

Tables des matières

Partie I : Introduction générale	1
I.1 Interactions non covalentes dans les protéines	4
I.1.1 Interactions ioniques	4
I.1.2 Interactions de type dipolaire	5
I.1.2.1 Interactions ion-dipôle	6
I.1.2.2 Interactions dipôle-dipôle	7
I.1.2.3 Dipôles induits	8
I.1.2.3.1 Interaction ion-dipôle induit	8
I.1.2.3.2 Interaction dipôle-dipôle induit	9
I.1.2.3.3 Interactions de van der Waals	9
I.1.3 Liaisons hydrogènes	9
I.1.4 Interactions hydrophobes	11
I.1.5 Conclusions	12
I.2 Structure des protéines	12
I.2.1 Structure primaire	13
I.2.2 Structure secondaire	15
I.2.2.1 Caractéristiques de la structure secondaire	15
I.2.2.2 Méthodes spectroscopiques pour l'étude de la structure secondaire	17
I.2.2.2.1 Dichroïsme circulaire	17
I.2.2.2.2 Spectroscopie infrarouge (IR)	18
I.2.2.2.3 Spectroscopie de fluorescence	19
I.2.3 Structure tertiaire	20
I.2.3.1 Caractéristiques de la structure tertiaire des protéines	20
I.2.3.2 Méthodes classiques pour l'étude de la structure tertiaire des protéines	21
I.2.3.2.1 Cristallographie aux rayons X	21
I.2.3.2.2 Spectroscopie RMN	22
I.2.4 Structure quaternaire	23
I.2.4.1 Caractéristiques de la structure quaternaire des protéines	23
I.2.4.2 Méthodes classiques pour l'étude de la structure quaternaire des protéines et des macroassemblages	24
I.2.4.2.1 Chromatographie d'exclusion stérique	24
I.2.4.2.2 Diffraction aux petits angles (SAXS)	24
I.2.4.2.3 Microscopie électronique	25
I.3 Interactions protéine-ligand	26
I.3.1 Introduction aux interactions non covalentes protéine-ligand	26
I.3.2 Méthodes d'études des interactions protéine-ligand	29
I.3.2.1 Méthodes thermochimiques	29
I.3.2.1.1 Analyse calorimétrique par dosage isotherme (ITC)	29
I.3.2.1.2 Analyse calorimétrique différentielle (DSC)	30
I.3.2.2 Méthodes spectroscopiques	31
I.3.2.2.1 Fluorimétrie	31
I.3.2.2.2 RMN	32
I.3.2.2.3 Résonance de plasmons de surface	32
I.3.2.3 Autres méthodes	32
I.3.2.3.1 Electrophorèse capillaire d'affinité	32
I.3.2.3.2 Dialyse à l'équilibre	33
I.4 Etudes des complexes non covalents par spectrométrie de masse	34
I.4.1 Généralités de l'electrospray des protéines	34
I.4.2 Etudes des complexes protéine-ligand par spectrométrie de masse	35
I.4.2.1 Observation du complexe	36

I.4.2.2	Détermination des constantes d'interaction	36
I.4.3	<i>Etudes structurales des complexes protéiques par spectrométrie de masse</i>	39
I.4.3.1	Mesure de masse des complexes protéiques	40
I.4.3.2	Etude de la structure des complexes protéiques	41
I.4.3.3	Etude de la dynamique structurale des complexes protéiques	42
I.5	Conclusion	42
I.6	Bibliographie	43

Partie II : Etude des complexes protéine-ligand intacts par spectrométrie de masse.....51

II.A Déconvolution des interactions spécifiques et non spécifiques.....55

II.A.1	Remarques générales	57
II.A.2	Abstract	59
II.A.3	Introduction	59
II.A.4	Experimental	61
II.A.4.1	<i>Materials</i>	61
II.A.4.2	<i>Mass spectrometry</i>	61
II.A.4.3	<i>Deconvolution model</i>	62
II.A.5	Results and discussion	65
II.A.6	Conclusions	79
II.A.7	References	80

II.B Influence des paramètres expérimentaux sur les constantes de dissociation.....83

II.B.1	Introduction	85
II.B.2	Interaction entre CK et ses ligands dans différents tampons	86
II.B.2.1	<i>Etats de charge</i>	86
II.B.2.2	<i>Interaction CK-ADP ou CK-ATP dans différents tampons</i>	87
II.B.3	Interaction CK-ligand à différents pH	92
II.B.4	Comparaison des constantes obtenues par différents techniques	95
II.B.5	Conclusions	100
II.B.6	Bibliographie	101

Partie III – Développement de méthodes pour l'étude de la structure des protéines.....105

III.A Fragmentation des protéines entières dans la cellule ICR par dissociation par capture d'électrons.....109

III.A.1	Introduction	111
III.A.2	Matériel et méthodes	113
III.A.3	Fragmentation de l'ubiquitine	114
III.A.4	Fragmentation de la myoglobine	117
III.A.5	Conclusions	119
III.A.6	Bibliographie	120

III.B Modification des protéines par le diéthylpyrocarbonate (DEPC).....123

III.B.1	Introduction	125
---------	--------------	-----

III.B.2 Réactivité du DEPC	128
III.B.2.1 Identification des peptides modifiés	129
III.B.2.2 Identification des sites de modification	132
III.B.2.3 Modifications non spécifiques	133
III.B.3 Etude des protéines par modification chimique : approche « bottom-up »	135
III.B.3.1 Modification de la myoglobine intacte par le DEPC	135
III.B.3.2 Identification des sites de modification par l'approche « bottom-up »	136
III.B.4 Etude des protéines par modification chimique : approche « top-down »	140
III.B.4.1 Identification des sites de modification par une approche « top-down »	140
III.B.4.2 Détermination du pourcentage de modification	143
III.B.5 Comparaison des différents états structuraux des protéines par modification différentielle	145
III.B.5.1 Dénaturation et modification de la CK entière	146
III.B.5.2 Comparaison des peptides modifiés entre les deux formes	147
III.B.6 Conclusions	152
III.B.7 Bibliographie	153

III.C Structure de protéines étudiées par échanges hydrogène/deutérium.....157

III.C.1 Introduction	159
III.C.2 Résultats	164
III.C.2.1 Substance P	164
III.C.2.2 Myoglobine	168
III.C.3 Conclusions	171
III.C.4 Bibliographie	171

Partie IV : Oligomères de la protéine du prion ovine (PrP).....177

IV.1 Introduction	179
IV.2 Généralités sur les oligomères de la PrP	184
IV.3 Etude des oligomères intacts par spectrométrie de masse	186
IV.3.1 Introduction	186
IV.3.2 FT-ICR MS	187
IV.3.3 Q-TOF	188
IV.3.4 Conclusions sur les expériences de spectrométrie de masse en conditions non dénaturantes	191
IV.4 Etude structurale des oligomères par échanges H/D et spectrométrie de masse FT-ICR	191
IV.4.1 Introduction	191
IV.4.2 Analyses préalables	192
IV.4.3 Echanges H/D du monomère et des oligomères de la protéine prion ovine	194
IV.4.3.1 Résultats	194
IV.4.3.2 Conclusions	202
IV.4.4 Comparaison des résultats obtenus par échange H/D avec la structure de la PrP ^{Sc}	204
IV.4.5 Site d'interaction entre les sous-unités des oligomères	205
IV.5 Conclusions	207
IV.6 Bibliographie	208

Partie V : Conclusion générale.....213

Partie VI : Matériel et méthodes 217

VI.1	Instrumentation	219
VI.1.1	Ionisation par électronébulisation	219
VI.1.1.1	La formation des ions en electrospray	219
VI.1.1.1.1	La formation de gouttelettes chargées	220
VI.1.1.1.2	Evaporation du solvant et explosions coulombiennes	220
VI.1.1.2	Le nanospray	221
VI.1.2	Spectrométrie de masse FT-ICR	222
VI.1.3	Etude de complexes non covalents par spectrométrie de masse	225
VI.1.3.1	Conditions nécessaires à l'observation d'interactions non covalentes par spectrométrie de masse	226
VI.1.3.1.1	Préparation des échantillons	226
VI.1.3.1.2	Optimisation des paramètres instrumentaux	227
VI.1.3.1.3	Analyse des résultats	228
VI.2	Méthodes expérimentales générales	229
VI.2.1	Spectrométrie de masse	229
VI.2.2	Dessalage par microcolonnes C ₁₈	229
VI.3	Modification chimique par le DEPC	229
VI.3.1	Modification d'un mélange de peptides par le DEPC	230
VI.3.2	Modification de la myoglobine	230
VI.3.3	Modification de l'ubiquitine	230
VI.3.4	Modification de la créatine kinase	230
VI.4	Echanges hydrogène/deutérium	231
VI.4.1	Calcul des hydrogènes échangés	231
VI.4.2	Echange H/D substance P	232
VI.4.3	Echange H/D myoglobine	232
VI.5	Oligomères de la protéine prion	232
VI.5.1	Formation des oligomères	232
VI.5.2	Chromatographie d'exclusion stérique	233
VI.5.3	Purification des oligomères en TEAA et lyophilisation	233
VI.5.4	Echanges H/D sur le monomère et les oligomères de la protéine prion	233
VI.5.4.1	Remarques générales	233
VI.5.4.2	Méthode détaillée	234
VI.6	Bibliographie	235

Abréviations

Å : Ångström ($1\text{Å} = 10^{-10}\text{ m}$)

AA : Acétate d'ammonium

AB : Bicarbonate d'ammonium

ADN : Acide Désoxyribonucléique

ADP : Adénosine Diphosphate

ARN : Acide Ribonucléique

ATP : Adénosine Triphosphate

BIRD : Blackbody Infrared Radiation Dissociation (dissociation par rayonnement infra-rouge de corps noir)

CAD : Collisional Activated Dissociation (dissociation par activation collisionnelle)

CD : Circular Dichroism (dichroïsme circulaire)

CID : Collisional Induced Dissociation (dissociation induite par collision)

CJD : Creutzfeldt Jacob Disease (maladie de Creutzfeldt Jacob)

CK : Créatine kinase

CRM : Charge Residue Model (modèle des charges résiduels)

Da : Dalton (unité de mesure de masse moléculaire, équivalente au gramme/mole)

DEPC : Diéthylpyrocarbonate

DNA : voir ADN

DSC : Differential Scanning Calorimetry

ECD : Electron Capture Dissociation (dissociation par capture d'électrons)

ELISA : Enzyme Linked Immunosorbent Assay

ESI : Electrospray Ionization (ionisation par électronébulisation)

ESST : Encéphalopathies Subaigues Transmissibles

eV : électron volt, $1\text{ eV} = 1.602 \cdot 10^{-19}\text{ J}$

FAB : Fast Atom Bombardement (bombardement par des atomes rapides)

FID : Free Induction Decay (décroissance en induction libre)

FT : Fourier Transformation (transformée de Fourier)

FT-IR : Fourier Transform Infrared Spectroscopy (spectroscopie infra rouge à transformée de Fourier)

FTMS : Fourier Transform Mass Spectrometry (spectrométrie de masse à transformée de Fourier)

ICR : Ion Cyclotron Resonance (résonance cyclotronique ionique)

IC : Ionisation Chimique

IE : Ionisation Electronique

IEM : Ion evaporation mechanism (mécanisme d'évaporation d'ions)

IR : Infrarouge

ITC : Isothermal Titration Calorimetry

H/D : Hydrogène/Deutérium

HDX : Hydrogen/deuterium exchange (échanges hydrogène/deutérium)

IRMPD : Infrared Multiphoton Dissociation (dissociation par activation vibrationnelle)

LC : Liquid Chromatography (chromatographie liquide)

MALDI : Matrix-Assisted Laser Desorption Ionization (désorption/ionisation laser assistée par matrice)

MS : Mass Spectrometry (spectrométrie de masse)

MW : Molecular Weight (masse moléculaire)

m/z : rapport masse sur nombre de charges

NMR : voir RMN

pH : potentiel Hydrogène

pD : potentiel Deutérium

PrP : Protéine du prion

PrP^C : Protéine du prion, forme cellulaire

PrP^{Sc} : Protéine du prion, forme *Scrapie*

Q-TOF : Quadrupole-Time of Flight (quadrupôle-temps de vol)

rf : radiofréquence

RMN : Résonance Magnétique Nucléaire

RNA : voir ARN

SAXS : Small Angle X-ray Scattering (diffraction à rayons X aux petits angles)

SDS-PAGE : Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis (gel d'électrophorèse au polyacrylamide, en présence de dodécylsulfate de sodium)

SNC : Système Nerveux Central

SORI : Sustained Off-Resonance Irradiation (irradiation continue hors resonance)

SPR : Surface Plasmon Resonance (résonance de plasmon de surface)

TEAA : Acétate de Triéthylammonium

TEAB : Bicarbonate de Triéthylammonium

TEAF : Formiate de Triéthylammonium

Partie I

Introduction générale

Les protéines sont des molécules constitutives et essentielles des systèmes biologiques. L'étymologie de leur nom, qui est dérivé du grec *πρωτεος* (« *proteos* ») signifiant de première importance [1], souligne ce rôle central qu'elles occupent dans le monde du vivant. Elles interviennent à tous les niveaux, et on trouve parmi leurs fonctions la catalyse (enzymes), le transport de molécules (hémoglobine), les déplacements intracellulaires (fuseau de division), la structuration mécanique (collagène dans la peau ou les os), la défense immunitaire (anticorps), la communication nerveuse ou hormonale et le contrôle du développement et de la différenciation [2]. Au niveau moléculaire, à cette variété de fonctions correspond une multitude de structures, de concentrations, de localisation, et surtout d'interactions avec les divers constituants du milieu les environnant (autres protéines, ADN, ARN, métabolites, membranes lipidiques, sels, ...). De par cette multiplicité de fonctions et de structures, on pourrait penser qu'une étude individualisée des structures et fonctions de chacune des protéines ne rendrait pas compte du fonctionnement du système complexe qu'est la cellule ou l'organisme entier. De ce point de vue, une approche holistique du problème par le biais de la biologie des systèmes (« *systems biology* ») semblerait plus appropriée [3]. Toutefois, les protéines présentent en général une fonction bien spécifique dans la cellule et elles se prêtent par conséquent bien à une approche réductionniste dans laquelle le système complexe « global » est décomposé en éléments dont on cherche à comprendre les caractéristiques.

L'étude des protéines peut elle-même se décomposer en plusieurs niveaux : d'une vision intégrée où l'on considère simultanément les éléments structuraux et fonctionnels d'une protéine donnée à des décompositions spécifiques des caractéristiques de la molécule. L'approche du chimiste est de ce point de vue la plus réductrice : la protéine peut être considérée comme une simple molécule, assemblage d'atomes reliés par des liaisons. Au premier ordre, du point de vue chimique les liaisons les plus importantes sont les liaisons covalentes ; or il apparaît facilement que la diversité des types de liaisons covalentes qu'on retrouve chez les protéines est extrêmement faible : ces molécules sont composées de répétitions de 20 résidus (parfois modifiés) reliés par un seul type de liaison, la liaison peptidique. Au second ordre, le chimiste doit considérer l'action de liaisons dites « faibles » qu'on appelle également non covalentes, par opposition aux précédentes. Or bien que les énergies de liaisons individuelles mises en jeu soient nettement plus faibles ($5\text{-}40\text{ kJ mol}^{-1}$ par rapport à $200\text{-}800\text{ kJ mol}^{-1}$ pour les liaisons covalentes [4]), la diversité des fonctions chimiques (acides, bases, donneur/accepteur d'hydrogène, chaînes aliphatiques ou aromatiques) présentes dans les résidus d'une protéine en font des éléments essentiels d'une

compréhension des protéines. Ces liaisons interviennent de fait à tous les niveaux de la structure des protéines, du repliement de la protéine lors de sa synthèse et de sa maturation jusqu'à sa forme fonctionnelle. De même, elles interviennent dans sa fonction, que ce soit pour reconnaître et fixer un substrat ou un partenaire ou dans des changements structuraux conduisant à une fonction (par exemple le transport transmembranaire).

Jusqu'ici, on a dans ce texte regroupé sous la dénomination de liaisons non covalentes un certain nombre d'interactions de natures physico-chimique diverses. Il convient donc dans un premier temps de décrire la nature physico-chimique des principaux types de liaisons non covalentes observées pour des protéines. Dans un second temps, la décomposition usuelle des différents types de structures observées pour des protéines sera décrite, en s'attachant à rendre compte à chaque niveau du type de liaisons pouvant être mis en jeu et des différentes techniques expérimentales permettant d'avoir un accès à ces structures. Comme écrit plus haut, les interactions non covalentes interviennent également au-delà de la simple structuration des protéines dans leurs interactions avec des ligands, rejoignant ainsi l'aspect plus fonctionnel d'une protéine. Il convient ainsi de mettre à part, dans une troisième section de cette introduction ces interactions entre une protéine et un ou plusieurs ligands ainsi que les méthodes d'étude permettant d'y avoir accès. Finalement, dans une dernière section, l'utilisation de la spectrométrie de masse, la technique particulière utilisée dans cette thèse pour l'étude des interactions non covalentes dans les protéines sera présentée.

I.1. Interactions non covalentes dans les protéines

On peut subdiviser les interactions non covalentes en quatre grandes catégories [4] : dans la première, on trouve les interactions électrostatiques entre groupements portant des charges nettes positives et négatives, qui correspondent aux interactions non covalentes les plus fortes. On peut classer dans un second groupe, toutes les interactions électrostatiques de type dipolaire. Le troisième type de liaisons est à relier au précédent : c'est celui des liaisons hydrogènes. Finalement, on trouve dans une catégorie un peu à part les liaisons dites hydrophobes.

I.1.1. Interactions ioniques

L'interaction ionique entre deux groupements portant une charge nette positive ou négative (soit au sein d'une molécule soit entre deux molécules) est la plus forte des interactions non covalentes. Ces interactions peuvent conduire à un effet attractif (charges

opposées) comme répulsif (charges du même signe). L'énergie d'interaction peut atteindre quelques centaines de kJ par mole, proche de l'énergie d'une liaison covalente. L'origine de ces interactions est l'attraction ou la répulsion coulombienne, donnée par la loi de Coulomb (équation I.1) :

$$E = \frac{q_1 \cdot q_2}{4\pi \cdot \epsilon_0 \cdot \epsilon \cdot r} \quad (\text{eq. I.1})$$

où q_1 et q_2 sont les deux charges, r est la distance les séparant, ϵ_0 est la permittivité du vide et ϵ est la permittivité relative du milieu. Dans l'eau, ϵ vaut 80, ce qui conduit à une forte diminution de cette énergie d'interaction en solution par rapport au vide. Il est très important de prendre en compte cette variation de ϵ dans différents milieux pour l'analyse des interactions non covalentes de protéines par spectrométrie de masse : lors de l'étape initiale de désolvatation, les analytes passent d'un solvant aqueux ($\epsilon = 80$) au vide ($\epsilon = 1$). Les interactions coulombiennes seront donc fortement affectées par cette désolvatation.

Quels sont les groupes chargés au sein d'une protéine ? Dans une protéine, les extrémités N et C ainsi que les groupements chimiques de certaines chaînes latérales sont susceptibles d'être protonés ou déprotonés suivant le pH du milieu. Par exemple, à pH acide, les groupements « basiques » (Lys, Arg, His) seront généralement protonés et porteront donc une charge positive. A pH basique, les groupes acides (Glu, Asp) sont susceptibles d'être déprotonés et portent donc généralement une charge négative. A pH neutre, la plupart du temps proche du pH physiologique de la protéine, celle-ci présente simultanément des charges positives et négatives ouvrant la possibilité d'interactions ioniques intra- et inter-moléculaires.

I.1.2. Interactions de type dipolaire

Une molécule n'a pas besoin d'une charge nette pour participer à des interactions électrostatiques : la présence d'un moment dipolaire (Fig. I.1) permanent μ suffit à faire interagir les molécules polaires via des liaisons non covalentes.

Pour des protéines, plusieurs niveaux de dipôles peuvent exister : la liaison peptidique est un exemple de tel dipôle (Fig. I.2), mais l'assemblage parallèle de plusieurs dipôles dans une hélice alpha conduit à un macrodipôle de moment nettement plus important. Finalement la présence de charges de signes opposés dans une protéine définit également des dipôles via la distribution de charges [5].

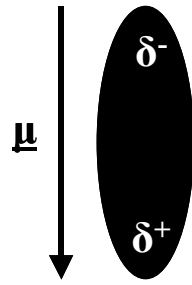


Fig. I.1 : Dipôle dans une molécule polaire.

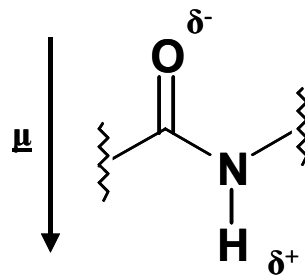


Fig. I.2 : Moment dipolaire de la liaison peptidique.

Un dipôle interagit avec son environnement électrostatique (ions, autres dipôles, quadripôles, dipôle induit, etc.) La force de ces interactions dépend surtout de la nature du partenaire et suit l'ordre ion > dipôle > quadripôle > dipôle induit. Les dipôles interagissent moins que les ions entre eux.

I.1.2.1. Interactions ion-dipôle

Dans un champ électrique, par exemple le champ généré par un ion, un dipôle va s'aligner en fonction de la charge de l'ion de façon à créer une force attractive entre l'ion et le dipôle (Fig. I.3).

Les interactions ion-dipôle, bien que moins fortes que des interactions ion-ion, sont à la fois importantes pour la structure intrinsèque d'une protéine (interaction électrostatiques entre des charges au sein d'une protéine) et pour l'interaction entre la protéine et le solvant (par exemple, l'interaction entre une chaîne latérale de la lysine chargée positivement et des molécules d'eau qui entourent la protéine).

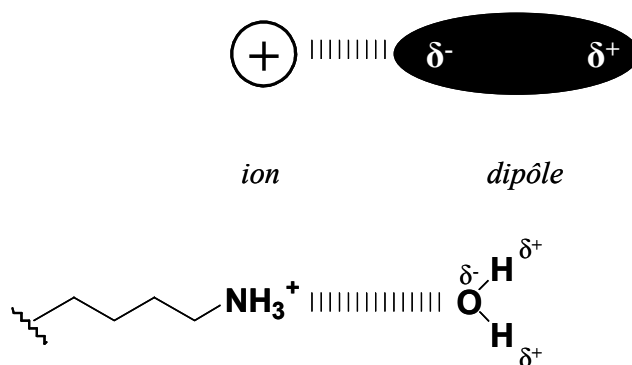


Fig. I.3 : Principe de l'interaction ion-dipôle (en haut) et exemple de l'interaction entre la chaîne latérale chargée positivement de la lysine et une molécule d'eau (en bas).

I.1.2.2. Interactions dipôle-dipôle

En l'absence d'un champ électrique extérieur, deux dipôles interagissent entre eux. Deux dipôles linéaires s'alignent soit tête-bêche (voir Fig. I.4) soit colinéaires, alignés de façon à maximiser les interactions attractives (maximum global ou local). Par contre un alignement parallèle côte à côte ou colinéaire opposé conduit à une forte répulsion.

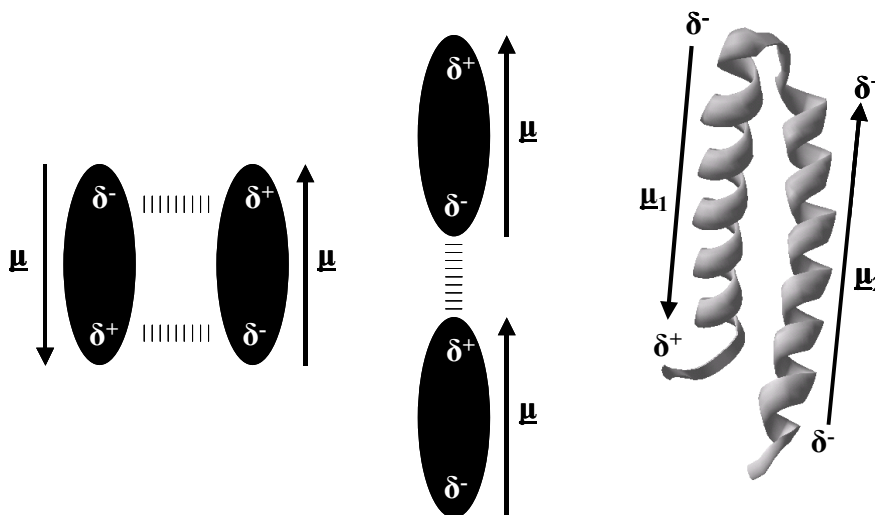


Fig. I.4 : Principe d'interaction entre deux dipôles. Interaction tête-bêche (à gauche), interaction colinéaire (au milieu) et exemple d'une interaction dipôle-dipôle au sein d'une protéine : deux hélices alpha anti-parallèles (à droite).

L'exemple des hélices alpha dans les structures de protéines illustre parfaitement les interactions dipolaires (voir Fig. I.4). Une hélice alpha est un élément structural secondaire

d'une protéine (voir I.2.2), où chaque spire de l'hélice est reliée à la suivante par l'alignement des moments dipolaires des liaisons peptidiques. Ceci conduit à la formation d'un macrodipôle, somme des moments dipolaires individuels de toutes les liaisons peptidiques de la chaîne [6]. Deux hélices peuvent elles-mêmes s'aligner de façon antiparallèle pour maximiser l'interaction entre macrodipôles.

Comme pour la section précédente, la présence d'un moment dipolaire dans l'eau conduit à des interactions de ce type entre la protéine et le solvant.

I.1.2.3. Dipôles induits

Il n'est pas nécessaire qu'une molécule possède une charge permanente ou un dipôle permanent pour qu'elle soit impliquée dans des interactions de type électrostatique : la présence d'un champ électrique (généré par une charge ou un dipôle extérieur) peut suffire à déplacer les électrons des orbitales d'une molécule. Ce déplacement se traduira par l'apparition d'un moment dipolaire dans la molécule en présence d'un champ électrique. La propriété physicochimique qui caractérise la facilité avec laquelle un dipôle peut être induit dans une molécule est la polarisabilité. L'interaction avec des dipôles induits est moins forte qu'avec des dipôles permanents puisqu'une partie de l'énergie d'interaction est utilisée pour l'induction du dipôle [6]. Il est important de noter qu'à la différence des charges et des dipôles permanents, les interactions impliquant un dipôle induit dépendent peu de l'orientation dans l'espace de la molécule dans laquelle le dipôle est induit ; ceci est très important lorsqu'on les compare aux interactions liées à un dipôle permanent qui sont extrêmement directionnelles comme montré ci-dessus. On subdivise ces types d'interactions suivant la nature de la charge extérieure générant le champ électrique.

I.1.2.3.1. Interaction ion-dipôle induit

Si on considère une molécule simple qui ne porte pas de charges et qui n'a pas de dipôle permanent (Fig. I.5, A), la distribution des électrons et des charges positives des noyaux est initialement homogène. La présence d'un cation à proximité de cette molécule va perturber la distribution des électrons : les électrons seront attirés par le cation. Sous l'effet de cette migration, une charge nette négative apparaîtra du côté proche du cation, tandis que l'appauvrissement en électrons du côté opposé se traduira par l'apparition d'une charge positive nette de ce côté. La présence du cation a donc créé un dipôle dans la molécule, qui interagit avec la charge de façon attractive comme un dipôle permanent.

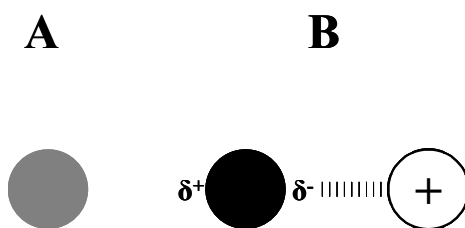


Fig. I.5 : Dans une molécule avec une distribution homogène des électrons (A), un dipôle induit est formé par la présence d'une charge externe (B).

I.1.2.3.2. Interaction dipôle-dipôle induit

Pour l'interaction entre un dipôle et un dipôle induit, si on considère qu'un dipôle peut s'exprimer sous la forme de deux charges permanentes situées à ses extrémités, on voit que le même raisonnement que pour le cas précédent peut être suivi. Bien entendu, comme l'interaction dipôle-dipôle est plus faible qu'une interaction dipôle-ion, l'énergie d'interaction est ici encore plus faible.

I.1.2.3.3. Interactions de van der Waals

Même en absence de groupes chargés (ions, dipôles permanents), tous les atomes et molécules interagissent entre eux. Cette interaction est le résultat d'interactions mutuelles entre des dipôles induits : si la distribution moyenne des électrons pour une molécule est homogène, la molécule présente un moment dipolaire instantané non nul puisque la distribution instantanée des électrons n'est pas homogène. Ces fluctuations instantanées dans deux molécules proches induisent mutuellement un dipôle dans chacune d'entre elles, dipôles alignés de façon antiparallèles pour maximiser l'interaction, dite interaction de van der Waals. Même si les forces d'interactions sont nettement plus faibles pour ces interactions de van der Waals que pour les autres, elles jouent un rôle important pour les biomolécules, en particulier lorsque les polarisabilités sont élevées (interactions entre groupes aromatiques). De même, la stabilité de la bicouche lipidique des membranes biologiques est assurée entre autres par l'interaction de van der Waals entre les chaînes hydrocarbonées des lipides.

I.1.3. Liaisons hydrogènes

La liaison hydrogène correspond à une liaison entre deux atomes électronégatifs, assurée par un atome d'hydrogène partagé entre les deux. Une liaison hydrogène peut être considérée comme un cas particulier de liaison électrostatique de type charge-dipôle, où une charge

partielle négative serait portée par l'accepteur A, et le dipôle constitué d'une charge partielle négative sur le donneur D et d'une charge partielle positive sur l'atome d'hydrogène. Ces éléments permettent d'expliquer la géométrie linéaire des liaisons hydrogènes, mais ne rendent pas compte du fait qu'au niveau orbitalaire, il y a réellement mise en commun des électrons de A et D pour former la liaison, ce qui lui donne un caractère plus proche de celui d'une liaison covalente.



Les distances et la force des liaisons hydrogènes dépendent de l'électronégativité de l'accepteur et du donneur : plus elle est élevée, plus les distances seront courtes et plus la force d'interaction sera grande. La force des liaisons hydrogènes est généralement de l'ordre de 8-40 kJ/mol à température ambiante [6].

Une seconde raison pour classer à part les liaisons hydrogènes est leur prévalence dans les liaisons non covalentes entre molécules biologiques et leur importance structurale. Dans les protéines, on les retrouve à la base de la structuration en hélices alpha et feuillets beta, par des interactions entre des groupes carbonyles (C=O, donneurs) et des N-H d'amide (accepteurs) (Fig. I.6). De nombreux autres groupes des chaînes latérales sont également susceptibles de servir de donneurs (-OH, -NH, plus rarement -SH) ou d'accepteurs (O, N et de façon moins importante S).

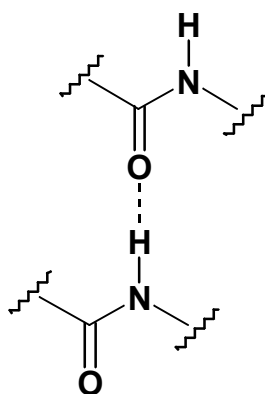


Fig. I.6 : Liaison hydrogène entre deux liaisons peptidiques.

I.1.4. Interactions hydrophobes

Dans une solution aqueuse, une molécule apolaire tend à interagir avec d'autres molécules apolaires plutôt que de s'entourer de molécules d'eau. On nomme cette propriété hydrophobicité et cela conduit à des interactions entre molécules qu'on appelle des interactions hydrophobes.

L'origine thermodynamique de l'interaction hydrophobe peut s'expliquer de la façon suivante. Considérons l'enthalpie libre d'une solution contenant de l'eau et des molécules non polaire (équation I.2) :

$$\Delta G = \Delta H - T\Delta S \quad (\text{eq. I.2})$$

Comme pour tout processus, il faut examiner les contributions enthalpiques (ΔH) et entropiques (ΔS) pour comprendre les éléments favorables ou défavorables à une interaction. Du point de vue enthalpique, l'énergie d'interaction est forte entre les molécules d'eau (réseau de liaisons hydrogènes) alors qu'elle est plus faible entre l'eau et les molécules apolaires, ce qui conduit à la faible solubilité des molécules apolaires dans l'eau. La présence d'une molécule apolaire dans de l'eau est donc défavorable enthalpiquement. Du point de vue entropique la situation est également défavorable : paradoxalement, la présence d'une molécule apolaire dans l'eau accroît la structuration de l'eau, conduisant à une diminution de l'entropie. On peut comprendre ce phénomène en observant que l'insertion d'une molécule apolaire dans de l'eau nécessite de créer une cavité dans le solvant en rompant des liaisons hydrogènes, d'insérer la molécule apolaire dans cette cavité et de réorganiser le réseau de liaison hydrogène pour contourner cet « obstacle ». Pour maximiser les liaisons hydrogènes dans cette dernière étape, les molécules d'eau doivent adopter une géométrie très précise pour envelopper la molécule apolaire. C'est ceci qui conduit à une structuration locale plus importante de l'eau et donc à une diminution de l'entropie.

Dans ces conditions, les deux termes, enthalpie et entropie, sont défavorables à la présence de molécules apolaires dans l'eau : par conséquent, l'eau pousse au regroupement de ces molécules apolaires entre elles, reliées par des liaisons de type van der Waals. Il est important de noter que la plus grande partie de l'énergie d'interaction est due non pas à ces interactions de van der Waals (qui sont faibles) mais aux contributions défavorables à la solvation de ces molécules apolaires dans l'eau. Donc en absence d'eau, comme par exemple dans le vide du spectromètre de masse, seule subsiste la part de l'énergie d'interaction due aux interactions de van der Waals.

Pour les protéines, ces interactions hydrophobes conduisent à un regroupement des résidus apolaires, généralement vers le centre de la protéine, ou encore à l'insertion de ces régions hydrophobes dans d'autres régions hydrophobes de la cellule comme dans les membranes lipidiques.

I.1.5. Conclusions

On constate dans cette partie que la majorité des interactions responsables de la structure des protéines sont d'origine électrostatique. On pourrait donc penser que la résolution des équations électrostatiques permettrait d'approcher simplement le problème de la structure des protéines. Mais il faut noter que ces différentes interactions sont extrêmement nombreuses et multiples dans une même protéine, et qu'une approche analytique est difficilement envisageable. De même, des approches par approximation, comme par exemple de considérer un solvant homogène via une constante diélectrique globale, sont difficiles à mettre en œuvre pour des protéines qui présentent très peu d'homogénéité par la diversité des résidus les composant. Il convient par conséquent d'aborder le problème des interactions non covalentes dans les protéines à un niveau plus élevé, tout en se basant sur la décomposition présentée ici.

I.2. Structure des protéines

Comme on l'a vu plus haut, la structuration des protéines est un élément essentiel pour comprendre leurs fonctions multiples, mais en même temps cette diversité est apportée par un nombre de briques de bases relativement pauvre : une protéine est composée d'un assemblage d'une vingtaine de types différents d'acides aminés (Fig. I.7), tous de stéréochimie L, reliés par une liaison peptidique.

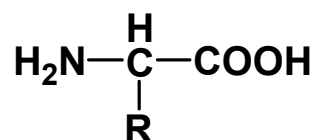


Fig. I.7 : Structure générale d'un acide α -aminé. La chaîne latérale (R) est l'élément variable qui en module les propriétés physicochimiques.

Les chaînes latérales des 20 acides aminés apportent des propriétés physicochimiques intrinsèques et jouent un rôle clé dans la structuration des protéines. Sans faire un exposé

exhaustif des propriétés des acides aminés, il est important de rappeler que les 20 acides aminés constituant des protéines peuvent être classés suivant leurs constitutions chimiques. Dans un premier groupe, on inclut les chaînes latérales aliphatiques : glycine (Gly, G)¹, alanine (Ala, A), valine (Val, V), leucine (Leu, L), isoleucine (Ile, I) et proline (Pro, P). La chaîne latérale de la proline, reliée à son groupe amine (Fig. I.8) confère une rigidité plus importante à la chaîne peptidique à proximité de ce résidu.

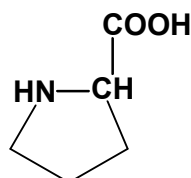


Fig. I.8 : Structure de la proline.

On peut classer dans un second groupe les acides aminés aromatiques : phénylalanine (Phe, F), tyrosine (Tyr, Y) et tryptophane (Trp, W). On distingue ensuite les acides aminés basiques (lysine [Lys, K], arginine [Arg, R], histidine [His, H]) et acides (acide aspartique [Asp, D], acide glutamique [Glu, E], asparagine [Asn, N], glutamine [Gln, Q]). Les deux derniers groupes consistent en les acides aminés hydroxylés (serine [Ser, S], thréonine [Thr, T]) et ceux qui contiennent du soufre (cystéine [Cys, C], méthionine [Met, M]).

Il est d'usage de séparer quatre niveaux structuraux lorsqu'on se réfère à la structure des protéines, progressant d'une vision linéaire de la molécule vers son assemblage tridimensionnel et à ses interactions avec d'autres molécules. Ces différents niveaux sont caractéristiques de la structure mais également de méthodes d'étude qui leur sont adaptées, ce qui permettra une présentation parallèle des éléments structuraux et des techniques expérimentales donnant accès aux structures.

I.2.1. Structure primaire

Le groupe α -carboxyl d'un acide aminé est susceptible de se lier avec le groupe α -amine d'un second acide aminé (Fig. I.9) pour former un amide CO-NH, formant un dipeptide, relié par ce qu'on appelle une liaison peptidique.

¹ Les abréviations entre parenthèses correspondent à la nomenclature en trois lettres et en une lettre des acides aminés qui seront utilisées dans la suite de ce manuscrit.

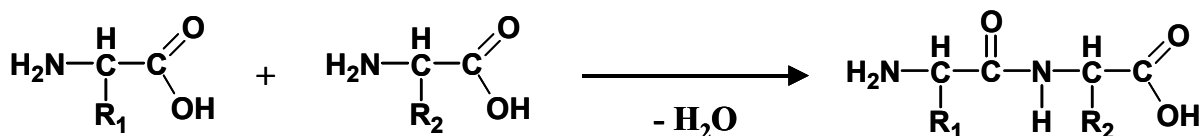


Schéma I.9 : Formation d'une liaison peptidique entre deux acides aminés.

Ce dipeptide peut à son tour former des liaisons peptidiques avec d'autres acides aminés, conduisant à des tripeptides, tetrapeptides, etc. jusqu'à une longue chaîne d'acides aminés, ce qui définit un polypeptide. La succession linéaire des acides aminés dans ce polypeptide constitue ce qu'on appelle la structure primaire de la protéine. Il est d'usage de représenter cette structure primaire par la séquence des acides aminés de l'amine libre à gauche vers l'acide carboxylique libre à droite. Par exemple, la description GFVAL correspond à l'enchaînement des acides aminés glycine-phénylalanine-valine-alanine-leucine.

On peut noter que l'enchaînement des acides aminés constitue quasiment le principal élément structural covalent dans une protéine. Une exception d'importance est celle des ponts disulfures : une liaison -S-S- peut en effet se former entre deux chaînes latérales de cystéine (Fig. I.10) :

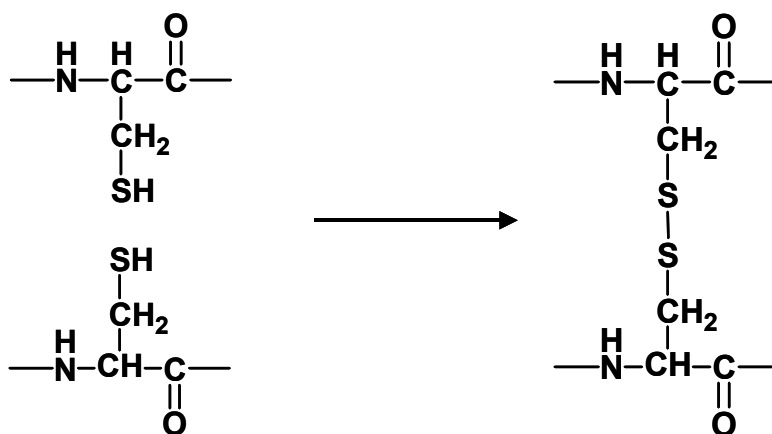


Fig. I.10 : Formation d'un pont disulfure entre deux cystéines.

D'un point vu chimique et physicochimique, la structure primaire d'une protéine est surtout caractérisée par une propriété : la séquence des acides aminés. De cette séquence peut être déduite sa masse moléculaire, caractéristique qui peut être étudiée en détail par spectrométrie de masse. Des informations plus détaillées sur l'étude de la structure primaire des protéines par spectrométrie de masse peuvent être trouvées dans [7].

I.2.2. Structure secondaire

I.2.2.1. Caractéristiques de la structure secondaire

Une chaîne polypeptidique ne se structure en général pas de façon aussi répétitive et régulière qu'une double hélice d'ADN, par exemple. Par contre on retrouve des motifs généraux qui sont présents dans presque toutes les protéines connues. Deux des motifs les plus importants de ce point de vue sont l'hélice α et le feuillet β .

Hélice α

Dans une hélice α , la chaîne polypeptidique présente un alignement hélicoïdal des acides aminés (Fig. I.11). L'hélice comporte 3,6 résidus par spire et le déplacement axial est de 1,50 Å par résidu, ce qui correspond à un écart de 5,41 Å entre les spires [6]. L'hélice est stabilisée par des liaisons hydrogènes du groupement carbonyle (C=O) du résidu n avec le groupement amide (N - H) du résidu $n+3$. Dans ces liaisons hydrogènes, la distance entre le N et le O est de 2,86 Å. De plus, les liaisons hydrogènes sont quasiment toutes parallèles à l'axe de l'hélice. L'effet additif de ces moments dipolaires conduit à un moment dipolaire important pour l'ensemble de l'hélice, l'extrémité N-terminal étant positive et l'extrémité C-terminal négative.

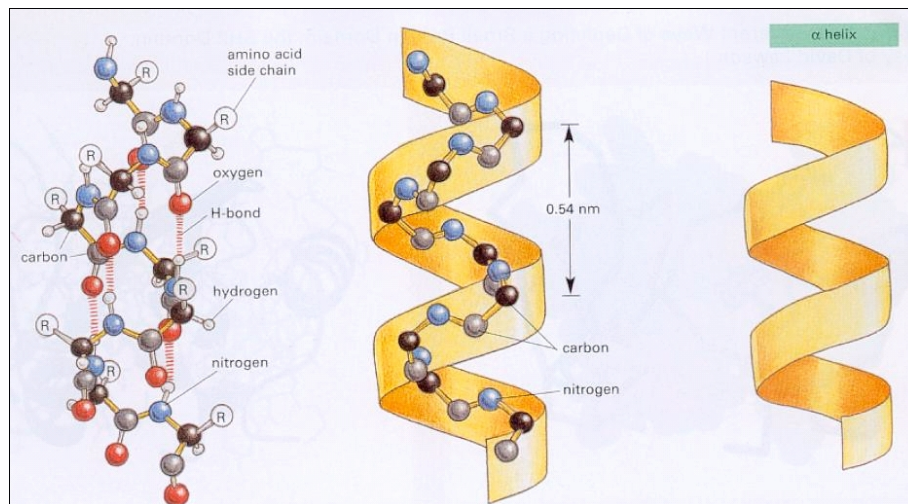


Fig. I.11 : Schéma d'une hélice α (copié de [8]).

Le contenu en hélice α des protéines peut varier entre 0 et 100%. Par exemple, la chymotrypsine n'a pratiquement pas d'hélice α , tandis que la myoglobine ou l'hémoglobine comportent presque 75% de leur structure sous forme d'hélices α [2].

Les hélices alpha sont très stables et peuvent être considérées comme des « cylindres rigides » qui participent au squelette d'une protéine. Elles imposent ainsi des contraintes sur la structure tridimensionnelle de la protéine (voir I.2.3).

Feuillet β

Le deuxième motif structural couramment observé dans les protéines est le feuillet β . Comme pour les hélices α , des liaisons hydrogènes entre des groupements carbonyles et des groupements amides de la liaison peptidique des différentes parties de la chaîne polypeptidique forment la base structurale de ce motif (Fig. I.12). Toutefois, les acides aminés entre lesquels les liaisons hydrogènes se forment ne sont généralement pas proches dans la structure primaire, car ce sont deux chaînes (ou plus) qui viennent se placer parallèlement l'une à l'autre. Dans les structures des protéines, les feuillets β sont représentés par des flèches (Fig. I.12). Les chaînes polypeptidiques qui constituent des feuillets β sont souvent alignées de façon anti-parallèle, bien qu'il existe aussi des exemples de feuillets β dans lesquels les différentes chaînes sont alignées de façon parallèle.

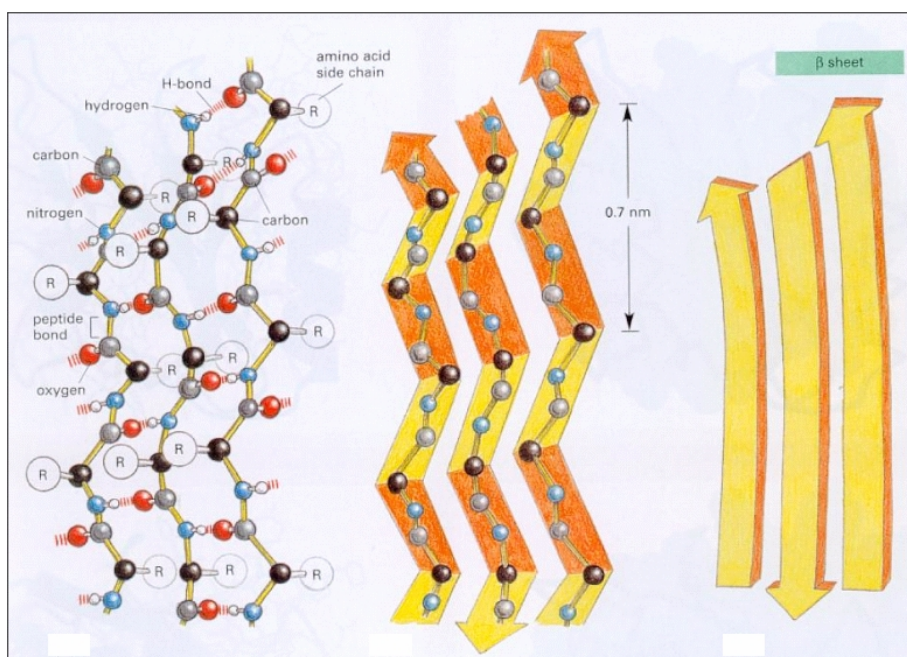


Fig I.12 : Schéma d'un feuillet β (copié de [8]).

Bien entendu, les hélices α et les feuillets β ne sont pas les seuls motifs structuraux secondaires, il existe une grande variété de motifs (β -turn, coiled coil, etc.) moins courants. Une description détaillée, qui dépasse les limites de cette thèse, peut être trouvée dans [9].

Comme les hélices alpha, des feuillets bêta, par leur rigidité, imposent des contraintes à la structure tridimensionnelle d'une protéine (voir I.2.3).

I.2.2.2. Méthodes spectroscopiques pour l'étude de la structure secondaire

Les protéines ont des bandes d'absorption dans des gammes de longueurs d'onde compatibles avec les méthodes spectroscopiques (IR, UV, fluorescence). L'absorbance d'un groupe chimique donné dépend fortement de son environnement, ce qui fait que les techniques spectroscopiques sont des méthodes de choix pour l'étude de la structure secondaire des protéines. On rappellera dans cette section les principes de quelques techniques d'études de la structure secondaire par des méthodes spectroscopiques : le dichroïsme circulaire, la spectroscopie d'absorption infrarouge (IR) et la fluorescence.

I.2.2.2.1. *Dichroïsme circulaire*

Le dichroïsme circulaire (CD) est la mesure de la différence d'absorbance pour des ondes polarisées circulaires droite et gauche (dans la gamme de 170-700 nm). C'est une caractéristique des molécules chirales. La plupart des molécules biologiques, y compris les protéines, possèdent des chromophores chiraux et les spectres CD correspondants contiennent donc des informations sur l'arrangement tridimensionnel. La spectroscopie CD est surtout utilisée afin de mesurer la teneur globale en éléments structuraux secondaires des peptides et des protéines (en sondant dans l'UV lointain, de 190-250 nm) et la chiralité liée à des arrangements spécifiques des chaînes aromatiques ainsi que des ponts disulfures dans une protéine (en sondant dans l'UV proche, de 250 à 300 nm).

Le chromophore le plus abondant dans les peptides et les protéines est la liaison peptidique, qui présente deux bandes d'absorption dans l'UV : une transition $\pi \rightarrow \pi^*$ à 190 nm et une transition $n \rightarrow \pi^*$ à 210 nm [10]. Les différentes formes de structures secondaires conduisent à des arrangements des liaisons peptidiques qui sont asymétriques et spécifiques. Les spectres CD dans l'UV lointain sont spécifiques et caractéristiques de différentes formes de structures secondaires. Des protéines qui contiennent surtout des hélices α (protéines $\alpha\alpha$) présentent des bandes fortes à 222 nm et 208 nm, qui sont absentes des spectres CD des protéines qui contiennent surtout des feuillets β . Ces dernières sont caractérisées par un minimum simple dans le spectre dont la position varie dans la gamme 210-230 nm. Par contre, les spectres CD ne permettent pas de différencier ces deux éléments structuraux dans des protéines qui les contiennent tous les deux. [11].

Dans l'UV proche, les chromophores présents dans les protéines sont essentiellement les groupes aromatiques (Phe, Tyr et Trp). Les spectres CD sont plus complexes et plus difficiles à analyser que dans l'UV lointain. Ils peuvent toutefois apporter une première idée qualitative de la structure tertiaire de la protéine, en précisant que les changements dans le spectre CD ne sont pas toujours corrélés à des changements importants de la structure tridimensionnelle de la protéine.

Les expériences de dichroïsme circulaire sont relativement simples et les résultats permettent d'obtenir une image globale de la structure secondaire de la protéine. Ils ne permettent par contre pas d'obtenir une structure détaillée et la technique du dichroïsme circulaire est souvent utilisée en complément d'autres expériences.

I.2.2.2.2. *Spectroscopie infrarouge (IR)*

La spectroscopie vibrationnelle, plus particulièrement la spectroscopie d'absorption infrarouge (IR), est également un outil puissant pour obtenir des informations sur la structure secondaire des protéines. Il convient de noter tout d'abord que le grand nombre de modes normaux présents dans une protéine conduit à une superposition de multiples bandes, ce qui rend les spectres extrêmement complexes. Mais une vision globale de l'allure des bandes permet néanmoins d'obtenir des informations sur les structures secondaires.

Plusieurs bandes d'absorption sont associées à la vibration des groupements spécifiques d'une protéine. La bande *amide I* (proche de 1650 cm^{-1}) est utilisée le plus souvent pour l'analyse de la structure secondaire [12]. Dans la majorité des protéines, on observe une bande amide I large, qui est une superposition des bandes amides I des différents éléments structuraux secondaires. La forme caractéristique du spectre IR dépend de la structure secondaire d'une protéine (hélice α versus feuille β), de façon analogue au dichroïsme circulaire. Les bandes *amides II* (vers 1550 cm^{-1}) et *amides III* ($1400\text{-}1200\text{ cm}^{-1}$) dépendent également de la structure secondaire, mais la corrélation entre la structure et la fréquence est moins nette que pour la bande amide I.

La spectroscopie IR présente également l'avantage de donner accès à des résolutions temporelles de l'ordre de la microseconde [13]. On peut ainsi étudier des catalyses enzymatiques, des réactions de transfert de charge dans des systèmes redox [14] ou encore des transferts de proton dans des protéines [15]. L'inconvénient majeur de la technique est l'interférence avec l'absorption IR de l'eau dans le même domaine, qui rend les solutions aqueuses de protéines difficiles à étudier.

1.2.2.2.3. Spectroscopie de fluorescence

La fluorescence est une autre technique spectroscopique classique d'étude du comportement de macromolécules en solution. Les trois acides aminés aromatiques (Phe, Tyr et Trp) présentent une fluorescence intrinsèque et peuvent être utilisés comme sondes intramoléculaires de la conformation, de la dynamique et des interactions intermoléculaires d'une protéine. Le tryptophane (Trp) est le chromophore le plus utilisé : la fluorescence de l'indole est très sensible à l'environnement moléculaire (absorption entre 275 et 295 nm, émission entre 320 et 350 nm), ce qui en fait un bon marqueur pour l'étude des modifications conformationnelles ou des interactions d'une protéine avec d'autres molécules [16]. En théorie, si la structure de la protéine était connue, les modifications de fluorescence du tryptophane pourraient apporter des éléments structuraux à une résolution atomique [16]. En pratique, le spectre de fluorescence d'une protéine est le résultat d'une multitude de contributions qui ne peuvent être modélisées aussi précisément.

Dans sa forme la plus simple, la fluorescence permet une mesure de la structure par l'effet de quench de la fluorescence par le solvant. Pour une protéine complètement dénaturée, les résidus tryptophanes sont exposés au solvant et la fluorescence de l'indole est réduite par des collisions avec les molécules du solvant. Par contre, dans une protéine repliée, le noyau indole se retrouvera généralement dans un noyau hydrophobe avec les autres résidus aromatiques. Dans ces conditions, l'intensité de la fluorescence augmente considérablement et de plus, du fait de la constante diélectrique réduite dans la région repliée, le maximum de fluorescence est déplacé [17]. De ce fait, le spectre d'émission en fluorescence apporte des informations sur les éléments structuraux secondaires comportant des résidus aromatiques.

On peut noter un second usage de la fluorescence, la technique dite du FRET, pour « Fluorescent resonance energy transfer » qui permet, en introduisant deux fluorophores susceptibles de transférer l'énergie absorbée de l'un vers l'autre, de mesurer la distance entre ces deux fluorophores [18]. Il convient toutefois de noter que de tels couples sont rarement présents dans l'état natif d'une protéine et qu'il est nécessaire d'ajouter biologiquement ou chimiquement ces chromophores aux protéines à étudier.

Grâce à sa bonne sensibilité (de l'ordre d'une micromole pour la plupart des configurations), la fluorescence peut être utilisée pour l'étude de la structure et du comportement de macromolécules en solution. Les concentrations nécessaires sont de l'ordre de quelques μM . Un des développements les plus récents est l'émergence de la spectroscopie sur molécule unique (acronyme anglais : « Single molecule spectroscopy »). Par cette méthode, des biopolymères individuels sont étudiés dans des conditions physiologiques [19],

ce qui permet d'obtenir des informations uniques sur des sous-populations dans un échantillon hétérogène [20]. L'inconvénient principal de la fluorescence est la nécessité d'avoir des chromophores adaptés, ce qui la limite à des parties de la protéine comportant généralement un tryptophane ou nécessite l'ajout d'un tel chromophore sur les protéines à étudier.

I.2.3. Structure tertiaire

I.2.3.1. Caractéristiques de la structure tertiaire des protéines

La structure tertiaire d'une protéine correspond à sa structure tridimensionnelle d'ensemble, assemblage dans l'espace des hélices α , feuillets β et autres éléments de la structure secondaire. Dans des conditions physicochimiques définies (par exemple, à l'intérieur d'une cellule), une protéine a normalement une structure tridimensionnelle fixée. Il faut noter que cette structure (en solution) n'est pas totalement rigide : la chaîne polypeptidique subit des mouvements thermiques et est susceptible de se déformer lors d'interactions avec d'autres molécules (voir I.3).

Quelles sont les bases énergétiques qui conduisent une protéine à prendre une structure tridimensionnelle bien définie ? Comme on l'a vu plus haut, en dehors des ponts disulfures qui jouent également un rôle dans la structure tridimensionnelle des protéines, la plupart de ces interactions sont de nature non covalentes. La plupart des types d'interactions non covalentes peut y jouer un rôle. En théorie, la simple connaissance de la structure primaire apportant la nature des acides aminés et leur enchaînement pourrait permettre de calculer ces interactions et de prédire ces structures. En pratique, bien que les progrès dans ce domaine soient constants [21], la détermination théorique *ab initio* d'une structure de protéine sans références expérimentales reste encore impossible de par la complexité du système.

Les interactions non covalentes permettent toutefois d'expliquer quelques règles générales sur le repliement des protéines [2]. C'est ainsi que « l'intérieur » d'une protéine (c'est-à-dire la partie qui n'est pas accessible au solvant) consiste presque entièrement en résidus apolaires (les acides aminés aliphatiques et aromatiques). En revanche, la « surface » d'une protéine (c'est-à-dire les parties en contact avec le solvant) est constituée essentiellement des résidus polaires ou chargés, comme des lysines et des arginines. Ce partage est une conséquence des interactions hydrophobes (voir I.1.4) qui conduisent les résidus ayant des chaînes latérales apolaires à interagir entre eux plutôt qu'avec le solvant. En revanche, les résidus polaires peuvent former facilement des liaisons hydrogènes avec le solvant et sont donc plus

susceptibles d'être localisés à la surface d'une protéine. La formation des liaisons hydrogènes intramoléculaires contribue essentiellement à la formation des éléments structuraux secondaires (voir I.2.2.1). La plupart des protéines présentent une structure plutôt globulaire, globalement sphérique [6]. On doit noter que cette structure tridimensionnelle est relativement stable et évolue seulement progressivement en fonction des conditions du milieu (pH, température, force ionique, etc.). Toutefois, au-delà d'une certaine limite, la structure est brutalement altérée : la protéine est dénaturée et perd pratiquement toujours sa fonction.

I.2.3.2. Méthodes classiques pour l'étude de la structure tertiaire des protéines

Pour comprendre l'architecture d'une protéine en détail, l'idéal est de pouvoir déterminer la position exacte de chaque atome de sa structure. Deux méthodes sont particulièrement adaptées à cela : la cristallographie aux rayons X et la résonance magnétique nucléaire (RMN).

I.2.3.2.1. *Cristallographie aux rayons X*

Une protéine est trop petite pour être observée directement par microscopie optique puisque sa taille ne suffit pas à diffuser de la lumière dans les longueurs d'onde du « visible ». En revanche, les distances interatomiques dans une protéine (environ 1,5 Å) sont du même ordre de grandeur que la longueur d'onde des rayons X. La diffusion de rayons X par les atomes d'une protéine est donc possible. Dans un cristal régulier, les protéines forment un arrangement périodique, suivant des axes de l'espace tridimensionnel, de cellules unitaires définies comme étant le plus petit élément représentant le cristal entier. Des translations suivant les axes d'une cellule unitaire permettent de reconstituer l'ensemble du cristal. Chaque atome d'une protéine a donc exactement la même position dans toutes les cellules unitaires. Le cristal est ainsi composé de plans d'atomes. Ces plans d'atomes, ou plutôt les nuages d'électrons correspondants, produisent un motif de diffraction lors de l'irradiation du cristal avec des rayons X [22]. Le motif de diffraction contient les informations sur la position de tous les atomes de la cellule unitaire. Il est donc possible par l'analyse de ce motif de calculer les angles de diffraction ainsi que les distances interatomiques pour reconstituer la structure moléculaire.

L'avantage principal de cette technique est sa haute résolution. Bien que la procédure de traitement des données soit relativement complexe, la cristallographie aux rayons X permet d'obtenir des informations très précises sur la structure moléculaire des protéines, souvent à résolution atomique ($\leq 1,2$ Å) [23]. A des résolutions encore plus élevées ($\leq 0,8$ Å), les

atomes d'hydrogène deviennent visibles et les atomes ne peuvent plus être considérés comme des sphères à cause de la déformation des nuages électroniques. Une telle visualisation des électrons permet de déterminer la distribution des charges et la polarisation des liaisons [24]. On obtient ainsi une image très détaillée de la structure d'une protéine.

Toutefois, la technique présente des inconvénients. En premier lieu, il est souvent difficile d'obtenir des cristaux avec une taille suffisante pour la cristallographie. Il est donc nécessaire de produire des quantités non négligeables de protéine pour trouver des conditions optimales de cristallisation. Il faut également disposer de plusieurs cristaux identiques car ils sont souvent dégradés en présence des rayons X (par chauffage ou par des radicaux libres). On peut s'interroger sur la possibilité d'obtenir des informations dynamiques à partir d'une structure forcément statique (le cristal). La réponse dépend surtout de la nature de la dynamique en question : s'il s'agit d'une dynamique sans mouvements à grande échelle, il est possible de caractériser des événements dynamiques. Par exemple, il est possible d'étudier une voie métabolique en piégeant des états intermédiaires. On peut également soit piéger des espèces non stables [25], soit utiliser des méthodes à résolution temporelle [26]. En revanche, les mouvements à grande échelle sont pratiquement toujours bloqués dans le cristal.

1.2.3.2.2. Spectroscopie RMN

La spectroscopie RMN est la deuxième méthode principale pour l'étude de la structure tridimensionnelle des protéines. Elle est basée sur l'absorption de la radiation à fréquence radio par des noyaux atomiques. Les niveaux d'énergie de spin des noyaux sont séparés dans un champ magnétique (effet Zeeman). Sous l'effet d'une radiofréquence, chaque noyau d'une molécule va résonner à une fréquence spécifique (fréquence de Larmor). Chaque noyau d'une protéine a sa propre fréquence de Larmor, qui dépend de l'environnement local du noyau (par exemple protection par des champs magnétiques locaux). Un spectre RMN contient ainsi des pics correspondants aux résonances de tous les noyaux de l'échantillon, chacun avec une fréquence (caractérisée par le déplacement chimique δ) caractéristique de son environnement chimique. De plus, la surface sous un pic est proportionnelle au nombre des noyaux résonnant à cette fréquence, ce qui permet de quantifier les résidus identifiés et caractérisés.

Un facteur très important permettant la détermination de la structure des protéines est la corrélation entre des noyaux d'une molécule. La perturbation locale des nuages d'électrons par le spin des noyaux conduit à un couplage scalaire entre des noyaux qui sont liés de façon covalente avec un maximum de quatre liaisons entre eux. L'origine de ce couplage est le changement des niveaux d'énergie des états de spin des noyaux en question pour conduire à

des nouveaux niveaux d'énergie, en fonction de l'alignement des spins des noyaux (parallèle ou anti-parallèle). Le nombre et la nature des noyaux voisins produit des motifs et des constantes de couplage spécifiques qui permettent d'attribuer une structure covalente. Dans les protéines, chaque chaîne latérale d'un acide aminé présente un motif de couplage scalaire caractéristique, ce qui est appréciable pour l'attribution des signaux RMN aux structures d'une protéine.

Un phénomène très important pour la détermination de la structure tridimensionnelle d'une protéine est le couplage « à travers l'espace ». Un noyau qui a été excité va se relaxer vers l'état fondamental par différents mécanismes. S'il y a d'autres noyaux, possédant un spin, suffisamment proches ($< 5 \text{ \AA}$ pour ^1H), un couplage dipolaire entre les deux noyaux permet la relaxation vers l'état fondamental. Ce couplage est connu sous le terme d'effet Overhauser nucléaire (nOe). Cet effet dépend de la distance r entre les deux noyaux et est proportionnel à $1/r^6$. Les nOe permettent donc de mesurer des distances nécessaires à la détermination de la structure d'une protéine.

En plus de la résolution spatiale qu'apporte les nOe, la spectroscopie RMN permet de travailler avec des protéines en solution et donc d'obtenir directement des informations sur la dynamique de la protéine en question. Un premier inconvénient de la méthode est la gamme de masse limitée : la limite supérieure pour l'attribution complète d'une protéine est autour de 30 kDa, à cause de la superposition des résonances et des temps de relaxation défavorables pour les grandes biomolécules. Un second inconvénient est la concentration relativement élevée : des concentrations de l'ordre du millimolaire sont généralement utilisées. Cela rend naturellement cette technique très consommatrice en échantillon, mais peut aussi dans certains cas induire des perturbations de la structure quaternaire des protéines (agrégation ou précipitation de certaines protéines)

I.2.4. Structure quaternaire

I.2.4.1. Caractéristiques de la structure quaternaire des protéines

Certaines protéines sont constituées d'un assemblage de plusieurs chaînes polypeptidiques. On parle alors pour chaque chaîne d'une sous-unité de la protéine. Dans ces cas, chaque sous-unité possède sa structure tertiaire propre et on parle de structure quaternaire pour la structure d'ensemble. Quelques exemples de protéines possédant une structure quaternaire sont l'hémoglobine, qui contient quatre sous-unités, et le ribosome, qui comporte un grand nombre de protéines ainsi que des ARN. Ces assemblages peuvent être parfaitement

bien définis, mais peuvent aussi être transitoires tout en étant fonctionnellement importants. En effet, dans l'environnement cellulaire, les protéines seront rarement isolées, et ces interactions avec d'autres partenaires sont essentielles à leur fonction. Dans d'autres cas, comme dans les macroassemblages amyloïdes, la formation d'un complexe entre des protéines peut conduire à la formation d'assemblages potentiellement pathogènes. Ces exemples soulignent l'importance de l'étude des structures quaternaires des protéines ainsi que des macroassemblages.

I.2.4.2. Méthodes classiques pour l'étude de la structure quaternaire des protéines et des macroassemblages

La grande diversité et la grande complexité structurale de ces macroassemblages constitue un véritable défi analytique. Les techniques classiques pour étudier ces aspects de la biologie structurale des protéines sont d'une part la chromatographie d'exclusion stérique, et d'autre part des méthodes basées sur la diffraction des rayons X (diffraction aux petits angles) et la microscopie électronique.

I.2.4.2.1. Chromatographie d'exclusion stérique

La chromatographie d'exclusion stérique permet de séparer des macroassemblages suivant leur taille (ou plutôt leur volume hydrodynamique) à la différence des autres méthodes chromatographiques qui utilisent plutôt l'affinité chimique avec un support. Les colonnes sont remplies d'un matériau caractérisé par une distribution de pores moléculaires de tailles différentes. Les macromolécules de l'échantillon entrent et sortent de ces pores selon leur taille. Les molécules les plus petites peuvent entrer dans un plus grand nombre de pores et sont donc retardées alors que les plus grosses éluent plus rapidement. En conséquence, les macromolécules sont séparées selon leurs dimensions.

Néanmoins, il faut noter que le volume hydrodynamique est différent pour deux macromolécules de même masse moléculaire mais d'architecture différente : une macromolécule peu structurée aura un plus grand volume hydrodynamique qu'une macromolécule de même masse moléculaire mais de structure très globulaire. La première sera donc plus rapidement éluée que la deuxième, bien que les deux masses soient identiques.

I.2.4.2.2. Diffraction aux petits angles (SAXS)

La diffraction aux petits angles (acronyme anglais : « Small angle X-ray scattering », SAXS) est une technique qui repose sur la diffraction élastique des rayons X par des protéines

et des macroassemblages dans la gamme 10-1000 Å [27]. Bien qu'il ne soit pas possible d'obtenir une résolution atomique par cette technique, elle apporte des informations structurales précieuses à basse résolution (forme globale de la protéine, structure tertiaire et quaternaire).

L'hétérogénéité fréquente des échantillons (coexistence de plusieurs espèces de protéine ou inhomogénéité conformationnelle) représente un défi majeur pour cette technique. Dans ce cas, en effet, un profil de diffusion contient des contributions de tous les composants. Cette difficulté peut être surmontée par des techniques de traitement des données afin de caractériser des conformères individuels de protéines [28] ou bien des agrégats de protéines [29].

I.2.4.2.3. Microscopie électronique

La microscopie électronique est une technique d'imagerie qui permet d'obtenir des informations sur des complexes macromoléculaires à basse résolution. Au départ, on n'obtient qu'une image bidimensionnelle de l'assemblage posé sur un support. Toutefois, la structure tridimensionnelle peut être reconstruite en considérant de multiples projections bidimensionnelles sous des angles différents [30]. Cette méthode permet d'étudier des macroassemblages ou des protéines entières, à condition que la masse moléculaire soit au moins quelques centaines de kDa [31]. Il est possible d'obtenir une résolution de l'ordre de 10-30 Å [32]. Pour des macroassemblages de protéines, on peut envisager de compléter une structure à basse résolution avec des structures à haute résolution des protéines individuelles (obtenues, par exemple, par cristallographie aux rayons X). Il est ainsi possible d'obtenir des informations structurales à une résolution presque atomique de structures sub-cellulaires [33]. Cette technique permet l'étude de différents états conformationnels d'assemblages macromoléculaires par imagerie de complexes « piégés » à différents états du changement conformationnel [31, 34]. La microscopie électronique permet également d'obtenir des informations sur des macromolécules et leurs assemblages dans le contexte cellulaire par imagerie de la structure biologique entière [35].

Quelques exemples d'études de complexes macromoléculaires par cette technique comprennent la structure du ribosome à une résolution de 11,5 Å [36], de virus [37] et de l'environnement cytosolique total de cellules eukaryotes [38] ainsi que l'étude d'une interaction protéine chaperone-substrat [39, 40].

I.3. Interactions protéine-ligand

Dans la section précédente, l'importance des interactions non covalentes dans la mise en place de la structure des protéines a été présentée. La fin de la section précédente montrait que le champ des interactions non covalentes allait bien au-delà de la structure des protéines elles-mêmes pour inclure également les interactions des protéines avec d'autres éléments présents dans la cellule. Or ce rôle est très important d'un point de vue biochimique car presque toutes les transformations chimiques au sein d'un système biologique sont effectuées par des enzymes et passent généralement par une étape de reconnaissance du substrat. Sans faire un exposé global du domaine de l'enzymologie, on s'intéressera à l'aspect thermodynamique des premières étapes d'une catalyse enzymatique, considérées du point de vue de la fixation d'un ligand sur une protéine. Ensuite, différentes méthodes classiques pour l'étude des interactions non covalentes entre une protéine et un ligand seront présentées, avant de passer à l'objet de cette thèse, leur étude par spectrométrie de masse.

I.3.1. Introduction aux interactions non covalentes protéine-ligand

Les propriétés biologiques d'une protéine dépendent de ses interactions avec d'autres molécules. Par exemple, des anticorps se fixent sur des protéines ou des peptides qu'ils reconnaissent très spécifiquement et qui peuvent provenir de virus ou de bactéries, l'hexokinase fixe du glucose et de l'ATP pour catalyser une réaction entre les deux, et les molécules d'actine se fixent entre elles pour s'assembler en filaments. Toutes les protéines interagissent, ou forment des liaisons, avec d'autres molécules. Dans certains cas, ces liaisons sont très solides, dans d'autres, elles sont faibles ou de courte durée. Toutefois, l'interaction présente toujours une grande spécificité dans le sens où une protéine ne peut normalement fixer que quelques molécules parmi les milliers, de types différents, présentes autour d'elle dans la cellule. On appelle la molécule fixée à la protéine (ion, petite molécule ou bien une autre macromolécule) un ligand de cette protéine.

