure 51 : Comparaison de résolution entre une analyse MS native réalisée sur un microTOF de Bruker (b et d) et une orbitrap Qexactive plus EMR (a et c). A l'aide de la haute résolution, il a été possible de déterminer l'état de charge des ions sans ambiguïté afin de confirmer l'identité de chaque oligomère.**

## **2.5 Conclusion et Perspectives**

En conclusion, les expériences par MS native ont permis de confirmer que le domaine NCBD est sujet à l'agrégation, et ce, rapidement en solution. Dans des conditions d'analyse contrôlées, il a été possible de suivre les différents états d'oligomérisation individuellement et il en est ressorti que l'hélice N-terminale était prédisposée à médier l'autoassemblage intermoléculaire des agrégats hélicoïdaux de la protéine NCBD.

Après cette 1<sup>ière</sup> étape d'agrégation, les protéines vont évoluer vers une structure plus stable sous forme de feuillet- $\beta$  qui peut elle-même évoluer vers une structure fibreuse de type amyloïdogénique en particulier à un pH acide. A un pH neutre, cet effet peut être mis en évidence en présence de l'énantiomère D-NCBD.

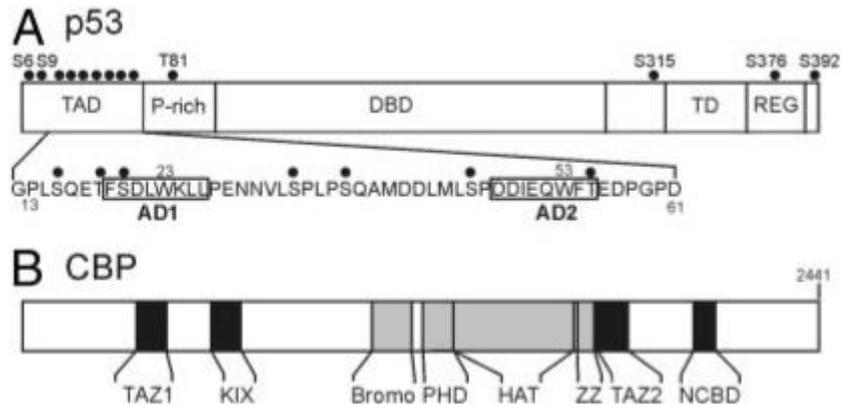
Le suivi d'un échantillon de L-NCBD par MS native sur une période de 4 semaines n'a révélé aucun changement dans le profil d'oligomérisation. Mais une perspective intéressante serait de suivre l'évolution de ce changement de conformation par MS native couplée à la mobilité ionique. De plus, des expériences visant à étudier l'oligomérisation de NCBD à pH acide sont envisagées.

## **3. Etude des interactions entre NCBD et TAD par MS native**

### **3.1 Introduction**

Parallèlement à l'étude d'oligomérisation de NCBD, nous avons utilisé la MS native pour mieux caractériser les interactions entre NCBD et son partenaire TAD (Transactivation Domain) qui est le domaine N-terminal de transactivation de la protéine P53. La protéine P53 est un facteur de transcription régulant de nombreuses fonctions cellulaires importantes telles que la régulation du cycle cellulaire, l'autophagie ou l'apoptose.<sup>[36]</sup> C'est un suppresseur de tumeurs essentiel, reconnu depuis longtemps pour réprimer le cancer grâce à l'induction de programmes ou d'apoptose en réponse à une pléthore de signaux de stress cellulaire différents.<sup>[37]</sup> Près de 60% des cancers humains sont liés à une mutation de P53 ou un dysfonctionnement de sa régulation.

Cette protéine se compose de cinq domaines tels que le domaine de transactivation (TAD), le domaine riche en proline, le domaine de liaison à l'ADN, le domaine de tétramérisation et le domaine régulateur.



**Figure 52 : Organisation du domaine de p53 et CBP/p300. (A) Domaines de p53. TAD (domaine de transactivation N-terminal), riche en P (riche en proline), DBD (domaine de liaison à l'ADN), TD (domaine de tétramérisation) et REG (domaine de réglementation C-terminal). L'emplacement des motifs AD1 et AD2 est indiqué sur une séquence d'acides aminés partielle ; les sites connus de phosphorylation sont indiqués par des points. (B) Domaines de CBP / p300. Les domaines qui interagissent avec le domaine de transactivation p53, TAZ1 (résidus 340 - 439), KIX (586 - 672), TAZ2 (1764 - 1855), NCBD (2059 - 2117) sont affichés en noir. Figure extrait de [38]**

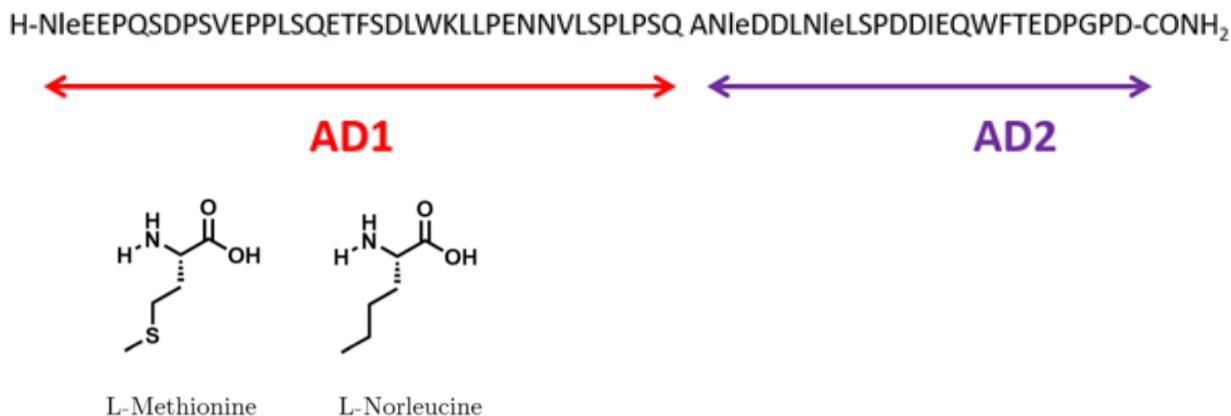
Le domaine TAD est un domaine N-terminal acide qui est intrinsèquement désordonné et responsable de la liaison à divers facteurs de transcription. Des études ont par ailleurs montré que l'activité de p53 est contrôlée par son interaction avec HDM2 et les coactivateurs transcriptionnels tels que la protéine régulatrice CREB-Binding Protein (CBP).<sup>[20, 38]</sup>

Le domaine TAD de la protéine p53 interagit via ses deux régions actives : le domaine d'activation 1 (AD1) de la zone d'acide aminé 1 à 42 et le domaine d'activation 2 (AD2) de la zone d'acide aminé 43 à 63.<sup>[39]</sup> Dans les cellules non stressées, AD2 se lie aux domaines CBP tandis que AD1 se lie à HDM2. TAD forme ainsi un complexe ternaire via AD1 et AD2 conduisant à la dégradation de p53 par ubiquitination. Le stress génotoxique va ensuite entraîner la phosphorylation des acides aminés thréonine et serine de AD1 affaiblissant ainsi son interaction avec HDM2. Cet affaiblissement conduit alors à une liaison préférentielle de TAD au CBP et à l'activation de p53.<sup>[40]</sup>

La structure de TAD en interaction avec NCBD a été résolue par RMN montre un noyau structural formé de 5 hélices  $\alpha$  (2 hélices sur TAD et 3 hélices sur NCBD)<sup>[20]</sup>. Outre ce cœur, le reste du complexe est désordonné et le mécanisme de reconnaissance moléculaire lors de la formation du complexe n'est pas encore compris. Au Laboratoire de Chimie des Protéines, des expériences de dichroïsme circulaire ont montré que le p53-TAD synthétisé chimiquement n'est pas structuré et l'amélioration du caractère hélicoïdal lors de l'ajout du NCBD confirme la formation du complexe. Selon l'analyse par titrage calorimétrique isotherme, la protéine p53-TAD possède deux sites de liaison pour la protéine NCBD, ce qui pourrait signifier que deux molécules de NCBD se lient à une seule molécule de TAD. Mais aucune information directe n'est disponible pour confirmer la stœchiométrie ni l'homogénéité du complexe formé. Afin de mieux comprendre l'interaction entre TAD et NCBD en particulier sous l'effet de la phosphorylation de TAD, nous avons choisi d'utiliser la MS native. Ainsi, les stœchiométries d'interaction entre NCBD et les deux domaines d'activation de TAD (AD1 et AD2) ont été déterminées, puis l'influence de la phosphorylation sur le complexe NCBD/TAD a été évaluée. Ces dernières expériences sont encore en cours.

Nos collaborateurs ont ainsi dans un premier temps effectué la synthèse chimique totale (de TAD) basée sur la ligature chimique native de trois segments peptidiques. Puis, dans un second temps, des expériences pour caractériser sa liaison au domaine de liaison du coactivateur nucléaire (NCBD) de la protéine de liaison au CREB ont été effectuées.<sup>[4]</sup>

Séquence de TAD synthétique :

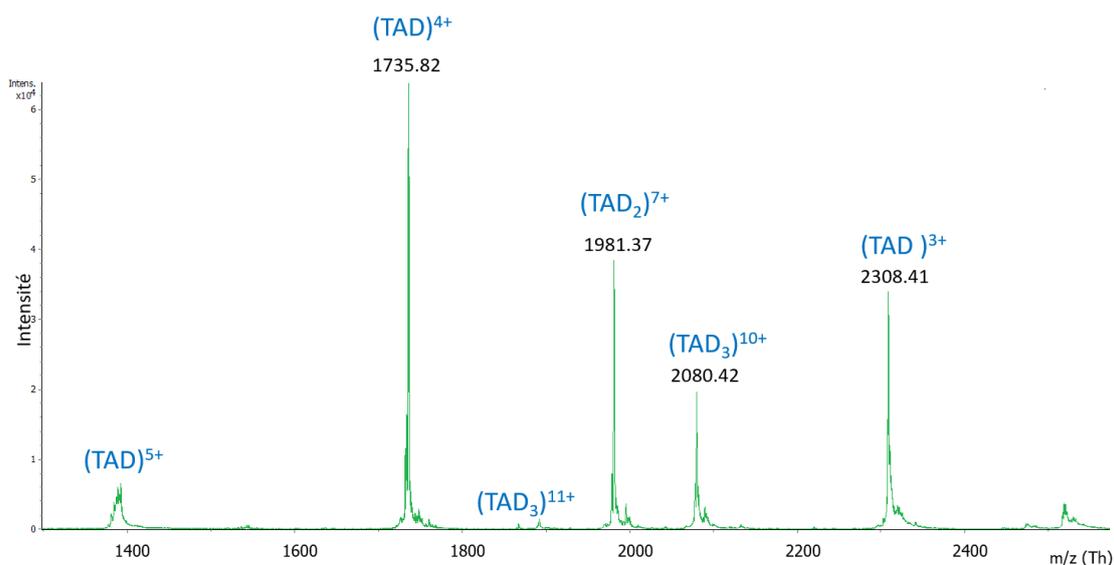


**Figure 53 :** Les méthionines ont été remplacées par des résidus Norleucine, quasi isostérique pour empêcher l'oxydation lors de la synthèse et des analyses ultérieures.

### **3.2 Etude des interactions entre NCBD et TAD par MS native**

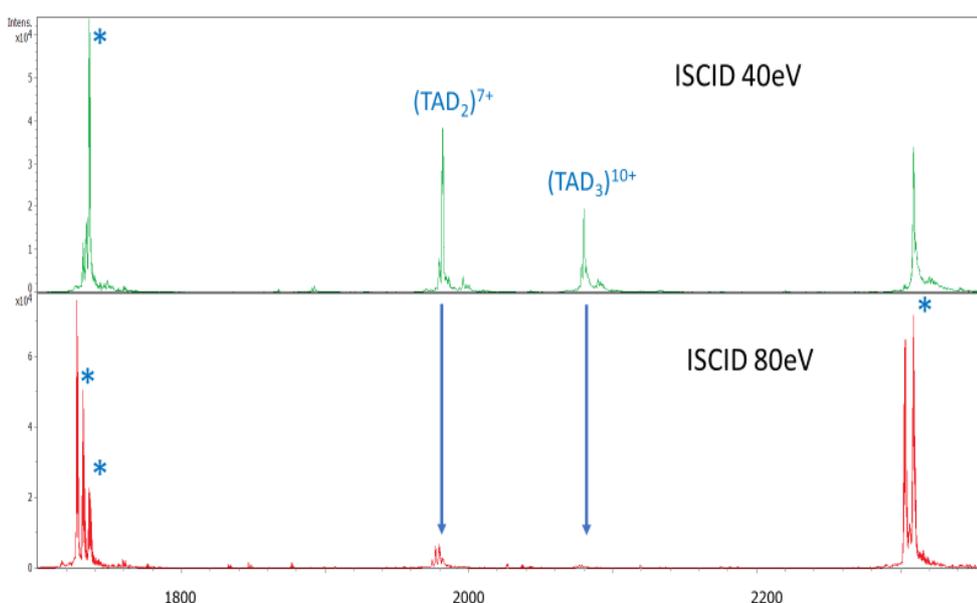
#### **3.2.1 Caractérisation des composés seuls**

Tout comme précédemment avec NCBD, nous avons optimisé les paramètres instrumentaux et les conditions d'analyse sur le composé TAD. Celui-ci a été solubilisé dans une solution d'acétate d'ammonium concentré à 100mM à pH 6.8. Puis, l'échantillon a été incubé pendant 5min à 37°C avant d'être analysé par MS native à une concentration de 150µM.



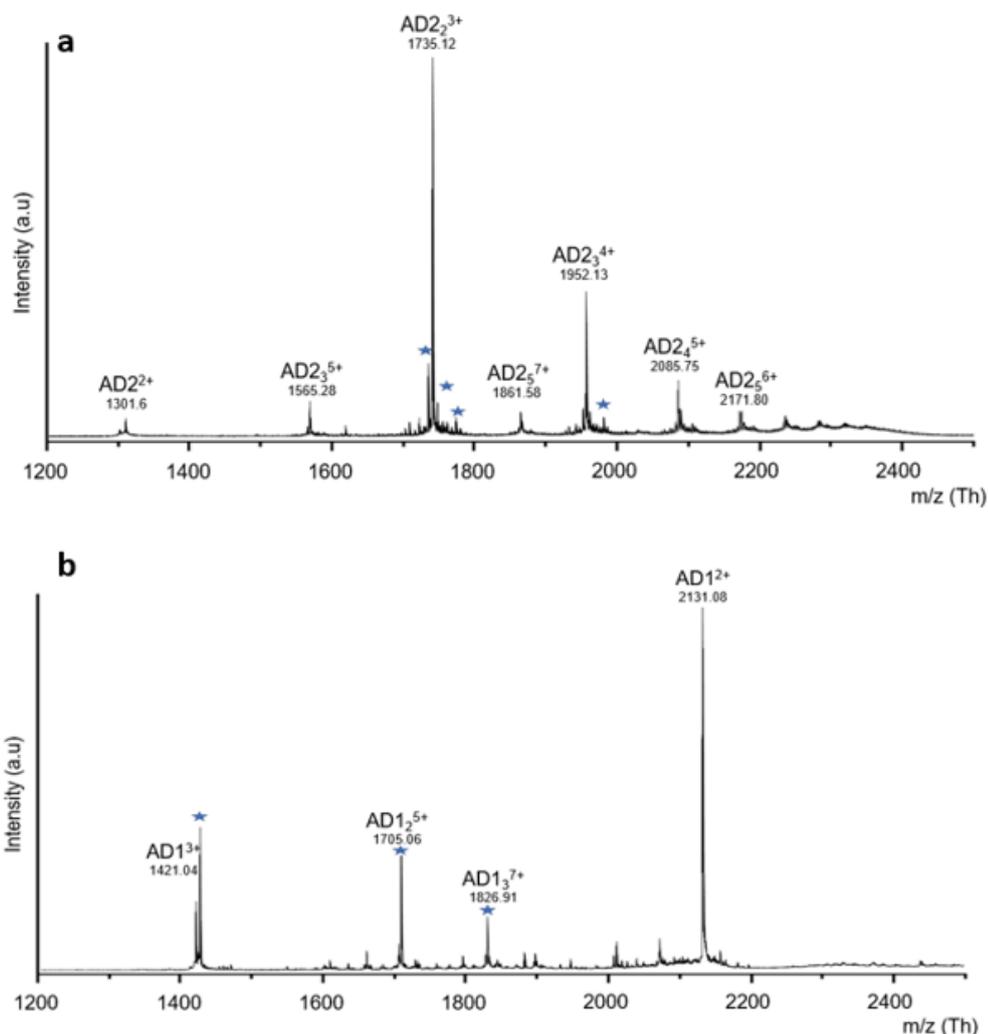
**Figure 54 : Spectre MS native de TAD. Détection majoritaire de sa forme monomérique à différents états de charge.**

Après analyse par MS native, le domaine TAD a été observé avec une masse de  $6905.2 \pm 0,2$  Da sous plusieurs états de charges et majoritairement sous forme monomérique. Des ions dimériques et trimériques sont également détectés montrant une certaine propriété de TAD à s'agréger. Comme cela a été exposé avec NCBD, les conditions d'interface du spectromètre de masse ont été ajustées pour qu'aucune dissociation en phase gazeuse ne se produise. Ceci est atteint pour une valeur ISCID de 40eV. Dans ces conditions, toutes les espèces sont détectées sous forme d'adduits d'ammonium (2 adduits de  $\text{NH}_4^+$  sur le monomère  $m/z$  1735,82 chargés  $4^+$  ; 3 adduits de  $\text{NH}_4^+$  sur le dimère  $m/z$  1981,93 chargés  $7^+$  ; 4 adduits de  $\text{NH}_4^+$  sur le trimère  $m/z$  2080,41 chargés  $10^+$ ). Une augmentation de l'énergie interne des ions dans l'interface (ISCID = 80eV) provoque la diminution de ces adduits mais également la dissociation des dimères et trimères (figure 55).



**Figure 55 : Spectre MS native de TAD obtenu en fonction de l'énergie de collision in source (ISCID) à 40eV et à 80eV. Les étoiles représentent la présence d'adduit de  $\text{NH}_4^+$ .**

Les deux domaines d'activation (AD1 et AD2) ont également été caractérisés dans les mêmes conditions.



**Figure 56 : Spectre de masse native des domaines activations de TAD (a) de AD2 à 150 μM et (b) de AD1 à 150 μM. Les étoiles représentent les adduits de NH<sub>4</sub><sup>+</sup>.**

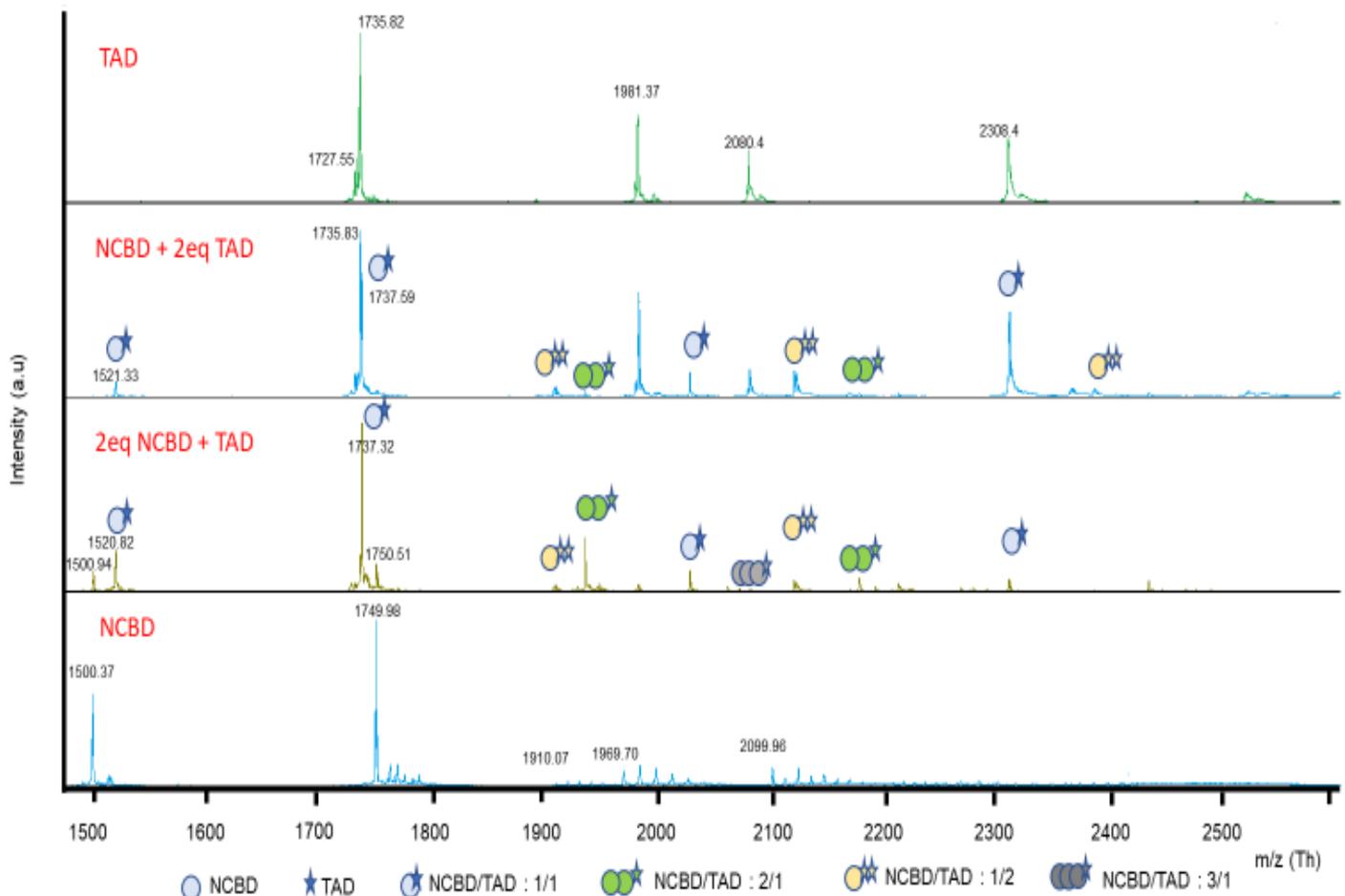
Les spectres de masse révèlent un comportement différent de AD1 et AD2 face à leurs propriétés d'oligomérisation. Alors que AD1 est majoritairement détecté sous forme monomérique ( $M = 4260.1 \pm 0,2$  Da), AD2 apparaît sous forme de dimère avec un ion 3-fois chargé caractéristique à m/z 1735.12 (figure 56). Des formes secondaires minoritaires ont été observées : dimère (m/z 1705.06 chargés 5+), et trimère (m/z 1826.91 chargés 7+) pour AD1 et trimère (1952.13 chargés 4+, 1565 chargés 5+) tétramère (2085.75 chargés 5+), mais également sous forme pentamériques (2171.80 chargés 6+) pour AD2. Le domaine AD2 montre ainsi une tendance à s'agréger très rapidement en solution.

Ainsi aussi bien NCBD que TAD (via son domaine AD2 principalement) ont la capacité de s'agréger en solution pour former rapidement des oligomères stables dans le temps. Mais qu'en est-il de la stœchiométrie d'interaction entre ces 2 partenaires ? Pour cela, nous avons

analysé par MS native la complexation entre TAD et NCBD, ainsi qu'avec les sous-domaines AD1 et AD2 afin de voir comment évoluent les oligomères lors de l'interaction et vérifier si la complexation entre TAD et NCBD s'effectue sur l'ensemble des deux domaines ou seulement sur un sous-domaine comme il a été proposé dans des études antérieures.<sup>[4]</sup>

### 3.2.2 Caractérisation du complexe NCBD-TAD

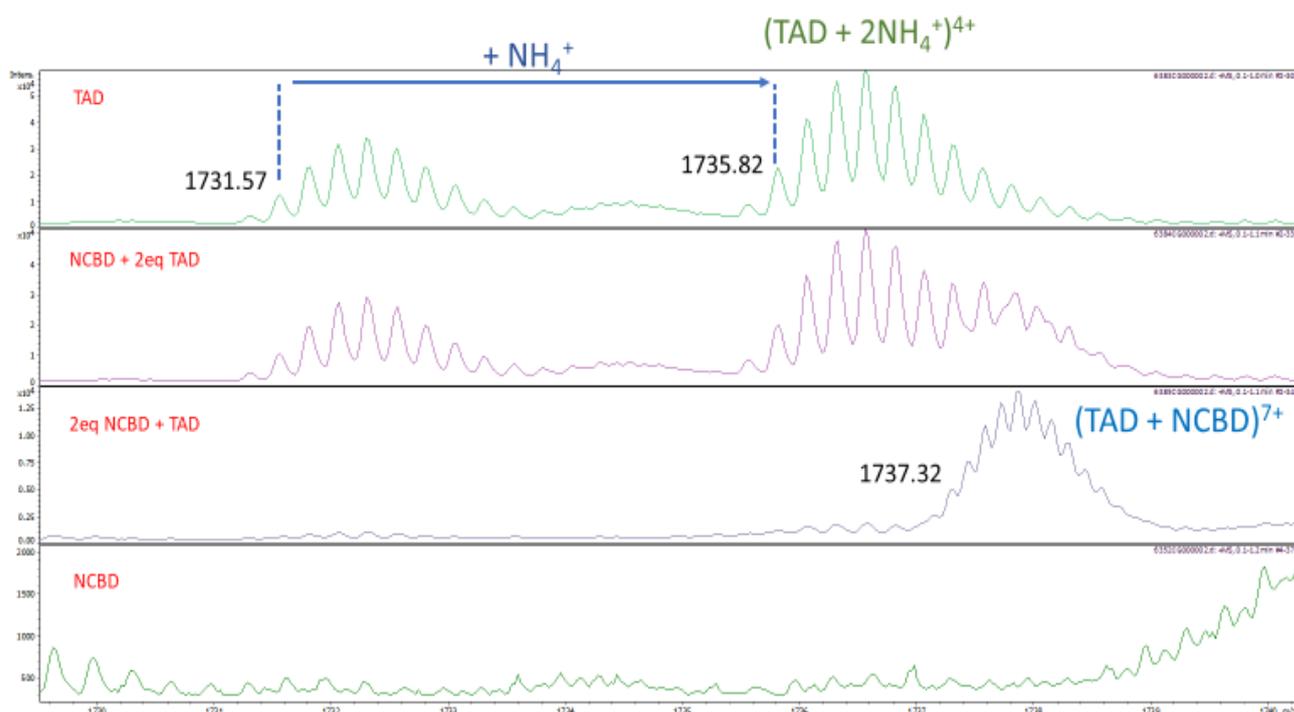
TAD et NCBD sont inversement chargés : TAD possède 15 résidus acides et NCBD 6 résidus basiques. Pour étudier l'interaction entre TAD et NCBD, nous avons mélangé différents ratios de L-NCBD et de TAD et suivi par MS native l'évolution des espèces grâce à leurs ions multichargés caractéristiques. En particulier, nous avons effectué l'analyse du complexe NCBD + TAD selon un ratio (2/1) et selon un ratio (1/2) où chacun des partenaires était mis en excès à une concentration en protéine de 150  $\mu$ M. Vu la concentration élevée d'analyse (nécessaire à la mise en évidence d'oligomères) et en tenant compte de la solubilité de NCBD, il n'a pas été possible de tester des ratios de mélange significativement plus importants.



**Figure 57 : Spectre de masse native de l'interaction entre le domaine TAD de p53 et le domaine NCBD de CBP selon un ratio NCBD/TAD (2/1) et selon un ratio NCBD/TAD (1/2). Dans un souci de visibilité sur le spectre, seules les espèces complexées ne sont ici annotées, les pics vierges de symboles correspondent à TAD ou NCBD libre.**

La figure 57 montre les spectres de masse natifs obtenus pour ses deux échantillons dans des conditions d'analyse strictement identiques à une énergie d'interface de 40eV. Dans les deux spectres, des complexes correspondant à l'interaction entre TAD et NCBD sont détectés et différentes stœchiométries d'interaction sont déterminées. Comme cela avait été montré pour l'étude de NCBD, la détermination de l'état de charge des ions par l'interprétation des profils isotopiques s'est avérée cruciale pour identifier la nature des complexes formés.

En particulier, la discrimination entre l'ion 4-fois chargé à  $m/z$  1735.8 (TAD + 2 adduits de  $\text{NH}_4^+$ ) et l'ion 7-fois chargé à  $m/z$  1737.3 (NCBD/TAD : 1/1) requiert une résolution suffisante pour les identifier sans ambiguïté (cf. figure 58).

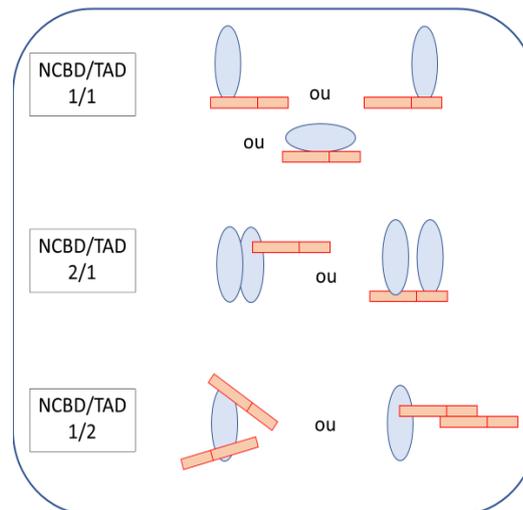


**Figure 58 : zoom de la région  $m/z$  illustrant la résolution nécessaire pour discriminer les espèces libres  $\text{TAD}^{4+}$  et complexées  $(\text{NCBD}/\text{TAD} : 1/1)^{7+}$ .**

L'espèce majoritaire est un complexe NCBD/TAD : 1/1, et ce, quelle que soit le partenaire placé en excès en solution ce qui indique que cette interaction est spécifique. Des ions correspondants à des complexes NCBD/TAD : 1/2 et 2/1 sont également observés dans les deux spectres de masse, avec des abondances relatives qui suivent le ratio introduit en solution. En revanche, aucun complexe impliquant des oligomères de NCBD supérieurs au dimère n'est observé ce qui montre que l'équilibre monomère/multimère est déplacé par la présence de TAD et que les agrégats de NCBD sont déstabilisés lors de l'interaction avec TAD. Seul un ion peu intense à  $m/z$  2061.5 correspondant à un complexe NCBD/TAD : 3/1 est observé en présence d'excès en NCBD. De façon intéressante, les clusters de NCBD libres ne sont plus observés non plus.

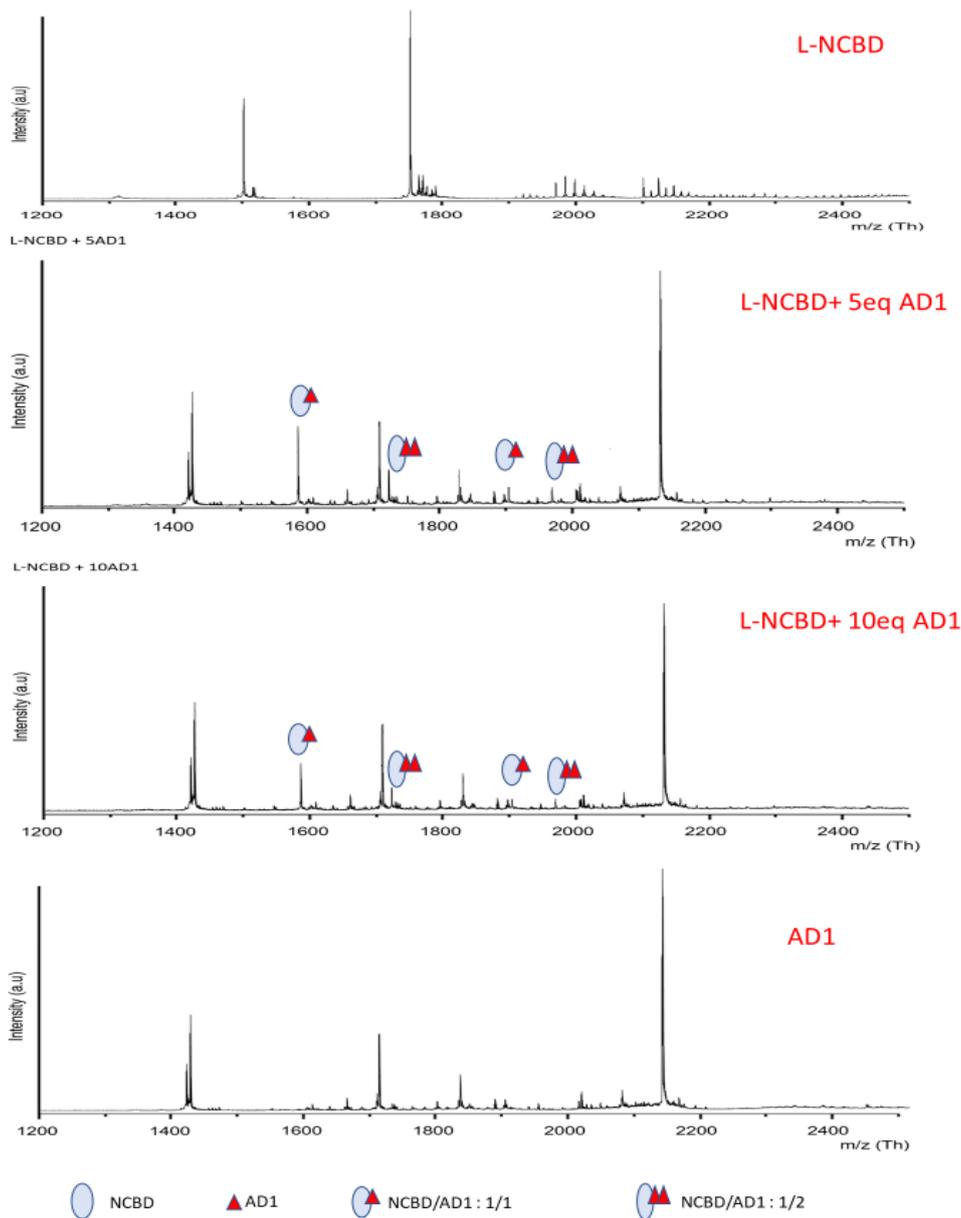
Ces résultats MS permettent ainsi de montrer qu'une interaction entre le domaine NCBD de la protéine régulatrice CREB-Binding Protein (CBP) et le domaine TAD de p53 était réalisée et formait majoritairement un complexe de stœchiométrie 1/1.

Toutefois, à partir de la seule mesure de la masse moléculaire, il n'est pas possible de préciser si les deux domaines de TAD (AD1 et AD2) entrent en interaction avec NCBD et plusieurs cas de figures peuvent être imaginés (figure 59).



**Figure 59 : Schéma montrant que plusieurs hypothèses peuvent expliquer les stœchiométries des complexes formés entre TAD et NCBD telles qu'observées sur le spectre de masse**

Pour éliminer certaines hypothèses, nous avons analysé de la même manière les interactions entre NCBD et les domaines AD1 et AD2 isolés pour déterminer si chacun des sous-domaines est capable d'interagir avec NCBD. Pour cela, nous avons travaillé avec un excès molaire de 5 équivalences et de 10 équivalences en AD2 ou en AD1 à une concentration protéique de 30 $\mu$ M. Une étude de la stabilité en phase gazeuse de ces édifices a été menée et montre que ceux-ci sont particulièrement résistants à la collision dans l'interface. En effet, les abondances relatives ne varient pas entre 0 et 80eV et de la fragmentation covalente se produit pour une énergie de collision supérieure. Ceci est en accord avec des interactions intermoléculaires de type ponts salin qui se trouvent renforcées dans le spectromètre de masse après désolvatation mais ne nous donne aucune information sur une éventuelle topologie préférentielle.



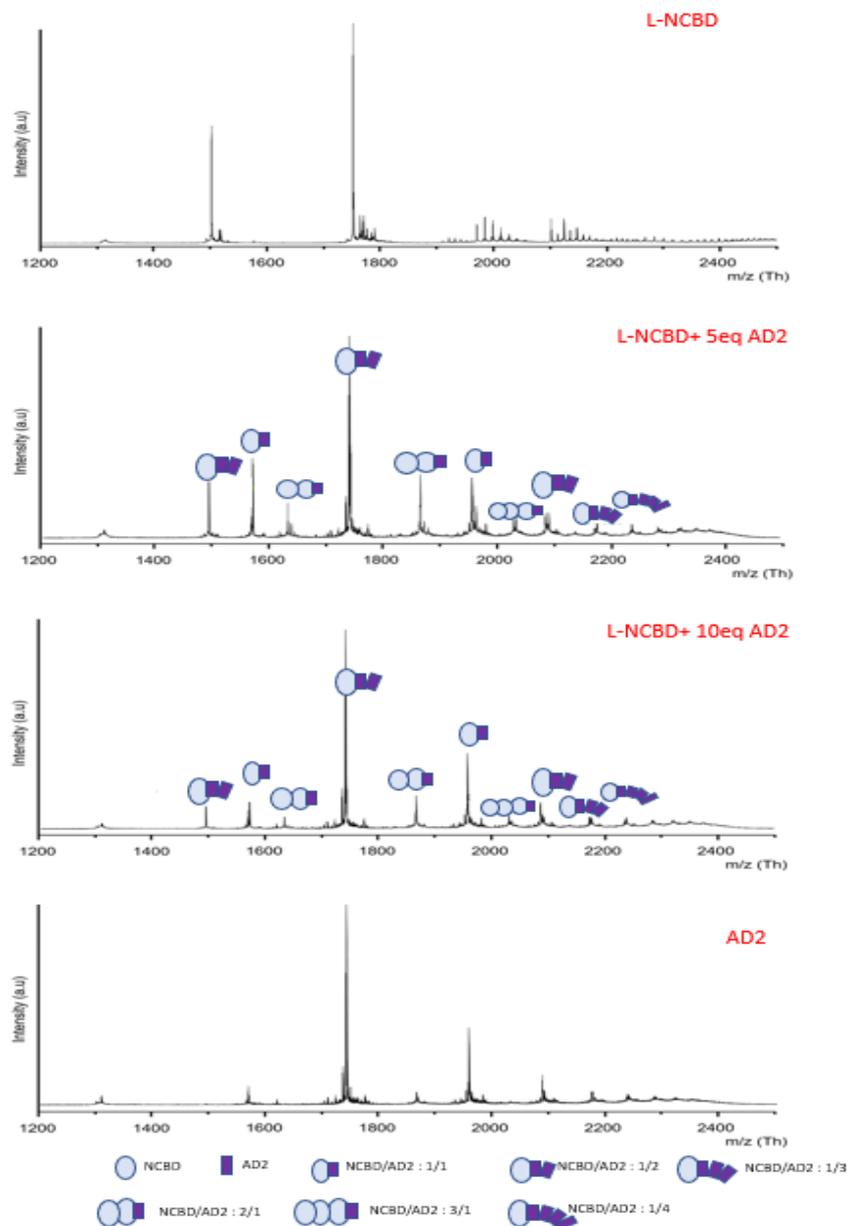
**Figure 60 : Spectre de masse native montrant l'interaction du domaine d'activation AD1 de p53 avec le domaine NCBD de la protéine CBP (a) à un ratio 1/5 (NCBD/AD1) et (b) à un ratio 1/10 (NCBD/AD1).**

Les spectres obtenus (figures 60 et 61) montrent que les deux domaines AD1 et AD2 présentent une affinité pour NCBD puisque des interactions ont été observées dans les deux échantillons. Aucune différence significative liée à un effet de concentration de AD1 ou de AD2 n'est observée dans la nature des complexes résultants. En ce qui concerne la complexation avec AD1, le spectre est plutôt simple et des ions correspondant à des complexes de type NCBD/AD1 : 1/1 et 1/2 sont détectés. En ce qui concerne l'interaction avec AD2, de nombreuses stœchiométries co-existent en solution comme cela est indiqué sur la figure 61. Ces complexes sont formés par l'interaction d'une protéine de NCBD avec plusieurs domaines d'activation AD2 (NCBD/nAD2 avec n=1 à 4) mais des ions révélant une interaction entre AD2 et des clusters de NCBD sont également observés (NCBD/AD2 : 2/1 et 3/1). Tout comme nous l'avions noté sur TAD, aucun signal correspondant aux oligomères de NCBD agrégés (trimères

et tétramères) n'est présent sur les spectres de masse ce qui semble indiquer que l'ajout d'AD1 comme d'AD2 induit leur dissociation en solution en faveur de la formation de complexes hétéromériques NCBD/AD2 et NCBD/AD1.

Alors que les conditions expérimentales sont strictement identiques, le comportement différent de NCBD lorsqu'il est mis en contact avec AD1 ou AD2 montre que la complexation n'est pas un phénomène d'association statistique mais qu'il est guidé par la séquence.

Il n'est pas possible de savoir à partir de la mesure de la masse moléculaire où sont positionnées toutes les molécules d'AD2.



**Figure 61 : Spectre de masse native montrant l'interaction du domaine d'activation 2 de p53 avec le domaine NCBD de la protéine CBP (a) à un ratio 1/5 (NCBD/AD2) et (b) à un ratio 1/10 (NCBD/AD2).**

En conclusion :

- NCBD interagit avec TAD pour former un complexe de stœchiométrie 1/1, qui co-existe avec des édifices NCBD/TAD : 1/2 et 2/1 dont les abondances relatives suivent la composition en protéine de la solution
- AD1 et AD2 forment des oligomères en solution avec une agrégation plus forte pour AD2 (5-mères) que pour AD1 (trimères)
- Les deux sous-domaines sont capables d'interagir avec NCBD

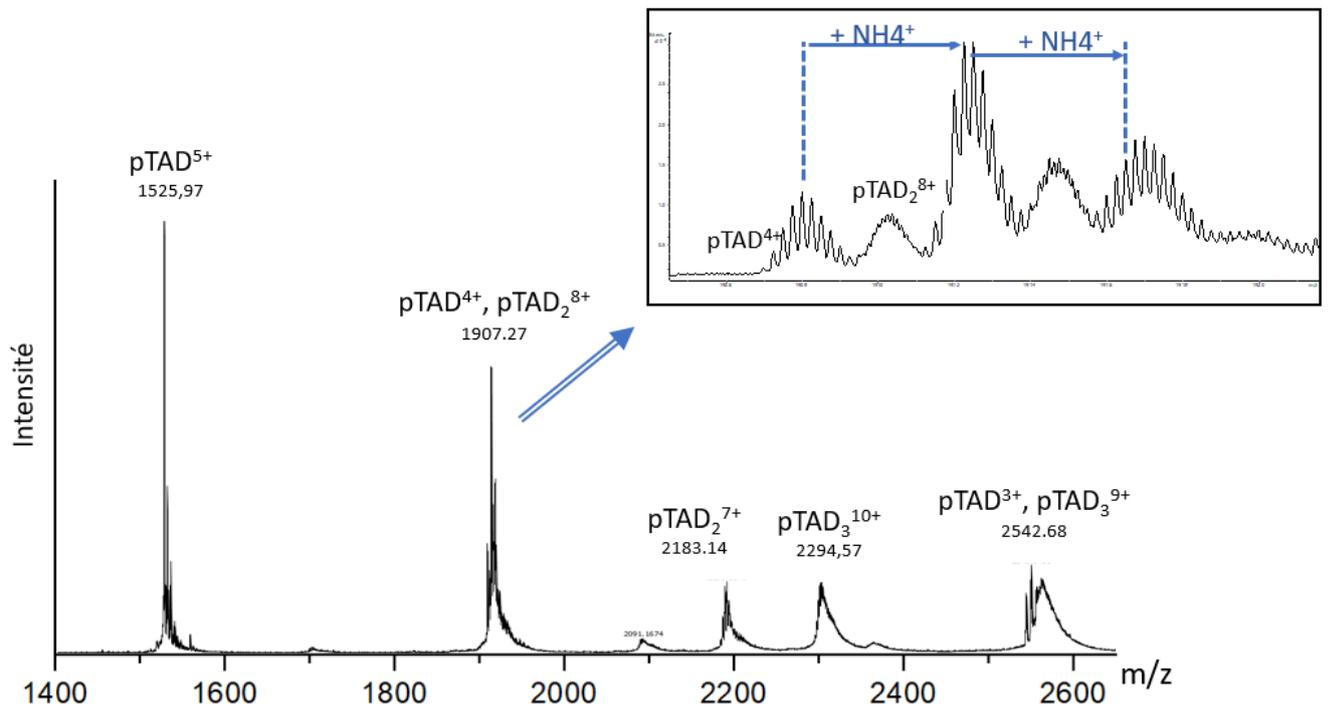
**3.3 Caractérisation de l'interaction entre NCBD et la protéine TAD phosphorylée.**

Ayant mis au point des conditions analytiques nous permettant de mettre en évidence des interactions entre TAD et NCBD, nous avons cherché à savoir quel serait l'impact de la phosphorylation de TAD sur la stabilité du complexe NCBD/TAD. En effet, il a été montré que la phosphorylation de TAD jouait un rôle dans l'activité de p53<sup>[41]</sup>.

Très peu de protéines en oncologie moléculaire ont autant de sites de modifications post-traductionnelles (PTM) que p53. Très tôt, des études ont montré que cette phosphorylation s'effectuait sur les acides aminés sérine et thréonine et qu'une seule phosphorylation permettait de doubler l'affinité d'interaction de p53 envers le domaine TAZ1 de CBP. Cette affinité entre p53 et TAZ1 augmente avec le degré de phosphorylation jusqu'à atteindre 10 fois sa valeur <sup>[42, 43]</sup>. Il a ainsi été observé que la phosphorylation avait un pouvoir de modulation important sur les interactions de p53 avec chacun des domaines de liaison de la protéine CBP avec des degrés différents selon le nombre et le type d'acide aminé phosphorylé<sup>[42]</sup>.

Parmi les 17 sites de phosphorylations de p53, 9 se situent sur le domaine TAD (7 sérines et 2 thréonines). Plusieurs travaux dans la littérature ont montré qu'une phosphorylation sur la Ser15, Ser20, Ser33, Ser37, Ser46, Thr18 et Thr55 individuellement régulait l'interaction entre p53 et quatre domaines de CBP<sup>[42]</sup>.

Afin d'examiner l'impact de la phosphorylation sur l'association de TAD et de NCBD, nos collaborateurs ont choisi d'étudier un scénario extrême dans lequel TAD est phosphorylé sur ces 9 sites potentiels (pTAD). L'objectif était ainsi de comparer l'efficacité d'interaction de TAD et de pTAD vis-à-vis de leur complexation à NCBD.



**Figure 62 : Spectre MS native du domaine TAD phosphorylé (p-TAD). Détection majoritaire de sa forme monomérique à différents états de charge. Dans l'encadré se trouve un agrandissement de la zone 1905-1925 m/z qui permet de visualiser les adduits d'ammonium (+17Da) ainsi que la co-détection du monomère chargé 4+ et du dimère chargé 8+.**

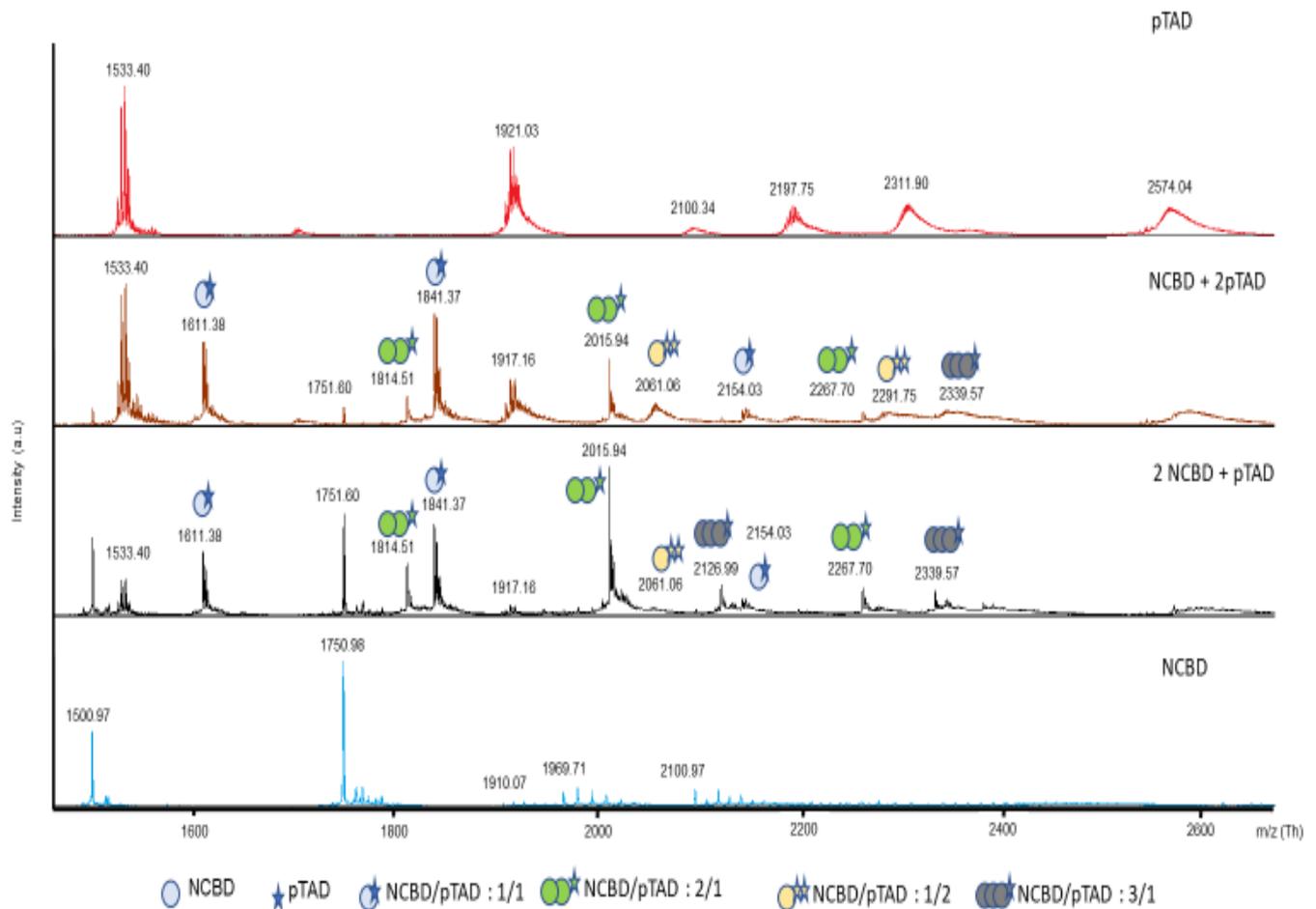
La figure 62 représente le spectre de masse natif de pTAD à 150µM. Il révèle que pTAD existe majoritairement sous forme de monomère (masse mesurée de  $7624.92 \pm 0.2$  Da en parfait accord avec la masse attendue d'après la séquence). De nombreux adduits d'ammonium (+17 Da) sont détectés, ce qui peut s'expliquer aux vues du caractère très acide de ce composé et de la présence de 9 groupements phosphates. Néanmoins, nous étions surpris que ceux-ci ne soient pas éliminés sous forme d'ammoniac lors du passage des ions dans l'interface comme c'est le cas pour les oligonucléotides<sup>[44]</sup>. Des signaux correspondant à un dimère (ion 7+ à m/z 2183.14) et à un trimère (ion 10+ à m/z 2294.57) sont détectés également sous forme d'adduits d'ammonium. Tout comme cela avait été le cas pour NCBD, la présence de ces adduits nous aide pour l'assignation de la charge des ions et donc l'attribution de leur masse.

Lorsque NCBD est rajouté en solution, des ions reflétant des complexes de plusieurs stœchiométries sont observés, dont l'abondance relative varie avec la concentration protéique (figure 63).

En présence de 2eq de pTAD, l'espèce majoritaire correspond à un complexe NCBD/pTAD : 1/1 dont les ions 7-fois et 8-fois chargés (m/z=1841.37 et 1611.38 respectivement) sont caractéristiques. Des complexes de type NCBD/pTAD : 2/1 et 1/2 sont également interprétés. Il est intéressant de noter que l'abondance relative de ces deux complexes (2/1 et 1/2) ne suit pas la concentration des espèces en solution puisque le complexe 1/2 est minoritaire en présence d'un excès de pTAD.

En présence de 2eq de NCBD, des complexes de stœchiométries NCBD/pTAD : 1/1, 1/2, 2/1 et 3/1 sont détectées. L'espèce majoritaire correspond ici au complexe de stœchiométrie

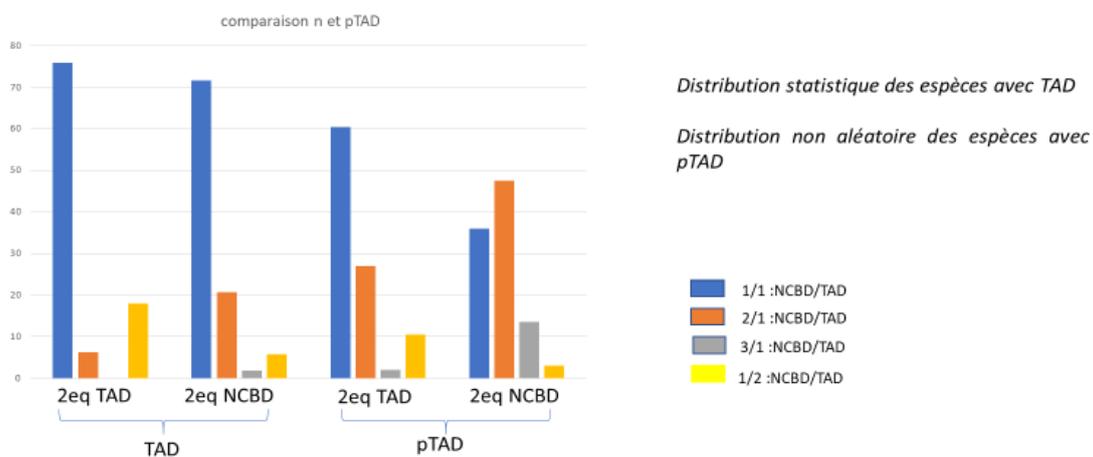
NCBD/pTAD : 2/1. Tout comme nous l'avons remarqué avec TAD non phosphorylé, les phénomènes d'oligomérisation de NCBD (mais aussi de TAD) semblent déstabilisés lorsque les deux partenaires sont mis en contact.



**Figure 63 : Spectre de masse native de l'interaction entre le domaine pTAD de p53 et le domaine NCBD de CBP selon un ratio NCBD/pTAD (1/2) (a) et selon un ratio NCBD/pTAD (2/1).**

Afin de mieux visualiser la répartition des espèces complexées et pouvoir comparer le comportement de NCBD en présence de TAD versus pTAD, nous avons estimé à partir des spectres ci-dessus les proportions relatives de chaque stœchiométrie en tenant compte en 1<sup>ière</sup> approximation de leur intensité. Bien sûr, en fonction de sa stœchiométrie et de sa conformation, chaque complexe présentera un facteur de réponse qui lui est propre. L'idée n'est pas ici de déterminer les abondances relatives des édifices au sein d'un même échantillon mais plutôt de comparer les distributions selon que TAD est phosphorylé ou non. Sur la figure 64, on observe que l'hétérodimère NCBD/TAD :1/1 est l'espèce prédominante, sauf dans l'échantillon pTAD + 2eq de NCBD. Ainsi, bien que possédant deux sous-domaines capables de lier NCBD s'ils sont étudiés séparément (AD1 et AD2), la complexation n'est pas quantitative. Plusieurs hypothèses peuvent être proposées : (i) l'un des sous-domaines AD1 ou AD2 possède une forte affinité pour NCBD mais une fois la 1<sup>ière</sup> molécule de NCBD fixée, le second sous-domaine n'est plus accessible ou bien un changement de conformation

diminue son affinité ; (ii) la molécule de NCBD interagit de façon concomitante avec les deux sous-domaines. Une idée serait de placer AD1 et AD2 en compétition vis-à-vis de NCBD et de déterminer si un complexe ternaire s'établit. La distribution des espèces change lorsque TAD est phosphorylé, notamment le complexe NCBD/pTAD : 2/1 est fortement stabilisé. Soit la phosphorylation de TAD induit un changement de conformation qui rend les deux sous-domaines accessibles à la complexation, soit la phosphorylation induit une augmentation de l'affinité pour un cluster de NCBD. Ce renforcement de l'interaction entre les deux partenaires confirme ainsi les résultats décrits dans la littérature où il a été démontré qu'une seule phosphorylation a permis de doubler l'affinité d'interaction de p53 envers le domaine TAZ1 de CBP<sup>[42, 45, 46]</sup>.



**Figure 64 : Distribution statistique des espèces.**

Des études sont actuellement en cours sur les sous-domaines individuels phosphorylés (pAD1 et pAD2) afin de vérifier si les deux sous-domaines présentent une augmentation d'affinité pour NCBD.

## 4. Conclusions et Perspectives

Cette étude a montré que la MS native est un outil clé pour aider la compréhension du processus d'association entre TAD et NCBD, malgré le fait que les deux protéines soient partiellement désordonnées. Nous avons observé que NCBD était non seulement propice à l'agrégation, mais également que son hélice N-terminale jouait un rôle important dans la médiation de l'auto-assemblage intermoléculaire du domaine. De plus, là où la spectroscopie IR et la microscopie électronique ont montré l'évolution des édifices vers une structure sous forme feuillet  $\beta$  et donc d'obtenir une structure plus stable de morphologie fibreuse en particulier à pH acide, la MS a d'autre part montré qu'à pH neutre, l'agrégation de NCBD pourrait s'effectuer rapidement, à faible concentration et qu'une interaction entre des protéines hétérochirales L- et D-NCBD était réalisable.

Dans un second temps, la MS native a permis de mieux comprendre le mécanisme d'interaction entre le domaine TAD non phosphorylé de la protéine p53 et NCBD. Il a ainsi été observé que la complexation se faisait selon une stœchiométrie préférentielle 1:1 (NCBD :TAD) et plus minoritairement 2 :1 (NCBD :TAD). De plus, les clusters de NCBD n'ont plus été observés en présence de TAD ce qui suggère que l'interaction avec TAD déstabilise l'équilibre monomère/multimère.

Enfin, nous avons pu confirmer que les interactions entre NCBD et TAD étaient renforcées lorsque TAD était phosphorylé et que pTAD pouvait interagir avec deux molécules de NCBD. Afin de mieux caractériser le rôle de chacune des phosphorylations de pTAD, des analogues du TAD dont le squelette a été modifié par des résidus 1,2,3 triazoles 1,4 substitués ont été synthétisés par nos collaborateurs du Laboratoire de Chimie des Protéines. Ces triazoles remplacent la liaison amide sans grande perturbation en termes de propriétés géométriques, stériques et électroniques<sup>[47]</sup>. Ces analogues de TAD présentent 2 modifications sur le sous-domaine AD1 et 2 modifications sur AD2. L'étude de leur comportement par MS native pourrait permettre de déterminer quelle position de phosphorylation peut perturber la formation du complexe.

La MS native est donc un outil adéquat pour étudier les phénomènes d'agrégation des protéines. Outre l'optimisation des conditions douces d'interface, la résolution instrumentale s'est avérée un paramètre capital pour l'interprétation des données. En effet, l'identification de l'état de charge est nécessaire à la détermination de la masse des ions. Notre microTOF-QII possédant une résolution moyenne (10000), il est très difficile d'identifier les ions avec un état de charge supérieur au 7+. C'est pourquoi il serait intéressant d'effectuer ses analyses sur un spectromètre de masse avec une plus haute résolution afin de confirmer l'identité de certains ions.

Le projet a démarré il y a un an et est toujours en cours. Il serait ainsi intéressant de poursuivre l'étude sur une échelle de temps plus longue (sur plusieurs semaines) et à des pH acides afin de voir s'il y aurait possibilité de confirmer les résultats observés en microscopie TEM ou par spectroscopie IR concernant l'évolution de l'agrégation.

## 5. Bibliographie

1. Rambaran, R.N. and L.C. Serpell, *Amyloid fibrils: abnormal protein assembly*. Prion, 2008. **2**(3): p. 112-117.
2. Xu, S., et al., *Characterization of tau fibrillization in vitro*. Alzheimer's & dementia : the journal of the Alzheimer's Association, 2010. **6**(2): p. 110-117.
3. Chiti, F. and C.M. Dobson, *Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade*. Annual Review of Biochemistry, 2017. **86**(1): p. 27-68.
4. Baral, A., et al., *Chemical synthesis of transactivation domain (TAD) of tumor suppressor protein p53 by native chemical ligation of three peptide segments*. Tetrahedron, 2019. **75**(6): p. 703-708.
5. Lu, Y., et al., *Native Mass Spectrometry Analysis of Oligomerization States of Fluorescence Recovery Protein and Orange Carotenoid Protein: Two Proteins Involved in the Cyanobacterial Photoprotection Cycle*. Biochemistry, 2017. **56**(1): p. 160-166.
6. Heuvel, R.H.H.v.d. and A.J.R. Heck, *Native protein mass spectrometry: from intact oligomers to functional machineries*. Current Opinion in Chemical Biology, 2004. **8**(5): p. 519-526.
7. Esposito, G., et al., *Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation*. Protein Sciences, 2000. **9**(5): p.831-845.
8. De Simone, A., et al., *Investigating in Vitro Amyloid Peptide 1-42 Aggregation: Impact of Higher Molecular Weight Stable Adducts*. ACS omega, 2019. **4**(7): p. 12308-12318.
9. Wang, W., E.N. Kitova, and J.S. Klassen, *Influence of Solution and Gas Phase Processes on Protein-Carbohydrate Binding Affinities Determined by Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. Analytical Chemistry, 2003. **75**(19): p. 4945-4955.
10. Breuker, K. and F.W. McLafferty, *Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(47): p. 18145-18152.
11. Chan, H.M. and N.B. La Thangue, *p300/CBP proteins: HATs for transcriptional bridges and scaffolds*. Journal of Cell Sciences, 2001. **114**: p.2363-2373.
12. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development*. Genes & development, 2000. **14**(13): p.1553-1577.
13. Giordano, A. and M.L. Avantaggiati, *p300 and CBP: partners for life and death*. Journal of Cellular Physiology, 1999. **181**:p. 218-230.
14. Kamei, Y., et al., *A CBP Integrator Complex Mediates Transcriptional Activation and AP-1 Inhibition by Nuclear Receptors*. Cell, 1996. **85**(3): p. 403-414.
15. Westin, S., et al., *Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators*. Nature, 1998. **395**(6698): p. 199-202.
16. Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator*. Nature, 2008. **451**: p.846-850.
17. Dancy, B.M., et al., *Live-cell studies of p300/CBP histone acetyltransferase activity and inhibition*. ChemBioChem, 2012. **13**(14): p2113-2121.
18. Park, S., et al., *Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation*. Proceedings of the National Academy of Sciences, 2017. **114**(27): p. E5335.
19. Karlsson, E., et al., *Coupled Binding and Helix Formation Monitored by Synchrotron-Radiation Circular Dichroism*. Biophysical Journal, 2019. **117**(4): p. 729-742.
20. Lee, C.W., et al., *Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein*. Biochemistry, 2010. **49**(46): p. 9964-9971.
21. Waters, L., et al., *Structural diversity in p160/CREB-binding protein coactivator complexes*. Journal of Biological Chemistry, 2006. **281**(21): p.14787-14795.

22. Demarest, S.J., et al., *Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators*. *Nature*, 2002. **415**: p. 549-553.
23. Garcia, A.M., et al., *Aggregation and amyloidogenicity of nuclear coactivator binding domain of CREB-binding protein*. *Chemistry – A European Journal*, 2020. **26**: p.9889-9899.
24. Gallardo, R., N.A. Ranson, and S.E. Radford, *Amyloid structures: much more than just a cross- $\beta$  fold*. *Current Opinion in Structural Biology*, 2020. **60**(7): p.7-16.
25. Eichner, T. and S.E. Radford, *Understanding the complex mechanisms of  $\beta$ 2-microglobulin amyloid assembly*. *The FEBS Journal*, 2011. **278**(20): p. 3868-3883.
26. Dasari, A.K.R., et al., *Transthyretin Aggregation Pathway toward the Formation of Distinct Cytotoxic Oligomers*. *Scientific Reports*, 2019. **9**(1): p. 33.
27. Stefanis, L.,  *$\alpha$ -Synuclein in Parkinson's disease*. *Cold Spring Harbor perspectives in medicine*, 2012. **2**(2): p. a009399-a009399.
28. Sengupta, U., A.N. Nilson, and R. Kaye, *The Role of Amyloid- $\beta$  Oligomers in Toxicity, Propagation, and Immunotherapy*. *EBioMedicine*, 2016. **6**: p. 42-49.
29. Stefani, M., *Protein aggregation diseases: toxicity of soluble prefibrillar aggregates and their clinical significance*. *Methods Molecular Biology*, 2010. **648**(25): p.25-41.
30. Devitt, G., et al., *Conformational Evolution of Molecular Signatures during Amyloidogenic Protein Aggregation*. *ACS Chemical Neuroscience*, 2019. **10**(11): p. 4593-4611.
31. Raskatov, J.A. and D.B. Teplow, *Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins*. *Scientific Reports*, 2017. **7**(1): p. 12433.
32. Hayden, E.Y., et al., *Identification of key regions and residues controlling A $\beta$  folding and assembly*. *Scientific Report*, 2017. **7**: p. 1-16.
33. Dutta, S., et al., *New insights into differential aggregation of enantiomerically pure and racemic A $\beta$ 40 systems*. *Peptide Science*, 2019. **111**(6): p. e24139.
34. Torbeev, V., et al., *Chiral recognition in amyloid fiber growth*. *Journal of Peptide Science*, 2016. **22**(5): p. 290-304.
35. Nagy-Smith, K., et al., *Molecular, Local, and Network-Level Basis for the Enhanced Stiffness of Hydrogel Networks Formed from Coassembled Racemic Peptides: Predictions from Pauling and Corey*. *ACS Central Science*, 2017. **3**(6): p. 586-597.
36. Raj, N. and L.D. Attardi, *The Transactivation Domains of the p53 Protein*. *Cold Spring Harbor perspectives in medicine*, 2017. **7**(1): p. a026047.
37. Lane, D. and A. Levine, *p53 Research: the past thirty years and the next thirty years*. *Cold Spring Harbor perspectives in biology*, 2010. **2**(12): p. a000893-a000893.
38. Ferreon, J.C., et al., *Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2*. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. **106**(16): p. 6591-6596.
39. Wells, M., et al., *Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain*. *Proceedings of the National Academy of Sciences*, 2008. **105**(15): p. 5762.
40. Lee, C.W., et al., *Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation*. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. **107**(45): p. 19290-19295.
41. Jenkins, L.M.M., et al., *p53 N-terminal phosphorylation: a defining layer of complex regulation*. *Carcinogenesis*, 2012. **33**(8): p. 1441-1449.
42. Teufel, D.P., M. Bycroft, and A.R. Fersht, *Regulation by phosphorylation of the relative affinities of the N-terminal transactivation domains of p53 for p300 domains and Mdm2*. *Oncogene*, 2009. **28**(20): p. 2112-2118.
43. Lee, C.W., et al., *Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation*. *Proceedings of the National Academy of Sciences*, 2010. **107**(45): p. 19290.

44. Potier, N., et al., *Negative electrospray ionization mass spectrometry of synthetic and chemically modified oligonucleotides*. *Nucleic Acids Research*, 1994. **22**(19): p. 3895-3903.
45. Jenkins, L.M.M., et al., *Two distinct motifs within the p53 transactivation domain bind to the Taz2 domain of p300 and are differentially affected by phosphorylation*. *Biochemistry*, 2009. **48**(6): p. 1244-1255.
46. Lee, C.W., et al., *Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP*. *Biochemistry*, 2009. **48**(10): p. 2115-2124.
47. Valverde, I.E., et al., *Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne–Azide Ligations*. *Angewandte Chemie International Edition*, 2012. **51**(3): p. 718-722.

# Chapitre V : Développement de la spectrométrie de masse pour la caractérisation structurale de complexe biologique par pontage chimique (XL-MS)

Dans ce chapitre, nous décrivons comment l'approche de pontage chimique couplé à la spectrométrie de masse (XL-MS) a été développée afin de caractériser les complexes biologiques non-covalents dans le but d'identifier les zones de proximités entre les partenaires. Nous détaillerons les optimisations que nous avons réalisées afin d'établir un protocole robuste qui pourrait être appliqué de façon rationnelle à un échantillon dont les connaissances structurales sont limitées. Ce développement a été réalisé sur un complexe protéique hétérodimérique issu de la famille des Récepteurs Nucléaires (RNs) : Le Récepteur de l'Acide Rétinoïque (RAR) en interaction avec le Récepteur aux Rétinoïdes X (RXR).

Cette étude s'inscrit dans le cadre d'une collaboration avec l'équipe du Dr. Natacha Rochel au Département de Biologie Structurale Intégrative de l'IGBMC (UMR 7104, Illkirch-Graffenstaden).

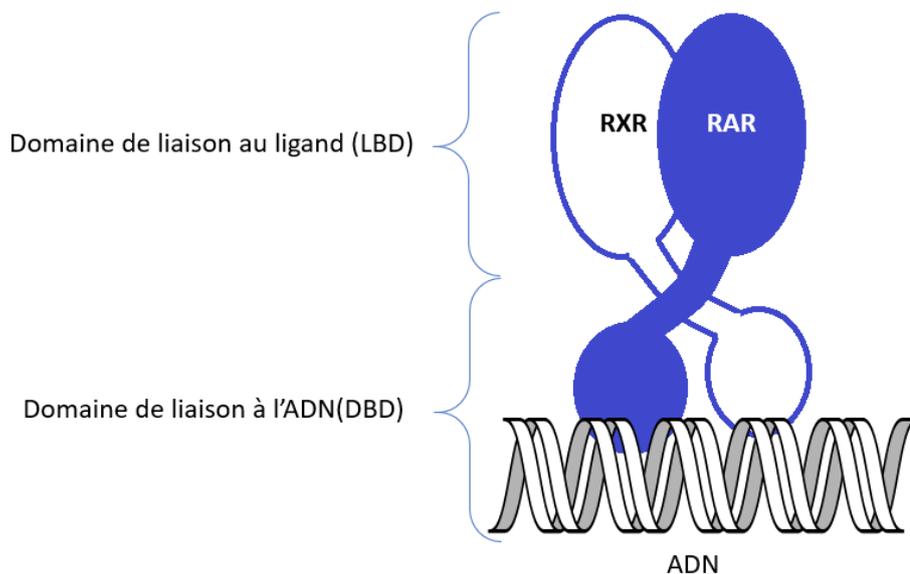
## 1. Introduction

Même si la méthodologie XL-MS repose sur la même technologie que l'approche d'analyse protéomique et que le protocole expérimental devient mieux contrôlé, les scientifiques rencontrent toujours de nombreuses difficultés (dans le choix de l'agent pontant le plus approprié ou des critères clés pour une interprétation fiable des spectres de fragmentation MS/MS des peptides pontés.<sup>[1]</sup>) et des obstacles au succès de l'étude peuvent se glisser à chaque étape de l'expérience. La mise au point du protocole de pontage, l'optimisation des conditions de MS associées, mais surtout l'interprétation des données, ont représenté 70% des activités de recherches de mon travail de thèse.

Ainsi le 1er objectif de mon projet fût d'optimiser les différentes étapes analytiques de l'approche XL-MS afin d'établir un protocole robuste et reproductible. Pour cette étape de développement, nous avons choisi comme modèle d'étude un complexe protéique hétérodimérique issu de la famille des Récepteurs Nucléaires (RNs) : Le Récepteur de l'Acide Rétinoïque (RAR) en interaction avec le Récepteur aux Rétinoïdes X (RXR). Outre son importance biologique, ce système présente plusieurs avantages pour cette optimisation méthodologique :

- Taille moyenne (80kDa) et organisation en domaines structuraux (Cf Chapitre 5.2)
- Protocole de production et purification éprouvé par nos collaborateurs permettant de disposer de quantités de matériel à l'échelle structurale (centaine de microgrammes)
- Cet hétérodimère est stable et a déjà été étudié par MS native entre nos mains<sup>[2]</sup>
- Des données structurales sur les sous-unités isolées ou sur l'hétérodimère sont disponibles dans la littérature ou en interne au laboratoire

Une fois notre protocole XL-MS révisé et optimisé, nous souhaitons l'appliquer pour caractériser l'interaction entre l'hétérodimère RAR/RXR et différents éléments de réponse nucléotidiques<sup>[3]</sup> reflétant l'interaction fonctionnelle entre les RNs et l'ADN et déterminer si l'approche XL-MS permet de mettre en évidence un changement conformationnel résultant.



**Figure 65 : Représentation schématique de RAR/RXR.**

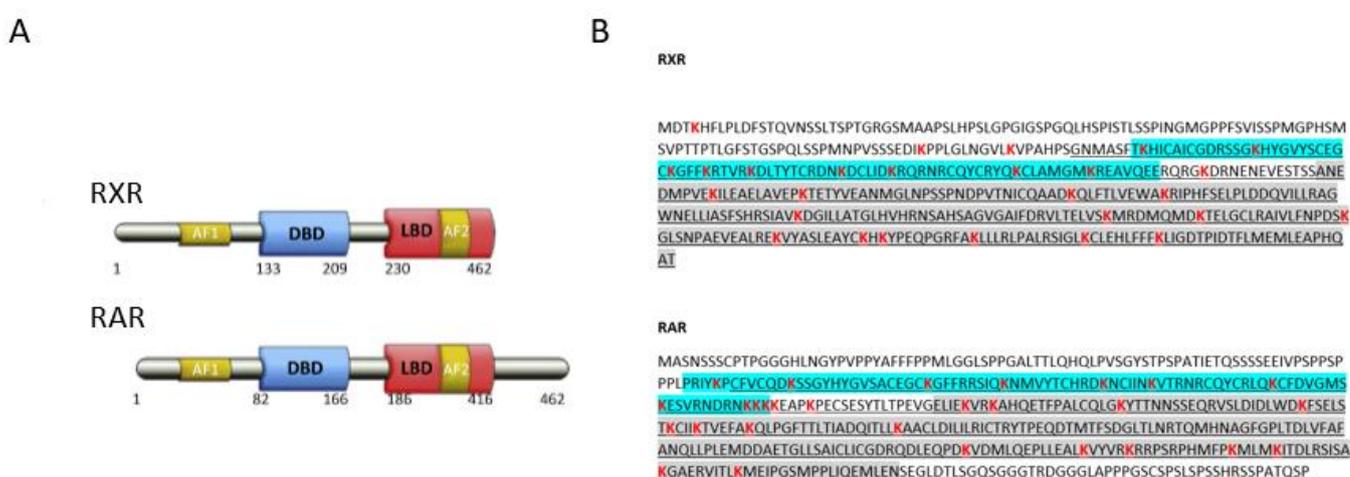
Ainsi nous visons deux objectifs par cette étude :

- Un objectif à « visée connaissance biologique » (l'association entre RAR/RXR et l'ADN s'opère-t-elle avec un changement de conformation protéique et si oui, à quel niveau ? Echelle moléculaire ou atomique ?)
- Un objectif à « visée fondamentale » (quelle résolution peut-on atteindre par l'approche XL-MS ?)

Dans ce projet de recherche, nos collaborateurs mènent en parallèle des études par la DRX.<sup>[4]</sup> Cependant, les différentes conditions de cristallisation testées jusqu'à présent n'ont pas permis d'observer la formation de cristaux pour l'hétérodimère. De plus, les analyses effectuées par d'autres techniques restent délicates. En effet, le complexe faisant une taille d'environ 80kDa, celui-ci est trop gros pour être analysé par RMN et est trop petit pour être analysé en cryo-EM. C'est pourquoi des analyses complémentaires effectués par MS nous semblent prometteuses.

