



HAL
open science

Développements en spectrométrie de masse pour l'étude des complexes biologiques

Nha Thi Nguyen Huynh

► **To cite this version:**

Nha Thi Nguyen Huynh. Développements en spectrométrie de masse pour l'étude des complexes biologiques. Chimie analytique. Université de Strasbourg, 2015. Français. NNT : 2015STRAF045 . tel-01291503

HAL Id: tel-01291503

<https://theses.hal.science/tel-01291503>

Submitted on 21 Mar 2016

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

THÈSE présentée par :

Nha Thi NGUYEN HUYNH

soutenue le : **12 octobre 2015**

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

Discipline/Spécialité : Chimie/Chimie analytique

**Développements en spectrométrie de
masse pour l'étude des complexes
biologiques**

THÈSE dirigée par :

Dr. LEIZE-WAGNER Emmanuelle Directrice de recherche, Université de Strasbourg

RAPPORTEURS :

Pr. LAPREVOTE Olivier

Rapporteur externe, Université Paris Descartes

Dr. REDEKER Virginie

Rapporteur externe, Institut des Neurosciences Paris-Saclay

AUTRE MEMBRE DU JURY :

Pr. DEJAEGERE Annick

Présidente du jury, Université de Strasbourg

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES
UMR 7140

THÈSE présentée par :
Nha Thi NGUYEN HUYNH

soutenue le : **12 octobre 2015**

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

Discipline/Spécialité : Chimie/Chimie analytique

**Développements en spectrométrie de
masse pour l'étude des complexes
biologiques**

THÈSE dirigée par :

Dr. LEIZE-WAGNER Emmanuelle Directrice de recherche, Université de Strasbourg

RAPPORTEURS :

Pr. LAPREVOTE Olivier

Rapporteur externe, Université Paris Descartes

Dr. REDEKER Virginie

Rapporteur externe, Institut des Neurosciences Paris-Saclay

AUTRE MEMBRE DU JURY :

Pr. DEJAEGERE Annick

Présidente du jury, Université de Strasbourg

„Wer kämpft, kann verlieren. Wer nicht kämpft, hat schon verloren.“

« Celui qui combat peut perdre. Celui qui ne combat pas a déjà perdu. »

(Bertolt Brecht)

Remerciements

Je souhaite remercier tout d'abord Pr. Annick Dejaegere, Pr. Olivier Lapr votte et Dr. Virginie Redeker d'avoir accept  d' valuer des travaux que j'ai pu r aliser pendant ma th se.

Ce travail de th se a  t  r alis  au Laboratoire de Spectrom trie de Masse des Interactions et des Syst mes (LSMIS), sous la direction du Dr. Emmanuelle Leize-Wagner. J'adresse vivement mes remerciements   Emmanuelle. Merci de m'avoir fait confiance et soutenue, et de m'avoir guid e pendant ma th se. Je souhaite remercier sinc rement Dr. No lle Potier pour m'avoir encadr  tout au long de ces trois ann es. Merci de m'avoir fait d couvrir la spectrom trie de masse supramol culaire, des connaissances qu'elle m'a transmises, de ses pr cieux conseils et de sa gentillesse.

Je tiens  galement   remercier les collaborateurs   l'IGBMC : Dr. Val rie Lamour, Dr. Patrick Schultz, Dr. Grigory Sharov, Dr. Natacha Rochel, Dr. Judit Osz, Dr. Anna Belorusova, Carole Peluso-Iltis... ainsi que ceux de SYBIO : Pr. Philippe Compain, Dr. Anne Bodlenner et J r my Schneider, avec qui j'ai eu l'occasion de travailler sur diff rents projets de recherche. Merci de votre patience, m me face   mes nombreuses interrogations ou lors de mes nombreuses demandes en  chantillon ! Les collaborations avec vous ne m'ont pas seulement permis de mener   bien ces travaux de th se, mais aussi d' largir mes horizons en biologie. Merci  galement   P lagie Fichter et Cl ment Potel qui ont contribu  pendant leur stage au succ s de ces projets, ainsi que les membres de la plateforme prot omique de l'IGBMC.

Je tiens  videmment   remercier tous les membres du laboratoire. Merci   H l ne pour son aide durant ma th se et ses conseils avis s sur la science comme sur la vie. En particulier, je remercie l' quipe des filles du laboratoire : Marianne pour ses conseils en Excel et ses tr s bons plats libanais, C cile ma coll gue d'ann e de th se pour son fameux cake sal  et son aide m me en tant que « docteur junior », et Margaux pour ses desserts fabuleux (entre autres). Bien s r je n'oublie pas Micha l et sa bonne humeur, Rabah et ses pauses caf , Yannis le « faux permanent » ou « faux jeune » (au choix), Maxime et Aur lien les geeks. Mention particuli re   C cile et Margaux pour la relecture de mon manuscrit ;   Yannis et Rabah qui viennent toujours acheter de quoi d jeuner avec moi (avoue Yannis je vais te manquer pour cela !). Merci  galement   Armelle, St phanie, Nathalie, Nassur, Soumia, M lanie. Merci aux anciens stagiaires du laboratoire et leur souffle de jeunesse. Et le douzi me  tage de la tour de chimie c'est aussi nos « colocataires » de la chimie quantique que j'ai c toy s pendant ces trois ans.

Je souhaite remercier les membres de la plateforme prot omique de l'IBMC : Philippe Hammann, Lauriane Kuhn et Johana Chicher. Merci d'avoir r pondu   mes interrogations ou lors de mes demandes au secours concernant les bases de donn es, les serveurs et les logiciels bio-informatiques. Je garderai en souvenir votre gentillesse et votre bonne humeur dans la vie de tous les jours.

Je tiens  galement   remercier de nombreux professeurs, ma tres de conf rences et chercheurs avec qui j'ai eu le privil ge de travailler pendant mes trois ann es de monitorat. Merci au Dr. Marco Cecchini de m'avoir accord  son temps pour discuter non seulement des travaux dirig s mais aussi de mes probl matiques scientifiques. Merci au Dr. Aur lie Guenet de m'avoir guid e en

tant que tutrice du monitorat, ainsi que de ses précieux conseils pour mener à bien des séances de travaux dirigés.

Merci aux différents doctorants que j'ai rencontrés pendant les congrès, formations ou autres. Des échanges avec vous (qui ne sont pas seulement scientifiques) m'ont permis de voir ce qui se fait dans divers domaines, des sciences expérimentales aux sciences humaines en passant par la psychologie ou même le sport. Et il ne faut pas le dire trop fort : je dois avouer qu'on aime bien se plaindre de nos thèses respectives !

Merci aux différentes personnes dont j'ai pu faire connaissance et avec qui j'ai eu l'occasion de discuter et croiser tout au long de ces trois années.

J'adresse également mes remerciements au Pr. Lysiane Richert. Le hasard a fait que j'ai eu l'opportunité de croiser votre chemin et j'en suis très contente. Sans vous, vos encouragements et des échanges sur la thèse, je pense que cette aventure n'aurait pas été possible.

Mes pensées vont aussi à Daniel Laulanné et à sa famille. Vous êtes des personnes formidables que j'ai eu la chance de connaître. Je suis ravie que nous sommes toujours restés en contact même si nous avons peu d'occasion de nous revoir. Comme vous le savez déjà, pour moi vous faites partie de la famille.

Enfin, je remercie mes proches de leur soutien. Merci à mes amis pour les moments passés ensemble. Merci à mes parents et à ma sœur qui m'ont toujours accordé la liberté d'action, ont toujours cru en moi et m'ont encouragée non seulement pendant ces trois années de thèse mais aussi tout au long de mes études. Et surtout merci à Florent qui m'a toujours soutenue « dans les moments difficiles de la vie ». Merci de sa patience, de son écoute, de nombreuses relectures et d'autres aides informatiques surtout lors de la rédaction de ce manuscrit. Promis je te rendrai la pareille quand ce sera ton tour.

Sommaire

Liste des abréviations et symboles.....	9
Introduction générale	11
Partie I : Introduction bibliographique	15
Chapitre 1 : Spectrométrie de masse et complexes biologiques non-covalents : état de l'art	17
1.1. Importance de l'étude des complexes biologiques non-covalents.....	19
1.2. Avantages et inconvénients des techniques d'analyse structurale conventionnelles.....	20
1.3. Apport de la spectrométrie de masse supramoléculaire « classique » : ionisation électrospray (ESI-MS).....	22
1.3.1. Instrumentation.....	22
1.3.2. Paramètres expérimentaux et instrumentaux spécifiques à l'étude des complexes non-covalents.....	25
1.3.2.1. Préparation spécifique de l'échantillon pour la préservation des interactions non-covalentes et compatible avec l'analyse en ESI-MS	25
1.3.2.2. Paramètres instrumentaux spécifiques à l'étude des complexes non-covalents.....	27
1.3.2.3. Les sources nano électrospray.....	29
1.3.3. Informations apportées par la spectrométrie de masse dans l'étude des complexes biologiques non-covalents.....	29
1.3.3.1. La stoechiométrie.....	29
1.3.3.2. La spécificité.....	30
1.3.3.3. La sélectivité.....	30
1.3.3.4. La constante d'affinité	30
1.4. Quelques approches complémentaires	31
1.4.1. La mobilité ionique couplée à la MS.....	31
1.4.2. L'échange hydrogène/deutérium (H/D) couplé à la MS.....	32
1.4.3. Le marquage chimique couplé à la MS	33
1.4.4. La protéolyse ménagée couplée à la MS	34
1.4.5. Le pontage chimique couplé à la MS	35

Chapitre 2 : Stratégie complémentaire voire alternative à l'ESI-MS pour l'étude des complexes biologiques : pontage chimique couplé à la MS 43

2.1. La spectrométrie de masse par désorption-ionisation laser assistée par matrice (MALDI-MS).....	45
2.1.1. Instrumentation.....	45
2.1.2. Difficultés d'étude des composés de haute masse moléculaire	47
2.1.2.1. La détection des ions à haut rapport m/z	48
2.1.2.2. La détection des complexes non-covalents par MALDI-MS.....	49
2.2. Application du couplage pontage chimique - spectrométrie de masse à l'étude des édifices non-covalents.....	51
2.2.1. Principe du pontage chimique couplé à la spectrométrie de masse (CX-MS).....	51
2.2.2. Mise en œuvre de la réaction de pontage	52
2.2.2.1. Les critères de choix d'agent de pontage.....	52
2.2.2.2. Les agents de pontage réagissant sur les amines.....	55
2.2.2.3. Les agents de pontage réagissant sur les thiols.....	57
2.2.2.4. Les agents de pontage photoréactifs (non-sélectifs).....	58
2.2.2.5. Les longueurs du bras espaceur	58
2.2.2.6. Optimisation de la réaction de pontage	59
2.2.3. Analyse des produits de pontage par MS.....	60
2.2.3.1. Production des peptides après pontage.....	60
2.2.3.2. Les types de peptide obtenus	60
2.2.3.3. Amélioration de la détection en MS des peptides pontés.....	62
2.2.4. Traitement des données obtenues par les logiciels dédiés au pontage.....	68
2.2.5. Informations apportées par le pontage chimique couplé à la MS.....	70
2.2.5.1. La stœchiométrie du complexe.....	70
2.2.5.2. Les zones de proximité entre les sous-unités.....	71
2.2.6. Contraintes et limites	72
2.3. Conclusion de l'étude bibliographique	74

Partie II : Résultats et discussion 83

Chapitre 3 : Etude par CX-MS de la stœchiométrie et de la topologie du complexe non-covalent SAGA HAT..... 85

3.1. Contexte de l'étude : le complexe SAGA HAT	87
3.2. Etudes préliminaires avec la BSA.....	88
3.3. Mise au point de la réaction de pontage.....	91
3.3.1. Choix de l'agent de pontage	91
3.3.1.1. Le glutaraldéhyde.....	91

3.3.1.2. Le bis[sulfosuccinimidyl] suberate (BS3)	93
3.3.2. Optimisation des conditions expérimentales.....	96
3.4. Choix et utilisation des logiciels d'identification	99
3.4.1. Logiciel pLink.....	100
3.4.2. Logiciel xQuest	101
3.5. Contraintes, avantages et limites	104
3.6. Publication : Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex.....	107
3.7. Perspective : application à l'étude des complexes protéiques membranaires ?	134
3.8. Conclusion	136
Chapitre 4 : Suivi par ESI-MS de la formation des complexes de RAR-RXR avec différents ADNs	139
4.1. Contexte de l'étude : les complexes RAR-RXR-ADN	141
4.2. Mise au point de l'étude des complexes RAR-RXR-ADN par ESI-MS.....	142
4.2.1. Difficultés de l'étude des complexes de type protéine-ADN	142
4.2.2. Optimisation des paramètres instrumentaux	144
4.3. Stratégie de l'étude des complexes RAR-RXR-ADN	146
4.4. Publication : Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry.....	148
4.5. Conclusion	159
Chapitre 5 : Mesure par ESI-MS des constantes de dissociation des complexes RXR-ligand.....	161
5.1. Contexte de l'étude : les complexes RXR-ligand.....	163
5.2. Modèles mathématiques du calcul des constantes de dissociation	164
5.3. Difficultés et mise au point des expériences de titrage de RXR-ligand suivies par ESI-MS.....	171
5.4. Mise au point et difficultés du calcul des constantes de dissociation pour trois ligands utilisés : 9-cis-RA, BMS 649 et R-DHRA.....	173
5.4.1. Problème de conformation de la protéine RXR ?	175
5.4.2. Problème de solubilité du ligand ?	175
5.5. Conclusion	177
Conclusion générale et perspectives	179
Annexes	183

Liste des abréviations et symboles

[...]	Concentration
°C	Degré Celsius
Δ	Delta, différence entre deux valeurs
ACN	Acétonitrile
AcONH ₄	Acétate d'ammonium
ADN	Acide désoxyribonucléique
BS3	Bis[sulfosuccinimidyl] suberate
BSA	Albumine sérique bovine
CX-MS	Pontage chimique couplé à la spectrométrie de masse
Da	Dalton
ESI	Ionisation par électrospray
HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
Kd	Constante de dissociation
LC	Chromatographie liquide
M	Mole par litre
MALDI	Désorption-ionisation laser assistée par matrice
μ L	Microlitre
μ s	Microseconde
MS	Spectrométrie de masse
MS/MS	Spectrométrie de masse en tandem
<i>m/z</i>	Rapport masse/charge
nanoLC	Nano chromatographie liquide
ppm	Partie par million
Q	Quadripôle
RAR	Récepteur de l'acide rétinoïque
RARE	Elément de réponse à l'acide rétinoïque
RXR	Récepteur X des rétinoïdes
SA	Acide sinapinique
SAGA HAT	Spt-Ada-Gcn5 acétyltransférase, module histone acétyl transférase
SDS-PAGE	Electrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium
TFA	Acide trifluoroacétique
TOF	Analyseur à temps de vol
Tris	Trishydroxyméthylaminométhane
V	Volt

Introduction générale

Depuis la publication du premier spectre de masse démontrant la détection d'un complexe protéine-hème en 1991 [1], il a été montré que la spectrométrie de masse (MS) pouvait constituer un outil de choix pour étudier les complexes biologiques non-covalents, et intégrer ainsi l'ensemble des méthodes utilisées pour les études d'analyse structurale (diffraction par rayon X, résonance magnétique nucléaire, microscopie électronique...). En particulier, la MS à ionisation électrospray (ESI-MS) a montré de grandes capacités à déterminer l'existence d'un complexe et de leurs stœchiométries, ou encore à analyser la dynamique des interactions [2]. Cependant la difficulté à transférer intacts certains complexes protéiques dans les tampons classiquement utilisés en ESI-MS nous incite aujourd'hui à développer de nouveaux protocoles d'analyse pour permettre leur étude par MS.

Une des approches complémentaires voire alternatives est d'utiliser la désorption-ionisation laser assistée par matrice (MALDI-MS) grâce à sa plus grande tolérance vis-à-vis des sels non volatils. Cependant, même si le MALDI-MS permet l'analyse des complexes non-covalents, il requiert d'opérer dans des conditions particulières de pH, de solvant, de matrice et du tir laser, qui sont pour l'instant limitées à quelques exemples de systèmes supramoléculaires. Dans ce contexte, plusieurs stratégies visant à l'obtention d'informations structurales ont été développées. Parmi celles-ci, la stratégie du pontage chimique mise en place en amont de l'analyse en MS émerge comme une alternative intéressante pour étudier les édifices non-covalents [3]. A mon arrivée au laboratoire, quelques études suggéraient la possibilité de caractériser des systèmes pontés par MS. Mais de nombreuses difficultés ont été évoquées, telles qu'un faible rendement de réaction de pontage et un manque cruel d'outil bio-informatique pour l'interprétation. C'est dans ce contexte qu'ont démarré mes travaux de thèse.

Ainsi, l'objectif de ces travaux de thèse est d'approfondir le développement de la MS supramoléculaire pour l'étude de complexes biologiques non-covalents. Nous nous sommes attachés à mettre au point des méthodologies afin de repousser les limites de la technique et d'obtenir des informations de plus en plus détaillées sur la structure et la dynamique des complexes. L'accent a été mis sur les cas où l'utilisation de l'ESI-MS supramoléculaire « classique » était limitée : des systèmes peu solubles dans les tampons d'ammonium, ainsi que des complexes mettant en jeu des oligonucléotides difficiles à dessaler de façon appropriée.

Dans la première partie de ce manuscrit consacrée à l'étude bibliographique, nous rappellerons au sein du chapitre 1 l'importance de l'étude des complexes biologiques non-covalents ainsi que quelques méthodes conventionnelles dédiées à leur analyse structurale. Puis nous mentionnerons l'état de l'art dans le domaine de la MS supramoléculaire et son apport à l'étude des édifices macromoléculaires. Enfin, nous discuterons de quelques pistes alternatives à la MS supramoléculaire « classique » permettant d'apporter des informations complémentaires. Dans le chapitre 2, nous nous intéresserons particulièrement au pontage chimique couplé à la MS, son principe, sa mise en œuvre et le traitement des données par des outils bio-informatiques.

La seconde partie de ce manuscrit est réservée aux résultats expérimentaux obtenus, leur interprétation et la discussion. Le développement de la MS a été réalisé sur différents systèmes biologiques ; les résultats sur trois systèmes en particulier sont présentés dans ce manuscrit :

- Chapitre 3 : Développement d'une approche de pontage chimique couplé à la MS (CX-MS) pour l'étude du complexe non-covalent SAGA HAT (Spt-Ada-Gcn5 Acétyltransferase, module Histone Acétyl Transferase) humain. Nous avons mis en place un protocole robuste de CX-MS afin de déterminer la stœchiométrie et d'aborder l'analyse topologique de ce complexe.
- Chapitre 4 : Développement d'une approche par ESI-MS pour étudier les complexes formés par RAR-RXR (récepteur de l'acide rétinoïque - récepteur X des rétinoïdes) avec différents ADN. L'ESI-MS a été utilisé pour détecter différents états d'oligomérisation présents en solution et suivre leur formation en fonction de la quantité et de la nature de l'ADN.
- Chapitre 5 : Développement d'une approche de mesure par ESI-MS de la constante de dissociation pour l'étude des complexes RXR-ligand. Notre objectif était de déterminer les constantes de dissociation des complexes de RXR avec trois ligands différents par des expériences de titrage.

Bibliographie

- [1] Katta, V., Chait, B.T., Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* 1991, 113, 8534–8535.
- [2] Hilton, G.R., Benesch, J.L.P., Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. R. Soc. Interface* 2012, 9, 801–816.
- [3] Merkley, E.D., Cort, J.R., Adkins, J.N., Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J. Struct. Funct. Genomics* 2013, 14, 77–90.

Partie I

Introduction bibliographique

Dans cette première partie, nous présenterons l'état d'avancement des connaissances actuelles dans l'étude des complexes biologiques non-covalents, particulièrement dans le domaine de la spectrométrie de masse.

L'introduction bibliographique se divise en deux chapitres :

- **Chapitre 1** : Spectrométrie de masse et complexes biologiques non-covalents : état de l'art
- **Chapitre 2** : Stratégie complémentaire voire alternative à l'ESI-MS pour l'étude des complexes biologiques : pontage chimique couplé à la MS

Chapitre 1 : Spectrométrie de masse et complexes biologiques non-covalents : état de l'art

Le fonctionnement des systèmes biologiques met en jeu des mécanismes de reconnaissance moléculaire et d'assemblage à travers des interactions non-covalentes. Outre le fait d'assurer l'intégrité structurale des biomolécules, ces interactions réversibles jouent un rôle important dans les processus cellulaires comme la division des cellules, la traduction et la transcription des gènes, ou encore la régulation des processus biologiques majeurs.

Dans ce chapitre, nous aborderons l'importance de l'étude des complexes biologiques non-covalents. Puis nous mentionnerons quelques méthodes conventionnelles pour étudier ces édifices, leurs avantages et inconvénients. Ensuite nous verrons le fonctionnement de la spectrométrie de masse (MS) supramoléculaire, les paramètres expérimentaux et instrumentaux spécifiques, et l'apport de la MS à l'étude de ces complexes. Enfin, nous discuterons de quelques pistes alternatives permettant d'apporter des informations complémentaires dans l'étude structurale des complexes non-covalents.

1.1. Importance de l'étude des complexes biologiques non-covalents

Dans les années 1940 s'est développé le paradigme du « tout génétique » (selon les termes de Henri Atlan [1]) ou de la biologie moléculaire, qui stipule la relation « un gène-une protéine-une fonction », c'est-à-dire qu'un gène code pour une protéine qui exerce une fonction [2]. Ceci implique que la connaissance du code génétique est suffisante en elle-même pour connaître entièrement les protéines et de ce fait, grâce au lien linéaire entre chaque protéine et sa fonction, la nature des organismes. Des efforts importants ont donc été réalisés pour déchiffrer le programme génétique des organismes vivants, on peut en citer par exemple le Projet Génome Humain [3].

Puis cette idée selon laquelle on peut tout comprendre et tout expliquer à propos des organismes vivants en se reposant uniquement sur le code génétique a été progressivement remplacée par une perspective plus complexe. Dans les années 1990, avec les efforts pour cartographier les réseaux d'interaction au niveau du protéome [4,5] émerge le paradigme de la biologie des réseaux, biologie des systèmes ou biologie intégrative [6,7]. Cette approche stipule que ce n'est pas une protéine mais plutôt des complexes de biomolécules, voire le réseau de ces complexes, qui font office d'intermédiaire entre le génotype et le phénotype. La structure et la topologie de ces réseaux, modulées par des influences internes (modifications génomiques) ou externes (milieu environnemental), déterminent le phénotype [8]. Dans ce sens, toute perturbation dans le fonctionnement des complexes, voire des réseaux de complexes, peut entraîner des perturbations dans le phénotype (Figure 1).

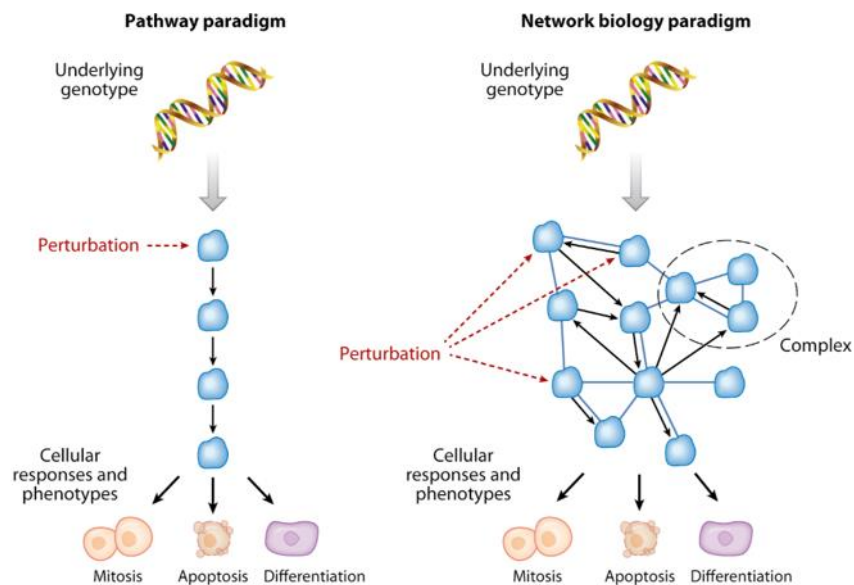


Figure 1. Les paradigmes de la biologie « linéaire » et de la biologie des réseaux [8].

Outre un intérêt de compréhension scientifique, les études sur les complexes biologiques et les réseaux moléculaires ont permis une meilleure appréhension des maladies. On peut mentionner

ainsi les recherches sur des complexes biologiques de la maladie d'Alzheimer [9], de la mucoviscidose [10] ou les infections virales [11,12] comme pour les virus de l'hépatite C [13], les virus de la grippe H1N1 [14], ou encore dans le cas du HIV [15,16]. L'implication des réseaux moléculaires dans le développement du cancer [17] et sa compréhension permettent de postuler la réponse des cellules cancéreuses aux médicaments et donc le développement pharmaceutique [18].

Le plus souvent, des sous-unités des complexes biologiques sont en interaction les unes avec les autres par des interactions non-covalentes réversibles, ce qui permet la régulation des processus biologiques. Ces interactions peuvent impliquer des liaisons électrostatiques, des liaisons de van der Waals, des liaisons hydrogène et l'effet de groupe d'hydrophobicité (en présence des molécules d'eau environnantes). Elles assurent non seulement la structure tertiaire et quaternaire de grandes biomolécules comme les ADN, ARN, protéines, carbohydrates *etc.* mais aussi l'intégrité structurale des complexes de type protéine-protéine, protéine-ligand, protéine-ADN, ADN-ADN, et ADN ou ARN-ligand [19].

En raison de l'importance de ces édifices biologiques non-covalents, des efforts considérables ont été déployés en termes de développements techniques, de méthodologie, d'approches et d'outils d'analyse.

1.2. Avantages et inconvénients des techniques d'analyse structurale conventionnelles

De nombreuses approches permettent d'aborder l'étude des complexes biologiques. On peut citer notamment :

- Les méthodes spectroscopiques (UV, fluorescence)
- La calorimétrie par titration isotherme (ITC)
- La diffraction des rayons X (DRX)
- La résonance magnétique nucléaire (RMN)
- Le dichroïsme circulaire (CD)
- La résonance plasmonique de surface (SPR)
- La microscopie électronique (EM)

Les méthodes spectroscopiques (UV, fluorescence) mettent en évidence les interactions moléculaires entraînant une modification de l'absorbance de la molécule d'intérêt [20]. Cependant, elles requièrent la présence de chromophore dans au moins l'une des sous-unités du complexe, et dont l'absorbance doit être affectée suite à l'interaction.

L'ITC permet d'obtenir les paramètres thermodynamiques des interactions, comme les variations d'enthalpie libre de liaison (ΔH), d'énergie libre (ΔG), d'entropie (ΔS) ; la constante de dissociation K_d , le nombre de sites de liaison et la stœchiométrie [21]. Cependant, les expériences d'ITC demandent une quantité non négligeable d'échantillon : de l'ordre de quelques milligrammes à

quelques dizaines de milligrammes de protéine. De plus, elles nécessitent que la variation de chaleur liée à l'interaction soit suffisamment grande pour être détectée. Enfin, l'analyse par ITC devient difficile en cas d'interactions complexes impliquant des coopérativités de liaison.

La DRX donne des informations sur la structure tridimensionnelle des composés biologiques sans limite de taille, et peut localiser de manière précise les atomes. Les principales difficultés sont la grande consommation de matériel et l'obtention de cristaux [22]. De la qualité cristalline de ces derniers dépend la qualité de l'interprétation des résultats. En outre, la DRX requiert un ordonnancement (atomique, moléculaire ou macromoléculaire) et il est possible que l'empilement des cristaux ne reflète pas la conformation majeure en solution.

La RMN permet d'avoir des informations structurales proches de l'échelle atomique des complexes protéiques. Cependant, plus la taille des macromolécules biologiques augmentent, plus on se heurte à deux difficultés majeures. D'une part, les signaux s'élargissent en raison d'une plus grande relaxation dipolaire ; d'autre part, ils peuvent se chevaucher menant à plus de complexité dans les spectres à interpréter [23]. Par conséquent, l'étude par RMN des complexes de plus de 100 kDa devient difficile.

Le CD permet de caractériser les structures des composés asymétriques, qu'il s'agisse de petites molécules ou d'édifices supramoléculaires ; et permet de suivre des changements dans la conformation des biomolécules en fonction de l'environnement (pH, température...) [24]. Cependant, contrairement à la DRX et la RMN qui sont capables de donner des informations structurales à l'échelle atomique, les informations obtenues par le CD sont de plus faible résolution, même si l'analyse en CD est plus rapide et consomme moins d'échantillon [25]. De plus, les tampons aqueux couramment utilisés (comme les tampons acétate, sulfate ou Tris) peuvent absorber la lumière et donc interférer avec l'échantillon.

La SPR permet de détecter les interactions moléculaires et donne l'accès à la constante de dissociation K_d ainsi qu'aux constantes cinétiques d'association k_{on} et de dissociation k_{off} . Néanmoins, l'immobilisation de l'échantillon sur la surface recouverte de métal durant l'analyse peut altérer les propriétés et la flexibilité conformationnelle donc la fonction des composés biologiques [26]. Les principales difficultés sont la mise en œuvre de l'analyse (il faut entre autres avoir une couche homogène de l'analyte) ainsi que les phénomènes d'adsorption non spécifique des molécules sur la surface.

L'EM est largement utilisée en biologie, notamment en parallèle à la DRX, pour l'étude structurale des complexes avec une résolution de l'ordre de 10 Å. La difficulté majeure consiste en l'optimisation de l'intensité du faisceau d'électron afin de ne pas dégrader l'échantillon biologique, tout en limitant le bruit pour avoir une bonne précision et un contraste suffisant [27]. De plus, des complexes hétérogènes ou dynamiques peuvent limiter la résolution obtenue.

Toutes ces méthodes peuvent être utilisées en parallèle et seront choisies en fonction des critères comme la quantité de matériel disponible ; l'homogénéité, l'affinité et la taille du complexe d'intérêt ; et les questions biologiques posées.

Dans les années 1990, avec le développement des ionisations ESI (ionisation électrospray – « electrospray ionisation ») et MALDI (la désorption-ionisation laser assistée par matrice – « matrix assisted laser desorption-ionisation »), la spectrométrie de masse (MS) s'ouvre au monde de la biologie [28,29]. Elle apparaît aujourd'hui comme une méthode complémentaire aux analyses

structurales classiques, et s'intègre désormais parmi les techniques de choix pour la caractérisation structurale des complexes biologiques non-covalents.

1.3. Apport de la spectrométrie de masse supramoléculaire « classique » : ionisation électrospray (ESI-MS)

1.3.1. Instrumentation

Les études sur les édifices supramoléculaires sont réalisées avec les sources d'ionisation douces, dont font partie l'ESI-MS et le MALDI-MS, capables de maintenir les interactions non-covalentes lors du passage en phase gazeuse. Cependant, les études en MALDI-MS restent en faible nombre par rapport à l'ESI-MS, et nous essayerons de comprendre les raisons de cette limitation dans le chapitre 2. Dans ce chapitre, nous nous focalisons sur l'ESI-MS qui est majoritairement utilisée pour analyser les édifices non-covalents [30,31].

Tout d'abord, l'ionisation douce et l'obtention d'ions multichargés par ESI-MS sont particulièrement adaptées à l'étude des complexes de macromolécules. Plusieurs études sur ces complexes supramoléculaires par ESI-MS ont été reportées, que ce soit sur les complexes de type protéine-protéine [32–34], protéine-ADN [35–37], ou protéine-ligand [38–40].

Un spectromètre de masse comporte quatre sous-unités :

- La source d'ionisation où les analytes sont nébulisés, désolvatés et ionisés.
- L'interface où les ions formés sont transmis à l'analyseur.
- L'analyseur de masse où les ions sont séparés selon leur rapport masse/charge (m/z).
- Le détecteur où l'impact des ions en signaux électriques traités est traduit et mis sous la forme finale d'un spectre de masse.

Dans la source, le processus d'ionisation en ESI-MS a lieu à pression atmosphérique et comporte plusieurs étapes : i) la formation de gouttelettes chargées ; ii) la fission coulombienne de gouttelettes chargées en fine gouttelettes ; iii) l'émission des ions en phase gazeuse (Figure 2).

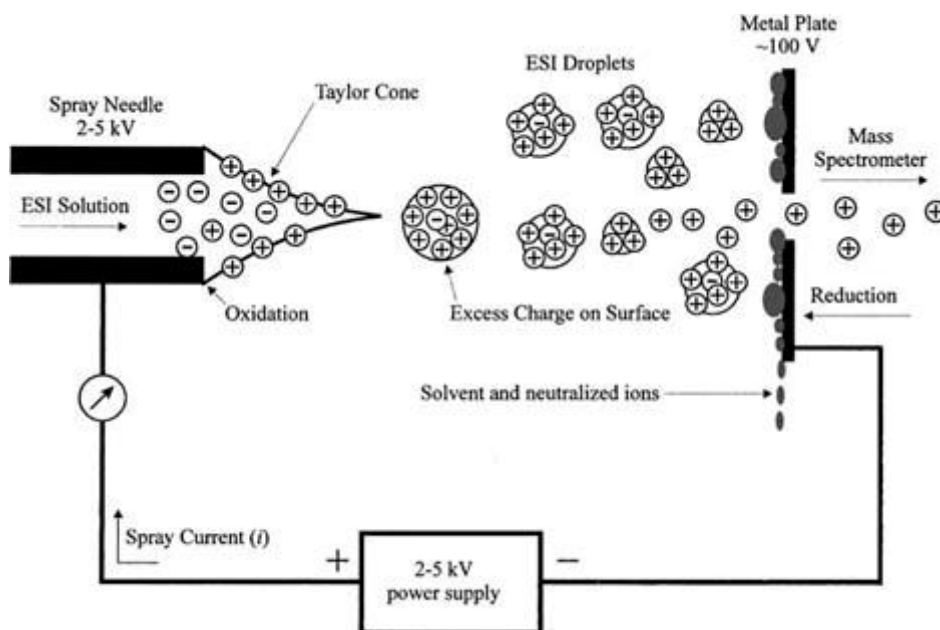


Figure 2. Représentation schématique du processus d'ionisation électrospray en mode positif [41].

L'interface assure le transfert des ions formés dans la source à pression atmosphérique à l'analyseur soumis à un vide poussé (de l'ordre de 10^{-6} - 10^{-7} mbar selon les appareils). La première partie de l'interface comporte les lentilles électrostatiques qui peuvent être associées ou remplacées par les « funnels ». Les ions progressent dans l'interface par la différence de pression ainsi que les différences de potentiel appliquées. Dans la seconde partie, l'interface est munie de multipôles (quadripôles, hexapôles ou octopôles) fonctionnant en mode radio-fréquence (RF) permettant la focalisation des ions vers l'analyseur. Dans cette zone de moyenne pression (de l'ordre du mbar), les ions sont susceptibles d'être accélérés et d'impacter les molécules de gaz résiduelles. Nous allons voir en 1.3.2.2 le rôle crucial de l'interface dans la transmission et le maintien de l'intégrité des espèces non-covalentes.

L'analyseur a pour rôle de séparer et déterminer le rapport m/z des ions. L'analyseur à temps de vol (TOF pour « Time Of Flight ») est le plus couramment utilisé lors des analyses en MS native grâce à sa grande gamme dynamique permettant de mesurer des espèces de grandes masses (de l'ordre de MDa). Il mesure les ions en fonction de leur temps de parcours dans le tube de vol libre de champ. Les ions qui ont les plus faibles rapports m/z vont arriver en premier, tandis que ceux qui ont les m/z plus grands vont mettre plus de temps pour arriver au détecteur. La gamme de masse utilisable d'un analyseur TOF est en théorie illimitée ; mais cette dernière est réduite dans la plupart des cas à 20 000 - 25 000 m/z pour assurer une précision et une résolution nécessaires à la mesure des masses détectées. Cependant, d'autres analyseurs de type Orbitrap comme par exemple le Exactive modifié EMR (« Extended Mass Range ») de Thermo Scientific [42-45] ou le Q-TOF modifié pour une meilleure sélection et transmission des ions de haut m/z [46-48] ont démontré également leur potentiel dans ce type d'approche.

Une fois les ions arrivés au détecteur, ce dernier assure la conversion des courants ioniques en courant électriques. Le détecteur à galettes de micro-canaux (« Microchannel Plate » ou MCP) est le plus souvent utilisé dans des appareils ESI-MS. Il est constitué de micro-canaux assemblés en

galettes et placés perpendiculairement aux faisceaux d'ions. Lorsque les ions heurtent les parois d'un micro-canal, une émission en cascade d'électrons se produit et le signal de l'ion incident est amplifié (Figure 3).

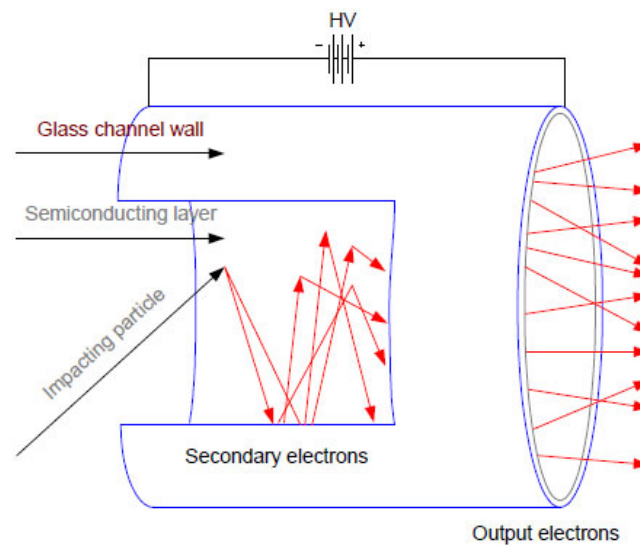


Figure 3. Représentation schématique du fonctionnement d'un micro-canal (source : Bruker).

La Figure 4 représente la configuration du spectromètre de masse de type ES-TOF MicroTOF de Bruker Daltonics utilisé lors de ces travaux de thèse. Les paramètres importants pour l'étude des complexes non-covalents sont encadrés.

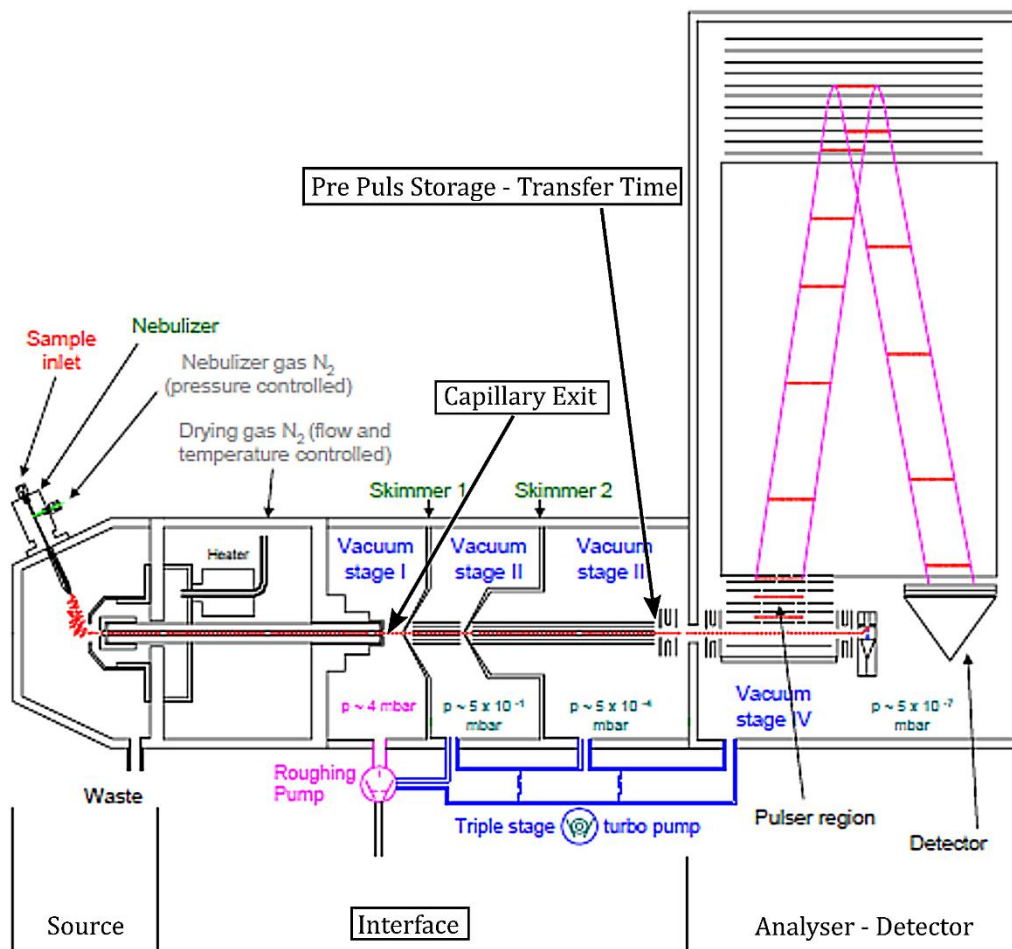


Figure 4. Représentation schématique du spectromètre de masse ES-TOF MicrOTOF. Les paramètres importants pour l'étude des complexes non-covalents sont encadrés (source : adapté de Bruker Daltonics).

1.3.2. Paramètres expérimentaux et instrumentaux spécifiques à l'étude des complexes non-covalents

1.3.2.1. Préparation spécifique de l'échantillon pour la préservation des interactions non-covalentes et compatible avec l'analyse en ESI-MS

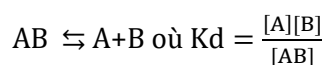
Premièrement, pour pouvoir analyser les complexes non-covalents en MS, il faut se placer dans les conditions d'existence de ces complexes tout en tenant compte de l'intolérance de l'ESI-MS à la présence de sels non-volatils. Afin de mimer les conditions *in vivo*, de manière générale, les échantillons biologiques sont stabilisés *in vitro* dans des tampons contenant un sel non volatil (NaCl, KCl *etc.*), ce qui n'est pas compatible avec l'analyse en ESI-MS. 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 masse moléculaire rendant souvent l'interprétation impossible. De plus, ces sels non volatils entrent en compétition de charges avec l'analyte et entraînent le phénomène de suppression du signal de l'analyte lors du processus d'ionisation par électrospray [49–52].

L'utilisation de sels d'ammonium s'est avérée être un bon compromis : ces tampons volatils sont à la fois compatibles avec l'ionisation électrospray et permettent d'assurer le maintien d'une certaine force ionique et du pH indispensable à l'échantillon biologique pour garder sa conformation native [29,53]. Ils assurent le maintien des édifices non-covalents tout en minimisant l'élargissement des pics sur le spectre de masse. Les sels d'ammonium les plus utilisés sont l'acétate d'ammonium (pH 6.8) et le bicarbonate d'ammonium (pH 7.5), à des concentrations variant de quelques mM à des centaines de mM, voire jusqu'à 1 M dans certain cas [29,53–56]. Récemment, l'équipe de Lars Konermann a démontré que les tampons bicarbonate d'ammonium mènent à la formation de bulles de gaz de CO₂ qui peuvent dénaturer les protéines lors de l'analyse en ESI-MS en conditions natives [57]. Par conséquent, l'acétate d'ammonium reste le tampon de choix malgré son faible pouvoir tampon à pH neutre.

Une étape d'échange de tampon est donc essentielle avant l'analyse en ESI-MS. Elle permet d'échanger le tampon en sortie de la purification ou de conservation d'échantillon biologique contre le tampon ammonium. Les techniques les plus couramment utilisées sont l'ultracentrifugation (comme par exemple les colonnes Centricon de Millipore et Vivaspinn de Vivascience) et la filtration sur gel (on peut citer les colonnes NAP 5 de GE Healthcare ou Biospin de Biorad). L'échange de tampon n'est toutefois pas aisé. D'une part, il faudrait adapter la technique à utiliser suivant les cas, un échantillon peut se dégrader lors d'ultracentrifugation tout comme il est possible d'avoir des pertes de matériel sur les colonnes de filtration. D'autre part, le compromis réside toujours en un échange convenable de tampon qui implique plusieurs cycles d'échange mais qui peuvent entraîner des pertes parfois non négligeable. Enfin, pour les complexes dont la présence du sel inorganique est nécessaire à la stabilisation et qui ne sont pas transférés proprement dans un tampon volatil, l'analyse par ESI-MS supramoléculaire « classique » s'avère être impossible.

Deuxièmement, il y a des précautions à prendre quant à la concentration d'étude. En effet, la formation des complexes dépend de la concentration des partenaires par rapport à la constante d'affinité. Pour un complexe AB, ceci se traduit par l'équilibre thermodynamique suivant :



Ainsi la concentration des partenaires doit être supérieure à K_d pour favoriser l'association non-covalente des espèces A et B. De plus, elle doit être suffisante pour que les espèces présentes puissent être détectées en MS. Cependant, une concentration trop importante en analyte peut mener à des agrégats non-spécifiques en phase gazeuse [58], il est donc important de vérifier quels sont les complexes observés à différentes concentrations, et de travailler à la concentration minimale permettant d'observer le complexe.

Enfin, des paramètres instrumentaux du spectromètre de masse au niveau de l'interface et de l'analyseur influencent également les espèces détectées.

1.3.2.2. Paramètres instrumentaux spécifiques à l'étude des complexes non-covalents

Pour pouvoir être observés lors des analyses en ESI-MS, 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 source-analyseur.

Depuis les premiers travaux en MS supramoléculaire [28,29,53,59–61], des études ont montré qu'il était possible de préserver la conformation native des protéines et de leurs complexes en phase gazeuse [62–64]. Cependant, il ne faut pas oublier que des distorsions de structure peuvent avoir lieu lors du passage de la phase condensée à la phase gazeuse. Lors de la désolvatation des molécules d'eau, les interactions électrostatiques et les liaisons hydrogène se retrouvent renforcées, tandis que l'effet hydrophobe est affaibli et les liaisons de van der Waals restent inchangées. Par conséquent, si l'effet hydrophobe joue un rôle prépondérant dans le maintien des contacts intermoléculaires, il sera probablement difficile d'observer les complexes intacts lors de l'analyse en ESI-MS [65,66].

D'autre part, lors du transfert de la source à l'analyseur, l'énergie interne des ions peut être suffisamment importante pour dissocier le complexe [67]. Cette énergie interne provient :

- De la température du gaz de séchage : elle est à adapter en fonction du solvant utilisé. Pour les échantillons biologiques hydrosolubles, cette température est de l'ordre de 160°C - 180°C et a peu d'influence sur le maintien de l'édifice, sauf dans le cas des duplexes d'ADN où cette température doit être plus basse pour ne pas les dissocier [68,69].
- De l'énergie de collisions avec les molécules de gaz : cette énergie est déterminée par l'intensité du champ électrique dans l'interface. La tension d'accélération (Capillary Exit) est la différence entre le potentiel modulable à la sortie du capillaire de transmission et le potentiel du premier skimmer du micrOTOF. Elle induit un champ électrique dirigeant et accélérant les ions jusqu'à l'analyseur. Une proportion de l'énergie cinétique est ensuite convertie en énergie interne lors des collisions avec les molécules de gaz environnant, provoquant l'évaporation du solvant. Les collisions contre les molécules de gaz favorisent donc la désolvatation des ions en les dissociant de leurs adduits de solvant ou de sels. Si cette tension d'accélération n'est pas suffisante, les ions seront moins bien désolvatés et l'intensité du signal considérablement diminuée. L'augmentation de cette tension permet d'améliorer la désolvatation des ions et leur transfert jusqu'à l'analyseur, mais une intensité trop élevée aboutit à des collisions trop énergétiques qui peuvent induire la dissociation des interactions non-covalentes. Par conséquent, elle est à optimiser pour chaque système d'étude.
- De la pression dans l'interface : lorsque la pression augmente, le libre parcours moyen des ions diminue. Ainsi ils vont subir beaucoup de collisions mais de faibles énergies, ce qui permet de les ralentir pour une meilleure focalisation et assure leur thermalisation réduisant le risque de dissociation [70]. Dans certaines interfaces, il y a donc la possibilité d'augmenter la pression pour favoriser la transmission des ions de haut rapport m/z [71,72].

Par conséquent, il est important d'optimiser ces paramètres 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.

En revanche, il y a des cas où la dissociation en phase gazeuse est provoquée intentionnellement, pour obtenir des informations sur la nature et la force des interactions entre les sous-unités, pour comparer les associations spécifiques de type protéine-ligand [73-75], ou pour avoir des informations structurales [76,77].

Enfin, les espèces ayant des rapports m/z très éloignés vont épouser différentes trajectoires dans le spectromètre de masse. Pour pouvoir observer des complexes biologiques de grande masse, il faudrait optimiser des paramètres de stockage (« pre-pulse storage ») et de transmission (« transfer time ») en faveur de la détection des ions lourds. De même, il est possible d'augmenter les amplitudes de voltage appliqué sur les guides d'ions des multipôles dans l'interface afin de mieux focaliser les ions lourds vers l'analyseur. La Figure 5 tirée de nos travaux montre un exemple de cette observation préférentielle dans l'étude des complexes de récepteurs rétinoïques à l'ADN *RAR β 2*. 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 .

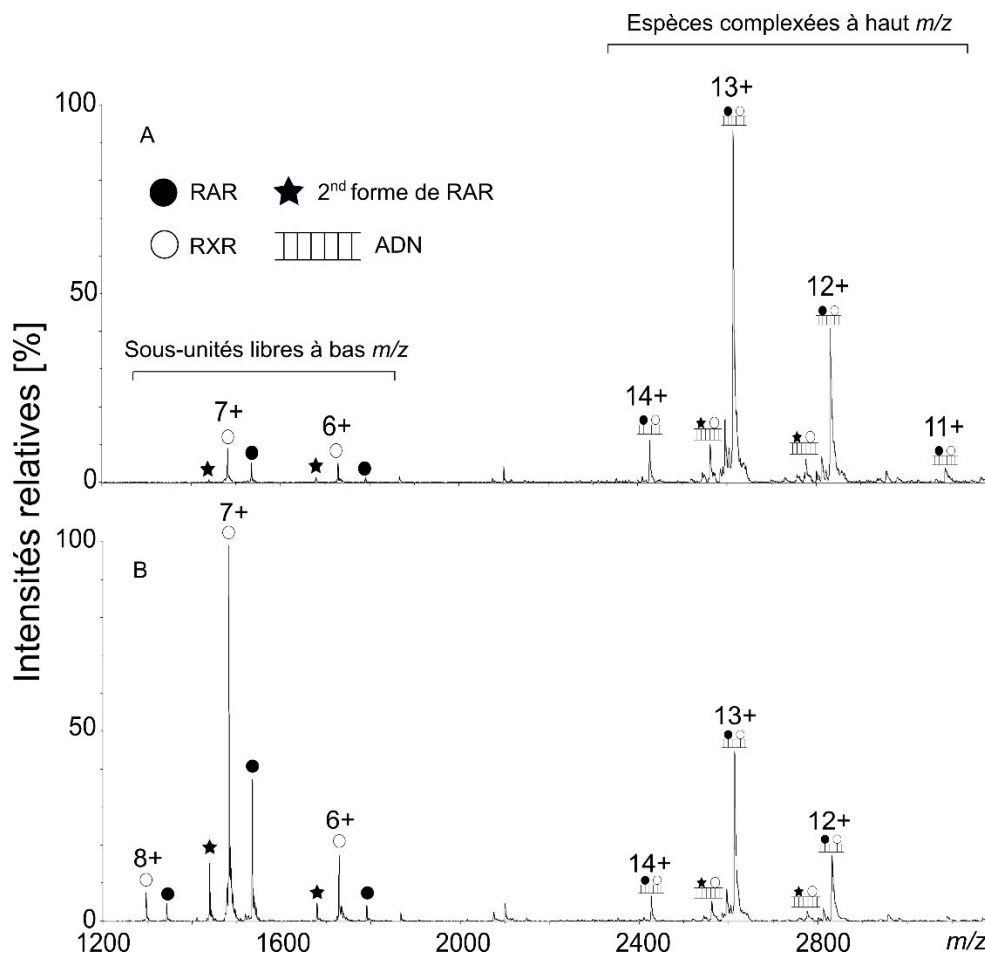


Figure 5. Spectres de masse des récepteurs rétinoïques RAR-RXR incubés avec 0,25 équivalent de *RAR β 2*. (A) Pre-pulse à 40 μ s, (B) Pre-pulse à 30 μ s. Dans les deux cas, la tension de cône (« capillary exit ») est à 150 V.

Pour toutes ces raisons, il est extrêmement important d'optimiser les paramètres d'analyse pour s'assurer que le complexe étudié est maintenu intact en phase gazeuse, notamment dans la zone d'interface du spectromètre de masse. Ces paramètres sont très discriminants et il n'est pas toujours évident de les optimiser afin de refléter au mieux les espèces présentes en solution.

1.3.2.3. Les sources nano électrospray

Les débits utilisés en MS supramoléculaire en infusion directe « classique » sont de l'ordre de quelques $\mu\text{L}/\text{min}$. Les quantités nécessaires pour réaliser des acquisitions en MS peuvent s'avérer importantes et contraignantes lorsque la quantité d'échantillon biologique est limitée.

Pour pallier des problèmes de quantité d'échantillon et de sensibilité, la source nano electro spray (nanoESI) a été développée par Wilm et Mann [78]. Elle est constituée d'un capillaire de nébulisation de plus faible diamètre et capable de réaliser des injections avec des débits de l'ordre d'une vingtaine de nL/min . Ce type de source permet donc de consommer très peu d'échantillon par analyse (quelques microlitres suffisent). Ensuite, il a été remarqué qu'il peut être un atout pour l'analyse des édifices non-covalents. En effet, en plus de la faible quantité d'échantillon nécessaire lors de l'analyse, il permet d'améliorer la désolvatation, de réduire les phénomènes de suppression d'ions et ainsi d'améliorer le signal MS [79–81].

Dès lors, les sources nanoESI se sont largement développées, que ce soit pour l'infusion directe comme par exemple la CaptiveSpray de Bruker Daltonics, le système NanoMate d'Advion BioSciences ou pour un couplage avec une méthode séparative telle que la nano chromatographie liquide (nanoLC) ou une électrophorèse capillaire (CESI). Cependant, il reste difficile à mettre en œuvre car les capillaires peuvent être bouchés plus facilement.

1.3.3. Informations apportées par la spectrométrie de masse dans l'étude des complexes biologiques non-covalents

Le grand avantage de la MS est qu'elle permet d'identifier de façon individuelle différentes espèces présentes simultanément en solution, contrairement à d'autres techniques qui ne peuvent donner qu'une vue globale de l'échantillon. Ceci permet donc de démontrer l'existence des différents d'état d'oligomérisation ou des complexes de stœchiométries différentes. La MS permet ainsi de suivre la dynamique d'interaction en fonction des conditions expérimentales [31], ou encore des effets de coopérativité négatifs ou positifs [82–84].

1.3.3.1. La stœchiométrie

L'information la plus directe qu'on puisse tirer des spectres de masse est la coexistence des différentes espèces présentes et la stœchiométrie des complexes préformés en solution. Des analyses en conditions natives donnent l'accès à la masse totale du complexe ; en connaissant la

masse de chaque sous-unité (cette information est aussi accessible par les analyses en MS en conditions dénaturantes) et avec une précision de mesure de masse suffisante, nous pouvons déterminer l'identité et le nombre de chaque sous-unités formant le complexe [85–87]. Afin de déterminer la stœchiométrie sans ambiguïté, il est crucial de vérifier que le complexe n'a pas été dissocié partiellement lors du passage en phase gazeuse ou lors de la transmission dans l'interface du spectromètre de masse.

1.3.3.2. La spécificité

La spécificité d'une interaction se traduit par la fixation préférentielle d'une molécule sur un site donné de son partenaire. Les analyses en MS peuvent donner des indications pour déterminer si une interaction est spécifique [88]. Si la stœchiométrie observée pour un complexe est multiple et aléatoire, ceci peut indiquer que la fixation est non spécifique. Dans le cadre d'une interaction spécifique, le spectre de masse reste inchangé lors d'un ajout croissant de substrat dans la solution de récepteur, mais varie lors de modifications des conditions expérimentales (comme le pH, la nature de chaque partenaire, le solvant *etc.*)

1.3.3.3. La sélectivité

La sélectivité du complexe formé peut être étudiée par des expériences de compétition en mettant en jeu un récepteur et plusieurs substrats dans le mélange [89]. Si la formation du complexe n'est pas sélective, le récepteur va s'associer indifféremment avec les entités présentes. Si l'un des substrats est sélectif, il va s'associer préférentiellement avec le récepteur même en présence d'autres substrats.

1.3.3.4. La constante d'affinité

Afin d'avoir des informations sur la constante de formation des complexes, nous pouvons réaliser les expériences de titration en MS en conditions natives [90,91]. De manière générale, un des partenaires est maintenu en quantité fixe, et la quantité de l'autre partenaire varie. Les intensités des pics correspondant aux entités libres et aux complexes sont ensuite tracées en fonction des concentrations ajoutées, puis une régression non linéaire de ces données permettra le calcul des constantes d'association (ou de dissociation). Les constantes relatives peuvent aussi être déterminées par des expériences de compétition [92]. Le chapitre 5 donne plus de détails à ce sujet, notamment au travers de l'étude des constantes de dissociation du récepteur X des rétinoïdes (RXR) avec différents ligands.

1.4. Quelques approches complémentaires

Afin d'avoir des informations complémentaires à l'ESI-MS supramoléculaire « classique » telles que les indications sur la conformation dépliée/globulaire des biomolécules ou les zones en interaction entre les sous-unités, depuis une dizaine d'années se développent les approches alternatives. Nous pouvons citer par exemple :

- La mobilité ionique couplée à la MS (IM-MS)
- L'échange hydrogène/deutérium (H/D) couplé à la MS (HDX-MS)
- Le marquage chimique couplé à la MS
- La protéolyse ménagée couplée à la MS (LP-MS)
- Le cross-link ou pontage chimique couplé à la MS (CX-MS)

1.4.1. La mobilité ionique couplée à la MS

La mobilité ionique (IM-MS) consiste à séparer les ions selon leur mobilité, qui dépend de leur taille et leur forme. Les structures de conformation compacte arrivent plus vite que les structures dépliées. En mesurant le temps de dérive d'ions dans le tube de dérive rempli de gaz sous un champ électrique, il est possible d'évaluer leur section efficace de collision (« collision cross-section » – CCS) et d'avoir ainsi une indication sur la conformation des ions en phase gazeuse. Trois techniques sont utilisées en IM-MS (Figure 6) : DTIMS (drift-time ion mobility spectrometry), TWIMS (travelling-wave ion mobility spectrometry) et FAIMS (field-asymmetric ion mobility spectrometry). Parmi les systèmes IM-MS commerciaux, nous pouvons mentionner le Synapt de Waters basé sur le TWIMS et le 6560 IMS-Q-TOF de Agilent basé sur le DTIMS.

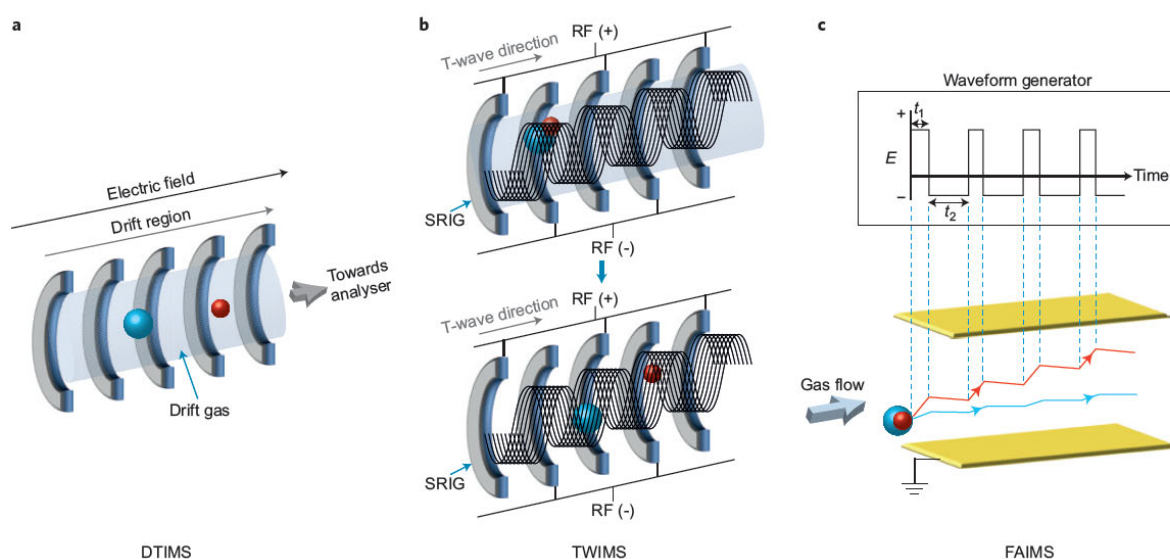


Figure 6. Représentation schématique des techniques utilisées en IM-MS [93].

De nombreuses revues ont montré le potentiel d'IM-MS à fournir des informations structurales des édifices non-covalents [94–99]. L'équipe de C. Robinson a ainsi démontré que les édifices biologiques non-covalents peuvent être transférés de la phase aqueuse à la phase gazeuse en maintenant leur structure quaternaire [100]. Grâce au faible temps d'analyse (de l'ordre de la milliseconde), l'IM-MS a été également utilisée avec succès pour étudier les intermédiaires et des transitions conformationnelles [101–103]. On peut aussi citer les travaux de Shvartsburg *et al.* sur la séparation des conformères de protéine [104], ou les études sur les changements de conformation induits par les associations avec le ligand [105,106]. De plus, l'IM-MS peut être efficace 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- β dont l'agrégation mène aux maladies neurodégénératives [107,108]. L'équipe de A. Heck a également publié des études en IM-MS sur les capsides de virus de l'hépatite B d'une masse de 4 MDa [109].

Cependant, quelques inconvénients de l'IM-MS sont à relever. Il a été observé que des complexes de protéine 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 [110]. 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 [93]. Enfin, l'utilisation de cette technique implique l'achat d'un appareil dédié.

1.4.2. L'échange hydrogène/deutérium (H/D) couplé à la MS

L'échange H/D consiste à faire incuber l'échantillon avec une solution de D₂O pour que ses hydrogènes labiles puissent s'échanger avec les deutériums de la solution (Figure 7). Les hydrogènes capables de s'échanger sont les hydrogènes des chaînes latérales liés aux hétéroatomes (N, O et S), les hydrogènes des côtés N-terminal et C-terminal, et les hydrogènes amidiques des liaisons peptidiques, mais seuls les échanges des hydrogènes amidiques sont observables [111].

Les hydrogènes impliqués dans des liaisons hydrogènes intramoléculaires et les hydrogènes enfouis au cœur de la structure des édifices supramoléculaires vont s'échanger plus lentement. En revanche, les hydrogènes exposés à la surface accessible au solvant et qui ne sont pas impliqués dans des liaisons hydrogènes peuvent s'échanger avec les deutériums du solvant de manière rapide. 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 de leurs complexes [112].

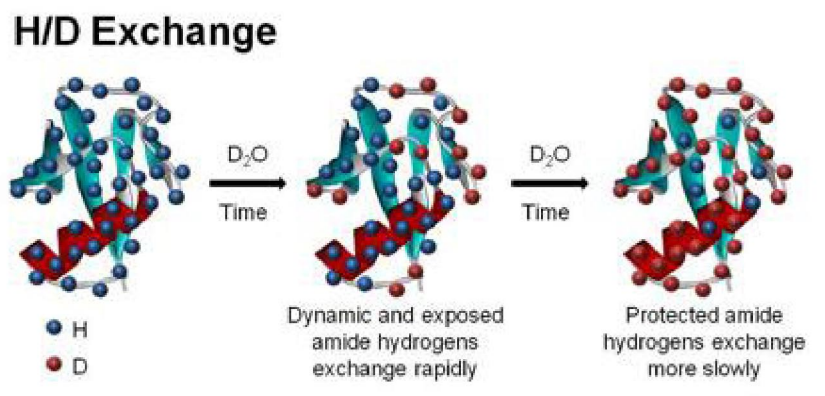


Figure 7. Représentation schématique du fonctionnement de l'échange H/D (source : National High Magnetic Field Laboratory).

Historiquement, le RMN est la méthode utilisée pour suivre les échanges H/D, jusqu'à ce que les équipes de Chait [113] et de Smith [114] initient le couplage avec la MS (HDX-MS). Dès lors, le HDX-MS a fait ses preuves dans les études sur les interactions de type protéine-ligand [115–118], protéine-protéine [119,120], sur la détermination des changements dans l'accessibilité du solvant dus à la formation du complexe [111,121], sur les dynamiques de dépliement/repliement des protéines et leurs mécanismes d'assemblage [122,123]. Enfin, le HDX-MS peut aussi faciliter la cristallogénèse en identifiant les régions flexibles ayant une faible structuration et une grande vitesse d'échange isotopique [97,124].

Après l'exposition au solvant deutéré, vient l'étape importante de « quenching » pour conserver des taux de deutération les plus proches possibles de ceux à la fin de la période de 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. Pour éviter les rééchanges, les étapes d'analyse qui suivent doivent être aussi maintenues à ces conditions. De plus, les rééchanges sont possibles avec l'humidité dans l'air. Ceci constitue des contraintes expérimentales indéniables à ce type d'approche.

1.4.3. Le marquage chimique couplé à la MS

En parallèle à l'échange H/D, la topologie de surface des protéines et de leurs complexes peut être étudiée par marquage chimique suivi de l'analyse en MS [125]. Cette approche consiste à utiliser les agents de marquage qui réagissent avec les acides aminés cibles. Ainsi, la protéine est marquée seule et en présence de ses partenaires. Les changements conformationnels liés à la formation du complexe vont modifier l'accessibilité et la réactivité des sites vis-à-vis du réactif. Ainsi l'analyse différentielle en MS après la protéolyse permettra l'identification des résidus impliqués dans l'interface d'interaction.

Dès lors, le marquage chimique a permis d'étudier les complexes de types protéine-ligand [126], protéine-ADN [127] ou protéine-protéine. Le choix de l'agent de marquage est important car

il ne doit pas affecter la structure de la protéine ni ses interactions avec ses partenaires [128]. De plus, son temps de réaction est crucial. Certains agents utilisés demandent un temps de réaction de quelques minutes à quelques heures : ce temps assez long peut ne pas être adapté aux complexes dynamiques car à chaque fois que les derniers se dissocient, les sites d'interaction sont accessibles aux agents de marquage. Pour y remédier, il existe des possibilités de marquage rapide en utilisant le rayonnement synchrotron pour générer les radicaux hydroxyles HO· qui oxydent des acides aminés des complexes dans l'intervalle de temps de la milliseconde [126]. En résumé, ces conditions expérimentales non triviales entraînent des problèmes de reproductibilité et d'automatisation, et cette approche ne permet de donner que les informations concernant la topologie de surface.

1.4.4. La protéolyse ménagée couplée à la MS

La protéolyse ménagée (« limited proteolysis ») repose sur le principe que les informations de structure peuvent être déduites à partir de la détermination de la protection des protéines contre la protéolyse enzymatique. 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 les partenaires sont protégées contre la protéolyse. En revanche, les régions non structurées, flexibles ou qui sont exposées au solvant sont moins protégées et plus susceptibles à la digestion. Ainsi la comparaison des fragmentations résultantes de la protéolyse et des vitesses de digestion en présence et en absence des partenaires de la protéine d'intérêt permettront d'obtenir des informations sur les régions accessibles aux protéases, sur la flexibilité de la protéine en solution, les zones d'interaction ou encore les changements conformationnels liés à l'interaction [129]. Cette approche peut également être utile pour savoir si les changements de conformation induits par la liaison avec un premier partenaire inhibent la liaison avec un deuxième partenaire sur un autre site d'interaction [130].