La capacité d'une protéine à fixer avec sélectivité et avec une grande affinité ses ligands repose sur la formation d'interactions non covalentes, qui, compte tenu de la faiblesse individuelle des interactions, doivent être en nombre suffisant. Du point de vue structural, cet objectif peut être atteint par la complémentarité des surfaces et des sites d'interaction des deux entités (Fig. I.13). La spécificité sera alors assurée à la fois par une structure particulière et par un certain nombre de liaisons non covalentes qui n'ont pas nécessairement de spécificité propre.

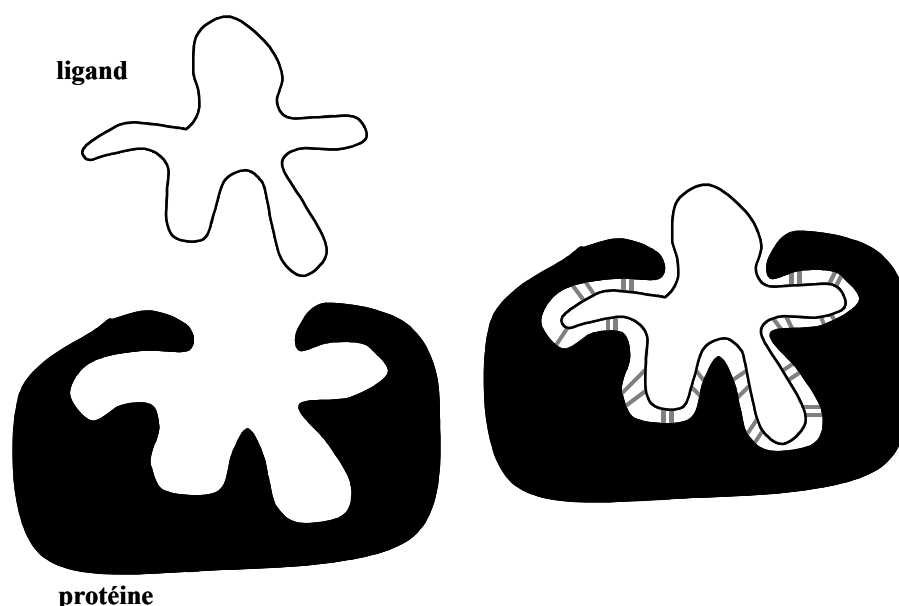


Fig. I.13 : La complémentarité des surfaces moléculaires entre une protéine (noir) et son ligand (blanc) permet la formation d'un grand nombre d'interactions non covalentes (symbolisées en gris)

La région d'une protéine qui fixe un ligand est appelée le site d'interaction. Il peut s'agir d'une cavité dans la surface de la protéine qui contient un arrangement séquentiel et/ou structural particulier d'acides aminés, ou d'un domaine exposé correspondant à un enchaînement très spécifique d'acides aminés, reconnaissant avec une très grande spécificité une autre séquence particulière. Une même protéine peut présenter plusieurs sites d'interaction de types différents, pour plusieurs ligands différents.

Dans une cellule, les molécules subissent des mouvements thermiques aléatoires, conduisant statistiquement à un grand nombre de rencontres entre elles. Si deux molécules qui s'approchent ont des surfaces peu compatibles, peu de liaisons non covalentes se forment. Par conséquent, le complexe entre ces deux molécules est susceptible de se dissocier rapidement. En revanche, la formation d'un grand nombre de liaisons non covalentes, conduit à une plus grande stabilité du complexe, d'où une durée de vie plus longue. On constate la présence d'interactions d'autant plus fortes que la fonction biologique du complexe nécessite une interaction longue. Un exemple en est le ribosome, qui est un complexe non covalent ayant une très longue durée de vie entre des molécules d'ARN et des protéines.

La force de l'interaction entre une protéine (P) et son ligand (L) est caractérisée par la constante de dissociation (K_d) ou par la constante d'association (K_a), reliées à l'énergie d'interaction par la loi d'action de masse ($\Delta G_a = -RT \ln K_a$) :

$$K_d = \frac{1}{K_a} = \frac{[P] \cdot [L]}{[PL]} \quad (\text{eq. I.3})$$

où [PL] est la concentration d'équilibre du complexe protéine-ligand et [P] et [L] correspondent aux concentrations de la protéine et du ligand dans la solution à l'équilibre. Comme pour tout équilibre chimique, il s'agit ici d'un équilibre dynamique, où le nombre d'évènements d'association est égal au nombre de dissociations.

La plupart des protéines exercent leurs fonctions par la fixation d'un ou de plusieurs ligands. Par exemple, une molécule d'actine structure la cellule en formant des filaments par interaction avec d'autres molécules d'actine. Pour d'autres protéines, la fixation d'un ligand n'est que la première étape nécessaire à leur fonctionnement, par exemple pour les enzymes. Les enzymes accomplissent toutes des transformations chimiques qui impliquent la dissociation et la formation de liaisons covalentes au sein d'un système biologique. Elles fixent des ligands (*substrats*) et les transforment chimiquement en *produits*. Les enzymes accélèrent considérablement des réactions chimiques (par des facteurs 10^6 ou plus) sans être modifiées, c'est-à-dire qu'elles agissent comme des *catalyseurs*.

Une cellule contient des milliers d'enzymes, la plupart d'entre elles actives en parallèle. Par leurs activités catalytiques, les enzymes forment un réseau très complexe de voies métaboliques. Chacune de ces voies est composée d'une succession de réactions chimiques. La complexité du système nécessite des contrôles afin de le réguler. Une des régulations les plus importantes est l'inhibition par fixation d'un ligand appelé *inhibiteur*. Cette inhibition est une régulation négative : l'activité de l'enzyme est réduite en présence d'inhibiteur. Symétriquement, les enzymes peuvent également subir une régulation positive, c'est-à-dire que leur activité peut être accrue par la fixation d'un ligand *activateur*. La régulation par un inhibiteur peut s'effectuer de deux façon : soit en se fixant dans le site de fixation du substrat (inhibiteur compétitif), ce qui a pour conséquence de diminuer la concentration en protéine libre, soit en se fixant sur un autre site d'interaction conduisant à un changement structural de la protéine.

Finalement, il faut noter qu'une protéine (souvent multimérique) peut présenter plusieurs sites d'interaction pour un même ligand. Dans ces situations, plusieurs cas sont à envisager suivant l'équivalence et l'indépendance des sites d'interactions. En particulier, la fixation simultanée de ligands peut être favorisée (coopérativité positive des sites), défavorisée (inhibition) ou les différents sites peuvent rester indépendants.

I.3.2. Méthodes d'études des interactions protéine-ligand

Compte tenu de l'importance des interactions entre les protéines et leurs ligands au sein des systèmes biologiques, il est incontournable de les étudier afin d'obtenir des informations sur le fonctionnement de ces systèmes. Dans ce contexte, se posent trois questions importantes :

1. Quelle est la stœchiométrie du complexe en question ?
2. Quelle est la valeur de la constante d'équilibre pour l'interaction ?
3. Pour des protéines avec des sites multiples : les sites sont-ils coopératifs ?

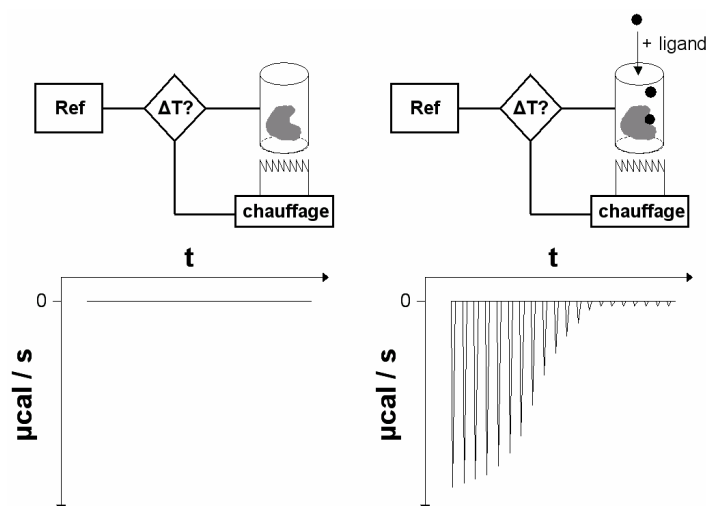
Pour pouvoir répondre à ces questions, il existe plusieurs méthodes, en dehors de la spectrométrie de masse. Les méthodes les plus importantes et les plus fréquemment utilisées sont présentées et discutées dans la suite de cette partie : les méthodes thermochimiques (ITC, DSC), les méthodes spectroscopiques (fluorescence, SPR, RMN) et les méthodes d'équilibre (EC, dialyse à l'équilibre).

I.3.2.1. Méthodes thermochimiques

I.3.2.1.1. Analyse calorimétrique par dosage isotherme (ITC)

L'ITC est une technique permettant la mesure directe des constantes d'interaction par une méthode thermochimique. L'interaction entre une protéine et un ligand est soit exothermique, soit endothermique. L'ajout d'un ligand dans une solution contenant la protéine libre va ainsi soit libérer soit absorber de la chaleur. Cette quantité de chaleur peut être mesurée par un calorimètre et donne accès directement à l'enthalpie libre d'association, qui est liée à la constante d'interaction (Fig. I.14).

Cette technique étant un dosage, elle donne également directement le nombre de sites d'interaction et une mesure de la coopérativité le cas échéant. Son inconvénient majeur est sa faible sensibilité : elle nécessite des concentrations de l'ordre de 10 mM, ce qui correspond à environ 1 mg d'une protéine moyenne.



Schema I.14 : Schéma du principe de l'ITC en haut. La partie du bas représente la mesure du chauffage à apporter lors de l'ajout d'une solution contenant le ligand. C'est à partir de ces données calorimétriques que sont calculés les paramètres thermochimiques de la réaction.

I.3.2.1.2. Analyse calorimétrique différentielle (DSC)

L'analyse calorimétrique différentielle (acronyme anglais : differential scanning calorimetry, DSC) est une technique permettant d'étudier les transitions structurales induites thermiquement. Elle sert à étudier les transitions entre différentes conformations de macromolécules biologiques (par exemple entre les structures globulaires et dépliées d'une protéine). Par DSC, on mesure la capacité thermique (C_p) d'une solution de cette macromolécule en fonction de la température. La transition est caractérisée par un fin pic centré à la température de transition T_m . L'interaction d'une protéine avec des ligands conduit à des changements de la stabilité thermique de la protéine, ce qui conduit ainsi à un changement de la température de transition [41]. La nouvelle température de transition est supérieure à la température de transition sans ligand pour la protéine globulaire. La mesure des constantes d'interaction par DSC est indirecte puisque la constante de dissociation est estimée suivant l'équilibre comparé entre les formes globulaires (libres et avec ligands) et dépliées (sans ligands) de la protéine plutôt que directement sur des formes libres et avec ligands. Cette technique permet également d'obtenir des informations sur la stœchiométrie des complexes protéine-ligand pour des protéines qui peuvent fixer plusieurs ligands [42], mais comme l'ITC elle présente l'inconvénient de nécessiter une concentration élevée en protéines.

I.3.2.2. Méthodes spectroscopiques

Les méthodes spectroscopiques sont basées sur le changement de propriétés spectroscopiques des protéines lors de la fixation d'un ligand. La protéine avec ligand présente donc un spectre différent de la protéine libre. De ces différences on peut remonter aux concentrations en protéines libres et avec ligands et donc aux constantes d'équilibre.

I.3.2.2.1. Fluorimétrie

L'étude des interactions protéine-ligand par fluorimétrie repose sur le changement des spectres d'émission des fluorophores (voir I.2.2.2.3) dans la protéine lors de la fixation d'un ligand à proximité des fluorophores. La fixation du ligand conduit soit à un changement en intensité du maximum du spectre d'émission, soit à un déplacement du maximum du spectre d'émission (Fig. I.15).

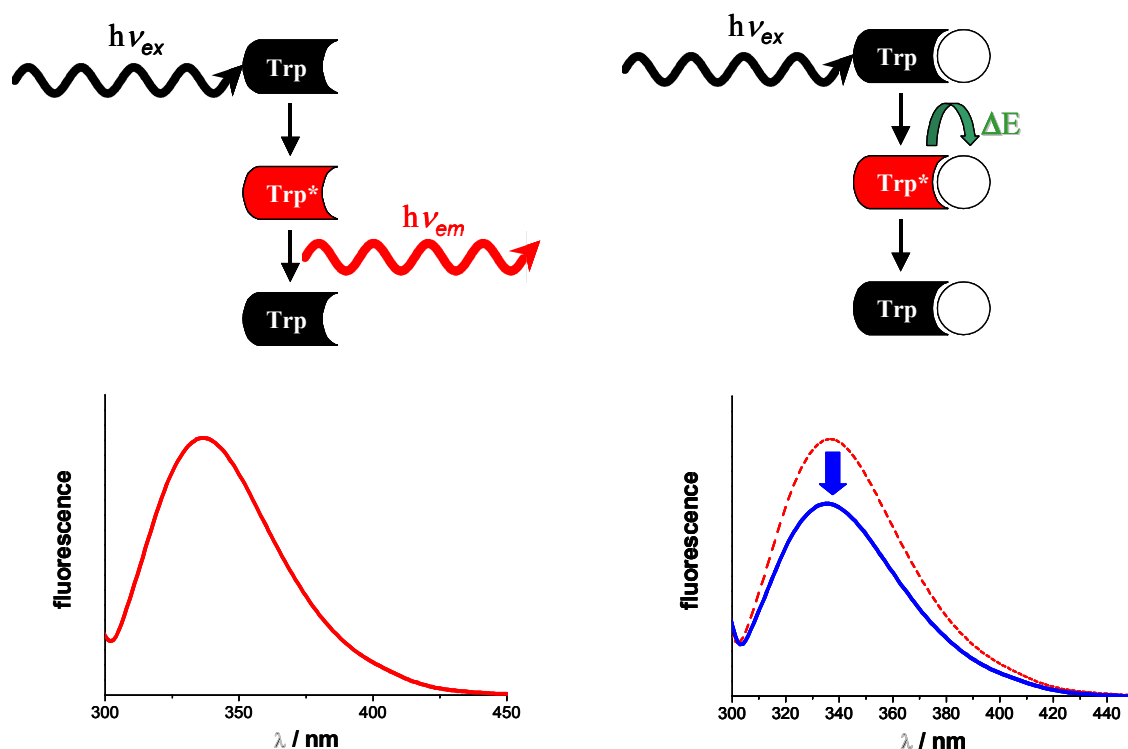


Fig. I.15 : Principe de la fluorimétrie pour l'études d'interaction protéine-ligand. Dans le schéma de droite, on constate que l'énergie du photon absorbé est susceptible d'être transférée du tryptophane vers le ligand fixé, ce qui conduit à un quench de la fluorescence et donc à une réduction de l'intensité de la fluorescence du tryptophane en présence de ligand.

Dans les deux cas, il est possible de déterminer la constante d'interaction et d'obtenir des informations sur la stœchiométrie et la coopérativité du système protéine-ligand analysé. Il est cependant nécessaire que les interactions soient toutes spécifiques, c'est-à-dire par exemple que les spectres d'émission des fluorophores qui ne sont pas impliqués dans l'interaction avec le ligand ne changent pas lors de la fixation du ligand.

I.3.2.2.2. RMN

La RMN (voir aussi I.2.3.2.2) est une méthode très puissante pour la caractérisation des interactions protéine-ligand. Lors de la fixation d'un ligand sur une protéine, plusieurs paramètres qui influencent les spectres RMN sont susceptibles de changer : les déplacements chimiques ainsi que les vitesses de relaxation et de diffusion [43]. Il est donc possible d'identifier les ligands qui se fixent sur la protéine ainsi que leur affinité pour la protéine. Les concentrations de protéine utilisées sont de l'ordre de 50-500 μM .

I.3.2.2.3. Résonance de plasmons de surface

La résonance de plasmons de surface (SPR) est une méthode de mesure de la liaison d'un ligand sur une protéine absorbée à la surface d'une couche métallique. Le système de détection est basé sur une variation de l'indice de réfraction à proximité (environ 300 nm) d'une interface métal-solution lors de la fixation d'un ligand [44]. Les processus moléculaires qui conduisent au phénomène du changement de l'indice de réfraction sont de nature quantique et relativement compliqués. Pour une revue sur la théorie de la SPR, voir [45] et un exemple intéressant, sur l'étude de l'interaction entre les protéines HIV-1 gp120 et CD4 est donné dans [46].

Un problème de la technique SPR est que des petits ligands (avec une masse moléculaire inférieure à 1000 Da) ne donnent pas de résultats satisfaisants.

I.3.2.3. Autres méthodes

I.3.2.3.1. Electrophorèse capillaire d'affinité

L'électrophorèse capillaire d'affinité est une technique particulière d'électrophorèse capillaire permettant d'obtenir des constantes d'équilibre. Cette méthode est basée sur les différences de vitesses de migration d'une protéine libre par rapport à la protéine en complexe avec un ligand dans un champ électrique. La mobilité électrophorétique (μ) d'une protéine est associée à sa masse moléculaire (M) et à sa charge (Z) par une relation d'une forme

approximative de $\mu \sim Z/M^{2/3}$. Si la protéine fixe un ligand chargé z fois d'une masse m qui est petite par rapport à la masse de la protéine, le changement de μ dû au changement de masse [de $M^{2/3}$ à $(M+m)^{2/3}$] est relativement petit par rapport au changement de μ dû au changement de charge (de Z à $Z \pm z$). Le complexe protéine-ligand va donc migrer à une vitesse différente de celle de la protéine libre. La mesure des temps de migration en fonction de la concentration en ligand permet d'obtenir des informations sur K_d [47].

L'électrophorèse capillaire d'affinité présente quelques inconvénients : tout d'abord, elle est difficile à appliquer pour les systèmes qui n'ont pas une stœchiométrie 1 : 1 [48]. Un autre inconvénient de la méthode est l'impossibilité de déterminer la concentration en ligand libre [48].

I.3.2.3.2. Dialyse à l'équilibre

La dialyse à l'équilibre (acronyme anglais : equilibrium dialysis) repose sur la détermination des concentrations des différentes espèces moléculaires à l'équilibre (voir équation I.3) afin de déterminer directement les constante(s) d'interaction ainsi que la stœchiométrie du complexe et la coopérativité entre plusieurs sites. Dans une expérience de dialyse à l'équilibre standard, deux chambres de même volume sont séparées par une membrane qui est perméable pour le ligand, mais pas pour la protéine. Le ligand est placé dans une des chambres, et la protéine est placée dans l'autre. Lorsque le ligand diffuse à travers la membrane, une partie des ligands est susceptible d'interagir avec la protéine. Cette diffusion ainsi que la fixation du ligand à la protéine continuera jusqu'à l'équilibre où la concentration en ligand libre dans les deux chambres est la même. Cette concentration en ligand est ensuite déterminée par une méthode adaptée et permet de déterminer les caractéristiques de l'interaction avec la protéine (K_d , stœchiométrie, coopérativité, etc.).

Pour obtenir des résultats fiables, il faut utiliser des concentration en protéine aux environs de 30-100 μM , ce qui est moins élevé que pour d'autres méthodes, mais qui correspond à une sensibilité relativement faible par rapport à ce que permet la spectrométrie de masse. La difficulté réside aussi dans la détermination de la concentration en ligand : il n'existe pas toujours une méthode directe adaptée, ce qui conduit fréquemment à l'utilisation de radio-isotopes [49] pour déterminer la concentration du ligand.

I.4. Etudes des complexes non covalents par spectrométrie de masse

Quel est l'intérêt de la spectrométrie de masse pour l'étude des complexes non covalents ? Le développement de sources d'ionisation dites « douces » (l'electrospray et le MALDI), a ouvert à la spectrométrie de masse le domaine de l'étude des protéines. Par rapport aux autres techniques utilisées jusque là, elle apporte plusieurs avantages. Tout d'abord, elle est très sensible : il est possible d'obtenir un spectre avec quelques picomoles, voire même femtomoles de protéine. Deuxièmement, la spectrométrie de masse permet d'obtenir des masses moléculaires très précises, à quelques ppm près, ce qui est nettement mieux que les informations obtenues par SDS PAGE (erreurs de l'ordre du kDa). Un troisième avantage réside dans la vitesse avec laquelle les informations sont obtenues. Dans le cadre de l'étude des interactions non covalentes, la spectrométrie de masse est susceptible de fournir des informations précises sur la stœchiométrie des complexes (nombre de ligands fixés) ainsi que des informations sur les constantes d'interaction (dans des conditions bien contrôlées).

Dans cette partie, la technique de l'electrospray de protéines, essentielle pour l'étude d'interactions non covalentes par spectrométrie de masse sera brièvement présentée (voir la partie VI pour les détails), et sera suivie d'un tour d'horizon des applications de la spectrométrie de masse pour l'étude des complexes non covalents de protéines et plus particulièrement pour l'étude de complexes protéine-ligand et de complexes protéiques.

I.4.1. Généralités de l'electrospray des protéines

Les spectres de masse des protéines ionisées par electrospray présentent une enveloppe de pics (voir Fig. I.16), correspondant à une famille d'ions de formule $[M+nH]^{n+}$ où n est l'état de charge de la protéine. La distribution des états de charge peut être corrélée à l'état de repliement plus ou moins important de la protéine, comme proposée par Chowdhury *et al.* [50]. Un exemple est l'analyse par spectrométrie de masse electrospray de l'ubiquitine dans diverses conditions. L'ubiquitine est une petite protéine d'environ 8650 Da. En conditions dénaturantes (Fig. I.16, spectre du haut), on observe une distribution d'états de charge assez large, centrée sur l'ion correspondant à l'état de charge 10+ de la protéine. Ceci reflète la structure ouverte de la protéine.

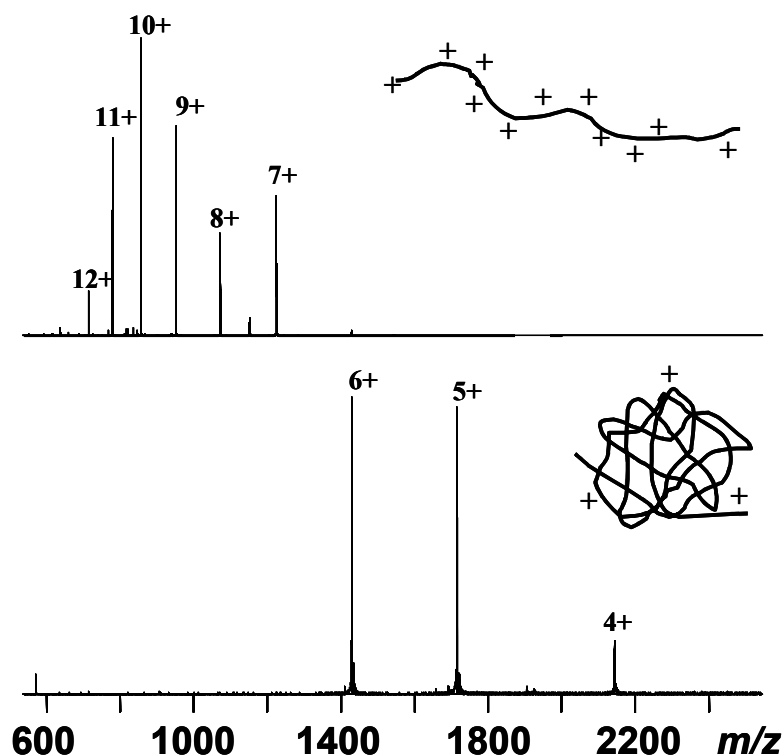


Fig. I.16 : Spectre electrospray de l'ubiquitine en conditions dénaturantes (en haut, conditions expérimentales : 2 μ M ubiquitine dans un mélange eau/méthanol/acide formique 49,9 : 49,9 : 0,2) et en conditions non dénaturantes (en bas, conditions expérimentales : 10 μ M ubiquitine dans un tampon bicarbonate d'ammonium pH=7,9 10 mM).

En conditions de solvant non dénaturantes (Fig. I.16, spectre du bas), on obtient une distribution d'états de charge beaucoup plus réduite, centrée sur l'ion correspondant à l'état de charge 5+. Ce résultat reflète une structure plus compacte de l'ubiquitine, et donc plus proche de son état natif. Il faut toutefois noter que l'information fournie par l'observation de la distribution d'états de charge ne peut être que qualitative, dans la mesure où de nombreux paramètres, notamment instrumentaux, peuvent avoir une influence notable.

I.4.2. Etudes des complexes protéine-ligand par spectrométrie de masse

Concernant les études des complexes protéine-ligand, il faut distinguer deux niveaux différents : tout d'abord, une analyse *qualitative* pour savoir si le ligand interagit avec la protéine en question. Un second stade est de tenter d'obtenir des informations *quantitatives* sur ces complexes, notamment des constantes d'interaction entre la protéine et les ligands. Ces deux questions seront traitées indépendamment dans la suite du chapitre.

I.4.2.1. Observation du complexe

Une des premières questions posées aux spectrométristes de masse concernant les interactions non covalentes est la possibilité de déterminer l'existence d'une interaction. Du point de vue de la mesure, la spectrométrie de masse peut sembler idéale pour ce genre d'étude puisqu'une interaction entre une protéine et un ligand peut se traduire par l'observation d'un pic dans le spectre de masse correspondant à l'adduit. Cette technique est particulièrement adaptée à la recherche de nouveaux médicaments [51], mais il faut garder à l'esprit que l'interaction a lieu en solution et que l'observation de la masse se fait dans le vide, sur une espèce ionisée.

Ainsi, une interaction présente en solution peut ne pas être détectée en phase gazeuse. C'est le cas par exemple des espèces non covalentes pour lesquelles des interactions hydrophobes (voir I.1.3.2) entrent en jeu. Des travaux ont ainsi mis en évidence l'impossibilité de relier les résultats obtenus en phase gazeuse à ceux obtenus en solution pour ce type de composés [52, 53].

A l'inverse, des interactions principalement électrostatiques (voir I.1.1, I.1.2) seront renforcées en phase gazeuse, et un complexe, même peu stable en solution, peut être détecté facilement par spectrométrie de masse. Compte tenu de ces éléments, il serait illusoire de vouloir relier des stabilités en phase gazeuse, étudiées par exemple en modifiant la tension à l'interface, aux stabilités en solution.

Dans certains cas, le problème est encore plus complexe, puisque le type d'interaction en phase gazeuse peut être différent de celui en solution. C'est le cas des complexes entre des cyclodextrines et des acides aminés aromatiques [54] : en solution, l'interaction est principalement de type hydrophobe entre le groupe aromatique et la cavité de la cyclodextrine, alors qu'en phase gazeuse, il s'agit d'une interaction ion-dipôle à l'extérieur de la cavité. Ceci signifie que des complexes inexistant en solution (notamment ceux impliquant des acides aminés non aromatiques) sont tout de même observés en phase gazeuse, donnant ainsi naissance à des faux positifs.

I.4.2.2. Détermination des constantes d'interaction

Le problème est encore plus complexe lorsqu'on cherche à obtenir des informations quantitatives à partir des spectres de masse. En effet, les seules informations disponibles aisément sont les intensités relatives des différentes espèces sur le spectre. Encore faut-il pouvoir les relier à des concentrations relatives en solution, rien ne prouvant que ces deux valeurs soient identiques. Pour cela, il faudrait que le rendement d'ionisation et les efficacités de transmission et de détection des différentes espèces soient comparables. Dans certaines

conditions, il est possible d'en faire l'hypothèse : par exemple dans le cas d'une protéine complexée ou non à un ligand de petite taille qui ne modifie pas sensiblement la structure de la protéine, et donc son rendement d'ionisation. On peut alors assimiler les intensités relatives dans le spectre aux concentrations relatives en solution, et en connaissant les concentrations initiales en solution, remonter aux concentrations absolues de protéine libre ou complexée. La connaissance de ces concentrations permet ensuite de déterminer les constantes d'interaction entre la protéine et son ligand.

Pour que ces déterminations de constantes d'interaction soient valables, il est indispensable que la stœchiométrie du complexe observée en phase gazeuse reflète la situation en solution. Dans ce contexte, il est important de rappeler que différentes situations peuvent se produire lors du passage du complexe de la solution à la phase gazeuse (Fig. I.17).

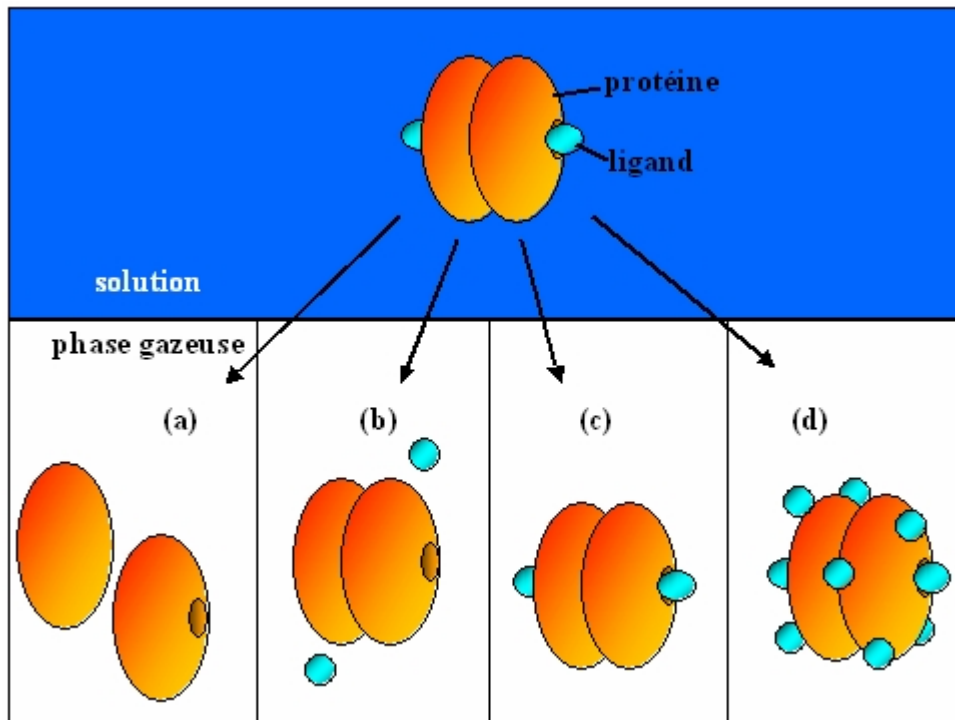


Fig. I.17 : Représentation des différents évènements qui peuvent avoir lieu lors du passage d'un complexe protéine-ligand en phase gazeuse. (a) perte de l'interaction entre des sous-unités de la protéine ; (b) perte de l'interaction entre la protéine et ses ligands ; (c) le complexe reste intact et la stœchiométrie ne change pas ; (d) fixation d'autres ligands (interactions non spécifiques).

Considérons une protéine dimérique (Fig. I.17, en orange) qui interagit avec deux ligands (Fig. I.17, en bleu clair) en solution. Lors du passage de ce complexe en phase gazeuse, quatre évènements principaux peuvent se produire sur la stœchiométrie du complexe :

- (a) Les interactions non covalentes entre les sous-unités de la protéine ne sont pas suffisamment fortes pour maintenir l'intégrité de la protéine en l'absence du solvant. L'interaction entre les sous-unités est perdue lors de la désolvatation et on observe la présence de monomères de la protéine dans le spectre.
- (b) La protéine entière reste intacte en phase gazeuse, mais l'interaction avec les ligands est perdue (pour les raisons possibles, voir I.4.2.1).
- (c) Dans des conditions expérimentales adaptées, le complexe entier est observé en phase gazeuse. Le site d'interaction précis entre la protéine et ses ligands peut différer de celui existant en solution, mais sans altérer la stœchiométrie du complexe. Seul ce cas permet la détermination de constantes d'association.
- (d) Des interactions non spécifiques sont observées, et se traduisent par une stœchiométrie surévaluée. Ce problème est souvent observé lors de l'utilisation de concentrations élevées de ligand.

La fragmentation de l'analyte, correspondant aux cas (a) et (b), peut souvent être contournée par l'optimisation des conditions expérimentales au niveau du spectromètre. Néanmoins, les interactions non spécifiques (cas d) constituent un problème plus délicat. Bien que ce problème puisse en théorie être simplement résolu en diminuant la concentration du ligand [55], la sensibilité des spectromètres de masse ne permet pas toujours de résoudre cette difficulté, en particulier lorsque les interactions protéine-ligand sont faibles. Il a été proposé de distinguer les interactions spécifiques et non spécifiques par fragmentation sélective des complexes non spécifiques en phase gazeuse, tout en gardant les interactions spécifiques [56], mais il a été montré que des interactions non spécifiques peuvent être plus fortes que des interactions spécifiques en phase gazeuse [57]. Il n'y a donc pas de méthode générale pour distinguer ces types d'interaction par spectrométrie de masse uniquement.

Daniel et al. ont montré l'intérêt que présente la détermination de constantes d'interaction par spectrométrie de masse en passant en revue des exemples de la littérature [58]. Ils constatent que pour la plupart des systèmes étudiés, les résultats obtenus par spectrométrie de masse présentent une bonne corrélation avec les valeurs obtenues directement en solution. La gamme des constantes de dissociation étudiées par des techniques basées sur la spectrométrie

de masse est de 1-880 μM [59-66]. Pour les systèmes avec $K_d < 1 \mu\text{M}$, il est difficile de déterminer précisément une constante puisque la quasi totalité du ligand se fixe à la protéine, ce qui aboutit à des problèmes lors du calcul des constantes. En revanche, pour les systèmes avec $K_d > 100 \mu\text{M}$, il est nécessaire de travailler avec un large excès de ligand pour observer la formation d'un complexe, et on risque ainsi la formation de complexes non spécifiques.

Du fait de toutes ces limitations, la spectrométrie de masse n'est pas nécessairement la méthode idéale pour déterminer des constantes d'interaction. Elle permet toutefois dans des cas particuliers d'obtenir une constante d'interaction avec une consommation de protéine relativement minime par rapport aux autres techniques.

I.4.3. Etudes structurales des complexes protéiques par spectrométrie de masse

Dans le cadre de travaux portant sur des complexes protéiques, une question fréquente est celle de la détermination de la structure et de la topologie (arrangement des différents partenaires du complexe les uns par rapport aux autres) du complexe. Quelles sont les stratégies pour répondre à ce problème, qui pourrait être classifié comme études sur la structure tertiaire et quaternaire des protéines, par spectrométrie de masse ?

Comme discuté précédemment, les deux techniques reines de la détermination structurale sont la cristallographie aux rayons X et la RMN. Toutefois, les limitations de ces techniques, en particulier leur faible sensibilité, leur limitation en gamme de masse, la difficulté pour obtenir des cristaux, ont été soulignées précédemment. Il reste donc un certain champ libre pour une approche de ces problématiques structurales par spectrométrie de masse, technique qui ne dispose pas d'une aussi bonne résolution spatiale, mais qui apporte sensibilité et rapidité par rapport aux deux techniques précédentes.

Les techniques permettant d'aborder la structure quaternaire (voir I.2.4.2) des protéines souffrent, elles, comme indiqué précédemment, de différents défauts : la chromatographie d'exclusion stérique permet de séparer des composants, mais est faiblement résolutive quant à la masse précise des complexes ainsi séparés, ce qui rend parfois impossible la détermination d'une stœchiométrie. De plus, la structure quaternaire est, comme la structure secondaire, dynamique, et les techniques comme le SAXS ou la microscopie électronique rendent difficilement compte de cette dynamique. La spectrométrie de masse apporte un moyen simple et efficace de déterminer des masses moléculaires et donc des stœchiométries. Nous verrons qu'en codant une information structurale sous forme d'une variation de masse, il est

également possible d'utiliser la spectrométrie de masse pour obtenir des informations structurales, statiques ou mêmes dynamiques.

Pour évaluer le potentiel de la spectrométrie de masse dans tous les domaines mentionnés, il est donc important de présenter comment la spectrométrie de masse permet d'obtenir des informations sur :

1. la masse moléculaire d'un complexe protéique
2. la structure tridimensionnelle d'une protéine ou d'un complexe protéique
3. la dynamique structurale d'une protéine ou d'un complexe protéique

I.4.3.1. Mesure de masse des complexes protéiques

Encore une fois, il est important de rappeler ici que la mesure de masse se fait sur un complexe initialement présent en solution, mais après passage en phase gazeuse. Comme pour les complexes protéine-ligand (voir I.4.3.1), le changement d'environnement est susceptible de modifier les différentes contributions énergétiques des interactions non covalentes (augmentation des interactions électrostatiques, disparition des interaction hydrophobes,), ce qui impose des contraintes pour l'observation des complexes non covalents protéine-protéine. De plus, la détermination des constantes d'interaction pour des complexes protéiques est plus difficile que pour des complexes protéine-ligand puisque l'efficacité d'ionisation d'une protéine est susceptible d'être considérablement modifiée lors de la complexation avec une autre protéine.

Malgré ces contraintes, quelques revues récentes de la littérature montrent l'intérêt de la spectrométrie de masse pour la détermination des masses moléculaires de complexes protéiques entiers [67-70]. Les plus grandes masses moléculaires atteintes actuellement sont de l'ordre de quelques MDa et comprennent des hémocyanines de 2,2 MDa de crabes abyssaux [71], le ribosome de *Thermus thermophilus* de 2,33 MDa [69] et le complexe multiprotéique de l'uréase de *Helicobacter pylori* à 4,26 MDa [68] qui constitue la plus grande masse de complexe protéique mesurée par spectrométrie de masse actuellement.

La technique d'ionisation de référence pour étudier les complexes protéiques est l'electrospray (ou sa variante, le nanospray). Dans certains cas particuliers, il est même possible de les étudier par MALDI [72], bien que cette technique ne soit probablement pas généralisable à tous les systèmes. Quelques techniques émergentes pour l'étude d'assemblages macromoléculaires ont été développées récemment ou sont en cours de

développement comme la mobilité ionique [73], l'« electrosonic spray ionization » (ESSI) [74] ou le « liquid beam ionization/desorption » (LILBID) [75].

Pour les complexes protéiques qui survivent au passage en phase gazeuse, la spectrométrie de masse est ainsi devenue un outil très puissant pour obtenir des informations sur leur masse moléculaire et donc pour déterminer leurs stœchiométries.

Il faut noter que la spectrométrie de masse permet, en plus de la détermination de stœchiométrie de complexes, d'identifier également les protéines présentes dans un complexe. Une technique récente permettant de répondre à cette question est la purification du complexe par affinité suivie de sa décomposition et de l'identification des protéines résultantes [76]. En suivant cette stratégie, Ho *et al.* ont identifiés de nombreux complexes protéiques de *Saccharomyces cerevisiae* [77].

I.4.3.2. Etude de la structure des complexes protéiques

Une fois déterminée la composition d'un complexe protéique, la question suivante qui vient à l'esprit est l'agencement spatial des différents partenaires les uns par rapport aux autres. Différentes possibilités existent pour étudier ces structures par spectrométrie de masse [70]. Tout d'abord, il est possible d'introduire des liaisons covalentes entre les différents partenaires du complexe par des réactifs multifonctionnels (« cross-linking »). Après dissociation et digestion enzymatique du complexe en solution, les peptides qui ont été « cross-linkés » sont identifiés par spectrométrie de masse. La longueur maximale de la chaîne du « cross-linker » étant connue, il est possible d'obtenir des contraintes spatiales pour la construction du complexe. Dans un exemple, Rappsilber *et al.* ont proposé grâce à cette technique un modèle tridimensionnel pour un complexe du pore nucléaire chez la levure [78].

Une autre possibilité est de passer par une modification chimique de la surface accessible en solution, et de mesurer ces modifications par spectrométrie de masse. Les régions d'une protéine en interaction avec d'autres protéines ne sont pas accessibles en solution et *a priori* ne seront pas modifiées. En comparant des cartes d'accessibilité pour des protéines libres et complexées, il devient possible de reconstruire la topologie du complexe. La modification chimique peut être réversible ou irréversible. Les échanges hydrogène/deutérium (H/D) [79-81] des amides de la liaison peptidique constituent des modifications réversibles tandis que la modification chimique des chaînes latérales par des réactifs chimiques particuliers (comme l'acétanhydride, le diéthylpyrocarbonate ou le phenylglyoxal) [81, 82] est irréversible, ce qui impose moins de contraintes au niveau de la rapidité de l'analyse après modification. Ces stratégies seront détaillées dans les parties correspondantes de la thèse.

I.4.3.3. Etude de la dynamique structurelle des complexes protéiques

Finalement, la même approche par modification chimique peut être utilisée pour étudier la dynamique structurelle des protéines ou des complexes protéiques. La caractérisation des modifications dans des conditions expérimentales différentes permet de remonter aux changements structuraux. En appliquant cette stratégie, Dumoulin *et al.* ont étudié le changement conformationnel du lysozyme induisant la formation d'agrégats multimériques [83]. Un autre exemple est l'étude des changements d'interaction protéine-protéine lors de la maturation des capsides virioniques de HIV-1 par échanges H/D [84]. De plus, cette technique permet également d'obtenir des informations sur les changements conformationnels d'une protéine lors de la fixation d'un petit ligand [85-87].

I.5. Conclusion

La spectrométrie de masse est devenue aujourd'hui un outil très puissant pour l'analyse des protéines. Grâce aux qualités de cette technique (sensibilité, vitesse, précision), il est non seulement possible d'obtenir des informations à tous les niveaux structuraux des protéines (de la structure primaire à la structure quaternaire), mais également sur leur interaction avec d'autres protéines et des ligands.

L'étude des interactions non covalentes en question se fait soit directement en phase gazeuse, soit indirectement par une modification chimique de certaines régions de la protéine en solution, qui correspond à des changements en masse moléculaire qui sont ensuite détectés par spectrométrie de masse.

Dans le cadre de cette thèse, nous avons souhaité déterminer quelle pouvait être l'utilité de la précision en masse, de la résolution et de capacité de fragmentation nouvelles, comme l'ECD, apportées par la spectrométrie de masse FT-ICR dans le cadre de ces deux approches, qui ont été utilisées pour répondre à des problématiques particulières.

Tout d'abord, dans l'étude des interactions protéine-ligand par spectrométrie de masse, les problèmes des interactions non spécifiques et de la comparaison des résultats obtenus par spectrométrie de masse avec ceux obtenus par d'autres méthodes ont été abordés.

Dans une deuxième partie, différentes méthodes pour l'étude des protéines par spectrométrie de masse FT-ICR ont été mises au point, notamment la modification chimique des protéines par le diéthylpyrocarbonate, les échanges hydrogène/deutérium et la fragmentation des protéines entières en phase gazeuse, une approche dite « Top-Down ».

Dans une troisième partie, la technique des échanges H/D a été appliquée à un modèle particulier : la protéine du prion ovin. Cette protéine forme, dans certaines conditions expérimentales *in vitro*, des oligomères, dont la structure n'est pas encore connue. L'approche *via* la spectrométrie de masse a permis d'obtenir des informations sur leur structure et la dynamique conformationnelle lors de la formation des oligomères.

I.6. Bibliographie

- [1] H. Hartley, Origin of the word 'Protein', *Nature* **1951**, 168, 244.
- [2] L. Stryer, *Biochimie*. 4. Aufl., Spektrum Akademischer Verlag, Heidelberg (1995).
- [3] H. Kitano, Systems biology: a brief overview, *Science* **2002**, 295, 1662-1664.
- [4] J. Israelachvili, *Intermolecular & Surface Forces*. Academic Press, New York and London (1992).
- [5] M. Laberge, Intrinsic protein electric fields: basic non covalent interactions and relationship to protein-induced Stark effects, *Biochim. Biophys. Acta* **1998**, 1386, 305-330.
- [6] T. E. Creighton, *Proteins – Structures and molecular properties*. 2nd Ed., W. H. Freeman and Company, New York (1993).
- [7] J. R. Chapman, ed., *Mass Spectrometry of Proteins and Peptides*, Meth. Mol. Biol., Vol. 146, Humana Press, Totowa (2000).
- [8] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of The Cell*. 4. ed., Garland Science (Taylor & Francis Group), New York (2002).
- [9] C. Branden, J. Tooze, *Introduction to Protein Structure*, 2^{ème} edition, Garland Publishing, New York (1999).
- [10] N. Sreerama, R. W. Woody, Circular dichroism of peptides and proteins, In: *Circular Dichroism: Principles and Applications*. R. W. Woody (ed.), Wiley-VCH, New York (2000), pp. 601-620.
- [11] N. Sreerama, S. Venyaminov, R. Woody, Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy, *Protein Sci.* **1999**, 8, 370-380.
- [12] A. Barth, C. Zscherp, What vibrations tell us about protein, *Q. Rev. Biophys.* **2002**, 35, 369-430.
- [13] C. Kötting, K. Gerwert, Proteins in action monitored by time-resolved FTIR spectroscopy, *Chem. Phys. Chem.* **2005**, 6, 1-8.

- [14] C. Zscherp, A. Barth, Reaction-induced infrared difference spectroscopy for the study of protein reaction mechanisms, *Biochemistry* **2001**, *40*, 1875-1883.
- [15] F. Garczarek, K. Gerwert, Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy, *Nature* **2006**, *439*, 109-112.
- [16] Y. Chen, M. D. Barkley, Toward understanding tryptophan fluorescence in proteins, *Biochemistry* **1998**, *37*, 9976-9982.
- [17] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Kluwer Academic/Plenum, New York (1999).
- [18] B. W. van der Meer, G. Coker III, S.-Y. Chen, *Resonance Energy Transfer: Theory and Data*. VCH, New York (1994).
- [19] A. M. Kelley, X. Michalet, S. Weiss, Chemical physics: single-molecule spectroscopy comes of age, *Science* **2001**, *292*, 1671-1672.
- [20] S. Weiss, Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy, *Nat. Struct. Biol.* **2000**, *7*, 724-729.
- [21] C. R. Matthews (éd.), Protein Folding Mechanisms, *Adv. Prot. Chem.*, Vol. 53, Academic Press, New York (2000).
- [22] G. Rhodes, *Crystallography made crystal clear. A guide for users of macromolecular models*. Academic Press, San Diego (2000).
- [23] A. Schmidt, V. S. Lamzin, Veni, vidi, vici – atomic resolution unravelling the mysteries of protein function, *Curr. Opin. Struct. Biol.* **2002**, *12*, 698-703.
- [24] P. I. Lario, A. Vrielink, Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution, *J. Am. Chem. Soc.* **2003**, *125*, 12787-12794.
- [25] G. A. Petsko, D. Ringe, Observation of unstable species in enzyme-catalyzed transformations using protein crystallography, *Curr. Opin. Chem. Biol.* **2000**, *4*, 89-94.
- [26] K. Moffat, Time-resolved biochemical crystallography: a mechanistic perspective, *Chem. Rev.* **2001**, *101*, 1569-1581.
- [27] M. E. Wall, S. C. Gallagher, J. Trewhella, Large-scale shape changes in proteins and macromolecular complexes, *Annu. Rev. Phys. Chem.* **2000**, *51*, 355-380.
- [28] L. Chen, K. O. Hodgson, S. Doniach, A lysozyme folding intermediate revealed by solution X-ray scattering, *J. Mol. Biol.* **1996**, *261*, 658-671.
- [29] D. J. Segel, D. Eliezer, V. Uversky, A. L. Fink, K. O. Hodgson, S. Doniach, Transient dimer in the refolding kinetics of cytochrome c characterized by small-angle X-ray scattering, *Biochemistry* **1999**, *38*, 15352-15359.

- [30] V. M. Unger, Electron cryomicroscopy methods, *Curr. Opin. Struct. Biol.* **2001**, *11*, 548-554.
- [31] J. Frank, Single-particle imaging of macromolecules by cryo-electron microscopy, *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 303-319.
- [32] W. Chiu, A. McGough, M. B. Sherman, M. F. Schmid, High-resolution electron cryomicroscopy of macromolecular assemblies, *Trends Cell. Biol.* **1999**, *9*, 154-159.
- [33] L. A. Amos, Focusing-in on microtubules, *Curr. Opin. Struct. Biol.* **2000**, *10*, 236-241.
- [34] H. R. Saibil, Conformational changes studied by cryo-electron microscopy, *Nat. Struct. Biol.* **2000**, *7*, 711-714.
- [35] A. C. Steven, U. Aebi, The next ice age: cryo-electron tomography of intact cells, *Trends Cell. Biol.* **2003**, *13*, 107-110.
- [36] I. S. Gabashvili, R. K. Agrawal, C. M. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, P. Penczek, Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution, *Cell* **2000**, *100*, 537-549.
- [37] L. Tang, J. E. Johnson, Structural biology of viruses by the combination of electron cryomicroscopy and X-ray crystallography, *Biochemistry* **2002**, *41*, 11517-11524.
- [38] O. Medalia, I. Weber, A. S. Frangakis, D. Nicastro, G. Gerisch, W. Baumeister, Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography, *Science* **2002**, *298*, 1209-1213.
- [39] S. Falke, M. T. Fisher, E. P. Gogol, Classification and reconstruction of a heterogeneous set of electron microscopic images: a case study of GroEL-substrate complexes, *J. Struct. Biol.* **2001**, *133*, 203-213.
- [40] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, H. R. Saibil, ATP-bound states of GroEL captured by cryo-electron microscopy, *Cell* **2001**, *107*, 869-879.
- [41] G. Bruylants, J. Wouters, C. Michaux, Differential scanning calorimetry in life science : thermodynamics, stability, molecular recognition and application in drug design, *Curr. Med. Chem.* **2005**, *12*, 2011-2020.
- [42] M. S. Celej, S. A. Dassie, M. Gonzalez, M. L. Bianconi, G. D. Fidelio, Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins, *Anal. Biochem.* **2006**, *350*, 277-284.
- [43] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules, *J. Am. Chem. Soc.* **1997**, *119*, 12257-12261.

- [44] D. J. O'Shannessy, M. Brigham-Burke, K. K. Soneson, P. Hensley, I. Brooks, Determination of rate and equilibrium binding constants for macromolecular interactions by surface plasmon resonance, *Meth. Enzym.* **1994**, *240*, 323-349.
- [45] H. Raether, *Surface plasmons on smooth and rough surfaces and on grating*. Springer-Verlag, Berlin (1988).
- [46] M. Brigham-Burke, J. R. Edwards, D. J. O'Shannessy, Detection of receptor-ligand interactions using surface plasmon resonance: model studies employing the HIV-1 gp120/CD4 interaction, *Anal. Biochem.* **1992**, *205*, 125-131.
- [47] Y.-H. Chu, L. Z. Avila, H. A. Biebuyck, G. M. Whitesides, Use of affinity capillary electrophoresis to measure binding constants of ligands to proteins, *J. Med. Chem.* **1992**, *35*, 2915-2917.
- [48] M. H. A. Busch, L. B. Carels, H. F. M. Boelens, J. C. Kraak, H. Poppe, Comparison of five methods for the study of drug-protein binding in affinity capillary electrophoresis, *J. Chromatogr. A* **1997**, *777*, 311-328.
- [49] J. Masuoka, J. Hegenauer, B. R. Van Dyke, P. Saltman, Intrinsic stoichiometric equilibrium constants for the binding of zinc(II) and copper(II) to the high affinity site of serum albumin, *J. Biol. Chem.* **1993**, *268*, 21533-21537.
- [50] S. K. Chowdhury, V. Katta, B. T. Chait, Probing conformational changes in proteins by mass spectrometry, *J. Am. Chem. Soc.* **1990**, *112*, 9012-9013.
- [51] G. Deng, G. Sanyal, Applications of mass spectrometry in early stages of target based drug discovery, *J. Pharm. Biomed. Anal.* **2006**, *40*, 528-538.
- [52] C. V. Robinson, E. W. Chung, B. B. Kragelund, J. Knudsen, R. T. Aplin, F. M. Poulsen, C. M. Dobson, 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.
- [53] Q. Wu, J. Gao, D. Joseph-McCarthy, G. B. Sigal, J. E. Bruce, G. M. Whitesides, R. D. Smith, Carbonic anhydrase-inhibitor binding: from solution to the gas phase, *J. Am. Chem. Soc.* **1997**, *119*, 1157-1158.
- [54] J. D. Cunniff, P. Vouros, False positives and the detection of cyclodextrin inclusion complexes by electrospray mass spectrometry, *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 437-447.
- [55] R. D. Smith, K. J. Light-Wahl, B. E. Winger, J. A. Loo, Preservation of non covalent associations in electrospray ionization mass spectrometry: multiply charged polypeptide and protein dimers, *Org. Mass Spectrom.* **1994**, *29*, 169-175.
- [56] O. Laprévotte, P. Ducrot, C. Thal, L. Serani, B. C. Das, Stereochemistry of electrosprayed ions. Indoloquinolizidine derivatives, *J. Mass Spectrom.* **1996**, *31*, 1149-1155.

- [57] W. Wang, E. Kitova, J. S. Klassen, Determination of protein-oligosaccharide binding by nanoelectrospray Fourier-transform ion cyclotron resonance mass spectrometry. *Meth. Enzym.* **2003**, 362, 376-396.
- [58] J. M. Daniel, S. D. Friess, S. Rajagopalan, S. Wendt, R. Zenobi, Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry, *Int. J. Mass Spectrom.* **2002**, 216, 1-27.
- [59] K. B. Turner, N. A. Hagan, D. Fabris, Inhibitory effects of archetypical nucleic acid ligands on the interactions of HIV-1 nucleocapsid protein with elements of psi-RNA, *Nucleic Acids Res.* **2006**, 34, 1305-1316.
- [60] A. Delobel, E. Graciet, S. Andreescu, B. Gontero, F. Halgand, O. Laprévotte, Mass spectrometric analysis of the interactions between CP12, a chloroplast protein, and metal ions: a possible regulatory role within a PRK/GAPDH/CP12 complex, *Rapid Commun. Mass Spectrom.* **2005**, 19, 3379-3388.
- [61] K. de Vriendt, K. Sandra, T. Desmet, W. Nérinckx, J. van Beeumen, B. Devreese, Evaluation of automated nano-electrospray mass spectrometry in the determination of non covalent protein-ligand complexes, *Rapid Commun. Mass Spectrom.* **2004**, 18, 3061-3067.
- [62] V. Henrotte, S. Laurent, V. Gabelica, L. V. Elst, E. de Pauw, R. N. Muller, Investigation of non covalent interactions between paramagnetic complexes and human serum albumin by electrospray mass spectrometry, *Rapid Commun. Mass Spectrom.* **2004**, 18, 1919-1924.
- [63] S. W. Bligh, T. Haley, P. N. Lowe, Measurement of dissociation constants of inhibitors binding to Src SH2 domain protein by non covalent electrospray ionization mass spectrometry, *J. Mol. Recognit.* **2003**, 16, 139-148.
- [64] S. M. Clark, L. Konermann, Determination of ligand-protein dissociation constants by electrospray mass spectrometry-based diffusion measurements, *Anal. Chem.* **2004**, 76, 7077-7083.
- [65] J. M. Daniels, G. McCombie, S. Wendt, R. Zenobi, Mass spectrometric determination of association constants of adenylate kinase with two noncovalent inhibitors, *J. Am. Soc. Mass Spectrom.* **2003**, 14, 442-448.
- [66] A. Ayed, A. N. Krutchinsky, W. Ens, K. G. Standing, H. W. Duckworth, Quantitative evaluation of protein-protein and ligand-protein equilibria of a large allosteric enzyme by electrospray ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **1998**, 12, 339-344.
- [67] R. H. H. van den Heuvel, A. J. R. Heck, Native protein mass spectrometry: from intact oligomers to functional machineries, *Curr. Opin. Chem. Biol.* **2004**, 8, 519-526.

- [68] A. J. R. Heck, R. H. H. van den Heuvel, Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom. Rev.* **2004**, *23*, 368-389.
- [69] J. L. Benesch, C. V. Robinson, Mass spectrometry of macromolecular assemblies: preservation and dissociation, *Curr. Opin. Struct. Biol.* **2006**, *16*, 245-251.
- [70] F. Sobott, C. V. Robinson, Protein complexes gain momentum, *Curr. Opin. Struct. Biol.* **2002**, *12*, 729-734.
- [71] S. Sanglier, E. Leize, A. van Dorsselaer, F. Zal, Comparative ESI-MS study of approximately 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.
- [72] A. M. Distler, J. Allison, C. Hiser, L. Qin, Y. Hilmi, S. Ferguson-Miller, Mass spectrometric detection of protein, lipid and heme components of cytochrome c oxidase from *R. sphaeroides* and the stabilization of non covalent complexes from the enzyme, *Eur. J. Mass Spectrom.* **2004**, *10*, 295-308.
- [73] J. A. Loo, B. Berhane, C. S. Kaddis, K. M. Wooding, Y. Xie, S. L. Kaufman, I. V. Chernushevich, Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 998-1008.
- [74] Z. Takats, J. M. Wiseman, B. Gologan, R. G. Cooks, Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure, *Anal. Chem.* **2004**, *76*, 4050-4058.
- [75] A. Wattenberg, F. Sobott, B. Brutschy, Detection of intact hemoglobin from aqueous solution with laser desorption mass spectrometry, *Rapid Commun. Mass Spectrom.* **2000**, *14*, 859-861.
- [76] A. Dziembowski, B. Séraphin, Recent developments in the analysis of protein complexes, *FEBS Lett.* **2004**, *556*, 1-6.
- [77] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier et al., Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry, *Nature* **2002**, *415*, 180-183.
- [78] J. Rappsilber, S. Siniosoglou, E. C. Hunt, M. Mann, A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry, *Anal. Chem.* **2000**, *72*, 267-275.
- [79] J. R. Engen, D. L. Smith, Investigating the higher order structure of proteins: hydrogen exchange, proteolytic fragmentation, and mass spectrometry, *Meth. Mol. Biol.* **2000**, *146*, 112.
- [80] C. S. Maier, M. L. Deinzer, Protein conformations, interactions, and H/D exchange, *Meth. Enzymol.* **2005**, *402*, 312-360.

- [81] J. Borch, T. J. Jorgensen, P. Roepstorff, Mass spectrometric analysis of protein interactions, *Curr. Opin. Chem. Biol.* **2005**, 9, 509-516.
- [82] K. L. Bennett, T. Matthiesen, P. Roepstorff, Probing protein surface topology by chemical surface labeling, crosslinking, and mass spectrometry, *Meth. Mol. Biol.* **2000**, 146, 113-131.
- [83] M. Dumoulin, D. Canet, A. M. Last, E. Pardon, D. B. Archer, S. Muyldermans, L. Wyns, A. Matagne, C. V. Robinson, C. Redfield et al., Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations, *J. Biol. Chem.* **2005**, 346, 773-788.
- [84] J. Lanman, T. T. Lam, M. R. Emmett, A. G. Marshall, M. Sakalian, P. E. Prevelige, Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange, *Nat. Struct. Mol. Biol.* **2004**, 11, 676-677.
- [85] M. J. Chalmers, S. A. Busby, B. D. Pascal, Y. He, C. L. Hendrickson, A. G. Marshall, P. R. Griffin, Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry, *Anal. Chem.* **2006**, 78, 1005-1014.
- [86] S. Akashi, K. Takio, Characterization of the interface structure of enzyme-inhibitor complex by using hydrogen-deuterium exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Protein Sci.* **2000**, 9, 2497-2505.
- [87] D. King, M. Lumpkin, C. Bergmann, R. Orlando, Studying protein-carbohydrate interactions by amide hydrogen/deuterium exchange mass spectrometry, *Rapid Commun. Mass Spectrom.* **2002**, 16, 1569-1574.

Partie II

Etudes des complexes protéine-ligand intacts par spectrométrie de masse

Cette première partie traitera de l'étude par spectrométrie de masse des constantes d'interactions entre une protéine (la créatine kinase) et deux de ses ligands (l'ADP et l'ATP). Nous aborderons dans un premier temps le problème des interactions non spécifiques formées au cours du processus d'électrospray avant de comparer les résultats obtenus par spectrométrie de masse avec ceux obtenus par d'autres méthodes. Nous nous intéresserons plus particulièrement aux conditions expérimentales pouvant influencer les résultats.

L'intérêt de la spectrométrie de masse pour l'étude des interactions non covalentes protéine-ligand est bien établi (voir introduction générale). Dans le cadre de cette thèse, un des objectifs était d'évaluer l'intérêt de la spectrométrie de masse FT-ICR pour ce type d'étude.

Une première question abordée était l'intérêt de la haute résolution, dans des cas où m/z serait élevé. Le système choisi pour aborder cette question était le système de la créatine kinase (CK). Cette enzyme est un homodimère de 86 kDa qui catalyse l'interconversion ATP + créatine en ADP + phosphocréatine et n'avait pas encore été étudiée par spectrométrie de masse. Les deux ligands ATP et ADP diffèrent en masse de 80 Da, et l'observation simultanée des deux ligands fixés à la protéine nécessite une résolution supérieure à 40 ppm, ce qui est le cas de la spectrométrie de masse FT-ICR (Fig. II.1).

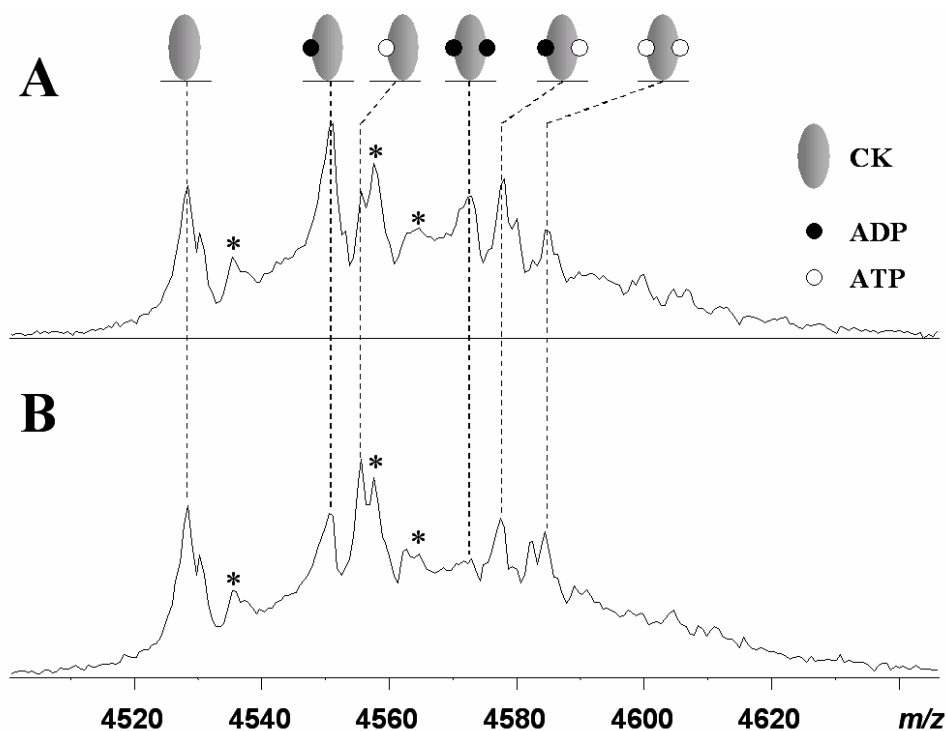


Fig. II.1 : Dosage simultané de la CK avec de l'ADP et de l'ATP dans du bicarbonate d'ammonium 10 mM (pH 7.9). (A) 20 μ M CK + 40 μ M ADP + 40 μ M ATP (état de charge 19+), (B) 20 μ M CK + 20 μ M ADP + 60 μ M ATP (état de charge 19+). Les étoiles (*) correspondent à des adduits (probablement des sels).

Un dosage avec à la fois de l'ATP et de l'ADP (Fig. 1) met en évidence l'observation simultanée de différents complexes (CK+ADP, CK+ATP, CK+2 ADP, CK+ADP+ATP, CK+2 ATP) (Fig. 1, A), mais aussi un changement des intensités relatives des différents

complexes dans le spectre en fonction de la concentration relative des ligands en solution (Fig. 1, B). Ces expériences permettent de vérifier la spécificité de la majeure partie des interactions observées, puisqu'on n'aurait pas observé cette « chasse » dans le cas d'interactions non spécifiques. Sur la base de ces travaux préliminaires, nous avons poursuivi ce travail avec pour objectif de répondre aux questions suivantes :

- L'interaction entre la protéine et ses ligands, observée en phase gazeuse, est-elle représentative de la stœchiométrie en solution ?
- Peut-on déterminer la valeur de la constante d'interaction pour la comparer à celles trouvées dans la littérature ?
- Pour compléter la question précédente : quelle peut être l'influence des paramètres expérimentaux (tampon, pH,) sur la constante d'interaction, et est-ce que cela peut être mesuré par spectrométrie de masse ?

Lors du dosage de la CK par de l'ADP ou de l'ATP, des interactions non spécifiques sont apparues. Dans un premier chapitre nous présenterons le développement d'un modèle statistique rendant possible la séparation de la contribution des interactions non spécifiques et des interactions spécifiques, ce qui permet le calcul des constantes d'interaction. Ce chapitre ayant été publié¹, il est reproduit intégralement dans cette thèse. Une comparaison des constantes de dissociation obtenues par spectrométrie de masse avec des valeurs de la littérature conduit essentiellement à constater l'extrême diversité des conditions expérimentales utilisées par les différents groupes. Par conséquent, dans une seconde partie, l'effet des conditions expérimentales (essentiellement tampon et pH) sur les constantes de dissociation mesurées par spectrométrie de masse a été étudié.

¹ T. Daubenfeld, A.-P. Bouin, G. van der Rest, A deconvolution method for the separation of specific versus non-specific interactions in noncovalent protein-ligand complexes analyzed by ESI-FT-ICR mass spectrometry, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1239-1248.

Chapitre II.A

Déconvolution des interactions
spécifiques et non spécifiques

II.A.1. Remarques générales

L'étude des interactions non covalentes protéine-ligand par spectrométrie de masse est fréquemment compliquée par la présence d'interactions non spécifiques qui perturbent la mesure de la stœchiométrie du complexe formé entre une protéine et ses ligands. Lorsque c'est le cas, les résultats obtenus en phase gazeuse ne permettent pas d'obtenir des informations correctes sur la composition du complexe en solution et par conséquent de quantifier cette interaction par le calcul des constantes d'équilibre.

En présence de ces interactions non spécifiques, et à condition qu'il n'y ait pas de décomposition du complexe en solution lors du passage en phase gazeuse, il faut distinguer deux situations quant au nombre de ligands : le nombre de ligands initialement fixés en solution (s) et le nombre de ligands non spécifiques (n) ajoutés au complexe lors de la désolvatation. Le nombre total de ligands mesurés par spectrométrie de masse pour un complexe donné (l) est alors évidemment la somme de ces deux nombres ($l = s + n$). Pour établir une constante d'équilibre, seule la contribution s est intéressante, mais la difficulté est qu'il n'existe aucun moyen direct pour séparer par spectrométrie de masse ces deux contributions.

Toutefois, on peut remarquer que si on prend l'ensemble des complexes mesurés, la distribution des interactions non spécifiques devrait suivre une distribution statistique sur un grand nombre de sites, alors que la distribution des interactions spécifiques dépend du système étudié. Si on a des hypothèses sur le système étudié, comme par exemple la présence de deux sites spécifiques et indépendants, on peut par conséquent prédire chacune de ces distributions à partir de deux paramètres : le nombre moyen de ligands fixés de façon spécifique ($\bar{s}$) suivant une distribution binomiale à deux sites (Tableau II.A.1) et le nombre moyen de ligands fixés de façon non spécifique ($\bar{n}$) suivant une distribution de Poisson (Tableau II.A.1).

Tableau II.A.1 : Exemples de distributions pour le cas spécifique sur deux sites équivalents et indépendants et pour le cas non spécifique. Ces distributions ne sont pas normalisées.

Distribution	0	1	2	3	4
Binomiale à deux sites	1	$\frac{\bar{s}}{1 - \bar{s}/2}$	$\frac{\bar{s}^2}{4 \cdot (1 - \bar{s}/2)^2}$	-	-
Poisson	1	$\bar{n}$	$\frac{\bar{n}^2}{2}$	$\frac{\bar{n}^3}{6}$	$\frac{\bar{n}^4}{24}$

Il convient ensuite de noter que le nombre total moyen de ligands ($\bar{l}$) est un paramètre expérimental qui relie les deux moyennes $\bar{s}$ et $\bar{n}$ (équation II.A.1). Par conséquent ces deux distributions peuvent être totalement décrites en considérant la valeur de $\bar{l}$ et un paramètre ajustable $\bar{s}$.

$$\bar{l} = \frac{\sum_{i=0}^n I_i \cdot i}{\sum_{i=0}^n I_i} = \bar{s} + \bar{n} \quad (\text{eqn. II.A.1})$$

Dans cette équation, i est le nombre de ligands fixés à la protéine, I_i est l'intensité du pic correspondant au complexe de la protéine avec i ligands fixés. Les deux distributions (spécifiques et non spécifiques) sont convoluées dans le cas où il y a superposition des deux effets. Ceci se traduit, si on prend l'exemple du modèle de la créatine kinase, par le fait que l'ion correspondant à la protéine ayant fixée un ligand (R_l) est la superposition de deux contributions (Equation II.A.2) d'un ligand spécifique (R_l^1) et d'un ligand non spécifique (R_l^0). Les pics suivants (2, 3 ou 4 ligands) sont, de façon plus complexe, également la superposition de contributions spécifiques, par exemple pour 2 ligands suivant l'équation II.A.3.

$$R_l = R_l^1 + R_l^0 = \frac{\bar{s}}{(1 - \bar{s}/2)} + (\bar{l} - \bar{s}) \quad (\text{eqn. II.A.2})$$

$$R_2 = R_2^2 + R_2^1 + R_2^0 = \frac{\bar{s}^2}{4 \cdot (1 - \bar{s}/2)^2} + \frac{\bar{s}}{(1 - \bar{s}/2)} \cdot (\bar{l} - \bar{s}) + \frac{(\bar{l} - \bar{s})^2}{2} \quad (\text{eqn. II.A.3})$$

Le point à retenir est qu'en ajustant la valeur de $\bar{s}$, on peut trouver une distribution théorique correspondant à celle observée expérimentalement. Une fois cette valeur de $\bar{s}$ optimisée, il découle facilement du tableau II.A.1 qu'on peut remonter aux seules contributions spécifiques et par conséquent aux constantes d'équilibre.

