

# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Université Toulouse 3 - Paul Sabatier

---

Présentée et soutenue par  
**Manon BATISTA**

Le 2 octobre 2020

**Métabolisme de l'ARN chez les archées : dialogue entre l'hélicase à ARN ASH-Ski2, l'exoribonucléase 5'- 3' aRNase J et l'exosome 3'-5' à ARN, à proximité du ribosome.**

---

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **BIOLOGIE STRUCTURALE ET FONCTIONNELLE**

Unité de recherche :

**LMGM - Laboratoire de Microbiologie et Génétique Moléculaires**

Thèse dirigée par

**Béatrice CLOUET-D'ORVAL et Valérie GUILLET**

Jury

Mme Emmanuelle SCHMITT, Rapporteur

Mme Tamara BASTA-LE BERRE, Rapporteur

M. Fabien DARFEUILLE, Rapporteur

Mme Gwennaële FICHANT, Examinatrice

Mme Béatrice CLOUET-D'ORVAL, Directrice de thèse

Mme Valérie Guillet, Co-directrice de thèse



## **Métabolisme de l'ARN chez les Archées :**

### **Dialogue entre l'Hélicase à ARN ASH-Ski2, l'Exoribonucléase 5'-3' aRNase J et l'Exosome 3'-5' à ARN, à proximité du Ribosome.**

La compréhension du métabolisme de l'ARN, c'est-à-dire tout événement du cycle de la vie des molécules d'ARN, de leur synthèse à leur dégradation, est fondamentale. Ces mécanismes permettent aux cellules de s'adapter rapidement à un environnement variable. Bien que de nombreuses voies de régulation aient été établies chez les eucaryotes et les bactéries, ces processus ont été peu examinés chez les archées. Malgré les informations fragmentaires disponibles sur les protéines qui assurent la maturation et la dégradation des ARN, de nombreuses questions demeurent en suspens et les voies métaboliques de l'ARN dans les cellules d'archées restent, pour la plupart, à découvrir. Les archées sont considérées comme formant un domaine du vivant singulier, distinct des eucaryotes et des bactéries, qui est désormais considéré comme constituant des modèles d'une importance universelle pour l'étude des mécanismes moléculaires conservés. L'équipe d'accueil "RNA biology of Archaea" développe un domaine de recherche innovant pour identifier les machines moléculaires et les enzymes impliquées dans le métabolisme de l'ARN chez les Archées. Ainsi les ribonucléases (RNases) de type  $\beta$ -CASP ont été identifiées comme des enzymes ubiquitaires chez les Archées. Dans ce contexte le travail de thèse a consisté (i) à répertorier les familles de RNases (structures et fonction) connues chez les Archées pour la rédaction d'une revue (Clouet-d'Orval et al 2018) et (ii) à identifier et caractériser les partenaires protéiques de l'exoRNase 5'-3' de la famille  $\beta$ -CASP, aRNase J, chez les Thermococcales. Les résultats obtenus, par des approches de protéomiques, biochimiques et enzymatiques, montrent que l'aRNase J est engagée spécifiquement avec l'hélicase ASH-Ski2 et l'exosome à ARN à proximité du ribosome (résultats publiés : Phung et al NAR 2020) et que le partenaire ASH-Ski2 est une hélicase à ARN, avec une polarité 3'-5' dont l'activité prépondérante est l'appariement de brins (annealing) (rédaction en cours : Batista et al, en préparation). L'ensemble des données obtenues qui sont confortées par des approches phylogénomiques, offrent de nouvelles perspectives quant à l'identification de nouvelles machineries moléculaires impliquées dans le métabolisme de l'ARN chez les Archées.

## RNA metabolism in Archaea

### Dialogue between the ASH-Ski2 RNA helicase, the 5'-3' exoribonuclease aRNase J and the RNA exosome at the vicinity of the ribosome

A deep understanding of RNA metabolism, which includes any event in the life cycle of RNA molecules, such as their synthesis, folding/unfolding, modification, processing and degradation, is critical. The continuous breakdown and synthesis of mRNAs allow cells to rapidly adapt to changing environment. RNA biology became an important and fast-evolving field at the center of post-transcriptional regulation of gene expression. Although numerous regulation pathways have been established in Eukarya and Bacteria, these processes have been hardly examined in Archaea. Despite fragmented information gathered on the proteins that maturate and degrade RNAs within archaeal cells, many open questions remain and RNA pathways in Archaea are, for the most part, still a mystery. Archaea are considered to form a separate domain of life, distinct from Eukarya and Bacteria. In-depth phylogenomic studies give strong evidences that Eukarya and Archaea are sister groups with closely related genetic information systems (translation, transcription, replication, DNA recombination and reparation). Archaea are now viewed as models with universal importance to study the molecular mechanisms that are conserved between Eukarya and Archaea. The "RNA biology of Archaea" group has developed an innovative area of research to identify RNA processing machineries in Archaea. In this context, the thesis work consisted (i) in listing the RNase families (structure and function) known in the Archaea for the writing of a review (Clouet-d'Orval et al 2018) and (ii) in identifying and characterizing the protein partners of exoRNase 5'-3' of the family  $\beta$ -CASP, aRNase J, in Thermococcales. The results obtained, using proteomic, biochemical and enzymatic approaches, show that aRNase J is specifically engaged with the ASH-Ski2 helicase and the RNA exosome in the vicinity of the ribosome (published results: Phung et al NAR 2020) and that the ASH-Ski2 partner is an RNA helicase, with a 3'-5' polarity, whose predominant activity is strand assembly (annealing) (writing in progress: Batista et al, in preparation). All the data obtained, which are supported by phylogenomic approaches, offer new prospects for the identification of new molecular machinery involved in RNA metabolism in Archaea.



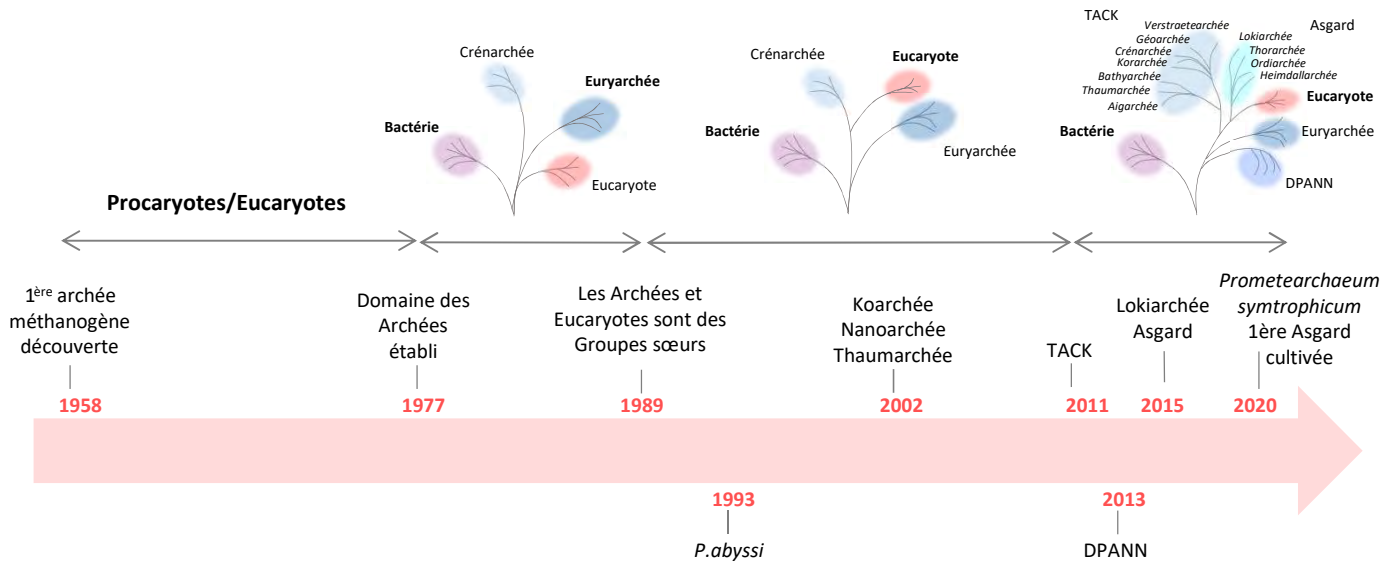
INTRODUCTION .....	6
I. Les Archées.....	8
1. Généralités .....	8
1.1 Découverte et classification .....	8
1.2 Les modèles d'études .....	12
2. La Réplication et la Réparation de l'ADN.....	13
2.1 La Réplication de l'ADN .....	13
2.2 La réparation de l'ADN .....	17
3. La Transcription .....	18
3.1 L'ARN polymérase .....	20
3.2 L'initiation.....	20
3.3 L'élongation.....	21
3.3 La terminaison .....	21
3.4 Les Facteurs de régulation de la transcription .....	23
4. La traduction.....	25
4.1 L'Initiation.....	28
4.2 L'Elongation .....	28
4.3 Terminaison et recyclage des ribosomes .....	29
4.4 Les voies de dégradation des protéines.....	29
II. La biologie des ARN chez les Archées.....	32
1. Les ARN messagers.....	34
2. Les ARN de transfert.....	36
3. Les ARN ribosomiques.....	37
4. Les ARN à boîte C/D et H/ACA.....	39
5. Les ARN CRISPR.....	39
6. Les protéines liant l'ARN .....	40
6.1 Les protéines Sm-like/Hfq .....	40
6.2 La protéine L7Ae.....	41
III. Ribonucléases chez les Archées .....	41
1. Les EndoRNases.....	42
2. Les ExoRNases 5'-3' : la famille $\beta$ -CASP .....	44
2.1 Le groupe CPSF .....	46
2.2 Le groupe RNase J.....	47
3. Les ExoRNases 3'-5' .....	48