La protéolyse ménagée a été utilisée avec succès pour déterminer des interfaces d'interactions entre les sous-unités des complexes de type protéine-protéine [131], protéine-ADN [129] et protéine-ligand [130]. Bien entendu, les conditions de réaction (choix de l'enzyme et de sa concentration, temps d'incubation) sont primordiales. En général, la protéolyse ménagée est conduite avec différentes protéases (trypsine, thermolysine, Glu-C, Lys-C, Asp-N, chymotrypsine, *etc.*) ; le choix des protéases adéquats peut être facilité si les connaissances préalables concernant les zones d'interaction sont à disposition [130]. La cinétique de protéolyse est également suivie à différents temps d'incubation ; le suivi de la protéolyse et l'analyse des produits peuvent être effectués par MALDI-MS ou ESI-MS [129].

Parmi les inconvénients de cette approche, nous pouvons citer le cas des complexes de type protéine-ADN : les interactions non spécifiques entre les acides aminés chargés positifs de la protéine et les groupements phosphates chargés négatifs de l'ADN peuvent constituer une barrière de protection de la protéine contre la protéolyse, ce qui complique la corrélation entre la protection à la protéolyse et l'information structurale [129]. Dans le cas des équilibres entre les formes liée et non liée de la protéine, la protéolyse peut avoir lieu sur les formes libres et liées, ce qui mènerait à des ambiguïtés dans l'interprétation des données [129]. Pour pallier ce problème, le partenaire d'interaction peut être introduit en excès pour que tous les sites de liaison possibles de la protéine soient occupés ; mais il faut être prudent afin que ceci n'induisse pas les interactions non spécifiques [130].

1.4.5. Le pontage chimique couplé à la MS

Le pontage chimique consiste à lier de façon covalente les sous-unités qui sont en interaction non-covalente ou qui sont proches, par des réactions chimiques avec l'agent de pontage (ou « cross-linker »). Les agents de pontage sont constitués de fonctions réactives qui peuvent être identiques ou différents. Chaque type d'agents de pontage réagit sur des acides aminés différents ; le choix s'effectue donc en fonction des acides aminés cibles et de leur présence dans la séquence des protéines d'intérêt. La réaction de pontage étant protéine dépendante, sa mise en œuvre constitue un processus d'essai-erreur [132] : il s'agit de trouver la nature et la quantité de l'agent de pontage adéquates pour une réaction de pontage réussie.

Une fois les complexes protéiques stabilisés de façon covalente, on peut déterminer sa stœchiométrie et suivre l'avancement de réaction de pontage par gel d'électrophorèse 1D (SDS-PAGE) ou par MALDI-MS [133,134]. Outre la stœchiométrie, les informations sur les zones en interactions et leurs contraintes de distance peuvent aussi être obtenues suite à une fragmentation dans le spectromètre de masse (stratégie « top-down ») [135] ou une digestion enzymatique de l'échantillon ponté en gel ou en solution (stratégie « bottom-up ») [132,136], mais ceci demande l'utilisation de moteurs de recherche dédiés à l'identification des peptides pontés. Enfin, les pontages *in vivo* sont également possibles pour identifier les partenaires d'interaction [137].

A mon arrivée au laboratoire, quelques études développées suggéraient la possibilité de caractériser des systèmes pontés, mais de nombreuses difficultés sont dès lors évoquées. Parmi celles-ci, on peut mentionner notamment un faible rendement de pontage ou un pontage incomplet qui rend l'interprétation difficile, et un manque cruel d'outil bio-informatique pour l'interprétation des résultats. Ainsi, les principales difficultés de cette approche concernent essentiellement l'optimisation des conditions de réaction et l'interprétation des données.

Nous développerons plus en détails le pontage chimique couplé à la MS dans les chapitres 2 et 3, notamment à travers l'application de cette approche pour déterminer la stœchiométrie et les zones en interaction d'un complexe recombinant humain SAGA HAT.

Bibliographie

- [1] Atlan, H., *La fin du "tout génétique" ? : vers de nouveaux paradigmes en biologie*, Editions Quae, 1999.
- [2] Beadle, G.W., Tatum, E.L., Genetic Control of Biochemical Reactions in *Neurospora*. *Proc. Natl. Acad. Sci. U. S. A.* 1941, 27, 499–506.
- [3] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., et al., Initial sequencing and analysis of the human genome. *Nature* 2001, 409, 860–921.
- [4] Finley, R.L., Brent, R., Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. U. S. A.* 1994, 91, 12980–12984.
- [5] Bartel, P.L., Roecklein, J.A., SenGupta, D., Fields, S., A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* 1996, 12, 72–77.
- [6] Strohmaier, R.C., The coming Kuhnian revolution in biology. *Nat. Biotechnol.* 1997, 15, 194–200.
- [7] Arkin, A.P., Schaffer, D.V., Network News: Innovations in 21st Century Systems Biology. *Cell* 2011, 144, 844–849.
- [8] Bensimon, A., Heck, A.J.R., Aebersold, R., Mass Spectrometry–Based Proteomics and Network Biology. *Annu. Rev. Biochem.* 2012, 81, 379–405.
- [9] Rembach, A., Stingo, F.C., Peterson, C., Vannucci, M., et al., Bayesian graphical network analyses reveal complex biological interactions specific to Alzheimer's disease. *J. Alzheimers Dis. JAD* 2015, 44, 917–925.
- [10] Guggino, W.B., Stanton, B.A., New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* 2006, 7, 426–436.
- [11] Calderwood, M.A., Venkatesan, K., Xing, L., Chase, M.R., et al., Epstein–Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U. S. A.* 2007, 104, 7606–7611.
- [12] Mendez-Rios, J., Uetz, P., Global approaches to study protein–protein interactions among viruses and hosts. *Future Microbiol.* 2010, 5, 289–301.
- [13] de Chasse, B., Navratil, V., Tafforeau, L., Hiet, M.S., et al., Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 2008, 4, 230.
- [14] Shapira, S.D., Gat-Viks, I., Shum, B.O.V., Dricot, A., et al., A physical and regulatory map of host–influenza interactions reveals pathways in H1N1 infection. *Cell* 2009, 139, 1255–1267.
- [15] Ptak, R.G., Fu, W., Sanders-Beer, B.E., Dickerson, J.E., et al., Short Communication: Cataloguing the HIV Type 1 Human Protein Interaction Network. *AIDS Res. Hum. Retroviruses* 2008, 24, 1497–1502.
- [16] Jäger, S., Gulbahce, N., Cimermanic, P., Kane, J., et al., Purification and characterization of HIV–human protein complexes. *Methods San Diego Calif* 2011, 53, 13–19.
- [17] Kreger, P.K., Lauffenburger, D.A., Cancer systems biology: a network modeling perspective. *Carcinogenesis* 2010, 31, 2–8.
- [18] Pe'er, D., Hacohen, N., Principles and Strategies for Developing Network Models in Cancer. *Cell* 2011, 144, 864–873.
- [19] Chen, F., Applying MALDI-ToF Mass Spectrometry to Study Non-covalent Interactions. Thèse ETH Zürich, 2014.
- [20] Sirajuddin, M., Ali, S., Badshah, A., Drug–DNA interactions and their study by UV–Visible, fluorescence spectroscopies and cyclic voltametry. *J. Photochem. Photobiol. B* 2013, 124, 1–19.
- [21] Ghai, R., Falconer, R.J., Collins, B.M., Applications of isothermal titration calorimetry in pure and applied research—survey of the literature from 2010. *J. Mol. Recognit.* 2012, 25, 32–52.
- [22] Giegé, R., A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* 2013, 280, 6456–6497.

- [23] Göbl, C., Madl, T., Simon, B., Sattler, M., NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* 2014, 80, 26–63.
- [24] Ranjbar, B., Gill, P., Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review. *Chem. Biol. Drug Des.* 2009, 74, 101–120.
- [25] Kelly, S.M., Jess, T.J., Price, N.C., How to study proteins by circular dichroism. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 2005, 1751, 119–139.
- [26] Homola, J., Surface Plasmon Resonance Sensors for Detection of Chemical and Biological Species. *Chem. Rev.* 2008, 108, 462–493.
- [27] Milne, J.L.S., Borgnia, M.J., Bartesaghi, A., Tran, E.E.H., et al., Cryo-electron microscopy: A primer for the non-microscopist. *FEBS J.* 2013, 280, 28–45.
- [28] Katta, V., Chait, B.T., Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* 1991, 113, 8534–8535.
- [29] Ganem, B., Li, Y.T., Henion, J.D., Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J. Am. Chem. Soc.* 1991, 113, 6294–6296.
- [30] Loo, J.A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* 1997, 16, 1–23.
- [31] Sharon, M., How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? *J. Am. Soc. Mass Spectrom.* 2010, 21, 487–500.
- [32] Pinkse, M.W.H., Maier, C.S., Kim, J.-I., Oh, B.-H., Heck, A.J.R., Macromolecular assembly of *Helicobacter pylori* urease investigated by mass spectrometry. *J. Mass Spectrom.* 2003, 38, 315–320.
- [33] Heck, A.J.R., van den Heuvel, R.H.H., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* 2004, 23, 368–389.
- [34] Boeri Erba, E., Barylyuk, K., Yang, Y., Zenobi, R., Quantifying Protein–Protein Interactions Within Noncovalent Complexes Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 2011, 83, 9251–9259.
- [35] Veenstra, T.D., Electrospray Ionization Mass Spectrometry: A Promising New Technique in the Study of Protein/DNA Noncovalent Complexes. *Biochem. Biophys. Res. Commun.* 1999, 257, 1–5.
- [36] Hanson, C.L., Robinson, C.V., Protein-Nucleic Acid Interactions and the Expanding Role of Mass Spectrometry. *J. Biol. Chem.* 2004, 279, 24907–24910.
- [37] Akashi, S., Osawa, R., Nishimura, Y., Evaluation of protein-DNA binding affinity by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2005, 16, 116–125.
- [38] Touboul, D., Maillard, L., Grässlin, A., Moumne, R., et al., How to Deal with Weak Interactions in Noncovalent Complexes Analyzed by Electrospray Mass Spectrometry: Cyclopeptidic Inhibitors of the Nuclear Receptor Coactivator 1-STAT6. *J. Am. Soc. Mass Spectrom.* 2009, 20, 303–311.
- [39] Maple, H.J., Garlish, R.A., Rigau-Roca, L., Porter, J., et al., Automated Protein–Ligand Interaction Screening by Mass Spectrometry. *J. Med. Chem.* 2012, 55, 837–851.
- [40] Boeri Erba, E., Petosa, C., The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* 2015, 24, 1176–1192.
- [41] Cech, N.B., Enke, C.G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* 2001, 20, 362–387.
- [42] Rose, R.J., Damoc, E., Denisov, E., Makarov, A., Heck, A.J.R., High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* 2012, 9, 1084–1086.
- [43] Rosati, S., Bremer, E.T. van den, Schuurman, J., Parren, P.W., et al., In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs* 2013, 5, 917–924.
- [44] Maple, H.J., Scheibner, O., Baumert, M., Allen, M., et al., Application of the Exactive Plus EMR for automated protein–ligand screening by non-covalent mass spectrometry. *Rapid Commun. Mass Spectrom.* 2014, 28, 1561–1568.

- [45] Gautier, V., Boumeester, A.J., Lössl, P., Heck, A.J.R., Lysine conjugation properties in human IgGs studied by integrating high-resolution native mass spectrometry and bottom-up proteomics. *PROTEOMICS* 2015, n/a-n/a.
- [46] Sobott, F., Hernández, H., McCammon, M.G., Tito, M.A., Robinson, C.V., A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Anal. Chem.* 2002, 74, 1402–1407.
- [47] van den Heuvel, R.H.H., van Duijn, E., Mazon, H., Synowsky, S.A., et al., Improving the Performance of a Quadrupole Time-of-Flight Instrument for Macromolecular Mass Spectrometry. *Anal. Chem.* 2006, 78, 7473–7483.
- [48] Snijder, J., Rose, R.J., Veessler, D., Johnson, J.E., Heck, A.J.R., Studying 18 Mega Dalton Virus Assemblies with Native Mass Spectrometry. *Angew. Chem. Int. Ed Engl.* 2013, 52, 4020–4023.
- [49] Tang, L., Kebarle, P., Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. *Anal. Chem.* 1993, 65, 3654–3668.
- [50] Enke, C.G., A Predictive Model for Matrix and Analyte Effects in Electrospray Ionization of Singly-Charged Ionic Analytes. *Anal. Chem.* 1997, 69, 4885–4893.
- [51] Constantopoulos, T.L., Jackson, G.S., Enke, C.G., Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 1999, 10, 625–634.
- [52] Tang, K., Page, J.S., Smith, R.D., Charge Competition and the Linear Dynamic Range of Detection in Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2004, 15, 1416–1423.
- [53] Ganem, B., Li, Y.T., Henion, J.D., Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray mass spectrometry. *J. Am. Chem. Soc.* 1991, 113, 7818–7819.
- [54] Heck, A.J.R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* 2008, 5, 927–933.
- [55] Sterling, H.J., Batchelor, J.D., Wemmer, D.E., Williams, E.R., Effects of Buffer Loading for Electrospray Ionization Mass Spectrometry of a Noncovalent Protein Complex that Requires High Concentrations of Essential Salts. *J. Am. Soc. Mass Spectrom.* 2010, 21, 1045–1049.
- [56] Banerjee, S., Mazumdar, S., Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* 2012, 2012.
- [57] Hedges, J.B., Vahidi, S., Yue, X., Konermann, L., Effects of Ammonium Bicarbonate on the Electrospray Mass Spectra of Proteins: Evidence for Bubble-Induced Unfolding. *Anal. Chem.* 2013, 85, 6469–6476.
- [58] Terrier, P., Tortajada, J., Buchmann, W., A Study of Noncovalent Complexes Involving Single-Stranded DNA and Polybasic Compounds Using Nanospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2007, 18, 346–358.
- [59] Baca, M., Kent, S.B.H., Direct observation of a ternary complex between the dimeric enzyme HIV-1 protease and a substrate-based inhibitor. *J. Am. Chem. Soc.* 1992, 114, 3992–3993.
- [60] Allen, M.H., Hutchens, T.W., Electrospray-ionization mass spectrometry for the detection of discrete peptide/metal-ion complexes involving multiple cysteine (sulfur) ligands. *Rapid Commun. Mass Spectrom. RCM* 1992, 6, 308–312.
- [61] Ganguly, A.K., Pramanik, B.N., Tsarbopoulos, A., Covey, T.R., et al., Mass spectrometric detection of the noncovalent GDP-bound conformational state of the human H-ras protein. *J. Am. Chem. Soc.* 1992, 114, 6559–6560.
- [62] Clemmer, D.E., Hudgins, R.R., Jarrold, M.F., Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* 1995, 117, 10141–10142.
- [63] Collings, B.A., Douglas, D.J., Conformation of Gas-Phase Myoglobin Ions. *J. Am. Chem. Soc.* 1996, 118, 4488–4489.

- [64] Hunter, C.L., Mauk, A.G., Douglas, D.J., Dissociation of Heme from Myoglobin and Cytochrome b5: Comparison of Behavior in Solution and the Gas Phase. *Biochemistry (Mosc.)* 1997, 36, 1018–1025.
- [65] Robinson, C.V., Chung, E.W., Kragelund, B.B., Knudsen, J., et al., Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein–CoA Ligand Binding and Assembly. *J. Am. Chem. Soc.* 1996, 118, 8646–8653.
- [66] Wu, Q., Gao, J., Joseph-McCarthy, D., Sigal, G.B., et al., Carbonic Anhydrase-Inhibitor Binding: From Solution to the Gas Phase. *J. Am. Chem. Soc.* 1997, 119, 1157–1158.
- [67] Gabelica, V., Pauw, E.D., Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom. Rev.* 2005, 24, 566–587.
- [68] Wan, K.X., Gross, M.L., Shibue, T., Gas-phase stability of double-stranded oligodeoxynucleotides and their noncovalent complexes with DNA-binding drugs as revealed by collisional activation in an ion trap. *J. Am. Soc. Mass Spectrom.* 2000, 11, 450–457.
- [69] Gupta, R., Kapur, A., Beck, J.L., Sheil, M.M., Positive ion electrospray ionization mass spectrometry of double-stranded DNA/drug complexes. *Rapid Commun. Mass Spectrom.* 2001, 15, 2472–2480.
- [70] Schmidt, A., Bahr, U., Karas, M., Influence of Pressure in the First Pumping Stage on Analyte Desolvation and Fragmentation in Nano-ESI MS. *Anal. Chem.* 2001, 73, 6040–6046.
- [71] Tahallah, N., Pinkse, M., Maier, C.S., Heck, A.J.R., 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 Commun. Mass Spectrom.* 2001, 15, 596–601.
- [72] Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R.H.H., Heck, A.J.R., Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int. J. Mass Spectrom.* 2007, 268, 198–206.
- [73] Potier, N., Barth, P., Tritsch, D., Biellmann, J.-F., Van Dorsselaer, A., Study of Non-Covalent Enzyme-Inhibitor Complexes of Aldose Reductase by Electrospray Mass Spectrometry. *Eur. J. Biochem.* 1997, 243, 274–282.
- [74] Cubrilovic, D., Biela, A., Sielaff, F., Steinmetzer, T., 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. *J. Am. Soc. Mass Spectrom.* 2012, 23, 1768–1777.
- [75] Rogniaux, H., Van Dorsselaer, A., Barth, P., Biellmann, J.F., et al., Binding of aldose reductase inhibitors: correlation of crystallographic and mass spectrometric studies. *J. Am. Soc. Mass Spectrom.* 1999, 10, 635–647.
- [76] Schwartz, B.L., Bruce, J.E., Anderson, G.A., Hofstadler, S.A., et al., Dissociation of tetrameric ions of noncovalent streptavidin complexes formed by electrospray ionization. *J. Am. Soc. Mass Spectrom.* 1995, 6, 459–465.
- [77] Felitsyn, N., Kitova, E.N., Klassen, J.S., Thermal dissociation of the protein homodimer ecotin in the gas phase. *J. Am. Soc. Mass Spectrom.* 2002, 13, 1432–1442.
- [78] Wilm, M., Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* 1996, 68, 1–8.
- [79] Sobott, F., Robinson, C.V., Protein complexes gain momentum. *Curr. Opin. Struct. Biol.* 2002, 12, 729–734.
- [80] Benesch, J.L.P., Sobott, F., Robinson, C.V., Thermal Dissociation of Multimeric Protein Complexes by Using Nanoelectrospray Mass Spectrometry. *Anal. Chem.* 2003, 75, 2208–2214.
- [81] Hilton, G.R., Benesch, J.L.P., Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. R. Soc. Interface* 2012, 9, 801–816.
- [82] Edwards, M.J., Williams, M.A., Maxwell, A., McKay, A.R., Mass spectrometry reveals that the antibiotic simocyclinone D8 binds to DNA gyrase in a “bent-over” conformation: evidence of positive cooperativity in binding. *Biochemistry (Mosc.)* 2011, 50, 3432–3440.

- [83] Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A., Sharon, M., Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2013, 110, 7235–7239.
- [84] Lin, H., Kitova, E.N., Klassen, J.S., Measuring positive cooperativity using the direct ESI-MS assay. Cholera toxin B subunit homopentamer binding to GM1 pentasaccharide. *J. Am. Soc. Mass Spectrom.* 2014, 25, 104–110.
- [85] Fitzgerald, M.C., Chernushevich, I., Standing, K.G., Whitman, C.P., Kent, S.B., Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 1996, 93, 6851–6856.
- [86] Zhang, Z., Krutchinsky, A., Endicott, S., Realini, C., et al., Proteasome Activator 11S REG or PA28: Recombinant REG α /REG β Hetero-oligomers Are Heptamers1. *Biochemistry (Mosc.)* 1999, 38, 5651–5658.
- [87] Sanglier, S., Leize, E., Dorselaer, A., Zal, F., Comparative ESI-MS study of ~2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. *J. Am. Soc. Mass Spectrom.* 2003, 14, 419–429.
- [88] Lengqvist, J., Mata de Urquiza, A., Perlmann, T., Sjövall, J., Griffiths, W.J., Specificity of receptor-ligand interactions and their effect on dimerisation as observed by electrospray mass spectrometry: bile acids form stable adducts to the RXR α . *J. Mass Spectrom. JMS* 2005, 40, 1448–1461.
- [89] Keller, K.M., Breeden, M.M., Zhang, J., Ellington, A.D., Brodbelt, J.S., Electrospray ionization of nucleic acid aptamer/small molecule complexes for screening aptamer selectivity. *J. Mass Spectrom.* 2005, 40, 1327–1337.
- [90] Daniel, J.M., McCombie, G., Wendt, S., Zenobi, R., Mass spectrometric determination of association constants of adenylate kinase with two noncovalent inhibitors. *J. Am. Soc. Mass Spectrom.* 2003, 14, 442–448.
- [91] Jecklin, M.C., Touboul, D., Bovet, C., Wortmann, A., Zenobi, R., Which Electrospray-Based Ionization Method Best Reflects Protein-Ligand Interactions Found in Solution? A Comparison of ESI, nanoESI, and ESSI for the Determination of Dissociation Constants with Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2008, 19, 332–343.
- [92] Cheng, X., Chen, R., Bruce, J.E., Schwartz, B.L., et al., Using Electrospray Ionization FTICR Mass Spectrometry To Study Competitive Binding of Inhibitors to Carbonic Anhydrase. *J. Am. Chem. Soc.* 1995, 117, 8859–8860.
- [93] Lanucara, F., Holman, S.W., Gray, C.J., Evers, C.E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* 2014, 6, 281–294.
- [94] Jurnecko, E., Barran, P.E., 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* 2010, 136, 20–28.
- [95] Uetrecht, C., Rose, R.J., Duijn, E. van, Lorenzen, K., Heck, A.J.R., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 2010, 39, 1633–1655.
- [96] Zhong, Y., Hyung, S.-J., Ruotolo, B.T., Ion mobility-mass spectrometry for structural proteomics. *Expert Rev. Proteomics* 2012, 9, 47–58.
- [97] Konijnenberg, A., Butterer, A., Sobott, F., Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 2013, 1834, 1239–1256.
- [98] Niu, S., Rabuck, J.N., Ruotolo, B.T., Ion mobility-mass spectrometry of intact protein-ligand complexes for pharmaceutical drug discovery and development. *Curr. Opin. Chem. Biol.* 2013, 17, 809–817.
- [99] Maurer, M.M., Donohoe, G.C., Valentine, S.J., Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity. *Analyst* 2015.
- [100] Ruotolo, B.T., Giles, K., Campuzano, I., Sandercock, A.M., et al., Evidence for macromolecular protein rings in the absence of bulk water. *Science* 2005, 310, 1658–1661.

- [101] Bernstein, S.L., Wytttenbach, T., Baumketner, A., Shea, J.-E., et al., Amyloid β -Protein: Monomer Structure and Early Aggregation States of A β 42 and Its Pro19 Alloform. *J. Am. Chem. Soc.* 2005, 127, 2075–2084.
- [102] Gidden, J., Bushnell, J.E., Bowers, M.T., Gas-phase conformations and folding energetics of oligonucleotides: dTG- and dGT-. *J. Am. Chem. Soc.* 2001, 123, 5610–5611.
- [103] Gidden, J., Bowers, M.T., Gas-phase conformational and energetic properties of deprotonated dinucleotides. *Eur. Phys. J. - At. Mol. Opt. Plasma Phys.* 2002, 20, 409–419.
- [104] Shvartsburg, A.A., Danielson, W.F., Smith, R.D., High-Resolution Differential Ion Mobility Separations Using Helium-Rich Gases. *Anal. Chem.* 2010, 82, 2456–2462.
- [105] Hyung, S.-J., Robinson, C.V., Ruotolo, B.T., Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* 2009, 16, 382–390.
- [106] Hopper, J.T.S., Oldham, N.J., Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom.* 2009, 20, 1851–1858.
- [107] Bernstein, S.L., Dupuis, N.F., Lazo, N.D., Wytttenbach, T., et al., Amyloid- β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* 2009, 1, 326–331.
- [108] Bleiholder, C., Do, T.D., Wu, C., Economou, N.J., et al., Ion Mobility Spectrometry Reveals the Mechanism of Amyloid Formation of A β (25–35) and Its Modulation by Inhibitors at the Molecular Level: Epigallocatechin Gallate and Scyllo-inositol. *J. Am. Chem. Soc.* 2013, 135, 16926–16937.
- [109] Uetrecht, C., Versluis, C., Watts, N.R., Wingfield, P.T., et al., Stability and Shape of Hepatitis B Virus Capsids In Vacuo. *Angew. Chem. Int. Ed Engl.* 2008, 47, 6247–6251.
- [110] Hall, Z., Politis, A., Bush, M.F., Smith, L.J., Robinson, C.V., Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* 2012, 134, 3429–3438.
- [111] Zhou, B., Zhang, Z.-Y., Application of hydrogen/deuterium exchange mass spectrometry to study protein tyrosine phosphatase dynamics, ligand binding, and substrate specificity. *Methods San Diego Calif* 2007, 42, 227–233.
- [112] Cravello, L., ETUDES STRUCTURALES DES PROTEINES PAR SPECTROMETRIE DE MASSE COUPLEE AUX ECHANGES HYDROGENE/DEUTERIUM ET A LA RETICULATION CHIMIQUE. Thèse Université Joseph-Fourier - Grenoble I, 2005.
- [113] Katta, V., Chait, B.T., Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom. RCM* 1991, 5, 214–217.
- [114] Zhang, Z., Smith, D.L., Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci.* 1993, 2, 522–531.
- [115] Xiao, H., Kaltashov, I.A., Eyles, S.J., Indirect assessment of small hydrophobic ligand binding to a model protein using a combination of ESI MS and HDX/ESI MS. *J. Am. Soc. Mass Spectrom.* 2003, 14, 506–515.
- [116] Yan, X., Broderick, D., Leid, M.E., Schimerlik, M.I., Deinzer, M.L., Dynamics and Ligand-Induced Solvent Accessibility Changes in Human Retinoid X Receptor Homodimer Determined by Hydrogen Deuterium Exchange and Mass Spectrometry†. *Biochemistry (Mosc.)* 2004, 43, 909–917.
- [117] Garcia, R.A., Pantazatos, D., Villarreal, F.J., Hydrogen/Deuterium Exchange Mass Spectrometry for Investigating Protein-Ligand Interactions. *ASSAY Drug Dev. Technol.* 2004, 2, 81–91.
- [118] Huzil, J.T., Chik, J.K., Slysz, G.W., Freedman, H., et al., A Unique Mode of Microtubule Stabilization Induced by Peloruside A. *J. Mol. Biol.* 2008, 378, 1016–1030.
- [119] Houde, D., Arndt, J., Domeier, W., Berkowitz, S., Engen, J.R., Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* 2009, 81, 5966–5966.

- [120] Betts, G.N., van der Geer, P., Komives, E.A., Structural and Functional Consequences of Tyrosine Phosphorylation in the LRP1 Cytoplasmic Domain. *J. Biol. Chem.* 2008, 283, 15656–15664.
- [121] Mandell, J.G., Baerga-Ortiz, A., Falick, A.M., Komives, E.A., Measurement of solvent accessibility at protein-protein interfaces. *Methods Mol. Biol. Clifton NJ* 2005, 305, 65–80.
- [122] Eyles, S.J., Kaltashov, I.A., Methods to study protein dynamics and folding by mass spectrometry. *Methods* 2004, 34, 88–99.
- [123] Pan, J., Rintala-Dempsey, A.C., Li, Y., Shaw, G.S., Konermann, L., Folding Kinetics of the S100A11 Protein Dimer Studied by Time-Resolved Electrospray Mass Spectrometry and Pulsed Hydrogen–Deuterium Exchange†. *Biochemistry (Mosc.)* 2006, 45, 3005–3013.
- [124] Pantazatos, D., Kim, J.S., Klock, H.E., Stevens, R.C., et al., Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc. Natl. Acad. Sci. U. S. A.* 2004, 101, 751–756.
- [125] Suckau, D., Mak, M., Przybylski, M., Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc. Natl. Acad. Sci. U. S. A.* 1992, 89, 5630–5634.
- [126] Borch, J., Jørgensen, T.J., Roepstorff, P., Mass spectrometric analysis of protein interactions. *Curr. Opin. Chem. Biol.* 2005, 9, 509–516.
- [127] Marie, G., Serani, L., Laprévotte, O., Cahuzac, B., et al., Differential chemical labeling of the AlcR DNA-binding domain from *Aspergillus nidulans* versus its complex with a 16-mer DNA target: Identification of an essential tryptophan involved in the recognition and the interaction with the nucleic acid. *Protein Sci.* 2001, 10, 99–107.
- [128] Mendoza, V.L., Vachet, R.W., Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* 2009, 28, 785–815.
- [129] Cohen, S.L., Chait, B.T., Ferré-D'Amaré, A.R., Burley, S.K., Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci.* 1995, 4, 1088–1099.
- [130] Shields, S.J., Oyeyemi, O., Lightstone, F.C., Balhorn, R., Mass spectrometry and non-covalent protein-ligand complexes: confirmation of binding sites and changes in tertiary structure. *J. Am. Soc. Mass Spectrom.* 2003, 14, 460–470.
- [131] Gervasoni, P., Staudenmann, W., James, P., Plückthun, A., Identification of the Binding Surface on β -Lactamase for GroEL by Limited Proteolysis and MALDI-Mass Spectrometry. *Biochemistry (Mosc.)* 1998, 37, 11660–11669.
- [132] Back, J.W., de Jong, L., Muijsers, A.O., de Koster, C.G., Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* 2003, 331, 303–313.
- [133] Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* 2006, 25, 663–682.
- [134] Gingras, A.-C., Gstaiger, M., Raught, B., Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 2007, 8, 645–654.
- [135] Petrotchenko, E.V., Borchers, C.H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* 2010, 29, 862–876.
- [136] Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., et al., High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2000, 97, 5802–5806.
- [137] Singh, P., Panchaud, A., Goodlett, D.R., Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Anal. Chem.* 2010, 82, 2636–2642.

Chapitre 2 : Stratégie complémentaire voire alternative à l'ESI-MS pour l'étude des complexes biologiques : pontage chimique couplé à la MS

Dans le chapitre 1, nous avons vu qu'outre les techniques d'analyse structurale conventionnelles, l'investigation des édifices biologiques non-covalents bénéficie grandement de la spectrométrie de masse supramoléculaire « classique ». L'ESI-MS est traditionnellement utilisée pour étudier les édifices supramoléculaires car elle offre la possibilité de conserver des interactions non-covalentes. Elle requiert les étapes de dessalage et d'échange de tampon pour éliminer les sels non-volatils incompatibles avec le processus d'ionisation électrospray. Cependant, la difficulté à transférer intacts certains complexes protéiques dans les tampons classiquement utilisés en ESI-MS nous incite aujourd'hui à développer de nouveaux protocoles d'analyse pour permettre leur étude par MS.

L'une de ces approches utilise le pontage chimique pour lier de façon covalente les sous-unités des complexes en amont de l'analyse en MS. Dans le cas des complexes dont la présence des sels non-volatils est nécessaire à la stabilisation, nous nous sommes orientés vers le MALDI-MS qui est plus tolérante vis-à-vis des sels.

Dans ce chapitre 2, avant d'aborder le pontage chimique, nous allons tout d'abord décrire le fonctionnement du MALDI-MS pour comprendre les difficultés d'étude des composés de haute masse par cette technique. Ensuite, nous verrons le principe du pontage chimique, les étapes expérimentales, la mise en œuvre et les critères garantissant une réaction de pontage réussie. Puis nous allons discuter en détail de l'analyse des produits de pontage par MS et du traitement des données par des outils bio-informatiques. Enfin, nous nous consacrerons aux informations structurales que nous pouvons en tirer.

2.1. La spectrométrie de masse par désorption-ionisation laser assistée par matrice (MALDI-MS)

2.1.1. Instrumentation

Le principe et l'acronyme de la spectrométrie de masse par désorption-ionisation laser assistée par matrice (MALDI-MS) ont été introduits en 1985 par Karas *et al.* [1] et ont été suivis par plusieurs applications aux composés biologiques [2-4]. Parmi celles-ci, nous pouvons mentionner la publication des premiers spectres de masse obtenus par MALDI-MS pour les protéines dont les masses excèdent les 10 kDa [5] et 100 kDa [6,7].

Lors d'analyse en MALDI-MS, l'échantillon est d'abord cristallisé avec un large excès de matrice. Pour cela, il est déposé sur une cible qui est un support métallique conducteur et inerte chimiquement. Il existe plusieurs modes de dépôt, les deux plus couramment utilisés est le dépôt goutte séchée (l'analyte et la matrice sont mélangés soit dans un tube soit directement sur cible, le mélange déposé est ensuite laissé évaporer à l'air ambiant pour être cristallisé) et le dépôt couche mince (la matrice est déposée d'abord et laissée sécher sur cible, puis l'analyte est appliqué par-dessus cette couche de matrice et laissé évaporé à l'air ambiant). Le dépôt couche mince est considéré comme permettant de générer des cristaux plus homogènes que le dépôt goutte séchée et donc d'améliorer la résolution et la sensibilité de l'analyse MS [8]. De plus, il offre la possibilité d'automatiser les dépôts d'échantillon, par exemple lors du couplage avec un système de chromatographie liquide ou d'électrophorèse capillaire.

Ensuite, l'irradiation laser sur ce dépôt de mélange analyte-matrice sous un vide poussé (environ 10^{-7} mbar) permet de générer des ions en phase gazeuse et ces ions sont guidés vers l'analyseur grâce à un champ électrique (Figure 1). Ainsi, le rôle de la matrice n'est pas seulement d'absorber l'énergie du faisceau laser afin de protéger l'analyte, mais aussi d'aider à la désorption, au transfert de la phase solide à la phase gazeuse, et à l'ionisation de l'analyte [9]. Il n'existe pas de matrice universelle mais certains types de matrices sont connus pour certaines applications. Par exemple, l'acide 2,5-dihydroxybenzoïque (DHB), l'acide α -cyanohydroxycinnamique (HCCA) et l'acide sinapinique acid (SA) sont les matrices les plus couramment utilisées pour les protéines et les peptides. Pour les acides nucléiques, ce sont plutôt le 6-aza-2-thiothymine (ATT), le 2,4,6-trihydroxyacétophénone (THAP), l'acide 3-hydroxypicolinique (3-HPA), l'acide picolinique (PA) et le DHB [10].

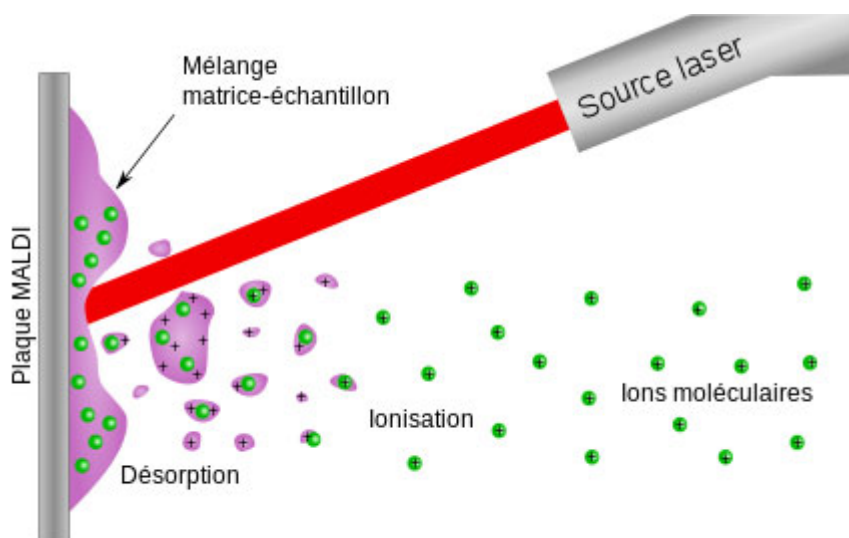


Figure 1. Représentation schématique de la désorption-ionisation MALDI [11].

Bien que l'ionisation MALDI ait été introduit et développé depuis plus d'une vingtaine d'années en spectrométrie de masse, son mécanisme d'ionisation n'est toujours pas complètement résolu à cause de l'implication d'une multitude de processus physiques et chimiques [8,9]. Globalement, le mécanisme d'ionisation introduit par Zenobi et Knochenmuss implique deux étapes [12] :

- L'ionisation primaire se produit pendant ou juste après le pulse laser. La génération des ions primaires reste toujours très controversée, nous pouvons néanmoins en citer deux modèles principaux : l'ionisation photochimique [13] où la formation des cations radicaux de matrice par photoionisation précède la formation d'autres ions; et l'ionisation cluster (ou modèle du « lucky survivor ») [14–16] où l'irradiation laser induit l'éjection des clusters d'ions considérés comme préformés en phase solide. En fonction du modèle considéré, les ions primaires formés peuvent être des ions de matrices, d'analyte, ou des deux.
- L'ionisation secondaire a lieu dès l'apparition des ions primaires. Des réactions de transfert de protons et de cations [12,17], ou des réactions de transfert d'électrons et d'anions [18] dans la plume de désorption/ablation génèrent les ions secondaires et mènent à l'ionisation d'analyte. Cette ionisation secondaire produit majoritairement des ions monochargés.

Les ions formés progressent ensuite dans le champ électrique appliqué entre la cible et la lentille électrostatique vers l'analyseur. Différents types d'analyseur peuvent être couplés à l'ionisation MALDI, nous pouvons mentionner par exemple l'analyseur résonance ionique cyclotronique à transformée de Fourier [19], l'analyseur piège à ions [20,21] ou l'analyseur à secteur magnétique [22]. Cependant, l'analyseur à temps de vol (TOF pour « Time Of Flight ») est le plus souvent utilisé avec l'ionisation MALDI. Il mesure les ions en fonction de leur temps de parcours dans le tube de vol libre de champ. Les ions de plus faibles rapports m/z vont arriver en premier, tandis que ceux qui ont les m/z plus grands vont mettre plus de temps pour arriver au détecteur. Comme développé dans le chapitre 1, la gamme de masse d'analyseur TOF est en théorie illimitée. Ainsi, il est particulièrement adéquat pour l'ionisation MALDI qui génère

principalement les ions monochargés donc de grands rapports m/z . En outre, dans ce type d'analyseur, les ions sont accumulés puis envoyés par paquets dans le tube de vol. Ainsi cet analyseur est bien adapté à l'ionisation pulsée du MALDI. Une fois le tube de vol traversé, les ions arrivent au détecteur où les courants ioniques sont convertis en signal électrique menant au spectre de masse. Comme pour les appareils ESI-MS, le détecteur classiquement utilisé en MALDI-MS est le détecteur MCP.

Enfin, un spectromètre de masse MALDI-TOF peut fonctionner selon deux modes:

- Mode linéaire : les ions sont expulsés par paquet et traversent « linéairement » le tube de vol jusqu'au détecteur. Cependant, de la dispersion spatiale et la différence dans l'énergie cinétique acquise par les ions de même m/z en résulte une faible résolution.
- Mode réflectron (Figure 2): le réflecteur électrostatique ralentit, réfléchit puis accélère les ions vers le détecteur. Cet ajustement en temps de parcours permet de corriger les dispersions spatiales et cinétiques des ions de même rapport m/z améliorant la résolution. Néanmoins, ce mode de fonctionnement ne peut s'appliquer qu'à des molécules de moins de 10 000 Da.

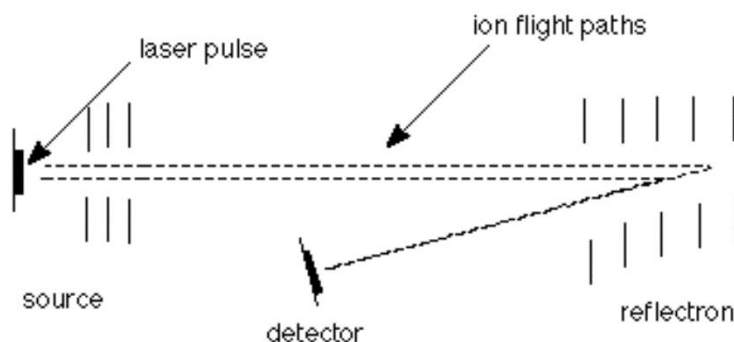


Figure 2. Représentation schématique du mode réflectron du MALDI-TOF (source : adapté de University of Wisconsin – Madison).

2.1.2. Difficultés d'étude des composés de haute masse moléculaire

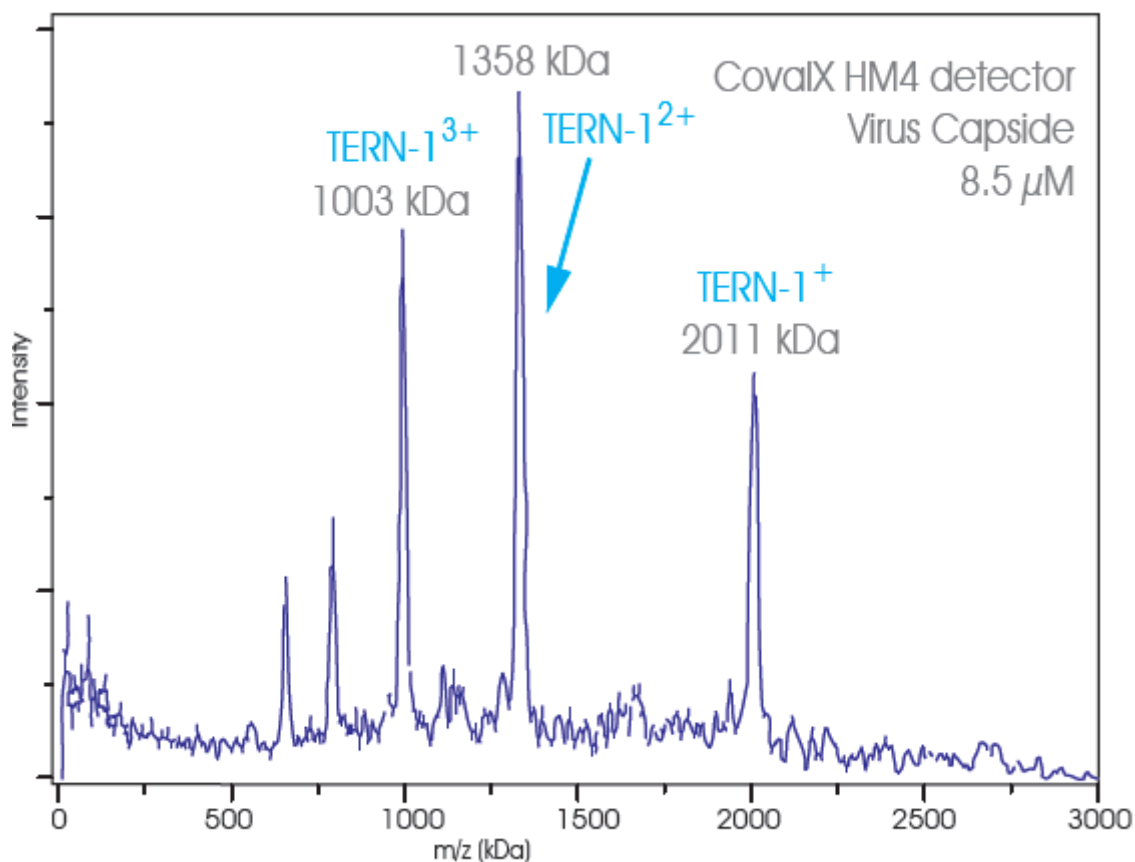
Bien que le MALDI soit une source d'ionisation douce, l'étude des édifices supramoléculaires par MALDI-MS reste en faible nombre par rapport à l'ESI-MS à cause de deux difficultés majeures : i) la détection des ions à haut rapport m/z , ii) la conservation de l'intégrité des complexes non-covalents lors de l'analyse dans le spectromètre de masse.

2.1.2.1. La détection des ions à haut rapport m/z

Comme mentionné précédemment, le détecteur MCP est le détecteur classiquement utilisé en MALDI-TOF. Cependant, la vitesse des ions de grand rapport m/z est plus faible, ce qui mène à une faible efficacité de conversion de ces ions en électrons. Dès lors, le détecteur MCP souffre d'un problème de sensibilité dans la gamme de masse élevée.

Pour pallier ce problème, l'entreprise Comet AG (Flamatt, Suisse) a développé un cryodétecteur permettant de détecter les ions de haute masse avec une bonne sensibilité [23]. Cependant, ce détecteur demande beaucoup d'entretien coûteux et chronophage, et n'est plus commercialisé [10,24]. Le seul type de détecteur de haute masse disponible commercialement est basé sur la technologie « ion conversion dynode » (ICD) dont fait partie celui développé par l'entreprise CovalX AG (Zürich, Suisse). Les ions de haute masse arrivant au détecteur CovalX percutent la dynode de conversion et sont convertis en ions secondaires. Ensuite, ces ions secondaires sont ré-accélérés et détectés par le multiplicateur « secondary electron multiplier » (SEM) [25].

Ce dispositif possède une plus grande sensibilité vis-à-vis des ions de haute masse et permet l'analyse des composés jusqu'aux alentours de 2 MDa avec la HM4, la génération la plus récente du détecteur (Figure 3). Sur le détecteur HM1 disponible au laboratoire, nous avons détecté des espèces jusqu'aux alentours de 800 kDa.



En pratique et dans un appareil de type MALDI-TOF, l'installation de ce détecteur ne demande pas beaucoup de modifications particulières dans la configuration existante du spectromètre de masse conventionnel. Lors de son utilisation, ce détecteur s'insère par les commandes de contrôle devant le détecteur MCP classique sans que le vide soit cassé [25]. De cette manière, nous pouvons continuer à utiliser le détecteur MCP pour les ions de faible m/z (en mode linéaire ou réflectron) tout en ayant la possibilité d'utiliser le détecteur CovalX (uniquement en mode linéaire) pour les ions de grand m/z , le tout sur le même spectromètre de masse à ionisation MALDI.

2.1.2.2. La détection des complexes non-covalents par MALDI-MS

Même si le MALDI-MS permet l'analyse des composés de grandes masses grâce au développement instrumental, un problème subsiste quant à la détection des complexes non-covalents par cette technique. En effet, des interactions non-covalentes peuvent être dissociées lors des étapes de la préparation d'échantillon et de la désorption-ionisation.

D'une part, les matrices classiquement utilisées et mélangées avec l'échantillon lors du dépôt sur cible sont des acides organiques dissous dans un mélange d'eau/solvant organique acidifié avec du TFA (acide trifluoroacétique). Dans ces conditions, la plupart voire la totalité des complexes non-covalents sont dissociés [8]. Pour mieux préserver les interactions non-covalentes, les matrices non-acides ont été introduites dans les années 1990. On peut mentionner entre autres le 6-aza-2-thiothymine [26], le 2,6-dihydroxyacétophénone [27], le 2,4,6-trihydroxyacétophénone [28] ou le 4-nitroaniline [29]. D'autres matrices légèrement acides comme le 6-azathiothymine (ATT) ont été également utilisées pour analyser les complexes non-covalents d'acides nucléiques [30]. Cependant, peu d'exemples d'étude utilisent ces matrices non-acides, probablement à cause d'un problème de sensibilité : elles donnent des spectres de moindre qualité comparés aux matrices acides [10]. D'autres stratégies consistent à continuer l'utilisation des matrices acides, mais à faire l'impasse sur l'utilisation de solvant organique, et/ou de neutraliser le pH de la solution de matrice ou du mélange matrice/analyte. Par exemple, F. Song [31] a pu analyser avec succès les complexes de type protéine-protéine en utilisant une solution aqueuse d'acide sinapinique (SA) dans du citrate d'ammonium ajusté à pH 7 avec de l'hydroxyde d'ammonium.

D'autre part, lors de la désorption-ionisation en MALDI, le signal correspondant aux complexes non-covalents intacts peut être observé dans des conditions particulières lors du premier tir laser. Ce phénomène s'appelle le « first shot phenomenon » et a été mentionné par Rosinke *et al.* en 1995 [32]. Ensuite, trois modèles explicatifs ont été proposés par Cohen *et al.* en 1997 [33], ils sont représentés sur la Figure 4 :

- (A) Lors de la croissance cristalline, les molécules de grandes tailles « ségrèguent » vers la surface des cristaux. Ainsi les complexes se retrouvent majoritairement à la surface, contrairement à des sous-unités de plus petite taille.
- (B) Lors du processus de cristallisation, l'échantillon est distribué de manière homogène dans les cristaux. Cependant, les complexes de protéine qui restent intacts

sont localisés uniquement à la surface des cristaux. A l'intérieur de ces cristaux, les complexes sont présents sous forme dissociée.

- (C) Le complexe survit à la cristallisation. Le pulse laser ablate les molécules de complexe à la surface des cristaux qui seront ionisées et détectées. Cependant, à cause de son irradiation, les autres molécules de complexe restantes autour du cratère sont dissociées. Par conséquent, les tirs laser qui suivent ne résultent qu'en la détection des sous-unités dissociés.

Même si reporté dans quelques publications [32–35], le « first shot phenomenon » reste un phénomène particulier qui dépend de plusieurs facteurs (matrice, préparation d'échantillon, analyte) et reste pour l'instant loin d'être général ou généralisable [35].

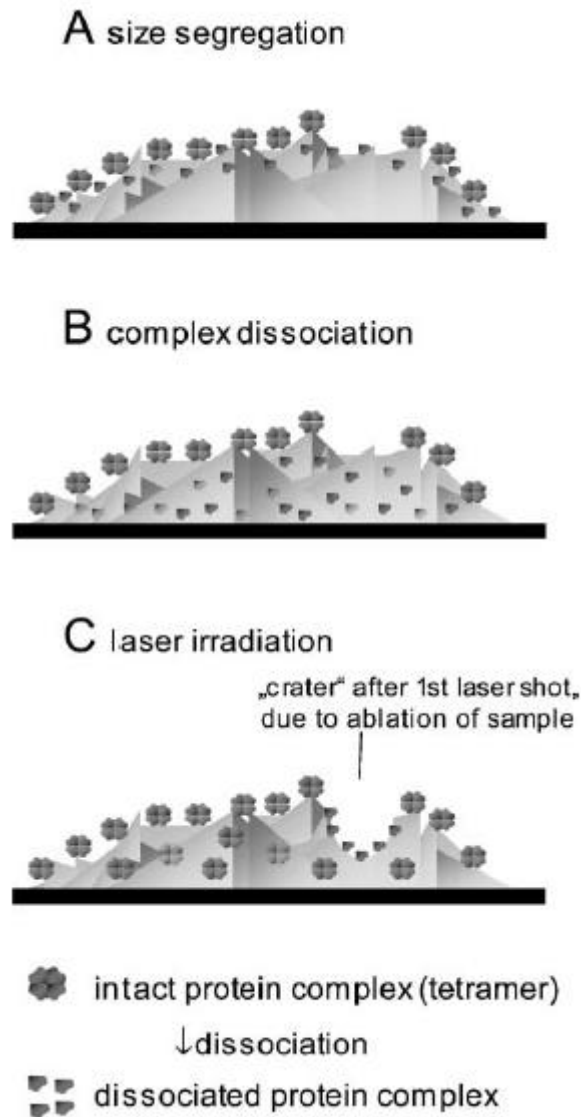


Figure 4. Les hypothèses du "first shot phenomenon" [35].

Pour résumer, même si le MALDI-MS permet l'analyse des complexes non-covalents, il faut tout de même opérer dans des conditions particulières de pH, de solvant, de matrice et du tir laser,

qui sont pour l'instant limités à quelques exemples de systèmes supramoléculaires. Dans ce contexte, la stratégie du pontage chimique mise en place en amont de l'analyse en MS émerge comme une alternative intéressante pour étudier les édifices non-covalents.

2.2. Application du couplage pontage chimique - spectrométrie de masse à l'étude des édifices non-covalents

2.2.1. Principe du pontage chimique couplé à la spectrométrie de masse (CX-MS)

Les modifications chimiques de protéines, dont fait partie le pontage chimique, datent des années 1920 [36,37], mais le développement de ce dernier ne commença qu'à partir des années 70 [38,39]. Le pontage chimique ou réticulation chimique (« cross-link ») consiste à lier de façon covalente les sous-unités qui sont en interaction non-covalente ou qui sont proches par l'intermédiaire de l'agent de pontage (ou l'agent réticulant, « cross-linker »). Le pontage peut concerner des protéines, des peptides, des acides nucléiques *etc.* Puis à partir des années 90 et avec les développements de la MS, la stratégie de coupler le pontage chimique avec la spectrométrie de masse émerge [40]. 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 CX-MS s'intègre dans l'éventail des approches d'analyse structurale. Cependant, son application restait encore très limitée à cause du faible rendement de pontage, du manque de sensibilité des instruments et des difficultés dans l'interprétation des données.

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 [41], l'essor de cette approche s'illustre par la publication de revues toujours plus nombreuses [40–45]. En juillet 2015, 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 CX-MS au fil des années (Figure 5). Parmi celles-ci, nous pouvons citer quelques exemples d'application du CX-MS sur de larges édifices supramoléculaires. Les architectures de l'ARN polymérase II de masse 670 kDa contenant 15 sous-unités, et de l'ARN polymérase I de masse 589 kDa contenant 14 sous-unités ont été étudiées avec succès par les équipes de Rappsilber et d'Aebersold [46,47]. Le CX-MS a permis de déterminer la structure du ribosome d'*E. coli*, un complexe de 53 protéines et de masse 2,5 MDa [48], ainsi que celles des protéasomes 19S et 26S [49–51]. Particulièrement, Lasker *et al.* [47] ont appliqué une approche intégrative où différentes données, y compris celles obtenues par CX-MS, ont été utilisées pour déterminer l'architecture moléculaire du protéasome 26S du *Schizosaccharomyces pombe* faisant 2,5 MDa. Un autre exemple concerne le complexe chaperonine TRiC/CCT (TCP-1 ring complex, chaperonin containing TCP-1) de 1 MDa, dont l'arrangement des sous-unités a été révélé par CX-MS et modélisation [52,53].

Néanmoins, il n'existe pas de protocole « standard » pouvant être transféré d'un système biologique à l'autre, et nous allons voir que des expériences de contrôle sont indispensables.

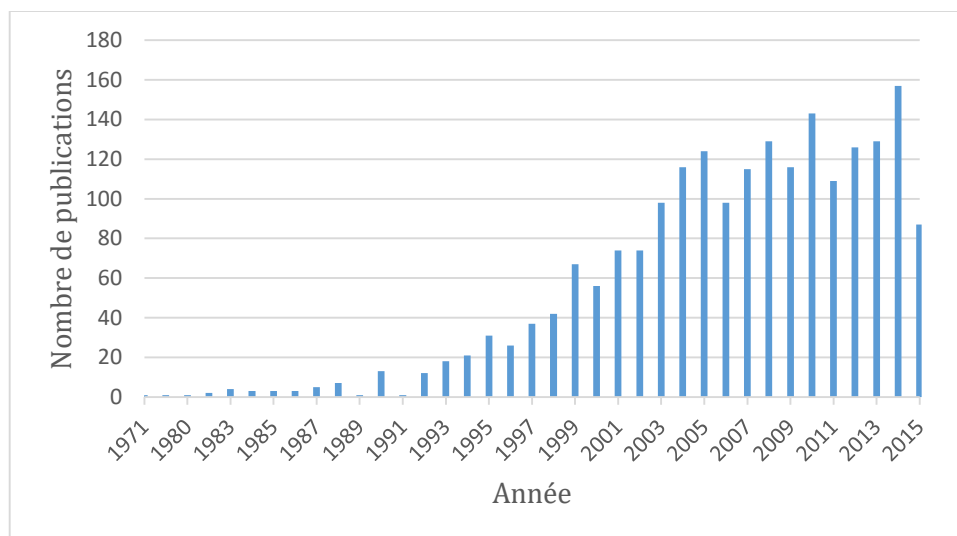


Figure 5. Nombre d'articles scientifiques publiés chaque année sur le pontage chimique couplé à la MS répertoriés dans la base PubMed (recherche effectuée en juillet 2015, mots-clés « cross-linking mass spectrometry »).

L'approche CX-MS peut se résumer en trois étapes majeures [54] représentées en Figure 6:

- La réaction de pontage chimique
- La production de peptides pontés qui seront analysés en MS
- L'identification des peptides pontés par des logiciels bio-informatiques

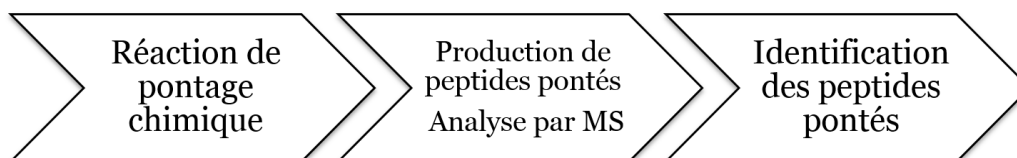


Figure 6. Représentation schématique des étapes de CX-MS.

2.2.2. Mise en œuvre de la réaction de pontage

2.2.2.1. Les critères de choix d'agent de pontage

Les agents de pontage (ou « cross-linkers ») sont constitués de deux fonctions réactives. Les fonctions réactives de ces agents sont séparées par une longueur de bras espaceurs. Ces fonctions réactives peuvent être identiques : on parle d'agent de pontage homobifonctionnel. Ils

vont créer des liaisons covalentes entre deux groupements fonctionnels identiques. Dans le cas où les fonctions réactives sont différentes, les agents de pontage sont dits hétérobifonctionnels. Ils vont lier de façon covalente les groupements fonctionnels différents des protéines, par exemple entre une amine et un thiol. Optionnellement, une troisième fonction réactive peut être introduite sur le bras espaceur (Figure 7). Il permet de créer un pontage supplémentaire avec une troisième protéine, ou de servir à l'isolation, la détection, ou l'identification des peptides pontés [40].

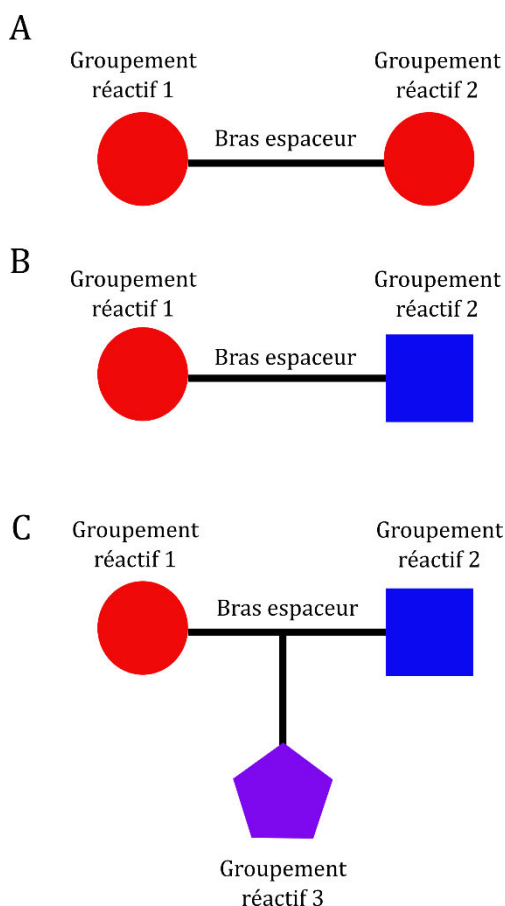


Figure 7. Différents types d'agent de pontage: (A) Agent de pontage homobifonctionnel, (B) Agent de pontage hétérobifonctionnel, (C) Agent de pontage trifonctionnel.

Les exigences concernant un agent de pontage sont nombreuses [40]:

- Sa réactivité : un agent de pontage doit être suffisamment réactif pour induire une certaine quantité de pontages.
- Sa sélectivité : un agent de pontage doit présenter une spécificité chimique et réagir sur des acides aminés définis. En effet, trop de ponts peuvent déformer la conformation native de la protéine et peuvent être difficiles à interpréter *a posteriori*.
- Ses conditions de réaction : un agent de pontage doit pouvoir réagir sous des conditions douces pour ne pas dénaturer les échantillons biologiques.

Il existe des centaines d'agents de pontages disponibles commercialement et par synthèse dans des laboratoires de recherche ([55–58], Thermo Scientific Pierce Protein Biology www.lifetechnologies.com). Le Tableau 1 présente de manière non-exhaustive quelques-uns de ces types d'agents de pontage.

Tableau 1. Liste des principaux types d'agent de pontage et leurs acides aminés cibles, adapté de [40,41,59].

Fonctions réactives	Acides aminés cibles
Haloacétyl	Cystéine
Alpha-haloacétyl	Cystéine
Maléimide	Cystéine
Thiopyridyl disulfide	Cystéine
Acryl	Cystéine
Aldéhyde	Lysine, N-terminal
Iminoester	Lysine, N-terminal
Imidoester	Lysine, N-terminal
Ester activé (succinimidyl, nitrophenyl, sulfotetrafluorophenyl)	Lysine, N-terminal
Ester HOSu, ester phénolique	Lysine, N-terminal
Glyoxal	Arginine
Dicarbonyl	Arginine, lysine
Carbodiimide	Acide aspartique, acide glutamique
Dihydrazide	Acide aspartique, acide glutamique
Phénylazide (photoréactif)	Non sélectif
Tétrafluorophénylazide (photoréactif)	Non sélectif
Benzophénone (photoréactif)	Non sélectif
Arylazide, diazirine (photoréactif)	Non sélectif

Comme chaque type d'agents de pontage réagit sur des acides aminés différents, les agents de pontage sont à choisir en fonction des acides aminés cibles et de leur présence dans la séquence des protéines d'intérêt. Dans le cadre des protéines hydrosolubles, les principaux groupements cibles sont des amines primaires (des lysines et des acides aminés N-ter) et des thiols (des cystéines) [41]. Le choix des agents de pontage porte également sur le bras espaceur dont la longueur régit la distance maximale entre deux groupements fonctionnels. Enfin, il faut garder à l'esprit que la réaction de pontage étant protéine- et complexe-dépendante, sa mise en œuvre constitue un processus d'essai-erreur [42,43], afin de trouver la nature d'agents de pontage adéquate pour une réaction de pontage réussie.

2.2.2.2. Les agents de pontage réagissant sur les amines

Les groupements fonctionnels les plus communément ciblés sont les amines des lysines et du N-terminal. Ceci peut s'expliquer par plusieurs facteurs [40,60] : i) leur haute réactivité, ii) leur emplacement souvent à l'extérieur des chaînes peptidiques ou près de la surface des protéines hydrosolubles, iii) leur présence à une fréquence assez élevée sur la séquence peptidique, ce qui permet d'avoir de nombreux points de réactions répartis de façon assez homogène.

L'acylation et l'alkylation sont les deux mécanismes réactionnels mis en jeu lors du pontage des amines [43]. Ils donnent des liaisons amides ou imides stables. Nous pouvons mentionner quelques fonctions réactives qui ciblent les amines :

- Les esters N-hydroxysuccinimide (NHS) : introduits il y a plus d'une trentaine d'années, ces esters NHS mettent en jeu les réactions d'acylation. Cependant, la plupart d'entre eux ne sont pas solubles en milieu aqueux des biomolécules. Pour y remédier, soit nous utilisons une solution mère très concentrée en ester NHS dans un solvant organique pour ensuite la diluer dans le milieu réactionnel, soit nous utilisons leurs analogues hydrosolubles qui sont les esters sulfo-NHS.

Lors de l'acétylation, les esters NHS réagissent sur les amines ϵ des lysines et les amines α du N-terminal, forment une liaison amide ou imide stable et libèrent le groupement NHS ou sulfo-NHS (Figure 8). Malgré leur grande spécificité vis-à-vis des amines, les esters NHS peuvent aussi réagir sur d'autres groupements nucléophiles. Cependant, la réactivité des serines est très lente comparée à celle plus rapide des amines, et les produits de la réaction sur les histidines (N-acylimidazole) sont très instables [61]. Quant aux cystéines, deux cas se présentent. Soit elles ne peuvent pas réagir avec les esters NHS car impliquées dans les ponts disulfure. Soit elles réagissent mais les produits obtenus (les thioesters) sont instables et peuvent réagir avec les amines pour former l'amide [40]. Ensuite, un pH alcalin (pH 8,4) favorise les ponts entre les lysines, tandis qu'un pH plutôt acide favorise réactions des esters NHS sur les tyrosines [61,62]. Les réactions des esters NHS sont aussi observées avec les sulfhydryls déprotonés des cystéines, les hydroxyls déprotonés des serines et thréonine [63,64], ou les groupements guanidine des arginines à un pH plus élevé [65].

En résumé, les esters NHS peuvent être considérés comme spécifiques aux amines des lysines et du N-terminal à condition que le pH soit contrôlé avec précaution utilisant les tampons phosphates, bicarbonates, HEPES ou borates [40]. Cependant, il est important de garder à l'esprit que lorsque les résidus acides aminés (autres que les lysines) cités ci-dessus sont particulièrement proches l'un de l'autre, le pontage peut avoir lieu même si ce n'est pas la réaction la plus favorable. Ceci peut être particulièrement utile et devrait être pris en compte en fonction du complexe d'intérêt. En effet, les pontages créés entre des résidus autres que les lysines peuvent apporter des informations structurales complémentaires, surtout dans des régions du complexe où peu de lysines sont présentes.

Enfin, l'hydrolyse des esters NHS les rend inactifs et augmente avec le pH [56,66]. Leur temps de demi-vie est de l'ordre de l'heure dans un milieu aqueux à un pH physiologique (pH 7.0-7.5) [43]. Par conséquent, des précautions sont à prendre lors de leur conservation (en minimisant les contacts avec l'humidité environnant) et de leur utilisation (pour les esters sulfo-NHS, en les utilisant tout de suite après dilution dans un

solvant aqueux). Une approche alternative consiste à utiliser les agents de pontage possédant des fonctions réactives N-succinimidyl carbamate qui sont moins sujets à l'hydrolyse que les esters NHS [58].

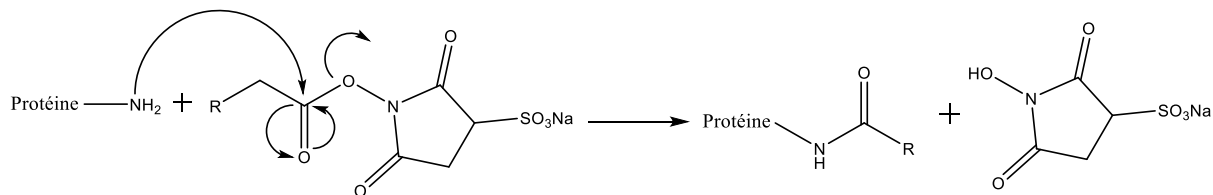


Figure 8. Réaction de pontage des esters NHS sur des protéines.

- Les esters imides : les esters imides font partie des agents de pontage les plus anciens [67] et les plus spécifiques pour modifier les amines primaires. Ils possèdent peu de réactivité vis-à-vis des groupes nucléophiles autres que les amines dans la gamme de pH entre 7 et 10. Ils sont hydrosolubles, mais subissent une dégradation due à l'hydrolyse. Leur temps de demi-vie dans un milieu aqueux est inférieur à 30 minutes [68]. L'avantage des esters imides est qu'ils permettent de garder la charge positive sur le produit réactionnel amidine à pH physiologique [43] (Figure 9).

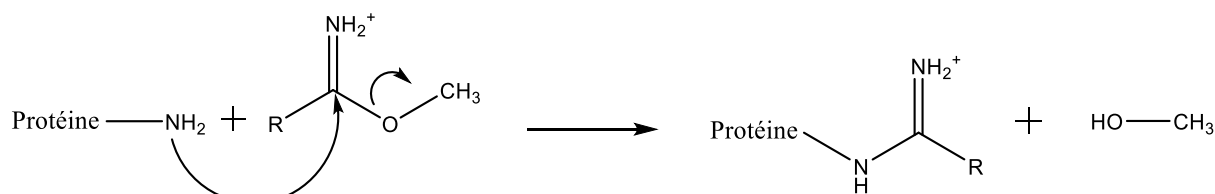


Figure 9. Réaction de pontage des esters imides sur des protéines.