II.A.2. Abstract

A method to separate specific and non-specific non-covalent interactions observed in ESI mass spectra between a protein and its ligands is presented. Assuming non-cooperative binding, the specific ligand binding is modeled as a statistical distribution on identical binding sites. For the non-specific fraction we assume a statistical distribution on a large number of “non-specific” interacting sites. The model was successfully applied to the non-covalent interaction between the protein Creatine kinase (CK) and its ligands ADP and ATP that both exhibit non-specific binding in the mass spectrum. The two sequential dissociation constants obtained by applying our method are $K_{1,diss} = 11.8 \pm 1.5 \mu\text{M}$ and $K_{2,diss} = 48 \pm 6 \mu\text{M}$ for ADP. For ATP, the constants are $K_{1,diss} = 27 \pm 7 \mu\text{M}$ and $K_{2,diss} = 114 \pm 27 \mu\text{M}$. All constants are in good correlation with reported literature values. The model should be valuable for systems with a large dissociation constant that require high ligand concentrations and thus have increased potential of forming non-specific adducts.

II.A.3. Introduction

The function of most, if not all, proteins is defined by their interaction with ligands (e.g. other proteins, DNA, RNA, small metabolites). These interactions are mediated by non-covalent interactions (hydrogen bonds, ionic and van der Waals as well as hydrophobic interactions). In order to understand the biological role of a protein it is important to obtain detailed information about these interactions.

Electrospray ionization (ESI) mass spectrometry [1] has gained interest as a rapid and sensitive tool to derive binding constants for non-covalent interactions [2-6]. The main advantages of the mass spectrometry (MS) based approach are its sensitivity (only fmoles of protein are needed), the high mass accuracy and the possibility to obtain direct information about the stoichiometry of biomolecular complexes. However, one has to keep in mind that when using MS based methods, the observation of the interaction in question is made in the gas phase. The solution phase equilibria thus could have been significantly altered upon transfer to the gas phase due to the modified energetic environment (like the increase of ionic interactions and loss of hydrophobic interactions in the vacuum). One of the main problems encountered in the MS approach is the binding of ligands upon transfer of the analyte to the gas phase, thus giving rise to so-called non-specific interactions [7-14] that alter the initial (i.e. solution phase) stoichiometry of the complex (e.g. due to increase of ligand concentration during the evaporation of the droplet in the ESI process). In such a case, the gas phase observation does not directly reflect the liquid phase equilibrium. In general, non-specific

interactions may arise when a large excess of ligand is employed, as it becomes necessary for systems with weak binding constants ($K_{\text{diss}} > 1 \mu\text{M}$). In order to study systems with weak binding constants, it is thus necessary to find means to separate specific from non-specific binding. As both specific and non-specific complexes have the same molecular mass, it is not possible to differentiate between them directly from the mass spectrum. One possibility to separate specific and non-specific interactions might be to detach the non-specifically bound ligands from the protein in the gas phase, for example by using blackbody infrared radiation dissociation (BIRD) [11]. However, this approach only yields reasonable results if the gas-phase interaction of the protein with the specifically bound ligands is much stronger than its interaction with the non-specifically bound ligands [8]. Klassen and coworkers have shown an example of a complex containing specific and non-specific interactions where non-specific binding is stronger than specific binding [9]. It is thus not generally possible to separate a mixture of specific and non-specific interactions using mass spectrometry techniques alone.

The phenomenon of non-specific binding was encountered in the course of a study centered on the non-covalent interaction between rabbit muscle creatine kinase (CK) and its physiological ligands adenosine diphosphate (ADP) and adenosine triphosphate (ATP). CK is an 86 kDa homodimeric protein that is involved in the energy balance of muscle cells by catalyzing the following reaction:



CK has been extensively studied by traditional biochemical techniques since the 1960s [15] and a comprehensive review of the studies would be beyond the scope of this article. However, two elements are worth notice: it has been established that CK has two binding sites. The binding constants of the CK/ADP interaction reported in the literature cover a wide range of values ($K_{\text{diss,ADP}}$ ranging from 0.5 μM to 500 μM [15-20]). The binding constants show a dependence on the presence of other molecular species (such as Mg^{2+} or creatine) that are involved in the biological role of the protein. However, most of the reported literature values, especially those obtained in the absence of other ligand, are in the range of 70-140 μM [17-20]. For ATP, the interaction with CK has been found to be less strong, the dissociation constant being around 300-500 μM [20]. However, for both ligands, it is not yet clear if the two binding sites are cooperative or not.

It was thus of interest to probe the non-covalent interaction between CK/ADP and CK/ATP, respectively, through the use of an MS technique, which allows for the simultaneous determination of the binding constant for both sites.

The ESI MS observations on the CK/ADP solutions as well as the CK/ATP solutions showed evidence for non-specific binding. In order to determine the specific binding constants, we built a mathematical model that allows deconvolution between specific and non-specific interactions. This allowed the determination of the two binding constants. The model is based on a binomial distribution of ligands on two putative specific binding sites and a Poisson distribution of ligands on putative non-specific binding sites. The latter distribution, which has also been reported for protein-carbohydrate interactions [10], proved to be a good approximation for a system with a number as small as 10 non-specific binding sites and should be applicable to systems exhibiting similar behaviour.

II.A.4. Experimental

II.A.4.1. Materials

Rabbit muscle Creatine kinase (CK) was obtained from Roche and desalted using Micro Bio-Spin 30 columns (BioRad) with a 40 kDa MW cutoff equilibrated with 10 mM NH_4HCO_3 (pH 7.9), giving stock solutions of 200 μM protein. ADP and ATP were obtained from Sigma and desalted using a BioRad AG50WX8 ion-exchange resin. ADP solutions (10 mM in milliQ grade water) were stored at -20°C . ATP solutions were freshly prepared daily. The concentration of the ligand stock solution was adjusted to 200 μM with 10 mM NH_4HCO_3 (pH 7.9) prior to titration experiments.

II.A.4.2. Mass spectrometry

Experiments were performed on a Bruker APEX III FT-ICR mass spectrometer equipped with a 7.0 T actively shielded magnet and an unmodified Apollo electrospray source. Titration experiments were performed by preparing solutions containing 4 μM of CK and ligand concentration ranging from 1 μM to 60 μM in 10 mM NH_4HCO_3 . Each point was at least repeated three times. Samples were introduced by direct infusion electrospray ionization with a flow rate of 1.5 $\mu\text{L}/\text{min}$. Instrumental parameters were adjusted in order to optimize detection of high m/z species and to preserve the non-covalent interaction between the protein subunits and between the protein and its ligands, respectively. Various conditions were tried

(see Results and Discussion) and the typical changes as compared to standard instrumental conditions are as follows: increase of the capillary exit voltage (250 V), increase of the hexapole radiofrequency (rf) amplitude, increase of the excitation parameters and increase of the gated-trapping delay. The nebulizing gas temperature was left at 140°C.

ADP or ATP binding to CK results in a less than 1% change in the molecular mass of the protein. This might affect the local structure around the protein's active site, but is not assumed to significantly alter its ionization and detection efficiencies during the MS measurement (however, this assumption has to be validated for each protein-ligand system in question as some ligands may have a significant effect on the physicochemical properties of the protein). Under this assumption, the intensities in the mass spectrum reflect the relative concentrations in solution. The normalized intensities can thus be used to calculate the dissociation constant, K_i according to eqn. II.A.1 [9]:

$$K_i = \frac{[P]_{eq} [L]_{eq}^i}{[PL_i]_{eq}} = \frac{\left([L]_0 - [P]_0 \cdot \sum_{j=1}^i j \cdot R_j / \left(1 + \sum_{j=1}^i R_j \right) \right)^i}{R_i} \quad (\text{eqn. II.A.4})$$

Here, $[P]_{eq}$, $[L]_{eq}$ and $[PL_i]_{eq}$ are the solution phase equilibrium concentrations of protein, ligand and the complex, respectively. $[L]_0$ and $[P]_0$ are the initial concentrations of ligand and protein, respectively, i is the number of bound ligands and R_j is the ratio of the peak heights of the signals corresponding to the protein complex with j ligands and the free protein, respectively.

II.A.4.3. Deconvolution model

In the presence of non-specific interactions, the experimentally measured ratios R_i include contributions from both specific and non-specific complexes. They can be represented as a sum of different contributions (eqn. II.A.2):

$$R_i = \frac{I_i^0}{I_0^0} + \frac{I_i^1}{I_0^0} + \dots + \frac{I_i^i}{I_0^0} = R_i^0 + R_i^1 + \dots + R_i^i = \sum_{j=0}^i R_i^j \quad (\text{eqn. II.A.5})$$

where I_i^j and R_i^j are respectively the intensity and the intensity ratio for the species with i bound ligands of which j ligands are specifically bound. As the value for R_i contains

contributions from both specific and non-specific interactions, eqn. B.1 cannot be employed on them to determine the binding constants for the specific interactions. In order to obtain correct equilibrium binding constants, the sum of the specific contributions (which will be denoted R_j^*) has to be used (eqn. II.A.3):

$$R_j^* = \frac{I_j^j}{\sum_{i=0}^{\infty} I_i^0} + \frac{I_{j+1}^j}{\sum_{i=0}^{\infty} I_i^0} + \frac{I_{j+2}^j}{\sum_{i=0}^{\infty} I_i^0} + \dots = \frac{I_0^0}{\sum_{i=0}^{\infty} I_i^0} \cdot (R_j^j + R_{j+1}^j + R_{j+2}^j + \dots) = \frac{\sum_{i=j}^{\infty} R_i^j}{\sum_{i=0}^{\infty} R_i^0} \quad (\text{eqn. II.A.6})$$

Assuming statistical binding for both the specific and the non-specific interactions we can model the ligand binding as a convolution of two distributions. First, a binomial distribution on a limited number of specific binding sites (corresponding to specific ligand binding) and second, a statistical distribution on a large number of binding sites (corresponding to non-specific ligand binding), as similar Poisson-type distribution were observed for non-specific interactions under similar experimental conditions [10]. Denoting the average number of specifically occupied binding sites with $\bar{s}$ and considering that the ligand binding follows a binomial distribution, one can calculate the probability of the protein having j ligands bound specifically (p_j^{spec}) on a finite number of specific binding sites, s , relative to the probability of the protein having no specifically bound ligand (p_0^{spec}) according to eqn. II.A.4:

$$\frac{p_j^{\text{spec}}}{p_0^{\text{spec}}} = \frac{s!}{j! \cdot (s-j)!} \cdot \left(\frac{\bar{s}/s}{1-\bar{s}/s} \right)^j \quad (\text{eqn. II.A.7})$$

A similar approach can be followed for the non-specifically bound ligands: denoting the average number of non-specifically occupied binding sites with $\bar{n}$, the corresponding probability of the protein having k ligands (with $k = i - j$) bound non-specifically (p_k^{ns}) on a defined number of non-specific binding sites, n , relative to the probability of the protein having no non-specifically bound ligands can be calculated using eqn. II.A.5:

$$\frac{p_k^{\text{ns}}}{p_0^{\text{ns}}} = \frac{n!}{k! \cdot (n-k)!} \cdot \left(\frac{\bar{n}/n}{1-\bar{n}/n} \right)^k \approx \frac{\bar{n}^k}{k!} \quad (\text{eqn. II.A.8})$$

In this latter case, the number of non-specific binding sites cannot be assumed *a priori*. However, if the number of binding sites n is sufficiently large and if the ratio $\bar{n}/n$ is sufficiently small (which means that relatively few ligands occupy the potential binding sites), this limitation can be overcome by the approximation in the last part of eqn. II.A.5. This approximation is equivalent to assuming that non-specifically bound ligands follow a Poisson distribution. The validity of this approximation will be discussed in the “Results and Discussion” section.

The convolution between specific and non-specific contributions is given by the product of these probabilities in eqn. II.A.6:

$$R_i^j = \frac{p_j^{\text{spec}}}{p_0^{\text{spec}}} \frac{p_{i-j}^{\text{ns}}}{p_0^{\text{ns}}} = \frac{s!}{j!(s-j)!} \left(\frac{\bar{s}/s}{1 - \bar{s}/s} \right)^j \frac{n!}{(i-j)!(n-i+j)!} \cdot \left(\frac{\bar{n}/n}{1 - \bar{n}/n} \right)^{i-j} \quad (\text{eqn. II.A.9})$$

$$\approx \frac{s!}{j!(s-j)!} \left(\frac{\bar{s}/s}{1 - \bar{s}/s} \right)^j \frac{(\bar{n})^{i-j}}{(i-j)!}$$

In this expression, R_i^j only depends of four variables: s , the number of specific binding sites, $\bar{s}$, the average number of specifically bound ligands, n , the number of non-specific binding sites and $\bar{n}$, the average number of non-specifically bound ligands. The numbers of specific and non-specific binding sites are conditions that depend on the system under consideration: the first one (s) requires *a priori* knowledge or clear experimental evidence, the second one (n), being much more difficult to evaluate and usually not known *a priori*, has to be approximated somehow. This is the reason for which the Poisson approximation was used and proved useful as will be shown in the following section. $\bar{s}$ and $\bar{n}$ are related to the experimentally obtained average number of ligands $\bar{l}$ through equation II.A.7:

$$\bar{l} = \bar{s} + \bar{n} \quad (\text{eqn. II.A.10})$$

By combining equations II.A.2, II.A.6 and II.A.7, the experimentally measured ratios R_i can be expressed as a function of s , n , $\bar{l}$ and $\bar{s}$ (eqn II.A.8) and under the conditions stated above these ratios can be approximated as a function of s , $\bar{l}$ and $\bar{s}$.

$$R_i = \sum_{j=0}^i R_i^j = \sum_{j=0}^i \frac{s!}{j!(s-j)!} \left(\frac{\bar{s}/s}{1-\bar{s}/s} \right)^j \frac{n!}{(i-j)! \cdot (n-i+j)!} \cdot \left(\frac{(\bar{l}-\bar{s})/n}{1-(\bar{l}-\bar{s})/n} \right)^{i-j} \quad (\text{eqn. II.A.11})$$

$$\approx \sum_{j=0}^i \frac{s!}{j!(s-j)!} \left(\frac{\bar{s}/s}{1-\bar{s}/s} \right)^j \frac{(\bar{l}-\bar{s})^{i-j}}{(i-j)!}$$

$\bar{s}$ can be used as an adjustable parameter to fit the experimentally observed distribution at each ligand concentration used. Corrected ratios R_j^* are then obtained straightforward by reinserting the R_i^j obtained by eqn. II.A.6 into eqn. II.A.3. The corrected values for the ratios for the specifically bound ligands (R_j^*) obtained by this procedure can then be used in order to obtain the equilibrium dissociation constants by applying eqn. II.A.1.

II.A.5. Results and discussion

The mass spectrum of 4 μM CK in 10 mM NH_4HCO_3 buffer (pH 7.9) shows a series of peaks that can be attributed to a charge state distribution of the creatine kinase (CK) homodimer with charge states ranging from 22+ to 18+ (Fig. II.A.1). Thus the interaction between the monomers is conserved upon transfer of the complex to the gas phase. The mass as determined by the experiment is 85929 Da which is in good agreement with the theoretical average mass (85943 Da). Proteins analyzed by ESI-MS under denatured conditions exhibit large bell-shaped charge state distributions [21]. On the other hand, narrow charge state distributions at relatively high m/z , as observed in our experiments, are believed to reflect at least a non-denatured, if not native-like structure, as originally proposed by Chowdhury and coworkers [22].

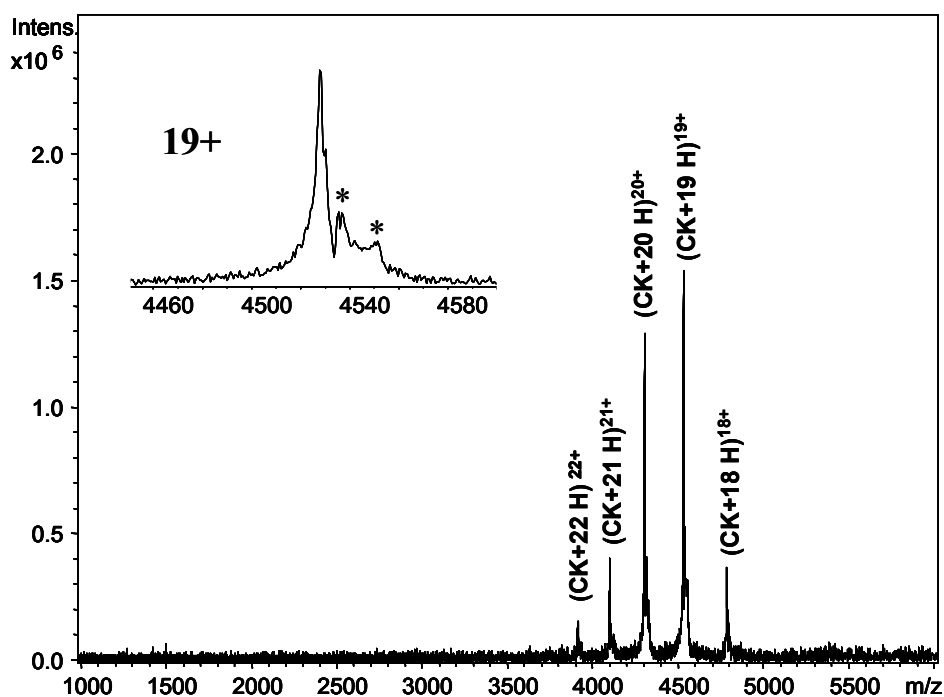


Fig. II.A.1: 4 μM CK in 10 mM NH_4HCO_3 . The inset shows the charge state 19+ with adducts that were not removed during desalting and the electrospray process (*).

When titrated with ADP in solution, ESI-MS of the resulting complexes yielded essentially the same charge states as in the case of the homodimer with additional peaks appearing that correspond to the ligand-bound CK (Fig. II.A.2). However, the ligand-bound forms of CK displayed a slightly higher charge state than the free protein, manifesting itself in a variation of ligand occupancy between different charge states in the spectrum (Fig. II.A.2, middle and bottom). The phenomenon of charge state variation between free and ligand-bound forms of the protein has been observed and described by Wang and coworkers [9] and seems to be a general phenomenon in ESI-MS analysis of noncovalent complexes [7, 9, 23-25].

Among the parameters that are critical to ensure that as little fragmentation as possible occur during the electrospray process are the potentials applied to accelerate and collide the ions with gas in the electrospray interface. A concern was that increasing the potential between the desolvation capillary exit and the first skimmer (called CapExit), a prerequisite in order to observe the ions, could lead to fragmentation of the protein-ligand complex. As shown in figure II.A.2, varying this potential on a range from 200 V to 400 V does not significantly alter the ratio of bound versus apo-complexes. In order to limit such effects nevertheless, the value chosen for the measurements (250 V) was the lowest that provided

sufficient S/N ratios. A similar series of measurements varying the desolvation gas temperature led to the conclusion to leave the typical temperature used in the interface (140°C) unchanged. The conservation of interactions in the protein-ligand complex over a broad range of conditions in the electrospray interface of the instrument indicates that there is no partial decomposition of the protein-ligand complex in the gas-phase induced by these conditions, although it does not completely rule out such decompositions occurring at other stages of the electrospray process.

As free and ligand-bound forms of CK displayed different charge state distributions, the R values in our spectra were calculated as the sum of peak heights (at all charge states) of the CK-ADP complexes in question, divided by the sum of peak heights (at all charge states) of the apo-protein (free CK), following the approach taken by Wang and coworkers [9]. As the intensity of ion signals in ICR spectra is proportional to the ion charge [26], this approach also included a correction of the intensities by dividing the signal intensity by the charge [9]. The intensities of the ligand-bound form of the protein had to be further corrected as their peak positions matched those of the adduct peak observed in the spectrum of the free protein (visualized by the right-hand star in the inset of Fig. II.A.1). It was assumed that this adduct peak was also present in the ligand-bound forms of the protein and did not influence ligand-binding.

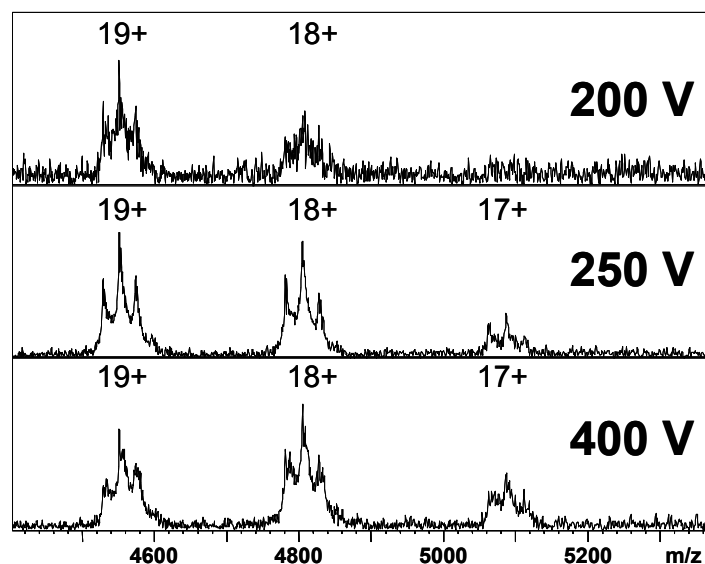


Fig. II.A.2: Dependence of interaction between CK (4 μ M) and ADP (20 μ M) on the CapExit voltage. The number above the peaks indicates the charge state, the CapExit voltage is shown on the right hand side of each spectrum.

When ADP was added at equimolar concentration to the CK solution, a maximum of two attached ligands was observed (Fig. II.A.3a). This is in agreement with the fact that CK has two active sites that can interact with ADP. When the ligand concentration was increased to 12 μM (corresponding to a 1:3 molar ratio of CK:ADP), a low intensity peak appeared corresponding to CK with three bound ADP (Fig. II.A.3b). At 60 μM ADP (a 1:15 molar ratio of CK:ADP) a maximum of four ADP were bound to CK (Fig. II.A.3c). These results could not simply be explained by a model with merely two specific binding sites.

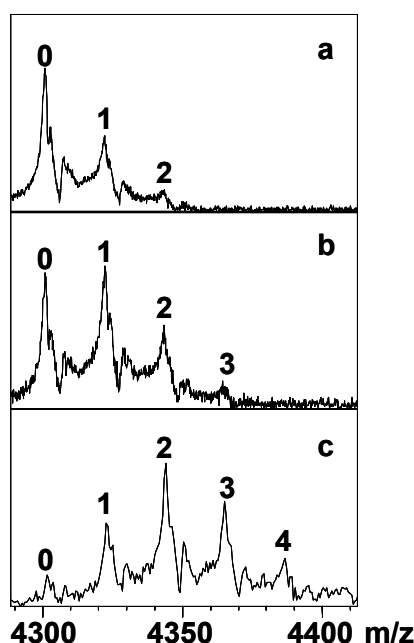


Fig. II.A.3: Titration of 4 μM CK with ADP. Numbers of ligands bound are indicated above the peaks. (a) 20+ charge state of 4 μM CK + 4 μM ADP, (b) 20+ charge state of 4 μM CK + 12 μM ADP, (c) 20+ charge state of 4 μM CK + 60 μM ADP.

The titration was performed with 4 μM CK and ligand concentrations ranging from 1 μM to 60 μM ADP (Table II.A.1). Experiments were performed three times in order to check for variation between individual measurements as has been reported for nanoESI-MS of protein-carbohydrate complexes [9]. For the CK-ADP interaction, R_i values did not show substantial differences between individual experiments (data not shown), as indicated by the relative low standard deviation (Table II.A.1). The experimental error in the measure of the R_i values increased with ligand concentration, which was expected as the intensity of the free protein decreases with increasing ligand concentration, increasing the error in calculation of the R_i values.

Table II.A.1: Experimental ratios and average number of bound ligands for the system CK+ADP. Values as well as errors correspond to the average of three individual measurements.

[ADP] ₀ / μM	R_1	R_2	R_3	R_4	$\bar{l}$
1	0.06 ± 0.04	-	-	-	0.05 ± 0.03
2	0.14 ± 0.05	0.01 ± 0.01	-	-	0.13 ± 0.05
4	0.31 ± 0.06	0.02 ± 0.02	-	-	0.27 ± 0.06
6	0.45 ± 0.04	0.09 ± 0.02	-	-	0.41 ± 0.03
8	0.56 ± 0.03	0.14 ± 0.02	-	-	0.49 ± 0.04
10	0.74 ± 0.05	0.22 ± 0.03	-	-	0.60 ± 0.04
12	0.80 ± 0.04	0.29 ± 0.05	0.02 ± 0.01	-	0.69 ± 0.05
16	1.13 ± 0.04	0.57 ± 0.09	0.10 ± 0.01	-	0.92 ± 0.04
20	1.33 ± 0.05	0.70 ± 0.15	0.16 ± 0.03	-	1.01 ± 0.06
40	2.73 ± 0.25	2.73 ± 0.44	0.87 ± 0.17	0.24 ± 0.05	1.55 ± 0.09
60	3.61 ± 0.85	5.02 ± 1.24	2.20 ± 0.90	0.90 ± 0.09	1.87 ± 0.16

A possible explanation for these data would be the occurrence of some non-specific binding added to the two specific binding sites present on the protein as known from other work on this system [15,19]. However, it is not possible to corroborate this hypothesis by visual inspection of the experimental data alone.

In order to investigate this hypothesis, a first element is to rule out that the binding stemmed from entirely statistical binding on more than two sites (which means that there are not as expected two specific sites with a larger affinity than other sites on the protein). The experimentally obtained R_i values at each ligand concentration were compared to a statistical distribution on an unknown number of binding sites that can be calculated according to eqn. (5). The result shows that the experimental data cannot be interpreted as non-specific interactions alone (Fig. II.A.4a, closed triangles, Fig. II.A.4b, white bars).

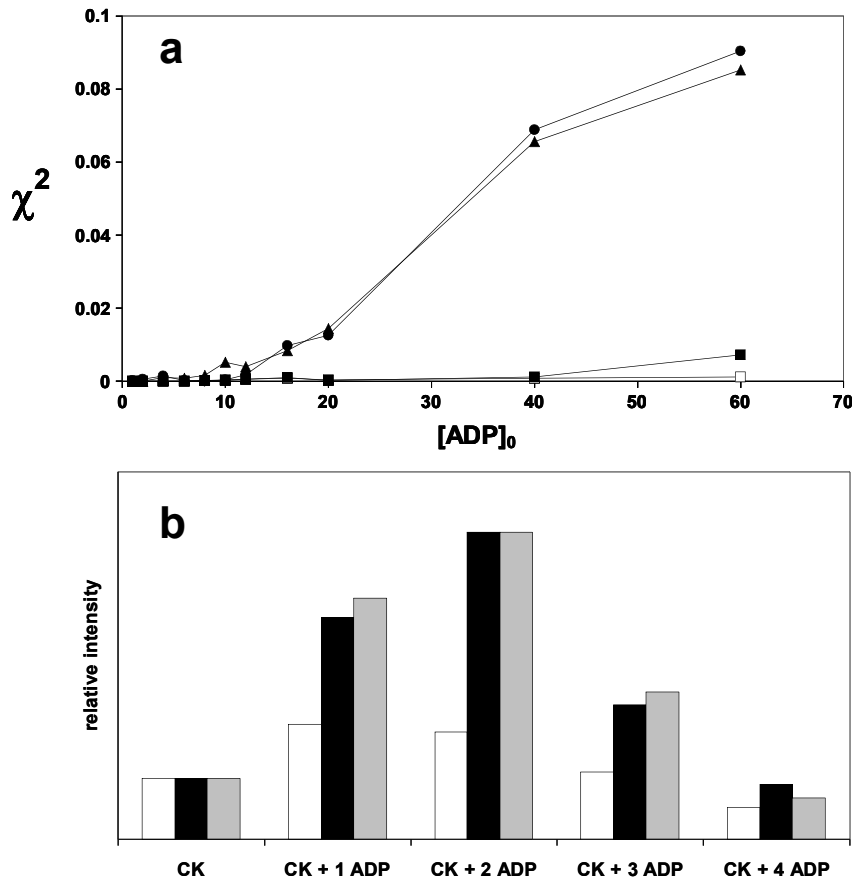


Fig. II.A.4: (a) Deviation of experimentally obtained spectra from spectra obtained by fitting with different models (closed circles: only specific interactions with two specific binding sites are considered [eqn. 4]; closed triangles: only non-specific interactions on a large number of binding sites are considered [eqn. 5]; closed squares: model based on two specific and two non-specific interaction sites [eqn. 8, first part, with $s=2$ and $n=2$]; open squares: model based on two specific and a large number of non-specific interaction sites [eqn. 8, last part, with $s=2$]). At a given concentration of ADP, the χ^2 value represents the sum of the squares of the differences between experimental and calculated intensities divided by the total intensity. (b) The dataset 4 μM CK + 60 μM ADP fitted with different models (black bars: experimental data, white bars: only non-specific interaction sites taken into account; grey bars: two specific and a large number of non-specific interaction sites taken into account).

The experimental data should thus be considered to represent a superposition of at least two different types of protein-ligand interactions. The simplest hypothesis is to assume that one can be related to the specific binding of ligands in the active site of the enzyme and the other to non-specific binding of ligands probably attached during the electrospray process. It is of interest to separate the contributions of these two types of interactions, since only the

specific binding to the catalytic site has to be taken into account for the evaluation of the solution phase binding constants. The model developed in the “Experimental” section, which should be general enough to be applied to other cases for which non-specific binding is observed, is based on the deconvolution of two statistical distributions, the first (named “specific”) limited to a finite number of binding sites known *a priori*, the second (named “non-specific”) can either be applied to a defined number of binding sites or to an unknown number of binding sites in the case where this number is large enough and provided that there are few ligands bound to these sites.

As for the specific interactions, CK has been shown to be a protein for which the two binding sites are equivalent and independent under the experimental conditions employed in our study [17, 19, 27]. The distribution of ligands in the binding sites can therefore be described by a binomial distribution on two sites according to eqn. 4.

As for the nonspecific fraction of the interaction, the origin is less clear: there is some evidence that it could be related to the electrospray process used for the production of ions [7, 10]. Considering the charge residue model (CRM) originally proposed by Dole and coworkers [28] that is likely to be valid for the ion formation in the electrospray source for large biomolecules [29], one could assume aggregation of ligands to the protein in the course of the evaporation of a finite size droplet in the electrospray source in the final stages of the droplet evaporation process. Considering that the ligand concentration is small enough, the Poisson distribution is a good approximation for analyzing the resulting data. Similar Poisson distributions have been observed for purely non-specific interactions [10].

If the non-specific binding occurred in solution by the interaction of the ligands with the surface of the protein prior to the electrospray process, there could be a limited number of available binding sites. However, there is little evidence for means to determine this number of sites. Increasing the concentration of ligand above 80 μM leads to the observation of a fifth and sixth ligand binding, which shows that at least four supplementary binding sites should be considered.

A model based on the presence of 2 specific and of only 2 non-specific binding sites (Fig. II.A.4a, closed squares) gave a much better fit of the experimental data than with non-specific interactions alone. However, at ligand concentrations above 40 μM , where non-specific binding is most likely to occur, this model gave still a poor fit. An increase in the number of non-specific binding sites to 10 resulted in a fit (data not shown) that was of the same quality than the fit obtained by applying the model based on 2 specific binding sites and the Poisson approximation for non-specific binding (Fig. II.A.4a, open squares): these two models differ

by less than 2% on the values of the binding constants. In light of the low occupancy of the ligand binding sites ($\bar{l} = 1.87$ at 60 μM), the Poisson distribution as employed in the deconvolution model is thus a good approximation for the non-specific binding, with the advantage that using a Poisson distribution does not require an *a priori* knowledge of the number of ligands that are non-specifically bound.

The deconvolution model detailed in the “Experimental” section has therefore been applied to the experimental data obtained for the CK+ADP system with $s = 2$ and with the Poisson distribution approximation for the non-specific binding. The $\bar{s}$ parameter was adjusted so that the R_i values given by eqn. II.A.8 fitted the experimental ratios given in Table II.A.1. Once this is done, one can apply directly equation II.A.3 as outlined in the experimental section in order to calculate the fitted ratios (Table II.A.2). $K_{1,\text{diss}}$ and $K_{2,\text{diss}}$ are the sequential dissociation constants for the equilibria $\text{CK-ADP} \rightleftharpoons \text{CK} + \text{ADP}$ and $\text{CK-ADP}_2 \rightleftharpoons \text{CK-ADP} + \text{ADP}$, respectively.

However, application of the deconvolution model leads to values of $K_{1,\text{diss}}$ and $K_{2,\text{diss}}$ that are, by inherent definitions of the model (non-cooperative binding), in a fixed ratio of 4, thus not allowing for any analysis of the validity of the model. The quality of the results was thus evaluated by partially deconvoluting the experimental ratios, removing the non-specific contribution ($\bar{n}$) from the experimental ratios (bearing in mind that, for instance, $R_1 = R_1^1 + \bar{n}$). $K_{1,\text{diss}}$ and $K_{2,\text{diss}}$ were then calculated based on these corrected experimental ratios (Table II.A.2, values in parenthesis) instead of using the purely fitted ratios. Within this latter approach, a flaw in the assumptions of the model would manifest itself by significant deviations from expected outcomes. For example, cooperativity between both sites (being contradictory to the assumptions underlying our model) would lead to a significant deviation in the ratio $K_{2,\text{diss}}/K_{1,\text{diss}}$, that is expected to be 4 in the complete absence of cooperativity. The values presented in parenthesis in table II.A.2 are obtained by following the latter approach.

Table II.A.2: Data obtained after deconvolution of specific and non-specific interactions. Values in parenthesis for R_1^* , R_2^* , $K_{1,diss}$ and $K_{2,diss}$ are those used to check the consistency of the model (see text). Errors for R_1^* , R_2^* , $K_{1,diss}$ and $K_{2,diss}$ are derived from equations 6, 3 and 1 assuming that the error on $\bar{s}$ and $\bar{n}$ is equal to the error on $\bar{l}$.

[ADP] ₀ / μM	$\bar{s}$	$\bar{n}$	R_1^*	R_2^*	$K_{1,diss}$ / μM	$K_{2,diss}$ / μM	$K_{2,diss} / K_{1,diss}$
1	0.05	-	0.05 (0.06) $\pm$ 0.03	-	14.4 (14.0) $\pm$ 34.0	-	-
2	0.12	0.00(5)	0.13 (0.13) $\pm$ 0.06	0.004 (0.004) $\pm$ 0.002	11.3 (11.2) $\pm$ 8.4	-	-
4	0.26	0.00(3)	0.30 (0.30) $\pm$ 0.10	0.02 (0.02) $\pm$ 0.01	9.7 (9.7) $\pm$ 3.5	-	-
6	0.27	0.14	0.36 (0.35) $\pm$ 0.06	0.03 (0.04) $\pm$ 0.01	14.0 (13.7) $\pm$ 2.5	56 (47) $\pm$ 55	4.0 (3.4)
8	0.33	0.16	0.46 (0.46) $\pm$ 0.09	0.05 (0.06) $\pm$ 0.01	14.3 (14.0) $\pm$ 1.3	57 (48) $\pm$ 22	4.0 (3.4)
10	0.46	0.14	0.69 (0.68) $\pm$ 0.11	0.10 (0.12) $\pm$ 0.02	11.9 (11.7) $\pm$ 1.1	48 (44) $\pm$ 19	4.0 (3.7)
12	0.43	0.26	0.71 (0.70) $\pm$ 0.14	0.10 (0.15) $\pm$ 0.02	14.7 (14.1) $\pm$ 1.2	59 (46) $\pm$ 27	4.0 (3.3)
16	0.55	0.37	1.10 (1.09) $\pm$ 0.15	0.21 (0.30) $\pm$ 0.03	12.6 (12.1) $\pm$ 0.6	50 (48) $\pm$ 22	4.0 (4.0)
20	0.66	0.34	1.4 (1.4) $\pm$ 0.3	0.34 (0.41) $\pm$ 0.07	12.4 (12.1) $\pm$ 0.7	50 (57) $\pm$ 36	4.0 (4.7)
40	1.10	0.45	3.8 (3.6) $\pm$ 1.0	2.34 (2.48) $\pm$ 0.64	9.2 (9.9) $\pm$ 0.8	37 (50) $\pm$ 16	4.0 (5.1)
60	1.25	0.63	6.3 (5.6) $\pm$ 3.1	5.2 (5.4) $\pm$ 2.6	8.9 (9.8) $\pm$ 1.8	35 (56) $\pm$ 29	4.0 (5.7)

It is of interest to notice that the individual contribution of each different type of species (specific and non-specific) can be computed according to equation II.A.6. This allows giving an example of the deconvolution process based on the 4 μM CK + 60 μM ADP experimental data (Figure II.A.5). Each experimental peak can be divided into a sum of different contributions (black: no specifically bound ligand, R_i^0 ; grey: one specifically bound ligand, R_i^1 ; white: two specifically bound ligands, R_i^2). This figure also shows that the non-specific binding accounts for about 50 % of the protein – two ligand intensity. This would have led to a completely underestimated dissociation constant if the deconvolution model had not been used.

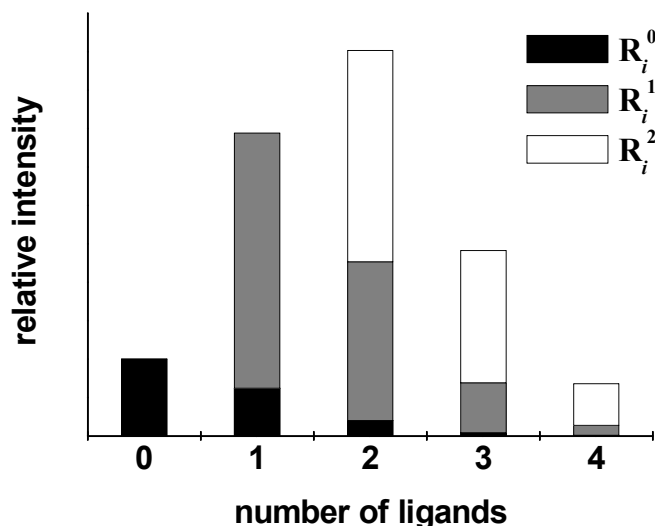


Fig. II.A.5: Contributions of species with zero specifically bound ligands (black), one specifically bound ligand (grey) and two specifically bound ligands (white) to the experimental intensities of 4 μM CK incubated with 60 μM ADP.

The data obtained by applying the deconvolution model to the experimental data are in good agreement with the experimental data (Fig. II.A.4a, closed squares, Fig. II.A.4b grey bars), especially at higher concentrations where the fraction of non-specific binding is most likely to increase. The average dissociation constants calculated from the fitted data in the ligand range from 1 μM to 60 μM ADP are $K_{1,\text{diss,ADP}} = 11.8 \pm 1.5 \mu\text{M}$ and $K_{2,\text{diss,ADP}} = 48 \pm 6 \mu\text{M}$, respectively. The ratio of $K_{2,\text{diss,ADP}}/K_{1,\text{diss,ADP}}$ is of course of 4 in the case where the model is strictly applied as expected for two equivalent and independent binding sites according to the multi-equilibrium formalism for non-cooperative binding sites [9,30]. In parenthesis, one can see that the ratio does not deviate significantly from 4 within the error margin when one simply removes non-specific contributions, removing some constraints from the model. The poor quality of the measurement of $K_{2,\text{diss}}$ at the highest concentrations arises from the fact that the intensity of the peak corresponding to the apo-protein decreases with increasing concentrations, leading to an increase in the error on the measurement of the ratios for individual measurements. However, as stated above, R_i values did not vary dramatically between different series of experiments. The K_d values obtained by our method are of the same order of magnitude as the reported literature values of about $K_1=50 \mu\text{M}$ and $K_2=200 \mu\text{M}$ that were obtained in the absence of other ligands than the one in question [17,20]. Different experimental conditions very probably account for the remaining difference with respect to the literature values. The data obtained by applying our model are in agreement with two non-

cooperative binding sites of CK for ADP under the experimental conditions employed: the model assumes non-cooperativity between the two sites, but as discussed above, a deviation of the ratio $K_{2,diss,ADP}/K_{1,diss,ADP}$ could be expected if the model was not valid. This does not completely rule out the presence of weak cooperativity between the sites since in this case the deviation from our non-cooperative model would be well below the experimental error margin. However, a strong cooperativity between the binding sites would have led to a poor fit with significant deviations. There might also be other experimental conditions (e.g. solution composition) where the sites behave in a cooperative manner [17].

The non-specific binding, as characterized by $\bar{n}$, increases with the ligand concentration and contributes to 25-40% (from as low as 6 μ M ADP concentration) of the total binding. Interestingly, already at a protein:ligand ratio of 1:3, the non-specific interactions contribute to 38% of the overall ligand binding (Table II.A.2). In the spectrum, the experimental indication for non-specific binding to occur is only a small peak corresponding to CK with three bound ligands (Fig. II.A.3b). Our result shows that non-specific interactions contribute also to the peak corresponding to CK with one and two ligands bound, respectively, and can thus not be neglected. Therefore, already at relatively low ligand concentration, non-specific binding should be taken into consideration.

As the model was built around the CK/ADP system, we subsequently aimed at studying the CK/ATP interaction in order to further validate our approach. ATP is a second physiological ligand of CK and as it is chemically similar to ADP, it was expected to show similar behaviour during the ESI experiments. The ESI mass spectra of CK titrated with ATP show the same charge state distribution as the uncomplexed dimer and exhibit no dissociation of the ligand upon varying the capillary exit voltage (data not shown). In the mass spectrum, peaks appeared that were attributed to ATP-bound CK (Fig. II.A.6). For a given ligand concentration, the extent of ATP binding to the protein was significantly lower than the extent of ADP binding, as indicated by the lower abundance of peaks attributed to ligand-bound protein. As in the case of ADP, upon increasing the concentration of ATP, observation of a peak corresponding to CK having three bound ATP provides clear experimental evidence for nonspecific binding (Fig. II.A.6c).

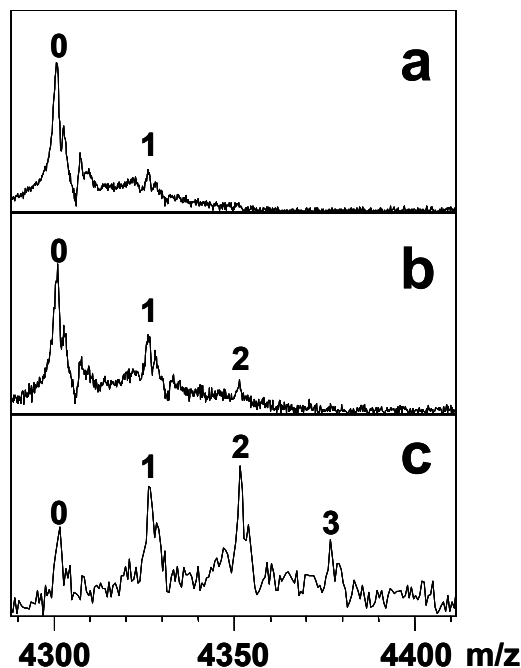


Fig. II.A.6: Titration of 4 μM CK with ATP. Numbers of ligands bound are indicated above the peaks. (a) 20+ charge state of 4 μM CK + 4 μM ATP, (b) 20+ charge state of 4 μM CK + 12 μM ATP, (c) 20+ charge state of 4 μM CK + 60 μM ATP.

The data obtained upon titration of CK with ATP were analyzed by applying our deconvolution model (Tab. II.A.3). For each data point, the model gave a good fit within the experimental error range (data not shown). The final data argue, as was the case for ADP, for the presence of nonspecific interactions from as low as 12 μM on, where there is no experimental evidence for this binding mode (Tab. II.A.3, Fig. II.A.6).

The binding constants for the CK/ATP interaction obtained after application of our model are $K_{1,\text{diss,ATP}} = 27 \pm 7 \mu\text{M}$ and $K_{2,\text{diss,ATP}} = 114 \pm 27 \mu\text{M}$, respectively, assuming that there is no cooperativity. The constants obtained by the MS approach are thus significantly smaller than those reported in the literature ($K_{\text{diss}} = 300 - 500 \mu\text{M}$). These differences are most probably due to the different experimental setup and conditions employed. However, MS analysis and deconvolution of specific and nonspecific interactions display the correct qualitative order of the interaction constants ($K_{\text{diss,ADP}} < K_{\text{diss,ATP}}$).

Table II.A.3: Experimental and fitted data obtained after titration of CK with ATP.

[ATP] ₀ / μM	R ₁	R ₂	R ₃	R ₄	$\bar{l}$
1	0.03 $\pm$ 0.03	-	-	-	0.03 $\pm$ 0.03
2	0.09 $\pm$ 0.02	-	-	-	0.08 $\pm$ 0.02
4	0.16 $\pm$ 0.04	-	-	-	0.14 $\pm$ 0.03
6	0.22 $\pm$ 0.02	-	-	-	0.18 $\pm$ 0.05
8	0.35 $\pm$ 0.03	0.02 $\pm$ 0.02	-	-	0.29 $\pm$ 0.03
10	0.40 $\pm$ 0.10	0.04 $\pm$ 0.02	-	-	0.34 $\pm$ 0.07
20	0.72 $\pm$ 0.09	0.23 $\pm$ 0.03	-	-	0.60 $\pm$ 0.04
30	0.87 $\pm$ 0.15	0.33 $\pm$ 0.09	-	-	0.69 $\pm$ 0.08
40	1.08 $\pm$ 0.29	0.61 $\pm$ 0.14	0.14 $\pm$ 0.03	-	0.96 $\pm$ 0.08
60	1.64 $\pm$ 0.37	1.29 $\pm$ 0.36	0.36 $\pm$ 0.18	0.10 $\pm$ 0.01	1.30 $\pm$ 0.16

[ATP] ₀ / μM	$\bar{s}$	$\bar{n}$	R_1^*	R_2^*	K _{1,diss} / μM	K _{2,diss} / μM
1	0.03	0.00	0.03 $\pm$ 0.03	-	-	-
2	0.08	0.00	0.08 $\pm$ 0.02	-	20 $\pm$ 6	-
4	0.14	0.00	0.15 $\pm$ 0.04	-	24 $\pm$ 7	-
6	0.18	0.00	0.20 $\pm$ 0.08	0.01 $\pm$ 0.004	26 $\pm$ 11	104 $\pm$ 45
8	0.29	0.00	0.34 $\pm$ 0.06	0.02 $\pm$ 0.005	20 $\pm$ 4	82 $\pm$ 15
10	0.32	0.02*	0.39 $\pm$ 0.13	0.04 $\pm$ 0.01	23 $\pm$ 8	90 $\pm$ 33
20	0.43	0.17	0.66 $\pm$ 0.11	0.09 $\pm$ 0.01	28 $\pm$ 4	111 $\pm$ 18
30	0.52	0.18	0.83 $\pm$ 0.23	0.14 $\pm$ 0.04	34 $\pm$ 9	135 $\pm$ 35
40	0.44	0.52	0.95 $\pm$ 0.28	0.13 $\pm$ 0.04	40 $\pm$ 11	162 $\pm$ 45
60	0.68	0.62	1.91 $\pm$ 0.96	0.49 $\pm$ 0.25	30 $\pm$ 14	120 $\pm$ 57

The data obtained allow for direct comparison of the average number of nonspecific interactions in either case. A simple plot of $\bar{n}$ versus $[\text{L}]_{eq}$ (L = ADP, ATP) shows that the two ligands show a similar behaviour in the experimental error range (Fig. II.A.7). In either case, the average number of nonspecific interactions increases with increasing ligand concentration.

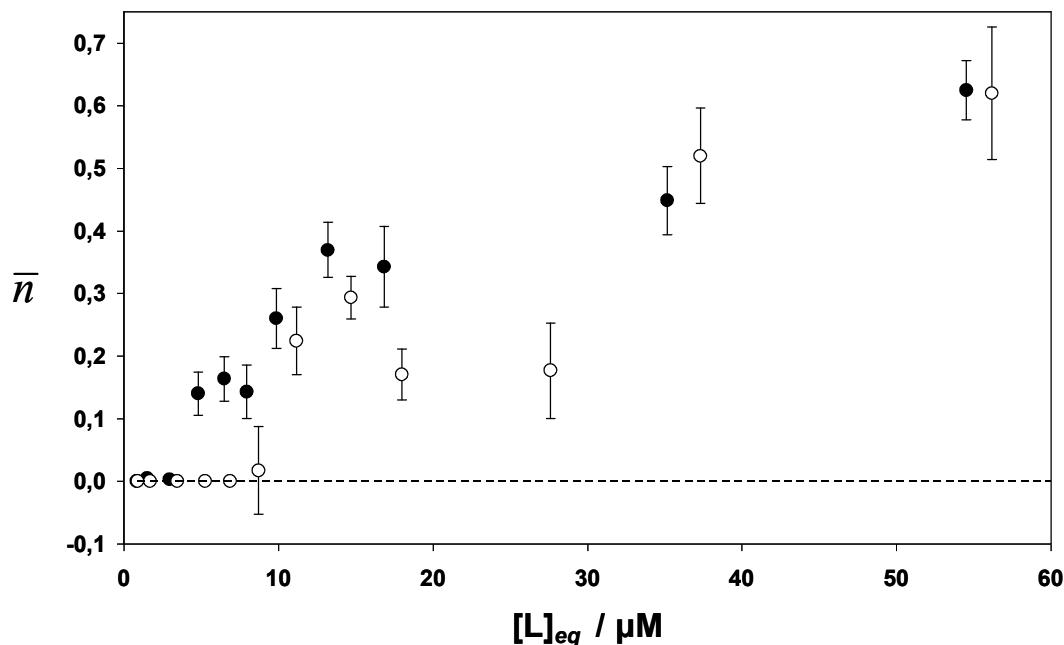


Fig. II.A.7: Average number of nonspecific interactions as a function of free ligand concentration for the CK/ADP (closed circles) and CK/ATP equilibria (open circles), respectively.

These data show another important result obtained by the application of our deconvolution model: the degree of nonspecific binding is less sensitive to the kind of ligand than the specific binding mode. As ADP and ATP differ by only one phosphate group, they are not expected to behave very differently with respect to the non-specific binding mode. The specific binding, on the other hand, varies significantly between the two ligands (Tab. II.A.2 and II.A.3), and thus is very likely to reflect the solution binding affinity.

As for the origin of the non-specific interactions, there are two possible sources: they may be generated during the electrospray process or they can already be present in solution. In either case, simple models for the formation of these non-specific adducts (droplet evaporation or multiple equilibria with many sites and identical binding constants) lead to expect linear dependence of the average number of bound ligands $\bar{n}$ versus the free ligand concentration [10]. However, the relatively large error in the value of $\bar{n}$ does not allow for a detailed discussion of this point. It is thus not generally possible to address the question of the origin or the exact nature of the non-specific interactions. These elements might already point to the limits of the modeling of the non-specific interactions as presented in our deconvolution model. In order to precisely take into account any possible process leading to the formation of

non-specific interactions in the deconvolution model, more adjustable parameters would be required. However, a deconvolution model with only one adjustable parameter, as employed in the present study, seemed to be reasonable in light of the experimental data.

II.A.6. Conclusions

The study of non-covalent interactions by electrospray mass spectrometry is hampered by the presence of non-specific interactions already present in solution or formed during the electrospray process [10]. The propensity of forming non-specific interactions increases with ligand concentration. Thus, by using the titration method [9] one should work with low ligand concentration to minimize the influence of non-specific binding. However, when studying systems with large dissociation constants (10^{-5} to 10^{-3} M) it is necessary to increase ligand concentration above a certain threshold value in order to observe the formation of complexes. Using the deconvolution model presented here it should be possible to increase the range of ligand concentration about ten-fold. The approach taken by us allows for separation of specific and non-specific interactions. The model allows to calculate equilibrium constants containing only the contribution from specific interactions. The model is limited to systems for which there is initial knowledge of the specific interactions that are expected. In its present form, it is limited to systems with binding sites that show no cooperativity, i.e. that are equivalent and independent. By introducing the ratio of constants as an additional parameter, it would be possible to extend the model for systems that show cooperativity. However, with an increasing number of adjustable parameters, the quality of the MS data (especially the S/N ratio) becomes even more important. Fitting experimental data to a model containing only one adjustable parameter (in our case the average number of specifically bound ligands, $\bar{s}$) does not impose this constraint.

II.A.7. References

- [1] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 464-471.
- [2] Loo, J. A. Electrospray Ionization Mass Spectrometry: a Technology for Studying Noncovalent Macromolecular Complexes. *Int. J. Mass Spectrom.* **2000**, *200*, 175-186.
- [3] Veenstra, T. D. Electrospray Ionization Mass Spectrometry in the Study of Biomolecular Non-Covalent Interactions. *Biophys. Chem.* **1999**, *79*, 63-79.
- [4] Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. Quantitative Determination of Noncovalent Binding Interactions using Soft Ionization Mass Spectrometry. *Int. J. Mass Spectrom.* **2002**, *216*, 1-27.
- [5] Smith, R. D. Evolution of ESI-Mass Spectrometry and Fourier Transform Ion Cyclotron Resonance for Proteomics and other Biological Applications. *Int. J. Mass Spectrom.* **2000**, *200*, 509-544.
- [6] Heck, A. J. R.; Jørgensen, T. J. D. Vancomycin in Vacuo. *Int. J. Mass Spectrom.* **2004**, *236*, 11-23.
- [7] Robinson, C. V.; Chung, E. W.; Kragelund, B. B.; Knudsen, J.; Aplin, R. T.; Poulsen, F. M.; Dobson, C. M. 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.
- [8] Peschke, M.; Verkerk, U. H.; Kobarle, P. Features of the ESI Mechanism that Affect the Observation of Multiply Charged Noncovalent Protein Complexes and the Determination of the Association Constant by the Titration Method. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1424-1434.
- [9] Wang, W.; Kitova, E. N.; Klassen, J. S. Determination of Protein-Oligosaccharide Binding by Nanoelectrospray Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry. *Meth. Enzym.* **2003**, *362*, 376-396.
- [10] Wang, W.; Kitova, E. N.; Klassen, J. S. Nonspecific Protein-Carbohydrate Complexes Produced by Nanoelectrospray Ionization. Factors Influencing Their Formation and Stability. *Anal. Chem.* **2005**, *77*, 3060-3071.
- [11] Wang, W.; Kitova, E. N.; Sun, J.; Klassen, J. S. Blackbody Infrared Radiative Dissociation of Nonspecific Protein-Carbohydrate Complexes Produced by Nanoelectrospray Ionization: The Nature of the Noncovalent Interactions. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1583-1594.

- [12] Pinske, M. W. H.; Heck, A. J. R.; Rumpel, K.; Pullen, F. Probing Noncovalent Protein-Ligand Interactions of the cGMP-Dependent Protein Kinase Using Electrospray Ionization Time of Flight Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1392-1399.
- [13] Sundquist, G.; Benkestock, K.; Roeraade, J. Investigation of Multiple Binding Sites on Ribonuclease A Using Nano-Electrospray Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1011-1016.
- [14] Benkestock, K.; Edlund, P.-O.; Roeraade, J. Electrospray Ionization Mass Spectrometry as a Tool for Determination of Drug Binding Sites to Human Serum Albumin by Noncovalent Interaction. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1637-1643.
- [15] Watts, D. C. Creatine Kinase (Adenosine 5'-Triphosphate-Creatine Phosphotransferase). In *The Enzymes, Vol. VIII*. Boyer, P. D., ed. Academic Press: New York and London, 1973, pp 383-455.
- [16] Borders, C. L.; Snider, M. J.; Wolfenden, R.; Edminston, P. L. Determination of the Affinity of Each Component of a Composite Quaternary Transition-State Analogue Complex of Creatine Kinase. *Biochemistry* **2002**, *41*, 6995-7000.
- [17] McLaughlin, A. C. The Interaction of 8-Anilino-1-naphtalenesulfonate with Creatine Kinase. *J. Biol. Chem.* **1974**, *249*(5), 1445-1452.
- [18] Burbaum, J. J.; Knowles, J. R. Internal Thermodynamics of Enzymes Determined by Equilibrium Quench: Values of K_{int} for Enolase and Creatine Kinase. *Biochemistry* **1989**, *28*, 9306-9317.
- [19] Hornemann, T.; Rutishauser, D.; Wallimann, T. Why Is Creatine Kinase a Dimer? Evidence for Cooperativity Between the Two Subunits. *Biochim. Biophys. Acta* **2000**, *1480*(1-2), 365-373.
- [20] Kuby, S. A.; Mahowald, T. A.; Noltmann, E. A. Studies on Adenosine Triphosphate Transphosphorylases. IV. Enzyme-Substrate Interactions. *Biochemistry* **1962**, *1*, 748-762.
- [21] Loo, J. A.; Ogorzalek Loo, R. G. Electrospray Ionization Mass Spectrometry of Peptides and Proteins. In *Electrospray Ionization Mass Spectrometry*. Cole, R. B., ed. John Wiley and Sons, Inc.: New York, 1997, pp 385-419.
- [22] Chowdhury, S. K.; Katta, V.; Chait, B. T. Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* **1990**, *112*, 9012-9013.
- [23] Loo, J. A.; Hu, P.; McConnell, P.; Mueller, W. T.; Sawyer, T. K.; Thanabal, V. A Study of Src SH2 Domain Protein-Phosphopeptide Binding Interactions by Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 234-243.

- [24] Greig, M. J.; Gaus, H.; Cummins, L. L.; Sasmor, H.; Griffey, R. H. Measurement of Macromolecular Binding Using Electrospray Mass Spectrometry. Determination of Dissociation Constants for Oligonucleotide-Serum Albumin Complexes. *J. Am. Chem. Soc.* **1995**, *117*, 10765-10766.
- [25] Gabelica, V.; De Pauw, E.; Rosu, F. Interaction between Antitumor Drugs and a Double-stranded Oligonucleotide Studied by Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **1999**, *34*, 1328-1337.
- [26] Jurchen, J. J.; Garcia, D. E.; Williams, E. R. Further studies on the origins of asymmetric charge partitioning in protein homodimers. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1408-1415.
- [27] Forstner, M.; Kriechbaum, M.; Laggner, P.; Wallimann, T. Structural Changes of Creatine Kinase upon Substrate Binding. *Biophys. J.* **1998**, *75*, 1016-1023.
- [28] Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular beams of macroions, *J. Chem. Phys.* **1968**, *49*, 2240-2249.
- [29] Kebarle, P.; Ho, Y. On the Mechanism of Electrospray Mass Spectrometry. In *Electrospray Ionization Mass Spectrometry*. Cole, R. B., ed. John Wiley and Sons, Inc.: New York, 1997, pp.3-63.
- [30] Tanford, C. Multiple Equilibria. In *Physical Chemistry of Macromolecules*. Tanford, C., ed. John Wiley and Sons, Inc.: New York and London, 1961, pp. 526-586.

Chapitre II.B

Influence des paramètres
expérimentaux sur les constantes de
dissociation

II.B.1. Introduction

Après avoir établi la part d'interactions spécifiques et la part d'interaction non spécifiques dans l'étude par spectrométrie de masse des complexes CK/ATP et CK/ADP, nous avons pu calculer les constantes de dissociation associées à chaque complexe. En comparant ces valeurs avec celles trouvées dans la littérature, nous avons noté des différences importantes, ce dont nous discuterons plus loin. Il faut cependant noter qu'il existe également une grande variabilité dans les constantes relevées dans la littérature. Un examen plus approfondi des expériences ayant permis de déterminer ces constantes nous a montré que les techniques utilisées étaient assez variées, mais aussi que pour une même technique, les conditions de pH, de tampon, de concentrations en différents sels et de température pouvaient varier considérablement. L'utilisation de la spectrométrie de masse n'autorise qu'une faible variabilité dans les concentrations en sels non volatils ainsi que dans la nature des tampons utilisés. Il nous a cependant paru intéressant d'étudier l'impact de modifications dans les conditions expérimentales utilisées pour l'analyse des systèmes CK/ATP et CK/ADP pour mieux situer les effets pouvant influencer les valeurs de constantes, et pour valider l'utilisation de la spectrométrie de masse pour l'étude des interactions non covalentes.

Dans un premier temps, l'effet du tampon a été étudié en changeant la nature des espèces utilisées. Dans ce contexte, il est important de rappeler que le bicarbonate d'ammonium n'est pas le seul tampon permettant l'étude des interactions non covalentes par spectrométrie de masse. La stabilité des protéines qui contiennent plusieurs sous-unités peut dépendre du tampon utilisé. Par exemple, il a été montré que la stabilité du tétramère de l'hémoglobine dans des expériences de spectrométrie de masse est considérablement améliorée en utilisant le bicarbonate de triéthylammonium (TEAB) au lieu du bicarbonate d'ammonium (AB) [1]. L'étude des interactions CK-ADP et CK-ATP a ainsi été menée dans différents tampons pour rechercher une influence du tampon sur l'interaction entre la protéine et ses ligands. Ceci nous a amené à étudier également les effets du pH sur les constantes d'association, pour finalement comparer les différentes valeurs obtenues à celles déjà mesurées dans la littérature à la lumière de ces données supplémentaires.

II.B.2. Interaction entre la CK et ses ligands dans différents tampons

II.B.2.1. Effet sur les états de charge

L'interaction entre la CK et ses ligands ADP et ATP a été étudiée dans différents tampons pour analyser l'influence du tampon sur l'interaction protéine-ligand. Les études ont été effectuées dans les tampons suivants : bicarbonate d'ammonium (AB), bicarbonate de triéthylamine (TEAB), acétate d'ammonium (AA), acétate de triéthylamine (TEAA) et formiate de triéthylamine (TEAF). La concentration des tampons était de 10 mM, le pH a été ajusté à pH 8.

Dans un premier temps, on peut déjà simplement regarder pour la protéine seule l'effet du changement de tampon sur les états de charge (Fig. II.B.1) : il apparaît nettement que l'état de charge moyen dans les tampons AB et AA (autour de 19+) est plus élevé que pour le TEAB, le TEAA ou le TEAF (autour de 12+ ou 13+). Ceci avait déjà été observé pour d'autres protéines [2]. Cette observation est qualitativement en accord avec les acidités de l'ammonium (NH_4) et de la triéthylamine protonée (HNEt_3^+). L'ammonium ($\text{pK}_a = 9,25$) étant plus acide que la triéthylamine ($\text{pK}_a = 11,01$), les observations expérimentales peuvent s'expliquer par une protonation plus importante dans le cas de l'ammonium.

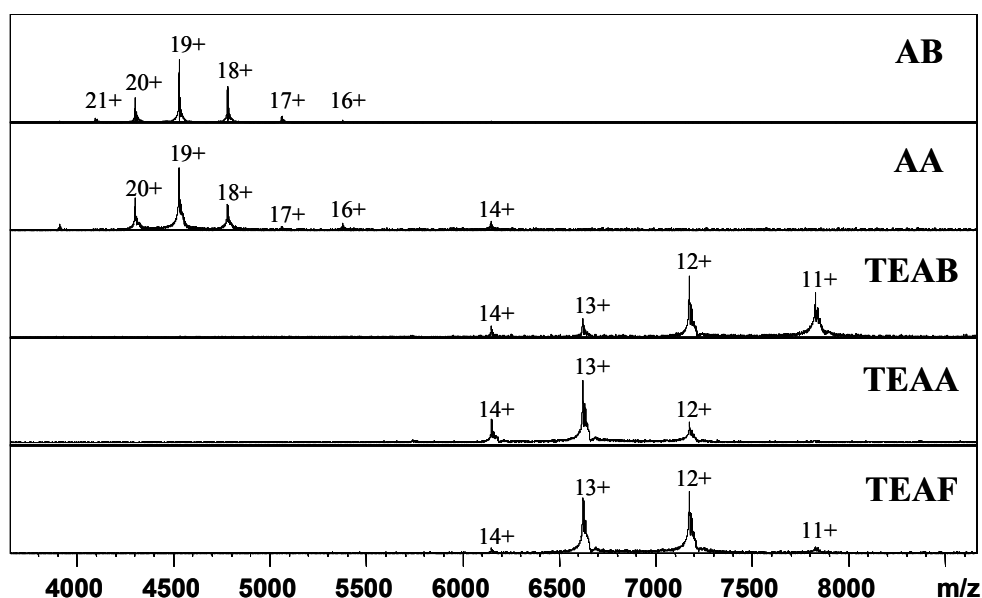


Fig. II.B.1 : 4 μM CK dans différents tampons : 10 mM AB (pH 8), 10 mM AA (pH 8), 10 mM TEAB (pH 8), 10 mM TEAA (pH 8) et 10 mM TEAF (pH 8).

Quelque soit le tampon, seule la forme dimérique de la protéine a été observée, ce qui indique que l'interaction entre les deux monomères reste intacte lors du processus d'électrospray dans tous les tampons. De plus, les distributions des états de charge dans tous les tampons sont restreintes, ce qui est caractéristique de protéines en conditions non dénaturantes. Dans le cas de la CK, la stabilité du complexe protéique ne dépend donc pas de la nature du tampon.

Il a été montré récemment que l'état de charge moyen des protéines multichargées et leur surface sont corrélés de façon linéaire avec une pente qui dépend du tampon [3], sur la base de l'attribution d'une structure sphérique aux protéines en phase gazeuse. Les résultats obtenus avec la CK dans les différents tampons sont en accord avec une structure globulaire de la protéine en phase gazeuse suivant ce modèle. Sur la base du nombre de charges portées, il n'y aurait pas de modification structurale importante entre les différents tampons.

II.B.2.2. Interaction CK-ADP ou CK-ATP dans différents tampons

Les interactions non covalentes entre la CK et ses ligands ADP et ATP ont été étudiées dans les différents tampons par un dosage de la protéine avec les ligands.

Lors du dosage de la CK avec l'ADP et l'ATP, les états de charge dans les différents tampons ne varient pas suite à la fixation d'un ligand, ce qui indique que la structure du complexe n'est pas trop différente de la structure de la protéine non complexée. Par contre, l'interaction varie de façon importante en fonction du tampon (Fig. II.B.2).

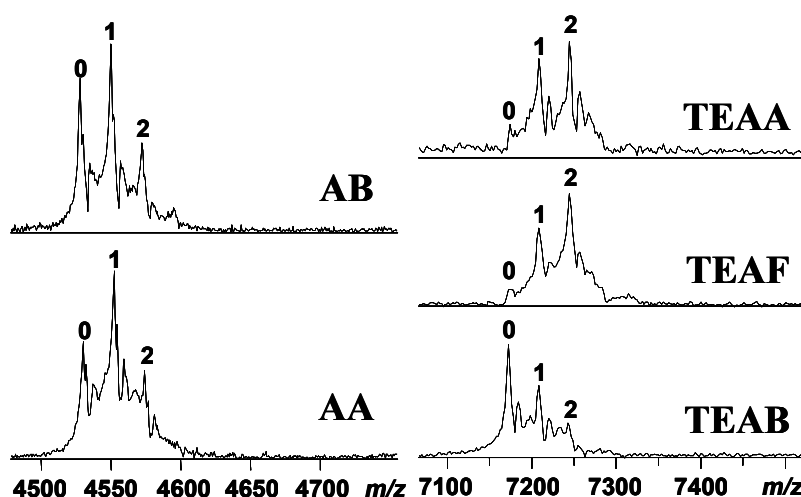


Fig. II.B.2 : 4 μ M CK + 16 μ M ADP dans différents tampons. A gauche : AB et AA, état de charge 19+. A droite : TEAA, TEAF et TEAB, état de charge 12+. Le nombre de ligands fixés par la protéine est indiqué au-dessus des pics.

Sur la base de l'observation de ces spectres, l'interaction semble du même ordre de grandeur dans les tampons AB et AA. La comparaison avec les résultats obtenus dans les tampons TEAA et TEAF suggère que dans les tampons comportant de la triéthylamine, l'interaction devient plus forte : à un rapport protéine:ligand 1:4, il n'y a presque plus de protéine non complexée, les pics correspondant à la protéine en complexe avec l'ADP sont les plus abondants dans le spectre. De plus, les interactions non spécifiques semblent plus faibles dans TEAA et TEAF que dans AA ou AB, ce qui est indiqué par la quasi absence de pics correspondant à la protéine avec plus de deux ligands fixés. Toutefois, le TEAB fait figure d'exception : à la différence de ce qui est observé dans les deux autres tampons comportant de la triéthylamine, l'interaction dans le TEAB semble moins forte que dans les tampons AA et AB (Fig. II.B.2).

Cette observation qualitative indique également que l'interaction entre CK et ADP observée en phase gazeuse ne peut pas être liée complètement à une interaction entre les charges positives portées par la protéine en phase gazeuse et des ligands chargés négativement, car certaines des formes de la protéine moins chargées fixent mieux l'ADP que des formes plus chargées.

L'analyse quantitative des interactions confirme ces observations qualitatives (Tab. II.B.1). La première constante de dissociation dans AB et AA est de l'ordre de 11-13 μM . Dans TEAA et TEAF, la première constante de dissociation est de l'ordre de 2-3 μM , reflétant la meilleure fixation du ligand dans ce tampon. En revanche, les constantes obtenues pour le tampon TEAB reflètent l'interaction plus faible entre la CK et l'ADP dans ce tampon : la première constante de dissociation est de l'ordre de 25 μM , indiquant une interaction plus faible. Toutefois, la deuxième constante est de l'ordre de 37 μM , et donc du même ordre de grandeur que dans AB et AA.

Tab. II.B.1 : Constantes de dissociation pour l'interaction CK-ADP en fonction du tampon.

tampon	$K_{1,ADP} / \mu\text{M}$	$K_{2,ADP} / \mu\text{M}$
AB	11,2	44,8
AA	12,9	51,6
TEAA	2,7	15,6
TEAF	1,8	13,2
TEAB	24,7	37,4

L'interaction entre la CK et l'ATP change aussi en fonction du tampon (Fig. II.B.3). Il faut noter que la qualité des données expérimentales a été inférieure à celle du système CK/ADP, probablement à cause de la meilleure fixation de l'ADP à la protéine qui permet d'obtenir des signaux intenses avec des concentrations moins importantes de ligand. On retrouve de nombreux points communs avec le comportement de l'interaction entre la CK et l'ADP en fonction du tampon. Comme pour l'ADP, l'interaction entre la protéine et son ligand augmente fortement dans le TEAA et le TEAF. L'interaction dans le TEAB semble un peu plus faible que dans tous les autres tampons. Dans le TEAA et le TEAF, on observe au maximum deux ligands fixés à la protéine à une concentration totale en ligand de 16 μM .

