



**HAL**  
open science

# Étude du mécanisme d'action du protéasome PAN-20S par diffusion de neutrons aux petits angles résolue en temps

Emilie Mahieu

► **To cite this version:**

Emilie Mahieu. Étude du mécanisme d'action du protéasome PAN-20S par diffusion de neutrons aux petits angles résolue en temps. Biophysique. Université Grenoble Alpes, 2019. Français. NNT : 2019GREAY073 . tel-02930469

**HAL Id: tel-02930469**

**<https://theses.hal.science/tel-02930469>**

Submitted on 4 Sep 2020

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

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

## **THÈSE**

Pour obtenir le grade de

**DOCTEUR DE LA COMMUNAUTE UNIVERSITE  
GRENOBLE ALPES**

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 25 mai 2016

Présentée par

**Emilie MAHIEU**

Thèse dirigée par **Frank GABEL** et **Bruno FRANZETTI**

préparée au sein de l'**Institut de Biologie Structurale**  
**Groupe Extrêmophiles et Grands Assemblages Moléculaires**  
dans l'**École Doctorale de Physique**

# **Étude du mécanisme d'action du protéasome PAN-20S par diffusion de neutrons aux petits angles résolue en temps**

Thèse soutenue publiquement le **17 décembre 2019**,  
devant le jury composé de :

**Dr. Véronique RECEVEUR-BRECHOT**

Présidente du jury et rapportrice, Directrice de recherche CNRS  
Provence et Corse

**Dr. Didier FLAMENT**

Rapporteur, Directeur de recherche IFREMER Centre de Bretagne

**Dr. Dominique DURAND**

Examinatrice, Directrice de recherche CNRS Île de France Gif-sur-Yvette

**Dr. Véronique ROSSI**

Examinatrice, Maître de conférences Université Grenoble-Alpes





# Remerciements

Je remercie tout d'abord les membres du jury pour avoir accepté d'examiner ce mémoire et pour le nouveau regard qu'ils ont apporté sur mon travail.

Je tiens à remercier mon directeur de thèse, Frank Gabel, pour m'avoir toujours encouragée durant ces trois années de thèse et pour l'optimisme dont il a fait preuve tout au long ce projet malgré les hauts et les bas. Merci à Bruno Franzetti, mon co-directeur de thèse, d'avoir cru en moi dès mon premier stage chez ELMA et pour toutes les discussions scientifiques. Je les remercie tous deux pour la liberté qu'ils m'ont laissée, à la fois pour tester mes hypothèses au laboratoire et dans mes choix de formation.

Je tiens à remercier Jacques Covès pour tout le soutien inestimable qu'il m'a apporté, aussi bien au laboratoire que pendant la relecture de ce manuscrit. J'ai pu désacraliser l'enzymologie et, je l'espère, maîtriser la machine. Merci pour la bonne humeur, les blagues et les belles photos !

Je voudrais remercier les membres de mon comité de suivi de thèse, Anne Martel, Caroline Mas et Cécile Breyton, pour leurs conseils précieux.

Merci à Teresa Carlomagno pour les discussions scientifiques et à Georg Krüger pour les nombreux échanges de mail qui nous ont tous deux aidés à mieux comprendre nos protéines. Je remercie chaleureusement toutes les personnes avec qui j'ai eu la chance de travailler et qui m'ont aidée dans ce projet : Martine Moulin pour son encadrement pendant la deutération des protéines, Anne Martel pour son aide pendant les mesures SANS, et Nico Carl pour les heures passées à travailler sur les scripts d'analyse de données SANS. Je remercie également Caroline Mas pour le SEC-MALLS et l'ITC, Luca Signor pour la spectrométrie de masse, Aline Le-Roy pour l'AUC, Daphnal Fenel pour la microscopie électronique et Michel Thépaut pour les divers sauvetages du quotidien. Merci à Thibault, Marion, et Francesca. Cela m'a fait plaisir de partager ce projet avec vous ! J'espère avoir été une bonne encadrante, j'ai de mon côté beaucoup appris en travaillant avec vous.

Je remercie les membres de l'équipe l'ELMA, grâce à qui aller au travail a toujours été agréable. Merci à Alexandre, Domi, Éric, et Mylène, pour les conversations autour du café et la bonne humeur ambiante. Je voudrais remercier spécialement Alexandra pour le mois d'été passé toutes les deux à l'IBS, et Sylvain pour les cafés du matin de ma première année de thèse et les quelques jours en Irlande. J'ai eu la chance de partager au cours de ces trois années le bureau 463 avec des personnes qui ont rendu cet endroit chaleureux, plein de thé et de plantes pas toujours au meilleur de leur forme. Merci à Gaëlle pour sa gentillesse et tous ses conseils, Jennifer pour l'instauration des pauses banane, et Zaynab et Lorenzo pour les innombrables conversations nourriture. Merci à Tiger, la mascotte du bureau, pour sa sérénité inspirante pendant la rédaction.

Je souhaiterais remercier Quentin, Jean-Mathieu, Kévin, Daniel, Julie, Kyprianos, Laura et Ana pour tous les moments partagés au quotidien à l'IBS. Je remercie spécialement Quentin pour les soirées Scythe et les pauses passées ensemble pendant ces derniers mois de rédaction. Je remercie Sandra, Marie-Léonor, Éric, Nicolas et Edouard pour les soirées du mardi et les week-ends au Monteynard. Merci Vita, Aljoša, Juliette et Andrea, pour leur cuisine délicieuse et les nombreux verres bus ensembles. Je fais également un clin d'œil à Lindsay, Tetiana et Sebastian.

Je remercie de tout cœur ma famille, pour m'avoir toujours encouragée dans toutes mes décisions, même lorsque celles-ci m'éloignent géographiquement d'elle. Enfin, je remercie Nico pour le

soutien infailible et la compréhension dont il fait preuve dans tous les aspects de notre vie ensemble.

# Table des matières

<b>Table des matières</b> .....	<b>1</b>
<b>Liste des abréviations</b> .....	<b>7</b>
<b>Contexte du sujet de thèse</b> .....	<b>9</b>
<b>Chapitre 1 : Introduction</b> .....	<b>11</b>
1. Enjeux de l'étude des grands assemblages macromoléculaires .....	11
1.1. Les grands assemblages macromoléculaires .....	11
1.2. Techniques pour l'étude de la structure et des mouvements fonctionnels des grands assemblages macromoléculaires .....	11
2. Le domaine Archaea.....	13
2.2. Importance des archées dans les écosystèmes .....	17
2.3. Intérêt des protéines issues des archées hyperthermophiles pour les études de biologie structurale .....	21
3. Présentation des systèmes protéasomes.....	26
3.1. Importance des protéases intracellulaires .....	26
3.2. Mécanisme général de la protéolyse par les protéasomes .....	28
3.3. Le protéasome 26S des eucaryotes .....	30
3.4. Les protéasomes des bactéries, mitochondries et chloroplastes.....	36
3.5. Les protéasomes des archées .....	39
4. Structure et mode d'action des protéines AAA+ .....	41
4.1. Les protéines de la superfamille des AAA+ .....	41
4.2. Particularités des protéases AAA+ .....	44
5. Structure et mode d'action du protéasome archéen PAN-20S .....	46
5.1. Histoire de l'étude du protéasome archéen PAN-20S.....	46
5.2. Processus de dépliement des protéines et hydrolyse de l'ATP .....	49
6. Problématiques et objectifs de la thèse .....	56
<b>Chapitre 2 : Présentation des méthodes biophysiques principales utilisées : La diffusion de rayons X et de neutrons aux petits angles</b> .....	<b>59</b>
1. La diffusion aux petits angles (SAS) .....	59
1.1. Théorie .....	60
1.2. Interprétation des données .....	65
2. La diffusion de rayons X aux petits angles (SAXS).....	69
2.1. Histoire du rayonnement synchrotron.....	69
2.2. Propriétés des rayons X.....	69
2.3. Instrumentation à l'échelle d'un synchrotron : l'exemple de l'ESRF .....	70
2.4. Instrumentation à l'échelle d'une ligne de faisceau : l'exemple de BM29 à l'ESRF.....	71
3. La diffusion de neutrons aux petits angles (SANS) .....	73
3.1. Histoire des sources de neutrons .....	73

3.2.	Propriétés des neutrons .....	73
3.3.	Instrumentation à l'échelle d'une source : l'exemple de l'ILL.....	74
3.4.	Instrumentation à l'échelle d'une ligne de faisceau : l'exemple de D22 à l'ILL.....	77
3.5.	La variation de contraste .....	79
3.6.	Applications du SANS en biologie .....	80
<b>Chapitre 3 : Matériel et méthodes biochimiques .....</b>		<b>81</b>
1.	Expression des protéines d'intérêt.....	81
1.1.	Gènes d'intérêts, provenance des plasmides et transformation .....	81
1.2.	Expression des protéines hydrogénées .....	82
1.3.	Expression des protéines deutérées.....	83
2.	Purification des protéines d'intérêt.....	85
2.1.	Lyse des cellules.....	85
2.2.	Caractéristiques des protéines et choix des tampons.....	85
2.3.	Purifications des protéines .....	86
2.4.	Électrophorèses sur gel SDS-PAGE.....	90
3.	Effet de la température sur la perte de l'étiquette d'adressage « <i>ssrA</i> » de la GFP <sub>ssrA</sub> .....	91
3.1.	Test d'expression .....	91
3.2.	Effet de la température de stockage .....	91
4.	Mesure de la consommation de l'ATP par PAN.....	91
4.1.	Théorie de l'enzymologie .....	91
4.1.3.	Représentations graphiques .....	93
4.1.4.	Signification de <i>k<sub>cat</sub></i> , <i>K<sub>M</sub></i> et <i>k<sub>cat</sub>/K<sub>M</sub></i> .....	94
4.2.	Mesure de l'affinité de PAN pour l'ATP.....	94
4.3.	Mesure de la consommation de l'ATP au cours du temps .....	95
4.4.	Effet d'inhibition de l'ADP sur l'hydrolyse de l'ATP.....	95
4.5.	Effet de l'ajout de GFP <sub>ssrA</sub> sur la consommation d'ATP par PAN.....	95
5.	Mesures de la dénaturation de la GFP <sub>ssrA</sub> par spectroscopie de fluorescence .....	96
5.1.	Contexte des mesures par spectroscopie de fluorescence .....	96
5.2.	Criblage des stœchiométries de GFP <sub>ssrA</sub> , PAN, particule 20S et d'ATP.....	97
5.3.	Mesures fonctionnelles avec différentes conditions contrôle .....	99
5.4.	Effets de l'ATP, ADP et ATP <sub>γ</sub> S sur l'activité du protéasome PAN-20S.....	99
5.5.	Effet d'inhibition de l'ADP sur le dépliement de la GFP <sub>ssrA</sub> par PAN.....	99
5.6.	Effets des variations de température sur l'activité du système protéasome PAN-20S.....	99
5.7.	Effet de la streptavidine monovalente sur l'activité de dépliement de la GFP <sub>ssrA</sub> par PAN 100	
5.8.	Mesures de la dénaturation de la GFP <sub>ssrA</sub> _Nter-biotine et de la GFP <sub>ssrA</sub> _Cter-biotine ...	100
5.9.	Mesures de la dénaturation de la GFP <sub>ssrA</sub> _Nter-streptavidine et de la GFP <sub>ssrA</sub> _Cter- streptavidine.....	100
6.	Évaluation de la dégradation de la GFP <sub>ssrA</sub> par Western-blot anti-GFP .....	101
6.1.	Principe du Western-blot .....	101
6.2.	Western-blot avec le système GFP <sub>ssrA</sub> , PAN et particule 20S .....	103

6.3. Western-blot avec le système GFPssrA_Nter-streptavidine ou GFPssrA_Cter-streptavidine, PAN et particule 20S .....	103
<b>Chapitre 4 : Matériel et méthodes biophysiques .....</b>	<b>105</b>
1. Diffusion de rayons X aux petits angles .....	105
1.1. Protocole expérimental d'une collecte de données SAXS sur BM29.....	105
2. Diffusion de neutrons aux petits angles résolue en temps (TR-SANS) et combinaison avec la spectroscopie de fluorescence en ligne.....	107
2.1. Instrumentation et protocole expérimental.....	107
2.2. Analyse des données .....	110
3. Diffusion de la lumière .....	111
3.1. Diffusion statique (ou élastique) de la lumière (SLS).....	112
3.2. Diffusion dynamique (ou quasi-élastique) de la lumière (DLS) .....	113
3.3. Protocole expérimental des mesures SLS-DLS réalisées.....	114
3.4. Protocole expérimental des mesures SEC-MALLS réalisées .....	115
4. Spectrométrie de masse .....	115
4.1. Introduction à la spectrométrie de masse.....	116
4.2. MALDI-MS et ESI-MS .....	116
4.3. Analyse des protéines purifiées par ESI-MS .....	118
4.4. Analyse des produits de dégradation par MALDI-TOF-MS .....	118
5. Microscopie électronique à transmission en coloration négative.....	119
5.1. Introduction à la microscopie électronique à transmission .....	119
5.2. Visualisation de la protéine PAN.....	120
1.1. Visualisation de la particule 20S .....	120
1.2. Interaction entre PAN et la particule 20S.....	120
<b>Chapitre 5 : Étude des protéines isolées .....</b>	<b>121</b>
1. Liste des différentes protéines purifiées.....	121
2. Étude de la particule régulatrice PAN.....	122
2.1. Rappel du contexte.....	122
2.2. Qualité de la protéine PAN purifiée mesurée par spectrométrie de masse.....	122
2.3. Étude de la protéine PAN par SAXS.....	123
2.4. Effet de la concentration d'ATP sur l'état oligomérique de PAN.....	125
3. Étude de la particule catalytique 20S .....	130
3.1. Problématique de l'assemblage de la particule 20S .....	130
3.2. Qualité de la particule 20S purifiée estimée par spectrométrie de masse .....	131
3.3. Étude de la particule 20S par SEC-MALLS.....	132
3.4. Étude de la particule 20S étudiée par SAXS .....	133
3.5. Visualisation de la particule 20S par microscopie électronique à coloration négative.....	133
3.6. Interaction avec PAN visualisée par microscopie électronique à coloration négative.....	134
4. Étude de la GFPssrA .....	135
4.1. Introduction à la fluorescence et à la protéine fluorescente verte GFP .....	135
4.2. Qualité de la protéine GFPssrA purifiée estimée par spectrométrie de masse.....	142

4.3.	Étude de l'effet de la température sur la perte de la séquence d'adressage « ssrA ».....	143
4.4.	Étude de la GFPssrA par SAXS .....	144
5.	Étude de la GFPssrA_Nter-biotine, de la GFPssrA_Cter-biotine et de la streptavidine monovalente 145	
5.1.	Présentation des protéines étudiées.....	145
5.2.	Qualité des GFPssrA_Nter-biotine et GFPssrA_Cter-biotine purifiées estimée par spectrométrie de masse .....	146
5.3.	Qualité de la streptavidine monovalente purifiée estimée par spectrométrie de masse.....	147
6.	Étude de la GFPssrA_Nter-streptavidine et de la GFPssrA_Cter-streptavidine.....	148
6.1.	Présentation des protéines étudiées.....	148
6.2.	Étude par SEC-MALLS des protéines des complexes GFPssrA_Nter-streptavidine et GFPssrA_Cter-streptavidine.....	149
6.3.	Étude par SAXS des protéines des complexes GFPssrA_Nter-streptavidine et GFPssrA_Cter-streptavidine.....	150
7.	Conclusion du chapitre.....	153
<b>Chapitre 6 : Mesures biochimiques sur le système GFPssrA, protéasome PAN-20S.....</b>		<b>155</b>
1.	Activité de dépliement et de dégradation : mesures par spectroscopie de fluorescence et de dégradation de la GFPssrA par Western-blot anti-GFP .....	155
1.1.	Avant-propos .....	155
1.2.	Mesures par spectroscopie de fluorescence .....	156
1.3.	Étude de la dégradation de la GFPssrA visualisé par Western-blot anti-GFP.....	159
2.	Recherche par spectroscopie de fluorescence des conditions de stœchiométries optimales entre les composants du mélange réactionnel en vue des expériences de TR-SANS.....	160
2.1.	Criblage des stœchiométries GFPssrA et PAN à différentes concentrations d'ATP et de MgCl <sub>2</sub> 160	
2.2.	Criblage des stœchiométries GFPssrA, PAN et 20S à 100 mM ATP.....	163
2.3.	Conclusions du criblage.....	164
3.	Effets de l'ATP, ADP et ATPγS sur l'activité du protéasome PAN-20S mesurés par spectroscopie de fluorescence.....	165
4.	Détection des produits de dégradation générés par le système par spectrométrie de masse .....	166
5.	Effets des variations de température sur l'activité du protéasome PAN-20S mesurés par spectroscopie de fluorescence.....	170
6.	Étude de la consommation de l'ATP par PAN .....	171
6.1.	Éstimation de l'affinité de PAN et de la GFPssrA pour l'ATP .....	171
6.2.	Mesure de la consommation d'ATP au cours du temps .....	173
6.3.	Effet d'inhibition de l'ADP sur l'activité ATPase de PAN .....	176
6.4.	Effet de l'ajout de GFPssrA sur l'hydrolyse de l'ATP par PAN .....	178
7.	Conclusion générale du chapitre .....	180
<b>Chapitre 7 : Diffusion de neutrons aux petits angles résolue en temps sur le système GFPssrA, protéasome PAN-20S .....</b>		<b>181</b>
1.	Rappel du contexte de l'expérience .....	181
2.	Mesures TR-SANS et de fluorescence.....	182
2.1.	Préambule.....	182

2.2.	Courbes SANS au cours du temps pour les deux conditions SANS .....	183
2.3.	Évolution des paramètres de $I(0)$ et $R_g$ au cours du temps pour les deux conditions SANS 186	
2.4.	Pourcentages finaux des paramètres mesurés .....	188
2.5.	Dépendance temporelle des différents paramètres mesurés .....	190
2.6.	Comparaison entre les agrégats générés par PAN et les agrégats dus à la dénaturation thermique de la d-GFPssrA.....	192
3.	Discussion et conclusion générale du chapitre .....	195
<b>Chapitre 8 : Mesures biochimiques et TR-SANS sur le système GFPssrA-streptavidine, protéasome PAN-20S .....</b>		<b>199</b>
1.	Contexte de l'étude.....	199
1.1.	Stratégie employée dans la littérature .....	199
1.2.	Stratégie employée dans cette étude .....	201
1.3.	Conditions expérimentales choisies.....	202
2.	Activité de dénaturation suivie par spectroscopie de fluorescence .....	203
2.1.	GFPssrA libre, GFPssrA_Nter-biotine et GFPssrA_Cter-biotine.....	203
2.2.	GFPssrA libre et streptavidine monovalente.....	205
2.3.	GFPssrA_Nter-streptavidine et GFPssrA_Cter-streptavidine .....	206
2.4.	Effet de l'ajout de GFPssrA libre .....	207
3.	Discussion des mesures biochimiques réalisées.....	210
4.	Diffusion de neutrons aux petits angles en temps résolu (TR-SANS).....	211
4.1.	Préambule .....	211
4.2.	Mesures TR-SANS avec la d-GFPssrA_Nter-biotine et d-GFPssrA_Nter-streptavidine ....	212
4.3.	Mesures TR-SANS avec la GFPssrA_Cter-biotine et GFPssrA_Cter-streptavidine .....	214
4.4.	Conclusions à retenir des mesures TR-SANS.....	217
5.	Discussion et conclusion générale du chapitre .....	218
<b>Chapitre 9 : Conclusion générale et perspectives.....</b>		<b>221</b>
<b>Références .....</b>		<b>225</b>



# Liste des abréviations

AAA+ : ATPases Associated with diverse cellular Activities

ADN : Acide Désoxyribonucléique

ADP : Adénosine diphosphate

ARN : Acide Ribonucléique

ARNr : Acide Ribonucléique ribosomale

ATP : Adénosine triphosphate

BM29 : Bending Magnet

CCD : Charge-Coupled Device

Cdc48 : Cell division control protein

Cryo-EM : Cryo-microscopie électronique

DEAE : Diethylaminoethyl

D-lab : deuteration lab

E. coli : Escherichia coli

ELMA : Extremophiles and Large Molecular Assemblies

EM : Electron Microscopy

ESRF : European Synchrotron Radiation Facility

FPLC : Fast Protein Liquid Chromatography

GFP : Green Fluorescent Protein

Hsp : Heat Shock Protein

IBS : Institut de Biologie Structurale

ILL : Institut Laue Langevin

IPTG : Isopropyl  $\beta$ -D-1-thiogalactopyranoside

LB : Lysogeny broth

MET : Microscopie électronique à transmission

Ntn : N-terminal nucleophile hydrolase

NTP : Nucleoside Triphosphate

OB : Oligonucleotide/Oligosaccharide-binding fold

PAN : Proteasome Activating Nucleotidase  
PDB : Protein Data Bank  
P-loop : Phosphate loop  
RI : Refractive Index  
RMN : Résonance Magnétique Nucléaire  
Rpn : Regulatory Particle non-ATPase  
Rpt : Regulatory Particle triple-A ATPase  
S : Svedberg  
SAMP : Small Archaeal Modifier Protein  
SANS : Small Angle Neutron Scattering  
SAXS : Small Angle X-ray Scattering  
SRH : Second Region of Homology  
SSD : Sensor and Substrate Discrimination domain  
TEA : Triéthanolamine  
Tris : Tris(hydroxymethyl)aminomethane  
Ub : Ubiquitine  
UbaA : Ubiquitin-like activating protein of archaea  
UBP : Ubiquitin-like Protein  
UPP : Ubiquitin Proteasome Pathway  
TR-SANS : Time-Resolved Small Angle Neutron Scattering  
VCP : Valosin-Containing Protein

# Contexte du sujet de thèse

La dégradation ciblée des protéines, aussi appelée protéolyse, est aujourd'hui vue comme un mécanisme de régulation cellulaire d'importance centrale. Cependant, l'acceptation de cette idée n'a pas été aisée et l'histoire de la découverte des protéasomes illustre bien la difficulté de changer les paradigmes en biologie. De vives discussions ont eu lieu à partir des années 1940 afin de déterminer si les protéines au sein des cellules étaient stables au cours du temps ou si elles étaient renouvelées<sup>1-3</sup>. Entre sa découverte au milieu des années 1950 et jusqu'au milieu des années 1970, le lysosome était considéré comme l'organite dans lequel les protéines intracellulaires sont dégradées<sup>4,5</sup>. Le système d'ubiquitine a ensuite été découvert à la fin des années 1970<sup>6,7</sup>. Une nouvelle ère dans le domaine de la dégradation des protéines a débuté à la fin des années 1980 avec l'identification des réactions et des enzymes impliquées dans la cascade ubiquitine-protéasome<sup>2,7</sup>. Il est aujourd'hui bien établi que la dégradation protéique est assurée majoritairement par des systèmes protéolytiques ATP-dépendants, dont le protéasome est le principal acteur. Des études ont permis de démontrer que la dégradation continue des protéines est nécessaire pour réguler les concentrations intracellulaires des enzymes qui contrôlent toutes les réactions métaboliques, ainsi que le contenu général de toutes les autres protéines, en réponse aux modifications physiologiques<sup>8</sup>. La protéolyse permet également d'éliminer les protéines identifiées comme anormales par des protéines chaperons<sup>9</sup>. Ces études ont été accompagnées de nombreuses recherches sur les mécanismes spécifiques et hautement régulés de la dégradation de protéines par les protéasomes des eucaryotes, bactéries et archées. Ces protéasomes possèdent chacun leur propre mode de reconnaissance des protéines cibles et de régulation, bien que leur structure ainsi que leurs mécanismes de dépliage ATP-dépendant et de dégradation des protéines substrat reposent sur des bases communes<sup>10</sup>.

Le travail présenté ici a été réalisé dans le cadre d'une thèse en biologie/biophysique des grands assemblages moléculaires avec comme but principal de **comprendre les mécanismes structuraux qui permettent le dépliage et la dégradation de protéines cibles en utilisant comme modèle le protéasome PAN-20S archéen**. Ce projet a été soutenu par un financement ANR (« ProtStretch ») du groupe ELMA à l'Institut de Biologie Structurale (IBS), dédié à l'étude de la dynamique fonctionnelle du protéasome PAN-20S archéen. **Cette thèse avait également pour objectif de développer les études biophysiques résolues en temps par diffusion de neutrons aux petits angles : TR-SANS, de l'anglais « Time-Resolved Small Angle Neutron Scattering »**. Pour cela nous avons étudié le dépliage et la dégradation par le protéasome de *Methanocaldococcus jannaschii* d'un substrat modèle fluorescent doté d'une étiquette d'adressage au protéasome, la GFP<sub>ssrA</sub>. Le projet dans lequel s'inscrit cette thèse est une collaboration avec une équipe de RMN qui a pour but d'étudier le même système par RMN. Ce manuscrit ne présentera que les études ayant permis la réalisation d'expérience de diffusion de neutrons aux petits angles résolues en temps, qui était le cœur du projet de thèse.

Le travail réalisé dans la première partie de la thèse a permis de maîtriser la qualité des protéines isolées et la biochimie du protéasome PAN-20S pour le dépliage et la dégradation de la GFP<sub>ssrA</sub> dans les conditions expérimentales nécessaires aux mesures TR-SANS. La spectroscopie de fluorescence *in situ* sur la ligne D22 de l'ILL, la deutération et la variation de contraste ont été combinées pour les mesures TR-SANS. Cette approche a permis de suivre en temps réel et de

façon spécifique le dépliement et la dégradation de la GFP<sub>ssrA</sub> par le protéasome PAN-20S. Ce travail a permis d'apporter des éléments en faveur d'un modèle expliquant le fonctionnement du protéasome PAN-20S et des systèmes AAA+ protéolytiques en général. Il représente aussi un fort intérêt méthodologique avec l'utilisation du TR-SANS, couplée avec la spectroscopie optique, comme outil pour l'étude des grands assemblages macromoléculaires.

Ce manuscrit comprend un **premier chapitre** d'introduction qui décrit d'abord rapidement les méthodes structurales permettant l'étude de grands assemblages macromoléculaires, puis l'intérêt d'utiliser des machineries macromoléculaires provenant d'archées hyperthermophiles pour de telles études. Une étude bibliographique présente ensuite l'importance de la dégradation intracellulaire des protéines et l'importance des protéases ATP-dépendantes dans la protéolyse intracellulaire. Les protéasomes des eucaryotes, bactéries et archées sont ensuite présentés et le protéasome PAN-20S provenant de l'archée hyperthermophile *Methanocaldococcus jannaschii* est décrit de façon approfondie avant d'exposer les objectifs principaux de ce projet de thèse.

Le **deuxième chapitre** de ce manuscrit décrira les techniques de diffusion de rayons X et de neutrons aux petits-angles (respectivement SAXS et SANS) de façon approfondie. Les **chapitres 3 et 4** détailleront respectivement les aspects biochimiques de la préparation des échantillons et les techniques biophysiques utilisées.

Les résultats seront présentés dans plusieurs chapitres. La caractérisation des protéines isolées utilisées lors de ce projet de thèse est présentée dans le **chapitre 5**. Les études biochimiques sur le système GFP<sub>ssrA</sub>, protéasome PAN-20S sont exposées dans le **chapitre 6** et les mesures TR-SANS sur ce système dans le **chapitre 7**. Une stratégie a été employée dans le but de bloquer la GFP<sub>ssrA</sub> au cours de son dépliement par PAN. Pour cela, une molécule de streptavidine monovalente a été associée proche de l'extrémité N-terminale ou C-terminale de la GFP<sub>ssrA</sub> dans le but de provoquer une gêne stérique. La caractérisation biochimique du système GFP<sub>ssrA</sub>-streptavidine, protéasome PAN-20S et les mesures TR-SANS associées sont décrites dans le **chapitre 8**. Les résultats obtenus seront discutés au fur et à mesure au sein de chaque chapitre et une conclusion générale sera dressée dans le **chapitre 9**.

# Chapitre 1 :

## Introduction

### 1. Enjeux de l'étude des grands assemblages macromoléculaires

#### 1.1. Les grands assemblages macromoléculaires

Pour accomplir des fonctions cellulaires complexes, certaines protéines forment des assemblages macromoléculaires qui peuvent contenir des protéines, des acides nucléiques, des groupements prosthétiques et des cofacteurs, des lipides, des sucres, et des composés inorganiques. Ces grands édifices agissent comme des machines moléculaires. Ces machineries remplissent un large éventail de fonctions biologiques : métabolisme, transport moléculaire, transduction du signal, régulation de l'expression des gènes ainsi que, dans le cas de la problématique de cette thèse, la dégradation des protéines intracellulaires<sup>11</sup>. Pour exercer leurs fonctions, elles peuvent s'associer à des systèmes cellulaires spécifiques tels que des organites, des membranes ou des éléments du cytosquelette. Ces complexes sont souvent composés de plus d'un type de protéine, et chaque protéine peut être présente en plusieurs exemplaires. En plus de leur complexité structurale, les grands assemblages macromoléculaires ne sont pas des objets statiques. Ils adoptent souvent des compositions et des conformations différentes en réponse aux changements de leur environnement et ils présentent une large gamme de mouvements moléculaires locaux qui peuvent être importants pour leur fonction. La détermination de la structure des assemblages intacts peut révéler des conformations et des interactions intermoléculaires qui ne sont présentes que dans le contexte de l'assemblage complet<sup>12</sup>.

La première étape de l'étude d'une machine cellulaire consiste souvent à déterminer la structure de ses composants individuels, comme une seule protéine ou un domaine protéique avec ses ligands, soit de façon isolée, soit dans un complexe moléculaire. Cela se fait le plus souvent par cristallographie aux rayons X et par spectroscopie de résonance magnétique nucléaire (RMN) en solution. Cependant, il est très intéressant de déterminer la structure des assemblages intacts dans des conditions proches de l'environnement physiologique de l'organisme vivant. De telles études sont nécessaires pour révéler les conformations, les interactions intermoléculaires et les mouvements moléculaires locaux qui ne sont présents que dans le contexte de l'assemblage complet. Ces travaux sont indispensables pour expliquer le mode d'action des machineries complexes. De telles études posent des défis techniques pour les principales techniques utilisées en biologie structurale<sup>13,14</sup>.

#### 1.2. Techniques pour l'étude de la structure et des mouvements fonctionnels des grands assemblages macromoléculaires

Cette partie aborde brièvement les différentes méthodes utilisées en biologie structurale et discute leur capacité à étudier de grands assemblages macromoléculaires et les mouvements fonctionnels des protéines qui les composent.

Parmi les différentes techniques de biologie structurale moderne, la cryo-microscopie électronique (cryo-EM) a réalisé des avancées spectaculaires au cours des dernières années, grâce entre autres à la combinaison de meilleurs détecteurs et outils d'analyse<sup>15,16</sup>. En cryo-EM, les assemblages macromoléculaires sont étudiés à l'état congelé dans un film mince de glace amorphe, préparé par la congélation rapide de l'échantillon biologique. Les assemblages n'ont pas besoin d'être solubles ou sous forme cristalline. Plusieurs centaines d'images sont utilisées pour reconstituer la structure de la particule d'intérêt<sup>15</sup>. Des structures à haute résolution (souvent proche de 3,5 Å) peuvent généralement être obtenues pour des systèmes jusqu'à des poids moléculaires très élevés (>11 MDa)<sup>17</sup>. De plus, la classification d'images permet de trier un ensemble de données cryo-EM contenant des particules de conformation différentes, chaque classe correspondant à un état fonctionnel différent<sup>16</sup>. Ces « images instantanées » des protéines peuvent alors permettre de reconstituer le mécanisme de réactions enzymatiques. Cependant, la résolution des cartes de densité cryo-EM peut être limitée lorsque l'assemblage moléculaire présente une hétérogénéité ou une flexibilité conformationnelle avant la congélation, ou lorsqu'il adopte une orientation préférentielle dans la couche de support de glace amorphe<sup>18</sup>.

Dans le domaine de la diffraction des rayons X, l'introduction de nouveaux instruments révolutionnaires, en particulier les lasers à électrons libres, permettent des applications entièrement nouvelles<sup>19</sup>. De plus, les développements continus en matière de faisceaux automatisés, de faisceaux nanofocalisés ou de cristallographie en série, continuent également à maintenir le rôle prépondérant de la cristallographie aux rayons X comme outil de détermination de structure, avec un débit plus important et de meilleures données provenant de petits cristaux<sup>20</sup>. La cristallographie aux rayons X, particulièrement avec la cristallographie en série et les lasers à électrons libre, peut permettre de capturer des « images instantanées » des protéines dans différentes conformations et de reconstituer les réactions à partir de ces différentes images statiques. La technologie des lasers à électrons libres tend aujourd'hui à suivre des réactions en temps réel grâce à sa rapidité d'acquisition des données<sup>21</sup>. Cependant, la principale limitation de la cristallographie aux rayons X est l'exigence de cristaux tridimensionnels bien ordonnés pour obtenir un fort diagramme de diffraction qui permet de résoudre la structure à haute résolution. Dans le cas d'un système à plusieurs composants tel qu'un complexe macromoléculaire, il est très difficile de trouver les conditions de cristallisation et de croissance des cristaux puisque ces systèmes n'ont pas une structure unique bien définie et qu'il est difficile d'en obtenir dans un état pur et stable<sup>22</sup>.

La spectroscopie de résonance magnétique nucléaire (RMN) en solution fournit des informations sur la structure et la dynamique locale des protéines, ainsi que sur leurs propriétés de liaison<sup>23,24</sup>. La détermination des structures *ab initio* par RMN est généralement limitée à des particules d'environ 40 kDa, en raison de la diminution rapide de la qualité du signal de RMN (élargissement de la ligne) lorsque le poids moléculaire augmente<sup>25</sup>. Les études structurales par RMN en solution sont basées sur des spectres de haute qualité enregistrés avec une bonne sensibilité et résolution. Avec l'augmentation de la masse moléculaire, ces exigences sont difficiles à atteindre en raison de l'élargissement et du chevauchement des pics. Au cours des 15 dernières années, l'application de la deutération sélective ou complète des hydrogènes des protéines, accompagnée d'un traitement plus efficace et automatisé des données, a facilité et accéléré la détermination des structures<sup>26-29</sup> et ont étendu la limite d'application à des particules de 1 MDa. En particulier, l'approche méthyl-TROSY<sup>26,29,30</sup> qui utilise la protonation sélective et le marquage <sup>13</sup>C des groupes méthyle de l'Ile,

Leu et Val, a permis des études structurales et dynamiques sur de grands assemblages macromoléculaires<sup>25</sup>.

La RMN à l'état solide, avec la méthode de « rotation à l'angle magique » (MAS, de l'anglais « *Magic-Angle Spinning* »), offre de nombreux avantages par rapport aux autres techniques. Il n'y a pas de limite de taille théorique à son application, bien que des problèmes de sensibilité et de résolution surviennent avec l'augmentation du poids moléculaire<sup>13,31</sup>. Les échantillons peuvent être notamment sous forme soluble, sédimentée, ou de nanocristaux<sup>32-34</sup>. Au cours de la dernière décennie, la RMN MAS a fait des progrès considérables. Il est maintenant possible déterminer les structures et caractériser la dynamique de systèmes complexes, y compris de grands assemblages de protéines, à résolution atomique<sup>13,31,35,36</sup>. Les éléments limitants de cette méthode, la sensibilité et la résolution, sont influencés par divers facteurs tels que de la taille et la dynamique des protéines ou l'homogénéité de l'échantillon. Ces problèmes peuvent être atténués grâce l'instrumentation utilisée et un choix approprié de marquage isotopique<sup>25,31</sup>.

Les techniques de diffusion de rayons X et de neutrons aux petits angles, qui sont les méthodes principales utilisées lors de ce projet de thèse sont décrites en détail dans le chapitre 2. Ces techniques permettent d'obtenir des informations structurales à basse résolution sur des protéines telles que leur masse moléculaire, taille et leur forme. Ces approches mesurent des populations de protéines en solution et peuvent caractériser les équilibres conformationnels au sein de ces populations. Elles ont l'avantage de permettre d'étudier les grands assemblages moléculaires dans des conditions proches des conditions physiologiques et de nombreuses conditions d'échantillonnage peuvent être testées de façon relativement simple. Enfin, les données peuvent être acquises de façon résolue dans le temps et peuvent ainsi permettre de suivre des réactions enzymatiques<sup>37</sup>. Dans le cas de la diffusion de neutrons aux petits angles, l'utilisation de la variation de contraste H<sub>2</sub>O/D<sub>2</sub>O couplée à la deutération sélective de parties du complexe d'étude permettent d'obtenir des informations ciblées sur la partie d'intérêt (voir chapitre 2, partie 3.5).

Pour aborder la complexité des grands assemblages macromoléculaires, une seule technique est insuffisante, et toutes les techniques structurales décrites précédemment présentent des avantages et des inconvénients complémentaires. La combinaison de plusieurs méthodes est aujourd'hui une pratique courante dans tous les domaines de la biologie structurale. On parle alors de biologie structurale intégrative<sup>14,38</sup>.

Il est important de noter que les études structurales réalisées sur les grands assemblages macromoléculaires, notamment les études visant à réaliser des avancées méthodologiques, mettent bien souvent à profit les avantages fournis par les protéines issus d'organismes hyperthermophiles<sup>35,39,40</sup> (voir partie 2.3.2). C'est le cas de ce projet de thèse qui avait pour objectif d'étudier le protéasome PAN-20S issu de l'archée hyperthermophile *Methanocaldococcus jannaschii* par diffusion de neutrons aux petits angles résolue en temps.

## 2. Le domaine Archaea

L'arbre du vivant décrit comment les êtres vivants sont apparentés et quelle serait leur origine. L'histoire de la classification des êtres vivants, depuis la première classification dans l'Antiquité par Aristote, est marquée de nombreux bouleversements et remises en question. Avec l'avènement de la microbiologie, les cellules eucaryotes (possédant un noyau entouré d'une membrane et des organites individualisés tels que les mitochondries, réticulum endoplasmique et lysosome) ont été

opposées aux cellules procaryotes (sans noyau et organites et plus petites que les cellules eucaryotes). Dans les années 1970, un arbre du vivant à cinq domaines, ou règnes, était décrit : les procaryotes (Monera), les eucaryotes unicellulaires (protistes), et les pluricellulaires champignons, plantes et animaux (figure 1)<sup>41</sup>.

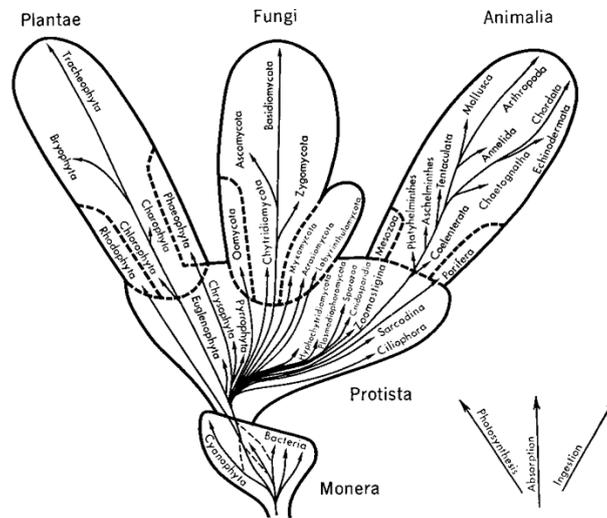


Figure 1 : Arbre du vivant à cinq règnes tel qu'admis dans les années 1970. Whittaker, R. H. (1969).

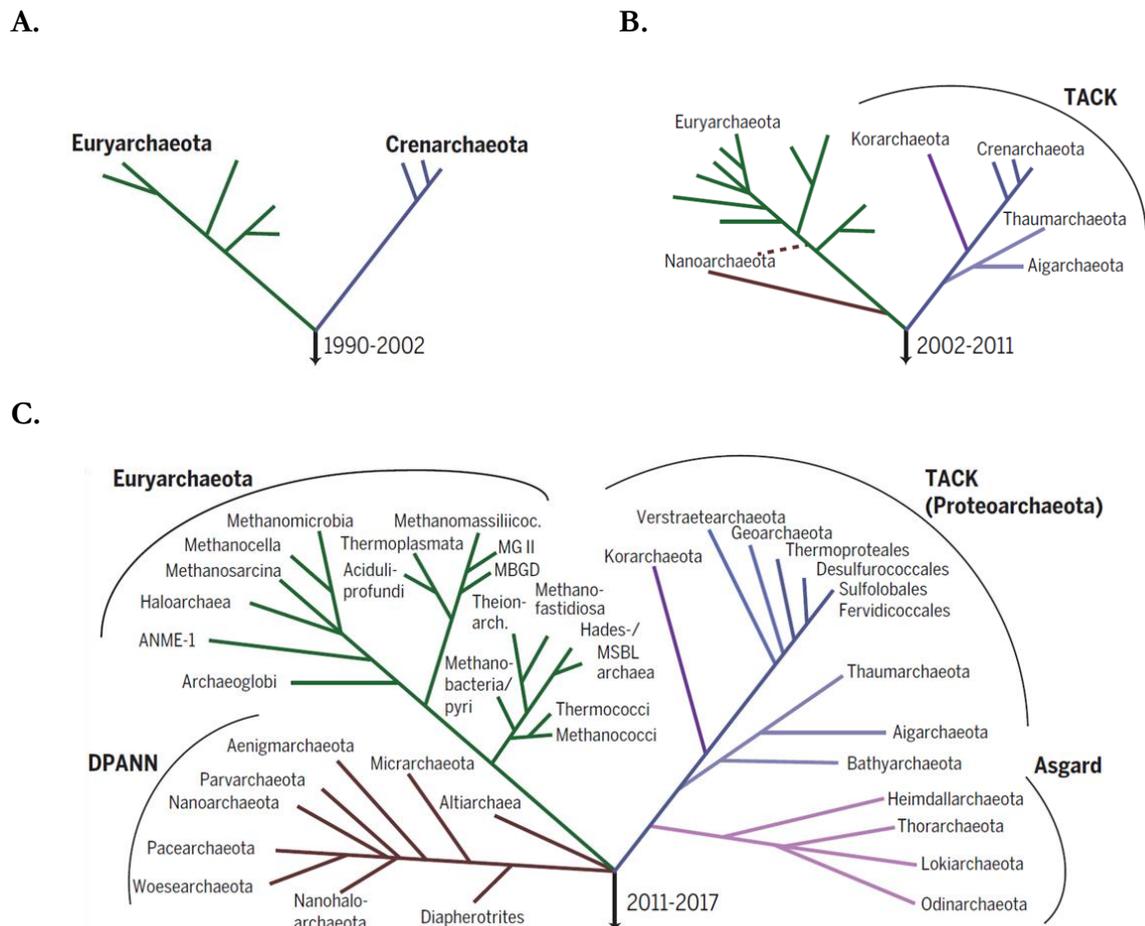
Ce découpage du monde du vivant proposait un scénario simple de l'histoire évolutive : les procaryotes seraient des organismes primitifs desquels auraient évolué les protistes, puis les trois règnes pluricellulaires. Cette vision a été balayée par la découverte de l'extraordinaire diversité du monde microbien ainsi que par la découverte des archées qui a bouleversé la classification du vivant à la fin des années 1970. Dans ce nouveau contexte, l'origine des eucaryotes est sujette à discussion.

### 2.1.1. Découverte des archées

En 1977, au cours de l'étude comparative de séquences de l'ARN ribosomal (ARNr) 16S des êtres vivants, Carl Woese a découvert l'existence d'un groupe de procaryotes distinct de celui des bactéries<sup>42</sup>. Bien que leur taille, leur forme et leur organisation cellulaire (cellules procaryotes à ADN circulaires organisés en opérons, et possédant des plasmides) était similaire à celle des bactéries, des caractéristiques inédites dans les séquences des ARNr conduisirent à proposer une nouvelle classification du vivant en trois domaines : Eukarya, Eubacteria et Archaeobacteria<sup>42</sup>. Au cours des années qui suivirent, l'existence de similitudes avec les eucaryotes, associée à des caractères uniques partagés par tous les membres de ce nouveau domaine, comme la composition particulière de leur paroi en « S-Layer », ont poussé Woese et ses collègues à renommer les 3 domaines Eukarya, Bacteria, et Archaea<sup>43</sup>. Par de nombreux aspects, les archées sont considérées aujourd'hui phylogénétiquement et fonctionnellement plus proches des eucaryotes que des bactéries, notamment dans les processus informationnels (tels que la réplication, la transcription et la traduction), la structure de la chromatine, les voies de sécrétions, certains mécanismes de division cellulaire, mais aussi le système protéolytique protéasome qui a été étudié lors cette thèse<sup>44-46</sup>.

### 2.1.2. Classification des archées et impact sur l'arbre du vivant

La phylogénie complexe du domaine Archaea étant encore un sujet d'étude en plein essor, l'arbre du vivant archéen n'a pas cessé d'évoluer depuis leur reconnaissance comme troisième domaine du vivant (figure 2).



**Figure 2 : Représentation schématique de la façon dont la forme de l'arbre du vivant archéen s'est développé cours du temps.** Spang, A., et al. (2017)<sup>53</sup>. **A.** Entre 1990 et 2002, seuls deux superphyla étaient connus (Euryarchaeota et Crenarchaeota). **B.** D'autres phyla ont été identifiés entre 2002 et 2011. **C.** Depuis 2011, diverses lignées d'archées supplémentaires ont été découvertes, stimulées par les progrès des technologies de séquençage et l'utilisation des techniques de la métagénomique et de la génomique unicellulaire.

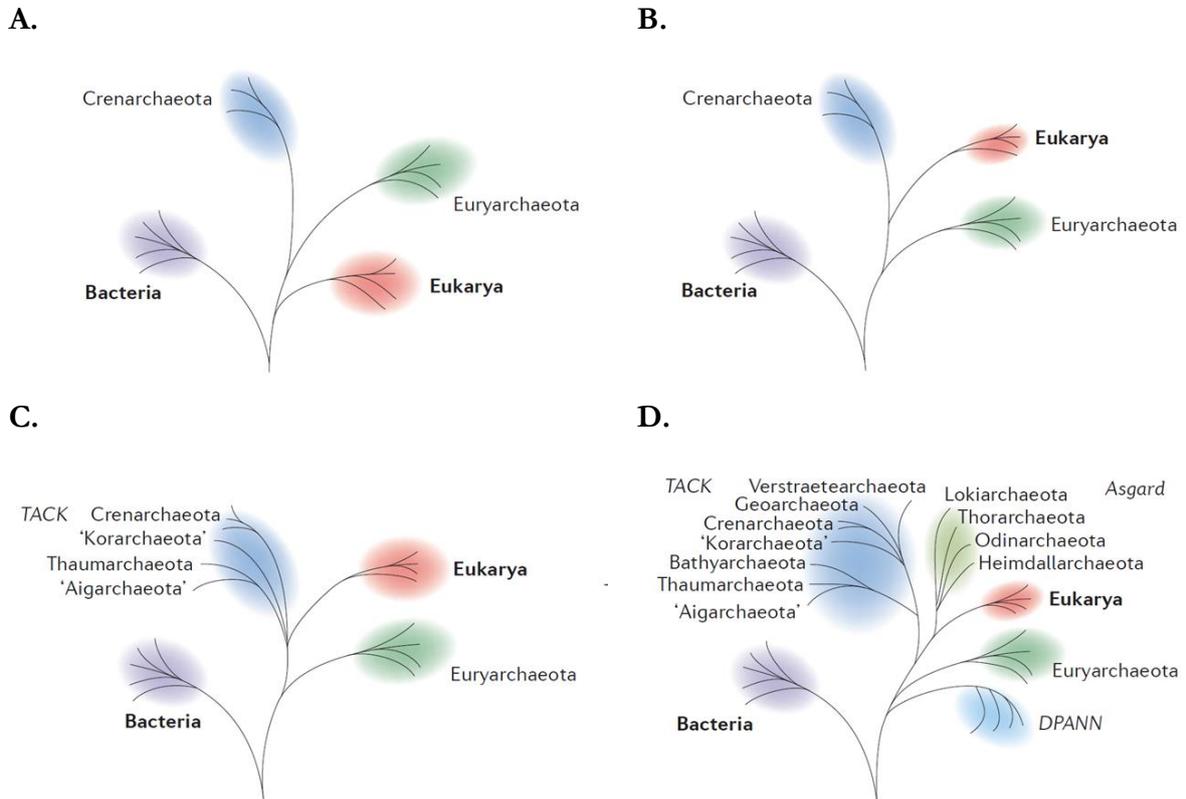
Entre 1990 et 2002 (figure 2.A), il n'y avait seulement que deux superphyla (sous-domaines) d'archées : Euryarchaeota et Crenarchaeota. Par conséquent, la majorité des séquences de gènes d'ARNr 16S détectés dans les études environnementales ont été initialement attribués à l'un ou l'autre de ces deux superphyla<sup>43</sup>.

Entre 2002 et 2011 (figure 2.B), les Crenarchaeota ont été regroupées avec les Thaumarchaeota<sup>47</sup>, les Aigarchaeota<sup>48</sup> et les Korarchaeota<sup>49</sup> pour former le superphylum des « TACK »<sup>50</sup>. De plus, les Nanoarchaeota<sup>51</sup> ont été reconnues comme formant un autre phylum.

La connaissance de la diversité des archées s'est encore récemment améliorée avec les développements dans le domaine de la métagénomique et de la génomique unicellulaire qui permettent de reconstituer des génomes d'archées directement à partir d'échantillons environnementaux sans culture préalable. Les archées sont divisées en quatre superphyla (figure 2.C) : les Euryarchaeota, les nouvelles TACK (ou Proteoarchaeota)<sup>52</sup>, les Asgard et les DPANN<sup>53</sup>. Les TACK comprennent aujourd'hui trois phyla supplémentaires à ceux précédemment mentionnés : les Geoarchaeota<sup>54</sup>, les Bathyarchaeota<sup>55</sup>, et les Verstraetearchaeota<sup>56</sup>. Les DPANN comprennent au moins neuf phyla<sup>57,58</sup> dont les Nanoarchaeota précédemment mentionné. Les

phyla décrits comme appartenant aux Asgard comprennent les Lokiarchaeota, les Thorarchaeota, les Odinararchaeota, et les Heimdallarchaeota<sup>59-61</sup>.

La taxonomie actuelle des groupes d'archées reconnus est sujette à controverse principalement en raison de l'absence de lignes directrices claires pour la classification des microorganismes<sup>62</sup>. Les efforts futurs qui utiliseront des critères bien définis fondés sur des cadres phylogénomiques et génomiques comparatifs aideront à établir une taxonomie robuste du domaine Archaea.



**Figure 3 : Représentation schématique de l'évolution de l'arbre du vivant au cours des 40 dernières années.** Eme, L., et al. (2017)<sup>70</sup>. L'arbre des trois domaines (A) suggérait qu'Archaea, Bacteria, et Eukarya représentaient les domaines primaires de la vie. Une topologie à deux domaines a ensuite été proposée (B) et mise à jour progressivement avec les TACK (C) puis les Asgard (D). Les bactéries et les eucaryotes sont indiquées respectivement en violet et en rouge, tandis que le vert et le bleu représentent les lignées d'archées. Les noms de domaine et les noms des superphyla sont écrits respectivement en gras et en italique. Les noms des phyla sont indiqués entre guillemets.

L'évolution de l'arbre phylogénétique des archées mène de façon logique à l'évolution de l'arbre du vivant (figure 3), et les relations évolutives entre les domaines Eukarya et Archaea sont aujourd'hui sujettes à débat. Dans l'arbre du vivant à trois domaines proposés initialement, Archaea et Eukarya représentaient chacun un groupe monophylétique et partageaient un ancêtre commun unique excluant les bactéries<sup>43,63</sup> (figure 3.A). En revanche, un arbre du vivant à deux domaines a été proposé très tôt, et la topologie de cet arbre a subi plusieurs changements. Lorsque l'on pensait que le domaine Archaea se composait uniquement des superphyla Euryarchaeota et Crenarchaeota (figure 3.B), une relation étroite entre Eukarya et du Crenarchaeota était envisagée<sup>64,65</sup>. Au début des années 2010, il a été proposé que les eucaryotes se ramifiaient au sein du superphylum TACK (figure 3.C), ou en représentaient un groupe frère<sup>46,50,66-69</sup>. Les analyses phylogénomiques incluant aux archées le superphylum Asgard ont fortement suggéré que les eucaryotes provenaient de l'intérieur des archées Asgard ou qu'ils représentaient un groupe frère<sup>58,59</sup> (figure 3.C). En effet, les génomes

des archées Asgard codent un répertoire de protéines que l'on ne trouvait jusqu'à présent que chez les eucaryotes (ESP de l'anglais « *Eukaryotic Signature Proteins* »), notamment impliquées dans le trafic membranaire, la formation/transport de vésicules, l'ubiquitine et la formation de cytosquelette<sup>59</sup>.

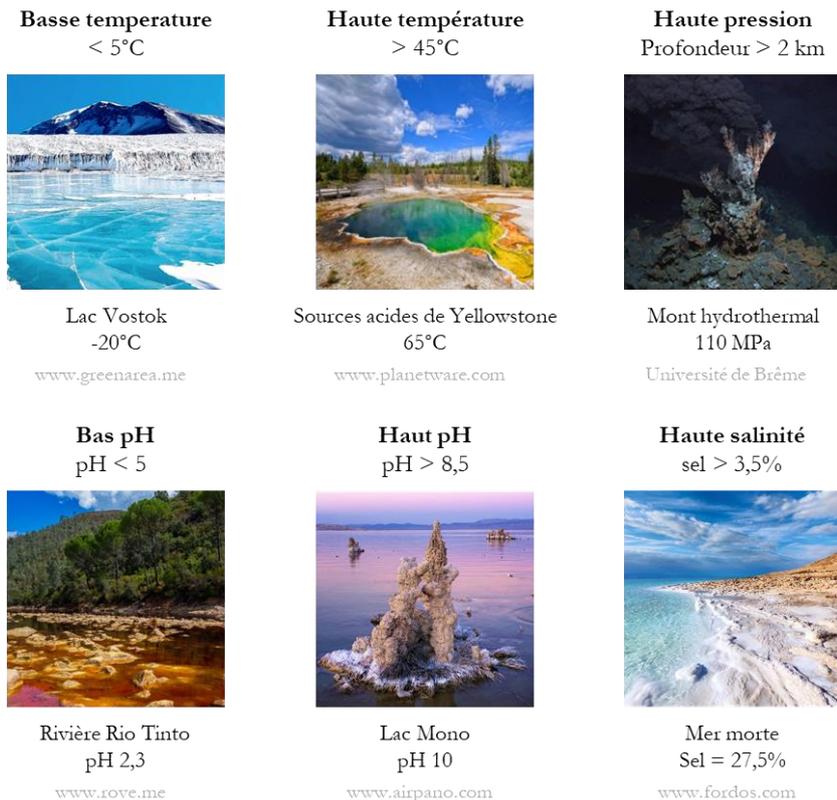
Tous ces travaux sont néanmoins controversés, car la classification des eucaryotes par rapport aux archées est fortement affectée par le choix des marqueurs moléculaires évolutifs utilisés pour construire l'arbre phylogénétique<sup>71-73</sup>. L'appartenance des eucaryotes au domaine Archaea et sa place éventuelle au sein de celui-ci continueront encore d'être débattus. Selon les marqueurs évolutifs utilisés, deux topologies d'arbres peuvent être obtenues : un arbre dans lequel Archaea et Eukarya ne forment qu'un seul domaine, et un où ils forment deux domaines distincts. Cette controverse a permis de relancer un intense débat sur l'origine des eucaryotes. La culture d'archées jusqu'alors incultivables permettra de mieux décrire leurs caractéristiques physiologiques et pourrait nous permettre de mieux comprendre l'origine et l'évolution des eucaryotes. Leur étude devrait permettre de tester les hypothèses émises concernant leur identité et de trancher entre les deux hypothèses d'arbre du vivant. En ce sens, la première archée Asgard a très récemment été isolée et cultivée<sup>74</sup>.

## 2.2. Importance des archées dans les écosystèmes

### 2.2.1. Les microorganismes extrémophiles

La majorité des organismes étudiés prospèrent dans des conditions dites « modérées » de température, pH, salinité, disponibilité de l'eau, niveau d'oxygène, pression et disponibilité en carbone. Ces paramètres dits « modérés » sont anthropocentriques, avec une température 37 °C, un pH de 7,4, salinité de 0,9 % à 3 %, et 1 atm (0,1 MPa) de pression<sup>75</sup>. Ces conditions sont confortables pour les êtres humains et sont idéales pour la culture d'*Escherichia coli* et sont souvent appelées « normales » ou « physiologiques ». Dans ce contexte, un environnement extrême peut être défini de façon anthropocentrique comme un environnement caractérisé par des conditions difficiles, au-delà des paramètres « modérés »<sup>76</sup>.

Dans les années 1950, les premiers travaux de microbiologie sur les environnements extrêmes ont mené à la découverte de microorganismes vivant à haute pression<sup>77</sup>. Au cours du siècle dernier, l'exploration d'environnements extrêmes tels que les cheminées hydrothermales profondes, les régions polaires glaciales ou les bassins géothermiques brûlants, a montré que la plupart d'entre eux étaient colonisés par des microorganismes, appelés « extrémophiles »<sup>78,79</sup>. Ces microorganismes tolèrent ou ont besoin de conditions physico-chimiques « extrêmes » pour leur croissance. Les conditions extrêmes peuvent être divisées en deux catégories : les conditions physiques extrêmes par exemple la température (de -20°C à 122°C), les rayonnements ionisants (rayonnements à haute énergie y compris les rayons X, les particules  $\alpha$ ) et la pression (jusqu'à 110 MPa) et les conditions géothermiques extrêmes, par exemple la dessiccation (limitation de l'eau), la salinité (de 10% à 40%, soit 100 à 400 g de sel par litre d'eau), le pH (de 0 à 12,8), les espèces oxygénées ou le potentiel redox (contenu en oxygène)<sup>76,80</sup> (figure 4).



**Figure 4 : Exemples d'environnements extrêmes.**

Les extrémophiles sont classés en fonction des conditions extrêmes dans lesquelles ils se développent. Le tableau 1 propose une liste non exhaustive des types d'extrémophiles. Ces microorganismes sont souvent tolérants à de multiples conditions extrêmes, et sont alors appelés "polyextrémophiles". Par exemple, *Halobacterium salinarum* NRC-1 est une archée capable de croître entre 2,6 et 5,1 NaCl et très résistante à la dessiccation, au rayonnement UV et à une limitation en oxygène<sup>81,82</sup>.

Il existe une grande variété d'organismes extrémophiles et c'est dans le domaine des Archaea que l'on retrouve les organismes les plus extrémophiles<sup>83-85</sup>.

Les organismes hyperthermophiles, qui nous intéressent tout particulièrement en biologie structurale (voir partie 2.3.2), ont besoin de températures avoisinant la température d'ébullition de l'eau pour se développer<sup>80</sup> (tableau 1). La plupart des zones terrestres et marines à haute température contenant de l'eau abritent des communautés microbiennes hyperthermophiles complexes<sup>86</sup>. Le premier organisme hyperthermophile, *Methanothermobacter feravidus*, a été découvert en 1981<sup>87</sup>. Seulement 3 ordres d'organismes hyperthermophiles connus aujourd'hui sont des bactéries : *Thermotogales*, *Aquificales* et *Feravidobacterium*<sup>88,89</sup>. Le reste des organismes hyperthermophiles sont des archées<sup>88</sup>.

**Tableau 1 : Classification non exhaustive et exemples d'extrêmophiles.** Adapté de Rothschild, L. J., & Mancinelli, R. L. (2001)<sup>76</sup>.

Paramètre environnemental	Type	Définition	Exemples
Température	Hyperthermophile	Croissance > 80°C	<i>Pyrolobus fumarii</i>
	Thermophile	Croissance 60°C-80°C	<i>Synechococcus lividus</i>
	Mésophile	Croissance 15°C-60°C	<i>Homo sapiens</i>
	Psychrophile	Croissance < 15°C	<i>Psychrobacter</i>
Radiation	Radiorésistant	Survit à des radiations ionisantes élevées	<i>Deinococcus radiodurans</i>
Pression	Piézosensible	Supporte > 1 atm avec une diminution de la croissance	<i>Escherichia coli</i>
	Piézotolérant	Supporte > 1 atm avec une croissance similaire	<i>Saccharomyces cerevisiae</i>
	Piézophile	Croissance optimale > 1 atm	<i>Pyrococcus abyssi</i>
Vide	/	Tolère le vide	<i>Hypsibius dujardini</i>
Dessiccation	Xérophile	Tolère l'anhydrobiose	<i>Artemia salina</i>
Salinité	Halophile	Croissance en fortes concentrations de sel (2-5 M NaCl)	<i>Halobacterium salinarum</i>
pH	Alkaliphile	Croissance pH 8,5-11	<i>Bacillus firmus</i>
	Acidophile	Croissance pH 0-5	<i>Sulfolobus acidocaldarius</i>
Oxygène	Anaérobe	Ne tolère pas l'oxygène	<i>Methanocaldococcus jannaschii</i>
	Microaérophile	Oxygène à des concentrations inférieures à celle de l'atmosphère	<i>Helicobacter pylori</i>
	Aérobe		<i>Homo sapiens</i>

### 2.2.2. Les archées dans la biosphère

Les archées ont longtemps été associées à l'extrêmophilie. Parmi les premières archées découvertes et étudiées, beaucoup vivent dans des conditions extrêmes. Une hypothèse populaire lors de leur découverte était que les archées étaient des organismes primitifs, proches de ceux vivants aux premiers âges de la Terre<sup>43</sup>. Comme mentionné dans la partie précédente, les archées<sup>90</sup> peuvent coloniser des environnements extrêmes comme des geysers ou monts hydrothermaux<sup>90</sup>, des environnements océaniques froids tels que les mers polaires<sup>91</sup>, des environnements salins tels que la mer morte et les lacs salés<sup>92</sup> et des environnements alcalins et acides<sup>93</sup>.

L'accès récent aux génomes de microbes non cultivables, qui représentent plus de 90% des communautés microbiennes, a permis de mettre en évidence une biodiversité archéenne insoupçonnée<sup>94,95</sup>. Ce sont ces travaux qui depuis dix ans ont permis de montrer que les archées sont présentes dans les sols, le permafrost, les sédiments, les océans<sup>94,96,97</sup> et représentent même une

partie du microbiome humain. Ces archées mésophiles appartiennent principalement aux organismes méthanogènes du phylum *Euryarchaeota* (dont *Methanobrevibacter smithii*) en ce qui concerne le microbiote intestinal<sup>98</sup> et au phylum des *Thaumarchaeota* pour le microbiote de la peau<sup>99,100</sup>.

Certains métabolismes tels que la méthanogénèse sont spécifiques des archées<sup>101</sup>. De plus, le métabolisme d'archées est un élément non négligeable du cycle géochimique de notre planète et influence de nombreux écosystèmes<sup>53,102</sup>. Les archées ont également un caractère symbiotique, notamment les archées méthanogènes présentes dans le tractus digestif des mammifères<sup>103,104</sup>, et les archées appartenant au phylum DPANN qui dépendraient principalement d'autres microorganismes pour satisfaire leur besoin en nucléotides, acides aminés et lipides<sup>105</sup>. Bien que l'on trouve des archées dans les différents microbiotes des êtres humains, aucune association claire entre archées et maladie humaine n'a pu être établie, bien que des soupçons grandissants s'accumulent et poussent à de nouvelles et intenses recherches. Une hypothèse populaire suggère que les méthanogènes stimulent la croissance par syntrophie d'organismes pathogènes. La syntrophie est une relation symbiotique obligatoire entre deux espèces où l'une se nourrit des produits de l'autre. Cette hypothèse est particulièrement privilégiée dans le cas de *Methanobrevibacter oralis* et de la maladie parodontale<sup>106</sup>. Plusieurs études également ont investigué le lien potentiel entre méthanogènes, notamment *Methanobrevibacter smithii*, et obésité<sup>107</sup>, et toutes ont noté que les sujets obèses possédaient un taux de méthanogènes significativement plus élevé que les sujets de poids normal. Une des théories est que les méthanoarchées de l'intestin pourraient augmenter la digestion des fibres et donc l'apport de calories et l'accumulation de dépôts gras menant à l'obésité chez les individus au régime alimentaire riche en fibres<sup>108,109</sup>.

Dans leur ensemble, les archées sont ubiquitaires et on estime qu'elles représenteraient 20 % de la biomasse terrestre<sup>110</sup>. Elles doivent donc être considérées, non plus comme des organismes marginaux, mais comme des acteurs à part entière de l'évolution du vivant, des microbiotes de nombreuses espèces animales, et des cycles géochimiques du carbone et de l'azote<sup>111</sup>.

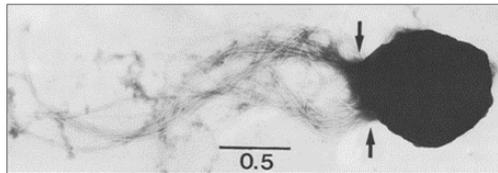
### 2.2.3. Présentation de l'organisme à l'origine des protéines étudiées dans la thèse : *Methanocaldococcus jannaschii*

L'archée *Methanocaldococcus jannaschii* dont proviennent les protéines étudiées dans ce projet a été isolée pour la première fois à partir d'un prélèvement de sédiments d'une cheminée hydrothermale (fumeur blanc), au niveau de la dorsale Est-Pacifique (21°N), à proximité de la pointe de la Basse Californie au Mexique, à une profondeur de 2600 mètres<sup>112</sup>. C'est un organisme anaérobie obligatoire, hyperthermophile et piézophile. Il peut se développer entre 50 et 90 °C, avec un optimum à 85 °C et un pH optimum proche de 6,0. Sa croissance est favorisée par une pression de 750 atm (75 MPa) à 86°C<sup>113</sup>. De plus, il ne peut croître qu'avec du dioxyde de carbone et de l'hydrogène comme sources d'énergie primaire, et produit alors du méthane comme sous-produit métabolique<sup>112</sup>.

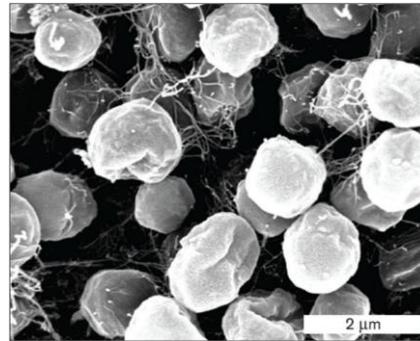
À son identification<sup>112</sup> *M. jannaschii* a d'abord été classée dans le genre des *Methanococcus* dans le phylum Euryarchaeota. Cependant, afin de mieux séparer les souches thermophiles et mésophiles de ce genre, une partie des espèces a été répartie dans un nouveau genre, *Methanocaldococcus* (*caldo*, signifiant « chaleur »), avec *M. jannaschii* comme souche type<sup>114</sup>.

La cellule de *M. jannaschii* se présente sous la forme de coques irrégulières, avec des plis dans sa paroi évoquant la forme d'un grain de raisin sec, d'environ 1,5  $\mu\text{m}$  de diamètre (figure 5). Elle possède deux faisceaux de flagelles polaires. Sa paroi est constituée d'une membrane cytoplasmique, et d'une couche de surface « S-layer » d'environ 20 nm d'épaisseur<sup>112,115</sup>.

**A.**



**B.**



**Figure 5 : Images de *Methanocaldococcus jannaschii* par microscopie électronique.** **A.** Archée *M. jannaschii* observée par microscopie électronique à transmission en coloration négative. Les dimensions sont données en  $\mu\text{m}$ . La cellule possède deux faisceaux de flagelles (flèches). La cellule apparaît sombre parce que le sel de coloration négatif a pénétré à travers l'enveloppe cellulaire brisée et a été piégé à l'intérieur de la cellule. Jones, W. J., *et al.* (1983)<sup>112</sup>. **B.** Image par microscopie électronique à balayage représentative de cellules de *M. jannaschii*. D'après Slonczewski, J. L., & Foster, J. W. (2013)<sup>116</sup>.

En 1996, le génome de *M. jannaschii* a été complètement séquencé<sup>117</sup>. Cela représentait le premier séquençage de génome complet d'un représentant du domaine des Archaea ainsi que d'un méthanogène. Au total, 1738 cadres ouverts de lecture ont été identifiés dans un grand chromosome et deux éléments extra-chromosomiques plus petits. Ce séquençage a permis de mettre en évidence les relations évolutives complexes entre les archées, les eucaryotes et les bactéries. La majorité des gènes liés à la production d'énergie, à la division cellulaire et au métabolisme se sont révélés similaires à ceux que des bactéries, tandis que la plupart des gènes impliqués dans la transcription, la traduction et la réplication sont apparus semblables à ceux que l'on trouve chez les eucaryotes.

### 2.3. Intérêt des protéines issues des archées hyperthermophiles pour les études de biologie structurale

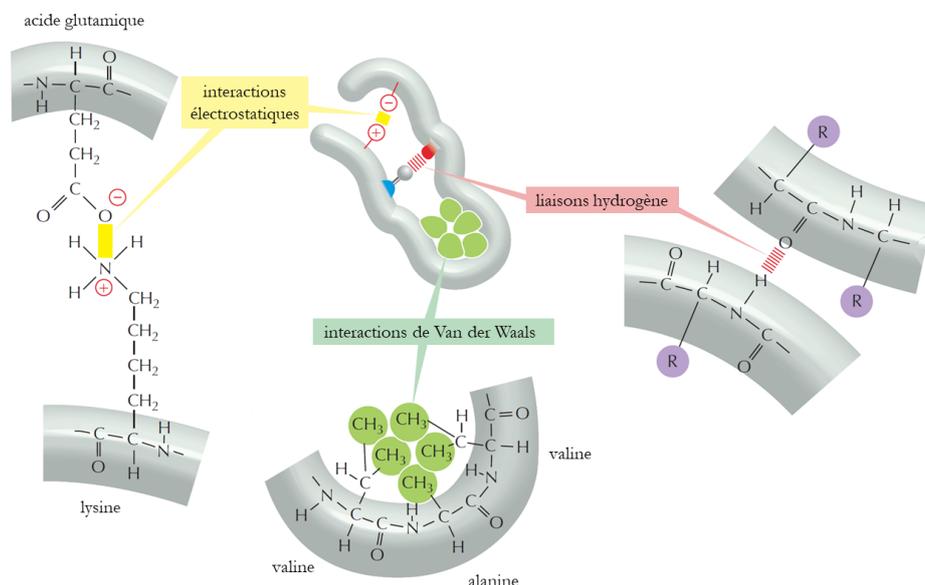
#### 2.3.1. Thermostabilité des protéines provenant d'archées hyperthermophiles

Dans leur habitat naturel, les organismes hyperthermophiles sont exposés à des températures de l'ordre de 100°C qui, combinées aux radiations et à l'accumulation de radicaux libres, représentent un défi pour l'intégrité des composants de la cellule : membranes, ADN, ARN et protéines. Dans ce chapitre nous présentons les propriétés de stabilité et de thermoactivation des protéines hyperthermophiles qui ont été mises à profit dans cette thèse pour réaliser les études biophysiques en temps réel sur le protéasome archéen.

Pour toutes les enzymes, la catalyse est un processus thermo-dépendant : si l'environnement est trop froid les mouvements sont ralentis et ne permettent plus une catalyse efficace, si la température est trop élevée, la structure de l'enzyme peut être altérée, jusqu'à la dénaturation. Beaucoup d'enzymes issues de microorganismes hyperthermophiles se distinguent par leur aptitude à réaliser

des réactions biochimiques à des températures très élevées. Les protéines sont des objets dynamiques et il apparaît que leur aptitude à réaliser des réactions biochimiques à haute température résulte d'une rigidification accrue de l'édifice moléculaire combinée à une flexibilité locale qui permet les mouvements nécessaires à l'accomplissement de la catalyse. Ainsi, des études de dynamique moléculaire réalisées sur des cellules entières ou des protéines isolées ont montré une rigidification des édifices macromoléculaires thermophiles ou hyperthermophiles comparativement à leurs homologues thermophiles<sup>118,119</sup>. Ce phénomène est une des raisons pour lesquelles la plupart des enzymes thermophiles ne se dénaturent pas à haute température et sont extrêmement robustes à température modérée, ce qui rend leur manipulation aisée à température ambiante.

Les bases moléculaires responsables de la stabilité des édifices macromoléculaires ont été bien décrites. On n'observe pas de différences importantes entre une enzyme thermophile vis-à-vis de son homologue mésophile<sup>88</sup>. Les séquences de protéines homologues hyperthermophiles et mésophiles sont typiquement similaires à 40 à 85%<sup>120</sup>, leurs structures tridimensionnelles sont superposables<sup>121-123</sup>, et elles ont les mêmes mécanismes catalytiques<sup>124,125</sup>. L'augmentation de la thermostabilité doit plutôt se trouver dans un petit nombre de mutations très spécifiques qui, souvent, ne respectent pas de règles évidentes comprises aujourd'hui<sup>126</sup>. Ces mutations ont pour effet de moduler des paramètres connus pour être à la base du repliement des protéines. L'effet hydrophobe est l'un des principaux déterminants de la stabilité des protéines. Le repliement des protéines est principalement dû à l'enfouissement des groupes non polaires (hydrophobes) dans une structure globulaire, loin des molécules d'eau du solvant<sup>127</sup>. Une fois repliée, la stabilité est encore modulée par les interactions entre les groupes qui ont été mis en contact. Dans les protéines, les interactions Van der Waals et les liaisons hydrogènes sont les plus importantes. Les liaisons ioniques, aromatiques et certains ponts disulfures contribuent à la stabilité. La figure 6 représente les différents types de liaisons non covalentes qui aident les protéines à se replier.



**Figure 6 : Trois types de liaisons non covalentes aident les protéines à se replier.** Bien qu'une seule de ces liaisons soit assez faible, plusieurs d'entre elles agissent ensemble pour créer un solide arrangement de liaisons, comme dans l'exemple montré. R est utilisé comme désignation générale d'une chaîne latérale d'acides aminés. Adapté de Alberts, B., *et al.* (2008)<sup>11</sup>.

Des études comparatives ont été menées entre homologues thermophiles et mésophiles pour déterminer les mécanismes d'adaptation intrinsèques aux protéines à partir des séquences primaires (grâce au séquençage complet des génomes) ou des structures tridimensionnelles<sup>128-130</sup>. Plusieurs facteurs à l'origine des propriétés dans des environnements thermiques extrêmes ont pu être extraits. Comme mentionné précédemment, les interactions hydrophobes sont essentielles dans le repliement et la stabilité des protéines. Les interactions coopératives au sein du cœur hydrophobe pourraient jouer un rôle crucial dans l'augmentation de la stabilité thermique<sup>131</sup>. Des interfaces essentiellement hydrophobes ont été observées entre les sous-unités d'enzymes thermostables<sup>132,133</sup>.

Les interactions électrostatiques jouent également un rôle clé dans la thermostabilité. Ainsi, une surface réduite exposée au solvant est autre un facteur qui peut favoriser la stabilité conformationnelle à haute température<sup>134</sup>. De larges réseaux d'interactions ioniques ou de liaisons hydrogène sont présents aux interfaces entre sous-unités des protéines thermostables afin d'éviter l'exposition au solvant<sup>135-137</sup>. En particulier, le seul facteur que l'on retrouve à ce jour dans l'ensemble des structures de protéines thermostables est l'augmentation du nombre d'interactions ioniques et leur organisation en réseau<sup>88,138</sup>. Aux températures élevées, les molécules d'eau sont de plus en plus désordonnées et n'interagissent plus avec les résidus chargés, qui se réorientent pour former des interactions électrostatiques, qui sont alors stabilisantes<sup>139</sup>. Dans certaines protéines d'organismes hyperthermophiles, des interactions multiples entre résidus aromatiques favorisent la thermostabilité<sup>129</sup>. Le dipôle des résidus aromatiques (Phe, Trp et Tyr), crée par la charge partielle négative des électrons  $\pi$  et par la charge partielle positive et sur les extrémités C-H<sup>140</sup>, permet des interactions entre résidus aromatiques, et également entre résidus aromatiques et les chaînes latérales des résidus Lys et Arg. Lorsque la température augmente, l'oxydation des Cys est facilitée et conduit à la cassure des ponts disulfures. Pour cette raison, la proportion des Cys est en général faible dans les protéines thermostables. Pour les protéines thermostables dont la proportion de Cys est équivalente ou supérieure à celle des mésophiles, les Cys impliquées dans les ponts disulfures sont protégées grâce à leur enfouissement dans la protéine<sup>141,142</sup>. Il est de plus en plus rapporté que les liaisons hydrogène, ioniques, électrostatiques, aromatiques et les ponts disulfures, particulièrement dans les surfaces exposées aux solvants des protéines sont importantes pour la stabilité thermique<sup>143,144</sup>. Par conséquent, le rapport entre acides aminés chargés et non chargés pourrait être un facteur important dans la stabilité de certaines protéines. La présence d'acides aminés chargés, en particulier à la surface de la protéine, augmenteraient les interactions ioniques qui assureraient la stabilité thermique des protéines<sup>138,144,145</sup>.

Les boucles et parties mobiles des protéines sont également des zones critiques pour la thermostabilité. La grande majorité des boucles dans une structure relie des éléments de structure secondaire. La proportion de résidus Pro et Gly dans les boucles, qui restreignent et augmentent respectivement la flexibilité conformationnelle de la chaîne polypeptidique, semble dépendre de la température de fonctionnement de l'enzyme. Les mutations de type Gly  $\rightarrow$  X ou X  $\rightarrow$  Pro (où X représente un acide aminé quelconque) diminuent le nombre de conformations accessibles à la chaîne polypeptidique, et ont été recensées dans plusieurs protéines thermostables<sup>88</sup>. Pour mieux résister aux hautes températures, les éléments les plus instables (boucles et régions N et C terminales) des protéines hyperthermophiles interagissent plus avec le reste de la protéine au travers de liaisons hydrogènes<sup>146</sup>, liaisons ioniques<sup>134,136,147</sup>, et/ou ponts disulfures (particulièrement vrai pour les *Crenarchaeae*)<sup>148</sup> supplémentaires. Les boucles exposées en surface peuvent également être raccourcies<sup>144,149,150</sup>.

Les études de stabilité protéique sont généralement menées sur des solutions de protéines diluées *in vitro*. Cependant, ces conditions ne représentent pas la situation réelle à l'intérieur d'une cellule. Ainsi, outre les modifications intrinsèques des protéines que nous avons décrites précédemment, il a été démontré que les facteurs extrinsèques contribuent également à la stabilité des protéines dans le contexte d'une cellule biologique<sup>151</sup>. Souvent c'est la liaison à des cofacteurs, au substrat ou à des partenaires au sein d'un complexe supramoléculaire qui vont permettre la stabilisation du système. L'encombrement macromoléculaire et les petites molécules naturelles comme les métabolites et les dérivés sucres jouent également un rôle important dans la stabilité des protéines hyperthermophiles<sup>152,153</sup>.

Alors que la plupart des espèces vivant sur Terre ne peuvent supporter des variations de température de plus de quelques dizaines de degrés, les organismes hyperthermophiles sont exposés à des variations des paramètres physique et chimique de très grande amplitude. En particulier, les organismes provenant des monts hydrothermaux tel que *M. jannaschii* sont soumis aux variations des flux hydrothermaux et à des chocs thermiques violents de 300°C à 4°C<sup>154</sup>. Pour résister, les hyperthermophiles présentent une variété d'adaptations cellulaires permettant de garantir l'intégrité du génome et du protéome. Ainsi, l'étude des réponses transcriptionnelles chez *M. jannaschii* a montré que les voies de repliement et de dégradation des protéines étaient mobilisées pendant un stress thermique chaud (de 85°C à 95°C) ou froid (de 85°C à 65°C)<sup>155</sup>. Chez les archées, plusieurs indices suggèrent que le protéasome peut participer à la réponse au stress et au système de contrôle de qualité des protéines<sup>156</sup>. La plupart des protéasomes purifiés des archées sont en effet extrêmement stables et actifs dans des conditions non favorables à d'autres protéines issues du même organisme. Ainsi le protéasome halophile 20S d'*Haloarcula marismortui* est toujours stable et actif à des concentrations en sel pouvant descendre jusqu'à 0,5 M, valeur qui constitue des conditions dénaturantes pour d'autres protéines halophiles<sup>157</sup>. De même, le protéasome d'autres archées résiste à des agents oxydants ou à des températures qui sont supérieures à l'optimum de croissance des microorganismes<sup>88</sup>. Ainsi, le système PAN-protéasome qui est l'objet d'étude de cette thèse est lui-même un acteur important de l'adaptation extrémophile.

En conclusion, les mécanismes responsables de la thermostabilité sont divers et chaque famille de protéines utilise des stratégies différentes en fonction des caractéristiques propres à son repliement. Souvent il suffit de peu de substitutions d'acides aminés pour modifier drastiquement la thermostabilité d'une protéine<sup>126</sup>. Ces modifications ne suffisent pas toujours et des mécanismes cellulaires contribuent aussi à la thermoadaptation.

Les paragraphes précédents ont fourni une image statique de la thermostabilité des protéines. Toutefois la stabilisation des édifices ne suffit pas à expliquer le fonctionnement des enzymes à haute température. L'état natif d'une protéine dans les conditions physiologiques n'est pas une espèce moléculaire unique, mais est mieux décrit par un ensemble de conformations statistiquement peuplées. Dans ce contexte, la « flexibilité » fait référence à ces conformations fluctuantes d'énergie comparable qui peuvent être converties en une série de petits changements. Plus la flexibilité est grande, plus le nombre de conformations est important et plus la barrière énergétique pour l'interconversion entre elles est faible<sup>127</sup>. Le concept de « l'état correspondant »<sup>158,159</sup> postule que les homologues des protéines font preuve d'une flexibilité comparable à leurs températures physiologiques pour remplir leur fonction biologique. Des études structurales et phylogénomiques réalisées au laboratoire sur la famille enzymatique les malate-

lactate deshydrogénases suggèrent que les substitutions d'acide aminés surviennent au cours de l'évolution sur certains « point chauds » des structures tertiaires et quaternaires pour parvenir à modifier le paysage énergétique d'une protéine de manière à préserver leur flexibilité fonctionnelle malgré la rigidification apparente de l'ensemble du système. Ainsi, l'adaptation extrêmophile semble résulter d'un compromis entre stabilité de la structure et une dynamique locale appropriée<sup>119</sup>. Ces mécanismes expliquent en partie pourquoi les protéines des hyperthermophiles sont peu actives des températures « basses ». Il faut toutefois noter que certaines familles de protéines hyperthermophiles maintiennent une activité biochimique significative dans une large gamme de températures. C'est le cas du protéasome PAN-20S étudié dans cette thèse. Celui-ci possède des optima pour ses différentes activités enzymatiques (hydrolyse de l'ATP, dépliement des protéines substrats, protéolyse) que l'on situe au-delà des 80°C. Il conserve toutefois une activité ralentie dans une large plage de température sub-optimale. C'est cette caractéristique que nous avons exploitée pour ralentir la réaction et la rendre compatible avec des études biophysiques en temps réel.

### 2.3.2. Intérêt des protéines hyperthermophiles pour les études biochimiques et biophysiques

Le laboratoire ELMA (de l'anglais « *Extremophiles and Large Molecular Assemblies* ») à l'Institut de Biologie Structurale (IBS) de Grenoble étudie principalement des protéines formant de grands assemblages macromoléculaires, issues d'archées hyperthermophiles. Les protéines hyperthermophiles sont particulièrement intéressantes et de plus en plus utilisées pour des études structurales impliquant de grands assemblages. Tout en ayant une structure quaternaire similaire à leurs homologues eucaryotes, elles présentent de grands avantages pour la biologie structurale des grands assemblages.

Comme mentionné précédemment, la proximité évolutive entre les archées et les eucaryotes font de nombre de leurs protéines des homologues structuraux. Cependant, les machineries cellulaires issues d'archées (hyperthermophiles ou non) sont en général moins complexes en terme de sous-unités que leur homologues eucaryotes et représentent donc des cibles intéressantes pour comprendre les mécanismes dans lesquelles elles sont impliquées<sup>156</sup>. C'est le cas du système PAN-20S, qui est une version moins complexe que le protéasome 26S des eucaryotes (voir partie 5).

Les propriétés des protéines issues d'organismes hyperthermophiles permettent de mettre en œuvre des protocoles de purifications à température ambiante. De plus, les protéines thermophiles sont généralement stables contre les produits chimiques dénaturants tels que les détergents, les surfactants, les agents oxydants et les protéases<sup>160</sup>. Dans le cas d'expression de protéines recombinantes, une première étape de purification des extraits protéiques d'*Escherichia coli* par un traitement thermique permet d'éliminer les contaminants mésophiles. Enfin, le rendement en protéine est généralement haut, ce qui est un avantage non négligeable puisque la plupart des méthodes de biologie structurale (cristallographie, RMN, SAXS, SANS) nécessite une grande quantité en protéine d'intérêt.

Dans le cas d'études d'enzymes, il est possible de contrôler l'activité en utilisant une température plus basse que la température optimale d'activité. Cette approche est particulièrement pertinente pour l'étude de machineries cellulaires qui travaillent sur des substrats complexes (protéines, ADN, ARN...) dans de longues gammes de temps (moteurs moléculaires, chaperonnes, protéasomes,

etc.). On pourra non seulement activer les réactions enzymatiques par la température lors d'études biochimiques et structurales, mais aussi ralentir ces réactions pour qu'elles soient analysables aux échelles de temps nécessaires aux techniques utilisées, notamment le SANS résolu en temps (TR-SANS), où une mesure est réalisée toutes les 30 secondes<sup>40</sup>.

Dans le cadre de ce projet, l'utilisation de la machinerie protéolytique PAN-20S issue de l'archée hyperthermophile *M. jannaschii* comporte ces avantages pratiques. En effet, les études par TR-SANS avec de la deutération et de la variation de contraste ont été mises en place et réalisées, couplées avec des études par spectroscopie de fluorescence (voir chapitre 7). Ces expériences nécessitent une grande quantité de solution protéique ainsi qu'une stabilité du système et un contrôle de son activité par la température. Ces études auraient été difficilement réalisables avec un système mésophile.

### 3. Présentation des systèmes protéasomes

#### 3.1. Importance des protéases intracellulaires

L'ensemble des protéines d'une cellule (le protéome) détermine la fonction spécifique et le comportement des cellules biologiques à tout moment. Le protéome doit donc être régulé et remodelé étroitement et spécifiquement tout au long du cycle de vie cellulaire, pendant la différenciation cellulaire dans les organismes multicellulaires et en réaction à des stimuli internes ou externes<sup>8</sup>. Bien que la régulation de l'expression génique représente un premier mécanisme efficace pour ajuster le type et le nombre de protéines produites par une cellule, la dégradation contrôlée des protéines permet aux cellules de remodeler rapidement le protéome existant lors d'étapes ultérieures<sup>11</sup>. La dégradation intracellulaire des protéines (la protéolyse intracellulaire) est donc un processus essentiel qui a lieu dans tous les organismes, de la cellule eucaryote aux microorganismes.

L'idée que tous les composants de la matière vivante, en particulier les protéines, sont dans un état perpétuel de renouvellement, n'est apparue que relativement récemment<sup>1</sup>. Au milieu du 20<sup>ème</sup> siècle, il était admis que la protéolyse intracellulaire était un processus aléatoire. Plus de cinquante ans d'intenses études ont finalement prouvé que la dégradation des protéines est un processus très complexe et étroitement régulé dans le temps et l'espace, qui joue un rôle extrêmement important dans la grande majorité des voies métaboliques<sup>161</sup>. La protéolyse intracellulaire est en effet un processus faisant intervenir des réseaux de partenaires et des voies multiples<sup>162,163</sup>.

Au cours de la vie d'une cellule, quasiment toutes ses protéines sont plusieurs fois renouvelées par dégradation et resynthèse<sup>164</sup>. Les vitesses de renouvellement sont très différentes d'une protéine à l'autre, avec des demi-vies allant de quelques minutes jusqu'à plusieurs jours<sup>8</sup>.

Certaines protéines spécifiques ont un taux de dégradation élevé, y compris celles impliquées dans des régulations métaboliques, dans la transcription, la division cellulaire, la différenciation et la réparation de l'ADN<sup>165</sup>. La dégradation ciblée de ces protéines est donc nécessaire pour réguler les concentrations intracellulaires des enzymes qui contrôlent toutes les réactions métaboliques. La protéolyse intracellulaire est aussi vitale pour permettre aux cellules de s'adapter aux changements environnementaux grâce au contrôle du contenu général des protéines<sup>166</sup>.

Les protéines dysfonctionnelles ou superflues doivent également être identifiées, ciblées et détruites par les machines protéolytiques cellulaires<sup>2,167</sup>. Les protéines anormales comprennent entre autres

celles récemment synthétisées et qui ne se replient pas correctement ainsi que les protéines qui s'accumulent dans la cellule après exposition à un stress environnemental ou des dommages oxydatifs. Elles comprennent également les protéines produites avec des erreurs pendant la transcription ou la traduction (tableau 2)<sup>8</sup>.

**Tableau 2 : Types d'anomalies pouvant toucher les protéines et cause de ces anomalies.** D'après Goldberg, A. L. (2003)<sup>8</sup>.

Type d'anomalie	Causes
Protéines incomplètes	Mutations non-sens, incorporation de la puromycine, terminaisons prématurées, clivages protéolytiques
Protéines faux-sens	Mutations, incorporation d'analogues d'acides aminés, erreurs biosynthétiques
Sous-unités libres de complexes multimériques	Sous-unités synthétisées en excès
Dommages post synthétiques	Radicaux oxygénés, dénaturation intracellulaire
Ingénierie génétique	Fusions de gènes, changement de cadre de lecture, localisations incorrectes

Une des caractéristiques des protéines dénaturées est qu'elles s'agrègent et deviennent insolubles en raison de la tendance des domaines hydrophobes à s'associer entre eux. De même, lorsque la production des protéines mal repliées dépasse la capacité de dégradation de la cellule, ces polypeptides forment souvent des agrégats intracellulaires avant leur dégradation rapide<sup>168,169</sup>. La formation de ces agrégats intracellulaires se produit lorsque les systèmes protéolytiques de la cellule ainsi que les chaperonnes moléculaires, qui normalement empêchent l'agrégation (par exemple Hsp 70/40)<sup>170,171</sup> ou resolubilisent (Hsp 104)<sup>172</sup> les micro-agrégats, ne peuvent pas faire face au taux de production des molécules dépliées. Des corps d'inclusions similaires se retrouvent dans diverses maladies neurodégénératives, dans lesquelles ces agrégats sont ubiquitinés et physiquement associés aux protéasomes, suggérant fortement un échec dans la machinerie de dégradation de la cellule<sup>168,173</sup>.

On parle généralement de protéases ou de peptidases pour définir les enzymes protéolytiques. Une protéase est une enzyme capable de dégrader une protéine en hydrolysant les liaisons peptidiques liant les résidus, tandis que la définition d'une peptidase est généralisée à tout type de peptide<sup>174</sup>. Les protéases et les peptidases représentent des constituants protéiques majeurs des cellules (3 à 10 % des protéines) et se répartissent en deux groupes selon leur besoin ou non en énergie. On distingue les protéases ATP-dépendantes et les peptidases ATP-indépendantes. Les protéases ATP-dépendantes interviennent essentiellement dans la dégradation des protéines dites régulatrices et dans l'élimination des protéines défectueuses. Elle est assurée par des machines protéolytiques qui séquestrent des sites actifs dans une chambre protégée pour le clivage des liaisons peptidiques. Ces enzymes sont connues en tant que protéases AAA+ (de l'anglais « *ATPases Associated with diverse cellular Activities* »), en raison de la présence d'un complexe de régulation AAA+ qui reconnaît les substrats spécifiques et utilise l'énergie chimique de l'hydrolyse de l'ATP pour déplier mécaniquement la protéine cible et assurer son transfert dans la chambre de dégradation<sup>175</sup>. L'inhibition ou l'inactivation de la majorité des protéases ATP-dépendantes entraîne une sensibilité au stress, et l'accumulation des protéines défectueuses dans le cytoplasme. De plus, la majorité de ces protéases sont induites en conditions de stress<sup>176,177</sup>. Les peptidases ATP-indépendantes

interviennent quant à elles pour réaliser l'étape finale de la protéolyse *in vivo*. En ce sens, elles sont chargées de dégrader les petits oligopeptides issus de la protéolyse dépendante de l'énergie en acides aminés réutilisables. Toutefois, un grand nombre de peptidases sont impliquées dans la régulation du fonctionnement cellulaire (activation et maturation des protéines, signalisation, etc.)<sup>178</sup>.

Comme la synthèse et la dégradation des protéines sont des processus qui consomment beaucoup d'ATP, des mécanismes hautement sélectifs garantissent que seules les protéines affectées à la dégradation soient éliminées. Si cette régulation échoue, les cycles inutiles de synthèse, dégradation et renouvellement des protéines consomment l'ATP des cellules et les protéines non fonctionnelles s'accumulent sous forme d'agrégats cytotoxiques<sup>8,179-181</sup>.

### 3.2. Mécanisme général de la protéolyse par les protéasomes

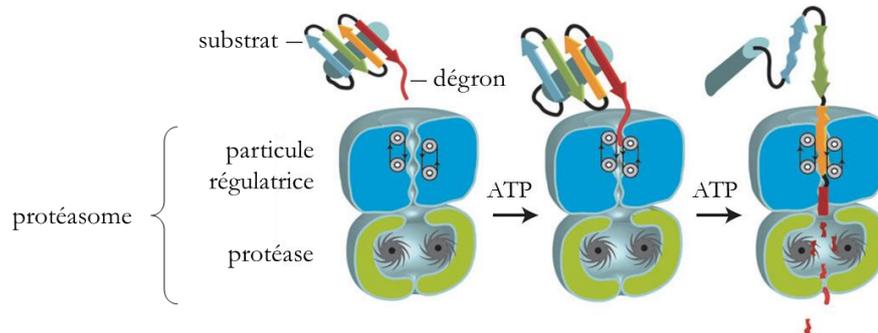
La machinerie qui détruit de façon ciblée les protéines est le protéasome, une protéase dépendante de l'ATP présente en de nombreuses copies dispersées dans la cellule<sup>182</sup>. Par exemple, les protéasomes représentent jusqu'à 1 % des protéines totales dans certains types de cellules de mammifères<sup>183</sup>. Les protéasomes ont été identifiés chez les bactéries, eucaryotes et archées<sup>184</sup>. Chaque protéasome est constitué d'un cylindre central creux avec un canal central, formant un ensemble composé de quatre anneaux heptamériques empilés avec une stœchiométrie  $\alpha_7\beta_7\beta_7\alpha_7$ <sup>185</sup>. Cet arrangement séquestre les sites actifs dans un compartiment central formé par les sous-unités, limitant l'accès aux seules protéines qui présentent un signal de dégradation. Le nucléophile catalytique, un résidu thréonine<sup>186</sup>, est le résidu N-terminal de trois sous-unités  $\beta$  ( $\beta_1$ ,  $\beta_2$  et  $\beta_5$ ) après auto-catalyse d'un propeptide<sup>187</sup>. Leurs poches de liaison aux substrats ont des préférences pour les acides aminés hydrophobes, basiques et acides et possèdent une activité de clivage de type chymotrypsines, trypsines et caspases, respectivement. Les sous-unités  $\alpha$  guident les substrats vers la chambre protéolytique à travers un canal central étroit nommé «  $\alpha$ -annulus », qui est obstrué à l'état de repos par leurs extrémités N-terminales flexibles qui forment une « porte » fermée (chez les eucaryotes, voir partie 5.2.2 pour plus de détails)<sup>188</sup>.

Étant donné que les sites catalytiques de la protéolyse sont séquestrés à l'intérieur de la particule catalytique, ils sont difficilement accessibles aux protéines natives repliées<sup>182</sup>. Chaque extrémité du cylindre peut être associée à un grand complexe protéique appelé « particule régulatrice » qui contient à sa base un anneau composé de six sous-unités (voir partie 4). Ces particules régulatrices appartiennent à la famille des protéines AAA+ et sont contrôlées par l'hydrolyse de l'ATP. Elles sont nécessaires pour déplier et transférer les substrats dans les particules catalytiques centrales à travers un canal central étroit<sup>175</sup>.

Les interfaces entre l'anneau hexamérique ATPase et la protéase heptamérique suggèrent un décalage de symétrie qui implique des changements dynamiques sous-jacents aux mécanismes de transfert des protéines à travers les canaux centraux de la particule régulatrice et de la protéase<sup>175</sup>.

Une machine aussi efficace que le protéasome nécessite d'être étroitement régulée. Elle doit être en particulier capable de distinguer les protéines à éliminer de celles qui sont nécessaires à la cellule. Les particules régulatrices du protéasome agissent comme une porte à l'entrée de la particule protéolytique, et seules les protéines marquées y généralement sont transférées. En général, des caractéristiques structurales superficielles spécifiques, dites dégrons ou étiquette d'adressage, caractérisent les protéines destinées à la destruction et sont reconnues par les cofacteurs associés aux machines protéolytiques<sup>182</sup>. Les dégrons peuvent varier considérablement, allant des résidus

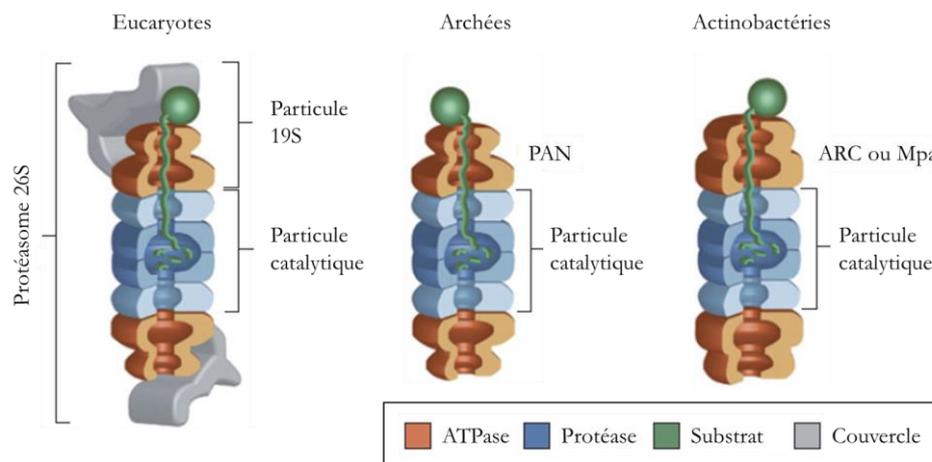
d'acides aminés phosphorylés aux terminaisons amino ou carboxyle exposées. Souvent, les dégrons résultent de la fixation par modification post-traductionnelle d'une petite molécule sur la protéine à dégrader<sup>10</sup>. La figure 7 illustre de façon simplifiée le mécanisme de dégradation des substrats par la machinerie protéolytique.



**Figure 7 : Mécanisme général de la protéolyse par les protéasomes.** Le dégron (étiquette d'adressage) est initialement reconnu par la particule régulatrice AAA+. Une région non structurée du substrat s'engage dans le canal de l'anneau AAA+. Les cycles de fixation et d'hydrolyse de l'ATP au sein de l'anneau AAA+ aboutissent à des changements conformationnels qui déplient le substrat et permettent son passage à travers le canal dans la chambre catalytique de la protéase. Adapté de Sauer, R. T., & Baker, T. A. (2011)<sup>182</sup>.

Une propriété importante du protéasome, et une des raisons de sa complexité, est la processivité de son mécanisme. Contrairement à une protéase « simple » qui coupe la chaîne polypeptidique d'un substrat juste une fois avant de se dissocier, le protéasome maintient la protéine substrat entière sans la relâcher jusqu'à ce qu'elle soit complètement dégradée en oligopeptides par des réactions successives<sup>189</sup>.

Les protéasomes sont hautement conservés au sein du vivant, à la fois au niveau structural et fonctionnel. Nous considérerons dans la suite de cette étude une division du monde du vivant en trois domaines, eucaryotes, bactéries et archées, comme illustré dans la figure 8.



**Figure 8 : Schéma des protéasomes des eucaryotes, archées et bactéries.** Tous les protéasomes sont composés d'une particule catalytique 20S (CP) formée de quatre des anneaux heptamériques empilés de sous-unités  $\alpha$  et  $\beta$ . Les CP peuvent s'associer aux particules régulatrices AAA+ ATPases, qui déplient et transfèrent les protéines du substrat dans la CP par un mécanisme dépendant de l'ATP. Chez les eucaryotes, six sous-unités différentes d'ATPase forment l'anneau hexamérique de la coiffe 19S, qui s'associe aux CP pour former 26S protéasomes. Le complexe 19S peut être séparé en sous-complexes de base et de couvercle. Les particules régulatrices AAA+ des archées (PAN) et des actinobactéries (ARC ou MPA) s'assemblent en cycles homo-hexamériques et sont associées aux CP in vitro, mais les

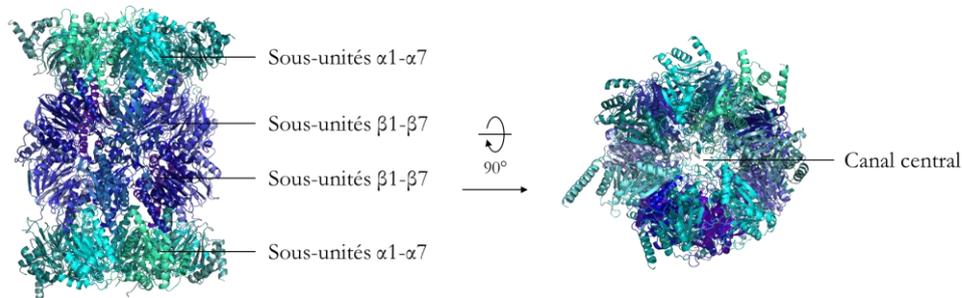
preuves que ces ATPases interagissent avec leurs CP correspondantes *in vivo* sont limitées. Adapté de Maupin-Furlow, J. (2012)<sup>156</sup>.

### 3.3. Le protéasome 26S des eucaryotes

Chez les eucaryotes, le protéasome 26S se compose d'environ 40 sous-unités protéiques différentes, divisées en une particule catalytique (protéasome 20S) et une particule activatrice (19S).

#### 3.3.1. La particule catalytique 20S

La partie centrale du protéasome 26S (avec « S » pour « Svedberg », l'unité de mesure du taux de sédimentation par ultracentrifugation), appelée particule catalytique ou protéase 20S. La particule catalytique 20S est formée par un empilement de quatre anneaux heptamériques : deux anneaux  $\beta$  intérieurs identiques, constitués chacun de sept sous-unités  $\beta$  différentes ( $\beta$ 1-7) et deux anneaux  $\alpha$  extérieurs identiques constitués chacun de sept sous-unités  $\alpha$  différentes ( $\alpha$ 1-7) comme représenté dans la figure 9.



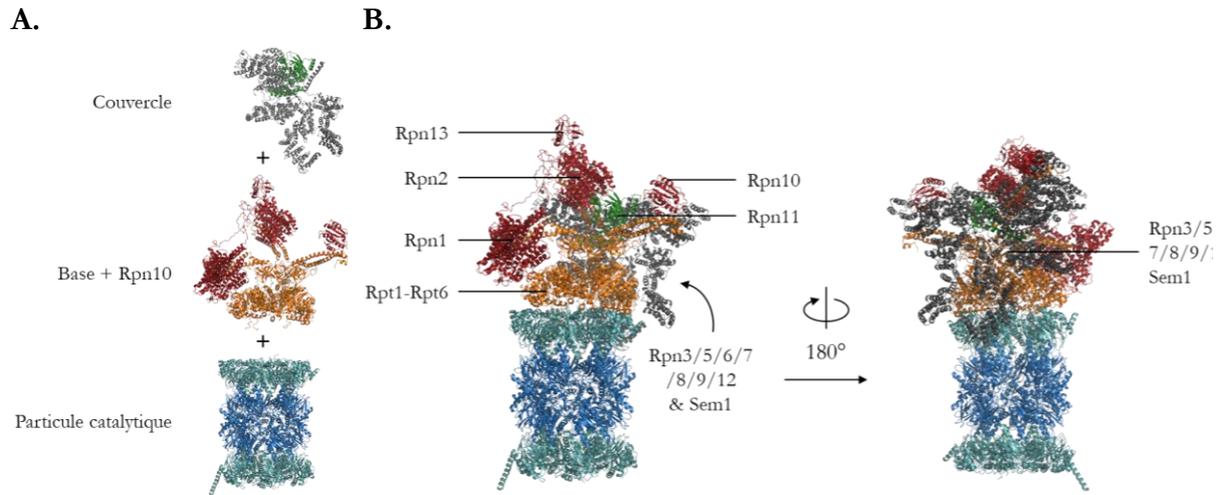
**Figure 9 : Structure de la particule catalytique 20S du protéasome 26S eucaryote.** Vues latérale (gauche) et supérieure (droite) de la particule catalytique 20S du protéasome 26S de *Saccharomyces cerevisiae* (PDB 6FVT). Les sous-unités sont représentées de différentes couleurs.

#### 3.3.2. La particule activatrice 19S

Chez les eucaryotes, les protéasomes 20S sont connus pour s'associer avec trois différentes familles d'activateurs qui donnent accès à la chambre protéolytique centrale. La famille la plus conservée contient l'activateur AAA+ 19S eucaryotes (« particule régulatrice/PA700 »). L'activateur 19S est constitué de plusieurs sous-unités avec des compositions et des fonctions distinctes. Il reconnaît les substrats poly-ubiquitinés et peut se lier à l'une ou aux deux extrémités du protéasome 20S pour former le protéasome 26S. Les deux autres familles d'activateurs sont non-ATP dépendants et sont moins conservés que les activateurs ATP-dépendants. Il s'agit des complexes 11S (PA28/REG, PA26) et PA200/Blm10. Ces facteurs, leurs substrats ainsi que leurs fonctions biologiques sont moins étudiés, bien que les mécanismes qu'ils utilisent pour activer le protéasome aient été caractérisés. A l'inverse des activateurs 11S et Blm10/PA200, la fonction biologique de la particule régulatrice eucaryote 19S est bien établie. Elle est également responsable de la sélection, du dépliement et du transfert des protéines substrat à éliminer, en particulier ceux modifiés par conjugaison à une chaîne poly-ubiquitine<sup>190</sup>.

Chaque côté de la particule catalytique 20S est couverte d'une particule régulatrice 19S qui régule la fonction protéolytique de la protéase. Les particules sont adjacentes au protéasome 20S en raison du nombre différent de leur sous-unités en contact respectives<sup>191,192</sup> (figure 10). La particule

régulatrice 19S contient 19 sous-unités et est divisée en deux sous-complexes, la base et le couvercle<sup>164</sup>.



**Figure 10 : Structure du protéasome 26S eucaryote.** Le protéasome 26S est composé de trois sous-complexes : la particule catalytique (avec les sous-unités  $\alpha$  en bleu clair et les sous-unités  $\beta$  en bleu foncé), la base (avec les sous-unités AAA+ Rpt1-Rpt6 en orange et Rpn2 et les sous-unités de liaison à l'ubiquitine Rpn1, Rpn2 et Rpn13 en rouge), et le couvercle (avec Rpn3, Rpn5, Rpn6, Rpn7, Rpn8, Rpn9, Rpn12 et Sem1 en gris et la DUB Rpn11 en vert). Le récepteur d'ubiquitine Rpn10 est représenté avec la base (rouge). La base et le couvercle forment la particule 19S, qui peut s'associer aux deux extrémités de la particule catalytique 20S. Par clarté, une seule particule 19S a été représentée dans cette figure. **A.** Les trois sous-complexes sont représentés individuellement. **B.** La structure entière du protéasome 26S est représentée (PDB 6FYT). L'orientation à gauche permet de voir l'entrée du canal central et Rpn11, et l'orientation à droite, tournée de 180°, souligne le sous-ensemble du couvercle.

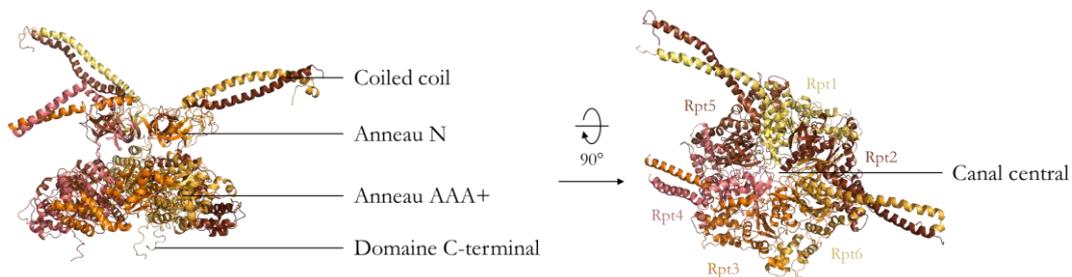
Le tableau 3 récapitule la composition en sous-unités des différents sous-complexes de la particule 19S chez la levure et chez l'Homme, ainsi que la fonction de chacune des sous-unités. Elles seront décrites chez la levure dans le reste de cette étude.

**Tableau 3 : Nom et fonction des sous-unités protéasomes.** Abréviations : ND, non disponible ; UBL, semblable à l'ubiquitine (de l'anglais « Ubiquitin-Like »). D'après Bard, J. A., *et al.* (2018)<sup>193</sup>.

Sous-complexe	<i>Saccharomyces cerevisiae</i>	<i>Homo sapiens</i>	Fonction
Base	Rpn1	PSMD2/S2	Liaison à Ubp6, ubiquitine et UBL
	Rpn2	PSMD1/S1	Liaison à l'ubiquitine Structurale
	Rpn13	ADRM1	Liaison à l'ubiquitine et UBL
	Rpt1	PSMC2/S7	ATPase
	Rpt2	PSMC1/S1	ATPase
	Rpt3	PSMC4/S6	ATPase
	Rpt4	PSMC6/S10	ATPase
	Rpt5	PSMC3/S6a	ATPase
Couvercle	Rpn3	PSMD3/S3	Structurale
	Rpn5	PSMD12	Structurale
	Rpn6	PSMD11/S9	Structurale

	Rpn7	PSMD6/S10	Structurale
	Rpn8	PSMD7/S12	Structurale
	Rpn9	PSMD13/S11	Structurale
	Rpn11	PSMD14/Poh1/Pad1	Désubiquitinase
	Rpn12	PSMD8/S14	Structurale
	Sem1	PSMD9/Dss1/Rpn15	Structurale
Cofacteurs supplémentaires	Rpn10	PSMD4/S5a	Liaison à l'ubiquitine et UBL
Désubiquitinases associées	Ubp6	Usp14	Désubiquitinase
	ND	Uch37	Désubiquitinase

La base comprend trois sous-unités non-ATPase appelées Rpn1, Rpn2 et Rpn13 (Rpn de l'anglais « *Regulatory Particle Non-ATPase* »). Rpn1 et Rpn2 contiennent de grands solénoïdes  $\alpha$  (paires empilées d'hélices  $\alpha$ ) qui fournissent plusieurs sites de liaison pour l'ubiquitine et les protéines de type ubiquitine UBL (de l'anglais « *Ubiquitin-Like* ») dans le cas de Rpn1 et un site de liaison pour le récepteur Rpn13 d'ubiquitine dans le cas de Rpn2<sup>194,195</sup>. Une sous-unité supplémentaire de récepteur d'ubiquitine, Rpn10, n'est pas considérée comme faisant partie de la base ou du couvercle en soi, mais comme faisant une liaison (relativement faible) entre les deux sous-complexes dans la particule régulatrice assemblée<sup>164</sup>. Au centre de la base se trouvent six sous-unités distinctes AAA+, Rpt1-Rpt6 (de l'anglais « *Regulatory Particle AAA ATPases* »). Les six sous-unités AAA+ du 19S s'assemblent en deux anneaux empilés : l'anneau hexamérique formé par les domaines AAA+ (anneau AAA+) est symétrique et positionné à proximité de la particule catalytique, tandis que les domaines N-terminaux OB (liaison oligonucléotide/oligosaccharide) forment un deuxième anneau composé d'un trimère de dimères (anneau N) avec trois « coiled-coils » (ou superhélice, composée chacun de 2 hélices  $\alpha$ ) (figure 11). Les Rpts forment donc un complexe hexamérique qui déplie les substrats, ouvre la porte de la particule régulatrice, et guide finalement les substrats dépliés vers la particule catalytique<sup>196</sup>. L'ordre topologique des sous-unités AAA+ est Rpt1 / Rpt2 / Rpt6 / Rpt3 / Rpt4 / Rpt5, tel que représenté dans la figure 11.



**Figure 11 : Structure des sous-unités Rpts du protéasome 26S eucaryote.** Vues latérale (gauche) et supérieure (droite) de l'hexamère Rpt1-Rpt6 du protéasome 26S de *Saccharomyces cerevisiae* (PDB 6FVI). Les sous-unités sont représentées de différentes couleurs.