- Les carbodiimides : ils forment des liaisons amides entre les amines et les carboxyliques [69] (Figure 10). L'agent de pontage le plus connu de cette famille est l'EDC (1-ethyl-3-(3-diméthylaminopropyl)carbodiimide). Ces carbodiimides sont également appelés les agents de pontage « de longueur nulle » (« zero-length ») car ils n'introduisent pas de distance supplémentaire entre les partenaires d'interaction après la réaction de pontage. Ainsi ils sont privilégiés pour lier des groupements fonctionnels proches (< 3 Å).

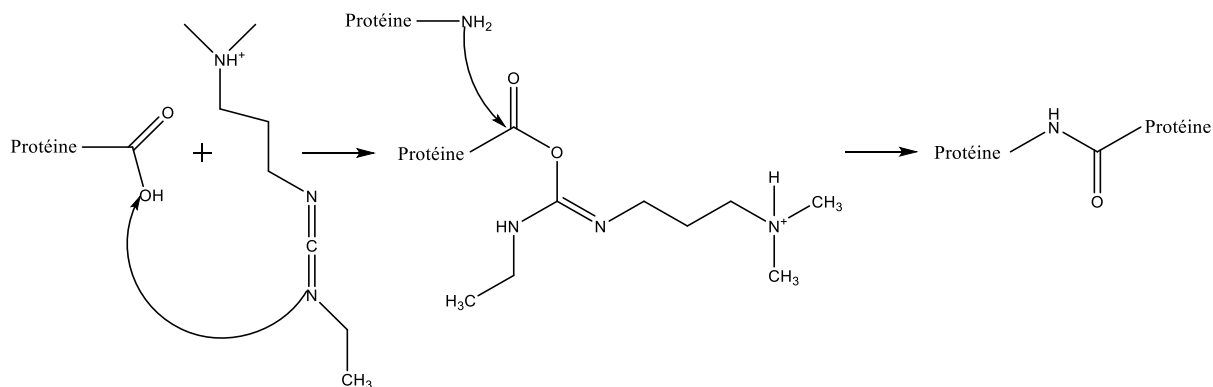


Figure 10. Réaction de pontage de l'EDC sur des protéines.

2.2.2.3. Les agents de pontage réagissant sur les thiols

Les groupes réactifs envers les thiols des cystéines constituent le deuxième groupe d'agents de pontage des plus couramment utilisés. L'alkylation et l'échange de ponts disulfures en sont les mécanismes réactionnels [43].

- Les maléimides : ils ajoutent un thiolate sur le groupement fonctionnel selon une addition de Michaël de type 1,4 [43] (Figure 11). Ils sont spécifiques aux thiols dans la gamme de pH de 6,5 à 7,5 [70,71]. Ils peuvent aussi réagir sur des amines [72]; cette réaction est cependant 1000 fois plus lente que la réaction sur les thiols à pH 7 [73].

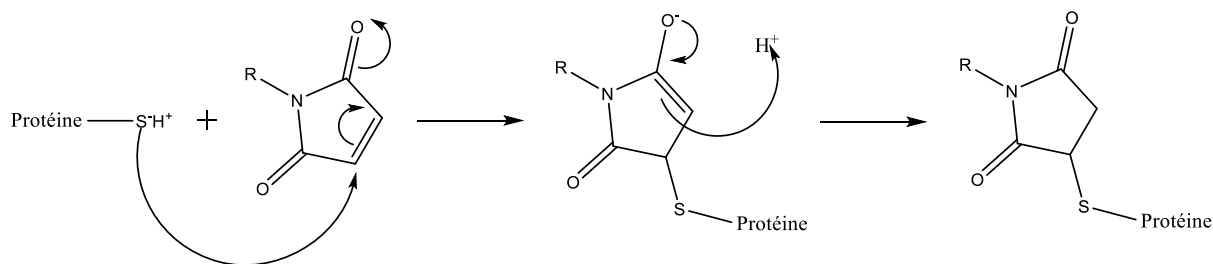


Figure 11. Réaction de pontage des maléimides sur des protéines.

- Les réactifs permettant l'échange de ponts disulfures : les cystéines impliquées dans des ponts disulfures peuvent participer à une réaction d'échange (Figure 12). Ces agents de pontage vont réagir sur la fonction thiol sur le pont disulfure, ce qui mène à une rupture de la liaison S-S et la formation d'un nouveau pont. Des agents réducteurs des thiols comme le DTT ou le 2-mercaptoéthanol peuvent aussi être utilisés en amont pour faciliter la réaction de pontage [40]. Ces échanges au niveau des soufres ont lieu dans une large gamme de conditions, de pH acide à pH basique [43].

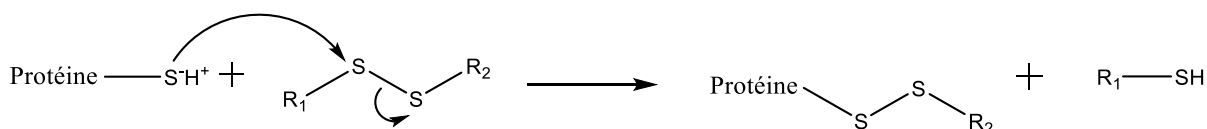


Figure 12. Réaction de pontage permettant l'échange de ponts disulfure avec les cystéines.

2.2.2.4. Les agents de pontage photoréactifs (non-sélectifs)

Les agents de pontage photoréactifs réagissent sur les groupements fonctionnels lors de l'exposition à la lumière UV. 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 [74,75]. Les agents les plus utilisés sont les arylazides (Figure 13). 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 sur des doubles liaisons au niveau des liaisons C-H ou N-H [43]. On peut aussi mentionner d'autres agents de pontage photoréactifs comme la benzophénone et ses dérivés [74,76].

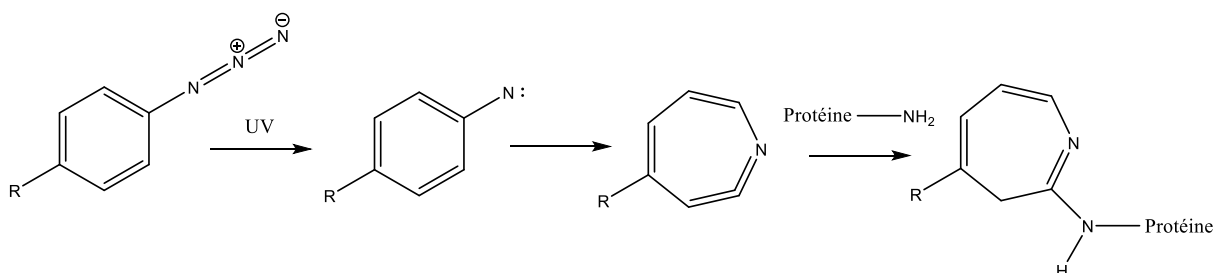


Figure 13. Réaction de pontage des arylazides photoréactifs sur des protéines.

2.2.2.5. Les longueurs du bras espaceur

Comme mentionné précédemment, le choix d'agent de pontage réside non seulement dans la nature de ses fonctions réactives mais aussi dans la longueur du bras espaceur. Cette chaîne carbonée sépare des deux extrémités réactives et est introduite lors du pontage entre deux groupements fonctionnels pontés. En s'ajoutant à la flexibilité des groupements fonctionnels, la longueur du bras espaceur détermine et donne une première approximation de la distance des ponts [44]. Plus cette longueur est importante, plus des groupements fonctionnels qui ne sont pas en interaction ou proches pourraient être liés. Ceci crée plus de ponts et entraîne plus d'hétérogénéité dans des espèces pontées. Enfin, un bras espaceur long confère à l'agent de

pontage plus d'encombrement stérique, ce qui gêne son accès au cœur de la protéine ou des complexes.

Outre les agents de pontage de longueur nulle, la plupart des agents de pontage possèdent des bras espaceurs dont la longueur est donnée par les fabricants. Néanmoins, une étude réalisée par Green *et al.* en 2001 a permis de donner les longueurs moyennes des 32 agents de pontage qui s'avèrent d'être souvent inférieures aux longueurs théoriques [77]. Le Tableau 2 présente quelques-unes de ces distances obtenues pour les agents de la famille des esters NHS : disuccinimidyl glutarate (DSG), dithiobis(succinimidylpropionate) (DSP), disuccinimidyl suberate (DSS), disuccinimidyl tartarate (DST), ethyleneglycol bis-(succinimidylsuccinate) (EGS) et bis(2-[succinimidooxycarbonyloxy]ethyl)sulfone (BSOCOES).

Tableau 2. Longueur des agents de pontage de type NHS-ester.
* Voir référence Pierce Chemicals (1999). Adapté de [77].

Agent de pontage	Distance moyenne N-N (Å)	Gamme de distance N-N (Å)	Distance donnée par le fabricant (Å) *
DSG	6,22	3,12 – 7,49	7,7
DSP	8,04	4,63 – 10,65	12
DSS	8,88	5,58 – 11,42	11,4
DST	4,06	2,45 – 5,84	6,4
EGS	9,11	3,23 – 14,79	16,1
BSOCOES	10,62	6,81 – 12,43	13

2.2.2.6. Optimisation de la réaction de pontage

La réaction de pontage doit être conduite dans des solutions compatibles à la préservation structurale des complexes supramoléculaires d'intérêt. De plus, ces tampons ne doivent pas interférer avec les réactions de pontage et ne doivent pas réagir avec des agents de pontage. Par exemple, dans le cas des agents de pontage de type ester NHS, un tampon Tris n'est pas compatible et un échange de tampon s'impose dès lors.

En outre, le pontage chimique peut induire des biais dans la conformation des protéines. En effet, plus les protéines sont modifiées par le pontage, plus leurs structures peuvent s'éloigner de celles en état natif [40]. Par conséquent, il est crucial d'optimiser les conditions de réaction afin de limiter le nombre de modifications induites par la réaction de pontage.

Pour cela, différentes conditions de réaction (concentration, ratio protéine/agent de pontage, temps de réaction, température) doivent être testées. La cinétique de réaction peut être suivie par gel d'électrophorèse SDS-PAGE ou par MALDI-MS [43,78]. Dans le cas du pontage d'un complexe protéique, les conditions où apparaît la formation de multimères non spécifiques doivent être écartées [41]. L'objectif est d'avoir le meilleur rendement possible de pontage mais à la plus faible concentration possible d'agents réactifs. Ceci a pour but d'éviter un pontage excessif conduisant à des perturbations dans la structure des protéines ou des complexes, ces distorsions conformationnelles pouvant induire d'agrégations ou d'associations non-spécifiques [78,79].

2.2.3. Analyse des produits de pontage par MS

2.2.3.1. Production des peptides après pontage

L'étape qui suit la réaction de pontage est la production de peptides pontés. Il existe deux principales approches. Dans l'approche « top-down », juste après la réaction de pontage, le complexe biologique intact subit une fragmentation MS/MS directement dans le spectromètre de masse. Cependant, l'efficacité d'ionisation et de fragmentation étant liée à la taille et à la nature de l'échantillon, cette approche s'avère être difficilement applicable aux complexes protéiques de grande taille [40,41].

La deuxième approche « bottom-up » réside en une digestion chimique ou enzymatique du complexe ponté. La réaction de pontage peut être contrôlée par gel d'électrophorèse ou par MALDI-MS, et la digestion peut être réalisée soit en gel soit en solution. L'équipe d'Andrea Sinz a remarqué que la digestion en solution est beaucoup plus efficace que la digestion en gel [73]. Dans tous les cas, la spécificité d'enzyme utilisée pour la digestion est un autre facteur important. Les enzymes de faible spécificité permettent l'obtention de peptides de plus petite taille préférables pour l'analyse en MS. Cependant, augmenter le nombre de coupures enzymatiques va générer un plus grand nombre de peptides, ce qui complique encore plus le nombre de combinaisons potentiels de séquence et donc l'identification des peptides pontés. C'est pourquoi Petrotchenko et Borchers proposent d'utiliser plutôt un mélange de deux enzymes spécifiques [41].

Enfin, pour bénéficier des avantages et limiter les inconvénients de chacune de ces deux approches, une autre stratégie consiste à combiner les deux (« middle-down »). Le complexe biologique ponté subit une protéolyse ménagée ou un clivage limité. Le mélange peptidique résultant contient des peptides de taille raisonnable adaptés à une analyse « top-down » [80,81].

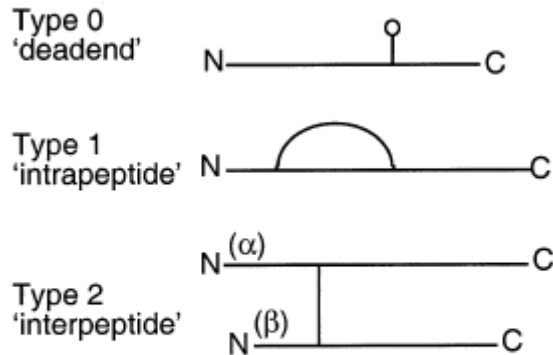
2.2.3.2. Les types de peptide obtenus

Après la réaction de pontage et la digestion enzymatique, outre les peptides natifs (peptides n'ayant pas été modifiés par l'agent de pontage), 3 types de peptides pontés peuvent être obtenus (Figure 14):

- Peptide « mono-link » (ou « dead-end », ou pontage de type 0) : peptide où seule une extrémité de l'agent de pontage a réagi.
- Pontage intra-peptide (« intra-link » ou peptide cyclique, ou pontage de type 1) : les deux extrémités fonctionnelles de l'agent de pontage ont réagi sur deux résidus d'acides aminés d'un même peptide.
- Pontage inter-peptide (« inter-link » ou pontage de type 2) : peptide où les deux extrémités fonctionnelles de l'agent de pontage ont réagi sur deux résidus d'acides aminés de deux peptides différents.

Rappsilber a également mentionné en 2011 [82] les peptides de type 3. Sous cette appellation sont regroupés les multipontages issus des combinaisons de pontages des peptides des types 0, 1 et 2 dont la nomenclature plus détaillée avait été proposée par Schilling *et al.* [83]. Cependant, étant donné leur complexité et la difficulté à les interpréter, nous ne les prenons pas en compte dans le cadre de ce travail.

(a) Single modifications



(b) Multiple modifications

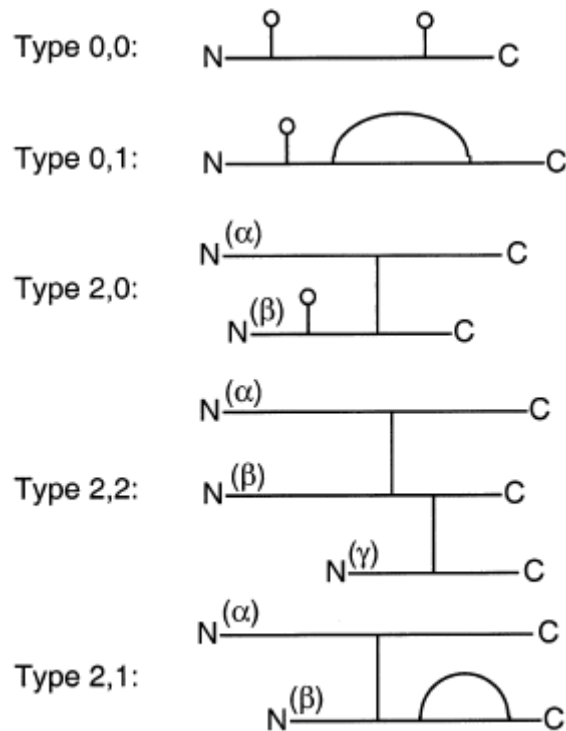


Figure 14. (a) Types de peptides pontés. (b) Dans le cas des modifications multiples, les peptides sont classés selon leurs longueurs et leurs masses avec $\alpha > \beta > \gamma$. La position des peptides dans la séquence de la protéine du N-terminal au C-terminal détermine l'ordre des deux chiffres désignant le type de pontage [83].

2.2.3.3. Amélioration de la détection en MS des peptides pontés

L'analyse du mélange peptidique en MS/MS doit faire face à un problème de détection des peptides pontés. En effet, ces derniers sont présents en faible quantité comparé aux peptides natifs, d'autant plus que la diversité des types de peptides modifiés par le pontage mentionnés précédemment diminue encore plus leur abondance. Beaucoup d'efforts sont consacrés à l'amélioration de cette détection, une des solutions à ce problème consiste à ajouter une étape de séparation ou d'enrichissement des peptides pontés [54,84–86].

En effet, la séparation des produits pontés, tant au niveau protéique qu'au niveau peptidique, peut aider grandement à la détection et l'identification. Avant l'étape de digestion, il est possible d'introduire une étape supplémentaire pour séparer les molécules de complexe ponté des molécules de complexe non ponté. Ceci permet de n'avoir plus tard que des peptides issus uniquement du complexe ponté. Ainsi, il existe des possibilités de séparer des complexes pontés avant digestion par électrophorèse ou par gel-filtration. Ensuite, après l'étape de digestion, les peptides obtenus sont souvent séparés par chromatographie liquide en amont de l'analyse MS. Cette ultime séparation décomplexifie le mélange peptidique et réduit les phénomènes de suppression d'ions lors d'analyse en MS [41].

Une autre stratégie pour ne détecter que les produits pontés consiste à utiliser les techniques d'enrichissement par chromatographie d'échange cation (SCX), chromatographie d'exclusion (SEC) ou par affinité [78]. Les tags d'affinité de type biotine peuvent être incorporés sur l'agent de pontage [87,88]. Après la digestion, une purification d'affinité sur colonne d'avidine permet d'enrichir en produits pontés. L'utilisation des anticorps anti-agent de pontage sont également possibles pour cette étape d'enrichissement [41].

Outre l'enrichissement, une autre possibilité consiste à introduire une signature reconnaissable en masse afin de faciliter la détection des peptides modifiés par pontage. Dans certains cas, cette stratégie rend également possible la distinction entre peptides « mono-link » et peptides pontés.

a. Les marquages isotopiques

Deux types de marquages sont à discriminer : les marquages d'agent de pontage et les marquages de protéine.

- Marquages isotopiques d'agent de pontage : le principe du marquage isotopique consiste à substituer un ou plusieurs atomes de l'agent de pontage par leurs isotopes lourds stables. De par ce principe, les agents de pontage « de longueur nulle » ne sont pas concernés, car ils n'insèrent pas d'identité chimique entre les groupements fonctionnels lors de la création du pont. Le plus souvent, un mélange équimolaire d'agents de pontage lourds et légers est utilisé. Comme les groupements fonctionnels ont autant de chance de réagir avec un agent de pontage léger que son homologue lourd, les signaux des peptides ainsi pontés sont présents sous forme de doublets avec un ratio 1 : 1 (Figure 15). Ceci permet de les distinguer des peptides non modifiés dont les signaux sont présents sous forme de singulets. Dès lors, la sélection des ions

précurseurs des peptides pontés lors de la fragmentation MS/MS est facilitée de même que l'interprétation des spectres de masse [54].

En général ce sont des atomes d'hydrogène ^1H qui sont remplacés par du deutérium ^2D , ou des ^{12}C par des ^{13}C . On peut par exemple citer les agents BSP, BSG et BSS marqués d0-d4 [89], les agents DSA et DMA marqués d0-d8 [90]. Collins *et al.* ont aussi utilisé des agents doublement marqués à l' ^{18}O [91].

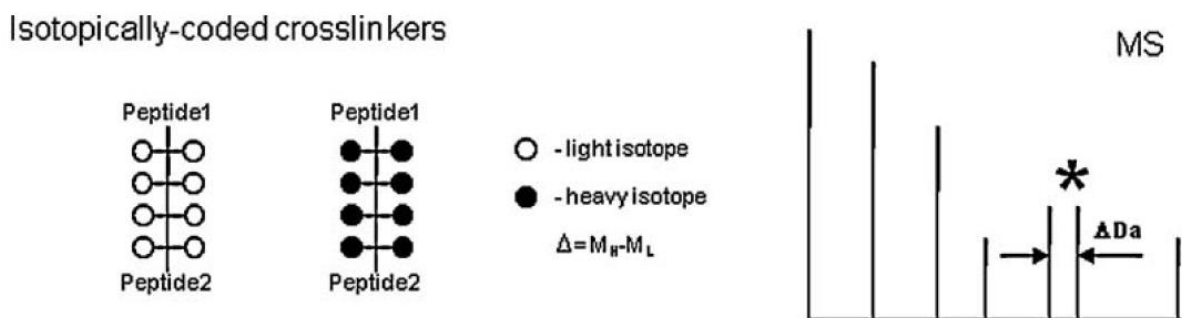


Figure 15. Principe du marquage isotopique des agents de pontage [41].

La quantité d'isotopes lourds introduits au sein de l'agent de pontage conditionne la différence de masse entre les deux pics du doublet. Dans le cas des mélanges d0-d4 et d0-d8, cet écart est respectivement de 4 Da et 8 Da. Idéalement, cette différence doit être ni trop faible pour permettre la séparation des deux pics pour des ions multichargés, ni trop importante pour permettre la sélection de tous les isomères lors d'analyse MS-MS. Enfin, le principal argument contre l'utilisation de cette approche de marquage réside dans le fait que le signal des peptides modifiés est diminué car divisé entre les formes léger et lourd.

- Marquages isotopiques de protéine : au lieu de marquer des agents de pontage, les possibilités de marquer les protéines pendant ou après la digestion enzymatique sont également envisageables. Nous pouvons marquer soit l'extrémité N-terminal soit l'extrémité C-terminal des peptides.

Premièrement, le marquage N-terminal exploite la propriété particulière des ponts inter-peptides par rapport aux autres peptides : la possession de deux extrémités N-terminaux. Avec l'utilisation d'un mélange équimolaire de réactifs lourd et léger, ce marquage mène à l'incorporation de deux groupes marqués isotopiquement, tandis que les autres peptides n'en auront qu'un seul. Les spectres de masse des ponts interpeptidiques contiendront un triplet de pics d'intensité 1 : 2 : 1 (Figure 16). Un inconvénient de cette technique est l'augmentation de faux positifs liés aux modifications possibles des amines ϵ des lysines. Pour y remédier, Chen *et al.* proposent de bloquer les lysines par méthylation avant la digestion enzymatique [92],

Petrotchenko *et al.* proposent de combiner le marquage N-terminal des protéines avec le marquage des agents de pontage [93].

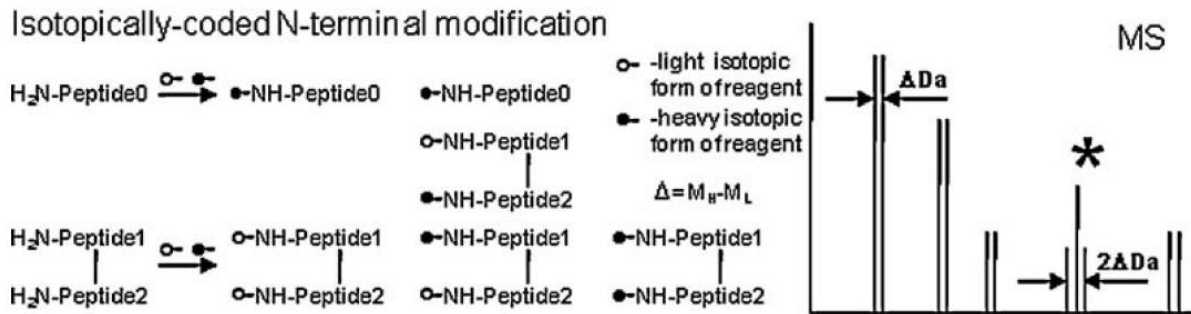


Figure 16. Principe du marquage N-terminal [41].

Deuxièmement, le marquage C-terminal profite du fait que lors de la digestion, les atomes d'oxygène des peptides peuvent s'échanger avec des atomes d'oxygène de l'eau. Avec l'utilisation de l'eau marquée ^{18}O dans le milieu réactionnel, les ponts inter-peptides vont avoir deux fois plus d'oxygènes marqués que les autres types de peptide. Par conséquent, leur signal dans le spectre de masse présentera une différence de masse de 8 Da, tandis que les peptides natifs ou pontés de façon intramoléculaire auront une différence de 4 Da [94,95] (Figure 17).

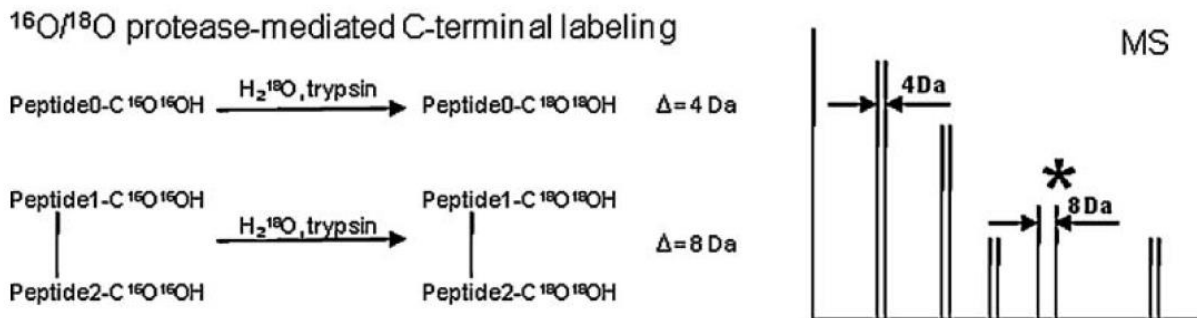


Figure 17. Principe du marquage C-terminal [41].

Enfin, une autre option est de marquer les protéines avec les isotopes stables comme le ^{15}N . Un mélange des formes marquées (H pour « heavy ») et non marquées (L pour « light ») des protéines subit des étapes de pontage et de digestion. Les ponts intermoléculaires auront quatre combinaisons possibles de formes isotopiques : LL, LH, HL et HH, lorsque les autres peptides n'en auront que deux : LL et HH. Ainsi nous

pouvons mieux distinguer les ponts intermoléculaires grâce au triplet ou quadruplet de pics [96].

b. La fluorescence

Outre les marquages isotopiques, les marquages fluorescents sur des agents de pontage ont été également développés. Initialement non fluorescent en solution, l'agent de pontage le devient quand ses deux fonctions réactives ont réagi. Ceci permet la détection par fluorescence des ponts interpeptides en chromatographie liquide en amont de l'analyse en MS [97]. Dans une autre étude, Wine *et al.* ont utilisé l'agent de pontage contenant une fonction réactive de type ester NHS, une fonction photoréactive, et entre les deux un pont disulfure avec un groupement fluorescent [98]. La première réaction de pontage a lieu sur l'amine primaire, suivie par l'activation du groupement photoréactif. Ensuite, une coupure du pont disulfure élimine le marquage fluorescent des protéines qui n'ont pas été pontées. Les produits pontés seront ensuite détectés par imagerie fluorescente sur gel (Figure 18).

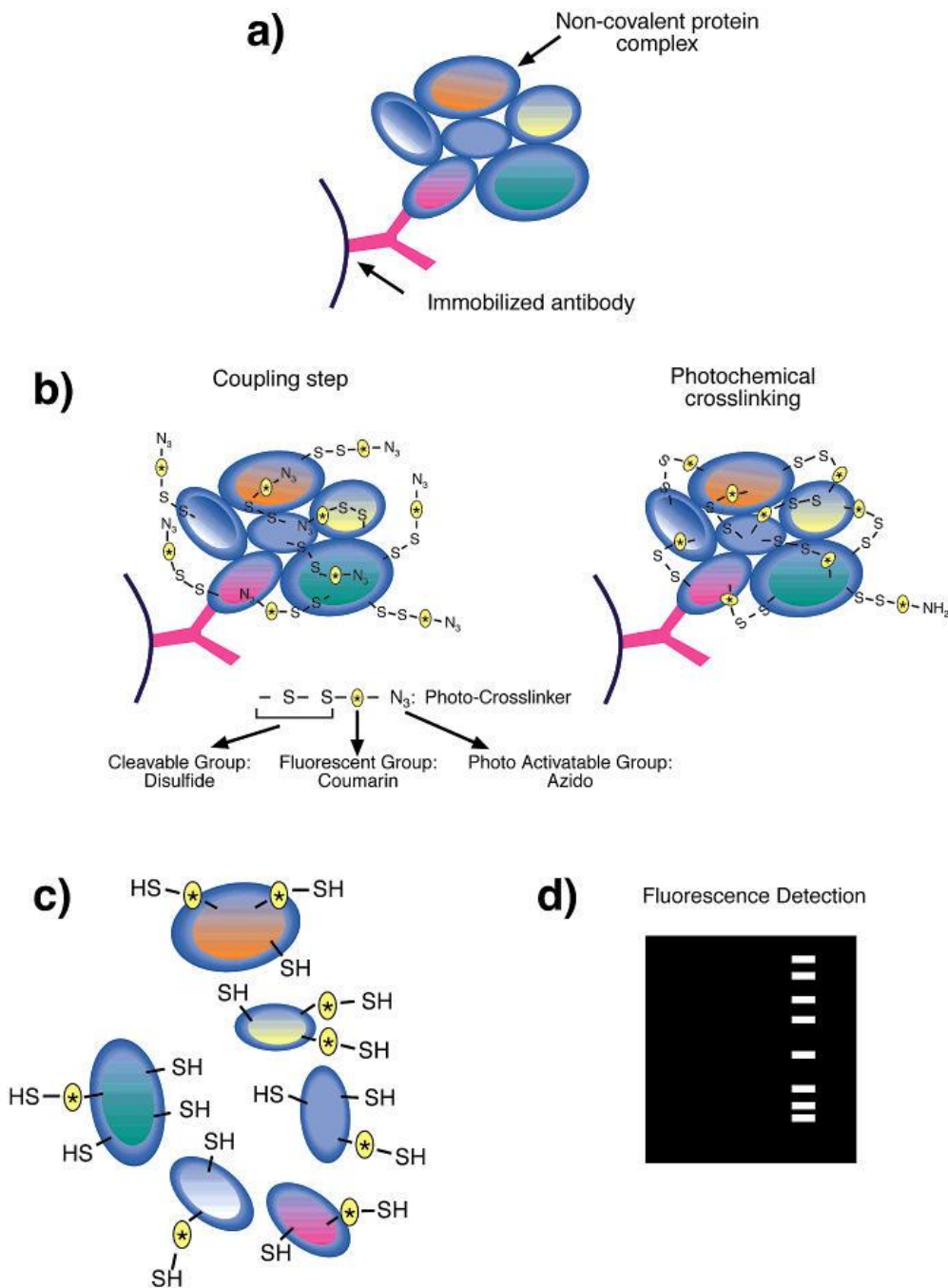


Figure 18. Utilisation des agents de pontage fluorescents par Wine *et al.*
 (a) Co-immunoprécipitation du complexe protéique non-covalent. (b) Pontage par le groupement ester NHS puis par le groupement photoréactif. (c) Réduction des groupements disulfures.
 (d) Séparation du complexe marqué par SDS-PAGE et détection de la fluorescence [98].

c. Les agents de pontages clivables

Le principe de cette fonctionnalisation est de cliver l'agent de pontage. Traditionnellement, l'agent de pontage clivable chimiquement utilisé pour cette approche est le DTSSP (3,3'-dithiobis(sulfosuccinimidyl propionate)) [62,99]. Il contient un pont disulfure pouvant être réduit par du dithiothréitol (DTT). Les peptides modifiés par le pontage auront un changement de masse entre avant et après réduction. De plus, les différences de masse observées avant et après réduction entre les différents types de pontage permettent de distinguer ces derniers entre eux [62] (Figure 19).

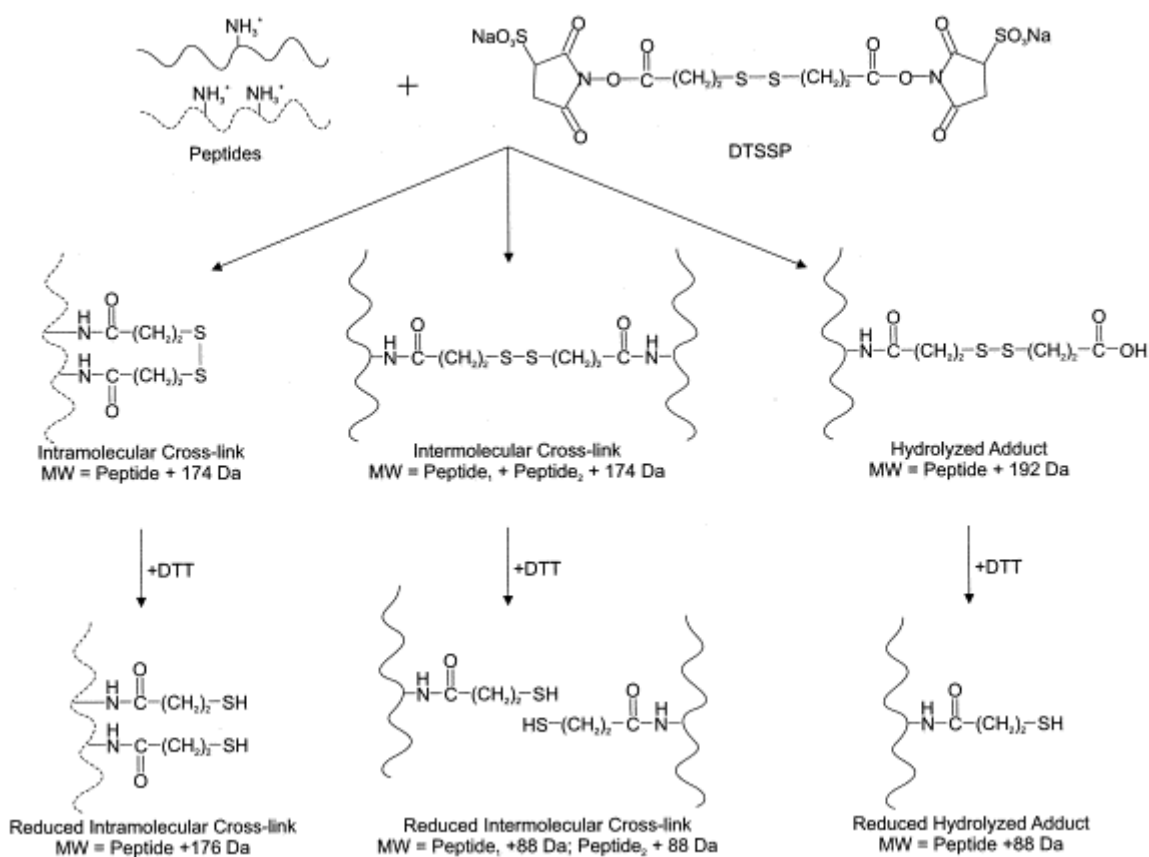


Figure 19. Réaction du DTSSP. La masse d'un pont intramoléculeire est égale à la somme de la masse d'un peptide + 174 Da, celle d'un pont intermoléculeire est égale à la somme des masses des deux peptides + 174 Da. Le « mono-link » a une masse égale à la somme de la masse du peptide + 192 Da. Après réduction par le DTT, l'augmentation de masse est de + 176 Da pour le pont intramoléculeire, et de + 88 Da pour les autres [62].

Il est également possible d'utiliser les agents de pontage photo-clivables [100] ou CID-clivables [101,102]. Par exemple, pour les agents de pontage clivables en CID, les groupes labiles sont introduits dans la structure de l'agent de pontage. Lors de la fragmentation CID, ces groupes

« reporters » produisent les ions « reporter » spécifiques dont le signal sur le spectre de masse permet de mieux distinguer les peptides contenant le pontage (Figure 20).



Figure 20. Principe d'utilisation des ions "reporter" [41].

Enfin, il est à noter que des efforts continuent à être déployés pour concevoir et synthétiser les nouveaux agents de pontage plus efficaces et qui améliorent la qualité des spectres de masse [57,58,103].

2.2.4. Traitement des données obtenues par les logiciels dédiés au pontage

Après l'analyse des peptides par MS, vient l'étape d'identification des produits pontés. Ce traitement de données constituait pendant longtemps le goulet d'étranglement du CX-MS. Ceci est dû principalement à la complexité du mélange peptidique issu de la réaction de pontage, à la grande quantité de données à traiter, ainsi qu'à la multitude de combinaisons théoriques possibles de peptides pontés.

Bien que les logiciels d'identification classiques tels que Mascot et ProteinScape permettent l'introduction des modifications pour l'identification des peptides « linéaires », aucun ne permet d'identifier les séquences des deux peptides liés par le pontage. Dans ce contexte, de nombreux logiciels de recherche dédiés au pontage sont développés au sein des laboratoires de MS [44,45,104]. Le Tableau 3 liste quelques-uns de ces logiciels et leurs liens de téléchargement.

Tableau 3. Aperçu de quelques logiciels dédiés à l'identification des peptides pontés. Adapté de [44,45].

Nom du logiciel	Site web	Référence
BLinks	http://brucelab.gs.washington.edu/BLinks.php	[105]
CLPM	http://search.cpan.org/~crabtree/BioX-CLPM-0.01/lib/BioX/CLPM/Base.pm	[106]
Crux	http://noble.gs.washington.edu/proj/crux	[107]
DXMSMS	http://creativemolecules.com/CM_Software.htm	[108,109]
MassMatrix	http://www.massmatrix.net	[110]
CrossWork/MassAI	http://www.massai.dk/	[111]
MS-Bridge (partie de UCSF Prospector)	http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msbridgestandard	[112]
PeptideMap (partie de PROWL)	http://prowl.rockefeller.edu/prowl/peptidemap.html	[113]
pLink	http://pfind.ict.ac.cn/software/pLink	[114]
Pro-Crosslink	https://c4c.uwc4c.com/express_license_technologies/pro-crosslink-version-1	[115]
ProteinXXX/GPMAW	http://www.gpmaw.com/html/downloads.html	[116,117]
VIRTUAL-MSLAB	Contact du développeur (ldk@science.uva.nl)	[118]
X-Link	-	[96]
X-Links	http://www.wsu.edu/proteomics/	[119]
X!Link	Contact du développeur (yjlee@iastate.edu)	[79]
xQuest/xProphet	http://prottools.ethz.ch/orinner/public/htdocs/xquest	[120]
StavroX/MeroX	http://www.stavrox.com/	[121,122]
XLPM	http://binf-app.host.ualr.edu/~mihir/cgi-bin/xlpm.cgi	[123]
X-Link Identifier	http://du-lab.org/	[124]
FINDX	http://findxlinks.blogspot.de/	[125]
xComb	https://persephone.rxlab.umaryland.edu/xcomb.php	[126]
XiQ	http://xiq.rappsilberlab.org	[127]

De manière générale, leur principe de fonctionnement est basé sur celui bien établi pour la protéomique classique. A partir d'une base de données contenant les séquences des protéines, ils génèrent les spectres théoriques des combinaisons possibles de pontages. Ensuite ils comparent les spectres expérimentaux avec les spectres théoriques pour sélectionner les séquences qui « matchent » les mieux. Néanmoins, chacun de ces logiciels permet l'identification des peptides modifiés avec des agents de pontage donnés et possède son propre algorithme de recherche sur une base de données. Certains peuvent traiter les données obtenues avec les expériences de pontage utilisant le marquage isotopique, tandis que d'autres identifient les peptides obtenus avec le pontage sans marquage (« label-free »). De même, les fonctions de calcul de score ne sont pas systématiquement implantées. Ainsi, le Tableau 4 recense les propriétés de quelques-uns de ces logiciels.

Tableau 4. Les propriétés de quelques logiciels dédiés à l'identification des peptides pontés. (*) Basé sur la première publication, () A l'échelle du protéome. Adapté de [45].**

Nom du logiciel	Application (*)	Type de données		Base de données		Calcul de score
		MS	MS/MS	Petite	Large (**)	
GPMaw	Marqué	✓	✓	✓	-	-
xComb	Label-free	-	✓	✓	-	-
X!Link	Label-free	-	✓	✓	-	✓
XLink-Identif	Label-free	-	✓	✓	-	✓
StavroX/MeroX	Label-free	✓	✓	✓	-	✓
BLinks	Label-free	✓	✓	✓	-	✓
pLink	Label-free	✓	✓	✓	✓	✓
xQuest/xProphet	Marqué	✓	✓	✓	✓	✓
XiQ	Marqué	✓	-	-	-	-

Petrochenko *et al.* [41] considèrent que pour être vraiment efficace, chaque nouvel agent de pontage ou approche méthodologique développée doit être supporté par un logiciel spécialement dédié. Malgré les progrès continus, ces logiciels souffrent toujours d'un manque d'automatisation et de compatibilité entre différents formats de données. Pour toutes ces raisons, il est considéré qu'à ce jour aucun logiciel ne satisfasse tous les critères requis pour analyser une large variété de types de données pouvant être obtenues par CX-MS [45].

Par conséquent, dans un protocole expérimental de CX-MS, le logiciel d'identification est à choisir en fonction de l'agent de pontage utilisé. Du temps doit également être consacré au tri et à l'utilisation de ces outils bio-informatiques. De manière générale, ces logiciels sont en développement continu. En cas de dysfonctionnements, les contacts avec les développeurs sont conseillés pour les signaler comme nous avons pu le faire dans le cadre de ces travaux de thèse.

Enfin, toujours en raison de la dispersion et de l'absence de « normalisation » des logiciels dédiés à l'identification des peptides pontés, une validation manuelle s'impose. Elle permet de valider des résultats issus de ces logiciels et de les comparer d'un logiciel à l'autre. Là encore, il n'existe pas de protocole général ou de critères normalisant la validation manuelle. Nous détaillerons nos critères de validation dans le chapitre 3.

2.2.5. Informations apportées par le pontage chimique couplé à la MS

2.2.5.1. La stœchiométrie du complexe

Classiquement, la réaction de pontage est suivie sur gel d'électrophorèse SDS-PAGE. Cette méthode plutôt sensible permet de suivre la disparition de bandes correspondantes aux sous-unités libres et l'apparition de bande de plus haut poids moléculaire correspondante au complexe

ponté. Néanmoins, il est difficile d'assigner une masse à cette bande, la migration sur gel étant affectée par la conformation ou encore la charge du complexe ponté.

Dès lors, une autre approche consiste à suivre la réaction de pontage par MALDI-MS. L'analyse en MALDI-MS est intéressante dans la mesure où elle peut confirmer l'existence et donner des informations sur la stœchiométrie du complexe ponté.

En effet, nous avons vu en 2.1.2.2 que de manière générale et sans conditions particulières, lors d'une analyse en MALDI-MS, des interactions non-covalentes sont dissociées lors des étapes de la préparation d'échantillon et de la désorption-ionisation. Dès lors, en mesurant la masse de l'échantillon avant pontage, nous pouvons détecter individuellement les sous-unités constituant le complexe et avons l'accès à leur masse. Ensuite, après la réaction de pontage, l'analyse en MALDI-MS fournit la masse du complexe ponté où ses sous-unités sont liées les unes aux autres par des réactions covalentes. La masse ainsi détectée correspond à la somme des masses des sous-unités impliquées dans le complexe. Avec une résolution suffisante, nous pouvons en déduire la quantité de chaque sous-unité et la stœchiométrie du complexe.

Bien entendu, ceci implique qu'à l'issue de la réaction de pontage, la stœchiométrie du complexe doit rester fidèle à celle en état natif. Autrement dit, le pontage ne doit pas générer les associations non-spécifiques [78]. Des conditions de réactions de pontage (concentration, ratio protéine/agent de pontage, température, temps de réaction) doivent être optimisées et la cinétique de la réaction suivie afin de s'assurer que les informations structurales qui en découlent sont fidèles à la réalité.

2.2.5.2. Les zones de proximité entre les sous-unités

Après la réaction de pontage, l'échantillon subit une protéolyse et le mélange de peptides résultant est ensuite analysé en ESI-MS/MS. A l'issue de l'identification, chaque type de peptides peut nous donner des informations structurales à des niveaux différents.

Les pontages inter-peptides intramoléculaires et intermoléculaires sont les plus exploités. En effet, l'identification de ces peptides permettra de déterminer des acides aminés pontés et donc les peptides soit en interaction, soit suffisamment proches dans la conformation du complexe. Ainsi nous avons accès à leurs contraintes de distance au sein d'une même protéine ou entre les protéines différentes [54,78,128,129]. Si les pontages intramoléculaires donnent des informations sur le repliement de chaque protéine lorsqu'elle est en interaction, les pontages intermoléculaires aident à déterminer les zones de proximité entre les partenaires.

Néanmoins, il faut garder à l'esprit qu'en s'ajoutant à la flexibilité des complexes biologiques et des acides aminés cibles, la longueur du bras espaceur des agents de pontage détermine la distance des ponts [44]. Les zones de proximité et les contraintes de distance déterminées par les pontages dépendent donc grandement de l'agent de pontage utilisé.

Les peptides « mono-link » et les peptides natifs ont été longtemps délaissés lors des études de complexes supramoléculaires par CX-MS. Ils sont considérés comme ne pouvant donner aucune ou très peu d'information structurale [40,41]. Ceci se traduit par des stratégies développées dans la littérature pour les exclure lors d'enrichissement des produits pontés [78,88]

ou lors d'analyse en MS/MS [40]. Néanmoins, nous pensons qu'ils ont beaucoup plus à offrir qu'au premier abord.

D'une part, les peptides « mono-link » donnent des indications quant à l'accessibilité de la surface des protéines au solvant et à l'agent de pontage [40], et à la réactivité des différents sites des protéines. Dans ce sens, lorsque la même séquence d'acides aminés est également identifiée dans des peptides pontés, ces peptides « mono-link » renforcent la confiance en des identifications de pontages intra- et intermoléculaires.

D'autre part, les peptides natifs ou non-modifiés par l'agent de pontage peuvent être également exploités à bon escient. Au cas où certaines zones dans la séquence protéique ne sont pas identifiées comme pontées, il serait intéressant de vérifier si elles ont été identifiées comme étant des peptides natifs. Si c'est le cas, cela pourrait suggérer que ces zones restent accessibles au solvant, à l'agent de pontage et à l'enzyme, et l'absence de pontage peut traduire un manque d'interactions, de proximité ou de réactivité des acides aminés cibles. En revanche, si ces zones ne sont identifiées ni comme peptides pontés ni comme peptides natifs, cela peut être une indication de l'inaccessibilité au solvant et à l'enzyme, comme par exemple des zones enfouies au cœur du complexe. Bien entendu, les informations obtenues par ces peptides natifs et « mono-links » ne sont que des indications et sont à utiliser avec précaution.

Enfin, après l'identification des produits pontés, les résultats obtenus peuvent être combinés avec ceux d'autres techniques (comme par exemple la microscopie électronique, la diffraction des rayons X) pour mener à bien l'élucidation structurale des édifices supramoléculaires [51,130]. Les informations sur les contraintes de distance peuvent également être insérées dans des algorithmes de modélisation moléculaire pour ériger des modèles de structure [131,132]. Ceci est également une manière élégante de combiner les données expérimentales et les données théoriques pour la prédiction et l'affinement de la structure des biomolécules.

2.2.6. Contraintes et limites

Malgré les développements continus, l'approche de pontage chimique couplé à la spectrométrie de masse est loin d'être triviale. Tout d'abord, cette complexité se reflète particulièrement à travers un manque d'uniformisation du vocabulaire et des lexiques utilisés. Par exemple, tandis que certains utilisent le sigle CX-MS (pour « chemical cross-linking mass spectrometry »), d'autres parlent de CL-MS (pour « cross-linking mass spectrometry ») ou encore XL-MS (pour « cross-linking mass spectrometry »). De même, nous pouvons trouver différentes façons d'orthographier le mot « pontage » : crosslink, cross-link, cross-linking *etc.* Les noms des peptides obtenus après pontage sont aussi complexes : soit ils sont numérotés (peptides de types 0, 1, 2, 3) ; soit ils sont plus explicités (« inter-link », « intra-link », « monolink » ou « dead-end »).

Ensuite, il n'existe pas de protocole expérimental défini, ou en tout cas généralisable comme en protéomique classique, principalement parce que la réaction de pontage est protéine-dépendante et complexe-dépendante. Des précautions doivent aussi être prises lors de la mise en œuvre de la réaction de pontage pour qu'elle soit optimale, c'est-à-dire avec un rendement suffisant sans induire d'agrégations non-spécifiques.

De plus, le pontage chimique, en créant des liaisons covalentes entre les sous-unités, introduit de l'hétérogénéité parmi les molécules du complexe ponté. Lors de l'analyse par MALDI-MS après la réaction de pontage et avant la digestion, ceci a pour conséquence une diminution de l'intensité du signal et un élargissement des pics correspondant au complexe dans le spectre de masse. Un compromis est donc à trouver quant à la puissance du laser utilisée. Elle doit être suffisante pour pouvoir détecter le complexe, mais ne doit pas être trop importante pour ne pas élargir encore plus les pics.

Lors d'analyse en MS/MS, la grande taille des peptides pontés (par rapport aux peptides natifs) constitue également un inconvénient. D'une part, l'énergie interne est distribuée à travers un grand nombre de liaisons, il y a donc moins d'énergie par liaison et donc moins de coupure de liaison, ce qui peut mener à une fragmentation incomplète [41]. D'autre part, la fenêtre m/z utilisée est limitée pour arriver à un bon compromis entre le nombre de cycles d'analyse (donc nombre de peptides analysés) et la précision de mesure de masse. Par rapport aux peptides ne contenant pas de pont, même en étant multichargés, le rapport m/z des peptides pontés ont plus de chance de dépasser la fenêtre de masse paramétrée. Ceci sans mentionner les phénomènes de suppression d'ions liés à des peptides plus facilement ionisés, ou de la dépendance des données (« data-dependence ») des analyses car les ions les plus abondants sont sélectionnés en premier pour la fragmentation [78]. Toutes ces raisons compliquent la détection des peptides pontés en MS/MS.

Ensuite, l'identification des peptides pontés présente également deux difficultés majeures. Premièrement, après l'analyse en MS/MS, l'opérateur se retrouve souvent avec l'énorme quantité totale de spectres de masse où les peptides pontés sont présents en faible abondance par rapport aux peptides natifs [78] ; cette faible quantité est encore plus réduite par la diversité des types de peptides modifiés par le pontage mentionnés précédemment. Deuxièmement, il existe une multitude de combinaisons théoriques de peptides pontés. Considérant un nombre de peptides natifs n , le nombre théorique de peptides pontés est $\frac{n^2+n}{2}$ [86,114], ce qui constitue une quantité considérable. Une masse mesurée peut donc correspondre à un large nombre de séquences de peptides pontés théoriques et rend l'identification difficile (le nombre de matchs aléatoires augmente, le temps de traitement par les logiciels est allongé *etc.*). Par conséquent, le nombre de peptides pontés identifiés dépend grandement de la précision de mesure de masse. Il a été mentionné dans la littérature qu'une précision d'au moins ± 10 ppm est nécessaire [84,112,133]. Dans le cadre de ces travaux de thèse, nous avons remarqué que 97 % des pontages identifiés ont une erreur aux alentours de 2 ppm. Ainsi, l'accès à un spectromètre de masse de haute précision (par exemple de type Orbitrap) est presque une condition *sine qua non*. Dans ce sens, les progrès instrumentaux permanents pour avoir de meilleures capacités d'analyse, résolution et précision de mesure de masse vont continuer à contribuer à l'essor du CX-MS.

En outre, comme mentionné précédemment, un logiciel d'identification dédié capable d'être automatisé et qui minimise les interventions manuelles serait désirable, particulièrement lors de l'interprétation des données MS/MS. Aujourd'hui, il n'existe pas de critères de validation d'identification des peptides pontés standardisés malgré un nombre croissant de publications en CX-MS. Enfin, il faut garder à l'esprit que les difficultés du CX-MS augmentent avec la complexité du système étudié, qui elle-même dépend de la taille et de l'hétérogénéité des complexes biologiques [78].

2.3. Conclusion de l'étude bibliographique

Depuis l'apparition de la spectrométrie de masse supramoléculaire, son potentiel est sans cesse repoussé dans l'analyse des interactions non-covalentes des composés biologiques. Ainsi, elle permet entre autres l'accès à la détection, la stœchiométrie, la spécificité et les constantes d'affinité de ces interactions. Cependant, certaines limites demeurent, notamment concernant l'étude des complexes régis par l'effet hydrophobe, ou les complexes ne pouvant pas être transférés dans des tampons volatils.

Outre l'ESI-MS supramoléculaire classique, les approches alternatives ont été proposées pour obtenir des informations complémentaires telles que les indications sur la conformation dépliée/globulaire des biomolécules ou les zones en interaction entre les sous-unités. Parmi celles-ci, le pontage chimique couplé à la spectrométrie de masse (CX-MS) a montré ses preuves et apparaît comme une approche prometteuse. Néanmoins, sa mise en œuvre n'est pas triviale, chaque étape expérimentale est cruciale, dépend du stade précédent et contribue au succès de l'étude. L'accent doit être mis sur le choix de l'agent de pontage, le contrôle des conditions de réaction, l'optimisation des analyses par MS et le traitement des données [45].

Dans ce contexte, ces travaux de thèse s'inscrivent dans la volonté de repousser les limites de la spectrométrie de masse supramoléculaire, que ce soit par MALDI-MS en couplage avec un pontage chimique, ou par ESI-MS en conditions natives. Ainsi, les différents systèmes sur lesquels nous avons été amenés à travailler nous ont permis d'appliquer et d'améliorer des approches méthodologiques et de répondre à des questions biologiques.

Bibliographie

- [1] Karas, M., Bachmann, D., Hillenkamp, F., Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* 1985, 57, 2935–2939.
- [2] Karas, M., Bahr, U., Laser desorption mass spectrometry. *TrAC Trends Anal. Chem.* 1986, 5, 90–93.
- [3] Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Process.* 1987, 78, 53–68.
- [4] Tanaka, K., Waki, H., Ido, Y., Akita, S., et al., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 1988, 2, 151–153.
- [5] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 1988, 60, 2299–2301.
- [6] Karas, M., Bahr, U., Hillenkamp, F., UV laser matrix desorption/ionization mass spectrometry of proteins in the 100 000 dalton range. *Int. J. Mass Spectrom. Ion Process.* 1989, 92, 231–242.
- [7] Karas, M., Bahr, U., Ingendoh, A., Hillenkamp, F., Laser Desorption/Ionization Mass Spectrometry of Proteins of Mass 100 000 to 250 000 Dalton. *Angew. Chem. Int. Ed. Engl.* 1989, 28, 760–761.
- [8] Hillenkamp, F., Peter-Katalinic, J., *MALDI MS: a practical guide to instrumentation, methods and applications*, Wiley-VCH, Weinheim 2007.
- [9] Dreisewerd, K., The Desorption Process in MALDI. *Chem. Rev.* 2003, 103, 395–426.
- [10] Chen, F., Applying MALDI-ToF Mass Spectrometry to Study Non-covalent Interactions. Thèse ETH Zürich, 2014.
- [11] Mème, A., Développement de La Spectrométrie de Masse Pour L'étude D'édifices Non-Covalents En Chimie et En Biologie. Thèse Université de Strasbourg, 2010.
- [12] Zenobi, R., Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* 1998, 17, 337–366.
- [13] Ehring, H., Karas, M., Hillenkamp, F., Role of photoionization and photochemistry in ionization processes of organic molecules and relevance for matrix-assisted laser desorption ionization mass spectrometry. *Org. Mass Spectrom.* 1992, 27, 472–480.
- [14] Karas, M., Glückmann, M., Schäfer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* 2000, 35, 1–12.
- [15] Krüger, R., Pfenninger, A., Fournier, I., Gluckmann, M., Karas, M., Analyte incorporation and ionization in matrix-assisted laser desorption/ionization visualized by pH indicator molecular probes. *Anal. Chem.* 2001, 73, 5812–5821.
- [16] Karas, M., Krüger, R., Ion Formation in MALDI: The Cluster Ionization Mechanism. *Chem. Rev.* 2003, 103, 427–440.
- [17] Knochenmuss, R., Zenobi, R., MALDI Ionization: The Role of In-Plume Processes. *Chem. Rev.* 2003, 103, 441–452.
- [18] McCarley, T.D., McCarley, R.L., Limbach, P.A., Electron-Transfer Ionization in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* 1998, 70, 4376–4379.
- [19] Sheng, L.-S., Covey, J.E., Shew, S.L., Winger, B.E., Campana, J.E., Matrix-assisted laser desorption ionization Fourier-transform mass spectrometry. *Rapid Commun. Mass Spectrom.* 1994, 8, 498–500.
- [20] Jonscher, K., Currie, G., McCormack, A.L., Yates, J.R., Matrix-assisted laser desorption of peptides and proteins on a quadrupole ion trap mass spectrometer. *Rapid Commun. Mass Spectrom. RCM* 1993, 7, 20–26.

- [21] Schwartz, J.C., Bier, M.E., Matrix-assisted laser desorption of peptides and proteins using a quadrupole ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* 1993, 7, 27–32.
- [22] Hill, J.A., Annan, R.S., Biemann, K., Matrix-assisted laser desorption ionization with a magnetic mass spectrometer. *Rapid Commun. Mass Spectrom. RCM* 1991, 5, 395–399.
- [23] Wenzel, R.J., Matter, U., Schultheis, L., Zenobi, R., Analysis of Megadalton Ions Using Cryodetection MALDI Time-of-Flight Mass Spectrometry. *Anal. Chem.* 2005, 77, 4329–4337.
- [24] Yanes, O., Avilés, F.X., Wenzel, R., Nazabal, A., et al., Proteomic Profiling of a Snake Venom Using High Mass Detection MALDI-TOF Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2007, 18, 600–606.
- [25] Wenzel, R., Röhling, U., Nazabal, A., Hillenkamp, F., A Detector Device for High Mass Ion Detection, a Method for Analyzing Ions of High Mass and a Device for Selection Between Ion Detectors. WO/2009/086642, 2009.
- [26] Lecchi, P., Le, H.M., Pannell, L.K., 6-Aza-2-thiothymine: a matrix for MALDI spectra of oligonucleotides. *Nucleic Acids Res.* 1995, 23, 1276–1277.
- [27] Zehl, M., Allmaier, G., Ultraviolet matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of intact hemoglobin complex from whole human blood. *Rapid Commun. Mass Spectrom.* 2004, 18, 1932–1938.
- [28] Thiede, B., Janta-Lipinski, M. von, Noncovalent RNA–peptide complexes detected by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 1998, 12, 1889–1894.
- [29] Jespersen, S., Niessen, W.M.A., Tjaden, U.R., van der Greef, J., Basic matrices in the analysis of non-covalent complexes by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* 1998, 33, 1088–1093.
- [30] Lecchi, P., Pannell, L.K., The detection of intact double-stranded DNA by MALDI. *J. Am. Soc. Mass Spectrom.* 1995, 6, 972–975.
- [31] Song, F., A Study of Noncovalent Protein Complexes by Matrix-Assisted Laser Desorption/Ionization. *J. Am. Soc. Mass Spectrom.* 2007, 18, 1286–1290.
- [32] Rosinke, B., Strupat, K., Hillenkamp, F., Rosenbusch, J., et al., Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) of membrane proteins and non-covalent complexes. *J. Mass Spectrom.* 1995, 30, 1462–1468.
- [33] Cohen, L.R.H., Strupat, K., Hillenkamp, F., Analysis of quaternary protein ensembles by matrix assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 1997, 8, 1046–1052.
- [34] Horneffer, V., Strupat, K., Hillenkamp, F., Localization of Noncovalent Complexes in MALDI-Preparations by CLSM. *J. Am. Soc. Mass Spectrom.* 2006, 17, 1599–1604.
- [35] Wortmann, A., Pimenova, T., Alves, S., Zenobi, R., Investigation of the first shot phenomenon in MALDI mass spectrometry of protein complexes. *Analyst* 2007, 132, 199–207.
- [36] Sumner, J.B., Graham, V.A., The nature of insoluble urease. *Exp. Biol. Med.* 1925, 22, 504–506.
- [37] Means, G.E., Feeney, R.E., Chemical modifications of proteins: history and applications. *Bioconjug. Chem.* 1990, 1, 2–12.
- [38] Bailey, A.J., Peach, C.M., Fowler, L.J., Chemistry of the collagen cross-links. Isolation and characterization of two intermediate intermolecular cross-links in collagen. *Biochem. J.* 1970, 117, 819–831.
- [39] Fasold, H., Klappenberger, J., Meyer, C., Remold, H., Bifunctional Reagents for the Crosslinking of Proteins. *Angew. Chem. Int. Ed. Engl.* 1971, 10, 795–801.
- [40] Paramelle, D., Miralles, G., Subra, G., Martinez, J., Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* 2013, 13, 438–456.
- [41] Petrotchenko, E.V., Borchers, C.H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* 2010, 29, 862–876.
- [42] Back, J.W., de Jong, L., Muijsers, A.O., de Koster, C.G., Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* 2003, 331, 303–313.

- [43] Sinz, A., Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* 2003, 38, 1225–1237.
- [44] Sinz, A., Arlt, C., Chorev, D., Sharon, M., Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Sci.* 2015, 24, 1193–1209.
- [45] Tran, B.Q., Goodlett, D.R., Goo, Y.A., Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 2015.
- [46] Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., et al., Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* 2010, 29, 717–726.
- [47] Jennebach, S., Herzog, F., Aebersold, R., Cramer, P., Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* 2012, 40, 5591–5601.
- [48] Lauber, M.A., Reilly, J.P., Structural Analysis of a Prokaryotic Ribosome Using a Novel Amidinating Cross-Linker and Mass Spectrometry. *J. Proteome Res.* 2011, 10, 3604–3616.
- [49] Kao, A., Randall, A., Yang, Y., Patel, V.R., et al., Mapping the Structural Topology of the Yeast 19S Proteasomal Regulatory Particle Using Chemical Cross-linking and Probabilistic Modeling. *Mol. Cell. Proteomics MCP* 2012, 11, 1566–1577.
- [50] Bohn, S., Beck, F., Sakata, E., Walzthoeni, T., et al., Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *Proc. Natl. Acad. Sci. U. S. A.* 2010, 107, 20992–20997.
- [51] Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., et al., Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, 1380–1387.
- [52] Kalisman, N., Adams, C.M., Levitt, M., Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci.* 2012, 109, 2884–2889.
- [53] Leitner, A., Joachimiak, L.A., Bracher, A., Mönkemeyer, L., et al., The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Struct. England* 2012, 20, 814–825.
- [54] Walzthoeni, T., Leitner, A., Stengel, F., Aebersold, R., Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* 2013, 23, 252–260.
- [55] Wong, S.S., *Chemistry of Protein Conjugation and Cross-Linking*, CRC Press, 1991.
- [56] Hermanson, G.T., *Bioconjugate Techniques*, Academic Press, 2013.
- [57] Paramelle, D., Cantel, S., Enjalbal, C., Amblard, M., et al., A new generation of cross-linkers for selective detection by MALDI MS. *Proteomics* 2009, 9, 5384–5388.
- [58] Nury, C., Redeker, V., Dautrey, S., Romieu, A., et al., A Novel Bio-Orthogonal Cross-Linker for Improved Protein/Protein Interaction Analysis. *Anal. Chem.* 2015, 87, 1853–1860.
- [59] Leitner, A., Joachimiak, L.A., Unverdorben, P., Walzthoeni, T., et al., Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* 2014, 111, 9455–9460.
- [60] Cravello, L., ETUDES STRUCTURALES DES PROTEINES PAR SPECTROMETRIE DE MASSE COUPLEE AUX ECHANGES HYDROGENE/DEUTERIUM ET A LA RETICULATION CHIMIQUE. Thèse Université Joseph-Fourier - Grenoble I, 2005.
- [61] Leavell, M.D., Novak, P., Behrens, C.R., Schoeniger, J.S., Kruppa, G.H., Strategy for selective chemical cross-linking of tyrosine and lysine residues. *J. Am. Soc. Mass Spectrom.* 2004, 15, 1604–1611.
- [62] Swaim, C.L., Smith, J.B., Smith, D.L., Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *J. Am. Soc. Mass Spectrom.* 2004, 15, 736–749.
- [63] Ouhia, A., Rene, L., Guilhem, J., Pascard, C., Badet, B., A new diazoacylating reagent: preparation, structure, and use of succinimidyl diazoacetate. *J. Org. Chem.* 1993, 58, 1641–1642.

- [64] Guo, Z.-F., Sun, Y., Zheng, S., Guo, Z., Preferential Hydrolysis of Aberrant Intermediates by the Type II Thioesterase in Escherichia coli Nonribosomal Enterobactin Synthesis: Substrate Specificities and Mutagenic Studies on the Active-Site Residues†. *Biochemistry (Mosc.)* 2009, 48, 1712–1722.
- [65] Gourand, F., Mercey, G., Ibazizène, M., Tirel, O., et al., Chemical Delivery System of Metaiodobenzylguanidine (MIBG) to the Central Nervous System. *J. Med. Chem.* 2010, 53, 1281–1287.
- [66] Besselink, G. a. J., Beugeling, T., Bantjes, A., N-Hydroxysuccinimide-activated glycine-sepharose. *Appl. Biochem. Biotechnol.* 1993, 43, 227–246.
- [67] Hartman, F.C., Wold, F., Bifunctional Reagents. Cross-Linking of Pancreatic Ribonuclease with a Diimido Ester¹. *J. Am. Chem. Soc.* 1966, 88, 3890–3891.
- [68] Hunter, M.J., Ludwig, M.L., The Reaction of Imidoesters with Proteins and Related Small Molecules. *J. Am. Chem. Soc.* 1962, 84, 3491–3504.
- [69] Grabarek, Z., Gergely, J., Zero-length crosslinking procedure with the use of active esters. *Anal. Biochem.* 1990, 185, 131–135.
- [70] Smyth, D.G., Blumenfeld, O.O., Konigsberg, W., Reactions of N-ethylmaleimide with peptides and amino acids. *Biochem. J.* 1964, 91, 589–595.
- [71] Kim, Y., Ho, S.O., Gassman, N.R., Korlann, Y., et al., Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjug. Chem.* 2008, 19, 786–791.
- [72] Brewer, C.F., Riehm, J.P., Evidence for possible nonspecific reactions between N-ethylmaleimide and proteins. *Anal. Biochem.* 1967, 18, 248–255.
- [73] Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* 2006, 25, 663–682.
- [74] Lee, H.S., Dimla, R.D., Schultz, P.G., Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. *Bioorg. Med. Chem. Lett.* 2009, 19, 5222–5224.
- [75] Forné, I., Ludwigsen, J., Imhof, A., Becker, P.B., Mueller-Planitz, F., Probing the Conformation of the ISWI ATPase Domain With Genetically Encoded Photoreactive Crosslinkers and Mass Spectrometry. *Mol. Cell. Proteomics MCP* 2012, 11.
- [76] Prestwich, G.D., Dormán, G., Elliott, J.T., Marecak, D.M., Chaudhary, A., Benzophenone Photoprobes for Phosphoinositides, Peptides and Drugs. *Photochem. Photobiol.* 1997, 65, 222–234.
- [77] Green, N.S., Reisler, E., Houk, K.N., Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. *Protein Sci. Publ. Protein Soc.* 2001, 10, 1293–1304.
- [78] Merkley, E.D., Cort, J.R., Adkins, J.N., Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J. Struct. Funct. Genomics* 2013, 14, 77–90.
- [79] Lee, Y.J., Lackner, L.L., Nunnari, J.M., Phinney, B.S., Shotgun Cross-Linking Analysis for Studying Quaternary and Tertiary Protein Structures. *J. Proteome Res.* 2007, 6, 3908–3917.
- [80] Novak, P., Young, M.M., Schoeniger, J.S., Kruppa, G.H., A top-down approach to protein structure studies using chemical cross-linking and Fourier transform mass spectrometry. *Eur. J. Mass Spectrom. Chichester Engl.* 2003, 9, 623–631.
- [81] Novak, P., Haskins, W.E., Ayson, M.J., Jacobsen, R.B., et al., Unambiguous assignment of intramolecular chemical cross-links in modified mammalian membrane proteins by Fourier transform-tandem mass spectrometry. *Anal. Chem.* 2005, 77, 5101–5106.
- [82] Rappsilber, J., The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* 2011, 173, 530–540.
- [83] Schilling, B., Row, R.H., Gibson, B.W., Guo, X., Young, M.M., MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* 2003, 14, 834–850.

