



HAL
open science

Etude de la structure et du mécanisme d'action du complexe unfoldase PAN, un activateur du protéasome 20S

Ziad Ibrahim

► **To cite this version:**

Ziad Ibrahim. Etude de la structure et du mécanisme d'action du complexe unfoldase PAN, un activateur du protéasome 20S. Biochimie, Biologie Moléculaire. Université Grenoble Alpes, 2016. Français. NNT : 2016GREAY027 . tel-01616844

HAL Id: tel-01616844

<https://theses.hal.science/tel-01616844v1>

Submitted on 15 Oct 2017

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 COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

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

Arrêté ministériel : 7 août 2006

Présentée par

Ziad IBRAHIM

Thèse dirigée par **Frank GABEL** et
codirigée par **Michael HAERTLEIN**

préparée au sein de l'**Institut de Biologie Structurale** et de
l'**Institut Laue-Langevin**
dans l'**École Doctorale de Physique**

Etude de la structure et du mécanisme d'action du complexe unfoldase PAN, un activateur du protéasome 20S

Thèse soutenue publiquement le **Judi 14 Avril 2016**,
devant le jury composé de :

Dr. Vincent VILLERET (Rapporteur)

(Institut de Recherche Interdisciplinaire, Lille)

Dr. Olivier COUX (Rapporteur)

(Centre de Recherche de Biochimie Macromoléculaire, Montpellier)

Prof. Eva PEBAY-PEYROULA (Examinatrice)

(Institut de Biologie Structurale, Grenoble)

Dr. Javier PEREZ (Examinateur)

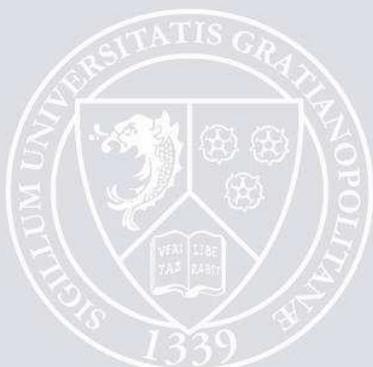
(Synchrotron SOLEIL, Paris)

Dr. Frank GABEL (Directeur de thèse)

(Institut de Biologie Structurale et Institut Laue-Langevin, Grenoble)

Dr. Michael HAERTLEIN (Co-Directeur de thèse)

(Institut Laue-Langevin, Grenoble)



REMERCIEMENTS

Ma thèse s'est déroulée à l'institut de biologie structurale (IBS) et à l'institut Laue Langevin (ILL) de Grenoble. Je tiens à remercier Bruno Franzetti directeur de l'équipe ELMA à l'IBS pour l'opportunité qu'il m'a offert de travailler dans son équipe et pour m'avoir donné les moyens pour effectuer mes travaux ainsi que pour toutes les discussions pendant mes années de thèses qui m'ont beaucoup aidé à bien effectuer mon travail et à avoir des résultats satisfaisants.

Mes remerciements les plus chaleureux vont à Frank Gabel mon directeur de thèse qui m'a accueilli sans me connaître et m'a ainsi offert l'occasion de faire mes preuves. Tout au long de ces trois ans et demi il a suivi mon travail avec beaucoup d'attention et m'a permis de m'ouvrir au monde de la recherche. Je tiens également à remercier Michael Haertlein et Susana Teixeira, mes co-encadrants de l'ILL qui ont suivi l'évolution de mes travaux au jour le jour. Leur apport à ce travail fut très important et leur aide a été également cruciale dans l'aboutissement des différents aspects de l'étude. Je tiens à remercier également Matteo Colombo, un ancien postdoc de l'équipe ELMA qui m'a accordé beaucoup de son temps au début de ma thèse pour m'initier aux différents aspects du projet, notre travail ensemble pendant deux ans a conduit à la première publication de ma thèse.

Beaucoup d'autres personnes ont permis à cette thèse d'aboutir en me permettant d'accéder à leurs laboratoires et leurs instruments et en participant activement à des expériences. Je tiens ici à les remercier. Il s'agit de : Trevor Forsyth, le directeur de laboratoire de deutération à l'ILL ; Martine Moulin pour son grand support et son aide pendant la deutération de mes protéines à l'ILL ; Anne Martel, responsable de la ligne D22 à l'ILL pour sa grande contribution dans les manips SANS en particulier dans la partie développement instrumentale ; Guy Schoehn du groupe microscopie électronique de l'IBS comme collaborateur principal dans le projet surtout dans la partie concernant les études structurales par Cryo-EM ; Daphna Fenel de la plateforme microscopie électronique ; Adam Round et Martha Brennich de la ligne de lumière BM29 à l'ESRF, Grenoble ; Didier Flament du centre Ifremer de Brest ; Nicole Thielens, responsable de la plateforme SPR de l'IBS ; Henry Kim pour son aide dans les expériences SPR ; Emilie Mahieu pour son aide pendant son stage de master 1 dans la préparation des échantillons pour les expériences SANS ; et tous les autres

membres de l'équipe ELMA pour toutes les discussions scientifiques précieuses. Leur bonne humeur et leur gentillesse resteront d'excellents souvenirs de ma thèse.

Une grande reconnaissance va à Olivier Coux et Vincent Villeret qui ont accepté d'être rapporteur de mon manuscrit, leurs commentaires pertinents m'ont donné une nouvelle vision de mon travail. Je remercie également Eva Pebay-Peyroula qui a présidé le jury de soutenance et Javier Perez pour accepter d'être membre examinateur de mon jury.

Un petit clin d'œil à Elodie Henriet, nous nous sommes rencontrés pendant ma thèse et nous avons créé une histoire qui pourrait durer une vie.

A Clara Najem pour les beaux moments que nous avons passé ensemble à Grenoble et à la grande chaleur d'amitié que nous avons partagé.

A tous mes amis de Grenoble avec lesquels j'ai passé des moments inoubliables.

Enfin je remercie ici mes parents, Joséphine et Ibrahim, ainsi que mes frères, Jad et Nehmé et toute la famille. Sans les efforts de mes parents en particulier, rien de ce que j'ai accompli n'aurait été possible. Je leur témoigne ici tout mon amour.

Etude de la structure et du mécanisme d'action du complexe unfoldase PAN, un activateur du protéasome 20S

RESUME

La dégradation intracellulaire des protéines est un processus fondamental qui a lieu dans tous les organismes, depuis les bactéries jusqu'aux êtres humains. Cette dégradation est nécessaire pour réguler les concentrations intracellulaires d'enzymes qui contrôlent toutes les réactions métaboliques, ainsi que le contenu général de toutes les autres protéines, en réponse aux modifications physiologiques. Un état d'équilibre dynamique entre synthèse et dégradation se crée ainsi dans lequel la concentration intracellulaire d'une protéine peut être modulée, notamment par l'acteur principal de l'élimination des protéines intracellulaires dans les trois domaines de la vie, le protéasome.

Le travail présenté dans cette thèse porte sur le complexe d'activation du protéasome des cellules d'archées (PAN). PAN est un complexe hexamérique énergie-dépendant découvert chez les archées et impliqué dans le dépliement des protéines substrat pour faciliter leur dégradation par le protéasome 20S. Il fait part de la famille des AAA+ (« ATPases Associated with diverse cellular Activities ») unfoldases. A ce jour, les études du complexe PAN assemblé n'ont pas réussi à révéler sa structure à cause des difficultés rencontrées au niveau de la préparation et la stabilité des échantillons.

La première partie de mon travail de thèse a permis de déterminer une structure du complexe PAN hexamérique de *Pyrococcus horikoshii* par cryo-microscopie électronique et un modèle pseudo-atomique. De plus, l'étude des différents états conformationnels du complexe induits par la liaison du nucléotide ont permis de gagner plusieurs informations sur le mécanisme des unfoldases AAA+ et de proposer un mode d'action du complexe PAN et de son interaction avec le protéasome 20S. La seconde partie de ma thèse a été dédiée à élucider des questions portant sur la dynamique et les changements conformationnels des systèmes AAA+ unfoldases en général et du complexe PAN en particulier pendant le dépliement des protéines substrats. La méthode de variation de contraste en diffusion de neutrons aux petits angles (SANS), couplée avec de la spectroscopie de fluorescence en ligne, a été appliquée à l'étude en temps réel du processus de dépliement d'une protéine, une variante de la GFP (« Green Fluorescent Protein ») par le système PAN de *Methanococcus jannaschii* à l'échelle de quelques secondes. Cette approche innovatrice a permis de révéler, pour la première fois, des mouvements de contraction de PAN pendant le dépliement du substrat induits par l'hydrolyse

de l'ATP et suivis par une relaxation de la molécule à la fin du processus. Le mécanisme de dépliement par PAN semble être un mouvement de pompage péristaltique qui entraîne un dépliement directionnel du substrat.

L'ensemble de mes résultats de thèse contribue à mieux comprendre la structure et le mécanisme d'action de la machine PAN dans les cellules et la dynamique fonctionnelle des complexes AAA+ à l'origine de leurs fonctions biologiques. Finalement, ce travail montre que l'approche SANS en temps réel, couplée à la fluorescence en ligne, développée pendant ma thèse, pourra être appliquée pour l'étude des changements conformationnels (de l'ordre de quelques secondes) d'un grand nombre d'autres machineries cellulaires en solution.

ABSTRACT

Intracellular protein degradation is a fundamental process occurring in all organisms, from bacteria to humans. The degradation of proteins is necessary for regulating the intracellular levels of enzymes that control all metabolic reactions, as well as the general content of all other proteins, in response to physiological changes. A state of dynamic equilibrium between synthesis and degradation is thus created wherein the intracellular concentration of a protein can be modulated, in particular by the proteasome which is a key player in the elimination of intracellular proteins in the three domains of life.

The work presented in this thesis deals with the proteasome activating complex from archaeal cells (PAN). PAN is an energy-dependent hexameric complex discovered in archaea and involved in the unfolding of protein substrates to facilitate their degradation by the 20S proteasome. It belongs to the AAA+ (« ATPases Associated with diverse cellular Activities ») unfoldases protein family. All previous structural studies on the assembled PAN complex have failed in revealing the structure of the whole complex because of the difficulties encountered during sample preparation and stabilization.

In the first part of this work we determined a cryo-electron microscopy structure and a pseudo-atomic model of the hexameric PAN complex from *Pyrococcus horikoshii*. In addition, the study of the different conformational states of the PAN complex induced by nucleotide binding helped in gaining precious information about the AAA+ unfoldases mechanism and to propose a mode of action of the PAN complex and its interaction with the 20S proteasome. The second part of the study led to elucidate the question about the dynamic and the conformational changes of the AAA+ unfoldases in general and the PAN complex in particular during the unfolding of substrate proteins. The method of contrast variation in Small Angle Neutron Scattering (SANS), coupled with online fluorescence spectroscopy, was applied to study, in real-time, the unfolding process of a substrate protein, a variant of the GFP (Green Fluorescent Protein) by the PAN complex from *Methanococcus jannaschii* with a time scale of few seconds. This innovative approach allowed to reveal, for the first time, contraction movements of PAN during substrate unfolding induced by ATP hydrolysis and followed by a relaxation of the molecule at the end of the process. The unfolding mechanism processed by PAN appears to be a peristaltic pumping motion that leads to a directional unfolding of the substrate.

The whole work presented in this thesis contributes to understand the structure and the mechanism of action of the PAN molecular machine inside cells and to have a clearer idea

about the dynamics of the AAA+ complexes at the origin of their biological functions. Finally, this study shows that the real-time SANS approach, coupled with online fluorescence, and developed during my thesis, can be applied to study conformational changes (of the order of few seconds) of a large number of other cellular machineries in solution.

TABLE DES MATIERES

| | | |
|----------|--|----|
| 1 | INTRODUCTION | 1 |
| 1.1 | Avant-propos..... | 1 |
| 1.2 | Rôles de la dégradation des protéines et importance du système Ubiquitine-Protéasome | 3 |
| 1.3 | Intérêt des archées hyperthermophiles pour l'étude du système protéolytique .. | 7 |
| 1.3.1 | Liens des archées avec les eucaryotes..... | 7 |
| 1.3.2 | Intérêt des systèmes issus des archées hyperthermophiles pour l'étude des grands assemblages de la protéolyse..... | 9 |
| 1.3.3 | Présentation des archées d'origine des protéines étudiées dans la thèse | 12 |
| 1.4 | Présentation des systèmes protéasomes archéens et eucaryotes | 16 |
| 1.4.1 | Description générale (eucaryotes/archées)..... | 16 |
| 1.4.2 | Les complexes 20S (eucaryotes/archées)..... | 19 |
| 1.4.3 | Les systèmes d'adressage aux protéasomes (ubiquitination et sampylation)..... | 21 |
| 1.4.4 | Les complexes de régulation du protéasome..... | 26 |
| 1.5 | Structure et fonction des AAA+ protéolytiques régulatrices du protéasome | 31 |
| 1.5.1 | Diversité des AAA+ | 31 |
| 1.5.2 | Les AAA+ ATPases protéolytiques | 35 |
| 1.5.3 | Structure des AAA+ ATPases protéolytiques | 41 |
| 1.5.4 | Mode d'action des unfoldases AAA+ | 45 |
| 1.6 | Le complexe AAA+ PAN des archées : une version simplifiée du 19S des eucaryotes | 48 |
| 1.7 | Problématiques et objectifs de la thèse | 55 |
| 1.7.1 | Objectif 1: étude structurale et fonctionnelle du complexe PAN hexamérique . | 57 |
| 1.7.2 | Objectif 2 : développement d'une approche de diffusion de neutrons aux petits angles (SANS) en temps réel pour étudier le processus de dépliement des protéines par les complexes AAA+ ATPases | 58 |
| 2 | DIFFUSION DE NEUTRONS AUX PETITS ANGLES (SANS) | 59 |
| 2.1 | Théorie | 59 |
| 2.1.1 | Le neutron | 59 |
| 2.1.2 | Interaction neutron-noyau et section efficace | 60 |

| | | |
|------------|--|-----|
| 2.2 | Diffusion de neutrons aux petits angles en mode statique et variation de contraste | 63 |
| 2.3 | Diffusion de neutrons aux petits angles en temps réel et combinaison avec la spectroscopie de fluorescence en ligne | 71 |
| 2.3.1 | Instrumentation | 71 |
| 2.3.2 | Démarche expérimentale | 72 |
| 2.3.3 | Réduction et interprétation des données | 78 |
| 2.4 | Intérêt de la technique SANS dans le cas de notre projet | 79 |
| 3 | MATERIELS ET METHODES BIOPHYSIQUES | 82 |
| 3.1 | Résonance Plasmon de Surface (SPR) | 82 |
| 3.1.1 | Théorie | 82 |
| 3.1.2 | Démarche expérimentale | 85 |
| 3.1.3 | Interprétation des données | 86 |
| 3.2 | Microscopie électronique | 86 |
| 3.2.1 | Coloration négative | 86 |
| 3.2.2 | Cryo-microscopie électronique | 86 |
| 3.2.3 | Analyse des données microscopie électronique | 87 |
| 3.3 | Diffusion de rayons X aux petits angles (SAXS) | 88 |
| 3.3.1 | Démarche expérimentale | 88 |
| 3.3.2 | Interprétation des données | 89 |
| 4 | MATERIELS ET METHODES BIOCHIMIQUES | 90 |
| 4.1 | Expression et purification des protéines hydrogénées | 90 |
| 4.1.1 | Synthèse des gènes, clonage et transformation | 90 |
| 4.1.2 | Expression | 91 |
| 4.1.3 | Purification et hexamérisation | 91 |
| 4.2 | Expression et purification des protéines deutérées | 94 |
| 4.2.1 | Adaptation des bactéries et culture par fermentation | 94 |
| 4.2.2 | Purification et hexamérisation | 98 |
| 4.3 | Activité enzymatique | 99 |
| 4.3.1 | Dépliage de la GFP _{ssrA} par <i>Ph</i> PAN et <i>Mj</i> PAN | 99 |
| 4.3.2 | Optimisation des conditions de dépliage | 99 |
| 5 | RESULTATS | 101 |
| 5.1 | Etude structurale et fonctionnelle du complexe de régulation du protéasome de l'archée <i>Pyrococcus horikoshii</i> | 101 |

| | | |
|------------|---|------------|
| 5.1.1 | Présentation du travail | 101 |
| 5.1.2 | Structure of the assembled PAN unfoldase reveals new insights into the molecular architecture and mode of action of proteasome regulatory ATPases..... | 104 |
| 5.1.3 | Discussion et perspectives..... | 150 |
| 5.2 | Etude des mécanismes de dépliement et de dégradation des protéines par le système PAN-Protéasome de l'archée <i>Methanocaldococcus Jannaschii</i> : mise au point d'une approche en temps réel par diffusion de neutrons aux petits angles | 152 |
| 5.2.1 | Présentation du travail | 152 |
| 5.2.2 | Small angle neutron scattering reveals protein substrate processing by a AAA+ unfoldase at real-time..... | 154 |
| 5.2.3 | Discussion et perspectives..... | 188 |
| 6 | CONCLUSIONS GENERALES ET PERSPECTIVES | 190 |
| 7 | REFERENCES BIBLIOGRAPHIQUES..... | 193 |

LISTE DES ABBRÉVIATIONS

ELMA: Extremophiles and Large Molecular Assemblies

ATP: Adénosine triphosphate

ADP: Adénosine diphosphate

PAN: Proteasome Activating Nucleotidase

GFP: Green Fluorescent Protein

Rpt: Regulatory Particle triple-A ATPase

ILL: Institut Laue Langevin

IBS: Institut de Biologie Structurale

Ifremer : Institut français de recherche pour l'exploitation de la mer

D-lab : deuteration lab

AAA+: ATPases Associated with diverse cellular Activities

ADN: Acide Désoxyribonucléique

ARN: Acide Ribonucléique

ARNr : Acide Ribonucléique ribosomale

E. coli: *Escherichia coli*

Hsp: Heat Shock Protein

S: Svedberg

CP: Core Particle

RP: Regulatory Particle

Rpn: Regulatory Particle non-ATPase

OB: Oligonucleotide/Oligosaccharide-binding fold

EM: Electron Microscopy

nm: nanometer

Ntn: N-terminal nucleophile hydrolase

ROS: Reactive Oxygen Species

UPP: Ubiquitin Proteasome Pathway

UPS: Ubiquitin Proteasome System

PA: Proteasome Activator

NTP: Nucleoside Triphosphate

P-loop: Phosphate loop

PDB: Protein Data Bank

SSD: Sensor and Substrate Discrimination domain

SRH: Second Region of Homology
NSF: N-Ethylmaleimide-Sensitive Factor
LCA: Last Common Ancestor
LECA: Last Eukaryotic Common Ancestor
SANS: Small Angle Neutron Scattering
SAXS: Small Angle X-ray Scattering
Ub: Ubiquitin
SAMP: Small Archaeal Modifier Protein
VCP: Valosin-Containing Protein
Cdc48: Cell division control protein
UbaA: ubiquitin-like activating protein of archaea
SPR: Surface Plasmon Resonance
ITC: Isothermal Titration Calorimetry
RI: Refractive Index
RU: Resonance Unit
NTA: Nitrilotriacetic acid
WT: Wild Type
EDTA: Ethylenediaminetetraacetic acid
CCD: Charge-Coupled Device
BM29: Bending Magnet
ESRF: European Synchrotron Radiation Facility
FPLC: Fast Protein Liquid Chromatography
LB: Lysogeny broth
IPTG: Isopropyl β -D-1-thiogalactopyranoside
DEAE: Diethylaminoethyl
TEA: Triéthanolamine

1 INTRODUCTION

1.1 Avant-propos

Alors que la synthèse des protéines a été et est toujours l'objet d'incessantes enquêtes, leur dégradation, connue sous le nom de protéolyse, est comparativement moins étudiée. En effet, les études de biologie cellulaire contemporaines se sont concentrées sur la synthèse protéique, en ont défini les différentes étapes, de la transcription à la traduction, et en ont précisé le rôle respectif du génome et des ribosomes. Comparativement moins d'études ont été consacrées à la protéolyse, c'est-à-dire la dégradation orchestrée de ces mêmes protéines, l'élimination des déchets et la réutilisation éventuelle des restes.

Les structures protéiques des cellules sont en perpétuel renouvellement. En permanence, il y a une dynamique de synthèse et de dégradation, l'aboutissement étant l'entretien du fonctionnement cellulaire, c'est-à-dire la production de biomatériau spécifique, à un moment donné de sa vie et de son évolution. Il est bien établi que cette dégradation protéique est assurée par des systèmes protéolytiques ATP-dépendants. Ces machineries protéolytiques sont à la fois très variées, spécifiques et puissantes. Elles détruisent rapidement les structures des protéines cibles par un processus hautement régulé dont les mécanismes commencent à être déchiffrés et les différentes étapes alternant les feux rouges ou verts, autorisant ou non la protéolyse, représentent des cibles thérapeutiques potentielles. La protéolyse s'adresse à des protéines anormales ou identifiées comme anormales par des molécules types, dites « chaperonnes » qui servent de référence ou de « patron » et confirment ou non la conformité de la molécule en question. Ainsi sont identifiées les erreurs de synthèse, les protéines tronquées, mal repliées sur elles-mêmes, offrant en trois dimensions un visage atomique différent des protéines « normales ». Dans les mêmes conditions, ces dernières sont aussi triées et recyclées, gardant les peptides ou acides aminés normaux et éliminant les déchets qui sont rejetés hors de la cellule. Cette activité protéolytique n'est possible qu'au travers d'une dépense d'énergie fournie par l'ATP.

Le travail présenté rentre dans le cadre d'une thèse en biologie/biophysique des grands assemblages moléculaires avec comme but principal de porter une étude biochimique, structurale et fonctionnelle sur le complexe d'activation du protéasome ATP-dépendant PAN (Proteasome Activating Nucleotidase). PAN est un complexe hexamérique ATP-dépendant régulateur du protéasome 20S chez les cellules d'archées et homologue des six sous-unités ATPasiques de la base (Rpt 1-6) du système 19S activateur du protéasome eucaryote. L'étude

du système PAN offre alors plusieurs avantages pour mieux comprendre les mécanismes fonctionnels de son homologue eucaryote plus complexe.

Cette thèse a été soutenue par un financement « ANR Archelyse » du groupe ELMA à l'IBS, dédié pour l'étude de la protéolyse intracellulaire chez les archées, et par une bourse de thèse de l'Institut Laue Langevin (ILL). Le travail a été réalisé en deux étapes : 1) une étude structurale et biochimique déjà engagée au sein du groupe ELMA sur le complexe PAN de *Pyrococcus horikoshii* (*PhPAN*), ce travail interdisciplinaire a combiné la biochimie, la diffusion de rayons X aux petits angles (SAXS) et la cristallographie (réalisée par Matteo Colombo, équipe ELMA) avec de la cryo-microscopie électronique en collaboration avec le groupe microscopie électronique de l'IBS. En parallèle, une étude *in vivo* avec différentes souches d'archées a été effectuée par Didier Flament à l'Ifremer de Brest. Dans le cadre de cette partie du projet, j'ai effectué toutes les préparations et les caractérisations biochimiques du système PAN, la mesure et le traitement des données SAXS ainsi que la génération des modèles pseudo-atomiques du complexe PAN en se servant des cartes de densité Cryo-EM, des données cristallographiques et des données SAXS en solution. Ce travail a abouti à un premier article soumis et qui sera présenté dans le premier chapitre des résultats de cette thèse. 2) La deuxième partie du travail correspond à une étude biophysique par diffusion de neutrons aux petits angles (SANS) en temps réel du complexe PAN de *Methanocaldococcus jannaschii* (*MjPAN*). Le travail structural réalisé dans la première partie de la thèse a fourni des hypothèses sur le fonctionnement des AAA+ unfoldases et a permis de bien maîtriser la biochimie du système biologique PAN. Dans cette deuxième partie, à part du travail biochimique sur *MjPAN* pour bien maîtriser le système et la production des échantillons de bonne qualité, j'ai deutéré mes protéines au sein du laboratoire de deutération de l'ILL (D-lab) et j'ai mesuré toutes mes données SANS sur la ligne D22 à l'ILL. Les résultats obtenus sont très prometteurs, surtout au niveau de la compréhension du mécanisme fonctionnel du système PAN et des systèmes AAA+ unfoldases en général. En outre, ce travail a également aidé au développement instrumental sur la ligne D22 de l'ILL avec la combinaison de la spectroscopie de fluorescence *in situ* ainsi que le raccourcissement du temps d'exposition jusqu'à 30 secondes, utilisé pour la mesure de nos données en temps réel pendant l'activité enzymatique du complexe PAN. Les résultats obtenus seront bientôt soumis dans un deuxième article déjà rédigé et qui sera présenté dans le deuxième chapitre des résultats. Dans le chapitre introduction de la thèse je présente d'abord une étude bibliographique sur la dégradation des protéines en général et l'importance du système ubiquitine-protéasome ainsi que le système protéasome des archées et des eucaryotes et l'intérêt des archées

hyperthermophiles pour l'étude des systèmes protéolytiques. En deuxième partie je présente les systèmes AAA+ protéolytiques régulatrices du protéasome et je finis par introduire le système PAN des archées comme une version simplifiée du complexe 19S eucaryote avant de décrire les deux objectifs principaux de ce projet de thèse. Les résultats seront présentés sous forme de deux chapitres en format articles, précédés par des chapitres détaillant les aspects biochimiques de la préparation des échantillons et des techniques biophysiques utilisées, en insistant sur les techniques neutroniques.

1.2 Rôles de la dégradation des protéines et importance du système Ubiquitine-Protéasome

La dégradation intracellulaire des protéines est un processus fondamental qui a lieu dans tous les organismes, depuis les bactéries jusqu'aux êtres humains. L'idée que tous les composants de la matière vivante, en particulier les protéines, sont dans un état perpétuel de renouvellement, n'est apparue qu'avec Schoenheimer (Schoenheimer & Clarke, 1942). Selon la théorie de Schimke (Schimke, 1970), 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. La protéolyse intracellulaire est au cœur de la fonction des cellules et est vitale pour la capacité des cellules à s'adapter aux changements environnementaux. Bien que la plupart des protéines cellulaires soient habituellement stables, il y a certaines protéines spécifiques qui 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 (Maupin-Furlow 2000). Lorsque ces taux de dégradation sont couplés avec un contrôle de la transcription, les cellules sont capables de répondre efficacement aux changements en modifiant rapidement la concentration d'enzymes à des moments et à des endroits précis dans la cellule. Non seulement les protéines 'normales' sont spécifiquement ciblées pour l'hydrolyse, mais aussi les protéines anormales mal repliées par les chaperonnes sont également éliminées (Fig. 1) (Goldberg, 1972). Les protéines anormales comprennent celles récemment synthétisées et qui ne se replient pas correctement ainsi que les protéines qui s'accumulent dans la cellule après exposition à un choc thermique ou des dommages oxydatifs. Elles comprennent aussi les protéines produites avec des erreurs pendant la transcription ou la traduction, par des mutations génétiques ou des dommages de l'ARN (Goldberg, 2003). Les sous-unités des complexes protéiques sont eux aussi souvent

des cibles de dégradation dans le cas des perturbations dans leur stœchiométrie (Table 1). Les systèmes protéolytiques jouent également un rôle important dans l'augmentation globale de l'hydrolyse de la plupart des protéines pendant le besoin des cellules à reconstituer leur stock en acides aminés. Au cours de la vie d'une cellule, quasiment toutes ses protéines sont plusieurs fois renouvelées par dégradation et resynthèse, le processus de dégradation est hautement sélectif et étroitement régulé (Glickman & Ciechanover, 2002), et les protéines sont détruites à des vitesses de « turnover » très différents d'une à l'autre, avec des demi-vies allant de quelques minutes jusqu'à plusieurs jours (Goldberg, 2003). Dans le foie de rat, par exemple, certaines protéines ont une demi-vie de l'ordre de dix minutes et d'autres ont une demi-vie d'un mois (Schimke, 1968).

| 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 d'oxygène, dénaturation intracellulaire |
| Ingénierie génétique | Fusions de gènes, changement de cadre de lecture, localisations incorrectes |
| Mauvais repliements des protéines | Stress environnemental, dysfonctionnement des systèmes chaperon |

Table 1 : Les protéines anormales sont dégradées rapidement dans les cellules (Goldberg, 2003).

Des modifications structurales des protéines conduisent à une dégradation plus rapide. La capacité des cellules à dégrader sélectivement les protéines mal-repliées a été démontrée pour la première fois il y a plus de 40 ans (Goldberg, 1972) avec la découverte que les divers traitements qui perturbent le repliement correct des protéines causent leur hydrolyse rapide. Ces résultats ont conduit au constat que la structure des protéines ne détermine pas seulement leurs propriétés catalytiques mais aussi leurs stabilité intracellulaire (Goldberg, 1974).

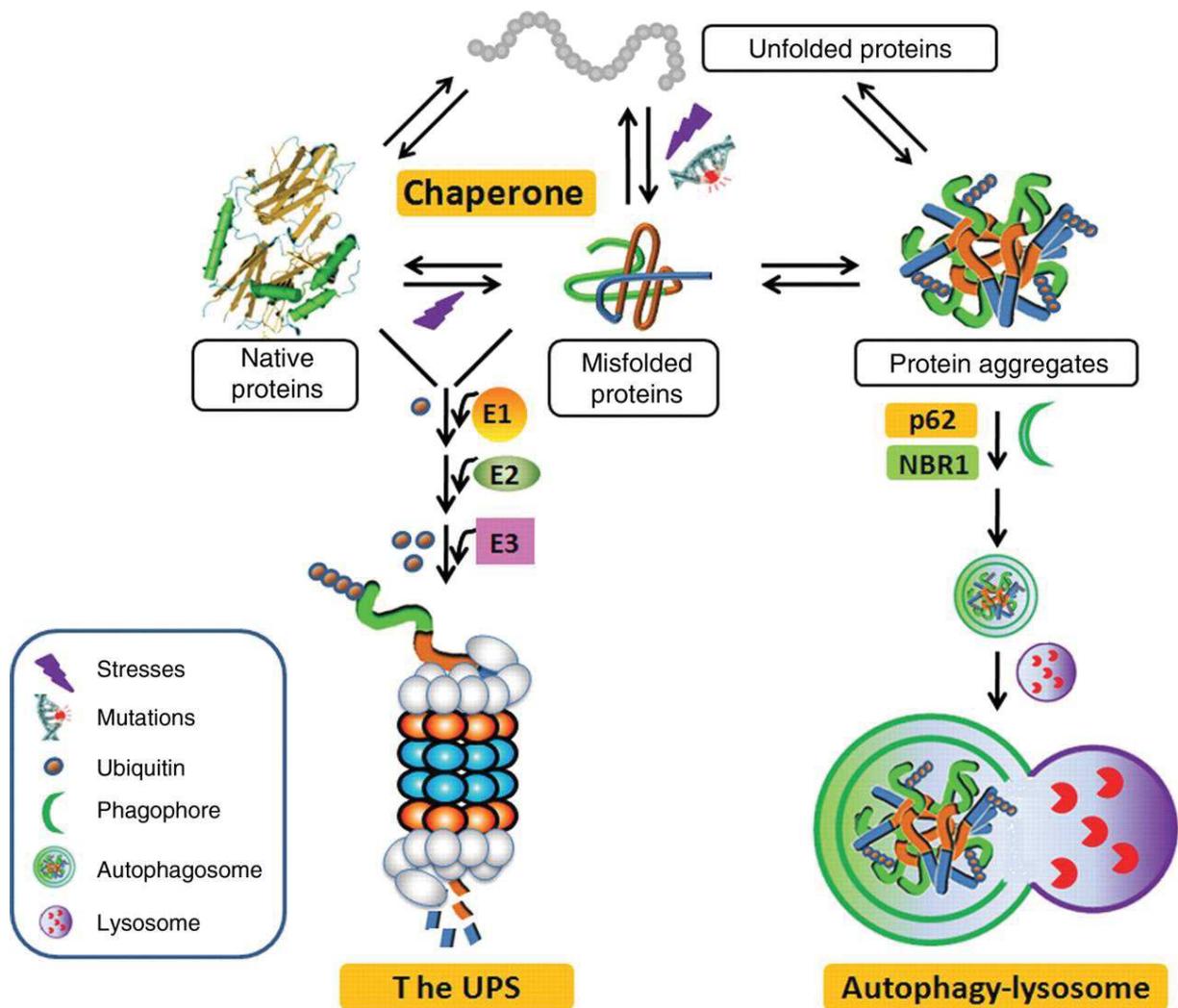


Figure 1 : Représentation schématique du système Ubiquitine/Protéasome (UPS) et du système autophagie-lysosome impliqué dans le contrôle qualité des protéines intracellulaires (Su & Wang, 2010).

Une des caractéristiques des protéines dénaturées est qu'elles s'agrègent et quittent la solution (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 (Sherman & Goldberg, 2001) (Klemes et al, 1981) (Prouty, 1972). La formation de ces corps d'inclusions se produit lorsque les systèmes protéolytiques de la cellule ainsi que les chaperonnes moléculaires, qui normalement empêchent l'agrégation (par exemple Hsp 70/40) (Frydman, 2001) (Hartl & Hayer-Hartl, 2002) ou resolubilisent (Hsp 104) (Glover & Lindquist, 1998) les microagrégats, ne peuvent pas faire face au taux de production des molécules dépliées. Des corps d'inclusions similaires se retrouvent dans divers maladies neurodégénératives, dans lesquelles ces inclusions sont ubiquitinés et physiquement associés

aux protéasomes (Sherman & Goldberg, 2001), suggérant fortement un échec dans la machinerie dégradative de la cellule (Sherman & Goldberg, 2001). Lorsque les cellules mammifères sont traitées avec des inhibiteurs du protéasome, ces inclusions apparaissent rapidement, ce qui indique que plusieurs substrats du protéasome dans les cellules normales sont des protéines mal repliées (Sherman & Goldberg, 2001). Si la production de telles protéines anormales s'arrête et la dégradation est normalement procédée, toutes les cellules, des bactéries (Prouty, 1972) jusqu'aux neurones (Yamamoto et al, 2000) peuvent éliminer les inclusions protéiques. Les mécanismes des cellules mammifères pour la solubilisation et la digestion de ces protéines ne sont pas connus, et une meilleure compréhension de ce processus pourrait avoir des applications thérapeutiques pour diverses maladies liées aux protéines mal repliées.

La dégradation spécifique des protéines intracellulaires est assurée par des machines protéolytiques qui séquestrent des sites actifs pour le clivage des liaisons peptidiques dans une chambre protégée. Ces enzymes sont connues en tant que protéases AAA+, en raison de la présence d'un complexe AAA+ unfoldase qui reconnaît les substrats spécifiques et utilise l'énergie chimique de l'hydrolyse de l'ATP pour déplier mécaniquement la protéine cible et assure sa translocation dans la chambre de dégradation (Olivares et al, 2016). Les protéases AAA+ existent chez les bactéries (ClpXP, ClpAP, HslUV), chez les archées (PAN, Cdc48) ainsi que chez les eucaryotes (les Rpt 1-6 du protéasome 26S). Chez les eucaryotes, l'un des acteurs majeurs de la protéolyse intracellulaire est un système multi-enzymatique sophistiqué, le système Ubiquitine/Protéasome (Finley et al, 2012), qui fonctionne en deux grandes étapes (Fig. 1). Le substrat est d'abord « marqué » par conjugaison covalente de chaînes d'ubiquitine (Ub), une petite protéine ubiquitaire très conservée de 76 acides aminés, grâce à une cascade enzymatique spécialisée. Ces chaînes permettent ensuite la reconnaissance et la dégradation des molécules polyubiquitinylées par un complexe protéolytique de 2 000 kDa, le protéasome 26S. La spécificité et l'efficacité du système Ub/Protéasome découlent de ce fonctionnement en deux temps. En effet, la multiplicité des enzymes d'ubiquitination permet une grande précision dans la reconnaissance des substrats. A l'inverse, le dénominateur commun représenté par la chaîne de poly-Ub simplifie le problème de la reconnaissance de la multitude des protéines dégradées par le protéasome 26S.

La dégradation par le protéasome 26S des protéines ubiquitinylées peut intervenir dans le cytoplasme et le noyau et est dépendante de l'ATP (Coux et al, 1996) (Voges et al, 1999). Le protéasome 26S, dont le rôle prépondérant dans la biologie des eucaryotes supérieurs a été

clairement établi, est formé par l'association de deux sous-complexes: un cœur protéolytique, le protéasome 20S, et un complexe régulateur appelé complexe 19S (Fig. 7).

1.3 Intérêt des archées hyperthermophiles pour l'étude du système protéolytique

1.3.1 Liens des archées avec les eucaryotes

Il existe de nombreuses fonctionnalités qui sont spécifiques pour les eucaryotes et peuvent être retracées jusqu'au dernier ancêtre eucaryote commun (LECA, « last eukaryotic common ancestor ») (Fig. 2), comme le noyau, le système endomembranaire (Baptiste et al, 2005), la mitochondrie (Embley & Martin, 2006), les introns du spliceosome (William Roy & Gilbert, 2006), les chromosomes linéaires avec des télomères synthétisés par des télomérases (Nakamura & Cech), le sexe méiotique (Ramesh et al, 2005), la synthèse du stérol (Desmond & Gribaldo, 2009), structure de la cytokinèse (Eme et al, 2009), et la capacité de la phagocytose (Jékely, 2003). Cependant, le « timing », les mécanismes et l'ordre des événements qui ont conduits à ces caractéristiques cellulaires et génétiques clés ne sont pas clairs. Par conséquent, connaître l'origine de la lignée ancestrale menant au LECA est une question clé.

Depuis leur découverte il y a 30 ans, les archées ont été montrées pour avoir un lien dans l'évolution avec les eucaryotes (Fig. 2) (Gribaldo et al, 2010). En particulier, de nombreux composants d'archée des systèmes impliqués dans la transmission et l'expression de l'information génétique (traduction, réplication et transcription, également appelés « systèmes d'informations ») montrent une similarité de séquence élevée avec leurs homologues eucaryotes qu'avec leurs homologues bactériens, et dans plusieurs cas sont absents chez les bactéries. Par exemple, plus de 30 protéines ribosomiques sont spécifiquement partagées entre les archées et les eucaryotes mais ne sont pas présentes chez les bactéries (Lecompte et al, 2002), et de nombreux facteurs de traduction chez les archées montrent une similitude avec leurs homologues eucaryotes (Londei, 2005). Les ARN polymérases des archées sont très liées à leurs homologues eucaryotes, en termes de composition en sous-unités et de structures (Werner, 2008), et utilisent des promoteurs très similaires et des facteurs de transcription basals pour l'initiation (Bell et al, 2001). Enfin, les quatre activités principales impliquées dans la réplication de l'ADN- l'initiation à l'origine, l'amorçage des fragments d'Okazaki, la synthèse des nouveaux brins et la séparations des brins de l'ADN- sont effectuées par des

ensembles d'enzymes qui sont partagées entre les archées et les eucaryotes mais ne sont pas homologues avec les systèmes bactériens (Leipe et al, 1999).

Un certain nombre de découvertes chez les archées ont confirmé le lien dans l'évolution de leurs systèmes d'informations avec ceux des eucaryotes, telles que la présence de petits ARN nucléolaires qui modifient l'ARN ribosomal (Omer et al, 2000), un exosome (pour la maturation et la dégradation de l'ARN) avec une grande similitude avec son homologue eucaryote (Hartung & Hopfner, 2009) et des histones semblables aux eucaryotes (Čuboňová et al, 2005). Plusieurs 'systèmes opérationnels' (systèmes impliqués dans les fonctions de « housekeeping ») semblent également être partagés entre archées et eucaryotes, y compris les ATPases vacuolaires (Gogarten et al, 1989), le système protéolytique (Ubiquitine-Protéasome chez les eucaryotes et SAMP-PAN-Protéasome chez les archées), les voies de sécrétions (Gribaldo & Brochier-Armanet, 2006) ainsi qu'un système de division cellulaire chez les archées récemment identifié et qui présente une homologie au complexe de « sorting » endosomal chez les eucaryotes (Lindås et al, 2008).

En outre, les archées et les eucaryotes peuvent avoir un lien évolutif dans leurs systèmes métaboliques (Armengaud et al, 2003; Maupin-Furlow, 2011). Cependant, les archées ont également de nombreuses caractéristiques spécifiques. Notamment, les membranes sont constituées par des lipides liés avec des liaisons éther, la liaison glycérol-phosphate a une stéréochimie opposée (glycérol-1-phosphate) par rapport à celles qu'on trouve chez les bactéries et les eucaryotes (glycérol-3-phosphate) et est également synthétisée par une enzyme non liée (Peretó et al, 2004). Les archées ont également des métabolismes uniques tels que la méthanogénèse (Baptiste et al, 2005), des enzymes uniques telles que les topoisomérases spécifiques de l'ADN (Forterre et al, 2007) et les ADN polymérases (Ishino et al, 1998) ainsi que des structures uniques de la surface cellulaire (Ng et al, 2008).

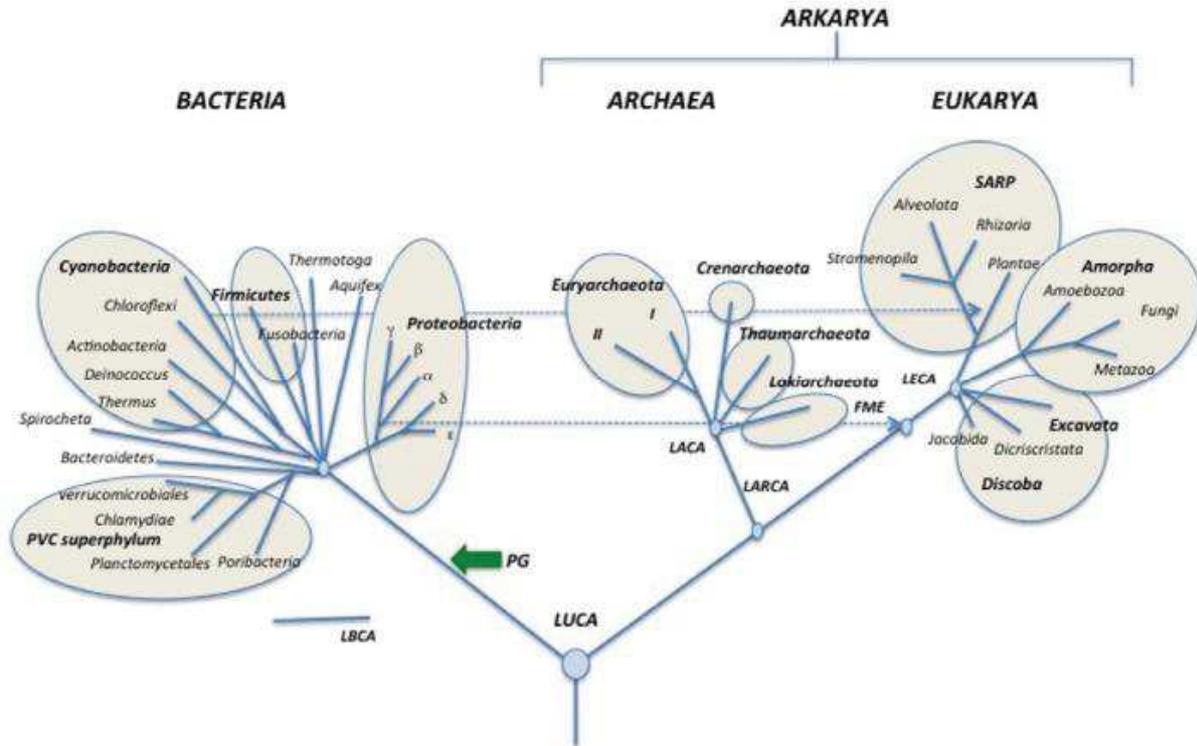


Figure 2 : L’arbre universel du vivant : version adaptée de l’arbre de Woese et al, 1990 et mise à jour par Forterre, 2015.

Bacteria, *Archaea* et *Eucarya* constituent les trois domaines du vivant actuellement reconnus. Arkarya est une proposition pour regrouper le domaine *Archaea* avec *Eucarya* pour un arbre à deux domaines. Les termes LUCA, LBCA, LARCA, LACA et LECA désignent le dernier ancêtre commun de chaque embranchement (Last Universal / Bacterial / Arkarya / Archaeal / Eucaryotic Common Ancestor) (Forterre, 2015).

1.3.2 Intérêt des systèmes issus des archées hyperthermophiles pour l’étude des grands assemblages de la protéolyse

Le laboratoire ELMA (« Extremophiles and Large Molecular Assemblies ») à l’Institut de Biologie Structurale (IBS) de Grenoble utilise principalement des archées pour étudier la fonction de la protéolyse. Ce choix est guidé par plusieurs facteurs. Tout d’abord, les archées présentent un petit génome et ont une position particulière dans l’évolution qui permet plus facilement d’identifier les systèmes et les mécanismes universels qui étaient déjà présents pour faire fonctionner les cellules primitives. De plus, leur proximité évolutive, associée au fait que le système PAN-Protéasome des archées est clairement une version moins complexe que le protéasome 26S des eucaryotes, devrait permettre de mieux comprendre des aspects fondamentaux de la biologie de la protéolyse (Maupin-Furlow, 2013). En ce qui concerne la microbiologie des archées, le nombre de génomes actuellement séquencés et disponibles est

significatif, et des outils génétiques ont été récemment développés (Leigh et al, 2011). Reste le problème des cultures anaérobies et le dégagement de H₂S rendant la manipulation des *Thermococcales* délicate. Le choix de ces organismes marins est cependant motivé par le fait que, à travers l'étude de leur fonction protéolytique, on peut aussi aborder des questions sur l'adaptation spécifique de ces organismes. En effet, ils sont constamment exposés à des variations importantes de leur environnement (variations des flux hydrothermaux par exemple) qui les exposent à des stress physico-chimiques et nutritionnels fréquents. Ceci les oblige à une grande plasticité métabolique. Dans ces circonstances, la fonction de protéolyse doit être fortement sollicitée, soit pour assurer le "contrôle qualité" des protéines *via* leur élimination ciblée, soit pour détruire des enzymes devenues inutiles, soit pour fournir de l'énergie à la cellule *via* la dégradation des protéines (autophagie), soit pour utiliser la matière organique (peptides) de l'environnement comme source d'énergie.

L'utilisation des protéines des archées hyperthermophiles présente également de grand avantage pour la biologie structurale des grands assemblages. Tout d'abord leurs propriétés thermophiles (Fig. 3) permettent de mettre en œuvre des protocoles de purifications à température ambiante et, dans le cas d'expression de protéines recombinantes, de réaliser une première étape de purification des extraits protéiques d'*Escherichia coli* par un choc thermique permettant d'éliminer les contaminants mésophiles. Ceci permet de s'affranchir de la tentation d'utiliser des stratégies de chromatographie d'affinité utilisant des étiquettes (His-tags etc.). La thermostabilité et la thermoactivation des complexes hyperthermophiles facilitent aussi la préparation d'échantillons monodisperses et stables dans le temps pour les études structurales par microscopie électronique ou les études en solution par diffusion de neutrons aux petits angles (SANS), effectuées durant ce projet de thèse. Notamment, la possibilité de contrôler et de ralentir l'activité par une stratégie de thermoactivation a permis une approche SANS en temps réel (ordre de secondes), voir chapitre 2.3. Il est clair que ces études auraient été presque impossibles à réaliser avec des systèmes mésophiles.

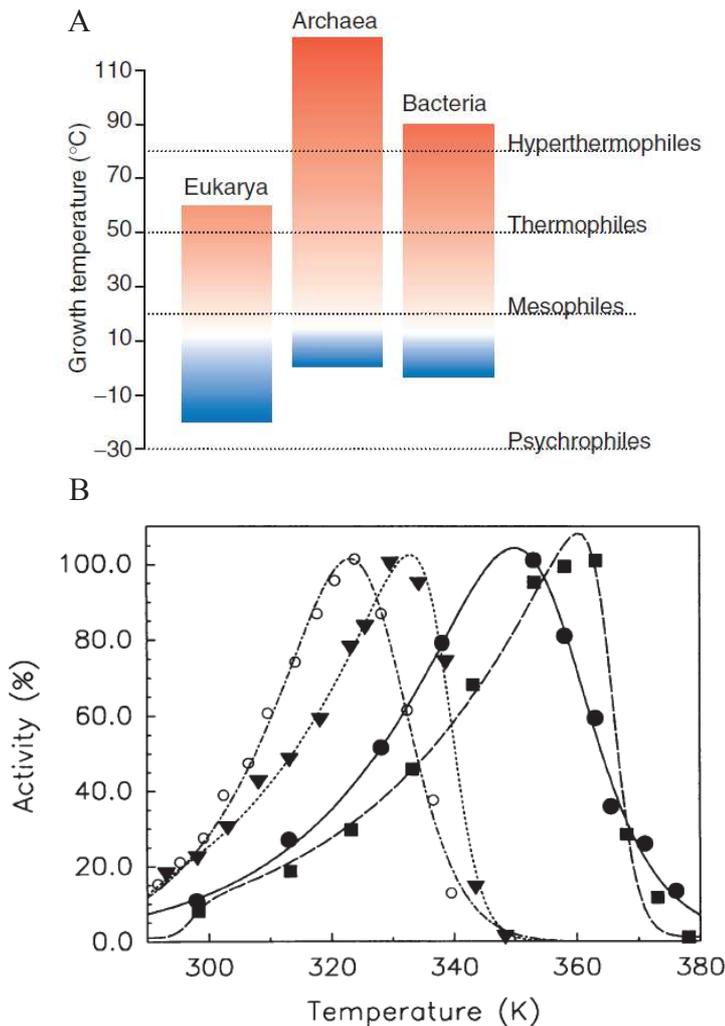


Figure 3 : Distribution phylogénétique de la thermophilie et activité à différentes températures des différentes hydrogénases.

(A) La capacité de croissance à haute température est courante chez les archées, relativement rare chez les bactéries et pratiquement absente chez les eucaryotes (Hickey & Singer, 2004). (B) *Methanoterris igneus* (carrés), *Methanococcus jannaschii* (cercle remplis), *Methanococcus thermolithotrophicus* (triangles) et *Methanococcus maripaludis* (cercles vides). Les deux premiers organismes sont hyperthermophiles, *M. thermolithotrophicus* est un thermophile et *M. maripaludis* est mésophile. Les enzymes présentent une activité maximale à différentes températures comme elles sont adaptées aux différentes températures de leurs habitats (Hei & Clark, 1994).

En ce qui concerne la machinerie protéolytique, les archées codent pour des protéasomes très similaires à ceux des eucaryotes par leur structure et leurs fonctions de base (Maupin-Furlow et al, 2006). Les études du protéasome des archées couplées avec ceux des eucaryotes ont beaucoup facilité la compréhension des mécanismes moléculaires gouvernant la dégradation contrôlée des protéines par ces nanomachines compartimentées. Les archées fournissent des modèles plus simples pour faciliter notre compréhension des systèmes eucaryotes plus complexes et beaucoup de connaissances sur les mécanismes moléculaires eucaryotes ont été acquises par des études biochimiques et génétiques des systèmes d'archées. Ceci est exemplifié par la première structure du protéasome 20S de l'archaea *Thermoplasma acidophilum* (Fig. 8) (Lowe et al, 1995).

Bien que beaucoup d'informations soient connues sur la protéolyse médiée par le protéasome, de nombreuses questions restent à résoudre. Pour certaines de ces questions les systèmes d'archées donnent des modèles simples qui pourraient améliorer notre compréhension sur la

fonction du protéasome eucaryote (Maupin-Furlow et al, 2006). En particulier, la composition des sous-unités de la particule régulatrice AAA+ ATPases chez les archées qui s'associe avec la particule catalytique 20S est plus simple comparée au système 19S eucaryote (Fig. 12 et 27) et fournit donc beaucoup de perspectives dans la compréhension de la relation structure-fonction du protéasome eucaryote.

Dans le cadre de ce projet, l'utilisation du complexe d'activation du protéasome PAN des cellules d'archées hyperthermophiles *Pyrococcus horikoshii* et *Methanocaldococcus jannaschii* comportaient plusieurs avantages pratiques: en effet, des études biophysiques sophistiquées sur le système par microscopie électronique et par SANS en temps réel, 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 sur le substrat modèle utilisé (GFPssrA). Toutes ces expériences nécessitent une grande quantité de solution protéique homogène et monodisperse ainsi qu'une stabilité du système et un contrôle de son activité par thermoactivation qu'on ne trouve pas souvent chez les complexes eucaryotes.

1.3.3 Présentation des archées d'origine des protéines étudiées dans la thèse

Mon travail de thèse porte sur le système d'activation du protéasome PAN, issu des archées hyperthermophiles *Pyrococcus horikoshii* (pour l'étude structurale en Cryo-EM, cristallographie et SAXS), *Pyrococcus abyssi* (pour l'étude *in vivo* de la forme courte de PAN) et *Methanocaldococcus jannaschii* (pour l'étude biophysique par SANS en temps réel). Pour apprécier l'intérêt général en biologie des machines moléculaires des archées hyperthermophiles il est nécessaire de présenter les trois souches qui sont à l'origine de notre système PAN et l'origine de leurs caractères hyperthermophiles.

Pyrococcus horikoshii

L'archée hyperthermophile *Pyrococcus horikoshii* (Fig. 4) est une *Euryarchaea* de l'ordre des thermococcales. Ce groupe comprend des organismes qui se développent tous dans des sources hydrothermales marines, en surface ou dans les abysses. Les sources hydrothermales océaniques résultent de remontées d'eau infiltrée dans les fissures du plancher océaniques et chauffée du fait d'une activité volcanique. L'activité tectonique est à l'origine de la formation de cheminées appelées « fumeurs noirs » dans lesquels le fluide hydrothermal chargé en minéraux jaillit à une température de 400 °C dans une eau à 2 °C. La forte pression

hydrostatique empêche l'eau de bouillir à ces profondeurs. Les effluents acides sont riches en métaux Fe(II) et Mn(II), et contiennent de fortes concentrations en H₂S, H₂ et CO₂ (Zierenberg et al, 2000). Ce sont ces composés chimiques réduits qui fournissent les accepteurs et donneurs d'électrons utilisés par les *Thermococcales* pour réaliser une production primaire par chimiosynthèse. Au contact de l'eau de mer, les minéraux dissous précipitent et forment des édifices creux qui peuvent atteindre plusieurs mètres. Les *Thermococcales* abyssales (*Pyrococcus horikoshii*, *Pyrococcus abyssi*, *Pyrococcus yayanosi*, *Thermococcus barophilus*, etc.) proviennent de l'intérieur de ces cheminées. La souche de *P. horikoshii* provient de sources hydrothermales situées à 1395 m de profondeurs au fond de la faille d'Okinawa dans l'Océan pacifique (González et al, 1998). Elle est capable de se développer à des températures comprises entre 80 et 102 °C, avec un optimum à 98 °C. Le pH optimal est 7 et les concentrations de NaCl entre 1 et 5% avec un optimum à 2,4% NaCl (González et al, 1998). *P. horikoshii* peut fermenter les peptides et les sucres en dioxyde de carbone, hydrogène et acides gras. Si une source de polysulfite est présente, les *Thermococcales* dégagent de l'hydrogène sulfure. Pour le catabolisme du glucose, les *Thermococcales* utilisent un chemin modifié de la voie d'Embden Meyerhoff (Imanaka et al, 2006). Les *Thermococcales* comme les halophiles sont capables d'utiliser le catabolisme des peptides et des acides aminés comme seule source de carbone et d'énergie. Ceci est permis par les amino-transférases qui convertissent les acides aminés en leur 2-oxoacides correspondants par une réaction de transamination avec un autre 2-oxoacide (Namboori & Graham, 2008). Certains acides aminés comme le glutamate, l'aspartate et l'alanine n'ont pas besoin d'être modifiés pour être transaminés et alimenter le métabolisme.

Pyrococcus horikoshii

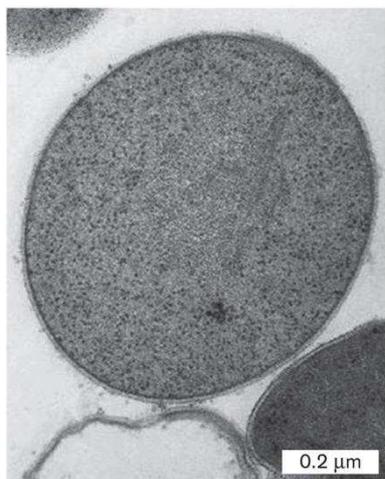


Figure 4 : Image par microscopie électronique de transmission d'une section mince de *P. horikoshii* (González et al, 1998)

Pyrococcus abyssi

L'archée hyperthermophile *Pyrococcus abyssi* a été identifiée au niveau des sources chaudes. Elle a été isolée dans une cheminée hydrothermale, prélevée à 2000 mètres de profondeur dans le Bassin Nord Fidji (Océan Pacifique) (Erauso et al, 1993). C'est une archée marine, chimioorganotrophe stricte. Elle est capable de se développer à des températures comprises entre 67 et 102 °C, avec un optimum à 96 °C, pour un pH neutre. Les cellules de *P. abyssi* sont des coques de 0,8 à 2 µm (Fig. 5). La structure de la paroi est similaire à celle de *Methanococcus jannaschii* avec une membrane cytoplasmique, entourée d'une double « S-layer » (Erauso et al, 1993). Dans son environnement, *P. abyssi* est soumise à une énorme pression hydrostatique (entre 200 et 300 fois la pression atmosphérique). Cependant, et de manière surprenante, elle est parfaitement capable de se développer à la pression atmosphérique. Ce sont donc des micro-organismes barophiles facultatifs. Mais les fortes pressions accélèrent leur croissance.

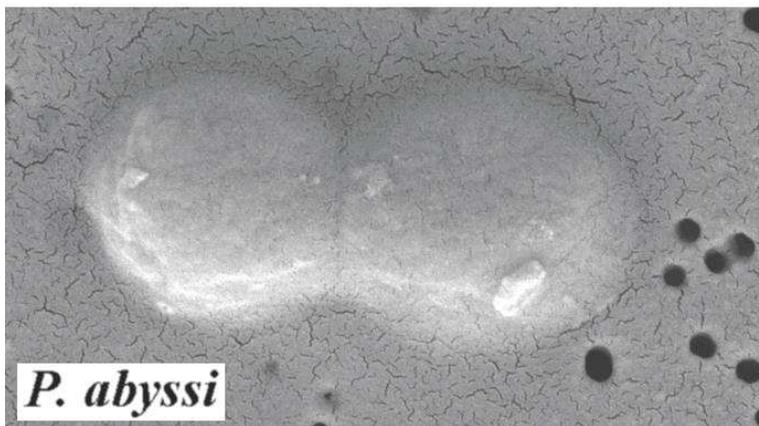


Figure 5 : Image par microscopie électronique à balayage de *P. abyssi* (Orange, 2008)

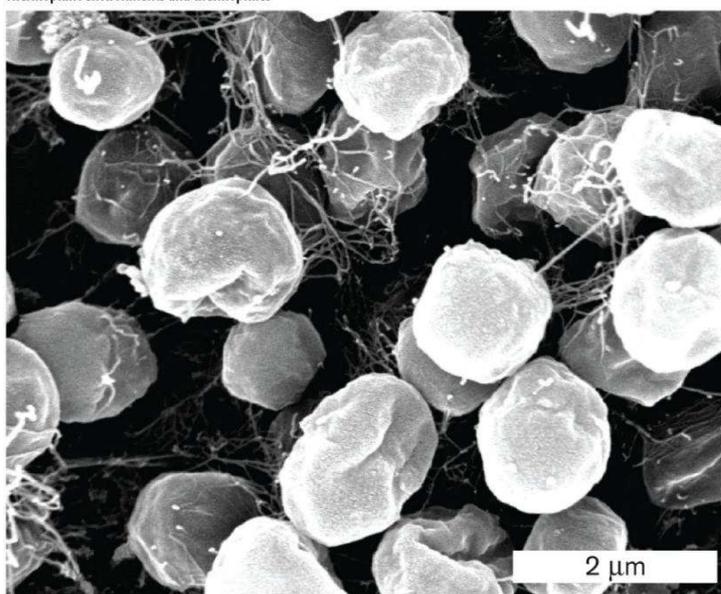
Methanococcus jannaschii

L'archée méthanogène hyperthermophile *Methanocaldococcus jannaschii* a été isolée sur un prélèvement de cheminée hydrothermale, au niveau de la dorsale Est-Pacifique, à proximité de la pointe de la Basse Californie (Mexique), à une profondeur de 2500 mètres (Jones et al, 1983). C'est une archée méthanogène, autotrophe, hyperthermophile, comme on en trouve un certain nombre dans le domaine des archées. Elle se présente sous la forme de coques irrégulières, avec des plissures dans sa paroi évoquant la forme d'un grain de raisin sec, de taille légèrement inférieure au micron (Fig. 6). Elle peut se développer entre 45 et 86 °C, avec un optimum à 80 °C. Sa paroi est constituée d'une membrane cytoplasmique, et d'une couche de surface «S-layer » d'environ 20 nm d'épaisseur (Jones et al, 1983) (König, 2001). Elle possède deux faisceaux de flagelles polaires. Depuis son identification (Jones et al, 1983), *M.*

jannaschii était classée dans le genre des *Methanococcus*. 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* (noter l'ajout du 'caldo', pour chaleur), avec *Methanocaldococcus jannaschii* comme souche type (Graham et al, 2001).

Le génome du *M. jannaschii* a été le premier parmi les génomes des archaebactéries dont on a obtenu le séquençage complet (Bult et al, 1996) ce qui révéla des relations évolutives complexes entre les archées, les eucaryotes et les bactéries: seulement moins de la moitié des gènes trouvés dans son génome pouvait, en fait, être reliée à ceux des autres organismes. Des études ont montré que le métabolisme du *M. jannaschii* et sa croissance à haute température sont favorisés par la pression jusqu'à 750 bar et que les limites pour la croissance sont atteintes à 90 °C soit à basse ou à haute pression (Jones et al, 1983). Ces résultats suggèrent que *M. jannaschii* est effectivement un organisme barophile et pas seulement barotolérant. Cette conclusion semble être aussi confirmée par des études des réactions enzymatiques qui ont montré que l'application d'une pression de 500 bar peut augmenter jusqu'à cinq fois le temps moyen de vie de l'hydrogénase de *M. jannaschii* à 90 °C (Hei & Clark, 1994) et jusqu'à trois fois celui de la protéase à 125 °C (Michels & Clark, 1997). Néanmoins, des études récentes ont montré que la pression inhibait la fonction du 20S protéasome du *M. jannaschii* (Frankenberg et al, 2003), suggérant ainsi que dans certains cas la moindre activité d'une protéine à la pression native pourrait être un mécanisme de régulation conférant des avantages à la cellule entière. Par conséquent, un comportement bien plus complexe de la cellule par rapport aux hautes pressions et aux hautes températures devra être envisagé.

Thermophilic environments and thermophiles



Microbiology: An Evolving Science, Third Edition Figure 5.4c
Copyright © 2014 W. W. Norton & Company, Inc.

B. BOONYARATANAKORNKIT ET AL., UC BERKELEY

Figure 6 : Image par microscopie électronique à balayage représentative de *M. jannaschii* (Slonczewski & Foster, 2013)

1.4 Présentation des systèmes protéasomes archéens et eucaryotes

1.4.1 Description générale (eucaryotes/archées)

La machinerie protéolytique de la voie de dégradation des protéines médiée par l'ubiquitine chez les eucaryotes est le protéasome 26S (Frankland-Searby, 2012). C'est un grand complexe protéique dynamique qui est hautement conservé chez les eucaryotes, à la fois au niveau structural et fonctionnel. Il se compose de plus de 30 sous-unités protéiques différentes et a un poids moléculaire de plus de 2 MDa. La Figure 7 montre ses composants principaux divisés en plusieurs groupes fonctionnels. La partie centrale (rouge) a la forme d'un tonneau et constitue la particule catalytique (CP) ou la protéase 20S qui est formée par un empilement de quatre anneaux heptamériques : deux anneaux β intérieurs identiques, formés chacun par sept sous-unités β différents (β 1-7) et deux anneaux α extérieurs identiques formés chacun par sept sous-unités α différents (α 1-7). Chaque côté de la particule catalytique 20S est couverte d'une particule régulatrice (RP) 19S qui régule la fonction protéolytique de la protéase (20S). La particule régulatrice est également divisée en deux sous-complexes, la base et le couvercle (« lid »). La base a six sous-unités ATPases différentes, Rpt1-6, et quelques sous-unités non-ATPases, tels que Rpn1, Rpn2 et Rpn13. Le couvercle contient plus de dix autres sous-unités Rpn (« regulatory particle non-ATPase »). L'association entre la base et le couvercle est principalement médiée par la sous-unité Rpn10 (Glickman et al, 1998). Parce que le lien entre la base et le couvercle est relativement faible, sous certaines conditions biochimiques, le couvercle peut être séparé de la base pour former un « lidless » protéasome.

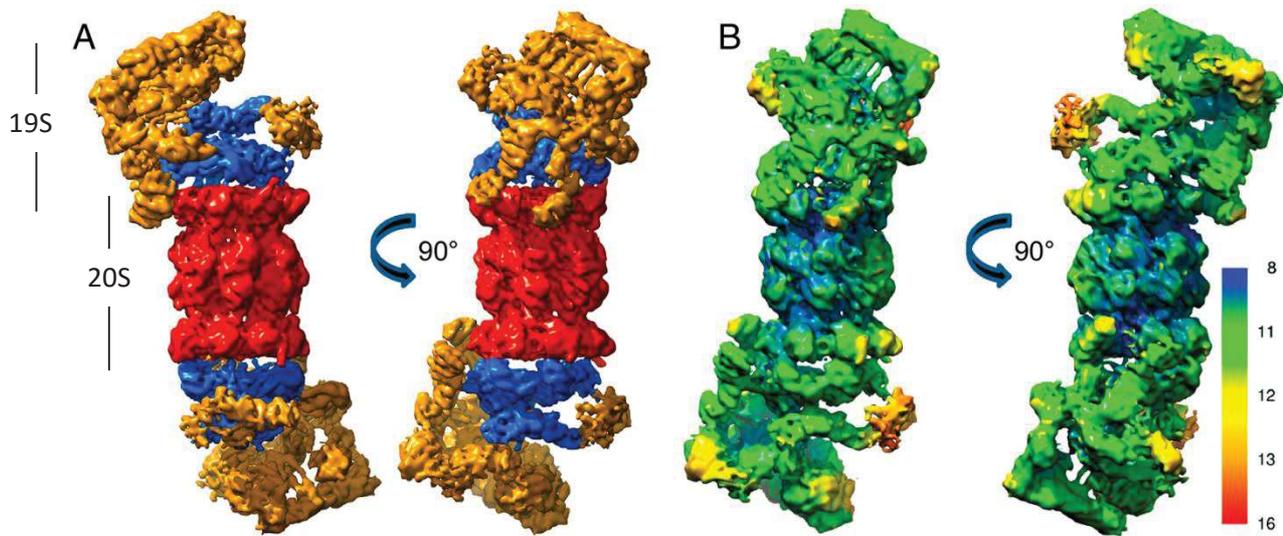


Figure 7 : Carte Cryo-EM du protéasome 26S de *Schizosaccharomyces pombe*.

(A) carte de densité Cryo-EM du protéasome 26S de *S. pombe* à 8,4 Å de résolution, montrée en deux vues liées par une rotation de 90° autour d'un axe de symétrie P7 de la particule catalytique (CP) (CP : rouge ; hexamère AAA-ATPase : bleu ; sous-unités Rpn : doré). (B) la surface de la carte Cryo-EM est colorée en fonction de la résolution local en Å, comme spécifié dans la barre de couleur. (Lasker et al, 2012).

Bien que la structure atomique de la particule catalytique 20S, comme un complexe protéique complet, soit très bien caractérisée (Groll et al, 1997; Lowe et al, 1995), des structures haute résolution de la particule régulatrice 19S sont principalement limitées aux sous-unités individuelles, ou même aux domaines isolés. Même avec le progrès considérable réalisé dans la technologie de toutes les méthodes de détermination des structures, il n'a toujours pas été possible de cristalliser l'ensemble du protéasome 26S. En revanche, des progrès significatifs ont été réalisés vers la détermination de la structure atomique des différents composants isolés de ce complexe. A un niveau de résolution plus basse, la cryo-microscopie électronique (Cryo-EM) a été utilisée pour obtenir des enveloppes 3D du protéasome 26S (Lander et al, 2012) (Lasker et al, 2012) (Fig. 7).

La base de la particule 19S régule la fonction du 20S en stimulant ses activités de protéase, reconnaissant les substrats, les dépliant d'une manière ATP-dépendante et permettant la translocation des substrats dépliés dans la particule catalytique. Le couvercle exerce des fonctions plus complexes et plus diversifiées au cours du processus de dégradation des protéines, impliquant des récepteurs de l'ubiquitine et des enzymes de déubiquitination. La dégradation des protéines ubiquitinylées nécessite le couvercle, mais les fonctions précises de nombreuses sous-unités du couvercle sont moins claires que ceux des sous-unités de la base

ou de la particule catalytique 20S. La plupart des composants du couvercle sont également impliqués dans des activités non-protéolytiques. Récemment, plusieurs sous-unités du 19S ont été identifiées, la plupart d'entre elles sont associées avec le couvercle (Besche et al, 2009). En correspondant aux fonctions de chaque sous-complexe, la particule catalytique 20S est structurellement le complexe le plus stable, tandis que le couvercle est le plus dynamique et le moins stable. Ainsi on peut comprendre qu'il y ait plus d'informations structurales sur la particule catalytique 20S que sur le couvercle de la particule régulatrice 19S.

Les archées contiennent un protéasome fonctionnel et un système de conjugaison protéique (ubiquitin-like) qui est lié au système Protéasome-Ubiquitine (UPS) des eucaryotes, ce système est appelé 'samylation' (SAMP : Small Archaeal Modifier Protein). Comme les eucaryotes, les archées utilisent le protéasome comme un système majeur de la protéolyse énergie-dépendante (Maupin-Furlow et al, 2006). Toutes les espèces d'archées avec des génomes séquencés codent pour une particule catalytique 20S composée de sous-unités α et β ainsi que pour un régulateur AAA+ (PAN et/ou Cdc48) supposé de s'associer à la particule catalytique pour aider à la dégradation des protéines (Maupin-Furlow et al, 2006) (Bar-Nun & Glickman, 2012). Les archées contiennent également la protéase Lon de type B liée à la membrane (Besche & Zwickl, 2004). Cependant, contrairement aux bactéries, les archées manquent les homologues de la protéase FtsH et dans la plupart des cas manquent également les protéases HslUV, Clp et Lon de type A (Besche & Zwickl, 2004). Globalement, la répartition des protéases énergie-dépendantes chez les archées ressemble plus aux eucaryotes qu'aux bactéries ou qu'aux organelles « bacterial-like » des eucaryotes (mitochondries, chloroplastes). Alors que les particules catalytiques 20S des archées n'ont jamais été purifiées en complexe avec les particules régulatrices de leurs organismes natifs, les particules catalytiques des archées peuvent s'associer avec des régulateurs ATPases *in vitro* (Fig. 25). Des études de cette association ont été lancées après la libération de la première séquence d'un génome d'archée (*M. jannaschii*), qui a montré qu'il code pour un homologue des sous-unités de Rpt 1-6 du protéasome 26S eucaryote (Bult et al, 1996). L'homologue, appelé PAN (pour Proteasome Activating Nucleotidase), a été synthétisé et purifié dans *E. coli* et il est indispensable pour la dégradation des protéines substrats (par exemple la caséine) par la particule catalytique y compris celle de *T. acidophilum* (Zwickl et al, 1999) et *M. jannaschii* (Wilson et al, 2000). *M. jannaschii* (*Mj*) *Mj*PAN forme un relativement simple complexe ATPase homohexamérique, et par conséquent a servi comme un modèle idéal pour comprendre comment les protéines Rpts fonctionnent dans la régulation du protéasome eucaryote (da Fonseca et al, 2012).

Tandis que toutes les archées codent pour la partie catalytique du protéasome, il y en a plusieurs qui ne codent pas pour des homologues Rpt/PAN. Que ce soit ou non des ATPases au-delà de PAN peuvent réguler ces particules catalytiques reste à déterminer. Des membres de la sous-famille des protéines AAA+ Cdc48/VCP/p97 sont universellement distribués parmi les archées et chez les eucaryotes, ces protéines de type Cdc48 sont associées au système Ubiquitine-Protéasome dans la dégradation des protéines associées au réticulum endoplasmique (ERAD) (Stolz et al, 2011). Le complexe Cdc48 ATPase des archées a été purifié et a montré une activité unfoldase ATP-dépendante pour stimuler la dégradation des protéines en utilisant des mutants de la particules catalytiques avec une porte d'entrée ouverte (Gerega et al, 2005). Plus récemment, la Cdc48 a montré une capacité d'associer avec la particule catalytique et de stimuler la dégradation des peptides et des protéines (Barthelme & Sauer, 2012). Les particules catalytiques des archées semblent alors être régulées par plusieurs types de protéines AAA+.

1.4.2 Les complexes 20S (eucaryotes/archées)

L'architecture du protéasome 20S est conservée entre les archées et les eucaryotes. Les protéasomes 20S sont des particules cylindriques de 15 nm de longueur et de 11 à 12 nm de diamètre organisées en quatre anneaux autour d'un canal central (Fig. 8). Il renferme des sites catalytiques confinés au sein de la structure ce qui constitue un premier niveau de régulation de son activité. La détermination de la structure cristallographique de la particule catalytique 20S de l'archée *T. acidophilum* a été une avancée majeure, révélant une structure cylindrique de quatre anneaux empilés, avec sept sous-unités α dans chacun des deux anneaux externes et sept sous-unité β dans chacun des deux anneaux centraux (Lowe et al, 1995). Les centres catalytiques ont été localisés dans la chambre centrale. Des études biochimiques et structurales des complexes d'inhibition ont révélé en outre des éléments essentiels du mécanisme catalytique du N-terminal nucléophile (Seemüller et al, 1995). Les archées et les bactéries codent généralement pour une seule sous-unité α et une seule sous-unité β pour former les quatre anneaux heptamériques ($\alpha\beta\alpha\beta$) du protéasome 20S. Les eucaryotes codent sept sous-unités α différentes (α 1-7) et sept sous-unité β différentes (β 1-7), qui occupent des positions uniques pour assembler un complexe heptamérique 20S pseudo-symétrique, tel que révélé par la structure cristallographique de la particule catalytique 20S de la levure *Saccharomyces cerevisiae* (Groll et al, 1997) (Fig. 8 B). Cette structure ainsi que d'autres complexes inhibiteurs associés ont montré la spécificité catalytique des trois sous-unités β 1,

$\beta 2$ et $\beta 5$ chez les eucaryotes, qui possèdent des activités de type caspase, trypsine et chymotrypsines, respectivement. La structure cristallographique du protéasome 20S bovin a indiqué que tous les protéasomes 20S eucaryote ont des structures très similaires (Unno et al, 2002).

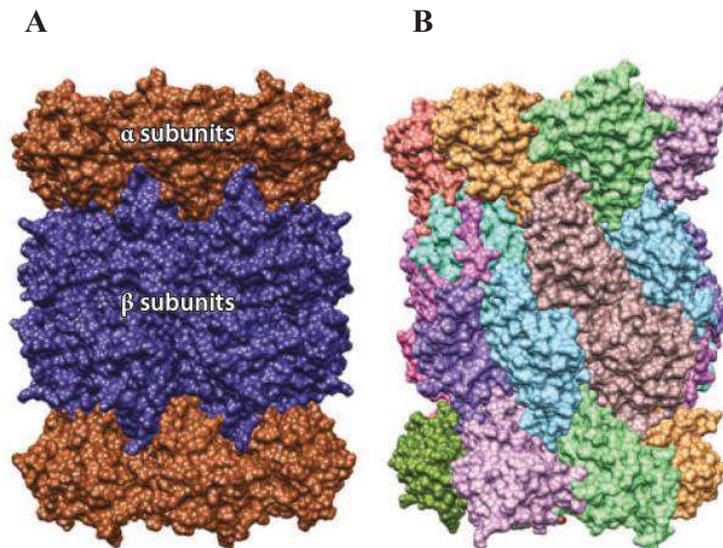


Figure 8 : Comparaison entre la structure du protéasome 20S chez les archées et celle du protéasome eucaryote.