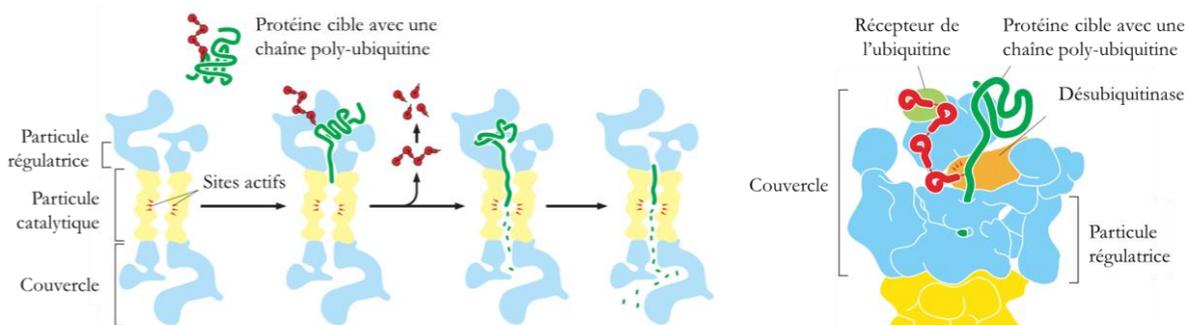
Les Rpts se ressemblent essentiellement au niveau de leur domaine central ATPasique bien qu'elles présentent une divergence au niveau de leur région N-terminale<sup>197-199</sup> (voir partie 4). Ces sous-unités ATPasiques semblent exercer des fonctions non redondantes dans la cellule puisque des mutations au niveau du domaine ATPasique chez les différentes sous-unités Rpts entraînent des phénotypes différents<sup>200</sup>.

Le sous-ensemble de couvercle agit comme un échafaudage qui soutient un côté de la base<sup>196</sup> et comprend sept sous-unités ayant une fonction principalement structurale d'échafaudage (Rpn3, Rpn5, Rpn6, Rpn7, Rpn8, Rpn9, Rpn12) ainsi que la sous-unité Rpn11. Rpn11 est une désubiquitinase DUB (de l'anglais « *Deubiquitinase* ») dépendante du  $Zn^{2+}$  de la famille JAMM et est responsable de l'élimination des chaînes d'ubiquitine liées au substrat avant leur entrée dans l'AAA+ ATPase<sup>201</sup>.

Les enzymes désubiquitinases DUBs sont donc des régulateurs importants des niveaux d'ubiquitine<sup>202</sup>. Les DUBs éliminent les molécules liées au carboxyle C-terminal du dernier résidu d'Ub (Gly76) et englobent un large groupe d'enzymes qui hydrolysent diverses liaisons (ester, peptide et isopeptide). Les DUBs sont un large groupe de protéases divisé en cinq sous-familles comprenant entre autres les métalloprotéases JAB1/MPN/MOV34 (JAMM/MPN+ ou JAMM), les Ub hydrolases C-terminales (UCHL, de l'anglais « *Ub C-terminal hydrolases* ») et les protéases spécifiques de l'Ub (UBPs, de l'anglais « *Ub-specific Proteases* »)<sup>203</sup>. Des protéases de type Ub (ULP, de l'anglais « *Ub-Like Proteases* ») qui éliminent les protéines de type Ub (UBLs, de l'anglais « *Ub-Like* ») des protéines cibles ont également été décrites<sup>204</sup>. La majorité de ces enzymes utilisent la fonction thiol de la cystéine comme nucléophile dans leur site actif pour l'hydrolyse et sont donc des protéases de type cystéine<sup>205</sup>. Par contre, les DUBs moins courants de la sous-famille JAMM utilisent un ion métallique catalytique pour leur activité<sup>205</sup>.

Le protéasome contient également un ou deux DUBs supplémentaires à Rpn11 associées de façon stable, Ubp6 et Uch37. Ubp6 est une UBP qui interagit avec Rpn1 et sépare les chaînes ubiquitine surnuméraires des substrats<sup>206</sup>. L'UCHL Uch37 (aussi appelée UCHL5) s'associe au récepteur d'ubiquitine Rpn13 et fonctionne probablement en clivant ou en modifiant les chaînes distales d'ubiquitine sur des substrats protéasomes<sup>207,208</sup>.

En résumé, au cours du processus de dégradation des protéines par le protéasome, les substrats protéiques poly-ubiquitinés sont reconnus par les récepteurs de l'ubiquitine du protéasome. Avant que les substrats soient complètement dépliés par les sous-unités Rpts et dégradés par la particule catalytique 20S, les enzymes de dé-ubiquitination du protéasome (Rpn11) enlèvent la chaîne de poly-ubiquitine des substrats (figure 12).



**Figure 12 : Schéma de la dégradation processive des protéines par le protéasome. A.** Le chapeau protéasomique reconnaît les protéines marquées par une chaîne de polyubiquitine et les transfère ensuite dans le cylindre central, où elles sont digérées. A un stade précoce, l'ubiquitine est séparée du substrat et est recyclée. Le transfert dans le cylindre central du protéasome est médié par un anneau d'ATPases qui déplie la protéine substrat au fur et à mesure qu'elle est transférée à travers l'anneau et dans le cylindre central. **B.** Structure générale du chapeau protéasomique. Le chapeau comprend un récepteur de l'ubiquitine qui maintient en place une protéine ubiquitylée pendant qu'elle commence à

être transférée dans le cylindre central, et une ubiquitine hydrolase sépare l'ubiquitine de la protéine à dégrader. Adapté de Alberts, B., et al. (2008)<sup>11</sup>.

### 3.3.3. L'ubiquitination

L'ubiquitination (ou ubiquitinylation) est une modification post-traductionnelle qui consiste en l'attachement covalent d'une molécule d'ubiquitine (chaîne de 76 résidus) sur une protéine cible.

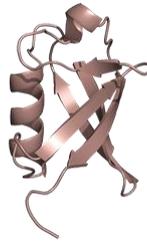


Figure 13 : Structure de l'ubiquitine. PDB 1UBQ.

Elle est appelée ainsi parce qu'elle est localisée dans tous les compartiments subcellulaires de toutes les cellules eucaryotes, où plus de 80% des protéines cellulaires sont sujettes à ce processus<sup>209</sup>. L'attachement de l'ubiquitine se fait *via* une liaison isopeptidique qui se met en place entre le résidu glycine se trouvant à l'extrémité C-terminale (G76) de la molécule d'ubiquitine et une Lys de la protéine cible. Tout substrat peut être étiqueté avec une seule molécule d'ubiquitine, c'est la mono-ubiquitination, ou par une chaîne d'ubiquitines plus ou moins longue résultant en la poly-ubiquitination du substrat. La molécule d'ubiquitine contient sept résidus Lys, et chacun peut accueillir une molécule d'ubiquitine lors de la formation des chaînes de poly-ubiquitine. Il faut quatre ubiquitines pour qu'une protéine puisse être dégradée<sup>210</sup>.

Le marquage des protéines par l'ubiquitine se fait en plusieurs étapes et implique l'intervention séquentielle de trois enzymes : une enzyme activatrice (E1), une enzyme de conjugaison (E2) et une ligase (E3)<sup>211</sup>. La figure 14 résume le processus d'ubiquitination des protéines à dégrader.

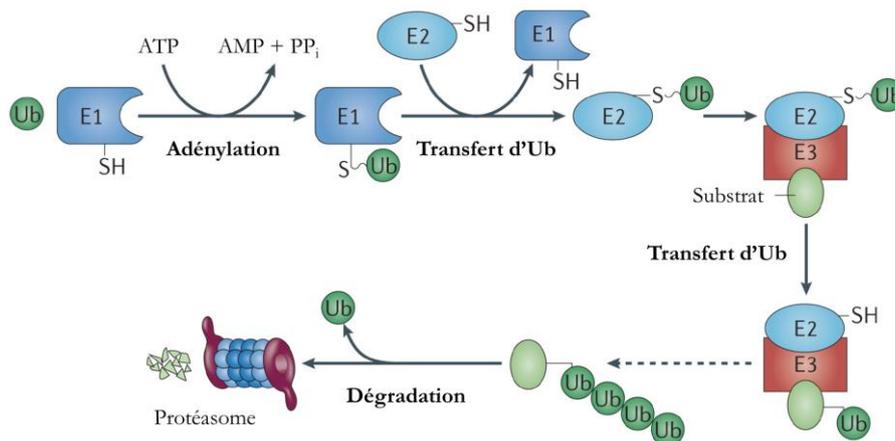


Figure 14 : Mécanisme d'ubiquitination. L'ubiquitination est un signal commun pour les protéasomes eucaryotes 26S et implique une cascade d'enzymes E1 (enzyme d'activation), E2 (enzyme de conjugaison) et E3 (ligase). Dans cette cascade, E1 (plus l'ATP) adénylate d'abord le carboxylate carboxy-terminal de l'ubiquitine (Ub), formant l'Ub-AMP, puis forme un intermédiaire thioester Ub (E1-Ub). L'ubiquitine est transférée de E1 à E2, puis à la protéine cible avec l'aide de E3. Typiquement, une liaison isopeptidique est formée entre le carboxylate C-terminal de l'ubiquitine et le groupe  $\epsilon$ -amino d'une chaîne latérale Lys de la protéine substrat ou de la chaîne ubiquitine croissante (les chaînes ubiquitine liées à Lys48 sont des signaux courants des protéasomes 26S). Les enzymes désubiquitylantes

des protéasomes 26S libèrent et recyclent l'ubiquitine pendant la dégradation des protéines du substrat. PP<sub>i</sub>, pyrophosphate inorganique. Adapté de Maupin-Furlow, J. (2012)<sup>156</sup>.

La molécule d'ubiquitine (Ub) doit être activée de manière à pouvoir former une liaison isopeptidique. Cette activation est réalisée selon une réaction séquentielle commençant par la liaison d'un ATP sur l'Ub conduisant à la formation d'un intermédiaire ubiquitine-adénylate qui sert de donneur d'Ub à l'enzyme E1. Parallèlement, une molécule d'ATP se fixe sur l'enzyme E1, provoquant un changement conformationnel permettant une meilleure affinité de l'enzyme E1 pour l'Ub. La Gly C-terminale de l'Ub est activée de manière ATP-dépendante et subit ensuite une attaque nucléophile par l'atome de soufre sous forme de sulfure d'une Cys se situant sur le site actif de l'enzyme E1. Cela a pour conséquence la formation d'une liaison thiol-ester entre la Gly C-terminale de l'Ub et la Cys de l'enzyme E1. Chaque enzyme E1 active est capable de porter deux molécules d'Ub, l'une sous forme de thiol ester et l'autre sous forme d'adénylate<sup>212</sup>. Cette molécule d'Ub est alors transférée sur la Cys du site actif de l'enzyme E2 par une liaison thioester covalente<sup>213</sup>. Lorsqu'elles sont liées à certaines E3 ligases, les E2 peuvent catalyser directement le transfert de la molécule d'Ub sur le substrat ou sur les résidus lysine de l'Ub elle-même, permettant ainsi la poly-ubiquitination. La réaction impliquant les E3 ligases se déroule en deux étapes : l'interaction de l'E3 ligase avec le substrat suite au signal d'ubiquitination et la liaison covalente d'une ou plusieurs molécules d'Ub au substrat. L'E3 ligase est impliquée aussi bien dans la catalyse de la première molécule d'Ub au substrat que dans la liaison des autres molécules d'Ub successives permettant la synthèse d'une chaîne de poly-ubiquitination, lorsque cela est requis<sup>214</sup>.

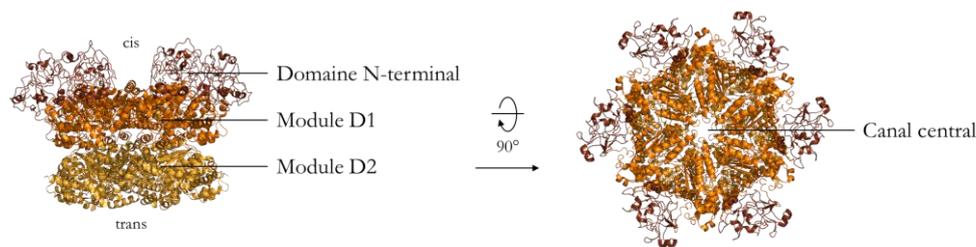
Bien que la séquence d'acides aminés hautement conservée d'Ub soit limitée aux eucaryotes, une superfamille de protéines de type ubiquitine (UBL, de l'anglais « *Ubiquitin-Like* ») est connue pour être présente dans tous les domaines de la vie avec des fonctions diverses. Parmi elles, la protéine FAT10 est aussi impliquée dans la protéolyse chez les eucaryotes selon un mécanisme similaire<sup>215</sup>. En général, les protéines UBL ne partagent pas beaucoup d'identité de séquence avec l'ubiquitine. Cependant, les protéines UBL ont un repliement en tonneau  $\beta$  et ont souvent un motif diglycine C-terminal qui est nécessaire à l'activité<sup>216</sup>.

Les Ub et UBL ne constituent pas le seul moyen d'adressage des protéines à la protéolyse par le protéasome 26S. Il existe des protéines dont la dégradation se passe indépendamment de l'ubiquitination comme l'ornithine décarboxylase impliquée dans la synthèse des polyamines<sup>217</sup> et la protéine p21 inhibitrice des kinases dépendantes des cyclines<sup>218</sup>.

#### 3.3.4. Un protéasome alternatif ATP-dépendant avec la protéine AAA+ Cdc48

La protéine AAA+ Cdc48, aussi appelée p97, VCP ou TER94, est l'une des protéines intracellulaires les plus abondantes et est très conservée chez toutes les espèces eucaryotes.

Cdc48 est composé d'un domaine régulateur N-terminal suivi de deux domaines AAA+ (figure 15). Les deux domaines AAA+ forment deux anneaux hexamériques empilés partageant un canal central connecté<sup>219</sup>, et l'activité ATPase des deux domaines AAA+ est essentielle<sup>220</sup>.



**Figure 15 : Structure de la protéine Cdc48/p97.** Vues latérale (gauche) et supérieure (droite) de la protéine Cdc48/p97 humaine (PDB 5FTK). Cdc48 est un homo-hexamère dont chaque monomère comprend un domaine N-terminal (N) (marron) et deux domaines AAA+ ATPase : D1 (orange foncé) et D2 (orange clair). Le côté N-terminal (D1) du canal central est appelé côté cis, et le côté C-terminal (D2) est appelé côté trans.

Cdc48 est situé dans le cytosol et le noyau, et joue un rôle essentiel dans de nombreux processus cellulaires, y compris entre autres dans la voie de dégradation ERAD du réticulum endoplasmique (de l'anglais « *Endoplasmic Reticulum-Associated Degradation* »), dans les systèmes de contrôle de la qualité des protéines associés aux mitochondries, ou dans la prévention de l'agrégation des protéines<sup>221</sup>. La fonction principale de Cdc48 est de séparer les protéines substrat des complexes protéiques et des membranes des organites. Ces protéines libérées sont ensuite transférées à la machine de dégradation ou repliées pour être recyclées. Dans le système protéasomique ubiquitine et 26S, Cdc48 fonctionne alors en amont de la dégradation protéasomique<sup>220</sup>. Plusieurs études récentes ont suggéré que le Cdc48 peut se lier directement à la particule catalytique 20S, formant un complexe protéasomique Cdc48-20S appelé aussi « protéasome alternatif »<sup>220,222,223</sup>. Cependant, le protéasome Cdc48-20S ne présente aucune activité de dégradation *in vitro* pour une protéine substrat modèle, Cdc48 seul présentant une activité intrinsèquement faible<sup>224</sup>.

### 3.4. Les protéasomes des bactéries, mitochondries et chloroplastes

Selon la taille du génome, les bactéries possèdent généralement de deux à cinq protéases AAA+, non-protéasomiques (sauf chez les actinobactéries qui possèdent un protéasome). De même, Les organites eucaryotes d'origine bactérienne (mitochondries et chloroplastes) possèdent souvent plusieurs protéases AAA+<sup>175</sup>. Il est important de présenter ici les protéases non-protéasomiques car leur structure et leur mode d'action présumés seront discutés par la suite par rapport au protéasome PAN-20S archéen étudié lors de cette thèse.

#### 3.4.1. Diversité des protéasomes bactériens

Les protéases AAA+ présentes dans les bactéries, les mitochondries et les chloroplastes comprennent la famille des protéases Clp, HslUV (aussi connu sous le nom de ClpYQ), Lon et FtsH<sup>182</sup>, mais aussi le protéasome Mpa-20S et les protéines de type Cdc48/p97<sup>225</sup>.

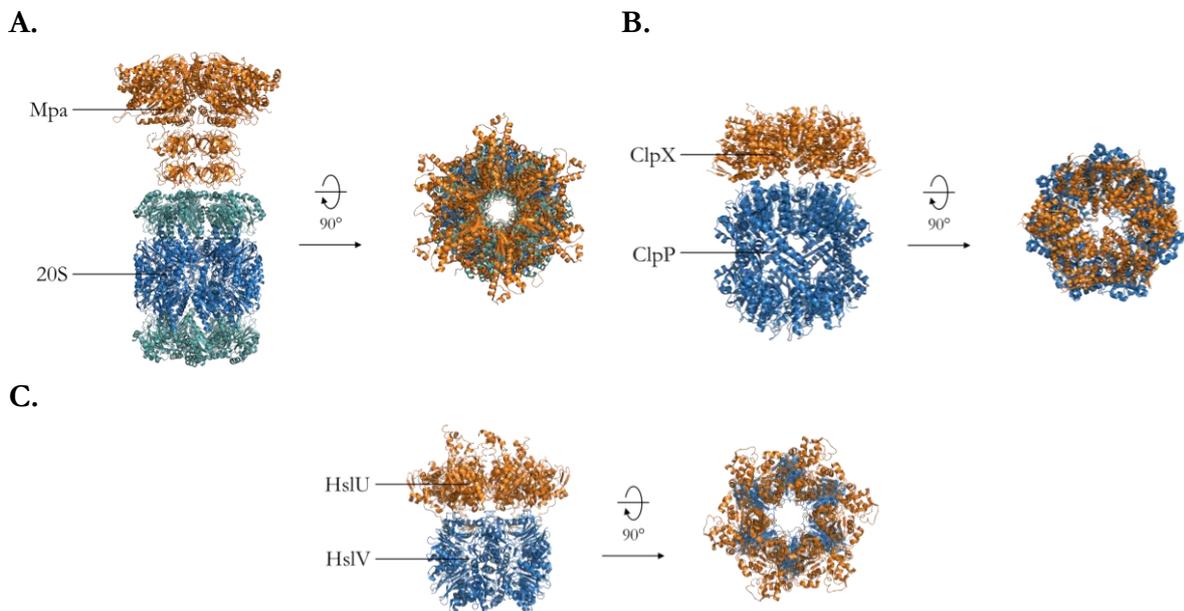
La famille des protéases Clp code des ATPases hexamériques distinctes (ClpA ou ClpX chez les bactéries à Gram-négatif et ClpC ou ClpE chez les bactéries à Gram-positif). Ces particules régulatrices s'associent à ClpP, une protéase séquestrée de 14 sous-unités pour former ClpXP, ClpAP, ClpCP ou ClpEP<sup>182,226</sup>. La protéase HslV se compose de deux anneaux hexamériques et s'associe avec l'hexamère HslU. Dans le cas de Lon et FtsH, l'activité ATPasique est liée à un domaine intégré dans le complexe lui-même, les deux domaines étant codés sur une seule chaîne polypeptidique<sup>227,228</sup>.

La plupart des phyla bactériens utilisent ClpXP, ClpAP ou ClpCP, ClpEP, HslUV, Lon et FtsH pour la dégradation dépendante de l'ATP des protéines, tandis que les actinobactéries utilisent également le protéasome 20S<sup>225,229</sup>. La particule 20S (similaire à la particule 20S eucaryote) et HslV utilisent une thréonine nucléophile pour le clivage des liaisons peptidiques. ClpP utilise une triade catalytique Ser-His-Asp, Lon utilise une paire de Ser-Lys tandis que FtsH utilise un site actif Asp-Zn<sup>2+</sup><sup>182</sup>.

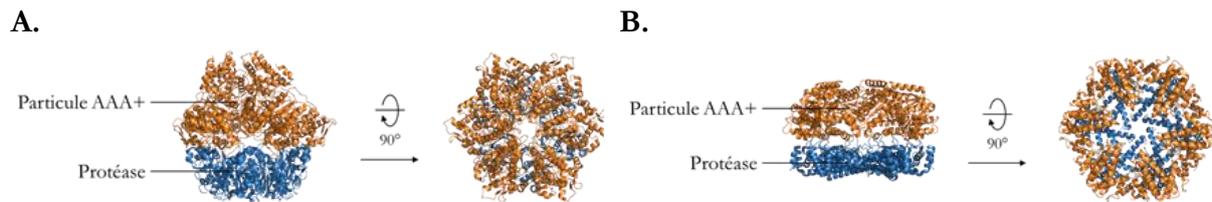
Les particules régulatrices AAA+ de la particule catalytique 20S sont Mpa (de l'anglais « *Mycobacterial Proteasomal ATPase* ») et son homologue ARC (de l'anglais « *AAA+ ATPase forming Ring shaped Complexes* »). Mpa et ARC constituent des anneaux hexamériques homologues du complexe formé par les protéines AAA+ (Rpt1-Rpt-6) de la base de la particule régulatrice 19S eucaryote. La protéine ARC étant peu caractérisée, nous ne parlerons dans la suite que de la protéine Mpa.

La particule catalytique 20S est similaire à celle des eucaryotes, sous forme d'un complexe empilé de deux anneaux  $\beta$  intérieurs flanqués d'un anneau  $\alpha$  de chaque côté. Les actinobactéries sont donc les seules bactéries possédant un protéasome, sous forme de protéasome Mpa-20S<sup>230</sup>. Les actinobactéries possèdent également un homologue de Cdc48/p97 appelé Cpa (de l'anglais « *Cdc48-like Protein of Actinobacteria* »)<sup>231,232</sup>. Tous les membres des actinobactéries qui codent pour Cpa hébergent également les gènes des sous-unités protéasomales, bien que certaines actinobactéries possèdent le protéasome mais pas Cpa.

Les figures 16 et 17 représentent les structures des différentes protéases AAA+ bactériennes.



**Figure 16 : Structure des protéases AAA+ bactérienne ayant une particule régulatrice distincte.** Les protéines régulatrices AAA+ sont représentées en orange et les sous-unités de type  $\alpha$  et  $\beta$  des protéases sont représentées respectivement en vert et bleu. **A.** Protéasome Mpa-20S de *Mycobacterium tuberculosis*. Les structures de Mpa et de la particule catalytique 20S ont été obtenues de façon indépendante (respectivement PDB 5KZF et PDB 3MI0). **B.** Protéase AAA+ ClpXP d'*Escherichia coli* (PDB 1TYF). Les structures de ClpX et de ClpP ont été obtenues de façon indépendante (respectivement PDB 3HWS et PDB 1YG6). **C.** Protéase AAA+ HslUV de *Haemophilus influenzae* (PDB 1G3I).



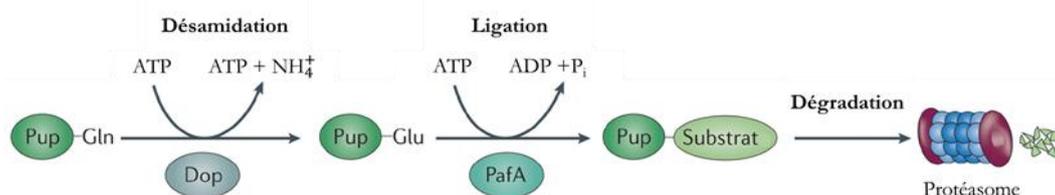
**Figure 17 : Structure des protéases AAA+ bactérienne fusionnées avec leur particule régulatrice.** Les protéines régulatrices AAA+ sont représentées en orange et les protéases sont représentées en bleu. **A.** Protéase Lon de *Yersinia pestis* (PDB 6ON2). **B.** Protéase FtsH de *Thermotoga maritima* (PDB 3KDS).

L'architecture des protéases AAA+ bactériennes est simple. Par exemple, les protéases HslV et ClpP contiennent respectivement douze et quatorze sous-unités identiques, homologues aux sous-unités  $\beta$  du protéasome 20S, disposées en deux anneaux empilés, tandis que le protéasome 20S est composé de sous-unités  $\alpha$  et  $\beta$  identiques dans un arrangement similaire au protéasome 20S eucaryote. Les particules régulatrices sont généralement des homo-hexamères. Structurellement, les sous-unités AAA+ ressemblent aux homologues eucaryotes, les Rpts. Ils contiennent tous un domaine en hélice  $\alpha$  proche de l'extrémité N-terminale, suivi du domaine OB (anneau N) et d'un anneau AAA+<sup>233</sup>.

Les rôles des protéases AAA+ spécifiques peuvent changer d'un organisme à l'autre. L'utilisation de plusieurs protéases AAA+ par des cellules bactériennes ou des organites eucaryotes peut, en partie, être liée aux divers emplacements subcellulaires des substrats. Par exemple, le FtsH bactérien et ses homologues mitochondriaux sont liés à la membrane, et certains de leurs substrats sont des protéines membranaires périphériques ou intégrales<sup>234,235</sup>, alors que la plupart des protéases Lon dans les bactéries sont cytoplasmiques, et celles dans les mitochondries fonctionnent dans la matrice<sup>236</sup>.

### 3.4.2. Les différentes voies d'adressage aux protéasomes bactériens

Les protéases bactériennes n'utilisent pas de récepteurs d'ubiquitine, car la signalisation de l'ubiquitine n'existe pas chez les bactéries<sup>237</sup>. Cependant, la protéine Pup de type ubiquitine (UBL) impliquée avec le protéasome Mpa-20S chez les actinobactéries cible les protéines par monopupylation pour leur dégradation. Les régions coiled-coils en N-terminal de l'anneau N de Mpa reconnaissent alors l'étiquette de dégradation Pup<sup>238</sup>. Cette voie de modification appelée pupylation n'a aucun lien génétique avec l'ubiquitination eucaryote et représente un exemple de convergence évolutive<sup>239</sup>. La figure 18 résume le processus de pupylation des protéines à dégrader.



**Figure 18 : Mécanisme de pupylation.** Pupylation et protéolyse à médiation protéasomique chez les actinobactéries. Lors de la pupylation, la Gln carboxy-terminale de la protéine de type ubiquitine (Pup) est désamidifiée en Glu par Dop. Le PafA peut alors fixer le Pup aux substrats, ce qui favorise leur dégradation protéasomique. Une fois conjugué à des substrats protéiques, Pup se lie aux domaines coiled-coils de l'ATPase protéasomique (Mpa chez les mycobactéries), et une région du Pup est convertie d'un état désordonné à une hélice  $\alpha$  (non représentée). P<sub>i</sub>, phosphate inorganique. Adapté de Maupin-Furlow, J. (2012)<sup>156</sup>.

Il existe également une variété de dégrons codés dans la séquence primaire qui transforment une protéine en substrat, comme l'étiquette *ssrA* (voir ci-après).

Chez *Escherichia coli*, Lon serait responsable de la dégradation d'environ 50 % des protéines mal repliées<sup>227</sup>, ce qui suggère que la protéase reconnaît les motifs généraux des protéines mal repliées avec une faible spécificité de séquence. En revanche, la protéase bactérienne ClpXP est beaucoup plus sélective et nécessite une séquence de dégron spécifique, telle que l'étiquette *ssrA*, pour reconnaître un substrat<sup>240</sup>. Ces deux enzymes illustrent différents mécanismes utilisés pour assurer la dégradation des protéines de mauvaise qualité.

L'un des exemples les mieux étudiés dans le domaine de la protéolyse régulée des bactéries est la reconnaissance de la séquence signal *ssrA* par ClpXP. Chez *E. coli* une interruption de la traduction peut conduire à l'addition d'une étiquette peptidique de 11 résidus (AANDENYALAA) du côté C-terminal de la chaîne naissante<sup>241</sup>. Cette modification est médiée par un ARN *ssrA* (appelé également 10Sa ARN et ARN<sub>tm</sub>) et marque la protéine pour la protéolyse. L'ARN *ssrA* fonctionne à la fois comme un ARN<sub>m</sub> et un ARN<sub>t</sub><sup>242</sup>. Lorsque les ribosomes rencontrent un problème au niveau de l'extrémité 3' de l'ARN<sub>m</sub> endommagé, un ARN *ssrA* chargé d'alanine vient se lier au ribosome comme un ARN<sub>t</sub> et ajoute l'alanine à la chaîne naissante ralentie. La traduction continue après l'ARN<sub>m</sub> par un petit cadre ouvert de lecture (ORF, de l'anglais « *Open Reading Frame* ») qui code pour le peptide de dégradation *ssrA* à l'extrémité C-terminale. Ce système fournit à la fois une méthode pour éviter l'accumulation des ribosomes à la fin des ARN<sub>m</sub> défectueux et un mécanisme général de contrôle qualité qui permet à la cellule de se débarrasser des fragments de protéines incomplètes qui aboutissent à des activités cellulaires inappropriées<sup>243</sup>.

La protéase Lon est quant à elle capable de reconnaître largement les protéines mal repliées. Comparée aux autres protéases dépendantes de l'énergie, la protéase Lon a la plus faible capacité de dépliement<sup>244</sup>, mais un répertoire de substrat étonnamment diversifié. En effet, Lon semble reconnaître les résidus hydrophobes sur des substrats mal repliés qui sont généralement enfouis dans la structure native comme étant son principal déterminant de reconnaissance plutôt qu'un motif de séquence unique<sup>245</sup>. Pour évaluer correctement la qualité des protéines, Lon doit être capable de faire la distinction entre les protéines mal repliées et celles qui sont dans des états de repliement intermédiaires. Il semble que la durée de vie plus courte des régions hydrophobes exposées d'une protéine en voie vers l'état natif fixe la limite inférieure des échelles de temps où Lon reconnaît une protéine de mauvaise qualité. En conséquence, le complexe Lon-substrat a une cinétique suffisamment transitoire pour être compatible avec cette discrimination, sinon Lon détruirait par erreur les protéines en cours de repliement. Ainsi, en combinant une faible capacité de dépliement avec un large spectre de reconnaissance, la protéase Lon gagne en sélectivité dans la reconnaissance de protéines vraiment mal repliées.

### 3.5. Les protéasomes des archées

#### 3.5.1. Les différents protéasomes archéens

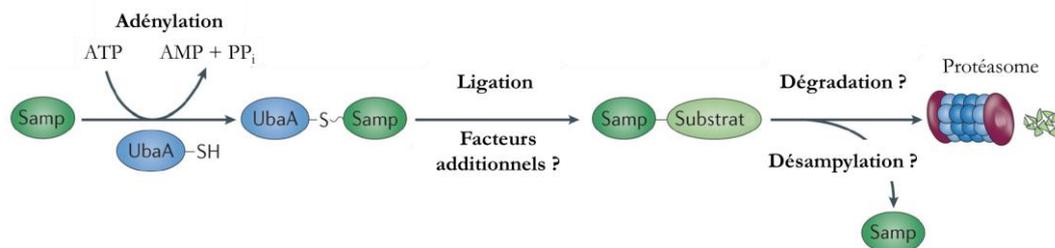
Les archées contiennent un protéasome fonctionnel qui est similaire au protéasome des eucaryotes. Toutes les espèces d'archées dont le génome a été séquencé codent pour une particule catalytique 20S homologue à la version eucaryote composée de sous-unités  $\alpha$  et  $\beta$ , ainsi qu'une ou plusieurs particule(s) régulatrice(s) AAA+<sup>246,247</sup>. Plusieurs protéasomes d'archée ont été caractérisés et la

majorité d'entre eux possèdent une activité uniquement de type chymotrypsine, notamment celui de *Haloferax volcanii* et de *T. acidophilum*, cependant les protéasomes 20S des archées *Methanosarcina thermophila* et *Methanocaldococcus jannaschii* ont montré des activités peptidases inhabituelles de type caspase et trypsine<sup>246</sup>. Globalement, les protéases AAA+ chez les archées ressemblent plus aux eucaryotes qu'aux bactéries ou qu'aux organelles des eucaryotes. Cependant, tout comme les bactéries, les protéasomes archéens ont une architecture plus simple que le protéasome 26S de eucaryotes, avec particule catalytique 20S composée dans la plupart des cas de sous-unités  $\alpha$  et  $\beta$  formant des anneaux homo-heptamériques et des particules régulatrices homo-hexamériques.

De multiples et distinctes particules AAA+ ont été retrouvées au sein d'organismes archéens, indiquant qu'un réseau de particules régulatrices permet d'élargir le spectre de substrats pouvant être prises en charge par les protéasomes<sup>248</sup>. On retrouve parmi elles les ubiquitaires Cdc48/p97, ainsi que VAT (de l'anglais « *Valosin-containing protein-like* »), complexe homologue aux protéines de type Cd48, AMA et PAN<sup>248-251</sup>. AMA est homologue à FtsH au niveau du domaine de ATPase et possède un domaine N-terminal homologue au domaine N des protéines de type Cd48<sup>252</sup>. Cette particule n'est présente que chez *Archaeoglobus fulgidus* et les archées méthanogènes. PAN est un complexe hexamérique homologue à celui formé par les protéines AAA+ (Rpt1-Rpt-6) de la base de la particule régulatrice 19S eucaryote. La plupart des archées codent une seule protéine PAN<sup>253</sup>. Cependant, une étude sur le complexe PAN de l'archée *Halobacterium salinarium* a mis en évidence deux isoformes de PAN (PAN A et PAN B), distincts en structure, régulation et fonction biologique, et ayant des affinités différentes pour l'association avec le protéasome 20S<sup>254</sup>. La partie 5 décrit la particule PAN et le protéasome PAN-20S de façon approfondie. Les archées contiennent également la protéase Lon de type B liée à la membrane<sup>255</sup>. Cependant, contrairement aux bactéries, les archées ne possèdent pas les homologues de la protéase FtsH et dans la plupart des cas, les protéases HslUV, Clp et Lon de type A sont également absentes chez ces organismes<sup>255</sup>.

### 3.5.2. Les différentes voies d'adressage aux protéasome archéens

Chez les archées, il existe une forme de modification post-traductionnelle nommée sampylation qui ressemble à l'ubiquitination chez les cellules eucaryotes. Les protéines nommées SAMP (de l'anglais « *Small Archaeal Modifier Protein* ») sont des UBL (de l'anglais « *Ubiquitin Like* ») (voir partie 3.3.3) fixées de façon covalente aux protéines cibles<sup>256</sup>. La figure 19 résume le processus de sampylation des protéines à dégrader.



**Figure 19 : Mécanisme de sampylation.** Comme pour l'ubiquitination, les petites protéines modificateurs (Samps) des archées sont adénylées à leur carboxylate carboxy-terminal par une enzyme activatrice E1 (UbaA) et transférées aux chaînes latérales Lys des substrats de protéines. Il reste à déterminer si d'autres facteurs (autres que l'E1) sont nécessaires pour assurer une sélection adéquate des protéines cibles et si les protéines sampylées sont dégradées par les protéasomes. Bien que des chaînes Samp2 liées à Lys58 aient été détectées, il n'est pas clair si ces chaînes sont ancrées aux protéines du substrat (non montré). PP<sub>i</sub>, pyrophosphate inorganique. Adapté de Maupin-Furlow, J. (2012)<sup>156</sup>.

La sampylation a été identifiée pour la première fois chez *Haloferax volcanii*. Deux parmi trois protéines UBL (SAMP1/2) de *H. volcanii* ont été trouvées attachées de façon covalente à des protéines cibles à travers des liaisons non-thiol<sup>256</sup>. La formation de ces conjugués protéiques implique une enzyme homologue à l'enzyme E1 présente chez les eucaryotes, nommée UbaA (de l'anglais « *Ubiquitin-like Activating protein of Archaea* »). L'enzyme UbaA est requise pour l'activation et la fixation du côté C-terminal de l'Ubl aux protéines cibles par analogie avec les systèmes eucaryotes<sup>257</sup>. Les liaisons isopeptidiques sont formées entre le groupe carboxyle de l'extrémité C-terminale des protéines SAMP des archées et le groupe ε-amino des résidus Lys sur les protéines cibles<sup>258</sup>. Les archées codent aussi pour des homologues des protéines de-ubiquitination (DUBs) ainsi que d'autres isopeptidases qui éliminent les molécules liées au dernier résidu des protéines Ub/UBL<sup>202</sup>. Les protéines DUBs retrouvées chez les archées appartiennent toutes à la sous-famille JAMM/MPN+ dont fait partie la sous-unité Rpn11 du protéasome 26S. La protéine HvJAMM1 de *H. volcanii* a été caractérisée récemment et a montré une activité de clivage des protéines SAMP-modifiées en catalysant l'élimination des SAMP1/2 ajoutés à des protéines<sup>259</sup>.

Cependant, un lien direct entre la sampylation et la dégradation des protéines par le protéasome n'a pas été clairement établi, et il n'y a toujours aucune preuve directe d'une association physique entre les protéines SAMP et le protéasome archéen<sup>260</sup>. Un homologue des protéines SAMP qui présente une similarité de structure et de séquence en acides aminés avec la protéine UBL eucaryote Urm1 (de l'anglais « *ubiquitin-related modifier-1* ») a été récemment identifié dans *Sulfolobus acidocaldarius*. Cette protéine interagit directement avec la particule catalytique 20S et est dégradée par ce complexe protéase. La fixation covalente de Urm1 à un substrat tel que la GFP fusionnée à Urm1 (GFP-Urm1) entraîne la reconnaissance de la protéine modifiée par la protéine régulatrice PAN de manière ATP-dépendante et sa dégradation par le particule catalytique 20S. Contrairement au système de dégradation eucaryote, où l'ubiquitine est clivée de la cible et recyclée, la protéine archéenne Urm1 est détruite *in vitro* par le protéasome 20S<sup>260</sup>. Une hypothèse est que les protéasomes dégradent également d'autres types de tags appelés « dégrons ». Conformément à cette hypothèse, les protéasomes d'archées dégradent les protéines en absence de la sampylation<sup>258</sup>. Il est également intéressant de noter que des études ont montré que le dégron bactérien *ssrA* permet d'adresser les substrats protéiques pour leur reconnaissance et dépliement par PAN *in vitro*<sup>250,261,262</sup>.

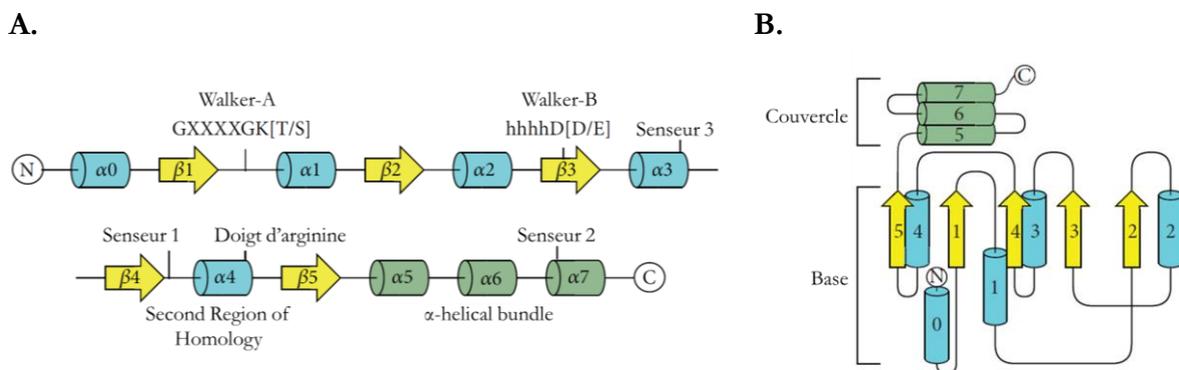
Les homologues SAMP et UbaA, sont communs chez les archées, suggérant que la sampylation est universelle dans ce domaine de la vie. Néanmoins, des homologues des enzymes E2 et E3 n'existent pas dans la plupart des archées (y compris *H. volcanii*)<sup>263</sup>. La sampylation est donc généralement considérée comme une forme simplifiée de l'ubiquitination chez les archées<sup>258</sup>. Cependant, la découverte par bioinformatique d'une cascade enzymatique complète comprenant les enzymes E1, E2 et E3 de modification de l'Ub dans *C. subterraneum*, et plus récemment dans les archées Asgard, suggère que le système sophistiqué de conjugaison de certaines classes de protéines par Ub n'est pas limité aux eucaryotes<sup>264,265</sup>.

## 4. Structure et mode d'action des protéines AAA+

### 4.1. Les protéines de la superfamille des AAA+

Toutes les cellules utilisent des complexes protéiques sophistiqués qui couplent énergie chimique de l'hydrolyse des nucléosides triphosphates (NTP) à la génération de force mécanique<sup>266</sup>. Ces complexes fonctionnent de la même manière qu'un moteur, les NTP étant le carburant qui est le

plus souvent utilisé pour surmonter les barrières thermodynamiques. Basé sur cette similitude, ces complexes protéiques sont appelés « machines moléculaires »<sup>267</sup>. Les protéines AAA+ forment une superfamille vaste et diversifiée qui catalysent l'hydrolyse de l'ATP et utilisent l'énergie ainsi obtenue pour effectuer des travaux mécaniques. Ces protéines se rassemblent généralement sous forme de complexes d'hexamères cycliques qui sont impliqués dans le remodelage énergie-dépendant des macromolécules<sup>268</sup>. Cela représente une grande variété de fonctions cellulaires différentes comme le dépliement et la dégradation des protéines, la biogénèse des peroxysomes, la biosynthèse de la bactériochlorophylle ainsi que la recombinaison, la réplication et la réparation de l'ADN<sup>269</sup>. Les protéines AAA+ comprennent le moteur moléculaire dynéine, les hélicases impliquées dans la réplication de l'ADN, les chélatases du métal, et les protéines régulatrices du protéasome. En raison de leurs fonctions diverses, les protéines AAA+ peuvent être trouvées dans la plupart des compartiments subcellulaires des cellules eucaryotes, ainsi que chez les archées, bactéries et virus<sup>10</sup>.

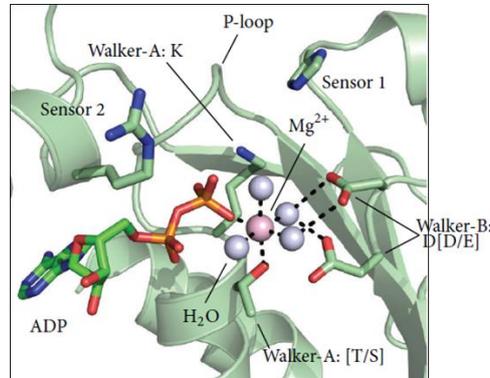


**Figure 20 : Le motif AAA+.** **A.** Topologie du motif AAA+  $\alpha$ - $\beta$ - $\alpha$  et emplacement des éléments du site actifs sont présentés en séquence primaire et en structure secondaire. Les hélices et les brins à l'intérieur du noyau  $\alpha/\beta$  sont colorés respectivement en bleu et jaune. Les hélices du domaine C-terminal sont colorées en vert. **B.** Le domaine AAA+ le plus simple est caractérisé par une topologie  $\alpha$ - $\beta$ - $\alpha$  avec un domaine C-terminal en  $\alpha$ -helical bundle. Les hélices et les brins à l'intérieur du domaine de la base sont représentés par des cylindres bleus et des flèches jaunes, respectivement, et les hélices du domaine du couvercle sont représentées par des cylindres verts. Adapté de Miller, J. M., & Enemark, E. J. (2016)<sup>270</sup>.

La superfamille AAA+ est un sous-ensemble de la plus grande superfamille des protéines P-loop<sup>247</sup>. Tous les membres de la famille P-loop sont structurellement similaires et possèdent un repliement distinct  $\alpha/\beta$ . Le domaine AAA+ contient 200-250 acides aminés avec un feuillet  $\beta$  central dans l'ordre  $\beta 5$ - $\beta 1$ - $\beta 4$ - $\beta 3$ - $\beta 2$  (figure 20). Le feuillet  $\beta$  est entouré des deux côtés d'hélices  $\alpha$  pour former un empilement à trois étages  $\alpha$ - $\beta$ - $\alpha$ . De nombreuses protéines AAA+ ont un faisceau d'hélices  $\alpha$  en C-terminal («  $\alpha$ -helical bundle ») en plus du noyau  $\alpha$ - $\beta$ - $\alpha$ . Les rôles fonctionnels de ce motif sont variés et comprennent la formation d'un couvercle sur le site de liaison des nucléotides et un rôle d'interactions de sous-unités dans les complexes protéiques oligomères. Le domaine AAA+ contient des motifs de séquence primaire utilisés à l'origine pour établir la superfamille AAA+<sup>271</sup>.

Comme toutes les NTPases à P-loop, les protéines AAA+ ont des résidus de motifs Walker-A et Walker-B qui sont essentiels pour la liaison et l'hydrolyse de l'ATP. Le motif Walker-A consiste en une séquence GXXXXXXGK[T/S], où X est un acide aminé quelconque et le résidu C-terminal est soit une Thr, soit une Ser. Structurellement, ce motif forme une boucle entre  $\beta 1$  et  $\alpha 1$  dans la topologie AAA+<sup>268</sup>. Ce motif est la P-loop et est l'une des séquences les mieux conservées des protéines AAA+. Le motif Walker-B au sein de  $\beta 3$  est caractérisé par la séquence hhhhD[D/E],

où h représente tout résidu hydrophobe et le résidu C-terminal est un Asp ou Glu. L'hydrolyse de l'ATP catalysée par les protéines AAA+ dépend du résidu Glu du Walker-B. Ce résidu conservé est la base catalytique qui active une molécule d'eau pour l'attaque nucléophile du phosphate  $\gamma$  pendant l'hydrolyse de l'ATP<sup>270</sup>. La figure 21 représente l'interaction entre une molécule d'ADP et les motifs Walker-A et Walker-B.



**Figure 21 : Site actif des protéines AAA+.** Les résidus du site actif sont positionnés avec précision pour lier les nucléotides et  $Mg^{2+}$ . La Lys et les résidus Ser/Thr du Walker-A lient respectivement les phosphates  $\beta$  et  $\gamma$  du nucléotide lié ainsi que l'ion  $Mg^{2+}$ . L'hydrolyse est due à l'extraction d'un proton d'une molécule d'eau par le résidu Glu du motif Walker B, activant ainsi la molécule d'eau pour une attaque nucléophile ultérieure sur le phosphate  $\gamma$  de l'ATP lié. Les lignes pointillées indiquent les interactions moléculaires. La molécule d'ADP liée et les caractéristiques critiques du site actif sont représentées en bâtonnet, les molécules d'eau en sphères bleu clair et l'ion magnésium en sphère magenta. Adapté de Miller, J. M., & Enemark, E. J. (2016)<sup>270</sup>.

Toutes les protéines AAA+ contiennent une région C-terminale au motif Walker-B appelée la deuxième région d'homologie (de l'anglais « *Second Region of Homology* »). Cette région s'étend sur 15 à 20 résidus pour inclure une partie de  $\beta 4$ , la totalité de l'hélice  $\alpha 4$  et la boucle reliant  $\alpha 4$  à  $\beta 5$ . La SRH contient les motifs du senseur 1 et du doigt d'arginine, tous deux requis pour l'hydrolyse de l'ATP. Ces caractéristiques coordonnent l'hydrolyse des nucléotides et propagent les changements conformationnels associés à l'hydrolyse des nucléotides entre les sous-unités des complexes protéiques AAA+<sup>272</sup>. En raison de l'importance fonctionnelle de cette région et de son absence dans les autres NTPases de type Walker, il s'agit d'une caractéristique déterminante de la famille superfamille AAA+.

Le motif du senseur 1 est situé à l'extrémité N-terminale du SRH dans la boucle reliant  $\beta 4$  à  $\alpha 4$ . Il s'agit d'un résidu polaire qui est le plus souvent de l'Asn mais qui peut aussi être une Ser, Thr ou Asp<sup>268</sup>. Il est structurellement situé entre les motifs Walker-A et Walker-B et fonctionne de façon coordonnée avec le Glu Walker-B pour orienter correctement la molécule d'eau qui exerce une attaque nucléophile sur le phosphate  $\gamma$  de la molécule ATP liée<sup>273</sup>. Le site AAA+ ATPase se trouve à l'interface de sous-unités adjacentes dans un complexe protéique. Les résidus Walker-A/B et senseur 1 du site de l'ATPase sont tous situés sur la même sous-unité, tandis que le doigt arginine est issu de la sous-unité voisine. Pour toutes les structures connues d'oligomères AAA+, ce résidu d'Arg se projette dans le site de liaison et d'hydrolyse ATP d'une sous-unité adjacente. Sur la séquence primaire, le doigt d'arginine est situé près de l'extrémité C-terminale de la SRH et est situé dans la boucle entre  $\alpha 4$  et  $\beta 5$ <sup>270</sup>.

Deux autres caractéristiques du site actif ont été identifiées à partir de l'examen des séquences et des structures de la protéine AAA+. Le senseur 2 est conservé dans toutes les protéines AAA+

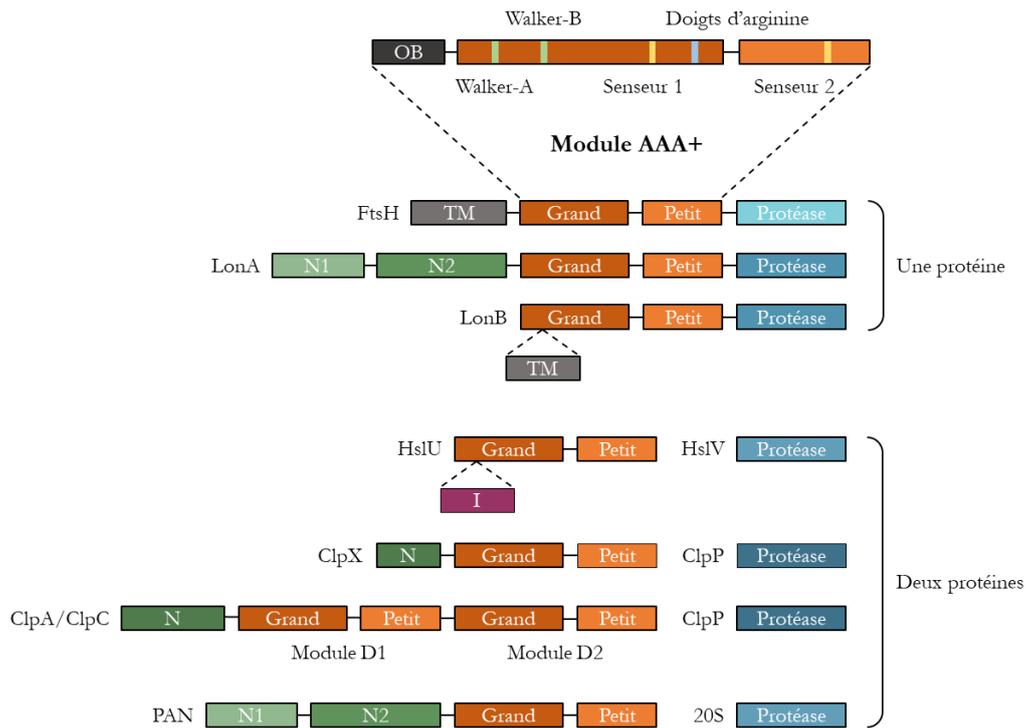
sous forme d'Arg ou de Lys et se trouve près du début de  $\alpha 7$ . Il sert de médiateur des changements de conformation associés à un cycle de liaison et d'hydrolyse de l'ATP<sup>269</sup>. Ceci se produit généralement par l'interaction directe d'un résidu du senseur 2 avec le phosphate  $\alpha$  de la molécule d'ATP liée. Le senseur 3 est sous forme de résidu d'Arg ou d'His. Ce senseur joue un rôle dans la stabilisation des interfaces des sous-unités qui sont nécessaires la fixation de l'ATP<sup>270</sup>.

Le domaine N-terminal  $\alpha 0$  du module AAA+ présente des caractéristiques particulières. Cette région contient généralement une Gly qui forme un bouchon à l'extrémité N-terminale<sup>268</sup>. Sur la base d'alignements de séquences, les protéines ont été classées en groupes en fonction du résidu après cette Gly. Par exemple, où les protéines de type Cdc48/p97 possèdent une autre Gly, et HslU et les Clp ont un résidu hydrophobe<sup>274</sup>. La région N-terminale à ces deux résidus est appelée « N-linker » et contribue à la poche de liaison à l'ATP et sert à relier le domaine AAA+ à d'autres domaines d'une protéine<sup>274</sup>. En raison du positionnement du N-linker directement adjacent au site de liaison ATP et entre les domaines d'une protéine, ce motif peut jouer un rôle direct dans le couplage de l'hydrolyse de l'ATP aux changements conformationnels.

Les protéines appartenant à la famille des protéines AAA+ sont donc structurellement similaires, mais fonctionnellement diverses en raison des éléments N- ou C-terminaux ajoutés au noyau  $\alpha$ - $\beta$ - $\alpha$  conservé et parfois des insertions d'éléments structuraux secondaires spécifiques à l'intérieur de ce noyau. Par conséquent, les membres de la superfamille AAA+ ont été classés dans des sous-familles basées sur leur séquence spécifique et leurs propriétés structurales<sup>268</sup>. Parmi elles, le clade 3 représente une famille de protéines qui sont fonctionnellement apparentées et forment des structures cycliques hexamériques avec un canal central et qui sont souvent associées aux protéases AAA+. Une caractéristique fonctionnelle déterminante de cette famille est une fonction de remodelage des protéines. Les membres de cette famille AAA+ sont définis par une courte insertion d'hélice  $\alpha$  entre  $\alpha 2$  et  $\beta 2$ . Cet élément structural forme une boucle qui est positionnée près du canal central de l'hexamère qui se lie au substrat protéique. Les membres du clade 3 AAA+ sont eux-mêmes différents sur le plan fonctionnel, grâce aux contributions d'éléments extérieurs au noyau  $\alpha$ - $\beta$ - $\alpha$ . Ainsi, le clade 3 peut être subdivisé en familles dont celles de protéases AAA+ telles que FtsH-, Protéasomal-, Cdc48 -, et ClpABC. Ces familles sont définies par les caractéristiques des domaines qui sont N- ou C-terminal au noyau AAA+.

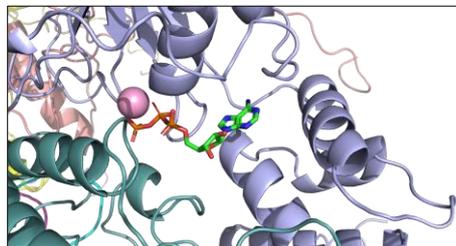
#### 4.2. Particularités des protéases AAA+

Les protéases AAA+ peuvent être divisées en familles distinctes selon les séquences de leur particule régulatrice ATPase, protéase et domaine(s) auxiliaire(s) : ClpXP, ClpAP, ClpCP, HslUV, Lon, FtsH, PAN/20S, Cdc48/20S et le protéasome 26S<sup>182</sup>. La plupart des protéases AAA+ possède un seul module AAA+ (figure 22). Le domaine AAA+ existe en deux copies dans certains membres de la famille AAA+, par exemple ClpA et ClpC et Cdc48/p97 contiennent deux modules AAA+ (désignés D1 et D2) formant deux anneaux distincts<sup>275</sup>.



**Figure 22 : Domaines des protéases AAA+.** Chaque protéase contient un ou deux modules AAA+, chacun est constitué d'un grand et d'un petit domaine et des domaines additionnels spécifiques à chaque famille. Le domaine protéase et le module AAA+ sont présents sur le même polypeptide pour FtsH, LonA et LonB. La protéase et les modules AAA+ sont deux polypeptides distincts dans HslUV, ClpXP, ClpAP, ClpCP et PAN/20S.

Les protéases AAA+ ne présentent pas toutes une homologie significative au niveau de la séquence primaire, ni en ce qui concerne le mode de clivage des protéines. Les conformations des domaines AAA+ sont quant à elles hautement conservées et l'ATP ou l'ADP se lie dans une poche entre le grand et le petit domaine AAA+ d'une sous-unité et le grand domaine AAA+ d'une sous-unité voisine (figure 23). Il y a six sites potentiels de liaison d'ATP/ADP dans un anneau hexamérique constitué d'un seul module AAA+. En outre, la liaison et l'identité du nucléotide modifie la rotation entre le grand et le petit domaine AAA+ d'une sous-unité fournissant un mécanisme pour lier les cycles ATPase et générer les changements conformationnels dans l'anneau hexamérique (voir partie 5.2.3)<sup>182</sup>.



**Figure 23 : Poche nucléotidique de PAN.** Chaque poche nucléotidique est formée par le grand domaine d'une sous-unité AAA+ (violet) et le petit domaine de la sous-unité adjacente (bleu) au sein de la particule régulatrice PAN (PDB 6HEA). La molécule d'ATP est représentée en bâtonnet et l'ion magnésium en sphère magenta.

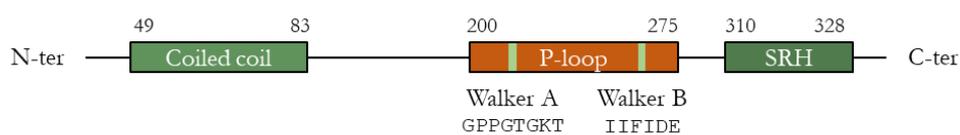
Le protéasome PAN-20S a été étudié dans ce projet de thèse dans le but de mieux comprendre le mode d'action de dépliement et dégradation des substrats protéiques.

## 5. Structure et mode d'action du protéasome archéen PAN-20S

### 5.1. Histoire de l'étude du protéasome archéen PAN-20S

En 1996, le séquençage complet de l'archée méthanogène *Methanocaldococcus jannaschii* a révélé la présence de deux gènes (nommés S4 et S8) homologues des gènes codant pour le 19S ATPases<sup>117</sup>. Pour tester s'il régulaient la protéolyse de façon ATP-dépendante, le gène S4 a été exprimé dans *E. coli* et le produit de 50 kDa (PAN) a été purifié et caractérisé<sup>276</sup>. Le complexe PAN purifié existe en solution sous forme d'un anneau homo-hexamérique<sup>277</sup> et possède une activité ATPase  $Mg^{2+}$ -dépendante<sup>276</sup>. Dans sa caractérisation initiale, il a été décrit que PAN favorisait la dégradation sélective des protéines possédant peu de structure tertiaire (comme la caséine et la RNase A oxydée possédant une étiquette *ssrA* à leur extrémité C-terminale) en présence de l'ATP<sup>276</sup>. Par la suite, il a été montré que PAN participe à la dégradation des protéines globulaires telles que la GFP possédant une étiquette *ssrA* à leur extrémité C-terminale par une réaction nécessitant l'hydrolyse de l'ATP<sup>261</sup>.

La séquence primaire de PAN a révélé que cette protéine appartient à la famille des AAA+ (voir partie 4) et contient un seul domaine AAA+ (résidus 200 à 342) avec un motif P-loop (qui comprend les motifs Walker A et B) et un motif SRH (« *Second Region of Homology* ») du côté C-terminal (figure 24).



**Figure 24 : Représentation schématique de la séquence primaire de PAN de *Methanocaldococcus jannaschii*.** PAN possède des caractéristiques typiques d'une AAA-ATPase, un domaine P-loop avec les motifs Walker A et Walker B et une seconde région d'homologie (SRH) à son extrémité C-terminale. La partie N-terminale de PAN contient une région « coiled-coil ».

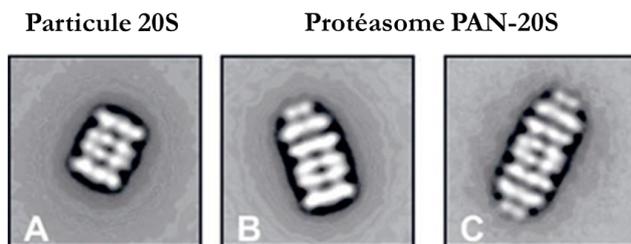
PAN partage 41-45% de similarité de séquence avec les sous-unités du 19S ATPases humain et de la levure<sup>276</sup>, avec un certain nombre de résidus conservés entre PAN et le 19S ATPases dans les motifs P-loop et SRH. De plus, PAN et le 19S eucaryote possèdent tous deux un motif « coiled-coil » à leur extrémité N-terminale<sup>276</sup>.

Comme mentionné dans la partie 3.5.1, le protéasome archéen PAN-20S présente une architecture plus simple que le protéasome 26S eucaryote. En raison de ces similarités et de la conservation associée des propriétés clés du protéasome PAN-20S, mais aussi de par les nombreux avantages qu'offrent un assemblage macromoléculaire provenant d'archée hyperthermophile (voir partie 2.3.2), les études au cours de ces deux dernières décennies sur protéasome PAN-20S ont été réalisées principalement dans le but d'élucider le mode d'action du protéasome 26S eucaryote. Ainsi, la première structure de cristallographie particule catalytique 20S de l'archée *T. acidophilum* en 1995 a été une avancée majeure, révélant la structure cylindrique de quatre anneaux empilés, avec sept sous-unités  $\alpha$  dans chacun des deux anneaux externes et sept sous-unité  $\beta$  dans chacun des deux anneaux centraux<sup>185</sup>. Les centres catalytiques ont été localisés dans la chambre centrale et des études biochimiques et structurales ont révélé des éléments essentiels du mécanisme catalytique du N-terminal nucléophile<sup>186</sup>.

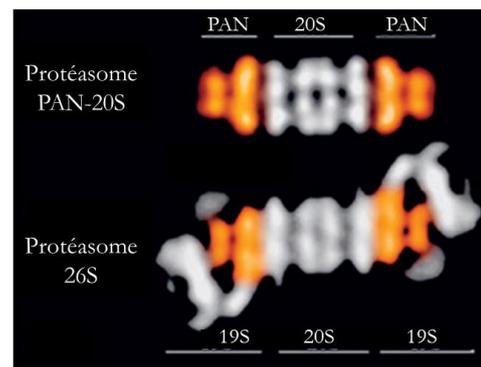
Jusqu'à très récemment, la plupart de la connaissance structurale sur les sous-unités AAA+ ATPases du complexe 19S venaient du complexe d'activation du protéasome (PAN)<sup>278</sup>. L'étude des complexes protéolytiques AAA+ des archées, notamment le protéasome PAN-20S, se sont avérées très précieuses pour comprendre la structure et les mécanismes enzymatiques du protéasome 26S eucaryote<sup>279</sup>.

Bien que l'association d'un complexe AAA+ avec une particule protéolytique semble être une caractéristique commune du système protéolytique AAA+, les particules catalytiques 20S des archées n'ont jamais été purifiées en complexe avec les particules régulatrices de leurs organismes natifs et une association entre PAN et la particule 20S a été difficile à observer par des approches biochimiques standards<sup>280</sup>. Cette difficulté à visualiser le protéasome PAN-20S assemblé a d'abord conduit certains chercheurs de suggérer un mécanisme alternatif pour expliquer la stimulation de la dégradation des protéines par PAN agissant sur le substrat sans formation du complexe PAN-20S<sup>281</sup>. Il a ensuite été suggéré que le complexe avec le protéasome 20S pourrait être difficile à détecter, car il est de courte durée (par rapport au complexe entre la particule 19S et le protéasome 20S eucaryote)<sup>282</sup>. L'association entre PAN et le protéasome 20S d'archée a été démontrée en 2005 par immunoprécipitation, résonance de plasmons de surface (SPR) et microscopie électronique (figure 25)<sup>282</sup>.

A, B, C.



D.



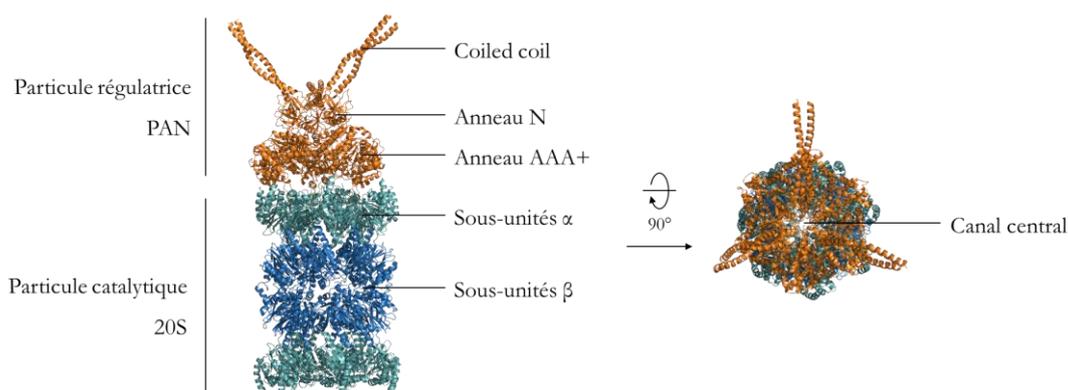
**Figure 25 : Analyse par microscopie électronique du complexe PAN-20S et de la particule 20S d'archée.** Des images du protéasome 20S d'archée (A), du complexe PAN-20S d'un seul côté du 20S (B) et du complexe PAN-20S des deux côtés du 20S (C). (D) Des particules négativement colorées du complexe PAN-20S (deux côtés) sont comparées au protéasome 26S mammifère. Adapté de Smith, D. M., *et al.* (2005)<sup>282</sup>.

La formation du complexe entre PAN et le 20S a été observée en présence d'ATP et ses analogues non-hydrolysables (AMPPNP et ATP $\gamma$ S) mais aucune interaction entre les deux particules n'a observée en présence d'ADP ou de n'importe quel autre nucléotide. Il semblerait donc que l'association de PAN avec le protéasome 20S soit favorisée lors de la liaison d'ATP et réduite par hydrolyse de l'ATP.

Ces études de microscopie électronique sur le complexe PAN-20S ont démontré que PAN peut s'associer au 20S à une ou aux deux extrémités. Ces études ont permis de montrer que PAN s'associe physiquement avec les protéasomes 20S pour que la dégradation ATP-dépendante des protéines puisse se produire<sup>282</sup>. La comparaison des séquences C-terminales de PAN dans diverses archées ainsi que du 19S ATPase eucaryote montre que l'avant dernier résidu est une Tyr chez PAN de 9 espèces d'archée différents (dans 5 autres espèces d'archée c'est une phénylalanine), cette

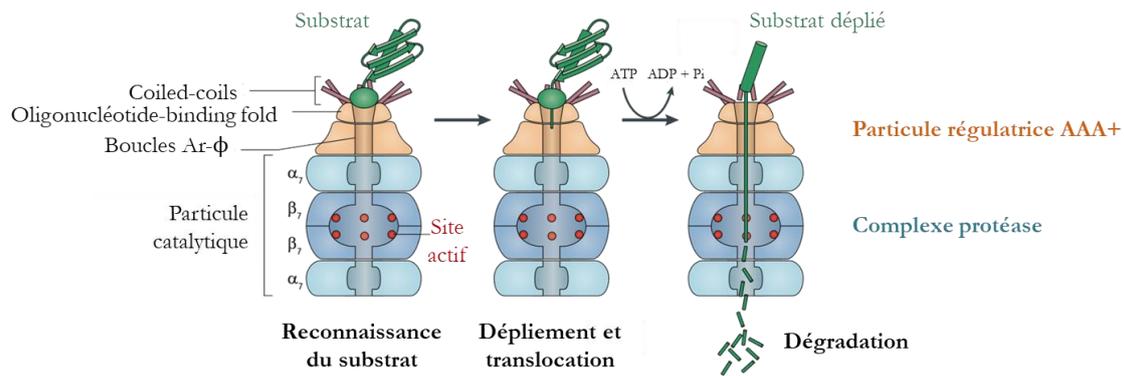
Tyr se retrouve également dans les quatre sous-unités ATPasiques du complexe 19S : Rpt1, Rpt2, Rpt3 et Rpt5 (mais pas dans Rpt4 et Rpt6) chez de nombreux eucaryotes. Un autre résidu conservé est un résidu hydrophobe qui précède la Tyr dans tous les PAN des 14 espèces d'archée ainsi que dans les Rpt2, Rpt3, Rpt5 et Rpt6 chez les eucaryotes. Ainsi ce motif C-terminal conservé appelé HbYX, avec Y la Tyr (ou Phe chez certaines archées) et X un résidu variable, apparaît dans presque tous les protéasomes ATPases connus<sup>283</sup>. Ce motif peut pénétrer dans une poche de fixation des sous-unités  $\alpha$  de la particule catalytique, stabilisant ainsi la formation du complexe entre la particule régulatrice et la particule catalytique. PAN ne s'associe *in vitro* que de façon transitoire à la particule 20S<sup>284</sup>, à moins qu'un pont de cystine (ou « *cross link* ») ne stabilise l'amarrage du motif HbYX dans la poche de fixation des sous-unités  $\alpha$  de la particule catalytique<sup>285</sup>.

Les images de microscopie électronique du protéasome PAN-20S ont permis de vérifier expérimentalement que PAN fonctionne en tant qu'un complexe hexamérique<sup>282</sup>. Bien que la structure du complexe PAN entier n'ait pas été résolue par cristallographie, les structures cristallines des domaines de *Methanocaldococcus janaschii* et d'*Archaeoglobus fulgidus* de PAN ont été obtenues en 2009<sup>262,278</sup>. Comme décrit dans la partie 3.3.2 pour les sous-unités AAA+ du 19S eucaryote, l'anneau hexamérique formé par les domaines AAA+ est positionné à proximité de la particule catalytique et transfère la protéine cible dans la particule catalytique d'une manière ATP-dépendante<sup>262,278</sup>. Cette disposition a été confirmée par la structure du protéasome PAN-20S d'*Archaeoglobus fulgidus* obtenue par cryo-microscopie électronique<sup>286</sup> (figure 26).



**Figure 26 : Structure du protéasome PAN-20S.** Vues latérale (gauche) et supérieure (droite) du protéasome PAN-20S d'*Archaeoglobus fulgidus* (PDB 6HEA). La particule régulatrice PAN est représentée en orange et la particule catalytique 20S est représentée en bleue.

Le canal central de PAN est aligné avec le canal central de la particule catalytique pour former un long tunnel avec une ouverture étroite pour le passage des protéines substrats dans la chambre protéolytique centrale. Les séquences N-terminales adoptent une conformation en hélices  $\alpha$  pour s'associer en trois coiled-coils (avec une hélice  $\alpha$  à chaque sous-unité associée par paires). Pour cela, un résidu Pro conservé à la base de chaque l'hélice N-terminale adopte une conformation *cis*. Cette introduction d'un coude dans l'hélice permet la formation de coiled-coil avec sa sous-unité voisine. Les six sous-domaines OB (de l'anglais « *Oligonucleotide / oligosaccharide-Binding fold* ») forment l'anneau N dont le canal central relativement étroit ( $\sim 15$  Å de diamètre) semble fonctionner comme un filtre en restreignant l'accès aux protéines repliées d'entrer dans le canal central de l'ATPase (figure 27). Pour déplier et transférer une protéine substrat, les boucles internes Ar- $\Phi$  (voir partie suivante) déplacent la protéine cible vers son extrémité C-terminale de PAN<sup>286,287</sup>.



**Figure 27 : Modèle de protéolyse basé sur le complexe d'activation du protéasome (PAN) archéen et la particule catalytique 20S.** Le domaine « coiled-coil » du côté N-terminal de chaque sous-unité de PAN forme une paire avec la sous-unité voisine, les trois paires de coiled-coil sortent de la face distale du domaine ATPase et entourent un pore formé par le domaine OB (« oligonucléotide/oligosaccharide binding fold ») qui peut servir comme point d'entrée pour les substrats protéiques à parcourir le canal ATPase. Les boucles aromatique-hydrophobe (Ar- $\phi$ ) au sein du canal ATPase peuvent exercer une force mécanique vers le bas sur les substrats protéiques avec un processus piloté par l'hydrolyse de l'ATP. Le dépliection de la protéine se produit avec la répétition de ces cycles d'hydrolyse et avec le domaine OB fournissant une plateforme rigide. Adapté de Maupin-Furlow, J. (2012)<sup>156</sup>.

## 5.2. Processus de dépliection des protéines et hydrolyse de l'ATP

### 5.2.1. Importance des boucles Ar- $\phi$

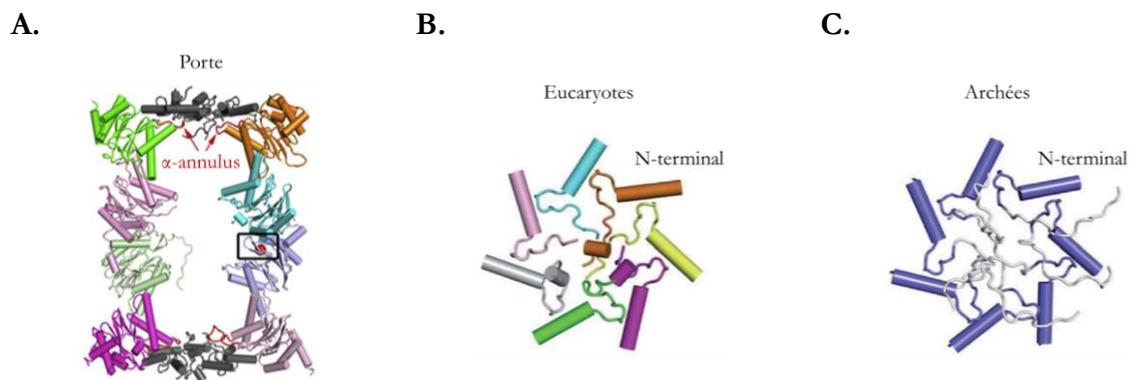
Les boucles Ar- $\phi$  aussi appelées « pore-1 loop » et « pore-2 loop » contiennent un motif constitué par une chaîne latérale aromatique Ar (une Tyr) suivie par une chaîne latérale hydrophobe  $\phi$ . Ces boucles jouent des rôles importants dans la reconnaissance du substrat, et leur fonction est cruciale dans le processus de dépliection et de transfert des protéines substrats en interagissant avec des régions hydrophobes des protéines dépliées<sup>288,289</sup>. Elles sont situées dans la région la plus étroite du canal ATPase formé par le domaine AAA+, à proximité de l'anneau N formé par les domaines OB<sup>290,291</sup>.

L'anneau N rigide exerce une force passive pour le dépliection de la protéine substrat en fournissant un canal qui bloque le mouvement des protéines repliées à travers le canal pour que les boucles Ar- $\phi$  appliquent une force mécanique sur la protéine<sup>262</sup>. Similairement à d'autres protéines AAA+, les boucles Ar- $\phi$  de PAN assurent le dépliection des protéines en appliquant une force mécanique sur les protéines substrats dans le canal ATPase par un mouvement ATP-dépendant du domaine AAA+<sup>262,289</sup>. En effet, le moteur ATPase guide les protéines substrats vers la particule catalytique 20S grâce aux mouvements de pagayage des boucles Ar- $\phi$  qui dépassent dans le canal central et interagissent avec la protéine substrat<sup>292-296</sup>.

### 5.2.2. Importance de l' $\alpha$ -annulus, de la porte de la particule catalytique et des motifs HbXY de la particule régulatrice

Les substrats rentrent dans le protéasome à travers un canal situé dans le centre de chaque anneau heptamérique  $\alpha$ <sup>297</sup>. L'ouverture de ce canal est contrôlée par deux éléments structuraux. Un premier obstacle est le canal étroit connu sous le nom de «  $\alpha$ -annulus » et situé légèrement en dessous de la surface de l'anneau des sous unités  $\alpha$ . Comme ce canal a une ouverture fixe de 13 Å de diamètre, les substrats doivent être dépliés avant qu'ils puissent entrer dans le protéasome. Ce mécanisme

permet de contrôler l'accès à la chambre catalytique du protéasome et empêche la dégradation non sélective des protéines. Le deuxième obstacle est une porte obstruée, formée à partir des résidus N-terminaux des sous-unités  $\alpha$  du protéasome. Le mécanisme de fermeture de cette porte est différent entre archées et eucaryotes (figure 28). Chez les eucaryotes, la fermeture adopte une conformation asymétrique définie par les séquences d'acides aminés uniques à chaque sous-unité  $\alpha$ . Cette structure limite l'entrée, même des petites molécules. Son importance est telle qu'elle est conservée des levures jusqu'aux mammifères<sup>298</sup> et que toute perturbation structurale causée par une mutation provoque des anomalies dans la survie chez la levure<sup>299</sup>. Chez les archées, les 12 résidus du côté N-terminal des sous-unités  $\alpha$  du protéasome sont très flexibles et sont situés sur la porte d'entrée ainsi qu'ils peuvent s'étendre à travers l'anneau  $\alpha$  pour empêcher l'accès des protéines repliées au protéasome, cela explique pourquoi le protéasome chez les archées a une activité plus élevée avec les petits substrats peptidiques<sup>250,300</sup>.



**Figure 28 : Porte et  $\alpha$ -annulus de la particule catalytique 20S.** A. Vue en coupe d'un côté du protéasome 20S de *S. cerevisiae*. La région entourant un site actif est indiquée par un rectangle. La région de la porte fermée est colorée en gris. L' $\alpha$  annulus, juste à l'intérieur de la porte, est une ouverture formée par des boucles (rouge) dans la sous-unité  $\alpha$ . B & C. Les extrémités N-terminales des sous-unités  $\alpha$  ferment l'ouverture des particules catalytiques de façon différentes pour l'entrée de protéines substrats. Les particules catalytiques eucaryotes sont contrôlées par les extrémités N-terminales bien ordonnées des sous-unités  $\alpha 2$ ,  $\alpha 3$  et  $\alpha 4$ , qui forment de nombreuses liaisons hydrogène et de van der Waals (B). Les portes des particules catalytiques archéennes sont désordonnées (les résidus en blanc sont très mobiles) (C). Adapté de Stadtmueller, B. M., & Hill, C. P. (2011)<sup>301</sup>.

Ainsi, la dégradation des tri- ou tétra-peptides par le protéasome 20S archéen n'a pas besoin de la présence de PAN et de l'ATP, probablement parce que ces petits peptides peuvent facilement diffuser dans la particule catalytique. Cependant, la présence de PAN et l'hydrolyse de l'ATP est nécessaire pour la dégradation de la GFP<sub>sSrA</sub>, un substrat globulaire replié<sup>261</sup>. Par conséquent, une porte fermée existe donc dans la particule catalytique 20S qui permet l'entrée des petits peptides, mais les protéines globulaires nécessitent la présence de PAN et l'hydrolyse de l'ATP<sup>250</sup>. Par contre, si la GFP<sub>sSrA</sub> est d'abord dénaturée par un traitement acide, ou pour les substrats contenant naturellement peu de structures tertiaires tels que la  $\beta$ -caséine, l'hydrolyse de l'ATP n'est pas nécessaire pour leur transfert dans la particule 20S par PAN. Ainsi, l'ouverture de la porte sans hydrolyse ATP peut permettre un transfert efficace de larges protéines si elles ont été préalablement dépliées, mais le dépliement par PAN nécessite l'hydrolyse de l'ATP<sup>282</sup>.

Le domaine C-terminal contient un motif conservé « hydrophobe-Tyr-X (HbYX) » avec X représentant n'importe quel acide aminé (voir partie 5.1). Ce motif peut pénétrer dans une poche de fixation des sous-unités  $\alpha$  de la particule catalytique, stabilisant ainsi la formation du complexe

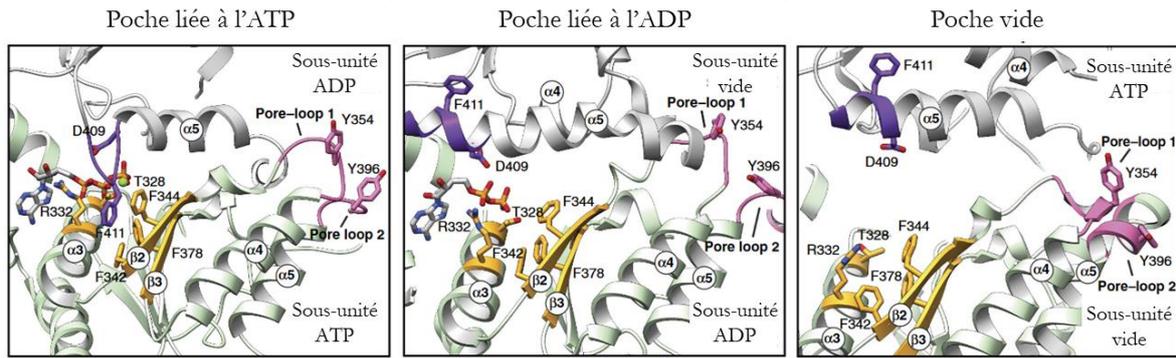
entre la particule régulatrice et la particule catalytique et la conformation ouverte de la porte formée par les extrémités N-terminales de sous-unités  $\alpha$ <sup>283</sup>.

### 5.2.3. Influence des cycles d'hydrolyse de l'ATP

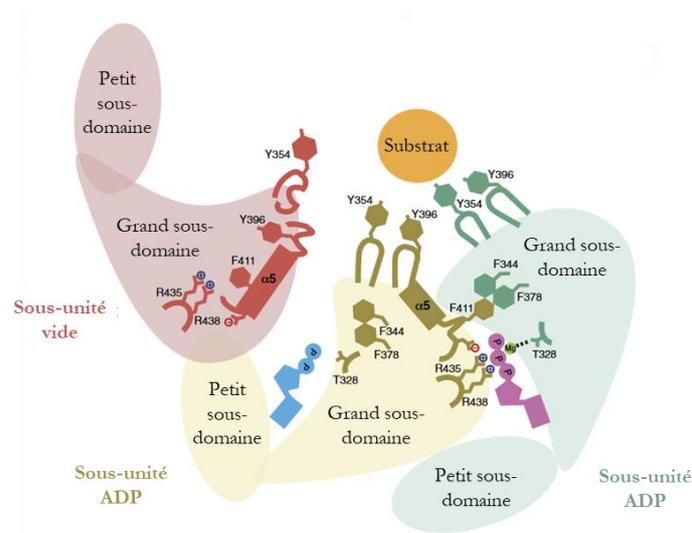
La compréhension des mécanismes de fonction du protéasome nécessite des connaissances approfondies sur sa structure tridimensionnelle, en particulier celle de la particule régulatrice, qui est considérablement plus complexe et flexible que la particule catalytique. Il n'a toujours pas été possible à ce jour de cristalliser l'ensemble du protéasome PAN-20S ou la particule régulatrice PAN entière isolée à cause des difficultés rencontrées lors de la purification du complexe PAN en grande quantité sous sa forme hexamérique monodisperse ainsi que la flexibilité et l'instabilité de la molécule. Cependant, les progrès récents de la cryo-microscopie électronique ont permis très récemment d'obtenir les structures quasi atomiques du protéasome 26S et du protéasome PAN-20S (figure 26). Ainsi, des études approfondies ont permis d'élucider en détail la structure des protéasomes de levures et de mammifères et ont fourni des informations sur les changements conformationnels pendant leur cycle fonctionnel<sup>201,294,302-307</sup>. De plus, il existe actuellement divers modèles séquentiels, concertés ou aléatoires visant à relier l'hydrolyse de l'ATP par l'anneau AAA+ au dépliement et au transfert des substrats protéiques<sup>308-313</sup>. Cependant, la plupart des informations disponibles sont encore insuffisantes pour comprendre complètement les mécanismes, et, dans le cas des protéases, ni l'architecture et les principes mécaniques. La structure du protéasome PAN-20S d'*Archaeoglobus fulgidus* en présence d'ATP non hydrolysable (ATP $\gamma$ S) a été élucidé très récemment par cryo-microscopie électronique à résolution quasi atomique (PAN à  $\sim 4.9$  Å, et la particule catalytique 20S à  $\sim 4.1$  Å)<sup>286</sup> (figure 26). Le mécanisme expliquant comment les mouvements des boucles Ar- $\Phi$  alimentés par l'hydrolyse de l'ATP entraînent le transfert du substrat a pu être ainsi étudié de façon plus fine.

Comme mentionné dans la partie 4.2, les nucléotides se lient aux AAA+ ATPases dans une poche formée à l'interface des grands et petits sous-domaines d'une ATPase et du grand sous-domaine de la sous-unité voisine. Trois types de poches nucléotidiques (ouvertes, fermées et intermédiaires) peuvent être associés à trois états nucléotidiques coexistants, comme c'est le cas pour l'ATPase YME1 mitochondriale homologue de FtsH<sup>310</sup>. Ces trois états nucléotidiques distincts sont en corrélation directe avec les trois modes d'interaction observés entre les Tyr des boucles Ar- $\Phi$  et les protéines substrats, comme représenté de façon détaillée dans la figure 29. L'ATP lié à une sous-unité est détecté par la sous-unité adjacente grâce à deux doigts d'arginine et un motif inter-sous-unités (ISS, de l'anglais « *InterSubunit Signaling motif* ») formant un pont entre les sous-unités à travers la poche de liaison nucléotidique. La fixation de l'ISS positionne les Tyr des boucles Ar- $\Phi$  de la sous-unité liée à l'ATP pour se lier à la protéine substrat. La perte du phosphate  $\gamma$  de l'ATP libère les doigts d'arginine, rétracte l'ISS et repositionne les boucles de pores pour les éloigner du substrat. L'absence de nucléotides dans la sous-unité rompt la coordination des deux côtés et séquestre les boucles de pores en hélices loin du substrat. En résumé, les sous-unités liées à l'ATP se lient directement au substrat, la sous-unité liée à l'ADP s'associe faiblement avec le substrat, tandis que la sous-unité vide est complètement désengagée, comme illustré dans la figure 29. Il y a donc une corrélation directe entre les états nucléotidiques et l'engagement du substrat dans la particule régulatrice AAA+.

A.



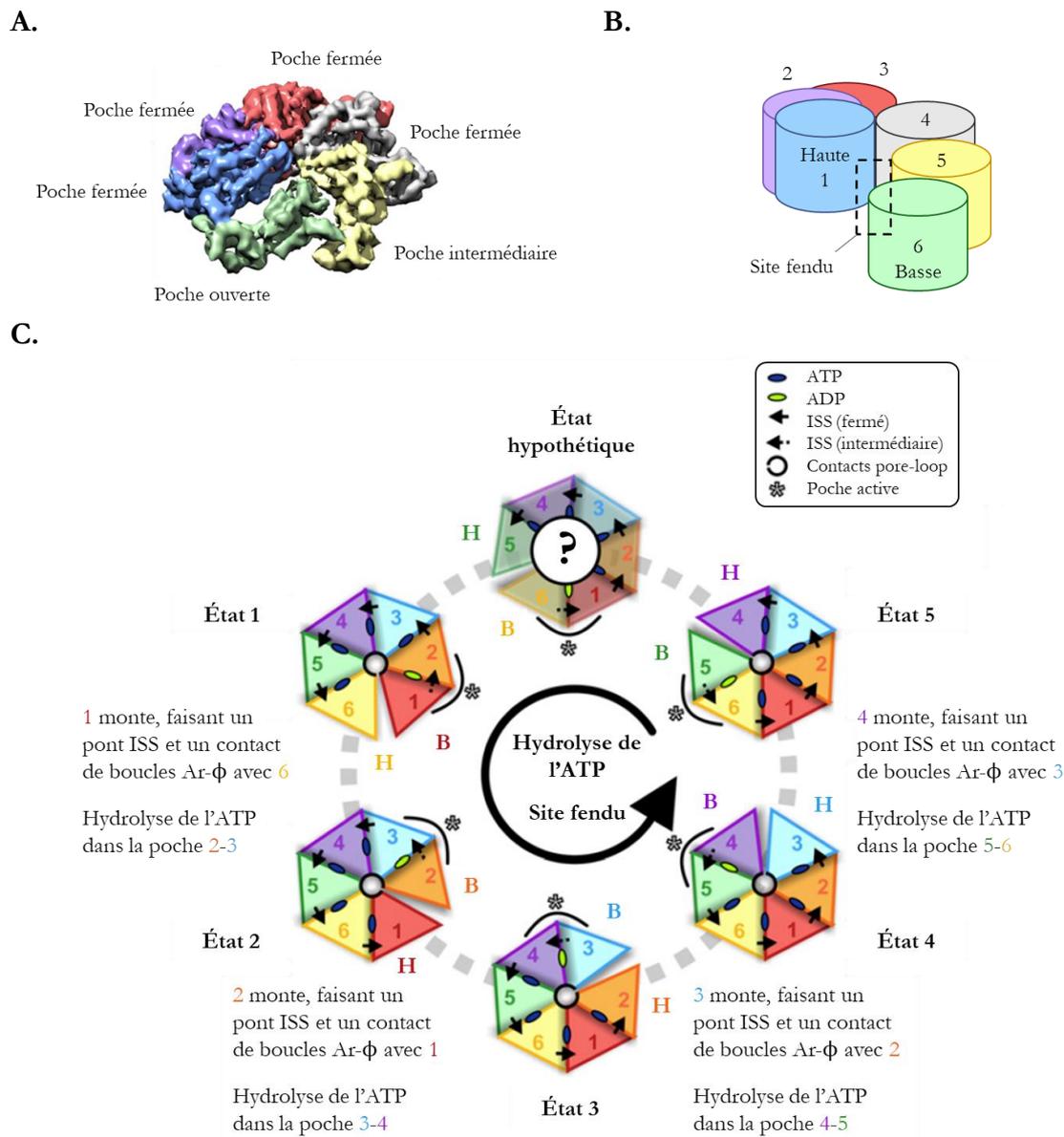
B.



**Figure 29 : Réarrangements conformationnels dans le domaine de l'ATPase YME1 mitochondriale.** **A.** Les trois conformations d'état nucléotidiques sont alignées pour montrer les principaux réarrangements qui affectent allostériquement la conformation des boucles Ar- $\phi$ . Dans la poche ATP (panneau de gauche), la boucle ISS de la sous-unité adjacente (violette) s'étend vers le feuillet  $\beta$  contenant la Phe (jaune). Dans cette conformation, les boucles Ar- $\phi$  (roses) s'étendent vers le canal central et interagissent fortement avec le substrat. Dans la poche liée à l'ADP (panneau central), la perte du phosphate  $\gamma$  entraîne une rétraction de la boucle de l'ISS du feuillet  $\beta$  contenant la Phe, et la boucle de l'ISS devient partie de l'hélice  $\alpha 5$  de la sous-unité adjacente. Avec la sous-unité liée à l'ADP, l'hélice  $\alpha 3$  se rapproche du feuillet  $\beta$  contenant Phe, et les Tyr des boucles Ar- $\phi$  interagissent plus faiblement avec le substrat. En l'absence de nucléotides (panneau de droite), les Tyr des boucles Ar- $\phi$  sont complètement dissociées du substrat et sont incorporés dans les hélices  $\alpha 4$  et  $\alpha 5$ . **B.** Représentation schématique des modifications conformationnelles dépendantes des nucléotides affectant l'ensemble de l'ATPase avec les mouvements décrits en (A). Lorsque l'ATP est lié dans la poche nucléotidique, les Tyr des boucles Ar- $\phi$  interagissent fortement avec le substrat. L'hydrolyse de l'ATP entraîne un affaiblissement de la liaison substrat-Tyr et la perte du nucléotide se traduit par une libération complète du substrat. Adapté de Puchades, *et al.* (2017)<sup>310</sup>.

Le modèle actuel décrit que le cycle ATPase progresse sous forme de rotation comme un escalier en colimaçon<sup>310</sup>, et affecte allostériquement la conformation des boucles Ar- $\phi$  qui interagissent avec la protéine substrat. Les conformations des sous-unités sous forme d'escalier en spirale (tel que représenté dans la figure 29) sont typiques de la clade 3 de la superfamille des AAA+ et ont déjà été signalées pour Cdc48/p97<sup>314</sup>, Hsp104<sup>308</sup>, Vps4<sup>309</sup>, YME1<sup>310</sup>, Rpts<sup>201,294,305,306</sup>, etc. Cinq conformations en spirale tournante ont été identifiées dans l'anneau AAA+ de PAN, révélant un

cycle d'ATPase presque complet<sup>286</sup>. Un modèle d'hydrolyse de l'ATP autour de l'anneau AAA+ de PAN a été proposé (figure 30).



**Figure 30 : Changements de conformations des sous-unités de PAN lié à l'hydrolyse de l'ATP.** **A.** Structure de l'anneau AAA+ de PAN illustrant la communication entre les différentes sous-unités. Les différentes sous-unités sont représentées en différentes couleurs. **B.** Diagramme représentant les sous-unités de PAN en différentes couleurs permettant de visualiser la forme en escalier en colimaçon et le site fendu. **C.** Modèle du cycle ATPase dans les protéasomes PAN-20S. Les cinq états identifiés de l'anneau AAA+ et l'état hypothétique sont représentés de façon schématisée. Dans chaque état, la sous-unité dans les positions la plus haute et la plus basse est indiquée respectivement par H et B, et la poche active (lié à l'ADP) est indiquée par des astérisques. Adapté de Majumder, P., et al. (2019)<sup>286</sup>.

Chez PAN, il existe deux contacts majeurs entre les sous-unités pour former les poches nucléotidiques : le pont ISS et un « contact de boucles Ar- $\phi$  ». Les contacts de boucles Ar- $\phi$  dans les anneaux AAA+ n'ont pas été signalés plus tôt. Dans le cas de PAN, une séparation d'environ 10 Å entre les bouts d'hélice adjacents de  $\alpha 5$  juxtapose les boucles de pores pour interagir mutuellement. Ces contacts limitent la flexibilité des boucles Ar- $\phi$  et positionnent les Tyr de façon appropriée pour interagir avec le substrat. L'ISS et les contacts de boucles Ar- $\phi$  sont inexistantes

dans le site fendu (figure 30). Avec la rotation de l'escalier en spirale, ces contacts sont créés et détruits. Les mouvements entre les sous-unités PAN reflètent l'état d'occupation des poches de liaison nucléotidique.

Les sous-unités formant la poche de liaison nucléotidique grâce à un pont ISS occupé par l'ADP est la poche intermédiaire aussi appelée poche active, où l'hydrolyse de l'ATP a lieu. En raison de l'hydrolyse de l'ATP dans la poche P1-P2, les résidus des doigts d'arginine de P1 (pour « Poche 1 ») perdent le contact avec le nucléotide, et les hélices  $\alpha 6$  (pore-2) et  $\alpha 5$  (pore-1) se déplacent vers le haut, entraînant le mouvement des boucles Ar- $\Phi$  (pore-1 loop et pore-1 loop). Le pont ISS de P1 est incorporé dans son hélice  $\alpha 6$ , et un site fendu est formé entre P1 et P2, avec P1 occupant la position la plus haute et P2, la position la plus basse dans l'escalier hélicoïdal AAA+. Le détachement du pont ISS de P1 à P2 signale le début de l'hydrolyse de l'ATP dans la poche P2-P3, et le cycle se poursuit séquentiellement dans un sens antihoraire autour de l'anneau AAA+. En parallèle, l'ATP est rechargé dans le site fendu (maintenant P1-P2). Ainsi, lorsque l'ATP dans la poche P2-P3 est hydrolysé et que les hélices  $\alpha 6$  (pore-2) et  $\alpha 5$  (pore-1) de P2 se déplacent vers la position la plus haute de l'escalier, la poche P1-P2 acquiert une conformation fermée, le pont ISS de P1 rétablissant le contact avec P2. La rotation de l'escalier en colimaçon entraîne des mouvements rythmiques des hélices  $\alpha 6$  (pore-2) et  $\alpha 5$  (pore-1) qui, à leur tour, permettent aux boucles Ar- $\Phi$  de propulser le substrat dans la particule catalytique.

En résumé, le cycle d'hydrolyse ATP séquentiel est étroitement coordonné. L'hydrolyse dans la sous-unité inférieure liée à l'ATP supprime la coordination entre les doigts d'arginine adjacents, l'ISS et les contacts de boucles Ar- $\Phi$ , repositionnant la sous-unité maintenant post-hydrolyse à la position la plus basse de l'escalier, ce qui, à son tour, déclenche une hydrolyse dans la sous-unité inférieure suivante liée à l'ATP. La perte de coordination des deux côtés de la sous-unité liée à l'ADP rompt l'interaction du substrat et déplace la sous-unité de l'hexamère, où elle peut libérer l'ADP et se lier de nouveau à de l'ATP en haut de l'escalier. La répétition de ce cycle entraîne un transfert progressif du substrat dans la chambre protéolytique grâce aux boucles Ar- $\Phi$ .

Le degré élevé de conservation structurale chez la famille des particules régulatrices AAA+ protéolytiques suggère que ce mécanisme peut être conservé au sein de tous les membres de cette famille. Les bases structurales du cycle séquentiel autour de l'anneau AAA+ lié à l'hydrolyse de l'ATP sont donc probablement conservées chez les protéases AAA+. Un mécanisme similaire a été récemment décrit pendant le dépliage de substrats par cryo-microscopie électronique par le protéasome 26S de levure et humain<sup>305,306</sup>.

Les mouvements des sous-unités régulent allostériquement les mouvements de l'anneau N et l'ouverture de la porte de la particule catalytique. En maintenant la position décalée de PAN fixe par rapport à la particule catalytique, l'escalier AAA+ pivote d'une sous-unité par état. La liaison de PAN stabilise la porte de la particule catalytique dans une conformation ouverte grâce à l'insertion du motif HbYX dans les poches de fixation des sous-unités  $\alpha$  de la particule catalytique. Un contact entre la sous-unité de PAN la plus basse et la partie N-terminale des sous-unités  $\alpha$  de la particule catalytique a également lieu. Ce contact tourne avec la rotation du site fendu. Toutefois, il n'a pas été possible d'identifier sans ambiguïté les résidus impliqués dans la formation du contact, ni le rôle de ce contact. De tels contacts ne sont pas observés dans le cas du protéasome 26S. Cependant, des études antérieures avec ClpXP ont montré que les boucles N-terminales de ClpP

forment des contacts hydrophobes avec les pore-2 loop (Ar- $\phi$ ) de ClpX. Ces contacts varient dynamiquement avec l'état nucléotidique des sous-unités ClpX individuelles et facilitent le dépliement efficace des protéines<sup>315</sup>. Cela suggère un effet allostérique de l'hydrolyse de l'ATP sur les contacts entre PAN et la particule catalytique, et joue probablement un rôle dans le transfert du substrat à travers le canal central.

#### 5.2.4. Modalités de prise en charge des substrats protéiques

Le transfert d'un substrat protéique de la particule régulatrice AAA+ vers la particule catalytique peut commencer par son extrémité N-terminale ou C-terminale, selon l'emplacement de l'étiquette de reconnaissance<sup>240,316</sup>. Certaines études ont suggéré que PAN déplie les protéines à sa surface avant leur transfert dans la particule catalytique 20S<sup>317</sup>. Des travaux plus récents soutiennent un modèle dans lequel les protéines sont dépliées de façon simultanée à leur transfert ATP-dépendant à travers l'anneau ATPase<sup>262,290</sup>. En accord avec ce modèle, des études ont montré que les cycles d'hydrolyse de l'ATP provoquent des changements de conformation dans l'anneau AAA+ alors transmis aux boucles Ar- $\phi$  interagissant et transférant le substrat protéique (voir partie précédente).

La protéase AAA+ la plus utilisée pour l'étude du processus de dépliement et de dégradation des protéines substrat est le système ClpXP bactérien<sup>175</sup>. La spectroscopie de force à l'échelle de la molécule unique a été utilisée pour visualiser directement l'activité mécanique de ClpX. La fixation de ClpXP sur une bille immobilisée et la fixation d'un substrat protéique à domaines multiples sur une autre bille permettent au substrat protéique d'être capturé par la protéase. Au fur et à mesure que la réaction de dégradation progresse, le dépliement alimenté par l'ATP et le transfert subséquent de chaque domaine modifient la distance entre les deux billes, fournissant une information directe sur l'activité de dépliement de ClpXP<sup>318,319</sup>. Le dépliement des multiples domaines protéiques se fait généralement de manière très rapide (entre de 1 et 5 ms) et coopérative, la constante de temps de dépliement correspondant à la constante de temps de l'hydrolyse de l'ATP. Cela indique qu'un événement d'hydrolyse entraîne le dépliement de la protéine substrat<sup>318,319</sup>, tout comme décrit dans la partie précédente. Cependant, les domaines protéiques stables (sur le plan mécanique) nécessitent un temps de dépliement plus lent et de nombreux ATP sont hydrolysés avant que ces substrats protéiques soient dépliés<sup>316,319,320</sup>. Dans ces cas, la fonction de l'anneau AAA+ suit un modèle probabiliste où chaque tentative de dépliement a une probabilité constante (généralement faible) de déplier un domaine protéique<sup>312</sup>.

ClpXP dégrade certains substrats protéiques à un taux proportionnel au taux d'hydrolyse de l'ATP. Par exemple, si le taux d'hydrolyse de l'ATP est réduit de 10% en utilisant des concentrations d'ATP en dessous de  $K_M$ , le taux de dépliement/dégradation est également réduit de 10%<sup>289</sup>. Pour d'autres protéines substrat, tels que la GFPssrA, ClpXP cesse le processus de dégradation lorsque la concentration d'ATP est sous une valeur seuil critique<sup>289,321,322</sup>. La prise en charge de ces substrats protéiques exige un taux minimum d'hydrolyse d'ATP, suggérant que de multiples événements d'hydrolyse coordonnés sont nécessaires avant que le dépliement se produise<sup>289</sup>. Dans ce cas-là, il est possible qu'un seul mouvement déplie seulement une partie du substrat, qui nécessite une action ultérieure sur la structure restante pour un dépliement total. Dans ce contexte, il a été suggéré que l'adhérence au substrat protéique est réduite lorsque l'anneau ClpX n'est que partiellement saturé en ATP, ce qui affecte les protéines les plus stables qui nécessitent une adhérence forte pour être dépliées<sup>323</sup>. ClpXP peut même libérer des protéines partiellement dégradées<sup>244,289,316,324</sup>. La libération

de protéines substrat stables (sur le plan mécanique) partiellement dégradées peut se produire en présence d'un autre substrat protéique moins stable. ClpXP peut alors libérer des protéines partiellement dégradées possédant une queue flexible d'environ 40 acides aminés, ce qui correspond la distance du haut du canal ClpX jusqu'aux sites actifs du ClpP. Ces protéines libérées échappent alors à une dégradation supplémentaire puisqu'elles ne contiennent plus de dégrons<sup>324</sup>. Il a été proposé que la libération de substrats protéiques stables au profit de la prise en charge de protéines substrat moins stables permettrait à ClpXP de dégrader préférentiellement les substrats protéiques les moins stables dans une population. Ce mécanisme empêche de piéger l'enzyme dans des cycles de tentative de dépliement et assure que l'énergie de l'hydrolyse de l'ATP soit utilisée efficacement pour la dégradation des protéines<sup>316</sup>.

Aucune étude similaire n'a été réalisée sur le protéasome PAN-20S. Les informations recueillies sur la protéase AAA+ ClpXP sont probablement valables pour toutes les protéases AAA+ de par le degré élevé de conservation structurale chez la famille des particules régulatrices AAA+ protéolytiques. Cependant, des études sur d'autres protéases AAA+, dont le protéasome PAN-20S, pourraient apporter de nouvelles informations sur le mode d'action des protéines de cette famille. Des études en solution sur ces systèmes pourraient de plus compléter les connaissances apportées par ces études réalisées à l'échelle de la molécule unique sur un système immobilisé.

Très peu d'études structurales existent sur les substrats pendant leur dépliement et leur transfert au sein de protéasomes. Les récentes études par cryo-microscopie électronique des protéasomes 26S de levure et humain sur le processus de transfert des substrats protéiques ont permis de visualiser des protéines substrat engagées dans les canaux centraux de la particule régulatrice et de la particule catalytique, jusqu'au site actif<sup>305,306</sup>. Cependant, la cryo-microscopie électronique propose des images instantanées des systèmes afin de reconstituer des cycles complets et cette technique ne permet pas de suivre en temps réel les changements structuraux ayant lieu. L'aspect « statique » des études structurales réalisées jusqu'à présent rend nécessaire de développer et d'appliquer des techniques biophysiques en solution et résolues en temps pour suivre simultanément la liaison et l'hydrolyse de l'ATP ainsi que les changements conformationnels de la machinerie protéolytique et des substrats protéiques pendant le processus de dépliement et la dégradation.

## 6. Problématiques et objectifs de la thèse

Les nombreuses expériences biochimiques, biophysiques et structurales ont considérablement amélioré notre compréhension des mécanismes moléculaires de la dégradation des protéines au cours des dernières décennies.

Les expériences à l'échelle de la molécule unique sur le système ClpXP ont mis en lumière les relations entre le dépliement et le transfert de substrats protéiques de façon corrélée à l'hydrolyse de l'ATP (voir partie 5.2.4). Les études structurales par cryo-microscopie électronique sur les protéasomes ont quant à elles révélé des modèles de liaison à l'ATP et des changements conformationnels concomitants (voir partie 5.2.3). Ces études ont fourni de nouvelles informations sur la corrélation entre l'hydrolyse ATP, les changements conformationnels coordonnés des sous-unités des particules régulatrices AAA+, et l'interaction et la progression du substrat de la particule régulatrice vers la particule catalytique. Très récemment, des structures de résolution atomique du protéasome 26S de la levure et humain avec un substrat engagé ont été obtenues. Bien que ces études aient permis de rapides et considérables progrès dans nos connaissances des protéases

AAA+, on ne possède que peu de données structurales en solution sur le protéasome et ses substrats. En particulier, des informations structurales directes sur les substrats, leurs états dépliés putatifs, les produits de dégradation, ainsi que l'évolution de leurs populations respectives au cours du temps pendant le processus de dégradation, demeurent rares à ce jour<sup>40</sup>. Dans ce contexte le protéasome PAN-20S des archées hyperthermophile est un objet d'étude de choix.

**Cette thèse a pour objectif principal d'apporter des informations sur le mode d'action du protéasome PAN-20S de *M. jannaschii*, et en particulier sur le processus de dépliement des substrats protéiques.** Ce système présente tous les avantages des protéines provenant d'organismes hyperthermophiles (voir partie 2.3.2), mis à profit pour les études biophysiques par diffusion de neutrons aux petits angles résolue en temps (TR-SANS, de l'anglais « *Time-Resolved Small Angle Neutron Scattering* »). Un substrat modèle a été utilisé lors de ces études. Il s'agit de la protéine fluorescente verte (GFP, de l'anglais « *Green Fluorescent Protein* ») modifiée par une étiquette d'adressage ssrA à l'extrémité C-terminale (GFPssrA). L'étude des changements conformationnels de la GFPssrA pendant le processus de dépliement et de dégradation par le système protéolytique PAN-20S était l'objectif principal de cette thèse. **Le processus de dépliement et de dégradation de la GFPssrA par le protéasome PAN-20S a été suivi en temps réel par des mesures biochimiques et par une approche biophysique novatrice, le TR-SANS.**

Le SANS permet d'étudier les structures et les changements de conformation des protéines solubles et des complexes qu'elles forment<sup>325,326</sup>. Contrairement au SAXS, technique plus courante et plus répandue, le SANS est capable de distinguer différentes protéines au moyen de leur deutération sélective<sup>327</sup> et de la variation du contraste des solvants (échange H<sub>2</sub>O/D<sub>2</sub>O)<sup>328-330</sup>. En particulier, les protéines naturelles (c'est-à-dire hydrogénées) peuvent être rendues « invisibles » à environ 42 % de D<sub>2</sub>O et le signal SANS peut être attribué presque exclusivement aux protéines deutérées<sup>37</sup>. Compte-tenu de la complexité du système étudié en termes de taille et de sa grande flexibilité en solution, le SANS, combinée avec la deutération et la variation de contraste, est une des rares techniques capables de fournir des données structurales en solution en distinguant les différents partenaires. En se basant sur le principe de la variation de contraste (voir chapitre 2, partie 3.5), la GFPssrA utilisée lors des mesures SANS a été deutérée, tandis que les complexes PAN et 20S étaient hydrogénés. Le tampon contenait quant à lui 42% D<sub>2</sub>O et 58% H<sub>2</sub>O (figure 39). Ainsi, le signal du système protéolytique étant masqué, la courbe de diffusion enregistrée correspond uniquement au signal de la protéine GFPssrA deutérée.

Au cours des vingt dernières années, le SAXS a subi une importante évolution au niveau du développement instrumental et des applications aux systèmes biologiques<sup>331,332</sup>. Cependant, le SANS étant moins utilisé que le SAXS, beaucoup moins de développements ont été réalisés au niveau des instruments dédiés aux applications biologiques. **Ce projet de thèse qui couple la spectroscopie de fluorescence en ligne et le suivi en temps réel d'une réaction biologique présente donc un fort intérêt méthodologique.** Il s'agit du développement du projet précédent concrétisé par la thèse de Ziad Ibrahim.

Le court temps d'exposition de 30 secondes utilisé pour la collecte des données de chaque courbe SANS mesurée a permis de suivre la cinétique de dépliement et de dégradation de façon résolue en temps pendant les 45 minutes de réaction. Ce temps d'exposition est particulièrement intéressant lors des 10 premières minutes de réaction où l'essentiel de la réaction a lieu. Les données recueillies pendant cette période recèleront donc beaucoup d'informations. De plus, la mesure de la

fluorescence de la GFP<sub>ssrA</sub>, couplée à la collecte des courbes SANS, a fourni des informations supplémentaires sur le processus en faisant la connexion entre le dépliement de la GFP<sub>ssrA</sub> du point de vue structural et la perte de sa fluorescence due à la déstabilisation du chromophore. Il faut également souligner que **l'utilisation d'un système d'archée hyperthermophile comme modèle d'étude fournit un avantage essentiel en permettant la thermoactivation du système PAN-20S, contrôlant ainsi la cinétique enzymatique d'hydrolyse de l'ATP et le processus de dépliement et de dégradation.**

Cette approche, combinée avec d'autres techniques biochimiques et structurales utilisées durant cette thèse (mesures de la consommation d'ATP, spectrométrie de masse, SAXS, microscopie électronique) a constitué un outil puissant pour réaliser une caractérisation fonctionnelle du protéasome PAN-20S qui nous a permis de mieux comprendre le mode d'action de cette machine moléculaire.

En effet, on ne sait toujours pas de façon claire si le protéasome PAN-20S déplie progressivement ses substrats en libérant des produits partiellement dépliés puis en les dégradant dans une seconde étape ou bien si le dépliement et la dégradation des substrats est un processus linéaire et concerté en une seule étape. Une seule machinerie PAN-20S pourrait également utiliser plusieurs mécanismes alternatifs, comme il a été décrit pour la protéase ClpXP. L'utilisation de techniques biophysiques en solution et résolues en temps pour suivre les changements conformationnels des substrats protéiques pendant le processus de dépliement et la dégradation, couplée à des mesures en temps réel sur l'hydrolyse de l'ATP, pourrait permettre de mieux comprendre le mode d'action de ces machines moléculaires.

Au cours de cette thèse, les protéines du système ont été produites au laboratoire. Un travail d'optimisation et de caractérisation approfondie a été réalisé afin de maîtriser la qualité de leur préparation et notamment de contrôler l'état oligomérique de la particule régulatrice PAN (chapitre 5). Des études ont ensuite permis de caractériser l'activité du système GFP<sub>ssrA</sub>, protéasome PAN-20S afin de définir les conditions (concentrations des composants, température, temps de mesure) permettant une activité optimale pour les mesures TR-SANS. Des mesures d'hydrolyse de l'ATP par PAN ont également été réalisées afin de coupler ces informations à celles sur le dépliement et dégradation de la GFP<sub>ssrA</sub> (chapitre 6). Ce travail de biochimie a été essentiel pour déterminer les conditions expérimentales optimales pour les mesures TR-SANS et obtenir des résultats exploitables sans biais expérimentaux. Une fois ce travail effectué, les mesures en TR-SANS ont alors pu être appliquées au système GFP<sub>ssrA</sub>, protéasome PAN-20S (chapitre 7). Pour compléter cette étude principale, une seconde stratégie a été développée visant à bloquer la GFP<sub>ssrA</sub> au cours de son dépliement par PAN. Pour cela, une molécule de streptavidine monovalente a été attachée à l'extrémité N-terminale ou C-terminale de la GFP<sub>ssrA</sub> dans le but de provoquer un encombrement stérique<sup>317</sup> (chapitre 5). L'activité du système GFP<sub>ssrA</sub>-streptavidine, protéasome PAN-20S dans les conditions d'études nécessaires aux mesures TR-SANS a ensuite été caractérisée (chapitre 8) avant de réaliser de nouvelles mesures en TR-SANS (chapitre 9).