- [84] Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., et al., Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteomics* 2010, 9, 1634–1649.
- [85] Luo, J., Fishburn, J., Hahn, S., Ranish, J., An Integrated Chemical Cross-linking and Mass Spectrometry Approach to Study Protein Complex Architecture and Function. *Mol. Cell. Proteomics* 2011, 11, M111.008318.
- [86] Stengel, F., Aebersold, R., Robinson, C.V., Joining Forces: Integrating Proteomics and Cross-linking with the Mass Spectrometry of Intact Complexes. *Mol. Cell. Proteomics* 2012, 11, R111.014027.
- [87] Hurst, G.B., Lankford, T.K., Kennel, S.J., Mass spectrometric detection of affinity purified crosslinked peptides. *J. Am. Soc. Mass Spectrom.* 2004, 15, 832–839.
- [88] Trester-Zedlitz, M., Kamada, K., Burley, S.K., Fenyő, D., et al., A Modular Cross-Linking Approach for Exploring Protein Interactions. *J. Am. Chem. Soc.* 2003, 125, 2416–2425.
- [89] Müller, D.R., Schindler, P., Towbin, H., Wirth, U., et al., Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* 2001, 73, 1927–1934.
- [90] Pearson, K.M., Pannell, L.K., Fales, H.M., Intramolecular cross-linking experiments on cytochrome c and ribonuclease A using an isotope multiplet method. *Rapid Commun. Mass Spectrom.* 2002, 16, 149–159.
- [91] Collins, C.J., Schilling, B., Young, M., Dollinger, G., Guy, R.K., Isotopically labeled crosslinking reagents: resolution of mass degeneracy in the identification of crosslinked peptides. *Bioorg. Med. Chem. Lett.* 2003, 13, 4023–4026.
- [92] Chen, X., Chen, Y.H., Anderson, V.E., Protein Cross-Links: Universal Isolation and Characterization by Isotopic Derivatization and Electrospray Ionization Mass Spectrometry. *Anal. Biochem.* 1999, 273, 192–203.
- [93] Petrotchenko, E.V., Serpa, J.J., Borchers, C.H., Use of a Combination of Isotopically Coded Cross-Linkers and Isotopically Coded N-Terminal Modification Reagents for Selective Identification of Inter-peptide Crosslinks. *Anal. Chem.* 2010, 82, 817–823.
- [94] Back, J.W., Notenboom, V., de Koning, L.J., Muijsers, A.O., et al., Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling. *Anal. Chem.* 2002, 74, 4417–4422.
- [95] Huang, B.X., Kim, H.-Y., Dass, C., Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2004, 15, 1237–1247.
- [96] Taverner, T., Hall, N.E., O'Hair, R.A.J., Simpson, R.J., Characterization of an Antagonist Interleukin-6 Dimer by Stable Isotope Labeling, Cross-linking, and Mass Spectrometry. *J. Biol. Chem.* 2002, 277, 46487–46492.
- [97] Sinz, A., Wang, K., Mapping Protein Interfaces with a Fluorogenic Cross-Linker and Mass Spectrometry: Application to Nebulin–Calmodulin Complexes. *Biochemistry (Mosc.)* 2001, 40, 7903–7913.
- [98] Wine, R.N., Dial, J.M., Tomer, K.B., Borchers, C.H., Identification of components of protein complexes using a fluorescent photo-cross-linker and mass spectrometry. *Anal. Chem.* 2002, 74, 1939–1945.
- [99] Bennett, K.L., Kussmann, M., Björk, P., Godzwon, M., 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 Sci. Publ. Protein Soc.* 2000, 9, 1503–1518.
- [100] Yang, L., Tang, X., Weisbrod, C., Munske, G., et al., pcPIR, a photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. *Anal. Chem.* 2010, 82, 3556–3566.
- [101] Back, J.W., Hartog, A.F., Dekker, H.L., Muijsers, A.O., et al., A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes. *J. Am. Soc. Mass Spectrom.* 2001, 12, 222–227.

- [102] Chowdhury, S.M., Munske, G.R., Tang, X., Bruce, J.E., Collisionally Activated Dissociation and Electron Capture Dissociation of Several Mass Spectrometry-Identifiable Chemical Cross-Linkers. *Anal. Chem.* 2006, 78, 8183–8193.
- [103] Lascoux, D., Paramelle, D., Subra, G., Heymann, M., et al., Discrimination and Selective Enhancement of Signals in the MALDI Mass Spectrum of a Protein by Combining a Matrix-Based Label for Lysine Residues with a Neutral Matrix. *Angew. Chem. Int. Ed.* 2007, 46, 5594–5597.
- [104] Lee, Y.J., Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol. Biosyst.* 2008, 4, 816–823.
- [105] Hoopmann, M.R., Weisbrod, C.R., Bruce, J.E., Improved Strategies for Rapid Identification of Chemically Cross-linked Peptides Using Protein Interaction Reporter Technology. *J. Proteome Res.* 2010, 9, 6323–6333.
- [106] Tang, Y., Chen, Y., Lichti, C.F., Hall, R.A., et al., CLPM: A Cross-Linked Peptide Mapping Algorithm for Mass Spectrometric Analysis. *BMC Bioinformatics* 2005, 6, S9.
- [107] McIlwain, S., Singh, P., Draghicescu, P., Goodlett, D.R., Noble, W.S., Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J. Proteome Res.* 2010, 9, 2488–2495.
- [108] Bich, C., Bovet, C., Rochel, N., Peluso-Iltis, C., et al., Detection of Nucleic Acid–Nuclear Hormone Receptor Complexes with Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2010, 21, 635–645.
- [109] Chen, F., Gerber, S., Korkhov, V.M., Mireku, S., et al., On the Efficiency of NHS Ester Cross-Linkers for Stabilizing Integral Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* 2014, 26, 493–498.
- [110] Xu, H., Zhang, L., Freitas, M.A., Identification and Characterization of Disulfide Bonds in Proteins and Peptides from Tandem MS Data by Use of the MassMatrix MS/MS Search Engine. *J. Proteome Res.* 2008, 7, 138–144.
- [111] Rasmussen, M.I., Refsgaard, J.C., Peng, L., Houen, G., Højrup, P., CrossWork: Software-assisted identification of cross-linked peptides. *J. Proteomics* 2011, 74, 1871–1883.
- [112] Clauser, K.R., Baker, P., Burlingame, A.L., Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* 1999, 71, 2871–2882.
- [113] Fenyő, D., A software tool for the analysis of mass spectrometric disulfide mapping experiments. *Comput. Appl. Biosci. CABIOS* 1997, 13, 617–618.
- [114] Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., et al., Identification of cross-linked peptides from complex samples. *Nat. Methods* 2012, 9, 904–906.
- [115] Gao, Q., Xue, S., Doneanu, C.E., Shaffer, S.A., et al., Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. *Anal. Chem.* 2006, 78, 2145–2149.
- [116] Peri, S., Steen, H., Pandey, A., GPMW – a software tool for analyzing proteins and peptides. *Trends Biochem. Sci.* 2001, 26, 687–689.
- [117] Nielsen, T., Thaysen-Andersen, M., Larsen, N., Jørgensen, F.S., et al., Determination of protein conformation by isotopically labelled cross-linking and dedicated software: Application to the chaperone, calreticulin. *Int. J. Mass Spectrom.* 2007, 268, 217–226.
- [118] de Koning, L.J., Kasper, P.T., Back, J.W., Nessen, M.A., et al., Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. *FEBS J.* 2006, 273, 281–291.
- [119] Anderson, G.A., Tolic, N., Tang, X., Zheng, C., Bruce, J.E., Informatics Strategies for Large-Scale Novel Cross-Linking Analysis. *J. Proteome Res.* 2007, 6, 3412–3421.
- [120] Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., et al., Identification of cross-linked peptides from large sequence databases. *Nat. Methods* 2008, 5, 315–318.
- [121] Götze, M., Pettelkau, J., Schaks, S., Bosse, K., et al., StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* 2011, 23, 76–87.

- [122] Götze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., et al., Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. *J. Am. Soc. Mass Spectrom.* 2014, 26, 83–97.
- [123] Jaiswal, M., Crabtree, N.M., Bauer, M.A., Hall, R., et al., XLPM: efficient algorithm for the analysis of protein-protein contacts using chemical cross-linking mass spectrometry. *BMC Bioinformatics* 2014, 15, S16.
- [124] Du, X., Chowdhury, S.M., Manes, N.P., Wu, S., et al., Xlink-Identifier: An Automated Data Analysis Platform for Confident Identifications of Chemically Cross-linked Peptides using Tandem Mass Spectrometry. *J. Proteome Res.* 2011, 10, 923–931.
- [125] Söderberg, C.A.G., Lambert, W., Kjellström, S., Wiegandt, A., et al., Detection of Crosslinks within and between Proteins by LC-MALDI-TOF/TOF and the Software FINDX to Reduce the MSMS-Data to Acquire for Validation. *PLoS ONE* 2012, 7.
- [126] Panchaud, A., Singh, P., Shaffer, S.A., Goodlett, D.R., xComb: A Cross-Linked Peptide Database Approach to Protein–Protein Interaction Analysis. *J. Proteome Res.* 2010, 9, 2508–2515.
- [127] Fischer, L., Chen, Z.A., Rappsilber, J., Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* 2013, 88, 120–128.
- [128] Redeker, V., Pemberton, S., Bienvenut, W., Bousset, L., Melki, R., Identification of Protein Interfaces between α -Synuclein, the Principal Component of Lewy Bodies in Parkinson Disease, and the Molecular Chaperones Human Hsc70 and the Yeast Ssa1p. *J. Biol. Chem.* 2012, 287, 32630–32639.
- [129] Redeker, V., Bonnefoy, J., Le Caer, J.-P., Pemberton, S., et al., A region within the C-terminal domain of Ure2p is shown to interact with the molecular chaperone Ssa1p by the use of cross-linkers and mass spectrometry. *FEBS J.* 2010, 277, 5112–5123.
- [130] Greber, B.J., Boehringer, D., Leitner, A., Bieri, P., et al., Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* 2014, 505, 515–519.
- [131] Chen, Y., Ding, F., Dokholyan, N.V., Fidelity of the Protein Structure Reconstruction from Inter-Residue Proximity Constraints. *J. Phys. Chem. B* 2007, 111, 7432–7438.
- [132] Schulz, D.M., Ihling, C., Clore, G.M., Sinz, A., Mapping the Topology and Determination of a Low-Resolution Three-Dimensional Structure of the Calmodulin–Melittin Complex by Chemical Cross-Linking and High-Resolution FTICRMS: Direct Demonstration of Multiple Binding Modes†. *Biochemistry (Mosc.)* 2004, 43, 4703–4715.
- [133] Clavier, S., Du, X., Sagan, S., Bolbach, G., Sachon, E., An integrated cross-linking-MS approach to investigate cell penetrating peptides interacting partners. *EuPA Open Proteomics* 2014, 3, 229–238.

Partie II

Résultats et discussion

Dans cette deuxième partie, nous présenterons les résultats obtenus au cours de ces travaux de thèse ainsi que nos interprétations.

La partie résultats et discussion comporte trois chapitres :

- **Chapitre 3** : Etude par CX-MS de la stœchiométrie et de la topologie du complexe non-covalent SAGA HAT
- **Chapitre 4** : Suivi par ESI-MS de la formation des complexes de RAR-RXR avec différents ADNs
- **Chapitre 5** : Mesure par ESI-MS de la constante de dissociation des complexes RXR-ligand

Chapitre 3 : Etude par CX-MS de la stœchiométrie et de la topologie du complexe non-covalent SAGA HAT

Dans l'introduction bibliographique, nous avons vu que l'approche de pontage chimique couplé à la spectrométrie de masse (CX-MS) est une alternative intéressante pour explorer les complexes biologiques non-covalents ainsi que les zones de proximité entre les partenaires.

Dans ce chapitre, nous décrirons comment nous avons mis en place un protocole permettant d'appliquer cette approche à l'étude de la stœchiométrie et de la topologie du complexe non-covalent SAGA HAT (Spt-Ada-Gcn5 Acétyltransferase, module Histone Acétyl Transferase) humain. Nous détaillerons des optimisations de conditions de détection en MALDI-MS et de réaction de pontage. Ensuite, nous discuterons des étapes d'identification et d'interprétation des données MS/MS. Nous verrons comment les informations obtenues par CX-MS ont été exploitées pour dresser une carte d'interactions entre les sous-unités et un premier modèle structural du complexe. Enfin, nous mentionnerons des contraintes, avantages et limites rencontrés au cours de cette étude.

3.1. Contexte de l'étude : le complexe SAGA HAT

Dans le cadre d'une collaboration avec l'équipe du Dr. Patrick Schultz et Dr. Valérie Lamour au Département de Biologie Structurale Intégrative de l'IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France), nous avons travaillé sur un sous-complexe recombinant de SAGA HAT (Spt-Ada-Gcn5 Acétyltransferase, module Histone Acétyl Transferase) humain. Le complexe SAGA est impliqué dans le processus de transcription, il permet l'acétylation des résidus lysines des histones, ce qui déroule les chromatines et rend l'ADN accessible (Figure 1). Chez l'humain, le complexe SAGA, d'une masse de 1,8 MDa, comporte 18 sous-unités [1,2], le sous-complexe SAGA HAT comporte quant à lui 4 sous-unités : SGF29 (33 238 Da), ADA2B (48 470 Da), ADA3 (48 902 Da) et GCN5 (93 926 Da, avec le tag Histidine 97 102 Da).

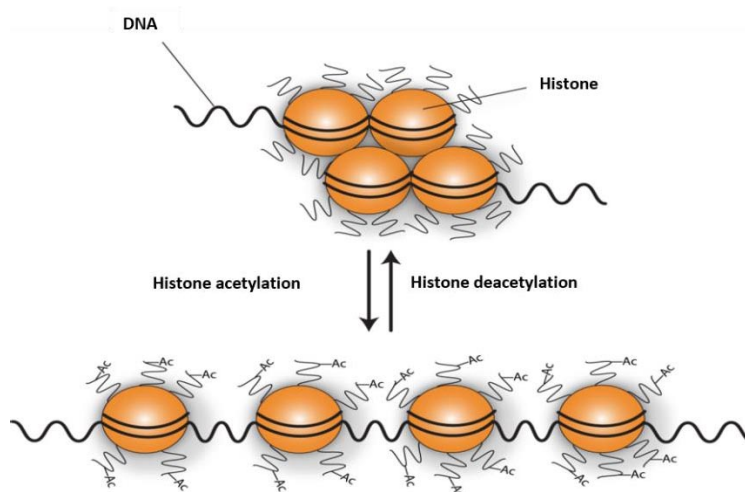


Figure 1. Représentation schématique de l'acétylation des histones permettant l'activation de la transcription.

L'étude de ce sous-complexe par les techniques d'analyse structurale conventionnelles par nos collaborateurs a rencontré quelques difficultés. Premièrement, il n'a pas été possible jusqu'à ce jour d'obtenir des cristaux de ce sous-complexe permettant leur étude en DRX. Deuxièmement, les images EM sont floues et le manque de détails ne permet pas une détermination structurale de qualité. Ces difficultés sont liées probablement à la grande hétérogénéité et la flexibilité du complexe. Enfin, seulement quelques structures atomiques de certains domaines structuraux de chaque sous-unité sont connues, ainsi la structure de SAGA HAT humain n'est pas résolue à ce jour. Dans ce contexte, l'approche de CX-MS apparaît comme une alternative intéressante : nous espérons que les informations obtenues par CX-MS permettront d'affiner les résultats de l'EM en combinant avec la modélisation moléculaire.

L'objectif de cette étude est donc de mettre en place un protocole de pontage couplé à la MS afin de vérifier la stœchiométrie de SAGA HAT estimée par filtration sur gel et d'aborder l'analyse topologique en complémentarité avec d'autres techniques d'analyse structurale (modélisation, microscopie électronique). Ainsi, la combinaison de ces techniques permettra de

proposer des hypothèses quant aux zones de contact entre les partenaires et à l'agencement des sous-unités les unes par rapport aux autres, voire de proposer un modèle structural à basse résolution de SAGA HAT.

Les étapes de l'approche CX-MS utilisée dans ce projet sont représentées en Figure 2.

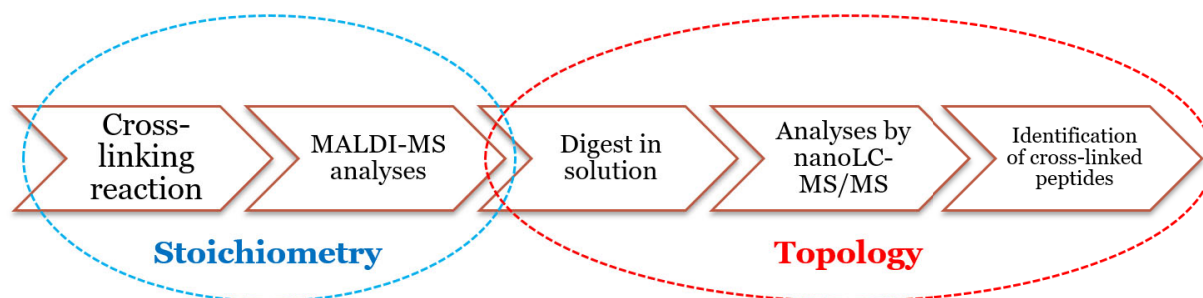


Figure 2. Diagramme du processus d'analyse utilisé pour déterminer la stœchiométrie et la topologie du complexe SAGA HAT.

3.2. Etudes préliminaires avec la BSA

Les études préliminaires ont été conduites avec la BSA (albumine sérique bovine) qui est disponible commercialement et en abondance par rapport à l'échantillon SAGA HAT. Il fait également partie des espèces utilisées comme calibrant en MALDI-MS en mode positif grâce à ses pics monochargé (66 431 Da) et dichargé (33 216 Da) du monomère, ainsi que le pic monochargé du dimère (132 861 Da). L'acide sinapinique (SA) communément utilisé pour la détection des biomolécules de haut m/z a été choisi comme matrice.

Premièrement, pour une solution de BSA à 150 μM dans l'eau, nous avons comparé les signaux obtenus par le détecteur MCP et par le détecteur de haute masse CovalX (HM1), tous les deux installés sur le même appareil MALDI Autoflex II TOF/TOF (Bruker) fonctionnant en mode linéaire. Il en résulte que l'utilisation du détecteur CovalX améliore grandement la détection du dimère de la BSA à 132 kDa sur le même dépôt, tout comme elle permet un gain de 10 fois sur l'intensité du signal du monomère (Figure 3). En effet, comme expliqué en chapitre 2, ce détecteur basé sur la technologie « ion conversion dynode » (ICD) possède une plus grande sensibilité vis-à-vis des ions de haut m/z . Par conséquent, les analyses qui suivent ont été réalisées avec le détecteur CovalX.

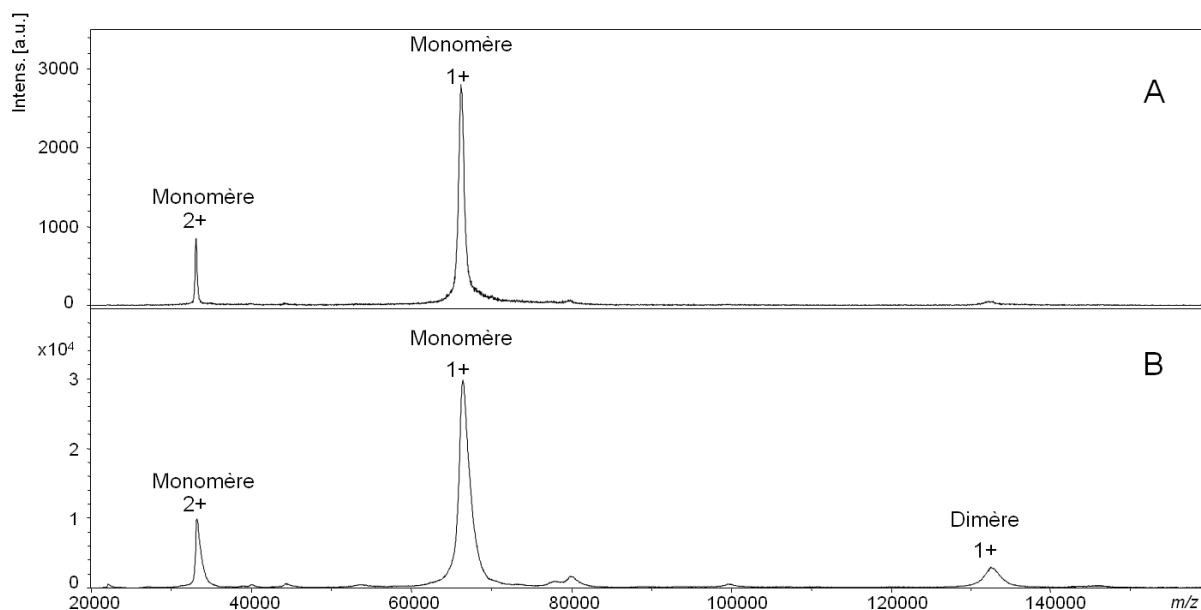


Figure 3. Spectres MALDI-MS de la BSA à 150 μM dans H₂O (matrice SA, dépôt goutte séchée). (A) Avec le détecteur MCP, (B) Avec le détecteur de haute masse CovalX HM1.

Le chlorure de sodium (NaCl) est indispensable pour conserver l'intégrité structurale du sous-complexe SAGA HAT. Par conséquent, des tests ont été réalisés afin d'avoir une idée concernant l'impact de la concentration de NaCl sur la détection des ions en MALDI-MS. Sur la Figure 4 se trouvent les spectres de la BSA à 1 μM dans 20 mM HEPES pH 8,0, sans et avec 100 mM NaCl. La présence du NaCl à 100 mM diminue l'intensité du signal obtenu d'un facteur 2, mais permet encore de détecter un bon signal aux alentours de la gamme de masse de la BSA.

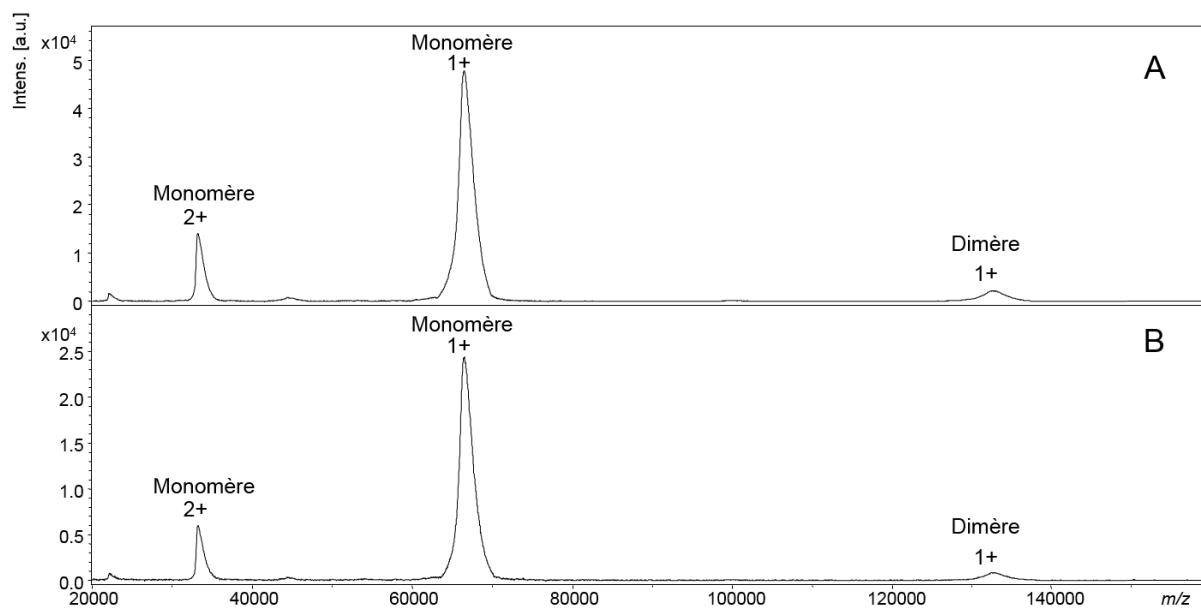


Figure 4. Spectres MALDI-MS de la BSA à 1 μM (matrice SA, dépôt goutte séchée) dans (A) 20mM HEPES, (B) 20mM HEPES + 100mM NaCl.

Les conditions de dépôts MALDI ont été également testées. Parmi les types de dépôts usuels (goutte séchée, couche mince, double couche), il s'est avéré que le dépôt goutte séchée permet d'obtenir les meilleurs résultats. Cela peut s'expliquer par le fait que le dépôt goutte séchée permet une meilleure incorporation de l'échantillon au sein des cristaux de la matrice, d'où une meilleure désorption-ionisation dans le cas de notre étude, et qu'il est plus robuste vis-à-vis un grand nombre de tirs laser comparé aux dépôts couche mince et double couche.

De même, la façon de déposer (mélange matrice-analyte sur cible ou mélange matrice-analyte dans un tube avant de déposer) est également optimisée. Les meilleurs résultats sont obtenus avec le mélange direct sur cible ; dans le cas du mélange préalable, le dépôt homogène est probablement gêné par de la précipitation ou une cristallisation dans le tube.

Nous avons aussi essayé de varier la concentration en matrice, car nous nous demandons si une plus grande concentration de matrice aide à la détection des composés de grandes masses. Ainsi différentes solutions de matrices SA à 5 mg/mL, 10 mg/mL et 20 mg/mL dans H₂O/ACN/0,1% TFA ont été testées (Figure 5). Il en résulte que la solution de SA à 10 mg/mL donne le meilleur signal dans le cadre de cette étude. Ce gain est faible mais nous pouvons nous attendre à ce que l'accumulation des faibles gains améliora l'analyse de SAGA HAT.

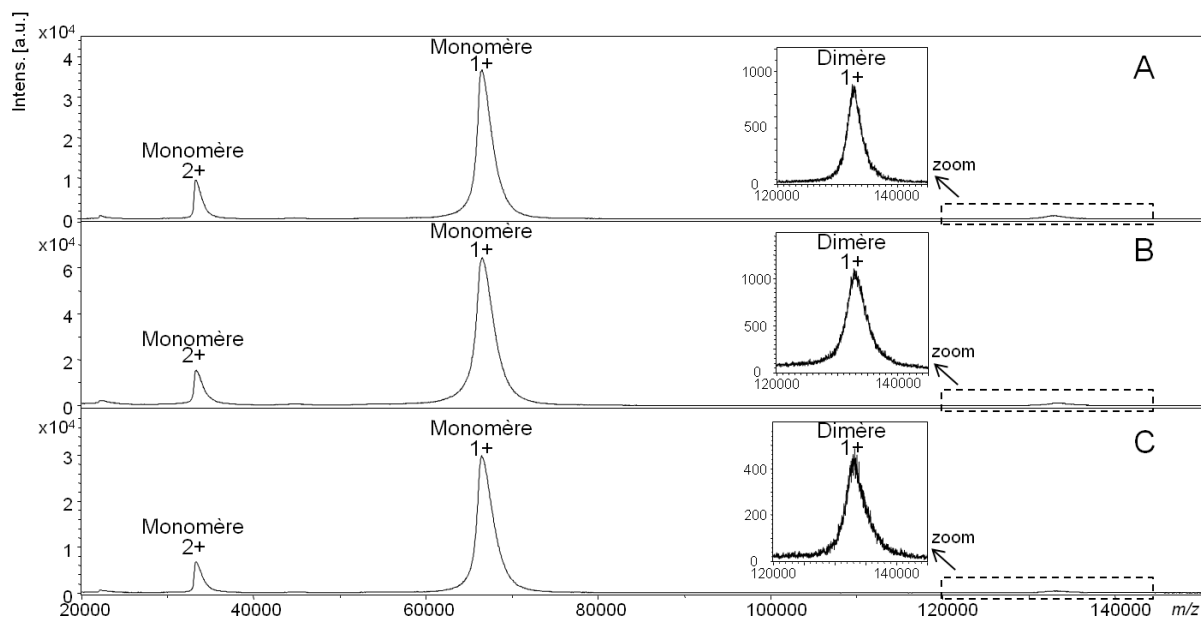


Figure 5. Spectres MALDI-MS de la BSA à 0,5 μ M dans H₂O avec différentes concentrations de matrice SA, dépôt goutte séchée. (A) SA à 5 mg/mL, (B) SA à 10 mg/mL, (C) SA à 20 mg/mL.

Des tests ultérieurs de solvant de matrice plus acidifié (avec 0,3% TFA au lieu de 0,1% TFA) donnent également de meilleurs signaux que dans H₂O/ACN/0,1% TFA, ceci grâce à une meilleure protonation et ionisation de l'échantillon.

En résumé, à partir de ces études préliminaires et pour la suite du projet, nous utilisons la matrice SA à une concentration de 10 mg/mL dans H₂O/ACN/0,3% TFA avec le mode de dépôt goutte séchée. Dans ces conditions, la Figure 6 montre le spectre MALDI-MS obtenu pour le complexe SAGA HAT avant le pontage, où sont détectés seulement les ions à 33 197, 48 773 et

97 197 ± 10 Da correspondant aux sous-unités séparées. L'ion à 33 197 Da correspond à SGF29, l'ion à 48 773 Da correspond à ADA2B et ADA3 qui sont co-détectées, et l'ion à 97 197 Da correspond à GCN5. Comme attendu, sans pontage, le complexe est dissocié et n'est pas détecté en MALDI-MS. Cependant, les conditions d'analyse sont désormais optimisées et seront utilisées par la suite lors du pontage.

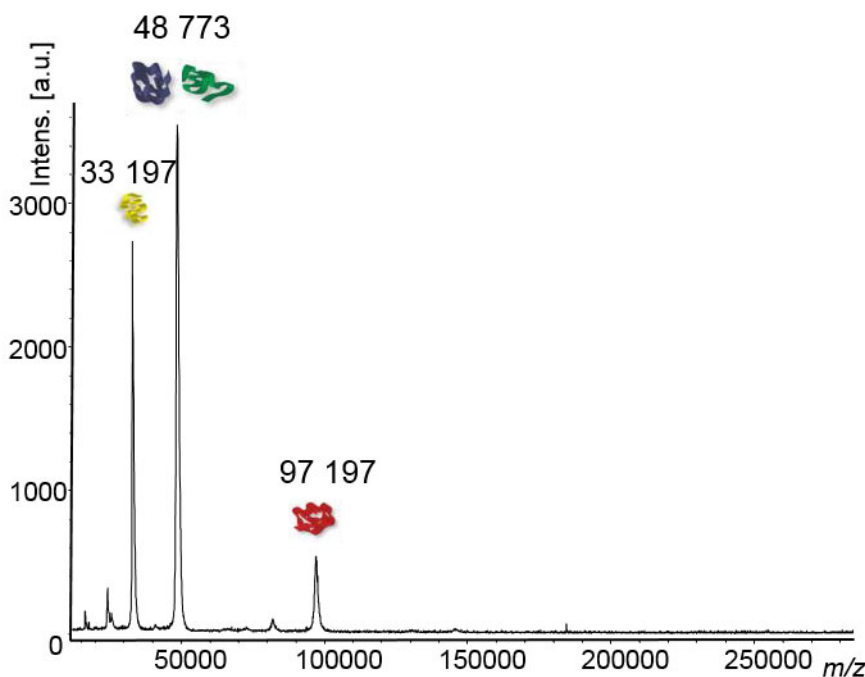


Figure 6. Spectre MALDI-MS du complexe SAGA HAT à 8 μM dans 20mM HEPES + 100mM NaCl avant le pontage (matrice SA, dépôt goutte séchée).

3.3. Mise au point de la réaction de pontage

3.3.1. Choix de l'agent de pontage

3.3.1.1. Le glutaraldéhyde

Le glutaraldéhyde ($\text{C}_5\text{H}_8\text{O}_2$, masse molaire 100,12 g/mol, N° CAS 111-30-8), reporté comme agent de pontage dans les années 60, est soluble dans l'eau, l'alcool et d'autres solvants organiques. Son caractère hydrosoluble convient au pontage des échantillons biologiques. Il est reconnu comme un agent de pontage universel, très efficace car il réagit sur plusieurs fonctions telles qu'amine, thiol, phénol, et imidazole [3]. Il est présent sous au moins 13 formes monomériques et polymériques (Figure 7). Aussi en raison de sa polymérisation par propagation intramoléculaire-intermoléculaire, le pontage par glutaraldéhyde risque de ne pas être spécifique. Les mécanismes de pontage par glutaraldéhyde, ou plutôt par ses différentes formes mono- et

polymériques, sont complexes (les principales réactions reportées sont de type addition de Michael et réaction de Schiff) et toujours en débat [4,5]. De plus, l'efficacité, les formes de glutaraldéhyde présentes en solution, ainsi que les réactions ayant lieu dépendent du pH du milieu.

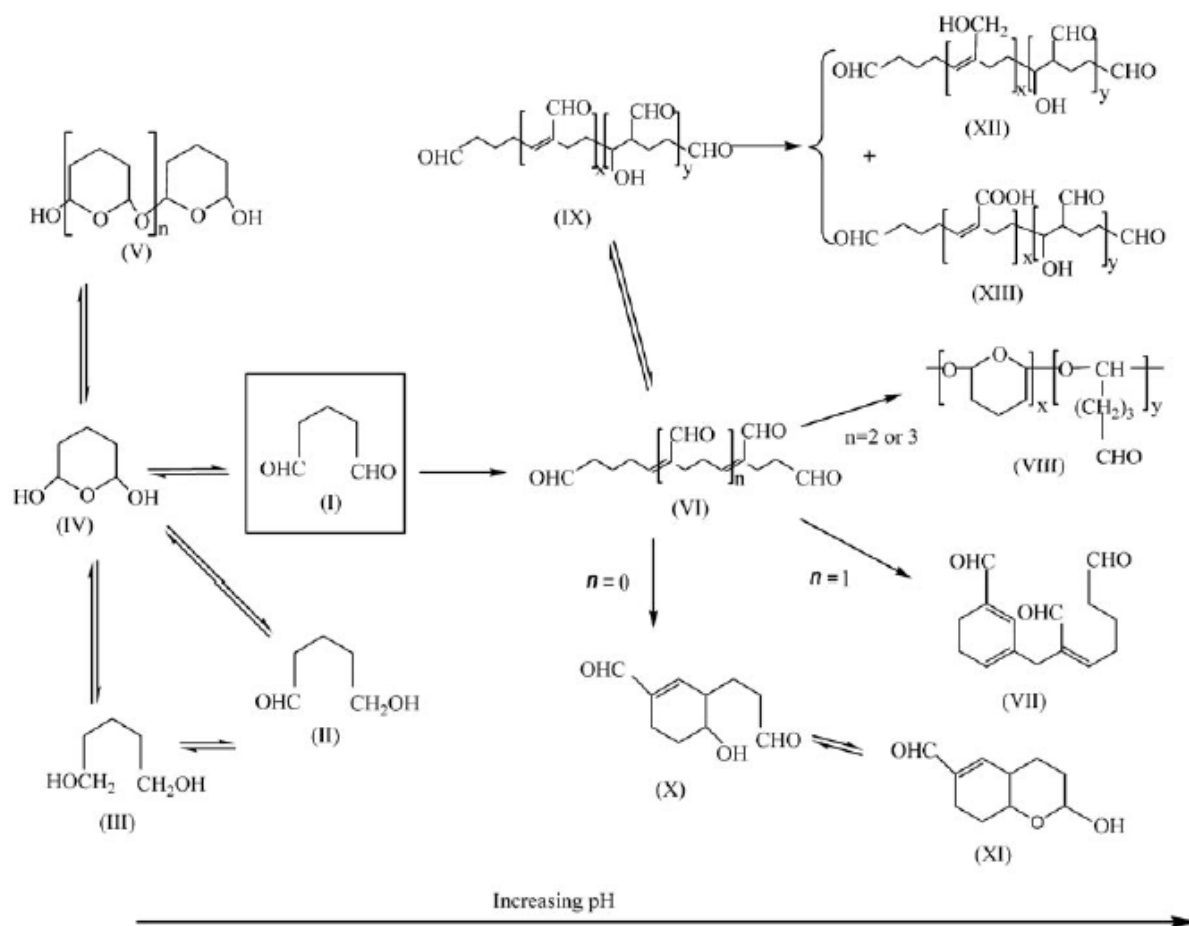


Figure 7. Les 13 formes possibles du glutaraldéhyde en milieu aqueux [4].

Les études préliminaires de pontage au cours de ces travaux de thèse ont été commencées avec le glutaraldéhyde en raison de sa très grande efficacité et de son utilisation réussie en microscopie électronique par nos collaborateurs. Les temps de réaction de pontage avec le glutaraldéhyde sont très courts, de quelques minutes en général. Pour le pontage du SAGA HAT, les quantités de glutaraldéhyde que nous avons testées allant de 0,01% à 0,5% en volume. Les températures de pontage peuvent être à 4°C ou à température ambiante. Pour arrêter la réaction de pontage, nous ajoutons au milieu réactionnel du Tris à 1 M. Un exemple de spectre du SAGA HAT ponté avec du glutaraldéhyde est montré en Figure 8.

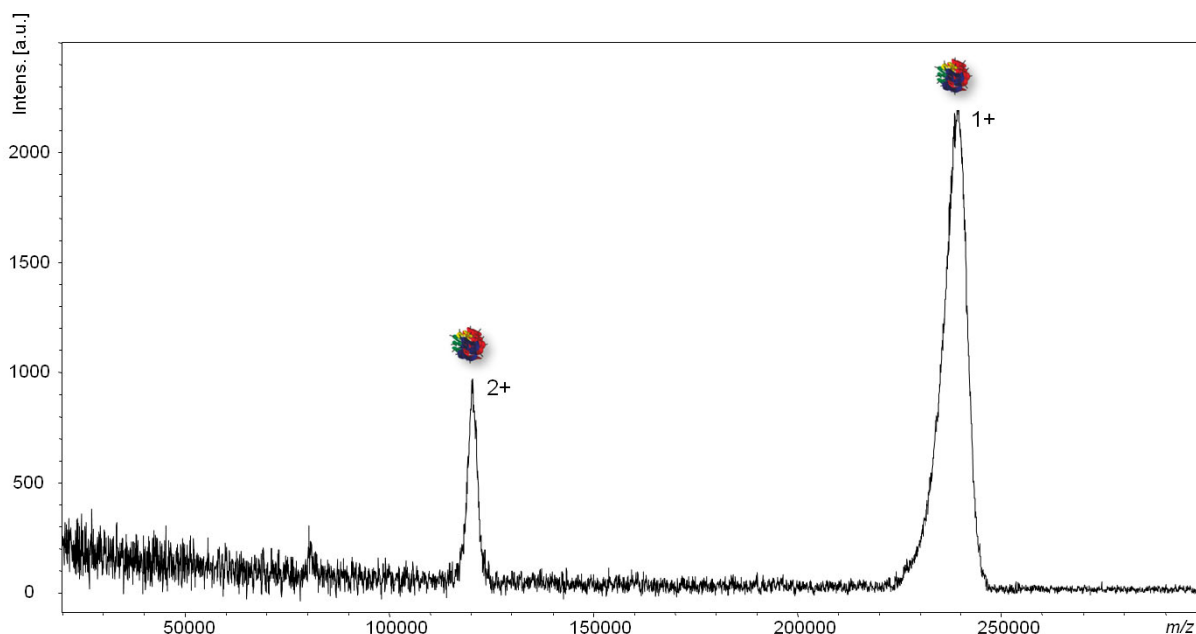


Figure 8. Spectre MALDI-MS du complexe SAGA HAT à 8 μM dans 20mM HEPES + 100mM NaCl après le pontage avec le glutaraldéhyde (matrice SA, dépôt goutte séchée). La réaction de pontage a été réalisée avec du glutaraldéhyde à 0,05 % pendant 2 min à température ambiante.

Cependant, la polymérisation du glutaraldéhyde ainsi que son manque de spécificité limitent ses applications dans la détermination structurale. Les formes polymériques du glutaraldéhyde génèrent beaucoup d'hétérogénéités dans le complexe ponté, ce qui se traduit par un élargissement des pics et une diminution de l'intensité du signal sur le spectre de masse. De plus, elles rendent difficile l'identification des peptides modifiés par le pontage ainsi que l'interprétation des distances entre les groupements pontés [6]. Ainsi, afin d'avoir plus de spécificité et plus d'informations sur les zones en interaction, nous avons choisi de continuer les études de pontage avec un autre agent plus spécifique.

3.3.1.2. Le bis[sulfosuccinimidyl] suberate (BS3)

Parmi les agents de pontage de la famille des esters NHS, nous avons sélectionné le BS3 (Figure 9). En effet, cet agent de pontage portant le groupe réactif succinimidyl offre plusieurs avantages pour démarrer une étude :

- Il cible les amines des lysines et du N-terminal. Comme développé précédemment dans le chapitre 2, ces derniers sont les plus ciblés en raison de leur haute réactivité, leur emplacement près de la surface des protéines hydrosolubles, et leur présence à une fréquence assez élevée sur la séquence peptidique. Dans le cadre de notre étude, le sous-complexe SAGA HAT est constitué justement de protéines hydrosolubles, qui possèdent en tout 164 lysines.

- Il est disponible commercialement, sous formes hydrosoluble (BS3) et organosoluble (DSS, disuccinimidyl suberate).
- Différentes longueurs du bras espaceur existent : par exemple, cette longueur est de 11,4 Å pour le BS3 et 7,7 Å pour son homologue BS2G (bis[sulfosuccinimidyl] glutarate).
- Il existe sous forme d'agent de pontage léger (d0) ou lourd (d4, où 4 atomes d'hydrogène sont remplacés par 4 atomes de deutérium). Cela offre la possibilité de marquage isotopique facilitant la détection et l'interprétation des peptides pontés.
- La réussite de la réaction de pontage réalisée avec le BS3 a été reportée dans plusieurs études [7-11].

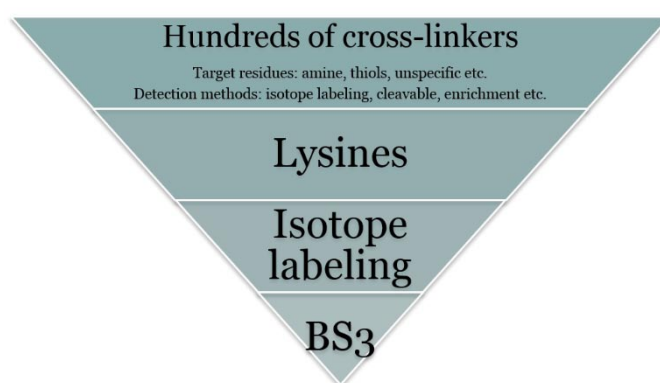


Figure 9. Choix de l'agent de pontage BS3 parmi les agents de pontage existants.

Le BS3 (formule chimique $C_{16}H_{18}N_2Na_2O_{14}S_2$, masse molaire 572,43 g/mol, N° CAS 82436-77-9, longueur du bras espaceur 11,4 Å) est un agent de pontage homobifonctionnel dont la fonction réactive est l'ester N-hydroxysuccinimide (NHS). Il est soluble en milieu aqueux, mais est également hydrolysé dans l'eau et se dégrade en milieu basique. Le temps de demi-vie des esters NHS est de l'ordre de l'heure dans un milieu aqueux à un pH physiologique (pH 7.0-7.5) et diminue avec l'augmentation du pH [12-14]. Sa formule chimique est représentée en Figure 10.

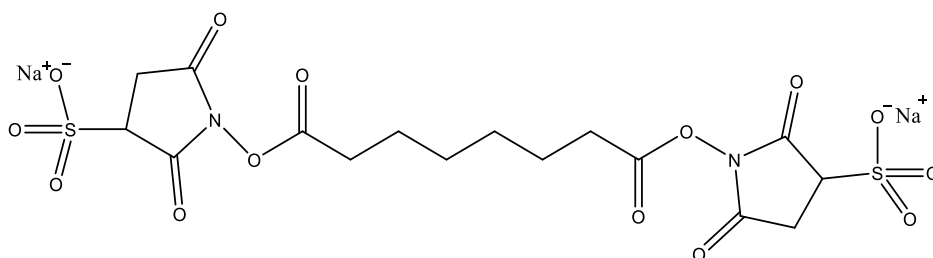


Figure 10. Formule chimique du BS3.

Il réagit essentiellement sur les amines primaires comme des lysines et les N-terminaux par substitution nucléophile dans la gamme de pH 7-9 (Figure 11), mais est aussi capable de réagir sur d'autres acides aminés en fonction du pH. Il est rapporté que des réactions secondaires peuvent avoir lieu entre le BS3 et les tyrosines, histidine, arginines [15]; toutefois le BS3 reste spécifique aux lysines tant que le pH est contrôlé par un tampon adéquat.

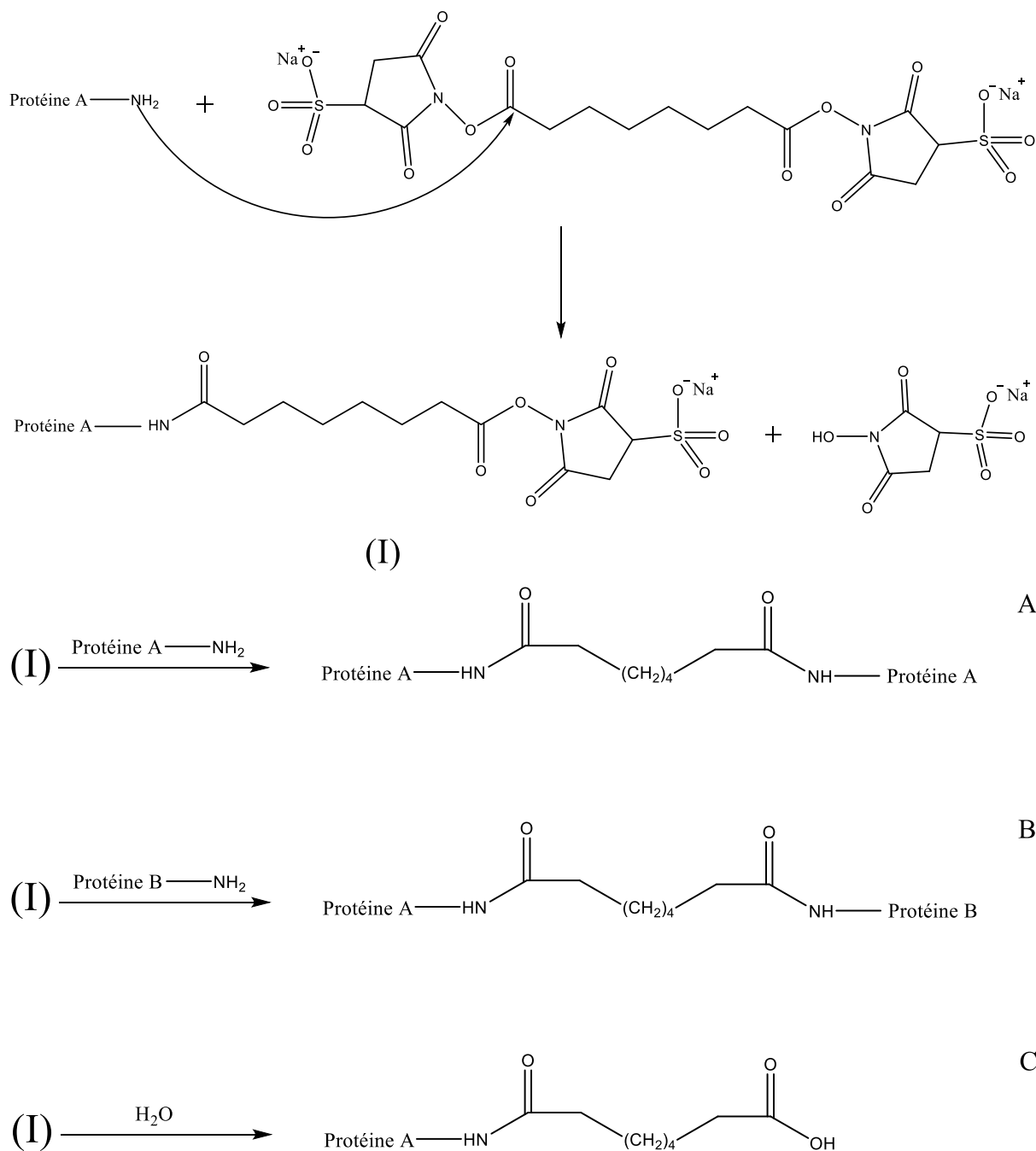


Figure 11. Réaction du BS3 sur les protéines. (A). Pont intra-peptidique, (B). Pont inter-peptidique, (C). Mono-link.

La longueur du bras espaceur du BS3 est de 11,4 Å ; en combinant cette distance avec la longueur de la chaîne latérale des lysines (6–6,5 Å) et la flexibilité des protéines en solution, la distance maximale entre les deux Cα des protéines susceptibles d'être pontés est de l'ordre de 30 Å [11,16].

Dans le cadre de notre étude, nous avons utilisé un mélange équimolaire de BS3 léger (BS3-d0) et BS3 lourd (BS3-d4, où 4 atomes d'hydrogène sont remplacés par 4 atomes de deutérium comme représentés sur la Figure 12). Ce marquage isotopique permet de mieux détecter et identifier les peptides pontés par la suite : certains seront doublement identifiés par les formes pontées avec l'agent de pontage léger et lourd. Cependant, le risque est de perdre en intensité car la quantité de ces peptides sera partagée entre ces deux formes.

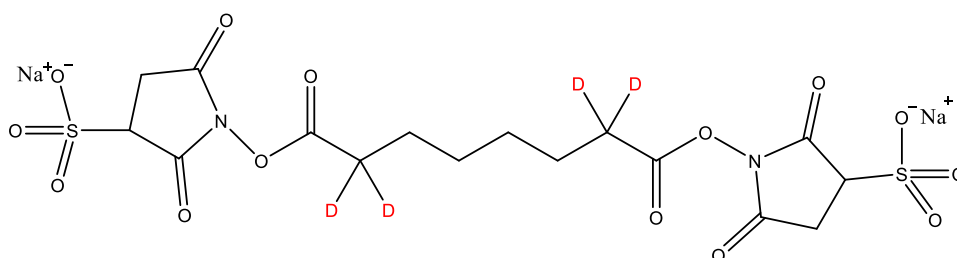


Figure 12. Formule chimique du BS3-d4.

Les trois produits de réaction peuvent apporter des informations utiles à la construction d'un modèle 3D du complexe. Dans un premier temps, nous nous sommes concentrés sur les ponts interprotéiques. Les « mono-links » peuvent être reliés à l'accessibilité au solvant mais peuvent aussi provenir d'un pontage incomplet où l'hydrolyse est en compétition avec la réaction de pontage.

3.3.2. Optimisation des conditions expérimentales

Dans le cadre de notre étude, nous avons conduit la réaction de pontage sur une solution de SAGA HAT d'environ 2 mg/mL (8 μM). Cette concentration d'échantillon est nécessaire pour une bonne détection en MALDI-MS et surtout une bonne détection de peptides pontés après digestion.

A la sortie de la purification et pour la conservation, le complexe SAGA HAT se trouve dans un tampon contenant 20 mM HEPES-KOH pH 8,0, 500 mM NaCl, 10% glycérol, 1 mM DTT, 0,1 mM EDTA. Même si le MALDI-MS est tolérant vis-à-vis des sels, il est important de limiter leur concentration au minimum afin d'assurer une bonne détection du signal de l'échantillon. A partir des études préliminaires sur la BSA en 3.2, nous avons vu qu'il y a possibilité de détecter un bon signal dans le tampon contenant 20mM HEPES pH 8,0 et 100 mM NaCl, compatible avec la conservation de l'intégrité de SAGA HAT selon nos collaborateurs.

La Figure 13 présente des spectres MALDI-MS du complexe SAGA HAT ponté avec différentes options d'échange de tampon. Tout d'abord, dans le tampon d'origine, aucun signal n'est détecté (Figure 13A). Ceci peut être expliqué par la présence du glycérol, visqueux et faiblement volatil, qui perturbe la cristallisation de la matrice avec l'échantillon [17], et confirme la nécessité de réaliser un échange de tampon avant la détection en MALDI-MS. Ensuite, il est important de savoir quelle option d'échange de tampon (avant ou après pontage) permet une meilleure détection du complexe ponté :

- Avec un échange de tampon seulement après pontage, nous n'observons aucun signal (Figure 13B). Notre hypothèse est que le glycérol serait « emprisonné » dans le complexe ponté et ne pourrait plus être éliminé correctement lors de l'échange de tampon. Il serait intéressant de conduire des études plus approfondies pour vérifier si ceci se reproduit sur d'autres systèmes biologiques.
- Un échange de tampon avant pontage permet une bonne détection du signal du complexe ponté grâce à une bonne élimination du glycérol et des sels (Figure 13C).
- Enfin, un échange de tampon avant et après pontage n'améliore pas le signal (Figure 13D), probablement car le gain dû à l'élimination du sel résiduel est contrecarré par la perte de l'échantillon lors de l'échange.

Ces résultats nous ont conduits à choisir un échange de tampon avant pontage.

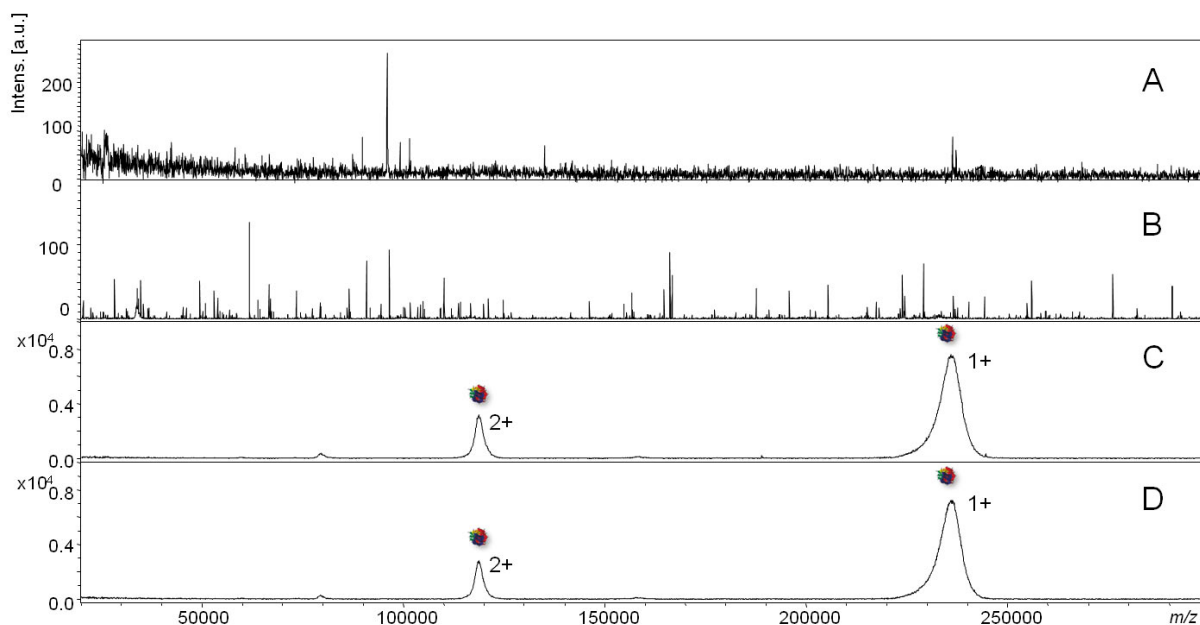


Figure 13. Spectres MALDI-MS du complexe SAGA HAT ponté avec différentes options d'échange de tampon dans 20mM HEPES pH 8,0 + 100 mM NaCl (matrice SA, dépôt goutte séchée). (A) Sans échange de tampon, (B) Echange de tampon après pontage, (C) Echange de tampon avant pontage, (D) Echange de tampon avant et après pontage.

Afin d'avoir un bon rendement de pontage en évitant des agrégations non-spécifiques, une optimisation des conditions de réaction est également indispensable.

Nous avons testé le pontage avec le BS3 aux différents ratios molaires protéine/BS3 (1:50, 1:75, 1:200) et aux différents temps de réaction (30 min, 2 h). La Figure 14 montre les spectres obtenus pour un temps de réaction de 30 min. Au ratio 1 : 50, des pics correspondant aux sous-unités libres sont encore observés. Au ratio 1 : 75, nous ne détectons plus de sous-unités mais seulement des pics liés au complexe ponté. Au ratio 1 : 200, le grand excès d'agent de pontage induit beaucoup d'hétérogénéités dans les espèces pontées menant à un élargissement des pics du complexe et une diminution de l'intensité.

Ainsi, nous avons sélectionné les conditions de pontage du SAGA HAT avec le BS3 à un ratio protéine/BS3 1 : 75 (ce qui est équivalent à un ratio molaire de lysine/BS3 de 1:0,5 car SAGA HAT contient 164 lysines en tout) et un temps de réaction de 30 min à température ambiante. Ceci est en accord avec les conditions de réaction typiquement observées pour le pontage avec les agents de type ester NHS (concentration de protéine de l'ordre de la micromolaire, 20 à 1000 fois en excès d'agents de pontage, temps de réaction d'environ 30 min à température ambiante ou 2 h dans la glace [18]). La réaction de pontage est arrêtée avec une solution de Tris à 25mM dans le volume final.

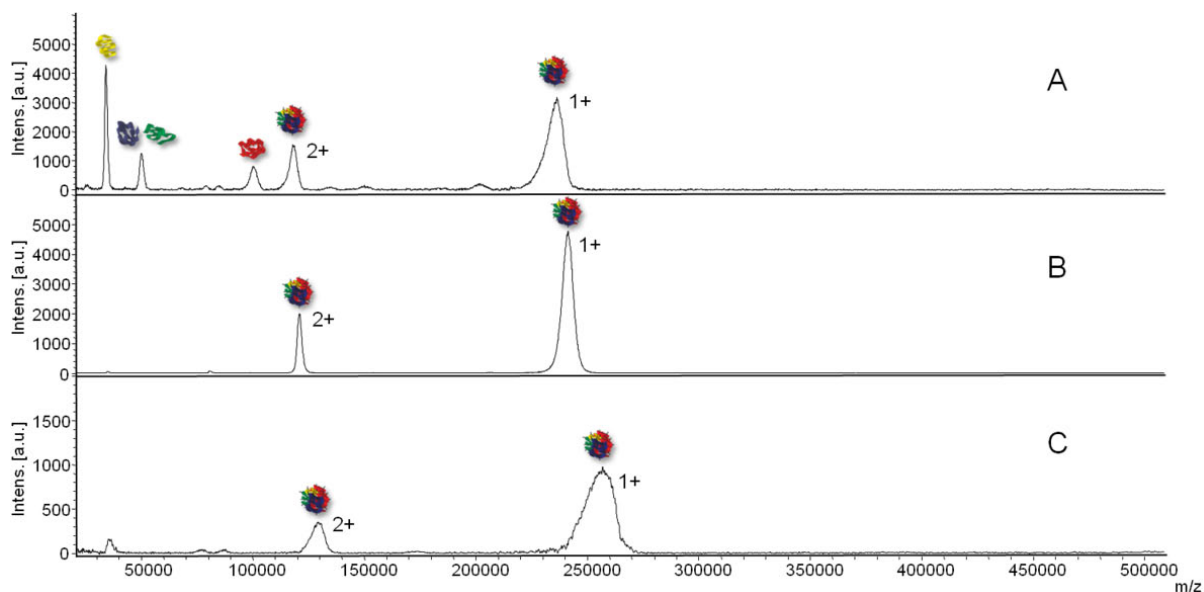


Figure 14. Spectres MALDI-MS du SAGA HAT ponté avec différents ratios protéine/BS3 et un temps de réaction de 30 min (matrice SA, dépôt goutte séchée). (A) Protéine/BS3 1 : 50, (B) Protéine/BS3 1 : 75, (C) Protéine/BS3 1 : 200. De ces résultats, les conditions optimales de pontage ont été fixées à un ratio 1 : 75 [19].

Enfin, comme toute étude par MALDI-MS, la puissance du laser et le nombre de tirs ont également été optimisés afin d'arriver à un bon compromis entre l'intensité du signal et la largeur des pics observés. Dans notre cas, le meilleur signal est obtenu avec une puissance laser de 53% et une accumulation de 3000 tirs.

Après le pontage, une digestion tryptique en solution a été réalisée pour générer des produits pontés. Là encore, la quantité de trypsine et le temps de digestion ont été optimisés ; sont retenus un ratio d'enzyme/protéine de 1 : 5 et une incubation à 37°C pendant 3h.

Après digestion, l'échantillon est analysé en nanoLC-MS/MS dans un spectromètre de masse LTQ Orbitrap Elite (Thermo Scientific) en mode positif couplé à un nanoRSLC U3000 (Thermo Scientific). 1 pmol de l'échantillon est injecté sur une pré-colonne (C18 Dionex Acclaim PepMap100, 75 µm i.d. x 2 cm, 3 µm, 100 Å) avec H₂O/ACN/acide formique (98/2/0,1) à 5 µL/min pendant 3,5 minutes. Ensuite, la séparation chromatographique a lieu sur une colonne C18 (Thermo Scientific Accucore, 75 µm i.d. x 50 cm, 2.6 µm, 150 Å) à 40°C avec 0,1% acide formique dans l'eau (phase mobile A) et 0,08% acide formique dans 80% ACN (phase mobile B). Le gradient dure 150 min (5 à 50% en phase B pendant 120 min, 50 à 99% en phase B pendant 1 min, 99% en phase B pendant 10 min, 99 à 5% en phase B pendant 1 min, 5% en phase B pendant 18 min) avec un débit de 200 nL/min. Puis les peptides sont analysés sur l'Orbitrap Elite en fragmentation HCD. L'accumulation des ions a été fixée à 100 ms et 4×10^4 ions. Les précurseurs de charge +1 sont exclus. La fenêtre d'isolation des ions précurseurs a été fixée à 2 Da, l'énergie de collision HCD à 30% pour la fragmentation. Le spectromètre de masse fonctionne en mode « full scan » dans la gamme de 300–1200 m/z avec une résolution de 240 000 à 400 m/z , puis les quinze ions les plus abondants sont soumis à l'analyse MS/MS dans l'orbitrap avec une résolution de 15 000. Les peptides sélectionnés pour la MS/MS sont exclus pendant 30 s. Enfin, les données brutes de masse sont converties en format MGF et MZML par le logiciel Proteome Discoverer™ (Thermo Scientific, version 1.4) pour l'identification.

3.4. Choix et utilisation des logiciels d'identification

Comme détaillé dans le chapitre précédent, il existe une multitude de logiciels dédiés à l'identification des peptides pontés développés dans des laboratoires de recherche. L'équipe de David Goodlett a répertorié les caractéristiques de quelques-uns de ces logiciels dans le Tableau 1. Dans notre étude, nous avons utilisé en parallèle pLink [20] et xQuest [21]. La différence dans leur algorithme de recherche permet d'augmenter la confiance et est très utile pour vérifier leur bon fonctionnement.

Tableau 1. Les propriétés et applications de quelques logiciels dédiés à l'identification des peptides pontés. (*) Basé sur la première publication, () A l'échelle du protéome. Adapté de [22].**

Nom du logiciel	Applic- ation (*)	Type de données		Base de données		Calcul de score	Réf- érence	Site web
		MS	MS ²	Petite	Large (**)			
GPMaw	Marqué	✓	✓	✓	-	-	[23]	http://www.gpmaw.com
xComb	Label-free	-	✓	✓	-	-	[24]	https://goodlett.umaryland.edu/xcomb.php
X!Link	Label-free	-	✓	✓	-	✓	[25]	Contact du développeur (yjlee@iastate.edu)
XLink-Identifieur	Label-free	-	✓	✓	-	✓	[26]	http://www.du-lab.org
StavroX/ MeroX	Label-free	✓	✓	✓	-	✓	[27,28]	http://www.stavrox.com
BLinks	Label-free	✓	✓	✓	-	✓	[29]	http://brucelab.gs.washington.edu/BLinks.php
pLink	Label-free	✓	✓	✓	✓	✓	[20]	http://pfind.ict.ac.cn/software/pLink
xQuest/ xProphet	Marqué	✓	✓	✓	✓	✓	[21]	http://proteomics.ethz.ch/cgi-bin/xquest2.cgi/index.cgi
XiQ	Marqué	✓	-	-	-	-	[11]	http://129.215.14.65/xiq

3.4.1. Logiciel pLink

Le logiciel pLink est développé au sein de l'équipe de Si-Min He, à l'Institute of Computing Technology (Chinese Academy of Sciences, Pékin, Chine). Il fonctionne sous le système d'exploitation Windows, et est téléchargeable à partir du lien suivant : <http://pfind.ict.ac.cn/software/pLink>. Pour l'utiliser, il faut ouvrir la console sous Windows et entrer les commandes.

Durant ces travaux de thèse, la version 1.15 du logiciel a été utilisée. Il possède l'outil pLabel permettant de visualiser l'interprétation des spectres MS/MS. Son algorithme (Figure 15) permet d'identifier des peptides qui sont pontés par un agent de pontage lourd ou léger. Les données à traiter doivent être sous format MGF.

Le fonctionnement de pLink peut être résumé comme suit : parmi les spectres de masse soumis à l'interrogation (« raw data »), le logiciel commence par faire un tri pour n'en considérer que ceux de qualité susceptibles d'être identifiés (« spectra quality filtering »). S'ensuit une étape de traitement des spectres (« preprocessing of spectra ») où les pics correspondants aux ions précurseurs, aux pertes de H₂O et NH₃, les pics du bruit de fond, *etc.* sont mis de côté et ne seront pas considérés dans l'identification. Ensuite, avant la recherche de toutes les combinaisons de peptide ponté α - β (où le peptide α est ponté au peptide β), se trouve l'étape de « prescoring » où pLink regarde les séquences possibles de α séparément en considérant β comme une modification sur la lysine (et *vice versa*) et attribue un score à chaque identification. Les 500 premiers candidats α et β ayant les meilleurs scores seront considérés pour être appariés en paires de peptides pontés. Puis le logiciel identifie la séquence entière du peptide ponté et attribue le score affiné

(« fine scoring »). Dans les étapes de « prescoring » et « fine scoring », les données expérimentales issues du « preprocessing » sont comparées aux données théoriques issues de la base de données (« protein database »). Pour les systèmes peu complexes, l'étape de « prescoring » peut être omise. Ensuite, le score de l'étape « fine scoring » est utilisé pour calculer la valeur de « e-value » pour chaque peptide ponté identifié. Enfin, le logiciel évalue le FDR (« False Discovery Rate »), crée le rapport d'identification et annote les spectres.

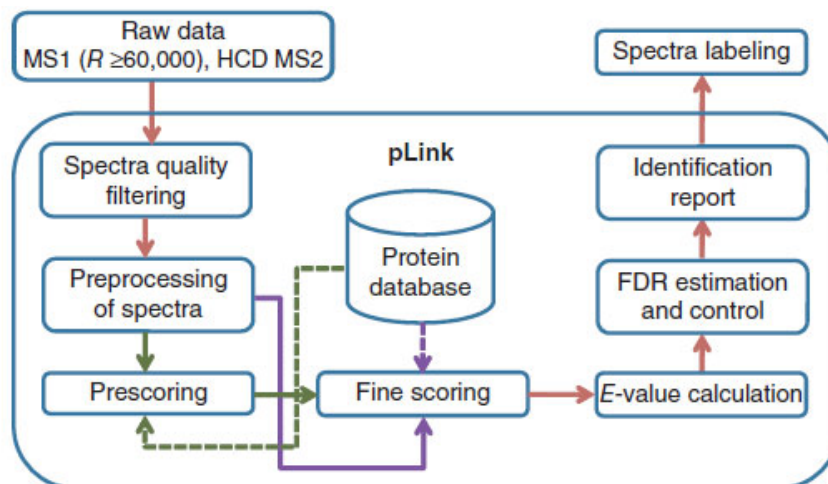


Figure 15. Représentation schématique de l'algorithme de pLink [20].