(A) vue de côté du protéasome 20S de l'archée *T. acidophilum* (PDB ID : 1PMA). Les deux anneaux aux extrémités comprennent sept sous-unités α identiques, et les deux autres anneaux à l'intérieur comprennent sept sous-unités β identiques. (B) Vue de côté du protéasome 20S eucaryote de *S. cerevisiae* (PDB ID : 1RYP). Chacun des sept sous-unités α différentes et sept sous-unités β différentes occupe une position unique dans leurs anneaux respectifs. La structure entière a une symétrie P7 liant les moitiés, supérieur et inférieur chacun à l'autre (Kish-Trier & Hill, 2013).

Les protéasomes appartiennent à la famille des Ntn Hydrolases (N-terminal nucleophile hydrolase). Le clivage des liaisons peptidiques est assuré par une attaque nucléophile de la thréonine N-terminale des sous-unités β (Seemüller et al, 1995). Pour étudier l'activité du protéasome, de petits peptides synthétiques sont en général utilisés pour caractériser son activité biochimique. Plusieurs protéasomes d'archée ont été caractérisés et la majorité d'entre eux possèdent une activité chymotrypsin-like notamment celui de *Haloferax volcanii* et de *T. acidophilum*, cependant les protéasomes 20S des archées *Methanosarcina thermophila* et *M. jannaschii* ont montré des activités peptidasiques inhabituelles de type caspase-like et trypsin-like (Maupin-Furlow et al, 2006). Les protéasomes semblent dégrader les protéines d'une manière processive dans la cellule, c'est-à-dire qu'ils dégradent complètement une protéine en oligopeptides avant d'attaquer une autre protéine. Cependant certaines protéines ne sont pas dégradées par le protéasome d'une façon processive, elles sont clivées en plusieurs fragments.

C'est le cas des facteurs NF-kappaB p105 et Nf-kappaB p100 (Ciechanover et al, 2001) et des facteurs d'inhibition de la traduction (Baugh & Pilipenko, 2004).

Une grande variété de structures cristallographiques du protéasome 20S en complexe avec des inhibiteurs a été étudiée, parce que son inhibition représente une approche pour le traitement du cancer. L'inhibiteur 'bortezomib' est actuellement approuvé pour le traitement de myélome multiple et le lymphome à cellules du manteau (Huber & Groll, 2012).

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 (Maupin-Furlow, 2011). 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 ce qui constitue des conditions dénaturantes pour d'autres protéines halophiles, notamment la malate déshydrogénase qui se déplie à bas sel (Franzetti et al, 2002). De même que le protéasome d'autres archées résistent à des agents oxydant, et à des températures qui sont supérieures à l'optimum de croissance des microorganismes (Wilson et al, 1999). D'autre part, Shockley et al. ont montré une régulation transcriptionnelle positive des sous-unités β du protéasome au cours d'un stress thermique chez *P. furiosus* (Shockley et al, 2003). Une autre étude montre que l'inhibition partielle du protéasome (80% des sous-unités β bloquées par l'action de l'inhibiteur) résulte en une incapacité des cellules à surmonter un stress thermique (Ruepp et al, 1998).

1.4.3 Les systèmes d'adressage aux protéasomes (ubiquitination et sampylation)

Les cellules eucaryotes ont développé des voies de contrôle pour éviter l'accumulation des protéines aberrantes et pour détruire spécifiquement des facteurs de régulation. Pour cela, les cellules identifient, étiquettent et dirigent les protéines cibles vers leur destruction principalement via le système Ubiquitine-Protéasome qui est l'un des mécanismes les plus élégants d'adressage des protéines eucaryotes à la protéolyse (Fig. 9).

L'ubiquitine est une chaîne de 76 résidus qui est ajoutée *en trans* sur les résidus lysines de substrats protéiques dans une réaction énergie dépendante (Hershko, 1988). La dégradation des protéines intracellulaires via le système Ubiquitine-Protéasome est un processus extrêmement complexe, contrôlé dans le temps et très régulé. Il joue un rôle dans une variété de processus cellulaires de base. La dégradation d'une protéine par cette voie implique deux

étapes distinctes et successives : (1) le marquage du substrat par la fixation covalente de plusieurs molécules d'ubiquitine (chaîne polyubiquitine) pour former le signal protéolytique et (2) la dégradation de la protéine marquée par le complexe du protéasome 26S avec libération d'ubiquitine libre et réutilisable catalysée par les enzymes de recyclage de l'ubiquitine (Fig. 9). La conjugaison de l'ubiquitine, une chaîne peptidique très conservée au cours de l'évolution, sur le substrat protéique se produit par l'intermédiaire d'un mécanisme de cascade en trois étapes. Dans un premier temps, l'enzyme d'activation E1 active l'ubiquitine par une réaction ATP-dépendante pour générer un intermédiaire thiol ester de haute énergie, E1-S~ubiquitine, où l'ubiquitine est liée à un résidu cystéine interne de l'enzyme E1. Une de plusieurs enzymes E2 (protéine porteuse d'ubiquitine) transfère la fraction d'ubiquitine activée à partir d'E1 via un autre intermédiaire thiol ester de haute énergie, E2-S~ubiquitine, au substrat qui est spécifiquement liée à un E3, un membre de la famille des protéines ubiquitine-ligase. Plusieurs familles d'enzymes E3 ont été décrites.

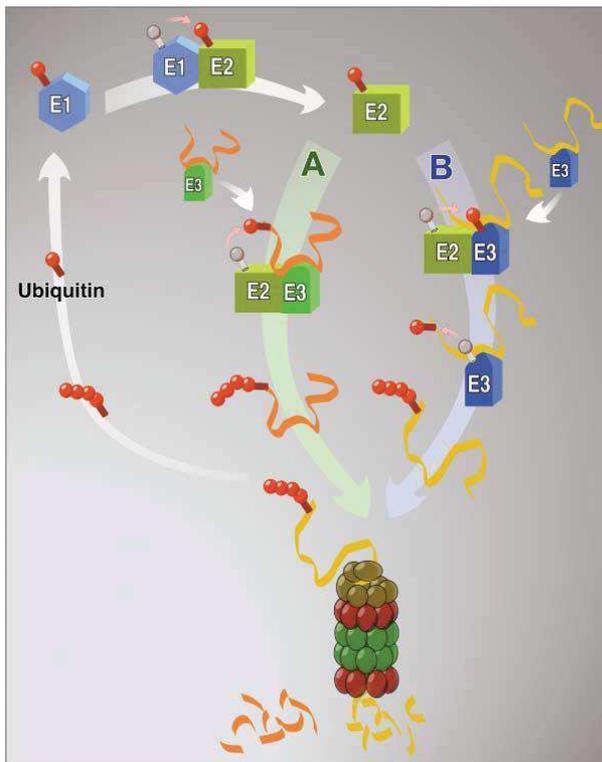


Figure 9 : le système Ubiquitine-Protéasome.

L'ubiquitine est tout d'abord activée par l'enzyme d'activation d'ubiquitine E1 à un intermédiaire thiol-ester de haute énergie. Elle est ensuite transférée, toujours sous forme d'un intermédiaire de haute énergie, à l'enzyme E2 (connue comme enzyme de conjugaison de l'ubiquitine). De E2, elle peut être transférée directement sur le substrat qui est lié spécifiquement à un membre de la famille des protéines d'ubiquitine ligase, E3 (A). Ou bien transférée tout d'abord pour E3 (B), pour générer encore un autre intermédiaire thiol-ester de haute énergie avant d'être transférée à la cible E3-substrat. Des fragments d'ubiquitines supplémentaires sont ajoutés successivement à ceux précédemment ajoutés par un mécanisme similaire pour générer une chaîne de polyubiquitine. Le substrat polyubiquitiné se lie spécifiquement au complexe du protéasome 26S : le substrat est dégradé en courts peptides, et l'ubiquitine (libre et réutilisable) est libéré par les enzymes de deubiquitination (Ciechanover & Brundin, 2003).

L'ubiquitination ne constitue pas le seul moyen d'adressage des protéines à la protéolyse par le protéasome 26S. Il existe à cette date un nombre de protéines dont la dégradation se passe indépendamment de l'ubiquitination comme l'ornithine décarboxylase impliquée dans la synthèse des polyamines (Hoyt & Coffino, 2004) et la protéine p21 inhibitrice des kinases dépendantes des cyclines (Chen et al, 2004).

Chez les archées, il existe une forme de modification post-traductionnelle nommée 'sampylation' qui ressemble à l'ubiquitination chez les cellules eucaryotes (Fig. 10). La sampylation a été identifiée chez les archées *H. volcanii* en se basant sur la conclusion que des protéines homologues à l'ubiquitine (Ubl : « ubiquitin like proteins »), nommées SAMP (« small archaeal modifier protein »), sont fixées de façon covalente aux protéines cibles (Humbard et al, 2010) et qu'une enzyme homologue à l'enzyme d'activation de l'ubiquitine E1 chez les eucaryotes, nommées UbaA (« ubiquitin-like activating protein of archaea »), est requise pour cette modification de la protéine (Miranda et al, 2011). Les homologues, SAMP et UbaA, sont communs chez les archées, suggérant que la sampylation est universelle dans ce domaine de la vie. Des homologues des enzymes E2 (Ub-conjugating enzyme) et E3 (Ub ligase enzyme) n'existent pas dans la plupart des archées (y compris *H. volcanii*) (Burroughs et al, 2009). La sampylation est donc considérée comme une forme simplifiée de l'ubiquitination. Le halophile *H. volcanii* a fourni un modèle idéal pour examiner si une forme simplifiée de l'ubiquitination existe chez les archées ou non. Il a un système génétique établi et une séquence complète du génome, qui a permis l'identification d'un seul homologue de l'enzyme E1 (l'enzyme UbaA) et de plusieurs protéines Ubls avec un domaine β -grasp et un motif diglycine à l'extrémité C-terminale (Fig. 10) (Maupin-Furlow, 2013). 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 (Humbard et al, 2010). La formation de ces conjugués protéiques nécessite la fonction du gène *ubaA* et le motif diglycine C-terminale de SAMP1/2 (Miranda et al, 2011). Ainsi, l'enzyme homologue de E1 (UbaA) est utilisée pour l'activation et la fixation du côté C-terminal de l'Ubl aux protéines cibles par analogie avec les systèmes eucaryotes. Les liaisons isopeptidiques sont formées entre le groupe carboxyle de l'extrémité C-terminale des protéines Ubls des archées et le groupe ϵ -amino des résidus lysines sur les protéines cibles (Maupin-Furlow, 2013). Les archées codent aussi pour des homologues des protéines de de-ubiquitination (DUBs) chez les eucaryotes ainsi que d'autres isopeptidases qui éliminent les molécules liées au dernier résidu des protéines Ub/Ubl (Katz et al, 2010). Les DUBs sont maintenant considérées en tant que régulateurs importants de l'ubiquitination et de la signalisation cellulaire dans les cellules eucaryotes (Amerik &

Hochstrasser, 2004). Dans les archées, les homologues des protéines DUBs appartiennent à la sous-famille JAMM/MPN+, un homologue DUB de l'archée *Archaeoglobus fulgidus* (*AfJAMM*) a été caractérisé au niveau structural et a fourni un cadre pour la modélisation structural des isopeptidases eucaryotes JAMM/MPN+, y compris la sous-unité Rpn11 du protéasome 26S. Un autre membre de la sous-famille JAMM/MNP+ des archées, le *HvJAMM1* de *H. volcanii* a été caractérisé récemment et a montré une activité pour cliver les protéines SAMP-modifiées (Hepowitz et al, 2012). *HvJAMM1* est purifiée en monomère et catalyse l'élimination des SAMP1/2 à partir des protéines liées en isopeptide ou d'une façon linéaire.

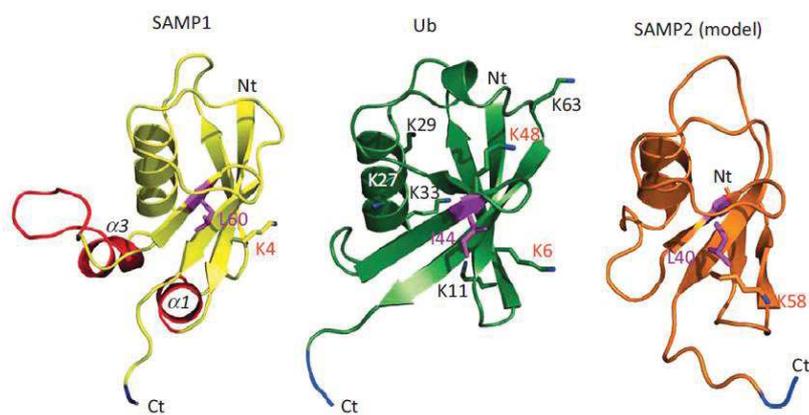


Figure 10 : Comparaison structural entre SAMP1/2 et l'ubiquitine (Ub).

Ile44 de l'ubiquitine nécessaire pour l'interaction protéine-protéine, ainsi que les analogues structuraux Leu60 de SAMP1 et Leu40 de SAMP2, sont mis en évidence en rose en modèle « stick ». Le motif diglycine en C-terminal, y compris le groupe α -carboxyl de l'Ub et SAMP1/2 qui forme une liaison isopeptidique avec les résidus lysines des protéines, est marqué en bleu. Les hélices $\alpha 1$ et $\alpha 3$ supplémentaires de SAMP1 sont colorées en rouge. Les résidus lysine sont aussi présentés en modèle « stick » (à l'exception Lys64 de SAMP2) (Maupin-Furlow, 2013) .

Bien que les protéasomes soient essentiels chez les eucaryotes et les archées (au moins chez *H. volcanii*), les systèmes Ub/Ubl de ces deux domaines de vie ont un impact différent sur la fonction cellulaire (Miranda et al, 2011) (Kimura & Tanaka, 2010). La sampylation est dispensable, alors que l'ubiquitination est requise pour la division cellulaire. Un lien direct entre la sampylation et le protéasome n'as pas encore été établi. Cependant, les suppressions des gènes du protéasome augmentent le niveau des protéines conjuguées aux SAMPs dans *H. volcanii* (Humbard et al, 2010), et ce système de conjugaison protéique contrôle la protéolyse médiée par le protéasome dans les autres domaines de la vie (pupylation chez les bactéries et ubiquitination chez les eucaryotes) (Burns & Darwin, 2010). Une hypothèse est que les protéasomes dégradent les protéines sampylées et autres types de tags « degrons », avec ce dernier stimulant le turnover des protéines qui doivent être enlevées des cellules d'archées

pour se diviser. Conformément à cette possibilité, les protéasomes d'archées dégradent les protéines en absence de la sampylation (par exemple, une queue hydrophobe du côté C-terminal stimule la dégradation de la GFP *in vitro* (Benaroudj et al, 2003)). Il est également possible que les protéasomes ont des rôles non protéolytiques essentiels chez les archées. Comme les protéasomes eucaryotes qui ont des fonctions non protéolytiques dans le processus de la transcription, y compris l'initiation, l'élongation, le traitement de l'ARNm et la dynamique de la chromatine (Yao & Ndoja, 2012).

Les archées semblent alors être hautement dépendant du système protéasome comme leur protéase énergie-dépendante majeur. En se basant sur de nombreuses études qui ont étudié comment les complexes PAN et la particule catalytique des archées fonctionnent au niveau atomique, un mécanisme de la façon dont les protéasomes dégradent les protéines dépliées ou repliées commence à être élaboré. L'amélioration de notre compréhension de la façon dont les protéines sont ciblées pour la dégradation dans les archées promet d'être un domaine de recherche passionnant. La sampylation est probablement un système simplifié pour le marquage covalent des protéines qui semble avoir des racines dans l'évolution à l'ubiquitination ainsi que d'autres voies qui jouent un rôle important dans la fonction des cellules eucaryotes.

Dans le cadre de cette thèse et pour adresser la protéine substrat utilisée (GFP_{ssrA}) au complexe PAN, nous avons utilisés un « degron » bactérien, le peptide 'ssrA'. Chez *E. coli* une interruption de la traduction peut conduire à l'addition d'une queue peptidique de 11 résidus (AANDENYALAA) du côté C-terminal de la chaîne naissante (Gottesman et al, 1998). Cette modification est médiée par un ARN *ssrA* (appelé également 10Sa ARN et ARN_{tm}) et marque la protéine pour la protéolyse. L'ARN *ssrA* fonctionne à la fois comme un ARNm et un ARN_t (Keiler et al, 1996). Lorsque les ribosomes rencontrent un problème au niveau de l'extrémité 3' de l'ARNm endommagé, un ARN *ssrA* chargé d'alanine vient se lier au ribosome comme un ARN_t et ajoute l'alanine à la chaîne naissante ralentie. La traduction passe après de l'ARNm à une petite « open reading frame » (ORF) dans le *ssrA* qui code pour le peptide de dégradation en C-terminale. Ce système fournit à la fois une méthode pour éviter l'accumulation des ribosomes à la fin des ARNm défectueux et un mécanisme général de control qualité qui permet à la cellule de se débarrasser des fragments de protéines incomplètes qui aboutissent à des activités cellulaires inappropriées. Des cellules dépourvues de l'ARN *ssrA* se développent plus lentement et montrent un certain degré de sensibilité à la température (Komine et al, 1994). Dans le cas du système PAN d'archée, des études précédentes ont montré que le « degron » *ssrA* permet d'adresser les substrats protéiques pour

leur reconnaissance et dépliement par PAN *in vitro*. D'autres études ont également montré un rôle des boucles conservées chez PAN et d'autre AAA+ unfoldases « pore 1 loop » dans la reconnaissance et l'interaction avec le peptide ssrA (Benaroudj & Goldberg, 2000) (Benaroudj et al, 2003) (Zhang et al, 2009b).

1.4.4 Les complexes de régulation du protéasome

Les substrats rentrent dans le protéasome à travers un pore situé dans le centre de chaque anneau heptamérique α (Wenzel & Baumeister, 1995). L'ouverture de ce pore est contrôlée par deux éléments structuraux. Un premier obstacle est le canal étroit connu sous le nom de ' α -annulus' et situé légèrement en dessous de la surface de l'anneau (Fig.11B). 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 fermée, formée à partir des résidus N-terminal des sous-unités α du protéasome. Le mécanisme de fermeture de cette porte est différent entre archées et eucaryotes. Chez les eucaryotes, la fermeture adopte une conformation asymétrique définie par les séquences d'acides aminés uniques à chaque sous-unité α . Cette structure limite l'entrée même des petites molécules, son importance est qu'elle est conservée des levures jusqu'aux mammifères (Unno et al, 2002) et que toute perturbation structurale causée par une mutation provoque des anomalies dans la survie chez la levure (Bajorek et al, 2003). Chez les archées, les 12 résidus du côté N-terminal des sous-unités α du protéasome sont très flexibles et sont situés sur la porte d'entrée ainsi qu'ils peuvent s'étendre à travers l'anneau α 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 (Benaroudj et al, 2003) (Religa et al, 2010).

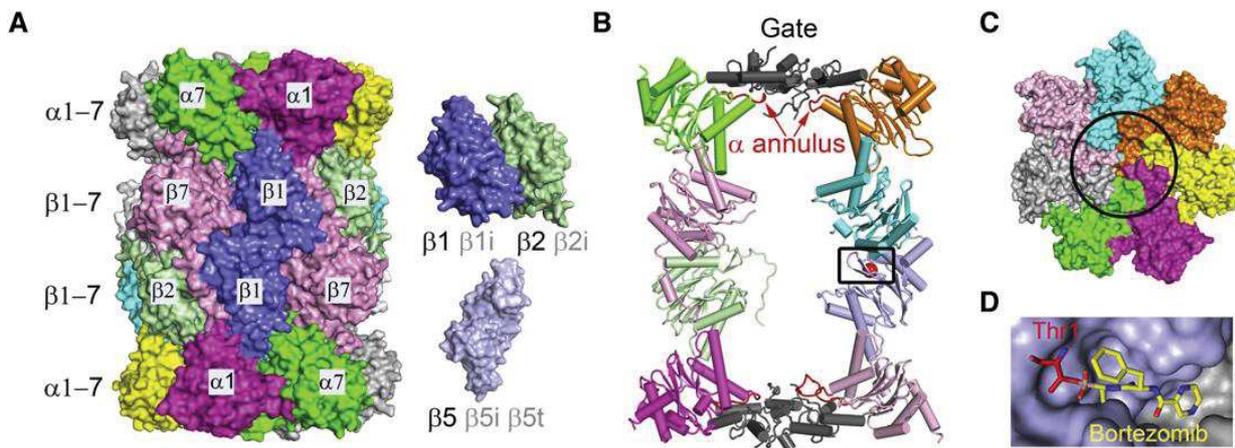


Figure 11 : La structure du protéasome 20S et de sa porte d'entrée.

(A) représentation de surface du protéasome 20S de *S. cerevisiae* (Groll et al, 1997). Les sous-unités sont marquées avec différentes couleurs. Les sous-unités β montrées à droite indiquent que d'autres homologues peuvent être exprimés pour certaines sous-unités. Les protéasomes 20S chez les archées et les bactéries sont très semblables, mais généralement constitués d'un seul type des sous-unités α et β avec une symétrie heptamérique. (B) vue en coupe d'un côté du protéasome 20S de *S. cerevisiae*. La région entourant un site actif est indiquée par une case. La région de la porte fermée est colorée en gris. L' α annulus, juste à l'intérieur de la porte, est une ouverture formée par des boucles (rouge) dans la sous-unité α . (C) vue de dessus de la surface du protéasome 20S. La zone de la porte est entourée par un cercle. (D) « close-up » du site actif $\beta 5$ de *S. cerevisiae* en complexe avec l'inhibiteur bortezomib (Groll, 2006). Correspond à la région en case dans (B). (Stadtmueller & Hill, 2011)

Comme les sites protéolytiques du protéasome sont séquestrés dans un endroit fermé en forme de tonneau, les activateurs du protéasome sont nécessaires pour faciliter l'accès du substrat et donc la dégradation complète des protéines. Pour cela la partie catalytique du protéasome est associée à des activateurs ATP dépendants (AAA+ ATPase) et non ATP dépendants (Stadtmueller & Hill, 2011). 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 19S eucaryotes (« Regulatory Particle/RP/PA700 »), avec son homologue PAN chez les archées. Ces facteurs exigent l'hydrolyse de l'ATP pour aider à la dégradation des substrats protéiques. L'activateur 19S est constitué de plusieurs sous-unités avec des compositions et des fonctions distinctes (Fig. 12). Il reconnaît les substrats polyubiquitinés 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 les complexes 11S (PA28/REG, PA26) et PA200/Blm10. Ces facteurs sont non-ATP dépendants et sont moins conservés que les activateurs ATP-dépendants, et leurs substrats ainsi que leurs fonctions biologiques sont moins clairs, bien que les mécanismes qu'ils utilisent pour activer le protéasome aient été mieux caractérisés.

A l'inverse des activateurs 11S et Blm10/PA200, la fonction biologique de la particule régulatrice eucaryote 19S ATP-dépendante est bien établie pour être la sélection, le

conditionnement et la livraison des substrats pour la protéolyse, en particulier ceux modifiés par conjugaison à une chaîne polyubiquitine (Finley, 2009). Elle contient 19 sous-unités stœchiométriques et est divisée en deux parties, le couvercle (« lid ») et la base (Glickman et al, 1998). L'activité du complexe protéolytique 20S dans les cellules eucaryotes est alors régulée par son association au complexe 19S. Les 19 sous-unités de la particule régulatrice sont six sous-unités AAA-ATPase (Rpt1-6) et 13 sous-unités non-ATPase (Rpn1-3, Rpn5-13, et Rpn15). Les récepteurs de l'ubiquitine (Ub) Rpn10 et Rpn13 reconnaissent les substrats polyubiquitinés (Husnjak et al, 2008) (Schreiner et al, 2008). La métalloprotéase Rpn11 déubiquitine les substrats avant leur dégradation (Verma et al, 2002). Les fonctions des autres sous-unités Rpn restent à établir.

Comme mentionné ci-dessus, la plupart des sous-unités de la particule régulatrice 19S sont non-ATPases. Ils exercent des fonctions très diverses qui sont critiques dans le processus de dégradation des substrats ubiquitinés. Plusieurs d'entre eux sont également impliqués dans d'autres processus cellulaires non protéolytiques. Au cours du processus de dégradation des protéines par le protéasome, les substrats protéiques polyubiquitinés sont reconnus par les récepteurs de l'ubiquitine du protéasome (Rpn10 et Rpn13). Avant que les substrats soient complètement dépliés par les ATPases du protéasome et dégradés par la particule catalytique 20S, Les enzymes de dé-ubiquitination du protéasome (Rpn11) enlèvent la chaîne de polyubiquitine des substrats marqués d'une manière très régulée. Les emplacements exacts de la plupart des récepteurs de l'ubiquitine et les enzymes de dé-ubiquitination ne sont pas clairs, mais ils peuvent être étroitement associés l'un à l'autre pour effectuer une série d'événements hautement coordonnés, allant de la reconnaissance du substrat, l'élimination de la chaîne d'ubiquitine et le dépliement des substrats globulaires.

Les AAA-ATPases ou les Rpts (pour « regulatory particle AAA ATPases ») forment un noyau hexamérique qui déplie les substrats, ouvre la porte de la particule régulatrice, et déplace éventuellement les substrats dépliés vers la particule catalytique (Lasker et al, 2012). La plupart de la connaissance structurale sur les sous-unités AAA-ATPases du complexe 19S jusqu'à présent vient de l'homologue de la particule régulatrice chez les archées, le complexe d'activation du protéasome (PAN) (Djuranovic et al, 2009) (Zhang et al, 2009a), plus de détails sur le régulateur PAN archéen, qui est le système étudié dans ma thèse, seront présentés dans le chapitre 1.6.

Les six sous-unités AAA-ATPase du 19S s'assemblent en deux anneaux empilés : l'anneau formé par les domaines AAA+ (AAA-ring) est symétrique et positionné à proximité de la particule catalytique, tandis que les domaines N-terminaux forment un deuxième anneau

(trimère de dimères) avec des paires de « coiled-coils » (anneau N) (Fig. 12). L'anneau N déplie les substrats, tandis que l'anneau AAA+ fait déplacer les substrats dans la particule catalytique d'une manière ATP-dépendante (Djuranovic et al, 2009) (Zhang et al, 2009a). Le travail difficile pour obtenir des informations sur la structure du complexe 19S est un sujet d'effort intense et de débat. Un modèle attractif fourni par microscopie électronique et des analyses informatiques (Bohn et al, 2010) (Förster et al, 2010) place les ATPases adjacentes au protéasome 20S (Fig. 12). Dans ce modèle, les sous-unités du 19S forment les deux grands sous-ensembles, la base et le couvercle. La base comprend les six sous-unités ATPases (les Rpts), deux protéines d'échafaudage (Rpn1-2) et plusieurs protéines (Rpn10, Rpn13) impliquées dans la reconnaissance et la transformation de l'ubiquitine. Tandis que le couvercle comporte au moins une enzyme de dé-ubiquitination, Rpn11, et huit autres (Rpn3,5-9,12,15) dont les fonctions ne sont pas encore caractérisées. Un autre modèle alternatif suggère que les protéines Rpn1 et Rpn2 forment un tour au-dessus de la porte d'entrée du protéasome 20S (Rosenzweig et al, 2008). Bien que, comme mentionné ci-dessus, le modèle dans lequel l'anneau des ATPases occupe cette position est généralement préféré (Förster et al, 2010). L'ordre topologique des sous-unités AAA-ATPase est Rpt1 / Rpt2 / Rpt6 / Rpt3 / Rpt4 / Rpt5, tel qu'il est déterminé sur la base des interactions protéines-protéines (Förster et al, 2009) (Fig. 12).

Plusieurs rôles ont été attribués aux sous-unités ATPasiques du complexe 19S régulateurs du protéasome eucaryote, notamment via leur capacité à utiliser l'énergie de l'ATP pour induire des changements conformationnels (Fig. 13). Il a été proposé que ces ATPases puissent utiliser le même mode d'action des autres AAA+ notamment en ce qui concerne leur oligomérisation en un complexe hexamérique (Glickman et al, 1998), la liaison au protéasome 20S (Peters et al, 1994), la reconnaissance et la liaison des substrats (Pickart, 1997) (Fu et al, 1999) (Beal et al, 1998), et leur dépliage des substrats (Braun et al, 1999) (Strickland et al, 2000). Les Rpts peuvent assister dans la translocation de substrats protéiques dépliés dans la chambre centrale catalytique du protéasome (Larsen & Finley). Les ATPases peuvent utiliser l'énergie de l'ATP pour induire un changement conformationnel au niveau de l'ouverture de la chambre protéolytique qui est complètement fermée chez la levure (Groll et al, 1997) et y contrôler l'accès. Les sous-unités ATPasiques sont présentes dans toutes les cellules eucaryotes: levure, mammifères, plantes (Glickman et al, 1998) et sont plus proches de leurs homologues d'une autre espèce qu'au sein d'une même espèce.

Les différentes Rpts se ressemblent essentiellement au niveau de leur domaine central ATPasique formés par les motifs Walker A et B. Elles présentent une divergence au niveau de

leur région N-terminale. Il existe cependant un domaine « coiled-coil » chez toutes les Rpts. La région N-terminale est importante pour la liaison des sous-unités entre elles et avec les substrats (Gorbea et al, 2000) (Richmond et al, 1997) (Inobe & Genmei, 2015).

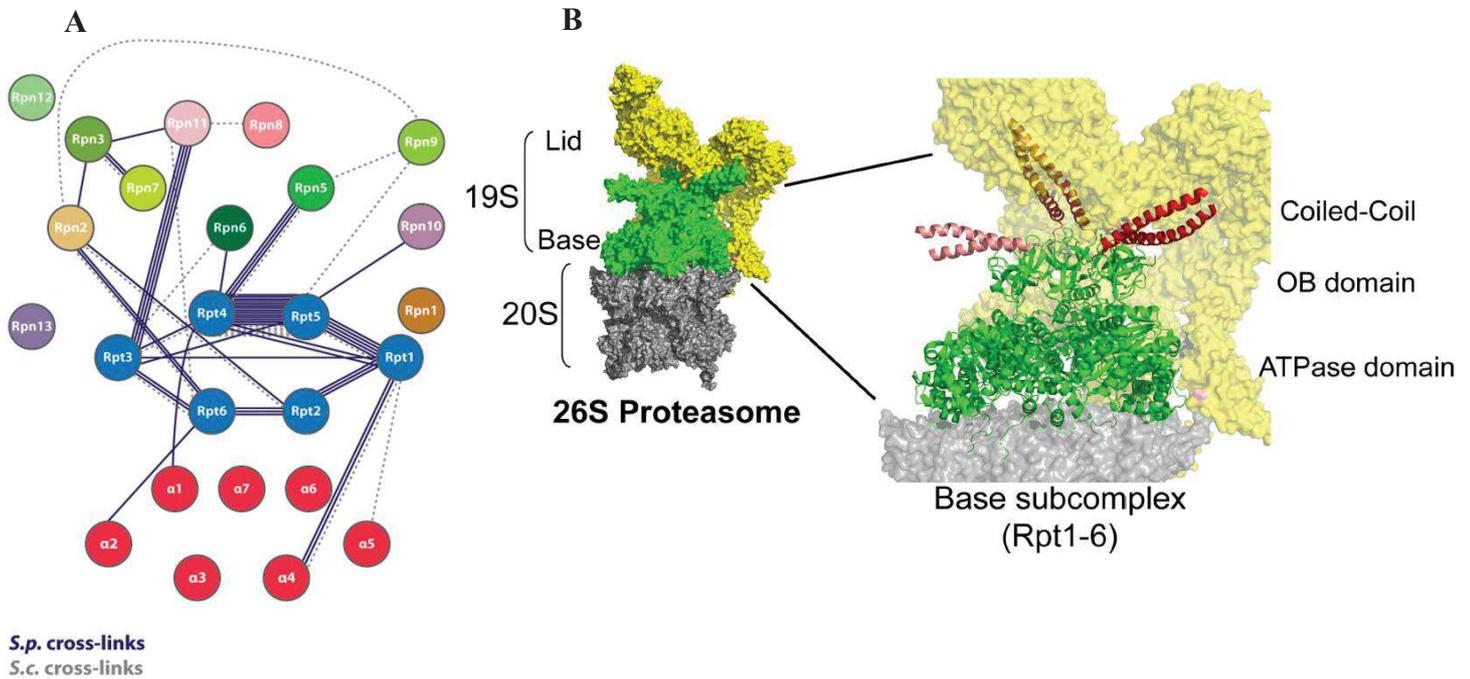


Figure 12 : Architecture des différents sous-unités du protéasome 26S de *S. pombe* et *S. cerevisiae* et structure des sous-unités Rpt de la levure.

(A) Cinquante-cinq paires de lysines réticulées des sous-unités du protéasome 26S de *S. pombe* (*S. cerevisiae*) sont représentées par un réseau d'interaction. (B) Structure du protéasome 26S. (Gauche) surface de la particule régulatrice 19S liée à la particule catalytique 20S (PDB ID 4B4T). La particule régulatrice 19S qui contient les sous-unités Rpt AAA+ ATPase (vert) et les sous-unités non-ATPase (jaune) s'associe aux deux extrémités de la particule 20S (gris). (Droite) Vue agrandie des sous-unités Rpt AAA+ ATPase (vert). Les coiled-coils du côté N-terminal formés par Rpt1-Rpt2 (rouge clair), Rpt4-Rpt5 (rouge) et Rpt3-Rpt6 (rouge foncé) sont colorés. Les structures sont produites par Pymol (Inobe & Genmei, 2015).

Les Rpts constituent des gènes essentiels comme l'a montré la délétion des gènes chez la levure (Gordon et al, 1993) (Russell et al, 1996). Les mutants sont létaux et ne sont pas restaurés par la surexpression des autres ATPases. Les sous-unités ATPasiques semblent exercer des fonctions non redondantes dans la cellule (Rubin et al, 1998) car des mutations au niveau du domaine ATPasique chez les différentes sous-unités Rpts entraînent des phénotypes différents. Il existe cependant une coopération entre les sous-unités ATPasiques du 19S pour la dégradation des substrats, car la dégradation d'un substrat spécifique peut être inhibée par des mutations au niveau des différentes 19S ATPases. L'activité peptidasique du protéasome 20S est stimulée lorsqu'il s'associe à son complexe régulateur indiquant que même les peptides qui n'ont pas besoin d'être dépliés sont aussi régulés par le complexe 19S et une

seule mutation provoque une inhibition de l'activité peptidasique. La mutation au niveau du domaine ATPasique de Rpt2 a montré une inhibition dramatique dans l'activité peptidasique du protéasome, ce qui suggère que cette sous-unité se situe sur la porte d'entrée de la cavité centrale protéolytique (Köhler et al, 2001).

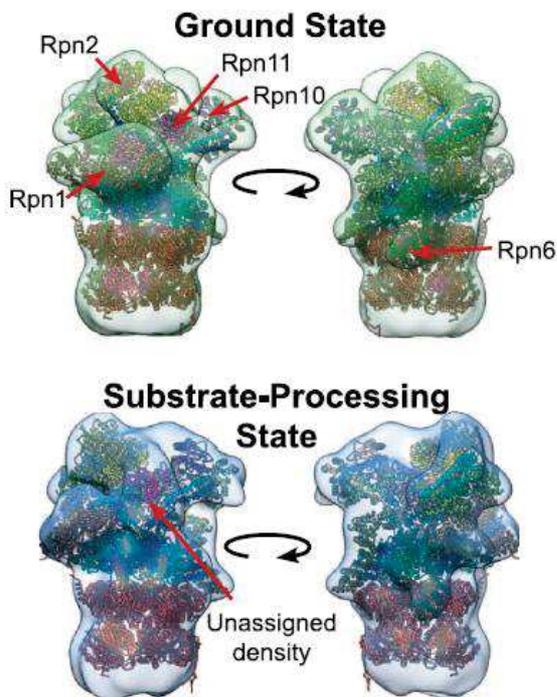


Figure 13 : fit en « rigid body » des modèles atomiques des sous-unités du complexe 19S.

Des modèles atomiques des sous-unités du protéasome 26S de *S. cerevisiae* fittés dans la carte de densité électronique en état fondamental (Ground State) du protéasome (haut) et en état du traitement du substrat (Substrate-Processing State) (bas). Les sous-unités Rpn9/5/6/7/3/12 sont colorées en vert avec différentes nuances. Rpn8/Rpn11 en magenta claire et foncé, Rpn10 et Rpn13 en violet, Rpn1 en brun, Rpn2 en jaune. L'hexamère AAA-ATPase en bleu et la particule catalytique en rouge. Les flèches rouges indiquent les Rpns sélectionnés et la densité non-attribuée (Asano et al, 2015).

1.5 Structure et fonction des AAA+ protéolytiques régulatrices du protéasome

1.5.1 Diversité des AAA+

Les ATPases (ou NTPases) jouent un rôle crucial dans la transformation de l'énergie chimique en événements biologiques. Il existe de nombreux types de NTPase dans la cellule, la classe la plus commune est celle des NTPases type Walker (« Walker-type-NTPases ») qui sont définies par leur motifs Walker A (ou P-loop) et B conservés, intervenant dans la fixation et l'hydrolyse des NTPs.

Les protéines AAA+ (« ATPases Associated with diverse cellular Activities ») forment une superfamille vaste et diversifiée trouvée dans tous les organismes. Ces protéines se rassemblent généralement sous forme de complexes d'hexamères cycliques qui sont impliqués dans le remodelage énergie-dépendant des macromolécules (Iyer et al, 2004). Les membres de la superfamille AAA+ contiennent un module ATPase très conservé de 200-250

acides aminés qui comprend un domaine de noyau $\alpha\beta\alpha$ où se trouvent les motifs Walker A et B de la P-loop NTPases (Ogura & Wilkinson, 2001).

Les protéines AAA+ sont impliquées dans une grande variété de fonctions cellulaires différentes (Fig. 14) comme le dépliement et la dégradation des protéines, la biogénèse des peroxyosomes, la biosynthèse de la bactériochlorophylle ainsi que la recombinaison, la réplication et la réparation de l'ADN (Ogura & Wilkinson, 2001). Les protéines AAA+ comprennent le moteur moléculaire 'la dynéine', les hélicases impliquées dans la réplication de l'ADN, les chélatases du métal, et les protéines associées au protéasome. En raison de leurs fonctions diverses, les protéines AAA+ peuvent être trouvées dans la plupart des compartiments subcellulaires de cellules eucaryotes, ainsi que chez les archées, bactéries et virus (Table 2). Il y a peu de corrélation entre le clade dont une protéine AAA+ appartient avec une activité de remodelage spécifique. Ceci suggère que l'évolution des protéines AAA+ a impliqué une émergence initiale d'un petit nombre de « clades » AAA+ définis que, par la suite, ce nombre est élargi et adapté pour permettre le traitement d'une grande variété de cibles. En outre, l'émergence des protéines partenaires et des cofacteurs a augmenté la diversité fonctionnelle des protéines AAA+ (Iyer et al, 2004).

Les analyses des structures et séquences révèlent que la superfamille AAA+ a subi des divergences considérables à la fois avant et depuis l'apparition du dernier ancêtre commun des trois domaines de la vie, eucaryote, archée et bactérie (Iyer et al, 2004) (Neuwald et al, 1999) (Ammelburg et al, 2006). Des études phylogénétiques basées sur des informations de la séquence et de la structure divisent la superfamille des AAA+ en groupes, clades et familles définis (Erzberger & Berger, 2006) (Ammelburg et al, 2006). Les clades au sein de chaque groupe (Table 2) sont différenciés sur la base de la présence d'éléments structuraux distincts à l'intérieur et autour du noyau AAA+. Cette classification met en évidence le fait que beaucoup de ces lignées AAA+ ont évolué pour acquérir leurs différences fonctionnelles uniques. Les différents clades entrent dans cinq groupes majeurs comme indiqué dans la Table 2. Les membres de chacun des principaux groupes au sein de la superfamille des AAA+ peuvent être trouvés dans les trois grands domaines de la vie (Table 2).

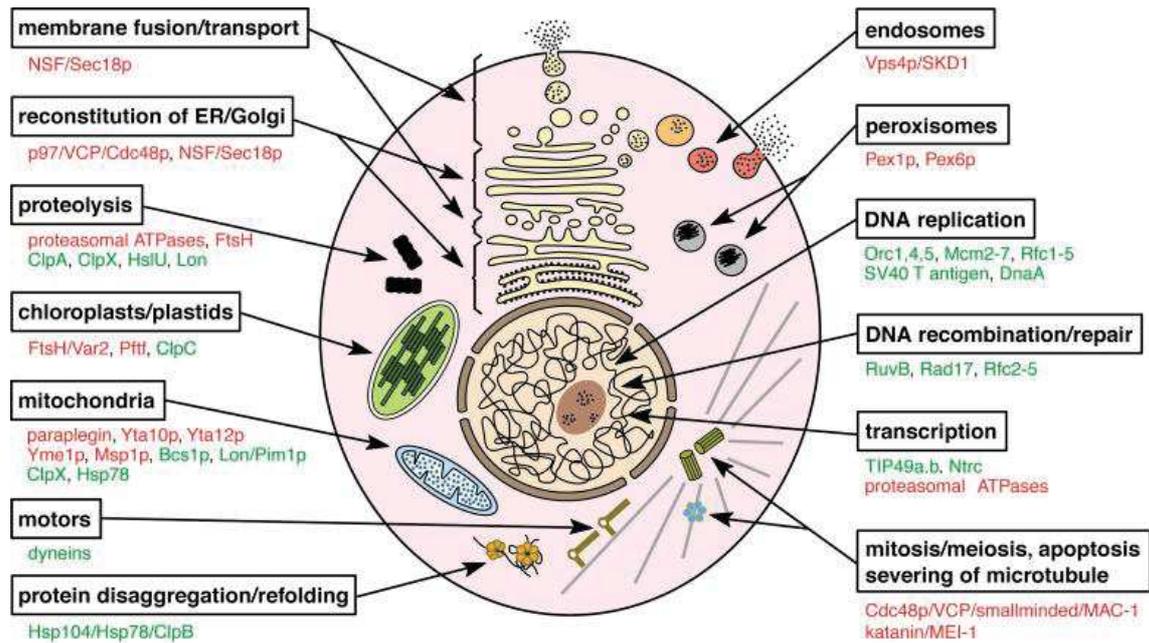


Figure 14 : Les diverses fonctions cellulaires des protéines AAA+.

Un diagramme schématique d'une cellule eucaryote hypothétique avec des organelles et assemblages macromoléculaires représentatives. La localisation subcellulaire des protéines AAA+ et les processus auxquels elles participent sont indiquées. Les membres procaryotes sont aussi inclus. Les membres AAA sont en rouge et les autres membres AAA+ sont en vert (Ogura & Wilkinson, 2001).

Les protéines AAA+ présentent une remarquable diversité des mécanismes d'action. Au cœur de cette diversité est le moteur moléculaire des AAA+. La fixation, l'hydrolyse et la détection de l'ATP sont médiées par un certain nombre de différents motifs et d'éléments de séquences au sein du module AAA+. En général, l'hydrolyse est proposée d'associer l'extraction d'un proton d'une molécule d'eau par le résidu glutamate du motif Walker B, activant ainsi la molécule d'eau pour une attaque nucléophile ultérieure sur le γ -phosphate de l'ATP lié. La lysine conservée et les résidus sérine/thréonine du Walker A acte pour lier les β - et les γ -phosphate du nucléotide lié et de l'ion Mg^{2+} , respectivement (Ogura & Wilkinson, 2001) (Leipe et al, 2003). Au cours du processus d'hydrolyse, les domaines N-terminal et C-terminal du module AAA+ se déplacent l'un par rapport à l'autre ce qui génère une force mécanique qui peut être utilisée pour affecter les événements de remodelage des molécules associées (Ogura & Wilkinson, 2001). Comme la plupart des protéines AAA+ fonctionnent en tant qu'assemblages oligomériques, et comme les communications inter-sous-unités existent entre les membres de ces assemblages, l'hydrolyse des nucléotides tout au long de l'anneau permet aux assemblages AAA+ de fonctionner de manière efficace. Ce moteur moléculaire central est adapté à une grande variété de fonctions. Cela a généralement été accompli par des modifications structurales directes au sein du/des module (s) AAA+ eux-mêmes, et/ou par la

présence des domaines supplémentaires aux extrémités N-terminal et C-terminal du module AAA+ dans une protéine. Dans certains cas, ces domaines supplémentaires sont insérés dans le module AAA+. Par conséquent, les membres de la superfamille AAA+ ont évolué d'une manière telle qu'ils peuvent reconnaître une énorme variété de différents substrats et des partenaires fonctionnels, permettant ainsi à l'énergie de l'hydrolyse de nucléotide à être dirigée vers les différents évènements de remodelage. En effet, les divers mécanismes utilisés par les protéines AAA+ sont aussi divers que le nombre de familles AAA+ qui existent.

| Classification and localization of AAA+ proteins | | | | | |
|--|---------------------|-------------------------|---|-----------------------------------|--|
| Classification | | | Cellular localization and evolutionary distribution | | |
| Group | Clade | Family | Localization | Distribution* | General function |
| Extended AAA | Classical AAA | FtsH | Chloroplast, mitochondria and bacterial membrane | E/B | Protein unfolding and degradation |
| | | Katanin | Cytosol | E | Microtubule severing and disassembly |
| | | NSF/CDC48 | Cytosol | E/A | Membrane fusion/ubiquitin system |
| | | Pex1/6 | Peroxisome | E | Peroxisome biogenesis |
| | | Bcs1p | Mitochondria | E | Cytochrome <i>b_c</i> assembly |
| | Other | PAAA | Cytosol/nucleus | E/A | Proteasome-associated protein unfolding and degradation |
| | | Rubisco activase | Chloroplast | E | Activation of Rubisco |
| | | Rvb | Cytosol and nucleus | E/A | Diverse (for example, DNA recombination/repair, transcription, snoRNP assembly) |
| | | ClpAB-D1 | Cytosol, mitochondria and chloroplast | E/A/B | Protease-associated protein unfolding and degradation/protein disaggregation |
| | | SpoVK | Cytosol | B | Unknown (sporulation associated) |
| | | Ycf2 | Chloroplast | E | Unknown |
| | | AFG1 | Mitochondria | E/B | Unknown |
| | | Viral helicase | Virus | Viruses | DNA recombination and repair |
| | | HEC | Clamp loader | HolB/DnaX | Cytosol |
| RFC | Nucleus | | | E/A | DNA replication and repair |
| WHIP | Cytosol and nucleus | | | E/B | DNA replication and repair |
| Initiation | DnaA/C | | Cytosol | B | DNA replication |
| | ORC/Cdc6 | | Cytosol and nucleus | E/A | DNA replication |
| Other | RuvB | | Cytosol | B | DNA recombination |
| | IstB | | Cytosol | B | DNA transposition |
| PACTT | HCL | HolA | Cytosol | B | DNA replication and repair |
| | | HslU/ClpX | Cytosol and mitochondria | E/B | Protease-associated protein unfolding and degradation |
| | Helix 2 insert | ClpAB-D2/Torsin | Cytosol, mitochondria, ER and chloroplast | E/A/B | Protease-associated protein unfolding and degradation/protein disaggregation/ER complex assembly |
| | | LonA | Cytosol and mitochondria | E/B | Protein unfolding and degradation |
| | | MCM | Cytosol and nucleus | E/A | DNA replication (helicase activity) |
| | | McrB/Unc-53 | Cytosol and nucleus | E/A/B | DNA restriction/unknown |
| | | Midasin | Nucleus | E | Maturation and nuclear export of ribosomes |
| | | MoxR | Cytosol | A/B | Protein complex assembly |
| | | Chelatase | Cytosol and chloroplast | E/A/B | Metal insertion |
| | | σ^{54} Activator | Cytosol | B | Transcriptional activation |
| | | YifB | Cytosol | B | Unknown |
| | | ComM | Cytosol | B | Unknown |
| | | Other | Dynein heavy chain | Cytosol | E |
| LonB | Membrane | | A | Protein unfolding and degradation | |
| ExeA | | Cytosol | B | Type II secretion | |
| STAND | | Varied | E/A/B | Varied | |

*A, B and E refer to archaea, bacteria and eukaryotes, respectively.

Table 2 : Classification et localisation cellulaire des protéines AAA + (Snider & Houry, 2008).

1.5.2 Les AAA+ ATPases protéolytiques

Les protéases et les peptidases représentent des constituants majeurs dans la cellule (3 à 10 % des protéines de la cellule) 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. 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. La majorité de ces protéases sont induites en conditions de stress (Hecker et al, 1996) (Krüger et al, 2000). Les peptidases ATP-indépendantes interviennent notamment dans les fonctions dites de ménage (« housekeeping »). Leur rôle est notamment de 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.

Les protéases ATP-dépendantes ont été identifiées dans tous les organismes analysés et dans les trois domaines du vivant : Bactérie, Eucaryote et Archée (Schmidt et al, 1999). Elles sont classées en plusieurs familles selon l'homologie de séquences de leurs sous-unités protéolytiques (ClpXP, ClpAP, ClpCP, HslUV, Lon, FtsH, PAN/20S, Cdc48/20S et le proteasome 26S) (Fig. 15) (Sauer & Baker, 2011), assurent le contrôle qualité des protéines et sont utilisées dans les circuits de régulation de toutes les cellules. Ces machines contiennent un compartiment protéolytique, avec des sites actifs protégés dans une chambre intérieure, et un anneau hexamérique appartenant à la famille des AAA+ ATPases (Sauer & Baker, 2011). Les protéines substrats s'attachent à l'anneau, soit directement, soit par l'intermédiaire de protéines adaptatrices. Une région non structurée du substrat s'engage dans le pore de l'anneau AAA+, et les cycles de fixation et d'hydrolyse de l'ATP aboutissent à des changements conformationnels qui dénaturent le substrat et permettent le passage du polypeptide déplié à travers le pore dans la chambre catalytique (Fig. 16).

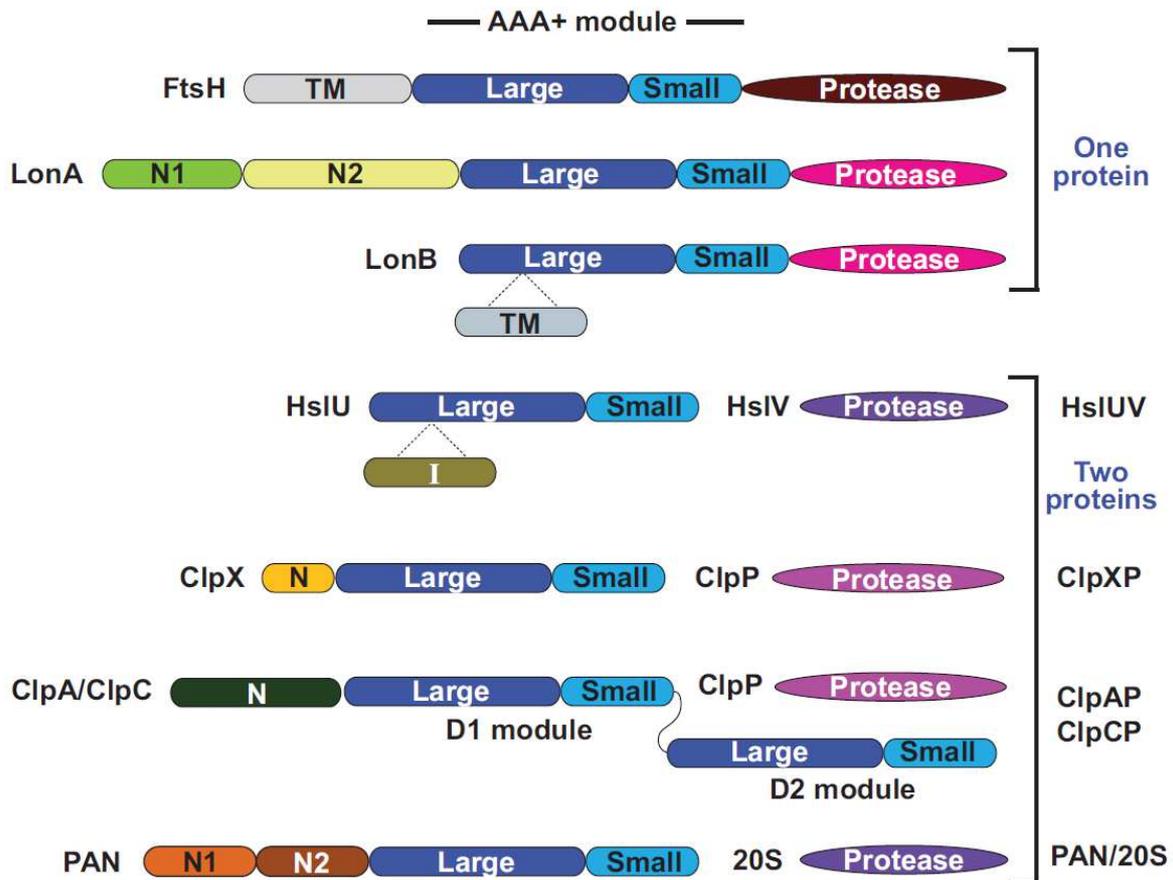


Figure 15 : Structure des domaines des protéases AAA+. Chaque protéase contient un ou deux modules AAA+, chacun est constitué d'un grand et d'un petit domaine et des domaines additionnels spécifiques à chaque famille. Le domaine protéase et le module AAA+ sont présents sur le même polypeptide pour FtsH, LonA et LonB. La protéase et les modules AAA+ sont deux polypeptides distincts dans HslIUV, ClpXP, ClpAP, ClpCP et PAN/20S (Sauer & Baker, 2011).

Les protéases ATP-dépendantes ne présentent pas toutes une homologie significative au niveau de la séquence primaire, ni en ce qui concerne le mode de clivage des protéines. Elles présentent plusieurs caractéristiques communes, notamment leur assemblage constitué de larges complexes multimériques « compartimentés » formés par l'association d'un complexe protéolytique et d'un complexe régulateur de type ATPasique (Fig. 17). Les sites actifs responsables de la protéolyse sont confinés au sein d'une chambre dont l'accès est contrôlé par les ATPases gardant les portes d'entrée de la chambre protéolytique. C'est cette architecture « compartimentée » qui confère à la protéolyse une part de sa spécificité et sa dépendance à l'ATP pour la dégradation d'un substrat.

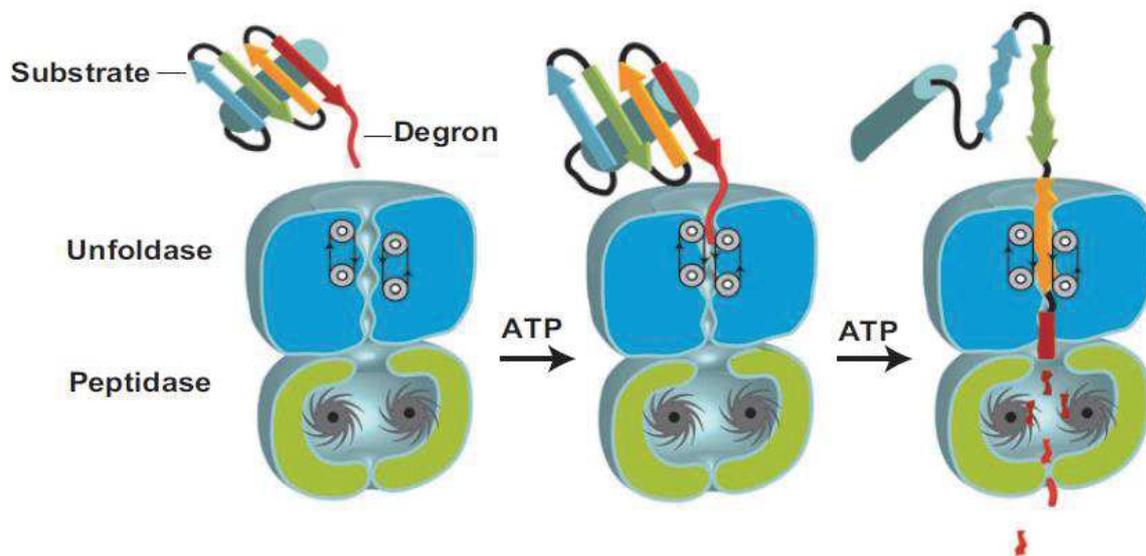
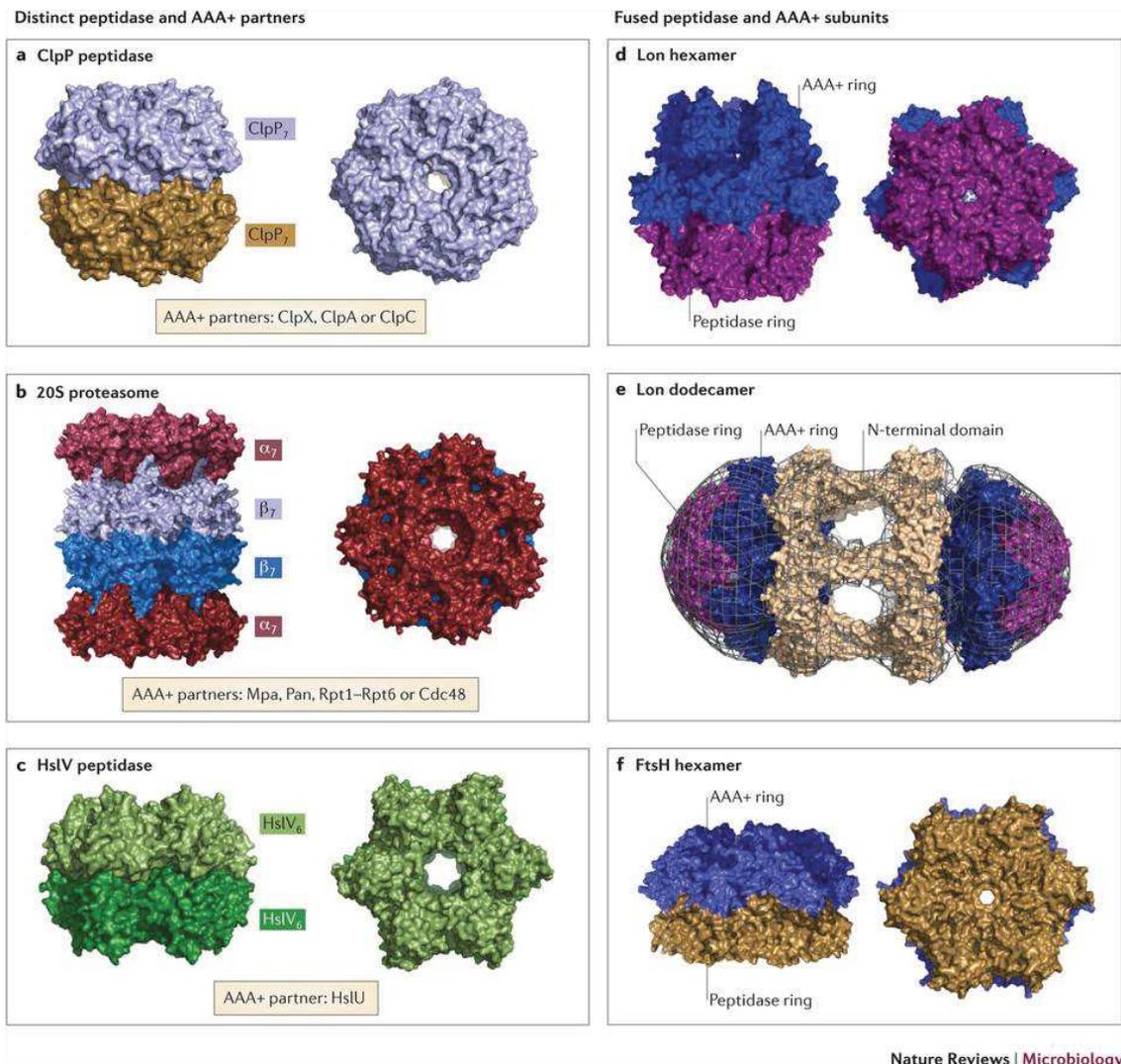


Figure 16 : Mécanisme de base d'une protéase AAA+. Le degron ou le tag de dégradation est initialement reconnu par l'hexamère unfoldase AAA+. Des cycles répétitifs d'hydrolyse de l'ATP alimentent le dépliement du substrat et son translocation via le pore centrale de l'enzyme dans la chambre catalytique du compartiment peptidase associé (Sauer & Baker, 2011).

La ressemblance au niveau de l'architecture des protéases ATP-dépendantes a été soulignée pour la première fois en 1995 lorsque les données structurales du complexe protéolytique ClpP d'*E. coli* et du protéasome 20 S archéobactérien sont devenues disponibles (Lowe et al, 1995) (Kessel et al, 1995). En effet, ces deux protéases présentent la même structure quaternaire en forme de tonneau : ClpP est formée par deux anneaux heptamériques, le protéasome 20S est formé par l'agencement de quatre anneaux heptamériques (Fig. 17). C'est une structure auto-compartmentée où les sites actifs sont confinés à l'intérieur d'une chambre dont l'ouverture ne dépasse pas quelques nanomètres de diamètre (De Mot et al, 1999). L'accès aux cavités internes de ces protéases est en général réservé à des petits peptides ou des protéines complètement dépliées par les complexes d'activation qui précèdent les complexes catalytiques et qui peuvent donc passer à travers les pores. Le complexe protéolytique possède intrinsèquement une activité peptidasique limitée et a besoin pour dégrader des protéines natives d'interagir avec une autre machinerie capable de reconnaître les protéines substrats, de les déplier afin qu'elles soient accessibles aux sites catalytiques. Ces complexes peuvent être complètement formés de sous-unités ATPasiques (ClpA, ClpX, HslU, PAN, Cdc48) ou partiellement (le complexe régulateur 19S du protéasome eucaryote). Dans le cas de FtsH et de Lon, l'activité ATPasique est liée à un domaine intégré dans le complexe lui-même (Chung & Goldberg, 1981) (Stahlberg et al, 1999) (Suzuki et al, 1997). Les domaines ou sous-unités ATPasiques sont les régulateurs de l'activité protéolytique

portant des activités de type chaperonne moléculaire. Elles appartiennent à la large famille des AAA+ qui utilisent de l'ATP comme source d'énergie pour conduire des fonctions de dépliement et de désassemblage (Neuwald et al, 1999) (Lupas & Martin, 2002).



Nature Reviews | Microbiology

Figure 17 : Des peptidases auto-compartmentés sont les composants de dégradation des AAA+ protéases. Dans chaque panneau, une seule structure est représentée. a) La peptidase ClpP d'*Escherichia coli* (PDB code 1TYF) se compose de deux anneaux heptamériques, elle utilise une triade catalytique Ser-His-Asp pour le clivage des liaisons peptidiques et fonctionne avec l'un des trois partenaires AAA+ homohexamériques (ClpX, ClpA ou ClpC). Dans différentes espèces, ClpP peut être composée de 14 sous-unités identiques, des anneaux homomériques distincts ou un mélange de sous-unités dans chaque anneau. b) le protéasome 20S de *T. acidophilum* (PDB 1PMA) a une structure de $\alpha_7\beta_7\beta_7\alpha_7$, il utilise une thréonine nucléophile pour le clivage de liaison peptidique, il s'associe avec le *Mycobacterium* proteasome-associated ATPase (Mpa) homohexamérique chez les bactéries, avec le proteasome activating nucleotidase (PAN) ou Cdc48 homohexamérique chez les archées, ou l'anneau hétérohexamérique Rpt1-Rpt6 dans le protéasome 26S eucaryote. Les anneaux α et β ont 7 sous-unités identiques chez les bactéries et les archées ou 7 sous-unités α et β différentes chez les eucaryotes. c) la peptidase HslV de *Haemophilus influenzae* (PDB 1G3I) se compose de deux anneaux homohexamériques (chaque sous-unité est homologue à une sous-unité β du protéasome 20S), elle utilise une thréonine nucléophile pour le clivage des liaisons peptidiques et s'associe avec l'homohexamère HslU. La protéase Lon homohexamérique de *Thermococcus onnurineus* (PDB 3K1J) est assemblée à partir de sous-unités dans lequel le module AAA+ est fusionné avec le domaine peptidase et utilise un pair de Ser-Lys pour le clivage de liaison peptidique. e) deux hexamères Lon de *E. coli* peuvent s'associer pour former un dodécamère, qui est stabilisé par des interactions du domaine N-terminal qui forment des portails de ~ 45 Å dans le lumen de l'enzyme. Le panneau montre les structures de *E. coli* et de *Bacillus subtilis* (PDB 3LJC et 3M6A, respectivement) modélisées dans une carte de densité électronique à basse résolution. f) la protéase FtsH homohexamérique de *Thermotoga maritima* (PDB 3KDS) s'assemble à partir de sous-unités dans lequel le module AAA+ est fusionné au domaine peptidasique. FtsH utilise un site actif Asp-Zn²⁺ pour l'hydrolyse de la liaison peptidique (Olivares et al, 2016).

1.5.3 Structure des AAA+ ATPases protéolytiques

Le trait caractéristique de la famille des AAA+ est un domaine central ATPasique de 200 à 250 acides aminés, pouvant être présent en une ou deux copies et qu'on appelle le module AAA (Koonin, 1993) (Latterich & Patel, 1998) (Swaffield et al, 1995). Ce domaine est formé de deux sous-domaines: le premier du côté N-terminal nommé P-loop et un autre du côté C-terminal nommé SSD «Sensor and Substrate-Discrimination Domain» (Fig. 18). Le sous-domaine P-loop inclut le motif Walker A (GX₂GXGKT où X représente n'importe quel acide aminé) impliqué dans la liaison au cofacteur Mg²⁺ et la formation des liaisons hydrogènes avec les nucléotides, et le motif Walker B ou « DEXD box » qui serait aussi impliqué dans la liaison au Mg²⁺ et l'hydrolyse des nucléotides (Confalonieri & Duguet, 1995). Le sous-domaine C-terminal est riche en hélices alpha nommé aussi SSD dont le rôle n'est pas encore bien connu. Ce sous-domaine permet la classification de la superfamille des AAA+ en plusieurs familles comprenant les protéines Clp/Hsp100, la famille des protéases Lon et la famille plus classique nommé AAA. Cette famille inclut les sous-unités ATPasiques du protéasome, et la protéase membranaire FtsH. Les membres de la famille partagent une région très conservée nommée SRH (« Second Region of Homology ») caractérisée par la séquence suivante [(T/S)-(N/S)-X₅-DXA-X₂-R-X₂-RX-(D/E). X n'importe quel acide aminé]. Les AAA+ contiennent d'autres domaines très variables dans leurs régions N-terminales et C-terminales dont la composition dépend de la fonction de la protéine. Ces domaines peuvent être des domaines protéolytiques : exemple du domaine C-terminal de Lon et FtsH (Stahlberg et al, 1999) (Suzuki et al, 1997), des domaines de liaison aux complexes protéolytiques par exemple la région C-terminal de ClpA, HslU (Seong et al, 2002). Les domaines N-terminaux peuvent constituer des domaines d'interaction avec d'autres protéines ou de liaisons de substrats (Dougan et al, 2002) (Gorbea et al, 2000).

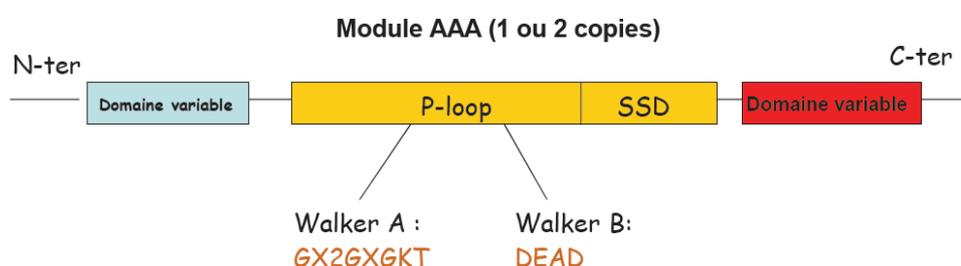


Figure 18: Représentation schématique de différents domaines structuraux d'une protéine AAA+.

Toute protéine AAA+ possède un module AAA (en jaune) qui constitue la signature de la famille des AAA+. Ce module inclut un sous-domaine nommé P-loop qui comprend deux motifs : le walker A et le walker B impliqués dans la liaison et l'hydrolyse des nucléotides. Un autre sous-domaine nommé SSD qui est riche en hélices alpha dont le rôle n'est pas encore bien connu. Les protéines AAA+ comprennent deux domaines (bleu et rouge) qui varient selon la fonction de chaque famille des protéines AAA+ (Chamieh, 2005).

Les structures cristallographiques des hexamères AAA⁺ sont connues pour HslU, FtsH, LonB ClpX et Cdc48/P97, et les structures des modules AAA⁺ en sous-unités simples sont connues pour ClpA, LonA et PAN (Sauer & Baker, 2011). Ces structures montrent que les conformations des domaines AAA⁺ sont hautement conservées et l'ATP ou l'ADP se lie dans une serrure entre le grand et le petit domaine AAA⁺ d'une sous-unité et le grand domaine AAA⁺ d'une sous-unité voisine (Fig. 19). 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 avec les changements conformationnels dans l'anneau hexamérique (Sauer & Baker, 2011). Il y a six sites potentiels de liaison d'ATP/ADP dans un anneau hexamérique constitué d'un seul module AAA⁺ (Fig. 19 & 20), mais des études en solution pour ClpX, HslU et PAN montrent qu'un maximum de quatre sites sont occupés jusqu'à saturation nucléotidique (Hersch et al, 2005). En outre, des études sur les hexamères de ClpX liés d'une manière covalente indiquent que l'hydrolyse des nucléotides dans une seule sous-unité est suffisante pour conduire les changements conformationnels nécessaires pour le dépliement, la translocation et la dégradation des protéines (Martin et al, 2005). Ces résultats suggèrent que le cycle ATPase commence par une enzyme avec quatre ATP liées, le produit après hydrolyse est une enzyme avec trois ATP et une ADP, puis retourne à l'état initial après la dissociation de l'ADP et la fixation de l'ATP.

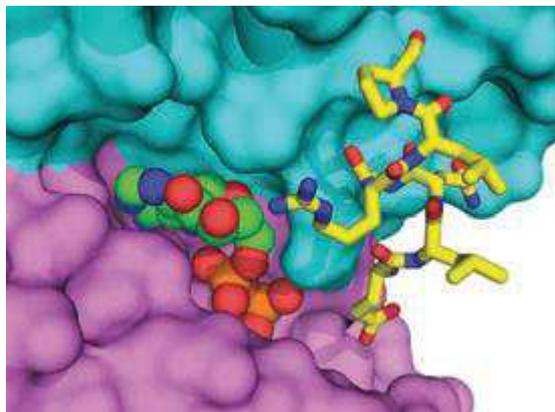


Figure 19 : liaison du nucléotide.

La poche pour la liaison ATP/ADP est située entre le grand (magenta) et le petit (cyan) domaine AAA⁺ d'une sous-unité. Des parties de ces domaines sont indiquées par une représentation de surface du complexe HslU d'*E. coli* (PDB code 1HQY, Chaîne E); l'ADP est indiqué par une représentation « space-filling ». La poche de liaison se compose essentiellement d'acides aminés de la boîte II (résidus 18-22) et du motif Walker A (résidus 57-66) dans le grand domaine ainsi que la première boucle/hélice (résidus 335/343) et du motif sensor II (résidus 390-393) du petit domaine. Des contacts additionnels avec le nucléotide du grand domaine d'une sous-unité voisine (résidus 321-327; représentation en stick) aident à la coordination des communications entre les sous-unités (Sauer & Baker, 2011).

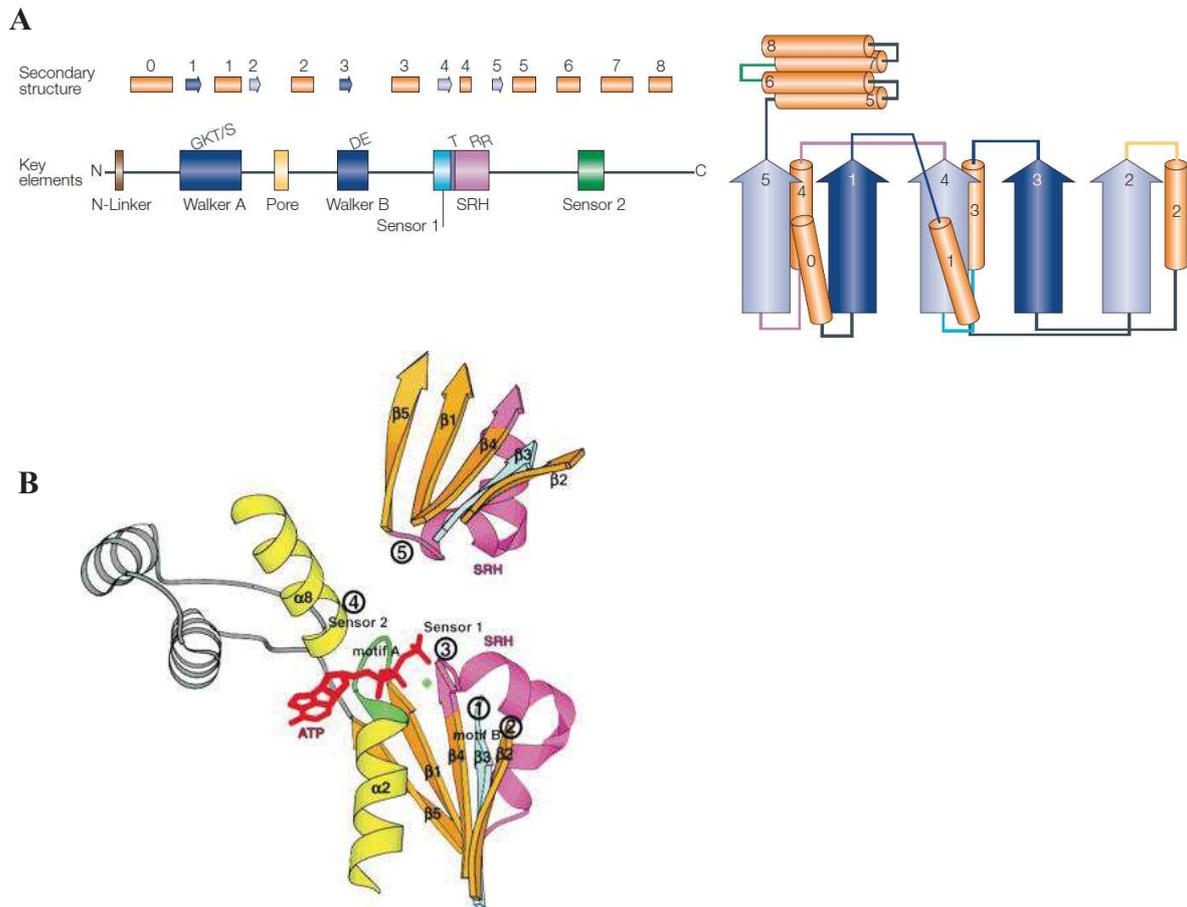


Figure 20 : Organisation d'un domaine AAA+.

(A) (Gauche) La structure secondaire et les éléments principaux d'un domaine AAA+ sont représentés dans un diagramme linéaire. Les positions approximatives d'éléments clés - le N-linker, Walker A et Walker B, Sensor 1 et Sensor 2 et la deuxième région d'homologie (SRH) - sont présentées sous la position des hélices (rectangles) et feuilletts β (flèches). (Droite) un diagramme topologique généralisé pour un domaine AAA+. La coloration des éléments de structure secondaire est compatible avec les éléments de l'image à gauche en haut et la coloration des boucles entre les hélices et les feuilletts est cohérente avec la coloration des éléments de l'image à gauche en bas (Hanson & Whiteheart, 2005). (B) Anatomie d'un site de liaison de l'ATP dans un AAA+ ATPases. La figure montre un traçage en « ribbon » des éléments de structures secondaires sélectionnés à partir d'une paire de molécules adjacentes de l'hexamère NSF-D2 (PDB 1D2N). L'adénine et son ion magnésium associé sont en rouge et vert respectivement. Les walkers A et B sont colorés en vert et cyan, respectivement. Le domaine SRH est en magenta. Les différents segments de la structure qui contribuent au groupe catalytique et/ou sensor des protéines AAA+ ou des familles de protéines liées sont indiqués par des nombres encadrés. Les hélices et les boucles autour du noyau de la structure β ont été omises pour plus de clarté (Ogura & Wilkinson, 2001).