3.1 L'aRNase R .....	49
3.2 L'Exosome à ARN .....	49
IV. Les Hélicases à ARN chez les Archées .....	52
1. Les Familles SF1-SF2 .....	56
2. Les hélicases Ski2-like .....	58
V. Publication 1 : Clouet d'Orval B., Batista M. et al. 2018 .....	64
Projet de thèse .....	65
MATERIELS ET METHODES .....	66
I. Construction des vecteurs d'expression .....	67
1. Clonage avec le kit InFusion® (Ozyme).....	69
2. Clonage par PCR inverse.....	69
3. Mutagénèse dirigée.....	70
II. Production des protéines recombinantes.....	70
1. Surexpression des protéines recombinantes.....	71
2. Lyse et préparation de l'extrait brut .....	71
3. Purification des protéines sans étiquette : <i>Pab</i> -aRNase J, <i>Pab</i> -Hel308, <i>Pab</i> -ASH-Ski2 et ses variants.....	71
4. Purification des protéines avec étiquette Histidine: <i>Pab</i> -aRNase J-(His) <sub>6</sub> , (His) <sub>6</sub> - <i>Pab</i> -Rrp41, (His) <sub>6</sub> - <i>Pab</i> -Csl4, (His) <sub>6</sub> - <i>Pab</i> -ASH-Ski2 et ses variants .....	72
5. Quantification des protéines.....	73
6. Conservation des protéines.....	73
III. Pull down pour établir les réseaux d'interactions des protéines (His) <sub>6</sub> - <i>Pab</i> -Rrp41 et (His) <sub>6</sub> - <i>Pab</i> -Csl4 .....	73
1. Pull down.....	73
2. Analyse protéomique .....	74
IV. Test d'interactions <i>in vitro</i> .....	74
1. Co-purification de mélange de culots protéiques recombinés sur colonne de Nickel en système FPLC.....	74
2. Co-purification de mélange de protéines recombinantes purifiées sur colonne de Nickel en système FPLC.....	75
3. Co-purification de mélange de protéines purifiées sur bille de Nickel .....	75
4. Mélange de protéines purifiées en système HPLC.....	75
V. Tests enzymatiques.....	76
1. Radiomarquage des substrats .....	77
2. Test ATPase.....	77
3. Test d'affinité pour les acides nucléiques .....	77
4. Test d'appariement et déroulement des brins d'acides nucléiques .....	78

VI. Tests Biophysiques et structurales.....	79
1. Diffusion dynamique de la lumière en solution (Dynamic Light Scattering) .....	79
2. SEC-MALS (Size Exclusion Chromatography – Multi Angle Light Scattering) .....	79
3. Fluorimétrie à balayage différentiel (Differential Scanning Fluorimetry) .....	80
4. Essais de cristallisation .....	80
5. Modélisation structurale des protéines .....	82
RESULTATS.....	83
I. Publication 2 : Phung D.K., Etienne C., Batista M. et al. 2020.....	84
II. Etudes des réseaux d’interactions d’aRNase J, ASH-Ski2 et des sous unités Rrp41 et Csl4 de l’Exosome à ARN.....	86
1. Technique dite de « pull down » .....	86
2. Réseaux d’interactions d’aRNase J et ASH-Ski2 .....	88
3. Réseaux d’interactions des sous unités de l’exosome Rrp41 et Csl4.....	94
4. Conclusions.....	102
III. Interaction entre partenaires du réseau d’interaction de <i>Pab</i> -aRNase J.....	102
1. Co-purification par affinité de protéines recombinantes Proie/Appât.....	102
2. <i>Pab</i> -aRNase J et <i>Pab</i> -ASH-Ski2 forment un complexe stable <i>in vitro</i> .....	104
3. <i>Pab</i> -aRNase J et la sous-unité de la coiffe <i>Pab</i> -Csl4 de l’exosome à ARN forment un complexe stable <i>in vitro</i> .....	110
4. Etudes préliminaires de co-purification de <i>Pab</i> -aRNase J et <i>Pab</i> -ASH-Ski2 avec la sous-unité 30S purifiée du ribosome de <i>P. abyssi</i> .....	114
IV. Caractérisation biochimique de <i>Pab</i> -ASH-Ski2 .....	116
1. Activité ATPasique .....	118
2. Liaison aux acides nucléiques .....	122
3. Activité hélicase.....	124
4. Conclusions.....	130
V. Etudes structurales d’ASH-Ski2 et aRNase J.....	132
1. Purification de <i>Pab</i> -aRNase J et <i>Pab</i> -ASH-Ski2 .....	132
2. Homogénéité des échantillons protéiques <i>Pab</i> -aRNase J et <i>Pab</i> -ASH-Ski2 .....	136
3. Etudes structurales préliminaires par cristallographie .....	139
DISCUSSION .....	142
ANNEXE .....	157



# INTRODUCTION



**Figure 1. Historique de la découverte des Archées et évolution de l'arbre du vivant.** Plusieurs dates clé sont indiquées, notamment celles de l'identification des quatre grands groupes d'archées (Euryarchée, TACK, DPANN et Asgard).

Cette introduction de thèse est composée de quatre grandes parties. La première partie présente le domaine des Archées et décrit les mécanismes moléculaires fondamentaux de la maintenance de l'ADN à la traduction en passant par la transcription. La deuxième partie détaille les différents types d'ARN et leur processing chez les Archées. La troisième partie présente ce qui est connu sur les ribonucléases d'Archées. La quatrième partie se focalise sur les Hélicases à ARN.

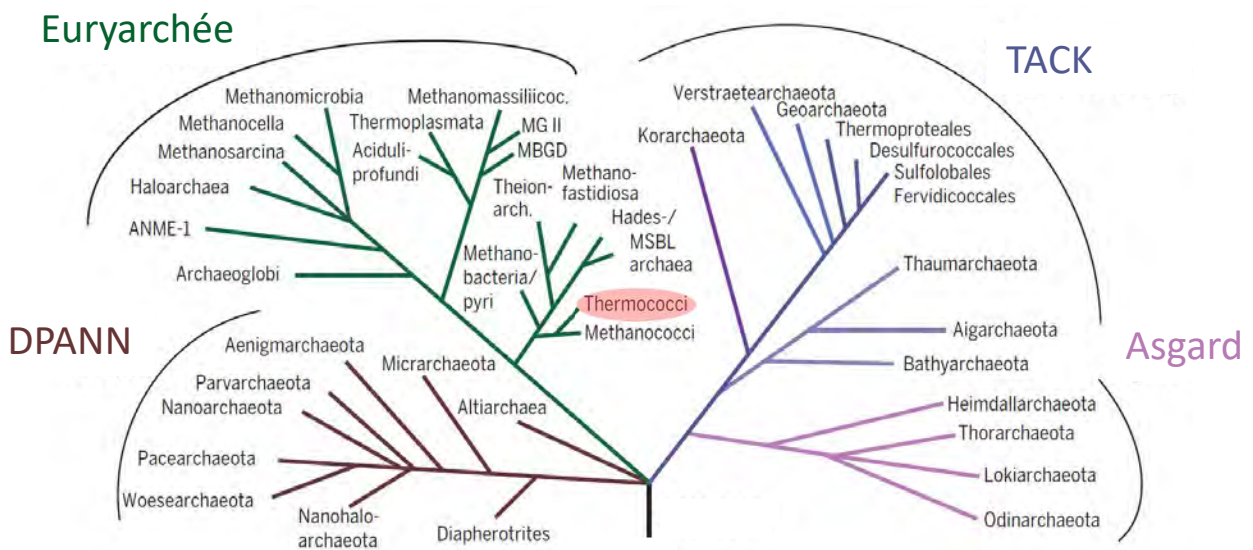
## I. Les Archées

### 1. Généralités

#### 1.1 Découverte et classification

En 1977, Carle Woese et Georges E. Fox révolutionnent la biologie de l'évolution en proposant un arbre du vivant en trois domaines: les Eucaryotes, les Bactéries et les Archées (**Figure 1**) (Woese and Fox 1977; Woese et al. 1990). Les premières archées identifiées sont majoritairement extrémophiles avec des habitats peu propices au développement de bactéries et eucaryotes. Elles sont retrouvées dans des habitats anoxiques comme les marécages (Méthanogènes), dans des eaux hypersalines comme la mer morte (Halophiles), à des températures très élevées comme dans les cheminées hydrothermales des fonds marins ou les sources chaudes comme dans le parc de Yellowstone (Thermophiles), ou à l'opposé dans des conditions de froid permanent comme les lacs secs de l'Antarctique (Psychrophiles) et à des pH extrêmes (Acidophiles et Alkaphiles). A présent, les archées sont considérées comme des organismes ubiquitaires présent dans tous les milieux terrestres, soit environ 20% de la biomasse des océans (DeLong and Pace 2001), ainsi que dans les microbiotes humain, de la peau, et des systèmes digestif et respiratoire (Drancourt et al. 2017; Mihajlovski et al. 2010; Oxley et al. 2010; Lurie-Weinberger and Gophna 2015; Gaci et al. 2014; Brugère et al. 2014; Moissl-Eichinger et al. 2018; Koskinen et al. 2017). Leur impact sur la santé humaine reste cependant à déterminer.