Lors de l'utilisation de pLink, les données nécessaires pour la recherche sont présentes par défaut dans des fichiers .ini (« enzyme.ini » pour les enzymes, « instrument.ini » pour les paramètres instrumentaux, « xlink.ini » pour les acides aminés cibles des agents de pontage *etc.*). Puis les paramètres de recherche sont entrés par l'utilisateur dans le fichier « pLink.ini ». Enfin, il faut ouvrir la console sous Windows pour lancer l'exécution du « plink.ini ». A la fin de la recherche, le dossier de résultats est généré automatiquement.

Dans le cadre de notre étude, le temps de recherche de pLink est de quelques heures pour environ 20 000 spectres MS/MS avec un ordinateur de bureau fonctionnant sous Windows XP (Windows XP Professional Version 2002 Service Pack 3, 32 bits, Intel Core 2 Duo CPU E6550 @ 2.33GHz, 2.97Go RAM).

3.4.2. Logiciel xQuest

Le logiciel xQuest est développé au sein de l'équipe de Ruedi Aebersold à l'Institute of Molecular Systems Biology (ETH Zürich, Suisse). La version téléchargeable sur <http://proteomics.ethz.ch/cgi-bin/xquest2.cgi/index.cgi> fonctionne sous système d'exploitation Linux. Dans notre cas avec un ordinateur fonctionnant sous Windows, nous avons téléchargé au préalable une machine virtuelle telle que VMware Player pour ce faire.

Son algorithme repose sur l'identification des peptides qui sont pontés à la fois par un agent de pontage léger et un agent de pontage lourd (marqué isotopiquement). Il est donc nécessaire d'utiliser au préalable un mélange d'agents de pontage non marqué et marqué. En présence d'un mélange équimolaire des agents de pontages lourd et léger, un peptide a autant de chance de réagir avec un agent de pontage léger qu'un agent de pontage lourd. Par conséquent, les peptides pontés fourniront des ions dont la différence de masse correspond au marquage. Dans notre cas, nous avons utilisé un mélange de BS3 léger et BS3 lourd où 4 atomes d'hydrogène ont été remplacés par 4 atomes de deutérium, le peptide ponté avec l'agent de pontage lourd aura une différence de masse de 4,0246 Da par rapport au même peptide mais ponté avec l'agent de pontage léger.

Sur la Figure 16 est représenté le fonctionnement de xQuest sur les paires de peptides modifiées avec les agents de pontage lourds et léger. (i) Les spectres MS/MS sont comparés, les ions qui sont présents dans les deux spectres (en vert) sont nommés ions communs (« common ions »), les ions présentant une différence de masse correspondant au marquage (en rouge) sont nommés « cross-link ions ». (ii) Ces deux groupes d'ions sont séparés. (iii) Les « common ions » sont utilisés pour lister les séquences de peptides possibles. (iv) Les peptides candidats issus de l'étape (iii) sont identifiés en considérant l'ensemble des « common ions » et les « cross-link ions ». (v) Seuls les candidats qui « matchent » avec la masse de l'ion précurseur M_r sont retenus et le spectre est annoté en (vi).

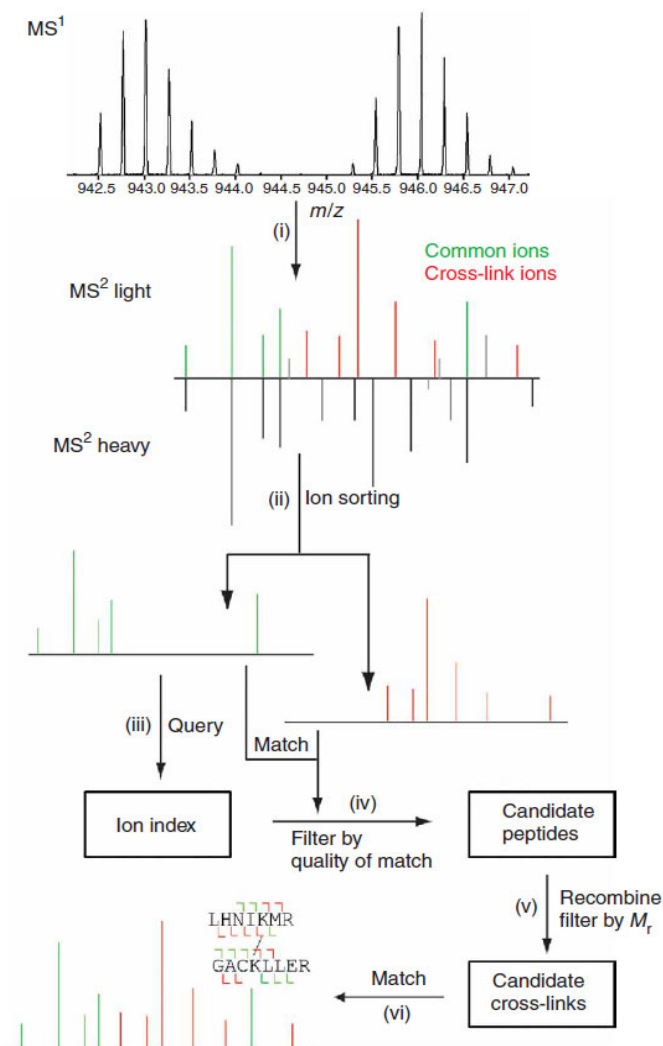


Figure 16. Représentation schématique de l'algorithme de xQuest [21].

Pour être traitées par xQuest, les données MS doivent être sous format MZXML. Si l'utilisation de pLink est plus aisée car le résultat d'identification est généré automatiquement, l'utilisateur doit intervenir à chaque étape de fonctionnement de xQuest. Brièvement, il faut tout d'abord créer la base de données decoy avec « xdecoy.pl ». Ensuite, dans « xmm.def » et « xquest.def », l'utilisateur modifie des paramètres de recherche comme la tolérance MS et MS/MS, la masse de l'agent de pontage, l'écart de masse entre l'agent de pontage non marqué et marqué *etc.* Il lance ensuite la recherche dans le terminal de commande. Enfin, les résultats de la recherche dans « xquest.xml » doivent être copiés dans le dossier résultat et son accès autorisé dans le terminal de commande.

Dans le cadre de notre étude, le temps de recherche de xQuest est de 1-2 h (plus rapide que pLink) pour environ 20 000 spectres MS/MS avec un ordinateur de bureau fonctionnant sous Windows XP (Windows XP Professional Version 2002 Service Pack 3, 32 bits, Intel Core 2 Duo CPU E6550 @ 2.33GHz, 2.97Go RAM).

La différence principale entre pLink et xQuest est que pLink permet d'identifier des peptides modifiés par un agent de pontage marqué ou non marqué, tandis que xQuest identifie seulement des peptides qui sont pontés à la fois par un agent de pontage non marqué et un agent de pontage marqué (Figure 17).

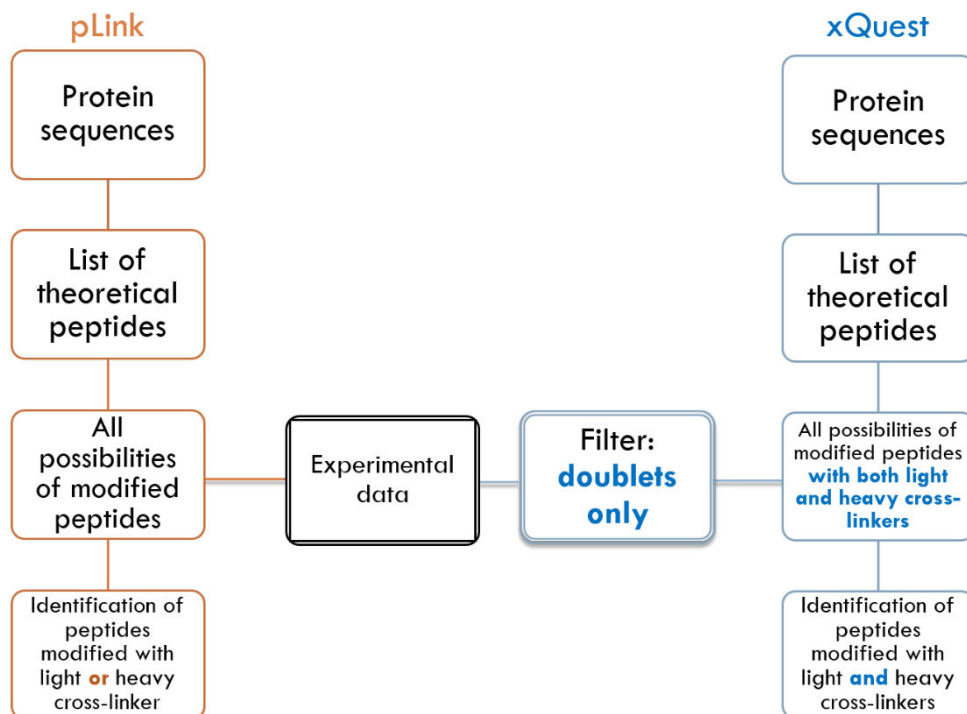


Figure 17. Comparaison des fonctionnements de pLink et xQuest.

3.5. Contraintes, avantages et limites

Premièrement, la réaction de pontage est protéine-dépendante et complexe-dépendante. Dans ce sens, sa mise en œuvre ainsi que son optimisation constitue un processus d'essai-erreur [8,12], afin de trouver la nature d'agent de pontage ainsi que les conditions expérimentales adéquates pour une réaction de pontage réussie.

Deuxièmement, l'utilisation ainsi que l'interprétation des résultats de pLink et xQuest a demandé également beaucoup de temps pour diverses raisons :

- Les formats de données : les données MS doivent être sous forme de fichier .mgf pour pLink, et .mzXML pour xQuest. Pour cela, l'outil MSConvert du logiciel ProteoWizard (version 3.0.4743) a été utilisé pour convertir les fichiers de données MZML au format mzXML (Figure 18).

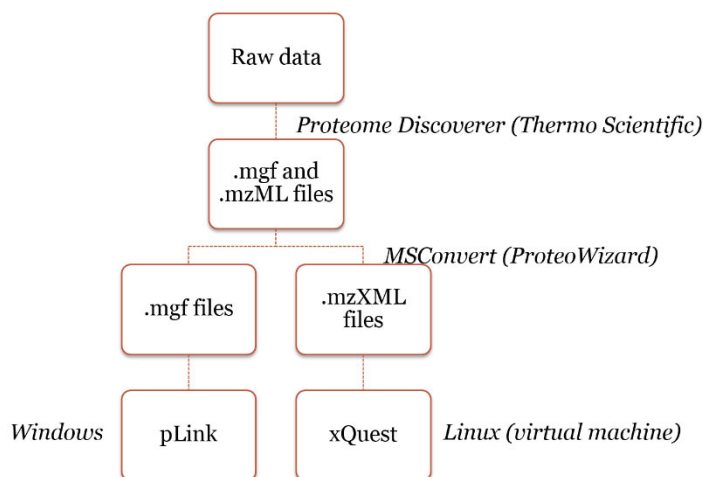


Figure 18. Conversion des formats de données utilisée dans ces travaux de thèse pour l'utilisation de pLink et xQuest.

- Le mode d'utilisation des logiciels : ces deux logiciels ne possèdent pas d'interface d'utilisateur. Pour xQuest, il faut entrer les lignes de commande des étapes une par une dans la console sous Linux (création de la banque de protéine decoy, identification, mise en forme du fichier de résultats). Pour pLink, l'utilisation est plus aisée car l'utilisateur n'est pas tenu à commander chaque étape, néanmoins il faut quand même passer par la console sous Windows. Pour un utilisateur novice peu habitué aux commandes informatiques, ceci pourrait constituer une difficulté. Dans ce cas, l'utilisation parallèle de ces deux logiciels permet entre autres de vérifier leur bon fonctionnement.
- Détermination de seuils de score à partir desquels les identifications sont considérées comme fiables : en effet, il n'existe pas de « règle d'or » ou de critères normalisés concernant les seuils de confiance. C'est pourquoi plusieurs essais ont été effectués pour déterminer nous-même ces seuils. Pour cela, nous avons répété les identifications, mais avec une base de données contenant les séquences de protéines aléatoires, ou encore avec une fausse masse d'agent de pontage. De ces tests, dans le cas de notre étude, nous avons défini le score seuil de xQuest à 25 et de pLink à E-06. Il est à noter que ce seuil est relatif, il dépend de la qualité des résultats et des scores obtenus pour les peptides, et est donc à adapter en fonction des données obtenues et des systèmes biologiques étudiés.
- Validation manuelle : même si les logiciels font le plus grand travail d'identification, il est indispensable d'effectuer une validation manuelle de l'identification. Ceci permet de vérifier notamment s'il y a bien au moins 3 acides aminés consécutifs identifiés, et si les fragments identifiés sont de bonne intensité. De cette validation manuelle, dans ce travail de thèse, environ 10% des peptides identifiés ont été éliminés (ce qui est comparable au 14% éliminé par validation manuelle dans une étude similaire de Han *et al.* [30]).

Troisièmement, l'algorithme de recherche de xQuest passe par la création d'un filtre (« filter ») qui ne retient que les spectres des peptides pontés à la fois par les agents de pontage lourd et léger. Néanmoins, ce couple n'est pas toujours détecté par le spectromètre de masse, il peut n'y avoir que des peptides pontés lourds sans leurs homologues légers et *vice versa*. C'est pourquoi l'utilisation parallèle de ce logiciel avec pLink permet de s'assurer de ne pas manquer ces peptides en question.

D'autre part, il est possible que la conformation des complexes étudiés soit flexible. Ainsi, les ponts sont formés en fonction de la conformation de départ à l'état natif du complexe. Dans le cas des complexes flexibles, les mêmes pontages peuvent ne pas exister d'un réplikat à l'autre. Cependant, une zone de proximité n'est en général pas déterminée par un seul mais plusieurs pontages, d'où l'importance de vérifier si les mêmes zones sont retrouvées.

Enfin, il faut garder à l'esprit que les zones de proximités trouvées dépendent de l'agent de pontage choisi (fonctions réactives, longueur du bras espaceur). Par conséquent, il est intéressant d'utiliser différents agents de pontage pour avoir des informations complémentaires. Dans le même objectif, il est également envisageable d'utiliser d'autres protéases que la trypsine, ou d'autres modes de fragmentation comme par exemple l'ETD (« electron-transfer dissociation ») lors d'analyse en MS/MS.

Les résultats de cette étude ont fait l'objet d'une publication dans le journal scientifique à comité de lecture *Protein Science* (Wiley Blackwell, 2015).

3.6. Publication : Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex

Protein Science **2015**, 24 (8), 1232–1246. DOI: 10.1002/pro.2676



Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex

Nha-Thi Nguyen-Huynh,¹ Grigory Sharov,² Clément Potel,^{1†} Pélégie Fichter,² Simon Trowitzsch,³ Imre Berger,³ Valérie Lamour,² Patrick Schultz,² Noëlle Potier,^{1*} and Emmanuelle Leize-Wagner¹

¹Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) UMR 7140 CNRS/Université de Strasbourg - "Chimie de la Matière Complexe", 1 Rue Blaise Pascal, 67008 Strasbourg, France

²Integrated Structural Biology Department, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104, INSERM U964, 1 rue Laurent Fries, 67404, Illkirch, France

³European Molecular Biology Laboratory (EMBL), Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble, France

Received 22 December 2014; Accepted 3 March 2015

DOI: 10.1002/pro.2676

Published online 9 March 2015 proteinscience.org

Abstract: Understanding the way how proteins interact with each other to form transient or stable protein complexes is a key aspect in structural biology. In this study, we combined chemical cross-linking with mass spectrometry to determine the binding stoichiometry and map the protein–protein interaction network of a human SAGA HAT subcomplex. MALDI-MS equipped with high mass detection was used to follow the cross-linking reaction using bis[sulfosuccinimidyl] suberate (BS3) and confirm the heterotetrameric stoichiometry of the specific stabilized subcomplex. Cross-linking with isotopically labeled BS3 d0-d4 followed by trypsin digestion allowed the identification of intra- and intercross-linked peptides using two dedicated search engines: pLink and xQuest. The identified interlinked peptides suggest a strong network of interaction between GCN5, ADA2B and ADA3 subunits; SGF29 is interacting with GCN5 and ADA3 but not with ADA2B. These restraint data were com-

Abbreviations: amu, atomic mass units; BSA, bovine serum albumin; CX, chemical cross-linking; DOPE, discrete optimized protein energy; EM, electron microscopy; ESI-MS, electrospray ionization mass spectrometry; H/D exchange, hydrogen/deuterium exchange; HCD, higher-energy collisional dissociation; IM, ion mobility; IMAC, immobilized metal affinity chromatography; LC-MS/MS, liquid chromatography coupled with mass spectrometry in tandem; MALDI-MS, matrix assisted laser desorption/ionization mass spectrometry; MOI, multiplicity of infection; NMR, nuclear magnetic resonance; PBS, phosphate buffered saline; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC, size exclusion chromatography; TFA, trifluoroacetic acid; XRD, X-ray diffraction.

Additional Supporting Information may be found in the online version of this article.

[†]*Present address:* Clément Potel's present address is Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands
Nha-Thi Nguyen-Huynh and Grigory Sharov contributed equally to this work.

Grant sponsors: Fondation pour la Recherche Médicale, the Institut National de la Santé et de la Recherche Médicale, the Centre National pour la Recherche Scientifique, the Strasbourg University, the Association pour la Recherche sur le Cancer and the Agence Nationale pour la Recherche (DiscoverID grant), the Labex INRT, the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; Grant sponsor: INSTRUMENT as part of the European Strategy Forum on Research Infrastructures (ESFRI); Grant sponsor: European Commission (EC) Framework Programme (FP) 7 project ComplexINC; Grant number: 279039.

*Correspondence to: Noëlle Potier; Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 CNRS/Université de Strasbourg - "Chimie de la Matière Complexe", 1 Rue Blaise Pascal, 67008 Strasbourg, France. E-mail: npotier@unistra.fr

bined to molecular modeling and a low-resolution interacting model for the human SAGA HAT sub-complex could be proposed, illustrating the potential of an integrative strategy using cross-linking and mass spectrometry for addressing the structural architecture of multiprotein complexes.

Keywords: cross-linking mass spectrometry; SAGA HAT subcomplex; proteomics; protein– protein interaction

Introduction

Most if not all biological processes are catalyzed by dedicated interaction networks. Proteins rarely exist alone in the cell but are involved in macromolecular complexes and elucidation of the assembling dynamics of these multiprotein blocks is of major importance to understand and modulate cellular dysfunction.

In the last few years, integrative approaches have attracted considerable interest. This might be partly because of the realization that, despite their own limitations, traditional structural methods have complementary attributes, and combining respective information would definitely help in exploring the architecture, dynamics, and function of macromolecular complexes. Data arising from various biophysical techniques may be gathered on the bound as well as on the free states of the multiprotein system, and may help generating testable structural models or understanding self-assembling of the protein building blocks.^{1,2} Input of physical theories and statistical data arising from sequence or structural databases homology studies, might also be added to best refine the proposed model.³ In this context, mass spectrometry (MS) was shown as a high-potential technique for characterizing intact multiproteins assemblies, especially for determining the complex stoichiometry or monitoring dynamic changes.⁴ More recently, MS-based technologies, such as IM,⁵ H/D exchange,⁶ chemical cross-linking (CX),⁷ have emerged to assist these structural building activities. In particular cross-linking, by providing important spatial restraint information, is emerging as a promising approach for generating residue level topological data.⁸ Revealing precisely which residues are linked yields proximity information that might increase the resolution of the proposed structural model. An increasing number of studies were reported during the last decade,^{9–11} including very large multiprotein complexes such as RNA polymerase II¹² and protein–drug interactions¹³ or illustrating the integration of CX-MS to theoretical structural knowledge arising from modeling,^{14,15} EM,¹⁶ or XRD.¹⁷ The choice of cross-linkers is crucial because the success of cross-linking reaction depends on their reactivity and selectivity as well as their spacer-arm lengths which rule the maximum distance between two cross-linked residues.^{18,19} Identifying the cross-linked peptides is also still very challenging since they often are present at a small concentration in solution compared to unmodified peptides. Indeed, multiple cross-linked products may occur for one reactive residue or the

reaction may be incomplete. Therefore, several enrichment methods that will improve MS detection have been successfully developed including affinity selection,²⁰ strong cation exchange (SCX) chromatography,²¹ H/D labeling or exclusion list of low charge states peptides. Using cross-link reagent bearing a HCCA (α -cyano-4-hydroxycinnamic acid) moiety might also be very efficient allowing specific MALDI signal enhancement of such cross-linked peptides compared to unmodified peptides.^{22,23} Furthermore, the CX-MS workflow requires the use of search engines dedicated to identification of cross-linked peptides. A large number of algorithms are available in the literature^{24–27} but the main challenge still resides in assigning the confidence of a match.

The multisubunit Spt-Ada-Gcn5 Acetyltransferase (SAGA) complex investigated here is a paramount example of a transcriptional coactivator.²⁸ Transcription of protein coding genes by the RNA polymerase II (RNAPII) enzyme is a tightly regulated process that requires the assembly of a Pre-Initiation Complex (PIC) at the promoter of transcribed genes. Sequence-specific transcriptional activators and post-translational modifications of nucleosomal histones contribute to the recruitment of multisubunit coactivator complexes acting as bridging factors between the activators and the PIC.^{29,30} In humans, the 1.8 MDa SAGA coactivator complex is composed of 18 distinct subunits^{31,32} and adopts a modular organization as evidenced by genetic complementation studies,³³ and by electron microscopy models.^{34,35} More recently, quantitative proteomics established a subunit interaction network and segmented SAGA into four stable modules.³⁶ SAGA harbors a Histone Acetyl Transferase (HAT) activity carried by the Gcn5 subunit and modulated by the Ada2 and Ada3 subunits. Together with Sgf29, these subunits form the HAT module which is a major regulator of histone H3 acetylation in yeast cells.^{37,38} Gcn5 contains a bromodomain which binds acetylated lysines in histone tails,³⁹ while Sgf29 contains a Tudor domain capable of binding H3K4me2/3, another hall mark of actively transcribed chromatin.⁴⁰

In this paper, we describe how CX-MS has been successfully used to determine the binding stoichiometry and topology of the human SAGA HAT sub-complex. We integrated these data with molecular modeling in order to determine the molecular interactions between the four subunits.

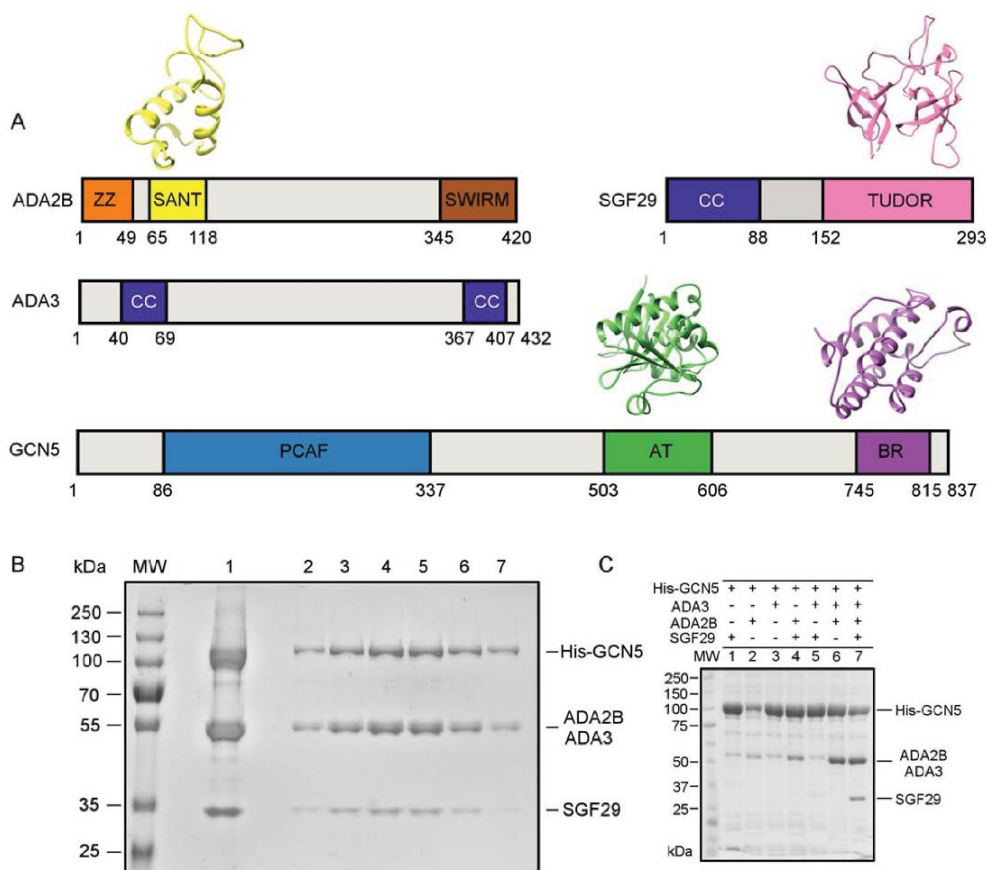


Figure 1. Purification and biochemical characterization of the SAGA HAT subcomplex. (A) Maps of the conserved structural domains within the SAGA HAT subunits. The domains are identified as follows: ZZ—type zinc finger domain; SANT—histone binding domain found in Swi3, ADA2, N-Cor and TFIIB; SWIRM—eukaryotic domain found in proteins implicated in chromatin remodeling and gene expression; CC—coiled-coil domain; TUDOR—structural motif involved in binding methylated arginine or lysine residues; PCAF—P300/CBP-associated factor homology domain; AT—acetyltransferase domain; BR—bromodomain. The known atomic structures of the conserved domains are shown (SANT—PDB 1X41, TUDOR—PDB 3MEA, AT—PDB 1Z4R, BR—PDB 3D7C). (B) SDS-PAGE analysis of the purified SAGA HAT module. MW molecular weight marker, lane 1 purified SAGA HAT after the metal chelating affinity column, lanes 2–7—peak fractions of the size exclusion chromatography. (C) Co-expression of histidine-tagged GCN5 with different combinations of the other SAGA HAT subunits followed by pull down experiments.

Results

Production of the human SAGA HAT module

The sequence analysis of the SAGA HAT subunits revealed conserved homology domains and for some of them an atomic model has been determined by X-ray crystallography [Fig. 1(A)]. To analyze the SAGA HAT subcomplex, we produced recombinant hGCN5, hADA2B, mADA3, and hSGF29 from a single virus in insect cells using the Multibac technology.⁴¹ The mouse ADA3 sequence is 99.3% identical to its human homologue with the exception of three mutations (T110A, V126L, and I226M). The complex was histidine-tagged on GCN5 and purified over a metal chelating affinity column followed by size exclusion chromatography (SEC). SEC of the HAT complex showed a single peak consistent with a homogeneous protein assembly and the SDS-PAGE analysis indicated that all four polypeptides co-elute in the same

fractions [Fig. 1(B)]. Because of their close molecular weight ADA2B and ADA3 co-migrate on a 10% acrylamide gel.

Co-expression experiments

In order to investigate the interactions between the catalytic subunit GCN5 and the other subunits of the SAGA HAT module, the histidine-tagged GCN5 was co-expressed in isolation or with different combinations of the ADA3, ADA2B, or SGF29 subunits. The resulting complex was purified and analyzed as described above. The ADA2B and ADA3 subunits interact poorly in isolation with GCN5 [lanes 2 and 3 in Fig. 1(C)] but form a heterotrimeric complex when expressed together [lane 6 in Fig. 1(C)]. SGF29 incorporates only in a complex formed by GCN5, ADA2B, and ADA3 [lane 7 in Fig. 1(C)] and does not bind stably to GCN5 alone [lane 1 in Fig. 1(C)].

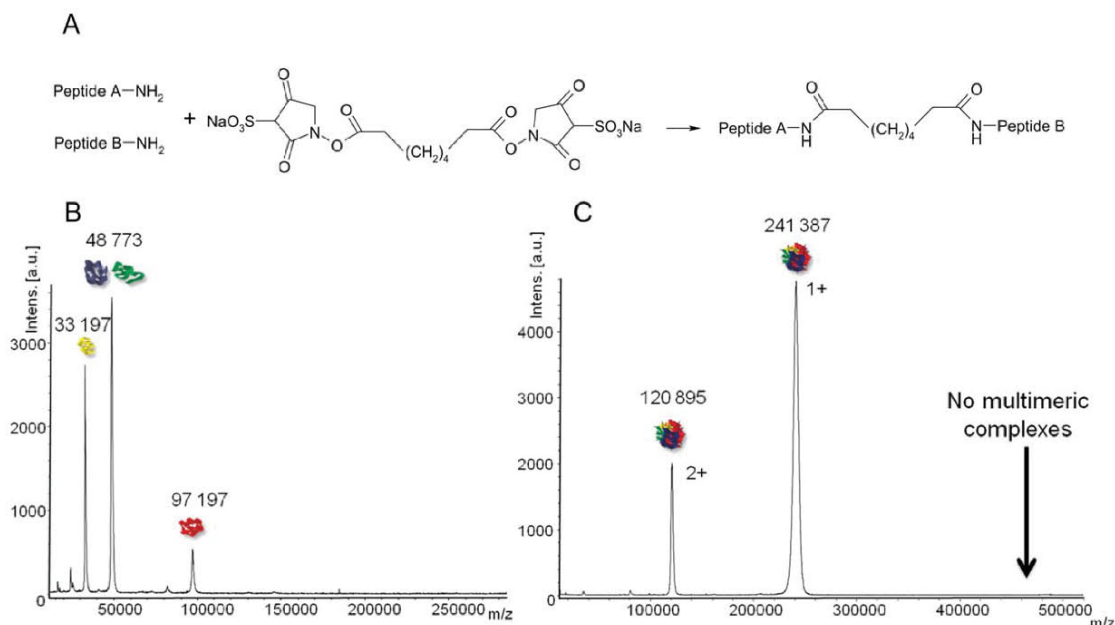


Figure 2. (A) Chemical formula of two peptide chains connected through the BS3 cross-linker. (B,C) MALDI-MS spectra of the complex before (B) and after (C) cross-linking with BS3. A unique stoichiometry at 241,387 Da is detected after cross-linking.

Characterization of the cross-linking reaction by MALDI-MS

The cross-linking reaction was performed with bis[sulfosuccinimidy] suberate (BS3). The reaction conditions (cross-linker to protein ratios, times, and temperatures) were optimized so that cross-linker reacts on the whole sample, but at a quite low cross-linker concentration to avoid perturbation on structure which can lead to aggregation or nonspecific interactions⁴² (see Materials and Methods, Supporting Information Fig. 1). The complex stoichiometry as well as the cross-linking kinetics were followed by MALDI-MS analysis using high mass detection.⁴³ Before cross-linking, ions at 33,197, 48,773, and $97,197 \pm 10$ Da were detected, which corresponds to the theoretical masses of the four subunits (ADA2B and ADA3 are co-detected since their mass difference is too small to be distinguished by MALDI-MS) [Fig. 2(A)]. After cross-linking, molecular mass of about 241,387 Da \pm 200 Da was measured. From this mass, two possible stoichiometries could be formally deduced, i.e. 1/1/1/1 (of each subunit) or 1/2/1 (GCN5/ADA3 or ADA2B/SGF29) [Fig. 2(B)]. A bottom-up approach by LC-MS/MS (see further) will definitively allow to conclude in favor of the first hypothesis. The signal corresponding to the cross-linked entire complex is quite broad, which results in a quite low mass accuracy but sufficient to propose a tetrameric binding stoichiometry without ambiguity. This may be because of the heterogeneity of the cross-linked molecules but also to a certain amount of salt adducts arising from the 100 mM NaCl containing buffer. No signal corresponding to other stoichiometries (homomers or high-order mul-

timers) neither to intermediates reflecting incomplete cross-linking reaction or assembly process was observed. This suggests that there was no residual aggregation and that the detected complex with the expected stoichiometry is formed by specific interactions.

Determination of the proximity map in SAGA HAT by LC-MS/MS

In order to get insight into the “in-proximity areas” (through identification of the spatially close lysines), the cross-linked complex was digested with trypsin and the resulting peptide mixture was then directly analyzed by LC-MS/MS. As described by Müller *et al.*,⁴⁴ a mixture of the light form (d0) and the isotopically labeled form (d4) of BS3 was used for the cross-linking reaction in order to facilitate the detection of low abundant cross-linked products. In this method, peptides containing the cross-linker will appear as a doublet in the mass spectrum with a mass shift of 4.025 amu.

Trypsin activity did not appear to be affected by cross-linking as shown by the high sequence coverages obtained after classical Mascot database search (from 55% for SGF29 up to 91% for ADA3). Whereas peptides that have been modified by a partially hydrolyzed cross-linker (type 0 cross-link or dead-end peptides¹⁸) are easily identified by performing standard database search in Mascot, two open-source algorithms, pLink²⁶ and xQuest,²⁷ were chosen, for both intra- and interlinkages determination. Whereas pLink is dedicated to the identification of peptides modified with either light or heavy cross-linker, xQuest is based on the determination of

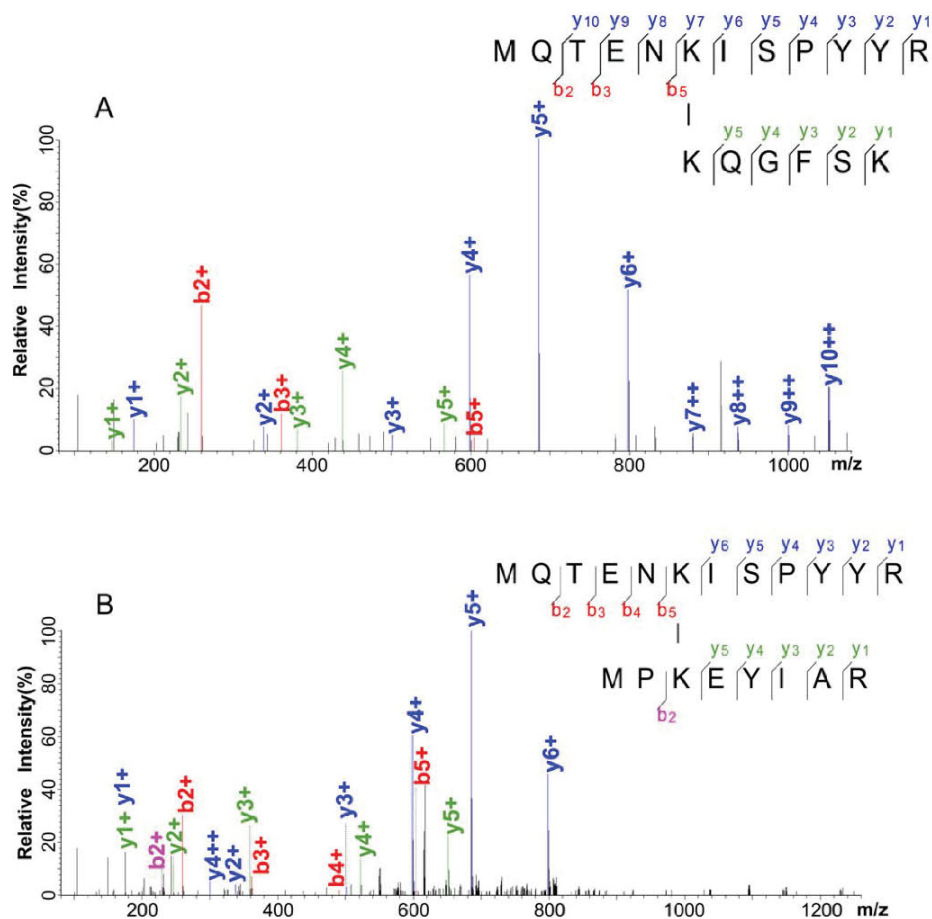


Figure 3. MS/MS spectra of two intercross-linked peptides as identified by pLink (score E-10). The K-K bridge may be (A) or not (B) detected as an ion fragment.

peptides modified with both light and heavy cross-linkers. The request was repeated using a database containing random protein sequences (instead of the sequences of four subunits of SAGA HAT) and false cross-linker mass in order to evaluate false positive matches as well as the associated scorings for both algorithms. From such experiment, background threshold scores were extracted, under which there is a serious risk that a good fit with the MS/MS data occurred by chance. Scores higher than 25 for xQuest and E-06 for pLink were then chosen as thresholds for positive identification. In addition, manual validation was systematically performed to further validate the identification (about 10% were manually removed). Cross-linked peptides follow the general rules of CID-peptide fragmentation⁴⁵ and thus, their fragmentation spectra should provide fragment ions arising from both bridged peptides upstream and downstream of the linkage. Two examples of MS/MS spectrum assignment by pLink are shown in Figure 3. In both cases, the corresponding peptides were unambiguously identified as reflected by the large series of consecutive y-ions and b-ions. In some cases, ions containing the link-

age product will also be detected increasing the identification relevance as shown in Figure 3(A) at $m/z = 879.4828$.

A total number of 97 unique inter-cross-links (indicating the linkage of unique lysine-lysine pairs) have been found with pLink and xQuest (Table I). Among them, 23 were identified by xQuest only and 43 by pLink only. These results are the combination of three technical replicates (displaying 80% overlap) and one biological replicate (40% overlap). It should be noticed that the proximity areas were strictly equivalent in all replicates. The overlap discrepancy might come from: (i) a lower MS/MS intensity spectrum that could not be unambiguously assigned, (ii) a flexibility of the complex structure that slightly modify the distance between reactive lysines leading to different links, or (iii) from a competitive reactivity of two neighboring lysines, yielding cross-linked peptides with different masses.

A linkage map has been produced from these results to visualize intrasubunit links and inter-subunit links (Fig. 4). This map clearly highlights several discrete inter-linked areas with high

Table I. List of intermolecular cross-links between SAGA HAT subunits. No score means the cross-linked peptide was not identified or not validated manually.

GCN5-ADA3					
Linked K in GCN5	Linked K in ADA3	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	85	5.26E-14	55	2614.3754	-0.4
314	147	3.65E-12		2905.4569	1.1
314	221		45.71	2654.3527	-0.9
314	354	1.56E-13	44.66	3218.6297	-0.7
361	85	5.47E-10		1803.0770	0.8
366	78	4.89E-07		1229.6615	-0.7
366	85	7.13E-07		1491.8961	0.7
366	421	5.47E-09		1400.7874	-0.1
536	85	1.74E-06		2835.6122	0.2
561	78	2.27E-07		1622.8750	-0.1
561	85	1.15E-08	50.64	1877.0595	-0.1
561	109	1.15E-09		2560.3219	1.3
561	147	1.37E-13		2172.1661	-0.6
580	85	1.28E-11	45.79	1856.1238	-0.5
649	85	8.37E-10	51.03	1563.9136	-0.7
649	124	2.08E-10	45.04	1630.9042	-0.3
649	304		36.54	2227.1581	-7.6
657	405		27.06	1937.1456	-7.9
660	85	1.29E-06		1455.8925	-0.8
707	85	1.03E-08		1617.0026	-0.3
713	364		35.84	2059.1279	-0.5
748	418		32.19	1957.0869	-0.8
750	418		44.43	1957.0843	-2.1
784	78	9.97E-07		2081.0428	2.5
816	85		35.8	1959.1605	-3.3

GCN5-ADA2B					
Linked K in GCN5	Linked K in ADA2B	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	328		44.78	2626.3341	-0.4
314	331	5.50E-07	57.22	2514.2217	-0.8
314	378	7.20E-09	42.5	3089.6396	-3.3
314	390		40.44	2753.3972	-0.5
361	390	4.30E-06		1942.0999	-2.7
536	214	2.43E-06		2651.5058	-0.7
572	214	2.85E-08		1664.9613	-1.1
572	220	7.75E-13		1720.9987	-0.8
580	250		38.81	2154.2570	-7.7
580	250		35.23	1870.0746	-1.6
702	250		29.07	2645.4865	-3.5

GCN5-SGF29					
Linked K in GCN5	Linked K in SFG29	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	40	4.31E-11		3585.7811	-1.3
314	50	2.09E-08	38.84	3410.6459	0.2
561	40	2.71E-16		2852.4903	1.1
561	50	5.72E-12	44.79	2673.3288	-0.2
580	40	5.47E-09		2827.5303	1.9
580	50	5.78E-10	51.09	2652.3947	0.1
649	40	1.06E-16	34.72	2539.3443	-0.2
649	50	4.39E-10	43.37	2360.1818	-0.7
649	79	1.49E-09		1404.7976	-0.5
649	112	6.92E-08		1365.7438	1.2
816	40	4.67E-07		2934.5975	-6.7

Table I. Continued

GCN5-SGF29					
Linked K in GCN5	Linked K in SFG29	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
816	50	1.39E-14		2755.4366	0.5
855	50	1.85E-06		2638.3675	-0.3
ADA3-ADA2B					
Linked K in ADA3	Linked K in ADA2B	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
147	235	7.92E-13	45.89	2644.4160	0.4
147	390		42.1	2033.1002	-0.8
194	331	3.05E-09	41.76	2069.1205	-0.1
194	378	7.49E-06		2640.5114	0.0
194	390		35.29	2304.2634	-3.0
210	319	3.62E-06		2446.2276	0.6
210	319	3.33E-08	30.06	3172.6412	-6.4
210	322	1.63E-10	39.13	2884.4199	-0.1
210	322	2.88E-12	40.76	3016.5401	-0.1
210	328	4.43E-06	56.43	2673.3658	0.2
210	331	7.99E-08	50.25	2557.2294	-0.1
210	378	8.13E-06		3132.6454	-0.6
210	390		33.39	2796.4038	-0.2
221	319	4.79E-07		1431.8233	1.6
221	322	9.36E-09	297.22	1874.0409	1.1
221	322	6.31E-15	45.29	1998.1108	-0.4
221	328	3.39E-14	42.52	1654.9365	0.5
221	331	1.54E-09	44.44	1542.8210	-2.8
221	373	1.63E-08		2087.2605	0.8
221	378		30.64	2118.2407	-0.3
221	390		41.6	1781.9974	-1.4
223	328	3.72E-07		2270.3145	0.1
223	378	2.86E-06		2729.5941	0.0
242	328	1.06E-06		1783.0202	-0.8
242	378	2.96E-11	45.58	2246.3254	0.2
242	390	9.38E-06	44.71	1910.0836	0.5
304	378	4.59E-06		2741.5227	-2.4
304	390		42.97	2405.2822	0.3
312	390		26.06	2977.5704	-1.0
354	390		40.72	2346.2754	-0.6
364	282	3.79E-07		1597.9274	0.2
372	390	2.14E-07	41.31	1940.0690	0.1
397	403		29.07	2645.4865	1.0
ADA3-SGF29					
Linked K in ADA3	Linked K in SGF29	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
77	50	6.23E-14	33.99	3586.8165	0.2
78	50	4.17E-07		2141.0575	1.3
85	40	2.34E-08		2574.4029	-0.1
85	50	1.70E-09	47.29	2399.2672	0.0
85	50		39.28	2555.3450	-9.1
97	50		34.14	3176.6059	-0.6
109	40	4.30E-08		3257.6652	-0.3
147	40	1.57E-11		2869.5094	0.4
147	50	4.08E-10	45.14	2690.3476	-0.4
372	40	1.38E-11	30.73	2772.4517	-1.3
372	50	1.06E-06		2597.3158	-0.6
397	50	4.09E-07		2355.2078	-0.1
421	40	6.71E-06		2487.3193	0.3
421	50	7.04E-07		2312.1834	0.4
421	99	2.34E-07		2501.3789	0.6

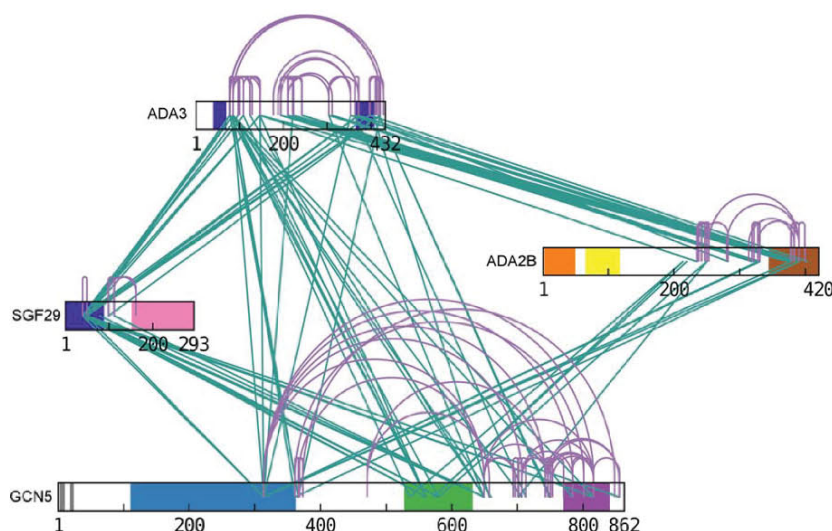


Figure 4. Inter- (green) and intra- (purple) molecular cross-linking map within the SAGA HAT subunits obtained with the BS3 d0-d4 cross-linking agent. The represented cross-links are deduced from both xQuest and pLink identification. Colors are consistent with Figure 1(A).

significance as shown by many contacts that have been detected for each of them. In particular, extensive cross-linking was observed between ADA3 and ADA2B suggesting a strong network of connections between these two subunits while no cross-links were detected between SGF29 and ADA2B. Remarkably, among the 22 lysines equally distributed on the SGF29 sequence, two lysines (K40 and K50) appeared to be mostly involved in inter-cross-linked products, suggesting that the bridging reaction occurred in a very specific way. On the other hand, no cross-links could be detected for the N-terminal half of GCN5 containing most of the PCAF domain, the Tudor domain of SGF29, and the ZZ-SANT domains of ADA2B.

Molecular modeling

Atomic structures of several domains composing the HAT subunits are available [SANT, Tudor, acetyltransferase, and bromo domains, Fig. 1(A)] or can be predicted by homology modeling of closely related sequences (PCAF, ZZ, SWIRM). Intrasubunit cross-links were used as additional spatial restraints helping to predict tertiary models of the SAGA HAT subunits since the relative positions of their domains are not known. For each subunit, domains with known or modelled structure were positioned in a way that minimizes the distances between the cross-linked lysines. That was achieved with a rigid body docking approach where the domain structure was not deformed.

For GCN5 the intricate interaction network has allowed to position the PCAF, acetyltransferase and bromo domains satisfying the maximum theoretical distances between cross-linked lysines [Fig. 5(A)]. The only exception is the connection between lysines

K361 and K801 [Fig. 5(B)]. K361 residue is situated within a disordered region between PCAF and acetyltransferase domains, that region is flexible and thus the lysine residue may occupy a different position. The whole GCN5 subunit is heavily cross-linked with the exception of N-terminus including mostly PCAF domain. According to the secondary structure prediction this part of the protein is mostly helical. We note that about 40% of GCN5 sequence, comprising the gaps between three known domains, is unstructured and we could not predict tertiary structure of these areas. In case of ADA2B protein most of the cross-links were detected in C-terminal part. ZZ and SANT domains that are known to stimulate nucleosome binding are not involved in any interactions [Fig. 5(C,D)]. The middle part of ADA2B does not have a specific structure and represent around 45% of total sequence. The complete structure of ADA3 was predicted by comparative modeling and docking of four domains (see Materials and Methods section) since no homologs with known structures were found. Secondary structure prediction showed that this protein is mostly helical with two short coiled-coil domains at the termini [Fig. 5(E,F)]. In the case of SGF29 we have modelled the structure of the coiled coil domain (3–88), but unfortunately this model is of low quality since the protein does not display enough intralinks that would allow more accurate structure modeling.

In order to obtain an initial low-resolution model of SAGA HAT subcomplex, all four subunits were positioned one with respect to the others by minimizing the distances between inter-subunit cross-linked lysines and by avoiding any steric clashes between the atomic models [Fig. 6(A,B)]. This model of the SAGA HAT module is still missing

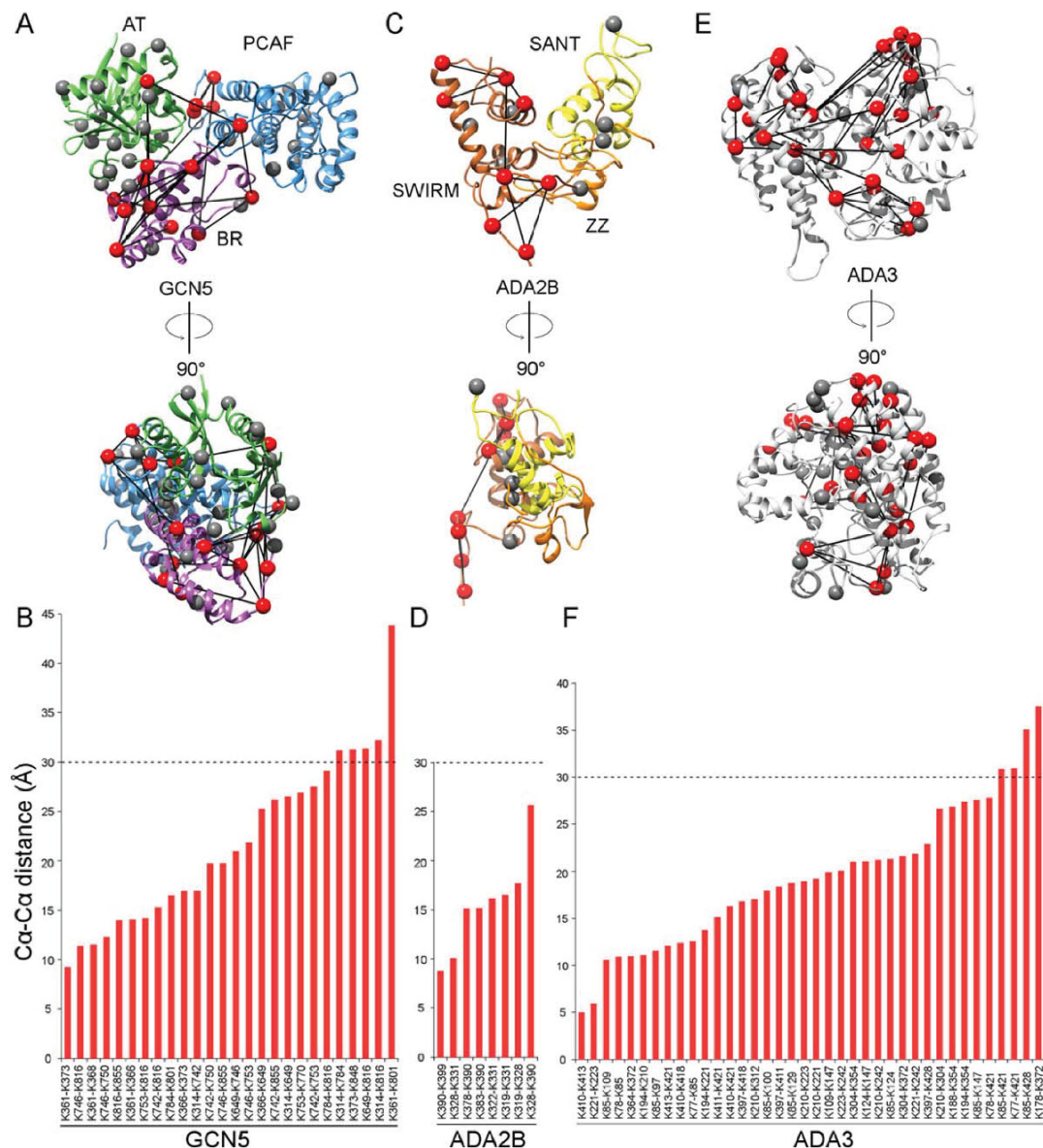


Figure 5. Tertiary subunit architecture modelled by intrasubunit cross-links. (A, C, E) Modeling of the conserved domains of GCN5, ADA2B, and ADA3, respectively. Red spheres correspond to lysine residues involved in intrasubunit cross-links, grey spheres show the nonmodified lysines and black lines represent the distance between the cross-linked residues. (B, D, F) Calculated distances between all cross-linked residues of GCN5, ADA2B, and ADA3, respectively. The 30Å threshold is indicated by a dotted line.

several unstructured parts of subunits and might be further improved by molecular dynamics methods since here only rigid-body docking was used.

Discussion

In this project, chemical cross-linking-mass spectrometry was used in combination with modeling in order to build a low-resolution structural model of the SAGA HAT subcomplex. Even with advances in mass spectrometry technology or with cross-linking reagents improvements, data interpretation is still challenging and the problem of extracting effective cross-linked peptide from random matching remains.

Strategy and validation of the identification of cross-linked peptides

MALDI-TOF allowed a straightforward comparison between control and cross-linked samples and thus a direct determination of the binding stoichiometry. While ESI-MS is shown to be an appropriate method to characterize noncovalent complexes in terms of existence and stoichiometry,⁴⁶ it could not be used in this particular case, since the SAGA HAT subcomplex could not be transferred properly to a volatile ammonium acetate buffer. Chemical cross-linking appeared as a nice alternative by covalently stabilizing the complex before its characterization by mass

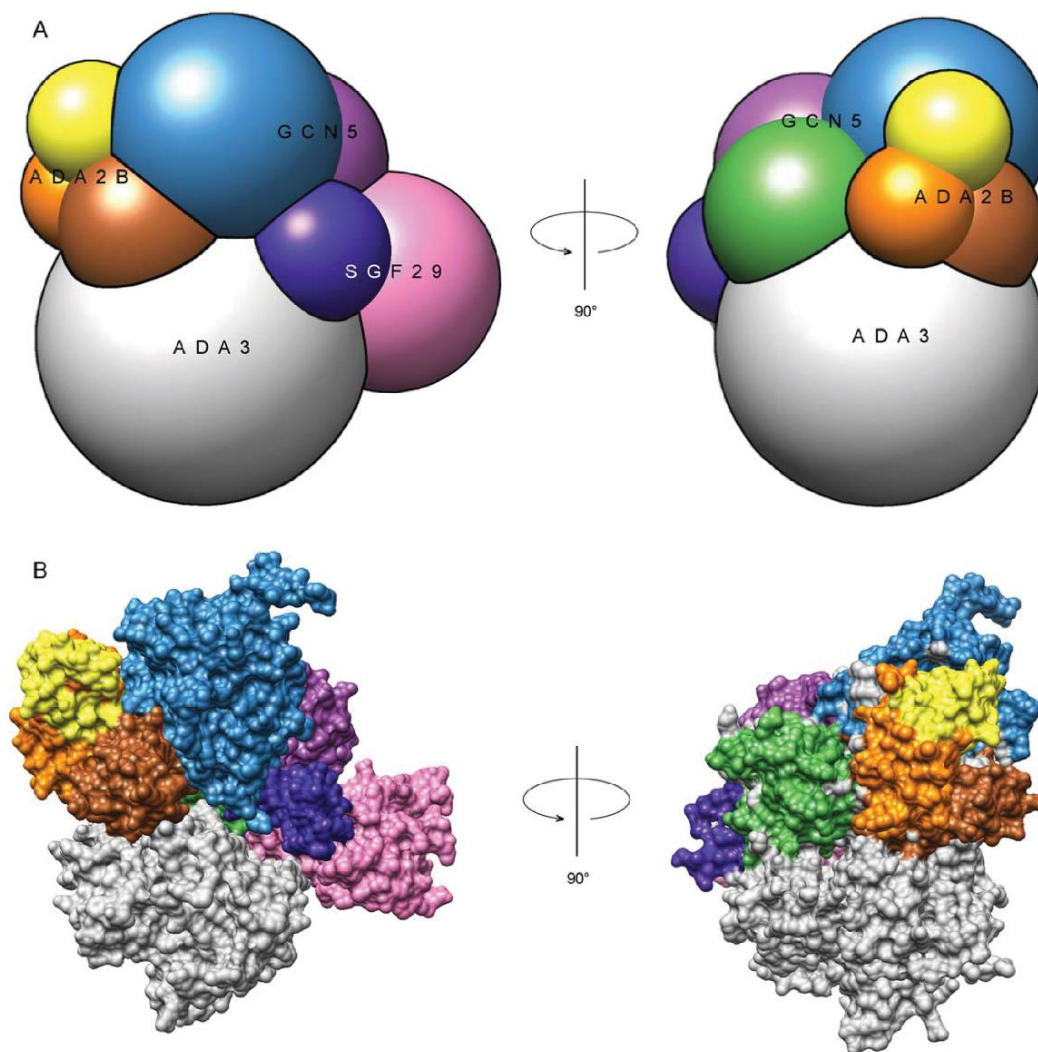


Figure 6. Quaternary organization of the SAGA HAT module. (A) Schematic representation of the 3D organization of the conserved SAGA HAT subunit domains obtained upon distance minimization of intersubunit cross-linked lysine residues. (B) Proposed 3-D model of the SAGA HAT module in which the atomic models of the conserved domains are positioned and oriented according to the cross-linking network and avoiding structural overlap. The model represents 63% of the SAGA HAT residues.

spectrometry so that it could withstand denaturing conditions and be analyzed by methods that usually lead to complex dissociation. A bottom-up approach was then used to look for the cross-linked peptides and establish a linkage map. As already described in the literature,^{47,48} mass spectrometer yielding high-resolution measurement of peptide and fragment masses but also able to precisely determine the charged states, appeared to be essential for the success of the database search. The large number of bridged products that are arising from all possible pair-wise combination of peptides containing a lysine should be considered, which generates a very large theoretical database and increases the risk of random matches if mass accuracy is not sufficient. In our case, the parameters of mass spectrometer (mass range of 300–1200 m/z) allowed us to analyze a large majority of digested peptides present in solu-

tion without compromising the high mass accuracy and high resolution. All matched peptides displayed a mass accuracy within 6 ppm from which 97% were lower than 2 ppm.

Interestingly, the overall connectivity pattern of the pLink- and xQuest-extracted maps is very convergent (Supporting Information Fig. 4). Each of the putative linked areas has been observed with both algorithms and combining the results of both searches yielded a more expressive network map. Some limitations of one algorithm might be compensated by the other (pLink requires generous fragmentation spectra of the cross-linked peptide, whereas xQuest needs quality fragmentation spectra for both d0 and d4 modified peptide).

Several observations strengthen the validation of our putative linkage map. (i) Each linked area is defined by multiple peptide pairs (for example, 6

and 7 links respectively are detected to propose a connection between the N- and C-termini of ADA3 with the N-terminal domain of SGF29). (ii) Most of the time, there is a correlation between inter and intrasubunit linkages. For example, the lysine K40 of SGF29 is in the vicinity of the lysines K85 and K421 of ADA3 whereas intrasubunit link between K85 and K421 of ADA3 strongly suggests that these two lysines are brought close to each other by the tertiary fold of the subunit. (iii) The analysis of deletion mutants have shown an interaction between yAda2b residues 167–434 (including the SWIRM domain) with C-terminus of Ada3⁴⁹ which is fully consistent with our observations. (iv) Yeast two hybrid approaches demonstrated an interaction between the N-termini of SGF29 and of ADA3^{50–53} which is also revealed by our experiments.

The absence of observed cross-linked peptides might also be worth to exploit but it is not straightforward to unambiguously correlate no cross-linking with distant lysines (i.e. higher than the known 30 Å maximum distance of C α –C α atoms of peptides that can be cross-linked by BS3^{54,55}). Indeed, various situations could prevent reaction with the cross-linker: (i) there is no lysine in a particular domain (only four lysines in the ADA2B N-terminus domain); (ii) the lysines may not be accessible to the cross-linker (buried in the structure) or not reactive because of chemical environment. Looking at dead-end peptides could likely help to address this purpose (see Supporting Information Table 3). In our case, 53% of the lysines in the GCN5 PCAF domain have been shown to possibly react with BS3, meaning that they are accessible to the solvent but without forming cross-links. On the other hand, only two lysines among 12 in SGF29 Tudor domain have been detected as dead-end modified lysine, strongly suggesting that these lysines are not available for cross-linking reaction. Interestingly these noncross-linked domains are often involved in interactions with modified histone tails and are likely to be exposed to bind nucleosomes thus being less prone to interactions with subunits of the HAT module.^{37,56} In such situation, the use of cross-linkers containing other types of reactive functions (photo-reactive cross-linkers, sulfhydryl cross-linkers etc.) is also conceivable. In these cases, it may be important to adapt the cross-links identification software to be used. (iii) The mass/charge of cross-linked peptides may be higher than 1200 *m/z* to be selected by the MS for fragmentation, or too short (less than three amino acids) to give sufficient information for being recognized during database search; (iv) the fragmentation mass spectrum is of poor quality to be unambiguously interpreted.

Biological relevance of our linkage map

We compared our results with the recently published linkage map of the full yeast SAGA.⁵⁷ The yeast and

human subunits show a large degree of sequence conservation except for hGCN5 which contains an additional PCAF domain at its N-terminus and the yAda3 subunit which is twice as large as hADA3. Despite a different lysine distribution in the sequences from two organisms, the network of domain interactions is highly conserved with the notable exception of ySgf29 which shows a large number of intrasubunit links and only interacts with yAda3 while hSGF29 has a strong interaction network with the most domains of hGCN5 (Fig. 4). The interaction between yeast Sgf29 and Gcn5 was previously demonstrated by systematic deletion analysis.⁵⁸ The same regions of hGCN5 make extensive cross-links with hADA3 along its whole sequence while in yeast the link between these two subunits is limited to a single domain in the C-terminal part of yAda3. Rather than reflecting the differences in primary sequence between the two organisms, we suggest that these differences in the cross-linking network reflect a conformational change between the free HAT module and when it is incorporated into the complete SAGA complex. Interestingly, the domains that show less cross-links within the HAT module in the yeast complex are strongly involved in interactions with other SAGA subunits. As shown by Han *et al.*,⁵⁷ in yeast Ada3 has few intra- and interlinks with subunits of HAT module, at the same time revealing numerous cross-links with other proteins like Taf5, Taf6, Taf12, Spt7, Sgf73, and Ada1. We show that in human ADA3 is extensively cross-linked to the other HAT subunits: ADA2B, SGF29, and GCN5. Altogether these results suggest that the HAT module is reorganized upon integration in the SAGA complex.

The histidine pull down assays [Fig. 1(C)] do not completely recapitulate the cross-linking experiments and it appears that proximity does not imply interaction. Cross-linking results indicate that the N-terminus of SGF29 is close to GCN5 but this proximity is not enough to form a stable pair-wise interaction. SGF29 requires the presence of both ADA3 and ADA2B to establish a network of interactions leading to a HAT subcomplex. Likewise the ADA2B and ADA3 subunits incorporate significantly into the subcomplex only when they are co-expressed together with GCN5. The last observation is partially supported by the fact that yAda2 plays a critical role in the formation of the HAT module and its association with the overall SAGA complex.³⁶

The interaction between CX-MS and molecular modeling is very promising. The data generated by CX-MS will provide associations and restraints information that will help to eliminate inconsistent theoretical models. In combination with complementary techniques, CX-MS is moving to integrate a general pipeline that aim to iteratively improve the accuracy of the proposed structural model with a

final objective of determining the closest one to the real state of the complex at a particular moment of its life. Large, heterogeneous or dynamic complexes would benefit from such an alternative approach since traditional structural biology methods are often difficult to carry through.

Materials and Methods

SAGA HAT expression and purification

The constructs for the expression of the HAT module of SAGA (hGCN5, hADA2B, mADA3, hSGF29) were generated using baculovirus Multibac system.⁴¹ Briefly, hGCN5 with TEV-cleavable 6xHis-tag on N-terminus was cloned into pFL acceptor plasmid, the other three genes encoding hADA2B, mADA3, and hSGF29 were cloned into donor plasmids (pIDS, pIDC and pIDK respectively), which were subsequently combined via Cre-Lox recombination. The total construct was integrated via Tn7 transposition into a bacmid with YFP at the backbone as an expression level marker. The bacmid was isolated and used for infection of 1 L Sf21 insect cells (160 mL of virus for 1 L of Sf21 insect cells in Sf900II serum-free media, MOI 15–20). YFP fluorescence was used to monitor overall expression levels after the day of infection. Cells were harvested in ice cold PBS with 10% glycerol when YFP levels reached a plateau (72 h). All further purification steps were done at 4°C.

Cell pellet from 1 L of culture was dissolved in 40 mL of lysis buffer (20 mM HEPES-KOH pH 8.0, 500 mM NaCl, 10% glycerol, Protease Inhibitor Cocktail (Roche)) and mildly sonicated on ice for 3 min. Cell extract was further cleared by ultracentrifugation for 1 h at 50,000 g, HAT subcomplexes were purified by His-tag using HiTrap Chelating HP 5 mL column (GE Healthcare), previously loaded with nickel. After washing steps with increasing concentration of imidazole, HAT subcomplexes were eluted at 250 mM imidazole as confirmed by 10% SDS-PAGE. Peak elutions were pooled and further purified on Superdex S200 16/60 GL (GE Healthcare) gel filtration column. Complexes were eluted in a similar buffer (20 mM HEPES-KOH pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM DTT, 0.1 mM EDTA) and analyzed by 10% SDS-PAGE. Peak fractions were pooled, concentrated on Amicon 30 kDa column, then flash frozen in liquid nitrogen and stored at –80°C.

During purification steps, protein concentrations were controlled by Bradford assay,⁵⁹ their quality by SDS-PAGE⁶⁰ and the identity of the subunits by Western Blotting.⁶¹

Pull-down assays

Constructs for expression of different subunit combinations were generated using Multibac system and

expressed in insect cells as described before. Small scale purifications using histidine-tag on GCN5 were performed in batch and eluates were analyzed on 10% SDS-PAGE.

Cross-linking and MALDI-MS data acquisition

Before cross-linking reaction, the complex was subjected to buffer exchange to 20 mM HEPES, 100 mM NaCl in 4°C to eliminate glycerol (Micro Bio-Spin chromatography column, exclusion limit 6 kDa). An equimolar mixture of the BS3 cross-linking reagent (Thermo Scientific) in the normal form (d0) and isotopically labeled form (d4) was prepared by dissolving in water. Cross-linking reactions were carried out on 30 µL of a 2 mg/mL protein complex at various protein to cross-linker molar ratios namely 1:50, 1:75, 1:200 at two incubation times (30 min, 2 h). From these assays, the optimal conditions for cross-linking reaction were fixed at a ratio protein:cross-linker 1:75 (equivalent to a lysine:cross-linker molar ratio of 1:0.5 since SAGA HAT subcomplex contains 164 lysine residues) for 30 min at room temperature. The reaction was quenched with Tris 25 mM final concentration at pH 7.5. The intact and cross-linked complexes were analyzed by MALDI-MS (Autoflex II TOF/TOF, Bruker) equipped with a HM1 high mass detector (CovalX, Switzerland). Sinapinic acid (SA) matrix at 10 mg/mL in H₂O/ACN/TFA (50/50/0.3) and a dried drop spotting method were chosen. The analyses were performed in a linear acquisition mode and the MALDI calibration was achieved with a solution of BSA at 5 µM using the monocharged ions of the monomer and dimer of BSA (Sigma-Aldrich).

LC-MS/MS data acquisition

After cross-linking, the protein complex was digested with trypsin (Sequencing Grade Modified, Promega) at a 1:5 enzyme to substrate molar ratio and incubated for 3 h at 37°C. The reaction was quenched with addition of TFA at final concentration of 0.1%. MALDI analysis was performed as a control after digestion, in order to ensure that there was no more signal corresponding to the entire complex. LC-MS/MS data were acquired on an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific) equipped with a nano electrospray ionization source in positive ion mode coupled with a Rapid Separation LC system (nanoRSLC U3000, Thermo Scientific), using HCD fragmentation mode. The isolation window of precursor ion was fixed at 2 Da, and the HCD collision energy at 30% for the fragmentation. A full-scan survey MS spectrum (300–1200 *m/z*) was carried out with resolution of 240,000 at *m/z* 400, then the fifteen most abundant ions detected were submitted to a MS/MS experiment in Orbitrap at resolution 15,000.