Comme indiqué ci-dessus, dans certains membres de la famille AAA+ le domaine ATPasique existe en deux copies, par exemple ClpA et ClpC contiennent deux modules AAA+ (désignés D1 et D2). La microscopie électronique et la cristallographie ont montré que les modules D1 et D2 forment des anneaux distincts (Kessel et al, 1995) (Guo et al, 2002). Cet arrangement double-lobé est aussi une caractéristique de la famille des chaperonnes ClpB/Hsp100. Bien que la raison d'avoir deux modules AAA+ dans ces enzymes n'est pas claire, des expériences biochimiques suggèrent que les deux anneaux ATPases peuvent fonctionner indépendamment

comme des protéines unfoldases avec le module D2 étant un unfoldase plus puissant (Kress et al, 2009).

Une autre étape du cycle protéolytique est la liaison des ATPases à leur partenaire protéolytique. Les complexes ATPasiques régulateurs de la protéolyse lient le complexe protéolytique sous forme hexamérique à l'une ou deux extrémités (Fig. 21), et les deux complexes communiquent alors via un canal central comme l'ont montré les images par microscopie électronique du complexe PAN-20S (Smith et al, 2005), Cdc48-20S (Barthelme et al, 2014) et la structure cristallographique du complexe HslU-HslV (Sousa et al, 2000). La liaison des protéases aux complexes ATPases est un processus dynamique et requiert la présence du Mg^{2+} et de l'ATP (Smith et al, 2005) ainsi que dans le cas où le complexe ATPasique est déjà lié à son substrat, l'affinité du complexe ATPasique pour son partenaire protéolytique augmente considérablement (Joshi et al, 2004) (Azim et al, 2005).

Des études cristallographiques et biochimiques soulignent l'importance de la région C-terminale du complexe ATPasique comme étant le domaine impliqué dans l'interaction avec son partenaire protéolytique (Seong et al, 2002) (Song et al, 2000). Le domaine C-terminal contient un motif conservé « hydrophobe-tyrosine-X (Hb-Y-X) » avec X représentant n'importe quel acide aminé, nécessaire pour la formation du complexe entre la particule régulatrice et la particule catalytique (RP-CP) et pour l'ouverture de la porte de la particule catalytique (Smith et al, 2007). La comparaison des séquences C-terminales de PAN dans divers archées ainsi que du 19S ATPase eucaryote montre que l'avant dernier résidu est une tyrosine dans PAN de 9 espèces d'archée différents (dans 5 autres espèces d'archée c'est une phénylalanine), cette tyrosine se retrouve également dans les quatre sous-unités ATPasiques du complexe 19S : Rpt1, Rpt2, Rpt3 et Rpt5 (mais pas dans Rpt4 et Rpt6) chez les humains, les rats, les souris, la drosophile, l'arabidopsis, les nématodes (sauf Rpt1) et la levure. Un autre résidu conservé est un résidu hydrophobe qui précède la tyrosine dans tous les PAN des 14 espèces d'archée ainsi que dans les Rpt2, Rpt3, Rpt5 et Rpt6 chez les eucaryotes. Ainsi le motif C-terminal conservé, HbYX, où Hb est le résidu hydrophobe, Y la tyrosine (ou phénylalanine chez certains archées) et X un résidu variable apparaît dans presque tous les protéasomes ATPases connus (Smith et al, 2007). Une telle séquence ne se trouve pas dans la partie C-terminale des activateurs ATP-indépendants, PA26 et PA28 α , β ou γ . Le motif HbYX est nécessaire pour la formation du complexe PAN-20S, des mutations dans le motif HbYX pourraient bloquer la porte d'ouverture soit en empêchant la formation du complexe PAN-20S, soit en permettant sa formation, mais en interférant le mécanisme d'ouverture de la porte.

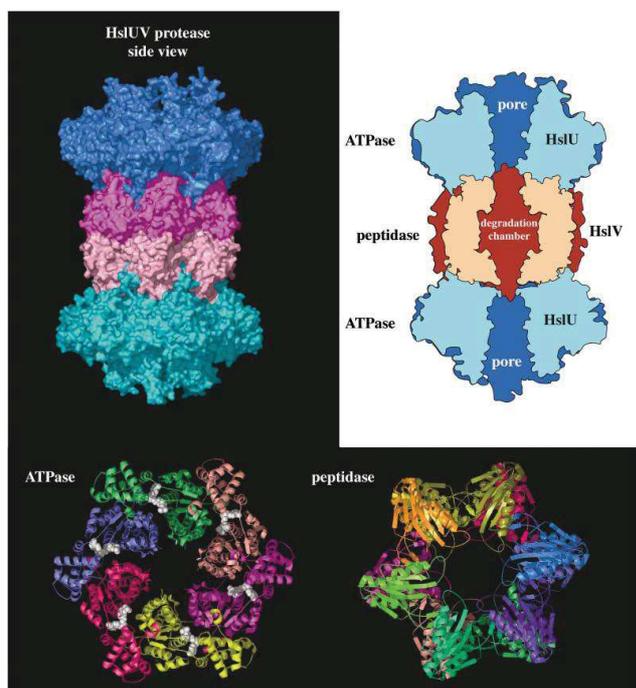


Figure 21 : Vue structurale de la protéase ATP-dépendante HslUV.

La figure est basée sur la structure signalée dans Sousa et al., 2000 (Sousa et al, 2000) et représente la structure cristallographique résolue de HslU-HslV. Le panneau en haut à gauche montre le modèle de surface de HslUV. Le complexe ATPasique HslU est coloré en bleu/vert et HslV coloré en rose/pourpre. En haut à droite, une coupe longitudinale montrant l'ouverture du complexe à l'extérieur par le pore du complexe ATPasique HslU et la chambre protéolytique de HslV. Les panneaux du bas montrent une vue axiale de l'ATPase HslU et de la protéase HslV. Les molécules d'ATP sont colorés en blanc (Sauer et al, 2004).

1.5.4 Mode d'action des AAA+ unfoldases

Les protéases AAA+ peuvent être considérées comme des machines de translocation de polypeptides (Sauer & Baker, 2011). Cette activité ATP-dépendante est essentielle pour le bobinage des substrats dénaturés à travers le pore axial de l'anneau AAA+ vers la chambre de dégradation d'une protéase associée. Il semble que la reconnaissance, l'accrochage et la translocation d'un 'tag' attaché à une protéine repliée (tel le degron « ssrA » présenté en chapitre 1.4.3) fournisse un moyen d'appliquer une force de dépliement en essayant de tirer une protéine de grande taille à travers un canal étroit (Sauer & Baker, 2011). Des cycles de liaison et d'hydrolyse de l'ATP provoquent des changements dans la conformation de l'anneau AAA+, qui sont transmis vers le substrat, au moins en partie, par des boucles (« pore loops ») qui se projettent dans le pore axial et qui contiennent un motif constitué par une chaîne latérale aromatique suivie par une chaîne latérale hydrophobe (Fig. 22). Le remplacement du résidu aromatique dans cette séquence avec une petite chaîne latérale n'empêche pas l'hydrolyse de l'ATP, mais élimine effectivement la dégradation ATP-dépendante dans de nombreux AAA+ protéases (Song et al, 2000) (Siddiqui et al, 2004) (Graef & Langer, 2006).

Après qu'un segment peptidique du substrat est engagé par le pore axiale d'une AAA+ unfoldase, quelques cycles de translocation tirent la protéine contre l'anneau hexamérique et l'entrée du pore (Sauer & Baker, 2011). A ce stade, la prochaine étape de translocation

appliquera une force de dépliement à l'élément structural attaché au 'tag'. Plusieurs résultats sont alors possibles : a) la structure se déplie en coopération permettant de procéder la translocation et la dégradation, b) la protéine résiste le dépliement et se dissocie c) la protéine résiste le dépliement, la machine de translocation est bloquée et le substrat reste lié.

Les « pores loops » jouent également des rôles importants dans la reconnaissance du substrat, et l'établissement de leur fonction est crucial dans la compréhension des défauts de translocation/dépliement. Pour ClpX par exemple, tels défauts ont été observés après mutation du résidu aromatique en seulement deux des six sous-unités dans un hexamère lié par des liaisons covalentes (Martin et al, 2008). A des concentrations saturantes en substrat, les taux de translocation et de dépliement sont plus lents pour le mutant lorsque la mutation des pores est dans une sous-unité qui pourra hydrolyser l'ATP. Le coût en hydrolyse d'ATP pour la translocation d'un polypeptide déplié augmente trois fois pour le mutant par rapport au wild type, alors que le coût de dépliement d'une protéine native est 18 fois plus élevé (Sauer & Baker, 2011) . Ces résultats suggèrent que de nombreux coups d'alimentation (« power stroke ») dans l'enzyme mutée n'ont pas réussi à déplacer le substrat vers la particule catalytique.

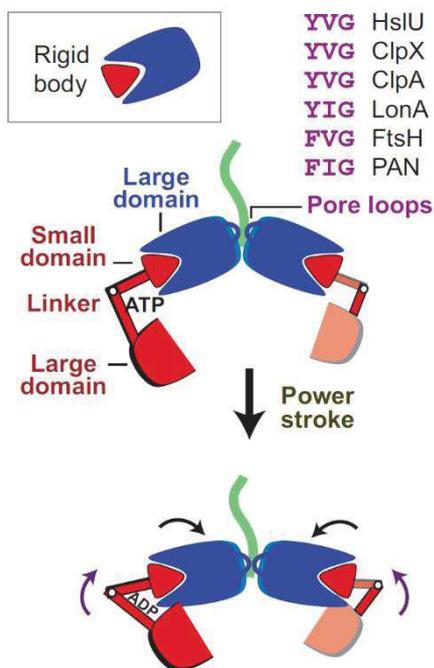


Figure 22 : des boucles hautement conservées dans le pore axial de l'unfoldase entrent en contact avec le substrat pour la translocation.

Des séquences homologues YVG, YIG, FVG, FIG, sont trouvées dans les boucles du pore axial « pore loops » de différentes unfoldases AAA+ (HslU, ClpX, ClpA, LonA, FtsH et PAN). La puissance de dépliement semble être causée par des changements nucléotide-dépendant de la rotation entre le grand et le petit domaine AAA+ d'une sous-unité ce qui entrainerait des mouvements de l'anneau AAA+ entier et la translocation de la chaîne polypeptidique (Sauer & Baker, 2011).

La capacité des AAA+ unfoldases à déplier les protéines natives est remarquable (Kim et al, 2000) (Kenniston et al, 2003). Par exemple, la GFP (Green Fluorescent Protein) est dépliée en solution avec une demi-vie de plusieurs années à température ambiante, alors que certaines AAA+ unfoldases peuvent la déplier dans quelques secondes (Zhang et al, 2009b).

Néanmoins, certaines protéases AAA+ n'arrivent pas à déplier et donc dégrader certaines protéines mêmes lorsqu'elles portent des tags de dégradation accessibles, suggérant que certaines unfoldases AAA+ sont plus puissantes que d'autres (Koodathingal et al, 2009) (Herman et al, 2003). En plus, la force de dépliement d'une enzyme spécifique peut aussi dépendre de la stabilité du substrat. Par exemple, ClpXP déplie la GFP beaucoup plus vite que Lon le fait, alors que c'est l'inverse pour la titine (Kim et al, 2000) (Gur & Sauer, 2008). Certaines de ces différences peuvent être liées à la capacité de l'enzyme à déplacer rapidement une partie du substrat après un premier événement de dépliement.

Le taux de dépliement/dégradation de certains substrats modèles par une AAA+ protéase est linéairement en corrélation avec le taux d'hydrolyse de l'ATP (Martin et al, 2008). 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%. Ce comportement indique que l'hydrolyse d'un seul ATP commence à être déjà impliquée dans le processus de dépliement. Cependant, la dégradation de certains substrats exige un taux minimum ou un seuil d'hydrolyse d'ATP, suggérant que de multiples événements d'hydrolyse coordonnés sont nécessaires avant que le dépliement se produise (Martin et al, 2008). Dans ce cas-là, il est possible qu'un seul mouvement déplie seulement une partie du substrat, qui nécessite une translocation rapide et une force ultérieure sur la structure restante pour un dépliement total. Les taux auxquels les différentes AAA+ protéases hydrolysent l'ATP au cours du dépliement de substrats varient considérablement (Koodathingal et al, 2009) (Kenniston et al, 2003) et cette variation peut expliquer comment une enzyme peut apparaître à être plus puissante qu'une autre pour le même substrat. Pour Lon, l'identité du 'tag' de dégradation « degron » peut également modifier la vitesse à laquelle un substrat natif est déplié (Gur & Sauer, 2009).

Comme indiqué ci-dessus, seulement quelques sous-unités de l'anneau hexamérique des AAA+ ClpX, HslU et PAN lient l'ATP à un moment donné. Reste à savoir si ces sous-unités peuvent se convertir en sous-unités libres et vice versa pendant que l'hydrolyse de l'ATP progresse et si d'autres hexamères AAA+ protéolytiques contiennent des sous-unités qui lient l'ATP et d'autres non. Plusieurs résultats démontrent la communication entre les sous-unités AAA+, y compris la coopérativité positive en liaison et hydrolyse d'ATP (Yakamavich et al, 2008). Des expériences utilisant un variant de ClpX avec une seule chaîne et différentes combinaisons de sous-unités actives et inactives indiquent que l'hydrolyse de l'ATP dans une seule sous-unité peut conduire au dépliement et translocation des protéines (Martin et al, 2005).

Le mécanisme exact expliquant comment les mouvements des boucles alimentés par l'hydrolyse de l'ATP entraînent la translocation du substrat n'est toujours pas établi et nécessitera probablement des études biophysiques plus sophistiquées en temps réel pendant l'hydrolyse de l'ATP et le dépliement/translocation du substrat.

1.6 Le complexe AAA+ PAN des archées : une version simplifiée du 19S des eucaryotes

Les études des complexes protéolytiques ATP-dépendantes des archées et des bactéries se sont avérées très précieuses pour comprendre la structure et les mécanismes enzymatiques du protéasome 26S eucaryote (Voges et al, 1999), notamment celles du protéasome 20S des archées et de son complexe de régulation ATPase, PAN, ont été très informatifs. Il semblerait que ce système ancestral ait évolué avant que la dégradation des protéines chez les eucaryotes soit devenue liée à l'ubiquitination pour améliorer la sélectivité et la régulation du processus (Maupin-Furlow et al, 2006). En raison de la conservation des propriétés clés du complexe 19S eucaryote chez le complexe PAN des archées, son étude offre de nombreux avantages pour l'élucidation de la fonction du protéasome 26S eucaryote.

En 1996, le séquençage complet de l'archée méthanogène *Methanococcus jannaschii* a révélé la présence de deux gènes (nommés S4 et S8) qui sont des homologues à des gènes codant pour le 19S ATPases (Bult et al, 1996). Pour tester s'ils régulent la protéolyse en se servant de l'ATP, le gène S4 a été exprimé dans *E. coli* et le produit de 50 kDa (PAN) a été purifié et caractérisé par Zwickl et al. (1999) (Zwickl et al, 1999). La séquence primaire de PAN contient un seul domaine AAA (résidus 200 à 342), y compris plusieurs caractéristiques de cette famille d'ATPase: un motif P-loop (qui comprend les motifs Walker A et B) et un motif SRH (« Second Region of Homology ») du côté C-terminal (Fig. 23). PAN partage 41-45% de similarité de séquence avec les sous-unités du 19S ATPases humain et de la levure (Zwickl et al, 1999). Comme vu par alignement de séquence avec les sous-unités Rpt 1-6 du 19S de la levure, un certain nombre de résidus conservés dans PAN et le 19S ATPases se trouvent dans les motifs P-loop et SRH (Fig. 24). PAN et les 19S ATPases ont aussi un motif « coiled-coil » à leur extrémité N-terminale (Zwickl et al, 1999). En 2005, une étude sur le complexe PAN de l'archée halophile *H. salinarium* a été lancée au sein de l'équipe ELMA. Les résultats obtenus ont indiqué la présence de deux isoformes du complexe PAN chez *H. salinarium* avec des modifications au niveau de l'extrémité N-terminale, les deux isoformes (PAN A et PAN B) ont montré des affinités différentes pour l'association avec le protéasome 20S ce qui peut

jouer un rôle dans la régulation de la protéolyse (Chamieh et al, 2008). D'autres études ont également montré qu'il y a une régulation transcriptionnelle positive des sous-unités α et β du protéasome ainsi que des deux ATPases régulateurs du protéasome, PAN A et PAN B, comme réponse au stress salin ou thermique des cellules de *H. salinarium* (Chamieh et al, 2012).

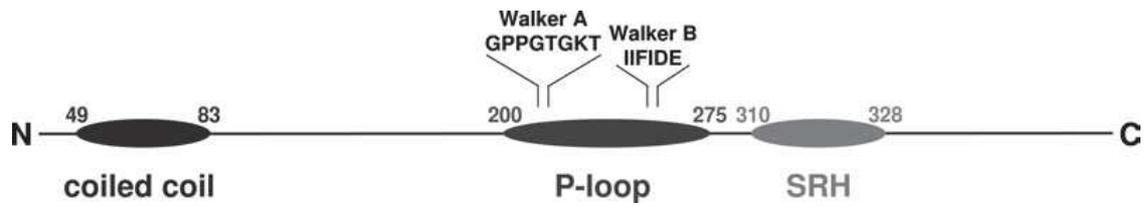


Figure 23 : représentation schématique de la séquence primaire de PAN.

PAN présente 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 » (Zwickl et al, 1999).

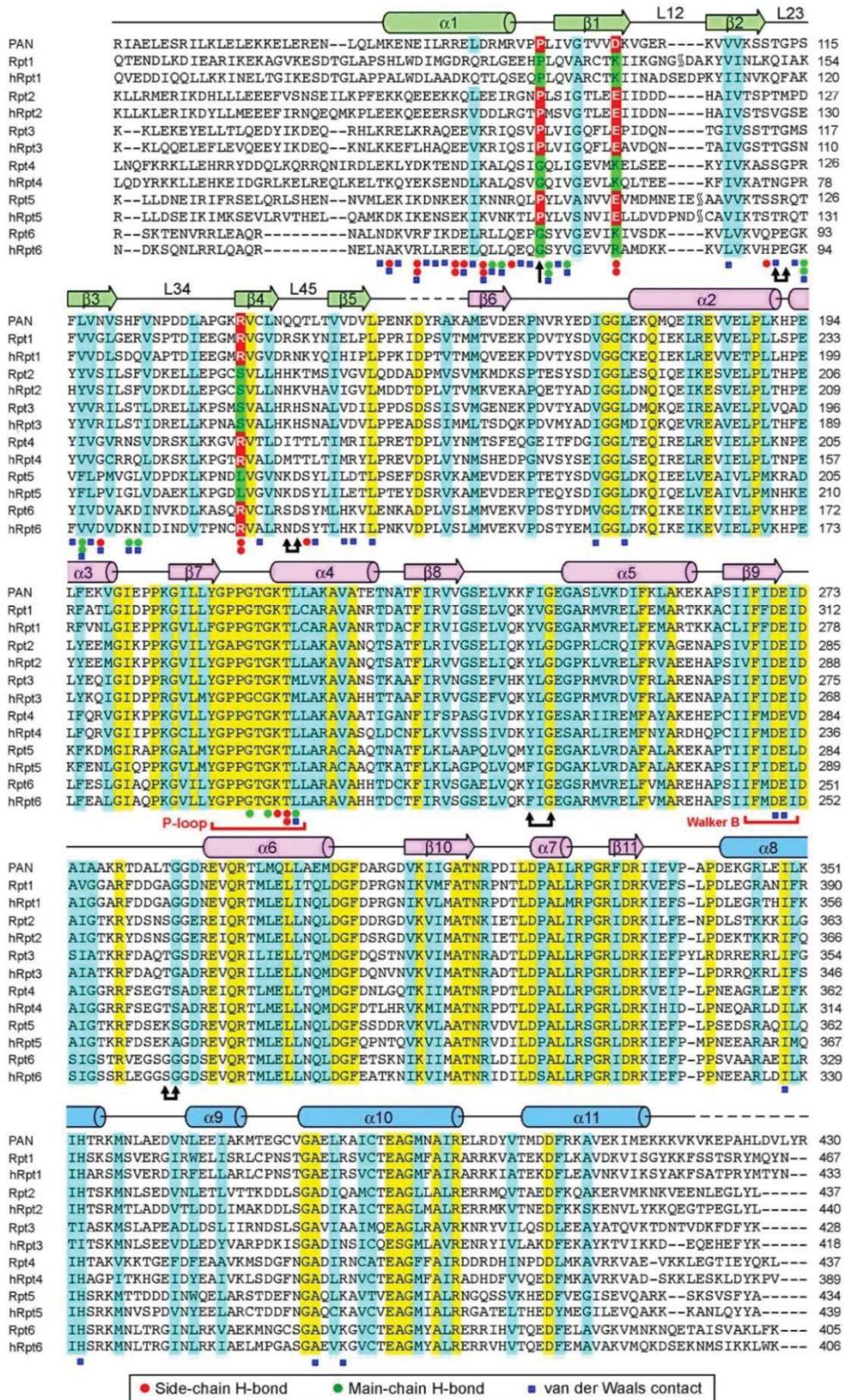


Figure 24 : alignement de séquence entre PAN et les AAA+ ATPases des protéasomes de la levure et humaines.

Les séquences de Rpt1 jusqu'à Rpt6 de levure sont dérivées de *S. cerevisiae*. Les éléments de la structure secondaire de PAN sont indiqués au-dessus des séquences. Les acides aminés invariants et conservés sont colorés en jaune et cyan respectivement. Les doubles flèches indiquent les résidus prédits pour aligner le canal de translocation de substrat. La boucle située entre $\beta 8$ et $\alpha 5$ est connue comme pore loop 1 (pore-1 loop, ou Ar- Φ loop) et celle entre $\alpha 8$ et $\alpha 9$ comme pore loop 2 (pore-2 loop). Les boucles du domaine OB sont appelés L23 (Zhang et al, 2009a).

Lorsque le complexe PAN est purifié jusqu'à homogénéité, il existe en solution sous forme d'un complexe homo-oligomérique qui se présente sous forme d'un anneau hexamérique (Wilson et al, 2000) et possède une activité ATPase Mg^{2+} -dépendante (Zwickl et al, 1999). Dans sa caractérisation initiale, PAN en présence de l'ATP, favorise la dégradation sélective des protéines qui manquent de structure tertiaire, comme la caséine et la RNase A oxydée (Zwickl et al, 1999). Par la suite, il a été montré que PAN participe à la dégradation des protéines globulaires telles que la GFP par une réaction nécessitant l'hydrolyse de l'ATP (Benaroudj & Goldberg, 2000). En fait PAN, par elle-même, pourrait catalyser le dépliement ATP-dépendant des protéines globulaires stables. Pour l'instant, aucun processus de dépliement ATP-dépendant a été démontré avec des particules 19S pures ou des complexes 26S.

Bien que l'association d'un complexe ATPase chaperon-like avec une particule protéolytique semble être une caractéristique commune à plusieurs systèmes de dégradation intracellulaire de protéine (le protéasome 26S et les complexes bactériens ClpAP, ClpXP, HslUV), une association entre PAN et la particule 20S a été difficile à observer par des approches biochimiques standards (Benaroudj et al, 2007). Cette difficulté à démontrer un tel complexe a 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 (Förster et al, 2003). 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) (Smith et al, 2005). Smith et ses collaborateurs (Smith et al, 2005) ont démontré l'association entre PAN et le protéasome 20S d'archée en utilisant l'immunoprécipitation, la résonance de plasmons de surface (SPR) et la microscopie électronique (Fig. 25). 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'est observée en présence d'ADP ou de n'importe quel autre nucléotide. Donc, l'association de PAN avec le protéasome 20S est favorisée lors de la liaison d'ATP et réduite par hydrolyse de l'ATP.

Des études de microscopie électronique sur le complexe PAN-20S ont démontré que PAN peut s'associer au 20S à une ou aux deux extrémités (Fig. 25). PAN apparaît alors comme une structure à deux anneaux, avec un grand anneau intérieur et un anneau extérieur plus petit et ressemble à un 'chapeau' qui couvre le cylindre 20S. La structure du complexe PAN-20S est remarquablement similaire à un protéasome 26S sans les composants du couvercle du 19S et les sous-unités non-ATPase de la base (Fig. 25D). En effet, les images du complexe PAN-20S

peuvent être exactement superposées sur les densités dans le complexe 26S. Ces études montrent que PAN s'associe physiquement avec les protéasomes 20S pour que la dégradation ATP-dépendante des protéines puisse se produire. La dégradation des tri- ou térapeptides par les protéasomes 20S des archées 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. Cette incapacité à stimuler la dégradation de petits peptides distingue PAN du complexe 19S qui stimule la dégradation ATP-dépendante de petits peptides (Köhler et al, 2001). Cependant Smith et ses collaborateurs (Smith et al, 2005) ont montré que PAN et l'ATP stimulent la dégradation des peptides de sept résidus ou plus. Donc, un pore fermé existe chez les particules catalytiques d'archée qui permet l'entrée des petits peptides, mais les peptides plus longs comme les protéines sont exclus en absence de PAN et de l'ATP (Benaroudj et al, 2003).

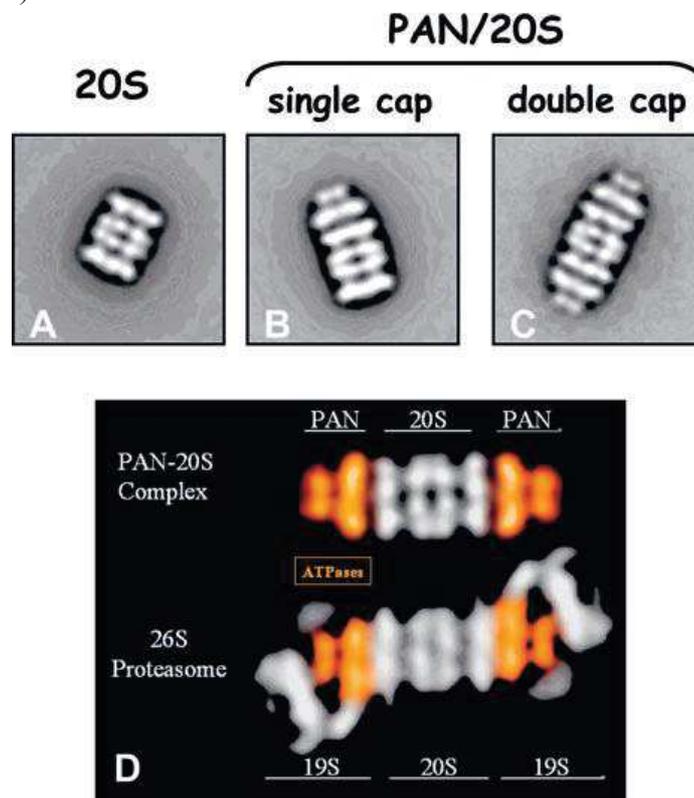


Figure 25 : analyse par microscopie électronique du complexe PAN-20S et du complexe 20S d'archée. Des images du protéasome 20S d'archée (A), du complexe PAN-20S d'un seul côté du 20S (B) et du complexe PAN-20S des deux côtés du 20S (C). (D) Des particules négativement colorées du complexe PAN-20S (deux côtés) sont comparées au protéasome 26S mammifère (Smith et al, 2005).

Par combinaison des informations structurales et biochimiques, une vue complète de la façon dont les composants ATPases du protéasome fonctionnent dans le dépliement des protéines et la translocation dans la particule catalytique pour la dégradation commence à être plus clair : en se basant sur des images de microscopie électronique du complexe PAN:20S et par

spectrométrie de masse des complexes PAN purifiés, on constate que PAN fonctionne en tant qu'un complexe hexamérique (Smith et al, 2005) avec son domaine C-terminal du côté de la face cylindrique des quatre heptamères des sous-unités α et β qui forme la partie catalytique (Fig. 25). 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. Le côté N-terminal de PAN est positionné de l'autre côté avec des domaines en « coiled coils » (CC) (une de chaque sous-unité) associés par paires, pour former trois domaines allongés pour recruter les protéines substrats (ou éventuellement des partenaires protéiques). Le domaine OB (« oligonucleotide/oligosaccharide-binding fold ») forme une plateforme stable et un pore relativement étroit (~15 Å de diamètre) pour fonctionner comme un filtre et restreindre l'accès aux protéines repliées d'entrer dans le canal central de l'ATPase (Fig. 26).

Certaines études suggèrent que PAN déplie les protéines à sa surface en l'absence de translocation (Navon & Goldberg, 2001). Des travaux plus récents soutiennent un modèle dans lequel les protéines sont dépliées par une translocation énergie-dépendante à travers l'anneau ATPase qui peut être couplée au passage de la protéine dépliée dans la partie catalytique pour la destruction (Zhang et al, 2009a) (Matouschek, 2003) (Fig. 26). En accord avec ce modèle, la boucle Ar- Φ (où Ar pour un acide aminé aromatique quelconque et Φ pour un acide aminé hydrophobe) également connu pour « pore-1 loop » ligne le passage centrale du domaine AAA+ et définit un des points de constriction à proximité du domaine OB dans le domaine ATPase (Zhang et al, 2009a) (Matouschek, 2003). Similairement à d'autres protéines AAA+, la boucle Ar- Φ de PAN facilite le dépliement des protéines en tirant les protéines substrats dans le canal ATPase par un mouvement ATP-dépendant du domaine AAA+ (Martin et al, 2008; Zhang et al, 2009a). Le domaine OB exerce une force passive pour le dépliement des protéines en fournissant un pore avec une plateforme stable qui bloque le mouvement de la structure des protéines repliées à travers le pore pour que les boucles Ar- Φ tirent la protéine (Zhang et al, 2009a). Enfin du compte, la protéine dépliée est transportée pour la dégradation par les sites actifs protéolytiques protégés à l'intérieur de la chambre centrale de la particule catalytique.

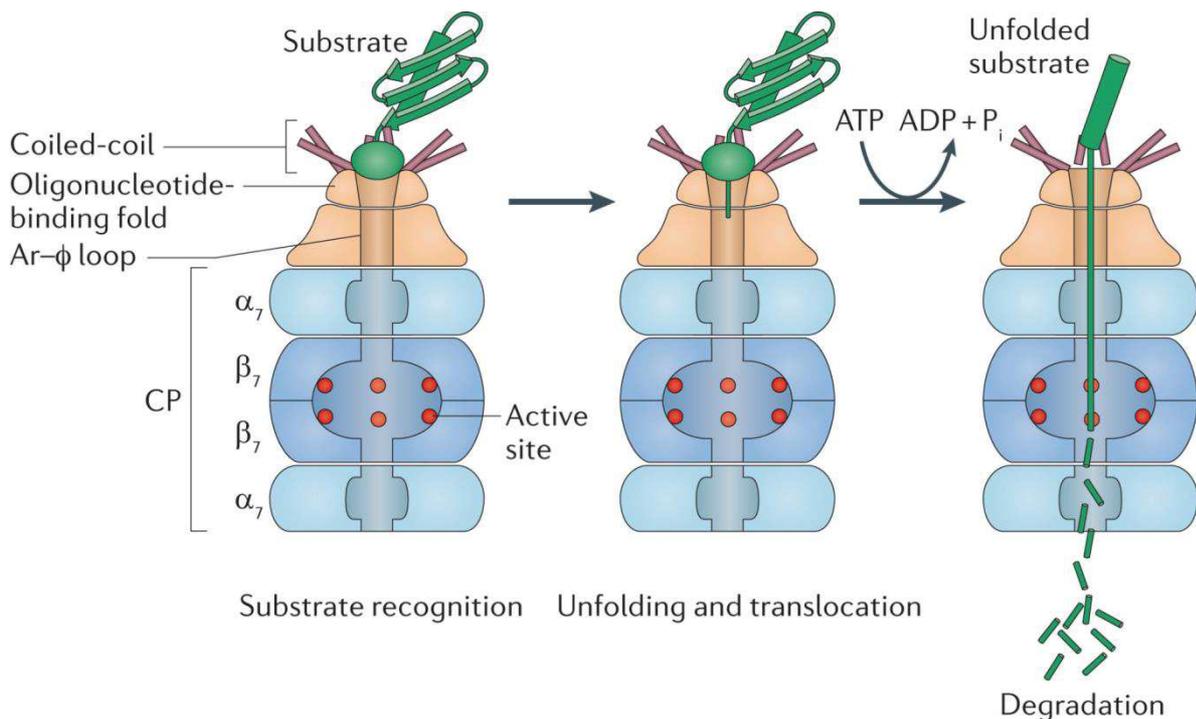


Figure 26: Modèle de protéolyse basé sur le complexe d'activation du protéasome (PAN) archéen et la particule catalytique 20S (CP).

Le domaine « coiled-coil » du côté N-terminal de chaque sous-unité de PAN forme un pair 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 (« oligonucleotide/oligosaccharide binding fold ») qui peut servir comme point d'entrée pour les substrats protéiques à parcourir le canal ATPase. Une boucle aromatique-hydrophobe (Ar – Φ) au sein du canal ATPase peut tirer les substrats protéiques vers le bas avec un processus piloté par l'hydrolyse de l'ATP. Le dépliement de la protéine se produit avec la répétition de ces cycles d'hydrolyse et avec le domaine OB fournissant une plateforme rigide (Maupin-Furlow, 2011).

La compréhension des mécanismes de fonction du protéasome nécessite des éclaircissements sur sa structure tridimensionnelle, en particulier celle de la particule régulatrice, qui est considérablement plus complexe que la particule catalytique. Les structures des particules catalytiques ont été résolues pour *T. acidophilum* (Lowe et al, 1995), *Archaeoglobus fulgidus* (Groll & Clausen, 2003), *Mycobacteria tuberculosis* (Hu et al, 2006), *S. cerevisiae* (Groll et al, 1997), et quelques mammifères (Unno et al, 2002). Les structures des deux composants de la particule régulatrice eucaryote Rpn13 (Schreiner et al, 2008) et le C-terminal de Rpt3 (Nakamura et al, 2007), sont également connus. Toutefois, de nombreux efforts restent à fournir pour résoudre la structure atomique du 19S eucaryote et de son homologue chez les archées, PAN. A 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, il n'y a toujours pas de données structurales complètes sur le complexe PAN assemblé. Les seules données structurales qui existent correspondent à des images de microscopie électronique par coloration négative (Smith et al, 2005) (Fig. 25), à un

modèle Cryo-EM à basse résolution (20 Å) (Medalia et al, 2009) montrant PAN sous sa forme dodécamérique ainsi qu'un modèle pseudo-atomique généré par la superposition de la structure atomique des domaines N et C-terminal de PAN avec l'hexamère de HslU bactérien (Zhang et al, 2009a) (Fig. 27). Malgré les informations structurales apportées par les modèles proposés antérieurement, un modèle structurel fiable du complexe PAN sous sa forme complète hexamérique est toujours inexistant.

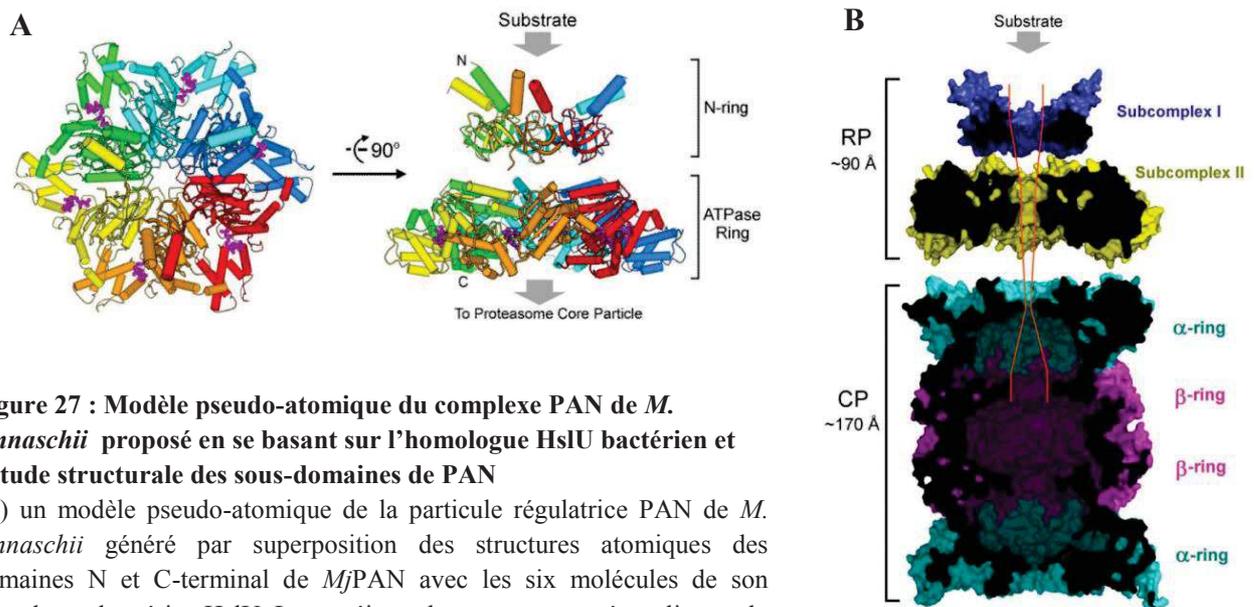


Figure 27 : Modèle pseudo-atomique du complexe PAN de *M. jannaschii* proposé en se basant sur l'homologue HslU bactérien et l'étude structurale des sous-domaines de PAN

(A) un modèle pseudo-atomique de la particule régulatrice PAN de *M. jannaschii* généré par superposition des structures atomiques des domaines N et C-terminal de *MjPAN* avec les six molécules de son homologue bactérien HslU. La protéine substrat est supposée se lier sur la face distale comme indiqué. (B) une vue complète du protéasome de *M. jannaschii*. Le modèle du protéasome complet est coupé au milieu pour montrer le chemin vers la chambre catalytique dans les anneaux β . La mise en place des domaines N-ter et C-ter de PAN et leur position par rapport à la particule catalytique 20S sont basées sur les images de microscopie électronique du complexe PAN-20S de Smith et al. (Smith et al, 2005) (Zhang et al, 2009a).

1.7 Problématiques et objectifs de la thèse

La combinaison de plusieurs études structurales et biochimiques a éclairci les mécanismes de base de plusieurs enzymes AAA+ unfoldases impliquées dans le dépliement et/ou le remodelage des protéines. Dans la plupart des cas, et malgré la large gamme des techniques et des systèmes biologiques utilisées (notamment bactériens), il n'est pas encore connu, au niveau structural détaillé, comment les protéines cibles sont initialement reconnues par les enzymes et comment les changements de conformations couplés à la liaison et à l'hydrolyse de l'ATP ainsi qu'à la libération du produit (ADP/Pi) peuvent conduire au dépliement et à la translocation des protéines substrats. Une structure à haute résolution (ou quasi-atomique) du complexe 19S régulateur du protéasome eucaryote ou de son homologue PAN archéen est

toujours manquante ce qui empêche notre compréhension de la façon dont ces régulateurs ATPasiques affectent et régulent la fonction des complexes catalytiques 20S. Les problèmes rencontrés au niveau de la préparation des échantillons et de l'assemblage des complexes ont ralenti l'apparition des données structurales interprétables sur les complexes régulateurs du protéasome eucaryote et archéen.

Des structures à haute résolution par Cryo-EM et cristallographie pourraient alors répondre à certaines de ces questions et éclaircir certains aspects dans la relation structure-fonction des régulateurs de la protéolyse. Cependant, et vu l'aspect « statique » de ces techniques, il sera nécessaire de développer et d'appliquer des techniques biophysiques en temps réel en solution pour suivre simultanément la liaison et l'hydrolyse de l'ATP ainsi que les changements conformationnels et l'activité mécanique pendant le dépliement des substrats protéiques par les AAA+ unfoldase.

Les facteurs qui déterminent si les AAA+ protéases dégradent progressivement ou libèrent des substrats partiellement processés sont encore mal compris. En outre, le mécanisme de dépliement utilisé n'est toujours pas connu et si c'est vraiment en « power-stroke » (comme proposé pour les systèmes Clp bactériens), ou si une seule enzyme AAA+ unfoldase peut utiliser plusieurs mécanismes alternatifs. Au final, notre connaissance du répertoire des substrats naturels et des « degrons » reconnus par le régulateur AAA+ d'archée (PAN) est limitée, de même que notre compréhension sur la régulation de la machine et des substrats par des adaptateurs ou par des réactions de modifications.

Ma thèse s'inscrit dans la thématique de la protéolyse intracellulaire et de la biophysique des grands assemblages moléculaires. Elle a pour objectifs d'apporter des informations biochimiques, fonctionnelles et surtout structurales sur les complexes d'activation du protéasome (PAN) des cellules d'archées hyperthermophiles *Pyrococcus horikoshii* et *Methanococcus jannaschii* par l'utilisation d'une approche de biologie structurale intégrative en combinant de la microscopie électronique (coloration négative et Cryo-EM) avec de la diffusion de rayons X aux petits angles (SAXS) et de la cristallographie (voir les chapitres 3 et 4 de la thèse pour la présentation des techniques utilisées). Le premier but était de déchiffrer les bases structurales et moléculaires du complexe PAN responsable du dépliement des substrats protéiques pour préparer leur dégradation par la partie catalytique du protéasome (20S). Un deuxième objectif de mon sujet de thèse était d'utiliser une approche élégante et innovatrice de diffusion de neutrons aux petits angles (SANS) en temps réel, combinée avec des mesures de fluorescence en ligne (voir chapitre 2). Le but de cette deuxième partie était de suivre 'en direct' le processus de dépliement du substrat modèle utilisé, la GFP_{ssrA}, par PAN

afin de mieux comprendre les changements de conformations ainsi que la dynamique du système nécessaire pour le dépliement des substrats.

1.7.1 Objectif 1: étude structurale et fonctionnelle du complexe PAN hexamérique

La première tâche de cette partie du projet de thèse était de maîtriser l'état oligomérique et fonctionnel du complexe PAN. Une étude structurale du complexe PAN thermostable de *Pyrococcus horikoshii* permettait de comprendre la fonction de son homologue 19S eucaryote plus complexe. Des analyses des extraits cellulaires des organismes hyperthermophiles *P. horikoshii* et *Pyrococcus abyssi* ont montré la présence d'une forme tronquée du côté N-terminal de PAN exprimée *in vivo*. Cette forme tronquée du complexe a été exprimée et purifiée d'une façon recombinante dans *Escherichia coli* sous sa forme hexamérique active (voir chapitre 5.1). Nos données biochimiques montrent que la forme tronquée possède aussi une activité de dépliement contre la GFPssrA, le substrat modèle utilisé.

La deuxième tâche dans cette partie de la thèse était la caractérisation structurale du complexe PAN hexamérique de *P. horikoshii* (*PhPAN*). Pour cela, les propriétés structurales du *PhPAN* hexamérique ont été analysées par cryo-microscopie électronique (Cryo-EM) et par diffusion de rayons X et de neutrons aux petits angles (SAXS/SANS) en solution. Une reconstruction 3D de PAN a été générée à 13 Å de résolution. Pour mieux comprendre le rôle du nucléotide dans les changements de conformation de *PhPAN*, deux autres modèles Cryo-EM à basse résolution ont été générés en présence ou en absence de l'ATP γ S. Finalement, et pour générer un modèle pseudo-atomique du complexe, les sous-domaines de *PhPAN* ont été caractérisés par cristallographie aux rayons X (domaine C-terminal ATPase monomérique) et par SAXS (domaine N-terminal hexamérique) en solution. La combinaison de ces approches structurales intégratives a permis la construction d'un modèle pseudo-atomique de *PhPAN* qui porte beaucoup d'informations sur le mécanisme d'action de cette grande machinerie cellulaire. Les résultats de cette première partie de la thèse font l'objet d'un manuscrit en révision dans le journal « Structure » et sont présentés comme chapitre 5.1 de la thèse.

1.7.2 Objectif 2 : développement d'une approche de diffusion de neutrons aux petits angles (SANS) en temps réel pour étudier le processus de dépliement des protéines par les complexes AAA+ ATPases

In vitro, après interaction avec le substrat protéique par reconnaissance du peptide signal ('degron' ou 'degradation tag'), PAN assure par un processus ATP-dépendant le dépliement des protéines avant leur translocation dans la chambre catalytique du protéasome *via* un canal étroit (Fig. 26). Vu l'absence d'une structure à haute résolution du complexe PAN, les mécanismes de ses fonctions sont très peu compris. Des nano-machines similaires à PAN ont été caractérisées biochimiquement et par microscopie électronique (Sauer et al, 2004) mais le mécanisme de dépliement ATP-dépendant reste peu connu: quels sont les motifs structuraux reconnus sur la protéine à déplier? Existe-t-il des changements de conformations du complexe PAN nécessaires pour assurer le processus de dépliement? Comment la forme structurée du substrat évolue pendant le dépliement par PAN et dans quel état le substrat est relâché après dépliement? Quel est le rôle structural et conformationnel joué par le nucléotide (ATP/ATP γ S) pour alimenter la réaction de dépliement et quel est le rôle du tag *ssrA* dans le mécanisme? Dans cette deuxième partie de ma thèse, une approche de SANS en temps réel a été développée pour aborder ce type de questions.

Le système PAN étudié dans cette partie du projet provient de l'archée hyperthermophile *Methanococcus jannaschii*. Le système est très stable à température ambiante comme expliqué ci-dessus et facile à purifier dans des milligrammes nécessaires pour les études biophysiques par SANS. Un avantage supplémentaire par rapport au *PhPAN* est la stabilité et la solubilité jusqu'à des concentrations élevées (>20 mg/ml) très utiles du point de vue technique pour la bonne préparation des expériences SANS.

Dans le cadre de ce projet le choix de la technique SANS pour l'étude du processus de dépliement était essentiel sur plusieurs points : 1) la diffusion aux petits angles des neutrons est l'une des rares techniques (à part la RMN et SAXS) qui donne des informations structurales et dynamiques (en termes de changements conformationnels) sur des molécules biologiques en solution. 2) Comme le système étudié est composé de deux partenaires protéiques, SANS permet de les distinguer par deutération et variation de contraste, ce qui nous a permis de suivre les changements conformationnels de la GFP*ssrA* et de PAN individuellement pendant le dépliement.

Un objectif méthodologique important de ma thèse était de coupler des mesures de fluorescence avec les mesures SANS en temps réel (de l'ordre de quelques secondes) sur la ligne D22 de l'Institut Laue-Langevin (ILL), permettant de suivre les changements de conformation soit de la GFPssrA soit de PAN ainsi que la diminution de la fluorescence de la GFPssrA sur le même échantillon. Ce type de mesure a été mis au point spécialement pour cette étude et a été validé pour la première fois dans ce projet. Les résultats seront soumis prochainement et sont présentés comme manuscrit en préparation (voir chapitre 5.2). Cette approche pourra être utilisée dans le futur sur d'autres projets utilisant des protéines fluorescentes (ou marquées par des molécules fluorescentes) pour des études dynamiques et fonctionnelles en temps réel d'une variété de machines moléculaires (chaperonne, unfoldase...).

2 DIFFUSION DE NEUTRONS AUX PETITS ANGLES (SANS)

2.1 Théorie

2.1.1 Le neutron

Le neutron a été découvert en 1932 par James Chadwick (Chadwick, 1932) et a été utilisé pour la première fois comme sonde en 1946 par Clifford Shull. Les neutrons avec les protons forment les noyaux de tous les atomes dans la nature (à l'exception l'atome ^1H). Les neutrons sont des particules subatomiques, sans charge électrique avec une masse légèrement supérieure à celle des protons. Ils ont une masse équivalente à 1839 électrons (1.674928×10^{-27} kg) et une durée de vie de 15 minutes (885.9 s) hors du noyau et possèdent une longueur d'onde qui suit la relation suivante:

$$\lambda = \frac{h}{mv} \quad (1.1)$$

Où h est la constante de Planck, m la masse du neutron et v sa vitesse.

La mécanique quantique dit que le neutron est une particule mais possède également une nature ondulatoire qui lui permet de se comporter comme une onde, ce qui fait de lui un puissant outil d'investigation, y compris pour la réflexion, la diffraction et la spectroscopie. Étant également une particule son énergie cinétique est donnée par la relation :

$$E = \frac{1}{2} mv^2 = \frac{\hbar^2 k^2}{2m} \quad (1.2)$$

Avec k le nombre d'onde ($k = 2\pi / \lambda$) l'énergie est aussi donnée par l'expression suivante:

$$E = \frac{\hbar^2}{2m\lambda^2} \quad (1.3)$$

En raison de leur neutralité électrique, les neutrons peuvent pénétrer dans la matière. Lorsque les neutrons sont produits par un réacteur, les neutrons utilisés pour l'étude de la matière molle sont modérés par une source froide ce qui leur confère une longueur d'onde qui se situe typiquement entre 3 et 25 Å correspondant à une gamme d'énergie située entre 0,1 et 10 meV (Lindner, 2002). Pour information, les rayons X possèdent eux une énergie de l'ordre du keV, soit 10^6 ordres de grandeur de plus. Du fait de cette gamme d'énergie, les neutrons permettent à la fois de sonder les structures par des phénomènes d'interférence mais également la dynamique par spectroscopie en échangeant une fraction importante de leur énergie avec l'échantillon.

2.1.2 Interaction neutron-noyau et section efficace

Prenons la diffusion élastique d'un faisceau de neutrons à partir d'un seul noyau. Dans ce cas nous traitons le noyau comme étant rigidement fixe à l'origine des coordonnées sans échange d'énergie avec le faisceau incident (Fig. 28).

Nous pouvons représenter le faisceau de neutrons avec une onde plane

$$\psi_i = e^{ikz} \quad (1.4)$$

Ou z est la distance du noyau dans la direction de propagation. L'onde diffusée sera sphérique avec la fonction d'onde suivante (k' est le nombre d'onde sortante):

$$\psi_s = -\frac{b}{r} e^{ik'r} \quad (1.5)$$

La longueur de diffusion b des noyaux varie de façon aléatoire avec le tableau périodique. Elle varie aussi entre les isotopes d'un même élément. Un exemple utile est le ^1H et le ^2H (hydrogène et deutérium respectivement). L'hydrogène a une longueur de diffusion cohérente de $-3,74 \times 10^{-5}$ Å et le deutérium $6,67 \times 10^{-5}$ Å. Ainsi la longueur de diffusion d'une molécule peut être modifiée par le remplacement de l'hydrogène par le deutérium et donc potentiellement moduler son signal par rapport à d'autres composants dans le système (voir paragraphe suivant). Cette technique de variation de contraste est l'un des principaux avantages de la diffusion des neutrons par rapport aux rayons X.

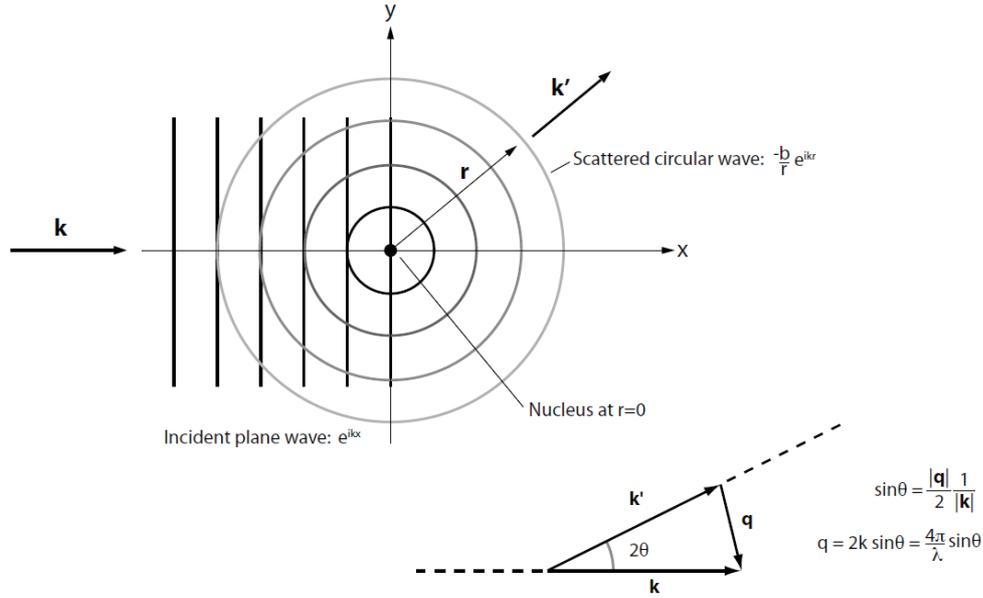


Figure 28 : Diffusion élastique des neutrons d'un noyau fixe (Pynn, 1990).

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

$$\psi_s = -\sum_i \left(\frac{b_i}{r}\right) e^{ik'r} e^{iq \cdot r} \quad (1.6)$$

$q = k' - k$ est le vecteur de diffusion avec k et k' les vecteurs d'onde des neutrons incidents et diffusés ($k = k'$ pour la diffusion élastique).

La section efficace de diffusion σ est une mesure de la taille d'un atome tel qu'il apparait pour un neutron et mesure donc l'efficacité avec laquelle les neutrons sont diffusés par le noyau. Dans une expérience de diffusion de neutrons, un faisceau de neutrons d'énergie E est incident sur un groupe d'atomes (une solution de protéines par exemple) (Fig. 29). En supposant une diffusion élastique (l'énergie des neutrons ne change pas au cours de l'interaction), nous pouvons mettre en place un détecteur pour compter les neutrons diffusés dans l'angle solide $d\Omega$ dans la direction θ, ϕ (Fig. 29). La section efficace différentielle est défini par:

$$\frac{d\sigma}{d\Omega} = \frac{\text{nombre de neutrons diffusés par sec dans } d\Omega \text{ dans la direction } \theta, \phi}{\phi d\Omega} \quad (1.7)$$

Où ϕ est le nombre de neutrons incidents par unité de surface par seconde, appelé flux incident. La section efficace de diffusion totale est définie par l'équation :

$$\sigma_s = \frac{\text{nombre total de neutrons diffusés par sec}}{\phi} \quad (1.8)$$

La section efficace différentielle d'un assemblage d'atomes est :

$$\frac{d\sigma}{d\Omega}(q) = \frac{1}{N} \left| \sum_i^N b_i e^{iq \cdot r_i} \right|^2 \quad (1.9)$$

Par intégration de la distribution de la densité de longueurs de diffusion à travers la totalité de l'échantillon et en normalisant par le volume de l'échantillon :

$$\frac{d\Sigma}{d\Omega}(q) = \frac{N}{V} \frac{d\sigma}{d\Omega}(q) = \frac{1}{V} \left| \int \rho(r) e^{iq \cdot r} dr \right|^2 \quad (1.10)$$

Ce résultat est connu comme l'équation de 'Rayleigh-Gans' et nous montre que la diffusion aux petits angles vient d'une hétérogénéité de la densité de longueurs de diffusion $\rho(r)$. $\Sigma = \sigma/V$ est la section efficace macroscopique. L'intégrale est la transformée de Fourier de la distribution de la densité de longueurs de diffusion et la section efficace différentielle est proportionnelle au carré de son amplitude. Ce dernier fait signifie que toutes les informations de phase sont perdues et qu'il est impossible d'effectuer la transformée de Fourier inverse pour obtenir la densité de longueurs de diffusion à partir de la section efficace macroscopique qui est une fonction du vecteur de diffusion, q .

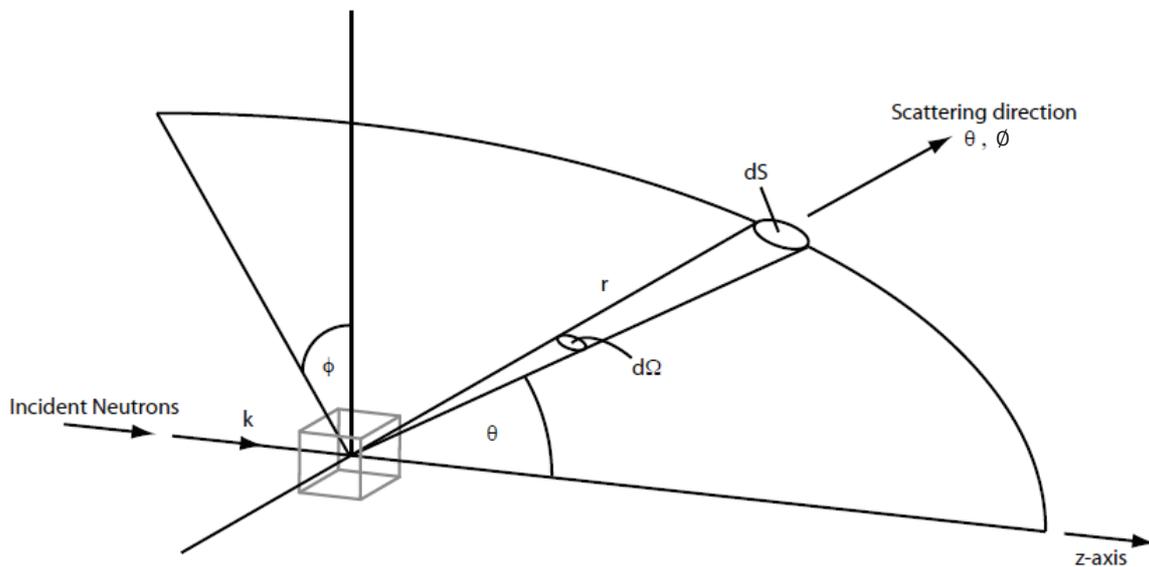


Figure 29 : Géométrie d'une expérience de diffusion (Squires, 2012).

2.2 Diffusion de neutrons aux petits angles en mode statique et variation de contraste

Les méthodes de diffusion aux petits angles sont des techniques permettant d'obtenir des informations sur la taille et la forme d'objets contenus dans une solution avec une sensibilité pour des macromolécules allant de quelques kDa jusqu'aux MDa et d'une dimension linéaire de 10 jusqu'à ~ 1000 Å. En biologie, elle permet d'obtenir des informations à basse résolution sur l'état d'oligomérisation et la forme des macromolécules biologiques (Petoukhov & Svergun, 2007). Deux principales techniques de diffusion aux petits angles sont utilisées en biologie structurale: la diffusion de rayons X (SAXS: « Small Angle X-ray Scattering ») et la diffusion de neutrons (SANS: « Small Angle Neutron Scattering »). Le principe de ces expériences est simple: une solution de la macromolécule à étudier est placée dans un faisceau parallèle et monochrome de neutrons (ou de rayons X). Les objets de la solution diffusent le rayonnement incident et la diffusion au voisinage du faisceau directe est enregistrée par un détecteur (Fig. 30). Dans le cas de rayons X, la diffusion résulte de l'interaction entre les nuages électroniques des atomes et les photons et dans le cas des neutrons, de leur interaction avec les noyaux des atomes. On collecte l'intensité du signal diffusé en fonction de l'angle par rapport au faisceau incident. Ce signal, qui résulte d'une moyenne sur toutes les orientations des macromolécules étudiées, donne une information sur leur taille et forme ou leur "enveloppe". En fonction de l'échantillon, ces techniques de diffusion permettent des reconstructions avec des précisions pouvant atteindre ~ 10 - 20 Å de résolution, en particulier pour des objets possédant des symétries internes (enveloppes de virus, protéines multimériques...).

Il est préférable d'effectuer les expériences de diffusion aux petits angles (SAS : « Small Angle Scattering ») avec des solutions idéales (c'est-à-dire suffisamment diluées pour éviter les interactions entre les composants individuels) des molécules individuelles ou bien des complexes de plusieurs molécules. Dans ce cas, et si l'échantillon est monodisperse, c'est-à-dire constituée d'un seul type de particules, des informations structurales sur les particules individuelles peuvent être obtenues. Cela comprend leur poids moléculaire MW et leur rayon de giration R_g (Eq. 1.14, 1.15), ainsi que leur reconstruction tridimensionnelle. Dans le cas des solutions polydisperses idéales, des valeurs moyennes de R_g et MW peuvent être calculées.

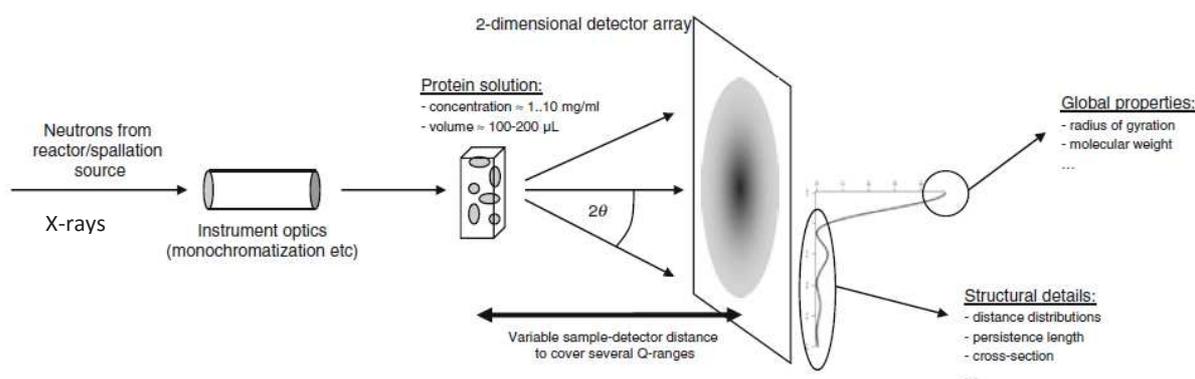


Figure 30 : Représentation schématique d'un dispositif expérimental d'une expérience de diffusion aux petits angles de neutrons ou rayons X (Gabel, 2012).

Comme expliqué ci-dessus, les neutrons ont les propriétés des ondes, comme l'amplitude et la longueur d'onde. Comme ils passent à travers la matière, des ondes secondaires sont générées par interaction avec les atomes individuels et la diffusion cohérente qui en résulte interfère d'une façon constructive ou destructive. Les clichés de diffraction des protéines dans un cristal utilisés en cristallographie sont le résultat d'une telle interférence entre des ondes diffusées par des atomes au sein des protéines qui sont régulièrement espacées dans la maille du cristal. Pour des solutions idéales (les positions des particules ne sont pas corrélées), la diffusion aux petits angles est générée par l'interférence des ondes secondaires cohérentes diffusées par des atomes situés au sein d'une seule molécule (Svergun et al, 2013). Pour interpréter des données de diffusion en termes de paramètres structuraux d'une molécule unique, le signal de diffusion doit être mesuré à partir d'un échantillon monodisperse de particules identiques. La monodispersité d'une solution protéique signifie que la protéine existe en solution sous forme d'une seule espèce oligomérique (par exemple monomère ou dimère) et que la solution protéique est libre d'agrégats et d'oligomères non-spécifiques. La préparation des échantillons est donc une étape critique importante (Fig. 31).

Le profil de diffusion est normalement décrit par l'intensité I en fonction de l'amplitude du vecteur de diffusion ou le « momentum transfer » q . En supposant que les neutrons incidents (ou les photons) n'échangent pas d'énergie avec l'échantillon, seule la diffusion élastique peut être considérée. Si k' et k sont les vecteurs d'onde du faisceau diffusé et incident respectivement, et l'amplitude du vecteur de diffusion $q = k' - k$, est alors (Fig. 28) :

$$|q| = \frac{4\pi}{\lambda} \sin\theta \quad (1.11)$$

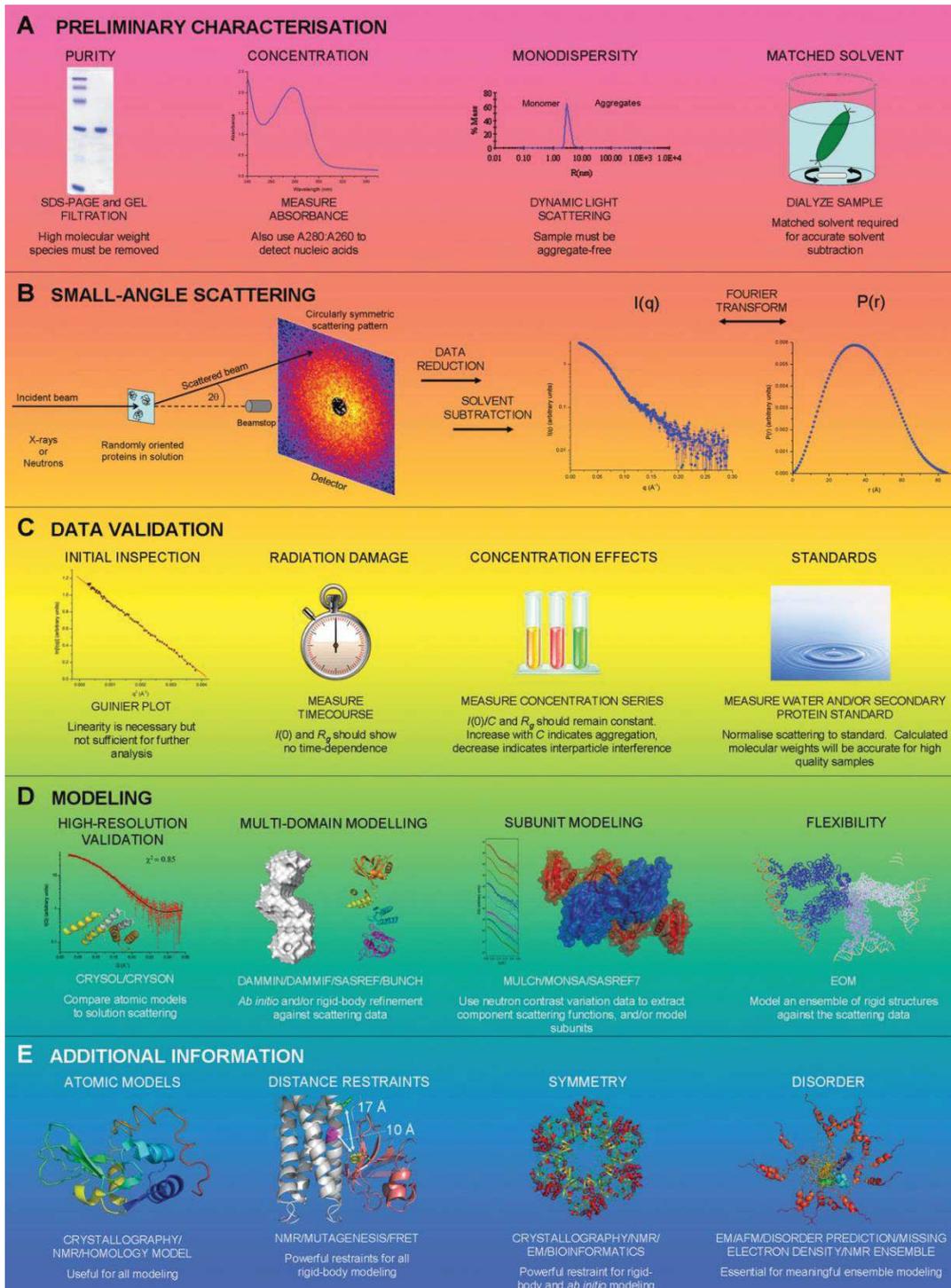


Figure 31 : un plan à suivre pour une manip de diffusion aux petits angles (Jacques & Trehella, 2010).

Où λ est la longueur d'onde du rayonnement incident, et θ est la moitié de l'angle entre le faisceau incident et diffusé (Fig. 30). La relation entre la particule et sa contribution au profile de diffusion peut être mathématiquement exprimée comme suit (Dmitri & Michel, 2003):

$$I(q) = \langle \left| \int (\rho_{protein} - \rho_{solvent}) e^{i\vec{q}\vec{r}} dV \right|^2 \rangle \quad (1.12)$$

Où $\rho_{protein} - \rho_{solvent}$ est la différence de densité de longueurs de diffusion entre la protéine et le solvant (c'est-à-dire le contraste $\Delta\rho$). L'intensité diffusée correspond à la transformée de Fourier de la densité de longueurs de diffusion des noyaux ρ_n ou de la densité électronique ρ_e . Pour une solution isotrope (les particules sont orientées d'une façon arbitraire et indépendante) l'intensité diffusée est symétrique d'une manière azimutale autour du faisceau incident. Si de plus la solution est monodisperse et idéale, l'intensité totale diffusée est simplement le produit du nombre de particules N_n par la diffusion d'une particule $i_n(q)$, moyennée sur toutes ses orientations.

$$I(q) = \sum_n N_n \langle i_n(q) \rangle \quad (1.13)$$

A partir d'un profil de diffusion aux petits angles de bonne qualité, il y a deux paramètres relatifs à la taille et à la forme de la particule qui peuvent être calculés avec une grande précision, l'intensité directe à angle nul ($2\theta=0$), ou $I(0)$, et le rayon de giration R_g . $I(0)$ ne peut pas être mesurée expérimentalement à cause du « beamstop », cependant $I(0)$ peut être déterminé par extrapolation en utilisant l'approximation de Guinier (Eq. 1.15). $I(0)$ est proportionnelle au nombre de particules par unité de volume (N) et le carré du volume de la particule (V). Elle peut être exprimée en termes de masse par unité de volume (concentration en mg/ml), le poids moléculaire (MW), le contraste ($\Delta\rho$) et le volume partiel spécifique (v) de la particule (Jacrot & Zaccai, 1981) :

$$I(0) \sim N(\Delta\rho V)^2 = \frac{c\Delta\rho^2 v^2 MW}{N_A} \quad (1.14)$$

N_A étant le nombre d'Avogadro.

Le rayon de giration R_g de la molécule est un indicateur de sa compacité. Des objets du même volume mais avec des formes différentes ont des valeurs de R_g différentes. Le R_g donne alors une information sur la distribution de masses (ou plus précisément du contraste avec le solvant) au sein d'une particule. Un moyen utile pour déterminer les valeurs de R_g et de $I(0)$ est l'approximation de Guinier ou le « Guinier plot ». Guinier a montré qu'à très petits angles, la diffusion dépend de la taille selon une loi générale indépendante de la forme et de la symétrie de la particule (Rice, 1956).

$$I(q) = I(0) \exp\left(-\frac{1}{3} R_g^2 q^2\right) \quad (1.15)$$

Ainsi, un fit linéaire de $\ln[I(q)]$ par rapport à q^2 donne le R_g et $I(0)$ à partir de la pente et l'intersection avec l'ordonnée à l'origine, respectivement. Pour des protéines globulaires cette relation est valable pour $qR_g < 1,3$, alors que pour des formes plus allongées la limite supérieure est plus faible. La nécessité d'avoir un Guinier plot linéaire pour les protéines globulaires est un moyen rapide de diagnostic de la qualité de l'échantillon.

Une autre méthode pour déterminer R_g et $I(0)$ qui utilise le profil de diffusion entier est de les calculer à partir de la fonction de distribution de distances entre des éléments de volume à l'intérieur de la particule $P(r)$ (Dmitri & Michel, 2003). Cette fonction est calculée par une transformation inverse :

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) e^{i\vec{q}\vec{r}} dq \quad (1.16)$$

R_g est donc calculé par:

$$R_g^2 = \frac{\int P(r) r^2 dr}{2 \int P(r) dr} \quad (1.17)$$

Et $I(0)$ est proportionnel à l'air sous $P(r)$:

$$I(0) = 4\pi \int_0^{Dmax} P(r) dr \quad (1.18)$$

Afin de mesurer le signal d'un objet, il est nécessaire que celui-ci ait un contraste avec le milieu dans lequel il se trouve. Durant ma thèse, les mesures ont toujours été effectuées en solution, il était donc nécessaire que les objets que nous voulions sonder aient un contraste par rapport au solvant. Le contraste se définit par une différence de densité de longueurs de diffusion ρ . Cette densité se calcule à partir de la somme des longueurs de diffusion cohérente b_i des atomes qui composent l'objet et de son volume V :

$$\rho = \frac{\sum_i b_i}{V} \quad (1.19)$$

Le carré du contraste, qui intervient dans l'intensité diffusée (Eq. 1.12), s'écrit :

$$\Delta\rho^2 = (\rho_1 - \rho_2)^2 \quad (1.20)$$

Où ρ_1 est la densité de longueur de diffusion de la particule et ρ_2 celle du milieu. Les valeurs des noyaux (ρ_n) s'expriment généralement en 10^{-10} cm^{-2} (en cas des mesures SANS) et celles de la densité électronique ρ_e en $e/\text{\AA}^3$ (en cas des mesures SAXS).

Nos mesures étant toutes faites en milieux aqueux, nous donnerons ici les valeurs de densité de longueur de diffusion pour l'eau, les contributions dues aux petits solutés tels les ions de sel étant négligeables à des concentrations modérées pour le SANS (Svergun et al, 2013). L'eau légère H_2O possède une densité de longueur de diffusion de $-0.56 * 10^{-10} \text{ cm}^{-2}$ et l'eau lourde D_2O possède une densité de longueur de diffusion de $6.33 * 10^{-10} \text{ cm}^{-2}$ (Fig. 32). On

appelle le point de croisement entre la ligne de référence du solvant ($\text{H}_2\text{O}:\text{D}_2\text{O}$) et les lignes correspondant aux densités de longueurs de diffusion des macromolécules (Fig. 32) leur « contrast match point ». A ce point, les macromolécules deviennent « invisible » en moyenne.

La grande différence entre ces deux valeurs est due à l'hydrogène qui suivant son état isotopique (^1H ou ^2H) possède une longueur de diffusion négative ou positive, c'est-à-dire il y a un changement de phase du neutron diffusé par 180 degrés (ou π) entre les deux. Cette propriété fait de la diffusion de neutrons une technique très adaptée pour la mesure d'objets organiques et donc possédant beaucoup d'hydrogène comme les protéines. En effet, pour un système qui doit être mesuré en solution aqueuse, il sera judicieux de se placer dans un solvant qui possède une densité de longueurs de diffusion la plus éloignée possible de l'objet sondé afin de maximiser le contraste et donc le signal total mesuré, qui est proportionnel au carré à cette différence. Ainsi un objet hydrogéné sera mesuré dans l'eau lourde afin de maximiser son contraste.

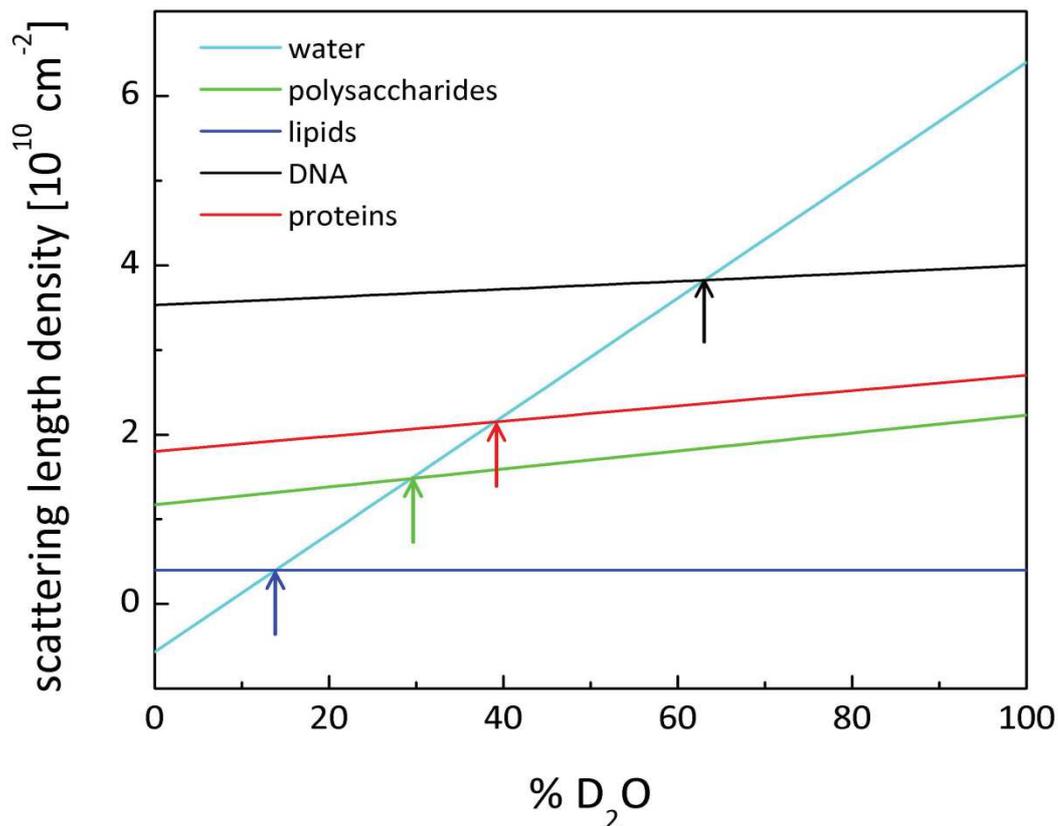


Figure 32 : la densité de longueurs de diffusion moyenne des molécules biologiques en fonction de la concentration en D_2O dans le solvant. Les conditions de « contrast match point » $\text{D}_2\text{O}/\text{H}_2\text{O}$ sont marquées par des flèches pour chaque type de molécules (Aurel Radulescu FRM II / TUM).