De manière générale, la forme et la taille des cellules d'archées sont similaires à celles des bactéries, néanmoins les parois des cellules d'archées ne possèdent pas de peptidoglycane (Kandler and Hippe 1977) et contrairement aux membranes bactériennes, leurs membranes contiennent des lipides à base d'isoprène fixés au glycérol-1-phosphate par des liaisons éther (Langworthy et al. 1972; Tourte et al. 2020). Les génomes d'archées possèdent une forte

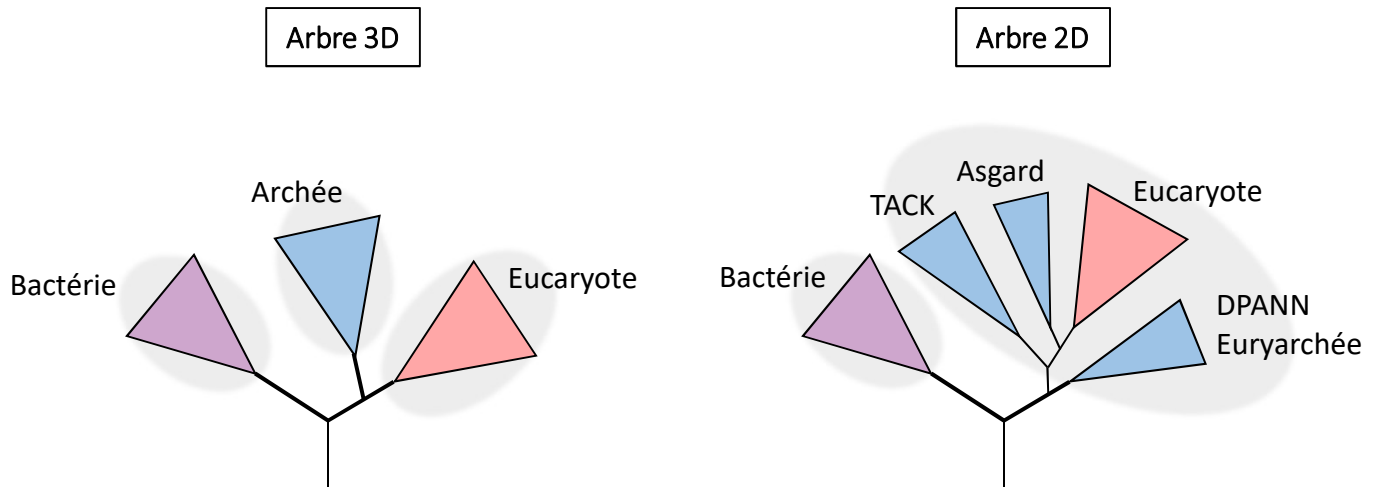


**Figure 2. Les clades d'Archées.** Les lignées d'archées sont aujourd'hui divisées en quatre clades : les Euryarchées (vert), TACK (bleu), DPANN (marron) et Asgard (violet) (Spang et al 2017).



densité de gènes et consistent en un chromosome circulaire dont la taille varie entre 0,49 (*Nanoarchaeum equitans*) et 5.75 Mpb (*Methanosarcina acetivorans*) (Kellner et al. 2018; Wang et al. 2015; Koonin and Wolf 2008). Ils peuvent disposer de plasmides dont 60 ont été isolés à ce jour (Wang et al. 2015).

Pendant plus de 20 ans, les Archées ont été classifiées en deux groupes (clades): les Euryarchées et les Crénarchées (Olsen and Woese 1993). A partir de 2002, avec le développement des techniques de séquençage à haut débit et l'essor des analyses phylogénomiques, plusieurs nouveaux groupes tels que les Korarchées, Nanoarchées et Thaumarchées ont été identifiés (**Figure 1**). A partir de 2011, la phylogénie des Archées s'enrichit avec une nouvelle classification qui établit le groupe des TACK qui comprend les Thaumarchées, Aigarchées, Crénarchées et Koarchées et le groupe des DPANN, qui sont caractérisés par de petits génomes (~0,5 à 1,5 Mpb) et constitués des Diaphérotrites, Parvarchées, Aenigmarchées, Nanoarchées et Nanohaloarchées (Kellner et al. 2018). Depuis, les données actuelles avec plus de 2133 génomes d'archées disponibles sur le site du NCBI, ont permis d'élargir encore cette diversité et permettent de préciser cette phylogénie avec l'identification des Géoarchées, Bathyarchées et Verstraetearchées au sein des TACK et de nouvelles lignées affiliées au groupe des Euryarchées (Eme et al. 2018; Eme et al. 2017; Spang et al. 2017). En 2015, avec la découverte de la nouvelle branche des Lokiarchées retrouvées dans des sédiments marins près du château de Loki (cheminées hydrothermales au milieu de l'océan Atlantique) et à partir de données de métagénomiques, la phylogénie des archées ouvre de nouvelles perspectives sur l'ancêtre commun des eucaryotes et des archées (Spang et al. 2015) (**Figure 2**). A présent il est proposé que le nouveau clade des Asgards, composé des Lokiarchées, Thorarchées, Odinararchées et Heimallarchées serait la branche la plus proche des eucaryotes (Zaremba-Niedzwiedzka et al. 2017). A partir de cette nouvelle classification, deux scénarios évolutifs ont été proposés selon les approches phylogénomiques et suscitent de nombreux débats (Gribaldo and Brochier-Armanet 2020). Le premier scénario, appelé modèle 3D, propose trois branches du vivant (**Figure 3**) avec un ancêtre commun précédant la diversification des archées pour les lignées sœurs archées et eucaryotes (Zhu et al. 2019; Da Cunha et al. 2018; Forterre 2013). Le second scénario, appelé modèle 2D, propose un arbre à deux branches avec le domaine des eucaryotes émergeant du domaine des Asgards (**Figure 3**) (Williams et al. 2020; Eme and Ettema 2018; Spang et al. 2018; Eme et al. 2018; Gribaldo



**Figure 3. Schéma des deux modèles proposés de l'évolution de l'arbre du vivant.**

Le scénario de l'arbre du vivant avec trois branches (3D) et celui avec l'arbre à deux branches (2D) sont représentés. Dans l'arbre 2D le domaine des archées est une branche sœur des eucaryotes. Dans l'arbre 3D, les eucaryotes émergent dans la branche des archées et sont proches des Asgards.

and Brochier-Armanet 2020; Dacks et al. 2016; Klinger et al. 2016; Hug et al. 2016; Doolittle 2020). Ce scénario est basé sur la découverte de gènes codant pour des protéines de signatures eucaryotes (ESP) présents dans tous les membres de la lignée des Asgards (Spang et al. 2018; Eme et al. 2018; Zaremba-Niedzwiedzka et al. 2017) et plus précisément du groupe des Heimdallarchées (Williams et al. 2020; Zaremba-Niedzwiedzka et al. 2017; Schleper and Sousa 2020). Récemment une souche de Lokiarchée du groupe des Asgard, nommée *Promethearchaeum syntrophicum*, a pu être isolée à partir de sédiments prélevés dans des eaux profondes japonaises et est maintenant cultivable en laboratoire (Imachi et al. 2020). Cette avancée majeure ouvre de nombreuses perspectives et permettra d'apporter de nouveaux éléments vers la compréhension de la relation évolutive archées-eucaryotes (Schleper and Sousa 2020).