## 2. RAR/RXR comme modèle : mise en place d'un protocole de pontage chimique

Pour démarrer cette étude de développement, nous avons décidé d'appliquer dans un 1<sup>er</sup> temps un protocole standard issu de la littérature<sup>[5]</sup> mettant en œuvre un agent pontant couramment utilisé en XL-MS : le bis(sulfosuccinimidyl) suberate (BS3, Cf chapitre III). Celui-ci nous a semblé intéressant car disponible commercialement pour un nombre d'essais élevé, possédant une longueur de bras espaceur intermédiaire en 1<sup>ère</sup> intention et se fixant majoritairement sur les lysines.<sup>[6, 7]</sup> qui est un constituant majeur de nos deux séquences protéiques.



**Figure 66 : (A) Représentation schématique des différents domaines observés sur RAR et RXR. (B) Séquence des protéines RAR et RXR. L'ensemble des lysines du complexes ont été mises en rouge. Le soulignement correspond au domaine LBD (surligné en bleu) et DBD (surligné en gris).**

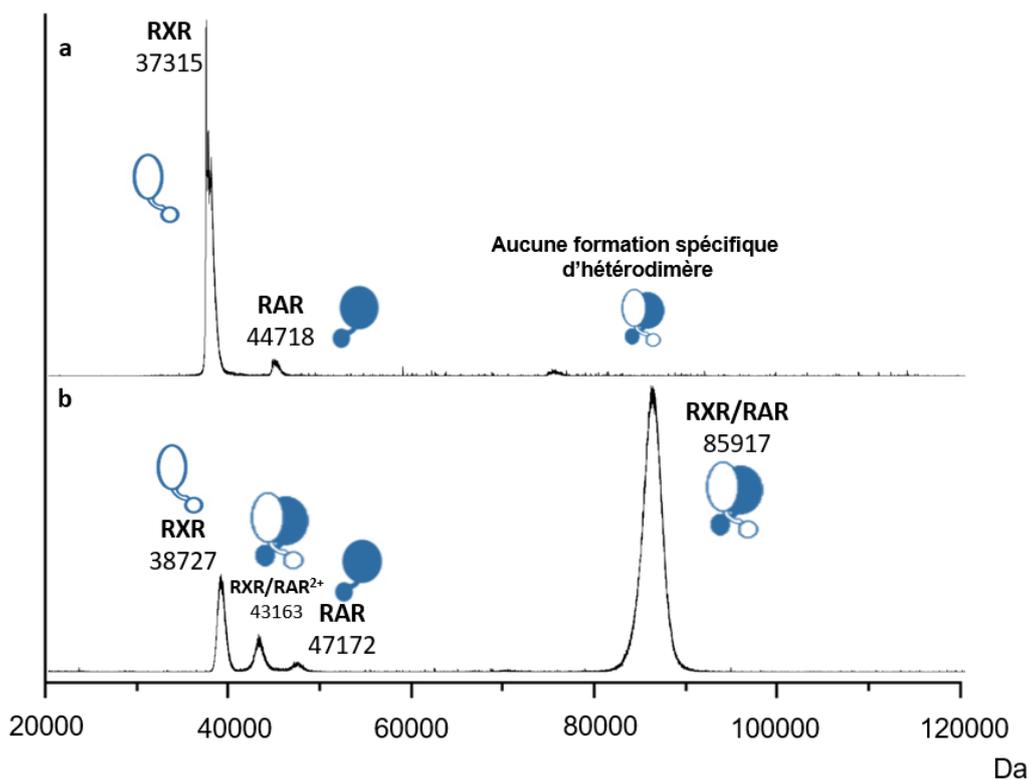
Alors que les pontages intermoléculaires permettent de comprendre comment interagissent les différentes sous-unités protéiques au sein d'un complexe, les pontages intramoléculaires permettent de positionner les différents domaines présents au sein de la protéine en apportant ainsi des informations complémentaires sur la structure.<sup>[5, 8, 9]</sup> Dans un souci d'efficacité et de spécificité (éliminer les risques d'interprétation erronée entre pontage intramoléculaire et pontage homodimérique potentiel de RXR), nous avons décidé de nous focaliser sur le pontage intermoléculaire entre RAR et RXR.

### 2.1 Suivi de la réaction de pontage par MALDI-MS

Dans le cadre de cette étude, nous avons réalisé la réaction de pontage sur notre complexe RAR/RXR concentré à 4.1 mg/ml soit 40 $\mu$ M à température ambiante pendant 30min en utilisant un ratio protéine/agent pontant 1 :50. La réaction de pontage est alors suivie par MALDI-MS. Cette technique est rapide et avait déjà montré son potentiel lors d'une précédente étude XL-MS au laboratoire<sup>[10]</sup>. Cette concentration relativement élevée d'échantillon est favorable pour une bonne détection en MS ou en SDS-PAGE mais aussi afin d'augmenter les chances de détecter efficacement les peptides pontés très minoritaires après digestion durant cette phase d'optimisation.

Après purification, le complexe RAR/RXR est obtenu dans un tampon salin constitué de Hepes 20mM pH7.5, NaCl 75mM, KCl 75mM, Glycérol 1%, MgSO4 4mM, TCEP 1mM. Les 150 mM de sels sont nécessaires à une bonne solubilisation de l'hétérodimère et le glycérol est systématiquement ajouté pour favoriser la stabilité et le stockage de l'échantillon. Bien que ce tampon soit efficace pour les analyses en SDS-PAGE, celui-ci est incompatible avec les analyses en MALDI-MS principalement en raison de la présence du glycérol qui perturbe la cristallisation de la matrice avec l'échantillon et empêche ainsi l'analyse. C'est pourquoi, celui-ci a été éliminé avant analyse par MS.

Un échange de tampon est ainsi réalisé à l'aide de cartouche de gel d'exclusion (BioSpin PM6kDa, Biorad) avant introduction de l'échantillon dans la source du spectromètre de masse. Cet échange peut être réalisé indifféremment avant ou après réaction de pontage. En effet, le glycérol ne semble pas perturber la réaction de pontage comme le montre l'intensité du pic correspondant à l'hétérodimère cross-linké (Figure 67) et comme cela avait déjà été observé lors de l'étude précédente sur le complexe SAGA. Pour des raisons pratiques et afin de comparer aisément les différentes conditions de pontage, nous avons choisi de réaliser le dessalage des échantillons systématiquement en amont du plan d'expérience, avant le pontage. Après échange, le complexe est ainsi présent dans un tampon Hepes à 20mM avec 150mM NaCl à pH 7.5.



**Figure 67 : Suivi cinétique de la réaction de pontage par MALDI-MS (a) Spectre MS avant pontage (b) Spectre MS après pontage par le BS3 (ratio protéine/BS3 : 1/50 ; 30min à T°C ambiante). L'échantillon est analysé selon la méthode de la goutte séchée en utilisant l'acide sinapinique (SA) comme matrice concentrée à 10 mg/mL sur un instrument de type Autoflex Speed (Bruker)**

Des ions à  $m/z$   $44718 \pm 70$  Da et  $37315 \pm 70$  Da sont observés en l'absence de pontage. Ces ions correspondent respectivement aux sous-unités RAR et RXR qui constituent le

complexe. En effet, les interactions non-covalentes qui maintiennent l'intégrité du complexe d'origine sont rompues durant l'analyse MALDI-MS et les deux sous-unités sont ainsi détectées individuellement. La différence d'intensité de signal entre les deux récepteurs nucléaires s'explique par un pouvoir d'ionisation différent et donc une réponse du détecteur qui diffère entre RAR et RXR, même constat si l'échantillon est analysé par ESI. Après pontage, un signal à  $m/z$   $85917 \pm 200$  Da est observé. Cette masse reflète l'existence du complexe hétérodimérique RAR/RXR stabilisé covalentement par le cross-linker selon une stœchiométrie 1 :1. L'écart-type de mesure est relativement important d'une part à cause de la présence de sels dans l'échantillon (élargissement des signaux) mais aussi car l'intensité du laser est augmentée afin de favoriser la détection des ions dimériques. L'intensité du signal montre ainsi une bonne efficacité de réaction (après 30min de réaction 70-80% du complexe est ponté).

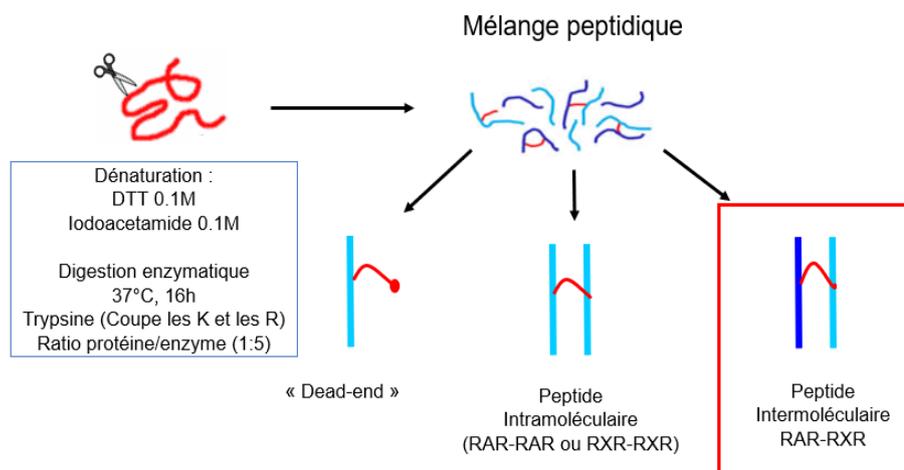
A ce stade, il n'est pas possible de déduire si les monomères détectés sont issus d'une réaction de pontage incomplète ou bien d'un équilibre monomère/hétérodimère en solution. Des expériences en MS native ont donc été effectuées (avant pontage), mais le transfert de l'échantillon dans un tampon d'acétate d'ammonium fut délicat et il s'est avéré difficile d'obtenir des spectres de masse de bonne qualité, ne permettant pas de trancher entre ces deux hypothèses.

A partir de ces spectres, aucun pic correspondant à un homodimère de type RAR/RAR (MW  $\approx$  89.4 kDa) ou RXR/RXR (MW  $\approx$  74.6 kDa) ou encore à une autre stœchiométrie d'interaction n'est observé démontrant ainsi la spécificité de la réaction de pontage et nous prouvant que l'espèce cross-linkée reflète bien le complexe initialement présent en solution.

## **2.2 Digestion enzymatique : formation de peptides pontés pour analyse MS**

Comme décrit dans le chapitre II, afin de caractériser les zones protéiques pontées, la stratégie protéomique Bottom-up a été appliquée.

Dans un souci d'éliminer les sels et excès d'agents de pontage avant introduction dans notre système nanoLC, nous avons dans un premier temps précipité nos protéines pontées avant de les digérer en suivant le protocole proposé par la plateforme protéomique de l'IBMC (voir annexe 5). Après digestion à la trypsine, les échantillons sont ensuite évaporés et resolubilisés dans de l'eau avec 0,1% d'acide formique (FA) avant d'effectuer l'analyse en nanoLC-MS/MS.



**Figure 68 : Schéma montrant les différents types de peptides pontés formés après digestion. Nous nous sommes focalisés sur les peptides pontés intermoléculaires.**

Comme nous l'avons souligné dans le chapitre bibliographie, un instrument de MS alliant une haute précision de mesure de masse, une bonne sensibilité et une grande vitesse d'acquisition des données est nécessaire d'une part pour pallier aux effets de coélution possible des peptides, et d'autre part pour permettre la sélection des ions pontés peu intenses et l'obtention de spectres de fragmentation suffisamment précis pour compenser la statistique de peptides pontés théoriques<sup>[11, 12]</sup>. Les données brutes sont ensuite converties en format MGF (Mascot Generic Format) par le logiciel Proteome Discoverer™ pour requête dans les logiciels d'interprétation.

Une interprétation protéomique classique a été réalisée à l'aide de l'algorithme de recherche MASCOT afin d'évaluer l'efficacité de la digestion par la trypsine sur le complexe ponté. Ainsi, le complexe RAR/RXR a été identifié avec 94% de recouvrement de séquence pour RAR et 86% de séquence pour RXR. Ce recouvrement de séquence est parfaitement suffisant pour identifier correctement le complexe protéique et nous indique que la digestion trypsine a été effective malgré le pontage.

### **2.3 Identification des peptides pontés**

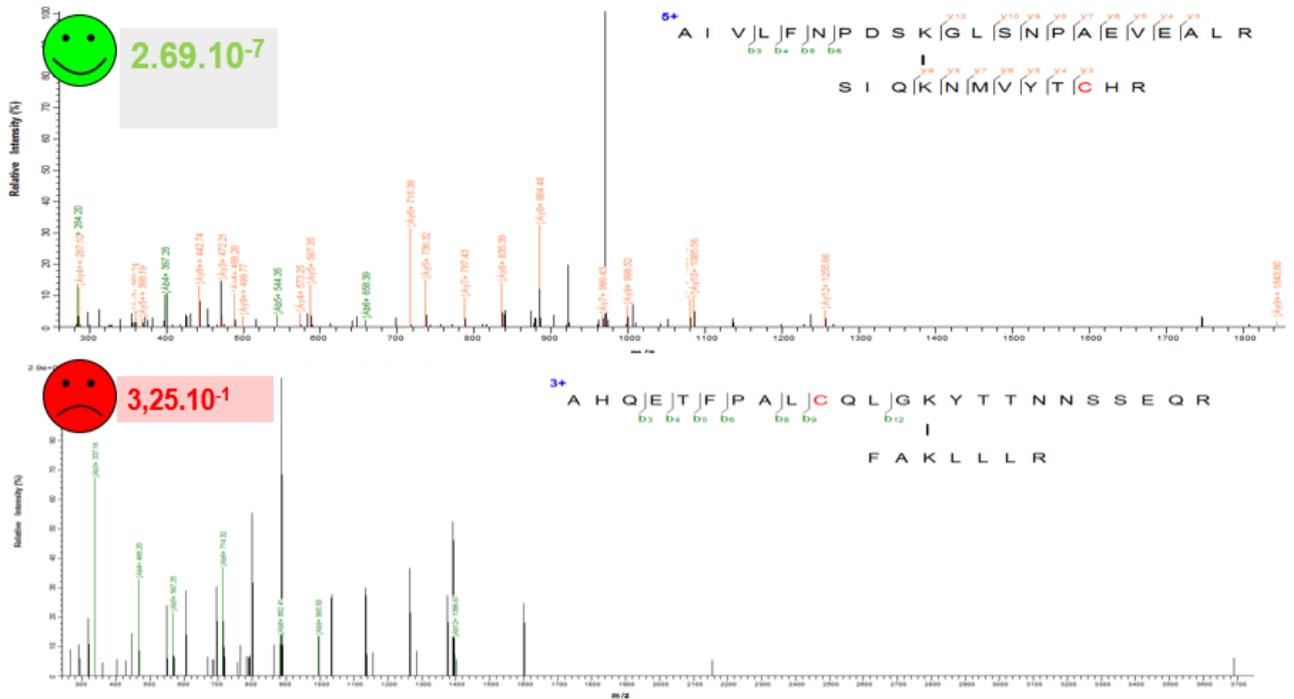
Comme déjà décrits dans le chapitre III de la partie bibliographie, des logiciels dédiés à l'identification des peptides pontés sont requis. Ceci est dû à la grande complexité du mélange peptidique obtenu après pontage ainsi qu'à la multitude de combinaisons théoriques possibles de peptides pontés. Ainsi, au début de ma thèse, beaucoup d'algorithmes étaient disponibles en ligne mais malgré nos efforts, peu nous permettaient d'obtenir des résultats robustes entre nos mains. Le logiciel pLink avait démontré un bon potentiel lors de l'analyse du complexe SAGA, c'est donc naturellement vers lui que nous nous sommes tournés en première intention.

En utilisant des paramètres de requête standards (erreur tolérée sur les spectres MS et MS/MS à 10ppm et à 20 ppm respectivement, nombre de coupure partielle de l'enzyme à 4, un minimum de 5 acides aminés par peptide, FDR à 5%), 37 peptides pontés dont 20 étaient des peptides intermoléculaires ont été proposés par pLink (Cf tableau 6). Sur les 37 peptides pontés proposés par pLink, 34 ont été observés avec une précision de mesure <2ppm sur l'ion précurseur.

**Tableau 6 : Liste de l'ensemble des peptides pontés détectés par pLink. En jaune sont identifiés les peptides intermoléculaires.**

Position des acides aminés pontés sur les protéines	Peptides	Score	Charge	Masse ion précurseur	Types de peptides
RXR (381)-RXR (181)	AIVLFNPDGKLSNPAEVEALR(10)-DNKDCLIDKR(9)	4.213445e-008	4	3753.95	intra
RXR (381)-RXR (165)	AIVLFNPDGKLSNPAEVEALR(10)-KDLYTCR(1)	1.922596e-015	4	3533.83	intra
RXR (381)-RAR (118)	AIVLFNPDGKLSNPAEVEALR(10)-SIQKNMVTYCHR(4)	2.691832e-007	5	4014.05	inter
RXR(381)-RAR(390)	AIVLFNPDGKLSNPAEVEALR(10)-SISAKGAER(5)	8.811572e-009	4	3395.82	inter
RXR (381)-RXR (165)	AIVLFNPDGKLSNPAEVEALREK(10)-KDLYTCR(1)	7.713615e-021	4	3790.96	intra
RXR(381)-RAR(390)	AIVLFNPDGKLSNPAEVEALREK(10)-SISAKGAER(5)	1.698535e-009	5	3652.95	inter
RAR(155)-RXR(160)	CFDVGMSKESVR(8)-GFFKR(4)	7.367377e-005	4	2206.07	inter
RAR(155)-RXR(165)	CFDVGMSKESVR(8)-KDLYTCR(1)	1.081178e-010	4	2608.22	inter
RAR(128)-RXR(165)	DNKCIINKVTR(2)-KDLYTCR(1)	2.942936e-006	4	2554.31	inter
RAR(134)-RXR(175)	DNKCIINKVTR(8)-DNKDCLIDKR(3)	3.206857e-014	5	2774.42	inter
RAR(134)-RXR(165)	DNKCIINKVTR(8)-KDLYTCR(1)	4.136504e-010	4	2554.31	inter
RAR (134)-RAR (390)	DNKCIINKVTR(8)-SISAKGAER(5)	5.229637e-003	3	2416.29	intra
RXR (364)-RXR (407)	DMQMDKTELGLR(6)-HKYPEQGR(2)	4.047485e-001	4	2845.34	intra
RXR (364)-RXR (165)	DMQMDKTELGLR(6)-KDLYTCR(1)	5.297877e-013	3	2790.29	intra
RXR (364)-RAR (390)	DMQMDKTELGLR(6)-SISAKGAER(5)	2.475634e-003	3	2652.27	inter
RXR (181)-RXR (417)	DNKDCLIDKR(9)-FAKLLLR(3)	2.989973e-002	4	2274.26	intra
RXR (181)-RAR (390)	DNKDCLIDKR(9)-SISAKGAER(5)	9.459326e-006	3	2332.19	inter
RXR (181)-RXR (194)	DNKDCLIDKR(9)-YQKCLAMGMK(3)	5.141640e-022	3	2643.27	intra
RAR(171)-RXR(407)	EAPKPECSYTLTPEVGELIEK(4)-HKYPEQGR(2)	1.185963e-001	4	3854.87	inter
RAR(171)-RXR(165)	EAPKPECSYTLTPEVGELIEK(4)-KDLYTCR(1)	4.282860e-012	4	3799.82	inter
RAR(118)-RXR(165)	GFFRRSIQKNMVTYCHRDK(9)-RTVRKDLTYTCRDNK(5)	7.497403e-002	5	4506.30	inter
RXR(213)-RXR(201)	GKDRNENEVESTSSANEDMPVEK(2)-CLAMGMKR(7)	1.376223e-001	5	3668.66	intra
RXR(213)-RXR(165)	GKDRNENEVESTSSANEDMPVEK(2)-KDLYTCR(1)	6.894686e-010	4	3758.71	intra
RXR(213)-RAR(118)	GKDRNENEVESTSSANEDMPVEK(2)-SIQKNMVTYCHR(4)	1.147724e-008	5	4238.93	inter
RXR(165)-RXR(201)	KDLYTCR(1)-CLAMGMKR(7)	1.199750e-005	4	2160.04	intra
RXR(165)-RXR(160)	KDLYTCR(1)-GFFKR(4)	1.011355e-014	4	1847.94	intra
RXR (165)-RAR (390)	KDLYTCR(1)-SISAKGAER(5)	1.494738e-010	3	2112.07	inter
RAR(118)-RXR(201)	SIQKNMVTYCHR(4)-CLAMGMKR(7)	8.342400e-010	4	2640.26	inter
RAR (118)-RXR (181)	SIQKNMVTYCHR(4)-DNKDCLIDKR(9)	5.214884e-003	4	2950.43	inter
RAR(118)-RXR(160)	SIQKNMVTYCHR(4)-GFFKR(4)	1.178363e-003	4	2328.17	inter
RAR (118)-RXR (165)	SIQKNMVTYCHR(4)-KDLYTCR(1)	2.858785e-005	5	2730.31	inter
RAR (118)-RAR (390)	SIQKNMVTYCHR(4)-SISAKGAER(5)	3.647694e-007	4	2592.30	intra
RAR(390)-RXR(417)	SISAKGAER(5)-FAKLLLR(3)	3.302204e-004	4	1916.13	inter
RAR (390)-RAR (390)	SISAKGAER(5)-SISAKGAER(5)	4.491236e-002	3	1974.06	intra
RXR (145)-RXR (201)	SSGKHGYYVYCEGCK(4)-CLAMGMKR(7)	3.863667e-019	3	2822.25	intra
RXR (356)-RXR (417)	VLTELVSKMR(8)-FAKLLLR(3)	4.914015e-014	3	2173.31	intra

Ces peptides ont été sélectionnés par pLink à partir de leur score. Avec pLink, plus le score est proche de 0, plus la confiance en l'identification des spectres MS/MS sera acceptable.<sup>[13]</sup> Voici dans la figure 69 un exemple de spectres MS/MS correspondant à une interprétation telle que présentée par pLabel, le logiciel de visualisation des spectres associés à pLink.



**Figure 69 : Exemple de spectre MSMS des peptides pontés montrant la corrélation entre le score et la qualité d'identification. Le spectre du haut correspond au peptide intermoléculaire RXR381-RAR118**

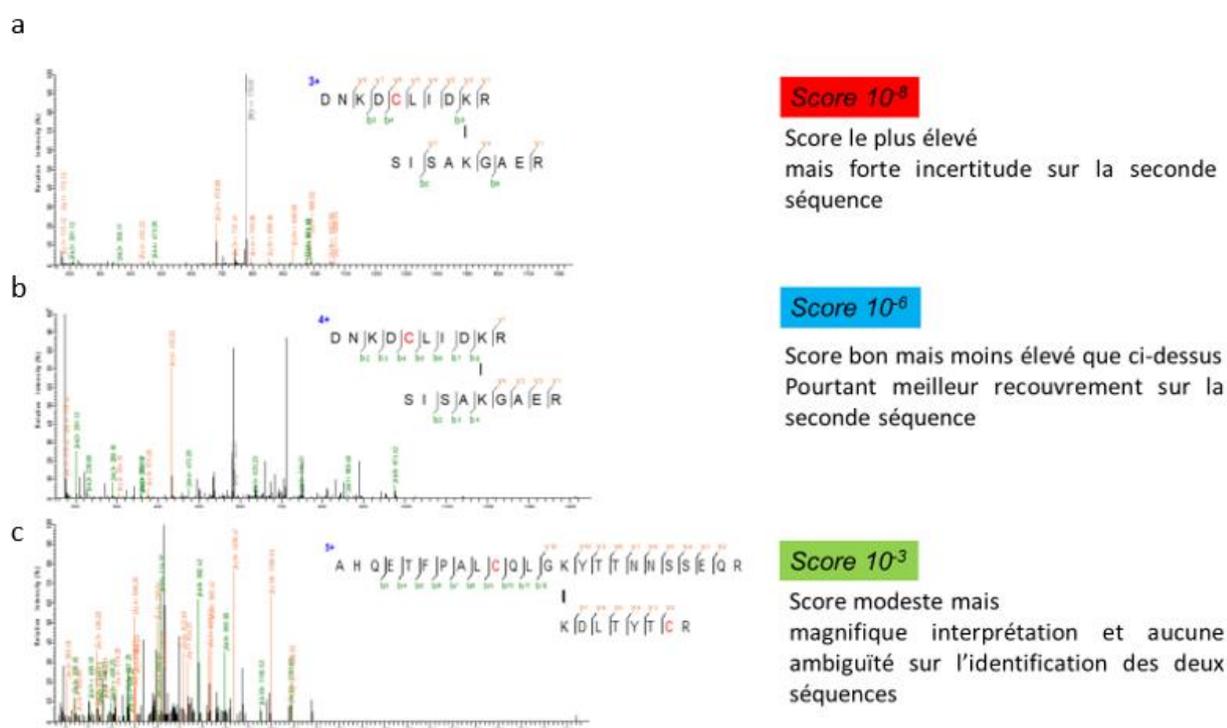
Le spectre du haut correspond au spectre de fragmentation d'un ion chargé 5 fois de rapport  $m/z = 802.81$ . Une majorité du signal a été corrélée aux ions fragments b et y de deux peptides tryptiques issus de RXR et de RAR pontés par les lysines (RXR381 et RAR118). Le bon recouvrement de séquence sur les deux peptides pontés en accord avec un bon score de  $2.69 \cdot 10^{-7}$ , nous conforte dans une bonne interprétation du logiciel et dans la confiance de la réelle existence de ce peptide ponté dans notre digest peptidique après pontage. A l'inverse, le spectre du bas, issu de la fragmentation d'un ion chargé 3 fois de rapport  $m/z = 1240.02$ , montre une majorité de signaux non interprétés sur le spectre MS/MS, un peptide  $\beta$  non reconnu, associés à un score de  $3.25 \cdot 10^{-1}$ , qui suggère que la proposition d'interprétation du logiciel pour ce spectre est peu fiable.

Dès lors, nous nous sommes attachés à rechercher une relation entre score d'identification et pertinence de l'identification. Comment s'assurer de la bonne assignation des ions fragments par le logiciel d'interprétation et ainsi, de la présence effective d'un peptide ponté dans notre mélange peptidique de départ ? Notre 1<sup>ère</sup> démarche fut de rechercher une valeur seuil de score, en dessous de laquelle l'identification ne serait pas validée. Pour cela, nous avons lancé les requêtes en utilisant les mêmes paramètres de recherche que précédemment décrits pour l'analyse mais en incluant volontairement une erreur qui devrait bloquer une interprétation rationnelle (masse erronée de l'agent de pontage, base de données protéiques erronée). Malheureusement, les scores attribués par pLink lors de ces recherches falsifiées n'étaient pas significativement réduits au point de créer un « score seuil » de validation dans l'interprétation. D'autre part, un examen attentif des spectres de fragmentation MS/MS nous a très vite révélé qu'il n'y avait pas de relation évidente entre la qualité de l'interprétation d'un peptide ponté et le score attribué par l'algorithme.

Voici quelques exemples illustrant une incohérence entre le score attribué par le logiciel et l'œil humain subjectif de l'analyste :

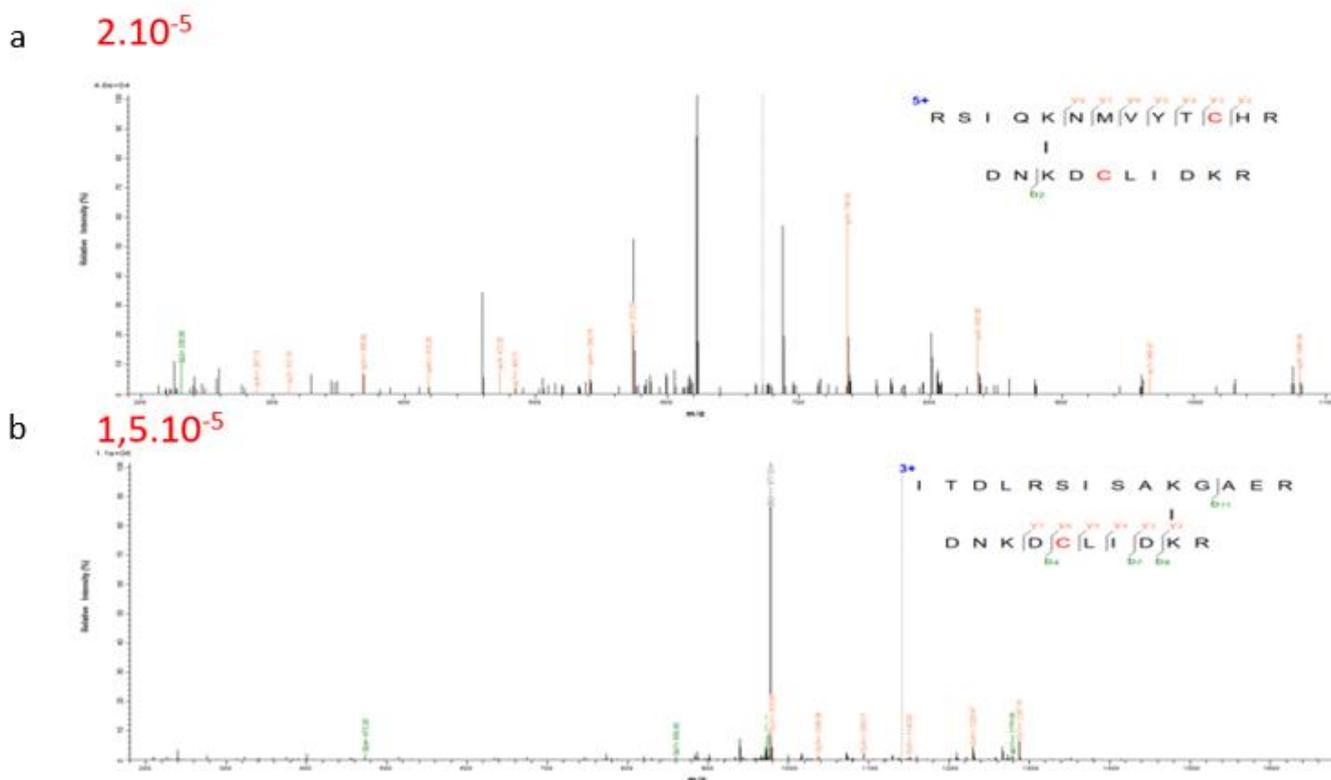
✦ *Le score ne reflète pas toujours un bon recouvrement de séquence sur les deux séquences du peptide ponté :*

La figure 70 nous montre que la valeur des scores attribuée par pLink n'est pas linéaire avec la confiance en l'interprétation telle que l'analyste la qualifie. En effet, un signal correspondant aux ions fragments b et y permettant de recouvrir quasiment l'intégralité de séquence des peptides pontés est observé sur les spectres 70b et 70c alors que les scores ( $10^{-6}$  et  $10^{-3}$ ) sont respectivement inférieurs au spectre 70a (score de  $10^{-8}$ ) pour lequel l'identification du second peptide reste ambiguë.



**Figure 70 : Relation non linéaire entre le score attribué par pLink v2.5 et la confiance en l'interprétation. Représentation par pLabel des spectres de fragmentation de peptides pontés suite à une requête dans le logiciel pLink. Tous ces spectres sont issus d'une même analyse LC-MS/MS du complexe RAR/RXR pontés par BS3.**

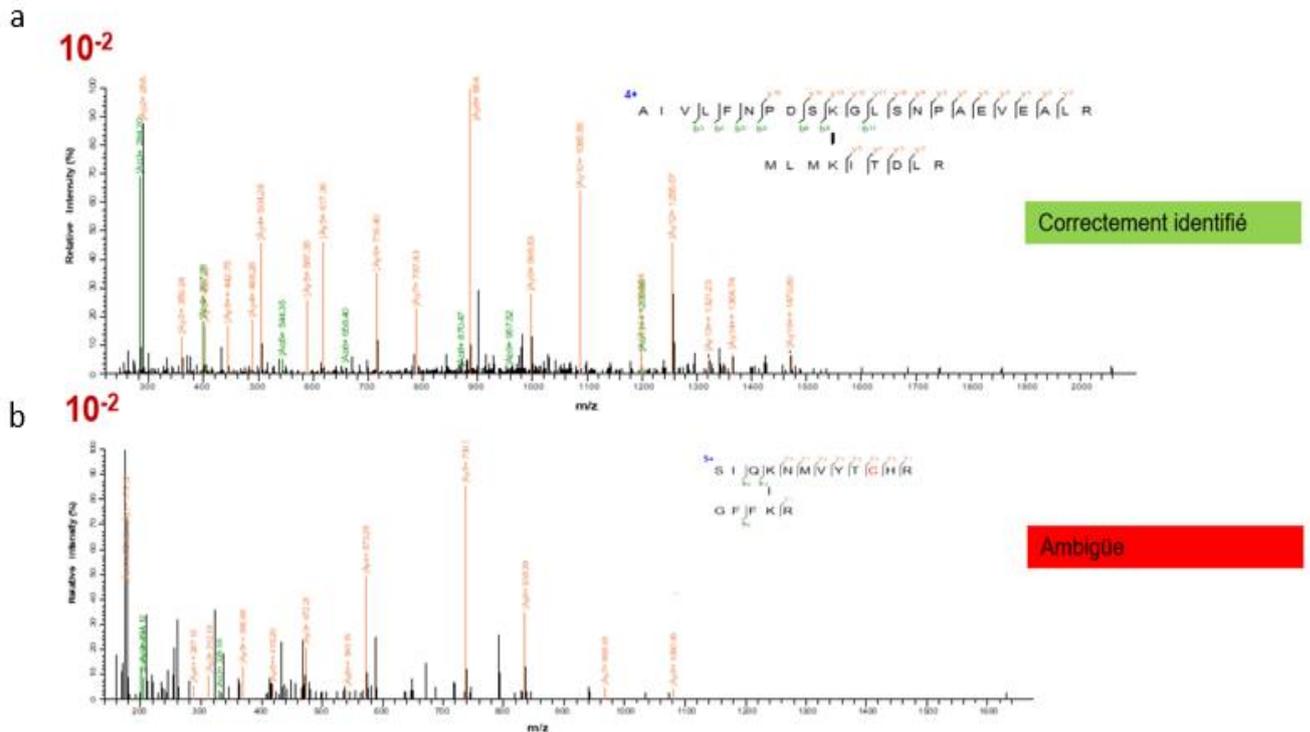
Ce cas de figure où la séquence d'un seul des deux peptides pontés est correctement identifiée malgré un score pLink convenable ( $>10^{-5}$ ) fut très répandu lors de nos analyses sur le complexe RAR/RXR mais aussi sur d'autres systèmes biologiques (complexe NuA4) et ce, quels que soient le cross-linker ou la méthode d'acquisition par MS choisis. La figure 71 montre deux exemples supplémentaires de spectres dont les ions fragments ne permettent d'identifier avec certitude qu'une seule des deux séquences réticulées. Alors que plusieurs combinaisons de peptides cross-linkés peuvent conduire à des masses très proches, il est crucial de disposer d'ions fragments qui valident la séquence des deux peptides connectés.



**Figure 71 : Exemple de spectres de fragmentation correspondant à des peptides pontés caractérisés par pLink avec des scores de  $2 \cdot 10^{-5}$  (a) et de  $1,5 \cdot 10^{-5}$  (b). Malgré un bon score d'identification, les peptides  $\beta$  (spectre a) et  $\alpha$  (spectre b) sont incertains**

✦ A score égal, la confiance en l'identification n'est pas équivalente :

La figure 72 représente deux spectres de fragmentation dont l'interprétation par pLink est jugée similairement fiable puisque le score attribué est équivalent ( $10^{-2}$ ). Néanmoins, alors que du signal correspondant à des ions fragments des deux séquences pontées est bien observé en 72a, aucun signal n'est observé pour l'attribution de la séquence peptide  $\beta$  en 72b. Il est dans ce cas délicat de confirmer la séquence du second peptide rendant l'identification du peptide ponté ambiguë.



**Figure 72 : Exemple de peptides pontés identifiés avec des scores pLink similaires ( $10^{-2}$ ) mais une certitude visuelle différente.**

Ainsi, cette étude par le logiciel pLink, nous a montré qu’une analyse purement automatisée, bien que nécessaire à la vue du set colossal de données MS/MS, doit encore être considérée avec précaution pour une identification robuste des peptides cross-linkés. De nombreuses incertitudes ou l’utilisation non vérifiée de filtres ou seuil de score, peuvent être à l’origine d’une mauvaise interprétation.<sup>[14]</sup> Ces mauvaises interprétations de cross-links se traduiront par de mauvaises conclusions du point de vue structural et ainsi par la génération (ou la reconstruction) d’un modèle structural moins précis, voire faux.

C’est pourquoi nous avons décidé d’effectuer systématiquement une validation manuelle, stricte et rigoureuse afin de ne conserver que les propositions de séquences pontées à fort taux de confiance.

## **2.4 La validation manuelle des peptides pontés comme moyen de contrôle de la qualité d’identification des spectres ?**

Alors qu’un grand nombre de logiciels d’interprétation de données de cross-links sont disponibles dans la littérature<sup>[15]</sup> et que le nombre de publications dans ce domaine augmente chaque année, aucun critère universel n’était décrit au démarrage de ce projet quant à la validation des données d’interprétation de spectres MS/MS d’ions pontés.

Après notre 1<sup>ère</sup> expérience avec pLink, nous avons répertorié un certain nombre de critères qui nous ont semblé importants pour effectuer une validation manuelle :

- Combien d’ions fragments doivent être assignés sur les spectres MS/MS pour qu’un peptide ponté soit considéré comme identifié ? Combien d’acides aminés consécutifs doivent représenter chaque peptide connecté ?
- Les ions fragments doivent-ils être détectés sur un ou sur les deux peptides ?

- La précision de mesure de masse de l'ion précurseur est-elle importante pour valider un peptide ponté ?
- La redondance des spectres MS/MS identifiés pour un peptide ponté est-elle nécessaire ?
- Doit-on définir une intensité seuil pour la détection des ions fragments ?
- Doit-on définir un taux seuil d'interprétation par spectre MS/MS, en d'autres termes un ratio seuil entre l'intensité du signal interprété par rapport à l'intensité du signal total) ?

Plus difficile encore : quelle est l'importance de ces différents critères sur l'identification des peptides pontés ? Quel poids leur accorder lors du calcul du score ou de la validation visuelle ?

Répondre à chacune de ces questions n'est pas facile car très subjectif selon l'analyste et d'autre part, les cas de figure de spectres MS/MS rencontrés sont très diversifiés selon les séquences, les tailles ou encore la concentration des peptides analysés.

Aussi, au fur et à mesure de l'expérience acquise lors de ce travail et également suite aux discussions avec des scientifiques d'autres laboratoires impliqués dans cette même problématique (symposium of structural protéomique Berlin2018, Göttingen2019), nous avons été amenés à choisir des valeurs seuils qui nous ont semblé raisonnables. Ces valeurs ont par la suite été utilisées lors de toutes nos validations manuelles d'interprétation.

Voici les paramètres de validation manuelle que nous avons choisis :

- Minimum 4 acides aminés consécutifs identifiés sur chacun des deux peptides
- Les ions fragments qui servent à l'identification ne sont pas issus du bruit de fond.
- Des ions fragments pontés (c'est-à-dire contenant l'agent ponté à la seconde séquence) renforcent la validation mais ne sont pas impératifs.
- L'erreur de mesure de masse sur l'ion précurseur doit être inférieure à 2 ppm
- La majorité du signal observé sur le spectre MS/MS doit être interprété mais la présence de signal non assigné n'est pas un critère principal de rejet de l'interprétation si tous les critères précédents sont respectés (effet de coélution/cofragmentation possible)
- La redondance soit par l'état de charge de l'ion précurseur, soit par une digestion trypsique incomplète, renforce la confiance en l'interprétation des données de MS/MS, mais n'est considérée pour l'identification que si les critères précédents sont recouverts

Afin de tester ces paramètres, nous avons appliqué notre méthodologie sur une protéine standard dont la structure est décrite : la Serum Albumin Bovine (BSA). La réaction de cross-link a été réalisée exactement comme décrite précédemment sur l'hétérodimère RAR/RXR avec du BS3, puis une recherche des peptides de pontage a été lancée sur le logiciel pLink.

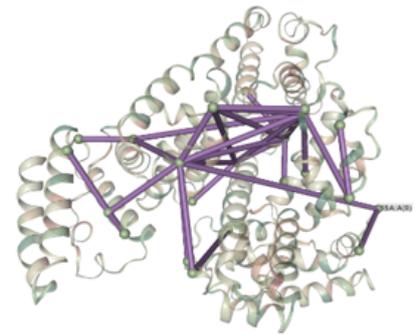
a

Peptides	Charge	Proteine	Score
ALKAWSVAR(3)-LSQKFPK(4)	3	BSA(211)-BSA(221)	1.070741e-010
ALKAWSVAR(3)-SLGKVGTR(4)	3	BSA(211)-BSA(431)	1.960884e-007
ATEEQLKTVMENFVAFVDK(7)-KQTALVELLK(1)	3	BSA(544)-BSA(524)	3.694891e-010
ATEEQLKTVMENFVAFVDK(7)-KQTALVELLKHKPK(1)	6	BSA(544)-BSA(524)	1.234518e-008
ATEEQLKTVMENFVAFVDK(7)-QIKKQKQALVELLK(4)	4	BSA(544)-BSA(524)	7.607259e-006
ATEEQLKTVMENFVAFVDKCCAAADDK(7)-KQTALVELLK(1)	4	BSA(544)-BSA(524)	3.359311e-011
CASIQKFGGER(6)-ALKAWSVAR(3)	3	BSA(204)-BSA(211)	2.715391e-011
CASIQKFGGER(6)-CCTKPESER(4)	3	BSA(204)-BSA(439)	2.217057e-008
CASIQKFGGER(6)-LSQKFPK(4)	3	BSA(204)-BSA(439)	1.615211e-009
CASIQKFGGER(6)-SLGKVGTR(4)	4	BSA(204)-BSA(431)	6.457899e-006
CCTKPESER(4)-ALKAWSVAR(3)	4	BSA(439)-BSA(211)	2.674581e-004
CCTKPESER(4)-LSQKFPK(4)	3	BSA(439)-BSA(221)	2.274513e-009
CCTKPESER(4)-SLGKVGTR(4)	3	BSA(439)-BSA(431)	5.458920e-010
CCTKPESERMPCTEDYLSLILNR(4)-SLGKVGTR(4)	5	BSA(114)-BSA(431)	2.483566e-007
DDSPDLPKLPDPNTLCEDEFK(8)-SLGKVGTR(4)	3	BSA(114)-BSA(431)	3.186046e-008
DDSPDLPKLPDPNTLCEDEFK(8)-SLGKVGTR(4)	5	BSA(114)-BSA(431)	1.850595e-005
DDSPDLPKLPDPNTLCEDEFK(8)-SLGKVGTR(4)	3	BSA(114)-BSA(431)	2.342645e-008
DTHKSEIAHR(1)-SLGKVGTR(4)	5	BSA(1)-BSA(431)	2.600797e-002
DTHKSEIAHR(1)-FKDLGEEHF(2)	4	BSA(4)-BSA(12)	2.669368e-001
EKVLASSAR(6)-LSQKFPK(4)	4	BSA(187)-BSA(221)	1.614620e-003
EKVLASSARQR(2)-LSQKFPK(4)	3	BSA(187)-BSA(221)	1.614620e-003
FKDLGEEHF(2)-DTHKSEIAHR(1)	4	BSA(12)-BSA(1)	2.975957e-009
KQTALVELLK(1)-SLGKVGTR(4)	4	BSA(524)-BSA(431)	5.171424e-008
KVPQVSTPTLVEVSR(1)-HKPKATEEQLK(4)	5	BSA(413)-BSA(537)	2.662763e-008
KVPQVSTPTLVEVSR(1)-LSQKFPK(4)	3	BSA(413)-BSA(221)	3.166289e-007
KVPQVSTPTLVEVSR(1)-SLGKVGTR(4)	4	BSA(413)-BSA(431)	5.680651e-009
LAKEYATELEECCK(3)-ALKAWSVAR(3)	3	BSA(350)-BSA(211)	3.729786e-010
LCVLHEKTPVSEK(7)-CASIQKFGGER(6)	3	BSA(465)-BSA(204)	5.232144e-012
LCVLHEKTPVSEK(7)-CASIQKFGGER(6)	4	BSA(465)-BSA(204)	7.177325e-008
LFTFHADICTLPDTEKQIK(16)-KQTALVELLK(1)	3	BSA(520)-BSA(524)	4.038876e-004
LFTFHADICTLPDTEKQIK(16)-KQTALVELLKHKPK(1)	7	BSA(520)-BSA(524)	1.064893e-004
LKECCDKP(7)-LSQKFPK(4)	3	BSA(280)-BSA(221)	2.422731e-007
LKPDNTLCEDEFK(2)-SLGKVGTR(4)	3	BSA(116)-BSA(431)	7.832925e-011
LKPDNTLCEDEFK(2)-SLGKVGTR(4)	5	BSA(116)-BSA(431)	9.687929e-008
LKPDNTLCEDEFK(2)-NECFLSHKDDSPDLPK(8)	4	BSA(116)-BSA(106)	1.317356e-003
LKPDNTLCEDEFK(2)-SLGKVGTR(4)	3	BSA(116)-BSA(431)	2.393079e-008
LKPDNTLCEDEFK(2)-SLGKVGTR(4)	4	BSA(116)-BSA(431)	1.537425e-006
LRCASIQKFGGER(8)-ALKAWSVAR(3)	5	BSA(204)-BSA(211)	2.223274e-008
LRCASIQKFGGER(8)-LSQKFPK(4)	5	BSA(204)-BSA(221)	1.240706e-008
LSQKFPK(4)-SLGKVGTR(4)	3	BSA(221)-BSA(431)	3.305709e-010
LVTDLTKVHK(7)-LSQKFPK(4)	3	BSA(239)-BSA(221)	2.185797e-008
LVTDLTKVHK(7)-LSQKFPK(4)	4	BSA(242)-BSA(221)	2.243856e-004
NECFLSHKDDSPDLPKLPDPNTLCEDEFK(16)-SLGKVGTR(4)	4	BSA(114)-BSA(431)	1.442381e-007
NECFLSHKDDSPDLPKLPDPNTLCEDEFK(16)-SLGKVGTR(4)	4	BSA(114)-BSA(431)	2.542907e-006
QEPERNECFLSHKDDSPDLPK(13)-LKPDPNTLCEDEFK(2)	4	BSA(106)-BSA(116)	2.196512e-004
QEPERNECFLSHKDDSPDLPK(13)-LKPDPNTLCEDEFK(2)	5	BSA(106)-BSA(116)	1.230726e-001
SHCIAEVEKDAIPENLPLTADFAEDK(9)-CCTKPESER(4)	4	BSA(294)-BSA(439)	7.725918e-008
SHCIAEVEKDAIPENLPLTADFAEDK(9)-CCTKPESER(4)	4	BSA(294)-BSA(439)	1.587736e-006
VGTRCCTKPESER(8)-LSQKFPK(4)	3	BSA(439)-BSA(221)	2.022144e-004
VHKECCHGDLLECCADDR(3)-ALKAWSVAR(3)	5	BSA(242)-BSA(211)	6.130527e-008
VHKECCHGDLLECCADDRADLAK(3)-ALKAWSVAR(3)	4	BSA(242)-BSA(211)	3.661786e-010
VHKECCHGDLLECCADDRADLAK(3)-LSQKFPK(4)	3	BSA(242)-BSA(221)	5.918036e-006
VTKCTESLVNR(3)-LSQKFPK(4)	3	BSA(474)-BSA(221)	1.854033e-001
YICDNQDTSISLKL(12)-ECCDKP(5)	3	BSA(273)-BSA(280)	2.904847e-008
YICDNQDTSISLKL(12)-LSQKFPK(4)	3	BSA(273)-BSA(221)	3.793033e-009

b

Séquence 1	Séquence 2	Position 1	Position 2	Distance
LFTFHADICTLPDTEKQIK	QIKKQKQALVELLK	520	524	6,42
FGERALKAWSVAR	VHKECCHGDLLECCADDR	211	242	9,73
SLGKVGTR	KPKATEEQLK	431	537	11,36
CASIQKFGGERALK	LCVLHEKTPVSEK	204	465	13,24
CASIQKFGGERALK	FGERALKAWSVAR	204	211	13,47
SLGKVGTR	CCTKPESER	431	439	13,51
FGERALKAWSVAR	RLAKEYATELEECCK	211	350	13,54
YICDNQDTSISLKL	ECCDKP	273	280	13,82
QIKKQKQALVELLK	ATEEQLKTVMENFVAFVDK	524	544	14,33
FGERALKAWSVAR	LSQKFPK	211	221	15,45
SLGKVGTR	QIKKQKQALVELLK	431	524	17,37
SHCIAEVEKDAIPENLPLTADFAEDK	CCTKPESER	294	439	18,38
LSQKFPK	LVTDLTKVHK	221	239	19,35
KVPQVSTPTLVEVSR	SLGKVGTR	413	431	19,49
DTHKSEIAHR	FKDLGEEHF	1	12	19,54
LSQKFPK	ECCDKP	221	280	19,72
KDDSPDLPK	SLGKVGTR	114	431	19,9
LSQKFPK	CCTKPESER	221	439	20,54
KDDSPDLPKLPDPNTLCEDEFK	SLGKVGTR	116	431	21,27
NECFLSHKDDSPDLPK	KDDSPDLPKLPDPNTLCEDEFK	106	116	24,48
LSQKFPK	SLGKVGTR	221	431	26,43
CASIQKFGGERALK	LSQKFPK	204	221	26,44
CASIQKFGGERALK	SLGKVGTR	204	431	27,15
FGERALKAWSVAR	SLGKVGTR	211	431	29,18
FGERALKAWSVAR	CCTKPESER	211	439	29,65
CASIQKFGGERALK	CCTKPESER	204	439	33
LSQKFPK	KVPQVSTPTLVEVSR	221	413	37,09
DTHKSEIAHR	SLGKVGTR	1	431	54,47

c



**Figure 73 : Résultats XL-MS obtenues après analyse de la BSA. (a) Liste des peptides pontés identifiés par pLink 2.7. En bleu sont notés les peptides identifiés après validation manuelle et en blanc les peptides qui ont été rejetés après validation. (b) Détermination des distances entre les zones pontées à partir du modèle cristallographique obtenu sur PDB. (4F5S). (c) Implémentation des données XL-MS sur la structure cristallographique.**

Lors de cette étude, nous avons réussi à identifier 55 peptides pontés uniques à l'aide de pLink. Après validation manuelle, nous avons éliminé 18 peptides pontés. Ainsi, après validation manuelle, 33% des peptides identifiés par pLink n'ont pas été conservés dans notre tableau de résultats.

Afin de vérifier la pertinence de ce tableau de résultats, les données ont été implémentées sur la structure cristallographique de la BSA (PDB : 4F5S) afin de voir si les distances résultant des sites pontés validés étaient cohérentes (figure 73b). Après analyse, l'ensemble des distances mesurées entre les différents peptides pontés était compris entre 6.42 et 54.47Å dont 86% sont comprises entre 9 et 30 Å. Deux peptides pontés présentent des distances extrêmes : entre la lysine 520 et la lysine 524 (distance de 6.4 Å) et entre le N-terminal et la lysine 431 (54.5Å.). Ce problème de distance peut s'expliquer par le fait que les analyses XL-MS sont réalisées en solution où la structure de la protéine est en mouvement.

Lors de ce mouvement protéique, il n'est pas improbable que la réaction de pontage ait été réalisée au moment où les deux acides aminés étaient à une distance atteignable par l'agent pontant. Nos données sont en parfait accord avec des études antérieures de la littérature qui ont reporté que la réaction de pontage avec le BS3 se réalise pour des sites distants entre environ 8 et 35 Å.<sup>[16]</sup> Cette expérience nous prouve qu'avec nos conditions expérimentales, nous avons la possibilité d'obtenir des informations structurales cohérentes par rapport aux données structurales déjà bien connues sur la BSA.<sup>[17, 18]</sup> et que nous pouvons les appliquer sur notre système RAR/RXR.

## 2.5 Résultats validés sur RAR/RXR : Visualisation des peptides pontés sur XiView

Le tableau 7 présente la liste des peptides pontés identifiés par le logiciel pLink pour le complexe RAR/RXR puis validés manuellement en utilisant les critères de validation cités précédemment. Une grande majorité des cross-links proposés par pLink nous semble correctement attribuée. Néanmoins, environ 20% de propositions trop ambiguës selon nos critères de validation sont éliminés du tableau de résultats et ne seront donc pas considérés pour extraire les informations de contraintes géométriques tridimensionnelles et en déduire des données structurales fiables.

**Tableau 7 : Liste des peptides pontés intermoléculaires détectés par pLink. En bleu sont notés les peptides identifiés après validation manuelle. En blanc sont notés les peptides non validés après validation manuelle.**

Position des acides aminés pontés sur les protéines	Peptides	Score	Charge	Masse ion précurseur
RXR (381)-RAR (118)	AIVLFNPDSKGLSNPAEVEALR(10)-SIQKNMVYTCR(4)	2.691832e-007	5	4014.05
RXR(381)-RAR(390)	AIVLFNPDSKGLSNPAEVEALR(10)-SISAKGAER(5)☒	8.811572e-009☒	4	3395.82
RXR(381)-RAR(390)	AIVLFNPDSKGLSNPAEVEALREK(10)-SISAKGAER(5)☒	1.698535e-009☒	5	3652.95
RAR(155)-RXR(160)	CFDVGMSKESVR(8)-GFFKR(4)☒	7.367377e-005	4	2206.07
RAR(155)-RXR(165)	CFDVGMSKESVR(8)-KDLTYTCR(1)☒	1.081178e-010☒	4	2608.22
RAR(128)-RXR(165)	DKNCIINKVTR(2)-KDLTYTCR(1)☒	2.942936e-006☒	4	2554.31
RAR(134)-RXR(175)	DKNCIINKVTR(8)-DNKDCLIDKR(3)☒	3.206857e-014☒	5	2774.42
RAR(134)-RXR(165)	DKNCIINKVTR(8)-KDLTYTCR(1)☒	4.136504e-010☒	4	2554.31
RXR (364)-RAR (390)	DMQMDKTELGLCLR(6)-SISAKGAER(5)	2.475634e-003	3	2652.27
RXR (181)-RAR (390)	DNKDCLIDKR(9)-SISAKGAER(5)	9.459326e-006	3	2332.19
RAR(171)-RXR(407)	EAPKPECSesyTLTPEVGELIEK(4)-HKYPEQPGR(2)☒	1.185963e-001☒	4	3854.87
RAR(171)-RXR(165)	EAPKPECSesyTLTPEVGELIEK(4)-KDLTYTCR(1)☒	4.282860e-012☒	4	3799.82
RAR(118)-RXR(165)	GFFRRSIQKNMVYTCRDRDK(9)-RTVRKDLTYTCRDNK(5)☒	7.497403e-002☒	5	4506.30
RXR(213)-RAR(118)	GKDRNENEVESTSSANEDMPVEK(2)-SIQKNMVYTCR(4)☒	1.147724e-008☒	5	4238.93
RXR (165)-RAR (390)	KDLTYTCR(1)-SISAKGAER(5)	1.494738e-010	3	2112.07
RAR(118)-RXR(201)	SIQKNMVYTCR(4)-CLAMGMKR(7)☒	8.342400e-010☒	4	2640.26
RAR (118)-RXR (181)	SIQKNMVYTCR(4)-DNKDCLIDKR(9)	5.214884e-003	4	2950.43
RAR(118)-RXR(160)	SIQKNMVYTCR(4)-GFFKR(4)☒	1.178363e-003☒	4	2328.17
RAR (118)-RXR (165)	SIQKNMVYTCR(4)-KDLTYTCR(1)	2.858785e-005	5	2730.31
RAR(390)-RXR(417)	SISAKGAER(5)-FAKLLLR(3)☒	3.302204e-004☒	4	1916.13

Afin de pouvoir établir des cartes de proximités à partir de ces résultats et pouvoir mieux visualiser les cross-links dans l'espace, des outils de visualisation ont été développés.<sup>[19-21]</sup> Malheureusement, la majorité d'entre eux requiert une étape préliminaire de retraitement de données. En effet, chaque logiciel d'identification nous fournit des résultats sous des formats différents (.html, .csv, .zhm) qu'il faut par la suite retraiter manuellement afin d'obtenir un fichier

adapté aux paramètres entrants exigés par les outils de visualisation. Ce retraitement de données reste long et délicat et est très chronophage.

XiView est un logiciel de visualisation des peptides pontés développé par l'équipe de J. Rappsilber ([https://xiview.org/xiNET\\_website/index.php](https://xiview.org/xiNET_website/index.php)). Il permet d'observer les sites pontés identifiés à partir des données MS/MS sur des cartes 2D de type circulaire ou linéaire mais également de confronter nos résultats aux données extraites de structures 3D obtenues à haute résolution par DRX ou issues de modélisation moléculaire.



Figure 74 : Image représentant les différentes fonctions de XiView (image extrait du site).

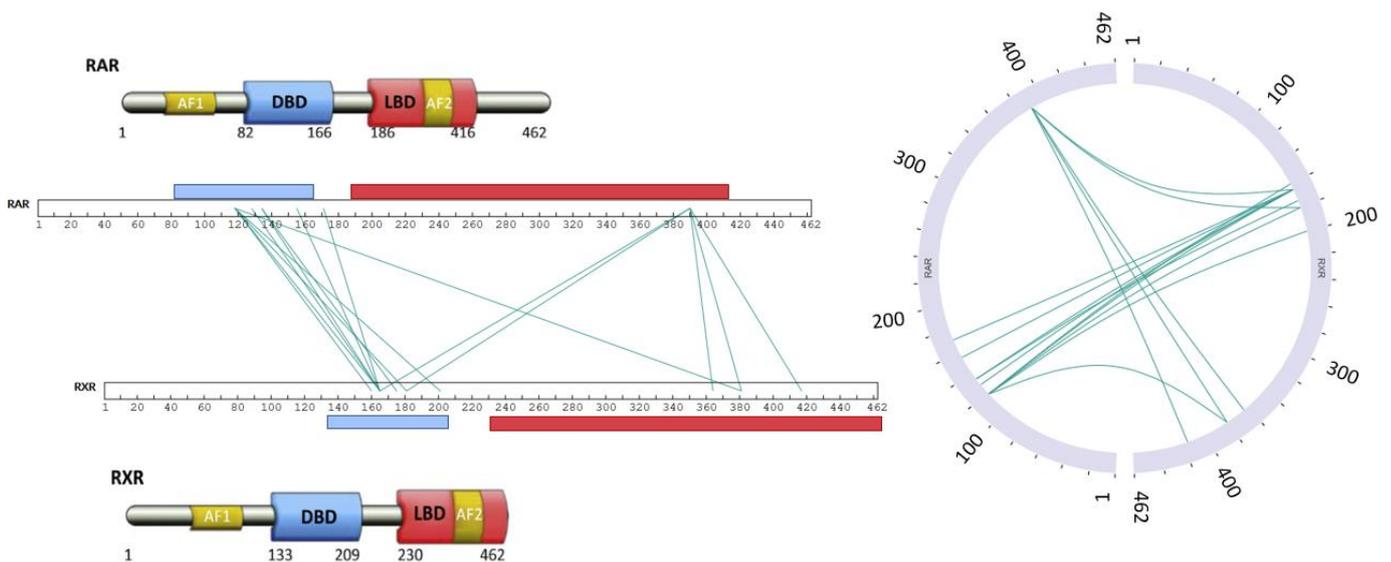
Pour utiliser XiView, il faut au préalable convertir nos données de peptides pontés en un fichier csv (Figure 75) où sont répertoriées les séquences pontées, la position du pontage, ainsi que les scores d'identification. La conversion des données en format .csv reste tout de même manuelle en fonction des logiciels d'interprétation utilisés. Ensuite, le logiciel est très convivial. Nous avons rapidement accès aux représentations linéaires ou circulaires des zones protéiques cross-linkées et un curseur de visualisation dépendant du score d'identification s'avère très utile pour repérer les cross-links à fort taux de confiance. Un accès direct au spectre de fragmentation interprété pour chaque cross-link est également possible ce qui nous permet d'être plus intuitifs sur la validation que la seule valeur du score.

PeakListFileName	ScanId	Protein1	Protein2	PepSeq1	PepSeq2	LinkPos1	LinkPos2	Charge	Crosslinker ModMass	Score
data.mgf	5203	Eaf7	Eaf7	SRTPVNGIGYR	KRIDNSDESDRSD R	7	1	5	138.068	16.532
data.mgf	6547	Eaf7	Eaf7	SRTPVNGIGYR	LKDQDDVDGSSR	7	2	3	138.068	15.967

Figure 75 : Exemple de fichiers .csv utilisé comme format d'entrée pour utiliser XiView.

Une 1<sup>ère</sup> carte a ainsi été établie pour le complexe RAR/RXR ponté au BS3 à partir des données pLink après validation manuelle (tableau 7). Chaque trait représente un peptide ponté unique. La redondance n'est donc pas considérée dans cette visualisation. Cette carte nous montre que les cross-links ne sont pas repartis de façon homogène sur les séquences protéiques bien que des lysines et sérines soient réparties régulièrement sur l'ensemble des séquences primaires des deux partenaires. Au contraire il y a des « nœuds » de pontage comme par exemple entre la zone RAR118 et RXR165 ou encore RAR390 et RXR381. Une interaction stable a notamment été observée entre l'acide aminé 381 de RXR et l'acide aminé 390 de RAR où de nombreux spectres MS/MS (en moyenne une vingtaine de spectres) associés à ce pontage ont été identifiés.

Ceci est en accord avec une structure native homogène de l'hétérodimère RAR/RXR plutôt qu'avec un complexe partiellement dénaturé pour lequel des lysines (ou sérine/thréonine) se rapprocheraient de manière aléatoire générant des cross-links sur l'ensemble des deux partenaires. Les pontages correspondant à ces nœuds reflètent la proximité entre les domaines DBD de RAR et RXR d'une part et les domaines LBD d'autre part ce qui est en accord avec les connaissances structurales des RNs.



**Figure 76 : Première carte de proximité obtenue sur le complexe RAR-RXR**

Cette expérience de pontage de RAR/RXR par le BS3 a été reproduite plusieurs fois. Il est à noter que la reproductibilité technique (c'est-à-dire l'analyse LC-MS/MS répéter 3 fois à partir du même échantillon biologique ponté et protéolysé) s'est avérée très bonne puisqu'elle fut de 78%. La reproductibilité biologique (c'est-à-dire l'analyse effectuée à partir de complexe biologique issu de productions protéiques différentes) est plus délicate à évaluer car il faut distinguer la reproductibilité au niveau peptidique et celle au niveau zone de pontage.

Ce point sera discuté plus en détail ultérieurement (partie II). Au niveau zone de pontage, la reproductibilité est excellente puisque nous observons systématiquement des peptides cross-linkés entre les mêmes régions protéiques :

- RAR-DBD (118-134) et RXR-BDB (165-181)
- RAR-DBD ((118-134) et RXR- LBD (380-381)
- RAR-LBD (386-390) et RXR-DBD (181)
- RAR-LBD (386-390) et RXR-LBD (364-381)

Cependant, nous avons été surpris du nombre très modeste de peptides cross-linkés uniques qui ont été identifiés par pLink et à partir de ces données, trop peu d'informations de structures pourront être déduites. C'est pourquoi nous avons décidé de réaliser un travail d'optimisation à la fois au niveau expérimental (conditions de la réaction de pontage et de génération des peptides pontés) mais aussi au niveau interprétation des données (choix du logiciel d'analyse et critère de validation des identifications) afin de vérifier si davantage d'informations structurales pouvaient être déduites.

## **2.6 Conclusions**

Ainsi, une 1<sup>ère</sup> expérience de pontage du complexe RAR/RXR par le BS3 nous a apporté plusieurs informations :

- Une analyse par MALDI-MS permet rapidement de vérifier que la réaction de pontage du complexe hétérodimérique RAR/RXR par le BS3 s'effectue avec un bon rendement de pontage et qu'aucune autre stœchiométrie reflétant des cross-links non spécifiques n'est observée. Ceci nous indique que la chimie réactionnelle de pontage utilisée ici est prometteuse et que des sites se trouvent à proximité des interfaces de dimérisation.
- Des peptides pontés ont été identifiés avec suffisamment d'indices de confiance par un logiciel spécifiquement dédié à cette application pLink.
- L'ensemble des peptides identifiés par pLink sont notés selon un score. Néanmoins, ce score n'est pas suffisant pour valider l'interprétation et une étape de validation manuelle par l'expérimentateur est nécessaire pour éliminer les propositions de peptides pontés trop ambiguës ou fausses.
- Le nombre de peptides pontés est relativement modeste et ne permet pas de sonder la structure de l'hétérodimère avec précision. C'est pourquoi nous avons entamé une optimisation du protocole dans le but d'augmenter le nombre de cross-links intermoléculaires validés.

## **3. Enrichissement en données de peptides pontés : sonder de façon plus précise la structure**

L'objectif était double :

- Identifier en plus grand nombre et avec une meilleure confiance les peptides pontés.
- Croiser les informations structurales obtenues à l'aide d'autres agents pontants de chimie et de longueur de bras espaceur différent.

Pour cela, nous sommes intervenus à plusieurs étapes du protocole expérimental ainsi qu'au niveau de l'étape d'interprétation.

### **3.1 Optimisation des paramètres réactionnels**

#### **3.1.1 Le milieu réactionnel :**

Des études préliminaires au laboratoire (stage master M2 de Steve Hessmann) ont montré que le temps de la réaction de pontage (30min, 1h, 2h30) ainsi que le ratio protéine/BS3 (1/25, 1/50, 1/250) influent peu sur le rendement de pontage tel que suivi par MALDI-MS. Par manque de temps d'accessibilité à l'orbitrap, nous n'avons pas pu comparer chacune de ces conditions du point de vue du nombre de peptides pontés, mais aucune amélioration n'a été observée en laissant la réaction se poursuivre de 30min à 1h avec un ratio protéine/BS3 : 1/50.

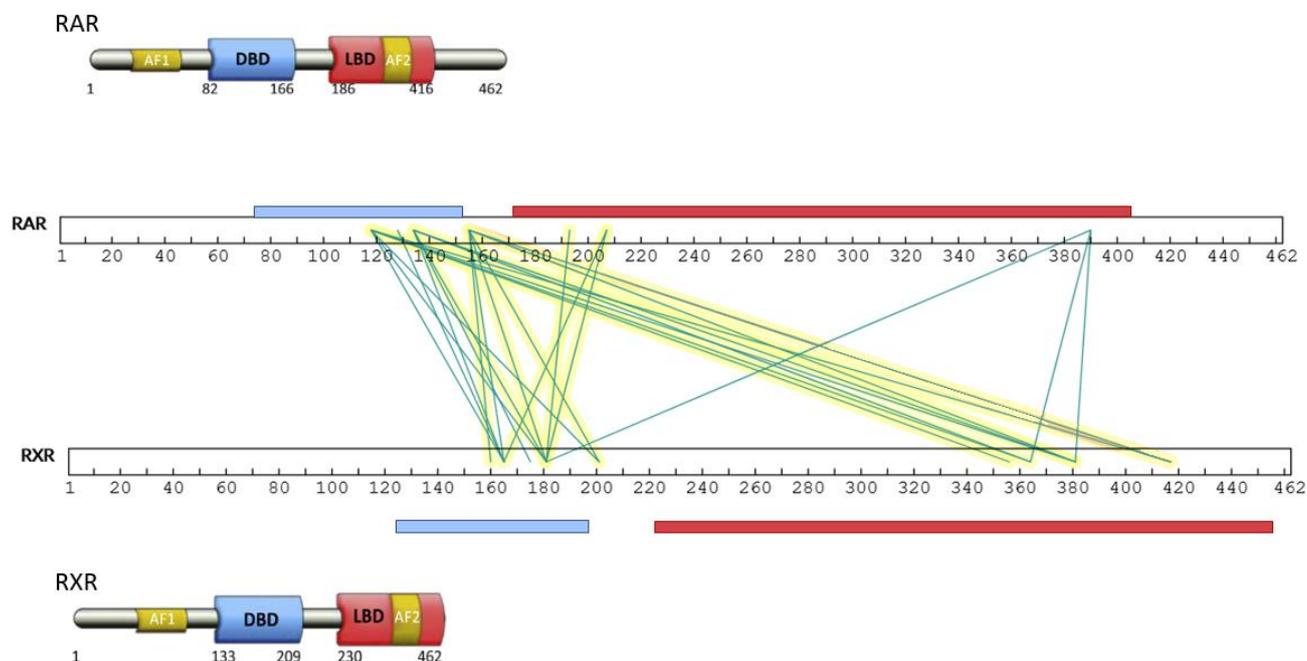
#### **3.1.2 Enrichissement des peptides pontés par chromatographie d'échange d'ions (SCX)**

Un bon rendement de pontage tel que celui que nous observons par MALDI avec le BS3 n'est pas un gage que la concentration de chaque peptide ponté individuellement est élevée. En effet, les peptides pontés uniques (issus d'une haute multiplicité de sites rendant les produits réticulés très hétérogènes) sont largement minoritaires par rapport aux peptides linéaires issus de la digestion tryptique des deux partenaires<sup>[22-24]</sup>. Ainsi, enrichir notre milieu en peptides cross-linkés permettra d'améliorer leur intensité en MS et donc d'une part d'augmenter les chances de sélection des ions précurseurs cross-linkés (davantage de peptides cross-linkés détectés) et d'autre part d'améliorer la qualité des spectres de fragmentation résultants (davantage de confiance en l'interprétation).

Nous avons décrits précédemment (Cf chapitre III de la partie bibliographie) que deux procédés peuvent être implémentés dans le protocole afin de partitionner notre mélange peptidique en séparant les peptides linéaires (globalement plus courts et peu chargés) des peptides cross-linkés (globalement plus lourds et plus chargés) : soit par utilisation de la chromatographie liquide d'exclusion stérique (SEC)<sup>[25]</sup>, soit par chromatographie échangeuse de cation (SCX).<sup>[26, 27]</sup> Alors que la SEC sépare les peptides selon leur taille et leur forme, la SCX sépare les ions selon leur charge.<sup>[25]</sup>

D'après les publications, nous n'avons pas décelé de différences d'efficacité reportées pour ces deux stratégies d'enrichissement, c'est pourquoi nous avons décidé d'utiliser un système de purification basé sur la SCX. Pour cela, nous avons utilisé des cartouches commerciales remplies de phase stationnaire contenant une phase greffée avec un revêtement d'acide benzène sulfonique (ThermoFisher) qui permettent de travailler rapidement à l'échelle de nos quantités d'échantillons et nous avons effectué une étude comparative du même échantillon RAR/RXR ponté par le BS3 avant et après enrichissement sur cartouche SCX (Cf partie expérimentale). L'identification des peptides pontés a été faite avec l'aide du logiciel pLink et une validation manuelle a été systématiquement réalisée. Alors que 16 peptides intermoléculaires sont identifiés avant enrichissement, nous obtenons maintenant 27 peptides intermoléculaires (en jaune sur le tableau 8) pour l'échantillon enrichi après SCX. Sur les 16 peptides observés sans enrichissement, quatre (RXR165-RAR171, RXR165-RAR390, RXR160-RAR118, RXR417-RAR390) n'ont pas été retrouvés après enrichissement). Outre une augmentation de 63% du nombre de peptides cross-linkés identifiés, on observe également une augmentation des scores d'identification dans la majorité

des cas, avec parfois des améliorations très significatives (par exemple un score passant de  $10^{-5}$  à  $10^{-11}$  pour le peptide RXR165-RAR118 ou encore de  $10^{-3}$  à  $10^{-9}$  pour le peptide RXR181-RAR118).



**Figure 77 : Carte de proximité obtenue sur le complexe RAR-RXR après SCX. En jaune sont représentés les peptides pontés uniquement identifiés après SCX. La proximité entre RAR-DBD et RXR-LBD est maintenant illustrée par de nombreux peptides cross-linkés**

Cette tendance à obtenir davantage de peptides pontés et avec un meilleur score d'identification sera d'autant plus significative que le système biologique d'intérêt est gros et complexe. En effet, plus un complexe sera lourd, plus le nombre de sites potentiellement pontés sera élevé et donc, plus le mélange peptidique risque d'être hétérogène. Dans ce cas, un enrichissement en peptides pontés sera crucial pour obtenir une liste fournie en données de réticulation. Une observation similaire sera également illustrée lors de l'analyse d'un complexe formé de 12 sous-unités : NuA4 (Cf chapitre VII).

Toutefois, il est à noter qu'entre nos mains, le protocole d'enrichissement par cartouches SCX ne s'est pas avéré parfaitement reproductible. En effet, plusieurs préparations se sont avérées infructueuses, voire totalement stériles, avec des résultats d'identification en peptides cross-linkés à la fois pauvres en nombre et en qualité. Ceci était peut-être dû à une mauvaise homogénéité des cartouches de phases ou encore à une perte trop importante de matériel lors des étapes de rinçage et de dessalage des cartouches (Annexe 6). C'est pourquoi il serait intéressant de configurer un système automatisé à l'aide d'une chromatographie liquide équipée d'une colonne SCX pour optimiser un fractionnement reproductible et adapté au système biologique considéré. D'autre part, si les quantités de matériel biologique le permettent, nous aurions une visualisation par détection UV qui nous fournirait un indice de qualité du digest peptidique ponté en amont de l'analyse LC-MS/MS.