## Chapitre 2 :

# Présentation des méthodes biophysiques principales utilisées : La diffusion de rayons X et de neutrons aux petits angles

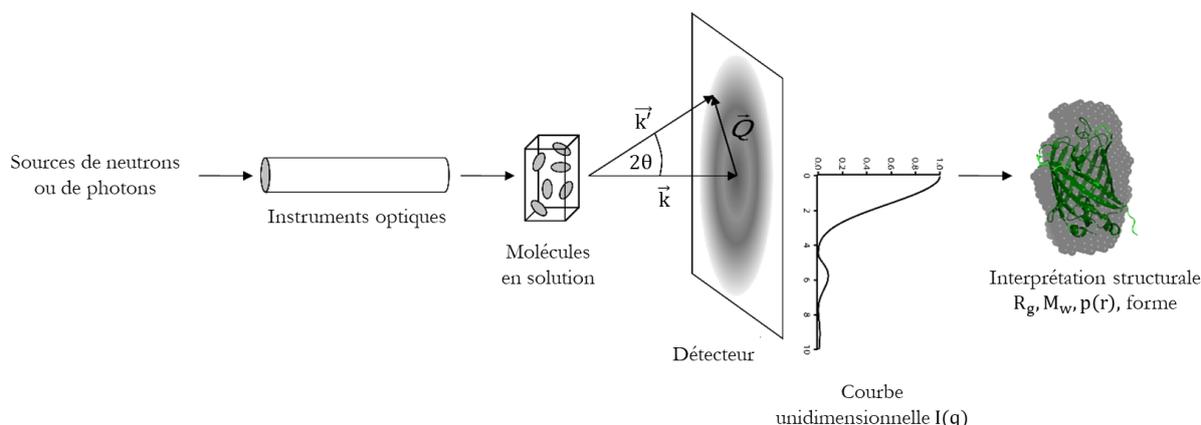
### 1. La diffusion aux petits angles (SAS)

Les méthodes de diffusion aux petits angles sont des techniques utilisées dans le cas d'études biologiques pour obtenir des informations structurales à basse résolution sur des molécules en solution<sup>328,333-335</sup>. Tout comme la lumière visible, les rayons X et les neutrons peuvent être diffusés lorsqu'ils interagissent avec la matière. Lors de la diffusion de rayons X aux petits angles (SAXS, de l'anglais « *Small Angle X-ray Scattering* »), la diffusion résulte de l'interaction entre les photons et les nuages électroniques des atomes. Lors de la diffusion de neutrons aux petits angles (SANS, de l'anglais « *Small Angle Neutron Scattering* »), la diffusion provient de l'interaction entre les neutrons et les noyaux des atomes. La diffusion aux petits angles (SAS, de l'anglais « *Small Angle Scattering* ») permet d'obtenir des informations sur les distances au sein des particules lorsque celles-ci sont plus longues que les distances inter-atomes. Les échelles de longueur sondées sont de l'ordre de la longueur d'onde du rayonnement utilisé, qui varie entre 1 Å pour les synchrotrons à 6-15 Å pour les neutrons.

Lors d'une expérience de diffusion aux petits angles, une solution tampon contenant des macromolécules est placée dans un faisceau monochrome de rayons X ou de neutrons. Les objets en solution diffusent alors le rayonnement incident et la diffusion à proximité du faisceau direct est enregistrée sur un détecteur (figure 31). De façon simplifiée, l'image obtenue sur le détecteur a pour centre le faisceau incident et, les molécules en solution ayant des orientations aléatoires, leurs intensités diffusées forment un cliché avec une symétrie concentrique par rapport au centre. Ces cercles concentriques représentent donc les intensités de diffusion de l'échantillon en fonction de l'angle de diffusion. La moyenne radiale de l'intensité diffusée autour du faisceau incident permet alors d'obtenir la courbe de diffusion unidimensionnelle de l'échantillon. Aux petits angles de la courbe de diffusion, les structures globales de la protéine telles que sa forme dominent le signal, tandis que les propriétés structurales de courtes distances au sein la protéine sont observées aux plus grands angles de la courbe.

Pour réaliser une mesure de diffusion aux petits angles dont les données sont exploitables, il est nécessaire que les molécules en solution soient assez diluées pour éviter les interactions inter-particulaires, on parle alors d'une solution « idéale ». Des informations structurales sur les particules en solution peuvent alors être obtenues. Ces informations incluent la masse moléculaire moyenne des particules  $M_w$ , leur rayon de giration  $R_g$  moyen ainsi qu'une reconstruction de leur enveloppe à une résolution de l'ordre du nanomètre. Si l'échantillon en solution est monodisperse, c'est-à-dire

composé que de particules complètement identiques (même type de particule et même état d'oligomérisation), toutes ces informations correspondront à celles de la particule d'intérêt.



**Figure 31 : Représentation schématique d'un dispositif expérimental d'une expérience de diffusion aux petits angles.** Adapté de Mahieu, E., & Gabel, F. (2018)<sup>37</sup>.

Les techniques de diffusion aux petit-angles sont parfaitement complémentaires aux informations structurales fournies par cristallographie, résonance magnétique nucléaire (RMN) et microscopie électronique<sup>325</sup>. Lors des études à basse résolution, l'avantage principal du SAS est de donner des informations directes sur la structure globale des molécules en solution, sans les artefacts spécifiques à la cristallographie ou à la microscopie électronique et sans les restrictions de taille liées à la RMN (voir partie 1.2). On pourra notamment étudier des protéines flexibles, dynamiques et/ou complexes avec l'avantage non négligeable de pouvoir contrôler le tampon et de l'ajuster à des conditions proches des conditions natives des protéines étudiées telles que la concentration en sels, la température, le pH, la pression, ou l'ajout de substrats.

### 1.1. Théorie

Cette partie a été rédigée en se basant principalement sur les livres *Elementary scattering theory: for X-ray and neutron users*<sup>336</sup> et *Experimental neutron scattering*<sup>337</sup>.

### 1.1.1. Géométrie d'une expérience de diffusion

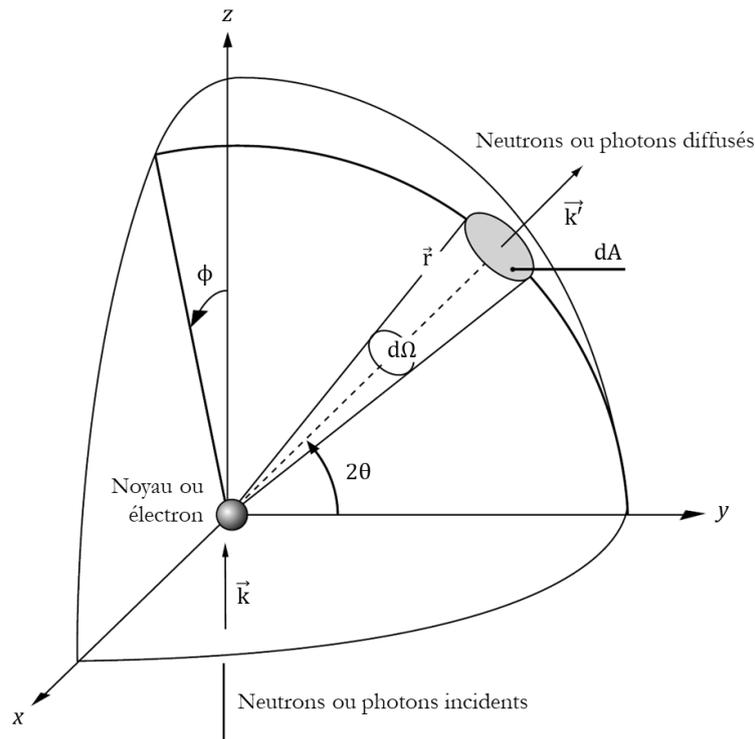


Figure 32 : Géométrie d'une expérience de diffusion. Adapté de Carlile, C. J. (2013)<sup>337</sup>.

La figure 32 illustre la géométrie d'une expérience de diffusion. Un neutron ou photon incident, spécifié par son vecteur d'onde  $\vec{k}$ , où  $k = 2\pi/\lambda$ , est diffusé dans un nouvel état ayant un vecteur d'onde  $\vec{k}'$ . L'origine des coordonnées est à la position du noyau ou de l'électron et le neutron ou photon est diffusé selon un vecteur  $\vec{k}$ . La direction de diffusion est définie par l'angle  $\phi$  et par l'angle  $2\theta$  entre les faisceaux incidents et diffusés. La diffusion se produit dans un cône élémentaire d'angle solide  $d\Omega$ . Si la diffusion est élastique, l'amplitude du vecteur d'onde est inchangée lors de la diffusion, c'est-à-dire  $|\vec{k}| = |\vec{k}'|$ . Si la diffusion est inélastique, il y a un échange d'énergie entre le neutron et l'échantillon : quand le neutron gagne de l'énergie  $|\vec{k}| < |\vec{k}'|$ , et quand il perd de l'énergie  $|\vec{k}| > |\vec{k}'|$ .

Une fois que les neutrons ou photons ont interagi avec l'échantillon, la mesure des neutrons ou photons diffusés permet de déterminer la section efficace de diffusion des neutrons ou photons.

La section totale  $\sigma_{tot}$  est définie par :

$$\sigma_{tot} = \frac{\text{nombre de neutrons ou photons diffusés dans toutes les directions par seconde}}{\text{flux incident } I_0}$$

Équation 1

De façon simplifiée, la section efficace de diffusion  $\sigma$  est une mesure de la taille d'un atome tel qu'il apparaît pour un neutron ou photon et mesure donc l'efficacité avec laquelle les neutrons ou photons sont diffusés par le noyau ou les électrons.

### 1.1.1. Diffusion élastique et transfert d'impulsion

Le profil de diffusion est usuellement décrit par l'intensité  $I$  en fonction de l'amplitude du vecteur de diffusion ou le transfert d'impulsion  $\vec{q}$ . Dans le cas de la diffusion aux petits angles, on suppose que les photons ou les neutrons incidents n'échangent pas d'énergie avec l'échantillon, on ne considère alors que la diffusion élastique. Selon cette hypothèse, les faisceaux incidents et diffusés ont la même longueur d'onde  $\lambda$ . Si  $\vec{k}$  et  $\vec{k}'$  sont respectivement les vecteurs d'onde des faisceaux incidents et diffusés, et que le vecteur de diffusion  $\vec{q} = \vec{k}' - \vec{k}$ , alors :

$$q = |\vec{q}| = \frac{4\pi}{\lambda} \sin\theta$$

Équation 2

où  $2\theta$  est l'angle de diffusion mesuré. Durant l'expérience, la longueur d'onde  $\lambda$  du faisceau incident est fixe. La valeur de  $q$  ne dépend donc que de l'angle  $2\theta$  et les variations d'angle de diffusion sont représentées par  $q$ .

### 1.1.2. Section efficace différentielle

La quantité de base qu'une expérience de diffusion vise à mesurer la fraction de particules incidentes qui diffusent dans diverses directions, telles que définies par les coordonnées polaires sphériques  $2\theta$  et  $\phi$  de la figure 32.

En supposant une diffusion élastique, un détecteur est installé pour compter les neutrons ou photons diffusés dans l'angle solide  $d\Omega$ . Si les neutrons ou les photons forment un flux constant de particules entrantes, le flux incident,  $I_0$ , est habituellement le nombre de neutrons ou photons frappant l'unité de surface de l'échantillon par unité de temps, où la surface est considérée comme étant perpendiculaire au faisceau incident (avec des unités de  $m^2 \cdot s^{-1}$ ). Leur taux d'arrivée dans la direction de  $2\theta$  et  $\phi$  jusqu'au détecteur qui sous-tend  $d\Omega$  (en stéradian sr), peut également être exprimé comme  $R$  par unité de temps ( $s^{-1}$ ).

Le ratio :

$$\frac{\text{Nombre dévié par } (2\theta, \phi) \text{ par seconde par unité d'angle solide}}{\text{Nombre de particules incidentes par unité de surface de faisceau}} = \frac{R(2\theta, \phi)}{I_0 d\Omega}$$

Équation 3

est alors étroitement lié à la fonction de diffusion d'intérêt. Il est connu sous le nom de section efficace différentielle,  $d\sigma/d\Omega$ , en  $m^2 \cdot sr^{-1}$  (surface par unité d'angle solide).

La définition précise de la section efficace différentielle utilisée varie légèrement selon le domaine d'application, mais est souvent citée par atome, ou par molécule, à travers une division par le nombre d'unités de diffusion d'intérêt,  $N$ , de l'échantillon :

$$\frac{d\sigma}{d\Omega} = \frac{R(2\theta, \phi)}{NI_0 d\Omega}$$

Équation 4

où, à proprement parler, l'égalité se situe dans la limite de  $d\Omega \rightarrow 0$ .

Dans les figures 32 et 33, l'angle solide  $d\Omega$  couvert par le détecteur sur l'échantillon est  $dA/r^2$  où  $dA$  est la zone grisée. Le nombre de neutrons diffusés par seconde dans cet angle solide est le flux multiplié par la surface ou  $|\psi_f|^2 \times dA$ , et qui est d'après l'équation 3 égal à  $I_0(dA/r^2)(d\sigma/d\Omega)$ . On a donc :

$$\frac{d\sigma}{d\Omega} = |\psi_f|^2 r^2$$

Équation 5

### 1.1.3. Diffusion par un noyau ou électron fixe

La description de la partie précédente sur la section efficace différentielle portait sur la description des données de diffusion. Décrivons maintenant comment la structure de l'échantillon, au niveau atomique, est liée aux diffusions mesurées.

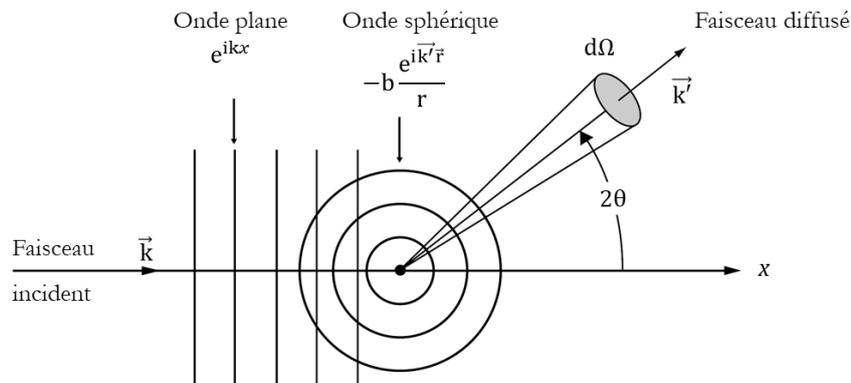


Figure 33 : Schéma de diffusion d'une onde plane de neutrons ou photons par un seul diffuseur. Adapté de Carlile, C. J. (2013)<sup>337</sup>.

Considérons la diffusion par un seul noyau ou électron fixe (ne se déplace pas et n'échange pas d'énergie). Une onde plane incidente de neutrons ou photons se déplaçant dans la direction  $x$  comme représenté dans la figure 33 est déterminée par la fonction onde :

$$\psi_i = e^{ikx}$$

Équation 5

où  $i^2 = -1$  et  $k = 2\pi/\lambda$  est le nombre d'onde.

La probabilité de trouver un neutron ou photon dans un volume  $dV$  est  $|\psi_i|^2 dV$ . Si  $|\psi_i|^2 = 1$ , l'Équation se réfère à une densité d'un neutron ou photon par unité de volume dans tout l'espace.

L'onde diffusée par un noyau ou électron isolé, comme indiqué dans la figure 33, est sphérique avec la fonction d'onde suivante,  $k'$  étant le nombre d'onde sortante :

$$\psi_f = -b \frac{e^{ik'r}}{r}$$

Équation 6

$r$  est la distance par rapport au noyau ou électron diffusant et  $b$  est connue sous le nom de « longueur de diffusion » du noyau ou de l'électron et varie selon les éléments du tableau périodique (et dans le cas des neutrons entre les isotopes d'un même élément) (figure 39). Le signe « - » de l'équation 6 est adopté pour s'assurer que la plupart des valeurs de  $b$  pour les éléments soient positives.

#### 1.1.4. Diffusion par un ensemble de noyaux ou électrons

Après avoir traité le cas d'un seul noyau, si l'on considère un assemblage tridimensionnel de noyaux ou d'électrons tout en maintenant l'hypothèse de la diffusion élastique, l'onde diffusée résultante peut être écrite sous la forme suivante :

$$\psi_s = - \sum_j \left( \frac{b_j}{r} \right) e^{i\vec{k}'\vec{r}} e^{i\vec{q}\cdot\vec{r}}$$

Équation 7

Où  $\vec{q} = \vec{k}' - \vec{k}$  est le vecteur de diffusion avec  $\vec{k}$  et  $\vec{k}'$  les vecteurs d'onde des neutrons incidents et diffusés.

Le nombre de neutrons ou photons diffusés par seconde est le flux fois la surface :

$$\frac{b^2}{r^2} v \cdot 4\pi r^2 = (4\pi b^2) v$$

Équation 8

Ainsi, on obtient :

$$\sigma_{\text{tot}} = \frac{I_f}{I_0} = 4\pi b^2$$

Équation 9

$\sigma_{\text{tot}}$  étant la section efficace du noyau ou de l'électron vue par le neutron ou le photon (en  $\text{m}^2$ ).

##### 1.1.1. Densité de longueurs de diffusion et intensité de diffusion

Selon que les rayons X ou les neutrons sont utilisés, l'intensité de diffusion mesurée correspond à la transformée de Fourier de la densité de longueurs de diffusion du noyau  $\rho_n$  ou de la densité électronique  $\rho_e$ , intégrée sur le volume de particule  $V$ . Cependant, comme les particules sont en solution,  $\rho$  doit être remplacée par  $\Delta\rho$ , le contraste entre la densité de longueurs de diffusion de la particule et celle de la solution tampon  $\rho_s$ , qui est supposée être homogène. La densité de longueurs de diffusion  $\Delta\rho$  se calcule avec les longueurs de diffusion  $b_j$  et le volume  $V_j$  (à la position  $r$ ) que chaque atome occupe au sein d'une particule. Comme la particule est immergée dans une solution de densité de longueurs de diffusion  $\rho_s$  il faut en retirer celle-ci :

$$\Delta\rho = \frac{b_j}{V_j} - \rho_s$$

Équation 10

Dans une solution idéale où les particules sont orientées de façon isotrope, l'intensité de diffusion est en symétrie sphérique autour du faisceau incident. L'intensité de diffusion totale  $I(\mathbf{q})$  peut être alors écrite comme la somme des intensités de diffusion des composants individuels de la solution moyennée sur toutes les orientations possibles (noté  $\langle \dots \rangle$ ) :

$$I(\mathbf{q}) \propto \left\langle \left| \int \Delta\rho \cdot e^{i\mathbf{q}\cdot\mathbf{r}} dV \right|^2 \right\rangle$$

Équation 11

## 1.2. Interprétation des données

### 1.2.1. Poids moléculaire $M_w$

L'intensité diffusée à l'angle zéro ( $2\theta = 0$ ) notée  $I(0)$  est proportionnelle à la masse moléculaire  $M_w$  et à la concentration  $c$  des particules en solution, à condition que les particules aient la même densité de longueurs de diffusion (i.e. la composition chimique moyenne des particules est identique)<sup>338</sup>.

L'intensité à  $\mathbf{q} = 0$  ne peut être mesurée directement, cette zone correspondant au faisceau direct étant masquée par le dispositif d'arrêt de faisceau.  $I(0)$  doit par conséquent être extrapolée depuis les données expérimentales grâce à l'approximation de Guinier (voir partie suivante). Si l'échantillon est monodisperse et que la masse moléculaire  $M_w$  ou la concentration  $c$  est connu, le paramètre inconnu  $M_w$  ou  $c$  peut être obtenu à partir du  $I(0)$  en utilisant l'équation 12<sup>333</sup> :

$$I(0) \propto M_w c$$

Équation 12

### 1.2.2. Rayon de giration $R_g$

En mécanique, le rayon de giration  $R_g$  représente la distance par rapport à l'axe de rotation d'un corps à laquelle il faut placer un point de masse égale à celle du corps pour qu'il ait le même moment d'inertie que celui-ci. Le  $R_g$  d'une molécule est donc un indicateur de sa taille et de sa compacité. Lors de mesures SAS, le  $R_g$  décrit la distribution de la densité de diffusion (par rapport au tampon) autour du centre de densité de diffusion de la molécule. Des objets de mêmes masses mais de formes différentes ont des valeurs de  $R_g$  différentes. Par exemple, un objet sphérique a un rayon de giration plus petit qu'un objet allongé ou que tout autre objet de même masse.

L'approximation de Guinier<sup>339</sup> permet d'obtenir une estimation des intensités  $I(\mathbf{q})$  aux très petits angles, autrement dit pour  $\mathbf{q}R_g \leq 1$ . La diffusion dépendra de la taille de la particule selon une loi générale indépendante de sa forme et de sa symétrie<sup>333</sup>.

$$I(\mathbf{q}) = I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right)$$

Équation 13

En traçant  $\ln(I(\mathbf{q}))$  en fonction de  $q^2$ , les données sont alors sous la forme d'une ligne droite avec une pente de  $-R_g^2/3$  dans la zone de validité de l'approximation de Guinier ( $\mathbf{q}R_g \leq 1$ ), et le rayon de giration peut ainsi être déduit. Pour des protéines globulaires, cette relation est valable pour

$qR_g \leq 1,3$  puisque la zone linéaire est généralement plus large<sup>326</sup>. En extrapolant la droite à l'ordonnée,  $I(0)$  peut être déduit.

Si la solution est monodisperse, le rayon de giration obtenu avec l'équation 13 peut être associé à la forme et à la taille (si la masse moléculaire est connue) de la particule en solution. Dans le cas d'un système polydisperse, la valeur de  $R_g$  est influencée par l'ensemble des particules en solution et sa valeur sera alors une moyenne.

Les valeurs  $q$  plus élevées contiennent des informations sur la forme moléculaire. Pour les macromolécules repliées, l'intensité de la diffusion diminue selon la loi de Porod<sup>340</sup> :

$$I(q) \propto q^{-4}$$

Équation 14

Une autre méthode pour déterminer  $R_g$  et  $I(0)$  consiste à utiliser la fonction de distribution de distances entre les éléments de volume à l'intérieur de la particule, dite fonction de distribution des distances par paires  $P(r)$ <sup>341</sup>. Cette fonction correspond à la probabilité des distances interatomiques et utilise donc le profil de diffusion entier et non pas seulement les valeurs à faible  $q$ . Il s'agit de la transformée de Fourier inverse de la courbe des intensités de diffusion  $I(q)$  d'un échantillon :

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) e^{iqr} dq$$

Équation 15

Dans le cas de solutions monodisperses, il est décrit que :

$$I(q) = 4\pi \int_0^{D_{\max}} P(r) \frac{\sin(qr)}{qr} dr$$

Équation 16

$I(0)$  est donc proportionnel à l'aire sous  $P(r)$  :

$$I(0) = 4\pi \int_0^{D_{\max}} P(r) dr$$

Équation 17

La limite maximale de l'intégrale est  $D_{\max}$ , qui représente la distance maximale entre deux points dans une même particule.

L'équation 18 décrit comment  $R_g$  est relié à  $P(r)$ . Pour l'appliquer, il est nécessaire d'avoir des données avec des statistiques élevées et sans erreurs systématiques<sup>333</sup> :

$$R_g^2 = \frac{\int P(r) r^2 dr}{2 \int P(r) dr}$$

Équation 18

### 1.2.3. Traitement des données et analyse standard en biologie

Les données brutes issues d'une mesure SAS sont réduites par une série de corrections qui comprennent la normalisation de l'image par rapport à l'intensité du faisceau incident, l'étalonnage de  $q$ , la correction pour la transmission de l'échantillon et la soustraction de l'arrière-plan, le masquage des pixels indésirables, et l'étalonnage de l'intensité, suivies d'une intégration radiale.

La suite ATSAS<sup>342</sup> contient plusieurs des programmes spécialement conçus pour analyser les données SAXS et SANS des biomacromolécules en solution une fois les données réduites et constitue la suite de programmes majoritairement utilisée dans le domaine. Les programmes présentés ci-dessous font partie de la suite ATSAS.

Les intensités du tampon sont soustraites des intensités des échantillons respectifs en utilisant le logiciel PRIMUS<sup>343</sup> afin d'obtenir la courbe de diffusion SAS. Différentes informations peuvent être obtenues selon la région de la courbe étudiée en utilisant ce même logiciel (voir figure 34) :

- Les rayons de giration  $R_g$  des échantillons sont extraits en utilisant l'approximation de Guinier<sup>339</sup> (équation 13). L'information du  $R_g$  peut être extraite des premiers points de la courbe en calculant la pente d'une droite les alignant. La possibilité d'aligner une droite sur ces points et de valider l'approximation de Guinier renseigne sur l'état de l'échantillon (agrégé ou non).
- Le contenu en flexibilité de l'échantillon peut être étudié par le tracé de la courbe de Kratky. Les macromolécules globulaires suivent la loi de Porod (équation 14) et ont des courbes en forme de cloche. Les molécules allongées, comme les protéines dépliées, n'ont pas ce pic et sont en légère augmentation ou ont un plateau dans la gamme  $q$  plus large.
- L'analyse de la fonction  $P(r)$  (équation 18) renseigne sur la forme des protéines en solution. Alors que cette courbe a une forme de cloche pour une protéine globulaire, elle présentera une queue plus longue pour les protéines allongées et plusieurs pics pour une protéine multi-domaine. La distance maximale  $D_{max}$  de la protéine peut également être estimée à partir de cette analyse des distances, représentant la distance extrapolée à  $P(r) = 0$ .

Il est courant d'analyser les données obtenues par SAS non pas en étudiant les résultats d'une seule concentration mais en combinant une faible et une forte concentration. Une mesure SAS sur une faible concentration apporte des informations précises aux petits angles, appelée la zone ou région de Guinier, alors que les informations aux plus grands angles ne sont pas accessibles car les données sont trop bruitées. En revanche, les mesures à forte concentration sont moins précises dans la région de Guinier (risque d'agrégation ou de répulsion) mais plus résolutive et donc précises sur le facteur de forme. Une courbe chimère peut être réalisée en fusionnant deux courbes sur une région couvrant quelques dizaines à une centaine de points, juste après la région de Guinier.

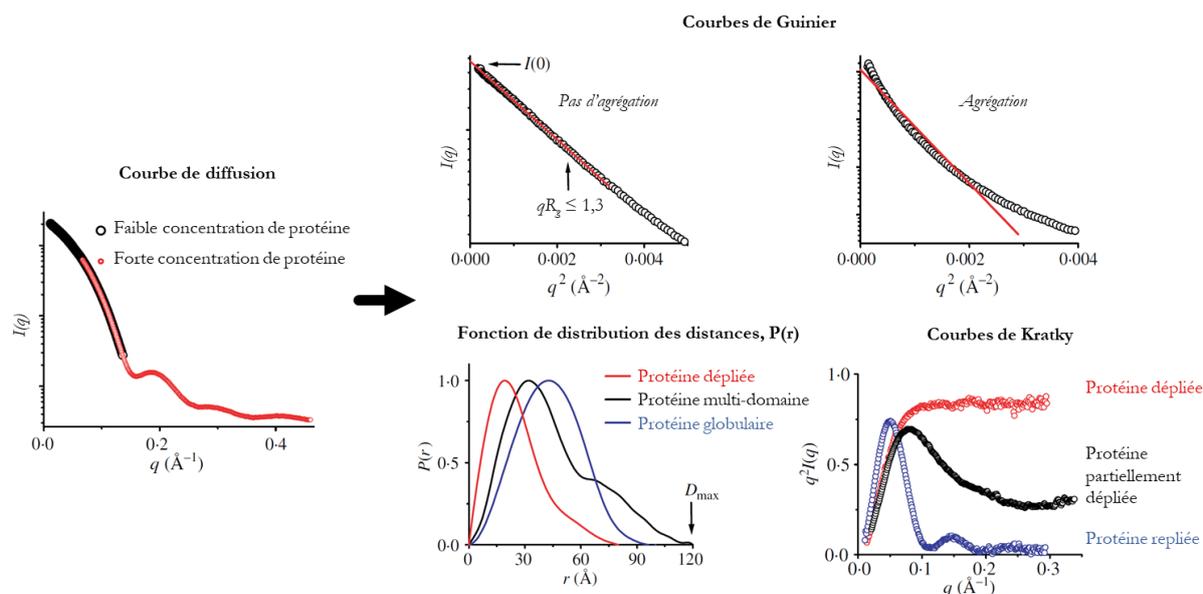


Figure 34 : Courbes théoriques de diffusion et d'analyse SAS. Adapté de Putnam, C. D., et al. (2007)<sup>326</sup>.

Le programme CHROMIXS permet de tracer un ensemble de données SAXS, généré lors d'une mesure SEC-SAXS ou SEC-SANS (chromatographie d'exclusion de taille en ligne SAXS ou SANS), en fonction du temps d'éluion. Cela permet de sélectionner des zones de tampon et échantillons d'intérêt et d'établir une courbe SAS, qui sera ensuite traitée par le logiciel PRIMUS.

Pour la visualisation structurale ainsi que la validité des modèles, les programmes DAMMIF<sup>344</sup> et DAMMIN<sup>345</sup> sont utilisés pour générer des modèles *ab initio* en billes en se basant sur la valeur de  $D_{max}$  calculée par le logiciel GNOM<sup>346</sup>.

Le programme OLIGOMER<sup>343</sup> est utilisé pour déterminer les fractions volumiques de différentes espèces d'un mélange de protéines. Les facteurs de forme de chaque espèce peuvent être générés par fmaker et utilisés par OLIGOMER pour le calcul des proportions.

Etant donné que l'intensité de diffusion observée pendant une expérience de SAS est la moyenne sur toutes les orientations des transformées de Fourier des intensités de longueur de diffusion  $\rho_n$ , ou des densités électroniques  $\rho_e$  des particules en solution, il est possible de calculer les courbes de diffusion théorique si  $\rho_n(\vec{r})$  ou  $\rho_e(\vec{r})$  sont connus. Les programmes CRY SOL<sup>347</sup> et CRYSON<sup>18</sup>, respectivement pour les rayons X et les neutrons, permettent de calculer les courbes de diffusion théorique de protéines à partir des fichiers PDB correspondants. Les courbes ainsi obtenues sont d'une grande aide pour l'interprétation des profils de diffusion expérimentaux, ou encore pour vérifier si la structure d'un fichier PDB correspond à la structure de la protéine en solution.

Le programme SASREF<sup>348</sup> effectue la modélisation de la structure d'un complexe formé de sous-unités dont la structure est connue par rapport à un ensemble de données SAS (courbes SAS des sous-unités isolées et courbe SAS du complexe). Les données SAS des sous-unités doivent d'abord être traitées par le programme CRY SOL ou CRYSON.

## 2. La diffusion de rayons X aux petits angles (SAXS)

Cette partie a été écrite en se basant principalement sur des informations fournies par le site internet de l'ESRF consulté en juillet 2019 ([www.esrf.eu](http://www.esrf.eu)), notamment la page spécifique à l'instrument SAXS BM29.

### 2.1. Histoire du rayonnement synchrotron

Les rayons X ont été découverts en 1895 par Wilhelm Röntgen, qui a montré qu'ils étaient capables de traverser le corps humain. Aujourd'hui, ils sont largement utilisés en imagerie médicale parce qu'ils ont une grande profondeur de pénétration à travers les matériaux et qu'ils sont sélectivement absorbés par les parties du corps ayant la plus forte densité électronique, comme les os. La principale différence entre le rayonnement synchrotron et les rayons X utilisés en imagerie médicale est la brillance. La brillance peut être définie comme le quotient de l'intensité lumineuse diffusée par une surface éclairée, dans une direction déterminée, par l'aire de cette surface projetée sur un plan normal à cette direction. Plus la brillance est élevée, plus la précision de l'image obtenue est grande. Une source synchrotron est cent milliards de fois plus brillante qu'une source de rayons X hospitalière.

Le rayonnement synchrotron a été observé pour la première fois chez General Electric aux États-Unis en 1947. Au départ, les rayons X du synchrotron étaient considérés comme un effet indésirable, parce qu'ils provoquaient une perte d'énergie des particules, mais elle a ensuite été reconnue dans les années 1960 comme une lumière aux propriétés exceptionnelles. A la fin des années 1970, l'intérêt d'utiliser les synchrotrons pour produire des rayons X extrêmement lumineux a été reconnu, aboutissant à la construction de l'ESRF (de l'anglais « *European Synchrotron Radiation Facility* ») au début des années 1990 et, peu après, de deux autres synchrotrons de « troisième génération » (*Advanced Photon Source* aux États-Unis et *SPring-8* au Japon).

### 2.2. Propriétés des rayons X

Les rayons X sont des ondes électromagnétiques constituées de photons tout comme la lumière visible, mais situées à l'extrémité du spectre électromagnétique, entre la lumière ultraviolette et les rayons gamma.

La longueur d'onde des rayons X, d'environ un dixième de nanomètre, est comparable aux distances interatomiques, ce qui rend les rayons X appropriés pour l'étude des atomes et des liaisons. On distingue les rayons X « durs » et rayons X « mous » selon leur énergie. Il est plus courant de parler d'énergie plutôt que de longueur d'onde. Les synchrotrons, dont l'ESRF, produisent des rayons X durs à des longueurs d'onde de 0,1 à 1 Å, c'est à dire à des énergies comprises entre 10 et 120 keV. En raison de leur énergie plus élevée, les rayons X durs pénètrent plus profondément dans la matière que les rayons X mous dont l'énergie est inférieure à 10 keV, qui eux sont facilement absorbés par une épaisseur millimétrique de matière solide. Du fait de l'énergie importante des photons, les rayons X sont des rayonnements dits « ionisants », provoquant des ionisations dans les atomes qu'ils traversent. Ce phénomène est exploité en fluorescence X, tandis qu'il constitue un désavantage pour les autres techniques utilisant les rayons X puisque ces ionisations ont un effet délétère sur les échantillons (appelé dommage de radiation).

## 2.3. Instrumentation à l'échelle d'un synchrotron : l'exemple de l'ESRF

Lorsqu'un électron en mouvement change de direction, il émet de l'énergie sous forme de rayonnement électromagnétique. Lorsque l'électron accélère ou décélère rapidement, l'énergie émise est à la longueur d'onde des rayons X. C'est sur ce principe que repose les synchrotrons. Un synchrotron accélère les électrons jusqu'à une énergie extrêmement élevée pour les faire ensuite régulièrement changer de direction. Les rayons X qui en résultent sont émis sous forme de douzaines de faisceaux minces, chacun dirigé vers une ligne de faisceau près de l'accélérateur.

### 2.3.1. Le canon à électrons, l'accélérateur linéaire et l'anneau d'accélération

Les électrons sont émis par un canon à électrons, un dispositif similaire aux tubes cathodiques que l'on trouve dans les anciens téléviseurs. Ils sont d'abord accélérés dans un accélérateur linéaire (Linac) jusqu'à une énergie de 200 MeV, puis transmis à l'anneau d'accélération de 300 m de circonférence où ils sont accélérés pour atteindre une énergie de 6 GeV. Ces électrons à haute énergie sont ensuite injectés dans un grand anneau de stockage où ils circulent dans un environnement sous vide, à énergie constante.

### 2.3.2. L'anneau de stockage et ses aimants

L'anneau de stockage est un tube de 844 m de circonférence où les électrons tournent à une vitesse proche de celle de la lumière. Lorsque les électrons se déplacent au sein de l'anneau, ils traversent différents types d'aimants et produisent alors des rayons X.

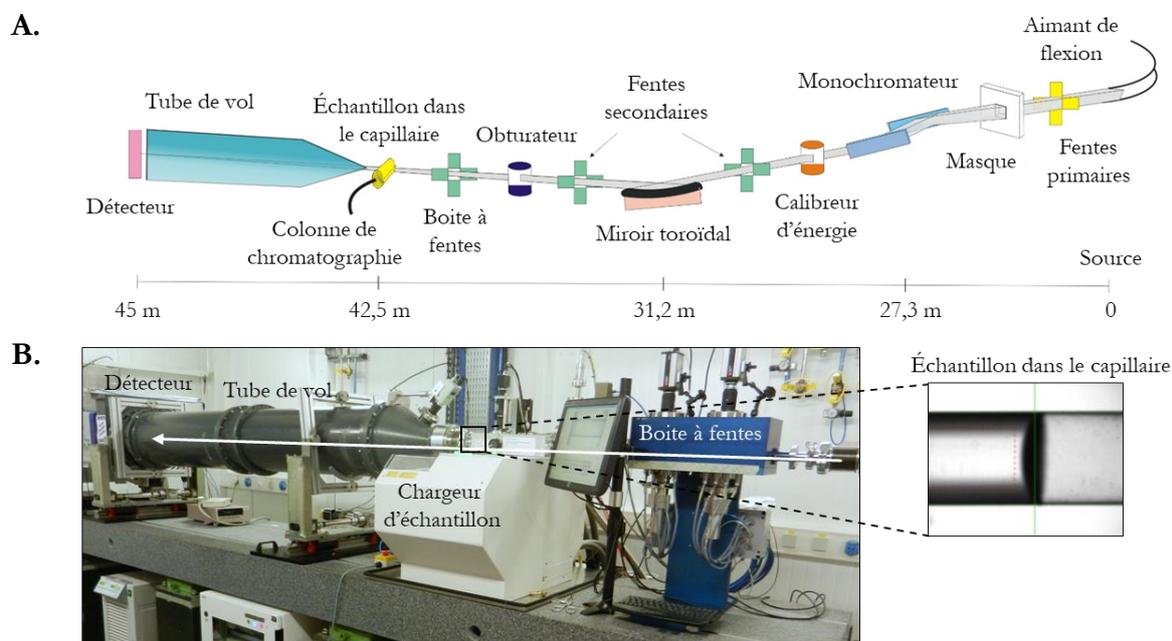
L'anneau de stockage comprend de façon alternative 32 sections droites et 32 courbes. Dans chaque section incurvée, deux grands aimants de flexion forcent le chemin des électrons. Dans chaque section droite, plusieurs aimants de focalisation assurent que les électrons restent près de leur trajectoire orbitale idéale. Les sections droites accueillent également les onduleurs, où sont produits les faisceaux intenses de rayons X.

La fonction principale des aimants de flexion est de dévier les électrons de leur trajectoire. Lorsque les électrons sont déviés de leur trajectoire droite alors qu'ils passent à travers ces aimants, ils émettent un jet de rayons X tangentiellement à la trajectoire initiale. Le rayonnement synchrotron d'un aimant de flexion couvre un spectre large et continu, des micro-ondes aux rayons X durs, et il est beaucoup moins concentré, ou brillant, que le faisceau fin des rayons X d'un onduleur. Les onduleurs sont des structures magnétiques composées d'un réseau complexe de petits aimants obligeant les électrons à suivre une trajectoire ondulée. Le rayonnement émis à chaque pli consécutif se chevauche et interfère avec celui des autres pliages. Cela génère un faisceau de rayonnement beaucoup plus focalisé, ou brillant, que celui généré par un seul aimant.

Les rayons X émis par les électrons sont dirigés vers les lignes de faisceaux qui entourent l'anneau de stockage dans le hall expérimental. Chaque ligne de faisceaux est conçue pour être utilisée avec une technique spécifique ou pour un type de recherche spécifique. Une ligne de faisceaux se compose d'une cabine optique, d'une cabine expérimentale et d'une cabine de contrôle, alignées en ligne.

## 2.4. Instrumentation à l'échelle d'une ligne de faisceau : l'exemple de BM29 à l'ESRF

Une ligne de faisceau pour un instrument SAXS comprend essentiellement les parties suivantes (figure 35), qui sont décrites ici dans l'exemple de la ligne de faisceaux BM29 à l'ESRF.



**Figure 35 : Ligne de faisceau BM29 de l'ESRF.** Schéma d'instrumentation (A) et photo de la zone expérimentale de la ligne BM29 à l'ESRF avec l'installation classique (B) ([www.esrf.eu](http://www.esrf.eu)).

### 2.4.1. La cabine des instruments optiques

La première tâche de la ligne de faisceau est de modifier le faisceau de rayons X brut fourni par un anneau de flexion. Des instruments optiques sont utilisés pour focaliser le faisceau et sélectionner l'énergie souhaitée, soit environ 12,5 keV (0,9919 Å). Comme le faisceau transporte beaucoup d'énergie sur une très petite surface, ces optiques sont conçues pour supporter une grande quantité de chaleur.

Un monochromateur sélectionne une longueur d'onde définie à partir de l'émission d'un anneau de flexion. Le monochromateur à double réflexion produit un faisceau réfléchi monochromatique dont la longueur d'onde est donnée par la loi de Bragg,  $2d\sin\theta = n\lambda$ , où  $d$  est la distance interréticulaire (distance entre deux plans cristallographiques),  $\theta$  le demi-angle de déviation (moitié de l'angle entre le faisceau incident et la direction du détecteur) et  $n$  l'ordre de diffraction (nombre entier).

L'élément de focalisation de la ligne de faisceau est le miroir toroïdal cylindrique qui réfléchit verticalement vers le haut. Ce miroir est situé à 31,2 m de la source et concentre le rayonnement monochromatique sur le détecteur à 13,5 m en aval.

Des fentes primaires délimitent le faisceau blanc à 4 mm par 4 mm pour protéger la mécanique du monochromateur tout en utilisant la taille maximale possible du faisceau à exploiter par l'optique (monochromateur et miroir). Ces fentes sont constituées de matériaux très absorbants (tungstène W ou tantale Ta). Des deux fentes secondaires sont placées après le monochromateur et le miroir

afin d'éliminer du faisceau de toute diffusion parasite, qui est très intense autour du faisceau monochromatique, ce qui est crucial pour atteindre de très petits angles.

Trois axes d'atténuateurs de faisceaux blancs sont situés juste après les fentes primaires. Habituellement, aucun atténuateur de faisceau blanc n'est utilisé. Ces atténuateurs permettent de protéger le monochromateur des basses énergies. Des atténuateurs monochromatiques sont montés aussi loin que possible de la position de l'échantillon pour diminuer la diffusion parasite qui pourrait en résulter. Ils sont introduits automatiquement dans le faisceau si la transmission est inférieure à 100 % (faisceau complet), par exemple en cas de dommages dus au rayonnement.

#### 2.4.2. La cabine expérimentale

La ligne de faisceau permet de placer un échantillon dans le faisceau de rayons X.

Un obturateur est utilisé avec un axe de rotation réglé verticalement et avec une ouverture totale de 5 mm. Deux capteurs lisent la position de la pièce en rotation et donnent un signal de retour qui permet de mesurer le temps d'exposition de l'échantillon. L'obturateur est situé dans le vide au début de la zone expérimentale aussi loin que possible de la position de l'échantillon afin de réduire toute diffusion parasite de celui-ci sur le détecteur.

Une boîte à fentes se trouve au début de la table expérimentale juste après l'obturateur. La combinaison de faisceaux incidents très intenses et les imperfections des fentes utilisées en amont conduiraient à une diffusion parasite importante. Cette diffusion parasite serait concentrée autour du faisceau direct et constituerait un sérieux obstacle à l'obtention de données de diffusion de bonne qualité à faible  $q$  pour les échantillons faiblement diffusants. Afin de réduire cette diffusion indésirable, une des fentes supplémentaires et un trou d'épingle est inclus dans l'installation. Deux premières fentes définissent le faisceau et une troisième fente, la fente de garde, est alignée le plus près possible du faisceau direct intense sans le toucher, car cela provoquerait à nouveau une diffusion parasite. L'échantillon est placé juste après la fente de protection. Le signal de diffusion peut alors être enregistré de manière fiable.

Un chargeur d'échantillons automatique permet la manipulation de liquides. Les échantillons dans le passeur d'échantillons peuvent être stockés dans une plaque de 96 puits et/ou dans des tubes. Les cellules de stockage et d'exposition peuvent être thermorégulées indépendamment l'une de l'autre, de 4 à 40 °C et de 4 à 60 °C, respectivement. Le nettoyage et le chargement de l'échantillon prennent environ 30 secondes chacun. Un système de chromatographie d'exclusion stérique *in-situ* est également disponible. Un porte-échantillon, généralement un capillaire en quartz thermostaté, contient alors la solution à étudier. La taille du faisceau à l'échantillon est d'environ 700  $\mu\text{m}$  par 700  $\mu\text{m}$ .

Le long trajet du faisceau entre l'échantillon dans le capillaire et le détecteur se produit sous vide grâce au tube de vol. Un dispositif d'arrêt de faisceau est situé à la fin du tube. Ce dispositif en matériau absorbant le faisceau direct l'empêche de frapper le détecteur et de l'endommager de part de son intensité. Il se compose d'un cylindre creux en tungstène de 3 mm de diamètre.

Un détecteur de rayons X (PILATUS 1M) positionné immédiatement après le tube de vol enregistre l'intensité de la diffusion en fonction de l'angle de diffusion  $2\theta$ . La distance entre l'échantillon et le détecteur est fixée à 2,867 m. L'instrument couvre une plage  $q$  de 0,025 à 5  $\text{nm}^{-1}$ . Pour autant que la diffusion de l'échantillon soit isotrope (indépendante de la direction de  $q$ ), l'intensité est obtenue

en faisant la somme de tous les éléments détectés au même rayon  $r$  du centre du faisceau. Connaissant la distance entre l'échantillon et le détecteur et la longueur d'onde, la courbe unidimensionnelle  $I(r)$  contre  $r$  est ensuite convertie en  $I(q)$  contre  $q$  (équation 2).

### 3. La diffusion de neutrons aux petits angles (SANS)

Cette partie a été écrite en se basant principalement sur des informations fournies par le livre *Experimental neutron scattering*<sup>337</sup> et par le site internet de l'ILL consulté en juillet 2019 ([www.ill.eu](http://www.ill.eu)), notamment les pages spécifiques aux instruments SANS D11 et D22.

#### 3.1. Histoire des sources de neutrons

Alors que les premières expériences de diffusion étaient menées à l'aide de neutrons provenant de sources radioactives, la fission dans un réacteur de recherche nucléaire constitue aujourd'hui la source la plus courante de neutrons. Le travail mené pendant la Seconde Guerre mondiale sur l'isolation des matières fissiles des minéraux bruts et des éléments combustibles usés a permis de déboucher sur des programmes nucléaires permettant d'avoir aujourd'hui de puissants réacteurs de recherche. La spallation des atomes lourds s'effectue quant à elle à l'aide de sources basées sur des accélérateurs, dont la construction a commencé près de quarante ans après l'installation des premiers instruments de diffusion des neutrons dans réacteurs de recherche. Les réacteurs s'approchent maintenant des flux neutroniques maximaux technologiquement réalisables, et les sources d'accélérateurs pulsés n'ont pas encore atteint cette limite et offrent la méthode la plus prometteuse pour réaliser des flux encore plus élevés.

#### 3.2. Propriétés des neutrons

La quasi-totalité de la matière qui nous entoure est constituée de neutrons liés. Le neutron libre est une particule instable, subissant une désintégration à trois corps en un proton, un électron et un anti-neutrino, avec une demi-vie d'environ 15 minutes. Le neutron peut être représenté sous la forme d'une onde ou d'une particule. Par conséquent, les études de diffraction des neutrons ou des excitations collectives dans les cristaux sont interprétées en supposant que le neutron est une onde. Lorsqu'il s'agit de la production et de la détection des neutrons, il est nécessaire de considérer le neutron comme une particule.

L'énergie des neutrons, aussi appelée température neutronique, indique l'énergie cinétique d'un neutron libre, habituellement donnée en eV. Tout comme les rayons X, il est plus courant de parler d'énergie plutôt que de longueur d'onde. La distribution de l'énergie des neutrons est ensuite adoptée à la distribution de Maxwell-Boltzmann (décrite dans la partie 3.3) connue pour l'agitation thermique (figure 37). Plus la température est élevée, plus l'énergie cinétique du neutron libre est grande. La classification de l'énergie des neutrons est décrite de la façon suivante :

- Neutrons rapides : énergie supérieure à 1 eV, 0,1 MeV.
- Neutrons épithermiques : énergie comprise entre 1 eV et 10 keV.
- Neutrons lents : énergie inférieure ou égale à 0,4 eV.
- Neutrons chauds : énergie d'environ 0,2 eV.
- Neutrons thermiques : énergie d'environ 0,025 eV.
- Neutrons froids : énergie comprise entre  $5 \times 10^{-5}$  eV et 0.025 eV.
- Neutrons très froids : énergie comprise entre  $3 \times 10^{-7}$  eV et  $5 \times 10^{-5}$  eV.

- Neutrons ultra-froids : énergie de l'ordre de  $10^{-7}$  eV.

Les neutrons ont un certain nombre de propriétés qui les distinguent des rayons X. Les neutrons offrent un choix de longueur d'onde d'inférieure à 1 Å jusqu'à supérieure à 15 Å, avec des énergies faibles et une absorption négligeable dans la plupart des matières, entraînant une pénétration élevée et des dommages de radiation négligeables. De plus, le rayonnement neutronique étant très pénétrant, il permet d'examiner les propriétés d'échantillons de matière dure (métaux, minerais, etc.) de plusieurs centimètres d'épaisseur et de sonder des échantillons enfermés dans des fours, cryostats ou cellules de pression.

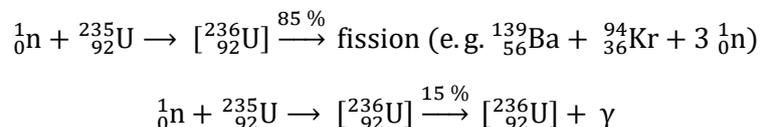
Un avantage significatif des neutrons est qu'ils sont diffusés par les noyaux atomiques, ce qui donne des amplitudes indépendantes de la masse atomique, contrairement aux rayons X qui sont diffusés par les électrons. Cependant, les neutrons sont diffusés différemment par les isotopes d'un même élément. Les effets isotopiques, en particulier la grande différence entre l'amplitude de diffusion de l'hydrogène et du deutérium (voir partie suivante), permettent de marquer différentes parties d'une structure et fournissent des applications puissantes. Cette deutération sélective, couplée à la technique de variation de contraste, est en grande partie responsable du succès du SANS, en particulier dans l'étude des systèmes biologiques.

Les neutrons sont produits par diverses réactions nucléaires, dont la fission, la photofission, la spallation et la fusion. Pour les expériences de diffusion, les plus importantes sont la fission et la spallation.

### 3.3. Instrumentation à l'échelle d'une source : l'exemple de l'ILL

L'ILL est à ce jour la source de neutrons la plus puissante au monde pour les expériences de diffusion. Le flux thermique au niveau des tubes de faisceaux est de  $1,5 \times 10^{15} \text{ cm}^{-2} \cdot \text{sec}^{-1}$ , fournissant des faisceaux de neutrons aux instruments dans la salle principale du réacteur et dans les deux salles de guidage des neutrons.

$^{235}\text{U}$  a été le premier isotope découvert comme fissile, c'est à dire capable de subir une réaction de fission après absorption d'un neutron thermique. Généralement, lorsque le noyau de  $^{235}\text{U}$  subit une fission, le noyau se divise en deux noyaux plus petits, avec des neutrons thermiques et la de la libération d'énergie sous forme de chaleur et de rayons gamma selon les réactions suivantes :



C'est sur ce principe que reposent les réacteurs nucléaires. Les réacteurs nucléaires se composent de quatre composants de base : le combustible, le modérateur, le liquide de refroidissement et le blindage. Le combustible contient la matière fissile, qui peut se présenter sous sa forme naturelle (tel que  $^{238}\text{U}$ ) ou être enrichie isotopiquement dans un isotope fissile (tel que  $^{235}\text{U}$ ) ou sous forme de composé (tel que le dioxyde d'uranium). Le cœur du réacteur du Réacteur à Haut Flux (RHF) de l'ILL est un élément combustible unique d'uranium de 10 kg, enrichi à 93 % en  $^{235}\text{U}$ . La fission thermique de  $^{235}\text{U}$  se produit avec la production, en moyenne, de 2,7 neutrons rapides et la désintégration simultanée du noyau de l'uranium en deux fragments de fission de masse inégale. Ces fragments sont principalement riches en neutrons et sont hautement radioactifs : ils se

désintègrent par l'intermédiaire d'une cascade d'émissions bêta (c'est-à-dire l'expulsion d'électrons) jusqu'à leur stabilisation. La réaction de fission devient auto-entretenu à condition qu'à partir de ce rendement de 3 neutrons, un neutron provoque une fission supplémentaire dans un autre noyau  $^{235}\text{U}$ . Les 2 neutrons restants sont absorbés ou s'échappent de la surface du réacteur. Ce sont ces derniers neutrons qui forment la source de neutrons lents pour les expériences de diffusion. Le réacteur (figure 36) est contrôlé au moyen d'une barre absorbant les neutrons, faite d'un alliage argent-indium-cadmium, située au sein d'une cavité cylindrique centrale dans l'élément combustible et qui est progressivement retirée du cœur au fur et à mesure que l'uranium est consommé. Il est également doté de 5 barres de sécurité, qui sont également des dispositifs d'absorption des neutrons et servent à arrêter le réacteur en cas d'urgence.

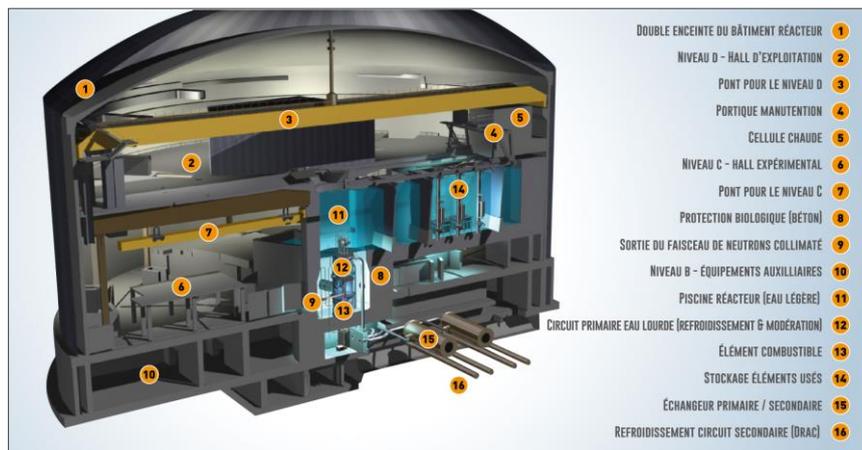
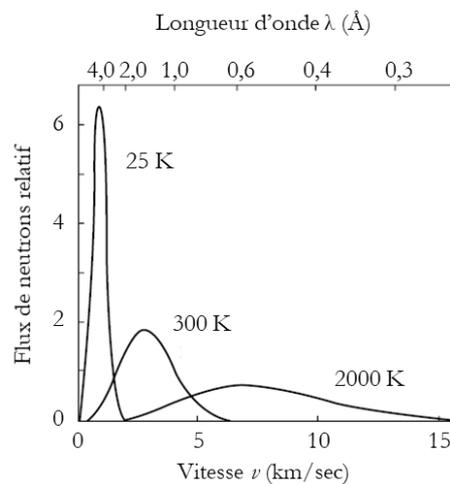


Figure 36 : Schéma du réacteur de l'ILL. Les différentes parties du réacteur sont numérotées de 1 à 16 ([www.ill.eu/](http://www.ill.eu/)).

Le liquide de refroidissement est généralement de l'eau légère bien que l'on utilise parfois un gaz, comme le dioxyde de carbone ou l'hélium. Dans le cas de l'ILL, de l'eau légère est pompée à travers les ailettes internes du cœur à une vitesse de 15,5 m/sec pour éliminer les 58 MW de chaleur générés par le processus de fission. Le modérateur est un matériau de faible masse atomique, comme l'eau légère, l'eau lourde et ou le graphite. Dans le cas de l'ILL, l'eau légère du liquide de refroidissement agit comme modérateur principal, mais un modérateur à eau lourde encercle le cœur, et sert également à réfléchir les neutrons dans le cœur pour modérer davantage et poursuivre la réaction en chaîne. En effet, le but du modérateur est de ralentir les neutrons de leur énergie initiale MeV vers les énergies meV où la probabilité de fission est beaucoup plus élevée. En raison de la présence du modérateur, le cœur d'un réacteur à fission thermique peut avoir plusieurs mètres de diamètre, alors que le cœur d'un réacteur à neutrons rapides peut mesurer seulement 20 cm.

L'émission bêta intense de fragments de fission a deux importantes conséquences. La première est que les éléments combustibles des réacteurs nucléaires deviennent rapidement hautement radioactifs et un bouclier épais autour du noyau est nécessaire, non seulement pour la sécurité du personnel, mais aussi pour réduire l'impact sur l'environnement. La deuxième conséquence est que de nombreux fragments de fission et leurs produits de désintégration sont très absorbants pour les neutrons et peuvent arrêter la réaction en chaîne. Le réservoir d'eau lourde contenant le noyau est situé dans une piscine remplie d'eau déminéralisée qui assure un blindage contre les rayonnements neutroniques et gamma produits par le noyau.

Le spectre énergétique des neutrons émis par un réacteur dépend de la température des modérateurs entourant la source, car le « gaz » neutronique est en équilibre thermique avec les matériaux des modérateurs. La figure 37 montre le flux pour des températures de modérateur de 25, 300 et 2 000 K. La plupart des réacteurs, dont celui de l'ILL, sont conçus pour une température de modérateur d'environ 300 K, correspondant à la courbe centrale de la figure 37. Trois composants situés à proximité immédiate du cœur permettent également de produire des neutrons chauds, froids et ultra-froids de respectivement 10 km/s, 700 m/s et 10 m/s respectivement). Ces composants sont une source chaude, comprenant une sphère de graphite maintenue à une température de 2400 K, et deux sources froides, dont la plus grande est une sphère contenant 20 litres de deutérium maintenu à l'état liquide à 25 K, dans laquelle les neutrons sont ralentis à l'énergie souhaitée par une succession de collisions avec les atomes de deutérium. Pour certaines expériences, telles que la diffusion aux petits angles, l'utilisation d'une source froide permet d'augmenter le flux à des longueurs d'onde supérieures à 3 Å. Dans d'autres cas, l'utilisation une source chaude qui diminue le flux à des longueurs d'onde inférieures à 0,8 Å (figure 37).



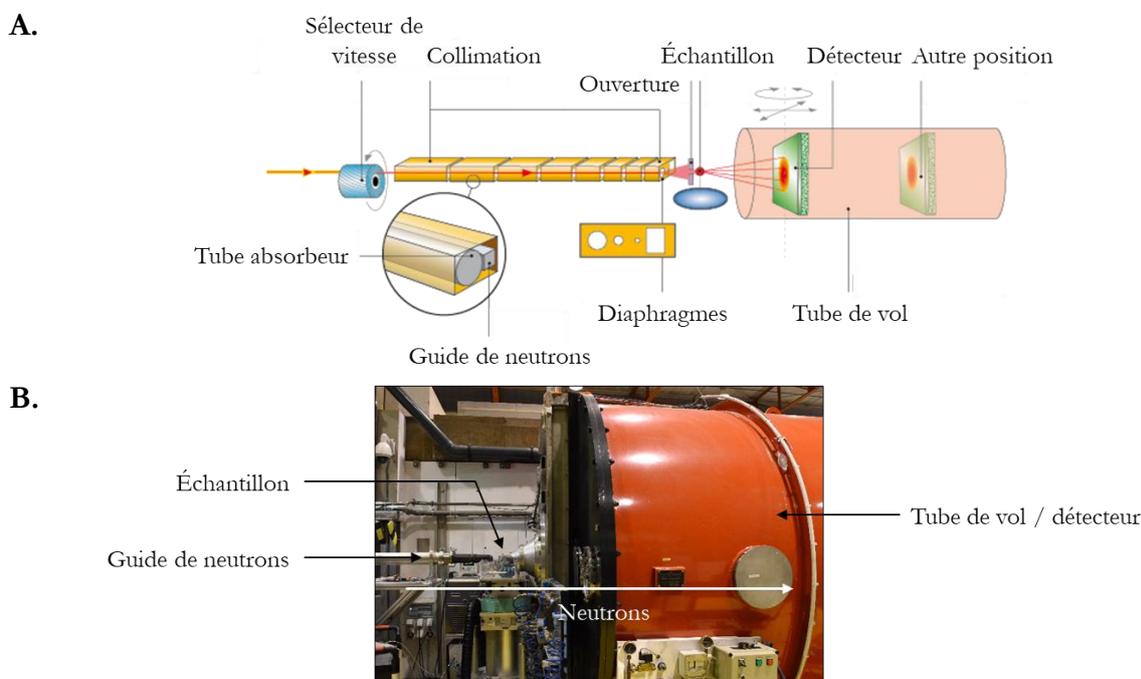
**Figure 37 : Distribution de flux Maxwell-Boltzmann pour les modérateurs à trois températures différentes.** Adapté de Carlile, C. J. (2013) <sup>337</sup>.

Les neutrons sont extraits de l'intérieur de la cuve du réacteur par une vingtaine de tubes à rayons, dont certains sont dirigés vers la source chaude ou une des sources froides. La plupart des tubes ne sont pas orientés radialement vers le cœur du réacteur mais sont tangents à celui-ci. Cela permet de réduire au minimum les effets de bruit de fond des flux élevés de rayons gamma générés à la fois par la réaction de fission et par les fragments de fission en décomposition. Ces tubes, qui s'étendent à l'extérieur de la piscine du réacteur sous forme de guides neutroniques, alimentent ensuite en neutrons une quarantaine de zones expérimentales, situées jusqu'à 100 m du réacteur dans deux halls expérimentaux. Les tubes ou guides conducteurs de neutrons permettent aux instruments de mesure d'être situés bien loin de la source elle-même et dans des régions plus spacieuses à faible bruit de fond, et permettent d'installer un grand nombre d'instruments. Ces guides, généralement en verre et recouverts de métaux tels que des couches de nickel et titane, exploitent le fait que l'indice de réfraction  $n$  des neutrons est inférieur à l'unité pour la plupart des matériaux, de sorte que le phénomène de réflexion externe totale peut se produire à la limite entre le matériau et un vide. Le faisceau de neutrons est alors transporté loin de la source de neutrons

sans perte d'intensité, et lorsque le tube de guidage est légèrement courbé en arc, il agit comme un filtre discriminant contre les neutrons de plus grande énergie et les rayons gamma.

### 3.4. Instrumentation à l'échelle d'une ligne de faisceau : l'exemple de D22 à l'ILL

Une ligne de faisceau pour un instrument SANS comprend essentiellement les parties suivantes (figure 38), qui sont décrites ici dans l'exemple de la ligne de faisceaux D22 à l'ILL.



**Figure 38 : Ligne de faisceau D22 de l'ILL.** Schéma du diffractomètre petits angles D22 ([www.ill.eu](http://www.ill.eu)) (A) et photos de la ligne D22 (B) à l'ILL.

Les neutrons de la source froide verticale de l'ILL passent par un guide de neutrons avant d'atteindre la ligne D22. Une bande de longueur d'onde relativement étroite est alors choisie par un sélecteur de vitesse parmi les neutrons. Les neutrons qui frappent cet appareil ont une très large distribution de vitesses (ou longueurs d'onde) et sont considérés comme un « faisceau blanc ». Le sélecteur de vitesse agit alors comme un monochromateur. Il est similaire à une turbine, avec un tambour rotatif de 25 cm de long où des pales sont fixées à une distance constante l'une de l'autre et disposées en hélice selon un axe de rotation. Le nombre de pales et leur espacement déterminent l'étalement des vitesses (longueurs d'onde) autour de la vitesse moyenne atteinte. Cet « écart de longueur d'onde » est généralement d'environ 10%. Seule la fraction des neutrons du « faisceau blanc » ayant la vitesse souhaitée peut passer à travers le dispositif pour une vitesse de rotation donnée du sélecteur, alors que les neutrons plus rapides ou plus lents sont absorbés sur les surfaces des pales généralement recouvertes de bore  $^{10}\text{B}$  lors de leur passage dans le sélecteur. Les longueurs d'onde techniquement possibles sur la ligne de faisceau D22, déterminées par la vitesse de rotation du tambour, peuvent se situer entre 4,5 et 40 Å.

La distance entre la source et l'échantillon est choisie par un système de collimation composé de huit sections de guides en verre mobiles. Le rôle des collimateurs est de transporter efficacement les neutrons à l'échantillon en garantissant l'efficacité de transport du flux neutronique maximal. En règle générale, la longueur du système de collimation est identique à la distance entre

l'échantillon et le détecteur. Cela peut être mieux compris si l'on considère un instrument SANS comme un appareil photographique à sténopé : pour obtenir une image nette, l'objet doit être au moins à la même distance de l'écran de détection que de la source de radiation.

Des diaphragmes sont alors utilisés pour donner une forme au faisceau. Ce sont de simples ouvertures de différentes tailles dans un disque généralement constitué de cadmium Cd, comme des fenêtres, positionnées à des points spécifiques à travers la collimation entre le sélecteur de vitesse et l'échantillon.

L'instrument D22 possède une gamme d'équipements afin d'adapter l'environnement expérimental selon les besoins des utilisateurs. Pour les mesures en conditions statiques par exemple, de nombreux supports thermorégulés sont disponibles, ainsi que des fours, pompes à vide et dispositifs de pression. Un système de chromatographie d'exclusion stérique in-situ (SEC-SANS) est également disponible<sup>349</sup>. L'absorbance, la fluorescence, ou la diffusion de la lumière peut également être mesurée simultanément sur les échantillons. Un instrument SAXS est également en cours d'installation. La partie 2.1 du chapitre 4 décrit l'instrumentation pour les mesures effectuées lors de cette thèse.

Le neutron étant une sonde très faible en énergie, la diffusion des neutrons ne perturbe pas les propriétés de l'échantillon. Cependant, la faiblesse de son interaction avec la matière rend également les neutrons difficiles à détecter directement. Les neutrons thermiques produisent une ionisation négligeable et ne peuvent être détectés que par la production d'une ionisation secondaire après absorption des neutrons. Le processus d'absorption conduit à la formation d'un noyau composé, qui se désintègre soit par l'émission d'un rayon gamma, soit par une division du noyau composé en deux noyaux chargés. Un type de détecteur le plus courant utilise comme matériau absorbant un gaz, tel que  $^3\text{He}$ . C'est le cas de l'instrument D22. La réaction de capture nucléaire se déroule de la façon suivante :

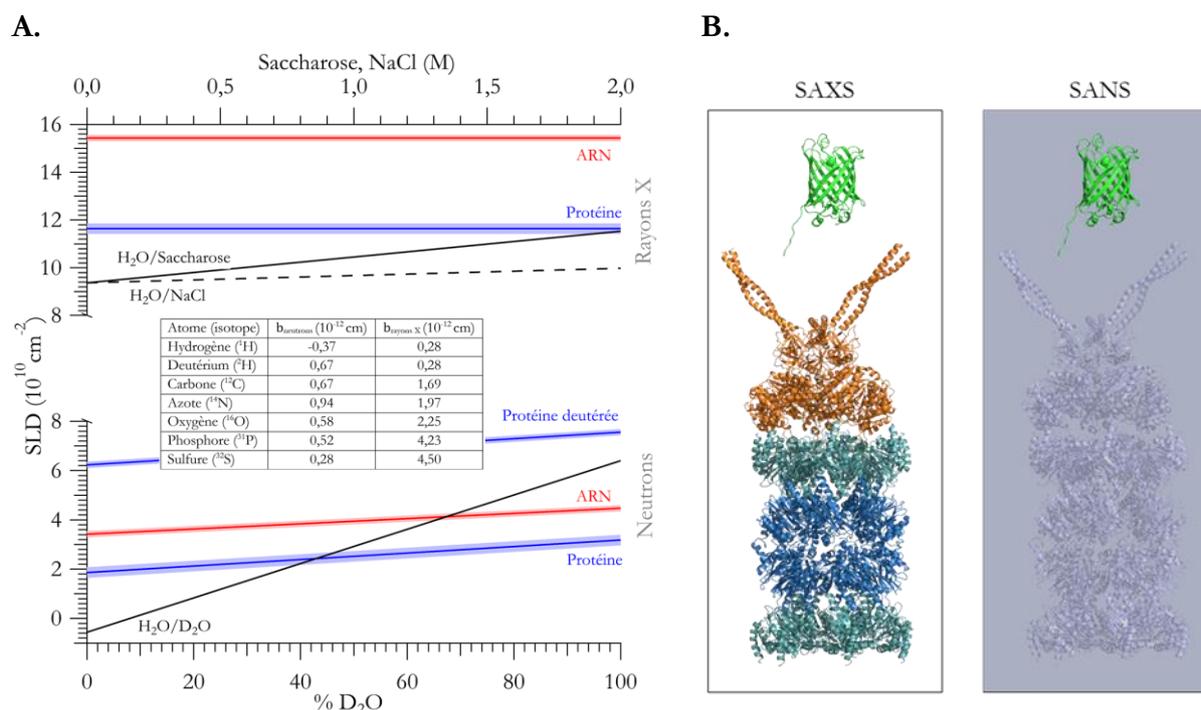


Un détecteur de gaz est généralement construit avec le gaz sensible aux neutrons enfermé dans un tube métallique cylindrique qui forme la cathode du détecteur. L'anode est un fin fil central tendu le long de l'axe du tube et est isolée de la cathode par des joints en céramique. Dans un détecteur  $^3\text{He}$ , le gaz  $^3\text{He}$  est pressurisé pour augmenter l'efficacité de détection. Un gaz additif, tel que le krypton ou le méthane, est inclus pour produire plus d'ionisation et aussi pour refroidir la décharge dans le plus court laps de temps possible. La réaction nucléaire dans le gaz libère de l'énergie exothermiquement, et les produits de la réaction se séparent dans des directions opposées. Les particules produites provoquent une ionisation dans le gaz et cette ionisation est collectée par un fil anodique à haute tension. La ligne D22 possède un détecteur d'une surface active de  $1 \text{ m}^2$ . Il se déplace à l'intérieur d'un tube à vide de 2,5 m de large et de 20 m de long fournissant des distances échantillon-détecteur de 1,1 m à 17,6 m. Afin d'éviter d'endommager le détecteur, la partie la plus intense du faisceau de neutrons (partie du faisceau qui n'est pas diffusée par l'échantillon) est interceptée par un arrêt de faisceau. Ce dispositif est composé de matériau absorbant les neutrons, souvent constitué de cadmium Cd, et est positionné au centre du tube en ligne avec le détecteur. L'instrument couvre une plage  $q$  totale de  $4 \times 10^{-4}$  à  $0,44 \text{ \AA}^{-1}$  sans décalage du détecteur ou  $0,85 \text{ \AA}^{-1}$  avec décalage du détecteur. Pour autant que la diffusion de l'échantillon soit isotrope, l'intensité est obtenue en faisant la somme de tous les éléments détectés au même rayon  $r$  du centre du

faisceau. Connaissant la distance entre l'échantillon et le détecteur et la longueur d'onde, la courbe unidimensionnelle  $I(r)$  en fonction de  $r$  est ensuite convertie en  $I(q)$  en fonction de  $q$  (équation 2).

### 3.5. La variation de contraste

Il existe une grande différence entre la longueur de diffusion  $b$  du deutérium ( $b_D = 0,667 \times 10^{-12}$  cm) et celle de l'hydrogène ( $b_H = -0,374 \times 10^{-12}$  cm), la dernière valeur étant négative alors celles des autres atomes des macromolécules biologiques sont positives. Cette différence de longueur de diffusion est utilisée en SANS lors de la préparation des mélanges de solvants isotopiques comme le mélange  $H_2O-D_2O$  et lors de la deutération des biomolécules. En considérant les valeurs de densité de longueur de diffusion pour l'eau, les contributions dues aux petits solutés tels les ions de sel sont négligeables pour le SANS à des concentrations modérées<sup>325</sup>.



**Figure 39 : La variation de contraste.** **A.** Densités de longueur de diffusion de diverses molécules biologiques en fonction de la concentration de  $D_2O$  (pour les neutrons) et de saccharose/NaCl (pour les rayons X) dans le solvant. L'ADN (non représenté) présente des valeurs très similaires à celle de l'ARN. L'ARN et l'ADN perdeutérés (non représentés) ont des valeurs très similaires à celles des protéines perdeutérées. Encart : longueurs de diffusion des neutrons et des rayons X des atomes et des isotopes les plus courants présents dans les molécules biologiques. Adapté de Mahieu, E., & Gabel, F. (2018)<sup>37</sup>. **B.** Schéma représentatif du contraste des protéines étudiées en SAXS et en SANS dans un tampon à 42%  $D_2O$  dans le solvant où la GFP<sub>ssrA</sub> est deutérée et le protéasome PAN-20S est hydrogéné.

Il existe une variation de la longueur de diffusion entre les différents atomes constituant les macromolécules biologiques. De ce fait, les grandes classes de molécules biologiques ont des densités de longueur de diffusions (SLD de l'anglais « *Scattering Length Density* ») différentes (équation 11). Les acides nucléiques, ADN et ARN, ont une SLD plus grande que celle des protéines, qui ont elles-mêmes une SLD plus grande que celle des lipides<sup>329,350</sup>. Ceci est représenté graphiquement sur la figure 39, on peut y voir les SLD de l'ARN, des protéines et des lipides, représentés en fonction de la concentration en  $D_2O$ . Les points où la droite de l'eau et celles des différentes molécules se croisent sont appelés points de correspondance, ils correspondent aux pourcentages de  $D_2O$  nécessaires pour atteindre un contraste nul pour ces molécules. On remarque

qu'à 0% D<sub>2</sub>O, la densité de longueurs de diffusion des protéines est supérieure à celle de l'eau, tandis qu'à environ 42 % v/v D<sub>2</sub>O, le point de correspondance est atteint et le rapport s'inverse.

L'avantage qu'en retire la diffusion de neutrons est la possibilité de distinguer la diffusion des différents constituants d'un système complexe, à condition que ces constituants aient des densités de diffusion de neutrons différentes. Dans le cas d'un complexe composé de plusieurs protéines, la longueur de diffusion d'un des partenaires peut être modifiée sélectivement par le procédé de deutération. La substitution d'un noyau par l'un de ses isotopes n'introduit pas une perturbation significative des propriétés chimiques du matériel, mais peut conduire à une forte variation de la longueur de diffusion. Les protéines contiennent une grande proportion d'hydrogènes qui peuvent, dans la plupart des cas, être facilement remplacés par l'isotope de deutérium. En faisant ensuite varier la densité de longueurs de diffusion du solvant par mélange H<sub>2</sub>O-D<sub>2</sub>O, la diffusion d'un partenaire spécifique est mesurée de façon sélective. A 0% D<sub>2</sub>O, le partenaire hydrogéné et le partenaire deutéré ont une densité de longueur de diffusion différente de celle du solvant. Cependant, à environ 42% v/v D<sub>2</sub>O, la densité de longueur de diffusion du partenaire hydrogéné est identique à celle de l'eau et seul le partenaire deutéré contribue au signal<sup>329,351</sup>.

### 3.6. Applications du SANS en biologie

La revue « *Biological small angle neutron scattering: recent results and developments* » a été publiée dans le journal « *Acta Crystallographica Section D* » en 2018<sup>37</sup> dans le cadre de cette thèse et résume les récentes avancées de la diffusion des neutrons aux petit angles (SANS) appliquée en biologie.

Au cours de ces dernières années, le SANS a été de plus en plus utilisé par la communauté de la biologie structurale pour obtenir des informations à basse résolution sur les complexes macromoléculaires biologiques en solution. En combinaison avec le marquage au deutérium et la variation du contraste des solvants (échange H<sub>2</sub>O/D<sub>2</sub>O), le SANS fournit des informations sur les composants individuels de grands complexes hétérogènes, qui sont complémentaires aux informations fournies par la cristallographie, la résonance magnétique nucléaire et la microscopie électronique. Les progrès et les développements des applications biologiques des SANS au cours de la dernière décennie sont passés en revue dans cet article. La revue couvre les résultats scientifiques de systèmes biologiques, tels que les complexes protéine-protéine, les complexes protéine-ARN/ADN-protéine et les protéines membranaires. En outre, une vue d'ensemble des développements récents en matière d'instruments, d'environnement d'échantillonnage, de marquage au deutérium et de logiciels est présentée. Enfin, des perspectives du SANS appliqué en biologie sont présentées dans le contexte des approches de biologie structurale intégrative.

# Chapitre 3 :

## Matériel et méthodes biochimiques

### 1. Expression des protéines d'intérêt

Une protéine peut être obtenue soit à partir de son organisme d'origine, soit par l'expression de son gène dans un autre organisme, habituellement la bactérie *Escherichia coli* (*E. coli*). Pour les protéines étudiées, la seconde méthode est la plus avantageuse en termes de facilité de culture et de rendement. En effet, *E. coli* est un organisme bien caractérisé avec de nombreuses lignées cellulaires, vecteurs plasmidiques et outils de biologie moléculaire disponibles. Ceci permet de plus une production élevée de protéines hétérologues<sup>11</sup>. Pour obtenir une protéine à partir d'*E. coli*, l'ADN codant pour la protéine d'intérêt est inséré dans un vecteur plasmidique, qui est à son tour introduit dans l'organisme hôte. La protéine d'intérêt, dite « recombinante », sera alors exprimée en utilisant la machinerie transcriptionnelle et traductionnelle de l'organisme hôte. La protéine recombinante sera séparée des protéines de l'organisme par différentes étapes de purifications. La surexpression des protéines recombinantes facilite beaucoup la production de solutions protéiques pures nécessaires pour toute étude fonctionnelle, biophysique ou structurale.

L'optimisation des protocoles d'expression et de purification des différentes protéines étudiées a constitué une partie importante de ce projet de thèse. Les protocoles décrits ici sont ceux retenus suite à cette optimisation. Sauf mention du contraire, j'ai purifié toutes les protéines présentées dans cette partie.

#### 1.1. Gènes d'intérêts, provenance des plasmides et transformation

Les séquences d'ADN codant pour les protéines d'intérêts provenant de *Methanocaldococcus jannaschii* (*M. jannaschii*) (PAN et sous-unités  $\alpha$  et  $\beta\Delta 6$  de la particule 20S) ou de *Aequorea victoria* (*A. victoria*) (GFPssrA, GFPssrA\_G10C et GFPssrA\_H231C) ont été optimisées par la société *GeneCust Europe* pour éliminer la différence d'usage des codons entre l'organisme d'origine et *E. coli*. Les gènes ont ensuite été synthétisés chimiquement et insérés chacun entre deux sites de restriction par *GeneCust Europe* dans un plasmide (pET-30a ou pETDuet-1) commandé à la société *Novagen*. Les gènes codant les sous-unités  $\alpha$  et  $\beta\Delta 6$  de la particule 20S ont été insérés dans le même plasmide (pETDuet-1).

Les gènes en provenance de *Streptomyces avidinii* (*S. avidinii*) codant pour la version sauvage et triple mutante de la streptavidine ont été commandés sur Addgene ([www.addgene.org](http://www.addgene.org)) sous forme optimisée pour l'expression des codons rares chez *E. coli* et insérés dans le plasmide pET21a.

Les vecteurs plasmidiques utilisés contiennent tous un promoteur fort T7 (inhibé par le répresseur LacI) en amont de leur site de multi-clonage. Ils possèdent également un site de fixation des ribosomes ainsi qu'un gène conférant la résistance à un antibiotique et un autre codant le répresseur LacI. L'ajout d'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) dans le milieu de culture permet de lever l'inhibition par LacI et de transcrire les gènes sous le contrôle du promoteur T7. On dit que le promoteur T7 est « inductible à l'IPTG ».

Le tableau 4 récapitule les informations sur les gènes des protéines d'intérêt et leurs vecteurs.

**Tableau 4 : Tableau récapitulatif des informations concernant les gènes des protéines d'intérêt et leurs vecteurs.**

	PAN	Particule 20S		GFP <sub>ssrA</sub>	GFP <sub>ssrA_G10C</sub>	GFP <sub>ssrA_H231C</sub>	Streptavidine sauvage	Streptavidine mutante
		Sous-unité $\alpha$	Sous-unité $\beta\Delta 6$					
<b>Organisme de provenance</b>	<i>M. jannaschii</i>	<i>M. jannaschii</i>		<i>A. victoria</i>	<i>A. victoria</i>	<i>A. victoria</i>	<i>S. avidinii</i>	<i>S. avidinii</i>
<b>Mutations et étiquettes</b>	étiquette poly-histidine en C-ter	étiquette poly-histidine en N-ter	$\Delta 6$ (délétée de 6 acides aminés N-ter)	A206K, étiquette ssrA en C-ter	A206K, G10C, étiquette ssrA en C-ter	A206K, H231C, étiquette ssrA en C-ter	étiquette poly-histidine en C-ter	N23A, S27D, S45A
<b>Vecteur</b>	pET-30a	pET-Duet-1		pET-30a	pET-30a	pET-30a	pET-21a	pET-21a
<b>Enzymes de restriction</b>	NdeI / XhoI	NcoI / AflIII	NdeI/XhoI	NdeI / NotI	NdeI / NotI	NdeI / NotI	NdeI / HindIII	NdeI / HindIII
<b>Résistance</b>	Kanamycine	Ampicilline		Kanamycine	Kanamycine	Kanamycine	Ampicilline	Ampicilline

Les plasmides contenant le gène des protéines d'intérêt ont été insérés séparément dans la souche d'*E. coli* BL21(DE3) compétente (*Agilent Technologies*). Ce système permet la surproduction de protéines dont les gènes sont sous le contrôle du promoteur du phage T7<sup>352</sup>. L'insert chromosomique DE3 est issu du phage T7 et code pour l'ARN polymérase T7. Il est lui-même sous le contrôle du promoteur *lacUV5*, inductible à l'IPTG.

Pour chaque protéine d'intérêt, 1  $\mu$ L de plasmides (100 ng/ $\mu$ L) contenant le gène d'expression de la protéine a été ajouté à 50  $\mu$ L de bactéries compétentes BL21(DE3) supplémentées par 1,7  $\mu$ L de  $\beta$ -mercaptoéthanol. Les cellules ont ensuite été incubées 30 minutes à 4°C puis transformées par choc thermique de 30 secondes à 42°C dans un bain marie. Après incubation de 2 minutes à 4°C, 250  $\mu$ L de milieu LB (Luria-Bertani), préalablement chauffé à 37°C, a été ajouté. La culture a été incubée 1 heure à 37°C sous agitation (200 rpm) puis 100  $\mu$ L ont étéensemencés sur des boîtes de Pétri LB-Agar contenant de l'antibiotique approprié (kanamycine à 50  $\mu$ g/mL (*Sigma*) ou carbénicilline à 75  $\mu$ g/mL au lieu d'ampicilline (*Sigma*)) (tableau 4). Les boîtes de Pétri ont été incubées une nuit dans une étuve à 37°C.

Milieu LB : tryptone (10 g/L), extrait de levure (5 g/L), NaCl (10 g/L).

Milieu LB-Agar : agar (20 g/L), tryptone (10 g/L), extrait de levure (5 g/L), NaCl (10 g/L).

## 1.2. Expression des protéines hydrogénées

Les cultures de bactéries ont toutes été réalisées dans du milieu LB contenant l'antibiotique approprié pour sélectionner les cellules possédant le vecteur plasmidique. Pour cela, une colonie de bactéries prélevée sur boîte de Pétri a servi à ensemencer une préculture de 50 mL de milieu LB contenant l'antibiotique approprié. L'incubation dure une nuit à 37 °C sous agitation (160 rpm). La préculture est alors diluée 100 fois dans du milieu LB contenant l'antibiotique approprié (8 mL de préculture dans 800 mL de milieu dans un erlenmeyer stérile de 2 L à ailettes). Les bactéries ont été cultivées à 37 °C sous agitation (160 rpm) et leur croissance a été suivie par mesure de la densité optique à 600 nm. Lorsque celle-ci a atteint 0,6 (phase exponentielle de croissance après environ 3

heures d'incubation), la surexpression des protéines a été induite par ajout de 1 mM IPTG. La culture induite a alors été incubée sous agitation (160 rpm) pendant 4 h à 37°C pour PAN et les deux versions de la streptavidine ou une nuit à 20°C pour la particule 20S et les différentes versions de la GFPssrA. Les cellules ont ensuite été culottées par centrifugation à 3000 x g pendant 20 minutes à 4 °C. Les culots bactériens ont été utilisés immédiatement pour la purification ou congelés à -80 °C.

### 1.3. Expression des protéines deutérées

La deutération des molécules biologiques implique la croissance d'une culture bactérienne (le plus souvent *E. coli*) dans de l'eau lourde (D<sub>2</sub>O) supplémentée avec des sources de carbone deutérées ou hydrogénées, en fonction du niveau de deutération requis. La biomasse est ensuite récoltée et la molécule deutérée est purifiée.

La deutération permet l'utilisation des techniques physiques basées sur les neutrons comme le SANS pour l'étude de la relation structure-fonction des protéines, ADN, ARN ou d'autres molécules biologiques. L'hydrogène est un élément majeur et essentiel dans la composition des molécules biologiques y compris les protéines. Comme les neutrons interagissent différemment avec les atomes d'hydrogène ou de deutérium, le contraste entre une protéine hydrogénée et une protéine deutérée est suffisant pour séparer les deux contributions et donc les techniques de diffusion de neutrons sont très utiles pour l'étude des macromolécules biologiques surtout dans le cas de complexes protéine-protéine (voir chapitre 2).

Les extraits cellulaires d-GFPssrA, d-GFPssrA\_G10C et d-GFPssrA\_H231C ont été produits avec l'aide de Martine Moulin au sein du laboratoire de deutération de l'ILL (D-Lab) selon un protocole identique pour les trois protéines.

#### 1.3.1. Adaptation des bactéries

Les plasmides et la souche bactérienne utilisés pour la production des protéines deutérées ont été les mêmes que ceux utilisés pour l'expression des protéines hydrogénées (partie 1.1). Les milieux composés de D<sub>2</sub>O et d'une source carbonée deutérée (d8-glycérol) étant très coûteux, une stratégie de culture à haute densité en fermenteur a été utilisée. Cela permet d'obtenir un haut rendement de biomasse et par conséquent un haut rendement de protéines recombinantes deutérées pour un faible volume de culture (1,6 L)<sup>327,353</sup>.

Le milieu choisi pour la fermentation est le milieu minimum Enfors<sup>354</sup> que l'on peut produire sous sa forme hydrogénée ou deutérée. Le glycérol présent dans le milieu est la source unique de carbone. Dans sa forme deutérée, l'utilisation conjointe de D<sub>2</sub>O et de glycérol deutéré permet d'avoir un taux de deutération proche de 100% (appelé « perdeutération »)<sup>355</sup>. La souche *E. coli* BL21(DE3) étant prototrophe (capable de synthétiser tous les composés nécessaires à son développement), elle peut pousser dans un milieu minimum.

Milieu Enfors (préparé en H<sub>2</sub>O ou D<sub>2</sub>O) : 6,86 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1,56 g/L KH<sub>2</sub>SO<sub>4</sub>, 6,01 g/L Na<sub>2</sub>HPO<sub>4</sub> anhydre, 0,49 g/L (NH<sub>4</sub>)<sub>2</sub>-H-Citrate, 1 mM MgSO<sub>4</sub> (H<sub>2</sub>O ou D<sub>2</sub>O), 1% solution minérale 1000X (H<sub>2</sub>O ou D<sub>2</sub>O), 5 g/L glycérol 100% (hydrogéné ou deutéré), 30 µg/mL kanamycine (H<sub>2</sub>O ou D<sub>2</sub>O).

Solution minérale 1000X (préparée en H<sub>2</sub>O ou D<sub>2</sub>O) : 0,5 g/L CaCl<sub>2</sub>.2H<sub>2</sub>O, 16,7 g/L FeCl<sub>3</sub>.6H<sub>2</sub>O, 0,18 g/L ZnSO<sub>4</sub>.7H<sub>2</sub>O, 0,16 g/L CuSO<sub>4</sub>.5H<sub>2</sub>O, 0,15 g/L MnSO<sub>4</sub>.4H<sub>2</sub>O, 0,18 g/L CoCl<sub>2</sub>.6H<sub>2</sub>O,

20,1 g/L EDTA.Na<sub>2</sub>. La préparation de la solution sous sa forme deutérée a été précédée d'un échange de tous les hydrogènes labiles contenus dans les poudres commerciales.

Le milieu d'expression doit être perdeutééré (100% deutéré) afin que les protéines soient elles-mêmes perdeutéérées. Afin d'optimiser l'adaptation et la croissance des bactéries en milieu Enfors minimum perdeutééré, les étapes suivantes ont été effectuées :

- Transformation et croissance des bactéries sur une boîte de LB-agar contenant de la kanamycine (50 µg/mL) (partie 1.1).
- Inoculation d'une colonie dans 10 mL de milieu Enfors hydrogéné sur la nuit à 37 °C.
- Inoculation au 1/10 de 10 mL de milieu Enfors deutéré de la culture précédente. Cette étape est réalisée 5 à 6 fois pour obtenir une croissance optimale en milieu minimum.
- Inoculation au 1/10 de 150 mL de milieu Enfors de la culture précédente. Cette dernière étape permet d'obtenir la préculture nécessaire pour inoculer le fermenteur qui servira après pour la croissance des bactéries et l'expression de la protéine.

Des tests d'expression ont été réalisés en parallèle de l'étape 3 afin de d'optimiser les conditions de croissance (temps et température d'induction) permettant le meilleur rendement des protéines exprimées en milieu perdeutééré.

### 1.3.2. Culture par fermentation

La culture a été réalisée en biofermenteur Labfors II (*Infors*) de 2,5 L, inoculée par une préculture au 1/10. Durant toute la fermentation, le pD, stabilisé à 6,9, l'oxygène dissous dans le milieu (pO<sub>2</sub>), avec pour valeur de consigne 30%, et la température à 30 °C ont été contrôlés et régulés.

La fermentation s'est déroulée généralement en trois phases :

- La phase de *Batch* : le but de cette première phase a été d'initier la croissance des bactéries. La source de carbone est alors le d8-glycérol présent dans le milieu Enfors en concentration limitante sans addition supplémentaire dans le milieu de culture.
- La phase de *Fed-Batch* : une fois que tout le glycérol du milieu consommé, une acidification progressive du milieu jusqu'au point de consigne (pD=6,9) et une augmentation brutale de la pO<sub>2</sub> (> 70 %) ont été observées. L'augmentation de pO<sub>2</sub> indique l'arrêt de la consommation d'oxygène de la bactérie, par manque de source de carbone. Le milieu a alors été supplémenté à intervalles réguliers en d8-glycérol tout au long de cette deuxième phase. La supplémentation du milieu en source de carbone a été réalisée avec du milieu Enfors deutéré contenant 120 g/L de glycérol deutéré. Cette phase de *Fed-Batch* avait pour but d'entretenir la croissance des bactéries pour augmenter d'une façon contrôlée la biomasse tout en évitant la production d'acides (acide acétique par exemple). L'intervalle entre chaque ajout de milieu et la quantité délivrée ont été définis de manière à ne pas restreindre la croissance bactérienne. En parallèle de l'ajout de d8-glycérol, le pD a été maintenu à 6,9 par ajout de NaOD via une pompe contrôlée automatiquement.
- La phase d'induction : une fois l'absorbance à 600 nm entre 10 et 12, une induction à 1 mM IPTG (D<sub>2</sub>O) a été effectuée. Après une nuit et à une absorbance à 600 nm d'environ 16, les bactéries ont été récoltées par centrifugation à 3 000 x g pendant 20 minutes à 4 °C. En moyenne, 50 à 70 g de culot cellulaire ont été récoltés pour 1,6 L de milieu dans le fermenteur. Par comparaison, un litre de milieu LB à 37 °C produit 3 à 4 g de culot cellulaire.

## 2. Purification des protéines d'intérêt

### 2.1. Lyse des cellules

Les cellules issues de 800 mL de culture hydrogénée ou 10 g de cellules issues de culture deutérée ont été remises en suspension dans 50 mL tampon de lyse supplémenté par 0,25 mg/mL de lysozyme (*Euromedex*), 0,05 mg/mL de DNase I grade II (*Roche*), 0,2 mg/ml de RNase (*Roche*), 1 mg/mL de Pefabloc SC (*Roche*) et 0,01 mg/mL de MgSO<sub>4</sub>. Les cellules ont été lysées à l'aide d'un broyeur cellulaire à 4 cycles à 18 kpsi à 7 °C (*Constant System LTD*).

Tampon de lyse pour PAN, la particule 20S, et les versions de la streptavidine : 20 mM Tris-HCl pH 7,5, 150 mM NaCl, 0,1% Triton X-100, 5 mM imidazole.

Tampon de lyse pour les différentes versions de la GFPssrA : 20 mM Tris-HCl pH 7,5, 150 mM NaCl, 0,1% Triton X-100.

### 2.2. Caractéristiques des protéines et choix des tampons

Le tableau 5 récapitule les caractéristiques de différentes protéines étudiées (masse moléculaires, point isoélectrique, etc.)

**Tableau 5 : Tableau récapitulatif des informations concernant les protéines d'intérêt.** Les valeurs de masses moléculaires, points isoélectriques, coefficients d'extinction et absorbance de 0,1% correspondant à 1 g/L ont été obtenues grâce au programme ExPASy ProtParam Tool®. Les colonnes GFPssrA-biotine et GFPssrA-streptavidine concernent respectivement les GFPssrA\_Nter-biotine & GFPssrA\_Cter-biotine et GFPssrA\_Nter-streptavidine & GFPssrA\_Cter-streptavidine. \*Valeurs ajustées expérimentalement et utilisées lors de la préparation de la protéine PAN.

	PAN (hexamère)	Particule 20S	GFPssrA	Streptavidine monovalente	GFPssrA - biotine	GFPssrA - streptavidine
<b>Masse moléculaire</b> (kDa)	298,4	751,3	27,7	54,3	27,9	82,2
<b>Point isoélectrique</b> pI	5,80	6,90	5,50	6,00	5,50	5,80
<b>Coefficient</b> <b>d'extinction <math>\epsilon</math></b> (L.mol <sup>-1</sup> .cm <sup>-1</sup> )	45825 79385*	296415	22015	1678	22015	1897
<b>Abs 0,1% (=1 g/L)</b> (M <sup>-1</sup> .cm <sup>-1</sup> )	0,154 0,266*	0,395	0,796	3,090	0,796	2,316

Ces paramètres ont permis de choisir les colonnes de chromatographies utilisées et caractéristiques du tampon de travail utilisé pour la purification de toutes les protéines étudiées, puis comme tampon pour les tests fonctionnels par la suite. Plusieurs tampons ont été testés (variations de concentrations en sels NaCl ou KCl, concentrations en MgCl<sub>2</sub>, variations de pH) dans le but principal d'optimiser l'état oligomérique de PAN, mais aucun tampon n'a été trouvé meilleur que celui utilisé dans les travaux précédents (thèse de Ziad Ibrahim).

En fonction des types de chromatographies utilisées, différents éléments ont été ajoutés au tampon de travail, tel que de l'imidazole ou du NaCl. Pour les protéines préparées en vue des mesures

SANS, le tampon de travail de la dernière étape de purification de chaque protéine contenait 42% de D<sub>2</sub>O.

L'absorbance à 280 nm a été mesurée pour déterminer la concentration des solutions protéiques grâce à un NanoDrop® (*Thermo Scientific NanoDrop® 2000 Spectrophotometer*) mesurant l'absorbance à 280 nm et calculant la concentration grâce à la loi de Beer-Lambert ( $A = \epsilon \cdot l \cdot c$  où  $l = 1$  cm). Les valeurs des masses moléculaires, coefficient d'extinction et absorbance de 0,1% correspondant à 1 g/L utilisées pour l'estimation de la concentration de chaque protéine étudiée sont rassemblées dans le tableau 5.

Tampon de travail : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

### 2.3. Purifications des protéines

Pour toutes les colonnes de chromatographie présentées dans cette partie, le système de purification ÄKTAFPLC® (*GE Healthcare*) a été utilisé et l'absorbance à 280 nm a été suivie grâce au détecteur UV-visible de l'appareil. Dans les étapes impliquant de la GFP<sub>ssrA</sub>, l'absorbance à 509 nm a également été suivie.

#### 2.3.1. Purification de la particule régulatrice PAN

Les protéines provenant de l'organisme hyperthermophile *M. jannaschii* sont généralement plus thermostables que les protéines d'*E. coli*. Un traitement thermique de 20 minutes à 70 °C a été effectué sur l'extrait cellulaire après la lyse comme une première étape de purification. Ce traitement a permis d'éliminer une grande partie de protéines d'*E. coli* instables à cette température. La solution a ensuite été clarifiée par centrifugation à 10000 x g pendant 1 h à 4 °C.

Le surnageant a été filtré (0,2 µM) et chargé sur une colonne de chromatographie d'affinité HisTrap Ni<sup>2+</sup> (*GE Healthcare*) préalablement équilibrée dans le tampon de travail contenant 5 mM d'imidazole. L'étiquette poly-histidine de la protéine recombinante PAN interagit avec le nickel chélaté sur la résine de la colonne, tandis que la plupart des autres protéines en solution n'ont pas été retenues par la colonne. Deux étapes de lavage avec le tampon de travail contenant 20 mM puis 50 mM d'imidazole pour 5 volumes de colonne (VC) ont permis d'éliminer les protéines interagissant faiblement avec la résine par compétition avec l'imidazole pour le nickel. Les protéines en interaction plus forte ont alors été éluées avec un gradient linéaire d'imidazole de 10 VC (50-400 mM d'imidazole dans le tampon de travail). PAN a été éluée à environ 150 mM d'imidazole, et la solution protéique a été dialysée contre 5 L de tampon de travail pendant une nuit.

La solution a ensuite été de nouveau filtrée (0,2 µM) et chargée sur une colonne de chromatographie échangeuse d'anions Ressource Q (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. La chromatographie par échange d'ions sépare les protéines selon leurs groupes chargés respectifs. Selon le pI de PAN mentionné dans le tableau 5, la protéine est chargée négativement au pH de 7,5 du tampon de travail. Les charges négatives de PAN entrent alors dans des interactions ioniques avec des charges positives fixées sur la matrice de la colonne, entraînant la rétention de la protéine sur celle-ci. Après lavage avec le tampon de travail par 4 VC pour éliminer les protéines en interaction faible, les protéines liées ont été éluées avec un gradient NaCl linéaire de 15 VC (100-300 mM NaCl dans le tampon de travail) par compétition croissante des ions Cl<sup>-</sup> pour neutraliser les charges positives de la résine. PAN a été éluée à environ 180 mM de NaCl.

Après concentration du pic de PAN avec un concentrateur Amicon ® Ultra 50 kDa MWCO (*Millipore*) jusqu'à un volume inférieur à 10 mL par centrifugation à 5000 x g à 15°C, la protéine a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis chargée par aliquotes de 500 µL sur une colonne de chromatographie d'exclusion de taille Superose 6 Increase 10/300 GL (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. Lors de cette technique, plus les protéines possèdent un rayon hydrodynamique de petite taille, plus elles sont retardées dans les pores des billes de la résine. Les protéines sont alors séparées selon leur masse moléculaire et leur taille. Les fractions correspondant au pic d'éluion de la protéine ont été combinées et concentrées à la concentration souhaitée de la même manière que précédemment. À cette étape, la protéine PAN est pure sur gel d'électrophorèse (partie 2.4) et d'après les mesures par spectrométrie de masse (partie 4.3). Les protéines purifiées ont été conservées à -20 °C pour des utilisations ultérieures.

### 2.3.2. Purification de la particule catalytique 20S

Les cellules utilisées pour la coexpression des sous-unités  $\alpha$  et  $\beta\Delta 6$  de la particule catalytique 20S (voir tableau 5) dans le but d'obtenir une particule catalytique 20S assemblée ont été lysées (partie 2.1). L'extrait cellulaire après la lyse des cellules a subi un traitement thermique de 20 minutes à 85°C, puis a été centrifugé, filtré et chargé sur une colonne de chromatographie d'affinité HisTrap Ni<sup>2+</sup> (*GE Healthcare*) selon le protocole décrit dans la partie 2.3.1. La particule 20S a été éluée à environ 200 mM d'imidazole. La solution protéique a été dialysée contre 5 L de tampon de travail pendant une nuit, puis concentrée avec un concentrateur Amicon ® Ultra 100 kDa MWCO (*Millipore*) jusqu'à un volume inférieur à 10 mL par centrifugation à 5000 x g à 15°C. La protéine a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis purifiée par une colonne de chromatographie d'exclusion de taille Superose 6 Increase 10/300 GL (*GE Healthcare*) de la même manière que décrit dans la partie 2.3.1. Les fractions correspondant au pic d'éluion de la protéine ont été combinées et concentrées à la concentration souhaitée de la même manière que précédemment. À cette étape, la particule 20S est quasiment pure sur gel d'électrophorèse (partie 2.4) et d'après les mesures par spectrométrie de masse (partie 4.3), avec une contamination de sous-unités  $\alpha$  auto-assemblées indissociable de la protéine. Les protéines purifiées ont été conservées à -20 °C pour des utilisations ultérieures.

### 2.3.3. Purification de la GFPssrA

Le protocole décrit ci-dessous a été adapté d'un protocole décrit dans la littérature<sup>356</sup>. Toutes les étapes de la purification de la GFPssrA ont été réalisées à 4°C. La GFPssrA deutérée (d-GFPssrA) a été purifiée selon le même protocole et dans les mêmes tampons.

Le surnageant issu de la lyse des cellules a été supplémenté par 100 mM de triéthanolamine base (TEA base) et 1,6 M de sulfate d'ammonium (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (saturation finale de 40%). Après incubation d'une heure à 4°C, les protéines précipitées ont été éliminées par centrifugation à 4°C pendant 20 minutes à 4000 x g. Le surnageant a alors été supplémenté par 2,8 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (saturation finale de 70%). Une extraction à l'éthanol a été réalisée sur cette solution.

Pour cela, 1/4 de volume d'éthanol ( $\geq 99,8\%$ ) a été ajouté et la solution a été agitée vigoureusement pendant une minute. Les deux phases ont été séparées par centrifugation à 3000 x g pendant 10 minutes à 4°C et la phase organique (du dessus) a été prélevée. 1/16 de volume d'éthanol a ensuite été ajouté à la solution. La solution a été de nouveau secouée vigoureusement pendant une minute puis centrifugée de façon identique. La phase organique a été de nouveau prélevée, avec laquelle

une extraction au butan-1-ol a été effectuée en ajoutant 1/4 de volume de butan-1-ol. La solution a été agitée vigoureusement 30 secondes et centrifugée comme précédemment. La phase aqueuse (du dessous) a alors été récupérée. Un volume équivalent de chloroforme a été ajouté, et la solution vigoureusement secouée 30 secondes et centrifugée de la même manière. La phase aqueuse (du dessus) contenant la GFPssrA a été prélevée. Un volume égal de solution de 1,2 M de  $(\text{NH}_4)_2\text{SO}_4$  (saturation finale de 30%), préparée dans de l'eau distillée, a été ajouté à la phase organique et la solution a été vigoureusement secouée 30 secondes et centrifugée comme précédemment. La phase aqueuse (du dessus) a alors été prélevée et combinée à la solution de GFPssrA précédemment récupérée.

La solution protéique contenant la GFPssrA soluble a ensuite été centrifugée à 10000 x g pendant 10 minutes à 4°C, puis chargée par aliquotes de 1 mL sur une colonne Phenyl Sepharose 6 Fast Flow (*GE Healthcare*), équilibrée avec un tampon 20 mM Tris-HCl pH 7,5 et 0,8 M  $(\text{NH}_4)_2\text{SO}_4$  (saturation finale de 20%). Dans ce tampon à haute force ionique, les molécules d'eau de l'enveloppe d'hydratation des protéines sont mobilisées pour hydrater les anions et les cations issus de la dissociation du sel. Il en résulte une modification de l'organisation des molécules d'eau autour des protéines et ce changement d'environnement est thermodynamiquement favorable à l'établissement d'interactions hydrophobes (liaisons hydrogènes, de van der Waals ou hydrophobes) entre les régions hydrophobes à la surface des protéines et le groupement hydrophobe phénol de la matrice. Une protéine est alors d'autant plus retenue qu'elle possède des motifs hydrophobes en surface (comme c'est le cas de la GFPssrA). Après un lavage avec le même tampon par 3 VC afin d'éliminer les protéines non fixées, l'élution des protéines a été obtenue en faisant passer 3 VC de tampon de travail, qui est de force ionique plus faible. Les acides aminés chargés ou polaires à la surface des protéines peuvent de nouveau établir des liaisons hydrogène avec le tampon. Les protéines sont donc de nouveau solubles dans le tampon et sont éluées de la colonne.

Après concentration du pic de GFPssrA élue avec un concentrateur Amicon® Ultra 5 kDa MWCO (*Millipore*) jusqu'à un volume inférieur à 10 mL par centrifugation à 5000 x g à 4°C, la protéine a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis chargée par aliquotes de 500 µL sur une colonne de chromatographie d'exclusion de taille Superose 12 10/300 GL (*GE Healthcare*) de la même manière que décrit dans la partie 2.3.1. Les fractions correspondant au pic d'élution de la protéine ont été combinées et concentrées à la concentration souhaitée de la même manière que précédemment. À cette étape, la GFPssrA est pure sur gel (partie 2.4) et d'après les mesures par spectrométrie de masse (partie 4.3), malgré une fraction de la population ayant perdu l'étiquette d'adressage ssrA indissociable de la forme intacte. Les protéines purifiées ont été conservées à -20 °C pour des utilisations ultérieures.

#### 2.3.4. Purification de la streptavidine monovalente

La streptavidine monovalente, constituée d'une sous-unité capable de lier la biotine et de trois sous-unités en étant incapables, a été purifiée par Jacques Covès selon un protocole déjà décrit dans la littérature<sup>357</sup>.

La protéine a été concentrée avec un concentrateur Amicon® Ultra 10 kDa MWCO (*Millipore*) par centrifugation à 5000 x g à 4°C jusqu'à la concentration souhaitée, et visualisée pure sur gel

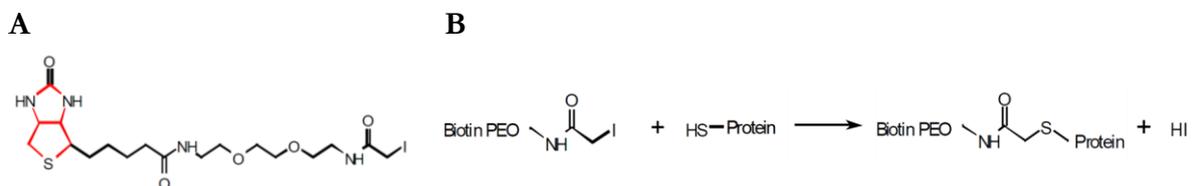
d'électrophorèse. Les protéines purifiées ont été conservées à -20 °C pour des utilisations ultérieures.

### 2.3.5. Purification de la GFPssrA\_Nter-biotine et de la GFPssrA\_Cter-biotine

Les GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine ont été préparées indépendamment selon un protocole identique, dont toutes les étapes ont été réalisées à 4°C.

Les protéines mutantes GFPssrA\_G10C et la GFPssrA\_H231C ont été purifiées selon le même protocole que la GFPssrA (partie précédente). Les protéines deutérés (d-GFPssrA\_G10C et d-GFPssrA\_H231C) ont été purifiées en utilisant le même protocole et les mêmes tampons. À l'issue de la purification, la solution protéique a été incubée une nuit à 4°C en présence de 10 mM DTT afin d'éviter la dimérisation des protéines par formation de ponts disulfure.

Afin d'éliminer le DTT, la solution a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis chargée par aliquotes de 1 mL sur une colonne de chromatographie d'exclusion de taille HiTrap Desalting (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. La protéine a été éluée par passage de 5 VC de tampon de travail, puis concentrée grâce à un concentrateur Amicon® Ultra 5 kDa MWCO (*Millipore*) jusqu'à un volume inférieur à 10 mL par centrifugation à 5000 x g à 4°C. La protéine a alors été immédiatement incubée en présence de biotine polyethyleneoxide (PEO) iodoacetamide (*Sigma*) pendant une nuit à 4°C, avec un ratio de 10 molécules de biotine pour une molécule de GFPssrA mutante. À pH 7,5, le groupe iodoacétamide de la biotine PEO réagit spontanément avec le groupe thiol de la cystéine, permettant une liaison covalente entre la biotine et la GFPssrA-G10C ou GFPssrA-H231C (figure 40). Les protéines résultantes sont alors appelées respectivement GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine.



**Figure 40 : Réaction de biotinylation avec la biotine polyethyleneoxide iodoacetamide.** A. Structure chimique semi-développée de la Biotine Polyethyleneoxide (PEO) Iodoacetamide. B. Représentation schématique de la liaison covalente entre la Biotine PEO Iodoacétamide et un groupe thiol d'une cystéine. D'après la fiche de description du produit (*Sigma B2059*).

L'excès de biotine a ensuite été éliminé par colonne HiTrap Desalting et la protéine éluée a de nouveau été concentrée comme précédemment jusqu'à la concentration souhaitée. À cette étape, les protéines GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine sont pures sur gel d'électrophorèse (partie 2.4) et d'après les mesures par spectrométrie de masse (partie 4.3), malgré une fraction de la population de GFPssrA ayant perdu l'étiquette d'adressage ssrA indissociable de la forme intacte. Les protéines ont été conservées à -20 °C pour des utilisations ultérieures ou utilisées directement pour la formation de complexe GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine.

### 2.3.6. Préparation de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine

La GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine ont été préparées indépendamment selon un protocole identique, dont toutes les étapes ont été réalisées à 4°C.

La GFPssrA\_Nter-biotine ou GFPssrA\_Cter-biotine (partie précédente) a été incubée avec la streptavidine monovalente (partie 2.3.4) pendant une nuit à 4°C, avec un ratio d'une molécule de GFPssrA-biotine pour une molécule de streptavidine monovalente dans le but de lier une molécule de GFPssrA-biotine à une molécule de streptavidine monovalente. La biotine est capable de se lier spontanément de façon non covalente à la sous-unité sauvage de la streptavidine monovalente avec une affinité très forte ( $K_D = 10^{-14}$ )<sup>358</sup>.

La solution protéique a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis chargée par aliquotes de 1 mL sur une colonne de chromatographie d'exclusion de taille HiLoad 16/600 Supedex 75 PG (*GE Healthcare*) de la même manière que décrit dans la partie 2.3.1. Les fractions correspondant au pic d'éluion de la protéine ont été combinées et concentrées à la concentration souhaitée grâce à un concentrateur Amicon® Ultra 10 kDa MWCO (*Millipore*). À cette étape, les complexes GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine sont purs sur gel d'électrophorèse (partie 2.4) et d'après les mesures par spectrométrie de masse (partie 4.3), malgré une fraction de la population de GFPssrA ayant perdue l'étiquette d'adressage ssrA indissociable de la forme intacte. Les protéines purifiées ont été conservées à -20 °C pour des utilisations ultérieures.

## 2.4. Électrophorèses sur gel SDS-PAGE

Chaque étape de l'expression et de la purification des protéines d'intérêt a été analysée sur gel d'électrophorèse en conditions dénaturantes. Les gels à 15% de polyacrylamide ont été coulés en deux parties : d'abord un gel de séparation (15% polyacrylamide préparé en tampon 1 M Tris-HCl pH8,8 en présence de 0,1% SDS) puis un gel de concentration (5% polyacrylamide préparé en tampon 1 M Tris-HCl pH 6,8 en présence de 0,1% SDS) dans lequel un peigne a été inséré.

Lors des migrations électrophorétiques, 10 µL de chaque échantillon ont été mélangés avec 5 µL de tampon de dénaturation, contenant notamment du DTT réduisant les ponts disulfures entre les cystéines et du SDS rompant les interactions hydrophobes intra- et inter-protéiques tout en chargeant les protéines négativement. Les échantillons ont ensuite été chauffés pendant 10 minutes à 100°C, puis déposés à côté de 5 µL de marqueur de taille (*Lifetechnologies PageRuler® Prestained Proteins Ladder*). Après migration des échantillons pendant 45 minutes à 30 mA dans un tampon de migration, les gels ont été colorés dans un bain de coloration d'InstantBlue (*Expedeon*) puis rincés dans de l'eau distillée.

Tampon de dénaturation : 50 mM Tris-HCl pH6,8, 8 M urée, 75 mM DTT, 3% SDS, 0,05% bleu de bromophénol, 2 M thiourée.

Tampon de migration : 25 mM Tris, 200 mM Glycine, 0,1 % SDS.

### 3. Effet de la température sur la perte de l'étiquette d'adressage « ssrA » de la GFP<sub>ssrA</sub>

#### 3.1. Test d'expression

Différentes conditions d'expression ont été testées afin de déterminer les conditions optimales pour l'expression de la GFP<sub>ssrA</sub> intacte. Plusieurs temps d'expression après induction à l'IPTG (5h, 20h, 22h et 24h) ont été testés à 20°C et à 30°C tel que décrit dans la partie 1.2. Des échantillons ont été déposés sur gel SDS-PAGE à 15% d'acrylamide et la migration électrophorétique a été réalisée selon le protocole décrit partie 2.4.