## **1.2 Les modèles d'études**

Malgré les avancées technologiques dans le domaine du séquençage à haut débit qui ont permis d'élargir notre vision sur la diversité des archées, les difficultés à mettre des souches d'archées en culture et le manque de marqueurs génétiques ont été un frein pour le développement d'outils génétiques (Farkas et al. 2013; Adam et al. 2017; Baker et al. 2020). A présent il est possible de transformer certaines lignées de Crénarchées et d'Euryarchées, de déléter ou remplacer des gènes avec la technique de Pop-in/Pop-out et d'exprimer une protéine portant une étiquette *in vivo* (Allers and Mevarech 2005; Farkas et al. 2013; Atomi et al. 2012). Chez les Crénarchées, les modèles les plus développés sont les Sulfolobales hyperthermophiles comme par exemple *Sulfolobus solfataricus* (Wagner et al. 2012). Chez les Euryarchées, divers modèles sont développés chez les Méthanogènes comme *Methanosarcina acetivorans* (Kohler and Metcalf 2012), les Halophiles comme *Haloferax volcanii* (Leigh et al. 2011) et les Thermococcales comme *Thermococcus kodakarensis* ou *Thermococcus barophilus* (Thiel et al. 2014).

En ce qui concerne les Thermococcales, elles sont divisées en trois genres principaux : Pyrococcus, Thermococcus et Palaeococcus qui se développent de manière hétérotrophes et sous anaérobiose stricte à des températures allant de 80 à 100°C (Cossu et al. 2015). Malgré leurs conditions de croissance extrêmes les Thermococcales affichent un temps de génération très court, de l'ordre de 23 min (Gorlas et al. 2013).

*Pyrococcus abyssi* est le modèle d'étude qui a été utilisé dans ces travaux de thèse. Cette archée hyperthermophile a été isolée pour la première fois dans des cheminées hydrothermales situées au nord des Fidji à une profondeur de 2000m (Erauso et al. 1993). Sa température de croissance optimale est de 96°C avec un temps de dédoublement de 33min. En laboratoire elle est cultivée à 90°C dans des conditions en anaérobiose à un pH de 6.4. Le génome de *Pyrococcus abyssi* est l'un des premiers à avoir été séquencé et annoté (Cohen et al. 2003). Elle dispose d'un chromosome d'environ 1,7Mpb avec un pourcentage en GC autour de 44%, code pour plus de 1800 protéines et possède un plasmide (pGT5). Même si les outils génétiques n'ont pas encore été développés pour cette souche, elle est considérée comme organisme modèle dans l'étude de la maintenance de l'ADN (Pluchon et al. 2013). La stabilité et la résistance de ses protéines thermophiles en fait un organisme modèle pour la biochimie.

## **2. La Réplication et la Réparation de l'ADN**

### **2.1 La Réplication de l'ADN**

La réplication de l'ADN est le processus qui permet la duplication et le transfert de l'information génétique d'une cellule mère à deux cellules filles. Ce mécanisme est conservé dans les trois domaines du vivant et requiert de nombreux facteurs qui assurent une duplication de l'ADN de manière précise et rapide au sein d'un complexe appelé le « réplisome ». Pendant longtemps, l'étude de la réplication et de la réparation de l'ADN chez les archées a été entravée par un manque d'outils génétiques et biophysiques. Cependant, au cours de la dernière décennie, de nouveaux outils ont été développés dans plusieurs espèces d'archées (Zatopek et al. 2018; Pluchon et al. 2013). Même si les archées disposent d'un chromosome circulaire, comme les bactéries, leur machinerie de réplication ressemble à celle des eucaryotes avec certaines spécificités (Raymann et al. 2014; Kelman and Kelman 2014; Yao and O'Donnell 2016).

#### *Les origines de réplication*

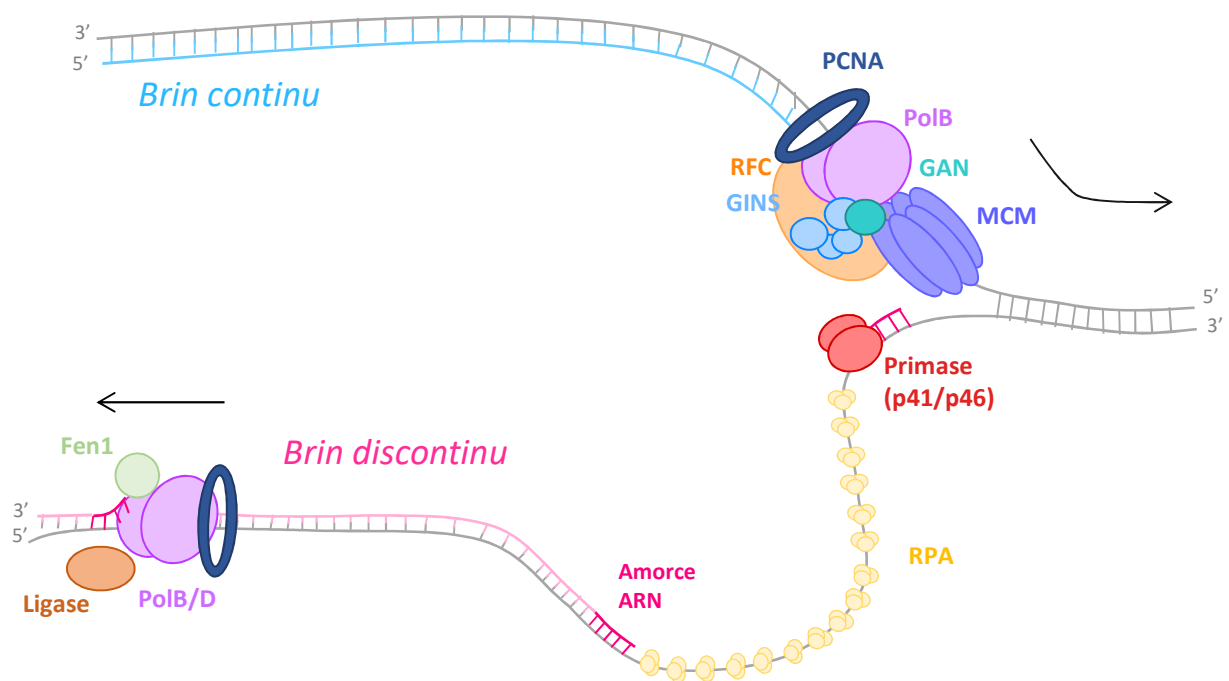
La réplication de l'ADN commence à des sites spécifiques sur le chromosome, appelés origines de réplication. Les origines de réplication sont présentes dans des régions intergéniques riches en résidus A/T et sont composées de séquences spécifiques de liaison reconnues par des

facteurs protéiques à proximité des gènes codant pour l'Origin recognition complex (ORC). Les archées possèdent une ou plusieurs origines de réplication selon les espèces (Kelman and Kelman 2018). Par exemple, le chromosome de *Pyrococcus abyssi* contient une seule origine (Matsunaga et al. 2003) tandis que ceux de *Sulfolobus acidocaldarius* et *Sulfolobus solfataricus* en ont trois (Lundgren et al. 2004; Robinson et al. 2004). De études récentes ont permis de montrer que les origines de réplication de *Thermococcus tokodaii* et *Haloferax volcanii* ne sont pas essentielles et que l'initiation de la réplication utilise un mécanisme de recombinaison homologue en parallèle (Gehring et al. 2017; Kelman and Kelman 2018). Ces observations nécessitent cependant d'être approfondies.

### *L'initiation*

Chez les archées, la réplication est initiée par la formation du complexe protéique Cdc6-MCM à l'origine de réplication. Cdc6 est homologue à la protéine Cdc6 eucaryote et permet le recrutement de l'hélicase MCM (MiniChromosome Maintenance) qui déroule localement le duplexe d'ADN et initie la formation de la fourche de réplication (**Figure 4**) (Sakakibara et al. 2009) (Kelman et al. 1999; Costa and Onesti 2009; Kelman et al. 2020). Ce complexe recrute ensuite les facteurs GINS et GAN, qui jouent un rôle essentiel dans l'établissement et le maintien des fourches de réplication (Marinsek et al. 2006; MacNeill 2010). GAN (pour GINS-associated nuclease) est une protéine de la famille des phosphoestérases DHH et est homologue à la protéine Cdc45 eucaryote (Makarova et al. 2012; Oyama et al. 2016; Marinsek et al. 2006; Nagata et al. 2017).

Une fois que l'hélicase MCM a déroulé l'ADN double brin (ADNdb), les protéines RPA (pour replication protein A) se lient et recouvrent l'ADN simple brin (ADNsb) exposé (**Figure 4**) pour empêcher la reformation du duplex et protéger l'ADN contre les attaques de nucléases ou modifications chimiques (Dickey et al. 2013). Dans la majorité des Aigarchées et Crenarchées, les protéines RPA sont similaires aux SSB bactériennes. Les génomes de Koarchées et Thaumarchées contiennent à la fois des homologues bactériens et eucaryotes (Wadsworth and White 2001; Raymann et al. 2014). Les autres lignées d'archées contiennent un à plusieurs homologues de RPA eucaryote, comme chez *Pyrococcus furiosus* et *Pyrococcus abyssi* qui comprennent un complexe trimérique RPA41/RPA14/RPA32 (Komori and Ishino 2001) ou *Methanocaldococcus jannaschii* qui n'en possède qu'une seule (Kelly et al. 1998).