Protéines	Peptides	Score	Charge	Masse ion Précurseur
RAR(207)-RXR(181)	AHQETFPALCQLGKYTTNNSSEQR(14)-DNKDCLIDKR(9)	2.099777e-001	4	4478.18
RAR(207)-RXR(165)	AHQETFPALCQLGKYTTNNSSEQR(14)-KDLTYTCR(1)	3.316451e-007	4	3973.88
RXR(381)-RAR(155)	AIVLFNPD SKGLSNPAEVEALR(10)-CFDVGMSKESVR(8)	7.243272e-005	4	3891.95
RXR(381)-RAR(134)	AIVLFNPD SKGLSNPAEVEALR(10)-DKNCIINKVTR(8)	4.735223e-003	4	3838.05
RXR(381)-RAR(118)	AIVLFNPD SKGLSNPAEVEALR(10)-SIQKNM VYTCHR(4)	2.439226e-008	4	4014.05
RXR(381)-RAR(390)	AIVLFNPD SKGLSNPAEVEALR(10)-SISAKGAER(5)	5.555037e-006	3	3395.81
RXR(381)-RAR(390)	AIVLFNPD SKGLSNPAEVEALREK(10)-SISAKGAER(5)	1.698535e-009	5	3652.95
RAR(155)-RXR(201)	CFDVGMSKESVR(8)-CLAMGMKR(7)	4.028464e-005	3	2518.17
RAR(155)-RXR(181)	CFDVGMSKESVR(8)-DNKDCLIDKR(9)	1.453987e-006	4	2828.33
RAR(155)-RXR(417)	CFDVGMSKESVR(8)-FAKLLLR(3)	3.820523e-005	4	2412.28
RAR(155)-RXR(160)	CFDVGMSKESVR(8)-GFFKR(4)	1.535547e-004	4	2206.07
RAR(155)-RAR(390)	CFDVGMSKESVR(8)-SISAKGAER(5)	1.228197e-004	4	2470.20
RAR(155)-RXR(165)	CFDVGMSKESVR(8)-KDLTYTCR(1)	7.618282e-010	3	2608.22
RAR(134)-RXR(175)	DKNCIINKVTR(8)-DNKDCLIDKR(3)	4.088102e-008	4	2774.42
RAR(134)-RXR(181)	DKNCIINKVTR(8)-DNKDCLIDKR(9)	5.070782e-007	5	2774.42
RAR(134)-RXR(165)	DKNCIINKVTR(8)-KDLTYTCR(1)	7.105975e-011	3	2554.31
RXR(364)-RAR(134)	DMQMDKTELGCLR(6)-DKNCIINKVTR(8)	5.521824e-003	4	3094.51
RXR(364)-RAR(118)	DMQMDKTELGCLR(6)-SIQKNM VYTCHR(4)	6.142095e-009	3	3270.51
RXR(364)-RAR(390)	DMQMDKTELGCLR(6)-SISAKGAER(5)	5.669952e-008	3	2652.27
RXR(181)-RAR(390)	DNKDCLIDKR(9)-SISAKGAER(5)	1.632775e-008	3	2332.19
RAR(193)-RXR(181)	KAHQETFPALCQLGK(1)-DNKDCLIDKR(9)	3.282042e-009	4	3141.58
RXR(165)-RAR(128)	KDLTYTCR(1)-DKNCIINK(2)	5.757860e-009	4	2198.09
RAR(118)-RXR(201)	SIQKNM VYTCHR(4)-CLAMGMKR(7)	6.445106e-012	3	2640.26
RAR(118)-RXR(181)	SIQKNM VYTCHR(4)-DNKDCLIDKR(9)	1.650149e-009	3	2950.43
RAR(118)-RXR(417)	SIQKNM VYTCHR(4)-FAKLLLR(3)	2.362235e-006	5	2534.37
RAR(118)-RXR(165)	SIQKNM VYTCHR(4)-KDLTYTCR(1)	4.050888e-011	3	2730.31
RAR(118)-RXR(356)	SIQKNM VYTCHR(4)-VLTELVSKMR(8)	2.818888e-003	4	2849.49

RAR/RXR + BS3  
Après Enrichissement SCX  
(27 peptides)

Protéines	Peptides	Score	Charge	Masse ion précurseur
RXR(381)-RAR(118)	AIVLFNPD SKGLSNPAEVEALR(10)-SIQKNM VYTCHR(4)	2.143218e-006	5	4014.05
RXR(381)-RAR(390)	AIVLFNPD SKGLSNPAEVEALR(10)-SISAKGAER(5)	8.811572e-009	4	3395.82
RXR(381)-RAR(390)	AIVLFNPD SKGLSNPAEVEALREK(10)-SISAKGAER(5)	1.698535e-009	5	3652.95
RAR(155)-RXR(165)	CFDVGMSKESVR(8)-KDLTYTCR(1)	1.081178e-010	4	2608.22
RAR(128)-RXR(165)	DKNCIINKVTR(2)-KDLTYTCR(1)	2.942936e-006	4	2554.31
RAR(134)-RXR(175)	DKNCIINKVTR(8)-DNKDCLIDKR(3)	3.206857e-014	5	2774.42
RAR(134)-RXR(165)	DKNCIINKVTR(8)-KDLTYTCR(1)	4.136504e-010	4	2554.31
RXR(364)-RAR(390)	DMQMDKTELGCLR(6)-SISAKGAER(5)	2.475634e-003	3	2652.27
RXR(181)-RAR(390)	DNKDCLIDKR(9)-SISAKGAER(5)	9.459326e-006	3	2332.19
RAR(171)-RXR(165)	EAPKPECSSESYLTPEVGLIEK(4)-KDLTYTCR(1)	4.282860e-012	4	3799.82
RXR(165)-RAR(390)	KDLTYTCR(1)-SISAKGAER(5)	1.494738e-010	3	2112.07
RAR(118)-RXR(201)	SIQKNM VYTCHR(4)-CLAMGMKR(7)	8.342400e-010	4	2640.26
RAR(118)-RXR(181)	SIQKNM VYTCHR(4)-DNKDCLIDKR(9)	5.214884e-003	4	2950.43
RAR(118)-RXR(160)	SIQKNM VYTCHR(4)-GFFKR(4)	1.178363e-003	4	2328.17
RAR(118)-RXR(165)	SIQKNM VYTCHR(4)-KDLTYTCR(1)	2.858785e-005	5	2730.31
RAR(390)-RXR(417)	SISAKGAER(5)-FAKLLLR(3)	3.302204e-004	4	1916.13

RAR/RXR + BS3  
Avant Enrichissement SCX  
(16 peptides)

**Tableau 8 : Tableau de comparaison des peptides pontés identifié avant et après enrichissement par SCX. Les peptides en jaune représentent les peptides pontés nouvellement identifiés après enrichissement par cartouche SCX.**

### 3.2 Optimisation des paramètres d'analyse LC-MS/MS

Dans la mesure où notre système biologique est peu complexe (formé de deux partenaires), nous nous sommes focalisés davantage sur les paramètres d'acquisition par MS que sur les paramètres de séparation chromatographique nanoLC. Nous n'avons malheureusement pas pu réaliser une étude exhaustive de l'influence des paramètres d'acquisition en MS et MS/MS par manque de temps d'accessibilité à l'instrument. Néanmoins, il nous a paru important de vérifier que notre méthode d'acquisition était robuste, c'est pourquoi nous avons comparé notre liste de peptides cross-linkés précédemment obtenue à celles obtenues après modification du temps maximal d'injection en MS/MS (200 et 500 ms), de l'énergie de collision (valeur discrète à 30eV ou 3 paliers à 25, 30, 35eV) et de la charge du précurseur sélectionné (>2+ ou >3+). Le tableau 9 répertorie les paramètres d'acquisition MS

et montre que la méthode nommée « Min3+ enhanced » a permis d'obtenir le plus grand nombre de peptides cross-linkés intermoléculaires. De façon inattendue, augmenter le temps maximal d'injection à 500 ms a conduit à une chute radicale du nombre de spectres MS/MS dont l'interprétation a abouti à l'identification d'un peptide cross-linké.

**Tableau 9 : Comparaison des méthodes d'acquisition MS appliquées à un échantillon cross-linké de RAR/RXR par du BS3.**

Method	Classical	Min3+500ms	Min3+ enhanced stepped	Min 4+	Min 4+ 500ms
<b>General</b>					
Lock mass injections	10 s	10 s	10 s	10 s	10 s
method duration	160 min	160 min	160 min	160 min	160 min
<b>FullMS</b>					
Resolution	70,000	70,000	70,000	70,000	70,000
AGC target	3,00E+06	3,00E+06	3,00E+06	3,00E+06	3,00E+06
Maximum IT	150 ms	150 ms	150 ms	150 ms	150 ms
<b>MSMS</b>					
Resolution	17,500	17,500	17,500	17,500	17,500
AGC target	5,00E+06	5,00E+06	5,00E+06	5,00E+06	5,00E+06
Maximum IT	200 ms	500 ms	200 ms	200 ms	500ms
TopN	8	8	8	8	8
Scan range	200 - 2000 m/z	200 - 2000 m/z	200 - 2000 m/z	200 - 2000 m/z	200 - 2000 m/z
Collision Energy (N)CE	30 eV	30 eV	25,30,35 eV	30 eV	30 eV
Intensity threshold	5,00E+04	5,00E+04	5,00E+04	5,00E+04	5,00E+04
Charge Exclusion	unassigned, 1, 2	unassigned, 1, 2	unassigned, 1, 2	unassigned, 1, 2,3	unassigned, 1, 2,3
<b>Nombre de peptides pontés Détectés via pLink</b>	<b>17</b>	<b>4</b>	<b>23</b>	<b>7</b>	<b>5</b>

La méthode d'acquisition « Min3+ enhanced » fut choisie pour la suite des expériences mettant en jeu l'analyse de complexes cross-linkés.

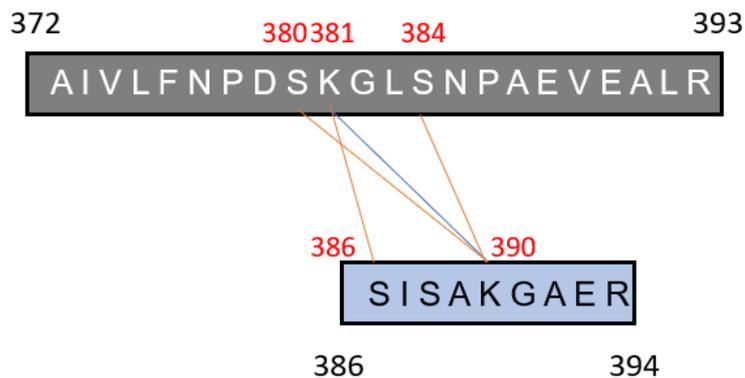
### **3.3 Optimisation du choix de l'agent pontant**

#### **3.3.1 Les esters NHS (comme BS3) ne se limitent pas aux pontages entre deux lysines mais peuvent également impliquer des sérine, thréonine et tyrosine.**

Les agents pontants de la famille des NHS dont font partie le BS3 sont connus dans la littérature pour réaliser des liaisons spécifiques sur les amines primaires des protéines telles que sur les lysines ou sur l'amine N-terminale de la séquence. Toutefois, il a été démontré en 2008 par Andrea Sinz que la liaison des NHS était également réalisable sur les fonctions hydroxy des sérines, thréonines ou tyrosines.<sup>[28]</sup> La spécificité de la réaction est guidée par le pH du milieu réactionnel : en condition légèrement acide (pH6) le groupement hydroxy des tyrosines réagit préférentiellement alors qu'à pH plus alcalin (pH8) ce sont les amines des lysines qui seront attaquées en premier.<sup>[28-30]</sup> Par exemple, une étude de pontage sur des laminines a montré qu'environ 76.3% de la réaction de pontage était réalisée avec les amines des lysines mais également que 12.5% impliquaient des sérines, 4.3% des tyrosines, 3.9% sur le N-terminal et 3% sur les thréonines.<sup>[28]</sup> Considérer cette hypothèse dans nos requêtes d'identification de peptides pontés permettra de mettre en évidence de nouveaux sites de pontage ou bien de confirmer des zones de proximité déjà suggérées à partir de pontage lysine-lysine.

Le complexe RAR-RXR est constitué de 55 lysines, de 90 sérines, de 56 thréonines et de 21 tyrosines. En limitant nos recherches sur les lysines, nous risquons donc de perdre énormément d'informations. C'est pourquoi, pour la suite des études, sérines, thréonines et tyrosines furent systématiquement considérées comme sites de pontage potentiels.

En tenant compte des sérines, thréonines et tyrosines comme sites de pontage potentiels avec le BS3, nous avons pu identifier 29 peptides intermoléculaires au lieu des 16 précédemment obtenus si la recherche se limite aux pontages lysine/lysine. Dans notre cas, tous les peptides supplémentaires ont été interprétés comme résultant d'un pontage dissymétrique impliquant une lysine et une sérine (figure 78). Bien que deux sérines proches soient disponibles (RAR386 et RXR380 par exemple), nous n'avons pas détecté de peptides pontés de type sérine/sérine.



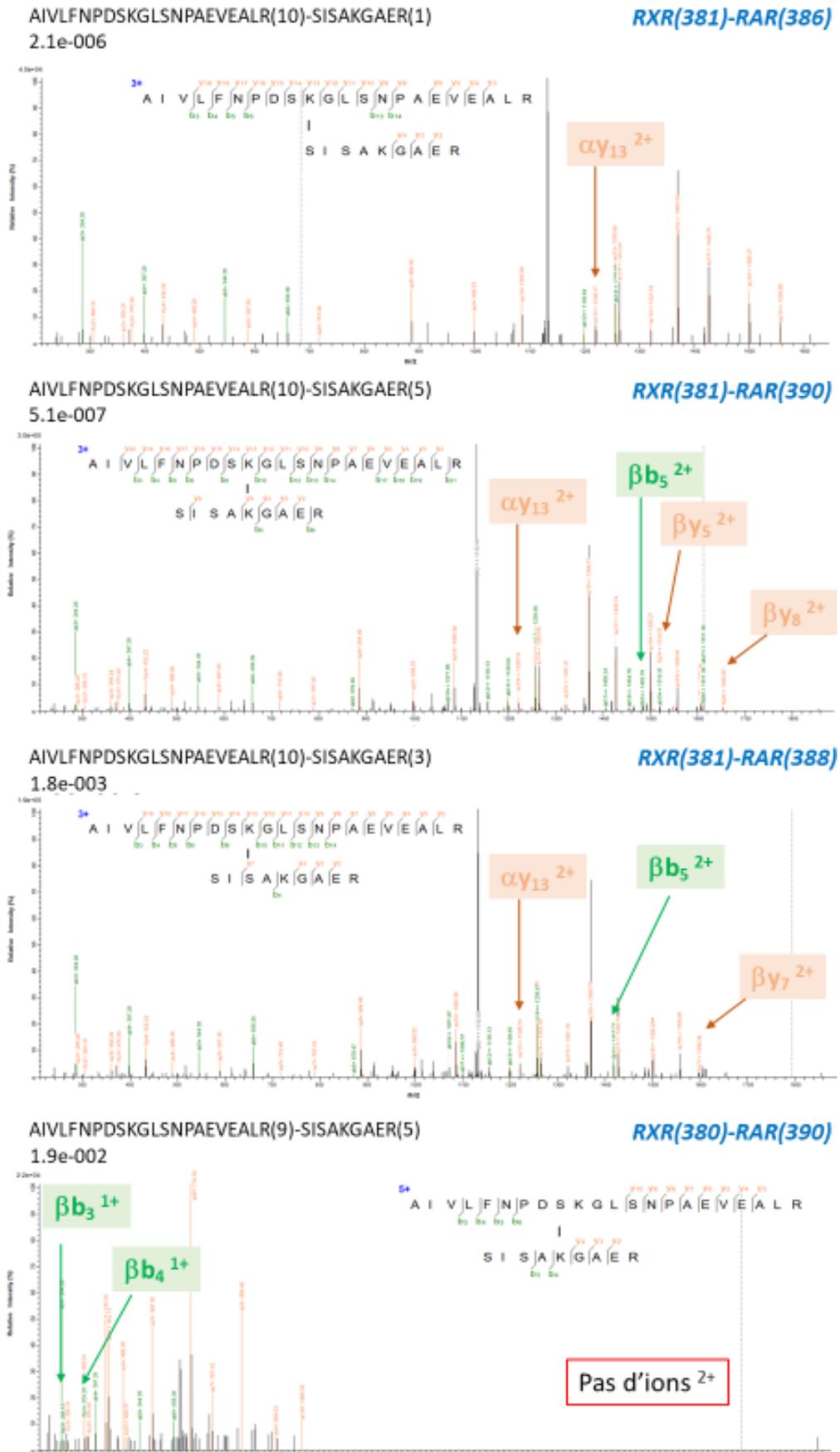
**Figure 78 : Schéma montrant la complémentarité des sites de pontages détectés sur un même peptide ponté. En rouge les peptides pontés entre lysine et sérine et en bleu, le peptide ponté identifié à partir des deux lysines.**

### Comment distinguer les pontages K-S et K-K ?

Si l'identification des séquences cross-linkées dans ce cas n'est en général pas remis en cause, la détermination du site de pontage (au niveau de l'acide aminé) peut être délicate lorsque plusieurs sites réactifs potentiels se situent sur le même peptide tryptique. C'est le cas notamment du peptide cross-linké de la figure 79 (RXR<sub>372-393</sub> – RAR<sub>386-394</sub>). 3 sites de pontage potentiels se situent sur RXR (S380, K381 et S384) et 3 sites de pontage potentiels se situent sur RAR (S386, S388 et K390) ce qui représente 9 propositions de connexion potentielles.

La figure 79 nous montre comment la recherche d'ions fragments discriminants nous aidera à trancher parmi l'une ou l'autre des propositions de pontage. En effet, l'observation d'un ion fragment de type  $\alpha\gamma 13^{2+}$  ( $m/z = 1220.17$ ) contenant le peptide  $\beta$  cross-linké permet d'assigner la K381 de RXR comme site de pontage préférentiel au détriment de la S380. De la même façon, l'ion fragment  $\beta\gamma 8^{2+}$  ( $m/z = 1654.87$ ) et l'ion  $\beta\gamma 5^{2+}$  ( $m/z = 1519.31$ ) démontre le pontage sur la K390 de RAR. Sur le spectre du bas, on observe qu'il n'y a aucun ion fragments doublement chargé dans la région haute du spectre ( $1200 < m/z < 1600$ ). Ces ions doublement chargés correspondent souvent à des ions fragments cross-linkés, c'est-à-dire contenant encore une fraction des deux séquences pontées. Ici, les ions fragments  $\beta\beta 3^{1+}$  ( $m/z = 288.15$ ) et  $\beta\beta 4^{1+}$  ( $m/z = 359.26$ ) suggèrent que le pontage a eu lieu sur la K390 de RAR. En revanche, aucun ion discriminant ne permet de trancher entre la S380 et la K381 de RXR.

Il est à noter que certains de ces ions fragments cross-linkés et doublement chargés sont d'intensités très faibles et que ces interprétations ne doivent être considérées que comme des propositions qu'il sera nécessaire de confirmer par reproductibilité ou encore par l'utilisation d'une enzyme de digestion possédant une spécificité d'hydrolyse intéressante si la séquence protéique le permet.



**Figure 79 : Spectres de fragmentation illustrant différents sites de pontage possibles à partir des mêmes séquences  $\alpha$  et  $\beta$  trypsiniques connectées.**

Ces peptides cross-linkés nouvellement identifiés en prenant en compte un pontage potentiel sur les serines/thréonines/tyrosines n'ont pas abouti à la mise en évidence d'une nouvelle zone de proximité sur RAR/RXR. En revanche, ils nous ont permis de confirmer une zone de pontage déjà observée mais faiblement représentée précédemment par un pontage entre lysines, ce qui rend notre carte de proximité ainsi déduite plus fiable.

### 3.3.2 Modifier l'agent pontant

Un bon rendement de pontage tel que celui que nous observons par MALDI avec le BS3 n'est pas un gage que le nombre de peptides intermoléculaires est élevé.

Afin de disposer de données de pontage plus riches, nous avons exploré l'avantage d'une modification de l'agent de pontage (i) du point de vue de sa chimie réactionnelle (s'il n'y a pas de lysines proches dans les zones d'interfaces intermoléculaires), (ii) mais aussi du point de vue de la longueur du bras espaceur (les lysines situées au niveau de l'interface intermoléculaire peuvent être proches ou plus éloignées les unes des autres).

Néanmoins, il n'existe pas de règles permettant de choisir le réactif optimal en fonction du système biologique d'intérêt (selon sa complexité ou sa taille par exemple). C'est pourquoi le choix du cross-linker s'avère parfois délicat, notamment lorsque ces études sont menées dans un but d'assistance à la modélisation structurale tridimensionnelle. En effet, l'objectif ici n'est pas de déterminer qui interagit avec qui au niveau protéique (c'est-à-dire quel partenaire interagit avec quelle sous-unité), mais plutôt de sonder des petites zones d'interface protéique et donc d'être en mesure de déterminer quels acides aminés sont proches les uns des autres.

Nous cherchons donc ici à enrichir la carte d'acides aminés connectés par le BS3 obtenue précédemment en accédant à quelques zones spatiales supplémentaires qui possèdent peu de lysines, ou bien des lysines éloignées ou encore orientées de façon inadéquate pour une réaction efficace par le BS3. Pour cela, nous avons choisi :

- un agent pontant qui a la même chimie réactionnelle que le BS3 mais avec une longueur de bras espaceur différent : BS(PEG)5 (homobifonctionnel ; 21.7 Å).<sup>[31]</sup>

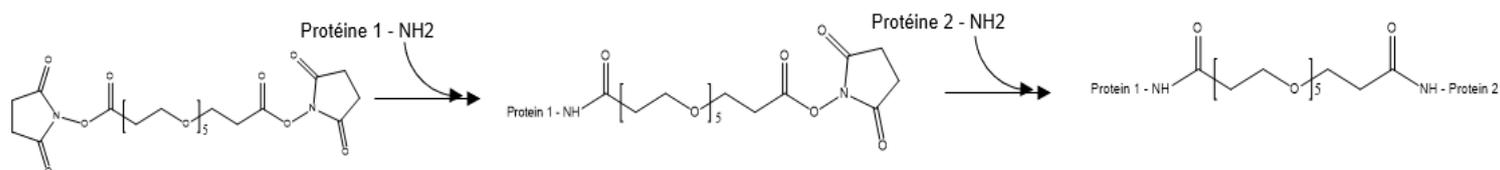


Figure 80 : Mécanisme réactionnel du pontage des protéines avec le BS(PEG)<sub>5</sub>.

- un agent pontant possédant une chimie réactionnelle différente de celle du BS3: LC-SMCC (hétérobifonctionnel Lysine/cystéine ; 16.2 Å)<sup>[32]</sup> RAR et RXR possèdent 18 et 13 cystéines respectivement

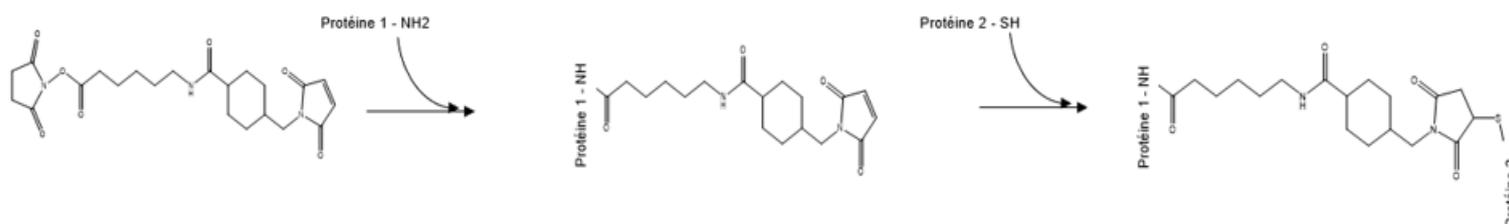
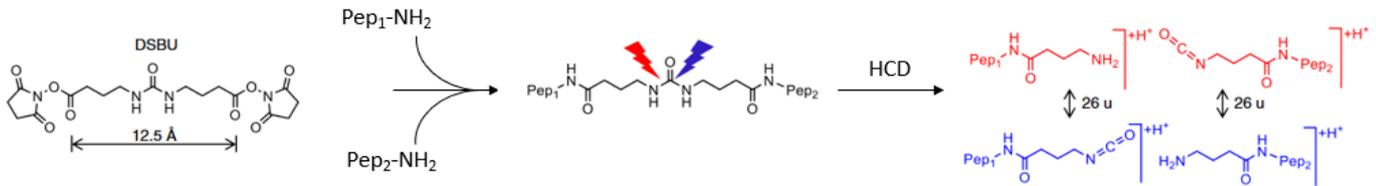


Figure 81 : Mécanisme réactionnel du pontage des protéines avec le LC-SMCC.

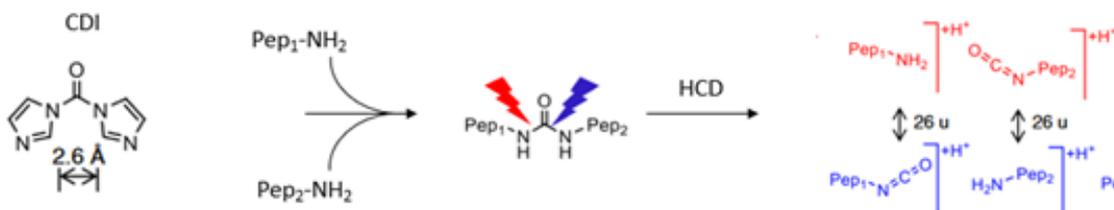
- deux agents pontants MS-clivable : Carbonyldiimidazole CDI [33] (homobifonctionnel lysine/lysine ; 2.6 Å) et disuccinimidyl dibutyric urea DSBU (homobifonctionnel lysine/lysine ; 12.5 Å)<sup>[34]</sup>

Ces agents pontant ont la particularité de générer des ions fragments spécifiques lors de leurs analyses MS facilitant ainsi leur validation d'identification<sup>[34-36]</sup>.



**Figure 82 : Mécanisme réactionnel du DSBU ainsi que son mécanisme de fragmentation dans les cellules HCD. Adapté de [35]**

Alors que le DSBU possède une longueur de bras espaceur similaire à celle du BS3, le CDI est un agent pontant clivable de type « zero-length ». Il s'agit d'un agent pontant offrant les contraintes de distances les plus petites (généralement <10 angströms) car il relie de manière covalente deux chaînes latérales sans ajouter de bras espaceur. Intuitivement, on pourrait penser qu'une longueur de bras espaceur plus petit va nous permettre de zoomer dans la structure et donc d'obtenir des informations structurales de plus haute résolution. En réalité, la réaction impose des contraintes stériques de distance et d'orientation des sites réactifs au niveau interfacial ce qui peut constituer une condition limitante.



**Figure 83 : Mécanisme réactionnel du CDI ainsi que son mécanisme de fragmentation dans les cellules HCD. Adapté de [35]**

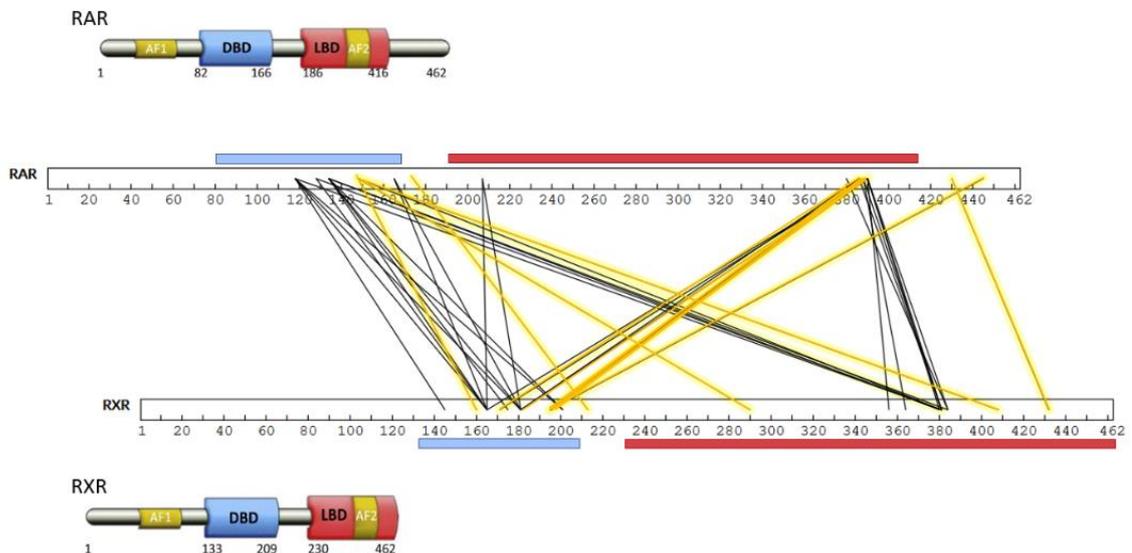
Au cours de cette étude, la réaction de pontage a été effectuée dans des conditions expérimentales strictement identiques afin de pouvoir comparer directement les efficacités de pontage des différents agents réticulants (30min à température ambiante à un ratio protéine/agent pontant à 1/50). L'interprétation des résultats a été réalisée par pLink (version 2.9) et les spectres MS/MS identifiés ont tous été vérifiés visuellement et validés à partir des critères de validation détaillée précédemment. L'ensemble des résultats est présenté dans le tableau 10.

### 3.3.2.1 Comparaison entre LC-SMCC et BS3 : Chimie réactionnelle différente

Tout comme avec le BS3, la réaction de pontage a été suivie par MALDI. Les spectres ont révélé dans ce cas un taux de réticulation plus modeste qu'avec le BS3 (environ 30-40% en comparant les intensités relatives entre le dimère et les monomères). Ceci est bien sûr approximatif car les facteurs de réponse des 3 espèces coexistantes en solution sont différents. Néanmoins, une comparaison des spectres MALDI obtenus pour le BS3 et pour le LC-SMCC dans des conditions d'analyse identiques nous confirme que le rendement de réaction est beaucoup moins important avec le LC-SMCC.

Après digestion trypsique, nous avons obtenu 11 peptides cross-linkés uniques. Sur ces 11 peptides, 3 impliquent la cystéine RXR195 (peptide CLAMGMK) et le peptide RAR386-390 (SISAKGAER). A partir des spectres de fragmentation, nous n'avons pas trouvé d'ions discriminants le site cross-linké sur RAR386-390 et il est difficile de dire si tous ces ions correspondent au même peptide cross-linké mais avec une interprétation par pLink différente, ou bien s'ils reflètent des cross-links mettant en jeu des sérines différentes sur le même peptide de RAR.

Dans les deux hypothèses, cette zone de proximité est bien représentée, soit par un même ion cross-linké et sélectionné dans le spectromètre de masse à plusieurs reprises (donc il s'agit d'un ion relativement intense), soit par des pontages multiples sur des sérines proches (donc ces deux peptides sont réellement proches). Ce pontage est particulièrement intéressant car il vient renforcer les 1<sup>ers</sup> résultats obtenus lors du pontage BS3 en montrant également une proximité entre le domaine DBD de RXR et le domaine LBD de RAR. Les 8 autres peptides cross-linkés (RXR160-RAR148, RXR171-RAR390, RXR195-RAR445, RXR213-RAR174, RXR290-RAR148, RXR381-RAR148, RXR408-RAR148 et RXR432-RAR430) sont également en accord avec la carte de proximité obtenue avec BS3.



**Figure 84 : Carte d'interaction entre RAR et RXR. En orange sont présent les peptides pontés par LC-SMCC et en noir ceux pontés par le BS3.**

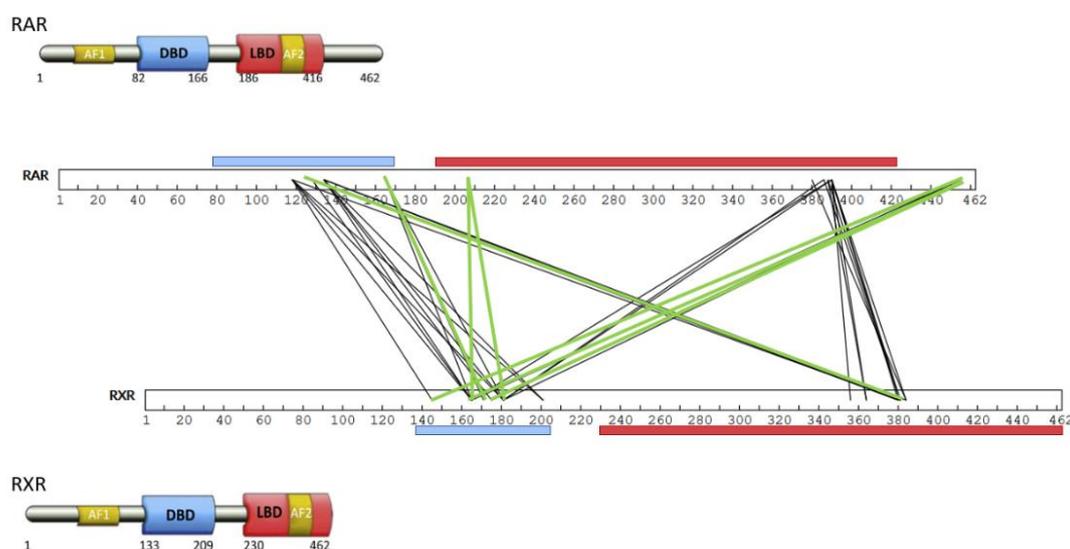
### 3.3.2.2 Comparaison entre le BS(PEG)<sub>5</sub> et le BS3 : même chimie réactionnelle mais longueur de bras espaceur différent

Dans ce cas, l'analyse par MALDI après pontage a montré des spectres très similaires à ceux observés avec le BS3, suggérant que le rendement de pontage fut équivalent.

28 peptides uniques cross-linkés par BS(PEG)<sub>5</sub> ont été obtenus par le logiciel pLink. Comme montré sur le tableau 10, les peptides cross-linkés par le BS(PEG)<sub>5</sub> ainsi identifiés sont, pour une grande majorité (20 peptides), identiques à ceux identifiés avec le BS3. Les 8 peptides différents se situent tous dans des zones de proximité déjà repérées avec le BS3 (voir la carte de proximité résultante (figure 85)). A l'inverse, 9 peptides uniques observés lors du pontage BS3 n'ont pas été identifiés. Parmi ces 9 peptides, 3 impliquent une connexion entre RXR<sub>364</sub> et le peptide SISAKGAER aux positions S386, S388 et K390. A partir des spectres MS/MS, il a été difficile de savoir si l'ensemble des ions détectés correspondent à un seul peptide ponté qui aurait été interprété par pLink de manière différente ou bien s'ils reflètent des pontages chimiques qui ont mis en jeu des sérines différentes sur le même peptide.

Les autres peptides manquant après pontage avec BS(PEG)<sub>5</sub> étaient observés entre le domaine DBD de RXR et le domaine DBD de RAR (RXR145-RAR118, RXR165-RAR134, RXR165-207, RXR201-RAR134). Il est difficile à ce stade de savoir si cette observation peut se traduire comme une donnée distancielle car cette zone de proximité est toutefois représentée par 4 autres peptides cross-linkés (RAR118 – RXR165, RAR118 – RXR201, RAR128 – RXR165 et RAR165 – RXR165).

Ainsi, de façon surprenante, modifier la longueur du bras espaceur de 11.4 Å (BS3) à 21.7 Å (BS(PEG)<sub>5</sub>) ne bouleverse pas de façon significative la carte de proximité entre RAR et RXR. 20 peptides uniques sont identifiés avec les deux agents de pontage ce qui montre que, (i) la présomption de ces zones de proximité est fiable (ii) la structure du BS(PEG)<sub>5</sub> est suffisamment flexible pour réagir sur des sites proches ou (iii) la structure protéique est suffisamment flexible pour que les mouvements protéiques placent les lysines/sérines dans une position favorable pour les deux agents pontants.



**Figure 85 : Carte de proximité montrant les peptides pontés observés par le BS3 et le BS(PEG)<sub>5</sub>. En vert sont les peptides uniquement observés en présence de BS(PEG)<sub>5</sub>**

### 3.3.2.3 Comparaison entre le BS3 et le CDI : même chimie réactionnelle mais longueur du bras espaceur plus petit

De la même manière qu'avec le BS(PEG)<sub>5</sub>, nous voulions voir si nous pouvions détecter des zones pontées supplémentaires à plus proche distance avec un agent pontant ayant un bras espaceur plus petit. Alors que nous démarrions le transfert de notre protocole vers des agents pontant MS-clivables, nous avons choisi le CDI.

L'analyse par MALDI du complexe RAR/RXR ponté par le CDI nous a révélé que la réaction de réticulation n'était pas très efficace (rendement de 20% environ). Néanmoins, nous avons tout de même réalisé la digestion trypsique puis l'analyse par LC-MS/MS. Malheureusement, seuls 6 peptides pontés ont été identifiés. Afin de vérifier que le milieu réactionnel ne pouvait être mis en cause et que le peu de réticulation observé était lié aux contraintes de distances intrinsèques au complexe RAR/RXR, nous avons lancé une requête d'analyse protéomique classique via le logiciel Mascot. Cette analyse nous a permis de vérifier d'une part que l'hydrolyse trypsique s'était opérée normalement et d'autre part que la réaction sur les amines des lysines (ou hydroxyle des sérines) s'était effectuée massivement. La figure 86 nous montre l'ensemble des acides aminés modifiés par le CDI (peptides « dead-end ») mais non pontés car la réaction d'hydrolyse de la seconde tête réactive du cross-linker a été plus rapide que la réaction de réticulation. Ceci semble suggérer que dans le cas du complexe RAR/RXR, peu de lysine/sérine se trouve dans une configuration de distance et d'orientation propice à un pontage par le CDI. Cette expérience n'a toutefois pas été optimisée en termes de conditions du milieu réactionnel (ratio protéine/CDI, T°C, pH, ...) et mériterait d'être reproduite.

#### RXR

MDTKHFLPLDFSTQVNSSLTSPTRGSMMAAPSLHP LGPGIGSPGQLHSPISLSSPINGMGPPFSVISSPMGPHSM  
SVPTTPTLGFSTGSPQLSSPMNPVSSSEDIKPPGLNGVLKVPAPHS GNMA SFTKHICAICGDRSSGKHYGVVYSCGCG  
KGGFKRTVRKDLTYTCRDNKDCLIDKRQRNRCQYCRYQ KCLAMGM KREAVQEERQRGKDRNENEVESTSSANED  
MPVEKILEAEAVEPKTETYVEANMGLNPSSPNDPVTNICQAADKQLFTLVWAKRIPHFSELPLDDQVILLRAGW  
NELLIASFSHRSIKVDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVS KM RDMQMDKT ELGCLRAIVLFPD SKGL  
SNPAEVEALREK VYAS LEAYCKHKYPEQGRFAKLLRLPALRSIGLK CLEHLFFFKLGIDTPIIDFLMEMLEAPHQAT

#### RAR

MASNSSSCTPGGGHNGYVPPYAFFFFPMLGGLSPPGALTTLQHQLPVSGYSTPSPATIQSSSSEEIVPSPSP  
PPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRR SIQK NMVYTCHRDKNCIINKVTRNRCQYRQLQKCFDVGMS  
KESVRNDRNKKKKEAPKPECESESYTLTPEVGELIEKVRKAHQETFPALCQLGKYTTNNSSEQRVSLDIDLWDFSELST  
KCIKTVEFAKQLPGFTTLTIADQITLLKAAACLDLILRICTRYTPEQDQMTMFS DGLTLNRTQMHNAGFGPLTDLVFAFA  
NQLLPLEMDAETGLLSAICLIGDRQDLEQPKVDMLQEPLLEALKVYVRKRRPSRPHMFPKMLMKITDLRSISAK  
GAERVITLMEIPGSMPPLIQEMLENSEGLDLSGQSGGGTRDGGGLAPPPGSCSPSLSPSSHRSSPATQSP

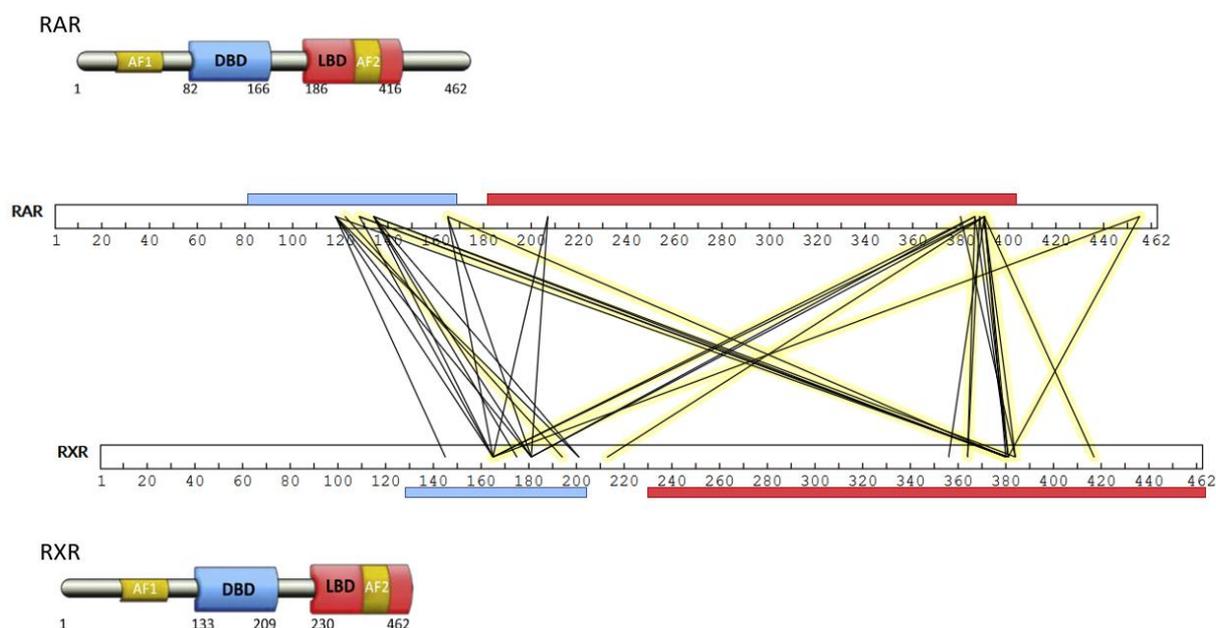
**Figure 86 : Séquence de RAR-RXR montrant les sites dead end observés pour le CDI (en jaune)**

### 3.3.2.4 Comparaison entre le BS3 et le DSBU : Efficacité dans l'identification des peptides pontés

Le DSBU et le BS3 possèdent une chimie réactionnelle identique, une longueur de bras espaceur équivalent (11.4 Å pour le BS3 et 12.5 Å pour DSBU). Ils devraient ainsi réagir de façon très similaire sur notre complexe RAR/RXR. L'avantage significatif à utiliser le DSBU est sa capacité à générer des ions fragments spécifiques et caractéristiques du pontage lors de l'analyse par MS/MS. Par cette expérience, nous avons voulu profiter de cet avantage énorme lors de l'identification des peptides pontés pour (i) faciliter la validation manuelle systématique et (ii) apporter une confirmation indépendante des peptides cross-linkés que nous avons déjà validés avec le BS3. La présence de ces ions fragments caractéristiques du pontage apportera un poids important dans la validation de peptides pour lesquels le recouvrement de séquence est jugé un peu faible notamment.

Au moment de cette étude, nous disposons de la version pLink 2.9 qui n'est pas encore adaptée pour l'étude des peptides pontés clivables en MS/MS. La seule option pour étudier les agents clivables par pLink est d'effectuer des analyses en MS3 ce qui n'a pas été réalisable avec l'Orbitrap Qexactive plus. C'est pourquoi pour effectuer cette comparaison, nous avons utilisé MeroX, un logiciel qui a été développé afin d'identifier les peptides pontés clivables et qui est donc plus adapté que pLink.<sup>[35]</sup>

Nous avons ainsi pu identifier 31 peptides intermoléculaires uniques lors du pontage avec le DSBU dont 21 sont identiques à ceux obtenus avec le BS3. Les 10 peptides non communs avec le BS3 se situent tous dans les zones de proximité précédemment décrites.



**Figure 87 : Carte de proximité des peptides pontés par le DSBU. En jaune sont représentés les peptides uniquement observés par le DSBU. En noir sont représentés les peptides observés de manière commune entre le DSBU et le BS3.**

### 3.3.2.5 Conclusion

Cette étude nous a montré qu'il est possible et utile de travailler avec différents agents pontants afin de consolider les résultats de pontage obtenus par MS avant de les partager avec nos collaborateurs pour des approches de modélisation structurale. En effet, tant que la reproductibilité à l'échelle peptidique est aléatoire, l'interprétation des données de fragmentation peut s'avérer délicate. Ainsi, la redondance observée avec différents agents pontants peut être considérée comme un gage de fiabilité des séquences pontées retenues.

**Tableau 10 : liste des peptides pontés intermoléculaires obtenus par plusieurs agents pontant interprétés par pLink et MeroX après validation manuelle. Les peptides en jaune sont les peptides qui ont été observés avec plusieurs agents pontants.**

Peptide1	Protein1	Peptide2	Protein2	Charge	Position 1	Position 2	Score	Commentaire	BSPEG5	LC-SMCC	BS3	CDI	DSBU
SSGKHYGVYSCGCK	RXR	SIQKNMVYTCR	RAR	5	145	118	6.355526e-004	K-K			X		
HYGVYSCGCK	RXR	SIQKNMVYTCRDK	RAR	4	151	118	32	K-K				X	
KDLTYTCR	RXR	SIQKNMVYTCR	RAR	5	165	118	3.206864e-005	K-K	X		X		X
DNKDCIDDKR	RXR	SIQKNMVYTCR	RAR	5	181	118	7.073574e-009	K-K	X		X		X
CLAMGMKR	RXR	SIQKNMVYTCR	RAR	5	201	118	2.309203e-005	K-K	X		X		X
AIVLFPDSKGLSNPAEVEALR	RXR	SIQKNMVYTCR	RAR	5	381	118	1.116240e-009	K-K	X		X		X
YQKCLAMGMKR	RXR	NMVTYTCR	RAR	3	194	122	30	K-K					X
DLTYTCRDNKDCIDDKR	RXR	NMVTYTCR	RAR	4	175	123	28	K-K				X	
KDLTYTCR	RXR	DKNCIINK	RAR	3	165	128	7.335703e-002	K-K	X		X		X
AIVLFPDSKGLSNPAEVEALR	RXR	DKNCIINK	RAR	4	380	128	1.154554e-005	S-K	X				X
AIVLFPDSKGLSNPAEVEALR	RXR	DKNCIINK	RAR	4	381	128	2.117775e-009	K-K	X		X		X
KDLTYTCR	RXR	DKNCIINKVTR	RAR	4	165	134	1.867912e-012	K-K			X		
DNKDCIDDKR	RXR	DKNCIINKVTR	RAR	4	175	134	1.067229e-009	K-K	X		X	X	X
DNKDCIDDKR	RXR	DKNCIINKVTR	RAR	4	181	134	4.108882e-010	K-K	X		X		X
CLAMGMKR	RXR	DKNCIINKVTR	RAR	3	201	134	3.789042e-002	K-K			X		
AIVLFPDSK	RXR	NCIINKVTR	RAR	3	380	134	3.161551e-001	S-K			X		
GLSNPAEVEALR	RXR	NCIINKVTR	RAR	3	384	134	1.604836e-002	S-K	X		X		X
GFFKR	RXR	CFDVGMSK	RAR	3	160	148	2.037266e-006	K-C		X			
IPHFSELPDDQVILLR	RXR	CFDVGMSKESVR	RAR	3	290	148	3.572727e-002	S-C			X		
AIVLFPDSKGLSNPAEVEALR	RXR	CFDVGMSK	RAR	3	381	148	4.092297e-006	K-C			X		
YPEQPGR	RXR	CFDVGMSKESVR	RAR	3	408	148	1.973452e-001	Y-C			X		
KDLTYTCR	RXR	NDRNKKK	RAR	4	165	165	1.429992e-001	K-K	X		X		X
EAPKPECSYTLTPEVGELIEK	RXR	KDLTYTCR	RAR	4	171	165	5.109150e-003	K-K	X				
DNKDCIDDKR	RXR	KAHQETFPALCQLGK	RAR	5	181	165	3.284669e-009	K-K	X		X		X
AIVLFPDSKGLSNPAEVEALR	RXR	NDRNKKK	RAR	4	380	165	26	K-K					X
EKVYASLEAYCK	RXR	KKEVPK	RAR	3	395	167	13	K-K				X	
EKVYASLEAYCK	RXR	KKEVPK	RAR	3	403	167	14	K-K				X	
VYASLEAYCKHKYPEQPGR	RXR	EAPKPECSYTLTPEVGELIEK	RAR	3	399	171	12	S-K				X	
GKDRNENEVESTSSANEDMPVEK	RXR	EAPKPECSYTLTPEVGELIEK	RAR	6	213	174	1.275212e-005	K-C		X			
KDLTYTCR	RXR	AHQETFPALCQLGKYTTNNSSEQ	RAR	5	165	207	2.598932e-004	K-K			X		
DNKDCIDDKR	RXR	AHQETFPALCQLGKYTTNNSSEQ	RAR	5	181	207	1.142147e-006	K-K	X		X		
GLSNPAEVEALR	RXR	MLMKITDLR	RAR	3	384	380	2.098583e-001	S-K	X				X
KDLTYTCR	RXR	ITDLRSISAK	RAR	3	165	386	24	K-K				X	X
KDLTYTCR	RXR	SISAKGAER	RAR	4	165	386	12	K-K					X
DNKDCIDDKR	RXR	SISAKGAER	RAR	3	181	386	6.387207e-001	K-S	X		X		X
CLAMGMK	RXR	SISAKGAER	RAR	3	195	386	1.039912e-004	C-S		X			
DMQMDKTELGLCLR	RXR	SISAKGAER	RAR	4	364	386	146	K-K			X		X
AIVLFPDSKGLSNPAEVEALR	RXR	SISAKGAER	RAR	3	381	386	2.022695e-003	K-S	X		X	X	X
CLAMGMK	RXR	SISAKGAER	RAR	3	195	388	8.476017e-005	C-S		X			
DMQMDKTELGLCLR	RXR	SISAKGAER	RAR	3	364	388	1.753665e-002	K-S			X		
AIVLFPDSKGLSNPAEVEALR	RXR	SISAKGAER	RAR	3	381	388	3.659226e-002	K-S	X		X		X
KDLTYTCR	RXR	SISAKGAER	RAR	3	165	390	1.342006e-009	K-K	X		X		X
DLTYTCR	RXR	SISAKGAER	RAR	3	171	390	1.983058e-002	C-K		X			
DNKDCIDDKR	RXR	SISAKGAER	RAR	3	181	390	1.381270e-003	K-K	X		X		X
CLAMGMK	RXR	SISAKGAER	RAR	3	195	390	1.156253e-004	C-K		X			
QRGKDR	RXR	SISAKGAER	RAR	3	213	390	182	K-K					X
VLTSLVSKMR	RXR	SISAKGAER	RAR	4	355	390	22	S-K				X	
VLTSLVSKMR	RXR	SISAKGAER	RAR	4	356	390	6.956368e-001	K-K	X		X	X	X
DMQMDKTELGLCLR	RXR	SISAKGAER	RAR	4	364	390	14	K-K			X	X	X
AIVLFPDSKGLSNPAEVEALREK	RXR	SISAKGAER	RAR	5	380	390	6.532381e-003	S-K	X		X	X	
AIVLFPDSKGLSNPAEVEALR	RXR	SISAKGAER	RAR	5	381	390	1.435353e-005	K-K	X		X	X	X
GLSNPAEVEALR	RXR	MLMKITDLRSISAKGAER	RAR	4	384	390	7.991130e-001	S-K			X		X
AIVLFPDSKGLSNPAEVEALR	RXR	SISAKGAER	RAR	5	384	390	1.104023e-003	S-K	X				
FAKLLLR	RXR	SISAKGAER	RAR	4	417	390	54	K-K				X	X
CLEHLFFF	RXR	SISAKGAER	RAR	3	432	430	8.257338e-006	C-S		X			
CLAMGMK	RXR	DGGGLAPPGSCSPSLSPSSNR	RAR	3	195	445	2.494549e-006	K-C		X			
SSGKHYGVYSCGCK	RXR	SSPATQSP	RAR	3	145	455	4.586031e-002	K-S	X				
KDLTYTCR	RXR	SSPATQSP	RAR	3	165	455	1.015630e-006	K-S	X				X
DNKDCIDDKR	RXR	SSPATQSP	RAR	3	175	455	6.162635e-006	K-S	X				
DNKDCIDDKR	RXR	SSPATQSP	RAR	3	181	455	1.282949e-016	K-S	X				
AIVLFPDSKGLSNPAEVEALR	RXR	SSPATQSP	RAR	3	381	455	211	K-S					X

### **3.4 Optimisation des paramètres d'interprétation**

Enfin, une part importante de l'optimisation du protocole concerne la validation des identifications de peptides cross-linkés. Cet aspect fut pendant longtemps le véritable frein à l'expansion de cette approche XL-MS. Alors que de nombreux logiciels dédiés au pontage ont été développés et se montrent disponibles open source<sup>[37]</sup>, en pratique, peu se sont avérés robustes et répétables entre nos mains : (i) selon le logiciel considéré, les formats d'entrée des données sont différents (.mgf pour pLink et .raw pour XlinkX), (ii) les fichiers mgf générés par les différents fournisseurs de MS ne sont pas parfaitement homogènes (Data Analysis v5.1 de Bruker introduit des commentaires qu'il faut supprimer), ce qui nécessite des manipulations de conversion des fichiers de données pour les rendre compatibles avec le logiciel choisi, (iii) certains logiciels ne sont pas adaptés à une requête dans de grandes bases de données protéiques (à l'échelle du protéome), (iv) certains logiciels ne sont pas adaptés à un changement des conditions expérimentales de cross-link (cross-linker MS-clivable ou non, protéase de digestion autre que la trypsine par exemple). D'autre part, les algorithmes utilisés pour les requêtes étant différents, il est possible que les résultats d'identifications diffèrent également, soit dans le nombre soit dans la qualité des interprétations. Ainsi, aucun logiciel ne satisfait à tous les critères pour analyser une large variété de données obtenues par XL-MS.

Afin de choisir le logiciel le mieux approprié mais également afin d'évaluer le gain à utiliser plusieurs logiciels par set de données, nous nous sommes attardés à comparer les résultats obtenus par 4 logiciels dont 3 sont open source (pLink, XiSearch, MeroX) et XlinkX commercialisé par la société ThermoFisher Scientific.

#### **3.4.1 Présentation des différents logiciels**

##### **3.4.1.1 XiSearch[38]**

Il s'agit d'un logiciel open source développé par l'équipe de Juri Rappsilber. Il s'agit d'un logiciel couramment utilisé pour identifier les peptides pontés non clivables à partir des spectres de fragmentation MS/MS. Pour identifier les peptides pontés, XiSearch identifie dans un premier temps les deux séquences branchées séparément en considérant le pontage comme une grande modification chimique non spécifiée. Dans un 1<sup>er</sup> temps, les spectres MS sont traités par un « pré-scoring » où uniquement ceux ayant un rapport signal/bruit correct sont gardés. Les peptides sont identifiés en deux étapes. Dans la première étape, le ou les candidats pour le peptide  $\alpha$  sont identifiés à partir des fragments d'ions observés sur le spectre MSMS. Dans un second temps, la masse du peptide  $\beta$  est déduite à partir de la différence entre la masse entre l'ion précurseur et le peptide  $\alpha$  candidat dans l'étape précédente en tenant compte bien sûr de la masse de l'agent pontant.

$$M_{peptide\beta} = M_{peptideXL} - (M_{peptide\alpha} + MX_{ler})$$

Où  $M_{peptide\beta}$  correspond à la masse du peptide  $\beta$  ;  $M_{peptide\alpha}$  correspond à la masse du peptide  $\alpha$ ,  $M_{peptideXL}$  correspond à la masse du peptide ponté et  $MX_{ler}$  correspond à la masse implémentée par l'agent pontant.

L'identification du peptide ponté est confirmée par comparaison de la masse des ions fragments théoriques et expérimentaux. Dans un dernier temps, une stratégie « target-decoy » est réalisée afin de calculer le taux de faux positifs en utilisant un second logiciel : XiFDR.

#### 3.4.1.2 MeroX[36]

MeroX : logiciel open source développé par l'équipe d'Andrea Sinz qui avait déjà développé StavroX<sup>[39]</sup>, logiciel couramment utilisé dans le domaine pour identifier les peptides pontés par des agents non clivables. MeroX possède les mêmes caractéristiques mais permet en plus d'identifier efficacement les peptides pontés MS-clivables en identifiant des fragments spécifiques à la liaison clivable sur les spectres MS2.

MeroX va dans un premier temps identifier tous les peptides pontés théoriques observables sur le complexe. Puis, les masses théoriques sont comparées avec les masses expérimentales issues des données MSMS. Par la suite, uniquement les spectres MSMS dont les masses précurseurs ont coïncidé avec les spectres théoriques sont conservés. Le logiciel calcule ensuite la masse de tous les ions fragments de type b et y théoriquement présents sur la séquence peptidique avant de les comparer aux spectres MSMS expérimentaux. Un score est par la suite attribué après interprétation des résultats. Dans un dernier temps, une stratégie « target-decoy » est réalisée afin de calculer le taux de faux positifs.

#### 3.4.1.3 XlinkX[40]

Il s'agit d'un logiciel développé par le laboratoire de A. Heck puis racheté par la firme ThermoFisher incluse comme un logiciel tiers du logiciel Proteome Discoverer<sup>TM</sup>. Ce logiciel permet d'identifier les peptides pontés aussi bien à partir d'agent non clivable que d'agent MS-clivable à partir de donnée issue soit d'analyses classiques de MS2 en utilisant comme cellule de collision l'HCD, soit en effectuant des analyses de types MS3 avec l'HCD ou encore en appliquant une analyse hybride où la MS2 est obtenus après HCD et la MS3 après ETD.

XlinkX effectue une recherche en deux étapes consécutives :

- Dans la première étape, XlinkX recherche la masse de tous les ions précurseurs de chaque peptide ponté en considérant le pontage comme une modification chimique présente sur les acides aminés et après avoir comparé les résultats avec une banque de données contenant l'ensemble des peptides pontés théoriques.
- Dans la seconde étape, tous les spectres qui contiennent au moins deux masses liées à deux peptides différents sont considérés comme des spectres potentiels de peptides pontés. Ces spectres sont par la suite soumis à une analyse séquentielle où l'ensemble des ions sont identifiés par rapport à une banque de données théorique constituée de l'ensemble des peptides pouvant être pontés. Un score est alors attribué en fonction de la qualité du spectre. En dernier lieu, la stratégie « target-decoy » est réalisée afin de calculer le taux de faux positifs.

### **3.4.2 Comparaison des logiciels d'interprétation sur notre système RAR/RXR ponté par le BS3**

Pour cette étude comparative, le même fichier de données en format mgf a été utilisé ainsi que les mêmes critères de requête d'identification : la tolérance du précurseur de masse a été fixée à 10 ppm et la tolérance MS/MS à 20 ppm, un minimum de 5 acides aminés par peptide et en considérant que le pontage puisse être réalisé sur les sérines et les thréonines. Chaque ensemble de données a été produit en utilisant un FDR de 5%.

Les résultats sont présentés dans le tableau 11. En l'absence de validation manuelle, 45 peptides pontés uniques ont été proposés avec pLink, 45 peptides uniques avec XlinkX 40 avec XiSearch et 27 avec MeroX. Alors que nous avons montré précédemment qu'il n'y a pas de relation linéaire entre le score et la confiance en l'identification au sein d'un même logiciel, il est également difficile de comparer transversalement l'identification d'un même peptide par tous les logiciels car l'attribution des scores est très différente selon le logiciel utilisé. En effet, tant l'algorithme de recherche que la valeur du score résultant diffèrent d'un logiciel à l'autre. Afin de nous aider à évaluer les logiciels, nous avons examiné tous les spectres MS/MS résultant d'une identification de peptide ponté et réalisé une validation manuelle.

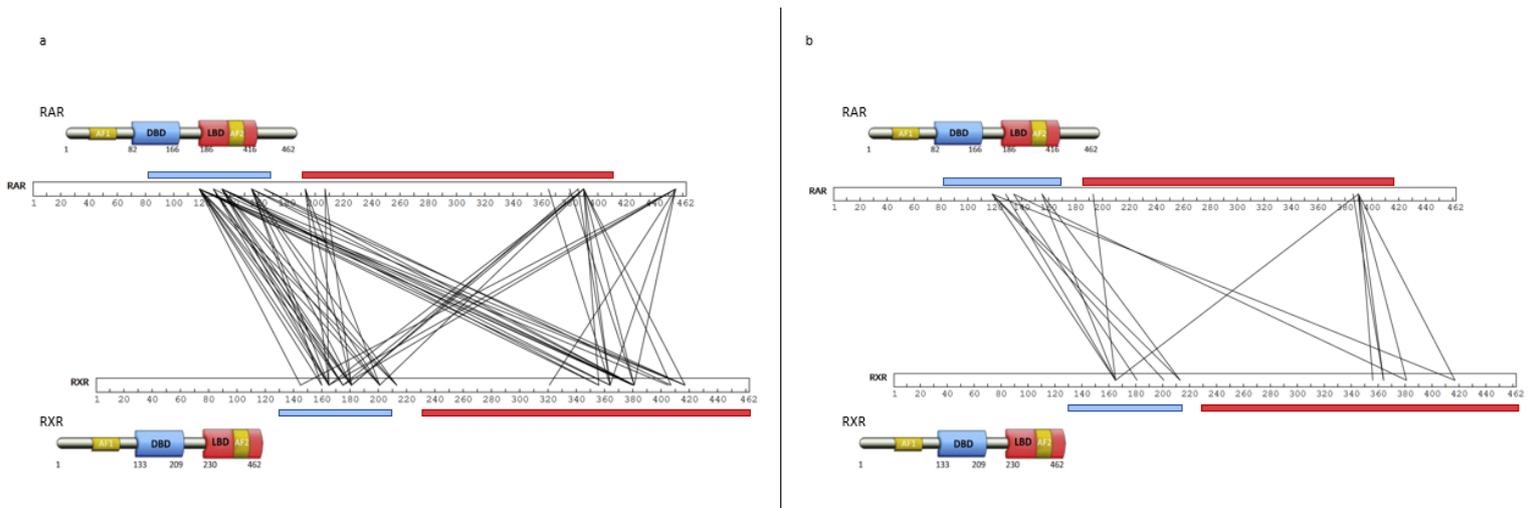
Il en ressort que :

- Le nombre et la nature des peptides pontés identifiés diffèrent d'un logiciel à l'autre.
- La validation manuelle a un effet plus ou moins drastique sur la liste finale des résultats selon le logiciel : ainsi, 13% des peptides identifiés par XiSearch furent éliminés après validation manuelle, 22% pour pLink, 5% pour XlinkX et 7% pour MeroX. pLink est le logiciel qui fournit le plus de peptides pontés candidats mais pour lequel la validation manuelle est la plus significative (le nombre de spectres MS/MS rejetés est le plus important). MeroX et XlinkX sont les logiciels pour lesquels nous observons le moins de rejet lors de la validation manuelle (7 et 5% respectivement)
- Certains peptides ne sont identifiés que par un seul logiciel (16 peptides), d'autres peptides sont reconnus par l'ensemble des 4 logiciels (16 peptides). Nous avons attribué à chaque identification, arbitrairement un score de 50 points par logiciel. En émettant l'hypothèse qu'un spectre MS/MS interprété de façon identique par plusieurs logiciels indépendants a davantage de poids dans nos critères de validation qu'un spectre MS/MS observé par un seul logiciel, ce score arbitraire nous permet d'extraire rapidement les interprétations non ambiguës.
- Il n'y a pas de relation linéaire entre le score d'identification et le nombre de logiciels qui ont identifié ce peptide. Par exemple le peptide RAR390-RXR356 est vu par les 4 logiciels alors que les scores sont modestes (6.04,  $10^{-5}$  pour XiSearch et pLink respectivement)

La validation manuelle des spectres MSMS reste donc un moyen incontournable pour s'assurer de la qualité d'identification des spectres MS/MS quel que soit le logiciel considéré. Toutefois, pLink fut dans notre cas d'étude, le logiciel fournissant le plus de faux-positifs. Ceci peut s'avérer une réelle difficulté pour l'étude de systèmes biologiques plus grands voire à l'échelle du protéome où la validation manuelle ne sera plus humainement réalisable.

L'utilisation plusieurs logiciels est particulièrement intéressante lorsque la validation manuelle est litigieuse. En effet, lorsque l'interprétation des spectres MS/MS par des logiciels

différents converge vers un même peptide cross-linké, cela nous rassure d'une part sur le choix de nos critères de validation manuelle mais aussi sur la confiance en la liste définitive de résultats fiables. L'instauration d'un score arbitraire (compris ici entre 50 et 200) va également permettre de faciliter la visualisation des peptides pontés sur les cartes de proximités par XiView en fonction d'un critère de confiance (figure 88). Il nous semble important de pouvoir isoler les peptides cross-linkés parfaitement validés sans pour autant perdre l'information de peptides cross-linkés candidats qui mériteraient d'être confortés par un cross-linker ou une protéase différente.



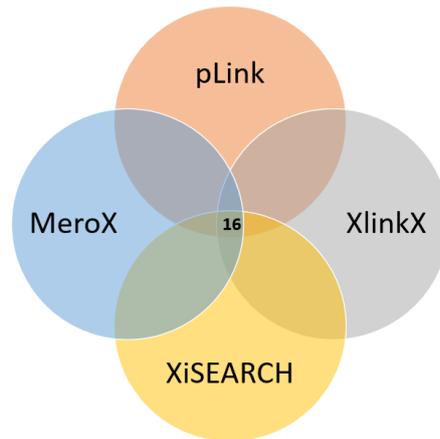
**Figure 88 : Carte de proximité obtenue en considérant (a) un score arbitraire de 50 (peptides interprétés par au moins un logiciel) et (b) pour un score de 200 (les peptides ont été détectés par les 4 logiciels).**

**Tableau 11 : Comparaison des peptides cross-linkés identifiés par 4 logiciels. En jaune sont représentés les peptides identifiés par au moins 2 logiciels. Entre parenthèses est noté le nombre de spectres MSMS qui ont conduit à la même identification.**

Peptides	Protéine1	Peptide Position 1	Protéine2	Peptide Position 2	xiSEARCH	MeroX	pLink	XlinkX	Score Arbitraire
SSPATQSP - SSGKHYGVYSECGCK	RAR	455	RXR	145		145(2)			50
SSPATQSP - KDLYTYCR	RAR	455	RXR	165		26(2)			50
DNKCIINK - DNKCLIDKR	RAR	128	RXR	175		120 (2)			50
SSPATQSP - DNKCLIDKR	RAR	455	RXR	175		19(2)			50
CFDVGMSKESVR - DNKCLIDKR	RAR	155	RXR	181				106,76 (8)	50
AHQETFPALCQLGKYTTNNSSEQR - DNKCLIDKR	RAR	207	RXR	181			1.512268e-005 (4)		50
CFEVGMSKESVR - CLAMGMKREAVQEER	RAR	155	RXR	201	6,44 (3)				50
DNKCIINKVTR - VLTELVSKMR	RAR	134	RXR	356			2.347400e-006 (3)		50
SIQKNMVTYCHR - DMQMDKTELGLCLR	RAR	118	RXR	364				135,16 (5)	50
DNKCIINKVTR - DMQMDKTELGLCLR	RAR	134	RXR	364				41,65 (2)	50
DMQMDKTELGLCLR - KRRPSR	RAR	365	RXR	364	7,58 (2)				50
CFDVGMSKESVR - AIVLFNPD SKGLSNPAEVEALREK	RAR	155	RXR	381				81,78 (1)	50
NDRNKK - AIVLFNPD SKGLSNPAEVEALR	RAR	164	RXR	381	6,67 (1)				50
MLMKITDLR - AIVLFNPD SKGLSNPAEVEALR	RAR	380	RXR	381			6.456695e-003 (3)		50
ESVRNDRNKKK - HKYPEQPGFR	RAR	164	RXR	407	7,79 (2)				50
DNKCIINKVTR - FAKLLLR	RAR	134	RXR	417			1.865275e-003 (1)		50
SIQKNMVTYCHR - SSGKHYGVYSECGCK	RAR	118	RXR	145			1.761603e-004 (5)	69,38 (3)	100
AHQETFPALCQLGKYTTNNSSEQR - KDLYTYCR	RAR	207	RXR	165			1.318391e-005 (4)	72(2)	100
SIQKNMVTYCHR - DNKCLIDKR	RAR	118	RXR	175				44,14 (1)	100
ITDLRSISAKGAER - DNKCLIDKR	RAR	390	RXR	175	13,67 (14)				100
DNKCIINK - DNKCLIDKR	RAR	128	RXR	181	12,01 (3)			101,61 (7)	100
SISAKGAER - DNKCLIDKR	RAR	390	RXR	181			4.200932e-006 (15)	120,29 (14)	100
DNKCIINKVTR - CLAMGMKR	RAR	134	RXR	201			1.232491e-004 (2)	95,44 (2)	100
SISAKGAER - CLAMGMKR	RAR	390	RXR	201			7.208745e-007 (4)	80,38 (3)	100
SSPATQSP - SIAVKGILLATGLHVHR	RAR	455	RXR	321	18 (4)	48 (7)			100
SSPATQSP - MRDQMDKTELGLCLR	RAR	455	RXR	364	19 (2)	33 (5)			100
DNKCIINK - AIVLFNPD SKGLSNPAEVEALR	RAR	128	RXR	381	6,61 (1)			56,01 (3)	100
SSPATQSP - AIVLFNPD SKGLSNPAEVEALREK	RAR	455	RXR	381	17 (1)	24 (2)			100
SIQKNMVTYCHR - VYASLEAYCKHKYPEQPGFR	RAR	118	RXR	405			9.682422e-002 (4)	86,36 (2)	100
SISAKGAER - HKYPEQPGFR	RAR	390	RXR	407	6,26 (3)		7.579237e-001(1)		100
CFDVGMSKESVR - CLAMGMKR	RAR	155	RXR	201	17(15)			161,77 (6)	100
SIQKNMVTYCHR - VLTELVSKMR	RAR	118	RXR	356		30 (1)		40,04(1)	100
DNKCIINKVTR - KDLYTYCR	RAR	134	RXR	165	9,89 (5)		7.117773e-007 (20)	90,29 (5)	150
DNKCIINKVTR - DNKCLIDKR	RAR	134	RXR	175	14,13 (10)		1.722728e-006 (5)	142,11 (4)	150
SIQKNMVTYCHR - DNKCLIDKR	RAR	118	RXR	181	17,09 (9)		5.590011e-004 (22)	168,11 (22)	150
KAHQETFPALCQLGK - DCLIDKR	RAR	193	RXR	181	16,27 (5)		3.269682e-005 (21)	135,36 (19)	150
SISAKGAER - DNKCLIDKR	RAR	386	RXR	181	12,24 (10)		2,568732e-006 (7)	35(2)	150
SIQKNMVTYCHR - AIVLFNPD SKGLSNPAEVEALR	RAR	118	RXR	380	18,26 (5)		3.779402e-009 (7)	52,00 (2)	150
SIQKNMVTYCHR - AIVLFNPD SKGLSNPAEVEALR	RAR	118	RXR	381	18,50 (8)		5.593745e-009 (16)	108,09 (15)	150
CFDVGMSKESVR - FAKLLLR	RAR	155	RXR	417		194 (1)	3.820523e-005(1)	128,65 (1)	150
SIQKNMVTYCHR - KDLYTYCR	RAR	118	RXR	165	17,25 (5)	105 (9)	2.153072e-007 (25)	150,89 (25)	200
DNKCIINK - KDLYTYCR	RAR	128	RXR	165	15,03 (7)	44 (3)	1.293574e-006 (5)	135,81 (3)	200
CFDVGMSKESVR - KDLYTYCR	RAR	155	RXR	165	16(10)	64 (6)	1.066215e-005(18)	124,34(20)	200
KAHQETFPALCQLGK - KDLYTYCR	RAR	193	RXR	165	6,12 (4)	245 (2)	1.397893e-009 (13)	139,69 (4)	200
SISAKGAER - KDLYTYCR	RAR	390	RXR	165	13,37 (15)	69 (2)	2.272138e-012 (16)	161,68 (15)	200
DNKCIINKVTR - DNKCLIDKR	RAR	134	RXR	181	14,72 (7)	46 (3)	5.132564e-006 (18)	129,29 (10)	200
SIQKNMVTYCHR - CLAMGMKR	RAR	118	RXR	201	16,50 (4)	20 (1)	7.806897e-001 (40)	154,74 (36)	200
SIQKNMVTYCHR - GKDRNENEVESTSSANEDMPVEK	RAR	118	RXR	213	17,30(5)	100(2)	3.512005e-003(8)	140,07 (14)	200
CFDVGMSKESVR - GKDRNENEVESTSSANEDMPVEK	RAR	155	RXR	213	19(5)	35(4)	1.562242e-003(8)	135,35 (4)	200
SISAKGAER - VLTELVSKMR	RAR	390	RXR	356	6,04 (4)	69 (2)	1.403138e-005 (23)	106,84 (10)	200
SISAKGAER - DMQMDKTELGLCLR	RAR	386	RXR	364	9,21 (2)	44 (2)	3.193104e-005 (11)	21 (5)	200
SISAKGAER - MRDQMDKTELGLCLR	RAR	390	RXR	364	8,65 (2)	75 (15)	1.559511e-002 (5)	91,09 (6)	200
DNKCIINKVTR - AIVLFNPD SKGLSNPAEVEALR	RAR	134	RXR	381	16,55 (3)	50 (3)	7.868259e-005 (2)	44,45 (1)	200
SISAKGAER - AIVLFNPD SKGLSNPAEVEALR	RAR	390	RXR	381	17,63 (9)	80 (18)	1.428066e-012 (25)	91,09(75)	200
SIQKNMVTYCHR - FAKLLLR	RAR	118	RXR	417	7,65 (1)	35 (2)	2.463987e-008 (5)	93,60 (4)	200
SISAKGAER - FAKLLLR	RAR	390	RXR	417	13,85 (8)	242 (5)	8.971482e-010 (19)	100,91(7)	200

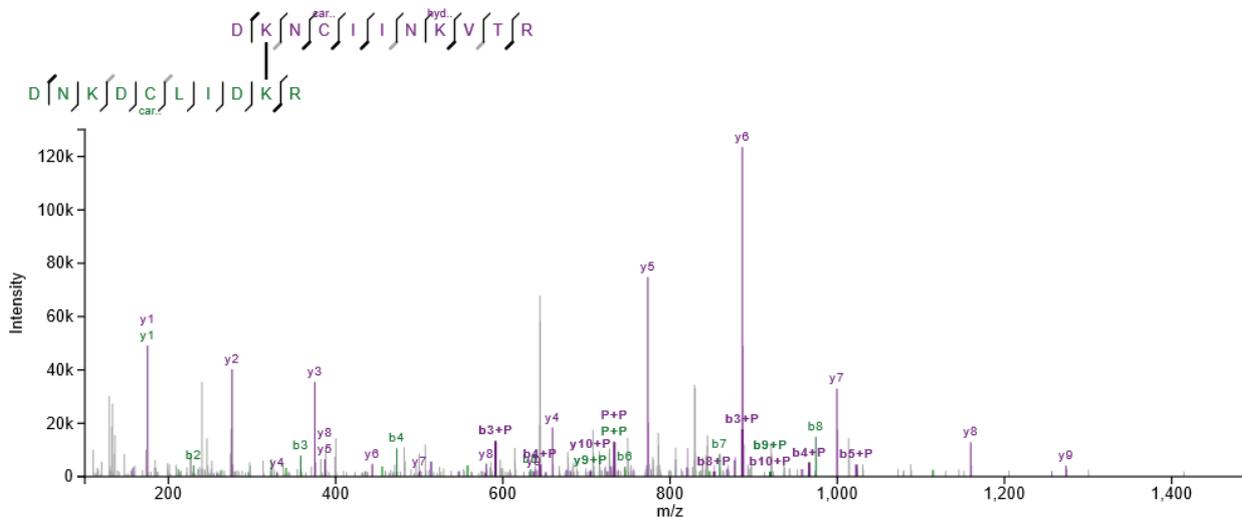
**La complémentarité d'identification des logiciels :**

Après validation manuelle, 21% des peptides identifiés ont été observés par les 4 logiciels, 24% par 3 logiciels et 30% par deux logiciels. Ainsi, seulement 25% des identifications ne sont observées par un seul logiciel.



**Figure 89 : Comparaison des paires inter peptidique réticulées uniques identifiées de RAR / RXR par MeroX, pLink, XiSearch et XlinkX.**

En tout, 75% des peptides pontés validés ont été observés par au moins deux logiciels, mais pas systématiquement les mêmes. Cette complémentarité permet d'enrichir la liste des peptides cross-linkés proposés et ainsi de gagner en informations structurales. En effet, le peptide RAR (128) - RXR (181) n'aurait pas été vu si nous avions utilisé que pLink alors qu'il a été visualisé et identifié correctement par XiSearch et XlinkX. La figure 90 montre que l'interprétation de ce peptide par XiSearch est non ambiguë.