MS data analysis

Two search engines, dedicated to cross-linked peptides xQuest (v2.1.1, <http://proteomics.ethz.ch/cgi-bin/xquest2.cgi/index.cgi>) and pLink (v1.15, <http://pfind.ict.ac.cn/software/pLink>), were used. In the xQuest searches, the precursor mass tolerance was set at 10 ppm as well as the fragment mass tolerance for common- and cross-linked ions. In pLink, the tolerance for MS1 matching was fixed at 10 ppm and the filter for peptide tolerance at 10 ppm. Each data set was produced using a FPR of 5%. All spectra of putative cross-linked peptides were manually controlled before positive identification. The bridged peptides were confidently identified when the majority of the observed fragment ions were assigned and when more than three consecutive fragment ions were matching for both linked peptides. The interaction maps between subunits were drawn via xiNET-Crosslink Viewer (<http://crosslinkviewer.org>), developed in Rappsilber laboratory.

SAGA HAT sequence analysis and structural modeling

Multisequence alignment of SAGA HAT subunits from different species allowed to identify the conserved structural domains. Homology modeling for the domains without known structure was used to predict the atomic models, otherwise the available crystal structures were found in PDB database. Intramolecular cross-linking data provided the distance restraints for further improving the homology models of each domain. Complete models of GCN5, ADA2B, and ADA3 were generated using rigid-body fitting with imposed cross-linking restraints (maximum distance between cross-linked lysines was set to 30 Å). The quality of the final models was verified using several bioinformatics algorithms, all models received good to fair scores (Supporting Information Table 4). The best predicted atomic models of all subunits were positioned and oriented according to the intersubunit cross-linking network and avoiding structural overlap, thus resulting into the 3-D model of the SAGA HAT module.

Disclosure: The authors declare no competing financial interest and no conflict of interest.

Acknowledgments

N-T. N-H. is supported by a PhD Scholarship from the French MESR (Ministère de l'Enseignement Supérieur et de la Recherche). The authors would like to thank the Proteomic Platform of IGBMC where the purchase of the LTQ Orbitrap Elite mass spectrometer was supported by the ARC Foundation, and the Proteomic Platform of IBMC (Institut de Biologie Moléculaire et Cellulaire) in Strasbourg, France, for their contribution to this project.

References

- Purdy MD, Bennett BC, McIntire WE, Khan AK, Kasson PM, Yeager M (2014) Function and dynamics of macromolecular complexes explored by integrative structural and computational biology. *Curr Opin Struct Biol* 27:138–148.
- Göbl C, Madl T, Simon B, Sattler M (2014) NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc* 80:26–63.
- Rodrigues JPGLM, Bonvin AMJJ (2014) Integrative computational modeling of protein interactions. *FEBS J* 281:1988–2003.
- Heck AJR (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* 5:927–933.
- Lanucara F, Holman SW, Gray CJ, Evers CE (2014) The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* 6:281–294.
- Konermann L, Tong X, Pan Y (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J Mass Spectrom* 43:1021–1036.
- Merkley ED, Cort JR, Adkins JN (2013) Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J Struct Funct Genomics* 14:77–90.
- Stengel F, Aebersold R, Robinson CV (2012) Joining Forces: Integrating Proteomics and Cross-linking with the Mass Spectrometry of Intact Complexes. *Mol Cell Proteomics* 11:R111.014027.
- Singh P, Panchaud A, Goodlett DR (2010) Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Anal Chem* 82:2636–2642.
- Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R (2010) Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Mol Cell Proteomics* 9:1634–1649.
- Rappsilber J (2011) The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* 173:530–540.
- Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalishman Y, Liu X, Gibbons BJ, Levitt M, Kornberg RD (2013) Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342:1238724.
- Müller MQ, Sinz A (2012) Chemical cross-linking and high-resolution mass spectrometry to study protein–drug interactions. *Methods Mol Biol* 803:205–218.
- Kalishman N, Adams CM, Levitt M (2012) Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc Natl Acad Sci* 109:2884–2889.
- Lössl P, Kölbl K, Tänzler D, Nannemann D, Ihling CH, Keller MV, Schneider M, Zaucke F, Meiler J, Sinz A (2014) Analysis of nidogen-1/laminin γ 1 interaction by cross-linking, mass spectrometry, and computational modeling reveals multiple binding modes. *PLoS ONE* 9:e112886.
- Thierbach K, von Appen A, Thoms M, Beck M, Flemming D, Hurt E (2013) Protein Interfaces of the Conserved Nup84 Complex from *Chaetomium thermophilum* Shown by Crosslinking Mass Spectrometry and Electron Microscopy. *Structure* 21:1672–1682.

17. Kim SJ, Fernandez-Martinez J, Sampathkumar P, Martel A, Matsui T, Tsuruta H, Weiss TM, Shi Y, Markina-Inarrairaegui A, Bonanno JB, Sauder JM, Burley SK, Chait BT, Almo SC, Rout MP, Sali A. (2014) Integrative Structure–Function Mapping of the Nucleoporin Nup133 Suggests a Conserved Mechanism for Membrane Anchoring of the Nuclear Pore Complex. *Mol. Cell. Proteomics* 13:2911–2926.
18. Paramelle D, Miralles G, Subra G, Martinez J (2013) Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* 13:438–456.
19. Sutherland BW, Toews J, Kast J (2008) Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. *J Mass Spectrom* 43:699–715.
20. Rhode BM, Hartmuth K, Urlaub H, Lührmann R (2003) Analysis of site-specific protein-RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *RNA* 9:1542–1551.
21. Fritzsche R, Ihling CH, Götze M, Sinz A (2012) Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun Mass Spectrom* 26:653–658.
22. Lascoux D, Paramelle D, Subra G, Heymann M, Geourjon C, Martinez J, Forest E (2007) Discrimination and selective enhancement of signals in the MALDI mass spectrum of a protein by combining a matrix-based label for lysine residues with a neutral matrix. *Angew Chem Int Ed* 46:5594–5597.
23. Paramelle D, Cantel S, Enjalbal C, Amblard M, Forest E, Heymann M, Geourjon C, Martinez J, Subra G (2009) A new generation of cross-linkers for selective detection by MALDI MS. *Proteomics* 9:5384–5388.
24. Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kühn U, Sinz A (2011) StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J Am Soc Mass Spectrom* 23:76–87.
25. Rasmussen MI, Refsgaard JC, Peng L, Houen G, Højrup P (2011) CrossWork: Software-assisted identification of cross-linked peptides. *J Proteomics* 74:1871–1883.
26. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ (2012) Identification of cross-linked peptides from complex samples. *Nat Methods* 9:904–906.
27. Rinner O, Seebacher J, Walzthoeni T, Mueller L, Beck M, Schmidt A, Mueller M, Aebersold R (2008) Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5:315–318.
28. Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* 11:1640–1650.
29. Grünberg S, Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* 38:603–611.
30. Hahn S, Young ET (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* 189:705–736.
31. Nagy Z, Tora L (2007) Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* 26:5341–5357.
32. Spedale G, Timmers HTM, Pijnappel WWMP (2012) ATAC-king the complexity of SAGA during evolution. *Genes Dev* 26:527–541.
33. Eisenmann DM, Chapon C, Roberts SM, Dollard C, Winston F (1994) The *Saccharomyces cerevisiae* SPT8 gene encodes a very acidic protein that is functionally related to SPT3 and TATA-binding protein. *Genetics* 137:647–657.
34. Brand M, Laurent C, Mallouh V, Tora L, Schultz P (1999) Three-dimensional structures of the TAFII-containing complexes TFIID and TFIIIC. *Science* 286:2151–2153.
35. Wu P-YJ, Ruhlmann C, Winston F, Schultz P (2004) Molecular Architecture of the *S. cerevisiae* SAGA Complex. *Mol Cell* 15:199–208.
36. Lee KK, Sardu ME, Swanson SK, Gilmore JM, Torok M, Grant PA, Florens L, Workman JL, Washburn MP (2011) Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Mol Syst Biol* 7:503.
37. Bian C, Xu C, Ruan J, Lee KK, Burke TL, Tempel W, Barsyte D, Li J, Wu M, Zhou BO, Fleharty BE, Paulson A, Allali-Hassani A, Zhou JQ, Mer G, Grant PA, Workman JL, Zang J, Min J (2011) Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* 30:2829–2842.
38. Durant M, Pugh BF (2006) Genome-wide relationships between TAF1 and histone acetyltransferases in *Saccharomyces cerevisiae*. *Mol Cell Biol* 26:2791–2802.
39. Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ, Workman JL (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* 111:369–379.
40. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, Mann M (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142:967–980.
41. Trowitzsch S, Bieniossek C, Nie Y, Garzoni F, Berger I (2010) New baculovirus expression tools for recombinant protein complex production. *J Struct Biol* 172:45–54.
42. Lee YJ, Lackner LL, Nunnari JM, Phinney BS (2007) Shotgun Cross-Linking Analysis for Studying Quaternary and Tertiary Protein Structures. *J Proteome Res* 6:3908–3917.
43. Wenzel R, Röhling U, Nazabal A, Hillenkamp F (2009) A detector device for high mass ion detection, a method for analyzing ions of high mass and a device for selection between ion detectors. Available from: Patent WO/2009/086642.
44. Müller DR, Schindler P, Towbin H, Wirth U, Voshol H, Hoving S, Steinmetz MO (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem* 73:1927–1934.
45. Gaucher SP, Hadi MZ, Young MM (2006) Influence of crosslinker identity and position on gas-phase dissociation of Lys–Lys crosslinked peptides. *J Am Soc Mass Spectrom* 17:395–405.
46. Hilton GR, Benesch JLP (2012) Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J R Soc Interface* 9:801–816.
47. Mann M, Hendrickson RC, Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 70:437–473.

48. Schubert P, Hoffman MD, Sniatynski MJ, Kast J (2006) Advances in the analysis of dynamic protein complexes by proteomics and data processing. *Anal Bioanal Chem* 386:482–493.
49. Candau R, Berger SL (1996) Structural and functional analysis of yeast putative adaptors. Evidence for an adaptor complex in vivo. *J Biol Chem* 271:5237–5245.
50. Kurabe N, Katagiri K, Komiya Y, Ito R, Sugiyama A, Kawasaki Y, Tashiro F (2007) Deregulated expression of a novel component of TFTC/STAGA histone acetyltransferase complexes, rat SGF29, in hepatocellular carcinoma: Possible implication for the oncogenic potential of c-Myc. *Oncogene* 26:5626–5634.
51. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569–4574.
52. Rolland T, Taşan M, Charlotteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruysinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabási AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M (2014) A proteome-scale map of the human interactome network. *Cell* 159:1212–1226.
53. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437:1173–1178.
54. Fischer L, Chen ZA, Rappsilber J (2013) Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J Proteomics* 88:120–128.
55. Merkle ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN (2014) Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Sci* 23:747–759.
56. Gamper AM, Kim J, Roeder RG (2009) The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis. *Mol Cell Biol* 29:266–280.
57. Han Y, Luo J, Ranish J, Hahn S (2014) Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J* 33:2534–2546.
58. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C (2010) The genetic landscape of a cell. *Science* 327:425–431.
59. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* 72:248–254.
60. Laemmli UK (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227:680–685.
61. Burnette WN (1981) “Western Blotting”: Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radiiodinated protein A. *Anal Biochem* 112:195–203.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

LC-MS/MS data acquisition

1 pmol of digested peptides was injected onto a pre-column (C18 Dionex Acclaim PepMap100, 75 μm i.d. x 2 cm, 3 μm , 100 \AA) with H₂O/ACN/FA (98/2/0.1) at 5 $\mu\text{L}/\text{min}$ for 3.5 minutes. Chromatographic separation was carried out on a C18 column (Thermo Scientific Accucore, 75 μm i.d. x 50 cm, 2.6 μm , 150 \AA) at 40°C by using 0.1% formic acid in water (mobile phase A) and 0.08% formic acid in 80% acetonitrile (mobile phase B) with a 150 min-long gradient (5 to 50% phase B for 120 min, 50 to 99% phase B for 1 min, 99% phase B for 10 min, 99 to 5% phase B for 1 min, 5% phase B for 18 min) at a flow rate of 200 nL/min. The peptides were then analyzed on a Thermo Scientific Orbitrap Elite with HCD fragmentation. The ion accumulation was set at 100 ms and 4 x 10⁴ ions. The precursors with unknown charge state or a charge state of +1 were excluded. Peptides selected for MS/MS acquisition were then placed on an exclusion list for 30 s using the dynamic exclusion mode to limit duplicate spectra. The minimum signal count was set at 500.

MS data analysis

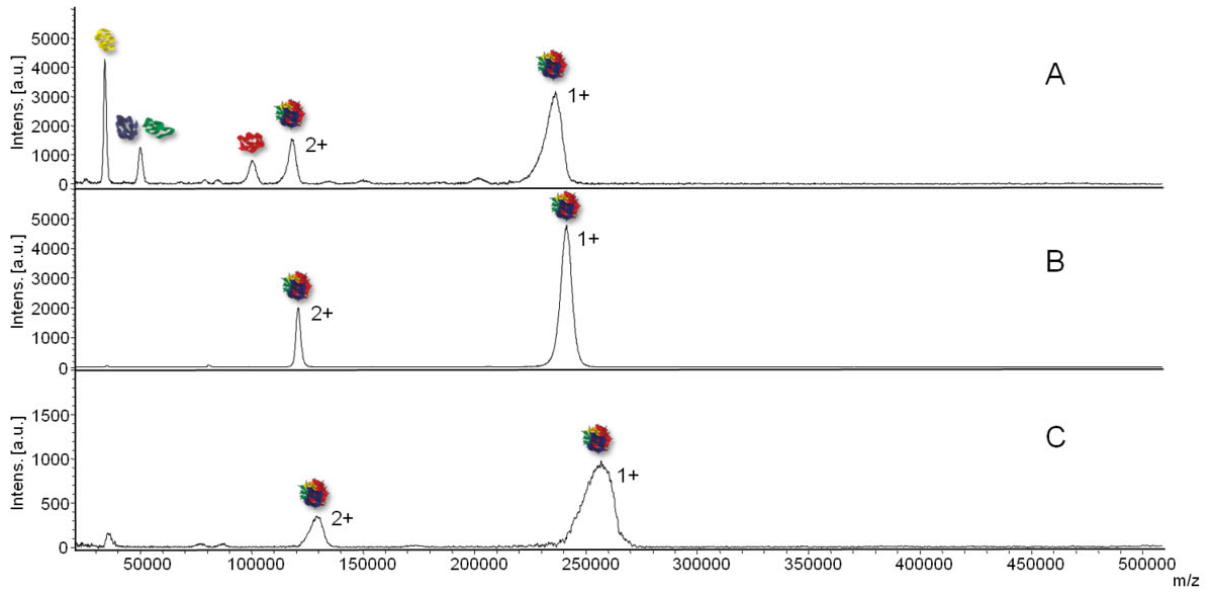
The MS raw data were converted to MGF and MZML files using the Proteome Discoverer™ Software (Thermo Scientific, v1.4). Then the MZML files were converted to mzXML files by use of MSConvert tools (ProteoWizard 3.0.4743). For the identification by pLink and xQuest, the database was constituted of the sequences of four subunits (UniProt), the sequence of GCN5 contains additionally a histidine-tag.

SAGA HAT sequence analysis and structural modeling

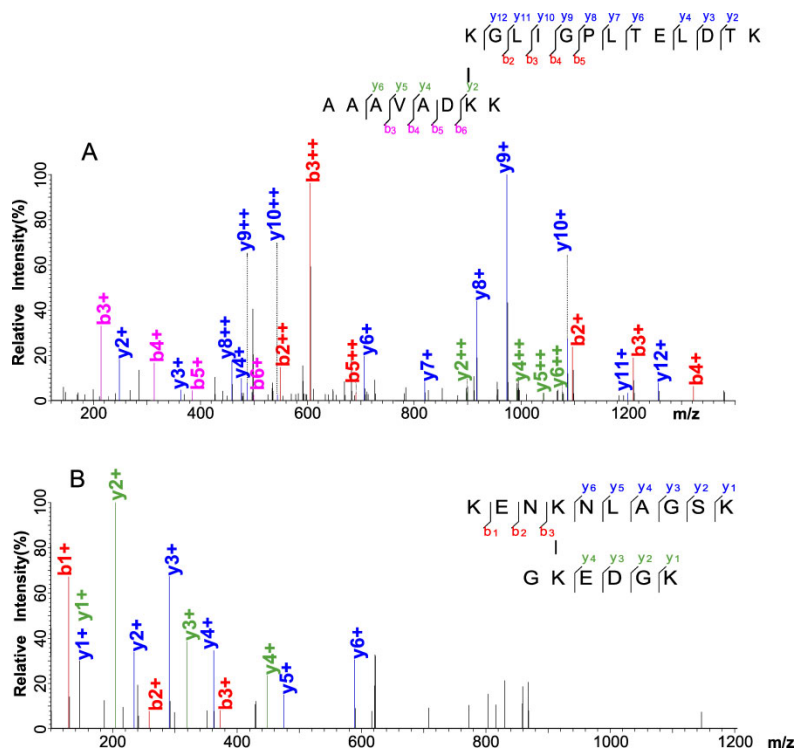
Secondary structure analysis of SAGA HAT subunits was done in PSIPRED v3.3 (<http://bioinf.cs.ucl.ac.uk/psipred>) to determine helical and beta-sheet containing regions and MARCOIL v1.0 (<http://toolkit.tuebingen.mpg.de/marcoil>) to predict coiled-coil domains. To identify conserved domains and residues sequence alignment of yeast, drosophila, mouse and human SAGA HAT subunits was performed using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Homology searches and template-based structure predictions were done with HHPRED (<http://toolkit.tuebingen.mpg.de/hhpred>) using HMM PDB database with default settings. ADA2B ZZ and SWIRM domains, GCN5 PCAF domain were searched and aligned with homologous proteins with known structures. One to several best scoring templates was chosen to create PIR alignment file with "Create model" function in HHPRED. These alignments were subsequently used as input files for Modeller v9.14 to create 3D homology models. Defaults settings were used except additional distance restraints were added based on cross-linking data (distance between C α atoms of cross-linked lysines was set below theoretical maximum of 30 \AA). Ten models were generated for each protein domain and the best scoring model (according to DOPE score) was further refined with automatic loop refinement function in Modeller with the same distance restraints. This allowed improving the disordered

loop regions of each domain. Protein structures of four overlapping domains of ADA3 protein (1-116, 110-189, 183-277 and 273-432) were generated by Robetta (<http://robetta.bakerlab.org>) with default threshold and settings. Complete models of GCN5, ADA2B and ADA3 were generated using rigid-body fitting Patchdock program (<http://bioinfo3d.cs.tau.ac.il/PatchDock/>) with defaults thresholds, except additional cross-linking restraints were added as before. Ten top scoring results of molecular docking were further refined with FireDock (<http://bioinfo3d.cs.tau.ac.il/FireDock>). Visualization of the results and distance measurements were performed in UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>). The quality of the final models was verified using ProQ, QMEAN, SelectPro and Verify3D algorithms, all models received good to fair scores (Supplementary Table 4).

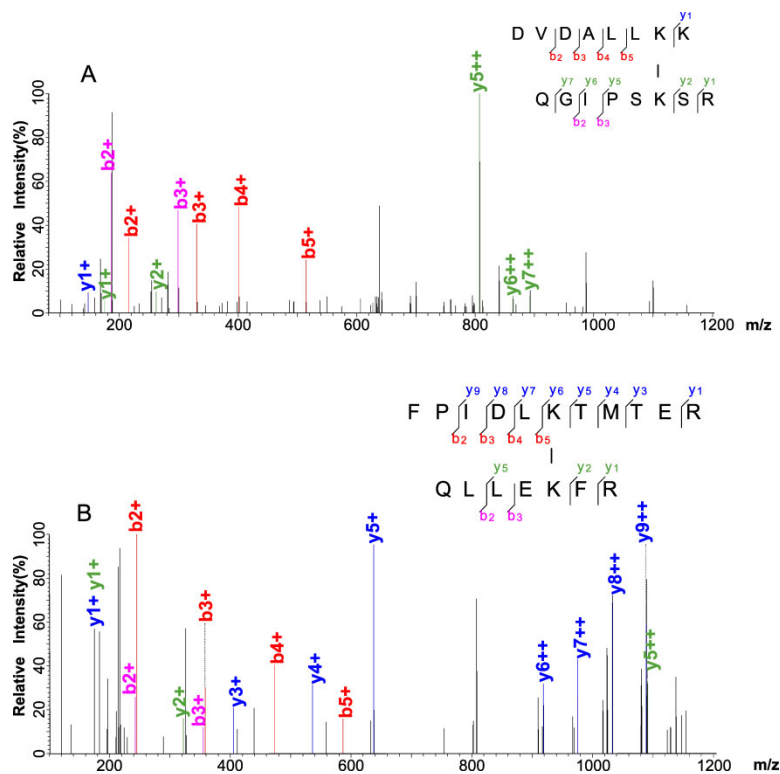
SUPPLEMENTARY FIGURES AND TABLES



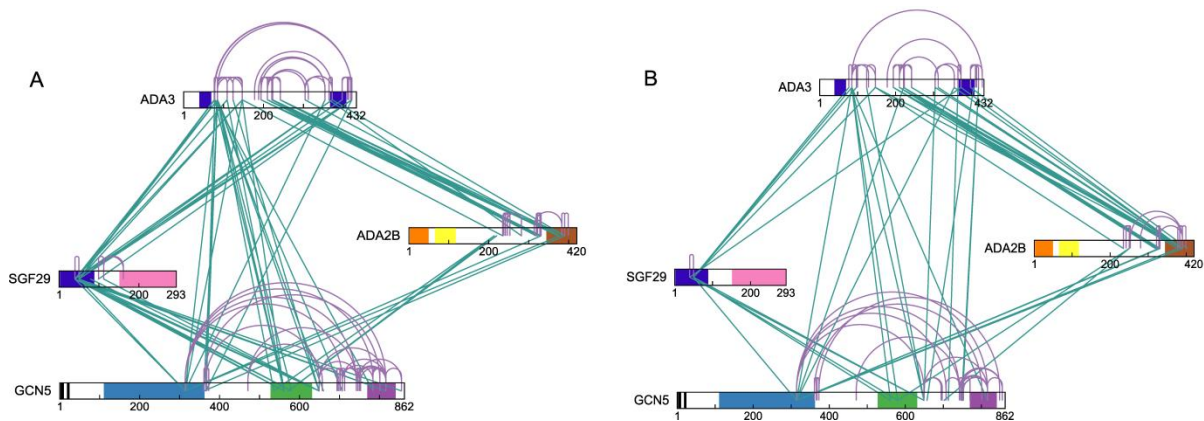
Supplementary Figure 1. MALDI-MS spectra of the cross-linked complex with different ratios protein:cross-linker BS3 for 30 min incubation time. (A) Protein:cross-linker 1:50. (B) Protein:cross-linker 1:75. (C) Protein:cross-linker 1:200. The optimal conditions for cross-linking reaction were fixed at 1:75.



Supplementary Figure 2. MS/MS spectra of intramolecular cross-linked peptides as identified by pLink. (A) Intralink within ADA3, score E-10. (B) Intralink within ADA2B, score E-10.



Supplementary Figure 3. MS/MS spectra of low scored cross-linked peptides as identified by pLink. (A) Interlink between ADA3-ADA2B, score E-06. (B) Intralink within GCN5, score E-06.



Supplementary Figure 4. Cross-linking map within the SAGA HAT subunits obtained by (A) pLink, (B) xQuest.

Supplementary Table 1. Detailed list of intermolecular cross-links between SAGA HAT subunits. The bolded underlined lysines are cross-linked. Sequence1 and Sequence2 are the sequences of the linked peptides.

GCN5-ADA3							
Linked K in GCN5	Linked K in ADA3	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	85	SQAEDVATY K VNYTR	FL K LGR	5,26E-14	55	2614,3754	-0,4
314	147	SQAEDVATY K VNYTR	IP K NDAPNR	3,65E-12		2905,4569	1,1
314	221	SQAEDVATY K VNYTR	AAAVAD K K		45,71	2654,3527	-0,9
314	354	SQAEDVATY K VNYTR	QAEL K ALSAHNR	1,56E-13	44,66	3218,6297	-0,7
361	85	QLLE K FR	FL K LGR	5,47E-10		1803,0770	0,8
366	78	VE K DK	K GDR	4,89E-07		1229,6615	-0,7
366	85	VE K DK	FL K LGR	7,13E-07		1491,8961	0,7
366	421	VE K DK	TL K ER	5,47E-09		1400,7874	-0,1
536	85	GIEFHVIGNSLTP K ANR	FL K LGR	1,74E-06		2835,6122	0,2
561	78	MP K EYIAR	K GDR	2,27E-07		1622,8750	-0,1
561	85	MP K EYIAR	FL K LGR	1,15E-08	50,64	1877,0595	-0,1
561	109	MP K EYIAR	LEG K TGHGPGPGPGR	1,15E-09		2560,3219	1,3
561	147	MP K EYIAR	IP K NDAPNR	1,37E-13		2172,1661	-0,6
580	85	TLAL K DGR	FL K LGR	1,28E-11	45,79	1856,1238	-0,5
649	85	K QGFSK	FL K LGR	8,37E-10	51,03	1563,9136	-0,7
649	124	K QGFSK	S KNVQPK	2,08E-10	45,04	1630,9042	-0,3
649	304	K QGFSK	NQN K PFSVPHTK		36,54	2227,1581	-7,6
657	405	DI K VPK	IMAAR K KR		27,06	1937,1456	-7,9
660	85	VP K SR	FL K LGR	1,29E-06		1455,8925	-0,8
707	85	K QAQIR	FL K LGR	1,03E-08		1617,0026	-0,3
713	364	K VYPGLSCFK	K HDLR		35,84	2059,1279	-0,5
748	418	E K GKELK	DQAW K TLK		32,19	1957,0869	-0,8
750	418	E K GKELK	DQAW K TLK		44,43	1957,0843	-2,1
784	78	K SEAPDYEVIR	K GDR	9,97E-07		2081,0428	2,5
816	85	K LFVADLQR	FL K LGR		35,8	1959,1605	-3,3
GCN5-ADA2B							
Linked K in GCN5	Linked K in ADA2B	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	328	SQAEDVATY K VNYTR	NLAG S KR		44,78	2626,3341	-0,4
314	331	SQAEDVATY K VNYTR	G KEDGK	5,50E-07	57,22	2514,2217	-0,8
314	378	SQAEDVATY K VNYTR	TIH K DHLQK	7,20E-09	42,5	3089,6396	-3,3
314	390	SQAEDVATY K VNYTR	QGIP S KSR		40,44	2753,3972	-0,5
361	390	QLLE K FR	QGIP S KSR	4,30E-06		1942,0999	-2,7
536	214	GIEFHVIGNSLTP K ANR	L KER	2,43E-06		2651,5058	-0,7
572	214	LVFDP K HK	L KER	2,85E-08		1664,9613	-1,1
572	220	LVFDP K HK	K NIAR	7,75E-13		1720,9987	-0,8
580	250	TLAL K DGR	RKIT K EK		38,81	2154,2570	-7,7
580	250	TLAL K DGR	IT K EK		35,23	1870,0746	-1,6
702	250	K LIERK	IT K EK		29,07	2645,4865	-3,5
GCN5-SGF29							
Linked K in GCN5	Linked K in SFG29	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	40	SQAEDVATY K VNYTR	SEHNLVNIQ K THER	4,31E-11		3585,7811	-1,3
314	50	SQAEDVATY K VNYTR	MQ TEN KISPYR	2,09E-08	38,84	3410,6459	0,2
561	40	MP K EYIAR	SEHNLVNIQ K THER	2,71E-16		2852,4903	1,1
561	50	MP K EYIAR	MQ TEN KISPYR	5,72E-12	44,79	2673,3288	-0,2
580	40	TLAL K DGR	SEHNLVNIQ K THER	5,47E-09		2827,5303	1,9
580	50	TLAL K DGR	MQ TEN KISPYR	5,78E-10	51,09	2652,3947	0,1
649	40	K QGFSK	SEHNLVNIQ K THER	1,06E-16	34,72	2539,3443	-0,2
649	50	K QGFSK	MQ TEN KISPYR	4,39E-10	43,37	2360,1818	-0,7
649	79	K QGFSK	K ALDK	1,49E-09		1404,7976	-0,5
649	112	K QGFSK	K TMR	6,92E-08		1365,7438	1,2
816	40	K LFVADLQR	SEHNLVNIQ K THER	4,67E-07		2934,5975	-6,7
816	50	K LFVADLQR	MQ TEN KISPYR	1,39E-14		2755,4366	0,5
855	50	L KEGGLDK	MQ TEN KISPYR	1,85E-06		2638,3675	-0,3
ADA3-ADA2B							
Linked K in ADA3	Linked K in ADA2B	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
147	235	IP K NDAPNR	DYNLVP A FLG K DK	7,92E-13	45,89	2644,4160	0,4
147	390	IP K NDAPNR	QGIP S KSR		42,1	2033,1002	-0,8
194	331	IPPLG K HYSQR	G KEDGK	3,05E-09	41,76	2069,1205	-0,1
194	378	IPPLG K HYSQR	TIH K DHLQK	7,49E-06		2640,5114	0,0
194	390	IPPLG K HYSQR	QGIP S KSR		35,29	2304,2634	-3,0
210	319	WAQEDLLEE K KGAR	K ENK	3,62E-06		2446,2276	0,6
210	319	WAQEDLLEE K KGAR	R K ENKNLAGSK	3,33E-08	30,06	3172,6412	-6,4
210	322	WAQEDLLEE K KGAR	EN K NLAGSK	1,63E-10	39,13	2884,4199	-0,1
210	322	WAQEDLLEE K KGAR	KEN K NLAGSK	2,88E-12	40,76	3016,5401	-0,1
210	328	WAQEDLLEE K KGAR	NLAG S KR	4,43E-06	56,43	2673,3658	0,2
210	331	WAQEDLLEE K KGAR	G KEDGK	7,99E-08	50,25	2557,2294	-0,1

Chapitre 3 : Etude par CX-MS de la stœchiométrie et de la topologie du complexe SAGA HAT

ADA3-ADA2B (continued)							
Linked K in ADA3	Linked K in ADA2B	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
210	378	WAQEDLLEE Q KDGAR	THI K DHLQK	8,13E-06		3132,6454	-0,6
210	390	WAQEDLLEE Q KDGAR	QGIP S KSR		33,39	2796,4038	-0,2
221	319	AAAVAD K KK	K ENK	4,79E-07		1431,8233	1,6
221	322	AAAVAD K KK	EN K NLAGSK	9,36E-09	297,22	1874,0409	1,1
221	322	AAAVAD K KK	KEN K NLAGSK	6,31E-15	45,29	1998,1108	-0,4
221	328	AAAVAD K KK	NLAG S KR	3,39E-14	42,52	1654,9365	0,5
221	331	AAAVAD K KK	G KEDGK	1,54E-09	44,44	1542,8210	-2,8
221	373	AAAVAD K KK	YVTV K THIK	1,63E-08		2087,2605	0,8
221	378	AAAVAD K KK	THI K DHLQK		30,64	2118,2407	-0,3
221	390	AAAVAD K KK	QGIP S KSR		41,6	1781,9974	-1,4
223	328	K GLIGPLTELDTK	NLAG S KR	3,72E-07		2270,3145	0,1
223	378	K GLIGPLTELDTK	THI K DHLQK	2,86E-06		2729,5941	0,0
242	328	DVDALL K KK	NLAG S KR	1,06E-06		1783,0202	-0,8
242	378	DVDALL K KK	THI K DHLQK	2,96E-11	45,58	2246,3254	0,2
242	390	DVDALL K KK	QGIP S KSR	9,38E-06	44,71	1910,0836	0,5
304	378	NQN K PFSVPHTK	THI K DHLQK	4,59E-06		2741,5227	-2,4
304	390	NQN K PFSVPHTK	QGIP S KSR		42,97	2405,2822	0,3
312	390	NQN K PFSVPHTKSLESR	QGIP S KSR		26,06	2977,5704	-1,0
354	390	QAE L KALSAHNR	QGIP S KSR		40,72	2346,2754	-0,6
364	282	K HDLR	E KMLR	3,79E-07		1597,9274	0,2
372	390	LA K EEVSR	QGIP S KSR	2,14E-07	41,31	1940,0690	0,1
397	403	K IMAAR	K RILNFLTESGWISR		29,07	2645,4865	1,0
ADA3-SGF29							
Linked K in ADA3	Linked K in SGF29	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
77	50	VLEAETQILTDWQD K KK	MQTEN K ISPYR	6,23E-14	33,99	3586,8165	0,2
78	50	K GDR	MQTEN K ISPYR	4,17E-07		2141,0575	1,3
85	40	FL K LGR	SEHNLVNIQ K THER	2,34E-08		2574,4029	-0,1
85	50	FL K LGR	MQTEN K ISPYR	1,70E-09	47,29	2399,2672	0,0
85	50	RFL K LGR	MQTEN K ISPYR		39,28	2555,3450	-9,1
97	50	DHELGAPP K HGKPK	MQTEN K ISPYR		34,14	3176,6059	-0,6
109	40	LEG K TGHGPGPGGR	SEHNLVNIQ K THER	4,30E-08		3257,6652	-0,3
147	40	IP K NDAPNR	SEHNLVNIQ K THER	1,57E-11		2869,5094	0,4
147	50	IP K NDAPNR	MQTEN K ISPYR	4,08E-10	45,14	2690,3476	-0,4
372	40	LA K EEVSR	SEHNLVNIQ K THER	1,38E-11	30,73	2772,4517	-1,3
372	50	LA K EEVSR	MQTEN K ISPYR	1,06E-06		2597,3158	-0,6
397	50	K IMAAR	MQTEN K ISPYR	4,09E-07		2355,2078	-0,1
421	40	TL K ER	SEHNLVNIQ K THER	6,71E-06		2487,3193	0,3
421	50	TL K ER	MQTEN K ISPYR	7,04E-07		2312,1834	0,4
421	99	TL K ER	IA K IAGLYNDSEPPR	2,34E-07		2501,3789	0,6

Supplementary Table 2. Detailed list of intramolecular cross-links between SAGA HAT subunits. The bolded underlined lysines are cross-linked. Sequence1 and Sequence2 are the sequences of the linked peptides.

GCN5-GCN5							
Linked K in GCN5	Linked K in GCN5	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
314	649	SQAEDVATY K VNYTR	K QGFSK		39,59	2575,2896	-0,9
314	707	SQAEDVATY K VNYTR	K QAQIR	9,81E-08	46,79	2624,3554	-0,2
314	742	SQAEDVATY K VNYTR	ETGW K PLGK	5,26E-08	36,5	2896,4580	-0,9
314	784	SQAEDVATY K VNYTR	K SEAPDYEVIR	5,16E-10		3350,6306	1,2
314	816	SQAEDVATY K VNYTR	K LFVADLQR	1,36E-12	46,81	2970,5450	0,2
361	366	QLLE K FR	VE K DK	1,03E-09		1691,9758	-1,5
361	368	QLLE K FR	DKLVPE K	6,68E-09		1902,1126	0,4
361	373	QLLE K FR	LVPE K FR	3,29E-07	47,42	1811,0658	-0,6
366	373	FRVE K DK	LVPE K FR		39,34	1799,0284	-1,2
368	361	VEKD K LVPE K FR	RQLLE K FR		35,02	2566,4723	-9,2
368	361	DKLVPE K	QLLE K FR		44,42	1898,0857	-1,0
373	361	DKLVPE K FR	QLLE K FR	2,10E-14	57,08	2058,2137	0,3
373	366	LVPE K FR	VE K DK	3,58E-08		1495,8609	-0,1
471	657	TLPENLTLEDA K FR	DI K VPK	2,98E-08		2339,3248	-1,4
471	742	TLPENLTLEDA K FR	ETGW K PLGK	3,09E-06	27,54	2651,4177	0,2
649	366	K QGFSK	VE K DK	7,43E-06		1452,8124	0,3
649	697	K QGFSK	Q K EIK		35,89	1588,9178	-0,7
694	649	IPYTELSHII K K	K QGFSK	3,98E-12	42,7	2276,3080	0,8
694	701	IPYTELSHII K K	EII K K	2,54E-09		2208,3132	-0,2
694	702	IPYTELSHII K K	K LIER	3,60E-06		2236,3194	0,2
697	649	Q K EIK	K QGFSK	2,77E-07		1588,9187	-0,7
707	701	K QAQIR	EII K K	2,22E-08		1509,9242	-1,0
707	702	K QAQIR	K LIER	1,01E-06		1541,9553	-0,6
713	707	K VYPGLSCFK	K QAQIR	1,84E-09		2021,1131	-3,7
742	707	ETGW K PLGK	K QAQIR	2,37E-13	48,82	1899,0878	-1,0
742	750	ETGW K PLGK	E K G K ELK		46,76	1983,1016	-1,3
742	816	ETGW K PLGK	K LFVADLQR		46,91	2241,2484	-1,7
742	816	ETGW K PLGKEK	K LFVADLQR		43,6	2498,3893	-0,2
742	855	ETGW K PLGKEK	L K EGGLIDK		34,12	2381,3111	-4,0
746	649	ETGW K PLGKEK	K QGFSK	7,28E-07		2103,1363	0,8
746	707	ETGW K PLGKEK	K QAQIR	4,74E-11		2156,2253	0,4
746	750	ETGW K PLGKEK	G K ELK	1,24E-06		1983,1040	-0,4
746	816	ETGW K PLGKEK	K LFVADLQR	5,40E-20	43,04	2498,3893	-0,2
746	855	ETGW K PLGKEK	L K EGGLIDK	3,13E-06	41,16	2381,3205	0,1
753	742	GKEL K DPDQLYTTLK	ETGW K PLGKEK	5,88E-07		3157,6909	0,0
753	742	EL K DPDQLYTTLK	ETGW K PLGKEK		40,49	2972,5712	-1,1
753	746	EL K DPDQLYTTLK	ETGW K PLGKEK	2,50E-06		2976,5995	-1,0
753	816	EL K DPDQLYTTLK	K LFVADLQR	6,88E-13	40,98	2789,5199	-0,6
753	816	GKEL K DPDQLYTTLK	K LFVADLQR		34,12	2974,6384	0,2
770	753	NLLAQI K SHPSAWPFMEPVK	EL K DPDQLYTTLK		26,58	3993,0957	-0,1
784	707	K SEAPDYEVIR	K QAQIR	1,79E-09		2353,2578	-0,7
784	801	K SEAPDYEVIR	FPIDL K TMTER	4,99E-14		2956,4891	0,9
784	816	K SEAPDYEVIR	K LFVADLQR	9,68E-12		2699,4470	-1,0
801	361	FPIDL K TMTER	QLLE K FR	1,06E-06	51,3	2420,3136	0,2
816	649	K LFVADLQR	K QGFSK	4,48E-09	37,95	1924,1082	0,4
816	707	K LFVADLQR	K QAQIR	2,88E-08		1973,1722	1,1
816	742	K LFVADLQR	ETGW K PLGK	3,16E-17		2245,2770	0,4
816	855	K LFVADLQR	L K EGGLIDK	3,54E-10	40,21	2198,2673	0,3
848	373	CASALE K FFYFK	DKLVPE K FR		30,13	2574,3460	-3,7
ADA3-ADA3							
Linked K in ADA3	Linked K in ADA3	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
77	85	VLEAETQILTDWQD K K	FL K LGR	2,70E-07	35,67	2786,5211	-0,2
77	421	VLEAETQILTDWQD K K	TL K ER	1,86E-08		2699,4380	0,1
85	78	FL K LGR	K GDR	8,77E-08		1348,8127	-0,3
85	421	FL K LGR	TL K ER	2,07E-11		1515,9136	0,0
97	85	DHELGA P PKHG K PK	FL K LGR		36,04	2380,3330	-2,1
100	85	DHELGA P PKHG K PK	RFL K LGR		38,78	2536,4174	-8,5
109	85	LEG K TGHGPGPGGR	FL K LGR	5,56E-07		2286,2595	-1,5
109	147	LEG K TGHGPGPGGR	IP K NDAPNR	1,10E-09		2577,3410	-0,8
124	85	SK N VQPK	FL K LGR	9,73E-11		1669,9878	-0,5
129	85	NVQ P KIQEYFTDDPIDVPR	FL K LGR		38,76	3272,7109	0,8
147	85	IP K NDAPNR	FL K LGR	4,37E-07	49,79	1898,1038	-1,5
147	124	IP K NDAPNR	SK N VQPK	1,16E-13	44,55	1965,0943	-0,3
178	372	TLEELL K PPEDA A EHYK	LAK E EVSR	6,20E-11	30	3108,5888	0,7
188	354	TLEELL K PPEDA A EHY K IPPLGK	QAEL K ALSAHNR	2,29E-12		4120,1843	-1,2
194	221	IPPL K HYSQR	AA A VAD K K	5,72E-08	45,37	2209,2519	0,3
194	354	TLEELL K PPEDA A EHY K IPPLG K HYSQR	RQAEL K ALSAHNR	2,85E-08		4948,5834	1,4
210	194	WAQEDLLEE K QDGAR	IPPL K HYSQR	4,08E-10	38,14	3223,6561	0,1
210	221	WAQEDLLEE K QDGAR	AA A VAD K K	9,69E-07	38,24	2829,4808	-1,0
210	221	WAQEDLLEE K QDGAR	AA A VAD K KK		32,87	2825,4543	-0,6

Chapitre 3 : Etude par CX-MS de la stœchiométrie et de la topologie du complexe SAGA HAT

ADA3-ADA3 (continued)							
Linked K in ADA3	Linked K in ADA3	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
210	223	WAQEDLLEE Q KD GAR	K GLIGPLTELDTK	1,46E-08		3308,7139	0,7
210	242	WAQEDLLEE Q KD GAR	DVDALL KK	1,90E-07	36,62	2825,4446	0,1
210	304	WAQEDLLEE Q KD GAR	NQN K PFSVPHTK	2,97E-09		3324,6674	1,8
221	242	AAAVAD KK	DVDALL KK		47,06	1811,0398	-0,4
223	221	K GLIGPLTELDTK	AAAVAD KK	9,47E-10	28,55	2294,3078	-0,8
223	242	K GLIGPLTELDTK	DVDALL KK	6,01E-16	38,12	2422,3945	0,4
242	221	DVDALL KK	AAAVAD KK	8,65E-15		1811,0403	0,6
304	354	NQN K PFSVPHTK	QAEL K ALS AHN R	5,99E-15	38,29	2870,5150	0,2
304	372	NQN K PFSVPHTK	LA K EEVSR	2,45E-09	30,49	2464,3073	0,3
312	210	NQN K PFSVPHT K SLES R	WAQEDLLEE Q KD GAR	2,09E-06	29,25	3896,9593	-0,2
372	364	LA K EEVSR	K HDLLR	2,75E-20	57,94	1853,0670	-0,1
411	397	K EKDQAWK	K IMAAR	9,24E-12	48,55	1858,0135	0,0
411	421	K EKDQAWK	TL K ER	1,43E-07		1814,9889	-0,5
413	410	E K DQAWK	TPT KK	3,91E-06		1618,8866	-0,4
413	410	TPT K KE K DQAWK	TPT K KE K		39,65	2427,3510	5,6
413	421	E K DQAWK	TL K ER	1,31E-08		1686,8940	0,5
418	397	DQAW K TLK	K IMAAR	2,57E-12	53,85	1815,0069	-0,4
418	410	EKDQAW K TLK	TPT KK	1,74E-06		1961,1133	-0,5
421	78	TL K ER	K GDR	2,44E-07		1257,7040	0,1
421	410	TL K ER	TPT KK	5,98E-08		1356,7976	-0,3
428	85	ESIL K LLDG	FL K LGR		35,56	1857,0957	-1,0
428	397	TLKERESIL K LLDG	K IMAAR		30,37	2440,4072	-0,6
ADA2B-ADA2B							
Linked K in ADA2B	Linked K in ADA2B	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
235	253	DYNLVP AFLG K DK	EE K ELR	2,10E-08	40,36	2419,2634	0,5
237	242	D K KEK	E K ALK	3,81E-06		1371,7972	0,1
237	247	D K KEK	K ITK	1,72E-07		1272,7652	0,7
238	242	D K KEK	E K ALK	2,83E-08		1371,7972	1,1
238	247	D K KEK	K ITK	9,25E-06		1272,7652	-0,2
242	247	E K ALK	K ITK	4,36E-06		1213,7645	-0,5
242	253	E K E K ALK	EE K ELR	2,13E-09	52,26	1784,9886	0,1
250	253	K IT K EEK	EE K ELR		31,09	1814,9965	-1,3
253	238	EE K ELR	D K KEK	5,47E-07		1586,8514	0,3
253	242	EE K ELR	E K ALK	6,09E-06		1527,8507	0,4
253	242	IT K EE K ELR	E K ALK	5,67E-07		1870,0774	-1,1
253	247	EE K ELR	K ITK	8,05E-07		1432,8437	-0,6
253	250	EE K ELR	IT K EEK		34,11	1686,8928	-6,6
280	390	EFDDL FENMH K EK	QGIP S KSR		39,28	2690,2995	-0,6
282	247	E K M LR	K ITK	9,49E-06		1305,7990	-0,4
314	383	H K REKR	DHL Q KR	2,17E-06		1786,9913	-5,8
322	331	EN K NLAGSK	G K EDGK		306,33	1729,8794	-3,1
322	331	KEN K NLAGSK	G K EDGK	1,76E-14	49,1	1862,0045	-0,5
322	331	KEN K NLAGSK	RG K EDGK		26,7	2014,0787	-1,0
328	319	NLAG S KR	K ENK	2,19E-07		1399,7782	0,0
328	331	NLAG S KR	G K EDGK	7,86E-09	47,48	1518,8302	0,1
331	319	G K EDGK	K ENK	3,76E-07		1291,6919	-1,2
378	390	TH K DHLQK	QGIP S KSR		41,78	2217,2841	-0,2
390	253	QGIP S KSR	EE K ELR		36,79	1811,9725	-0,9
390	328	QGIP S KSR	NLAG S KR		38,11	1753,9810	0,6
390	383	QGIP S KSR	DHL Q KR	2,23E-10	45,56	1809,0157	0,0
399	390	LPSYLD K VLK	QGIP S KSR	8,86E-07	50,7	2184,2526	0,4
SGF29-SGF29							
Linked K in SGF29	Linked K in SGF29	Sequence1	Sequence2	Score in pLink	Score in xQuest	Mass (Da)	Mass precision (ppm)
40	50	SEHNLVNI Q KTHER	MQTEN K ISPYR	2,16E-17	36,09	3370,6721	-0,2
99	112	IAA K IAGLYNDSEPPR	K TMR	2,29E-06		2406,2877	-0,1
99	161	IAA K IAGLYNDSEPPR	PGD K VAAR	3,36E-11		2664,4234	-0,8
161	112	PGD K VAAR	K TMR	2,88E-08		1488,8382	-0,2

Supplementary Table 3. Detailed list of dead-end, inter-linked, intra-linked and non-cross-linked lysines.

K position in GCN5	GCN5			
	Non-cross-linked	Dead-end	Inter-linked	Intra-linked
116		X		
126		X		
127	X			
130	X	X		
138		X		
144		X		
149				
152		X		
213		X		
220	X			
228	X	X		
232	X			
253	X			
267				
284	X			
294	X			
314	X	X	X	X
361	X	X	X	X
366	X	X	X	X
368	X	X		X
373	X	X		X
383	X	X		
458	X	X		
471	X	X		X
536	X	X	X	
561		X	X	
572		X	X	
574		X		
580		X	X	
613	X			
624	X	X		
629	X			
648		X		
649		X	X	X
654		X		
657		X	X	X
660	X	X	X	
668	X			
694	X	X		X
695	X	X		
697	X	X		X
701		X		X
702	X		X	X
707	X	X	X	X
713	X	X	X	X
722		X		
742	X	X		X
746	X	X		X
748	X	X	X	
750	X	X	X	X
753	X	X		X
763	X	X		
770	X	X		X
783		X		
784	X	X	X	X
801		X		X
816		X	X	X
848				X
853		X		
855		X	X	X
862		X		

K position in ADA3	ADA3			
	Non-linked	Dead-end	Inter-linked	Intra-linked
5	X			
15	X	X		
21	X	X		
77	X	X	X	X
78	X	X	X	X
85		X	X	X
97	X	X	X	X
100		X		X
102		X		
103		X		
105	X	X		
109	X	X	X	X
122	X	X		
124	X	X	X	X
129	X	X		X
147	X	X	X	X
178	X	X		X
188	X	X		X
194	X	X	X	X
210	X	X	X	X
221	X	X	X	X
222	X	X		
223	X	X	X	X
235	X	X		
242	X	X	X	X
243	X	X		
288	X			
304	X	X	X	X
312	X	X	X	X
319	X	X		
348	X	X		
354	X	X	X	X
363	X			
364	X	X	X	X
372	X	X	X	X
397	X	X	X	X
404	X			
405	X		X	
410	X	X		X
411		X		X
413	X	X		X
418	X	X	X	X
421	X	X	X	X
428	X	X		X

K position in ADA2B	ADA2B			
	Non-linked	Dead-end	Inter-linked	Intra-linked
6	X			
7	X			
121	X	X		
202	X			
212				
214		X	X	
220			X	
235	X	X	X	X
237				X
238		X		X
240		X		
242		X		X
245		X		
247		X		X
250		X	X	X
253		X		X
258				
269		X		
280	X	X		X
282	X	X	X	X
287				
301	X	X		

ADA2B (continued)				
K position in ADA2B	Non-linked	Dead-end	Inter-linked	Intra-linked
314				X
317		X		
319		X	X	X
322		X	X	X
328		X	X	X
331	X	X	X	X
335	X	X		
356	X	X		
373	X	X	X	
378	X	X	X	X
383	X	X		X
390	X	X	X	X
399	X	X		X
402		X		
403		X	X	
SGF29				
K position in SGF29	Non-linked	Dead-end	Inter-linked	Intra-linked
22	X	X		
40	X	X	X	X
50	X	X	X	X
58		X		
67	X			
79		X	X	
83	X	X		
88	X	X		
99		X	X	X
112	X	X	X	X
136	X			
140	X			
161				X
167		X		
189	X			
200	X			
217	X			
230	X			
288		X		
291				
292				
293				

Supplementary Table 4. Quality evaluations of domain models. For each homology model of the protein domains the quality assessment was performed in several bioinformatics software packages.

Domain	Residue number	Templates used	HHpred	% identity	ProQ LGscore	QMEAN	SelectPro	Verify3D
hADA2B SWIRM	330-418	2aqe_A	99,9	25	1,403	0,781	0,661	0,17
hADA2b ZZ	7-56	2e5r_A	99,2	30	0,576	0,573	0,357	0,19
hGCN5 PCAF	177-399	1w9c_A 3m1i_C	NA	NA	1,892	0,382	0,622	0,2
mADA3	1-432	NA	NA	NA	2,756	0,432	0,783	0,21

3.7. Perspective : application à l'étude des complexes protéiques membranaires ?

Malgré les quelques exemples de protéines membranaires ou de complexes de protéines membranaires étudiés avec succès en ESI-MS [31,32] ou en MALDI-MS [33,34], l'investigation des protéines membranaires par MS reste délicate à cause de l'utilisation des détergents nécessaires à leur solubilisation. Une approche récemment appliquée avec succès est d'étudier des complexes de protéines membranaires par CX-MS [35-37].

Dans ce contexte, nous avons voulu étendre l'approche de pontage développée sur SAGA HAT sur des systèmes de protéines membranaires, et de voir si le pontage marche toujours même en présence de grande concentration de détergent (au-dessus de la CMC, concentration micellaire critique). Pour cela, nous avons conduit les études préliminaires sur un complexe hydrosolubles supportant une forte présence en détergent.

Avant mon arrivée au laboratoire, des essais avaient montré le potentiel du kit de pontage K100 proposé par CovalX pour ponter avec succès le complexe RAR α Δ AB-RXR α Δ AB (récepteur de l'acide rétinoïque - récepteur X des rétinoïdes). De plus, ce complexe supporte et reste formé à grande concentration de détergent. Par conséquent, nous avons décidé de continuer les tests de pontage sur ce complexe en grande concentration de détergent (10 mM de CHAPS).

Le kit de pontage K100 a été optimisé et est commercialisé par CovalX AG (Zürich, Suisse). C'est un mélange de 3 agents de pontage (Figure 19):

- SBAT (1,1'-(suberoyldioxy) bisazabenzotriazole, longueur du bras espaceur 11,4 Å)
- SBBT (1,1'-(suberoyldioxy) bisbenzotriazole, longueur du bras espaceur 11,4 Å)
- GBAT (glutaroyldioxy bisazabenzotriazole, longueur du bras espaceur 7,7 Å)

Ces agents de pontage réagissent sur les amines primaires de la même façon que les esters NHS. Cependant, la présence des groupements aromatiques dans leurs structures permet une plus grande réactivité grâce à la délocalisation des électrons [38]. Ainsi, le mélange de ces trois agents de pontage de différentes longueurs du bras espaceur capables de lier spécifiquement des complexes de protéines permet d'augmenter l'efficacité du pontage [39]. Il est vendu en kit sous format de tubes de 2 mg contenant entre autres sa solution de solubilisation (diméthylformamide ou DMF). Il est conseillé pour stabiliser les complexes de masse de 10-100 kDa.

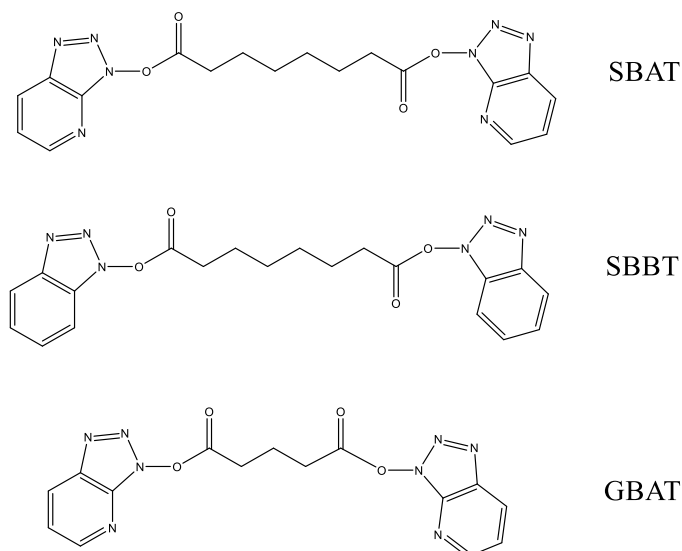


Figure 19. Formules chimiques du K100.

De même que pour le complexe SAGA HAT, des optimisations de conditions de réaction de pontage et de détection MALDI-MS ont été effectuées. Sur la Figure 20 sont présentés des spectres MALDI-MS du complexe RAR-RXR après pontage dans différents tampons :

- (A) 20 mM HEPES + 2 mM CHAPS + 300 mM NaCl
- (B) 20 mM HEPES + 10 mM CHAPS + 300 mM NaCl

En présence de 10 mM CHAPS (au-dessus de la CMC du CHAPS qui est de 6 mM) et de 300mM NaCl, le pic à 86 kDa correspondant au complexe ponté est détecté, ce qui suggère que la réaction de pontage a lieu même en grandes concentrations du détergent CHAPS et du NaCl.

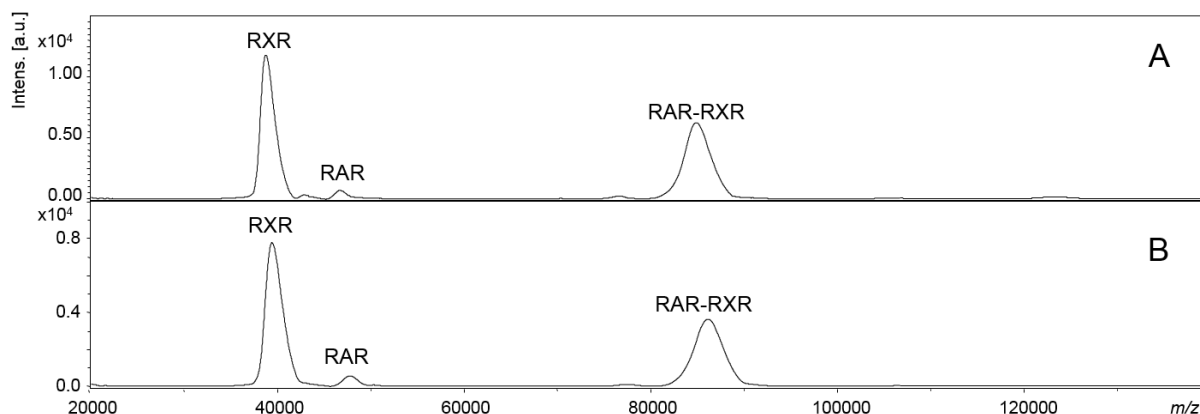


Figure 20. Spectre MALDI-MS du complexe RAR-RXR à 5 μ M après pontage avec le K100 (matrice SA, dépôt goutte séchée) dans (A) 20 mM HEPES + 2 mM CHAPS + 300 mM NaCl, (B) 20 mM HEPES + 10 mM CHAPS + 300 mM NaCl. La réaction de pontage a été conduite avec le K100 à 2 mg/mL pendant 30 min à température ambiante.

Ainsi, ce résultat encourage grandement des études à venir en CX-MS dans le cadre d'une collaboration avec l'équipe du Dr. Jean-Pierre Simorre à l'Institut de Biologie Structurale de

Grenoble sur un complexe de protéines membranaires PBP1B (« Penicilling Binding Protein 1B ») demandant de grandes concentrations en sel et en détergent.

3.8. Conclusion

Notre travail de pontage couplé à la MS a été mis à profit dans le but de dresser une carte d'interactions entre les sous-unités de SAGA HAT, ce qui a permis à nos collaborateurs à l'IGBMC d'affiner par modélisation moléculaire un premier modèle à basse résolution du complexe SAGA HAT.

Afin de compléter cette carte et à partir des mêmes données expérimentales MS/MS obtenues, nous pourrions modifier les paramètres d'interrogation des logiciels d'identification, comme par exemple en autorisant les pontages entre les acides aminés autres que les lysines. En effet, comme discuté en chapitre 2, même si les agents de pontage de type ester NHS réagissent majoritairement sur les lysines, les réactions secondaires sur les serines, tyrosines *etc.* peuvent avoir lieu. Dans ce sens, les pontages créés entre des résidus autres que les lysines peuvent apporter des informations structurales complémentaires, surtout dans des régions du complexe où peu de lysines sont présentes. Il serait également intéressant d'utiliser d'autres techniques de fragmentation, comme par exemple l'ETD (« Electron-Transfer Dissociation ») ou l'ETHcD (Electron-Transfer and higher-energy collision Dissociation), pour obtenir d'autres jeux de données. Une autre perspective consisterait à réaliser la réaction de pontage de SAGA HAT avec d'autres types d'agent de pontage. En effet, dans notre étude, nous avons identifié les zones de proximité grâce aux pontages par le BS3. Une réaction de pontage avec des agents contenant d'autres fonctions réactives (agents de pontage photoréactifs, agents de pontage réagissant sur des thiols *etc.*) permettrait d'obtenir d'autres zones de proximité entre d'autres acides aminés. Il serait également envisageable d'ajouter une étape d'enrichissement avant l'analyse du mélange peptidique en MS/MS, comme par exemple la chromatographie d'échange cation (SCX), afin de mieux détecter les peptides pontés. Enfin, se placer dans des conditions de pontage minimal (avec des sous-unités non pontées complètement) serait aussi utile, afin d'obtenir un minimum d'hétérogénéités dans les formes pontées du complexe.

Les premiers tests de réaction de pontage dans un milieu très concentré en sel et en détergent encouragent les études à venir sur des complexes de protéines membranaires par CX-MS. Dans ce sens, il serait intéressant de transférer des étapes réalisées dans le cadre de ces travaux de thèse sur les types de système dont l'investigation est délicate par la MS supramoléculaire « classique », comme des complexes protéiques membranaires ou des systèmes biologiques dont la structure est flexible.

En résumé, ce travail a permis de confirmer le potentiel du pontage chimique couplé à la MS pour étudier la stœchiométrie et les zones de proximité des sous-unités des complexes biomoléculaires. Cette méthode s'intègre dans l'éventail des techniques de biologie structurale permettant l'accès à la structure des protéines et à leurs interactions. Ceci permettra d'améliorer la précision des modèles structuraux et de s'approcher le plus possible à l'état natif des complexes à un moment donné de leurs vies. On parle aujourd'hui d'approche intégrative, et le couplage CX-MS y jouera un rôle important.

Bibliographie

- [1] Nagy, Z., Tora, L., Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* 2007, 26, 5341–5357.
- [2] Spedale, G., Timmers, H.T.M., Pijnappel, W.W.M.P., ATAC-king the complexity of SAGA during evolution. *Genes Dev.* 2012, 26, 527–541.
- [3] Habeeb, A.F.S.A., Hiramoto, R., Reaction of proteins with glutaraldehyde. *Arch. Biochem. Biophys.* 1968, 126, 16–26.
- [4] Migneault, I., Dartiguenave, C., Bertrand, M.J., Waldron, K.C., Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques* 2004, 37, 790–806.
- [5] Wine, Y., Cohen-Hadar, N., Freeman, A., Frolow, F., Elucidation of the mechanism and end products of glutaraldehyde crosslinking reaction by X-ray structure analysis. *Biotechnol. Bioeng.* 2007, 98, 711–718.
- [6] Chen, F., Applying MALDI-ToF Mass Spectrometry to Study Non-covalent Interactions. Thèse ETH Zürich, 2014.
- [7] Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., et al., High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2000, 97, 5802–5806.
- [8] Back, J.W., de Jong, L., Muijsers, A.O., de Koster, C.G., Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* 2003, 331, 303–313.
- [9] Huang, B.X., Kim, H.-Y., Dass, C., Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2004, 15, 1237–1247.
- [10] Miteva, Y.V., Budayeva, H.G., Cristea, I.M., Proteomics-Based Methods for Discovery, Quantification, and Validation of Protein–Protein Interactions. *Anal. Chem.* 2013, 85, 749–768.
- [11] Fischer, L., Chen, Z.A., Rappsilber, J., Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* 2013, 88, 120–128.
- [12] Sinz, A., Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* 2003, 38, 1225–1237.
- [13] Hermanson, G.T., *Bioconjugate Techniques*, Academic Press, 2013.
- [14] Besselink, G. a. J., Beugeling, T., Bantjes, A., N-Hydroxysuccinimide-activated glycine-sepharose. *Appl. Biochem. Biotechnol.* 1993, 43, 227–246.
- [15] Paramelle, D., Miralles, G., Subra, G., Martinez, J., Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* 2013, 13, 438–456.
- [16] Merkley, E.D., Rysavy, S., Kahraman, A., Hafen, R.P., et al., Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Sci.* 2014, 23, 747–759.
- [17] Hillenkamp, F., Peter-Katalinic, J., *MALDI MS: a practical guide to instrumentation, methods and applications*, Wiley-VCH, Weinheim 2007.
- [18] Merkley, E.D., Cort, J.R., Adkins, J.N., Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J. Struct. Funct. Genomics* 2013, 14, 77–90.
- [19] Nguyen-Huynh, N.-T., Sharov, G., Potel, C., Fichter, P., et al., Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex. *Protein Sci.* 2015, 24, 1232–1246.
- [20] Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., et al., Identification of cross-linked peptides from complex samples. *Nat. Methods* 2012, 9, 904–906.

- [21] Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., et al., Identification of cross-linked peptides from large sequence databases. *Nat. Methods* 2008, 5, 315–318.
- [22] Tran, B.Q., Goodlett, D.R., Goo, Y.A., Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 2015.
- [23] Peri, S., Steen, H., Pandey, A., GPMW – a software tool for analyzing proteins and peptides. *Trends Biochem. Sci.* 2001, 26, 687–689.
- [24] Panchaud, A., Singh, P., Shaffer, S.A., Goodlett, D.R., xComb: A Cross-Linked Peptide Database Approach to Protein–Protein Interaction Analysis. *J. Proteome Res.* 2010, 9, 2508–2515.
- [25] Lee, Y.J., Lackner, L.L., Nunnari, J.M., Phinney, B.S., Shotgun Cross-Linking Analysis for Studying Quaternary and Tertiary Protein Structures. *J. Proteome Res.* 2007, 6, 3908–3917.
- [26] Du, X., Chowdhury, S.M., Manes, N.P., Wu, S., et al., Xlink-Identifier: An Automated Data Analysis Platform for Confident Identifications of Chemically Cross-linked Peptides using Tandem Mass Spectrometry. *J. Proteome Res.* 2011, 10, 923–931.
- [27] Götze, M., Pettelkau, J., Schaks, S., Bosse, K., et al., StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* 2011, 23, 76–87.
- [28] Götze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., et al., Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. *J. Am. Soc. Mass Spectrom.* 2014, 26, 83–97.
- [29] Hoopmann, M.R., Weisbrod, C.R., Bruce, J.E., Improved Strategies for Rapid Identification of Chemically Cross-linked Peptides Using Protein Interaction Reporter Technology. *J. Proteome Res.* 2010, 9, 6323–6333.
- [30] Han, Y., Luo, J., Ranish, J., Hahn, S., Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J.* 2014, 33, 2534–2546.
- [31] Barrera, N.P., Bartolo, N.D., Booth, P.J., Robinson, C.V., Micelles Protect Membrane Complexes from Solution to Vacuum. *Science* 2008, 321, 243–246.
- [32] Barrera, N.P., Isaacson, S.C., Zhou, M., Bavro, V.N., et al., Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods* 2009, 6, 585–587.
- [33] Trimpin, S., Deinzer, M.L., Solvent-Free MALDI-MS for the Analysis of a Membrane Protein via the Mini Ball Mill Approach: Case Study of Bacteriorhodopsin. *Anal. Chem.* 2007, 79, 71–78.
- [34] Marty, M., Das, A., Sligar, S., Ultra-thin Layer MALDI Mass Spectrometry of Membrane Proteins in Nanodiscs. *Anal. Bioanal. Chem.* 2012, 402, 721–729.
- [35] Chen, F., Gerber, S., Heuser, K., Korkhov, V.M., et al., High-Mass Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry of Integral Membrane Proteins and Their Complexes. *Anal. Chem.* 2013, 85, 3483–3488.
- [36] Chen, F., Gerber, S., Korkhov, V.M., Mireku, S., et al., On the Efficiency of NHS Ester Cross-Linkers for Stabilizing Integral Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* 2014, 26, 493–498.
- [37] Chen, F., Gülbakan, B., Weidmann, S., Fagerer, S.R., et al., Applying mass spectrometry to study non-covalent biomolecule complexes. *Mass Spectrom. Rev.* 2015, n/a–n/a.
- [38] Bich, C., Maedler, S., Chiesa, K., DeGiacomo, F., et al., Reactivity and Applications of New Amine Reactive Cross-Linkers for Mass Spectrometric Detection of Protein–Protein Complexes. *Anal. Chem.* 2010, 82, 172–179.
- [39] Bich, C., Nazabal, A., Wenzel, R., Zenobi, R., Cross-Linking Reagents for Molecular Interactions Analysis. Patent N° WO2010136539, 2010.

Chapitre 4 : Suivi par ESI-MS de la formation des complexes de RAR-RXR avec différents ADNs

Nous avons abordé dans le chapitre 1 les informations apportées par la spectrométrie de masse supramoléculaire « classique » par ESI-MS dans l'étude des complexes biologiques non-covalents. Dans ce chapitre, nous expliquerons comment nous avons adapté cette approche pour étudier les complexes formés par RAR-RXR (récepteur de l'acide rétinoïque - récepteur X des rétinoïdes) avec différents ADNs. Tout d'abord, nous avons défini des conditions expérimentales permettant de détecter des espèces qui co-existent en solution et mis en lumière la relation entre leur abondance relative et le ratio molaire protéine/ADN. Ensuite, nous verrons comment le suivi de la dynamique d'assemblage et de désassemblage des différents états d'oligomérisation a été réalisé. Enfin, nous démontrerons l'influence de la séquence et de la longueur du spacer des ADNs sur la formation de l'hétérodimère RAR-RXR-ADN.