Nous venons de voir que le signal mesuré est directement lié au contraste neutronique entre l'objet et le solvant. Par conséquent, dans un cas de mélange à trois composants (deux protéines et le solvant par exemple), il y aura deux contrastes différents entre les deux objets et le solvant, amenant un signal contenant les contributions des deux objets par rapport au solvant et entre eux (s'ils sont en interaction). Afin de segmenter le signal il est possible d'effectuer la variation de contraste via un marquage isotopique. Nous venons de voir que l'hydrogène et le deutérium ont des longueurs de diffusion très différentes et la variation de contraste se sert de cette propriété pour distinguer des composants. La condition nécessaire pour cette expérience est que les densités de longueur de diffusion des deux objets mesurés se situent entre celles de l'eau lourde et légère et assez loin l'un de l'autre pour que le contraste soit assez élevé. Il suffit ensuite de faire un mélange D_2O/H_2O dans des proportions telles que ce mélange atteigne la densité de longueur de diffusion d'un des composants, celui-ci perdant son contraste avec le solvant, il ne reste plus que le contraste de l'autre objet envers le solvant et donc la mesure ne concernera plus que lui.

Alors quel est le résultat pratique de la discussion ci-dessus? Imaginons un système à deux phases tel que celui présenté dans la figure 33. Il se compose de deux phases incompressibles de densités de longueur de diffusion différentes ρ_1 et ρ_2 .

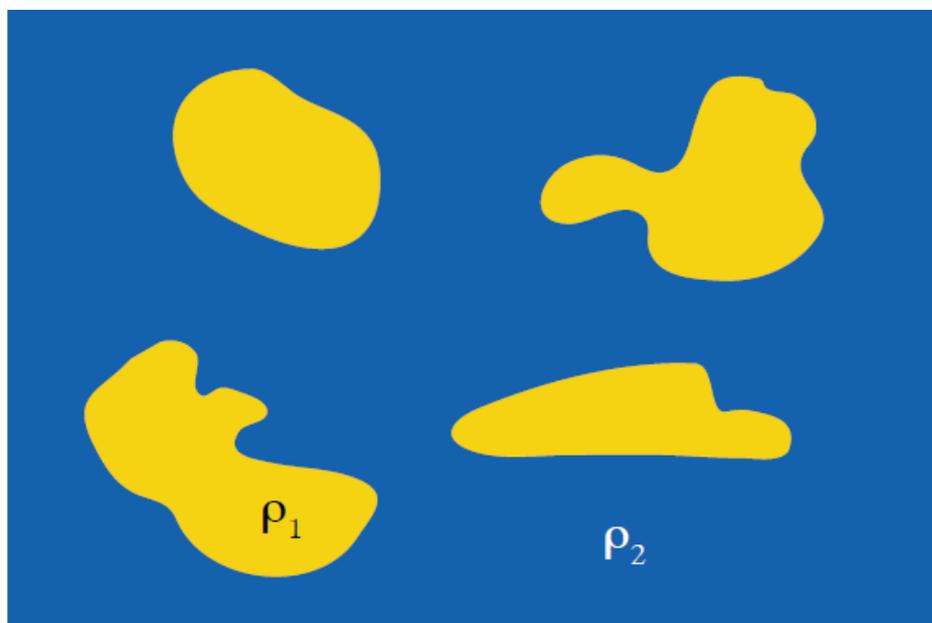


Figure 33 : un système contenant deux phases avec des densités de longueur de diffusion ρ_1 et ρ_2 .

Ainsi

$$V = V_1 + V_2 \quad (1.21)$$

$$\rho(r) = \begin{pmatrix} \rho_1 & \text{en } V_1 \\ \rho_2 & \text{en } V_2 \end{pmatrix} \quad (1.22)$$

En prenant l'équation de l'intensité de diffusion

$$I(q) = \left| \int \rho(r) e^{i\vec{q}\vec{r}} dr \right|^2 \quad (1.23)$$

Et en divisant le volume total en deux sous-volumes

$$I(q) = \left| \int \rho_1 e^{i\vec{q}\vec{r}} dr_1 + \int \rho_2 e^{i\vec{q}\vec{r}} dr_2 \right|^2 \quad (1.24)$$

$$I(q) = \left| \int \rho_1 e^{i\vec{q}\vec{r}} dr_1 + \rho_2 \left\{ \int e^{i\vec{q}\vec{r}} dr - \int e^{i\vec{q}\vec{r}} dr_1 \right\} \right|^2 \quad (1.25)$$

Donc

$$I(q) = (\rho_1 - \rho_2)^2 \left| \int e^{i\vec{q}\vec{r}} dr_1 \right|^2 \quad (1.26)$$

Prenant le cas d'un assemblage de deux molécules (protéine/protéine, protéine/ADN ou protéine/ARN) (Fig. 34 et Fig. 35).

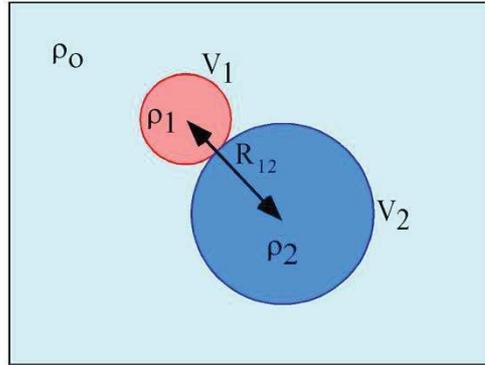


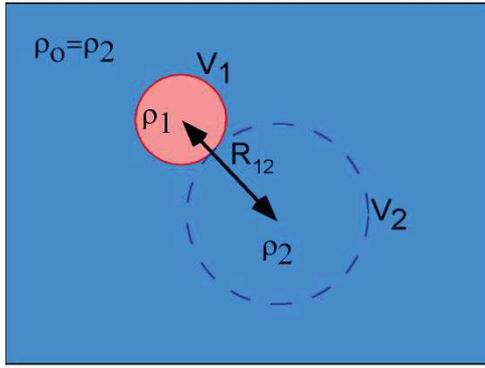
Figure 34 : exemple d'un système de complexe de deux molécules. ρ_0 (densité du solvant), ρ_1 et ρ_2 les densités des deux molécules du complexe.

$$I(q) = \left\langle \left| \int (\rho(r) - \rho_0) e^{i\vec{q}\vec{r}} dr \right|^2 \right\rangle \quad V = V_1 + V_2 \quad (1.27)$$

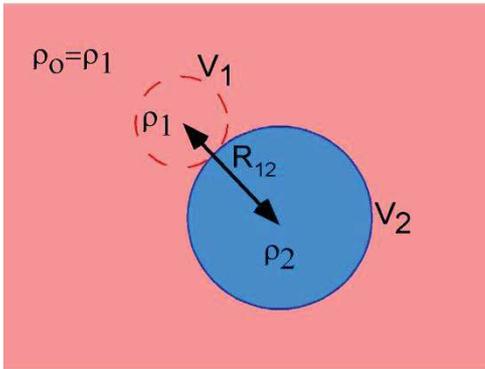
$$I(q) = \left\langle \left| (\rho_1 - \rho_0) \int e^{i\vec{q}\vec{r}} dr_1 + (\rho_2 - \rho_0) \int e^{i\vec{q}\vec{r}} dr_2 \right|^2 \right\rangle \quad (1.28)$$

$$I(q) = (\rho_1 - \rho_0)^2 \langle |F_1(Q)|^2 \rangle + (\rho_2 - \rho_0)^2 \langle |F_2(Q)|^2 \rangle + 2(\rho_1 - \rho_0)(\rho_2 - \rho_0) |F_{12}(Q)|^2 \quad (1.29)$$

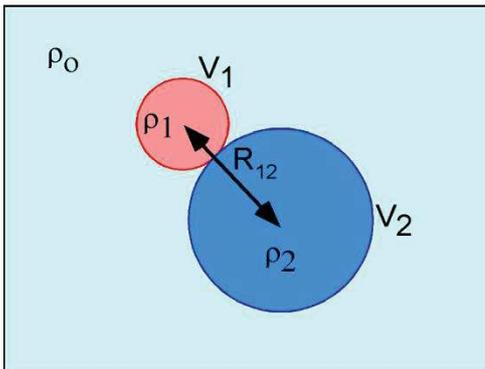
Avec $F_1(Q)$ et $F_2(Q)$ les facteurs de forme des deux molécules et $F_{12}(Q)$ le facteur de forme du complexe des deux molécules.



$$I_1(Q) = (\rho_1 - \rho_0)^2 \langle |F_1(Q)|^2 \rangle \quad (1.30)$$



$$I_2(Q) = (\rho_2 - \rho_0)^2 \langle |F_2(Q)|^2 \rangle \quad (1.31)$$



$$I(Q) = (\rho_1 - \rho_0)^2 \langle |F_1(Q)|^2 \rangle + (\rho_2 - \rho_0)^2 \langle |F_2(Q)|^2 \rangle + 2(\rho_1 - \rho_0)(\rho_2 - \rho_0) |F_{12}(Q)|^2$$

Figure 35 : les différentes possibilités de contraste et leurs équations d'intensité respectives.

2.3 Diffusion de neutrons aux petits angles en temps réel et combinaison avec la spectroscopie de fluorescence en ligne

2.3.1 Instrumentation

Sur la ligne D22 de l'Institut Laue Langevin (ILL), Grenoble, France, un faisceau de neutrons monochromatiques ($\frac{\Delta\lambda}{\lambda} = 10\%$) est collimaté et envoyé sur l'échantillon (<http://www.ill.eu/instruments-support/instruments-groups/instruments/d22>). Les neutrons sont ensuite diffusés par l'échantillon dans différentes directions (Fig. 36). Ils sont ensuite mesurés par un détecteur en fonction de leur position en deux dimensions. La ligne D22 est

équipée d'un multidétecteur (^3He) qui a une des surfaces les plus larges de toutes les lignes SANS qui existent actuellement (surface active de 1 m^2), avec une taille des pixels de $0,8 \times 0,8\text{ cm}$. Le détecteur se compose de 128 tubes linéaires verticaux (Reuter-Stokes®), disposés verticalement avec un espacement de 8 mm . Il se déplace à l'intérieur d'un tube à vide de $2,5\text{ m}$ de largeur et 20 m de longueur fournissant une distance échantillon-détecteur de $1,1$ jusqu'à $17,6\text{ m}$. D22 couvre ainsi une gamme totale de q de 4×10^{-4} jusqu'à $0,44\text{ \AA}^{-1}$ (sans détecteur décalé) ou jusqu'à $0,85\text{ \AA}^{-1}$ (avec détecteur décalé) dans les conditions standards (<http://www.ill.eu/instruments-support/instruments-groups/instruments/d22/characteristics/>).

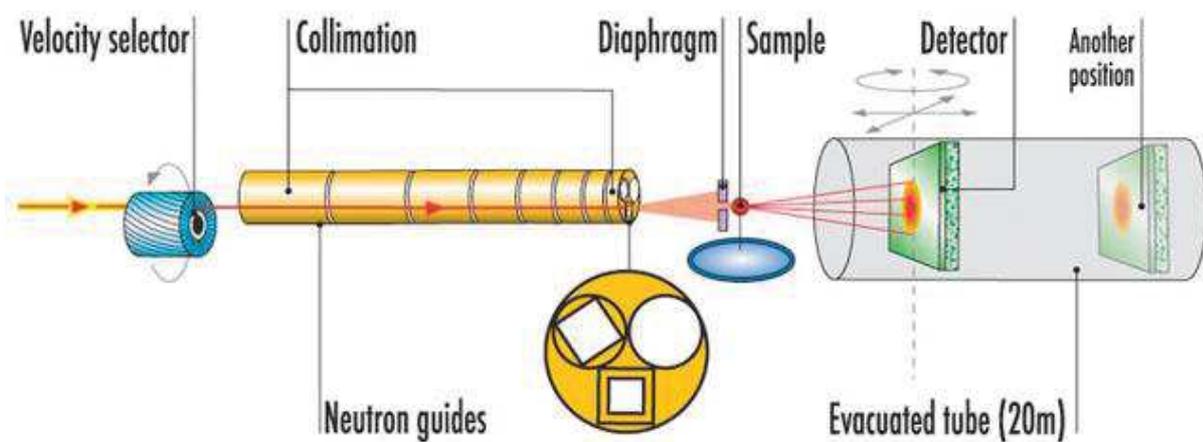


Figure 36 : Schéma du diffractomètre petits angles D22 (haut) et photos de la ligne D22 (bas) à l'Institut Laue Langevin (ILL, Grenoble).

Toutes nos expériences SANS ont été effectuées sur le diffractomètre petits angles D22 à l'ILL (sauf une expérience pour détermination du « contrast match point » sur D11). Pour les expériences standards en mode statique, dans le cas de la détermination du « contrast match point » et des mesures sur les protéines isolées (*PhPAN*, *MjPAN* et *GFPssrA*), $200\text{ }\mu\text{l}$

d'échantillon ont été pipetés dans des cuvettes de quartz (Hellma 100QS) de 1 mm de trajet optique et placés dans un support de plusieurs échantillons dont la température a été maintenue constante pendant toute l'expérience. Deux configurations de l'instrument ont été choisies, l'échantillon a été placé à une distance détecteur/collimateur de 2/2 m pour la GFPssrA, 5,6/5,6 m pour *Mj*PAN et 4/4 m pour *Ph*PAN. La longueur d'onde était de 6 Å dans tous les cas.

Pour bien contrôler le « contrast match point » et pour obtenir la valeur précise du pourcentage de D₂O dans le tampon qui permet de masquer le signal des protéines hydrogénées, les deux protéines hydrogénées (h), hPAN et hGFPssrA, ont été mesurées avec une série de différents pourcentages de D₂O dans le tampon, les valeurs de la racine carré de $I(0)$ ont été plottées en fonction du % de D₂O, et la valeur provenant de l'intersection de la droite avec l'axe des abscisses a été prise comme le « contrast match point ». Dans les deux cas les protéines hydrogénées étaient complètement masquées à 42% D₂O. Des données SANS en mode statique ont été ensuite collectées sur les protéines deutérées (d) en présence et en absence de leurs partenaires hydrogénées à 42% D₂O dans le tampon, pour comparer la superposition du signal et vérifier l'élimination complète de la contribution des protéines hydrogénées dans le signal final.

Pour les mesures SANS en mode cinétique (réaction de dépliement), 300 µl du mélange réactionnel (tables 3 et 4) ont été préparés et pipetés dans les cuvettes de quartz (Hellma 100QS) de 1 mm de trajet optique quelques secondes avant de les placer dans le support déjà ajusté à la température convenable de la cinétique enzymatique de dépliement (50, 55, 60 ou 65 °C). Le support contenant les cuvettes SANS sur la ligne D22 était thermostaté par l'intermédiaire d'un bain-marie qui couvre une large gamme de température (de 4 jusqu'à 95 °C).

Une fois la cuvette placée dans le support tempéré, la réaction de dépliement de la GFPssrA par PAN a commencé et la collecte des données SANS a été tout de suite lancée dès le début de la réaction (quelques secondes entre placer la cuvette dans le rack, fermer la porte du « hutch » et lancer le début de la collecte) avec des boucles de 30 secondes de temps d'exposition pour chaque courbe SANS enregistrée. La réaction de dépliement a été lancée pour une durée de 90 minutes résultant à un totale de 180 courbes de diffusion en mode cinétique. Deux mélanges réactionnels principaux ont été mesurés dans un milieu contenant 42% D₂O permettant d'éliminer la contribution des protéines hydrogénées (h) présentes en solution en gardant que le signal des protéines deutérées (d):

- dGFP/hPAN avec une distance détecteur/collimateur de 2/2 m.
- dPAN/hGFP avec une distance détecteur/collimateur 5,6/5,6 m.

Cinétique n°1 : dGFP/hPAN

| Protéine/ composant | Concentrations | Volume | Tampon |
|--------------------------------|-----------------------|---------------|----------------------|
| hPAN | 10 mg/ml | 150 μ l | H ₂ O |
| dGFP | 2 mg/ml | 60 μ l | D ₂ O |
| ATP/ATP γ S | 100 mM | 66 μ l | D ₂ O |
| MgCl ₂ | 100 mM | 12 μ l | H ₂ O |
| Tampon | / | 12 μ l | H ₂ O |
| Totale | / | 300 μ l | 42% D ₂ O |

Table 3: Composition de l'échantillon dGFP/hPAN. Pour le contrôle dGFP, hPAN est remplacé par du Tampon H₂O. Tampon : Tris/HCl pH 7,5 (20 mM), MgCl₂ (10 mM), NaCl (100 mM).

Cinétique n° 2 : dPAN/hGFP

| Protéine/ composant | Concentration | Volume | Tampon |
|--------------------------------|----------------------|---------------|----------------------|
| dPAN | 3 mg/ml | 90 µl | H ₂ O |
| hGFP | 4,8 mg/ml | 72 µl | H ₂ O |
| ATP/ATPγS | 100 mM | 66 µl | D ₂ O |
| MgCl ₂ | 100 mM | 12 µl | H ₂ O |
| Tampon | / | 60 µl | D ₂ O |
| Totale | / | 300 µl | 42% D ₂ O |

Table 4 : Composition de l'échantillon dPAN/hGFP. Pour le contrôle dPAN, hGFP est remplacé par du Tampon H₂O. Tampon : Tris/HCl pH 7,5 (20 mM), MgCl₂ (10 mM), NaCl (100 mM).

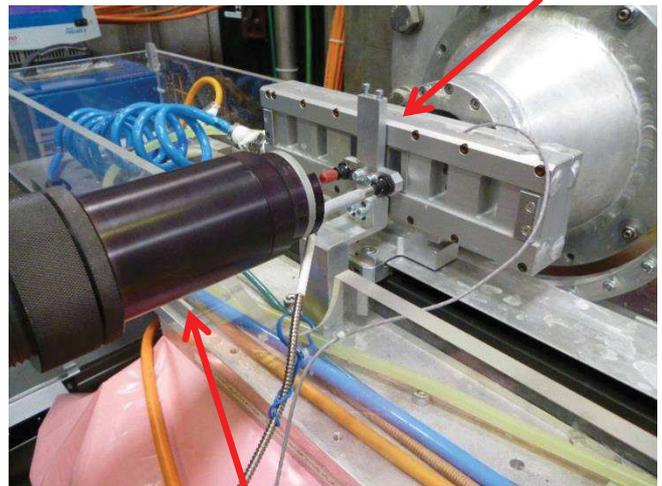
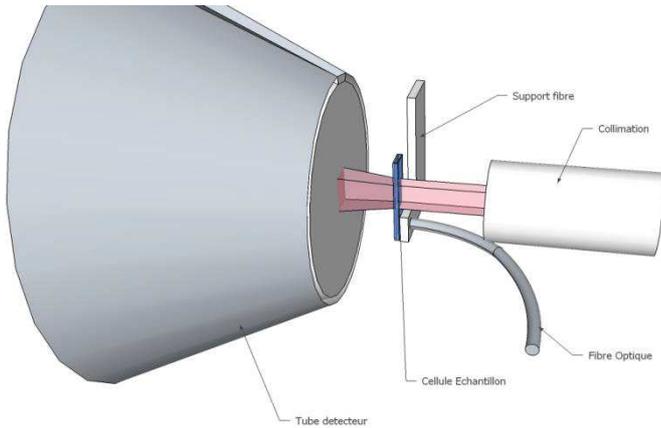
Ces deux cinétiques ont été ensuite mesurées dans plusieurs conditions: présence d'ATP ou d'ATPγS ou en absence du nucléotide; à différentes températures: 50, 55, 60 et 65 °C; en présence du tag ssrA; différentes stœchiométries PAN:GFPssrA. Dans toutes ces conditions les volumes ont été ajustés pour être à 42% D₂O dans le volume réactionnel final.

En parallèle des mesures cinétiques SANS sur *Mj*PAN, GFPssrA et leur complexe, des mesures de fluorescence en ligne ont été effectuées sur les mêmes échantillons et en même temps sur l'instrument D22. Un montage spécial d'un spectrofluorimètre *in situ* (Ocean Optics, 65000 pro) a été mis en place (Fig. 37), en collaboration avec Anne Martel la responsable de l'instrument D22: la lumière provenant d'une lampe à deutérium à travers une fibre optique (IDIL Fibres Optiques) passe d'abord au travers d'un filtre qui ne laisse passer que la longueur d'onde d'excitation souhaitée correspondant à la gamme d'absorption du fluorophore de la GFPssrA (400 nm). Une partie seulement de la fluorescence émise par la GFPssrA traverse à nouveau la fibre optique. Afin de précisément sélectionner la longueur d'onde d'émission que l'on souhaite analyser (509 nm), un deuxième filtre a été placé avant le détecteur. 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. Les fibres optiques ont été branchées sur les cuvettes SANS contenant le volume réactionnel. Les spectres d'émission de fluorescence ont été enregistrés avec une résolution de 1 spectre/seconde en même temps que les courbes de diffusion SANS. Les données de fluorescence ont été ensuite sauvegardées dans les mêmes fichiers que les données brutes SANS (fichier .nxs). Ce double setup SANS/Fluorescence nous a permis de suivre le processus de dépliage en temps réel avec des données structurales et de les corrélérer avec les pertes de la fluorescence de la GFPssrA due au dépliage.

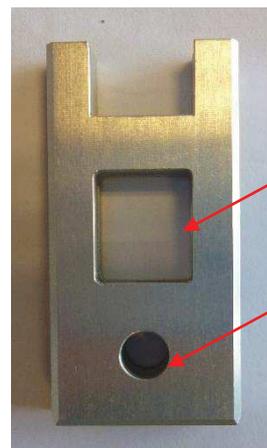
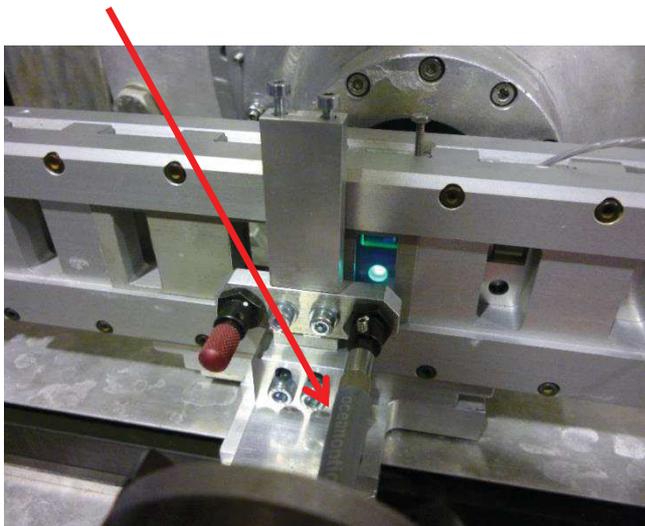
Outre les échantillons, les tampons correspondants, un standard de bore, une cuvette de quartz vide et un faisceau de neutron à vide ont été mesurés pour la réduction des données. Les temps d'exposition ont variés entre 30 s (échantillons en mode cinétique), 5 min (bore), 20 min (cuvette vide à 2 m et tampon à 2 m) et 30 min (cuvette vide à 5,6 m et tampon à 5,6 m). Les transmissions ont été mesurées pour 2 min pour chaque échantillon/tampon et pour le faisceau vide. Les temps d'expositions dans les mesures standard (« statique ») étaient de 30 minutes pour *PhPAN* et 5 minutes pour *MjPAN*.

Cuvette thermostatée



Faisceau de neutrons incidents

Fibres optiques



Faisceau neutrons

Fluorescence

Figure 37 : design et images du setup fluorimètre/SANS monté sur la ligne D22 de l'Institut Laue Langevin (ILL, Grenoble).

Le faisceau des neutrons incidents ainsi que les fibres optiques d'excitation et d'émission de la fluorescence ciblent le même échantillon pendant le processus du dépliage. Le portoir spécial utilisé pour la cuvette de quartz contenant deux trous pour les neutrons et la fluorescence est aussi montré en bas à droite.

2.3.3 Réduction et interprétation des données

Les données SANS brutes ont été réduites (par rapport à l'efficacité du détecteur, le bruit de fond électronique et la diffusion du porte-échantillon) et intégrées sur une image bidimensionnelle en utilisant le logiciel GRASP de l'ILL (Dewhurst, 2002). Les spectres d'émission de la fluorescence de la GFP_{ssrA} ont été extraits à partir des fichiers Nexus en utilisant le logiciel HDFview (<https://www.hdfgroup.org/products/java/hdfview/>).

Pour les données SANS et SAXS les intensités du tampon ont été soustraites des intensités des échantillons respectives en utilisant le logiciel PRIMUS (Konarev et al, 2003). Les rayons de girations R_g de tous les échantillons ont été extraits en utilisant l'approximation de Guinier (Guinier, 1939) (Eq. 1.15) avec le même programme. La validité de l'approximation de Guinier a été vérifiée dans chaque cas (sauf pour les agrégats de la GFP). Pour la validation des modèles pseudo-atomiques et des modèles Cryo-EM, CRY SOL/CRYSON (Svergun et al, 1995) (Svergun et al, 1998) ont été utilisés pour ajuster les courbes de diffusion théoriques calculées à partir des modèles dans les courbes SAXS/SANS expérimentales. Comme l'intensité diffusée observée dans les expériences SAS est la transformé de Fourier moyennée sur toutes les orientations de la densité de longueurs de diffusion ρ_n ou de la densité électronique ρ_e des particules dans la solution, les courbes de diffusion théoriques peuvent être calculées si ρ_n ou ρ_e sont connus. Avec l'aide des programmes CRY SOL (pour les rayons X) et CRYSON (pour les neutrons), qui font partie du paquet ATSAS (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>), les courbes de diffusion théoriques peuvent être calculées à partir des fichiers pdb. Le programme EM2DAM (Petoukhov et al, 2012) a été utilisé pour générer les fichiers pdb à partir des fichier CCP4 des cartes de densité électronique. Les courbes obtenues sont très utiles pour l'interprétation des données expérimentales ou pour vérifier si le modèle structural du fichier pdb correspond à la structure de la protéine en solution.

Pour la visualisation structurale ainsi que la validité des modèles, DAMMIF (Franke & Svergun, 2009) et DAMMIN (Svergun, 1999) ont été utilisés pour générer des modèles *ab initio* en billes en utilisant GNOM (Svergun, 1992). Les masses moléculaires et l'état d'oligomérisation des complexes ont été mesurés à partir des données SAXS par calibration relative avec une protéine standard, la BSA, à 4 mg/ml en utilisant les intensités $I(0)$ respectives (Mylonas & Svergun, 2007).

Pour l'expérience de dépliage de la dGFP_{ssrA} par hPAN (Fig. 3, chapitre 5.2), le programme OLIGOMER (Konarev et al, 2003) a été utilisé pour déterminer les fractions

volumiques de chaque espèce pendant le dépliement (agrégats et GFPssrA natives). Les facteurs de forme de chaque espèce ont été générés par fmaker (du package OLIGOMER) et utilisés par OLIGOMER pour le calcul des proportions.

2.4 Intérêt de la technique SANS dans le cas de notre projet

L'étude de la dynamique et des changements conformationnels du complexe PAN et de son substrat modèle, la GFPssrA, pendant le processus de dépliement était un objectif principal lors de ma thèse. Dans les deux dernières décennies la diffusion des rayons X aux petits angles (SAXS) a subi une énorme évolution au niveau du développement instrumentale et applications aux systèmes biologiques (Thomas et al, 2016). Toutefois, SANS était moins utilisée que SAXS et il y avait moins de développement au niveau des instruments dédiés aux applications biologiques. Vu la complexité du système PAN:GFPssrA en terme de taille (300 kDa) et de sa grande flexibilité en solution, SANS, combinée avec deutération et variation de contraste, est une des rares techniques capables de fournir des données structurales en solution en permettant de distinguer les différents partenaires. En se basant sur le principe de la variation de contraste en SANS (paragraphe 2.2) et en prenant comme exemple le système utilisé lors de ma thèse (*Mj*PAN:GFPssrA), lorsque une des deux protéines est deutérée et l'autre hydrogénée, la protéine deutérée possède la densité de longueurs de diffusion la plus élevée alors que la protéine hydrogénée possède une densité de longueurs de diffusion égale à un mélange 42% D₂O /58% H₂O dans le tampon et ainsi un tel mélange masquera donc le signal de la protéine hydrogénée et la courbe de diffusion enregistré dans ce cas correspond uniquement au signal de la protéine deutérée sans contribution de son partenaire hydrogénée en solution (Fig. 38). Ceci permet de mesurer des informations structurales sur chaque protéine dans le complexe et donc suivre sa dynamique et son changement conformationnel après formation du complexe et pendant le processus de dépliement en temps réel. Deux combinaisons du complexe (*Mj*PAN:GFPssrA) ont été effectuées, *Mj*PAN deutérée et GFPssrA hydrogénée avec 42 % D₂O dans le tampon. Dans ce cas la GFPssrA est masquée et seul le changement conformationnel de *Mj*PAN est suivie ou bien *Mj*PAN hydrogénée et GFPssrA deutérée avec 42 % D₂O dans le tampon et dans ce cas c'est le signal de *Mj*PAN qui est masqué et le changement conformationnel de la GFPssrA est suivi pendant le dépliement. Les deux combinaisons utilisées sont schématisées dans la figure 39.

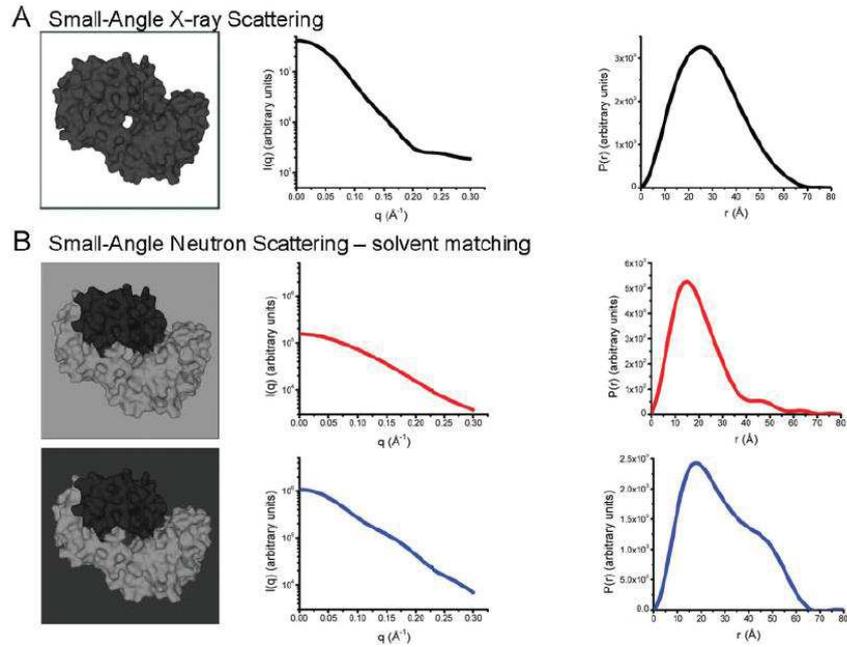


Figure 38 : diffusion aux petits angles (neutrons et rayon X) sur un complexe formé entre une protéine deutérée et une autre non-deutérée : l'expérience de la variation de contraste. La figure illustre la variation de contraste en utilisant une structure (PDB : 3GMR) d'une 'T-cell surface glycoprotein CD1d1' en complexe avec 'beta-2-microglobulin'. (A) les protéines ont chacune la même densité électronique et donc la même densité de diffusion des rayons X. Le profil de SAXS par conséquent donne des informations sur la forme du complexe entier comme une particule de contraste uniforme (représenté en noir). (B) le contraste de diffusion de neutrons des composants deutérés (noir) et non-deutérés (gris) est différent, et il est donc possible de mesurer la diffusion de chaque composant individuel par 'contrast matching' en changeant la composition H/D du solvant. La protéine non-deutérée sera 'matchée' à environ 40% D₂O, alors que la protéine deutérée sera 'matchée' à 90% D₂O, cette valeur dépend du niveau de deutération de la protéine (60% dans cet exemple) (Jacques & Trewella, 2010).

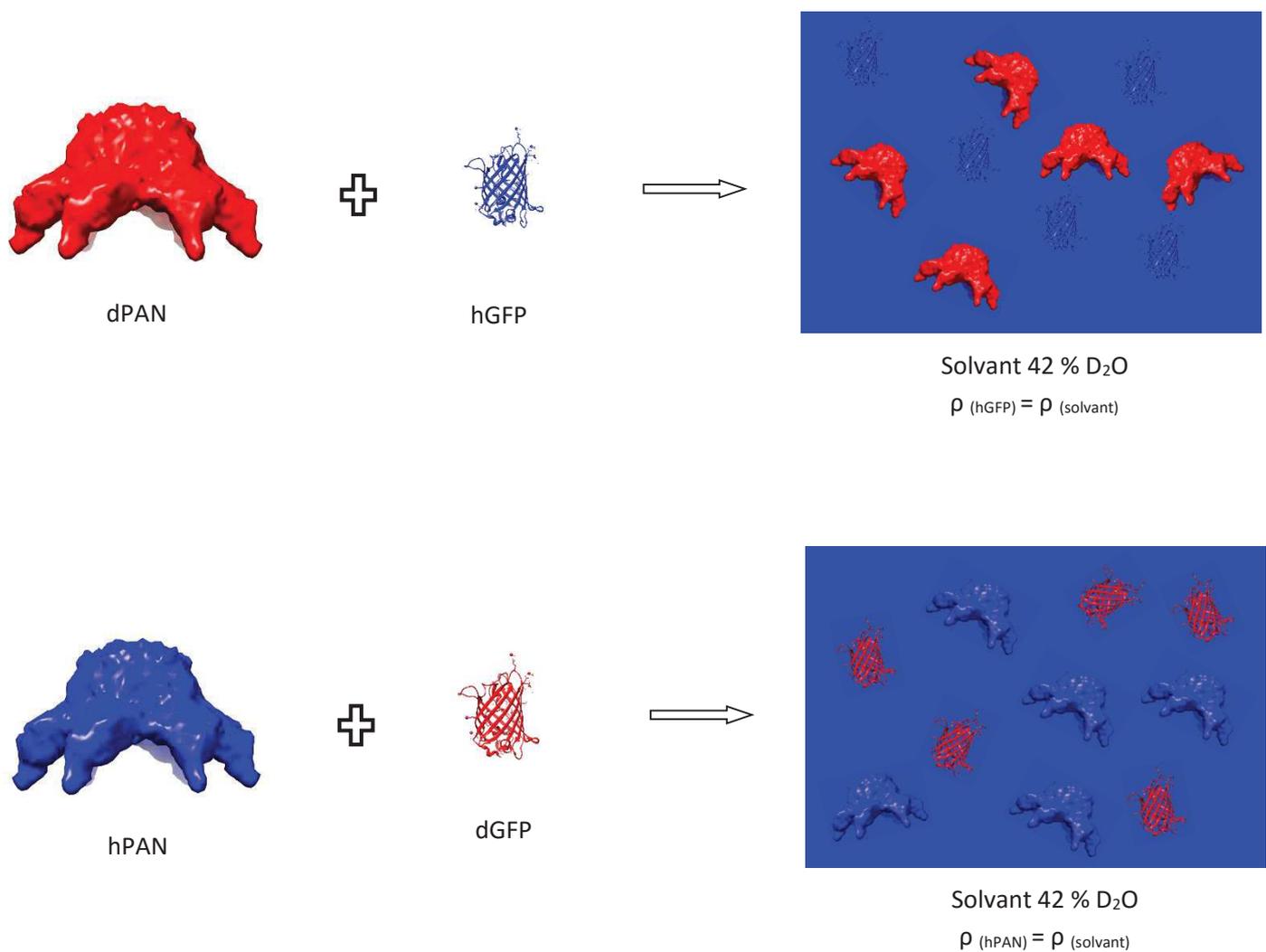


Figure 39 : représentation schématique des deux possibilités de variation de contraste utilisées dans notre cas d'étude.

Le court temps d'exposition de 30 secondes utilisé pour la collecte des données sur les partenaires deutérés lors de la cinétique de dépliement a permis de collecter un grand nombre de profils SANS pendant un temps court de dépliement essentiel pour suivre le processus avec une courte échelle de temps portant beaucoup d'informations structurales et dynamiques sur le complexe PAN et son substrat pendant les toutes premières minutes du début de la réaction. Il faut également souligner que l'utilisation des systèmes d'archées hyperthermophiles comme modèles d'étude dans le cadre de notre étude en temps réel par SANS a fourni un avantage essentiel pour le projet en raison de la thermoactivation du complexe PAN permettant de mieux contrôler la cinétique enzymatique et le processus de dépliement.

Pour conclure, des mesures de fluorescence de la GFP_{ssrA}, couplées à la collecte des courbes SANS (Fig. 37), ont fourni beaucoup d'informations supplémentaires sur le processus

permettant de faire la connexion entre le dépliement de la GFP_{ssrA} du point de vue structural et la perte de sa fluorescence due à la déstabilisation du fluorophore, protégé à l'intérieur du tonneau β de la protéine. La combinaison de cette approche couplée, combinée avec d'autres techniques biochimiques et structurales utilisées durant ma thèse (microscopie électronique, SAXS, cristallographie aux rayons X, SPR, voir chapitres 3 et 4) a constitué un outil puissant pour réaliser une caractérisation structurale, fonctionnelle et dynamique du système PAN qui nous a permis d'éclaircir le mode d'action de cette machine moléculaire et de gagner des informations précieuses sur le mécanisme des AAA⁺ unfoldases et du complexe d'activation du protéasome eucaryote (19S) (voir les deux chapitres résultats 5.1 et 5.2).

3 MATÉRIELS ET MÉTHODES BIOPHYSIQUES

3.1 Résonance Plasmon de Surface (SPR)

Après validation de l'activité de dépliement de la GFP_{ssrA} par *Mj*PAN et le rôle joué par le tag *ssrA* dans ce processus (chapitre 4.3), une étude plus fine de l'interaction entre les deux protéines est devenue nécessaire pour comprendre plus en détail les étapes qui précèdent le dépliement et plus particulièrement l'interaction entre les deux partenaires ainsi que les facteurs qui l'affectent. Cette étude d'interaction est aussi une expérience préparatoire pour le contrôle qualité de notre système (*Mj*PAN:GFP_{ssrA}) (qui sera ensuite étudié par SANS), pour le calcul de la constante d'affinité (K_D) et pour les constantes cinétiques (K_a et K_d) du complexe dans différentes conditions.

Plusieurs systèmes existent et ont été largement utilisés dans des études *in vitro* impliquant des interactions protéine-protéine, y compris SPR (surface plasmon resonance), et ITC (isothermal titration calorimetry). SPR permet de mesurer des interactions intermoléculaires en temps réel dans un système d'écoulement contrôlé. La technique est très sensible et capable de mesurer des interactions avec des constantes d'affinités comprises entre 100 mM et 1 pM, nécessite une faible quantité de protéine (quelques μ g sont suffisants pour mesurer l'affinité entre *Mj*PAN et la GFP_{ssrA}), et peut fournir des cinétiques précises ainsi que des informations sur l'affinité de l'interaction (Katsamba et al, 2002).

3.1.1 Théorie

La résonance plasmon de surface (SPR) est un phénomène physique qui peut être appliqué pour mesurer des interactions biomoléculaires et a été utilisé dans certains cas dans des études impliquant des interactions protéine-protéine et protéine-ARN/ADN (Katsamba et al, 2002)

(Park et al, 2000). Plusieurs entreprises produisent des biocapteurs pour la SPR, Biacore (GE Healthcare Bio-Sciences AB) a été le premier à produire un instrument SPR commercial disponible et est toujours l'instrument le plus largement utilisé. Le principe de base du système comporte un prisme sur un côté de la surface de transduction recouverte d'un mince film d'or (50–100 nm) et un « flow channel » de l'autre côté (non illuminé) de la surface (Fig. 40).

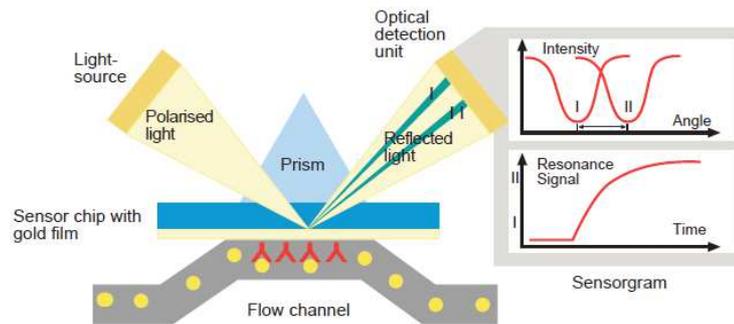


Figure 40 : Le principe de base du système Biacore utilisé dans nos études d'interactions protéine-protéines. Le système comporte un prisme sur un côté de la surface recouverte d'un film mince d'or et d'un canal d'écoulement de l'autre côté de la surface. La lumière pénètre dans le prisme, elle se réfléchit après sur une large gamme d'angles d'incidence et l'intensité de la lumière réfléchi est ensuite mesurée par rapport à l'angle d'incidence par l'unité de détection optique. Cependant, à un certain angle, en particulier, la lumière est absorbée par la surface du biocapteur en tant que plasmons qui se propagent dans le milieu sous forme d'ondes évanescentes. Il existe plusieurs facteurs qui peuvent influencer l'angle sous lequel cela se produit, l'indice de réfraction du matériel dans le canal d'écoulement à proximité de la surface est l'un de ces facteurs. L'adsorption des molécules sur la surface du biocapteur ainsi que les interactions moléculaires proches de la surface conduisent à des modifications dans l'indice de réfraction, et donc un décalage dans l'angle de résonance (I à II). Ce changement angulaire est exprimé en RU (Resonance Units) dans le sensorgramme (Biacore Technology Brochure, GE Healthcare AB).

Une source de lumière polarisée entre dans le prisme, la lumière est ensuite réfléchi sur une large gamme d'angles d'incidences et son intensité est ensuite mesurée par rapport à l'angle incident par l'unité de détection optique. Cependant, à un certain angle, en particulier, la lumière est absorbée par la surface de la puce du biocapteur provoquant une chute de l'intensité de la lumière réfléchi. Une résonance de plasmons de surface est observée à cet angle comme les plasmons se propagent dans le milieu sous forme d'ondes évanescentes. Il existe plusieurs facteurs qui peuvent influencer l'angle de réflexion dont l'indice de réfraction (RI) du matériel dans le « flow channel » à proximité de la surface (à moins de 200 nm) étant l'un d'entre eux. L'adsorption des molécules à la surface du biocapteur ainsi que les interactions moléculaires proches de la surface conduisent à des changements dans le RI et donc un déplacement de l'angle de résonance. Ce changement d'angle est exprimé en unité

de résonance (RU), qui est directement proportionnelle à la variation de masse à proximité de la surface (Karlsson, 2004).

Le setup de base pour la puce du biocapteur de Biacore optimisé pour cette étude implique l'immobilisation des molécules de *MjPAN* his taggées du côté C-terminal sur un capteur d'acide nitriloacétique (NTA) (Sensor Chip NTA, Biacore) (Fig. 41). Comme les molécules de l'analyte, GFPssrA et wild type GFP (wtGFP), injectées dans le système d'écoulement atteignent la surface du biocapteur, la liaison se produit conduisant à une modification de l'indice de réfraction du matériel à proximité de la surface, ce qui entraîne à son tour à la variation de l'angle de réflexion de la lumière incidente. Ce changement est détecté au niveau de l'unité de détection optique et s'affiche sous forme de sensorgrammes (RU vs. temps) dans le logiciel de contrôle, qui peut ensuite être utilisé pour des analyses de cinétique ou d'affinité de l'interaction entre les deux partenaires (Katsamba et al, 2002) (Fig. 42).

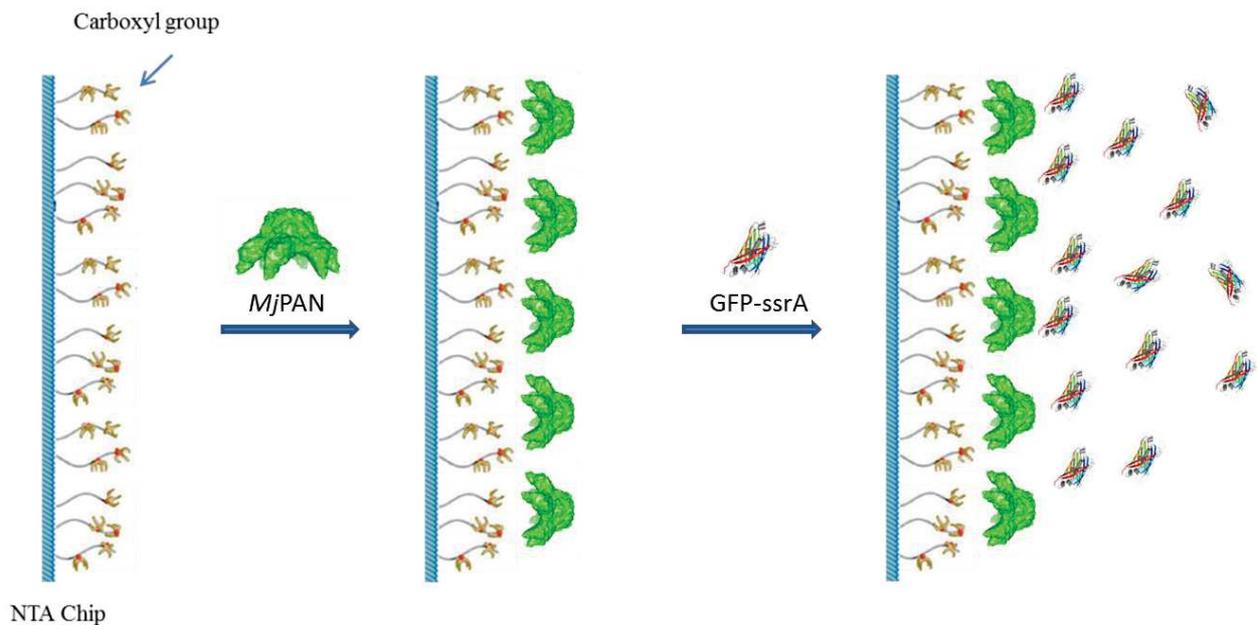


Figure 41 : Interaction protéine-protéine sur la surface du biocapteur. *MjPAN* His taggé du côté C-terminal (en vert) a été immobilisée sur la surface du chip NTA et les protéines analytes (GFPssrA et wtGFP) ont été injectées dans le tampon à un débit constant. Quand les protéines atteignent la surface du chip liant *MjPAN*, la liaison produite entre *MjPAN* et GFP mène à un changement d'indice de réfraction, qui est enregistré et affiché comme sensorgramme.

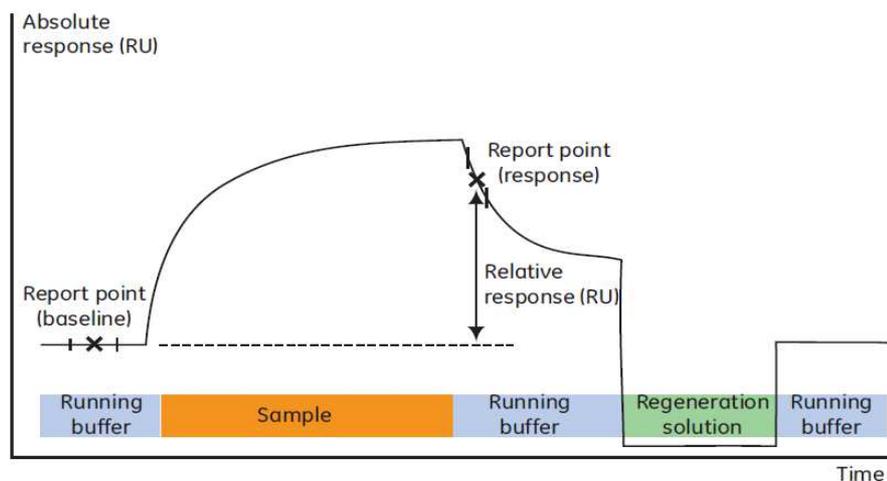


Figure 42 : Le sensorgramme. Une illustration d'un sensorgramme montrant les différentes phases de l'événement de liaison : l'association et la dissociation suivie d'une régénération de la surface pour enlever les analytes restants et apporter de nouveau la réponse au niveau de la référence (Biacore Sensor Surface Handbook, GE Healthcare Bio-Sciences AB).

3.1.2 Démarche expérimentale

La dynamique de l'interaction entre *MjPAN* et GFP_{ssrA} a été caractérisée par SPR en utilisant l'instrument T200 (GE Healthcare Bio-Sciences AB). Le principe de base de la surface du biocapteur utilisé est illustré dans la figure 42. La protéine PAN de *M. jannaschii* taggée avec 6 histidines du côté C-terminal (chaque monomère) est utilisée dans l'étude. Avant immobilisation de PAN, une solution de NiCl₂ à 0,5 mM a été injectée sur la surface du biocapteur avec un débit de 10 µl/min pendant 1 min pour la fixation du Ni²⁺ sur la surface, PAN est ensuite immobilisée par les tags histidines sur le biocapteur NTA (Sensor Chip NTA, Biacore), déjà chargé avec le nickel. La protéine sous sa forme hexamérique est diluée jusqu'à une concentration finale de 100 nM dans le tampon HBS (10mM HEPES, pH 7,5, 100 mM NaCl, 10 mM MgCl₂) supplémenté avec du 50 µM EDTA et 0,025 % Tween 20, et injectée sur la surface à 10 µl/min à température ambiante pour générer une réponse entre 500 et 1000 RU (Resonance Units). Les protéines GFP (*ssrA* et wild type) ont été diluées dans le même tampon à une gamme de concentration de 500, 1000, 2000, 3000, 4000, 5000 et 6000 nM pour GFP_{ssrA} et de 100, 200, 300, 400 et 500 pour wtGFP et injectées à 45 °C avec un débit de 50 µl/min pour 2 ou 3 min. La régénération de la surface pour se débarrasser des protéines fixées est effectuée en injectant pendant 1 min une solution de 350 mM EDTA à 30/50 µl/min pour chélater le métal (Ni²⁺), le ligand (PAN) et l'analyte (GFP), la surface est ensuite rechargée avec du Ni²⁺ pour un deuxième cycle d'immobilisation et d'interaction.

3.1.3 Interprétation des données

Les analyses des concentrations protéiques ont été réalisées en duplicate et le signal de base a été soustrait de chaque ensemble de données à partir d'une surface de référence ne contenant pas la protéine *Mj*PAN. Les données sont analysées en utilisant le modèle d'interaction de Langmuir avec une stœchiométrie 1:1 (Day et al, 2013) dans le cas de la GFP_{ssrA} en utilisant le logiciel d'évaluation du Biacore T200 (GE Healthcare Bio-Sciences AB) pour déterminer la cinétique (les constantes d'association et de dissociation K_a et K_d) ainsi que l'affinité (K_D) de l'interaction *Mj*PAN:GFP_{ssrA}. Dans le cas du wtGFP, le modèle de stœchiométrie 1:1 de Langmuir (Day et al, 2013) n'était pas applicable à cause de l'interaction aléatoire non spécifique entre la wtGFP et *Mj*PAN en absence du tag *ssrA*, donc aucune valeur quantitative n'est calculée dans ce cas pour caractériser l'interaction et les résultats sont présentés qualitativement comme contrôle montrant le rôle du tag *ssrA* dans l'interaction et la spécificité de PAN d'interagir avec les protéines globulaires sans avoir spécifiquement l'activité de dépliement.

3.2 Microscopie électronique

3.2.1 Coloration négative

Afin de contrôler les différents états oligomériques des complexes PAN de *P. horikoshii* et *M. jannaschii* après nos différentes étapes de purifications, 4 μ l des échantillons purifiés (~ 0,1 mg/ml) à la sortie de la colonne Superose 6 10/300 GL (GE Healthcare) de gel filtration (voir détails en chapitre 4) ont été chargés sur une grille contenant un film mince de carbone amorphe (Franzetti et al, 2002). Les échantillons ont été colorés en utilisant 2% de sodium silicotungstate, pH 7,5 et séchés à l'air. Les images étaient prises dans des conditions à faible dose avec un microscope électronique T12 FEI qui travaille à 120 kV et avec un grossissement nominale de 40 000 en utilisant une caméra CCD Orius SC1000.

3.2.2 Cryo-microscopie électronique

Pour les échantillons de la cryo-microscopie électronique, 4 μ l des échantillons purifiés (~ 0,1 mg/ml) à la sortie de la colonne Superose 6 10/300 GL (GE Healthcare) de gel filtration en présence d'ATP γ S ou en absence de nucléotide ont été chargés sur une grille Quantifoil R2/1 (Quantifoil Micro Tools GmbH, Germany) et vitrifiée en utilisant un Mark IV (FEI). La grille congelée a été transférée dans un microscope électronique Polara travaillant à 300 kV (ATP-free *Ph*PAN) ou dans un microscope électronique F20 (FEI) à 200kV (ATP γ S-bound

PhPAN). Les images ont été prises dans des conditions à faible dose (moins de $20 \text{ e}^-/\text{\AA}^2$) et avec un grossissement nominal de 78 000 sur un camera CCD GATAN ultrascan (ATP-free *PhPAN*) ou avec un grossissement nominal de 50 000 sur camera CCD Eagle FEI (ATP γ S-bound *PhPAN*). Les tailles des pixels correspondantes au niveau de l'échantillon sont 1,5 \AA (ATP-free *PhPAN*) et 2,15 \AA (ATP γ S-bound *PhPAN*). Un total de 30 000 (ATP-free *PhPAN*) et 20 000 (ATP γ S-bound *PhPAN*) particules ont été sélectionnées en utilisant une procédure de sélection de particules semi-automatique avec le EMAN boxer routine (Ludtke et al, 1999) et extraites dans 128 x 128 et 90 x 90 cases de pixels respectivement. Les modèles ont été corrigés avec CTFFIND3 (Mindell & Grigorieff, 2003) et Bsoft (Heymann et al, 2008) et filtrés jusqu'à 25 et 8 \AA . Une analyse de projection par SPIDER (Frank et al, 1996) a été effectuée. Le modèle initial a été généré par rétro-projection d'une vue de côté de la particule en appliquant une symétrie C6. Un nombre total de 372 re-projections équidistantes et 50 cycles ont été utilisés pour obtenir la reconstruction 3D finale (les paramètres d'orientation des images brutes et reconstruction 3D sont stables). Chaque cycle est composé des re-projections du modèle, un alignement en utilisant le fichier de 25 \AA filtré et des reconstructions 3D en utilisant le fichier de 8 \AA filtré. Seul les meilleurs 15 000 images ou 10 000 images (« crosscorrelation criteria ») ont été utilisées pour la reconstruction. La résolution des structures est entre 13 \AA et 15 \AA (0,5 « Fourier-shell correlation criteria ») pour l'ATP-free *PhPAN* et l'ATP γ S-bound *PhPAN* respectivement.

3.2.3 Analyse des données microscopie électronique

Les modèles générés par cryo-microscopie électronique ont tout d'abord été validés par SAXS en solution (voir paragraphe 3.3) après génération des fichiers pdb de chaque modèle par le logiciel EM2DAM (Petoukhov et al, 2012). Les dimensions des deux modèles (ATP-free et ATP γ S-bound) ont été ensuite mesurées par Pymol (Delano, 2002), comparées et présentées pour mieux comprendre les changements de conformation liés à la liaison du nucléotide. La carte de densité électronique de *PhPAN* en absence de nucléotide a été utilisée pour générer le modèle pseudo-atomique du complexe en fittant les structures cristallographiques hautes résolutions des domaines N-terminal (hexamérique) et C-terminal (monomérique) dans la carte par le logiciel Chimera (Pettersen et al, 2004). Un modèle pseudo-atomique hypothétique de la forme complète du *PhPAN* (« full length *PhPAN* ») a été généré en ajoutant les résidus manquants des « coiled coil » du côté N-terminal prédits en utilisant le programme PHYRE2 (Kelley et al, 2015). Ce modèle (full length *PhPAN*) a été

également fitté dans la carte de densité Cryo-EM de la base (Rpt 1-6) du complexe 19S activateur du protéasome eucaryote segmenté de la structure Cryo-EM du protéasome 26S de *Sacharomyces cerevisiae* (9 Å, EMDB-1992) (Fig. 5, chapitre 5.1).

3.3 Diffusion de rayons X aux petits angles (SAXS)

3.3.1 Démarche expérimentale

Toutes les expériences SAXS ont été effectuées sur la ligne de lumière BM29 (BioSAXS beamline) à l'ESRF (European Synchrotron Radiation Facility) de Grenoble, France, en utilisant un détecteur 2D Pilatus (Pernot et al, 2013) (Fig. 43). La distance entre l'échantillon et le détecteur a été réglée à 2,867 m. L'énergie des rayons X utilisés est de 12,5 keV qui correspond à une longueur d'onde de 0,9919 Å. Les échantillons ont été mesurés en mode statique ou en mode FPLC (« Fast Protein Liquid Chromatography ») (Pernot et al, 2013). Dans le cas de FPLC-SAXS, une colonne Superose 6 10/300 GL (GE Healthcare) d'exclusion de taille a été montée sur la ligne. L'échantillon de *Mj*PAN a été injecté sur la colonne sous sa forme dodécamérique ou hexamérique à une concentration de 5 mg/ml et a été élué avant l'exposition aux rayons X. Les images ont été ensuite collectées sur tous les pics à la sortie de la colonne, l'échantillon a été élué sur la colonne avec un débit de 0,5 ml/min et le temps d'exposition était de 4 s par image. En mode statique les échantillons ont été mesurés à température ambiante à une concentration entre 0,6 et 1 mg/ml pour *Ph*PAN et 3 à 4 mg/ml pour *Mj*PAN. Le temps d'exposition était de 10 images avec 1 s par image pour l'échantillon et le tampon. Les mesures ont été effectuées en mode d'écoulement ou les échantillons ont été poussés à travers un capillaire à un débit constant pour éviter les dommages d'irradiation.

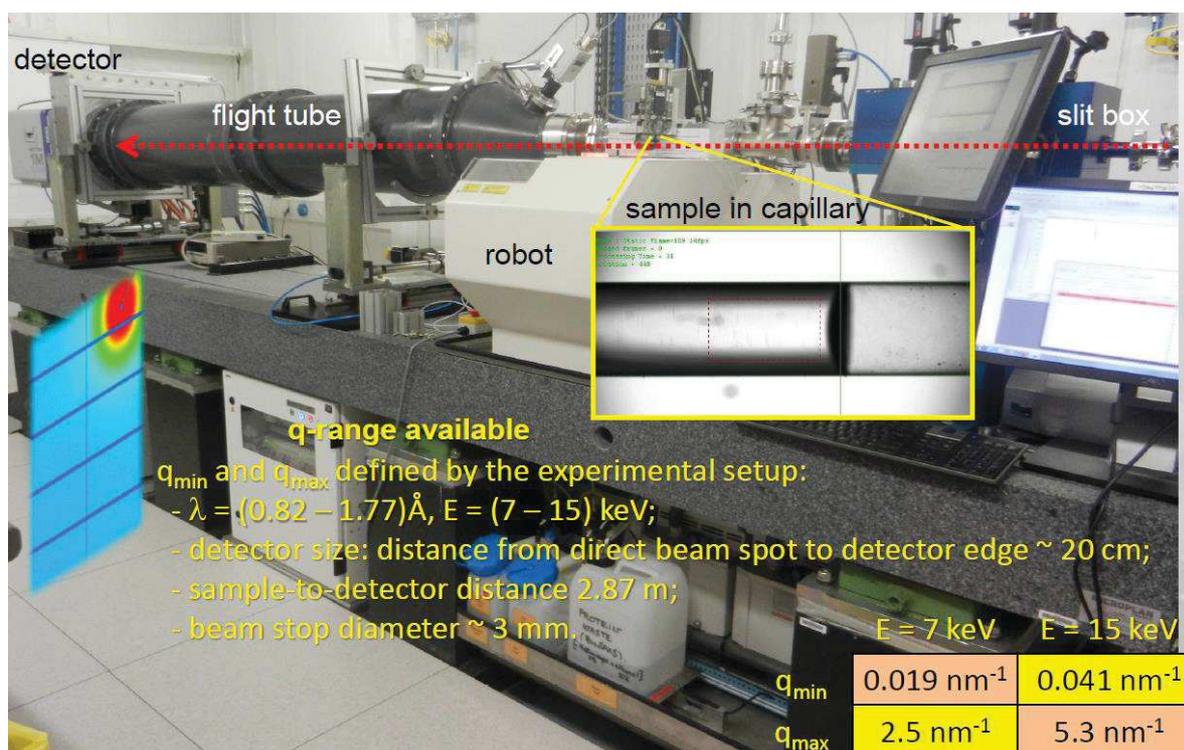


Figure 43 : Image du « setup » et caractéristiques de la ligne BM29 à l'ESRF, Grenoble. (Présentation Petra Pernot, HSC16, 15 Septembre 2014).

3.3.2 Interprétation des données

Pour la collecte et la réduction des données SAXS, un logiciel spécifique de la ligne BM29 de l'ESRF (BsxCuBE) (Pernot et al, 2013) a été utilisé d'une façon automatisée. Les intensités de diffusion 1D des échantillons et des tampons ont été exprimées en fonction du module du vecteur de diffusion q . Les mêmes logiciels ont été utilisés pour le traitement des données SAXS et SANS (voir paragraphe 2.3.3). Les courbes de diffusion SAXS expérimentales ont été utilisées pour valider les modèles cryo-microscopie électronique de *PhPAN* ainsi que l'état oligomérique de son domaine N-terminal et pour générer des modèles *ab initio* de ce complexe. D'autres données SAXS mesurées également sur les dodecamères et hexamères de *PhPAN* (en mode statique) et *MjPAN* (en mode FPLC-SAXS) ont permis la détermination des paramètres biophysiques et structurales (R_g , $I(0)$, MW) de ces complexes.

4 MATÉRIELS ET MÉTHODES BIOCHIMIQUES

4.1 Expression et purification des protéines hydrogénées

Une protéine peut être obtenue soit de l'organisme originel soit par l'expression de son gène dans un autre organisme, habituellement la bactérie *Escherichia coli*. Pour les protéines *PhPAN* et *MjPAN*, la seconde méthode est la plus avantageuse en termes de temps et de rendement. Des progrès au niveau de la biologie moléculaire et de la technologie des ADN recombinants ont permis l'expression des gènes d'un organisme dans un organisme différent (Schleif, 1993). Pour ce faire, l'ADN codant pour la protéine d'intérêt est inséré dans un vecteur, qui à son tour est introduit dans l'organisme hôte qui va produire la protéine. Par conséquent, il est possible d'insérer des séquences d'ADN dans le vecteur qui seront ensuite surexprimées en protéines en utilisant la machinerie transcriptionnelle et traductionnelle de l'organisme hôte accueillant l'ADN recombinant. L'organisme hôte est ensuite mis en culture dans le milieu approprié et exprime la protéine recombinante qui sera séparée des protéines de l'organisme par différentes étapes de purifications. La surexpression des protéines recombinantes facilite beaucoup la production de solution protéique pure nécessaire pour toutes les études fonctionnelles, biophysiques et structurales.

4.1.1 Synthèse des gènes, clonage et transformation

Les gènes codants pour les protéines *MjPAN* et GFPssrA ont été générés chez *GeneCust Europe* (<http://www.genecust.com>) à partir des séquences d'ADN synthétiques optimisées pour les codons rares dans *E. coli* et clonés dans le plasmide de surexpression pET30a. Le gène de *MjPAN* est inséré entre les deux sites de restriction NdeI et XhoI permettant l'expression d'une protéine avec un tag poly-histidine à l'extrémité C-terminale, le gène codant pour la GFPssrA est inséré entre les deux sites de restriction NdeI et NotI pour l'expression d'une protéine sans tag. Le gène codant pour *PhPAN* est cloné dans le plasmide pET41c entre les deux sites de restriction NdeI et BamHI à partir de l'ADN génomique en utilisant un protocole de PCR standard. Les deux vecteurs pET41c et pET30a sont sous le contrôle du promoteur T7 pour l'expression des protéines d'intérêt et portent un gène de résistance à la kanamycine. La souche bactérienne *E. coli* BL21 (DE3) est ensuite transformée par les vecteurs pET41c et pET30a contenant les gènes d'intérêt par choc thermique à 42 °C etensemencée sur une boîte de Pétri (LB-agar + Kana). La boîte de Pétri est conservée sur la nuit dans une étuve à 37 °C.

4.1.2 Expression

La culture de bactéries est réalisée dans du milieu LB-kana (Lysogeny broth/Kanamycine) (contenant de la kanamycine, 50 µg/ml) qui permet de sélectionner les cellules surexprimant les protéines. Une colonie de bactéries prélevée sur une boîte de Pétri permet d'ensemencer 30 à 50 ml de milieu LB-kana, qui, après une nuit d'incubation à 37 °C sous agitation (220 rpm), constitue une pré-culture. Celle-ci est alors diluée 100 fois dans du milieu LB-kana (10 ml dans 1 l dans un Erlenmeyer stérile de 2 l). Les bactéries sont cultivées à 37 °C sous agitation (200 rpm) et leur multiplication est suivie par mesure de l'absorbance de la solution à 600 nm. Lorsque la densité optique atteint 0,6 (phase exponentielle précédant la phase stationnaire) après environ 2 à 3 h d'incubation, la surexpression des protéines est induite par ajout d'IPTG pour obtenir une concentration finale de 1 mM. La culture induite est alors incubée sous agitation pendant 4 h, puis centrifugée pendant 15 min à 4500 rpm à 4 °C dans un rotor Beckman JA10. Les culots bactériens sont utilisés immédiatement pour la purification ou congelés à -80 °C.

4.1.3 Purification et hexamérisation

Les bactéries resuspendues dans du tampon de lyse (20 mM Tris/HCl pH 7,5, 150 mM NaCl) sont lysées par cinq cycles de sonication (40 s ON / 40 s OFF) à l'aide d'une sonde à 100% d'activité (Branson Sonifer® 150) après ajout (dans le cas des protéines PAN) du lysozyme, DNaseI, MgSO₄, Pefabloc, RNase, une tablette de 'complete' (Protease Inhibitor Cocktail Tablets) et du triton X-100. Vu que les protéines PAN sont isolées des organismes hyperthermophiles *Pyrococcus horikoshii* et *Methanococcus jannaschii*, elles sont plus thermostables que les protéines d'*E. coli* et fonctionnent de manière optimale à haute température. Pour cela un choc thermique de 15 minutes à 85 °C est effectué sur les cellules après la lyse comme une étape de purification supplémentaire. Ce traitement permet d'éliminer le maximum de protéines d'*E. coli* non thermostables à cette température. La solution est ensuite centrifugée à 12 000 rpm pendant 1 heure (PAN) ou 20 min à 14 000 rpm (GFPssrA) dans un rotor Beckman JA20.

Purification de PhPAN :

Le surnageant est déposé sur une colonne DEAE échangeuse d'anion faible (volume VC = 20 ml), équilibrée avec du tampon 20 mM Tris/HCl pH 8, 10 mM MgCl₂, 150 mM NaCl et 1 mM DTT à un débit de 2 ml/min. Elle est lavée avec le même tampon après fixation de la

protéine puis les protéines sont éluées avec un gradient de chlorure de sodium (NaCl) allant de 150 mM jusqu'à 500 mM. La protéine sort à une concentration de NaCl de 300 mM, les fractions contenant la protéine sont rassemblées. La solution protéique est ensuite diluée avec le même tampon jusqu'à une concentration finale de 150 mM de NaCl et déposée de nouveau sur une colonne MonoQ 5/50 GL échangeuse d'anion forte (GE Healthcare) (VC = 1 ml) à un débit de 1 ml/min équilibrée et lavée dans le même tampon contenant 150 mM NaCl. La protéine est ensuite éluée avec un gradient de 150 mM jusqu'à 500 mM NaCl. Les fractions contenant la protéine sont congelées à -20 °C ou déposées immédiatement sur une colonne de gel filtration. Après concentration de la solution de *Ph*PAN par un Centricon® de cut-off 50 kDa, la protéine est alors déposée par injection de 500 µl sur une colonne de gel filtration Superose 6 10/300 GL (GE Healthcare) équilibrée avec le même tampon (20 mM Tris/HCl pH 8, 10 mM MgCl₂, 150 mM NaCl et 1 mM DTT) au débit de 0,5 ml/min. Les fractions contenant la *Ph*PAN sont réunies, concentrées et stockées à -20 °C pour des utilisations ultérieures.

Purification de *Mj*PAN :

Après la lyse des cellules et la centrifugation, le surnageant est déposé sur une colonne de Nickel (HiTrap Chelating HP, GE Healthcare) (VC = 5 ml) équilibrée avec du tampon contenant 20 mM Tris/HCl pH 7,5, 10 mM MgCl₂, 150 mM NaCl et 5 mM imidazole. La colonne est lavée avec le même tampon contenant 50 mM puis 100 mM imidazole (10 volumes colonne chaque lavage). La protéine est éluée dans 20 ml de tampon d'éluion contenant 500 mM imidazole. Après une dialyse pendant la nuit contre un tampon ne contenant pas d'imidazole, la protéine est ensuite purifiée sur une colonne MonoQ 5/50 GL échangeuse d'anion puis une colonne Superose 6 10/300 GL d'exclusion de taille comme décrit pour *Ph*PAN. Les fractions contenant la *Mj*PAN sont réunies, concentrées et stockées à -20 °C pour des utilisations ultérieures.

Comme déjà mentionné dans le chapitre « Introduction », le complexe d'activation du protéasome PAN est actif sous sa forme hexamérique. Dans nos conditions de purification les deux complexes PAN de *P. horikoschii* et *M. janaschii* sont purifiés sous forme d'un équilibre dodécamère/hexamère. Pour obtenir uniquement la forme hexamère des complexes PAN et une solution homogène et monodisperse de la protéine nécessaire pour nos études biochimiques et structurales, en particulier la microscopie électronique et le SAXS/SANS, la solution protéique est chauffée après la dernière colonne de gel filtration pendant 30 min à 60 °C en présence de 4 mM ATP et 10 mM MgCl₂ déjà présent dans le tampon d'éluion. La

protéine est chargée de nouveau sur colonne Superose 6 10/300 GL sans ATP dans le tampon et l'état d'oligomérisation final est vérifié via le profil d'élution (chromatogramme) et par microscopie électronique en coloration négative.

Purification de la GFPssrA :

100 mM du triéthanolamine base (TEA base) avec 40% (w/v) de sulfate d'ammonium $(\text{NH}_4)_2\text{SO}_4$ sont ajoutés au surnageant obtenu après la lyse et la centrifugation des cellules (comme expliqué ci-dessus). Après incubation d'une heure à 4 °C, les protéines précipitées sont éliminées par centrifugation à 4 °C pendant 20 min à 6 000 rpm dans un rotor JA20 (Beckman J2-MC Centrifuge). 20% (w/v) du $(\text{NH}_4)_2\text{SO}_4$ est ensuite ajouté au surnageant. Une première extraction à l'éthanol est réalisée. Pour cela 1/4 de volume d'éthanol (Fluka, $\geq 99,8\%$) est ajouté. La solution est secouée vigoureusement pendant une minute. Les deux phases sont séparées par centrifugation 10 min à 4 500 rpm (Beckman Coulter Alegarx®-30R Centrifuge). La phase organique (du dessus) contenant la GFPssrA est récupérée. 1/16 de volume d'éthanol est ensuite ajouté. La solution est de nouveau secouée vigoureusement pendant une minute et centrifugée de nouveau 10 min à 4 500 rpm (Beckman Coulter Alegarx®-30R-Centrifuge). La phase organique est récupérée, avec laquelle on effectue une seconde extraction au butan-1-ol. Pour cela 1/4 de volume de butan-1-ol est ajouté. La solution est secouée vigoureusement 30 s et centrifugée 10 min à 4 500 rpm (Beckman Coulter Alegarx®-30R-Centrifuge). A cette étape, la GFPssrA est dans la phase aqueuse (du dessous), qui est récupérée. Un volume équivalent de chloroforme est ajouté. La solution est secouée vigoureusement 30 s et centrifugée 10 min à 4 500 rpm (Beckman Coulter Alegarx®-30R-Centrifuge). La phase aqueuse contenant la GFPssrA est récupérée, et 20% (w/v) du $(\text{NH}_4)_2\text{SO}_4$ y est ajouté. La solution protéique contenant la GFPssrA soluble est déposée sur une colonne Phenyl Sepharose 6 Fast Flow (GE Healthcare) (VC ~ 18 ml) équilibrée avec un tampon 20 mM Tris/HCl pH 7,5, 20% (w/v) $(\text{NH}_4)_2\text{SO}_4$ à un débit de 1 ml/min. Elle est lavée avec le même tampon (3 volumes colonne) puis les protéines sont éluées avec du 20 mM Tris/HCl pH 7,5 sans $(\text{NH}_4)_2\text{SO}_4$. Les fractions contenant les protéines sont rassemblées et congelées à -20 °C.

La colonne Phenyl Sepharose (GE Healthcare) à forte concentration en sulfate d'ammonium, fixe les protéines par précipitation non dénaturante sur le support (chromatographie d'adsorption). Les protéines se fixent au gel par des liaisons hydrogènes, des liaisons de van der Waals ou hydrophobes. La baisse de la concentration en sulfate d'ammonium permet de resolubiliser les protéines qui sont ainsi éluées de la colonne.

4.2 Expression et purification 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_2O) supplémenté 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 et caractérisée.

La deutération permet l'utilisation des techniques biophysiques basées sur les neutrons comme la diffusion de neutrons aux petits angles (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 paragraphe 2.2). Les échantillons dPAN (*Mj*PAN) et dGFPssrA (d = deutéré) ont été produits avec l'aide de Martine Moulin et Michael Härtle in au sein du laboratoire de deutération de l'ILL, le D-LAB.

4.2.1 Adaptation des bactéries et culture par fermentation

Les plasmides et la souche bactérienne utilisés pour la production des protéines deutérées sont les mêmes que ceux utilisés pour l'expression des protéines hydrogénées (voir paragraphe 4.1.1). Les milieux composés de D_2O et d'une source carbonée (d_8 -glycérol) étant très coûteux, une stratégie de culture à haute densité « high cell density culture » est utilisée. Cela permet d'obtenir un haut rendement de biomasse par volume et par conséquent un haut rendement de protéines recombinantes deutérées. Le milieu choisi pour la fermentation est le milieu minimum Enfors (Xu et al, 1999) (Table 5 et 6) que l'on peut produire sous sa forme hydrogénée ou deutérée. La souche utilisée BL21(DE3) ne présentant pas d'auxotrophie peut pousser dans ce type de milieu. Pour garantir la stabilité du plasmide recombinant, par conséquent l'expression de la protéine, le marqueur de résistance à la kanamycine est préféré à l'ampicilline. En effet, la β -lactamase responsable pour la résistance à l'ampicilline a tendance à traverser la membrane extérieure vers le milieu, conduisant à une perte de sélection pour les bactéries recombinantes.

| Milieu Enfors | Quantité (Pour 1 litre) |
|---|----------------------------|
| (NH ₄) ₂ SO ₄ | 6,86 g |
| KH ₂ SO ₄ | 1,56 g |
| Na ₂ HPO ₄ anhydre | 6,48 g |
| (NH ₄) ₂ -H-Citrate | 0,49 g |
| MgSO ₄ 1M (D ₂ O) | 1 ml |
| Solution minérale 1000X (D ₂ O) | 1 ml |
| Glycérol 100% (deutééré) | 5 g |
| Kanamycine 10 mg/ml (D ₂ O) | 3 ml |
| D ₂ O | qsp 1 litre |

Table 5 : Composition du milieu minimum Enfors pour la croissance des bactéries. Le glycérol fait office de source unique de carbone. L'utilisation conjointe de D₂O et de glycérol deutérée permet d'avoir un taux de deutération proche de 100%. qsp : quantité suffisante pour.

| Solution minérale | Concentration (g/litre) |
|--------------------------------------|----------------------------|
| CaCl ₂ .2H ₂ O | 0,5 |
| FeCl ₃ .6H ₂ O | 16,7 |
| ZnSO ₄ .7H ₂ O | 0,18 |
| CuSO ₄ .5H ₂ O | 0,16 |
| MnSO ₄ .4H ₂ O | 0,15 |
| CoCl ₂ .6H ₂ O | 0,18 |
| EDTA.Na ₂ | 20,1 |

Table 6 : Composition de la solution minérale 1000X. La préparation de la solution sous sa forme deutérée sera précédée d'un échange de tous les hydrogènes labiles contenus dans les poudres commerciales. La solution du MgSO₄ et la solution minérale sous leurs formes deutérées nécessitent un échange de tous les hydrogènes labiles. Les poudres sont solubilisées dans du D₂O et évaporées plusieurs fois avec un 'évaporateur rotatif'.