**Figure 4. Schéma de la réplication d'archée.**

L'hélicase MCM déroule l'ADNdb de 3' en 5' et forme la fourche de réplication. Le brin continu (bleu) est synthétisé par la Polymérase B (PolB) aidé de PCNA. Le brin discontinu (rose) est polymérisé de façon discontinue en fragments d'Okazaki initiés par la primase qui synthétise des amorces d'ARN. Les séquences ARN sont ensuite éliminées grâce à la polymérase et Fen1 et les brins d'ADN sont reliés grâce à la Ligase. Les protéines RPA maintiennent et protègent l'ADNsb.

### *L'élongation*

L'ADN polymérase et le reste des composants du réplisome sont ensuite recrutés pour démarrer la synthèse bidirectionnelle de l'ADN. Chez les archées, la synthèse de l'ADN est réalisée par deux ADN polymérases, PolB et PolD, qui utilisent des amorces ARN pour synthétiser le brin complémentaire et possèdent une activité de correction exonucléasique de 3' en 5' (Greenough et al. 2015). Comme les ADN polymérases ne se déplacent que dans le sens 5' vers 3' et que le duplexe d'ADN déroulé est antiparallèle, un brin est synthétisé de façon continue (brin continu) tandis que l'autre est copié de façon discontinue (brin discontinu) sous la forme d'une série de fragments ARN d'Okazaki synthétisé par la primase (p41 et p46) (**Figure 4**) (Bocquier et al. 2001; Chemnitz Galal et al. 2012; Frick and Richardson 2001). Le facteur PCNA (pour proliferating cell nuclear antigen) est recruté par le facteur de réplication C (RFC) et s'associe à l'ADN polymérase pour lui conférer une haute processivité. PCNA forme un anneau trimérique qui encercle l'ADNdb et se lie ensuite à l'ADN polymérase pour une synthèse de brin d'ADN rapide et efficace. PCNA interagit également avec d'autres facteurs du métabolisme des acides nucléiques et du cycle cellulaire, le plus souvent via son motif PIP et est également impliquée dans la réparation de l'ADN (Pan et al. 2011). L'endonucléase Fen1 et la ligase à ADN sont ensuite recrutées pour éliminer l'amorce ARN et sceller l'ADN respectivement. (Balakrishnan and Bambara 2013)

### *La terminaison de la réplication*

À la fin du processus de réplication, les deux fourches (sur un chromosome circulaire, ou des fourches adjacentes sur un chromosome linéaire) entrent en collision et la synthèse de l'ADN s'arrête. Les données sur la terminaison de la réplication d'archée sont limitées, la nature circulaire de leurs chromosomes suggère un mécanisme similaire à celui des bactéries qui se termine dans une région *ter* du chromosome à l'aide de la protéine Tus (Duggin et al. 2011). Cependant ni la présence d'une séquence *ter*-like ni celle d'une protéine homologue à Tus n'a été décrite chez les archées.

## **2.2 La réparation de l'ADN**

La réplication de l'ADN peut être influencée par des facteurs environnementaux, endogènes ou externes, entraînant des dommages et des altérations de la molécule d'ADN, c'est pourquoi le processus de réparation de l'ADN est nécessaire pour garantir l'intégrité de l'information génétique portée par le chromosome. Les archées vivent souvent dans des environnements exposés à des températures, une salinité, une pression ou pH extrêmes, dans lesquels les dommages physiques à l'ADN sont fréquents, nécessitant des voies de réparation de l'ADN particulièrement robustes.

Les données observées jusqu'à présent suggèrent que les archées possèdent plusieurs voies de réparation comme la réparation par excision de bases (BER), la réparation par excision de nucléotides (NER), la recombinaison homologue (HR) ou non homologue (NHEJ) (Grogan 2015; White and Allers 2018; Zatopek et al. 2018). Elles possèdent un certain nombre de gènes codant des protéines impliquées dans la réparation de l'ADN qui diffère selon les espèces (White and Allers 2018). La voie NER permet d'éliminer des lésions volumineuses qui déforment l'ADN. La plupart des archées possèdent au moins quelques gènes NER de type eucaryote, comme XPD et XPB, mais leur fonction n'est pas encore claire (Fujikane et al. 2010). La voie BER est utilisée lors des dommages de bases individuelles, causés par exemple par une oxydation ou une méthylation. Ce sont les dommages les plus courants. Cette voie peut faire intervenir de nombreuses enzymes dont la nucléase Fen1, la lyase RP (Grasso and Tell 2014) et PCNA. En effet PCNA joue un rôle important dans les réparations de l'ADN en interagissant notamment avec l'endonucléase AP (Kiyonari et al. 2009), l'ADN glycosylase (Kiyonari et al. 2008) et la nucléase XPF (Roberts et al. 2003). La RNase HII permet l'excision des ribonucléotides intégrés par erreur à la place d'un nucléotide et le brin d'ADN est ensuite pris en charge par la PolB et Fen1 (Heider et al. 2017).

Les voies HR et NHEJ interviennent après une cassure double brin. La voie HR permet également de relancer la réplication de l'ADN lorsque les fourches sont bloquées au niveau d'un dommage à l'ADN (Gehring et al. 2017). Les cassures doubles brins doivent être traitées par des exonucléases pour générer des extrémités 3' d'ADNsb. Dans la voie HR les protéines Mre11, Rad50, NurA et HerA permettent la maturation des extrémités double brin (Ahdash et al. 2017; Delmas et al. 2013). L'extrémité 3' d'ADNsb est ensuite liée par la recombinase RadA qui est impliquée dans l'invasion du brin au niveau d'une séquence homologue et forme ainsi



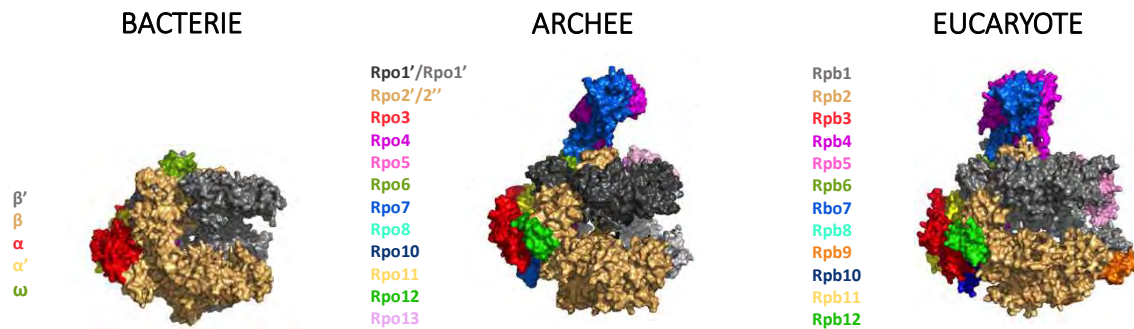
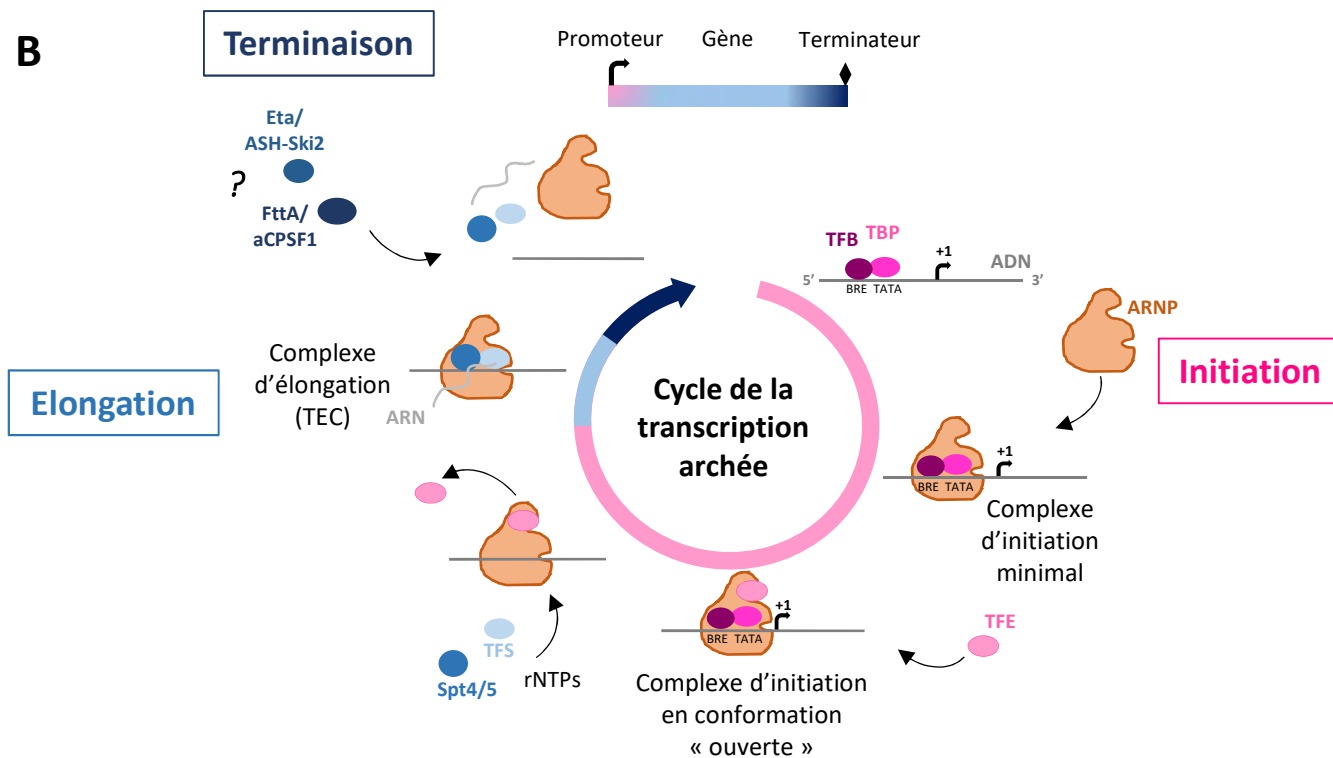
la double jonction d'Holliday (Morrical 2015). Cette conformation pourrait être reconnue par l'hélicase Hel308 (aussi nommée Hjm) chez les archées. Hel308 est une hélicase de la famille Ski2 qui est capable *in vitro* de dérouler les jonctions de Holliday en présence d'ATP (Guy and Bolt 2005). Des expériences de pull-down et chromatographie d'exclusion ont montré qu'Hel308 de *Sulfolobus Tokodaii*, interagit avec l'endonucléase spécifique des jonctions de Holliday, Hjc (Li et al. 2008) et plus récemment chez *Sulfolobus islandicus* les données démontrent une interaction directe entre Hel308 et PINA, une ATPase impliquée dans les jonctions de Holliday (Zhai et al. 2018). Chez l'homme, l'activité de Hel308 est localisée au niveau des fourches de réplication avec la protéine de réparation Rad51 et son activité de déroulement des duplexes d'ADN est stimulée par la protéine RPA (Tafel et al. 2011). Ces données suggèrent qu'Hel308 pourrait être requise pour le redémarrage de la réplication, lorsque la machinerie a été stoppée suite à un dommage de l'ADN.