#### 3.2. Effet de la température de stockage

L'influence de la température de stockage de la GFP<sub>ssrA</sub> purifiée a été étudiée. Pour cela, des échantillons de GFP<sub>ssrA</sub> purifiée à 1 mg/mL ont été conservés pendant un mois à température ambiante, 4°C, -20°C ou -80°C. Les échantillons ont alors été déposés sur gel SDS-PAGE à 15% d'acrylamide et la migration électrophorétique a été réalisée tel que décrit partie 2.4. Les échantillons ont également été mesurés par spectrométrie de masse comme décrit dans la partie 4.3.

## 4. Mesure de la consommation de l'ATP par PAN

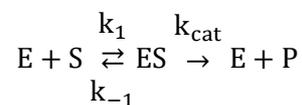
### 4.1. Théorie de l'enzymologie

#### 4.1.1. Notion de vitesse de réaction

Une vitesse de réaction en phase liquide est conventionnellement exprimée en variation de la concentration de produit (ou de substrat) en fonction du temps car ce sont les concentrations qui caractérisent l'état d'avancement d'une réaction par rapport à son équilibre. La vitesse de réaction caractérise donc l'aspect cinétique de la réaction enzymatique. Elle s'exprime en M.s<sup>-1</sup>.

#### 4.1.2. Relation de Michaelis-Menten

Le modèle Michaelis et Menten permet de décrire le comportement d'un grand nombre d'enzymes<sup>359</sup>. Dans ce modèle, une enzyme E fixe un substrat S de manière réversible pour former un complexe intermédiaire ES. Une réaction irréversible en résulte et mène au relargage du produit P et de l'enzyme libre :



Équation 19

Dans certains cas, la conversion de ES en E et P peut comporter plusieurs étapes et/ou n'est pas nécessairement irréversible. En début de la réaction, la concentration en substrat S n'a pas encore eu le temps de varier significativement, on peut donc considérer qu'elle est constante, ce qui simplifie les calculs. De plus, la concentration en produit P de la réaction est encore faible, ce qui permet de négliger une réaction inverse et éventuellement l'inhibition de l'enzyme par le produit.

La vitesse de la réaction enzymatique  $v$  peut être alors écrite :

$$v = - \frac{d[S]}{dt} = \frac{d[P]}{dt} = k_{\text{cat}}[ES]$$

Équation 20

$v = \frac{d[P]}{dt}$  est, pour chaque temps  $t$ , la pente de la tangente à la courbe  $[P] = f(t)$ .

L'évolution de la quantité de complexe  $[ES]$  est le résultat de la différence entre le taux de formation du complexe  $[ES]$  et de sa dissociation :

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_{\text{cat}}[ES]$$

Équation 21

Si la vitesse de la réaction est constante, cela implique que  $[ES]$  est constant, soit :

$$\frac{d[ES]}{dt} = 0$$

Équation 22

Cette hypothèse est justifiée par le fait que la concentration en enzyme  $E$  est faible par rapport aux concentrations en substrat  $S$  et produit  $P$ , et donc la concentration en n'importe quel intermédiaire catalytique  $ES$  est faible, et cette concentration ne peut pas varier beaucoup en valeur absolue. Ainsi, selon l'équation 21, on a :

$$k_1[E][S] - k_{-1}[ES] - k_{\text{cat}}[ES] = 0$$

Équation 23

On peut alors exprimer  $[ES]$  en fonction de la quantité d'enzyme totale  $[E_t]$ ,  $[S]$  et des constantes  $k_1$ ,  $k_{-1}$  et  $k_{\text{cat}}$  :

$$[ES] = \frac{[E_t][S]}{(k_{-1} + k_{\text{cat}})/k_1 + [S]}$$

Équation 24

La vitesse initiale de la réaction,  $v_0$ , peut alors être exprimée en fonction des quantités  $[E_t]$  et  $[S]$  mesurées expérimentalement, c'est l'équation de Michaelis-Menten :

$$v_0 = k_{\text{cat}}[ES] = \frac{k_{\text{cat}}[E_t][S]}{K_M + [S]}$$

Équation 25

Où  $K_M$  est la constante de Michaelis définie comme suit :

$$K_M = \frac{k_{-1} + k_{\text{cat}}}{k_1}$$

Équation 26

La vitesse maximale de la réaction  $v_{\text{max}}$  est atteinte lorsque le substrat est en excès (concentration saturante) et que tous les sites actifs de l'enzyme sont occupés. Dans ce cas  $[E_t] = [ES]$ , donc :

$$v_{\max} = k_{\text{cat}}[E_t]$$

Équation 27

L'équation de Michaelis-Menten s'écrit alors :

$$v = \frac{v_{\max}[S]}{K_M + [S]}$$

Équation 28

Dans le cas où  $k_{\text{cat}}$  est très petit devant  $k_{-1}$ , la constante de Michaelis devient alors :

$$K_M \approx \frac{k_{-1}}{k_1} = K_S$$

Équation 29

Où  $K_S$  est la constante de dissociation du complexe  $[ES]$ . Dans ce cas, le  $K_M$  s'apparente à une mesure de l'affinité de l'enzyme pour son substrat.

### 4.1.3. Représentations graphiques

#### Représentation de Michaelis-Menten

En fixant  $[E]$  et en faisant croître  $[S]$ , on constate que la vitesse croît et tend vers une limite  $v_{\max}$  quand toute l'enzyme est saturée de substrat, c'est-à-dire quand toute l'enzyme est sous forme de complexe  $ES$ .

Ainsi, si  $[S]$  devient très grande,  $K_M \ll [S]$  et  $v_0 = v_{\max}$ , tel que représenté dans la figure 41.

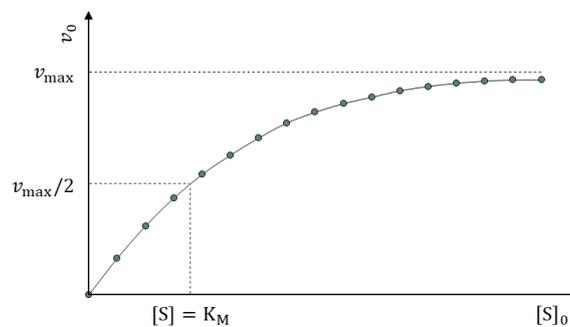


Figure 41 : Représentation de Michaelis-Menten sous forme d'hyperbole rectangulaire.

#### Représentation de Lineweaver & Burk

La représentation graphique de Michaelis-Menten ne donnant qu'une valeur approximative des paramètres  $K_M$  et  $v_{\max}$ , la représentation en double inverse  $1/v_0 = f(1/[S]_0)$  est utilisée, telle que présentée dans la figure 42. Les paramètres de  $[S]_0$ ,  $v_0$ ,  $v_{\max}$ ,  $K_M$  sont alors reliés par l'équation suivante :

$$\frac{1}{v_0} = \frac{1}{v_{\max}} + \frac{K_M}{v_{\max}} \frac{1}{[S]_0}$$

Équation 30

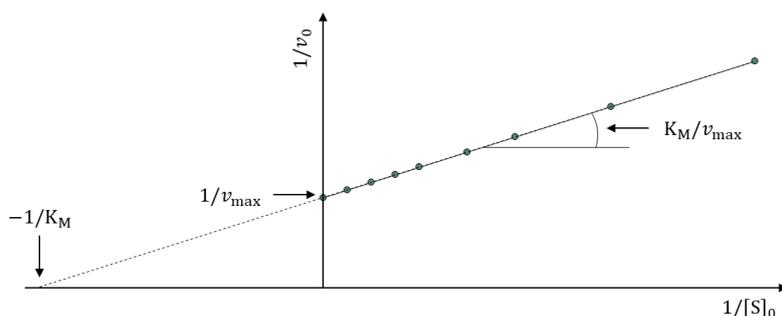


Figure 42 : Représentation Lineweaver & Burk sous forme linéaire.

#### 4.1.4. Signification de $k_{\text{cat}}$ , $K_M$ et $k_{\text{cat}}/K_M$

La constante catalytique  $k_{\text{cat}}$  représente le nombre maximum de molécules de substrat transformées par molécules d'enzyme en une seconde. L'unité de  $k_{\text{cat}}$  est donc  $\text{s}^{-1}$ . Elle traduit l'efficacité cinétique de l'enzyme.

La représentation de Michaelis-Menten montre que  $K_M$  est égal à la concentration en substrat avec laquelle la réaction atteint la moitié de sa vitesse maximale. La constante de  $K_M$  est souvent considérée comme l'affinité de l'enzyme pour son substrat. Elle correspond à la constante de dissociation apparente du substrat. Cependant, cette relation n'est valable que pour les réactions dans lesquelles  $k_{\text{cat}}$  est très petit devant  $k_{-1}$ , ce qui conduit à l'équation 28.

C'est le rapport entre ces deux valeurs ( $k_{\text{cat}}/K_M$ ) qui représente le mieux l'efficacité catalytique d'une enzyme.

#### 4.2. Mesure de l'affinité de PAN pour l'ATP

PAN est une enzyme catalysant la réaction d'hydrolyse de l'ATP suivante :



Pour mesurer l'affinité de PAN pour l'ATP, des mesures de concentration en phosphate inorganique libre en solution ont été réalisées à différentes concentrations d'ATP. Pour cela, différentes conditions ont été mesurées à 1 mM, 2,5 mM, 5 mM, 7,5 mM, 10 mM, 12,5 mM, 15 mM et 20 mM ATP. Les conditions mesurées en triplicata sont :

- 1  $\mu\text{M}$  PAN
- 1  $\mu\text{M}$  PAN + 1  $\mu\text{M}$  GFPssrA
- 1  $\mu\text{M}$  GFPssrA

Les mélanges réactionnels ont été préparés dans la glace dans un tampon contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 100 mM  $\text{MgCl}_2$  dans un volume final de 20  $\mu\text{L}$ . Ils ont ensuite été incubés à 55°C pendant 2 minutes avant d'être replacés dans la glace afin d'arrêter la réaction. Pour le dosage, ils ont été dilués 6 fois dans le tampon de travail avant de prélever 4  $\mu\text{L}$  et d'ajouter 36  $\mu\text{L}$  de tampon de dosage et 200  $\mu\text{L}$  du réactif contenu dans le kit Sigma MAK113. La quantification est basée sur la formation d'un complexe entre le molybdate, le vert de malachite et l'orthophosphate libre qui absorbe à 620-640 nm<sup>360,361</sup>. Après 30 minutes d'incubation à température ambiante, l'absorbance à 630 nm est mesurée à l'aide d'un spectrophotomètre lecteur de plaques

(BioTek Synergy® H4 Hybrid Multi-Mode Microplate Reader). Une courbe standard de phosphate réalisée avec différentes quantités de phosphate inorganique a été utilisée pour estimer la quantité de phosphate libérée lors de l'hydrolyse de l'ATP.

#### 4.3. Mesure de la consommation de l'ATP au cours du temps

Des tests enzymatiques (hors faisceau neutrons) ont été réalisés avec les protéines dans les concentrations suivantes afin de mesurer la consommation d'ATP par PAN :

- Condition SANS 1 – 70  $\mu\text{M}$  GFPssrA, 20  $\mu\text{M}$  PAN et 20  $\mu\text{M}$  particule 20S (tableau 12)
- Condition SANS 2 – 180  $\mu\text{M}$  GFPssrA, 5  $\mu\text{M}$  PAN et 30  $\mu\text{M}$  particule 20S (tableau 13)

Pour cela, la quantité de phosphate inorganique générée par PAN a été mesurée en fonction du temps. Des mélanges réactionnels ont été préparés avec 100 mM ATP & 200 mM  $\text{MgCl}_2$  dans un tampon contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM  $\text{MgCl}_2$  dans un volume final de 20  $\mu\text{L}$  dans les conditions suivantes en triplicata :

- Tampon seul
- GFPssrA seule
- PAN seul
- Particule 20S seule
- GFPssrA, PAN et particule 20S

Les réactions ont été incubées à 55°C et suivies pendant 1 heure. Pour chaque temps de mesure choisi, 20  $\mu\text{L}$  de mélange réactionnel ont été placés dans la glace pour arrêter la réaction. Pour le dosage, les échantillons ont été dilués 150 fois dans le tampon de travail avant de prélever 4  $\mu\text{L}$  et d'ajouter 36  $\mu\text{L}$  de tampon de dosage et 200  $\mu\text{L}$  de réactif contenu dans le kit Sigma MAK113. Après 30 minutes d'incubation à température ambiante, l'absorbance à 630 nm a été mesurée et la quantité de phosphate libérée estimée comme décrit précédemment (partie 4.2).

#### 4.4. Effet d'inhibition de l'ADP sur l'hydrolyse de l'ATP

La quantité de phosphate inorganique générée par PAN en présence de GFPssrA a été mesurée en fonction du temps dans la condition SANS 1 (tableau 12). Des mélanges réactionnels ont été préparés avec 200 mM  $\text{MgCl}_2$  dans un tampon contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM  $\text{MgCl}_2$  dans un volume final de 20  $\mu\text{L}$  dans les conditions suivantes en triplicat :

- 20 mM ATP
- 80 mM ADP
- 20 mM ATP + 80 mM ADP

Les mélanges ont été incubés à 55°C et les réactions suivies pendant 30 min avec un mélange réactionnel de 20  $\mu\text{L}$  pour chaque point de temps mesuré. Pour le dosage, les échantillons ont été traités de la même manière que décrit dans la partie 4.2. Un test de dénaturation de la GFPssrA en fonction du temps par mesure de la fluorescence a été réalisé dans les mêmes conditions selon le protocole décrit en partie 5.

#### 4.5. Effet de l'ajout de GFPssrA sur la consommation d'ATP par PAN

La quantité de phosphate inorganique générée par PAN avec ou sans GFPssrA a été mesurée en fonction du temps. Des mélanges réactionnels ont été préparés avec 100 mM ATP & 200 mM

MgCl<sub>2</sub> ou 10 mM ATP & 200 mM MgCl<sub>2</sub> dans un tampon contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub> dans un volume final de 20 µL dans les conditions suivantes :

- 1 µM PAN
- 5 µM PAN
- 10 µM PAN
- 20 µM PAN
- 1 µM PAN + 180 µM GFPssrA
- 5 µM PAN + 180 µM GFPssrA
- 10 µM PAN + 180 µM GFPssrA
- 20 µM PAN + 180 µM GFPssrA

Les mélanges ont été incubés à 55°C et les réactions suivies pendant 10 min avec un mélange réactionnel de 20 µL pour chaque point de temps mesuré. Pour le dosage, les échantillons ont été traités de la même manière que décrit dans la partie 4.2.

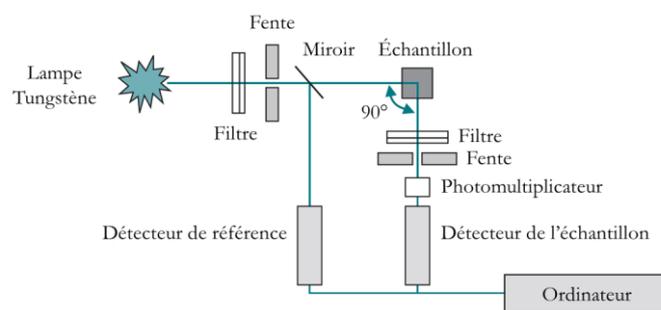
## 5. Mesures de la dénaturation de la GFPssrA par spectroscopie de fluorescence

### 5.1. Contexte des mesures par spectroscopie de fluorescence

PAN permet le dépliement de la GFPssrA en présence d'ATP et de MgCl<sub>2</sub>. Lorsque la particule 20S est également présente en solution, la GFPssrA est alors dégradée.

La fluorescence de la GFPssrA est un indicateur de son état de repliement, puisque la fluorescence de la protéine est intrinsèquement liée à sa structure, le chromophore étant protégé par la structure en tonneau β. Lorsque la GFPssrA est dénaturée (par conséquent dépliée ou dégradée), la protéine perd sa fluorescence. La dénaturation de la GFPssrA peut alors être suivie par mesure de la diminution de l'émission de fluorescence à  $\lambda_{em} = 509$  nm après excitation à  $\lambda_{ex} = 400$  nm à l'aide d'un spectrofluorimètre. La partie 4 du chapitre 5 décrit de façon plus approfondie le phénomène de fluorescence et la protéine GFPssrA.

La spectroscopie de fluorescence a l'avantage de fonctionner de façon précise à de très faibles concentrations de protéine fluorescente, là où la spectroscopie UV/Vis est moins précise à de si faibles concentrations. Cela permet l'utilisation de microcuvettes ou de plaques 96 puits (utilisées lors de ce projet) contenant peu de volume d'échantillon. La sélectivité spectrale est l'un des principaux facteurs qui ajoutent à la sensibilité élevée des mesures de fluorescence. En raison du décalage de Stokes (partie 4.1.1 du chapitre 5), la longueur d'onde de la lumière émise est différente de celle de la lumière d'excitation. Une autre caractéristique utilise le fait que la fluorescence est émise dans toutes les directions. En plaçant le détecteur à un angle de 90° par rapport à l'incidence de la lumière d'excitation, le bruit de fond du faisceau incident est réduit. Cette géométrie est utilisée au lieu de placer le détecteur sur la ligne de la lumière d'excitation à un angle de 180°. Un schéma de spectrofluorimètre typique est illustré figure 43.



**Figure 43 : Schéma d'un spectrofluorimètre à géométrie à 90°.** Les chemins optiques sont représentés par des lignes vertes. Adapté de Wilson, K., & Walker, J. (2010)<sup>362</sup>.

La lumière d'une source d'excitation passe par un filtre qui sélectionne la longueur d'onde du faisceau excitant, puis par l'échantillon. Une partie de la lumière est alors absorbée, induisant la fluorescence du chromophore. Une partie de la lumière de fluorescence est ensuite concentrée un second filtre, qui est le plus souvent placé à un angle de 90° par rapport à la lumière d'excitation et permet d'analyser l'émission de fluorescence. La lumière est ensuite captée par un détecteur. Le contrôle de la température est nécessaire pour une mesure précise car l'intensité de l'émission d'un chromophore dépend de la température de la solution.

## 5.2. Criblage des stœchiométries de GFP<sub>sstA</sub>, PAN, particule 20S et d'ATP

L'optimisation des conditions de la réaction de dépliement et de dégradation de la GFP<sub>sstA</sub> par la machinerie protéolytique était une étape importante pour définir des conditions expérimentales pour les mesures SANS.

Dans un premier temps, différentes concentrations de GFP<sub>sstA</sub>, PAN, ATP et MgCl<sub>2</sub> ont été combinées selon les concentrations décrites dans le tableau 6.

**Tableau 6 : Différentes concentrations des différents protéines et composants combinés lors du criblage des conditions d'affinité sans particule 20S.** Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

Protéine/composant	Masse moléculaire	Concentration finale	
PAN	298437,8 Da	1 μM	0,30 mg/mL
		5 μM	1,5 mg/mL
		10 μM	3 mg/mL
		20 μM	6 mg/mL
		45 μM	13,4 mg/mL
		70 μM	20,9 mg/mL
GFP <sub>sstA</sub>	27666 Da	1 μM	0,03 mg/mL
		5 μM	0,15 mg/mL
		10 μM	0,3 mg/mL
		20 μM	0,5 mg/mL
		45 μM	1,5 mg/mL
		70 μM	2 mg/mL
		125 μM	3 mg/mL
		180 μM	5 mg/mL

ATP & MgCl <sub>2</sub>	/	10 mM & 20 mM	/
		50 mM & 100 mM	
		100 mM & 200 mM	

Dans un second temps, différentes concentrations de GFPssrA, PAN, et particule 20S ont été combinées à 100 mM ATP et 200 mM MgCl<sub>2</sub> selon les concentrations décrites dans le tableau 7.

**Tableau 7 : Différentes concentrations des différents protéines et composants combinés lors du criblage des conditions d'affinité avec particule 20S.** Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

Protéine/composant	Masse moléculaire	Concentration finale	
PAN	298437,8 Da	1 µM	0,30 mg/mL
		5 µM	1,5 mg/mL
		10 µM	3 mg/mL
		20 µM	6 mg/mL
Particule 20S	747584,5 Da	1 µM	0,75 mg/mL
		5 µM	3,7 mg/mL
		10 µM	7,5 mg/mL
		20 µM	15 mg/mL
GFPssrA	27666 Da	1 µM	0,03 mg/mL
		5 µM	0,15 mg/mL
		10 µM	0,3 mg/mL
		20 µM	0,5 mg/mL
		45 µM	1,5 mg/mL
		70 µM	2 mg/mL
		125 µM	3 mg/mL
		180 µM	5 mg/mL

Après préchauffage de l'instrument à 55°C, les réactions ont été préparées à 4°C sans PAN dans un tampon de réaction contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub>. Les échantillons ont été préparés dans des plaques 96 puits permettant une mesure de la fluorescence par le dessous, avec un film transparent autocollant sur le dessus pour éviter l'évaporation des échantillons (*Corning® 96 Well Black with Clear Flat Bottom Polystyrene*).

La dénaturation de la GFPssrA a été suivie ( $\lambda_{\text{ex}} = 400 \text{ nm}$  et  $\lambda_{\text{em}} = 509 \text{ nm}$ ) à l'aide d'un spectrofluorimètre lecteur de plaques (*BioTek Synergy® H4 Hybrid Multi-Mode Microplate Reader*). La calibration de l'instrument (*autogain*) a été effectuée sur les échantillons sans PAN afin que la réaction de dépliement ne commence pas avant le début de la mesure. La plaque de mesure a ensuite été refroidie à température ambiante et la protéine PAN a ensuite été ajoutée dans les réactions devant contenir la protéine. Des contrôles négatifs de GFPssrA, PAN et particule 20S seuls ont été mesurés. La courbe de dépliement de la GFPssrA a alors été enregistrée au cours du temps avec un point de mesure toutes les 40 secondes pendant 1 heure à 55 °C. Les courbes de fluorescence de la GFPssrA en fonction du temps ont ensuite été générées en normalisant chaque courbe en définissant le premier point de mesure comme correspondant à 100% de fluorescence.

### 5.3. Mesures fonctionnelles avec différentes conditions contrôle

Des mesures ont été réalisées dans les deux conditions SANS sélectionnées :

- Condition SANS 1 – 70  $\mu\text{M}$  GFPssrA, 20  $\mu\text{M}$  PAN et 20  $\mu\text{M}$  particule 20S (tableau 12)
- Condition SANS 2 – 180  $\mu\text{M}$  GFPssrA, 5  $\mu\text{M}$  PAN et 30  $\mu\text{M}$  particule 20S (tableau 13)

Les différentes conditions testées pour chaque condition SANS en triplicata sont :

- GFPssrA seule en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$
- GFPssrA et PAN en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$
- GFPssrA, PAN et particule 20S en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$
- GFPssrA et particule 20S en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$
- GFPssrA et PAN
- GFPssrA, PAN et particule 20S

Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2.

### 5.4. Effets de l'ATP, ADP et ATP $\gamma$ S sur l'activité du protéasome PAN-20S

Des tests fonctionnels ont été effectués à 100 mM d'ATP, ADP ou ATP $\gamma$ S en présence de 200 mM  $\text{MgCl}_2$  dans les conditions SANS 1 et 2 (tableaux 12 et 13).

Pour chacune des conditions SANS, des mesures de GFPssrA seule, en présence de PAN et en présence de PAN et particule 20S ont été réalisées. Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2.

### 5.5. Effet d'inhibition de l'ADP sur le dépliement de la GFPssrA par PAN

Des mesures ont été réalisées dans la condition SANS 1 (tableau 12) sur la GFPssrA seule et la GFPssrA en présence de PAN dans les conditions suivantes en triplicata :

- 20 mM ATP (& 200 mM  $\text{MgCl}_2$ )
- 80 mM ADP (& 200 mM  $\text{MgCl}_2$ )
- 20 mM ATP et 80 mM ADP (& 200 mM  $\text{MgCl}_2$ )

Les échantillons ont été préparés et mesurés pendant 30 minutes de la même manière que décrit dans la partie 5.2.

### 5.6. Effets des variations de température sur l'activité du système protéasome PAN-20S

Des mesures ont été effectuées à différentes températures (30, 40, 45, 50, 55, 60 et 65°C) en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$  dans les conditions SANS 1 et 2 (tableaux 12 et 13).

Pour chacune des conditions SANS, des mesures de GFPssrA seule, en présence de PAN et en présence de PAN et particule 20S ont été réalisées. Après préchauffage de l'instrument à la température souhaitée, les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2. Après 1 heure de réaction, la plaque a été incubée à température ambiante

pendant 10 minutes et la suite de courbe de dépliement de la GFPssrA a alors été mesurée pendant 20 minutes dans les mêmes conditions que précédemment.

### 5.7. Effet de la streptavidine monovalente sur l'activité de dépliement de la GFPssrA par PAN

Des mesures ont été réalisées sur la GFPssrA seule et la GFPssrA en présence de PAN (25  $\mu$ M GFPssrA et 25  $\mu$ M PAN) avec ou sans streptavidine monovalente (25  $\mu$ M) en présence de 100 mM ATP & 200 mM MgCl<sub>2</sub>. Les échantillons ont été préparés et mesurés en triplicata de la même manière que décrit dans la partie 5.2.

### 5.8. Mesures de la dénaturation de la GFPssrA\_Nter-biotine et de la GFPssrA\_Cter-biotine

Des tests fonctionnels ont été réalisés avec les protéines dans les concentrations suivantes :

- Condition SANS 3 – 25  $\mu$ M GFPssrA\_Nter-biotine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 14)
- Condition SANS 4 – 25  $\mu$ M GFPssrA\_Cter-biotine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 15)
- Condition « contrôle » – 25  $\mu$ M GFPssrA, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S

Les différentes conditions testées pour chaque condition SANS en triplicata sont les mêmes que celles listées dans la partie 5.3. Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2.

### 5.9. Mesures de la dénaturation de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine

#### 5.9.1. Mesures fonctionnelles avec différentes conditions contrôle

Des mesures ont été effectuées avec les protéines dans les concentrations suivantes :

- Condition SANS 5 – 25  $\mu$ M GFPssrA\_Nter-streptavidine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 14)
- Condition SANS 6 – 25  $\mu$ M GFPssrA\_Cter-streptavidine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 15)

Les différentes conditions testées pour chaque condition SANS en triplicata sont les mêmes que celles listées dans la partie 5.3. Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2.

#### 5.9.2. Effet de l'ajout d'une quantité équimolaire de GFPssrA libre

Des tests fonctionnels ont été réalisés avec les protéines dans les concentrations suivantes :

- Condition SANS 5 – 25  $\mu$ M GFPssrA\_Nter-streptavidine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 14)
- Condition SANS 6 – 25  $\mu$ M GFPssrA\_Cter-streptavidine, 25  $\mu$ M PAN et 25  $\mu$ M particule 20S (tableau 15)

- Condition « contrôle » : 25  $\mu\text{M}$  GFPssrA, 25  $\mu\text{M}$  PAN et 25  $\mu\text{M}$  particule 20S

Les conditions testées en triplicata en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$  sont :

- GFPssrA\_Nter-streptavidine seule
- GFPssrA\_Nter-streptavidine, PAN et particule 20S
- GFPssrA\_Cter-streptavidine seule
- GFPssrA\_Cter-streptavidine, PAN et particule 20S
- GFPssrA libre seule
- GFPssrA libre
- PAN et particule 20S

Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2. Après 1 heure de réaction, la plaque a été incubée à température ambiante pendant 10 minutes. Chaque condition a ensuite été supplémentée de GFPssrA libre à 25  $\mu\text{M}$  et les échantillons ont de nouveau été mesurés pendant 1 heure dans les mêmes conditions que précédemment.

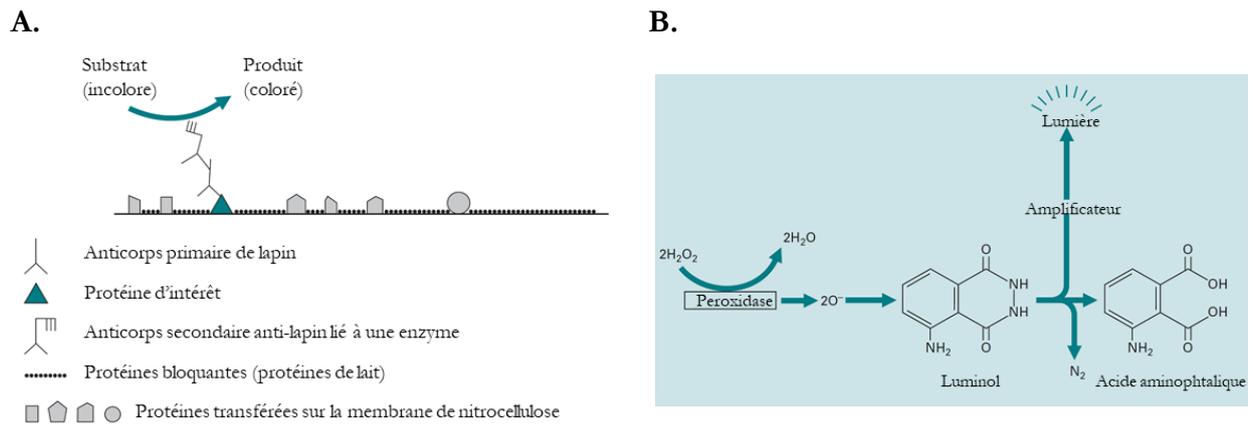
### 5.9.3. Effet de l'ajout d'une quantité cinq fois supérieure de GFPssrA libre

Les différentes conditions testées en triplicata en présence de 100 mM ATP & 200 mM  $\text{MgCl}_2$  sont les mêmes que celles listées dans la partie 5.9.29. Les échantillons ont été préparés et mesurés de la même manière que décrit dans la partie 5.2. Après 1 heure de réaction, la plaque a été incubée à température ambiante pendant 10 minutes. Chaque condition a ensuite été supplémentée de GFPssrA libre à 125  $\mu\text{M}$  et les échantillons ont de nouveau été mesurés pendant 1 heure dans les mêmes conditions que précédemment.

## 6. Évaluation de la dégradation de la GFPssrA par Western-blot anti-GFP

### 6.1. Principe du Western-blot

La technique de SDS-PAGE (partie 2.4) permet le fractionnement d'un mélange de protéines pendant le processus d'électrophorèse. Il est possible d'utiliser ce fractionnement pour examiner les protéines dénaturées. La première étape de l'analyse par Western-blot consiste à transférer les protéines séparées du gel de polyacrylamide à une membrane de nitrocellulose. Le transfert des protéines du gel à la nitrocellulose est réalisé par une technique appelée « électroblotting ». Dans cette méthode, un sandwich de gel et de nitrocellulose est comprimé dans une cassette et immergé dans un tampon entre deux électrodes parallèles. Un courant qui passe perpendiculairement au gel provoque l'électrophorèse des protéines du gel vers la feuille de nitrocellulose. La membrane de nitrocellulose avec la protéine transférée est appelée « blot ». Une fois transférée sur la nitrocellulose, la protéine d'intérêt peut être détectée parmi toutes les protéines séparées (figure 44).



**Figure 44 : Principe du Western-blot avec chimiluminescence.** **A.** Utilisation d'anticorps secondaires liés à des enzymes dans l'immunodétection de blots protéiques. L'anticorps primaire (de lapin par exemple) détecte d'abord la protéine d'intérêt sur le blot. L'anticorps anti-lapin couplé à la peroxydase du raifort détecte ensuite l'anticorps primaire. Enfin, l'ajout d'un substrat enzymatique entraîne le dépôt d'un produit coloré au site de la protéine d'intérêt sur le blot. **B.** Utilisation de la chimiluminescence pour détecter la peroxydase du raifort. Adapté de Wilson, K., & Walker, J. (2010)<sup>362</sup>.

Le blot est d'abord incubé dans une solution protéique, par exemple 10% (m/v) de sérum-albumine bovine ou 5% (m/v) de lait en poudre non gras (la technique dite du blocage), qui bloque tous les sites de liaisons hydrophobes restant sur la feuille nitrocellulose et empêche ainsi la liaison non spécifique d'anticorps à la membrane elle-même. Le blot est ensuite incubé dans une dilution d'un anticorps primaire dirigé contre la protéine d'intérêt. Cette molécule d'anticorps se lie au blot si elle détecte son antigène, identifiant ainsi la protéine d'intérêt.

Afin de visualiser cette interaction, le blot est incubé dans une solution d'un anticorps secondaire, qui est dirigé contre l'anticorps de l'espèce qui a fourni l'anticorps primaire. Par exemple, si l'anticorps primaire a été élevé chez un lapin, l'anticorps secondaire serait l'anticorps anti-lapin. Cet anticorps secondaire est étiqueté afin de permettre que l'interaction de l'anticorps secondaire avec l'anticorps primaire puisse être visualisée sur le blot. Les molécules d'anticorps anti-espèces sont facilement disponibles sur le marché, avec un choix de différentes étiquettes. L'une des méthodes de détection les plus courantes consiste à utiliser un anticorps secondaire lié à une enzyme. Dans ce cas, après traitement avec un anticorps secondaire lié par une enzyme, le blot est incubé dans une solution d'enzyme-substrat. L'enzyme transforme alors le substrat en un produit coloré insoluble qui est précipité sur la nitrocellulose. La présence d'une bande colorée indique donc la position de la protéine d'intérêt. Une enzyme utilisée généralement est la peroxydase de raifort, qui, avec le peroxyde d'hydrogène  $\text{H}_2\text{O}_2$  comme substrat, oxyde le 3-amino-9-éthylcarbazole en un produit brun insoluble ou le 4-chloro-1-naphthol en un produit bleu insoluble. Une autre méthode de détection de la peroxydase de raifort consiste à utiliser la méthode de la chimiluminescence accrue (ECL de l'anglais « *Enhanced chemiluminescence* »). En présence de  $\text{H}_2\text{O}_2$  et du substrat chimioluminescent luminol, la peroxydase de raifort oxyde le luminol avec production simultanée de lumière, dont l'intensité est augmentée de 1000 fois par la présence d'un activateur chimique. L'émission lumineuse peut être alors visualisée par un imageur à caméra CCD. C'est cette seconde méthode qui a été utilisée dans cette étude.

## 6.2. Western-blot avec le système GFPssrA, PAN et particule 20S

Les produits réactionnels des réactions principales des conditions SANS 1 et 2 issus des tests de dépliement de la GFPssrA de la partie 5.3 ont été analysés par Western-blot anti-GFP, à savoir :

- GFPssrA seule en présence de 100 mM ATP & 200 mM MgCl<sub>2</sub>
- GFPssrA et PAN en présence de 100 mM ATP & 200 mM MgCl<sub>2</sub>
- GFPssrA, PAN et particule 20S en présence de 100 mM ATP & 200 mM MgCl<sub>2</sub>

Les échantillons ont été dilués dans du tampon à 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub> afin de déposer 7.5 ng de GFPssrA pour chaque échantillon sur gel SDS-PAGE 15% acrylamide. La migration électrophorétique a été réalisée selon le protocole décrit partie 2.4.

Le gel a été ensuite incubé 10 minutes dans le tampon de transfert. En parallèle, une membrane de nitrocellulose découpée en proportions identiques au gel a été activée 10 secondes dans de l'éthanol, puis rincée 5 minutes dans de l'eau distillée avant d'être incubée 10 minutes dans du tampon de transfert. Les échantillons ont été transférés sur la membrane pendant 1 heure à 100 V dans du tampon de transfert. La membrane a ensuite été rincée 3 fois 5 minutes dans du TBS-Tween, avant d'être saturée pendant 1 heure par incubation à température ambiante dans du TBS-Tween contenant 5% (p/v) de lait en poudre écrémé (*Regilait*). Après rinçage 3 fois 5 minutes dans du TBS-Tween, la membrane a été incubée 1h à température ambiante dans du TBS-Tween contenant l'anticorps primaire anti-GFP (1/5000 – *abcam*<sup>®</sup>), puis de nouveau rincée 3 fois 5 minutes dans du TBS-Tween. La membrane a enfin été incubée 1 heure à température ambiante dans du TBS-Tween contenant l'anticorps secondaire anti-lapin conjugué à la HRP (1/2000 – *abcam*<sup>®</sup>) et rincée 3 fois 5 minutes dans du TBS-Tween avant de révéler le signal par trempage de la membrane dans les solutions de détection A (luminol) et B (peroxyde) dans un rapport de 1:1 (*Amersham*<sup>®</sup> *ECL*<sup>®</sup> – *GE Healthcare*). La membrane a été exposée environ 30 secondes et le signal a été visualisé grâce à l'aide d'un imageur à caméra CCD (*Kodak*). Le temps d'exposition est ajusté selon le contraste signal/bruit.

Tampon de transfert 10X : Trizma-base (30,3 g/L), glycine (144 g/L). Conserver à 4°C.

Tampon de transfert 1X : tampon de transfert 10X (10%), méthanol. Conserver à 4°C.

Tris Buffered Saline 5X (TBS 5X), pH7,6 : Trizma-base (12,1 g/L), chlorure de sodium (40,0 g/L). Ajuster le pH à 7,6 avec de l'acide chlorhydrique. Conserver à 4°C.

TBS-Tween : TBS 5X (20%), Tween<sup>™</sup> 20 (0,1%). Conserver à 4°C.

## 6.3. Western-blot avec le système GFPssrA\_Nter-streptavidine ou GFPssrA\_Cter-streptavidine, PAN et particule 20S

Les produits réactionnels des réactions principales des conditions SANS 5 et 6 provenant des tests de dépliement de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine de la partie 5.9.1 ont été analysés par Western-blot anti-GFP, soit pour chacune :

- GFPssrA-streptavidine seule avec 100 mM ATP & 200 mM MgCl<sub>2</sub>
- GFPssrA-streptavidine et PAN avec 100 mM ATP & 200 mM MgCl<sub>2</sub>
- GFPssrA-streptavidine, PAN et particule 20S avec 100 mM ATP & 200 mM MgCl<sub>2</sub>

Les échantillons ont été traités exactement comme décrit dans la partie précédente.



# Chapitre 4 :

## Matériel et méthodes biophysiques

### 1. Diffusion de rayons X aux petits angles

De nombreuses mesures de SAXS ont été réalisées à l'ESRF sur les différentes protéines dans le cadre de propositions BAG (de l'anglais « *Block Allocation Group* ») de l'IBS. J'ai pu effectuer l'acquisition et l'analyse des données de façon autonome.

#### 1.1. Protocole expérimental d'une collecte de données SAXS sur BM29

Lors de mesures SAXS, l'échantillon peut être directement chargé dans le capillaire (voir partie 2.4.2) et exposé aux rayons X (mode statique) ou bien chargé sur une colonne d'exclusion de taille, branchée en amont du capillaire (SEC-SAXS)<sup>363</sup>.

Dans le premier cas, les échantillons de protéine à mesurer ont été préparés avant l'expérience. Pour déterminer s'il y a des effets interparticulaires expérimentaux (agrégation/répulsion), une série de concentrations pour chaque protéine d'intérêt a été préparée dans le tampon de travail (au moins 3 concentrations de 1 mg/mL à 20 mg/mL). Une fois sur place, les échantillons ont été centrifugés pendant quelques minutes à 4°C pour éliminer d'éventuels agrégats et les concentrations ont été mesurées à nouveau. Chaque échantillon a été déposé dans un tube et placé dans le chargeur d'échantillon, qui en a prélevé 40 µL. Le tampon exact de la protéine a également été mesuré pour servir de blanc lors de l'analyse des données. Le temps d'exposition était de 1 seconde par image avec 10 images par échantillon et tampon. Les mesures ont été effectuées en mode écoulement où les échantillons sont poussés à travers le capillaire à un débit constant pour éviter les dommages d'irradiation.

Dans le second cas (SEC-SAXS), l'acquisition des données a été réalisée à la sortie de la colonne et une seule injection par protéine, de haute concentration, a été nécessaire. Le temps d'exposition était généralement de 4 secondes par image. Un avantage de cette méthode est la séparation des différentes espèces et agrégats en solution. Seul le pic d'intérêt est étudié, sans contamination, et le tampon de travail est directement utilisé comme blanc d'acquisition.

À l'ESRF, la collecte et la réduction des données brutes SAXS est effectuée automatiquement par le logiciel BsxCuBE, spécifique de la ligne BM29<sup>363</sup>. Les intensités de diffusion à une dimension des échantillons et des tampons sont alors exprimées en fonction du module du vecteur de diffusion  $q$ .

Les protéines mentionnées dans le tableau 8 ont été mesurées à plusieurs concentrations en mode statique. Les concentrations récapitulées dans le tableau correspondent aux deux concentrations utilisées lors de l'analyse des données pour générer la courbe chimère SAXS de chaque protéine en utilisant le logiciel PRIMUS (voir partie 1.2.3 du chapitre 2).

**Tableau 8 : Protéines mesurées par SAXS en mode statique.** Les courbes SAXS aux faibles concentrations et fortes concentrations ont été utilisées lors de l'analyse des données pour générer la courbe chimère SAXS de chaque protéine. Les courbes aux faibles concentrations ont été utilisées pour la région de Guinier et les courbes aux fortes concentrations pour les plus grands angles. La courbe chimère a été réalisée en fusionnant les deux courbes de concentrations sur une région couvrant quelques dizaines à une centaine de points, juste après la région de Guinier. Pour la streptavidine monovalente, la courbe SAXS à 1 mg/mL a été utilisée telle quelle.

	Faible concentration (mg/mL)	Forte concentration (mg/mL)
GFPssrA	2,8	10,1
GFPssrA_Nter-biotine	0,9	9,0
GFPssrA_Cter-biotine	0,8	9,2
Streptavidine monovalente	1,0	1,0
GFPssrA_Nter-streptavidine	1,2	20
GFPssrA_Cter-streptavidine	1,0	10,1
Particule 20S	1,0	6,7

La protéine PAN a été mesurée en mode statique à 0,6, 1, 1,5, 2,9, 6, 7,8, 11,5 et 16,2 mg/mL dans le tampon de travail afin de suivre l'évolution du rayon de giration  $R_g$  et de l'intensité à angle nul  $I(0)$  en fonction de la concentration en protéine. Les valeurs de  $R_g$  et  $I(0)$  ont été obtenues en utilisant le logiciel PRIMUS (voir partie 1.2.3).

La protéine PAN a également été mesurée en SEC-SAXS. Pour cela, une solution protéique de 300  $\mu$ L à 9,7 mg/mL préparée dans le tampon de travail a été injectée sur la colonne Superose 6 Increase (*GE Healthcare*) préalablement équilibrée dans le même tampon. Les données collectées par le détecteur UV et par SAXS lors de l'élution de la protéine dans le tampon de travail ont été analysées à l'aide du logiciel CHROMIXS (voir partie 1.2.3). La courbe d'absorbance à 280 nm a été normalisée par rapport à la valeur maximale du pic d'élution et tracée en fonction du volume d'élution et le  $R_g$  correspondant a été représenté par une ligne en fonction du volume d'élution. La courbe SAXS expérimentale de PAN a été générée à partir du profil SEC-SAXS par le même logiciel (moyenne des points de mesure du volume d'élution de 13,5 à 14,5 mL) et l'analyse de cette courbe a été réalisée grâce au logiciel PRIMUS.

**Tableau 9 : Logiciel et fichier PDB utilisés pour chaque protéine étudiée.**

	Logiciel utilisé	Fichier PDB utilisé
GFPssrA	CRYSOL	1EMB
GFPssrA_Nter-biotine	CRYSOL	1EMB
GFPssrA_Cter-biotine	CRYSOL	1EMB
Streptavidine monovalente	CRYSOL	3RY1
GFPssrA_Nter-streptavidine	SASREF	1EMB & 3RY1
GFPssrA_Cter-streptavidine	SASREF	1EMB & 3RY1
PAN	CRYSOL	6HEA
Particule 20S	CRYSOL	6HEA

La courbe SAXS théorique de chaque protéine a été générée par le logiciel CRY SOL ou SASREF et comparée avec la courbe expérimentale (voir partie 1.2.3). Le tableau 9 récapitule le logiciel et le fichier PDB utilisés pour chaque protéine étudiée.

Tampon de travail : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

## 2. Diffusion de neutrons aux petits angles résolue en temps (TR-SANS) et combinaison avec la spectroscopie de fluorescence en ligne

Les mesures de TR-SANS (de l'anglais « *Time-Resolved Small Angle Neutron Scattering* ») ont été réalisées en octobre 2018 et août 2019. J'ai pu effectuer l'acquisition et l'analyse des données avec l'assistance d'Anne Martel (responsable de l'instrument D22), Nico Carl (pour les scripts Python) et Frank Gabel (directeur de thèse).

### 2.1. Instrumentation et protocole expérimental

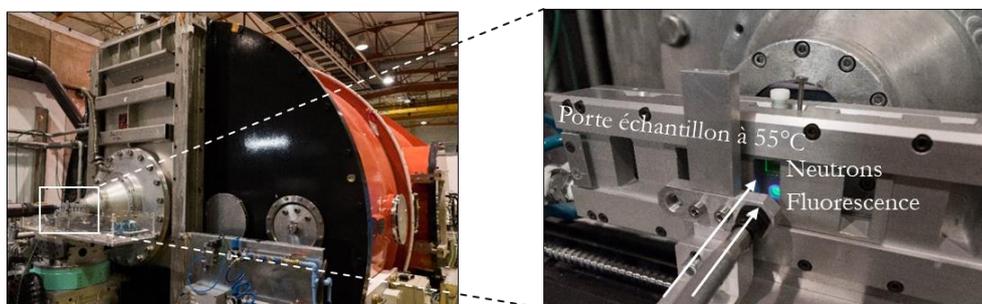


Figure 45 : Photos du dispositif expérimental sur la ligne D22 à l'ILL.

Tableau 10 : Liste des échantillons mesurés en TR-SANS sur le système GFPssrA, protéasome PAN-20S. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O.

Échantillon	Contenu en protéines	Contenu en ATP & MgCl <sub>2</sub>
1	d-GFPssrA isolée	100 mM ATP & 200 mM MgCl <sub>2</sub>
2	d-GFPssrA et h-PAN	100 mM ATP & 200 mM MgCl <sub>2</sub>
3	d-GFPssrA, h-PAN et particule h-20S	100 mM ATP & 200 mM MgCl <sub>2</sub>
4	d-GFPssrA et particule h-20S	100 mM ATP & 200 mM MgCl <sub>2</sub>
5	d-GFPssrA et h-PAN	0 mM ATP & 200 mM MgCl <sub>2</sub>
6	d-GFPssrA, h-PAN et particule h-20S	0 mM ATP & 200 mM MgCl <sub>2</sub>

Tableau 11 : Liste des échantillons mesurés en TR-SANS sur le système GFPssrA-streptavidine, protéasome PAN-20S. Tous les échantillons contenaient 100 mM ATP et 200 mM MgCl<sub>2</sub>. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O.

Échantillon	Contenu en protéines
1	d-GFPssrA_Nter-biotine isolée
2	d-GFPssrA_Cter-biotine isolée
3	d-GFPssrA_Nter-streptavidine isolée
4	d-GFPssrA_Cter-streptavidine isolée

5	d-GFPssrA_Nter-biotine et h-PAN
6	d-GFPssrA_Cter-biotine et h-PAN
7	d-GFPssrA_Nter-streptavidine et h-PAN
8	d-GFPssrA_Cter-streptavidine et h-PAN
9	d-GFPssrA_Nter-biotine, h-PAN et particule h-20S
10	d-GFPssrA_Cter-biotine, h-PAN et particule h-20S
11	d-GFPssrA_Nter-streptavidine, h-PAN et particule h-20S
12	d-GFPssrA_Cter-streptavidine, h-PAN et particule h-20S

**Tableau 12 : Composition de la condition SANS 1.** Composition de l'échantillon d-GFPssrA, h-PAN et particule h-20S en condition SANS 1. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O. Pour les échantillons sans un ou plusieurs composants, ces composants ont été remplacés par du tampon.

Protéine / composant (h: hydrogéné, d: deutéré)	Concentration initiale	Concentration finale	Volume
d-GFPssrA	1,8 mM (50 mg/mL)	72 µM (2 mg/mL)	12,4 µL
h-PAN	67 µM (20 mg/mL)	20 µM (6 mg/mL)	93 µL
h-20S	58 µM (45 mg/mL)	20 µM (16 mg/mL)	111,8 µL
ATP	455 mM	100 mM	68 µL
MgCl <sub>2</sub>	2,5 M	200 mM	24,8 µL

**Tableau 13 : Composition de la condition SANS 2.** Composition de l'échantillon d-GFPssrA, h-PAN et particule h-20S en condition SANS 2. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O. Pour les échantillons sans un ou plusieurs composants, ces composants ont été remplacés par du tampon.

Protéine / composant (h: hydrogéné, d: deutéré)	Concentration initiale	Concentration finale	Volume
d-GFPssrA	1,8 mM (50 mg/mL)	180 µM (5 mg/mL)	31 µL
h-PAN	67 µM (20 mg/mL)	5 µM (1,5 mg/mL)	23,2 µL
h-20S	58 µM (45 mg/mL)	30 µM (23,5 mg/mL)	163 µL
ATP	455 mM	100 mM	68 µL
MgCl <sub>2</sub>	2,5 M	200 mM	24,8 µL

**Tableau 14 : Composition des conditions SANS 3 et 4.** Composition des échantillons d-GFPssrA-biotine, h-PAN et particule h-20S en condition SANS 3 et 4. La condition SANS 3 contenait de la d-GFPssrA\_Nter-biotine et la condition SANS 4 de la d-GFPssrA\_Cter-biotine. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O. Pour les échantillons sans une ou plusieurs protéines, ces protéines ont été remplacées par du tampon.

Protéine / composant (h: hydrogéné, d: deutéré)	Concentration initiale	Concentration finale	Volume
d-GFPssrA-h-biotine	896 µM (25 mg/mL)	25 µM (0,7 mg/mL)	8,7 µL
h-PAN	107 µM (32 mg/mL)	26 µM (7,8 mg/mL)	75,6 µL

h-20S	58 $\mu$ M (45 mg/mL)	26 $\mu$ M (20 mg/mL)	138 $\mu$ L
ATP	500 mM	100 mM	62 $\mu$ L
MgCl <sub>2</sub>	3 M	200 mM	20,7 $\mu$ L
Tampon	/	/	5,1 $\mu$ L

**Tableau 15 : Composition des conditions SANS 5 et 6.** Composition de l'échantillon d-GFPssrA-streptavidine, h-PAN et particule h-20S en condition SANS 5 et 6. La condition SANS 3 contenait de la d-GFPssrA\_Nter-streptavidine et la condition SANS 4 de la d-GFPssrA\_Cter-streptavidine. Composition du tampon : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> dans 42% D<sub>2</sub>O. Pour les échantillons sans une ou plusieurs protéines, ces protéines ont été remplacées par du tampon.

Protéine / composant (h: hydrogéné, d: deutéré)	Concentration initiale	Concentration finale	Volume
d-GFPssrA-h-biotine	550 $\mu$ M (45 mg/mL)	24 $\mu$ M (2 mg/mL)	13,8 $\mu$ L
h-PAN	107 $\mu$ M (32 mg/mL)	26 $\mu$ M (7,8 mg/mL)	75,6 $\mu$ L
h-20S	58 $\mu$ M (45 mg/mL)	26 $\mu$ M (20 mg/mL)	138 $\mu$ L
ATP	500 mM	100 mM	62 $\mu$ L
MgCl <sub>2</sub>	3 M	200 mM	20,7 $\mu$ L

Toutes les expériences de TR-SANS ont été effectuées sur la ligne D22 à l'ILL (figure 45). Les échantillons listés dans le tableau 10 et dans le tableau 11 ont été préparés dans un tampon à 42% de D<sub>2</sub>O contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub> permettant de masquer le signal des protéines hydrogénées.

Chaque mélange réactionnel a été préparé juste avant sa mesure dans un volume final de 310  $\mu$ L selon les conditions décrites dans le tableau 12 pour la condition SANS 1, tableau 13 pour la condition SANS 2, tableau 14 pour les conditions SANS 3 et 4, et tableau 15 pour les conditions SANS 5 et 6. L'échantillon a ensuite été immédiatement injecté dans une cuvette de quartz de 1 mm de trajet optique et placé dans le support ajusté à 55°C. Tous les échantillons ont été mesurés avec une distance collimateur/détecteur de 2,0/2,8 m. Pour les réactions d-GFPssrA et h-PAN contenant 100 mM ATP & 200 mM MgCl<sub>2</sub>, une mesure indépendante supplémentaire a été effectuée avec une distance collimateur/détecteur de 8,0/8,0 m afin d'accéder à plus d'informations structurales sur les agrégats de d-GFPssrA.

La réaction de dépliement et de dégradation de la d-GFPssrA par le système protéolytique étant déclenchée immédiatement une fois la cuvette placée dans le support tempéré, la collecte des données a été démarrée le plus rapidement possible (une vingtaine de secondes de délai entre le positionnement de la cuvette dans le portoir et le début de la collecte). Dans les conditions SANS 1 et 2, chaque courbe SANS a été issue de 30 secondes de temps d'exposition, avec un total de 90 courbes de diffusion par échantillon pour 45 minutes de réaction. Dans les conditions SANS 3 à 6, chaque courbe SANS obtenue à partir de 20 secondes de temps d'exposition, avec un total de 90 courbes de diffusion par échantillon pour 30 minutes de réaction.

Pour recueillir un ensemble complet de données, plusieurs mesures de références ont été effectuées lors de chaque session de collecte de données dans les mêmes configurations collimateur/détecteur. En plus de la diffusion de l'échantillon dans sa cuve, des mesures ont été réalisées sur la cuve

d'échantillon vide, le tampon avec et sans ATP & MgCl<sub>2</sub>, le bruit de fond indépendant de l'échantillon, et sur une mesure de la réponse ou du rendement du détecteur. Le bruit de fond indépendant de l'échantillon comprend le bruit électronique dans les circuits de détection et de comptage ainsi que la détection de neutrons ne passant pas à travers l'échantillon (fond présent dans le hall expérimental). Ceci est mesuré en remplaçant l'échantillon par une feuille de cadmium, absorbant fortement les neutrons. Les transmissions ont été mesurées pour chaque échantillon et tampon ainsi que pour la cuvette vide et le faisceau direct. La transmission peut être définie comme la partie du faisceau incident qui traverse l'échantillon sans diffusion ni absorption. Elle est mesurée en atténuant le faisceau primaire (afin que le faisceau direct puisse être mesuré directement sans saturation ni détérioration du détecteur) et en calculant le rapport des intensités avec et sans l'échantillon dans le faisceau. À l'ILL, les mesures brutes SANS sont réduites par les expérimentateurs en utilisant des logiciels tels que GRASP<sup>364</sup> ou LAMP<sup>365</sup>, développés au même institut.

Simultanément aux mesures cinétiques SANS, des mesures de fluorescence en ligne ont été effectuées sur l'instrument D22. Pour cela, un spectrofluorimètre *in situ* a été mis en place en collaboration avec Anne Martel, la responsable de l'instrument D22<sup>40</sup>. La lumière provenant d'une lampe à deutérium circulant dans une fibre optique passe d'abord à travers un filtre sélectionnant la longueur d'onde d'excitation correspondant à la gamme d'absorption du chromophore de la GFPssrA (400 nm). Une partie seulement de la fluorescence émise par la GFPssrA traverse à nouveau la fibre optique. Un deuxième filtre a été placé avant le détecteur afin de sélectionner précisément la longueur d'onde d'émission à analyser (509 nm). L'intensité du signal émis a été mesurée au niveau d'un détecteur qui convertit l'énergie des photons de fluorescence en un signal numérique. La fibre optique a été installée au contact de la cuvette contenant le mélange réactionnel. Cet environnement expérimental permet de suivre le processus de dépliement et de dégradation en temps réel avec des données structurales et de les corrélérer avec les pertes de la fluorescence de la GFPssrA dues à son dépliement.

## 2.2. Analyse des données

La réduction des données et l'intégration radiale sur une image bidimensionnelle ont été réalisées à l'aide du logiciel LAMP. Pour chaque échantillon des conditions SANS 1 et 2, les données ont été réduites pour chaque courbe issue de 30 secondes de temps d'exposition, et également sous forme moyennée de 4 courbes consécutives afin d'obtenir des courbes issues de 2 minutes de temps d'exposition. Dans le cas des conditions SANS 3 à 6, les données ont été réduites pour chaque courbe issue de 20 secondes de temps d'exposition, et aussi sous forme moyennée de 3 courbes consécutives pour obtenir des courbes issues de 90 secondes de temps d'exposition. Les courbes de diffusion des tampons ont été soustraites des courbes d'échantillons respectives grâce à un script Python (Nico Carl).

Les rayons de giration  $R_g$  et les intensités diffusées à angle nul  $I(0)$  ont été extraits à l'aide de l'approximation de Guinier (équation 13) pour chaque courbe expérimentale grâce à un second script Python (Nico Carl). La validité de l'approximation de Guinier ( $qR_g \leq 1,3$ ) a été respectée dans tous les cas à l'exception des échantillons d-GFPssrA en présence de h-PAN où des agrégats ont été observés. Dans ce cas, les trois points de données les plus bas ont été utilisés pour extraire une estimation inférieure de  $R_g$  et  $I(0)$ .

Les courbes d'évolution de  $I(0)$  et du  $R_g$  en fonction du temps des échantillons contenant de la d-GFPssrA en présence de h-PAN et de la particule h-20S des conditions SANS 1 et 2 ont été comparées respectivement à une courbe bi-exponentielle (équation 31) et linéaire (équation 32) en prenant en compte les barres d'erreur.

$$f(t) = y_0 + A_1 e^{-\frac{t}{t_1}} + A_2 e^{-\frac{t}{t_2}}$$

**Équation 31**

Où  $y_0$  est l'intensité initiale à  $t = 0$  minutes,  $A_1$  et  $A_2$  sont les amplitudes de diminution exponentielle avec des temps caractéristiques respectif  $t_1$  et  $t_2$ . La même fonction a également été utilisée les courbes d'hydrolyse de l'ATP en fonction du temps (partie 4.3).

$$f(x) = ax + b$$

**Équation 32**

Où  $a$  est la pente et  $b$  est l'intensité initiale à  $t = 0$  minutes.

Pour les conditions SANS 1 et 2, la courbe SANS théorique la GFP a été générée par le logiciel CRYSON à partir du fichier PDB 1EMB (avec une partie flexible ajoutée à l'extrémité C-terminale) et comparée à la première courbe expérimentale ( $t = 0,5$  minutes) (voir partie 1.2.3).

Tous les spectres ont ensuite été extraits par le programme h5py pour Python (<http://www.h5py.org>) et ajustés à fonction gaussienne afin de déterminer l'intensité de fluorescence à 509 nm. L'erreur de fluorescence a été estimée à partir de l'écart-type des paramètres d'ajustement. Les courbes d'évolution de l'intensité de fluorescence en fonction du temps des échantillons contenant de la d-GFPssrA en présence de h-PAN et de particule h-20S des conditions SANS 1 et 2 ont été comparées respectivement à une courbe bi-exponentielle (équation 31) et exponentielle simple (équation 33) en prenant en compte les barres d'erreur.

$$f(t) = y_0 + A_1 e^{-\frac{t}{t_1}}$$

**Équation 33**

### 3. Diffusion de la lumière

La diffusion de la lumière est une technique d'analyse spectroscopique non destructive permettant d'obtenir des informations sur les propriétés des macromolécules, notamment la masse moléculaire, les dimensions et les coefficients de diffusion, ainsi que les propriétés d'interaction des protéines.

Les techniques de diffusion de la lumière utilisent généralement des lasers à haute monochromaticité, à focale étroite et à forte intensité. Lorsque la lumière incidente frappe une macromolécule, la lumière diffusée possède en majorité la même longueur d'onde que la lumière incidente (diffusion élastique de la lumière) tandis que le reste de la lumière diffusée possède une longueur d'onde supérieure ou inférieure à celle de la lumière incidente (diffusion non élastique de la lumière).

### 3.1. Diffusion statique (ou élastique) de la lumière (SLS)

La diffusion statique de la lumière est également connue sous le nom de diffusion de Rayleigh et consiste à mesurer l'intensité de la lumière diffusée par une solution à différents angles par rapport au faisceau laser incident. Cette technique est très analogue au SAXS et SANS et utilise un autre type de rayonnement et une autre longueur d'onde. Elle est donc susceptible à des dimensions des particules différentes.

L'intensité de diffusion des macromolécules est proportionnelle à la masse moléculaire  $M_w$  au carré, et donc idéale pour la détermination de ce paramètre, puisque l'apport de petites molécules de solvant peut être négligé.

Dans une solution idéale, les macromolécules sont entièrement indépendantes les unes des autres et la diffusion de la lumière peut être décrite de la façon suivante :

$$\frac{I_\theta}{I_0} = R_\theta = P(q)KcM_w$$

Équation 34

où  $I_\theta$  est l'intensité de la lumière diffusée à l'angle  $\theta$ ,  $I_0$  est l'intensité de la lumière incidente,  $K$  est une constante proportionnelle au carré de l'incrément d'indice de réfraction,  $c$  est la concentration et  $R_\theta$  le rapport Rayleigh.  $P(q)$  décrit la dépendance angulaire de la lumière diffusée.

Pour les solutions non idéales, il est possible de quantifier les interactions moléculaires comme second coefficient viral  $A_2$ , qui est une mesure de la force des interactions entre les molécules, grâce à la dépendance à la concentration de l'intensité diffusée<sup>366</sup> :

$$\frac{Kc}{R_\theta} = \frac{1}{P(q)} \left( \frac{1}{M_w} + 2A_2c \right)$$

Équation 35

Pour déterminer la masse moléculaire à partir de la diffusion de la lumière, trois paramètres doivent être mesurés : l'intensité de la lumière diffusée sous différents angles, la concentration de la protéine et l'augmentation de l'indice de réfraction spécifique du solvant. L'instrumentation minimale requise est une source lumineuse (laser), un détecteur à diffusion de lumière laser multi-angles (MALLS), ainsi qu'un détecteur d'indice de réfraction. Ces instruments peuvent être utilisés sur des échantillons en solution en mode statique, mais peuvent également être connectés à une colonne de chromatographie liquide à haute performance pour permettre la détermination en ligne de la masse moléculaire des macromolécules éluées. La chromatographie utilisée est alors la chromatographie d'exclusion de taille (SEC), et la combinaison de ces méthodes est connue sous le nom de SEC-MALLS. Contrairement à la chromatographie d'exclusion de taille conventionnelle, la détermination de la masse moléculaire à partir du MALLS est indépendante du volume d'éluion de la macromolécule. C'est un avantage précieux, car le temps de rétention d'une macromolécule sur la colonne d'exclusion de taille peut dépendre de sa forme.

L'indice de polydispersité de la solution (ou au sein d'un pic d'éluion dans le cas du SEC-MALLS) peut être déterminée par le rapport de la masse moléculaire  $M_w$  sur la masse moléculaire moyenne

$M_n$ . La masse moléculaire moyenne  $M_n$  est la masse moléculaire moyenne statistique de toutes les protéines de l'échantillon, et est définie comme :

$$M_n = \frac{\sum N_i M_i}{\sum N_i}$$

Équation 36

où  $M_i$  est la masse moléculaire d'une particule et  $N_i$  est le nombre de particules de cette masse moléculaire.  $M_n$  peut être mesurée par des méthodes qui déterminent le nombre de molécules dans un échantillon d'un poids donné.

La masse moléculaire moyenne en nombre  $M_w$  est quant à elle définie par :

$$M_w = \frac{\sum N_i M_i^2}{\sum N_i M_i}$$

Équation 37

Par rapport à  $M_n$ ,  $M_w$  prend plus en compte la masse moléculaire d'une particule dans la détermination des contributions à la moyenne de la masse moléculaire. Plus la particule est massive, plus elle contribue à  $M_w$ .  $M_w$  est déterminée par des méthodes sensibles à la taille moléculaire plutôt qu'à leur nombre, telles les techniques de diffusion de la lumière et des rayons X et neutrons aux petits angles.

Un échantillon monodisperse ne contenant qu'un seul type de molécule d'une masse moléculaire définie, la masse moyenne est indépendante de la méthode de calcul de la moyenne. L'indice de polydispersité  $M_w/M_n$  sera alors de 1. Si l'échantillon contient un mélange d'espèces de masses moléculaires différentes, la masse moyenne dépendra de la méthode du calcul de la moyenne et la polydispersité sera différente de 1.

### 3.2. Diffusion dynamique (ou quasi-élastique) de la lumière (DLS)

L'intensité de la lumière diffusée fluctue au cours du temps. Cette fluctuation est due aux petites molécules en mouvement brownien dans les solutions. La lumière diffusée subit alors de l'interférence avec la lumière diffusée par la même molécule au cours du temps (fonction d'autocorrélation ci-après). Cette fluctuation d'intensité contient des informations sur l'échelle de temps du mouvement des diffuseurs. La diffusion dynamique de la lumière mesure la distribution de longueur d'onde de la lumière diffusée par les particules au cours du temps. Cette technique permet la détermination des coefficients de diffusion et l'évaluation de l'aggrégation des protéines.

La fonction d'autocorrélation est mesurée dans un premier temps et permet de comparer le signal mesuré à lui-même, avec un petit décalage temporel. Une modélisation graphique permet d'extraire un temps caractéristique de décroissance de cette fonction. Si l'autocorrélation décroît rapidement (taux faible), c'est que le signal mesuré varie rapidement. C'est le cas lorsque les particules se déplacent assez vite, donc qu'elles sont de petites tailles (plus mobiles). L'inverse de ce temps

caractéristique (dit temps de relaxation) est lié au coefficient de diffusion des particules par la relation suivante :

$$\frac{1}{\tau} = Dq^2$$

Équation 38

avec  $D$  le coefficient de diffusion,  $q$  le vecteur d'onde et  $\tau$  le temps de relaxation.

Le coefficient de diffusion est lié à la taille des particules par une équation connue sous le nom de relation de Stokes-Einstein :

$$R_H = \frac{k_B T}{6\pi\eta_S D}$$

Équation 39

où  $R_H$  désigne le rayon hydrodynamique,  $k_B$  la constante de Boltzmann,  $T$  la température et  $\eta_S$  la viscosité du solvant.

Le paramètre d'intérêt est le rayon hydrodynamique, ou rayon de Stokes, qui est la taille d'une particule sphérique qui aurait le même coefficient de diffusion dans une solution ayant la même viscosité. Cependant, le rayon hydrodynamique décrit une particule idéalisée et peut différer significativement de la taille physique réelle d'une macromolécule. En effet, la plupart des protéines ne sont pas sphériques et leur rayon hydrodynamique dépend donc de leur forme.

### 3.3. Protocole expérimental des mesures SLS-DLS réalisées

Les analyses SLS-DLS (de l'anglais « *Static Light Scattering-Dynamic Light Scattering* ») ont été réalisées grâce à la plateforme PSCM géré conjointement par l'ESRF et l'ILL. J'ai réalisé les mesures de façon autonome tandis que l'analyse des données a été réalisée en ayant recours au service de la plateforme (Nico Carl).

Quatre conditions de tampon de la protéine PAN ont été testées par SLS-DLS :

- 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub> en présence de 100 mM ATP
- 20 mM Tris-HCl pH 7,5, 500 mM NaCl et 10 mM MgCl<sub>2</sub> en présence de 100 mM ATP
- 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub>
- 20 mM Tris-HCl pH 7,5, 500 mM NaCl et 10 mM MgCl<sub>2</sub>

Pour chaque condition expérimentale, une gamme de concentration de PAN a été préparée dans le tampon correspondant pour réaliser un point de mesure à 0,5, 0,75, 1 et 1,5 mg/mL. Les solutions protéiques et leurs tampons respectifs ont ensuite été filtrés (MWCO 0,1 μM) pour éliminer les poussières et les artefacts de la solution. Les mesures ont été effectuées à 55°C.

Les mesures ont été effectuées avec un goniomètre (*ALV CGS-3*) équipé d'un laser HeNe (632,8 nm) à des angles de diffusion allant de 30° à 150° ( $q = 6,8 \times 10^{-4} \text{ \AA}^{-1} - 2,6 \times 10^{-3} \text{ \AA}^{-1}$ ). Du toluène a été utilisé pour l'étalonnage de l'intensité absolue. Le toluène, les tampons et les échantillons ont été mesurés dans des cellules de quartz cylindriques de 1 cm. La viscosité et la réfraction de l'eau à 55°C ont été utilisées pour tous les échantillons et un incrément d'indice de réfraction de  $dn/dc = 0,184 \text{ mL/g}$  a été estimé à partir de la séquence des acides aminés<sup>367</sup>. Les données DLS ont été

analysées à l'aide de la méthode des cumulants<sup>368</sup>. Les données SLS ont été analysées à l'aide de l'équation de Zimm (équation 35) et extrapolées à la concentration zéro<sup>366</sup>.

### 3.4. Protocole expérimental des mesures SEC-MALLS réalisées

De nombreuses mesures de SEC-MALLS (plateforme ISBG à l'IBS, Caroline Mas) ont été effectuées sur les différentes protéines étudiées au cours de cette thèse. J'ai pu effectuer l'acquisition et l'analyse des données de façon autonome.

#### 3.4.1. Mesures sur la protéine PAN à différentes concentrations d'ATP

La protéine PAN a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis incubée 30 minutes à 55°C dans le tampon de travail supplémenté par 0, 5, 10, 25, 50, 60, 75 ou 100 mM ATP avec une concentration de 1,5 mg/mL PAN en solution.

Un volume 50 µL de chaque échantillon a ensuite été injecté dans une colonne Superose 6 10/300 GL Increase (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. Les données collectées par le détecteur UV (*Hitachi Elite LaChrom UV detector L-2400*), le réfractomètre (*Wyatt Optilab T-rex*) et l'instrument MALLS (*Wyatt Dawn Heleos II*) lors de l'élution de la protéine dans le tampon de travail ont été analysées à l'aide du logiciel ASTRA 6. Les courbes d'indice de réfraction, d'absorbance à 280 nm et de diffusion de la lumière ont été normalisées par rapport à la valeur maximale du pic d'élution et tracées en fonction du volume d'élution et la masse moléculaire en kDa correspondant aux différents pics d'élution a été représentée par une ligne en fonction du volume d'élution.

Tampon de travail : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

#### 3.4.2. Mesures sur la particule 20S

La particule 20S a été préparée à 1 mg/mL puis centrifugée à 10000 x g à 4°C pendant 10 minutes. Un volume 50 µL a ensuite été injecté dans une colonne Superose 6 10/300 GL Increase (*GE Healthcare*) préalablement incubée dans le tampon de travail. Les données issues de la mesure ont été analysées de la même manière que dans la partie précédente.

#### 3.4.3. Mesures sur la GFPssrA\_Nter-streptavidine et la GFPssrA\_Cter-streptavidine

La GFPssrA\_Nter-biotine et la GFPssrA\_Cter-biotine ont été préparées à 6 mg/mL, tandis que la GFPssrA\_G10C-streptavidine et la GFPssrA\_Cter-streptavidine ont été préparées à 4 mg/mL. Les protéines ont été centrifugées à 10000 x g à 4°C pendant 10 minutes. Un volume 50 µL de chaque échantillon a ensuite été injecté dans une colonne Superose 12 10/300 GL (*GE Healthcare*) préalablement incubée dans le tampon de travail. Les données issues des mesures ont été analysées de la même manière que dans les parties précédentes.

## 4. Spectrométrie de masse

Des analyses en spectrométrie de masse ont été effectuées au cours de cette thèse en ayant recours aux services de la plateforme ISBG à l'IBS (Luca Signor) pour l'analyse des échantillons et le traitement des données, bien que j'aie participé à certaines mesures et à l'analyse des données.

## 4.1. Introduction à la spectrométrie de masse

La spectrométrie de masse (*MS* de l'anglais « *Mass Spectrometry* ») est un outil analytique utilisé pour déterminer les masses de différents composés dans un échantillon. Cette technique est utilisée pour identifier les composés d'un mélange, ce qui peut inclure la détection d'impuretés dans un échantillon ou la détermination du mélange de protéines dans un échantillon. Lors de cette technique, les molécules d'un échantillon sont converties en ions gazeux qui sont ensuite séparés dans un spectromètre de masse en fonction de leur rapport masse/charge ( $m/z$ ) et détectés. Le spectre de masse est un graphique de l'abondance relative des ions à chaque rapport  $m/z$ .

La spectrométrie de masse ne mesure que la masse des particules chargées et les composés neutres sont invisibles. Toutes les méthodes de MS utilisent 3 étapes fondamentales se déroulant sous vide : l'ionisation (source d'ions), la séparation et isolation des ions (analyseur de masse) et la détection (détecteur d'ions).

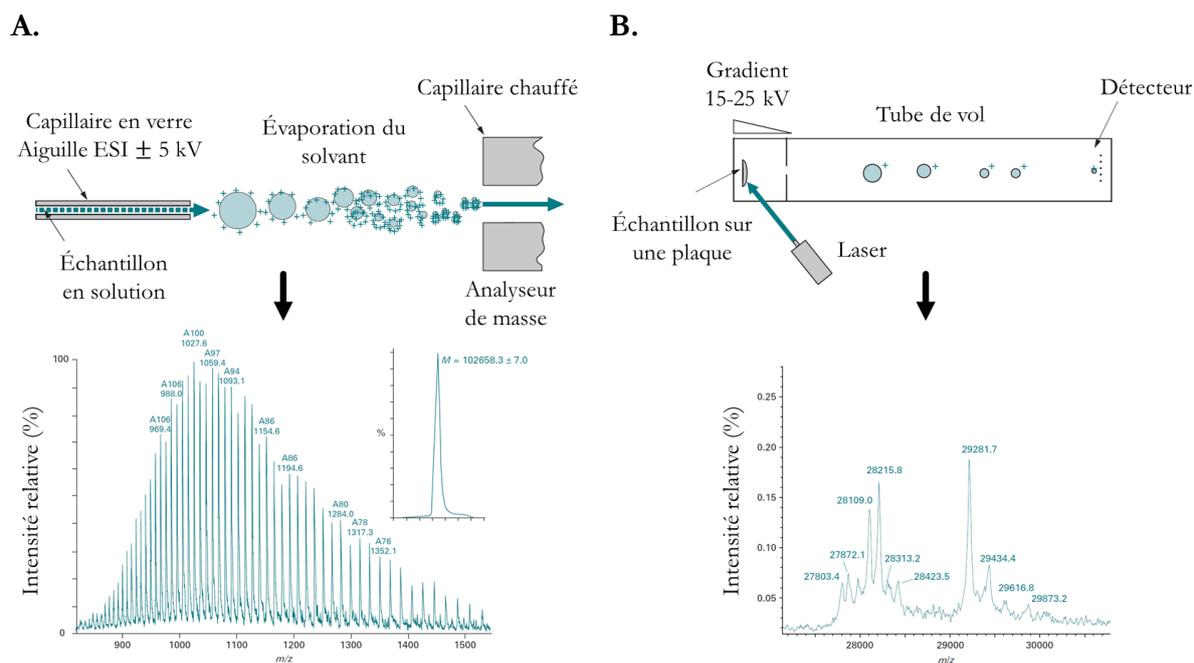
- L'ionisation : l'échantillon est vaporisé et ionisé. Si l'échantillon ne peut pas être vaporisé sans se décomposer, il ne peut pas être analysé par spectrométrie de masse. L'ionisation par électrospray (*ESI* de l'anglais *ElectroSpray Ionization*) et l'ionisation par désorption laser assistée par matrice (*MALDI* de l'anglais *Matrix-Assisted Laser Desorption Ionization*) permettent d'ioniser des biomolécules sans les détruire. Les ions de l'analyte sont appelés "ions moléculaires".
- La séparation et isolation des ions : une fois que les ions quittent la source d'ions, ils passent dans un analyseur de masse, dont la fonction est de séparer les ions et de mesurer leur rapport  $m/z$ . Deux techniques sont communément utilisées, faisant intervenir le temps de vol (*TOF* de l'anglais *Time Of Flight*) ou un analyseur quadripolaire (Q). Dans le cas d'un analyseur TOF, les ions sont analysés en fonction du temps qu'ils mettent à traverser un tube libre de champ et de longueur variable. Le temps de vol est directement proportionnel au rapport  $m/z$  des ions. Les analyseurs quadripolaires sont de leur côté constitués de quatre électrodes identiques, couplées deux à deux, permettant de séparer les ions en fonction de leur rapport  $m/z$  et ne laisser passer qu'un type d'ions à la fois.

Dans la plupart des instruments, un type particulier d'ionisation est couplé à un analyseur de masse particulier. L'ESI est couplé à un analyseur quadripolaire (limite  $m/z$  de 3000) et le MALDI est couplé à la détection TOF (limite  $m/z$  de 500000). Sur les spectromètres de dernière génération, ces deux techniques sont couplées (Q-TOF).

- La détection : Les ions séparés par l'analyseur de masse frappent le détecteur d'ions, qui enregistre combien d'ions le frappent pour chaque rapport  $m/z$  spécifique. Généralement, le résultat est un diagramme à barres verticales dont l'axe des abscisses représente l'augmentation en  $m/z$  et dont l'axe des ordonnées représente l'abondance relative.

## 4.2. MALDI-MS et ESI-MS

La source d'ionisation est la composante critique de toute méthode de spectrométrie de masse, si bien que les deux techniques d'ionisation les plus populaires en biologie, l'ESI et le MALDI, ont valu à leurs inventeurs le prix Nobel de Chimie en 2002. Ces deux techniques ont révolutionné la spectrométrie de masse biologique parce qu'elles sont douces, c'est à dire elles ionisent les biomolécules sans les décomposer. La figure 46 résume le principe des deux techniques.



**Figure 46 : Principe de l'ESI-MS et du MALDI-TOF-MS. A.** Le spectre ESI d'une protéine de grande masse moléculaire (gauche) est constitué des espèces d'ions annotées par l'état de charge, par exemple avec 99, 100, 101 charges, etc. avec la valeur  $m/z$  associée. L'encadré montre le « spectre déconvolué ». La multiplication des ions chargés entraîne des pics multiples qui rendraient plus difficile l'interprétation des masses de mélanges mais permet l'analyse de masse protéique avec précision. **B.** Le spectre MALDI-TOF d'un mélange de protéines (droite) est presque exclusivement constitué d'ions à charge unique représentant les espèces d'ions moléculaires des protéines constitutives. Adapté de Wilson, K., & Walker, J. (2010)<sup>362</sup>.

L'ESI permet la production d'ions en phase gazeuse à partir d'un échantillon en solution (solvant volatil). L'échantillon, préalablement passé sur une colonne de chromatographie en phase inverse, est amené jusqu'à la source. L'application d'un champ électrique intense ( $10^6 \text{ V.m}^{-1}$ ) entre l'extrémité du capillaire et une contre-électrode, provoque alors, sous flux d'azote à haute température, l'évaporation du solvant et la formation d'ions en phase vapeur. Les molécules les plus petites produisent généralement des ions à charge unique, mais les ions à charges multiples sont souvent formés à partir de biomolécules plus grosses, contrairement au MALDI, ce qui donne des rapports  $m/z$  suffisamment petits pour l'analyseur quadripolaire. Ainsi, des masses de larges protéines peuvent être mesurées avec précision dans l'ESI-MS bien que la limite  $m/z$  de mesure, soit de 3000.

Dans le MALDI, un échantillon est mélangé avec un matériau de matrice cristallin absorbant les UV, et déposé sur une plaque cible métallique, où il cocrystallise au séchage. La plaque est ensuite insérée dans l'instrument MS, où elle est placée sous vide et frappée avec un laser UV. Des impulsions de lumière laser d'une durée de quelques nanosecondes provoquent une excitation et une vaporisation rapides de la matrice cristalline et l'éjection de la matrice et des ions analyte en phase gazeuse. Ceci génère un ensemble d'ions de matrice et d'analyte qui sont analysés dans un analyseur de masse TOF. En général, le MALDI confère une charge unique aux protéines, ce qui simplifie et complique à la fois l'analyse en aval. D'une part, le calcul de masse est trivial, car  $m/z = m$  pour  $z$  de +1. L'avantage particulier du MALDI est alors la capacité de produire des ions de grande masse, avec une haute sensibilité. Les protéines de  $m/z$  jusqu'à 500000 peuvent être mesurées avec précision par TOF, et c'est une technique précieuse pour examiner les mélanges.

Cependant, la charge unique rend également plus difficile l'analyse des protéines intactes, car leur grande taille pousse leurs valeurs  $m/z$  en dehors de la zone de détection de la plupart des spectromètres de masse.

### 4.3. Analyse des protéines purifiées par ESI-MS

Une analyse en spectrométrie de masse a été réalisée sur les protéines purifiées suivantes : GFPssrA (50 mg/mL), GFPssrA\_Nter-biotine (20 mg/mL), GFPssrA\_Cter-biotine (20 mg/mL), PAN (20 mg/mL), particule 20S (40 mg/mL) et streptavidine monovalente (20 mg/mL). Toutes les protéines ont été préparées dans un tampon contenant 20 mM Tris-HCl pH7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub> et un volume de 5 µL a été donné à la plateforme de spectrométrie de masse pour l'analyse des échantillons et traitement des données.

La spectrométrie de masse ESI à chromatographie liquide (LC/ESI-MS) a été réalisée sur un spectromètre LC-TOF 6210 couplé à un système HPLC (*Agilent Technologies*). Tous les solvants utilisés étaient de qualité HPLC (Chromasolv, *Sigma-Aldrich*), avec le tampon A à 0,03% acide trifluoroacétique (TFA) dans de l'eau et le tampon B à 95% acétonitrile, 5% eau et 0,03% TFA. Juste avant l'analyse, les échantillons de protéines ont été dilués dans des conditions acides de dénaturation (tampon A) jusqu'à une concentration finale de 5 µM. Les échantillons ont ensuite été dessalés par chromatographie en phase inverse C8 (Zorbax 300 SB-C8, 5 µm, 300 µm ID5 mm, *Agilent Technologies*) pendant 3 minutes à un débit de 50 µL/min avec 100% de tampon A puis élués avec 70% de tampon B à un débit de 50 µL/min pour la détection MS. L'acquisition de la MS a été effectuée en mode ion positif dans la gamme 300 à 3200  $m/z$ . Les spectres MS ont été acquis et les données traitées avec le logiciel MassHunter (*Agilent Technologies*) et avec le logiciel GPMAW (*Lighthouse Data*).

### 4.4. Analyse des produits de dégradation par MALDI-TOF-MS

Une analyse en spectrométrie de masse a été effectuée sur la réaction GFPssrA, PAN et particule 20S à 200 mM MgCl<sub>2</sub> en présence ou en absence de 100 mM ATP dans la condition SANS 2 (tableau 13). Les mélanges réactionnels ont été préparés dans un tampon contenant 20 mM Tris-HCl pH 7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub> dans un volume final de 50 µL et incubés pendant 1 heure à 55°C.

La dénaturation de la GFPssrA a été suivie par spectroscopie de fluorescence selon le protocole décrit dans la partie 5.3 lors de l'incubation à 55°C afin de s'assurer de l'activité du système protéolytique. Les échantillons ont ensuite été donnés à la plateforme de spectrométrie de masse pour l'analyse de l'échantillon et le traitement des données.

Les spectres de masse des échantillons ont été mesurés avec un spectromètre de masse Autoflex (*Bruker Daltonics*) en mode linéaire à ions positifs. L'étalonnage externe de la masse de l'instrument a été effectué à l'aide d'étalons de calibration (*Bruker Daltonics*) offrant une précision de masse < 1000 ppm. Les échantillons ont été mélangés avec une matrice d'acide sinapinique (Sigma ; 10 mg/mL dans l'eau/acétonitrile/acide trifluoroacétique, 50/50/0,1, v/v/v) dans un rapport de 1:1 (v:v) et 1-2 µL de ce mélange ont été déposés sur la plaque et séchés à l'air. Les données des spectres de masse ont été traitées avec le logiciel Flexanalysis (*Bruker Daltonics*).

Un gel SDS-PAGE a également été réalisé. Pour cela, les protéines GFPssrA, PAN et la particule 20S ont été mélangées à 1 mg/mL (équivalant à 36 µM, 3 µM et 1 µM respectivement) soit en

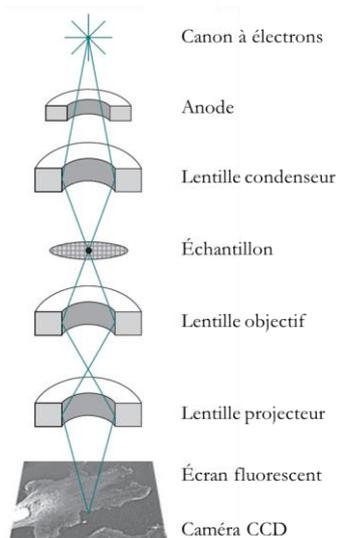
présence de 100 mM ATP & 200 mM MgCl<sub>2</sub>, soit sans ATP. Les deux échantillons ainsi que les trois protéines isolées (GFP<sub>ssrA</sub>, PAN et la particule 20S à 1 mg/mL) ont été incubées 1 heure à 55°C avant d'être déposés sur gel SDS-PAGE à 15% d'acrylamide. La dénaturation de la GFP<sub>ssrA</sub> a été suivie par spectroscopie de fluorescence selon le protocole décrit dans la partie 5.3 et la migration électrophorétique a été réalisée de la même manière que décrit partie 2.4.

## 5. Microscopie électronique à transmission en coloration négative

Les analyses de microscopie électronique à transmission (MET) en coloration négative réalisées au cours de cette thèse ont été effectuées en ayant recours aux services de la plateforme ISBG à l'IBS (Daphna Fenel) pour le traitement des échantillons et la visualisation par microscopie.

### 5.1. Introduction à la microscopie électronique à transmission

La figure 47 représente un schéma simplifié de MET. L'équivalent de la source de lumière dans un MET est un canon à électrons. Lorsqu'une haute tension (10-100 kV) passe entre la cathode et l'anode, une pointe de borure de lanthane (dans le cas du microscope utilisé) émet des électrons. Les électrons chargés négativement traversent un trou dans l'anode pour former un faisceau d'électrons. Le faisceau d'électrons traverse alors des lentilles électromagnétiques. La focalisation du faisceau d'électrons est obtenue en changeant la tension à travers les lentilles électromagnétiques. Lorsque le faisceau d'électrons traverse l'échantillon, certains électrons sont dispersés tandis que d'autres sont focalisés par l'objectif du projecteur sur un écran fluorescent ou enregistrés avec un film photographique ou une caméra numérique. Les électrons ont un pouvoir de pénétration limité, ce qui signifie que les échantillons doivent être minces (50-100 nm) pour pouvoir les traverser.



**Figure 47** : Schéma d'un microscope électronique à transmission. Un faisceau d'électrons d'un canon à électrons est focalisé sur l'échantillon par des lentilles électromagnétiques. L'échantillon est placé sur une grille de microscope électronique. L'image est agrandie avec une lentille d'objectif électromagnétique et projetée sur un détecteur avec la lentille du projecteur. Le détecteur est un écran fluorescent. Toutes ces étapes sont réalisées sous vide. Le trajet des électrons à travers le microscope électronique est représenté en bleu. Adapté de Wilson, K., & Walker, J. (2010)<sup>362</sup>.

Un échantillon est adsorbé sur une grille métallique recouverte d'un film de carbone fin. Les principaux constituants des molécules biologiques (carbone, hydrogène, nitrogène, etc.) ne sont pas très denses en électrons et n'interagissent donc pas avec le faisceau d'électrons. L'échantillon est donc coloré négativement en l'enrobant dans un matériau dense en électrons. Une solution contenant un agent contrastant (ou colorant), tels que l'acétate d'uranyle, le molybdate d'ammonium ou le silicotungstate de sodium, est ajouté sur la grille pendant quelques secondes puis absorbé. L'agent contrastant se fixe préférentiellement à la surface des particules adsorbées, créant une empreinte des particules dans le colorant. Les particules apparaissent alors plus claires

que ce qui l'entoure, d'où le nom de coloration négative. Ce ne sont alors pas les particules elles-mêmes qui sont observées, mais leur empreinte.

La MET en coloration négative est une technique simple et rapide pour déterminer la morphologie globale des macromolécules et permet d'obtenir des images très contrastées. La taille minimale d'une macromolécule pouvant être observée par MET en coloration négative est de 100 kDa et la meilleure résolution pouvant être obtenue, à partir de reconstructions tridimensionnelles, est d'environ 15 Å<sup>369</sup>.

## 5.2. Visualisation de la protéine PAN

La protéine PAN à 1,5 mg/mL a été centrifugée à 10000 x g à 4°C pendant 10 minutes, puis incubée 30 minutes à 55°C dans le tampon de travail supplémenté par 100 mM ATP. Un volume 500 µL de chaque échantillon a ensuite été injecté dans une colonne Superose 6 10/300 GL Increase (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. Une fraction de l'épaulement à gauche du pic d'éluion et une fraction à droite du pic d'éluion ont été prélevées et confiées à la plateforme de microscopie électronique afin d'être traitées et analysées par la technique de coloration négative.

Pour cela, un volume 4 µL de chaque échantillon d'une concentration d'environ 0,1 mg/mL (tampon de travail) a été déposé sur une grille contenant un film mince de carbone amorphe. Les échantillons ont été colorés en utilisant 2% de silicotungstate de sodium, pH 7,5 puis séchés à l'air. Les photographies ont été prises dans des conditions à faible dose avec un microscope électronique T12 FEI à 120 kV et avec un grossissement nominal de 49000 en utilisant une caméra CCD Orius SC1000.

Tampon de travail : 20 mM Tris-HCl pH 7,5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>.

### 1.1. Visualisation de la particule 20S

La particule 20S à 5 mg/mL a été centrifugée à 10000 x g à 4°C pendant 10 minutes, et un volume 500 µL a été injecté dans une colonne Superose 6 10/300 GL Increase (*GE Healthcare*) préalablement équilibrée dans le tampon de travail. Une fraction à gauche du pic d'éluion a été prélevée et confiée à la plateforme de microscopie électronique pour traitement et analyse par la technique de coloration négative de la même manière que précédemment.

### 1.2. Interaction entre PAN et la particule 20S

La protéine PAN et la particule 20S ont été incubées 15 minutes à 70°C avec 2,5 molécules de PAN (0,34 µM soit 0,1 mg/mL) pour une particule 20S (0,14 µM soit 0,1 mg/mL) en présence de 50 µM d'ATPγS dans le tampon de travail. L'échantillon a ensuite été laissé à refroidir à température ambiante avant d'être confié à la plateforme de microscopie électronique pour être traité et analysé par la technique de coloration négative de la même manière que précédemment.

## Chapitre 5 :

# Étude des protéines isolées

Ce chapitre décrit le travail de caractérisation des protéines d'intérêt dans le but de certifier la qualité des protéines purifiées et de maîtriser leur comportement en solution avant de les utiliser pour des tests fonctionnels.

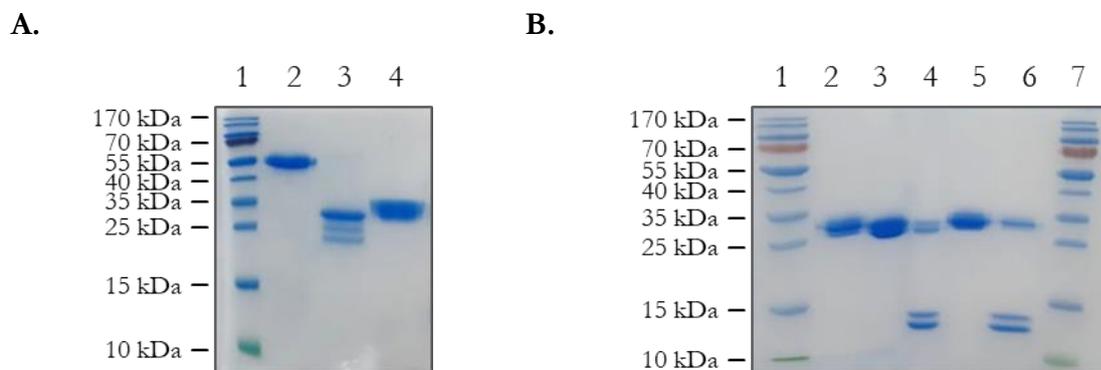
### 1. Liste des différentes protéines purifiées

Les protéines produites lors de ce projet sont les suivantes :

- PAN
- Particule 20S
- Streptavidine monovalente
- GFPssrA
- GFPssrA\_Nter-biotine
- GFPssrA\_Cter-biotine
- GFPssrA\_Nter-streptavidine
- GFPssrA\_Cter-streptavidine

Chacune de ces protéines sera présentée puis étudiée dans une partie dédiée.

La pureté des protéines produites a été estimée par gel SDS-PAGE à 15% d'acrylamide (figure 48), où les protéines migrent en fonction de leur masse moléculaire.



**Figure 48 : Protéines étudiées déposées sur SDS-PAGE**, 15 % acrylamide, coloration à l'Instant Blue. Environ 2  $\mu$ g de chaque échantillon protéique a été déposé sur gel. **A.** Le puits 1 correspond aux marqueurs de masse moléculaire. Les échantillons sont numérotés de 2 à 4 et correspondent aux conditions : PAN, particule 20S, GFPssrA **B.** Les puits 1 et 7 correspondent aux marqueurs de masse moléculaire. Les échantillons sont numérotés de 2 à 6 et correspondent aux conditions : (2) GFPssrA, (3) GFPssrA\_Nter-biotine, (4) GFPssrA\_Nter-streptavidine, (5) GFPssrA\_Cter-biotine, (6) GFPssrA\_Cter-streptavidine.

Les protéines ont été produites pures puisque l'on n'observe que les bandes correspondant à celles-ci sur gel SDS-PAGE à 15% d'acrylamide. Le gel étant en conditions dénaturantes, les monomères des protéines ont été visualisés. La protéine PAN (figure 48.A, 2) présente une bande correspondant à son monomère (49,6 kDa). La particule 20S (figure 48.A, 3) présente 3 bandes correspondant au monomère de sa sous-unité  $\alpha$  (29,9 kDa), au monomère de sa sous-unité  $\beta\Delta 6$  (23,7 kDa), et à une forme plus légère sur gel correspondant probablement à une forme protéolysée de la sous-unité  $\beta\Delta 6$ . Tous les échantillons contenant de la GFPssrA (figure 48.A, 4 ; figure 48.B,

2), GFPssrA\_G10C constituante de la GFPssrA\_Nter-biotine et GFPssrA\_Nter-streptavidine (figure 48.B, 3 & 4) ou de la GFPssrA\_H231C constituante de la GFPssrA\_Cter-biotine et GFPssrA\_Cter-streptavidine (figure 48.B, 5 & 6) présentent deux bandes sur gel. La première bande correspond à la forme intacte de la protéine (27,7 kDa, 27,7 kDa et 27,6 kDa respectivement) et la deuxième bande correspond à une forme plus légère de la protéine correspondant probablement à une forme protéolysée. Les échantillons contenant de la streptavidine monovalente (figure 48.B, 4 & 6) présentent deux bandes, une correspondant à la sous-unité sauvage de la protéine (14,2 kDa) et l'autre correspondant à la forme triple mutante (13,4 kDa). La forme sauvage est plus lourde que la forme mutante car elle possède une étiquette poly-histidines.

Les protéines ont été étudiées séparément et les résultats sont présentés dans les différentes parties de ce chapitre.

## 2. Étude de la particule régulatrice PAN

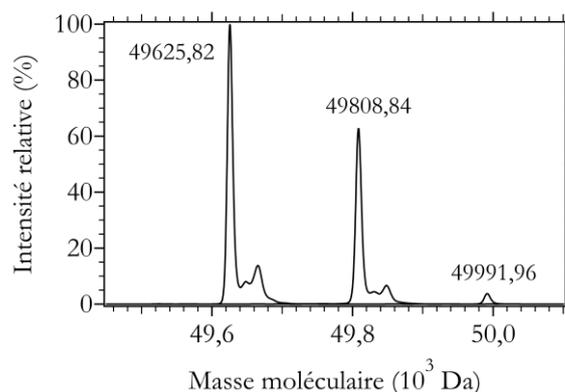
### 2.1. Rappel du contexte

Il a été décrit dans les études sur PAN que la protéine est active sous forme d'hexamère (voir introduction). Cependant, la solution protéique purifiée est constituée d'un mélange de plusieurs formes oligomériques de PAN, avec un état conformationnel majoritaire sous forme de dodécamère, ou dimère d'hexamères (travaux de thèse de Ziad Ibrahim<sup>40</sup>).

La première tâche de ce projet de thèse était de comprendre les conditions d'oligomérisation de PAN et d'obtenir une protéine isolée sous forme d'hexamères en solution.

### 2.2. Qualité de la protéine PAN purifiée mesurée par spectrométrie de masse

La pureté de la solution protéique purifiée a été estimée par spectrométrie de masse. Les échantillons ont été mesurés par ionisation par électrospray (ESI). Les données issues de l'analyse sont présentées dans la figure 49 sous forme de spectre de masse avec l'abondance relative à chaque masse moléculaire. Les masses moléculaires estimées sont très fiables tandis que l'abondance relative des différentes espèces est une information qualitative.



**Figure 49 : Spectre de masse de la protéine PAN purifiée.** Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire.

Le profil de spectrométrie de masse présente un pic principal correspondant à une masse moléculaire de 49625,8 Da et un pic secondaire correspondant à une masse moléculaire plus lourde

de 183 Da (soit 49808,8 Da). La spectrométrie de masse ayant lieu dans des conditions dénaturantes, ces masses moléculaires estimées devraient correspondre à des monomères de protéines.

D'après la séquence en acides aminés, la masse moléculaire théorique d'un monomère de PAN est de 49754,65 Da. La perte de la méthionine initiale à l'extrémité N-terminale lors de la traduction de la protéine, qui est un phénomène courant<sup>370</sup>, constitue une différence de 131 Da par rapport à la masse moléculaire attendue, la masse moléculaire serait donc alors de 49623,65 Da. Cette masse moléculaire théorique correspond à la masse moléculaire du pic principal. Le second pic traduit la présence d'un adduit de 183 Da, lié à l'ajout d'inhibiteur de protéase Pefabloc dans le tampon de lyse. Cet adduit ne devrait pas avoir de conséquences sur les caractéristiques de celle-ci (structure et activité).

On peut donc conclure que **la protéine PAN produite est pure en solution**. En effet, aucun contaminant n'a été détecté et la masse moléculaire estimée correspond à celle d'un monomère de PAN. La protéine purifiée a pu être utilisée pour la suite de l'étude.

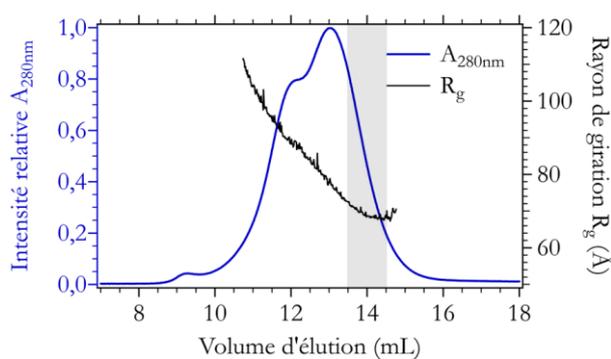
On considérera pour les autres analyses en spectrométrie de masse que tout pic secondaire de masse moléculaire plus lourde 183 Da par rapport au pic principal correspond à un adduit provenant du Pefabloc.

## 2.3. Étude de la protéine PAN par SAXS

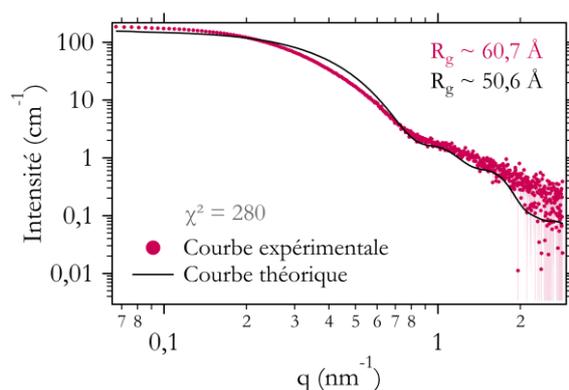
### 2.3.1. Mesures SEC-SAXS

La protéine PAN a été étudiée par SAXS. La solution protéique étant composée de plusieurs formes oligomériques de PAN, la protéine a été chargée sur une colonne d'exclusion de taille et des mesures SAXS ont été réalisées au fur et à mesure de l'élution de la protéine. La figure 50.A représente le profil de SEC-SAXS obtenu. L'absorbance à 280 nm ainsi que le rayon de giration  $R_g$  ont été représentés en fonction du volume d'élution. Le rayon de giration a été calculé et représenté grâce au logiciel CHROMIXS. La figure 50.B représente la courbe SAXS de PAN générée par le logiciel CHROMIXS à partir du profil SEC-SAXS en moyennant des points de mesure du volume d'élution de 13,5 mL à 14,5 mL (zone grisée). La courbe théorique de la protéine PAN hexamérique calculée à partir de la structure cryo-EM (voir partie 5 du chapitre 1) de la protéine grâce au logiciel CRY SOL est également représentée. La figure 50.C illustre la zone de Guinier de la courbe expérimentale ayant permis de calculer le  $R_g$  et  $I(0)$  grâce à l'approximation de Guinier (équation 13).

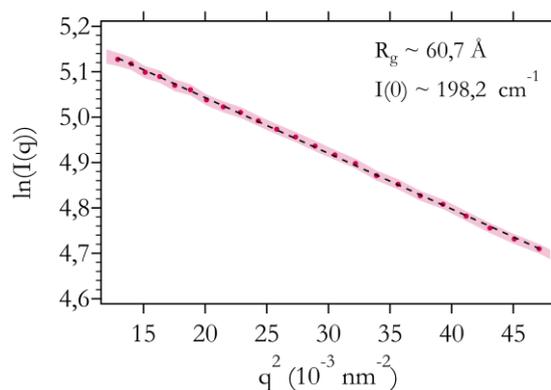
A.



B.



C.



**Figure 50 : Mesures SAXS sur la protéine PAN.** **A.** Profil SEC-SAXS de PAN mesuré sur BM29 (ESRF). La courbe d'absorbance à 280 nm (bleu) a été tracée en fonction du volume d'élution à travers une colonne Superose 6 10/300 GL Increase. Les intensités ont été normalisées par rapport à la valeur maximale du pic d'élution. Le rayon de giration en Å correspondant a été calculé grâce à l'approximation de Guinier (équation 13) est représenté par une ligne noire grâce au logiciel CHROMIXS. Le volume d'élution de 13,5 mL à 14,5 mL est représenté par une zone grisée. **B.** Courbe SAXS expérimentale de PAN générée à partir du profil SEC-SAXS par le logiciel CHROMIXS (moyenne des points de mesure du volume d'élution de 13,5 mL à 14,5 mL) et courbe SAXS théorique calculée par le logiciel CRY SOL à partir de la structure cryo-EM de PAN (PDB 6HEA). **C.** Région de Guinier de la courbe expérimentale SAXS de PAN. Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans les figures B et C, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

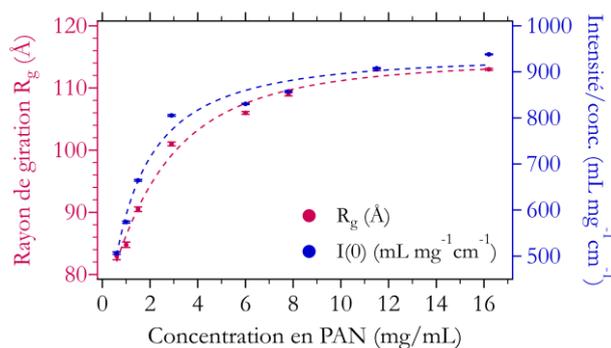
Le profil d'élution de la protéine (figure 50.A) présente un pic principal correspondant probablement à la forme hexamérique de PAN avec un épaulement sur la gauche correspondant probablement à des formes oligomériques plus large (comme décrit dans la partie suivante). De façon logique, le rayon de giration diminue au fur et à mesure de l'élution des protéines. Les profils SAXS des fractions de 13,5 mL à 14,5 mL, où le rayon de giration est stable et qui contiendraient PAN sous forme d'hexamères, ont été moyennés grâce au logiciel CHROMIXS.

Le profil de la courbe SAXS mesurée (figure 50.B) est en désaccord avec la courbe théorique pour tous les vecteurs de diffusion  $q$  représentés. Les décalages observés peuvent s'expliquer par la flexibilité conformationnelle de la partie N-terminale de PAN (voir partie 5 du chapitre 1) et de l'étiquette poly-histidines présente à l'extrémité C-terminale de chaque monomère de la protéine (soit 6 fois). Cependant, le rayon de giration obtenu expérimentalement (60,7 Å) semble cohérent par rapport à la valeur théorique (50,6 Å) compte tenu de l'ajout des étiquettes poly-histidines (figure 50.C).

On peut donc conclure que la **protéine PAN sous forme hexamérique est présente à droite du profil d'éluion et possède un rayon de giration en accord avec la valeur théorique.**

### 2.3.2. Effet de la concentration de PAN sur son état oligomérique étudié par SAXS

Des mesures SAXS en mode statique (« batch ») ont été réalisées sur les solutions de PAN de concentrations croissantes en absence d'ATP afin d'investiguer l'effet de la concentration de la protéine sur son état oligomérique. La figure 51 représente les valeurs de  $R_g$  et de  $I(0)$  en fonction de la concentration en PAN calculées grâce à l'approximation de Guinier (équation 13).



**Figure 51 : Valeurs des rayons de giration moyens et intensités à angle nul moyennes mesurés pour des solutions de PAN à différentes concentrations.**

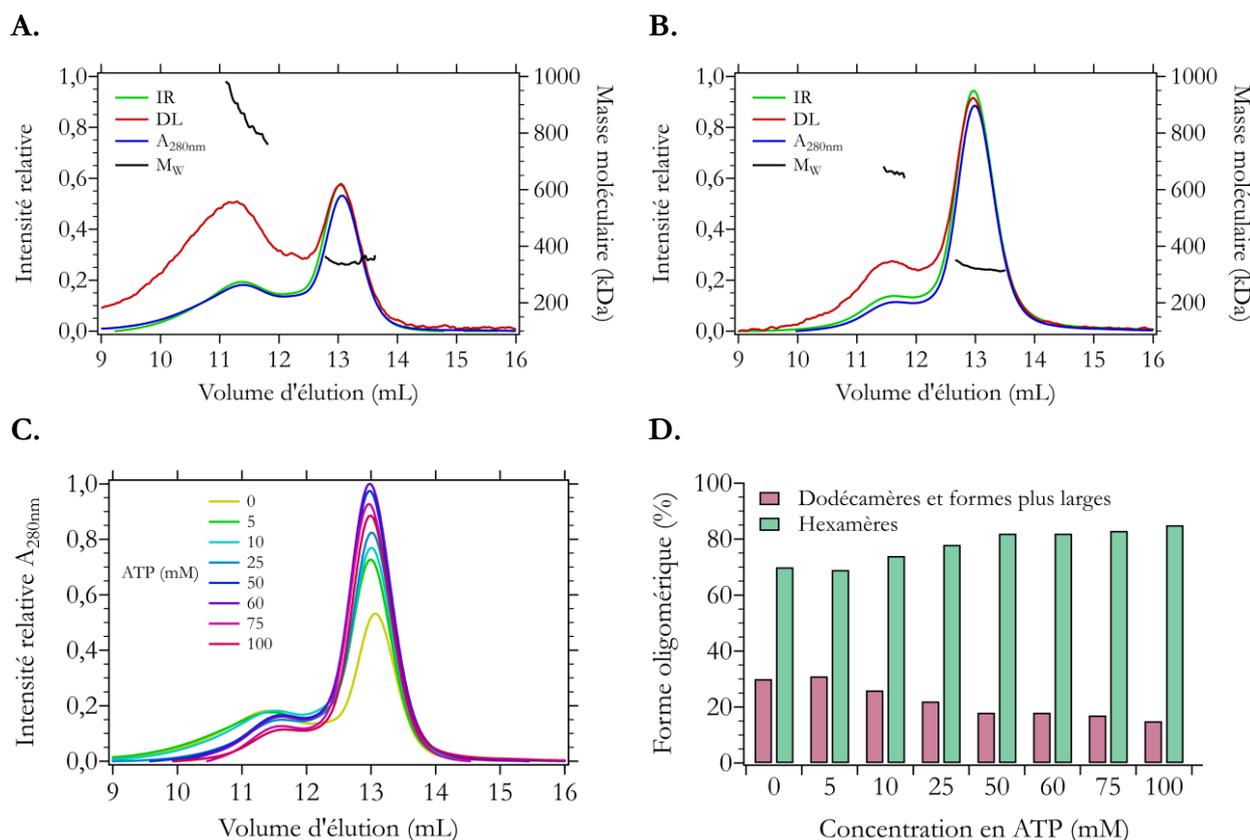
D'après l'équation 12, la masse moléculaire moyenne  $M_w$  peut être calculée à partir des valeurs de  $I(0)$  et de concentration. La valeur de  $I(0)$  présentée dans la figure 51 étant déjà normalisée par la concentration, les valeurs présentées sont directement proportionnelles à la masse moléculaire  $M_w$  (en kDa). Comme expliqué dans la partie 1.2.2, le rayon de giration est quant à lui un indicateur de la compacité de la protéine.

On peut observer que les valeurs de  $I(0)$  et de  $R_g$  augmentent avec la concentration de PAN. Cela indique que la masse moléculaire et la compacité des protéines en solution augmentent avec la concentration en protéine. À 1 mg/mL, la valeur de  $I(0)$  correspondant à une masse moléculaire d'environ 500 kDa indique que la protéine PAN est déjà sous forme d'un mélange d'hexamères (298 kDa) et de dodécamères (596 kDa). De plus, la valeur de  $R_g$  d'environ 82 Å est supérieure à celle de l'hexamère de PAN (voir partie précédente). On peut donc conclure que **la protéine PAN s'oligomérisse lorsque sa concentration augmente et que la protéine est déjà sous forme d'un mélange d'hexamères et de dodécamères à 1 mg/mL.**

## 2.4. Effet de la concentration d'ATP sur l'état oligomérique de PAN

### 2.4.1. Effets de la concentration d'ATP étudiés par SEC-MALLS

L'effet de la concentration d'ATP sur l'état oligomérique de PAN (à 1,5 mg/mL) a été étudié par SEC-MALLS. Pour cela, différentes concentrations d'ATP (0, 5, 10, 25, 50, 60, 75 ou 100 mM) ont été incubées avec la protéine PAN pendant 30 minutes à 55°C dans un tampon à 20 mM Tris-HCl pH7,5, 100 mM NaCl et 10 mM  $\text{MgCl}_2$ , puis chaque solution a été mesurée par SEC-MALLS. La figure 52 présente les résultats obtenus.



**Figure 52 : Profil SEC-MALLS de la protéine PAN.** A & B. Profil SEC-MALLS de PAN à 0 mM ATP (A) et 100 mM ATP (B). Les courbes d'indice de réfraction (vert), d'absorbance à 280 nm (bleu) et de diffusion de la lumière (rouge) ont été tracées en fonction du volume d'élution à travers une colonne Superose 6 10/300 GL Increase. Les intensités ont été normalisées par rapport à la valeur maximale de l'échantillon le plus concentré. La masse moléculaire en kDa correspondant aux différents pics d'élution est représentée par une ligne noire. C. Profil SEC-MALLS de PAN à 0, 5, 10, 25, 50, 60, 75 et 100 mM ATP. Les courbes représentent l'absorbance à 280 nm en fonction du volume d'élution à travers une colonne Superose 6 10/300 GL Increase aux différentes concentrations d'ATP. Les intensités ont été normalisées par rapport à la valeur maximale de l'échantillon le plus concentré. D. Pourcentage de protéines PAN sous forme d'hexamères ou sous forme de dodécamères (et formes plus large) d'après les masses moléculaires estimées et les masses calculées par le logiciel ASTRA.

Les figures 52.A et 52.B représentent les profils de SEC-MALLS de la protéine PAN à respectivement 0 et 100 mM ATP. L'absorbance à 280 nm, l'indice de réfraction et de diffusion de la lumière sont représentés en fonction du volume d'élution. Les courbes d'absorbance à 280 nm et d'indice de réfraction sont quasiment superposables tandis que les courbes de diffusion de la lumière sont plus intenses aux volumes d'élution plus faibles (où les protéines éluées ont des masses moléculaires plus élevées). Cela est dû au fait que l'intensité de la diffusion de la lumière d'une solution est directement proportionnelle à la masse moléculaire moyenne de ses composants. Les profils SEC-MALLS sont similaires pour les deux concentrations d'ATP avec un pic principal correspondant à une protéine éluée vers 13 mL. On remarque qu'un pic minoritaire vers 11 mL est présent à 0 mM ATP et ne reste plus que sous forme d'épaulement à 100 mM ATP. Ces deux pics ou épaulements correspondent à des formes oligomériques différentes de PAN. La masse moléculaire des protéines éluées a été calculée à partir de l'indice de réfraction et de la diffusion de la lumière (équation 35) et est également représentée sous forme de ligne en fonction du volume d'élution pour chaque oligomère de protéine élué.