Afin d'optimiser l'adaptation et la croissance des bactéries en milieu Enfors minimum deutéré, les étapes suivantes ont été effectuées (Fig. 44) :

- ✓ Croissance des bactéries sur une boîte de LB-agar.
- ✓ 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.

Comme le milieu d'expression doit être perdeutééré (milieu 100% deutéré), les passages sont effectués en premier lieu en milieu hydrogéné, pour ensuite être faits en milieu deutéré.

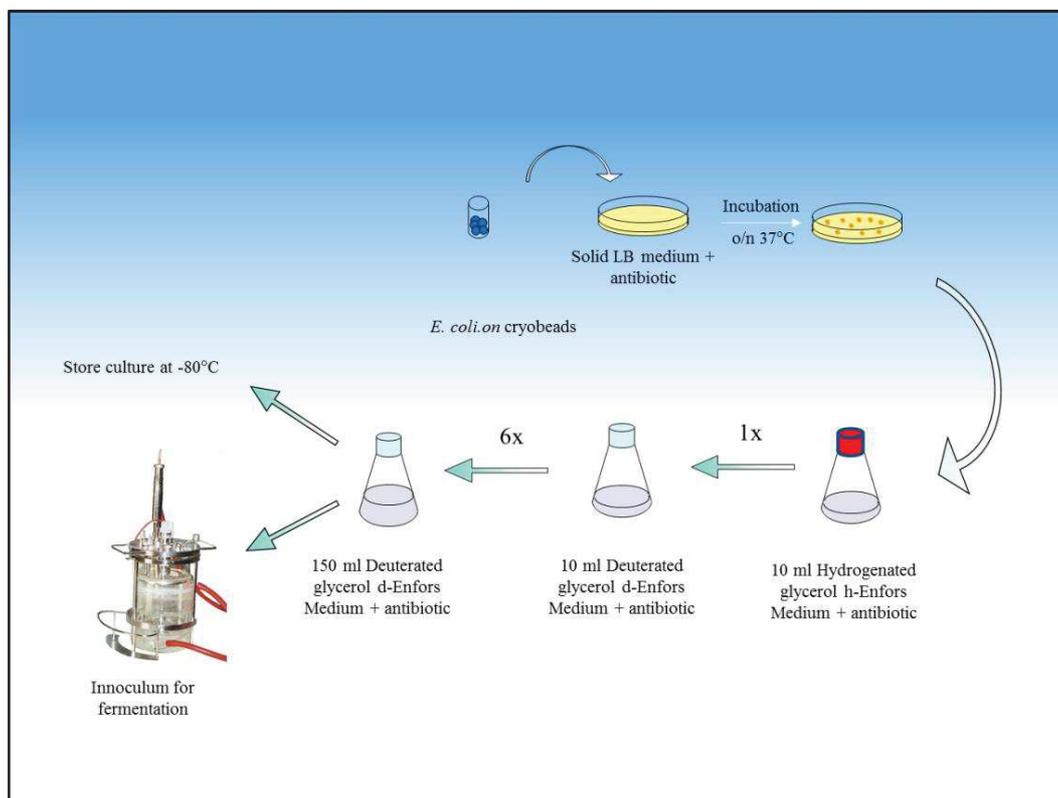


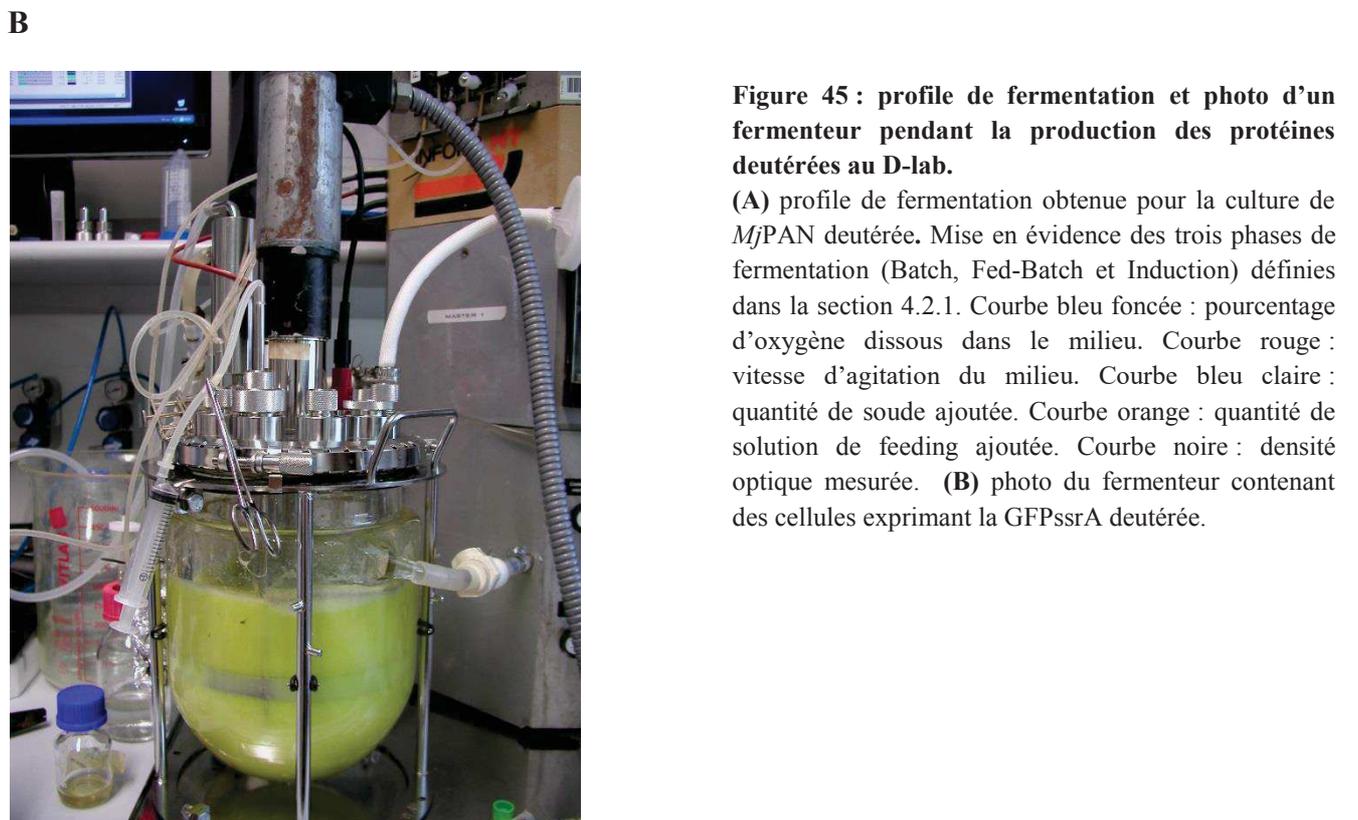
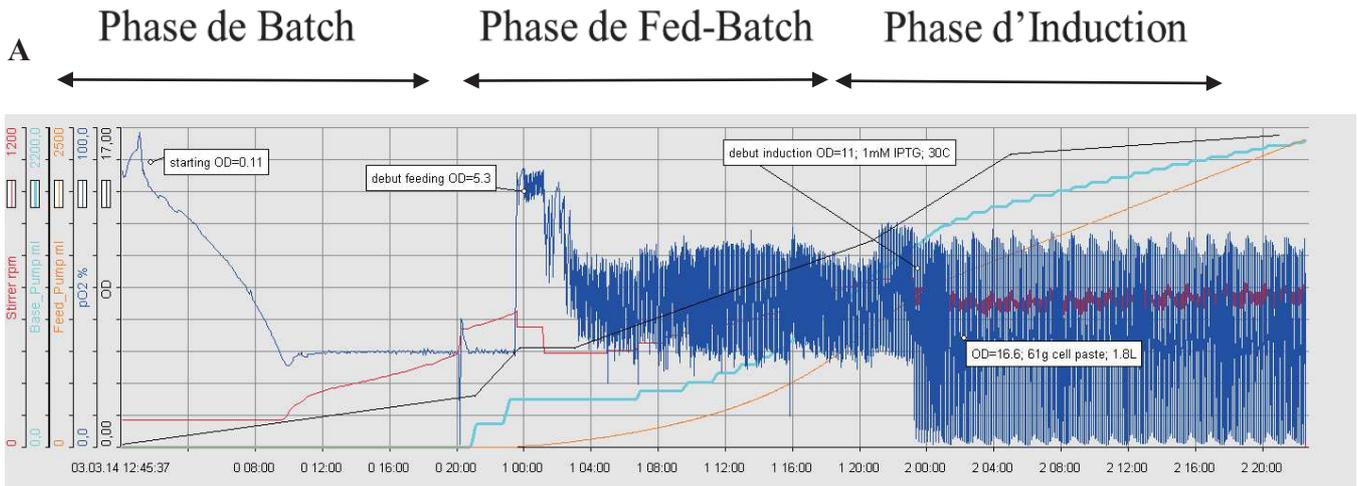
Figure 44 : adaptation des cellules de *E. coli* au deutérium (Michael Haertlein, D-lab ILL).

La culture est faite 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, la pO₂ du milieu (oxygène dissous dans le milieu), avec pour valeur de consigne 30%, et la température (30 °C) sont contrôlés et régulés. La fermentation se déroule en trois phases (Fig. 45):

La première est qualifiée de « Batch »: le but de cette première phase est d'initier la croissance des bactéries. La source de carbone est le d8-glycérol présent dans le milieu de culture en concentration limitante de 5g/l.

Une fois que tout le glycérol consommé, on observe à une acidification progressive jusqu'au point de consigne ($pD=6,9$) du milieu et une augmentation brutale de la pO_2 . Cette augmentation indique l'arrêt de la consommation d'oxygène de la bactérie, par manque de source de carbone. Il devient alors nécessaire de supplémenter à intervalles régulières le milieu en glycérol tout au long de la deuxième phase, dite de « Fed-Batch ». Cette seconde phase consiste à soutenir la croissance des bactéries et à augmenter d'une façon contrôlée la biomasse tout en évitant la production d'acides (acide acétique par exemple). La supplémentation du milieu en source de carbone se fait avec une solution d-Enfors (D_2O) avec 120 g/l de glycérol deutéré, appelée 'solution de feeding'. L'intervalle entre chaque ajout de milieu et la quantité délivrée doivent être définis de manière à ne pas restreindre la croissance bactérienne. En parallèle de l'ajout de glycérol, le pD est maintenu à 6,9 par ajout de NaOD via une pompe contrôlée automatiquement.

La phase d'« Induction »: une fois la densité optique (DO) ayant atteint 10-12, une induction à 1 mM IPTG (D_2O) est effectuée. Cette induction se poursuit pendant une nuit. Cette durée est définie préalablement grâce aux tests d'expression réalisés à petite échelle en milieu deutéré. Les bactéries sont récoltées vers des densités optiques proches de 16. En moyenne, 50 à 70 g de pâte bactérienne est récoltée pour 2 l de milieu dans le fermenteur. Par comparaison, un litre de milieu LB à 37 °C produit 3-4 g de pâte bactérienne.

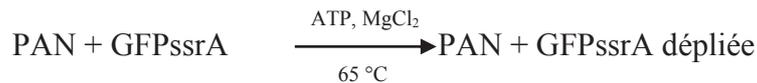


4.2.2 Purification et hexamérisation

Les protéines deutérées *Mj*PAN et GFPssrA ont été purifiées (et hexamérisées, pour PAN) selon le même protocole et dans les mêmes tampons que les protéines hydrogénées (voir paragraphe 4.1.3). Tous les tampons utilisés pendant la purification des protéines deutérées sont préparés avec du H₂O.

4.3 Activité enzymatique

La protéine PAN est en effet une enzyme qui catalyse le dépliement des protéines substrat en présence d'ATP comme source d'énergie et Mg^{2+} comme cofacteur (Benaroudj & Goldberg, 2000). Durant ma thèse le substrat utilisé était la GFPssrA, une fusion entre la protéine native (« wild type », wt) GFP « Green Fluorescent Protein », et un tag de 11 résidus (tag « ssrA » : AANDENYALAA) du côté C-terminal permettant l'adressage et la reconnaissance de la protéine par PAN (Weber-Ban et al, 1999):



4.3.1 Dépliement de la GFPssrA par *Ph*PAN et *Mj*PAN

La mesure de l'activité de PAN chez *P. horikoshii* et *M. jannaschii* consiste à suivre la diminution de l'émission de fluorescence par la GFPssrA à 509 nm après incubation avec PAN et excitation à 400 nm traduisant le dépliement de la GFPssrA. Le tampon de réaction utilisé contenait 20 mM Tris/HCl pH 7,5, 100 mM NaCl, 1mM DTT, 10 mM $MgCl_2$ et 4 mM ATP. La courbe de dépliement de la GFPssrA a été mesurée dynamiquement tout au long du processus avec le spectrofluorimètre « Synergy H4 Hybrid Multi-Mode Microplate Reader » (BioTek) à une λ_{ex} 400 nm et λ_{em} 509 nm et une température de 65 °C. La GFPssrA a été diluée avec le tampon de la réaction jusqu'à une concentration finale de 1 μ M, en absence ou en présence de PAN à différents ratios. Pour comparer l'habilité de dépliement par PAN avec de l'ATP ou de l'ATP γ S, la diminution de la fluorescence totale de la GFPssrA a été mesurée à 65 °C pendant 60 minutes. Le même test a été effectué en présence de la GFP non taggée, en absence d'ATP ou $MgCl_2$ ou en présence d'ATP γ S.

4.3.2 Optimisation des conditions de dépliement

L'optimisation des conditions de la réaction de dépliement de la GFPssrA par *Mj*PAN pour s'adapter aux conditions expérimentales SANS était une étape nécessaire pour suivre la cinétique des deux molécules en temps réel pendant le processus de dépliement. Pour ce but, différentes conditions ont été testées pour les protéines hydrogénées et deutérées avec les deux combinaisons nécessaires pour les mesures SANS (dGFPssrA:hPAN et hGFPssrA:dPAN) à 42 % D_2O dans le tampon. Le but de ces tests était d'optimiser le temps de la réaction de dépliement à l'échelle de temps d'exposition de chaque molécule au faisceau de neutron en gardant une concentration suffisante pour obtenir un signal analysable. Une

gamme de réactions a été testée par fluorescence « offline » à différentes températures de réaction (40, 45, 50, 55, 60 et 65 °C), différentes concentrations d'ATP et différentes concentrations et stœchiométries de protéines (*Mj*PAN et GFPssrA) hydrogénées et deutérées. Les conditions choisies pour être utilisées pendant les expériences de dépliement en temps réel avec le SANS étaient :

Pour (dGFPssrA : hPAN): la réaction a été lancée à 55 °C avec un ratio de ~ 2:1, à 2 mg/ml dGFPssrA (72 µM) et 10 mg/ml dPAN (33 µM), 100 mM ATP, 100 mM MgCl₂ et 42% D₂O dans le tampon réactionnel.

Pour (hGFPssrA : dPAN): la réaction a été lancée à 60 °C avec un ratio de 17:1, à 4,8 mg/ml hGFPssrA (173 µM) et 3 mg/ml dPAN (10 µM), 100 mM ATP, 100 mM MgCl₂ et 42% D₂O dans le tampon réactionnel.

Ces deux conditions ont été choisies pour être utilisées pendant les expériences SANS en raison de leur corrélation à la fois avec les conditions de la réaction enzymatique et les limitations expérimentales des mesures SANS. Plusieurs autres contrôles ont été testés et utilisés, plus de détails ont été présentés dans le paragraphe 2.3.2.

5 RESULTATS

Dans ce chapitre je présente les résultats obtenus au cours de ma thèse sur les protéines PAN étudiées des deux archées hyperthermophiles *P. horikoshii* et *M. jannaschii* avec les différentes approches utilisées en partant de la biochimie des systèmes jusqu'aux études structurales et dynamiques par diffusion de neutrons (SANS) et de rayons X (SAXS) aux petits angles, microscopie électronique et cristallographie aux rayons X (réalisée par Matteo Colombo). Les résultats seront présentés en deux chapitres: un premier chapitre est dédié à la caractérisation biochimique et structurale du système PAN de *P. horikoshii* par microscopie électronique combinée avec des études SAXS/SANS et de la cristallographie aux rayons X. Dans le deuxième chapitre je présente l'étude dynamique en temps réel réalisée sur le système PAN de *M. jannaschii* par SANS en combinaison avec la deutération (réalisée au D-LAB) et la variation de contraste ainsi que les nouveaux aspects méthodologiques qui ont accompagné le projet sur la ligne D22 de l'Institut Laue Langevin (ILL).

5.1 Etude structurale et fonctionnelle du complexe de régulation du protéasome de l'archée *Pyrococcus horikoshii*

5.1.1 Présentation du travail

Vu la simplicité structurale relative du complexe PAN d'archées, son organisation en sous-unités homologues et sa facilité d'expression dans *E. coli*, les propriétés biochimiques et structurales du complexe ont été intensivement étudiées. Il a servi comme modèle de son homologue eucaryote plus complexe et a donné un système expérimental puissant pour investiguer le rôle de l'ATP et des mécanismes biochimiques impliqués dans le processus de translocation du substrat dans la particule catalytique 20S (Smith et al, 2006). Ces études sont très précieuses pour éclaircir la structure et les mécanisme enzymatiques du protéasome 26S (Voges et al, 1999) suite à la conservation des propriétés clés du complexe PAN par rapport à son homologue eucaryote 19S, plus complexe.

Cependant et malgré plusieurs efforts, la structure atomique de PAN et du 19S ne sont toujours pas déterminées. Il est maintenant clair que PAN et la famille des AAA+ ATPases bactériens (Lon, HslU, ClpA et ClpX) utilisent l'ATP en plusieurs étapes au cours de la dégradation des protéines. Une hypothèse pour les AAA+ ATPases est que des changements de conformations induits par l'hydrolyse de l'ATP seraient à l'origine du dépliement des protéines, l'ouverture de la porte d'entrée du protéasome 20S et la translocation du substrat

(Sauer & Baker, 2011). Plusieurs preuves expérimentales indiquent que PAN subit des changements de conformations après liaison et/ou hydrolyse d'ATP. Un exemple est que l'association de l'ATP γ S ou de l'AMPPNP stabilise l'association de PAN avec la particule catalytique 20S (ainsi que ClpA avec ClpP et HslU avec HslV), la liaison de l'ATP doit donc induire une conformation de PAN qui aura une plus grande affinité pour le 20S. En outre, l'association entre PAN et le 20S n'est pas évidente à démontrer en absence de nucléotides (Smith et al, 2005; Wang et al, 2001).

Tous les efforts par cristallographie aux rayons X pour résoudre la structure de PAN et le complexe 19S ou de définir les changements de conformations lors de la liaison d'ATP n'ont pas eu de succès jusqu'à présent tandis que des données existent sur les structures de HslU (Wang et al, 2001) et de ClpX (Glynn et al, 2009) bactériens. Une compréhension complète des mécanismes moléculaires de dégradation des protéines par le protéasome exigera des informations structurales détaillées sur les formes nucléotide-libre et nucléotide-lié de PAN et du 19S eucaryote. Plusieurs hypothèses concernant les différentes étapes du processus de dépliement et de dégradation des substrats protéiques ainsi que le rôle de l'ATP, ont été obtenues à travers des études biochimiques sur le complexe PAN-20S et des études biochimiques et structurales sur les protéases bactériennes ATP-dépendantes (Navon & Goldberg, 2001) (Benaroudj et al, 2003) (Zhang et al, 2009a). Elles peuvent être résumées comme suit:

- ✓ La liaison du nucléotide sur PAN favorise l'association entre l'anneau ATPase et le complexe 20S.
- ✓ La formation du complexe déclenche l'ouverture de la porte de l'anneau α du protéasome 20S.
- ✓ La liaison du substrat protéique induit un changement de conformation dans PAN qui active l'hydrolyse de l'ATP.
- ✓ Des cycles répétés d'hydrolyse de l'ATP catalysent le dépliement des protéines globulaires.
- ✓ Le polypeptide déplié peut diffuser à travers l'anneau ATPase (dans sa forme ATP-lié) et la porte ouverte de l'anneau α .
- ✓ Le polypeptide dans la chambre centrale de la particule 20S est dégradé progressivement à des petits peptides.

Plusieurs questions détaillées sur la fidélité de ce régime et son applicabilité au complexe 26S eucaryote restent incertaines, nos travaux sur le complexe PAN nous ont permis de répondre à une partie de ces questions. Une compréhension complète du mode d'action de ces systèmes

protéolytiques aura besoin de plusieurs études structurales plus approfondies sur le complexe PAN associé avec le 20S et sur le protéasome 26S eucaryote pour mieux comprendre et déchiffrer les différentes étapes du mécanisme d'action de dépliement ATP-dépendant des substrats protéiques par les AAA+ unfoldases régulateurs de la protéolyse.

Dans cette première publication, une analyse de l'extrait cellulaire de *P. horikoshii* et *P. abyssi* a montré la présence d'une forme courte de PAN exprimée *in vivo*. Nos données biochimiques montrent que cette forme courte présente une activité de dépliement contre la GFP_{ssrA} et les propriétés structurales du PhPAN ont été analysées par cryo-microscopie électronique, cristallographie aux rayons X et diffusions de rayons X et de neutrons aux petits angles en solution (SAXS/SANS). La combinaison de ces approches structurales intégratives a permis la construction d'un modèle pseudo-atomique de PAN et d'obtenir des informations précieuses sur l'architecture de cette grande machine moléculaire ainsi que son mécanisme et son mode d'action. En outre, les changements de conformation de PhPAN induits par la liaison du nucléotide ont été également étudiés par génération de deux modèles Cryo-EM de PhPAN en présence et en absence d'ATP γ S permettant de mieux comprendre les différentes conformations adoptées par PAN et de proposer un mécanisme et un mode d'action décrivant le processus de dépliement alimenté par la liaison et l'hydrolyse de l'ATP et de suggérer un mécanisme d'interaction entre PAN et son partenaire catalytique 20S.

5.1.2 Structure of the assembled PAN unfoldase reveals new insights into the molecular architecture and mode of action of proteasome regulatory ATPases

Structure of the assembled PAN unfoldase reveals new insights into the molecular architecture and mode of action of proteasome regulatory ATPases

Ziad Ibrahim^{1,2,#}, Matteo Colombo^{1,#}, Gaëlle Hogrel³, Didier Flament³, Michael Härtlein², Susanna Teixeira², Eric Girard¹, Guy Schoehn¹, Frank Gabel^{1,2,*} and Bruno Franzetti^{1,*}

¹ Université Grenoble Alpes, CNRS, Commissariat à l'Energie Atomique, Institut de Biologie Structurale F-38044 Grenoble, France

² Institut Laue-Langevin, 71 Avenue des Martyrs, F-38000 Grenoble, France

³ Ifremer, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes Plouzané, France

contributed equally

* Corresponding authors: frank.gabel@ibs.fr, bruno.franzetti@ibs.fr

SUMMARY

PAN (proteasome activating nucleotidase) is the archaeal homolog of the eukaryotic 26S proteasome regulatory ATPases and promotes ATP-dependent protein unfolding and translocation into the 20S proteasome. Insight into the structure and function of this important representative of proteasomal ATPases has been very limited up to date. Here, we report a pseudo-atomic model of the assembled hexameric PAN particle, based on single particle Cryo-EM reconstruction, crystallography and solution scattering (SAXS/SANS), as well as *in vivo* expression and functional studies. Our models reveal a novel hexapod dome-shape with six legs attached to a central, hollow, half-spherical body enclosing a large cavity. Moreover, large conformational changes were observed by Cryo-EM in response to ATP γ S binding suggesting a structural mechanism for substrate unfolding and translocation as a function of nucleotide binding/hydrolysis as well as for interaction with the proteolytic core particle.

INTRODUCTION

The 26S proteasome is a 2.5 MDa cellular machine that is predominantly responsible for intracellular protein breakdown (Voges et al, 1999). The destruction of specific proteins by the ubiquitin-26S proteasome pathway allows to maintain a healthy proteome by eliminating misfolded or damaged proteins that, if overlooked, may give rise to toxic aggregates (Goldberg, 2003). Defective proteasome machinery is thus associated with ageing and a number of degenerative and hereditary diseases (Sherman & Goldberg, 2001; Vilchez et al, 2014). The proteasome system also plays a critical role in regulating many cellular processes by engaging a variety of key regulatory cellular proteins such as transcription and cell cycle factors (Lipford & Deshaies, 2003) in an ATP-dependent degradation. Misregulation of these processes is linked to the development of cancer and proteasome regulatory complexes represent important pharmaceutical targets (Grigoreva et al, 2015; Navon & Ciechanover, 2009). The 26S proteasome consists of a 20S cylindrical core particle (CP) and a 19S regulatory particle (RP) (Voges et al, 1999). The 20S core particle is formed by four stacked heptameric rings that house the proteolytic sites within a central chamber (Baumeister et al, 1998). The two outer and inner rings each consist of a single type or 7 different types of α and β subunits for archaea and eukaryotes, respectively. In order to avoid indiscriminate degradation of cytosolic proteins, substrate access to the 20S proteolytic chambers (composed of the β subunits rings) is limited by the α subunits forming an entrance gate (α -gate) that restricts the access to unfolded proteins (Lowe et al, 1995). The targeting and the access of ubiquitylated proteins to the 20S particle is a highly selective process achieved by the 19S regulatory particle (Bar-Nun & Glickman, 2012). The 19S machinery is devoted to ubiquitinated protein recognition, deubiquitination, unfolding and α -gate opening to translocate the substrates to the 20S core particle (Pack et al, 2014). It comprises a lid devoted to recognition of target proteins and a base harboring the ATP-dependent unfoldase activity (Lander et al, 2012; Lasker et al, 2012). The lid is formed by nine non-ATPase subunits (Rpn) responsible for recognition and deubiquitination of protein substrates. The base harbors three non-ATPase subunits and six different AAA+ ATPase subunits (Rpt1-Rpt6). The Rpt proteins form a hexameric complex that fulfills multiple tasks: interaction with the 20S peptidase core, promotion of α rings gate opening, catalysis of the energy-dependent unfolding of the target proteins and their translocation into the 20S degradation chamber (Kim et al, 2013) (Smith et al, 2005). Thus, determining the mode of action of the Rpt rings is essential to understand the regulation of the proteasome activity. However, there is still limited structural information on the internal architecture of the eukaryotic Rpt unfoldase

rings. Moreover, the molecular mechanism coupling the nucleotide hydrolysis with the unfoldase activity is currently unclear (Förster et al, 2014; Sauer & Baker, 2011).

In eukaryotes, the multi-subunit composition and the high dynamic properties of the proteolytical AAA+ ATPase represent major bottlenecks to obtain high-resolution structures of individual Rpt subunits as well as the assembled unfoldase ring. Moreover, the flexibility of the eukaryotic 19S sub-complex is presumably reflected in multiple conformations (Asano et al, 2015), which in turn contribute to the difficulty of generating high-resolution structural models of these complexes. Archaeal cells share many important features of cell biology with eukaryotes (Maupin-Furlow et al, 2006), to whom they are evolutionarily related (Woese, 1998) (Raymann et al, 2015), and are therefore good model systems. The proteasome activating nucleotidase (PAN) is the archaeal counterpart of the eukaryotic regulatory AAA+ ATPase hexameric ring. The PAN hexamer is formed by a single type of subunit that shares 41-45% sequence identity with the eukaryotic Rpts subunits. Due to its homogeneous composition, thermostability and ease of expression in *Escherichia coli*, PAN offers many technical advantages over the more complex eukaryotic system. The PAN complex from *Methanococcus jannaschii* (MjPAN) is the best characterized biochemically (Smith et al, 2005) (Zwickl et al, 1999) (Benaroudj & Goldberg, 2000) (Zhang et al, 2009b). It displays protein unfoldase and chaperone activities *in vitro* (Benaroudj & Goldberg, 2000). However, there is a dramatic lack of structural and dynamical information on the assembled complex. A pseudo-atomic model of the MjPAN complex was proposed by Zhang et al. (Zhang et al, 2009a), based on X-ray data obtained from two unassembled subdomains generated by limited proteolysis: the hexameric N-terminus (residues 74-150) and the monomeric ATPase domain (residues 158-418) (Wilson et al, 2000; Zwickl et al, 1999). In this model, the ATPase hexameric ring was modeled by the bacterial HslU structure (Bochtler et al, 2000; Sousa et al, 2000) and the hexameric N-terminal domain was placed over the ATPase ring, based on low resolution electron microscopy images from Smith et al (Smith et al, 2005). The proposed MjPAN pseudo-atomic structure was used as a homology model to determine the structural organization of the AAA+ ATPases subunits in the Cryo-Electron Microscopy (Cryo-EM) structures of human and yeast 26S proteasome (Beck et al, 2012; da Fonseca et al, 2012). In the resulting model for Rpt hexamers, the eukaryotic C-terminal domain architecture was found to be more open than the one proposed in the PAN model (Zhang et al, 2009a). Thus, there is a lack of information relative to the coupling between the N-terminal and the C-terminal domains and there is a need for more accurate structural data on functional,

assembled PAN complexes, in particular regarding the structural rearrangements of PAN upon ATP binding and hydrolysis.

A major obstacle to obtain high resolution structural details on PAN complex was the difficulty to produce monodisperse and soluble populations of recombinant hexamers. Here, we demonstrate the existence of a shorter version of PAN, lacking the first 34 amino acids at the N-terminal, and that is predominantly accumulated in Thermococcales. In particular, the short version of *Pyrococcus horikoshii* PAN (*PhPAN*) was found to be stable *in vitro* and we developed protocols to obtain monodisperse populations of active hexamers. The structure of the *PhPAN* hexamer was determined by single particle Cryo-EM in ATP-free and ATP γ S-bound conformations and compared with small-angle scattering data in solution. Moreover, we determined the X-ray structure of the *PhPAN* C-terminal domain. The integrative structural biology approach allowed us to generate a pseudo-atomic model of the assembled PAN complex which displays a significantly different architecture compared to previous models (Medalia et al, 2009; Zhang et al, 2009a). The proposed model also reveals important features conserved between archaeal and eukaryotic proteolytical AAA+ ATPases which clearly distinguishes them from bacterial systems. The structural data of *PhPAN* in the ATP-free state and the ATP γ S-bound state shed light on the ATP-dependent unfolding process (Iosefson et al, 2015; Martin et al, 2008; Navon & Goldberg, 2001): important contraction motions, associated with nucleotide-binding, were observed, indicating conformational transitions coupled to ATP-binding and hydrolysis. These observations are consistent with previously proposed models explaining ATP-dependent unfolding processes (Martin et al, 2008; Navon & Goldberg, 2001) and suggest that substrate proteins are unfolded on the surface of PAN.

RESULTS

Identification, purification and functional characterization of a native, hexameric PAN complex from Pyrococcus horikoshii

It was shown previously that in the extreme halophilic archaeon *Halobacterium salinarium* the use of a second translation initiation site resulted in the accumulation of a short version of PAN proteins (Chamieh et al, 2012). Likewise, in *Pyrococcus horikoshii* and *Pyrococcus abyssi* PAN (*Ph*PAN and *Pa*PAN), primary sequence analyses revealed the existence of an alternative translation initiation site at positions 35 (Met³⁵) and 32 (Met³²), respectively (Fig. S1). These observations raised the question of the length of PAN proteins expressed in thermococcales cells. Total protein extracts obtained from *P. horikoshii* and *P. abyssi* cells were analyzed by western blot with anti-PAN specific antibody and the positions of the immunoreactive bands were compared to those obtained for the recombinant full-length and short variants of *Ph*PAN and *Pa*PAN proteins. The results showed that, under normal growth conditions, the short PAN version is the only form expressed *in vivo* (Fig. 1A and B). In the case of *P. abyssi*, kinetic studies of the gene product accumulation indicated that this shorter form of PAN is produced at all stages of cell growth (Fig. 1C, red triangles). Unlike the full-length protein, the short PAN version from *P. horikoshii* was highly soluble when expressed in *Escherichia coli* and remained soluble after purification by two anion exchange chromatography steps. Gel filtration profiles showed that the majority of the purified complexes were assembled as dodecamers (Fig. 1D, blue curve), which corresponded to two stacked hexameric rings as indicated by negative stain electron microscopy micrographs (Fig. 1E, pool A). By heating *Ph*PAN dodecamers at 60 °C in the presence of ATP and MgCl₂, monodisperse, hexameric *Ph*PAN particles were obtained (Fig. 1D, black curve and Fig 1E, pool B). Small angle X-ray (SAXS) and neutron (SANS) scattering studies on the dodecameric and hexameric *Ph*PAN complex in solution, in the presence or absence of ATP and ATP γ S, revealed the important role played by ATP in inducing PAN hexamerization (Supplemental Information and Fig. S2). Unfoldase assays, using GFPssrA as a substrate, were performed to assess the functional state of the purified hexameric PAN populations (Benaroudj, 2000). The addition of threefold molar excess of *Ph*PAN resulted in a > 90% loss of the initial GFPssrA fluorescence (Fig. 1F). In contrast, no unfolding of GFPssrA was observed when non-hydrolysable ATP γ S was added instead of ATP. Thus, the short version of PAN from *P. horikoshii*, is expressed *in vivo* and can catalyze the energy-dependent unfolding of a globular protein in its hexameric state.

Single-particle Cryo-EM reconstruction of the ATP-free PhPAN hexamer reveals structural features that differ markedly from those of other AAA+ unfoldases

To date, no atomic-resolution structure is available for the PAN complex. The purified *PhPAN* hexamers remained stable over time (> 48 h) at concentrations compatible with electron microscopy, small angle scattering and biochemical assays. In negative-stain EM micrographs, *PhPAN* hexamers assumed preferentially random orientations (Fig. 1E, pool B), some particles with top-view orientations clearly showing the overall architecture of the complex as a symmetric, single hexameric ring. Single particle Cryo-EM reconstruction yielded the first structure of the assembled hexameric PAN regulatory particle at 10-15 Å resolution (Fig. 2). The particle displays a hexapod dome-shape with six legs attached to a central, hollow, half-spherical body. The top-view of the particle shows a small entrance (~10Å), arranged on the 6-fold symmetry axis (Fig. 2, left, and inset), while the bottom-view reveals a large semi spherical internal cavity with a ~ 71 Å opening diameter to the C-terminal side that is exposed to the 20S catalytic particle of the proteasome (Fig 2, right). The particle orientation and the localization of each domain was determined by fitting the high resolution crystal structures of the N-terminal and C-terminal domains inside the Cryo-EM map (Fig. 3). This structure is markedly different from previous 3D reconstructions of the whole PAN complex (Medalia et al, 2009). The compact, central body of the complex displays a diameter of ~ 60 Å and a height of ~ 27 Å (Fig. 2). Overall, the height of the *PhPAN* particle is ~ 80 Å and the diameter is maximal at the C-terminal side with a value of ~122 Å. Moreover, the large internal cavity and the ATPase module (C-terminal) conformations are very different from previously proposed PAN models as well as from other hexameric ATPases such as HslU (Bochtler et al, 2000) and ClpX (Glynn et al, 2009). Indeed, the bacterial AAA+ unfoldases and the previously proposed PAN models lacked this wide internal cavity and the protruding legs on the C-terminal domain that was always shown as a flat side. The structural data presented here highlight the scaffold role of the N-terminal sub-domain, from which the ATPase sub-domains extend to interact with the 20S particle (Smith et al, 2005). Indeed, the constriction of the α ring of the core particle is too narrow to allow access of peptide chains and exists in a closed state (Fig. 4C). Opening of the gated channel of the α ring is thought to be induced by binding of the regulatory particle (RP) (Groll et al, 2000). The C-terminal residues encompassing the HbYX motif of PAN (Fig. 3) were shown to bind directly to the 20S core particle (Förster et al, 2005; Rabl et al, 2008; Smith et al, 2007) driving the α -gate opening and substrate processing. Interestingly, by superimposing the *PhPAN* Cryo-EM map on the atomic structure of the 20S proteasome, it is noteworthy

that the ATPase domains do not match with the α -subunit ring pockets to induce gate opening (Fig. 4C, left), suggesting a conformational change of the *PhPAN* C-terminal domain, driven by ATP.

Generation of a pseudo-atomic, hexameric PhPAN model combining Cryo-EM, X-ray crystallography and SAXS data

AAA+ ATPases are ATP-dependent molecular machines with highly conserved oligomeric structures (Baker & Sauer, 2006; Lupas & Martin, 2002; Sauer et al, 2004). Surprisingly, the Cryo-EM structure of *PhPAN* at 10-15 Å shows a different arrangement of the C-terminal ATPase domains compared to other AAA+ ATPases. In order to obtain insight into the *PhPAN* structure, we expressed and purified its individual N-terminal (residues 1-101) and C-terminal (residues 102-365) domains and determined the X-ray structure of the monomeric C-terminal domain (*PhPAN*-Cter) at 1.94 Å resolution (Fig. 3A). In addition, by a SAXS approach, we found that the hexameric N-terminal domain (*PhPAN*-Nter) displays high structural homology with the X-ray structure of the PAN N-terminal domain from *Methanococcus jannaschii* (Fig. S3A).

Together, the data allowed the generation of a pseudo-atomic model of the whole *PhPAN* particle. To this end, the hexameric model of the N-terminal sub-domain of *MjPAN* (PDB 3H43) (Zhang et al, 2009a) and the X-ray structure of *PhPAN*-Cter were fitted into to Cryo-EM map (Fig. 3B) using the program UCSF Chimera (Pettersen et al, 2004). The six oligonucleotide/oligosaccharide-binding fold (OB-fold) of the N-terminal domain are well placed within the Cryo-EM density (correlation coefficient $CC_{\text{overall}} = 0.88$), while the coiled-coil domains are not visible. Indeed, the Cryo-EM map corresponds to the short version of the *PhPAN* complex lacking the first 34 residues that form a large part of the coiled-coils as predicted by bioinformatics tools (PHYRE2 Protein fold Recognition Server) (Kelley et al, 2015). The C-terminal side of *PhPAN* in the Cryo-EM map presented here differs significantly from the hexameric ring of other AAA+ machineries (ClpX, HslU). By fitting the monomeric *PhPAN*-Cter structure in the Cryo-EM map, a model of the AAA+ ring was generated (Fig. 3B, red & green). The correlation coefficient of each *PhPAN*-Cter molecule fitted into the Cryo-EM map of the whole *PhPAN* was 0.72. In this model the pore I (Ar- Φ loop, residues 180-182) and II (residues 220-222) loops are at the interface contacts between *PhPAN*-Cter subunits. Similarly to the Cryo-EM structure, the generated pseudo-atomic model has a dome-shaped structure with a wide internal cavity (Fig. 3B). In the pseudo-atomic model, The opening at the C-terminal side, allowing access to the unfolded substrate,

has a 50 Å diameter and appears dramatically wider than the opening of other hexameric AAA+ unfoldases like the bacterial HslU (7 Å) (Bochtler et al, 2000; Sousa et al, 2000). The larger ring has a diameter of 120 Å and thus PAN resembles a “top hat” capping the 20S particle.

While the Cryo-EM structure, and the pseudo-atomic model based on it, provide valuable insight into the structure of the free PAN complex, they represent static snapshots of specific conformations due to the experimental conditions (Round et al, 2013). By recording SAXS curves from hexameric *Ph*PAN, we obtained important complementary data in solution which mimic the PAN physiological environment more closely and encode information on its assembly state and conformation. The SAXS curve back calculated from the Cryo-EM densities of the ATP-free PAN complex using the program EM2DAM (Petoukhov et al, 2012) and fitted with CRY SOL (Svergun et al, 1995), indicates a fair agreement between the EM model and the experimental SAXS data (Fig. 3C, red curve) with a χ value of 1.67. The pseudo-atomic *Ph*PAN model, based on the Cryo-EM density, was also validated by SAXS and fitted the experimental data well (Fig. 3C, green curve) with a χ value of 1.36. Both radii of gyration were also in good agreement; 49 Å for the pseudo-atomic model and 45 Å for the Cryo-EM model compared to the experimental value of 51 Å. The minor discrepancies at intermediate Q-values might be due to difference in the sequences and residue numbers between the rigid building blocks (340 residues) and the construct in solution (365 residues). In addition, *ab initio* low-resolution envelopes (program DAMMIF (Franke, 2009)) were generated from the hexameric *Ph*PAN SAXS data, imposing P6 symmetry, and revealed a hat-shaped structure with six legs attached to a central body (Fig. S5). The solution shapes and dimensions were compatible with the Cryo-EM and pseudo-atomic models. Finally, an alternative model was built and tested by superimposing the *Ph*PAN-Cter structure with the bacterial HslU hexameric ring (Fig. S4). The hexameric N-terminal domain of *Mj*PAN was positioned above the generated AAA+ ring, forming the *Ph*PAN-HslU model as described in (Zhang et al, 2009a). This model displayed only a poor fit ($\chi = 2.32$) against the experimental SAXS data (Fig. S4, red curve) and could be discarded, thus strongly suggesting that the conformation observed in the Cryo-EM map reported in this manuscript is present in solution. In conclusion, experimental SAXS data were very useful to probe the accuracy of the pseudo-atomic models based on Cryo-EM. To our knowledge, this is the first time that a pseudo-atomic model of the PAN complex is validated in solution. Finally, the *Ph*PAN Cryo-EM structure reported here was generated in the absence of ATP and its conformational state is not fit to interact with the 20S particle (Fig. 4C), thus providing a structural explanation for

previous studies that have shown that PAN-20S interaction requires ATP binding/hydrolysis (Smith et al, 2005). Taken together, these observations suggest that significant conformational changes occur after nucleotide-binding to PAN.

The PAN complex switches between two conformations in response to nucleotide binding

Despite their crucial role in many cellular pathways, the modes of action of the AAA+ superfamily members are still not completely deciphered, and the link between the ATPase catalytic cycle and the conformational changes responsible for the biological activity of these proteins has yet to be elucidated (Rouiller et al, 2000). In the ATP-free *Ph*PAN Cryo-EM map presented here, the C-terminal domain does not match the α -ring of the 20S core particle (Fig. 4C), in line with previous work that demonstrated that ATP is required for the PAN-20S interaction (Smith et al, 2005). In order to elucidate the structural basis of the ATP-driven transition we determined the Cryo-EM structure of the ATP γ S-bound *Ph*PAN at 25 Å resolution (Fig. 4B). In the presence of ATP γ S, PAN was presumably found to be locked in a conformation similar to that of the ATP-bound state (Medalia et al, 2009). Figure 4 highlights the changes induced upon ATP γ S binding: although the overall topology of PAN did not change, the internal cavity of the complex adopted a more compact conformation upon nucleotide binding (>10 Å contraction of its diameter). Moreover, the size and orientation of the C-terminal side ‘legs’ are different. Indeed, in the nucleotide-bound state, the PAN ‘legs’ are relatively flat compared to the ATP-free *Ph*PAN where the complex shows longer and more inclined C-terminal domains. Moreover, the C-terminal extensions change their orientation and move upward upon ATP γ S binding.

Concomitantly with the C-terminal structural modifications, the overall height of the ATP γ S-bound *Ph*PAN complex is ~60 Å, i.e. 20 Å less than the ATP-free complex (Fig. 4). As mentioned above, the ATPase domains of the nucleotide-free PAN do not match the α -subunit ring diameter of the 20S catalytic complex. Interestingly, after ATP γ S binding, the diameter of the C-terminal nucleotidase domain switched from ~122 Å to ~111 Å which is almost the same diameter as the α -proteasome ring (~112 Å) (Fig. 4C), thus facilitating the interaction between both particles necessary to induce gate opening and substrate translocation. These data provide a structural explanation for the ATP-dependence of the PAN-20S interaction in agreement with previous biochemical findings (Smith et al, 2005). Finally, the theoretical SAXS curve of the Cryo-EM PAN model in its ATP γ S-bound state (Fig. S4, green curve), was compared against experimental SAXS data from ATP-free *Ph*PAN hexamer in solution. The agreement was very poor, supporting that major conformational changes occur in PAN

upon ATP γ S binding in solution (37 Å for the theoretical radius of gyration of the PAN complex in its ATP γ S bound state vs. 51 Å of the ATP-free PAN).

Taken together, our data suggest that PAN is a highly dynamic macromolecular complex switching between two major conformational states represented as a “ground state” and a “substrate-processing state” which is adopted upon nucleotide binding.

DISCUSSION

An active, short version of the PAN protein is present in vivo

Despite the major role played by the proteolytic AAA+ proteins in the regulation of intracellular proteolysis, there is little information available regarding their physiological isoforms (Chamieh et al, 2008). In the present paper, we show that PAN accumulates in *Pyrococcus horikoshii* and *Pyrococcus abyssi* cells with a shorter N-terminal domain lacking the first 34 and 31 amino acids, respectively (Fig. 1A-C). Moreover, the short PAN displayed substrate unfolding activity (Fig. 1F). These results are in line with previous findings that demonstrated that short variants of PAN proteins, lacking the first 38 (*HsPANA*) and 35 (*HsPANB*) amino acids, are expressed in the extreme halophilic archaeon *Halobacterium salinarium* (Chamieh et al, 2012). In the latter organism, RNA mapping revealed the existence of shorter transcripts suggesting the use of a second, internal translation initiation site (Chamieh et al, 2008). Likewise, the bacterial proteolytic regulatory AAA+ proteins ClpA and ClpX (that assemble with the ClpP peptidase) were also detected *in vivo* as two isoforms (Gerth et al, 2004) and internal translation initiation mechanisms have been found for ClpA unfoldases (Park et al, 1993). Together, these observations indicate that N-terminal heterogeneity and the use of a second translation initiation site represent widespread regulatory strategies for prokaryotic ATPases implicated in intracellular proteolysis even though the functional significance of PAN N-terminal shortening remains unclear.

In *H. salinarum* the N-terminal truncated variant was found to display better association with the 20S proteasome in cellular extracts (Chamieh et al, 2008). Similarly, the N-terminal truncated variant of the Cdc48 unfoldase displays higher affinity for the 20S proteasome core particle compared to the full-length Cdc48 (Barthelme & Sauer, 2012). Moreover, the N-terminal subdomain of *PhPAN* assembled as a dodecameric complex in solution, similarly to the *MjPAN* N-terminal domain (Zhang et al, 2009a). These observations strongly support the role played by the N-terminal domain of PAN in the stability and the control of the oligomerization state of the PAN complex. Structural studies on the PAN complex from *Methanococcus jannaschii* (Zhang et al, 2009a) showed that the N-terminal part of the protein corresponds to a flexible domain of approximately 150 residues located outside the hexameric rings when the complex is assembled. This domain possesses a coiled-coil motif that displays chaperone activity and it also devoted to substrate recognition (Lee et al, 2003). Thus, the *in vivo* N-terminal truncation may also represent a way to modify the biochemical activities of the PAN proteolytic regulators as well as their substrate specificity.

ATP and heat induce dissociation of PAN dodecamers

While the hexameric ring architecture represents the active state of AAA+ proteases (Sauer & Baker, 2011) we found that the purified *Ph*PAN hexamers self-assembled into stable, double-stacked hexamers which might be explained by the fact that PAN dodecamers represent a storage form of the protein *in vivo*. Indeed, Wilson et al. have shown that PAN produced in *M. jannaschii* is purified from the cell lysate as a distinct 550 kDa complex showing that the protein is associated as dodecamers *in vivo* (Wilson et al, 2000). Our results also highlight that PAN hexamerization is induced by ATP hydrolysis and not only by binding of the nucleotide as demonstrated by SANS experiments showing that ATP γ S failed in promoting PAN hexamerization but only induces small conformational changes in the dodecamer (Fig. S2B). Thus, ATP hydrolysis is needed to dissociate the stable dodecamers into free active hexamers.

The structure of the proteasome regulatory particle PAN reveals a novel architecture, conserved between archaea and eukarya

The Cryo-EM structure presented here provides the first structure of the assembled hexameric PAN regulatory particle (Fig. 2). A pseudo-atomic model was generated by fitting the high-resolution X-ray crystallography structures of the N-terminal and C-terminal domains into the Cryo-EM density maps and subsequently validated by SAXS analysis in solution (Fig. 3). While the overall architecture of the PAN complex presented here is consistent with previous low resolution negative stain EM studies which indicated a two ring architecture (Smith et al, 2005; Smith, 2005), it is markedly different from previous medium resolution 3D models obtained from the whole PAN complex from *M. jannaschii* (Medalia et al, 2009; Zhang et al, 2009a) in terms of the wide internal cavity exposed to the 20S catalytic particle and the elongated C-terminal domains that are missing in previous PAN models. Indeed, we show here that the *Ph*PAN complex consists of a hexapod structure with six legs protruding from the C-terminal domain around a central body and containing a large cavity, widely open toward the 20S proteasome. This internal organization has, to our knowledge, not been observed previously. Instead, in the pseudo-atomic model proposed by Zhang et al. (Zhang et al, 2009a), access to the 20S proteasome appeared to be mediated by a small pore formed by the conserved aromatic phenylalanine residues in the pore 1 loop (Ar- Φ loop) of the C-terminal ATPases domain. The model, however, was built based on a superposition of the C-terminal ATPase domain of PAN with the bacterial AAA+ regulatory particle HslU. Likewise, the protruding C-terminal domain architecture existing in our model is not shared with bacterial regulatory particles (Bochtler et al, 2000; Glynn et al, 2009). A second

important and novel feature are the extended legs that may play an important role in the docking and interaction with the 20S catalytic particle as they contain the critical hydrophobic-tyrosine-X (HbYX) motif on their C-termini, conserved in the eukaryal Rpts, that docks into pockets in the 20S to stimulate the opening of a gated substrate entry channel (Yu et al, 2010).

Our structural data on the PAN complex show that its architecture differs from those of the bacterial regulatory particles ClpX and HslU in terms of the two ring architecture, internal cavity and the protruding C-terminal domain (Bochtler et al, 2000; Glynn et al, 2009). Furthermore, the PAN structure fits into the Cryo-EM envelope determined for the eukaryotic ring complex (Fig. 5), built by the six ATPases subunits situated at the base of the human and yeast 19S particles (Rpt1-6) (Lander et al, 2012), underlining a high degree of conservation between archaeal and eukaryotic unfoldases. In addition, they suggest similar mechanisms for substrate unfolding and nucleotide-induced conformational changes between proteasome regulators from these two domains of life, and justify the use of the "simple" PAN complex as a model system to study the basic mode of action of the eukaryotic proteasome regulatory complex.

Large structural transitions in response to nucleotide binding: implications for unfoldase mechanisms and PAN-proteasome coupling

In order to better understand the conformational switching that enable substrate unfolding in AAA+ unfoldases, we assessed the different conformational states of the hexameric *Ph*PAN complex associated with nucleotide binding by Cryo-EM. Two different conformations of PAN were found in the ATP-free state and in the presence of ATP γ S (Fig. 4). Upon ATP γ S binding, an important, global contraction of the complex occurs. The ATP γ S-bound state represents a locked conformation of PAN during the ATP binding/hydrolysis process and therefore a static snapshot of one specific state. The C-terminal domain adopts a more compact conformation which fits the α proteasome diameter (Fig. 4B), most likely explaining the necessity of nucleotide binding to induce the PAN-20S interaction as observed previously by biochemical studies (Smith et al, 2005).

The contracted conformation of PAN in the presence of ATP γ S probably correspond to an intermediate state of the molecule during the substrate unfolding process. However, the contracting motions may affect only monomers separately and not occur necessarily in a cooperative way. Cycles of transitions between "relaxed" and "compact" conformations probably take place during nucleotide hydrolysis and may be associated with protein

unfolding. This observation is consistent with Cryo-EM studies of the native 26S proteasome from *Saccharomyces cerevisiae* showing that major conformational changes occur in the regulatory particle upon ATP γ S binding and during substrate processing, suggesting that large scale motions facilitate substrate unfolding (Matyskiela et al, 2013; Śledź et al, 2013). The remarkable conformational changes and contraction motions observed in the PAN complex in response to nucleotide-binding support a mode of action consistent with the one proposed for the bacterial unfoldase ClpX (Iosefson et al, 2015): in this model, the unfoldase would pull on exposed hydrophobic patches, presented by the substrate, by using power strokes driven by the ATPase cycle and leading to the contraction and relaxation cycles of the complex and to unfolding of the polypeptide substrates. However, our pseudo-atomic model (Fig. 3B) indicates that the OB ring, formed by the six PAN N-terminal domains, adopts a compact conformation which delimitates a small pore (~ 12-15 Å) preventing the passage of globular proteins into the PAN internal cavity. Indeed, biochemical studies performed by Navon and Goldberg (Navon & Goldberg, 2001) have indicated that protein unfolding is initiated on the surface of the OB domain ring. It is therefore reasonable to propose a mechanism in which the important structural transitions powered by ATP hydrolysis observed here may initiate the unfolding process of the substrate proteins at the PAN surface and that a coordinated movement of the conserved pore loops (YIG in *Ph*PAN), situated within the large internal cavity, would grip concurrently the incoming protein and apply a pulling force. Alteration of the hydrophobic character of the internal surface of PAN may also result in a greater interaction with hydrophobic surfaces of the partially unfolded protein, dissolving the substrate step by step on the internal surface of the complex.

Several models have been proposed to explain the concerted action of proteolytical unfoldases with peptidase core particles leading to the complete destruction of substrate proteins (Smith et al, 2005; Zhang et al, 2009b). Since free PAN hexamers can unfold globular proteins such as GFP_{ssrA}, the unfoldase activity of the PAN system does not require 20S binding. In fact, in archaea the PAN-proteasome complex is a labile complex *in vivo* (Smith et al, 2005) and it has been suggested that, in the cytosol, free PAN hexamers can perform substrate binding and ATP-dependent protein unfolding and/or chaperone activity (Benaroudj & Goldberg, 2000). However, we showed here that the conformational changes induced by ATP hydrolysis favor a more compact form with “contracted” C-terminal domain of PAN which could promote the C-terminal α -helix to dock into the 20S proteasome (Rabl et al, 2008) allowing the opening of the α -ring and the translocation of the unfolded substrate through the axial channel of the core particle where peptide bond cleavage occurs. This mechanism is supported by surface

plasmon resonance (SPR) experiments that demonstrated the ATP-dependence of the PAN-20S interaction (Smith et al, 2005). On the other hand, the association with the 20S α rings would accommodate the large ATP-induced motions needed for the unfolding process. Therefore, the most likely mechanism that we can propose, based on our data, is that substrate unfolding and translocation in the 20S particle are two dissociated mechanisms but that substrate unfolding promotes the transient association of the PAN complex to the surface of the 20S proteasome. This model is consistent with the mechanism proposed by Navon and Goldberg who found that proteins are unfolded before their translocation into the proteasome takes place and that, if translocation of GFPssrA through ATPase ring was prevented, unfolding of GFP still occurs (Navon & Goldberg, 2001). Other work also showed that, if a globular protein is first denatured, ATP is no longer necessary for translocation and degradation (Smith et al, 2005). Together, these observations lead us to favor a mechanism where PAN functions on the surface of the 20S proteasome but without any stepwise translocation. To confirm this mechanism, further dynamical and structural studies are needed on the PAN-substrate and PAN-proteasome complexes. The structural and mechanistic data provided here allowed gaining valuable insights into the mechanism of these molecular machines that sheds light on the 26S proteasome mode of action, in particular the nucleotide-dependent conformational switching of the 19S eukaryotic regulatory particle during substrate processing motorized by the ATPase activity of the Rpts base subunits.

EXPERIMENTAL PROCEDURES

***Pyrococcus horikoshii* and *Pyrococcus abyssi* cell cultivation**

The hyperthermophilic microorganisms *P. horikoshii* and *P. abyssi* were cultivated as described Godfroy et al., 2006 (Godfroy et al, 2006). For more details, see Supplemental Experimental Procedures.

Proteins

Proteins were expressed and purified using standard protocols. For more details, see Supplemental Experimental Procedures.

X-ray crystallography

*Ph*PAN-Cter crystals were grown using standard protocols. The data set was collected at the European Synchrotron Radiation Facility (ESRF) beamline ID29 (Grenoble, France). For more details see Supplemental Experimental procedures.

SAXS/SANS

X-ray and neutron scattering data were collected at the BioSAXS beamline BM29 at the ESRF (Grenoble, France) and on the D22 instrument at the Institut Laue-Langevin (ILL) (Grenoble, France) respectively. For more details see Supplemental Experimental Procedures.

Negative-stain and Cryo-electron microscopy

See Supplemental Experimental Procedures.

Unfolding activity

The *Ph*PAN unfolding activity was measured as described in Zhang et al., 2009 (Zhang et al, 2009b) by using GFPsrA as a model substrate. For more details see Supplemental Experimental Procedures.

AUTHOR CONTRIBUTIONS

Z.I. and M.C. conducted experiments and wrote the paper. G.H. and D.F. conducted experiments. M.H., S.T. and E.G. supervised work. G.S. conducted and designed experiments. F.G. and B.F. designed experiments, supervised work and wrote the paper.

ACKNOWLEDGMENTS

We acknowledge financial support from the Agence Nationale de la Recherche (grant ‘Archelyse’ ANR-12-BSV8-0019-0). Ziad Ibrahim was supported by an ILL PhD fellowship. Matteo Colombo was supported by a French National Research Agency postdoctoral fellowship.

This work used the platforms of the Grenoble Instruct Center (ISBG : UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The electron microscope facility is supported by the Rhône-Alpes Region, the Fondation Recherche Medicale (FRM), the Fonds FEDER and the GIS-Infrastructures en Biologie Sante et Agronomie (IBISA). SAXS data were recorded at the ESRF BioSAXS beamline BM29 (local contact: Martha Brennich) and SANS data at the ILL instrument D22 (local contact: Anne Martel) using the BAG system.

FIGURE LEGENDS

Figure 1. Pyrococcales express a short and active isoform of PAN *in vivo*.

(A) *P. horikoshii* proteins were resolved on 12% SDS-PAGE: (1) *P. horikoshii* total extracts, (2) protein extracts from *E. coli* cells expressing recombinant full length *PhPAN* and (3) the shorter isoform of *PhPAN*. **(B)** *P. abyssi* total extracts (1) and purified recombinant short *PaPAN* (2) were resolved on 12% SDS-PAGE. The different isoforms of all PAN extracts were detected by western blot with the corresponding anti-PAN antibody. **(C)** Kinetics of the *PaPAN* gene expression in a *P. abyssi* cell line. Growth pattern of *P. abyssi* cells at different time points: in 10 hours of cell culture, cell population increased to 4.5×10^8 cells/mL reaching a plateau. Inset: expression of the PAN gene at different stages of the cell growth (red triangles). *P. abyssi* total extracts at different stages of the cell growth were resolved on 12% SDS-PAGE and analysed by western blot showing that the short *PaPAN* is always produced. **(D)** Size exclusion chromatography profiles of the dodecameric (blue) and the hexameric (black) forms of the short *PhPAN*. Hexamerization of *PhPAN* dodecamers was obtained by heating the protein complex for 30 min at 60 °C in the presence of 4 mM ATP. The transition from dodecameric *PhPAN* (pool A) to hexameric *PhPAN* (pool B) was confirmed by single particle negative staining electron microscopy micrographs **(E)** and by SAXS and SANS (see Supplemental Information). **(F)** *PhPAN* catalyzes ATP-dependent unfolding of the globular protein GFPssrA. Time course of fluorescence changes of GFPssrA in the unfolding buffer at 65 °C in the presence of *PhPAN* hexamers at 3:1 molar ratio together with 4 mM ATP (red curve) or 4 mM ATP γ S (green curve). The black curve shows GFPssrA fluorescence quenching at 65 °C, due to the GFPssrA reversible thermal denaturation, in the absence of *PhPAN*.

Figure 2. Single particle Cryo-Electron Microscopy 3D reconstruction of ATP-free *PhPAN* hexamer. Cryo-EM model of the ATP-free *PhPAN* hexamer. From left to right: top view, side view and bottom view of the electron density of *PhPAN* hexamer at 10-15 Å resolution. The entrance, N-terminal and C-terminal domains and the internal cavity are shown.

Figure 3. Pseudo-atomic model of the *PhPAN* complex.

(A) Structure of the *PhPAN*-Cter, solved by X-ray crystallography. The α/β fold and the helical domain of the *PhPAN* nucleotidase domain are shown in red and green, respectively. ADP is bound to the hinge region between these two domains. **(B)** The crystal structures of

M. jannaschii PAN N-terminal domain ring (*Mj*PAN-Nter) (PDB ID code 3H43) (blue) and of *Ph*PAN-Cter (red & green) in cartoon representation fit well into the *Ph*PAN electron density map (correlation coefficients 0.72 and 0.88, respectively), as shown in side and top views allowing the localization of its N-terminal and C-terminal domains and the generation of a pseudo-atomic model of the *Ph*PAN complex. (C) Experimental SAXS curve of the *Ph*PAN hexamer in solution (blue points) and the back-calculated SAXS curves (CRY SOL) from the ATP-free *Ph*PAN Cryo-EM electron density map (EM2DAM) (red curve) and of the pseudo-atomic *Ph*PAN structure (green curve). The experimental SAXS curve superposes reasonably with the back-calculated ones from the *Ph*PAN pseudo-atomic model ($\chi = 1.36$) and the *Ph*PAN Cryo-EM map ($\chi = 1.67$), indicating that the structural arrangement of both models is similar in solution. Inset: Guinier plot of the experimental SAXS data displaying a straight line, which supports the monodispersity of the sample.

Figure 4. PAN structural transition in response to nucleotide binding.

Top and bottom views of the ATP-free *Ph*PAN (A) or in the presence of ATP γ S (B). The two models are displayed at 25 Å resolution (resolution of the ATP γ S *Ph*PAN dataset). ATP γ S-*Ph*PAN shows flattened C-terminal domains compared to the ATP-free *Ph*PAN in which the C-terminal domains are more elongated. The C-terminal domain, N-terminal domain and the internal cavity diameters as well as the overall height of the complex are shown in both cases indicating a clear compaction of the molecule upon nucleotide binding. (C) Suggested models of the *Ph*PAN-20S proteasome (*Thermoplasma acidophilum* (PDB ID code 1PMA)) assembly with ATP-free *Ph*PAN (left) and ATP γ S-*Ph*PAN (right). The ATP γ S-*Ph*PAN displays a compatible diameter to interact with the 20S proteasome, while the ATP-free *Ph*PAN does not.

Figure 5. Comparison between the archaeal and eukaryal AAA+ unfoldases. A hypothetical pseudo-atomic model of the full length *Ph*PAN complex was generated by adding the predicted missing coiled coil residues (shown in red) on the N-terminal side of the generated pseudo-atomic model using the program PHYRE2 (Kelley et al, 2015). The generated complete model was also fitted into the Cryo-EM density map of the eukaryotic 19S base (Rpt 1-6 ATPases subunits) from *Saccharomyces cerevisiae* generated from the Cryo-EM structure of the complete 26S proteasome (9 Å, EMDB-1992).

FIGURES

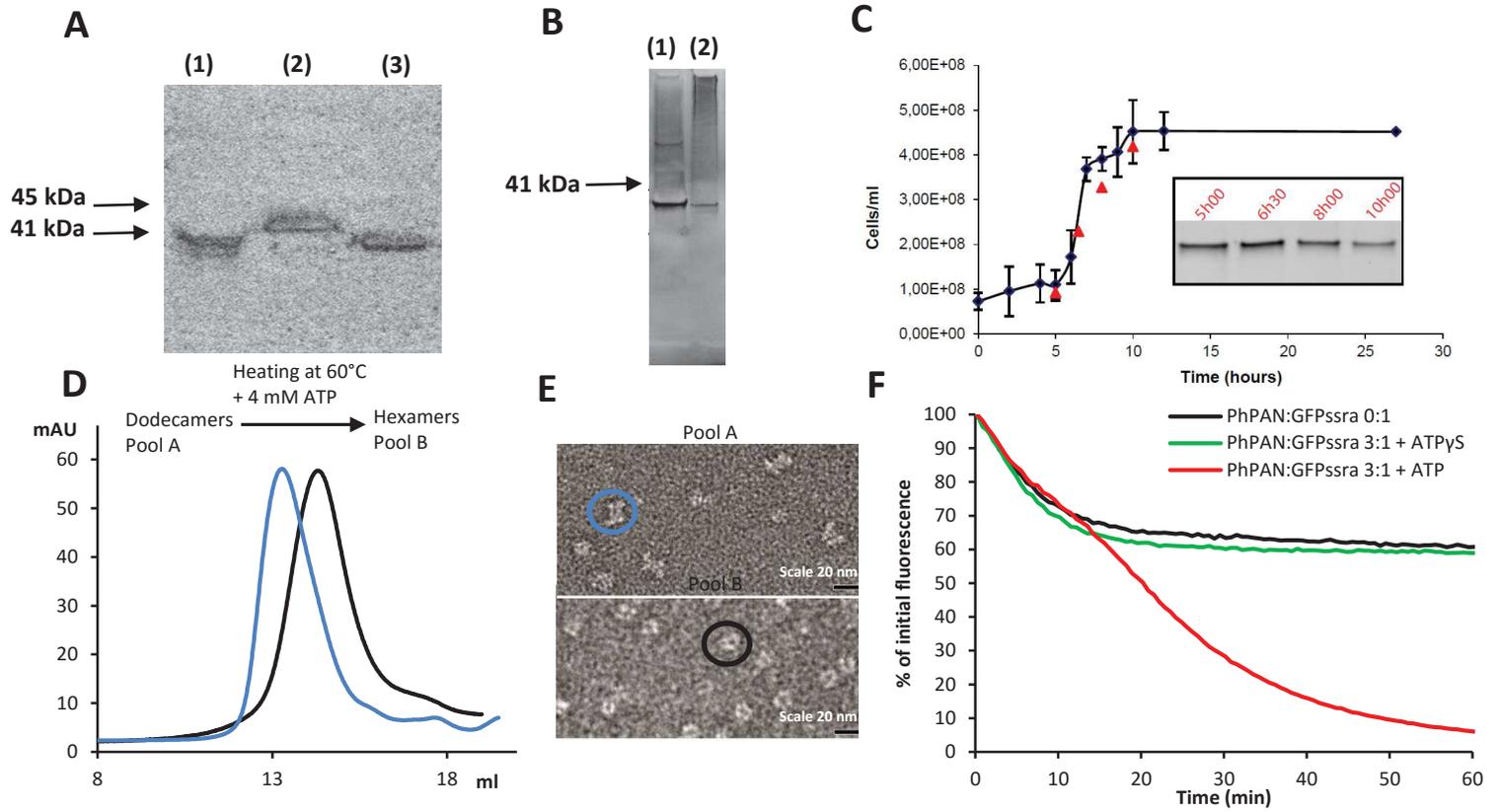


Figure 1. Pyrococcales express a short and active isoform of PAN *in vivo*.

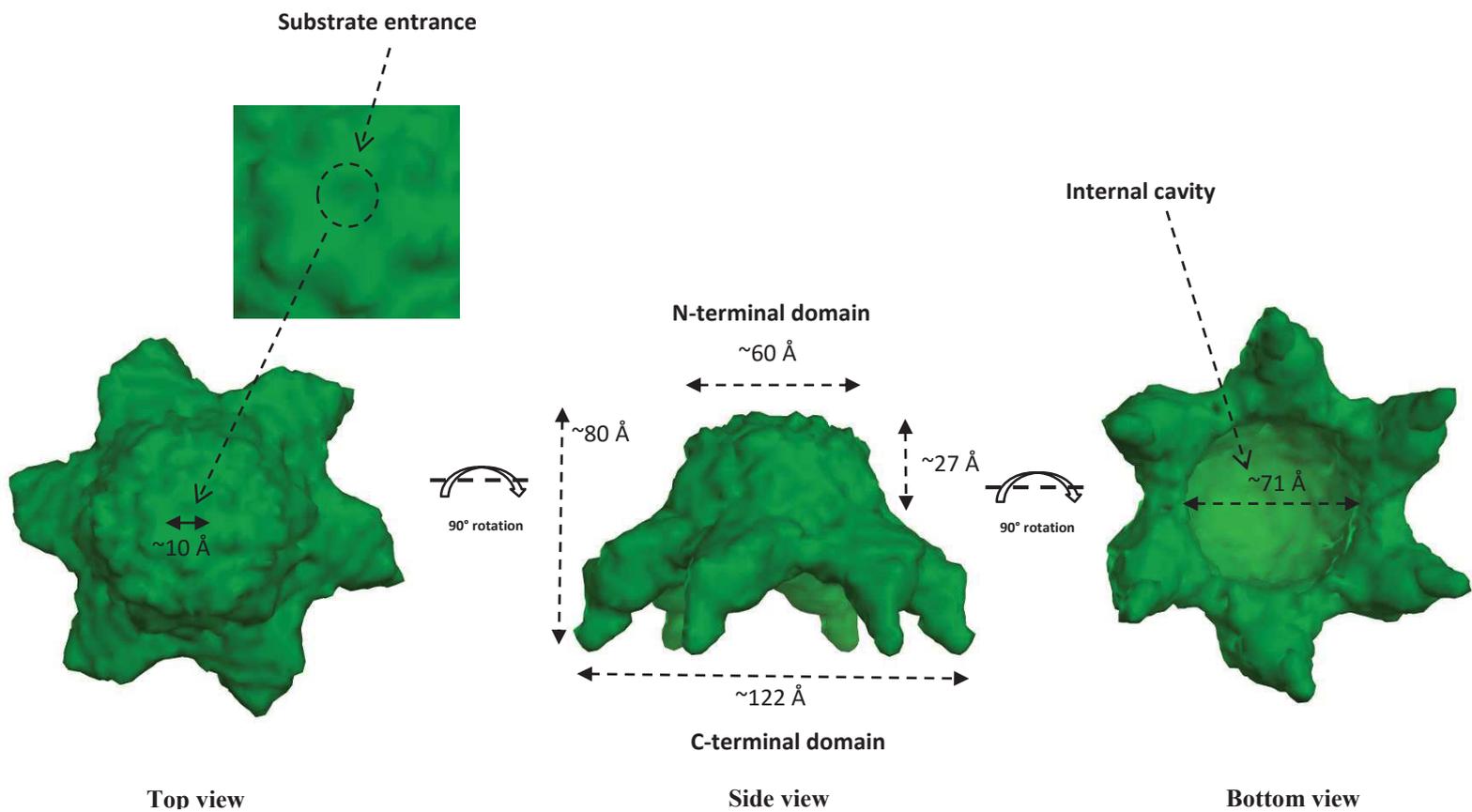


Figure 2. Single particle Cryo-Electron Microscopy 3D reconstruction of ATP-free *PhPAN* hexamer.

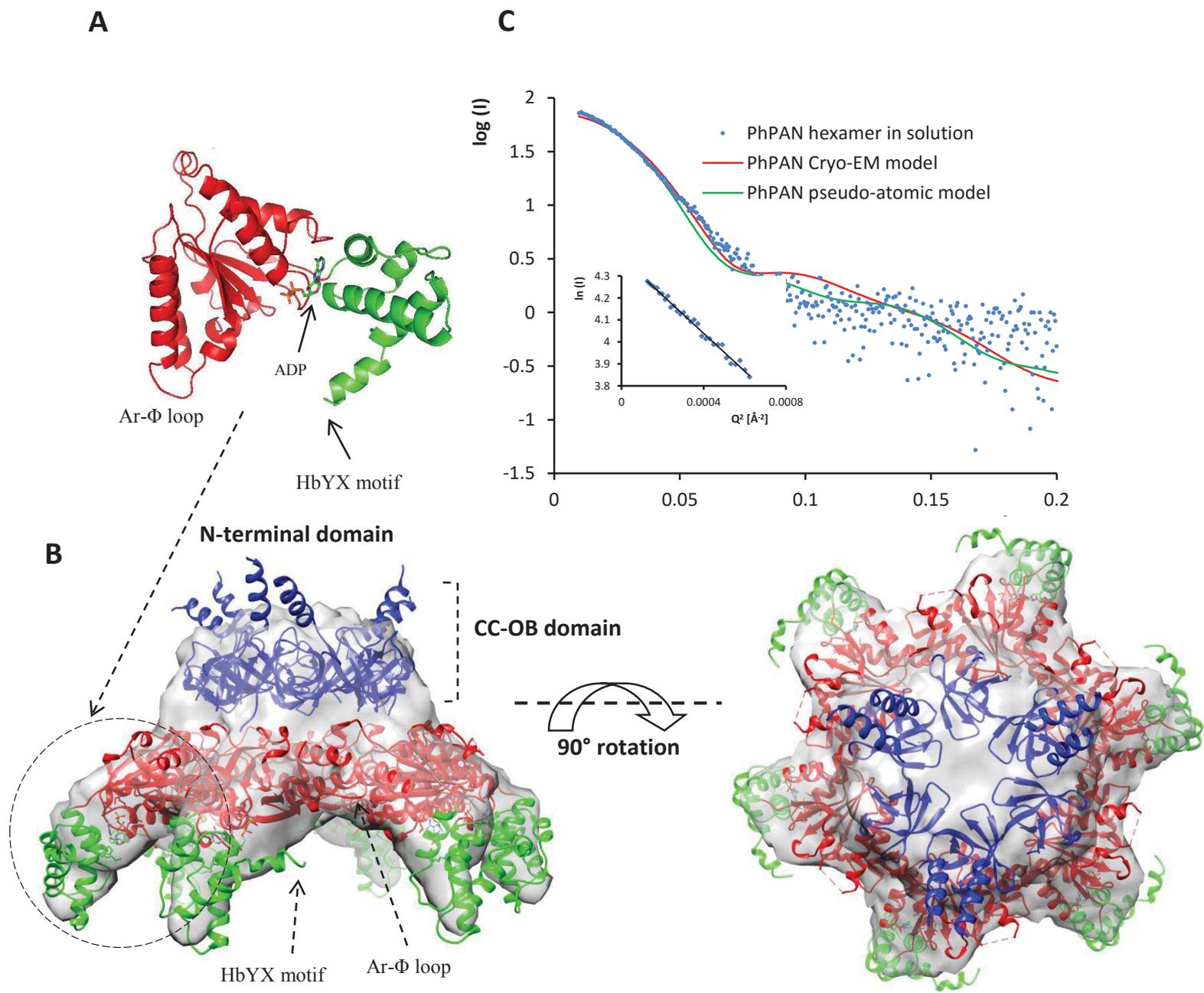


Figure 3. Pseudo-atomic model of the *Ph*PAN complex.