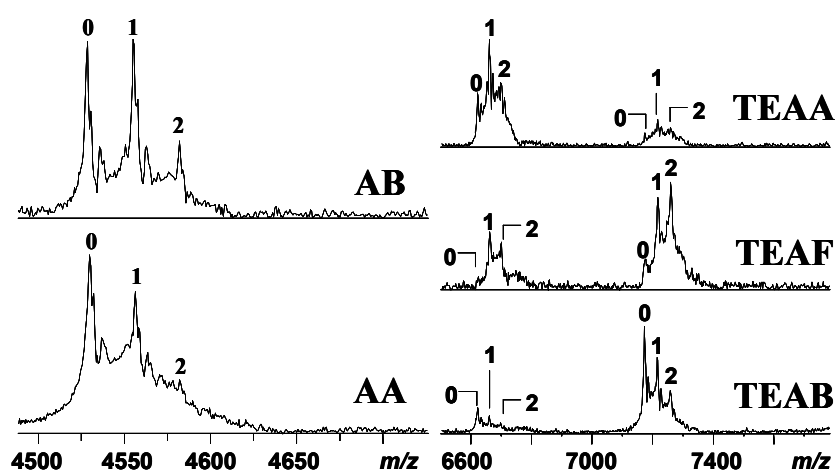


Fig. II.B.3 : 4 μM CK + 16 μM ATP dans différents tampons. A gauche : AB et AA, état de charge 19+. A droite : TEAB, TEAA et TEAF, état de charge 12+ (à droite)/13+ (à gauche). Le nombre de ligands fixés sur la protéine est indiqué au-dessus des pics.

Les résultats quantitatifs (Tab. II.B.2) confirment en partie les résultats qualitatifs : dans AB, AA, TEAA et TEAF, l'interaction entre la CK et l'ATP est moins forte que l'interaction entre la CK et l'ADP. En revanche, dans le TEAB, l'interaction entre la CK et l'ATP est plus forte que l'interaction entre la CK et l'ADP, contrairement à ce que le spectre laissait supposer.

Tab. II.B.2 : Constantes de dissociation pour l'interaction CK-ATP en fonction du tampon.

tampon	$K_{1,ATP} / \mu\text{M}$	$K_{2,ATP} / \mu\text{M}$
AB	26	103
AA	26	40
TEAA	6	21
TEAF	3	13
TEAB	16	29

Bien que la faible qualité des données obtenues avec l'ATP ne permette pas d'obtenir des informations détaillées sur ce système protéine-ligand, il est évident que la nature du tampon joue un rôle très important dans le cas de ces interactions non covalentes.

Quelles sont les explications possibles pour ce comportement ? L'anion seul n'a visiblement pas d'influence majeure sur l'interaction puisque dans AA et AB, les constantes sont du même ordre de grandeur, bien que les anions soient différents. Pour le TEAA et le TEAF, on observe le même phénomène.

Différentes hypothèses pourraient expliquer le résultat obtenu. Une première possibilité serait que la structure de la protéine, notamment du site actif, change légèrement en fonction du tampon (le dimère de la protéine étant la seule forme observée dans le spectre, des changements de grande ampleur de la structure sont peu probables). Il est possible que des changements mineurs de la structure ne déstabilisent pas les interactions entre les deux sous-unités, mais conduisent à des changements au niveau des sites d'interaction des ligands. La seule mesure de la masse moléculaire ne permet pas de répondre à cette question.

Une seconde hypothèse envisageable serait la précipitation soit de la protéine soit du ligand dans certains tampons, ce qui modifierait les concentrations de ces espèces moléculaires en solution. Les concentrations des différentes espèces moléculaires ont été vérifiées par spectrophotométrie, par l'absorption à 280 nm pour la CK et à 259 nm pour l'ADP et l'ATP. Ces mesures ont invalidé cette hypothèse.

Une troisième hypothèse serait la différence de qualité de l'effet de tampon pour ces différents sels à pH 8. Les courbes de dosage théoriques montrent des différences entre les différents tampons : les solutions de AB, de AA et de TEAB sont des tampons efficaces à pH 8, tandis que le TEAA et le TEAF n'ont pas un effet tampon important dans cette zone de pH (Fig. II.B.4). Ces dernières solutions seraient donc plus sensibles à toutes les perturbations de pH lors de l'expérience.

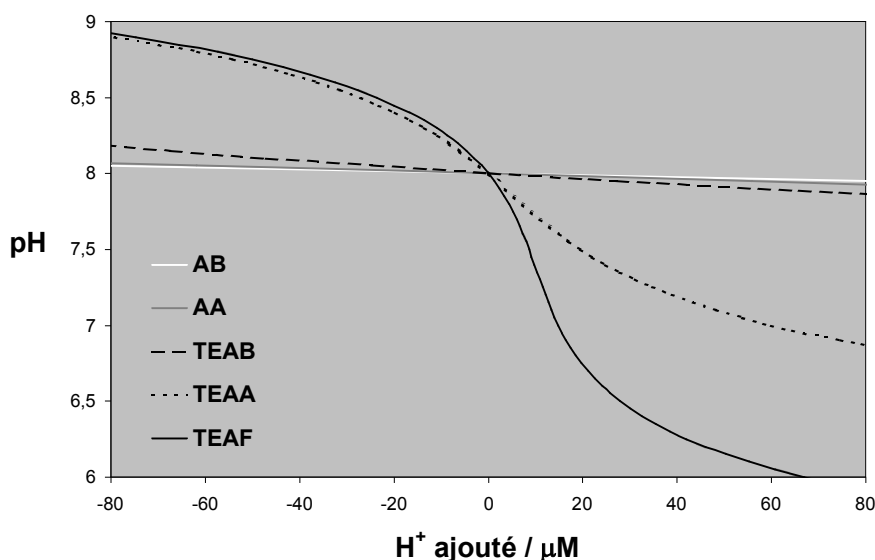


Fig. II.B.4 : Courbes de dosage théoriques pour les différentes solutions utilisées.

D'après la littérature l'ADP ainsi que l'ATP se fixent mieux à la CK lorsque le pH est réduit de pH 8,0 à pH 6,6 [4]. Pour les systèmes CK-ADP et CK-ATP, la diminution de pH se traduirait ainsi par une augmentation de l'interaction et des constantes de dissociation plus faibles. De ce fait, au vu des résultats précédents, on peut s'interroger sur la possibilité d'une perturbation du pH de la solution pendant l'expérience. Il a été vérifié que l'ajout des ligands à la solution de protéine ne perturbe pas nettement le pH : avant l'electrospray, la solution est donc bien à pH 8. Une variation du pH ne peut avoir lieu que dans la seringue, la ligne de transfert, ou pendant le processus d'electrospray. Soulignons ici quelques particularités de l'electrospray [5] : lors d'une nébulisation en mode positif, les gouttelettes dirigées vers le spectromètre de masse comportent un excédent en cations, tandis que la solution restant dans le capillaire est enrichie en anions. Pour rétablir l'électroneutralité dans le capillaire, il est nécessaire que des cations soient formés. Plusieurs réactions électrochimiques pourraient conduire à ceci, par exemple l'oxydation du métal du capillaire suivant l'équation générale $M \rightarrow M^{n+} + n e^-$, mais aussi d'autres réactions électrochimiques impliquant des espèces en solution. Une des réactions possibles conduit à la formation de H^+ : $H_2O \rightarrow 2 H^+ + \frac{1}{2} O_2 + 2 e^-$. Il est difficile de savoir quels types de réactions s'effectuent réellement dans le capillaire, mais il serait envisageable que des H^+ soient formés, ce qui conduirait à une acidification de la solution. Sur la base du courant mesuré pendant les expériences (10^{-7} A), on peut estimer une concentration maximale supplémentaire de 40 μM de H^+ . Ce surplus de protons, qui ne sont pas présents initialement dans la solution pourrait conduire à des pH finaux différents

dans les différents tampons : pour AA, AB et TEAB, le pH resterait autour de pH 8, tandis que dans le cas de TEAA (pH_{final} 7,2) et TEAF (pH_{final} 6,3), on aurait une diminution significative du pH (Fig. ??). Cette explication est également en accord avec l'ordre relatif des constantes obtenues en TEAA et TEAF. Dans le TEAF, les constantes de dissociation sont inférieures à celles obtenues dans le TEAA pour l'ADP et l'ATP, ce qui pourrait s'expliquer par la diminution plus importante du pH dans le TEAF.

Les évolutions observées pour les constantes de dissociations CK-ADP et CK-ATP dans les tampons TEAA et TEAF par rapport aux tampons AA et AB pourraient donc peut-être s'expliquer par un changement de pH, ces tampons ayant un faible effet tampon à pH 8. Par contre il est d'ores et déjà évident que cette explication ne suffirait pas dans le cas du TEAB, dont l'effet tampon est comparable, à pH 8, à celui de AA et AB alors que l'interaction entre la CK et ses ligands ADP et ATP semble avoir un comportement très différent.

II.B.3. Interaction CK-ligand à différents pH

Pour vérifier si l'hypothèse d'une modification du pH dans le processus electrospray pouvait expliquer une partie de nos résultats, les constantes de dissociation ont été mesurées en faisant varier le pH. L'objectif était principalement de vérifier si les effets du pH déjà décrits dans la littérature [4] pouvaient également être observés par spectrométrie de masse. Dans ce but, des solutions de 10 mM AB ont été préparées à des pH légèrement acide (pH 6,6), neutre (pH 7,2) et légèrement basique (pH 8,0). Le bicarbonate d'ammonium a été choisi comme tampon car ce sel tamponne relativement bien dans ces régions de pH : un ajout de $40 \mu\text{M H}^+$ (maximum estimé pour l'electrospray) aurait un effet ne dépassant pas une variation de 0,05 unités de pH.

Quelque soit le pH du tampon utilisé, seul le dimère a été observé, sans influence notable sur la distribution des états de charge. Ceci indique que la stabilité du dimère n'est pas affectée par le pH dans cette gamme. Une diminution du pH en dessous de 6,6 conduit à la dissociation du dimère de la CK, comme observé par Liang et al. [6].

Lorsqu'on observe l'évolution de l'interaction (Fig. II.B.5) de l'ADP ou de l'ATP avec la CK en fonction du pH, on observe une meilleure fixation des ligands lorsque le pH diminue, comme attendu.

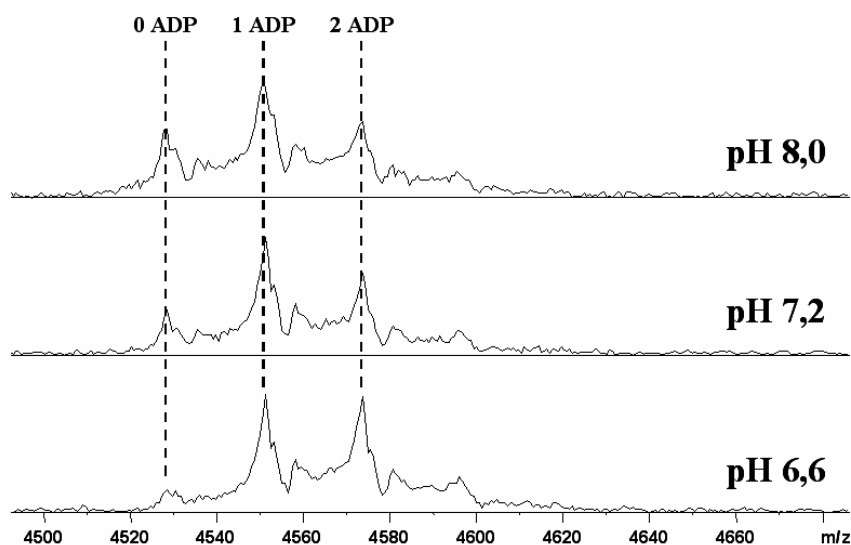


Fig. II.B.5 : 4 μM CK + 24 μM ADP, dans AB, état de charge 19+, à différents pH.

Les observations expérimentales indiquent donc qu'en diminuant le pH, l'ADP comme l'ATP se fixent mieux à la protéine ce qui correspond à ce qui avait été décrit initialement par Watts [4]. Ces données étayent ainsi l'hypothèse selon laquelle la variation des constantes en fonction du tampon pourrait être liée au pH de la solution dans le capillaire d'électrospray, qui est plus susceptible de diminuer dans les solutions moins bien tamponnées à pH 8 (TEAA et TEAF).

Ce résultat qualitatif se confirme en calculant les constantes de dissociation pour des solutions à différents pH (Tab. II.B.3). Du fait de la présence d'interactions non spécifiques, les constantes ont été calculées suivant le modèle de déconvolution présenté précédemment.

Tab. II.B.3 : Constantes de dissociation pour les interactions CK-ADP et CK-ATP en fonction du pH.

pH	$K_{1,ADP} / \mu\text{M}$	$K_{2,ADP} / \mu\text{M}$
8,0	11,2	44,8
7,2	7,3	29,2
6,6	3,6	14,4

pH	$K_{1,ATP} / \mu\text{M}$	$K_{2,ATP} / \mu\text{M}$
8,0	25,7	102,7
7,2	12,3	49,2
6,6	7,3	29,3

La contribution des interactions non spécifiques varie de façon intéressante en fonction du pH (Tab. II.B.4). Pour l'ADP, la part des interactions non spécifiques diminue en baissant le pH, allant de 0,26 à pH 8,0 jusqu'à presque zéro (0,01) à pH 6,6. En revanche, pour l'ATP, on constate le contraire : la contribution des interactions non spécifiques augmente en baissant le pH, de 0,16 (pH 8,0) jusqu'à 0,28 (pH 6,6). L'origine précise de ce comportement n'est pas plus connue que l'origine des interactions non spécifiques elles-mêmes.

Tab. II.B.4 : Part des interactions non spécifiques pour 4 μM CK + 16 μM ligand (ADP, ATP) à des pH différents.

pH	$\bar{n}_{\text{ADP}}$	$\bar{n}_{\text{ATP}}$
8,0	0,26	0,16
7,2	0,14	0,21
6,6	0,01	0,28

Finalement, on peut s'interroger sur la nature des interactions qui conduisent à cette variation des constantes d'interaction en fonction du pH. Les équilibres CK/ADP et CK/ATP ne dépendent pas directement du pH (puisque la concentration de H^+ ne figure pas explicitement dans l'expression de la constante d'équilibre), mais il faut noter que la forme majoritaire d'ADP et d'ATP en solution dépend du pH. La diminution du pH favorise la formation des formes protonées des ligands (HADP^{2-} et HATP^{3-}) aux dépens des formes non protonées (ADP^{3-} et ATP^{4-}). Le pH peut ainsi modifier la forme du substrat présente dans la solution, or, il a été observé que HATP^{3-} a un effet inhibiteur sur la réaction catalysée par la CK [4]. Ceci pourrait s'expliquer par une interaction plus forte entre la protéine et son ligand protoné par rapport au ligand dans sa forme non protonée. De même, il a été montré que ATP^{4-} interagit moins bien avec la CK [4], ce qui est cohérent avec nos résultats. On ne trouve pas de résultats équivalents dans la littérature pour l'ADP hormis le fait que l'interaction entre la CK et l'ADP augmente en diminuant le pH [4], mais on peut supposer qu'il en est probablement de même pour l'ADP que pour l'ATP.

Une seconde explication de nos observations pourrait être une modification du site actif lors de la diminution de pH. En particulier, des résidus basiques situés au niveau du site actif et impliqués dans l'interaction avec les ligands [7] pourraient être protonés lorsque le pH diminue. Parmi ceux-ci, les histidines His191 et His296 sont les plus susceptibles de changer d'état de protonation dans la gamme de pH 6,6-8,0. Une protonation des résidus His191 et His296 conduirait à une interaction électrostatique plus forte avec les groupes phosphates des ligands, augmentant ainsi l'interaction.

Si tous les résultats présentés sur les évolutions du pH sont cohérents entre la littérature et nos propres travaux, il faut noter cependant qu'une modification de pH ne suffit pas à expliquer la variation des constantes de dissociation en changeant de tampon : dans la section précédente, on a estimé le pH final dans les solutions au minimum à 7,2 pour le TEAA et 6,3 pour le TEAF. Les constantes de dissociation pour le premier ligand sont, dans ces deux tampons, $K_{1,TEAA} = 2,7 \mu\text{M}$ et $K_{1,TEAF} = 1,8 \mu\text{M}$. Ces constantes évoluent en fonction du pH de la façon suivante : $K_{1,pH\ 7,2} = 7,3 \mu\text{M}$ et $K_{1,pH\ 6,6} = 3,6 \mu\text{M}$. Les résultats sont similaires pour la seconde constante d'interaction et pour l'ATP. L'hypothèse d'un effet du pH dans les tampons était donc séduisante, mais elle ne peut, à elle seule, rendre compte complètement des observations expérimentales, même pour les tampons ayant un effet de pH faible à pH 8.

En conclusion, on peut noter que la détermination des constantes de dissociation pour les systèmes CK-ADP et CK-ATP a pu être menée par spectrométrie de masse dans le bicarbonate d'ammonium et que ces constantes suivent la même évolution en fonction du pH que ce qui était décrit dans la littérature [4]. Ces résultats vont dans le bon sens pour interpréter les différences observées entre les différents tampons, mais ils ne suffisent pas à rendre compte complètement des phénomènes pouvant avoir lieu dans différents tampons. Il resterait à vérifier sur d'autres systèmes dépendant différemment du pH si de telles évolutions sont également observées en faisant varier les tampons. On peut toutefois déjà retenir de ce travail l'importance du choix du tampon pour étudier des interactions entre des protéines et des ligands. D'après la littérature, les constantes d'interaction mesurées par spectrométrie de masse sont globalement en accord avec celles obtenues par d'autres techniques, mais les expériences correspondantes ont majoritairement été réalisées en acétate ou bicarbonate d'ammonium, tampons qui ont un effet tampon fort dans la gamme de pH habituellement utilisée pour ce type d'étude.

II.B.4. Comparaison des constantes obtenues par différentes techniques

Pour valider notre approche de détermination des constantes des équilibres CK/ADP et CK/ATP par spectrométrie de masse, il aurait été souhaitable de pouvoir comparer les valeurs mesurées avec des valeurs de référence. Or, ces valeurs de référence n'existent pas, puisque les interactions en question n'ont jamais été étudiées dans exactement les mêmes conditions que celles que nous avons utilisées. Une étude approfondie de la littérature révèle que ces constantes ont, en fait, été mesurées par un grand nombre de techniques différentes dans de

nombreuses conditions différentes (Tab. II.B.5). La comparaison avec nos résultats est rendue d'autant plus difficile que de grands écarts existent entre les valeurs de la littérature : de 0,5 μM à 500 μM pour l'équilibre CK-ADP et de 3,9 μM à 3260 μM pour l'équilibre CK-ATP.

Tab. II.B.5 : Constantes de dissociation pour les équilibres CK-ADP (en haut) et CK-ATP (en bas). Dans les cas où K_1 et K_2 ont été déterminés indépendamment, K_2 est la valeur donnée entre parenthèses. + indique la présence, - l'absence de l'ion correspondant.

$K_d / \mu\text{M}$	Mg^{2+}	Cr/PCr	NO_3^-	<u>ADP</u> pH	Méthode
135	+	-	-	8,3	fluorimétrie[8]
132	+	+	-	8,3	fluorimétrie[8]
143	+	-	+	8,3	fluorimétrie[8]
13,7	+	+	+	8,3	fluorimétrie[8]
170	+	+	-	8,0	photométrie[4]
110	-	-	-	9,0	fluorimétrie[9]
90	+	-	-	9,0	fluorimétrie[9]
84	+	+	-	9,0	fluorimétrie[9]
4 (22)	-	+	+	9,0	fluorimétrie[9]
0,5 (11)	+	+	+	9,0	fluorimétrie[9]
40	+	+	-	8,0	quench à l'équilibre [10]
40	+	+	-	6,8	quench à l'équilibre [10]
80	+	+	-	7,2	quench à l'équilibre [10]
20	+	+	-	7,6	quench à l'équilibre [10]
10	+	+	-	8,4	quench à l'équilibre [10]
500	+	+	-	7,0	photométrie[11]
70	+	-	-	7,9	dialyse à l'équilibre / gradient de sédimentation[12]
100	-	-	-	7,9	dialyse à l'équilibre / gradient de sédimentation [12]

$K_d / \mu\text{M}$	Mg^{2+}	Cr/PCr	NO_3^-	<u>ATP</u> pH	Méthode
1200	+	+	-	8,0	photométrie[4]
710	+	+	-	9,0	pH-stat[13]
700	+	+	-	8,0	pH-stat[14]
12,5	-	-	-	7,0	analyse calorimétrique par dosage isotherme (ITC) [15]
3,9	+	-	-	7,0	analyse calorimétrique par dosage isotherme (ITC) [15]
600	+	+	-	8,0	quench à l'équilibre [10]
300	+	+	-	6,8	quench à l'équilibre [10]
700	+	+	-	7,2	quench à l'équilibre [10]
200	+	+	-	7,6	quench à l'équilibre [10]
300	+	+	-	8,4	quench à l'équilibre [10]
3260	+	+	-	8,0	photométrie[11]
300	+	-	-	7,9	dialyse à l'équilibre / gradient de sédimentation [12]
500	-	-	-	7,9	dialyse à l'équilibre / gradient de sédimentation [12]

Plusieurs raisons peuvent expliquer cette diversité de résultats : tout d'abord tous les résultats n'ont pas été réalisés au même pH, ce qui a une influence importante, comme nous l'avons vu dans la partie précédente. Ensuite, la nature du tampon n'étant pas toujours la même, ceci peut modifier l'interaction. L'étude des interactions non covalentes par spectrométrie de masse impose la contrainte d'utiliser un tampon volatil, ce qui n'est pas le cas pour les techniques « classiques » utilisées pour étudier ces interactions. Il n'est par conséquent pas simple de se replacer dans des conditions déjà étudiées.

Il y a un élément de variabilité important à souligner : toutes ces études pourraient ne pas avoir été menées avec la même structure de l'enzyme. Il existe d'après la littérature différents états d'oxydation dans la CK obtenue commercialement [16]. Le résidu Cys282, localisé dans le site actif de la protéine [7], est probablement le résidu oxydé [17, 18]. Ces différents états d'oxydation présentent une activité enzymatique différente [17-20]. Si l'on suppose que ces différentes formes peuvent présenter également une interaction différente avec les ligands ADP et ATP, en accord avec les observations de Trundle et Cunningham [19], il serait possible d'expliquer au moins en partie les différences entre les valeurs de la littérature par la nature de la protéine utilisée. Une autre modification envisageable serait la nucléotidylation du site actif de l'enzyme [21]. A l'heure actuelle, ces hypothèses structurales sont en cours d'investigation par spectrométrie de masse.

Finalement, une dernière hypothèse quant à l'origine de la diversité des constantes d'interaction est la variété des techniques utilisées. Parmi toutes ces techniques, la fluorimétrie est la technique utilisée par le plus grand nombre d'équipes de recherche et pourrait donc être considérée comme la méthode de référence pour la CK. Les constantes de dissociation obtenues par fluorimétrie sont plus élevées que celles obtenues par spectrométrie de masse (Tab. II.B.5). Pour l'ADP, la plupart des constantes ont des valeurs entre 40 et 140 μM , ce qui correspond à une valeur de K_1 de l'ordre de 20-70 μM en absence de coopérativité entre les deux sites. Ces valeurs semblent être indépendantes des conditions expérimentales, et, pour la plupart, ne changent pas considérablement en présence d'autres ions, bien qu'il existe des variations notables (Tab. II.B.5). La constante obtenue par spectrométrie de masse ($K_{1,\text{ADP}} = 12,5 \mu\text{M}$) correspond à une interaction nettement plus forte que ce qui est rapporté dans la littérature, soit 70 μM si l'on considère l'ensemble des conditions, ou 50 μM si l'on ne considère que les résultats obtenus en l'absence d'inhibiteurs ou d'activateurs connus de la réaction enzymatique.

Pour l'ATP, la plupart des valeurs rapportées dans la littérature sont de l'ordre de 200-700 μM , ce qui correspond à une valeur de $K_{1,\text{ATP}}$ de l'ordre de 100-350 μM en absence de

coopérativité entre les deux sites. Les constantes obtenues par spectrométrie de masse sont de l'ordre de 17-30 μM (à pH 7,9), ce qui est largement hors de la gamme de la majorité des constantes rapportées dans la littérature. Les résultats obtenus par notre approche sont par contre assez proches des constantes obtenues par ITC [15], qui ont été obtenues dans des conditions particulièrement proches ($K_{\text{I,ATP}} = 6,3 \mu\text{M}$ à pH 7,0 à comparer à $K_{\text{I,ATP}} = 12,3 \mu\text{M}$ à pH 7,2 par spectrométrie de masse). On peut toutefois noter que même si les valeurs absolues diffèrent assez fortement dans la littérature, l'ordre relatif entre l'ADP et l'ATP est toujours conservé et est en accord avec nos résultats.

A l'exposé de ces résultats on peut se poser la question de savoir si les expériences faites avec des techniques différentes ont réellement mesuré le même état du complexe protéine-ligand. Nos résultats par spectrométrie de masse semblent en bon accord avec l'ITC mais diffèrent notablement de ceux obtenus par la technique la plus couramment utilisée, la fluorimétrie. Les techniques de fluorimétrie [22] et d'ITC [23, 24] ont été présentées en détail dans l'introduction (voir I.3.2.1.1 et I.3.2.2.1). Il faut simplement rappeler ici que la fluorimétrie mesure en fait simplement l'interaction entre un ligand et un ou plusieurs tryptophanes dont la fluorescence est modifiée par l'ajout de ce ligand, tandis que l'ITC est la mesure directe du bilan thermochimique d'une addition de ligand. Par conséquent ces deux techniques ne mesurent pas la même observable, et dans les deux cas, c'est sous réserve d'hypothèses implicites que ces observables peuvent être reliées à une constante d'équilibre.

Ainsi, dans le cas de la fluorimétrie, seule l'interaction entre le ligand et le fluorophore du tryptophane conduit à un signal. Des ligands qui seraient fixés à la protéine mais qui ne seraient pas à proximité du fluorophore ne conduiraient pas à un changement de fluorescence et ne seraient pas mesurés. On pourrait imaginer des ligands approchant du site actif de la protéine, mais sans être en interaction avec les fluorophores (Fig. II.B.6, A), la protéine pouvant déjà interagir avec le ligand sans que les fluorophores soient impliqués. A cette interaction pourrait déjà correspondre une part substantielle de l'énergie d'interaction entre les partenaires, ce qui se traduirait par une variation d'enthalpie et de la masse, détectables par ITC et par spectrométrie de masse. L'entrée en contact du ligand avec le ou les fluorophores pourrait ne correspondre qu'à un second équilibre, ayant sa propre énergie d'interaction, ce qu'on peut schématiser par deux équilibres successifs :



La fluorimétrie ne mesurerait ainsi que le bilan global $K_{\text{diss,fluo}} = 1/(K_A K_B)$, la spectrométrie de masse le rapport $([P,L]_A + [P,L]_B)/P$ et l'ITC un bilan énergétique global pour ces deux réactions.

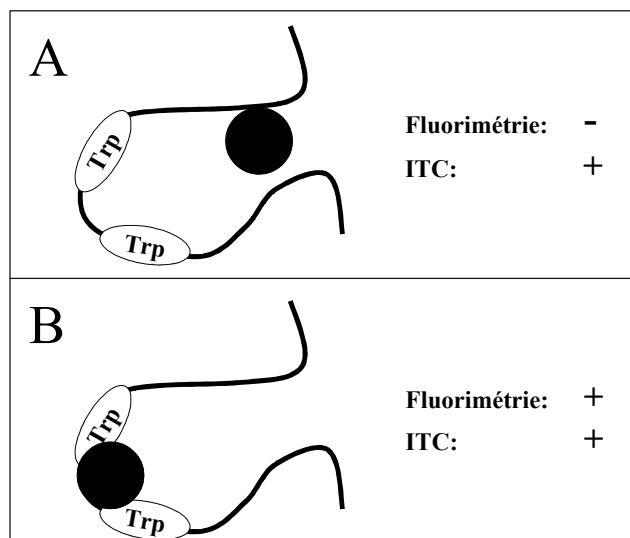


Fig. II.B.6 : Deux situations possibles de fixation du ligand dans le site actif d'une protéine. (A) Le ligand (en noir) est déjà en contact avec la protéine, mais n'est pas suffisamment proche des tryptophanes pour induire une variation de la fluorescence. (B) Le ligand est en contact avec les tryptophanes présents dans le site actif.

Il faut noter que les tryptophanes ont été symbolisés comme étant au cœur du site d'interaction, mais cela n'a pas besoin d'être le cas : il suffirait qu'existe la possibilité pour le ligand de se fixer dans deux positions, l'une à proximité des tryptophanes et l'autre plus loin pour que ce raisonnement soit possible.

On peut noter que le cas de la CK avec l'ADP et l'ATP ne serait pas un exemple isolé : un effet similaire est observé pour le système cyclophiline 18 humaine (hCyp18) et cyclosporine A (CsA). Les constantes de dissociation obtenues par ITC sont de l'ordre de 1,6-17 nM (suivant le pH) [25], tandis qu'avec la fluorimétrie, la constante de dissociation mesurée est de l'ordre de 30-205 nM [26]. Il n'y a toutefois pas de raisons pour que ce phénomène soit général : la fluorimétrie et l'ITC peuvent être en accord, comme par exemple dans le cas de la glutathione et de la glutathione S-transférase, où les constantes de dissociation obtenues sont de 175 μ M (fluorimétrie) et de 189 μ M (ITC) [27]. Dans le cadre de notre hypothèse, ceci traduirait simplement l'inexistence d'un second site de fixation éloigné des fluorophores.

Cette hypothèse peut permettre de rationaliser les résultats obtenus : en présence de deux sites d'interaction d'énergies comparables, la répartition entre les deux sites peut ne pas favoriser celui qui conduit à un quench de la fluorescence. Les expériences d'ITC et de spectrométrie de masse mesureraient de façon similaire un bilan, homologue à une constante d'interaction globale, ce qui est observé quand on compare nos résultats de spectrométrie de masse à ceux d'ITC. Par contre, la fluorimétrie pourrait mesurer une constante d'association plus faible, simplement parce que tous les ligands fixés n'interagissent pas nécessairement avec les tryptophanes. A l'appui de cette hypothèse, on peut noter que nous n'avons pas trouvé de cas dans la littérature où une constante d'association mesurée par ITC ou spectrométrie de masse soit inférieure à celle mesurée par fluorimétrie.

II.B.5. Conclusions

Les constantes d'équilibre pour les systèmes CK-ADP et CK-ATP obtenues par spectrométrie de masse dépendent du tampon utilisé. Le cas du TEAB est un peu particulier puisque le comportement de notre système n'y suit pas les mêmes règles que dans les autres tampons. Nous n'avons pas d'explication à cela, mais il faut noter que l'intensité des spectres obtenus en TEAB était beaucoup plus faible que dans les autres tampons. De ce fait, l'erreur sur les constantes calculées est importante et ces résultats sont moins fiables que ceux obtenus dans les autres tampons. Si l'on excepte le cas du bicarbonate de triéthylammonium, il semble qu'on puisse séparer les tampons utilisés en deux catégories, d'un côté les tampons ammoniums (AA et AB), de l'autre deux tampons triéthylammoniums (TEAA et TEAF). Pour ces quatre tampons, les valeurs de constantes présentent une corrélation avec l'effet tampon des différents sels à pH 8 : le TEAA et le TEAF stabilisent très peu le pH dans ces conditions tandis que les deux autres sont de meilleurs tampons. Ceci conduit à l'hypothèse d'un effet électrochimique se produisant dans l'électrospray qui perturberait le pH. Une étude de l'effet du pH sur les constantes de dissociation a permis de montrer que l'évolution des constantes allait dans le bon sens pour expliquer l'effet des différents sels, mais cette explication n'est pas suffisante : en évaluant la quantité maximale de protons produits par les phénomènes électrochimiques, on ne peut arriver à expliquer complètement la variation des constantes de dissociations observées dans les deux tampons TEAA et TEAF. Ces résultats montrent en tout cas que le choix d'un tampon qui soit volatil mais qui présente également de bonnes capacités en terme de tampon dans la gamme de pH à laquelle on souhaite travailler est essentiel.

La comparaison des constantes de la littérature entre elles et avec nos résultats expérimentaux est difficile, car les conditions expérimentales relevées dans la littérature sont extrêmement variées. Plusieurs hypothèses sont possibles pour expliquer cette variété, entre autres choses une variabilité au niveau de la protéine elle-même. Toutefois ces variabilités ne permettent pas de rendre compte des écarts observés entre la fluorimétrie et deux techniques, l'ITC et la spectrométrie de masse. Une hypothèse séduisante, mais qui reste à démontrer sur plus que deux exemples pourrait être que ces techniques ne présentent en fait pas la même observable : en particulier, s'il existe plusieurs états différents d'attachement d'un ligand à une protéine, des techniques comme la fluorimétrie les distingue tandis que l'ITC ou la spectrométrie de masse en font une mesure globale.

II.B.6. Bibliographie

- [1] D. Lemaire, G. Marie, L. Serani, O. Laprévotte, Stabilization of gas-phase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer, *Anal. Chem.* **2001**, 73, 1699-1706.
- [2] F. Halgand, O. Laprévotte, Mean charge state and charge state distribution of proteins as structural probes. An electrospray ionisation mass spectrometry study of lysozyme and ribonuclease A, *Eur. J. Mass Spectrom.* **2001**, 7, 433-439.
- [3] M. Hauteux, N. Hue, A. Du Fou de Kerdaniel, A. Zahir, V. Malec, O. Laprévotte, Under non-denaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface, *Int. J. Mass Spectrom.* **2004**, 231, 131-137.
- [4] D. C. Watts, Creatine Kinase (Adenosine 5'-triphosphate-creatine phosphotransferase), dans: *The Enzymes*, P. D. Boyer (Ed.), Academic Press NY and London (1973), Vol. VIII, pp. 383-455.
- [5] G. J. van Berkel, The electrolytic nature of electrospray, dans: *Electrospray Ionization Mass Spectrometry*, R. B. Cole (Ed.), John Wiley & Sons, Inc., NY (1997), pp. 65-105.
- [6] Y. Liang, F. Du, S. Sanglier, B.-R. Zhou, Y. Xia, A. Van Dorselaer, C. Maechling, M.-C. Kilhoffer, J. Haiech, Unfolding of rabbit muscle creatine kinase induced by acid. A study using electrospray ionization mass spectrometry, isothermal titration calorimetry, and fluorescence spectroscopy, *J. Biol. Chem.* **2003**, 278, 30098-30105.

- [7] S. D. Lahiri, P.-F. Wang, P. C. Babbitt, M. J. McLeish, G. L. Kenyon, K. N. Allen, The 2.1 Å structure of Torpedo californica creatine kinase complexed with the ADP-Mg²⁺-NO₃⁻ - creatine transition-state analogue complex, *Biochemistry* **2002**, *41*, 13861-13867.
- [8] C. L. Borders, Jr., M. J. Snider, R. Wolfenden, P. L. Edminston, Determination of the affinity of each component of a composite quaternary transition-state analogue complex of creatine kinase, *Biochemistry* **2002**, *41*, 6995-7000.
- [9] A. C. McLaughlin, The interaction of 8-Anilino-1-naphtalenesulfonate with creatine kinase. Evidence for cooperativity of nucleotide binding, *J. Biol. Chem.* **1974**, *249*, 1445-1452.
- [10] J. J. Burbaum, J. R. Knowles, Internal thermodynamics of enzymes determined by equilibrium quench: values for K_{int} for enolase and creatine kinase, *Biochemistry* **1989**, *28*, 9306-9317.
- [11] T. Hornemann, D. Rutishauser, T. Wallimann, Why is creatine kinase a dimer? Evidence for cooperativity between the two subunits, *Biochim. Biophys. Acta* **2000**, *1480*, 365-373.
- [12] S. A. Kuby, T. A. Mahowald, E. A. Noltmann, Studies on adenosine triphosphate transphosphorylases. IV. Enzyme-substrate interactions, *Biochemistry* **1962**, *1*, 748-762.
- [13] E. T. Maggio, G. L. Kenyon, Properties of a CH₃-blocked creatine kinase with altered catalytic activity. Kinetic consequence of the presence of the blocking group, *J. Biol. Chem.* **1977**, *252*, 1202-1207.
- [14] R. Furter, E. M. Furter-Graves, T. Wallimann, Creatine kinase: the reactive cysteine residue is required for synergism but is nonessential for catalysis, *Biochemistry* **1993**, *32*, 7022-7029.
- [15] M. Forstner, C. Berger, T. Wallimann, Nucleotide binding to creatine kinase: an isothermal titration microcalorimetry study, *FEBS Letters* **1999**, *461*, 111-114.
- [16] D. Favre, B. Muellhaupt, Reactivation of creatine kinase by dithiothreitol prior to use in an in vitro translation extract, *ALTEX* **2005**, *22(4)*, 259-264.
- [17] J.-R. Kim, H. W. Yoon, K.-S. Kwon, S.-R. Lee, S. G. Lee, Identification of proteins containing cysteine residues that are sensitive to oxidation by hydrogen peroxide at neutral pH, *Anal. Biochem.* **2000**, *283*, 214-221.
- [18] A. M. Hurne, C. L. L. Chai, P. Waring, Inactivation of rabbit muscle creatine kinase by reversible formation of an internal disulfide bond induced by the fungal toxin gliotoxin, *J. Biol. Chem.* **2000**, *275*, 25202-25206.
- [19] D. Trundle, L. W. Cunningham, Iodine oxidation of the sulfhydryl groups of creatine kinase, *Biochemistry* **1969**, *8*, 1919-1925.

- [20] Y. J. Suzuki, J. D. Edmondson, G. D. Ford, Inactivation of rabbit muscle creatine kinase by hydrogen peroxide, *Free Radic. Res. Commun.* **1992**, *16*(2), 131-136.
- [21] S. S. M. David, B. E. Haley, ATP nucleotidylation of creatine kinase, *Biochemistry* **1999**, *38*, 8492-8500.
- [22] M. R. Eftink, Fluorescence methods for studying equilibrium macromolecule-ligand interactions, dans: *Meth. Enzymol.*, L. Brand et M. L. Johnson (Eds.), Academic Press NY (1997), Vol. 273, pp. 221-257.
- [23] H. F. Fisher, N. Singh, Calorimetric methods for interpreting protein-ligand interactions, dans: *Meth. Enzymol.*, M. L. Johnson et G. K. Ackers (Eds.), Academic Press NY (1995), Vol. 259, pp. 194-221.
- [24] M. M. Pierce, C. S. Raman, B. T. Nall, Isothermal titration calorimetry of protein-protein interactions, *Methods* **1999**, *19*, 213-221.
- [25] J. Fanghänel, G. Fischer, Thermodynamic characterization of the interaction of human cyclophilin 18 with cyclosporin A, *Biophys. Chem.* **2003**, *100*, 351-366.
- [26] P. Gallo, F. Rossi, M. Saviano, C. Pedone, G. Colonna, R. Ragone, Specific interaction between bovine cyclophilin A and synthetic analogues of cyclolinopeptide A, *J. Biochem.* **1998**, *124*, 880-885.
- [27] E. Ortiz-Salmerón, Z. Yassin, M. J. Clemente-Jimenez, F. J. Las Heras-Vazquez, F. Rodriguez-Vico, C. Barón, L. Garcia-Fuentes, Thermodynamic analysis of the binding of glutathione to glutathione S-transferase over a range of temperatures, *Eur. J. Biochem.* **2001**, *268*, 4307-4314.

Partie III

Développement de méthodes pour l'étude de la structure des protéines

Dans cette seconde partie, nous présenterons la mise au point de méthodes destinées à l'étude de structures de protéine ou de complexes protéiques par spectrométrie de masse FT-ICR. A la différence de la partie précédente, il ne s'agit plus ici d'étudier les complexes entiers en phase gazeuse, mais d'utiliser la spectrométrie de masse comme outil d'analyse de modifications en solution du complexe. Ces méthodes dérivent de méthodes déjà existantes, mais ont dû être adaptées à nos conditions expérimentales.

Comme souligné dans l'introduction, la spectrométrie de masse a pris une place importante parmi les techniques d'étude de structure des protéines. Dans la première partie de cette thèse, les complexes protéiques ont été étudiés par une approche où ils sont engendrés intacts en phase gazeuse après désolvatation et ionisation dans la source de l'instrument. Si ces approches permettent d'avoir des informations liées à la stœchiométrie des assemblages en solution, elles n'apportent pas un grand nombre d'informations sur leur structure en solution. La deuxième partie de cette thèse porte donc sur ce second aspect. Comme indiqué dans l'introduction, la spectrométrie de masse ne permet qu'une mesure de masse en phase gazeuse. On peut donc soit tenter d'analyser des éléments structuraux mais sous forme ionisée et en phase gazeuse, soit convertir des informations structurales en solution sous forme de variations de masse observables via le spectromètre de masse.

Dans ce cadre, la spectrométrie de masse FT-ICR peut apporter deux avantages par rapport aux autres techniques de spectrométrie de masse : la précision en masse et la résolution très élevée, et des méthodes dissociatives spécifiques telles que la dissociation par capture d'électrons. Jusqu'ici la plupart des études structurales sur des complexes protéiques ont été menées en utilisant une approche dite « bottom-up », où, après modification de la masse, la protéine ou le complexe protéique est dissocié et soumis à une étape de protéolyse avant analyse par spectrométrie de masse. Or la précision sur la mesure de masse des instruments FT-ICR permet d'envisager une autre approche, dite « top-down », où l'on peut se passer de l'étape de protéolyse : les complexes protéiques sont dissociés et les protéines entières sont introduites dans le spectromètre de masse. La fragmentation se fait directement par activation en phase gazeuse. L'avantage d'une telle approche est multiple, en particulier lorsque la digestion peut ne pas conduire à une bonne couverture de la protéine. En particulier, dans le cas de modifications chimiques qui perturbent l'action de la protéine (cas de modifications de lysines par le DEPC par exemple), on n'est plus tributaire du choix de l'enzyme. Le premier chapitre détaille la mise au point d'une méthode ECD/Top-Down par FT-ICR. Dans le chapitre suivant, une méthode de modification covalente utilisant le diéthylpyrocarbonate (DEPC) est présentée : il s'agissait à la fois de vérifier la spécificité de la méthode quant aux acides aminés marqués et de vérifier sa capacité à apporter des informations structurales. Les deux approches, « bottom-up » et « top-down » sont comparées dans le cas de modifications de protéines par le DEPC.

Finalement, un dernier chapitre présente un travail de mise au point des échanges H/D étudiés par FT-ICR. Il s'agit ici, non plus d'introduire une modification de masse sous forme

d'une modification covalente, mais sous forme d'un remplacement des hydrogènes d'amides accessibles au solvant par des deutériums. Bien que cette technique soit relativement classique et bien décrite dans la littérature, il a fallu vérifier que les protocoles étaient adaptés aux conditions disponibles au laboratoire, et en particulier regarder si la précision apportée par un FT-ICR permettait de se passer d'une étape de séparation chromatographique pour l'analyse des peptides. L'application de cette dernière méthode aux oligomères de la protéine du prion ovien fait l'objet de la dernière partie de cette thèse.

Chapitre III.A

Fragmentation des protéines entières
dans la cellule ICR par dissociation par
capture d'électrons

III.A.1. Introduction

L'approche classique pour l'analyse structurale de protéines par spectrométrie de masse est une approche partant d'un découpage de la protéine d'intérêt en peptides par l'action d'une protéase. L'analyse des masses des peptides par spectrométrie de masse permet de remonter à la caractérisation de la protéine d'où sa dénomination d'approche « bottom-up ». Néanmoins, cette approche présente quelques inconvénients intrinsèques lorsqu'on ne cherche pas uniquement à identifier la protéine par une recherche en base de donnée mais également à la caractériser complètement. Tout d'abord, le processus de digestion, les enzymes utilisées et le problème de purification et d'introduction dans le spectromètre de masse font qu'on ne retrouve en général pas assez de peptides pour donner une couverture de l'ensemble de la séquence de la protéine. De plus, pour identifier le résidu portant une modification post-traductionnelle, un séquençage par MS/MS du peptide portant la modification est nécessaire et n'est possible que si l'intensité du pic correspondant est suffisante. Finalement, les protéines présentent souvent une hétérogénéité lorsque de multiples modifications sont possibles et l'approche « bottom-up » ne permet d'en rendre compte que partiellement.

Une technique alternative pour la caractérisation de protéines est la fragmentation par une approche dite « top-down », où l'on aborde le problème de la caractérisation de la protéine par la fragmentation dans le spectromètre de masse de la protéine entière. Les premiers exemples d'utilisation de cette technique remontent à Loo et al [1] et sont arrivés peu après le développement des sources electrospray et MALDI [1]. Dans ce travail précurseur, la fragmentation de la ribonucléase A bovine (13 kDa) sous l'effet de collisions (CID) a été observée dans un spectromètre de masse de type triple-quadrupôle. En 1994, le groupe de McLafferty introduit une approche similaire sur des spectromètres de masse FT-ICR. L'activation des ions nécessaire à la fragmentation est apportée soit par une excitation radiofréquence (rf) résonante de l'ion parent suivie de collisions dissociatives (CID) [2], soit par excitation *rf* non résonante « sustained off-resonance irradiation CID » (SORI-CID) [2] ou encore par absorption multiple de photons infrarouge produits par un laser à CO₂ (Infrared Multiphoton Dissociation, IRMPD) [3]. Ces techniques de collision ont été mises à profit pour fragmenter aussi bien des protéines telles que l'ubiquitine (8 kDa) ou l'anhydrase carbonique (29 kDa), que des oligonucléotides. La mise en évidence et l'exploitation du phénomène de dissociation de peptides suite à la capture d'un électron (Electron Capture Dissociation, ECD) [4-6] ont apporté un nouvel outil pour la dissociation de biomolécules dans les spectromètres de masse FT-ICR. Une des principales difficultés de l'approche « top-down » est que de très nombreux fragments sont engendrés par la fragmentation d'une

protéine entière en comparaison avec ce qui est formé pour des peptides : il faut tenir compte à la fois du nombre plus important de sites de coupure potentiels, mais également de la multiplication des états de charge des fragments. De ce fait, l'analyse d'un spectre « top-down » nécessite de disposer d'une excellente sensibilité, d'une excellente résolution et, pour l'attribution correcte des fragments, d'une excellente précision sur la mesure de masse.

La spectrométrie de masse FT-ICR, et en particulier l'accroissement de sensibilité apporté par l'optimisation du transfert des ions dans la cellule, apporte ces éléments indispensables. La combinaison de l'ECD avec un FT-ICR par une approche « top-down » a été mise à profit aussi bien pour identifier de nombreux types de modifications post-traductionnelles [7] (un exemple récent étant les histones humaines [8-11]) que pour obtenir des informations sur les structures secondaires et tertiaires de protéines en phase gazeuse [12]. De plus, la technique peut également être mise à profit pour révéler des informations sur la structure des protéines en solution. Le groupe de Young et al. a présenté des études sur des protéines préalablement liées chimiquement par des réactifs bifonctionnels (« Crosslinker ») en solution et ensuite fragmentées en phase gazeuse pour en tirer des informations sur leurs structures [13, 14].

Dissociation des peptides et protéines par ECD

Le processus de dissociation par capture d'électron a lieu en soumettant des ions portant plusieurs charges positives à une irradiation par un faisceau d'électrons thermiques (< 1 eV d'énergie cinétique). Les ions multichargés $[M + nH]^{n+}$, produits par une source d'électrospray, peuvent se recombiner avec un électron pour conduire, au moins de façon intermédiaire, à une espèce réduite $[M + nH]^{(n-1)+}$. L'énergie déposée dans le système par la capture de l'électron permet à l'intermédiaire d'évoluer, soit directement par une perte de $H^{\bullet}$, soit, via des transferts d'hydrogènes, vers une fragmentation, ou encore de se stabiliser sous sa forme réduite. Cette dernière voie est particulièrement observée dans le cas de systèmes de taille importante : ceci s'interprète comme une redistribution de l'énergie interne sur les nombreux modes de vibration internes du système, rendant le temps de vie de l'intermédiaire supérieur au temps d'observation, voire même au temps de désexcitation radiative.

Expérimentalement, les fragments observés sont essentiellement des ions de formes c et z par rupture de la liaison N- C_{α} (Fig. III.A.1). Ceci montre une différence majeure de ce mode de dissociation par rapport aux méthodes où l'apport d'énergie se fait par excitation vibrationnelle (CID ou IRMPD) d'une espèce (multi-)protonée. Le mécanisme précis de cette fragmentation atypique reste encore un point de débat [5], néanmoins on peut souligner

l'importance qu'il y a à considérer que la fragmentation a lieu dans une espèce radicalaire, excédentaire en H.

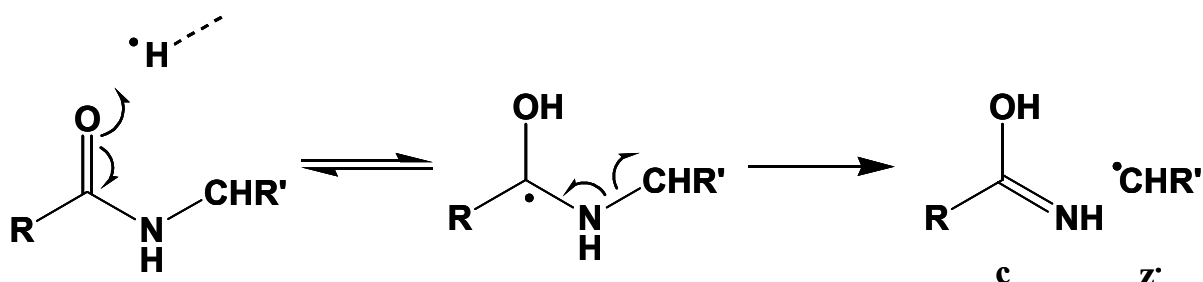


Fig. III.A.1 : Version simplifiée de l'un des mécanismes proposés pour l'ECD conduisant à la formation des ions *c* et *z'*.

Dans le cadre de cette thèse, l'approche « Top-Down » a été utilisée en particulier pour les travaux portant sur les modifications sélectives d'acides aminés par le DEPC. Il s'avère qu'une approche de type « bottom-up » ne permettait pas d'identifier et de localiser précisément les sites de modification (voir chapitre III.B). Novak et al. avaient mis en évidence l'intérêt d'une approche « Top-Down » sur l'ubiquitine modifiée par l'acétate de N-hydroxysuccinimidyle (NHS) sur un spectromètre de masse hybride Q-FTICR par fragmentation par collision [15]. A notre connaissance, la fragmentation par ECD n'a jamais été utilisée auparavant pour identifier des modifications artificielles des acides aminés afin d'obtenir des informations sur leur réactivité ainsi que sur la structure des protéines. Toutefois, avant d'entreprendre l'étude des sites modifiés par le DEPC par une approche ECD – Top-down, nous avons souhaité vérifier si l'approche top-down était applicable à deux protéines : l'ubiquitine et la myoglobine.

III.A.2. Matériel et méthodes

L'ubiquitine et la myoglobine (Sigma) ont été préparées à des concentrations de 10-20 μM dans une solution $\text{H}_2\text{O}/\text{MeOH}/\text{acide formique}$ 49,9 : 49,9 : 0,2 (v/v/v) et analysées par injection directe par nano-ESI-FTMS. Pour l'ubiquitine, l'état de charge 10+ a été sélectionné (16+ pour la myoglobine). L'ECD a été effectué par irradiation des ions pendant 50 ms avec une énergie des électrons de l'ordre de 1-2 eV. 200 scans ont été accumulés.

III.A.3. Fragmentation de l'ubiquitine

Le spectre ECD de l'ubiquitine (Fig. III.A.2) a été réalisé sur le spectromètre de masse FT-ICR, et on constate qu'un grand nombre de fragments sont ainsi obtenus. De fait, l'analyse de ce spectre en utilisant les outils des logiciels Xmass et BioTools (Bruker Daltonics, Brême) n'a pas conduit à un résultat interprétable, très probablement à cause des nombreux fragments présents à des états de charge différents. Le spectre a donc été analysé à la main en y recherchant directement des massifs correspondant aux fragments c et z. Les fragments ont été identifiés par leur masse moléculaire et par la correspondance des massifs isotopiques expérimentaux et théoriques. Pour y arriver, la haute résolution de la spectrométrie de masse FT-ICR était primordiale, surtout dans les cas où des massifs isotopiques se chevauchaient (Fig. III.A.2, encart).

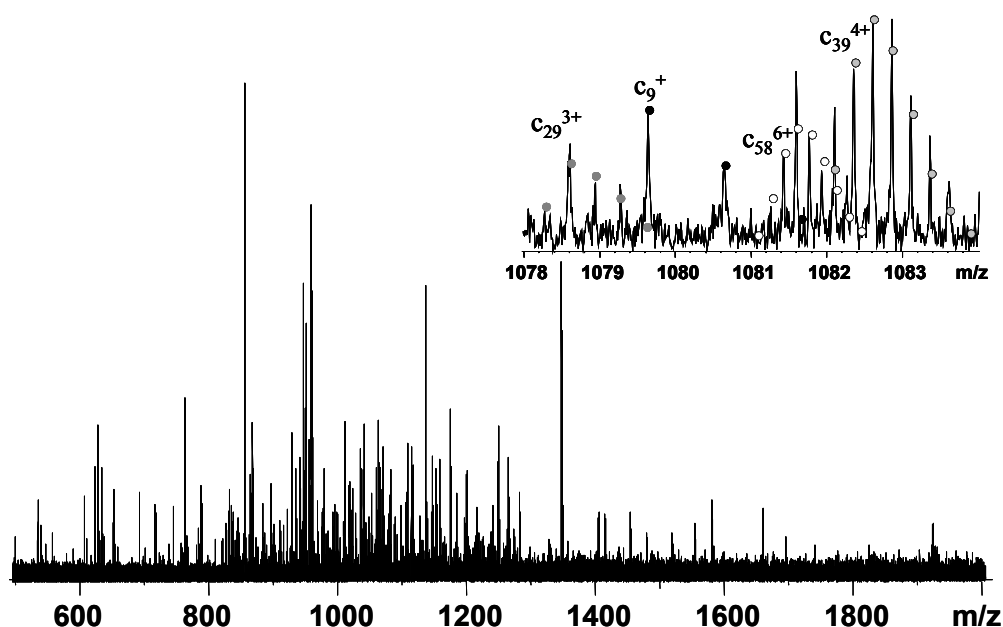


Fig. III.A.2 : Spectre ECD de l'ubiquitine (état de charge 10+, 10 μ M, 50 ms temps d'irradiation à environ 1-2 eV). L'encart détaille une petite région du spectre (m/z 1078 – 1084) où de nombreux chevauchements sont apparents. Les intensités relatives théoriques sont indiquées par des points.

Dans le cas de l'ubiquitine, 76 fragments de la protéine (40 ions c et 36 ions z) ont été identifiés par cette approche (Tab. III.A.1, Fig. III.A.3). Les masses inférieures à m/z 500 n'ont pas été détectées, car la technique de piégeage dynamique utilisée pour améliorer le rapport S/B conduit à un bruit de fond chimique trop important dans cette région du spectre.

De ce fait, nous n'avons pas pu obtenir d'informations sur cette partie du spectre, ce qui explique l'absence des ions c_1 à c_3 et z_1 à z_4 . Les plus petits ions identifiés sont donc c_4 et z_5 . Pour la série des ions c , on trouve presque tous les fragments possibles du côté N-terminal jusqu'à c_{44} . L'absence des ions c_{18} , c_{36} et c_{37} s'explique par le fait que la rupture de la liaison N- C_α au niveau d'une proline ne conduit pas à une fragmentation de la chaîne polypeptidique (puisque la proline est cyclique). Au delà, on ne trouve plus que quelques fragments c : c_{58} , c_{59} , c_{62} , c_{65} et c_{75} . Les intensités relatives des pics des ions correspondants dans le spectre sont faibles. Il apparaît donc une forme de discrimination en faveur des fragments conduisant à des ions de basses masses moléculaires. La série z présente un comportement similaire : presque tous les fragments de z_5 à z_{45} sont présents. Au-delà, seuls quelques fragments sont observés à une faible intensité relative : z_{48} , z_{52} , z_{69} . Plusieurs éléments peuvent expliquer ces faibles efficacités de détection de fragments de grande taille : une dilution sur un plus grand nombre de massifs isotopiques ou encore des fragmentations multiples. En effet, ces fragments de grande taille sont plus chargés que les plus petits. Par conséquent la recombinaison avec des électrons est plus probable. Or si un fragment capture un électron, il se refragmente à son tour, conduisant à un fragment interne et un ion terminal de plus petite taille.

Tab. III.A.1 : Fragments de l'ubiquitine identifiés après dissociation par capture d'électrons.

fragment	$M_{\text{exp}} / \text{Da}$	δ / ppm	fragment	$M_{\text{exp}} / \text{Da}$	δ / ppm	fragment	$M_{\text{exp}} / \text{Da}$	δ / ppm
c4	536,3	-7	c32	3588,0	6	z18	2080,2	12
c5	635,4	8	c33	3716,0	-1	z19	2195,2	16
c6	763,5	19	c35	3902,2	14	z20	2282,3	11
c7	864,5	17	c38	4209,3	10	z21	2395,3	-4
c8	977,6	18	c39	4324,4	10	z22	2496,4	4
c9	1078,6	17	c40	4452,4	10	z23	2652,5	18
c10	1135,7	11	c41	4580,4	-3	z24	2709,5	14
c11	1263,7	1	c43	4849,7	7	z25	2824,5	12
c13	1477,9	0	c44	4962,8	15	z26	2953,6	13
c14	1578,9	13	c58	6481,5	5	z29	3322,8	3
c15	1692,0	-16	c59	6643,6	6	z30	3379,9	16
c16	1821,1	16	c62	6998,8	17	z31	3450,9	10
c17	1920,2	17	c65	7345,0	13	z32	3597,9	1
c19	2146,2	17	c75	8502,7	10	z33	3711,0	9
c20	2233,3	6	z5	541,3	-6	z35	3980,2	13
c21	2348,3	10	z6	654,4	11	z36	4108,3	16
c22	2449,3	-3	z8	866,6	14	z37	4236,4	17
c23	2562,4	6	z9	1003,6	-5	z38	4351,4	12
c24	2691,4	-4	z10	1116,7	-4	z43	4844,7	13
c25	2805,5	-4	z11	1217,8	7	z44	4972,8	15
c26	2904,6	4	z12	1304,8	11	z45	5216,8	13
c27	3032,7	7	z13	1433,8	10	z48	5458,0	10
c28	3103,7	13	z14	1561,9	16	z52	5869,3	18
c29	3231,8	14	z15	1690,0	23	z69	7697,3	16
c30	3344,9	9	z16	1803,1	19			
c31	3473,0	10	z17	1917,1	10			



Fig. III.A.3 : Représentation des fragments c et z observés sur la séquence primaire de l'ubiquitine.

La complémentarité entre les séries c et z permet toutefois une très bonne couverture de la protéine : 61 ruptures de liaisons sur 75 possibles sont observées dans au moins une des deux séries. Sur les 14 restantes, 3 correspondent à des liaisons du côté N-terminal de prolines pour lesquelles on n'attend pas de fragmentations. Cette excellente couverture de séquence apportée par cette approche ECD/Top-Down serait tout à fait utile pour la recherche de modifications de certains résidus : la probabilité de pouvoir identifier un site de modification précis est assez élevée, ce qui valide le choix de cette approche.

III.A.4. Fragmentation de la myoglobine

Malgré l'apparent succès obtenu pour l'ubiquitine, il faut tempérer ce résultat en notant que l'ubiquitine n'est qu'une protéine de 8 kDa, ce qui en fait un système très simple pour une telle approche. Comme indiqué dans l'introduction, en l'absence d'un laser infrarouge, cette approche reste toutefois limitée à des systèmes dans lesquels il est possible d'apporter suffisamment d'énergie interne pour observer des dissociations. Pour tenter de situer cette limite dans nos conditions instrumentales, la même étude sur une protéine de taille plus importante, la myoglobine (17 kDa) a été effectuée.

De façon générale, les résultats obtenus pour la myoglobine sont très similaires à ceux obtenus pour l'ubiquitine : le spectre obtenu présente un degré de complexité similaire (Fig. III.A.4), avec également des superpositions de massifs isotopiques, et les fragments de masses inférieures à m/z 500 n'ont pas été détectés. De la même façon, des séries c (de c_5 à c_{46}) et z (de z_7 à z_{34}) presque complètes pour les extrémités sont observées, alors que les fragments de plus grandes tailles sont moins fréquemment présents (c_{51} , c_{52} , c_{56} , c_{59} , c_{62} , c_{67} , z_{44} , z_{45} , z_{48} , z_{56} , z_{57} , z_{75}).

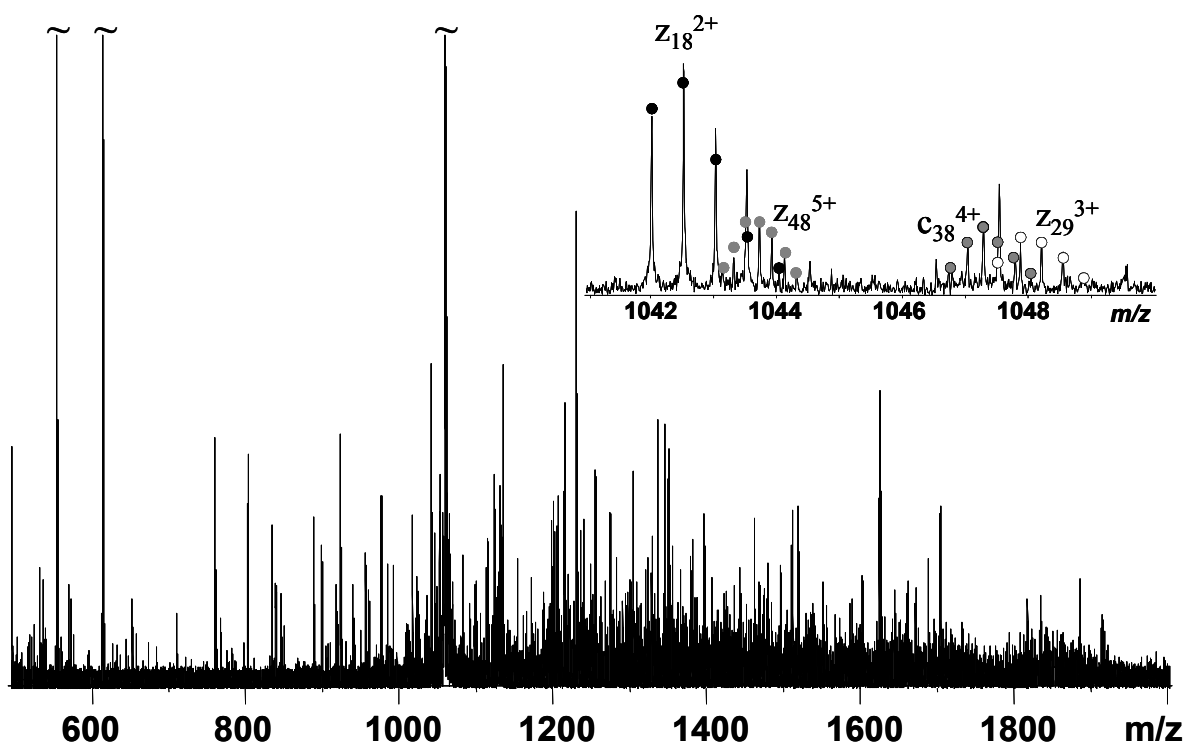


Fig. III.A.4 : Spectre ECD de la myoglobine (état de charge 16^+ , $10 \mu\text{M}$, 50 ms temps d'irradiation d'environ $1\text{-}2 \text{ eV}$). L'encart détaille la région m/z 1041-1050 avec les différents ions qui y sont identifiés. Les intensités relatives théoriques sont indiquées avec des points.



Fig. III.A.5 : Représentation des fragments c et z sur la séquence primaire de la myoglobine.

A la différence de l'ubiquitine, on observe cette fois-ci (Fig. III.A.5) que les séries c et z ne se chevauchent plus. Il y a même une partie centrale de la protéine (acides aminés 67-96) qui n'est couverte que par une seule fragmentation. Ceci pourrait s'expliquer par l'hypothèse faite plus haut suivant laquelle l'absence de fragments de taille importante proviendrait d'une capture multiple d'électrons sur des espèces de taille importante fortement chargées. De plus, il avait été observé qu'en absence d'activation vibrationnelle [16], les zones les plus

structurées en phase gazeuse d'une protéine se fragmentent difficilement par ECD. Par conséquent, si la localisation précise d'une modification sur les régions les plus externes de la séquence primaire de la protéine semble aisée, une telle localisation sur la partie centrale sera plus aléatoire, dépendant de l'existence de fragments apportant l'information utile dans cette région.

III.A.5. Conclusions

La comparaison des résultats pour l'ubiquitine et la myoglobine situe bien les limitations d'une approche « top-down » par ECD. Pour les deux protéines, des séries quasiment complètes des fragments correspondant aux 40 à 50 résidus extrémaux sont observées. Au-delà de 50 résidus, la présence de fragments dans le spectre diminue fortement, à la fois en nombre mais également en intensité pour ceux qui sont observés. L'origine de cette décroissance n'est pas bien établie : il pourrait s'agir d'un effet de multifragmentation, de limitation de la détection, mais également d'effets de structuration conduisant à la subsistance de liaisons non covalentes en phase gazeuse insuffisamment activées par la capture d'électron. On s'attend à ce que ce dernier effet devienne de plus en plus important avec l'augmentation de la taille de la protéine : la taille conduisant en effet d'une part à une plus grande structuration de l'assemblage en phase gazeuse par des liaisons non covalentes, mais également à une augmentation des degrés de liberté du système sur lesquels peut se diluer l'énergie apportée par la capture de l'électron. Une solution à ce dernier problème serait d'irradier les analytes en phase gazeuse avec un laser infrarouge conduisant à une augmentation de l'énergie interne des protéines par absorption de photons. Les liaisons non covalentes sont ainsi pré- ou post-dissociées et l'ECD conduit simplement à la rupture des liaisons covalentes du squelette peptidique.

Finalement, ce qu'il faut retenir de ce travail est essentiellement que l'approche « top-down » avec ECD sera très utile pour séquencer des extrémités C et N-terminales de protéines. Dans les parties plus centrales d'une protéine, la localisation de modifications sera plus aléatoire, dépendant de la présence de fragmentations dans les régions permettant une localisation des modifications recherchées. Les difficultés de retraitement des spectres et la nécessité d'optimiser les conditions de fragmentations ont été également mises en évidence dans ce travail. Le retraitement des spectres ECD est long, car il a été fait à la main. Il conviendra de rechercher des outils informatiques adaptés, basés par exemple sur l'algorithme THRASH ou ses variantes [17]. Les conditions de fragmentation seront à optimiser, mais pour

cela il sera certainement indispensable de disposer d'un laser infrarouge installé sur le FT-ICR de façon à partir d'un système où les liaisons non covalentes sont dissociées par activation vibrationnelle.

III.A.6. Bibliographie

- [1] J. A. Loo, C. G. Edmonds, R. D. Smith, Primary Sequence Information from Intact Proteins by Electrospray Ionization Tandem Mass Spectrometry, *Science* **1990**, *248*, 201-204.
- [2] M. W. Senko, J. P. Speir, F. W. McLafferty, Collisional Activation of Large Multiply Charged Ions Using Fourier Transform Mass Spectrometry, *Anal. Chem.* **1994**, *66*, 2801-2808.
- [3] D. P. Little, J. P. Speir, M. W. Senko, P. B. O'Connor, F. W. McLafferty, Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing, *Anal. Chem.* **1994**, *66*, 2809-2815.
- [4] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.
- [5] R. A. Zubarev, K. F. Haselmann, B. Budnik, F. Kjeldsen, F. Jensen, Towards an understanding of the mechanism of electron-capture dissociation: a historical perspective and modern ideas, *Eur. J. Mass Spectrom.* **2002**, *8*, 337-349.
- [6] R. A. Zubarev, Electron-capture dissociation tandem mass spectrometry, *Curr. Opin. Biotech.* **2004**, *15*, 12-16.
- [7] Y. Ge, B. G. Lawhorn, M. ElNaggar, E. Strauss, J.-H. Park, T. P. Begley, F. W. McLafferty, Top Down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry, *J. Am. Chem. Soc.* **2002**, *124*, 672-678.
- [8] J. J. Pesavento, Y.-B. Kim, G. K. Taylor, N. L. Kelleher, Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by Top Down Mass Spectrometry, *J. Am. Chem. Soc.* **2004**, *126*, 3386-3387.
- [9] N. Siuti, M. J. Roth, C. A. Mizzen, N. L. Kelleher, J. J. Pesavento, Gene-specific characterization of human histone H2B by electron capture dissociation, *J. Prot. Res.* **2006**, *5*, 233-239.
- [10] C. E. Thomas, N. L. Kelleher, C. A. Mizzen, Mass spectrometric characterization of human histone H3: a bird's eye view, *J. Prot. Res.* **2006**, *5*, 240-247.

- [11] M. T. Boyne II, J. J. Pesavento, C. A. Mizzen, N. L. Kelleher, Precise characterization of human histones in the H2A gene family by Top Down Mass Spectrometry, *J. Prot. Res.* **2006**, 5, 248-253.
- [12] H. Oh, K. Breuker, S. K. Sze, Y. Ge, B. K. Carpenter, F. W. McLafferty, Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy, *Proc. Natl. Acad. Sci.* **2002**, 99, 15863-15868.
- [13] G. H. Kruppa, J. Schoeninger, M. M. Young, A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry, *Rapid Commun. Mass Spectrom.* **2003**, 17, 155-162.
- [14] P. Novak, M. M. Young, J. S. Schoeninger, G. H. Kruppa, A top-down approach to protein structure studies using chemical cross-linking and Fourier transform mass spectrometry, *Eur. J. Mass Spectrom.* **2003**, 9, 623-631.
- [15] P. Novak, G. H. Kruppa, M. M. Young, J. Schoeninger, A top-down method for the determination of residue-specific solvent accessibility in proteins, *J. Mass Spectrom.* **2004**, 39, 322-328.
- [16] D. M. Horn, Y. Ge, F. W. McLafferty, Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins, *Anal. Chem.* **2000**, 72, 4778-4784.
- [17] D. M. Horn, R. A. Zubarev, F. W. McLafferty, Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules, *J. Am. Soc. Mass Spectrom.* **2000**, 11, 320-332.

Chapitre III.B

Modification des protéines par le
diéthylpyrocarbonate (DEPC)

III.B.1. Introduction

L'interaction entre une protéine et ses ligands (autres protéines, petites molécules, ARN, etc.) peut avoir une influence sur la structure de cette protéine. De ce fait, la détermination de la constante d'interaction entre une protéine et son ligand est une étape importante pour la caractérisation du complexe ainsi formé, mais ne suffit pas à rendre compte de l'ensemble des phénomènes se produisant lors de l'interaction. Il sera donc tout aussi important d'obtenir des informations plus précises sur la structure de la protéine étudiée seule ou en complexe avec ses partenaires.

Comment la structure d'une protéine peut-elle être étudiée par spectrométrie de masse ?