**Figure 90 : Peptide RAR (128) -RXR (181) interprété par XiSearch. +P est la masse de l'ion précurseur.**

La pertinence des résultats nécessite une bonne reproductibilité :

Pour utiliser de manière pertinente les données de XL-MS, il est impératif d'effectuer des expériences de répétabilité<sup>[3]</sup> :

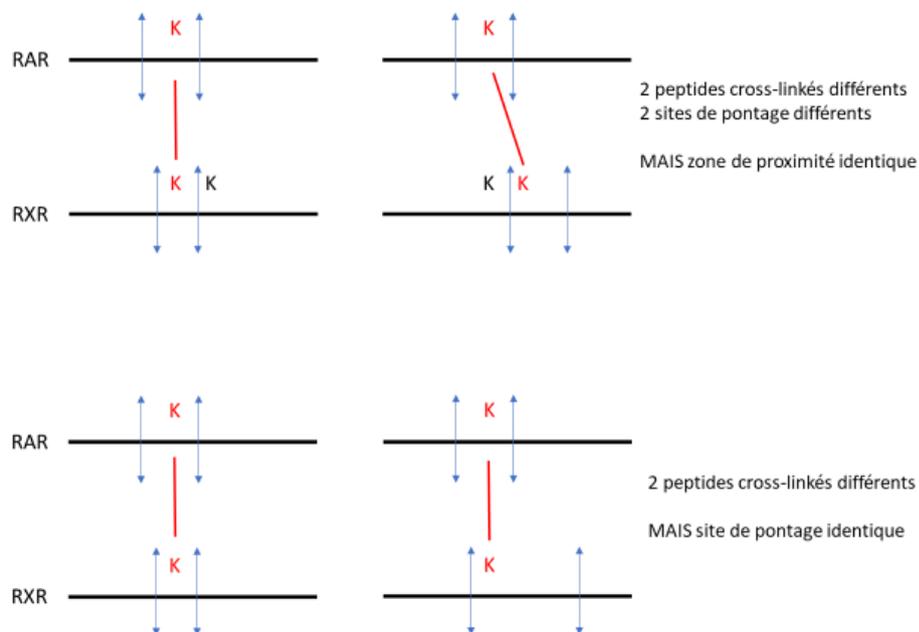
- (i) Notre protocole d'analyse (pontage covalent et analyse par LC-MS/MS) est-il robuste ? Il s'agit ici d'évaluer la reproductibilité technique.
- (ii) Le système sur lequel on travaille est-il conformationnellement robuste d'une préparation d'échantillon à l'autre ? Il s'agit ici d'évaluer la reproductibilité biologique.

Les études de répétabilité réalisées sur notre complexe ponté ont abouti à un taux de 70% de répétabilité technique au niveau peptidique sur la base de 3 injections du même digest peptidique (mesurée uniquement sur le pontage avec le BS3). Ce pourcentage peut s'expliquer par le fait que d'une analyse LC-MS/MS à l'autre, les ions parents sélectionnés peuvent ne pas être exactement les mêmes.

La reproductibilité biologique (c'est-à-dire l'analyse effectuée à partir de complexe biologique issu de productions protéiques différentes) a été réalisée à partir de 5 préparations d'échantillons différentes, à des moments différents. Elle est plus délicate à évaluer car il faut distinguer la reproductibilité au niveau peptidique et celle au niveau zone de pontage. En effet, pour une même zone de proximité, plusieurs peptides de masses précurseur différentes peuvent être présents dans le mélange peptidique.

Plusieurs paramètres sont susceptibles d'influer sur la reproductibilité biologique :

- i- La réaction de pontage peut s'effectuer avec un rendement différent et ainsi certains peptides cross-linkés ne seront pas systématiquement formés (le ratio protéine/cross-linker peut légèrement varier car la concentration en protéine ou en cross-linker n'est pas strictement identique),
- ii- Le pontage s'effectue entre des sites réactifs situés à une distance équivalente mais positionnés différemment sur la séquence protéique (ce qui conduit, après trypsinolyse, à des peptides pontés différents),
- iii- La trypsinolyse n'est pas strictement identique d'un complexe ponté à l'autre (selon la nature de l'acide aminé modifié lysine ou sérine/thréonine, la trypsine n'atteindra pas les mêmes sites de protéolyse,
- iv- La conformation du complexe est flexible ou bien plusieurs conformations coexistent en solution (les sites réactifs potentiels changent au cours du temps de l'expérience ce qui conduit également à des peptides pontés de nature différente)



**Figure 91 : représentation schématique illustrant la différence entre site de pontage (position des acides aminés pontés dans la séquence protéique) et zone de proximité (les acides aminés pontés ne sont pas identiques mais situés très proches dans l'espace tridimensionnel de la structure protéique). Les traits rouges entre 2 K représentent les cross-links, les flèches bleues représentent les hydrolyses tryptiques.**

Au niveau peptidique, celle-ci a été déterminée à 51% ce qui semble en accord avec ce qui est décrit dans la littérature<sup>[41]</sup>. Au niveau zone de pontage, la reproductibilité est excellente puisque nous observons systématiquement des peptides cross-linkés entre les mêmes régions protéiques :

- RAR-DBD (118-134) et RXR-BDB (165-181)
- RAR-DBD ((118-134) et RXR- LBD (380-381)
- RAR-LBD (386-390) et RXR-DBD (181)
- RAR-LBD (386-390) et RXR-LBD (364-381)

### **3.5 Conclusion**

En conclusion, la redondance observée avec différents agents pontants peut être considérée comme un gage de fiabilité des séquences pontées retenues. L'utilisation d'un enrichissement en peptides pontés par SCX, la mise en place d'une liste de critères de validation et l'utilisation de plusieurs logiciels d'interprétation se sont avérées essentielles et nous ont permis d'identifier de manière non ambiguë quelques zones d'interactions entre les RAR et RXR. Ce protocole ainsi défini sera utilisé pour la suite de notre projet.

Toutefois, plusieurs points nous semblent d'ores et déjà importants à souligner :

-Il est crucial de connaître la stœchiométrie du complexe ponté et de vérifier l'état d'avancement de la réaction de pontage. Ainsi, dans le cas de systèmes biologiques de grande taille, le recours à un spectromètre de masse de type MALDI équipé de détecteur haute masse pourrait compenser la limite de détection des gels de SDS-PAGE.

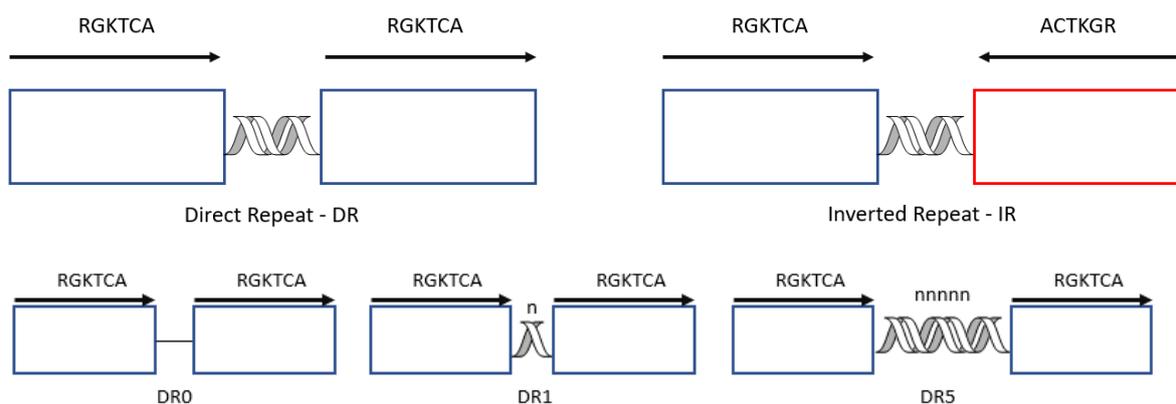
-la mise en place d'une séparation SCX (ou SEC) en amont de l'analyse LC-MS permettra d'identifier de manière plus efficace les zones pontées sur les complexes.

-le protocole de validation des peptides pontés en particulier lorsque le logiciel n'a identifié correctement qu'un seul des deux peptides reste à standardiser.

## **4. Etude des changements conformationnels de RAR-RXR lors de sa liaison avec l'ADN**

### **4.1 Le contexte**

Le complexe RAR/RXR agit comme un régulateur de la transcription qui active ou réprime les gènes. Il se lie spécifiquement à l'ADN, au niveau de séquences RAREs (retinoic acid response element).<sup>[3, 42]</sup> Les RAREs sont des fragments d'ADN qui contiennent deux répétitions d'une séquence hexamérique consensus de type 5'-PuG(G/T)TCA-3' (avec Pu= nucléotide présentant une purine comme base azotée : G/A) séparées par un nombre variable de nucléotides (n de 0 à 8)<sup>[4]</sup>. Ce motif hexamérique peut être répété de manière directe (« direct repeat - DR ») ou de manière indirecte (« inverted repeat » - IR). Ainsi, un RARE composé de deux répétitions directes du motif hexamérique séparées par 5 nucléotides sera noté DR5.



**Figure 92 : Représentation schématique des RAREs de type "Direct Repeat" à gauche et de type "Inverted Repeat" à droite. R=G/A et K=G/T**

Par une approche d'immunoprécipitation de la chromatine suivie par du séquençage (ChIP-Seq), des sites DR1, DR2 et DR5 comprenant la séquence consensus ont été repérés dans plusieurs types cellulaires. Des éléments de type DR0 ont également été reportés. De façon surprenante, l'association de RAR/RXR sur DR0 ne module pas l'expression des gènes *in vitro*<sup>[4, 43]</sup>. L'efficacité de liaison de RAR-RXR est plus faible sur les DR1, plus forte sur les DR2 et encore plus forte sur les DR5<sup>[44]</sup>. Ces derniers représentent la majorité des RAREs naturels. Alors que la fixation des récepteurs nucléaires sur les RAREs active l'expression des gènes cibles de l'acide rétinoïque (AR), il a aussi été rapporté que lorsque le dimère RAR-RXR se lie à des RAREs de type DR1, il y a un changement de son activité qui bascule de celle d'un activateur à celle d'un répresseur de transcription<sup>[45]</sup>.

Ainsi, même si l'activité de RAR/RXR est modulée par son interaction avec différents cofacteurs (coactivateurs ou corépresseur) ou ligands (agonistes, antagonistes), les RAREs sur lesquels l'hétérodimère va se lier, semblent également avoir une influence dans le recrutement de ces cofacteurs. En particulier, le nombre de nucléotides espaçant les deux motifs hexamériques pourrait être impliqué<sup>[46]</sup>. Comprendre le mécanisme de reconnaissance sur les éléments de réponse d'ADN ainsi que l'activation résultante de RAR/RXR nécessite de parfaire nos connaissances quant à la structure qu'adopte l'hétérodimère lorsqu'il se lie à l'ADN.

Depuis plusieurs années, des informations structurales indirectes sont obtenues par différentes techniques (gel natif d'électrophorèse, SAXS, RMN, MS, ...) mais aucune structure atomique n'est encore disponible sur l'hétérodimère RAR/RXR natif. Par MS non dénaturante, nous avons pu montrer un comportement différent des DBDs de RAR et RXR selon l'élément de réponse considéré : une forte hétérodimérisation de RAR et RXR en présence de DR5 est facilement détectée tandis qu'aucune spécificité d'interaction n'a été observée avec l'élément de réponse DR0<sup>[2]</sup>. Des études par cristallographie aux rayons X sont également effectuées avec persévérance mais se heurtent à une cristallogénèse capricieuse notamment si l'on travaille avec les séquences entières des récepteurs. Une structure atomique est néanmoins disponible pour l'hétérodimère RAR $\alpha$ /RXR $\alpha$  DBD<sup>[47]</sup> ainsi que pour l'hétérodimère RAR $\beta$ /RXR $\alpha$   $\Delta$ AB en présence de DR1<sup>[42]</sup>.

Pour cette étude, nous ne nous intéresserons pas aux interactions entre l'ADN et les protéines. Notre objectif ici était de déterminer si l'approche par XL-MS pourrait nous aider à mettre en évidence un changement de structure de l'hétérodimère selon la séquence des éléments de réponse auxquels il se lie, afin de savoir s'il existe une relation entre la structure adoptée par les récepteurs et leur activité. En parallèle, cette étude porterait une dimension plus fondamentale en nous fournissant des données de « résolution spatiale » de l'approche XL-MS, c'est-à-dire, sur l'ampleur des mouvements structuraux accessibles par cette technique.

Dans un 1<sup>er</sup> temps, nous allons chercher à identifier si certaines zones du complexe se sont réarrangées en présence d'ADN (dans notre cas en présence d'un oligonucléotide de type DR5) et dans un second temps, nous chercherons s'il est possible de mettre en évidence une différence dans le réarrangement conformationnel de l'hétérodimère selon le type d'ADN utilisé (DR5, DR1 ou DR0).

```

RARb2  DR5  AGGGTTCAccgaaAGTTCACT
          TCCAAGTggccttTCAAGTGA

Ramp2  DR1  TGAGTTCAaGGGTCAAT
          ACTCAAGTtCCCAAGTTA
  
```

**Figure 93 : Séquence de RAREs utilisés dans cette étude. Les deux sites hexanucléotidique sont notés en rouge. Les nucléotides présents entre les deux demi-sites sont notés en minuscule.**

#### **4.2 Réalisation du pontage sur le complexe RAR/RXR/ADN**

Nous avons travaillé dans des conditions de pontage chimique en tenant compte des optimisations exposées précédemment. 3 types d'éléments de réponse ont été considérés : un DR5 (RARb2DR5), un DR1 (Ramp2DR1) et un DR0 (Hoxb13DR0). Les expériences réalisées avec DR0 n'ont pas pu être reproduites, nous ne présenterons donc ici que les résultats obtenus avec des oligonucléotides DR5 et DR1.

Les oligonucléotides de type DR1 et de type DR5 ont été incubés avec un excès de 2 équivalences molaires en présence d'hétérodimère RAR/RXR  $\Delta$ AB. Puis, la réaction de pontage a été réalisée par l'ajout de BS3 selon un ratio 1 :50 pendant 30min à Tamb. Après digestion par la trypsine et enrichissement des peptides par colonnes SCX, les fractions peptidiques ont été analysées par LC-MS/MS. Un échantillon standard d'hétérodimère en l'absence d'ADN a été traité en parallèle.

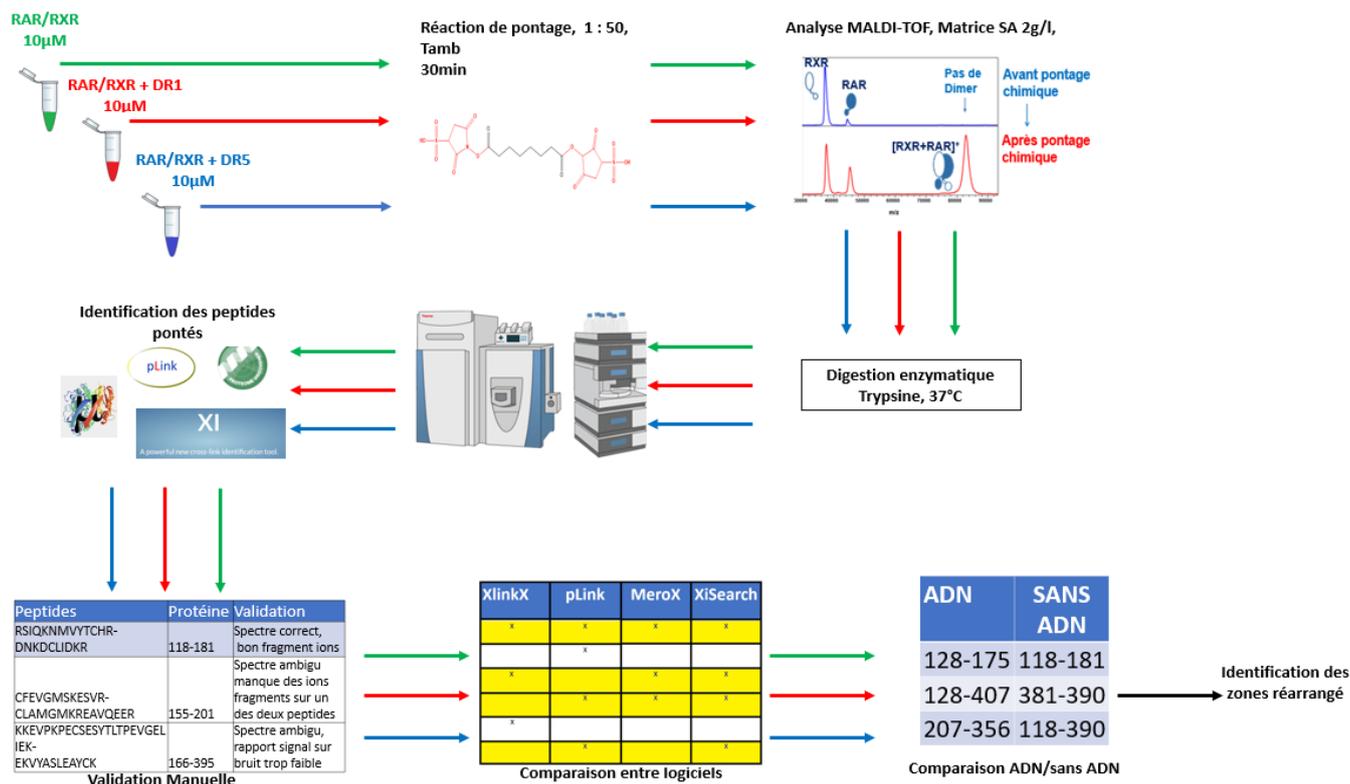


Figure 94 : Protocole d'analyse XLMS pour l'identification des zones réarrangé en présence d'ADN.

#### 4.3 Identification des peptides pontés : application de la validation manuelle et de l'identification par les différents logiciels.

L'identification des peptides pontés a été réalisée à l'aide des 4 logiciels que nous avons déjà évalués pLink, XiSearch, MeroX et XlinkX. Une validation manuelle a systématiquement été appliquée. Les critères de requêtes et de validation sont ceux précédemment exposés à savoir les 4 ions consécutifs observés sur les deux peptides ainsi qu'une erreur de masse sur l'ion précurseur <2ppm.

##### 4.3.1 Comparaison des zones de proximité de RAR/RXR sans et avec ADN

Cette étude différentielle a été réalisée 3 fois pour chacun des 3 échantillons (RAR/RXR, RAR/RXR + DR5, RAR/RXR + DR1) en utilisant 3 échantillons complexés nouvellement préparés. Le taux de reproductibilité biologique fut de 55% au niveau peptidique pour les échantillons complexés à l'ADN, et de 51% pour l'hétérodimère non complexé à l'ADN. Nous avons choisi de comparer les 3 réplicats d'analyse de façon indépendante plutôt que de les combiner. En effet, le taux de reproductibilité biologique nous a paru faible et dans cette 1<sup>ère</sup> démarche d'analyse qualitative, il nous a semblé plus judicieux de ne pas additionner simplement les tableaux de résultats. Ce faible taux de reproductibilité biologique peut être dû aux variations potentiellement induites par l'expérimentateur lors du protocole XL-MS d'une série d'expériences à l'autre mais peut également être lié à des préparations en protéine pas toujours parfaitement équivalente en termes d'homogénéité structurale, ce point sera discuté ultérieurement.

Sur le tableau 12 sont représentés les résultats obtenus pour la série d'analyse qui a donné les meilleurs résultats (nombre de peptides pontés plus élevé souvent associé également à des scores d'identification plus hauts). Ainsi, 53 peptides pontés intermoléculaires ont été identifiés pour le complexe RAR/RXR en l'absence d'ADN, 47 peptides pontés en présence de DR5 et 42 peptides pontés en présence de DR1. En complément de la validation manuelle systématiquement effectuée, un score arbitraire reflétant le nombre de logiciels qui ont conduit à l'interprétation du peptide cross-linké considéré est également indiqué (50 points par logiciel). On remarque que tous les peptides ont été identifiés par au moins 2 logiciels (score de 100 minimum, à l'exception du peptide RAR155-RXR417). En jaune, sont soulignés les peptides qui n'ont été identifiés strictement qu'en présence d'ADN (soit avec DR5 soit avec DR1). Ainsi, 20 peptides cross-linkés sont observés en présence d'ADN uniquement, 18 peptides cross-linkés ne sont observés qu'en absence d'ADN et 35 peptides cross-linkés sont communément observés avec et sans ADN.

Nous nous sommes focalisés dans un 1<sup>er</sup> temps sur les résultats obtenus en présence de DR5 car cet élément de réponse possède une bonne affinité pour RAR/RXR et déclenche efficacement l'activité de transcription des gènes correspondants<sup>[48]</sup>. Au niveau peptidique, on constate que 18 peptides cross-linkés n'ont été observés que sur l'échantillon qui contenait l'ADN DR5. A l'inverse, 25 peptides cross-linkés ne sont détectés que sur le complexe sans ADN. La visualisation différentielle de ces peptides selon la présence de DR5 peut s'expliquer de plusieurs manières :

- Certaines réticulations ne peuvent plus se réaliser en présence d'ADN car les acides aminés impliqués dans le pontage ont bougé.
- Certaines réticulations se réalisent en présence d'ADN car le changement de conformation associé à sa liaison avec RAR/RXR rend certains acides aminés accessibles et davantage proches.
- Certains peptides pontés n'ont pas été interprétés correctement.

Comme nous l'avons vu dans le paragraphe précédent, l'interprétation des résultats au niveau peptidique est compliquée car les acides aminés réticulés d'un échantillon à l'autre (tels que détectés par XL-MS) peuvent être différents sans que cela ne soit lié à un changement significatif de la conformation du complexe. Ainsi, nous avons élargi notre vision de la réticulation à l'échelle des zones de proximité. Trois modifications sont ainsi repérées :

- (i) La réticulation entre RAR118 et la zone RXR356-417 semble déstabilisée en présence de DR5
- (ii) RAR128 devient accessible à la réticulation en présence de DR5 et de nombreux peptides cross-linkés en résultent
- (iii) La réticulation entre RAR194-234 et RXR356-381 semble s'établir en présence de DR5

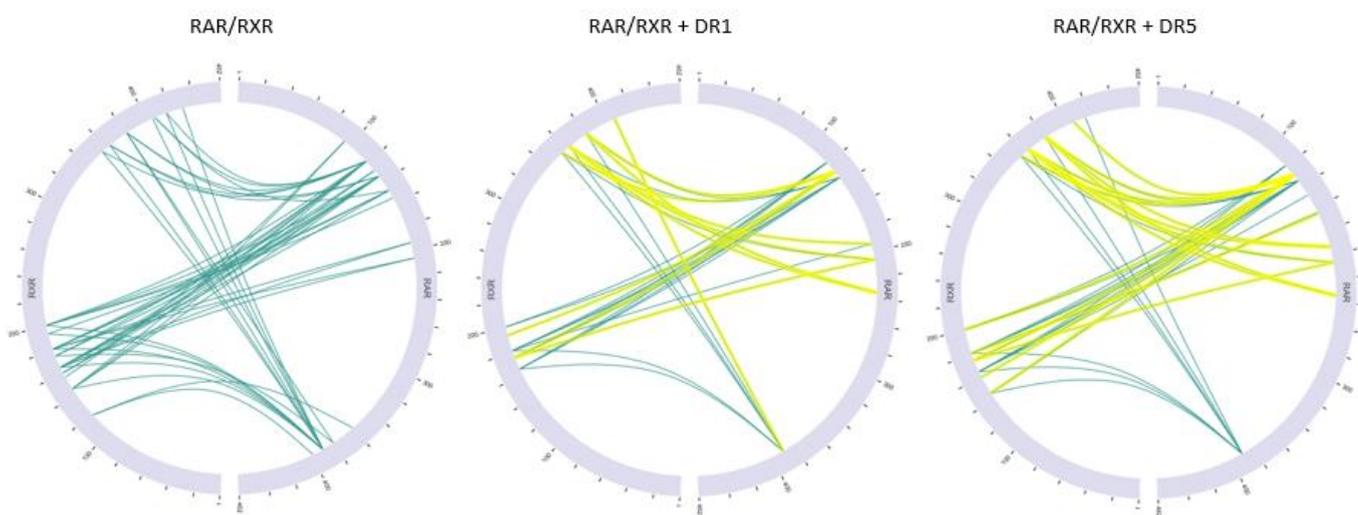
Ces modifications sont significatives car elles résultent de la détection de plusieurs sites cross-linkés convergeant vers la même conclusion. Par exemple, deux peptides montrent la proximité de RAR207 et RAR234 vers RXR356-381 en présence de DR5. De même, 4 peptides cross-linkés illustrent un changement d'accessibilité sur RAR118 lors de la complexation avec DR5. RAR118 se situe dans le DBD de RAR, il est ainsi possible qu'il soit impliqué dans l'interaction avec DR5 (ou qu'il se situe proche de l'acide aminé impliqué) ce qui constituerait une gêne stérique pour la réaction de pontage, ou bien, un mouvement

peptidique induit par l'association au DR5 éloigne suffisamment RAR118 pour concrétiser le pontage.

Il est à noter que ces 3 zones de pontage différentielles sont systématiquement observées d'un réplicat à l'autre, même si la nature des sites impliqués dans ces pontages n'est pas identique.

#### **4.3.2 Comparaison des zones de proximité lors de l'association avec DR5 ou avec DR1**

A partir des données du tableau 12, on constate que les résultats sont très similaires pour les complexes formés en présence de DR5 ou de DR1. Il y a 69% d'identité au niveau peptidique entre ces deux échantillons et aucune différence significative de zone de proximité n'a été clairement observée.



**Figure 95 : Carte de proximité des différents complexes étudiés en présence d'ADN DR5, d'ADN DR1 et sans ADN. En jaune sont les peptides uniquement observés en présence d'ADN.**

Ainsi, à partir d'une analyse XL-MS qualitative et en utilisant le pontage par le BS3, il n'a pas été possible de mettre en évidence une variation de proximité entre le complexe formé avec DR5 et DR1. Une analyse XL-MS quantitative ou encore l'utilisation d'un cross-linker possédant une chimie réactionnelle différente devront être envisagées pour déterminer si la résolution de l'approche XL-MS pourrait permettre d'illuminer les mouvements structuraux susceptibles d'être à l'origine du changement d'activité décrit dans la littérature.

**Tableau 12 : Comparaison des peptides pontés identifiés pour des échantillons en présence et en absence d'ADN. En jaune sont les peptides pontés observés uniquement sur le complexe en présence d'ADN.**

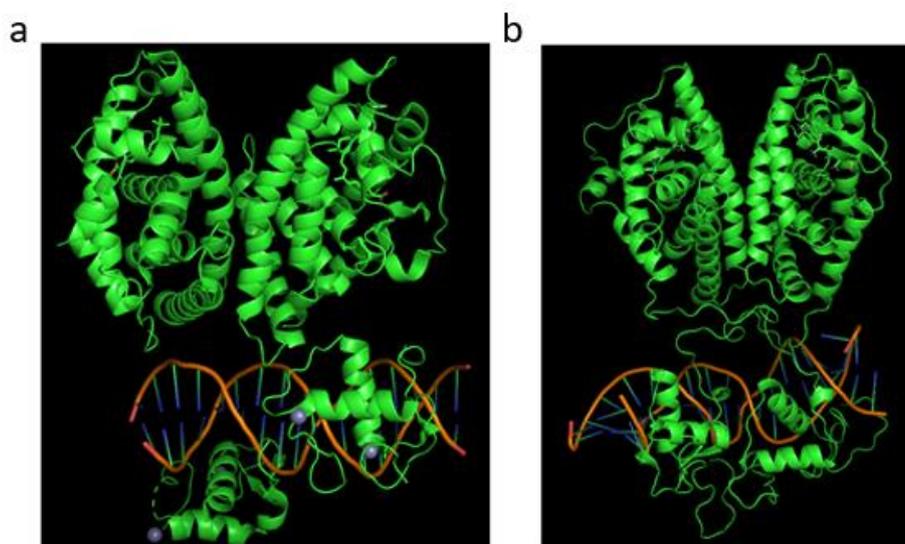
Peptide1	Protéine1	Position Peptide1	Peptide2	Protéine2	Position Peptide 2	DR1	DR5	BS3	Score arbitraire
SIQKNMVYTCR	RAR	118	SSGKHYGVYSCEGCKGFFKR	RXR	145			X	200
SIQKNMVYTCR	RAR	118	GFFKR	RXR	160			X	150
SIQKNMVYTCR	RAR	118	KDLTYTCR	RXR	165	X	X	X	200
SIQKNMVYTCR	RAR	118	DNKDCLIDKR	RXR	175	X		X	200
SIQKNMVYTCR	RAR	118	DNKDCLIDKR	RXR	181	X	X	X	200
SIQKNMVYTCR	RAR	118	CLAMGMKR	RXR	201	X	X	X	150
SIQKNMVYTCR	RAR	118	GKDRNENEVESTSSANEDMPVEK	RXR	213	X	X		200
SIQKNMVYTCR	RAR	118	VLTELVSKMR	RXR	356			X	150
SIQKNMVYTCR	RAR	118	DMQMDKTELGLR	RXR	364			X	150
SIQKNMVYTCR	RAR	118	AIVLFPDSKGLSNPAEVEALR	RXR	381			X	200
SIQKNMVYTCR	RAR	118	FAKLLLR	RXR	417			X	100
DKNCIINKVTR	RAR	128	SSGKHYGVYSCEGCKGFFKR	RXR	145		X		200
DKNCIINKVTR	RAR	128	GFFKR	RXR	160		X		150
DKNCIINKVTR	RAR	128	KDLTYTCR	RXR	165	X	X	X	150
DKNCIINKVTR	RAR	128	DNKDCLIDKR	RXR	175	X	X		200
DKNCIINKVTR	RAR	128	DNKDCLIDKR	RXR	181	X	X	X	100
DKNCIINKVTR	RAR	128	YQKCLAMGMK	RXR	194	X			100
DKNCIINKVTR	RAR	128	CLAMGMKR	RXR	201		X		100
DKNCIINK	RAR	128	GKDRNENEVESTSSANEDMPVEK	RXR	213	X		X	200
DKNCIINKVTR	RAR	128	VLTELVSKMR	RXR	356	X	X		100
DKNCIINKVTR	RAR	128	DMQMDKTELGLR	RXR	364		X		150
DKNCIINKVTR	RAR	128	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X	X	100
DKNCIINKVTR	RAR	134	SSGKHYGVYSCEGCKGFFKR	RXR	145		X	X	100
DKNCIINKVTR	RAR	134	GFFKR	RXR	160			X	100
DKNCIINKVTR	RAR	134	KDLTYTCR	RXR	165	X	X	X	150
DKNCIINKVTR	RAR	134	DNKDCLIDKR	RXR	175	X	X	X	150
DKNCIINKVTR	RAR	134	DNKDCLIDKR	RXR	181	X	X	X	150
DKNCIINKVTR	RAR	134	CLAMGMKR	RXR	201	X	X	X	150
DKNCIINKVTR	RAR	134	VLTELVSKMR	RXR	356		X		100
DKNCIINKVTR	RAR	134	DMQMDKTELGLR	RXR	364	X	X	X	100
DKNCIINKVTR	RAR	134	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X	X	100
DKNCIINKVTR	RAR	134	FAKLLLR	RXR	417	X		X	150
LQKCFEVGMSK	RAR	148	KDLTYTCR	RXR	165		X	X	100
LQKCFEVGMSK	RAR	148	YQKCLAMGMK	RXR	194			X	100
LQKCFEVGMSK	RAR	148	CLAMGMKR	RXR	201			X	100
CFDVGMSKESVR	RAR	155	GFFKR	RXR	160		X	X	100
CFDVGMSKESVR	RAR	155	KDLTYTCR	RXR	165	X	X	X	200
CFDVGMSKESVR	RAR	155	DNKDCLIDKR	RXR	181	X	X	X	200
CFDVGMSKESVR	RAR	155	CLAMGMKR	RXR	201	X	X	X	150
CFDVGMSKESVR	RAR	155	GKDRNENEVESTSSANEDMPVEK	RXR	213	X		X	200
CFDVGMSKESVR	RAR	155	FAKLLLR	RXR	417			X	50
NKKK	RAR	164	DNKDCLIDKR	RXR	181		X		100
NKKK	RAR	165	DNKDCLIDKR	RXR	181		X	X	100
KAHQETFPALCQLGK	RAR	194	KDLTYTCR	RXR	165			X	100
KAHQETFPALCQLGK	RAR	194	DNKDCLIDKR	RXR	181	X		X	100
KAHQETFPALCQLGK	RAR	194	DMQMDKTELGLR	RXR	364	X	X		100
KAHQETFPALCQLGK	RAR	194	AIVLFPDSKGLSNPAEVEALR	RXR	381		X		100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	KDLTYTCR	RXR	165			X	100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	DNKDCLIDKR	RXR	175	X	X		100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	DNKDCLIDKR	RXR	181			X	100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	VLTELVSKMR	RXR	356	X	X		100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	DMQMDKTELGLR	RXR	364	X	X		100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	DMQMDKTELGLR	RXR	364	X	X		100
KAHQETFPALCQLGKYTTNNSSEQR	RAR	207	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X		100
KFSELSTKCIK	RAR	234	DMQMDKTELGLR	RXR	364	X	X		100
KFSELSTKCIK	RAR	234	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X		100
KMLMKITDLR	RAR	380	DNKDCLIDKR	RXR	175			X	150
KMLMKITDLR	RAR	380	AIVLFPDSKGLSNPAEVEALR	RXR	381			X	150
SISAKGAER	RAR	390	SSGKHYGVYSCEGCKGFFKR	RXR	145		X	X	150
SISAKGAER	RAR	390	KDLTYTCR	RXR	165	X	X	X	200
SISAKGAER	RAR	390	DNKDCLIDKR	RXR	181	X	X	X	200
SISAKGAER	RAR	390	YQKCLAMGMK	RXR	194			X	200
SISAKGAER	RAR	390	CLAMGMKR	RXR	201			X	200
SISAKGAER	RAR	390	GKDRNENEVESTSSANEDMPVEK	RXR	213	X			150
SISAKGAER	RAR	390	VLTELVSKMR	RXR	356	X	X	X	200
SISAKGAER	RAR	390	DMQMDKTELGLR	RXR	364	X	X	X	200
SISAKGAER	RAR	390	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X	X	200
SISAKGAER	RAR	390	FAKLLLR	RXR	417		X	X	200
SISAKGAER	RAR	390	SIGLKCLEHLFFF	RXR	431			X	100
SSPATQSP	RAR	455	SSGKHYGVYSCEGCK	RXR	145	X	X		100
SSPATQSP	RAR	455	KDLTYTCR	RXR	165	X		X	100
SSPATQSP	RAR	455	DNKDCLIDKR	RXR	181	X	X		100
SSPATQSP	RAR	455	GKDRNENEVESTSSANEDMPVEK	RXR	213		X		100
SSPATQSP	RAR	455	SIAVKDGILLATGLHVHR	RXR	321	X	X	X	100
SSPATQSP	RAR	455	DMQMDKTELGLR	RXR	364	X		X	100
SSPATQSP	RAR	455	AIVLFPDSKGLSNPAEVEALR	RXR	381	X	X	X	100

### **4.3.3 Implémentation des données de XL-MS avec les éléments de structures disponibles : vers un modèle structural affiné ?**

#### 4.3.3.1 Implémentation des données et contraintes de distances

Une fois les zones de pontages identifiées, l'objectif était d'implanter les données obtenues par XL-MS sur les données structurales de RAR/RXR disponible afin d'établir un modèle le plus proche possible de la structure in vivo.<sup>[49-51]</sup> En effet, il a été montré dans la littérature que les contraintes de distance issues des données XL-MS peuvent être confrontées aux modèles structuraux dérivés d'autres techniques d'analyse telles que la cryo-EM, la SAXS, afin d'affiner la modélisation.<sup>[16, 50, 52]</sup> Ainsi, à l'aide de calculs de modélisation, il a été possible d'obtenir des modèles plus résolutifs.<sup>[53, 54]</sup>

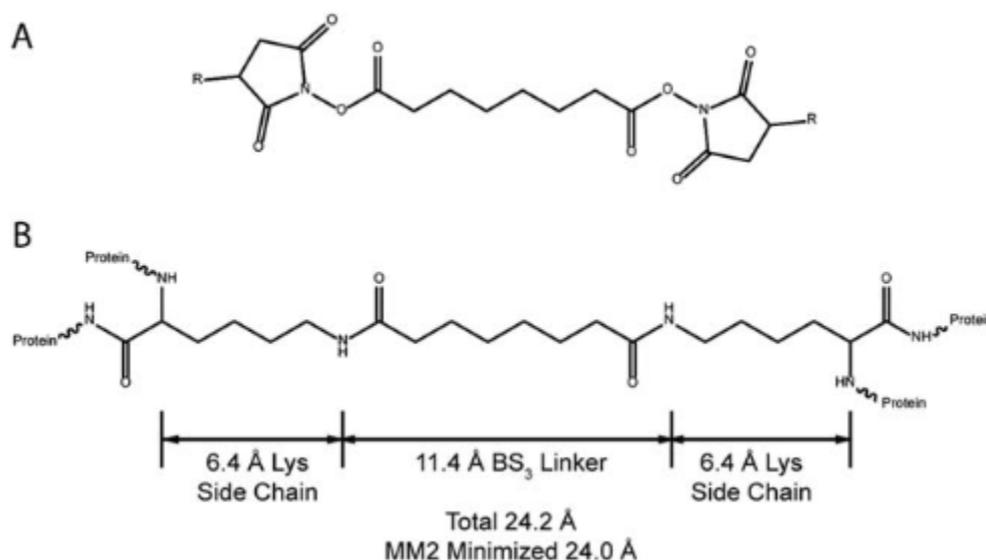
Dans notre cas, les données structurales disponibles ont été obtenues par SAXS (données non publiées). Nos collaborateurs disposent ainsi de plusieurs modèles structuraux, à basse résolution, qui proposent les meilleures convergences de modélisation à partir des jeux de données expérimentales. En particulier, 3 structures nommées C1-2, C2-2 et C3-2 obtenues sur le complexe en présence d'ADN DR5, seront considérées pour cette étude. En parallèle, nous avons également choisi de confronter nos données à la structure obtenue en présence d'ADN DR1 par cristallographie aux rayons X avec une résolution de 3.5 Å décrite dans la littérature<sup>[42]</sup>. Nous ne disposons d'aucun modèle structural fiable de l'hétérodimère multidomaine en l'absence d'ADN car celui-ci est vraisemblablement très flexible.



**Figure 96 : Structures du complexe RAR/RXR en présence d'ADN. (a) structure cristallographique de RAR/RXR avec DR1. (b) Structure SAXS de RAR/RXR avec DR5.**

Notre démarche consiste à localiser les peptides cross-linkés que nous avons pu identifier sur les structures disponibles et à évaluer si les contraintes de distances imposées par l'agent pontant sont compatibles ou non avec la position des acides aminés correspondants. Pour l'instant, nous nous sommes focalisés sur les résultats obtenus à partir du pontage par le BS3 car c'est avec ce cross-linker que nous disposons des données de XL-MS les plus fiables selon nous. Celui-ci se fixe sur des sites spécifiques (K, S, T) et possède un bras espaceur de longueur 11.4 Å. Ainsi, pour qu'un pontage s'établisse, il faut que les deux sites protéiques réactifs soient suffisamment proches pendant un temps suffisamment long et raisonnablement accessible au solvant.

Merkley et collaborateurs ont montré que le pontage par le BS3 s'établissait entre deux acides aminés dont les carbones C $\alpha$  étaient éloignés jusqu'à 24.2 Å (Cf figure 97). En tenant compte des probables mouvements conformationnels des protéines, il est admis de considérer une contrainte de distance entre les C $\alpha$  des acides aminés pontés de 30 Å [16]



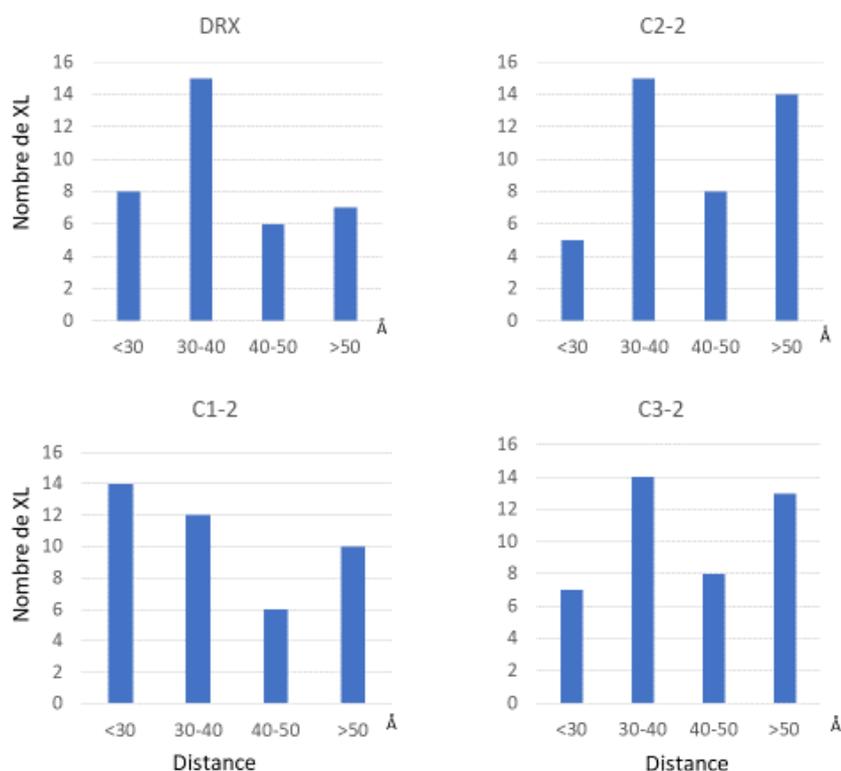
**Figure 97 : Structure chimique de l'agent pontant de la famille des esters NHS. (A) R=H, structure du disuccinimidylsuberate (DSS) ; R=SO $_3^-$ , structure du bis(sulfosuccinimidyl)suberate (BS3). (B) formation d'une liaison pontée entre deux lysines et l'agent pontant.[16]**

Le tableau 13 regroupe les distances entre les sites réticulés des peptides pontés que nous avons identifiés selon le modèle structural considéré (structure cristallographique DRX ou structures SAXS C1-2, C2-2, C3-2). Notre première impression est que, quels que soient les modèles structuraux considérés, beaucoup des distances mesurées se situent au-delà des 30 Å autorisés pour un pontage par le BS3. En effet, seulement 33 % des peptides cross-linkés identifiés relient des sites séparés par une distance  $d < 30$  Å si l'on considère la structure C1-2, 12%, 16% et 22% pour les structures C2-2, C3-2 et DRX respectivement ! Le modèle structural C1-2 est néanmoins celui qui semble le plus compatible avec nos données de cross-link (c'est également le modèle qui converge le mieux avec les données SAXS).

**Tableau 13 : Tableaux regroupant les distances des peptides pontés identifiés à partir de différents modèles. En vert sont notées les distances compatibles avec à la longueur d'extension maximale du BS3.**

Protein1	SeqPos1	Protein2	SeqPos2	DRX	C1-2	C2-2	C3-2
RAR	118	RXR	165	34.77	22.95	32.94	26.70
RAR	118	RXR	181	39.75	37.89	41.45	35.39
RAR	118	RXR	201	34.37	37.19	46.91	40.12
RAR	118	RXR	213	-	29.55	36.19	39.35
RAR	128	RXR	145	34.91	41.56	61.28	59.05
RAR	128	RXR	160	27.40	32.02	51.17	49.40
RAR	128	RXR	165	26.35	25.87	45.29	44.17
RAR	128	RXR	175	-	23.98	38.95	37.25
RAR	128	RXR	181	43.54	37.45	49.21	41.94
RAR	128	RXR	201	32.30	31.48	53.90	54.06
RAR	128	RXR	356	46.02	57.91	62.83	93.08
RAR	128	RXR	364	42.87	49.00	49.59	88.35
RAR	128	RXR	381	32.99	53.82	62.94	91.91
RAR	134	RXR	145	24.14	43.14	51.14	50.23
RAR	134	RXR	165	22.49	29.66	39.19	38.32
RAR	134	RXR	175	-	20.90	29.07	34.44
RAR	134	RXR	181	37.47	30.09	34.11	31.18
RAR	134	RXR	201	19.73	36.51	47.05	48.38
RAR	134	RXR	356	56.64	62.96	70.43	84.77
RAR	134	RXR	364	53.96	53.57	59.18	82.77
RAR	134	RXR	381	50.10	66.66	68.86	85.69
RAR	148	RXR	165	32.32	16.63	30.01	30.14
RAR	155	RXR	160	36.17	17.29	30.03	26.95
RAR	155	RXR	165	39.62	9.66	23.00	21.18
RAR	155	RXR	181	48.76	26.07	33.49	25.43
RAR	155	RXR	201	38.40	23.63	35.64	31.58
RAR	164	RXR	181	-	38.92	36.07	36.62
RAR	165	RXR	181	-	37.40	35.90	36.04
RAR	194	RXR	364	39.63	40.16	39.61	39.51
RAR	194	RXR	381	37.94	36.68	37.37	37.11
RAR	207	RXR	175	-	74.16	63.32	60.65
RAR	207	RXR	356	45.32	45.24	47.05	46.96
RAR	207	RXR	364	52.04	52.19	54.77	54.31
RAR	234	RXR	364	46.42	46.08	45.83	47.28
RAR	234	RXR	381	36.74	36.20	35.85	38.13
RAR	390	RXR	145	62.79	65.13	59.06	47.63
RAR	390	RXR	165	60.28	53.48	54.43	57.47
RAR	390	RXR	181	65.98	74.98	73.75	66.16
RAR	390	RXR	356	25.07	22.73	23.23	25.58
RAR	390	RXR	364	32.53	30.93	31.02	33.58
RAR	390	RXR	381	19.15	20.74	18.89	20.62
RAR	390	RXR	417	24.66	23.90	24.61	26.14
RAR	455	RXR	145	-	-	-	-
RAR	455	RXR	181	-	-	-	-
RAR	455	RXR	213	-	-	-	-
RAR	455	RXR	321	-	-	-	-
RAR	455	RXR	381	-	-	-	-

Un nombre important de cross-links réticulant des acides aminés éloignés de plus de 50 Å sont également observés quelle que soit la structure considérée, ce qui semble peu compatible avec la longueur du bras espaceur du BS3. Ces peptides correspondent à une réticulation croisée du RAR LBD avec le RXR DBD (par exemple RAR390-RXR165 ou RAR390-RXR181) ou du RAR DBD avec le RXR LBD (par exemple les peptides RAR134-RXR364 ou RAR128-RXR381).



**Figure 98 : Distribution des distances entre deux acides aminés pontés en fonction de la structure.**

Plusieurs hypothèses peuvent être émises pour commenter ces résultats :

- Le peptide ponté n'a pas été correctement identifié car les données MS ont été mal interprétées : ceci est ponctuellement possible mais semble peu probable sur l'ensemble des données incohérentes aux vues des spectres MS/MS qui ont été validés manuellement.
- Les modèles structuraux disponibles ne sont pas justes : cette hypothèse est également possible ponctuellement mais ne nous semble pas la plus réaliste aux vues du nombre élevé d'acides aminés mal positionnés et repartis sur plusieurs domaines des deux partenaires.
- La conformation du complexe (avec ou sans ADN) est perturbée lors de la préparation des échantillons (avant pontage), notamment lors de l'étape de dessalage sur colonne de gel d'exclusion. Nous avons l'expérience de ce type de dessalage sur les sous-domaines isolés (LBD ou DBD) mais il est possible que les protéines entières se comportent différemment.
- La conformation du complexe a été modifiée lors de la réaction de pontage chimique : quelques travaux ont reporté l'investigation d'une possible perturbation de la structure protéique durant la réaction de pontage. Ils n'ont observé aucune perturbation majeure de la conformation protéique lorsque l'excès molaire de cross-

linker était égal au nombre de sites réactifs sur la protéine<sup>[55, 56]</sup>. Néanmoins, dans le cas de l'hétérodimère RAR/RXR, cette hypothèse est envisagée car, même si la liaison avec l'ADN doit stabiliser le complexe, la conformation de l'hétérodimère multidomaines est flexible. Ainsi, au cours de l'évolution de la réaction de pontages covalents, des modifications irréversibles peuvent s'opérer, rendant alors certaines zones accessibles à des pontages qui seraient irréalistes sur la conformation native d'origine.

- Il existe plusieurs conformations en solution et nous observons ici une vue globale de toutes ces conformations.

Afin de vérifier ces différentes hypothèses, plusieurs études complémentaires ont été effectuées. Elles ont principalement fait appel à des techniques indépendantes du cross-link et visaient à vérifier l'intégrité structurale des échantillons avant et durant le pontage.

#### 4.3.3.2 Modification de la conformation du complexe après le pontage ? Utilisation d'une méthode complémentaire à la MS : la spectroscopie IR

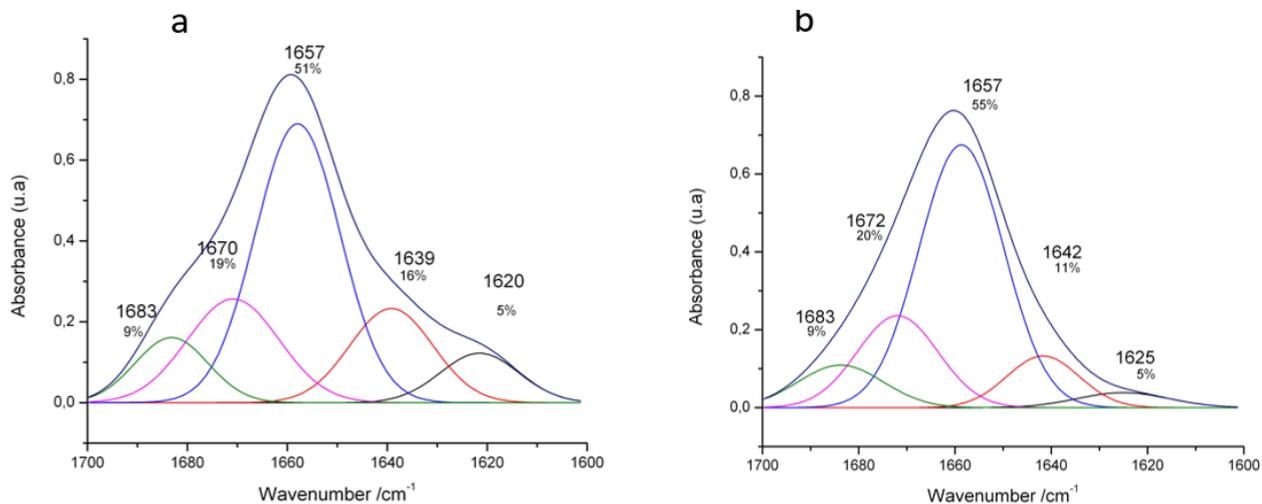
Afin de vérifier si un changement conformationnel est réalisé lors du pontage chimique, une technique analytique indépendante a été utilisée. Il s'agit de la spectroscopie infrarouge à transformée de Fourier (FTIR). Comme nous l'avons exposé dans la partie bibliographie, la détermination de la structure secondaire est rendue possible grâce à la décomposition de la bande amide I détectée dans la région spectrale 1700-1600  $\text{cm}^{-1}$  et est associée à l'élongation de la liaison C=O des groupements amides.<sup>[57]</sup>

Tout comme en MS, une faible quantité d'échantillon est nécessaire et l'analyse s'effectue facilement dans des tampons aqueux.<sup>[58]</sup> Nous avons donc choisi d'utiliser cette approche pour analyser la structure secondaire du complexe RAR/RXR avant et après pontage afin de contrôler si un changement conformationnel se produisait.

Pour cela, nous avons travaillé avec notre échantillon de RAR/RXR concentré à 10  $\mu\text{M}$  dans le tampon classiquement utilisé pour nos expériences de pontage c'est-à-dire 20mM d'HEPES, de 150mM de NaCl à pH 7.5.

Au cours de cette étude, nous avons utilisé un spectromètre de type Vertex 70 (Bruker) fonctionnant par transformée de Fourier. Pour les analyses réalisées en IR (3 000  $\text{cm}^{-1}$  à 700  $\text{cm}^{-1}$ ) le spectromètre est équipé d'une source Global<sup>TM</sup>, d'une séparatrice en bromure de potassium (KBr) et d'un détecteur MCT (Tellure de mercure-Cadmium) refroidi à l'azote liquide. L'ouverture de la fenêtre est de 8 mm, la fréquence de balayage est de 40 kHz et le nombre de scans pour chacun des spectres est de 256. Un accessoire de réflexion totale atténuée est placé dans le compartiment échantillon pour réaliser les analyses. Les échantillons sont ensuite analysés en IR moyen pour déterminer la structure secondaire du complexe RAR/RXR avant et après pontage par du BS3 dans des conditions strictement identiques à celles utilisées en XL-MS.

La déconvolution des bandes amides est réalisée avec le logiciel OriginLab 9.0 et les bandes individuelles sont analysées en courbes gaussiennes, avec une optimisation consécutive des amplitudes, des positions des bandes et de leur largeur à mi-hauteur. Le pourcentage de chacune des structures secondaires est déterminé, avec une erreur de 3,5%, par déconvolution (tableau 14).



**Figure 99 : Déconvolution de la bande Amide I (1700 - 1600 cm<sup>-1</sup>) obtenu par spectroscopie infrarouge montrant la structure secondaire du complexe RAR / RXR (a) et du complexe RAR / RXR réticulé par BS3 (b) Aucune évolution flagrante de la structure secondaire en présence de BS3.**

**Tableau 14 : Pourcentages des éléments de la structure secondaire de RAR/RXR avant et après pontage par le BS3 déduits de la figure 99. La précision des pourcentages est estimée à ± 3.5%**

Echantillon	Helice $\alpha$ + structure désorganisée /% (1657 + 1640)	Feuillet $\beta$ /% (1620)	$\beta$ - Coudes (1670-1680)
RAR/RXR	67	5	28
RAR/RXR ponté	66	5	29

La bande la plus intense présente à 1657 cm<sup>-1</sup> correspond ainsi à la vibration C=O des hélices  $\alpha$ . Une seconde bande détectable à 1639-1642 cm<sup>-1</sup> correspond aux structures désorganisées. Les bandes à 1620 cm<sup>-1</sup> et celles comprises entre 1670 cm<sup>-1</sup> et 1690 cm<sup>-1</sup> correspondent à la structure secondaire des feuillets  $\beta$  et des  $\beta$ -coudes respectivement.

En l'absence de pontage par le BS3, la structure secondaire majoritaire détectée pour l'échantillon d'hétérodimère RAR/RXR correspond à la structure hélices  $\alpha$  + désorganisées qui constitue 67% des structures secondaires détectées. La structure feuillet  $\beta$  a été détecté à 5% et la structure  $\beta$ -coudes à 28%.

Après pontage par le BS3, malgré le fait que la forme de la bande amide soit différente, nous obtenons des pourcentages très similaires des structures secondaires (66% de la structure du complexe ponté correspond à la structure hélices  $\alpha$  + désorganisées, 5% à la structure feuillet  $\beta$  et 29% à la structure  $\beta$ -coudes). La structure secondaire du complexe ne semble donc pas affectée de façon significative par la réticulation au BS3 dans nos conditions d'analyse. Une étude similaire a été menée en parallèle sur la BSA et les résultats ont

également confirmé que la structure secondaire de la BSA ne semblait pas être modifiée après pontage par le BS3.

En perspective, nous prévoyons d'utiliser l'IR lointain pour obtenir des informations sur d'éventuels changements conformationnels sur notre complexe en présence d'ADN.

#### 4.3.3.3 Modification de la conformation des échantillons d'une purification à l'autre.

Avoir connaissance de la qualité conformationnelle des échantillons (stœchiométrie de complexation, stabilité des interactions, constante d'affinité) et être en mesure de vérifier cette qualité structurale sont très importants pour interpréter nos données de XL-MS et en évaluer la pertinence. Un système biologique partiellement décomplexé ou dénaturé sera à l'origine d'une forte variabilité dans la nature des séquences cross-linkées et ainsi d'une baisse du taux de reproductibilité biologique voire d'une mauvaise interprétation des résultats.

Puisque les expériences par IR semblaient indiquer que l'échantillon ne subissait pas de modifications majeures suite au pontage, nous avons cherché à savoir si les incohérences dans nos données XL-MS ne seraient pas liées à un paramètre en amont de l'expérience analytique, c'est-à-dire au cours de la purification (ou stockage) de l'hétérodimère.

#### MS native du complexe RAR/RXR/ADN :

Dans un 1<sup>er</sup> temps, profitant de l'expertise acquise au laboratoire en spectrométrie de masse native, nous avons pensé contrôler l'homogénéité et la stœchiométrie d'interaction de l'hétérodimère RAR/RXR en l'absence et en présence d'ADN. La mesure de la masse moléculaire des espèces détectées en conditions non dénaturantes nous permettrait de vérifier que l'hétérodimère est quantitativement formé et que l'interaction avec les différents éléments de réponse (DR0, DR1 et DR5) s'effectue avec la stœchiométrie attendue.

Pour cela, comme exposé dans le chapitre III de la partie bibliographie, nous avons dessalé l'échantillon dans un tampon compatible avec l'ESI-MS, à savoir l'acétate d'ammonium AcONH<sub>4</sub> à 200mM pH 6.8. Ce tampon avait préalablement été utilisé avec succès pour l'étude des domaines LBD et DBD individuellement<sup>[2]</sup>. Pour cela, des systèmes de dessalage par centrifugation (Vivaspin) ou par gel d'exclusion (colonne BioSpin) ont été testés, sans succès ! Malgré nos efforts, le dessalage dans ce tampon a systématiquement échoué et l'intensité résultante sur les spectres de masse natifs ne nous a pas permis de conclure quant à l'homogénéité en stœchiométrie des complexes présents en solution. Nous nous sommes donc tournés vers un suivi sur gel natif d'électrophorèse.

#### Gel d'électrophorèse natif du complexe RAR/RXR/ADN

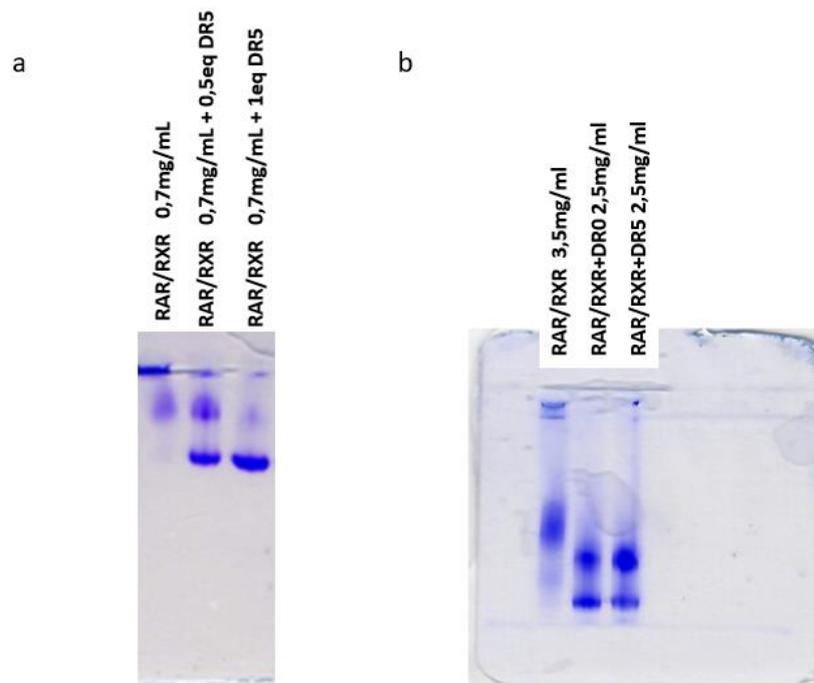
Très utilisée en biochimie, l'électrophorèse sur gel est une technique analytique permettant de séparer les molécules en fonction de leur charge et leur taille en les faisant migrer à travers un gel par application d'un champ électrique<sup>[59, 60]</sup>

Les gels natifs, également appelés gels non dénaturants, analysent les protéines qui sont encore dans leur état replié. Ainsi, la mobilité électrophorétique dépend non seulement de la charge des protéines, mais également de leur forme physique et de leur taille.<sup>[61]</sup>

Pour vérifier l'homogénéité des espèces en solution, les échantillons contenant l'hétérodimère seul ou en présence d'ADN, avant et après pontage par le BS3, ont été analysés sur un gel natif constitué de 4-15% de polyacrylamide.

Un premier dépôt du complexe RAR/RXR seul concentré à 3.5mg/mL a été réalisé afin d'observer la conformation initiale du complexe avant l'ajout d'agent pontant et avant sa liaison avec le partenaire oligonucléotidique. Une bande diffuse est observée, liée à la flexibilité du complexe multidomaines. Deux autres dépôts du complexe RAR/RXR lié à l'ADN DR0 à 2.5mg/mL et RAR/RXR lié à l'ADN DR5 à 2.5mg/mL ont été réalisés. En présence d'ADN, la bande observée sur le gel est plus fine (car la fixation de l'ADN sur l'hétérodimère rend la structure du complexe plus compacte) et la migration est plus rapide.

La figure 100 montre les gels natifs obtenus pour 2 préparations d'hétérodimères. Le gel (a) montre une bande unique pour le complexe RAR/RXR en présence d'1eq de DR5, démontrant une fixation du DR5 quantitative sur l'hétérodimère. En revanche, le gel (b), réalisé sur une préparation protéique différente, montre une migration non standard, avec deux bandes détectées, reflétant l'existence de plusieurs stœchiométries et conformations en solution. Ainsi, les préparations 1 et 2 se comportent différemment vis-à-vis de l'interaction avec l'ADN. Nos collaborateurs recherchent encore l'origine de ce comportement inhabituel sur gel.



**Figure 100 : Gel natif 4-15% avec différents dépôts de RAR/RXR sans ADN et avec ADN DR5 et DR0. (a) Gel natif obtenu lors d'une préparation antérieure. (b) gel natif obtenu lors de la préparation de nos échantillons.**

Quel que soit le RARE considéré (DR0 et DR5), le constat de deux bandes sur gel natif est sans appel.

La présence de deux conformations dans le même échantillon pose ainsi un réel problème. En effet, il est impossible de distinguer les deux conformations lors du pontage chimique sans faire appel à une analyse quantitative comparative.

**C'est pourquoi il est extrêmement important de bien contrôler l'échantillon purifié avant d'utiliser les données !**

En conclusion, la XL-MS a ainsi été appliquée sur le complexe RAR-RXR en présence d'ADN afin de mettre en évidence les réarrangements conformationnels protéiques lors de sa liaison avec l'ADN. Malgré l'observation de quelques différences entre l'hétérodimère libre et complexé à l'ADN, nous ne sommes pas parvenus à tirer des conclusions pertinentes sur un éventuel changement de conformation.

Très peu de peptides uniques ont été observés en présence du complexe avec ADN. De plus, une reproductibilité biologique incertaine a également été constatée. Des études effectuées par spectroscopie IR ont montré que le pontage n'entraînait pas de changement conformationnel du complexe. Toutefois, un problème lors de la purification du complexe a été détecté remettant ainsi en cause l'ensemble des études réalisées. C'est pourquoi au début d'une étude, il est impératif de contrôler l'échantillon purifié afin de ne pas obtenir des faux résultats. Des nouvelles études sur le complexe RAR/RXR en présence d'ADN seront réalisées afin de reprendre la recherche de changements conformationnels.

Afin de pouvoir mettre en évidence un changement conformationnel ou la coexistence de plusieurs conformations en solution, nous avons entrepris de mettre en place une méthodologie de XL-MS quantitative au laboratoire.

## 5. Mise en évidence d'un réarrangement conformationnel par analyse quantitative.

### **5.1 Les différents protocoles d'analyses quantitatives**

L'analyse quantitative des peptides pontés par spectrométrie de masse (QXL-MS) est une technique relativement nouvelle pour étudier la dynamique structurale des protéines et des complexes protéiques.<sup>[62-66]</sup>

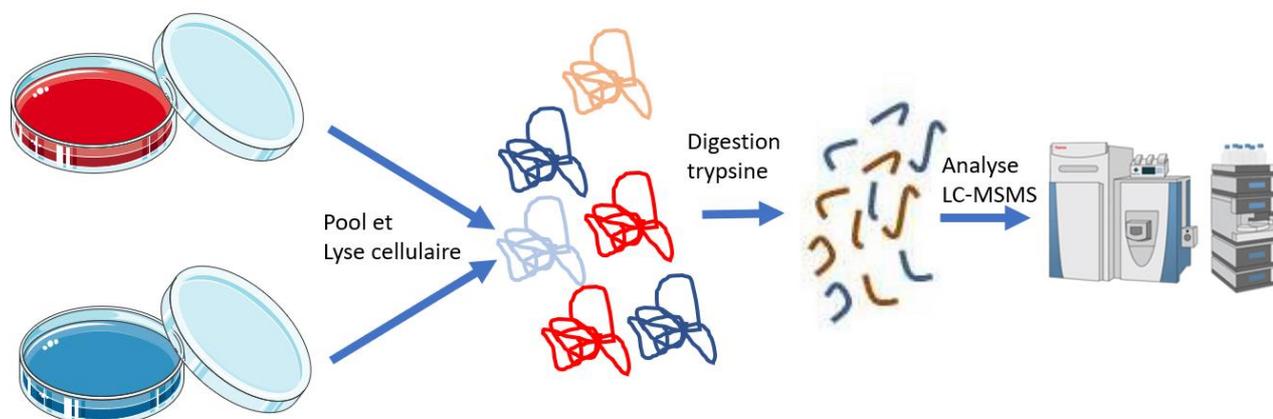
La capacité de la QXL-MS à révéler le changement conformationnel des protéines a été développée sur un petit ensemble de systèmes modèles dont les structures étaient déjà connues. Comme dans les analyses protéomiques quantitatives classiques, il existe deux stratégies principales pour quantifier les peptides : le marquage isotopique et les approches « label free ».

### 5.1.1 L'analyse quantitative des peptides par marquage isotopique

La quantification basée sur le marquage isotopique est appliquée afin de comparer deux états conformationnels. Un des échantillons sera alors marqué à l'aide d'isotopes lourds et l'autre sera marqué par des isotopes légers. L'analyse en MS permet de repérer des paires peptides lourds / légers dont l'abondance relative peut être calculée et reliée à l'abondance du peptide correspondant en solution. Ce marquage isotopique peut être réalisé de différentes manières :

#### 5.1.1.1 Au niveau peptidique :

Le marquage est alors réalisé le plus tôt possible afin de limiter les erreurs de quantification inévitables liées à la manipulation en parallèle des échantillons à comparer. Un des marquages peptidiques les plus utilisés est le marquage par la technique SILAC (Stable Isotope Labelling by Amino Acids in Cell Culture).<sup>[67]</sup> Cette méthode est aujourd'hui considérée comme étant la méthode de marquage la plus stable et la plus performante afin de quantifier les protéines extraites de cellules en culture. Le principe est le suivant : les cellules sont cultivées dans un milieu contenant soit des acides aminés naturels (marquage léger) soit dans un milieu contenant des acides aminés marqués isotopiquement (marquage lourd). Les populations cellulaires sont par la suite mélangées et traitées comme un seul échantillon afin de limiter les erreurs de quantification. Puis, après avoir extrait et digéré les protéines, l'analyse est réalisée classiquement par MS. La quantification est basée sur l'extraction des courants d'ions. Les peptides ainsi marqués auront les mêmes propriétés mais vont différer par leurs masses, ce qui permettra de reconnaître la provenance de chaque peptide.



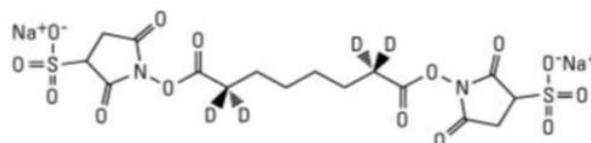
**Figure 101 : Protocole expérimental SILAC**

Cette technique a été utilisée en QXL-MS afin d'étudier les changements dans les structures protéiques et les interactions dans les cellules de carcinome multirésistantes.<sup>[67]</sup>

D'autres études ont également été réalisées en effectuant le marquage des peptides pendant la digestion enzymatique. Ainsi, après pontage chimique, l'un des deux conformères est digéré par la trypsine présente dans de l'eau normale tandis que l'autre est digéré dans  $H_2^{18}O$ , pour ajouter une masse de 8 Da aux peptides. Puis, les deux conformations sont mélangées avant d'être analysées par LC-MS/MS<sup>[68]</sup>.

### 5.1.1.2 L'analyse quantitative par utilisation d'agent pontant marqué

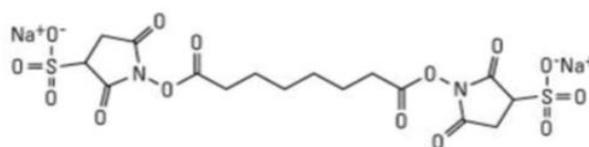
La stratégie alternative la plus utilisée en QXL-MS consiste à marquer l'agent de pontage par des isotopes lourds.<sup>[69-71]</sup> Dans ce cas, les deux états conformationnels sont pontés séparément en utilisant soit l'agent pontant non marqué soit un agent pontant marqué par les isotopes lourds (exemple du BS3-d0 ou -d4 en figure 101).<sup>[72]</sup> Ainsi, seuls les peptides modifiés par l'agent pontant seront marqués et donc, apparaîtront sous forme de doublets sur le spectre de masse. Outre l'aspect quantitatif, cette approche permet de repérer rapidement les produits marqués par le cross-linker mais elle est coûteuse et requiert la synthèse de l'agent pontant marqué (pas toujours disponible commercialement).



**BS3-d4**

MW : 576,45

Mass agent pontant : 142,09



**BS3-d0**

MW : 572,43

Mass agent pontant : 138,07

**Figure 102 : Exemple d'agent pontant marqué isotopiquement et disponible commercialement.**

Une limitation de cette approche de marquage isotopique est qu'elle ne s'applique que pour la comparaison d'un nombre limité d'échantillons (souvent deux) et qu'elle requiert une séparation des conformations des protéines les plus stables avant la réaction de pontage.

### 5.1.2 La quantification label free

La quantification « label free » (sans marquage) peut être réalisée à partir de n'importe quels échantillons et n'importe quels agents pontant. Les échantillons sont traités en parallèle mais en suivant des procédures strictement identiques. Pour chaque peptide ponté d'intérêt, son abondance est comparée par la mesure des intensités de signal à travers les acquisitions individuelles de MS.<sup>[73, 74]</sup> Par conséquent, la quantification label free peut être menée sur plusieurs échantillons en parallèle et ne nécessite pas nécessairement une pré-séparation des conformères avant le pontage, du moment qu'ils peuvent être distingués facilement après pontage chimique.

Deux approches de quantification label free sont décrites dans la littérature<sup>[75]</sup> :

- La quantification par comptage spectral (Spectral Count) qui repose sur le comptage du nombre de fois que tous les peptides correspondant à une protéine ont été séquencés.<sup>[76-78]</sup> Plus la protéine est abondante, plus le nombre de peptides tryptiques disponibles pour le séquençage est élevé, ce qui entraîne plus d'événements MS / MS, appelés comptages spectraux. Néanmoins cette approche est difficilement applicable en XL-MS car les peptides cross-linkés sont souvent les moins intenses et ils risquent d'être sélectionnés dans une analyse et pas dans une autre, sans que cela ne doive être interprété comme reflétant un changement conformationnel.

- La quantification selon l'intensité du signal spectrométrique en MS (XIC).<sup>[79-81]</sup> Cette méthode repose sur la mesure de l'espace tridimensionnel de l'intensité des ions des peptides, de leur m/z et du temps d'éluion chromatographique. Cette technique est la plus utilisée car elle est plus précise et adaptée pour mesurer les abondances relatives de protéines d'abondance moyenne contrairement au spectral count.

### **5.1.3 Développement d'outils informatique**

À ce jour, les applications de QXL-MS ont été limitées aux quelques laboratoires pionniers qui ont développé la méthode. Une raison majeure qui empêche la QXL-MS de trouver une application plus large est le manque d'outils de traitement des données.

Les intensités des peptides pontés sont obtenues sur la base de leurs chromatogrammes ioniques extraits (XIC). Ces XIC doivent alors être générés manuellement pour l'ensemble des peptides pontés identifiés. Ce qui prend beaucoup de temps et devient ingérable lorsqu'il s'agit d'ensembles de données complexes et volumineuses.

Encore aujourd'hui, des efforts considérables sont consacrés à l'automatisation du traitement des données QXL-MS. De nouveaux logiciels permettant l'analyse quantitative ont ainsi vu le jour tel que Xtract<sup>[65]</sup> qui permet de quantifier les données issues de XQuest/XProphet<sup>[7]</sup>

Le développement de logiciels assurant l'analyse quantitative des peptides pontés identifiés est réalisé selon deux approches :

- La première approche consiste à développer des outils dédiés afin de quantifier les données de pontage tels que XiQ.<sup>[71]</sup>
- La seconde approche vise à exploiter des logiciels de protéomique quantitative déjà établis, dont certains ne sont pas prévus pour analyser les données obtenues par XL-MS comme par exemple les logiciels Pinpoint<sup>[64]</sup> (Thermo Fisher Scientific) et les logiciels open source tel que MaxQuant<sup>[71]</sup> ou Skyline<sup>[82]</sup>.

## **5.2 Mise en place d'un protocole expérimental pour l'analyse quantitative des peptides pontés.**

Alors que les résultats obtenus sur le complexe RAR/RXR en présence d'ADN semblent suggérer la coexistence de plusieurs conformations en solution, nous avons décidé à la fin de ma thèse de débiter la mise en place d'un protocole visant à l'analyse quantitative des peptides pontés.

Pour cela, nous avons choisi une stratégie basée sur l'utilisation d'agents pontants marqués. Le BS3 ayant présenté un bon rendement de pontage aussi bien sur l'hétérodimère seul ou complexé à l'ADN, nous avons tout naturellement utilisé du BS3 marqué au deutérium sur son bras espaceur pour cette analyse quantitative. Nous nous sommes basés en 1<sup>ère</sup> intention sur le protocole expérimental détaillé dans la publication du Dr. Rappsilber dans le journal *nature protocols* de 2019 avec l'ultime objectif de caractériser efficacement les changements structuraux au sein du complexe RAR/RXR lors de sa liaison avec l'ADN.<sup>[68]</sup> La purification du complexe RAR/RXR/ADN n'étant pas encore stabilisée, nous ne décrivons ici que la mise en place du protocole expérimental.

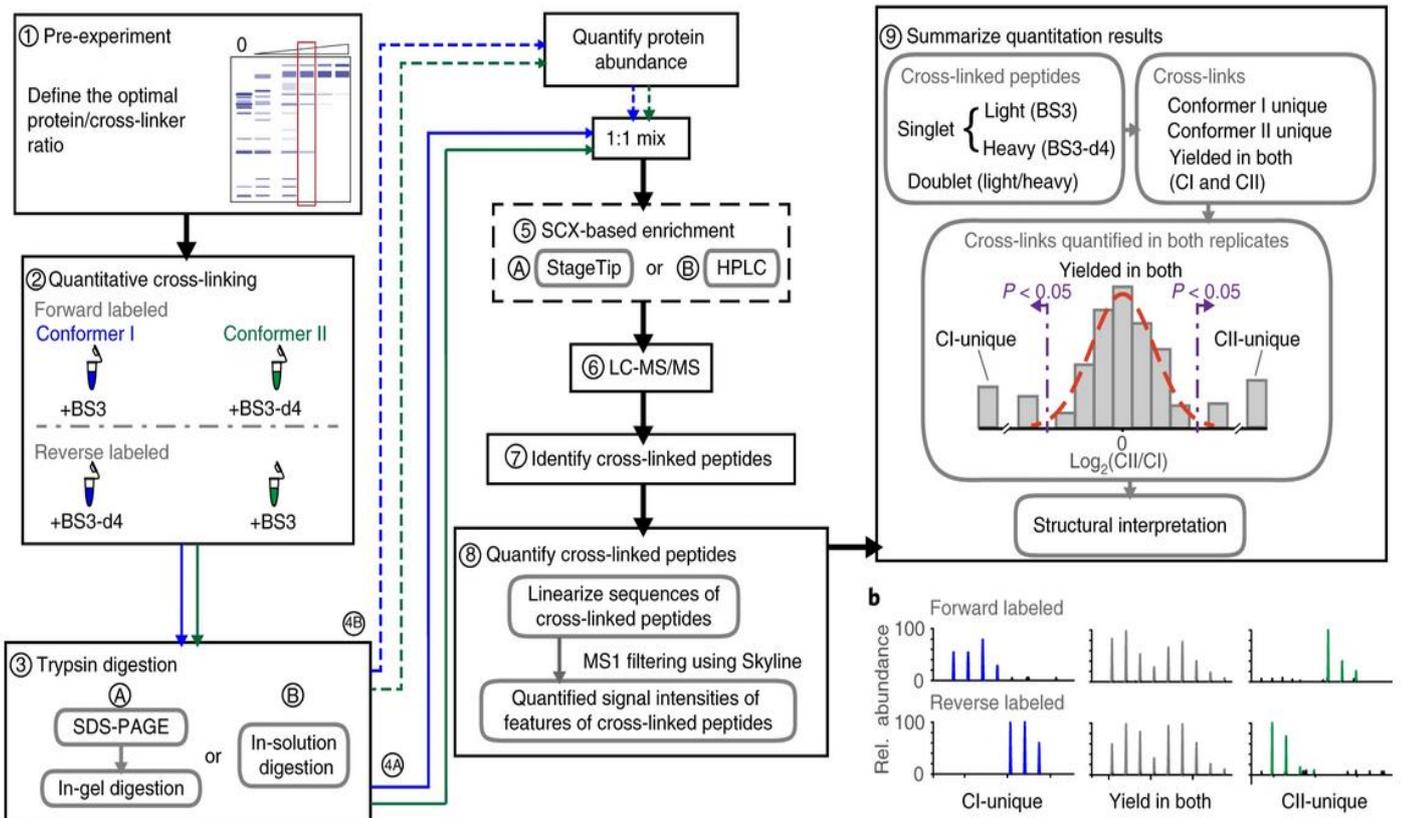


Figure 103 : Protocole expérimental de QXL-MS. Extrait de Chen et al 2015.