La voie NHEJ, qui est rare chez les archées, est initiée par protéine Ku qui recrute ensuite une nucléase, polymérase et ligase qui traitent et réparent la cassure double brin (Bartlett et al. 2013).

Les archées hyperthermophiles possèdent une topoisomérase unique, capable de former des « surenroulements » d'ADN, la réverse gyrase (Forterre 2002; Yang et al. 2020). Le rôle exact de ces surenroulements chez les hyperthermophiles n'est pas entièrement compris. Il est proposé que la réverse gyrase permettrait de maintenir l'ADN sous forme double brin ce qui pourrait le rendre plus résistant à divers processus induisant des dommages (Kampmann and Stock 2004) ou permettrait de réguler l'expression des gènes lors de l'initiation de la transcription (Bell et al. 1998).

### 3. La Transcription

Les archées utilisent un appareil de transcription de type eucaryote qui transcrit un génome relativement petit et dense, avec des gènes souvent organisés en opérons. Leur machinerie de transcription est cependant proche de celle de l'ARN polymérase II (ARNP II), enzyme responsable de la transcription des ARNm et ARNc eucaryotes (Grohmann and Werner 2011; Blombach et al. 2019). Les organismes modèles archées qui ont été les plus étudiés pour la transcription sont les thermophiles telles que les Sulfolobales, les Thermococcales et les Pyrococcales ainsi que les Halophiles (Smollett et al. 2017). Le cycle de transcription est

**A****B**

**Figure 5 . La machinerie de transcription archée.**

**A**-Structures de l'ARNP bactérienne (*Thermus aquaticus*, pdb 116V), d'archée (*Sulfolobus shibatae*, pdb 4B1O) et l'ARNP II eucaryote (*Sacharomyces cerevisiae*, pdb 1NT9) (Werner 2013). Les sous-unités homologues sont de la même couleur et démontrent le degré de similitude entre les trois moteurs de la transcription et, en particulier, l'ARNP archée et eucaryote. **B**- La transcription d'un gène est initiée au niveau du promoteur, se poursuit dans le corps du gène et s'achève au niveau du termineur. L'ARNP progresse de manière répétée dans le cycle, et des sous-ensembles spécifiques de facteurs d'initiation, d'élongation et potentiellement de terminaison (indiqué avec un point d'interrogation) aident l'ARNP dans ce processus.

généralement divisé en trois phases principales, l'initiation, l'élongation et la terminaison, qui sont modulées par des facteurs de transcription et qui sont soumises à une régulation.

### **3.1 L'ARN polymérase**

L'ARNP est une enzyme composée de plusieurs sous unités qui transcrit un ADN matrice en ARN. Alors que les bactéries et les archées ne contiennent qu'une seule ARNP, les eucaryotes en contiennent trois : l'ARNP I (dédiée aux ARNr), ARNP II (ARNm et ARNnc) et ARNP III (ARNt, ARNr et ARNnc). La première structure cristallographique de l'ARNP archée a été résolue en 2008 chez *Sulfolobus solfataricus* (Hirata et al. 2012; Hirata et al. 2008) (**Figure 5A**). La résolution des structures tridimensionnelles des ARNP des trois domaines du vivant suggère une évolution à partir d'une enzyme ancestrale. L'organisation architecturale commune comprend un noyau universel conservé qui est composé de « la plate-forme d'assemblage » (Rpo3/11/10/12), des sous-unités catalytiques (Rpo1'/1''-Rpo2'/2'') et une petite sous-unité impliquée dans la stabilité de l'enzyme (Rpo6) (Werner 2013). Toutes les ARNP ressemblent à une pince de crabe dont le site actif de l'enzyme est situé dans le sillon inférieur de la pince. Chez les archées et les eucaryotes, ce noyau a été supplémenté de sous-unités qui étendent la plate-forme d'assemblage (Rpo10 et 12) avec des sites additionnels de liaison à l'ADN (Rpo5 et 13) et des facteurs de transcription (Rpo4 et 7). La caractéristique marquante distinguant les ARNP archée et eucaryotes de l'ARNP bactérienne est la présence d'une tige (Rpo4/7) qui dépasse du noyau en angle droit par rapport à la direction de la transcription (**Figure 5A**) (Werner and Grohmann 2011; Werner 2007; Fouqueau et al. 2017).

### **3.2 L'initiation**

La transcription est initiée par le recrutement de facteurs d'initiation au niveau du promoteur. La plupart des promoteurs répertoriés chez les archées ont trois éléments : une boîte TATA, un élément B de reconnaissance (BRE, B-recognition element) et le motif initiateur (Inr). Les séquences TATA et BRE sont des motifs spécifiques reconnus par deux facteurs de transcription : TBP (TATA binding Protein) et TFB (Transcription factor B), respectivement. TBP et TFB d'archées sont homologues aux facteurs TBP et TFIIB eucaryotes et ont des fonctions similaires. Brièvement, la liaison du facteur TBP au promoteur, induit une cascade qui conduit au recrutement de protéines pour l'initiation de la transcription. TFB se lie au complexe ADN-

TBP, et le complexe ADN-TBP-TFB recrute l'ARNP (**Figure 5B**). Ce complexe ADN-TBP-TFB-ARNP représente la machinerie de transcription minimale chez les archées, et est suffisant pour initier la transcription *in vitro*. Le facteur d'initiation de la transcription E (TFE), qui est l'homologue du facteur d'initiation TFEII eucaryote, est ensuite recruté et stabilise la conformation ouverte du complexe de pré-initiation qui facilite la séparation des brins d'ADN et le chargement du brin matrice par l'ARNP (Blombach et al. 2015). L'Inr est un motif dinucléotidique qui est important pour le choix précis du site de départ de la transcription (Smollett et al. 2017; Blombach et al. 2019).

### **3.3 L'élongation**

L'élongation de la transcription est un processus discontinu qui est fréquemment interrompu par une pause ou un arrêt dus à la nature de la séquence, des protéines associées à l'ADN (chromatine formée d'histones) ou des lésions de l'ADN. De nombreux facteurs sont impliqués dans ces processus, notamment le facteur de clivage de la transcription TFS, qui est l'homologue de TFSII eucaryote, et le facteur de processivité Spt4/5 (Fouqueau and Werner 2017). La protéine Spt4/5 est recrutée pour faciliter la transition entre initiation et élongation, en stabilisant le complexe d'élongation (TEC) (**Figure 5B**). Spt4/5 est homologue à DSIF chez l'Homme et NusG chez la bactérie. C'est le seul facteur associé à l'ARNP qui conservé dans les trois domaines du vivant (Smollett et al. 2017).