Comme nous l'avons vu dans l'Introduction générale, une possibilité pour obtenir ces informations sera d'utiliser la réactivité des acides aminés accessibles en solution. Dans cette approche (Fig. III.B.1), la première étape consiste à marquer les résidus accessibles à la surface de la protéine par des réactifs spécifiques. Ensuite, la protéine sera digérée et les peptides résultants analysés par spectrométrie de masse. Les modifications des acides aminés sont détectées par la différence de masse moléculaire par rapport aux peptides non modifiés. Une approche différente consiste à faire de la MS/MS sur la protéine entière, approche appelée « top-down ». Avec les données obtenues par spectrométrie de masse, il est ainsi possible d'identifier les régions modifiées (donc accessibles) et non modifiées (donc protégées, soit à l'intérieur de la structure de la protéine soit par l'interaction avec un partenaire). Dans la figure III.B.1, la hauteur des barres de la carte d'accessibilité est proportionnelle au pourcentage de l'acide aminé correspondant ayant été modifié (plus haute est la barre, plus l'acide aminé est modifié). Si la couverture de séquence fournit suffisamment d'informations, il est finalement possible de reconstituer la topologie du complexe protéique.

Les modifications covalentes de la protéine peuvent être réversibles ou irréversibles. L'approche par modifications réversibles de référence est l'échange des hydrogènes d'amides par des deutériums [1]. Evidemment, ce type de modification est réversible puisque les deutériums peuvent être rétro-échangés par des hydrogènes en solution (voir chapitre III.C), ce qui impose des contraintes au niveau du temps d'analyse *post modification* : l'analyse doit se faire assez rapidement et à froid afin de diminuer le taux de rétro-échange. La modification irréversible de groupements chimiques des chaînes latérales des acides aminés par des réactifs spécifiques n'impose pas cette contrainte : le temps de traitement de l'échantillon après la modification pourra donc être plus long. Un inconvénient de cette approche pourrait être la

perturbation de la structure tridimensionnelle par la modification. En effet, les réactifs utilisés auront un encombrement stérique plus important qu'un hydrogène, et peuvent modifier la charge ou les fonctions chimiques des acides aminés marqués. De ce fait, l'interaction de cet acide aminé avec son environnement pourra être perturbée, ce qui peut affecter la structure de la protéine entière. Il sera donc souhaitable de limiter le nombre de modifications par protéine.

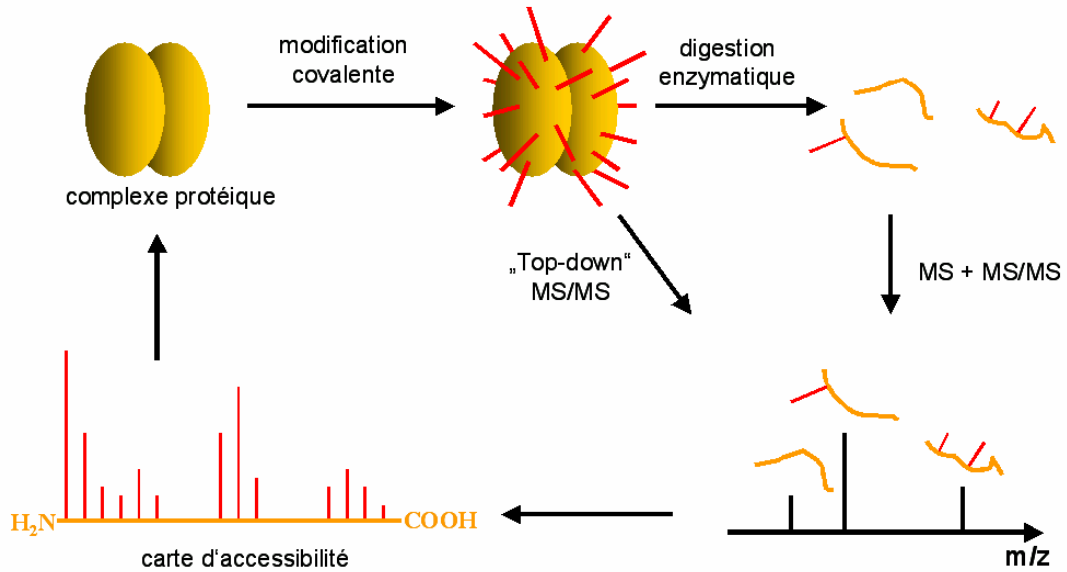


Fig. III.B.1 : Stratégie pour obtenir des informations sur la structure des protéines par modification chimique et analyse par spectrométrie de masse (explications voir texte).

Concernant les modifications irréversibles, deux approches sont développées actuellement : la modification par des réactifs monofonctionnels et la modification par des réactifs multifonctionnels. Cette dernière approche, dite « Crosslinking », ne sera pas traitée dans le cadre de cette thèse. Nous nous intéresserons dans ce chapitre aux modifications covalentes irréversibles des protéines par des réactifs monofonctionnels.

Il existe une grande variété de réactifs permettant de modifier différents acides aminés [2]. Les exemples de la littérature dans lesquels la modification chimique des protéines a été utilisée en combinaison avec la spectrométrie de masse couvrent une grande variété de questions, comme par exemple l'identification des résidus réactifs localisés dans le site actif de la protéine [3, 4]. Cependant, dans la plupart des cas, cette technique a été employée afin d'étudier l'accessibilité de la surface des protéines [5-12] et les variations structurales se produisant lors de la formation d'un complexe protéique [13-22]. Les modifications utilisées

peuvent être l'acétylation des lysines [16, 18, 21], la modification des arginines par des 1,2-diones [13, 18], l'iodation des tyrosines [20] et des modifications moins habituelles comme la glycinamidation des résidus acides [19] ou l'oxydation par des radicaux hydroxyles [15].

Pour nos études, nous avons choisi le diéthylpyrocarbonate (DEPC), réactif utilisé pour modifier spécifiquement les histidines à des valeurs de pH comprises entre 6 et 7 [2]. Ce réactif présente un intérêt particulier car un certain nombre de protéines contiennent des histidines dans leur site actif. De plus, cet acide aminé est assez fréquent pour obtenir des informations structurales, mais pas au point de compliquer excessivement l'analyse par une multitude de modifications.

La modification par le DEPC est couramment utilisée par les biochimistes pour inactiver des enzymes qui contiennent des histidines dans leur site actif. Dans ce cas, la spectrométrie de masse a été employée afin d'identifier les acides aminés modifiés, et déterminer ainsi le site actif de la protéine [23-25]. D'autres exemples présentent l'identification des sites d'interaction entre une protéine et un ligand. Parmi ces exemples, notons l'analyse de l'interaction entre la sélénoprotéine P et l'héparine [26] et de la complexation de la protéine prion par le Cu (II) [27].

Notre objectif a été tout d'abord de mettre au point la modification de protéines par le DEPC à pH 8, ce qui correspond par exemple au pH du bicarbonate d'ammonium utilisé lors des études d'interactions non covalentes par spectrométrie de masse. En effet, nous souhaitons pouvoir obtenir des informations structurales sur les protéines dans les mêmes conditions que celles employées lors de la détermination des constantes d'interaction par spectrométrie de masse.

Au-delà, notre objectif était d'évaluer le potentiel de la modification des protéines par le DEPC dans l'étude des structures de protéines.

Le mécanisme de la réaction est une substitution nucléophile qui conduit à une augmentation de la masse moléculaire des acides aminés modifiés de + 72.02 Da (Fig. III.B.2), correspondant à une carbéthoxylation. D'après ce mécanisme, la réactivité est susceptible de dépendre du pH utilisé lors de la réaction. Il est ainsi possible que l'état de protonation des résidus nucléophiles (par exemple Lys, Ser, Tyr), qui change en fonction du pH, ait un rôle important sur la réactivité vis-à-vis du DEPC.

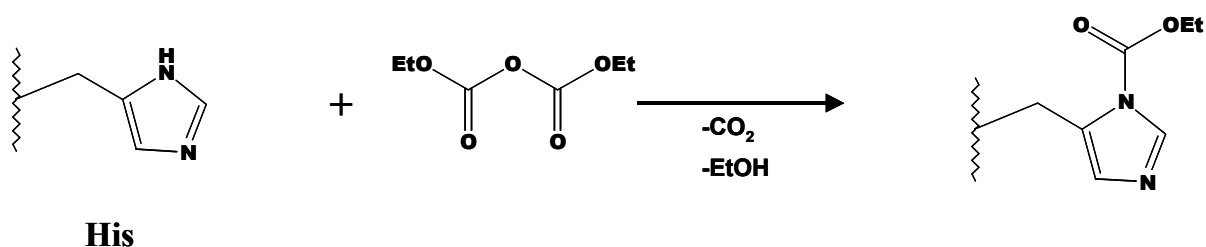


Fig. III.B.2 : Carbéthoxylation des histidines par le DEPC.

Tout d'abord, nous avons étudié la spécificité de la réaction avec le DEPC dans les conditions expérimentales dans lesquelles nous travaillons. Ensuite, nous avons développé deux approches différentes pour l'identification et la quantification des sites de modification : la première est basée sur la digestion enzymatique et l'analyse des peptides (approche « bottom-up »), la seconde sur la fragmentation de la protéine intacte en phase gazeuse (approche « top-down »). Dans la dernière partie, nous tenterons d'évaluer l'intérêt de la technique pour l'étude des complexes protéiques avec l'exemple de la créatine kinase, une protéine dimérique.

III.B.2. Réactivité du DEPC

D'après la littérature, le DEPC réagit principalement sur l'histidine, et est même devenu le réactif standard pour la modification de cet acide aminé [2]. Néanmoins, il a été observé que des lysines [28] ainsi que des sérines et tyrosines [29] sont aussi susceptibles d'être modifiées par le DEPC, très probablement à cause des groupements -NH_2 et -OH de leurs chaînes latérales qui pourraient également intervenir dans des substitutions nucléophiles. Vu le pH légèrement basique (pH 8) employé lors de nos expériences, la réactivité de certains acides aminés peut être modifiée par rapport à la plupart des résultats rapportés dans la littérature, obtenus à pH neutre ou légèrement acide (pH 6-7). Par exemple, les lysines pourraient être déprotonées, ce qui augmenterait leur réactivité. Il a donc tout d'abord été important d'examiner la spécificité de la réaction, un préalable pour la modification des protéines, qui constituent des systèmes complexes.

L'étude a été effectuée sur un mélange de peptides obtenu par digestion du cytochrome c. Cette approche présente plusieurs avantages. Tout d'abord, les peptides obtenus sont suffisamment petits pour que l'on puisse estimer que des effets de structure (autres que de la structure primaire) seront exclus. La réactivité avec le DEPC ne dépendra donc que de la

nature des acides aminés. Ensuite, un mélange suffisamment hétérogène (en termes de composition en acides aminés) permet d'obtenir le maximum d'informations possible sur la réactivité en une seule expérience. Finalement, l'identification des acides aminés modifiés par MS/MS sera plus facile pour des peptides, qui constituent des petits systèmes, que pour des protéines entières. Au cas où la réaction ne serait pas suffisamment spécifique, nous devrions obtenir une modification significative des peptides qui ne contiennent pas d'histidines.

Dans un premier temps, le mélange a été traité avec un excès de DEPC par rapport aux peptides (DEPC : peptides 750 :1 [n/n]). L'analyse des peptides a été effectuée en trois étapes principales : la première étape a été l'identification des peptides modifiés par l'augmentation caractéristique de leur masse moléculaire. Ensuite, les acides aminés modifiés ont été identifiés par des expériences en mode MS/MS, ce qui nous a permis d'obtenir également quelques informations sur la réactivité relative des différents résidus. Finalement, les résultats obtenus ont été examinés du point de vue de la spécificité de la réaction et de l'importance des modifications non spécifiques.

III.B.2.1. Identification des peptides modifiés

La carbéthoxylation d'un peptide par le DEPC conduit à une augmentation de sa masse moléculaire de 72,02 Da. Ce changement de la masse moléculaire se traduit par un déplacement caractéristique dans le spectre de masse pour le peptide modifié, ce qui permet une inspection visuelle du spectre pour confirmer si la réaction a bien eu lieu (Fig. III.B.3, encart).

Après une analyse plus approfondie du spectre du mélange modifié, 33 peptides du cytochrome c (modifiés ou non) ont été identifiés par leur masse moléculaire (Tab. III.B.1). Pour chaque peptide modifié, nous avons aussi vérifié que le massif isotopique correspondant était absent dans le spectre du mélange peptidique non modifié. On constate que tous les peptides qui contiennent des histidines ont été modifiés. Néanmoins, la plupart des peptides qui n'ont pas d'histidine dans leur séquence présentent également des modifications, par exemple le peptide GKK qui présente trois carbéthoxylations. Il est donc évident que les résultats ne peuvent pas être expliqués uniquement par une modification des histidines.

Ces résultats pourraient s'expliquer par la modification des lysines (qui portent un groupe $-NH_2$ dans leur chaîne latérale) ainsi que du N-terminal de chaque peptide. Il est important de noter que les expériences ont été effectuées à pH légèrement basique. Les lysines, ainsi que les N-terminaux, sont donc très probablement déprotonés et donc plus réactifs vis-à-vis du DEPC, ce qui expliquerait en partie la réactivité observée.

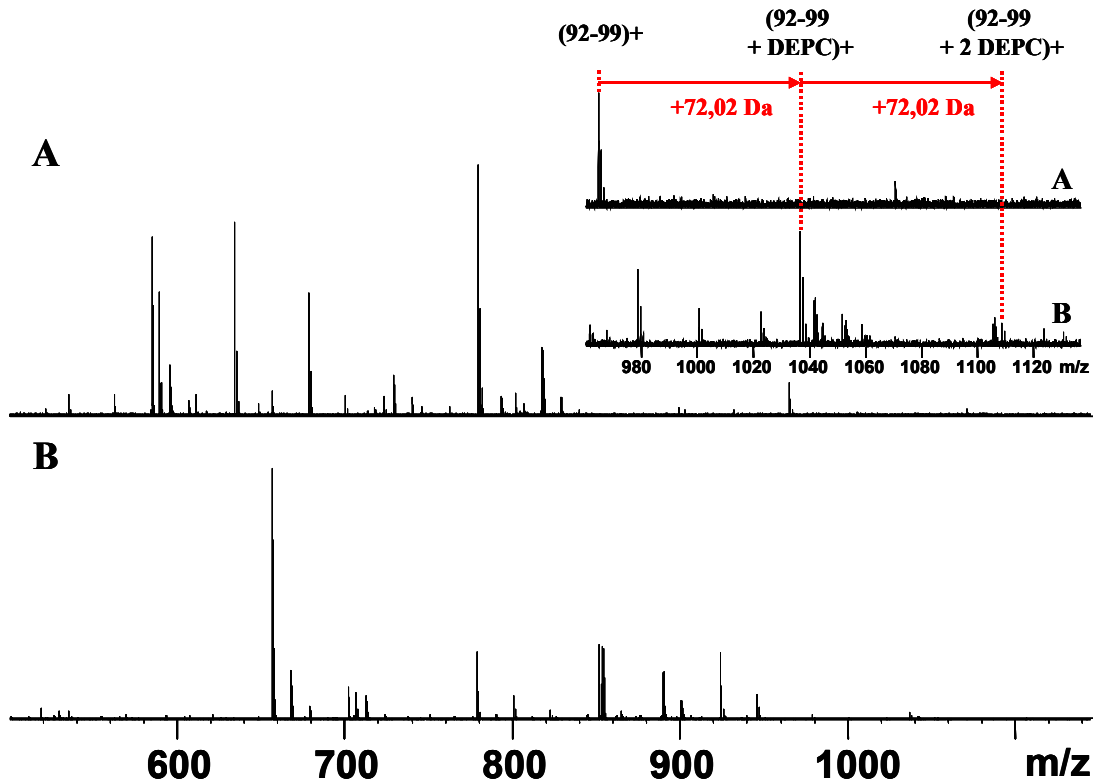


Fig. III.B.3 : Spectres du mélange peptidique avant (A) et après traitement avec le DEPC (B) présentant des différences significatives. L'encart montre l'exemple du peptide correspondant aux acides aminés 92-99 du cytochrome *c* qui a été modifié lors de la réaction avec le DEPC, avec le décalage en masse de 72,02 Da, caractéristique d'une carbéthoxylation.

Dans cette hypothèse, on constate une très bonne corrélation entre le nombre maximal $n_{\max}$ de modifications pour chaque peptide ($n_{\max} = 1 + n_{\text{Lys}} + n_{\text{His}}$, où n_{Lys} et n_{His} représentent le nombre de lysines et d'histidines dans la séquence) et le nombre de modifications déterminé dans l'expérience, n_{exp} (Tab. III.B.1). Sur les 33 peptides identifiés, seuls deux peptides (TGPNLHGLFGR et YIPGTK) présentent plus de modifications qu'attendu. Ces deux peptides ont un maximum de deux sites de modification et présentent chacun un maximum de trois carbéthoxylations. Toutefois, les pics qui correspondent à ces deux peptides portant trois modifications sont de très faible intensité dans le spectre. Il est possible qu'il s'agisse de modifications « non spécifiques », dues au rapport élevé DEPC : peptides qui a été utilisé. D'après le mécanisme de la réaction (substitution nucléophile), on pourrait donc imaginer que les résidus ayant des groupements nucléophiles (comme le groupement $-\text{OH}$ dans les thréonines, ou les tyrosines), bien qu'ayant une réactivité moins élevée que l'histidine et la lysine, pourraient être modifiés à ces hautes concentrations de DEPC. Ces modifications « non spécifiques » seront examinées en détail plus loin (voir III.B.2.3).

Tab. III.B.1 : Peptides du digestat du cytochrome c identifiés après modification par le DEPC. n_{max} correspond au nombre maximum de sites de modification, comme précisé dans le texte.

M / Da	séquence	acides aminés	n_{exp}	n_{max}	M_{theo} / Da	δ / ppm
547,286	GKK	6-8	3	3	547,285	2,0
905,523	KIFVQK	8-13	2	3	905,522	0,9
705,406	IFVQK	9-13	1	2	705,406	0,3
777,428	IFVQK	9-13	2	2	777,427	0,6
520,302	FVQK	10-13	0	2	520,301	1,5
1239,637	TGPNLHGLFGR	28-38	1	2	1239,636	1,1
1311,657	TGPNLHGLFGR	28-38	2	2	1311,657	0,0
1383,680	TGPNLHGLFGR	28-38	3	2	1383,678	1,2
1511,784	TGPNLHGLFGRK	28-39	3	3	1511,773	7,1
1727,798	KTGQAPGFSYTDANK	39-53	2	3	1727,800	-1,2
1527,686	TGQAPGFSYTDANK	40-53	1	2	1527,684	1,5
2080,964	GITWGEETLMEYLENPK	56-72	1	2	2080,966	-0,9
2209,061	GITWGEETLMEYLENPKK	56-73	1	3	2209,061	-0,1
2281,072	GITWGEETLMEYLENPKK	56-73	2	3	2281,082	-4,5
949,516	KYIPGTK	73-79	2	3	949,512	4,3
1021,538	KYIPGTK	73-79	3	3	1021,533	4,3
749,396	YIPGTK	74-79	1	2	749,396	0,0
821,416	YIPGTK	74-79	2	2	821,417	-1,0
893,438	YIPGTK	74-79	3	2	893,438	0,2
850,462	MIFAGIK	80-86	1	2	850,462	0,0
1050,579	MIFAGIKK	80-87	2	3	1050,578	0,7
1122,600	MIFAGIKK	80-87	3	3	1122,599	0,7
647,401	IFAGIK	81-86	0	2	647,401	1,1
719,426	IFAGIK	81-86	1	2	719,422	5,3
534,317	FAGIK	82-86	0	2	534,317	0,0
606,339	FAGIK	82-86	1	2	606,338	1,5
632,314	KGER	88-91	2	2	632,313	1,7
1577,832	KGEREDLIAYLK	88-99	2	3	1577,830	1,1
1649,855	KGEREDLIAYLK	88-99	3	3	1649,851	2,2
1035,549	EDLIAYLK	92-99	1	2	1035,549	0,2
1107,571	EDLIAYLK	92-99	2	2	1107,570	0,6
531,327	LKK	98-100	2	3	531,327	0,6
603,349	LKK	98-100	3	3	603,348	1,0

En résumé, nous avons mis en évidence la carbéthoxylation des peptides à pH 8. Dans les conditions expérimentales utilisées, les histidines, les lysines et le N-terminal des peptides semblent être les régions préférentiellement modifiées. La modification de ces acides aminés sera considérée comme « spécifique » par la suite, tandis que la modification de tout autre acide aminé sera considérée comme « non spécifique ». Des modifications de cette dernière catégorie sont susceptibles de se produire lors de l'utilisation d'un excès de réactif.

III.B.2.2. Identification des sites de modification

D'après les masses moléculaires des peptides modifiés et non modifiés du mélange, il semble qu'à pH 8, les histidines, les lysines et le N-terminal des peptides soient les sites les plus réactifs vis-à-vis du DEPC. Néanmoins, pour valider cette hypothèse, il a été nécessaire d'identifier les sites de modification le plus précisément possible à l'aide de la spectrométrie de masse en tandem. Les peptides modifiés présentant une intensité suffisante ont donc été fragmentés par SORI-CAD.

L'ion dichargé à m/z 656,83, identifié comme étant le peptide 28-38 (TGPNLHGLFGR) bicarbéthoxylé a été choisi pour illustrer les résultats obtenus par cette approche. Dans le spectre, 16 fragments b et y ont été identifiés, montrant sans ambiguïté que l'histidine ainsi que le N-terminal sont les sites portant une carbéthoxylation (Fig. III.B.4), ce qui correspond à la spécificité de la réaction proposée dans la section précédente.

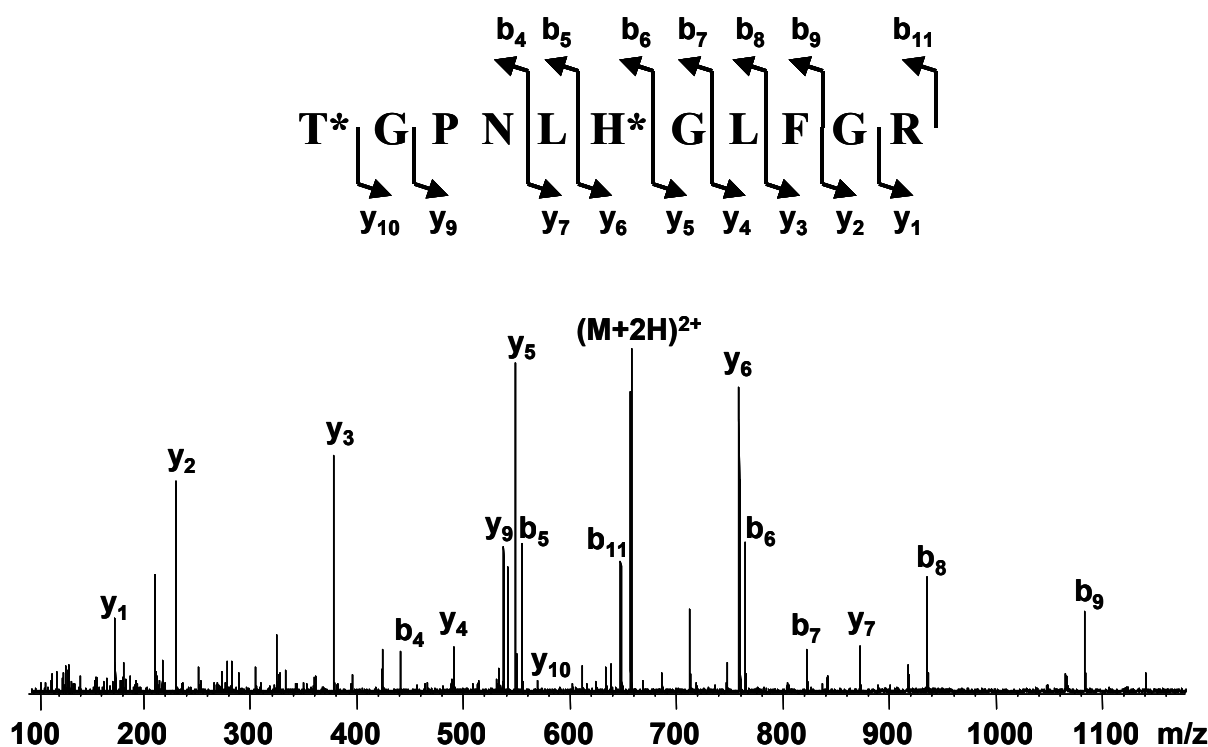


Fig. III.B.4 : Spectre MS/MS (SORI-CAD) du peptide 28-38 (TGPNLHGLFGR), qui porte deux carbéthoxylations (*). Les sites de modification ont été identifiés sans ambiguïté.

Pour les autres peptides étudiés par MS/MS, des résultats similaires ont été obtenus (Tab. III.B.2). Malheureusement, nous n'avons pas pu, par cette approche, mettre en évidence de lysines modifiées. Les peptides présentant des lysines potentiellement modifiées étaient soit

de trop faible intensité, soit la MS/MS sur ces peptides a été trop peu efficace, ou encore, il s'est avéré que le N-terminal a été identifié comme étant le résidu modifié. Le tableau III.B.2 montre trois peptides qui appartiennent à cette dernière catégorie : les peptides IFVQK, MIFAGIK et TGQAPGFSYTDANK présentent chacun une modification localisée à l'extrémité N-terminale, tandis que les lysines ne sont pas modifiées. Néanmoins, il est probable que le groupe $-NH_2$ de la chaîne latérale des lysines ait une réactivité similaire à celle du N-terminal des peptides, ce qui rend la modification de cet acide aminé probable.

Tab. III.B.2 : Peptides modifiés par le DEPC étudiés par spectrométrie de masse en tandem SORI-CAD.

séquence	n_{DEPC}	résidus modifiés
I*FVQK	1	I1 (N-ter)
M*IFAGIK	1	M1 (N-ter)
T*GPNLHGLFGR	1	T1 (N-ter)
T*GPNLH*GLFGR	2	T1 (N-ter), H6
T*GQAPGFSYTDANK	1	T1 (N-ter)

L'identification des sites de modification par MS/MS sur des peptides modifiés est ainsi en accord avec l'hypothèse selon laquelle les histidines, les lysines et les extrémités N-terminales sont les résidus les plus réactifs dans les conditions expérimentales utilisées.

III.B.2.3. Modifications non spécifiques

Il nous reste à évaluer l'importance des modifications qui ont été classées comme étant non spécifiques. Comme indiqué dans la section III.B.2.1, les peptides TGNLHGLFGR et YIPGTK présentent chacun un maximum de trois modifications d'après l'observation expérimentale, alors que chacun de ces deux peptides n'a que deux sites de modification. Cette observation a été attribuée aux modifications non spécifiques qui se sont très probablement produites lors de l'utilisation d'un excès de réactif (voir III.B.2.1).

Afin d'évaluer l'importance de ces modifications non spécifiques, le nombre de modifications moyen par peptide, $\bar{n}_{DEPC}$, a été calculé (selon l'équation III.B.1) en tenant compte des intensités des pics dans le spectre (en faisant l'hypothèse que, pour un peptide donné, la carbéthoxylation ne conduit pas à un changement important de l'efficacité d'ionisation ni de l'efficacité de détection) :

$$\bar{n}_{DEPC} = \frac{\sum_i i \cdot I_i}{\sum_i I_i} \quad (\text{eq. III.B.1})$$

Dans cette equation, i est le nombre de modifications et I_i l'intensité du pic correspondant. Les résultats de ce calcul montrent que le nombre moyen de modifications ne dépasse jamais le nombre maximal de modifications attendues. (Fig. III.B.5).

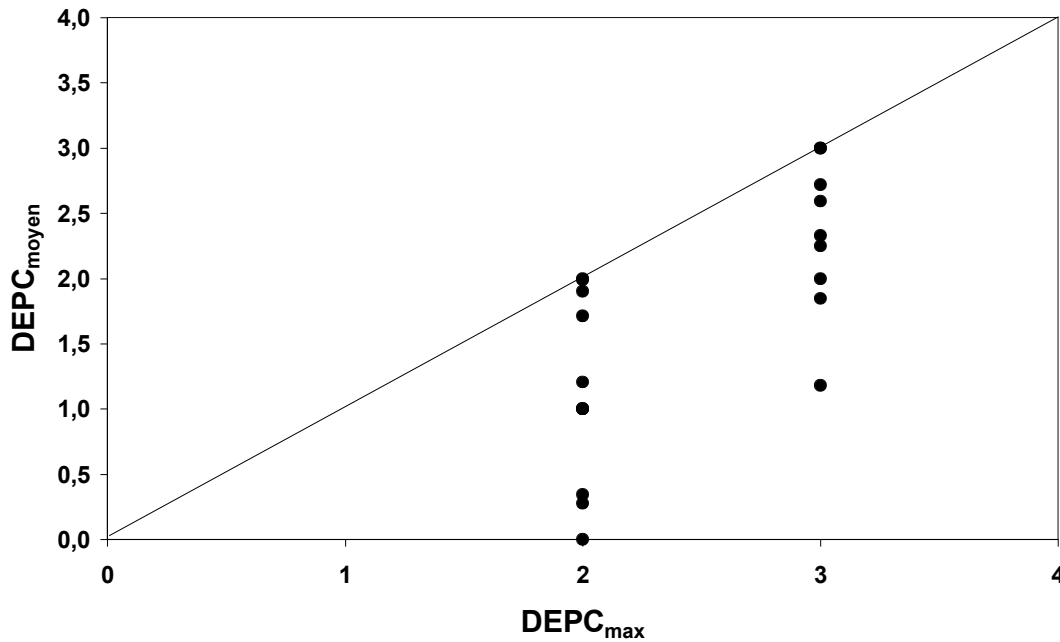


Fig. III.B.5 : Le nombre moyen de modifications ne dépasse pas le nombre maximal de modifications attendues ($DEPC_{moyen} = \bar{n}_{DEPC}$, $DEPC_{max} = n_{max}$).

Le résultat de cette analyse montre que les modifications non spécifiques ne concernent qu'une quantité négligable de peptides dans le mélange. Néanmoins, leur présence prouve qu'en utilisant un grand excès de DEPC, on risque de modifier des acides aminés autres que l'histidine, la lysine et le N-terminal. Ce résultat impose donc des contraintes sur la quantité de DEPC à utiliser lors de la modification des protéines entières. Un rapport DEPC : analyte plus faible sera préférable afin de minimiser autant que possible les problèmes liés aux modifications non spécifiques.

III.B.3. Etude des protéines par modification chimique : approche « bottom-up »

Après avoir analysé la réactivité du DEPC à pH 8 sur un mélange de peptides, la deuxième étape était la modification de protéines entières dans l'objectif d'obtenir des informations sur leur structure tridimensionnelle. La myoglobine a été choisie comme système modèle. Cette protéine de petite taille (17 kDa) devait conduire à un nombre suffisamment petit de peptides lors de la digestion pour ne pas compliquer l'analyse par spectrométrie de masse ni l'identification des sites modifiés, mais elle est suffisamment structurée pour présenter des réactivités différentes pour un acide aminé donné.

Les concentrations élevées de DEPC utilisées lors de l'étude de la réactivité du DEPC avec le mélange peptidique ne seront pas compatibles avec des études de structure des protéines par modification chimique. En effet, on risque d'avoir des modifications non spécifiques, comme cela a été observé dans le cas du mélange de peptides et par ailleurs, un trop grand nombre de modifications risque de perturber la structure tridimensionnelle de la protéine. La modification d'un acide aminé pourrait par exemple conduire à une modification de sa charge (ce qui est le cas pour la carbéthoxylation des lysines par le DEPC), ce qui aurait des conséquences non négligeables sur la structure de la protéine. Le rapport DEPC : protéine utilisé (10 :1) présente ainsi un bon compromis entre maximiser le nombre des modifications et minimiser le risque de perturbation de la structure. La réaction de modification a été arrêtée par ajout d'histidine. L'efficacité de ce « blocage » de la réaction a été démontré par l'ajout d'histidine avant la réaction : aucune modification de la myoglobine n'a été observée dans ce cas (Fig. III.B.6, en bas).

Dans un premier temps, la modification de la protéine entière sera étudiée. Ensuite, l'intérêt de l'approche « bottom-up », c'est-à-dire de l'identification des sites de modification par analyse des peptides modifiés, sera évalué.

III.B.3.1. Modification de la myoglobine intacte par le DEPC

Le spectre de masse de la myoglobine modifiée par le DEPC présente un maximum de quatre carbéthoxylations par protéine (Fig. III.B.6, en haut). Le pic le plus intense correspond à la population avec deux modifications, et on trouve en moyenne deux carbéthoxylations par protéine. Etant donné que la myoglobine présente 31 sites de modification potentiels (11 histidines, 19 lysines et le N-terminal), on peut estimer que la protéine n'est donc pas trop modifiée. Ce faible nombre de modifications ne devrait pas perturber significativement la structure tridimensionnelle de la protéine.

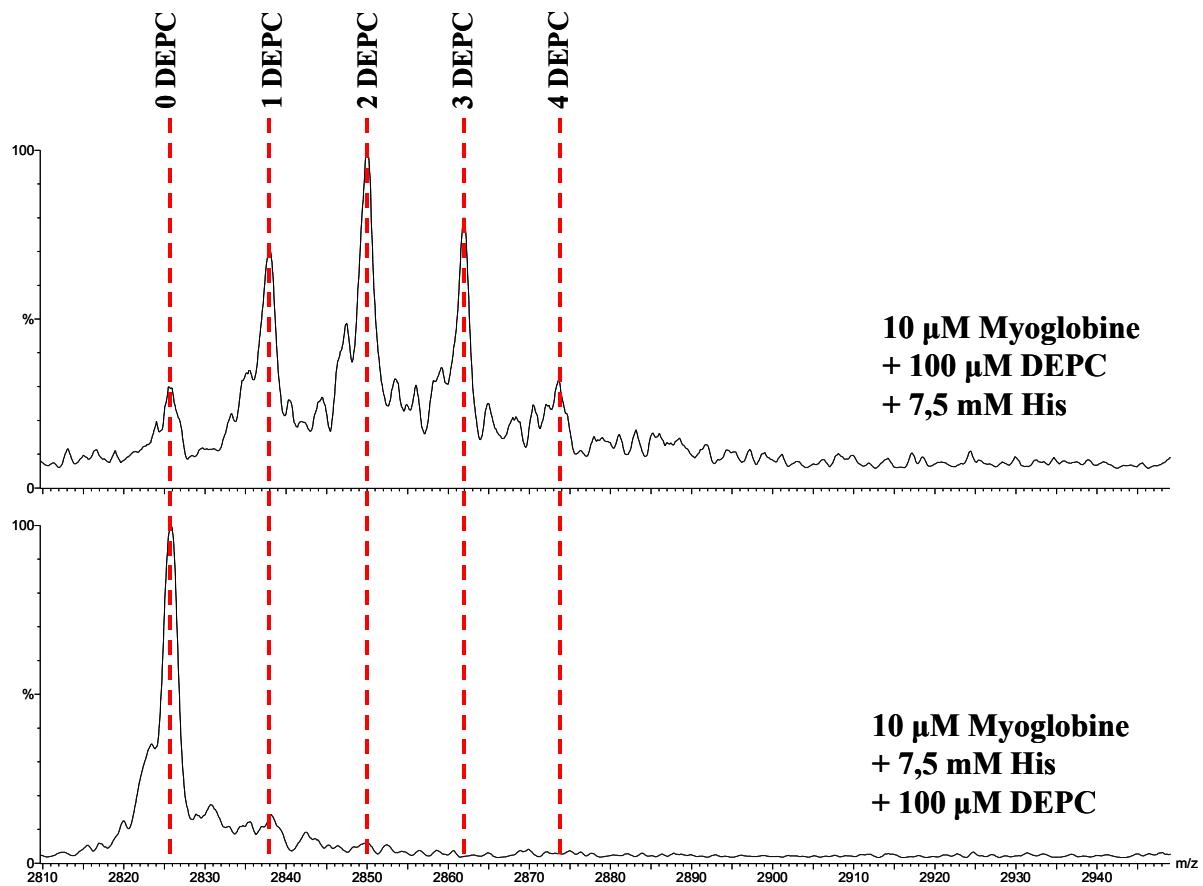


Fig. III.B.6 : Spectre de masse de la myoglobine modifiée par le DEPC. L'histidine a été ajoutée après la réaction (en haut) ou avant la réaction (en bas).

Cependant, le spectre de la protéine entière modifiée ne donne aucune information sur les sites de modification. Il est notamment impossible de savoir sur cette seule base si quatre résidus sont modifiés de façon spécifique ou si un plus grand nombre de résidus sont modifiés « statistiquement » conduisant à un nombre moyen de quatre modifications.

III.B.3.2. Identification des sites de modification par l'approche « bottom-up »

Afin d'étudier en détail les sites de modification, nous avons choisi dans un premier temps l'approche dite « bottom-up ».

La protéine a été digérée par deux protéases différentes : soit par la trypsine, soit par la GluC. La trypsine coupe du côté C-terminal des arginines (R) et des lysines (K). Une étude préliminaire centrée sur l'ubiquitine modifiée a montré que la trypsine coupe aussi après des lysines modifiées. Bien que cette réaction ne semble pas être la plus favorable, on ne peut pas

l'exclure par avance. Il est donc possible qu'une carbéthoxylation soit localisée sur une lysine au niveau C-terminal d'un peptide. La GluC, quant à elle, coupe du côté C-terminal de l'acide glutamique (E) et (plus rarement) du côté C-terminal de l'acide aspartique (D). Ces deux acides aminés ne sont pas modifiés par le DEPC et le clivage de la protéine ne sera donc pas modifié par la carbéthoxylation.

Le spectre de masse de la digestion trypsique de la myoglobine modifiée présente 153 massifs isotopiques. 87 d'entre eux ont pu être attribués à 40 peptides de la myoglobine qui présentent différents nombres de carbéthoxylations. La totalité des peptides modifiés et non modifiés couvre 100% de la séquence primaire de la protéine (Fig. III.B.7). Quant à la digestion par la GluC de la myoglobine modifiée, 22 peptides (modifiés ou non) de la myoglobine ont été retrouvés, ce qui correspond à une couverture de 66,6% de la séquence primaire de la protéine.

Les résultats obtenus sur la myoglobine montrent l'intérêt de la spectrométrie de masse FT-ICR pour ces études : grâce à sa précision en masse, il a été possible d'identifier les peptides modifiés malgré leur faible intensité, sans avoir besoin de recourir à la MS/MS. De même, la totalité des peptides a été analysée en un seul spectre (grâce à la haute résolution) sans nécessité d'une séparation préalable par LC-MS, ce qui simplifie le traitement des données.

Il est tout d'abord important de noter que pour tous les peptides observés, le nombre de modifications observées dans le spectre est inférieur ou égal au nombre maximal de modifications possibles (Fig. III.B.7). Il n'y a donc pas de modifications non spécifiques, ce qui est en accord avec l'hypothèse selon laquelle les modifications non spécifiques ne sont susceptibles de se produire qu'à haute concentration en DEPC.

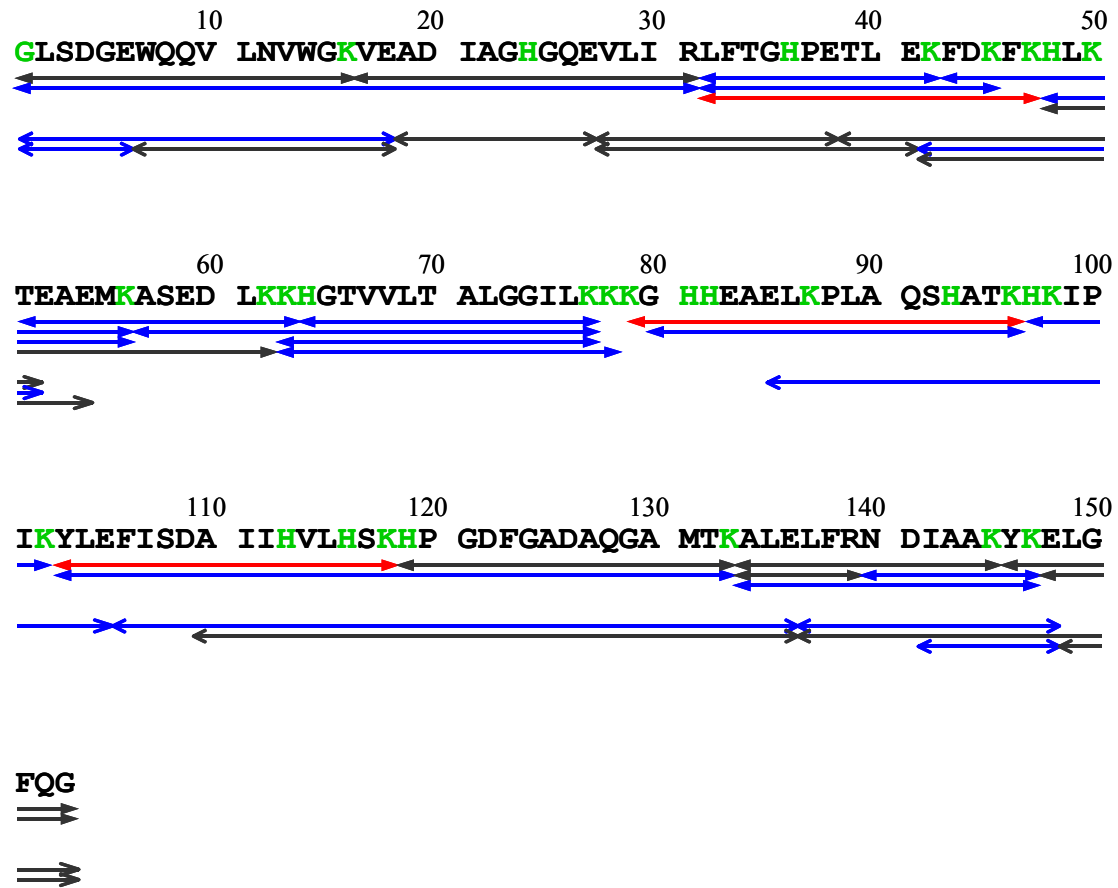


Fig. III.B.7 : Visualisation des peptides portant le nombre maximal de DEPC observés dans les spectres (flèches fermées en haut : digestion par la trypsine ; flèches ouvertes en bas : digestion par la GluC, gris : 0 carbéthoxylations, bleu : 1 carbéthoxylation, rouge : 2 carbéthoxylations). Les sites de modification potentiels sont représentés en vert.

Les résultats montrent qu'il y a au moins 11 sites modifiés dans la protéine (Fig. III.B.7). Ce nombre est plus élevé que le nombre maximal de modifications observées dans le spectre de la protéine entière. Le « profil » de modification observé dans le spectre de la protéine entière serait ainsi plutôt le résultat de modifications statistiques sur un certain nombre de sites potentiels que le résultat de la modification de quatre sites particuliers de la protéine.

De plus, même en utilisant plusieurs protéases, il n'a pas été possible d'identifier les sites de modification à un acide aminé près sans ambiguïté. La plupart des peptides modifiés présentent plus de sites de modification que de modifications observées (Fig. III.B.7). Une analyse plus précise des sites de modification par MS/MS n'était pas possible à cause des faibles intensités des pics correspondant aux peptides modifiés. Ces faibles intensités sont elles-mêmes la conséquence du faible rapport DEPC : protéine qui a été choisi afin de diminuer le risque de perturber la structure tridimensionnelle de la protéine.

Parmi tous les peptides identifiés, le peptide GLSDGE (1-6) portant une carbéthoxylation est le seul peptide pour lequel le nombre de modifications est égal au nombre de sites de modification, seule la glycine N-terminale étant susceptible d'être modifiée. Il serait donc possible d'attribuer également la modification des peptides 1-18 et 1-31 à une modification du N-terminal de la protéine, ce qui correspond aux résultats obtenus avec le mélange des peptides qui montrent que le N-terminal semble être le site le plus réactif. Néanmoins, d'autres exemples montrent qu'une telle affirmation est difficile. Par exemple, le peptide 42-52 présente une modification, tandis que le peptide 42-54 (qui contient aussi l'acide aminé modifié du peptide 42-52) ne porte aucune modification.

Un autre problème est la quantification de la modification. Il est possible qu'un acide aminé modifié se retrouve dans différents peptides après digestion enzymatique. Or, pour déterminer précisément le pourcentage de modification d'un acide aminé, il sera nécessaire de connaître les concentrations de tous les peptides modifiés et non modifiés qui contiennent l'acide aminé en question. Cependant, vues les différences d'efficacité d'ionisation pour les différents peptides (complicé par le fait que la modification par le DEPC est susceptible de la modifier significativement), il n'est pas possible de déterminer les concentrations des différents peptides. En conséquence, la quantification de la modification ne sera pas possible par une approche « bottom-up ».

En résumé, nous avons mis en évidence qu'avec un rapport DEPC : protéine 10 :1, on observe un maximum de quatre carbéthoxylations dans le spectre de la myoglobine. Après digestion de la protéine par la trypsine ou la GluC, les peptides modifiés, bien que présentant de faibles intensités dans le spectre, ont pu être identifiés sans ambiguïté grâce à la précision en masse de la spectrométrie de masse FT-ICR. Les observations sont en accord avec la réactivité du DEPC déterminée précédemment. Néanmoins, il n'a été pas possible de localiser avec précision les sites de modification sauf le N-terminal. De plus, vues les différences d'efficacité d'ionisation des différents peptides, il ne serait pas possible de quantifier la modification. L'ensemble de ces résultats obtenus sur la myoglobine modifiée montrent donc qu'il sera difficile d'obtenir des informations détaillées sur la réactivité des acides aminés, et donc sur la structure de la protéine par l'approche « bottom-up ».

III.B.4. Etude des protéines par modification chimique : approche « top-down »

Les problèmes rencontrés lors de l'approche « bottom-up » pour l'identification des sites de modification nous ont amenés à une stratégie différente : une approche dite « top-down ». Cette approche consiste en la fragmentation de la protéine entière en phase gazeuse. Les sites de modification sont ensuite identifiés par la différence de masse des fragments modifiés par rapport aux fragments non modifiés. Si le nombre de fragments est suffisant, les sites de modification peuvent être identifiés sans ambiguïté. Par ailleurs, le problème de la différence d'efficacité d'ionisation ne se pose plus. La technique utilisée pour la fragmentation des protéines en phase gazeuse est la dissociation par capture d'électrons (« Electron Capture Dissociation », ECD, voir chapitre « Top-Down »). Dans un premier temps, l'ubiquitine a été utilisée comme système modèle. L'ubiquitine présente suffisamment de sites de modification (9) et une masse assez faible pour permettre d'évaluer l'intérêt de cette méthode pour l'identification des sites de modification.

III.B.4.1. Identification des sites de modification par une approche « top-down »

Le spectre de l'ubiquitine modifiée présente des pics correspondant à l'ubiquitine portant entre 3 et 7 carbéthoxylations, le maximum étant 5 modifications. Les expériences ECD ont été effectuées avec les paramètres présentés dans la partie « Fragmentation des protéines par ECD » (partie D). Dans le spectre ECD, un grand nombre de pics a été observé, attribués en grande partie aux fragments de l'ubiquitine modifiés et non modifiés (Fig. III.B.8). Au total, 79 fragments ECD (ions c et z) de l'ubiquitine (modifiés et non modifiés) ont été retrouvés (Fig. III.B.8). Pour tous ces fragments, les erreurs de masse ont été de l'ordre de quelques dizaines de ppm, ce qui s'explique par la pression élevée dans la cellule ICR lors de l'expérience.

On observe au niveau du N-terminal (jusqu'au résidu E16) une série d'ions c quasiment complète (Fig. III.B.8). Entre les résidus E16 et P38, on observe ensuite encore quelques ions c, mais il reste des « trous » dans la séquence. On constate le même résultat au niveau du C-terminal : une série quasiment complète d'ions z de z_5 à z_{32} (F44) et quelques fragments jusqu'à z_{45} (D32). Les séries des ions c et z couvrent ainsi la totalité de la séquence de l'ubiquitine, ce qui permet d'obtenir en principe des informations précises sur les sites de

modification. Cependant, les fragments absents laissent déjà supposer que certaines modifications ne pourront pas être attribuées avec précision.

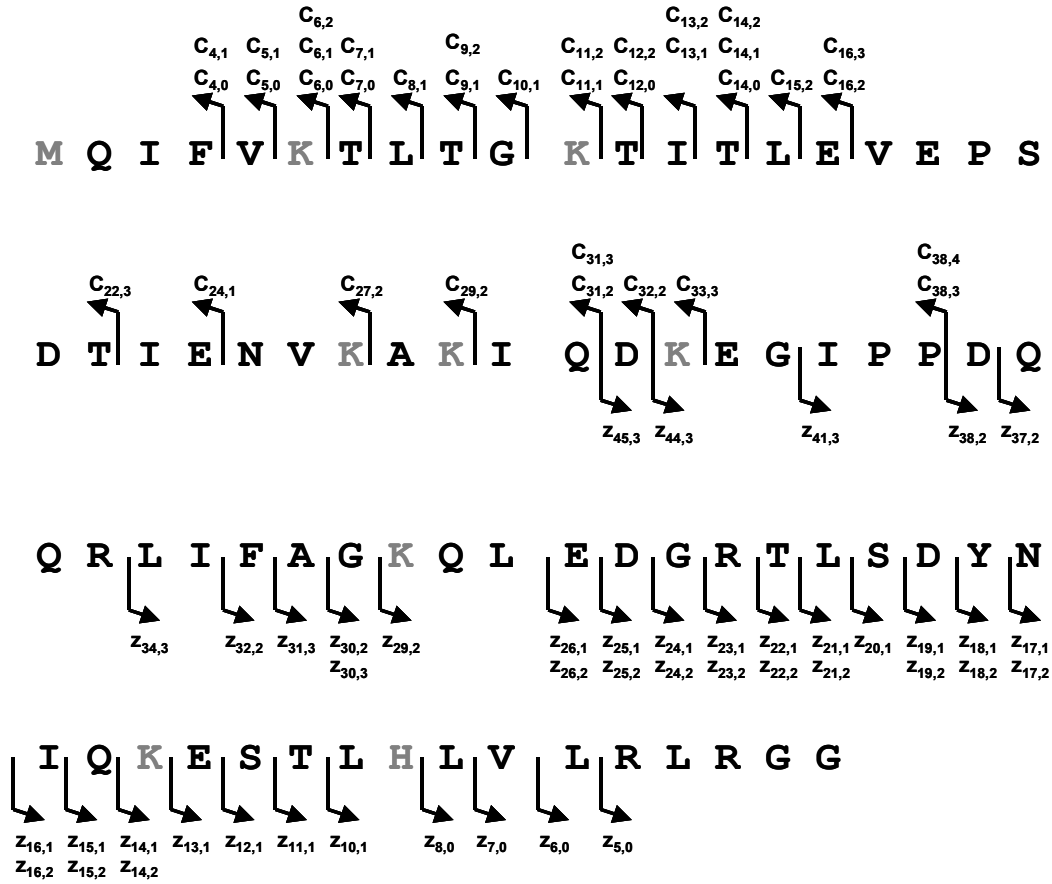


Fig. III.B.8 : Localisation des fragments ECD sur la séquence primaire de l'ubiquitine. Les sites potentiels de modification sont en gris. Pour les fragments $c_{i,j}$ et $z_{i,j}$, i indique l'endroit de la fragmentation dans la séquence et j le nombre de modifications observées.

L'ubiquitine présente 9 sites de modification potentiels au total (N-terminal, 7 lysines, 1 histidine). Le N-terminal est partiellement modifié, comme le montrent les ions c_4 et c_5 non modifiés et modifiés. K6 est également partiellement modifié. On trouve en effet deux fragments, $c_{6,2}$ et $c_{9,2}$, dans lesquels ce résidu est modifié, mais les fragments $c_{6,0}$ et $c_{7,0}$ prouvent que ce résidu n'est pas totalement modifié. De même pour K11, les fragments $c_{16,3}$ et $c_{22,3}$ montrent que ce résidu est modifié, mais, la présence des fragments $c_{12,0}$ et $c_{14,0}$, dans lesquels K11 n'est pas modifié, conduit à la conclusion que ce résidu est partiellement modifié, comme le N-terminal et K6. Pour les autres sites de modification potentiels (K27, K29, K33, K48, K63, H68), la série des ions c ne permet pas de tirer de conclusions précises. L'ion $c_{38,4}$ contient au moins une carbéthoxylation au niveau des résidus K27, K29 et K33. Cependant,

des modifications partielles ayant été observées au niveau du N-terminal, de K6 et de K11, il n'est pas possible de déterminer précisément le ou les site(s) de modification parmi ces résidus.

La série des ions z , permet d'obtenir des informations sur les modifications au niveau C-terminal. Les fragments identifiés suggèrent une modification complète de H68 : à partir de l'ion z_{10} (le premier ion qui contient H68), aucun ion z non modifié n'a été identifié. Ensuite, K63 est modifié partiellement, ce qui est montré par la présence des ions z_{14} à z_{26} présentant une ou deux modifications. H68 étant modifié complètement, la présence de fragments avec une seule modification entre z_{14} et z_{26} montre que K63 est modifié, mais pas complètement. Quant au résidu K48, il est aussi modifié, ce qui est prouvé par la présence des fragments $z_{30,3}$, $z_{31,3}$, $z_{34,3}$ et $z_{41,3}$. L'absence de fragments portant une seule modification à partir de z_{29} suggère que K48 peut être aussi complètement modifié. Cependant, d'après les conclusions concernant K63, il n'y a pas de preuve pour cette hypothèse. Pour les autres sites potentiels de modification, la série des ions z ne permet pas de tirer de conclusions puisque trois modifications ont été le maximum observé.

En conclusion, l'analyse des fragments ECD permet les conclusions suivantes sur la modification des résidus présentant des sites de modification potentiels :

N-terminal	partiellement modifié
K6	partiellement modifié
K11	partiellement modifié
K27, K29, K33	au moins un résidu partiellement modifié
K48	partiellement modifié
K63	partiellement modifié
H68	complètement modifié

Ces résultats obtenus sur l'ubiquitine, une petite protéine présentant peu de sites de modification potentiels, montrent que l'approche « top-down » permet d'obtenir des informations plus précises du point de vue de la localisation des sites modifiés que l'approche « bottom-up » (protéolyse et analyse des peptides). Pour les acides aminés localisés au niveau N- et C-terminal de la protéine, un grand nombre de fragments permet d'obtenir des informations précises sur les sites modifiés. Par contre, pour les sites de modification potentiels localisés au milieu de la séquence (K27, K29, K33), l'absence de fragments dans cette région ne permet pas de déterminer précisément leur état de modification. Ce problème

deviendra plus important pour des protéines de masses moléculaires plus élevées. Néanmoins, les résultats montrent qu'au niveau N- et C-terminal, les fragments permettent d'obtenir des informations précises sur les sites modifiés.

III.B.4.2. Détermination du pourcentage de modification

A la vue des résultats obtenus dans la section précédente, on peut maintenant se demander s'il est possible de quantifier la modification des différents résidus. En faisant l'hypothèse que la carbéthoxylation n'a aucun effet ni sur le mécanisme de fragmentation de l'ubiquitine, ni sur la détection des fragments d'ECD, les intensités relatives des formes modifiées et non modifiées de chaque fragment sont directement proportionnelles aux concentrations relatives des sites modifiés et non modifiés. Il serait donc possible de déterminer le pourcentage de modification pour les différents sites. Dans cette section, des exemples sont analysés afin d'examiner cette question.

Pour obtenir le pourcentage de modification du N-terminal, les intensités des ions c_4 et c_5 dans leurs états modifiés et non modifiés sont prises en compte (équation III.B.2) :

$$\%N_{ter} = \frac{I_{mod}}{I_{mod} + I_{non\ mod}} \cdot 100 \quad (\text{eq. III.B.2})$$

Dans cette equation, I_{mod} et I_{nonmod} sont les intensités des fragments modifiés et non modifiés. Pour c_4 , on obtient 39% de modification, tandis que pour c_5 , on obtient 17% de modification.

Des résultats similaires sont obtenus dans la partie C-terminale. H68 étant complètement modifié, le pourcentage de K63 pourrait être calculé avec une équation similaire (équation III.B.3) :

$$\%K63 = \frac{I_{2DEPC}}{I_{1DEPC} + I_{2DEPC}} \cdot 100 \quad (\text{eq. III.B.3})$$

Dans cette équation, I_{1DEPC} et I_{2DEPC} sont les intensités des pics correspondant aux fragments présentant une ou deux modifications. Pour les fragments z_{14} à z_{26} , on obtient des pourcentages de modification de K63 allant de 0% à 70% (Tab. III.B.3).

Tab. III.B.3 : Pourcentage de modification de K63 obtenu par analyse des intensités des fragments z_{14} au z_{26} portant une et deux modifications.

fragment	%K63
Z ₁₄	38,3
Z ₁₅	56,0
Z ₁₆	43,8
Z ₁₇	17,0
Z ₁₈	33,1
Z ₁₉	33,7
Z ₂₀	0,00
Z ₂₁	58,0
Z ₂₂	46,1
Z ₂₃	66,6
Z ₂₄	57,5
Z ₂₅	55,9
Z ₂₆	28,6

Ces résultats montrent que les fragments obtenus par dissociation par capture d'électrons ne permettent pas une analyse quantitative des sites de modifications. Les hypothèses faites pour calculer les pourcentages de modification dans ce modèle simple ne sont donc probablement pas valables. Il est ainsi possible que la carbéthoxylation joue un rôle dans le mécanisme de fragmentation, ce qui rend une analyse quantitative plus difficile. Il y a aussi une faible possibilité que d'autres acides aminés soient modifiés. Néanmoins, cette dernière hypothèse est peu probable d'après les résultats discutés dans les sections précédentes.

En résumé, les résultats obtenus en utilisant l'approche « top-down » pour l'identification des modifications montrent que la localisation des résidus modifiés, surtout au niveau N- et C-terminal, est beaucoup plus facile qu'avec l'approche « bottom-up ». Par contre, cette approche ne permet pas non plus de quantifier les modifications. De plus, cette approche est limitée par la masse moléculaire de la protéine étudiée (voir chapitre « top-down ») : pour des protéines ayant une masse moléculaire supérieure à 20 kDa, la fragmentation par ECD n'est pas assez efficace pour obtenir des résultats précis sur les sites de modification, à l'exception des régions N- et C-terminales.

III.B.5. Comparaison des différents états structuraux des protéines par modification différentielle

La modification par le DEPC peut aussi servir à obtenir des informations structurales sur des complexes protéiques, par exemple des protéines composées de différentes sous-unités. Dans cette approche, on modifie la protéine dans le complexe et dans son état monomérique (Fig. III.B.9). Le site d'interaction est plus accessible dans le monomère et y sera ainsi plus modifié que dans le complexe. Par des approches « bottom-up » (protéolyse, MS/MS des peptides) ou « top-down » (ECD), on obtiendra ensuite des informations détaillées sur le site de modification et par conséquent sur le site d'interaction entre, par exemple, les différents monomères d'une protéine dimérique ou bien sur le site d'interaction entre une protéine et son ligand. Par une approche voisine, Hager-Braun et Tomer ont mis en évidence le site d'interaction entre les protéines HIVgp120 et CD4 [13].

La créatine kinase, un homodimère, a été choisie comme système modèle. Tout d'abord, les conditions expérimentales ont été mises au point afin de casser les interactions entre les monomères sans dénaturer totalement la protéine et la modification par le DEPC des deux formes a été comparée pour les protéines intactes (c'est-à-dire sans digestion). Ensuite, les peptides obtenus après protéolyse ont été analysés afin d'obtenir des informations précises sur les sites de modification.

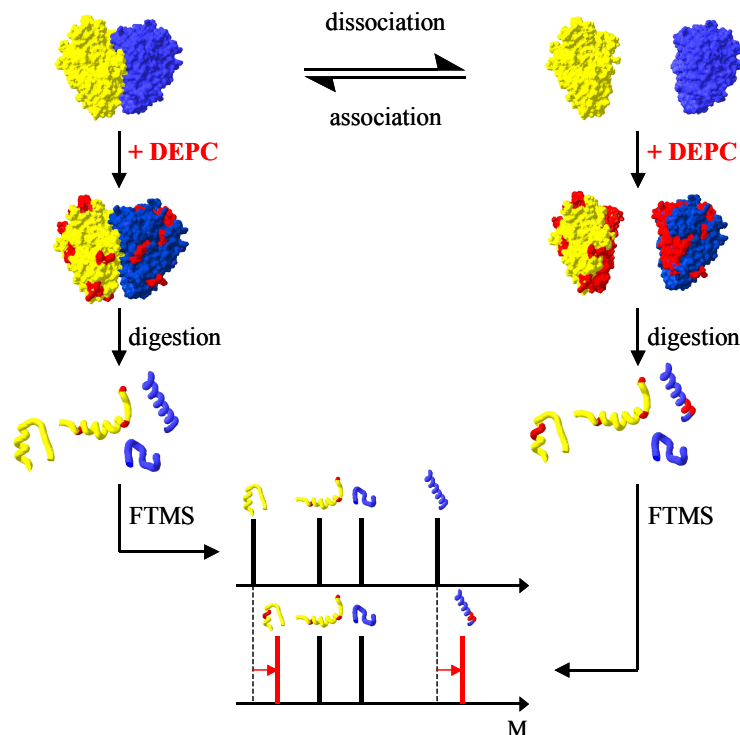


Fig. III.B.9 : Détermination du site d'interaction par modification différentielle du dimère par rapport au monomère.

III.B.5.1. Dénaturation et modification de la CK entière

L'interaction entre les monomères de la CK a été détruite par dénaturation thermique. La dénaturation a été suivie par spectrométrie de masse : avant incubation à 60°C, seul le dimère de la CK était présent dans le spectre. La masse déterminée expérimentalement est de 86154 Da. Après incubation à 60°C pendant 1 minute, le monomère ($M_{\text{exp}} = 43078$ Da) de la CK était la seule espèce présente dans le spectre. Ce résultat montre que les interactions non covalentes entre les monomères de la CK sont perdues lors de la dénaturation thermique partielle. On observe une distribution des états de charges relativement restreinte à m/z élevé (autour de m/z 3000) ce qui correspondrait à une structure plutôt globulaire ou non dénaturée [30, 31]. Toutefois, ceci ne prouve pas véritablement que le monomère ait conservé une structure native.

Le monomère ainsi que le dimère ont été ensuite modifiés par le DEPC. Le dimère a été dénaturé après modification afin de faciliter la comparaison des deux formes. La haute résolution de la spectrométrie de masse FT-ICR a permis de distinguer les pics correspondant à des espèces moléculaires modifiées par différents nombres de DEPC (Fig. III.B.10). Le monomère présente plus de modifications ($\bar{n}_{\text{DEPC}} = 17,5 \pm 0,8$) que le dimère ($\bar{n}_{\text{DEPC}} = 5,7 \pm 0,4$) par monomère (Fig. III.B.10). Ce résultat sur la protéine modifiée et non digérée est donc en accord avec l'hypothèse selon laquelle plusieurs sites potentiels ne sont pas accessibles à la modification dans le dimère.

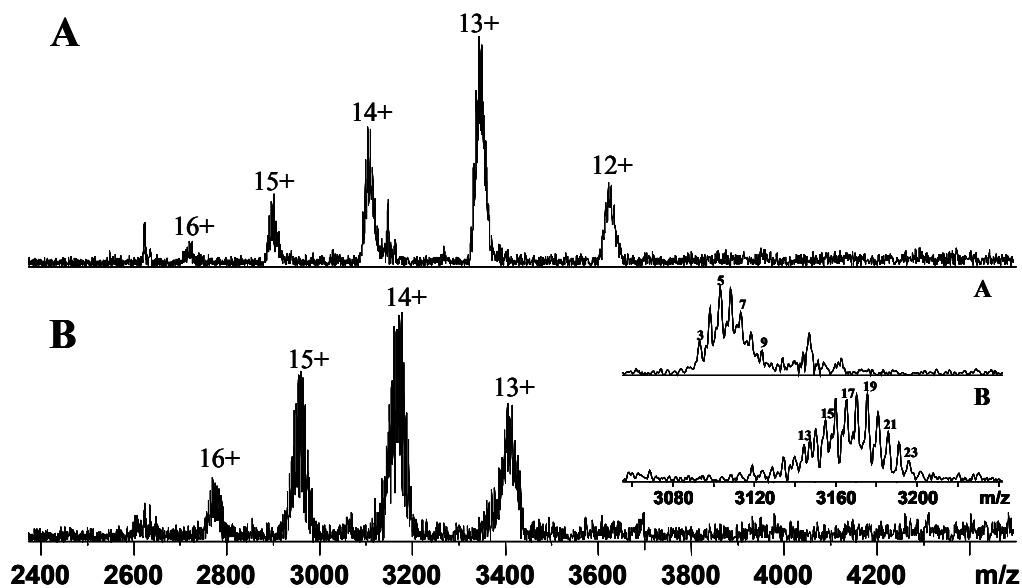


Fig. III.B.10 : Modification de la CK par le DEPC : dimère (A) et monomère (B). L'agrandissement de l'état de charge 14+ montre les différentes modifications en détail.

Toutefois, il serait possible que ce résultat soit dû à une différence de la solution. Il est en effet possible que le chauffage de la solution ait abouti à l'évaporation de CO_2 et de NH_3 (suite à la décomposition du NH_4HCO_3). La diminution de la concentration du tampon pourrait avoir une influence sur la réactivité du DEPC, qui est susceptible de diminuer en présence de NH_3 (d'après le mécanisme réactionnel).

Pour vérifier que les différences observées sont liées à une différence de structure du monomère entre les deux états de la protéine, nous avons donc effectué une étude sur un échantillon qui contenait le monomère et le dimère. Nous avons constaté la présence simultanée des espèces observées dans les spectres du dimère (Fig. III.B.10, A) et du monomère (Fig. III.B.10, B). Ce résultat démontre que les différences du nombre de modifications entre dimère et monomère sont liées aux différences de structure de la protéine, et non pas à des différences au niveau de la solution.

III.B.5.2. Comparaison des peptides modifiés entre les deux formes

Après avoir mis en évidence la modification de la protéine entière, le prochain objectif a été l'identification des peptides modifiés dans les deux échantillons (sans chercher à identifier précisément les acides aminés modifiés dans un premier temps), ce qui a été réalisé par analyse par spectrométrie de masse FT-ICR des digestions du dimère et du monomère par la trypsine. La digestion par la GluC des deux formes n'a pas fourni de résultats supplémentaires et ne sera donc pas présentée.

Dans le spectre du dimère digéré, 38 peptides ont été identifiés avec des nombres différents de carbéthoxylations. Ces peptides étaient absents dans un digestat de la CK non modifiée. Les peptides identifiés correspondent à une couverture de 69,2% de la séquence de la protéine. Dans tous les peptides modifiés, le nombre de modifications ne dépasse pas le nombre maximal de sites potentiels de modification. Dans le spectre du monomère digéré à la trypsine, 46 peptides de la CK avec différents nombres de carbéthoxylations ont été identifiés par leur masse moléculaire. Les peptides couvrent 76,3% de la séquence primaire de la protéine.

Quasiment tous les peptides identifiés dans le digestat du dimère sont aussi présents dans le spectre du digestat du monomère (Tab. III.B.4, colonne de gauche). La seule exception est le peptide 236-246 (modifié par deux carbéthoxylations), qui n'est présent que dans le spectre du dimère. Il est possible que la région correspondante de la protéine soit moins accessible dans le monomère. Cependant, l'intensité de ce peptide dans le spectre du digestat du dimère

est très faible. Il est donc susceptible d'être caché dans le bruit fond, et ce n'est donc pas une différence significative entre les deux spectres.

Plusieurs différences sont observées entre les deux spectres. Dans le spectre du monomère, on constate la présence d'un certain nombre de peptides absents dans le spectre du dimère (Tab. III.B.4, colonne de droite). Il est important de noter que ces peptides portent tous des carbéthoxylations, ce qui montre qu'ils sont accessibles. Cette observation correspond à une modification plus importante du monomère que du dimère (Fig. III.B.10). Ce résultat est cohérent avec l'augmentation attendue de l'accessibilité de certaines régions du monomère par rapport au dimère.

Qualitativement, les résultats obtenus avec la protéine digérée sont donc en accord avec ceux obtenus avec la protéine entière¹. Comme nous l'avons vu précédemment, il est difficile d'établir une quantification absolue de la modification. Par conséquent, nous avons cherché à établir une quantification relative par comparaison des deux spectres. Les peptides communs aux deux échantillons ont ainsi été comparés suivant l'approche utilisée par Hager-Braun et Tomer pour une modification différentielle des arginines [13]. Pour tous les peptides, le rapport de l'intensité du peptide dans le spectre du monomère sur l'intensité du peptide dans le spectre du dimère a été déterminé et tracé en fonction de la localisation du peptide dans la séquence de la protéine (Fig. III.B.11).

D'après les résultats obtenus, une différence d'intensité supérieure à deux est considérée comme significative. Pour la plupart des peptides, le rapport des intensités entre les deux formes de la protéine est autour de l'unité, indiquant que l'abondance ne varie pas ou très peu entre monomère et dimère. Pour les peptides portant des sites de modification potentiels, cela indique qu'il n'y a pas de changement important de la réactivité avec le DEPC lors de la dénaturation de la protéine. Pour les résidus jusqu'à 170 et entre 259-357, la réactivité ne semble pas être trop différente entre les deux formes, à l'exception des peptides trouvés en plus dans le spectre du monomère. Pour les peptides présentant une différence entre les deux formes, il est important de noter que tous les peptides qui ont une intensité plus élevée dans le cas du monomère sont des peptides modifiés. Ce résultat est aussi en accord avec le fait que le monomère présente plus de modifications que le dimère.

¹ Le terme « protéine entière » fait référence à la protéine avant digestion, et non pas à la structure quaternaire intacte (dimère).

Tab. III.B.4 : Peptides de la CK (monomère et dimère) identifiés après modification par le DEPC et digestion enzymatique. La position correspond à la localisation du peptide dans la séquence primaire de la protéine, n_{DEPC} au nombre de carbethoxylations du peptide.