À 0 mM ATP, la ligne correspondant à la masse moléculaire du pic minoritaire n'est pas plate, signifiant qu'il y a plusieurs formes oligomériques (polydispersité) au sein de ce pic avec une masse moléculaire de plus de 800 kDa. La ligne correspondant à la masse moléculaire du pic principal est quant à elle plate, suggérant que les protéines sont monodisperses au sein de ce pic, avec une masse moléculaire moyenne de 346 kDa. Ceci diffère de la masse moléculaire théorique de PAN qui est de 298 kDa. L'indice de polydispersité indique que cette population de protéine est légèrement polydisperse ( $M_w/M_n = 1,002$ ), suggérant que le pic d'éluion est principalement composé de protéines PAN sous forme d'hexamères avec une légère contamination de formes oligomériques plus larges.

À 100 mM ATP, la ligne correspondant à la masse moléculaire de l'épaule est relativement plate, suggérant une solution protéique monodisperse, avec une masse moléculaire moyenne de 663 kDa. L'indice de polydispersité indique que cette population de protéines est légèrement polydisperse ( $M_w/M_n = 1,002$ ). On peut supposer qu'il s'agit principalement de dodécamères de PAN, dont la masse moléculaire théorique est de 596 kDa, mélangés avec une faible quantité de formes oligomériques plus larges. La ligne correspondant à la masse moléculaire du pic principal est elle aussi relativement plate, avec une masse moléculaire moyenne de 326 kDa. Cela diffère toujours de la masse moléculaire théorique de PAN (298 kDa), même si la solution est moins polydisperse qu'à 0 mM ATP ( $M_w/M_n = 1,001$ ). Cela suggère qu'il y a toujours une légère contamination de la solution d'hexamères de PAN par une forme oligomérique plus large (probablement des dodécamères).

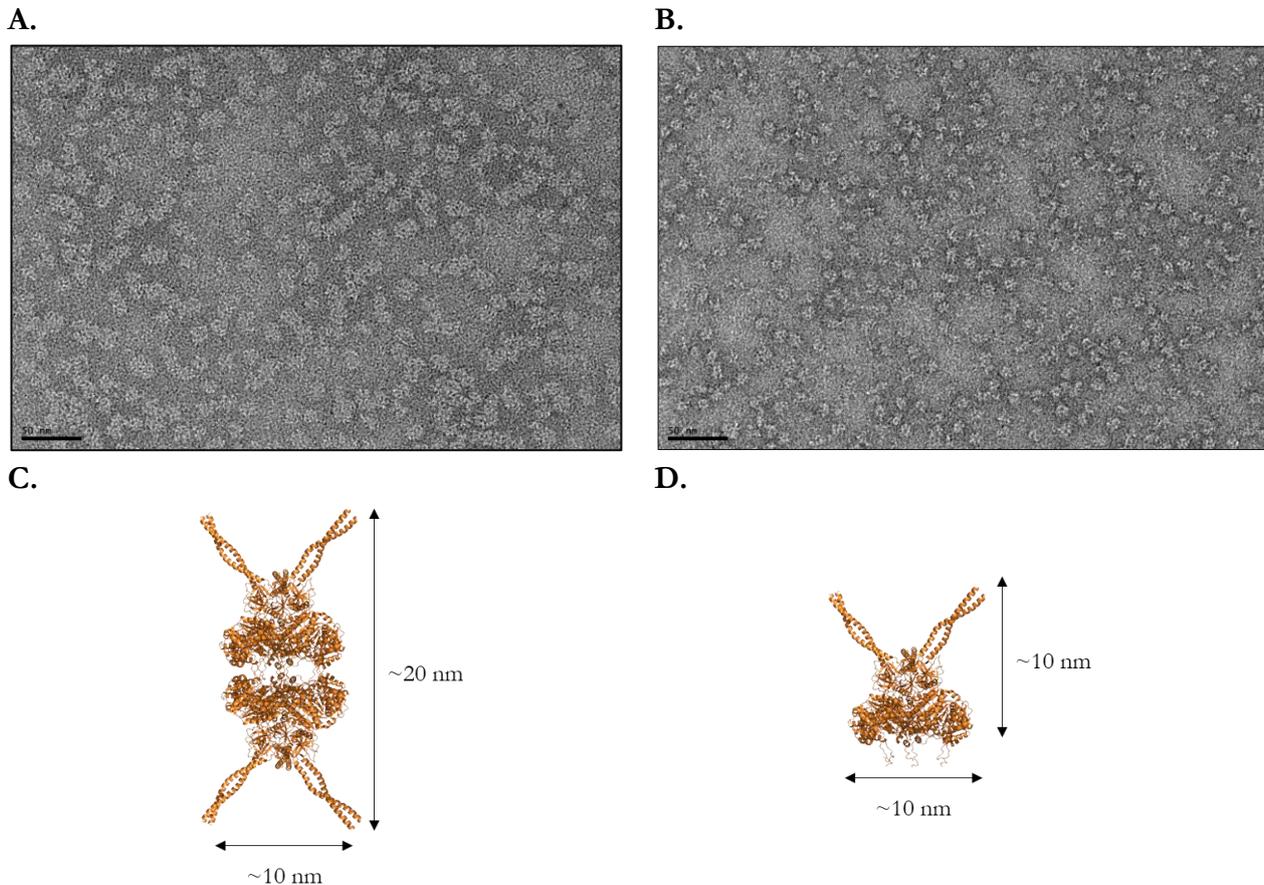
La figure 52.C présente le profil d'absorbance à 280 nm en fonction du volume d'éluion pour toutes les concentrations d'ATP testées. On observe de façon qualitative que plus la concentration d'ATP augmente, plus l'épaule (dodécamères ou formes plus larges) disparaît au profit du pic principal (hexamères). La figure 52.D représente ce phénomène de façon plus quantitative. La proportion d'hexamères et de dodécamères (ou formes plus larges) a été estimée pour chaque condition à partir de la masse des différents oligomères calculée par le logiciel ASTRA. La proportion de dodécamères et formes plus larges diminue au profit de la forme hexamérique lorsque la concentration d'ATP augmente. A partir de 50 mM ATP, une diminution de 50% des formes plus larges est observée au profit des hexamères. La condition 100 mM ATP semble la plus adéquate pour obtenir de la protéine sous forme d'hexamères, bien qu'il y ait toujours 15% de dodécamères.

On peut conclure qu'**incuber PAN à 55°C en présence d'ATP et de MgCl<sub>2</sub> permet d'hexamériser la protéine**, comme suggéré par les travaux précédents (thèse de Ziad Ibrahim). De plus, **100 mM ATP est la meilleure condition permettant d'hexamériser PAN malgré la présence de 15% de dodécamères**. Ce résultat pourrait être biaisé par le fait qu'il n'y ait pas d'ATP dans le tampon d'éluion et/ou que les mesures aient lieu à température ambiante et non à 55°C.

#### 2.4.2. Visualisation de la protéine PAN par microscopie électronique à coloration négative

Les différents oligomères de PAN ont été visualisés par microscopie électronique à transmission en coloration négative. Pour cela, deux échantillons ont été mesurés à partir des conditions identiques à l'obtention du profil SEC-MALLS en figure 52.B. Une fraction contenant du

dodécamère de l'épaule et une contenant de l'hexamère du pic principal ont été prélevés et visualisés et correspondent respectivement aux figures 53.A et 53.B.



**Figure 53 : Protéine PAN observée au microscope électronique à transmission.** Visualisation de l'état oligomérique de PAN par microscopie électronique à transmission en coloration négative (2% sodium silicotungstate) à un grossissement de 49000 X. Barre d'échelle : 50 nm. **A.** Dodécamères de PAN. **B.** Hexamères de PAN. **C.** Représentation artistique d'un dodécamère de PAN réalisée à partir du fichier PDB d'un hexamère de PAN (PDB 6HEA). **D.** Structure d'un hexamère de PAN (PDB 6HEA).

Les observations au microscope électronique à transmission montrent que les deux échantillons contiennent une population homogène de protéines, et permettent de visualiser des oligomères de PAN différents. La résolution de la microscopie en coloration négative ne permet pas de visualiser précisément la forme des protéines en solution. Cependant, la taille des particules visualisées semble correspondre respectivement à un dodécamère de PAN (figure 53.C) et à un hexamère de PAN (figure 53.D).

En conclusion, **les deux formes oligomériques de PAN peuvent être visualisées par microscopie électronique.** De plus, on peut visualiser **une population homogène d'hexamère de PAN après incubation à 55°C à 100 mM ATP** et sélection des fractions après chromatographie d'exclusion stérique.

#### 2.4.3. Effet des concentrations d'ATP et de NaCl étudiés par SLS-DLS

Différentes conditions de tampon ont été testées par SLS/DLS (variations de concentrations en sels NaCl, KCl, et MgCl<sub>2</sub>, variations de pH) dans le but d'optimiser l'hexamérisation de PAN. Cette technique a l'avantage de pouvoir mesurer de nombreuses conditions de tampon d'un échantillon

sans avoir à équilibrer pour chacune une colonne de chromatographie avant d'effectuer une mesure. De plus, l'instrument permet de réaliser des mesures à 55°C. Les informations obtenues sont alors issues de l'ensemble des protéines présentes en solution.

Parmi les diverses conditions de tampon testées, la condition la plus intéressante pour obtenir PAN sous forme d'hexamères a été un tampon similaire au tampon de travail (100 mM Tris-HCl pH7,5 et 10 mM MgCl<sub>2</sub>) avec 500 mM NaCl au lieu de 100 mM. On suppose qu'une haute concentration de sel limite les interactions électrostatiques entre les hexamères de PAN et permet de favoriser ainsi l'hexamérisation de la protéine. Les résultats des mesures de SLS/DLS à 55°C dans les conditions de tampon à 100 et 500 mM NaCl en présence de 100 mM ATP ou sans ATP sont représentés dans le tableau 16.

**Tableau 16 : Mesures SLS-DLS sur la protéine PAN.** Rayon hydrodynamique et masse moléculaire mesurées par SLS/DLS dans différentes conditions de tampon. Les erreurs données pour le rayon hydrodynamique et la masse moléculaire sont des erreurs statistiques et ne représentent pas une polydispersité.

Concentration en ATP (mM)	Concentration en NaCl (mM)	Rayon hydrodynamique (nm)	Masse moléculaire (kDa)
0	100	11,7 ± 0,1	700 ± 4
0	500	10,5 ± 0,1	521 ± 5
100	100	7,8 ± 0,1	305 ± 3
100	500	9,1 ± 0,1	361 ± 3

La technique de SLS/DLS permet de déterminer le rayon hydrodynamique et la masse moléculaire moyenne de toutes les espèces en solution. Cette méthode ne permet pas de distinguer les hexamères et dodécamères de PAN (dont le poids moléculaire est trop proche pour la technique) mais en renseignant sur les valeurs moyennes des paramètres mesurés, elle fournit une estimation de la proportion de chacune des formes oligomériques en solution.

On observe que sans ATP, la condition à 500 mM NaCl a permis d'obtenir un rayon hydrodynamique moyen et la masse moléculaire moyenne des espèces en solution plus faibles qu'à 100 mM NaCl. À 500 mM NaCl, la masse moléculaire moyenne est plus faible que la masse moléculaire théorique du dodécamères de PAN (596 kDa), suggérant qu'il s'agit d'un mélange d'hexamères et de dodécamères en solution. En présence de 100 mM ATP, le rayon hydrodynamique moyen et la masse moléculaire moyenne des espèces en solution diminuent pour les deux conditions de sel. Cependant, ces valeurs sont plus faibles à 100 mM qu'à 500 mM NaCl, et la masse moléculaire moyenne à 100 mM NaCl est très proche de la valeur théorique de l'hexamère de PAN (298 kDa). Les valeurs légèrement plus élevées à 500 mM NaCl indiquent la présence d'une faible quantité de dodécamères en solution.

On peut donc conclure dans les conditions testées que :

- **En absence d'ATP, l'utilisation de 500 mM NaCl permet de favoriser la formation d'hexamères de PAN.** Cela est probablement dû à des liaisons des ions du tampon avec les charges à la surface de PAN, limitant les interactions électrostatiques entre plusieurs hexamères.
- **À 100 mM NaCl, la présence de 100 mM ATP permet d'hexamériser toute la population de PAN.** Ce résultat diffère des résultats de SEC-MALLS (partie 2.4.1). Cette différence est sûrement due au fait que l'échantillon n'est pas dilué dans un tampon sans ATP lors de son

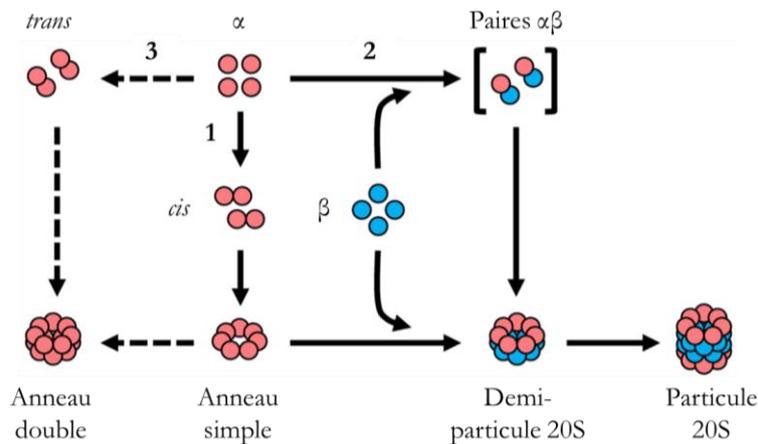
passage sur la colonne de chromatographie, et que la température est maintenue à 55°C, permettant à toute la population de PAN de rester sous forme d'hexamères. **La condition de tampon à 100 mM NaCl supplémentée de 100 mM ATP a donc été retenue pour hexamériser PAN et les expériences biochimiques et biophysiques.**

### 3. Étude de la particule catalytique 20S

#### 3.1. Problématique de l'assemblage de la particule 20S

La structure générale et les propriétés de la particule 20S de *M. jannaschii* ont été décrites dans l'introduction. La particule 20S est composée de deux anneaux de sous-unités catalytiques  $\beta$  pris entre deux anneaux de sous-unités  $\alpha^{246}$ . Les sous-unités  $\beta$  sont synthétisées *in vivo* comme des précurseurs et subissent une activation autocatalytique pour exposer la thréonine à l'extrémité N-terminale grâce au clivage des acides aminés 1 à 6 (appelés « propeptide »). La protéine  $\beta\Delta 6$  utilisée lors de cette étude ne possède pas le propeptide afin d'exposer cette thréonine<sup>277</sup>.

Lors de l'assemblage de la particule 20S recombinante, une partie des sous-unités  $\alpha$  s'assemblent irréversiblement sous la forme d'anneaux double non fonctionnels, empêchant la formation de protéasome 20S actif, tel qu'illustré dans la figure 54<sup>371</sup>. Aucune des méthodes de purification disponibles au laboratoire n'a pu permettre de séparer les particules 20S des molécules en anneau double de sous-unités  $\alpha$ .

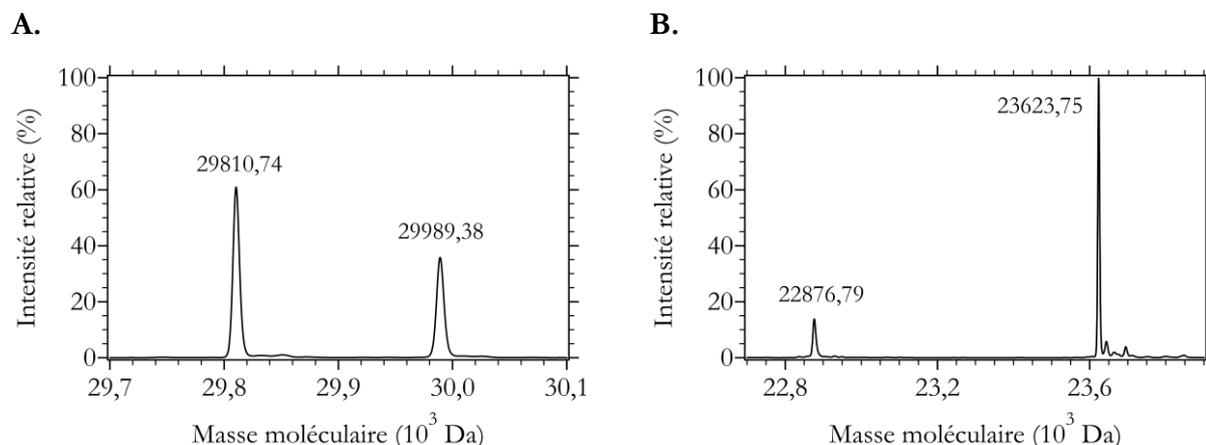


**Figure 54 : Réseau d'assemblage de la particule 20S de *Methanococcus maripaludis*.** Trois voies d'assemblage sont disponibles pour les sous-unités  $\alpha$ . Les sous-unités de  $\alpha$  peuvent interagir les unes avec les autres en cis (voie 1) menant à la formation d'un anneau simple. L'anneau simple sert de socle pour les sous-unités  $\beta$  jusqu'à ce qu'une demi-particule 20S se forme, qui se dimérise pour donner naissance à la particule 20S. Les sous-unités  $\alpha$  peuvent également interagir avec les sous-unités  $\beta$  pour former directement des demi-particules 20S (voie 2). Ici, le crochet désigne les hétérodimères  $\alpha\beta$  comme le précurseur le plus probable des demi-particules 20S. Les sous-unités  $\alpha$  peuvent interagir les unes avec les autres en trans (voie 3). Cela conduit à la formation d'un anneau double  $\alpha$  qui est structuellement analogue à une demi-particule 20S. L'anneau double peut également se former directement à partir d'anneaux simple. La formation de l'anneau double  $\alpha$  est un processus (lignes pointillées) qui rivalise avec les réactions menant à une particule 20S fonctionnelle (lignes pleines). Adapté de Panfair, D., et al. (2015)<sup>371</sup>.

Les résultats présentés ci-dessous ont été obtenus après optimisation des protocoles d'expression, par coexpression des sous-unités permettant d'éviter la formation d'anneaux double de sous-unités  $\alpha$ <sup>372</sup>, et de purification de la particule 20S afin de minimiser la contamination de la solution protéique par des anneaux double non fonctionnels de sous-unités  $\alpha$ .

### 3.2. Qualité de la particule 20S purifiée estimée par spectrométrie de masse

La pureté de la solution protéique purifiée a été estimée par spectrométrie de masse ESI. Les données issues de l'analyse sont présentées de la même façon que précédemment (partie 2.2) dans la figure 55.



**Figure 55 : Spectre de masse de la particule 20S purifiée.** Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire.

Deux pics principaux possédant chacun un pic secondaire sont visibles sur les spectres de masse. Dans la figure 55.A, le principal correspond à une masse moléculaire de 29810,74 Da, et le pic secondaire correspond à une masse moléculaire plus lourde de 179 Da (adduit de Pefabloc). Dans la figure 55.B, le pic principal correspond à une masse moléculaire de 23623,75 Da et le pic secondaire correspond à une masse moléculaire de 22876,9 Da. Ces masses moléculaires estimées devraient correspondre à des monomères des sous-unités de la particule 20S.

La masse moléculaire théorique du monomère de la sous-unité  $\alpha$  de la particule 20S est de 29941,35 Da. Avec la perte de la méthionine initiale à l'extrémité N-terminale (perte de 131 Da), la masse moléculaire serait donc alors de 29810,15 Da. Cette masse moléculaire théorique correspond à la masse moléculaire du second pic principal.

La masse moléculaire théorique du monomère de la sous-unité  $\beta\Delta 6$  de la particule 20S est de 23754,68 Da. La perte de la méthionine initiale aurait pour conséquence une masse moléculaire de 23623,48 Da. Cela coïncide avec la masse moléculaire du premier pic principal. Le pic secondaire associé correspond à la même molécule ayant perdu 6 acides aminés à l'extrémité N-terminale et 3 acides aminés à l'extrémité C-terminale. Cela pourrait par exemple être dû à une maturation lors de l'assemblage de la particule 20S, ce qui n'a pas été décrit dans la littérature, ou dû à une dégradation de la protéine par des protéases malgré l'ajout d'inhibiteurs de protéases dans le tampon de lyse. Cette espèce pourrait expliquer la troisième bande visible sur gel SDS-PAGE (voir partie 1).

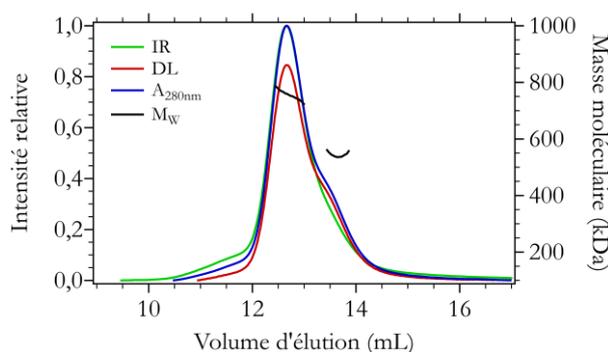
L'abondance relative des deux sous-unités (pics principaux et secondaires) est similaire, ce qui indique qu'il y a autant de sous-unités  $\alpha$  que  $\beta\Delta 6$ . Cela est logique puisque la particule 20S est composée d'un nombre identique de sous-unités  $\alpha$  et  $\beta\Delta 6$ .

On peut donc conclure de façon générale que **la particule 20S produite est pure en solution**. En effet, aucun contaminant n'a été détecté et la masse moléculaire estimée correspond à celle des monomères des sous-unités  $\alpha$  et  $\beta\Delta 6$ .

La sous-unité  $\beta\Delta 6$  étant celle responsable de la protéolyse, on peut s'interroger au sujet de l'éventuelle conséquence de la perte de certains de ces acides aminés sur l'activité des particules 20S purifiées. 14 monomères s'associent pour former une particule 20S. D'après la hauteur de pic observé, l'espèce tronquée présente une abondance relative plus faible par rapport à la forme intacte de la sous-unité  $\beta\Delta 6$  (environ 10%). Il y aurait donc seulement environ une sous-unité sur 14 ayant perdu des acides aminés. L'activité de la particule 20S purifiée ne devrait donc pas être impactée et la protéine purifiée a été utilisée pour la suite de l'étude.

### 3.3. Étude de la particule 20S par SEC-MALLS

Des mesures SEC-MALLS ont été réalisées afin d'estimer si la particule 20S est monodisperse en solution. La figure 56 représente les profils de SEC-MALLS de la particule 20S comme représenté précédemment.



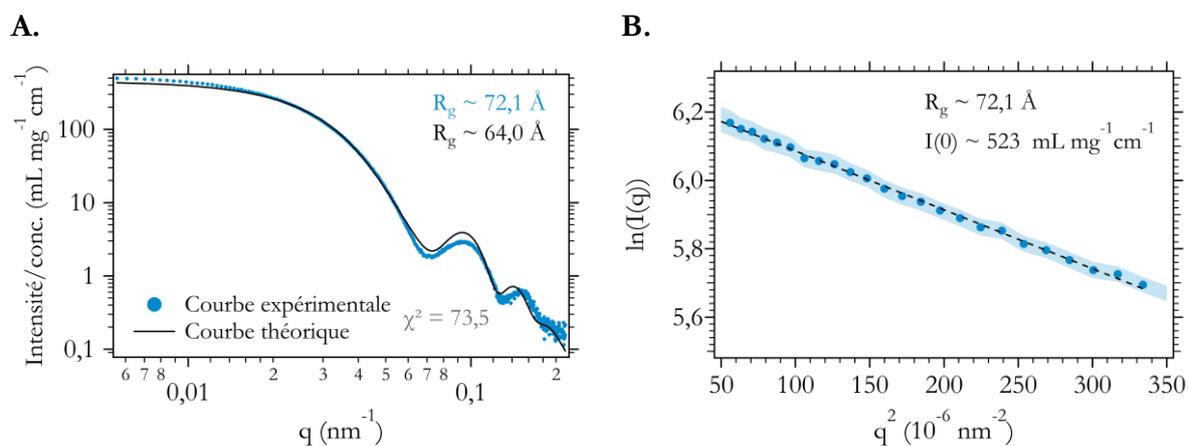
**Figure 56 : Profil SEC-MALLS de la particule 20S.** Les courbes d'indice de réfraction (vert), d'absorbance à 280 nm (bleu) et de diffusion de la lumière (rouge) ont été tracées en fonction du volume d'élution d'une colonne Superose 6 10/300 GL Increase. La masse moléculaire en kDa correspondant aux différents pics d'élution est représentée par une ligne noire.

Le profil SEC-MALLS présente un pic d'élution avec un épaulement sur la droite. La ligne correspondant à la masse moléculaire du pic d'élution n'est pas plate, et l'indice de polydispersité est d'environ  $M_w/M_n \approx 1,040$ , signifiant que la solution protéique est polydisperse. Cependant, la masse moléculaire moyenne est estimée à 756,6 kDa ( $\pm 0,8 \%$ ), ce qui coïncide avec la masse théorique de la particule 20S de 751,3 kDa. La ligne correspondant à la masse moléculaire de l'épaulement du pic d'élution n'est également pas plate, avec un rapport indicateur de la polydispersité  $M_w/M_n \approx 1,050$ . De plus, la masse moléculaire moyenne est estimée à 565,2 kDa ( $\pm 0,8 \%$ ). Cela est en accord avec l'hypothèse d'un mélange de particules 20S et de sous-unités  $\alpha$  assemblées sous formes d'anneau double (voir partie 3.1) dont la masse moléculaire théorique est de 418,9 kDa.

En conclusion, **la solution de particule 20S purifiée est contaminée par une faible quantité de sous-unités  $\alpha$  assemblées seules**. Pour la suite des travaux réalisés lors de cette thèse, seules les fractions à gauche du pic d'élution ont été prélevées pour limiter cette contamination. On peut estimer qu'une faible quantité de sous-unités  $\alpha$  assemblées seules était néanmoins toujours présente en solution.

### 3.4. Étude de la particule 20S étudiée par SAXS

La particule 20S a été étudiée par SAXS dans le but de mieux la caractériser. Les données sont représentées de la même manière que précédemment (partie 2.3) et la courbe théorique de la particule 20S a été calculée à partir de la structure cryo-EM<sup>286</sup> (voir partie 5 du chapitre 1) de la protéine grâce au logiciel CRY SOL.



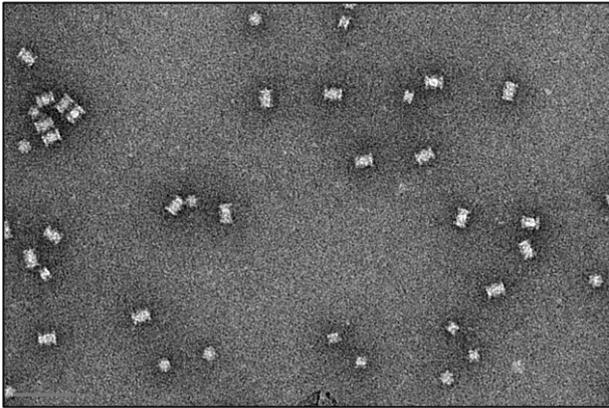
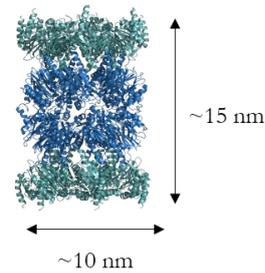
**Figure 57 : Mesures SAXS sur la particule 20S.** **A.** Courbe SAXS expérimentale de la particule 20S mesurée sur BM29 (ESRF) et courbe SAXS théorique calculée par le logiciel CRY SOL à partir de la structure cryo-EM de la particule 20S (PDB 6HEA). **B.** Région de Guinier de la courbe expérimentale SAXS de la particule 20S. Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans les deux figures, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

La courbe SAXS mesurée est relativement similaire à la courbe théorique pour tous les vecteurs de diffusion  $q$  représentés. Les décalages observés peuvent s'expliquer par la présence d'étiquettes poly-histidines flexibles présentes sur chaque sous-unité  $\alpha$  de la particule 20S (soit 14 fois). Les deux pics observés indiquent que la protéine est globulaire. Le rayon de giration obtenu expérimentalement (72,1 Å) semble cohérent par rapport à la valeur théorique (64,0 Å) si l'on prend en compte les nombreuses étiquettes poly-histidines ajoutées. On peut noter que la faible augmentation de l'intensité de diffusion aux valeurs  $q$  faibles ( $q < 0,015 \text{ nm}^{-1}$ ) indique une légère agrégation de l'échantillon<sup>40,333,334</sup>. D'après la valeur du  $I(0)$  présentée dans la figure 57.B,  $M_w$  serait d'environ 500 kDa (voir partie 2.3.2). Cela ne correspond pas à la masse moléculaire théorique de la particule 20S (751,3 kDa), bien que la valeur de  $I(0)$  soit similaire pour la courbe théorique (environ 525 mL.mg<sup>-1</sup>.cm<sup>-1</sup>). Cela est en accord avec l'hypothèse d'un mélange de particules 20S et d'anneaux double de sous-unités  $\alpha$  (voir partie 3.1 et 3.2) dont la masse moléculaire théorique est de 418,9 kDa.

On peut donc conclure que **la particule 20S possède un rayon de giration en accord avec la valeur théorique mais avec une masse moléculaire indiquant la présence d'anneaux double de sous-unités  $\alpha$ .**

### 3.5. Visualisation de la particule 20S par microscopie électronique à coloration négative

La particule 20S a été visualisée par microscopie électronique à transmission en coloration négative (figure 58.A).

**A.****B.**

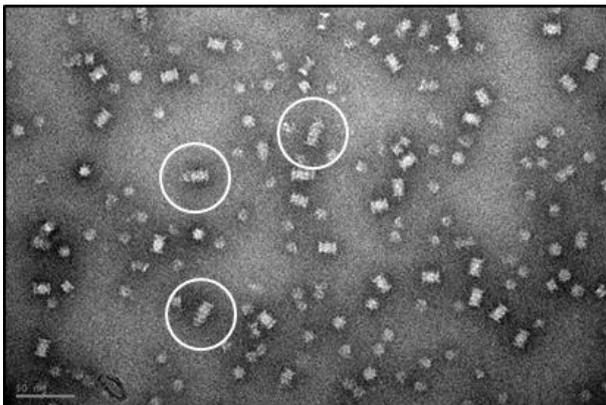
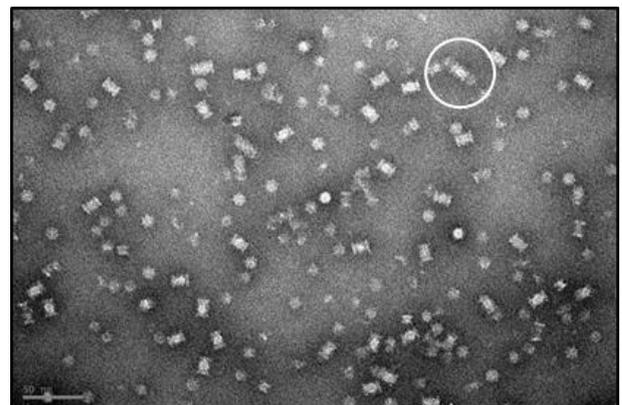
**Figure 58 : Particule 20S observée au microscope électronique à transmission. A.** Particules 20S observées au microscope électronique à transmission en coloration négative (2% sodium silicotungstate) à un grossissement de 49000 X. Barre d'échelle : 50 nm. **B.** Structure de la particule 20S (PDB 6HEA).

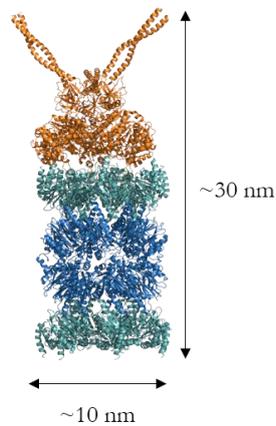
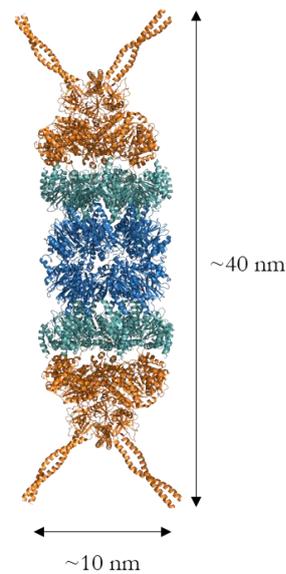
D'après les observations au microscope électronique à transmission, l'échantillon contient une population de protéines composée majoritairement de particule 20S. En effet, la forme similaire et les dimensions de la protéine observée sont similaires à la structure connue de la particule 20S (figure 58.B). On peut même distinguer deux anneaux centraux correspondant aux deux anneaux de sous-unités  $\beta\Delta 6$ , entourés de chaque côté d'un anneau correspondant aux anneaux de sous-unités  $\alpha$ . Une espèce minoritaire plus petite est également visualisée. D'après les résultats précédents et la littérature (voir partie 3.1), on peut affirmer qu'il s'agit de sous-unités  $\alpha$  assemblées sous forme d'anneau double.

Pour conclure, **la solution de particules 20S peut être visualisée et contient une minorité de sous-unités  $\alpha$  assemblées en anneau double.**

### 3.6. Interaction avec PAN visualisée par microscopie électronique à coloration négative

Une fois la qualité des solutions protéiques de la particule 20S et de PAN mise en évidence, l'objectif du travail était de visualiser la machinerie protéolytique PAN-20S assemblée. Les deux protéines ont donc été incubées 30 minutes en présence d'ATP $\gamma$ S (ATP non hydrolysable) puis la solution a été visualisée par microscopie électronique à transmission en coloration négative (figure 59.A et 59.B). En effet, il a été démontré que l'association transitoire de PAN avec le protéasome 20S est favorisée lors de la liaison à l'ATP et réduite par hydrolyse de l'ATP<sup>373</sup>.

**A.****B.**

**C.****D.**

**Figure 59 : Protéasome PAN-20S observé au microscope électronique à transmission. A & B.** Particules 20S incubées avec PAN et de l'ATP $\gamma$ S observés au microscope électronique à transmission en coloration négative (2% sodium silicotungstate) à un grossissement de 49000 X. Barre d'échelle : 50 nm. Les complexes où PAN est associé à la particule 20S sont entourés en blanc. **C & D.** Structure de la particule 20S avec PAN associé à une extrémité (**C**) ou à deux extrémités (**D**) (PDB 6HEA).

Les observations au microscope électronique à transmission de l'assemblage des complexes PAN-20S ont révélé qu'environ 10% des particules 20S s'associaient à la protéine PAN dans les conditions expérimentales testées. De plus, il semble que PAN puisse s'associer au protéasome 20S à une extrémité ou aux deux extrémités, comme décrit dans la littérature<sup>373</sup>, avec une forme et des dimensions similaires aux structures connues (figure 59.C et 59.D).

On peut donc conclure que **PAN et la particule 20S purifiées s'associent physiquement dans les conditions testées**. Le fait que seulement 10% des molécules soient associées peut être expliqué par le fait que le système protéolytique chez les archées est un complexe transitoire dont l'association PAN-20S n'a jamais été maintenue stable *in vitro* (voir partie 5.1 du chapitre 1).

## 4. Étude de la GFP<sub>sSrA</sub>

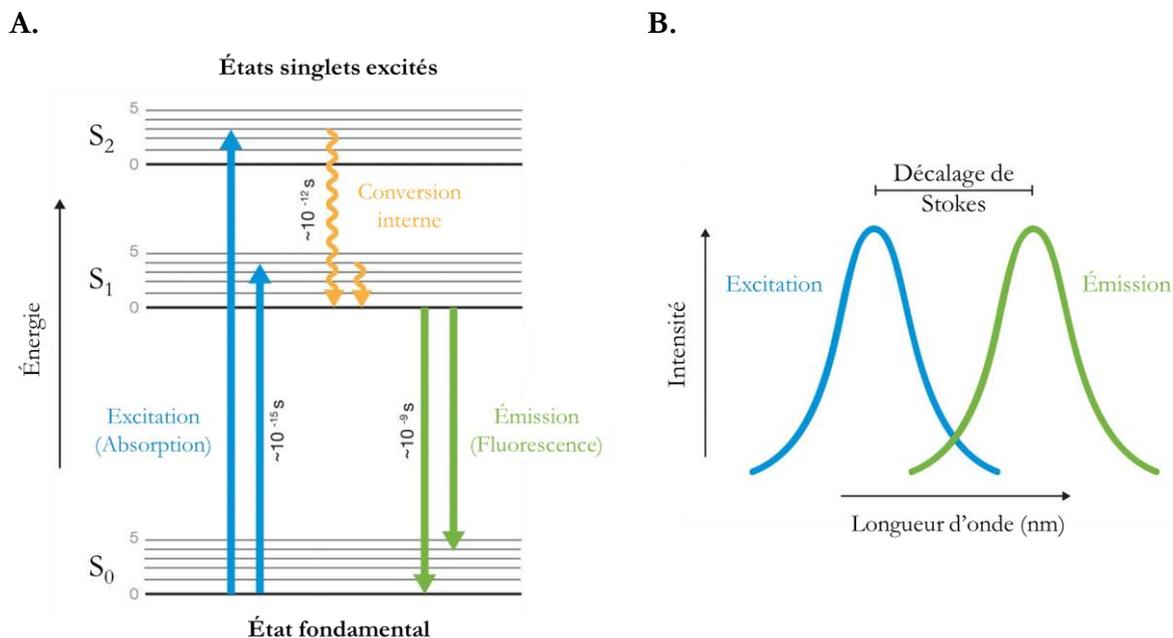
### 4.1. Introduction à la fluorescence et à la protéine fluorescente verte GFP

L'objectif de cette sous-partie est de donner quelques éléments qui permettront une meilleure compréhension des propriétés photophysiques de la GFP et de sa sensibilité environnementale. Cette partie a été écrite grâce au livre *Advanced Fluorescence Reporters in Chemistry and Biology*<sup>374</sup> et à la thèse de Romain Berardozi<sup>375</sup>.

#### 4.1.1. La fluorescence

Les électrons d'une molécule peuvent passer d'un niveau d'énergie électronique fondamental (noté  $E_0$ ) à un niveau électronique supérieur (notés  $E_1$ ,  $E_2$ ...) lorsque ceux-ci absorbent de l'énergie. Lorsqu'un électron est passé dans un niveau d'énergie supérieur, la molécule est dite excitée. Elle peut revenir à l'état fondamental en perdant cette énergie. Les molécules possèdent des sous-niveaux d'énergie dits vibrationnels et rotationnels, et la plupart des molécules transforment leur

énergie d'excitation en énergie de rotation ou de vibration. Cependant, certaines molécules ont la propriété d'émettre des photons lorsqu'elles se dés excitent. Elles émettent alors de la lumière, et on parle de luminescence moléculaire. Lorsque cette émission est consécutive à l'absorption d'un photon, il s'agit de photoluminescence, dont la forme la plus importante est la fluorescence. La relation entre l'absorption et l'émission de lumière d'une molécule, dite fluorophore ou fluorochrome, est illustrée dans le diagramme énergétique de Jablonski (figure 60.A). Lors de l'absorption des photons, un fluorophore est excité de son état fondamental ( $S_0$ ) à des états d'énergie électronique singlet plus élevés (par exemple  $S_1$  ou  $S_2$ ) à l'échelle de temps des femtosecondes. En quelques picosecondes, les électrons à l'état excité subissent une relaxation vibratoire non radiative et une conversion interne (c'est à dire entre états d'énergie électronique voisins) qui n'émettent pas de lumière pour retourner à l'état vibrationnel le plus bas de l'état  $S_1$ . À l'échelle nanoseconde, le fluorophore émet de la fluorescence en passant de l'état  $S_1$  vers un des niveaux vibrationnels de l'état fondamental. Il peut également retourner à l'état vibrationnel le plus bas de l'état fondamental par une relaxation vibratoire non radiative.



**Figure 60 :** **A.** Le diagramme énergétique de Jablonski illustre les états électroniques d'un fluorophore et les transitions entre eux. Les états électroniques sont indiqués par des lignes horizontales, les lignes épaisses représentant les états de base vibrationnels respectifs. Les transitions à l'intérieur et entre les états électroniques peuvent être non radiatives (flèches ondulées) ou radiatives (flèches droites), et se produisent à des échelles de temps différentes. Lorsque les électrons excités retournent à leur état fondamental à une échelle nanoseconde, ils émettent une fluorescence.  $S_0$ , état de masse ;  $S_1$  ou  $S_2$ , états d'énergie électronique singlet supérieure. **B.** L'émission de signaux de fluorescence se produit généralement à une longueur d'onde plus grande. Le décalage entre les longueurs d'onde d'excitation et d'émission est appelé décalage de Stokes. Adapté de Dunst, S., & Tomancak, P. (2019)<sup>376</sup>.

Lorsque les molécules d'un fluorophore reviennent à un état fondamental après avoir été excitées, elles ré-émettent les photons absorbés, produisant de la lumière. Ces derniers sont d'énergie inférieure aux photons d'excitation en raison d'une perte d'énergie due aux transitions non radiatives vers l'état vibrationnel le plus bas de l'état  $S_1$ . Lorsque les photons perdent de l'énergie, la longueur d'onde à laquelle ils sont émis augmente. La longueur d'onde de la lumière émise est

donc toujours supérieure à celle de la lumière absorbée. Ce déplacement est appelé décalage de Stokes (figure 60.B).

Plusieurs paramètres caractérisent les fluorophores (la figure 60.B résume les paramètres spécifiques à la GFP) :

- Les longueurs d'ondes de leurs pics d'excitation et d'émission.
- Leur durée de vie à l'état excité.
- L'efficacité de la fluorescence, appelée rendement quantique ( $\Phi$ ). Cette grandeur représente la probabilité d'émission d'un photon de fluorescence suite à l'absorption d'un photon et peut être estimée comme le rapport entre le nombre de photons émis par fluorescence et le nombre de photons absorbés. En raison de la compétition cinétique entre les différentes voies de désactivation radiatives et non radiatives de l'état excité  $S_1$ , l'émission d'un photon de fluorescence après excitation n'est qu'une possibilité parmi d'autres. La valeur de ce rapport varie de 0,05 à 1, avec 1 s'il y a autant de photons émis que de photons absorbés.
- Le coefficient d'extinction molaire ( $\epsilon$ ) lie la quantité de lumière absorbée, pour une longueur d'onde donnée, à la concentration molaire du fluorophore en solution. Plus il est grand, plus intense sera la fluorescence à intensité lumineuse incidente et rendement quantique égaux. Il est compris entre 5000 et 250000  $M^{-1}.cm^{-1}$ .
- Leur brillance, égale à  $\epsilon.\Phi$ . Elle caractérise donc l'intensité de la fluorescence pour une longueur d'onde donnée.

Les différentes caractéristiques de l'émission de fluorescence ne constituent la signature d'une molécule que sous certaines conditions. En effet, ces paramètres dépendent fortement de l'environnement local du fluorophore. Des changements plus ou moins importants peuvent se manifester sous l'effet de divers facteurs, tels que le pH, la polarité, la présence d'ions ou les éventuelles interactions avec d'autres molécules (quenchers).

Le quenching est un processus qui implique une diminution du rendement quantique d'un fluorochrome, souvent réversible. Ce phénomène provient d'une collision entre un fluorophore excité et une autre molécule. Cette collision produit une dissipation non radiative de l'énergie de retour de l'état excité à l'état fondamental, entraînant un arrêt de l'émission de la fluorescence par le fluorochrome du stade  $S_1$  à  $S_0$ . Des petites molécules comme l'oxygène, les halogènes, les amines ou les ions sont des agents quenchants par collision.

La température joue également un grand rôle dans l'émission de la fluorescence. À une longueur d'onde d'émission donnée, quand la température augmente, l'intensité de la fluorescence diminue. De plus, quand la température est basse, le risque de quenching diminue, car les molécules sont moins excitées et la probabilité de collision est plus faible.

#### 4.1.2. Origine et histoire de la GFP

La protéine fluorescente verte GFP (de l'anglais « *Green Fluorescent Protein* ») a été découverte et purifiée en 1962 par le groupe de Osamu Shimomura<sup>377</sup> comme protéine compagnon de l'aequorine, la fameuse protéine chimioluminescente bleue dépendante du calcium de la méduse *Aequorea victoria*. Au cours des deux décennies suivantes, il a été déterminé que l'aequorine et la GFP travaillent ensemble dans les organes clairs de la méduse pour convertir les signaux

luminescents induits par le calcium en une fluorescence verte caractéristique de l'espèce. Malgré des spéculations intéressantes, on ne comprend toujours pas la fonction biologique de la fluorescence de ces deux protéines, pourquoi les émissions vertes sont écologiquement plus intéressantes au bleu des émetteurs primaires, et dans quel intérêt biologique une GFP séparée a été sélectionnée évolutivement plutôt que de muter l'aequorine pour déplacer sa longueur d'onde d'émission<sup>378</sup>.

En 1992, Douglas Prasher a réussi à cloner le gène de la GFP et en 1994, Martin Chalfie a exprimé la séquence codante dans les cellules d'*E. coli*. Puis, au début du 21<sup>ème</sup> siècle, Roger Tsien a conçu des variantes de GFP de couleurs multiples. En 2008, Shimomura, Chalfie et Tsien ont reçu le prix Nobel de chimie pour ces réalisations. La première structure cristallographique de GFP a été résolue par James Remington en 1996. La protéine est aujourd'hui une des plus utilisées en biologie, notamment comme outil biotechnologique, particulièrement adapté en tant que protéine rapportrice pour la mesure du niveau d'expression et la cartographie spatiale de protéines d'intérêt chez de nombreuses espèces<sup>379</sup>.

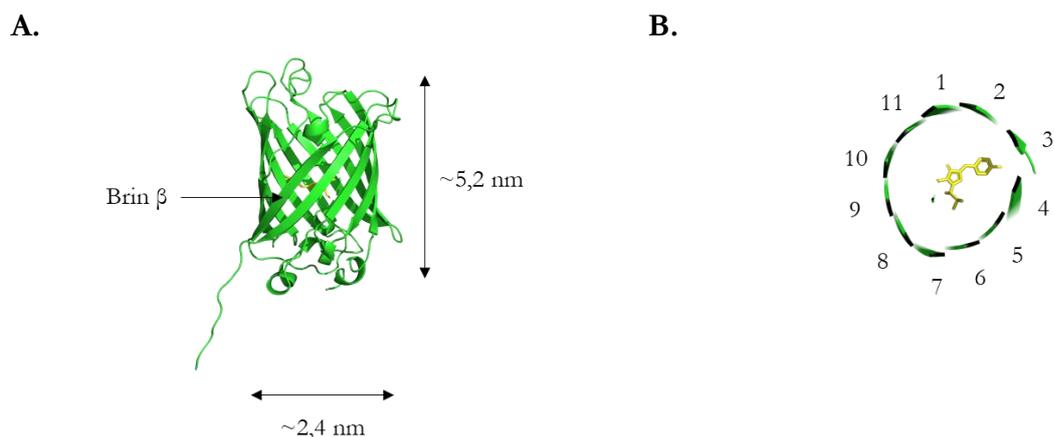
#### 4.1.3. Structure générale de la GFP

La GFP est composée de 11 brins  $\beta$  organisés en un tonneau soluble, formé de façon post-traductionnelle au cours du repliement de la protéine (figure 61.A). Les deux extrémités du tonneau  $\beta$  sont fermées par de courtes hélices  $\alpha$  formant des « couvercles ». L'extrémité N-terminale est engagée dans la formation d'une des hélices  $\alpha$  refermant le tonneau et joue un rôle central dans les phénomènes de repliement et de fluorescence comme le suggère la diminution drastique de fluorescence suite à la troncation d'un unique acide aminé en N-terminal<sup>380</sup>. À l'opposé, l'extrémité C-terminale forme une région flexible à l'extérieur du tonneau pouvant être modifiée sans importantes détériorations des propriétés de fluorescence. En général, la partie C-terminale tolère davantage les modifications si bien que les constructions en fusion C-terminale conduisent à moins d'artefacts<sup>381</sup>.

Le repliement de la GFP est fortement affecté par la température du milieu. La protéine étant issue d'une méduse vivant en eaux froides, elle est particulièrement thermo-intolérante au-delà de 25°C lors de sa production<sup>382</sup>. La sensibilité à la température est cependant limitée au processus de repliement. Une fois la GFP repliée correctement à basse température, elle reste stable et fluorescente à des températures atteignant au moins 65°C<sup>383</sup>.

Naturellement, la GFP a tendance à former des dimères<sup>384</sup>. Ces dimères résultent d'interactions hydrophobes entre les acides aminés apolaires (Ala206, Leu221 et Phe223) de deux monomères. Le remplacement de ces acides aminés par d'autres à chaîne latérale chargée (mutations par substitution Ala206Lys, Leu221Lys et Phe223Arg) réduit considérablement l'interaction et permet l'obtention d'un mutant monomérique.

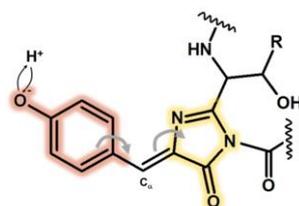
Le tonneau  $\beta$  est traversé par une hélice  $\alpha$  (figure 61) portant à son extrémité une triade d'acides aminés réagissant entre eux pour former un groupe hétérocycle à l'origine des propriétés d'absorption de la lumière : le chromophore (figure 61.B).



**Figure 61** : Structure de la protéine fluorescente GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale). **A.** Tonneau  $\beta$  composé de 11 brins antiparallèles contenant le chromophore en son centre. **B.** Vue en coupe transversale de l'intérieur du tonneau  $\beta$ . Les numéros représentent les 11 brins  $\beta$  formant le tonneau.

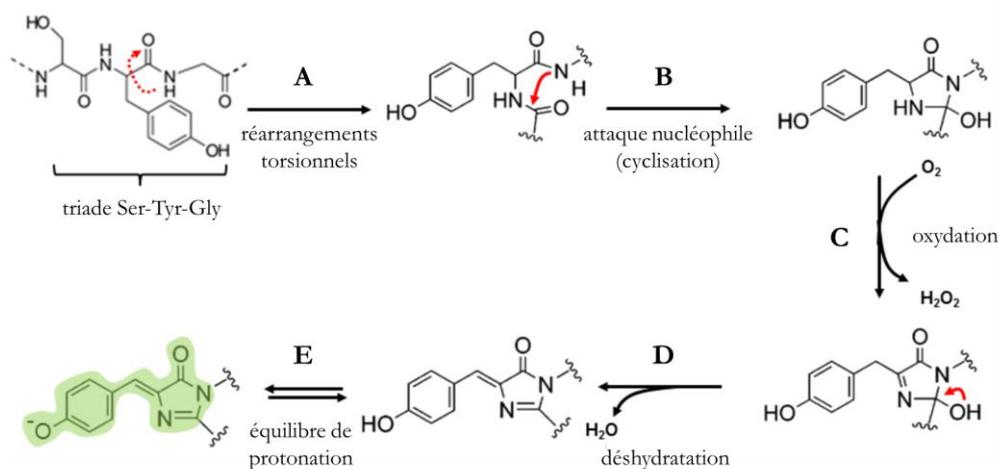
#### 4.1.4. Structure et étapes de maturation du chromophore

La réaction autocatalytique entre les trois acides aminés Ser65-Tyr66-Gly67 de la GFP, appelée « maturation », donne naissance à un 4-(p-hydrobenzylidène)-5-imidazolinone (p-HBDI) appelé chromophore. Le chromophore est composé de deux cycles, un homocycle aromatique hydrobenzyle (anneau  $\Phi$ ) et un hétérocycle imidazolinone (anneau I), reliés entre eux par un groupement propène dit pont méthylène en conformation *cis*-plane (figure 62). Les liaisons doubles des deux anneaux sont conjuguées par méthylène, permettant ainsi la délocalisation des électrons sur l'ensemble du chromophore à l'origine des propriétés de fluorescence.



**Figure 62** : Chromophore de la GFP. Le chromophore p-HBDI est composé de deux cycles : un homocycle aromatique (hydrobenzyle en orange) présentant une fonction titrable (OH/O<sup>-</sup>) et un hétérocycle azoté (imidazolinone en jaune) relié par un pont méthylène en conformation *cis*-plane. Le carbone méthylénique est nommé C<sub>α</sub>. Les flèches grises représentent les mouvements des électrons entre les deux cycles conjugués par le pont méthylène. Adapté d'une figure de Romain Berardozzi<sup>375</sup>.

Au cours de la première étape de la maturation, des réarrangements torsionnels affectent la triade d'acides aminés. Suite à cette étape, le carbonyle de la liaison peptidique entre la Ser65 et la Tyr66 pointe en direction de l'amine de la liaison peptidique entre la Gly67 et la Tyr66 (figure 63.A). Cette géométrie favorise l'attaque nucléophile du doublet non-liant de l'amine sur le carbone du carbonyle aboutissant à une cyclisation intra-moléculaire (figure 63.B). La cyclisation est favorisée par la présence de Gly (très faible encombrement stérique). L'amine portée par l'anneau I formée est ensuite oxydée en présence d'oxygène (figure 63.C). Une dernière étape de déshydratation permet d'enrichir le chromophore en électrons et la formation du pont méthylène entre l'anneau  $\Phi$  et I (figure 63.D). Cette liaison active alors la délocalisation des électrons le long du p-HBIDI responsable de la fluorescence.

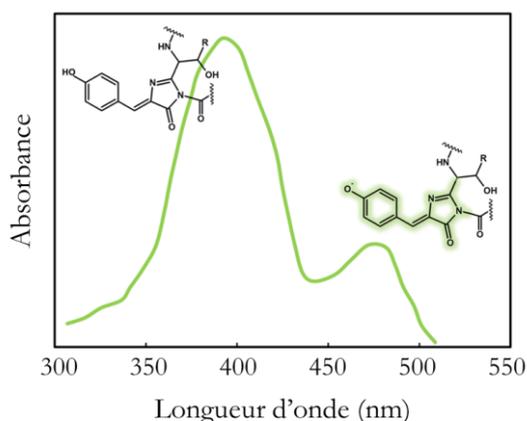


**Figure 63 : Maturation de la GFP.** Vue synoptique des réactions chimiques autocatalytiques intervenant lors de la maturation du chromophore. **A.** Réarrangements torsionnels de la triade d'acides aminés Ser65, Tyr66 et Gly67. **B.** Attaque nucléophile intra-moléculaire entre l'amine de la liaison peptidique 66-67 et le carbonyle de la liaison peptidique 65-66 aboutissant à une cyclisation intramoléculaire. **C.** Réaction d'oxydation de l'anneau I du chromophore **D.** Réaction de déshydratation aboutissant à la mise en place du pont méthylène conjugant les deux anneaux. Adapté d'une figure de Romain Berardozzi <sup>375</sup>.

La maturation étant particulièrement sensible aux conditions physico-chimiques de l'environnement, et la quantité de protéines fluorescentes fonctionnelles émettant des photons varie significativement selon le contexte.

Le chromophore de la GFP existe dans deux états : l'anneau  $\Phi$  porte un groupement protonable en équilibre entre une forme acide protonée (phénol) et une forme basique conjuguée déprotonée (phénolate). On parle d'équilibre de protonation (figure 63.E). Le spectre d'absorption de la GFP est alors constitué de deux bandes larges, non structurées. La forme protonée du chromophore (ou forme neutre) est l'état prédominant, et présente un pic majeur d'absorption à  $\sim 395$  nm. La forme non protonée (ou forme anionique) est moins répandue, et absorbe à environ  $\sim 475$  nm<sup>385</sup>. Seule la forme protonée anionique est fluorescente (figure 64). Le spectre d'émission, quant à lui, ne présente qu'une seule bande dont le maximum dépend de la longueur d'onde d'excitation. Une excitation à 395 nm conduit à l'obtention d'un spectre d'émission constitué d'une bande d'émission d'intensité faible, non structurée, située entre 420 et 470 nm et d'une bande intense centrée à 508 nm. Suite à une excitation dans la bande à 475 nm, seule la bande intense est observée dont le maximum d'émission est très légèrement déplacé vers le bleu et positionné à 503 nm.

L'incapacité du chromophore isolé de la GFP<sup>386</sup> et de la protéine dénaturée<sup>387</sup> à émettre de la lumière après excitation suggère que le repliement en tonneau  $\beta$  joue un rôle central dans le phénomène de la fluorescence. L'environnement protéique du chromophore permet d'une part de maintenir de la planéité du chromophore (essentielle à la fluorescence) et d'autre part de restreindre les phénomènes de quenching en compétition avec la voie radiative<sup>388</sup>.

**A****B**

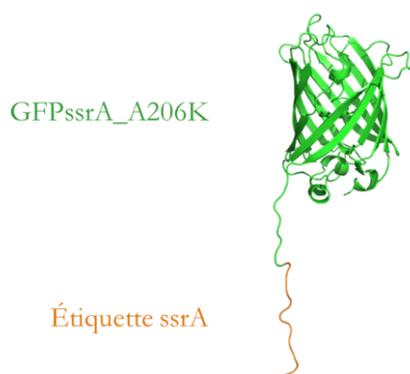
	Excitation (nm)	Émission (nm)	$\epsilon$ ( $M^{-1}.cm^{-1}$ )	$\phi_F$	Brillance ( $\epsilon.\phi$ , UA)
GFP neutre	395	508	25-30000	0,79	21600
GFP anionique	475	503	9,5-14000		~ 9400

**Figure 64** : Propriétés de fluorescence de la GFP. **A**. Spectre d'absorption de la GFP. Le spectre est composé d'une bande violette (maximum à 395 nm) et d'une bande bleue (maximum à 475 nm) correspondant respectivement aux formes neutre et anionique du chromophore. Adapté d'une figure de Romain Berardozi<sup>375</sup>. **B**. Tableau récapitulatif des propriétés photophysiques de la GFP.

#### 4.1.5. Fusion de la GFP avec l'étiquette d'adressage « *ssrA* »

Le substrat de la machinerie protéolytique utilisé a été adapté de la GFP\_A206K. Il s'agit donc d'un substrat modèle et non d'un substrat naturel de *M. jannaschii*. Ce choix repose sur les différentes propriétés de la protéine décrites dans les parties précédentes. En effet, le fait que la GFP soit une protéine thermostable est très utile puisque l'activité du système protéolytique sera déclenchée à 55°C lors des mesures fonctionnelles. On ne craint alors pas de dépliement thermique de la protéine, malgré une légère diminution de l'efficacité quantique de fluorescence à cette température. L'avantage principal est que le dépliement de la GFP est accompagné par une disparition de sa fluorescence, qui peut être suivi quantitativement pendant la réaction. On peut alors suivre l'activité du système protéolytique par mesure de la disparition de la fluorescence de la GFP.

L'étiquette d'adressage *ssrA* de 11 résidus (AANDENYALAA) (voir partie 3.4.2 du chapitre 1) a été ajoutée à l'extrémité C-terminale de GFP (GFP*ssrA*)<sup>389</sup> comme illustré dans la figure 65.

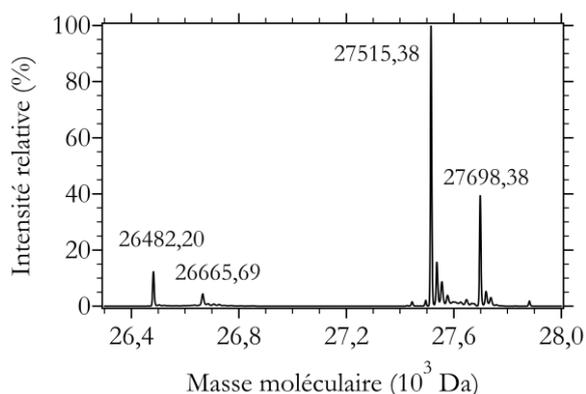


**Figure 65** : Structure de la protéine fluorescente GFP*ssrA* (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale).

Dans le cas du système protéolytique d'archées, des études précédentes ont montré que l'étiquette *ssrA* permet d'adresser les substrats à la machinerie de dégradation *in vitro*<sup>390</sup>. La GFP*ssrA* a donc été purifiée lors de ce projet de thèse.

#### 4.2. Qualité de la protéine GFP*ssrA* purifiée estimée par spectrométrie de masse

La pureté de la solution protéique purifiée a été estimée par spectrométrie de masse ESI. Les données issues de l'analyse sont présentées dans la figure 66 de la même façon que précédemment (partie 2.2).



**Figure 66 : Spectre de masse de la GFP*ssrA* purifiée.** Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire.

Le spectre de masse montre un premier pic d'une masse moléculaire de 27515,4 Da avec un pic secondaire correspondant à une masse moléculaire plus lourde de 183 Da (adduit de Pefabloc). Un second pic beaucoup moins intense d'une masse moléculaire de 26482,2 Da est présent, possédant lui-même un pic secondaire correspondant à une masse moléculaire plus lourde de 183 Da (adduit de Pefabloc).

En prenant en compte la perte de la méthionine initiale lors de la traduction de la protéine, la masse moléculaire théorique de la GFP*ssrA* est de 27535 Da. De plus, la maturation du chromophore (partie 4.1.4) est un phénomène entraînant une perte de 20 Da. La masse moléculaire serait donc alors de 27515 Da, ce qui coïncide exactement avec la masse moléculaire observée. En plus de ces phénomènes, la perte des 10 derniers acides aminés de l'extrémité C-terminale de la protéine (ANDENYALAA) entraînerait une masse moléculaire de 26482 Da, ce qui est exactement la masse moléculaire du pic minoritaire. Cette forme correspond à la seconde bande observée sur gel SDS-PAGE (partie 1).

Les 10 derniers acides aminés de l'extrémité C-terminale correspondent à la quasi-totalité de l'étiquette d'adressage *ssrA*. D'après l'abondance relative des différents pics, on peut donc conclure que **la GFP*ssrA* produite est pure avec une proportion d'environ 10% de la population ayant perdu l'étiquette d'adressage *ssrA***. On peut donc s'attendre à ce que cette proportion de protéines ne soit pas adressable au système protéolytique. La perte de l'étiquette d'adressage *ssrA* est un phénomène très problématique pour le projet puisqu'il a été observé que la GFP*ssrA* n'est pas prise en charge par le système protéolytique lorsque toute la population de GFP*ssrA* possède une étiquette tronquée.

Ces données correspondent à la meilleure qualité de la GFP<sub>ssrA</sub> obtenue, avec moins de 10% de la population de GFP<sub>ssrA</sub> ayant perdu l'étiquette d'adressage *ssrA*. La suite de cette partie présente l'étude ayant permis de comprendre comment optimiser la production de cette protéine pour limiter la perte de l'étiquette d'adressage.

#### 4.3. Étude de l'effet de la température sur la perte de la séquence d'adressage « *ssrA* »

La GFP<sub>ssrA</sub> (et les mutants GFP<sub>ssrA</sub>\_G10C et GFP<sub>ssrA</sub>\_H231C décrits plus tard) peut perdre la majorité de leur étiquette d'adressage. La perte de l'étiquette d'adressage peut être visualisée sur gel SDS-PAGE à 15% d'acrylamide sous la forme d'une seconde bande de masse moléculaire plus faible que la GFP<sub>ssrA</sub> intacte (figure 48) et par spectrométrie de masse (figure 66). J'ai constaté au cours de ce projet que les conditions d'expression de la protéine dans la bactérie *E. coli*, mais aussi les conditions de purification et de stockage, influençaient la perte de l'étiquette *ssrA*.

Différentes conditions d'expression ont été testées afin de déterminer quelles conditions étaient les plus favorables à l'expression d'une population de GFP<sub>ssrA</sub> intacte. Plusieurs temps d'expression après induction à l'IPTG (5, 20, 22 et 24h) ont été testés à 20 et à 30°C et des échantillons ont été déposés sur gel SDS-PAGE à 15% d'acrylamide (figure 67).



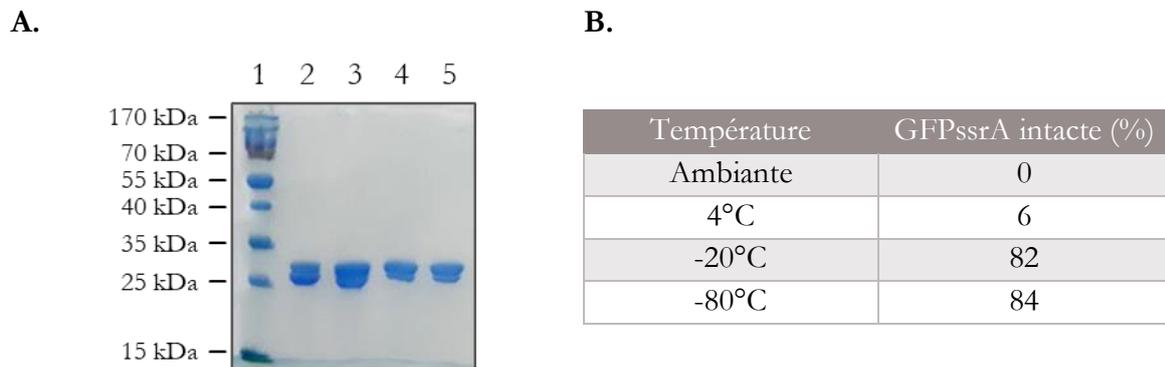
**Figure 67 : Test d'expression de la GFP<sub>ssrA</sub>.** SDS-PAGE, 15 % acrylamide, coloration à l'Instant Blue. Le puits 1 correspond aux marqueurs de masse moléculaire. Les échantillons sont numérotés de 2 à 10 et correspondent aux conditions : échantillon avant induction (2), induction à 20°C pendant 5h (3), 20h (4), 22h (5) ou 24h (6), et induction à 30°C pendant 5h (7), 20h (8), 22h (9) ou 24h (10).

On peut observer pour toutes les conditions d'expression testées une bande de masse moléculaire d'environ 27 kDa sur gel SDS-PAGE à 15% d'acrylamide. Cela correspond à la GFP<sub>ssrA</sub> surexprimée. Cependant, les bandes correspondant aux conditions d'expression à 30°C sont plus basses que celles des conditions d'expression à 20°C. Il semblerait donc qu'une partie de la GFP<sub>ssrA</sub> possède une séquence d'adressage tronquée lorsqu'elle est exprimée à 30°C, alors qu'une température de 20°C semble plus favorable pour l'expression de GFP<sub>ssrA</sub> intacte.

La GFP<sub>ssrA</sub> a été purifiée directement à partir du culot bactérien sans stockage de celui-ci à -80°C pour limiter l'action de protéases lors de la congélation et décongélation du culot bactérien. La GFP<sub>ssrA</sub> peut également perdre sa séquence d'adressage au cours de sa purification, l'ajout d'inhibiteurs de protéases dans le tampon de lyse ne semblant pas empêcher ce phénomène. Les nombreuses purifications de GFP<sub>ssrA</sub> réalisées au cours de cette thèse ont permis de déterminer que la qualité de la GFP<sub>ssrA</sub> dépendait directement de la température à laquelle était réalisée la purification. Une purification à température ambiante entraîne une perte de la séquence

d'adressage, alors qu'une purification à 4°C minimise ce phénomène. Malgré une température d'expression de 20°C et une purification à 4°C, environ 10 à 20% de la population de GFPssrA purifiée est sous forme tronquée.

L'influence de la température de stockage de la GFPssrA purifiée a également été étudiée. Des échantillons de GFPssrA purifiée ont été conservés pendant un mois à température ambiante, 4°C, -20°C ou -80°C puis déposés sur gel SDS-PAGE à 15% d'acrylamide (figure 68.A) et mesurés par spectrométrie de masse (figure 68.B).



**Figure 68 : Formes intacte et tronquée de la GFPssrA après 1 mois d'incubation à différentes températures.**  
**A.** SDS-PAGE, 15 % acrylamide, coloration à l'Instant Blue. Environ 2 µg de chaque protéine ont été déposés sur gel. Le puits 1 correspond aux marqueurs de masse moléculaire. Les échantillons sont numérotés de 2 à 5 et correspondent aux conditions : GFPssrA stockée 1 mois à température ambiante (2), 4°C (3), -20°C (4) ou -80°C (5). **B.** Estimation quantitative de la proportion de GFPssrA intacte dans les différentes conditions d'après les intensités relatives des pics observés en ESI-MS. La forme tronquée présente une perte des 10 derniers acides aminés de l'extrémité C-terminale de la protéine (étiquette ssrA).

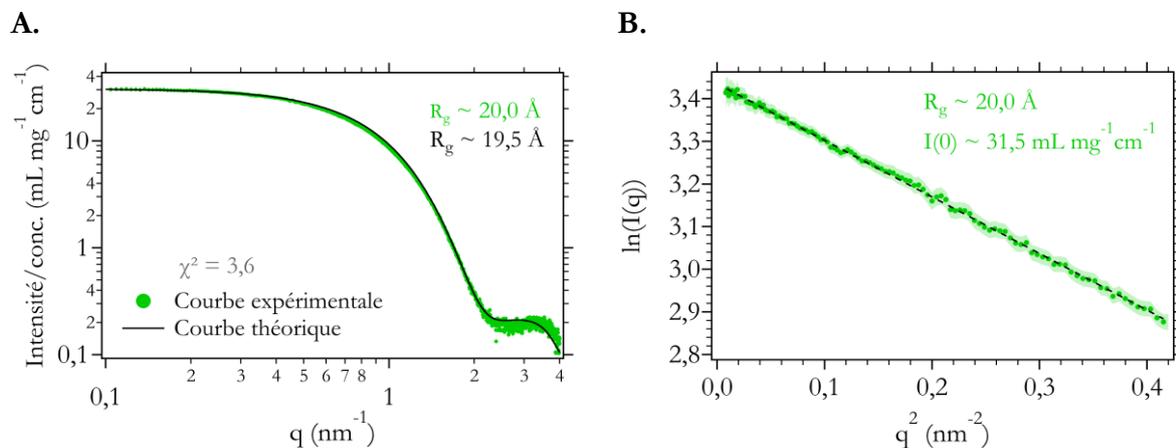
Lorsque la GFPssrA a été stockée à -20°C ou à -80°C, la bande de faible masse moléculaire correspondant à la GFPssrA sans étiquette d'adressage observée sur gel SDS-PAGE à 15% d'acrylamide est fine (et similaire à celle observée dans la GFPssrA fraîchement préparée). Lorsque la GFPssrA a été stockée à 4°C ou température ambiante, cette bande est épaisse. L'estimation de la proportion des différentes formes de GFPssrA par spectrométrie de masse a révélé qu'environ 80-85% de la population de GFPssrA stockée à -20°C et -80°C conservait son étiquette d'adressage (ce qui était la proportion initiale), tandis que 0-5% de la population de GFPssrA stockée à 4°C et à température ambiante la conservait. A l'instar de la température de d'expression et de purification, la température de stockage de la GFPssrA a donc également une importance cruciale.

On peut donc conclure de ces études que **la perte de l'étiquette d'adressage ssrA est liée à la température d'expression, de purification et de stockage de la protéine.** La protéine a été exprimée à 20°C, purifiée à 4°C et stockée à -80°C. On suppose que la troncature de l'étiquette d'adressage, toujours à la même position, est due à une coupure par une protéase qui, malgré la purification, reste en faible quantité dans la solution protéique. L'étiquette est en effet une séquence flexible accessible aux protéases. Le programme ExPaSy PeptideCutter n'a pas permis de déterminer de quelle protéase il pourrait s'agir d'après le site de coupure.

#### 4.4. Étude de la GFPssrA par SAXS

La GFPssrA a été étudiée par SAXS afin de vérifier la qualité du signal et des informations obtenues en prévision des mesures SANS. En effet, les mesures SANS suivront le signal de la GFPssrA

deutérée (voir chapitre 7). La figure 69.A représente la courbe SAXS de la GFPssrA ainsi que la courbe théorique de la GFPssrA calculée à partir de la structure cristalline d'une GFP avec une séquence flexible ajoutée à l'extrémité C-terminale (afin de représenter l'étiquette ssrA) grâce au logiciel CRY SOL. La figure 69.B représente la zone de Guinier de la courbe expérimentale ayant permis de calculer le  $R_g$  et  $I(0)$  grâce à l'approximation de Guinier (équation 13).



**Figure 69 : Mesures SAXS sur la GFPssrA.** **A.** Courbe expérimentale SAXS de la GFPssrA mesurée sur BM29 (ESRF) et courbe SAXS théorique calculée par le logiciel CRY SOL à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale). **B.** Région de Guinier de la courbe expérimentale SAXS de la GFPssrA. Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans les deux figures, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

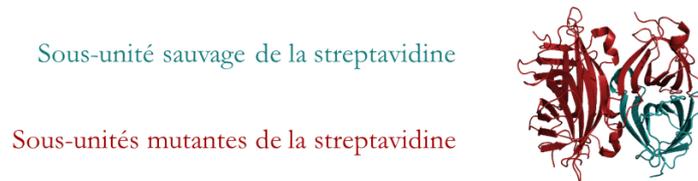
La courbe SAXS mesurée est superposable à la courbe théorique pour tous les vecteurs de diffusion  $q$  représentés. De plus, le rayon de giration obtenu expérimentalement (20 Å) est en bon accord avec le rayon de giration théorique (19,5 Å). Les valeurs de  $I(0)$  présentées dans les figures 69.B et 69.C indiquent que la masse moléculaire moyenne est de  $M_w$  serait d'environ 30 kDa (voir partie 2.3.2). Cela qui est en accord avec la masse moléculaire théorique de la GFPssrA (27,7 kDa) et avec les précédentes mesures (partie 4.1).

On peut donc conclure que **la GFPssrA produite permet d'obtenir un signal SAXS de qualité et d'obtenir des informations ( $R_g$  et  $M_w$ ) cohérentes avec les valeurs théoriques.** La GFPssrA deutérée a donc pu être préparée dans les mêmes conditions en prévision des mesures SANS.

## 5. Étude de la GFPssrA\_Nter-biotine, de la GFPssrA\_Cter-biotine et de la streptavidine monovalente

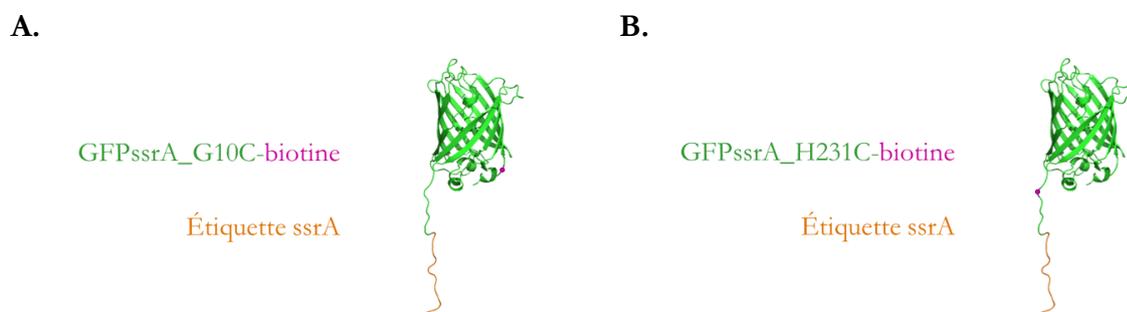
### 5.1. Présentation des protéines étudiées

Une approche utilisant de la GFPssrA couplée à de l'avidine par l'intermédiaire d'une molécule de biotine a été utilisée dans de précédents travaux pour bloquer la GFPssrA au cours de son dépliement par PAN<sup>317</sup>. Un tétramère d'avidine fixant quatre molécules de biotine et donc quatre GFPssrA, nous avons choisi d'utiliser une streptavidine monovalente au lieu de l'avidine, où une seule des quatre sous-unités est capable de lier la biotine<sup>262</sup>. Les raisons de ce choix (principalement la simplification de l'interprétation des données SANS) et le contexte de cette étude sont expliqués de façon plus approfondie dans le chapitre 8. De la streptavidine monovalente a donc été préparée au laboratoire (figure 70).



**Figure 70 : Structure de la streptavidine monovalente** (PDB 3RY1 streptavidine sauvage).

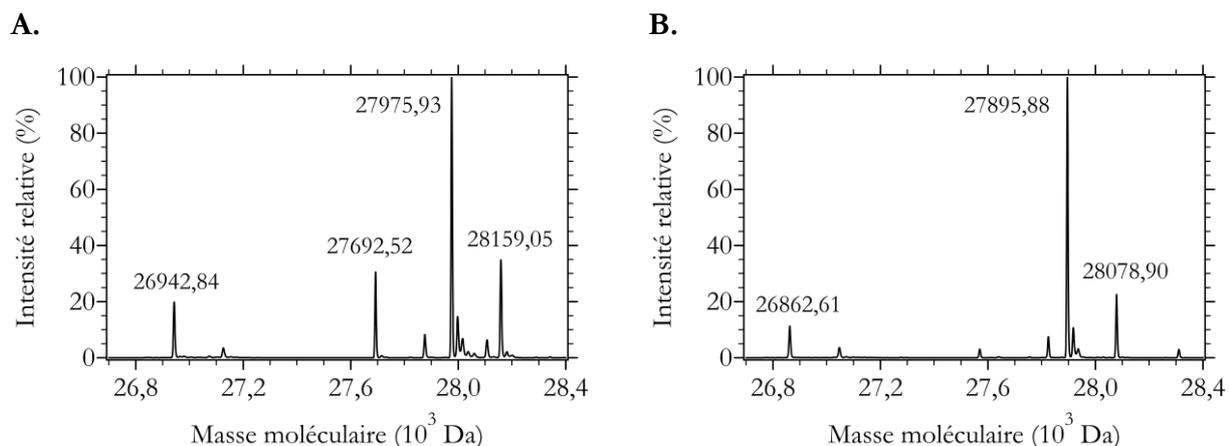
L'extrémité modifiée de la biotine utilisée (Biotin PEO Iodoacetamide) est réactive avec les groupements thiols, absents naturellement de la surface de la GFPssrA. Afin de permettre la liaison à la biotine, des mutants de la GFPssrA ont été conçus par substitution pour introduire une cystéine (thiol) proche de son extrémité N-terminale (GFPssrA\_G10C) ou C-terminale (GFPssrA\_H231C). Les raisons de ces choix sont également expliquées de façon plus approfondie dans le chapitre 8. Les GFPssrA mutées ont été purifiées et liées à la biotine (figure 71).



**Figure 71 : Structure de la protéine fluorescente GFPssrA** (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale).

## 5.2. Qualité des GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine purifiées estimée par spectrométrie de masse

La pureté des solutions protéiques purifiées a été estimée par spectrométrie de masse ESI. Les données issues de l'analyse sont présentées de la même façon que précédemment (partie 2.2) dans la figure 72.A. pour la GFPssrA\_Nter-biotine et dans la figure 72.B. pour la GFPssrA\_Cter-biotine.



**Figure 72 : Spectre de masse de la GFPssrA\_Nter-biotine purifiée (A) et de la GFPssrA\_Cter-biotine purifiée.** Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire.

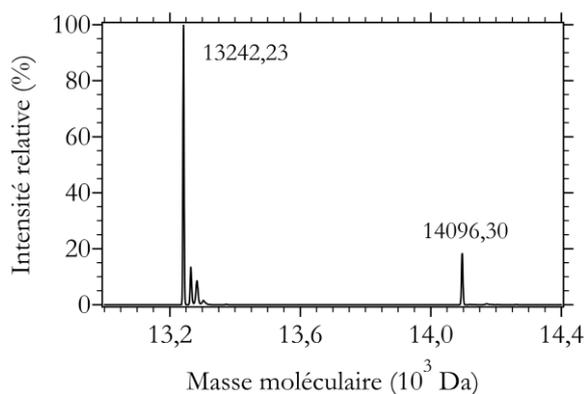
Les résultats montrent un pic principal d'une masse moléculaire de 27975,93 Da (figure 72.A) et de 27895,88 Da (figure 72.B) avec chacune un pic secondaire correspondant à une masse moléculaire plus lourde de 183 Da (adduits Pefabloc). Un second pic beaucoup moins intense est visible, correspondant à une masse moléculaire de 26942,84 Da (figure 72.A) et de 26862,61 Da (figure 72.B). Enfin, la figure 72.A présente un quatrième pic correspondant à une masse moléculaire de 27692,52 Da.

Les masses moléculaires théoriques de la GFPssrA\_G10C et de la GFPssrA\_H231C, avec la perte de la méthionine initiale et la maturation du chromophore (voir partie 5.1), sont respectivement de 27561,11 Da et de 27481,02 Da.. De plus, la biotine liée aux protéines a une masse moléculaire de 542,43 Da, mais sa liaison aux protéines entraîne une perte de 128 Da (perte d'une molécule d'iode et d'un hydrogène), la masse moléculaire ajoutée serait donc de 414,43 Da. Les masses moléculaires de la GFPssrA\_Nter-biotine et de la GFPssrA\_Cter-biotine seraient donc respectivement de 27975,54 Da et de 27895,43 Da, ce qui concorde avec les masses moléculaires observées des pics principaux. La perte des 10 derniers acides aminés de l'extrémité C-terminale des protéines entraînerait des masses moléculaires respectives de 26942,29 Da et 26862,2 Da, qui sont exactement les masses moléculaires des pics minoritaires observés. Enfin, le quatrième pic de la figure 72.A correspond à une molécule de GFPssrA\_Nter-biotine ayant perdu deux acides aminés à l'extrémité N-terminale et un acide aminé à l'extrémité C-terminale, probablement dû une dégradation de la protéine par des protéases malgré l'ajout d'inhibiteurs de protéases dans le tampon de lyse. Cette espèce, possédant toujours 10 des 11 acides aminés de l'étiquette d'adressage ssrA, devrait être prise en charge par la machinerie protéolytique.

D'après l'abondance relative des différents pics, on peut donc conclure que **les GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine produites sont pures avec une proportion d'environ 10% de la population ayant perdu l'étiquette d'adressage ssrA**. On peut donc s'attendre à ce que cette proportion de protéines ne soit pas adressable au système protéolytique. Ces protéines ont pu être utilisées pour la formation de complexe avec la streptavidine monovalente.

### 5.3. Qualité de la streptavidine monovalente purifiée estimée par spectrométrie de masse

La pureté de la solution protéique purifiée a été estimée par spectrométrie de masse ESI. Les données issues de l'analyse sont présentées dans la figure 73 tel que dans la partie 2.2.



**Figure 73 : Spectre de masse de la streptavidine monovalente purifiée.** Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire.

Le spectre de masse présente deux pics, le premier d'une masse moléculaire de 13242,12 Da et le second d'une masse moléculaire de 14095,95 Da. Ces masses moléculaires estimées devraient correspondre aux monomères des sous-unités de la streptavidine monovalente.

La masse moléculaire théorique du monomère d'une sous-unité sauvage de la streptavidine monovalente est de 14225,46 Da tandis que la masse moléculaire théorique du monomère de la sous-unité triple mutante est de 13371,60 Da. Avec la perte de la méthionine à l'extrémité N-terminale (131 Da) lors de la traduction de la protéine, les masses moléculaires seraient respectivement de 14094,5 Da et de 13240,4 Da. Ces masses moléculaires correspondent à celles observées.

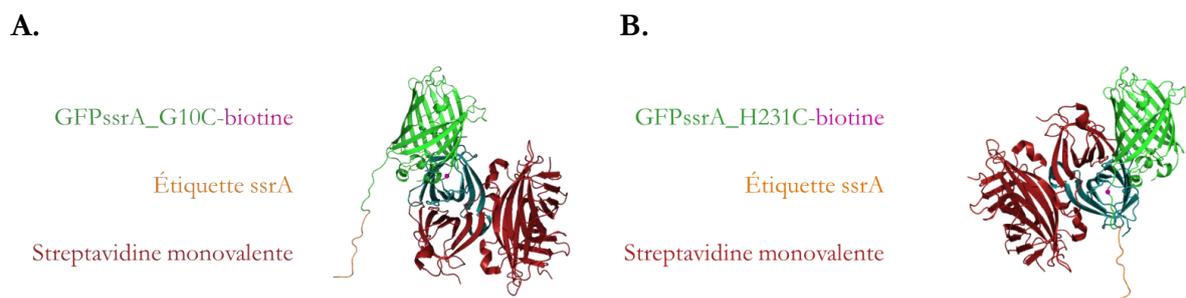
L'abondance relative de la sous-unité mutante est environ trois fois plus importante que celle de la sous-unité sauvage, suggérant que la streptavidine purifiée est bien monovalente (voir partie 5.1).

On peut donc conclure de façon générale que **la streptavidine monovalente produite est pure en solution**. En effet, aucun contaminant n'a été détecté et la masse moléculaire estimée correspond à celle des monomères des sous-unités sauvages et mutantes. La protéine purifiée a pu être utilisée pour la formation de complexe avec les GFPssrA biotinylées.

## 6. Étude de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine

### 6.1. Présentation des protéines étudiées

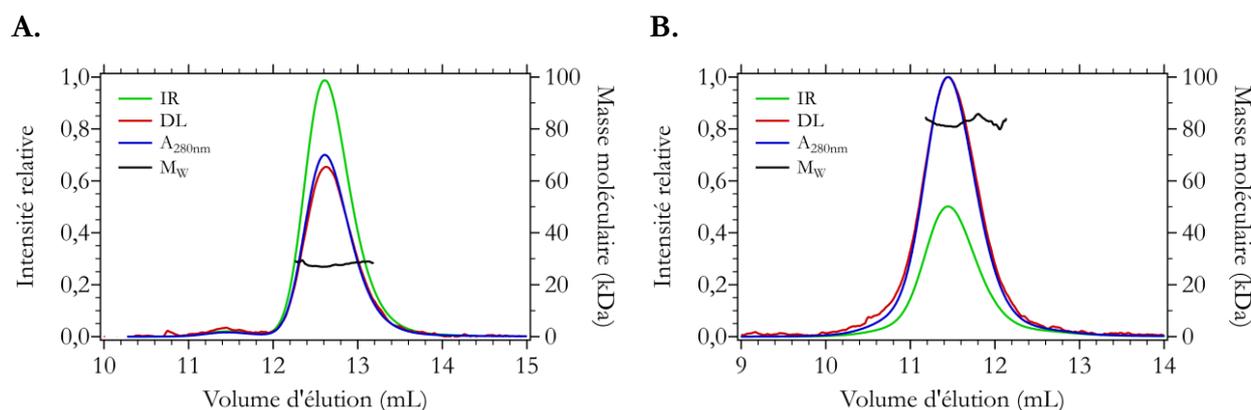
Une fois la qualité de la GFPssrA\_Nter-biotine, GFPssrA\_Cter-biotine et streptavidine monovalente assurée, les protéines ont été associées dans le but de produire de la GFPssrA\_Nter-streptavidine ou de la GFPssrA\_Cter-streptavidine. La liaison entre l'avidine et la biotine est l'interaction biochimique (non covalente) la plus forte retrouvée dans la nature ( $K_D = 10^{-15}$ )<sup>391</sup>, et la liaison entre la biotine et la sous-unité sauvage de la streptavidine monovalente est également très forte ( $K_D = 10^{-14}$ )<sup>358</sup>. La figure 74 illustre les deux complexes préparés.



**Figure 74 : Représentation de la GFPssrA\_Nter-streptavidine (A), de la GFPssrA\_Cter-streptavidine (B).** (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale). La sous-unité sauvage de la streptavidine monovalente est attachée proche de l'extrémité N-terminale de la GFPssrA\_G10C et proche de C-terminale de la GFPssrA\_H231C par l'intermédiaire de la biotine.

## 6.2. Étude par SEC-MALLS des protéines des complexes GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine

Des mesures SEC-MALLS ont été réalisées afin d'estimer la masse moléculaire des protéines GFPssrA-biotine et GFPssrA-streptavidine et déterminer s'il y a bien une molécule de GFPssrA\_Nter-biotine ou GFPssrA\_Cter-biotine couplée à une molécule de streptavidine. La figure 75 représente les profils de SEC-MALLS de la GFPssrA\_Cter-biotine et de la GFPssrA\_Cter-streptavidine comme représenté précédemment.



**Figure 75 : Mesures SEC-MALLS sur la GFPssrA\_Cter-biotine et la GFPssrA\_Cter-streptavidine.** Profil SEC-MALLS de la GFPssrA\_Cter-biotine (**A**) et de la GFPssrA\_Cter-streptavidine (**B**). Les courbes d'indice de réfraction (vert), d'absorbance à 280 nm (bleu) et de diffusion de la lumière (rouge) ont été tracées en fonction du volume d'élution à travers une colonne Superose 12 10/300 GL. Les intensités ont été normalisées par rapport à la valeur maximale de l'échantillon le plus concentré. La masse moléculaire en kDa correspondant aux différents pics d'élution ont été représentées par une ligne noire. Les profils pour la GFPssrA\_Nter-biotine et de la GFPssrA\_Nter-streptavidine sont similaires.

Tableau 17 : Masses moléculaires théoriques calculées à partir de la séquence en acides aminés des protéines étudiées grâce au programme ExPASy ProtParam Tool® et masses moléculaires moyennes expérimentales calculées par le logiciel ASTRA6 à partir des mesures SEC-MALLS.

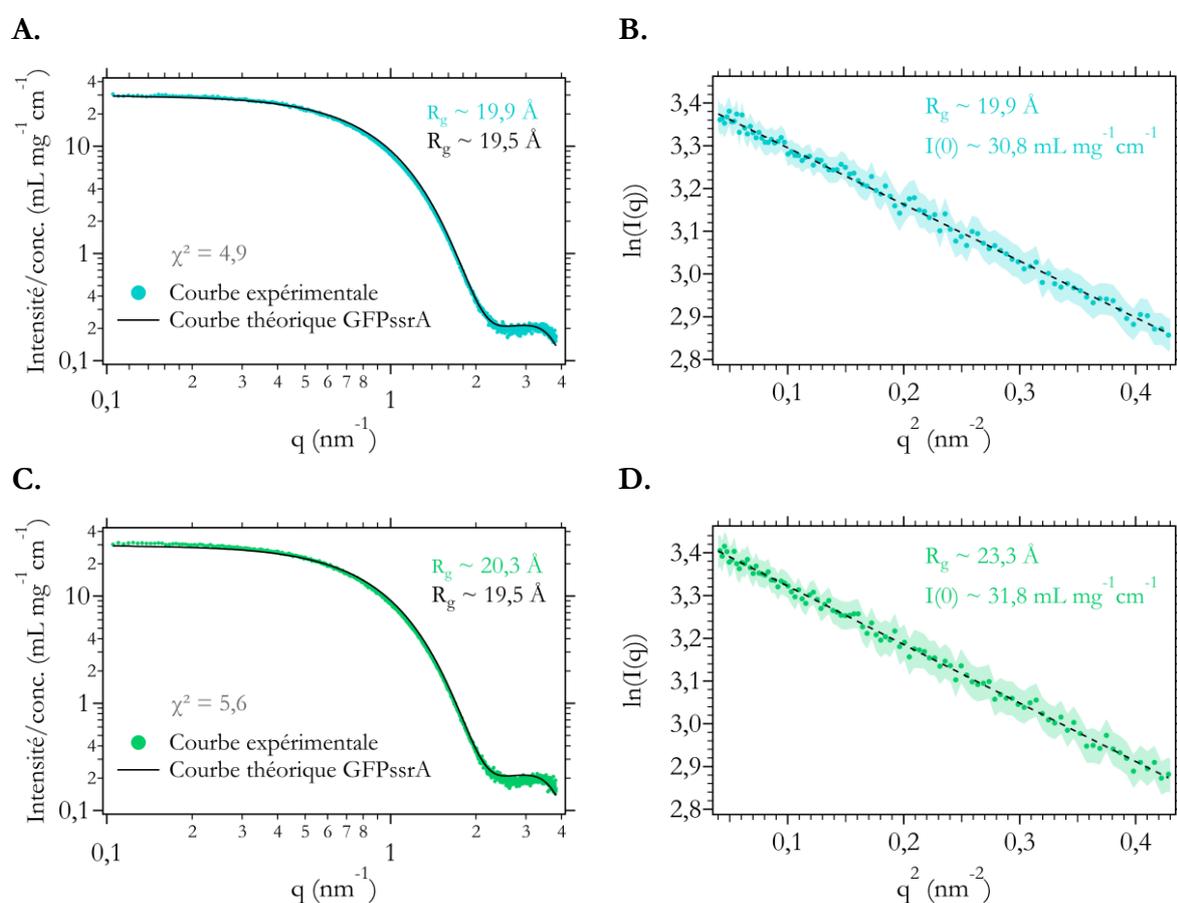
Échantillon	Masse moléculaire théorique (kDa)	Masse moléculaire expérimentale (kDa)
GFPssrA_Nter-biotine	27,9	27,8 (± 1,8 %)
GFPssrA_Cter-biotine	27,9	27,1 (± 2,17 %)
GFPssrA_Nter-streptavidine	82,2	81,2 (± 2,65 %)
GFPssrA_Cter-streptavidine	82,2	82,8 (± 1,14 %)

Un seul pic d'élution a été observé pour toutes les protéines mesurées, indiquant que les protéines sont pures en solution comme suggéré par les autres techniques. De plus, les pics d'élution sont symétriques avec la ligne correspondant à la masse moléculaire plate et un indice de polydispersité  $M_w/M_n \approx 1,000$ , ce qui signifie que les protéines sont monodisperses. Enfin, les masses moléculaires expérimentales coïncident avec les masses moléculaires théoriques. On peut donc conclure que :

- Les solutions protéiques de GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine et GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine sont pures et monodisperses.
- Il y a bien une molécule de GFPssrA\_Nter-biotine ou GFPssrA\_Cter-biotine liée à une molécule de streptavidine puisque la masse moléculaire mesurée correspond à une molécule de GFPssrA-biotine (27,9 kDa) et une molécule de streptavidine (54,3 kDa).

### 6.3. Étude par SAXS des protéines des complexes GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine

Les protéines GFPssrA-G10C-biotine et GFPssrA-Cter-biotine ont été étudiées par SAXS (figure 76) pour s'assurer de la qualité du signal et des informations obtenues en prévision des mesures SANS. Les données sont représentées de la même manière que pour la GFPssrA (partie 4.4).

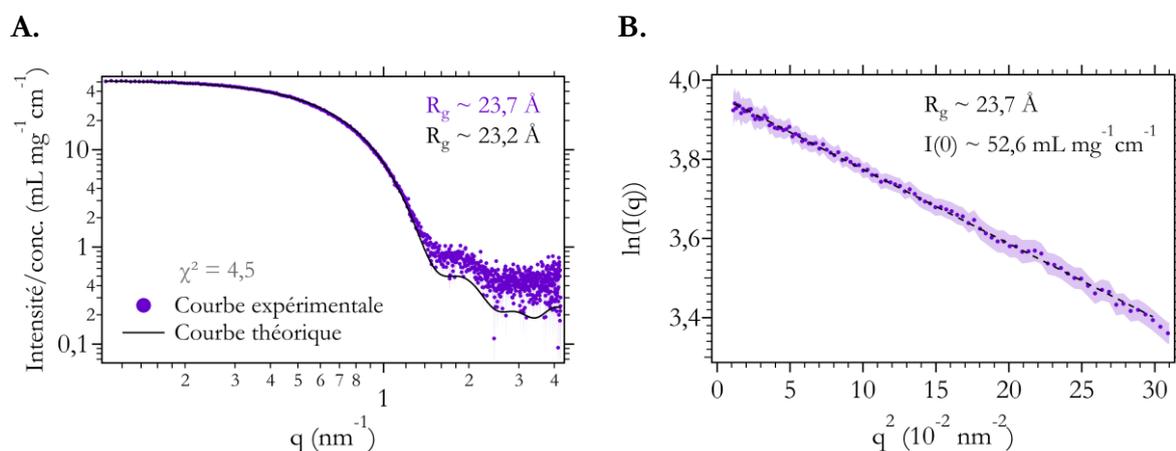


**Figure 76 : Mesures SAXS sur la GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine.** **A, C.** Courbes expérimentales SAXS de la GFPssrA\_Nter-biotine (**A**) et GFPssrA\_Cter-biotine (**B**) mesurées sur BM29 (ESRF) et courbe SAXS théorique calculée par le logiciel CRYSOLOG à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale). **B, C.** Région de Guinier des courbes expérimentales SAXS de la GFPssrA\_Nter-biotine (**C**) et GFPssrA\_Cter-biotine (**D**). Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans toutes les figures, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

Les courbes SAXS mesurées sont superposables à la courbe théorique pour tous les vecteurs de diffusion  $q$  représentés. De plus, les rayons de giration obtenus expérimentalement pour la GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine (respectivement 19,9 Å & 20,3 Å) sont en bon

accord avec le rayon de giration théorique (19,5 Å). Enfin, la masse moléculaire moyenne  $M_w$  calculée serait d'environ 30 kDa pour les deux protéines, ce qui correspond à la masse moléculaire théorique (27,9 kDa) et est cohérent avec les précédentes mesures (partie 5.1 et 6.2).

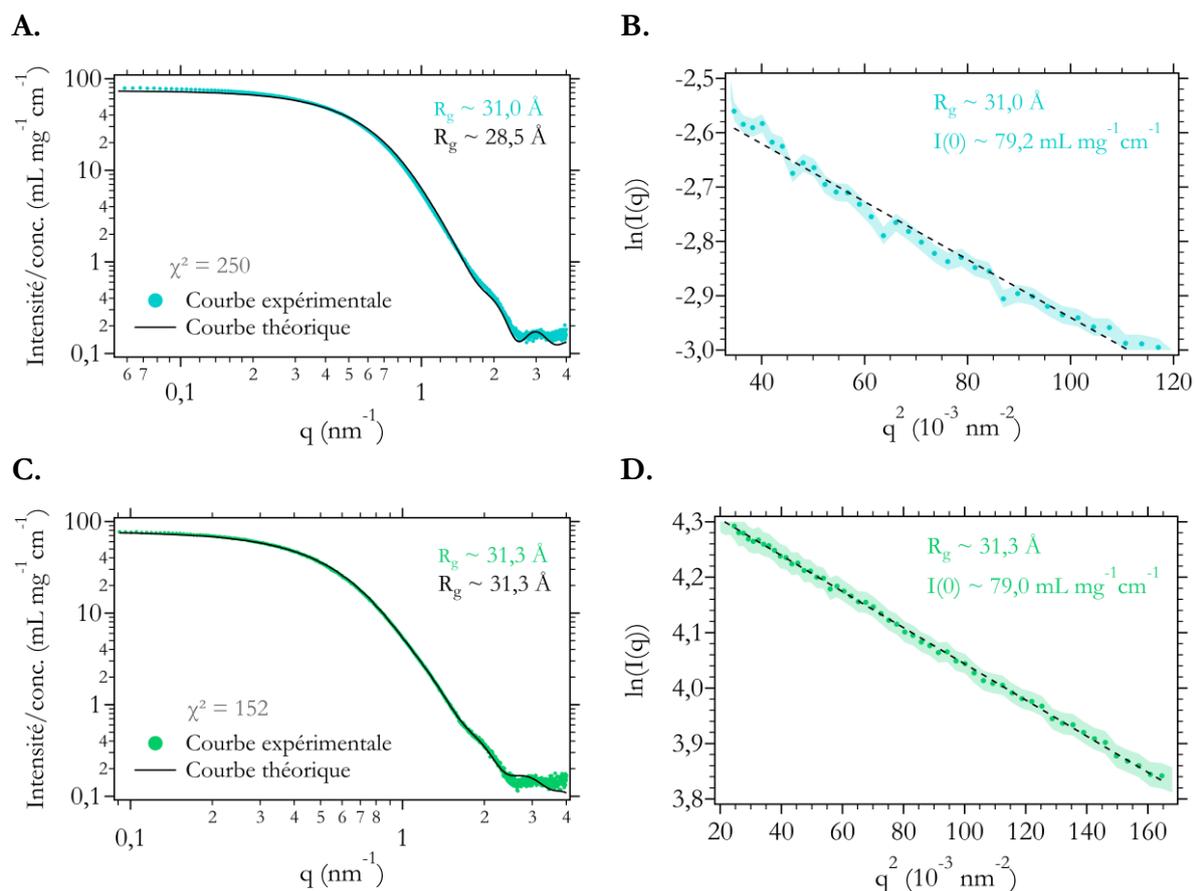
La streptavidine monovalente a également été mesurée par SAXS (figure 77) dans le but de mieux la caractériser. Les données sont représentées de la même manière que précédemment et la courbe théorique de la streptavidine a été calculée à partir de la structure cristalline de la streptavidine sauvage grâce au logiciel CRY SOL.



**Figure 77 : Mesures SAXS sur la streptavidine monovalente. A.** Courbe SAXS expérimentale de la streptavidine monovalente mesurée sur BM29 (ESRF) et courbe SAXS théorique calculée par le logiciel CRY SOL à partir de la structure cristalline de la streptavidine sauvage (PDB 3RY1). **B.** Région de Guinier de la courbe expérimentale SAXS de la streptavidine monovalente. Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans les deux figures, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

La courbe expérimentale SAXS obtenue apparaît superposée avec celle calculée jusqu'au vecteur de diffusion  $q = 0,12 \text{ nm}^{-1}$ . Le décalage à partir de  $q = 0,12 \text{ nm}^{-1}$  peut s'expliquer par la flexibilité conformationnelle de l'étiquette histidine de la sous-unité sauvage de la streptavidine monovalente, ou par une soustraction trop faible du tampon par rapport à l'échantillon (voir partie 1.2.3 du chapitre 2). Les mutations des sous-unités mutantes de la streptavidine monovalente ne devraient pas avoir d'impact sur le profil SAXS de la protéine, qui donne une information sur la structure globale de la protéine. De plus, le rayon de giration obtenu expérimentalement (23,7 Å) est cohérent avec le rayon de giration théorique (23,2 Å). Enfin, la masse moléculaire moyenne  $M_w$  calculée serait d'environ 50 kDa pour les deux protéines, ce qui correspond à la masse moléculaire théorique (54,3 kDa) et est cohérent avec les précédentes mesures (partie 5.2).

Les protéines GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine ont également été étudiées par SAXS (figure 78) afin de mieux les caractériser. Les données sont représentées de la même manière que précédemment et une courbe théorique générée grâce au logiciel SASREF pour chaque protéine à partir des structures cristallines de la streptavidine sauvage et de la GFP avec une séquence flexible ajoutée à l'extrémité C-terminale.



**Figure 78 : Mesures SAXS sur la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine. A, C.** Courbes expérimentales SAXS de la GFPssrA\_Nter-streptavidine (A) et GFPssrA\_Cter-streptavidine (B) mesurées sur BM29 (ESRF) et courbe SAXS théorique calculée par le logiciel SASREF à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) et de la structure cristalline de la streptavidine sauvage (PDB 3RY1). **B, C.** Région de Guinier des courbes expérimentales SAXS de la GFPssrA\_Nter-streptavidine (C) et GFPssrA\_Cter-streptavidine (D). Une droite superposée aux points expérimentaux est représentée par un trait pointillé. Dans toutes les figures, les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

Les courbes SAXS mesurées sont relativement bien superposables aux courbes théoriques jusqu'au vecteur de diffusion  $q = 2 \text{ nm}^{-1}$ . Le décalage à partir de  $q = 2 \text{ nm}^{-1}$  peut être dû à l'étiquette histidine présente sur la sous-unité sauvage de la streptavidine monovalente, ou dû au logiciel SASREF qui ne permet pas de fournir qu'une estimation approximative de la structure globale des complexes. Les rayons de giration obtenus expérimentalement pour la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine (respectivement  $31,0 \text{ \AA}$  &  $31,3 \text{ \AA}$ ) sont cohérents par rapport aux valeurs théoriques obtenues (respectivement  $28,5 \text{ \AA}$  &  $31,3 \text{ \AA}$ ). Enfin, la masse moléculaire moyenne  $M_w$  calculée serait d'environ  $79 \text{ kDa}$  pour les deux protéines, ce qui est en accord avec la masse moléculaire théorique ( $82,2 \text{ kDa}$ ) et cohérent avec les précédentes mesures (partie 6.2).

On peut donc conclure des différentes mesures SAXS que :

- **Les GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine produites permettent d'obtenir un signal SAXS de qualité et d'obtenir des informations ( $R_g$  et  $M_w$ ) cohérentes avec les valeurs théoriques.** Les GFPssrA-biotine deutérées ont donc pu être préparées dans les mêmes conditions en prévision des mesures SANS.

- **La streptavidine monovalente, ainsi que les GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine, offrent un signal SAXS de qualité avec des informations ( $R_g$  et  $M_w$ ) en accord avec les valeurs théoriques.**

## 7. Conclusion du chapitre

Le but principal de ce chapitre était de présenter toutes les protéines étudiées et d'exposer les résultats ayant permis une meilleure compréhension de leur comportement en solution et une optimisation de leur préparation.

La qualité des protéines purifiées a été validée par SDS-PAGE à 15% d'acrylamide et spectrométrie de masse ESI. Les mesures de SAXS, de SEC-MALLS et de microscopie électronique à transmission en coloration négative ont permis d'étudier les différentes protéines.

La protéine PAN a été préparée sous forme majoritairement hexamérique, sa forme active, par ajout de 100 mM ATP et incubation à 55°C dans le tampon de travail à 20 mM Tris-HCl pH7,5, 100 mM NaCl et 10 mM MgCl<sub>2</sub>. Les autres protéines ont été préparées dans le même tampon. La particule 20S a été préparée de façon à minimiser la contamination par les anneaux double de sous-unités  $\alpha$ . Les protéines GFPssrA, GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine ont été purifiées et stockées à basse température afin de limiter la perte de l'étiquette d'adressage ssrA à environ 10% de leur population. La streptavidine monovalente a été purifiée et chaque molécule a été liée à une GFPssrA\_Nter-biotine ou GFPssrA\_Cter-biotine afin d'obtenir respectivement de la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine. De plus, la qualité du signal SAXS de la GFPssrA, GFPssrA\_Nter-biotine et de la GFPssrA\_Cter-biotine a été validée en prévision des mesures SANS réalisées sur leur version deutérées.

La reconstitution du protéasome PAN-20S, même si partiellement visualisable en microscopie électronique, ainsi que la préparation des différentes versions de substrats, a permis d'envisager la réalisation d'études fonctionnelles. L'activité du système PAN, particule 20S et GFPssrA est décrite dans les chapitres 6 et 7, tandis que l'activité du système PAN, particule 20S et GFPssrA\_Nter-streptavidine ou GFPssrA\_Cter-streptavidine est décrite dans le chapitre 8.



## Chapitre 6 :

# Mesures biochimiques sur le système GFP<sub>ssrA</sub>, protéasome PAN-20S

Ce chapitre décrit les expériences effectuées afin de caractériser l'activité du système GFP<sub>ssrA</sub>, protéasome PAN-20S dans les conditions d'études nécessaires aux mesures SANS. Dans le contexte de notre étude *in vitro*, il est important de maîtriser les conditions d'activité du système et de définir les conditions permettant une activité optimale pour les mesures SANS. La concentration des composants, la température de la réaction, et le temps de mesure ont été sélectionnés suite aux différentes expériences présentées dans ce chapitre. Des mesures de la consommation d'ATP par PAN au cours du temps ainsi que l'identification des peptides produits par le protéasome PAN-20S ont permis d'apporter des informations complémentaires à celle apportée lors de l'étude par TR-SANS.

### 1. Activité de dépliement et de dégradation : mesures par spectroscopie de fluorescence et de dégradation de la GFP<sub>ssrA</sub> par Western-blot anti-GFP

#### 1.1. Avant-propos

La particule régulatrice PAN possède une activité ATPase et est nommée comme protéine AAA+<sup>247,392</sup>. L'énergie apportée par l'hydrolyse de l'ATP avec le cofacteur Mg<sup>2+</sup> lui permet de déplier les substrats protéiques<sup>261,262</sup>. Cette activité représente la fonction principale de PAN. Une fois les protéines substrat reconnues, elles sont dépliées puis transférées dans la partie catalytique 20S qui les découpe alors en peptides<sup>277</sup> (voir partie 5 du chapitre 1).

La GFP<sub>ssrA</sub> étant une protéine fluorescente, la mesure de la fluorescence est un bon indicateur pour suivre la dénaturation de ce substrat au cours du temps<sup>261,389</sup>. La fluorescence de la GFP<sub>ssrA</sub> a été mesurée grâce à un spectrofluorimètre lecteur de plaque avec une longueur d'onde d'excitation  $\lambda_{ex} = 400$  nm et d'émission  $\lambda_{em} = 509$  nm. Les réactions ont été suivies pendant une heure avec un point de mesure de la fluorescence toutes les 40 secondes. Afin de permettre l'activité de PAN et de la particule 20S, les mesures ont été réalisées à 55°C. Cette température a été choisie pour correspondre à des travaux précédents<sup>40</sup>.

Dans les conditions d'étude de l'activité de dépliement et de dégradation de la GFP<sub>ssrA</sub> par le protéasome PAN-20S, les concentrations relatives des différents protagonistes de la réaction sont très proches. De ce fait, la réaction échappe à une analyse de type michaelienne qui suppose une concentration d'enzyme beaucoup plus faible que la concentration en substrat afin de traiter de façon simple la vitesse initiale où la concentration en complexe enzyme-substrat est constante en fonction du temps ( $d[ES]/dt = 0$ , équation 22)<sup>359</sup>. Qui plus est, l'équation dite de Michaelis-Menten (équation 25) décrit des cinétiques de réaction où un substrat est transformé

irréversiblement en un produit en une étape<sup>359</sup> (voir partie 4.1 du chapitre 3). Ceci ne peut être appliqué à la dégradation de la GFPssrA qui nécessite un dépliement et un transfert préalable, le tout par le système complexe composé de PAN et de la particule 20S. La dégradation de la GFPssrA par le complexe PAN-20S sera donc traitée de manière qualitative comme le constat d'un mécanisme processif.

## 1.2. Mesures par spectroscopie de fluorescence

Deux conditions (appelées « conditions SANS 1 et 2 ») ont été adoptées pour les mesures SANS. Le choix de ces conditions provient des expériences décrites dans la partie 2 de ce chapitre. Les mesures présentées ici ont été réalisées après la sélection des conditions SANS 1 et 2 mais sont délibérément présentées en premier. Cela permettra au lecteur de ce manuscrit de mieux comprendre le test fonctionnel classiquement réalisé et l'analyse qui en découle.

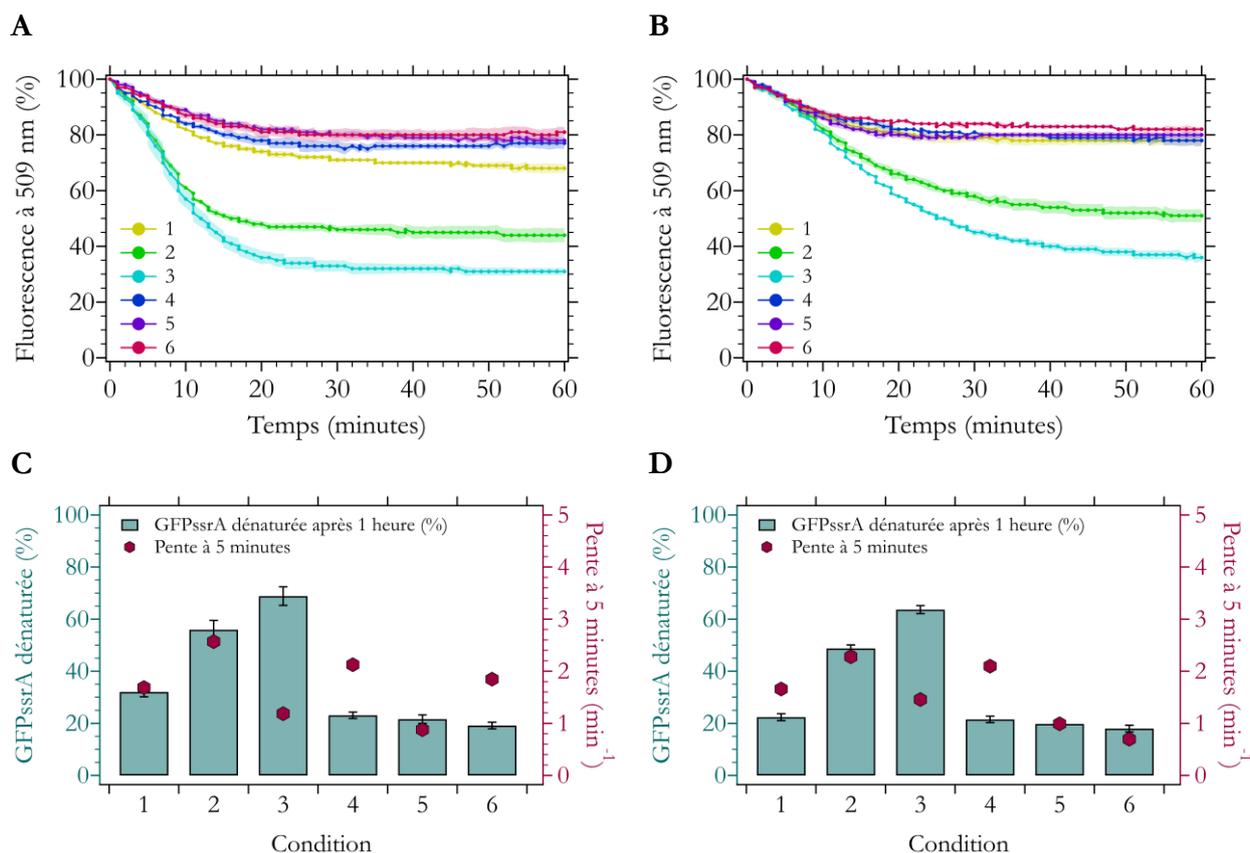
Un test a été effectué dans la condition SANS 1 (composition décrite partie 2.3) et dans la condition SANS 2 (composition décrite partie 2.3) avec différents contrôles. Les conditions suivantes ont été testées :

- (1) GFPssrA en présence de 100 mM ATP
- (2) GFPssrA et PAN en présence de 100 mM ATP
- (3) GFPssrA, PAN et particule 20S en présence de 100 mM ATP
- (4) GFPssrA et particule 20S en présence de 100 mM ATP
- (5) GFPssrA et PAN
- (6) GFPssrA, PAN et particule 20S

On considère tout au long de ce chapitre qu'en présence de 100 mM ATP, la concentration de MgCl<sub>2</sub> était de 200 mM.