Le protocole expérimental utilisé en QXL-MS peut être ainsi divisé en cinq modules :

Ce protocole vise à se standardiser en 6 étapes :

- La préexpérience,
- Le pontage chimique quantitatif
- La préparation d'échantillons pour la LC-MS
- L'acquisition de données LC-MS/MS
- L'identification des peptides pontés
- La quantification des peptides pontés.

### **5.2.1 La préexpérience**

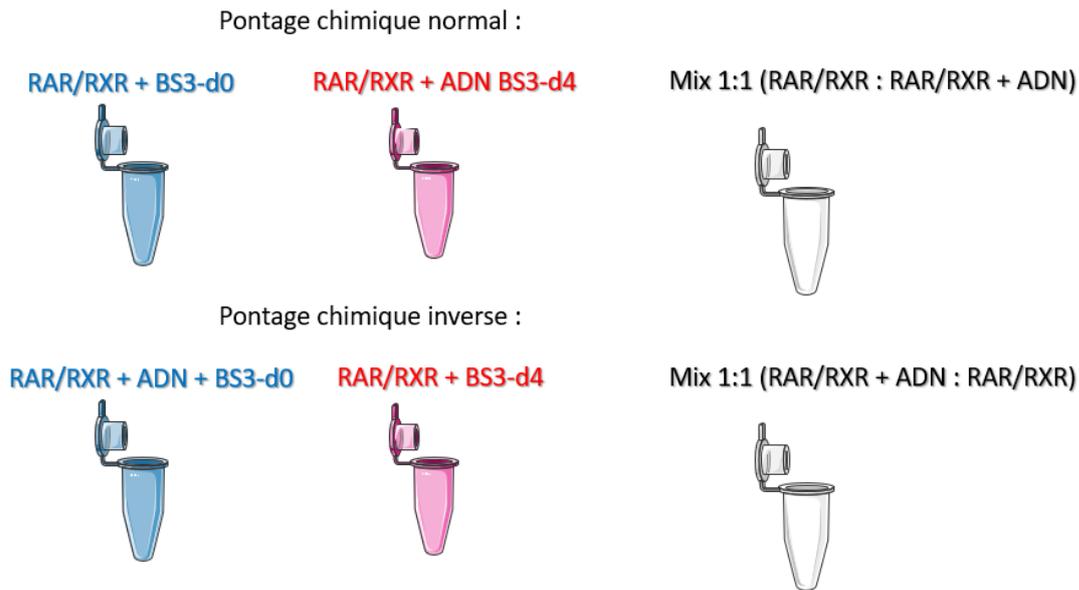
La préexpérience est une procédure clé pour l'identification des zones impactées par le changement conformationnel. Celle-ci est menée pour déterminer les conditions optimales de pontage. Cette préexpérience a déjà été réalisée sur le complexe RAR/RXR et nous a permis d'identifier les conditions optimales aussi bien en amont de l'analyse LC-MS/MS qu'au niveau de l'interprétation des spectres MS/MS de peptides pontés. La réaction de pontage a ainsi été réalisée à température ambiante pendant 30min à pH 7.5 dans un tampon HEPES à 25mM + NaCl 150mM.

### **5.2.2 Le pontage chimique pour l'analyse quantitative**

Le protocole expérimental utilisé pour effectuer la quantification est strictement identique à celui de la préexpérience, seul l'agent pontant différera entre les 2 échantillons quantifiés : l'un des échantillons sera traité par un cross-linker léger et l'autre échantillon par l'agent pontant correspondant lourd, c'est-à-dire, marqué isotopiquement. Dans notre cas, nous avons utilisé le BS3-D0 comme agent pontant léger et le BS3-D4 comme agent pontant lourd (cf. figure 102). Le marquage apporte une différence de masse de 4,025 Da qui permettra de repérer l'échantillon origine du peptide ponté identifié.

Ainsi l'agent léger (BS3-D0) a été ajouté pour ponter l'hétérodimère RAR/RXR et l'agent lourd (BS3-D4) pour ponter le complexe RAR/RXR+DR5. Afin de valider la quantification et de vérifier que l'agent de pontage marqué n'induit pas de biais conformationnel, une étape de pontage chimique inversé sera effectuée en parallèle.<sup>[64, 68]</sup> Ainsi, un second échantillon de RAR/RXR a été ponté avec l'agent lourd et le RAR/RXR+DR5 avec l'agent pontant léger. Les études antérieures ont montré que cette procédure est essentielle pour une quantification précise.<sup>[83]</sup>

Les complexes marqués par les cross-linkers légers et lourds sont ensuite mélangés à un rapport de 1:1, puis traités par une approche protéomique classique comme un seul échantillon. La détection de signaux sous forme de doublets séparés de 4,025 Da au niveau des ions précurseurs caractéristiques du marquage isotopique a un double avantage : (i) ils confirment l'identification des produits de pontage et (ii) ils permettent la quantification relative des peptides pontés pour chacun des conformères.<sup>[70]</sup> Les peptides pontés présents dans les deux conformères sont observés comme des signaux de doublet dans les deux répliqués, tandis que les peptides pontés qui sont uniques à l'un ou l'autre des conformères seront observés comme des signaux singulets avec un déplacement de masse relié au marquage isotopique du cross-linker dont il dérive.<sup>[64, 68, 70]</sup>



**Figure 104 : Exemple de marquage isotopique réalisé pour l'analyse quantitative.**

Nous obtenons alors 6 échantillons pontés : 3 pour le marquage isotopique standard et 3 par marquage inverse. Nous avons choisi d'analyser également les échantillons pontés par le BS3-D0 et BS3-D4 séparément car ils apportent une dimension de reproductibilité supplémentaire dans l'identification des peptides pontés d'intérêt.

### **5.2.3 Préparation des échantillons et analyses LC-MSMS**

#### 5.2.3.1 Préparation des échantillons

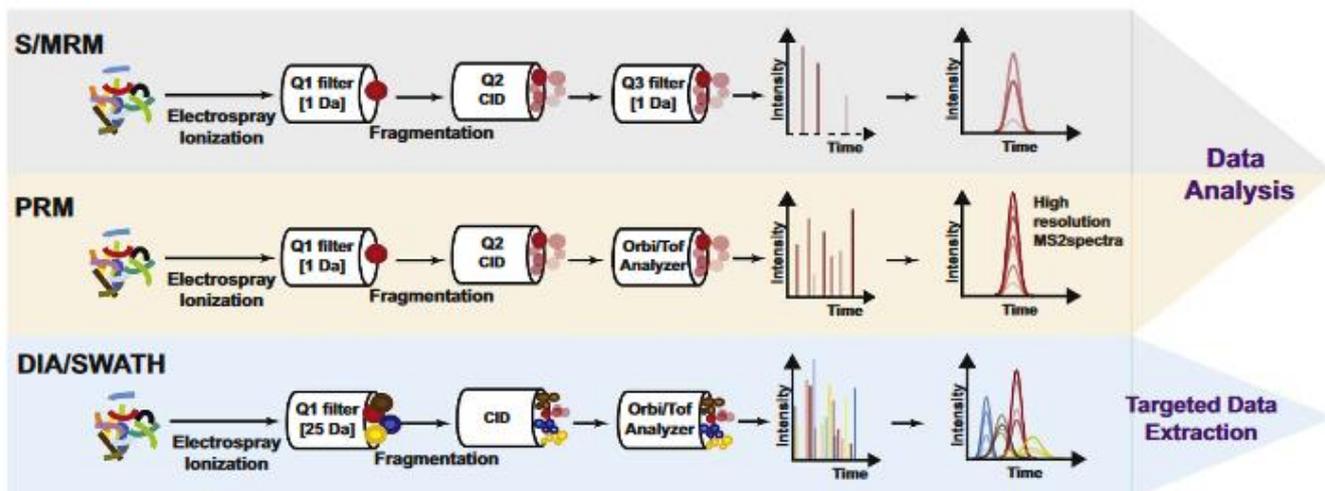
L'ensemble des échantillons sont digérés par la trypsine en utilisant un ratio de 1 :5 (protéine : enzyme) afin de digérer de manière efficace le complexe. L'ensemble des étapes de traitement de l'échantillon sont réalisées en parallèle et en faisant attention à bien respecter les bonnes proportions afin de limiter au maximum les erreurs de quantification sur le complexe.

#### 5.2.3.2 Analyse LC-MSMS : importance de la haute résolution pour l'analyse quantitative

Une fois les échantillons digérés, les analyses sont réalisées sur un système LC-MSMS à haute résolution de type Orbitrap. Les données MS/MS peuvent être acquises selon deux modes :

- Soit en utilisant le mode « Data Dependant Acquisition » (DDA)
- Soit en utilisant le mode « Data Independent Acquisition » (DIA)

En mode DDA, le spectromètre de masse sélectionne dans un premier temps les ions peptidiques les plus intenses puis ils sont fragmentés et analysés dans un second temps. Il s'agit de la méthode la plus classiquement utilisée en protéomique classique. Sinon, en mode DIA, pour chaque cycle, l'instrument se concentre sur une fenêtre de masse étroite de précurseurs et acquiert des données MS / MS de tous les précurseurs détectés dans cette fenêtre. Cette fenêtre de masse est ensuite échelonnée sur toute la plage de masse, collectant systématiquement les données MS / MS de tous les précurseurs détectés. En conséquence, l'acquisition DIA offre potentiellement une analyse plus complète des peptides que la DDA et peut en principe donner des spectres de masse en tandem de toutes les molécules ionisées après leur conversion en phase gazeuse.<sup>[84, 85]</sup>



**Figure 105 : Schéma explicative sur les principaux modes d'acquisition MSMS utilisé en protéomique. Extrait de Aebersold 2018.[84]**

Pour notre étude, nous avons décidé de travailler avec le mode DDA et avec une acquisition des spectres MS1 à haute résolution (Top 8). En effet, la modification de notre méthode d'acquisition MS devrait permettre de compenser la faible abondance des ions cross-linkés en se concentrant sur la sélection des ions précurseurs avec une charge  $\geq 3+$ .

#### **5.2.4 Identification des peptides pontés**

L'identification des peptides pontés reste une étape incontournable du protocole QXL-MS. Celle-ci est réalisée de manière indépendante par rapport à l'analyse quantitative.

Nous avons donc appliqué la méthodologie d'identification et de validation manuelle précédemment optimisée lors de l'étude qualitative en utilisant les mêmes logiciels (XiSearch<sup>[38]</sup>, XlinkX<sup>[40]</sup>, MeroX<sup>[35]</sup> et pLink<sup>[13]</sup>). Les résultats ont été validés manuellement de manière indépendante pour chaque logiciel avant d'être comparés. Pour l'analyse quantitative, nous avons choisi de ne considérer que les peptides pontés identifiés et validés manuellement par au moins deux logiciels (score arbitraire de 100).

### **5.2.5 Analyse quantitative des peptides pontés : Utilisation de Skyline**

Une des difficultés des études quantitatives des peptides pontés est le manque de logiciels permettant d'effectuer l'analyse quantitative de manière automatisée. Les logiciels utilisent des listes caractéristiques spécifiques pour effectuer l'analyse quantitative des peptides. Ces caractéristiques sont la valeur m/z de l'ion parent, la charge et le temps de rétention qui ont permis l'identification du peptide.<sup>[86]</sup>

Les logiciels permettant l'analyse quantitative acceptent des listes de caractéristiques observées sur les spectres MSMS des peptides pontés contenant la valeur m/z de l'ion parent identifié, la charge et le temps de rétention comme entrée. Cependant, les logiciels ont souvent deux limites majeures.<sup>[82]</sup>

Premièrement, la quantification d'un peptide ponté marqué isotopiquement doit être réalisée sur l'ensemble des réplicats.

Deuxièmement, la majorité des logiciels ne fournissent pas une interface permettant de visualiser les résultats de quantification.

Seulement deux logiciels répondent à ces différents critères : Skyline et Pinpoint. Dans ce protocole, nous avons utilisé Skyline car il s'agit d'un logiciel open source gratuit et qui possède plusieurs tutoriels détaillés proposés par les développeurs.

En paramètre d'entrée, Skyline requiert les séquences protéiques afin de calculer la masse des peptides. Cependant, celui-ci ne permet pas encore de calculer la masse des peptides pontés. C'est pourquoi il est nécessaire de convertir la séquence de ces peptides pontés en une forme linéaire non pontée qui a une masse identique et conserve les informations sur les sites de pontage. La linéarisation de la séquence des peptides pontés est effectuée simplement en utilisant un tableur tel que Microsoft Excel en amont de la quantification par Skyline. Les peptides linéarisés sont ensuite répertoriés dans un fichier texte contenant la séquence du peptide, le nom du fichier MS où il a été identifié, le numéro de balayage du spectre MS2 et la charge du peptide<sup>[41]</sup>

Ces informations permettent à Skyline de construire une bibliothèque de peptides pontés et de suivre leurs signaux au sein de l'analyse LC-MS/MS. La séquence ainsi que les états de charge identifiés sont utilisés pour calculer les valeurs « m/z » de la paire de peptides pontés, et le numéro de balayage du spectre MS2 définit le temps de rétention. Une valeur de m/z associée à un temps de rétention sert de coordonnée à Skyline pour identifier le signal d'une paire de peptides pontés à un état de charge donné (souvent appelé « caractéristique »). L'intensité du signal d'une caractéristique est mesurée comme l'aire de son pic chromatographique d'ion. L'intensité du signal d'une paire de peptides pontée est la somme des intensités de signal de toutes ses caractéristiques détectées.<sup>[87]</sup>

Lors de la quantification à l'aide d'agents de pontage marqués par des isotopes, les identifications des versions légère et lourde de chaque caractéristique doivent être incluses dans la bibliothèque d'entrée. Bien qu'une quantité estimée d'échantillon égale devrait être utilisée pour l'analyse LC-MS/MS pour chaque conforme, il est nécessaire de normaliser les intensités de signal des paires de peptides réticulés sur la base de l'abondance des protéines dans chaque échantillon.<sup>[38]</sup>

### **5.3 Quantification des résultats obtenus pour RAR/RXR**

Nous avons ainsi analysé les 6 échantillons séparément (Cf figure 104). L'interprétation des résultats nous a permis de détecter 40 peptides uniques pour la conformation RAR/RXR et 38 peptides uniques pour la conformation RAR/RXR + ADN pour les échantillons marqués dans le sens normal. Puis, nous avons linéarisé les peptides avant d'effectuer l'analyse quantitative par Skyline en appliquant le protocole mis en place dans la publication de J. Rappsilber.<sup>[68]</sup>

Une bibliothèque distincte doit être construite pour chaque agent pontant. Nous avons ainsi réalisé une première bibliothèque pour les peptides identifiés avec le BS3-D0 et une seconde avec le BS3-D4. Puis, deux autres bibliothèques pour le marquage inverse. A partir des paramètres d'entrée présents dans le logiciel, une quantification automatisée est réalisée.

Comme pour tout autre logiciel, de l'automatisation résulte des erreurs, principalement liées aux signaux MS1 mal isolés. Cela se produit plus souvent avec des signaux de faible intensité. Pour cela, une étape essentielle de la quantification consiste à inspecter et corriger manuellement les signaux peptidiques isolés pour la quantification.

Une fois la correction manuelle réalisée, les données sont exportées au format Excel où le quotient entre le signal du peptide léger et le signal du peptide lourd est calculé pour l'ensemble des peptides. Si le résultat est égal à 0, nous aurons un peptide correspondant à un conformère unique. Si ce résultat est différent de 0, celui-ci sera considéré comme un doublet et sera ainsi détecté dans les deux conformations. Ce calcul est réalisé pour le marquage normal et le marquage inverse. Les données sont par la suite combinées et les peptides observés uniquement dans un seul sens de marquage sont supprimés car ils possèdent une incertitude dans leur identification et dans leur quantification.

L'interprétation des résultats ne fait que commencer et à ce jour, nous nous sommes uniquement intéressés aux deux échantillons notés Mix (figure 104), c'est-à-dire les deux répliquats qui représentent le pontage isotopique soit sur l'hétérodimère/ADN soit inversement sur l'hétérodimère seul. Les peptides singulets du conformère I, c'est-à-dire uniquement observé pour RAR/RXR ou du conformère II (RAR/RXR+ADN) doivent être systématiquement détectés et quantifiés dans les deux répliquats. S'ils n'ont été quantifiés que dans un seul, le résultat ne sera pas considéré comme fiable et le peptide sera supprimé du tableau de résultat. Une moyenne de l'ensemble des répliquats est ensuite calculée sur l'ensemble des peptides quantifiés comme « doublet »

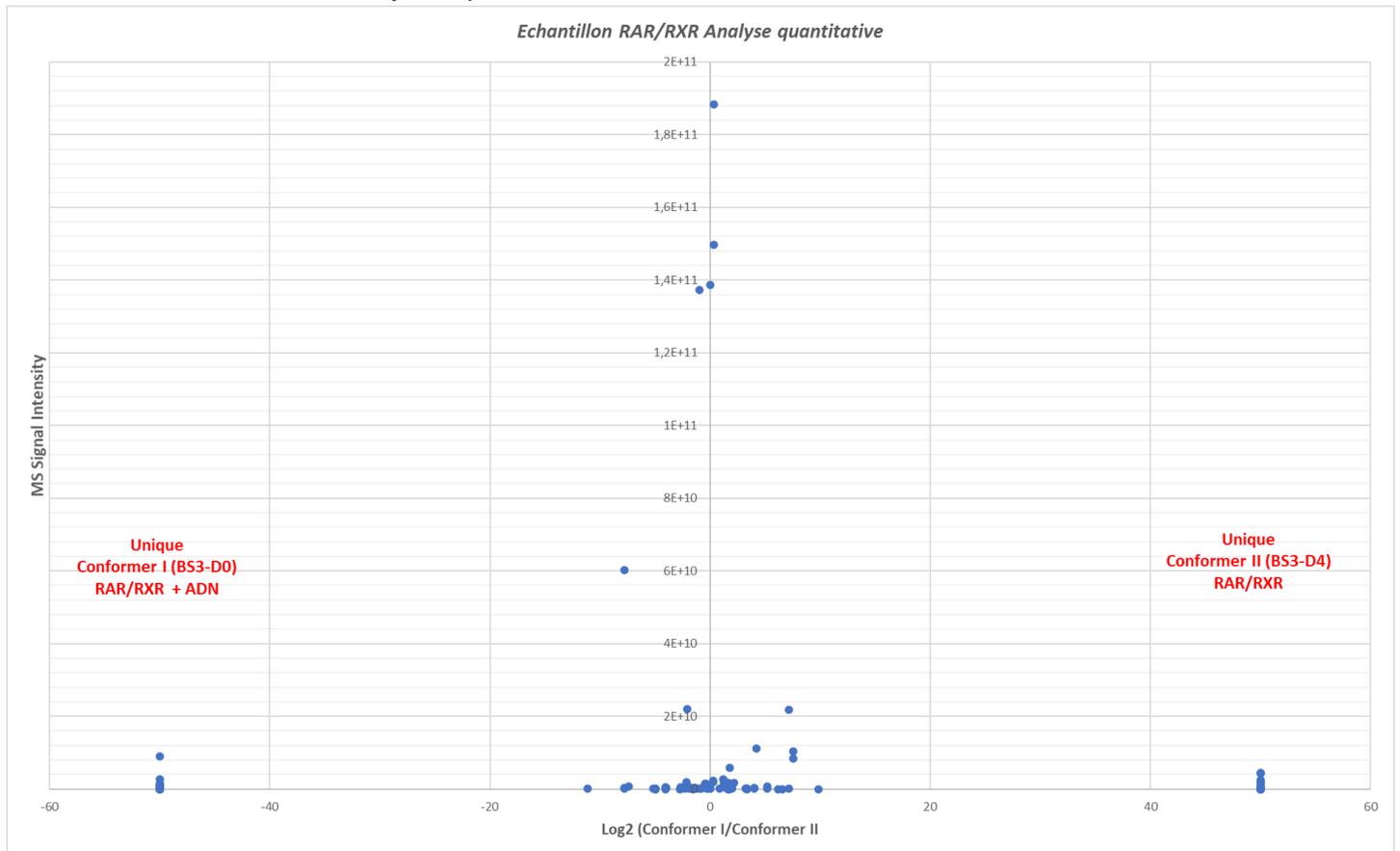
Les liaisons pontées uniquement détectées dans l'hétérodimère (avec le cross-linker D0) ou uniquement détectées dans le complexe avec ADN (avec le cross-linker D4) sont interprétées comme des différences de conformation. La différence conformationnelle peut également être reflétée comme des changements majeurs de signal entre les deux conformères. En revanche, les pontages issus des régions qui ont les mêmes caractéristiques structurales dans les deux conformères donnent lieu à un rapport conformère I/conformère II proche de 1. Dans certaines études, un seuil arbitraire est mis en place pour indiquer si une évolution de structure était détectée pour ces peptides.

Pour modéliser la distribution des valeurs des rapports de signaux conforme I/ conforme II, un graphique est réalisé à partir du logarithme du rapport de signaux Log2 (conforme I/ conforme II) des peptides pontés par rapport à l'intensité du signal. Cette modélisation mettra ainsi en valeur les données et définira le seuil pour les valeurs significatives.<sup>[88]</sup> (cf. figure 106)

**Tableau 15 : Liste des peptides pontés quantifiés dans l'ensemble des échantillons et observés comme singulet (RAR/RXR+ADN unique et RAR/RXR unique) et comme doublet (RAR/RXR+ADN préférentielle et RAR/RXR préférentielle)**

<b>RAR/RXR + ADN unique</b>	<b>Protéine1</b>	<b>Position1</b>	<b>Protéine2</b>	<b>Position2</b>
CFEVGMSKESVR-AIVLFNPDSKGLSNPAEVEALR	RAR	155	RXR	381
CFEVGMSKESVR-HKYPEQPGR	RAR	155	RXR	407
DKNCIINK-GFFKR	RAR	128	RXR	160
KKEVPK-VLTELVS KM RDMQMDKTELGLR	RAR	167	RXR	356
KKEVPK-VLTELVS KM RDMQMDKTELGLR	RAR	167	RXR	364
SIQKNMVYTCR-QRNRQYCRYQKCLAMGMK	RAR	118	RXR	194
KAHQETFPALCQLGK-YQKCLAMGMK	RAR	207	RXR	201
KKEVPK-DMQMDKTELGLR	RAR	167	RXR	356
PHMFPMMLMKITDLRSISAKGAER-AIVLFNPDSK	RAR	381	RXR	381
RPSRPHMFPMMLM-KDLTYTCR	RAR	377	RXR	165
<b>RAR/RXR unique</b>	<b>Protéine1</b>	<b>Position1</b>	<b>Protéine2</b>	<b>Position2</b>
AHQETFPALCQLGKYTTNNSSEQR-DNKDCLIDKR	RAR	207	RXR	175
DKNCIINKVTR-FAKLLLR	RAR	134	RXR	417
EVPKPECSSESYLTPEVGELIEK-AIVLFNPDSKGLSNPAEVEALR	RAR	171	RXR	381
KAHQETFPALCQLGK-DNKDCLIDKR	RAR	193	RXR	175
KEVPKPECSSESYLTPEVGELIEK-AIVLFNPDSKGLSNPAEVEALR	RAR	167	RXR	381
SIQKNMVYTCR-EKVYASLEAYCK	RAR	118	RXR	395
KKEVPK-YQKCLAMGMKR	RAR	167	RXR	194
<b>RAR/RXR + ADN Préférentielle</b>	<b>Protéine1</b>	<b>Position1</b>	<b>Protéine2</b>	<b>Position2</b>
SIQKNMVYTCR-DCLIDKR	RAR	118	RXR	181
DKNCIINK-KDLTYTCR	RAR	128	RXR	165
MLMKITDLRSISAKGAER-GLSNPAEVEALREK	RAR	390	RXR	395
DKNCIINKVTR-DCLIDKR	RAR	128	RXR	181
SISAKGAER-SIGLKCLEHLFFFK	RAR	390	RXR	431
<b>RAR/RXR Préférentielle</b>	<b>Protéine1</b>	<b>Position1</b>	<b>Protéine2</b>	<b>Position2</b>
KAHQETFPALCQLGK-KDLTYTCR	RAR	194	RXR	165
DKNCIINKVTR-DNKDCLIDKR	RAR	128	RXR	181
SISAKGAER-DNKDCLIDKR	RAR	390	RXR	175
AHQETFPALCQLGKYTTNNSSEQR-DNKDCLIDKR	RAR	207	RXR	181
SIQKNMVYTCR-DMQMDKTELGLR	RAR	118	RXR	364
SISAKGAER-DCLIDKR	RAR	390	RXR	181

Ainsi, après quantification et après avoir réassemblé l'ensemble des résultats, nous avons modélisé la distribution des valeurs. Celle-ci nous a ainsi permis de visualiser directement les peptides uniques observés dans une seule conformation et les peptides identifiés dans les deux complexes. Après quantification, nous avons ainsi observé 10 peptides uniques pour le complexe RAR/RXR liée à l'ADN, 7 peptides uniques détectés dans le complexe RAR/RXR. Enfin, 38 peptides ont été observés dans les deux conformations et correspondent à ceux détectés lors des analyses qualitatives.



**Figure 106 : Schéma montrant la distribution des valeurs des rapports de signaux pour le marquage inverse montrant la liste des peptides unique détectée pour chaque conformation.**

#### **5.4 Conclusion : L'analyse quantitative des peptides pontés**

Ainsi, nous sommes parvenus à mettre en place un protocole permettant de procéder à la quantification de peptides pontés par l'utilisation d'un cross-linker marqué isotopiquement, le BS3.

Néanmoins, Suite au problème de purification de l'échantillons, il est impossible de conclure sur la véracité biologique des résultats. Des études supplémentaires devront ainsi être effectuées dès que les problèmes rencontrés lors de la préparation de l'échantillon seront résolus.

La mise en place d'un protocole expérimental pour l'analyse quantitative des peptides pontés semble être une méthode indispensable afin d'identifier les changements conformationnels des structures. L'utilisation d'agents pontant marqués isotopiquement s'est avérée efficace afin de caractériser rapidement les peptides pontés observés dans les différentes structures. Toutefois, cette méthode comporte quelques limites. En effet, il existe très peu d'agent pontant marqué disponible commercialement ce qui peut limiter cette approche. De plus, la quantification par marquage isotopique fonctionne uniquement si deux états conformationnel sont détectés pour un complexe. Pour la suite du projet, il serait intéressant d'effectuer de l'analyse quantitative par label free afin de comparer les résultats obtenus avec et sans marquage isotopique.

Une deuxième difficulté rencontrée dans ce protocole est l'utilisation de logiciel non optimisé pour l'analyse quantitative des peptides pontés. En effet, durant ma thèse, très peu de logiciels dédiés à l'analyse quantitative des peptides pontés étaient disponibles. L'outil Skyline étant uniquement dédié pour l'analyse quantitative des peptides classiques, il a fallu convertir l'ensemble des peptides pontés en peptides linéaires en tenant compte des modifications présentes sur les deux peptides. Aujourd'hui, une première version de Skyline permettant la quantification des peptides pontés est en cours de développement (version 20.2).

## 6. Conclusion et Perspectives

En conclusion, nous avons au cours de ma thèse réussi à mettre en place un protocole analytique pour la caractérisation de peptides pontés par MS.

Un protocole XL-MS robuste et reproductible a ainsi été établi et une étude comparative systématique de plusieurs logiciels d'interprétation des données de XL-MS nous a permis de faire ressortir les paramètres importants à une identification la plus probable possible des acides aminés pontés. Les données XL-MS ont ensuite été comparées aux données SAXS pour aider à affiner un modèle 3D plus précis. Une étude de reproductibilité biologique a également permis de mettre en évidence qu'il est crucial de contrôler la « qualité de complexation » des échantillons utilisés pour l'analyse, c'est-à-dire vérifier que nous travaillons sur des échantillons parfaitement similaires d'une expérience à l'autre.<sup>[50, 51, 89]</sup>

La résolution de cette approche XL-MS n'est pas facile à évaluer, mais cette approche est très prometteuse en interaction avec d'autres techniques d'analyses structurales (notamment la cryo-EM ou les approches théoriques par modélisation moléculaire). En effet, cette technique est non limitante en termes de taille de complexes et grâce à un protocole d'analyse en solution, elle permet d'appréhender l'étude de changement conformationnel.<sup>[90]</sup>

## 7. Bibliographie

1. Merkley, E.D., J.R. Cort, and J.N. Adkins, *Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze*. Journal of structural and functional genomics, 2013. **14**(3): p. 77-90.
2. Nguyen-Huynh, N.-T., et al., *Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry*. Biophysical Chemistry, 2016. **210**: p. 2-8.
3. Balmer, J.E. and R. Blomhoff, *A robust characterization of retinoic acid response elements based on a comparison of sites in three species*. The Journal of Steroid Biochemistry and Molecular Biology, 2005. **96**(5): p. 347-354.
4. Moutier, E., et al., *Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology*. Journal of Biological Chemistry, 2012. **287**(31): p. 26328-26341.
5. O'Reilly, F.J. and J. Rappsilber, *Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology*. Nature Structural & Molecular Biology, 2018. **25**(11): p. 1000-1008.
6. Huang, B.X., H.-Y. Kim, and C. Dass, *Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry*. Journal of the American Society for Mass Spectrometry, 2004. **15**(8): p. 1237-1247.
7. Leitner, A., T. Walzthoeni, and R. Aebersold, *Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline*. Nature Protocols, 2014. **9**(1): p. 120-137.
8. Chu, F., D.T. Thornton, and H.T. Nguyen, *Chemical cross-linking in the structural analysis of protein assemblies*. Methods (San Diego, Calif.), 2018. **144**: p. 53-63.
9. Leitner, A., et al., *Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines*. Trends in Biochemical Sciences, 2016. **41**: p. 20-32.
10. Nguyen-Huynh, N.-T., et al., *Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex*. Protein science : a publication of the Protein Society, 2015. **24**(8): p. 1232-1246.
11. Stengel, F., R. Aebersold, and C.V. Robinson, *Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes*. Molecular & cellular proteomics : MCP, 2012. **11**(3): p. R111.014027-R111.014027.
12. Yang, B., et al., *Identification of cross-linked peptides from complex samples*. Nature Methods, 2012. **9**(9): p. 904-906.
13. Chen, Z.-L., et al., *A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides*. Nature Communications, 2019. **10**(1): p. 1-12.
14. Iacobucci, C. and A. Sinz, *To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry*. Analytical Chemistry, 2017. **89**(15): p. 7832-7835.
15. Yu, C. and L. Huang, *Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology*. Analytical chemistry, 2018. **90**(1): p. 144-165.
16. Merkley, E.D., et al., *Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances*. Protein science : a publication of the Protein Society, 2014. **23**(6): p. 747-759.
17. Bujacz, A., *Structures of bovine, equine and leporine serum albumin*. Acta Crystallographica Section D, 2012. **68**(10): p. 1278-1289.
18. Majorek, K.A., et al., *Structural and immunologic characterization of bovine, horse, and rabbit serum albumins*. Molecular immunology, 2012. **52**(3-4): p. 174-182.

19. Graham, M.J., et al., *xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data*. bioRxiv, 2019: p. 561829.
20. Kosinski, J., et al., *Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures*. Journal of Structural Biology, 2015. **189**(3): p. 177-183.
21. Riffle, M., et al., *ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data*. Journal of Proteome Research, 2016. **15**(8): p. 2863-2870.
22. Fritzsche, R., et al., *Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis*. Rapid Communications in Mass Spectrometry, 2012. **26**(6): p. 653-658.
23. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* Angewandte Chemie International Edition, 2018. **57**(22): p. 6390-6396.
24. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes*. Journal of Structural Biology, 2011. **173**: p. 530-540.
25. Leitner, A., et al., *Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography*. Molecular & cellular proteomics : MCP, 2012. **11**(3): p. M111.014126-M111.014126.
26. Schmidt, R. and A. Sinz, *Improved single-step enrichment methods of cross-linked products for protein structure analysis and protein interaction mapping*. Analytical and Bioanalytical Chemistry, 2017. **409**(9): p. 2393-2400.
27. Buncherd, H., et al., *Selective enrichment and identification of cross-linked peptides to study 3-D structures of protein complexes by mass spectrometry*. Journal of Proteomics, 2012. **75**(7): p. 2205-2215.
28. Kalkhof, S. and A. Sinz, *Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters*. Analytical and Bioanalytical Chemistry, 2008. **392**(1): p. 305-312.
29. Swaim, C.L., J.B. Smith, and D.L. Smith, *Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides*. Journal of the American Society for Mass Spectrometry, 2004. **15**(5): p. 736-749.
30. Leavell, M.D., et al., *Strategy for selective chemical cross-linking of tyrosine and lysine residues*. Journal of the American Society for Mass Spectrometry, 2004. **15**(11): p. 1604-1611.
31. Chen, F., et al., *On the Efficiency of NHS Ester Cross-Linkers for Stabilizing Integral Membrane Protein Complexes*. Journal of The American Society for Mass Spectrometry, 2015. **26**(3): p. 493-498.
32. Chen, Y., et al., *Structural Characterization of Cross-Linked Species in Trastuzumab Emtansine (Kadcyla)*. Bioconjugate Chemistry, 2016. **27**(9): p. 2037-2047.
33. Hage, C., et al., *The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis*. Angewandte Chemie-International Edition, 2017. **56**(46): p. 14551-14555.
34. Muller, M.Q., et al., *Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS*. Analytical Chemistry, 2010. **82**(16): p. 6958-6968.
35. Iacobucci, C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions*. Nature Protocols, 2018. **13**(12): p. 2864-2889.
36. Gotze, M., et al., *Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis*. Journal of the American Society for Mass Spectrometry, 2015. **26**(1): p. 83-97.
37. Tran, B.Q., D.R. Goodlett, and Y.A. Goo, *Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics*. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2016. **1864**: p. 123-129.

38. Mendes, M.L., et al., *An integrated workflow for crosslinking mass spectrometry*. *Molecular systems biology*, 2019. **15**(9): p. e8994-e8994.
39. Götze, M., et al., *StavroX--a software for analyzing crosslinked products in protein interaction studies*. *Journal of the American Society for Mass Spectrometry*, 2012. **23**(1) : p. 76-87.
40. Liu, F., et al., *Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification*. *Nature Communications*, 2017. **8**(1): p. 15473.
41. Muller, F., et al., *On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry*. *Journal of the American Society for Mass Spectrometry*, 2018. **29**(2): p. 405-412.
42. Chandra, V., et al., *The quaternary architecture of RAR $\beta$ -RXR $\alpha$  heterodimer facilitates domain-domain signal transmission*. *Nature Communications*, 2017. **8**(1): p. 868.
43. Chatagnon, A., et al., *RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements*. *Nucleic acids research*, 2015. **43**(10): p. 4833-4854.
44. Leid, M., P. Kastner P Fau - Chambon, and P. Chambon, *Multiplicity generates diversity in the retinoic acid signalling pathways*. *Trends in Biochemical Sciences*, 1992. **17**(10): p. 427-433.
45. Bastien, J. and C. Rochette-Egly, *Nuclear retinoid receptors and the transcription of retinoid-target genes*. *Genes*, 2004. **328**: p. 1-16.
46. Kurokawa, R., et al., *Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding*. *Nature*, 1994. **371**: p. 528-531.
47. Rastinejad, F., et al., *Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1*. *Embo Journal*, 2000. **19**(5): p. 1045-1054.
48. Leid, M., P. Kastner, and P. Chambon, *Multiplicity generates diversity in the retinoic acid signalling pathways*. *Trends in Biochemical Sciences*, 1992. **17**(10): p. 427-433.
49. Schmidt, C. and H. Urlaub, *Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies*. *Current Opinion in Structural Biology*, 2017. **46**: p. 157-168.
50. Bullock, J.M.A., et al., *Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry*. *Structure*, 2018. **26**(7): p. 1015-1024.e2.
51. Koukos, P.I. and A.M.J.J. Bonvin, *Integrative Modelling of Biomolecular Complexes*. *Journal of Molecular Biology*, 2020. **432**(9): p. 2861-2881.
52. Orban-Nemeth, Z., et al., *Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data*. *Nature Protocols*, 2018. **13**(3): p. 478-494.
53. Adikaram, P.R., et al., *Development of R7BP inhibitors through cross-linking coupled mass spectrometry and integrated modeling*. *Communications Biology*, 2019. **2**(1): p. 338.
54. Faini, M., F. Stengel, and R. Aebersold, *The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology*. *Journal of The American Society for Mass Spectrometry*, 2016. **27**(6): p. 966-974.
55. Chavez, J.D. and J.E. Bruce, *Chemical cross-linking with mass spectrometry: a tool for systems structural biology*. *Current opinion in chemical biology*, 2019. **48**: p. 8-18.
56. Mädler, S., et al., *Does chemical cross-linking with NHS esters reflect the chemical equilibrium of protein-protein noncovalent interactions in solution?* *Journal of the American Society for Mass Spectrometry*, 2010. **21**(10): p. 1775-1783.
57. Haris, P.I., *Infrared Spectroscopy of Protein Structure*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1095-1106.
58. Voronina, L., et al. *Mid-Infrared Spectroscopy and Mass Spectrometry Combined to Build a Comprehensive Lung Cancer Fingerprint of Blood Plasma*. in *2019 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)*. 2019.
59. Arndt, C., et al., *Native polyacrylamide gels*. *Methods in molecular Biology*, 2012. **869**: p.49-53.
60. Roelofs, J., et al., *Native Gel Approaches in Studying Proteasome Assembly and Chaperones*. *Methods in molecular biology (Clifton, N.J.)*, 2018. **1844**: p. 237-260.

61. Nowakowski, A.B., W.J. Wobig, and D.H. Petering, *Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions*. *Metallomics : integrated biometal science*, 2014. **6**(5): p. 1068-1078.
62. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes*. *Journal of Mass Spectrometry*, 2003. **38**(12): p. 1225-1237.
63. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* *Angewandte Chemie-International Edition*, 2018. **57**(22): p. 6390-6396.
64. Chen, Z., et al., *Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes*. *Wellcome open research*, 2016. **1**: p. 1-5.
65. Walzthoeni, T., et al., *xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry*. *Nature Methods*, 2015. **12**(12): p. 1185-1190.
66. Zheng, Q., et al., *Probing Protein 3D Structures and Conformational Changes Using Electrochemistry-Assisted Isotope Labeling Cross-Linking Mass Spectrometry*. *Journal of the American Society for Mass Spectrometry*, 2016. **27**(5): p. 864-875.
67. Chavez, J.D., et al., *Quantitative interactome analysis reveals a chemoresistant edgotype*. *Nature Communications*, 2015. **6**(1): p. 1-12.
68. Chen, Z.A. and J. Rappsilber, *Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes*. *Nature Protocols*, 2019. **14**(1): p. 171-201.
69. Müller, D.R., et al., *Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis*. *Analytical Chemistry*, 2001. **73**(9): p. 1927-1934.
70. Fischer, L., Z.A. Chen, and J. Rappsilber, *Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers*. *Journal of Proteomics*, 2013. **88**: p. 120-128.
71. Chen, Z.A., et al., *Quantitative Cross-linking/Mass Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant*. *Molecular & Cellular Proteomics*, 2016. **15**(8): p. 2769.
72. Schmidt, C. and C.V. Robinson, *A comparative cross-linking strategy to probe conformational changes in protein complexes*. *Nature Protocols*, 2014. **9**(9): p. 2224-2236.
73. Zhong, X., et al., *Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled Protein Interaction Reporter (PIR) Cross-Linkers*. *Journal of Proteome Research*, 2017. **16**(2): p. 720-727.
74. Yu, C., et al., *Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes*. *Analytical Chemistry*, 2016. **88**(20): p. 10301-10308.
75. Drabovich, A.P., et al., *Chapter 2 - Proteomic and Mass Spectrometry Technologies for Biomarker Discovery*, in *Proteomic and Metabolomic Approaches to Biomarker Discovery*, H.J. Issaq and T.D. Veenstra, Editors. 2013, Academic Press: Boston. p. 17-37.
76. Zybailov, B., et al., *Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae**. *Journal of Proteome Research*, 2006. **5**(9): p. 2339-2347.
77. Ishihama, Y., et al., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein*. *Molecular & Cellular Proteomics*, 2005. **4**(9): p. 1265-1272.
78. Trudgian, D.C., et al., *Comparative evaluation of label-free SING normalized spectral index quantitation in the central proteomics facilities pipeline*. *Proteomics*, 2011. **11**(14): p.2790-2797.
79. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control*. *Nature*, 2011. **473**: p. 337-342.
80. Belouah, I., et al., *Peptide filtering differently affects the performances of XIC-based quantification methods*. *Journal of Proteomics*, 2019. **193**: p. 131-141.

81. Andersen, J.S., et al., *Proteomic characterization of the human centrosome by protein correlation profiling*. *Nature*, 2003. **426**: p. 570-574.
82. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments*. *Bioinformatics*, 2010. **26**(7): p. 966-968.
83. Lau, H.-T., et al., *Comparing SILAC- and stable isotope dimethyl-labeling approaches for quantitative proteomics*. *Journal of proteome research*, 2014. **13**(9): p. 4164-4174.
84. Uzozie, A.C. and R. Aebersold, *Advancing translational research and precision medicine with targeted proteomics*. *Journal of Proteomics*, 2018. **189**: p. 1-10.
85. Pappireddi, N., L. Martin, and M. Wühr, *A Review on Quantitative Multiplexed Proteomics*. *ChemBioChem*, 2019. **20**(10): p. 1210-1224.
86. Petrotchenko, E.V., K.A.T. Makepeace, and C.H. Borchers, *DXMSMS Match Program for Automated Analysis of LC-MS/MS Data Obtained Using Isotopically Coded CID-Cleavable Cross-Linking Reagents*. *Current Protocols in Bioinformatics*, 2014. **48**(1): p. 8.18.1-8.18.19.
87. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments*. *Bioinformatics (Oxford, England)*, 2010. **26**(7): p. 966-968.
88. Zhang, G., et al., *Protein quantitation using mass spectrometry*. *Methods in molecular biology (Clifton, N.J.)*, 2010. **673**: p. 211-222.
89. Trnka, M.J., R. Pellarin, and P.J. Robinson, *Role of integrative structural biology in understanding transcriptional initiation*. *Methods*, 2019. **159**: p. 4-22.
90. Singh, P., A. Panchaud, and D.R. Goodlett, *Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique*. *Analytical Chemistry*, 2010. **82**(7): p. 2636-2642.

# Chapitre VI : Vers la caractérisation d'une protéine partiellement désordonnée par XL-MS : la protéine CBP

Dans ce chapitre, nous avons appliqué le protocole XL-MS, précédemment optimisé, sur une protéine partiellement désordonnée de la famille des coactivateurs nucléaires : la protéine CBP.

Cette étude s'inscrit dans le cadre d'une collaboration avec l'équipe du Dr. Natacha Rochel au Département de Biologie Structurale Intégrative de l'IGBMC (UMR 7104, Illkirch-Graffenstaden).

## 1. Introduction

### **1.1 Régions et protéines intrinsèquement désordonnées**

Pendant longtemps il a été admis qu'une structure tridimensionnelle fixe était essentielle à sa fonction. C'est le paradigme structure-fonction. Christian Anfinsen (prix Nobel en 1972) a suggéré que cette structure native serait unique et prédéterminée par la séquence en acides aminés<sup>[1]</sup>. Mais dès les années 1950, le concept d'adaptabilité conformationnel est apparu, proposant que les protéines ne doivent pas être considérées comme des objets tridimensionnels rigides.<sup>[2, 3]</sup> Grâce aux techniques d'analyse structurale comme la RMN ou la DRX notamment mais aussi par les outils de prédiction de structure bio-informatiques, l'hypothèse que le désordre structural était une exception se réfute et l'identification de protéine active et désordonnée dans leur forme native entraîne la révision du paradigme structure-fonction.<sup>[4, 5]</sup>

Le terme Région Intrinsèquement Désordonnée (Intrinsically-disordered Region IDR) a été introduit pour définir des régions protéiques ou des protéines (intrinsically disordered protein IDP) qui ne possèdent pas une structure tridimensionnelle bien définie ou qui existent sous la forme d'un ensemble de conformations de stabilités comparables<sup>[6]</sup> Les IDPs interviennent dans de nombreuses fonctions en interaction avec des protéines structurées (régulation de la transcription, transport membranaire, voie de signalisation.<sup>[7]</sup> Elles sont aussi impliquées dans certaines pathologies comme le cancer<sup>[8]</sup>, les maladies neurodégénératives, le diabète<sup>[9]</sup> ou encore les maladies à prion<sup>[10]</sup>. L'une des maladies provenant d'un mauvais repliement et de l'agrégation d'IDPs est la maladie d'Alzheimer. Elle implique l'agrégation non soluble de deux IDPs : amyloïde  $\beta$  et tau<sup>[11, 12]</sup>. De même, la maladie de Parkinson est liée à l'agrégation de l' $\alpha$ -synucléine dans les neurones, conduisant au déclin des fonctions cognitives et motrices.<sup>[13]</sup>

Dans ce chapitre, nous nous sommes intéressés à une protéine partiellement désordonnée, la CREB- Binding Protéine (CBP). Il s'agit d'une protéine coactivatrice permettant la régulation de la transcription cellulaire.<sup>[14]</sup>

## **1.2 Le rôle de CBP : protéine coactivatrice multitâche**

Les CREB (*CRE Binding protein*) sont des protéines de liaison aux CRE (*cAMP Responsive Elements*) qui sont des éléments de réponse à l'AMPc se situant dans la partie régulatrice des gènes dont la transcription est dépendante de la présence de l'AMPc. Ce sont des facteurs de transcription de type "*leucine-zipper*" qui régulent l'expression des gènes dans une pléthore de processus cellulaires.<sup>[15-18]</sup>

L'étape de transactivation est stimulée par phosphorylation par la protéine kinase A. L'état phosphorylé ou non de CREB n'influence pas sa fixation à l'ADN mais semble requise pour le recrutement de son coactivateur : la protéine CBP ("CREB Binding Protein"). CBP et la protéine homologue P300 constituent l'unique famille d'acétyltransférases KAT3<sup>[19]</sup>

CBP est un cofacteur transcriptionnel. Il agit comme "échafaudage" pour l'ensemble des molécules impliquées dans la transcription. Son interaction avec les facteurs de transcriptions spécifiques liés à leurs éléments de réponse au niveau du promoteur (comme RAR ou PPAR Peroxisome Proliferator Activated Receptor) permet son interaction avec d'autres facteurs de transcription généraux tels que TBP (TATA box Binding Protein - composant de TFIID), TFIIB, TFIIE et TFIIIF. CBP participe donc à la formation et à la stabilisation du complexe actif nécessaire pour la transcription par l'ARN polymérase II appelé complexe d'initiation de la transcription. CBP possède également une activité histones-acétylase sur des lysines de l'extrémité N-terminale des histones. Cette acétylation induit la déstabilisation des nucléosomes et permet le passage de l'ARN polymérase. CBP recrute des protéines appelées Nucleosome Assembly Proteins impliquées dans le contrôle de l'organisation des réseaux de nucléosomes formant la chromatine.

Il a été reporté que des mutations peuvent perturber leur activité catalytique. Ces mutations sont associées à un certain nombre de maladies humaines telles que le syndrome de Rubinstein-Taybi<sup>[20]</sup>, la leucémie<sup>[21]</sup> et le lymphome<sup>[22]</sup>. C'est pourquoi l'étude de ces protéines coactivatrices est importante afin de mieux comprendre comment les mécanismes de mutations sont réalisés.

Dans le cadre de notre projet, nous nous sommes intéressés à CBP pour sa capacité à médier l'activation transcriptionnelle des hétérodimères de RAR/RXR avec l'aide de la protéine SRC-1 (Steroid Receptor Coactivator 1).<sup>[23, 24]</sup> SRC-1 est connu pour interagir avec les récepteurs nucléaires à travers une région centrale constituée de trois motifs hélicoïdaux conservés de la séquence peptidique consensus LXXLL (avec L : leucine et X : acide aminé quelconque).<sup>[25]</sup> Elle interagit également avec CBP via son domaine C-terminal.<sup>[26]</sup> Il a également été montré que l'hétérodimère RAR/RXR pouvait interagir avec CBP de manière ligand dépendant au niveau du domaine N-terminal de CBP. Cette liaison entre CBP et l'hétérodimère s'établit suite à la fixation d'un 1<sup>er</sup> coactivateur SRC-1.<sup>[27]</sup> Cette interaction tripartite entre RAR CBP et SRC-1 entraîne une stabilisation du complexe qui induirait une augmentation de l'activité transcriptionnelle. Aucune information structurale n'est disponible sur le complexe RAR/RXR/SRC-1/CBP mais un modèle structural obtenu par cryo-EM est proposé pour un autre récepteur nucléaire, le récepteur aux estrogènes ER $\alpha$ , en interaction avec p300.<sup>[28]</sup>

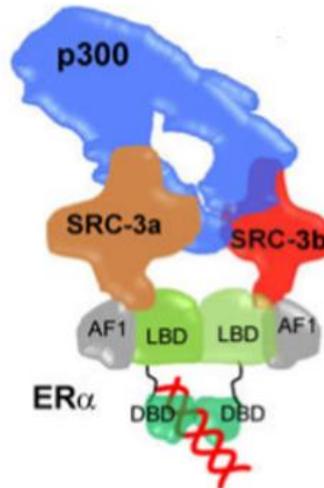


Figure 107 : Modèle structural proposé par cryo-EM du complexe formé par l'homodimère ER $\alpha$  (constitué des domaines AF1, LBD et DBD et fixé sur l'ADN), le coactivateur SRC-3 et p300 d'après[28]

### 1.3 CBP : une protéine désordonnée constituée de plusieurs sous-domaines

CBP est une protéine qui contient à la fois des régions intrinsèquement désordonnées et des domaines fonctionnels repliés. Le domaine histone acétyltransférase (HAT) est flanqué d'autres domaines, y compris le bromodomaine (BRD), la région riche en cystéine / histidine 2 (CH2) et la région riche en cystéine / histidine 3 (CH3), qui ont un rôle potentiel dans la régulation de l'activité de l'acétyltransférase et de la spécificité du substrat.<sup>[29-31]</sup> Des études ont par ailleurs démontré que le domaine BRD de CBP peut se lier efficacement aux lysines acétylées dans les histones H2A, H2B, H3 et H4, et contribue à la fonction du domaine HAT en acétylant les nucléosomes dans la chromatine native.<sup>[32]</sup> Le domaine HAT de CBP contient une boucle d'autorégulation riche en lysine qui lorsqu'elle est hypoacétylée lui confère la capacité d'inhiber l'activité de l'acétyltransférase.<sup>[33]</sup>

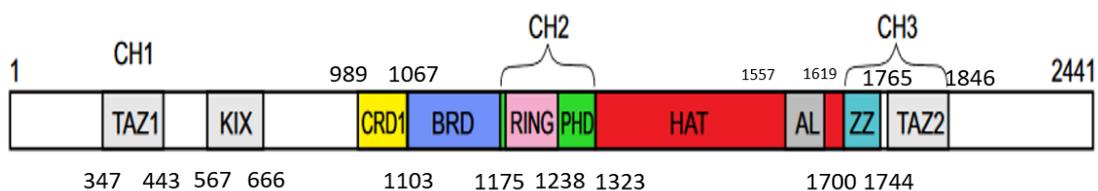
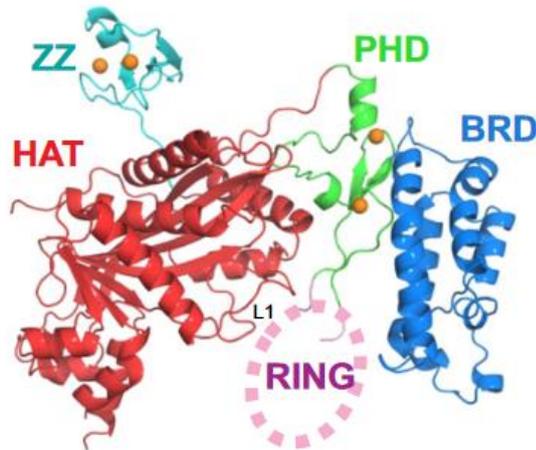


Figure 108 : Représentation des différents domaines de CBP. Extrait de [26]

Des études effectuées par DRX ont permis d'établir une première structure du noyau catalytique de CBP englobant les domaines BRD, CH2 (constitué des sous-domaines PHD et RING) et du domaine HAT.<sup>[34]</sup> Alors que quelques d'études ont été entrepris sur les domaines structurés (HAT, BRD, TAZ1, TAZ2, KIX), peu de travaux décrivent la partie C-terminale de la protéine CBP (domaine CH3).<sup>[35]</sup> Il faut noter qu'il est très difficile d'étudier cette protéine intacte, seule ou en interaction avec des partenaires car il s'agit d'une protéine lourde et désordonnée. Ainsi, le positionnement des sous-unités sur le cœur mieux structuré de la protéine reste délicat et peu abordé dans la littérature. C'est pourquoi, des méthodes

d'analyses complémentaires ont dû être utilisées pour effectuer la caractérisation de cette protéine.<sup>[36]</sup>



**Figure 109 : Structure cristalline du noyau catalytique de CBP souris. Les domaines BRD, PHD, HAT et ZZ sont représentés respectivement en bleu, vert, rouge et cyan. Les atomes de zinc sont représentés par des sphères orange. [37]**

Dans le cadre de ce projet, nous avons l'objectif d'obtenir des informations structurales sur la protéine CBP à l'aide de la XL-MS afin de positionner les différents domaines de la protéine et nous guider vers la modélisation d'une structure à plus haute résolution de la protéine. La caractérisation de cette protéine permettra par la suite de mieux comprendre son implication dans l'activation transcriptionnelle des hétérodimères de RAR/RXR.

#### **1.4 Difficultés de l'étude CBP**

La caractérisation de la protéine CBP s'est avérée délicate. En effet, nous avons fait face à plusieurs difficultés :

- Cette protéine possède des 7 domaines présentant des motifs structuraux (en couleur sur la figure 109) mais plus de 1400 acides aminés (ce qui représente 60% de la séquence protéique) sont prédits comme des régions intrinsèquement désordonnées (RIDs). De plus, le domaine HAT contient une région désordonnée (région AIL des acides aminés 1557 à 1619). A cela s'ajoutent peu d'informations structurales disponibles dans la littérature pour nous guider dans l'optimisation des paramètres réactionnels, notamment le choix très important du cross-linker.
- Lorsqu'on parle de IDPs, on s'attend à un désordre intrinsèque et donc que la protéine peut adopter de multiples conformations.
- Les IDPs établissent moins de liaisons intramoléculaires stabilisatrices que les protéines ordonnées et donc sont plus dynamiques ! Elles ne possèdent pas suffisamment d'acides aminés non polaires pour former un cœur hydrophobe comme c'est très souvent le cas chez les protéines ordonnées.
- Après expression et purification, la quantité comme la concentration de la protéine purifiée sont faibles ne permettant pas de pouvoir effectuer beaucoup de répliquats. D'autre part, et ceci est commun à l'étude de beaucoup d'IDPs, CBP est sensible à l'action des protéases. Si cela peut être exploité d'un point de vue structural pour

localiser les IDRs, cela risque de rendre l'échantillon encore plus hétérogène et ainsi de fausser l'interprétation de nos résultats.

- Enfin, à cause de la masse élevée de la protéine pontée (265 kDa), la réaction de pontage ne peut pas être suivie sur gel d'électrophorèse SDS-PAGE ni sur MALDI-MS en l'absence d'un détecteur de hautes masses.<sup>[38]</sup>

L'objectif de cette étude était ainsi double : (i) caractériser la protéine CBP afin de préciser l'organisation architecturale entre les domaines ce qui nous permettra de mieux comprendre son implication dans l'activation transcriptionnelle et (ii) évaluer l'efficacité du protocole XL-MS que nous avons mis en place sur le complexe RAR-RXR, sur des protéines désordonnées.

## 2. Protocole expérimental

### **2.1 Expression et purification de CBP**

La protéine CBP a été exprimée en appliquant le système de vecteurs d'expression de baculovirus. Un baculovirus est un virus enveloppé d'ADN qui a la possibilité d'infecter énormément d'espèces d'insectes et qui est communément utilisé pour surexprimer des protéines recombinantes dans des cellules d'insectes<sup>[39]</sup>.

CBP humain a été exprimé sous forme d'une protéine taguée (Histag+Flag) afin de faciliter la purification sur colonne d'affinité (colonne au nickel + Anticorps Anti-Flag). Une seconde étape de purification sur colonne d'exclusion de gel permet d'éliminer les agrégats éventuels.

Chaque étape de l'expression et de la purification est ensuite contrôlée par SDS-PAGE afin de s'assurer que la protéine CBP exprimée par baculovirus n'a pas subi de protéolyse. La concentration d'échantillon est par la suite déterminée par nanodrop. Une concentration proche de 0,3 – 0,5 mg/ml est classiquement obtenue (1 – 1,8 pmol/ $\mu$ L).

### **2.2 La réaction de pontage**

La réaction de pontage est effectuée à partir d'un échantillon de CBP concentré à 0.3mg/mL (environ 1.13 $\mu$ M) dans un tampon HEPES 25mM, NaCl 75mM, KCl 75mM et TCEP 1mM à pH 7.5 sur un échantillon frais juste après la dernière étape de gel filtration afin de limiter une agrégation artéfactuelle liée au stockage et pour limiter la dégradation protéolytique de l'échantillon. Nous avons dans un 1<sup>er</sup> temps choisi d'appliquer le protocole mis en place sur RAR/RXR avec des cross-linkers de type amino-réactif.

CBP contient 447 sites candidats au pontage amino-réactifs (116 lysines, 218 sérines et 113 thréonines). Après un 1<sup>er</sup> essai infructueux de pontage par du BS3 en ratio Protéine/BS3 : 1/50 (identification de 12 peptides uniques), nous avons réalisé les expériences suivantes avec davantage d'agents pontant, en ratio 1/200. Après 30min de réaction à température ambiante, la réaction est stoppée par l'ajout de Tris/HCl 25mM final.

Puis la protéine pontée a été précipitée afin d'éliminer les sels et excès de réactifs, dérivatisée à l'iodoacétamide et hydrolysée par une enzyme de digestion de façon classique en analyse protéomique. Après analyse par LC-MS/MS, les peptides pontés ont été identifiés en utilisant pLink, XlinkX, XiSearch et MeroX puis validés manuellement.

### **2.3 Optimisation de la digestion enzymatique**

En examinant la séquence de la protéine CBP, nous nous sommes rendu compte que les lysines et arginines (sites d'hydrolyse de la trypsine) n'étaient pas réparties régulièrement sur la séquence de CBP, avec des régions pauvres sur les parties N- et C-terminales ce qui nous limiterait dans la caractérisation de ces domaines après pontage. C'est pourquoi nous avons décidé de tester en parallèle de la trypsine, une seconde enzyme qui présente une spécificité différente, la chymotrypsine. La spécificité de la chymotrypsine est corrélée à l'hydrophobicité de l'acide aminé en amont de l'hydrolyse. Ainsi, cette dernière clive préférentiellement après les tryptophanes, tyrosines et phénylalanines, puis dans une moindre mesure, les leucines, méthionines et histidines. Comme pour la trypsine, cette enzyme est efficace à 37°C dans des conditions standards. Elle nous fournira des informations sur les régions N et C-terminales de CBP d'une part et permettra de détecter d'autres peptides pontés complémentaires à la digestion trypsique.

Une analyse des données de digestion après pontage au BS3 de CBP par le logiciel MASCOT révèle que la chymotrypsine peut hydrolyser la protéine après réticulation. Un pourcentage de recouvrement de séquence (en tenant compte des Dead end) de 50% a été obtenu, ce qui est comparable à celui calculé avec la trypsine de 64%. Il est intéressant de noter que les zones de recouvrement de séquence sont complémentaires (voir annexe 8)

### **2.4 Choix de deux agents pontant : optimisation du bras espaceur**

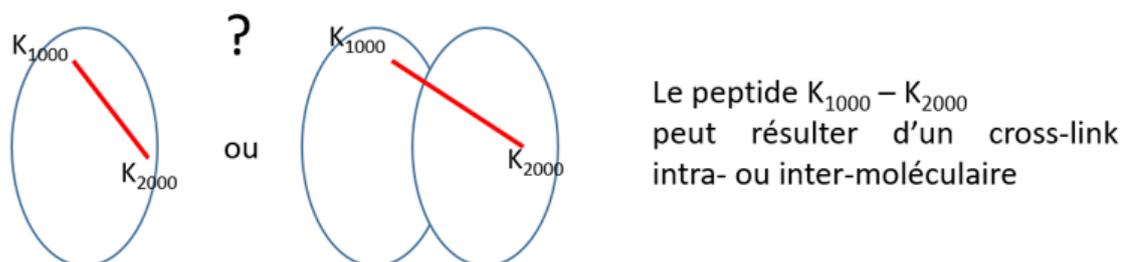
Afin d'optimiser nos chances de détecter des peptides pontés intramoléculaires pertinents sur CBP, nous avons décidé de tester un second agent pontant. L'option de choisir un cross-linker présentant une réactivité chimique différente a été envisagée mais écartée dans un 1<sup>er</sup> temps. En effet, une réactivité sur les thiols s'annonçait peu prometteuse aux vues du faible nombre de cystéines (50 cystéines, non réparties uniformément sur la séquence puisque 21 cystéines sont situées entre les AA 1660 et 1850). Une autre possibilité aurait été de travailler avec un agent pontant capable de se fixer sur les tyrosines.<sup>[40, 41]</sup> mais nous avons jugé le nombre de tyrosine présent sur la structure également trop faible (41 tyrosines réparties uniformément). Une expertise encore très limitée avec les cross-linkers photoréactifs moins spécifiques associée à une protéine de structure inconnue très flexible a finalement guidé notre choix vers un cross-linker dont on préserverait la spécificité amino-réactive mais dont on modifierait la longueur du bras espaceur. C'est pourquoi, nous avons retenu le BS(PEG)<sub>5</sub>. Grâce à une longueur de bras espaceur plus grand que le BS3, il devrait être possible d'aller fixer des sites susceptibles d'être plus éloignés sur une protéine en mouvement.

L'objectif de cette étude était ainsi d'optimiser le nombre de sites pontés sur CBP afin d'identifier quel domaine est proche de quel autre domaine.

### 3. Résultats et Discussion

#### **3.1 Détermination de la stœchiométrie d'oligomérisation potentielle par MALDI-MS**

Afin de pouvoir interpréter nos résultats de cross-link de façon pertinente, il nous a paru essentiel de vérifier la stœchiométrie de complexation de CBP. En effet, dans l'éventualité où CBP serait présente sous forme d'homodimère, il est alors impossible de faire la différence entre un peptide intramoléculaire ou intermoléculaire.



**Figure 110 : Représentation schématique de l'ambiguïté de l'identification des peptides pontés intramoléculaire et intermoléculaire.**

Pour suivre la stœchiométrie de complexation après cross-link, nous disposons au laboratoire de deux approches : l'électrophorèse sur gel SDS-PAGE et le MALDI-MS. Ni les gels d'acrylamide que nous n'utilisons ni le MALDI-MS dont nous disposons ne permettraient de détecter un complexe d'une taille équivalente à celle de CBP cross-linkée si la protéine formait un dimère ( $MM > 530$  kDa). C'est pourquoi nous avons pris contact avec l'équipe du Dr Stéphane Chaignepain du Centre de Génomique Fonctionnelle de Bordeaux, qui disposait d'un spectromètre de masse MALDI équipé d'un détecteur haute masse (CovalX<sup>®</sup>). La protéine CBP a ainsi été analysée avant et après pontage chimique par du BS3 ou du BS(PEG)<sub>5</sub> à une concentration de  $1.9 \mu\text{M}$  selon la méthode de la goutte séché en utilisant comme matrice l'acide sinapique (SA).

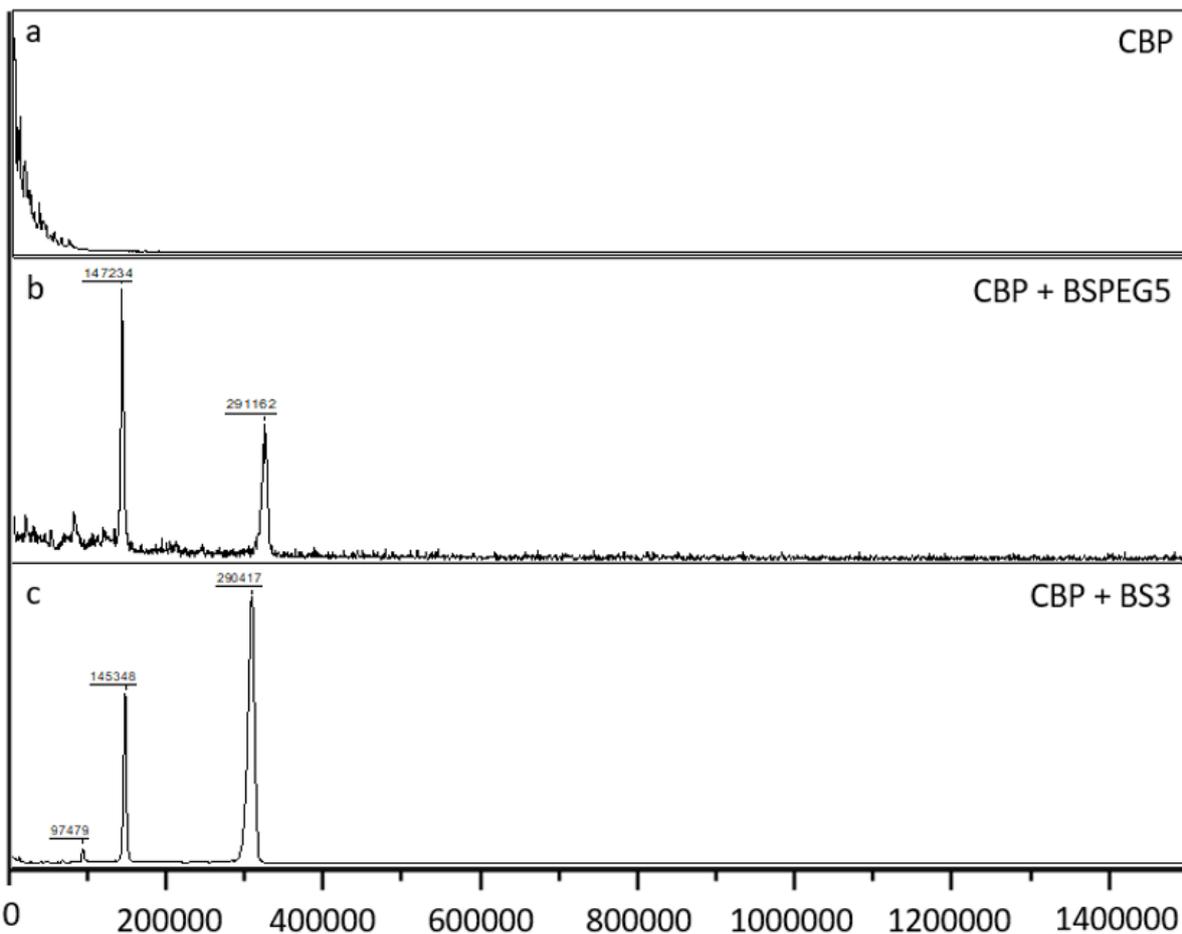


Figure 111 : Analyse MALDI-MS de la protéine CBP (a) avant pontage (b) après pontage avec le BS(PEG)<sub>5</sub>. (c) après pontage avec le BS3.

De façon surprenante, et malgré nos efforts (changement de matrice, concentration modifiée de la matrice, précipitation de l'échantillon au méthanol à froid dans l'objectif de reconcentrer la protéine), seuls les échantillons contenant CBP cross-linkée ont abouti à l'obtention d'un signal par MALDI. Une des explications possibles serait que la protéine se soit agrégée avant analyse durant le transport vers Bordeaux (1 jour) alors que les échantillons cross-linkés auraient été « stabilisés » par le pontage. En effet, la protéine est connue pour s'agréger assez facilement.

Les échantillons ayant subi un cross-link par BS3 et BSPEG ont généré un signal dont les masses moléculaires furent mesurées à 290,41 kDa et 291,16 kDa respectivement. Les ions observés à 145348 Da et à 97479 Da correspondent respectivement aux ions chargés 2+ et 3+. Des signaux relativement larges sont obtenus, ce qui est lié à la présence de sels en concentration non négligeable (Hepes 25mM, NaCl 75mM, KCl 75mM, agent de pontage en excès). Néanmoins, la précision de mesure de la masse est largement suffisante pour nous indiquer sans ambiguïté que les ions détectés correspondent à CBP sous forme monomérique. Aucun signal de masse supérieure n'est observé. Ceci indique que CBP n'a pas été réticulée sous forme de dimère et que chaque peptide cross-linké que nous interpréterons pourra être relié à un peptide ponté intramoléculaire, fournissant des données sur l'arrangement spatial des domaines au sein d'une même molécule de CBP. L'écart de masse entre la masse mesurée et la masse théorique de CBP (265181 Da) montre la réticulation par l'agent pontant sur la structure.

### **3.2 Identification des peptides pontés**

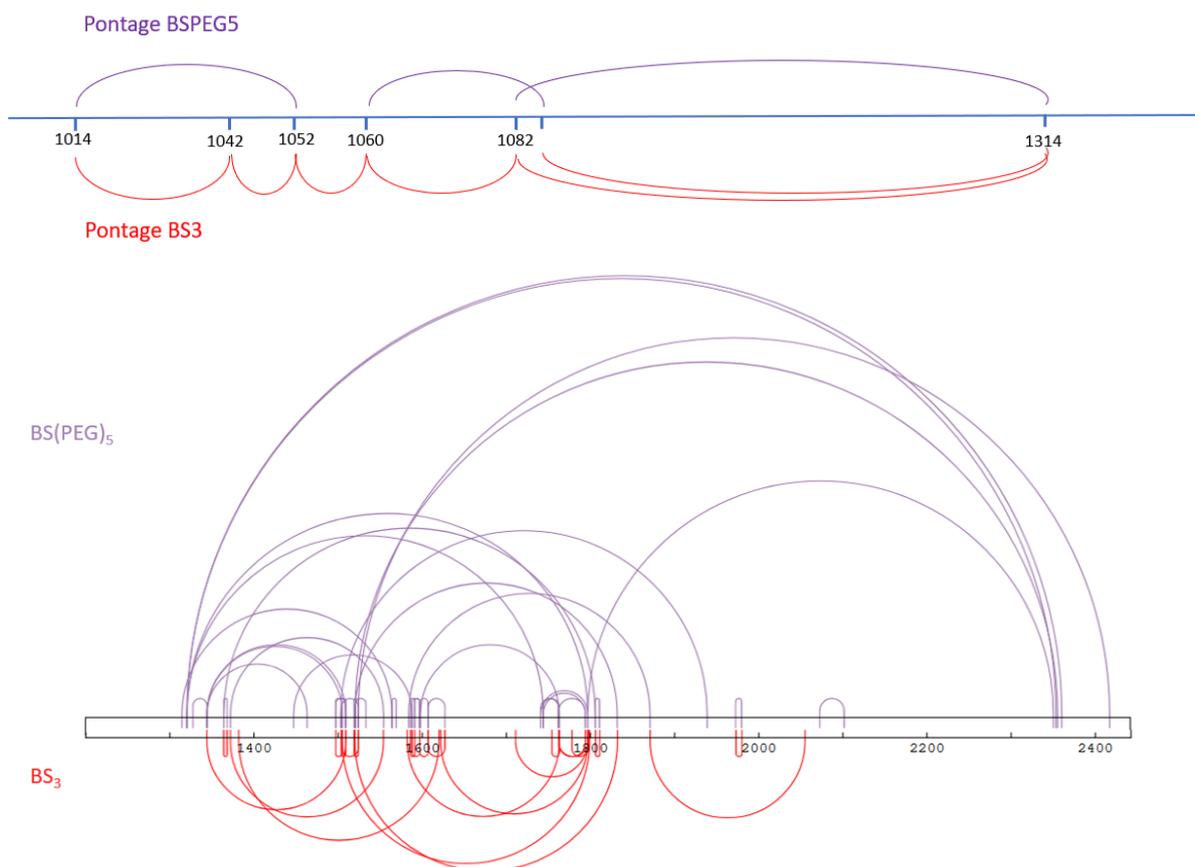
Par la suite, le complexe ponté a été digéré soit par la trypsine, soit par la chymotrypsine. L'objectif est ainsi de voir s'il est possible d'identifier de manière complémentaire les zones pontées sur le complexe.

Comme nous en avons noté l'intérêt pour RAR/RXR, nous avons fait appel à plusieurs logiciels de recherche dédiés à l'identification des peptides pontés (pLink, XlinkX, MeroX et XiSearch). Dans les recherches, la tolérance de masse de l'ion précurseur et des ions fragments ont été fixés à 10 et 20ppm respectivement. Chaque ensemble de données a été produit en utilisant un FDR de 5%. Tous les spectres identifiés ont été validés manuellement lorsqu'au moins quatre ions fragments consécutifs correspondant à chacune des deux séquences liées ont été observés. Un score arbitraire a été attribué en fonction de la redondance d'identification selon les logiciels (50 points si le peptide est interprété par un logiciel, 100 points par 2 logiciels, 150 points par 3 et 200 points par les 4 logiciels considérés). Les tableaux 16 et 17 montrent les résultats obtenus lors du pontage par le BS3 et par le BS(PEG)<sub>5</sub> après digestion par la trypsine respectivement. Dans ces tableaux, ne sont répertoriés que les peptides pontés uniques.

#### **3.2.1 Comparaison entre les échantillons pontés par le BS3 et le BSPEG**

Dans un 1<sup>er</sup> temps, nous nous sommes focalisés sur les échantillons obtenus par digestion trypsique. Après validation manuelle, 52 peptides pontés intramoléculaires uniques ont été interprétés lors du pontage par le BS3 et 67 peptides pontés uniques par le BS(PEG)<sub>5</sub>, ce qui correspond à 20% de rejet d'interprétations tous logiciels confondus. Les peptides pontés « doublon » qui reflètent une incertitude de positionnement de l'acide aminé ponté au sein d'une même séquence peptidique ont été éliminés du tableau. C'est notamment le cas lorsque deux lysines se suivent dans la séquence (par ex pour les peptides pontés 1497-1504 et 1497-1505 ou les peptides 1597-1606 et 1597-1607). Pour cette étude sur CBP, nous cherchons à obtenir des 1<sup>res</sup> informations sur la proximité de zones ou de domaines c'est pourquoi la notion de site ponté n'est pas encore pertinente. Parmi les 52 peptides obtenus en présence de BS3, 18 peptides résultent d'un pontage impliquant une sérine ou une thréonine et seulement 1/3 des peptides ont été identifiés par au moins deux logiciels (score arbitraire >50). Pour l'échantillon cross-linké au BS(PEG)<sub>5</sub>, il ressort 63 peptides à S/T et 46% de peptides identifiés par au moins deux logiciels.