<u>monomère/dimère</u>	
position	n_{DEPC}
32-39	0
32-39	1
32-40	1
86-95	0
107-129	1
116-129	0
116-129	1
138-147	1
156-169	0
156-171	1
170-176	1
172-176	0
172-208	1
177-208	0
177-208	1
215-222	1
215-235	1
215-241	2
223-235	0
223-235	1
236-246	1
242-250	1
252-258	1
259-264	0
259-265	1
265-291	2
292-303	1
292-303	2
298-306	1
307-313	0
319-340	0
320-340	0
341-357	0
358-364	0
358-365	1
369-380	0

<u>dimère</u>	
position	n_{DEPC}
236-246	2

<u>monomère</u>	
position	n_{DEPC}
15-39	3
9-14	1
86-104	1
105-115	2
107-115	1
107-129	3
170-208	1
170-208	2
177-208	2
304-313	1

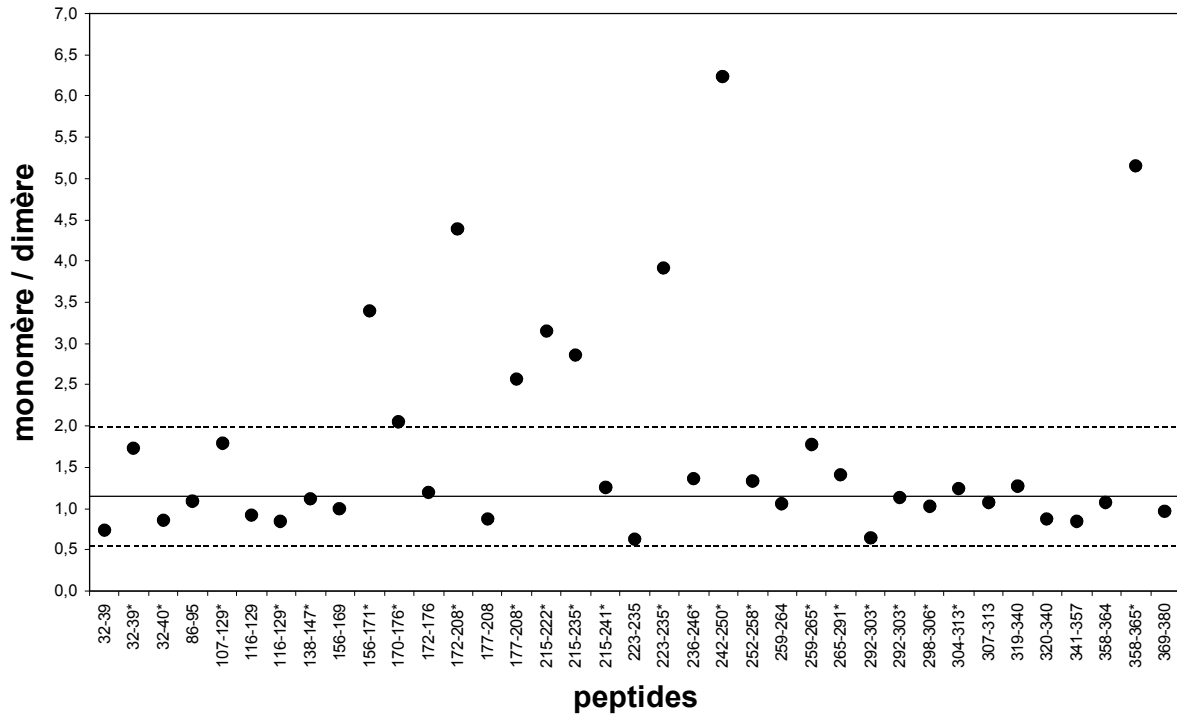


Fig. III.B.11 : Comparaison des intensités des peptides communs dans les spectres des digestions du monomère et du dimère de la CK modifiée par le DEPC.

En faisant l'hypothèse que tous les résidus potentiellement modifiables d'un peptide sont modifiés il est possible de visualiser la modification différentielle sur la structure de la CK qui a été obtenue par cristallographie aux rayons X (PDB : 2CRK [32]). Une différence significative de modification d'un résidu (Fig. III.B.11) est colorée en vert (Fig. III.B.12). En revanche, les résidus qui ne présentent pas une différence significative sont colorés en rouge (Fig. III.B.12). Les résidus qui n'ont pas pu être identifiés sont colorés en noir (Fig. III.B.12). Il est important de noter que cette approche implique l'hypothèse que la structure tridimensionnelle du monomère n'est pas modifiée lors de la dénaturation thermique, ce qui n'est pas certain. La visualisation sur la structure cristallographique pourrait ainsi être utile afin d'obtenir une première idée, mais ne représente probablement pas un modèle bien adapté pour caractériser les conditions réelles en solution.

Plusieurs informations peuvent être obtenues par cette approche. Tout d'abord, les résidus qui sont plus modifiés dans le monomère ne semblent pas localisés dans une région particulière de la protéine. La distribution de ces résidus n'est donc pas en accord avec l'hypothèse simple que seuls les résidus localisés dans le site d'interaction entre les deux sous-unités deviennent plus accessibles lors de la dénaturation.

Toutefois, il semble que les résidus potentiellement modifiables localisés autour du site actif sont plus modifiés dans le monomère que dans le dimère (dans la représentation à droite dans la Fig. III.B.12, ces résidus sont localisés autour des feuillets bêta au milieu de la structure). Ce résultat peut être interprété comme le signe d'une « ouverture », lors de la dénaturation thermique, du site actif, qui présente déjà une certaine accessibilité dans le dimère (pour permettre le fonctionnement de l'enzyme). Néanmoins, il est difficile de prouver cette hypothèse sur la seule base des résultats obtenus.

On peut également observer que certaines régions, comme par exemple le N-terminal, ne présentent pas de différence significative de la modification par le DEPC. Il est possible que ces régions de la protéine ne subissent pas de changements importants lors de la dénaturation, bien que ces résultats ne constituent pas une preuve suffisante de cette hypothèse.

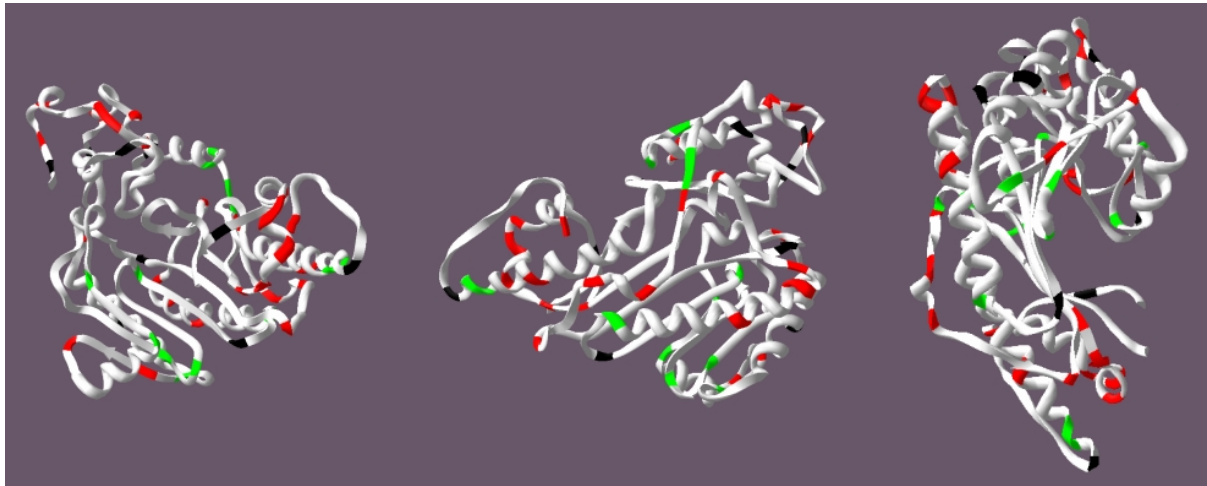


Fig. III.B.12 : Visualisation des modifications différentielles de la structure du monomère de la CK (vert : résidus plus modifiés sur le monomère ; rouge : pas de différence de modification entre dimère et monomère ; noir : résidu non trouvé ; gris : résidu non modifiable par le DEPC). Le monomère est présenté sous trois angles différents pour mieux visualiser les régions qui présentent des différences.

Au vu de ces résultats, la dénaturation thermique ne conduit donc pas seulement à la perte d'interaction entre les deux sous-unités, et ainsi à une augmentation de la réactivité des résidus localisés dans le site d'interaction vis-à-vis du DEPC, mais perturbe également d'autres régions de la structure. Lyubarev et al. ont proposé un modèle d'après lequel la perte des interactions entre les sous-unités de la CK lors du chauffage à 60°C aboutit directement à la dénaturation complète des monomères [33]. Même si nos observations indiquent plutôt une

dénaturation partielle, cette hypothèse expliquerait que les changements structuraux ne soient pas limités à l'interface entre les monomères.

Les différences observées entre les deux formes de la CK montrent donc l'utilité de la modification chimique « différentielle » des protéines par le DEPC pour suivre des modifications structurales.

En résumé, nous avons mis en évidence que le marquage chimique pourrait être utilisé comme indicateur de changements structuraux d'une protéine. Dans ce contexte, la comparaison de la modification entre différents états de la protéine entière (réplée versus déplée, avec ligand versus sans ligand, etc.) permet d'obtenir des informations globales. Les différences de modification des peptides, pour leur part, permettent de localiser plus précisément les régions de la structure de la protéine qui subissent un changement.

III.B.6. Conclusions

A pH 8, le DEPC réagit principalement avec des histidines, des lysines et le N-terminal d'un peptide. On peut observer des modifications non spécifiques si on utilise un large excès de DEPC. La réaction est donc assez spécifique pour être utile dans l'objectif d'étudier les protéines.

Le DEPC a été ensuite employé afin de modifier les protéines. L'analyse des protéines modifiées a été réalisée par différentes approches. L'approche « bottom-up » (digestion enzymatique, analyse des peptides) ne permettait ni de localiser précisément les modifications, ni de les quantifier. Cette approche a cependant permis de déterminer les régions qui subissent des modifications structurales lors de la dénaturation thermique de la créatine kinase (CK).

En revanche, la fragmentation de la protéine entière en phase gazeuse par dissociation par capture d'électron (ECD), approche dite « top-down », permet de localiser les modifications plus précisément, surtout au niveau N- et C-terminal. Cette approche est donc plus précise que l'approche « bottom-up ». Néanmoins, il ne serait, de façon générale, pas non plus possible de quantifier les modifications par cette approche. De plus, pour l'instant, l'utilisation de cette approche est limitée aux protéines avec des masses moléculaires inférieure à 20-25 kDa.

En conclusion, la modification spécifique des acides aminés par le DEPC est un outil analytique utile pour l'analyse des protéines, surtout des protéines composées de différentes sous-unités. Cette technique pourrait constituer un complément précieux pour l'analyse de la

structure des protéines. Cependant, l'analyse des données est complexe et les informations accessibles dépendent de la stratégie analytique employée.

III.B.7. Bibliographie

- [1] J. R. Engen, D. L. Smith, Investigating the higher order structure of proteins: hydrogen exchange, proteolytic fragmentation, and mass spectrometry, *Meth. Mol. Biol.* **2000**, 146, 95-112.
- [2] D. S. Tawfik, Ethoxyformylation of Histidine, dans : *The Protein Protocols Handbook*, J. M. Walker ed., Humana Press, Totowa (NJ) 2002, pp. 473-474.
- [3] T. D. Wood, Z. Guan, C. L. Borders Jr., L. H. Chen, G. L. Kenyon, F. W. McLafferty, Creatine kinase: essential arginine residues at the nucleotide binding site identified by chemical modification and high-resolution tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 3362-3365.
- [4] D. Delforge, B. Devreese, M. Dieu, E. Delaive, J. van Beeumen, Identification of lysine 74 in the pyruvate binding site of alanine dehydrogenase from *Bacillus subtilis*. *J. Biol. Chem.* **1997**, 272, 2276-2284.
- [5] D. H. Hambly, M. L. Gross, Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J. Am. Soc. Mass Spectrom.* **2005**, 16, 2057-2063.
- [6] J. Santrucek, M. Strohalm, V. Kadlcik, R. Hynek, M. Kodicek, Tyrosine residues modification studied by MALDI-TOF mass spectrometry. *Biochem. Biophys. Res. Commun.* **2004**, 323, 1151-1156.
- [7] M. Strohalm, J. Santrucek, R. Hynek, M. Kodicek, Analysis of tryptophan surface accessibility in proteins by MALDI-TOF mass spectrometry. *Biochem. Biophys. Res. Commun.* **2004**, 323, 1134-1138.
- [8] J. S. Sharp, J. M. Becker, R. L. Hettich, Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* **2004**, 76, 672-683.
- [9] M. Strohalm, M. Kodicek, M. Pechar, Tryptophan modification by 2-hydroxy-5-nitrobenzyl bromide studied by MALDI-TOF mass spectrometry. *Biochem. Biophys. Res. Commun.* **2003**, 312, 811-816.
- [10] J. F. Leite, M. Cascio, Probing the topology of the glycine receptor by chemical modification coupled to mass spectrometry. *Biochemistry* **2002**, 41, 6140-6148.

- [11] V. N. Lapko, X.-Y. Jiang, D. L. Smith, P.-S. Song, Surface topography of phytochrome A deduced from specific chemical modification with iodoacetamide. *Biochemistry* **1998**, *37*, 12526-12535.
- [12] D. Suckau, M. Mak, M. Przybylski, Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5630-5634.
- [13] C. Hager-Braun, K. B. Tomer, Characterization of the tertiary structure of soluble CD4 bound to glycosylated full-length HIVgp120 by chemical modification of arginine residues and mass spectrometric analysis. *Biochemistry* **2002**, *41*, 1759-1766.
- [14] G. J. Carven, L. J. Stern, Probing the ligand-induced conformational change in HLA-DR1 by selective chemical modification and mass spectrometric mapping. *Biochemistry* **2005**, *44*, 13625-13637.
- [15] G. Xu, K. Takamoto, M. R. Chance, Radiolytic modification of basic amino acid residues in peptides : probes for examining protein-protein interactions. *Anal. Chem.* **2003**, *75*, 6995-7007.
- [16] S. Izumi, H. Kaneko, T. Yamazaki, T. Hirata, S. Kominami, Membrane topology of guinea pig cytochrome P450 17alpha revealed by a combination of chemical modifications and mass spectrometry. *Biochemistry* **2003**, *42*, 14663-14669.
- [17] A. Scaloni, P. Ferranti, G. de Simone, G. Mamone, N. Sannolo, A. Malorni, Probing the reactivity of nucleophile residues in human 2,3-diphosphoglycerate/deoxy-hemoglobin complex by aspecific chemical modifications. *FEBS Lett.* **1999**, *452*, 190-194.
- [18] W. Fiedler, C. Borchers, M. Macht, S.-O. Deininger, M. Przybylski, Molecular characterization of a conformational epitope of hen egg white lysozyme by differential chemical modification of immune complexes and mass spectrometric peptide mapping. *Bioconjugate Chem.* **1998**, *9*, 236-241.
- [19] S. Akashi, M. Shirouzu, T. Terada, Y. Ito, S. Yokoyama, K. Takio, Characterization of the structural difference between active and inactive forms of the Ras protein by chemical modification followed by mass spectrometric peptide mapping. *Anal. Biochem.* **1997**, *248*, 15-25.
- [20] G. D. Hobba, B. E. Forbes, E. J. Parkinson, G. L. Francis, J. C. Wallace, The insulin-like growth factor (IGF) binding site of bovine insulin-like growth factor binding protein-2 (bIIGFBP-2) probed by iodination. *J. Biol. Chem.* **1996**, *271*, 30529-30536.
- [21] B. Ehrhard, R. Misselwitz, K. Welfle, G. Hausdorf, R. W. Glaser, J. Schneider-Mergener, Chemical modification of recombinant HIV-1 capsid protein p24 leads to the

release of a hidden epitope prior to changes of the overall folding of the protein. *Biochemistry* **1996**, *35*, 9097-9105.

[22] M. Ploug, H. Rahbek-Nielsen, V. Ellis, P. Roepstorff, K. Dano, Chemical modification of the urokinase-type plasminogen activator and its receptor using tetranitromethane. Evidence for the involvement of specific tyrosine residues in both molecules during receptor-ligand interaction. *Biochemistry* **1995**, *34*, 12524-12534.

[23] M. Tsubaki, K. Kobayashi, T. Ichise, F. Takeuchi, S. Tagawa, Diethyl pyrocarbonate modification abolishes fast electron accepting ability of cytochrome b561 from ascorbate but does not influence electron donation to monodehydroascorbate: identification of the modification sites by mass spectrometric analysis. *Biochemistry* **2000**, *39*, 3276-3284.

[24] M. O. Glocker, M. Kalkum, R. Yamamoto, J. Schreurs, Selective biochemical modification of functional residues in recombinant human macrophage colony-stimulating factor b (rhM-CSF b): identification by mass spectrometry. *Biochemistry* **1996**, *35*, 14625-14633.

[25] M. Lemaire, J.-M. Schmitter, E. Issakidis, M. Miginiac-Maslow, P. Decottignies, Essential histidine at the active site of sorghum leaf NADP-dependent malate dehydrogenase. *J. Biol. Chem.* **1994**, *269*, 27291-27269.

[26] R. J. Hondal, S. Ma, R. M. Caprioli, K. E. Hill, R. F. Burk, Heparin-binding histidine and lysine residues of rat selenoprotein P. *J. Biol. Chem.* **2001**, *276*, 15823-15831.

[27] K. Qin, Y. Yang, P. Mastrangelo, D. Westaway, Mapping Cu(II) binding sites in prion proteins by diethyl pyrocarbonate modification and matrix-assisted laser desorption-time of flight (MALDI-TOF) mass spectrometric footprinting. *J. Biol. Chem.* **2002**, *277*, 1981-1990.

[28] P. Dominici, B. Tancini, C. B. Voltattorni, Chemical modification of pig kidney 3,4-dihydroxyphenylalanine decarboxylase with diethyl pyrocarbonate. *J. Biol. Chem.* **1985**, *260*, 10583-10589.

[29] W. B. Melchior Jr., D. Fahrney, Ethoxyformylation of proteins. Reaction of ethoxyformic anhydride with α -chymotrypsin, pepsin and pancreatic ribonuclease at pH 4. *Biochemistry* **1970**, *9*, 251-258.

[30] S. K. Chowdhury, V. Katta, B. T. Chait, Probing conformational changes in proteins by mass spectrometry, *J. Am. Chem. Soc.* **1990**, *112*, 9012-9013.

[31] M. Hautreux, N. Hue, A. Du Fou de Kerdaniel, A. Zahir, V. Malec, O. Lapr v te, Under non-denaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface, *Int. J. Mass Spectrom.* **2004**, *231*, 131-137.

- [32] J. K. Rao, G. Bujacz, A. Wlodawer, Crystal structure of rabbit muscle creatine kinase. *FEBS Lett.* **1998**, 439, 133-137.
- [33] A. E. Lyubarev, B. I. Kurganov, V. N. Orlov, H.-M. Zhou, Two-state irreversible thermal denaturation of muscle creatine kinase. *Biophys. Chem.* **1999**, 79, 199-204.

Chapitre III.C

Structures de protéines étudiées par
échanges hydrogène/deutérium

III.C.1. Introduction

La structure d'une protéine est essentiellement déterminée par des interactions non covalentes intra- et intermoléculaires, elles-mêmes dépendantes de la séquence primaire de la protéine. Par conséquent, l'analyse de ces interactions non covalentes devrait permettre d'obtenir des informations sur la structure d'une protéine. Parmi les différentes interactions non covalentes, les liaisons hydrogènes jouent un rôle structurant primordial, en particulier parce qu'elles sont au cœur des hélices alpha et des feuillets beta, éléments de la structure secondaire des protéines, dont les interactions entre eux conduisent aux structures d'ordre supérieur (structures tertiaire et quaternaire). Dans des conditions expérimentales convenablement choisies, les groupements chimiques impliqués dans les liaisons hydrogènes sont susceptibles d'échanger leurs hydrogènes avec ceux du solvant. Il est donc intéressant d'exploiter cette propriété des liaisons hydrogènes pour en tirer des informations sur la liaison, et par la suite sur la structure de la protéine induite par ces liaisons ou par leur absence.

Une protéine présente deux types d'atomes d'hydrogène : ceux qui sont liés de manière covalente à un atome de carbone et qui ne peuvent s'échanger en conditions usuelles et ceux qui sont liés à des hétéroéléments (oxygène, azote, soufre), pour lesquels un échange est possible. On dénomme ces derniers « labiles » [1]. Les hydrogènes labiles peuvent être subdivisés en deux groupes en fonction de leurs cinétiques d'échange : tout d'abord les hydrogènes labiles des chaînes latérales qui s'échangent presque toujours à une vitesse très rapide (constantes cinétiques autour de 10^9 - 10^{11} s⁻¹), à l'exception des amides primaires de l'asparagine et de la glutamine. Les hydrogènes de la liaison peptidique, qui constituent un second groupe d'hydrogènes labiles, s'échangent plus lentement (constantes cinétiques de 10^3 s⁻¹ à moins de 10^{-4} s⁻¹) [1] et représentent la plus grande part des hydrogènes d'amides (en dehors des hydrogènes d'amides des chaînes latérales) dont les vitesses d'échange sont compatibles avec une détection par spectrométrie de masse (voir Fig. III.C.1).

Cependant, les hydrogènes de la liaison peptidique n'échangent pas tous à la même vitesse. La vitesse exacte d'échange pour un hydrogène d'une liaison peptidique donnée dépend de son environnement chimique (par exemple son accessibilité en solution). D'après Hvidt et Nielsen [1], on peut séparer les hydrogènes des liaisons peptidiques d'une protéine en deux groupes en fonction de leur cinétique d'échange : ceux qui forment des liaisons hydrogènes intramoléculaires ainsi que ceux qui sont localisés à « l'intérieur » d'une protéine (c'est-à-dire loin de la surface accessible au solvant ou dans des régions hydrophobes) s'échangent très lentement. En pratique, la cinétique d'échange pour ce type d'hydrogènes

n'est pas mesurable (quasiment pas d'échange en 24 h). En revanche, les hydrogènes d'amides qui sont localisés à la « surface » de la protéine, c'est-à-dire accessibles au solvant et ayant des liaisons hydrogènes avec des molécules de solvant, s'échangent à la même vitesse que ceux d'un peptide non structuré. Pour une conformation donnée d'une protéine, chaque hydrogène d'amide peut être classé dans un de ces deux groupes.

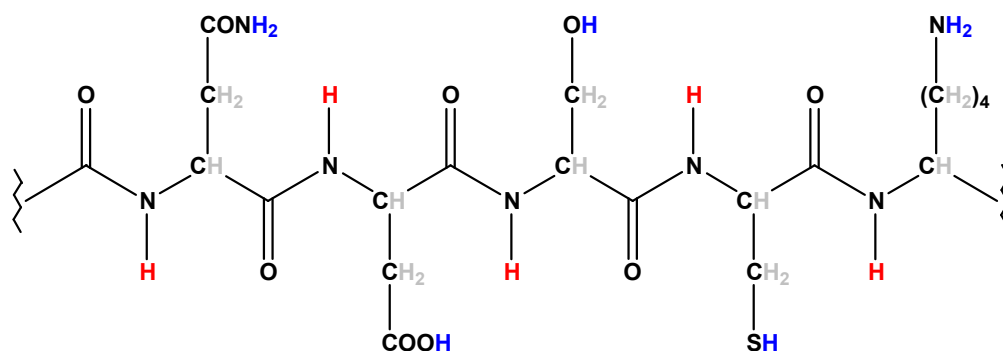


Fig. III.C.1 : Les hydrogènes dans une protéine s'échangent avec des vitesses différentes (pas d'échange : gris, échange rapide : bleu, échange à des vitesses compatibles avec la spectrométrie de masse : rouge).

Néanmoins, la structure tridimensionnelle d'une protéine en solution n'est pas figée. La protéine subit plutôt des fluctuations conformationnelles permanentes. Chaque hydrogène d'amide labile a donc une certaine probabilité de se trouver soit dans une conformation dite « fermée » (non accessible au solvant), soit dans une conformation « ouverte » (accessible au solvant). Puisque l'échange hydrogène/deutérium ne s'effectue que dans la conformation « ouverte », chaque hydrogène d'amide labile a ainsi une certaine probabilité d'échange. L'échange H/D global pour une protéine donnée est ainsi une fonction des probabilités individuelles pour chacun des atomes d'hydrogène. Ces probabilités, de leur côté, dépendent des conditions expérimentales (pH, température, force ionique, etc.), qui ont un effet sur la conformation globale de la protéine (ou plus précisément modifient les différentes populations de l'espace conformationnel de la protéine). Par conséquent, on constate, au niveau des échanges H/D, un changement des proportions d'hydrogènes échangés lorsque la protéine est analysée dans diverses conditions expérimentales. Il est ainsi possible d'obtenir des informations sur la structure ou plus précisément sur des différences structurales entre différentes formes d'une protéine en analysant la cinétique d'échange des hydrogènes de la liaison peptidique, à la fois à l'échelle globale de la protéine entière mais aussi à l'échelle d'un hydrogène d'amide particulier.

Il existe plusieurs techniques pour analyser les échanges H/D. Parmi ces techniques, la RMN est la méthode qui permet d'obtenir les informations les plus précises sur la localisation des échanges H/D puisqu'en théorie chaque hydrogène peut être analysé indépendamment. En pratique, néanmoins, cette technique est limitée à des protéines avec une masse moléculaire inférieure à environ 30 kDa. La FT-IR est une deuxième méthode pour analyser les échanges H/D. Lors de la substitution d'un hydrogène par un deutérium, la fréquence de la vibration N-H change suffisamment pour être observable par spectroscopie infrarouge. Cette technique n'est pas limitée par la masse moléculaire de la protéine (ou du complexe protéique), mais en général il n'est pas possible de localiser précisément la région de la protéine dans laquelle a eu lieu l'échange. Avec le développement des sources d'ionisation dites « douces », la spectrométrie de masse est devenue un outil puissant pour l'analyse des échanges H/D. Initialement, l'ionisation par FAB (« Fast Atom Bombardement ») a été utilisée pour analyser les échanges H/D [2], mais au début des années 1990, l'electrospray est devenu la méthode de choix pour ce type d'expériences. Les premiers exemples se focalisent surtout autour de petites protéines telles l'ubiquitine [3], la myoglobine [4] ou le cytochrome c [5], mais également sur des polypeptides synthétiques, ayant des structures secondaires différentes [6]. Finalement, il a été montré que l'ionisation par MALDI peut également être utilisée pour l'analyse d'échanges H/D [7-9]. Plusieurs revues récentes sont venues souligner l'importance prise par la technique des échanges H/D couplés à la spectrométrie de masse pour l'étude des structures secondaires, tertiaires et quaternaires de protéines de tailles très variables [10-13].

Quels sont les principaux champs d'application de cette technique couplée à la spectrométrie de masse ? De façon générale, elle permet d'obtenir des informations sur la structure tridimensionnelle d'une protéine [14, 15] et sur les changements structuraux intervenant dans différentes conditions expérimentales [16-19]. Cette technique a également été utilisée pour des assemblages non covalents impliquant des protéines, que ce soient pour des interactions protéine-ligand [8, 9, 20-26], ou pour des complexes protéiques [7, 27-32]. En ce qui concerne les interactions protéine-ligand, les exemples comme la protéine fixant le facteur de croissance insuline-like I (IGF-I) (63 kDa) [21], la protéine kinase A (40 kDa) [9] ou la tyrosine kinase de la rate (72 kDa) [25] montrent l'attrait de cette méthode pour les protéines de masses moléculaires élevées. Pour les complexes protéiques, des exemples tels que l'ATPase de la levure (370 kDa) [7] montrent que la technique est moins limitée, en terme de masse moléculaire, que d'autres techniques établies comme la RMN.

Ce dernier point devient indispensable pour l'analyse structurale de macroassemblages de protéines, dont en particulier les agrégats de la protéine prion [33-35] et d'autres agrégats amyloïdes [36-43]. La présence de tels agrégats a été reliée à un certain nombre de maladies dégénératives dont l'issue est fatale [44-46]. Ceci souligne le besoin d'obtenir des informations structurales sur ces assemblages ou sur les formes intermédiaires y conduisant afin de développer des stratégies diagnostiques et thérapeutiques visant ce type de maladies.

Du point de vue technique et méthodologique un premier enjeu à l'heure actuelle est l'amélioration de la résolution spatiale. Des approches pour arriver à cet objectif comprennent la combinaison de plusieurs protéases pour l'étape de digestion enzymatique des protéines [47] ou encore l'utilisation de la fragmentation en phase gazeuse de peptides deutérés [48-53]. Toutefois, l'utilisation de techniques de fragmentation reste un véritable défi puisque les deutériums risquent de se redistribuer sur le peptide lors de l'étape d'activation. Un second enjeu est la réduction de l'échange en retour [54], aux niveaux de l'échantillon, de la chromatographie et du spectromètre de masse. Pour finir, on peut noter un autre développement récent qui est l'utilisation conjointe des échanges H/D et de la spectrométrie de masse en conditions non dénaturantes afin d'étudier la cinétique de dépliement des protéines sur l'échelle de la milliseconde [55-58].

Quelle est la stratégie générale pour l'étude de la structure de protéines par échanges H/D et spectrométrie de masse ? L'approche se décompose en différentes étapes (Fig. III.C.2) Dans une première étape, les différentes formes de la protéine (par exemple monomère et dimère) sont incubées dans une solution qui contient des deutériums échangeables (D_2O , CH_3OD , ...). Dans cette solution, l'échange hydrogène/deutérium s'effectue. Ensuite, la protéine échangée est découpée par digestion enzymatique. La masse des peptides résultants est mesurée par spectrométrie de masse. Pour chaque peptide, le déplacement du massif isotopique donne accès au nombre moyen d'hydrogènes échangés, qu'on peut ramener à un pourcentage d'échange, ce qui permet ainsi d'obtenir une « carte d'accessibilité » des sections de la protéine correspondant aux peptides. La comparaison des échanges dans deux états différents permet de remonter à des informations sur les structures des complexes en question (flèches en pointillés dans Fig. III.C.2).

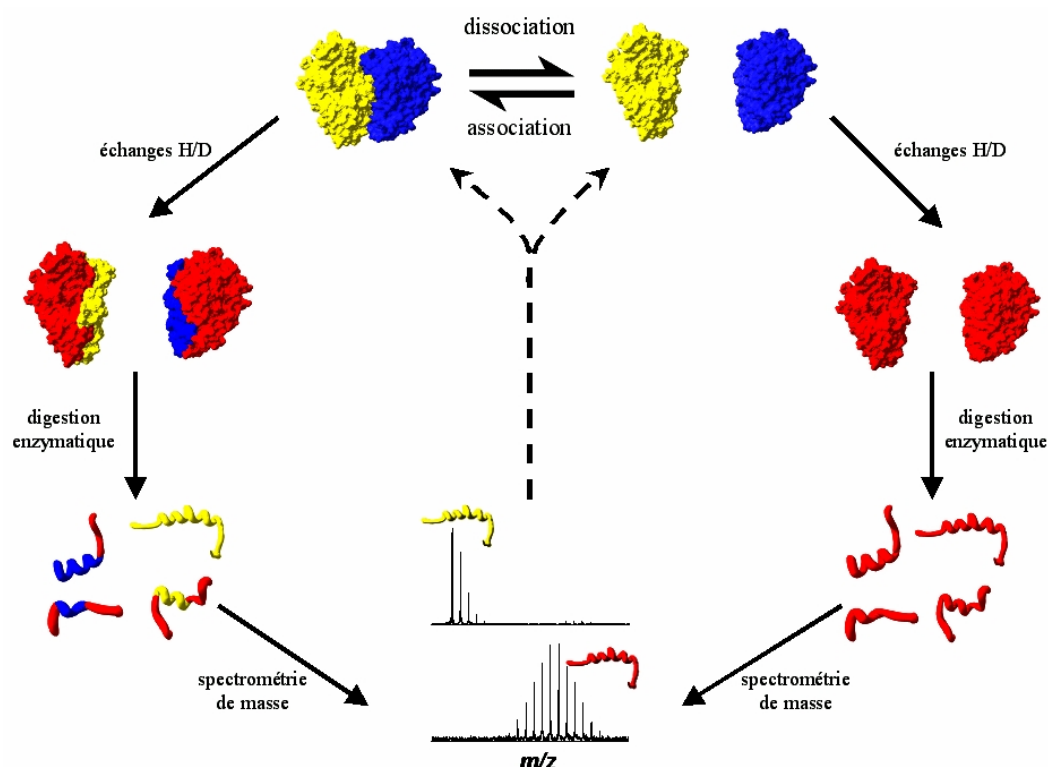


Fig. III.C.2 : Stratégie analytique pour l'analyse des différentes formes d'une protéine par échanges H/D (détails voir texte).

Un écueil important subsiste dans l'application générale de cette technique, celui de l'échange en retour (« *back exchange* ») des deutériums. Après incubation dans une solution qui contient du deutérium, la digestion enzymatique va conduire à l'ouverture de la protéine, risquant de perdre les informations d'accessibilité en échangeant des hydrogènes situés initialement dans des régions peu accessibles au solvant. Il est par conséquent important d'arrêter avant cette étape l'échange H/D. En sachant que les échanges d'hydrogènes d'amides sont fortement réduits à basse température et à un pH autour de pH 2,5, on ajoute en général une solution de H₂O « froide » (0°C) et « acide » (par exemple 4% acide formique).

Dans cette section, la mise au point des conditions expérimentales pour l'analyse des protéines et des complexes protéiques par échange H/D sur notre spectromètre de masse FT-ICR sera présentée. Notre choix a porté sur une analyse par injection directe des peptides obtenus par digestion enzymatique, sans séparation préalable par chromatographie liquide, profitant en cela des qualités de résolution et de précision en masse apportées par un spectromètre de masse FT-ICR. Des travaux antérieurs dans l'équipe ayant suggéré un possible échange en retour des deutériums dans la source nanospray de l'instrument, ce point

a été également vérifié au niveau de cette étude préalable. Finalement, les conditions optimales pour observer les échanges H/D dans nos conditions ont été déterminées.

III.C.2. Résultats

III.C.2.1. Substance P

La première étape de la mise au point des échanges hydrogène/deutérium consistait à vérifier, sur un peptide témoin, que le transfert en phase gazeuse ne perturbe pas l'intégration de deutérium réalisée en solution. La substance P (de séquence RPKPQQFFGLM-NH₂), un petit peptide (M = 1346,73 Da), a été choisie comme système test. La substance P présente 8 hydrogènes d'amides dans des liaisons peptidiques qui sont a priori des hydrogènes échangeables dans ce type d'expérience.

Le spectre de masse après ionisation par electrospray de la substance P présente un pic intense à 674,35 *m/z*, correspondant à la substance P dans son état de charge 2+. Ce pic correspond au plus intense de tous les états de charge (de 1+ à 2+) de la substance P dans les conditions expérimentales utilisées. Le massif isotopique correspondant à la substance P est déplacé vers les *m/z* élevés après 10 minutes d'incubation dans D₂O (Fig. III.C.3, B). Ce résultat indique non seulement que l'échange hydrogène/deutérium s'est effectué en solution, mais aussi qu'au moins une partie des deuteriums incorporés restent sur le peptide lors du transfert en phase gazeuse.

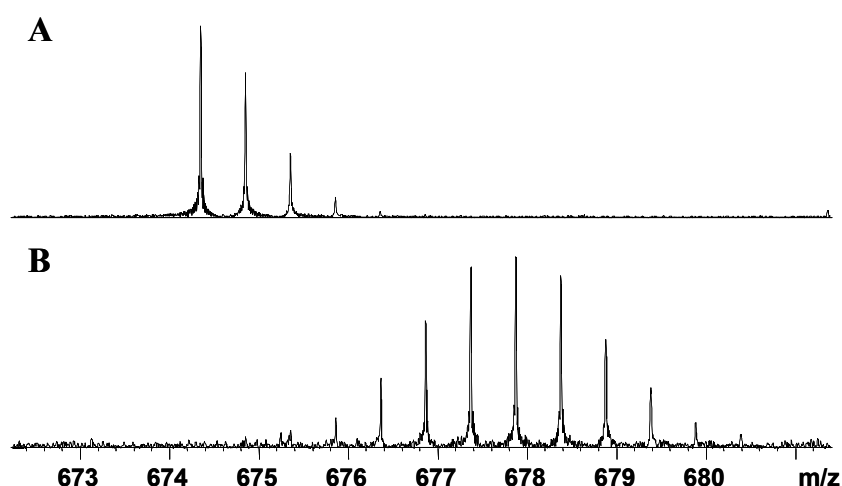


Fig. III.C.3 : Echange hydrogène/deutérium sur la substance P visualisé sur l'état de charge 2+ du peptide. (A) Avant incubation dans D₂O et (B) immédiatement après 10 minutes d'incubation dans D₂O.

Une analyse plus détaillée, en suivant le protocole décrit dans le chapitre expérimental montre qu'il y a en total 5,93 hydrogènes échangés, ce qui correspond à l'échange de 74,1% des hydrogènes échangeables. Les deutériums incorporés ne sont ainsi pas tous ré-échangés avant l'analyse par injection directe.

Pour estimer l'ordre de grandeur du ré-échange lors de ces expériences, un échantillon après incubation et arrêt de l'échange a été laissé dans la seringue d'électrospray et des spectres ont été mesurés à des temps donnés. Les spectres de masse montrent que le pourcentage de deutérium initialement incorporé dans la substance P diminue au fur et à mesure que le temps passe (Fig. III.C.4, A). Après 42 minutes, seuls 23% des hydrogènes échangeables sont encore échangés (à comparer aux 74% échangés initialement). Le pourcentage de deutériums incorporés dans l'échantillon semble converger vers 20%, ce qui correspond à la proportion relative de D₂O par rapport à H₂O (10% D₂O, 40% H₂O, 50 % acétonitrile). Il est important de noter qu'après huit minutes à température ambiante, le pourcentage a déjà diminué de 30%. Il est donc indispensable de minimiser les temps de manipulation à température ambiante d'un échantillon échangé.

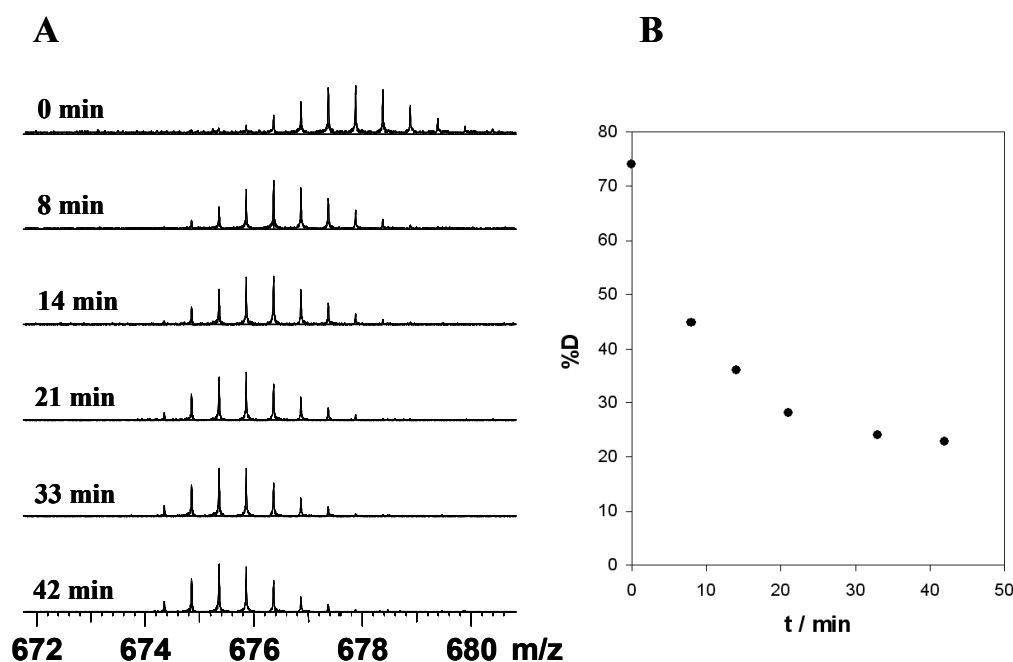


Fig. III.C.4 : Evolution du spectre de la substance P après incubation avec D₂O et ré-échange à température ambiante dans la seringue electrospray à des temps différents. (A) Evolution du spectre (état de charge 2+) et (B) pourcentage de deutériums incorporés en fonction du temps.

Pour mieux suivre les effets de la température pour les expériences d'échanges H/D, une expérience a été réalisée en conservant l'échantillon à 0°C (sur la glace). Des aliquots ont été pris à des temps différents et analysés par electrospray. L'examen des spectres montre qu'il n'y a quasiment pas de changement dans le massif isotopique, indiquant une réduction importante des échanges en retour (Fig. III.C.5, A). Une analyse détaillée des données confirme cette observation : après 43 minutes à 0°C, il n'y a que 10% des deutériums qui ont été ré-changés. Ce résultat confirme la littérature : en diminuant la température, le ré-échange est ralenti, et il est par conséquent raisonnable d'espérer conserver à 0°C un échantillon pendant des durées allant de 20 à 40 minutes si nécessaire.

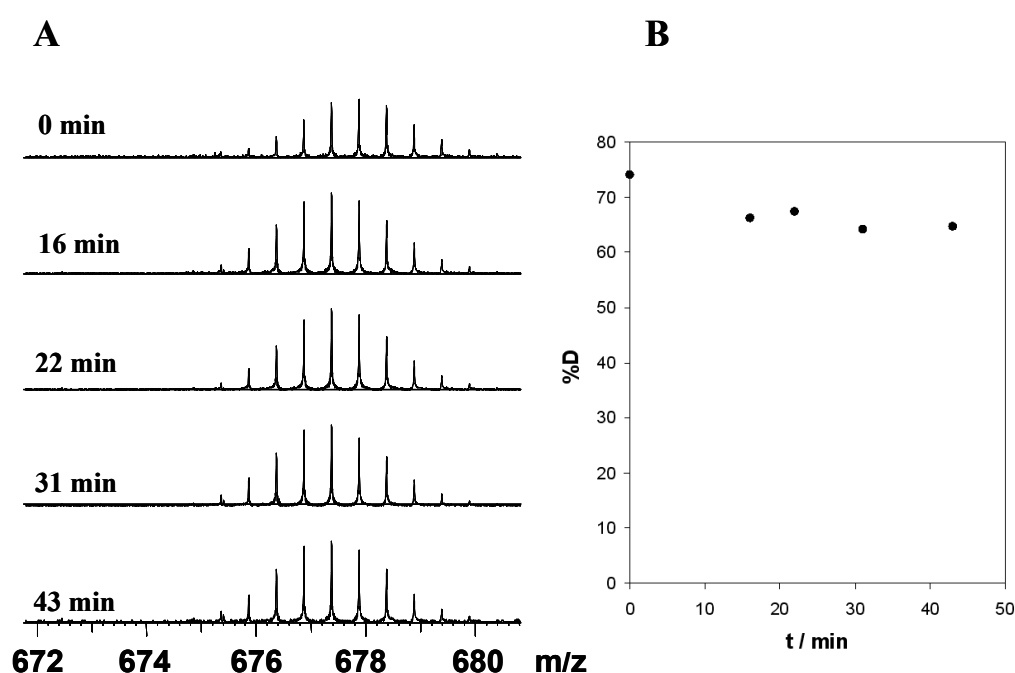


Fig. III.C.5 : Evolution du spectre de la substance P après incubation avec D_2O et ré-échange à 0°C (sur la glace) pendant des temps différents. (A) Evolution du spectre (état de charge 2+) et (B) pourcentage de deutériums incorporés en fonction du temps.

Il est ainsi possible d'étudier les échanges hydrogène/deutérium avec introduction par electrospray et conservation dans la glace. Cependant, pour une analyse des échanges H/D d'échantillons de protéines réelles, la quantité de matériel disponible est en général bien moins importante que dans le cas de la substance P. L'analyse de ces échantillons se fera donc normalement en utilisant la source nanospray. Sur la source ESI de l'APEX III, les deux sources ESI et nano-ESI diffèrent principalement du fait que l'aiguille nanospray est placée directement en face du capillaire de transfert, chauffé à 50°C. De plus la « chambre » dans

laquelle se trouve l'aiguille n'est pas isolée du milieu extérieur, ce qui met le spray en contact avec de l'air plus ou moins chargé d'humidité, alors que dans le cas de l'ESI le spray est en contact avec l'azote utilisé comme gaz nébulisateur. On attend donc un ré-échange plus probable à ce niveau dans le cas du nanospray. La substance P a servi comme système test pour évaluer les possibilités d'utilisation de la source nanospray pour l'étude d'échanges H/D en infusion directe. Pour réaliser cette étude, l'échantillon a été laissé dans l'aiguille nanospray et des spectres de masse ont été enregistrés à des temps différents.

L'examen des spectres montre qu'on observe une intégration de deutériums dans le peptide, comme dans le cas de l'électrospray (Fig. III.C.6, A). Il y a toutefois des différences majeures par rapport à l'électrospray : pour commencer, le nombre moyen d'hydrogènes échangés initialement est plus petit puisque seuls 50% des hydrogènes sont mesurés comme étant échangés. Si l'on regarde l'évolution de l'échange au cours du temps (Fig. III.C.6, B), il semble probable qu'immédiatement après avoir placé l'aiguille en face du capillaire de transfert, il y ait un ré-échange important expliquant ainsi l'écart avec l'électrospray. Après 10 minutes, plus de la moitié des hydrogènes échangés initialement sont ré-échangés ; après 20 minutes, seuls 10% des hydrogènes échangés au départ sont intacts et après 30 minutes, il n'y a quasiment plus de deutériums présents dans le peptide, la totalité ayant été ré-échangée.

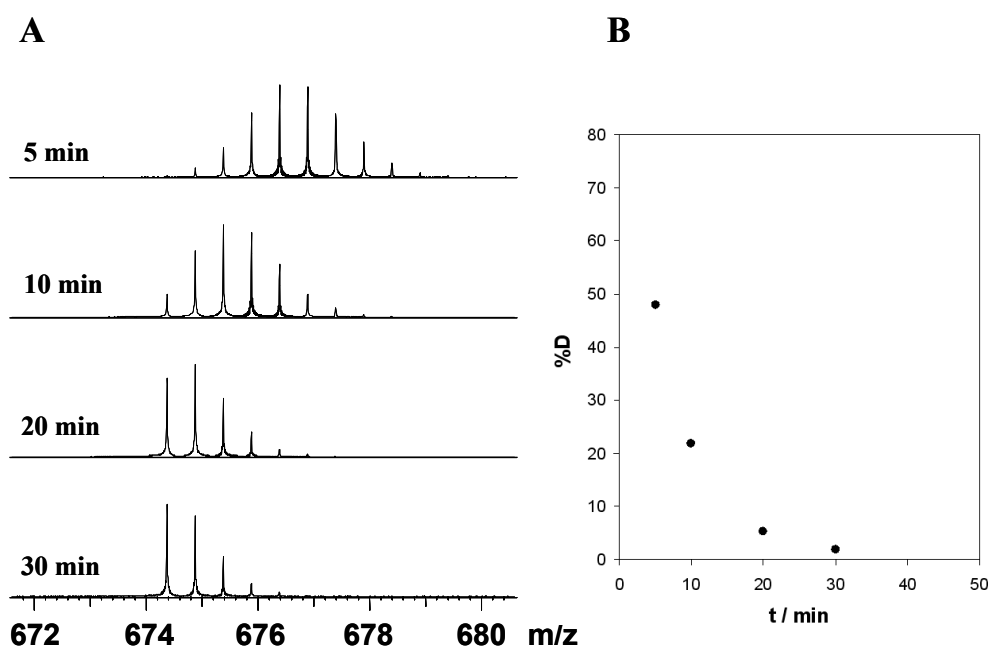


Fig. III.C.6: Evolution du spectre de la substance P après incubation avec D_2O et ré-échange à température ambiante dans l'aiguille nanospray en fonction du temps. (A) Evolution du spectre (état de charge 2+) et (B) pourcentage de deutériums incorporés.

Cette série d'expériences montre qu'en contrôlant attentivement les paramètres expérimentaux (dont en particulier la température) et qu'à condition de minimiser le temps d'analyse, il est possible de déterminer des taux d'échange H/D par nanospray sur notre instrument FT-ICR. Il convient néanmoins de rester prudents, car les difficultés expérimentales sont nombreuses, et la répétition des expériences de multiples fois pourra seule assurer que les paramètres expérimentaux restent reproductibles.

III.C.2.2. Myoglobine

Après avoir évalué l'approche de l'analyse des échanges H/D par injection directe par electrospray et nano-ESI-FT-ICR-MS, le point suivant était de mettre au point les conditions expérimentales pour l'analyse d'une protéine entière par échanges H/D et spectrométrie de masse, un préalable à l'analyse de complexes protéiques. La myoglobine a été choisie comme système test, puisque sa masse moléculaire est suffisamment petite pour assurer une analyse fiable par spectrométrie de masse. Pour une première évaluation de la méthode, le choix d'une protéine de taille plutôt modeste permet de limiter la complexité de l'analyse du spectre liée à un nombre trop important de peptides après la digestion à la pepsine. En revanche, la myoglobine est suffisamment structurée pour que l'on puisse observer des différences dans les cinétiques d'échange H/D contrairement à des peptides non structurés.

L'incorporation de deutérium dans la myoglobine a été mesurée après 1 minute, 30 minutes et 60 minutes d'incubation dans D₂O. Dans le spectre de la myoglobine digérée après échanges H/D, seuls les peptides correspondant aux résidus 1-11, 8-29, 12-29, 14-29 et 138-153 ont été retrouvés, ce qui correspond à une couverture de 29% de la séquence primaire de la protéine. L'abondance des autres peptides était probablement trop faible pour assurer leur détection. Pour ces différents peptides, les spectres de masse montrent clairement l'incorporation de deutériums, ce qui se traduit par une augmentation de la masse moléculaire et un déplacement du massif isotopique correspondant (Fig. III.C.7, A).

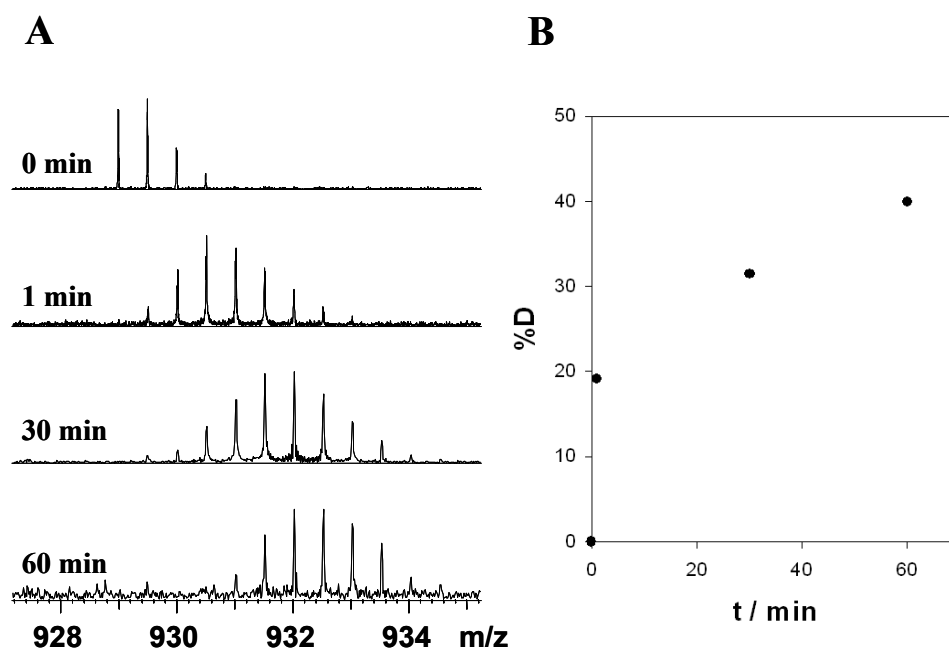


Fig. III.C.7 : Evolution du spectre du peptide 138-153 (état de charge 3+) de la myoglobine après incubation dans D_2O à température ambiante (environ $23^\circ C$) pendant des temps variables. (A) Evolution du spectre et (B) pourcentage de deutériums incorporés en fonction du temps.

La quantification des deutériums incorporés dans le peptide 138-153 correspond à ce qui était apparent dans le spectre : après une heure, 40% des hydrogènes d'amides ont été échangés, ce qui correspond à 6 hydrogènes sur les 15 possibles (Fig. III.C.7, B). Pour les autres peptides, des évolutions similaires sont observées. La quantification montre que les pourcentages d'échange varient entre 23% (peptide 14-29) et 40% (peptide 138-153) après une heure d'incubation (Tab. III.C.1). Par rapport à la substance P, un peptide non structuré pour lequel on attend un échange complet de tous les hydrogènes échangeables, et pour lequel on observait en nanospray un taux d'échange de 50% après 10 minutes d'incubation, le pourcentage d'échange est nettement plus faible. Ce résultat reflète très probablement les éléments structuraux de la myoglobine, qui est une protéine globulaire avec beaucoup d'hélices α . Or, d'après la théorie des échanges H/D pour des protéines [1], les hydrogènes qui sont impliqués dans des liaisons hydrogènes structurant les hélices α s'échangent très lentement, ce qui se traduit dans un faible taux d'échange. Une analyse plus détaillée des données obtenues montre que le peptide 1-11 a échangé 34% de ses hydrogènes d'amides. Le taux d'échange augmente ensuite pour les acides aminés 8-12 (46%) pour atteindre 93% pour les acides aminés 12-14. Les acides aminés 14-29 présentent seulement 23% d'échange. Ces

différences entre des peptides de la partie N-terminale sont surprenantes car tous ces acides aminés sont, d'après la structure obtenue par cristallographie aux rayons X, localisés dans une hélice alpha, et ne devraient donc pas être susceptibles d'échanger autant. Toutefois, il est possible d'expliquer ce résultat par la flexibilité élevée de la partie N-terminale de la myoglobine en solution par rapport à la structure dans le cristal.

Tab. III.C.1 : *Pourcentage de deutériums incorporés en fonction du temps pour les différents peptides de la myoglobine identifiés dans le spectre après échanges hydrogène/deutérium et digestion à la pepsine.*

	0 min	1 min	30 min	60 min
1-11	0	17,4	27,4	33,5
8-29	0	15,4	22,7	33,7
12-29	0	11,8	19,1	30,9
14-29	0	9,0	16,0	22,6
138-153	0	19,2	31,4	39,9

Le fait que les hydrogènes d'amides situés sur les résidus les plus internes (14-29) s'échangent nettement moins rapidement que ceux de la section 8-14 s'expliquerait aussi par cette dynamique de la myoglobine en solution. Les extrémités de la protéine sont susceptibles de présenter une dynamique plus importante que les résidus localisés plus à l'intérieur de la protéine. La séquence d'événements : rupture de la liaison hydrogène – échange H/D – reformation de la liaison hydrogène est ainsi plus probable pour les résidus qui sont localisés aux extrémités. Toutefois sans une exploitation plus complète de la cinétique d'échange et sans une meilleure couverture de la séquence de la protéine, ce résultat ne peut être que qualitatif.

Il est à noter que l'étape de digestion à la pepsine se fait dans un milieu comportant 40% de deutériums. Aucune incorporation de deutérium dans les peptides d'autolyse de la pepsine n'est toutefois observée, puisqu'on a travaillé dans des conditions limitant l'échange H/D au maximum. Par exemple, sur le peptide DRANNKVGLAPVA, on observe une incorporation de moins de 1% de deutérium. Cet élément présente un intérêt analytique : les peptides d'autolyse de la pepsine apparaissent par conséquent comme des peptides non échangés, ce qui permet d'une part de limiter l'ambiguïté sur certaines attributions et également de les utiliser comme étalons internes pour la calibration des spectres de masse.

En résumé, les résultats obtenus avec la myoglobine n'étaient que partiellement concluants : certes ils montraient la faisabilité d'une étude d'échanges H/D sur une protéine de taille modeste avec notre instrumentation et en suivant le protocole que nous avons choisi. La précision en masse de l'instrument apportait suffisamment d'informations pour pouvoir se baser uniquement sur une mesure de masse pour attribuer les peptides à la fois dans les spectres non échangés et dans ceux après échange. Toutefois il restait des éléments d'interrogation : nous n'avons pas pu valider la simple utilisation de la mesure de masse, sans LC-MS, pour l'étude de mélanges potentiellement très complexes du fait de la faible couverture de séquence. Sur la base de ces résultats, il a tout de même été choisi de passer à l'étude d'un système réel, celui d'oligomères du prion ovin qui sera détaillé dans la partie IV. Comme l'optimisation des conditions expérimentales, en particulier celles de la digestion et du désalage des échantillons sont à mettre au point individuellement pour chaque système, cet aspect du travail sera détaillé dans le chapitre correspondant.

III.C.3. Conclusions

En conclusion, l'analyse des échanges hydrogène/deutérium en solution d'un petit peptide (substance P) et d'une petite protéine (myoglobine) analysés par injection directe electrospray et nano-ESI-FTMS montre que cette approche pourra être utilisée pour l'analyse de complexes protéiques. Les deutériums incorporés dans le peptide ou la protéine sont conservés au moins en partie lors des étapes de préparation de l'échantillon ainsi que lors du passage en phase gazeuse aussi bien avec les sources electrospray que nanospray disponibles sur l'instrument. L'exemple de la myoglobine montre de plus que même si les conditions de conservation de l'échange ne sont pas parfaites, il est possible de conserver suffisamment d'informations pour relier les taux d'échanges observés aux structures initiales en solution des protéines. Il sera donc possible d'utiliser cette technique pour obtenir des informations structurales sur des complexes protéiques.

III.C.4. Bibliographie

- [1] A. Hvidt, S. O. Nielsen, Hydrogen exchange in proteins, *Adv. Prot. Chem.* **1966**, *21*, 287-386.
- [2] D. L. Smith, Z. Zhang, Y. Liu, Amide hydrogen exchange and mass spectrometry: a probe of high-order structure in proteins, *Pure & Appl. Chem.* **1994**, *66*, 89-94.

- [3] V. Katta, B. T. Chait, Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, *Rapid Commun. Mass Spectrom.* **1991**, 5, 214-217.
- [4] R. S. Johnson, K. A. Walsh, Mass spectrometric measurement of protein amide hydrogen exchange rates of apo- and holo-myoglobin, *Protein Sci.* **1994**, 3, 2411-2418.
- [5] D. S. Wagner, R. J. Andereg, Conformation of cytochrome c studied by deuterium exchange-electrospray ionization mass spectrometry, *Anal. Chem.* **1994**, 66, 706-711.
- [6] D. S. Wagner, L. G. Melton, Y. Yan, B. W. Erickson, R. J. Andereg, Deuterium exchange of alpha-helices and beta-sheets as monitored by electrospray ionization mass spectrometry, *Protein Sci.* **1994**, 3, 1305-1314.
- [7] A. Nazabal, M. Laguerre, J.-M. Schmitter, J. Vaillier, S. Chaignepain, J. Velours, Hydrogen/deuterium exchange on yeast ATPase supramolecular protein complex analyzed at high sensitivity by MALDI mass spectrometry, *J. Am. Soc. Mass Spectrom.* **2003**, 14, 471-481.
- [8] G. S. Anand, C. A. Hughes, J. M. Jones, S. S. Taylor, E. A. Komives, Amide H/2H exchange reveals communication between the cAMP and catalytic subunit-binding sites in the R'alpha subunit of protein kinase A, *J. Mol. Biol.* **2002**, 323, 377-386.
- [9] M. D. Andersen, J. Shaffer, P. A. Jennings, J. A. Adams, Structural characterization of protein kinase A as a function of nucleotide binding. Hydrogen-deuterium exchange studies using matrix-assisted laser desorption ionization-time of flight mass spectrometry detection, *J. Biol. Chem.* **2001**, 276, 14204-14211.
- [10] J. R. Engen, D. L. Smith, Investigating the higher-order structure of proteins. Hydrogen exchange, proteolytic fragmentation, and mass spectrometry, *Methods Mol. Biol.* **2000**, 146, 95-112.
- [11] C. S. Maier, M. L. Deinzer, Protein conformations, interactions and H/D exchange, *Methods Enzymol.* **2005**, 402, 312-360.
- [12] I. A. Kaltashov, Probing protein dynamics and function under native and mildly denaturing conditions with hydrogen exchange and mass spectrometry, *Int. J. Mass Spectrom.* **2005**, 240, 249-259.
- [13] L. S. Busenlehner, R. N. Armstrong, Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry, *Arch. Biochem. Biophys.* **2005**, 433, 34-46.
- [14] H. Mazon, O. Marcillat, E. Forest, C. Vial, Hydrogen/deuterium exchange studies of native rabbit MM-CK dynamics, *Protein Sci.* **2004**, 13, 476-486.

- [15] H. Mazon, O. Marcillat, E. Forest, C. Vial, Local dynamics measured by hydrogen/deuterium exchange and mass spectrometry of creatine kinase digested by two proteases, *Biochimie* **2005**, 87, 1101-1110.
- [16] S. J. Eyles, I. A. Kaltashov, Methods to study protein dynamics and folding by mass spectrometry, *Methods* **2004**, 34, 88-99.
- [17] H. Mazon, O. Marcillat, E. Forest, C. Vial, Denaturant sensitive regions in creatine kinase identified by hydrogen/deuterium exchange, *Rapid Commun. Mass Spectrom.* **2005**, 19, 1461-1468.
- [18] T. E. Wales, J. R. Engen, Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass Spectrom. Rev.* **2006**, 25, 158-170.
- [19] D. L. Smith, Y. Deng, Z. Zhang, Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry, *J. Mass Spectrom.* **1997**, 32, 135-146.
- [20] F. Wang, W. Li, M. R. Emmett, C. L. Hendrickson, A. G. Marshall, Conformational and dynamic changes of yersinia protein tyrosine phosphatase induced by ligand binding and active site mutation and revealed by H/D exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Biochemistry* **1998**, 37, 15289-15299.
- [21] H. Ehring, Hydrogen exchange/electrospray ionization mass spectrometry studies of structural features of proteins and protein/protein interactions, *Anal. Biochem.* **1999**, 267, 252-259.
- [22] M. M. Zhu, D. L. Rempel, Z. Du, M. L. Gross, Quantification of protein-ligand interactions by mass spectrometry, titration and H/D exchange: PLIMSTEX, *J. Am. Chem. Soc.* **2003**, 125, 5252-5253.
- [23] K. D. Powell, M. C. Fitzgerald, Accuracy and precision of a new H/D exchange- and mass spectrometry-based technique for measuring the thermodynamic properties of protein-peptide complexes, *Biochemistry* **2003**, 42, 4962-4970.
- [24] R. A. Garcia, D. Pantazatos, F. J. Villarreal, Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions, *ASSAY and Drug Development Technologies* **2004**, 2, 81-91.
- [25] M. I. Catalina, M. J. E. Fischer, F. J. Dekker, R. M. J. Liskamp, A. J. R. Heck, Binding of a diphosphorylated-ITAM peptide to spleen tyrosine kinase (Syk) induces distal conformational changes: a hydrogen exchange mass spectrometry study, *J. Am. Soc. Mass Spectrom.* **2005**, 16, 1039-1051.

- [26] M. J. Chalmers, S. A. Busby, B. D. Pascal, Y. He, C. L. Hendrickson, A. G. Marshall, P. R. Griffin, Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry, *Anal. Chem.* **2006**, 78, 1005-1014.
- [27] J. G. Mandell, A. M. Falick, E. A. Komives, Identification of protein-protein interfaces by decreased amide proton solvent accessibility, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 14705-14710.
- [28] J. R. Engen, T. E. Smithgall, W. H. Gmeiner, D. L. Smith, Comparison of SH3 and SH2 domain dynamics when expressed alone or in a SH(3+2) construct: the role of protein dynamics in functional regulation, *J. Mol. Biol.* **1999**, 287, 645-656.
- [29] T. T. Lam, J. K. Lanman, M. R. Emmett, C. L. Hendrickson, A. G. Marshall, P. E. Prevelige, Mapping of protein:protein contact surfaces by hydrogen/deuterium exchange, followed by on-line high-performance liquid chromatography-electrospray ionization Fourier-transform ion-cyclotron-resonance mass analysis, *J. Chromatogr. A* **2002**, 982, 85-95.
- [30] S. Kang, P. E. Prevelige Jr., Domain study of bacteriophage P22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange, *J. Mol. Biol.* **2005**, 347, 935-948.
- [31] E. A. Komives, Protein-protein interaction dynamics by amide H/2H exchange mass spectrometry, *Int. J. Mass Spectrom.* **2005**, 240, 285-290.
- [32] J. Borch, T. J. D. Jorgensen, P. Roepstorff, Mass spectrometric analysis of protein interactions, *Curr. Opin. Chem. Biol.* **2005**, 9, 509-516.
- [33] A. Nazabal, M. Bonneu, S. J. Saupe, J.-M. Schmitter, High-resolution H/D exchange studies on the HET-s218-295 prion protein, *J. Mass Spectrom.* **2005**, 40, 580-590.
- [34] A. Nazabal, M.-L. Maddelein, M. Bonneu, S. J. Saupe, J.-M. Schmitter, Probing the structure of the infectious amyloid form of the prion-forming domain of HET-s using high-resolution hydrogen/deuterium exchange monitored by mass spectrometry, *J. Biol. Chem.* **2005**, 280, 13220-13228.
- [35] A. Nazabal, S. Dos Reis, M. Bonneu, S. J. Saupe, J.-M. Schmitter, Conformational transition occurring upon amyloid aggregation of the HET-s prion protein of *Podospora anserina* analyzed by hydrogen/deuterium exchange and mass spectrometry, *Biochemistry* **2003**, 42, 8852-8861.
- [36] E. J. Nettleton, C. V. Robinson, Probing conformations of amyloidogenic proteins by hydrogen exchange and mass spectrometry, *Methods Enzymol.* **1999**, 309, 633-646.

- [37] P. Tito, E. J. Nettleton, C. V. Robinson, Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: a site-specific investigation by mass spectrometry, *J. Mol. Biol.* **2000**, *303*, 267-278.
- [38] I. Kheterpal, S. Zhou, K. D. Cook, R. Wetzel, A β amyloid fibrils possess a core structure highly resistant to hydrogen exchange, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13597-13601.
- [39] E. J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C. M. Dobson, C. V. Robinson, Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry, *Biophys. J.* **2000**, *79*, 1053-1065.
- [40] S. S.-S. Wang, S. A. Tobler, T. A. Good, E. J. Fernandez, Hydrogen exchange-mass spectrometry analysis of β -amyloid peptide structure, *Biochemistry* **2003**, *42*, 9507-9514.
- [41] I. Kheterpal, R. Wetzel, K. D. Cook, Enhanced correction methods for hydrogen exchange-mass spectrometric studies of amyloid fibrils, *Protein Sci.* **2003**, *12*, 635-643.
- [42] Z. P. Tao, P. Tito, C. V. Robinson, Site-specific hydrogen exchange of proteins: insights into the structure of amyloidogenic intermediates, *Methods Enzymol.* **2005**, *402*, 389-402.
- [43] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gairi, M. Feliz, E. Giralt, C. V. Robinson, C. M. Dobson, Molecular recycling within amyloid fibrils, *Nature* **2005**, *436*, 554-558.
- [44] C. M. Dobson, Protein folding and misfolding, *Nature* **2003**, *426*, 884-890.
- [45] C. M. Dobson, Getting out of shape, *Nature* **2002**, *418*, 729-730.
- [46] C. M. Dobson, In the footsteps of alchemists, *Science* **2004**, *304*, 1259-1262.
- [47] L. Cravello, D. Lascoux, E. Forest, Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins, *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2387-2393.
- [48] Y. Deng, H. Pan, D. L. Smith, Selective isotope labeling demonstrates that hydrogen exchange at individual peptide amide linkages can be determined by collision-induced dissociation mass spectrometry, *J. Am. Chem. Soc.* **1999**, *121*, 1966-1967.
- [49] M.-Y. Kim, C. S. Maier, D. J. Reed, M. L. Deinzer, Site-specific amide hydrogen/deuterium exchange in E. coli thioredoxins measured by electrospray ionization mass spectrometry, *J. Am. Chem. Soc.* **2001**, *123*, 9860-9866.
- [50] J. A. A. Demmers, D. T. S. Rijkers, J. Haverkamp, J. A. Kilian, A. J. R. Heck, Factors affecting gas-phase deuterium scrambling in peptide ions and their implications for protein structure determination, *J. Am. Chem. Soc.* **2002**, *124*, 11191-11198.

- [51] C. Hagman, P. Hakansson, J. Buijs, K. Hakansson, Inter-molecular migration during collisional activation monitored by hydrogen/deuterium exchange FT-ICR tandem mass spectrometry, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 639-646.
- [52] J. K. Hoerner, H. Xiao, A. Dobo, I. A. Kaltashov, Is there hydrogen scrambling in the gas phase? Energetic and structural determinants of proton mobility within protein ions, *J. Am. Chem. Soc.* **2004**, *126*, 7709-7717.
- [53] T. J. D. Jorgensen, H. Gardsvoll, M. Ploug, P. Roepstorff, Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation, *J. Am. Chem. Soc.* **2005**, *127*, 2785-2793.
- [54] M. Kipping, A. Schierhorn, Improving hydrogen/deuterium exchange mass spectrometry by reduction of the back-exchange effect, *J. Mass Spectrom.* **2003**, *38*, 271-276.
- [55] D. J. Wilson, L. Konermann, A capillary mixer with adjustable reaction chamber volume for millisecond time-resolved studies by electrospray mass spectrometry, *Anal. Chem.* **2003**, *75*, 6408-6414.
- [56] D. J. Wilson, L. Konermann, Mechanistic studies on enzymatic reactions by electrospray ionization MS using a capillary mixer with adjustable reaction chamber volume for time-resolved measurements, *Anal. Chem.* **2004**, *76*, 2537-2543.
- [57] J. Pan, D. J. Wilson, L. Konermann, Pulsed hydrogen exchange and electrospray charge-state distribution as complementary probes of protein structure in kinetic experiments: implications for ubiquitin folding, *Biochemistry* **2005**, *44*, 8627-8633.
- [58] B. M. Hossain, L. Konermann, Pulsed hydrogen/deuterium exchange MS/MS for studying the relationship between noncovalent protein complexes in solution and in the gas phase after electrospray ionization, *Anal. Chem.* **2006**, *78*, 1613-1619.

Partie IV

Oligomères de la protéine du prion ovin (PrP)

Dans cette troisième partie, nous présenterons l'application des méthodes décrites dans la partie précédente aux oligomères formés par la protéine du prion ovin. Cette protéine forme *in vitro* dans certaines conditions des oligomères dont la structure n'est pas connue. La petite taille des sous-unités monomériques, les données cristallographiques existant sur la forme monomérique et l'intérêt biologique évident qu'il y aurait à élucider ces structures oligomériques faisaient de cette protéine un bon candidat pour une première application des méthodes développées dans les parties précédentes.