4.1. Contexte de l'étude : les complexes RAR-RXR-ADN

Dans le cadre d'une collaboration avec l'équipe du Dr. Natacha Rochel au Département de Biologie Structurale Intégrative de l'IGBMC, nous avons travaillé sur les complexes formés par RAR-RXR DBD (récepteur de l'acide rétinoïque - récepteur X des rétinoïdes, domaine de liaison à l'ADN) avec différents éléments de réponse. Faisant partie de la famille des récepteurs nucléaires, le complexe RAR-RXR agit comme un régulateur de la transcription qui active ou réprime les gènes [1]. Cette régulation de l'expression des gènes se passe par l'association de RAR-RXR aux ADN-éléments de réponse RAREs (« RAR response element ») [2] (Figure 1).

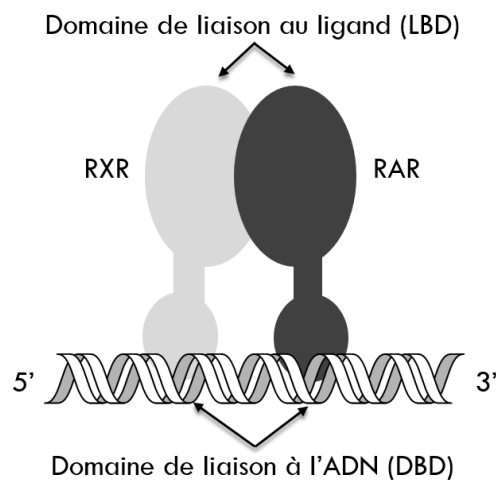


Figure 1. Représentation schématique de l'association de RAR-RXR à l'ADN.

Les RAREs contiennent une séquence hexamérique consensus 5'-RGKTCA-3' (avec R = A ou G, K = G ou T) pouvant être répétée de manière directe (DR - « direct repeat ») ou indirecte (IR - « inverted repeat ») comme représentée sur la Figure 2.

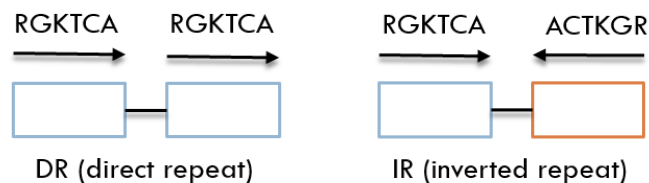


Figure 2. Représentation schématique des RAREs DR (« direct repeat ») et IR (« inverted repeat »).

Ces deux sites de fixation sont séparés par un nombre de paires de base variant de 0 (dans ce cas on parle par exemple de DR0) à 8 (DR8) [3] qui détermine la longueur du spacer de l'ADN (Figure 3). Les connaissances actuelles dans le domaine de la reconnaissance de l'ADN par RAR-RXR sont limitées aux complexes formés avec les éléments de type DR1 et DR5 [4-6].

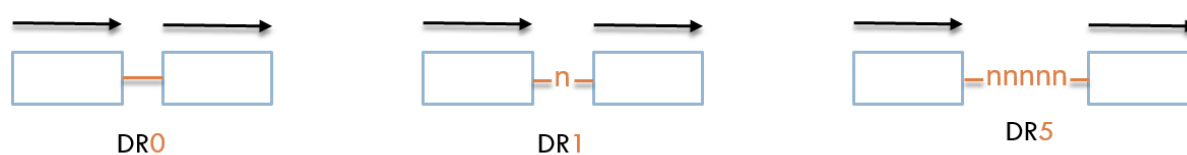


Figure 3. Représentation schématique de quelques exemples de longueurs du spacer de l'ADN.

Dans ce projet de recherche, nos collaborateurs souhaitent étudier la structure de la forme hétérodimérique (connue comme fonctionnelle) des complexes de RAR-RXR associés à différents éléments de réponse par la DRX. Cependant, les conditions de cristallisation déjà testées ne permettent pas d'obtenir des cristaux d'hétérodimère mais essentiellement des cristaux des formes homodimériques. Par conséquent, la question qui nous a été posée est de savoir quel élément de réponse induit l'hétérodimère avec RAR-RXR et dans quelles conditions expérimentales. Dans ce contexte, le choix de l'ESI-MS supramoléculaire s'est naturellement imposé grâce aux informations qu'elle permet d'obtenir. En effet, cette technique a fait ses preuves dans les études concernant l'existence, la stœchiométrie, ainsi que l'architecture des complexes protéiques [7]. En particulier, elle permet de suivre les différents états d'oligomérisation présents en solution, ainsi que les assemblages des partenaires en fonction du temps ou du milieu [8].

L'objectif de cette étude est donc d'explorer la composition et la dynamique d'interaction de RAR-RXR avec différents ADN. Dans ce sens, l'ESI-MS a été utilisée tout d'abord comme outil de screening des ADN afin de savoir lequel se lie et forme l'hétérodimère avec RAR-RXR. Ainsi, nous espérons que les informations obtenues par ESI-MS permettront d'affiner les conditions de cristallisation ou qu'elles pourront être incluses dans la modélisation moléculaire pour les études structurales de RAR-RXR-ADN.

4.2. Mise au point de l'étude des complexes RAR-RXR-ADN par ESI-MS

4.2.1. Difficultés de l'étude des complexes de type protéine-ADN

Bien que plusieurs études sur des complexes non-covalents par ESI-MS soient reportées, celles concernant des interactions de type protéine-ADN sont encore peu nombreuses [9–12]. Ceci pourrait s'expliquer par la grande polarité et la nature polyanionique de l'ADN. Les groupements phosphates tendent à se lier de manière non-spécifique aux cations alcalins présents dans le milieu [11] ; les adduits de sodium ou potassium alors formés persistent après l'ionisation et la désolvatation, ce qui a pour résultat l'élargissement du signal, rendant la mesure des masses et la détermination de la stœchiométrie plus difficiles [13]. Par conséquent, la préparation d'échantillon demande beaucoup d'attention.

En amont de l'analyse en ESI-MS, un échange de tampon est réalisé afin de transférer l'échantillon dans un solvant volatil compatible avec l'ionisation électrospray et capable de maintenir les interactions non-covalentes. Dans le cas de notre étude, nous avons utilisé un tampon d'acétate d'ammonium à 100 mM, pH 7,8. Les échanges ont donc été réalisés afin d'avoir des solutions mères de protéines et d'ADN dans ce tampon. Au début, ces échanges se sont avérés délicats à maîtriser du fait d'adduits restants qui élargissaient le signal sur le spectre de masse comme illustré en Figure 4. Les spectres de RAR seul et RXR seul (Figure 4A et Figure 4B) sont bien résolus ; néanmoins, le spectre des deux protéines incubées avec 1 équivalent d'ADN (Figure 4C) présente beaucoup d'adduits élargissant le signal et rendant la mesure des masses et la détermination de la stœchiométrie impossibles.

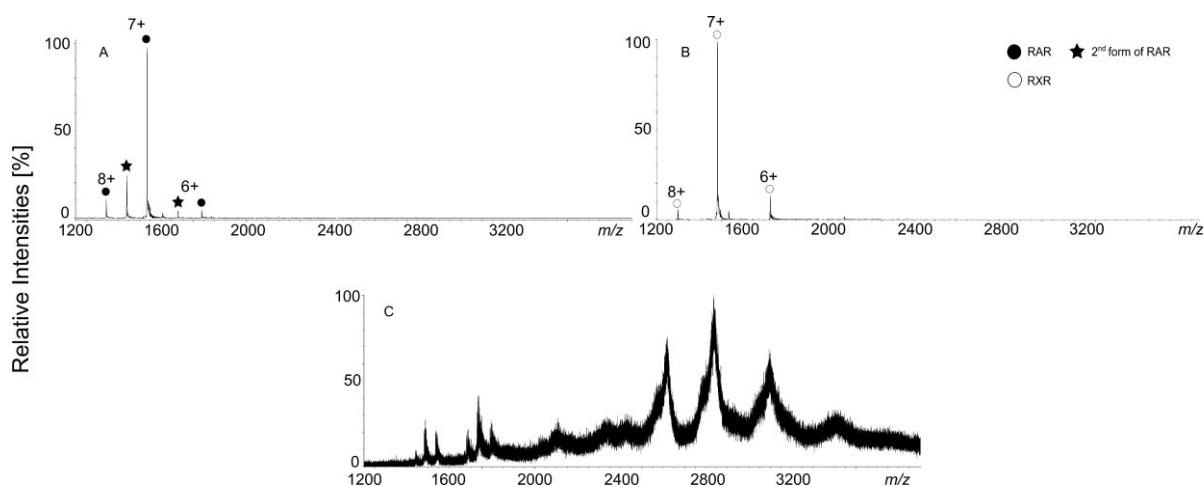


Figure 4. Spectres de masse de (A) RAR seul à 9 μ M, (B) RXR seul à 9 μ M, (C) RAR-RXR incubés avec 1 équivalent de RAR β 2 DR5.

Après optimisations, l'utilisation des colonnes de filtration sur gel Micro Bio-Spin (Biorad) avec deux passages successifs pour l'ADN seul a été concluante dans le cas de notre étude : elle nous a permis de nous affranchir de ces adduits, ce qui a été indispensable pour une détermination précise des masses des espèces détectées (Figure 5).

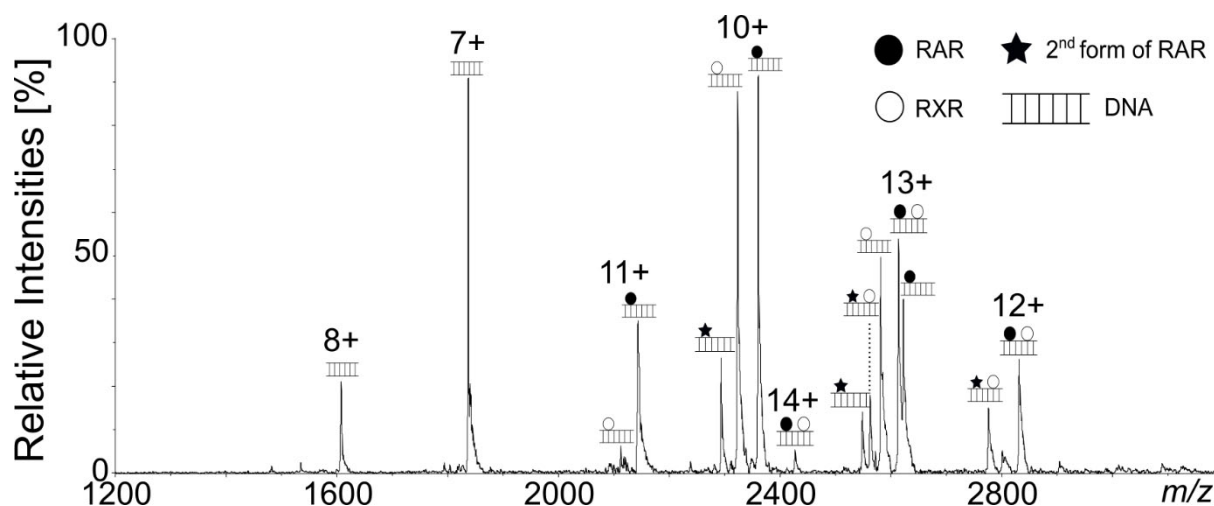


Figure 5. Spectres de masse des récepteurs rétinoïques RAR-RXR incubés avec 1,2 équivalent de *RARβ2* DR5 avec deux échanges de tampon successifs pour l'ADN seul.

D'autre part, la concentration d'échantillon lors de l'analyse doit être suffisante pour favoriser la formation des dimères de RAR-RXR et permettre de détecter le signal avec une bonne intensité, mais ne doit pas être trop importante pour éviter des associations non-spécifiques. Afin d'être non ambigu dans des cas de stœchiométries très hétérogènes, la concentration de protéine a été revue à la hausse par rapport à nos conditions initiales et a été fixée à 32 μ M.

4.2.2. Optimisation des paramètres instrumentaux

Dans ce projet, les analyses MS ont été réalisées en mode positif sur un spectromètre de masse MicrOTOF (BrukerDaltonics). Les paramètres instrumentaux ont été optimisés afin de maintenir les complexes protéine-ADN intacts lors du passage en phase gazeuse et lors de la transmission dans l'interface du spectromètre de masse.

Tout d'abord, il est important d'avoir la pression du gaz de nébulisation et le débit du gaz de séchage adéquats, afin de stabiliser le spray et d'assurer une bonne désolvatation des ions. Dans le cas de notre étude, la pression du gaz de nébulisation a été fixée à 0,6 bar et le débit du gaz de séchage à 6 L/min.

D'autre part, dans l'interface du spectromètre de masse où les ions sont transférés de la source à l'analyseur, l'énergie interne des ions peut être suffisamment importante pour dissocier le complexe [14]. La tension d'accélération (Capillary Exit) est un paramètre particulièrement déterminant : cette différence de potentiel entre la sortie du capillaire en verre et le premier skimmer du micrOTOF induit un champ électrique dans l'interface accélérant les ions. Cette énergie cinétique acquise va accroître l'énergie interne des ions seuls ou encore solvatés. Ainsi, l'augmentation de cette tension permet d'améliorer leur transfert jusqu'à l'analyseur et la désolvatation des ions, mais elle peut induire la dissociation des interactions non-covalentes et

donc la dissociation des complexes. Pour notre étude, nous avons fait varier cette tension de 100 V à 200 V, le meilleur signal étant obtenu pour une tension de 150 V.

Nous avons également adapté des paramètres de stockage (« pre-pulse storage time ») et de transmission (« transfer time ») pour mieux détecter des complexes protéine-ADN qui ont des rapports m/z plus élevés que les protéines libres. Avec un temps de transmission plus faible, on favorise la détection des formes de protéines libres, tandis qu'avec un temps de transmission plus élevé, on observe préférentiellement des formes complexées. La Figure 6 illustre cet effet discriminant auquel il faut faire attention pour pouvoir relier les signaux détectés sur les spectres de masse à l'abondance relative des espèces présentes. C'est un compromis entre une bonne détection des ions lourds et une meilleure représentativité des espèces présentes en solution. Dans le cas de notre étude, le « pre-pulse » a été fixé à 40 μs et le « transfer time » à 100 μs .

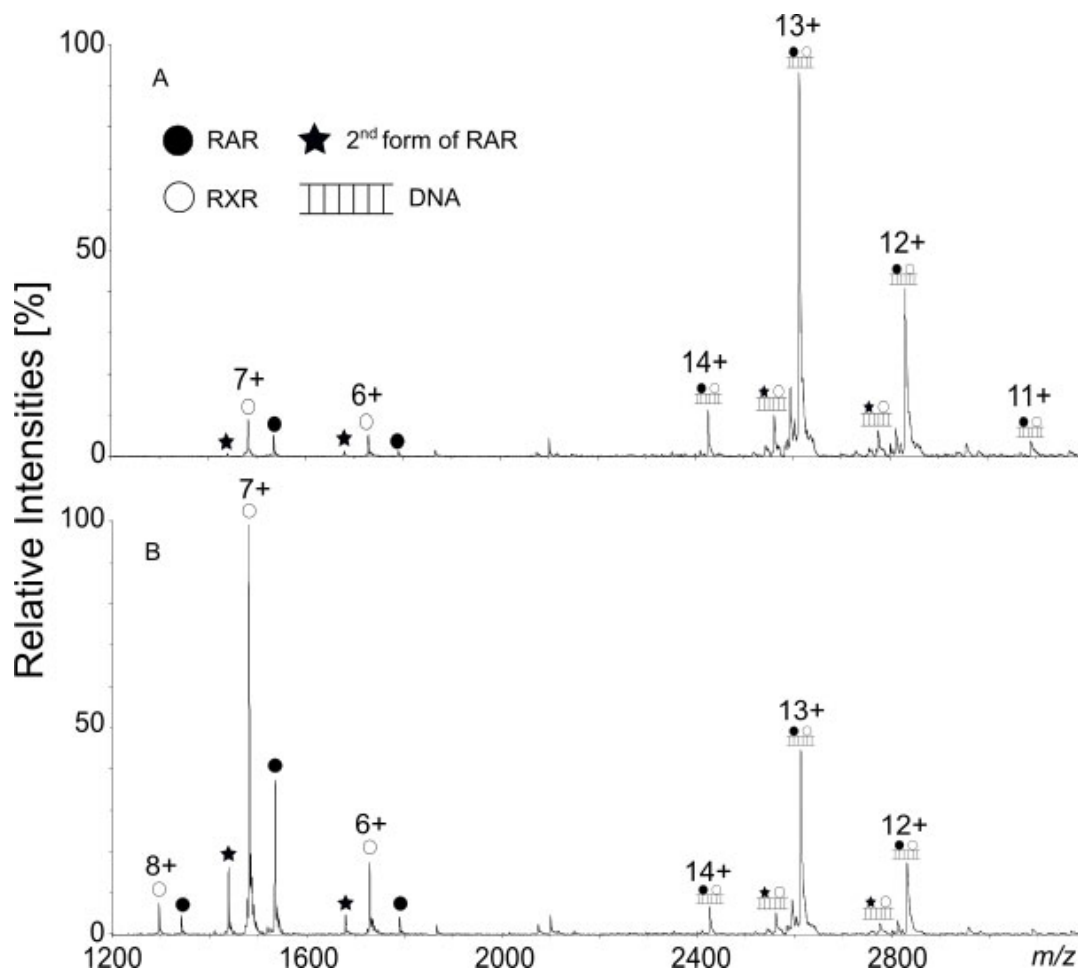


Figure 6. Spectres de masse des récepteurs rétinoïques RAR-RXR incubés avec 0,25 équivalent de RAR β 2 DR5. (A) Pre-pulse à 40 μs , (B) Pre-pulse à 30 μs .

4.3. Stratégie de l'étude des complexes RAR-RXR-ADN

Tout d'abord, nous avons optimisé les conditions expérimentales pour la détection des hétérodimères RAR-RXR-ADN. Au début du projet, nous avons pensé qu'un excès d'ADN favoriserait la formation des formes dimériques et donc de l'hétérodimère RAR-RXR lié à l'ADN. De manière surprenante, ce ne fut pas du tout le cas. Par conséquent, nous avons voulu suivre l'évolution des espèces détectées représentées en Figure 7 en fonction de la quantité d'ADN ajoutée. Ainsi, il s'est avéré que les abondances relatives des formes oligomériques dépendent spécifiquement du ratio molaire protéine/ADN.

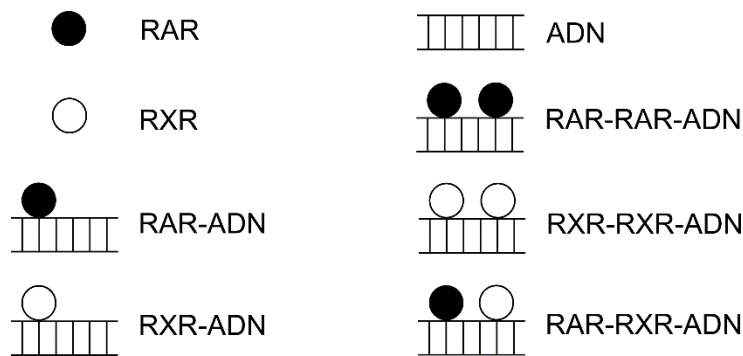


Figure 7. Représentation schématique des espèces détectées en ESI-MS dans l'étude de RAR-RXR-ADN.

Ensuite, nous avons cherché à savoir si les deux sites de fixation de l'ADN peuvent accueillir les deux protéines indépendamment de leur nature. Pour cela, des mélanges de RAR seul avec l'ADN (sans RXR) et de RXR seul avec l'ADN (sans RAR) ont été analysés. Il en résulte que pour les ADN testés et en cas de présence d'une seule protéine, les deux sites peuvent être occupés entièrement par RAR ou RXR pour former le monomère et l'homodimère correspondants.

Face à ces résultats, nous nous sommes demandé si une fois l'homodimère constitué, l'hétérodimère peut encore être formé. Pour cela, nous avons réalisé les expériences d'ordre d'ajout : RAR est d'abord incubé avec l'ADN avant l'ajout de RXR ou à l'inverse, nous incubons d'abord RXR avec l'ADN avant l'ajout de RAR. Les résultats ont montré qu'il y a un déplacement des homodimères pour former l'hétérodimère. Ceci a confirmé que l'occupation des sites de fixation est un processus dynamique [15], et que grâce à la coopérativité de l'hétérodimérisation et la stabilité des interactions protéiques avec l'ADN, la formation de l'hétérodimère est favorisée comparé à celle de l'homodimère [16].

Ensuite, nous avons voulu étudier l'influence de la séquence et de la longueur du spacer entre les deux sites sur la répartition des formes oligomériques. D'abord, nous avons analysé les mélanges de RAR-RXR avec différents ADNs possédant la même séquence (*RARβ2*) mais se distinguant par le nombre de nucléotides entre deux sites « direct repeat » (DR0, DR1, DR2, DR5). Il en résulte que pour *RARβ2*, plus la longueur du spacer est importante, plus l'hétérodimère est abondant. Ensuite, les expériences ont été réalisées avec les RAREs dont les séquences sont

différentes. Les résultats obtenus ont suggéré que la séquence de l'ADN joue également un rôle indéniable dans la répartition des formes complexées formées.

En résumé, le protocole expérimental développé en ESI-MS au cours de ce projet de recherche a permis de détecter les différentes espèces présentes en solution et de suivre l'assemblage et le désassemblage des oligomères formés. De plus, il a permis de suivre la dynamique d'interaction en fonction de l'ordre d'ajout, et de mettre en lumière l'influence du spacer et de la séquence de l'ADN dans la formation des complexes de RAR-RXR-ADN.

Les résultats de cette étude ont fait l'objet d'une publication dans le journal scientifique à comité de lecture *Biophysical Chemistry* (Elsevier, 2015).

4.4. Publication : Monitoring of the retinoic acid receptor-retinoid X receptor dimerization upon DNA binding by native mass spectrometry

Biophysical Chemistry **2015**. DOI: 10.1016/j.bpc.2015.10.006

ARTICLE IN PRESS

BIOCHE-05867; No of Pages 7

Biophysical Chemistry xxx (2015) xxx–xxx



Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: <http://www.elsevier.com/locate/biophyschem>

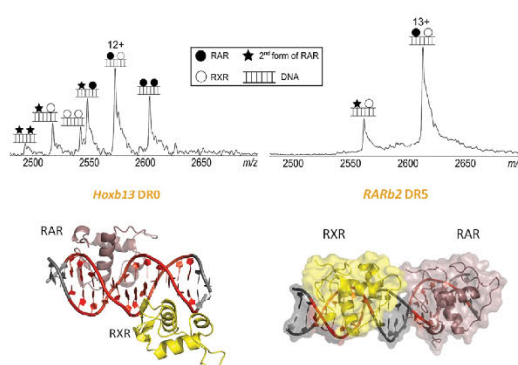
Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry

Nha-Thi Nguyen-Huynh^a, Judit Osz^b, Carole Peluso-Iltis^b, Natacha Rochel^b,
Noëlle Potier^{a,*}, Emmanuelle Leize-Wagner^a^a Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 CNRS/Université de Strasbourg - "Chimie de la Matière Complexe", 1 Rue Blaise Pascal, 67008 Strasbourg, France^b Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964, Centre National de Recherche Scientifique (CNRS) UMR 7104, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France

HIGHLIGHTS

- The composition of RAR–RXR complexes varies with direct repeat sequences.
- The heterodimerization of RAR–RXR bound to DR0 elements is non-cooperative.
- Protein/DNA molar ratios determine the relative abundance of RAR–RXR–DNA complexes.
- The dimerization of RAR–RXR upon DNA binding was monitored by native ESI-MS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 13 July 2015
 Received in revised form 9 October 2015
 Accepted 27 October 2015
 Available online xxx

Keywords:

Native mass spectrometry
 Dynamics study
 Protein–DNA interaction
 Nuclear receptors
 RAR–RXR complexes
 Retinoic acid receptor (RAR)
 Retinoid X receptor (RXR)

ABSTRACT

Identifying protein–DNA interactions is essential to understand the regulatory networks of cells and their influence on gene expression. In this study, we use native electrospray mass spectrometry (ESI-MS) to investigate how the heterodimerization of retinoic acid receptor–retinoid X receptor (RAR–RXR) is mediated by DNA sequence. In presence of various RAR response elements (RAREs), three oligomeric states of RAR–RXR DNA binding domains (DBDs) bound to RAREs (monomer, homo- or heterodimers) were detected and individually monitored to follow subunit assembly and disassembly upon RAREs' abundance or sequence. In particular, a cooperative heterodimerization was shown with *RARb2* DR5 (5 base pair spaced direct repeat) while a high heterogeneity reflecting random complex formation could be observed with the DR0 response elements, in agreement with native gel electrophoresis data or molecular modeling. Such MS information will help to identify the composition of species formed in solution and to define which DR sequence is specific for RAR–RXR heterodimerization.

© 2015 Published by Elsevier B.V.

Abbreviations: NR, nuclear receptor; RAR, retinoic acid receptor; RXR, retinoid X receptor; DNA, deoxyribonucleic acid; RA, all-trans retinoic acid (RA); DR, direct repeat; RARE, RAR response element; ESI-MS, electrospray ionization mass spectrometry; DBD, DNA binding domain; LBD, ligand binding domain; Kd, dissociation constant; ITC, isothermal titration calorimetry; MST, microscale thermophoresis.

* Corresponding author.

E-mail address: npotier@unistra.fr (N. Potier).<http://dx.doi.org/10.1016/j.bpc.2015.10.006>

0301-4622/© 2015 Published by Elsevier B.V.

Please cite this article as: N.-T. Nguyen-Huynh, et al., Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry, *Biophys. Chem.* (2015), <http://dx.doi.org/10.1016/j.bpc.2015.10.006>

1. Introduction

Proteins, such as many transcription factors that bind to specific DNA sequences, are essential for the proper regulation of gene expression. Identifying the specific sequences that each factor binds can help to elucidate regulatory networks within cells and how genetic variation can cause disruption of normal gene expression, which is often associated with diseases. Various techniques have been reported for the study of protein–DNA interactions such as chromatin immunoprecipitation (ChIP) [1], electrophoretic gel mobility shift assay (EMSA) [2], DNA pull-down assay [3], microplate capture and detection assay [4], reporter assay [5] or isothermal calorimetry, but several new high-throughput methods can provide comprehensive binding information much more rapidly. Combined with *in vivo* determinations of transcription factor binding locations, this information provides more detailed views of the regulatory circuitry of cells and the effects of variation on gene expression.

All-trans retinoic acid (RA), the naturally active vitamin A metabolite, exerts a wide range of effects on vertebrate development and plays a critical role in the homeostasis and pathophysiology of adult tissues [6,7]. RA exerts most of its pleiotropic effects through interaction with three members of the nuclear receptor (NR) superfamily, the all-trans retinoic acid receptors (RAR) α , β , γ that heterodimerize with the retinoid X receptors (RXR) α , β , γ . Heterodimer RAR–RXR acts as RA-dependent transcriptional regulators that either activate or repress target genes [8]. The RAR–RXR heterodimer regulates gene expression by binding to DNA response elements or RAREs, the best characterized of which comprise direct repeats of the consensus 5'-RGKTC A-3' (where R = A or G, and K = G or T) separated by 1, 2 or 5 nucleotides (DR1, DR2 or DR5) [9]. Genome wide analysis of RAR binding sites in mouse embryo fibroblasts (MEFs), embryonic stem (ES) and F9 embryonal carcinoma cells have revealed a variety of RARE motifs consisting of the classical DR1, DR2 and DR5 and also of direct repetition of the core motif without spacer (DR0), direct repeat of the core motif with 8 nucleotides space (DR8) or inverted repeat without spacer (IRO) elements [10]. In undifferentiated cells, high prevalence of the non-canonical DR0-containing RAREs are observed while in differentiated cells, the canonical DR5 motif is more prevalent [10]. *In vitro* EMSA analysis indicated that flanking and spacer sequence both influence RAR–RXR binding. Of particular interest are DR0 elements since RAR–RXR bound to DR0s are not able to modulate gene expression in transcriptional reporter assay [10]. The molecular basis of RAR–DNA recognition of non-canonical elements and their transcriptional specificities remain unknown. Current atomic level understanding of DNA recognition by RAR–RXR and RXR homodimer is limited to complexes with DR1 elements [11,12] and low resolution solution structures of RAR–RXR on the classical DR1 and DR5 elements [13]. A complementary method that would help characterizing the heterogeneity and composition of the nuclear receptor complexes upon DRs sequence and spacer length would be of great interest.

While traditional methods for determining the specificity of DNA-binding proteins are slow and laborious, mass spectrometry (MS) – due to their high sensitivity, selectivity and speed [14] – has emerged as a complementary method for characterizing intact assemblies [15]. Through improvements in instrumental technology and methodology, nondenaturing or native MS is now emerging as a new approach for getting insights into the existence, the stoichiometry or the architecture of protein complexes [16]. In particular, MS is a proven technique for following the different oligomeric states present in solution. This allows assessment of the study of protein complex dynamics by monitoring protein subunits assembly or disassembly over time or solution composition [17]. An increasing number of studies were reported since the last decade in the literature [18–20], including very large multiprotein complexes such as intact ribosomes [21] or proteins in interaction with a wide variety of partners including protein subunits, drug molecules, lipids, oligonucleotides. However, examples of protein–DNA interaction

studies by MS are under-represented in the literature [22–25]. The major problem encountered with the analysis of these complexes may lie in the very high polarity and polyanionic nature of the DNA. Due to its polyphosphate backbone, the oligonucleotides have a tendency to catch alkaline cations present in the medium, leading to formation of oligonucleotide–sodium or –potassium adducts that persist after ionization and desolvation. Quite broad *m/z* signals, if even a signal, are then observed preventing precise mass measurement and stoichiometry unambiguous assignment [26]. Moreover, particular care should be taken in the interpretation since oligonucleotides may undergo nonspecific binding to basic proteins [27].

In this paper, we describe how ESI-MS has been successfully used to monitor the composition of RAR–RXR DNA binding domains (DBD) complexes with various RAREs. In particular, classical DR5 were compared to non-canonical DR0 elements. This method proved to be very efficient for checking the stabilization of heterodimers on RARE sequences. We discovered a correlation between the protein/DNA molar ratios and the dynamics of assembly of binding complexes. To understand how non-canonical DR0 elements are recognized by RAR–RXR and to decipher the relationship between the RAR binding element type and the transcriptional response, molecular model building was used in complement to mass spectrometry characterization.

2. Materials and methods

2.1. Constructs, expression and purification

The HsRXR α -DBD (130–212) and HsRAR α -DBD (82–167) were expressed in fusion with thioredoxine and hexahistidine tags. Purifications of the RXR and RAR DBDs were performed as previously described [12] and included an affinity chromatography, fusion tag removal by thrombin proteolysis and gel filtration. The HsRAR α DNA binding domain–ligand binding domain DBD-LBD (82–462), and HsRXR α DBD-LBD (130–462) were expressed and purified as described elsewhere [13]. The oligonucleotide strands were purchased from SIGMA and annealed.

2.2. TBE native PAGE

The proteins were run on a continuous 6% polyacrylamide gel with constant current of 6 mA for 2 h at 4 °C in the TBE (Tris-Borate-EDTA) migration buffer. The unloaded gel was pre-run before the electrophoresis. The TBE buffer contained 0.089 M Tris base, 0.089 M boric acid (pH 8.3) and 2 mM Na₂EDTA. Approximately 5 μ g of protein was loaded per each lane.

2.3. Microscale thermophoresis

Microscale thermophoresis (MST) measurements were performed with a Monolith NT.115 instrument (NanoTemper Technologies GmbH, Munchen, Germany). The purified RAR α /RXR α DBDs-LBDs heterodimer was labeled with NT-647-NHS fluorescent dye following the manufacture protocol (Nanotemper Technologies). After the labeling, the excess of the free dye was eliminated by Illustra NAP-5 desalting column (GE Healthcare) and at the same time, the buffer was changed for either 100 mM ammonium acetate pH 7.8 or 50 mM HEPES pH 8.0, 100 mM NaCl, 2% glycerol. Each measurement consists of 16 reaction mixtures where the fluorescent-labeled RAR α /RXR α DBDs-LBDs concentration was constant and RAR β 2 DR5 concentrations were in serial dilution (1:1) through the 16 reactions. Depending on the measurements, the RAR α /RXR α DBDs-LBDs concentrations varied between 37 and 94 nM and the RAR β 2 DR5 serial dilution between 2 nM–55 μ M.

The measurements were performed using premium type glass capillaries (Nanotemper) at 40–100% LED excitation and 40–80% MST power, with a laser-on time of 30 s and a laser-off time of 5 s.

NanoTemper Analysis 1.4.27 software was used to fit the data and to determine the K_d .

2.4. MS analysis

Before analysis, the RXR DBD, RAR DBD and DNAs were subjected to buffer exchange to 100 mM ammonium acetate pH 7.8 at 4 °C (Micro Bio-Spin chromatography column, exclusion limit 6 kDa). Particular care was taken to the desalt DNAs separately twice. MS data were acquired in the positive ion mode on an electrospray time-of-flight mass spectrometer (MicroTOF, BrukerDaltonic, Germany). Denaturing conditions measurements were conducted by diluting the proteins to 10 μ M in H₂O/ACN (50/50) acidified with 1% formic acid to verify their purity and sequence integrity. Mass spectrometric analyses of the noncovalent complexes were carried out in 100 mM ammonium acetate pH 7.8. Mixture of proteins at 32 μ M and DNA at various equivalents (0.1, 0.25, 0.5, 1.2 equivalents) were infused at 4 μ L/min through a KDS100 syringe pump (KD Scientific, Holliston, MA, USA). The source and transfer parameters of the instrument were optimized in order to obtain the best spectrum without dissociating the complexes. Especially, the capillary exit voltage, which accelerates the ions in the middle pressure interface, was increased from 100 V to 200 V. The pre-pulse time, which represents a delay between transfer time and TOF pulser on for injection of ions into the TOF analyzer, was increased from 30 μ s to 40 μ s to optimize the transmission of high m/z ions (for more details see Results and Discussion part). Heterodimers were best obtained at the capillary exit of 150 V and the pre-pulse time at 40 μ s.

3. Results and discussion

3.1. MS method to follow the association dynamics of RAR–RXR–DNA complexes

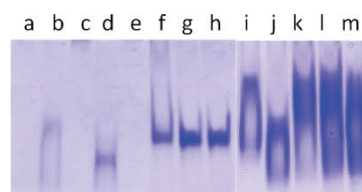
To clarify our understanding of the specificity of the RAR and RXR binding to DR0s, we compared the formation of RAR–RXR complexes by various techniques on 2 types of DR0s – the DR0 from *Msi* gene and the pseudo-DR0 motif from *Hoxb13* gene – to those formed on DR5s from *RARb2* and *F11r* genes (Fig. 1A). All experiments were performed using the DNA binding domains of RAR and RXR. By isothermal titration calorimetry (ITC), we previously have determined the DNA binding affinities of RAR–RXR DBD–LBDs constructs [10] and have shown that RAR–RXR have affinities values for DR0s in the same range as for DR5s (Supplementary Table 1). As MS experiments are performed in ammonium acetate buffer and ITC could not be performed in this buffer, we use another method, the microscale thermophoresis (MST) [28]. Using a titration approach, MST allowed us to determine the affinity constants of interactions in the binding equilibrium. In MST experiments, one of the binding partners is fluorescently labeled at a fixed concentration (a few nM), while the concentration of the unlabeled partner is varied from a high concentration, much above the expected dissociation constant (K_d), down to substoichiometric concentrations with respect to the labeled protein. Unlabeled *RARb2* DR5 was thus titrated into a fixed concentration of labeled RAR–RXR (Supplementary Fig. 1). Remarkably, the apparent K_d values calculated from MST experiments compare very well with K_d values obtained independently by ITC and are in the same range in the 2 buffers (Supplementary Table 1). The formation of the RAR–RXR DBD–DNA complexes was also followed by TBE native PAGE electrophoretic mobility (Fig. 1B). It shows that for DR5, only heterodimers are observed, while for *Hoxb13* DR0, we suspect that a mixture of species is present in the smeared bands. Similar results were observed for other DR0s tested. Mass spectrometry was then chosen to identify which complexes were formed under which conditions.

Denaturing MS was first performed to check the protein homogeneity and integrity. Mass spectra revealed that the RAR DBD was present under 2 forms: a major form at 10 607.1 Da and a minor one at 9

A

Hoxb13 pseudoDR0 GAAGGTCAAGGCCAAG
CTTCCAGTTCGGTTC
Msi DR0 AGGGGTCAAGGTCAA
TCCCAGTTCAGTTT
F11r DR5 AGGGGTCAagccaGGTTCACT
TCCCAGTtcggTCCAAGTGA
RARb2 DR5 AGGGTCAccgaaAGTTCACT
TCCAAGTggcttTCAAGTGA
RARb2-DR2 AGGGTCAagAGTTCACT
TCCAAGTtcTCAAGTGA
RARb2-DR1 AGGGTCAaAGTTCACT
TCCAAGTtTCAAGTGA
RARb2-DR0 AGGGTCAAGTTCACT
TCCAAGTTCAGTGA
RARb2-DR0rev AGGGTCAAGTTCACT
TCCAAGTTCAGTGA
Ramp2 DR1 TcAGTTCAaGGTCAAT
AcTCAAGTtCCCAGTTA

B



a RARa
b RARa-RARb2 DR5
c RXRa
d RXRa-RARb2 DR5
e RARa-RXRa
f RARa-RXRa-0.5eq RARb2 DR5
g RARa-RXRa-0.75eq RARb2 DR5
h RARa-RXRa-1eq RARb2 DR5
i RARa-1eq *Hoxb13* DR0
j RXRa-1eq *Hoxb13* DR0
k RARa-RXRa-0.5eq *Hoxb13* DR0
l RARa-RXRa-0.75eq *Hoxb13* DR0
m RARa-RXRa-1eq *Hoxb13* DR0

Fig. 1. (A) Sequences of the DNA response elements used in this study with the first hexanucleotide half-site shown in red and the second hexanucleotide half-site in red and underlined. The interspacer nucleotides are shown in black and lowercase; (B) TBE native PAGE gel of RAR and RXR DBDs. The free proteins do not migrate in the conditions of the gel. For DR0 REs the smeared bands and the presence of several bands indicate a mixture of species that cannot be resolved.

940.8 Da \pm 0.5 Da which may correspond to the loss of the remaining residues of the tag GSHM plus two amino acids (proline and arginine). RXR DBD was detected at 10 236.8 Da \pm 0.5 Da, which is in good agreement with the theoretical molecular mass. For detection of non-covalent complexes, solution as well as instrument conditions were carefully optimized (see Materials and Methods). In particular, the interface voltages implied in collision activation and ion transmission were adjusted using a well-characterized heterodimer complex formed between RAR–RXR and *RARb2* DR5. For this, an increasing ratio of *RARb2* DR5 (0.1, 0.25, 0.5, 1.2 equivalents) was added to a RAR–RXR mixture and the MS spectra recorded for each equivalent. Fig. 2 shows that different oligomeric states were observed upon solution conditions. At 0.1 and 0.25 equivalents of DNA, unbound RAR and RXR were observed in concomitance with a third species at 33 954.6 Da \pm 0.4 Da corresponding to a RAR–RXR–DR5 heterocomplex bound to DNA. Although no zinc was added to our ammonium acetate buffer, 2 Zn²⁺ remain bound to the free or bound proteins, which indicates that at least part of the protein conformation has been retained in vacuo during analysis. No species revealing the existence of RAR–RAR–DR5 or RXR–RXR–DR5 homodimer complexes were furthermore

4

N.-T. Nguyen-Huynh et al. / Biophysical Chemistry xxx (2015) xxx–xxx

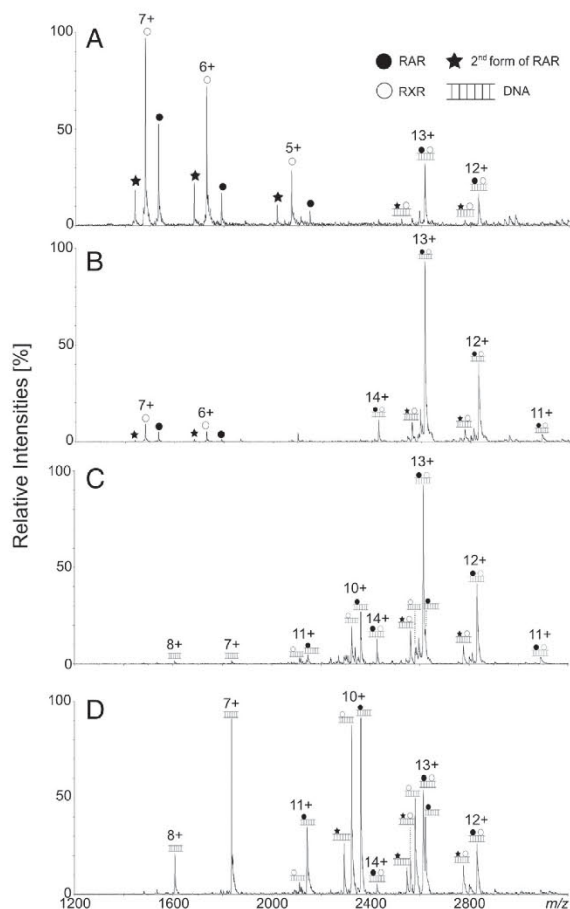


Fig. 2. Mass spectra of retinoic receptors RAR–RXR bound to DNA depending on different equivalents of DNA. (A) 0.1 equivalent of RAR/2 DR5; (B) 0.25 equivalent of RAR/2 DR5; (C) 0.5 equivalent of RAR/2 DR5; (D) 1.2 equivalent of RAR/2 DR5.

observed. As shown previously [29], MS spectra revealed that the heterodimeric interface is greatly favored in presence of DNA over homodimeric ones. In particular, heterodimerization of RAR–RXR DBD–LBD bound to DR5 elements was precedently shown by ESI-MS [30], but in the present work we used only DBD to avoid any bias by the LBD dimerization interface. These results strongly support structurally-specific interactions and revealed the different populations of complexes present in solution rather than unspecific aggregation in the gas phase. As the relative protein abundance decreased (0.5 and 1.2 eq. of DNA), RAR–DR5 and RXR–DR5 monomeric complexes emerged in the spectra while signal due to free proteins became negligible. Because of instrument effects, the relative abundance of free protein versus protein–DNA complexes could not be extracted nor evaluated from MS spectra (Supplementary Fig. 2). Indeed, both species display different m/z ratios and different conformations, two parameters that are known to affect the ion intensities. Light or heavy m/z ions will then have different trajectories inside the MS interface and thus different transmission rates. Transmission rates were optimized by modifying the interface voltage settings, in particular, the pre-pulse voltage and the transfer time (these 2 settings are responsible for the storage of the ions and their transmission inside the mass spectrometer). This might also explain the

apparently low intensity of free proteins at 0.1–0.5 equivalents of DNA. However, the spectra along the DNA titration could be compared as they were acquired under the same interface parameters.

To further investigate whether both DNA binding sites were able to bind to retinoic receptors independently of their nature (RAR or RXR), experiments consisting of RAR-only or RXR-only incubation with the RAR/2 DR5 were carried out. In each case, the spectra reveal that both DNA binding sites could be occupied by RAR or RXR (Fig. 3). Molecular masses of 34 319.0 Da and 33 584.7 Da \pm 0.3 Da corresponding to RAR–RAR–DR5 and RXR–RXR–DR5 homodimers complexes respectively were measured. The complexes formed by DNA with the major and minor forms of RAR show that the N-terminal degradation of the protein did not seem to affect the DNA recognition. From these data, it is not possible to determine whether both sites are bound with similar affinities, but this is beyond the scope of this study.

3.2. Impact of RAREs on heterodimerization

To determine whether the preformed homodimer–DNA complex could prevent the formation of the heterodimer–DNA complex, experiments consisting of addition of RAR and RXR to DNA sequentially were carried out. In particular, RAR was pre-incubated with the DR5 before addition of RXR (Fig. 4A and B) or in the opposite way, RXR–DR5 complex was pre-formed before addition of RAR (Fig. 4C and D). The spectra clearly show that all homodimer–DNA species are totally displaced in favor of heterodimer–DNA complex. Indeed, the occupancy of the sites is a dynamic process as already shown [31]. These results confirm that, thanks to the cooperativity of heterodimerization and the stabilization of subunits interactions by DNA, the formation of heterodimer–DNA is favored compared to homodimer–DNA species, as shown for classical DR5 and DR2 elements [32]. The same experiments were performed with *Hoxb13* DR0 yielding the same conclusion, i.e. the species present at equilibrium are identical independent of the order of partner addition (Supplementary Fig. 3). Because each species present in solution displays a different molecular mass, all oligomeric states could be followed individually. MS appeared in our case to be a powerful technique for monitoring the disappearance of homodimer–DNA forms and the concomitant formation of the corresponding heterodimer–DNA species. It is well suited to follow all fluctuations in subunit assembly or disassembly, and thus possesses a great potential to assess to binding dynamics or even to the assembly pathway if a hierarchy is required [17].

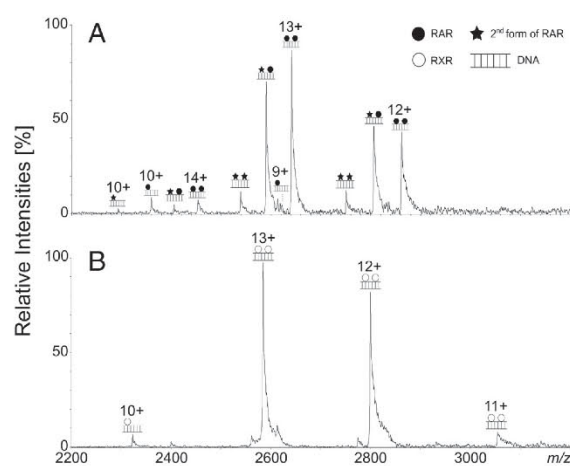


Fig. 3. Mass spectra of retinoic receptors bound to DNA. (A) RAR incubated with RAR/2 DR5; (B) RXR incubated with RAR/2 DR5.

ARTICLE IN PRESS

N.-T. Nguyen-Huynh et al. / Biophysical Chemistry xxx (2015) xxx–xxx

5

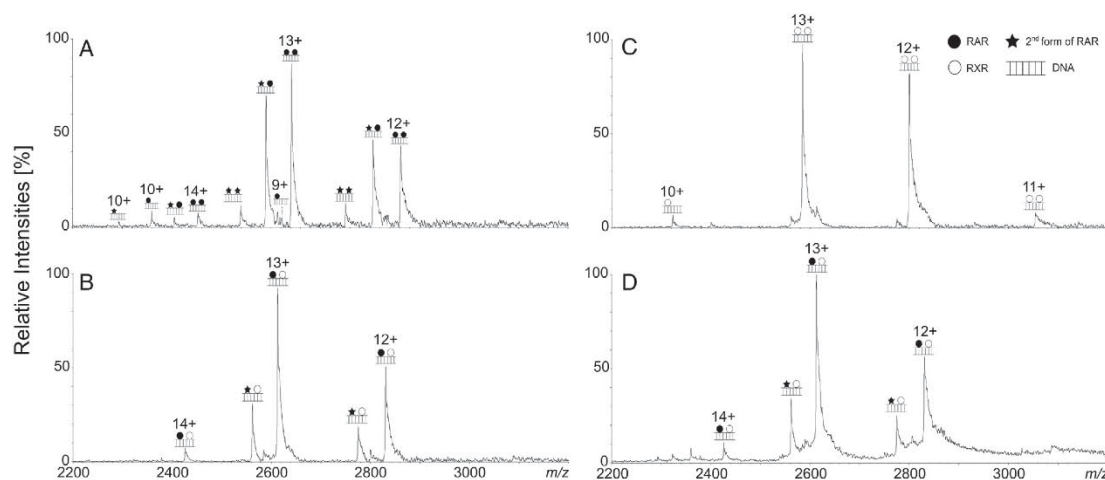


Fig. 4. Mass spectra of retinoic receptors bound to DNA depending on order of incubation. (A) RAR incubated with $RAR/\beta 2$ DR5; (B) RAR incubated with $RAR/\beta 2$ DR5 followed by an addition of RXR; (C) RXR incubated with $RAR/\beta 2$ DR5; (D) RXR incubated with $RAR/\beta 2$ DR5 followed by an addition of RAR.

In order to find out whether the length of the spacer between the two DNA binding sites was a determining factor, formation of the RAR–RXR–DNA complexes was followed after incubation of the proteins with various RAREs displaying the same sequence of the two half-sites ($RAR/\beta 2$) but differing by the number of nucleotides between the two direct repeat sites (DR0, DR1, DR2, DR5). All experiments were performed under the exact same conditions in terms of relative abundance in solution and mass spectrometer interface parameters. Mass spectra revealed the formation of different oligomeric states according to the considered DNA (Fig. 5). When DR0 was added to the solution, the spectrum showed a high level of heterogeneity and many stoichiometries could be measured. Monomer-, homodimer- and heterodimer–DNA complexes coexist without an evident favored stoichiometry, which would have been an indication of specificity in the binding interfaces. On the contrary, the relative abundance of the RAR–RAR and RXR–RXR homodimer–DR0 complexes versus the RAR–RXR heterodimer–DR0 (about 40% and 50% respectively) rather suggests random filling of the DNA binding sites in agreement with the observed species in the native gel (Fig. 1). Similar results were observed with another DR0 RE (*Msi*, Supplementary Fig. 4).

When DR1 is added, no signal corresponding to either monomer–DNA or to RAR homodimer–DNA complexes were observed. Only RXR homodimer- and RAR–RXR heterodimer–DR1 were detected. This result can be explained by the fact that for DR1, homodimer of RXR–DR1 is formed preferentially to the detriment of RAR–DR1 [33]. In presence of DR2 and DR5, a minor signal corresponding to RAR and RXR homodimer–DNA is observed while the main species corresponds to the heterodimer–DNA complex. From these results, an observation immediately emerges: while the heterodimeric form is always present, for $RAR/\beta 2$, the longer is the spacer, the more abundant is the formation of the corresponding RAR–RXR–DNA complex. As a consequence, the heterodimeric interface appears to be stabilized with a 5 base pair spaced DNA. Model building of the RAR–RXR heterodimer in complex with DR5 or DR0 (Fig. 6) indicates that for RAR–RXR in complex with a DR5 element, interactions between the 2 monomers will be formed, indicative of a cooperative dimer, as already modeled [34]. In contrast, in the complex with DR0 element, no interactions between the 2 monomers are formed; this is consistent with the binding to DR0 element as non-cooperative dimer providing an explanation of MS results. This non-cooperative dimer of RAR–RXR–DR0 may explain the inability of RAR–RXR bound to DR0 to activate gene expression in transcriptional reporter assay, as shown for glucocorticoid receptors with negative response elements [35].

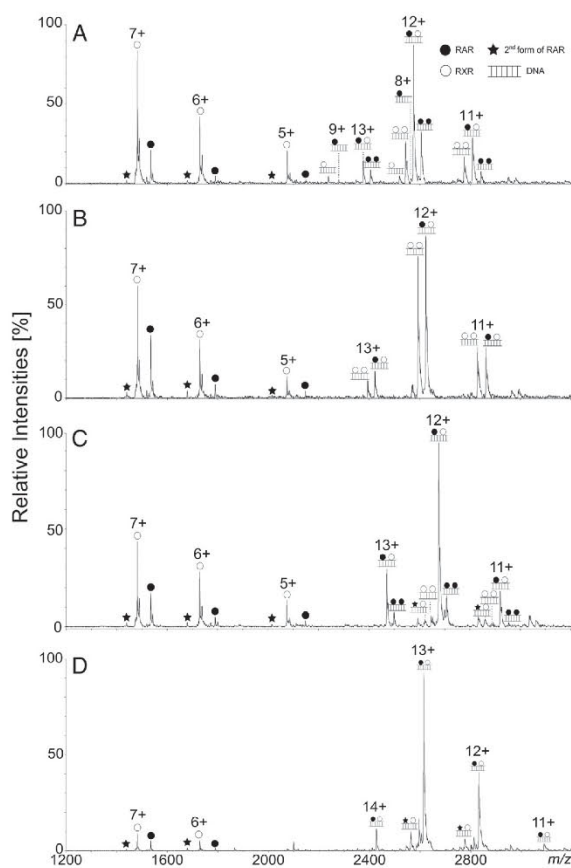


Fig. 5. Mass spectra of retinoic receptors RAR–RXR bound to DNA depending on the spacer length. (A) 0.25 equivalent of $RAR/\beta 2$ DR0; (B) 0.25 equivalent of $RAR/\beta 2$ DR1; (C) 0.25 equivalent of $RAR/\beta 2$ DR2; (D) 0.25 equivalent of $RAR/\beta 2$ DR5.

Please cite this article as: N.-T. Nguyen-Huynh, et al., Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry, *Biophys. Chem.* (2015), <http://dx.doi.org/10.1016/j.bpc.2015.10.006>

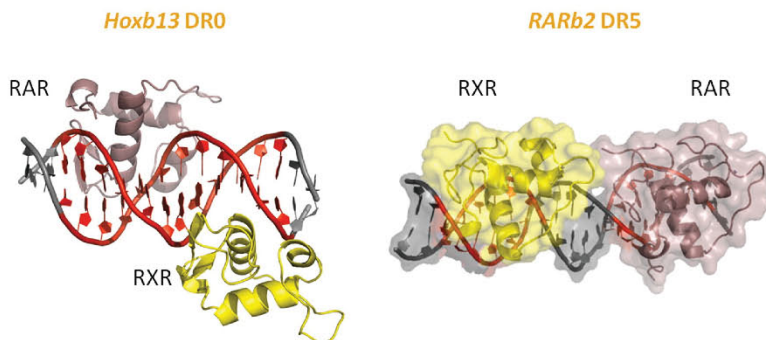


Fig. 6. 3D models of RAR–RXR DBDs bound to Hoxb13DR0 (left) and bound to RARb2 DR5 (right). The same color code of the DNA is used as Fig. 1. The RAR–RXR complexes were modeled based on crystal structures of RXR complexes (data not shown). On DR0, the 2 monomers lie on each side of the DNA making no protein–protein contact, while on DR5, the 2 monomers are side by side.

To further investigate how the heterodimerization is mediated on DNA, similar experiments were performed with RAREs displaying different sequences. Table 1 summarizes ratios of homodimer- and heterodimer–DNA complexes depending on the sequence and spacer length. These proportions were deduced from the relative intensities of the m/z ions with the approximation that all homodimer- and heterodimer–DNA complexes had similar response factors. As it can be noticed, heterodimer–DNA seems to be the major form regardless of the considered RAREs. This result suggests a poor selectivity in RARE targets when forming a stabilized favorable heterodimer interface. Except for RAR β 2 DR5, homodimers- and heterodimer–DNA coexist in solution, the homodimer/heterodimer ratio varies according to the DNA. This suggests that the sequence of DNA has influence over the repartition and the specificity of complexes formed with retinoic receptors, and this influence can be compensated by the difference in the sequence of the half sites.

4. Conclusion

This work illustrates the potential of MS to obtain insight into different species present in solution in a fast, robust and low sample-consuming way. The cooperativity of heterodimerization of RAR–RXR bound to DR5 elements was unambiguously highlighted, whereas the binding to DR0 elements is non-cooperative. The influence of protein/DNA molar ratios, sequence and length of DNA spacer on the relative abundance of RAR–RXR–DNA complexes was also highlighted. ESI-MS was crucial in allowing the study of retinoic receptor heterodimerization as a function of DNA. Thanks to its mass precision, assembly and disassembly of the different oligomeric states could be monitored as a function of the experimental conditions. This method is complementary to structural methods and will contribute to our understanding of the critical role the spacer plays in the specificity of the dimer. Although the first

crystal structure of NR DBD was reported in 1991 [36], the number of available crystal structures with diverse or natural REs remains very low, highlighting the need to find alternative structural techniques. The MS information about the dynamics and composition of dimers will help to define conditions favoring heterodimer for crystallization and structural analysis. Considering the strong heterodimerization interface in the RAR–RXR–LBDs that promotes the binding of the full-length receptors on these RAREs, similar MS analyses on the entire protein (and not only the DBD) will now be planned, possibly with co-regulators that may modulate the composition of the dimers, as already shown for PPAR–RXR heterodimer and RXR homodimer bound to PPRE [37]. This might help to better detect and understand the recruitment of co-activators and/or co-repressors, and determine the influence of LBD on the NR–DNA heterodimerization. In combination with complementary techniques, MS data will help to understand allosteric effects, follow changes in the heterodimerization surfaces and hence the nature of formed complexes.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bpc.2015.10.006>.

Acknowledgments

We thank the Proteomic Platform of IGBMC, the Structural Biology and Genomics platform (IGBMC) and Catherine Birck for help in the MST experiments, Anna Belorusova and Bruno Kieffer for discussion, and Roland H. Stote for the English correction. Nha-Thi Nguyen-Huynh is supported by a PhD Scholarship from the French MESR (Ministère de l'Enseignement Supérieur et de la Recherche). The project was supported by the Centre National pour la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Agence Nationale de Recherche (ANR-11-BSV8-023), the Association pour la Recherche sur le Cancer (ARC) (SFI20121205585), the Fondation pour la Recherche Médicale (FRM), Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions, the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01).

References

- [1] J.D. Nelson, O. Denisenko, K. Bomsztyk, Protocol for the fast chromatin immunoprecipitation (ChIP) method, *Nat. Protoc.* 1 (2006) 179–185, <http://dx.doi.org/10.1038/nprot.2006.27>.
- [2] L.M. Hellman, M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions, *Nat. Protoc.* 2 (2007) 1849–1861, <http://dx.doi.org/10.1038/nprot.2007.249>.
- [3] K.K. Wu, Analysis of protein–DNA binding by streptavidin–agarose pulldown, *Methods Mol. Biol.* 338 (2006) 281–290, <http://dx.doi.org/10.1385/1-59745-097-9:281>.

Table 1

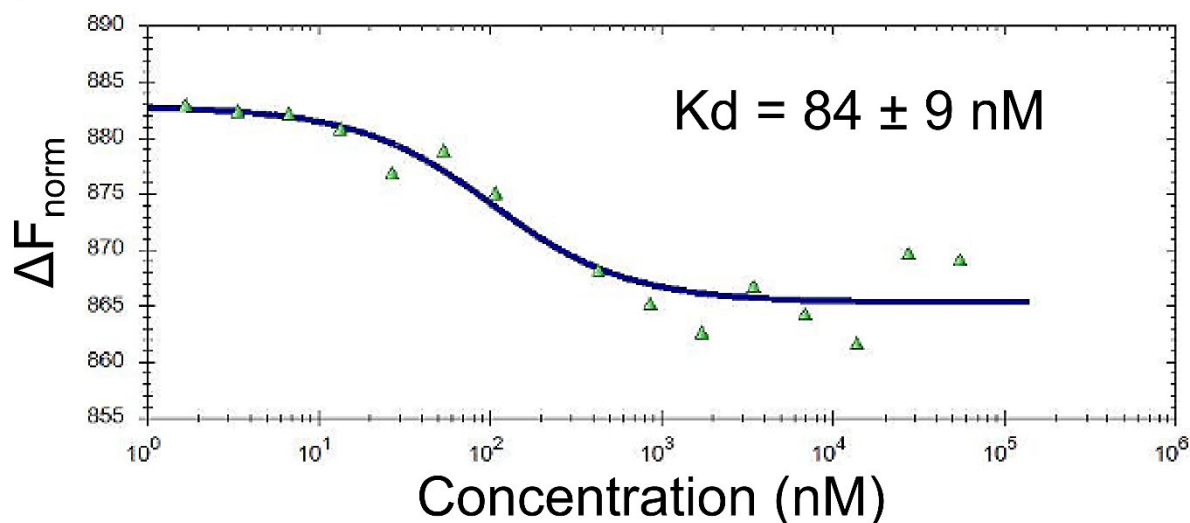
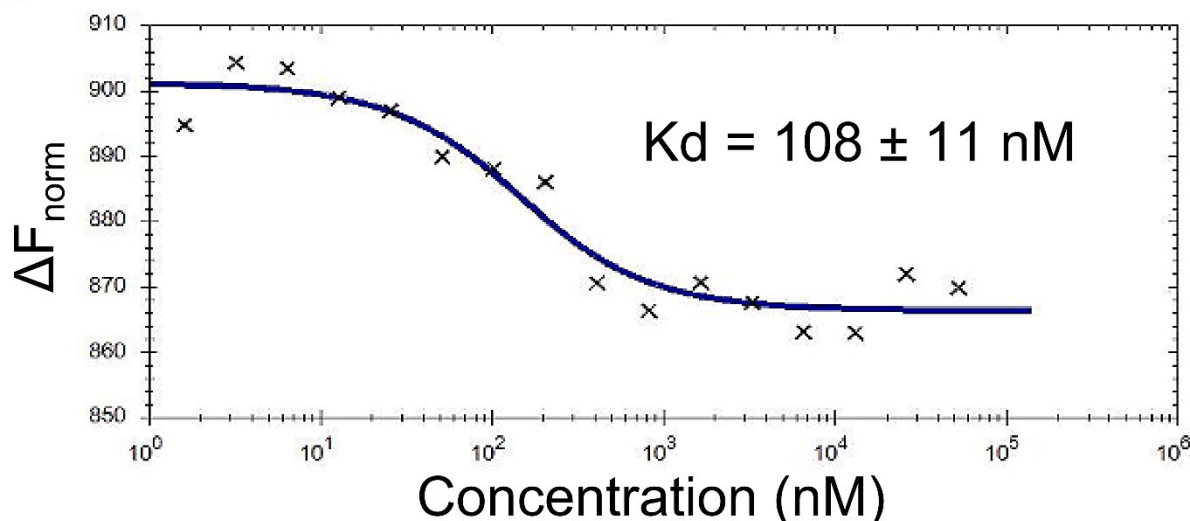
Relative quantity of homodimers and heterodimer of RAR–RXR bound to 0.25 equivalence of DNA with regard to the total quantity of existing dimers, depending on the sequence and spacer length: X = less than 20%, XX = between 20% and 50%, XXX = between 50% and 75%, XXXX = between 76% and 99%, XXXXX = 100%.

Sequence	Spacer length	Monomers	Homodimers	Heterodimer RAR–RXR
RAR β 2	DR0	Present	XX	XXX
	DR0–rev	Present	XX	XXX
	DR1	Absent	XX	XXX
	DR2	Absent	XX	XXXX
	DR5	Absent	–	XXXXXX
Hoxb13	DR0	Absent	XXX	XX
Msi	DR0	Absent	XX	XXX
F11r	DR5	Absent	XX	XXX

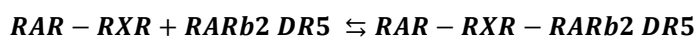
Please cite this article as: N.-T. Nguyen-Huynh, et al., Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry, *Biophys. Chem.* (2015), <http://dx.doi.org/10.1016/j.bpc.2015.10.006>

- [4] E. Jagelská, V. Brázda, S. Pospisilová, B. Vojtesek, E. Palecek, New ELISA technique for analysis of p53 protein/DNA binding properties, *J. Immunol. Methods* 267 (2002) 227–235.
- [5] D.R. Setzer, D.B. Schulman, C.V. Gunther, M.J. Bumbulis, Use of a reporter gene assay in yeast for genetic analysis of DNA-protein interactions, *Methods Mol. Biol.* 543 (2009) 219–241, http://dx.doi.org/10.1007/978-1-60327-015-1_15.
- [6] M. Mark, N.B. Ghyselinck, P. Chambon, FUNCTION OF RETINOIC NUCLEAR RECEPTORS: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis, *Annu. Rev. Pharmacol. Toxicol.* 46 (2006) 451–480, <http://dx.doi.org/10.1146/annurev.pharmtox.46.120604.141156>.
- [7] M. Mark, N.B. Ghyselinck, P. Chambon, Function of retinoic acid receptors during embryonic development, *Nucl. Recept. Signal.* 7 (2009) <http://dx.doi.org/10.1621/nrs.07002>.
- [8] C. Rochette-Egly, P. Germain, Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs), *Nucl. Recept. Signal.* 7 (2009) 1–18, <http://dx.doi.org/10.1621/nrs.07005>.
- [9] J.E. Balmer, R. Blomhoff, A robust characterization of retinoic acid response elements based on a comparison of sites in three species, *J. Steroid Biochem. Mol. Biol.* 96 (2005) 347–354, <http://dx.doi.org/10.1016/j.jsbmb.2005.05.005>.
- [10] E. Moutier, T. Ye, M.-A. Choukrallah, S. Urban, J. Osz, A. Chatagnon, et al., Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology, *J. Biol. Chem.* 287 (2012) 26328–26341, <http://dx.doi.org/10.1074/jbc.M112.361790>.
- [11] F. Rastinejad, T. Wagner, Q. Zhao, S. Khorasanizadeh, Structure of the RXR–RAR DNA-binding complex on the retinoic acid response element DR1, *EMBO J.* 19 (2000) 1045–1054, <http://dx.doi.org/10.1093/emboj/19.5.1045>.
- [12] J. Osz, A.G. McEwen, P. Poussin-Courmontagne, E. Moutier, C. Birck, I. Davidson, et al., Structural Basis of Natural Promoter Recognition by the Retinoic X Nuclear Receptor, *Sci. Rep.* 5 (2015) <http://dx.doi.org/10.1038/srep08216>.
- [13] N. Rochel, F. Ciesielski, J. Godet, E. Moman, M. Roessel, C. Peluso-Iltis, et al., Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings, *Nat. Struct. Mol. Biol.* 18 (2011) 564–570, <http://dx.doi.org/10.1038/nsmb.2054>.
- [14] F.W. McLafferty, E.K. Fridriksson, D.M. Horn, M.A. Lewis, R.A. Zubarev, Biomolecule mass spectrometry, *Science* 284 (1999) 1289–1290, <http://dx.doi.org/10.1126/science.284.5418.1289>.
- [15] C. Bich, R. Zenobi, Mass spectrometry of large complexes, *Curr. Opin. Struct. Biol.* 19 (2009) 632–639, <http://dx.doi.org/10.1016/j.sbi.2009.08.004>.
- [16] A.J.R. Heck, R.H.H. van den Heuvel, Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom. Rev.* 23 (2004) 368–389, <http://dx.doi.org/10.1002/mas.10081>.
- [17] M. Sharon, How far can we go with structural mass spectrometry of protein complexes? *J. Am. Soc. Mass Spectrom.* 21 (2010) 487–500, <http://dx.doi.org/10.1016/j.jasms.2009.12.017>.
- [18] P. Singh, A. Panchaud, D.R. Goodlett, Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique, *Anal. Chem.* 82 (2010) 2636–2642, <http://dx.doi.org/10.1021/ac100072a>.
- [19] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, et al., Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics, *Mol. Cell. Proteomics* 9 (2010) 1634–1649, <http://dx.doi.org/10.1074/mcp.R000001-MCP201>.
- [20] J. Rappsilber, The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, *J. Struct. Biol.* 173 (2011) 530–540, <http://dx.doi.org/10.1016/j.jsb.2010.10.014>.
- [21] H. Videler, L.L. Ilag, A.R.C. McKay, C.L. Hanson, C.V. Robinson, Mass spectrometry of intact ribosomes, *FEBS Lett.* 579 (2005) 943–947, <http://dx.doi.org/10.1016/j.febslet.2004.12.003>.
- [22] J.A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev.* 16 (1997) 1–23, [http://dx.doi.org/10.1002/\(SICI\)1098-2787\(1997\)16:1<1::AID-MA51>3.0.CO;2-L](http://dx.doi.org/10.1002/(SICI)1098-2787(1997)16:1<1::AID-MA51>3.0.CO;2-L).
- [23] T.D. Veenstra, Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes, *Biochem. Biophys. Res. Commun.* 257 (1999) 1–5, <http://dx.doi.org/10.1006/bbrc.1998.0103>.
- [24] C.L. Hanson, C.V. Robinson, Protein-nucleic acid interactions and the expanding role of mass spectrometry, *J. Biol. Chem.* 279 (2004) 24907–24910, <http://dx.doi.org/10.1074/jbc.R300037200>.
- [25] S. Akashi, R. Osawa, Y. Nishimura, Evaluation of protein-DNA binding affinity by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* 16 (2005) 116–125, <http://dx.doi.org/10.1016/j.jasms.2004.09.021>.
- [26] K. Bleicher, E. Bayer, Various factors influencing the signal intensity of oligonucleotides in electrospray mass spectrometry, *Biol. Mass Spectrom.* 23 (1994) 320–322, <http://dx.doi.org/10.1002/bms.1200230604>.
- [27] P. Agback, H. Baumann, S. Knapp, R. Ladenstein, T. Härd, Architecture of nonspecific protein–DNA interactions in the Sso7d–DNA complex, *Nat. Struct. Mol. Biol.* 5 (1998) 579–584, <http://dx.doi.org/10.1038/836>.
- [28] S.A.I. Seidel, P.M. Dijkman, W.A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, et al., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions, *Methods* 59 (2013) 301–315, <http://dx.doi.org/10.1016/j.jymeth.2012.12.005>.
- [29] C. Zechel, X.Q. Shen, J.Y. Chen, Z.P. Chen, P. Chambon, H. Gronemeyer, The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats, *EMBO J.* 13 (1994) 1425–1433.
- [30] C. Bich, C. Bovet, N. Rochel, C. Peluso-Iltis, A. Panagiotidis, A. Nazabal, et al., Detection of nucleic acid–nuclear hormone receptor complexes with mass spectrometry, *J. Am. Soc. Mass Spectrom.* 21 (2010) 635–645, <http://dx.doi.org/10.1016/j.jasms.2009.12.004>.
- [31] O. Givaty, Y. Levy, Protein sliding along DNA: dynamics and structural characterization, *J. Mol. Biol.* 385 (2009) 1087–1097, <http://dx.doi.org/10.1016/j.jmb.2008.11.016>.
- [32] P. Chambon, A decade of molecular biology of retinoic acid receptors, *FASEB J.* 10 (1996) 940–954.
- [33] Q. Zhao, S.A. Chasse, S. Devarakonda, M.L. Sierk, B. Ahvazi, F. Rastinejad, Structural basis of RXR–DNA interactions, *J. Mol. Biol.* 296 (2000) 509–520, <http://dx.doi.org/10.1006/jmbi.1999.3457>.
- [34] F. Rastinejad, T. Perlmann, R.M. Evans, P.B. Sigler, Structural determinants of nuclear receptor assembly on DNA direct repeats, *Nature* 375 (1995) 203–211, <http://dx.doi.org/10.1038/375203a0>.
- [35] W.H. Hudson, C. Youn, E.A. Ortlund, The structural basis of direct glucocorticoid-mediated transrepression, *Nat. Struct. Mol. Biol.* 20 (2013) 53–58, <http://dx.doi.org/10.1038/nsmb.2456>.
- [36] B.F. Luisi, W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, P.B. Sigler, Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA, *Nature* 352 (1991) 497–505, <http://dx.doi.org/10.1038/352497a0>.
- [37] A. Ijpenberg, N.S. Tan, L. Gelman, S. Kersten, J. Seydoux, J. Xu, et al., In vivo activation of PPAR target genes by RXR homodimers, *EMBO J.* 23 (2004) 2083–2091, <http://dx.doi.org/10.1038/sj.emboj.7600209>.