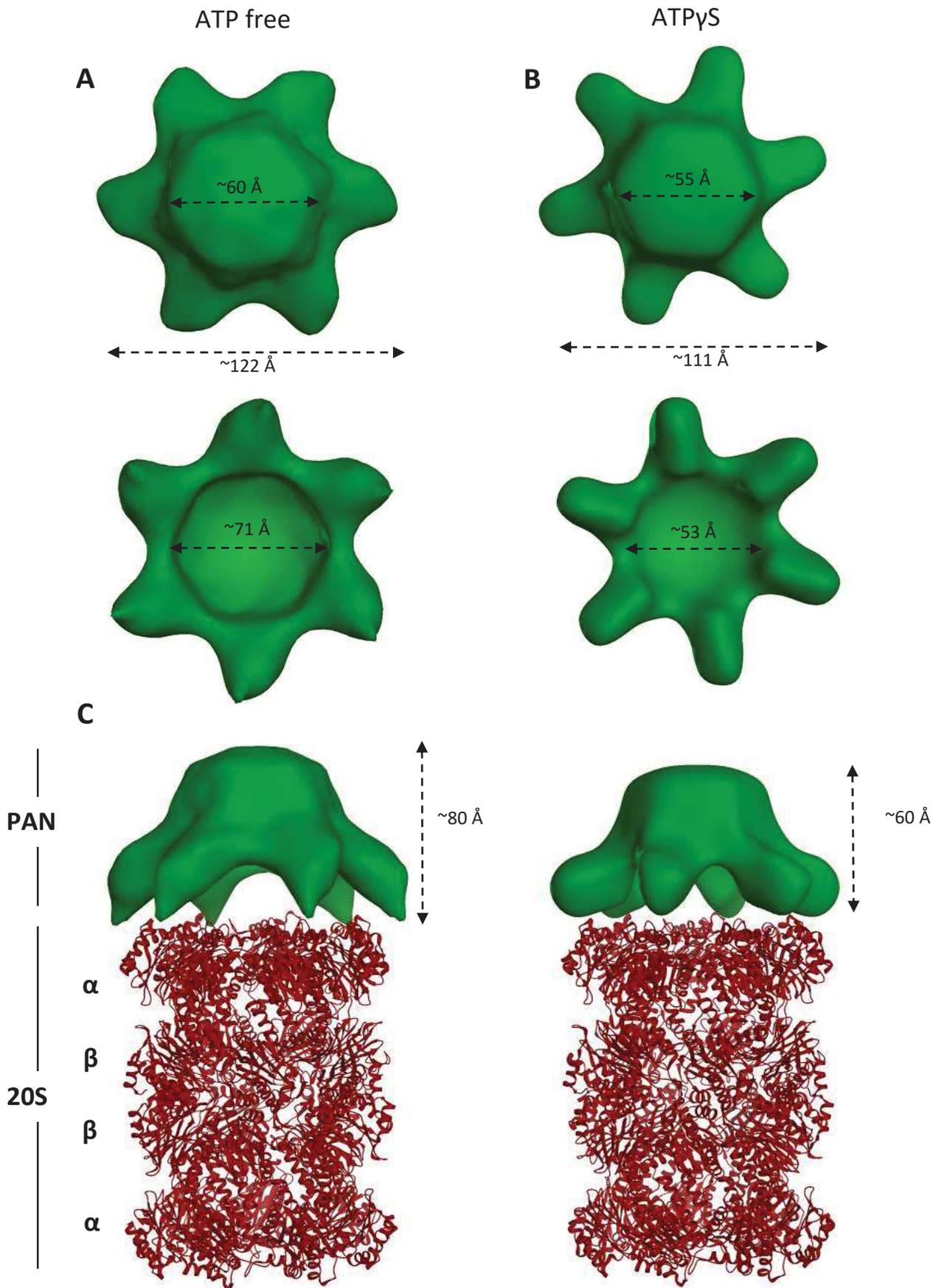
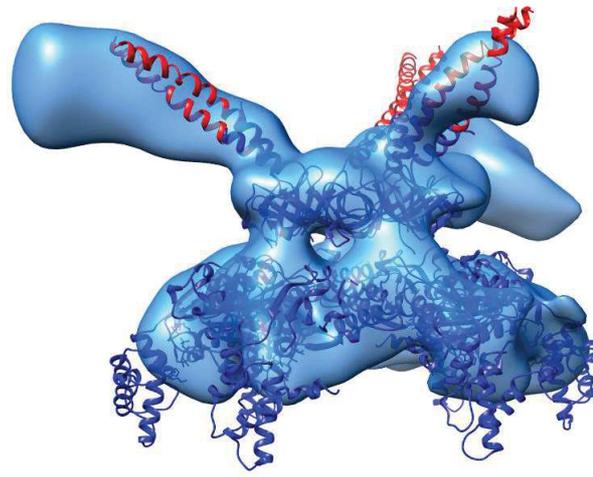
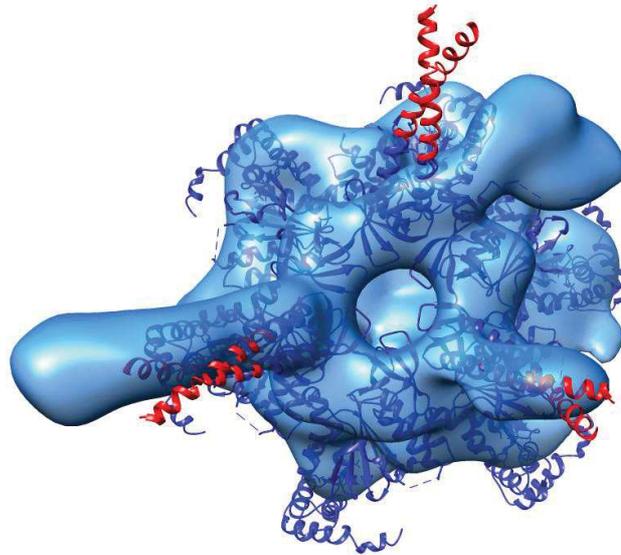


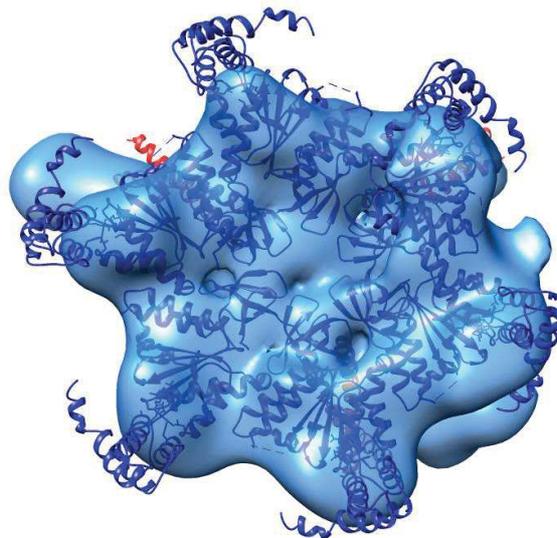
Figure 4. PAN structural transition in response to nucleotide binding.



Side view



Top view



Bottom view

Figure 5. Comparison between the archaeal and eukaryal AAA+ unfoldases.

SUPPLEMENTAL INFORMATION

Small angle scattering reveals different oligomerization states of *Ph*PAN linked to ATP hydrolysis

The observation of dodecameric and hexameric forms of *Ph*PAN (Fig. 1, D & E) prompted us to study their structural and biophysical properties in order to determine potential effects of the oligomerization state on activity, as well as the effect of ATP/ATP γ S and Mg²⁺ on hexamerization. To this end, small angle neutron (SANS) and X-ray (SAXS) scattering studies have been carried out on the proteins in solution. SAXS curves were recorded at room temperature on the dodecamer and the hexamer in solution. Without any prior heating with ATP and Mg²⁺, the experimental R_g obtained was 65 Å (Table S1), showing the presence of the dodecameric form in solution (Fig. S2A, green curve). The molecular weight of the dodecamer was measured by relative calibration to a 4 mg/ml BSA sample using the respective I(0) intensities (Mylonas & Svergun, 2007) and yielded 482 kDa (theoretical MW= 492 kDa). When the scattering curves were recorded on the PAN complex after heating at 60 °C in the presence of 4 mM ATP and 10 mM MgCl₂, the R_g decreased to 51 Å (Table S1), which is close to the theoretical R_g calculated from the Cryo-EM model (46 Å) and the one calculated from the generated atomic model (49 Å) (Fig. 3B), indicative of hexamerization of PAN dodecamers in solution. The hexamerization step was done immediately before the SAXS experiment (see Supplementary Experimental Procedures) due to the high instability of PAN hexamers in solution and in order to get rid of aggregates that distort the final SAXS signal. When the experiment was carried out at high temperature (60°C) by SANS in the absence of both ATP and MgCl₂ or in the presence of only one of them, the dodecamer was always present (data not shown) and no hexamerization was observed. These results validated the switch between the two oligomeric states, induced by heating with ATP hydrolysis and already observed on the Superose 6 column gel filtration profile (Fig. 1D) and by negative stain electron microscopy (Fig. 1E).

In order to better understand the role of nucleotide binding in the hexamerization and/or the conformational changes of the PAN complex, we carried out a SANS study at 40 °C in the presence or absence of ATP γ S, the non-hydrolysable ATP analog. Surprisingly, in the presence of ATP γ S, no hexamerization of the *Ph*PAN complex is observed, and PAN dodecamers are presumably locked in a closed conformation compared to the nucleotide-free dodecamers. Only conformational changes within the dodecamers without complete hexamerization were observed in this case (Fig. S2B). The ATP γ S bound form had an R_g of

63 Å and a D_{\max} of 235 Å, while the nucleotide free dodecamers had an R_g of 65 Å and a D_{\max} of 245 Å (Table S1).

These results highlight the fact that hexamerization and activation of *Ph*PAN requires not only nucleotide binding but also nucleotide hydrolysis in the presence of $MgCl_2$ at high temperature, in agreement with the fluorescence assays (Fig. 1F). PAN dodecamers showed higher stability in solution up to high concentrations while, once hexamerized, they were much less stable and showed a tendency for precipitation above 0.6-0.8 mg/ml. This fact has been confirmed by biochemical and SANS results showing that *Ph*PAN hexamers have a tendency to aggregate in solution when the protein is concentrated above 0.6-0.8 mg/ml (data not shown). The instability of the PAN hexamers, might be due to the high flexibility of the complex once it is hexamerized and is responsible for its biological activity.

X-ray structure of the *Ph*PAN -Cter

All our attempts to generate high-quality diffracting crystals of the complete *Ph*PAN complex failed. We succeeded in producing and purifying the C-terminal subdomain of *Ph*PAN comprising residues 102-365 (named *Ph*PAN-Cter). *Ph*PAN-Cter was purified in a monomeric state in agreement with previous studies showing that PAN oligomerization is driven by the N-terminal domain (Zhang et al, 2009a). The asymmetric unit revealed one single molecule of *Ph*PAN-Cter (Fig. 3A). The X-ray structure was refined using data up to 1.94 Å resolution. *Ph*PAN-Cter displays the typical AAA+ folding with an N-terminal P-loop NTPase $\alpha/\beta/\alpha$ subdomain connected to a smaller α -helical C-terminal domain. Clear positive electron density peaks were found in the Fo-Fc maps just after molecular replacement in the nucleotide binding pocket. This electron density was modeled as an adenosine di-phosphate molecule. These observations suggest that the monomeric C-terminal nucleotidase domain has the ability to hydrolyze the ATP into ADP. The phosphate in position β is coordinated by the Walker A motif (G P P G C G K T; 147-154) residues, Lys153 and Thr154 side chains and by the amide backbone of Gly150. The phosphate in position α is stabilized by the Leu219 amide backbone. The adenine base establishes H-bond with Gly109 amide backbone while the ribose interacts with His289. Furthermore, Asp270 of the Walker B motif binds two water molecules, while Thr154 binds a water molecule. It has been reported that Thr154 binds an Mg^{2+} ion in the presence of ATP (PDB 3M0E) (Chen et al, 2010; Ogura & Wilkinson, 2001) and it is fundamental for ATP hydrolysis. The electron density for residues 215-222 is missing, suggesting that this region is highly flexible. The electron density map of *Ph*PAN-Cter backbone terminates with Thr358, lacking electron density for the residues 359-365.

To determine the structural similarity of *Ph*PAN-Cter with other AAA+ modules, the PDB was submitted to DALI server, revealing a high similarity with *Mj*PAN C-terminal domain PDB 3H4M (rmsd of 1.3 Å for 244 C α) and for PDB 3WHK (rmsd of 1.3 Å for 238 C α). This latter is a chimeric construct that merge the large AAA+ module from PAN (Fig. 3A, red) of *P. furiosus* with the small AAA+ module of the eukaryotic Rpt-5 ATPase (Sato et al, 2014). PAN-Rpt5 contains an ATP molecule in the nucleotide binding site. This feature prompted us to compare the nucleotide binding site of *Ph*PAN-Cter and PAN-Rpt5. The residues involved in the ATP/ADP binding are superimposable. However, the C α of Asn253-Arg254-Pro255 (residue numbers relative to *Ph*PAN-Cter) that are the putative sensor I, are moved away from the nucleotide binding site by 1.2 Å, 1.75 Å and 1.65 Å, respectively, from the ATP bound form (PAN-Rpt5) to the ADP bound form (*Ph*PAN-Cter). Furthermore, compared to the *Mj*PAN C-terminal subdomain, *Ph*PAN-Cter contains a glycine residue in position 348 that determines a 54° deviation of the C-terminal α -helix. The C-terminal α helix of PAN contains the last three residues encompassing the HbYX motif that is needed for the proteasome α gate opening. The different conformations of the C-terminal α -helix of PAN observed here and in the *Mj*PAN subdomain (also in its ADP bound form), suggest an ATP driven mechanism for PAN-20S interaction, in which the ATP binding determines the movement of the C-terminal α -helix of PAN towards the residues 2-11 of the α subunit of the 20S core particle, where the α -gate may therefore be opened.

SAXS characterization of *Ph*PAN-Nter

Any attempts to crystallize *Ph*PAN-Nter (residues 1-101 of the N-ter short *Ph*PAN) were unsuccessful and SAXS measurements were performed instead. The R_g value obtained using the Guinier approximation was about 37 Å. Theoretical scattering curves were generated from the crystal structure of the *Mj*PAN-Nter (79 amino acids, PDB 3H43) by using CRY SOL (Svergun et al, 1995). In particular, a hexameric and two dodecameric models (Nter-Nter and Cter-Cter) were generated from the packing of the symmetry related molecules of the N-ter *Mj*PAN domain in the crystal structure. The models were then fitted and compared with the SAXS experimental data (Fig. S3). The hexameric form displayed the poorest fit ($\chi = 53.93$) while the back-calculated curves from the dodecameric forms, in particular the Cter-Cter arrangement, were in better agreement with the experimental data ($\chi = 12.23$ for the Nter-Nter and 3.84 for the Cter-Cter dodecamer).

The oligomerization state of the N-terminal domain complex was confirmed by the molecular weight measured by a relative calibration to a BSA sample, as described in the Supplemental

Experimental Procedures. This measurement yielded a value of 135 kDa that is in agreement with the theoretical molecular weight of 139 kDa of the dodecameric *Ph*PAN-Nter. These results strongly suggest that *Ph*PAN-Nter forms a dodecameric complex in solution (two stacked hexameric rings), similar to the *Mj*PAN-Nter N-terminal domain, with a high structural homology between the two domains for both species, and that the dimerization of the two hexamers is probably occurring on the C-terminal side. The residual discrepancy between the Cter-Cter model with the SAXS data might be due to the difference in length (101 and 79 amino acids), the amino acid composition (38% identity) between the two N-terminal domains and the conformational flexibility of these two fragments which produces a smearing of side-minima in SAXS curves of globular objects (Lindner, 2002). Unfortunately, this complex was too small to be studied by electron microscopy and crystallization was equally unsuccessful. When the N-terminal part is produced separately from the complex, it shows higher stability and less flexibility compared to the whole protein, suggesting that the occurring ATP hydrolysis on the *Ph*PAN C-terminal side is the major factor promoting the protein instability.

The high structural homology between the two N-terminal domains from *Mj*PAN and *Ph*PAN, confirmed by SAXS in solution, strongly support our choice of using the *Mj*PAN N-terminal domain for the modeling and the generation of the complete pseudo-atomic model of *Ph*PAN as described previously (Fig. 3B).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Pyrococcus horikoshii and *Pyrococcus abyssi* cell cultivation and western blot analysis

P. horikoshii and *P. abyssi* were grown on SME YP medium, pH 7.0, supplemented with PIPES (20 mM) and elemental sulfur (1 mg/ml) (Godfroy et al, 2006). Growth characterization of *P. abyssi* was carried out in duplicate using 40 ml of medium in 100 ml flasks. 4 ml of overnight culture was used as inoculums and the cell density was monitored over 27 h using a Thoma cell counting chamber. According to the growth curve obtained, cultures were sampled at 5, 6.5, 8 and 10 h after inoculation. Inside an anaerobic chamber, the cultures were transferred to anaerobic centrifuge tubes (Beckman) and centrifuged at low speed to settle the sulfur particles (2 min, 200 g at 4 °C). Still in anaerobic conditions, the supernatant was transferred to a new tube and centrifuged for 1 h at 15,000 g and 4 °C. The cell pellets were then washed in 10 ml of cold sea salts in 50 ml Falcon tubes and further centrifuged for 1 h at 15,000 g and 4 °C. Pellets were re-suspended in PBS buffer complemented with antiprotease (Complete free EDTA, Roche) at 1/2 (weight/volume) ratio. The cells were sonicated on ice for 5 min (0.5 sec ON/0.5 sec OFF, 35W) and the cell extracts recovered after centrifugation for 30 min at 15,000 g and 4 °C. The total protein concentration was determined by Bradford assays using BSA as standard. 25 µg of proteins in loading buffer (50 mM Tris pH 6.8, 8 M urea, 2 M thiourea, 75 mM DTT, 3% SDS, 0.05% bromofenol azul) were separated in SDS-PAGE 4-12% (Pierce). The proteins were transferred to a polyvinylidene difluoride PVDF low fluorescence membrane (Amersham) and probed with specific anti-PAN rabbit antibodies. Immunoreactive bands were visualized by fluorescence using the Typhoon 9400 fluorimager, as described in the manufacturer protocol [ECL+ (enhanced chemiluminescence) detection kit; Amersham]. The cultivation was performed overnight at 90 °C in 1 l serum vials containing 500 ml of medium under anaerobic conditions (N₂ gas phase). Cells were harvested by centrifugation and cell pellets were immediately stored at room temperature in water/isopropanol (50%/50% v/v) until utilization. To prepare the protein extracts, cell pellets were re-suspended in lysis buffer [50 mM Tris/HCl (pH 8.3) and 0.1% Triton X-100] by repeated passage through a 0.2 mm diameter needle. Cell debris were removed from the homogenate by centrifugation (30,000 g for 30 min). Protein was recovered from the supernatant and its concentration determined by the Bradford assay using BSA as the calibration standard. SDS loading buffer [50 mM Tris/HCl (pH 6.8), 100 mM DTT (dithiothreitol), 2% SDS, 0.1% Bromophenol Blue and 10% glycerol] was then added to the protein samples. Before electrophoretic separation (12% SDS-PAGE), the samples were heat-denatured at 95 °C for 10 min. The proteins were transferred

to a polyvinylidene difluoride PVDF membrane (Amersham) and probed with specific anti-PAN rabbit antibodies. Immunoreactive bands were visualized by chemiluminescence as described in the manufacturer's protocol [ECL® (enhanced chemiluminescence) detection kit; Amersham].

Recombinant protein cloning, expression and purification

A truncated form of PAN from *P. horikoshii* (*PhPAN*), lacking the first 34 amino-acids (Fig. S1), was synthesized and cloned by *GeneCust Europe* (<http://www.genecust.com>) into pET-41c vector, between the *Nde*I and *Bam*HI restriction sites, to promote the expression of the protein of interest without tag. *PhPAN* was over-expressed in *Escherichia coli* strain BL21 (DE3) Rosetta2 at 37 °C during 4 h. The cells were re-suspended in 50 ml Tris-HCl 20 mM pH 8.0, NaCl 150 mM and Triton X-100 0.1% and supplemented with 12.5 mg lysozyme (Euromedex), 2.5 mg DNase I grade II (Roche), 10 mg RNase (Roche), 50 mg Pefabloc SC (Roche) and 0.5 ml 2 M MgSO₄. Cells were disrupted by sonication with 10 cycles on/off of 30 s each at 4 °C. After heating at 85 °C during 15 min to eliminate most mesophilic proteins from the *E. coli* host, the lysate was clarified by centrifugation at 12,000 rpm for 1 h (JA20 rotor) and the supernatant was loaded onto a DEAE Sepharose CL-6B column followed by a 1 ml Mono Q 5/50 GL column (GE healthcare). Both columns were equilibrated in 150 mM NaCl, 20 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 1 mM DTT. After washing with the same buffer, bound proteins were eluted with a 20 CV linear salt gradient (0.15-0.35 M NaCl for the DEAE Sepharose and 0.15-0.5 M NaCl for the MonoQ 5/50 GL using the equilibration buffer). The fractions of the peak were combined and for further purification, the protein was loaded on a Superose 6 10/300 GL size exclusion column (GE healthcare) equilibrated in the buffer containing 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 1 mM DTT. Protein concentrations were determined using the Bradford assay (Bio-rad) and confirmed by SDS-PAGE. The purified proteins were stored at – 20 °C after the size-exclusion step. Before each use, PAN molecules were hexamerized and activated by heating at 60 °C for 30 min with 4 mM ATP and 10 mM MgCl₂ and the hexamers were re-injected on the Superose 6 10/300 GL size exclusion column as a last purification step and to monitor the complete hexamerization of the molecules.

The N-terminal (residues 1-101, *PhPAN*-Nter) and the C-terminal (residues 102-365, *PhPAN*-Cter) domains DNA of *PhPAN* were individually synthesized and cloned by *GeneCust Europe* (<http://www.genecust.com>) into pET-15b vector using *Nco*I/*Xho*I cloning sites. *PhPAN*-Nter and *PhPAN*-Cter were over-expressed in *E. coli* strain BL21 (DE3)

Rosetta2 at 37 °C during 4 h. The lysates were prepared following the same procedures as described above. The NaCl concentration of the clarified supernatants was adjusted to 70 mM. For *PhPAN-Cter*, the sample was loaded onto a pre-packed DEAE-sepharose anion exchange column, previously equilibrated with 2 CV of Tris-HCl 20 mM pH 8.0, 70 mM NaCl, 10 mM MgCl₂, 1 mM DTT. *PhPAN-Cter* was eluted with a 20 CV linear salt gradient between 70-400 mM NaCl. In a similar manner for the *PhPAN-Nter*, the supernatant was loaded onto a Resource S cation exchange chromatography column (GE healthcare) equilibrated with the same buffer and eluted with the same gradient. The fractions containing the proteins were checked by SDS-PAGE 15%, confirmed by N-terminal sequencing and pooled. The proteins were then concentrated with Amicon Ultra 3 kDa cut-off to a final volume of 500 µl and injected into a Superose 12 size exclusion column (GE Healthcare) for further purification. The two sub-domains were pooled and concentrated for crystallization (*PhPAN-Cter*) or SAXS (N-terminal domain) purposes.

The DNA sequence encoding the GFPssrA protein (containing a A206K mutation to minimize dimerization) was generated from synthetic DNA fragments and cloned into the over-expression plasmid pET-30a by *GeneCust Europe* (<http://www.genecust.com>). The resulting construct was used to transform *E. coli* strain BL21 (DE3) for the recombinant expression of the protein at 37 °C. GFPssrA was purified by ethanol extraction followed by Phenyl Sepharose 6 Fast Flow hydrophobic interaction chromatography as described in (Samarkina et al, 2009).

***PhPAN-Cter* crystallization, data collection, structure determination and refinement**

PhPAN-Cter (residues 102-365) was purified as described above and concentrated up to 7.6 mg/ml using an Amicon Ultra 3 kDa cut/off filter. Crystals were grown at 20 °C in NaCH₃COO 0.1 M pH 5.5, NH₄PO₄ 0.6 M using the hanging drop vapor diffusion method. *PhPAN-Cter* was mixed with an equal volume of the reservoir solution and diamond-like crystals grew up to average 200 µm size in several months. Crystals were cryo-protected by serial transfer into drops equilibrated against increasing concentrations of PEG400 (5-30%) in the reservoir buffer. The crystals were flash cooled in liquid nitrogen prior to data collection at 100K. A data set was collected (360°) at the ESRF beamline ID29, the diffraction extending up to 1.8 Å. Data were indexed and integrated with XDS (Kabsch, 2010) and scaled with SCALA. Intensities were converted to amplitude by TRUNCATE in the CCP4 suite (Winn et al, 2011). Based on a correlation coefficient between random half datasets (CC_{1/2}) of 59 (Karplus & Diederichs, 2015) (Karplus & Diederichs, 2012) , a R_{pim} value (Weiss, 2001)

of 57 % and a $I/\sigma(I)$ of 1.1 in the highest resolution shell as provided by SCALA, the final data resolution was set at 1.94 Å with C2 space group (Table S2). The *Ph*PAN-Cter structure was determined by molecular replacement with PHASER (McCoy, 2007), using the *Mj*PAN C-terminal domain as searching model (PDB ID 3H4M), depleted of waters and ligands, and all the experimental crystallographic reflections. A clear solution was determined, with a Translation Function Z-score of 16.9. The asymmetric unit contained a single monomer. After molecular replacement a round of rigid-body refinement was performed. The restrained refinement was carried out using data to 1.94 Å and the software package PHENIX-dev-1839 version (Afonine et al, 2012) resulting in a final $R_{\text{work}}= 22\%$ and $R_{\text{free}}= 24$ (see Table S2 for data collection and refinement statistics). During the refinement procedure TLS and individual ADP refinement were applied.

SAXS/SANS data collection and analysis

The small angle X-ray scattering (SAXS) measurements were performed at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The data were collected using a standard set up (automated sample mounting to a capillary by a robot and measurement of 10 frames with 1 s exposure time were acquired using the flow-through mode) (Pernot et al, 2013). The samples were measured at 3.3 mg/ml (N-terminal domain), 1.6 mg/ml (*Ph*PAN dodecamer) and 0.3 mg/ml (*Ph*PAN hexamer). The data processing and analysis was done by using standard ESRF beamline software (BSxCuBE) (Pernot et al, 2013) and PRIMUS (Konarev et al, 2003). The radii of gyration (R_g), forward scattered intensities ($I(0)$) and experimental scattering curves were compared to the back-calculated scattering curves from the Cryo-EM model using the programs EM2DAM (Petoukhov et al, 2012) and CRY SOL (Svergun et al, 1995) by means of a least χ -fit and by visual inspection of the curves. The molecular weights of the complexes (N-terminal domain and *Ph*PAN dodecamer) were measured by relative calibration to a 4 mg/ml BSA sample (Mylonas & Svergun, 2007).

The small angle neutron scattering (SANS) experiment was performed on the D22 instrument at the Institut Laue-Langevin (ILL, Grenoble, France). The instrument was configured with a sample-detector distance of 4 m, collimation length of 4 m and neutron wavelength (λ) of 6 Å. The samples were measured in Hellma 100Qs quartz cells of 1.00 mm path-length at 40 °C. Two samples were measured: the dodecameric (1 mg/ml) *Ph*PAN at 40 °C in the presence or absence of ATP γ S. The buffer, the empty cell, the empty beam and the boron were measured to perform data reduction (correction for detector efficiency, electronic noise, and sample

holder scattering). The data reduction and radial integration over a two-dimensional image were performed using standard ILL software (Dewhurst, 2002).

For both SAXS and SANS measurement the scattering curve of the buffer was subtracted from the respective sample scattering curve with the software PRIMUS (Konarev et al, 2003). The R_g and $I(0)$ values were extracted using the Guinier approximation (Guinier, 1939). The pair-wise distance distribution functions $P(r)$ were generated with GNOM (Svergun, 1992). The program DAMMIF (Franke & Svergun, 2009) was used to generate low resolution envelopes of *Ph*PAN from the SAXS data recorded on the PAN hexamers. We used several variations of the input parameters in the default mode to model the data. Data up to $Q_{\max}=0.10-0.15 \text{ \AA}^{-1}$ were included. The DAMMIF input file was generated using the program GNOM (Svergun, 1992), imposing the restraints $P(r=0)=0$ and $P(r=D_{\max})=0$ for the pair distance distribution function $P(r)$. The values of D_{\max} were varied to obtain the best fit of the data and a smooth $P(r)$ curve. As a quality control, the radii of gyration determined from the pair distribution analyses were compared with the one determined by the Guinier analysis.

Negative-stain electron microscopy and single particle 3D reconstruction

After the size exclusion column step of the purification, 4 μl of the dodecameric and the hexameric fractions of *Ph*PAN sample ($\sim 0.1 \text{ mg/ml}$) were deposited onto carbon-coated 400-mesh copper grids. The samples were stained using 2% sodium silicotungstate pH7.5 and air-dried. Images were taken under low-dose conditions in a T12 FEI electron microscope working at 120 kV and with a nominal magnification of 40,000 using an Orius SC1000 CCD camera.

For Cryo-EM, 4 μl of sample were loaded onto a Quantifoil R2/1 holey grid (Quantifoil Micro Tools GmbH, Germany), blotted for 2S at 100% humidity and vitrified using a Mark IV (FEI). The frozen grid was transferred to a Polara electron microscope working at 300 kV (ATP-free *Ph*PAN) or on a F20 (FEI) electron microscope working at 200 kV (ATP γ S-bound *Ph*PAN). The images were taken under low dose conditions (less than $20 \text{ e}^-/\text{\AA}^2$) and with a nominal magnification of 78,000 on a GATAN ultrascan CCD camera (ATP-free *Ph*PAN) or with a nominal magnification of 50,000 on an Eagle FEI CCD camera (ATP γ S-bound *Ph*PAN). The corresponding pixel sizes at the sample level are 1.5 \AA (ATP-free *Ph*PAN) and 2.15 \AA (ATP γ S-bound *Ph*PAN).

A total of 30,000 particles (ATP-free *Ph*PAN) and 20,000 (ATP γ S-bound *Ph*PAN) were selected using a semi-automatic particle selection procedure with the EMAN boxer routine (Ludtke et al, 1999) and extracted into respectively 128×128 and 90×90 pixel boxes. The

images were CTF-corrected with CTFFIND3 (Mindell & Grigorieff, 2003) and Bsoft (Heymann et al, 2008) and low-path-filtered to 25 and 8 Å.

A SPIDER (Frank et al, 1996) based projection matching analysis was performed. The initial model was generated by back-projecting a side view of the particle and applying a C6 symmetry. A total number of 372 equally spaced re-projections and 50 cycles were used to obtain the final 3D reconstruction (the orientation parameters of the raw images and 3D reconstruction were stable). Each cycle consisted of model re-projections, alignment using the 25 Å low-path-filtered stack file, and 3D reconstructions using the 8 Å low-path-filtered stack file. Only the best 15,000 images or 10,000 images (cross correlation criteria) were used for the 3D reconstruction. The resolution of the structure were found to be between 13 Å and 25 Å (0.5 Fourier-shell correlation criteria,) for ATP-free *Ph*PAN and ATP γ S-bound *Ph*PAN respectively (Fig. S6).

Enzyme assays

GFP_{ssrA} was used as a *Ph*PAN substrate. GFP_{ssrA} is a variant of the green fluorescent protein, with its C-terminus fused to the 11-residue peptide *ssrA* (AANDENYALAA) (Weber-Ban et al, 1999). In eubacteria, this specific hydrophobic C-terminal extension sequence is incorporated into nascent polypeptide chains when the ribosome is stalled by a lack of aminoacyl transfer RNA, so this 11-mer targets proteins for degradation. The protein degradation pathway follows a similar trajectory for all ATP-dependent proteases and the substrate specificity is usually determined through placement of degradation tags or degrons (Benaroudj & Goldberg, 2000; Nyquist & Martin, 2014). The selection of GFP as substrate was based on the observation that GFP is extremely heat stable, with a melting temperature of around 83 °C. In addition, the unfolding of GFP is accompanied by disappearance of fluorescence, which can be quantitatively monitored in real time over the 60 min time course of the assay (Zhang et al, 2009b).

The reaction buffer that was used contained 20 mM Tris-HCl (pH 8), 100 mM NaCl, 10 mM MgCl₂ and 4 mM ATP. The unfolding curve of GFP_{ssrA} was measured dynamically by Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek) at $\lambda_{\text{ex}} = 400$ nm and $\lambda_{\text{em}} = 509$ nm, with a temperature control at 65 °C. The GFP_{ssrA} was diluted with the reaction buffer to a final concentration of 1 μ M, in the presence and absence of PAN. As PAN is only active as an hexamer, the protein was first heated for 30 min at 60 °C in the presence of ATP and MgCl₂, purified by gel filtration and then used for the activity assay. To compare the

unfolding ability of PAN with ATP or ATP γ S, the total fluorescence decrease of GFPssrA was measured after incubation at 65 °C for 60 min.

SUPPLEMENTAL FIGURES LEGENDS

Figure S1. Protein sequence alignment of PAN from *P. horikoshii*, *P. abyssi*, *M. jannaschii* and *H. salinarium*. A second methionine initiation site (underlined in black) is present in all PAN sequences showing the expression of a full-length and a short form of the protein. The two isoforms of *Ph*PAN are indicated with arrows.

Figure S2. Small angle X-ray and neutron scattering study of the *Ph*PAN in the dodecameric and hexameric states. (A) Experimental SAXS curves of *Ph*PAN in solution collected at room temperature. Top: Scattering curve collected without any prior heating with ATP and MgCl₂ (green points) showing the presence of the dodecameric form of PAN in solution with an R_g of 65 Å; curve recorded on the *Ph*PAN complex after heating at 60°C with 4 mM ATP and 10 mM MgCl₂ (red points), showing the hexameric form of PAN (R_g = 51 Å). The back-transformed curves from the pair distance distribution functions P(r) (using the program GNOM) are displayed as black curves. Inset: Guinier plot of the dodecameric and the hexameric *Ph*PAN. Bottom: P(r) of *Ph*PAN in its dodecameric (green points) and hexameric (red points) forms obtained from the SAXS data with GNOM. (B) Experimental SANS curves of *Ph*PAN in solution with and without ATPγS. Top: Scattering curves collected at 40 °C with (red points) and without (green points) ATPγS and MgCl₂. Both curves display small differences over the whole angular range studied, indicating minor conformational changes in the dodecameric particles induced by the ATPγS in solution; the back-transformed curves from the pair distance distribution functions P(r) are displayed as black curves. Inset: Guinier plot of *Ph*PAN dodecamers in the presence or absence of ATPγS. Bottom: P(r) of *Ph*PAN in its dodecameric form with (red points) and without (green points) ATPγS obtained from the SANS data using the program GNOM.

Figure S3. SAXS study of the *Ph*PAN N-terminal domain in solution

(A) Experimental SAXS curve from the *Ph*PAN N-terminal domain (blue points) and the back-calculated scattering curves from the crystal structure of the *Mj*PAN-Nter (PDB 3H43) in its dodecameric Cter-Cter (green curve), dodecameric Nter-Nter (orange curve) and hexameric (red curve) forms. The experimental SAXS curve superposes well with the theoretical curve from the dodecameric Cter-Cter domain of *Mj*PAN. Inset: Guinier plot of *Ph*PAN-Nter domain and the structures of the *Mj*PAN N-ter domain on its hexameric (red), dodecameric Nter-Nter (orange) and dodecameric Cter-Cter (green) forms adapted from PDB

3H43. **(B)** *Ph*PAN N-terminal domain with *Mj*PAN N-terminal domain shares 38.7% sequence identity.

Figure S4. Small angle X-ray scattering (SAXS) study of *Ph*PAN in solution. Experimental SAXS curve of the *Ph*PAN hexamer in solution (blue points) and the back-calculated SAXS curves (CRY SOL) from the *Ph*PAN Cryo-EM electron density maps (EM2DAM) in the presence of ATP γ S (green curve) ($\chi = 4.68$) and the modeled *Ph*PAN-HslU structure (red curve) ($\chi = 2.32$) as described in (Zhang et al, 2009a). The experimental SAXS curve show poor agreement with both back-calculated curves indicating a difference in the structure and a large conformational change induced by ATP γ S binding. Inset: Guinier plot of the *Ph*PAN hexamer. The Cryo-EM model of *Ph*PAN in the presence of ATP γ S (green) and the modeled *Ph*PAN-HslU structure (red) are also shown.

Figure S5. *Ab initio* SAXS models of *Ph*PAN. Gallery of five *ab initio* envelopes of *Ph*PAN calculated from SAXS data using DAMMIF (Franke & Svergun, 2009) (green) with P6 symmetry imposed. Each model is represented as spheres and viewed from the side and the bottom. The SAXS models are compared with the Cryo-EM (10-15 Å resolution) (gray) and the pseudo-atomic model (blue). All three show good agreement regarding the overall structural features.

Figure S6. Fourier Shell Correlation curves.

Fourier Shell Correlation for the two EM structures which give a resolution of 13 Å for the ATP free structure (0.5 threshold) and 25 Å for the ATP γ S *Ph*PAN Cryo-EM structure.

Table S1. SAXS/SANS parameters of the dodecameric and the hexameric *Ph*PAN and its N-terminal domain. **(A)** The radii of gyration (R_g), extracted from the Guinier ranges ($Q_{\min}R_g \dots Q_{\max}R_g$), are shown in comparison to the values extracted by indirect Fourier transform using GNOM (Svergun, 1992). The two GNOM R_g values represent the direct/indirect-space values, respectively. **(B)** Radii of gyration R_g extracted from the program CRY SOL (Svergun et al, 1995) after fitting the back-calculated scattering curves from the Cryo-EM model and the high resolution crystal structures into the experimental SAXS data. The χ values of each fit are also presented.

Table S2. Data collection and refinement statistics for *PhPAN-Cter*.

Values in parenthesis are for the highest resolution shell.

$$^a R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_j |I_{\text{hkl},j} - \langle I_{\text{hkl}} \rangle|}{\sum_{\text{hkl}} \sum_j I_{\text{hkl},j}}$$
 where $I_{\text{hkl},j}$ is the j^{th} intensity measurement of reflection hkl

and $\langle I \rangle$ is the average intensity from multiple observations.

$$^b R_{\text{pim}} = \frac{\sum_{\text{hkl}} \sqrt{\frac{1}{n-1}} \sum_{j=1}^n |I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle|}{\sum_{\text{hkl}} \sum_i I_{\text{hkl},j}}$$
 with $I_{\text{hkl},j}$ is the j^{th} intensity measurements of reflection hkl

and $\langle I \rangle$ is the average intensity from multiple observations. n represents the multiplicity of the measurements.

^c $CC_{1/2}$ = Correlation coefficient between random half datasets.

$$^d R_{\text{work}} = \frac{\sum_{\text{hkl}} ||F_o| - |F_c||}{\sum_{\text{hkl}} |F_o|}$$
 for all data except 5% which were used for Rfree calculation.

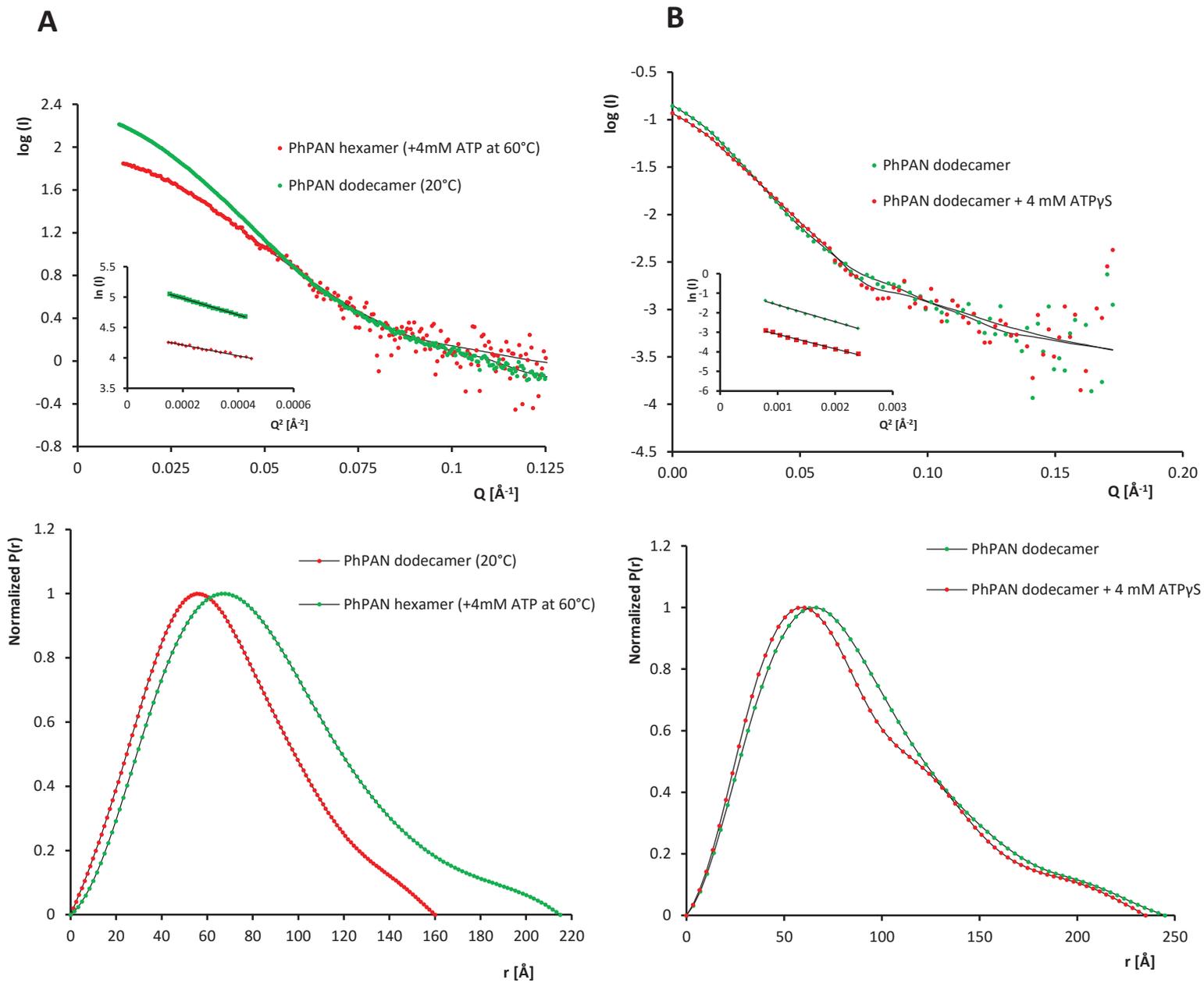
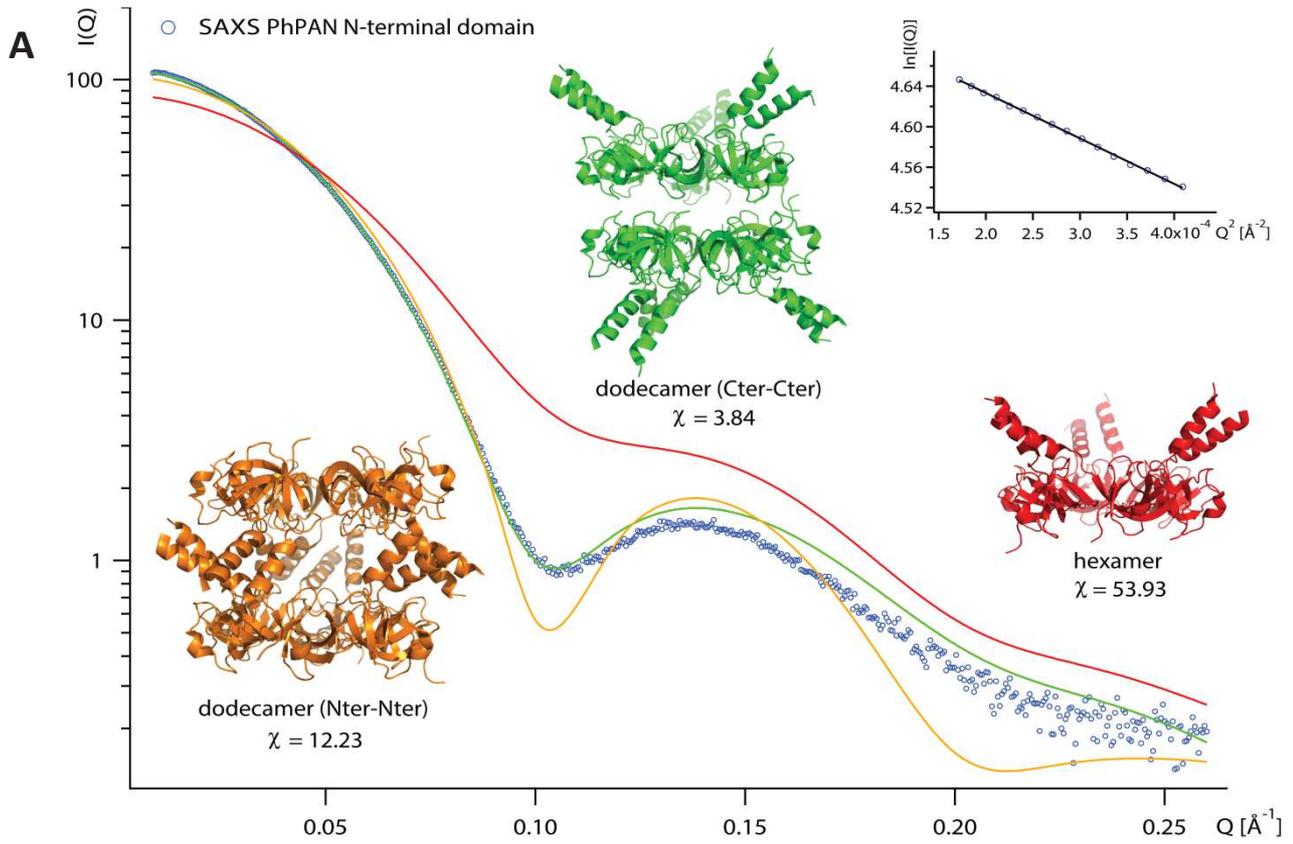


Figure S2. Small angle X-ray and neutron scattering study of the *Ph*PAN in the dodecameric and hexameric states.



B

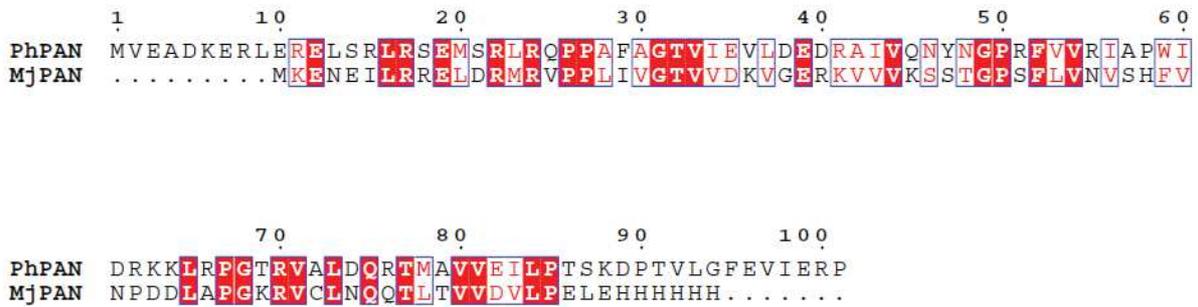


Figure S3. SAXS study of the *Ph*PAN N-terminal domain in solution.

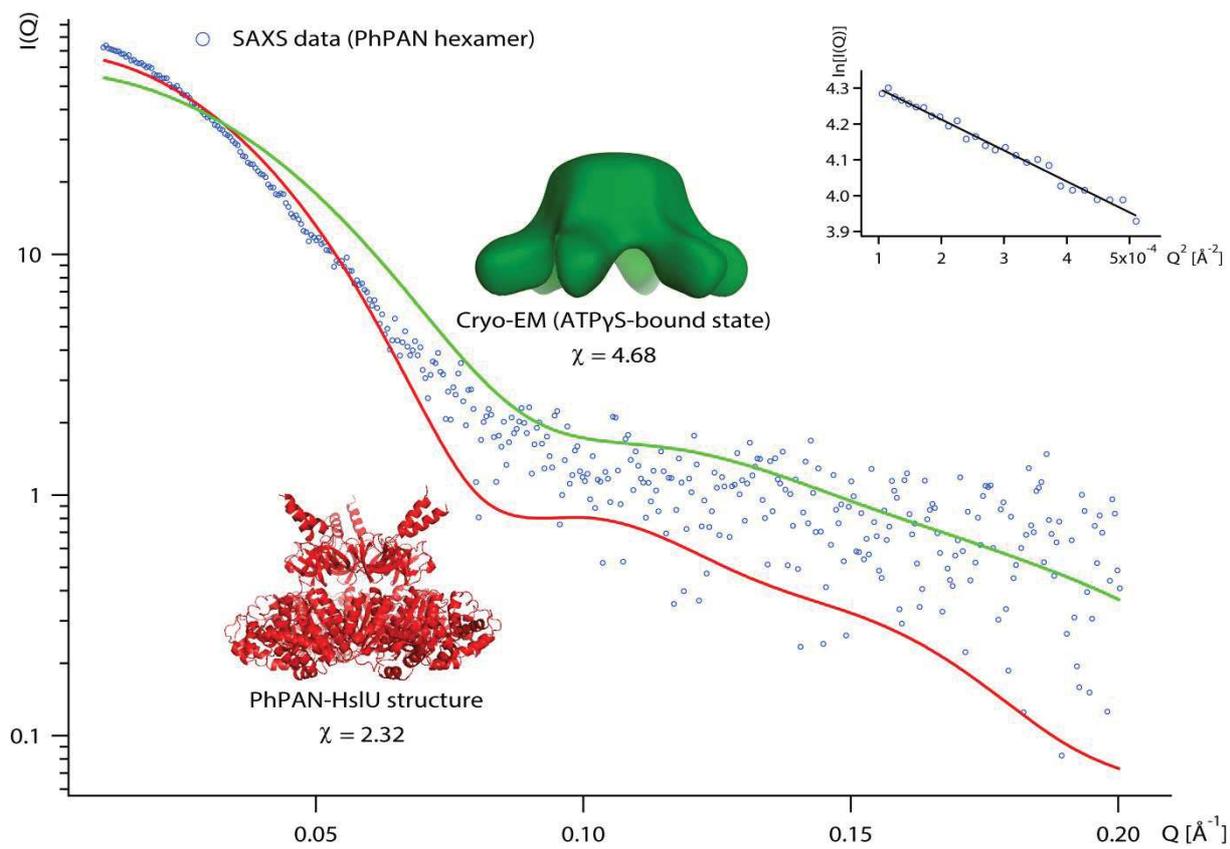


Figure S4. Small angle X-ray scattering (SAXS) study of *Ph*PAN in solution.

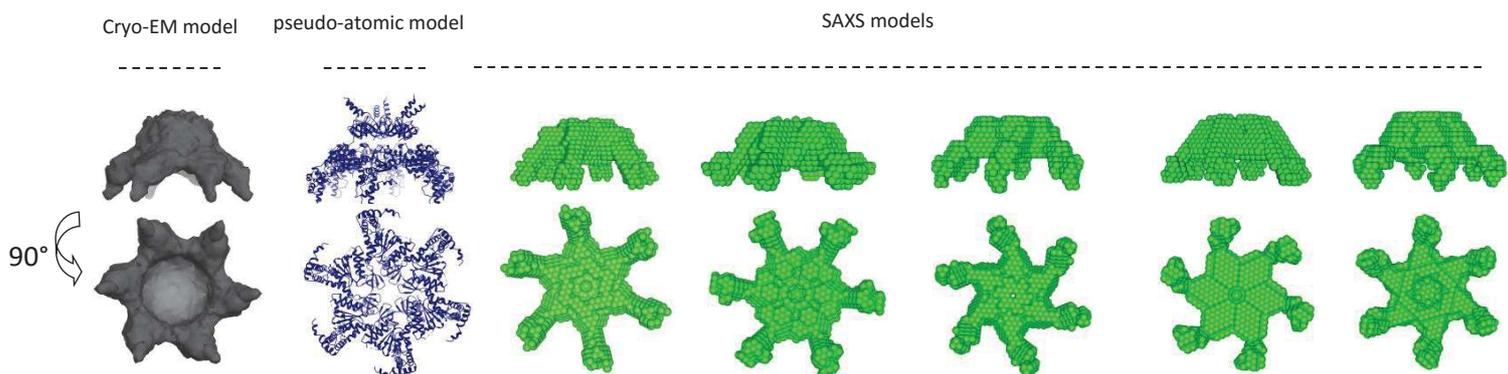


Figure S5. *Ab initio* SAXS models of *PhPAN*.

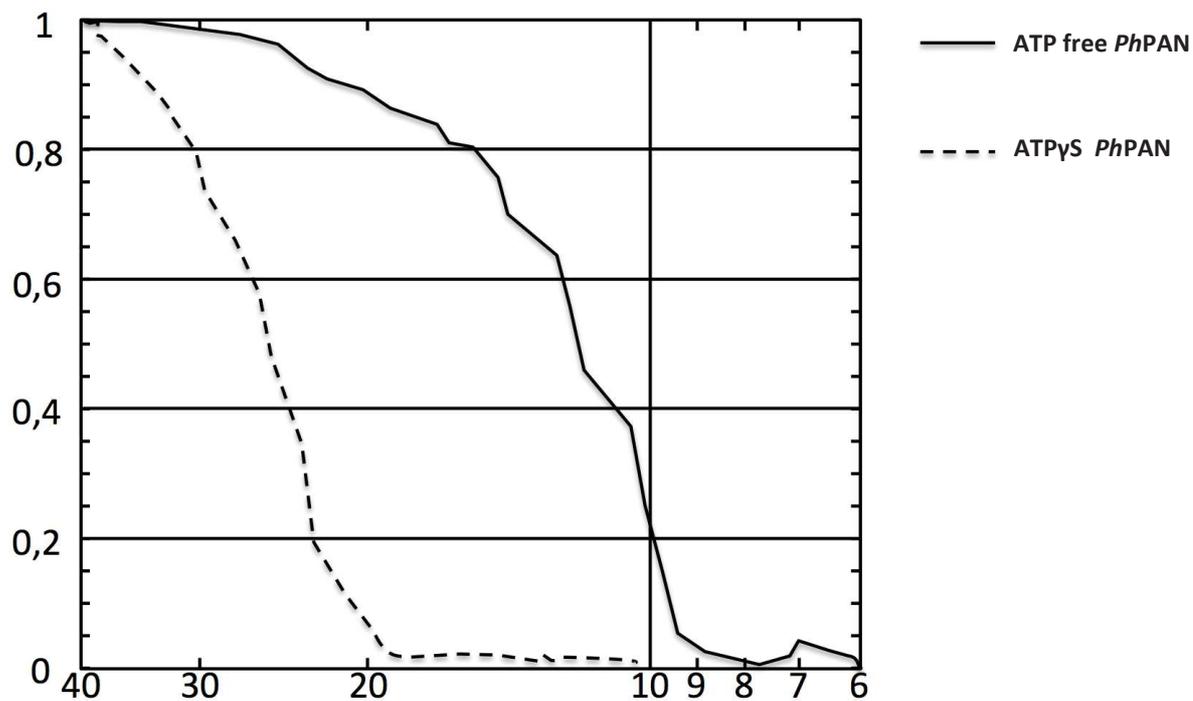


Figure S6. Fourier Shell Correlation curves for the two Cryo-EM structures.

A

| Sample (experimental data) | I (0) (Guinier) | R _g (Guinier) (Å) | Q _{min} R _g - Q _{max} R _g | R _g (GNOM) (Å) | D _{max} (GNOM) (Å) |
|---|--------------------|---------------------------------|--|------------------------------|--------------------------------|
| <i>Ph</i> PAN (SAXS) (dodecamer) | 194.1 ± 0.5 | 65 ± 1.7 | 0.75-1.26 | 65.8/65.6 | 215 ± 5 |
| <i>Ph</i> PAN (SAXS) (hexamer) | 80.3 ± 0.6 | 51.3 ± 0.1 | 0.57-1.18 | 52.1/52.0 | 160 ± 5 |
| <i>Ph</i> PAN (SANS) (dodecamer) | 0.2 ± 0.0 | 64.8 ± 0.8 | 0.85-1.55 | 70.7/70.3 | 245 ± 5 |
| <i>Ph</i> PAN (SANS) (dodecamer+ATPγS) | 0.1 ± 0.0 | 63.1 ± 0.9 | 0.83-1.51 | 68.5 /68.0 | 235 ± 5 |
| <i>Ph</i> PAN (SAXS) (N-ter domain) | 112.4 ± 0.1 | 36.7 ± 0.0 | 0.44-1.26 | 36.1/36.2 | 110 ± 5 |
| BSA (SAXS) Used for calibration | 65.5 ± 0.1 | 30.3 ± 0.0 | 0.68-0.91 | 30.9/30.9 | 92 ± 5 |

B

| Theoretical scattering curves | R _g (CRY SOL) (Å) | fitted in: (SAXS data) | χ |
|---|---------------------------------|----------------------------|-------|
| <i>Ph</i> PAN (Cryo-EM model) | 45.3 | <i>Ph</i> PAN hexamer | 1.67 |
| <i>Ph</i> PAN (Cryo-EM model + ATPγS) | 36.8 | <i>Ph</i> PAN hexamer | 2.46 |
| <i>Ph</i> PAN (pseudo-atomic model) | 49.1 | <i>Ph</i> PAN hexamer | 1.36 |
| <i>Ph</i> PAN-HslU (modeled structure) | 43.1 | <i>Ph</i> PAN hexamer | 2.32 |
| N-ter <i>Mj</i> PAN (hexamer) | 30.1 | <i>Ph</i> PAN N-ter domain | 53.93 |
| N-ter <i>Mj</i> PAN (dodecamer, Nter-Nter) | 33.1 | <i>Ph</i> PAN N-ter domain | 12.23 |
| N-ter <i>Mj</i> PAN (dodecamer, Cter-Cter) | 35.5 | <i>Ph</i> PAN N-ter domain | 3.84 |

Table S1. SAXS/SANS parameters of the dodecameric and the hexameric *Ph*PAN and its N-terminal domain.

| | |
|-------------------------------------|--|
| Structure | <i>Ph</i> PAN-Cter |
| Beam line | ID29 (ESRF) |
| Space Group | C121 |
| Unit Cell Constants (Å) | a= 114.34 b= 46.49 c= 56.43 $\alpha,\gamma=90^\circ$ $\beta=115.8^\circ$ |
| Resolution (Å) | 47.95-1.94 (2.05- 1.94) |
| Wavelength (Å) | 0.95372 |
| R _{merge} ^a (%) | 7.2 (124.8) |
| R _{pim} ^b (%) | 3.1 (57.6) |
| CC ^{1/2} ^c | 99.8 (59.4) |
| I/ σ (I) | 9.7 (1.1) |
| Completeness (%) | 97.7 (93.1) |
| Multiplicity | 6.8 (6.4) |
| Unique Reflections | 19463 (2687) |
| Refinement | |
| R _{work} ^d (%) | 22.3 |
| R _{free} ^d (%) | 23.9 |
| Number of atoms: | 2010 |
| Protein | 1965 |
| Water | 18 |
| heteroatoms | 27 |
| Ramachandran plot: | |
| Most favored region | 226 (92%) |
| Allowed region | 17 (6.9%) |
| Outliers | 2 (0.8 %) |

Table S2. Data collection and refinement statistics for *Ph*PAN-Cter.

5.1.3 Discussion et perspectives

Ce travail a apporté plusieurs nouvelles perspectives sur le complexe d'activation du protéasome chez les archées (PAN) qui pourraient également porter des informations importantes sur son homologue eucaryote plus sophistiqué (le complexe 19S), en particulier la base du complexe comportant les sous-unités formant l'hexamère ATPase (Fig. 7, chapitre 1.4). Nous avons tout d'abord montré qu'une forme courte du complexe PAN dépourvue de 34 et de 31 acides aminés de son côté N-terminal est présente *in vivo* chez *P. horikoshii* et *P. abyssi* respectivement. Cette forme courte montre une activité de dépliement des substrats protéiques globulaires (la GFPssrA) et une solubilité en solution absente chez la forme longue du complexe. Ces résultats indiquent que l'hétérogénéité de l'extrémité N-terminale représente probablement une stratégie de régulation chez les procaryotes pour les ATPases impliquées dans la protéolyse intracellulaire. Malheureusement, tous nos essais d'exprimer et de purifier la forme longue de *PhPAN* ont échoué à cause de son insolubilité en solution une fois que le complexe est exprimée dans *E. coli*. Cette difficulté nous a empêchés de comparer son activité et sa structure avec la forme courte détectée *in vivo*.

Nos études structurales sur la forme courte du complexe ont révélé une nouvelle architecture des AAA+ régulatrices de la protéolyse qui est plutôt conservée entre archées et eucaryotes (Fig. 5, chapitre 5.1) et qui est remarquablement différente des régulateurs bactériens (ClpX et HslU). Nous avons pu générer un modèle Cryo-EM à 10-15 Å de résolution (Fig. 2, chapitre 5.1), qui a ensuite été utilisé pour générer un modèle pseudo-atomique avec les structures cristallographiques du domaine C-terminal ATPase monomérique et du domaine N-terminal sous sa forme hexamérique. Les deux modèles (Cryo-EM et pseudo-atomique) ont ensuite été validés en solution par SAXS. Notre modèle consiste à une structure sous forme d'un hexapode avec six bras sortant du domaine C-terminal et un corps central contenant une grande cavité ouverte vers le protéasome 20S.

Cette nouvelle architecture n'a pas été observée dans les autres modèles de PAN à basse résolution déjà publiés et souligne une forte conservation structurale entre les complexes unfoldases des archées et des eucaryotes (Fig. 5, chapitre 5.1) indiquant des mécanismes similaires pour le dépliement des substrats protéiques ainsi que pour les changements de conformations induits par liaison et hydrolyse de nucléotides. Afin de mieux caractériser ces changements, nous avons généré deux modèles Cryo-EM basse résolution de *PhPAN* en absence ou en présence d'ATPγS (Fig. 4, chapitre 5.1). Une simple comparaison entre ces deux structures montre clairement le changement de conformation dramatique induit par la

liaison de l'ATP γ S qui a abouti à une contraction globale de la molécule et une diminution dramatique du volume de la cavité centrale ainsi que des diamètres des domaines N-terminal et C-terminal.

Nos données structurales et dynamiques ainsi que nos observations biochimiques ont été très utiles pour proposer un mécanisme et un mode d'action de PAN pendant le dépliement des substrats. Le mécanisme le plus probable qu'on peut proposer sur la base de nos données est que le dépliement du substrat se fait indépendamment de la translocation dans la partie catalytique 20S et que ce dépliement favorise l'association transitoire entre PAN et le 20S (Fig.4, chapitre 5.1). Ce modèle nous a amené à privilégier un mécanisme où PAN fonctionne sur la surface du protéasome 20S mais sans translocation par étape progressive. Bien que ce mécanisme semble être le plus réaliste, d'autres études dynamiques et structurales sont encore nécessaires sur les complexes PAN-substrat et PAN-20S pour mieux comprendre le processus de dépliement ainsi que le rôle de chaque partenaire. Les informations structurales et mécanistiques sur PAN fournies dans ce papier nous donnent pourtant des informations précieuses sur le mécanisme d'action de ces larges machines moléculaires ainsi que sur le mode d'action du protéasome 26S eucaryote et en particulier la partie ATPasique qui forme la base du complexe et qui joue le rôle du moteur qui fait fonctionner toute la machine 19S.

5.2 Etude des mécanismes de dépliement et de dégradation des protéines par le système PAN-Protéasome de l'archée *Methanocaldococcus Jannaschii* : mise au point d'une approche en temps réel par diffusion de neutrons aux petits angles

5.2.1 Présentation du travail

Le travail présenté dans le chapitre précédent sur l'étude structurale par cryo-microscopie électronique du complexe PAN de *P. horikoshii* a fourni des informations intéressantes sur l'architecture globale et la structure des complexes PAN régulateurs de la protéolyse chez les archées. Il a également apporté des informations sur la dynamique du complexe créée par la liaison du nucléotide (ATP γ S). Bien que l'enveloppe Cryo-EM et le modèle pseudo-atomique présentés fournissent des informations précieuses sur la structure et les mécanismes fonctionnels du complexe PAN libre, ces modèles représentent des conformations spécifiques et statiques en raison des conditions expérimentales de la microscopie électronique.

Les unfoldases AAA+ régulateurs des systèmes protéolytiques et les chaperonnes de désassemblage détruisent leurs protéines substrats via des mécanismes très dynamiques et soigneusement contrôlés. Des changements de conformation dans le complexe, alimentés par la liaison et l'hydrolyse de l'ATP créent une force de dépliement mécanique nécessaire pour tirer et déplier la protéine substrat attachée avant sa translocation dans la particule catalytique. Les modules AAA+ ATPases des particules régulateurs sont les moteurs qui animent les procédés mécaniques de dénaturation, de désassemblage et de translocation. Dans les machines AAA+, les cycles de liaison d'ATP, d'hydrolyse et de libération d'ADP/Pi entraînent des cycles de changements conformationnels dans l'enzyme (Sauer et al, 2004). Des protéines très stables, celles pour lesquelles le dépliement spontané peut prendre des mois, voire des années, sont dépliées en quelques minutes ou moins par la réaction enzymatique (Weber-Ban et al, 1999) (Kenniston et al, 2003). Certaines enzymes AAA+ génèrent des plus grandes forces de dénaturation et sont donc plus puissantes que d'autres en termes de dépliement des substrats protéiques. Il y a un grand progrès dans la compréhension des rôles et des mécanismes biologiques des AAA+ ATPases impliqués dans la dégradation et le dépliement des protéines. Ce progrès dépend de la disponibilité des informations sur les conformations tridimensionnelles de ces enzymes ainsi que sur la découverte et la caractérisation de nouveaux substrats bien adaptés pour les conditions expérimentales. Malgré tous ces progrès, de nombreuses questions importantes, tant au niveau structurelle et

dynamique, restent sans réponses. En effet, les techniques structurales classiques utilisées précédemment pour l'étude des AAA+ unfoldases restent limitées en termes d'informations dynamiques. Souvent seulement un « snapshot » statique d'un complexe protéique est visualisé malgré le fait que les changements de conformation sont souvent à la base des fonctions biologiques, comme est le cas des moteurs moléculaires AAA+ unfoldase. Les études biophysiques et structurales en temps réel pendant l'action des machines AAA+ unfoldases dans le dépliement des substrats protéiques sont toujours inexistantes et les changements de conformation de ces machines moléculaires à l'échelle de quelques secondes est un domaine d'étude important.

Afin de mieux comprendre la dynamique et les changements conformationnels du complexe PAN et de son substrat modèle la GFPssrA en temps réel en solution pendant le processus de dépliement ATP-dépendant nous avons mis au point et utilisé la diffusion de neutrons aux petits angles (SANS) avec de la deutération et la variation de contraste entre le complexe PAN et son substrat, couplée à la fluorescence en ligne (Fig. 37, chapitre 2.3). Cette étude nous a permis de suivre le processus de dépliement en focalisant séparément sur un des deux partenaires du système. En un premier temps des courbes SANS ont été enregistrées sur la GFPssrA deutérée pendant son dépliement par PAN, ensuite les changements de conformations de PAN ont été également suivis en temps réel pendant l'activité enzymatique. Ce projet porte également un aspect méthodologique vu le court temps d'exposition de 30 secondes utilisé pour le collecte des données SANS qui permet de suivre en temps réel le processus de dépliement ainsi que le couplage avec les mesures de la fluorescence en ligne sur la GFPssrA pendant l'activité enzymatique. La totalité des résultats obtenus ont apporté plusieurs informations précieuses sur le mécanisme fonctionnel du système PAN pendant le dépliement des substrats protéiques ainsi que sur le mode d'action des AAA+ unfoldases impliqués dans la régulation des systèmes protéolytiques. Nos résultats ont également validé l'applicabilité et la faisabilité du SANS en temps réel pour étudier des processus biologiques exécutés par des diverses machines moléculaires.

5.2.2 Small angle neutron scattering reveals protein substrate processing by a AAA+ unfoldase at real-time

Small angle neutron scattering reveals protein substrate processing by a AAA+ unfoldase at real-time

Ziad Ibrahim^{1,2}, Anne Martel², Martine Moulin², Henry Kim¹, Michael Härtlein², Bruno Franzetti¹, Frank Gabel^{1,2,*}

¹Université Grenoble Alpes, CNRS, Commissariat à l'Energie Atomique, Institut de Biologie Structurale F-38044 Grenoble, France

²Institut Laue-Langevin, 71 Avenue des Martyrs, F-38000 Grenoble, France

* Corresponding author: frank.gabel@ibs.fr

ABSTRACT

Energy-dependent proteases are responsible for the regulated destruction of proteins in bacteria, archaea and eukaryotes and AAA+ proteolytic machines degrade damaged misfolded and unneeded proteins to maintain protein homeostasis. This process involves the ATP-dependent unfolding of a target protein and its translocation into a self-compartmentalized proteolytic chamber, the proteasome. Even though their function is essential in numerous biological processes, the structural alterations of the AAA+ proteolytic machines and their protein substrates during the unfolding process remain unclear and only static pictures from crystallography or electron microscopy are available. Real-time, kinetic, structural methods to assay the conformational changes during enzyme catalyzed protein unfolding have not been carried out to date.

Here we describe and use a real-time small-angle neutron scattering (SANS) approach in solution, in combination with deuterium labeling and contrast variation, to follow the mechanical unfolding of a GFP protein substrate by the archaeal PAN AAA+ unfoldase. During unfolding, the substrate is pulled in a directional way by PAN and released (partially) unfolded in solution. Concomitantly, PAN adopted a contracted conformation induced by nucleotide binding and hydrolysis, followed by a slower expansion of the complex at the end of the reaction. These results assume that PAN unfolds their substrates in a peristaltic pumping mechanism. The approach used in the current paper could be broadly applicable for studying the mechanisms of AAA+ proteases and remodeling chaperone.

INTRODUCTION

Protein unfolding by cellular enzymes is necessary for subsequent degradation by AAA+ proteases and for the remodeling of macromolecular complexes (Prakash & Matouschek, 2004) (Sauer & Baker, 2011). In all kingdoms of life, AAA+ proteases are involved in specific regulatory tasks, including the degradation of cell-cycle components, transcription factors and housekeeping proteins that are misfolded, damaged or due to premature translation termination (Baytshok et al, 2015) . The degradation of most proteins in eukaryotic and prokaryotic cells is catalyzed by large proteolytic complexes that hydrolyze ATP and proteins in linked reactions (Baumeister et al, 1998) (Ciechanover, 2005) and the majority of our present knowledge about intracellular proteolysis comes from studies seeking to understand the biochemical basis of the ATP requirement (Ciechanover, 2005) (Goldberg, 2005) . AAA+ energy-dependent proteases are found across the three domains of life. Bacteria contain multiple AAA+ proteases (ClpXP, ClpAP, HslUV, Lon, FtsH and Mpa:20S), archaea have two cytoplasmic AAA+ proteases (PAN:20S and Cdc48:20S); and eukaryotes contain one cytosolic or nuclear AAA+ protease (the 26S proteasome) and homologs of AAA+ bacterial proteases in organelles (Baytshok et al, 2015). These proteolytic machines use cycles of ATP binding and hydrolysis by the hexameric AAA+ ATPase to unfold and degrade improperly synthesized, damaged or misfolded proteins and hydrolyze regulatory proteins that must be specifically removed or cleaved for cell signaling (Maupin-Furlow, 2011). The eukaryotic 26S proteasome is technically challenging for sophisticated biophysical studies due to its complexity in terms of subunit organization and the difficulty to express it as an assembled complex in *E. coli* strains (Rajagopala et al, 2012).

All AAA+ hexamers that translocate and unfold proteins contain related axial-pore loops, with strong conservation of an aromatic tyrosine, tryptophan or phenylalanine side chain at the first position and ATP-fueled changes in the conformation of the hexameric rings cause movement of the axial pore loop implicated in substrate translocation and unfolding. The aromatic side chains in one or more of the pore loops bind the degradation tag as the machine pulls the attached native substrate through the axial channel, forcing its denaturation by different proposed mechanisms (Martin et al, 2008) (Ibrahim et al., submitted) (Kim et al, 2015). Once unfolding is successful, axial pore-loop movements, coupled to conformational changes in the AAA+ unfoldase ring propel the translocating polypeptide through the pore in steps of five to eight residues per power stroke, facilitating processed degradation (Martin et al, 2008) (Kenniston et al, 2005) (Aubin-Tam et al, 2011) (Cordova et al, 2014). Despite the central importance of the nucleotide-dependent, coordinated conformational changes in the

AAA+ proteases and protein remodeling machines, very little is known about the structural arrangements during the potential coordination between nucleotide binding and hydrolysis, conformational changes, substrate unfolding and release (Iosefson et al, 2015). In particular, the biological systems and the experimental techniques used up to date were not able to resolve all issues related to the unfolding process and the ATP-driven protein unfolding mechanism remains unclear: what kind of conformational changes of the complexes are necessary for the unfolding process? How does the folding state of the substrate evolve as it is being processed? What kind of protein substrates (size, folding state) can be unfolded? Moreover, substrate processing by the AAA+ unfoldase machines remains unclear, e.g. structural and conformational modifications of substrate proteins as a function of time during the unfolding and the state in which they are released from the complex. The common feature believed to underlie the diverse functions of the AAA+ family of ATPases is their ability to undergo structural alterations during the ATP power stroke conformational changes that cause unfolding of proteins or disassembly of protein complexes (Hanson & Whiteheart, 2005) .

Protein unfolding and degradation by AAA+ unfoldases has been studied by combining a range of *in vitro* and *in vivo* techniques (Reuter & Maupin-Furlow, 2004): it has been traced by the disappearance of intact substrate on SDS gels (Benaroudj & Goldberg, 2000) the appearance of acid-soluble peptides (Baytshtok et al, 2015) and the loss of native fluorescence for proteins such as GFP (Weber-Ban et al, 1999). Other methods, including single-molecule optical trapping (Iosefson et al, 2015) or Förster resonance energy transfer (FRET) (Baytshtok et al, 2015), have been applied to monitor unfolding by AAA+ enzymes in the absence of proteolysis. While these techniques provide very accurate kinetic data on the reaction, they have not provided real-time structural and dynamical information of the AAA+ unfoldase enzyme or its protein substrates. Moreover, the high (and low) resolution structural studies which were carried out on different AAA+ proteasome regulators (Glynn et al, 2009; Matyskiela et al, 2013; Medalia et al, 2009) provided only static snapshots with one or two conformations of the unfoldase complexes in the presence or absence of nucleotide. A central unsolved challenge in dissecting the mechanisms of these complicated machines is to determine how ATP-binding and -hydrolysis coordinates the conformational changes that ultimately allow substrate unfolding. Although different models are conceivable, most proposals in the literature rely on multiple, untested assumptions, and the paucity of methods to test specific models has limited the insight into these machines.

Because of its structural simplicity with respect to the eukaryotic 19S system and its thermal stability, the archaeal PAN (proteasome activating nucleotidase) regulatory particle has

proven very useful to gain structural and mechanistic insights into AAA+ unfoldase systems (Lowe et al, 1995) (Akopian et al, 1997). The ATPase complex PAN functions in protein breakdown in association with the archaeal 20S proteasome. When both are mixed ATP, PAN stimulates the degradation of natively unfolded proteins and peptides longer than seven residues, whereas shorter peptides diffuse into the 20S proteasome unassisted (Zwickl et al, 1999). PAN forms a hexameric ring spanning a wide, central cavity (Ibrahim et al, submitted) and thus its general structure resembles that of the six ATPases of the 19S complex (Smith et al, 2005) (Ibrahim et al, submitted). It exhibits several typical chaperone properties, including the ability to bind unfolded polypeptides selectively and to prevent their aggregation (Strickland et al, 2000) (Benaroudj & Goldberg, 2000). In addition, in the presence of ATP, PAN efficiently unfolds certain globular substrates and translocates them into the 20S proteasome (Benaroudj & Goldberg, 2000). It displays 40-45% sequence homology to the six ATPases in the base of the eukaryotic 19S particle and appears to have regulated proteasome function prior to the linkage of ubiquitination and proteolysis (Zwickl et al, 1999).

In this paper, we present a novel and robust method to measure real-time conformational changes in solution on the time scale of seconds and apply it to monitor both the unfolding process of a protein substrate (GFP) as well as the concomitant conformational changes of the PAN unfoldase. To this end we used Small Angle Neutron Scattering (SANS), deuteration of GFPsrA and PAN, in combination with online fluorescence spectroscopy and temperature activation. SANS, combined with contrast variation ($H_2O:D_2O$ exchange in solution) and specific deuteration of proteins, is a particularly powerful technique to monitor conformational changes undergone by individual protein partners within a complex by highlighting or suppressing their respective signals (Jacrot, 1976) (Timmins & Zaccai, 1988). Moreover, by using the hyperthermophilic PAN complex from *Methanocaldococcus janaschii* and temperature activation we were able to monitor the PAN and GFP partners individually within their complex in real-time during the unfolding reaction, with a time-resolution of 30 seconds. On the basis of the results obtained, we propose and discuss a molecular mechanism in which ATP hydrolysis by PAN results in a coordinated contraction of the complex providing a power stroke in which the internal cavity concurrently grip and coordinately apply an unfolding force to the protein substrate that will be completely unfolded and released unstructured in solution. We found that the GFP substrate is pulled by PAN in a directional longitudinal process, allowing its unfolding and release in solution within a few seconds, resulting in its aggregation as a second step within a time scale shorter than one minute. During the process we monitored a peristaltic pumping motion of PAN with the same time

scale that drives the directional substrate unfolding and translocation. The contraction is followed by a relaxation of the complex to reach its initial conformation after total consumption of the ATP and both can be described by a bi-exponential law with time scales of several hundred seconds. These findings helped to explain several functional and mechanistic aspects of the PAN complex and the AAA+ unfoldases machines.