Il est intéressant de noter que les peptides pontés aux BS3 et BS(PEG)<sub>5</sub> convergent et permettent de confirmer les identifications (figure 112).



**Figure 112 : Comparaison des cartes de pontage générées par XiView lorsque le cross-link est réalisé par le BS(PEG)<sub>5</sub> (violet) et le BS3 (rouge). La schématisation du haut montre que des peptides uniques différents représentent le pontage BS3 et BS(PEG)<sub>5</sub> mais que les mêmes acides aminés sont accessibles dans les deux expériences. La schématisation du bas montre la complémentarité des peptides pontés entre les acides aminés 1200 et 2442 de CBP.**

La validation est souvent limitée car une seule des deux séquences pontées ( $\alpha$  ou  $\beta$ ) est identifiée sans ambiguïté. Utiliser un cross-linker clivable pourrait nous aider à confirmer que les spectres MS/MS considérés correspondent effectivement à des ions modifiés par le cross-linker et donc valider l'identité de ces peptides cross-linkés. De façon intéressante, en utilisant une précision de 2ppm sur l'ion précurseur ainsi que les éléments de séquence interprétée pour le peptide  $\alpha$ , une seule possibilité pour le peptide  $\beta$  est trouvée. Néanmoins, ces peptides pontés candidats n'ont pas été retenus dans notre tableau de résultats car il réside une incertitude sur leur légitimité.

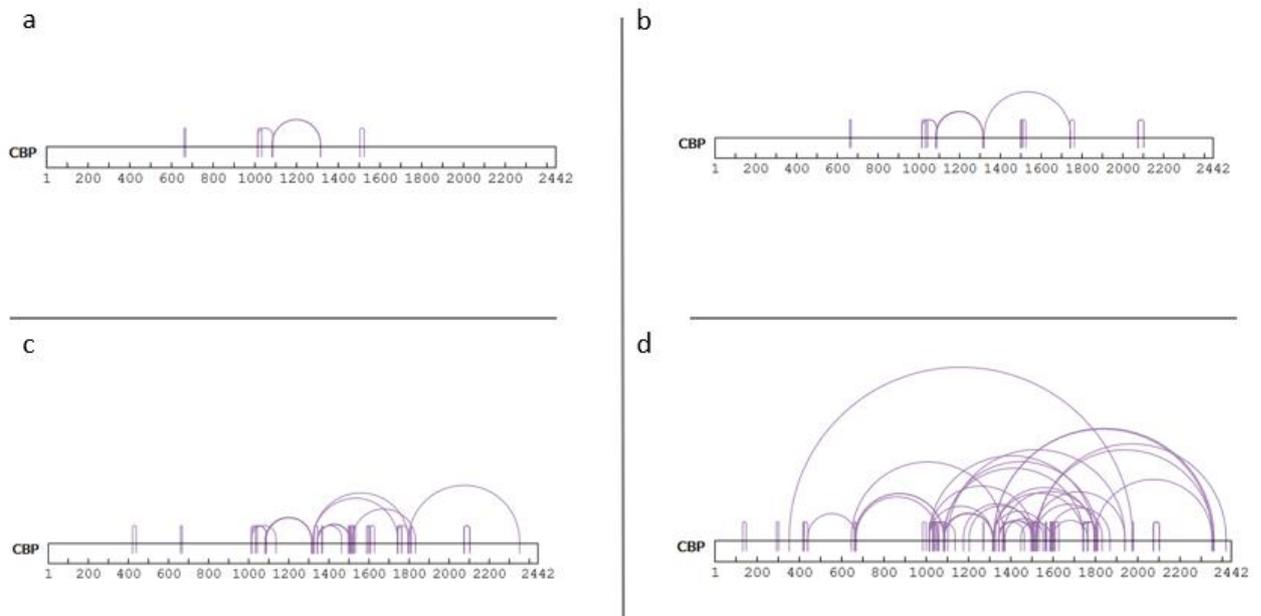
### **3.2.2 Carte de proximité des peptides pontés**

A partir de ces résultats, une première carte de proximité a été réalisée afin de visualiser les zones proches au sein du complexe. Ne disposant que de deux répliquats et sachant que CBP était très flexible, nous avons abordé l'interprétation de cette carte avec précaution. En particulier, nous l'avons décomposée selon la valeur du score arbitraire (attribué en fonction du nombre de logiciels qui ont permis l'identification des peptides) car celui-ci nous donnait un 1<sup>er</sup> gage de confiance sur l'interprétation des données MS/MS (figure 113).

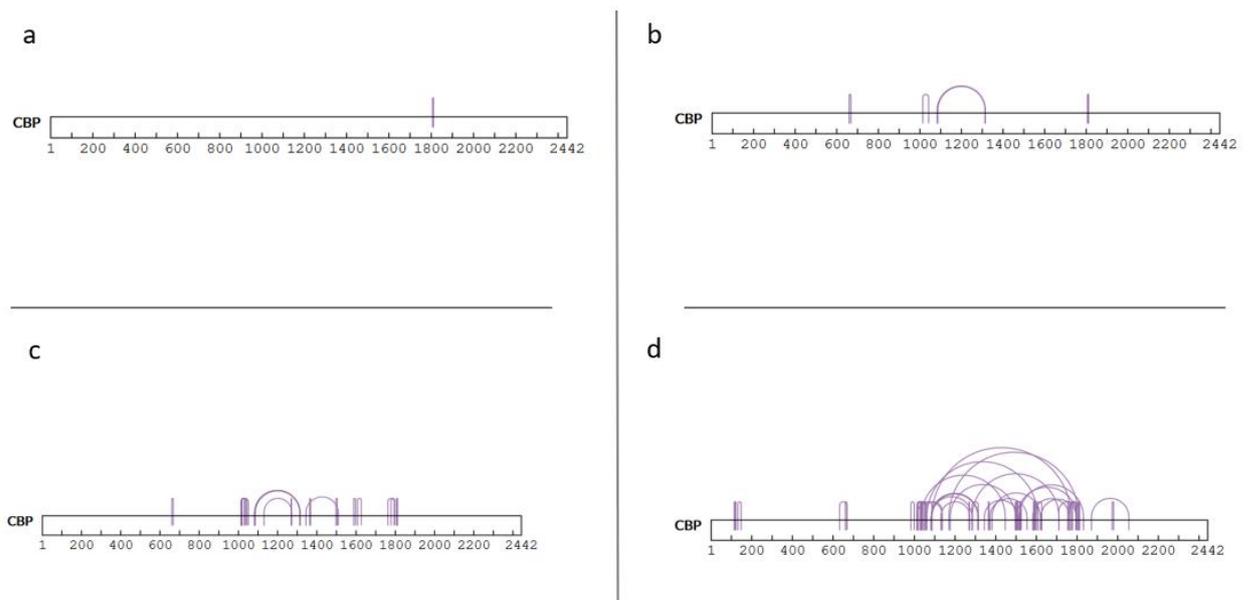


- Tous les peptides cross-linkés par BS3 sont situés sur le cœur de CBP alors que certains peptides cross-linkés par BS(PEG)<sub>5</sub> touchent des zones N- et C-terminales. Cette observation est intéressante car, les 2 cross-linkers ayant la même réactivité chimique, elle nous prouve que certains sites ne sont pas accessibles par BS3 et qu'ainsi les informations déduites du pontage par BS(PEG)<sub>5</sub> sont pertinentes.

- Quel que soit l'agent de pontage considéré, les peptides à haut score arbitraire (identiquement identifiés par plusieurs logiciels) se situent dans le cœur de CBP (structuré) alors que les peptides à bas score arbitraire (identifié par un seul logiciel) mettent en jeu des acides aminés situés dans des zones désordonnées. Ceci peut être lié au fait que les pontages se réalisent de façon plus reproductible dans les zones structurées que dans les zones désordonnées. Ainsi, les peptides issus des zones structurées sont présents en plus forte abondance ce qui génère des spectres MS/MS de meilleure qualité et donc une interprétation moins délicate.



**Figure 114 : Cartes de proximité de la protéine CBP pontée par le BS(PEG)<sub>5</sub> en fonction du score arbitraire. (a) Carte de proximité obtenue pour un score de 200 (b) pour un score de 150 (c) pour un score de 100 et (d) pour un score de 50.**



**Figure 115 : Cartes de proximité de la protéine CBP pontée par le BS3 observé en fonction du score. (a) Carte de proximité obtenue pour un score de 200 (b) pour un score de 150 (c) pour un score de 100 et (d) pour un score de 50.**

- Cette carte met en évidence plusieurs zones de proximité.

Concernant le cœur catalytique, nos résultats montrent des proximités entre des domaines qui sont en accord avec la littérature.<sup>[37]</sup> Des pontages intradomains, mais aussi interdomains sont identifiés. En particulier, CRD1 interagirait avec plusieurs domaines (KIX, TAZ2, PHD et BRD), mais également avec des régions désordonnées (983-999 ; 1060-1086 ; 1102-1940).

Les proximités entre HAT/TAZ2 ; BRD/PHD et TAZ2/ZZ sont représentés par de nombreuses réticulations. Trois peptides pontés ont également été identifiés entre le domaine HAT et le domaine AL qui sont connus dans la littérature pour être très proches.<sup>[37]</sup> Le domaine ZZ serait également positionné proche de plusieurs autres domaines comme AL et PHD mais surtout de TAZ2 pour lequel 5 peptides pontés sont détectés. Les résultats obtenus avec le BS(PEG)<sub>5</sub> sont très intéressants car, avec une longueur de bras espaceur presque deux fois plus grande que le BS3, il permet de sonder la présence de domaines à plus longue distance. En particulier la région C-terminale désordonnée est impliquée dans plusieurs pontages avec PHD, HAT mais aussi avec KIX. Au niveau N-terminal, des peptides pontés ont été détectés entre le domaine KIX et les domaines CRD1 et HAT suggérant que les extrémités N- et C-terminales de CBP serait repliées au niveau du cœur catalytique.

**Tableau 16 : Liste des peptides pontés identifiés entre les sous-unités de CBP pour l'échantillon ponté par le BS3.**

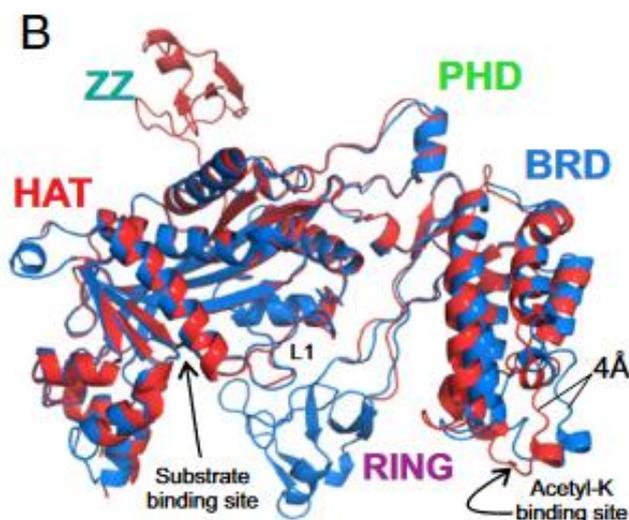
Peptide1	Position 1	Peptide2	Position 2	Score arbitraire	XLinkX	xiSEARCH	pLink	MeroX	CBP domaines
GGSGSSINPGIGNVSASSPVQC	114	SPLSQGDSSAPSLPK	121	50				X	desordered-desordered
QAASTSGPTPAASQALNPQAAQ	148	SPLSQGDSSAPSLPK	132	50				X	desordered-desordered
MENLVAYAK	632	ELEEKRR	668	50			X		KIX-KIX
ELEEKRR	668	IYKIQK	660	150	X	X	X		desordered - KIX
VPSPSSVASAETNSQQPGPDVFI	983	TETQAEDTEPDGSGKGEPR	999	50	X				desordered-CRD1
TETQAEDTEPDGSGKGEPR	1014	SEPMEVDEKKPEVK	1051	100		X	X		CRD1 - CRD1
TETQAEDTEPDGSGKGEPR	1014	NPMDLSTIKR	1136	50		X			CRD1-BRD
SEMMEEEDLQGASQVK	1030	SEMMEEEDLQGASQVK	1019	100		X	X		CRD1 - CRD1
SEMMEEEDLQGASQVKEETDIA	1030	KPEVK	1052	50			X		CRD1-CRD1
SEMMEEEDLQGASQVKEETDIA	1033	TETQAEDTEPDGSGKGEPR	1014	100	X		X		CRD1 - CRD1
SEMMEEEDLQGASQVKEETDIA	1033	SEPMEVDEKKPEVK	1043	100	X			X	CRD1 - CRD1
EETDIAEQKSEPMEVDEKKPEVI	1034	TAVYHEILIGYLEYVK	1447	50				X	CRD1-HAT
EETDIAEQKSEPMEVDEKKPEVI	1036	SEMMEEEDLQGASQVK	1030	50	X				CRD1-CRD1
EETDIAEQKSEPMEVDEK	1042	TETQAEDTEPDGSGKGEPR	1014	150	X	X		X	CRD1 - CRD1
EETDIAEQKSEPMEVDEK	1042	KPEVK	1052	50			X		CRD1-CRD1
SEPMEVDEKKPEVK	1051	NPMDLSTIKR	1136	50			X		CRD1-BRD
VEVKEEESSTNGTASQSTSPSC	1060	KPEVK	1052	50	X				CRD1-CRD1
KKIFKPEELR	1082	KTGRPR	1314	100	X		X		desordered - PHD
IFKPEELR	1086	KPEVK	1052	50			X		desordered-CRD1
IFKPEELR	1086	KTGRPR	1314	150		X	X	X	desordered - PHD
KIFKPEELR	1086	LSIQR	1771	50			X		desordered-TAZ2
QPVDPQLLIGIPDYFDIVKNPMD	1130	KNDTLDPEPFVDCK	1272	100	X			X	BRD-PHD
VYKFCSK	1176	IPKPKR	1497	50		X			BRD-HAT
NDTLDPEPFVDCKECCGR	1284	KKIFKPEELR	1083	50				X	PHD-BRD
NDTLDPEPFVDCKECCGR	1284	KTGRPR	1314	50			X		PHD-PHD
NDTLDPEPFVDCKECCGR	1284	FVYTCNECKHHVETR	1711	50			X		HAT - HAT
TVEVKPGMK	1368	VVASSDK	1364	100	X			X	HAT-HAT
FVDSGEMSESFYR	1382	KKPSMPNVSNDLSQKLYATMEK	1620	50		X			HAT-AL
IPKPKR	1497	KMLDK	1504	100	X		X		HAT-HAT
MLDKAFAER	1509	VNKFLR	1344	100		X	X		HAT-HAT
MLDKAFAER	1509	RLQEWYKK	1504	50			X		HAT-HAT
IIHDYKDIFKQATEDR	1524	LQEWYKK	1504	50			X		HAT-HAT
DIFKQATEDRLTSAK	1524	IIHDYK	1519	50	X				HAT-HAT
ELPYFEGDFWPNVLEESIKELEC	1554	TVEVKPGMK	1372	50			X		HAT-HAT
KNNKKTNK	1588	NKSSISR	1597	100		X	X		AL-AL
NNKKTNK	1591	NAKKK	1586	50			X		AL-AL
KPSMPNVSNDLSQKLYATMEK	1607	EEEESSSTNGTASQSTSPSPQRK	1060	50				X	CRD1-AL
KPSMPNVSNDLSQK	1607	NKSSISR	1597	50			X		AL-AL
KPSMPNVSNDLSQK	1607	LYATMEKHK	1627	100	X		X		AL-AL
WGLGLDDEGSSQGEPQSKSPC	1754	SPQESR	1763	50	X				desordered-desordered
WGLGLDDEGSSQGEPQSKSPC	1762	NANCSLPSQCKMK	1794	100	X		X		HAT-ZZ
SPQESR	1763	EESTAASETTEGSSQGDGSK	1582	50		X			AL-ZZ
NANCSLPSQCKMK	1797	LYATMEK	1622	50		X			TAZ2-HAT
NANCSLPSQCKMK	1797	FVYTCNECKHHVETR	1711	50		X			TAZ2-HAT
NANCSLPSQCKMK	1797	WGLGLDDEGSSQGEPQSKSPQESR	1762	50		X			TAZ2-ZZ
NANCSLPSQCKMK	1797	CIQSLVHACQCR	1778	100		X	X		TAZ2-TAZ2
NANCSLPSQCKMKR	1799	KMLDKAFAER	1505	50		X			TAZ2-HAT
VVQHTKGCKR	1806	KTNGGCPVCK	1811	200	X	X	X	X	TAZ2-TAZ2
GCKRKTNGGCPVCK	1809	KTSRVYK	1170	50			X		TAZ2-BRD
HQENKCPVPFCLNIKHKLRL	1832	IIHDYKDIFKQATEDRLTSAK	1520	50				X	TAZ2-HAT
MATMNTNR	1871	MPSVQPPR	2055	50			X		TAZ2-desordered
EAQQQHLR	1973	VNINNSMPPGR	1980	50				X	desordered-desordered

**Tableau 17 : Liste des peptides pontés identifiés entre les sous-unités de CBP pour l'échantillon ponté par le BS(PEG)<sub>5</sub>.**

Peptide1	Position 1	Peptide2	Position 2	Score arbitraire	XLinkX	xiSEARCH	pLink	MeroX	CBP Domains
QAASTSGPTPAASQALNPQAQK	148	SPLSQGDSSAPSLPK	132	50	X				desordered-desordered
MGITGNTSPFGQPFQAGGQP	291	QSMVNSLPTFPTDIK	302	50				X	desordered-desordered
QIISHWKNCTR	420	NASDKR	439	100		X	X		TAZ1-TAZ1
VEGDMYESANSR	646	QNHPEAGEVFRVVASDDK	1367	50		X			KIX-HAT
IQKELEEK	663	KKIFKPEELR	1082	50		X			KIX-CRD1
ELEEKR	668	NASDKR	439	50		X			KIX-TAZ1
ELEEKRR	668	IYKIQK	660	200	X	X	X	X	desordered - KIX
VPTSSVASAETNSQQPGDPVP	983	TETQAEDTEPDGEGSKGEP	999	50				X	desordered-CRD1
TETQAEDTEPDGEGSKGEP	1013	IFKPEELR	1086	200	X	X	X	X	CRD1-CRD1
TETQAEDTEPDGEGSKGEP	1014	SEPMEVDEKKPEVK	1051	50					CRD1-CRD1
TETQAEDTEPDGEGSKGEP	1014	NPMDLSTIKR	1136	100	X		X		CRD1-BRD
TETQAEDTEPDGEGSKGEP	1014	KTGRPR	1314	50			X		CRD1-PHD
TETQAEDTEPDGEGSKGEP	1014	KMLDK	1505	50			X		CRD1-HAT
TETQAEDTEPDGEGSKGEP	1014	NANCSLPCQKMK	1794	50			X		CRD1-TAZ2
SEMMEEDLQASQVKEETDIAE	1033	TETQAEDTEPDGEGSKGEP	1014	200	X	X	X	X	CRD1-CRD1
SEMMEEDLQASQVKEETDIAE	1033	SEPMEVDEKKPEVK	1043	150	X	X	X		CRD1-CRD1
EETDIAEQSEPMEVDEKKPEVK	1036	SEMMEEDLQASQVK	1030	50	X				CRD1-CRD1
KPEVKVEVK	1056	IQKELEEK	668	50			X		CRD1-KIX
VEVEEEEESSNGTASQSTSPSQ	1060	IFKPEELR	1086	50			X		CRD1-desordered
KIFKPEELR	1082	KTGRPR	1314	100	X		X		CRD1-PHD
KIFKPEELRQALMPTLEALYR	1083	MVKWGLGLDDEGSSQGEP	1744	50			X		CRD1-ZZ
IFKPEELR	1086	KTGRPR	1314	150		X	X	X	CRD1-PHD
VYKFCCK	1176	IPKPKR	1497	50			X		PHD-HAT
KYEFSPQLCCYQK	1203	KMLDK	1505	50			X		BRD-HAT
KTGRPR	1314	KKEESTAASETTEGSQGD	1564	50			X		PHD-AL
KENKFSAK	1320	PQSQPPHSSPSR	2354	50			X		PHD-desordered
KENKFSAK	1320	PQSQPPHSSPSR	2361	50			X		PHD-desordered
LGNHLEDRVKNKFLR	1344	FSAKR	1327	100	X		X		HAT-HAT
VNKFLR	1344	KMLDK	1505	100	X		X		HAT-HAT
VVASDDK	1364	VVQHTKGCKR	1806	50			X		HAT-TAZ2
TVEVKPGMK	1368	VVASDDK	1364	100	X			X	HAT-HAT
CLRTAVYHEILIGYLEYVKK	1447	KNNKK	1588	50			X		HAT-AL
KLGYVTGHIWACPPSEGDDYIFH	1463	VNKFLR	1344	100		X	X		HAT-AL
LQEWYK	1503	PTSQVPAPPPPAQPPAAVE	1939	50			X		HAT-desordered
LQEWYKK	1504	IPKPKR	1497	100		X	X		HAT-HAT
MLDKAFAER	1509	VNKFLR	1344	50		X			HAT-HAT
MLDKAFAER	1509	IPKPKR	1497	150	X	X	X		HAT-HAT
MLDKAFAER	1509	LQEWYKK	1504	100	X		X		HAT-HAT
IIHDYKIDIFK	1521	QATEDRLTSAK	1533	100	X			X	HAT-HAT
IIHDYKIDIFKQATEDR	1524	LQEWYKK	1504	200	X	X	X	X	HAT-HAT
DIFKQATEDRLTSAK	1524	IIHDYK	1519	100	X			X	HAT-HAT
DIFKQATEDRLTSAK	1524	SALSSLESLVGDITGDTLEK	2418	50			X		HAT-desordered
ELPYFEGDFWPNVLEESIKELEQ	1554	TVEVKPGMK	1372	50			X		HAT-HAT
ELPYFEGDFWPNVLEESIKELEQ	1564	EESTAASETTEGSQGD	1569	50			X		HAT-HAT
KNNKKTNK	1588	NKSSISR	1597	50			X		HAT-HAT
NNKKTNK	1591	NAKKK	1586	50			X		AL-AL
NNKKTNK	1591	NKSSISR	1597	50			X		AL-AL
KPSMPNVSNDLSQK	1607	NKSSISR	1597	100	X		X		AL-AL
KPSMPNVSNDLSQK	1607	LYATMEKHK	1627	100		X	X		AL-AL
MVKWGLGLDDEGSSQGEPSQK	1744	KENKFSAK	1320	150		X	X	X	ZZ-PHD
SHAHKMVKWGLGLDDEGSSQK	1744	SPQESR	1763	100	X		X		ZZ-desordered
MVKWGLGLDDEGSSQGEPSQK	1744	NANCSLPCQKMK	1794	100		X	X		ZZ-TAZ2
WGLGLDDEGSSQGEPSQKSPQ	1762	SHAHKMVK	1741	150	X	X	X		ZZ-TAZ2
WGLGLDDEGSSQGEPSQKSPQ	1762	NANCSLPCQKMK	1794	50			X		ZZ-TAZ2
SPQESRR	1763	SSISRANK	1598	50			X		ZZ-AL
NANCSLPCQKMK	1797	KENKFSAK	1320	100		X	X		TAZ2-HAT
NANCSLPCQKMK	1797	SHAHKMVK	1741	50			X		ZZ-TAZ2
NANCSLPCQKMK	1797	SPAPVQSPR	2351	100		X	X		TAZ2-desordered
GCKRKTNGGCPVCK	1809	IFKPEELR	1086	50			X		TAZ2-CRD1
RKTNGGCPVCK	1811	VVQHTKGCKR	1806	100		X	X		TAZ2-TAZ2
HCQENKCPVPFCLNIKHKL	1832	IIHDYKIDIFKQATEDRLTSAK	1520	100		X		X	TAZ2-HAT
MATMNT	1871	NAKKKNNK	1584	50			X		desordered-AL
PTSQVPAPPPPAQPPAAVEAFA	1940	PEELRQALMPTLEALYR	1102	50			X		desordered-CRD1
EAQQQHLRY	1973	VNINNSMPPGR	1980	50	X				desordered-desordered
VNINNSMPPGRGTMGTPGSON	1980	KLIQQQLVLLHAK	352	50			X		desordered-TAZ1
TLKSPSSPQQQQVLNLIK	2073	SNPQLMAAFIKQR	2102	50				X	desordered-desordered
PQSQPPHSSPSR	2356	IIHDYKIDIFK	1520	50			X		desordered-HAT

### 3.2.3 Implémentation des résultats XL-MS sur la structure cristallographique du cœur catalytique de P300

Quelques données obtenues par DRX sont disponibles dans la littérature sur le cœur catalytique des protéines CBP et P300 [30, 37]. Pour nous aider dans l'interprétation de ces 1<sup>ers</sup> résultats, nous disposons ainsi d'une structure sur CBP de type souris (PDB :4U7G), et d'une structure P300 de type humain (PDB : 4BHW). Ces études ont montré que les deux protéines adopteraient une structure très similaire. (cf. figure 116 ). A ce jour, peu d'études ont permis de mettre en évidence une différence majeure sur leurs rôles biologiques.[37, 42]



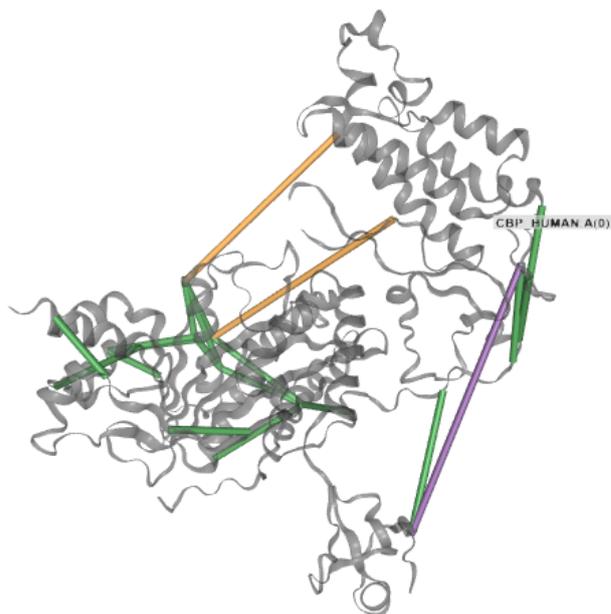
**Figure 116 : Superposition de la structure du cœur catalytique de CBP et de P300 de type souris montrant la légère différence de structure du cœur catalytique entre les deux protéines. Extrait de [37]**

Pour cette étude, nous avons travaillé avec une protéine CBP humaine. La bonne homologie de structure entre P300 et CBP de souris ainsi qu'une forte homologie de séquence entre CBP humains et souris (95% par Blast, Annexe 9) nous ont incité à choisir d'implémenter nos données XL-MS sur la structure cristallographique CBP de souris. Pour cela, nous avons rassemblé les peptides pontés issus du pontage BS3 et BS(PEG)<sub>5</sub> obtenus à partir de la digestion trypsique et utiliser le logiciel XiView pour localiser les pontages sur la structure cristallographique correspondante.

a

SeqPos1	SeqPos2	Highest Score	3D Distance
1082	1314	100.0000	14.05
1083	1314	200.0000	14.03
1083	1315	200.0000	16.79
1083	1744	50.0000	46.54
1086	1314	150.0000	22.09
1176	1497	50.0000	36.53
1203	1505	50.0000	42.25
1320	1744	150.0000	23.67
1327	1344	100.0000	16.58
1344	1463	100.0000	19.17
1344	1505	100.0000	19.99
1344	1509	100.0000	16.19
1364	1368	100.0000	13.12
1372	1554	50.0000	11.74
1497	1504	100.0000	10.34
1497	1505	50.0000	12.22
1497	1509	150.0000	18.12
1504	1509	100.0000	8.32
1504	1524	200.0000	15.09
1505	1524	100.0000	18.61
1519	1524	100.0000	9.94
1521	1533	100.0000	17.71

b



**Figure 117 : Implémentation des données XL-MS sur la structure cristallographique du cœur catalytique de CBP de type souris (PDB : 4U7G). (a) liste des peptides pontés identifiés sur la structure. (b) Implantation des pontages chimiques sur la structure du cœur catalytique de CBP de type souris.**

Sur les 87 peptides pontés uniques identifiés, seulement 19 peuvent être implémentés sur la structure de CBP, situés sur le cœur catalytique (figure 117). On remarque que toutes les distances mesurées pour ces peptides cross-linkés se situent entre 8 et 30 Å, excepté trois d'entre elles (peptide 1083-1744, 1176-1497 et 1203-1505) qui atteignent 46 Å. Ces distances sont ainsi parfaitement en accord avec la distance maximale d'élongation de nos deux agents pontant. En effet, il a été démontré que la distance d'élongation maximale pouvant être atteinte par le BS3 est de 35 Angströms.<sup>[43]</sup> Ces résultats sont très encourageants, car ils montrent que, dans la zone structurée de CBP (cœur catalytique), les données de cross-link sont cohérentes, et qu'ainsi notre protocole semble être non destructeur et approprié pour sa mission. Ces résultats nous encouragent donc à considérer les données XL-MS obtenues dans les zones désordonnées avec optimisme.

### **3.2.4 Comparaison entre la digestion trypsique et la digestion chymotrypsique**

Dans un second temps, nous avons comparé les données XL-MS obtenues selon l'enzyme de digestion utilisée afin de voir s'il était possible d'obtenir des informations structurales complémentaires sur la protéine.

Après validation manuelle des résultats, très peu de peptides pontés ont été identifiés après digestion chymotrypsique. En effet, 29 peptides pontés sont issus du pontage par le BS(PEG)<sub>5</sub> et uniquement 11 après pontage par le BS3. Soit la chymotrypsine procède difficilement à l'hydrolyse de CBP lorsqu'elle est pontée, soit l'identification des peptides pontés résultants est difficile par les logiciels choisis. En effet, la spécificité d'hydrolyse moins stricte que la trypsine augmente l'espace de recherche dans les bases de spectres cross-linkés théoriques générées par les logiciels et limite ainsi leur capacité à converger vers une solution fiable.

La majorité des peptides pontés ont été identifiés uniquement par un seul logiciel suggérant ainsi que l'interprétation des peptides pontés obtenus par digestion chymotrypsique est très délicate. En effet, uniquement 5 peptides pontés par le BS(PEG)<sub>5</sub> ont été identifiés par au moins deux logiciels et aucun de manière commune pour le BS3.

**Tableau 18 : Listes des peptides pontés obtenus avec le BS(PEG)<sub>5</sub> après digestion chymotrypsique. En jaune sont les peptides observés par plusieurs logiciels.**

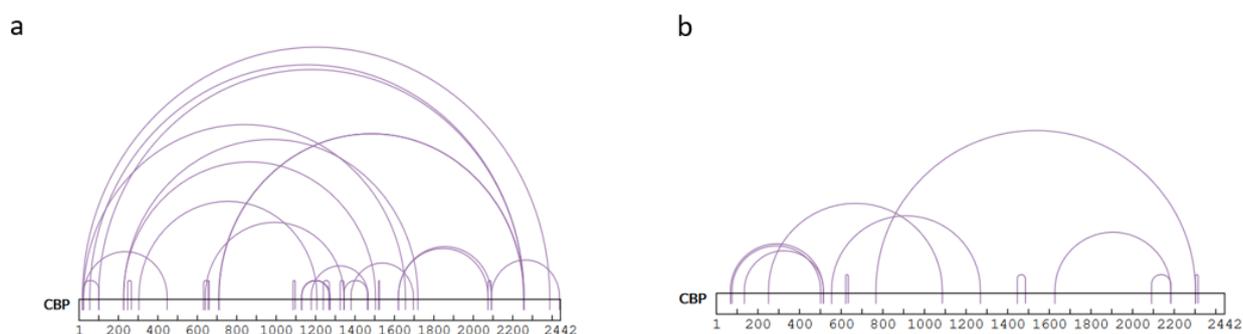
Peptide1	Position1	Peptide2	Position2	Score arbitraire	XLinkX	xiSEARCH	pLink	MeroX
SSPGFSANDSTDF	18	GGQAQGPNSANMASL	100	50		X		
SSPGFSANDSTDF	18	AVTMASSIDQGHGNPEQSAML	2388	50		X		
SANDSTDFGSLFDLENDL	23	GSPASGIQNTIGSVGTGQQNATSL	447	50		X		
SANDSTDFGSLFDL	23	LSCcmDLMDGRDAFL	1656	50		X		
NSGNLVPDAASKHKQL	55	QGSSMGQMAAQMGQL	2257	50		X		
PYPTPAMQGASSVL	226	YKKML	1504	50			X	
PTPAMQGASSVLAETL	227	HCTVCEYDL	1719	50		X		
NTAQAGGMAKM	266	TQVSPQM	247	50			X	
VNSLPTFPTDIKNTSVTNVPM	304	SPQTLCCYGKQLCTIPRDAAY	1207	50				X
VAYAKKVEGDMY	632	HLLAEKIY	657	50			X	
YESANSRDEYY	641	EDRVNKF	1344	50			X	
HLLAEKIY	657	KIQKEL	660	100	X		X	
SLPVNRMQVSQGMNSF	709	PLQGSSMGQMAAQMGQLGQMGQPGL	2255	50		X		
RQALMPTLEALY	1097	KPEEL	1086	50			X	
CSKLAEFVEQIEDPVMQSLGY	1178	VTGHIWACPPSEGDDY	1467	50		X		
EKKKNDTLDPEPF	1270	DIVKNPM	1130	50			X	
EKKKNDTLDPEPF	1270	HFCEKCF	1239	100	X		X	
EKKKNDTLDPEPF	1274	DIVKNPM	1130	50			X	
GNHLEDRVNKF	1344	SAKRL	1325	100	X		X	
GNHLEDRVNKF	1344	VKKLGY	1463	100	X		X	
VDSGEMSESF	1380	HTQGQDRF	1696	50		X		
AERIIHDYKDIF	1520	KQATEDRL	1524	100	X		X	
KSPSPQQQQVQL	2075	SQKLY	1620	50			X	
KSPSPQQQQVQL	2075	NILKSNPQLM	2091	50			X	
KSNPQLM	2092	SQKLY	1620	50			X	
QGSSMGQMAAQMGQL	2257	GGQAQGPNSANMASL	100	50		X		
QGSSMGQMAAQMGQLGQMGQPGL	2257	PVNRMQVSQGMNSF	711	50		X		
SLVGDTTGDITLEKF	2437	KSNPQLM	2092	50			X	

**Tableau 19 : Listes des peptides pontés obtenus par le BS3 après digestion chymotrypsique.**

Peptide1	Position1	Peptide2	Position2	Score arbitraire	XLinkX	xiSEARCH	pLink	MeroX
PDAASKHKQLSELLRGGSSINPGIGNVSASS	68	DTQLQPQVPGQQPAQPQTHQQMRTLNLPLGN	514	50			X	
LSQGDSSAPSLPKQAASSTSGPTPAASQALNPQ	135	DTQLQPQVPGQQPAQPQTHQQMRTLNLPLGN	514	50			X	
TGHAGLNTAQAGGMAKMoxGITGNTSPF	251	KPEELRQALMPTLEAL	1086	50				X
PQTHQQMRTLNLPLGNPNMNPAGGITDQO	501	GNVSASSPVQQLGGQAQGPNSANMoxAS	75	50		X		
KDRRMENLVAY	622	AKKVEGDMY	634	50	X			
EKKKNDTLDPEPF	1269	PTSLGATNPLM	554	50			X	
HCcmHPPDQKIPKPKRLQEWYKMoX	1485	RTAVYHEILIGY	1446	50				X
NPGHNPNMASM	2184	ATMEKHKEVF	1627	50			X	
NPGHNPNMASM	2184	KSNPQLMAAF	2091	50			X	
QQQQMKQIQISPGQPNMSPQQHML	2302	NSMGSVPGMAISPSRMPQPPNMM	767	50			X	
QQRILQQQMKQIQISPGQPNM	2302	SPQQHMLSGQPQASHLPQQIATSL	2315	50				X

Sur CBP cross-linkée avec le BS(PEG)<sub>5</sub>, la majorité des peptides pontés identifiés impliquent les extrémités N- et C-terminales désordonnées de la protéine. Très peu d'informations sont obtenues sur le cœur catalytique. Même si Mascot révèle la présence de peptides linéaires chymotrypsines, nous n'avons aucune idée de leur abondance et il n'est pas exclu que la protéolyse du cœur catalytique ponté ait été limitée. Des expériences complémentaires sont nécessaires pour pouvoir conclure.

Ces peptides ont essentiellement été obtenus au sein des zones désordonnées de la structure et ont uniquement été identifiés par un seul logiciel rendant leur validation ambiguë.



**Figure 118 : Carte de proximité de la protéine CBP obtenue après digestion par la chymotrypsine. (a) après pontage par le BS(PEG)<sub>5</sub>. (b) après pontage avec le BS3.**

En conclusion, la digestion chymotrypsique reste à optimiser et n'a pas permis de confirmer les zones pontées détectées auparavant après digestion trypsique. Une seconde possibilité serait d'effectuer une double digestion en utilisant deux enzymes différentes. En effet, des protocoles expérimentaux développés pour l'identification des peptides pontés ont mis en place l'utilisation de la double digestion afin d'améliorer l'efficacité d'identification des zones pontées sur les complexes biologiques. [44-46]

### **3.2.5 Difficultés d'obtenir des résultats reproductibles.**

Au cours de cette étude, nous n'avons travaillé qu'à partir de deux réplicats et nous avons obtenus une reproductibilité biologique de 36% pour les échantillons digérés par la trypsine. Cette faible reproductibilité biologique peut s'expliquer car :

- La protéine possède une structure flexible, elle est donc présente en solution sous la forme de multiples conformations. La flexibilité de la structure peut entraîner une modification ponctuelle des sites accessibles par l'agent pontant. Ainsi, chaque peptide cross-linké est présent en très faible abondance car au moment de la réaction de pontage, la structure n'est pas stabilisée.
- Une réactivité compétitive peut parfois être observée lorsque des sites pouvant être pontés sont très proches. C'est notamment le cas de CBP qui possède beaucoup de lysines proches sur la séquence primaire ainsi que de S / T. Cette proximité peut alors entraîner une diminution de l'abondance de chacun des peptides cross-linkés

qui sont déjà peu concentrés dans le mélange peptidique, ce qui dilue l'information résultante.

- L'efficacité de la protéolyse pourrait varier selon le degré et la position des pontages d'une préparation à l'autre

C'est pourquoi des études supplémentaires seront nécessaires afin d'améliorer la reproductibilité biologique et pouvoir en retirer des informations pertinentes. D'autre part, l'utilisation d'agents pontants clivables tels que le DSBU serait très certainement un atout afin de valider l'identité des peptides pontés.

## 4. Conclusions et Perspectives

Ainsi, un protocole XL-MS faisant appel à deux agents de pontage : le BS3 et le BS(PEG)<sub>5</sub> possédant des longueurs de bras espaceurs différentes (11.4 et 21.5Å respectivement) a permis d'obtenir des premières informations sur la structure d'une protéine partiellement désordonnée, CBP. Une comparaison des résultats XL-MS avec les éléments de structure DRX dont nous disposons montre que les distances entre acides aminés pontés au sein du cœur catalytique telles que déduites des expériences XL-MS sont cohérentes et semble reproductible avec les deux agents pontant. Ceci nous encourage à considérer avec intérêt les candidatures de peptides cross-linkés impliquant des régions désordonnées qui semblent être des zones moins reproductibles.

A partir de ces informations, une première carte de proximité a pu être obtenue suggérant la proximité entre certains domaines de CBP. Néanmoins, plusieurs réplicats devront être réalisés à partir d'agents pontants clivables afin de renforcer ces 1<sup>res</sup> hypothèses en appliquant une dimension d'analyse quantitative pour tenter de distinguer des ensembles conformationnels.

Avec cette étude, nous nous sommes ainsi rendu compte du potentiel de la XL-MS pour aborder la caractérisation des protéines désordonnées qui reste, quelle que soit la technique d'analyse, un défi.

## 5. Bibliographie

1. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains*. Science, 1973. **181**(4096): p. 223-230.
2. Karush, F., *Heterogeneity of the Binding Sites of Bovine Serum Albumin*. Journal of the American Chemical Society, 1950. **72**(6): p. 2705-2713.
3. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(2): p. 98-104.
4. Kriwacki, R.W., et al., *Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity*. Proceedings of the National Academy of Sciences, 1996. **93**(21): p. 11504.
5. Fletcher, C.M. and G. Wagner, *The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein*. Protein science : a publication of the Protein Society, 1998. **7**(7): p. 1639-1642.
6. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nature Reviews Molecular Cell Biology, 2005. **6**(3): p. 197-208.
7. Dunker, A.K., et al., *Intrinsic disorder and protein function*. Biochemistry, 2002. **41**(21): p. 6573-6582.
8. Uversky, V.N., et al., *Protein Intrinsic Disorder and Human Papillomaviruses: Increased Amount of Disorder in E6 and E7 Oncoproteins from High Risk HPVs*. Journal of Proteome Research, 2006. **5**(8): p. 1829-1842.
9. Goldsbury, C., et al., *Amyloid Fibril Formation from Full-Length and Fragments of Amylin*. Journal of Structural Biology, 2000. **130**(2): p. 352-362.
10. Telling, G.C., et al., *Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein*. Cell, 1995. p. 79-90.
11. Crowther, R.A. and M. Goedert, *Abnormal Tau-Containing Filaments in Neurodegenerative Diseases*. Journal of Structural Biology, 2000. **130**(2): p. 271-279.
12. Martinelli, A.H.S., et al., *Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies*. International journal of molecular sciences, 2019. **20**(6): p. 1322.
13. Spillantini, M.G. and M. Goedert, *The  $\alpha$ -Synucleinopathies: Parkinson's Disease, Dementia with Lewy Bodies, and Multiple System Atrophy*. Annals of the New York Academy of Sciences, 2000. **920**(1): p. 16-27.
14. Christopher, J.M., et al., *An evolving understanding of nuclear receptor coregulator proteins*. Journal of Molecular Endocrinology, 2013. **51**(3): p. T23-T36.
15. Chan, H.M. and N.B. La Thangue, *p300/CBP proteins: HATs for transcriptional bridges and scaffolds*. Journal of Cell Science, 2001. **114**(13): p.2363-2373.
16. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development*. Genes & development, 2000. **14**(13): p.1553-1577.
17. Giordano, A. and M.L. Avantaggiati, *p300 and CBP: partners for life and death*. (0021-9541).
18. Kamei, Y., et al., *A CBP Integrator Complex Mediates Transcriptional Activation and AP-1 Inhibition by Nuclear Receptors*. Cell, 1996. **85**(3): p. 403-414.
19. Arany Z Fau - Sellers, W.R., et al., *E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators*. Cell, 1994. **77**(6): p. 799-800.
20. Petrij, F., et al., *Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP*. Nature, 1995. **376**: p. 348-351.
21. Ganly, P., C.M. Walker Lc Fau - Morris, and C.M. Morris, *Familial mutations of the transcription factor RUNX1 (AML1, CBFA2) predispose to acute myeloid leukemia*. Leukemia & lymphoma, 2004. **45**(1): p. 1-10.
22. Pasqualucci, L., et al., *Inactivating mutations of acetyltransferase genes in B-cell lymphoma*. Nature, 2011. **471**(7337): p. 189-195.

23. Korzus, E., et al., *Transcription Factor-Specific Requirements for Coactivators and Their Acetyltransferase Functions*. Science, 1998. **279**(5351): p. 703.
24. Kurokawa, R., et al., *Differential Use of CREB Binding Protein-Coactivator Complexes*. Science, 1998. **279**(5351): p. 700.
25. Westin, S., et al., *Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators*. Nature, 1998. **395**(6698): p. 199-202.
26. Chakravarti, D., et al., *Role of CBP/p300 in nuclear receptor signalling*. Nature, 1996. **383**(6595): p. 99-103.
27. Sheppard, H.M., et al., *Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1*. Molecular and cellular biology, 2001. **21**(1): p. 39-50.
28. Yi, P., et al., *Structure of a biologically active estrogen receptor-coactivator complex on DNA*. Molecular cell, 2015. **57**(6): p. 1047-1058.
29. Park, S., et al., *The CH2 domain of CBP/p300 is a novel zinc finger*. FEBS letter, 2013. **587**(16): p. 2506-2511.
30. Delvecchio, M., et al., *Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation*. Nature Structural & Molecular Biology, 2013. **20**(9): p. 1040-1046.
31. Filippakopoulos, P., et al., *Histone recognition and large-scale structural analysis of the human bromodomain family*. Cell, 2012. **149**(1): p. 214-231.
32. Kraus, W.L., J.T. Manning Et Fau - Kadonaga, and J.T. Kadonaga, *Biochemical analysis of distinct activation functions in p300 that enhance transcription initiation with chromatin templates*. Journal of Bioscience and Bioengineering, 2020. **129**(1): p.6-15.
33. Thompson, P.R., et al., *Regulation of the p300 HAT domain via a novel activation loop*. Nature Structural & Molecular Biology, 2004. **11**(4): p.308-315.
34. Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator*. Nature, 2008. **451**: p. 846-850.
35. Dancy, B.M., et al., *Live-cell studies of p300/CBP histone acetyltransferase activity and inhibition*. Chembiochem, 2012. **13**(14): p. 2113-2121.
36. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*. Nature reviews. Molecular cell biology, 2015. **16**(1): p. 18-29.
37. Park, S., et al., *Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation*. Proceedings of the National Academy of Sciences, 2017. **114**(27): p. E5335.
38. Dyson, H.J. and P.E. Wright, *Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300*. The Journal of biological chemistry, 2016. **291**(13): p. 6714-6722.
39. Wickham, T.J. and G.R. Nemerow, *Optimization of Growth Methods and Recombinant Protein Production in BTI-Tn-5B1-4 Insect Cells Using the Baculovirus Expression System*. Biotechnology Progress, 1993. **9**(1): p. 25-30.
40. Leavell, M.D., et al., *Strategy for selective chemical cross-linking of tyrosine and lysine residues*. Journal of the American Society for Mass Spectrometry, 2004. **15**(11): p. 1604-1611.
41. Kalkhof, S. and A. Sinz, *Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters*. Analytical and Bioanalytical Chemistry, 2008. **392**(1): p. 305-312.
42. Kasper, L.H., et al., *Double null cells reveal that CBP and p300 are dispensable for p53 targets p21 and Mdm2 but variably required for target genes of other signaling pathways*. Cell cycle (Georgetown, Tex.), 2011. **10**(2): p. 212-221.
43. Merkley, E.D., et al., *Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances*. Protein science : a publication of the Protein Society, 2014. **23**(6): p. 747-759.

44. Mendes, M.L., et al., *An integrated workflow for crosslinking mass spectrometry*. *Molecular systems biology*, 2019. **15**(9): p. e8994-e8994.
45. Iacobucci, C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions*. *Nature Protocols*, 2018. **13**(12): p. 2864-2889.
46. Dau, T., et al., *Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry*. *Analytical Chemistry*, 2019. **91**(7): p. 4472-4478.



## Chapitre VII : XL-MS, une aide à la compréhension des données de cryo-EM : vers la génération d'un modèle structural du complexe NuA4

Ce travail a été effectué dans le cadre d'une collaboration avec l'équipe du Dr. Patrick Schultz (Département de Biologie structurale Intégrative de l'IGBMC à Illkirch) et du Dr. Luc Negroni (responsable de la Plateforme Protéomique de l'IGBMC à Illkirch)

### 1. Introduction

L'acétylation est l'une des modifications post-traductionnelles qui régulent de nombreux processus biologiques.<sup>[1-3]</sup> Il s'agit d'une modification qui neutralise les charges positives et modifie la taille de la chaîne latérale du résidu entraînant un changement de conformation des protéines modifiées et leurs modes d'interactions avec leurs molécules cibles en particulier avec l'ADN.

Les histones acétyltransférases (HAT) existent souvent dans de grands complexes multimériques tels que le complexe SAGA-HAT<sup>[4-6]</sup> et le complexe Nucléosome Acétyltransférase de H4 (NuA4).<sup>[7-9]</sup> Il s'agit d'enzymes qui sont associées à l'activation de la transcription notamment en agissant sur le remodelage et la décondensation de la chromatine (la structure au sein de laquelle l'ADN se trouve empaqueté) permettant ainsi l'exposition de nombreux sites de liaison pour les protéines régulant la transcription.<sup>[10]</sup>

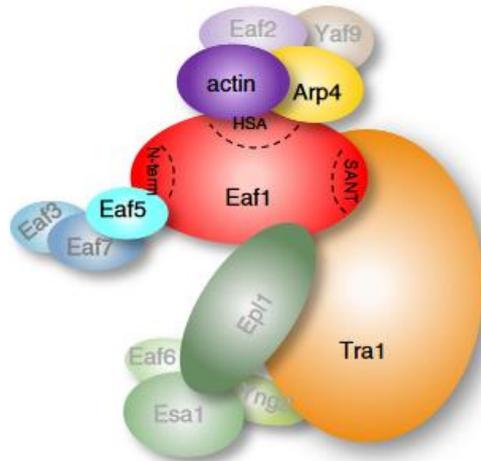
Le complexe NuA4 est un de ces complexes qui joue un rôle important dans les processus génomiques cruciaux en particulier dans la réparation et la transcription des dommages réalisés sur l'ADN.<sup>[11-13]</sup>

NuA4 est un complexe constitué de 13 sous-unités uniques dans les levures ayant une masse moléculaire de 1 MDa (figure 119). Une des caractéristiques de NuA4 est qu'il partage des sous-unités avec d'autres complexes protéiques impliqués dans les réactions de liaison à la chromatine et dans la modification des histones. Le complexe NuA4 est constitué de deux sous-complexes indépendants. Le premier se nomme piccolo et est constitué des sous-unités Esa1, Epl1, Yng2 et Eaf6.<sup>[14, 15]</sup> Le second sous-complexe se nomme TINTIN et est constitué des sous-unités Eaf5, Eaf7 et Eaf3.<sup>[16]</sup> Chez l'homme, les homologues de NuA4 et SWR1 sont physiquement associés pour former le complexe Tip60 qui regroupe les activités acétyltransférase et d'échange d'histone.<sup>[17]</sup>

Tra1, connue comme étant un module de recrutement dans le complexe SAGA.<sup>[18, 19]</sup> est la plus grosse sous-unité du complexe. Elle est constituée de 3744 acides aminés et joue un rôle important dans la régulation transcriptionnelle. Sa grande taille ainsi que le manque d'activité dans les complexes NuA4 et SAGA ont suggéré que Tra1 peut servir d'échafaudage pour l'assemblage du complexe. Les sous-unités Swc4, Yaf9, Arp4 et Act1 sont également connues pour jouer un rôle important dans le remodelage de la chromatine. La seule sous-unité totalement associée au complexe NuA4 est la sous-unité Eaf1 qui sert de plateforme et coordonne l'assemblage des quatre modules fonctionnels cités précédemment.<sup>[20]</sup>

Cependant, la compréhension de l'assemblage à l'échelle atomique de NuA4 a été jusqu'à présent entravée par des informations structurales limitées. Malgré des efforts qui se poursuivent, il n'a pas été possible d'obtenir des cristaux diffractants utilisables pour une étude par DRX. Quelques structures sur des sous-unités individuelles sont disponibles<sup>[21-23]</sup> Récemment, les structures cryo-EM de sous-complexes de NuA4 de *Saccharomyces cerevisiae* ont été reportées<sup>[17]</sup>: structure à 4.7 Å du complexe NuA4 en absence du module Piccolo et une structure à 7.6 Å en présence de Piccolo<sup>[17]</sup>. Ces modèles montrent que les modules TINTIN et Piccolo sont ancrés sur la sous-unité Tra1 et que la sous-unité Eaf1 constitue la plateforme de NuA4. Néanmoins, plusieurs zones de la carte de densité EM n'ont pas pu être attribuées à cette résolution, cela étant probablement lié à la présence de régions désordonnées.

En raison de la complexité élevée de sa composition, de la difficulté à exprimer un complexe natif en quantité compatible avec l'analyse structurale (quelques dizaines de µg à minima) et d'une inhomogénéité conformationnelle pouvant conduire à de l'agrégation partielle, aucune structure du système NuA4 humain dans son intégralité n'est disponible. C'est pourquoi des études intégratives combinant plusieurs techniques doivent être réalisées afin de construire, tel un puzzle, des modèles avec la meilleure résolution possible. Dans ce cadre, la XL-MS est une approche bien appropriée en permettant d'identifier les zones protéiques proches les unes des autres dans l'espace.



**Figure 119 : Représentation schématique du complexe NuA4. Extrait de Wang et al 2018.[17] Le sous-domaine Piccolo est représenté en vert.**

L'objectif de cette étude est de déterminer (i) l'identité des sous-unités qui interagissent et (ii) l'identité des domaines protéiques qui sont proches, afin de pouvoir obtenir des informations sur les sous-unités qui ne sont pas encore modélisées et poursuivre la construction d'un modèle structural actuellement proposé à partir d'images cryo-EM.

## 2. Matériels et méthode

### 2.1 La réaction de pontage chimique : utilisation du BS3

Les échantillons de NuA4 ont été exprimés dans un système *Pichia pasteuris*. Cette levure a la propriété de pousser en milieu très concentré, ce qui permet de limiter le volume de culture. L'extrait nucléaire des cellules est récupéré puis le complexe est purifié sur colonne d'affinité.<sup>[24]</sup> Il est intéressant de noter que le complexe ainsi purifié est un complexe endogène, non surexprimé. Il est constitué de 12 sous-unités que nous avons résumées dans le tableau 20 :

**Tableau 20 : Liste de l'ensemble des protéines présentent dans l'échantillon de NuA4.**

Protéine	Nombre d'acide aminés	Masse
Act1/Actin	376	41,6 kDa
Arp4/Actin	468	52,9 kDa
Uncharacterized bromodomain protein	568	64,7kDa
Eaf6	141	16,5kDa
Eaf1	1051	118,2kDa
Yaf9/Yeats4	219	25,1kDa
Yng2	261	29,8kDa
Swc4/Dmap1	565	64,7kDa
EPC/EPL1	752	86,6kDa
Esa1/Tip60	464	53,9kDa
TRA1	3825	437,2kDa
Eaf7	237	27,2kDa

Avant la réaction de pontage, un échange de tampon a été réalisé afin que le complexe soit présent dans un tampon approprié pour la réaction de cross-link : 20mM Hepes, 100mM NaCl à 4°C. Cet échange de tampon a été réalisé en utilisant des colonnes d'exclusion (MicroBiospin P 30 kDa).

La réaction de pontage a ainsi été réalisée avec le BS3 à la concentration de 3mM dans des conditions standards. En raison de la taille du complexe (~ 1MDa), il n'a pas été possible de suivre la réaction de pontage ni sur gel d'électrophorèse ni par MALDI-MS. Nous avons stoppé la réaction après 1h de réaction par l'ajout de Tris/Hcl 25mM. Après digestion par la trypsine, l'échantillon a été soumis à une analyse LC-MS/MS sur une orbitrap LTQ Elite en Top10. En 1<sup>ère</sup> intention, le logiciel pLink a été utilisé pour le traitement des données. En accord avec notre objectif de positionner les sous-unités dans la carte de densité EM, nous nous sommes focalisés en 1<sup>er</sup> lieu sur les pontages intermoléculaires.

Les résultats obtenus à partir de cette analyse furent décevants. En effet, seulement 42 peptides cross-linkés intermoléculaires uniques ont été identifiés dont une majorité impliquait les sous-unités EPL1 et Act1 ou Arp4, ou encore EPL1 et Tip60 déjà caractérisées. Ceci était donc peu informatif. Le relativement faible nombre de peptides cross-linkés intermoléculaires peut être lié à :

- Un taux de pontage insuffisant, ce qui semblait peu probable à la vue du nombre d'acides aminés modifiés de type « dead-end » obtenu par une analyse Mascot
- Une hydrolyse par la trypsine insuffisante sur le système cross-linké, réduisant ainsi la concentration déjà faible de chaque peptide cross-linké généré

- Un mélange peptidique si complexe que la dynamique de l'analyse LC-MS/MS ne parvient pas à refléter l'existence des peptides cross-linkés de faible abondance en mode d'acquisition DDA.

Afin d'améliorer notre set de données XL-MS, nous avons effectué une seconde expérience mettant en jeu un procédé d'enrichissement des peptides cross-linkés en amont de l'analyse MS. Deux concentrations de BS3 ont été choisies en tenant compte de quelques références de la littérature<sup>[25, 26]</sup> et de notre expérience de laboratoire : 3mM et 5mM. Ces deux analyses ont été effectuées à partir de la même préparation de NuA4.

## **2.2 Enrichissement des peptides cross-linkés par implémentation d'une étape de chromatographie SCX**

Nos collaborateurs de la Plateforme Protéomique de l'IGBMC ont mis en place un système chromatographique de séparation par échange d'ions à partir de colonnes « home-made » (2mm de diamètre interne, 15cm en longueur) en vue d'analyses protéomiques d'échantillons très complexes (analyse de type « shotgun »<sup>[27-29]</sup>. C'est donc naturellement vers ce système séparatif que nous nous sommes dirigés.

Le préfractionnement des peptides sur la colonne d'échange d'ions en SCX s'effectue par paliers, en introduisant des étapes d'élution par augmentation de la concentration en sels. Dans notre cas, 10 fractions d'élution ont été générées en faisant varier la concentration en KCl sur la colonne SCX entre 0 et 1000 mM. Chaque fraction d'élution a été dessalée sur cartouche de phase inverse C18 puis analysée par nanoLC-MS/MS sur l'orbitrap LTQElite en Top10 avec une résolution de 240 000 en MS1 et 15 000 en MS2.

## **2.3 Identification et positionnement des zones pontés de NuA4 après SCX**

Chacune des fractions de SCX a été analysée par LC-MS/MS pour les deux conditions de pontage 3mM et 5mM en BS3. Pour les requêtes, nous avons utilisé les mêmes critères que pour nos études sur les RNs et CBP et les 4 logiciels d'interprétation pLink, MeroX, XiSearch et XlinkX ont été confrontés (tolérance de masse de l'ion précurseur : 10ppm ; tolérance de masse des ions fragments :20ppm ; nombre de miscleavage : 4). Chaque ensemble de données a été produit en utilisant un FDR de 5%.

Une validation manuelle a été effectuée pour chaque spectre et a résulté en le rejet de 24% et 16% pour les pontages à 3mM et 5mM respectivement. Les données obtenues sur les deux séries de pontage à 3mM et 5mM de BS3 ont été traitées indépendamment puis rassemblées.

Les résultats montrent sans ambiguïté l'efficacité du préfractionnement des peptides par SCX en amont de l'analyse MS (cf. Tableau 21). En effet, ce sont maintenant 321 et 362 peptides intermoléculaires qui sont interprétés par pLink dans les conditions de pontage 3 et 5mM respectivement (avant validation manuelle). 2/3 des peptides cross-linkés correspondent à des peptides intramoléculaires et 1/3 à des peptides intermoléculaires.

**Tableau 21 : Nombre de peptides pontés identifiés par pLink (avant validation manuelle) en fonction de la fraction d'éluion après SCX.**

Fraction	Concentration KCl (mM)	3mM		5mM	
		Nombre de peptides identifiés par pLink	Nombre de peptides Interprotéine identifiés	Nombre de peptides identifiés par pLink	Nombre de peptides Interprotéine identifiés
0	0	2	1	51	13
1	50	1	0	13	2
2	75	3	1	22	13
3	100	10	0	22	1
4	125	32	3	32	3
5	200	42	16	44	17
6	250	195	51	176	56
7	400	236	78	348	88
8	600	241	75	301	86
9	800	168	50	215	52
10	1000	131	46	155	32

Ce tableau montre un enrichissement important pour les fractions 6 à 9, avec un taux d'identification de pontage encore élevé pour la fraction 10, ce qui laisse suggérer que certains peptides ont été retenus sur la phase échangeuse d'ions.

L'ensemble des résultats obtenus après traitement des données par les 4 logiciels pLink, XiSearch, MeroX et XlinkX est présenté dans un tableau récapitulatif (cf. Tableau 22).

Un score arbitraire a été attribué en fonction du nombre de logiciels qui ont conduit à une même interprétation par spectre MS/MS (« score 5mM » pour l'échantillon ponté par 5mM de BS3 et « score 3mM » pour l'échantillon ponté avec 3mM de BS3). Ce score arbitraire attribue 50 points par logiciel. Ainsi, sur l'ensemble de l'expérience (3 et 5mM) 212 peptides uniques ont été interprétés par XiSearch, 201, 192 et 64 par pLink, XlinkX et MeroX respectivement. 65% des peptides sont identifiés par au moins deux logiciels (score >50). Ceci est relativement faible et, comme nous l'avons observé pour les études précédentes, il est nécessaire de recourir à la combinaison de plusieurs algorithmes de traitement pour limiter les risques de perdre de l'information. La colonne notée « score présent 3 ET 5mM » correspond à la somme des scores arbitraires attribués pour un même peptide dans les deux conditions. Il reflète la reproductibilité à l'échelle peptidique pour les deux conditions de pontage. Ce score varie de 50 (peptide unique interprété par un seul logiciel dans une seule condition de pontage) à 400 (peptide unique interprété par les 4 logiciels et observé pour les deux conditions de pontage).

Outre l'inspection manuelle des spectres MS/MS, la validation des interprétations est consolidée par la redondance :

- Une même interprétation est fournie à partir de spectres MS/MS issus de différents états de charge
- Une même interprétation est proposée sur l'ion cross-linké sans et avec méthionine oxydée
- Une même interprétation est fournie par plusieurs logiciels (score arbitraire)
- Une même interprétation est retrouvée dans deux fractions de SCX consécutives

On observe que toutes les sous-unités sont représentées par plusieurs pontages intermoléculaires. 4 d'entre elles, Eaf1, EPL1, Tra1 et Tip60 sont impliquées dans 70% des pontages détectés.

**Tableau 22 : Liste de l'ensemble des peptides pontés identifiés sur NuA4 après pontage par du BS3 à 5mM et à 3mM. En orange apparaissent les peptides pontés identifiés par 4 logiciels, en bleu par 3 logiciels et en gris par 2 logiciels.**

Peptide 1	Protein 1	Peptide 2	Protein 2	Charge	Position 1	Position 2	Score 5mM	Fraction SCX	XLinkX	xiSEARCH	pLink	MeroX	score 3mM	score présent 3 ET 5mM
VATGKTHSTAYNSTLDPVA	TRA1	NKMoxNQVANR	Eaf1	4	3673	1017	200	7	X	X	X	X	200	400
TSMDLKR	Swc4/Dmap1	AVAEKLR	EPC/EPL1	3	424	448	200	7	X	X	X	X	200	400
NSAKSQGGMNVYDSDA	EPC/EPL1	VKISAPPERK	Act1/Actin	4	599	329	200	7;8	X	X	X	X	200	400
IEELKR	Eaf1	AVAEKLR	EPC/EPL1	3	268	448	200	7;9;10	X	X	X	X	200	400
TTKEDYQAIQR	TRA1	KYAQDEVPIRPDDPTDR	Eaf1	4	3177	173	200	7	X	X	X	X	150	350
TTKEDYQAIQR	TRA1	NKMNQVANR	Eaf1	4	3177	1017	200	6	X	X	X	X	150	350
SKDEPTDDKPR	Arp4/Actin	KWSLR	Eaf1	5	353	272	200	7	X	X	X	X	150	350
SHDAESEQEISPSKR	Incharacterize	SRTPVNKGIGYR	Eaf7	4	450	154	150	8	X	X	X	X	200	350
SESKTESR	TRA1	AVAEKLR	EPC/EPL1	3	428	448	150	6	X	X	X	X	200	350
SESKTESR	TRA1	KELHQK	Eaf6	4	428	15	150	6	X	X	X	X	200	350
NKMNQVANR	Eaf1	SQELKR	TRA1	4	1017	2880	200	6	X	X	X	X	150	350
KGYNsgNYPNR	EPC/EPL1	KELHQK	Eaf6	3	374	15	150	7	X	X	X	X	200	350
FASTFLAPVNVKSEEPYK	Incharacterize	LKQDDVDGSSR	Eaf7	4	490	170	150	8	X	X	X	X	200	350
VANPREEVKLPAAATEASISR	TRA1	KELHQK	Eaf6	5	3349	15	150	8	X	X	X	X	150	300
VANPREEVKLPAAATEASISR	TRA1	SSNGHSTPQVR	Eaf1	4	3349	943	150	8;9	X	X	X	X	150	300
TTKEDYQAIQR	TRA1	KGYNsgNYPNR	EPC/EPL1	3	3177	374	200	6	X	X	X	X	100	300
SHDAESEQEISPSKR	Incharacterize	LKQDDVDGSSR	Eaf7	3	450	170	200	6	X	X	X	X	100	300
SESKTESR	TRA1	IHQEKLQPGVSLR	Swc4/Dmap1	4	428	482	150	8	X	X	X	X	150	300
SESKTESR	TRA1	KYAQDEVPIRPDDPTDR	Eaf1	4	428	173	100	7;8	X	X	X	X	200	300
NSAKSQGGMNVYDSDA	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	4	599	215	150	7	X	X	X	X	150	300
NKMNQVANR	Eaf1	AVAEKLR	EPC/EPL1	4	1017	448	150	7	X	X	X	X	150	300
LKYDRDK(2)	Eaf1	AVAEKLR(5)	EPC/EPL1	4	706	448	100	7	X	X	X	X	150	300
KGYNsgNYPNR	EPC/EPL1	KRYQFTDNR	Eaf6	4	374	76	150	7	X	X	X	X	150	300
KELLHATR	TRA1	TEVDNENEHDDNDIFK	Eaf1	4	321	378	150	7	X	X	X	X	150	300
IEELKR	Eaf1	TSMDLKR	Swc4/Dmap1	3	268	424	150	7	X	X	X	X	150	300
GSEELAKNSPYLFEK(7)	Swc4/Dmap1	SESKTESR(4)	TRA1	4,5	104	428	100	6	X	X	X	X	200	300
FEDSSEKDDSDPYVcmFR	EPC/EPL1	EDEIKK	Esa1/Tip60	3	266	165	150	5;6;7;8;9	X	X	X	X	150	300
FEDSSEKDDSDPYVcmFR	EPC/EPL1	TGGSMIQSHNEVSKacVR	Esa1/Tip60	4	266	181	200	6;7	X	X	X	X	100	300
ELKQTR	EPC/EPL1	NSNDETRNDSLEDDTD	Eaf7	5	305	201	150	6	X	X	X	X	150	300
AVAEKLR	EPC/EPL1	TTKEDYQAIQR	TRA1	3	448	3177	150	6	X	X	X	X	150	300
VLPSAAGAR	EPC/EPL1	LDEWIEKDRIDYK	Esa1/Tip60	4	2	79	100	6;7;8;10	X	X	X	X	150	250
TQSTTKR	Eaf1	AVAEKLR	EPC/EPL1	3	638	448	150	6	X	X	X	X	100	250
TFNKLcmR	TRA1	KYAQDEVPIRPDDPTDRDL	Eaf1	5	2279	173	100	7;9;10	X	X	X	X	150	250
TASPEPIEIQEAK(1)	Esa1/Tip60	KISVK(1)	EPC/EPL1	4,5,6	2	15	100	7	X	X	X	X	150	250
SHDAESEQEISPSKR(15)	Incharacterize	TPKDR(3)	Eaf7	3	450	166	100	6	X	X	X	X	150	250
QTLQVFKQSDIADLETEDQC	EPC/EPL1	GQHICcmLTDQVVKQYDR	Esa1/Tip60	5	26	430	100	8	X	X	X	X	150	250
NSAKSQGGMNVYDSDA	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	3	599	221	100	7;8	X	X	X	X	150	250
LDEWIEKDR(7)	Esa1/Tip60	KISVK(1)	EPC/EPL1	4	79	15	50	7	X	X	X	X	200	250
KTDSSSGRR(1)	Arp4/Actin	DKALK(2)	Swc4/Dmap1	4,5	343	146	100	6	X	X	X	X	150	250
KPLIFLTDEEIKER(12)	Eaf1	SQELKR(5)	TRA1	4	351	2880	100	8,9	X	X	X	X	150	250
KISVK	EPC/EPL1	QEDTSVLDLGGKDSAGQDV	Esa1/Tip60	4	15	133	150	7	X	X	X	X	100	250
KGYNsgNYPNR(1)	EPC/EPL1	SQELKR(5)	TRA1	4	374	2880	100	6	X	X	X	X	150	250
KGYNsgNYPNR	EPC/EPL1	IEELKR	Eaf1	4	374	268	150	7	X	X	X	X	100	250
KGYNsgNYPNR	EPC/EPL1	SSNGHSTPQVR	Eaf1	4	374	943	150	6	X	X	X	X	100	250
KGYNsgNYPNR	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	4	374	221	100	8	X	X	X	X	150	250
KFELAAK(7)	Swc4/Dmap1	IEELKR(5)	Eaf1	4	351	268	100	8	X	X	X	X	150	250
KFELAAK	Swc4/Dmap1	KYAQDEVPIRPDDPTDR	Eaf1	4	345	173	100	9	X	X	X	X	150	250
KESDR	EPC/EPL1	LNQELTAMoxNPLKVR	Arp4/Actin	4	453	415	150	8	X	X	X	X	100	250
KELHQK	Eaf6	KESDR	EPC/EPL1	4	15	453	150	6	X	X	X	X	100	250
FEDSSEKacDSDPYVcmFR	EPC/EPL1	TGGSMoxIQSHNEVSKVR	Esa1/Tip60	4	263	182	150	7	X	X	X	X	100	250
FADSVLPKNIR	TRA1	NKMoxNQVANR	Eaf1	4	3368	1017	100	9	X	X	X	X	150	250
ESPSKESER	Incharacterize	LKQDDVDGSSR	Eaf7	4	429	170	150	5	X	X	X	X	100	250
ESPSKESER	Incharacterize	SRTPVNKGIGYR	Eaf7	5	429	154	100	7	X	X	X	X	150	250
DYDKYR	EPC/EPL1	KQFER	Esa1/Tip60	3	95	228	150	6	X	X	X	X	100	250
AVAEKLR(5)	EPC/EPL1	KYAQDEVPIRPDDPTDR(1)	Eaf1	4	448	173	150	8	X	X	X	X	100	250
AVAEKLR	EPC/EPL1	KELHQK	Eaf6	4	448	15	100	7	X	X	X	X	150	250
VPTYFPDEAKFEER	Arp4/Actin	MQKELTALAPSSMK	Act1/Actin	4	221	316	200	6;7	X	X	X	X	0	200
VLPSAAGAR	EPC/EPL1	SSNGHSTPQVR	Eaf1	3	2	943	100	8	X	X	X	X	100	200
VATSQNTTPVTNR	Eaf1	TTKEDYQAIQR	TRA1	3	800	3176	100	5	X	X	X	X	100	200
VATGKTHSTAYNSTLDPVA	TRA1	VLPSAAGAR	EPC/EPL1	4	3673	2	100	9	X	X	X	X	100	200
VANPREEVKLPAAATEASISR	TRA1	NKMoxNQVANR	Eaf1	5	3349	1017	150	9	X	X	X	X	50	200
VANPREEVKLPAAATEASISR	TRA1	VATSQNTTPVTNR	Eaf1	5	3349	801	100	8	X	X	X	X	100	200
TTKEDYQAIQR	TRA1	PEPTK	Eaf6	3	3177	2	100	6	X	X	X	X	100	200
TTKEDYQAIQR	TRA1	QITLTKAMoxQAR	Eaf1	4	3177	1010	0	8	X	X	X	X	200	200
TTKEDYQAIQR	TRA1	SSNGHSTPQVR	Eaf1	3	3177	943	100	5	X	X	X	X	100	200
TSMDLKR(6)	Swc4/Dmap1	KESDR(1)	EPC/EPL1	4,5	424	453	50	6	X	X	X	X	150	200
TSMDLKR(6)	Swc4/Dmap1	VIEEINVQDR(9)	Yaf9/Yeats4	3	424	201	50	7	X	X	X	X	150	200
TGGSMIQSHNEVSKVR(14)	Esa1/Tip60	VLPSAAGAR(1)	EPC/EPL1	4	182	2	100	9,1	X	X	X	X	100	200
SQELKR	TRA1	KYAQDEVPIRPDDPTDRDL	Eaf1	5	2880	173	50	8;10	X	X	X	X	150	200
RKEAIEVGHEVFLQR	EPC/EPL1	AKALQEEK	Yng2	4	322	80	100	9	X	X	X	X	100	200
NSAKSQGGMNVYDSDA	EPC/EPL1	QLLNHKGVVVPVYK	Arp4/Actin	4	599	198	200	8	X	X	X	X	0	200
KYAQDEVPIRPDDPTDR(1)	Eaf1	TSMDLKR(6)	Swc4/Dmap1	4	173	424	50	8	X	X	X	X	150	200
KWNALPQR(1)	TRA1	AVAEKLR(5)	EPC/EPL1	4	2829	448	50	10	X	X	X	X	150	200
KRHVLPDAK	Esa1/Tip60	IETGVKGEIEEHLQR	EPC/EPL1	6	439	51	50	7	X	X	X	X	150	200
KLAELVCVSR(1)	Eaf1	AVAEKLR(5)	EPC/EPL1	3,4	42	448	50	8	X	X	X	X	150	200
KGYNsgNYPNR	EPC/EPL1	KYAQDEVPIRPDDPTDR	Eaf1	4	374	173	150	7,8	X	X	X	X	50	200
KGYNsgNYPNR	EPC/EPL1	PEPTK	Eaf6	3	374	2	100	7	X	X	X	X	100	200
KESDR	EPC/EPL1	KYAQDEVPIRPDDPTDR	Eaf1	4	453	173	150	7	X	X	X	X	50	200
KELYSNIVMSGTTFMFGIA	Act1/Actin	NSAKSQGGMNVYDSDVDD	EPC/EPL1	4	292	599	200	8	X	X	X	X	0	200
IPDMDLITVSTVLNEKDEAIN	EPC/EPL1	NKMNQVANR	Eaf1	4	437	1017	200	7	X	X	X	X	0	200
IETGVKGEIEEHLQR	EPC/EPL1	HGYGKLIQFSYELSK	Esa1/Tip60	4	51	337	200	7	X	X	X	X	0	200
GFENFTKSSHSNNNPAAR(7)	Eaf6	TEGKVGSPK(4)	Esa1/Tip60	3,4	64	352	100	7	X	X	X	X	100	200
FEDSSEKDDSDPYVcmFR	EPC/EPL1	LRTGGSMIQSHNEVSKVR	Esa1/Tip60	4	266	182	50	7,8	X	X	X	X	150	200
ESPSKESER	Incharacterize	NSNDETRNDSLEDDTD	Eaf7	4	429	201	100	5	X	X	X	X	100	200
ENDKR(4)	Swc4/Dmap1	IEELKR(5)	Eaf1	3	167	268	100	6	X	X	X	X	100	200
ELTALAPSSMKVK	Act1/Actin	NSAKSQGGMNVYDSDVDD	EPC/EPL1	4	327	599	100	7	X	X	X	X	100	200