IV.1. Introduction

Les encéphalopathies subaiguës spongiformes transmissibles (ESST), ou maladies à prion sont des maladies neurodégénératives du système nerveux central à issue fatale systématique. Elles se manifestent chez l'Homme sous la forme des maladies de Creutzfeldt Jacob (CJD), syndrome de Gerstmann-Sträussler-Scheinker et Insomnie Fatale. Les ESST se rencontrent également chez de nombreuses autres espèces de mammifères, comme la vache, le mouton, le vison, et différents cervidés [1]. Le développement de la maladie semble reposer sur la propagation au sein des organismes atteints d'un agent pathogène transmissible. L'agent infectieux, isolé à partir des tissus infectés, ne peut pas être totalement inactivé par les procédures de destruction des acides nucléiques utilisées classiquement pour la stérilisation en microbiologie et en virologie, ce qui a conduit à l'hypothèse que l'agent infectieux n'était pas un virus ou une bactérie. Alors que la digestion par la nucléase micrococcalle ou la RNase I n'inactive pas l'agent de la tremblante (la forme de l'ESST chez le mouton), la digestion par la protéinase K (PK) et la trypsine ou l'exposition à de hautes concentrations de guanidine et sodium dodécylsulfate (agents dénaturants des protéines) fait disparaître le pouvoir infectieux des fractions contenant l'agent semi-purifié [2]. Dans les années 1980, l'équipe de Stanley Prusiner avait identifié une protéine hydrophobe dans des échantillons de l'agent infectieux de la tremblante [3, 4] et a donc suspecté cette protéine d'être au moins en partie responsable du pouvoir infectieux de ces échantillons.

Après digestion ménagée de l'agent purifié par la protéinase K, des fragments protéiques migrant entre 27 et 30 kDa sont détectés sur gel SDS PAGE [5]. Prusiner a proposé alors que l'agent infectieux, ou *prion* (**proteinaceous and infectious**) [2] serait constitué d'une protéine baptisée PrP (**P**rion **P**rotein) possédant une partie résistante à la digestion par la PK, nommée PrP²⁷⁻³⁰. Sous certaines conditions, la PrP²⁷⁻³⁰ forme des structures semblables aux fibres amyloïdes observées dans le cas d'autres maladies neurodégénératives comme la maladie d'Alzheimer [6]. Par d'autres expériences, il a été montré qu'un polypeptide ayant la même séquence que la protéine PrP est présent chez tous les mammifères [7-9]. Il s'agit d'une glycoprotéine membranaire de 23 kDa environ, extrinsèque et extracellulaire appelée PrP^C (C pour « cellulaire »), qui est exprimée de manière ubiquitaire dans les conditions normales. La protéine PrP qui co-purifie avec l'agent des ESST existe donc chez l'hôte sous sa forme PrP^C. Dans les conditions pathologiques, cette protéine formerait des agrégats dont la taille et la structure ne sont pas encore bien connues, et qui sont appelés PrP^{Sc} (Sc pour « Scrapie », l'acronyme anglais pour la tremblante). PrP²⁷⁻³⁰ serait donc le produit de clivage protéolytique de ces agrégats de grande taille, qui posséderaient un pouvoir infectieux. Alors

qu'aucune modification covalente ne différencie PrP^C et PrP^{Sc} [10], ces deux isoformes semblent adopter des arrangements tridimensionnels différents : la forme PrP^{Sc} est riche en feuillets β , insoluble, résistante aux protéases, alors que la forme PrP^C est riche en hélice α [11], soluble et sensible aux protéases. Prusiner a proposé un modèle selon lequel la PrP^C se transformerait en PrP^{Sc} (Fig. IV.1) [2]. Il s'agit essentiellement d'une transformation de la structure tridimensionnelle de la protéine PrP^C en présence de la PrP^{Sc} :

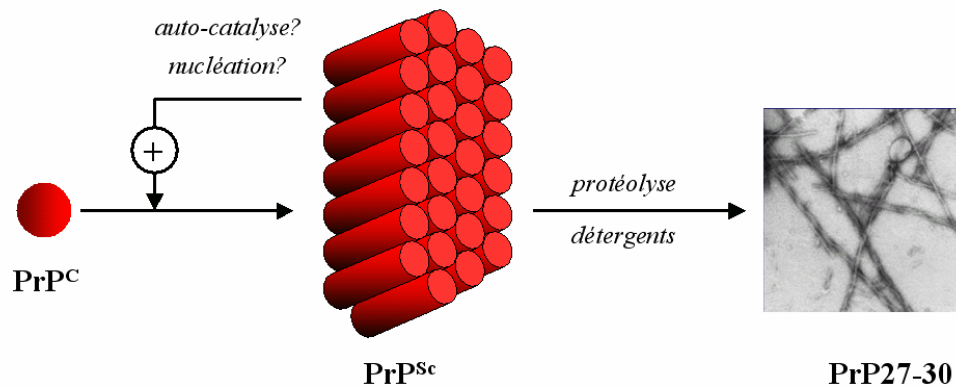


Fig. IV.1 : Schéma de la formation de PrP^{Sc} et des fibrilles amyloïdes à partir de PrP^C.

La protéine PrP possède ainsi des caractéristiques semblables à celles d'autres protéines associées à certaines neuropathologies comme la maladie d'Alzheimer ou la Chorée de Huntington, avec lesquelles les maladies à prions partagent plusieurs points communs :

- Le fragment infectieux de PrP^{Sc} généré par digestion à la protéinase K (PrP27-30) forme des fibres amyloïdes lorsqu'on l'extrait par certains détergents [12]. Des fibres amyloïdes comparables, constituées de la protéine A β 1-42, sont observées chez les patients atteints de la maladie d'Alzheimer [13].
- De larges dépôts protéiques essentiellement constitués de PrP sont souvent observés dans les cerveaux d'animaux atteints d'ESST. De manière similaire, des cerveaux touchés par la maladie d'Alzheimer présentent fréquemment des plaques protéiques riches en A β 1-42 [14].
- La pathologie cérébrale observée dans les maladies amyloïdes du système nerveux central (SNC) et les maladies à prions ont pour caractéristique une abondante perte neuronale [14].

Quelles sont les caractéristiques de l'agent infectieux ? Sous certaines conditions de purification et en présence de certains détergents, la protéine PrP²⁷⁻³⁰ forme des fibres amyloïdes [6] : lorsqu'on les observe en microscopie électronique, ces fibres présentent une morphologie comparable à celle des fibres amyloïdes observées dans la maladie d'Alzheimer (les dimensions des fibres de PrP²⁷⁻³⁰ sont 10-20 nm de diamètre et 100-200 nm de longueur). Les études de la PrP^{Sc} par spectroscopie infrarouge à transformée de Fourier (FTIR) [11, 15] révèlent une autre caractéristique amyloïde de cette protéine : la protéine PrP^{Sc} contient un taux de feuillet β important (30-50% de feuillet β , 20% d'hélice α) alors que la protéine PrP^C est majoritairement structurée en hélice α (40% d'hélice α , 3% de feuillet β). Les fibres de PrP^{Sc} purifiées présentent une diffraction des rayons X comparable à celle d'autres fibres amyloïdes caractérisées par cette même méthode [16] : cette diffraction caractéristique, et en particulier l'existence de raies de diffraction de 4,5 Å de résolution, suggère une architecture basée sur l'empilement des feuillets β , le plan de chaque feuillet étant orienté perpendiculairement à l'axe de la fibre. L'obtention récente de cristaux à deux dimensions a permis l'étude structurale de PrP^{Sc} par diffraction des électrons [17]. Cette étude révèle l'existence d'une symétrie d'ordre 3 dans la structure des cristaux de l'agent infectieux. Les résultats de cette étude ont été utilisés comme contrainte pour l'établissement d'un modèle structural de PrP^{Sc} (Fig. IV.2) [18].

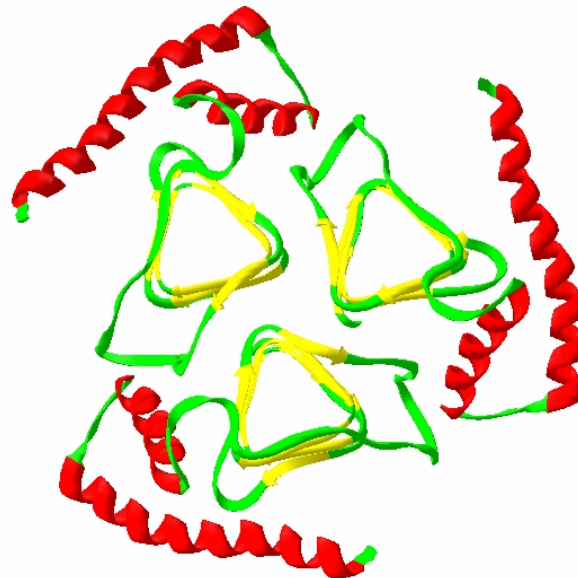


Fig. IV.2 : Modèle structural de la PrP^{Sc}, d'après [18]. L'unité asymétrique de PrP²⁷⁻³⁰ est ici un trimère. Les résidus 90-174 de chaque monomère de PrP sont organisés en une hélice β (au milieu).

Dans ce modèle, les résidus 90 à 174 de la partie correspondant à la PrP²⁷⁻³⁰ sont repliés en une hélice β , constituant l'interface entre les trois monomères d'un trimère, et les résidus 175 à 231 adoptent une structure identique à celle observée dans la PrP recombinante. Un modèle d'empilement en fibres amyloïdes de ce trimère est également proposé, mais celui-ci ne permet pas de prédire la diffraction expérimentale.

On a longtemps considéré que les agrégats amyloïdes formés par les protéines A β 1-42, Huntingtine et autres protéines responsables de maladies neurodégénératives pouvaient être intrinsèquement neurotoxiques. Ainsi, les identités flagrantes entre l'ESST et les amyloïdoses du SNC ont été à l'origine de l'hypothèse selon laquelle la PrP^{Sc} serait neurotoxique, en plus d'être infectieuse. Cependant, il est de plus en plus admis que les larges agrégats amyloïdes observés dans la maladie d'Alzheimer ou de Huntington n'ont aucune activité biologique : la formation de gros assemblages protéiques chez les organismes atteints de maladies amyloïdes ne seraient donc pas la cause de la pathologie elle-même. Sans remettre en cause l'implication des protéines amyloïdes dans les différentes pathologies auxquelles on les associe, cela pourrait signifier qu'un ou plusieurs intermédiaires amyloïdogènes seraient à l'origine de ces maladies [19]. Cette hypothèse est relayée par plusieurs expériences réalisées en culture cellulaire montrant que des oligomères de Transthyretine provoquent la mort cellulaire à des concentrations relativement faibles, alors que les fibres amyloïdes formées par les mêmes protéines sont dépourvues de tout effet cytopathogène dans les mêmes systèmes [20]. Dans le cas du prion, il est également concevable que les fibres amyloïdes formées *in vitro* ne représentent pas les espèces toxiques ou infectieuses formées au cours des ESST. Celles-ci sont peut-être à rechercher parmi les intermédiaires formés lors de l'amyloïdogenèse de PrP *in vitro*. Des états oligomériques de faible poids moléculaire ont été observés durant la purification de PrP^{Sc} *in vivo* [21, 22], mais bien qu'elles ne soient présentes que dans les tissus infectés, on ignore si ces formes ont un effet biologique. D'autres formes oligomériques peuvent être produites *in vitro* à partir de protéine PrP recombinante [23] sur lesquelles nous reviendrons plus loin. Il existe à l'heure actuelle peu de preuves à l'appui d'un possible rôle d'oligomères de PrP dans la pathologie cérébrale des ESST. De plus, on ne peut exclure une dérégulation de la fonction de PrP *in vivo* comme cause de la mort neuronale [24].

Quelles sont les informations expérimentales dont on dispose sur le changement conformationnel de la protéine PrP ? La première étude du repliement de la protéine PrP recombinante (PrP^{rec}) a été menée par l'équipe de Glockshuber sur la construction PrP(121-231) de souris correspondant au domaine globulaire de la protéine PrP entière [25]. Les

résultats obtenus montraient alors que la partie structurée de PrP peut se replier de façon autonome selon un mécanisme à deux états à pH neutre et basse force ionique. Dans ces conditions, la cinétique de repliement de la partie globulaire de PrP est très rapide ($<1 \mu\text{s}$) [26]. Plusieurs équipes montreront par la suite que le dépliement-repliement de la protéine PrP est irréversible dans certaines conditions. A haute force ionique, la dénaturation chimique de la protéine PrP(90-231) humaine est multi-étapes : dans ces conditions et à pH acide, la PrPrec se convertit en oligomères alors qu'à pH neutre, des fibres amyloïdes sont formées à long terme [27]. Des fibres amyloïdes de morphologie distincte de celles obtenues avec PrP(90-231) sont formées par la protéine PrP entière suivant le même protocole [28]. Une large gamme de conditions peut, en fait, mener à la formation de fibres amyloïdes par la protéine PrP [29-35] ; ces fibres sont plus résistantes que la forme monomérique de PrPrec à la digestion par la protéinase K et ont les mêmes propriétés de fixation de certains ligands (Rouge Congo, anticorps) que la PrP^{Sc}. Les formes amyloïdes de PrP recombinante ont donc été jugées structurellement proches des fibres amyloïdes formées par la PrP27-30. Ces critères sont cependant insuffisants pour affirmer l'identité des deux formes de fibres amyloïdes d'autant plus qu'une seule des conditions utilisées (incubation de la protéine recombinante à 37°C dans 3M urée, 0,2 M NaCl, 50 mM acétate de sodium pH 5.0) pour la conversion structurale de PrPrec a permis l'obtention d'une espèce infectieuse [36].

L'observation d'oligomères des protéines PrP murines et bovines *in vitro* a été rapportée dans la littérature [23, 27, 37-39]. Ces oligomères peuvent ou non être retrouvés sur la voie de l'amyloïdogenèse de PrP dans différents organismes *in vivo*, et les conditions de leur formation sont variables (dénaturation thermique, chimique ou à hautes pressions). Dans tous les cas, la distribution de masse des oligomères est très large, la stœchiométrie d'assemblage des plus petites espèces formées variant entre 8 et 15 sous-unités de PrP. Aucune de ces études ne relate un effet biologique de ces oligomères. Quant à leur type d'assemblage, trois questions importantes se posent:

- Quelle est la masse moléculaire des espèces oligomériques ?
- Quels sont les changements au niveau structural intervenant lors de la formation des oligomères ?
- Quels sont les sites d'interaction entre les monomères dans les systèmes oligomériques ?

Pour répondre à ces trois questions, nous avons choisi une approche par spectrométrie de masse. Les expériences ont été effectuées avec la variante ARQ de la PrP ovine. Ce nom fait référence aux acides aminés des positions 136 (A), 154 (R) et 171 (Q), positions qui présentent un polymorphisme chez le mouton.

Tout d’abord, les conditions expérimentales ont été mises au point pour purifier les différents oligomères utilisés lors de nos expériences. Les études structurales ont été effectuées avec ces oligomères purifiés.

Nous avons mis au point des conditions expérimentales permettant l’observation des espèces oligomériques les plus stables en phase gazeuse dans le but de déterminer leur masse, et donc leur état d’oligomérisation.

Pour répondre aux questions concernant les conversions structurales et le site d’interaction, une approche par échanges hydrogène/deutérium en solution, suivi par une digestion enzymatique et l’analyse des fragments par spectrométrie de masse FT-ICR a été choisie.

IV.2. Généralités sur les oligomères de la PrP

Comme nous l’avons vu précédemment, la conversion de la PrP^C en PrP^{Sc} correspond à une modification de la structure tridimensionnelle de la protéine. Or un changement conformationnel d’une structure A (par exemple PrP^C) en une structure B (par exemple PrP^{Sc}) implique au moins une séquence dépliement – repliement, passant par une espèce intermédiaire (PrP^{*}) :



Ces deux étapes permettent la rupture d’une liaison (dépliement) ainsi que la formation d’une nouvelle liaison (repliement) conduisant à la nouvelle structure. Dans le cas de la protéine PrP, l’enjeu global est l’étude de tous les chemins de dépliement/repliement qui conduisent à une structure autre que la structure native de la protéine PrP^C.

Après la dénaturation thermique de la PrP ovine à pH acide (55°C à pH 3,38), on constate la formation d’espèces oligomériques ayant des tailles définies qui apparaissent comme des pics résolus lors de l’analyse par chromatographie d’exclusion stérique [23] (Fig. IV.3). Selon l’ordre d’apparition dans le chromatogramme, les oligomères ont été appelés P1, P2, etc. (Fig. IV.3). Il a été possible de purifier les pics chromatographiques P1, P2 et P3. Toutefois, seuls

les oligomères P1 et P3 semblent être suffisamment stables pour permettre des études ultérieures, l'oligomère P2 se dissociant rapidement. En effet si l'on analyse de nouveau cet oligomère par chromatographie immédiatement après purification à 15°C on observe une quantité non négligeable de monomère dans la solution.

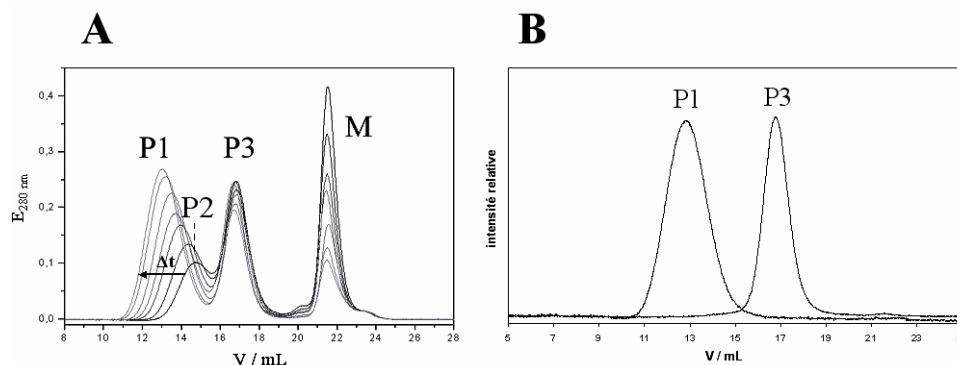


Fig. IV.3 : Analyse des oligomères de la PrP par chromatographie d'exclusion stérique. (A) Mélange des oligomères après différents temps d'incubation (M = monomère). (B) Nouvelle analyse des deux oligomères principaux P1 et P3 après purification (superposition des deux chromatogrammes).

Les différents oligomères sont formés selon le schéma cinétique présenté dans la figure IV.4 (publication en préparation). Selon ce schéma, les trois oligomères principaux sont formés indépendamment à partir de la PrP^C (forme native de la PrP) via des structures intermédiaires différentes.

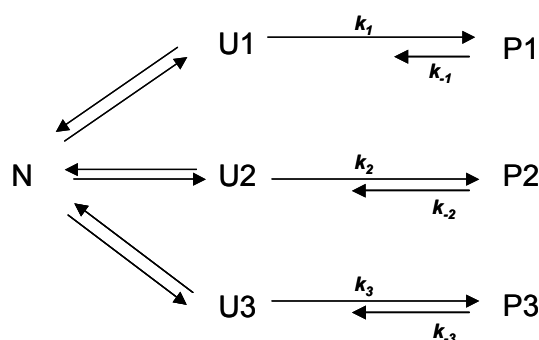


Fig. IV.4 : Schéma cinétique de la formation des oligomères P1, P2 et P3 à partir de la forme native de la PrP (N) via des voies parallèles avec des intermédiaires spécifiques (U1, U2 et U3) pour chaque oligomère.

Pour étudier la nature des oligomères plus en détail, nous nous sommes intéressés aux structures des différents oligomères identifiés. Afin d'obtenir des informations sur leur stoechiométrie, nous souhaitons tout d'abord obtenir la masse moléculaire de ces assemblages protéiques. Ensuite, les structures secondaires et tertiaires du monomère et des oligomères P1 et P3 ont été étudiées par échanges H/D et spectrométrie de masse pour obtenir des informations concernant les changements structuraux se produisant lors de l'oligomérisation. Les mêmes études sont susceptibles de fournir des informations sur la structure quaternaire des complexes, voire sur les sites d'interaction entre les sous-unités des complexes protéiques.

IV.3. Etude des oligomères intacts par spectrométrie de masse

IV.3.1. Introduction

La détermination de la masse moléculaire des oligomères intacts nécessite le maintien des interactions non covalentes entre les sous-unités lors de l'analyse. Les masses moléculaires des oligomères ont été estimées par Rezaei et al. [23] sur la base d'expériences de SAXS (« small angle X-ray scattering ») à 275 ± 28 kDa pour P3 et 730 ± 73 kDa pour P1 ce qui correspondrait à un 12-mer pour P3 et à un 36-mer pour P1. P1 pourrait être un trimère (2,7-mer) de P3 [23]. Cependant, il est important de noter que l'erreur expérimentale, qui est estimée être de l'ordre de 15%, d'après la masse volumique de l'assemblage, est relativement élevée ce qui ne permet pas de donner une stoechiométrie précise.

La détermination de la masse moléculaire des oligomères par spectrométrie de masse aurait l'avantage d'une plus grande précision sur la mesure de masse. Une masse moléculaire plus précise sera importante pour mieux répondre à la question de la topologie des oligomères ainsi que pour préciser le schéma cinétique de l'oligomérisation. Néanmoins, la mesure par spectrométrie de masse impose plusieurs contraintes. Tout d'abord, les oligomères doivent être stables dans un tampon volatil qui sera compatible avec la spectrométrie de masse. Ensuite, les oligomères doivent être suffisamment stables en phase gazeuse pour ne pas subir de fragmentations. La concentration des oligomères en solution doit être suffisante pour permettre la détection par spectrométrie de masse, mais ne doit pas être trop élevée afin de ne pas conduire à la formation de complexes non spécifiques.

Il a donc tout d'abord été nécessaire de trouver des conditions expérimentales qui soient compatibles avec la spectrométrie de masse et dans lesquelles les oligomères restent stables en solution avec une concentration suffisante pour la mesure de masse (voir partie VI).

Après avoir réalisé des tests dans différents tampons volatils, nous avons montré que le TEAA (triéthylammoniumacétate) était le seul tampon compatible avec la spectrométrie de masse dans lequel les oligomères restaient relativement stables. Malgré tout, seul l'oligomère P3 s'est révélé suffisamment stable dans un tampon 50 mM TEAA pour permettre l'étude par spectrométrie de masse. L'oligomère P1 forme des agrégats aux concentrations nécessaires pour l'analyse tandis que l'oligomère P2 se dissocie pour former du monomère. P3 a donc été choisi pour la détermination de la masse moléculaire des oligomères par spectrométrie de masse.

IV.3.2. FT-ICR MS

Le spectre de masse FT-ICR de l'oligomère P3 dans 50 mM TEAA pH 3,38 présente un grand nombre de pics et ne correspond pas à une espèce homogène (Fig. IV.5). Le calcul des masses moléculaires à partir du spectre montre la présence simultanée d'un certain nombre de multimères dont au moins les formes comportant 5, 6, 7, 8, 9 et 10 sous-unités (Tab. IV.1). La présence d'éléments comportant de 1 à 4 sous-unités est possible mais ne peut être prouvée sur la base de ce seul spectre, car ils sont masqués par la présence d'oligomères qui comportent un multiple du nombre de sous-unités. La différence entre les masses moléculaires expérimentales et théoriques correspond à une différence d'environ 17-18 Da par monomère, mais on doit noter que la précision de l'instrument dans cette gamme de m/z est faible.

Tab. IV.1 : Masses moléculaires théoriques et expérimentales des formes multimériques de la PrP observées dans le spectre de masse FT-ICR de l'oligomère P3. Les multimères indiqués avec une étoile (*) ne sont pas identifiés de façon certaine.

taille de l'oligomère	M _{theo} / Da	M _{exp} / Da
1*	22915	22933
2*	45830	45860
3*	68745	68798
4*	91660	91717
5	114575	114648
6	137490	137599
7	160405	160529
8	183320	183481
9	206235	206388
10	229150	229288

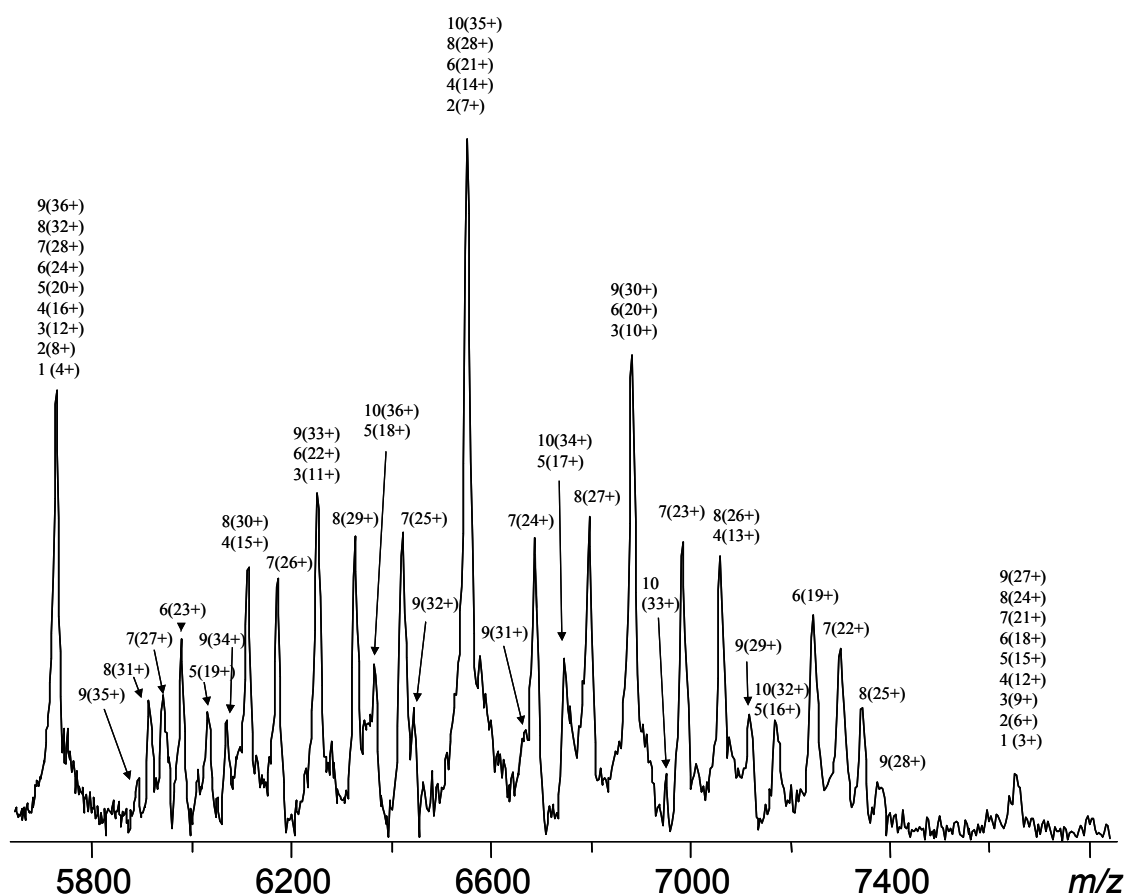


Fig. IV.5 : Spectre de masse de l'oligomère P3 dans 50 mM TEAA pH 3,38 obtenu par ESI-FT ICR MS.

Pour déterminer si la présence de ces nombreuses formes multimériques était due à une décomposition d'oligomères de plus grande taille qui aurait lieu à cause de la configuration particulière de la source electrospray, nous avons décidé de réaliser les mêmes expériences sur un autre spectromètre de masse.

IV.3.3. Q-TOF

Nous avons analysé l'oligomère P3 sur un instrument Q-TOF modifié pour permettre l'optimisation des paramètres de source (notamment la pression dans l'interface) pour la détection de complexes protéiques. Ces expériences ont été effectuées en collaboration avec le laboratoire du Prof. Albert Heck à l'université d'Utrecht (Pays-Bas).

Le spectre de l'oligomère P3 obtenu sur le Q-TOF présente, tout comme celui obtenu sur le FT-ICR, une hétérogénéité importante. (Fig. IV.6). On y trouve un grand nombre de pics correspondant à différentes formes multimériques de la protéine ($n = 6-11$) (Tab. IV.2).). Comme dans le cas du spectre obtenu sur le FT-ICR, la présence d'éléments comportant de 1

à 5 sous-unités est possible mais ne peut être prouvée sur la base de ce seul spectre, car ils sont masqués par la présence d'oligomères qui comportent un multiple du nombre de sous-unités. Le spectre présente ici la région de m/z 7500 à 11000 où se trouvent les ions les plus intenses. Dans la région de m/z inférieur à 7500, on observe des pics de plus faible intensité attribuables en grande partie au monomère.

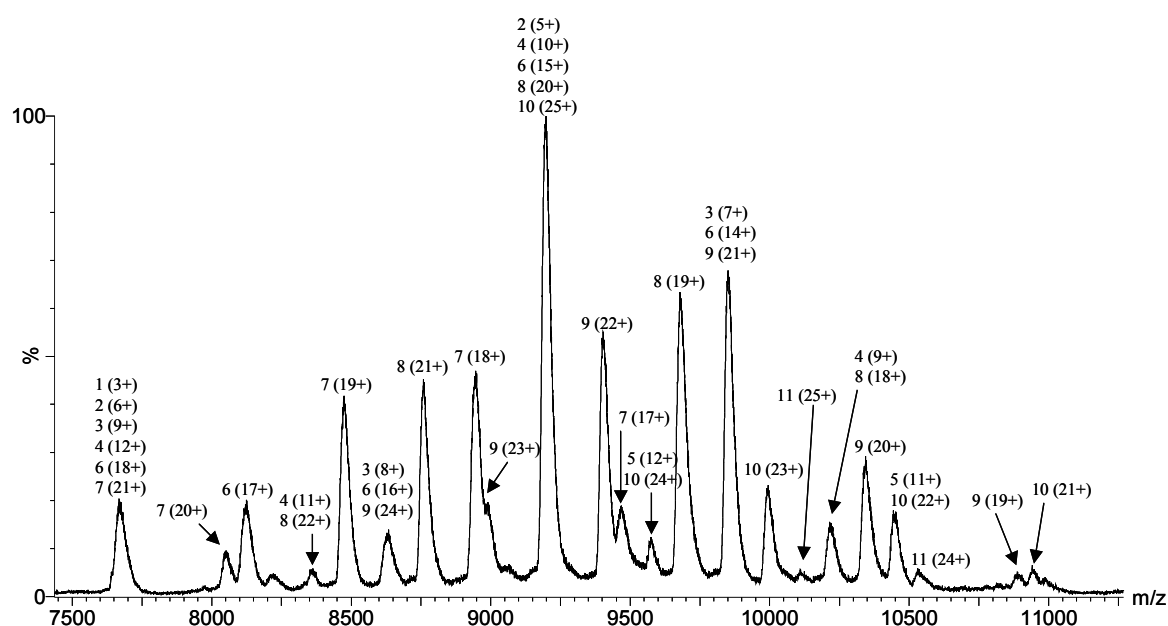


Fig. IV.6 : Détail de la région m/z 7500 à 11000 du spectre ESI-Q-TOF de l'oligomère P3 dans 50 mM TEAA pH 3,38 à un potentiel d'accélération de 200 V.

Tab. IV.2 : Masses moléculaires théoriques et expérimentales des oligomères de la PrP observés par spectrométrie de masse Q-TOF. Les multimères indiqués avec une étoile (*) ne sont pas identifiés de façon certaine.

taille de l'oligomère	$M_{\text{theo}} / \text{Da}$	$M_{\text{exp}} / \text{Da}$
1*	22915	23001
2*	45830	45994
3*	68745	69006
4*	91660	91973
5*	114575	114909
6	137490	138080
7	160405	160992
8	183320	183829
9	206235	206898
10	229150	229815
11	252065	252708

Bien que les résultats soient globalement comparables à ceux obtenus par spectrométrie de masse FT-ICR, la meilleure sensibilité du Q-TOF a permis d'observer des espèces minoritaires (11-mer notamment) qui n'étaient pas observables à l'aide du FT-ICR. De plus, cela nous a permis d'utiliser des concentrations plus faibles d'échantillon, ce qui limite le risque d'agrégation non spécifique. Toutefois, toutes les formes multimériques observées en FT-ICR sont également observées sur le Q-TOF.

Comment expliquer ce résultat ? Il est possible que les oligomères initialement présents en solution se fragmentent lors du passage en phase gazeuse. Dans ce cas, l'hétérogénéité observée dans le spectre ne serait pas un reflet de la solution. Il serait possible que les interactions non covalentes qui stabilisent des oligomères soient au moins en partie des interactions de nature hydrophobes, ce qui expliquerait l'instabilité des oligomères en phase gazeuse. En revanche, la présence de multimères en phase gazeuse ne peut s'expliquer que par la présence d'autres types d'interactions non covalentes (par exemple des interactions électrostatiques ou dipôlares). La co-existence au sein des oligomères de différents types d'interactions non covalentes pourrait au moins en partie expliquer les résultats obtenus. Une autre possibilité serait la rupture des interactions non covalentes entre des monomères à cause de la présence d'un trop grand nombre de charges sur les oligomères ce qui conduirait à une répulsion électrostatique importante entre les sous-unités. Dans ce cas, il serait aussi possible d'envisager une réorganisation des oligomères initialement présents en solution lors du passage en phase gazeuse, conduisant à la formation d'oligomères plus stables en phase gazeuse.

Une autre explication dans ce contexte tient compte du fait que les oligomères observés (avec le FT-ICR ou avec le Q-TOF) présentent tous une masse moléculaire plus élevée que la masse théorique (Tab. IV.1, Tab.IV.2), et que la masse des adduits dépend du rapport m/z et du potentiel d'accélération des ions dans l'interface. Ceci s'expliquerait par la présence d'adduits sur les oligomères. Il est possible que l'oligomère (ou les oligomères) constituant P3 nécessite la présence de ces adduits pour sa stabilité. Une hétérogénéité des adduits conduirait à une dilution des pics correspondants dans le spectre de masse. Il est donc nécessaire d'augmenter l'énergie interne des ions pour dissocier au moins une partie des adduits présents sur les oligomères. Or, l'énergie pour faire disparaître des adduits pourrait être supérieure à celle nécessaire pour fragmenter l'oligomère (ou les oligomères) de départ, ce qui conduirait à la présence d'une telle hétérogénéité des espèces dans le spectre.

Cependant, il est aussi possible que le mélange d'oligomères ait été préexistant en solution. Toutefois, une telle hétérogénéité en solution aurait eu tendance à élargir les pics lors de la chromatographie d'exclusion stérique, ce qui n'était pas le cas.

Pour répondre plus précisément à la question de l'origine de l'hétérogénéité des oligomères, nous envisageons de fractionner le pic correspondant à l'oligomère P3 et d'analyser les différentes fractions. Une différence entre ces fractions indiquerait une hétérogénéité de notre solution d'oligomère P3.

IV.3.4. Conclusions sur les expériences de spectrométrie de masse en conditions non dénaturantes

Du fait de sa plus grande stabilité dans nos conditions expérimentales, l'analyse de masse moléculaire par spectrométrie de masse a été réalisée sur l'oligomère P3. Nous avons observé une très forte hétérogénéité de l'état d'oligomérisation de la protéine, mais nous ne savons pas actuellement si cette hétérogénéité préexistait en solution ou si elle est le résultat d'une fragmentation dans la source d'electrospray. Pour répondre à cette question, nous envisageons d'analyser différentes fractions du pic P3, qui devraient refléter une éventuelle hétérogénéité en solution. Il semble cependant probable, d'après les différents résultats, que ces oligomères soient composés au minimum de 11 sous-unités.

IV.4. Etude structurale des oligomères par échanges H/D et spectrométrie de masse FT-ICR

IV.4.1. Introduction

Les oligomères de la protéine du prion ovin ont été caractérisés par une grande variété de méthodes biophysiques, comme la chromatographie d'exclusion stérique, la spectroscopie CD et la spectroscopie FT-IR [23]. Toutes ces méthodes permettent d'obtenir des informations globales sur la protéine entière. Néanmoins, il est également important d'obtenir des informations structurales plus fines sur certaines régions de la protéine soumises à des modifications lors de la formation des oligomères.

Des approches biochimiques ont été utilisées afin de répondre à ces questions, comme l'« epitope mapping » ou la mutagenèse dirigée. Ces expériences ont permis de montrer, par

exemple, que dans le cas de l'oligomère P3, la région qui correspond aux acides aminés 221-229 (localisés au C-terminal) n'est pas accessible [23].

Nous avons envisagé une approche complémentaire, basée sur le marquage des régions accessibles au solvant par échanges hydrogène/deutérium en solution, suivi d'une analyse par spectrométrie de masse. La mise au point de cette méthode a été présentée dans le chapitre III.C de ce manuscrit. Néanmoins, du fait des conditions expérimentales indispensables à la stabilité des oligomères, notamment le pH acide, il a été nécessaire d'adapter les conditions expérimentales pour ce système non standard (voir chapitre VI).

IV.4.2. Analyses préalables

Avant d'analyser les échanges H/D de la protéine du prion ovin dans ses différents états oligomériques, il a été nécessaire d'identifier les peptides obtenus après digestion par la pepsine. Dans le spectre de la protéine du prion ovin monomérique digérée par la pepsine pendant 5 minutes à 0°C (sur la glace) (Fig. IV.7), 45 peptides de la protéine ont été identifiés grâce à la haute résolution et la précision en masse de la spectrométrie de masse FT-ICR (Fig. IV.8). Ces peptides correspondent à une couverture de plus de 90% de la séquence primaire de la protéine. La figure IV.7 montre un exemple de la complexité du spectre MS, confirmant l'intérêt de la haute résolution de la spectrométrie de masse FT-ICR pour séparer les différents massifs isotopiques présents.

La plupart des peptides identifiés, en particulier les peptides les plus abondants, sont localisés dans la région 135-171. Les peptides correspondant aux autres régions de la protéine étant de plus faible intensité dans le spectre. Nous n'avons trouvé aucun peptide dans les régions 178-185 et 217-220, ce qui n'est pas surprenant puisqu'il existe un pont disulfure entre C182 et C217. La couverture de séquence (>90%) n'a pas pu être améliorée par l'utilisation d'autres protéases comme la protéase de type XIII ou la protéase de type XVIII, qui ont été utilisées avec succès sur d'autres systèmes en complément de la pepsine [40]. Les peptides les plus abondants identifiés dans le spectre sont tous localisés dans la même région de la protéine (acides aminés 135-171), ce qui pourrait indiquer une plus grande accessibilité de cette région de la protéine dans nos conditions expérimentales.

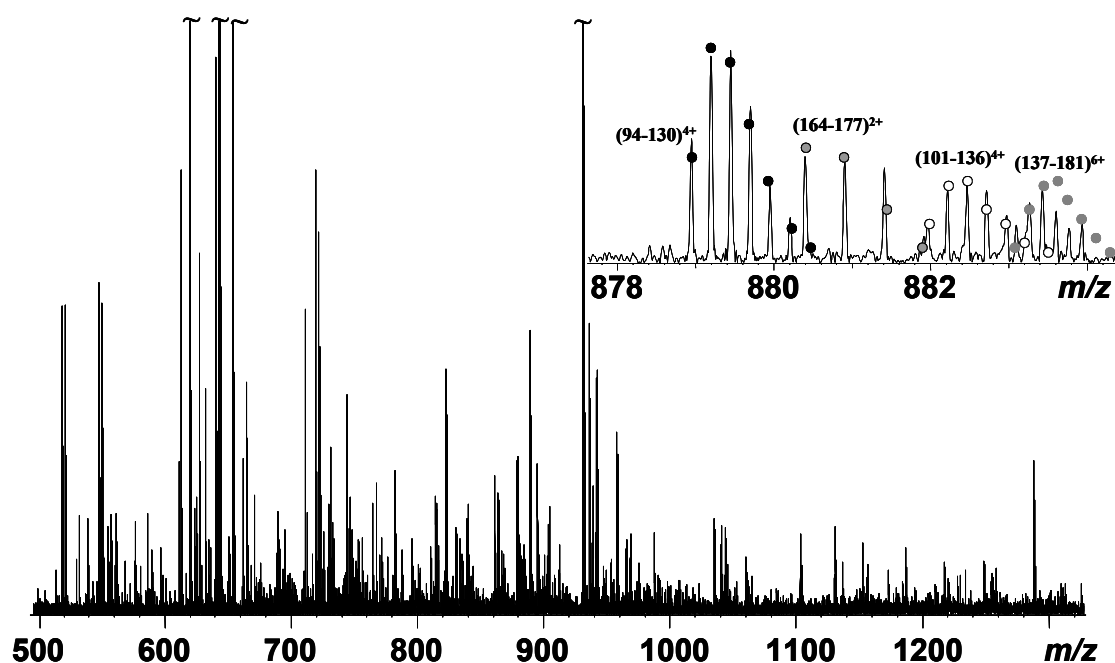


Fig. IV.7 : Spectre nano-ESI-FTMS de la protéine du prion ovin digérée par la pepsine. L'encart détaille une partie des peptides identifiés grâce à la haute résolution et à la précision en masse de la spectrométrie de masse FT-ICR.

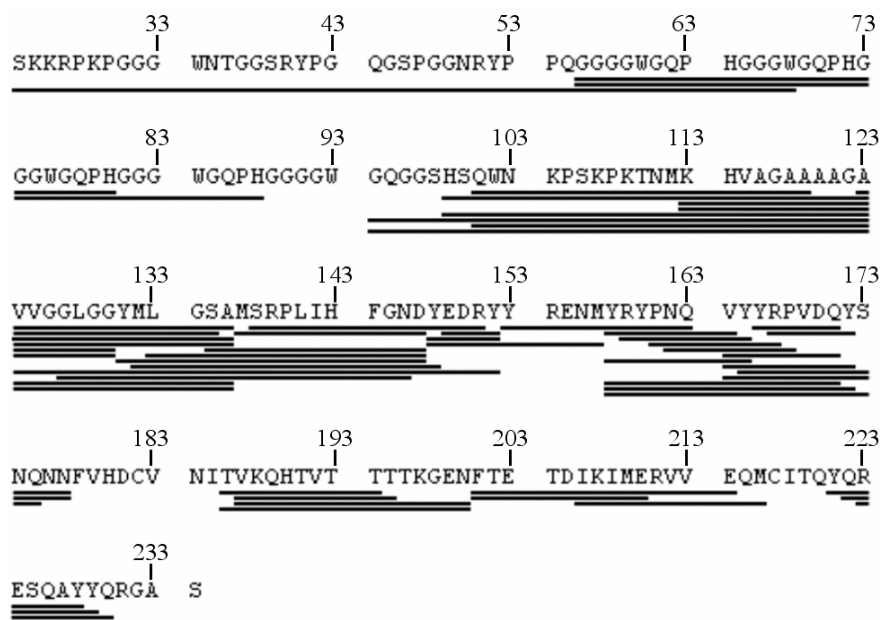


Fig. IV.8 : Représentation des peptides de la protéine du prion ovin après digestion par la pepsine. La séquence correspond aux acides aminés 24-234 du variant ARQ de la protéine.

D'après les expériences de spectroscopie IR [23], cette région S135-Q171 pourrait subir des modifications structurales importantes lors du processus d'oligomérisation et elle serait donc une des régions les plus intéressantes à étudier de façon détaillée par échanges H/D. Cependant, la plupart des études structurales des protéines par échanges hydrogène/deutérium analysés par spectrométrie de masse ont été réalisées à pH neutre ou légèrement basique afin de maximiser les échanges. Les oligomères de la protéine du prion n'étant stables qu'à pH acide, il a été nécessaire de vérifier la vitesse des échanges H/D à pH acide. Par ailleurs, la dépolymérisation des oligomères augmentant avec la température, il était nécessaire d'optimiser ce paramètre. Après avoir testé différentes conditions expérimentales, il a été décidé d'effectuer les échanges H/D à une température de 15°C ce qui est un bon compromis entre le maximum d'échange (qui nécessiterait une température maximale) et la stabilité des oligomères P1 et P3 dont nous avons vérifié la stabilité pendant au moins deux heures à 15°C. L'oligomère P2 n'est pas stable dans ces conditions expérimentales et ne constitue donc pas un système adapté à une étude par échanges H/D.

IV.4.3. Echanges H/D du monomère et des oligomères de la protéine du prion ovin

IV.4.3.1. Résultats

Le pourcentage d'échange de la forme monomérique ainsi que des oligomères P1 et P3 a été déterminé pour différents temps entre 0 et 120 minutes. Pour vérifier la reproductibilité des résultats, chaque point a été répété trois fois. Lors de l'analyse des échanges H/D du monomère de la protéine prion, 14 peptides de la protéine ont été retrouvés dans le spectre, ce qui correspond à une couverture de 55% de la séquence primaire de la protéine. Aucun peptide de la partie N-terminale ou des hélices *H2* et *H3* (voir Fig. IV.9) n'est présent dans le spectre bien que des peptides de ces régions aient été présents en faible abondance après une digestion par la pepsine de la protéine avant échange. Dans les spectres de masse des digestats des oligomères P1 et P3, seuls 9 peptides de la région S135-Q171 ont été observés, ce qui correspond à une couverture de 17,5% de la séquence primaire de la protéine.

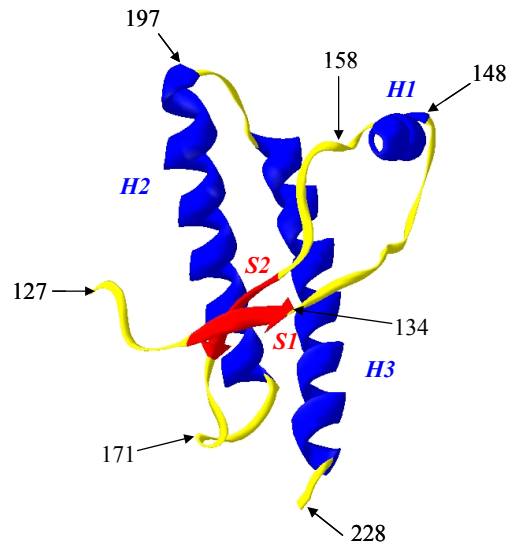


Fig. IV.9 : Structure tridimensionnelle de la partie G127-Y228 de la PrP ovine (variant ARQ). Les éléments de la structure secondaire sont indiqués (S : feuillet β ; H : hélice α)

Les différentes régions de la protéine du prion présentent des différences significatives d'incorporation de deutérium, à la fois pour un état donné de la protéine (par exemple le monomère), mais aussi entre différents états oligomériques. Ces résultats seront discutés pour les différentes régions de la protéine dans les paragraphes suivant.

Acides aminés 135-152

Les acides aminés S135-Y152 correspondent à une boucle située entre le feuillet beta *S1* et l'hélice *H1* ainsi qu'à la première partie de l'hélice *H1* (voir Fig. IV.9). Dans la région 135-147, on observe jusqu'à 25-30% d'échange après deux heures pour la forme monomérique de la PrP. Ce taux d'échange augmente légèrement pour les oligomères P1 et P3, mais on peut néanmoins considérer que l'accessibilité en solution de cette région change peu lors de l'oligomérisation. La structure de cette région de la protéine ne semble donc pas subir de changements significatifs lors de l'oligomérisation.

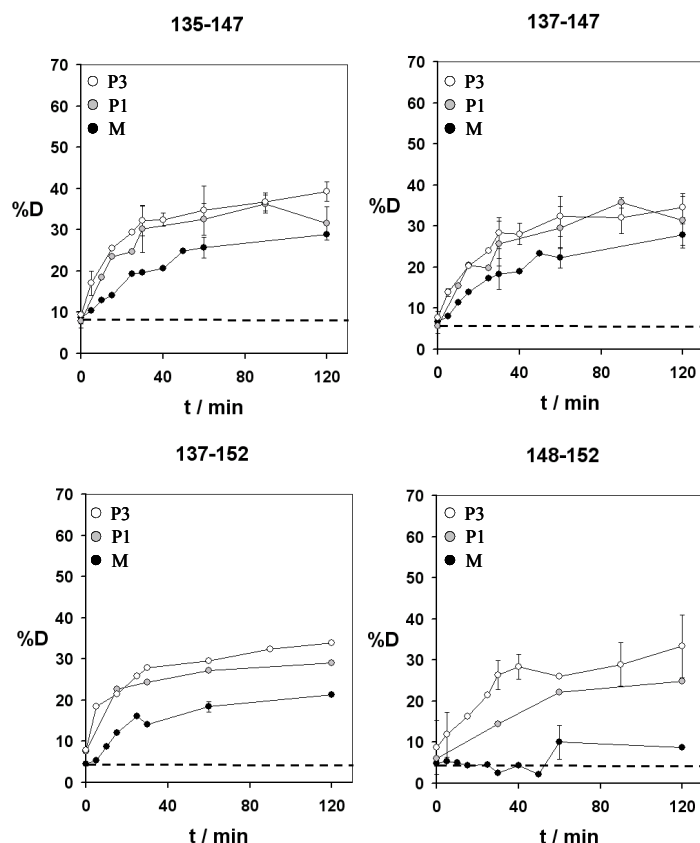


Fig. IV.10 : Pourcentage d'hydrogènes échangés pour les différentes formes de la PrP dans la région 135-152 en fonction du temps (point noir : monomère, M ; point gris : oligomère P1 ; point blanc : oligomère P3).

Le pourcentage de deutérium incorporé dans les peptides situés au début de l'hélice *HI* est significativement plus faible dans le monomère que celui incorporé dans les peptides adjacents.

Ainsi, le peptide M137-Y152 présente seulement 20% d'hydrogènes d'amides échangés après deux heures, et le peptide Y148-Y152, entièrement localisé dans l'hélice *HI*, moins de 10%. Le pourcentage d'échange à $t = 0$ étant de 5%, on peut considérer qu'il n'y a quasiment pas d'échange pour ce peptide dans l'échelle de temps considérée. Ces résultats montrent bien la différence entre la région non structurée (S135-D147) et une région structurée en début d'hélice α (Y148-Y152).

Les oligomères P1 et P3 présentent une augmentation significative de l'échange H/D pour les peptides M137-Y152 et Y148-Y152 par rapport aux résultats obtenus avec le monomère (Fig. IV.10). Cette région de la protéine devient donc plus accessible lors du processus d'oligomérisation.

Acides aminés 148-157

Dans le cas du monomère, les acides aminés de la région Y148-M157 sont essentiellement localisés dans l'hélice *H1* (Y148-E155). Cette région très structurée étant peu accessible au solvant, le pourcentage des hydrogènes échangés est très faible (<10%) pour tous les points expérimentaux obtenus pour le monomère (Fig. IV.11).

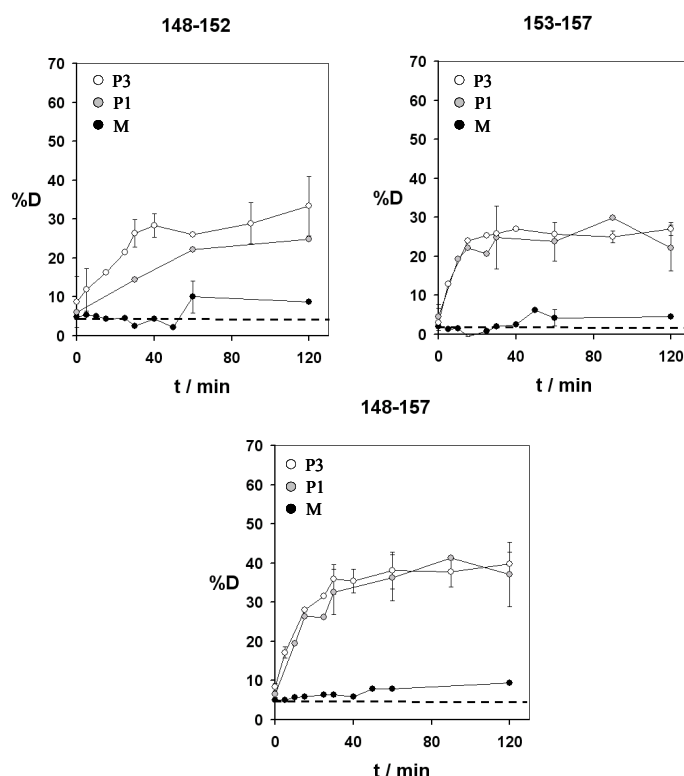


Fig. IV.11 : Pourcentage des hydrogènes échangés pour les différentes formes de la PrP dans la région 148-157 en fonction du temps (point noir : monomère, M ; point gris : oligomère P1 ; point blanc : oligomère P3).

Pour les oligomères P1 et P3, on constate une augmentation significative de l'échange H/D par rapport au monomère (Fig. IV.11), ce qui confirme les résultats observés pour les peptides S137-Y152 et Y148-Y152. Ce résultat indique une probable déstructuration de l'hélice *H1* dans les oligomères P1 et P3, ce qui serait en accord avec la perte de 7,6% en hélice α observée dans les expériences de spectroscopie IR [23]. L'hélice *H1* représentant environ 12% des hélices alpha de la protéine entière (d'après la structure tridimensionnelle obtenue par cristallographie aux rayons X), il est parfaitement vraisemblable qu'au moins une partie de l'hélice *H1* soit dénaturée lors du processus d'oligomérisation.

Acides aminés 158-171

Les peptides qui contiennent les acides aminés Y158-Q171 sont localisés dans la région située entre les hélices *H1* et *H2* qui contient également le feuillet bêta *S2* (V164-Y166) (voir Fig. IV.9). Les peptides de cette région présentent un pourcentage d'échange d'environ 15-20% après deux heures. Ce pourcentage est donc intermédiaire entre celui observé pour l'hélice *H1* et celui observé pour la région peu structurée S135-D147. La région Y158-Q171 étant localisée à l'intérieur de la protéine telle qu'elle est représentée d'après les expériences de cristallographie, l'accessibilité des hydrogènes d'amides est probablement plus faible que dans la région S135-D147. Dans le monomère, la région Y158-Y165 présente un pourcentage d'échange (15,7% après deux heures) légèrement inférieur à celui de la région Y164-Q171 (20,9% après deux heures).

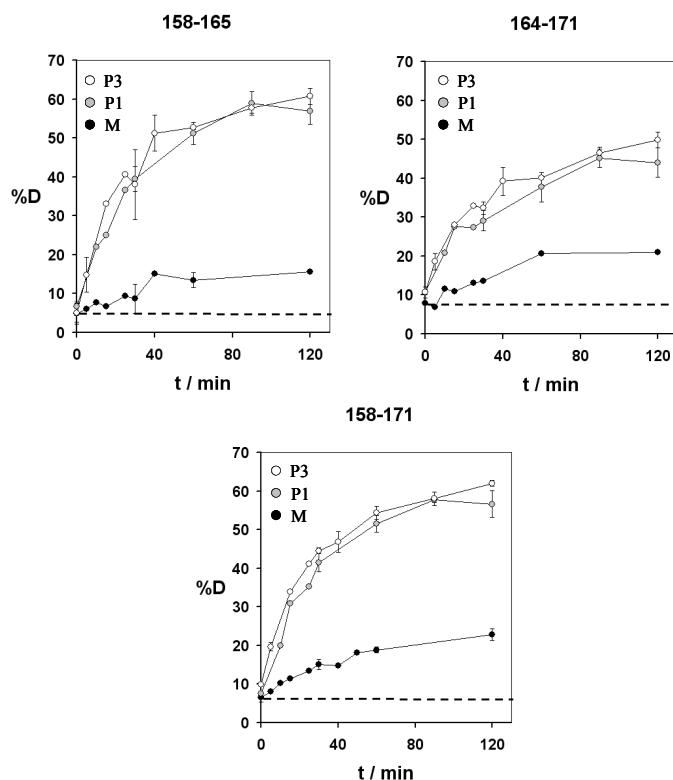


Fig. IV.12 : Pourcentage des hydrogènes échangés par la région 158-171 du monomère en fonction du temps.

Le pourcentage de deutérium incorporé par cette région dans les oligomères P1 et P3 est nettement plus élevé que dans le monomère (Fig. IV.12). En effet, le pourcentage de deutérium incorporé dans les peptides s'élève à plus que 50% (P1) et 60% (P3) pour les peptides Y158-Y165 et Y158-Q171 après deux heures d'échange (Fig. IV.12). La région 158-

171 est donc beaucoup plus accessible dans les oligomères P1 et P3 que dans le monomère. Contrairement à ce qui a été observé pour le monomère, le pourcentage d'échange dans cette région est même significativement plus élevé dans les oligomères que dans la région faiblement structurée 135-147. Une hypothèse pour expliquer cette importante augmentation de l'accessibilité dans la région 135-171 serait une « ouverture » de cette boucle conduisant à un éloignement de la partie *SI-HI-S2* du reste de la protéine. La boucle entre *HI* et *H2* ne serait donc plus « cachée » à l'intérieur de la protéine, mais en contact direct avec le solvant, ce qui expliquerait le pourcentage élevé d'échange dans cette région. Il est important de noter que le double mutant Y160C, M209C (qui contient un pont disulfure artificiel entre la boucle *HI-S2* et l'hélice *H3*) ne forme pas d'oligomères. L'absence de flexibilité dans la région de la boucle *HI-S2-H2* semble ainsi inhiber la formation des oligomères, ce qui renforce l'hypothèse d'une « ouverture » de cette région lors de la formation des oligomères.

Autres régions

Les peptides dont il a été discuté auparavant ont été observés dans les digestats de toutes les formes de la PrP (monomérique ainsi qu'oligomériques). Toutefois, certains peptides n'ont été trouvés que dans les digestats du monomère. Bien que ces peptides n'apportent pas d'informations sur les oligomères, les résultats obtenus sont en accord avec la structure connue de la forme monomérique de la PrP et valident ainsi l'approche choisie.

Acides aminés 94-136

Le peptide correspondant aux acides aminés G94-A136 a été identifié dans deux spectres de la forme monomérique de la PrP (0 minutes et 10 minutes). Il n'était pas présent dans les spectres des digestats des formes oligomériques de la PrP, si bien qu'il n'est pas possible de l'utiliser pour comparer cette partie de la séquence entre le monomère et les oligomères. Néanmoins, la comparaison avec le pourcentage d'échange des autres peptides (13% d'échange au maximum pour tous les autres peptides de la forme monomérique après 10 minutes) montre, pour ce peptide, un taux d'échange plus élevé (21% après 10 minutes). Cette observation est en accord avec la structure de la protéine puisque cette région est localisée dans la partie N-terminale, qui est très flexible et qui est donc susceptible de présenter un pourcentage d'échange élevé.

Acides aminés 186-214

Les peptides qui contiennent les acides aminés T186-E214 constituent la deuxième moitié de l'hélice *H2* (T186-L197), la boucle entre les hélices *H2* et *H3* (G198-T202) et la première partie de l'hélice *H3* (E203-E214). On observe une différence significative entre les régions avec une hélice et les régions qui ne contiennent pas d'hélices. Le pourcentage d'échange du peptide F201-D205, qui est essentiellement localisé dans la boucle *H2-H3* est le plus élevé observé (32,5% après deux heures d'échange) dans cette région. Bien que ce peptide n'ait pas pu être observé dans tous les spectres, les résultats indiquent que cette région est très accessible en solution. Les peptides T186-N200 et F201-E210 contiennent également chacun une partie de la boucle *H2-H3* et une partie des hélices *H2* (dans le peptide T186-N200) ou *H3* (dans le peptide F201-E210). Malgré la présence d'acides aminés appartenant à des hélices le pourcentage total d'échange dans ces peptides s'élève à environ 20% après deux heures d'échange, ce qui est comparable avec les résultats obtenus pour la région 158-171, qui est peu structurée. Ce résultat confirme celui obtenu sur le peptide F201-D205 qui indique un échange H/D très élevé dans la boucle *H2-H3*. Le peptide I206-E214, entièrement localisé dans l'hélice *H3* et donc très peu accessible en solution (tous les hydrogènes d'amides formant des liaisons hydrogènes), présente un pourcentage d'échange très faible (autour de 0-5%).

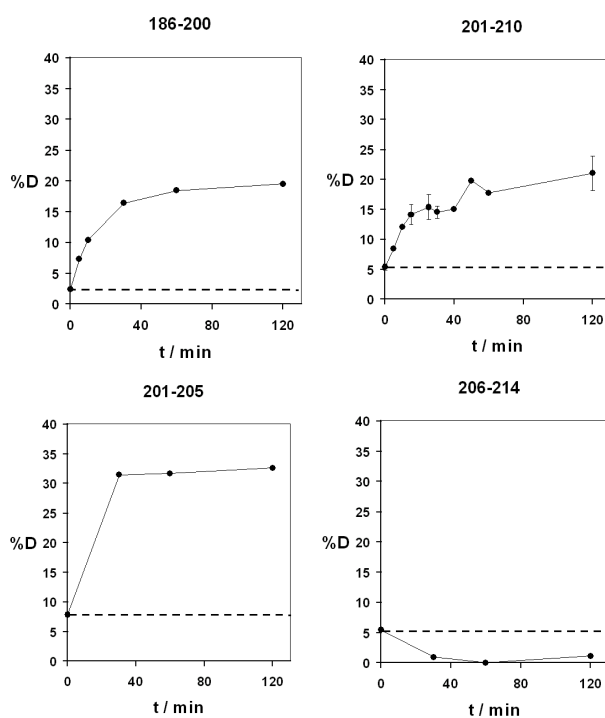


Fig. IV.13 : Pourcentage d'hydrogènes échangés dans la région 186-214 du monomère en fonction du temps.

Région 221-231

La masse observée pourrait correspondre à plusieurs peptides de la protéine : Y221-Y228, Q222-Y229, R223-Q230, E224-R231. Puisque tous ces peptides correspondent à la même région (Y221-R231), même s'il est impossible de préciser de quel(s) peptide(s) il s'agit, on peut l'analyser comme représentatif des échanges dans cette région C-terminale, notée « 221-231 ». Le pourcentage d'échange s'élève à 25% après deux heures d'échange, ce qui est comparable avec les valeurs obtenues pour les peptides de la région peu structurée S135-D147. Cette région de la protéine pourrait donc également être considérée comme relativement accessible, bien qu'elle semble contenir une hélice α d'après la structure obtenue par cristallographie aux rayons X [41]. Une hypothèse pour expliquer cette observation serait une perte partielle de la structure secondaire au niveau C-terminal. Par rapport à la structure cristallographique, la protéine est susceptible de présenter une flexibilité plus élevée en solution. Il serait donc possible que l'extrémité C-terminale soit plus accessible en solution que ce qu'on attendrait d'après la structure cristallographique.

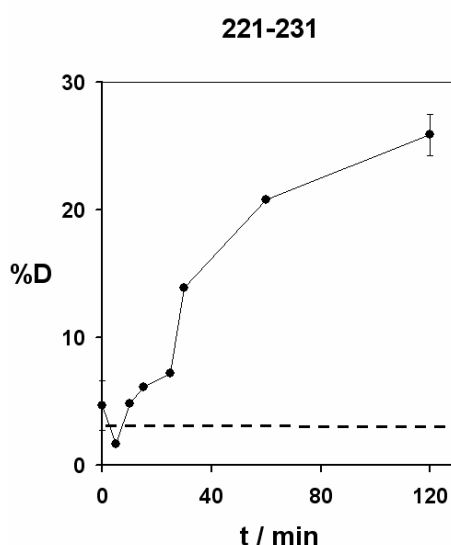


Fig. IV.14 : Pourcentage d'hydrogènes échangés dans la région 221-231 du monomère en fonction du temps.

IV.4.3.2. Conclusions

En résumé, les résultats obtenus par échanges H/D du monomère du variant ARQ de la protéine du prion ovin sont globalement en accord avec la structure obtenue par cristallographie aux rayons X (Fig. IV.15, Fig. IV.16). Les peptides entièrement localisés dans des régions comportant des hélices alpha (Y148-Y152, Y153-M157, Y148-M157, I206-E214) présentent un pourcentage d'échange très faible par rapport aux régions non structurées qui sont plus accessibles. Le peptide T186-N200 présente un pourcentage d'hydrogènes échangés assez élevé, malgré le fait qu'il contienne une grande partie de l'hélice *H2*, ce qui est probablement dû à la grande accessibilité de la boucle *H2-H3*. La région G127-G134 (qui contient *S1*) est représentée en rouge pour illustrer le pourcentage d'échange très élevé de cette région, discuté précédemment. La seule région de la partie globulaire de la protéine sur laquelle aucune information n'a pu être obtenue est la région située autour du pont disulfure (C182-C217).

Concernant les oligomères, les échanges H/D de P1 et P3 présentent des différences significatives par rapport à la forme monomérique de la protéine. En effet, l'hélice *H1* ainsi que la boucle entre *H1* et *H2* sont beaucoup plus accessibles en solution dans les oligomères que dans le monomère. Ces résultats reflètent vraisemblablement des changements conformationnels de la PrP ayant eu lieu lors de l'oligomérisation. Les pourcentages d'échange de tous les peptides observés sont plus élevés dans le cas de l'oligomère P3 que dans le cas de l'oligomère P1. Ce résultat s'est confirmé lors de la répétition des expériences. Une explication possible pour cette observation serait que la structure des sous-unités de l'oligomère P3 soit thermodynamiquement moins stable que celle de l'oligomère P1, ce qui avait déjà été démontré par des cinétiques d'oligomérisation analysées par chromatographie d'exclusion stérique.

En résumé, le bon accord des résultats obtenus par échanges H/D sur la forme monomérique avec la structure tridimensionnelle obtenue par cristallographie aux rayons X valide l'approche choisie. Des différences ont été observées entre les oligomères et le monomère, reflétant des changements structuraux dus à l'oligomérisation.

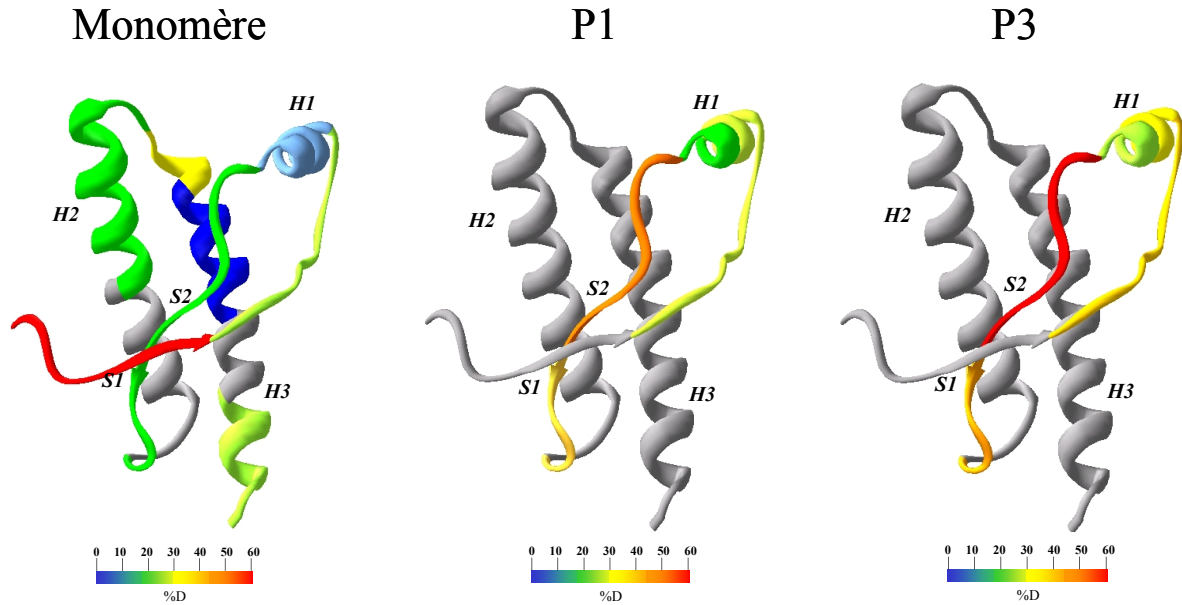


Fig. IV.15 : Pourcentage d'hydrogènes échangés dans les différentes régions du monomère et des oligomères de la PrP représentés sur la structure tridimensionnelle du monomère obtenue par cristallographie aux rayons X (du côté N-terminal, seuls les acides aminés à partir de G127 figurent sur la structure cristallographique). Les peptides des régions en gris n'ont pas été observés.

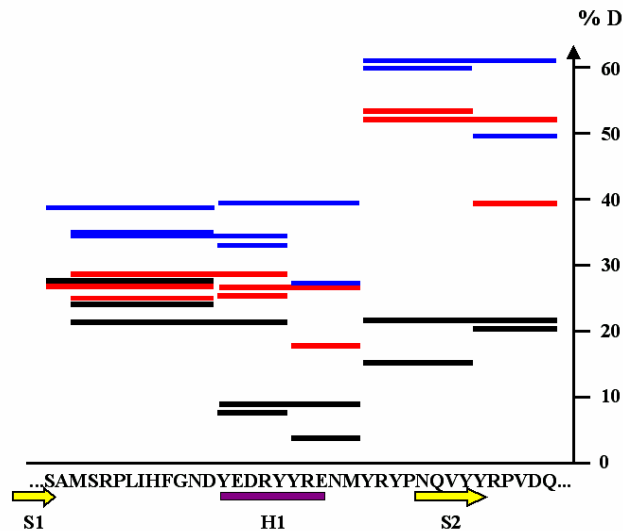


Fig. IV.16 : Comparaison du pourcentage de deutérium incorporé dans les peptides de la région 135-171 après un temps d'échange de 2 heures représenté sur la structure primaire de la protéine (noir : monomère ; rouge : oligomère P1 ; bleu : oligomère P3). Les éléments de structure secondaire sont indiqués en bas.