RESULTS

Purification and biophysical characterization of a functional hexameric PAN complex from *Methanocaldococcus jannaschii* and its interaction with GFP substrates

Sample quality of biomacromolecules is crucial for structural interpretation of solution scattering data (Jacques & Trewthella, 2010) and we characterized the PAN-GFP system by several biophysical techniques prior to the SANS experiments. The original gel filtration analysis showed that the purified PAN complexes were assembled as a mixture of dodecamers and hexamers (Fig. 1A, blue curve). By heating the dodecamers at 60 °C in the presence of ATP and MgCl₂, it was possible to isolate monodisperse hexameric *Mj*PAN particles that remained stable up to high concentration (~20 mg/ml) (Fig. 1A, black curve). The freshly hexamerized *Mj*PAN particles had a tendency to reassemble into dodecamers in the absence of ATP and high temperature. However, and during all our biophysical studies, we were working at high temperature (> 55 °C, except SPR) in the presence of ATP allowing to keep PAN in its hexameric form. These results, in addition to our previous results on the PAN complex from *Pyrococcus horikoshii* (Ibrahim et al, submitted), highlight the important role played by ATP hydrolysis, processed at high temperature, in the PAN hexamerization process and the generation of monodisperse, hexameric solutions. In addition, the gel filtration column fractions were analyzed by negative stain electron microscopy (Fig. 1A) and by FPLC-SAXS (Fig. S1) where X-ray scattering data were collected on the *Mj*PAN complexes at the output of the two peaks from the Superose 6 gel filtration column (see materials and methods) corresponding to the dodecameric and the hexameric *Mj*PAN, confirming the isolation of a monodisperse solution of the *Mj*PAN hexamers after heating with ATP. Indeed, the experimental SAXS data showed that the radius of gyration had decreased when the PAN complex was recorded after heating at 60 °C in the presence of 4 mM ATP and 10 mM MgCl₂ showing the hexamerization of the PAN complexes.

The hexameric *Mj*PAN was active at 60 °C, capable of unfolding more than 50% of the GFP_srA substrates with PAN in 6-fold molar excess (Fig. 1B, green line). During our studies on the unfolding activity of the hexameric *Mj*PAN complex, of particular interest was the

unexpected finding about the difference in the unfolding rate between the monomeric (~27 kDa) and the dimeric (~54 kDa) GFPssrA. We found that, upon incubation with PAN and ATP, monomeric GFPssrA was unfolded (Fig. 1B, green curve). However, when dimeric GFPssrA was used as substrate, after an initial loss (~ 30%) of the fluorescence signal during the first ~ 2 min of the reaction, the fluorescence subsequently regained almost the initial level after ~ 10 min (Fig. 1B, red curve) showing that, at this ATP amount, dimeric GFPssrA is too large to be completely unfolded by PAN and to pass through the opening of the complex. For the rest of our biophysical studies, the monomeric GFPssrA was used as the only substrate for the PAN complex.

In order to quantitatively characterize the interaction between *Mj*PAN and GFPssrA we used Surface Plasmon Resonance (SPR). Binding of GFPssrA to *Mj*PAN was observed in the presence of 2 mM ATP at 45 °C and was absent in channels lacking PAN particles (Fig. 1 C). The association between PAN and GFPssrA did not reach a steady state showing the lack of formation of stable complexes and indicated only transient interaction. Indeed, once we stopped flowing the GFPssrA over the chip, the dissociation phase of the complex had begun (Fig. 1 C) and the GFPssrA was slowly detaching from the immobilized PAN. By fitting a simple 1:1 Langmuir binding model (see Materials and Methods), we were able to determine the kinetics association k_a ($512.3 \text{ M}^{-1} \cdot \text{s}^{-1}$) and dissociation k_d (0.0022 s^{-1}) rate constants as well as the affinity K_D ($4.28 \text{ }\mu\text{M}$) of the *Mj*PAN:GFPssrA interaction. Interestingly, when wild type (wt) GFP, lacking the ssrA peptide tag on its C-terminal side, was flowed over the chip, the resulting sensorgram showed the completely different behavior of a fast association/dissociation process (Fig. 1 D). The interaction between wtGFP and PAN appeared to be non-specific and multiple GFP molecules were interacting with one molecule of PAN resulting in a higher signal with respect to the GFPssrA case (at the same amount of GFP flown over the chip). Taken together, these results confirm for the first time at a high degree of sensitivity that there is no stable complex formed between PAN and its substrates but only a transient specific interaction to start the unfolding.

Substrates are pulled and unfolded by PAN in a directional longitudinal process

By following the PAN-catalyzed unfolding of GFPssrA by real-time SANS in combination with online fluorescence (Fig. S8), we found that PAN continuously unfolded the GFP substrates within minutes, even in the absence of the 20S catalytic particle (Fig. 2 and 3). Thus, substrate unfolding by PAN does not only correspond to a weak structural destabilization of the GFPssrA leading to the loss of its fluorescence signal but to an

irreversible state of unfolding where the substrate was unable to refold again as seen for the dimeric GFP (Fig. 1B). Our real-time data from the *Mj*PAN:GFPssrA complex showed that the SANS profiles of GFPssrA were evolving continuously over time in 30 s intervals as the protein was being unfolded at 55 °C (Fig. 2) while in the absence of PAN, the scattering profiles of GFPssrA remained virtually unchanged over more than an hour at the same temperature (Fig. S3A). Concomitantly, unfolding resulted in a fast increase in substrate aggregates accompanied by a loss of the online fluorescence signal (Fig. 3). Within 30-60 seconds, GFP was started to being unfolded and losing its native, globular structure through intermediate, elongated structural states before its complete release in solution and subsequent aggregation, starting at about 60 seconds (Fig. 2). After 90 seconds, first GFP aggregates appear as seen through the scattering curves and the generated *ab initio* models. With increasing time, both SANS and fluorescence data show that the unfolded GFP molecules, once released by PAN, were unable to refold again in solution and, after losing their native structure and thus their solubility, interacted to form larger aggregates. Indeed, a continuous increase in zero angle scattering $I(0)$ (Table S1), volumes of the *ab initio* models, as well as the respective populations determined by the OLIGOMER fit (Fig. 3) illustrate the buildup of larger aggregates with time. However, it is important to stress that the model extracted from the first scattering curve (recorded after 30 s) yielded an equal volume as the one generated from the control sample in the absence of PAN (native GFP), but with a different conformation (more elongated particle) (Fig. 2, red model). Thus, after 30 s of the reaction, our sample is still not contaminated with aggregates and a population of the GFP molecules are unfolded (or partially unfolded), giving rise to the elongated shape envelope of the protein and the increase of D_{\max} . We were able to describe the GFP populations in our reaction volume during unfolding by PAN by two species (the folded/native GFP and the aggregates after 50 minutes) as a function of time (Fig. 3) (see Material and Methods). While after ~ 10 minutes most of isolated native GFP (green curve) was lost and clustered into large aggregates (red curve), the decrease of fluorescence (blue curve) was slower and it took more time (~ 50 min) for complete disappearance.

Several control conditions (use of non-hydrolysable ATP analogue ATP γ S and absence of nucleotides) were probed at different temperatures in order to assess the involvement of ATP in the unfolding process. The results show that ATP γ S is not sufficient for GFPssrA unfolding by PAN since the SANS profiles remained identical over 50 minutes at 55 °C (Fig. S5A). Similarly, when the experiment was carried out in the absence of nucleotides, no structural changes in GFP were observed by SANS (Fig. S5B). However, in both cases a small decrease

in the emission fluorescence at 509 nm (Fig. S5C and D) suggested a weak destabilization of GFP_{ssrA}. These unexpected findings indicate that, even in the absence of a hydrolysable nucleotide, PAN and its protein substrate interact *via* the *ssrA* tag.

In conclusion, the present findings indicate that PAN quickly releases completely unfolded substrates in solution when it functions alone without the involvement of other cellular partners like the 20S catalytic particles or other components such as molecular chaperones (Weber-Ban et al, 1999). Furthermore, the unfolding of protein substrates seems to be processed in a unidirectional way allowing the formation of long, unfolded substrates as intermediate states before the accumulation into larger aggregates as observed in the case of thermal unfolding of isolated GFP in a SANS control experiment (Fig. S4).

An ATP-induced reversible contraction of PAN is responsible for substrate unfolding

In analogy to GFP, we monitored (deuterated) PAN conformational changes by real-time SANS with 30 second time resolution at 60 °C in the presence of ATP and an excess of (hydrogenated) GFP_{ssrA} substrate, in parallel with online emission fluorescence measurements (Fig. S8). The SANS profiles of PAN showed a clear evolution within several minutes during the unfolding process towards a more compact form (Fig. 4, blue curve). Indeed, both Guinier analysis and $P(r)$ after 4 min 30 s showed a decrease of ~ 8 Å in the radius of gyration and more than 20 Å decrease in the maximum dimension D_{\max} with respect to the initial state after 30 s (Fig. 4 & 5). Forward scattered intensities $I(0)$ remained stable in the same time range (Table S1), indicating that the PAN oligomeric state remained intact and that the R_G differences corresponded to conformational changes within the hexamer. Moreover, the overall SANS curves at both times did not superpose on the whole q -range, thus validating conformational changes induced by nucleotide hydrolysis during substrate unfolding. The decrease of the radius of gyration followed an exponential behavior (Fig. 5). Interestingly, towards the end of the unfolding reaction, PAN adopted again its initial conformation. Indeed, the scattering curves recorded at the end of the unfolding reaction (50 min) superposed nicely over the whole q -range with the first curve recorded after 30 s of the reaction, indicating the same conformation of the PAN molecule (Fig. 4, pink curve). The Guinier analysis of the final curves yielded the same radius of gyration (~ 68 Å) as the first one (Fig. 5). Similarly, the maximum dimensions (D_{\max}) of the molecule at the end of the reaction calculated from the $P(r)$ also showed good correlation with the first curve (Fig. 4, inset).

In order to quantitatively characterize the contraction and the relaxation mechanism of PAN during substrate unfolding we used a two exponential function (Eq. 1, Material and Methods) that fitted the experimental R_G (extracted by Guinier plots) vs. time (Fig. 5, black curve). The characteristic time parameters τ_{contract} and τ_{relax} , describing the time course of PAN contraction and relaxation during substrate unfolding were 300 and 420 s, respectively. These findings indicate that, at the end of the reaction (~50 min), when all ATP molecules are consumed and the whole substrate is unfolded, PAN is relaxed and re-expanded again to reach its initial, none-contracted conformation. Indeed, the fluorescence signal after 50 min was negligible (Fig 5, inset) showing that most of the GFPssrA was unfolded and unfolding activity by PAN had ceased. Control experiments (not shown) indicated no clear conformational changes in PAN at low temperatures (50-55 °C) but only when the reactions were carried out at 60 and 65 °C. Further control experiments, without nucleotide or with ATP γ S, in the presence of GFPssrA, were performed at 60 °C and showed that PAN does not undergo conformational changes under these conditions (Fig. S6). Without nucleotide or in the presence of ATP γ S the scattering curves remained stable during the 50 min, indicating that in the absence of a hydrolysable nucleotide, PAN did not undergo remarkable conformational changes. The stable conformation observed in the presence of ATP γ S could be explained by the fact that PAN adopts a close conformation once ATP γ S is added to the protein solution and due to the absence of hydrolysis the complex remains locked in this conformation during all the measurement time.

DISCUSSION

GFP interaction, unfolding and release by PAN occurs within seconds

Our SPR study on the PAN-GFPssrA interactions indicate that the association between PAN and its substrate is short-lived and transient with the impossibility of forming a stable complex, which may also explain the fast unfolding and degradation processes observed inside biological cells (A L Goldberg & John, 1976). Of particular interest was the clear difference in interaction with PAN between the ssrA tagged and the wild type GFP, highlighting the crucial role played by the bacterial ssrA tag in addressing the protein substrates for the AAA+ unfoldases and controlling the start of the unfolding reaction (Fig. 1C-D). Although the ssrA tag is not a natural tag for PAN in archaea, it seems to have the same function with PAN *in vitro* by addressing the protein substrate for unfolding (Benaroudj & Goldberg, 2000). Surprisingly, PAN showed a tendency to interact with the non-tagged wtGFP protein in a non-specific way (Fig. 1B), a result which might be justified by the fact

that PAN is a ‘sticky’ protein, able to interact with a wide range of proteins inside the cell but only working in the presence of specific addressing system, like the SAMP or the Urm1 systems in archaea (Humbard et al, 2010) (Anjum et al, 2015).

Our real-time SANS study on GFP_{ssrA}, coupled with online fluorescence, shows that GFP starts to lose its native 3D structure in the presence of PAN within 1-2 minutes (Fig. 2, Fig. 3). Of particular interest was the finding that the average model 30 seconds after the beginning of the unfolding reaction (Fig. 2, red), displayed the same volume as the native GFP model in the absence of PAN (Fig. 2, gray) but with a more elongated shape. This shape elongation is compatible with a directional unfolding process of GFP, probably starting from the C-terminal side which is recognized via the *ssrA* tag to the N-terminal side of PAN. Concomitantly to the unfolding and loss of fluorescence, GFP aggregates appear in solution as indicated by the increase of the scattered intensity at small angles, $I(0)$ (Fig. 2, Table S1). The generated *ab initio* models display elongated shapes with increasing volumes over time as an indicator of the size distribution of GFP aggregates. The disappearance of fluorescence towards the end of the process (~ 50 min) implies total quenching of the fluorophore embedded into the β barrel of the folded GFP. Due to the absence of the 20S catalytic particle in our reaction volume, it can be assumed that the unfolded GFP_{ssrA} was released in solution without being degraded into small peptides. Interestingly, the continuously growing tubular forms and constant diameters of the shape envelopes (Fig. 1) suggest that PAN-unfolded GFP molecules, in the absence of associated 20S particles, spontaneously refold partially into some globular form (“molten globule”?) and aggregate in a “pearl necklace” arrangement, probably interacting via exposed hydrophobic patches. It is noteworthy to stress that the GFP aggregates due to unfolding by PAN are markedly different from the ones appearing after forced thermal denaturation (Fig. S4) suggesting distinct molecular unfolding mechanisms in both cases. In the absence of associated 20S particle, substrate proteins are released from PAN into the cytosol where they may either spontaneously refold if they are only weakly destabilized by PAN (case of the dimeric GFP, Fig. 1B), become bound by other chaperonins to help their refolding, or as seen in our *in vitro* SANS experiment that the dramatically unfolded substrate will interact to form large aggregates in the cytosol.

A reversible, nucleotide-induced global contraction of the PAN complex on a minute time scale is responsible and sufficient for its unfolding activity

The real-time SANS data reveal an ATP-induced, reversible contraction of the *Mj*PAN complex in solution (> 10% diminution of its radius of gyration and D_{\max}) during substrate

unfolding on the time scale of minutes (Fig. 4, Fig. 5, Table S1). To our best knowledge, this contraction has not been observed in previous studies describing AAA+ unfoldase structures and function and reveal a contracted conformation that seems to be required for substrate unfolding. So far, only rotations of the ATPase rings or movements of the pore loops have been described as responses to nucleotide binding and hydrolysis (Glynn et al, 2009). Even though these changes correspond only to a minor (~ 1 Å) decrease in the radius of gyration (calculated from ClpX bacterial AAA+ unfoldase in its nucleotide free (PDB ID 3HTE) and nucleotide bound (PDB ID 3HWS) states), it might suggest that a similar rotation of the nucleotide ring and the movement of the pore loops are responsible for the contraction and mode of action of PAN in our case. Our Cryo-EM study on the *Ph*PAN complex have revealed a global contraction of the molecule upon ATP γ S binding, the reduction of the molecule dimensions appeared in the C-terminal and N-terminal domain diameters as well as in the volume of the internal cavity.

The time course of the reversible PAN contraction could be very well described by a bi-exponential time behavior of the radius of gyration (Fig. 5), suggesting a two-state model between an open, relaxed state, and a closed, contracted state. Both PAN contraction and relaxation occur on a time scale of a few hundred seconds, and after relaxation PAN reaches again its initial conformation indicating that the conformational switch, fueled by ATP hydrolysis, is reversible and needed as a main mechanical force to allow cyclic substrate unfolding. Since nucleotide hydrolysis and release is associated with a global contraction (and consecutive expansion), the continuous conformational transition of the PAN subunits might be sufficient to promote unfolding of globular substrate proteins probably trapped (or partially trapped) within the internal cavity of PAN. However, due to the limitations of the SANS technique, we are unable to determine if the contraction is due to a global, symmetric movement of the six subunits or if it is a random, asymmetric movement concerning only some of the six subunits, depending on the number of bound nucleotides (fully/partially loaded nucleotide states). Our SANS and fluorescence data (Fig. 3) indicate, since GFP was in twofold stoichiometric excess, that PAN has the ability to unfold more than one GFP molecule with different cycles of ATP hydrolysis as we were working with a large excess of ATP (Fig. 5). Furthermore, of particular interest was the finding that at the end of the unfolding reaction, when all ATP molecules were hydrolyzed and the GFP ssrA was unfolded, the PAN complex was expanded again to reach its initial conformation. We therefore can conclude that the conformational alterations induced by ATP hydrolysis and detected by our approach should be sufficient to promote unfolding in a related manner.

Partial unfolding and release of non-translocated substrates

Of particular interest was the unexpected finding that large substrate proteins (dimeric GFP_{ssrA} in our case) were partially unfolded and released in solution where they were refolded again giving an increase in the fluorescence signal (Fig. 1B). In the presence of large protein substrates, PAN probably needs more ATP-driven energy for complete unfolding of substrates. Indeed, by using the same amount of ATP, PAN was able to irreversibly unfold the monomeric GFP while only destabilizing the dimeric GFP without complete irreversible unfolding (Fig. 1B). PAN may either need the help of other proteases for segmentation of large non-translocatable proteins into smaller substrates or the presence of the 20S proteasome. In the absence of these supports, the partially unfolded substrates are released from PAN into solution, where they may spontaneously refold (or partially refold), as seen in our case, or become bound by other components as chaperonins as described by Weber-Ban et al., (Weber-Ban et al, 1999) for the bacterial ClpA unfoldase. This action of unfolding and release may also be exerted by other AAA+ unfoldase family members: for example, in the dissociation by ClpX of the MuA transposase tetramer from DNA (Levchenko et al, 1995). In the case of smaller substrates less energy is needed (Iosefson et al, 2015) and the proteins are completely unfolded and translocated from the cavity of PAN into the hydrolytic cavity of the 20S protease, where they are completely degraded. Our findings on the unfolding by PAN, combined with previous results from Zhang et al. (Zhang et al, 2009b) on the role of the 20S Core Particle (CP), support the hypothesis of a probable role played by the CP in facilitating the unfolding and degradation of large stable non-translocatable substrates. Our results are also consistent with biochemical studies performed by Navon and Goldberg (Navon & Goldberg, 2001) showing that unfolding is initiated on the outer surface of the ATPase ring and that PAN can catalyze unfolding of non-translocatable substrates in a ATP-hydrolysis cycles of conformational changes independently from the translocation and degradation of the unfolded substrates. Consequently, our results favor a mechanism in which unfolding and translocation could be dissociated, however, binding of the 20S CP and translocation/degradation of unfolded substrates can accelerate and increase the rate of substrate unfolding by the PAN complex.

Real-time SANS approach to study AAA+ unfoldase machines and other substrate processing macromolecular machines

Previous studies measured PAN, ClpA or ClpX unfolding of GFPssrA by changes in native fluorescence (Benaroudj & Goldberg, 2000) (Weber-Ban et al, 1999), single molecule optical trapping (Cordova et al, 2014) and Förster resonance energy transfer (FRET) (Baytshtok et al, 2015). However, to the best of our knowledge, none of these methods have been used to determine structural parameters for unfolding by any of the studied AAA+ enzymes or their substrates and there is no simple way to assess real-time structural transitions of the unfolded substrate or the AAA+ unfoldase machines. The innovative real-time SANS approach used in this project, applied for the first time to a biological system, has several advantages with respect to more “classical” approaches. Due to the high performance of the D22 beamline at the ILL (Grenoble, France) we were able to run our experiment with modest sample requirements: small volumes (300 µl), low protein concentrations (2 mg/ml) with relatively small molecules (GFPssrA, ~ 28 kDa) and yet were able to reach short exposure times (30 s), suitable with the time scales of many biological processes. The time scale available can probably be reduced to a few seconds for larger molecules (> 50 kDa) and for slightly higher concentrations (5-10 mg/ml). Thus, this kinetic, real-time SANS approach could be systematically applied to a wide range of biological systems, in particular large molecular assemblies implying mechanical movements to ensure biological functions, including transcription processes, protein degradation (unfoldases and proteases) and refolding as well and the quality control system implying chaperone protein complexes.

MATERIAL AND METHODS

Recombinant protein cloning, expression and purification

The PAN gene from *Methanocaldococcus jannaschii* (MjPAN), optimized for rare codons in *Escherichia coli*, was synthesized and cloned by GeneCust Europe (<http://www.genecust.com>) into a pET30a vector using NdeI/XhoI cloning sites to promote the expression of the protein of interest with a poly-histidine tag on its C-terminal side. MjPAN was over-expressed in *E. coli* strain BL21 (DE3) at 37 °C during 4 h. The cells were re-suspended in 50 ml Tris-HCl 20 mM pH 7.5, NaCl 150 mM, Triton X-100 0.1% and Imidazole 5 mM and supplemented with 12.5 mg lysozyme (Euromedex), 2.5 mg DNase I grade II (Roche), 10 mg RNase (Roche), 50 mg Pefabloc SC (Roche) and 0.5 ml 2 M MgSO₄. Cells were disrupted by sonication with 10 cycles on/off of 30 s each at 4 °C. The lysate was clarified by centrifugation at 12,000 rpm for 1 h (JA20 rotor) and the supernatant was loaded onto a 5 ml Nickel column (HiTrap Chelating

HP, GE Healthcare) equilibrated in 20 mM Tris/HCl pH 7.5, 10 mM MgCl₂, 100 mM NaCl and 5 mM Imidazole. After washing with the same buffer first with 20 mM and then with 50 mM imidazole, bound proteins were eluted with 20 ml at 500 mM imidazole. After dialysis overnight against the same buffer without imidazole, the protein was purified by a 6 ml Resource Q anion exchange column (GE healthcare) equilibrated in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂. After washing with the same buffer, bound proteins were eluted with a 20 CV (column volumes) linear salt gradient (0.10-0.35 M NaCl using the equilibration buffer). The fractions of the peak were combined and for further purification, the protein was loaded on a Superose 6 10/300 GL size exclusion column (GE healthcare) equilibrated in the buffer containing 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂. Protein absorbance at 280 nm was used to determine the concentration of purified protein samples (NanoDrop 2000 spectrophotometer, Thermo Scientific). The purified proteins were stored at -20 °C after the size-exclusion step. Before each use, PAN molecules were hexamerized and activated by heating at 60 °C for 30 min with 4 mM ATP and 10 mM MgCl₂ and the hexamers were re-injected on the Superose 6 10/300 GL size exclusion column as a last purification step and to monitor the complete hexamerization of the molecules.

The DNA sequence encoding the GFPssrA protein (containing an A206K mutation to minimize dimerization) was generated from synthetic DNA fragments and cloned by *GeneCust Europe* (<http://www.genecust.com>) into the over-expression plasmid pET30a using NdeI/NotI cloning sites. GFPssrA is a variant of the green fluorescent protein, with its C-terminus fused to the 11-residue peptide ssrA (AANDENYALAA) (Weber-Ban et al, 1999). In eubacteria this 11-mer targets proteins for degradation and the substrate specificity is usually determined through placement of degradation tags or degrons (Benaroudj & Goldberg, 2000). The resulting construct was used to transform *E. coli* strain BL21 (DE3) for the recombinant expression of the protein at 37 °C. GFPssrA was purified by ethanol extraction followed by Phenyl Sepharose 6 Fast Flow hydrophobic interaction chromatography as described (Samarkina et al, 2009) and resulted in a monomeric form of the protein as checked by gel filtration and SANS. Dimeric GFPssrA was generated by keeping the protein solution at room temperature for few days.

Expression and purification of perdeuterated *Mj*PAN and GFPssrA proteins

Identical, engineered plasmids containing the genes of interest and encoding for both proteins, *Mj*PAN and GFPssrA, were used for the production of the hydrogenated and perdeuterated proteins. The perdeuterated proteins were expressed at the Institut Laue-Langevin Deuteration

Laboratory (Grenoble, France) in *E. coli* BL21 (DE3) bacterial strain. High-cell-density fermentation process in D₂O Enfors minimal medium (Artero et al, 2005) with d₈-glycerol (fully deuterated glycerol) as the carbon source and D₂O as solvent was used to grow bacteria at 30 °C to an OD₆₀₀ of 11, followed by induction of protein expression by isopropyl-β-D-thio-galactoside (IPTG) to a final concentration of 1 mM. Bacteria were harvested when an OD₆₀₀ of 16.6 was reached. The expressed perdeuterated proteins were purified using the same protocols as their hydrogenated counterparts. For purification of both hydrogenated and perdeuterated forms, H₂O-based buffer systems were used.

Surface plasmon resonance (SPR)

The PAN:GFP_{ssrA} interaction was monitored by surface plasmon resonance (SPR) with a Biacore T200 instrument (Biacore Inc.). His-tagged PAN was immobilized on the Ni²⁺-nitrilotriacetic acid (NTA) chip. First, NiCl₂ solution at 0.5 mM was injected on the surface at 10 μl/min for 1 min for Ni²⁺ binding on the surface. PAN was then immobilized via its His-tags on the sensor chip. The protein in its hexameric form was diluted to a final concentration of 100 nM in HBS buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM MgCl₂), supplemented with 50 μM EDTA, 0.025 % Tween 20, 5 mM imidazole and 2 mM ATP and injected on the surface at 10 μL/min at room temperature to generate a response between 500 and 1000 RU (Response Units). GFP proteins, *ssrA* tagged and wild type, were then diluted in the same buffer to concentrations of 500, 1000, 2000, 3000, 4000, 5000 and 6000 nM for GFP_{ssrA} and 100, 200, 300, 400 and 500 nM for wtGFP and injected at 45 °C with a flow rate of 50 μl/min for 2 or 3 min. The regeneration of the surface and dissociation of fixed proteins was carried out by injecting a solution of 350 mM EDTA at 30 μl/min during 1 min in order to chelate the metal (Ni²⁺), the ligand (PAN) and the analyte (GFP). The surface was then recharged with Ni²⁺ for a second cycle of immobilization and interaction. Analyses of protein concentrations were done in duplicate and any background signal from a NTA-only reference flow cell was subtracted from every data set. In the case of GFP_{ssrA}, data were analyzed using a simple 1:1 Langmuir interaction model with the Biacore T200 evaluation software (Biacore Inc.) to determine the kinetics (association/dissociation rate constants; k_a/k_d) as well as the affinity (K_D) of the PAN:GFP_{ssrA} interactions.

BioSAXS measurements

An on-line FPLC system (Viscotek GPCmax, Malvern Instruments) was attached directly to the sample-inlet valve of the BM29 beamline (ESRF, Grenoble, France) sample changer

(Round et al, 2013). Protein samples were loaded into vials and automatically injected onto the column (Superose 6 10/300 GL, GE Healthcare) via an integrated syringe system. Buffers were degassed and a flow rate of 0.5 ml/min at room temperature was used for all sample runs. The buffer was the same as described in the purification paragraph for the gel filtration column. Prior to each run, the column was equilibrated with 2 CV of buffer and the baseline was monitored. All data from the run were collected at a wavelength of 0.9919 Å using a sample-to-detector (PILATUS 1M, DECTRIS) distance of 2.81 m corresponding to a Q -range of 0.008–0.45 Å⁻¹. Approximately 550 frames with an exposure time of 4 s/frame were collected per sample run. Initial data processing was performed automatically using the EDNA pipeline (Incardona, 2009), generating radially integrated, calibrated and normalized one-dimensional profiles for each frame. Initial frames were averaged to create the reference buffer and the frames collected from each elution peak (10 frames/peak), corresponding to the scattering of an individual purified species, were also averaged and subtracted from the reference buffer using the program PRIMUS (Konarev et al, 2003). Radii of gyration R_G and pair distance distribution functions $P(r)$ were extracted as described in the “SANS data collection and analysis” section below.

SANS data collection and analysis

All SANS data sets were recorded on the D22 diffractometer at the Institut Laue Langevin (ILL, Grenoble, France). For the static control experiments (match-points and masking of hydrogenated proteins, Fig. S2), 200 µl of the sample solutions were measured at room temperature in quartz cuvettes (Hellma 100QS) with 1 mm path length. Two configurations of the instrument were chosen, the sample was placed with a collimator/detector distances of 2/2 m for the GFPssrA and 5.6/5.6 m for *Mj*PAN. The wavelength was 6 Å in all cases.

For the real-time kinetic experiments, 300 µl of the reaction mixture (to cover the neutron and the UV fluorescence beams) (Table 1 & 2) were prepared and pipetted in the quartz cells a few seconds before putting them in the thermostated rack already adjusted to the desired temperature needed for the enzymatic reaction. The unfolding reaction immediately started once the cuvette was placed in the rack. The scattering curves were recorded with 30 s exposure time, in total 180 scattering curves were recorded on 90 minutes reaction time for each condition. Two main reaction mixtures were measured in a medium containing 42% D₂O allowing the matching of the hydrogenated protein signals: dGFPssrA:hPAN with a collimator/detector distance of 2/2 m and dPAN:hGFPssrA with a collimator/detector distance 5.6/5.6 m (Table 2). Both kinetics were equally measured in different control experiments:

with ATP, ATP γ S or in the absence of nucleotide, at different temperatures (50, 55, 60 and 65 °C), at different PAN:GFPssrA stoichiometry or in the absence of one of the protein partners. The corresponding buffers, the empty cell, the empty beam and boron/cadmium were measured in the same collimator/detector setups to perform data reduction (correction for detector efficiency, electronic noise, and sample holder scattering). Data reduction and radial integration over a two-dimensional image were performed using standard ILL software (Dewhurst, 2002).

1st reaction: dGFPssrA/hPAN

| Protein/component | Final concentration | Volume | Buffer |
|--------------------|---------------------|-------------|----------------------|
| hPAN | 10 mg/ml | 150 μ l | H2O |
| dGFPssrA | 2 mg/ml | 60 μ l | D2O |
| ATP/ATP γ S | 100 mM | 66 μ l | D2O |
| MgCl ₂ | 100 mM | 12 μ l | H2O |
| Buffer | / | 12 μ l | H2O |
| Total | / | 300 μ l | 42% D ₂ O |

Table 1: Samples composition, dGFPssrA/hPAN.

Buffer composition: Tris/HCl pH 7.5 (20 mM), MgCl₂ (10 mM), NaCl (100 mM).

2nd reaction: dPAN/hGFPssrA

| Protein/component | Final concentration | Volume | Buffer |
|--------------------|---------------------|-------------|----------------------|
| dPAN | 3 mg/ml | 90 μ l | H2O |
| hGFP | 4.8 mg/ml | 72 μ l | H2O |
| ATP/ATP γ S | 100 mM | 66 μ l | D2O |
| MgCl ₂ | 100 mM | 12 μ l | H2O |
| Buffer | / | 60 μ l | D2O |
| Total | / | 300 μ l | 42% D ₂ O |

Table 2: Sample composition, dPAN/hGFPssrA.

Buffer composition : Tris/HCl pH 7.5 (20 mM), MgCl₂ (10 mM), NaCl (100 mM).

The scattering curves of the buffers were subtracted from the respective sample scattering curves with the software PRIMUS (Konarev et al, 2003). The radii of gyration, R_G , and forward scattered intensities, $I(0)$, were extracted using the Guinier approximation, i.e. by a

linear fit in an $\ln[I(q)]$ vs. q^2 plot (Guinier, 1939). The pair-wise distance distribution functions $P(r)$ were generated with GNOM (Svergun, 1992) by imposing $P(r=0)=0$ and $P(r=D_{\max})=0$ (D_{\max} = maximum distance found in the particle). The programs DAMMIN (Svergun, 1999) and DAMMIF (Franke & Svergun, 2009) were used to generate low resolution envelopes of GFPssrA and *Mj*PAN from the SANS data up to $Q_{\max}= 0.3-0.35 \text{ \AA}^{-1}$ and $0.12-0.14 \text{ \AA}^{-1}$, respectively. As a quality control, the radii of gyration determined from the $P(r)$ analyses were required to agree with the ones from the Guinier analyses. For the GFPssrA, the first experimental scattering curves recorded at the beginning of the unfolding were compared to the back-calculated scattering curves from the atomic model (PDB ID 3B3P) using the program CRYSON (Svergun et al, 1998). The program OLIGOMER (Konarev et al, 2003) was used to fit the experimental scattering curves from the multicomponent mixture of the GFPssrA during the unfolding by PAN (native folded GFPssrA and GFPssrA aggregates) by the form factors of each component calculated by the program ffmaker (from OLIGOMER) to find the volume fractions of each component in the mixture during different times of the unfolding.

The time dependence of the PAN radius of gyration during substrate processing (from the dPAN:hGFPssrA data) was fitted by a two-exponential function

$$f(t) = R_G \left\{ e^{\left(-\frac{t}{\tau_{\text{contract}}}\right)} + \left(1 - e^{\left(-\frac{t}{\tau_{\text{relax}}}\right)}\right) \right\} \quad (\text{Eq. 1})$$

by the program IGOR Pro (Gomez et al, 2002). R_G is the radius of gyration extracted by the Guinier fits, and τ_{contract} and τ_{relax} are the time relaxation constants of the PAN complex, describing its contraction and the relaxation, respectively, during substrate unfolding.

Online fluorescence measurement

In parallel to all kinetic SANS measurements, online fluorescence measurements were carried out. To this end, a special setup with an *in situ* spectrofluorimeter (Ocean Optics, 65000 pro) was implemented on the D22 beamline at the ILL (Grenoble, France) (Fig. S8). Fiber optics (IDIL Fibres Optiques) were connected to the SANS quartz cuvette containing the reaction volume and the excitation and emission wavelengths were filtered to be at 400 and 509 nm, respectively. The fluorescence emission spectrums were recorded at 1 spectrum/s and the fluorescence data were registered on the Nexus files (Maddison et al, 1997), together with the SANS data. All spectrums were subsequently extracted and analyzed by the program HDFview (<https://www.hdfgroup.org/products/java/hdfview/>).

Negative-stain electron microscopy

After the size exclusion column step of the purification, 4 μ l of the dodecameric and the hexameric fractions of *Mj*PAN sample (\sim 0.1 mg/ml) were deposited onto carbon-coated 400-mesh copper grids. The samples were stained using 2% sodium silicotungstate pH 7.5 and air-dried. Images were taken under low-dose conditions in a T12 FEI electron microscope working at 120 kV and with a nominal magnification of 40,000 using an Orius SC1000 CCD camera.

Enzyme assays

GFPssrA was used as an *Mj*PAN substrate, based on the observation that GFP is extremely heat stable, with a melting temperature of 83 °C (Nagy et al, 2004). In addition, the unfolding of GFP is accompanied by disappearance of fluorescence, which can be quantitatively monitored in real-time during the assay (Zhang et al, 2009b). The reaction buffer contained 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂ and 4 mM ATP. The unfolding curve of GFPssrA was measured dynamically by Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek) at $\lambda_{\text{ex}} = 400$ nm and $\lambda_{\text{em}} = 509$ nm, with a temperature control at 60 °C. The monomeric or dimeric GFPssrA was diluted with the reaction buffer to a final concentration of 1 μ M, in the presence and absence of PAN. As PAN is only active as a hexamer, the protein was first heated for 30 min at 60 °C in the presence of ATP and MgCl₂, purified by gel filtration and then used for the activity assay. To compare the unfolding ability of PAN, the total fluorescence decrease of GFPssrA was measured after incubation at 60 °C for 20 min.

ACKNOWLEDGEMENTS

We acknowledge financial support from the Agence Nationale de la Recherche (grants ‘Archelyse’ ANR-12-BSV8-0019-0 and “HYDROSAS” ANR-11-JSV5-0003). Ziad Ibrahim was supported by an ILL PhD fellowship. This work used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). We thank Daphna Fenel for carrying out electron microscopy experiments and Nicole Thielens for her help in carrying out the SPR experiments. SAXS data were recorded at the ESRF BioSAXS beamline BM29 (local contact: Martha Brennich) and SANS data at the ILL instrument D22 using the BAG systems and D11 (local contact: Peter Lindner). This work benefited from the activities of the DLAB consortium funded by the European Union under contracts HPRI-2001-50065 and RII3-CT-2003-505925, and from UK Engineering and Physical Sciences Research Council-funded activity within the ILL-EMBL Deuteration Laboratory under grants GR/R99393/01 and EP/C015452/1

FIGURES

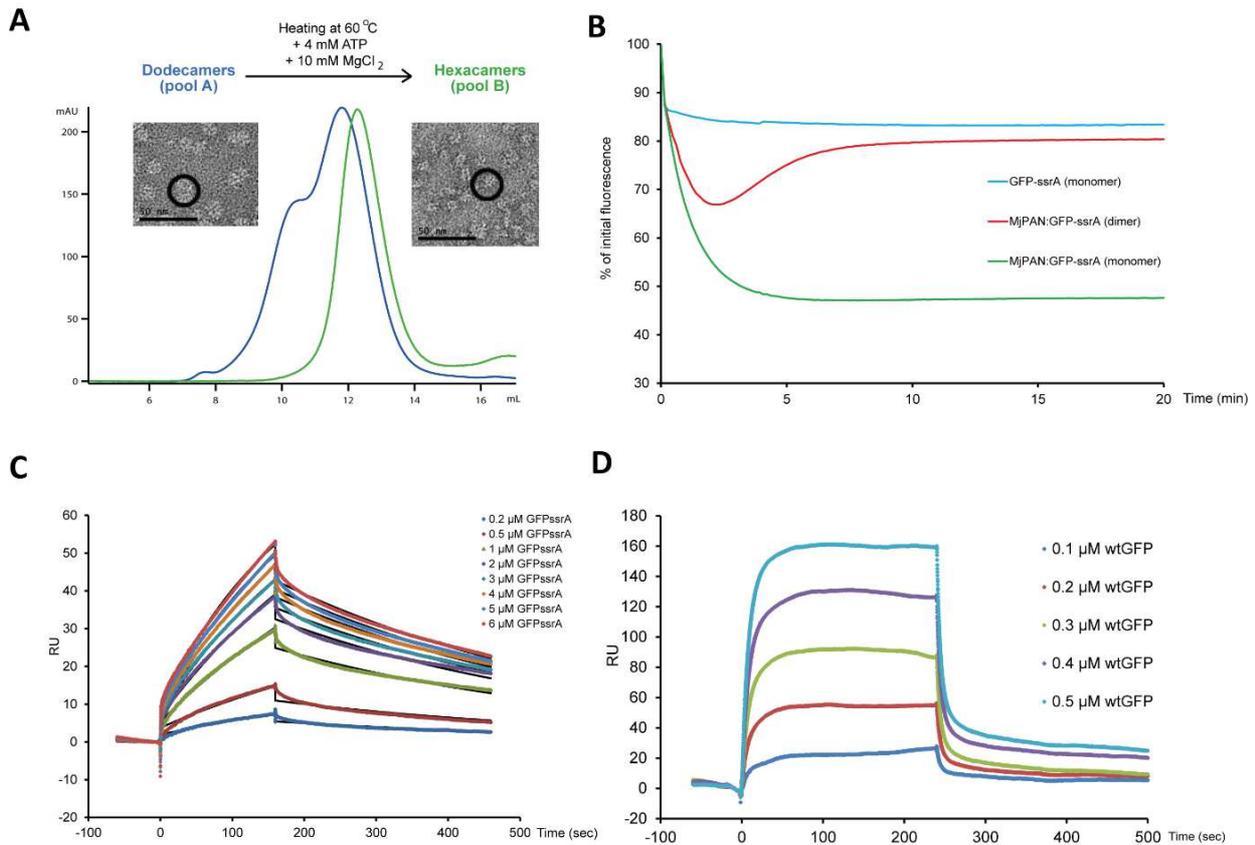


Figure 1: purification, functional and biophysical characterization of the *Mj*PAN complex.

(A) Superose 6 column chromatography purification profiles of the dodecameric/hexameric mixture (blue chromatogram) and the hexameric (black chromatogram) forms of *Mj*PAN complexes. Hexamerization of PAN dodecamers was induced by heating the complex for 30 min at 60 °C in the presence of 4 mM ATP and 10 mM MgCl₂. Negative stain transmission electron micrographs of the *Mj*PAN complex in its dodecameric (pool A) and hexameric (pool B) forms are shown next to the corresponding chromatogram. The elution volumes and the EM shapes of the purified complexes from pools A and B correspond to dodecamers and hexamers, respectively. The respective sample states were further confirmed by FPLC-SAXS (Fig. S1). (B) *Mj*PAN is unable to completely unfold dimeric GFP complexes in our experimental conditions. Time course of fluorescence intensity changes of 1 μM of monomeric (red curve) or dimeric (green curve) GFPssrA in the absence of PAN (blue curve) or in the presence of 6 μM *Mj*PAN. PAN was unable to unfold dimeric GFPssrA and the protein is released and refolded again after partial destabilization by PAN. All measurements were done at 60 °C in buffer in the presence of 4 mM ATP. (C, D) Determination of PAN:GFPssrA association by surface plasmon resonance (SPR). Sensorgrams of the *Mj*PAN complex to GFPssrA (C) and wtGFP (D) are shown. The sensorgrams are given as resonance response units (RU) as a function of time. The black curves superimposed on the sensorgrams represent the model fitted curves. The rate constants k_a and k_d were determined simultaneously as local fitting parameters from which the K_D was determined.

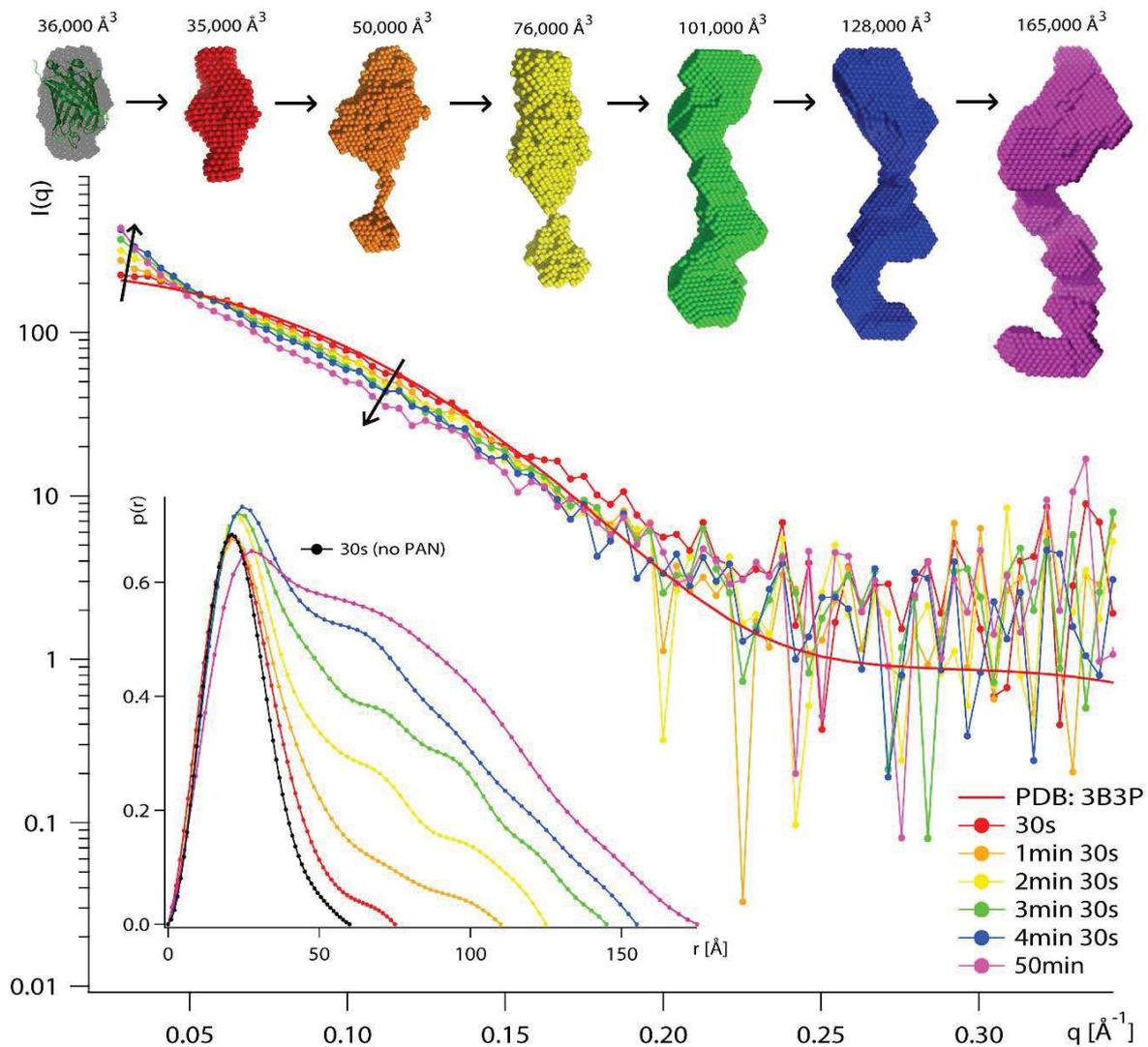


Figure 2: GFPssrA is released and aggregated in solution after unfolding by PAN.

Middle: SANS curves recorded on deuterated (d) dGFP (2 mg/ml) in the presence of hydrogenated (h) hPAN (10 mg/ml) and ATP (100 mM) in a 42% D₂O buffer at 55 °C during the first 50 minutes of the unfolding reaction. The scattering profiles show an evolution over time indicating that the GFPssrA is unfolded by PAN and released in solution. The back-calculated SANS curve (using CRYSON) from the GFP crystal structure (PDB ID 2B3P) is fitted to the 30 s experimental curve and shows good agreement. Top: *ab initio* envelopes of dGFP at different times of the unfolding process, the volume calculated from each model is shown on the top. The first model (gray) is generated from SANS data without hPAN (control experiment, see Fig. S3) and is superposed with the crystal structure. Inset, bottom left: pair-distance distribution functions, P(r), calculated using GNOM.

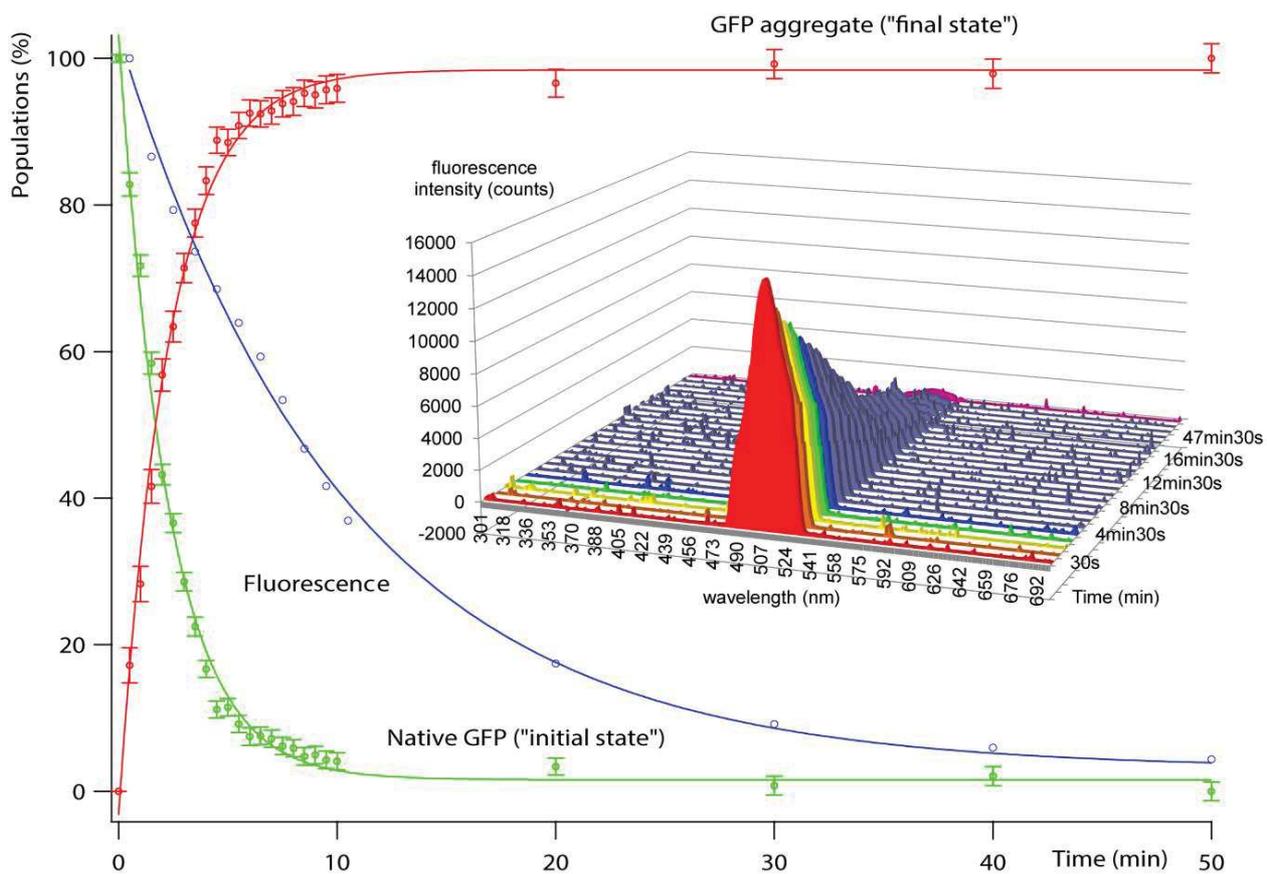


Figure 3: real-time quantification of native and aggregated GFPssrA populations during unfolding by PAN.

Curves showing the percentages of natively folded GFP (green) and its aggregates (red) during 50 minutes of the unfolding reaction. The population of each species at different time points of the reaction is calculated by the program OLIGOMER. Inset: UV fluorescence spectroscopy emission spectrums of GFPssrA in the presence of PAN and ATP at 55 °C recorded on the same samples and at the same time as the SANS measurements. The 1D representation of the GFP emission peak fluorescence at 509 nm is shown in blue, together with the GFP populations.

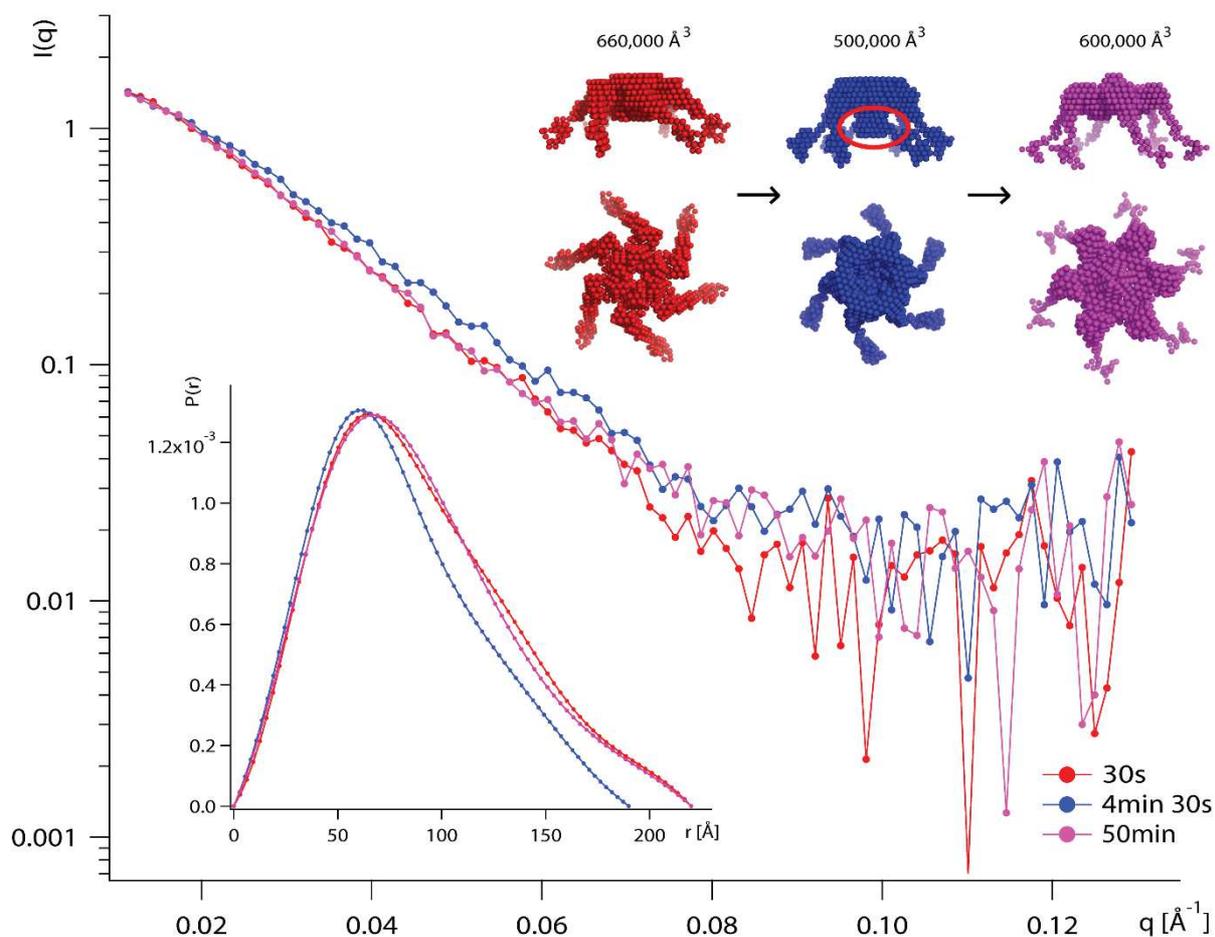


Figure 4: ATP-induced contraction followed by a re-expansion of PAN hexamers during substrate unfolding.

Middle: Scattering curves recorded on deuterated (d) dPAN (3 mg/ml) in the presence of hydrogenated (h) hGFP (5 mg/ml) at 60 °C with 30 s exposure time per curve during different times of the unfolding process. The first scattering curve recorded after 30 s (red) of the unfolding reaction is compared with the ones recorded after 4 min 30 s (blue) and 50 min (pink). Top, *ab initio* envelopes of dPAN-hGFP at 30 s, 4 min 30 s and 50 min, the volumes calculated from each model are shown on the top. Inset: pair distance distribution functions, $P(r)$, calculated using GNOM, of dPAN-hGFP at different times of the unfolding reaction show a clear contraction and relaxation of PAN during unfolding.

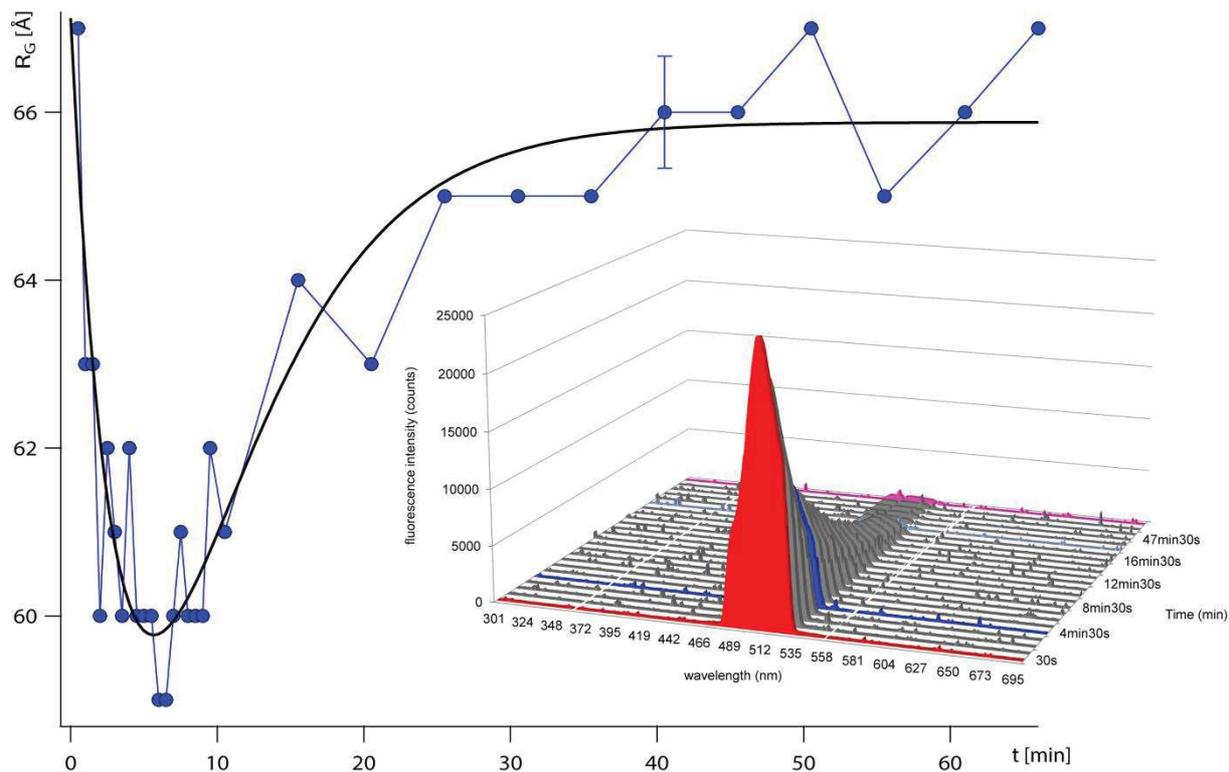


Figure 5: time dependence of the contraction and the re-expansion of PAN during substrate unfolding. Evolution of the radius of gyration (R_G) of PAN, calculated by the Guinier approach as a function of the unfolding reaction time during 60 min of the reaction. A two exponential function (Eq. 1) is fitted against the experimental data (black curve) allowing the extraction of the contraction and the relaxation constants of PAN. Inset: UV fluorescence spectroscopy emission spectra of the GFPssrA in the presence of PAN and ATP at 60°C recorded on the same samples and at the same time as the SANS measurements.

SUPPLEMENTARY FIGURES

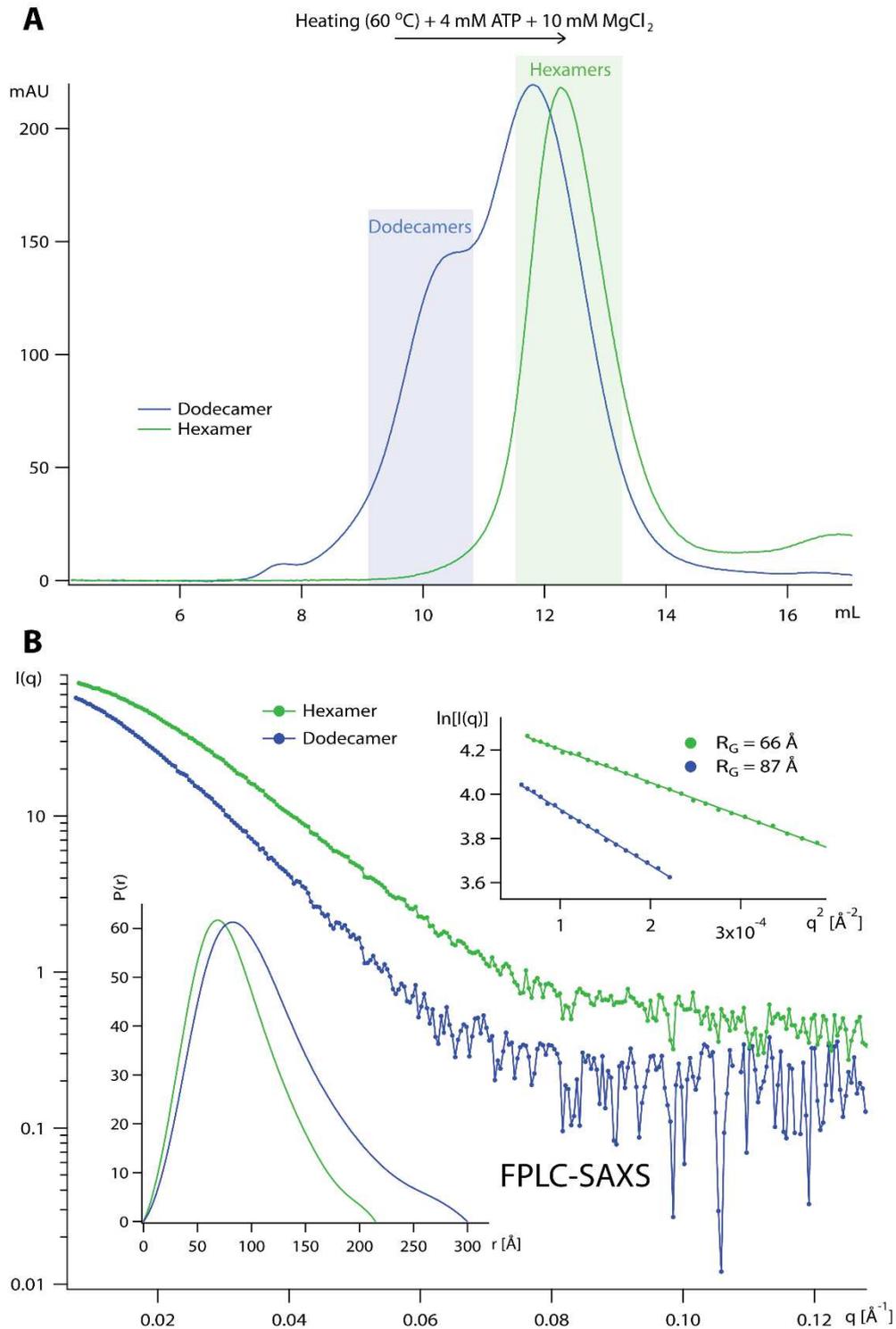


Figure S1: FPLC-SAXS study of the *MjPAN* complex in its dodecameric and hexameric state.

(A) Superose 6 column gel filtration profiles of the dodecameric and the hexameric *MjPAN*. The shaded rectangles indicate the frame zones used for averaging and analysis of the SAXS curves corresponding to both oligomeric forms. Individual frames were collected with an exposure time of 4 s. (B) Scattering curves, pair distance distribution functions $P(r)$ (GNOM) and Guinier plots from the dodecameric (blue points) and the hexameric (green points) *MjPAN*. The hexameric form of the complex was generated by heating the dodecamers at 60 °C with 4 mM ATP and 10 mM MgCl₂ before re-injecting the protein solution on the superose 6 column. Each scattering curve corresponds to the average of 10 frames collected on the desired peak area.

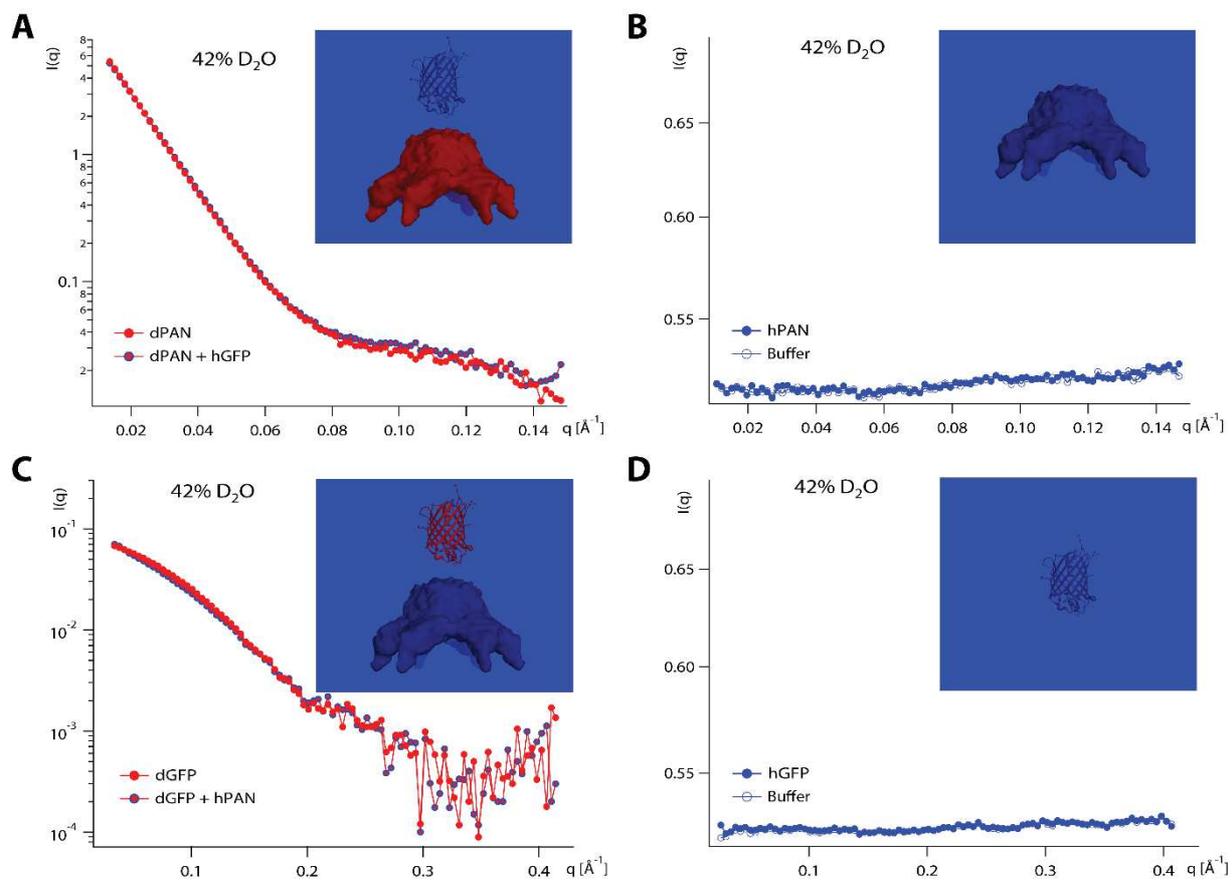


Figure S2: SANS control experiments on hydrogenated (h) and deuterated (d) proteins and their complexes validate the SANS strategy.

(A) dPAN (5 mg/ml) in the presence (blue points) or absence (red points) of hGFPssrA (1 mg/ml). (B) hPAN (2 mg/ml) (filled blue circles) superimposed with its corresponding buffer (open blue circles). (C) dGFPssrA (2 mg/ml) in the presence (blue points) or absence (red points) of hPAN (10 mg/ml). (D) hGFPssrA (4 mg/ml) (filled blue circles) superimposed with its corresponding buffer (open blue circles). All SANS data were recorded at 42% D_2O and the very good superposition of the respective curves over the total q range show a negligible contribution of the hydrogenated proteins to the final signal. All curves were measured at room temperature in a static mode and without ATP in a buffer containing 20 mM Tris/HCl pH 7.5, 100 mM NaCl and 10 mM $MgCl_2$.

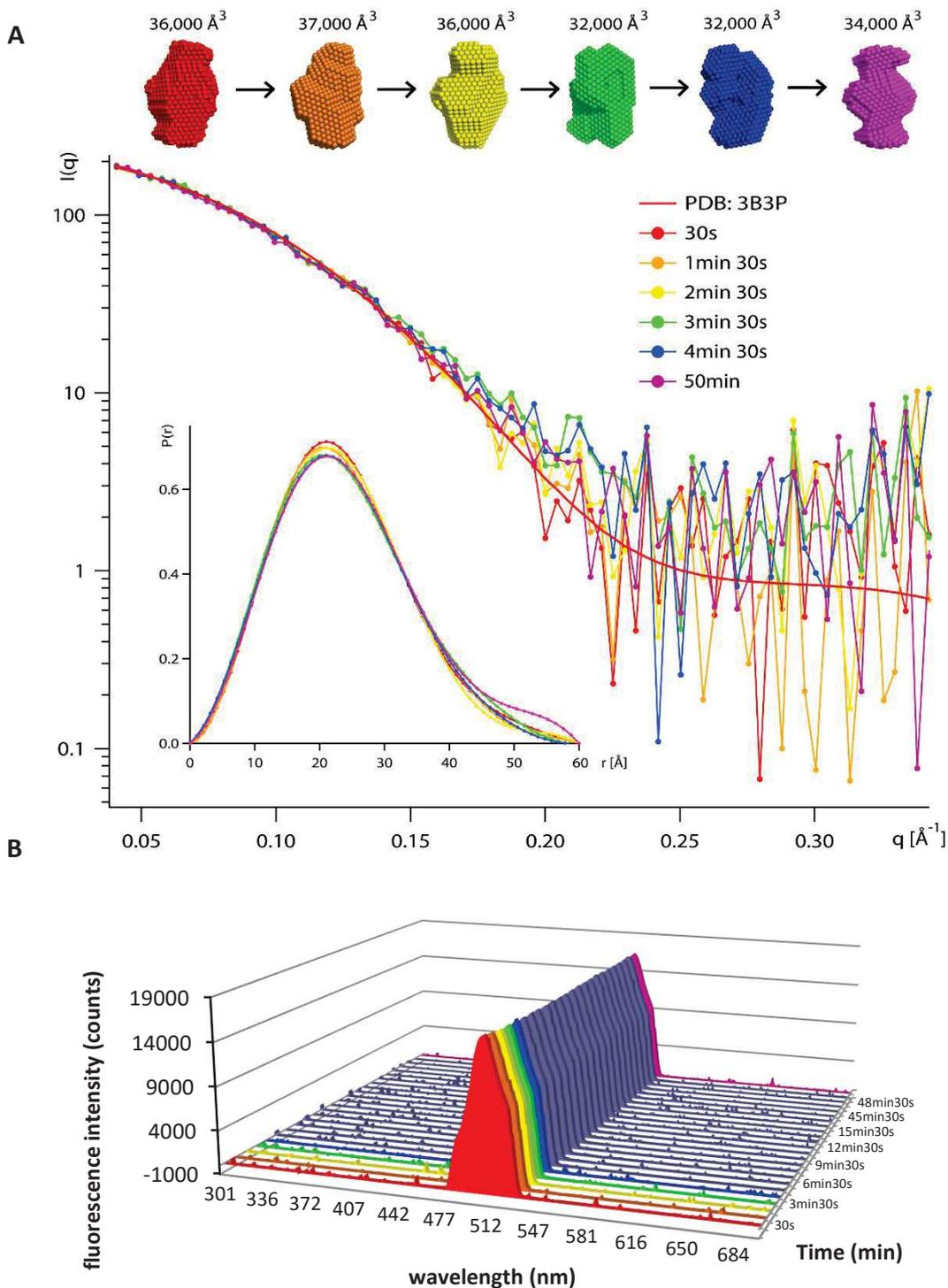


Figure S3: structural stability of isolated GFPssrA at 55 °C in solution.

(A) SANS curves of dGFPssrA (2 mg/ml) in the absence of hPAN recorded at 55 °C during the first 5 minutes and after 50 minutes with 30 second exposure time per curve. The scattering profiles were stable over time showing no thermal unfolding of isolated GFP. *Ab initio* envelopes of dGFPssrA, volumes calculated from each model and pair distance distribution functions $P(r)$ (GNOM) are equally shown. (B) UV fluorescence spectroscopy emission spectra of dGFPssrA in the absence of hPAN, measured on the same sample and at the same time as the SANS measurements show that the GFP fluorescence signal remained stable over time (same color code as in (A)).

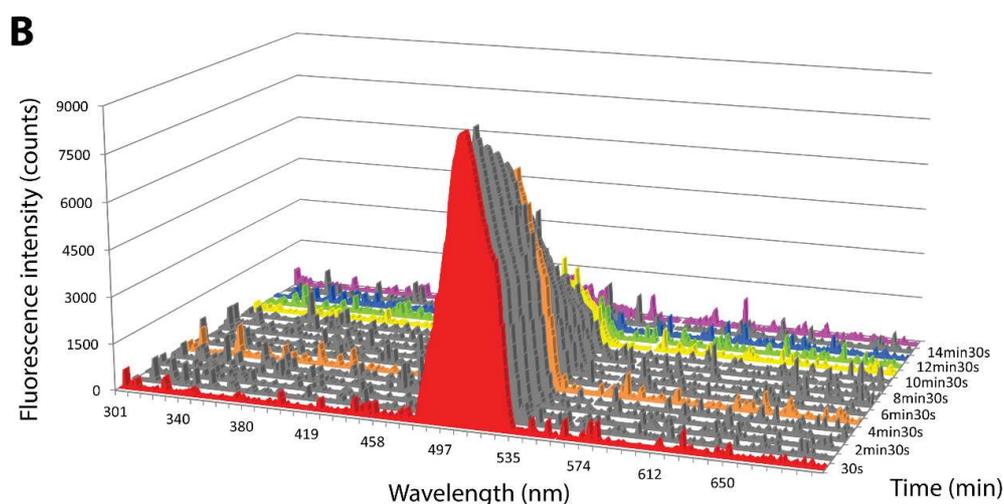
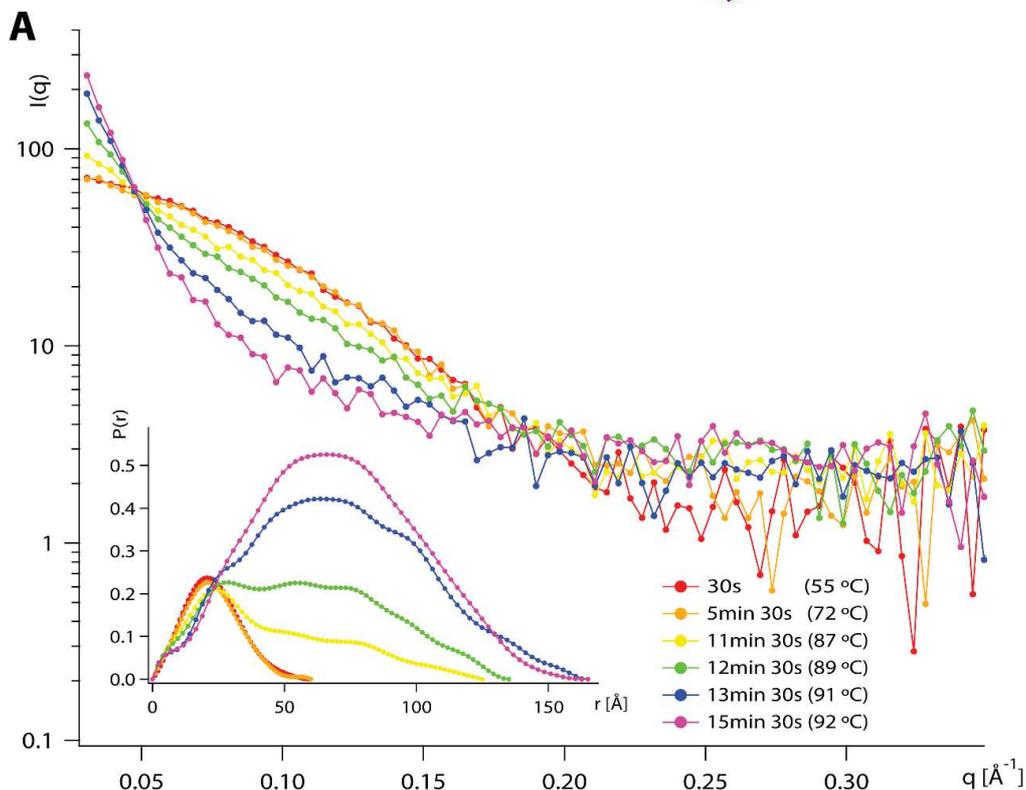
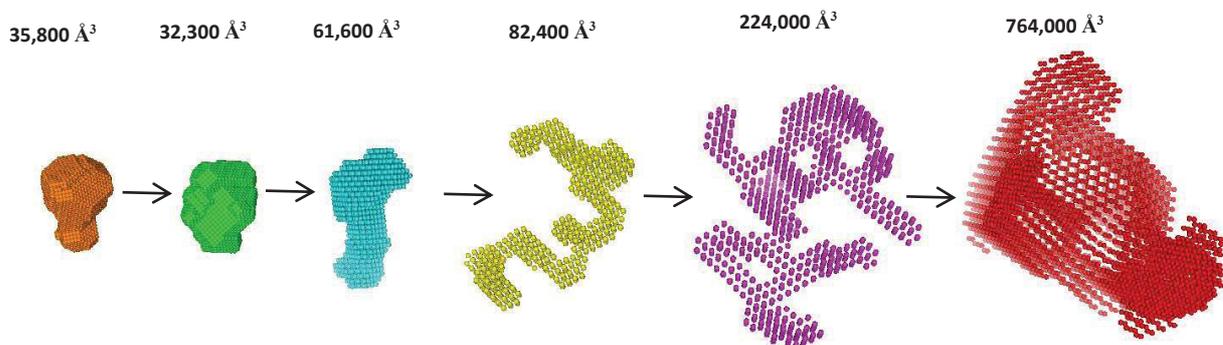


Figure S4: thermal unfolding of GFPssrA.