Les données représentées dans les figures 79.A, 1.B correspondent aux courbes d'évolution de la fluorescence en fonction du temps dans les différentes conditions testées et permettent d'avoir une appréciation qualitative de l'activité du système. Chaque courbe de fluorescence de la GFPssrA en fonction du temps a été normalisée en définissant le premier point de mesure comme correspondant à 100% de fluorescence.

Les figures 79.C, 79.D correspondent à l'analyse de ces courbes avec les valeurs sous forme d'histogrammes représentant le pourcentage de GFPssrA dénaturée ayant perdu sa fluorescence après 1 heure à 55°C pour les différentes conditions testées. Les valeurs représentées sous forme de points correspondant à la pente de la tangente à la courbe après 5 minutes de réaction. Cela représente la vitesse de dénaturation de la GFPssrA en début de réaction. Ces deux paramètres permettent d'estimer l'efficacité du système. Chaque condition a été mesurée en triplicata afin de s'assurer de la reproductibilité des mesures.



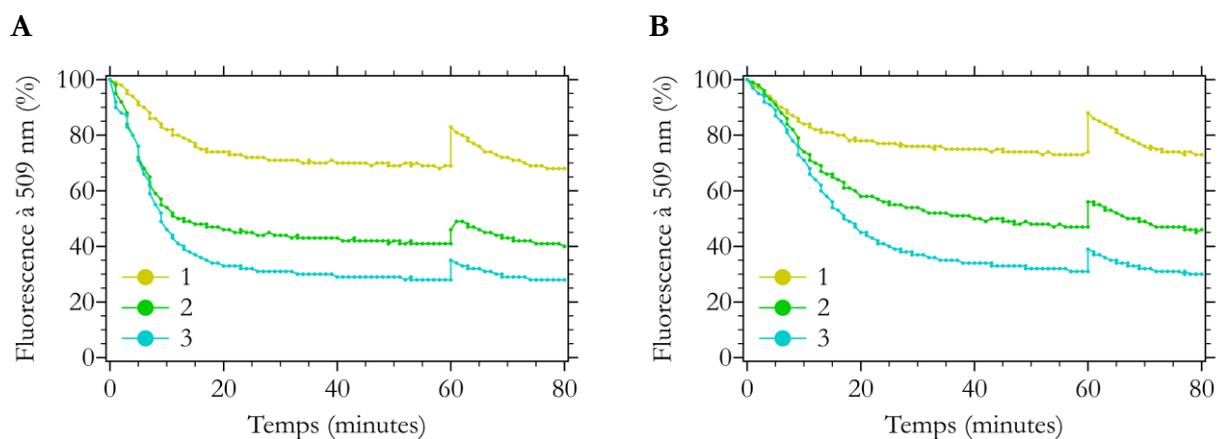
**Figure 79 : Mesure de l'activité du système à 55°C dans les deux conditions SANS.** Les figures A. et C. représentent les résultats pour la condition SANS 1 et les figures B. et D. représentent les résultats pour la condition SANS 2. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6, ce qui correspond aux réactions suivantes : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA, PAN et particule 20S, ou (4) GFPssrA et particule 20S, toutes quatre à 100 mM ATP, et (5) GFPssrA et PAN et (6) GFPssrA, PAN et particule 20S toutes deux sans ATP. **A, B.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire de la même couleur. **C, D.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les résultats obtenus dans les conditions SANS 1 et 2 montrent les mêmes tendances. Dans les conditions où la GFPssrA est seule (1) ou en présence de particule 20S avec 100 mM ATP (4), et dans toutes les conditions sans ATP (5 et 6), le pourcentage de GFPssrA dénaturée est de 20 à 30% tandis que la pente à 5 minutes est de 1 à 2  $\text{min}^{-1}$ . Ces valeurs correspondent à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C mais ne sont pas liées à une dénaturation de la protéine<sup>250,383</sup>. Il est intéressant de noter qu'une incubation de 10 minutes à température ambiante après une heure à 55°C permet de rétablir la majorité de la fluorescence initiale (figure 80).

En présence de PAN à 100 mM ATP (2), le pourcentage de GFPssrA dénaturée après 1 heure de réaction est d'environ 50% pour les deux conditions SANS tandis que la pente à 5 minutes est d'environ 3,5  $\text{min}^{-1}$  dans la condition SANS 1 et de 1,3  $\text{min}^{-1}$  dans la condition SANS 2. Cette augmentation de pourcentage de GFPssrA dénaturée en fin de réaction indique une activité de dépliement de la GFPssrA par PAN. La dénaturation de la GFPssrA est plus progressive dans les conditions SANS 2 avec une pente moins importante que dans les conditions SANS 1 mais un

pourcentage final de GFPssrA dénaturée similaire. Ce phénomène sera mieux décrit dans la partie 2.1.

Lorsque la GFPssrA est avec PAN et la particule 20S (3) à 100 mM ATP, le pourcentage de GFPssrA dénaturée en fin de réaction est d'environ 65% pour les deux conditions SANS tandis que la pente à 5 minutes est d'environ  $3,5 \text{ min}^{-1}$  dans la condition SANS 1 et de  $1,5 \text{ min}^{-1}$  dans la condition SANS 2. L'ajout de la particule 20S permet de visualiser dans les deux conditions une augmentation significative du pourcentage final de GFPssrA dénaturée sans observer néanmoins d'évolution significative de la pente à 5 minutes. Le système apparaît donc plus efficace lorsque la concentration de particules 20S augmente dans la solution. Il est intéressant de noter qu'en présence de PAN avec ou sans particules 20S, une incubation de 10 minutes à température ambiante ne permet de restaurer qu'une faible partie de la fluorescence (figure 79), suggérant que les molécules de GFPssrA ont été dénaturées de façon irréversible.



**Figure 80 : Effet de l'incubation à température ambiante après 1 heure à 55°C sur le rétablissement de la fluorescence à 509 nm de la GFPssrA dans les deux conditions SANS.** Les conditions testées sont numérotées de 1 à 3 selon les réactions suivantes en présence 100 mM ATP : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA, PAN et particule 20S. Après 1 heure à 55°C, les réactions ont été incubées à température ambiante pendant 10 minutes avant de les mesurer de nouveau de la même façon pendant 20 minutes. **A.** Résultats pour la condition SANS 1 (70  $\mu\text{M}$  GFPssrA, 20  $\mu\text{M}$  PAN et 20  $\mu\text{M}$  particule 20S). **B.** Résultats pour la condition SANS 2 (180  $\mu\text{M}$  GFPssrA, 5  $\mu\text{M}$  PAN et 30  $\mu\text{M}$  particule 20S).

On peut tirer les conclusions suivantes de ces mesures dans les conditions testées :

- **La présence d'ATP est nécessaire à l'activité du système protéolytique dans les conditions d'étude.** En effet, l'activité de dépliage de la GFPssrA par PAN (avec ou sans particule 20S) n'est détectée qu'en présence d'ATP, comme décrit la littérature<sup>261,277,392</sup>.
- **La particule 20S n'est pas suffisante à la dénaturation de la GFPssrA** puisqu'il n'y a pas d'activité visible de dénaturation de la GFPssrA en présence de particule 20S seule en présence d'ATP. Les protéines natives doivent donc être dépliées par PAN afin d'être prises en charge par la particule catalytique 20S, comme le suggère la littérature<sup>277,392</sup>.
- **La présence de particule 20S en solution permet de rendre le dépliage de la GFPssrA par PAN plus efficace.** Ce comportement pourrait être expliqué par un effet coopératif de la particule 20S avec PAN. Par exemple, la particule 20S permettrait à PAN de déplier les protéines substrat plus rapidement en les stabilisant pendant leur transfert<sup>286</sup>. Une autre hypothèse serait que la particule 20S permet de rendre la reconnaissance des substrats

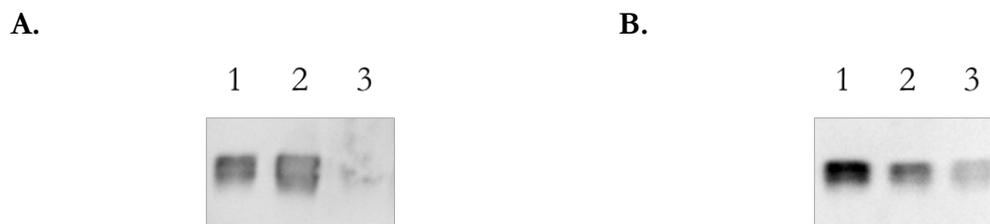
protéiques par PAN plus rapide en dégradant les protéines dépliées qui restent en solution en son absence<sup>261</sup>.

- **La GFPssrA seule présente une perte de fluorescence de 20-30% qui n'est pas liée à une dénaturation car elle est majoritairement réversible. Elle est due à la diminution de l'efficacité quantique de sa fluorescence à 55°C.** Il s'agit d'un phénomène amplement décrit dans la littérature<sup>374,383</sup>. Sur les 20-30% de la population perdant sa fluorescence, 15% des GFPssrA la retrouve après incubation 10 minutes à température ambiante.
- **La perte de fluorescence de la GFPssrA due à la dénaturation de la protéine en présence PAN et de particule 20S est majoritairement irréversible.** Une proportion de 5 à 10% de la population de la GFPssrA retrouve de la fluorescence après incubation 10 minutes à température ambiante, alors que 15% de la GFPssrA seule récupère sa fluorescence dans les mêmes conditions. La perte de fluorescence liée à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA ne peut donc pas être soustraite des mesures de dénaturation de la GFPssrA par le système protéolytique.

### 1.3. Étude de la dégradation de la GFPssrA visualisé par Western-blot anti-GFP

La spectroscopie de fluorescence ne permet pas de distinguer la différence entre les molécules de GFPssrA dépliées et les molécules de GFPssrA dégradées. L'immuno-détection par Western-blot avec un anticorps primaire anti-GFP permet de détecter sélectivement la GFPssrA native afin de suivre sa dégradation. Cette technique a l'avantage de démontrer facilement et rapidement en laboratoire si la machinerie protéolytique utilisée est fonctionnelle dans les conditions étudiées.

Un Western-blot avec un anticorps primaire anti-GFP a été réalisé à partir des produits finaux des trois principales réactions des deux conditions SANS (voir partie 2.3) en présence de 100 mM ATP (figure 81) : (1) GFPssrA, (2) GFPssrA et PAN et (3) GFPssrA, PAN et particule 20S.



**Figure 81 : Western-blot dans les deux conditions SANS 1 et 2** avec anticorps primaire anti-GFP et anticorps secondaire anti-lapin conjugué à la HRP après SDS-PAGE, 15 % acrylamide. Le signal a été révélé par une exposition de 30 secondes après trempage de la membrane dans des solutions de luminol et de peroxyde puis visualisé grâce à une caméra CCD. Les échantillons sont numérotés de 1 à 3 et représentent les conditions en présence 100 mM ATP : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA, PAN et particule 20S. **A.** Condition SANS 1. **B.** Condition SANS 2.

L'analyse est similaire pour les deux conditions SANS. Deux bandes sont visibles dans les échantillons contenant de la GFPssrA seule (1) ou avec PAN (2). Ces bandes correspondent à la GFPssrA intacte et GFPssrA ayant perdue une partie de l'étiquette d'adressage (voir partie 4 du chapitre 5), correspondant à de la GFPssrA. Dans les échantillons contenant le protéasome PAN-20S, les bandes correspondant à la GFPssrA disparaissent. Cela signifie que la GFPssrA a été en majorité dégradée à l'issue de la réaction et n'est ainsi plus visualisée par Western-blot.

On peut tirer quelques conclusions de cette analyse, dans les conditions testées :

- **La dégradation de la GFPssrA n'a lieu qu'en présence de la particule catalytique 20S dans les conditions testées**, comme le suggère la littérature<sup>390</sup>.
- **La GFPssrA est majoritairement dégradée après 1 heure à 55°C en présence de PAN et de la particule 20S** dans les deux conditions d'études, y compris lorsque la GFPssrA est en large excès par rapport à PAN.
- **Les deux formes de GFPssrA sont dégradées par le système protéolytique.** La forme plus légère correspondant à une population de GFPssrA ayant perdu une partie de l'étiquette d'adressage ssrA disparaît au même titre que celle correspondant aux GFPssrA intactes. Il est possible que la GFPssrA ayant perdu une partie de sa séquence signal soit tout de même adressable au protéasome PAN-20S car elle possède encore une partie de l'étiquette, qui est fixée sur la partie flexible C-terminale de la protéine. Or, il a été décrit que PAN reconnaît les protéines possédant une séquence flexible suivi d'un dégron<sup>280</sup>. Pour vérifier cette hypothèse, il faudrait générer plusieurs mutants de GFPssrA dont l'étiquette aurait été partiellement tronquée et réaliser des études fonctionnelles telles que celles décrites ici. On pourrait alors déterminer la taille minimale de l'étiquette d'adressage nécessaire au dépliement et à la dégradation de la protéine.

## 2. Recherche par spectroscopie de fluorescence des conditions de stœchiométries optimales entre les composants du mélange réactionnel en vue des expériences de TR-SANS

Le but principal du criblage réalisé était d'optimiser les conditions de concentration des différents éléments pour les mesures TR-SANS. En effet, des mesures en temps réel SANS requièrent d'avoir un temps d'exposition le plus court possible (quelques dizaines de secondes par point de mesure) tout en gardant une concentration suffisante en protéine d'intérêt pour obtenir un signal analysable. La réaction devait donc être efficace, mais suffisamment lente pour suivre la réaction avec un point de mesure toutes les 30 secondes pendant 1 heure. Pour cela, différentes concentrations de GFPssrA, PAN, particule 20S, ATP et MgCl<sub>2</sub> ont été combinées.

### 2.1. Criblage des stœchiométries GFPssrA et PAN à différentes concentrations d'ATP et de MgCl<sub>2</sub>

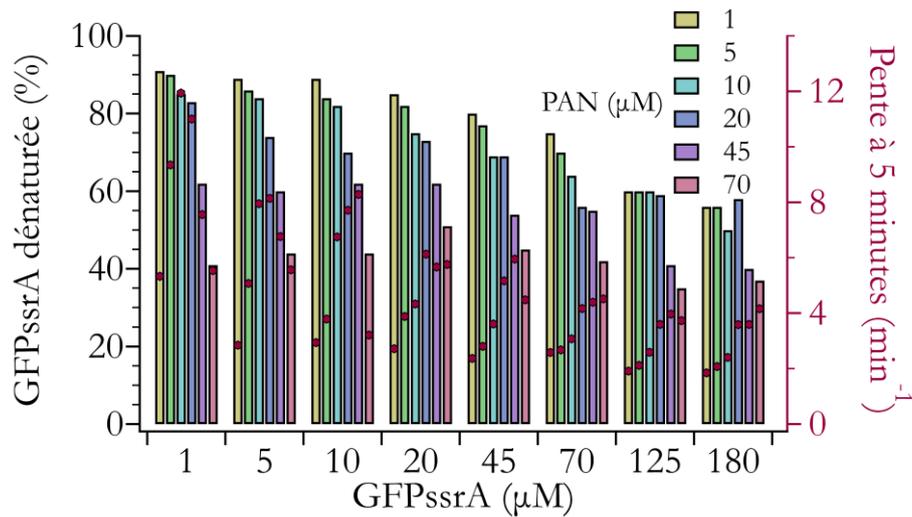
Dans un premier temps, différentes concentrations des composants suivants ont été combinées :

- **GFPssrA** : 1, 5, 10, 20, 45, 70, 125, 180 μM
- **PAN** : 1, 5, 10, 20, 45, 70 μM
- **ATP & MgCl<sub>2</sub>** : 10 mM ATP & 20 mM MgCl<sub>2</sub>, 50 mM ATP & 100 mM MgCl<sub>2</sub>, 100 mM ATP & 200 mM MgCl<sub>2</sub>. Les concentrations de MgCl<sub>2</sub> ont été fixées au double des concentrations d'ATP. Lorsque l'on mentionnera dans la suite de cette partie une concentration d'ATP, on considère que le MgCl<sub>2</sub> était également présent en solution.

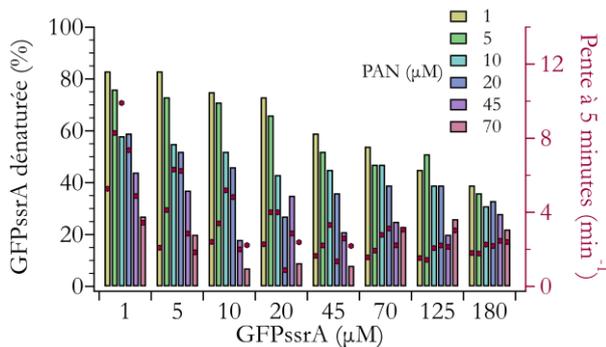
La figure 82 présente les données sous forme d'histogrammes correspondant au pourcentage de GFPssrA dénaturée après 1 heure à 55°C en fonction des différentes conditions testées, tandis que les données représentées sous forme de points correspondent à la pente à 5 minutes de réaction en fonction des différentes conditions testées. Étant donné le nombre de conditions à tester (192

conditions), chaque condition n'a été mesurée qu'une fois et on interprètera donc les résultats de façon qualitative.

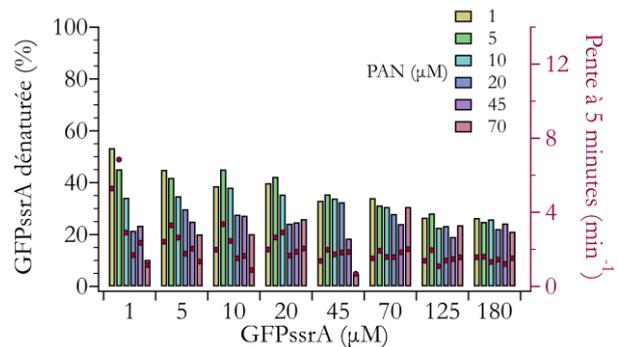
A



C



D



**Figure 82 : Criblage des conditions des stœchiométries de GFPssrA et PAN à différentes concentrations d'ATP** par mesure de l'activité du système à 55°C. **A.** Conditions à 100 mM ATP. **B.** Conditions à 50 mM ATP. **C.** Conditions à 10 mM ATP. Les données sont représentées sous forme d'histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) en fonction de la concentration de GFPssrA ( $\mu\text{M}$ ). Les barres de différentes couleurs représentent la concentration de PAN ( $\mu\text{M}$ ). La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les figures A, B & C présentent les mêmes réactions à différentes concentrations d'ATP. On observe de façon générale que plus la concentration d'ATP est élevée, plus le pourcentage de GFPssrA dénaturée et la pente à 5 minutes augmentent. Cela signifie donc que l'activité de dépliage de la GFPssrA par PAN est dépendante de la concentration d'ATP et que plus la concentration initiale d'ATP augmente, plus PAN a une activité de dépliage efficace.

À des concentrations de PAN et d'ATP constantes (courbes de couleur identique au sein de chaque figure A, B & C), on peut voir que plus la concentration de GFPssrA est élevée, plus le pourcentage de GFPssrA dénaturée et la pente à 5 minutes diminuent. Le système est donc moins efficace lorsque la concentration de GFPssrA augmente à des concentrations de PAN et d'ATP fixes. Cet effet est moins visible à 10 mM ATP (figure 82.C) puisque PAN est peu actif.

À des concentrations de GFPssrA et d'ATP fixes (séries de courbes de couleurs différentes au sein de chaque figure A, B & C), on observe que plus la concentration de PAN est élevée, plus le

pourcentage de GFPssrA dénaturée en fin de réaction diminue. Cependant, la pente à 5 minutes augmente, puis diminue lorsque la concentration de PAN augmente à des concentrations de GFPssrA et d'ATP fixes. À 100 mM ATP, la pente augmente aussi avec la concentration de PAN aux concentrations de GFPssrA élevées (70 à 180  $\mu$ M GFPssrA). Ces effets sont moins visibles à 10 mM ATP (figure 82.C) puisque PAN est peu actif. Il apparaît alors qu'à des concentrations faibles de PAN, la dénaturation de la GFPssrA est progressive au cours du temps, avec une pente plus faible mais un taux de dénaturation élevé en fin de réaction. Lorsque la concentration de PAN augmente, en particulier à 45 et 70  $\mu$ M, la dénaturation de la GFPssrA est rapide en début de réaction, avec un taux de dénaturation plus faible en fin de réaction.

Trois conclusions peuvent être tirées de ces mesures :

- **L'efficacité de PAN est corrélée à la concentration initiale d'ATP.** Cela est en accord avec de récentes observations<sup>393</sup> et est logique si l'on considère que l'activité de dépliement des substrats protéique par PAN est liée à son activité d'hydrolyse de l'ATP<sup>250,286,373</sup> (voir partie 5.2.3 du chapitre 1).
- **Le système est moins efficace lorsque la concentration de GFPssrA augmente.** Une hypothèse expliquant ce phénomène serait que tout l'ATP est consommé rapidement, ne permettant plus à PAN de reconnaître et déplier efficacement tous les substrats protéiques présents en solution. Cette seconde hypothèse est cohérente avec le fait que PAN est moins actif à faible concentration d'ATP. Il est également possible que la quantité de GFPssrA soit trop élevée pour pouvoir être dépliée en totalité par PAN.
- **Plus la concentration de PAN augmente, plus le dépliement de la GFPssrA est réalisé rapidement, avec un taux de dépliement de la GFPssrA plus faible en fin de réaction.** On peut supposer que la forte concentration de PAN dans la réaction entraîne une consommation d'ATP trop rapide pour permettre une activité de PAN suffisamment efficace pendant le reste de la réaction. Cette hypothèse est également cohérente avec le fait que PAN est moins actif à faible concentration d'ATP. L'apparition d'un plateau après une forte dénaturation de la GFPssrA en début de réaction, empêchant PAN de déplier les substrats restants, pourrait être dû à une concentration d'ATP trop faible, mais pourrait aussi résulter entre autres d'une accumulation d'ADP agissant comme inhibiteur.

L'hypothèse privilégiée est donc que l'activité de PAN est fortement dépendante de la concentration initiale d'ATP. On pourrait ajouter de l'ATP après une heure de réaction avec 180  $\mu$ M GFPssrA pour confirmer cette hypothèse. Cependant, l'ajout d'ATP à température ambiante (puisque'il faut ressortir la plaque de mesure du lecteur de plaque à 55°C pour y ajouter un élément) permet à une partie des GFPssrA dont le chromophore était « éteint » de retrouver leur fluorescence (figure 80), rendant difficile l'interprétation des résultats. On pourrait également utiliser un système de régénération de l'ATP. Cependant, cela aurait le désavantage de complexifier davantage le système pour les mesures SANS, d'autant plus que cette méthode utilise des enzymes mésophiles susceptibles d'être instable à 55°C. **On utilisera donc 100 mM ATP dans la suite de cette étude**, qui est la concentration maximum pouvant être utilisée (à cause de la limite de concentration du stock d'ATP et du volume disponible pour la réaction).

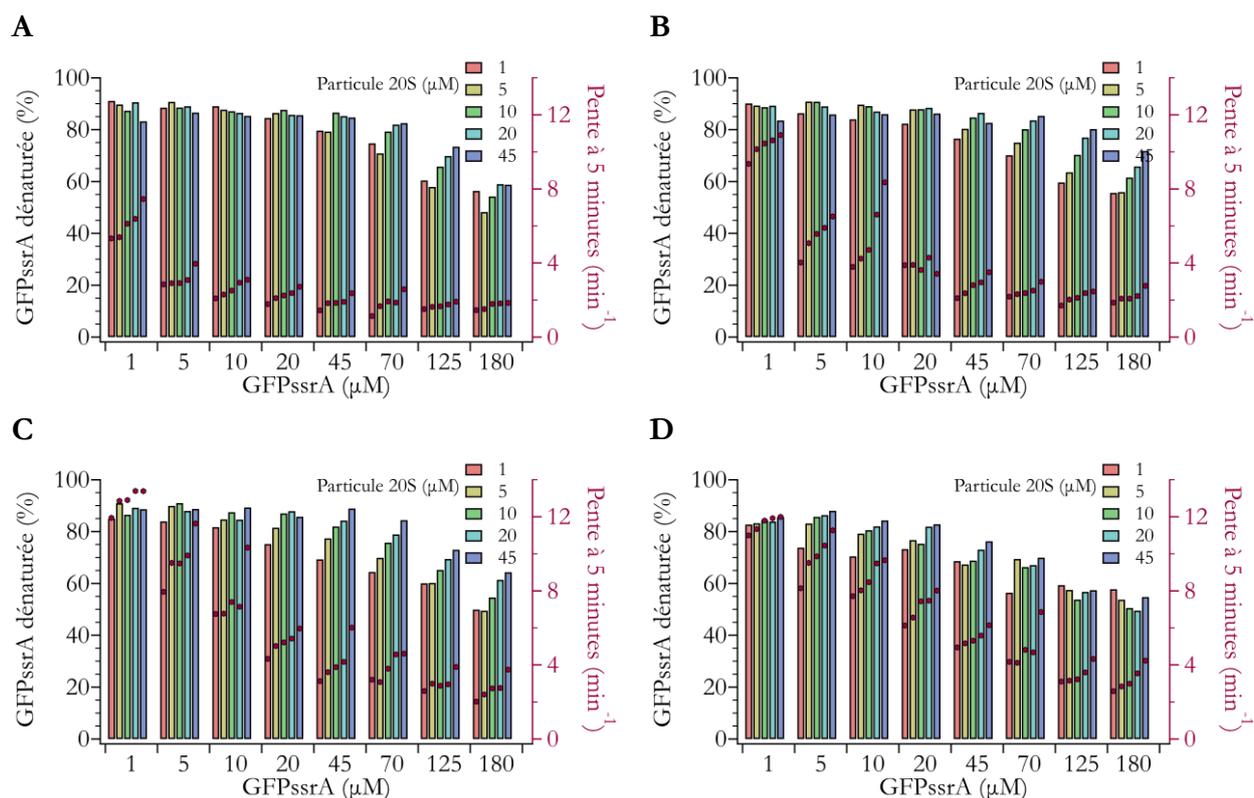
On utilisera pour la suite du criblage les concentrations de 1, 5, 10 et 20  $\mu$ M PAN puisque les concentrations de 45 et 70  $\mu$ M PAN sont trop élevées.

## 2.2. Criblage des stœchiométries GFPssrA, PAN et 20S à 100 mM ATP

Dans un second temps, différentes concentrations des éléments suivants ont été combinées à 100 mM ATP afin de déterminer les stœchiométries optimales compatibles avec la durée des expériences SANS et les contraintes de sensibilité de la méthode :

- **GFPssrA** : 1, 5, 10, 20, 45, 70, 125, 180  $\mu\text{M}$
- **PAN** : 1, 5, 10, 20  $\mu\text{M}$
- **Particule 20S** : 1, 5, 10, 20  $\mu\text{M}$

La figure 83 présente les données de la même manière que dans la partie précédente.



**Figure 83** : Criblage des conditions de stœchiométries de GFPssrA et particule 20S à différentes concentrations de PAN par mesure de l'activité du système à 55°C en présence de 100 mM ATP. **A.** Conditions à 1  $\mu\text{M}$  PAN. **B.** Conditions à 5  $\mu\text{M}$  PAN. **C.** Conditions à 10  $\mu\text{M}$  PAN. **D.** Conditions à 20  $\mu\text{M}$  PAN. Les données sont représentées sous forme d'histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) en fonction de la concentration de GFPssrA ( $\mu\text{M}$ ). Les barres de différentes couleurs représentent la concentration de particule 20S ( $\mu\text{M}$ ). La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les figures A, B, C & D présentent les mêmes concentrations de GFPssrA et particules 20S à différentes concentrations de PAN. On peut constater que plus la concentration de PAN augmente, plus le pourcentage final de GFPssrA dénaturée diminue alors que la pente à 5 minutes augmente. Cela concorde avec les résultats précédemment décrits (partie 2.1). Il apparaît donc que l'ajout de la particule catalytique 20S ne modifie pas cette tendance.

À des concentrations de PAN et particule 20S fixes (courbes de couleur identique au sein de chaque figure A, B, C & D), le pourcentage de GFPssrA dénaturée et la pente à 5 minutes diminuent lorsque la concentration de GFPssrA augmente. PAN est donc moins efficace lorsque la

concentration de GFPssrA augmente à des concentrations en ATP constantes, tout comme décrit dans la partie 2.1. L'ajout de la particule catalytique 20S ne modifie donc pas non plus cette tendance.

À des conditions de GFPssrA et de PAN constantes (séries de courbes de couleurs différentes au sein de chaque figure A, B, C & D), on constate que plus la concentration de la particule 20S augmente, plus le pourcentage de GFPssrA dénaturée et la pente à 5 minutes augmentent. Le système apparaît donc plus efficace lorsque la concentration de particules 20S augmente dans la solution.

On peut donc conclure de ces mesures que dans les conditions testées :

- **L'ajout de particule 20S n'a pas d'influence sur les phénomènes décrits en partie 2.1.**
- **L'efficacité de dépliement de la GFPssrA par PAN augmente avec la concentration de particule 20S.** Cette information complète la partie 1.2 qui décrivait que la présence de 20S augmentait l'efficacité de dépliement de la GFPssrA par PAN.

### 2.3. Conclusions du criblage

Pour rappel, les conditions de mesures SANS en temps réel nécessitent d'avoir une réaction efficace, mais suffisamment lente pour suivre la réaction avec un point de mesure toutes les 30 secondes. De plus, une concentration relativement élevée de GFPssrA est nécessaire pour avoir un signal analysable.

Les mesures réalisées ont permis de sélectionner un tampon contenant 100 mM ATP et 200 mM MgCl<sub>2</sub> ainsi que deux conditions de stœchiométrie :

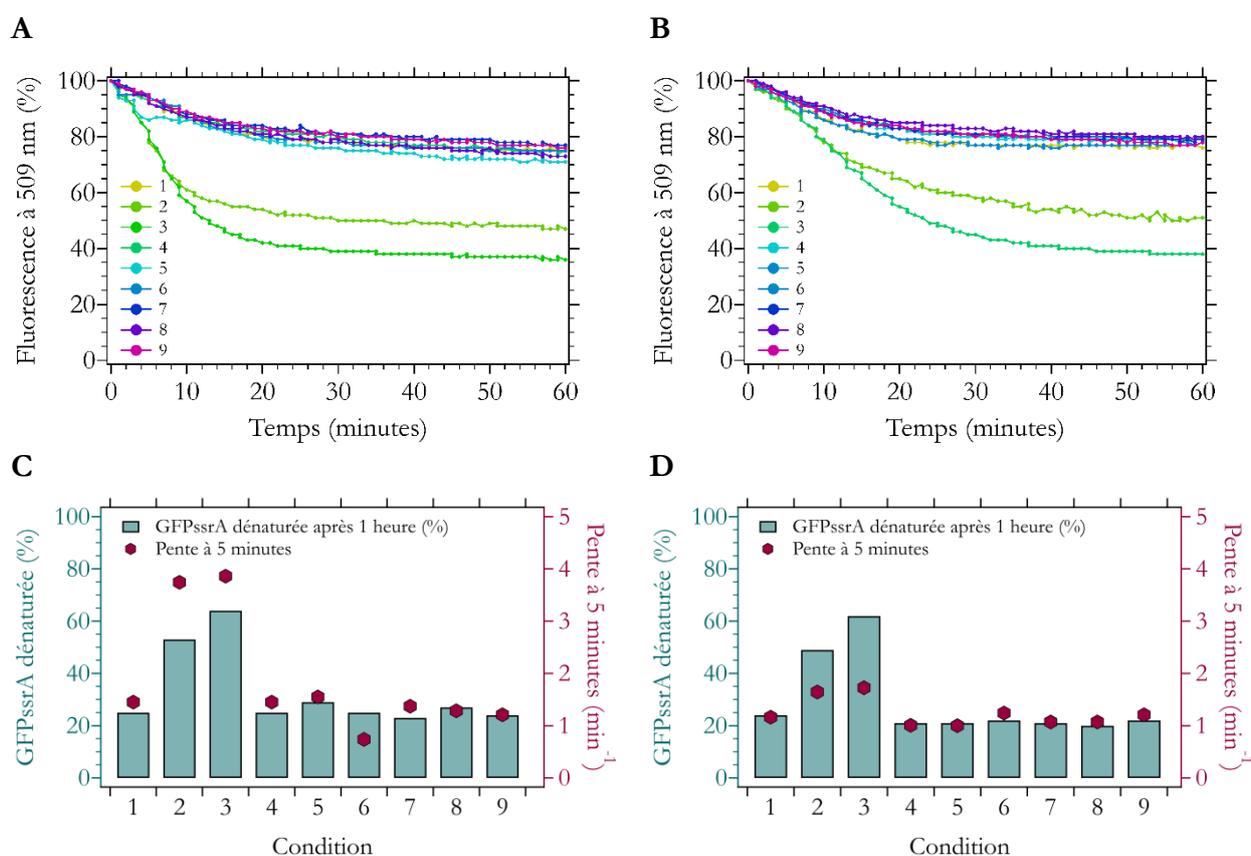
- **Condition SANS 1 – 70 μM GFPssrA, 20 μM PAN et 20 μM particule 20S** : cela correspond à la concentration minimum de GFPssrA (~ 2 mg/mL) permettant de visualiser un signal exploitable en SANS avec 30 secondes de temps d'exposition aux neutrons. La concentration de PAN choisie permet de suivre l'activité du système en ayant un point de mesure SANS toutes les 30 secondes. La concentration de particule 20S choisie correspond à la concentration maximum pouvant être utilisée compte-tenu des limitations des concentrations des stocks des protéines et du volume disponible. On pourra visualiser une différence significative d'efficacité de dénaturation en présence et en absence de particule 20S. Cette condition a également l'avantage d'être cohérente avec les précédents travaux effectués au laboratoire<sup>40</sup>.
- **Condition SANS 2 – 180 μM GFPssrA, 5 μM PAN et 30 μM particule 20S** : cette condition permet d'avoir un signal plus fort de la GFPssrA lors des mesures SANS, tout en observant un effet de PAN sans perte d'activité trop importante au cours du temps et un ajout de particule 20S ayant un effet drastique sur la dénaturation de la GFPssrA. La concentration de 30 μM de particule 20S n'a pas pu être testée lors de ce criblage, ne pouvant pas être réalisée dans toutes les conditions (limitation des concentrations des stocks des protéines et volume disponible). L'efficacité étant corrélée à l'augmentation de la concentration de particule 20S, cela permettra de renforcer l'effet de la particule 20S sur l'efficacité du système.

L'activité du protéasome PAN-20S a été amplement étudiée depuis la découverte de PAN en 1999<sup>261,277,392</sup>. Cependant, les conditions choisies pour les mesures SANS sont très éloignées de celles précédemment décrites dans la littérature, puisque ces mesures nécessitent des

concentrations de protéines beaucoup plus élevées. Différents tests ont donc été réalisés dans les deux conditions SANS choisies afin de s'assurer que la machinerie protéolytique fonctionne de la façon attendue.

### 3. Effets de l'ATP, ADP et ATP $\gamma$ S sur l'activité du protéasome PAN-20S mesurés par spectroscopie de fluorescence

Une étude fonctionnelle a été réalisée afin d'apprécier l'effet de différents nucléotides sur le système protéolytique. Pour cela, 100 mM d'ATP, ADP ou ATP $\gamma$ S (ATP non hydrolysable) ont été utilisés en présence de 200 mM MgCl<sub>2</sub> dans les deux conditions SANS sélectionnées dans la partie 2.3. Les données sont représentées dans les figures 84 de la même manière que dans la partie 1.2. Chaque condition n'ayant été mesurée qu'une fois, on interprétera les résultats de façon qualitative.



**Figure 84** : Effet de l'ATP, ADP et ATP $\gamma$ S sur l'activité du système à 55°C dans les deux conditions SANS en présence 200 mM MgCl<sub>2</sub>. Les figures A. et C. représentent les résultats pour la condition SANS 1 et les figures B. et D. représentent les résultats pour la condition SANS 2. Pour toutes les figures, les conditions testées sont numérotées de 1 à 9. Ceci correspond aux réactions suivantes : (1) GFPssrA, (2) GFPssrA et PAN, ou (3) GFPssrA, PAN et particule 20S, toutes trois en présence d'ATP, (4) GFPssrA, (5) GFPssrA et PAN, ou (6) GFPssrA, PAN et particule 20S, toutes trois en présence d'ADP, (7) GFPssrA, (8) GFPssrA et PAN, ou (9) GFPssrA, PAN et particule 20S toutes trois en présence d'ATP $\gamma$ S. **A, B.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. **C, D.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

L'analyse est semblable pour les deux conditions SANS. Dans les conditions où la GFPssrA est isolée en présence de 100 mM ATP (1) et dans toutes les conditions à 100 mM ADP et 100 mM ATP $\gamma$ S (4 à 9), le pourcentage final de GFPssrA dénaturée est de 20 à 30% tandis que la pente à 5

minutes est de 1 à 1,5 min<sup>-1</sup>. Ces valeurs correspondent à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C.

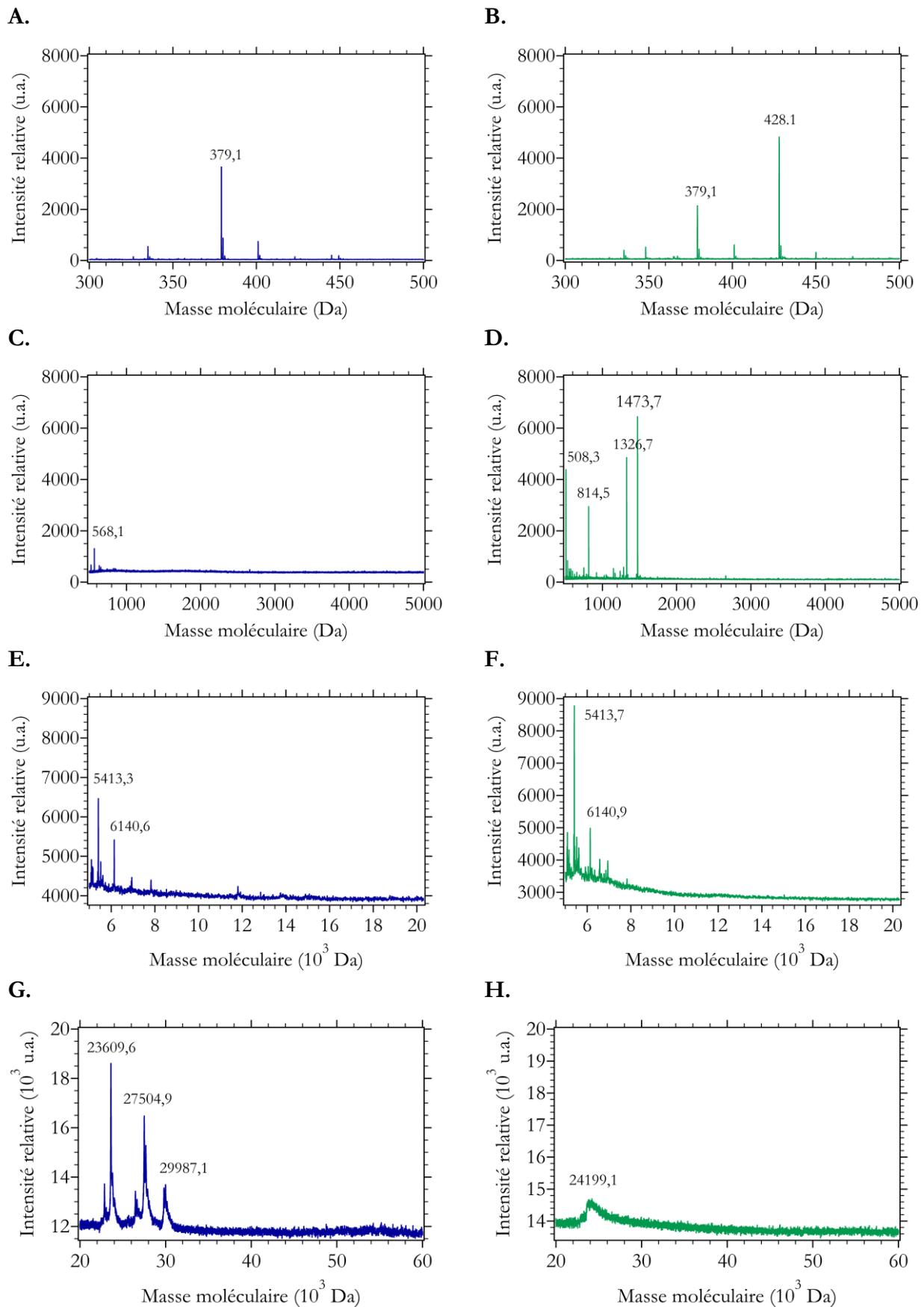
La GFPssrA avec PAN à 100 mM ATP (2) présente un pourcentage de GFPssrA dénaturée en fin de réaction est d'environ 50% pour les deux conditions SANS tandis que la pente à 5 minutes est d'environ 3,7 min<sup>-1</sup> dans la condition SANS 1 et de 1,6 min<sup>-1</sup> dans la condition SANS 2. En présence de la machinerie protéolytique complète à 100 mM ATP (3), le pourcentage de GFPssrA dénaturée en fin de réaction est d'environ 60% pour les deux conditions SANS tandis que la pente à 5 minutes est d'environ 3,8 min<sup>-1</sup> dans la condition SANS 1 et de 1,7 min<sup>-1</sup> dans la condition SANS 2. Ces valeurs indiquent une activité de dépliement de la GFPssrA par PAN qui est stimulée par l'ajout de la particule 20S, comme il a été décrit dans la partie 2.1.

Il n'y a donc pas d'activité visible de dénaturation de la GFPssrA par PAN et la particule 20S en présence d'ADP ou d'ATP non hydrolysable (ATP $\gamma$ S) tandis que c'est le cas en présence d'ATP. On peut conclure de ces mesures que **l'hydrolyse de l'ATP est nécessaire à l'activité du système protéolytique dans les conditions d'étude**, comme le suggère la littérature<sup>261,392</sup>.

#### 4. Détection des produits de dégradation générés par le système par spectrométrie de masse

La taille moyenne des produits de dégradation de la GFPssrA par le système protéolytique après une heure de réaction à 55°C a été estimée expérimentalement par spectrométrie de masse MALDI-TOF. Cette mesure n'a été réalisée que pour la condition SANS 2, qui a pour avantage de contenir un excès de molécules de GFPssrA comparé aux particules PAN et 20S (voir partie 2.3). Ceci permet une analyse plus simple des spectres de masse.

Deux échantillons contenant la GFPssrA, PAN et la particule 20S ont été préparés, l'un en présence de 100 mM ATP afin de visualiser les produits de dégradation de la GFPssrA par la machinerie protéolytique et l'autre ne contenant pas d'ATP afin que la GFPssrA reste intacte. Après une heure d'incubation à 55°C, les échantillons ont été mesurés par MALDI-TOF. Les données issues de l'analyse sont présentées dans la figure 85 sous forme de spectre de masse avec l'abondance relative à différentes gammes de masse moléculaire.



**Figure 85 : Spectres de masse de la GFPssrA, PAN et la particule 20S après une heure d'incubation à 55°C.** Les figures à droite (A, C, E, G) représentent la condition sans ATP et les figures à gauche (B, D, F, H) la condition

en présence de 100 mM ATP. Les données sont représentées sous forme de l'abondance relative en fonction de la masse moléculaire. **A & B.** Spectres de masses de 300 Da à 500 Da. **C & D.** Spectres de masses de 500 Da à 5 kDa. **E & F.** Spectres de masses de 5 kDa à 20 kDa. **G & H.** Spectres de masses de 20 kDa à 60 kDa.

Les pics principaux de chaque spectre de masse ont été comparés à la séquence de la GFPssrA. On peut noter que les valeurs retrouvées à la fois dans l'échantillon sans ATP (contrôle négatif) et avec 100 mM ATP n'ont pas été considérées. Les résultats peuvent être résumés de la façon suivante :

#### **GFPssrA, PAN et particule 20S sans ATP (figure 85.A)**

- Gamme de masse moléculaire très faible, faible et moyenne (300-500 Da, 500-5000 Da et 5000-20000 Da) :
  - Aucune espèce détectée
- Gamme de masse moléculaire haute (20000-60000 Da)
  - 23609,6 Da : sous-unité  $\beta\Delta 6$  de la particule 20S
  - 27504,9 Da : GFPssrA intacte
  - 29987,1 Da : sous-unité  $\alpha$  de la particule 20S

#### **GFPssrA, PAN et particule 20S avec 100 mM ATP (figure 85.B)**

- Gamme de masse moléculaire très faible (300-500 Da)
  - 428,1 Da : pas de correspondance avec la séquence de la GFPssrA
- Gamme de masse moléculaire faible (500-5000 Da)
  - 508,3 Da : ne correspond à aucune séquence de la GFPssrA
  - 814,5 Da : séquence de 6 acides aminés de la GFPssrA (NFKIRH et FKIRHN)
  - 1326,7 Da : séquence de 13, 10 et 12 acides aminés de la GFPssrA (SRYPDHMKRH, TGVVPILVELDGD et HYQQNTPIGDGP)
  - 1473,7 Da : séquence de 14 et 11 acides aminés de la GFPssrA (FTGVVPILVELDGD et FSRYPDHMKRH)
- Gamme de masse moléculaire moyenne et haute (5000-20000 Da et 20000-60000 Da)
  - Aucune espèce détectée

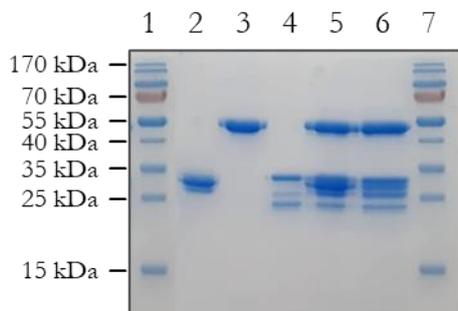
Des peptides correspondant à des séquences de la GFPssrA ont été observés uniquement dans l'échantillon de GFPssrA, PAN et la particule 20S contenant 100 mM ATP comme illustré figure 86, tandis que la GFPssrA intacte n'est détectée que dans l'échantillon de GFPssrA, PAN et la particule 20S sans ATP. Les peptides détectés correspondent en moyenne à des séquences d'une dizaine d'acides aminés.

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTL 60  
————— 1472,78 Da  
————— 1325,71 Da  
 VTTFSYGVCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLV 120  
————— 1472,78 Da  
————— 1325,64 Da  
 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLAD 240  
————— 813,46 Da  
————— 813,46 Da  
 HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAAGITHGMDELAAND  
————— 1325,60 Da  
 ENYALAA 247

**Figure 86 : Séquence en acides aminés de la GFPssrA.** Les peptides détectés en spectrométrie de masse sont soulignés en couleur avec la masse moléculaire correspondante.

L'échantillon de GFPssrA, PAN et la particule 20S sans ATP contient également la particule 20S (sous forme de monomères de ces deux sous-unités). On note que la protéine PAN n'est pas détectée car trop diluée dans la réaction pour être visualisée. De façon étonnante, l'échantillon de GFPssrA, PAN et la particule 20S contenant 100 mM ATP ne semble quant à lui pas contenir la particule 20S.

Pour vérifier que la particule 20S est toujours présente après la dégradation de la GFPssrA, un gel SDS-PAGE a été réalisé. Pour cela, les protéines GFPssrA, PAN et la particule 20S ont été mélangées à 1 mg/mL soit en présence de 100 mM ATP, soit sans ATP puis incubées 1 heure à 55°C avant d'être déposées sur gel SDS-PAGE à 15% d'acrylamide (figure 87).



**Figure 87 : SDS-PAGE avec de la GFPssrA, PAN et la particule 20S après une heure d'incubation à 55°C.** SDS-PAGE, 15 % acrylamide, coloration à l'InstantBlue. Les puits 1 et 7 correspondent aux marqueurs de masse moléculaire. Les échantillons sont numérotés de 2 à 6 et correspondent aux conditions : GFPssrA (2), PAN (3), particule 20S (4), GFPssrA, PAN et particule 20S en présence de 100 mM ATP (5) ou sans ATP (6). Les protéines GFPssrA, PAN et la particule 20S ont été mélangées à 1 mg/mL puis incubées 1 heure à 55°C avant d'être déposées sur gel.

Après 1 heure à 55°C, l'échantillon de GFPssrA, PAN et particule 20S sans ATP contient les 3 protéines dans des quantités similaires aux contrôles des protéines seules. L'échantillon de GFPssrA, PAN et particule 20S contenant 100 mM ATP présente des quantités de PAN et de particules 20S similaires aux contrôles avec les protéines seules et à l'échantillon sans ATP mais avec une forte diminution de la bande correspondant à la GFPssrA. La GFPssrA intacte disparaît donc au profit de forme(s) de plus faible(s) masse(s) moléculaire(s) non visible(s) sur gel SDS-PAGE à 15% d'acrylamide. De plus, on peut conclure que la particule 20S ne disparaît pas au cours

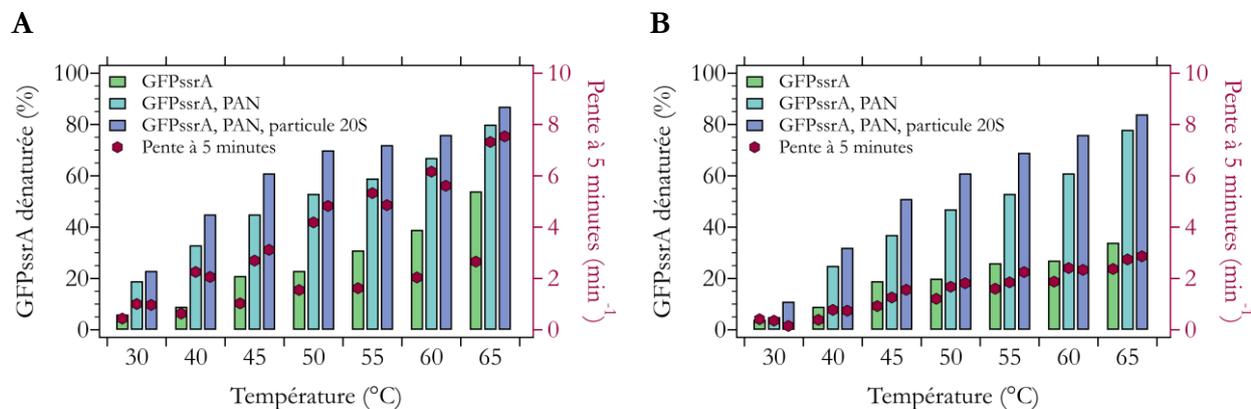
de la réaction. L'absence de particule 20S dans l'échantillon à 100 mM ATP en spectrométrie de masse était probablement un artefact expérimental.

En conclusion, la spectrométrie de masse MALDI-TOF démontre que dans les conditions testées :

- **Le système protéolytique dégrade la GFPssrA en peptides d'une dizaine d'acides aminés en moyenne après 1 heure à 55°C en présence de 100 mM ATP.** Ces valeurs sont cohérentes avec celles de la littérature qui décrit que les produits de dégradation de substrats de la machinerie protéolytique de *M. jannaschii* font de 3 à 30 résidus<sup>277</sup>.
- **Les deux formes de GFPssrA sont dégradées par le système protéolytique.** En effet, aucune GFPssrA intacte n'a été détectée dans le spectre de masse à 100 mM ATP. Cela est cohérents avec les résultats du Western-blot présenté partie 1.3. Il se peut également que la quantité de GFPssrA restée intacte soit trop faible pour être détectée.

## 5. Effets des variations de température sur l'activité du protéasome PAN-20S mesurés par spectroscopie de fluorescence

Des mesures d'activité ont été réalisées afin de visualiser l'effet de la température sur l'activité du système protéolytique. Pour cela, différentes températures ont été testées en présence de 100 mM ATP dans les deux conditions SANS sélectionnées dans la partie 2.3. La figure 88 présente de la même manière que dans les parties précédentes l'analyse de l'activité du système à différentes températures. Les différentes températures testées sont : 30, 40, 45, 50, 55, 60 et 65°C (limite de l'instrument). Chaque condition n'ayant été mesurée qu'une fois, on interprètera les résultats de façon qualitative.



**Figure 88 : Mesure de l'activité du système à différentes températures** dans les deux conditions SANS en présence 100 mM ATP. **A.** Résultats pour la condition SANS 1. **B.** Résultats pour la condition SANS 2. Les données sont représentées sous forme d'histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) aux différentes températures. Les barres de différentes couleurs représentent les conditions (GFPssrA, en présence de PAN ou en présence de PAN et de la particule 20S). La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Des résultats similaires ont été obtenus pour les deux conditions SANS. Lorsque la GFPssrA est isolée, le pourcentage de GFPssrA dénaturée après 1 heure augmente progressivement avec la température, de 5% à 30°C jusqu'à 55% à 65°C pour la condition SANS 1 et 35% pour la condition SANS 2. La pente à 5 minutes augmente également progressivement avec la température, de 0,5

$\text{min}^{-1}$  à  $30^\circ\text{C}$  jusqu'à  $2,5 \text{ min}^{-1}$  à  $65^\circ\text{C}$  pour les deux conditions SANS. Ces valeurs correspondent à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA.

Quand PAN est présent, le pourcentage de GFPssrA dénaturée en fin de réaction augmente progressivement avec la température, de 20% à  $30^\circ\text{C}$  jusqu'à 80% à  $65^\circ\text{C}$  pour la condition SANS 1 et de 5% à  $30^\circ\text{C}$  jusqu'à 80% pour la condition SANS 2. La pente à 5 minutes augmente également progressivement, de  $1 \text{ min}^{-1}$  à  $30^\circ\text{C}$  jusqu'à  $7,5 \text{ min}^{-1}$  à  $65^\circ\text{C}$  pour la condition SANS 1 et de  $0,5 \text{ min}^{-1}$  à  $30^\circ\text{C}$  jusqu'à  $3 \text{ min}^{-1}$  à  $65^\circ\text{C}$  pour la condition SANS 2. Avec PAN et la particule 20S, le pourcentage final de GFPssrA dénaturée augmente progressivement avec la température, de 25% à  $30^\circ\text{C}$  jusqu'à 85% à  $65^\circ\text{C}$  pour la condition SANS 1 et de 10% à  $30^\circ\text{C}$  jusqu'à 85% pour la condition SANS 2. La pente à 5 minutes augmente également progressivement, de  $1 \text{ min}^{-1}$  à  $30^\circ\text{C}$  jusqu'à  $7,5 \text{ min}^{-1}$  à  $65^\circ\text{C}$  pour la condition SANS 1 et de  $0,2 \text{ min}^{-1}$  à  $30^\circ\text{C}$  jusqu'à  $3 \text{ min}^{-1}$  à  $65^\circ\text{C}$  pour la condition SANS 2. L'activité de dépliement de la GFPssrA par PAN est stimulée par l'ajout de la particule 20S à toutes les températures testées.

La température a donc un effet important sur l'activité du système. **Le protéasome PAN-20S est plus efficace lorsque la température augmente, mais la diminution de l'efficacité quantique de la fluorescence de la GFPssrA est également plus importante.** La température de  $55^\circ\text{C}$  est appropriée pour notre étude puisqu'elle permet d'observer une activité de la machinerie protéolytique tout en limitant l'effet de diminution de l'efficacité quantique de la fluorescence de la GFPssrA<sup>374</sup> et la dénaturation thermique de la protéine, qui n'est pas stable au-delà de  $65^\circ\text{C}$ <sup>383</sup> (voir partie 4.1.4 du chapitre 5).

## 6. Étude de la consommation de l'ATP par PAN

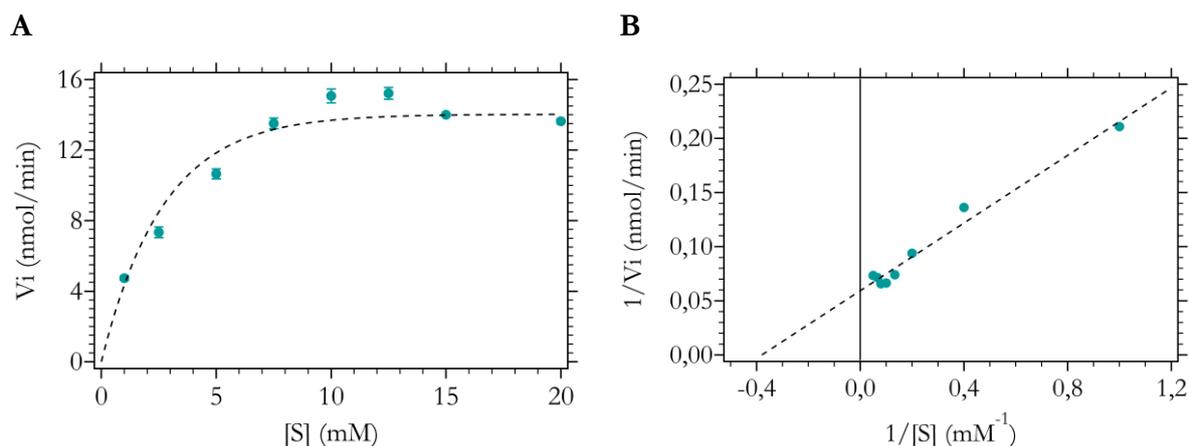
### 6.1. Estimation de l'affinité de PAN et de la GFPssrA pour l'ATP

PAN est une enzyme catalysant l'hydrolyse de l'ATP en ADP et phosphate inorganique. Nous avons souhaité estimer expérimentalement l'affinité de PAN pour l'ATP. Pour cela, la vitesse de réaction d'hydrolyse de l'ATP a été mesurée à différentes concentrations d'ATP (de 1 à 20 mM). Un second but de l'expérience était de déterminer si l'ajout de GFPssrA a une influence sur l'affinité de PAN pour l'ATP. Les conditions mesurées en triplicata étaient alors :

- 1  $\mu\text{M}$  PAN
- 1  $\mu\text{M}$  PAN + 1  $\mu\text{M}$  GFPssrA
- 1  $\mu\text{M}$  GFPssrA

La vitesse de réaction de l'hydrolyse de l'ATP a été déterminée en mesurant la quantité de phosphate inorganique libre en solution après 2 minutes de réaction à  $55^\circ\text{C}$ . Pour cela, les mélanges réactionnels ont été préparés dans la glace, puis incubés à  $55^\circ\text{C}$  avant d'être replacés dans la glace afin d'arrêter la réaction. Après addition de Malachite Green Phosphate Assay Kit, le dosage a ensuite été réalisé par mesure de l'absorbance à 630 nm traduisant la formation d'un complexe orthophosphate-molybdate-vert malachite. La quantité de phosphate libéré lors de l'hydrolyse de l'ATP a été estimée par comparaison avec une courbe standard de phosphate.

A titre d'exemple, la figure 89 montre les représentations de Michaelis-Menten (A) et Lineweaver & Burk (B) pour la réaction avec 1  $\mu\text{M}$  PAN.



**Figure 89 : Mesure de la vitesse initiale de l'hydrolyse de l'ATP par PAN (1  $\mu\text{M}$ ) en fonction de la concentration d'ATP** par mesure de la concentration de phosphate organique libre en solution après 2 minutes à 55°C. **A.** Représentation de Michaelis-Menten. La courbe en pointillé représente un fit d'une courbe hyperbole rectangulaire aux points de mesure. **B.** Représentation de Lineweaver & Burk. La droite en pointillé représente un fit d'une droite aux points de mesure.

On constate que la courbe A présente une allure hyperbolique tandis que la représentation en double inverse B a une allure de droite, suggérant que PAN a bien un comportement d'enzyme michaelienne pour l'hydrolyse de l'ATP. Les valeurs de  $v_{\max}$  et  $K_M$  ont été déterminées pour chacune des conditions à partir de la représentation de Lineweaver & Burk. Ces valeurs sont compilées dans le tableau 18.

**Tableau 18 : Valeurs de  $v_{\max}$  et  $K_M$  pour les différentes conditions** testées par mesure de la quantité de phosphate inorganique libre après 2 minutes à 55°C. Ces valeurs ont été déterminées grâce à la représentation en double inverse de Lineweaver & Burk. Le nombre de points ayant servi à la génération d'une droite ainsi que le coefficient de détermination  $R^2$  sont également indiqués.

	Nombre de points	$R^2$	$v_{\max}$ (nmol/min)	$K_M$ (mM)
1 $\mu\text{M}$ PAN	8	0,98	16,9	2,6
1 $\mu\text{M}$ PAN + 1 $\mu\text{M}$ GFPssrA	8	0,98	14,2	5,5
1 $\mu\text{M}$ GFPssrA	7	0,99	4,6	22,2

On observe dans un premier temps que le  $K_M$  apparent de PAN pour l'ATP est de 2,6 mM. Cette valeur est du même ordre de grandeur que celle d'environ 0,5 mM (à 80°C) décrite précédemment dans la littérature pour PAN et d'autres ATPase AAA+<sup>277</sup>. À 100 mM ATP lors des mesures *in vitro* sur le système modèle, la concentration d'ATP est 40 fois supérieure au  $K_M$ .

La GFPssrA semble avoir une influence négative sur l'affinité de PAN pour l'ATP car le  $K_M$  augmente (5,5 mM) avec l'ajout de 1  $\mu\text{M}$  GFPssrA. La vitesse d'hydrolyse maximale  $v_{\max}$  d'hydrolyse de l'ATP par PAN n'est par contre pas significativement influencée par l'ajout de GFPssrA. Cependant, la GFPssrA seule possède une activité ATPasique caractérisée par un  $K_M$  pour l'ATP de 22,2 mM. La GFPssrA a donc de façon étonnante une faible affinité pour l'ATP. Comme cette affinité pour l'ATP est environ 10 fois plus faible que celle de PAN et une  $v_{\max}$  3,5 fois moins importante, l'apparente hydrolyse de l'ATP par la GFPssrA ne devrait contribuer que

de façon mineure aux mesures d'activité d'hydrolyse de l'ATP par PAN lors des tests d'activité *in vitro*.

On peut donc tirer plusieurs conclusions de ces mesures :

- **À une concentration initiale de 100 mM ATP dans les conditions d'études SANS, soit 40 fois supérieure au  $K_M$ , l'activité d'hydrolyse de l'ATP par PAN est proche de la  $v_{max}$ .** L'enzyme fonctionne donc à capacité maximale (au moins en début de réaction) dans les conditions d'études. On pourra alors interpréter l'évolution des courbes de fluorescence sur le temps de mesure.
- **L'ajout de GFPssrA n'augmente pas l'affinité de PAN pour l'ATP.**
- **Étonnamment, la GFPssrA a une faible affinité pour l'ATP qui contribue de façon mineure aux mesures d'activité ATPase par PAN.**

On peut s'étonner de l'affinité relativement faible de PAN pour l'ATP, signifiant que la concentration d'ATP nécessaire à une activité importante de PAN est relativement élevée. Ceci est observé en général pour les enzymes dont le substrat est en concentration élevée dans leur environnement, ou qui ont un spectre de substrat large<sup>394</sup>. Le spectre de substrat de PAN est étroit (ATP, CTP et GTP) et on suppose que la concentration intracellulaire d'ATP à chaque instant est assez faible. À titre d'exemple, il a été déterminé que la valeur moyenne estimée de la concentration d'ATP chez *E. coli* était de 1,5 mM, avec des niveaux très différents de concentrations intracellulaires d'ATP selon les cellules<sup>395</sup>. Si cette valeur de concentration intracellulaire est similaire chez *M. jannaschii* et que les estimations d'affinité sont représentatives de la réalité physiologique, cela suggère que PAN ne fonctionne généralement pas en capacité maximum.

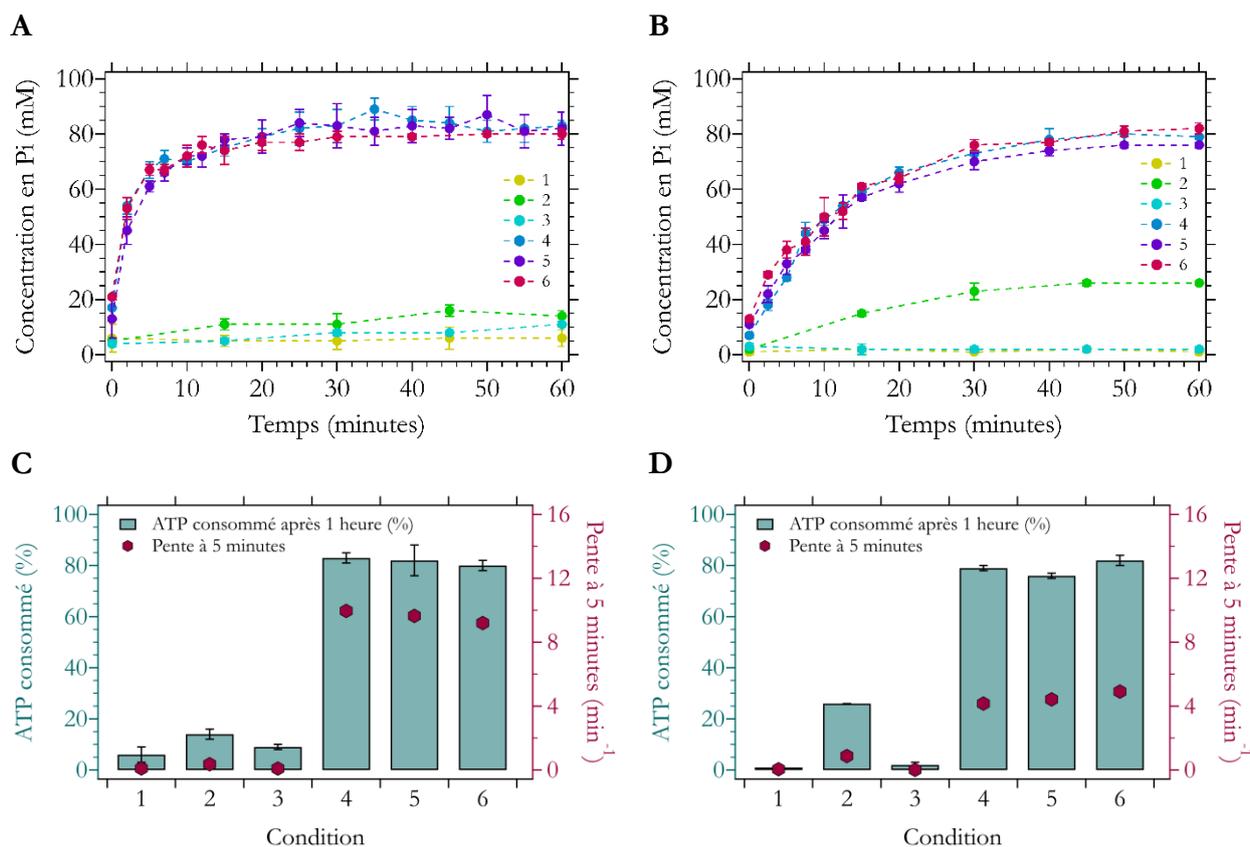
## 6.2. Mesure de la consommation d'ATP au cours du temps

Bien que des études biochimiques et structurales aient été réalisées sur la liaison et l'hydrolyse de l'ATP par PAN et sur le lien entre hydrolyse de l'ATP et dépliement des substrats protéiques<sup>250,286,396</sup> (voir partie 5.2.3 du chapitre 1), à notre connaissance aucune mesure de la consommation d'ATP au cours du temps par PAN n'a été publiée.

La consommation d'ATP par PAN au cours du temps a été suivie dans les deux conditions SANS d'étude (partie 2.3). La quantité de phosphate inorganique libre en solution au cours du temps a été mesurée pendant 1 heure de réaction à 55°C avec une concentration initiale de 100 mM ATP. Les conditions étudiées en triplicata sont :

- |                     |                                     |
|---------------------|-------------------------------------|
| ▪ (1) Tampon        | ▪ (4) PAN                           |
| ▪ (2) Particule 20S | ▪ (5) PAN et GFPssrA                |
| ▪ (3) GFPssrA       | ▪ (6) PAN, GFPssrA et particule 20S |

Les données représentées dans les figures 90.A, 90.B correspondent aux courbes d'évolution de la concentration de phosphate inorganique libre en fonction du temps dans les différentes conditions testées et permettent d'avoir une appréciation qualitative de la consommation de l'ATP par le système. Les figures 90.C et 90.D représentent le pourcentage d'ATP consommé après une heure de réaction à 55°C ainsi que la pente de la tangente à la courbe après 5 minutes de réaction à 55°C, équivalant à la vitesse de consommation de l'ATP en début de réaction. Ces deux paramètres permettent d'estimer l'efficacité du système.



**Figure 90 : Mesure de l'hydrolyse de l'ATP en fonction du temps à 55°C** dans les deux conditions SANS par mesure de la concentration de phosphate organique libre en solution avec des concentrations initiales de 100 mM ATP. Les figures A. et C. représentent les résultats pour la condition SANS 1 et les figures B. et D. représentent les résultats pour la condition SANS 2. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6 correspondant aux réactions suivantes : (1) tampon, (2) particule 20S, (3) GFPssrA, (4) PAN, (5) PAN et GFPssrA, et (6) PAN, GFPssrA et particule 20S. **A, B.** Courbes d'évolution la concentration de phosphate libre en solution en fonction du temps dans les différentes conditions testées. **C, D.** Histogrammes avec le pourcentage d'ATP consommé après 1 heure de réaction (par mesure de la concentration de phosphate libre en solution) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les résultats obtenus dans les conditions SANS 1 et 2 montrent les mêmes tendances. Dans les conditions où il n'y a que le tampon ou la particule 20S en solution (1 et 2), le pourcentage de phosphate inorganique libre après 1 heure à 55°C est inférieur à 10% tandis que la pente à 5 minutes est inférieure à  $0,1 \text{ min}^{-1}$ . Cela représente les conditions où il n'y a pas d'hydrolyse de l'ATP. Lorsque la GFPssrA est isolée en solution (3), le pourcentage de phosphate inorganique libre après 1 heure à 55°C est de 15% en condition SANS 1 et de 25% en condition SANS 2, tandis que la pente à 5 minutes est respectivement d'environ  $0,35$  et  $0,85 \text{ min}^{-1}$ . Il y a donc une faible activité d'hydrolyse de l'ATP en présence de GFPssrA, comme le suggèrent les résultats de la partie 6.1.

La protéine PAN seule (4), avec de la GFPssrA (5), et avec de la GFPssrA et de la particule 20S (6) présente un pourcentage de phosphate inorganique libre après 1 heure d'environ 80% dans les deux conditions SANS. La pente à 5 minutes est quant à elle entre  $9$  et  $10 \text{ min}^{-1}$  pour la condition SANS 1 et entre  $4$  et  $5 \text{ min}^{-1}$  pour la condition SANS 2. L'apparition de phosphate inorganique est plus progressive dans les conditions SANS 2 avec une pente moins importante que dans les conditions SANS 1. Toutefois, le pourcentage de phosphate inorganique libre après 1 heure est similaire dans les deux conditions SANS. Dans la condition SANS 1, l'ATP est consommée rapidement en début

de réaction pour atteindre un plateau après 15 minutes de réaction. Dans la condition SANS 2, il y a moins de PAN en solution. L'ATP est consommée moins rapidement que dans la condition SANS 1 en début de réaction, mais de façon continue au cours du temps jusqu'à atteindre un niveau final de consommation de l'ATP identique à la condition SANS 1 après 45 minutes.

On remarque que ce comportement de consommation d'ATP par PAN ressemble de façon frappante à celui de la dénaturation de la GFPssrA pour des conditions similaires (allure générale des courbes, pourcentage de consommation de l'ATP et de dénaturation de la GFPssrA après 1 heure, pente de la tangente de la courbe après 5 minutes). L'ATP semble donc l'élément limitant pour l'activité de dépliement de la GFPssrA, comme le suggèrent les observations de la partie 2.1., et il semblerait que le dépliement de la GFPssrA par PAN soit modulé étroitement par l'activité d'hydrolyse de l'ATP.

Trois conclusions principales peuvent être tirées :

- **La cinétique d'hydrolyse de l'ATP par PAN module la cinétique de dépliement de la GFPssrA.** Ce phénomène a été suggéré dans la littérature pour PAN<sup>250,286,393</sup> et a été décrit de façon approfondie pour la protéase AAA+ ClpXP<sup>289</sup> (voir partie 5.2.4 du chapitre 1). Les mesures SANS permettront de compléter les observations de cette expérience en suivant directement l'évolution structurale de la GFPssrA au cours du temps.
- **La GFPssrA a une faible activité d'hydrolyse de l'ATP.** Il est intéressant de noter que la GFP seule ou l'étiquette ssrA seule n'hydrolysent pas l'ATP dans des conditions similaires (travaux de Jacques Covès au sein de l'équipe ELMA).
- **L'hydrolyse de l'ATP par PAN n'est pas influencée par l'ajout de la GFPssrA ou de la particule 20S dans ces conditions expérimentales.**

Cela est en désaccord avec des résultats précédemment publiés et décrivant que PAN seul possède peu d'activité d'hydrolyse de l'ATP mais que cette activité est stimulée par l'ajout de GFPssrA, l'ajout de la particule 20S n'ayant pas d'effet<sup>250,262</sup>. Cependant, il est important de noter que les conditions d'étude choisies pour le SANS sont très différentes de celles des précédents travaux, utilisant 4  $\mu$ M GFPssrA, 2,8 nM PAN, 14 nM particule 20S et 1 mM ATP à 45°C<sup>250</sup> alors que le  $K_M$  apparent était de 0,5 mM<sup>277</sup>. Dans les études publiées, la concentration d'ATP utilisée est trop faible pour permettre une activité d'hydrolyse de l'ATP par PAN proche de la  $v_{max}$ . Dans ces conditions, des nuances peuvent être observées, telles que la stimulation de l'activité d'hydrolyse de l'ATP par PAN lorsque la GFPssrA est présente en solution. À une concentration initiale de 100 mM ATP dans les conditions d'études, cette activité est proche de la  $v_{max}$ , ne permettant pas d'observer de variation lors de l'ajout de GFPssrA.

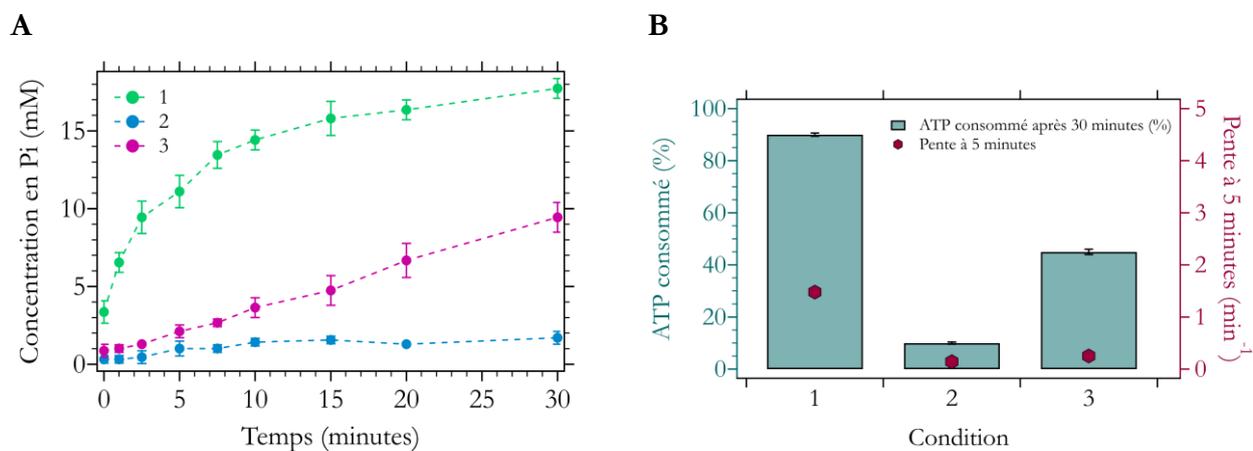
La consommation de l'ATP par PAN ne dépasse pas les 80% (80 mM sur les 100 mM initiaux) dans toutes les conditions testées. Avec les 20% restant, soit 20 mM, la concentration d'ATP est toujours environ 7,5 fois supérieure au  $K_M$  mesuré de PAN pour l'ATP (partie 6.1). PAN devrait donc toujours être à un niveau d'activité d'hydrolyse de l'ATP non négligeable. Cependant, la concentration du produit de réaction d'hydrolyse de l'ATP, l'ADP, peut avoir un effet d'inhibition sur l'activité d'hydrolyse de l'ATP une fois présent à 80 mM en solution, comme suggéré dans la littérature<sup>396</sup>. Une autre hypothèse serait que le dépliement de la GFPssrA par PAN nécessite un seuil minimum de concentration d'ATP, comme il a été observé pour la prise en charge de certains substrats protéiques par le système bactérien ClpXP<sup>289,321,322</sup>.

### 6.3. Effet d'inhibition de l'ADP sur l'activité ATPase de PAN

L'éventuel effet de l'inhibition de l'ADP sur l'hydrolyse de l'ATP par PAN en présence de GFPssrA a été investigué dans la condition SANS 1. La quantité de phosphate inorganique libre en solution au cours du temps a été mesurée pendant 30 minutes de réaction à 55°C en triplicata dans les conditions suivantes à 200 mM MgCl<sub>2</sub> :

- 20 mM ATP
- 80 mM ADP
- 20 mM ATP et 80 mM ADP, qui représentent les concentrations d'ATP et d'ADP en fin de réaction SANS dans les conditions d'études.

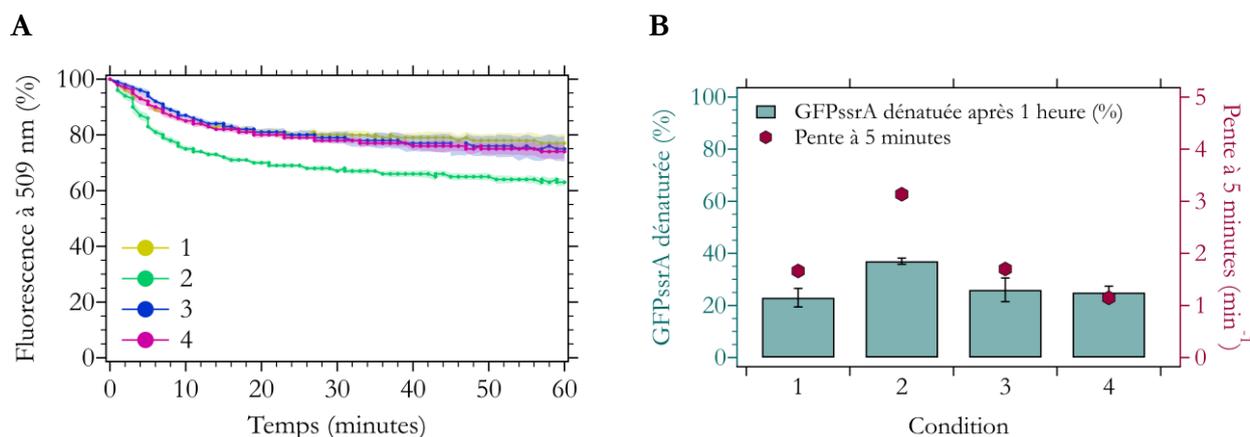
Les données sont présentées dans la figure 91 de façon similaire à la partie précédente.



**Figure 91 : Mesure de l'hydrolyse de l'ATP dans différentes conditions** par mesure de la concentration de phosphate organique libre en solution dans la condition SANS 1 à 200 mM MgCl<sub>2</sub>. Pour toutes les figures, les conditions testées sont numérotées de 1 à 3 correspondant concentrations initiales de nucléotides : (1) 20 mM ATP, (2) 80 mM ADP et (3) 20 mM ATP & 80 mM ADP. **A.** Courbe d'évolution la concentration de phosphate libre en solution en fonction du temps dans les différentes conditions testées. **B.** Histogramme avec le pourcentage d'ATP consommé après 30 minutes de réaction (par mesure de la concentration de phosphate libre en solution) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

À 80 mM ADP (2), le pourcentage de phosphate inorganique libre après 30 minutes à 55°C est de 10% et la pente à 5 minutes est de 0,15 min<sup>-1</sup>. Cela correspond aux valeurs expérimentales lorsqu'il n'y a pas d'hydrolyse de l'ATP (voir partie 6.2). À 20 mM ATP (1), le pourcentage de phosphate inorganique libre après 30 minutes à 55°C est de 90% tandis que la pente à 5 minutes est de 1,5 min<sup>-1</sup>. Il s'agit du contrôle positif d'hydrolyse de l'ATP dans les conditions de mesure. À 20 mM ATP et 80 mM ADP (3), le pourcentage de phosphate inorganique libre après 30 minutes à 55°C est de 45% et la pente à 5 minutes est de 0,25 min<sup>-1</sup>. L'activité d'hydrolyse de l'ATP par PAN est donc moins importante que pour le témoin positif, suggérant que la présence de 80 mM d'ADP en solution inhibe l'hydrolyse par PAN des 20 mM d'ATP.

Une mesure par spectroscopie de fluorescence de la dénaturation de la GFPssrA en fonction du temps a été réalisée dans les mêmes conditions (avec 70 μM GFPssrA) afin de déterminer l'effet de ces mêmes conditions sur la dénaturation de la GFPssrA. Les résultats sont représentés dans la figure 92 tel que précédemment.



**Figure 92** : Mesure de l'activité du système à 55°C dans la condition SANS 1 à 200 mM MgCl<sub>2</sub> dans différentes conditions. Pour toutes les figures, les conditions testées sont numérotées de 1 à 4 correspondant concentrations initiales de nucléotides : GFPssrA à (1) 20 mM ATP & 80 mM ADP et GFPssrA en présence de PAN à (2) 20 mM ATP, (3) 80 mM ADP et (4) 20 mM ATP & 80 mM ADP. **A**. Courbe d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire de la même couleur. **B**. Histogramme avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Lorsque la GFPssrA est seule (1), avec PAN à 80 mM ADP (3), ou avec PAN à 20 mM ATP et 80 mM ADP (4), le pourcentage de GFPssrA dénaturée après 30 minutes est d'environ 25% tandis que la pente à 5 minutes est de 1 à 1,7 min<sup>-1</sup>. Cela correspond à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C. Dans la condition où la GFPssrA est en présence de PAN à 20 mM ATP (2), le pourcentage de GFPssrA dénaturée après 30 minutes est de 37% tandis que la pente à 5 minutes est de 3 min<sup>-1</sup>. Il n'y a donc dénaturation de la GFPssrA par PAN que dans la condition en présence de 20 mM ATP seulement, et non en présence de 20 mM ATP et 80 mM ADP. L'effet de compétition/inhibition de l'ADP sur l'activité d'hydrolyse de l'ATP semble donc se répercuter sur l'activité de dépliage de la GFPssrA.

Deux conclusions principales sont à retenir :

- **L'ADP présent en solution à haute concentration a un effet d'inhibition sur l'hydrolyse de l'ATP par PAN.** Ces mesures expliquent que le système ne consomme que 80 mM ATP sur les 100 mM initiaux dans les conditions d'étude. Il est probable que l'ADP entre en compétition réversible/transitoire avec les sites de fixation de l'ATP de PAN. Cela est d'ailleurs suggéré dans la littérature<sup>323,396</sup>. Cela n'est pas représentatif des conditions physiologiques, où l'ADP est recyclé en ATP au sein des cellules. Dans ces circonstances, l'introduction d'un système de régénération de l'ATP serait intéressante. Cependant, comme mentionné précédemment (partie 2.1), cela compliquerait le système déjà complexe mesuré par SANS.
- **L'effet d'inhibition de l'ADP sur l'hydrolyse de l'ATP empêche l'activité de dépliage de la GFPssrA par PAN.** Ces résultats confirment que la cinétique de dépliage de la GFPssrA par PAN est corrélée à la cinétique d'hydrolyse de l'ATP par PAN comme suggéré par certaines études<sup>250,286,373,393</sup>.

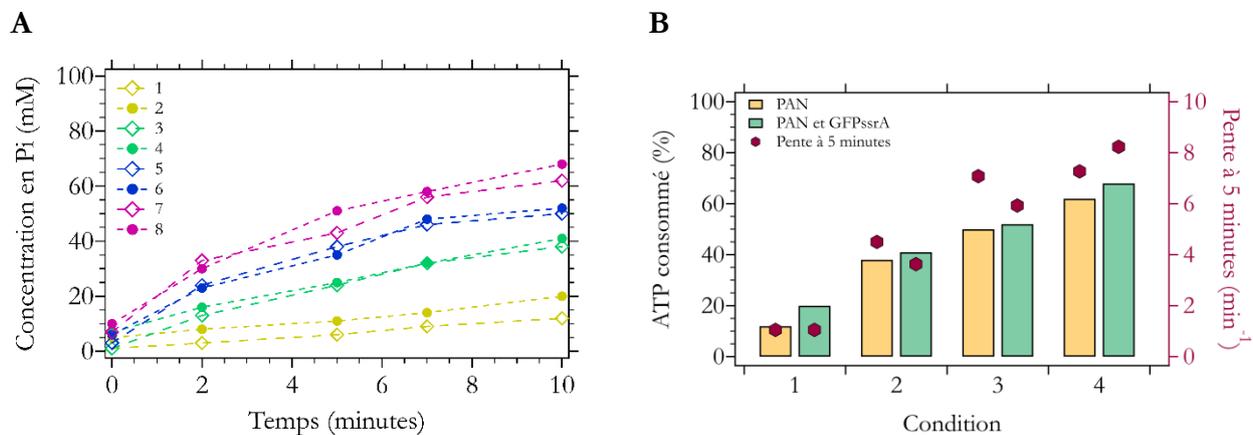
## 6.4. Effet de l'ajout de GFPssrA sur l'hydrolyse de l'ATP par PAN

### 6.4.1. Effet à 100 mM ATP

L'effet de la présence de GFPssrA en solution sur l'activité d'hydrolyse de l'ATP par PAN a été étudié. La quantité de phosphate inorganique libre en solution au cours du temps a été mesurée pendant 10 minutes de réaction à 55°C avec 100 mM ATP & 200 mM MgCl<sub>2</sub> dans les conditions suivantes :

- (1) 1 μM PAN
- (2) 1 μM PAN & 180 μM GFPssrA
- (3) 5 μM PAN
- (4) 5 μM PAN & 180 μM GFPssrA
- (5) 10 μM PAN
- (6) 10 μM PAN & 180 μM GFPssrA
- (7) 20 μM PAN
- (8) 20 μM PAN & 180 μM GFPssrA

Les données présentées dans la figure 93 de la même manière que dans la partie 6.2. Chaque condition n'a été mesurée qu'une fois et on interprètera donc les résultats de façon qualitative.



**Figure 93 : Mesure de l'hydrolyse de l'ATP en fonction du temps à 100 mM ATP & 200 mM MgCl<sub>2</sub> par mesure de la concentration de phosphate organique libre en solution dans différentes conditions. A.** Courbe d'évolution la concentration de phosphate libre en solution en fonction du temps dans les différentes conditions testées. Les conditions sont numérotées de 1 à 8 correspondant aux conditions à (1) 1 μM PAN, (2) 1 μM PAN & 180 μM GFPssrA, (3) 5 μM PAN, (4) 5 μM PAN & 180 μM GFPssrA, (5) 10 μM PAN, (6) 10 μM PAN & 180 μM GFPssrA, (7) 20 μM PAN et (8) 20 μM PAN & 180 μM GFPssrA. **B.** Histogramme avec le pourcentage d'ATP consommé après 10 minutes de réaction (par mesure de la concentration de phosphate libre en solution) dans les différentes conditions testées. Les conditions sont numérotées de 1 à 4 correspondant aux conditions à 1 μM (1), 5 μM (2), 10 μM (3) ou 20 μM PAN (4). Les barres jaunes correspondent à PAN seul et les barres vertes à PAN en présence de GFPssrA (180 μM). La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

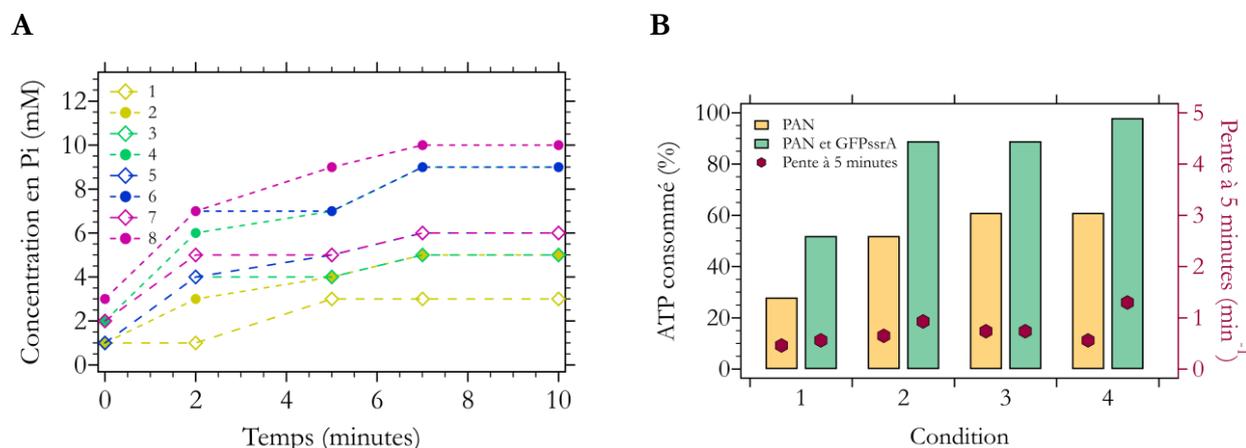
En absence de GFPssrA (série de couleur jaune, figure 93.B), plus la concentration de PAN augmente, plus le pourcentage de phosphate inorganique libre en fin de réaction et la pente à 5 minutes augmentent. Cela est logique puisque PAN possède une activité ATPasique, et comme précédemment décrit (partie 6.2), l'hydrolyse de l'ATP est plus efficace lorsque la concentration de PAN est importante.

L'ajout de GFPssrA pour une même concentration de PAN (série de couleur verte, figure 93.B) n'a pas d'impact significatif sur le pourcentage de phosphate inorganique libre en fin de réaction et

la pente à 5 minutes. Cela signifie que l'ajout de GFPssrA n'a pas d'effet de stimulation de l'hydrolyse de l'ATP par PAN aux concentrations de PAN testées et à 100 mM ATP.

#### 6.4.2. Effet à 10 mM ATP

Des mesures identiques à celle de la partie précédente ont été réalisées à 10 mM ATP & 200 mM MgCl<sub>2</sub>. Les données sont représentées tel que précédemment dans la figure 94.



**Figure 94 : Mesure de l'hydrolyse de l'ATP en fonction du temps à 10 mM ATP & 200 mM MgCl<sub>2</sub>** par mesure de la concentration de phosphate organique libre en solution dans différentes conditions. **A.** Courbe d'évolution la concentration de phosphate libre en solution en fonction du temps dans les différentes conditions testées. Les conditions sont numérotées de 1 à 8 correspondant aux conditions à (1) 1 μM PAN, (2) 1 μM PAN & 180 μM GFPssrA, (3) 5 μM PAN, (4) 5 μM PAN & 180 μM GFPssrA, (5) 10 μM PAN, (6) 10 μM PAN & 180 μM GFPssrA, (7) 20 μM PAN et (8) 20 μM PAN & 180 μM GFPssrA. **B.** Histogramme avec le pourcentage d'ATP consommé après 10 minutes de réaction (par mesure de la concentration de phosphate libre en solution) dans les différentes conditions testées. Les conditions sont numérotées de 1 à 4 correspondant aux conditions à 1 μM (1), 5 μM (2), 10 μM (3) ou 20 μM PAN (4). Les barres jaunes correspondent à PAN seul et les barres vertes à PAN en présence de GFPssrA (180 μM). La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

En absence de GFPssrA (série de couleur jaune, figure 94.B), le pourcentage de phosphate inorganique libre en fin de réaction augmente avec la concentration de PAN, et semble saturer après 10 μM PAN (pas d'augmentation significative entre 10 et 20 μM PAN). De la même façon qu'à 100 mM ATP, plus la concentration de PAN à 10 mM ATP est importante, plus l'hydrolyse de l'ATP est efficace.

Le pourcentage de phosphate inorganique libre en fin de réaction et la pente à 5 minutes augmentent drastiquement en présence de GFPssrA pour une même concentration de PAN (série de couleur verte, figure 94.B). L'ajout de GFPssrA a donc un effet de stimulation de l'hydrolyse de l'ATP par PAN aux concentrations de PAN testées et à 10 mM ATP.

Si l'on compare ces résultats à 10 mM ATP avec les mesures précédentes à 100 mM ATP (partie 6.4) on peut conclure que **la GFPssrA ne stimule l'activité d'hydrolyse de l'ATP de PAN que lorsque la concentration d'ATP est plus faible que celles des conditions d'études SANS**. À 10 mM ATP, la concentration est 4 fois supérieure au K<sub>M</sub> apparent mesuré (partie 6.1), ce qui ne suffit pas à être à la v<sub>max</sub>. Le système sera donc sensible à des variations de vitesse. Ces résultats sont similaires à ceux des précédents travaux où l'effet de stimulation de l'hydrolyse de l'ATP par

PAN avait été observé lors de l'ajout de GFP<sub>ssrA</sub> lorsque la concentration d'ATP ne permettait pas une activité de PAN à capacité maximale<sup>250,262</sup> (voir partie 6.2).

## 7. Conclusion générale du chapitre

Le but principal de ce chapitre était de définir les conditions permettant une activité optimale pour les mesures SANS (concentrations des composants, température, temps de mesure) et d'éliminer les incertitudes de manière à interpréter de façon rigoureuse les résultats SANS.

Le criblage des conditions de concentrations et des stœchiométries des différents composants du système par spectroscopie de fluorescence a permis de sélectionner deux conditions SANS :

- **Condition SANS 1 – 70  $\mu$ M GFP<sub>ssrA</sub>, 20  $\mu$ M PAN et 20  $\mu$ M particule 20S**
- **Condition SANS 2 – 180  $\mu$ M GFP<sub>ssrA</sub>, 5  $\mu$ M PAN et 30  $\mu$ M particule 20S**

Ces conditions vont pouvoir permettre de réaliser un point de mesure en SANS toutes les **30 secondes**, tout en ayant un signal provenant de la GFP<sub>ssrA</sub> exploitable. Les mesures de l'évolution de la **fluorescence à 509 nm** pendant 1 heure montrent également qu'un temps de mesure de **45 minutes** est suffisant pour suivre toute la réaction de dénaturation de la GFP<sub>ssrA</sub> par PAN (avec ou sans particule 20S).

La mesure du  $K_M$  apparent de PAN pour l'ATP et la quantification de l'hydrolyse de l'ATP au cours du temps ont confirmé que la concentration de **100 mM ATP** était adéquate pour les conditions d'étude. De plus, la quantification de l'hydrolyse de l'ATP dans les deux conditions SANS et les tests complémentaires réalisés ont permis démontrer que **la cinétique d'hydrolyse de l'ATP module la cinétique de dépliement de la GFP<sub>ssrA</sub> par PAN.**

La dégradation de la GFP<sub>ssrA</sub> par la particule 20S suite à son dépliement par PAN a pu être étudiée par Western-blot anti-GFP ainsi que par spectrométrie de masse MALDI-TOF. La particule 20S est donc active dans les conditions d'étude et génère des **peptides de 10 acides aminés en moyenne.**

Enfin, d'après la spectrométrie de fluorescence, la température de **55°C** correspond à la température appropriée pour moduler l'activité du système tout en limitant la diminution de l'efficacité quantique de la fluorescence de la GFP<sub>ssrA</sub>.

Le comportement des protéines dans les concentrations élevées est similaire à celui décrit dans la littérature. Les conditions mises au point vont pouvoir maintenant être appliquées aux expérimentations par SANS afin de suivre le dépliement et la dégradation de la GFP<sub>ssrA</sub> par la machinerie protéolytique dans des conditions optimales.

## Chapitre 7 :

# Diffusion de neutrons aux petits angles résolue en temps sur le système GFP<sub>ssrA</sub>, protéasome PAN-20S

Ce chapitre présente l'expérience centrale du projet de thèse : l'étude par TR-SANS couplée à la mesure de la fluorescence, du dépliement et de la dégradation d'une protéine substrat par le protéasome PAN-20S. Grâce à la variation de contraste et à la deutération de la protéine substrat GFP<sub>ssrA</sub>, les modifications structurales induites par la machinerie protéolytique ont pu être suivies au cours du temps. Les résultats obtenus ont permis de mieux comprendre le mode d'action du protéasome PAN-20S.

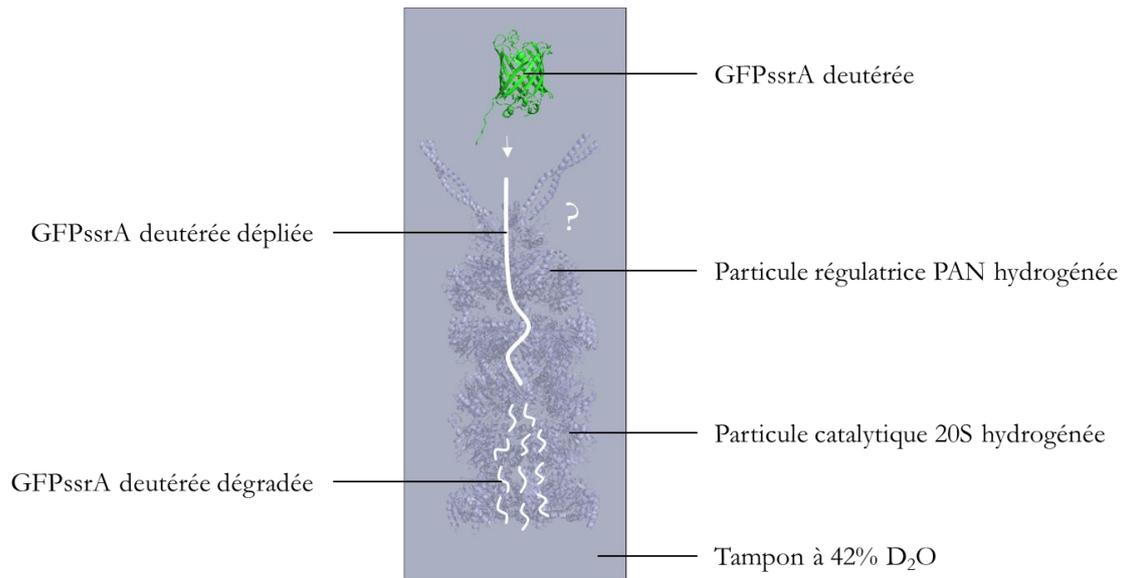
### 1. Rappel du contexte de l'expérience

En utilisant un protéasome provenant de l'archée hyperthermophile *M. jannaschii*, la température de 55°C a été utilisée pour activer la machinerie protéolytique et contrôler la vitesse du processus de dégradation pour la rendre analysable avec un point de mesure toutes les 30 secondes.

Le SANS permet d'étudier en solution les structures et les changements de conformation des édifices macromoléculaires. Contrairement à la diffusion des rayons X aux petits angles (SAXS), plus courante et plus répandue, le SANS est capable de distinguer différentes protéines grâce à la deutération des protéines et de la variation du contraste des solvants (échange H<sub>2</sub>O/D<sub>2</sub>O). En particulier, à environ 42 % de D<sub>2</sub>O, le signal provenant des protéines naturelles (c'est-à-dire hydrogénées) peut être masqué, et le signal SANS est alors attribué aux protéines deutérées (voir partie 3.5 du chapitre 2). Ce procédé de variation de contraste a été utilisé dans le présent travail pour masquer le signal de la machinerie protéolytique hydrogénée (h) (h-PAN et particule h-20S) et d'enregistrer exclusivement le signal de la protéine substrat deutérée (d) (d-GFP<sub>ssrA</sub>) pendant le processus de dégradation en solution.

Le substrat protéique utilisé est la protéine fluorescente GFP possédant une étiquette d'adressage *ssrA* (GFP<sub>ssrA</sub>) qui nous a permis d'obtenir des informations complémentaires sur l'état de repliement de la protéine grâce à la mesure de fluorescence *in situ*, en parallèle des données SANS. En effet, la GFP<sub>ssrA</sub> est fluorescente uniquement sous forme native et perd sa fluorescence lorsqu'elle est dénaturée (voir partie 4.1.3 du chapitre 5).

La figure 95 illustre le système étudié.



**Figure 95 : Représentation du système étudié par TR-SANS.** Dans un tampon à 42% D<sub>2</sub>O, le signal des protéines hydrogénées (PAN et particule 20S) est masqué et seul le signal de la protéine deutérée (GFPssrA) est enregistré. Le processus de dépliement et de dégradation de la GFPssrA deutérée par le protéasome PAN-20S peut alors être suivi de façon sélective en enregistrant uniquement le signal de la GFPssrA deutérée et de ses formes dérivées (native, agrégats, peptides, etc.). La GFPssrA deutérée sous forme native est illustrée en vert car elle possède une fluorescence verte. Une fois dénaturée, elle perd sa fluorescence, et est alors représentée en blanc sur la figure. La perte de fluorescence de la GFPssrA a été suivie en parallèle des mesures TR-SANS. Ces formes dénaturées, représentées en blanc, sont toujours visibles par SANS, car elles sont toujours deutérées. Les étapes du processus de dépliement et de dégradation sont le questionnement biologique que cette étude cherche à élucider, ce qui est représenté sur la figure par un point d'interrogation.

## 2. Mesures TR-SANS et de fluorescence

### 2.1. Préambule

Deux conditions ont été sélectionnées à la suite des études biochimiques présentées dans le chapitre 6 :

- Condition SANS 1 – 70  $\mu$ M GFPssrA, 20  $\mu$ M PAN et 20  $\mu$ M particule 20S
- Condition SANS 2 – 180  $\mu$ M GFPssrA, 5  $\mu$ M PAN et 30  $\mu$ M particule 20S

Les raisons du choix de ces deux conditions sont expliquées en détail dans la partie 2.3 du chapitre 6. Des mesures TR-SANS ont été réalisées sur la ligne de faisceaux D22 à l'ILL. Pour les deux conditions SANS, les échantillons suivants ont été mesurés :

- (1) d-GFPssrA en présence de 100 mM ATP
- (2) d-GFPssrA et h-PAN en présence de 100 mM ATP
- (3) d-GFPssrA, h-PAN et particule h-20S en présence de 100 mM ATP
- (4) d-GFPssrA et particule h-20S en présence de 100 mM ATP
- (5) d-GFPssrA et h-PAN
- (6) d-GFPssrA, h-PAN et particule h-20S

Tous les échantillons à 100 mM ATP contenaient également 200 mM MgCl<sub>2</sub>. Les échantillons 4, 5 et 6 étaient des contrôles négatifs. Par soucis de clarté, seuls les résultats obtenus pour les

échantillons 1, 2 et 3 des deux conditions SANS seront présentés dans ce manuscrit. Toutes les réactions présentées dans ce chapitre possédaient donc une concentration initiale de 100 mM ATP.

Les réactions ont été initiées par la disposition des échantillons dans le porte-échantillon à 55°C et les mesures ont été suivies pendant 45 minutes avec un point de mesure SANS et de la fluorescence à 509 nm toutes les 30 secondes. La figure 96 est une photographie des échantillons 1, 2 et 3 de la condition SANS 2 après leur mesure.

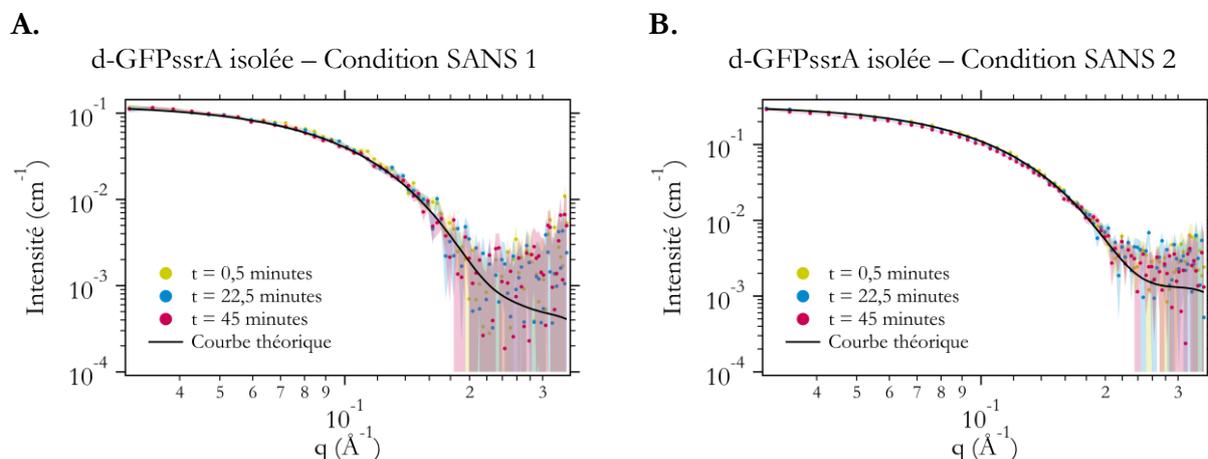


**Figure 96 : Photographie des principaux échantillons de la condition SANS 2** après 45 minutes de mesure par TR-SANS à 55°C. (1) d-GFPssrA, (2) d-GFPssrA en présence de h-PAN, (3) d-GFPssrA en présence de h-PAN et de la particule h-20S. Les trois échantillons ont été préparés avec une concentration initiale de 100 mM ATP.

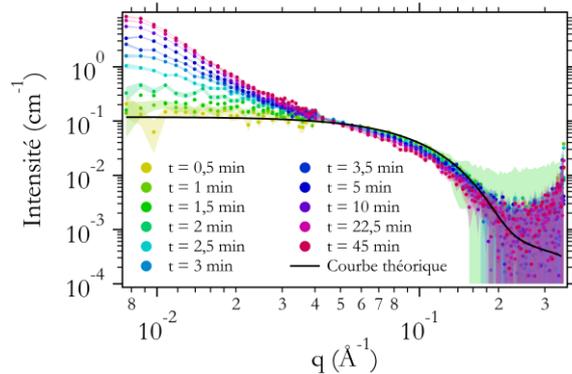
L'aspect macroscopique des échantillons après 45 minutes de réaction permet de visualiser à l'œil nu les conséquences de la réaction biochimique sur l'état des échantillons. L'échantillon de d-GFPssrA isolée (1) est limpide et arbore une couleur verte similaire à l'état initial (non représenté ici). L'échantillon de d-GFPssrA en présence de h-PAN (2) présente un aspect visiblement agrégé, tandis que l'échantillon de d-GFPssrA avec la machinerie protéolytique complète hydrogénée (3) est limpide avec une couleur verte très atténuée par rapport à celle de la d-GFPssrA isolée. Les mesures TR-SANS, associées à la mesure de la fluorescence à 509 nm, ont permis d'étudier les réactions à un niveau moléculaire au cours du temps.

## 2.2. Courbes SANS au cours du temps pour les deux conditions SANS

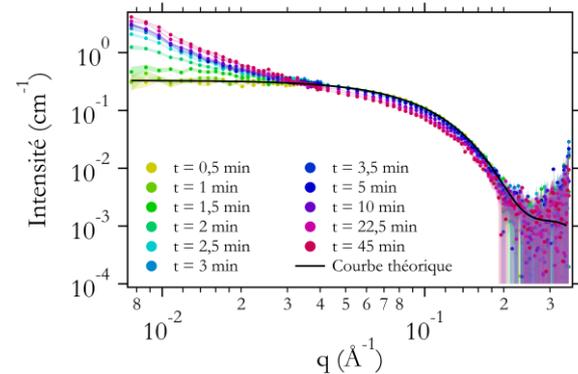
Dans le tampon à 42% D<sub>2</sub>O, les signaux de la protéine h-PAN et particule h-20S sont masqués, et le signal SANS provient uniquement de la d-GFPssrA. La figure 97 représente les courbes expérimentales SANS en fonction du temps des différents échantillons mesurés dans les deux conditions SANS.



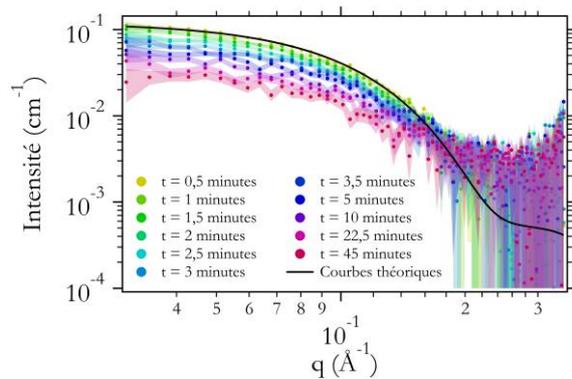
**C.** d-GFPssrA et h-PAN – Condition SANS 1



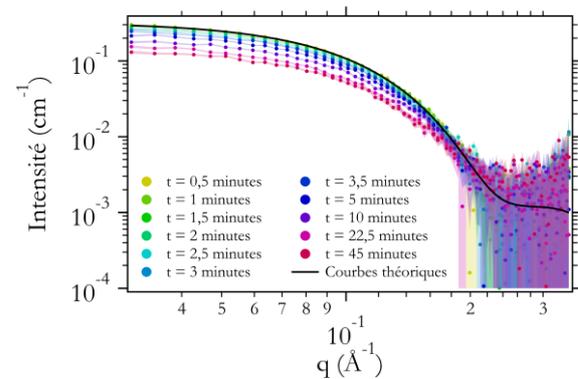
**D.** d-GFPssrA et h-PAN – Condition SANS 2



**E.** d-GFPssrA, h-PAN et particule h-20S – Cond. 1



**F.** d-GFPssrA, h-PAN et particule h-20S – Cond. 2



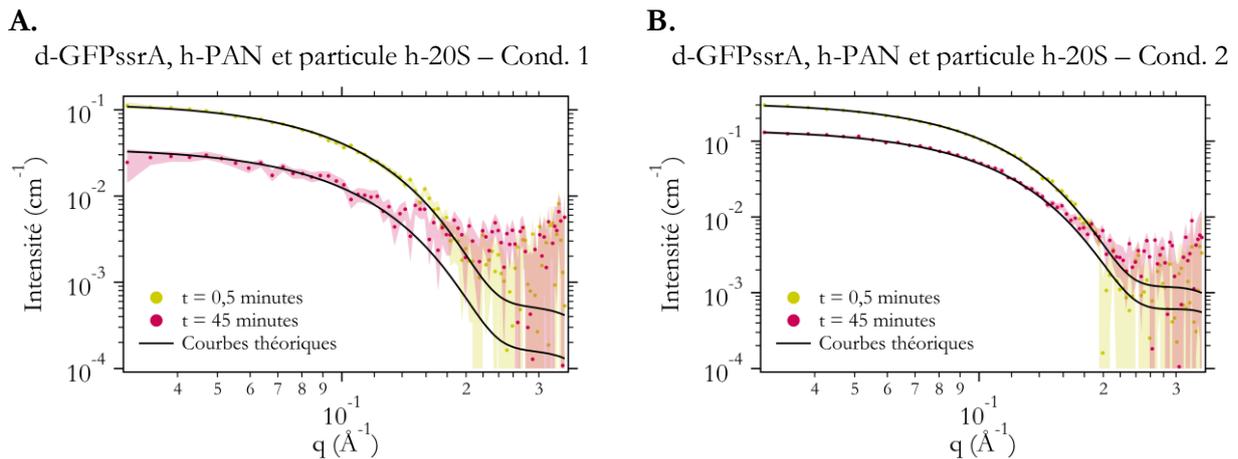
**Figure 97 : Courbes expérimentales SANS en fonction du temps sur le système GFPssrA, protéasome PAN-20S mesurées sur la ligne D22 (ILL).** Les échantillons des conditions SANS 1 et 2 sont représentés dans les panneaux de droite (A, C & E) et de gauche (B, D & F). **A & B.** d-GFPssrA isolée. **C & D.** d-GFPssrA en présence de h-PAN. **E & F.** d-GFPssrA en présence de h-PAN et de la particule h-20S. Pour chaque figure, une courbe SANS théorique calculée par le logiciel CRYSON à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) est représentée sous forme de courbe noire superposée à la première courbe expérimentale. Les barres d'erreur sont représentées pour chaque courbe expérimentale sous forme d'aire nuancée de la même couleur.