IV.4.4. Comparaison des résultats obtenus par échanges H/D avec la structure de la PrP^{Sc}

L'analyse des expériences d'échange H/D apporte des informations précieuses pour la compréhension de la structure des oligomères de la protéine du prion ovin. Il serait intéressant d'établir un lien entre ces résultats et la structure de la PrP^{Sc}. Une comparaison des accessibilités en solution des oligomères avec les éléments structuraux de la PrP^{Sc}, déduits de la structure proposée par Govaerts et al. [18], fait ressortir des similitudes intéressantes (Fig. IV.17). Bien que les résultats obtenus par échanges H/D sur les oligomères ne représentent qu'une partie relativement petite de la protéine entière, certains éléments sont notables :

- L'absence de l'hélice *H1* dans le modèle de la PrP^{Sc} est en accord avec les résultats obtenus par échanges H/D
- Le dépliement de la boucle *S1-H1-S2* et son écartement de la moitié *H2-H3*, est nécessaire pour aboutir à la structure de la PrP^{Sc}
- La région de la PrP^{Sc} qui correspond à la boucle *H1-H2* (Fig. IV.17, parties orange et rouge) présente une accessibilité élevée, ce qui est en accord avec nos résultats

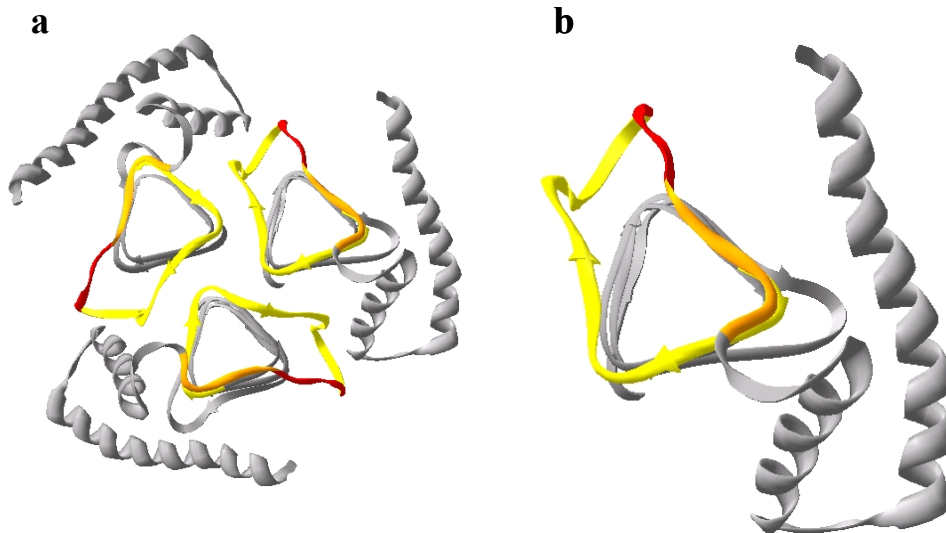


Fig. IV.17 : Représentation des résultats obtenus par échanges H/D sur la structure proposée pour PrP^{Sc} [18]. (a) unité de base ; (b) détail d'un monomère de l'unité de base. La coloration correspond aux résultats obtenus pour l'oligomère P3, avec en rouge les régions les plus accessibles et en jaune les régions les moins accessibles (pour P1, les résultats sont sensiblement les mêmes).

Il serait tentant d'en conclure que les informations structurales obtenues sur les oligomères sont également valables pour la PrP^{Sc}. Toutefois, malgré l'excellente concordance des résultats, une telle conclusion reste difficile :

- Les résultats n'ont pas été obtenus dans les mêmes conditions expérimentales. Pour les oligomères, les études ont été faites en solution, tandis que la structure de PrP^{Sc} a été déduite de données obtenues à partir de cristaux.
- Les oligomères ont été obtenus par dénaturation thermique *in vitro*, tandis que la PrP^{Sc} a été formée *in vivo*, ce qui correspond à un environnement biologique et des influences potentielles sur la structure considérablement différents
- Les oligomères et la PrP^{Sc} n'ont pas le même comportement biochimique : par exemple, contrairement à la PrP^{Sc}, les oligomères sont sensibles à la digestion enzymatique.

Néanmoins, il est important de noter que les résultats obtenus par échanges H/D ne sont pas incompatibles avec la structure proposée pour la PrP^{Sc}. Il serait intéressant d'approfondir cette question par des études ultérieures afin d'obtenir plus d'informations sur le rapport structural entre ces deux formes de la protéine du prion. Par exemple, on pourrait envisager d'effectuer les échanges H/D sur des agrégats de la PrP ayant une taille plus élevée ou directement sur la forme PrP^{Sc}. L'incorporation des deutériums pourrait ensuite être étudiée soit par digestion enzymatique et identification des peptides (si la digestion du complexe permet d'obtenir une bonne couverture de séquence) soit par fragmentation du monomère de la protéine en phase gazeuse, par exemple par une combinaison de l'ECD et de l'IRMPD (à condition qu'une dissociation des agrégats présents en solution soit possible afin d'obtenir des espèces monomériques en phase gazeuse).

IV.4.5. Site d'interaction entre les sous-unités des oligomères

Quelles sont les informations obtenues sur le site d'interaction entre les sous-unités grâce aux résultats des échanges H/D ? Tout d'abord, il est important de noter que pour les peptides observés, aucune diminution de l'échange H/D dans les oligomères n'a été observée. Cela indique que le site d'interaction n'est pas localisé dans les régions observées. Les régions du site d'interaction étant susceptibles d'être peu accessibles lors de la digestion enzymatique, cela expliquerait l'absence des peptides correspondant dans les spectres des digestats d'oligomères. Par contre, ces peptides devraient être présents dans les spectres des digestats

du monomère. Cela pourrait être le cas au niveau des hélices *H2* et *H3* : l'absence des peptides dans cette région après digestion des oligomères pourrait indiquer une accessibilité faible ou même une inaccessibilité des acides aminés dans cette région. Néanmoins, il n'est pas possible d'obtenir des informations sur le site d'interaction entre les sous-unités des oligomères sur la seule base des données que nous avons obtenues lors des échanges H/D.

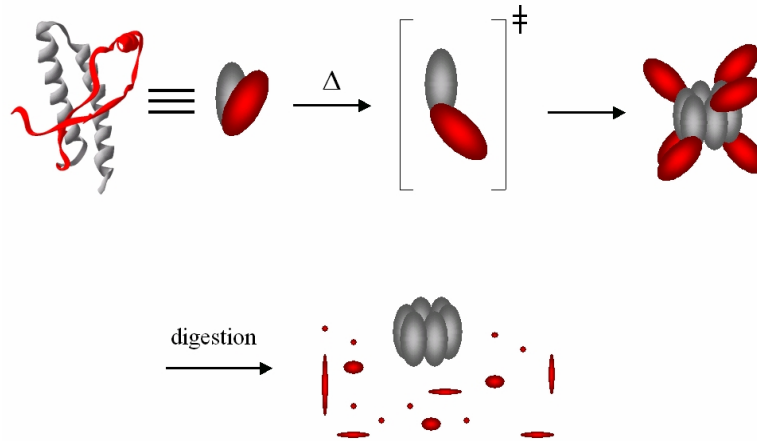


Fig. IV.18 : Modèle permettant d'expliquer l'absence des peptides de la région des hélices *H2* et *H3* dans le digestat des oligomères par l'interaction entre les sous-unités.

Il est important de noter qu'après délétion des acides aminés de la partie N-terminale (acides aminés 23 à 88) et de la partie comportant l'hélice *H1* et la boucle *S2* comprise entre les hélices *H1* et *H2* (acides aminés 141 à 176), la forme tronquée de la protéine du prion conserve ses propriétés infectieuses, ce qui signifie qu'elle peut encore se transformer en PrP^{Sc} [42]. Ceci signifie qu'une autre région de la protéine est fondamentale pour la transformation de la PrP^C en PrP^{Sc}. Dans l'hypothèse où cette transformation impliquerait une oligomérisation de la protéine, il apparaît donc évident que le site d'interaction ne peut pas se situer dans les régions délétées, mais pourrait se situer comme nous le proposons dans la région située autour des hélices *H2* et *H3*.

Nous avons réalisé un certain nombre d'expériences dans le but de valider cette hypothèse. En effet, dans le cas où la région des hélices *H2* et *H3* serait protégée de la digestion enzymatique par l'interaction entre les sous-unités, un fragment de grande taille correspondant à cette région devrait être détecté après la digestion par la pepsine. Aucun fragment correspondant n'a pu être observé, mais il est possible que la digestion n'ait été que partielle dans cette région et ait généré un grand nombre de peptides chacun en quantité trop

faible pour être détecté. Par ailleurs, il est possible que l'interaction soit maintenue et même renforcée par la perte de l'environnement protéique après la digestion, et que ces agrégats n'aient pas été dénaturés, ou que leur rendement d'ionisation ait été trop faible pour une détection par spectrométrie de masse. L'analyse du digestat pepsique des oligomères par chromatographie d'exclusion suivie de tests ELISA n'a pas non plus permis de mettre en évidence un fragment non digéré correspondant à la région située autour des hélices *H2* et *H3*. Nous n'avons donc obtenu que peu d'informations concernant la localisation du site d'interaction.

Toutefois, l'hypothèse d'une oligomérisation au niveau de la région des hélices *H2* et *H3* reste en accord non seulement avec les résultats des échanges H/D, mais également avec les données bibliographiques [42].

IV.5. Conclusions

Des conditions expérimentales ont été mises au point pour purifier et lyophiliser des oligomères de la PrP afin de permettre une étude par différentes techniques faisant appel à la spectrométrie de masse. La détermination de la masse moléculaire des oligomères par spectrométrie de masse n'a pas été possible à cause de l'instabilité des oligomères en solution (P2), lors de la concentration/lyophilisation (P1) ou lors du processus d'électrospray (P3).

En revanche, les résultats obtenus par échanges H/D et spectrométrie de masse FT-ICR montrent des différences significatives de la structure de la PrP entre ses états monomérique et oligomériques. Lors de l'oligomérisation, la partie correspondant à l'hélice *H1* et aux régions l'encadrant (*S1-H1-S2*) de la forme monomérique devient plus accessible. Ce résultat s'explique par le dépliement (partiel ou total) de l'hélice *H1* et par une « ouverture » de la protéine due à la flexibilité de la boucle *H1-S1-H2*. Ces résultats sont en accord avec les expériences de FT-IR et avec des résultats obtenus par mutagenèse dirigée.

Pour l'oligomère P3, on peut noter que tous les peptides observés présentent un pourcentage d'échange légèrement plus élevé que pour l'oligomère P1. Cette différence pourrait s'expliquer par la plus grande stabilité de l'oligomère P1 par rapport à P3, qui a été démontrée par chromatographie d'exclusion stérique.

Ces résultats montrent aussi que la région constituée des acides aminés 135-171 (comprenant l'hélice *H1*) ne correspond pas au site d'interaction des sous-unités dans les oligomères puisque le pourcentage de deutérium incorporé augmente dans les oligomères pour tous les peptides observés dans cette région.

L'absence des peptides localisés au niveau des hélices *H2* et *H3* pourrait s'expliquer par leur localisation dans le site d'interaction entre les sous-unités des oligomères. Toutefois, il n'a pas été possible d'en apporter la preuve dans le cadre de nos expériences.

Les résultats obtenus par échanges H/D sont compatibles avec la structure qui a été proposée pour la PrP^{Sc}. Bien que ce résultat ne soit pas la preuve d'une corrélation entre ces deux structures, il peut donner des pistes pour des études ultérieures.

IV.6. Bibliographie

- [1] B. Hörnlimann, Prionen und Prionkrankheiten, dans : *Prionen und Prionkrankheiten*, B. Hörnlimann, D. Riesner, H. Kretzschmar, éditeurs, Walter de Gruyter, Berlin (2001), pages 3-20.
- [2] S. B. Prusiner, Novel proteinaceous infectious particle cause scrapie, *Science* **1982**, 216, 136-144.
- [3] S. B. Prusiner, D. E. Garfin, S. P. Cochran, J. R. Baringer, W. J. Hadlow, C. M. Eklund, R. E. Race, Evidence for hydrophobic domains on the surface of the scrapie agent, *Trans. Am. Neurol. Assoc.* **1978**, 103, 62-64.
- [4] S. B. Prusiner, M. P. McKinley, D. F. Groth, K. A. Bowman, N. I. Mock, S. P. Cochran, F. R. Masiarz, Scrapie agent contains a hydrophobic protein, *Proc. Natl. Acad. Sci. USA* **1981**, 78, 6675-6679.
- [5] D. C. Bolton, M. P. McKinley, S. B. Prusiner, Identification of a protein that purifies with the scrapie prion, *Science* **1982**, 218, 1309-1311.
- [6] S. B. Prusiner, M. P. McKinley, K. A. Bowman, D. C. Bolton, P. E. Bendheim, D. F. Groth, G. G. Glenner, Scrapie prions aggregate to form amyloid-like birefringent rods, *Cell* **1983**, 35, 349-358.
- [7] S. B. Prusiner, D. F. Groth, D. C. Bolton, S. B. Kent, L. E. Hood, Purification and structural studies of a major scrapie prion protein, *Cell* **1984**, 38, 127-134.
- [8] B. Oesch, D. Westaway, M. Walchli, M. P. McKinley, S. B. Kent, R. Aebersold, R. A. Barry, P. Tempst, D. B. Teplow, L. E. Hood et al., A cellular gene encodes scrapie PrP27-30 protein, *Cell* **1985**, 40, 735-746.
- [9] B. Chesebro, R. Race, K. Wehrly, J. Nishio, M. Bloom, D. Lechner, S. Bergstrom, K. Robbins, L. Mayer, J. M. Keith et al., Identification of scrapie prion protein-specific mRNA in scrapie-infected and uninfected brain, *Nature* **1985**, 315, 331-333.

- [10] N. Stahl, M. A. Baldwin, D. B. Teplow, L. Hood, B. W. Gibson, A. L. Burlingame, S. B. Prusiner, Structural studies of the scrapie prion protein using mass spectrometry and amino acid sequencing, *Biochemistry* **1993**, 32, 1991-2002.
- [11] K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. Huang, R. J. Fletterick, F. E. Cohen, S. B. Prusiner, Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 10962-10966.
- [12] M. P. McKinley, R. K. Meyer, L. Kenaga, F. Rahbar, R. Cotter, A. Serban, S. B. Prusiner, Scrapie prion rod formation in vitro requires both detergent extraction and limited proteolysis, *J. Virol.* **1991**, 65, 1340-1351.
- [13] R. A. Somerville, Ultrastructural links between scrapie and Alzheimer's disease, *Lancet* **1985**, 1, 504-506.
- [14] A. Aguzzi, C. Haass, Games played by rogue proteins in prion disorders and Alzheimer's disease, *Science* **2003**, 302, 814-818.
- [15] B. W. Caughey, A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes, W. S. Caughey, Secondary structure analysis of the scrapie-associated protein PrP27-30 in water by infrared spectroscopy, *Biochemistry* **1991**, 30, 7672-7680.
- [16] J. T. Nguyen, M. Inouye, M. A. Baldwin, R. J. Fletterick, F. E. Cohen, S. B. Prusiner, D. A. Kirschner, X-ray diffraction of scrapie prion rods and PrP peptides, *J. Mol. Biol.* **1995**, 252, 412-422.
- [17] H. Wille, M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard, S. B. Prusiner, Structural studies of the scrapie prion protein by electron crystallography, *Proc. Natl. Acad. Sci. USA* **2002**, 289, 1163-1178.
- [18] C. Govaerts, H. Wille, S. B. Prusiner, F. E. Cohen, Evidence for assembly of prions with left-handed beta-helices into trimers, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 8342-8347.
- [19] B. Caughey, P. T. Lansbury, Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders, *Annu. Rev. Neurosci.* **2003**, 26, 267-298.
- [20] N. Reixach, S. Deechongkit, X. Jiang, J. W. Kelly, J. N. Buxbaum, Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture, *Proc. Natl. Acad. Sci. USA*, **2004**, 101, 2817-2822.
- [21] S. B. Prusiner, D. E. Garfin, S. P. Cochran, M. P. McKinley, D. F. Groth, W. J. Hadlow, R. E. Race, C. M. Eklund, Experimental scrapie in the mouse: electrophoretic and sedimentation properties of the partially purified agent, *J. Neurochem.* **1980**, 35, 574-582.

- [22] S. Tzaban, G. Friedlander, O. Schonberger, L. Horonchik, Y. Yedidia, G. Shaked, R. Gabizon, A. Taraboulos, Protease-sensitive scrapie prion protein in aggregates of heterogeneous sizes, *Biochemistry* **2002**, *41*, 12868-12875.
- [23] H. Rezaei, F. Eghiaian, J. Perez, B. Doublet, Y. Choiset, T. Haertle, J. Grosclaude, Sequential generation of two structurally distinct ovine prion protein soluble oligomers displaying different biochemical reactivities, *J. Mol. Biol.* **2005**, *347*, 665-679.
- [24] A. Aguzzi, M. Polymenidou, Mammalian prion biology: one century of evolving concepts, *Cell* **2004**, *116*, 313-327.
- [25] S. Hornemann, R. Glockshuber, Autonomous and reversible folding of a soluble amino-terminally truncated segment of the mouse prion protein, *J. Mol. Biol.* **1996**, *261*, 614-619.
- [26] G. Wildegger, S. Liemann, R. Glockshuber, Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates, *Nat. Struct. Biol.* **1999**, *6*, 550-553.
- [27] I. V. Baskakov, G. Legname, M. A. Baldwin, S. B. Prusiner, F. E. Cohen, Pathway complexity of prion protein assembly into amyloid, *J. Biol. Chem.* **2002**, *277*, 21140-21148.
- [28] O. V. Bocharova, L. Breydo, A. S. Parfenov, V. V. Salnikov, I. V. Baskakov, In vitro conversion of full-length mammalian prion protein produces amyloid form with physical properties of PrP(Sc), *J. Mol. Biol.* **2005**, *346*, 645-659.
- [29] S. Hornemann, R. Glockshuber, A scrapie-like unfolding intermediate of the prion protein domain PrP(121-231) induced by acidic pH, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6010-6014.
- [30] G. S. Jackson, L. L. Hosszu, A. Power, A. F. Hill, J. Kenney, H. Saibil, C. J. Craven, J. P. Waltho, A. R. Clarke, J. Collinge, Reversible conversion of monomeric human prion protein between native and fibrillogenic conformations, *Science* **1999**, *283*, 1935-1937.
- [31] W. Swietnicki, M. Morillas, S. G. Chen, P. Gambetti, W. K. Surewicz, Aggregation and fibrillization of the recombinant human prion protein huPrP90-231, *Biochemistry* **2000**, *39*, 424-431.
- [32] H. Rezaei, Y. Choiset, F. Eghiaian, E. Treguer, P. Mentre, P. Debey, J. Grosclaude, T. Haertle, Amyloidogenic unfolding intermediates differentiate sheep prion protein variants, *J. Mol. Biol.* **2002**, *322*, 799-814.
- [33] W. Q. Zou, N. R. Cashman, Acidic pH and detergents enhance in vitro conversion of human brain PrP^C to a PrP^{Sc}-like form, *J. Biol. Chem.* **2002**, *277*, 43942-43947.
- [34] J. Kazlauskaitė, N. Sanghera, I. Sylvester, C. Venien-Bryan, T. J. Pinheiro, Structural changes of the prion protein in lipid membranes leading to aggregation and fibrillization, *Biochemistry* **2003**, *42*, 3295-3304.

- [35] J. Torrent, M. T. Alvarez-Martinez, M. C. Harricane, F. Heitz, J. P. Liautard, C. Balny, R. Lange, High pressure induces scrapie-like prion protein misfolding and amyloid fibril formation, *Biochemistry* **2004**, *43*, 7162-7170.
- [36] G. Legname, I. V. Baskakov, H. O. Nguyen, D. Riesner, F. E. Cohen, S. J. DeArmond, S. B. Prusiner, Synthetic mammalian prions, *Science* **2004**, *305*, 673-676.
- [37] B. Y. Lu, J. Y. Chang, Isolation and characterization of a polymerized prion protein, *Biochem. J.* **2002**, *364*, 81-87.
- [38] F. Sokolowski, A. J. Modler, R. Masuch, D. Zirwer, M. Baier, G. Lutsch, D. A. Moss, K. Gast, D. Naumann, Formation of critical oligomers is a key event during conformational transition of recombinant syrian hamster prion protein, *J. Biol. Chem.* **2003**, *278*, 40481-40492.
- [39] J. Torrent, M. T. Alvarez-Martinez, F. Heitz, J. P. Liautard, C. Balny, R. Lange, Alternative prion structural changes revealed by high pressure, *Biochemistry* **2003**, *42*, 1318-1325.
- [40] H. Mazon, O. Marcillat, E. Forest, C. Vial, Local dynamics measured by hydrogen/deuterium exchange and mass spectrometry of creatine kinase digested by two proteases, *Biochimie* **2005**, *87*, 1101-1110.
- [41] F. Eghiaian, J. Grosclaude, S. Lesceu, P. Debey, B. Doublet, E. Treguer, H. Rezaei, M. Knossow, Insights into the PrPC --> conversion from the structure of antibody-bound ovine prion scrapie-susceptibility variants, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10254-10259.
- [42] I. V. Baskakov, C. Aagaard, I. Mehlhorn, H. Wille, D. Groth, M. A. Baldwin, S. B. Prusiner, F. E. Cohen, Self-assembly of recombinant prion protein of 106 residues, *Biochemistry* **2000**, *39*, 2792-2804.

Partie V

Conclusion générale

Les travaux menés au cours de cette thèse ont permis le développement, la mise au point et la validation de nouvelles méthodes analytiques pour l'étude de la structure de protéines et de leurs interactions non covalentes avec des ligands en utilisant la technique de la spectrométrie de masse FT-ICR. Ces méthodes ont ensuite été appliquées à des systèmes ayant un intérêt biologique, notamment pour étudier l'interaction entre la créatine kinase et ses ligands ADP et ATP et pour obtenir des informations structurales sur des oligomères de la protéine du prion ovin.

Un premier angle d'attaque a été de regarder si la résolution et la précision en masse du spectromètre de masse pouvaient être mises à profit pour l'étude directe de l'interaction d'une protéine avec des petits ligands. Nous avons ainsi montré qu'il était tout à fait possible de distinguer sur un dimère de CK de 86 kDa entre la fixation d'un ATP et d'un ADP (80 Da d'écart) et même d'observer une compétition entre ces ligands. Toutefois ce travail a également mis en évidence la difficulté que pose la détermination de constantes d'interaction du fait de l'apparition d'interactions non spécifiques lorsque les constantes de dissociation sont élevées. Un modèle permettant de séparer les contributions spécifiques des contributions non spécifiques, basé sur une approche statistique, a été mis au point pour contourner cette limitation. L'analyse des résultats obtenus comparés aux résultats de la littérature obtenus par d'autres techniques soulève de nombreuses questions. En particulier, elle nous conduit à formuler l'hypothèse d'une dépendance des résultats au type d'observable sondée par la méthode expérimentale. Il faudrait conforter ou infirmer cette hypothèse en examinant d'autres systèmes présentant également des variations importantes de constantes suivant les méthodes utilisées.

Un second accès à la structure est d'introduire une modification de masse propre à la structure en solution de l'espèce étudiée, en se basant sur l'accessibilité de la protéine à des réactifs présents dans le solvant ou au solvant lui-même. Ces techniques permettent de s'affranchir du problème de la désolvatation, mais posent la difficulté de détecter et localiser efficacement ces modifications de masse. Deux types de modifications ont été envisagés : d'une part la modification chimique par des réactifs spécifiques de résidus accessibles au solvant et d'autre part la technique plus classique des échanges hydrogène/deutérium avec le solvant. Dans le premier cas, nous avons développé une méthode basée sur la modification de résidus histidine et lysine par le diéthylpyrocarbonate (DEPC), et tenté de développer une approche « top-down » utilisant la fragmentation par ECD d'une protéine entière pour localiser les sites de modification et de la comparer avec l'approche « bottom-up » plus classique. Cette approche est spécifique des instruments de type FT-ICR car elle nécessite de

disposer du mode de fragmentation par ECD et d'une très bonne précision sur la mesure de masse des fragments. Cette approche a été validée sur l'ubiquitine par l'identification précise des résidus modifiés. Dans la configuration actuelle de notre instrument, cette approche « top-down » semble limitée à la localisation de modifications situées à proximité des extrémités N et C-terminales de la protéine et à des protéines d'une masse inférieure à 25 kDa. L'extension de cette gamme de masse et de la couverture de séquence pourrait se faire par l'ajout d'un laser IR permettant une préexcitation vibrationnelle des protéines ionisées dans la cellule du spectromètre de masse. L'approche bottom-up, quant à elle, a été validée en observant des modifications d'accessibilité lors de la dénaturation de la CK. Un second type de modifications de masse pouvant être utilisé est l'échange des hydrogènes d'amides par des deutériums par incubation dans D₂O. L'intérêt de la spectrométrie de masse FT-ICR dans une approche « bottom-up » classique de ce système est que sa précision en masse permet de se passer d'étapes de séparation chromatographique, ce qui, compte tenu des contraintes sur le temps d'analyse inhérentes à la technique des échanges H/D, constitue une amélioration importante.

Finalement, quelques unes des techniques mises au point ont été appliquées à l'étude des oligomères de la protéine du prion ovin. Après avoir mis au point les conditions expérimentales pour purifier les oligomères par chromatographie d'exclusion stérique, l'homogénéité des oligomères a été étudiée par spectrométrie de masse dans des conditions non dénaturantes sur différents instruments (ICR et Q - TOF). Pour l'instant, les résultats sont difficiles à interpréter : un mélange d'oligomères est observé dans les deux instruments, mais l'origine de celui-ci, en solution, ou par décomposition d'un multimère de taille plus importante lors de la désolvatation, n'est pas encore bien établie. Ce travail a toutefois jeté les bases d'une étude plus complète, en s'intéressant en particulier aux différentes fractions des pics chromatographiques séparés par exclusion stérique. Les structures des oligomères ont également été étudiées par échanges hydrogène/deutérium. Les différences de taux d'échange entre monomère et oligomères ont permis d'identifier les régions de la PrP qui subissent des changements structuraux lors de l'oligomérisation. Du fait de l'absence de certains peptides, il n'a pas été possible de localiser les sites d'interaction entre les sous-unités des oligomères. Toutefois, cette absence pourrait elle-même indiquer une accessibilité réduite aux enzymes utilisées pour la protéolyse des régions correspondantes dans les oligomères, ce qui donne des pistes pour caractériser plus en détail la structure de ces oligomères.

Partie VI

Matériel et méthodes

Nous avons regroupé dans cette dernière partie quelques méthodes utilisées dans différentes parties de ce travail et dont le détail aurait alourdi les résultats exposés dans les parties précédentes.

VI.1. Instrumentation

VI.1.1. Ionisation par électronébulisation

L'analyse de molécules par spectrométrie de masse nécessite d'ioniser ces molécules et de les transférer d'un milieu liquide ou solide à un vide poussé. Pour cela, différentes sources d'ionisation ont été développées, adaptées à différents types d'échantillons.

L'électronébulisation (ou electrospray) est une technique d'ionisation qui permet de transférer des macromolécules (par exemple des protéines) intactes en phase gazeuse. Grâce à l'absence de fragmentations induites lors de l'ionisation (comme dans le cas d'autres modes d'ionisation comme l'IE ou l'IC), l'electrospray est considéré comme une technique d'ionisation « douce ». Elle a été développée par John Fenn dans les années 1980 [1] et aujourd'hui, elle est (avec le MALDI) la méthode de référence pour la spectrométrie de masse des biomolécules. Les conditions d'ionisation « douces » permettent notamment de préserver les interactions non covalentes lors du passage en phase gazeuse, ce qui permet l'analyse de systèmes présentant de telles interactions. Dans cette section, les bases et les mécanismes de la technique d'electrospray seront rappelés. Ce préalable est nécessaire pour comprendre le contexte de l'étude des interactions non covalentes protéine-ligand et protéine-protéine par cette technique, qui sera discuté dans les sections suivantes.

VI.1.1.1. La formation des ions en electrospray

La formation des ions en electrospray se déroule en trois étapes principales :

1. émission des gouttelettes chargées à partir de la solution à l'extrémité d'un capillaire chargé
2. évaporation du solvant et fission des gouttelettes par explosions coulombiennes successives
3. émission des ions en phase gazeuse à partir des gouttelettes

Ces trois étapes vont être détaillées dans la suite de ce chapitre.

VI.1.1.1.1. La formation de gouttelettes chargées

L'étape initiale du processus d'electrospray est l'émission de gouttelettes chargées à partir d'une solution à l'extrémité d'un capillaire (Fig. VI.1). Ceci est obtenu grâce à l'application d'une forte différence de potentiel, de l'ordre de quelques kV, entre le capillaire et une contre électrode ce qui induit un champ électrique intense de l'ordre de 10^6 V/m.

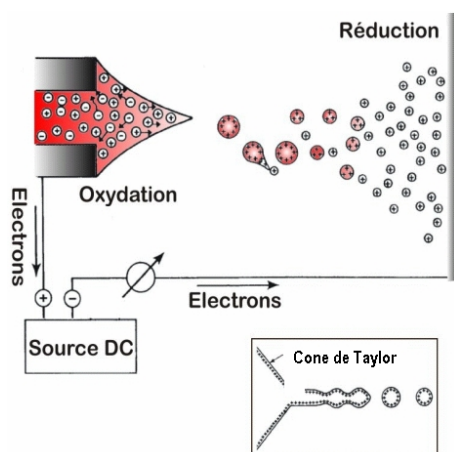


Fig. VI.1 : Formation des gouttelettes en electrospray, adapté de [2].

Lorsqu'on applique un champ électrique, la surface air/liquide présente une forme quasi-conique (Fig. VI.1). Cette déformation résulte de l'équilibre entre les forces électrostatiques et la tension de surface. En parallèle, une migration des charges se produit dans le liquide au niveau de l'extrémité du capillaire, ce qui est dû à un mouvement électrophorétique des électrolytes. Ainsi, en mode dit « positif », les cations migrent vers l'extrémité du capillaire et la surface du liquide présente une charge nette positive, alors que les anions se dirigent vers l'intérieur du liquide (Fig. VI.1). Quand les forces électrostatiques et la tension de surface s'équilibrent, la surface est conique, et nommée « cône de Taylor » [3]. Une légère augmentation du potentiel appliqué déstabilise le cône (Fig. VI.1, en bas), puisque la tension de surface devient trop faible pour conserver la stabilité de l'interface air/liquide, et celle-ci se rompt à l'extrémité du cône, où le champ électrique est maximal. La solution se disperse en gouttelettes, caractérisées par une charge nette positive (négative en mode d'ionisation négatif).

VI.1.1.1.2. Evaporation du solvant et explosions coulombiennes

Lors du trajet des gouttelettes entre l'extrémité du cône de Taylor et l'entrée de l'analyseur, leur taille diminue suite à l'évaporation du solvant. L'énergie nécessaire à celle-ci

est fournie par l'énergie thermique de l'air ambiant, ou par l'utilisation d'un gaz de nébulisation chauffé. Puisque la charge nette de la gouttelette reste constante pendant ce processus d'évaporation, la densité de charge augmente. Lorsque la densité de charge de surface approche la tension de surface, on tend vers la limite de Rayleigh [4], qui est définie comme le nombre de charges à la surface d'une gouttelette au-delà duquel les forces électrostatiques surpassent la tension de surface. Lorsque la limite de Rayleigh est atteinte, il y a fission de la gouttelette (ou « explosion coulombienne ») en gouttelettes de plus petites tailles. Ces nouvelles gouttelettes sont de nouveau soumises à l'évaporation du solvant jusqu'à ce que la limite de Rayleigh soit atteinte de nouveau, etc. Ce processus se répète jusqu'à ce que toutes les molécules de solvant entourant l'ion soient éliminées.

VI.1.1.2. Le nanospray

Certains échantillons, et en particulier des échantillons biologiques, ne sont disponibles qu'en très faibles quantités, de l'ordre de quelques picomoles. De ce fait, il est devenu nécessaire de diminuer le volume d'échantillon nécessaire et d'augmenter la sensibilité des sources electrospray. Des sources electrospray miniaturisées ont donc été développées : ce sont Wilm et Mann qui ont introduit la technique nanospray [5]. Cette technique consiste à introduire quelques microlitres de la solution à analyser dans une aiguille en verre recouverte d'un matériau conducteur. L'orifice de sortie de l'aiguille a un diamètre de 1 à 2 μm . On applique ensuite un potentiel de l'ordre de 1 kV sur l'aiguille, ce qui crée un champ électrique intense entre celle-ci et la contre électrode. Le processus de formation des ions est globalement le même que pour le processus electrospray « classique ».

Le nanospray a comme principal avantage de ne consommer que très peu d'échantillon : le débit, dans la plupart des configurations sans assistance pneumatique est très faible (de l'ordre de 20-100 nL/min), et quelques microlitres d'échantillon à une concentration de l'ordre du micromolaire permettent ainsi de travailler pendant une vingtaine de minutes. Cet avantage est particulièrement important dans le cas de l'analyse de protéines, où les quantités sont souvent limitées.

La technique du nanospray s'est donc imposée comme une méthode de choix dans l'analyse de biomolécules, malgré certains inconvénients, principalement des problèmes de reproductibilité liés à la grande sensibilité de la méthode vis à vis des conditions expérimentales [6, 7]. Par exemple, pour les études des interactions non covalentes, la tension appliquée sur l'échantillon (et ainsi l'énergie cinétique du complexe) a souvent une influence très importante sur la stabilité du complexe. Or, il est difficile de contrôler avec certitude la

position exacte de l’aiguille nanospray et donc le champ électrique imposé à l’échantillon. Pour nos études d’interactions non covalentes en conditions non dénaturantes, nous avons donc choisi de travailler en ionisation electrospray, où il est plus facile de contrôler ces paramètres.

VI.1.2. Spectrométrie de masse FT-ICR

Plusieurs types de spectromètres de masse existent et parmi ceux-ci nous avons utilisé en particulier dans ce travail la résonance cyclotronique ionique à transformée de Fourier (*Fourier-Transform Ion Cyclotron Resonance*, FT-ICR). Cette technique présente des qualités particulières par la très haute résolution et la précision en masse qui peuvent être atteintes [8]. Sans entrer dans les détails de la technique (pour une revue plus complète, voir [9, 10]), les bases physique et les principes fondamentaux sont rappelés ci-dessous.

La cellule d’analyse est localisée dans un champ magnétique intense (de l’ordre de 7T au laboratoire) généré par une bobine supraconductrice. Sous l’effet de ce champ magnétique, les ions présents dans la cellule FT-ICR ayant une composante de vitesse perpendiculaire aux axes du champ subissent la force de Lorentz (équation VI.1) :

$$\vec{F} = q\vec{v} \times \vec{B} \quad (\text{eq. VI.1})$$

L’ion est ainsi dévié suivant un arc de cercle. Si le champ magnétique est suffisamment intense ou suffisamment étendu, l’ion peut se retrouver sur une trajectoire circulaire (dans le plan xy) ce qui définit un mouvement cyclotron. (Fig. VI.2).

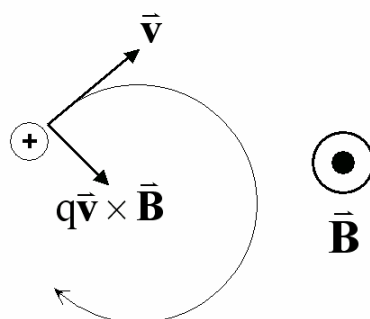


Fig. VI.2 : Mouvement cyclotron d’un ion dans un champ magnétique.

Le mouvement cyclotron est caractérisé par la vitesse angulaire, ω_c ou la fréquence cyclotron ν_c . En absence de perturbations externes de la trajectoire, ω_c se calcule en considérant la force de Lorentz égale à la force centrifuge (équation VI.2) :

$$qvB = \frac{mv^2}{r} \quad (\text{eq. VI.2})$$

Comme $\omega_c = v/r$, on obtient alors l'expression pour ω_c et ν_c (équation VI.3) :

$$\nu_c = \frac{\omega_c}{2\pi} = \frac{eB}{m/z} = \frac{1,535611 \cdot 10^7 B}{m/z} \quad (\text{eq. VI.3})$$

où dans la partie gauche, B est en teslas et m en u.m.a. Le rapport m/z d'un paquet d'ions est donc inversement proportionnel à la fréquence cyclotronique. La détermination des fréquences cyclotroniques de tous les ions dans une cellule ICR permet donc de déterminer les masses des analytes présents.

Comment la fréquence cyclotronique est-elle déterminée ? Une première étape consiste à amener les ions produits dans une source externe jusque dans la cellule par l'intermédiaire d'une optique de transfert. Cette séparation permet d'éloigner la source du champ magnétique, mais aussi de permettre un pompage efficace pour passer d'une ionisation à pression atmosphérique à un vide de 10^{-10} mbar dans la cellule. Ce transport s'achève en piégeant les ions dans la cellule par l'application d'un potentiel de piégeage suffisant (Fig. VI.3). Le mouvement des ions est ainsi restreint par le champ magnétique dans le plan xy et par les plaques de piégeage dans l'axe z.

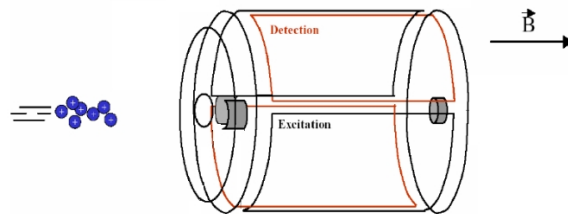


Fig. VI.3 : Représentation schématique d'une cellule ICR cylindrique. Les plaques d'excitation sont représentées en noir, les plaques de détection en rouge. Les plaques de piégeage se trouvent à gauche et à droite.

Le mouvement des ions immédiatement après piégeage est non cohérent et les ions sont situés à proximité de l'axe central de la cellule. Dans cette situation, il n'est pas possible de détecter directement un signal. Pour rendre la détection possible, il faut passer par une séquence excitation-détection. L'étape d'excitation s'effectue par l'irradiation des ions piégés dans la cellule par une radiation électromagnétique via des plaques d'excitation (voir Fig. VI.4). A la fréquence cyclotron d'ions de même m/z , cette irradiation conduit à une augmentation et une mise en phase des mouvements de ces ions. Il faut noter qu'une excitation trop intense ou trop longue aboutit à l'éjection des ions (par collision des ions avec les plaques), ce qui permet dans des conditions expérimentales contrôlées de sélectionner un ion avec un rapport m/z donné en éjectant tous les autres.

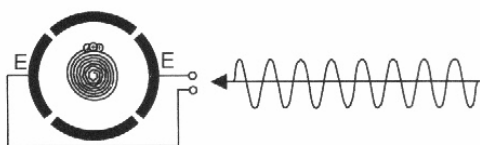


Fig. VI.4 : Excitation des ions piégés dans la cellule ICR (vu parallèlement à l'axe z).

Le mouvement des ions ainsi excités induit ensuite un courant dans les plaques de détection (voir Fig. VI.5). L'évolution temporelle de ce courant est enregistrée par l'électronique du système. Mathématiquement il s'agit d'une sinusoïde qui présente une décroissance exponentielle. La sinusoïde s'explique par le mouvement circulaire des ions qui s'approchent et s'éloignent périodiquement des plaques de détection. La décroissance du signal est une conséquence de la perturbation de la trajectoire des ions dans la cellule par des collisions successives avec le gaz résiduel. Le signal ainsi enregistré est appelé *FID* (« Free Induction Decay »).

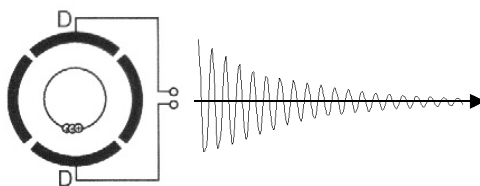


Fig. VI.5 : Détection des ions piégés dans la cellule ICR après excitation (vu parallèlement sur l'axe z).

Pour remonter au spectre de masse, il faut d'abord effectuer une transformation de Fourier sur le FID : ceci permet de convertir un signal temporel en signal dans le domaine des fréquences. Concrètement, les composantes de fréquences présentes dans le signal auront une intensité, qui dépendra de l'amplitude de ces fréquences donc du nombre de charges présentes dans le paquet d'ions. Le spectre de masse (intensité en fonction de m/z) est ensuite calculé par simple application de l'équation VI.3.

VI.1.3. Etude de complexes non covalents par spectrométrie de masse

Suite au développement de l'electrospray et du MALDI, qui ont permis l'analyse de protéines par spectrométrie de masse sans induire de fragmentations, des groupes de recherche se sont également interrogés sur l'intérêt de ces méthodes d'ionisation douces pour l'analyse de complexes non covalents. Les premiers à avoir publié des résultats dans ce domaine sont Ganem *et al.* [11]. Depuis, de nombreuses applications de la spectrométrie de masse electrospray et MALDI à l'étude de complexes non covalents ont été publiées, et plusieurs revues ont rapporté l'état de l'art dans le domaine [12-17]. Les complexes étudiés peuvent être des complexes entre petites molécules, entre protéines et petites molécules, entre protéines, ou entre protéines et acides nucléiques. Il est impossible de faire ici une revue exhaustive de tout ce qui a été fait dans ce domaine. Néanmoins, deux thématiques particulières concernant l'étude des complexes non covalents par spectrométrie de masse seront traitées plus en détail dans la suite : l'analyse des complexes protéine-ligand et l'analyse des complexes protéiques.

Ces résultats ne doivent pas faire oublier qu'il faut prendre de grandes précautions lorsqu'on s'intéresse aux complexes non covalents. Il s'agit en effet d'étudier *en phase gazeuse* des complexes existant initialement en solution, et les conditions de préparation des échantillons, les paramètres expérimentaux ainsi qu'une analyse critique des résultats jouent tous un rôle déterminant dans la réussite de l'expérience et la validité des conclusions qui en seront tirées.

Avant de discuter de l'analyse des complexes protéine-ligand et protéine-protéine par spectrométrie de masse, il est donc nécessaire de comprendre les deux étapes principales concernant l'expérience : la préparation de l'échantillon et l'optimisation des paramètres instrumentaux.

VI.1.3.1. Conditions nécessaires à l'observation d'interactions non covalentes par spectrométrie de masse

La nature même des interactions non covalentes rend leur observation par spectrométrie de masse difficile, même lorsque des méthodes d'ionisation dites « douces » sont utilisées. Il est indispensable d'utiliser des conditions expérimentales adéquates, de manière tout d'abord à conserver les interactions au sein de la solution à analyser (d'où la nécessité d'un choix judicieux de solvant), et ensuite à les conserver lors du passage en phase gazeuse. Dans ces conditions, il est toutefois capital de garder à l'esprit que certains types d'interactions non covalentes ne seront que très exceptionnellement maintenues en phase gazeuse : c'est notamment le cas des interactions hydrophobes.

Même si l'on trouve dans la littérature quelques exemples d'utilisation de l'ionisation MALDI pour étudier ce type de complexes [15, 18-22], nous nous focaliserons sur l'electrospray, qui reste la méthode de référence pour ce type d'études. En effet, au-delà de la nature même de l'échantillon (analytes co-cristallisés dans une matrice le plus souvent acide, ce qui implique une importante concentration de l'échantillon) et de l'énergie interne relativement importante conférée aux ions par l'ionisation MALDI, le fait que ce mode d'ionisation ait tendance à former des agrégats rend difficile la détermination de la spécificité des interactions observées. De plus, d'un point vue pratique, les ions en MALDI sont monochargés, ce qui rend difficile la détection des complexes non covalents ayant des masses moléculaires élevées sur des appareils limités dans leur gamme de m/z .

VI.1.3.1.1. Préparation des échantillons

La préparation des échantillons est sans doute l'étape la plus délicate en spectrométrie de masse de complexes non covalents. En effet, l'utilisation de l'electrospray impose des contraintes, principalement au niveau de la volatilité des solvants ou tampons utilisés pour solubiliser les protéines et de l'absence de sels. D'autre part, si l'on veut conserver les protéines dans leur conformation native, il est nécessaire d'utiliser des solutions aqueuses tamponnées à un pH physiologique (en général autour de 7), sans trace de solvants organiques.

C'est en raison de ces deux contraintes que l'analyse des complexes non covalents peut poser des problèmes : en effet, les tampons les plus couramment utilisés en biochimie contiennent des quantités non négligeables de cations alcalins (sodium ou potassium) ou d'ions phosphate. Ces espèces ne sont pas compatibles avec l'analyse des protéines par electrospray. Il est nécessaire d'utiliser dans ce cas des tampons volatils, les solutions de

bicarbonate d'ammonium ou d'acétate d'ammonium étant les tampons les plus courants. Dans le cas où l'échantillon initial se trouve dans un tampon inadéquat, il est nécessaire de le dessaler ou de le transférer dans un tampon compatible avec la spectrométrie de masse. On peut noter que certaines protéines ne conservent leur structure que grâce à la présence de sels non volatils, ou grâce à une force ionique du milieu très élevée, si bien qu'il sera parfois impossible de trouver des tampons permettant leur analyse par spectrométrie de masse.

L'influence du tampon sur les phénomènes en phase gazeuse, et plus particulièrement sur la protonation des protéines, a aussi été démontrée par Lemaire *et al.* [23] : en présence de bicarbonate de triéthylammonium (TEAB), on observe un déplacement de la distribution des états de charge vers les hautes valeurs de m/z (donc vers les faibles états de charge). Pour comprendre cette influence du cation, le paramètre à prendre en compte est la différence de basicité en phase gazeuse de la triéthylamine par rapport à l'ammoniaque. Le fait d'utiliser une base plus forte (la triéthylamine) permet de diminuer l'état de charge moyen de la protéine. Cette diminution de l'état de charge permet ainsi d'expliquer la plus grande stabilité des ions formés : d'une part, l'énergie cinétique des ions peu chargés est plus faible, ce qui implique une plus faible énergie transférée à l'ion lors des collisions avec le gaz résiduel, d'autre part, les ions portant peu de charges sont moins soumis aux répulsions coulombiennes intramoléculaires susceptibles de modifier la structure tertiaire de la protéine. La contrepartie de l'utilisation de ce type de tampons est la difficulté de l'analyse des ions de hautes valeurs de m/z , qui demande généralement une adaptation des conditions expérimentales.

VI.1.3.1.2. *Optimisation des paramètres instrumentaux*

Nébulisation de l'échantillon

L'utilisation de tampons aqueux pour solubiliser l'échantillon rend la nébulisation plus difficile, principalement en raison de la tension de surface importante de l'eau comparée à celle des solvants couramment utilisés tels que le méthanol ou l'acétonitrile. Il est donc nécessaire d'adapter les paramètres en conséquence : on modifiera principalement (selon les appareils) la température de la source, les débits des gaz de nébulisation et de séchage, et le débit de la solution. L'optimisation de ces paramètres est parfois complexe, car il faut favoriser la désolvatation, tout en maintenant l'intégrité des complexes non covalents lors des collisions entre l'analyte et les autres molécules présentes. Il faut donc optimiser l'énergie interne des ions pour avoir un bon compromis entre ces deux objectifs contradictoires.

Le transfert des espèces chargées en phase gazeuse

Le problème suivant dans l'analyse de complexes non covalents est le maintien des interactions non covalentes en phase gazeuse. Ce maintien va être fonction de divers paramètres instrumentaux :

- Tension d'accélération à l'interface : dans notre appareil FT-ICR, cette tension correspond à la différence de potentiel entre la sortie du capillaire de désolvatation et le premier *skimmer* (CapExit). Cette accélération se situe à la sortie du capillaire de désolvatation, dans une zone de l'instrument où la pression de gaz résiduel est encore relativement élevée. Le transfert de l'énergie cinétique apportée en énergie interne *via* des collisions avec le gaz résiduel permet un dernier chauffage collisionnel des ions. Un minimum d'énergie est nécessaire à la désolvatation, mais une quantité trop importante va apporter à l'ion un excès d'énergie interne pouvant conduire à une dissociation du système non covalent.
- Pression dans l'interface : la pression dans l'interface va influencer le libre parcours des ions dans cette zone, et donc le nombre de collisions. Des études ont mis en évidence le rôle positif de l'élévation de la pression sur la désolvatation des espèces, mais aussi sur le maintien des interactions non covalentes [24, 25]. Sur l'appareil utilisé lors de ces études (Bruker APEX III FT-ICR), il est cependant difficile de faire varier la pression de façon simple. Néanmoins, il faut tenir compte de l'effet de la pression pour mieux comprendre le comportement des complexes lors du passage en phase gazeuse.
- Transmission des ions de hauts rapports m/z : l'ionisation electrospray en conditions non dénaturantes génère, comme nous l'avons vu précédemment, des espèces de rapport m/z élevé, ce qui pose un problème de transmission de ces ions. Dans le cas d'un FT-ICR, il est surtout important de synchroniser le temps de vol des ions dans l'optique de transfert et « l'ouverture » de la cellule (*gated-trapping delay*) pour optimiser le piégeage des ions de rapport m/z élevé.

VI.1.3.1.3. Analyse des résultats

Les espèces observées par spectrométrie de masse sont des espèces en phase gazeuse. Toute la question lorsqu'on traite de problèmes d'ordre biologique est de savoir si les informations spectrales reflètent la situation en solution. Ce problème se pose au niveau de la structure tertiaire et quaternaire des protéines, ainsi qu'au niveau des informations

quantitatives (calculs de constantes d'interaction par exemple) qui peuvent être tirées des spectres. Ces problèmes seront discutés dans les sections suivantes, où l'on se focalisera tout d'abord sur l'étude des complexes protéine-ligand et ensuite sur l'étude des complexes protéiques.

VI.2. Méthodes expérimentales générales

VI.2.1. Spectrométrie de masse

La plupart des expériences de spectrométrie de masse (à part les expériences réalisées sur le Q-TOF et le triple quadripôle Quattro II) ont été effectuées sur un FT-ICR Bruker APEX III équipé d'un aimant supraconducteur de 7 T et d'une source Apollo (electrospray et nanospray). Les spectres ont été retraités et analysés par des logiciels fournis par le constructeur (Xmass, BioTools, DataAnalysis).

VI.2.2. Dessalage par microcolonnes C₁₈

Les microcolonnes ont été conditionnées 4 fois avec de l'acétonitrile, puis équilibrées 4 fois avec de l'acide formique 0,1%. L'échantillon est chargé par aspiration – refoulement (10 fois). Les sels sont lavés 3 fois avec de l'acide formique 0,1%. Les peptides sont finalement élués avec un mélange eau :méthanol :acide formique 49,9 :49,9 :0,2 (v/v/v).

VI.3. Modification chimique par le DEPC

La modification chimique par le DEPC de peptides ou de protéines a pour but la carbéthoxylation spécifique de certains résidus accessibles à la surface de la molécule. Cette modification induira un ajout de masse de 72,02 Da par acide aminé modifié, détectable par spectrométrie de masse. Nous présenterons ici les différents protocoles que nous avons utilisés pour la mise au point de cette méthode. De façon générale, on peut dire qu'il n'existe pas un protocole pour cette méthode, mais plusieurs protocoles qui seront adaptés aux conditions d'analyse utilisées (« bottom-up » ou « top-down ») et à l'échantillon (peptides ou protéines).

VI.3.1. Modification d'un mélange de peptides par le DEPC

La réaction de modification a été réalisée sur une solution de 8 μM de digestat de cytochrome c dans 10 mM NH_4HCO_3 (pH 7,9). La réaction a été initiée par addition de 1 μL de DEPC (60 mM dans EtOH), ce qui correspond à un rapport peptide : réactif de 1 :750. La réaction de modification s'est déroulée pendant 5 minutes à température ambiante. Ensuite, le surplus de réactif a été éliminé par dessalage avec des microcolonnes C_{18} (Agilent). Les peptides ont été analysés par nanospray dans des conditions standard.

VI.3.2. Modification de la myoglobine

10 μM myoglobine (concentration finale) dans 10 mM NH_4HCO_3 (pH 7,9) ont été incubés avec 100 μM DEPC (concentration finale) pendant 5 minutes à 0°C (sur la glace). La réaction a été arrêtée en ajoutant de l'histidine à une concentration finale de 7,5 mM. La moitié de l'échantillon a été analysée par ESI-MS sur un triple quadripôle (Quattro II, Micromass). L'autre moitié et un aliquot de la myoglobine non modifiée ont été digérés par la trypsine ou par l'endoprotéase GluC (V8) et analysés par nano-ESI-FT-ICR MS dans des conditions standard.

VI.3.3. Modification de l'ubiquitine

2 μL d'une solution de 100 μM d'ubiquitine ont été dilués dans 7 μL 10 mM NH_4HCO_3 (pH 7,9) et incubés avec 1 μL 200 mM DEPC (dans MeOH) pendant 5 minutes à température ambiante. Le mélange est ensuite dessalé sur des microcolonnes C_{18} (Agilent) et analysé par nano-ESI-FTMS.

Après sélection des pics correspondant à l'état de charge 8+, la protéine a été irradiée par un faisceau d'électrons d'une énergie de 1-2 eV pendant 50 ms. 200 scans ont été accumulés.

VI.3.4. Modification de la créatine kinase

5 μM CK dans 10 mM NH_4HCO_3 ont été chauffés pendant 1 minute à 60°C afin de casser les interactions entre les monomères. L'échantillon a été analysé par injection directe en ESI-FT-ICR-MS dans des conditions standard.

Pour la modification par le DEPC, 99 μL 20 μM CK (dimère ou monomère) dans 10 mM NH_4HCO_3 (pH 7,9) ont été incubés avec 1 μL DEPC 60 mM (dans MeOH) pendant 5 minutes à 0°C (sur la glace). Les échantillons ont été dessalés par des colonnes de filtration sur gel Bio-Spin P30 (BioRad) ayant une taille d'exclusion de 30 kDa (pré-équilibrées avec 10 mM

NH₄HCO₃) en suivant le protocole du fournisseur, et une moitié de chaque échantillon a été analysée par injection directe en ESI-FT-ICR-MS.

L'autre moitié a été digérée par la trypsine dans des conditions standard et analysée par injection directe en nano-ESI-FT-ICR-MS.

VI.4. Echanges hydrogène/deutérium

VI.4.1. Calcul des hydrogènes échangés

Pour chaque peptide observé dans le spectre, le nombre moyen de deutériums incorporés a été calculé comme étant la différence entre la masse moyenne du peptide ayant échangé des hydrogènes et celle du peptide natif (équation VI.4) :

$$\Delta\bar{m}(t) = \bar{m}(t) - \bar{m}(t = 0) \quad (\text{eq. VI.4})$$

Dans cette equation, $\bar{m}$ est la masse moyenne du peptide, calculée d'après les intensités des pics du massif isotopique observé dans le spectre (équation VI.5) :

$$\bar{m} = \frac{\sum_i I_i m_i}{\sum_i I_i} \quad (\text{eq. VI.5})$$

Dans cette équation, m_i est la masse de l'isotope i et I_i est l'intensité correspondante. Le pourcentage d'échange est donné par le rapport des valeurs de $\Delta\bar{m}$ sur le nombre d'hydrogènes potentiellement échangeables présents sur le peptide (équation VI.6) :

$$\%D(t) = \frac{\Delta\bar{m}(t)}{n_{AA} - n_{P\neq PI} - 1} \cdot 100 \quad (\text{eq. VI.6})$$

Dans cette équation, n_{AA} est le nombre des acides aminés du peptide et $n_{P\neq PI}$ est le nombre de prolines (qui n'ont pas d'hydrogènes d'amides échangeables).

VI.4.2. Echange H/D substance P

4 μL D_2O ont été ajoutés à 1 μL 1 mM substance P dans H_2O . Le mélange a été incubé pendant 10 minutes à température ambiante. Ensuite, 45 μL d'une solution d'acide formique 0,1% ont été ajoutés afin d'arrêter l'échange. Pour l'analyse par électrospray, 50 μL MeOH/acide formique 4% ont été ajoutés et l'échantillon a été analysé par injection directe. Pour l'analyse par nanospray, l'échantillon a tout d'abord été dessalé par une microcolonne C_{18} (Agilent) pour simuler des conditions expérimentales réelles et a ensuite été analysé par injection directe avec des aiguilles nanospray (Proxeon, Denmark).

VI.4.3. Echange H/D myoglobine

4 μL D_2O ont été ajoutés à 1 μL 100 μM myoglobine dans 10 mM NH_4HCO_3 (pH 7,9). Le mélange a été incubé pendant des temps variables à température ambiante et l'échange H/D a été arrêté par ajout de 4 μL d'acide formique 4% froid (gardé sur la glace à 0°C). Ensuite, 1 μL de pepsine 40 μM dans H_2O a été ajouté et l'échantillon a été incubé pendant 5 minutes à 0°C (sur la glace). Les peptides résultant de la digestion ont été dessalés sur une microcolonne C_{18} (Agilent), élués dans H_2O /MeOH/acide formique 49,9 : 49,9 : 0,2 (v/v/v) et analysés par injection directe par nano-ESI-FTMS.

VI.5. Oligomères de la protéine prion

VI.5.1. Formation des oligomères

12 mg de PrP (variant ARQ) ont été dissous dans 15 mM NaAcétate (pH 4,7) puis transférés dans 20 mM NaCitrate pH 3,38 par chromatographie d'exclusion stérique grâce à une colonne G25. La solution résultante a été utilisée pour les expériences suivantes.

Pour obtenir un échantillon contenant majoritairement l'oligomère P1, 45 μM de la solution de PrP dans 20 mM NaCitrate pH 3,38 ont été incubés à 55°C pendant 12h puis analysés par chromatographie d'exclusion stérique grâce à une colonne TSK 4000SW équilibrée avec 20 mM NaCitrate pH 3,38.

Pour obtenir un échantillon contenant majoritairement l'oligomère P3, 120 μM de la solution de PrP dans 20 mM NaCitrate pH 3,38 ont été incubés à 50°C pendant 10 minutes puis analysés par chromatographie d'exclusion stérique grâce à une colonne TSK 4000SW équilibrée avec 20 mM NaCitrate pH 3,38.

VI.5.2. Chromatographie d'exclusion stérique

La chromatographie d'exclusion stérique a été réalisée sur une chaîne Akta FPLC (Pharmacia) grâce à une colonne TSK4000SW (7 mm x 600 mm, Interchim). La colonne a été calibrée pour des masses moléculaires ainsi que des rayons de Stokes en utilisant les kits de calibration pour les basses et hautes masses moléculaires (Amersham). Avant chaque expérience, la colonne a été équilibrée avec au moins quatre volumes du tampon d'élution. Les expériences ont été effectuées à 20°C. Le débit a été fixé à 1 mL/min et l'élution de la protéine a été suivie par absorption UV à 220 nm et 280 nm.

VI.5.3. Purification des oligomères en TEAA et lyophilisation

L'oligomère P3 a été formé comme indiqué dans le paragraphe VI.5.1. La purification a été réalisée sur une colonne TSK4000 équilibrée avec 50 mM TEAA pH 3,38. Les fractions éluées entre 14,7 mL et 16,0 mL ont été collectées. La concentration était de l'ordre de 5-10 μ M suivant l'expérience. Les échantillons ont ensuite été lyophilisés et conservés à -20°C.

VI.5.4. Echanges H/D sur le monomère et les oligomères de la protéine prion

VI.5.4.1. Remarques générales

La concentration des oligomères après la purification, étape primordiale pour étudier les différentes espèces en détail, était de l'ordre de 5-15 μ M (équivalents monomériques). L'approche « classique » des échanges H/D, qui consiste en une dilution de la solution de protéine initiale dans D₂O, suivie par une étape d'arrêt de la réaction par dilution dans un tampon à pH acide, n'a pas donné des résultats satisfaisants. Il n'a pas été possible de reconcentrer suffisamment les échantillons pour obtenir un spectre d'intensité suffisante. Pour éviter une dilution aussi importante de l'échantillon, nous l'avons transféré directement dans un tampon deutéré par chromatographie d'exclusion stérique par des colonnes G-25 (Fig. VI.6).

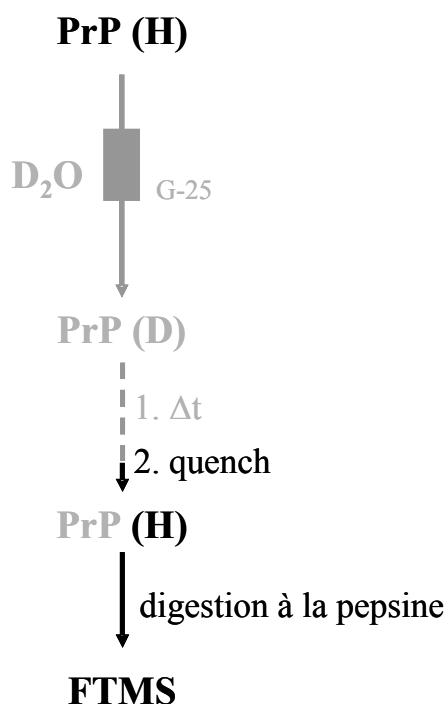


Fig. VI.6 : Principe des échanges hydrogène/deutérium utilisés pour des solutions avec des faibles concentrations initiales. Les étapes dans lesquelles la protéine est en présence de deutérium sont indiquées en gris.

En utilisant cette méthode, environ 50% de la protéine est retenue sur la colonne (mesuré par spectrophotométrie), diminuant la concentration initiale de 50%. Avec l'approche classique, la concentration de l'échantillon dans le tampon deutéré est seulement de 10% de la concentration initiale. Le deuxième avantage de cette approche est l'utilisation du tampon deutéré à la même concentration en sel que le tampon initial. Ainsi, la force ionique ne change pas lors du transfert des oligomères dans le tampon deutéré, éliminant la possibilité d'une déstabilisation des oligomères à cause d'un changement de force ionique.

VI.5.4.2. Méthode détaillée

100 μL d'une solution de 5-15 μM PrP variant ARQ (équivalent en monomère, les concentrations varient légèrement suivant la purification) dans 20 mM NaCitrate pH 3,38 ont été transférés dans 20 mM NaCitrate pD 3,00 à l'aide d'une colonne G-25 (MicroSpin G-25 columns, Amersham Biosciences). La colonne a été équilibrée avec six fois 400 μL 20 mM NaCitrate pD 3,00 à froid. La protéine dans le tampon deutéré a ensuite été incubée à 15°C pendant des durées variables. 20 μL d'acide formique 4% (à 4°C) ont alors été ajoutés à des aliquots de 20 μL d'échantillon afin de bloquer l'échange. 2 μL d'une solution de pepsine à

40 μM dans 20 mM NaCitrate pH 3,38 ont ensuite été ajoutés et l'échantillon a été digéré pendant 5 minutes sur la glace (0°C). Les peptides résultant ont été dessalés et concentrés par des microcolonnes C_{18} (Agilent) par élution dans 4 μL $\text{H}_2\text{O}/\text{MeOH}/\text{acide formique}$ 49,9 : 49,9 : 0,2 (v/v/v). Cette étape a été réalisée à température ambiante afin d'augmenter les interactions hydrophobes entre les peptides et la colonne C_{18} . Immédiatement après dessalage, l'échantillon a été analysé par nano-ESI-FTMS. Pour analyser le plus rapidement possible les échantillons après échanges H/D afin de minimiser le ré-échange, il n'a pas été possible d'optimiser la position de l'aiguille de nanospray et seulement 50 scans ont été enregistrés. Il est ainsi très probable que des peptides peu abondants dans l'échantillon n'aient pas pu être détectés à cause d'une trop faible intensité dans le spectre. Il est important de noter que tous les peptides présentent entre 2 et 8% d'hydrogènes échangés immédiatement après transfert dans le tampon deutéré. Il est probable que cet échange initial, qui semble être intrinsèque pour tous les peptides, se fasse lors de l'étape de transfert.

VI.6. Bibliographie

- [1] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* **1989**, 246, 64-71.
- [2] F. Lottspeich, H. Zorbas (eds.), *Bioanalytik*. Spektrum Akademischer Verlag, Heidelberg (1998).
- [3] G. I. Taylor, Disintegration of water drops in an electric field, *Proc. R. Soc. London A* **1964**, A280, 383-397.
- [4] Lord Rayleigh, On the equilibrium of liquid conducting masses charged with electricity, *Phil. Mag.* **1882**, 14, 184-186.
- [5] M. S. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source, *Anal. Chem.* **1996**, 68, 1-8.
- [6] V. Gabelica, C. Vreuls, P. Filée, V. Duval, B. Joris, E. DePauw, Advantages and drawbacks of nanospray for studying noncovalent protein-DNA complexes by mass spectrometry, *Rapid Commun. Mass Spectrom.* **2002**, 16, 1723-1728.
- [7] Y. Li, R. B. Cole, Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **2003**, 73, 5739-5746.
- [8] M. B. Comisarow, A. G. Marshall, Fourier transform ion cyclotron resonance spectroscopy, *Chem. Phys. Lett.* **1974**, 26, 489-490.

- [9] A. G. Marshall, C. L. Hendrickson, G. S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer, *Mass Spectrom. Rev.* **1998**, *17*, 1-35.
- [10] M. P. Barrow, W. I. Burkitt, P. J. Derrick, Principles of Fourier transform ion cyclotron resonance mass spectrometry and its applications in structural biology, *Analyst* **2005**, *130*, 18-28.
- [11] B. Ganem, Y.-T. Li, J. D. Henion, Detection of noncovalent receptor-ligand complexes by mass spectrometry, *J. Am. Chem. Soc.* **1991**, *113*, 6294-6296.
- [12] J. A. Loo, Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes, *Int. J. Mass Spectrom.* **2000**, *200*, 175-186.
- [13] T. D. Veenstra, Electrospray ionization mass spectrometry in the study of biomolecular non covalent interactions, *Biophys. Chem.* **1999**, *79*, 63-79.
- [14] B. Ganem, J. D. Henion, Going gently into flight: analyzing noncovalent interactions by mass spectrometry, *Bioorg. Med. Chem.* **2003**, *11*, 311-314.
- [15] G. Bolbach, Matrix-assisted laser desorption/ionization analysis of non covalent complexes: fundamentals and applications, *Curr. Pharm. Des.* **2005**, *11*, 2535-2557.
- [16] B. N. Pramanik, P. L. Bartner, U. A. Mirza, Y. H. Liu, A. K. Ganguly, Electrospray ionization mass spectrometry for the study of non covalent complexes: an emerging technology, *J. Mass Spectrom.* **1998**, *33*, 911-920.
- [17] J. A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev.* **1997**, *16*, 1-23.
- [18] A. S. Woods, J. C. Buchsbaum, T. A. Worrall, J. M. Berg, R. J. Cotter, Matrix-assisted laser-desorption/ionization of noncovalently bound compounds, *Anal. Chem.* **1995**, *67*, 4462-4465.
- [19] A. M. Distler, J. Allison, C. Hiser, L. Qin, Y. Hilmi, S. Ferguson-Miller, Mass spectrometric detection of protein, lipid and heme components of cytochrome c oxidase from *R. sphaeroides* and the stabilization of non covalent complexes from the enzyme, *Eur. J. Mass Spectrom.* **2004**, *10*, 295-308.
- [20] T. B. Farmer, R. M. Caprioli, Determination of protein-protein interactions by matrix-assisted laser desorption/ionization mass spectrometry, *J. Mass Spectrom.* **1998**, *33*, 697-704.
- [21] S. Z. Luo, Y. M. Li, W. Qiang, Y. F. Zhao, H. Abe, T. Nemoto, X. R. Qin, H. Nakanishi, Detection of specific noncovalent interaction of peptide with DNA by MALDI-TOF, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 28-31.

- [22] C. Borchers, K. B. Tomer, Characterization of the noncovalent complex of human immunodeficiency virus glycoprotein 120 with its cellular receptor CD4 by matrix-assisted laser desorption/ionization mass spectrometry, *Biochemistry* **1999**, 38, 11734-11740.
- [23] D. Lemaire, G. Marie, L. Serani, O. Laprévote, Stabilization of gas-phase non covalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer, *Anal. Chem.* **2001**, 73, 1699-1706.
- [24] A. Schmidt, U. Bahr, M. Karas, Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in nano-ESI-MS, *Anal. Chem.* **2001**, 73, 6040-6046.
- [25] N. Tahallah, M. Pinkse, C. S. Maier, A. J. R. Heck, 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.

RESUME

Les qualités analytiques de la spectrométrie de masse à résonance cyclotronique ionique et à transformée de Fourier, combinée à une source d'électronébulisation (ESI-FT-ICR) ont été mises à profit pour l'étude d'interactions non-covalentes dans des assemblages protéiques. Dans une première partie, l'utilisation de cette technique pour la détection de complexes protéiques intacts a été développée sur le système constitué de la créatine kinase (CK) et de ses ligands ADP et ATP, conduisant à l'observation de ces complexes en phase gazeuse. Les parts spécifiques et non spécifiques de ces interactions ont été séparées par un modèle statistique, conduisant à la détermination des constantes de dissociation ($K_{1,ADP}=11.8 \pm 1.5 \mu\text{M}$, $K_{2,ADP}=48 \pm 6 \mu\text{M}$, $K_{1,ATP}=27 \pm 7 \mu\text{M}$, $K_{2,ATP}=114 \pm 27 \mu\text{M}$) qui sont en accord avec la littérature. Dans une seconde partie, des techniques pour l'étude de l'accessibilité de surface de complexes de protéines par modification chimique en solution, aussi bien réversibles (échanges H/D) qu'irréversibles (modification spécifique d'histidine et de lysine par le DEPC) ont été mises au point et appliquées. Les sites de modification ont été localisés soit par une approche « bottom-up » (digestion enzymatique suivie d'une mesure de masse des peptides résultants) soit par une approche « top-down » (fragmentation en phase gazeuse de la protéine intacte par dissociation par capture d'électrons [ECD]). Dans le cadre de l'instrumentation actuelle, cette dernière approche a conduit à des résultats pour des protéines allant jusqu'à 17 kDa et pourrait être étendue jusqu'à des protéines de 25 kDa. Ces différentes stratégies analytiques ont été appliquées à des oligomères de la protéine du prion (PrP) obtenus par dénaturation thermique partielle *in vitro* de l'espèce monomérique. L'accroissement de l'incorporation de deutérium dans certaines régions de la forme oligomérique (par rapport au monomère) reflète un dépliement partiel de la protéine, en particulier au niveau d'une hélice α .

ABSTRACT

The outstanding analytical performances of electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR MS) were applied to study non-covalent interactions in protein complexes. In a first part, the application of this technique to the detection of intact protein complexes was developed on the system of Creatine kinase (CK) and its ligands ADP and ATP, leading to the observation of these complexes in the gas phase. Specific and non-specific contributions of these interaction were separated by means of a statistical model, yielding interaction constants ($K_{1,ADP}=11.8 \pm 1.5 \mu\text{M}$, $K_{2,ADP}=48 \pm 6 \mu\text{M}$, $K_{1,ATP}=27 \pm 7 \mu\text{M}$, $K_{2,ATP}=114 \pm 27 \mu\text{M}$) that are in fair agreement with literature values. In a second part, methods to study the surface accessibility of protein complexes by chemical modification in solution, either reversible (H/D exchange) or irreversible (specific modification of His and Lys by DEPC) were developed and implemented. The modification site was identified either by a bottom-up (enzymatic digestion followed by mass measurement of the resulting peptides) or a top-down approach (gas phase fragmentation of the entire protein by electron capture dissociation [ECD]). Within the framework of the current instrumentation, the latter approach yields reasonable results for proteins as large as 17 kDa and should be extendible to 25 kDa. Application of the different analytical strategies yielded insight into the structure and conformational dynamics of prion protein (PrP) oligomers that were obtained by partial thermal denaturation of the monomeric species *in vitro*. Increase in deuterium incorporation in some regions of the oligomeric species (compared to the monomer) reflects partial unfolding of the protein, as indicated by the unfolding of an α helix.