EKLDALENSPNLTQQQR	Eaf1	AVAEKLR	EPC_EPL1	4	988	448	50							X	150	200	
AVAEKLR(5)	EPC/EPL1	VANPREEVKLPAAATEASISR	TRA1	4	448	3349	100	9						X	100	200	
AVAEKLR	EPC/EPL1	KMoxLTER	Swc4/Dmap1	3	448	352	150	7				X	X	X	50	200	
AVAEKLR	EPC/EPL1	QEQONQER	TRA1	3	448	470	0	6			X	X	X	X	200	200	
AVAEKLR	EPC/EPL1	VNVLLTKR	TRA1	3	448	3782	100	8						X	100	200	
YYRDKFQPPGTYIR	EPC/EPL1	INEDFTKAK	Yng2	4	100	78	150	7;9				X	X	X	0	150	
VLPSAAGAR	EPC/EPL1	NKMoxNQVANR	Eaf1	3	2	1017	100	9				X	X		50	150	
VLPSAAGAR	EPC/EPL1	NSNDETRNNDSELDLDDTD	Eaf7	4	2	201	50	7				X			100	150	
VIEEINVQKDR(9)	Yaf9/Yeats4	KQVNMK(1)	Swc4/Dmap1	3	201	544	50	6						X	100	150	
TVTVMK	Eaf1	LNVNNKPTWWSYVEFSND	Swc4/Dmap1	3	357	70	150	7			X	X	X	X	0	150	
TTKEDYQAIQR	TRA1	KESDR	EPC/EPL1	4	3177	453	150	5			X	X	X	X	0	150	
TTKEDYQAIQR	TRA1	LKQDDVDVGGSSR	Eaf7	4	3177	170	150	5			X	X	X	X	0	150	
TTKEDYQAIQR	TRA1	TGTPNGNRPRGPNNSFTQER	Eaf1	6	3177	824	150	7			X	X	X	X	0	150	
TFNKLcR(4)	TRA1	PEPTK(1)	Eaf6	3,4	2279	2	100	9					X		50	150	
TFNKLcCmR	TRA1	AVAEKLR	EPC/EPL1	3	2279	448	100	8				X	X	X	50	150	
TEVDNENEHDDDDIFKD	Eaf1	SQELKR(5)	TRA1	3	378	2880	100	6			X			X	50	150	
SSNGHSTPQVR	Eaf1	SQELKR	TRA1	3	948	2880	0	6			X	X	X	X	150	150	
SQELKR	TRA1	KESDR	EPC/EPL1	3	2880	453	100	6				X	X	X	50	150	
SLTTDAYESALLEGKIVLYS	Eaf1	KFELAAK	Swc4/Dmap1	4	256	345	150	7				X	X	X	0	150	
SLDADITDRNSAKacSQQGN	EPC/EPL1	VKISAPPERK	Act1/Actin	6	597	329	150	8				X	X	X	0	150	
SLDADITDRNSAKacSQQGN	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	5	600	221	150	7				X	X	X	0	150	
SKDEPTDDKPR	Arp4/Actin	TSLLVQLTSSFDKR	Swc4/Dmap1	4	353	312	0	8			X	X	X	X	150	150	
SESKTESR	TRA1	IPDMDLITVSTVLNEKDEAL	EPC/EPL1	4	428	437	150	8			X	X	X	X	0	150	
SESKTESR	TRA1	SSNGHSTPQVR	Eaf1	4	428	944	50	6				X			100	150	
QRKESDR	EPC/EPL1	SQELKR	TRA1	4	453	2880	50	6			X				100	150	
NSAKacSQQGMNVYDSDV	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	4	597	221	50	6;7				X			100	150	
NKMNQVANR(2)	Eaf1	PEPTK(1)	Eaf6	4	1017	2	50	7					X		100	150	
NKMNQVANR	Eaf1	VNVLLTKR	TRA1	3	1017	3782	150	8			X	X	X	X	0	150	
LPKPDIVSVR	TRA1	KYAQDEVPIRPDDPTDR	Eaf1	4	3581	173	150	8			X	X	X	X	0	150	
LNQELTAMNPGKLVKVR	Arp4/Actin	KWLSLR	Eaf1	4	415	272	150	7			X	X	X	X	0	150	
KYAQDEVPIRPDDPTDR(1)	Eaf1	KMLTER	Swc4/Dmap1	4	173	352	0	8			X	X	X	X	150	150	
KWNALPQR	TRA1	RDPASELSASLDYQKGLFG	Eaf1	4	2829	520	150	7;8			X	X	X	X	0	150	
KPLPR	EPC/EPL1	LNQELTAMoxNPGKLVKVR	Arp4/Actin	4	515	415	150	10			X	X	X	X	0	150	
KPLIFLTDEEIK	Eaf1	TVKFK	Swc4/Dmap1	3	340	60	150	6			X	X	X	X	0	150	
KMLTER	Swc4/Dmap1	SHKSLTTDAYESALLEGK	Eaf1	4	352	241	0	7			X	X	X	X	150	150	
KLSEYEK	Eaf1	LNQELTAMNPGKLVKVR	Arp4/Actin	4	475	415	50	9			X				100	150	
KLSEYEK	Eaf1	NVTMDDKTSLLVQLTSSFD	Swc4/Dmap1	4	475	299	150	7			X	X	X	X	0	150	
KGYNSGNYPNR	EPC/EPL1	LNQELTAMNPGKLVKVR	Arp4/Actin	4	374	415	150	10			X	X	X	X	0	150	
KGYNSGNYPNR	EPC/EPL1	TPTIWLPODDKLLIR	Eaf1	4	374	559	150	8			X	X	X	X	0	150	
KGYNSGNYPNR	EPC/EPL1	VIEEINVQKDR	Yaf9/Yeats4	4	374	201	100	6			X	X			50	150	
KELLHATR	TRA1	KYAQDEVPIRPDDPTDR	Eaf1	4	321	173	150	9			X	X	X	X	0	150	
KDDMoxESTEDVSK	Swc4/Dmap1	SKDEPTDDKPR	Arp4/Actin	4	208	353	0	5			X	X			X	150	150
GSNGCcmYKR	TRA1	AVAEKLR	EPC/EPL1	3	3467	448	150	6				X	X	X	0	150	
GQSRDEPFAYTIQKLR	TRA1	NKMoxNQVANR	Eaf1	4	3570	1017	150	9				X	X	X	X	0	150
GFENFTKSSHSNNNPAAR	Eaf6	LSWKPPVFTASQLR	Esa1/Tip60	4	64	451	150	10			X	X	X	X	0	150	
FEDSSEKDDSDPYVCcmFR	EPC/EPL1	LDEWIEKDRIDYK	Esa1/Tip60	4	266	79	150	7			X	X	X	X	0	150	
FEDSSEKDDSDPYVCcmFR	EPC/EPL1	NKMoxNQVANR	Eaf1	4	266	1017	150	6				X	X	X	X	0	150
FEDSSEKacDDSDPYVCcmFR	EPC/EPL1	EGEELTREDEIKK	Esa1/Tip60	4	264	165	100	5;6			X	X	X	X	50	150	
FEDSSEKacDDSDPYVCcmFR	EPC/EPL1	TGSGMoxIQSHNEVSKacV	Esa1/Tip60	4	264	181	150	7			X	X	X	X	0	150	
FASTFLAPVNKSEEPYK	Jncharacterize	NSNDETRNNDSELDLDDTD	Eaf7	5	490	201	50	8				X			100	150	
FASTFLAPVNKSEEPYK	Jncharacterize	SRTPVNGKIGYR	Eaf7	4	490	154	150	7			X	X	X	X	0	150	
ENDKER	Swc4/Dmap1	RSKDEPTDDKPR	Arp4/Actin	3	167	360	100	6			X			X	50	150	
ELKQTR	EPC/EPL1	LESJDIEKESEGLLAPDDMI	Yng2	4	305	110	150	7			X	X	X	X	0	150	
ELKQTR	EPC/EPL1	NSNDETRNNDSELDLDDTD	Eaf7	5	305	201	0	6			X	X	X	X	150	150	
ELKQTR	EPC/EPL1	SRTPVNGKIGYR	Eaf7	4	305	154	0	8			X	X	X	X	150	150	
EKLDALENSPNLTQQQR	Eaf1	TTKEDYQAIQR	TRA1	4	988	3175	0	7			X			X	X	150	150
FEVKLPAATEASISR	TRA1	NKMoxNQVANR	Eaf1	4	3349	1017	0	7;9;10				X	X	X	X	150	150
EAKIVR	TRA1	VATSQNTTPVTNTR	Eaf1	3	3152	800	100	6				X	X	X	50	150	
AYEQVHOAKQK	EPC/EPL1	IEELKR	Eaf1	3	666	268	50	7				X			100	150	
AVAEKLR	EPC/EPL1	LNQELTAMNPGKLVKVR	Arp4/Actin	3	448	415	150	7			X	X	X	X	0	150	
AVAEKLR	EPC/EPL1	RDPASELSASLDYQKGLFG	Eaf1	5	448	520	100	10			X			X	50	150	
AVAEKLR	EPC/EPL1	SQELKR	TRA1	3	448	2880	50	7			X				100	150	
VPTYFPDEAKFEER(10)	Arp4/Actin	KGYNSGNYPNR(1)	EPC/EPL1	3	221	374	100	8			X				0	100	
VLPSAAGAR	EPC/EPL1	GQHIIcCmLTDQVVKQYDR	Esa1/Tip60	3	2	430	100	7				X	X	X	0	100	
VLPSAAGAR	EPC/EPL1	LVVPEQYSEEQFAPGELKV	Esa1/Tip60	3	2	32	100	7;10			X	X	X	X	0	100	
VLPSAAGAR	EPC/EPL1	TASPEPIEQEAK	Esa1/Tip60	3	2	2	100	10			X	X	X	X	0	100	
VLPSAAGAR	EPC/EPL1	VPTYFPDEAKFEER	Arp4/Actin	3	2	221	100	6				X	X	X	0	100	
VIEEINVQKDR(9)	Yaf9/Yeats4	KGYNSGNYPNR(1)	EPC/EPL1	4	201	374	100	6			X			X	0	100	
VIEEINVQKDR	Yaf9/Yeats4	KYAQDEVPIRPDDPTDR	Eaf1	4	201	173	100	8			X	X			0	100	
VATSQNTTPVTNTR	Eaf1	FKCcmNADEDAYR	TRA1	4	801	3316	100	7					X	X	0	100	
VANPREEVKLPAAATEASISR	TRA1	FEDSSEKDDSDPYVCcmFR	EPC/EPL1	5	3349	266	100	8			X	X			0	100	
VANPREEVKLPAAATEASISR	TRA1	KGYNSGNYPNR	EPC/EPL1	5	3349	374	100	8			X	X			0	100	
TVKFK	Swc4/Dmap1	LHIGLDDFKK	Eaf1	4	60	449	0	9			X	X			100	100	
TTKEDYQAIQR	TRA1	SSNGHSTPQVR	Eaf1	4	3177	944	100	5;6				X			X	0	100
TSMDLKR(6)	Swc4/Dmap1	APSTGK(1)	Yaf9/Yeats4	3	424	2	50	6					X		X	50	100
TPVNGKIGYR	Eaf7	FASTFLAPVNKSEEPYK	Jncharacterize	4	154	490	100	8			X			X	0	100	
TGSGMIQSHNEVSKVR	Esa1/Tip60	FEDSSEKDDSDPYVCcmFR	EPC/EPL1	4	182	266	100	7			X			X	0	100	
TFNKLcCmR	TRA1	IPDMDLITVSTVLNEKDEAL	EPC/EPL1	4	2279	437	100	7				X	X	X	0	100	
SSNGHSTPQVR(1)	Eaf1	VNVLLTKR(8)	TRA1	3,4,5	943	3782	100	7				X	X	X	0	100	
SSHSNNNPAAR	Eaf6	LSWKPPVFTASQLR	Esa1/Tip60	3	65	451	100	8			X	X	X	X	0	100	
SNKGDSTVK	Esa1/Tip60	VLPSAAGAR	EPC/EPL1	3	116	2	0	7				X	X	X	100	100	
SLDADITDRNSAKSQQGMN	EPC/EPL1	VKISAPPERK(2)	Act1/Actin	4	593	329	100	8				X	X	X	0	100	
SKTDSSSGRR	Arp4/Actin	RDKALKEQEAKEAEK	Swc4/Dmap1	5	342	146	0	7			X			X	100	100	
SHDAESEQEISPSKR	Jncharacterize	NSNDETRNNDSELDLDDTD	Eaf7	4	450	201	50	6;7				X			50	100	
SESKTESR	TRA1	GSEELAKNSPYLFEK	Swc4/Dmap1	4	428	104	100	6			X	X			0	100	
SESKTESR	TRA1	VLPSAAGAR	EPC/EPL1	3	428	2	100	6				X	X	X	0	100	
RSKKTDSSSGRR	Arp4/Actin	KWLSLR	Eaf1	4	343	272	100	8				X	X	X	0	100	
RQIFQTLALQGFSSQKDTL	TRA1	PEPTK(1)	Eaf6	4	2809	2	100	8			X			X	0	100	
RKEAIEVGHVEFQLR	EPC/EPL1	ELHQKVLKLR	Eaf6	4	322	20	100	8				X	X	X	0	100	



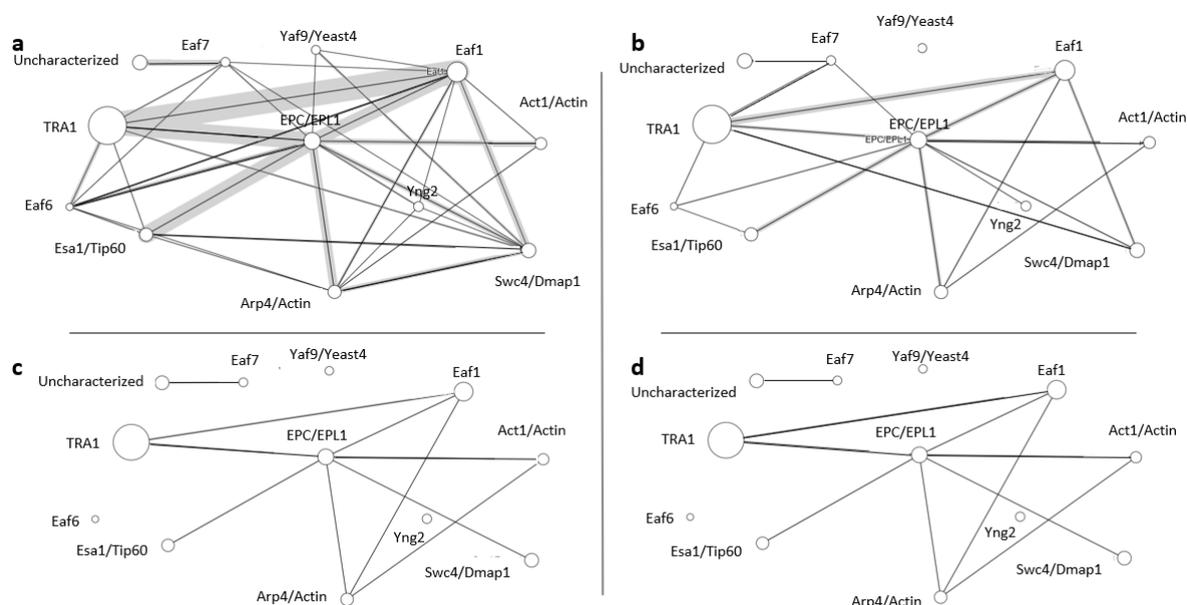
RSKKTDS5SGRR	Arp4/Actin	EQEAEAEK	Swc4/Dmap1	4	342	154	0	6	X				50	50
RSKKTDS5SGRR	Arp4/Actin	EREKRDALK	Swc4/Dmap1	6	342	143	0	7	X				50	50
RSKKTDS5SGR	Arp4/Actin	EYDFELKDLLECCmR	Swc4/Dmap1	4	343	132	50	6		X			0	50
RSKDPTDDKPR	Arp4/Actin	IMRSEKTESRQLLMVIDA	TRA1	6	353	428	0	9	X				50	50
RRNSSRINVQFER	Eaf1	KLAELYCcmVSR	Eaf1	3	677	42	50				X		0	50
QRKESDR	EPC/EPL1	RETERKQYLR	Swc4/Dmap1	3	453	319	50	8	X				0	50
QLLNHHKGVVPPVYK	Arp4/Actin	LKELEAEERKLVTVNQ	Yaf9/Yeats4	4	198	205	0	8	X				50	50
QDTADSVLKNNTNPQPQTR	TRA1	QPEDGSGTNDPDLQR	Eaf1	4	3209	25	50	8				X	0	50
QDTADSVLKNNTNPQPQTR	TRA1	QPEDGSGTNDPDLQR	Eaf1	3	3209	27	50	8				X	0	50
NVTMDDKTSLLVQLSSFD	Swc4/Dmap1	AVAEKLR(5)	EPC/EPL1	4	299	448	50	8	X		X		0	50
NSNDEDRNRNDSLEDDTD	Eaf7	KELHQQ	Eaf6	5	201	15	50	6		X			0	50
NQAQAQQQQQLQQQQLS	EPC/EPL1	VKISAPPER	Act1/Actin	4	403	329	50	7		X			0	50
NKMNQVANR	Eaf1	LPKPDIVSVR	TRA1	3	1017	3581	50	7;8	X				0	50
MTVDEPLDEFDFSTFEKSL	EPC/EPL1	VKISAPPERK	Act1/Actin	5	586	329	50	7		X			0	50
MKLPK	TRA1	NSAKSQQGMNVYDSVDD	EPC/EPL1	4	1452	599	0	7	X				50	50
MDIDVKEETK	TRA1	DRLKELEAEREK	Yaf9/Yeats4	4	2048	205	0	7	X				50	50
LSEYEK	Eaf1	TQSRFESPLSSAPTR	Yng2	4	477	164	50	5		X			0	50
LSDVSR	TRA1	KQFER	Esa1_Tip60	3	2429	228	0	7				X	50	50
LNQNAEPFKLHIGLDDFK	Eaf1	TVKFK	Swc4/Dmap1	5	440	60	50	7		X			0	50
LNQELTAMNPLKVR	Arp4/Actin	IEELKR	Eaf1	3	415	268	50	6;7		X			0	50
LKDQDDVDGSSR	Eaf7	EGEELTREDEIKK	Esa1/Tip60	4	170	165	0	6	X				50	50
LKDQDDVDGSSR	Eaf7	FASTFLAPVNVKSEPEYK	Jncharacterized	4	170	490	50	8;9	X				0	50
LIKRR	Esa1/Tip60	IETGVKEGEEEEHLQR	EPC/EPL1	4	437	51	0	7		X			50	50
LGSAEETEKEKSVSEDK	TRA1	DKTKSNK	Esa1_Tip60	3	2034	100	50	7				X	0	50
LEKEK	Swc4/Dmap1	AVAEKLR	EPC/EPL1	3	161	448	0	6	X				50	50
LEKEK	Swc4/Dmap1	SKDEPTDDKPR	Arp4/Actin	4	161	353	0	6		X			50	50
LEKEK	Swc4/Dmap1	WGSMLDAMRCKMRRR	Eaf1	4	161	669	0	8	X				50	50
KTDS5SGR	Arp4/Actin	EKLDALENSPNTLTPQQR	Eaf1	4	343	988	0	8	X				50	50
KRQRNGVSR	Yng2	KQTAMSR	Swc4/Dmap1	3	153	34	50	10	X				0	50
KPLPR	EPC/EPL1	IEELKR	Eaf1	3,4	515	268	0	10			X		50	50
KPLIFLTDEEIK	Eaf1	EKLNVNKPTWVSVFES	Swc4/Dmap1	4	340	70	50	8		X			0	50
KMTVDEPLDEFDFSTFEK	EPC/EPL1	LNQELTAMNPLKVR(13)	Arp4/Actin	3	568	415	50	7			X		0	50
KMLTER	Swc4/Dmap1	AVAEKLR	EPC/EPL1	3	352	448	50	7	X				0	50
KMLTER	Swc4/Dmap1	EYVPLRRSKTDS5SGR	Arp4/Actin	4	352	343	50	8	X				0	50
KLSEYEK	Eaf1	LPKSEKLSAEETEK	TRA1	3	475	2022	0	8	X				50	50
KLIMDR	Arp4/Actin	IKNTR	Esa1/Tip60	4	461	236	50	8	X				0	50
KISVK	EPC/EPL1	LVVPEQYSEEQFAPGELKV	Esa1/Tip60	4	15	32	50	7		X			0	50
KHTSDDK	Eaf1	TPVKNKGIYRTRR	Eaf7	4	686	154	0	8	X				50	50
KGYNSGNYPNR	EPC/EPL1	FESPLSSAPTR	Yng2	4	374	168	0	8		X			50	50
KGYNSGNYPNR	EPC/EPL1	NGVSR	Yng2	3	374	160	0	7			X		50	50
KGYNSGNYPNR	EPC/EPL1	TPVKNKGIYR	Eaf7	4	374	154	0	8		X			50	50
KFQTIANLISQISSNRFAS	Jncharacterized	WLSLEVAQAEARKAAEEK	TRA1	9	463	239	0	9	X				50	50
KFELAAK	Swc4/Dmap1	LNQELTAMNPLKVR	Arp4/Actin	4	345	415	50	7	X				0	50
KESDR(1)	EPC/EPL1	PEPTK(1)	Eaf6	4	453	2	50	6			X		0	50
KELHQQ	Eaf6	AVAEKLR	EPC/EPL1	3	15	448	50	7	X				0	50
IVKQPQDLKTLK	Jncharacterized	DPKDTSKRRTR	Eaf7	6	501	231	0	8	X				50	50
IPSKVSK	EPC/EPL1	EKSPVEK	TRA1	3	736	2511	50	10	X				0	50
IINEDFTKAK	Yng2	YYRDKFPPGTYIR	EPC/EPL1	4	78	100	50	7;9;10	X				0	50
IKQAMDK	Yaf9/Yeats4	ILNDDKRYTLLMGIVYEF	TRA1	5	187	3540	0	8	X				50	50
IKNTR(4)	Esa1/Tip60	LPKPDIVSVR(3)	TRA1	3	238	3581	50	7;8			X		0	50
IHQEKLPQGVSLR(5)	Swc4/Dmap1	LKGVSI(2)	Yaf9/Yeats4	4	482	10	50	9			X		0	50
IEELKR	Eaf1	LNQELTAMNPLKVR	Arp4/Actin	3	268	415	50	6;7	X				0	50
HEYGKLLIQFSYELSK	Esa1/Tip60	IETGVKEGEEEEHLQR	EPC/EPL1	5	337	51	50	7	X				0	50
GSELAKNSPYLFEK	Swc4_Dmap1	IETGVKEK	EPC_EPL1	4	104	47	0	7				X	50	50
GSNGCYKR(7)	TRA1	PEPTK(1)	Eaf6	4	3467	2	50	6			X		0	50
GLLHQEGSAEEKCcmVR	TRA1	KGYNSGNYPNR	EPC/EPL1	4	1640	374	50	7		X			0	50
GLLHQEGSAEEKCcmVR	TRA1	KYAQDEVPIRPDDPTDR	Eaf1	5	1640	173	0	9	X				50	50
FEDSSEKDDSDPYVCFR(7)	EPC/EPL1	SESKTESR(3)	TRA1	4,5	266	427	50	4			X		0	50
FEDSSEKDDSDPYVCFR(7)	EPC/EPL1	TNSELDWLKWPRR(9)	TRA1	3	266	1944	50	6			X		0	50
FEDSSEKDDSDPYVCFR	EPC/EPL1	FESPLSSAPTR	Yng2	4	266	168	50	7		X			0	50
FEDSSEKDDSDPYVCFR	EPC/EPL1	LVFYVHYQDFNKR	Esa1/Tip60	4	266	71	50	8	X				0	50
FEDSSEKDDSDPYVCFR	EPC/EPL1	LVVPEQYSEEQFAPGELKV	Esa1/Tip60	4	266	32	50	10	X				0	50
FEDSSEKDDSDPYVCFR	EPC/EPL1	RPVFWPKP	Esa1/Tip60	4	266	92	50	7	X				0	50
FASTFLAPVNVKSEPEYK	Jncharacterized	TPVKNKGIYR	Eaf7	4	490	154	50	7;8		X			0	50
EVKRR	EPC/EPL1	NSNDEDRNRNDSLEDDTD	Eaf7	5	320	201	0	6	X				50	50
ESKDGLK	Swc4/Dmap1	NSNDEDRNRNDSLEDDTD	Eaf7	5	87	201	50	6	X				0	50
ESKDGLK	Swc4/Dmap1	RSKDEPTDDKPR	Arp4/Actin	4	87	353	0	9	X				50	50
ESERKR	Jncharacterized	SESKTESR	TRA1	3	434	428	0	8	X				50	50
EREKR(4)	Swc4/Dmap1	EKSPVEK(2)	TRA1	4	143	2511	50	6			X		0	50
EREKR	Swc4/Dmap1	AVAEKLR	EPC/EPL1	3	143	448	50	7	X				0	50
ENDKER	Swc4/Dmap1	AEEQSSNKQDTADSVLKN	TRA1	4	167	3200	0	6	X				50	50
ELTALAPSSMKVK(11)	Act1/Actin	SLDADITDRNSAKSQQGM	EPC/EPL1	4	327	600	50	8			X		0	50
ELKQTR	EPC/EPL1	FESPLSSAPTR	Yng2	3	305	172	0	8		X			50	50
ELKQTR	EPC/EPL1	LSDVSRKK	TRA1	4	305	2435	0	9	X				50	50
EKLDALENSPNTLTPQQR	Eaf1	SOELKR	TRA1	4	988	2880	0	8			X		50	50
EKENDK	Swc4/Dmap1	EVFEQDIIACPNLETYSIKL	TRA1	4	163	3391	0	8	X				50	50
EKAGILGK	TRA1	ALKEQEAEAEKLEK	Swc4/Dmap1	4	2374	149	0	8	X				50	50
EGEELTREDEIKK	Esa1/Tip60	DDSDPYVCFR	EPC/EPL1	3	165	269	0	6		X			50	50
EYDFELKDLLECCR(8)	Swc4/Dmap1	50	Arp4/Actin	4	132	344	50	6			X		0	50
EAEKLEKEKENDK	Swc4/Dmap1	MDIDVKEETKEDIAERP	TRA1	4	158	2048	0	6	X				50	50
DENTEKK	Jncharacterized	FYKVSER	Swc4/Dmap1	3	263	281	0	8	X				50	50
AVAEKLR(5)	EPC/EPL1	KFELAAK(1)	Swc4/Dmap1	4	448	345	50	7			X		0	50
AVAEKLR	EPC/EPL1	LLEKCLR	TRA1	3	448	2188	0	9	X				50	50
AVAEKLR	EPC/EPL1	LPKPDIVSVR	TRA1	3	448	3581	50	8	X				0	50
AVAEKLR	EPC/EPL1	NKTIQALPEIYPHK	Eaf1	4	448	200	50	8	X				0	50
AVAEKLR	EPC/EPL1	NSNDEDRNRNDSLEDDTD	Eaf7	4	448	201	0	6	X				50	50
AVAEKLR	EPC/EPL1	YATGKTHSTAYNNSTLDP	TRA1	4	448	3673	50	7	X				0	50
AVAEKLR	EPC/EPL1	WSLRQPKR	Eaf1	3	448	279	0	10		X			50	50

## 2.4 Visualisation de la carte de proximité du complexe NuA4

A partir de ces résultats, une première carte de proximité a été réalisée afin de visualiser les zones proches au sein du complexe telles que déterminées par XL-MS. La figure 120 montre les zones de proximité obtenues pour le pontage par 5mM de BS3 en fonction du score arbitraire, c'est-à-dire selon la redondance d'identification par les logiciels. L'épaisseur du trait reflète le nombre de peptides cross-linkés représentatifs de la zone de proximité.

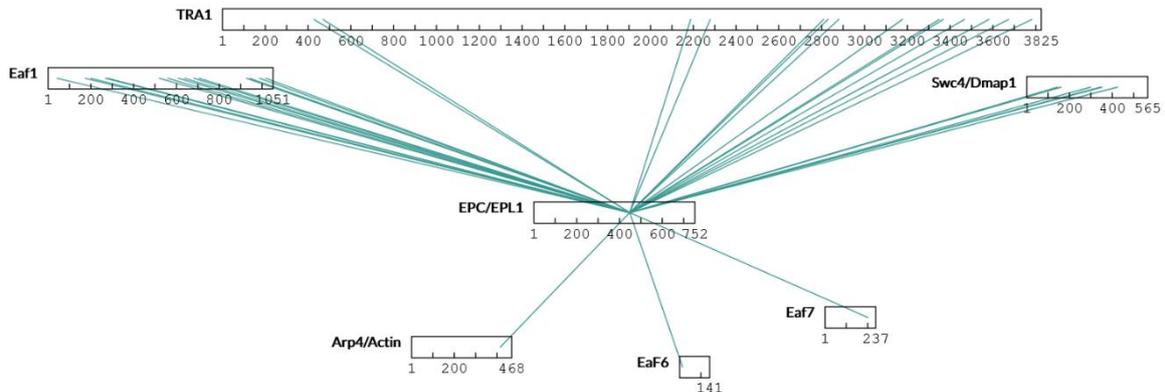
En particulier, un réseau de connexions entre les sous unités Tra1, Epc/Epl1 et Eaf1 est clairement visible. Sur cette carte de proximité, nous pouvons rapidement constater que la sous-unité EPC/EPL1 interagit avec toutes les sous-unités à l'exception de la protéine Uncharacterized bromodomain. Un réseau important a également été observé entre les sous-unités Esa1/Tip60 et Epc/Epl1 et Eaf6 montrant ainsi une interaction forte entre ses trois sous-unités représentant le sous-complexe piccolo de NuA4. Ce sous-complexe est également caractérisé par une interaction unique entre Yng2, la quatrième sous unité de piccolo et Epc/Epl1.

La majorité des zones d'interactions observées sur Eaf6 sont réalisées sur le domaine N-terminal (K2, K15 et K64) alors que les zones d'interactions semblent être davantage réparties sur l'ensemble de la séquence de Eaf1 et de Yng2. Une interaction unique a également été observée sur le bromodomaine avec la sous-unité Eaf7 montrant que ce domaine semble être assez isolé des autres sous-unités. La sous-unité Arp4/Actin interagit fortement avec Epc/Epl1, mais aussi avec Act1/Actin et avec les autres sous-unités du module Piccolo (Esa1/tip60, Eaf6 et Yng2). Pour Yng2, et Yaf9, très peu d'informations ont été déduites rendant leur positionnement plus délicat.



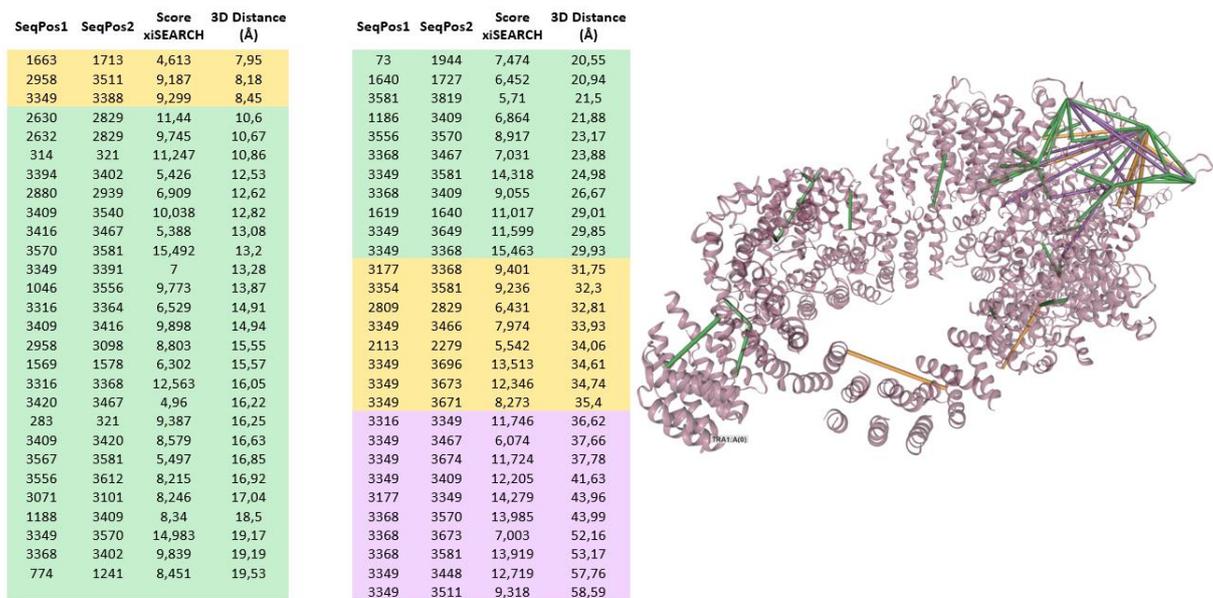
**Figure 120 : Carte de proximité (xiNET) du complexe NuA4 cross-linké par 5mM BS3 en fonction du score arbitraire. L'épaisseur du trait entre les sous-unités est proportionnelle au nombre de peptides pontés entre les sous-unités. Carte de proximité obtenue pour un score de 50 (a) ; score de 100 (b) ; score de 150 (c) ; score de 200 (d).**

De façon intrigante, certains acides aminés sont impliqués dans de nombreux pontages. C'est le cas notamment des lysines K374 et K448 de Epc/EpL1. La lysine K448 est réticulée à plusieurs sites de Eaf1 (acides aminés 173, 268, 520, 943, ...) ou de Swc4 (acides aminés 143, 161, 352, 424) et également à de nombreux sites sur Tra1 (acides aminés 428, 470, 3177, 3349, 3782, 2880, ...). En l'absence de données quantitatives, cela nous suggère que plusieurs conformations co-existent en solution ou bien que le cross-linking a perturbé le complexe en rendant des sites initialement éloignés, accessibles au pontage.



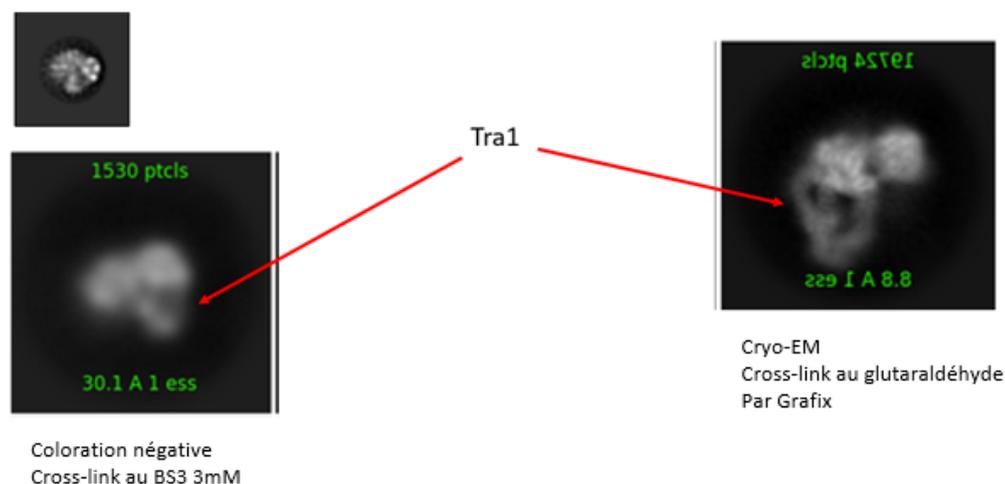
**Figure 121 : Visualisation du réseau de pontage entre la K448 de EPL1 et différentes sous-unités de NuA4.**

Afin de vérifier si la réaction de pontage n'avait pas modifié drastiquement la conformation de NuA4, nous avons décidé d'examiner les pontages intramoléculaires de la sous-unité Tra1 pour laquelle nous disposons d'une structure tridimensionnelle modélisée par nos collaborateurs. Sur la figure 122 sont représentés les cross-links impliquant Tra1 issus du pontage de NuA4 par 5mM BS3. Sur les 57 peptides intramoléculaires identifiés par XiSearch, 36 montrent une distance d'éloignement comprise entre 10 et 30 Å et sont donc cohérents avec le cross-linker BS3 utilisé. 10 peptides présentent des sites éloignés de plus de 35 Å, reflétant un changement conformationnel. Ces 10 peptides sont situés dans la partie C-terminale de la protéine entre les acides aminés 3349 et 3511 dont la plupart sont impliqués dans l'interface d'interaction avec Eaf1. Il est donc possible que la coexistence de plusieurs conformations du domaine C-terminal de Tra1 se soient répercutée par l'établissement de pontages non cohérents avec les autres sous-unités de NuA4.



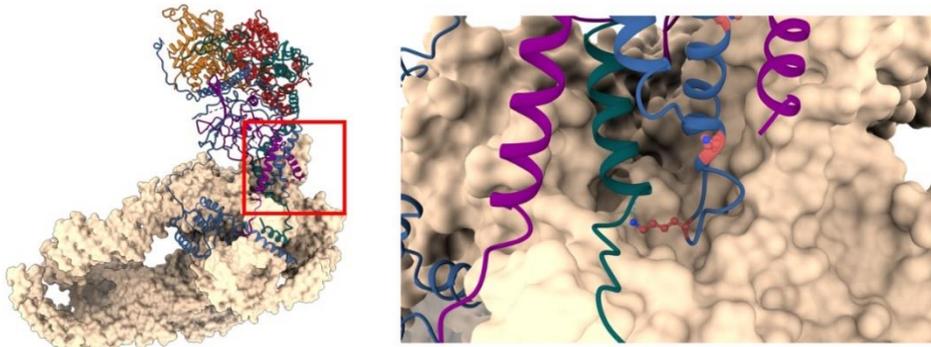
**Figure 122 : Implémentation des cross-links intramoléculeurs sur la structure cryo-EM de Tra1. En vert sont les distances atteintes lors du pontage par le BS3 comprises entre 10 et 30Å. En jaune sont les peptides observés avec une distance supérieure de 5Å. En violet sont les distances aberrantes (>30Å).**

Afin de vérifier que le cross-link par le BS3 n'est pas à l'origine d'un changement de conformation du complexe NuA4 ou d'une partielle dénaturation, nous avons entrepris une expérience de contrôle par cryo-EM. Pour cela, nos collaborateurs du Département de Biologie Structurale Intégrative de l'IGBMC ont analysé par coloration négative en EM un échantillon de NuA4 ponté par 3mM de BS3. Les images obtenues montrent bien Tra1, qui possède une asymétrie avec 2 lobes vides et un petit lobe. En revanche, le reste des sous-unités en interaction avec Tra1 est flou, ce qui indique que le complexe est plus flexible qu'il ne devrait, voire partiellement dénaturé.



**Figure 123 : Mise en évidence d'un comportement conformationnel anormal de NuA4 après pontage par 3mM de Bs3. L'intégralité du complexe est visualisée lors d'un cross-link au glutaraldéhyde par grafix.[30]**

La lysine K241 de Eaf1 est très importante pour la stabilité de NuA4. Elle est enfouie dans la structure de Tra1 et interagit fortement par liaison hydrogène avec Tra1 et Swc4. Ainsi, le complexe serait fortement déstabilisé si cette lysine était cross-linkée (figure 124). En examinant nos données, 1 seul peptide impliquant la K241 a été identifié dans l'échantillon ponté avec 3mM de BS3 dans la fraction SCX 7, il s'agit du peptide K241 – K352 (Eaf1 – Swc4). Ainsi, il ne semble pas que nos conditions de cross-link aient modifié significativement cet ancrage essentiel.



**Figure 124 : zoom de la structure cryo-EM de NuA4 sur l'ancrage de Eaf1 et Swc4 sur Tra1. La K241 de Eaf1 est impliquée dans des liaisons hydrogènes entre les trois sous-unités.**

Aux vues des résultats de l'expérience de contrôle par EM sur l'échantillon cross-linké au BS3, il semble donc difficile d'utiliser les résultats de XL-MS pour compléter la modélisation de NuA4. Une nouvelle expérience en utilisant des concentrations de BS3 plus faibles sont envisagées.

### 3. Conclusions et Perspectives.

Une expérience de pontage du complexe NuA4 par le BS3 a été menée afin d'identifier les domaines protéiques proches dans l'espace et ainsi compléter la modélisation entreprise par cryo-EM. Cette étude a démontré l'utilité d'une décomplexification du mélange peptidique en amont de l'analyse par MS. En effet, l'élution fractionnée d'une colonne de SCX a permis de multiplier par 8 le nombre de peptides intermoléculaires identifiés.

Ceci est particulièrement important lorsque l'on travaille sur des systèmes lourds et composés de plusieurs sous-unités car après pontage, le mélange peptidique est très complexe et les peptides cross-linkés en abondance faible par rapport aux peptides linéaires risquent de ne pas être sélectionnés pour les événements MS/MS.

Les résultats de XL-MS que nous avons obtenus ne sont pas cohérents avec les éléments de structure tridimensionnelle dont nous disposons et une expérience de contrôle par EM suggère que la réaction de pontage a déstabilisé NuA4. Ainsi, cette étude nous sensibilise au fait qu'il est essentiel de disposer d'expériences de contrôle afin de vérifier que les données de XL-MS ont préservé la mémoire de l'état conformationnel en solution.

Il serait intéressant de réaliser la réaction de pontage avec un autre type d'agent de pontage (MS-clivable notamment) à des concentrations douces pour éviter de déstructurer le complexe.

La combinaison entre la cryo-EM, la modélisation moléculaire et la XL-MS est une stratégie prometteuse en plein essor. Elle a été utilisée avec succès par Robinson et al afin de caractériser un médiateur de l'ARN polymérase II. Il s'agit d'un complexe biologique constitué de 52 protéines et ayant un poids moléculaire de 2.5 MDa. Dans cette étude, la XL-MS réalisée avec le 1,1'-(suberoyldioxy) bisazabenzotriazole (SBAT) a permis d'aboutir à une structure EM à une résolution comprise entre 6 et 11 Å<sup>[31]</sup>.

Les grands complexes hétérogènes ou dynamiques bénéficieraient ainsi d'une approche alternative aux méthodes traditionnelles utilisées en biologie structurale qui sont souvent difficiles à appliquer.

## 4. Bibliographie

1. Barnes, C.E., D.M. English, and S.M. Cowley, *Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription*. *Essays in biochemistry*, 2019. **63**(1): p. 97-107.
2. Narita, T., B.T. Weinert, and C. Choudhary, *Functions and mechanisms of non-histone protein acetylation*. *Nature Reviews Molecular Cell Biology*, 2019. **20**(3): p. 156-174.
3. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications*. *Cell Research*, 2011. **21**(3): p. 381-395.
4. Helmlinger, D. and L. Tora, *Sharing the SAGA*. *Trends in biochemical sciences*, 2017. **42**(11): p. 850-861.
5. Koutelou, E., C.L. Hirsch, and S.Y.R. Dent, *Multiple faces of the SAGA complex*. *Current opinion in cell biology*, 2010. **22**(3): p. 374-382.
6. Baker, S.P. and P.A. Grant, *The SAGA continues: expanding the cellular role of a transcriptional co-activator complex*. *Oncogene*, 2007. **26**(37): p. 5329-5340.
7. Doyon, Y. and J. Côté, *The highly conserved and multifunctional NuA4 HAT complex*. *Current Opinion in Genetics & Development*, 2004. **14**: p. 147-154.
8. Doyon, Y., et al., *Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans*. *Molecular and Cellular Biology*, 2004. **24**: p. 1884-1896.
9. Allard, S., et al., *NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p*. *The EMBO journal*, 1999. **18**(18): p. 5108-5119.
10. Sternglanz, R. and H. Schindelin, *Structure and mechanism of action of the histone acetyltransferase Gcn5 and similarity to other N-acetyltransferases*. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. **96**(16): p. 8807-8808.
11. Auger, A., et al., *Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants*. *Molecular and cellular biology*, 2008. **28**(7): p. 2257-2270.
12. Espinosa-Cores, L., et al., *Insights Into the Function of the NuA4 Complex in Plants*. *Frontiers in Plant Science*, 2020. **11**(125).
13. Steunou, A.-L., et al., *Combined Action of Histone Reader Modules Regulates NuA4 Local Acetyltransferase Function but Not Its Recruitment on the Genome*. *Molecular and cellular biology*, 2016. **36**(22): p. 2768-2781.
14. Selleck, W., et al., *The Saccharomyces cerevisiae Piccolo NuA4 histone acetyltransferase complex requires the Enhancer of Polycomb A domain and chromodomain to acetylate nucleosomes*. *Molecular and cellular biology*, 2005. **25**(13): p. 5535-5542.
15. Friis, R.M.N., et al., *A glycolytic burst drives glucose induction of global histone acetylation by picNuA4 and SAGA*. *Nucleic acids research*, 2009. **37**(12): p. 3969-3980.
16. Bhat, W., S. Ahmad, and J. Côté, *TINTIN, at the interface of chromatin, transcription elongation, and mRNA processing*. *RNA biology*, 2015. **12**(5): p. 486-489.
17. Wang, X.J., et al., *Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex*. *Nature Communications*, 2018. **9**.p. 1-11.
18. Helmlinger, D., et al., *Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex*. *The EMBO journal*, 2011. **30**(14): p. 2843-2852.
19. Díaz-Santín, L.M., et al., *Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution*. *eLife*, 2017. **6**: p. 1-20.

20. Mitchell, L., et al., *Functional dissection of the NuA4 histone acetyltransferase reveals its role as a genetic hub and that Eaf1 is essential for complex integrity*. *Molecular and cellular biology*, 2008. **28**(7): p. 2244-2256.
21. Wang, H., et al., *Structure of the transcription coactivator SAGA*. *Nature*, 2020. **577**(7792): p. 717-720.
22. Yan, Y., et al., *The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate*. *Nature Structural Biology*, 2002. **9**(11): p. 862-869.
23. Xu, P., et al., *The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism*. *Molecular Cell*, 2016. **63**(6): p. 965-975.
24. Ben-Shem, A., G. Papai, and P. Schultz, *Architecture of the multi-functional SAGA complex and the molecular mechanism of holding TBP*. *The FEBS Journal*, 2020. p.1-13.
25. Mintseris, J. and S.P. Gygi, *High-density chemical cross-linking for modeling protein interactions*. *Proceedings of the National Academy of Sciences*, 2020. **117**(1): p. 93.
26. Han, Y., et al., *Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex*. *The EMBO journal*, 2014. **33**(21): p. 2534-2546.
27. Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics*. *Chemical reviews*, 2013. **113**(4): p. 2343-2394.
28. McCormack, A.L., et al., *Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level*. *Analytical Chemistry*, 1997. **69**(4): p. 767-776.
29. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by Mass Spectrometry: Approaches, Advances, and Applications*. *Annual Review of Biomedical Engineering*, 2009. **11**(1): p. 49-79.
30. Stark, H., *Chapter Five - GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM*, in *Methods in Enzymology*, G.J. Jensen, Editor. 2010, Academic Press. p. 109-126.
31. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex*. *Cell*, 2016. **166**(6): p. 1411-1422.e16.

## Conclusion générale et perspectives

En conclusion, ce travail de thèse a montré que la MS se fraye aujourd'hui une place de choix dans le panel des techniques d'analyse structurale des protéines. La caractérisation de protéines ou complexes de protéines intacts par approche de MS native, ou encore l'utilisation de réticulation covalente pour stabiliser le complexe protéique d'intérêt par approche XL-MS permet d'accéder à des informations structurales de stœchiométrie ou de positionnement relatif des séquences protéiques dans l'espace qui sont cruciales pour affiner la modélisation de leur structure tridimensionnelle.

Dans la première partie de ce manuscrit consacrée à l'étude bibliographique, nous avons fait le point sur les différentes méthodes analytiques permettant la caractérisation structurale de complexe biologique non covalent en montrant leurs apports et leurs limites. Dans ce contexte, plusieurs méthodologies basées sur la MS ont proposé des alternatives ou des informations complémentaires visant, grâce à une approche intégrative, à construire des modèles structuraux toujours plus précis, plus complets, plus dynamiques, en d'autres termes, plus proches des structures *in vivo*. En particulier, des expériences de marquage chimique, d'échange hydrogène-deutérium (HDX-MS) ou de pontage chimique couplés à la MS fournissent des informations à faible résolution, mais qui combinées aux autres techniques d'analyse structurale, permettront d'aboutir à la construction de modèles à l'échelle atomique.

Dans un premier temps, nous avons adapté l'approche par MS native pour étudier le processus d'oligomérisation du domaine de liaison au coactivateur nucléaire de la protéine CREB (NCBD) par MS native. Cette étude nous a montré que la MS native n'est pas limitée à la caractérisation de complexes biologiques bien structurés, mais peut également apporter des informations précieuses sur des protéines à conformation flexible susceptibles de s'agréger. La MS native s'est montrée très appropriée pour suivre l'évolution des différentes stœchiométries co-existant en solution. Ainsi il a été montré que l'hélice N-terminale de NCBD jouait un rôle important dans la médiation de l'autoassemblage du domaine NCBD.

D'autre part, l'étude de l'interaction entre NCBD et le domaine TAD de la protéine p53 a montré que la complexation se faisait selon une stœchiométrie préférentielle 1 : 1 (NCBD : TAD) en induisant en parallèle la déstabilisation des oligomères de NCBD. Dans des conditions hyperphosphorylées de TAD, les interactions entre NCBD et pTAD s'avèrent renforcées avec la détection d'un complexe de stœchiométrie NCBD/pTAD : 2/1, montrant que pTAD pouvait interagir avec 2 molécules de NCBD. Dans le cadre de ces études, nous avons souligné l'intérêt à disposer d'un spectromètre de masse à haute résolution pour attribuer sans ambiguïté les états de charge des espèces détectées et ainsi en déduire leurs masses moléculaires respectives.

De plus, il serait intéressant de poursuivre l'étude sur une échelle de temps plus long (sur plusieurs semaines) et à des pH acides afin de voir s'il y aurait possibilité de confirmer les

résultats observés en microscopie TEM ou par spectroscopie IR concernant l'évolution de l'agrégation.

Dans un second temps, nous avons développé un protocole analytique pour la caractérisation structurale de complexes biologiques par pontage chimique couplée à la spectrométrie de masse (XL-MS). Ce protocole a été optimisé sur un complexe de la famille des récepteurs nucléaires : le complexe RAR/RXR. Un protocole XL-MS robuste et reproductible a ainsi été mis en place et une étude comparative systématique par plusieurs logiciels spécialisés dans l'interprétation des données de XL-MS nous a permis de faire ressortir les paramètres importants afin d'accéder à une identification la plus probable possible des acides aminés pontés, à savoir 5 ions fragments consécutifs (4 acides aminés) sur chacune des deux séquences réticulées et une redondance obtenue par la répétabilité des expériences ou l'utilisation de plusieurs logiciels d'interprétation. Néanmoins, cette redondance est encore subjective et exploitée arbitrairement par l'expérimentateur.

D'un point de vue biologique, les données XL-MS n'ont malheureusement pas pu être traduites en informations structurales pertinentes. Nous avons choisi ce complexe RAR/RXR comme modèle d'étude, en imaginant qu'un système simple formé de deux sous-unités nous permettrait de consolider rapidement notre expérience en XL-MS et aboutirait à la mise en évidence des changements conformationnels induits par la liaison avec l'ADN. En réalité, nous avons été confrontés à un manque de reproductibilité lié à la fois à un système flexible (dont nous n'avons peut-être pas anticipé l'ampleur) et un souci lors du processus d'expression protéique que nous n'avons pas identifié rapidement. Ainsi ce travail fut un réel succès du point de vue méthodologique puisqu'il nous a fait prendre conscience de l'importance de vérifier l'homogénéité et la stœchiométrie de complexation en amont de la réaction de cross-link et ceci lors de chaque expérience, ainsi que l'importance de vérifier l'interprétation automatique des spectres MS/MS. Ce projet se poursuit afin de pouvoir traduire cette rigueur analytique en informations structurales.

Pour la suite du projet, il nous semble utile de mettre en place en amont des analyses LC-MS une technique d'enrichissement automatisée afin d'obtenir des échantillons davantage concentrés en peptides pontés. Il serait également intéressant d'effectuer des analyses sur un spectromètre de masse équipé d'une cellule de mobilité ionique pour voir l'apport de cette méthode dans l'identification des peptides pontés davantage chargés que les peptides linéaires.

Enfin, le protocole ainsi développé a été appliqué à deux autres édifices biologiques :

- Le premier édifice est une protéine possédant une structure désorganisée faisant partie de la famille des coactivateurs nucléaires. Cette protéine coactivatrice CBP est une protéine à haut poids moléculaire (environ 230 000 Da) connue pour interagir avec le complexe RAR/RXR. L'objectif de cette étude était d'exploiter la XL-MS afin d'obtenir des informations structurales sur les domaines non structurés de CBP en identifiant des séquences protéiques proches dans l'espace et venir ainsi compléter les données obtenues par Cryo-EM et participer à l'élaboration d'un premier modèle structural.

- Les résultats XL-MS s'avèrent prometteurs car des peptides identifiés avec confiance impliquant des séquences désordonnées ont été observés. Néanmoins, des études supplémentaires seront nécessaires afin de reproduire nos résultats et confirmer les informations structurales obtenues sur cette protéine. Pour la suite du projet, il serait intéressant d'étudier la complexation entre le complexe RAR/RXR, le coactivateur SRC-1 et la protéine CBP afin de voir si un changement conformationnel s'opère et si certaines zones considérées jusqu'alors comme désordonnées, acquièrent une conformation stabilisée par la formation du complexe quaternaire. La limitation de cette étude reste aujourd'hui la production d'un complexe homogène en concentration compatible avec la XL-MS.
  
- Le second édifice est un complexe protéique qui joue un rôle important dans les processus génomiques cruciaux en particulier dans la réparation des dommages réalisés sur l'ADN. Il s'agit du complexe NuA4 constitué de 12 sous-unités ayant une masse moléculaire de 1 MDa. En raison de la complexité élevée de la composition et de la flexibilité de conformation du complexe NuA4 les analyses structurales antérieures étaient limitées à des modèles structuraux à faible résolution. C'est pourquoi notre objectif était ici d'utiliser la XL-MS pour d'obtenir des informations sur le positionnement des différentes sous-unités du complexe, voire des zones protéiques proches dans l'espace et venir ainsi compléter, tel un puzzle, les données modélisées par cryo-EM. Lors de cette étude, nous avons démontré qu'une étape de préfractionnement par SCX conduisant à un enrichissement des peptides cross-linkés était primordiale pour l'obtention de données augmentées à la fois en nombre et en qualité. En effet, cet enrichissement vise à compenser la faible abondance des peptides cross-linkés au sein d'un mélange peptidique très complexe. Malheureusement, les résultats de XL-MS que nous avons obtenus se sont avérés incohérents avec les éléments de structure tridimensionnelle dont nous disposions et une expérience de contrôle par EM a suggéré que la réaction de pontage aurait pu déstabiliser NuA4. Ainsi, cette étude nous a sensibilisés au fait qu'il est essentiel de disposer d'expériences de contrôle afin de vérifier que les données de XL-MS ont préservé la mémoire de l'état conformationnel en solution. Pour les prochaines expériences, il serait intéressant de réaliser la réaction de pontage avec un autre type d'agent de pontage (MS-clivable notamment) à des concentrations douces pour éviter de déstructurer le complexe.

Depuis le début des années 2000 où les 1<sup>ère</sup> approches XL-MS ont été reportées, ce champ d'application a montré un fort potentiel pour sonder la structure tridimensionnelle des protéines ou complexes biologiques et notre travail de thèse illustre également ce potentiel. Néanmoins, comme nous l'avons souligné dans ce manuscrit et comme il est également décrit dans la littérature, il reste encore de nombreux défis analytiques, en particulier :

- le choix du cross-linker approprié pour chaque système d'étude
- un contrôle rigoureux des conditions de réactions de pontage pour éviter un changement de conformation artificiel

- une interprétation standardisée des données de XL-MS qui limitent l'attribution de faux-positifs tout en fournissant le maximum d'informations structurales

Dans cet objectif, le développement de cross-linkers MS-clivables fut à notre avis une étape cruciale en améliorant la détection de peptides pontés spécifiquement. D'autre part, quel que soit le cross-linker choisi, l'identification non ambiguë des sites pontés repose toujours sur l'interprétation des données MS/MS, c'est pourquoi nous pensons qu'introduire une étape d'enrichissement des peptides cross-linkés en amont de l'analyse MS devrait également être privilégié.

L'avenir de cette approche XL-MS réside aussi dans l'étude de ce qui se passe vraiment dans la cellule en capturant les interactions protéine/protéines en temps réel *in vivo*. Quelques études ont d'ores et déjà montré la faisabilité du concept par l'utilisation du cross-linker PIR<sup>[1]</sup>, de cross-linker MS-clivable qui pénètre la membrane plasmique<sup>[2]</sup> ou encore par l'introduction d'acides aminés photoactivables directement lors de l'expression protéique au sein des cellules<sup>[3]</sup>. La difficulté de cette dernière approche réside d'une part dans le rendement d'incorporation de ces acides aminés modifiés qui reste limitante (surtout dans des organismes d'expression autres qu'*E. coli*) et dans l'analyse du set très complexe de données de peptides pontés. Il est intéressant de noter que des plateformes d'analyses *in vivo* s'établissent progressivement<sup>[4]</sup>. Dans cette optique, il faudra probablement faire appel à des algorithmes de requête plus robustes et fiables pour faire face à la grande gamme dynamique et la haute complexité du mélange peptidique généré *in vivo*, d'où l'importance de définir des critères d'identification de peptides pontés à haut degré de confiance.

### Bibliographie

1. Tang, X. and J.E. Bruce, *A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies*. *Molecular bioSystems*, 2010. **6**(6): p. 939-947.
2. Buncherd, H., et al., *A gas phase cleavage reaction of cross-linked peptides for protein complex topology studies by peptide fragment fingerprinting from large sequence database*. *Journal of Proteomics*, 2014. **108**: p. 65-77.
3. Götze, M., et al., *A Simple Cross-Linking/Mass Spectrometry Workflow to Study System-Wide Protein Interactions*. *bioRxiv*, 2019: p. 524314.
4. Kaake, R.M., et al., *A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells*. *Molecular & cellular proteomics : MCP*, 2014. **13**(12): p. 3533-3543.

# Liste des Communications scientifiques

## 4.1 Publications avec comité de lecture

- Garcia, A.M., **Giorgiutti, C.**, El Khoury, Y., Bauer, V., Spiegelhalter, C., Leize-Wagner, E., Hellwig, P., Potier, N. and Torbeev, V. (2020), "Aggregation and amyloidogenicity of nuclear coactivator binding domain of CREB-binding protein". *Chemistry – A European Journal*, Accepted Author Manuscript. doi:[10.1002/chem.202001847](https://doi.org/10.1002/chem.202001847)

## 4.2 Communications Orales

- 2018, march 26<sup>th</sup>, Réunion du Club Jeune de Spectrométrie de Masse (RCJSM), Saint-Martin-de-Londres (France) : **Chemical cross-linking for the characterization of retinoid acid receptor-retinoid X receptor** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Emmanuelle Leize-Wagner.
- 2018, April 17<sup>th</sup>, UMR 7140 meeting, Strasbourg (France) : **Chemical cross-linking for the characterization of retinoid acid receptor-retinoid X receptor** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Emmanuelle Leize-Wagner.
- 2018, November 29<sup>th</sup>, Université Franco-Allemande : French German Doctoral School Workshop - International Workshop 'Enzyme reactivities and their applications', Strasbourg (France) : **Chemical cross-linking a new way to probe protein complex** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Emmanuelle Leize-Wagner.
- 2019, Mai 8<sup>th</sup>, UMR 7140 meeting, Strasbourg (France) : **Chemical cross-linking a new way to probe protein complex** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Emmanuelle Leize-Wagner.
- 2019, September 16<sup>th</sup>, Congrès de Spectrométrie de Masse Appliqué à la Protéomique 2019 (SMAP2019), Strasbourg (France) **Towards the structural characterization of a retinoic acid (RAR/RXR) nuclear receptor complex** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Emmanuelle Leize-Wagner.
- 2019, December, Université Franco-Allemande : French German Doctoral School Workshop - International Workshop 'Shining light on biological systems and molecules', Strasbourg (France) : **Structural Characterization of biological complexes by Infrared Spectroscopy and Cross-linking coupled to Mass**

**Spectrometry** Christophe Giorgiutti, Judit Osz, Zahia Boubegtiten Carole Peluso-Iltis, Natacha Rochel, Noelle Potier, Petra Hellwig, Emmanuelle Leize-Wagner.

#### **4.3 Communications par Affiche**

- 2018, March 18<sup>th</sup>: European Mass Spectrometry Conference 2018 (EMSC 2018) Saarbruck (Germany) : **Cross-linking/Mass Spectrometry of Nuclear Receptor Complexes** Christophe Giorgiutti, Carole Peluso-Iltis, Judit Osz, Laurianne Kuhn, Philippe Hammann, Emmanuelle Leize-Wagner, Natacha Rochel, Noelle Potier,
  
- 2018, October 10<sup>th</sup>: Symposium of Structural Proteomics 2018 (SSP 2018), Berlin (Germany) : **Chemical cross-linking for the characterization of retinoid acid receptor-retinoid X receptor: correlation between cross-links distance and 3D structure** Christophe Giorgiutti, Carole Peluso-Iltis, Judit Osz, Laurianne Kuhn, Philippe Hammann, Emmanuelle Leize-Wagner, Natacha Rochel, Noelle Potier,
  
- 2019, September 16<sup>th</sup>: Congrès de Spectrométrie de Masse Appliqué à la Protéomique 2019 (SMAP2019): **Towards the structural characterization of a retinoic acid (RAR/RXR) nuclear receptor complex with XL-MS** Christophe Giorgiutti, Judit Osz, Carole Peluso-Iltis, Emmanuelle Leize-Wagner, Natacha Rochel, Noelle Potier,
  
- 2019, November 3<sup>rd</sup>: Symposium of Structural Proteomics 2018 (SSP 2018), Göttingen (Germany) : **Conformational study of a retinoic acid (RAR/RXR) nuclear receptor complex** Christophe Giorgiutti, Carole Peluso-Iltis, Judit Osz, Laurianne Kuhn, Philippe Hammann, Emmanuelle Leize-Wagner, Natacha Rochel, Noelle Potier,

#### **4.4 Enseignements délivrés en tant que moniteur (Faculté de chimie, université de Strasbourg)**

- Année 2017/2018 :

Travaux Pratique « Chimie Analytique » Master 1 CPAM et Master 1 SACEB. (Pr. Maurice Millet)

- Années 2018/2019 et 2019/2020 :

- o Travaux Pratique « Chimie Analytique » Licence 3 Chimie Biologie (Dr. Yannis François)
- o Travaux Pratique Chimie Analytique Licence 2 Métiers de la chimie (Dr. Yannis François)

#### **4.5 Autres**

Elu représentant des doctorants de l'UMR7140 de Septembre 2018 à Septembre 2020.

## Annexe 1 : Poster EMSC 2018 (Saarbrücken)

### Cross-linking/Mass Spectrometry of Nuclear Receptor Complexes

Christophe Giorgiutti<sup>a</sup>, Carole Peluso-Ittis<sup>b</sup>, Judit Osz<sup>b</sup>, Laurianne Kuhn<sup>c</sup>, Philippe Hammann<sup>d</sup>,  
Emmanuelle Leize-Wagner<sup>a</sup>, Natacha Rochel<sup>a</sup>, Noelle Potier<sup>a</sup>

<sup>a</sup>Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (I.S.M.S)  
UMR 7140 CNRS/UDS - "Centre de la Matière Complexée"  
1 Rue Elzévir Pascal - 67028 Strasbourg, France

<sup>b</sup>Département de Biologie Structurale Intégrative  
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)  
1 Rue Lissac Fric - 67404 Illkirch, France

<sup>c</sup>Plateforme Protéomique Strasbourg-Epinalde  
<sup>d</sup>Plateforme Protéomique Strasbourg-Epinalde  
Institut de Biologie Moléculaire et Cellulaire (IGBMC)  
15 Rue René Descartes - 67000 Strasbourg

### INTRODUCTION

Protein complexes formation is crucial for a majority of biological processes. Therefore, understanding the way how proteins interact with each other to form transient or stable protein complexes is a key point in structural biology [1]. ESI-MS has shown a high potential for characterizing intact assemblies, especially to determine the complex stoichiometry or monitoring dynamic changes [2]. However, finding conditions compatible with both complex stability and ESI-MS might be very challenging. Thanks to its tolerance to concentrated salt over ESI-MS, MALDI-MS can also be used after coupling with chemical cross-linking, the protein complex is thus covalently stabilized, which allows more flexibility regarding to the sample preparation [3].

In this project, we combined chemical cross-linking (XL-MS) with MALDI-MS and ESI-MS to determine the proximity interfaces of a nuclear receptor protein complex in order to complement our structural data set and also to further define the MS potential in this domain.

### MATERIALS AND METHODS

In order to get more information on the interacting area, cross-linking with bisulfosuccinimidyl suberate (BS3) in the proportion 75/1 cross-linker/complex at 2 mg/mL was performed. After 30 min of incubation at ambient temperature, the cross-linking reaction was quenched with Tris 1M, analysed by MALDI-MS (Autoflex Speed TOF, Bruker), and submitted to tryptic proteolysis. Digested peptides were then analysed by LC-MS/MS (LTQ-Orbitrap Elite, Thermo Fisher). Intra- and inter-crosslinked peptides were then identified using a dedicated search engine: pLink [4].

### RESULTS

#### 1. Stoichiometry

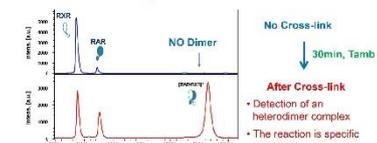


Figure 1. MALDI-MS spectra of the complex before (A) and after (B) cross-linking with BS3.

Sinapinic acid was used as matrix on our MALDI TOF. A mass of  $84994 \pm 200$  Da was measured which unambiguously reveals the cross-link of the RAR/RXR heterodimer. No signal corresponding to other stoichiometries (homodimers or higher-order structures) was detected.

#### 2. Identification of inter-crosslinked peptides

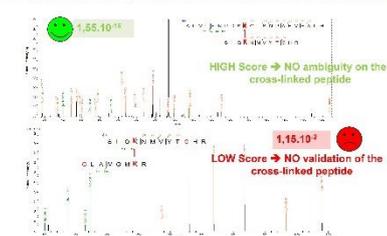
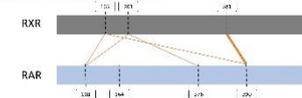


Figure 2. MS/MS spectra of two inter-crosslinked peptides as identified by pLink.

After tryptic digestion, nanoLC-MS/MS was performed on an orbitrap instrument. MASCOT was used for identification of unmodified peptides while pLink was chosen for the proposition of intra- and inter-crosslinked peptides. Manual validation was systematically performed to further validate the identification (about 10% were manually removed). A total number of 5 unique inter-crosslinks have been found with pLink. These results are the combination of three technical replicates (80% overlap) and one biological replicate (40% overlap).

Proteins	Sequence	pLink score
RXR(381)-RAR(390)	AVLFLPDSKGLSNPAEVEALR-SISAKGAER	2,02E-15
RXR(165)-RAR(118)	KDLTYTCR-SIQKNMVTYTCR	3,33E-10
RXR(201)-RAR(118)	CLAMGMKR-SIQKNMVTYTCR	1,41E-10
RXR(165)-RAR(390)	KDLTYTCR-SISAKGAER	2,86E-07
RXR(201)-RAR(376)	CLAMGMKR-PHMFPMKMLMK	2,80E-07

#### 3. Proximity map



In both cases, the corresponding peptides were identified with relevance – with or without the fragment containing cross-linking – thanks to several (>3) intense fragment ions and high mass precision on fragment ions (2 ppm)

From these results, a first linkage map has been produced to visualize inter-subunit links. The thickness of the line is proportional to the number of MS/MS spectra that have led to the matched peptides.

### CONCLUSION

The data generated by XL-MS will help to model iteratively the assembly of multiprotein systems. We hope that combined with complementary structural technique data, a low resolution map will be refined in order to reflect physiological structures as close as possible. These results illustrate the potential of XL-MS as part of an integrative strategy for addressing the architecture of multiprotein complexes.

[1] Hübner G. et al. *Nature* 2011, 478(7368): 83-91  
 [2] Giorgiutti C. et al. *Journal of Proteomics* 2017, 150: 1-13  
 [3] Wang L. et al. *Nat. Methods* 2012, 9, 664-4

\*The authors would like to thank the French Ministry of Health (ANR) for their contribution to this project.

## Annexe 2 : Poster SSP2018 (Berlin)

**Chemical cross-linking for the characterization of retinoid acid receptor-retinoid X receptor: correlation between crosslinks distance and 3D structure**

Christophe Giorgiutti<sup>a</sup>, Carole Peluso-Ilitis<sup>b</sup>, Judit Osz<sup>b</sup>, Laurianne Kuhn<sup>c</sup>, Philippe Hammann<sup>c</sup>, Emmanuelle Leize-Wagner<sup>a</sup>, Natacha Rochel<sup>b</sup>, Noelle Potier<sup>a</sup>

<sup>a</sup>Laboratoire de Spectrométrie de Masses des Interactions et des Systèmes (LSMIS)  
UMR 7140 CNRS/UDS - "Chimie de la Matière Complexée"  
1 Rue Blaise Pascal - 67008 Strasbourg, France

<sup>b</sup>Département de Biologie Structurale Intégrative  
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)  
1 Rue Laurent Fries - 67404 Illkirch, France

<sup>c</sup>Plateforme Protéomique Strasbourg Esplanade  
Institut de Biologie Moléculaire et Cellulaire (IBMC)  
15 Rue René Descartes - 67000 Strasbourg

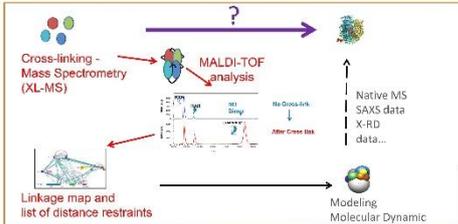


### INTRODUCTION

Protein complexes formation is crucial for a majority of biological processes. Therefore, understanding the way how proteins interact with each other to form transient or stable protein complexes is a key point in structural biology [1]. ESI-MS has shown a high potential for characterizing intact assemblies, especially to determine the complex stoichiometry or monitoring dynamic changes [2]. However, finding conditions compatible with both complex stability and ESI-MS might be very challenging. Thanks to its tolerance to concentrated salt over ESI-MS, MALDI-MS can also be used after coupling with chemical cross-linking; the protein complex is thus covalently stabilized, which allows more flexibility regarding to the sample preparation [3].

In this project, we combined chemical cross-linking (XL-MS) with MALDI-MS and ESI-MS to determine the proximity interfaces of a nuclear receptor protein complex in order to complement our structural data set and also to further define the MS potential in this domain.

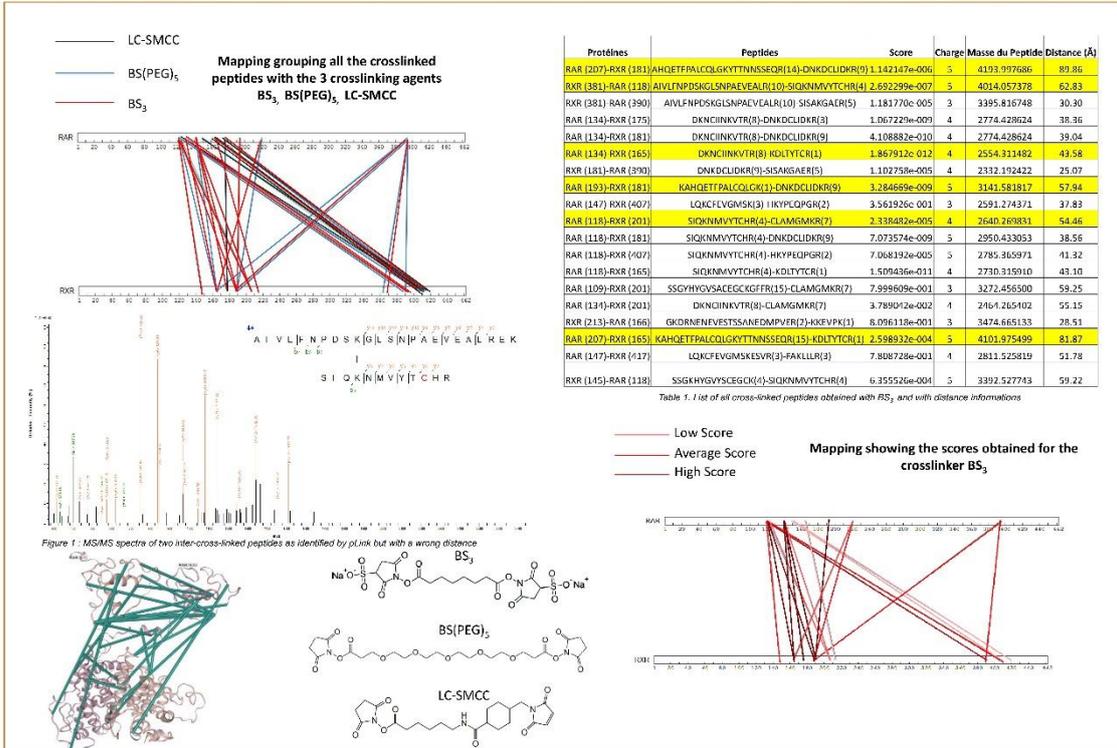
### MATERIALS AND METHODS



In order to get more information on the interacting area, cross-linking with bisulfosuccinimidyl suberate (**BS<sub>3</sub>**), pegylated bisulfosuccinimidylsuberate (**BS(PEG)<sub>2</sub>**) et succinimidyl-4-N-maleimidomethylcyclohexane-1-carboxy-6-amidocaproate (**LC-SMCC**) in the proportion 75/1 cross-linker/complex at 2 mg/mL was performed. After 1h30 min of incubation at ambient temperature, the cross-linking reaction was quenched with Tris 1M, analysed by MALDI-MS (Autoflex Speed TOF, Bruker), and submitted to tryptic proteolysis. Digested peptides were then analysed by LC-MS/MS (Orbitrap Exactive Plus, Thermo Fisher). Intra- and inter-crosslinked peptides were then identified using a dedicated search engine: pLink [4].



### RESULTS



### CONCLUSION

The data generated by XL-MS will help to model iteratively the assembly of multiprotein systems. We hope that combined with complementary structural technique data, a low resolution map will be refined in order to reflect physiological structures as close as possible. These results illustrate the potential of XL-MS as part of an integrative strategy for addressing the architecture of multiprotein complexes.

[1] McKen G. M. and Banerjee S. J., *J. H. Soc. Chromatogr* 2012, 9, 861-16.

[2] Heck A. J., *Nat. Methods* 2008, 5, 927-33

[3] Stengel T., Petersen M. R. and Robinson C. V., *Euk. Cell. Proteomics* 2012, 11, 1-13.

[4] Yang B. et al., *Nat. Methods* 2012, 9, 904-6

The authors would like to thank the Proteomics Platforms of IGBMC and IBMC (Institut de Biologie Moléculaire et Cellulaire) for their contribution to this project



## Annexe 4 : Poster SSP2019 (Göttingen)



### Conformational study of a retinoic acid (RAR/RXR) nuclear receptor complex

Christophe Giorgiutti<sup>a</sup>, Carole Peluso-Itlis<sup>b</sup>, Judit Osz<sup>b</sup>, Zahia Boubeglien<sup>d</sup>, Laurianne Kuhn<sup>c</sup>, Philippe Hamman<sup>c</sup>, Emmanuelle Leize-Wagner<sup>a</sup>, Petra Hellwig<sup>d</sup>, Natacha Roche<sup>c</sup>, Noelle Potier<sup>a</sup>

<sup>a</sup>Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) - UMR 7140 CNRS/UDS - "Chimie de la Matière Complexe" - University of Strasbourg, Strasbourg (France)

<sup>b</sup>Département de Biologie Structurale Intégrative - Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Strasbourg (France)

<sup>c</sup>Plateforme Protéomique Strasbourg Esplanade - Institut de Biologie Moléculaire et Cellulaire (IBMC) - Strasbourg (France)

<sup>d</sup>Laboratoire de Bioélectrochimie et Spectroscopie (LBS) UMR7140 CNRS/UDS - "Chimie de la Matière Complexe" - University of Strasbourg, Strasbourg (France)



#### Introduction

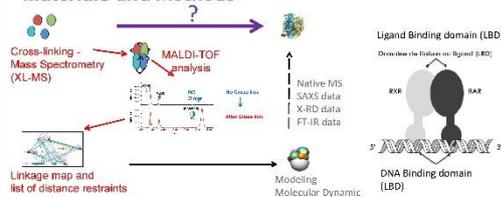
Protein complexes formation is crucial for a majority of biological processes. Therefore, understanding the way how proteins interact with each other to form transient or stable protein complexes is a key point in structural biology<sup>[1]</sup>. ESI-MS has shown a high potential for characterizing intact assemblies, especially to determine the complex stoichiometry or monitoring dynamic changes<sup>[2]</sup>. In addition to mass spectrometry, other structural techniques have been developed to perform structural characterization such as Fourier Transform Infrared (FT-IR) spectroscopy<sup>[3]</sup>. With this second analytical techniques, it was therefore possible to characterize biological complexes, in particular by obtaining information on the secondary structure of complexes. Combination of several analytical techniques therefore offers the possibility of obtaining complementary structural information on biological complexes that would allow us to better understand their biological properties.