### **3.3 La terminaison**

La terminaison de la transcription est importante pour éviter les molécules d'ARN aberrantes et pour aider au recyclage de l'ARNP. Chez les bactéries, les signaux de terminaison de la transcription dépendent d'une petite structure en épingle en amont d'une région poly(U) (Terminaison intrinsèque) ou nécessitent l'action d'une protéine de régulation, le plus souvent l'hélicase hexamérique Rho (Terminaison Rho-dépendante) (Kuehner et al. 2011; Porrua et al. 2016).

Chez les eucaryotes, la terminaison par l'ARNP II peut être déclenchée par différentes voies, en fonction des signaux de l'ARN naissant en 3' et des facteurs de terminaison présents à l'extrémité d'un gène (Porrua et al. 2016; Kuehner et al. 2011). La voie CPSF fait intervenir le complexe CPSF (Cleavage and Polyadenylation Specificity Factor) qui induit le clivage

endoribonucléolytique à l'extrémité 3' de l'ARNm naissant. Le produit de clivage en amont est polyadénylé, tandis que le produit de clivage en aval est dégradé par l'exoribonucléase 5'-3' XRN2 (ou Rat1 chez la levure). Chez la levure, la terminaison de la transcription de nombreux ARNnc, tels que les ARNsn ou ARNsno (impliqués respectivement dans l'épissage et la modification des ARNr) par l'ARNP II fait intervenir la voie NNS (Nrd1-Nab3-Sen1). L'ARNP III est capable de terminer la transcription lorsqu'elle rencontre une séquence poly(T) de longueur variable sur le brin, sans qu'il soit nécessaire de recourir à des facteurs accessoires (Porrua et al. 2016). Le processus de terminaison de la transcription chez les archées reste un débat. Plusieurs études ont été réalisées notamment chez les halobactéries, les méthanocoques et les méthanobactéries. Globalement, ces différentes recherches démontrent que les sites de terminaison identifiés comprennent généralement des segments d'oligo (dT) et dans peu de cas des éléments de séquence pouvant former une structure en épingle à cheveux en amont du site de terminaison (Brown et al. 1989). Les premières études *in vitro* ont été réalisées sur un ARNt de *Methanococcus thermolithotrophicus* et ont montré que la terminaison de la transcription dépendait de la présence d'un segment poly(T) et d'une région T-loop en amont (Thomm et al. 1993). Une terminaison indépendante d'éléments structuraux a ensuite été décrite *in vitro* pour le gène hypA1 chez *Pyrococcus furiosus* ainsi que pour les gènes rapporteurs *in vivo* chez *Thermococcus kodakarensis* (Santangelo et al. 2009; Spitalny and Thomm 2008). Chez *Methanococcus thermoautotrophicus* et *Methanococcus jannaschii*, la terminaison est influencée par la longueur du segment d'oligo (dT) (Santangelo and Reeve 2006; Hirtreiter et al. 2010). Chez *Sulfolobus acidocaldarius* et *Methanosarcina mazei*, les extrémités 3' des ARNs ont été étudiées *in vivo* à l'aide de données RNA-seq et ont révélé que la terminaison se déroulait immédiatement après un motif poly(U) sans structure secondaire ARN (Dar et al. 2016). Les différentes données exposent aussi des caractéristiques de terminaison spécifiques à chaque lignée. Par exemple, récemment, l'identification des sites de terminaison à l'échelle du génome dans *Haloferax volcanii* a confirmé que le segment oligo(dT) est un motif de terminaison et a identifié un motif supplémentaire (AGATC) en aval du site de terminaison (Berkemer et al. 2020). Il reste à définir si la terminaison chez les archées repose uniquement sur des séquences intrinsèques ou si des facteurs protéiques sont impliqués. Récemment, chez *Thermococcus kodakarensis*, la protéine Eta (nommée ASH-Ski2 dans ce travail de thèse) a été proposée comme impliquée

dans la terminaison de la transcription (Walker et al. 2017). Cette protéine est une hélicase de la famille des Ski2, comprenant des enzymes impliquées dans l'épissage et le contrôle qualité de l'ARN et la réplication de l'ADN (Johnson and Jackson 2013) et qui fait l'objet d'étude dans l'équipe et en particulier au cours de cette thèse. L'étude chez *Thermococcus kodakarensis* montre qu'elle déclenche la libération des transcrits des complexes d'élongation bloqués (Walker et al. 2017). Cependant, dans cette publication l'activité hélicase décrite semble plutôt lente et ne peut suivre les taux d'élongation de l'ARNP. Il a récemment été proposé que Eta soit plutôt impliquée dans le redémarrage après un arrêt de transcription (Maier and Marchfelder 2019). Mes travaux de thèse ont mis en évidence que l'hélicase ASH-Ski2 a une spécificité pour l'ARN et que son réseau d'interaction inclut des acteurs majeurs du métabolisme de l'ARN (l'exoribonucléases 5'-3' aRNase J et l'exosome à ARN). Nos études montrent la présence des sous-unités Rpo1 et Rpo2 dans le réseau d'interaction d'ASH-Ski2, ce qui pourrait en effet montrer un lien avec la machinerie de transcription (Phung et al. 2020). Chez les archées, la protéine aCPSF1 est l'homologue de CPSF73 eucaryote engagée dans le complexe de terminaison de la transcription CPSF (Phung et al. 2013). aCPSF1 a une activité endoribonucléolytique préférentielle aux dinucléotides CA et une activité exoribonucléolytique 5'-3' (Phung et al. 2013). Les protéines Rpo1 et Rpo2 ont été retrouvées dans le réseau d'interactions d'aCPSF1 de *Pyrococcus abyssi* (Phung Thèse 2017). Une étude récente a montré qu'aCPSF1 (nommée FttA dans la publication) est capable de cliver et libérer les transcrits naissant du complexe d'élongation chez *Thermococcus kodokaraensis* et propose un mécanisme dans lequel aCPSF1/FttA serait responsable de la terminaison de la transcription (Sanders et al. 2020). Cependant, il reste à montrer une interaction fonctionnelle entre l'ARNP et aCPSF1/FttA. Les mécanismes de la terminaison de la transcription chez les archées restent à déterminer et pourraient être variés selon les groupes d'archées.

### **3.4 Les Facteurs de régulation de la transcription**

Bien que avec la machinerie de transcription basale archéenne est proche de celle des eucaryotes, les facteurs de transcription de la régulation génique (TFs) sont similaires aux facteurs bactériens (Karr et al. 2017). Les génomes archéens sont compacts avec des régions intergéniques très courtes, ce qui limite l'espace pour les sites de liaison des facteurs de transcription régulateurs par rapport aux génomes eucaryotes. La majorité des TFs

procaryotes sont constitués de deux domaines de liaison un à l'ADN (DBD) et l'autre aux ligands (LBD), qui interagissent avec les signaux intracellulaires (Lemmens et al. 2019). Le LBD est responsable de la détection des changements environnementaux ou métaboliques, souvent par interaction avec des petites molécules exogènes, par exemple des ions métalliques ou des composés aromatiques, ou avec des métabolites endogènes tels que les sucres et les acides aminés. La liaison des ligands par allostérie induit des changements de conformation du TF qui affectent généralement les propriétés de liaison à l'ADN, soit en provoquant une dissociation de l'ADN, soit en stimulant l'affinité de liaison à l'ADN. Une banque de données des TF archées existent, limité actuellement à 45 génomes <http://bioinformatics.zj.cn/archaeatf>. Les familles de TFs les plus représentées chez les archées sont : Lrp/AsnC, MarR et TrmB (Lemmens et al. 2019) dont les structures ont été résolues (Karr et al. 2017). La famille « Lrp/AsnC » (Leucine-responsive regulatory protein/ asparagine synthase C) est la famille la plus abondante chez les archées et régulent les gènes impliqués dans le métabolisme des acides aminés, dans le métabolisme central ou dans les processus de transport (Brinkman et al. 2003; Peeters and Charlier 2010). La famille MarR (multiple antibiotic resistance regulator) chez les bactéries regroupe des régulateurs de divers gènes impliqués dans différents processus tels que la réponse au stress, le métabolisme ou la virulence. C'est une famille bien représentée chez les archées mais leurs fonctions biologiques restent élusives. La famille TrmB, a d'abord été considérée comme spécifique des archées et a finalement été identifiée de façon sporadique chez certaines bactéries (Karr et al. 2017).