SUPPLEMENTARY MATERIAL

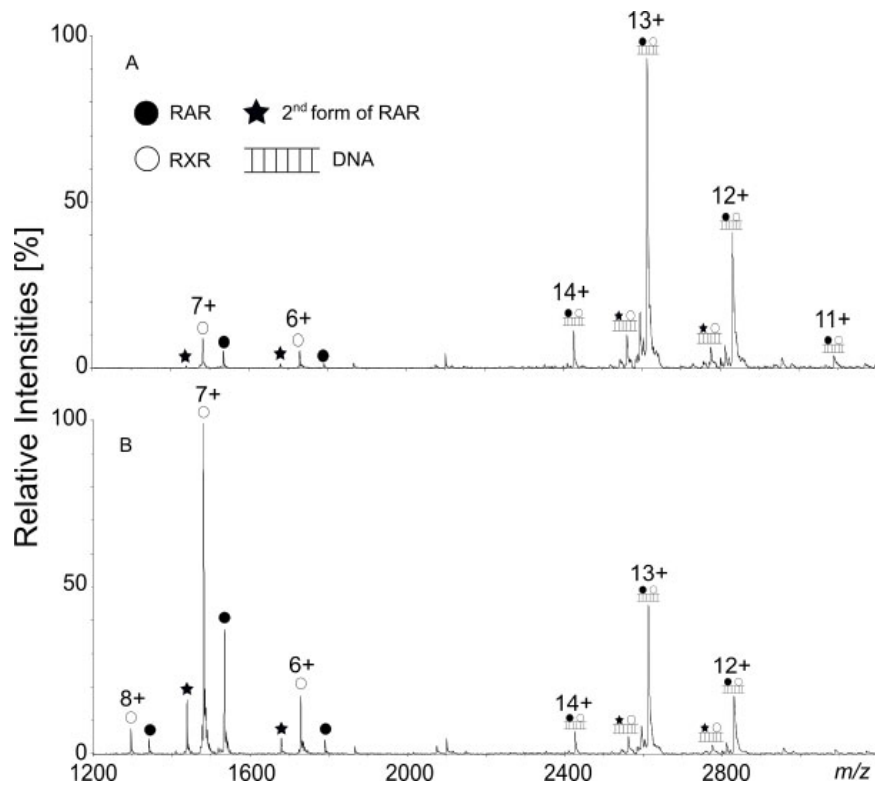
A

B


Supplementary Figure 1. DNA binding to RAR-RXR DBD-LBD measured by MST. Unlabeled *RARb2* DR5 was titrated into a fixed concentration of labeled RAR-RXR. The representative MST curves are shown in (A) 50 mM HEPES, pH 8.0, 100 mM sodium chloride, 2% glycerol and (B) 100mM ammonium acetate, pH 7.8 buffers. ΔF_{norm} = normalized fluorescence change. Three measurements were averaged and the standard deviation was determined. The K_d values were calculated by directly fitting the curve using the Hill function as described in [28]. In this study:

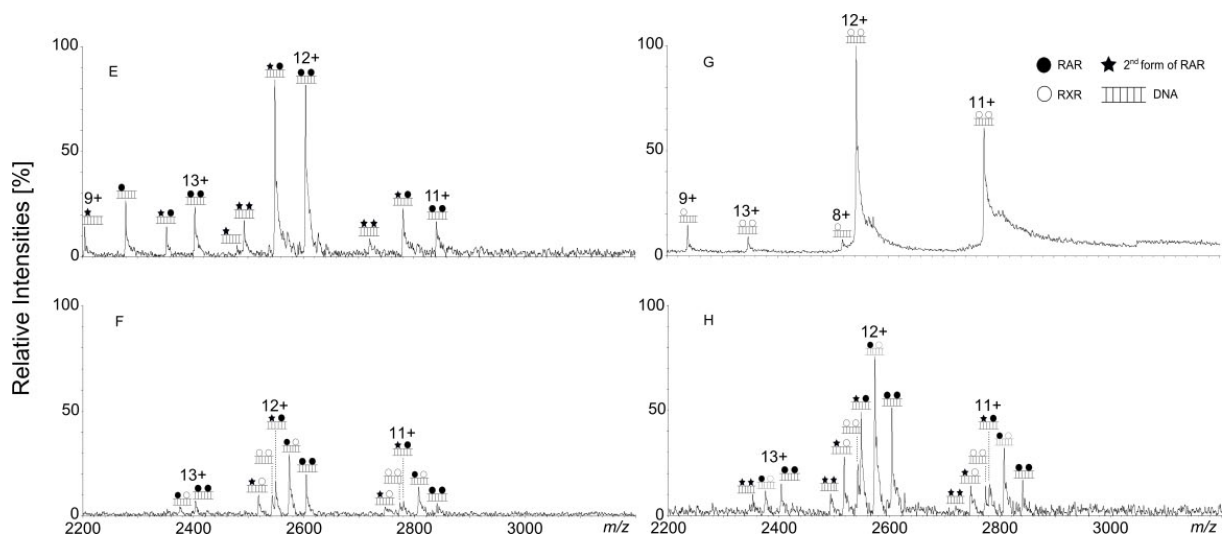


$$K_d = \frac{[RAR-RXR]_{\text{free}} \cdot [RARb2 DR5]_{\text{free}}}{[RAR-RXR-RARb2 DR5]}$$

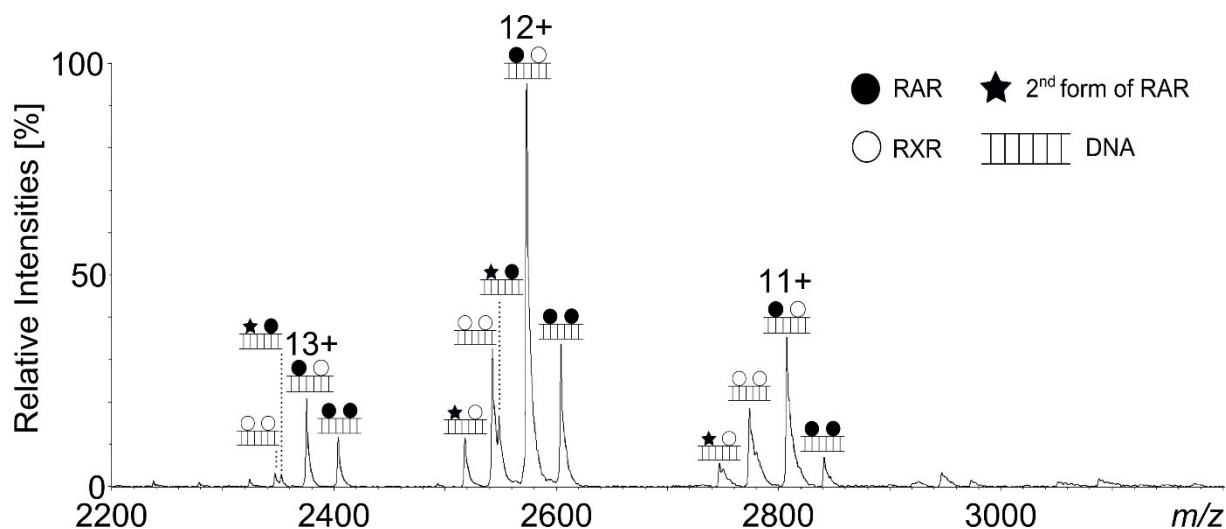
$$K_d = \frac{([RAR-RXR]^{\circ} - [RAR-RXR-RARb2 DR5]) \cdot ([RARb2 DR5]^{\circ} - [RAR-RXR-RARb2 DR5])}{[RAR-RXR-RARb2 DR5]}$$



Supplementary Figure 2. Mass spectra of retinoic receptors RAR-RXR bound to 0.25 equivalent of *RARβ2* DR5 illustrating the influence of the pre-pulse time of the mass spectrometer on the detected abundance of free protein versus protein-DNA complexes. (A) Capillary exit at 150 V and pre-pulse time at 40 μ s; (B) Capillary exit at 150 V and pre-pulse time at 30 μ s.



Supplementary Figure 3. (A) RAR incubated with *Hoxb13* DR0; (B) RAR incubated with *Hoxb13* DR0 followed by an addition of RXR; (C) RXR incubated with *Hoxb13* DR0; (D) RXR incubated with *Hoxb13* DR0 followed by an addition of RAR.



Supplementary Figure 4. Mass spectra of RAR-RXR bound to 0.25 equivalent of *Msi* DR0.

Supplementary Table 1. Kd values (nM) for apparent binding affinities of RAR-RXR and different RAR binding element oligonucleotides from MST. Buffer A: 50 mM HEPES, pH 8.0, 100 mM sodium chloride, 2% glycerol. Buffer B: 100mM ammonium acetate, pH 7.8. * Measured by ITC in Moutier *et al.*, 2012.

Element	Buffer A	Buffer B
<i>RAR</i> β 2 DR5	73 \pm 8 *	-
<i>Hoxb13</i> DR0	80 \pm 10	118 \pm 15
<i>Msi</i> DR0	116 \pm 15 *	-
<i>Msi</i> DR0	76 \pm 13 *	-
<i>Socs3</i> DR0	110 \pm 16 *	-

4.5. Conclusion

Notre travail a montré le grand potentiel de la MS supramoléculaire dans le cadre d'étude d'interactions protéine-ADN. Ainsi nous pouvons étudier la présence des trois états d'oligomérisation (monomère, homodimère et hétérodimère) de RAR-RXR en fonction des ADN différents. Les résultats obtenus ont mis en évidence la relation entre les ratios molaires protéine/ADN et la dynamique d'assemblage des complexes. En outre, ils ont aussi montré que la séquence et la longueur du spacer de l'ADN jouent un rôle important dans la coopérativité d'hétérodimérisation. Ceci est particulièrement utile pour le screening des ADN et pour déterminer les conditions d'obtention d'hétérodimères en amont de la cristallographie.

En se basant sur ce protocole, les expériences préliminaires sont déjà réalisées et seront poursuivies sur un élément de réponse de type DR8. L'étude des oligomères de RAR-RXR y est plus complexe car il est composé de 3 sites de fixation potentiels (DR2-DR0-IR). De plus, il serait intéressant de conduire l'étude dynamique avec les protéines entières qui contiennent les LBDs (« ligand binding domain », domaine de liaison au ligand) en plus des DBDs. Ceci permettra de travailler au plus proche des conditions *in vivo*, ainsi que d'explorer le mécanisme de recrutement des co-activateurs et/ou co-répresseurs. Enfin, il serait également envisageable de combiner l'étude par l'ESI-MS avec l'approche CX-MS, afin d'avoir des informations complémentaires sur les zones d'interaction.

En résumé, les résultats de ces travaux de thèse ont confirmé le potentiel de la MS supramoléculaire pour l'étude des édifices non-covalents ainsi que sa place parmi les techniques d'analyse structurale. En particulier, l'ESI-MS permet de suivre la dynamique d'interaction et de détecter les espèces présentes en solution, et peut être utilisée en amont et/ou en parallèle des autres techniques.

Bibliographie

- [1] Rochette-Egly, C., Germain, P., Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nucl. Recept. Signal.* 2009, 7, 1–18.
- [2] Balmer, J.E., Blomhoff, R., A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.* 2005, 96, 347–354.
- [3] Moutier, E., Ye, T., Choukrallah, M.-A., Urban, S., et al., Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology. *J. Biol. Chem.* 2012, 287, 26328–26341.
- [4] Rastinejad, F., Wagner, T., Zhao, Q., Khorasanizadeh, S., Structure of the RXR–RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.* 2000, 19, 1045–1054.
- [5] Osz, J., McEwen, A.G., Poussin-Courmontagne, P., Moutier, E., et al., Structural Basis of Natural Promoter Recognition by the Retinoid X Nuclear Receptor. *Sci. Rep.* 2015, 5.
- [6] Rochel, N., Ciesielski, F., Godet, J., Moman, E., et al., Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* 2011, 18, 564–570.
- [7] Heck, A.J.R., van den Heuvel, R.H.H., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* 2004, 23, 368–389.
- [8] Sharon, M., How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? *J. Am. Soc. Mass Spectrom.* 2010, 21, 487–500.
- [9] Loo, J.A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* 1997, 16, 1–23.
- [10] Veenstra, T.D., Electrospray Ionization Mass Spectrometry: A Promising New Technique in the Study of Protein/DNA Noncovalent Complexes. *Biochem. Biophys. Res. Commun.* 1999, 257, 1–5.
- [11] Hanson, C.L., Robinson, C.V., Protein-Nucleic Acid Interactions and the Expanding Role of Mass Spectrometry. *J. Biol. Chem.* 2004, 279, 24907–24910.
- [12] Akashi, S., Osawa, R., Nishimura, Y., Evaluation of protein-DNA binding affinity by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2005, 16, 116–125.
- [13] Bleicher, K., Bayer, E., Various factors influencing the signal intensity of oligonucleotides in electrospray mass spectrometry. *Biol. Mass Spectrom.* 1994, 23, 320–322.
- [14] Gabelica, V., Pauw, E.D., Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom. Rev.* 2005, 24, 566–587.
- [15] Givaty, O., Levy, Y., Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* 2009, 385, 1087–1097.
- [16] Chambon, P., A decade of molecular biology of retinoic acid receptors. *FASEB J.* 1996, 10, 940–954.

Chapitre 5 : Mesure par ESI-MS des constantes de dissociation des complexes RXR-ligand

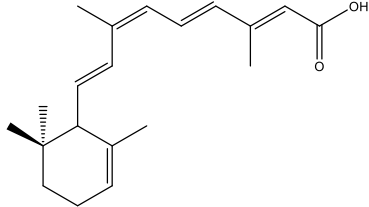
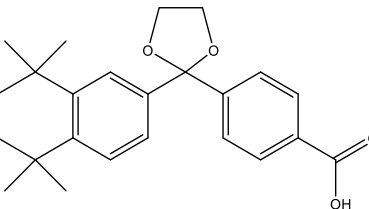
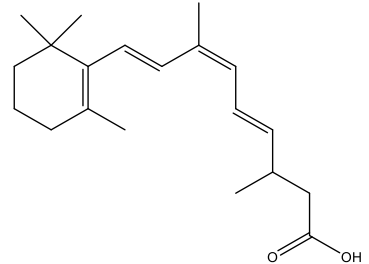
Dans le chapitre précédent, nous avons vu que l'ESI-MS supramoléculaire permettait de détecter des espèces présentes en solution ainsi que de suivre la dynamique d'assemblage des complexes non-covalents. Dans ce chapitre, nous allons exploiter une autre information pouvant être obtenue par l'ESI-MS : les constantes de dissociation des complexes. Pour cela, nous avons adapté cette approche pour mesurer les constantes de dissociation des complexes formés entre RXR et trois différents ligands. Tout d'abord, nous allons aborder des modèles mathématiques développés pour le calcul des constantes de dissociation par ESI-MS. Ensuite, nous mentionnerons les difficultés rencontrées ainsi que la mise au point des expériences de titrage. Enfin, nous discuterons des valeurs de constantes de dissociation obtenues ainsi que les propositions de solution aux problèmes de notre système d'étude.

5.1. Contexte de l'étude : les complexes RXR-ligand

Dans le cadre d'une seconde collaboration avec le Dr. Natacha Rochel au Département de Biologie Structurale Intégrative de l'IGBMC, nous avons étudié les complexes formés par RXR LBD (récepteur X des rétinoïdes, domaine de liaison au ligand) avec différents ligands. RXR fait partie de la famille des récepteurs nucléaires et est un facteur de transcription qui régule l'expression des gènes. Son fonctionnement est modulé de par son association avec des ligands agonistes, antagonistes ou synergistes [1].

Dans ce projet de recherche, nos collaborateurs s'intéressent à un ligand agoniste de RXR : le R-DHRA (l'énantiomère R de l'acide 9-cis-13,14-dihydrorétinoïque), récemment identifié comme ligand naturel chez les souris [2]. L'objectif de cette étude est de savoir si le R-DHRA active et se lie au RXR de la même façon que d'autres ligands existants qui sont le 9-cis-RA (acide 9-cis-rétinoïque, ligand naturel) et le BMS 649 (ligand synthétique). Le Tableau 1 regroupe les informations concernant ces ligands.

Tableau 1. Les différents ligands utilisés lors de cette étude.

Ligand	Structure	Formule brute	Masse (Da)	Solvant de conservation
9-cis RA N° CAS : 5300-03-8		$C_{20}H_{28}O_2$	300,4	Ethanol
BMS 649 N° CAS : 146670-40-8		$C_{24}H_{28}O_4$	380,5	Ethanol
R-DHRA -		$C_{20}H_{30}O_2$	302,5	Ethanol

Pour cela, différentes techniques permettant la détermination des constantes de dissociation ont été considérées. En ITC, la mesure est rendue difficile, dans le cas de notre étude, à cause des échanges de chaleurs avec le solvant organique : en raison de la faible quantité de ligand disponible, il n'est pas possible de garder la quantité de solvant organique fixe lors de la mesure. En quenching de fluorescence, nos collaborateurs ne peuvent vérifier que la poche de remplissage de RXR était vraiment inoccupée en absence du ligand. En ESI-MS supramoléculaire, cette vérification pourra se faire plus facilement ; en outre, cette technique a été reportée dans plusieurs études comme permettant de déterminer avec succès des constantes de dissociation des complexes protéine-ligand [3-7]. De plus, quelques études développées auparavant au sein de notre laboratoire ont également confirmé le potentiel de l'ESI-MS supramoléculaire pour la mesure des constantes de dissociation des complexes non-covalents de type protéine-protéine [8].

Dans ce contexte, nous avons voulu adapté cette approche pour déterminer les constantes de dissociation des complexes RXR-ligand par des expériences de titrage.

5.2. Modèles mathématiques du calcul des constantes de dissociation

Pour évaluer la stabilité des complexes non-covalents par MS, les méthodes les plus communément utilisées sont les expériences de compétition [3,5,6] et les expériences de titrage [9,10]. Pour notre étude, nous avons choisi les expériences de titrage pour déterminer les valeurs des constantes de dissociation de nos complexes. Dans cette méthode, la concentration de la protéine est gardée constante tandis que la concentration du ligand varie. Pour chaque concentration du ligand, l'intensité du signal correspondant à la protéine libre et l'intensité du signal correspondant au complexe sont mesurées sur le spectre de masse. A partir de ces données, une régression non linéaire effectuée sur les points de titrage permet de déterminer la valeur de la constante de dissociation pour chaque complexe.

Dans ce sens, deux approches ont été développées par Daniel *et al.* [9] et Jecklin *et al.* [10]. Si elles se basent sur le même principe de résolution d'une équation du second degré, elles diffèrent par les relations considérées, et ainsi l'expression finale et la valeur obtenue : constante d'association K_a pour Daniel *et al.*, et constante de dissociation K_d pour Jecklin *et al.*, K_d étant l'inverse de K_a . Nous détaillons ci-dessous les étapes de résolution non explicitées dans ces deux publications.

Nous notons $[PL]$ la concentration du complexe protéine-ligand, $[P]$ la concentration de protéine libre, et $[L]$ la concentration du ligand libre. Les concentrations totales sont représentées avec un « 0 » en indice.

Par définition, la constante d'association K_a s'écrit :

$$K_a = \frac{[PL]}{[P] \cdot [L]}$$

(Équation 1)

Par la conservation de la matière, la concentration totale de la protéine ($[P]_0$) est la somme des concentrations des espèces libres ($[P]$) et complexées ($[PL]$) :

$$[P]_0 = [P] + [PL] \rightarrow [P] = [P]_0 - [PL]$$

(Équation 2)

De même, pour le ligand :

$$[L]_0 = [L] + [PL] \rightarrow [L] = [L]_0 - [PL]$$

(Équation 3)

Daniel *et al.* posent R comme le rapport de la concentration du complexe sur la concentration de la protéine libre. Ainsi, R correspond au rapport de l'intensité du signal du complexe sur l'intensité du signal de la protéine libre, déterminées à partir des spectres de masse :

$$R = \frac{[PL]}{[P]}$$

(Équation 4)

De (Équation 1) et (Équation 4):

$$Ka = \frac{R}{[L]}$$

Avec (Équation 2) :

$$Ka = \frac{R}{[L]_0 - [PL]}$$

(Équation 5)

$$[PL] = [PL] \cdot \frac{[P]}{[P]} = \frac{[PL]}{[P]} = \frac{[PL]}{[P]} \cdot \frac{[P]_0}{[P]_0} = \frac{\frac{[PL]}{[P]} \cdot [P]_0}{\frac{[P]_0}{[P]}}$$

Avec (Équation 2) :

$$[PL] = \frac{\frac{[PL]}{[P]} \cdot [P]_0}{\frac{[P]_0}{[P]}}$$

En simplifiant le dénominateur par [P] :

$$[PL] = \frac{\frac{[PL]}{[P]}}{\frac{1 + \frac{[PL]}{[P]}}{1}} \cdot [P]_0$$

Avec (Équation 4) :

$$[PL] = \frac{R}{1 + R} \cdot [P]_0$$

En remplaçant cette expression dans (Équation 5) :

$$Ka = \frac{R}{[L]_0 - \frac{R}{1 + R} [P]_0}$$

$$Ka = \frac{R}{\frac{[L]_0(1 + R) - R[P]_0}{1 + R}}$$

$$Ka = \frac{R(1 + R)}{[L]_0(1 + R) - R[P]_0}$$

$$Ka = \frac{R^2 + R}{[L]_0(1 + R) - [P]_0 R}$$

$$Ka[L]_0(1 + R) - Ka[P]_0 R = R^2 + R$$

$$Ka[L]_0 + Ka[L]_0 R - Ka[P]_0 R = R^2 + R$$

$$-R^2 - R + Ka[L]_0 R - Ka[P]_0 R + Ka[L]_0 = 0$$

$$-R^2 + R(-1 + Ka[L]_0 - Ka[P]_0) + Ka[L]_0 = 0$$

$$\Delta = (Ka[L]_0 - Ka[P]_0 - 1)^2 - 4 \cdot (-1) \cdot Ka[L]_0$$

$$\text{Or } (Ka[L]_0 - Ka[P]_0 - 1)^2 \geq 0 \text{ et } 4 \cdot Ka[L]_0 > 0$$

$$\Delta = (Ka[L]_0 - Ka[P]_0 - 1)^2 - 4 \cdot (-1) \cdot Ka[L]_0 > 0$$

Il y a donc deux solutions :

$$\begin{cases} R_1 = \frac{-(-1 + Ka[L]_0 - Ka[P]_0) + \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2}}{2 \cdot (-1)} \\ R_2 = \frac{-(-1 + Ka[L]_0 - Ka[P]_0) - \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2}}{2 \cdot (-1)} \end{cases}$$

Examinons les deux solutions :

$$R_2 = \frac{1}{2} \cdot (-1 + Ka[L]_0 - Ka[P]_0 - \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2})$$

$$\text{Or } 4 \cdot Ka \cdot [L]_0 > 0$$

$$(Ka[L]_0 - Ka[P]_0 - 1)^2 < (Ka[L]_0 - Ka[P]_0 - 1)^2 + 4Ka[L]_0$$

Comme les deux côtés sont positifs, en prenant leurs racines carrées :

$$-1 - Ka[P]_0 + Ka[L]_0 < \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2}$$

$$-1 - Ka[P]_0 + Ka[L]_0 - \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2} < 0$$

$$R_2 < 0 \text{ (Impossible par définition de R)}$$

Par conséquent, nous ne gardons que R_1 comme solution :

$$R_1 = \frac{1}{2} \cdot \left(-1 + Ka[L]_0 - Ka[P]_0 + \sqrt{4Ka[L]_0 + (Ka[L]_0 - Ka[P]_0 - 1)^2} \right)$$

(Équation 6)

Ce qui est l'expression de Daniel *et al.*

Avec les mêmes étapes, Jecklin *et al.* utilisent l'expression de la constante de dissociation Kd à la place de Ka :

$$Kd = \frac{[P][L]}{[PL]}$$

$$R = \frac{[P]}{[PL]}$$

(Ici l'expression de R est l'inverse de celle définie par de Daniel *et al.*)

$$Kd = R \cdot [L] = R([L]_0 - [PL])$$

(Équation 7)

Comme avant :

$$[PL] = \frac{\frac{[PL]}{[P]}}{1 + \frac{[PL]}{[P]}} [P]_0$$

$$[PL] = \frac{\frac{1}{R}}{1 + \frac{1}{R}} [P]_0$$

$$[PL] = \frac{1}{R(1 + \frac{1}{R})} [P]_0$$

$$[PL] = \frac{1}{R\left(\frac{R+1}{R}\right)} [P]_0$$

$$[PL] = \frac{1}{1+R} [P]_0$$

En remplaçant $[PL]$ par son expression dans (Équation 7), on obtient :

$$Kd = R([L]_0 - \frac{1}{1+R} [P]_0)$$

$$Kd = \frac{R([L]_0(1+R) - [P]_0)}{1+R}$$

$$Kd + Kd.R = R[L]_0 + R^2[L]_0 - R[P]_0$$

$$-R^2[L]_0 + R(Kd - [L]_0 + [P]_0) + Kd = 0$$

$$\Delta = (Kd - [L]_0 + [P]_0)^2 - 4.(-1).[L]_0Kd$$

$$\text{Or } (Kd - [L]_0 + [P]_0)^2 \geq 0 \text{ et } 4.Kd.[L]_0 > 0$$

$$D'où \Delta = (Kd - [L]_0 + [P]_0)^2 + 4.Kd.[L]_0 > 0$$

Il y a donc deux solutions :

$$\begin{cases} R_1 = \frac{-Kd + [L]_0 - [P]_0 + \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}}{-2[L]_0} \\ R_2 = \frac{-Kd + [L]_0 - [P]_0 - \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}}{-2[L]_0} \end{cases}$$

Examinons les deux solutions :

$$R_1 = \frac{-Kd + [L]_0 - [P]_0 + \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}}{-2[L]_0}$$

$$R_1 = -\frac{-Kd + [L]_0 - [P]_0 + \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}}{2[L]_0}$$

Or $4 \cdot Kd \cdot [L]_0 > 0$

$$(Kd - [L]_0 + [P]_0)^2 < (Kd - [L]_0 + [P]_0)^2 + 4 \cdot Kd \cdot [L]_0$$

Comme les deux côtés sont positifs, en prenant leurs racines carrées :

$$Kd - [L]_0 + [P]_0 < \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}$$

$$0 < \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2} - Kd + [L]_0 - [P]_0$$

Donc en multipliant les deux côtés par $-\frac{1}{2[L]_0}$

$$R_1 < 0 \text{ (Impossible par définition de R)}$$

Par conséquent, nous ne prenons que R_2 comme solution :

$$R_2 = -\frac{-Kd + [L]_0 - [P]_0 - \sqrt{4Kd[L]_0 + (Kd - [L]_0 + [P]_0)^2}}{2[L]_0}$$

(Équation 8)

Ce qui est l'expression finale de Jecklin *et al.*

Que ce soit avec l'expression de Daniel *et al.* ou Jecklin *et al.*, en déterminant les rapports des intensités des signaux du complexe et de la protéine libre, et en connaissant les concentrations totales de la protéine et du ligand introduites, nous pouvons placer les points de titrage sur une courbe, avec la concentration totale en ligand en abscisse et en ordonnée le rapport $R : R = f([L]_0)$. Une régression non linéaire avec l'expression finale de R en fonction de $[L]_0$ permet de déterminer K_a ou K_d à partir des constantes de régression.

5.3. Difficultés et mise au point des expériences de titrage de RXR-ligand suivies par ESI-MS

Pour pouvoir utiliser les expressions de Daniel *et al.* ou Jecklin *et al.* où sont reliées les intensités du signal MS et les concentrations des espèces, il faudrait s'assurer que les rendements d'ionisation de la protéine libre et du complexe sont similaires. Ainsi, leurs facteurs de réponse sont comparables et les intensités des signaux peuvent être reliées à leur concentration en solution. Si ce n'est pas le cas, il faudrait prendre en compte les ratios des facteurs de réponse comme ont mentionné Gabelica *et al.* [4]. Dans notre cas, les masses des ligands BMS 649, 9-cis-RA et R-DHRA (successivement de 380,5 Da, 300,4 Da et 302,5 Da) sont beaucoup plus faibles devant la masse de la protéine qui est de 28 985 Da. Par conséquent, nous considérons que les masses de la protéine libre et du complexe protéine-ligand ainsi que leurs rapports m/z sont relativement proches, que leurs facteurs de réponse sont similaires, et que le changement conformationnel de la protéine induit par son association au ligand reste modeste.

Avant les analyses en ESI-MS, un échange de tampon est effectué pour avoir une solution de protéine dans le tampon d'acétate d'ammonium (AcONH_4) à 200 mM. Ce tampon permet de garder la configuration native de la protéine tout en étant compatible avec l'analyse en ESI-MS. Après l'échange, la concentration en protéine est vérifiée par l'expérience de Bradford. Les solutions mères de ligand quant à elles sont à 10^{-2} M dans l'éthanol.

Ensuite, la concentration de travail a été optimisée. En effet, elle doit être suffisante pour favoriser la formation du complexe et pour que les espèces présentes puissent être détectées en MS, mais une concentration trop importante peut mener à des associations non-spécifiques. De plus, dans notre cas d'étude où les solutions mères de ligands sont dans 100% d'éthanol, une contrainte supplémentaire s'impose : le solvant final du complexe protéine-ligand ne doit pas contenir plus de 2 % en éthanol dans le volume final pour ne pas dénaturer la protéine. Or, la concentration en protéine est gardée fixe et la concentration des ligands est augmentée pour les points de titrage ; le volume en éthanol est alors de plus en plus important dans la solution de mélange. Pour toutes ces raisons, nous avons optimisé et fixé la concentration de RXR à 5 μM , ce qui permet d'obtenir une bonne détection en MS et de varier les concentrations en ligand tout en restant inférieur à 2% en éthanol. Ainsi, les mélanges de protéine-ligand sont réalisés à 4°C dans 100 mM AcONH_4 , pH 8, avec la concentration de protéine fixe à 5 μM et la concentration du ligand variable.

D'autre part, les paramètres instrumentaux ont également été optimisés, particulièrement la tension d'accélération (« capillary exit »), pour avoir une bonne détection des espèces tout en évitant toute fragmentation du complexe. Dans ce sens, la valeur optimale du « capillary exit » est fixée à 60 V, celle du « transfer time » à 100 μs et celle du « pre-pulse storage » à 40 μs . De même,

nous avons pris soin d'effectuer les points de mesure dans la même journée, avec le même appareil et les mêmes paramètres.

Lors des analyses en ESI-MS, nous avons remarqué que les complexes adhèrent longtemps au capillaire d'injection, ce qui demande beaucoup de temps de rinçage entre chaque analyse et peut induire des erreurs dans l'intensité du signal MS détecté. Par conséquent, nous avons optimisé nos étapes de rinçage. Après chaque analyse, le premier rinçage est réalisé d'abord avec du tampon AcONH₄ à 100 mM pour bien solubiliser et enlever de la protéine. Ensuite, nous faisons passer une solution d'AcONH₄ à 100 mM contenant 10 % en éthanol pour mieux solubiliser et enlever le ligand. Enfin, du tampon AcONH₄ à 100 mM est injecté pour conditionner le circuit avant l'analyse suivante. Avec ces trois étapes, le rinçage est beaucoup moins chronophage et beaucoup plus efficace qu'un simple passage du tampon.

La Figure 1 représente les spectres de masse les plus représentatifs de RXR seule et des complexes de RXR-ligand. RXR est présente sous une forme majoritaire et plusieurs formes minoritaires dues à la gluconoylation, seule la forme majoritaire de la protéine est annotée.

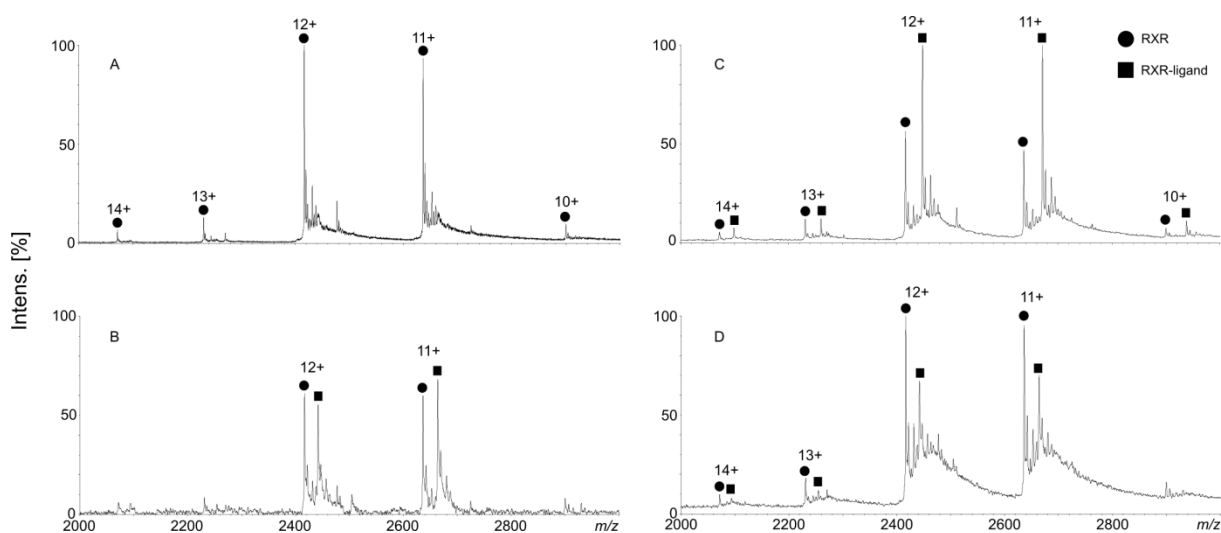


Figure 1. Spectres de masse ESI-MS en mode positif dans 100 mM AcONH₄ de (A) RXR seule à 5 μM, (B) RXR à 5 μM + 9-cis-RA à 25 μM, (C) RXR à 5 μM + BMS 649 à 50 μM, (D) RXR à 5 μM + R-DHRA à 150 μM. Incubation à 4°C pendant 15 min.

5.4. Mise au point et difficultés du calcul des constantes de dissociation pour trois ligands utilisés : 9-cis-RA, BMS 649 et R-DHRA

Après avoir obtenu des spectres des points de titrage, nous avons décidé d'utiliser l'expression de Jecklin *et al.* (Équation 8) pour notre détermination de constantes de dissociation. Pour ce faire, le logiciel KaleidaGraph (Synergy Software) a été utilisé pour réaliser la régression non linéaire.

Afin d'obtenir une bonne régression des données, après les études préliminaires, nous avons réalisé plus de points de titrage autour de la « flexion » de la courbe de régression. De même, comme le soulignent Daniel *et al.*, une bonne correction de la ligne de base des spectres de masse est importante car nous relevons les intensités des signaux de protéine libre et des complexes directement sur les spectres de masse. Par conséquent, plusieurs méthodes de correction de la ligne de base ont été testées. Dans la première méthode, nous déterminons manuellement la ligne de base pour chaque pic et la soustraction est ensuite appliquée. Dans la seconde, une valeur constante est soustraite des spectres de manière automatique par la fonction « baseline subtraction » du logiciel de traitement des spectres DataAnalysis (Bruker). Il en résulte que dans notre cas, la soustraction automatique de la ligne de base par le logiciel DataAnalysis est la méthode la plus robuste et la plus reproductible.

La Figure 2 représente les courbes obtenues à la suite des expériences de titrage des complexes RXR-ligand.

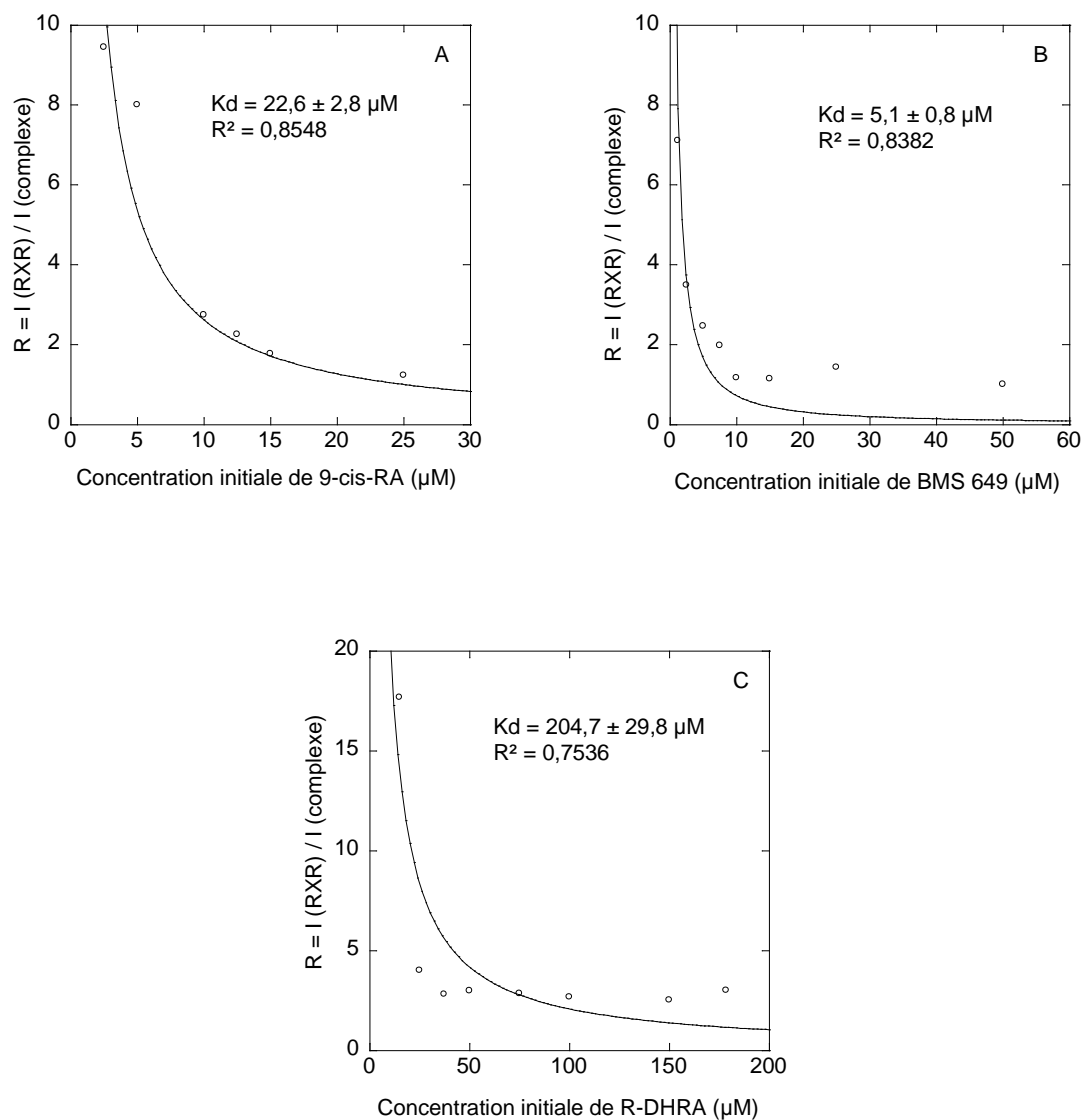


Figure 2. Courbes de titrage de RXR par (A) 9-cis-RA, (B) BMS 649, (C) R-DHRA avec 5 μM de RXR.

Le Tableau 2 résume les valeurs des constantes de dissociation qui découlent du titrage suivi par ESI-MS, ainsi que les constantes de dissociation évaluées par nos collaborateurs à l'aide des expériences de quenching de fluorescence.

Tableau 2. Les constantes de dissociation des complexes RXR-ligand déterminées par quenching de fluorescence et ESI-MS.

Complexe	Kd (nM) déterminée par quenching de fluorescence	Kd (nM) déterminée par ESI-MS
RXR – 9-cis-RA	20	22,6 x 10 ³
RXR – BMS 649	34	5,1 x 10 ³
RXR – R-DHRA	90	204,7 x 10 ³

Les valeurs calculées par l'ESI-MS diffèrent de trois ordres de grandeur des valeurs obtenues par nos collaborateurs. De plus, la faible valeur des coefficients de détermination (R^2) de la régression des points de titrage par ESI-MS traduit un manque d'adéquation entre le modèle mathématique utilisé et la distribution de nos points.

Face à ces problèmes, nous avons tout d'abord voulu vérifier le bon fonctionnement de notre logiciel de régression KaleidaGraph. Pour cela, une régression de nos données MS a été réalisée par nos collaborateurs avec le logiciel Origin (OriginLab). De cela, nous avons obtenu les mêmes valeurs de constantes de dissociation et les mêmes coefficients de détermination pour les deux logiciels, ce qui permet d'écarter l'hypothèse d'un dysfonctionnement de KaleidaGraph.

Par conséquent, nous avons envisagé différentes hypothèses explicatives concernant nos conditions expérimentales.

5.4.1. Problème de conformation de la protéine RXR ?

Cela ne nous paraît pas invraisemblable que dans la solution de protéine sur laquelle nous avons travaillé, la conformation de la protéine ait changé, soit lors de la purification, soit lors de son transfert dans le tampon AcONH₄ par échange de tampon. Par conséquent, les interactions entre RXR et les ligands auraient pu être perturbées. Afin de trancher sur ce point, il faudra travailler avec une autre préparation de protéine.

5.4.2. Problème de solubilité du ligand ?

Notre réflexion s'est tout d'abord portée sur les difficultés expérimentales rencontrées lors de l'analyse en ESI-MS. En effet, il a fallu rincer abondamment entre chaque analyse du fait de l'adhésion de l'échantillon sur le circuit d'injection. De plus, et ce malgré toutes nos précautions, le sprayer du spectromètre de masse s'est bouché une fois lors de nos expériences. Toutes ces constatations nous font émettre l'hypothèse d'un problème de solubilité des ligands dans notre tampon de travail. Par conséquent, nous avons souhaité doser le ligand dans le mélange avec la protéine dans 100 mM AcONH₄ par RMN pour chacun de nos points de titrage en ESI-MS pour une éventuelle correction des données. Malheureusement, nos concentrations de travail en ligand sont trop faibles pour un dosage en RMN. Néanmoins, pour avoir un ordre d'idée, nos collaborateurs ont

mesuré les concentrations du ligand BMS 649 ainsi que du R-DHRA seul à 200 μM dans 100 mM AcONH_4 + 5% EtOH (différent de notre solvant de travail : 100 mM AcONH_4 + 2% EtOH au maximum). Il s'est avéré que le BMS 649 est 16,7 fois moins concentré et le R-DHRA 20 fois moins.

Dans le cas de notre étude, nos collaborateurs ont réalisé des expériences de quenching de fluorescence avec les concentrations de travail autour de nM dans le tampon Tris 10 mM + NaCl 100 mM. Il serait possible que les ligands soient plus solubles dans ce tampon comparé à l' AcONH_4 , surtout avec des solutions 1000 fois moins concentrées qu'en ESI-MS.

D'autre part, dans les travaux de thèse réalisés par S. Sanglier-Cianferani sur les complexes RXR-ligand en ESI-MS [11], à 2,3 équivalent de ligand, le taux de fixation du 9-cis RA a été estimé à 75 % et du BMS 649 à 100 %. Or dans notre étude, aux alentours de cet équivalent, ces proportions sont de l'ordre de 25 % et 45 %. Cette différence semble suggérer un problème de ligands ou un souci concernant la protéine.

De plus, en ce qui concerne les ligands 9-cis RA et BMS 649, nos collaborateurs ont remarqué qu'il existe parfois de grandes différences d'un lot commercial à l'autre, ce qui par exemple influe fortement les expériences de cristallisation. Ces problèmes de reproductibilité peuvent signifier qu'un lot de moindre qualité peut avoir été employé lors de notre projet.

Tous ces faits tendent à confirmer l'hypothèse d'un problème de solubilité du ligand dans le mélange avec la protéine dans le tampon de travail (100 mM AcONH_4 + 2% EtOH). Cette mauvaise solubilité entraînerait une sur-estimation de la concentration totale du ligand et une sous-estimation de la concentration du complexe en solution, d'où une régression imparfaite lors de la détermination des constantes de dissociation.

Pour remédier à ce problème, plusieurs perspectives d'étude s'ouvrent à nous. Tout d'abord, nous pouvons travailler avec des solutions de ligand dans du DMSO (diméthylsulfoxyde) à la place de l'éthanol. Le DMSO étant un solvant qui solubilise mieux les ligands que l'éthanol, les ligands seront probablement mieux solubilisés dans 100 mM AcONH_4 + 2% DMSO que dans 100 mM AcONH_4 + 2% EtOH. Néanmoins, le DMSO est moins volatil et plus visqueux, sa présence va entraîner la formation d'adduits et diminuer l'intensité des signaux détectés en ESI-MS.

Nous pourrions également réaliser des expériences de compétition en incubant plusieurs ligands dans le même mélange avec RXR pour évaluer les constantes de dissociation relatives, ceci en supposant que les solubilités des ligands dans le tampon 100 mM AcONH_4 + 2% EtOH sont quasi identiques. Pour cela, les ligands utilisés doivent avoir des masses suffisamment distinctes pour pouvoir être différenciés lors d'analyse en ESI-MS. Dans notre cas d'étude, le BMS 649 de masse 380,5 Da peut être utilisé avec le R-DHRA (302,5 Da). Toutefois, cette approche ne permet pas d'obtenir les valeurs absolues des constantes de dissociation.

Il existe aussi la possibilité de vérifier l'association des ligands sur gel natif en incubant le mélange RXR-ligand avec des co-activateurs qui ne s'y lient qu'en présence du ligand. Enfin, nous pourrions travailler avec un lot différent de ligands pour s'affranchir du problème de reproductibilité entre des lots commerciaux. Il serait également intéressant de réaliser les expériences de quenching de fluorescence dans le tampon utilisé en ESI-MS, c'est-à-dire 100 mM AcONH_4 + 2% EtOH, afin d'étudier l'évolution des valeurs de constantes de dissociation selon la nature du tampon de travail.

5.5. Conclusion

Dans ce projet, nous cherchions à évaluer les constantes de dissociation des complexes de RXR-ligand par ESI-MS. Malgré les optimisations expérimentales et le traitement des données, les valeurs des constantes de dissociation obtenues sont différentes de trois ordres de grandeur de celles mesurées par nos collaborateurs par quenching de fluorescence. De plus, la faible valeur des coefficients de détermination de la régression de nos points de titrage suggère que nos données s'éloignent du modèle mathématique utilisé. Nos hypothèses qui pourraient l'expliquer sont un problème de solubilité des ligands dans le mélange avec la protéine dans le tampon d'acétate d'ammonium utilisé pour l'analyse en ESI-MS, et/ou un problème de conformation de protéine.

Quoiqu'il en soit, ces résultats tendent à montrer que la détermination de la constante de dissociation est possible par l'ESI-MS, mais que son succès dépend des systèmes étudiés.

Dans le cas des complexes protéiques hydrosolubles, cette détermination est facilitée grâce à la solubilité similaire des partenaires dans le même solvant. De plus, il est possible de détecter les partenaires protéiques libres sur le même spectre de masse, à la différence des systèmes protéine-ligand où nous ne pouvons détecter les signaux de la protéine libre et du ligand libre qu'avec différents paramètres instrumentaux.

Dans notre cas d'étude sur les complexes de type protéine-ligand, le problème de solubilité du ligand serait l'origine majeure des difficultés rencontrées dans l'évaluation de la valeur de constante de dissociation, et pourrait constituer la limite de notre approche. Pour nous affranchir de ce problème, nous pourrions envisager plusieurs perspectives, comme travailler avec d'autres lots de ligands, ou avec des solutions mères des ligands dans du DMSO au lieu de l'éthanol. Il serait également intéressant d'évaluer des constantes de dissociation relatives par des expériences de compétition, mais ceci implique l'hypothèse que les solubilités des ligands sont quasi identiques.

En résumé, la détermination de la constante de dissociation par ESI-MS est possible, mais notre étude suggère qu'elle ne peut être appliquée aux différents systèmes avec la même facilité. Dans certains cas, des problèmes de mise en œuvre peuvent survenir ; dès lors, la complexité de l'étude augmente ainsi que la nécessité d'effectuer des investigations plus approfondies. Ceci n'en reste pas moins intéressant, au contraire, il permet de mettre en lumière les difficultés pouvant être rencontrées, dont la résolution permettra de repousser encore plus loin le potentiel de la MS supramoléculaire.

Bibliographie

- [1] Dawson, M.I., Xia, Z., The retinoid X receptors and their ligands. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* 2012, 1821, 21–56.
- [2] Rühl, R., Krzyżosiak, A., Niewiadomska-Cimicka, A., Rochel, N., et al., 9-cis-13,14-Dihydroretinoic Acid Is an Endogenous Retinoid Acting as RXR Ligand in Mice. *PLoS Genet.* 2015, 11.
- [3] Jørgensen, T.J.D., Roepstorff, P., Heck, A.J.R., Direct Determination of Solution Binding Constants for Noncovalent Complexes between Bacterial Cell Wall Peptide Analogues and Vancomycin Group Antibiotics by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 1998, 70, 4427–4432.
- [4] Gabelica, V., Galic, N., Rosu, F., Houssier, C., De Pauw, E., Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry. *J. Mass Spectrom. JMS* 2003, 38, 491–501.
- [5] Tjernberg, A., Carnö, S., Oliv, F., Benkestock, K., et al., Determination of Dissociation Constants for Protein–Ligand Complexes by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 2004, 76, 4325–4331.
- [6] Erba, E.B., Zenobi, R., Mass spectrometric studies of dissociation constants of noncovalent complexes. *Annu. Rep. Sect. C Phys. Chem.* 2011, 107, 199–228.
- [7] El-Hawiet, A., Kitova, E.N., Arutyunov, D., Simpson, D.J., et al., Quantifying Ligand Binding to Large Protein Complexes Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 2012, 84, 3867–3870.
- [8] Mème, A., Développement de La Spectrométrie de Masse Pour L'étude D'édifices Non-Covalents En Chimie et En Biologie. Thèse Université de Strasbourg, 2010.
- [9] Daniel, J.M., McCombie, G., Wendt, S., Zenobi, R., Mass spectrometric determination of association constants of adenylate kinase with two noncovalent inhibitors. *J. Am. Soc. Mass Spectrom.* 2003, 14, 442–448.
- [10] Jecklin, M.C., Touboul, D., Bovet, C., Wortmann, A., Zenobi, R., Which Electrospray-Based Ionization Method Best Reflects Protein-Ligand Interactions Found in Solution? A Comparison of ESI, nanoESI, and ESSI for the Determination of Dissociation Constants with Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2008, 19, 332–343.
- [11] Sanglier-Cianferani, S., La Spectrométrie de Masse : Un Nouvel Outil Pour L'étude Des Interactions Faibles En Biologie. Thèse Université Louis Pasteur, Strasbourg, 2002.

Conclusion générale et perspectives

Dans la première partie de ce manuscrit consacrée à l'étude bibliographique, nous avons fait le point sur l'état d'avancement des connaissances en MS supramoléculaire. Même si l'ESI-MS reste un outil de choix pour étudier les complexes biologiques non-covalents, quelques-unes de ses limitations subsistent, comme par exemple la difficulté à transférer intacts certains complexes dans les tampons classiquement utilisés en ESI-MS. Dans ce contexte, la stratégie du pontage chimique couplé à la MS (CX-MS) émerge comme une alternative intéressante pour étudier les édifices non-covalents. Néanmoins, de nombreuses problématiques ont été soulevées, principalement en ce qui concerne l'optimisation des conditions de réaction et l'interprétation des données.

Dans ce contexte, ces travaux de thèse s'inscrivent dans la volonté de repousser le potentiel de la MS concernant l'analyse des complexes supramoléculaires, de développer les approches expérimentales et de répondre à des questions biologiques. Les résultats sur le développement de la MS ont été présentés à travers différents systèmes biologiques étudiés tout au long de cette thèse :

- Premièrement, nous avons développé avec succès une approche de CX-MS pour étudier le complexe non-covalent SAGA HAT. Son étude par la cristallographie et la microscopie électronique a rencontré des difficultés, liées probablement à la grande hétérogénéité et flexibilité du complexe. Dans ce contexte, nos objectifs étaient de déterminer la stœchiométrie ainsi que les zones de proximité de SAGA HAT par CX-MS. Pour cela, les conditions de réaction de pontage avec le BS3 (ratios agent de pontage/protéine, temps de réaction, températures) ont été optimisées, la cinétique de réaction suivie par MALDI-MS, et les peptides générés par digestion tryptique ont été analysés par nanoLC-MS/MS. Beaucoup de temps a ensuite été consacré à l'utilisation de pLink et xQuest, deux logiciels dédiés à l'identification des produits pontés, afin d'interpréter les données MS/MS obtenues. L'identification des peptides a été systématiquement validée manuellement. Les pontages inter- et intramoléculaires identifiés ont permis de dresser une carte des zones de proximité entre les sous-unités du complexe. De nombreux pontages entre ADA3 et ADA2B suggèrent de fortes interactions entre ces sous-unités, tandis qu'il n'existe aucun pontage entre SGF29 et ADA2B. Enfin, cette carte d'interactions a permis d'affiner par modélisation un premier modèle moléculaire à basse résolution du complexe SAGA HAT.

En perspective, nous cherchons maintenant à optimiser et appliquer cette méthodologie à d'autres systèmes, comme par exemple les protéines membranaires ou les protéines intrinsèquement non structurées. Dès lors, il serait intéressant de transférer des étapes réalisées dans le cadre de ces travaux de thèse sur des systèmes dont l'investigation est délicate par la MS supramoléculaire « classique ».

D'une façon plus générale, les résultats obtenus au cours de ce projet améliorent considérablement notre compréhension de l'approche CX-MS et permettent de mieux appréhender ses limites : sa mise en œuvre n'est pas triviale, chaque étape expérimentale est cruciale, dépend du stade précédent et contribue au succès de l'étude. L'accent doit être mis sur le choix de l'agent de pontage, le contrôle des conditions de

réaction, l'optimisation des analyses par MS et le traitement des données par des outils bio-informatiques adéquats.

- Deuxièmement, nous avons adapté l'approche ESI-MS supramoléculaire pour étudier les complexes formés par RAR-RXR DBDs avec différents ADN. Le choix de l'ESI-MS s'est imposé : cette technique permet de détecter les différentes espèces qui co-existent en solution ainsi que de suivre les assemblages des partenaires en fonction du temps ou du milieu. De par la nature polyanionique de l'ADN, l'étude des interactions de type protéine-ADN par ESI-MS est délicate à cause d'adduits alcalins qui élargissent le signal MS rendant la mesure des masses difficile voire impossible. Pour notre étude, après l'optimisation du protocole d'échange de tampon, nous avons optimisé les conditions d'observation de complexes d'intérêt puis effectué des expériences de titration, et suivi la dynamique d'interaction en fonction des conditions expérimentales. Les paramètres instrumentaux ont été également optimisés afin d'obtenir le meilleur signal sans dissocier les complexes.

Les expériences de titration suggèrent que la quantité d'ADN ajoutée joue un rôle important dans la formation des complexes avec RAR-RXR. En effet, un défaut d'ADN semble privilégier la formation d'hétérodimère RAR-RXR-ADN, tandis qu'en présence d'un excès d'ADN, d'autres espèces (monomères et homodimères) sont favorisées. Des expériences sur l'ordre d'ajout des protéines quant à elles ont permis de comprendre que : i) les deux sites d'interaction d'ADN sont susceptibles d'être occupés en totalité par la quantité de protéines présentes sans discrimination spécifique, ii) l'hétérodimérisation est coopérative. Enfin, nous avons aussi mis en lumière l'influence de la longueur et de la séquence des ADNs testés sur la répartition des espèces présentes.

Dans l'avenir, ce protocole ainsi mis au point sera utilisé pour conduire l'étude avec les protéines entières afin de travailler au plus proche des conditions physiologiques, ainsi que d'explorer le mécanisme de recrutement des co-activateurs ou co-répresseurs.

En résumé, nos résultats confirment le potentiel de l'ESI-MS dans la détection des différentes espèces qui co-existent en solution et qui sont difficiles à distinguer par d'autres méthodes d'analyse. En outre, elle permet de suivre la dynamique d'assemblage des complexes, et peut être utilisée en amont et/ou en parallèle des autres techniques.

- Enfin, nous avons souhaité adapter l'approche de mesure des constantes de dissociation par ESI-MS aux complexes RXR-ligands. Nous avons réalisé les expériences de titrage de la protéine par trois différents ligands (9-cis-RA, BMS 649 et R-DHRA), afin d'évaluer les valeurs des constantes de dissociation par une régression non linéaire de nos données. Les premiers résultats obtenus nous indiquent que nos données ne suivent pas le modèle mathématique utilisé, et que les valeurs des constantes de dissociation diffèrent de trois ordres de grandeur de celles mesurées par nos collaborateurs par quenching de fluorescence. Nos hypothèses qui pourraient l'expliquer sont un problème de solubilité des ligands dans le mélange avec la protéine dans le tampon d'acétate

d'ammonium utilisé pour l'analyse en ESI-MS, et/ou un problème de conformation de la protéine.

Face à cette difficulté, nous avons envisagé différentes perspectives. Notamment, il s'agit de travailler avec d'autres lots de ligands et de protéine, avec des solutions mères des ligands dans du DMSO au lieu de l'éthanol, et de s'assurer que la protéine est dans sa bonne conformation pour se lier correctement aux ligands. Il serait également intéressant d'évaluer des constantes de dissociation relatives par des expériences de compétition, mais ceci implique l'hypothèse que la part soluble des ligands est à peu près la même.

Quoiqu'il en soit, ces résultats ont soulevé un point important : la détermination de la constante de dissociation est possible par l'ESI-MS, mais elle ne peut être appliquée aux différents systèmes avec la même facilité. Dans certains cas, des problèmes de mise en œuvre peuvent survenir ; dès lors, la complexité de l'étude augmente ainsi que la nécessité d'effectuer des investigations plus approfondies.

En conclusion, les résultats de ces travaux de thèse ont permis de repousser le potentiel de la MS dans l'analyse des interactions non-covalentes des composés biologiques, que ce soit par MALDI-MS en couplage avec un pontage chimique, ou par ESI-MS en conditions natives. Ainsi, ces travaux de thèse nous permettent d'espérer de passer de la MS supramoléculaire dans des conditions « classiques » parfaitement adéquates et contrôlées (comme par exemple des systèmes bien solubles dans l'acétate d'ammonium, des complexes protéine-protéine) à des travaux sur des systèmes plus délicats (systèmes intégrant l'ADN difficile à dessaler, complexes flexibles, protéines membranaires ou encore protéines non structurées).

Annexes

Liste des communications scientifiques

1. Publications avec comité de lecture

- **Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex**

Nha-Thi Nguyen-Huynh, Grigory Sharov, Clément Potel, Pélagie Fichter, Simon Trowitzsch, Imre Berger, Valérie Lamour, Patrick Schultz, Noëlle Potier, Emmanuelle Leize-Wagner, *Protein Science* 2015, 24 (8), 1232–1246

- **Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry**

Nha-Thi Nguyen-Huynh, Judit Osz, Carole Peluso-Iltis, Natacha Rochel, Noëlle Potier, Emmanuelle Leize-Wagner, *Biophysical Chemistry* 2015

2. Communications orales

- **Chemical cross-linking and MALDI-MS for the characterization of intact protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

Workshop du Collège Franco-allemand CDFA 0407 - International Research Training Group GRK 1478, 9 décembre 2013, Strasbourg, France

- **Chemical cross-linking and MALDI-MS for the characterization of intact protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

Congrès français de Spectrométrie de Masse et d'Analyse Protéomique 2014, 30 juin-2 juillet 2014, Lyon, France

- **Chemical cross-linking coupled to mass spectrometry to determine the interaction network of protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

Workshop du Collège Franco-allemand CDFA 0407 - International Research Training Group GRK 1478, 5 novembre 2014, Strasbourg, France

- **Chemical cross-linking coupled to mass spectrometry to determine the interaction network of protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

Journée des Doctorants en chimie 2014, 7 novembre 2014, Strasbourg, France

- **Chemical cross-linking and mass spectrometry to determine the interaction network of protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

Symposium Chemical Complexity & Biology, 19-20 janvier 2015, Strasbourg, France

- **Interaction network of protein complexes by chemical cross-linking and mass spectrometry**

Nha-Thi Nguyen-Huynh, Grigory Sharov, Clément Potel, Pélagie Fichter, Simon Trowitzsch, Imre Berger, Valérie Lamour, Patrick Schultz, Noëlle Potier, Emmanuelle Leize-Wagner

Congrès français de Spectrométrie de Masse et d'Analyse Protéomique 2015, 15-18 septembre 2015, Ajaccio, France

3. Communications par affiche

- **Chemical cross-linking and MALDI-MS for the characterization of intact protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Emmanuelle Leize, Valérie Lamour, Noëlle Potier

7th European Proteomics Association Meeting, 14-17 Octobre 2013, Saint-Malo, France

- **Chemical cross-linking and MALDI-MS for the characterization of intact protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Emmanuelle Leize, Valérie Lamour, Noëlle Potier

Workshop on Native Mass Spectrometry and Ion Mobility, 8-9 avril 2014, Strasbourg, France

- **Chemical cross-linking and MALDI-MS for the characterization of intact protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

20th International Mass Spectrometry Conference, 24-29 août 2014, Genève, Suisse

- **Chemical cross-linking and mass spectrometry to determine the interaction network of protein complexes**

Nha-Thi Nguyen-Huynh, Pélagie Fichter, Grigory Sharov, Clément Potel, Patrick Schultz, Valérie Lamour, Noëlle Potier, Emmanuelle Leize-Wagner

63rd ASMS Conference on Mass Spectrometry & Allied Topics, 31 mai-4 juin 2015, St-Louis (MO), Etats-Unis

4. Enseignement en monitorat (Université de Strasbourg)

Année universitaire 2012-2013 :

- Travaux dirigés Cinétique Chimique, Licence 2 Chimie
(responsable : Pr. Paolo Samorì)
- Travaux dirigés Thermochimie, Licence 1 Chimie
(responsable : Pr. Rémi Barillon)
- Travaux dirigés Chimie pour les Sciences du Vivant, Licence 1 Sciences du Vivant
(responsable : Dr. Rachel Schurhammer)

Année universitaire 2013-2014 :

- Travaux dirigés Thermochimie, Licence 2 Sciences du Vivant
(responsable : Dr. Marco Cecchini)
- Travaux dirigés Chimie pour les Sciences du Vivant, Licence 1 Sciences du Vivant
(responsable : Dr. Rachel Schurhammer)

Année universitaire 2014-2015 :

- Travaux dirigés Thermochimie, Licence 2 Sciences du Vivant
(responsable : Dr. Marco Cecchini)
- Travaux dirigés Chimie pour les Sciences du Vivant, Licence 1 Sciences du Vivant
(responsable : Dr. Rachel Schurhammer)

Nha Thi NGUYEN HUYNH

Développements en spectrométrie de masse pour l'étude des complexes biologiques

Résumé

L'élucidation des interactions non-covalentes des complexes biologiques revêt d'une importance majeure dans la compréhension du fonctionnement cellulaire. L'objectif de ce travail de thèse est d'approfondir les développements de la spectrométrie de masse (MS) pour l'étude de ces complexes, que ce soit par MALDI-MS (la désorption-ionisation laser assistée par matrice) ou par ESI-MS (l'ionisation électrospray). Ce travail s'est articulé autour de trois axes : i) étude de la stœchiométrie et de la topologie du complexe SAGA HAT (Spt-Ada-Gcn5 Acétyltransferase, module Histone Acétyl Transferase) par pontage chimique couplé à la MS ; ii) suivi de la dimérisation des complexes formés par RAR-RXR (récepteur de l'acide rétinoïque - récepteur X des rétinoïdes) avec différents ADN ; iii) mesure de la constante de dissociation des complexes RXR-ligand. Les méthodologies développées ont permis de repousser le potentiel de la MS et d'obtenir des informations structurales des complexes biologiques.

Mots-clés : spectrométrie de masse, pontage chimique, complexes biologiques non-covalents, étude dynamique, interaction protéine-protéine, interaction protéine-ADN, interaction protéine-ligand

Summary

Elucidation of non-covalent interactions of biological complexes takes on great importance for the understanding of cellular function. The purpose of this thesis is a further development of mass spectrometry (MS) for the study of these complexes, either by MALDI-MS (matrix-assisted laser desorption-ionization) or by ESI-MS (electrospray ionization). This work was focused on three main lines: i) study of the stoichiometry and the topology of SAGA HAT (Spt-Ada-Gcn5 Acetyltransferase, Histone Acetyl Transferase module) complex by chemical cross-linking coupled to MS; ii) monitoring the dimerization of the complexes formed by RAR-RXR (retinoic acid receptor - retinoid X receptor) with different DNAs; iii) measuring the dissociation constant of RXR-ligand complexes. The developed methodologies made it possible to expand the potential of MS and get insight into structure of biological complexes.

Keywords: mass spectrometry, cross-linking, non-covalent biological complexes, dynamics study, protein-protein interaction, protein-DNA interaction, protein-ligand interaction