In this project, we combined chemical cross-linking (XL-MS)<sup>[4]</sup> with MALDI-MS and ESI-MS to determine the proximity interfaces of a nuclear receptor protein complex in order to complement our structural data set and also to further define the MS potential in this domain. With FT-IR, the objective was initially to answer the question if XL-MS would lead to artefactual conformational changes which would give false structural information on the complexes. The second objective was to characterize the complex in order to complement structural information of our complex.

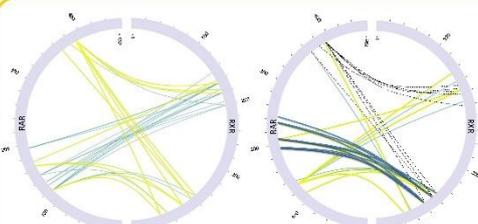
#### Motivation

- Characterisation of a nuclear receptor complex from the retinoic acid family (RAR/RXR) in presence of different interaction partner.
- Evaluation if the spatial resolution obtained makes it possible to describe the possible conformational changes upon DNA binding.

#### Materials and Methods



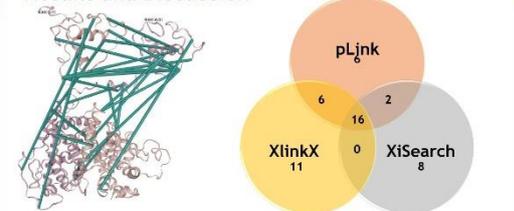
In order to get more information on the interacting area, cross-linking with bisulfosuccinimidyl suberate (BS<sub>3</sub>) and Disuccinimidyl Dibutyric Urea (DSBU) in the proportion 50/1 cross-linker/complex at 2 mg/mL was performed. After 30 min of incubation at ambient temperature, the cross-linking reaction was quenched with Tris 1M, analysed by MALDI-MS (Autoflex Speed TOF, Bruker), and submitted to tryptic proteolysis. Digested peptides were then fractionated by Strong cation exchange columns (SCX columns) and then analysed by LC-MS/MS (Orbitrap Exactive Plus, Thermo Fisher). Intra- and inter-crosslinked peptides were then identified using dedicated search engine: pLink<sup>[4]</sup>, XiSearch<sup>[5]</sup> and XlinkX<sup>[6]</sup>.



Cross-link present on the complex in absence of DNA. Cross-link identified with DSBU (Yellow) and BS3 (grey)

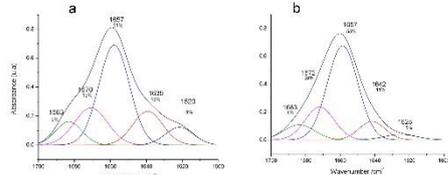
Cross-link only present on the complex in presence of DNA. (Purple) Cross-link not detected in presence of DNA have been added in dotted lines. Cross-link identified with DSBU (Yellow) and BS3 (grey)

#### Results and Discussion



Implementing XL-MS data on a structural SAXS Models

Comparison of identified unique cross-linked peptide pairs of RAR/RXR by pLink 2.3.5, XiSearch 1.6.751 and XlinkX 2.0



Deconvolution of Amide I Band (1700 – 1600 cm<sup>-1</sup>) obtained by Infrared Spectroscopy showing the secondary structure of RAR/RXR complex (a) and the cross-linked RAR/RXR complex (b) No blatant evolution of secondary structure in presence of Cross-linker

In order to be able to confirm the evolution of the structure in presence of DNA, the cross-link experiment was performed on 5 different biological samples preparation and injected 3 times consecutively into the mass spectrometer.

At the end of this study, only the most reproducible peptides were conserved. Moreover, only cross-linked peptides were selected if they were at least detected by 2 different software but with identical identification parameters. The results showed a technical reproducibility of 80% and a lower biological reproducibility around 45%.

#### Conclusion

A XL-MS protocol has been optimized after testing different cross-linkers and interpretation software in order to improve available structural models of a RAR/RXR complex. XL-MS data were then compared with SAXS data to help refinement towards a more precise 3D model. Resolution of the technique is not easy to evaluate but it might be promising for the study of larger protein complexes even occurring with protein conformational changes. FTIR spectroscopy seems to be an interesting tool for performing structural characterization of biological complex.

#### Perspectives

- Carry on the study of the complex in presence of DNA and/or coactivator.
- Analysis of a larger complex from the ribosomal family (Ribosome 70S).

#### References

- [1] Hillon, G. R.; Benesch, J. L. P. *J R Soc Interface* **2012**, *9*, 801-816.
- [2] Heck, A. J. R. *Nat Methods* **2008**, *5*, 927-935.
- [3] Barth, A. *Eur Biotechnol* **2007**, *17*, 1073-1101.
- [4] Rappasber, J. *Journal of Structural Biology* **2011**, *173*, 530-540.
- [5] Chen, Z. L. et al. *Nat Commun* **2019**, *10*.
- [6] Graham, M. J. et al. *BioRxiv* **2019**, 561829.
- [7] Liu, F. et al. *Nat Methods* **2017**, *14*, 539-539.

Christophe Giorgiutti would like to thank the Franter Research in Chemistry foundation (FRC Foundation) for the financing of his thesis.

## Annexe 5 : Protocole expérimentale de précipitation des protéines avant digestion

### ⇒ Préparation solution

- Bicarbonate d'ammonium ( $\text{NH}_4\text{HCO}_3$  M= 79.06 g/mol) 50 mM dans  $\text{H}_2\text{O}$
- Acétate d'ammonium (M= 77.08 g/mol) 0.1 M dans MeOH 100 %
- MeOH 80 % : 8 mL méthanol pure 100% + 2 mL  $\text{H}_2\text{O}$
- Acétate d'ammonium (M= 77.08 g/mol) 0.1 M dans MeOH 80 %

-

- Solution à préparer le jour même de la manip

-

- DTT à 0.1 M dans 50 mM de Bicarbonate d'ammonium
- Iodoacetamide 0.1 M dans 50 mM de Bicarbonate d'ammonium

$$V = m / M C$$

### ⇒ Précipitation

#### Précipitation :

- Pour 25 $\mu\text{L}$  de protéine cross-linké, on rajoute environ 5X le volume de solution d'acétate d'ammonium dans 100% MeOH glacé. (On peut rajouter un peu de tampon de bicarbonate d'ammonium si on n'a pas assez de volume) (environ 125 $\mu\text{L}$ )
- Mettre au congélateur overnight de préférence (> 2 H à -20 °C)
- Centrifuger à 12000RCF à 4 °C pendant 15 min

*Dans le culot on a les protéines + sels*

- Mettre papier sopalin
- Retourner le tube pour éliminer le surnageant

#### Lavage :

- Ajouter 125  $\mu\text{L}$  pour laver avec 0.1 M d'acétate d'ammonium dans 80 % MeOH congelé (*MeOH à 80 % permet de solubiliser les sels*)
- Centrifuger 15 min à 12000 RCF à 4°C
- Retourner le tube pour éliminer le surnageant (*verser verticalement le tube et le faire plusieurs fois*)

*Ne pas vortexer sinon risque de détruire le culot*

*Enlever MeOH avant de mettre au speed vac*

- Sécher au Speed Vac à T°amb pendant 15 min
- Reprendre le culot dans 100 µL de bicarbonate d'ammonium

Réduction : réduire pont S-S

- Ajouter 5 µL DTT à 0.1 M
- Vortex (*puissance max*)
- Laisser 10 min à 95°C
- Centrifuger 10 000 RCF pendant 15 s

Alkylation : iodo est en excès car doit réagir avec DTT et protéine.

- Ajouter 10 µL Iodoacetamide à 0.1 M
- Garder la solution d'iodoacetamide à l'obscurité avec papier aluminium
- Mettre papier alu autour du tube
- Attendre 25 min à T°ambiante = ne pas laisser + longtemps

Quencher la réaction : mettre la même quantité de DTT et iodo = stopper la réaction c-à-d bloquer les iodo restant dans la solution

- Ajouter 5 µL de DTT à 0.1 M
- Vortexer léger

Digestion :

- Dans un aliquot de 1 µg de trypsine, ajouter 5 µL de Bicarbonate d'ammonium à 50 mM
- Vortexer légèrement et centrifuger
- Ajouter 2 µL de trypsine
- Laisser à 37 °C toute la nuit
- Centrifuger à 12000 RCF pendant 15 min à 4°C (*pour éliminer les grosses particules*)
- Speed vac à sec de l'échantillon
- Mettre l'échantillon au congélateur à -20 °C
- Ajouter 100 µL d'acide formique à 0.1 % pour solubiliser les peptides obtenus.

*L'acide formique va protonner les peptides donc ils seront chargés (on aura des ions peptidique)*

## Annexe 6 : Protocole d'enrichissement SCX

### Column Conditioning :

- 1) Remove the white bottom cap and place the spin column inside a **2.0 mL sample tube**.
- 2) Spin the column at 5,000 g for 1 minute to collect the resin in the barrel; discard the flowthrough.
- 3) Add 300  $\mu$ L of Acetonitrile
- 4) Add 300  $\mu$ L of 1% formic acid and spin at 2,000 g for 1 minute; discard the flowthrough.
- 5) Add 300  $\mu$ L of 0.1% formic acid in water to the spin column and spin at 2,000 g for 1 minute; discard the flowthrough.

### Sample Loading :

- 1) Dissolve sample in 300  $\mu$ L of 0.1% formic acid solution.
- 2) Load the sample into the spin column and spin at 2,000 g for 1 minute; discard the flowthrough.
- 3) Add 300  $\mu$ L of 0.1% formic acid to the spin column and spin at 2,000 g for 1 minute; discard the flowthrough.
- 4) Repeat step (3).
- 5) Add 300  $\mu$ L of 30% ACN in 0.1% formic acid solution to the spin column and spin at 2,000 g for 1 minute; discard the flowthrough.
- 6) Repeat step (5).

### Sample Elution (use 2 mL sample tubes for fraction collection) :

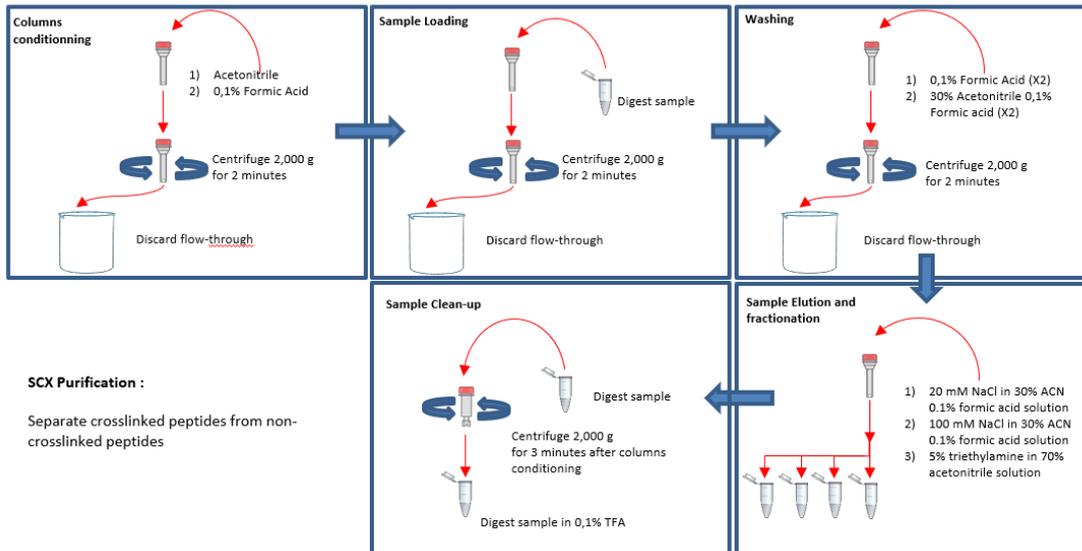
- 1) Add 300  $\mu$ L of 20 mM NaCl in 30% ACN 0.1% formic acid solution and spin at 2,000 g for 1 minute. Keep the eluate.
- 2) Repeat step (1) four more times.
- 3) Combine all eluates from steps (1) and (2) in a single sample tube. **This is the 20 mM NaCl fraction.**
- 4) Add 300  $\mu$ L of 100 mM NaCl in 30% ACN 0.1% formic acid solution and spin at 2,000 g for 1 minute. Keep the eluate.
- 5) Repeat step (1) four more times.
- 6) Combine all eluates from steps (4) and (5) in a single sample tube. **This is the 100 mM NaCl fraction.**
- 7) Add 300  $\mu$ L of 5% triethylamine in 70% acetonitrile solution to the spin column and spin at 2,000 g for 1 minute.
- 8) Repeat step (7) one more time, collecting the eluate into the same tube. **This is the crossed-linked peptide enriched fraction.**

### Sample Clean-up (do this for each fraction using high pH reversed-phase spin-columns; PN: 84868):

- 1) Dry the sample in a vacuum centrifuge.
- 2) Resuspend the sample (each fraction) in 300  $\mu$ L of 0.1% TFA.
- 3) ***\*Use high pH reversed-phase spin columns (product number 84868)\****

- After conditioning the column, add the sample to the spin-column and spin at 3,000 g for 2 minutes.
- 4) Add 300  $\mu$ L of 0.1% TFA to the spin column and spin at 3,000 g for 2 minutes.
- 5) Repeat step (4) two more times to fully desalt the sample.

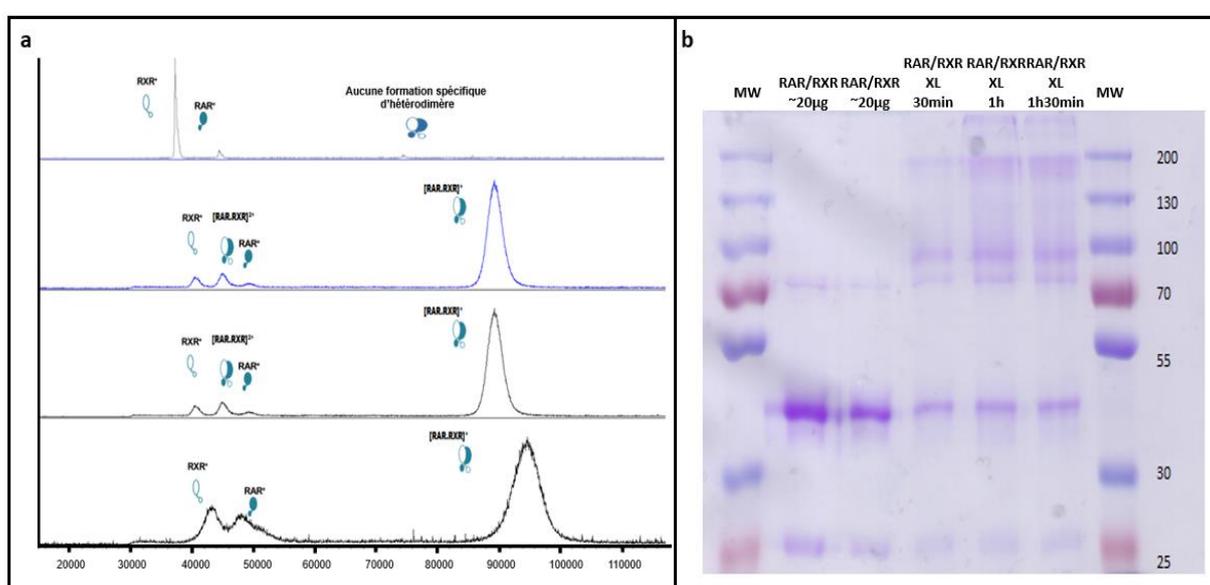
Elute with 70% ACN 0.1% TFA solution (two elutions with 300  $\mu$ L each time)



## Annexe 7 : Suivi et optimisation de la réaction de pontage

Les conditions expérimentales ont été optimisées afin que l'agent de pontage puisse potentiellement réagir sur l'ensemble de l'échantillon mais à une concentration plutôt faible d'agent de pontage. En effet, la concentration en agent pontant doit être la plus faible possible pour être certain que le composé ne perturbe pas la structure du complexe qui peut dans certains cas s'agréger et réaliser des interactions non spécifiques.

Ainsi, afin d'optimiser la réaction de pontage, des analyses par MALDI-MS et par SDS-PAGE du complexe RAR/RXR avant et après pontage ont été réalisées. Pour cela, un échange de tampon a ainsi été réalisé dans un premier temps afin d'éliminer une grande partie des sels et le glycérol. Après échange, le complexe est présent dans un tampon Hepes à 20mM avec 150mM NaCl à pH 7.5. Puis, dans un second temps, nous avons testé le pontage avec le BS3 avec différents ratios molaire protéine/agent pontant (1 :50 et 1 :200) et avec différents temps de réaction (30min, 1h30, 2h00).



**Figure 125 : (a) Spectres MALDI-MS du complexe avant et après pontage chimique avec le BS3 à différentes conditions réactionnelles (1) avant pontage (2) après 30min à  $T_{amb}$  à un ratio protéine/BS3 (1:50) (3) après 1h00 à  $T_{amb}$  à un ratio protéine/BS3 (1:50) (4) après 1h00 à  $T_{amb}$  à un ratio protéine/BS3 (1:250). La figure b représente l'analyse SDS-PAGE sur un gel d'électrophorèse à 10% acrylamide du complexe RAR-RXR purifié avant et après la réaction de pontage. La stœchiométrie du complexe ainsi que la réaction de pontage ont été cinétiquement suivies par MALDI-MS et SDS-PAGE (figure a). Ainsi, plusieurs ions du complexe ont été détectés avant pontage à  $44509 \pm 70$  Da et  $37300 \pm 70$  Da par gel et par MALDI-MS. Ces ions correspondent respectivement au composé RAR et RXR qui constituent le complexe. Après pontage, un poids moléculaire de  $84994 \pm 200$  Da a été détecté. Aucun dimère ou tétramère du complexe réticulé n'a été trouvé par MALDI-MS. Cette masse correspond à la formation hétérodimérique du complexe RAR/RXR ponté selon une stœchiométrie 1 :1. Le signal correspondant au complexe ponté complet est grand et très intense, montrant ainsi une bonne efficacité de réaction.**

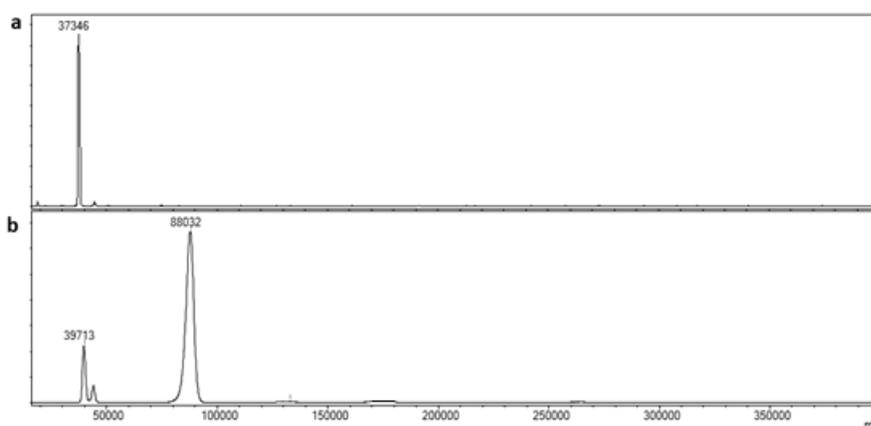
La stœchiométrie du complexe ainsi que la réaction de pontage ont été cinétiquement suivies dans un premier temps par MALDI-MS (figure a). Ainsi, plusieurs ions du complexe ont été détectés avant pontage à  $44718 \pm 70$  Da et  $37315 \pm 70$  Da par MALDI-MS. Aucuns ions du complexe hétérodimérique n'a alors été détecté. Ces ions correspondent respectivement au composé RAR et RXR qui constituent le complexe. Après pontage, un poids moléculaire de  $85917 \pm 200$  Da a été détecté. Aucun dimère ou tétramère du complexe ponté n'a été trouvé

par MALDI-MS. Cette masse correspond à la formation hétérodimérique du complexe RAR/RXR ponté selon une stœchiométrie 1 :1. Le signal correspondant au complexe ponté complet est grand et très intense, montrant ainsi une bonne efficacité de réaction.

De plus, un gel d'électrophorèse SDS-PAGE à 10% (figure b) a été effectué en parallèle des analyses MALDI-MS pour confirmer la formation du complexe ponté en utilisant une coloration bleue de Coomassie compatible avec la MS. Au sein du gel, nous avons effectué un premier dépôt du complexe avant que la réaction de pontage ne soit effectuée ainsi que d'autres dépôts du complexe ponté selon divers conditions réactionnelle afin d'optimiser les paramètres de la réaction (temps, rapport protéine/agent pontant). Plusieurs temps de réaction (30 min, 1 h et 1h30) et plusieurs rapports protéine / réticulant (1/50 et 1/250) ont été testés.

Les analyses effectuées par MALDI-MS et par SDS-PAGE ont ainsi montré que la réaction semble être réalisée après 30min avec un rapport protéine /agent pontant au 1/50 où environ 80% du complexe est ponté. En effet, sur le gel, on peut observer l'apparition d'une bande à environ 80 kDa qui semble correspondre au complexe RAR-RXR et qui confirme alors les données obtenues en MALDI-MS. Cependant, la formation d'un complexe de haut poids moléculaire (environ 160 kDa) peut également être observée sur le gel. Après analyse protéomique sur la bande de gel, il a été trouvé que le complexe formé était un tétramère de RXR qui peut être physiologiquement formable puisque RXR est connu pour former des complexes tétramériques.<sup>[100, 101]</sup> De plus, l'analyse MALDI ne montre qu'une spécificité de la réaction entre RXR et RAR car aucun autre complexe covalent entre RAR et RXR n'est formé.

Afin de confirmer que le complexe formé est uniquement hétérodimérique, une analyse MALDI-MS en présence d'un détecteur à haute résolution de type CovalX<sup>®</sup> a été réalisé par un laboratoire de Bordeaux. Il s'agit d'un détecteur de haute masse disponible commercialement et basé sur la technologie « ion conversion dynode ». Grâce à ce détecteur il a été possible d'effectuer des analyses MALDI de notre complexe RAR/RXR ponté au-delà de la limite habituelle du spectromètre (à environ 120 000Da) pour effectuer des analyses jusqu'à 1 MDa.



**Figure 126 : Spectre obtenu sur un MALDI-MS équipé d'un détecteur CovalX<sup>®</sup> de RAR/RXR10 $\mu$ M avant pontage (a) et après réaction de pontage réalisé pendant 30min à Tamb (b).**

Les spectres MALDI à haute résolution ont ainsi permis de confirmer ce qui a été observé sur le gel SDS-PAGE. Après réaction de pontage, nous avons pu observer une masse à 88032 correspondant à la masse de l'hétérodimère de RAR/RXR. Au cours des différentes

analyses MALDI réalisé sur cet appareil, aucune masse supérieure à  $88032 \pm Da$  a été détecté montrant ainsi que la formation du complexe hétérodimérique était une réaction spécifique.

Des études supplémentaires sur la formation de complexes à haut poids moléculaires détectables par SDS-PAGE sont encore en cours afin de confirmer la formation spécifique de l'hétérodimère.

Toutefois, il a été démontré dans une étude que les protéines pontées peuvent migrer de manière particulière sur les gels. En effet, les analyses de peptides pontés sur gel peuvent conduire à une compaction de la structure et entraîne une migration plus importante sur le gel. <sup>[102]</sup> Ce phénomène pourrait ainsi être observé sur le complexe RAR-RXR après pontage et expliquerait ainsi l'étrange migration observé sur le gel.

#### Paramètre optimisé :

Ainsi, nous avons sélectionné les conditions de pontage de RAR/RXR avec le BS3 à un ratio protéine/BS3 1 :50 et à un temps de réaction de 30min à température ambiante. Les conditions réactionnelles sont ainsi en accord avec les conditions de réaction typiquement observées pour le pontage avec le BS3 observé dans des autres études récentes. De plus, grâce au analyses MALDI et aux analyses effectuées par SDS-PAGE, il a été démontré que la réaction de pontage était une réaction spécifique.

Par la suite, la réaction de pontage est arrêtée avec une solution de Tris à 25mM en volume final. De plus, les conditions d'analyse MALDI ont également été optimisé afin d'arriver à avoir un rapport signal sur bruit correct. Le meilleur signal a été observé avec une puissance de 50% et avec une accumulation de 5000 tirs.

## Annexe 8 : Tableau analyse quantitative complet

Peptide Modified Sequence	Normalized Area	conformer I/Conformer I	Singlet or doublet	Conformer I/Conformer II	log2(Conformer/Conformer II)
AHQET[+160.1]FPALC[+114.1]QLGKYTTNNSSEORE[+29]NENEVESTS[+1]S[+160.1]ANEDMPVERILEAEALVEP	9,07E+08		0,00E+00 singlet		50
AHQETFPALCQLGK[+1]E[+25]DCLIDK[+1]ROR	3,87E+07		1,32E+00 Doublet	7,60E-01	-0,396080747
AHQETFPALCQLGK[+1]E[+25]YQK[+1]CLAMGMK	1,50E+07		8,24E-01 Doublet	1,21E+00	0,278949059
AHQETFPALCQLGK[+1]E[+29]KLYTYCRDNK[+1]DCLIDK	4,84E+07		0,00E+00 singlet		50
AHQETFPALCQLGK[+1]E[+29]YQK[+1]CLAMGMKREAVQEQER	1,24E+07		8,24E-01 Doublet	1,21E+00	0,278949059
AHQETFPALCQLGK[+1]YTTNNSSEORE[+29]DCLIDK[+1]R	5,09E+07		1,32E+00 Doublet	7,60E-01	-0,396080747
AHQETFPALCQLGK[+1]YTTNNSSEORE[+29]DNKDCLIDK[+1]R	3,14E+06		8,11E-02 Doublet	1,23E+01	3,623683485
CFEFGMSK[+1]ESVRE[+25]AIVLFPDPSK[+1]GLSNPAEVEALR	5,20E+07		0,00E+00 singlet		-50
CFEFGMSK[+1]ESVRE[+25]HK[+1]YPEQGR	2,05E+07		0,00E+00 singlet		-50
CFEFGMSK[+1]ESVRE[+25]NENEVESTSANEDMPVERILEAEALVEPK[+1]TETVVEANMGLNPPSPNDPVTNICQAM	1,81E+07		0,00E+00 singlet		-50
CFEFGMSK[+1]ESVRE[+29]EAVQEQERQGRK[+1]DR	5,51E+07		0,00E+00 singlet		50
CFEFGMSK[+1]ESVRE[+29]YQK[+1]CLAMGMK	2,90E+08		0,00E+00 singlet		50
CIHK[+1]TVFEFAKE[+29]K[+1]DLTYTCR	2,86E+09		0,00E+00 singlet		-50
CIHK[+1]TVFEFAKE[+29]CLAMGMK[+1]R	3,93E+06		0,00E+00 singlet		-50
COYCRLOK[+1]CFEFGMSKESVRE[+25]QLFTLVEWAK[+1]R	8,67E+09		0,00E+00 singlet		-50
DK[+1]N[+114.1]IINK[+155.1]VTRE[+25]AIVLFPDPSK[+1]GLSNPAEVEALR	1,89E+09		8,87E-01 Doublet	1,13E+00	0,172751944
DK[+1]N[+114.1]IINK[+156.1]VTRE[+25]DNKDC[+114.1]IIDI[+1]R	1,23E+10		4,90E-02 Doublet	3,72E+02	8,538323919
DK[+1]N[+114.1]IINK[+156.1]VTRE[+25]K[+1]DLTYTC[+114.1]R	1,38E+08		2,54E+00 Doublet	5,62E+00	2,489850164
DK[+1]N[+114.1]IINK[+156.1]VTRE[+25]SIAVK[+1]IDGILLATGLHVHR	1,50E+06		7,52E+02 Doublet	1,33E-03	-9,554452066
DK[+1]N[+114.1]IINK[+160.1]VTRE[+29]AIVLFPDPSK[+1]GLSNPAEVEALR	1,68E+09		8,87E-01 Doublet	1,13E+00	0,172751944
DK[+1]N[+114.1]IINK[+160.1]VTRE[+29]K[+1]DLTYTC[+114.1]R	2,46E+07		1,78E-01 Doublet	5,62E+00	2,489850164
DK[+1]N[+114.1]IINK[+25]K[+1]DLTYTC[+114.1]R	3,90E+08		1,99E+02 Doublet	5,04E-03	-7,633459413
DK[+1]N[+114.1]IINK[+29]K[+1]DLTYTC[+114.1]R	7,74E+10		1,99E+02 Doublet	5,04E-03	-7,633459413
DK[+1]N[+114.1]IINK[+29]SIAVK[+1]IDGILLATGLHVHR	1,13E+09		7,52E+02 Doublet	1,33E-03	-9,554452066
DK[+1]N[+114.1]IINKVTR[+25]DNK[+1]DC[+114.1]IIDI[+156.1]R	1,33E+10		1,82E-01 Doublet	5,51E+00	2,461485386
DK[+1]N[+114.1]IINKVTR[+29]AIVLFPDPSK[+1]GLSNPAEVEALR	3,10E+08		1,21E-01 Doublet	6,10E+00	2,607750185
DK[+1]N[+114.1]IINKVTR[+29]DNK[+1]DC[+114.1]IIDI[+160.1]R	2,41E+09		1,82E-01 Doublet	5,09E+00	2,348593855
DK[+1]N[+114.1]IINKVTR[+29]K[+1]DLTYTC[+114.1]R	3,50E+08		8,98E-01 Doublet	3,94E-01	-1,342484884
DK[+1]N[+114.1]IINKVTR[+29]SIAVK[+1]IDGILLATGLHVHR	2,56E+09		6,56E-01 Doublet	8,24E+00	3,043288338
DK[+1]N[+114.1]IINKVTR[+29]YQK[+1]GLSNPAEVEALR	8,65E+06		0,00E+00 singlet		50
DK[+1]N[+114.1]IINKVTR[+29]GFFK[+1]R	2,40E+06		0,00E+00 singlet		-50
DK[+1]N[+114.1]IINKVTR[+29]GFFK[+1]RTVTR	2,40E+06		0,00E+00 singlet		-50
DK[+1]N[+114.1]IINKVTR[+29]K[+1]DLTYTCR	9,74E+07		5,30E-01 Doublet	1,89E+00	0,914679652
DK[+1]N[+114.1]IINKVTR[+29]RTVTRK[+1]R	4,49E+07		0,00E+00 singlet		-50
DK[+1]N[+114.1]IINKVTR[+29]K[+1]DLTYTCR	5,17E+07		5,30E-01 Doublet	1,89E+00	0,914679652
DK[+1]N[+114.1]IINKVTR[+29]DCLIDK[+1]ROR	4,53E+08		0,00E+00 singlet		-50
DK[+1]N[+114.1]IINKVTR[+29]FAK[+1]LLLR	3,59E+07		0,00E+00 singlet		50
DK[+1]N[+114.1]IINKVTR[+29]SIAVK[+1]IDGILLATGLHVHR	2,83E+07		1,88E+01 Doublet	5,31E-02	-4,234258348
DK[+156.1]N[+114.1]IINKVTR[+25]K[+1]DLTYTC[+114.1]R	1,29E+07		4,91E-01 Doublet	5,25E+00	2,391291605
DKNCIINK[+1]E[+25]GFFK[+1]R	1,34E+07		5,39E+00 Doublet	1,86E-01	-2,430127582
DKNCIINK[+1]E[+29]GFFK[+1]RTVTR	7,23E+07		5,39E+00 Doublet	1,86E-01	-2,430127582
DKNCIINK[+1]VTR[+25]CLAMGMK[+1]R	1,09E+09		0,00E+00 singlet		-50
DKNCIINK[+1]VTR[+25]DCLIDK[+1]R	3,28E+06		3,78E-01 Doublet		-4,724131147
DKNCIINK[+1]VTR[+25]DMQMDK[+1]TELGLCLR	5,21E+08		0,00E+00 singlet		50
DKNCIINK[+1]VTR[+25]DNKDCLIDK[+1]R	3,30E+07		1,82E+01 Doublet	5,49E-02	-4,186832437
DKNCIINK[+1]VTR[+25]K[+1]DLTYTCR	7,37E+07		6,11E-01 Doublet	1,64E+00	0,71104359
DKNCIINK[+1]VTR[+25]VLTLSVK[+1]MR	2,05E+06		2,64E+01 Doublet	3,78E-02	-4,724131147
DKNCIINK[+1]VTR[+29]DCLIDK[+1]R	8,67E+07		2,64E+01 Doublet	3,78E-02	-4,724131147
DKNCIINK[+1]VTR[+29]DNK[+1]DCLIDK	4,03E+08		0,00E+00 singlet		50
DKNCIINK[+1]VTR[+29]DNKDCLIDK[+1]R	6,02E+08		1,82E+01 Doublet	5,49E-02	-4,186832437
DKNCIINK[+1]VTR[+29]K[+1]DLTYTCR	4,50E+07		4,50E-01 Doublet	1,64E+00	0,71104359
GFRRSISQK[+1]NMVYTCR[+29]CLAMGMK[+1]REAVQEQER	5,00E+05		0,00E+00 singlet		50
ITDLRSISAK[+1]GAERE[+25]AIVLFPDPSK[+1]R	4,59E+08		0,00E+00 singlet		-50
ITDLRSISAK[+1]GAERE[+25]TVRK[+1]DLTYTCRDNK	3,53E+07		1,55E-01 Doublet	6,46E+00	2,690654444
ITDLRSISAK[+1]GAERE[+25]VYASLEAYCK[+1]HKYPEQGRFAK	4,97E+07		0,00E+00 singlet		-50
ITDLRSISAK[+1]GAERE[+29]TVRK[+1]DLTYTCRDNK	8,47E+06		1,55E-01 Doublet		2,690654444
K[+1]AHQETFPALC[+114.1]QLGKE[+25]K[+1]DLTYTC[+114.1]R	2,53E+10		7,23E-03 Doublet	1,38E+02	7,111366747
K[+1]AHQETFPALC[+114.1]QLGKE[+29]K[+1]DLTYTC[+114.1]R	1,83E+08		7,23E-03 Doublet	1,38E+02	7,111366747
K[+1]AHQETFPALCQLGK[+29]K[+1]DLTYTCR	1,22E+07		6,79E-04 Doublet	1,47E+03	10,52385427
K[+1]KKYVPEK[+25]SIVLTVSKMRDMQMDK[+1]TELGLCLR	8,90E+06		0,00E+00 singlet		50
K[+1]RRPSRPHMFPKE[+29]K[+1]DLTYTCR	2,45E+06		0,00E+00 singlet		-50
KAHQETFPALCQLGK[+1]E[+25]YQK[+1]CLAMGMKREAVQEQER			0,00E+00 singlet		-50
KAHQETFPALCQLGK[+1]YTTNNSSEORE[+29]DNKDCLIDK[+1]R	1,99E+07		0,00E+00 singlet		50
KAHQETFPALCQLGK[+1]YTTNNSSEORE[+29]IDNKDCLIDK[+1]R	5,93E+06		0,00E+00 singlet		50
KEVPK[+1]PECSYTLTPEVGLIEKE[+25]DCLIDK[+1]R	5,65E+07		0,00E+00 singlet	0,00E+00	-50
KEVPK[+1]PECSYTLTPEVGLIEKE[+29]EAVQEQERQGRK[+1]DR	1,33E+06		0,00E+00 singlet		50
KEVPK[+1]PECSYTLTPEVGLIEKVRKE[+25]VYASLEAYCK[+1]HKYPEQGR	1,76E+07		0,00E+00 singlet		-50
KK[+1]IEVPK[+25]VLTLSVKMRDMQMDK[+1]TELGLCLR	1,60E+06		0,00E+00 singlet		-50
KK[+1]IEVPK[+29]YQK[+1]CLAMGMK	9,54E+06		0,00E+00 singlet		50
KK[+1]IEVPK[+29]SIVLTVSKMRDMQMDK[+1]TELGLCLR	4,56E+08		0,00E+00 singlet		-50
KK[+1]IEVPK[+25]DMQMDK[+1]TELGLCLR	2,01E+07		0,00E+00 singlet		-50
KK[+1]IEVPK[+25]EAVYASLEAYCK[+1]HKYPEQGR	2,91E+06		2,88E-03 Doublet	1,70E+02	7,411081578
KK[+1]IEVPK[+25]GLSNPAEVEALREK[+1]VYASLEAYCKHK	7,85E+06		0,00E+00 singlet		-50
KK[+1]IEVPK[+29]EAVYASLEAYCK[+1]HKYPEQGR	1,71E+07		5,88E-03 Doublet	1,70E+02	7,411081578
KKEVPK[+1]IPEC[+114.1]SIVLTVSKMRDMQMDK[+29]AIVLFPDPSK[+1]GLSNPAEVEALR	2,93E+09		0,00E+00 singlet		50
LQK[+1]CFEFGMSK[+29]DMQMDK[+1]TELGLCLR	6,94E+08		0,00E+00 singlet		50
LQK[+1]CFEFGMSKESVRE[+25]AIVLFPDPSK[+1]GLSNPAEVEALREK	7,88E+07		0,00E+00 singlet		-50
LQK[+1]CFEFGMSKESVRE[+25]EAVQEQERQGRK[+1]R	4,60E+07		0,00E+00 singlet		-50
LQK[+1]CFEFGMSKESVRE[+25]VPAHPSGNMASFTK[+1]IHICAIQGR	1,35E+07		0,00E+00 singlet		-50
LQK[+1]CFEFGMSKESVRE[+29]HK[+1]YPEQGR	8,87E+08		0,00E+00 singlet		50
LQKCFEFGMSK[+1]E[+29]MRDMQMDK[+1]TELGLCLR	8,73E+07		0,00E+00 singlet		50
LQKCFEFGMSK[+1]ESVRE[+29]VLTLSVK[+1]MRDMQMDKTELGLCLR	1,26E+07		0,00E+00 singlet		50
LQKCFEFGMSK[+1]ESVRE[+29]FAK[+1]LLLR	1,84E+06		0,00E+00 singlet		50
MLMK[+1]ITDLRE[+25]AIVLFPDPSK[+1]GLSNPAEVEALR	1,23E+08		2,37E+00 Doublet	4,22E-01	-1,243842064
MLMK[+1]ITDLRE[+25]K[+1]DLTYTC[+114.1]R	2,24E+08		2,24E+00 Doublet		-50
MLMK[+1]ITDLRE[+29]AIVLFPDPSK[+1]R	2,91E+08		2,37E+00 Doublet	4,22E-01	-1,243842064
MLMK[+1]ITDLRSISAKE[+29]VLTLSVK[+1]MRDMQMDKTELGLCLR	1,78E+07		0,00E+00 singlet		50
MLMK[+1]ITDLRSISAKGAERE[+25]AIVLFPDPSK[+1]R	8,11E+06		0,00E+00 singlet		-50
MLMK[+1]ITDLRSISAKGAERVTI[+25]QLFTLVEWAK[+1]R	1,12E+08		0,00E+00 singlet		50
MLMKITDLRSISAK[+1]GAERE[+25]GLSNPAEVEALREK[+1]R	4,25E+06		0,00E+00 singlet		-50
NCIINK[+1]E[+25]RTVTRK[+1]R	2,54E+06		9,01E+01 Doublet	1,11E-02	-6,493447454
NCIINK[+1]E[+29]RTVTRK[+1]R	2,29E+08		9,01E+01 Doublet	1,11E-02	-6,493447454

NCIINK(+1)VIRE(+25)DMQMDK(+1)	3.00E+08	0.00E+00	singlet	0	-
NCIINK(+1)VIRE(+25)K(+1)DLTYTCR	3.17E+08	1.27E+00	Doublet	7.87E-01	-0.345117104
NCIINK(+1)VIRE(+25)YQK(+1)CLAMGMKREAVQERQR	1.22E+07	0.00E+00	singlet	0	-50
NCIINK(+1)VIRE(+29)DKLIDK(+1)R	3.40E+08	4.88E+01	Doublet	2.05E-01	-5.607691168
NCIINK(+1)VIRE(+29)DNK(+1)DCLIDKR	3.19E+08	7.92E-01	Doublet	2.919399742	-
NCIINK(+1)VIRE(+29)DNKDCIDK(+1)R	3.54E+08	4.67E+00	Doublet	2.14E-01	-2.224685718
NCIINK(+1)VIRE(+29)K(+1)DLTYTCR	3.48E+08	1.27E+00	Doublet	7.87E-01	-0.345117104
NCIINK(+1)VIRE(+29)SIGLK(+1)CLEHFFKLGIDPIDTFLMEMLEAPHQMT	3.59E+09	0.00E+00	singlet	0	-50
NDRNKR(+1)KE(+25)AVLFPNPSK(+1)GLSNPAEVALREK	3.19E+08	0.00E+00	singlet	0	-50
NK(+1)KKKVEPKRE(+29)VYASLEAYCK(+1)HKYPEQGR	3.97E+08	0.00E+00	singlet	0	-50
NMVVTCIINK(+1)VIRE(+29)EK(+1)VYASLEAYCKHYPEQGR	3.20E+07	0.00E+00	singlet	0	-50
PECESYTLTPVEGLIEK(+1)VIRE(+25)DMQMDK(+1)TELGCLR	4.01E+07	0.00E+00	singlet	0	-50
PHMFPK(+1)MLMKRDLRSISAKGAERE(+29)AVLFPNPSK(+1)	3.32E+08	4.97E-01	Doublet	2.01E+00	3.007396623
PHMFPKMLMK(+1)ITDLRSISAKGAERE(+25)AVLFPNPSK(+1)	6.28E+08	4.97E-01	Doublet	2.01E+00	3.007396623
PHMFPKMLMK(+1)ITDLRSISAKGAERE(+29)PHSELPDDCVILLRAGWNELIASFSHRSIAVK(+1)DGILLATGLHVH	5.99E+06	0.00E+00	singlet	0	-50
PSRPHMFPK(+1)MLMKE(+25)IGYCVYQK(+1)CLAMGMKREAVQER	4.12E+06	0.00E+00	singlet	0	-50
PSRPHMFPK(+1)MLMKE(+25)HK(+1)YPEQGR	3.04E+08	0.00E+00	singlet	0	-50
PSRPHMFPK(+1)MLMKTDLRE(+29)SIYAVK(+1)DGILLATGLHVHRSNSAHSAGVGAIFDR	6.10E+07	0.00E+00	singlet	0	-50
QDLEQDPRVDMLEPELLEALK(+1)E(+25)DLTLVEVAK(+1)	3.48E+06	0.00E+00	singlet	0	-50
QDLEQDPRVDMLEPELLEALK(+1)E(+29)AVLFPNPSK(+1)GLSNPAEVALREKVYASLEAYCKHK	7.99E+07	0.00E+00	singlet	0	-50
QDLEQDPRVDMLEPELLEALK(+1)YVYRE(+25)FAK(+1)LLLR	4.76E+07	0.00E+00	singlet	0	-50
QLPSITTLTADOTLTK(+1)AACLIDILIRE(+29)MRDMQMDK(+1)ITELGCLR	5.06E+09	0.00E+00	singlet	0	-50
RPSRPHMFPK(+1)MLMKE(+25)AVLFPNPSK(+1)GLSNPAEVALRE	3.80E+08	2.58E+00	Doublet	3.88E-01	-1.367135581
RPSRPHMFPK(+1)MLMKE(+25)K(+1)DLTYTCR(+1)14.1JR	2.27E+10	0.00E+00	singlet	0	-50
RPSRPHMFPK(+1)MLMKE(+25)K(+1)DLTYTCR	3.81E+09	0.00E+00	singlet	0	-50
RPSRPHMFPK(+1)MLMKE(+29)AVLFPNPSK(+1)GLSNPAEVALRE	4.64E+08	2.58E+00	Doublet	3.88E-01	-1.367135581
RPSRPHMFPK(+1)MLMKE(+29)YQK(+1)CLAMGMKREAVQERQR	1.69E+08	3.73E+00	Doublet	2.68E-01	-1.898234283
RRPSRPHMFPK(+1)MLMKE(+25)ITDLRSISAK(+29)INENEVEST(+1)SSANEDMPVER	2.91E+07	0.00E+00	singlet	0	-50
RRPSRPHMFPK(+1)MLMKE(+25)YQK(+1)CLAMGMK	3.54E+07	3.73E+00	Doublet	2.68E-01	-1.898234283
RRPSRPHMFPK(+1)MLMKE(+29)NSAHSAGVGAIFDRVLTLSVK(+1)MRDMQMDKTELGCLR	4.99E+06	0.00E+00	singlet	0	-50
RSIOK(+1)NMLL(10)VYTC(+1)14.1JHRE(+29)AVLFPNPSK(+1)K	2.12E+08	0.00E+00	singlet	0	-50
RSIOK(+1)NMYTCGRDNK(+29)HK(+1)YPEQGR	7.61E+06	0.00E+00	singlet	0	-50
RSIOK(+1)NMYVTC(HRE(+25)YQK(+1)CLAMGMK	7.61E+06	0.00E+00	singlet	0	-50
RSIOKNMVVTC(HRDK(+1)NINCINKVIRE(+29)AVLFPNPSK(+1)	2.88E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)K(+1)14.1JAMGMK(+1)R	2.07E+10	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)DMQMDK(+1)ITELGCLR(+1)14.1JR	1.17E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)DMQMDK(+1)DK(+1)14.1JLIDK(+1)156.1JR	4.36E+09	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)DNKDCIDK(+1)14.1JLIDK(+1)R	8.68E+08	3.14E+01	Doublet	3.18E-02	-4.974356741
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)K(+1)DLTYTCR(+1)14.1JR	5.46E+08	2.09E+01	Doublet	4.78E-02	-4.385927479
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)SSGK(+1)HYGVYSC(+1)14.1JEGC(+1)14.1JK	2.84E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+25)VYASLEAY(+1)C(+1)14.1JHKHYPEQGR	2.07E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)AVLFPNPSK(+1)GLSNPAEVALRE	3.60E+08	8.66E+00	Doublet	1.15E-01	-3.11502088
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)AVLFPNPSK(+1)GLSNPAEVALREK	3.28E+08	5.78E+00	Doublet	1.23E-01	-3.022773325
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)EK(+1)VYASLEAY(+1)14.1JK	4.66E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)FAK(+1)LLLR	2.49E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)GLS(+1)NPAEVALREK	2.10E+09	0	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)K(+1)DLTYTCR(+1)14.1JR	1.34E+10	2.09E+01	Doublet	4.78E-01	-4.385927479
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)SIYAVK(+1)DGILLATGLHVH	3.13E+09	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(+1)14.1JHRE(+29)SIYAVK(+1)DGILLATGLHVH	2.59E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRDK(+25)AVLFPNPSK(+1)GLSNPAEVALRE	3.88E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRDK(+25)EKVYASLEAYCK(+1)HKYPEQGR	1.37E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRDK(+25)VPAHPSGNMASTP(+1)HICCAICDDR	3.21E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRDK(+29)AVLFPNPSK(+1)GLSNPAEVALRE	4.05E+08	3.00E+00	Doublet	2.03E-01	-2.298992173
SIOK(+1)NMYVTC(HRDKNCINKVIRE(+29)GLSNPAEVALREK(+1)	2.54E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRE(+25)AVLFPNPSK(+1)GLSNPAEVALREK	4.16E+07	1.23E+01	Doublet	2.03E-01	-2.298992173
SIOK(+1)NMYVTC(HRE(+25)CLAMGMK(+1)R	2.44E+09	2.09E+00	Doublet	4.78E-01	-4.385927479
SIOK(+1)NMYVTC(HRE(+25)DKLIDK(+1)R	1.43E+08	1.91E+02	Doublet	5.25E-03	-7.574302819
SIOK(+1)NMYVTC(HRE(+25)DNK(+1)DCLIDKR	7.08E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRE(+25)DNKDCIDK(+1)R	3.31E+10	2.09E+00	Doublet	4.78E-01	-4.385927479
SIOK(+1)NMYVTC(HRE(+25)GFKK(+1)R	7.25E+07	1.39E+00	Doublet	7.19E-01	-0.475263657
SIOK(+1)NMYVTC(HRE(+25)GNRCVYCVYQK(+1)CLAMGMK	1.45E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRE(+29)AVLFPNPSK(+1)GLSNPAEVALRE	9.10E+08	9.78E-02	Doublet	8.15E-02	-3.616204144
SIOK(+1)NMYVTC(HRE(+29)AVLFPNPSK(+1)GLSNPAEVALREK	9.33E+07	2.24E+00	Doublet	4.46E-01	-1.165973002
SIOK(+1)NMYVTC(HRE(+29)CLAMGMK(+1)R	3.84E+09	7.52E-01	Doublet	1.33E+00	0.411343016
SIOK(+1)NMYVTC(HRE(+29)DNKDCIDK(+1)R	2.73E+10	2.09E+00	Doublet	4.78E-01	-4.385927479
SIOK(+1)NMYVTC(HRE(+29)GFKK(+1)R	3.01E+08	1.39E+00	Doublet	7.19E-01	-0.475263657
SIOK(+1)NMYVTC(HRE(+29)HK(+1)YPEQGR	2.30E+08	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRE(+29)TVRCK(+1)DLTYTCRDNK	3.86E+07	0.00E+00	singlet	0	-50
SIOK(+1)NMYVTC(HRE(+29)VLTLSVK(+1)	5.99E+06	0	singlet	0	-50
SIOKNNVY(+1)TC(+1)14.1JHRE(+29)SIYAVK(+1)DGILLATGLHVH	5.15E+08	0.00E+00	singlet	0	-50
SIOKNNVY(+1)TC(+1)14.1JHRE(+29)AVLFPNPSK(+1)GLSNPAEVALRE	9.54E+08	9.78E-02	Doublet	8.15E-02	-3.616204144
SIOKNNVYTC(HRDK(+1)NINCINKVIRE(+29)VYASLEAYCK(+1)HKYPEQGR	3.57E+07	0.00E+00	singlet	0	-50
SISAK(+1)E(+25)AVLFPNPSK(+1)GLSNPAEVALREK	3.72E+08	3.33E-02	Doublet	3.00E+01	4.907822193
SISAK(+1)E(+29)AVLFPNPSK(+1)GLSNPAEVALREK	3.24E+07	3.33E-02	Doublet	3.00E+01	4.907822193
SISAK(+1)GAERE(+25)AVLFPNPSK(+1)	7.38E+07	2.15E+03	Doublet	4.66E-04	-11.06874557
SISAK(+1)GAERE(+25)AVLFPNPSK(+1)GLSNPAEVALRE	1.83E+11	8.68E-01	Doublet	1.15E+00	0.204221031
SISAK(+1)GAERE(+25)AVLFPNPSK(+1)GLSNPAEVALREK	3.68E+11	9.40E-01	Doublet	1.06E+00	0.088737227
SISAK(+1)GAERE(+25)AVLFPNPSKGLS(+1)NINPAEVALREK	1.87E+06	8.75E+04	Doublet	1.14E-05	-16.41732653
SISAK(+1)GAERE(+25)DCLIDK(+1)R	9.80E+07	3.93E-01	Doublet	2.55E+00	1.34923197
SISAK(+1)GAERE(+25)DMQMDK(+1)TELGCLR	3.73E+08	3.72E-01	Doublet	3.49E+02	0.574071047
SISAK(+1)GAERE(+25)DMQMDK(+1)TELGCLR	3.75E+08	3.40E-01	Doublet	1.690581459	-
SISAK(+1)GAERE(+25)DNK(+1)DK(+1)14.1JLIDK(+1)156.1JR	7.96E+09	4.83E-03	Doublet	2.07E+02	7.693207235
SISAK(+1)GAERE(+25)DNKDCIDK(+1)R	1.39E+08	1.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)FAK(+1)LLLR	3.20E+09	4.26E-01	Doublet	2.35E+00	1.229933876
SISAK(+1)GAERE(+25)GFKK(+1)R	2.36E+08	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)K(+1)DLTYTCR(+1)14.1JR	8.12E+08	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)K(+1)DLTYTCR	3.31E+08	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)SIYAVK(+1)DGILLATGLHVH	9.38E+08	1.04E+00	Doublet	1.82E+00	0.865709158
SISAK(+1)GAERE(+25)SIGLK(+1)CLEHFFK	9.47E+06	1.84E+01	Doublet	3.89E-02	-3.892160502
SISAK(+1)GAERE(+25)SIGLK(+1)CLEHFFK	2.90E+09	3.6E-01	Doublet	1.573195997	-
SISAK(+1)GAERE(+25)SSGK(+1)HYGVYSC(+1)14.1JEGC(+1)14.1JK	5.01E+08	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)VLTLSVK(+1)NMR	2.03E+09	4.09E-01	Doublet	2.45E+00	3.290337894
SISAK(+1)GAERE(+25)YQK(+1)C(+1)14.1JLAMGMK	2.14E+09	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+25)YQK(+1)C(+1)14.1JLAMGMK	6.05E+07	0.00E+00	singlet	0	-50
SISAK(+1)GAERE(+29)AVLFPNPSK(+1)GLSNPAEVALRE	1.58E+11	5.68E-01	Doublet	1.15E+01	0.204221031
SISAK(+1)GAERE(+29)AVLFPNPSK(+1)GLSNPAEVALREK	1.63E+11	9.70E-01	Doublet	1.03E+00	0.044083725
SISAK(+1)GAERE(+29)DCLIDK(+1)R	3.84E+07	3.93E-01	Doublet	2.55E+00	1.34923197
SISAK(+1)GAERE(+29)DMQMDK(+1)TELGCLR	3.16E+08	1.00E+00	singlet	0	-50
SISAK(+1)GAERE(+29)DNKDCIDK(+1)R	3.32E+08	1.0E+00	Doublet	0.141958559	-
SISAK(+1)GAERE(+29)FAK(+1)LLLR	1.37E+09	4.26E-01	Doublet	2.35E+00	1.229933876
SISAK(+1)GAERE(+29)SIGLK(+1)CLEHFFK	9.37E+08	1.84E+01	Doublet	3.89E-02	-3.892160502
SISAK(+1)GAERE(+29)SIGLK(+1)CLEHFFK	9.76E+08	3.3E-01	Doublet	9.71E-03	-6.686450115
SISAK(+1)GAERE(+29)VLTLSVK(+1)NMR	8.38E+08	4.09E-01	Doublet	2.45E+00	3.290337894

**Annexe 9 : Blast réalisé entre la séquence de CBP Humaine et la séquence de P300 Humain. En rouge sont noté les différences entre les deux séquences.**

Query	1	MAENLLD-GPPNPKRAKLSSPGFSAN--DSTDFGSLFDLENDLPDELIPNGGELGLLNSG	57
Sbjct	1	....VVEP...SA..P.....AL..SAS.G.....H.....-..ST....T.G.	59
Query	58	NL-----VPDAASKHKQLSELLRGGSGSSINPGIGN--VSASSPVQQGLGGQAQG	105
Sbjct	60	DINQLQTSLGM.Q.....S..SPNL.M.V.GPGQ.M.....A	108
Query	106	QPNSANMASLSAMGKSPLSQGDSSAPSLPKQAASTSGPTPAASQALNPQAQKQVGLATSS	165
Sbjct	109	.QS.PGLGLINS.V...MT.AGLTS.NM---GMG...NQGPT.S-----T.MMN.P	157
Query	166	PATSQTGPGICMNANFNQTHPGLLNSNSGHSLINQASQQAQVMNGSLGAAGRGRGAGMP	225
Sbjct	158	VNQPAM.MNTG...GM.---.M.AAGN.QGIMPN-----I.....-QN.Q	206
Query	226	YPTPAMQGASSSVLAETLTQVSPQMTGHAGLNTAQAGGMAKMGITGNTSPFGQPFSSQAGG	285
Sbjct	207	..N.G.-.SAGNL.T.P.Q.G....G.QT..RGP.P---L...MMN.PN.Y.S.YT.NP.	262
Query	286	QPMGATGVNPLASKQSMVNSLPTFPFDIKNTSVTNVPMNSQMOTSV----GIVP--TQA	339
Sbjct	263	.QI..S.LGL.IQT.TVLS.N.SP.AM.K.AVPGGGM...G.QPAPQVQQP.L.TPVA.G	322
Query	340	IATGP-TADPEKRKLIQQQLVLLLHAHKCQRREQANGEVRACSLPHCRTMKNVNLNHTHC	398
Sbjct	323	MGS.AH.....Q.N.....	382
Query	399	QAGKACQVAHCASSRQIISHWKNCTRHDPCVCLPLKNASDKRNQQTILGSPASGIQNTIG	458
Sbjct	383	.S..S.....G.....P..TGAPV.LG.P-S	441
Query	459	SVGTGQQNATSLSNPNPIDPSSMQRAYALGLPY-MNQPQTQLQPQVPGQQPAQP-QTHQ	516
Sbjct	442	.L.V...S.PN..TVSQ.....IE.....QV..MP..P.V.AKN..NQ..G.SP.	501
Query	517	QMRTLNLPLGNNPMNIPAGGITDDQPPNLISESALPTSLGATNPLMNDGNSNGNIGTLST	576
Sbjct	502	G...PMSNMSAS..GV-N..VGV--.T.S.L.D.M.HSAINSQ..M.---.ENASVPS.GP	555
Query	577	IPTAAPSSSTGVRKGWHEHVTQDLRSHLVHKLVAIFPTPDPAALKDRRMLNVLVAYAKKV	636
Sbjct	556	M....Q..T..I..Q...DI.....N.....R..	615
Query	637	EGDMYESANSRDEYYHLLAEKIYKIQKELEEKRRSRLHKQGILGNQPALPAPGAQP-PVI	695
Sbjct	616	.....N.A.....T..Q..NM.P.AAGMVPVSMN.G.NM	675
Query	696	PQAQPVRRPENGPLSLP-----VNRMQVSQGMNSFNPMSLGNVQLPQAPMGPRAA	744
Sbjct	676	G.P..GMTS...PD.SMIRGSVPNQMP.ITPQS.L.Q.GQ..MA-----P.IV..QT	730
Query	745	SPMNHVSQMNSMGSV-PGMAISPSRMPQPPNMMGAHTNNMMAQAPAQSQFLPQNQFPSSS	803
Sbjct	731	P.LQ.HG.LAQP.ALN.P.GYG.-.Q..S-----G.....T...QG	774
Query	804	-GAMSVGMGQPPAQTGVSSQGVPGAALP--NPLNMLGPQASQLPCPPVTQSPLH---PTP	857
Sbjct	775	MNVTNIPLAPSSG.AP...A.MSSSSC.VNS.IMPP.S.G.HIH..QLP.PA..QNS.S.	834
Query	858	PPASTAAGMPSLQHTTTPPGMTTPQPAAPTQPSTPVSSSGQTPTPTPGSVPSATQTQSTPT	917
Sbjct	835	V.SR.PT-----P.H...SIGAQ..P.T.I.APVTPPAMP-----PGP.S.ALHP	880
Query	918	VQAAAQAQVTPQPQTPVQPPSVATPQ-----SSQQQPTPVHAQPPGTPPL	961
Sbjct	881	PPRQTPTPP.T.LPQQ...SLP.A.SADQPQQQPRSQQSTAA.VPTPTA.LLPPQ.A...	940
Query	962	SQAAASIDNRVPTPSSVASAETNSQQPGPDVPVLEMKTETQAEDTEPDGKESKGEPRSEM	1021
Sbjct	941	..P.V..EGQ.SN.P.TS.T.V...AIAEQ.SQ.V.M.AKM.VDQ.E.ADTQP.DI..S	1000
Query	1022	MEEDLQGASQVKEETDIAEQKSEPMEVDEKKPEVKVEVKEEESSSNGTASQSTSPSQPR	1081
Sbjct	1001	KV..C-----M.ST.TE.RST.L.T.I...DQP.TSATQS.PA.G.SK	1045
Query	1082	KKIFKPEELRQALMPTLEALYRQDPESLPPFRQPVDPQLLGIPTYFDIVKNPMDLSTIKRK	1141
Sbjct	1046	.....S.....	1105
Query	1142	LDTGQYQEPWQYVDDVWLMFNNAWLYNRKTSRVYKFCSKLAEVFEQEIDPVMQSLGYCCG	1201
Sbjct	1106	.....I.....Y.....S.....	1165

Query	1202	RKYEFSPTLCCYKQLCTIPRDAAYYSYQNRHYHFCEKCFTEIQGENVTLGDDPSQPQT	1261
Sbjct	1166	..L.....T.....N.....S.S.....	1225
Query	1262	ISKDQFEKKNNDILDPEPFVDCKEGRKMHQICVLHYDIIWPSGFVDCNCLKKTGRPRKE	1321
Sbjct	1226	.N.E..S.R.....L.E.T.....HE...A....G....S.A.T... 1285	
Query	1322	NKFSAKRLQTTRLGNHLEDVRNKFLLRRQNHPEAGEVFVRVVASDDKTVEVKPGMKSRFVD	1381
Sbjct	1286	.....PS.....TP..N...D.....S...T....HA.....A..... 1345	
Query	1382	SGEMSESFYRTKALFAFEEIDGVDVCFVGMHVQYEGSDCPPNTRRVIISYLDLSIHFFR	1441
Sbjct	1346	...A.....L.....Q.....V.... 1405	
Query	1442	PRCLRTAVYHEILIGYLEYVKKLGYVTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQE	1501
Sbjct	1406	.K.....T..... 1465	
Query	1502	WYKMLDKAFAERI IHDYKDFIKQATEDRLTSAKELPYFEGDFWPNVLEESIKELEQEEE	1561
Sbjct	1466	.....VS...V..... 1525	
Query	1562	ERKKEESTAASETTEGSQGD SKNAKKNKKNKKNKSSISRANKKKPSMPNVSNDLSQKL	1621
Sbjct	1526	...R..N.-SN.S.DVTK.....S...L.G....G..... 1584	
Query	1622	YATMEKHKEVFFVIHLHAGPVINTLPPIVDPDPLLSCDLMDGRDAFLT LARDKHWFESSL	1681
Sbjct	1585	.....R.I...AA.S.....IP.....L..... 1644	
Query	1682	RRSKWSTLCMLVELHTQGQDRFVYTCNECKHHVETRWHTVCEYDLCINCYNKSHAHK	1741
Sbjct	1645	..AQ...M.....S.....T.....N.D.. 1704	
Query	1742	MVKWGLGLDDEGSSQGEFQSKSPQESRRLSIQRCIQSLVHACQCRNANCSLPSCQKMKRV	1801
Sbjct	1705	.E.L.....SNN.QAAATQ..GD..... 1764	
Query	1802	VQHTKGCKRKTNGGCPVCKQLIALCCYHAKHCQENKCPVPFCLNIKHKLRQQQIQHRLQQ	1861
Sbjct	1765	.....I.....Q.....L..... 1824	
Query	1862	AQLMRRMATMNRNVPQQLSPTSAPPPTPTQPSTPQTPQPPAQPSPVSMSPAGF	1921
Sbjct	1825	..ML.....S.QRTG.VG.QQGL.SPT.ATPT.PTGQQ.T...T.QPTSQPPTPPNSMP 1884	
Query	1922	PSVARTQPPTTVSTGKPTS-----QVPAPPPPAQPPAAVEAARQIEREAQQQHLR	1974
Sbjct	1885	.YLP...AAGP..Q..AAGQVTPPTPQT.Q..LPG.....M.M..Q.A.ET.RQMAH 1944	
Query	1975	VNINNSMPPGRITMGTPGSQMAPVSLNVPRPNQVSGFVMPSPMPGQ-----WQQAPLPQ	2028
Sbjct	1945	.Q.FQR--.IQHQ.PPMTPMAPMGMNPP.MTRGP..HLE.G.G.TGMQQQPP.S.GG... 2002	
Query	2029	-QQPMPGLPRPVISMQAQ-----AAVAGPRMPVQP--PRISPSALQDLLRTLKSPS	2078
Sbjct	2003	P..LQS.M...AMMSV..HGQPLNM.PQP.LGQVGIS.LK.GTV.QQ...N.....R... 2062	
Query	2079	SPQQQQVNLNLSNPQLMAAFIKQRTAKYVANQPGMQP-QPGL-QSQPGMQP-----QP	2131
Sbjct	2063	..L.....S..HA....L.....A...ANSN.QPI.G...MP.G...L..PTMPG.Q 2122	
Query	2132	GMHQQPQLNLNAMQAGVPRPG-----VPPQQQAMGGLNFPQGQALNIMNPGHNPNMASM	2185
Sbjct	2123	.V.SN.AM..M.P....Q.A.LPQQPQQQL.PP...MS..A.QM.M-----HNT. 2175	
Query	2186	NPQYREMLRRQLLQQQQQQQQQQQQQQQGSAGMAGGMAGHGQFQQPQGGYPPAMQ	2245
Sbjct	2176	PS.F.DI.....MM.....GAGP.IGP...N.N.....V-...QQ. 2224	
Query	2246	QQQRMQHLPLQGSSMGMAAQMGQLGQMGQPGLGADSTPNIQQALQQRILQQQQMKQOI	2305
Sbjct	2225	.RMQHMQ-----QMQQGNMG.I...P.----A...EAGASL-..Y...L.....M 2269	
Query	2306	GSPGQPNMSPQQHMLSGQPQASHLPGQQIATSLSNQVRSAPVQSPRPQSQPPHSSPSP	2365
Sbjct	2270	...V.....PN.A.SP..Q...PN.....Q..P..... 2329	
Query	2366	RIQPQSPHHVSPQTGSPHPGLAVTMASSIDQHLGNPEQSAMPLPQL	2412
Sbjct	2330	.M.....S.....VAAQ.NPME...FAS.D.NS...S... 2376	

**Annexe 10 : Blast réalisé entre la séquence de CBP Humaine et la séquence de CBP Mouse. En rouge sont noté les différences entre les deux séquences.**

Query	1	MAENLLDGPFPKRAKLSPPGFSANDSTDFGSLFDLENDLPDELIPNGGELGLLNSGNLV	60
Sbjct	1	..... <b>N</b> .....-.. <b>S</b> .....	59
Query	61	PDAASKHKQLSELLRGGSGSSINPGIGNVSASSPVQQGLGGQAQGPNSANMASLSAMGK	120
Sbjct	60	..... <b>T</b> ..... <b>G</b> .....	119
Query	121	SPLSQGDSSAPSLPKQAASTSGPTPAASQALNPQAQKQVGLATSSPATSQTGPICMNAN	180
Sbjct	120	.. <b>N</b> ..... <b>T.N</b> ..... <b>P</b> ..... <b>V</b> .....	179
Query	181	FNQTHPGLLNSNSGHSLINQASQQAQVMNGSLGAAGRGRGAGMPYPTPAMQGASSVLA	240
Sbjct	180	..... <b>M</b> .. <b>Q</b> ..... <b>A</b> ..... <b>T</b> .....	239
Query	241	ETLTQVSPQMTGHAGLNTAQAGGMAKMGITGNTSPFGQPFSSQAGGQPMGATGVNPLASK	300
Sbjct	240	..... <b>A</b> ..... <b>T</b> .. <b>M</b> .. <b>T</b> ..... <b>T</b> .. <b>Q</b> .....	299
Query	301	QSMVNSLPTFPTDIKNTSVINVPNMSQMQTSVGIVPTQAIATGPTADPEKRKLIQQQLVL	360
Sbjct	300	..... <b>A</b> ..... <b>T</b> ..... <b>L</b> .....	359
Query	361	LLHAHKCQRREQANGEVRACSLPHCRTMKNVNLNMTHTCQAGKACQVAHCASSRQIISHWK	420
Sbjct	360	.....	419
Query	421	NCTRHDPCVCLPLKNASDKRNQQTILGSPASGIQNTIGSVGTGQQNATSLSNPNPIDPSS	480
Sbjct	420	..... <b>A</b> .....	479
Query	481	MQRAYAALGLPYMNQPQTQLQPQVPGQQPAQPQTHQQMRTLNLPLGNNPMNIPAGGITTDQ	540
Sbjct	480	..... <b>PA</b> ..... <b>A</b> ..... <b>S</b> .....	539
Query	541	QPPNLISESALPTSLGATNPLMNDGNSNGNIGTLSTIPTAAPPSSSTGVRKGVHEHVTQDL	600
Sbjct	540	..... <b>S</b> .....	599
Query	601	RSHLVHKLVQAIFFPTPDPAALKDRRMLNVAAYAKKVEGDMYESANSRDEYYHLLAEKIYK	660
Sbjct	600	.....	659
Query	661	IQKELEEKRRLHKKQGILGNQPALPAPGAQPPVIPPQAQPVPPNGPLSLPVNRMQVSSQG	720
Sbjct	660	..... <b>S</b> ..... <b>P</b> .. <b>S</b> ..... <b>P</b> .....	719
Query	721	MNSFNPMSLGNVQLPQAPMGPRRAASPMNHSVQMNMSGVPGMAISPSRMPQPPNMMGAHT	780
Sbjct	720	..... <b>A</b> ..... <b>T.A</b> .....	779
Query	781	NNMMAQAPASQFLPQNQFPSSSGAMSV---GMGQPPAQTVGSQGVPGAALPNPLNMLG	837
Sbjct	780	.. <b>I</b> ..... <b>T.N</b> ..... <b>NSV</b> ..... <b>A</b> .. <b>A</b> ..... <b>A</b> .....	839
Query	838	PQASQLPCPPVTQSPLHPTPPPASTAAGMPSLQHTTTPPGMTPPQPAAPTQPSTPVSSSGQ	897
Sbjct	840	..... <b>P.A</b> ..... <b>-</b> .....	898
Query	898	TPTPTPGSVPSATQTQSTPTVQAAAQAQVTPQPQTPVQPPSVATPQSSQQQPTPVHAQPP	957
Sbjct	899	..... <b>A</b> ..... <b>T</b> .....	958
Query	958	GTPLSQAAASIDNRVPTPSSVASAETNSQQPGPDVPVLEMKTETQAEDTEPDGSKGEP	1017
Sbjct	959	..... <b>T</b> .. <b>S</b> ..... <b>M</b> ..... <b>V</b> .. <b>TD</b> .. <b>A</b> .. <b>E</b> .. <b>T</b> .....	1018
Query	1018	RSEMMEEDLQASQVKEETDIAEQKSEPMVEDEKKPEVKVEVEKEEESSNGTASQSTSP	1077
Sbjct	1019	..... <b>S</b> ..... <b>TT</b> ..... <b>E</b> ..... <b>A</b> ..... <b>N</b> .. <b>D</b> .....	1078
Query	1078	SQPRKKIFKPEELRQALMPTLEALYRQDPESLPPFRQPVDPQLLGIPIFYFDIVKNPMDLST	1137
Sbjct	1079	.....	1138
Query	1138	IKRKLDTGQYQEPWQYVDDVWLMFNNAWLYNRKTSRVYKFCSKLAEVFEQEIDPVMQSLG	1197
Sbjct	1139	.....	1198
Query	1198	YCCGRKYEFSPTLCCYGKQLCTIPRDAAYYSYQNRHYHFCEKCFTEIQGENVTLGDDPSQ	1257
Sbjct	1199	.....	1258

Query	1258	PQTTISKDQFEKKKNDTLDPEPFVDCKECGRKMHQICVLHYDIIWPSGFVCDNCLKKTGR	1317
Sbjct	1259	.....	1318
Query	1318	PRKENKFSAKRLQTTRLGNHLEDRVNKFLRRQNHPEAGEVFVRVVASDDKTVKPGMKS	1377
Sbjct	1319	.....	1378
Query	1378	RFVDSGEMSESFYRTKALFAFEEIDGVDVCFGMHVQEYGSDCPPNTRRVYISYLDISI	1437
Sbjct	1379	.....NTALIA.HQIQG.....	1438
Query	1438	HFFRPRCLRTAVYHEILIGYLEYVKKLGYVTGHIWACPPSEGDDYIFHCHPPDQKIPKPK	1497
Sbjct	1439	.....	1498
Query	1498	RLQEWYKMLDKAFAERIIHDYKDIKQATEDRLTSAKELPYFEGDFWPNVLEESIKELE	1557
Sbjct	1499	.....N.....N.....	1558
Query	1558	QEEEEERKKEESTAASETTEGSQGDSKNAKKKNNKKNKSSISRANKKKPSMPNVSNDL	1617
Sbjct	1559	.....P.....	1618
Query	1618	SQKLYATMEKHKEVFFVIHLHAGPVINTLPPIVDPDPLLSCDLMDGRDAFLTLARDKHWE	1677
Sbjct	1619	.....S.Q.....	1678
Query	1678	FSSLRRSKWSTLCMLVELHTQGQDRFVYTCNECKHHVETRWHCTVCEYDLCINCYNTKS	1737
Sbjct	1679	.....	1738
Query	1738	HAHKMVKWGLGLDDEGSSQGEPSKSPQESRRLSIQRCIQSLVHACQCRNANCSLPSCQK	1797
Sbjct	1739	.T.....	1798
Query	1798	MKRVVQHTKGCKRKTNGGCPVCKQLIALCCYHAKHCQENKCPVPFCLNIKHKLRQQQIQH	1857
Sbjct	1799	.....	1858
Query	1858	RLQQAQLMRRRMATMNRNVPQQSLPSPTSAPPGTPTQQPSTPQTPQPPAQPPSPVMSMS	1917
Sbjct	1859	.....N..	1918
Query	1918	PAGFPSVARTQPPTTVSTGKPTSQVPAPPPPAQPPAAVEAARQIEREAQQQQHLYRVNI	1977
Sbjct	1919	.....N.....I.A.....N.....A..	1978
Query	1978	NNSMPPGRTGMGTGPSQMAPVSLNVRPNQVSGPVMPSMPPGQWQQAPLPQQQPMPGLPR	2037
Sbjct	1979	..G.....A.....T..G.....S.....I.....M..	2038
Query	2038	FVISMQAQAAVAGPRMPSVQPPRSISPSALQDLLRTLKSPSSPQQQQVNLILKSNPQLM	2097
Sbjct	2039	..M.....N.....	2098
Query	2098	AAFIKQRTAKYVANQPGMQPQGLQSQPGMQPQGMHQPSLQNLNAMQAGVPRPGVPPQ	2157
Sbjct	2099	.....P	2158
Query	2158	QQAMGGLNPQGQALNIMNPGHNPNMASMNPPYREMLRRQLLQQQQQQQQQQQQQQQQG	2217
Sbjct	2159	.P.....TN.....V...-L.H.....N	2217
Query	2218	SAGMAGGMAGHGQFQQPQGGYPPAMQQQQRMQHLPQGS SMGQMAAQMGQLGQMGQP	2277
Sbjct	2218	..SL.....S.....A...-.....I.....P.....	2276
Query	2278	GLGADSTPNIQQALQQRILQQQQMKQQIGSPGQPNPMSPPQQHMLSGQPQASHLPGQQIAT	2337
Sbjct	2277	.....	2336
Query	2338	SLSNQVRSFAPVQSPRPQSQPPHSSPSPRIQPQSPHHVSPQTGSPHPGLAVTMASSIDQ	2397
Sbjct	2337	.....M..	2396
Query	2398	GHLGNPEQSAMPLPQLNTPRSALSSELVGDITGDTLEKFVEGL	2442
Sbjct	2397	.....N.....	2441