Les résultats obtenus dans les conditions SANS 1 et 2 montrent les mêmes tendances. Les courbes expérimentales SANS de la d-GFPssrA isolée (figure 97.A et 97.B) sont toutes identiques au cours du temps et sont superposables à la courbe théorique de la GFP native pour tous les vecteurs de diffusion  $q$  représentés. La d-GFPssrA isolée conserve donc sa structure native pendant 45 minutes à 55°C.

En présence de h-PAN (figure 97.C et 97.D), la d-GFPssrA présente une courbe expérimentale initiale ( $t = 0,5$  minutes) superposable à la courbe théorique de la GFP native pour tous les vecteurs de diffusion  $q$  représentés. Des changements significatifs dans les courbes SANS apparaissent ensuite progressivement au cours du temps. L'intensité de diffusion augmente aux valeurs  $q$  faibles ( $q < 0,03 \text{ \AA}^{-1}$ ) et diminue aux moyens angles ( $0,04 \text{ \AA}^{-1} < q < 0,1 \text{ \AA}^{-1}$ ), tout en restant stable aux angles plus élevés ( $q > 0,15 \text{ \AA}^{-1}$ ). Ce phénomène indique la formation progressive d'agrégats au cours de la réaction<sup>40,333,334</sup>.

Lorsque la d-GFPssrA est avec le protéasome h-PAN-h-20S (figure 97.E et 97F), la courbe SANS initiale ( $t = 0,5$  minutes) est en accord avec la structure native de la GFP. Les courbes SANS

évoluent ensuite progressivement au cours du temps. En effet, l'intensité de diffusion diminue aux faibles valeurs  $q$  ( $q < 0,15 \text{ \AA}^{-1}$ ) et augmente à des angles plus élevés ( $q > 0,15 \text{ \AA}^{-1}$ ). Les courbes SANS peuvent être raisonnablement bien superposées aux faibles angles au modèle GFP native tout au long de la réaction mais dévient aux angles plus élevés, pour lesquels l'intensité de diffusion est plus élevée que la courbe modèle. Cela est représenté dans la figure 98 sous forme de courbe théorique de la GFP superposée à la première et dernière courbe expérimentale. Cette augmentation de l'intensité de diffusion aux angles plus élevés indique l'apparition de produits plus petits qui se dispersent de façon plus homogène sur une large gamme de  $q$  et jusqu'à des angles plus élevés<sup>333,334</sup>.



**Figure 98 : Première (jaune) et dernière (rouge) courbe expérimentale SANS des échantillons d-GFPssrA en présence de h-PAN dans les conditions SANS 1 (A) et 2 (B).** Une courbe SANS théorique calculée par le logiciel CRYSON à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) est représentée sous forme de courbe noire superposée à chaque courbe expérimentale. Pour chaque courbe expérimentale, les barres d'erreur sont représentées sous forme d'aire nuancée de la même couleur.

On peut donc conclure des courbes expérimentales SANS que :

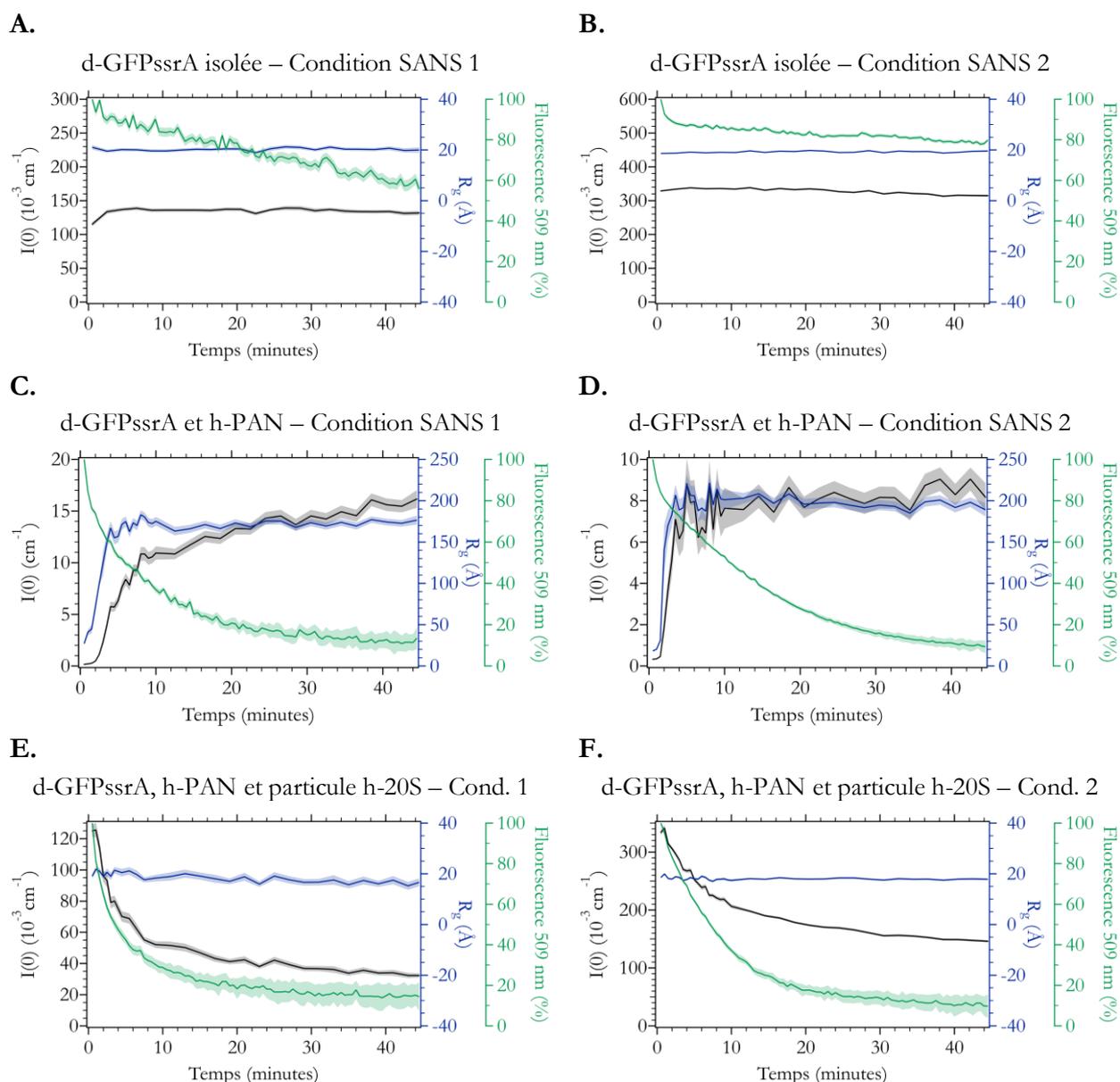
- **La d-GFPssrA isolée à 55°C est stable sous forme native au cours du temps.** Cela est en accord avec les mesures biochimiques précédemment réalisées (chapitre 6) et avec la photographie des échantillons (partie 2.1).
- **La présence de h-PAN provoque l'agrégation progressive de la d-GFPssrA.** Cela concorde avec la photographie des échantillons (partie 2.1) et avec l'étude ayant précédé ce projet de thèse<sup>40</sup>. Dans cette étude, l'enregistrement des données a été réalisé dans une configuration supplémentaire (distance collimateur/détecteur de 8,0/8,0 m) par rapport à la configuration utilisée précédemment (distance collimateur/détecteur de 2,0/2,8 m). Cela permet d'accéder à plus d'informations sur les agrégats de d-GFPssrA aux faibles valeurs  $q$  ( $0,07 \text{ \AA}^{-1} < q < 0,03 \text{ \AA}^{-1}$ ). Les informations obtenues seront discutées dans les parties suivantes.
- **La machinerie complète hydrogénée empêche l'apparition d'agrégats de d-GFPssrA et entraîne la disparition progressive de la d-GFPssrA native au profit de produits plus petits.** Cela est en accord avec les études biochimiques présentées dans le chapitre 6 et avec la photographie des échantillons (partie 2.1). Cela sera également discuté de façon plus approfondie dans les parties suivantes.

Certains paramètres ont pu être extraits des mesures TR-SANS et complétés par les mesures de la fluorescence à 509 nm de la GFPssrA.

### 2.3. Évolution des paramètres de $I(0)$ et $R_g$ au cours du temps pour les deux conditions SANS

Le rayon de giration  $R_g$  et l'intensité diffusée à l'angle zéro  $I(0)$  ont été extraits à partir des courbes expérimentales SANS grâce à l'approximation de Guinier (équation 13). Comme décrit dans la partie 1.2.2 du chapitre 2, le  $R_g$  d'une protéine est un indicateur de sa taille et de sa compacité, tandis que  $I(0)$  est proportionnelle à la masse moléculaire  $M_w$  et à la concentration  $c$  des protéines en solution.

La figure 99 représente, pour chaque échantillon, l'évolution au cours du temps de l'intensité diffusée à angle nul  $I(0)$  (noir), du rayon de giration  $R_g$  (bleu), et de la fluorescence à 509 nm (vert).



**Figure 99 :** Évolution au cours du temps de l'intensité diffusée à l'angle zéro  $I(0)$ , du rayon de giration  $R_g$ , et de la fluorescence à 509 nm des échantillons des conditions SANS 1 et 2, respectivement dans les panneaux de droite (A, C & E) et de gauche (B, D & F). **A & B.** d-GFPssrA isolée. **C & D.** d-GFPssrA en présence de h-PAN. **E & F.** d-GFPssrA en présence de h-PAN et de la particule h-20S. Pour toutes les réactions, un point de mesure de fluorescence à 509 nm, de  $R_g$  et de  $I(0)$  a été enregistré toutes les 30 secondes. Après 10 minutes de réaction, les

données SANS ont été moyennées pour avoir un point de mesure de  $R_g$  et de  $I(0)$  toutes les 2 minutes. Pour chaque courbe, les barres d'erreur sont représentées sous forme d'aire nuancée de la même couleur.

L'évolution du  $I(0)$ ,  $R_g$  et de la fluorescence à 509 nm est similaire dans les conditions SANS 1 et 2. Lorsque la d-GFPssrA est isolée en solution (figure 99.A et 99.B), les valeurs de  $R_g$  et de  $I(0)$  restent stables au cours du temps, ce qui indique que la taille et la masse moléculaire sont inchangées pendant l'expérience. De plus, la valeur du rayon de giration obtenu expérimentalement (environ 20 Å) est en bon accord avec le rayon de giration théorique (19,5 Å) (voir partie 4.4 du chapitre 5). La fluorescence diminue quant à elle d'environ 40% pour la condition SANS 1 et de 20% pour la condition SANS 2. Cette diminution modérée de la fluorescence peut être attribuée à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C (voir partie 1.2 du chapitre 6). La diminution de la fluorescence pour la condition SANS 1 apparaît relativement élevée mais l'allure de la courbe, avec une diminution linéaire et bruitée au cours du temps, semble indiquer un artefact expérimental. La d-GFPssrA reste donc sous forme native lorsqu'elle est isolée en solution pendant 45 minutes à 55°C, comme suggéré dans la partie précédente.

Les valeurs de  $R_g$  et de  $I(0)$  de la d-GFPssrA en présence de h-PAN (figure 99.C et 99.D) augmentent rapidement et drastiquement au cours du temps. La valeur du  $R_g$  augmente d'environ dix fois et la valeur du  $I(0)$  augmente d'environ cent fois. Cependant, malgré une mesure dans une configuration instrumentale supplémentaire (8 m/8 m) pour étudier les objets plus grands, l'approximation de Guinier n'est plus valide (puisque  $qR_g > 1,3$ ) après environ deux minutes de réaction. Cela empêche de déduire les valeurs réelles de  $R_g$  et de  $I(0)$ , qui sont alors sous-estimées. La fluorescence diminue en parallèle progressivement, indiquant une dénaturation de la d-GFPssrA. L'évolution de tous les paramètres moléculaires est donc compatible avec l'apparition rapide de grands agrégats de d-GFPssrA, comme suggéré dans la partie précédente.

Lorsque la d-GFPssrA est en présence de la machinerie protéolytique complète hydrogénée (figure 99.E et 99.F), l'intensité  $I(0)$  et la fluorescence diminuent progressivement. La diminution de la valeur du  $I(0)$  indique que la masse moléculaire moyenne des espèces en solution diminue, tandis que la diminution de la fluorescence signifie que la GFPssrA est dénaturée au cours du temps. En revanche, le  $R_g$ , qui est une mesure de la taille moyenne des espèces moléculaires et est fortement dominé par les grosses particules<sup>333</sup>, ne diminue quasiment pas. Ces données indiquent que la population de d-GFPssrA est dégradée progressivement par le protéasome h-PAN-h-20S, tandis que les d-GFPssrA natives restantes en solution dominent l'information du  $R_g$ . Cela est en accord avec les informations de la partie précédente.

Ainsi, on peut conclure de ces observations que :

- **La d-GFPssrA en présence de h-PAN forme des agrégats dont la taille dépasse les limites de détection de la technique.** Comme discuté dans le travail ayant précédé cette thèse<sup>40</sup>, cela signifie que **lorsqu'il n'est pas associé à la particule 20S, PAN déplie les protéines substrat et les relâche sous forme dénaturée en solution.** Ces protéines dénaturées présentent des caractéristiques structurales altérées, telles que des domaines hydrophobes exposés, qui les conduisent à s'agréger rapidement.
- **La d-GFPssrA est dégradée en présence de h-PAN et de h-20S et aucun agrégat n'est détecté.** Les mesures en spectrométrie de masse présentées dans la partie 4 du chapitre 6 ont révélé la présence de **peptides de l'ordre d'une dizaine d'acides aminés** en fin de réaction. Cela est cohérent avec l'évolution du  $I(0)$  qui indique la diminution de la masse moléculaire

moyenne des espèces en solution, mais aussi avec la détection dans la partie précédente d'une intensité de diffusion plus importante aux angles élevés par rapport à la courbe théorique.

L'absence d'agrégats en solution indique que **tous les substrats protéiques étant dépliés par h-PAN sont pris en charge directement par la particule catalytique h-20S**. En effet, une fraction même faible de substrats protéiques dépliés et relâchés en solution aurait formé des agrégats détectés par SANS, comme observé lorsque la d-GFPssrA est en présence de h-PAN seul.

L'évolution des paramètres pour les réactions où la d-GFPssrA est en présence de la machinerie protéolytique complète a été étudiée de façon plus approfondie.

#### 2.4. Pourcentages finaux des paramètres mesurés

Les études biochimiques sur l'hydrolyse de l'ATP présentées dans la partie 6.2 du chapitre 6 réalisées dans des conditions identiques à celles des expériences SANS ont permis de compléter les informations fournies par les mesures TR-SANS couplées à la spectroscopie de fluorescence. Le tableau 19 résume l'évolution des paramètres d'intérêt après 45 minutes de réaction pour la réaction d-GFPssrA en présence de h-PAN et h-20S dans les deux conditions SANS. Ces paramètres sont l'hydrolyse de l'ATP, l'intensité diffusée à angle nul  $I(0)$ , la fluorescence à 509 nm et le rayon de giration  $R_g$ .

**Tableau 19 : Valeurs initiales et finales (après 45 minutes à 55°C) de l'intensité diffusée à angle zéro  $I(0)$ , fluorescence à 509 nm et concentration en phosphate inorganique  $P_i$  pour les échantillons de GFPssrA contenant PAN et la particule 20S** dans les conditions SANS 1 et 2. Le pourcentage d'activité a été déduit pour chaque paramètre à partir de ces valeurs. (\*) La concentration initiale d'ATP était de 100 mM.

	Condition SANS 1	Condition SANS 2
	<b><math>I(0)</math></b>	
<b><math>I(0)</math> initial (<math>\text{cm}^{-1}</math>)</b>	$1,3 \cdot 10^{-1} \pm 3,8 \cdot 10^{-3}$	$3,3 \cdot 10^{-1} \pm 4,2 \cdot 10^{-3}$
<b><math>I(0)</math> final (<math>\text{cm}^{-1}</math>)</b>	$3,2 \cdot 10^{-2} \pm 1,3 \cdot 10^{-3}$	$1,4 \cdot 10^{-1} \pm 2,5 \cdot 10^{-3}$
<b>Perte de <math>I(0)</math> (%)</b>	<b>~74,1</b>	<b>~56,5</b>
	<b>Fluorescence</b>	
Fluorescence initiale (%)	$100 \pm 1,6$	$100 \pm 1,0$
Fluorescence finale (%)	$14,4 \pm 5,7$	$10,5 \pm 4,0$
<b>Perte de fluorescence (%)</b>	<b>~85,6</b>	<b>~89,5</b>
	<b>Hydrolyse de l'ATP</b>	
Concentration en $P_i$ initiale (mM)	$0 \pm 0$	$0 \pm 0$
Concentration en $P_i$ finale (mM)	$80 \pm 2$	$81 \pm 2$
<b>Hydrolyse de l'ATP (%) *</b>	<b>~80</b>	<b>~81</b>
	<b><math>R_g</math></b>	
<b><math>R_g</math> initial (Å)</b>	19,0	18,5
<b><math>R_g</math> final (Å)</b>	17,1	17,7
<b>Perte de <math>R_g</math> (%)</b>	<b>~10</b>	<b>~4,7</b>

Les paramètres de  $I(0)$  et de fluorescence sont tous deux des indicateurs du taux de dégradation de la d-GFPssrA tandis que la concentration en  $P_i$  est indicatrice du taux d'hydrolyse de l'ATP. Les valeurs obtenues sont à examiner de façon qualitative puisque les données de TR-SANS sont issues

d'une seule mesure. De plus, la mesure de l'hydrolyse de l'ATP a certes été effectuée dans les conditions SANS 1 et 2 mais dans des volumes plus petits et avec moins de points de mesures, puis les valeurs de concentration en  $P_i$  ont été déduites par comparaison à une courbe standard.

Dans la condition SANS 1 (gauche du tableau), le taux final d'hydrolyse de l'ATP est du même ordre de grandeur que la perte de fluorescence et de  $I(0)$  (respectivement environ 80%, 74,1% et 85,6%). Il y a donc un lien étroit entre les phénomènes d'hydrolyse de l'ATP et de dégradation des substrats. Toutes les molécules de d-GFPssrA dénaturées sont dégradées par la machinerie protéolytique.

Dans la condition SANS 2 (droite du tableau), le taux final d'hydrolyse de l'ATP est du même ordre de grandeur que la perte de fluorescence, tandis que la perte de  $I(0)$  est plus faible (respectivement environ 81%, 89,5% et 56,5%). La perte de fluorescence indique que la majorité de la population de d-GFPssrA est dénaturée en fin de réaction avec un taux similaire à l'hydrolyse de l'ATP (respectivement 89,5% et 81%). Cependant, la perte de  $I(0)$  est plus faible (environ 56,5%). Cela signifie que dans la condition SANS 2, presque toutes les d-GFPssrA sont dénaturées par h-PAN de façon modulée par l'hydrolyse de l'ATP, mais que seulement une partie des d-GFPssrA dénaturées (environ 60%) sont ensuite dégradées. Le fait que le  $R_g$  reste stable au cours de la réaction indique qu'il n'y a pas d'agrégation de la d-GFPssrA, comme décrit dans les parties précédentes. On peut en conclure que la forme déstabilisée de la d-GFPssrA est structuralement proche de la GFP native (à basse résolution) et ne présente pas de caractéristiques structurales altérées, telles que des domaines hydrophobes exposés, qui la mènerait à former des oligomères et/ou agrégats.

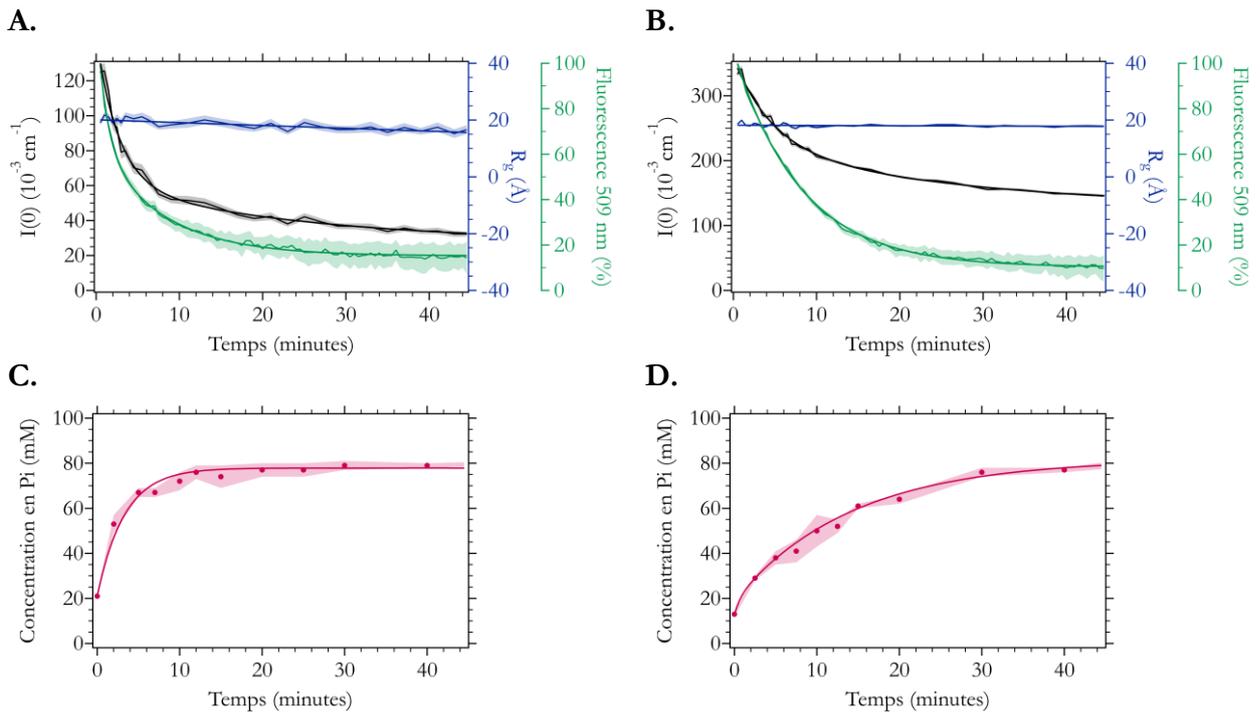
On peut donc conclure de ces données que :

- **Lorsqu'il y a 3,5 molécules d-GFPssrA pour un PAN (condition SANS 1), toutes les d-GFPssrA dénaturées par PAN sont dégradées par la particule catalytique 20S.**
- **Lorsqu'il y a 35 molécules d-GFPssrA pour un PAN (condition SANS 2), seulement 60% de d-GFPssrA dénaturées par PAN sont dégradées par la particule catalytique 20S. De plus, les GFPssrA déstabilisées qui n'ont pas été dégradées sont structuralement stables.**
- **Le taux final d'hydrolyse de l'ATP est corrélé à la dénaturation du substrat protéique par PAN dans les deux conditions.** Cela est accord avec les expériences réalisées dans le chapitre 6.

Ces phénomènes seront discutés de façon plus approfondie dans la partie 3. On peut aussi noter que la perte de fluorescence apparaît plus importante lors des mesures couplées au TR-SANS que lors des précédentes études au laboratoire (chapitre 6). Les différences constatées entre les deux dispositifs expérimentaux étaient le spectrofluorimètre utilisé et le contrôle de la température. Alors qu'il a été contrôlé que la température soit exactement à 55°C à l'intérieur de la cuvette contenant l'échantillon sur la ligne D22, la température du spectrofluorimètre utilisé au laboratoire a été contrôlée à l'intérieur de l'instrument et non directement dans l'échantillon. Une mesure indépendante de la température a permis de montrer que lorsque l'instrument était réglé à 55°C, la température de l'échantillon était d'environ 50°C. Il a été décrit dans la partie 5 du chapitre 6 que l'activité du protéasome PAN-20S dépendait fortement de la température. Cela peut donc expliquer les différences enregistrées.

## 2.5. Dépendance temporelle des différents paramètres mesurés

L'allure des courbes d'évolution des différents paramètres a également pu être étudiée en les comparant à des courbes modèles (linéaire, exponentielle ou bi-exponentielle) afin d'en déduire des constantes de temps caractéristiques. La figure 100 illustre les courbes d'évolution au cours du temps du  $I(0)$ ,  $R_g$ , et de la fluorescence (figures 100.A. et x.B) ou de la concentration de phosphate inorganique (figures 100.C. et 100.D) dans les deux conditions SANS telles que représentées précédemment. La courbe modèle utilisée pour chaque courbe expérimentale est représentée sous forme de ligne. Le tableau 20 résume les différents paramètres obtenus de cette analyse. Ces paramètres sont à considérer de façon qualitative pour les raisons énoncées dans la partie précédente.



**Figure 100 :** Évolution des différents paramètres ( $R_g$ ,  $I(0)$ , fluorescence et consommation d'ATP) dans les conditions SANS 1 et 2 pour les échantillons contenant de la GFPssrA en du protéasome PAN-20S. Les courbes expérimentales de  $I(0)$ , fluorescence et d'hydrolyse de l'ATP ont été ajustées à une fonction bi-exponentielle pour l'hydrolyse de l'ATP des conditions SANS 1 & 2 et la fluorescence de la condition SANS 1. La courbe expérimentale de fluorescence de la condition SANS 2 a été ajustée à une fonction exponentielle, et les courbes expérimentales de  $R_g$  pour les conditions SANS 1 & 2 ont été ajustées à une fonction linéaire. Le tableau 20 compile les différents paramètres extraits de ces ajustements. **A, B.** Évolution au cours du temps du  $I(0)$ ,  $R_g$  et de la fluorescence à 509 nm tel que représentés dans la figure 99. **C & D.** Hydrolyse de l'ATP au cours du temps tel que représentée dans la figure 90. Les courbes théoriques sont représentées pour chaque courbe expérimentale sous forme de ligne de la même couleur.

**Tableau 20: Paramètres relatifs au  $R_g$ ,  $I(0)$ , à la fluorescence et à la consommation d'ATP dans les conditions SANS 1 et 2 pour les échantillons contenant de la GFPssrA en présence de PAN et de particule 20S.** Les constantes de temps caractéristiques  $t_1$  et  $t_2$  pour le  $I(0)$  et l'hydrolyse de l'ATP des conditions SANS 1 & 2 et la fluorescence de la conditions SANS 1 ont été obtenues à l'aide d'une fonction bi-exponentielle ( $f(t) = y_0 + A_1 \cdot \exp(-t/t_1) + A_2 \cdot \exp(-t/t_2)$ ). La constante de temps caractéristique  $t_1$  pour la fluorescence de la condition SANS 2 a été obtenue à l'aide d'une fonction exponentielle ( $f(t) = y_0 + A_1 \cdot \exp(-t/t_1)$ ), et les paramètres  $a$  et  $b$  pour le  $R_g$  a ont été obtenus à l'aide d'une fonction linéaire ( $f(x) = a \cdot x + b$ ).

	Condition SANS 1	Condition SANS 2
	<b>I(0)</b>	
$\chi^2$	21,84	32,45
$y_0$	$0,03 \pm 0,01$	$0,13 \pm 0,01$
$A_1$	$0,03 \pm 0,00$	$0,10 \pm 0,01$
$A_2$	$0,08 \pm 0,01$	$0,13 \pm 0,02$
$t_1$ (minutes)	$23,95 \pm 12,90$	$27,79 \pm 12,90$
$t_2$ (minutes)	<b><math>2,66 \pm 0,43</math></b>	<b><math>4,57 \pm 0,67</math></b>
	<b>Fluorescence</b>	
$\chi^2$	5,64	16,05
$y_0$	$15,21 \pm 1,05$	$10,28 \pm 0,58$
$A_1$	$50,29 \pm 5,36$	$95,13 \pm 0,69$
$A_2$	$56,81 \pm 5,07$	/
$t_1$ (minutes)	$7,77 \pm 1,10$	<b><math>8,03 \pm 0,15</math></b>
$t_2$ (minutes)	<b><math>1,16 \pm 0,22</math></b>	/
	<b>Hydrolyse de l'ATP</b>	
$\chi^2$	15,68	7,21
$y_0$	$76,75 \pm 2,72 \cdot 10^5$	$81,75 \pm 1,52$
$A_1$	$1,11 \pm 2,72 \cdot 10^5$	$-62,66 \pm 5,43$
$A_2$	$-56,66 \pm 2,99$	$-6,10 \pm 6,29$
$t_1$ (minutes)	$1,13 \cdot 10^4 \pm 2,77 \cdot 10^9$	$14,23 \pm 1,95$
$t_2$ (minutes)	<b><math>3,30 \pm 0,41</math></b>	<b><math>0,72 \pm 4,59</math></b>
	<b>R<sub>g</sub></b>	
$\chi^2$	21,50	46,67
a	$20,11 \pm 0,38$	$18,20 \pm 0,11$
b	$-0,10 \pm 0,02$	$-0,01 \pm 0,01$

La diminution de I(0) et de la fluorescence ainsi que l'hydrolyse de l'ATP peuvent être décrites pour la condition SANS 1 comme un processus d'allure bi-exponentielle avec une constante de temps caractéristique lente ( $t_1$ ) et rapide ( $t_2$ ) (figure 100.A et 100.C et gauche du tableau 20). Pour la condition SANS 2, la diminution de I(0) et l'hydrolyse de l'ATP peuvent être décrites comme un processus d'allure bi-exponentielle avec une constante de temps caractéristique lente ( $t_1$ ) et rapide ( $t_2$ ), tandis que la fluorescence peut être décrite comme un processus d'allure exponentielle avec une constante de temps caractéristique ( $t_1$ ) (figure 100.B ex 100.D et droite du tableau 20).

Dans les deux conditions SANS, les constantes de temps rapides  $t_2$  de la consommation d'ATP sont similaires aux constantes  $t_2$  respectives extraites de la fluorescence et du I(0) (ou  $t_1$  dans le cas de la fluorescence de la condition SANS 2). En prenant en compte les marges d'erreurs, les constantes de temps rapides caractéristiques  $t_2$  sont entre 1 et 4 minutes pour la condition SANS 1 et entre 4 et 8 minutes pour la condition SANS 2. Cela est en accord avec les allures des courbes de fluorescence décrites dans le chapitre 6.

Dans l'ensemble, il existe donc une corrélation entre les trois ensembles de données expérimentales, qui affichent presque toutes une allure bi-exponentielle et des constantes de temps rapides  $t_2$  similaires pour chaque condition SANS. Cela est en faveur d'une relation directe entre les paramètres fonctionnels et structuraux. En effet, la dénaturation de la GFPssrA, suivie par perte

de la fluorescence, est similaire à la dégradation de la d-GFPssrA native, suivie par la diminution du I(0), et à la consommation d'ATP.

Les études du chapitre 6 avaient démontré que l'hydrolyse de l'ATP modulait la prise en charge de la GFPssrA. Le suivi du I(0) a permis de compléter ces données. Au début de la réaction (5 premières minutes pour la condition SANS 1 et 10 premières minutes pour la condition SANS 2), l'ATP est rapidement hydrolysé par PAN et la dénaturation et dégradation des GFPssrA natives (I(0) et fluorescence) a lieu à un rythme similaire. Après le début de la réaction, la majeure partie de l'ATP est transformée en ADP et en  $P_i$ , et l'hydrolyse de l'ATP se poursuit à un rythme beaucoup plus lent. La raison la plus probable de ce ralentissement est que, d'une part la diminution de la concentration en ATP ne permet pas de maintenir une vitesse de réaction maximale constante et, d'autre part, l'augmentation de la concentration d'ADP inhibe de façon compétitive l'hydrolyse de l'ATP en se fixant aux sites de liaison des nucléotides de PAN (voir partie 6.3 du chapitre 6)<sup>396</sup>.

Ces observations sont cohérentes avec les comportements décrits dans le chapitre 6, et indiquent que :

- **Le dépliement et la dégradation de la d-GFPssrA par le protéasome h-PAN-h-20S sont directement modulés par l'hydrolyse de l'ATP.** Ce phénomène avait été démontré dans la partie 6.3 du chapitre 6 grâce aux mesures de l'hydrolyse de l'ATP comparées aux mesures de fluorescence. Le suivi de d'intensité diffusée à angle zéro I(0) a permis de compléter ces données. Le suivi de d'intensité diffusée à angle zéro I(0) a permis de compléter ces données en apportant une information directe sur la structure de la d-GFPssrA.
- **La majorité du dépliement et de la dégradation de la d-GFPssrA par la machinerie protéolytique a lieu dans les premières minutes de réaction** comme indiqué par la constante de temps caractéristique rapide  $t_2$ . De plus, comme décrit dans la partie 2 du chapitre 6, la vitesse de réaction est liée à la quantité de PAN en solution.

## 2.6. Comparaison entre les agrégats générés par PAN et les agrégats dus à la dénaturation thermique de la d-GFPssrA

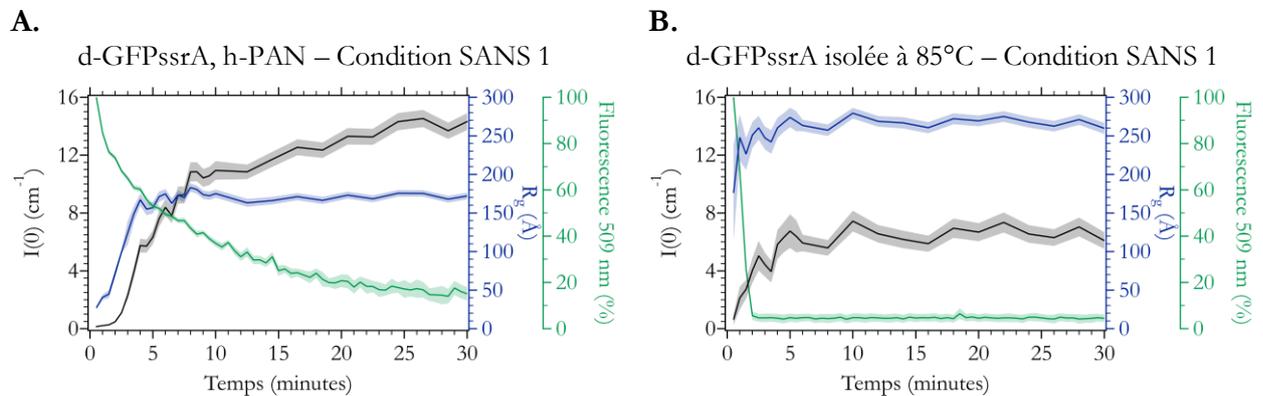
La d-GFPssrA en présence de h-PAN, formant des agrégats, a été comparée à de la d-GFPssrA agrégée suite à la dénaturation thermique de la protéine. Le but de cette comparaison était de déterminer si ces agrégats issus de deux processus différents ont une structure générale similaire, et si des informations peuvent être obtenues sur le mécanisme de PAN. Pour dénaturer la d-GFPssrA thermiquement, un échantillon de d-GFPssrA seule dans les conditions SANS 1 a été incubé 30 minutes à 85°C et mesuré par TR-SANS couplée à la spectroscopie de fluorescence de la même manière que les précédents échantillons. La figure 101 est une photographie de l'échantillon après sa mesure.

Après incubation à 85°C, l'échantillon arbore une apparence visiblement agrégée et la solution ne présente plus aucune couleur verte. Cet aspect est cependant différent de celui observé lorsque la d-GFPssrA est incubée avec h-PAN à 55°C (figure 96), qui présentait une apparence d'agrégats diffus avec une couleur verte résiduelle. **Cette différence d'aspect macroscopique indique que la réaction d'agrégation est différente dans les deux cas.**



**Figure 101 :** Photographie de l'échantillon de d-GFPssrA incubée à 85°C de la condition SANS 1 pendant 30 minutes de mesure par TR-SANS.

Les valeurs de l'intensité diffusée à l'angle zéro  $I(0)$  et du rayon de giration  $R_g$ , ont pu être extraits des mesures TR-SANS et complétés par les mesures de la fluorescence de la GFPssrA. Les résultats sont représentés dans la figure 102 de la même manière que précédemment.

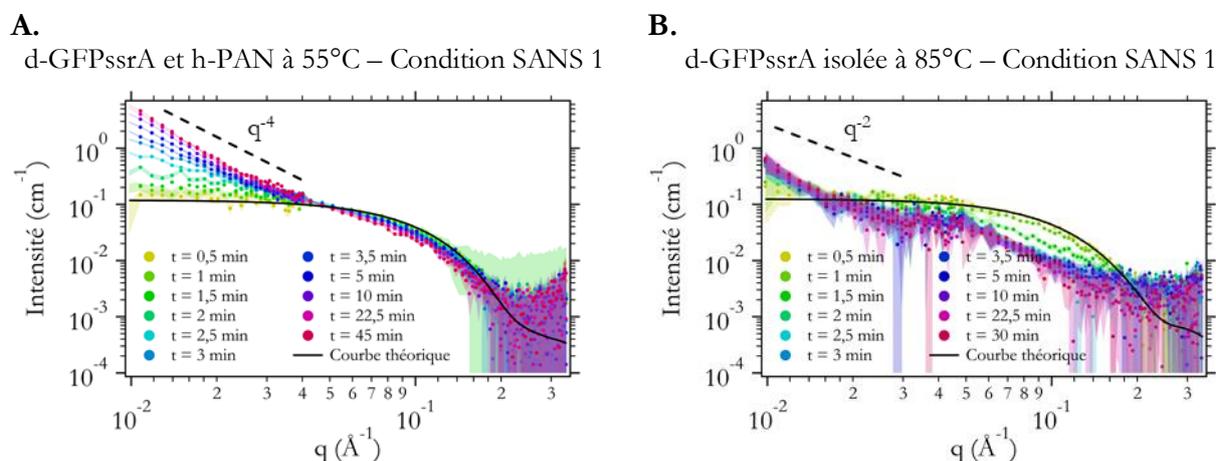


**Figure 102 :** Évolution au cours du temps de l'intensité diffusée à l'angle zéro  $I(0)$ , du rayon de giration  $R_g$ , et de la fluorescence à 509 nm de la d-GFPssrA format des agrégats dans la condition SANS 1. **A.** d-GFPssrA en présence de h-PAN à 55°C. **B.** d-GFPssrA incubée à 85°C. Pour les deux réactions, un point de mesure de fluorescence à 509 nm, de  $R_g$  et de  $I(0)$  a été enregistré toutes les 30 secondes. Après 10 minutes (A) ou 5 minutes de réaction (B), les données SANS ont été moyennées pour avoir un point de mesure de  $R_g$  et de  $I(0)$  toutes les 2 minutes. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur.

L'évolution des paramètres de la réaction d-GFPssrA en présence de h-PAN (figure 102.A) est déjà décrite dans la partie 2.3. L'approximation de Guinier pour la d-GFPssrA dénaturée thermiquement n'est plus valide dès le second point de mesure (puisque  $qR_g > 1,3$ ). Cela a également été observé pour la d-GFPssrA en présence de PAN. Les agrégats sont tellement grands qu'ils n'entrent plus dans la gamme de  $q$  mesurée et les valeurs de  $R_g$  et de  $I(0)$  obtenues sont fortement sous-estimées. Dans les deux cas, les valeurs apparentes de  $R_g$  et de  $I(0)$  augmentent rapidement et drastiquement au cours du temps. La perte de la fluorescence suit cependant une allure très différente pour les deux échantillons. Dans le cas de la d-GFPssrA dénaturée h-PAN (figure 102.A), la fluorescence diminue progressivement au cours du temps, indiquant une déstabilisation progressive de la d-GFPssrA. De plus, une faible partie de la population (~15%) reste fluorescente après 30 minutes. Lorsque la d-GFPssrA est thermiquement dénaturée (figure 102.B), la fluorescence disparaît complètement et quasi-instantanément, indiquant une dénaturation totale extrêmement rapide de la d-GFPssrA. **L'évolution de tous les paramètres moléculaires est donc compatible avec l'apparition rapide de grands agrégats de d-GFPssrA, mais qui apparaissent selon un processus différent.** Une analyse plus fine des

valeurs de  $R_g$  et de  $I(0)$  ne serait pas pertinente puisque nous n'avons pas accès aux dimensions réelles des deux types d'agrégats.

La figure 103 représente les courbes SANS en fonction du temps de la d-GFPssrA en présence de h-PAN et de la d-GFPssrA incubée à 85°C dans les conditions SANS 1.



**Figure 103 : Courbes expérimentales SANS en fonction du temps de la d-GFPssrA formant des agrégats dans la condition SANS 1. A.** d-GFPssrA en présence de h-PAN à 55°C. **B.** d-GFPssrA incubée à 85°C. Pour chaque figure, une courbe SANS théorique calculée par le logiciel CRYSON à partir de la structure cristalline d'une GFP (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) est représentée sous forme de courbe noire superposée à la première courbe expérimentale. Les barres d'erreur sont représentées pour chaque courbe expérimentale sous forme d'aire nuancée de la même couleur.

Les échantillons présentent une courbe expérimentale initiale ( $t = 0,5$  minutes) superposable à la courbe théorique de la GFP native pour tous les vecteurs de diffusion  $q$  représentés. Les deux présentent ensuite des courbes indiquant la formation progressive d'agrégats au long de la réaction<sup>40,333,334</sup>. L'intensité de diffusion augmente aux valeurs  $q$  faibles, c'est-à-dire  $q < 0,03 \text{ \AA}^{-1}$  pour les agrégats générés par PAN et  $q < 0,02 \text{ \AA}^{-1}$  pour les agrégats thermiques, et diminue aux moyens angles, soit  $0,04 \text{ \AA}^{-1} < q < 0,1 \text{ \AA}^{-1}$  pour les agrégats produits par PAN et  $0,02 \text{ \AA}^{-1} < q < 0,2 \text{ \AA}^{-1}$  pour les agrégats thermiques. L'intensité de diffusion reste toutefois stable aux angles plus élevés, c'est-à-dire  $q > 0,15 \text{ \AA}^{-1}$  pour les agrégats générés par PAN et  $q > 0,2 \text{ \AA}^{-1}$  pour les agrégats thermiques. Bien que les deux échantillons montrent un phénomène d'agrégation, les courbes évoluent de façon différente au cours du temps dans les deux cas. La d-GFPssrA avec h-PAN montre une forte augmentation de l'intensité de diffusion aux petits angles et une légère diminution aux moyens angles, alors que la d-GFPssrA incubée à 85°C présente une faible augmentation aux petits angles et forte diminution aux moyens angles. Cela indique que les agrégats formés et les structures qui les composent ont une organisation différente dans les deux cas.

Les pentes des courbes expérimentales à faible  $q$  et à moyen/haut  $q$  fournissent des informations sur la forme des agrégats. Pour les agrégats générés par PAN (figure 103.A), la faible diminution aux moyens angles ( $0,04 \text{ \AA}^{-1} < q < 0,1 \text{ \AA}^{-1}$ ), couplée à la diminution progressive de la fluorescence, montre que la d-GFPssrA conserve une forme proche de la forme native mais qu'elle s'agrège. La pente à faible  $q$  ( $q < 0,03 \text{ \AA}^{-1}$ ) est de  $q^4$ , indiquant que les agrégats sont compacts et homogènes<sup>397</sup>. En outre, les d-GFPssrA constituant ces agrégats semblent également garder une forme proche de leur forme native compacte, engendrant ainsi des agrégats compacts. Lors de l'agrégation thermique de la d-GFPssrA (figure 103.B), la forte diminution aux moyens angles ( $0,02 \text{ \AA}^{-1} < q < 0,2 \text{ \AA}^{-1}$ ) est révélatrice d'un changement de forme de la d-GFPssrA. Cette information, couplée à

celle de la fluorescence, indique que les d-GFPssrA ont été complètement dénaturées et perdent leur forme native. La pente à faible  $q$  ( $q < 0,02 \text{ \AA}^{-1}$ ) est de  $q^2$ . Une pente de  $q^2$  est généralement observée pour les polymères lorsqu'ils sont fortement gonflés par le solvant<sup>397</sup>. Pour une protéine, cela correspondrait alors à une dénaturation qui aurait pour effet de transformer la forme de la protéine d'une forme globulaire en une forme « gonflée ». Les agrégats sont eux-mêmes formés par des d-GFPssrA complètement dénaturées.

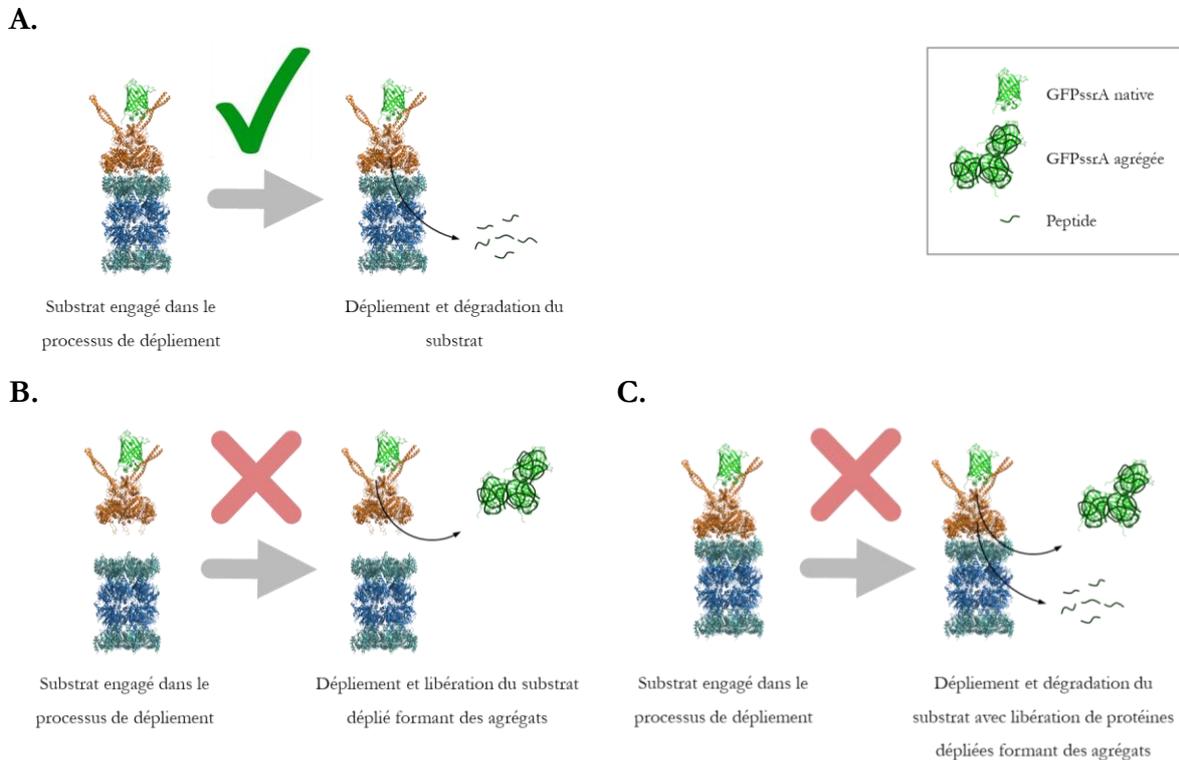
On peut donc conclure que **l'activité de dépliement de h-PAN génère des agrégats différents de ceux provoqués par une dénaturation thermique**. Les agrégats thermiques sont formés de d-GFPssrA complètement dénaturées tandis que **les agrégats générés par h-PAN sont constitués de d-GFPssrA conservant une forme compacte, avec une forme globale compacte et homogène**. Cela confirme ce qui a été décrit dans l'étude ayant précédé ce projet<sup>40</sup>. La configuration collimateur/détecteur (8,0 m/8,0 m) supplémentaire utilisée a permis de mieux apprécier la taille des agrégats, qui sont tellement grands qu'ils n'entrent toujours pas dans la gamme de  $q$  mesurée.

### 3. Discussion et conclusion générale du chapitre

Les données TR-SANS ont permis de suivre l'évolution de la protéine substrat d-GFPssrA en présence de la machinerie protéolytique hydrogénée avec une résolution temporelle de 30 secondes. Les mesures de fluorescence et de l'intensité de diffusion à angle nul  $I(0)$  montrent que, après une diminution rapide de la population de la d-GFPssrA native, le processus de dégradation se poursuit à un second rythme, plus lent. **L'hydrolyse de l'ATP est fortement corrélée à la dégradation des d-GFPssrA et module le processus de dépliement et de dégradation des substrats protéiques**. Ce phénomène a déjà été décrit lors de certaines des études biochimiques sur le protéasome PAN-20S<sup>250,277,373,393</sup> et suggéré par les récentes études par cryo-microscopie électronique du protéasome PAN-20S en l'absence de substrats<sup>286</sup> et du protéasome 26S en présence de substrats<sup>305,306</sup>. Cependant, il s'agit ici de la **première étude directe en solution suivant un substrat protéique dégradé par le protéasome PAN-20S, ceci en relation avec des paramètres biochimiques (fluorescence à 509 nm et hydrolyse de l'ATP)**.

Alors que les molécules de d-GFPssrA s'agrègent très rapidement suite à son dépliement par h-PAN, la présence de la machinerie protéolytique complète hydrogénée génère des produits de faible masse moléculaire sans aucune l'agrégation de d-GFPssrA. Cela est confirmé par les mesures en spectrométrie de masse qui ont détecté des peptides d'environ 10 acides aminés en fin de réaction. Même si l'association de la particule régulatrice PAN et de la particule 20S pour former le protéasome PAN-20S est connue pour être transitoire en présence d'ATP en solution *in vitro* (voir partie 5.1 du chapitre 1 et partie 3.6 du chapitre 5)<sup>373</sup>, un rapport stœchiométrique d'une particule régulatrice h-PAN pour une particule catalytique h-20S (condition SANS 1) est suffisant pour empêcher complètement l'apparition d'agrégats de d-GFPssrA. En effet, l'intensité de diffusion à angle nul  $I(0)$  étant proportionnelle à la masse moléculaire d'une espèce macromoléculaire donnée, elle est particulièrement sensible à la présence d'oligomères et d'agrégats, comme c'est le cas quand la d-GFPssrA est en présence de h-PAN seul. Une fraction, même faible, de produits libérés sous une forme d'état intermédiaire déstabilisé aurait été facilement détectable s'ils avaient formé des oligomères ou agrégats. La comparaison des données SANS en présence de h-PAN avec celles de h-PAN et particule h-20S montre donc que **les processus de dépliement, de transfert et de dégradation sont étroitement couplés lorsque l'étape initiale de dépliement est engagée**.

L'absence totale d'agrégats exclut une libération même minimale de substrats protéiques dépliés pendant le processus de dégradation, et suggère plutôt que **le substrat protéique est pris en charge par PAN lorsque le protéasome PAN-20S est déjà assemblé**. Cela est en accord avec les récentes structures par cryo-microscopie électronique de substrats engagés dans le protéasome 26S eucaryote<sup>305,306</sup>. Le modèle proposé est représenté dans la figure 104.



**Figure 104 : Modèle de dépliement et dégradation des substrats protéiques par le protéasome PAN-20S une fois le substrat protéique engagé dans PAN.** Schéma représentant les possibilités de mode d'action de la machinerie protéolytique. **A.** Le protéasome PAN-20S est déjà assemblé lorsque les substrats sont engagés dans le processus de dépliement. Les substrats protéiques seraient dépliés et dégradés sans libération de formes dénaturées formant des agrégats. **B.** Le protéasome PAN-20S n'est pas encore assemblé lors de l'engagement des substrats protéiques dans le processus de dépliement. Cela aurait pour conséquence la libération de substrats protéiques dépliés formant des agrégats. **C.** Le protéasome PAN-20S est déjà assemblé lorsque les substrats sont engagés dans le processus de dépliement et de dégradation. L'assemblage du protéasome complet pourrait permettre une libération de protéines dénaturées formant des agrégats, soit à cause d'une dissociation du protéasome PAN-20S, soit même lorsque le protéasome PAN-20S est assemblé. Les résultats obtenus ont permis de valider l'hypothèse A. Modèles générés à partir des PDB 6HEA & PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale.

Deux conditions de composition d'échantillon ont été testées (conditions SANS 1 et 2) et représentent deux modalités de prise en charge des substrats protéiques en fonction de la quantité de protéines à dégrader. **On considère ces situations comme ayant lieu en quantité « limitée » ou « abondante » en substrats protéique par rapport à PAN**, même s'il n'existe aucune information *in vivo* sur ce qui représenterait une quantité « abondante » en substrats protéiques pour PAN. Dans le premier cas étudié, avec 3,5 GFPssrA pour une particule régulatrice PAN et une particule catalytique 20S (condition SANS 1), **les substrats protéiques sont dépliés et dégradés de façon rapide et coordonnée**. Si les protéines sont relâchées sous forme déstabilisée, elles ne sont pas détectables par la technique. Dans le second cas étudié, avec 35 GFPssrA pour une particule régulatrice PAN et quatre particules catalytique 20S (condition SANS 2), **les protéines**

**substrat sont prise en charge par PAN, qui en relâche une partie sous forme dénaturée.** Seulement 60% des protéines dénaturées sont dégradées par la particule catalytique 20S.

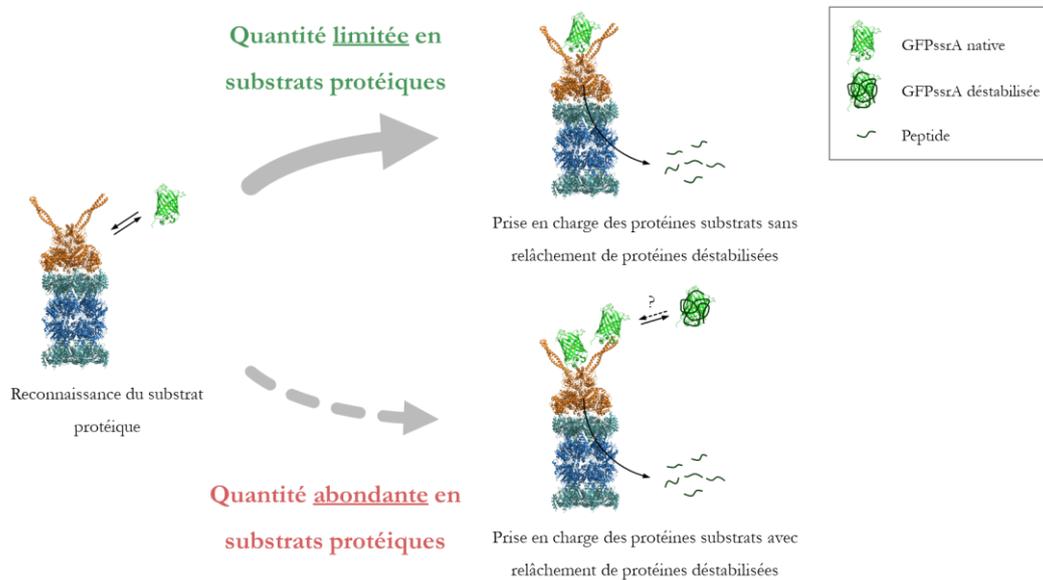
Les GFPssrA relâchées en solution ont été déstabilisées par PAN puisqu'elles ne sont plus fluorescentes, leur structure a donc été suffisamment fragilisée pour déstabiliser leur chromophore (voir partie 4.1.3 du chapitre 5)<sup>387</sup>. Cependant, elles ne sont pas distinguables de la GFPssrA native d'un point de vue structural (à basse résolution) et ne présentent pas de caractéristiques structurales altérées qui augmenteraient leur tendance à l'oligomérisation et/ou à l'agrégation. Ces GFPssrA déstabilisées et relâchées sont donc différentes des GFPssrA dépliées et forment des agrégats en présence de PAN seul. On peut donc supposer que les GFPssrA déstabilisées et relâchées n'ont pas été complètement dépliées par PAN et sont impliquées dans un processus différent.

Différentes hypothèses peuvent être avancées concernant l'état de ces GFPssrA relâchées. Ces protéines pourraient être légèrement déstabilisées suite à des tentatives infructueuses de dépliement de PAN. Cela est en accord avec le modèle probabiliste décrit pour ClpXP où chaque tentative de dépliement d'une protéine substrat a une probabilité identique de la déplier<sup>312</sup>. Ces tentatives de dépliement pourraient alors résulter en une légère déstabilisation de la GFPssrA. Une autre hypothèse est que les GFPssrA seraient partiellement prises en charge par le protéasome PAN-20S, qui les relâcherait sous forme partiellement dépliées et éventuellement partiellement dégradées. Cela a été décrit pour la prise en charge de protéines stables d'un point de vue mécanique par le système ClpXP<sup>244,289,316,324</sup>. Il a également été suggéré que la libération de substrats protéiques stables au profit de la prise en charge de protéines substrat moins stables permettrait à ClpXP de dégrader préférentiellement les substrats protéiques les moins stables dans une population<sup>316</sup>. Un mécanisme similaire pourrait avoir lieu pour le protéasome PAN-20S en présence d'une grande quantité de protéine à dégrader, la machinerie protéolytique « cherchant » les protéines les moins stables dans la population. N'ayant qu'un seul type de substrat lors de ces expériences en solution, une expérience complémentaire avec plusieurs types de substrats pourrait apporter des informations supplémentaires. Enfin, il est important de noter que ces expériences ont lieu en concentrations limitées d'ATP, qui est consommé par PAN au cours de la réaction. Il a été décrit pour ClpXP qu'une concentration minimum d'ATP était nécessaire pour déplier et dégrader certaines protéines, notamment la GFPssrA<sup>289,321,322</sup>. De plus, l'ADP accumulé au cours du temps entre en compétition pour les sites de liaison à l'ATP, empêchant ainsi le dépliement et la dégradation de la GFPssrA (voir partie 6.3 du chapitre 6). Il est possible que les concentrations en ATP et ADP influencent le phénomène de relargage des protéines dénaturées. Des mesures complémentaires en présence d'un système de régénération de l'ATP pourraient alors permettre de déterminer l'implication de la concentration d'ATP et d'ADP dans ce phénomène.

**Dans une situation physiologique d'excès d'un substrat protéique, la condition abondante pourrait représenter un mode de fonctionnement du protéasome PAN-20S.** Une telle situation pourrait avoir naturellement lieu dans les cellules en cas d'accumulation de protéines mal repliées de façon consécutive à des stress intenses, des défauts de la traduction, des mutations, des maladies dégénératives ou encore le vieillissement<sup>168,173</sup>. En effet, des agrégats intracellulaires ubiquitinylés et physiquement associés au protéasome 26S ont été retrouvés lors de diverses maladies neurodégénératives, suggérant fortement un échec du protéasome de la cellule<sup>168,173</sup>. Notons tout de même que la présence des sous-unités du couvercle (voir partie 3.3 du chapitre 1) de la particule régulatrice 19S du protéasome 26S rendent certains aspects de son fonctionnement

différent, certaines sous-unité couvercle ayant pour rôle de stabiliser les protéines substrat au cours de leur prise en charge par le protéasome 26S<sup>193</sup>.

La figure 105 représente le modèle proposé en fonction de la quantité de substrats à dégrader par le protéasome PAN-20S.



**Figure 105 : Modèle de prise en charge des substrats protéiques par le protéasome PAN-20S.** Schéma représentant les deux possibilités de processus de reconnaissance et de prise en charge des substrats par la machinerie protéolytique PAN-20S. En quantité limitée de protéines substrat (flèche en ligne continue), la GFPssrA reconnue par PAN est dépliée et dégradée en une seule étape selon un processus linéaire et concerté entre PAN et la particule 20S. En quantité abondante de protéines substrat (flèche en pointillé), une fois reconnues par PAN, certaines GFPssrA sont libérées sous forme déstabilisées. Modèle basé sur les PDB 6HEA & PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale.

Une publication présentant les résultats de ce chapitre est en phase finale de rédaction.

Les constantes de temps rapides  $t_2$  ont été calculées à partir des données de  $I(0)$ , fluorescence à 509 nm et d'hydrolyse de l'ATP des échantillons de d-GFPssrA en présence de la machinerie protéolytique complète (partie 2.5). Elles suggèrent que le mécanisme du protéasome h-PAN-h-20S est trop rapide pour détecter la d-GFPssrA sous forme partiellement dépliée au cours de son transfert de h-PAN vers la particule h-20S. **L'état intermédiaire de dépliement du substrat protéique doit être très transitoire et faiblement peuplé par rapport aux formes natives et aux produits de dégradation tout au long du processus de réaction.** Il est possible qu'il soit également difficilement détectable par la technique, qui mesure une moyenne des espèces moléculaires en solution. En effet, un échantillon contenant la d-GFPssrA et le protéasome h-PAN-h-20S est complexe de par la polydispersité des espèces de d-GFPssrA pouvant être présentes en solution (formes natives, formes dépliées, peptides).

La seconde partie du projet de thèse avait pour but de bloquer la d-GFPssrA dans un état de dépliement afin que la majorité de la population du substrat soit bloquée dans un état intermédiaire déplié au même moment. La stratégie développée et les mesures TR-SANS associées sont le sujet du chapitre suivant.

## Chapitre 8 :

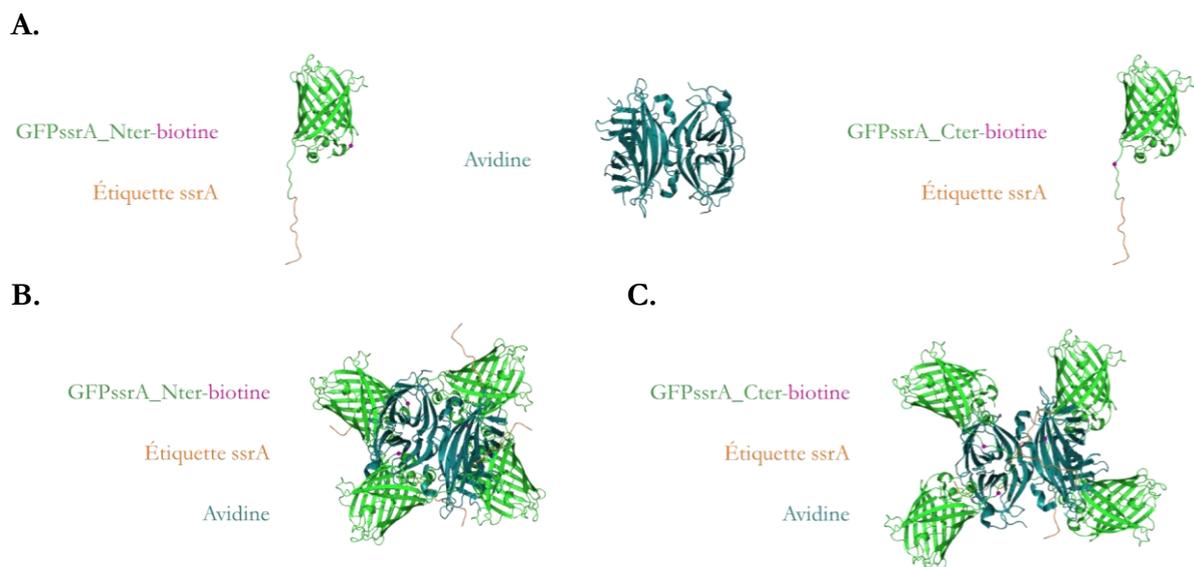
# Mesures biochimiques et TR-SANS sur le système GFPssrA-streptavidine, protéasome PAN-20S

Cette partie du projet de thèse avait pour objectif de développer un système permettant d'accéder à des informations sur le(s) état(s) intermédiaire(s) de dépliement de la protéine substrat GFPssrA. La stratégie envisagée était de bloquer le processus de dépliement par PAN en greffant un module streptavidine à l'extrémité N-terminale ou C-terminale de la GFPssrA. La première partie du chapitre décrit les expériences biochimiques effectuées afin de caractériser l'activité du protéasome PAN-20S avec les constructions GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine dans les conditions d'études nécessaires aux mesures TR-SANS. Les mesures TR-SANS réalisées sur ces mêmes systèmes sont décrites dans la seconde partie de ce chapitre.

### 1. Contexte de l'étude

#### 1.1. Stratégie employée dans la littérature

Les travaux présentés dans ce chapitre sont basés sur un article publié en 2001 par Navon & Goldberg<sup>317</sup>. L'hypothèse de cette étude était que si le processus de dépliement d'un substrat était une conséquence de son transfert au sein de PAN, alors la fixation d'une molécule volumineuse, l'avidine, à la GFPssrA devrait bloquer son transfert et empêcher son dépliement. Pour tester cette hypothèse, les auteurs ont utilisé une stratégie consistant à lier une molécule de biotine (~500 Da) à proximité de l'extrémité N-terminale ou C-terminale de la GFPssrA (figure 106).

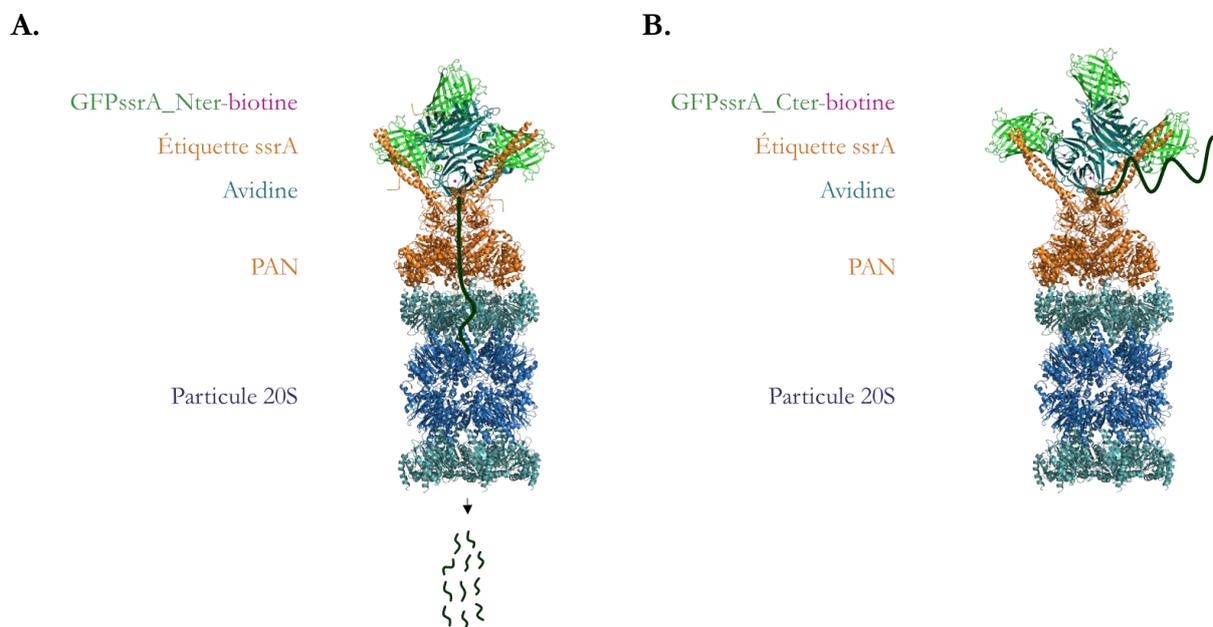


**Figure 106 : Représentation des différents composants de la GFPssrA\_Nter-avidine et de la GFPssrA\_Cter-avidine. A.** Représentation de la GFPssrA (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) avec une biotine liée proche de son extrémité N-terminale ou C-terminale et structure de l'avidine (PDB 3RY1). Une molécule d'avidine peut fixer quatre molécules de biotine, et donc quatre GFPssrA. **B & C.** Représentation artistique

d'une molécule d'avidine couplée à quatre GFPssrA par l'intermédiaire des molécules de biotine liées proche de leur extrémité N-terminale (B) ou C-terminale (C).

Pour cela, la GFPssrA a été mutée par substitution pour introduire une cystéine (capable de lier la biotine) à l'une ou l'autre de ses deux extrémités. Les molécules de biotine couplées à la GFPssrA ont alors pu reconnaître l'avidine (~50 kDa), un tétramère possédant quatre sites de liaison.

Navon & Goldberg ont montré que la GFPssrA était dépliée et dégradée par le protéasome PAN-20S quand une biotine ou une avidine (par l'intermédiaire d'une biotine) étaient fixées près de son extrémité N-terminale. La fixation d'une biotine près de l'extrémité C-terminale de la GFPssrA entraînait son dépliement et sa dégradation de façon similaire. Par contre, la fixation d'une avidine près de l'extrémité C-terminale de la GFPssrA a empêché sa dégradation, bien que la particule soit dépliée. Il a été alors déduit que l'avidine bloquait le passage de la GFPssrA à travers PAN, et que le substrat restait lié à PAN pendant environ une heure. La figure 107 illustre un modèle de la prise en charge par la machinerie protéolytique lors de la fixation d'une avidine proche de l'extrémité N-terminale ou C-terminale de la GFPssrA selon les observations décrites dans cette étude.



**Figure 107 :** Schémas représentant la prise en charge de la GFPssrA avec une avidine liée à proximité de son extrémité N-terminale (A) et de la GFPssrA avec une avidine liée proche de son extrémité C-terminale (B) par la machinerie protéolytique en présence d'ATP et de MgCl<sub>2</sub> selon le modèle décrit<sup>317</sup>. Selon ce modèle, la GFPssrA avec une avidine liée proche de son extrémité N-terminale serait dépliée et dégradée tandis que la GFPssrA avec une avidine liée proche de son extrémité C-terminale serait dépliée mais pas dégradée et ce substrat resterait fixé à PAN (bloqué) pendant environ une heure.

Ces travaux ont conduit à un modèle selon lequel le dépliement mécanique des protéines substrat entraînerait leur transfert à travers l'anneau AAA+ PAN dans le sens N-terminal vers C-terminal. Ce processus déstabiliserait les domaines globulaires en amont, ce qui mènerait au dépliement global des substrats à la surface de PAN. Les protéines seraient ainsi dépliées à la surface de PAN avant leur transfert, et l'utilisation de GFPssrA liée à une avidine proche de son extrémité C-terminale permettrait de bloquer la GFPssrA au cours de son transfert dans un état déplié.

Cependant, ces travaux reposaient en majeure partie sur des études biochimiques par spectrométrie de fluorescence et immuno-détection à l'aide d'anticorps anti-GFP. De plus, cette étude n'a pas été

poursuivie par des études structurales et fonctionnelles plus approfondies. Des études plus récentes, décrites dans l'introduction<sup>286,306,310,398</sup> (voir parties 5.2.3 et 5.2.4 du chapitre 1), considèrent également le dépliement des substrats protéiques comme une conséquence du processus de transfert au sein de PAN. Néanmoins, ces travaux ne supportent pas l'hypothèse selon laquelle les substrats sont dépliés à la surface de PAN puis transférés au sein de la particule 20S. Ils suggèrent plutôt un transfert des substrats par les boucles Ar- $\Phi$  (voir parties 5.2.3 du chapitre 1) directement associé à leur dépliement, sans déstabilisation des domaines globulaires en amont.

## 1.2. Stratégie employée dans cette étude

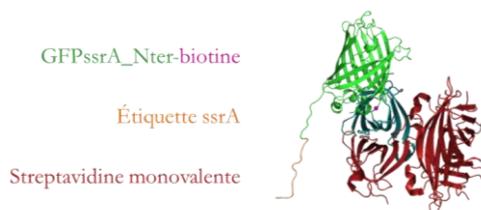
Au début de ce projet de thèse, l'utilisation de GFPssrA liée à une molécule d'avidine à proximité de son extrémité C-terminale nous a paru intéressante pour étudier un substrat protéique déplié et bloqué au début du processus de transfert. La formation du complexe PAN-GFPssrA-avidine stable pourrait refléter une étape intermédiaire dans le processus normal de transfert du substrat. Des études SANS sur un tel substrat permettrait alors d'obtenir des informations structurales sur le(s) état(s) intermédiaire(s) de dépliement de la GFPssrA.

Pour réaliser ce projet, il fallait simplifier le système utilisé. Dans le cas des mesures SANS, il serait problématique d'utiliser une molécule d'avidine hydrogénée liée à quatre molécules de GFPssrA deutérées. En effet, les quatre molécules de GFPssrA liées seraient très proches les unes des autres et pourraient générer une interférence inter-particulaire. Il y aurait alors une perte d'information sur le facteur de forme d'une seule molécule de GFPssrA.

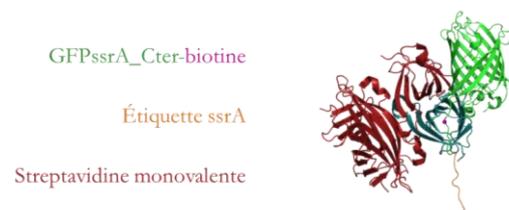
**A.**



**B.**



**C.**

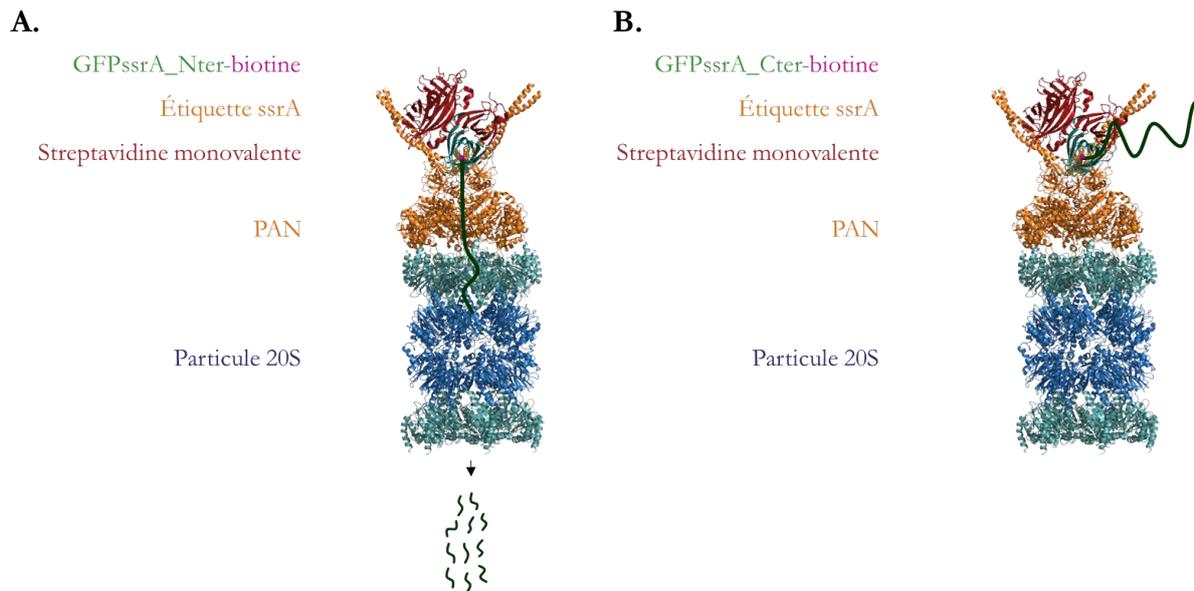


**Figure 108 : Représentation des différents composants de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine. A.** Représentation de la GFPssrA\_Nter-biotine, de la GFPssrA\_Cter-biotine (PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale) et de la streptavidine monovalente (sous-unités inactives en rouge et sous-unité active en vert) (PDB 3RY1). Une molécule de streptavidine monovalente peut fixer une molécule de biotine, et donc une GFPssrA. **B & C.** Représentation artistique d'une molécule de streptavidine monovalente couplée à une GFPssrA par l'intermédiaire d'une molécule de biotine liée à proximité de leur extrémité N-terminale (B) ou C-terminale (C).

Notre choix s'est donc porté sur la streptavidine monovalente. Cette particule permet de lier une seule biotine au lieu de quatre (voir partie 5.1 du chapitre 5) ce qui simplifie l'interprétation des données SANS. Deux protéines substrat ont été préparées à partir de deux GFPssrA mutantes, qui

ont été purifiées et liées à la biotine. La première, la GFPssrA\_Nter-biotine, permet de fixer une molécule de streptavidine à proximité de l'extrémité N-terminale de la GFPssrA, tandis que la seconde, la GFPssrA\_Cter-biotine, permet de lier une molécule de streptavidine proche de son extrémité C-terminale. Les positions des mutations (G10C pour la GFPssrA\_Nter, et H231C pour la GFPssrA\_Cter) ont été choisies pour imiter la position des mutations décrites dans l'article de référence<sup>317</sup>. La figure 108 représente les différentes protéines générées pour ce projet.

La figure 109 illustre les processus supposés avoir lieu une fois ces substrats au contact du protéasome PAN-20S en présence d'ATP et de MgCl<sub>2</sub> d'après l'article de référence.



**Figure 109 :** Schémas représentant la prise en charge de la GFPssrA\_Nter-streptavidine (A) et de la GFPssrA\_Cter-streptavidine (B) par la machinerie protéolytique en présence d'ATP et de MgCl<sub>2</sub> selon le modèle proposé<sup>317</sup>. Selon ce modèle, la GFPssrA\_Nter liée à la streptavidine monovalente serait dépliée et dégradée tandis que la GFPssrA\_Cter liée à la streptavidine monovalente serait dépliée mais pas dégradée. La GFPssrA\_Cter-streptavidine resterait fixée à PAN (bloquée) pendant environ une heure.

Notre objectif était de réaliser des mesures SANS de la même manière que celle décrite dans le chapitre 7, cette fois-ci avec la GFPssrA deutérée et la biotine, streptavidine monovalente et machinerie protéolytique hydrogénées. Cela pourrait permettre d'accéder à des informations structurales sur le substrat dans un état intermédiaire de dépliement.

### 1.3. Conditions expérimentales choisies

Comme mentionné dans le chapitre 6, les conditions de mesures SANS en temps réel nécessitaient d'avoir une réaction efficace, mais suffisamment lente pour suivre la réaction avec un point de mesure toutes les 30 secondes. De plus, une concentration relativement élevée de GFPssrA deutérée était nécessaire pour avoir un signal de diffusion de neutrons analysable. Néanmoins, d'autres contraintes expérimentales se sont ajoutées à ces critères. Souhaitant bloquer le substrat au cours de son dépliement par PAN pour visualiser un état intermédiaire déplié, l'idéal était d'avoir un excès de PAN par rapport à la GFPssrA afin que toutes les molécules de GFPssrA soient dans un état intermédiaire déplié. Il fallait donc avoir dans la réaction au moins autant de molécules de PAN que de molécules de GFPssrA. Cependant, il a été démontré dans la partie 2 du chapitre 6 qu'une concentration élevée de PAN entraînait une consommation d'ATP trop rapide pour

permettre la prise en charge de toutes les molécules de GFPssrA disponibles. Une concentration de PAN inférieure à 30  $\mu\text{M}$  a donc été utilisée. Le reste du volume disponible dans la réaction devait alors être occupé par la particule 20S, avec au moins une particule 20S pour une molécule de PAN. Compte-tenu de ces critères, quatre conditions de mesures SANS à 100 mM ATP & 200 mM  $\text{MgCl}_2$  ont été sélectionnées :

- **Condition SANS 3 – 25  $\mu\text{M}$  GFPssrA\_Nter-biotine, 25  $\mu\text{M}$  PAN et 25  $\mu\text{M}$  particule 20S**
- **Condition SANS 4 – 25  $\mu\text{M}$  GFPssrA\_Cter-biotine, 25  $\mu\text{M}$  PAN et 25  $\mu\text{M}$  particule 20S**

Ces deux conditions correspondent à des conditions « contrôles » lorsque les GFPssrA mutantes ne sont pas liés à la streptavidine monovalente. La concentration minimum de d-GFPssrA permettant de visualiser un signal exploitable en SANS avec 30 secondes de temps d'exposition aux neutrons est d'environ 70  $\mu\text{M}$  ( $\sim 2 \text{ mg/mL}$ ). Avec une concentration de 25  $\mu\text{M}$ , il faudra prolonger le temps d'exposition en conséquence.

- **Condition SANS 5 – 25  $\mu\text{M}$  GFPssrA\_Nter-streptavidine, 25  $\mu\text{M}$  PAN et 25  $\mu\text{M}$  particule 20S**
- **Condition SANS 6 – 25  $\mu\text{M}$  GFPssrA\_Cter-streptavidine, 25  $\mu\text{M}$  PAN et 25  $\mu\text{M}$  particule 20S**

Ces deux conditions correspondent aux conditions dans lesquelles la GFPssrA\_Cter-streptavidine serait bloquée dans un état intermédiaire au cours de son dépliement et où la GFPssrA\_Nter-streptavidine serait prise en charge normalement par la machinerie protéolytique. Différentes mesures fonctionnelles ont été réalisées afin de caractériser les conditions d'études sélectionnées.

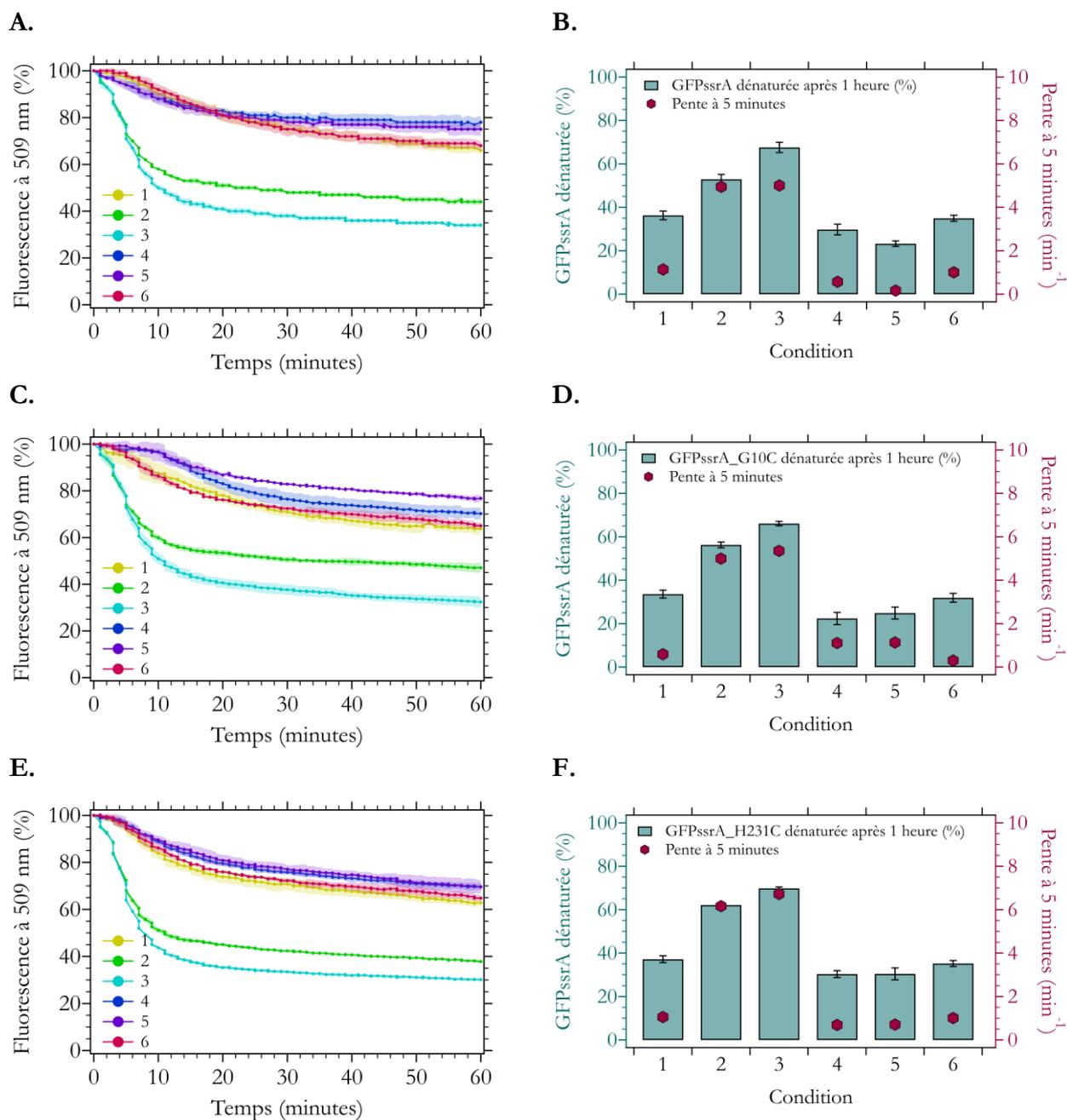
## 2. Activité de dénaturation suivie par spectroscopie de fluorescence

### 2.1. GFPssrA libre, GFPssrA\_Nter-biotine et GFPssrA\_Cter-biotine

Une étude fonctionnelle a été réalisée dans les conditions SANS 3 et 4 (voir partie précédente) et dans une condition similaire avec de la GFPssrA libre. Le but était de déterminer si les GFPssrA-biotine étaient prise en charge de la même manière que la GFPssrA libre. Les conditions testées sont :

- (1) GFPssrA en présence de 100 mM ATP
- (2) GFPssrA et PAN en présence de 100 mM ATP
- (3) GFPssrA, PAN et particule 20S en présence de 100 mM ATP
- (4) GFPssrA et particule 20S en présence de 100 mM ATP
- (5) GFPssrA et PAN
- (6) GFPssrA, PAN et particule 20S

On considère pour le reste de ce chapitre qu'en présence de 100 mM ATP, la concentration de  $\text{MgCl}_2$  était de 200 mM. Les données représentées dans les figures 110.A, 5.C et 5.E correspondent aux courbes d'évolution de la fluorescence en fonction du temps dans les différentes conditions testées pour respectivement la GFPssrA, GFPssrA\_Nter-biotine et la GFPssrA\_Cter-biotine. Les figures 110.B, 110.D et 110.F correspondent à l'analyse de ces courbes présentées de la même manière que dans le chapitre 6.



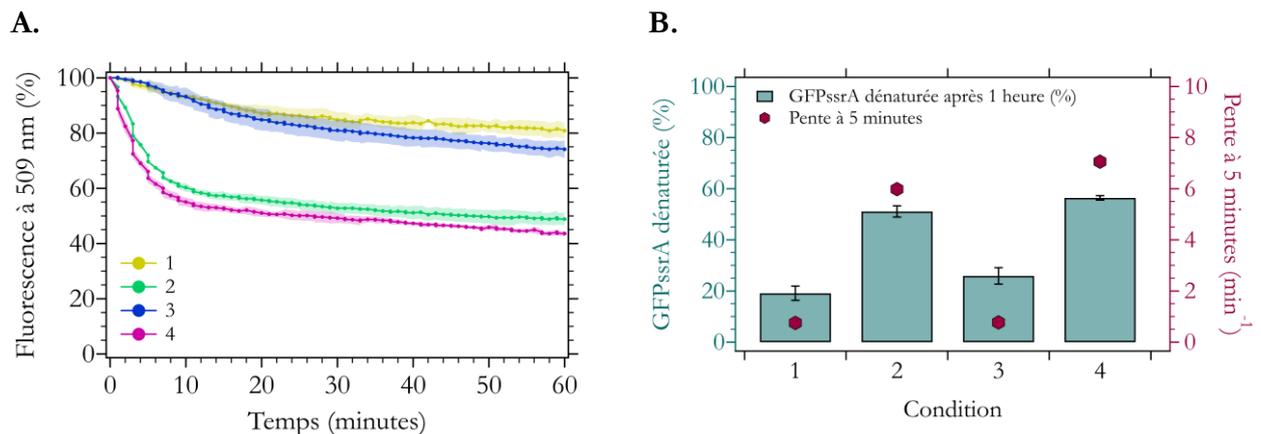
**Figure 110 : Mesure de l'activité du système à 55°C avec de la GFPssrA libre, GFPssrA\_Nter-biotine (formée à partir de la GFPssrA\_G10C) et GFPssrA\_Cter-biotine (formée à partir de la GFPssrA\_H231C) (conditions SANS 3 et 4).** Les figures A & B, C & D et E & F représentent respectivement les conditions avec de la GFPssrA libre, la GFPssrA\_Nter-biotine et la GFPssrA\_Cter-biotine. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6, ce qui correspond aux réactions suivantes : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA, PAN et particule 20S, ou (4) GFPssrA et particule 20S toutes quatre en présence de 100 mM ATP, (5) GFPssrA et PAN et (6) GFPssrA, PAN et particule 20S toutes deux sans ATP. **A, C et E.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur. **B, D et F.** Histogrammes avec le pourcentage de GFPssrA dénaturation après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

L'analyse est similaire pour les deux conditions SANS et la condition avec de la GFPssrA libre, et est semblable aux résultats présentés dans la partie 1.2 du chapitre 6. Dans les conditions où la

GFPssrA (libre ou liée à la biotine) est seule (1) ou en présence de particule 20S avec 100 mM ATP (4), et dans toutes les conditions sans ATP (5 et 6), le pourcentage de GFPssrA dénaturée est de 20 à 35% tandis que la pente à 5 minutes est de 0,5 à 2 min<sup>-1</sup>. Ces valeurs sont la conséquence de la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C<sup>383</sup>. Il s'agit des contrôles négatifs. En présence de PAN à 100 mM ATP (2), le pourcentage de GFPssrA dénaturée après 1 heure de réaction est de 55 à 60% et la pente à 5 minutes oscille entre 5 et 6 min<sup>-1</sup>. Lorsque la machinerie protéolytique est complète à 100 mM ATP (3), le système est plus efficace avec un pourcentage de GFPssrA dénaturée en fin de réaction entre 65 à 70% et une pente à 5 minutes de 5 à 7 min<sup>-1</sup>. En conclusion, la GFPssrA\_Nter-biotine et la GFPssrA\_Cter-biotine sont prises en charge de façon similaire à la GFPssrA libre par le protéasome PAN-20S. **La biotine seule n'empêche donc pas le dépliement et la dégradation des substrats protéiques.**

## 2.2. GFPssrA libre et streptavidine monovalente

Une expérience contrôle a été réalisée afin de déterminer si la streptavidine monovalente libre en solution avait un effet sur l'activité de dégradation de la GFPssrA par la machinerie protéolytique. Pour cela, 25 μM de streptavidine monovalente a été ajouté dans un mélange réactionnel contenant de la GFPssrA libre (25 μM) et PAN (25 μM) en présence de 100 mM ATP. Les données représentées dans les figures 111.A et 111.B de la même manière que précédemment.



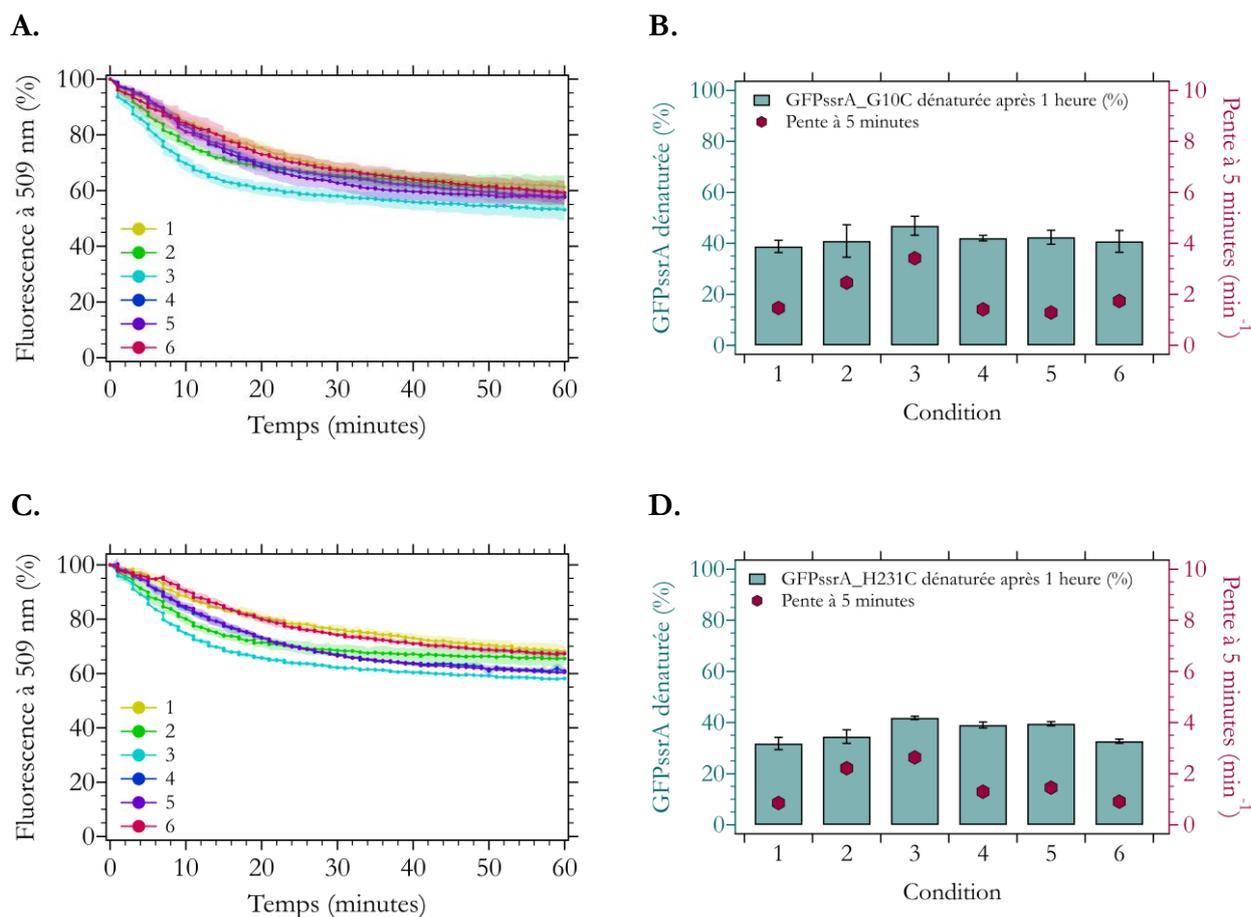
**Figure 111 : Mesure de l'activité du système à 55°C avec de la GFPssrA libre avec ou sans streptavidine monovalente** en présence de 100 mM ATP. Pour toutes les figures, les conditions testées sont numérotées de 1 à 4, ce qui correspond aux réactions suivantes : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA et streptavidine monovalente, ou (4) GFPssrA, streptavidine monovalente et PAN. **A.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur. **B.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Dans les conditions où la GFPssrA libre est seule (1) ou en présence de streptavidine monovalente libre avec 100 mM ATP (3), le pourcentage de GFPssrA dénaturée est de 20 à 25% tandis que la pente à 5 minutes est d'environ 0,75 min<sup>-1</sup> comme attendu pour les contrôles négatifs. Quand la GFPssrA est avec PAN (2) ou avec PAN et la streptavidine monovalente libre (4) à 100 mM ATP, le pourcentage de GFPssrA dénaturée après 1 heure de réaction est entre 50 et 55% et la pente à 5 minutes est entre 6 et 7 min<sup>-1</sup>. Il apparaît donc que l'activité de dépliement de la GFPssrA libre par PAN est similaire en absence ou en présence de streptavidine monovalente libre en solution. En

conclusion, la streptavidine monovalente isolée n'a pas d'effet significatif sur l'activité de dépliement de la GFPssrA.

### 2.3. GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine

Une fois les différentes expériences contrôle réalisées dans les parties 2.1 et 2.2, une étude biochimique a été effectuée dans la condition SANS 5 et 6 (streptavidine liée respectivement à l'extrémité N-terminale ou C-terminale de la GFPssrA) (voir partie 1.3). Les conditions testées sont les mêmes que dans la partie 2.1. Les données représentées dans les figures 112 tel que dans les parties précédentes.



**Figure 112 : Mesure de l'activité du système à 55°C avec de la GFPssrA\_Nter-streptavidine (formée à partir de la GFPssrA\_G10C) et GFPssrA\_Cter-streptavidine (formée à partir de la GFPssrA\_H231C) (conditions SANS 5 et 6).** Les figures A & B, et C & D représentent respectivement les conditions avec de la GFPssrA\_Nter-streptavidine et de la GFPssrA\_Cter-streptavidine. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6, ce qui correspond aux réactions suivantes : (1) GFPssrA, (2) GFPssrA et PAN, (3) GFPssrA, PAN et particule 20S, et (4) GFPssrA et particule 20S toutes quatre en présence de 100 mM ATP, et (5) GFPssrA et PAN et (6) GFPssrA, PAN et particule 20S toutes deux sans ATP. **A, C.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur. **B, D.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les résultats présentent des tendances semblables dans les deux conditions SANS. Dans toutes les conditions testées (1 à 6), le pourcentage de GFPssrA dénaturée en fin de réaction est d'environ

40%. Cependant, les valeurs de pente sont différentes selon les conditions testées. Pour les contrôles négatifs (1, 4, 5 et 6), la pente à 5 minutes est d'environ  $1 \text{ min}^{-1}$ . En présence de PAN ou de la machinerie protéolytique complète à 100 mM ATP (2 et 3), la pente à 5 minutes est entre 2,2 et  $3,5 \text{ min}^{-1}$ . Les GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine semblent donc peu prises en charge par la machinerie protéolytique. Même si la pente initiale est un peu plus forte pour les conditions 2 et 3 que pour les contrôles négatifs, la fluorescence finale n'est pas significativement différente comparée aux contrôles négatifs.

Un Western-blot avec un anticorps primaire anti-GFP a été réalisé à partir des produits finaux des trois principales réactions des conditions SANS 5 et 6 (figure 113) en présence de 100 mM ATP : (1) GFPssrA-streptavidine, (2) GFPssrA-streptavidine et PAN, (3) GFPssrA-streptavidine, PAN et particule 20S.



**Figure 113 : Western-blot dans les conditions SANS 5 (A) et 6 (B)** avec anticorps primaire anti-GFP et anticorps secondaire anti-lapin conjugué à la HRP après SDS-PAGE, 15 % acrylamide. Le signal a été révélé par une exposition de 30 secondes après trempage de la membrane dans des solutions de luminol et de peroxyde puis visualisé grâce à une caméra CCD. Les échantillons sont numérotés de 1 à 3 et représentent les conditions suivantes à 100 mM ATP : (1) GFPssrA-streptavidine, (2) GFPssrA-streptavidine et PAN, (3) GFPssrA-streptavidine, PAN et particule 20S.

Deux bandes sont visibles pour tous les échantillons. Ces deux bandes correspondent à de la GFPssrA intacte et de la GFPssrA ayant perdu une partie de l'étiquette signal ssrA (voir partie 4 du chapitre 5). Cela signifie que les GFPssrA liées à la streptavidine monovalente n'ont pas été dégradées à l'issue de la réaction comme l'est la GFPssrA sous forme libre.

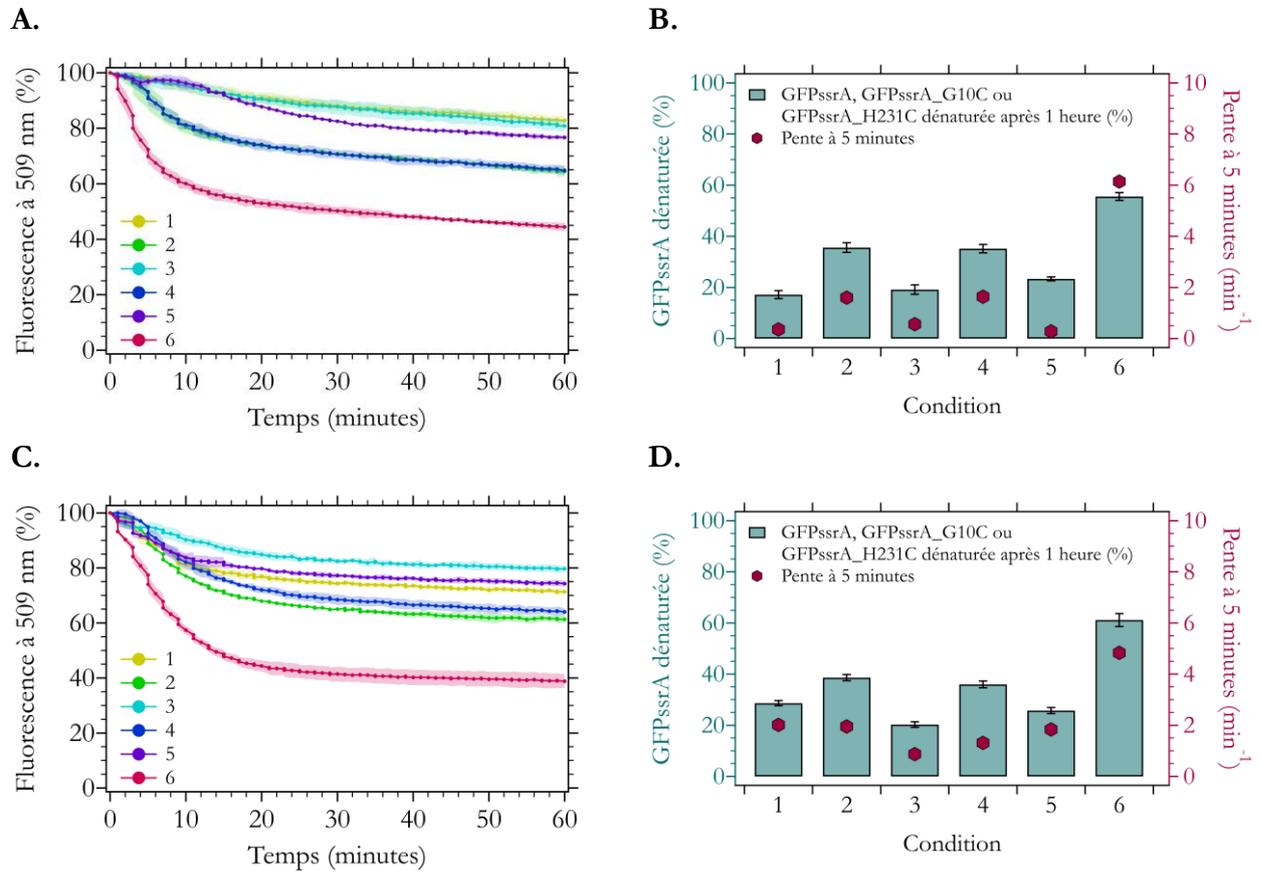
On peut donc conclure que **la GFPssrA\_Nter-streptavidine et la GFPssrA\_Cter-streptavidine ne sont soit pas efficacement prises en charge par le protéasome PAN-20S, soit la machinerie est bloquée au cours du dépliement et/ou du transfert de ces substrats protéique.** La streptavidine pourrait empêcher la reconnaissance et/ou la liaison de la GFPssrA par PAN, ou bien permettre ces étapes mais empêcher le dépliement et/ou le transfert à travers PAN.

## 2.4. Effet de l'ajout de GFPssrA libre

### 2.4.1. Ajout d'une quantité équimolaire de GFPssrA libre

Un test fonctionnel a été réalisé afin de déterminer si les GFPssrA-streptavidine, une fois mises en présence du protéasome PAN-20S, empêchaient la prise en charge de nouveaux substrats. Pour cela, des réactions dans les conditions SANS 5 et 6 (voir partie 1.3) et avec de la GFPssrA libre ont été réalisées tel que précédemment pendant une heure. De la GFPssrA libre a ensuite été ajoutée dans une quantité équimolaire à la GFPssrA-streptavidine, puis les réactions ont continué pendant une heure.

Les données représentées dans les figures 114.A et 114.C correspondent aux courbes d'évolution de la fluorescence en fonction du temps dans les différentes conditions testées, la figure 114.A pour les réactions initiales et la figure 114.B pour les réactions après ajout de la GFPssrA libre. Les figures 114.B et 114.D correspondent à l'analyse de ces courbes présentées de façon identique aux parties précédentes.



**Figure 114 : Mesure de l'activité du système à 55°C avec de la GFPssrA libre, GFPssrA\_Nter-streptavidine (formée à partir de la GFPssrA\_G10C) et GFPssrA\_Cter-streptavidine (formée à partir de la GFPssrA\_H231C) avec ajout après 1 heure de GFPssrA libre au ratio 1 protéines initiales : 1 GFPssrA libre.** Les figures A & B, et C & D représentent respectivement la réaction initiale et la réaction après ajout de la GFPssrA libre en présence de 100 mM ATP. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6, ce qui correspond aux réactions suivantes : (1) GFPssrA\_Nter-streptavidine, (2) GFPssrA\_Nter-streptavidine, PAN et particule 20S, (3) GFPssrA\_Cter-streptavidine, (4) GFPssrA\_Cter-streptavidine, PAN et particule 20S, (5) GFPssrA libre, (6) GFPssrA libre, PAN et particule 20S. **A, C.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur. **B, D.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.

Les figures 114.A et 114.B correspondent à la réaction initiale. Les résultats sont similaires à ceux des parties précédentes. Dans les conditions contrôle (1, 3 et 5), le pourcentage de GFPssrA dénaturée est de 17 à 25% tandis que la pente à 5 minutes est de 0,3 à 0,5 min<sup>-1</sup>, comme il est attendu. Quand la GFPssrA libre est en présence du protéasome PAN-20S à 100 mM ATP (6), le pourcentage de GFPssrA dénaturée après 1 heure de réaction est d'environ 55% et la pente à 5 minutes est d'environ 6 min<sup>-1</sup>. Enfin, les GFPssrA liées à la streptavidine avec la machinerie protéolytique à 100 mM ATP (2 et 4) présentent un pourcentage de GFPssrA dénaturée après 1

heure de réaction est d'environ 20% et une pente à 5 minutes est d'environ  $1,6 \text{ min}^{-1}$ . Comme décrit dans la partie précédente, le système apparaît donc bloqué en présence de GFPssrA-streptavidine.

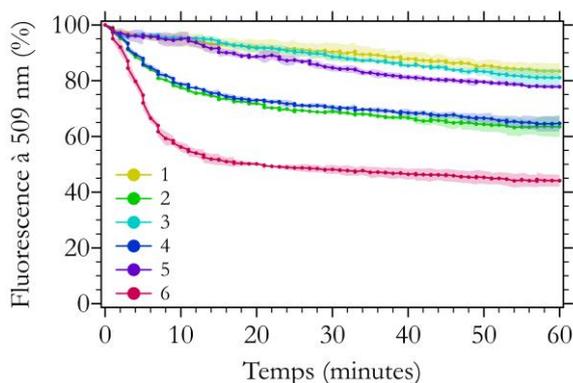
Après ajout de la GFPssrA libre (figures 114.C et 114.D), on peut interpréter les valeurs de la pente à 5 minutes qui correspondent à la dénaturation de la GFPssrA ajoutée mais pas les valeurs du pourcentage de GFPssrA dénaturée en fin de réaction, qui sont issues d'une moyenne des protéines substrat de la réaction initiale encore présentes et de la GFPssrA libre ajoutée. Pour les contrôles négatifs (1, 3 et 5), et pour les GFPssrA-streptavidine avec le protéasome PAN-20S à 100 mM ATP (2 et 4), la pente à 5 minutes est entre 1 et 2  $\text{min}^{-1}$ . Par contre, la GFPssrA libre en présence du protéasome PAN-20S à 100 mM ATP (6) présente une pente à 5 minutes d'environ  $5 \text{ min}^{-1}$ . La GFPssrA libre ajoutée après une heure d'incubation avec la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine est donc peu prise en charge comparée à lorsque le système a été incubé préalablement avec de la GFPssrA libre.

On peut donc conclure que **les GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine empêchent le protéasome PAN-20S de prendre en charge un nouveau substrat protéique si ce substrat est ajouté en même quantité que la GFPssrA-streptavidine**. Cela peut être parce que la streptavidine monovalente crée un encombrement stérique empêchant la prise en charge de nouveaux substrats, mais aussi parce que la GFPssrA liée à la streptavidine est bloquée sous forme liée à PAN au cours de son dépliement et/ou de son transfert et empêche un autre substrat d'être pris en charge par la machinerie protéolytique dans les conditions testées.

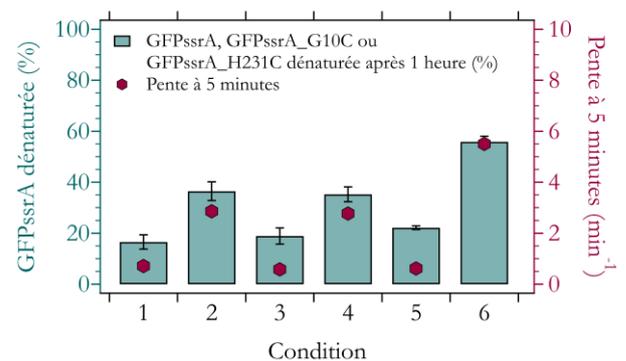
#### 2.4.2. Ajout d'une quantité cinq fois supérieure de GFPssrA libre

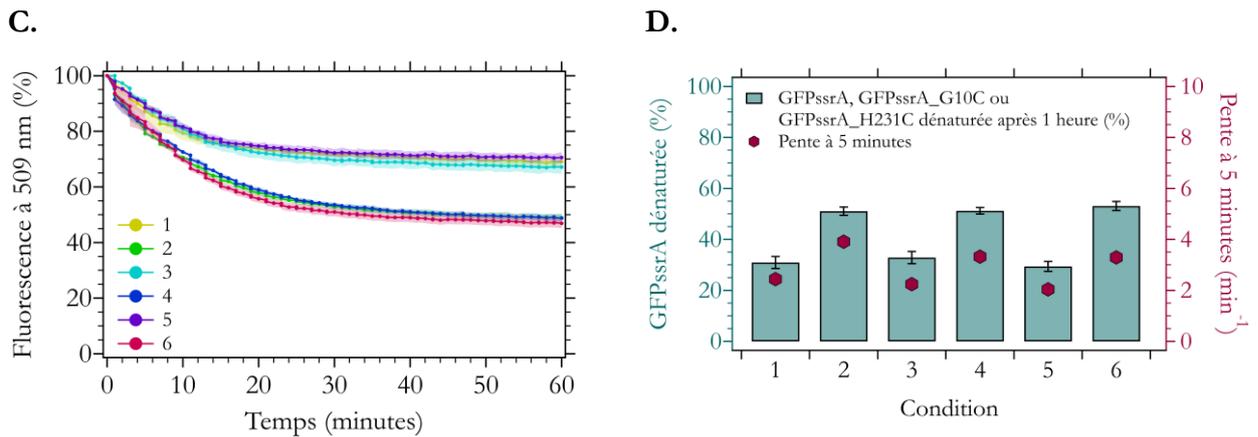
Une expérience a été réalisée afin déterminer si les GFPssrA-streptavidine, une fois incubées avec le protéasome PAN-20S, empêchaient la prise en charge de GFPssrA libre introduite en excès. Des mesures ont été effectuées de la même manière que dans la partie précédente, puis cinq fois plus de GFPssrA libre a ensuite été ajoutée par rapport aux protéines substrat initialement présentes. Les réactions ont de nouveau eu lieu pendant une heure. Les données obtenues sont représentées dans les figures 1115.A et 115.C de manière identique à la partie précédente.

**A.**



**B.**





**Figure 115 : Mesure de l'activité du système à 55°C avec de la GFPssrA libre, GFPssrA\_Nter-streptavidine (formée à partir de la GFPssrA\_G10C) et GFPssrA\_Cter-streptavidine (formée à partir de la GFPssrA\_H231C) avec ajout après 1 heure de GFPssrA libre au ratio 1 protéines initiales : 5 GFPssrA libre. Les figures A & B, et C & D représentent respectivement la réaction initiale et la réaction après ajout de la GFPssrA libre en présence de 100 mM ATP. Pour toutes les figures, les conditions testées sont numérotées de 1 à 6, ce qui correspond aux réactions suivantes : (1) GFPssrA\_Nter-streptavidine, (2) GFPssrA\_Nter-streptavidine, PAN et particule 20S, (3) GFPssrA\_Cter-streptavidine, (4) GFPssrA\_Cter-streptavidine, PAN et particule 20S, (5) GFPssrA libre, (6) GFPssrA libre, PAN et particule 20S. **A, C.** Courbes d'évolution de la fluorescence à 509 nm en fonction du temps dans les différentes conditions testées. Les barres d'erreur sont représentées pour chaque courbe sous forme d'aire nuancée de la même couleur. **B, D.** Histogrammes avec le pourcentage de GFPssrA dénaturée après 1 heure de réaction (par mesure de la diminution de la fluorescence à 509 nm) dans les différentes conditions testées. La pente de la tangente à la courbe après 5 minutes de réaction est représentée pour chaque condition par un point rouge.**

Les résultats pour la réaction initiale (figures 115.A et 115.B) sont identiques à ceux de la partie précédente. Lorsque de la GFPssrA est ajoutée en excès après une heure de réaction (figures 115.C et 115.D), la pente à 5 minutes est d'environ 2 min<sup>-1</sup> dans les conditions contrôle (1, 3 et 5). Pour la GFPssrA liée à la streptavidine ou la GFPssrA libre est en présence de la machinerie protéolytique à 100 mM ATP (2, 4 et 6), la pente à 5 minutes est entre 3 et 4 min<sup>-1</sup>.