(A) SANS data recorded on isolated dGFPssrA (2 mg/ml) at different times during a temperature increase from 55 to 95 °C. The scattering profiles show a pronounced evolution over time indicating that GFP is being unfolded and aggregates in solution. *Ab initio* envelopes of dGFPssrA, volumes calculated from each model and pair distance distribution functions $P(r)$ (GNOM), are equally shown. (B) UV fluorescence spectroscopy emission spectra of dGFPssrA measured during the thermal unfolding process on the same sample and at the same time as the SANS measurements show a clear decrease of the GFP fluorescence signal (same color code as in (A)).

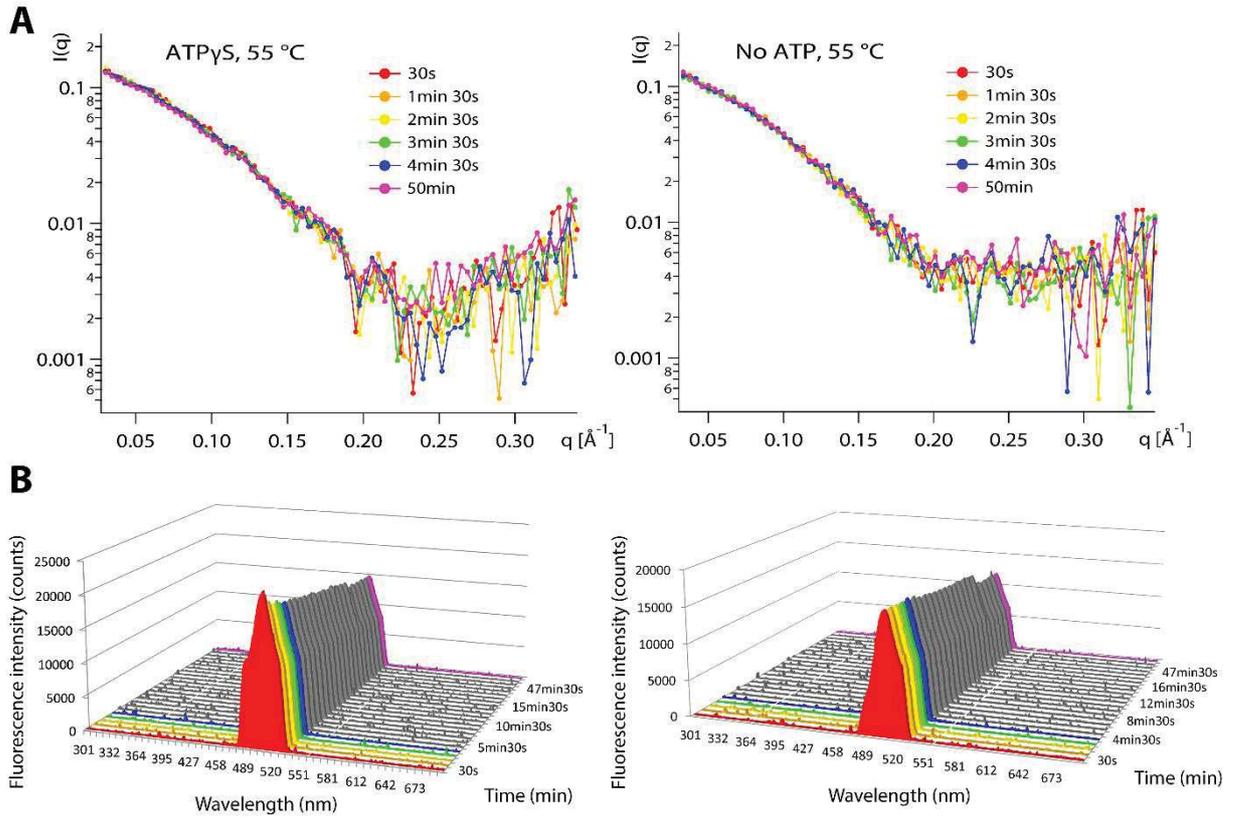


Figure S5: the presence of ATP γ S or the absence of nucleotide prevent GFPssrA unfolding.

(A) SANS curves recorded on dGFPssrA (2 mg/ml) in the presence of hPAN (10 mg/ml) without nucleotide (left) or with ATP γ S (right) at 55 °C during the first 5 minutes and after 50 minutes with 30 s exposure time per curve. The scattering profiles were stable over time showing no unfolding of GFP under these conditions. (B) UV fluorescence spectroscopy emission spectrums of GFPssrA in the presence of PAN without nucleotide (left) or with ATP γ S (right), recorded during the first 17 minutes of the unfolding reaction on the same samples and at the same time as the SANS measurements (same color code as in (A)).

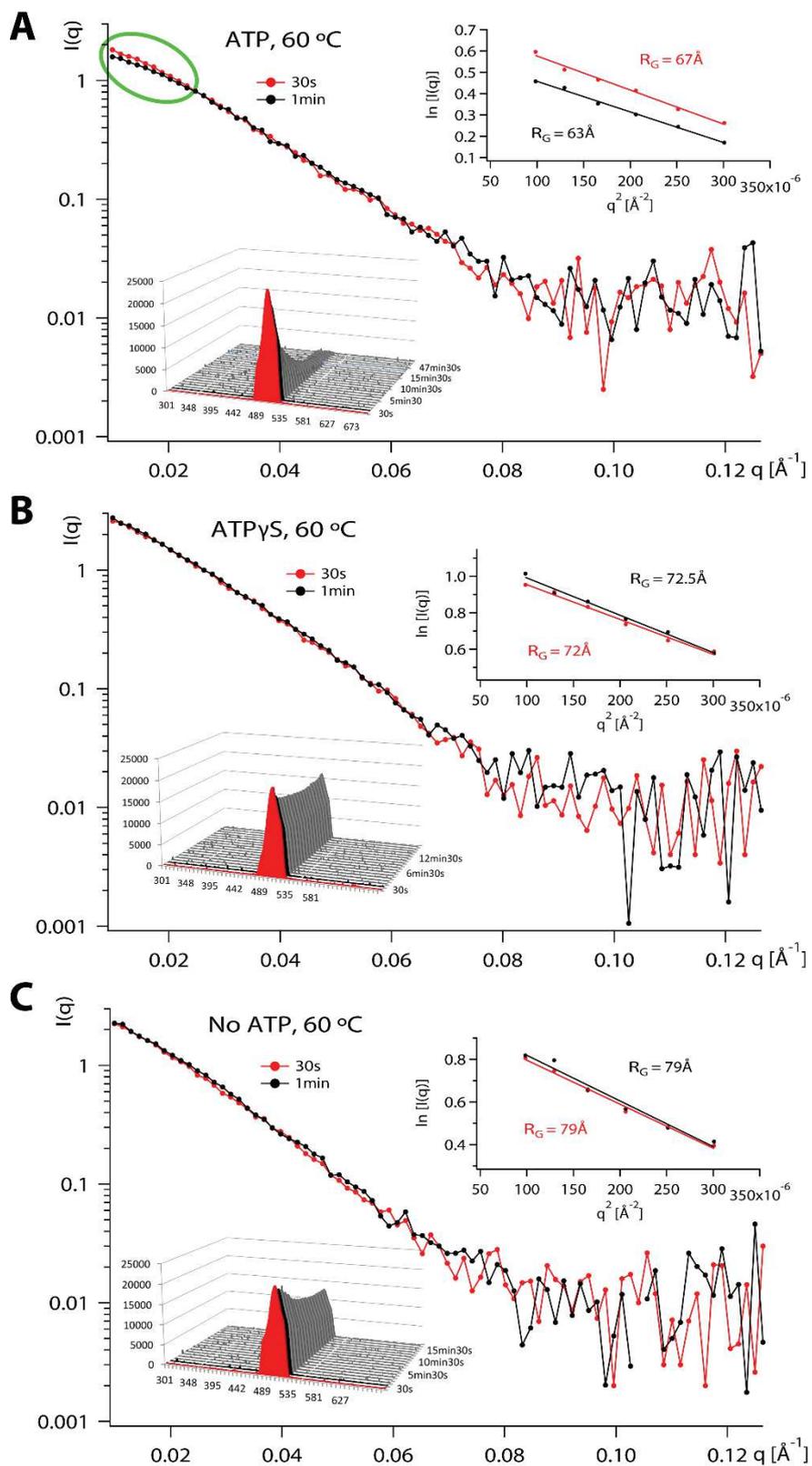


Figure S6: lack of conformational changes in the absence of hydrolysable ATP.

SANS curves recorded at 30 s and 1 min on dPAN (3 mg/ml) in the presence of hGFPssrA (5 mg/ml) at 60 °C in the presence of ATP (A), ATP γ S (B) or without nucleotide (C) with 30 s exposure time per curve. Insets: Guinier plots and UV fluorescence spectroscopy emission spectra of GFP in each condition recorded during the first 15 min of the unfolding reaction on the same samples and at the same time as the SANS measurements. In the absence of ATP or in the presence of ATP γ S, the fluorescence signal remained relatively stable over time with a very small decrease due to the interaction with PAN.



Figure S7: Images of the quartz SANS cuvettes containing PAN and its substrate GFP after the unfolding reaction and after the SANS experiment. Samples in each cuvette are indicated above. GFP aggregates after unfolding in the dPAN/hGFP and dGFP/hPAN samples are indicated by dashed circles.

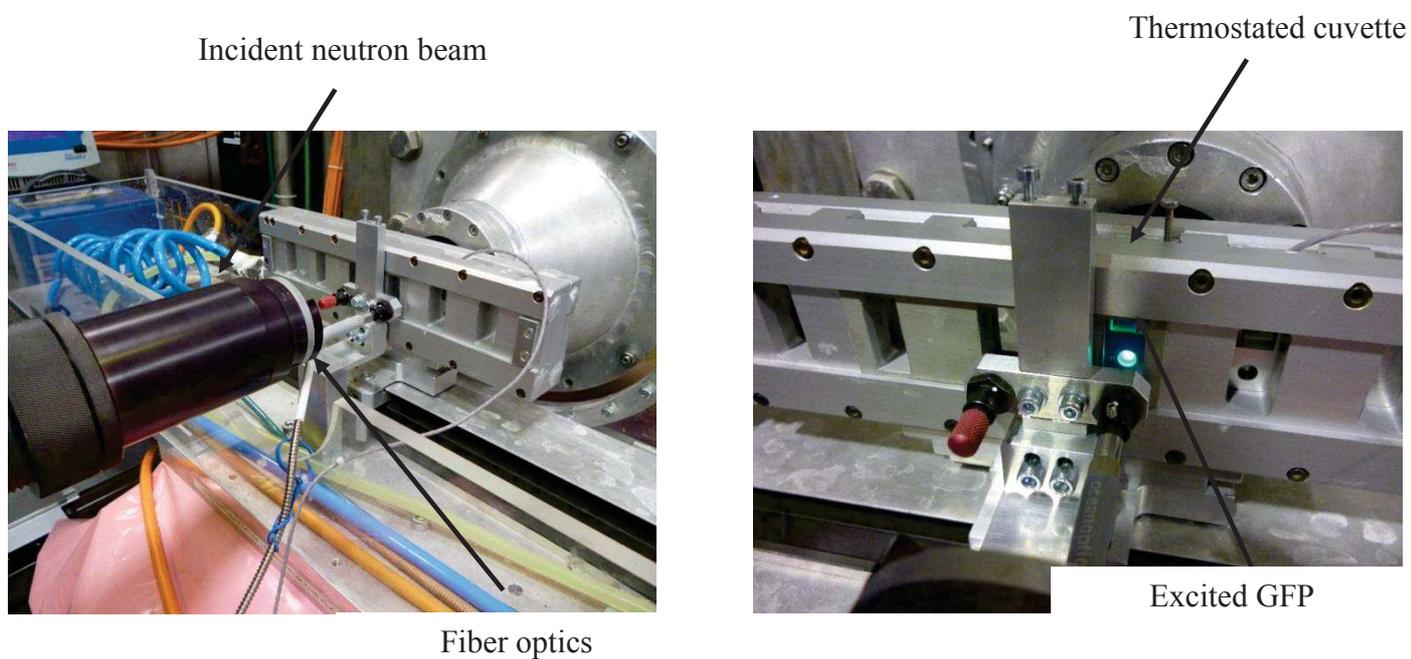


Figure S8: SANS/fluorimeter setup mounted on the D22 beamline at the Institut Laue Langevin (ILL, Grenoble, France)

The neutron beam, excitation and emission fluorescence fibers and the SANS cuvette containing GFPssrA are shown.

| | R_g (Å) | D_{max} (Å) | $I(0)$ (cm ⁻¹) | Model volume (Å ³) |
|---------------------------------|-----------|---------------|----------------------------|--------------------------------|
| dGFP/hPAN (ATP, 55°C) | | | | |
| 30s | 21 ± 0.7 | 75 ± 5 | 262 ± 6 | 35000 |
| 1min30s | 26 ± 1.0 | 110 ± 5 | 315 ± 10 | 50000 |
| 2min30s | 31 ± 8.0 | 125 ± 5 | 406 ± 11 | 76000 |
| 3min30s | 35 ± 0.6 | 145 ± 5 | 494 ± 13 | 101000 |
| 4min30s | 37 ± 0.6 | 155 ± 5 | 572 ± 14 | 128000 |
| 50min | 41 ± 0.6 | 175 ± 5 | 608 ± 15 | 165000 |
| dGFP (55°C) | | | | |
| 30s | 17 ± 0.6 | 60 ± 5 | 227 ± 5 | 36000 |
| 1min30s | 17 ± 0.6 | 60 ± 5 | 227 ± 5 | 37000 |
| 2min30s | 17 ± 0.6 | 60 ± 5 | 223 ± 4 | 36000 |
| 3min30s | 17 ± 0.6 | 60 ± 5 | 224 ± 4 | 32000 |
| 4min30s | 18 ± 0.6 | 60 ± 5 | 228 ± 5 | 32000 |
| 50min | 19 ± 0.6 | 60 ± 5 | 241 ± 5 | 34000 |
| dGFP/hPAN (no ATP, 55°C) | | | | |
| 30s | 19 ± 1.0 | 60 ± 5 | 0.1 ± 0.005 | ~36000 |
| 1min30s | 19 ± 0.7 | 60 ± 5 | 0.1 ± 0.030 | ~36000 |
| 2min30s | 21 ± 0.8 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 3min30s | 19 ± 0.9 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 4min30s | 20 ± 0.8 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 50min | 19 ± 1.3 | 60 ± 5 | 0.1 ± 0.006 | ~36000 |
| dGFP/hPAN (ATPγS, 55°C) | | | | |
| 30s | 18 ± 0.6 | 60 ± 5 | 0.2 ± 0.004 | ~36000 |
| 1min30s | 18 ± 0.8 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 2min30s | 19 ± 1.0 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 3min30s | 19 ± 0.8 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| 4min30s | 19 ± 0.6 | 60 ± 5 | 0.1 ± 0.003 | ~36000 |
| 50min | 20 ± 0.7 | 60 ± 5 | 0.1 ± 0.004 | ~36000 |
| dGFP (thermal unfolding) | | | | |
| 30s (55°C) | 18 ± 0.9 | 60 ± 5 | 79 ± 2 | 35800 |
| 5min30s (72°C) | 18 ± 0.6 | 60 ± 5 | 77 ± 2 | 32300 |
| 11min30s (87°C) | 31 ± 0.7 | 125 ± 5 | 125 ± 3 | 61600 |
| 12min30s (89°C) | 40 ± 0.5 | 135 ± 5 | 216 ± 5 | 82400 |
| 13min30s (91°C) | 53 ± 0.4 | 165 ± 5 | 466 ± 8 | 224000 |
| 15min30s (92°C) | 54 ± 0.3 | 165 ± 5 | 564 ± 10 | 764000 |
| dPAN/hGFP (ATP, 60°C) | | | | |
| 30s | 68 ± 3.0 | 220 ± 5 | 2.0 ± 0.05 | 660000 |
| 4min30s | 61 ± 3.0 | 190 ± 5 | 1.5 ± 0.04 | 500000 |
| 50min | 68 ± 2.0 | 220 ± 5 | 1.7 ± 0.04 | 600000 |
| dPAN/hGFP (noATP, 60°C) | | | | |
| 30s | 78 ± 2.0 | 240 ± 5 | 2.7 ± 0.06 | / |
| 1min | 79 ± 1.5 | 240 ± 5 | 2.8 ± 0.05 | / |
| dPAN/hGFP (ATPγS, 60°C) | | | | |
| 30s | 75 ± 2 | 235 ± 5 | 3 ± 0.07 | / |
| 1min | 75 ± 2 | 235 ± 5 | 3 ± 0.06 | / |

Table S1: SANS parameters of the GFPssrA and PAN calculated from the SANS data recorded during all the executed experiments. The radius of gyration (R_g) and the $I(0)$ intensity are extracted from the Guinier plots. The maximum dimension (D_{max}) is extracted from the $P(r)$ function by indirect Fourier transform using GNOM. The volume of each model is calculated by the program DAMMIN.

5.2.3 Discussion et perspectives

Le travail présenté dans cette partie de la thèse complète les études structurales en mode statique du complexe PAN de *P. horikoshii* présenté dans le chapitre précédent. Malgré toutes les données structurales précieuses fournies par cryo-microscopie électronique combinée avec de la cristallographie et SAXS, l'aspect dynamique et mécanistique en temps réel du mode d'action de PAN a nécessité des études biophysiques innovatrices en solution pour suivre « en direct » le processus de dépliement des substrats protéiques par le complexe PAN.

Le choix de la diffusion de neutrons aux petits angles (SANS) comme technique principale avec le court temps d'exposition de 30 secondes, combinée avec de la deutération et le changement de contraste entre les partenaires protéiques (PAN et son substrat GFPssrA) a fourni plusieurs informations sur le processus de dépliement au niveau des deux partenaires : la GFPssrA semble être dépliée par PAN d'une façon directionnelle longitudinale et relâchée rapidement en solution avec une échelle de temps de quelques secondes (Fig. 2 et 3, chapitre 5.2), tandis que le changement conformationnel principal de PAN concomitant est une contraction globale, induite par hydrolyse d'ATP et révélée par une diminution de son rayon de giration et de sa dimension maximale (Fig. 4 et 5, chapitre 5.2). Une fois l'ATP est hydrolysé et le substrat est déplié, PAN a été capable de se relaxer de nouveau pour atteindre sa conformation initiale non-contractée. En se basant sur tout ce qui a été déjà publié sur le mode d'action des complexes AAA+ unfoldases bactériens, et en particulier les systèmes 'Clp'(Sen et al, 2013), il est possible que PAN fonctionne avec un mécanisme similaire où l'hydrolyse de l'ATP alimente une rotation des sous-unités de l'héxamère et une contraction globale de la cavité centrale permettant aux résidus aromatiques et hydrophobe du « pore 1 loop » (Ar- Φ loop) de tirer progressivement le substrat protéique en le dépliant et en le translocant vers la particule protéolytique. Par contre, l'ampleur des changements conformationnels de plus de 10% du rayon de giration et de la dimension maximale de PAN suggèrent que, au-delà des mouvements rotationnels, des mouvements translationnels pourraient être impliqués.

Concernant les aspects méthodologiques de cette étude, nous avons pu développer plusieurs points au niveau de la technique: le couplage des mesures SANS avec de la spectroscopie de fluorescence « online » *in situ* a apporté plusieurs avantages permettant de corréler les informations structurales des courbes SANS avec les changements du signal de fluorescence de la GFP pendant le processus de dépliement aboutissant à une interprétation plus fidèle de nos données. En outre, le court temps d'exposition de 30 secondes a également permis de

suivre la réaction enzymatique avec une échelle de temps plus proche des processus biologiques. La puissance de la ligne D22 à l'ILL surtout au niveau du flux de neutrons incidents sur l'échantillon et la collimation du faisceau a beaucoup aidé à exécuter nos expériences avec ces conditions: faibles concentrations ($\sim 2\text{mg/ml}$), protéine de petite taille (GFPssrA, $\sim 27\text{ kDa}$), court temps d'exposition (30 s). En effet, c'est la première fois où une étude en temps réel par SANS est réalisée sur un système biologique et les résultats obtenus sont encourageantes au niveau de l'applicabilité technique de cette approche sur d'autres machines moléculaires impliquées dans divers processus biologiques.

6 CONCLUSIONS GENERALES ET PERSPECTIVES

Les deux objectifs principaux de cette thèse étaient tout d'abord de lancer une étude structurale et fonctionnelle sur le complexe PAN hexamérique régulateur du protéasome archéen dans le but d'apporter un maximum d'informations au niveau de la relation structure-fonction de ce système et de son homologue eucaryote qui reste relativement flou à cause du manque des données structurales à haute résolution. Un deuxième objectif était de développer et de tester une nouvelle approche en solution pour étudier la cinétique du processus de dépliement des protéines par les complexes AAA+ ATPases en temps réel par diffusion neutronique et de comprendre la dynamique fonctionnelle du complexe PAN induite par liaison et hydrolyse d'ATP et nécessaire pour le traitement des substrats protéiques. Les résultats obtenus au cours de ce projet (avec un manuscrit soumis et un autre en cours de soumission, voir chapitres 5.1 et 5.2) ont répondu à des questions biophysiques sur ces systèmes et ouvert la porte à plusieurs perspectives dans le cadre de l'étude structurale et dynamique du système PAN et de la machinerie protéolytique archéenne (SAMP-PAN-20S). Finalement, ils ont contribué au niveau du développement des approches cinétiques structurales pour les études en temps réel d'une large gamme de processus biologiques.

Les résultats obtenus au niveau de la caractérisation structurale du *Ph*PAN ont révélé la structure du complexe avec une résolution d'environ 13 Å par cryo-microscopie électronique. La structure présentée est remarquablement différente des autres modèles à basse résolution et pseudo-atomiques proposés à ce jour. En outre, et par comparaison avec les autres structures à haute résolution qui existent pour d'autres systèmes AAA+ unfoldase, comme les systèmes ClpX et HslU bactériens, PAN semble être également différente dans l'architecture globale, surtout au niveau de la large cavité à l'intérieur de la molécule et l'extension des bras de son domaine C-terminal qui n'existent pas chez les AAA+ unfoldases bactériens. La combinaison des différentes approches structurales utilisées (Cryo-EM, cristallographie et SAXS) nous a permis de générer un modèle pseudo-atomique du complexe et de confirmer les deux structures en solution. Contrairement aux homologues bactériens, les structures proposées pour PAN montrent une grande homologie avec la base ATPase du 19S eucaryote. En superposant le modèle pseudo-atomique du *Ph*PAN avec la carte de la base du 19S (Fig. 5, chapitre 5.1) on voit clairement la grande similarité au niveau des domaines « coiled-coil » du côté N-terminal de chaque molécule ainsi qu'au niveau de l'architecture en double « ring » (domaine ATPase et domaine OB) partagée entre le complexe PAN et les Rpts eucaryotes suggérant un mécanisme similaire pour le dépliement et la préparation des protéines substrats

avant leur translocation dans la particule catalytique 20S pour la dégradation. En plus, des informations structurales sur *Ph*PAN et l'étude des différentes conformations induites par la présence ou l'absence de nucléotide (ATP γ S) ont montré une contraction globale de la molécule, surtout au niveau de la cavité centrale et du domaine C-terminal responsable de l'interaction avec le protéasome 20S. Cette diminution du diamètre de l'anneau C-terminal du complexe pourra expliquer le rôle de la liaison et l'hydrolyse du nucléotide dans l'interaction entre PAN et son partenaire catalytique en plus du rôle que l'ATP joue dans le dépliement des substrats protéiques.

L'étude par diffusion neutronique en temps réel du dépliement du substrat par PAN est, selon notre connaissance, l'une des rares études qui a suivi en solution des altérations structurales en temps réel pendant une activité de dépliement ou de repliement de protéines substrats par des machines énergie-dépendantes de remodelage de protéines. Les résultats obtenus durant cette étude ont éclairci plusieurs aspects des mécanismes moléculaires détaillés impliqués dans la reconnaissance, la translocation et le dépliement du substrat par les AAA+ protéases. Concernant le substrat, la GFP_{ssrA} semble être tirée par PAN d'une façon directionnelle vers la cavité centrale du complexe aboutissant au dépliement progressif de la protéine. Vu l'absence de la particule catalytique 20S dans le cas de nos expériences SANS, une fois la GFP_{ssrA} était dépliée elle a été relâchée en solution pour former des larges agrégats. D'un autre côté, cette étude a également répondu à des questions sur comment les sous-unités de l'anneau AAA+ de PAN coordonnent la liaison et l'hydrolyse de l'ATP pour conduire les changements de conformation nécessaires pour effectuer le travail mécanique de dépliement. En effet, l'analyse de nos données SANS mesurées pendant le dépliement du substrat a montré que PAN adopte une contraction globale, traduite par une diminution de son rayon de giration et de sa dimension maximale. Cette contraction est essentiellement induite par liaison et hydrolyse de l'ATP et semble être le mouvement mécanique principal responsable du dépilement du substrat. La relaxation du complexe à la fin de la réaction pour atteindre sa conformation initiale a montré que PAN est très dynamique comme système et capable d'exécuter plusieurs cycles d'hydrolyse d'ATP pour le dépliement d'un grand nombre de substrat. La contraction du système PAN pendant le dépliement du substrat révélée par SANS est aussi en corrélation avec les deux modèles obtenus par cryo-microscopie électronique en absence et en présence d'ATP γ S. L'ensemble de ces données nous amènent à proposer un mécanisme d'action du complexe PAN en particulier et des autres systèmes AAA+ unfoldases en général. Ce mécanisme suggère un mouvement de pompage péristaltique induit par

l'hydrolyse d'ATP qui entraîne une translocation directionnelle du substrat à travers le pore et la cavité centrale de PAN pour faire passer la protéine déliée dans la particule catalytique.

Ce projet ouvre de nouvelles perspectives intéressantes dans l'étude des systèmes protéolytiques archéens ainsi que dans les études mécanistiques et structurales des AAA+ protéolytiques en utilisant le système PAN comme système modèle. Pour son étude structurale, des modèles cryo-microscopie électronique à une meilleure résolution sont attendus en collaboration avec l'équipe microscopie électronique de l'IBS. Une autre perspective sera d'identifier et de caractériser au niveau biochimique et structural des partenaires du système PAN homologues aux sous-unités Rpn du couvercle du complexe 19S eucaryote dont le but est d'assembler un système 19S-like régulateur du protéasome 20S chez les archées. Une troisième perspective sera de 'sampler' (voir chapitre 1.4.3) *in vitro* des substrats naturel de PAN de chez les archées pour étudier leur mécanisme de dépliement et de dégradation pour le système PAN-20S ainsi que le rôle du système 'SAMP' dans l'adressage des protéines substrats vers PAN.

Concernant les études biophysiques de la dynamique et du mode d'action des AAA+ unfoldases une étude sera lancée prochainement par RMN et SANS pour mieux comprendre avec plus en détail le mécanisme de dépliement ainsi que pour identifier plus spécifiquement les résidus du complexe PAN impliqués dans la reconnaissance, l'interaction et le dépliement des substrats protéiques. Actuellement nous sommes en train de purifier et d'assembler le protéasome 20S de *M. jannaschii* qui sera également utilisé dans le cadre de nos études biochimiques et biophysiques sur le système PAN.

Finalement, en ce qui concerne l'approche SANS en temps réel utilisée, il y a aussi des perspectives à réaliser surtout au niveau du temps d'exposition et la possibilité d'utiliser un temps encore plus court (de l'ordre de 5-10 secondes) pour suivre des mécanismes biologiques avec une résolution plus fine et suffisante de « trapper » les complexes étudiés dans plusieurs états conformationnel surtout au début de leur activité suffisant pour déchiffrer leurs mécanismes fonctionnels.

7 REFERENCES BIBLIOGRAPHIQUES

A L Goldberg a, John ACS (1976) Intracellular Protein Degradation in Mammalian and Bacterial Cells: Part 2. *Annual review of biochemistry* **45**: 747-804

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography* **68**: 352-367

Akopian TN, Kisselev AF, Goldberg AL (1997) Processive Degradation of Proteins and Other Catalytic Properties of the Proteasome from *Thermoplasma acidophilum*. *Journal of Biological Chemistry* **272**: 1791-1798

Amerik AY, Hochstrasser M (2004) Mechanism and function of deubiquitinating enzymes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1695**: 189-207

Ammelburg M, Frickey T, Lupas AN (2006) Classification of AAA+ proteins. *Journal of Structural Biology* **156**: 2-11

Anjum RS, Bray SM, Blackwood JK, Kilkenny ML, Coelho MA, Foster BM, Li S, Howard JA, Pellegrini L, Albers S-V, Deery MJ, Robinson NP (2015) Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon *Sulfolobus acidocaldarius*. *Nat Commun* **6**

Armengaud J, Fernandez B, Chaumont V, Rollin-Genetet F, Finet S, Marchetti C, Myllykallio H, Vidaud C, Pellequer J-L, Gribaldo S, Forterre P, Gans P (2003) Identification, Purification, and Characterization of an Eukaryotic-like Phosphopantetheine Adenylyltransferase (Coenzyme A Biosynthetic Pathway) in the Hyperthermophilic Archaeon *Pyrococcus abyssi*. *Journal of Biological Chemistry* **278**: 31078-31087

Asano S, Fukuda Y, Beck F, Aufderheide A, Förster F, Danev R, Baumeister W (2015) A molecular census of 26S proteasomes in intact neurons. *science* **347**: 439-442

Aubin-Tam M-E, Olivares Adrian O, Sauer Robert T, Baker Tania A, Lang Matthew J (2011) Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell* **145**: 257-267

Azim MK, Goehring W, Song HK, Ramachandran R, Bochtler M, Goettig P (2005) Characterization of the HslU chaperone affinity for HslV protease. *Protein Science : A Publication of the Protein Society* **14**: 1357-1362

Bajorek M, Finley D, Glickman MH (2003) Proteasome Disassembly and Downregulation Is Correlated with Viability during Stationary Phase. *Current Biology* **13**: 1140-1144

Baker TA, Sauer RT (2006) ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends in Biochemical Sciences* **31**: 647-653

Baptiste E, Susko E, Leigh J, MacLeod D, Charlebois R, Doolittle W (2005) Do orthologous gene phylogenies really support tree-thinking? *BMC Evolutionary Biology* **5**: 1-10

- Bar-Nun S, Glickman MH (2012) Proteasomal AAA-ATPases: Structure and function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**: 67-82
- Barthelme D, Chen JZ, Grabenstatter J, Baker TA, Sauer RT (2014) Architecture and assembly of the archaeal Cdc48-20S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E1687-E1694
- Barthelme D, Sauer RT (2012) Identification of the Cdc48•20S Proteasome as an Ancient AAA+ Proteolytic Machine. *Science (New York, NY)* **337**: 843-846
- Baugh JM, Pilipenko EV (2004) 20S Proteasome Differentially Alters Translation of Different mRNAs via the Cleavage of eIF4F and eIF3. *Molecular Cell* **16**: 575-586
- Baumeister W, Walz J, Zühl F, Seemüller E (1998) The Proteasome: Paradigm of a Self-Compartmentalizing Protease. *Cell* **92**: 367-380
- Baytshtok V, Baker TA, Sauer RT (2015) Assaying the kinetics of protein denaturation catalyzed by AAA+ unfolding machines and proteases. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 5377-5382
- Beal RE, Toscano-Cantaffa D, Young P, Rechsteiner M, Pickart CM (1998) The Hydrophobic Effect Contributes to Polyubiquitin Chain Recognition. *Biochemistry* **37**: 2925-2934
- Beck F, Unverdorben P, Bohn S, Schweitzer A, Pfeifer G, Sakata E, Nickell S, Plitzko JM, Villa E, Baumeister W, Förster F (2012) Near-atomic resolution structural model of the yeast 26S proteasome. *Proceedings of the National Academy of Sciences* **109**: 14870-14875
- Bell SD, Magill CP, Jackson SP (2001) Basal and regulated transcription in Archaea. *Biochemical Society Transactions* **29**: 392-395
- Benaroudj N, Goldberg AL (2000) PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nat Cell Biol* **2**: 833-839
- Benaroudj N, Goldberg, A. L., (2000) PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *NATURE CELL BIOLOGY* **2**: 833-839
- Benaroudj N, Smith D, Goldberg AL (2007) What the Archaeal PAN-Proteasome Complex and Bacterial ATP-Dependent Proteases Can Teach Us About the 26S Proteasome. In *Protein Degradation Series*, pp 215-247. Wiley-VCH Verlag GmbH & Co. KGaA
- Benaroudj N, Zwickl P, Seemüller E, Baumeister W, Goldberg AL (2003) ATP Hydrolysis by the Proteasome Regulatory Complex PAN Serves Multiple Functions in Protein Degradation. *Molecular Cell* **11**: 69-78
- Besche H, Zwickl P (2004) The Thermoplasma acidophilum Lon protease has a Ser-Lys dyad active site. *European Journal of Biochemistry* **271**: 4361-4365
- Besche HC, Haas W, Gygi SP, Goldberg AL (2009) Isolation of Mammalian 26S Proteasomes and p97/VCP Complexes Using the Ubiquitin-like Domain from HHR23B Reveals Novel Proteasome-Associated Proteins(). *Biochemistry* **48**: 10.1021/bi802198q

Bochtler M, Hartmann C, Song HK, Bourenkov GP, Bartunik HD, Huber R (2000) The structures of HslU and the ATP-dependent protease HslU-HslV. *NATURE* **403**: 800-805

Bohn S, Beck F, Sakata E, Walzthoeni T, Beck M, Aebersold R, Förster F, Baumeister W, Nickell S (2010) Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 20992-20997

Braun BC, Glickman M, Kraft R, Dahlmann B, Kloetzel P-M, Finley D, Schmidt M (1999) The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nat Cell Biol* **1**: 221-226

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb J-F, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk H-P, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *science* **273**: 1058-1073

Burns KE, Darwin KH (2010) Pupylation vs. Ubiquitylation: Tagging for Proteasome-Dependent Degradation. *Cellular microbiology* **12**: 424-431

Burroughs AM, Iyer LM, Aravind L (2009) Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer and ubiquitin conjugation. *Proteins* **75**: 895-910

Chadwick J (1932) The Existence of a Neutron. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **136**: 692-708

Chamieh H (2005) Characterisation of Pan A and Pan B : two proteasomal regulatory proteins from the halophilic archaeon. Université Joseph-Fourier - Grenoble I,

Chamieh H, Guetta D, Franzetti B (2008) The two PAN ATPases from *Halobacterium* display N-terminal heterogeneity and form labile complexes with the 20S proteasome. *Biochemical Journal* **411**: 387-397

Chamieh H, Marty V, Guetta D, Perollier A, Franzetti B (2012) Stress regulation of the PAN–proteasome system in the extreme halophilic archaeon *Halobacterium*. *Extremophiles* **16**: 215-225

Chen B, Sysoeva TA, Chowdhury S, Guo L, De Carlo S, Hanson JA, Yang H, Nixon BT (2010) Engagement of Arginine Finger to ATP Triggers Large Conformational Changes in NtrC1 AAA+ ATPase For Remodeling Bacterial RNA Polymerase. *Structure (London, England : 1993)* **18**: 1420-1430

Chen X, Chi Y, Bloecher A, Aebersold R, Clurman BE, Roberts JM (2004) N-Acetylation and Ubiquitin-Independent Proteasomal Degradation of p21Cip1. *Molecular Cell* **16**: 839-847

Chung CH, Goldberg AL (1981) The product of the lon (capR) gene in *Escherichia coli* is the ATP-dependent protease, protease La. *Proceedings of the National Academy of Sciences of the United States of America* **78**: 4931-4935

Ciechanover A (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* **6**: 79-87

- Ciechanover A, Brundin P (2003) The Ubiquitin Proteasome System in Neurodegenerative Diseases: Sometimes the Chicken, Sometimes the Egg. *Neuron* **40**: 427-446
- Ciechanover A, Gonen H, Bercovich B, Cohen S, Fajerman I, Israël A, Mercurio F, Kahana C, Schwartz AL, Iwai K, Orian A (2001) Mechanisms of ubiquitin-mediated, limited processing of the NF- κ B1 precursor protein p105. *Biochimie* **83**: 341-349
- Confalonieri F, Duguet M (1995) A 200-amino acid ATPase module in search of a basic function. *BioEssays* **17**: 639-650
- Cordova Juan C, Olivares Adrian O, Shin Y, Stinson Benjamin M, Calmat S, Schmitz Karl R, Aubin-Tam M-E, Baker Tania A, Lang Matthew J, Sauer Robert T (2014) Stochastic but Highly Coordinated Protein Unfolding and Translocation by the ClpXP Proteolytic Machine. *Cell* **158**: 647-658
- Coux O, Tanaka K, Goldberg AL (1996) Structure and Functions of the 20S and 26S Proteasomes. *Annual review of biochemistry* **65**: 801-847
- Čuboňová Lu, Sandman K, Hallam SJ, DeLong EF, Reeve JN (2005) Histones in Crenarchaea. *JOURNAL OF BACTERIOLOGY* **187**: 5482-5485
- da Fonseca Paula CA, He J, Morris Edward P (2012) Molecular Model of the Human 26S Proteasome. *Molecular Cell* **46**: 54-66
- Day ES, Capili AD, Borysenko CW, Zafari M, Whitty A (2013) Determining the affinity and stoichiometry of interactions between unmodified proteins in solution using Biacore. *Analytical Biochemistry* **440**: 96-107
- De Mot R, Nagy I, Walz J, Baumeister W (1999) Proteasomes and other self-compartmentalizing proteases in prokaryotes. *Trends in Microbiology* **7**: 88-92
- Delano WL. (2002) The PyMOL Molecular Graphics System.
- Desmond E, Gribaldo S (2009) Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution* **1**: 364-381
- Dewhurst C (2002) GRASP. http://www.ill.fr/lss/grasp/grasp_main.html.
- Djuranovic S, Hartmann MD, Habeck M, Ursinus A, Zwickl P, Martin J, Lupas AN, Zeth K (2009) Structure and Activity of the N-Terminal Substrate Recognition Domains in Proteasomal ATPases. *Molecular Cell* **34**: 580-590
- Dmitri IS, Michel HJK (2003) Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* **66**: 1735
- Dougan DA, Reid BG, Horwich AL, Bukau B (2002) ClpS, a Substrate Modulator of the ClpAP Machine. *Molecular Cell* **9**: 673-683
- Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *NATURE* **440**: 623-630
- Eme L, Moreira D, Talla E, Brochier-Armanet C (2009) A Complex Cell Division Machinery Was Present in the Last Common Ancestor of Eukaryotes. *PLOS ONE* **4**: e5021

- Erauso G, Reysenbach A-L, Godfroy A, Meunier J-R, Crump B, Partensky F, Baross J, Marteinsson V, Barbier G, Pace N, Prieur D (1993) *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Archives of Microbiology* **160**: 338-349
- Erzberger JP, Berger JM (2006) EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS. *Annual Review of Biophysics and Biomolecular Structure* **35**: 93-114
- Finley D (2009) Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annual review of biochemistry* **78**: 477-513
- Finley D, Ulrich HD, Sommer T, Kaiser P (2012) The Ubiquitin-Proteasome System of *Saccharomyces cerevisiae*. *Genetics* **192**: 319-360
- Förster A, Masters EI, Whitby FG, Robinson H, Hill CP (2005) The 1.9 Å Structure of a Proteasome-11S Activator Complex and Implications for Proteasome-PAN/PA700 Interactions. *Molecular Cell* **18**: 589-599
- Förster A, Whitby FG, Hill CP (2003) The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *The EMBO Journal* **22**: 4356-4364
- Förster F, Lasker K, Beck F, Nickell S, Sali A, Baumeister W (2009) An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochemical and Biophysical Research Communications* **388**: 228-233
- Förster F, Lasker K, Nickell S, Sali A, Baumeister W (2010) Toward an Integrated Structural Model of the 26S Proteasome. *Molecular & Cellular Proteomics : MCP* **9**: 1666-1677
- Förster F, Schuller J, Unverdorben P, Aufderheide A (2014) Emerging Mechanistic Insights into AAA Complexes Regulating Proteasomal Degradation. *Biomolecules* **4**: 774
- Forterre P (2015) The universal tree of life: an update. *Frontiers in Microbiology* **6**: 717
- Forterre P, Gribaldo S, Gabelle D, Serre M-C (2007) Origin and evolution of DNA topoisomerases. *Biochimie* **89**: 427-446
- Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Structural Biology* **116**: 190-199
- Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**: 342-346
- Franke D, Svergun, D. I., (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**: 342-346
- Frankenberg R, Andersson M, Clark D (2003) Effect of temperature and pressure on the proteolytic specificity of the recombinant 20S proteasome from *Methanococcus jannaschii*. *Extremophiles* **7**: 353-360
- Frankland-Searby S, Bhaumik, S. R., (2012) The 26S Proteasome Complex: An Attractive Target for Cancer Therapy. *Biochim Biophys Acta* **1825**: 64-76

- Franzetti B, Schoehn G, Garcia D, Ruigrok RW, Zaccai G (2002) Characterization of the proteasome from the extremely halophilic archaeon *Haloarcula marismortui*. *Archaea* **1**: 53-61
- Frydman J (2001) Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annual review of biochemistry* **70**: 603-647
- Fu H, Doelling JH, Rubin DM, Vierstra RD (1999) Structural and functional analysis of the six regulatory particle triple-A ATPase subunits from the Arabidopsis 26S proteasome. *The Plant Journal* **18**: 529-539
- Gabel F (2012) Small Angle Neutron Scattering for the Structural Study of Intrinsically Disordered Proteins in Solution: A Practical Guide. In *Intrinsically Disordered Protein Analysis*, Uversky VN, Dunker AK (eds), Vol. 896, 8, pp 123-135. Springer New York
- Gerega A, Rockel B, Peters J, Tamura T, Baumeister W, Zwickl P (2005) VAT, the *Thermoplasma* Homolog of Mammalian p97/VCP, Is an N Domain-regulated Protein Unfoldase. *Journal of Biological Chemistry* **280**: 42856-42862
- Gerth U, Kirstein J, Mostertz J, Waldminghaus T, Miethke M, Kock H, Hecker M (2004) Fine-Tuning in Regulation of Clp Protein Content in *Bacillus subtilis*. *JOURNAL OF BACTERIOLOGY* **186**: 179-191
- Glickman MH, Ciechanover A (2002) The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiological Reviews* **82**: 373-428
- Glickman MH, Rubin DM, Fried VA, Finley D (1998) The Regulatory Particle of the *Saccharomyces cerevisiae* Proteasome. *Molecular and Cellular Biology* **18**: 3149-3162
- Glover JR, Lindquist S (1998) Hsp104, Hsp70, and Hsp40: A Novel Chaperone System that Rescues Previously Aggregated Proteins. *Cell* **94**: 73-82
- Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009) Crystal structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**: 744-756
- Godfroy A, Postec A, Raven N (2006) Growth of hyperthermophilic microorganisms for physiological and nutritional studies. In *Extremophiles*, pp 93-107. Elsevier
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 6661-6665
- Goldberg AL (1972) Degradation of Abnormal Proteins in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **69**: 422-426
- Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. *NATURE* **426**: 895-899
- Goldberg AL (2005) Nobel Committee Tags Ubiquitin for Distinction. *Neuron* **45**: 339-344
- Goldberg AL, Dice, J. F., (1974) INTRACELLULAR PROTEIN DEGRADATION IN MAMMALIAN AND BACTERIAL CELLS. *Annu Rev Biochem* **43**: 835-869

- González JM, Masuchi Y, Robb FT, Ammerman JW, Maeder DL, Yanagibayashi M, Tamaoka J, Kato C (1998) *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* **2**: 123-130
- Gorbea C, Taillandier D, Rechsteiner M (2000) 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
- Gordon C, McGurk G, Dillon P, Rosen C, Hastie ND (1993) Defective mitosis due to a mutation in the gene for a fission yeast 26S protease subunit. *NATURE* **366**: 355-357
- Gottesman S, Roche E, Zhou Y, Sauer RT (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *GENES & DEVELOPMENT* **12**: 1338-1347
- Graef M, Langer T (2006) Substrate specific consequences of central pore mutations in the i-AAA protease Yme1 on substrate engagement. *Journal of Structural Biology* **156**: 101-108
- Graham DE, Kyrpides N, Anderson IJ, Overbeek R, Whitman WB (2001) Genome of *Methanocaldococcus* (methanococcus) *jannaschii*. In *Methods in Enzymology* Vol. Volume 330, pp 40-123. Academic Press
- Gribaldo S, Brochier-Armanet C (2006) The origin and evolution of Archaea: a state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**: 1007-1022
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Micro* **8**: 743-752
- Grigoreva TA, Tribulovich VG, Garabadzhiu AV, Melino G, Barlev NA (2015) *The 26S proteasome is a multifaceted target for anti-cancer therapies*, Vol. 6.
- Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, Glickman MH, Finley D (2000) A gated channel into the proteasome core particle. *Nat Struct Mol Biol* **7**: 1062-1067
- Groll M, Berkers, C. R., Ploegh, H. L., Ovaa, H., (2006) Crystal Structure of the Boronic Acid-Based Proteasome Inhibitor Bortezomib in Complex with the Yeast 20S Proteasome. *Structure* **14**: 451-456
- Groll M, Clausen T (2003) Molecular shredders: how proteasomes fulfill their role. *Current Opinion in Structural Biology* **13**: 665-673
- Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R (1997) Structure of 20S proteasome from yeast at 2.4Å resolution. *NATURE* **386**: 463-471
- Guinier A (1939) *La Diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques*. Masson, Paris
- Guo F, Maurizi MR, Esser L, Xia D (2002) Crystal Structure of ClpA, an Hsp100 Chaperone and Regulator of ClpAP Protease. *Journal of Biological Chemistry* **277**: 46743-46752
- Gur E, Sauer RT (2008) Recognition of misfolded proteins by Lon, a AAA(+) protease. *GENES & DEVELOPMENT* **22**: 2267-2277

- Gur E, Sauer RT (2009) Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 18503-18508
- Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**: 519-529
- Hartl FU, Hayer-Hartl M (2002) Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *science* **295**: 1852-1858
- Hartung S, Hopfner K-P (2009) Lessons from structural and biochemical studies on the archaeal exosome. *Biochemical Society Transactions* **37**: 83-87
- Hecker M, Schumann W, Völker U (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Molecular Microbiology* **19**: 417-428
- Hei DJ, Clark DS (1994) Pressure Stabilization of Proteins from Extreme Thermophiles. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **60**: 932-939
- Hepowitz N, Uthandi S, Miranda HV, Toniutti M, Prunetti L, Olivarez O, De Vera IMS, Fanucci GE, Chen S, Maupin-Furlow JA (2012) Archaeal JAB1/MPN/MOV34 metalloenzyme (HvJAMM1) cleaves ubiquitin-like small archaeal modifier proteins (SAMPs) from protein-conjugates. *Molecular Microbiology* **86**: 971-987
- Herman C, Prakash S, Lu CZ, Matouschek A, Gross CA (2003) Lack of a Robust Unfoldase Activity Confers a Unique Level of Substrate Specificity to the Universal AAA Protease FtsH. *Molecular Cell* **11**: 659-669
- Hersch GL, Burton RE, Bolon DN, Baker TA, Sauer RT (2005) Asymmetric Interactions of ATP with the AAA+ ClpX6 Unfoldase: Allosteric Control of a Protein Machine. *Cell* **121**: 1017-1027
- Hershko A (1988) Ubiquitin-mediated protein degradation. *Journal of Biological Chemistry* **263**: 15237-15240
- Heymann JB, Cardone G, Winkler DC, Steven AC (2008) Computational Resources for Cryo-Electron Tomography in Bsoft. *Journal of Structural Biology* **161**: 232-242
- Hickey DA, Singer GAC (2004) Genomic and proteomic adaptations to growth at high temperature. *Genome Biology* **5**: 117-117
- Hoyt MA, Coffino P (2004) Ubiquitin-proteasome system. *CMLS, Cell Mol Life Sci* **61**: 1596-1600
- Hu G, Lin G, Wang M, Dick L, Xu R-M, Nathan C, Li H (2006) Structure of the Mycobacterium tuberculosis proteasome and mechanism of inhibition by a peptidyl boronate. *Molecular Microbiology* **59**: 1417-1428
- Huber EM, Groll M (2012) Inhibitors for the Immuno- and Constitutive Proteasome: Current and Future Trends in Drug Development. *Angewandte Chemie International Edition* **51**: 8708-8720
- Humbard MA, Miranda HV, Lim J-M, Krause DJ, Pritz JR, Zhou G, Chen S, Wells L, Maupin-Furlow JA (2010) Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *NATURE* **463**: 54-60

- Husnjak K, Elsasser S, Zhang N, Chen X, Randles L, Shi Y, Hofmann K, Walters K, Finley D, Dikic I (2008) Proteasome subunit Rpn13 is a novel ubiquitin receptor. *NATURE* **453**: 481-488
- Imanaka H, Yamatsu A, Fukui T, Atomi H, Imanaka T (2006) Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Molecular Microbiology* **61**: 898-909
- Inobe T, Genmei R (2015) N-Terminal Coiled-Coil Structure of ATPase Subunits of 26S Proteasome Is Crucial for Proteasome Function. *PLOS ONE* **10**: e0134056
- Iosefson O, Nager AR, Baker TA, Sauer RT (2015) Coordinated gripping of substrate by subunits of a AAA+ proteolytic machine. *Nat Chem Biol* **11**: 201-206
- Ishino Y, Komori K, Cann IKO, Koga Y (1998) A Novel DNA Polymerase Family Found in Archaea. *JOURNAL OF BACTERIOLOGY* **180**: 2232-2236
- Iyer LM, Leippe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology* **146**: 11-31
- Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science : A Publication of the Protein Society* **19**: 642-657
- Jacrot B (1976) The study of biological structures by neutron scattering from solution. *Reports on Progress in Physics* **39**: 911
- Jacrot B, Zaccai G (1981) Determination of molecular weight by neutron scattering. *Biopolymers* **20**: 2413-2426
- Jékely G (2003) Small GTPases and the evolution of the eukaryotic cell. *BioEssays* **25**: 1129-1138
- Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS (1983) *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Archives of Microbiology* **136**: 254-261
- Joshi SA, Hersch GL, Baker TA, Sauer RT (2004) Communication between ClpX and ClpP during substrate processing and degradation. *Nat Struct Mol Biol* **11**: 404-411
- Kabsch W (2010) XDS. *Acta Crystallographica Section D: Biological Crystallography* **66**: 125-132
- Karlsson R (2004) SPR for molecular interaction analysis: a review of emerging application areas. *Journal of Molecular Recognition* **17**: 151-161
- Karplus PA, Diederichs K (2012) Linking Crystallographic Model and Data Quality. *science* **336**: 1030-1033
- Karplus PA, Diederichs K (2015) Assessing and maximizing data quality in macromolecular crystallography. *Current Opinion in Structural Biology* **34**: 60-68
- Katsamba PS, Park S, Laird-Offringa IA (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* **26**: 95-104

- Katz Elijah J, Isasa M, Crosas B (2010) A new map to understand deubiquitination. *Biochemical Society Transactions* **38**: 21-28
- Keiler KC, Waller PRH, Sauer RT (1996) Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. *science* **271**: 990-993
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols* **10**: 845-858
- Kenniston JA, Baker TA, Fernandez JM, Sauer RT (2003) Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine. *Cell* **114**: 511-520
- Kenniston JA, Baker TA, Sauer RT (2005) Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 1390-1395
- Kessel M, Maurizi MR, Kim B, Kocsis E, Trus BL, Singh SK, Steven AC (1995) Homology in Structural Organization Between E. coli ClpAP Protease and the Eukaryotic 26 S Proteasome. *Journal of molecular biology* **250**: 587-594
- Kim Y-C, Li X, Thompson D, DeMartino GN (2013) ATP Binding by Proteasomal ATPases Regulates Cellular Assembly and Substrate-induced Functions of the 26 S Proteasome. *Journal of Biological Chemistry* **288**: 3334-3345
- Kim Y-C, Snoberger A, Schupp J, Smith DM (2015) ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. *Nat Commun* **6**
- Kim Y-I, Burton RE, Burton BM, Sauer RT, Baker TA (2000) Dynamics of Substrate Denaturation and Translocation by the ClpXP Degradation Machine. *Molecular Cell* **5**: 639-648
- Kimura Y, Tanaka K (2010) Regulatory mechanisms involved in the control of ubiquitin homeostasis. *Journal of Biochemistry* **147**: 793-798
- Kish-Trier E, Hill CP (2013) Structural Biology of the Proteasome. *Annual Review of Biophysics* **42**: 29-49
- Klemes Y, Etlinger JD, Goldberg AL (1981) Properties of abnormal proteins degraded rapidly in reticulocytes. Intracellular aggregation of the globin molecules prior to hydrolysis. *Journal of Biological Chemistry* **256**: 8436-8444
- Köhler A, Cascio P, Leggett DS, Woo KM, Goldberg AL, Finley D (2001) The Axial Channel of the Proteasome Core Particle Is Gated by the Rpt2 ATPase and Controls Both Substrate Entry and Product Release. *Molecular Cell* **7**: 1143-1152
- Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H (1994) A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 9223-9227
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**: 1277-1282

König H (2001) Archaeal Cell Walls. In *eLS*. John Wiley & Sons, Ltd

Koodathingal P, Jaffe NE, Kraut DA, Prakash S, Fishbain S, Herman C, Matouschek A (2009) ATP-dependent Proteases Differ Substantially in Their Ability to Unfold Globular Proteins. *Journal of Biological Chemistry* **284**: 18674-18684

Koonin EV (1993) A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Research* **21**: 2541-2547

Kress W, Mutschler H, Weber-Ban E (2009) Both ATPase Domains of ClpA Are Critical for Processing of Stable Protein Structures. *Journal of Biological Chemistry* **284**: 31441-31452

Krüger E, Witt E, Ohlmeier S, Hanschke R, Hecker M (2000) The Clp Proteases of *Bacillus subtilis* Are Directly Involved in Degradation of Misfolded Proteins. *JOURNAL OF BACTERIOLOGY* **182**: 3259-3265

Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, Martin A (2012) Complete subunit architecture of the proteasome regulatory particle. *NATURE* **482**: 186-191

Larsen CN, Finley D Protein Translocation Channels in the Proteasome and Other Proteases. *Cell* **91**: 431-434

Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, Baumeister W (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences* **109**: 1380-1387

Latterich M, Patel S (1998) The AAA team: related ATPases with diverse functions. *Trends in Cell Biology* **8**: 65-71

Lecompte O, Ripp R, Thierry J-C, Moras D, Poch O (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Research* **30**: 5382-5390

Lee Y-Y, Chang C-F, Kuo C-L, Chen M-C, Yu CH, Lin P-I, Wu WF (2003) Subunit Oligomerization and Substrate Recognition of the *Escherichia coli* ClpYQ (HslUV) Protease Implicated by In Vivo Protein-Protein Interactions in the Yeast Two-Hybrid System. *JOURNAL OF BACTERIOLOGY* **185**: 2393-2401

Leigh JA, Albers S-V, Atomi H, Allers T (2011) Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiology Reviews* **35**: 577-608

Leipe D, Koonin E, Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. *Journal of molecular biology* **333**: 781-815

Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? *Nucleic Acids Research* **27**: 3389-3401

Levchenko I, Luo L, Baker TA (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *GENES & DEVELOPMENT* **9**: 2399-2408

Lindås A-C, Karlsson EA, Lindgren MT, Ettema TJG, Bernander R (2008) A unique cell division machinery in the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 18942-18946

- Lindner P, Zemb, Th., (2002) Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter. *Elsevier ed*
- Lipford JR, Deshaies RJ (2003) Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation. *Nat Cell Biol* **5**: 845-850
- Londei P (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiology Reviews* **29**: 185-200
- Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *science* **268**: 533-539
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology* **128**: 82-97
- Lupas AN, Martin J (2002) AAA proteins. *Current Opinion in Structural Biology* **12**: 746-753
- Martin A, Baker TA, Sauer RT (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *NATURE* **437**: 1115-1120
- Martin A, Baker TA, Sauer RT (2008) Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Molecular Cell* **29**: 441-450
- Matouschek A (2003) Protein unfolding — an important process in vivo? *Current Opinion in Structural Biology* **13**: 98-109
- Matyskiela ME, Lander GC, Martin A (2013) Conformational switching of the 26S proteasome enables substrate degradation. *Nat Struct Mol Biol* **20**: 781-788
- Maupin-Furlow J (2011) Proteasomes and protein conjugation across domains of life. *Nature Reviews Microbiology* **10**: 100-111
- Maupin-Furlow JA (2013) Ubiquitin-like proteins and their roles in archaea. *Trends in Microbiology* **21**: 31-38
- Maupin-Furlow W, H. L., Kaczowka, S. J., Ou, M. S., (2000) PROTEASOMES IN THE ARCHAEA: FROM STRUCTURE TO FUNCTION. *Frontiers in Bioscience* **5**: 837-865
- Maupin-Furlow JA, Humbard MA, Kirkland PA, Li W, Reuter CJ, Wright AJ, Zhou G (2006) Proteasomes from Structure to Function: Perspectives from Archaea. In *Current Topics in Developmental Biology* Vol. Volume 75, pp 125-169. Academic Press
- McCoy AJ, Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J., (2007) Phaser crystallographic software. *J Appl Cryst* **40**: 658–674
- Medalia N, Beer A, Zwickl P, Mihalache O, Beck M, Medalia O, Navon A (2009) Architecture and Molecular Mechanism of PAN, the Archaeal Proteasome Regulatory ATPase. *Journal of Biological Chemistry* **284**: 22952-22960
- Michels PC, Clark DS (1997) Pressure-Enhanced Activity and Stability of a Hyperthermophilic Protease from a Deep-Sea Methanogen. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **63**: 3985-3991

Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology* **142**: 334-347

Miranda HV, Nembhard N, Su D, Hepowitz N, Krause DJ, Pritz JR, Phillips C, Söll D, Maupin-Furlow JA (2011) E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *Proceedings of the National Academy of Sciences* **108**: 4417-4422

Mylonas E, Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography* **40**: s245-s249

Nakamura TM, Cech TR Reversing Time: Origin of Telomerase. *Cell* **92**: 587-590

Nakamura Y, Umehara T, Tanaka A, Horikoshi M, Padmanabhan B, Yokoyama S (2007) Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. *Biochemical and Biophysical Research Communications* **359**: 503-509

Namboori SC, Graham DE (2008) Acetamido Sugar Biosynthesis in the Euryarchaea. *JOURNAL OF BACTERIOLOGY* **190**: 2987-2996

Navon A, Ciechanover A (2009) The 26 S Proteasome: From Basic Mechanisms to Drug Targeting. *Journal of Biological Chemistry* **284**: 33713-33718

Navon A, Goldberg AL (2001) Proteins Are Unfolded on the Surface of the ATPase Ring before Transport into the Proteasome. *Molecular Cell* **8**: 1339-1349

Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. *Genome Research* **9**: 27-43

Ng SYM, Zolghadr B, Driessen AJM, Albers S-V, Jarrell KF (2008) Cell Surface Structures of Archaea. *JOURNAL OF BACTERIOLOGY* **190**: 6039-6047

Nyquist K, Martin A (2014) Marching to the beat of the ring: polypeptide translocation by AAA+ proteases. *Trends in Biochemical Sciences* **39**: 53-60

Ogura T, Wilkinson AJ (2001) AAA+ superfamily ATPases: common structure—diverse function. *Genes to Cells* **6**: 575-597

Olivares AO, Baker TA, Sauer RT (2016) Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat Rev Micro* **14**: 33-44

Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP (2000) Homologs of Small Nucleolar RNAs in Archaea. *science* **288**: 517-522

Orange F (2008) Experimental silicification of microorganisms. Applications in the search for evidence of life in early Earth and extraterrestrial rocks. Université d'Orléans,

Pack C-G, Yukii H, Toh-e A, Kudo T, Tsuchiya H, Kaiho A, Sakata E, Murata S, Yokosawa H, Sako Y, Baumeister W, Tanaka K, Saeki Y (2014) Quantitative live-cell imaging reveals spatio-temporal dynamics and cytoplasmic assembly of the 26S proteasome. *Nat Commun* **5**

Park S, Myszka DG, Yu M, Littler SJ, Laird-Offringa IA (2000) HuD RNA Recognition Motifs Play Distinct Roles in the Formation of a Stable Complex with AU-Rich RNA. *Molecular and Cellular Biology* **20**: 4765-4772

Park SK, Kim KI, Woo KM, Seol JH, Tanaka K, Ichihara A, Ha DB, Chung CH (1993) Site-directed mutagenesis of the dual translational initiation sites of the clpB gene of *Escherichia coli* and characterization of its gene products. *Journal of Biological Chemistry* **268**: 20170-20174

Peretó J, López-García P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends in Biochemical Sciences* **29**: 469-477

Pernot P, Round A, Barrett R, De Maria Antolinos A, Gobbo A, Gordon E, Huet J, Kieffer J, Lentini M, Mattenet M, Morawe C, Mueller-Dieckmann C, Ohlsson S, Schmid W, Surr J, Theveneau P, Zerrad L, McSweeney S (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of Synchrotron Radiation* **20**: 660-664

Peters JM, Franke WW, Kleinschmidt JA (1994) Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *Journal of Biological Chemistry* **269**: 7709-7718

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**: 342-350

Petoukhov MV, Svergun DI (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Current Opinion in Structural Biology* **17**: 562-571

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**: 1605-1612

Pickart CM (1997) Targeting of substrates to the 26S proteasome. *The FASEB Journal* **11**: 1055-1066

Prakash S, Matouschek A (2004) Protein unfolding in the cell. *Trends in Biochemical Sciences* **29**: 593-600

Prouty WF, Goldberg, A. L., (1972) Fate of Abnormal Proteins in *E. coli* Accumulation in Intracellular Granules before Catabolism. *Nature* **240**: 147-150

Rabl J, Smith DM, Yu Y, Chang S-C, Goldberg AL, Cheng Y (2008) Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Molecular Cell* **30**: 360-368

Rajagopala SV, Sikorski P, Caufield JH, Tovchigrechko A, Uetz P (2012) Studying protein complexes by the yeast two-hybrid system. *Methods (San Diego, Calif)* **58**: 392-399

Ramesh MA, Malik S-B, Logsdon Jr JM (2005) A Phylogenomic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis. *Current Biology* **15**: 185-191

Raymann K, Brochier-Armanet C, Gribaldo S (2015) The two-domain tree of life is linked to a new root for the Archaea. *Proceedings of the National Academy of Sciences* **112**: 6670-6675

- Religa T, Sprangers R, Kay L (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science (New York)* **328**: 98-102
- Reuter CJ, Maupin-Furlow JA (2004) Analysis of Proteasome-Dependent Proteolysis in *Haloferax volcanii* Cells, Using Short-Lived Green Fluorescent Proteins. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **70**: 7530-7538
- Rice SA (1956) Small angle scattering of X-rays. A. Guinier and G. Fournet. Translated by C. B. Wilson and with a bibliographical appendix by K. L. Yudowitch. Wiley, New York, 1955. 268 pp. \$7.50. *Journal of Polymer Science* **19**: 594-594
- Richmond C, Gorbea C, Rechsteiner M (1997) Specific Interactions between ATPase Subunits of the 26 S Protease. *Journal of Biological Chemistry* **272**: 13403-13411
- Rosenzweig R, Osmulski PA, Gaczynska M, Glickman MH (2008) The central unit within the 19S regulatory particle of the proteasome. *nature structural & molecular biology* **15**: 573-580
- Rouiller I, Butel VM, Latterich M, Milligan RA, Wilson-Kubalek EM (2000) A Major Conformational Change in p97 AAA ATPase upon ATP Binding. *Molecular Cell* **6**: 1485-1490
- Round A, Brown E, Marcellin R, Kapp U, Westfall CS, Jez JM, Zubieta C (2013) Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. *Acta Crystallographica Section D* **69**: 2072-2080
- Rubin DM, Glickman MH, Larsen CN, Dhruvakumar S, Finley D (1998) Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome. *The EMBO Journal* **17**: 4909-4919
- Ruepp A, Eckerskorn C, Bogyo M, Baumeister W (1998) Proteasome function is dispensable under normal but not under heat shock conditions in *Thermoplasma acidophilum*. *FEBS Letters* **425**: 87-90
- Russell SJ, Sathyanarayana UG, Johnston SA (1996) Isolation and Characterization of SUG2: A NOVEL ATPase FAMILY COMPONENT OF THE YEAST 26 S PROTEASOME. *Journal of Biological Chemistry* **271**: 32810-32817
- Samarkina ON, Popova AG, Gvozdk EY, Chkalina AV, Zvyagin IV, Rylova YV, Rudenko NV, Lusta KA, Kelmanson IV, Gorokhovatsky AY, Vinokurov LM (2009) Universal and rapid method for purification of GFP-like proteins by the ethanol extraction. *Protein Expression and Purification* **65**: 108-113
- Satoh T, Saeki Y, Hiromoto T, Wang Y-H, Uekusa Y, Yagi H, Yoshihara H, Yagi-Utsumi M, Mizushima T, Tanaka K, Kato K (2014) Structural Basis for Proteasome Formation Controlled by an Assembly Chaperone Nas2. *Structure* **22**: 731-743
- Sauer RT, Baker TA (2011) AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annual review of biochemistry* **80**: 587-612
- Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA, Hersch GL, Joshi SA, Kenniston JA, Levchenko I, Neher SB, Oakes ESC, Siddiqui SM, Wah DA, Baker TA (2004) Sculpting the Proteome with AAA+ Proteases and Disassembly Machines. *Cell* **119**: 9-18
- Schimke RT (1970) CHAPTER 32 - Regulation of Protein Degradation in Mammalian Tissues¹². In *Mammalian Protein Metabolism*, Munro HN (ed), pp 177-228. Academic Press

Schimke RT, Ganschow, R., and Doyle, D. (1968) Regulation of protein turnover in mammalian tissues. *Fed Proc* **27**

Schleif RF (1993) *Genetics and Molecular Biology*: Johns Hopkins University Press.

Schmidt M, Lupas AN, Finley D (1999) Structure and mechanism of ATP-dependent proteases. *Current Opinion in Chemical Biology* **3**: 584-591

Schoenheimer R, Clarke HT (1942) *The dynamic state of body constituents*, Cambridge, Mass.: Harvard University Press.

Schreiner P, Chen X, Husnjak K, Randles L, Zhang N, Elsasser S, Finley D, Dikic I, Walters K, Groll M (2008) Ubiquitin docking at the proteasome via a novel PH domain interaction. *NATURE* **453**: 548-552

Seemüller E, Lupas A, Zühl F, Zwickl P, Baumeister W (1995) The proteasome from *Thermoplasma acidophilum* is neither a cysteine nor a serine protease. *FEBS Letters* **359**: 173-178

Sen M, Maillard Rodrigo A, Nyquist K, Rodriguez-Aliaga P, Pressé S, Martin A, Bustamante C (2013) The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears. *Cell* **155**: 636-646

Seong IS, Kang MS, Choi MK, Lee JW, Koh OJ, Wang J, Eom SH, Chung CH (2002) The C-terminal Tails of HslU ATPase Act as a Molecular Switch for Activation of HslV Peptidase. *Journal of Biological Chemistry* **277**: 25976-25982

Sherman MY, Goldberg AL (2001) Cellular Defenses against Unfolded Proteins: A Cell Biologist Thinks about Neurodegenerative Diseases. *Neuron* **29**: 15-32

Shockley KR, Ward DE, Chhabra SR, Connors SB, Montero CI, M. Kelly R (2003) Heat Shock Response by the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **69**: 2365-2371

Siddiqui SM, Sauer RT, Baker TA (2004) Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *GENES & DEVELOPMENT* **18**: 369-374

Ślędź P, Unverdorben P, Beck F, Pfeifer G, Schweitzer A, Förster F, Baumeister W (2013) Structure of the 26S proteasome with ATP- γ S bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proceedings of the National Academy of Sciences* **110**: 7264-7269

Slonczewski JL, Foster JW (2013) *Microbiology: An Evolving Science (Third Edition)*: W. W. Norton.

Smith D, Benaroudj N, Goldberg A (2006) Proteasomes and their associated ATPases: a destructive combination. *J Struct Biol* **156**: 72 - 83

Smith DM, Chang S-C, Park S, Finley D, Cheng Y, Goldberg A (2007) Docking of the Proteasomal ATPases' C-termini in the 20S Proteasomes alpha Ring Opens the Gate for Substrate Entry. *Molecular Cell* **27**: 731-744

Smith DM, Kafri G, Cheng Y, Ng D, Walz T, Goldberg AL (2005) 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

Smith DM, Kafri, G., Cheng, Y., Ng D., Walz, T., Goldberg, A. L., (2005) 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

Snider J, Houry W (2008) AAA+ proteins: diversity in function, similarity in structure. *Biochem Soc Trans* **36**: 72 - 77

Song HK, Hartmann C, Ramachandran R, Bochtler M, Behrendt R, Moroder L, Huber R (2000) Mutational studies on HslU and its docking mode with HslV. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 14103-14108

Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB (2000) Crystal and Solution Structures of an HslUV Protease–Chaperone Complex. *Cell* **103**: 633-643

Squires GL (2012) *Introduction to the Theory of Thermal Neutron Scattering*: Cambridge University Press.

Stadtmueller BM, Hill CP (2011) Proteasome Activators. *Molecular Cell* **41**: 8-19

Stahlberg H, Kutejová E, Suda K, Wolpensinger B, Lustig A, Schatz G, Engel A, Suzuki CK (1999) Mitochondrial Lon of *Saccharomyces cerevisiae* is a ring-shaped protease with seven flexible subunits. *Proceedings of the National Academy of Sciences* **96**: 6787-6790

Stolz A, Hilt W, Buchberger A, Wolf DH (2011) Cdc48: a power machine in protein degradation. *Trends in Biochemical Sciences* **36**: 515-523

Strickland E, Hakala K, Thomas PJ, DeMartino GN (2000) Recognition of Misfolding Proteins by PA700, the Regulatory Subcomplex of the 26 S Proteasome. *Journal of Biological Chemistry* **275**: 5565-5572

Su H, Wang X (2010) The ubiquitin-proteasome system in cardiac proteinopathy: a quality control perspective. *Cardiovascular Research* **85**: 253-262

Suzuki CK, Rep M, van Dijl JM, Suda K, Grivell LA, Schatz G (1997) ATP-dependent proteases that also chaperone protein biogenesis. *Trends in Biochemical Sciences* **22**: 118-123

Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**: 495-503

Svergun D, Barberato C, Koch MHJ (1995) CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**: 768-773

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* **76**: 2879-2886

Svergun DI, Koch MHJ, Timmins PA, May RP (2013) *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*: OUP Oxford.

Svergun DI, Richard S, Koch MHJ, Sayers Z, Kuprin S, Zaccai G (1998) Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 2267-2272

- Swaffield JC, Melcher K, Johnston SA (1995) A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. *NATURE* **374**: 88-91
- Thomas B, Dominique D, Pierre R, Aurélien T, Patrice V, Javier P (2016) A Brief Survey of State-of-the-Art BioSAXS. *Protein & Peptide Letters* **23**: 1-1
- Timmins PA, Zaccai G (1988) Low resolution structures of biological complexes studied by neutron scattering. *European Biophysics Journal* **15**: 257-268
- Unno M, Mizushima T, Morimoto Y, Tomisugi Y, Tanaka K, Yasuko N, Tsukihara T (2002) Structure Determination of the Constitutive 20S Proteasome from Bovine Liver at 2.75 Å Resolution. *Journal of Biochemistry* **131**: 171-173
- Verma R, Aravind L, Oania R, McDonald WH, Yates JR, Koonin EV, Deshaies RJ (2002) Role of Rpn11 Metalloprotease in Deubiquitination and Degradation by the 26S Proteasome. *science* **298**: 611-615
- Vilchez D, Saez I, Dillin A (2014) The role of protein clearance mechanisms in organismal ageing and age-related diseases. *Nature Communications* **5**
- Voges D, Zwickl Pa, Baumeister W (1999) The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. *Annual review of biochemistry* **68**: 1015-1068
- Wang J, Song J, Franklin M, Kamtekar S, Im Y, Rho S, Seong I, Lee C, Chung C, Eom S (2001) Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* **9**: 177 - 184
- Weber-Ban EU, Reid BG, Miranker AD, Horwich AL (1999) Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *NATURE* **401**: 90-93
- Weiss M (2001) Global indicators of X-ray data quality. *Journal of Applied Crystallography* **34**: 130-135
- Wenzel T, Baumeister W (1995) Conformational constraints in protein degradation by the 20S proteasome. *Nat Struct Mol Biol* **2**: 199-204
- Werner F (2008) Structural evolution of multisubunit RNA polymerases. *Trends in Microbiology* **16**: 247-250
- William Roy S, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet* **7**: 211-221
- Wilson HL, Aldrich HC, Maupin-Furlow J (1999) Halophilic 20S Proteasomes of the Archaeon *Haloferax volcanii*: Purification, Characterization, and Gene Sequence Analysis. *JOURNAL OF BACTERIOLOGY* **181**: 5814-5824
- Wilson HL, Ou MS, Aldrich HC, Maupin-Furlow J (2000) 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
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A,

- Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**: 235-242
- Woese C (1998) The universal ancestor. *Proceedings of the National Academy of Sciences* **95**: 6854-6859
- Xu B, Jahic M, Enfors S-O (1999) Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of *Escherichiacoli*. *Biotechnology Progress* **15**: 81-90
- Yakamavich JA, Baker TA, Sauer RT (2008) Asymmetric nucleotide transactions of the HslUV protease. *Journal of molecular biology* **380**: 946-957
- Yamamoto A, Lucas JJ, Hen R (2000) Reversal of Neuropathology and Motor Dysfunction in a Conditional Model of Huntington's Disease. *Cell* **101**: 57-66
- Yao T, Ndoja A (2012) Regulation of gene expression by the ubiquitin-proteasome system. *Seminars in cell & developmental biology* **23**: 523-529
- Yu Y, Smith DM, Kim HM, Rodriguez V, Goldberg AL, Cheng Y (2010) Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions. *The EMBO Journal* **29**: 692-702
- Zhang F, Hu M, Tian G, Zhang P, Finley D, Jeffrey PD, Shi Y (2009a) Structural Insights Into The Regulatory Particle Of The Proteasome From *Methanocaldococcus jannaschii*. *Molecular Cell* **34**: 473-484
- Zhang F, Wu Z, Zhang P, Tian G, Finley D, Shi Y (2009b) Mechanism of Substrate Unfolding and Translocation by the Regulatory Particle of the Proteasome from *Methanocaldococcus jannaschii*. *Molecular Cell* **34**: 485-496
- Zhou G, Kowalczyk D, Humbard MA, Rohatgi S, Maupin-Furlow JA (2008) Proteasomal Components Required for Cell Growth and Stress Responses in the Haloarchaeon *Haloferax volcanii*. *JOURNAL OF BACTERIOLOGY* **190**: 8096-8105
- Zierenberg RA, Adams MWW, Arp AJ (2000) Life in extreme environments: Hydrothermal vents. *Proceedings of the National Academy of Sciences* **97**: 12961-12962
- Zwickl P, Ng D, Woo KM, Klenk H-P, Goldberg AL (1999) An Archaeobacterial ATPase, Homologous to ATPases in the Eukaryotic 26 S Proteasome, Activates Protein Breakdown by 20 S Proteasomes. *Journal of Biological Chemistry* **274**: 26008-26014