La GFPssrA libre ajoutée en excès après une heure d'incubation avec la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine est donc prise en charge de la même manière que lorsque le système a été incubé préalablement avec de la GFPssrA libre. Cela signifie que **les GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine n'empêchent pas de dégrader une nouvelle protéine substrat lorsque celle-ci est introduite en excès.** Cela suggère que la **potentielle interaction entre la GFPssrA-streptavidine et PAN est faible ou instable.**

### 3. Discussion des mesures biochimiques réalisées

Le but principal de ces études biochimiques était d'étudier le système GFPssrA-streptavidine, protéasome PAN-20S dans les conditions permettant une activité optimale pour les mesures TR-SANS. Quatre nouvelles conditions SANS ont été décrites et comparées à de la GFPssrA libre (partie 1.3). Les compositions d'échantillon ont permis de réaliser des points de mesure en TR-SANS toutes les minutes avec un signal provenant de la GFPssrA exploitable. Les mesures de l'évolution de la fluorescence à 509 nm pendant 1 heure ont montré qu'un temps de mesure de 30 minutes était suffisant, le but étant de visualiser un substrat bloqué dans un état intermédiaire de dépliement.

Le comportement des protéines dans les concentrations élevées employées n'est pas complètement similaire à celui décrit dans l'article de référence<sup>317</sup>. Il est néanmoins important de noter que les concentrations de protéines utilisées dans les travaux de référence sont très différentes de celles employées dans notre étude (0,1  $\mu$ M GFPssrA-avidine et 5 nM PAN dans l'étude de Navon & Goldberg contre 25  $\mu$ M GFPssrA-streptavidine et 25  $\mu$ M PAN dans notre étude), ce qui pourrait expliquer les différents résultats obtenus. Dans les conditions expérimentales décrites dans ce chapitre, on peut vraisemblablement supposer que la GFPssrA\_Nter-streptavidine et la GFPssrA\_Cter-streptavidine sont bloquées au cours de leur dépliement et/ou de leur transfert à travers PAN. Les mesures TR-SANS permettront de discriminer entre cette hypothèse et celle selon laquelle ces protéines substrat ne sont pas prises en charge par la machinerie protéolytique.

Les conditions testées dans cette première partie de chapitre vont pouvoir maintenant être appliquées aux expérimentations par SANS afin de déterminer si l'un et/ou l'autre substrat protéique a été bloqué dans un état intermédiaire de dépliement et, si oui, d'interpréter les résultats obtenus.

## 4. Diffusion de neutrons aux petits angles en temps résolu (TR-SANS)

### 4.1. Préambule

Des mesures TR-SANS ont été réalisées sur la ligne de faisceaux D22 à l'ILL. Les données ont été enregistrées au cours de deux sessions différentes en raison d'un problème instrumental (rupture du vide au niveau du sélecteur de vitesse) survenu lors de la première session. De plus, nous avons rencontré d'autres problèmes techniques mineurs lors de la première session d'enregistrement de données qui ont conduit à une perte d'information pour les trois premières minutes de TR-SANS et deux premières minutes de fluorescence. Le tableau 21 récapitule les échantillons mesurés lors de chacune des sessions.

**Tableau 21: Liste des échantillons mesurés au cours des deux sessions d'enregistrement de données TR-SANS pour le système GFPssrA-streptavidine, protéasome PAN-20S.**

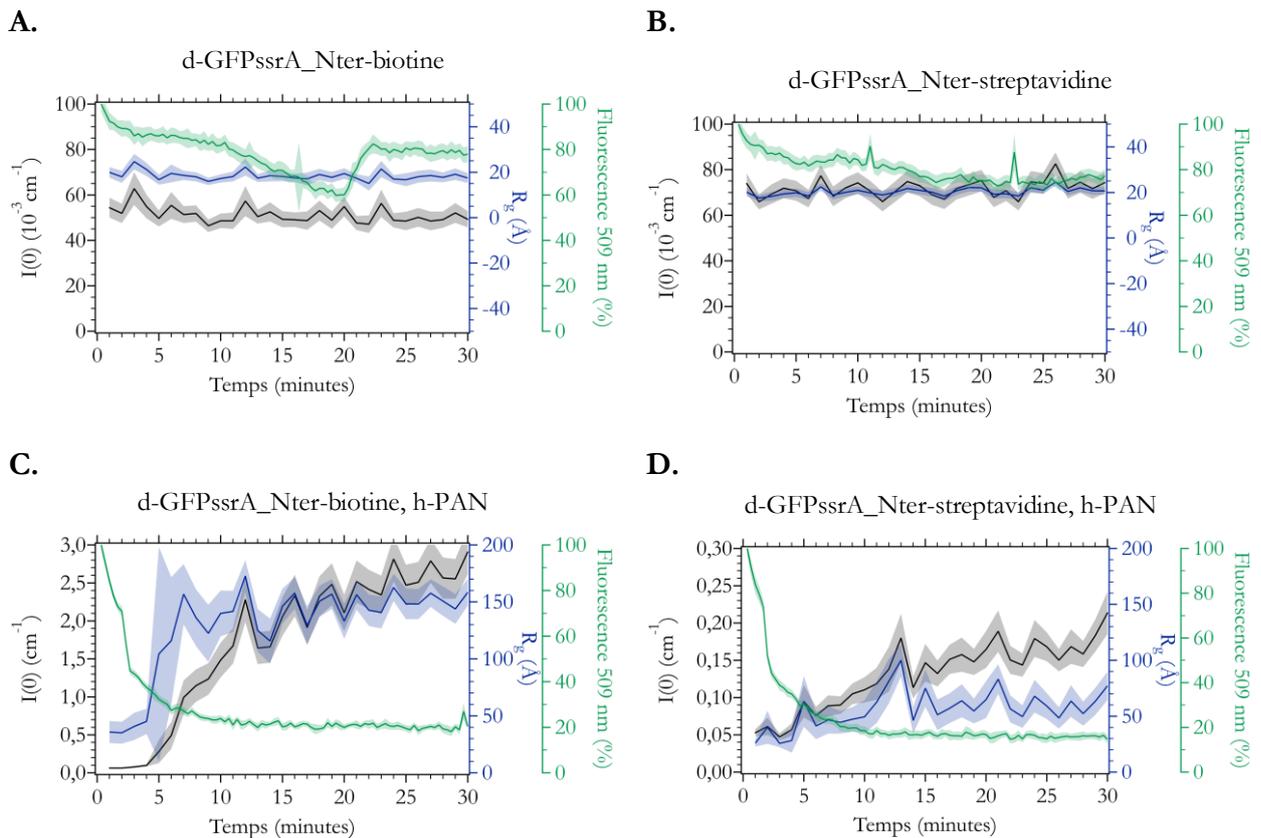
Échantillon		
Session 1	d-GFPssrA_Nter-biotine, protéasome h-PAN-h-20S	Condition SANS 3
	d-GFPssrA_Cter-biotine, protéasome h-PAN-h-20S	Condition SANS 4
	d-GFPssrA_Nter-streptavidine, protéasome h-PAN-h-20S	Condition SANS 5
	d-GFPssrA_Cter-streptavidine, protéasome h-PAN-h-20S	Condition SANS 6
Session 2	d-GFPssrA_Nter-biotine	Condition SANS 3
	d-GFPssrA_Nter-biotine, h-PAN	Condition SANS 3
	d-GFPssrA_Cter-biotine	Condition SANS 4
	d-GFPssrA_Cter-biotine, h-PAN	Condition SANS 4
	d-GFPssrA_Nter-streptavidine	Condition SANS 5
	d-GFPssrA_Nter-streptavidine, h-PAN	Condition SANS 5
	d-GFPssrA_Cter-streptavidine	Condition SANS 6
d-GFPssrA_Cter-streptavidine, h-PAN	Condition SANS 6	

Tous les échantillons souhaités ont néanmoins pu être mesurés au cours de ces deux sessions et les résultats sont présentés dans la partie suivante.

#### 4.2. Mesures TR-SANS avec la d-GFPssrA\_Nter-biotine et d-GFPssrA\_Nter-streptavidine

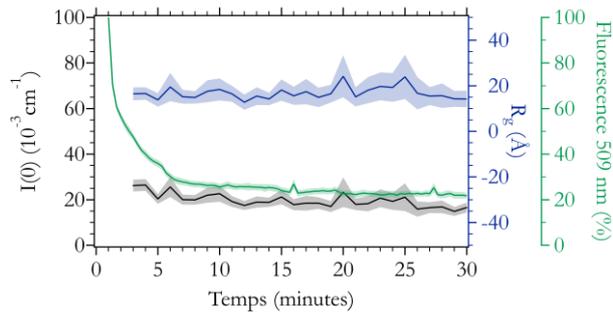
Des mesures TR-SANS ont été réalisées sur la d-GFPssrA\_Nter-biotine isolée (figure 116.A), en présence de h-PAN (figure 116.C) et en présence de h-PAN et de la particule h-20S (figure 116.E). Ces conditions correspondent à des contrôles où la GFPssrA est prise en charge par le protéasome h-PAN-h-20S. Des mesures identiques ont été effectuées sur la d-GFPssrA\_Nter-streptavidine isolée (figures 116.B, 116.C et 116.F). Tous les échantillons ont été préparés avec une concentration initiale de 100  $\mu\text{M}$  ATP & 200  $\mu\text{M}$   $\text{MgCl}_2$ . Les réactions ont été activées lors de leur disposition dans le porte-échantillon à 55°C et suivies pendant 30 minutes avec un point de mesure SANS toutes les minutes et un point de mesure de la fluorescence à 509 nm toutes les 20 secondes.

La figure 116 représente, pour chacun des échantillons, l'évolution de l'intensité à angle nul  $I(0)$  (noir), du rayon de giration  $R_g$  (bleu), et de la fluorescence à 509 nm (vert) au cours du temps. Dans le tampon à 42%  $\text{D}_2\text{O}$ , les signaux provenant de la biotine, streptavidine monovalente, protéine PAN et particule 20S hydrogénées sont masqués. Le  $R_g$  et  $I(0)$  correspondent alors exclusivement au signal de la GFPssrA deutérée et à ses formes dérivées. Les informations à déduire des valeurs  $R_g$  et de  $I(0)$  sont de la même nature que celles décrites dans la partie 2.3 du chapitre 7.

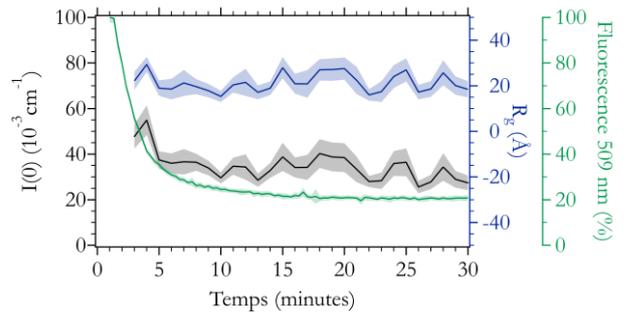


**E.**

d-GFPssrA\_Nter-biotine, h-PAN et particule h-20S

**F.**

d-GFPssrA\_Nter-streptavidine, h-PAN et particule h-20S



**Figure 116 : Évolution au cours du temps de l'intensité à angle nul  $I(0)$ , du rayon de giration  $R_g$ , et de la fluorescence à 509 nm des échantillons les conditions SANS 3 et 5, respectivement dans les panneaux de droite (A, C & E) et de gauche (B, D & F). A.** GFPssrA\_Nter-biotine. **B.** GFPssrA\_Nter-streptavidine. **C.** GFPssrA\_Nter-biotine en présence de PAN. **D.** GFPssrA\_Nter-streptavidine en présence de PAN. **E.** GFPssrA\_Nter-biotine en présence de PAN et de la particule 20S. **F.** GFPssrA\_Nter-streptavidine en présence de PAN et de la particule 20S. Un point de mesure a été réalisé toutes les 20 secondes pour la fluorescence à 509 nm et toutes les minutes pour le  $R_g$  et  $I(0)$ . Pour chaque courbe, les barres d'erreur sont représentées sous forme d'aire nuancée de la même couleur.

Lorsque la d-GFPssrA\_Nter-biotine (figure 116.A) ou la d-GFPssrA\_Nter-streptavidine (figure 116.B) est isolée en solution, les valeurs de  $R_g$  et de  $I(0)$  restent stables au cours du temps. La valeur du  $R_g$ , en moyenne de 20 Å, correspond à la valeur théorique du  $R_g$  de la GFP native (voir partie 4.4 du chapitre 5). La fluorescence diminue quant à elle d'environ 20%, ce qui correspond à la diminution de l'efficacité quantique de la fluorescence de la GFPssrA à 55°C. **La GFPssrA liée à la biotine ou à la streptavidine reste donc sous forme native lorsqu'elle est isolée en solution pendant 30 minutes à 55°C.**

Les valeurs de  $R_g$  et de  $I(0)$  de la d-GFPssrA\_Nter-biotine en présence de PAN (figure 116.C) augmentent rapidement et drastiquement au cours du temps. Comme pour la GFPssrA libre (voir partie 2.2 du chapitre 7), l'approximation de Guinier n'est plus valide après environ cinq minutes de réaction (puisque  $qR_g > 1,3$ ), ce qui empêche d'obtenir les valeurs réelles de  $R_g$  et de  $I(0)$ , qui sont ici sous-estimées. La fluorescence diminue en parallèle d'environ 80%. Ces tendances sont similaires à celle décrites dans la partie 2.2 du chapitre 7 pour la GFPssrA libre et indiquent que **la liaison de la biotine n'empêche pas son dépliement de la d-GFPssrA par h-PAN et son agrégation.**

La d-GFPssrA\_Nter-streptavidine en présence de PAN (figure 116.D) présente des valeurs  $R_g$  et de  $I(0)$  augmentant également au cours du temps. Cependant, l'augmentation de ces valeurs est limitée comparé à la d-GFPssrA\_Nter-biotine et l'approximation de Guinier reste valide tout au long de la réaction ( $qR_g \leq 1,3$ ), permettant d'interpréter les valeurs de  $R_g$  et de  $I(0)$ . Ainsi, ces deux valeurs augmentent progressivement au cours du temps jusqu'à atteindre un facteur trois entre le début et la fin de la réaction. En parallèle, la fluorescence diminue d'environ 80%. **La d-GFPssrA est donc dépliée par h-PAN et forme des oligomères de d-GFPssrA de faible masse moléculaire.**

Lorsque la d-GFPssrA\_Nter-biotine est en présence du protéasome h-PAN-h-20S (figure 116.E), le  $R_g$  reste stable au cours du temps avec une valeur correspondant à celle de la GFP native. La

valeur du  $I(0)$  diminue faiblement au cours du temps. Cependant, il est important de noter que les données des trois premières minutes de réaction n'ont pas pu être enregistrées (voir partie 4.1). On sait que la réaction de dépliage et de dégradation a lieu lors des premières minutes de réaction. De plus, les valeurs initiales du  $I(0)$  dans les échantillons de d-GFPssrA\_Nter-biotine isolée et en présence de h-PAN (voir tableau 22) indiquent qu'il y a eu une perte importante de la valeur du  $I(0)$  au cours des trois premières minutes de réaction, qui a ensuite continué de diminuer progressivement jusqu'à une perte d'environ 75%. La fluorescence diminue quant à elle d'environ 77%. **La d-GFPssrA est dégradée au cours du temps de façon similaire à la d-GFPssrA libre en condition SANS 1 (voir partie 2.2 du chapitre 7). Parmi les d-GFPssrA dénaturées par h-PAN, environ 97% sont dégradées par la particule catalytique h-20S**

La d-GFPssrA\_Nter-streptavidine en présence de la machinerie protéolytique complète hydrogénée (figure 116.F) présente également un  $R_g$  stable tout au long de la réaction avec une valeur correspondant à celle de la GFP native. Tout comme la réaction contrôlée avec la d-GFPssrA\_Nter-biotine, la valeur de  $I(0)$  diminue au cours du temps, avec une perte totale de d'environ 55% (voir tableau 22). La fluorescence diminue en parallèle d'environ 79%. On peut donc en conclure qu'**environ 69% des d-GFPssrA dénaturées par h-PAN sont dégradées par la particule h-20S.**

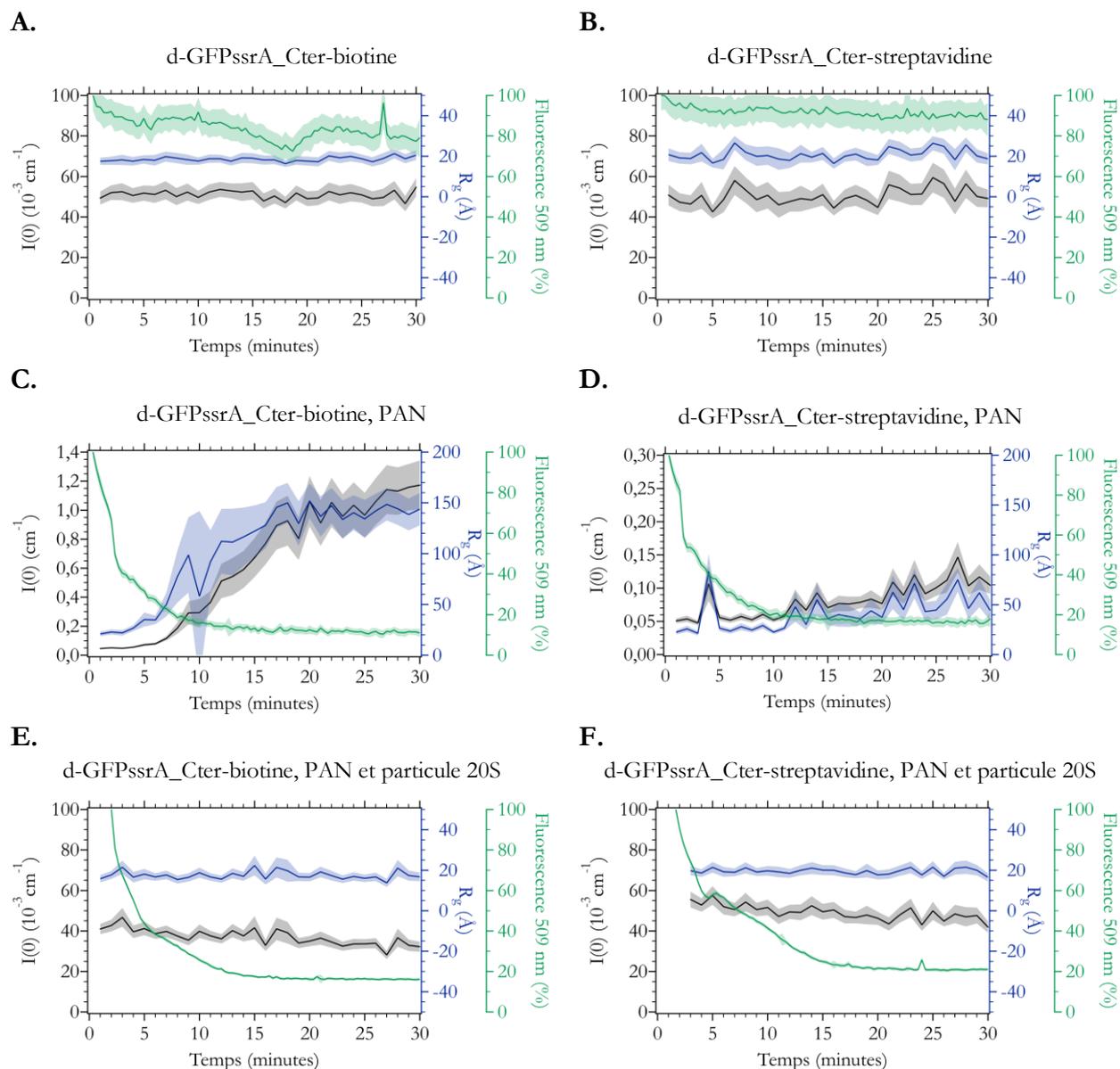
**Tableau 22: Valeurs initiales et finales de l'intensité diffusée à angle zéro  $I(0)$  et fluorescence à 509 nm pour les échantillons de d-GFPssrA\_Nter-biotine (condition SANS 3) et d-GFPssrA\_Nter-streptavidine (condition SANS 5) contenant h-PAN et la particule h-20S. Le pourcentage d'activité a été déduit pour chaque paramètre à partir de ces valeurs.**

	d-GFPssrA_Nter-biotine	d-GFPssrA_Nter-streptavidine
	<b><math>I(0)</math></b>	
<b><math>I(0)</math> initial d-GFPssrA (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	54,4 ± 4,4	73,8 ± 4,5
<b><math>I(0)</math> initial d-GFPssrA, protéasome h-PAN-h20S (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	26,2 ± 2,6	47,7 ± 5,3
<b><math>I(0)</math> final d-GFPssrA, protéasome h-PAN-h20S (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	13,5 ± 2,2	32,9 ± 6,7
<b>Perte de <math>I(0)</math> (%)</b>	~75	~55
	<b>Fluorescence</b>	
Fluorescence initiale (%)	100 ± 2,5	100 ± 2,3
Fluorescence finale (%)	22,5 ± 1,4	20,7 ± 0,9
<b>Perte de fluorescence (%)</b>	~77	~79

### 4.3. Mesures TR-SANS avec la GFPssrA\_Cter-biotine et GFPssrA\_Cter-streptavidine

Des mesures TR-SANS similaires à celles décrites dans la partie précédente ont été effectuées sur la GFPssrA\_Cter-biotine isolée (figure 117.A), en présence de PAN (figure 117.C) et en présence de PAN et de la particule 20S (figure 117.E), ainsi que sur la GFPssrA\_Cter-streptavidine dans les

mêmes conditions (figures 117.B, 117.C et 117.F). Les données sont représentées de la même manière que précédemment.



**Figure 117 : Évolution au cours du temps de l'intensité à angle nul  $I(0)$ , du rayon de giration  $R_g$ , et de la fluorescence à 509 nm des échantillons les conditions SANS 4 et 6, respectivement dans les panneaux de droite (A, C & E) et de gauche (B, D & F). A. d-GFPssrA\_Cter-biotine. B. d-GFPssrA\_Cter-streptavidine. C. d-GFPssrA\_Cter-biotine en présence de PAN. D. d-GFPssrA\_Cter-streptavidine en présence de PAN. E. d-GFPssrA\_Cter-biotine en présence de PAN et de la particule 20S. F. d-GFPssrA\_Cter-streptavidine en présence de PAN et de la particule 20S. Un point de mesure a été réalisé toutes les 20 secondes pour la fluorescence à 509 nm et toutes les minutes pour le  $R_g$  et  $I(0)$ . Pour chaque courbe, les barres d'erreur sont représentées sous forme d'aire nuancée de la même couleur.**

La plupart des échantillons de d-GFPssrA\_Cter-biotine et d-GFPssrA\_Cter-streptavidine ont des résultats semblables à ceux de la d-GFPssrA\_Nter-biotine et d-GFPssrA\_Nter-streptavidine (partie précédente).

**La d-GFPssrA\_Cter-biotine et la d-GFPssrA\_Cter-streptavidine isolées en solution restent stables sous forme native** au cours du temps avec des valeurs de  $R_g$  et de  $I(0)$  constantes, tandis que la fluorescence diminue d'environ 20% (figure 117.A et 117.B).

La d-GFPssrA\_Cter-biotine en présence de h-PAN (figure 117.C) présente des valeurs de  $R_g$  et de  $I(0)$  qui augmentent rapidement et drastiquement au cours de la réaction, qui sont de plus sous-estimées (approximation de Guinier non valide). La fluorescence diminue quant à elle d'environ 90%. Cela signifie que **la d-GFPssrA est dépliée par h-PAN et s'agrège**. Les valeurs de  $R_g$  et de  $I(0)$  de la d-GFPssrA\_Cter-streptavidine en présence de h-PAN (figure 117.D) augmentent également au cours du temps, de la même manière que pour la d-GFPssrA\_Nter-streptavidine dans les mêmes conditions. Ces deux valeurs augmentent progressivement au cours du temps jusqu'à atteindre un facteur deux (au lieu de trois) entre le début et la fin de la réaction, alors que la fluorescence diminue d'environ 90%. **La d-GFPssrA est donc dépliée par h-PAN et forme des oligomères de d-GFPssrA de faible masse moléculaire.**

La valeur  $R_g$  de la d-GFPssrA\_Cter-biotine en présence du protéasome h-PAN-h-20S (figure 117.E) reste stable au cours du temps et correspond à celle de la d-GFP native. Cependant, les données des trois premières minutes de réaction n'ont pas pu être enregistrées. Tout comme dans la partie précédente, la valeur de  $I(0)$  diminue au cours du temps. La perte totale de  $I(0)$  est d'environ 38%, tandis que la fluorescence diminue d'environ 84% (tableau 23). On peut donc conclure que **parmi les d-GFPssrA dénaturés par h-PAN, environ 45% sont ensuite dégradés par la particule h-20S.**

La d-GFPssrA\_Cter-streptavidine dans les mêmes conditions (figure 117.F) présente également un  $R_g$  stable tout au long de la réaction avec une valeur correspondant à celle de la GFP native. Alors que la perte totale de  $I(0)$  est d'environ 27%, la fluorescence diminue d'environ 80% (tableau 23). En conclusion, **environ 33% des d-GFPssrA dénaturées par h-PAN sont ensuite dégradées par la particule h-20S.**

**Tableau 23: Valeurs initiales et finales de l'intensité diffusée à angle zéro  $I(0)$  et fluorescence à 509 nm pour les échantillons de d-GFPssrA\_Cter-biotine (condition SANS 3) et d-GFPssrA\_Cter-streptavidine (condition SANS 5) contenant h-PAN et la particule h-20S. Le pourcentage d'activité a été déduit pour chaque paramètre à partir de ces valeurs.**

	d-GFPssrA_Cter-biotine	d-GFPssrA_Cter-streptavidine
	<b><math>I(0)</math></b>	
<b><math>I(0)</math> initial d-GFPssrA (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	49,3 ± 3,2	50,8 ± 5,1
<b><math>I(0)</math> initial d-GFPssrA, protéasome h-PAN-h20S (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	40,9 ± 2,5	55,5 ± 5,3
<b><math>I(0)</math> final d-GFPssrA, protéasome h-PAN-h20S (<math>10^{-3}</math> cm<math>^{-1}</math>)</b>	30,3 ± 2,4	40,6 ± 3,8
<b>Perte de <math>I(0)</math> (%)</b>	~38	~27
	<b>Fluorescence</b>	
Fluorescence initiale (%)	100 ± 1,6	100 ± 1,4
Fluorescence finale (%)	16,2 ± 0,6	20,4 ± 0,9

Perte de fluorescence (%)	~84	~80
---------------------------	-----	-----

#### 4.4. Conclusions à retenir des mesures TR-SANS

Les résultats obtenus par TR-SANS permettent de dresser les conclusions suivantes :

- **Les quatre substrats protéiques utilisés restent stables au cours du temps lorsqu'ils sont isolés en solution dans les conditions testées.** Cela est en accord avec les mesures réalisées sur la GFPssrA libre (chapitre 7).
- **La d-GFPssrA\_Nter-biotine et la d-GFPssrA\_Cter-biotine s'agrègent en présence de h-PAN et sont dégradées en présence de la machinerie protéolytique complète hydrogénée.** Il s'agit des résultats attendus pour ces réactions contrôles. En effet, cela est cohérent avec les résultats obtenus pour la GFPssrA libre (chapitre 7) et lors des études biochimiques (parties précédentes de ce chapitre).
- **La d-GFPssrA\_Nter-streptavidine et d-GFPssrA\_Cter-streptavidine sont dépliés et semblent former des oligomères de faible masse moléculaire en présence de PAN.** Les deux d-GFPssrA-streptavidine s'agrègent donc moins que leur équivalent d-GFPssrA-biotine. Malgré une oligomérisation, l'approximation de Guinier est validée tout au long de la réaction. Cela indique que la taille des structures formée est accessible par la technique. **La streptavidine semble donc avoir un effet protecteur contre l'agrégation de la d-GFPssrA.** Comme l'indique la diminution de fluorescence, la d-GFPssrA est dépliée par h-PAN, puis, comme le montre l'augmentation du I(0), forme des oligomères. Le fait que la diminution de fluorescence soit semblable à celle observée pour les d-GFPssrA-biotine suggère que h-PAN n'est pas affecté par la streptavidine dans son processus de dépliement. Il est alors probable que la présence de streptavidine provoque une gêne stérique en solution, empêchant de larges agrégats de se former. Il est intéressant de noter que la taille des oligomères formés varie légèrement entre les deux formes de d-GFPssrA-streptavidine. Les oligomères formés par la d-GFPssrA\_Nter-streptavidine étaient plus gros que ceux formés par la d-GFPssrA\_Cter-streptavidine, même s'il est délicat de discuter si cette différence est significative ou non.

**Tableau 24 : Valeurs de perte de fluorescence à 509 nm et de perte de I(0) pour les différents substrats protéiques testés.** Le pourcentage de d-GFPssrA dénaturées ensuite dégradées de chaque population de protéine substrat a été déduit de ces valeurs.

	Perte de fluorescence (%)	Perte de I(0) (%)	Protéines dénaturées ensuite dégradées (%)
d-GFPssrA_Nter-biotine	77	75	97
d-GFPssrA_Nter-streptavidine	79	55	69
d-GFPssrA_Cter-biotine	84	38	45
d-GFPssrA_Cter-streptavidine	80	27	33

- **En présence de la machinerie protéolytique complète hydrogénée, les quatre protéines substrat sont dénaturées par h-PAN mais dégradées par la particule catalytique h-20S selon des taux différents.** De plus, il n'y pas d'agrégation visible de la d-GFPssrA. Le tableau 24 résume les tendances pour les différents substrats protéiques testés en présence de h-PAN et de particule h-20S.

Ces derniers résultats nécessitent d'être discutés en lumière des études biochimiques précédemment réalisées.

## 5. Discussion et conclusion générale du chapitre

Les mesures TR-SANS indiquent que la d-GFPssrA\_Nter-streptavidine et d-GFPssrA\_Cter-streptavidine sont dépliées et forment des oligomères de faible masse moléculaire en présence de h-PAN. La taille des oligomères est probablement due à présence de la streptavidine en solution qui empêche de former de plus gros agrégats. En présence du protéasome h-PAN-h-20S, les substrats protéiques sont pris en charge différemment. Alors que toutes les molécules de d-GFPssrA\_Nter-biotine dénaturées par h-PAN sont dégradées par la particule h-20S, seulement la moitié des molécules de d-GFPssrA\_Cter-biotine dénaturées sont ensuite dégradées. En ce qui concerne les d-GFPssrA-streptavidine, environ deux tiers des molécules de d-GFPssrA\_Nter-streptavidine et un tiers des molécules de d-GFPssrA\_Cter-streptavidine dépliées sont ensuite dégradées. Il est difficile d'interpréter à partir de ces données si certains des substrats protéiques utilisés sont bloqués dans un état intermédiaire déplié au cours de leur processus dépliement et/ou de transfert à travers le protéasome h-PAN-h-20S. De plus, il ne nous paraît pas pertinent de réaliser une analyse plus fine à partir de ces données obtenues. En effet, les résultats obtenus sont basés sur des extrapolations puisque les trois premières minutes de mesures étaient manquantes dans les échantillons contenant la particule 20S. De plus, les courbes d'évolution de la fluorescence obtenues lors de ces expériences par TR-SANS dans les conditions SANS 5 et 6 (respectivement d-GFPssrA\_Nter-streptavidine et d-GFPssrA\_Cter-streptavidine) sont en désaccord avec les études biochimiques présentées en première partie de ce chapitre.

Les études biochimiques réalisées par mesure de la fluorescence de la GFPssrA avaient révélé que la fluorescence de la GFPssrA ne diminuait pas en présence de PAN ou de la machinerie protéolytique complète dans les conditions SANS 5 et 6. Pourtant, les mesures avaient été effectuées en utilisant exactement les mêmes composants dans des conditions similaires (seul le volume final de l'échantillon était plus faible). Les différences notables entre les mesures de fluorescence pendant aux mesures SANS et les précédentes étaient le spectrophotomètre utilisé et le contrôle de la température. Alors qu'il a été contrôlé que la température soit exactement à 55°C à l'intérieur de la cuvette contenant l'échantillon lors des mesures SANS, la température du spectrophotomètre utilisé au laboratoire a été contrôlée à l'intérieur de l'instrument mais pas directement dans l'échantillon. Or, une mesure indépendante a permis de montrer que lorsque l'instrument était réglé à 55°C, la température de l'échantillon était d'environ 50°C. Comme montré dans la partie 5 du chapitre 6, l'activité de la machinerie protéolytique dépend fortement de la température. Il se pourrait alors que la protéine PAN, ralentie par la prise en charge de substrats tels que la GFPssrA\_Nter-streptavidine et GFPssrA\_Cter-streptavidine, nécessite une température d'au moins 55°C pour avoir une efficacité significative. Les mesures biochimiques décrites dans l'article de référence<sup>317</sup> avaient pourtant lieu à 45°C, même si les concentrations différentes utilisées pourraient expliquer les différences observées.

Un échantillon contenant la d-GFPssrA-streptavidine et le protéasome h-PAN-h-20S est complexe de par la polydispersité des espèces de d-GFPssrA pouvant être présentes en solution (GFPssrA native, forme dépliée, peptides). Il existe alors un risque de surinterprétation non négligeable, qui doit être limité par des mesures rigoureuses. En l'absence de résultats reproductibles entre tests

biochimiques et les mesures par TR-SANS, et compte tenu des difficultés techniques rencontrées lors de mesures TR-SANS, interpréter les résultats obtenus apparaît délicat. Les tendances observées lors des mesures TR-SANS semblent néanmoins prometteuses. Il faudrait refaire des mesures biochimiques avec un meilleur contrôle de la température et de nouvelles mesures TR-SANS avec des mesures sur la totalité de la gamme de temps. Des résultats plus rigoureux permettraient alors de faire des interprétations sur le mécanisme du protéasome PAN-20S.



## Chapitre 9 :

### Conclusion générale et perspectives

L'objectif principal de cette thèse était de réaliser une étude biophysique sur le protéasome PAN-20S dans le but d'apporter de nouvelles informations sur les mécanismes employés par cette machine moléculaire pour déplier et dégrader les protéines substrat. Pour cela, une approche novatrice en solution par diffusion de neutrons aux petits angles résolue en temps (TR-SANS) couplée à de la spectroscopie de fluorescence a été appliquée.

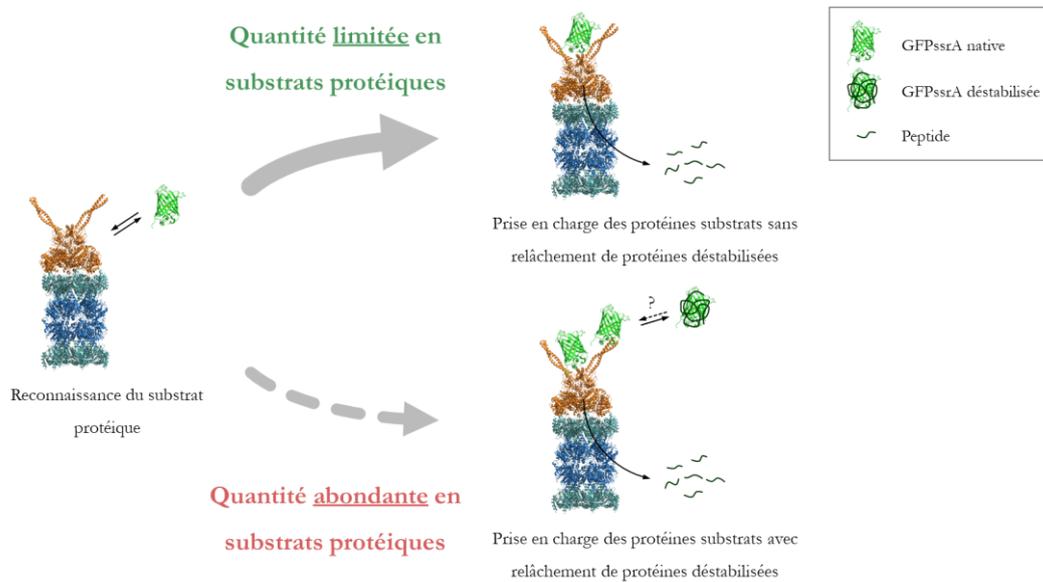
Les protéines ont d'abord été étudiées de façon isolées (chapitre 5), puis des études biochimiques ont permis de déterminer les conditions de mesures optimales aux études par TR-SANS (chapitre 6). Notamment, l'hydrolyse de l'ATP au cours du temps dans les deux compositions d'échantillon retenues a été suivie et mise en relation avec les données de TR-SANS et de mesure de la fluorescence. Les mesures TR-SANS couplées à de la spectrofluorimétrie (chapitre 7) ont permis de suivre de façon résolue en temps le dépliement et la dégradation de la GFP<sub>ssrA</sub> par la machinerie protéolytique. L'ensemble de ces données nous amènent à proposer un mécanisme d'action du protéasome PAN-20S en particulier, et des autres protéases AAA+ en général.

Deux situations ont pu être mises en évidence, illustrant les différentes modalités de prise en charge des substrats protéiques par le protéasome PAN-20S en fonction de la quantité de protéines à dégrader. Les deux compositions d'échantillon étudiées, l'une en quantité « limitée » et l'autre en quantité « abondante » en substrats protéiques par rapport à PAN (selon les considérations énoncées dans le chapitre 7), ont permis de proposer un modèle de prise en charge des protéines substrat par le protéasome PAN-20S.

Nous proposons que, dans une situation **en quantité limitée en substrats protéiques, un transfert direct du substrat de la particule régulatrice PAN vers la particule catalytique 20S** est réalisé lorsque la GFP<sub>ssrA</sub> est reconnue par PAN. En effet, toutes les GFP<sub>ssrA</sub> dépliées sont dégradées et ce mécanisme a lieu sans libération intermittente, ou avec une libération très limitée et non détectable, de substrats protéiques dépliés ou déstabilisés en solution. Par contre, lors d'une situation **en quantité abondante en protéines substrat, PAN relâche en solution une partie des GFP<sub>ssrA</sub> sous forme déstabilisée**. En effet, seulement un peu plus de la moitié des protéines dépliées sont dégradées par le protéasome PAN-20S. Les autres protéines sont alors sous forme déstabilisée non fluorescente. Ces formes déstabilisées de GFP<sub>ssrA</sub> sont structurellement similaires à la forme native de la GFP<sub>ssrA</sub> et ne forment pas d'agrégats.

Différentes hypothèses concernant l'état des GFP<sub>ssrA</sub> relâchées sont alors envisageables : ces protéines pourraient être légèrement déstabilisées suite à des tentatives infructueuses de dépliement de PAN, ou alors partiellement dépliées et éventuellement partiellement dégradées. De plus, on ne sait pas si ces protéines sont relâchées de façon libre en solution ou toujours associées à PAN. Les études de RMN, réalisées par des collaborateurs, permettront d'apporter des informations complémentaires à celles obtenues lors de cette étude. Il est possible que cette situation d'excès d'un substrat protéique ait lieu naturellement dans les cellules lors des conditions de stress intenses ou dans les cellules eucaryotes lors de l'accumulation de protéines mal repliées en cas de mutations,

maladies dégénératives, lors du vieillissement, etc. La figure 118 représente le modèle proposé suite à l'analyse des données.

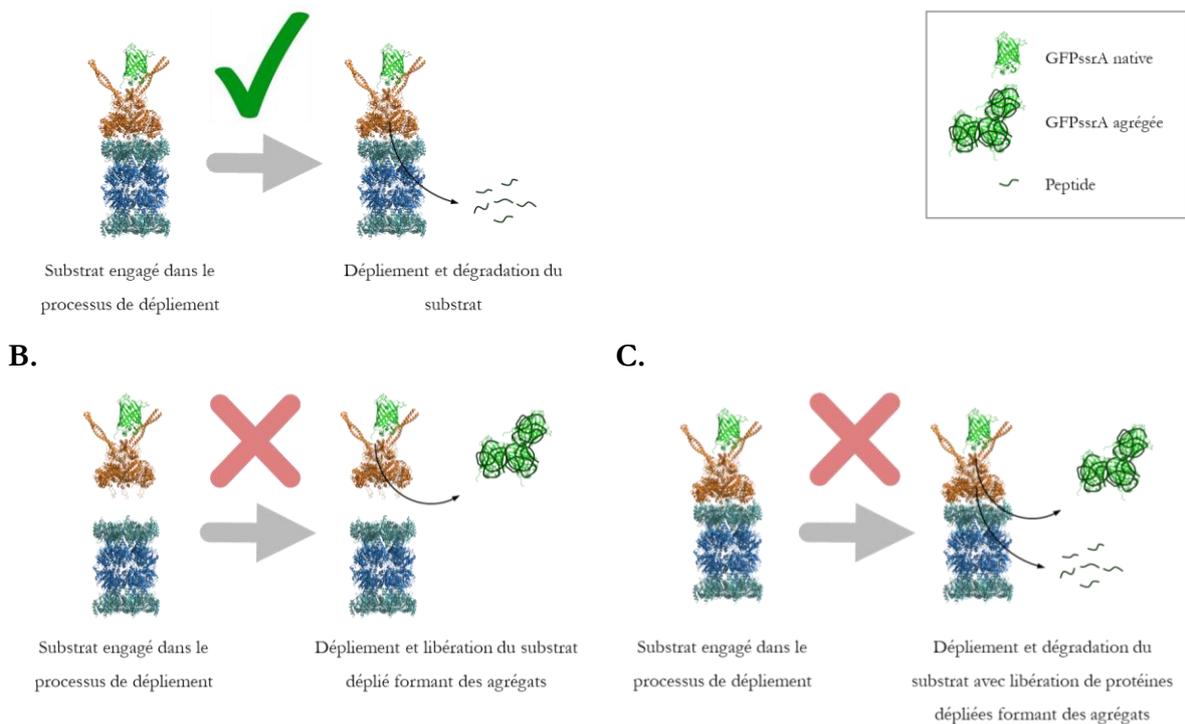


**Figure 118 : Modèle de prise en charge des substrats protéiques par le protéasome PAN-20S.** Schéma représentant les deux possibilités de processus de reconnaissance et de prise en charge des substrats par la machinerie protéolytique PAN-20S. En quantité limitée de protéines substrat (flèche en ligne continue), la GFPssrA reconnue par PAN est dépliée et dégradée en une seule étape selon un processus linéaire et concerté entre PAN et la particule 20S. En quantité abondante de protéines substrat (flèche en pointillé), une fois reconnues par PAN, certaines GFPssrA sont libérées sous forme déstabilisées. Modèle basé sur les PDB 6HEA & PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale.

Des agrégats de GFPssrA se forment lorsque la PAN déplie complètement une protéine et que la particule catalytique 20S n'est pas présente pour effectuer la dégradation subséquente. Il est à souligner que dans les deux situations étudiées (abondante et limitée en protéines substrat), aucun agrégat n'est détecté dans les échantillons où la GFPssrA est en présence de PAN et de particule 20S. **Les protéines substrat sont donc prises en charge par PAN lorsque le protéasome PAN-20S est déjà assemblé. La machinerie protéolytique reste ensuite assemblée** lorsque le substrat protéique est déplié par PAN et qu'il est transféré au sein de la particule catalytique 20S. Le protéasome PAN-20S pourrait rester assemblé par lui-même ou le substrat protéique en cours de transfert pourrait servir de fil connecteur pour maintenir les deux particules associées. En conséquence, une fois le processus de dépliement de la protéine substrat engagé, **il n'y a pas de relâchement possible** et le processus de destruction s'accomplit jusqu'à son terme. **Des peptides d'une dizaine d'acides aminés en moyenne sont alors générés.** Le modèle proposé est représenté dans la figure 119.

Ce modèle souligne la synergie fonctionnelle entre la protéine régulatrice AAA+ PAN et la particule catalytique 20S qui permet au système de ne pas relâcher des protéines qui seraient toxiques dans le cytosol en formant des corps d'inclusion. Cela suppose également que l'association PAN-20S doit être étroitement contrôlé dans les cellules pour éviter l'accumulation de protéines substrat dénaturées, et/ou que les formes libres de PAN sont inactives dans la cellule.

**A.**



**Figure 119 : Modèle de dépliement et dégradation des substrats protéiques par le protéasome PAN-20S une fois le substrat protéique engagé dans PAN.** Schéma représentant les possibilités de mode d'action de la machinerie protéolytique. **A.** Le protéasome PAN-20S est déjà assemblé lorsque les substrats sont engagés dans le processus de dépliement. Les substrats protéiques seraient dépliés et dégradés sans libération de formes dénaturées formant des agrégats. **B.** Le protéasome PAN-20S n'est pas encore assemblé lors de l'engagement des substrats protéiques dans le processus de dépliement. Cela aurait pour conséquence la libération de substrats protéiques dépliés formant des agrégats. **C.** Le protéasome PAN-20S est déjà assemblé lorsque les substrats sont engagés dans le processus de dépliement et de dégradation. L'assemblage du protéasome complet pourrait permettre une libération de protéines dénaturées formant des agrégats, soit à cause d'une dissociation du protéasome PAN-20S, soit même lorsque le protéasome PAN-20S est assemblé. Les résultats obtenus ont permis de valider l'hypothèse A. Modèles générés à partir des PDB 6HEA & PDB 1EMB avec une partie flexible ajoutée à l'extrémité C-terminale.

Ainsi, ces expériences mettent en évidence le mode d'action rapide et coordonné du protéasome PAN-20S. **Une fois le dépliement du substrat protéique engagé, les processus de transfert et de dégradation sont étroitement couplés.** Cela souligne le caractère de haute processivité de la machinerie protéolytique. Nous confirmons également que **l'hydrolyse de l'ATP module directement l'activité de dépliement et de dégradation des protéines substrat par le protéasome PAN-20S.**

Les résultats obtenus au cours de ce projet font l'objet d'un article en phase finale de rédaction et ouvrent de nombreuses perspectives dans le cadre de l'étude du protéasome PAN-20S.

Toutes les expériences de ce projet ont été réalisées avec un substrat protéique modèle, la GFPssrA, qui est une protéine monomérique sous forme de tonneau  $\beta$ . Il serait intéressant de réaliser de nouvelles expériences avec **d'autres substrats protéiques** tels que des oligomères, des protéines possédant des parties flexibles ou intrinsèquement dépliées pour étudier la capacité de la machinerie protéolytique à prendre en charge différentes protéines substrat et le mode d'action associé. L'utilisation simultanée de différents substrats protéiques permettrait de déterminer si le protéasome PAN-20S dégrade certains types de protéines de façon prioritaire par rapport à d'autres.

Notre compréhension de la **régulation du protéasome PAN-20S et de ses substrats protéiques par des adaptateurs ou par des réactions de modifications post-traductionnelles** est limitée. Identifier et caractériser au niveau biochimique et structural des partenaires de la machinerie protéolytique est donc un enjeu actuel important. De plus, plusieurs questions restent ouvertes quant au mécanisme de reconnaissance des substrats. Peu d'informations sont disponibles sur les **différents dégrons reconnaissables par PAN, ainsi que sur les voies dans lesquelles ils sont impliqués et les mécanismes par lesquels ils sont reconnus et pris en charge**. Par exemple, il n'est toujours pas décrit comment se fait la linéarisation par le protéasome PAN-20S de protéines substrat dont le dégron n'est pas positionné à l'extrémité N-terminale ou C-terminale, comme dans le cas des protéines naturellement sampylées et ubiquitinylées sur des lysines. Cela pose la question de l'existence *in vivo* d'un couvercle équivalent à celui du 19S du protéasome 26S eucaryote chez le protéasome PAN-20S qui permettrait d'orienter les substrats protéiques favorablement pour qu'une de leurs extrémités soit prise en charge par PAN. Dans ce contexte, il est envisagé de « sampyler » *in vitro* des substrats protéiques pour étudier leur mécanisme de dépliement et de dégradation par la machinerie protéolytique.

Ce projet est la **première étude structurale suivant une protéine substrat dégradée par le protéasome PAN-20S, ceci en relation avec des paramètres biochimiques**. Il s'agit d'une des rares études suivant en solution des modifications structurales en temps réel pendant une activité de remodelage de protéines par des protéines AAA+. Cette approche pourra être utilisée dans le futur pour d'autres études fonctionnelles en temps réel d'une variété de grands assemblages macromoléculaires comme les moteurs moléculaires, chaperonne, machinerie de réplication et réparation de l'ADN ou encore la machinerie de dégradation de l'ARN. Dans ce contexte, **les protéines hyperthermophiles demeurent des cibles d'étude à privilégier** compte tenu des temps d'acquisition nécessaires en TR-SANS. De futurs développements en TR-SANS rendront possible l'utilisation de temps d'exposition plus courts. En conséquence, l'acquisition d'une grande quantité de données expérimentales devrait stimuler le développement de **logiciels de réduction et d'analyse de données automatisés**. Les récentes avancées sur les lignes de faisceaux de neutrons, notamment la ligne D22 de l'ILL, rendront très bientôt possible de **coupler le TR-SANS à des mesures SAXS ou DLS**. Des informations sur l'ensemble des protéines en solution compléteront les informations sélectives apportées par la deutération d'une seule protéine ou sous-unité couplée à la variation de contraste. Des **systèmes d'injection automatisés** (de type « *stop flow* ») pourront également permettre de déclencher de manière mieux contrôlée les réactions biochimiques avec certains composants clés, de renouveler certaines réactions ou au contraire d'en figer d'autres aux étapes clés de certains processus moléculaires.

# Références

1. Schoenheimer, R. The dynamic state of body constituents. *The dynamic state of body constituents*. (1946).
2. Wolf, D. H. & Messen, R. Mechanisms of cell regulation – proteolysis, the big surprise. *FEBS Lett* **592**, 2515–2524 (2018).
3. Hogness, D. S., Cohn, M. & Monod, J. Studies on the induced synthesis of  $\beta$ -galactosidase in *Escherichia coli*: The kinetics and mechanism of sulfur incorporation. *Biochimica et Biophysica Acta* **16**, 99–116 (1955).
4. de Duve, C. Lysosomes, a new group of cytoplasmic particles. *Subcellular particles* **60**, 128–159 (1959).
5. Ciechanover, A. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* **6**, 79–87 (2005).
6. Wilkinson, K. D., Urban, M. K. & Haas, A. L. Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J. Biol. Chem.* **255**, 7529–7532 (1980).
7. Morange, M. What history tells us XLI. Ubiquitin and proteolysis. *J Biosci* **41**, 325–329 (2016).
8. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).
9. Groll, M., Bochtler, M., Brandstetter, H., Clausen, T. & Huber, R. Molecular Machines for Protein Degradation. *ChemBioChem* **6**, 222–256 (2005).
10. Yedidi, R. S., Wendler, P. & Enenkel, C. AAA-ATPases in Protein Degradation. *Front. Mol. Biosci.* **4**, (2017).
11. Alberts, B., Wilson, J. H. & Hunt, T. *Molecular Biology of the Cell* (Garland Science, New York). *USA*. 1601p (2008).
12. Demers, J.-P., Fricke, P., Shi, C., Chevelkov, V. & Lange, A. Structure determination of supra-molecular assemblies by solid-state NMR: Practical considerations. *Progress in Nuclear Magnetic Resonance Spectroscopy* **109**, 51–78 (2018).
13. Schanda, P. Relaxing with liquids and solids - A perspective on biomolecular dynamics. *J Magn Reson* **306**, 180–186 (2019).
14. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019).
15. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat Methods* **13**, 24–27 (2016).
16. Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018).
17. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450–457 (2015).
18. Glaeser, R. M. How good can cryo-EM become? *Nat Methods* **13**, 28–32 (2016).
19. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nature Chemistry* **10**, 31–37 (2018).
20. Coquelle, N. *et al.* Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Cryst D* **71**, 1184–1196 (2015).
21. Lehmkuhler, F. *et al.* Dynamics of soft nanoparticle suspensions at hard X-ray FEL sources below the radiation-damage threshold. *IUCr* **5**, 801–807 (2018).
22. Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995–1014 (2014).
23. Brutscher, B. *et al.* NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. in *Intrinsically Disordered Proteins Studied by NMR Spectroscopy* (eds. Felli, I. C. & Pierattelli, R.) vol. 870 49–122 (Springer International Publishing, 2015).
24. Delaforge, E. *et al.* Investigating the Role of Large-Scale Domain Dynamics in Protein-Protein Interactions. *Front Mol Biosci* **3**, 54 (2016).

25. Marchanka, A., Kreutz, C. & Carlomagno, T. Isotope labeling for studying RNA by solid-state NMR spectroscopy. *J. Biomol. NMR* **71**, 151–164 (2018).
26. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T<sub>2</sub> relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94**, 12366–12371 (1997).
27. Riek, R., Pervushin, K. & Wüthrich, K. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends in Biochemical Sciences* **25**, 462–468 (2000).
28. Riek, R., Wider, G., Pervushin, K. & Wuthrich, K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences* **96**, 4918–4923 (1999).
29. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced <sup>1</sup>H–<sup>13</sup>C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).
30. Kay, L. E. Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view. *Journal of Magnetic Resonance* **210**, 159–170 (2011).
31. Quinn, C. M. & Polenova, T. Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Quarterly Reviews of Biophysics* **50**, (2017).
32. Martin, R. W. & Zilm, K. W. Preparation of protein nanocrystals and their characterization by solid state NMR. *Journal of Magnetic Resonance* **165**, 162–174 (2003).
33. Bertini, I., Luchinat, C., Parigi, G. & Ravera, E. SedNMR: On the Edge between Solution and Solid-State NMR. *Acc. Chem. Res.* **46**, 2059–2069 (2013).
34. Bertini, I. *et al.* Solid-state NMR of proteins sedimented by ultracentrifugation. *PNAS* **108**, 10396–10399 (2011).
35. Gauto, D. F. *et al.* Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat Commun* **10**, 1–12 (2019).
36. Schubeis, T., Le Marchand, T., Andreas, L. B. & Pintacuda, G. <sup>1</sup>H magic-angle spinning NMR evolves as a powerful new tool for membrane proteins. *Journal of Magnetic Resonance* **287**, 140–152 (2018).
37. Mahieu, E. & Gabel, F. Biological small-angle neutron scattering: recent results and development. *Acta Crystallogr D Struct Biol* **74**, 715–726 (2018).
38. Cerofolini, L. *et al.* Integrative Approaches in Structural Biology: A More Complete Picture from the Combination of Individual Techniques. *Biomolecules* **9**, 370 (2019).
39. Appolaire, A. *et al.* Pyrococcus horikoshii TET2 Peptidase Assembling Process and Associated Functional Regulation. *J. Biol. Chem.* **288**, 22542–22554 (2013).
40. Ibrahim, Z. *et al.* Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. *Sci Rep* **7**, 40948 (2017).
41. Whittaker, R. H. New Concepts of Kingdoms of Organisms. *Science* **163**, 150–160 (1969).
42. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088–5090 (1977).
43. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS* **87**, 4576–4579 (1990).
44. Forterre, P., Brochier, C. & Philippe, H. Evolution of the Archaea. *Theoretical Population Biology* **61**, 409–422 (2002).
45. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1007–1022 (2006).
46. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *PNAS* **112**, 6670–6675 (2015).
47. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**, 245–252 (2008).

48. Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**, 3204–3223 (2011).
49. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* **105**, 8102–8107 (2008).
50. Guy, L. & Ettema, T. J. G. The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends in Microbiology* **19**, 580–587 (2011).
51. Huber, H., Hohn, M. J., Stetter, K. O. & Rachel, R. The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Research in Microbiology* **154**, 165–171 (2003).
52. Petitjean, C., Deschamps, P., López-García, P. & Moreira, D. Rooting the Domain Archaea by Phylogenomic Analysis Supports the Foundation of the New Kingdom Proteoarchaeota. *Genome Biol Evol* **7**, 191–204 (2014).
53. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).
54. Kozubal, M. A. *et al.* Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J* **7**, 622–634 (2013).
55. Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
56. Vanwonterghem, I. *et al.* Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* **1**, 1–9 (2016).
57. Castelle, C. J. *et al.* Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. *Current Biology* **25**, 690–701 (2015).
58. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
59. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
60. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
61. Seitz, K. W. *et al.* New Asgard archaea capable of anaerobic hydrocarbon cycling. *bioRxiv* 527697 (2019) doi:10.1101/527697.
62. Gribaldo, S. & Brochier-Armanet, C. Time for order in microbial systematics. *Trends in Microbiology* **20**, 209–210 (2012).
63. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *PNAS* **86**, 9355–9359 (1989).
64. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *PNAS* **81**, 3786–3790 (1984).
65. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
66. Kelly, J. J., Policht, K., Grancharova, T. & Hundal, L. S. Distinct Responses in Ammonia-Oxidizing Archaea and Bacteria after Addition of Biosolids to an Agricultural Soil. *Appl. Environ. Microbiol.* **77**, 6551–6558 (2011).
67. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).
68. Williams, T. A. & Embley, T. M. Archaeal “Dark Matter” and the Origin of Eukaryotes. *Genome Biol Evol* **6**, 474–481 (2014).
69. Lasek-Nesselquist, E. & Gogarten, J. P. The effects of model choice and mitigating bias on the ribosomal tree of life. *Molecular Phylogenetics and Evolution* **69**, 17–38 (2013).
70. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* **15**, 711–723 (2017).

71. Forterre, P. The Common Ancestor of Archaea and Eukarya Was Not an Archaeon. *Archaea* **2013**, (2013).
72. Cunha, V. D., Gaia, M., Gabelle, D., Nasir, A. & Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLOS Genetics* **13**, e1006810 (2017).
73. Cunha, V. D., Gaia, M., Nasir, A. & Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genetics* **14**, e1007215 (2018).
74. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *bioRxiv* 726976 (2019) doi:10.1101/726976.
75. Canganella, F. & Wiegel, J. Extremophiles: from abyssal to terrestrial ecosystems and possibly beyond. *Naturwissenschaften* **98**, 253–279 (2011).
76. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2001).
77. Zobell, C. E. & Morita, R. Y. BAROPHILIC BACTERIA IN SOME DEEP SEA SEDIMENTS1. *J Bacteriol* **73**, 563–568 (1957).
78. Aguilar, C. F. *et al.* Crystal structure of the  $\beta$ -glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability11Edited by R. Huber. *Journal of Molecular Biology* **271**, 789–802 (1997).
79. Antranikian, G., Vorgias, C. E. & Bertoldo, C. Extreme Environments as a Resource for Microorganisms and Novel Biocatalysts. in *Marine Biotechnology I* (eds. Ulber, R. & Le Gal, Y.) 219–262 (Springer Berlin Heidelberg, 2005). doi:10.1007/b135786.
80. Rampelotto, P. H. Resistance of Microorganisms to Extreme Environmental Conditions and Its Contribution to Astrobiology. *Sustainability* **2**, 1602–1623 (2010).
81. Leuko, S., Raftery, M. J., Burns, B. P., Walter, M. R. & Neilan, B. A. Global Protein-Level Responses of *Halobacterium salinarum* NRC-1 to Prolonged Changes in External Sodium Chloride Concentrations. *J. Proteome Res.* **8**, 2218–2225 (2009).
82. Kottemann, M., Kish, A., Iloanusi, C., Bjork, S. & DiRuggiero, J. Physiological responses of the halophilic archaeon *Halobacterium* sp. strain NRC1 to desiccation and gamma irradiation. *Extremophiles* **9**, 219–227 (2005).
83. Takai, K. & Horikoshi, K. Genetic Diversity of Archaea in Deep-Sea Hydrothermal Vent Environments. *Genetics* **152**, 1285–1297 (1999).
84. Daffonchio, D. *et al.* Stratified prokaryote network in the oxic–anoxic transition of a deep-sea halocline. *Nature* **440**, 203–207 (2006).
85. Andrei, A.-Ş., Banciu, H. L. & Oren, A. Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol Lett* **330**, 1–9 (2012).
86. Stetter, K. O. Hyperthermophiles in the history of life. *Phil. Trans. R. Soc. B* **361**, 1837–1843 (2006).
87. Stetter, K. O. *et al.* *Methanothermus fervidus*, sp. nov., a novel extremely thermophilic methanogen isolated from an Icelandic hot spring. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie* **2**, 166–178 (1981).
88. Vieille, C. & Zeikus, G. J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews* **65**, 1–43 (2001).
89. Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. Were the original eubacteria thermophiles? *Systematic and Applied Microbiology* **9**, 34–39 (1987).
90. Fiala, G. & Stetter, K. O. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* **145**, 56–61 (1986).
91. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001).
92. Magrum, L. J., Luehrsen, K. R. & Woese, C. R. Are extreme halophiles actually “bacteria”? *J Mol Evol* **11**, 1–8 (1978).

93. Ciaramella, M., Napoli, A. & Rossi, M. Another extreme genome: how to live at pH 0. *Trends in Microbiology* **13**, 49–51 (2005).
94. Reeve, J. N. & Schleper, C. Archaea: very diverse, often different but never bad? *Curr Opin Microbiol* **14**, 271–273 (2011).
95. Haegeman, B. *et al.* Robust estimation of microbial diversity in theory and in practice. *ISME J* **7**, 1092–1101 (2013).
96. DeLong, E. F. Everything in moderation: Archaea as ‘non-extremophiles’. *Current Opinion in Genetics & Development* **8**, 649–654 (1998).
97. Lepage, P. *et al.* A metagenomic insight into our gut’s microbiome. *Gut* **62**, 146–158 (2013).
98. Miller, T. L., Wolin, M. J., Macario, E. C. de & Macario, A. J. Isolation of *Methanobrevibacter smithii* from human feces. *Appl. Environ. Microbiol.* **43**, 227–232 (1982).
99. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biology* **12**, R50 (2011).
100. Probst, A. J., Auerbach, A. K. & Moissl-Eichinger, C. Archaea on Human Skin. *PLOS ONE* **8**, e65388 (2013).
101. Ferry, J. G. & Kestead, K. A. Methanogenesis. *Archaea* 288–314 (2007) doi:10.1128/9781555815516.ch13.
102. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**, 2407–2425 (2017).
103. Eckburg, P. B., Lepp, P. W. & Relman, D. A. Archaea and Their Potential Role in Human Disease. *Infection and Immunity* **71**, 591–596 (2003).
104. Gaci, N., Borrel, G., Tottey, W., O’Toole, P. W. & Brugère, J.-F. Archaea and the human gut: New beginning of an old story. *World J Gastroenterol* **20**, 16062–16078 (2014).
105. Castelle, C. J. & Banfield, J. F. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* **172**, 1181–1197 (2018).
106. Vianna, M. E., Conrads, G., Gomes, B. P. F. A. & Horz, H. P. T-RFLP-based *mcrA* gene analysis of methanogenic archaea in association with oral infections and evidence of a novel *Methanobrevibacter* phylotype. *Oral Microbiology and Immunology* **24**, 417–422 (2009).
107. Bang, C. & Schmitz, R. A. Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol Rev* **39**, 631–648 (2015).
108. Dridi, B., Raoult, D. & Drancourt, M. Archaea as emerging organisms in complex human microbiomes. *Anaerobe* **17**, 56–63 (2011).
109. Matarazzo, F. *et al.* The domain Archaea in human mucosal surfaces. *Clinical Microbiology and Infection* **18**, 834–840 (2012).
110. DeLong, E. F. & Pace, N. R. Environmental Diversity of Bacteria and Archaea. *Syst Biol* **50**, 470–478 (2001).
111. Hong, J.-K. & Cho, J.-C. Environmental Variables Shaping the Ecological Niche of Thaumarchaeota in Soil: Direct and Indirect Causal Effects. *PLOS ONE* **10**, e0133763 (2015).
112. Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**, 254–261 (1983).
113. Miller, J. F., Shah, N. N., Nelson, C. M., Ludlow, J. M. & Clark, D. S. Pressure and Temperature Effects on Growth and Methane Production of the Extreme Thermophile *Methanococcus jannaschii*. *Appl. Environ. Microbiol.* **54**, 3039–3042 (1988).
114. Graham, D. E., Kyrpides, N., Anderson, I. J., Overbeek, R. & Whitman, W. B. Genome of *Methanocaldococcus (methanococcus) jannaschii*. in *Methods in Enzymology* vol. 330 40–123 (Academic Press, 2001).
115. König, H., Rachel, R. & Claus, H. Proteinaceous Surface Layers of Archaea: Ultrastructure and Biochemistry. *Archaea* 315–340 (2007) doi:10.1128/9781555815516.ch14.

116. L. S., Joan & W, F., John. *Microbiology: An Evolving Science: Third International Student Edition*. (W. W. Norton & Company, 2013).
117. Bult, C. J. *et al.* Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
118. Tehei, M. & Zaccai, G. Adaptation to high temperatures through macromolecular dynamics by neutron scattering: Dynamic basis of thermoadaptation. *FEBS Journal* **274**, 4034–4043 (2007).
119. Zaccai, G. Ecology of Protein Dynamics. <https://www.ingentaconnect.com/content/ben/cpc/2013/00000003/00000001/art00004> (2013).
120. Davies, G. J., Gamblin, S. J., Littlechild, J. A. & Watson, H. C. The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins: Structure, Function, and Bioinformatics* **15**, 283–289 (1993).
121. Auerbach, G. *et al.* Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Structure* **5**, 1475–1483 (1997).
122. Chi, Y.-I. *et al.* Crystal structure of the  $\beta$ -glycosidase from the hyperthermophile *Thermosphaera aggregans*: insights into its activity and thermostability. *FEBS Letters* **445**, 375–383 (1999).
123. Hollien, J. & Marqusee, S. Structural distribution of stability in a thermophilic enzyme. *PNAS* **96**, 13674–13678 (1999).
124. Bauer, M. W., Driskill, L. E. & Kelly, R. M. Glycosyl hydrolases from hyperthermophilic microorganisms. *Current Opinion in Biotechnology* **9**, 141–145 (1998).
125. Zwickl, P., Fabry, S., Bogedain, C., Haas, A. & Hensel, R. Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. *Journal of Bacteriology* **172**, 4329–4338 (1990).
126. Han, H. *et al.* Improvements of thermophilic enzymes: From genetic modifications to applications. *Bioresource Technology* **279**, 350–361 (2019).
127. Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys.: Condens. Matter* **22**, 323101 (2010).
128. Irimia, A. *et al.* The 2.9 Å Resolution Crystal Structure of Malate Dehydrogenase from *Archaeoglobus fulgidus*: Mechanisms of Oligomerisation and Thermal Stabilisation. *Journal of Molecular Biology* **335**, 343–356 (2004).
129. Karlström, M. *Structure determination, thermal stability and catalytic mechanism of hyperthermostable isocitrate dehydrogenases*. (Biovetenskaper och näringslära / Biosciences and Nutrition, 2006).
130. Koutsopoulos, S., Oost, J. van der & Norde, W. Temperature-dependent structural and functional features of a hyperthermostable enzyme using elastic neutron scattering. *Proteins: Structure, Function, and Bioinformatics* **61**, 377–384 (2005).
131. Haney, P. J. *et al.* Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *PNAS* **96**, 3578–3583 (1999).
132. Hatanaka, H., Tanimura, R., Katoh, S. & Inagaki, F. Solution structure of ferredoxin from the thermophilic Cyanobacterium *Synechococcus elongatus* and its thermostability<sup>11</sup> Edited by P.E. Wright. *Journal of Molecular Biology* **268**, 922–933 (1997).
133. Wallon, G. *et al.* Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*<sup>11</sup> Edited by A. R. Fersht. *Journal of Molecular Biology* **266**, 1016–1031 (1997).
134. Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. & Rees, D. C. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463–1469 (1995).
135. Rice, D. W. *et al.* Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol Rev* **18**, 105–117 (1996).
136. Yip, K. *et al.* The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147–1158 (1995).

137. Kumar, S., Ma, B., Tsai, C.-J. & Nussinov, R. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins: Structure, Function, and Bioinformatics* **38**, 368–383 (2000).
138. Szilágyi, A. & Závodszy, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493–504 (2000).
139. Elcock, A. H. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins<sup>11</sup> Edited by B. Honig. *Journal of Molecular Biology* **284**, 489–502 (1998).
140. Feller, G. & Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1**, 200–208 (2003).
141. Choi, I.-G., Bang, W.-G., Kim, S.-H. & Yu, Y. G. Extremely Thermostable Serine-type Protease from *Aquifex pyrophilus* MOLECULAR CLONING, EXPRESSION, AND CHARACTERIZATION. *J. Biol. Chem.* **274**, 881–888 (1999).
142. Kumar, S. & Nussinov, R. How do thermophilic proteins deal with heat? *CMLS, Cell. Mol. Life Sci.* **58**, 1216–1233 (2001).
143. Trivedi, S., Gehlot, H. S. & Rao, S. R. Protein thermostability in Archaea and Eubacteria. *Genetics and Molecular Research* **12** (2006).
144. Tanaka, Y. *et al.* How Oligomerization Contributes to the Thermostability of an Archaeon Protein PROTEIN L-ISOASPARTYL-O-METHYLTRANSFERASE FROM *SULFOLOBUS TOKODAI*. *J. Biol. Chem.* **279**, 32957–32967 (2004).
145. Suhre, K. & Claverie, J.-M. Genomic Correlates of Hyperthermostability, an Update. *J. Biol. Chem.* **278**, 17198–17202 (2003).
146. Lieph, R., Veloso, F. A. & Holmes, D. S. Thermophiles like hot T. *Trends in Microbiology* **14**, 423–426 (2006).
147. Hashimoto, H. *et al.* Hyperthermostable Protein Structure Maintained by Intra and Inter-helix Ion-pairs in Archaeal O<sup>6</sup>-Methylguanine-DNA Methyltransferase. *Journal of Molecular Biology* **292**, 707–716 (1999).
148. Jorda, J. & Yeates, T. O. Widespread Disulfide Bonding in Proteins from Thermophilic Archaea. *Archaea* **2011**, (2011).
149. Thompson, M. J. & Eisenberg, D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability<sup>1</sup> Edited by I. B. Honig. *Journal of Molecular Biology* **290**, 595–604 (1999).
150. Yamagata, Y. *et al.* Entropic Stabilization of the Tryptophan Synthase  $\alpha$ -Subunit from a Hyperthermophile, *Pyrococcus furiosus* X-RAY ANALYSIS AND CALORIMETRY. *J. Biol. Chem.* **276**, 11062–11071 (2001).
151. Unsworth, L. D., Oost, J. van der & Koutsopoulos, S. Hyperthermophilic enzymes – stability, activity and implementation strategies for high temperature applications. *The FEBS Journal* **274**, 4044–4056 (2007).
152. Ramos, E., Lopez, T., Bosch, P., Asomoza, M. & Gomez, R. Thermal Stability of Sol-Gel Hydrotalcites. *Journal of Sol-Gel Science and Technology* **8**, 437–442 (1997).
153. Santos, H. & Costa, M. S. D. Compatible solutes of organisms that live in hot saline environments. *Environmental Microbiology* **4**, 501–509 (2002).
154. Wirth, R. Colonization of Black Smokers by Hyperthermophilic Microorganisms. *Trends in Microbiology* **25**, 92–99 (2017).
155. Boonyaratankornkit, B. B. *et al.* Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock. *Environmental Microbiology* **7**, 789–797 (2005).
156. Maupin-Furlow, J. Proteasomes and protein conjugation across domains of life. *Nat Rev Microbiol* **10**, 100–111 (2012).
157. Franzetti, B., Schoehn, G., Garcia, D., Ruigrok, R. W. H. & Zaccari, G. Characterization of the proteasome from the extremely halophilic archaeon *Haloarcula marismortui*. *Archaea* **1**, 53–61 (2002).

158. Somero, G. N. Proteins and Temperature. *Annual Review of Physiology* **57**, 43–68 (1995).
159. Tehei, M. *et al.* Adaptation to extreme environments: macromolecular dynamics in bacteria compared in vivo by neutron scattering. *EMBO reports* **5**, 66–70 (2004).
160. Habbeche, A. *et al.* Purification and biochemical characterization of a detergent-stable keratinase from a newly thermophilic actinomycete *Actinomyces keratinolyticus* strain Cpt29 isolated from poultry compost. *Journal of Bioscience and Bioengineering* **117**, 413–421 (2014).
161. Kudriaeva, A. A. & Belogurov, A. A. Proteasome: a Nanomachinery of Creative Destruction. *Biochemistry Mosc.* **84**, S159–S192 (2019).
162. Goldberg, A. L. & Dice, J. F. Intracellular Protein Degradation in Mammalian and Bacterial Cells. *Annu. Rev. Biochem.* **43**, 835–869 (1974).
163. Goldberg, A. L. & St. John, A. C. Intracellular Protein Degradation in Mammalian and Bacterial Cells: Part 2. *Annu. Rev. Biochem.* **45**, 747–804 (1976).
164. Glickman, M. H. Getting in and out of the proteasome. *Seminars in Cell & Developmental Biology* **11**, 149–158 (2000).
165. Maupin-Furlow, J., Wilson, H., Kaczowka, S. & Ou, M. Proteasomes in the archaea: from structure to function. *Frontiers in bioscience : a journal and virtual library* **5**, D837-65 (2000).
166. Schimke, R. T., Ganschow, R., Doyle, D. & Arias, I. M. Regulation of protein turnover in mammalian tissues. *Fed Proc* **27**, 1223–1230 (1968).
167. Ciechanover, A. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* **6**, 79–87 (2005).
168. Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29**, 15–32 (2001).
169. Klemes, Y., Etlinger, J. D. & Goldberg, A. L. Properties of abnormal proteins degraded rapidly in reticulocytes. Intracellular aggregation of the globin molecules prior to hydrolysis. *J. Biol. Chem.* **256**, 8436–8444 (1981).
170. Hartl, F. U. Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science* **295**, 1852–1858 (2002).
171. Frydman, J. Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annual Review of Biochemistry* **70**, 603–647 (2001).
172. Glover, J. R. & Lindquist, S. Hsp104, Hsp70, and Hsp40: A Novel Chaperone System that Rescues Previously Aggregated Proteins. *Cell* **94**, 73–82 (1998).
173. VerPlank, J. J. S., Lokireddy, S., Feltri, M. L., Goldberg, A. L. & Wrabetz, L. Impairment of protein degradation and proteasome function in hereditary neuropathies. *Glia* **66**, 379–395 (2018).
174. Barrett, A. J. & McDonald, J. K. Nomenclature: protease, proteinase and peptidase. *Biochem J* **237**, 935 (1986).
175. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat Rev Microbiol* **14**, 33–44 (2016).
176. Kruger, E., Witt, E., Ohlmeier, S., Hanschke, R. & Hecker, M. The Clp Proteases of *Bacillus subtilis* Are Directly Involved in Degradation of Misfolded Proteins. *Journal of Bacteriology* **182**, 3259–3265 (2000).
177. Hecker, M., Schumann, W. & Völker, U. Heat-shock and general stress response in *Bacillus subtilis*. *Molecular Microbiology* **19**, 417–428 (1996).
178. Duque-Magalhães, M. C. La protéolyse intracellulaire. *Biochimie* **66**, 653–662 (1984).
179. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology* **10**, 524–530 (2000).
180. Ciechanover, A. & Brundin, P. The Ubiquitin Proteasome System in Neurodegenerative Diseases: Sometimes the Chicken, Sometimes the Egg. *Neuron* **40**, 427–446 (2003).
181. Schmidt, M. & Finley, D. Regulation of proteasome activity in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 13–25 (2014).

182. Sauer, R. T. & Baker, T. A. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
183. Tanaka, K. & Ichihara, A. Half-life of proteasomes (multiprotease complexes) in rat liver. *Biochemical and Biophysical Research Communications* **159**, 1309–1315 (1989).
184. Schmidt, M. *et al.* Sequence information within proteasomal prosequences mediates efficient integration of  $\beta$ -subunits into the 20 S proteasome complex. Edited by R. Huber. *Journal of Molecular Biology* **288**, 117–128 (1999).
185. Lowe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539 (1995).
186. Seemüller, E. *et al.* Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* **268**, 579–582 (1995).
187. Seemüller, E., Lupas, A. & Baumeister, W. Autocatalytic processing of the 20S proteasome. *Nature* **382**, 468–470 (1996).
188. Rabl, J. *et al.* Mechanism of Gate Opening in the 20S Proteasome by the Proteasomal ATPases. *Molecular Cell* **30**, 360–368 (2008).
189. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. The Proteasome: Paradigm of a Self-Compartmentalizing Protease. *Cell* **92**, 367–380 (1998).
190. Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annual Review of Biochemistry* **78**, 477–513 (2009).
191. Bohn, S. *et al.* Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *PNAS* **107**, 20992–20997 (2010).
192. Förster, F., Lasker, K., Nickell, S., Sali, A. & Baumeister, W. Toward an Integrated Structural Model of the 26S Proteasome. *Molecular & Cellular Proteomics* **9**, 1666–1677 (2010).
193. Bard, J. A. M. *et al.* Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **87**, 697–724 (2018).
194. He, J. *et al.* The Structure of the 26S Proteasome Subunit Rpn2 Reveals Its PC Repeat Domain as a Closed Toroid of Two Concentric  $\alpha$ -Helical Rings. *Structure* **20**, 513–521 (2012).
195. Shi, Y. *et al.* Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**, aad9421 (2016).
196. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences* **109**, 1380–1387 (2012).
197. Gorbea, C., Taillandier, D. & Rechsteiner, M. Mapping Subunit Contacts in the Regulatory Complex of the 26 S Proteasome: S2 AND S5b FORM A TETRAMER WITH ATPase SUBUNITS S4 and S7. *Journal of Biological Chemistry* **275**, 875–882 (2000).
198. Richmond, C., Gorbea, C. & Rechsteiner, M. Specific Interactions between ATPase Subunits of the 26 S Protease. *J. Biol. Chem.* **272**, 13403–13411 (1997).
199. Inobe, T. & Genmei, R. N-Terminal Coiled-Coil Structure of ATPase Subunits of 26S Proteasome Is Crucial for Proteasome Function. *PLoS ONE* **10**, e0134056 (2015).
200. Rubin, D. M., Glickman, M. H., Larsen, C. N., Dhruvakumar, S. & Finley, D. Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome. *The EMBO Journal* **17**, 4909–4919 (1998).
201. Zhu, Y. *et al.* Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nat Commun* **9**, 1360 (2018).
202. Katz, E. J., Isasa, M. & Crosas, B. A new map to understand deubiquitination. *Biochem Soc Trans* **38**, 21–28 (2010).
203. Nijman, S. M. B. *et al.* A Genomic and Functional Inventory of Deubiquitinating Enzymes. *Cell* **123**, 773–786 (2005).
204. Li, S.-J. & Hochstrasser, M. The Yeast ULP2 (SMT4) Gene Encodes a Novel Protease Specific for the Ubiquitin-Like Smt3 Protein. *Molecular and Cellular Biology* **20**, 2367–2377 (2000).

205. Komander, D. Mechanism, Specificity and Structure of the Deubiquitinases. in *Conjugation and Deconjugation of Ubiquitin Family Modifiers: Subcellular Biochemistry* (ed. Groettrup, M.) 69–87 (Springer New York, 2010). doi:10.1007/978-1-4419-6676-6\_6.
206. Lee, B.-H. *et al.* USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature* **532**, 398–401 (2016).
207. Hamazaki, J. *et al.* A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *The EMBO Journal* **25**, 4524–4536 (2006).
208. Qiu, X.-B. *et al.* hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *The EMBO Journal* **25**, 5742–5753 (2006).
209. Husnjak, K. & Dikic, I. Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions. *Annual Review of Biochemistry* **81**, 291–322 (2012).
210. Hershko, A. Ubiquitin-mediated protein degradation. *J. Biol. Chem.* **263**, 15237–15240 (1988).
211. Pickart, C. M. & Cohen, R. E. Proteasomes and their kin: proteases in the machine age. *Nat Rev Mol Cell Biol* **5**, 177–187 (2004).
212. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
213. Ravid, T. & Hochstrasser, M. Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. *Nat Cell Biol* **9**, 422–427 (2007).
214. Hershko, A. & Ciechanover, A. The Ubiquitin System. *Annual Review of Biochemistry* **67**, 425–479 (1998).
215. Hipp, M. S., Kalveram, B., Raasi, S., Groettrup, M. & Schmidtke, G. FAT10, a Ubiquitin-Independent Signal for Proteasomal Degradation. *Molecular and Cellular Biology* **25**, 3483–3491 (2005).
216. van der Veen, A. G. & Ploegh, H. L. Ubiquitin-Like Proteins. *Annual Review of Biochemistry* **81**, 323–357 (2012).
217. Hoyt, M. A. & Coffino, P. Ubiquitin-free routes into the proteasome. *Cell. Mol. Life Sci.* **61**, 1596–1600 (2004).
218. Chen, T. *et al.* Clusterin-Mediated Apoptosis Is Regulated by Adenomatous Polyposis Coli and Is p21 Dependent but p53 Independent. *Cancer Res* **64**, 7412–7419 (2004).
219. DeLaBarre, B. & Brunger, A. T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat Struct Mol Biol* **10**, 856–863 (2003).
220. Esaki, M., Johjima-Murata, A., Islam, Md. T. & Ogura, T. Biological and Pathological Implications of an Alternative ATP-Powered Proteasomal Assembly With Cdc48 and the 20S Peptidase. *Front. Mol. Biosci.* **5**, 56 (2018).
221. Wolf, D. H. & Stolz, A. The Cdc48 machine in endoplasmic reticulum associated protein degradation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 117–124 (2012).
222. Barthelme, D., Jauregui, R. & Sauer, R. T. An ALS disease mutation in Cdc48/p97 impairs 20S proteasome binding and proteolytic communication: Cdc48 ALS Disease Mutation. *Protein Science* **24**, 1521–1527 (2015).
223. Barthelme, D. & Sauer, R. T. Identification of the Cdc48bullet20S Proteasome as an Ancient AAA+ Proteolytic Machine. *Science* **337**, 843–846 (2012).
224. Rothballer, A., Tzvetkov, N. & Zwickl, P. Mutations in p97/VCP induce unfolding activity. *FEBS Letters* **581**, 1197–1201 (2007).
225. Müller, A. U. & Weber-Ban, E. The Bacterial Proteasome at the Core of Diverse Degradation Pathways. *Front. Mol. Biosci.* **6**, 23 (2019).
226. Gur, E., Biran, D. & Ron, E. Z. Regulated proteolysis in Gram-negative bacteria — how and when? *Nat Rev Microbiol* **9**, 839–848 (2011).
227. Chung, C. H. & Goldberg, A. L. The product of the lon (capR) gene in Escherichia coli is the ATP-dependent protease, protease La. *PNAS* **78**, 4931–4935 (1981).

228. Lee, I. & Suzuki, C. K. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1784**, 727–735 (2008).
229. Striebel, F., Kress, W. & Weber-Ban, E. Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Current Opinion in Structural Biology* **19**, 209–217 (2009).
230. De Mot, R., Nagy, I., Walz, J. & Baumeister, W. Proteasomes and other self-compartmentalizing proteases in prokaryotes. *Trends in Microbiology* **7**, 88–92 (1999).
231. Unciuleac, M.-C., Smith, P. C. & Shuman, S. Crystal Structure and Biochemical Characterization of a Mycobacterium smegmatis AAA-Type Nucleoside Triphosphatase Phosphohydrolase (Msm0858). *Journal of Bacteriology* **198**, 1521–1533 (2016).
232. Ziemski, M. *et al.* Cdc48-like protein of actinobacteria (Cpa) is a novel proteasome interactor in mycobacteria and related organisms. *eLife* **7**, e34055 (2018).
233. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 2–14 (2012).
234. Ito, K. & Akiyama, Y. Cellular Functions, Mechanism of Action, and Regulation of Ftsh Protease. *Annual Review of Microbiology* **59**, 211–231 (2005).
235. Gerdes, F., Tatsuta, T. & Langer, T. Mitochondrial AAA proteases — Towards a molecular understanding of membrane-bound proteolytic machines. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 49–55 (2012).
236. Gur, E. The Lon AAA+ Protease. in *Regulated Proteolysis in Microorganisms* (ed. Dougan, D. A.) 35–51 (Springer Netherlands, 2013). doi:10.1007/978-94-007-5940-4\_2.
237. Jastrab, J. B. & Darwin, K. H. Bacterial Proteasomes. *Annual Review of Microbiology* **69**, 109–127 (2015).
238. Wang, T., Darwin, K. H. & Li, H. Binding-induced folding of prokaryotic ubiquitin-like protein on the Mycobacterium proteasomal ATPase targets substrates for degradation. *Nat Struct Mol Biol* **17**, 1352–1357 (2010).
239. Özcelik, D. *et al.* Structures of Pup ligase PafA and depupylase Dop from the prokaryotic ubiquitin-like modification pathway. *Nat Commun* **3**, 1–10 (2012).
240. Flynn, J. M., Neher, S. B., Kim, Y.-I., Sauer, R. T. & Baker, T. A. Proteomic Discovery of Cellular Substrates of the ClpXP Protease Reveals Five Classes of ClpX-Recognition Signals. *Molecular Cell* **11**, 671–683 (2003).
241. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development* **12**, 1338–1347 (1998).
242. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. *Science, New Series* **271**, 990–993 (1996).
243. Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *PNAS* **91**, 9223–9227 (1994).
244. Koodathingal, P. *et al.* ATP-dependent Proteases Differ Substantially in Their Ability to Unfold Globular Proteins. *J. Biol. Chem.* **284**, 18674–18684 (2009).
245. Gur, E. & Sauer, R. T. Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes Dev.* **22**, 2267–2277 (2008).
246. Maupin-Furlow, J. A. *et al.* Proteasomes from Structure to Function: Perspectives from Archaea. in *Current Topics in Developmental Biology* vol. 75 125–169 (Academic Press, 2006).
247. Bar-Nun, S. & Glickman, M. H. Proteasomal AAA-ATPases: Structure and function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 67–82 (2012).
248. Forouzan, D. *et al.* The Archaeal Proteasome Is Regulated by a Network of AAA ATPases. *J. Biol. Chem.* **287**, 39254–39262 (2012).

249. Rockel, B., Jakana, J., Chiu, W. & Baumeister, W. Electron cryo-microscopy of VAT, the archaeal p97/CDC48 homologue from *Thermoplasma acidophilum* 11 Edited by D. Rees. *Journal of Molecular Biology* **317**, 673–681 (2002).
250. Benaroudj, N., Zwickl, P., Seemüller, E., Baumeister, W. & Goldberg, A. L. ATP Hydrolysis by the Proteasome Regulatory Complex PAN Serves Multiple Functions in Protein Degradation. *Molecular Cell* **11**, 69–78 (2003).
251. Humbard, M. A. & Maupin-Furlow, J. A. Prokaryotic Proteasomes: Nanocompartments of Degradation. *MMB* **23**, 321–334 (2013).
252. Djuranovic, S., Rockel, B., Lupas, A. N. & Martin, J. Characterization of AMA, a new AAA protein from *Archaeoglobus* and methanogenic archaea. *Journal of Structural Biology* **156**, 130–138 (2006).
253. Reuter, C. J., Kaczowka, S. J. & Maupin-Furlow, J. A. Differential Regulation of the PanA and PanB Proteasome-Activating Nucleotidase and 20S Proteasomal Proteins of the Haloarchaeon *Haloferax volcanii*. *Journal of Bacteriology* **186**, 7763–7772 (2004).
254. Chamieh, H., Guetta, D. & Franzetti, B. The two PAN ATPases from *Halobacterium* display N-terminal heterogeneity and form labile complexes with the 20S proteasome. *Biochem. J.* **411**, 387–397 (2008).
255. Besche, H. & Zwickl, P. The *Thermoplasma acidophilum* Lon protease has a Ser-Lys dyad active site. *European Journal of Biochemistry* **271**, 4361–4365 (2004).
256. Humbard, M. A. *et al.* Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature* **463**, 54–60 (2010).
257. Miranda, H. V. *et al.* E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *PNAS* **108**, 4417–4422 (2011).
258. Maupin-Furlow, J. A. Archaeal Proteasomes and Sampylation. in *Regulated Proteolysis in Microorganisms* (ed. Dougan, D. A.) vol. 66 297–327 (Springer Netherlands, 2013).
259. Hepowit, N. L. *et al.* Archaeal JAB1/MPN/MOV34 metalloenzyme (HvJAMM1) cleaves ubiquitin-like small archaeal modifier proteins (SAMPs) from protein-conjugates. *Molecular Microbiology* **86**, 971–987 (2012).
260. Anjum, R. S. *et al.* Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon *Sulfolobus acidocaldarius*. *Nat Commun* **6**, 8163 (2015).
261. Benaroudj, N. & Goldberg, A. L. PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nat Cell Biol* **2**, 833–839 (2000).
262. Zhang, F. *et al.* Mechanism of Substrate Unfolding and Translocation by the Regulatory Particle of the Proteasome from *Methanocaldococcus jannaschii*. *Molecular Cell* **34**, 485–496 (2009).
263. Burroughs, A. M., Jaffee, M., Iyer, L. M. & Aravind, L. Anatomy of the E2 ligase fold: Implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *Journal of Structural Biology* **162**, 205–218 (2008).
264. Fuchs, A. C. D., Maldoner, L., Wojtynek, M., Hartmann, M. D. & Martin, J. Rpn11-mediated ubiquitin processing in an ancestral archaeal ubiquitination system. *Nat Commun* **9**, 2696 (2018).
265. James, R. H. *et al.* Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. *Nat Commun* **8**, 1–15 (2017).
266. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291–294 (1998).
267. Piccolino, M. Biological machines: from mills to molecules. *Nat Rev Mol Cell Biol* **1**, 149–152 (2000).
268. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology* **146**, 11–31 (2004).
269. Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure-diverse function. *Genes Cells* **6**, 575–597 (2001).
270. Miller, J. M. & Enemark, E. J. Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea* **2016**, 1–12 (2016).

271. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. *Genome Res.* **9**, 27–43 (1999).
272. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**, 519–529 (2005).
273. Erzberger, J. P. & Berger, J. M. EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
274. Smith, G. R., Contreras-Moreira, B., Zhang, X. & Bates, P. A. A link between sequence conservation and domain motion within the AAA+ family. *Journal of Structural Biology* **146**, 189–204 (2004).
275. Kessel, M. *et al.* Homology in Structural Organization Between *E. coli* ClpAP Protease and the Eukaryotic 26 S Proteasome. *Journal of Molecular Biology* **250**, 587–594 (1995).
276. Zwickl, P. *et al.* An Archaeobacterial ATPase, Homologous to ATPases in the Eukaryotic 26 S Proteasome, Activates Protein Breakdown by 20 S Proteasomes. *J. Biol. Chem.* **274**, 26008–26014 (1999).
277. Wilson, H. L., Ou, M. S., Aldrich, H. C. & Maupin-Furlow, J. Biochemical and Physical Properties of the *Methanococcus jannaschii* 20S Proteasome and PAN, a Homolog of the ATPase (Rpt) Subunits of the Eucaryal 26S Proteasome. *Journal of Bacteriology* **182**, 1680–1692 (2000).
278. Djuranovic, S. *et al.* Structure and Activity of the N-Terminal Substrate Recognition Domains in Proteasomal ATPases. *Molecular Cell* **34**, 580–590 (2009).
279. Voges, D., Zwickl, P. & Baumeister, W. The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. *Annual Review of Biochemistry* **68**, 1015–1068 (1999).
280. Benaroudj, N., Smith, D. & Goldberg, A. L. What the Archaeal PAN–Proteasome Complex and Bacterial ATP-Dependent Proteases Can Teach Us About the 26S Proteasome. in *Protein Degradation Series* 215–247 (John Wiley & Sons, Ltd, 2007). doi:10.1002/9783527619320.ch10a.
281. Forster, A. The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *The EMBO Journal* **22**, 4356–4364 (2003).
282. Smith, D. M. *et al.* ATP Binding to PAN or the 26S ATPases Causes Association with the 20S Proteasome, Gate Opening, and Translocation of Unfolded Proteins. *Molecular Cell* **20**, 687–698 (2005).
283. Smith, D. M., Chang, S.-C., Park, S., Finley, D. & Cheng, Y. Docking of the Proteasomal ATPases' C-termini in the 20S Proteasomes alpha Ring Opens the Gate for Substrate Entry. **30** (2008).
284. Barthelme, D. & Sauer, R. T. Bipartite determinants mediate an evolutionarily conserved interaction between Cdc48 and the 20 S peptidase. *Proc Natl Acad Sci USA* **110**, 3327–3332 (2013).
285. Barthelme, D., Chen, J. Z., Grabenstatter, J., Baker, T. A. & Sauer, R. T. Architecture and assembly of the archaeal Cdc48\*20S proteasome. *Proceedings of the National Academy of Sciences* **111**, E1687–E1694 (2014).
286. Majumder, P. *et al.* Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proc Natl Acad Sci U S A* **116**, 534–539 (2019).
287. Yu, Y. *et al.* Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome–ATPase interactions. *The EMBO Journal* **29**, 692–702 (2010).
288. Koga, N., Kameda, T., Okazaki, K. & Takada, S. Paddling mechanism for the substrate translocation by AAA+ motor revealed by multiscale molecular simulations. *PNAS* **106**, 18237–18242 (2009).
289. Martin, A., Baker, T. A. & Sauer, R. T. Diverse Pore Loops of the AAA+ ClpX Machine Mediate Unassisted and Adaptor-Dependent Recognition of *ssrA*-Tagged Substrates. *Molecular Cell* **29**, 441–450 (2008).
290. Matouschek, A. Protein unfolding — an important process in vivo? *Current Opinion in Structural Biology* **13**, 98–109 (2003).
291. Zhang, F. *et al.* Structural Insights into the Regulatory Particle of the Proteasome from *Methanocaldococcus jannaschii*. *Molecular Cell* **34**, 473–484 (2009).
292. Matyskiela, M. E., Lander, G. C. & Martin, A. Conformational switching of the 26S proteasome enables substrate degradation. *Nat Struct Mol Biol* **20**, 781–788 (2013).

293. Nyquist, K. & Martin, A. Marching to the beat of the ring: polypeptide translocation by AAA+ proteases. *Trends in Biochemical Sciences* **39**, 53–60 (2014).
294. Schweitzer, A. *et al.* Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc Natl Acad Sci USA* **113**, 7816–7821 (2016).
295. Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *PNAS* **111**, 5544–5549 (2014).
296. Śledź, P. *et al.* Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *PNAS* **110**, 7264–7269 (2013).
297. Wenzel, T. & Baumeister, W. Conformational constraints in protein degradation by the 20S proteasome. *Nat Struct Mol Biol* **2**, 199–204 (1995).
298. Unno, M. *et al.* The Structure of the Mammalian 20S Proteasome at 2.75 Å Resolution. *Structure* **10**, 609–618 (2002).
299. Bajorek, M., Finley, D. & Glickman, M. H. Proteasome Disassembly and Downregulation Is Correlated with Viability during Stationary Phase. *Current Biology* **13**, 1140–1144 (2003).
300. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR. *Science* **328**, 98–102 (2010).
301. Stadtmueller, B. M. & Hill, C. P. Proteasome Activators. *Molecular Cell* **41**, 8–19 (2011).
302. Eisele, M. R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell Reports* **24**, 1301–1315.e5 (2018).
303. Wehmer, M. *et al.* Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc Natl Acad Sci USA* **114**, 1305–1310 (2017).
304. Ding, Z. *et al.* High-resolution cryo-EM structure of the proteasome in complex with ADP-AlFx. *Cell Res* **27**, 373–385 (2017).
305. Peña, A. H. de la, Goodall, E. A., Gates, S. N., Lander, G. C. & Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, eaav0725 (2018).
306. Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49–55 (2019).
307. Beck, F. *et al.* Near-atomic resolution structural model of the yeast 26S proteasome. *PNAS* **109**, 14870–14875 (2012).
308. Gates, S. N. *et al.* Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273–279 (2017).
309. Su, M. *et al.* Mechanism of Vps4 hexamer function revealed by cryo-EM. *SCIENCE ADVANCES* **8** (2017).
310. Puchades, C. *et al.* Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358**, eaao0464 (2017).
311. Kim, Y.-C., Snoberger, A., Schupp, J. & Smith, D. M. ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. *Nat Commun* **6**, 8520 (2015).
312. Cordova, J. C. *et al.* Stochastic but Highly Coordinated Protein Unfolding and Translocation by the ClpXP Proteolytic Machine. *Cell* **158**, 647–658 (2014).
313. Smith, D. M., Fraga, H., Reis, C., Kafri, G. & Goldberg, A. L. ATP Binds to Proteasomal ATPases in Pairs with Distinct Functional Effects, Implying an Ordered Reaction Cycle. *Cell* **144**, 526–538 (2011).
314. Banerjee, S. *et al.* 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875 (2016).
315. Martin, A., Baker, T. A. & Sauer, R. T. Distinct Static and Dynamic Interactions Control ATPase-Peptidase Communication in a AAA+ Protease. *Molecular Cell* **27**, 41–52 (2007).
316. Kenniston, J. A., Baker, T. A. & Sauer, R. T. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *PNAS* **102**, 1390–1395 (2005).

317. Navon, A. & Goldberg, A. L. Proteins Are Unfolded on the Surface of the ATPase Ring before Transport into the Proteasome. *Molecular Cell* **8**, 1339–1349 (2001).
318. Aubin-Tam, M.-E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell* **145**, 257–267 (2011).
319. Maillard, R. A. *et al.* ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates. *Cell* **145**, 459–469 (2011).
320. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine. *Cell* **114**, 511–520 (2003).
321. Peterson, C. N., Levchenko, I., Rabinowitz, J. D., Baker, T. A. & Silhavy, T. J. RpoS proteolysis is controlled directly by ATP levels in Escherichia coli. *Genes Dev.* **26**, 548–553 (2012).
322. Nager, A. R., Baker, T. A. & Sauer, R. T. Stepwise Unfolding of a  $\beta$  Barrel Protein by the AAA+ ClpXP Protease. *Journal of Molecular Biology* **413**, 4–16 (2011).
323. Iosefson, O., Nager, A. R., Baker, T. A. & Sauer, R. T. Coordinated gripping of substrate by subunits of a AAA+ proteolytic machine. *Nature Chemical Biology* **11**, 201–206 (2015).
324. Vass, R. H. & Chien, P. Critical clamp loader processing by an essential AAA+ protease in *Caulobacter crescentus*. *PNAS* **110**, 18138–18143 (2013).
325. Svergun, D. I., Koch, M. H., Timmins, P. A. & May, R. P. *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. vol. 19 (Oxford University Press, 2013).
326. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
327. Haertlein, M. *et al.* Biomolecular deuteration for neutron structural biology and dynamics. in *Methods in enzymology* vol. 566 113–157 (Elsevier, 2016).
328. Feigin, L. A. & Svergun, D. I. *Structure analysis by small-angle X-ray and neutron scattering*. vol. 1 (Springer, 1987).
329. Jacrot, B. The study of biological structures by neutron scattering from solution. *Reports on progress in physics* **39**, 911 (1976).
330. Zaccai, G. & Jacrot, B. Small Angle Neutron Scattering. *Annual Review of Biophysics and Bioengineering* **12**, 139–157 (1983).
331. Bizien, T. *et al.* A Brief Survey of State-of-the-Art BioSAXS. <https://www.ingentaconnect.com/content/ben/ppl/2016/00000023/00000003/art00006> (2016).
332. Brosey, C. A. & Tainer, J. A. Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current Opinion in Structural Biology* **58**, 197–213 (2019).
333. Guinier, A., Fournet, G. & Yudowitch, K. L. Small-angle scattering of X-rays. (1955).
334. Glatter, O. & Kratky, O. *Small angle X-ray scattering*. (Academic press, 1982).
335. Kaler, E. W. & Brumberger, H. In Modern Aspects of Small-Angle Scattering. *Kluwer Academic, Dordrecht* **140**, 329 (1995).
336. Sivia, D. S. *Elementary scattering theory: for X-ray and neutron users*. (Oxford University Press, 2011).
337. Carlile, C. J. *Experimental Neutron Scattering: BTM Willis. Carlile*. (Oxford University Press, 2013).
338. Jacrot, B. & Zaccai, G. Determination of molecular weight by neutron scattering. *Biopolymers: Original Research on Biomolecules* **20**, 2413–2426 (1981).
339. Guinier, A. La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. in *Annales de physique* vol. 11 161–237 (EDP Sciences, 1939).
340. Porod, G. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. *Colloid & Polymer Science* **124**, 83–114 (1951).
341. Glatter, O. A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallography* **10**, 415–421 (1977).

342. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212–1225 (2017).
343. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography* **36**, 1277–1282 (2003).
344. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography* **42**, 342–346 (2009).
345. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal* **76**, 2879–2886 (1999).
346. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography* **25**, 495–503 (1992).
347. Svergun, D., Barberato, C. & Koch, M. H. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography* **28**, 768–773 (1995).
348. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical journal* **89**, 1237–1250 (2005).
349. Jordan, A. *et al.* SEC-SANS: size exclusion chromatography combined in situ with small-angle neutron scattering. *Journal of applied crystallography* **49**, 2015–2020 (2016).
350. Harroun, T. A., Wignall, G. D. & Katsaras, J. Neutron scattering for biology. in *Neutron Scattering in Biology* 1–18 (Springer, 2006).
351. Stuhrmann, H. B. Neutron small-angle scattering of biological macromolecules in solution. *Journal of applied crystallography* **7**, 173–178 (1974).
352. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **189**, 113–130 (1986).
353. Dunne, O. *et al.* Matchout deuterium labelling of proteins for small-angle neutron scattering studies using prokaryotic and eukaryotic expression systems and high cell-density cultures. *European Biophysics Journal* **46**, 425–432 (2017).
354. Xu, B., Jahic, M. & Enfors, S.-O. Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of Escherichiacoli. *Biotechnology progress* **15**, 81–90 (1999).
355. Meilleur, F., Weiss, K. L. & Myles, D. A. Deuterium labeling for neutron structure-function-dynamics analysis. in *Micro and Nano Technologies in Bioanalysis* 281–292 (Springer, 2009).
356. Yakhnin, A. V., Vinokurov, L. M., Surin, A. K. & Alakhov, Y. B. Green fluorescent protein purification by organic extraction. *Protein expression and purification* **14**, 382–386 (1998).
357. Howarth, M. & Ting, A. Y. Monovalent streptavidin expression and purification. *Nat. Protoc* (2008).
358. Howarth, M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3**, 267–273 (2006).
359. Michaelis, L. & Menten, M. L. The kinetics of the inversion effect. *Biochem. Z* **49**, 333–369 (1913).
360. Itaya, K. & Ui, M. A new micromethod for the colorimetric determination of inorganic phosphate. *Clinica chimica acta* **14**, 361–366 (1966).
361. Penney, C. L. A simple micro-assay for inorganic phosphate. *Anal. Biochem.* **75**, 201–210 (1976).
362. Wilson, K. & Walker, J. *Principles and techniques of biochemistry and molecular biology*. (Cambridge university press, 2010).
363. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of synchrotron radiation* **20**, 660–664 (2013).
364. Dewhurst, C. D. GRASP: Graphical Reduction and Analysis SANS Program. *ILL Report ILL03DE01T, Grenoble, France* (2003).
365. Richard, D., Ferrand, M. & Kearley, G. J. Analysis and visualisation of neutron-scattering data. *Journal of Neutron Research* **4**, 33–39 (1996).

366. Zimm, B. H. The scattering of light and the radial distribution function of high polymer solutions. *The Journal of Chemical Physics* **16**, 1093–1099 (1948).
367. Zhao, H., Brown, P. H. & Schuck, P. On the distribution of protein refractive index increments. *Biophys. J.* **100**, 2309–2317 (2011).
368. Frisken, B. J. Revisiting the method of cumulants for the analysis of dynamic light-scattering data. *Appl Opt* **40**, 4087–4091 (2001).
369. Lau, W. C. & Rubinstein, J. L. Single particle electron microscopy. in *Electron Crystallography of Soluble and Membrane Proteins* 401–426 (Springer, 2013).
370. Capecchi, M. R. Initiation of E. coli proteins. *Proc Natl Acad Sci U S A* **55**, 1517–1524 (1966).
371. Panfair, D., Ramamurthy, A. & Kusmierczyk, A. R. Alpha-ring Independent Assembly of the 20S Proteasome. *Sci Rep* **5**, (2015).
372. Panfair, D. & Kusmierczyk, A. R. Examining Proteasome Assembly with Recombinant Archaeal Proteasomes and Nondenaturing PAGE: The Case for a Combined Approach. *J Vis Exp* (2016) doi:10.3791/54860.
373. Smith, D. M., Benaroudj, N. & Goldberg, A. Proteasomes and their associated ATPases: A destructive combination. *Journal of Structural Biology* **156**, 72–83 (2006).
374. Demchenko, A. P. *Advanced Fluorescence Reporters in Chemistry and Biology I: Fundamentals and Molecular Design*. (Springer Science & Business Media, 2010).
375. Berardozzi, R. Etude photophysique des protéines fluorescentes photoconvertibles utilisées en microscopie de super-résolution. (Grenoble Alpes, 2016).
376. Dunst, S. & Tomancak, P. Imaging Flies by Fluorescence Microscopy: Principles, Technologies, and Applications. *Genetics* **211**, 15–34 (2019).
377. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* **59**, 223–239 (1962).
378. THE GREEN FLUORESCENT PROTEIN | Annual Review of Biochemistry. <https://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.67.1.509>.
379. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
380. Phillips, G. N. Structure and dynamics of green fluorescent protein. *Current Opinion in Structural Biology* **7**, 821–827 (1997).
381. Palmer, E. & Freeman, T. Investigation Into the use of C- and N-terminal GFP Fusion Proteins for Subcellular Localization Studies Using Reverse Transfection Microarrays. *Comp Funct Genomics* **5**, 342–353 (2004).
382. Siemering, K. R., Golbik, R., Sever, R. & Haseloff, J. Mutations that suppress the thermosensitivity of green fluorescent protein. *Current Biology* **6**, 1653–1663 (1996).
383. Bokman, S. H. & Ward, W. W. Renaturation of Aequorea green-fluorescent protein. *Biochemical and Biophysical Research Communications* **101**, 1372–1380 (1981).
384. Yang, F., Moss, L. G. & Phillips, G. N. The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246–1251 (1996).
385. Ward, W. W. Energy Transfer Processes in Bioluminescence. in *Photochemical and Photobiological Reviews: Volume 4* (ed. Smith, K. C.) 1–57 (Springer US, 1979). doi:10.1007/978-1-4684-3551-1\_1.
386. Forbes, M. W. & Jockusch, R. A. Deactivation Pathways of an Isolated Green Fluorescent Protein Model Chromophore Studied by Electronic Action Spectroscopy. *J. Am. Chem. Soc.* **131**, 17038–17039 (2009).
387. Webber, N. M., Litvinenko, K. L. & Meech, S. R. Radiationless Relaxation in a Synthetic Analogue of the Green Fluorescent Protein Chromophore. *J. Phys. Chem. B* **105**, 8036–8039 (2001).
388. S. Mooney, C. R. *et al.* Taking the green fluorescence out of the protein : dynamics of the isolated GFP chromophore anion. *Chemical Science* **4**, 921–927 (2013).

389. Weber-Ban, E. U., Reid, B. G., Miranker, A. D. & Horwich, A. L. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90–93 (1999).
390. Benaroudj, N., Tarcsa, E., Cascio, P. & Goldberg, A. L. The unfolding of substrates and ubiquitin-independent protein degradation by proteasomes. *Biochimie* **83**, 311–318 (2001).
391. Gitlin, G., Bayer, E. A. & Wilchek, M. Studies on the biotin-binding site of avidin. Tryptophan residues involved in the active site. *Biochem J* **250**, 291–294 (1988).
392. Zwickl, P. *et al.* An Archaeobacterial ATPase, Homologous to ATPases in the Eukaryotic 26 S Proteasome, Activates Protein Breakdown by 20 S Proteasomes. *J. Biol. Chem.* **274**, 26008–26014 (1999).
393. Snoberger, A., Anderson, R. T. & Smith, D. M. The Proteasomal ATPases Use a Slow but Highly Processive Strategy to Unfold Proteins. *Front Mol Biosci* **4**, (2017).
394. Wold, F. [1] Affinity labeling—An overview. in *Methods in Enzymology* vol. 46 3–14 (Academic Press, 1977).
395. Yaginuma, H. *et al.* Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Scientific Reports* **4**, 6522 (2014).
396. Kim, Y.-C., Snoberger, A., Schupp, J. & Smith, D. M. ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. *Nat Commun* **6**, 1–13 (2015).
397. Zemb, T. & Lindner, P. *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. (Elsevier, 2002).
398. de la Peña, A. H., Goodall, E. A., Gates, S. N., Lander, G. C. & Martin, A. Substrate-engaged 26 S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, eaav0725 (2018).

## Résumé

Les protéasomes sont de grands assemblages macromoléculaires ubiquitaires composés d'un complexe catalytique 20S et d'une particule régulatrice comprenant un module AAA-ATPase. Cette machine cellulaire est chargée de dégrader sélectivement les protéines intracellulaires pour permettre le renouvellement du protéome, éliminer les protéines défectueuses et contrôler de nombreuses fonctions biologiques. Le travail de cette thèse avait pour objectif de mettre à jour les mécanismes qui permettent aux complexes AAA-ATPase de déplier sélectivement les protéines substrat et de les transférer à la particule 20S, dans laquelle elles sont détruites. Pour cela, une approche novatrice a été utilisée en combinant la diffusion de neutrons aux petits angles résolue en temps (TR-SANS) avec la spectroscopie de fluorescence permettant de suivre l'activité biochimique. Le protéasome de l'archée hyperthermophile *Methanocaldococcus jannaschii*, a été utilisé comme système modèle. Il est composé de la protéase 20S et de la particule régulatrice AAA-ATPase PAN. Un variant de la protéine fluorescente GFP a été utilisé comme protéine substrat.

Les données obtenues montrent que l'activité de dépliement de PAN génère des formes de GFP dénaturée formant des agrégats. En revanche, l'association avec la particule 20S prévient la formation de ces espèces et indique qu'une fois le dépliement d'une protéine par PAN engagé, les processus de transfert dans le complexe 20S et de dégradation sont étroitement couplés. L'analyse des spectres de diffusion neutronique du substrat GFP montrent que la population de GFP native disparaît rapidement au profit des peptides générés par la protéase 20S, comme confirmé par une analyse en spectrométrie de masse. Cela démontre le caractère hautement processif du protéasome. Enfin, deux modes d'action de PAN ont été mis en évidence selon la quantité de protéines à dégrader par rapport au protéasome PAN-20S. Ces travaux permettent de valider expérimentalement un des modèles de fonctionnement du protéasome préalablement proposés et soulignent l'importance d'un contrôle de l'association des protéasomes *in vivo*. Cette étude met également en valeur l'intérêt de la technique TR-SANS pour étudier la dynamique fonctionnelle de grandes machines cellulaires.

Mots-clés : protéolyse, protéasome, grands assemblages moléculaires, diffusion de neutrons aux petits angles résolue en temps.

## Abstract

Proteasomes are large ubiquitous macromolecular assemblies composed of a 20S catalytic complex and a regulatory particle containing an AAA-ATPase module. This cellular machine is responsible for selectively degradation of intracellular proteins in order to allow proteome renewal, elimination of defective proteins and control of many biological functions. The objective of this thesis was to reveal the mechanisms by which the AAA-ATPase complexes selectively unfold substrate proteins and translocate them into the 20S particle, where they are destroyed. To this end, an innovative approach was used by combining time-resolved small-angle neutron scattering (TR-SANS) combined with fluorescence spectroscopy to monitor biochemical activity. The proteasome of the hyperthermophilic archaea *Methanocaldococcus jannaschii* was used as a model system. It is composed of the 20S protease of the regulatory particle AAA-ATPase PAN. A variant of the fluorescent protein GFP was used as a substrate protein.

The data obtained show that PAN unfolding activity generates denatured species of GFP forming aggregates. The association with the 20S particle prevents the formation of these species and indicates that once the unfolding of a substrate by PAN is engaged, translocation into the 20S complex and degradation processes are closely coupled. Analysis of the neutron scattering spectra of the GFP substrate reveal that the native GFP population is rapidly disappearing in favor of peptides generated by the 20S protease, as confirmed by mass spectrometric analysis. This demonstrates the highly processive nature of the proteasome. Finally, two modes of action of PAN have been identified depending on the amount of proteins to be degraded compared to the PAN-20S proteasome. This work allows to experimentally validate one of the proteasome function models previously proposed and emphasizes the importance of controlling the association of proteasomes *in vivo*. This study also highlights the interest of TR-SANS technique to study the functional dynamics of large cellular machines.

Key words: proteolysis, proteasome, large molecular assemblies, time resolved small-angle neutron scattering.