L'adaptation des cellules à des conditions physiologiques variables est aussi contrôlée par le remodelage local et global de la chromatine. Les systèmes de compaction de l'ADN sont très variables dans le domaine des archées (Sanders et al 2019 White et belle 2002) et plusieurs familles de protéines de la chromatine ont été caractérisées (Peeters et al. 2015). Les superfamilles Cren7 et Alba sont largement conservées et ont été proposées comme jouant un rôle dans la compaction du génomes (Aravind et al. 2003; Bell et al. 2002; Driessen and Dame 2013). La majorité des lignées d'archée, à l'exception des Crénarchées, codent les orthologues des histones H3 et H4 eucaryotes. Les histones d'archées ont une structure tertiaire similaire à celle des histones eucaryotes mais diffère une fois liée à l'ADN. Les histones eucaryotes forment des octamères sur l'ADN alors que les histones d'archées forment des

filaments de taille variable appelés Hypernucléosomes. (Nalabothula et al. 2013; Maruyama et al. 2013; Henneman et al. 2018). Une étude récente a révélé que les archées du genre *Sulfolobus* (Crénarchées) organisent leur génome en domaines définis par la protéine Coalescine (Takemata et al. 2019). Ces domaines interagissent pour former des compartiments définis de type A et B, correspondant à des régions de la chromatine active et inactive de la transcription respectivement, comme chez les eucaryotes (Takemata and Bell 2020).

#### 4. La traduction

La traduction est le processus de synthèse des protéines en décodant l'information génétique portée par l'ARNm. Les archées ne possédant pas de noyau, un couplage de la traduction et transcription est possible (French et al. 2007). La machinerie de synthèse des protéines est composée du ribosome, des ARNt aminoacylés et de facteurs protéiques qui assistent et modulent précisément les différentes étapes de la traduction pour assurer la synthèse des protéines. Même si les ARNm archées sont de types bactériens, les archées possèdent un appareil de traduction plus complexe, avec des facteurs et un ribosome proches de ceux décrits chez les eucaryotes (Lecompte et al. 2002; Londei 2005). Le processus de traduction est composé de trois étapes : l'initiation, l'élongation et la terminaison. Chacune de ces étapes fait intervenir des protéines G qui requièrent l'hydrolyse du GTP. A ce jour, l'étape la mieux décrite chez les archées est l'initiation, en contraste avec les étapes d'élongation et de terminaison (Benelli et al. 2017).

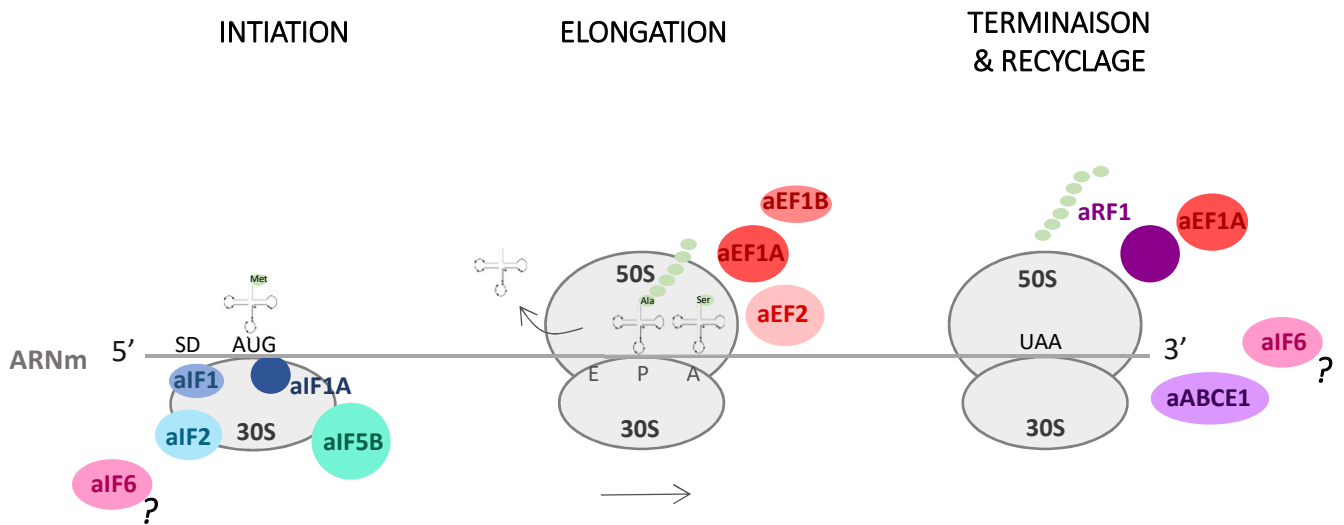
Les ribosomes sont des particules ribonucléoprotéiques composés des ARNr conservées qui portent l'activité catalytique (ribozyme) et de protéines ribosomiques. Ils sont constitués d'une grande et d'une petite sous-unité, les sous unités 30S et 50S chez les archées, formant un site de fixation des ARNt aminoacylés (site A), un site de fixation de ARNt déaminoacylé (site P) et un site de sortie de l'ARNt (site E). La sous unité 30S comprend l'ARN 16S et 25 protéines et permet de décoder la séquence de l'ARNm. La sous unité 50S contient les ARN 23S et 5S et 39 protéines ribosomiques qui structurent le site catalytique de la réaction de trans-peptidation. Les protéines ribosomiques d'archées sont soit universelles soit homologues à celles des eucaryotes (**Tableau 1**) (Lecompte et al. 2002; Armache et al. 2013; Ban et al. 2000; Ban et al. 2014). Récemment, une protéine du ribosome de *Pyrococcus abyssi*,



localisée à l'emplacement de la protéine eS21 dans le ribosome eucaryote a été identifiée et proposée comme spécifique des archées au vue de sa structure (Coueux et al. 2020). Les ARNr subissent des modifications, dont les plus abondantes sont la méthylation en 2'OH des riboses et la conversion de l'uridine en pseudouridine (Piekna-Przybylska et al. 2008; Ferreira-Cerca 2017). Ces modifications sont catalysées sur les pré-ARNr lors de la synthèse des ribosomes et semblent être importantes dans la fonction des ribosomes d'archées (Yip et al. 2013). Chez *Sulfolobus solfataricus*, les données observées suggèrent que les modifications des nucléotides aient un effet stabilisateur sur la structure du ribosome (Noon et al. 1998; Polikanov et al. 2015). Toutefois, les études fonctionnelles sur la maturation de l'ARNr chez les archées sont très limitées.

Vitesse de sédimentation		ARN	Protéines u	Protéines e	Protéine a
70S	30S	16S	15	9	1
	50S	23S, 5S	19	20	

**Tableau 1. Les ribosomes archées.** Le contenu en protéines ribosomales universelles (u), de type eucaryote (e) ou archée (a) sont indiqués pour *Pyrococcus abyssi*.



**Figure 6. Schéma synthétique des étapes de la traduction et de ses principaux facteurs chez les archées.** La traduction démarre par le recrutement de la petite sous unité du ribosome au niveau du codon de démarrage AUG et se termine au niveau d'un codon STOP (UAA,UAG ou UGA). Les séquences SD des ARNm sont indiquées. Les ribosomes sont schématisés en gris et les facteurs qui participent à chaque étapes de la traduction dans d'autres couleurs. Le point d'interrogation à côté du facteur aIF6 signifie que son rôle dans une étape spécifique de la traduction n'est pas connu, c'est pourquoi il est présenté dans les étapes d'initiation et de terminaison. Les ARNt aminoacylés et le peptides naissant sont schématisés.

#### **4.1 L'Initiation**

Pendant l'initiation de la traduction, les ribosomes doivent identifier le codon d'initiation (AUG) qui définit la phase ouverte de lecture (ORF) et adresser l'ARNt initiateur (ARNt<sub>i</sub><sup>Met</sup>). Le principe général de cette étape est le même à travers les trois domaines du vivant mais les mécanismes moléculaires diffèrent.

Les ARNm d'archées possèdent des séquences SD, des régions 5'-UTR très courtes ou en sont dépourvus (ARNm « Leaderless »). Brièvement l'initiation démarre par le recrutement de la petite sous unité du ribosome et des facteurs d'initiation aIF1, aIF1A, aIF2 et aIF5B qui sont homologues aux facteurs eucaryotes (Kyrpides and Woese 1998; Benelli and Londei 2011; Schmitt et al. 2019). aIF1 contribue à l'identification du codon d'initiation, aIF1A occupe le site A et aIF2 permet la liaison de l'ARNt<sub>i</sub> au site P. Une fois le codon initiateur reconnu, la grande sous unité 50S est recrutée à l'aide du facteur aIF5B (**Figure 6**). Le complexe de démarrage de la traduction chez *Pyrococcus abyssi* contenant tous les acteurs a été reconstitué *in vitro* et étudié par cryo-microscopie électronique (Coureux et al. 2016; Coureux et al. 2020). Un autre facteur aIF6, homologue du facteur eIF6, est certainement impliqué dans la traduction mais son rôle chez les archées comme chez les eucaryotes n'est pas clairement identifié (Benelli and Londei 2009).

#### **4.2 L'Elongation**

L'élongation de la traduction est elle-même divisée en trois étapes qui correspondent au décodage de l'ARNm, à la formation de liaisons peptidiques (trans-peptidation) et à la translocation. L'élongation qui est un processus conservé chez les bactéries et les eucaryotes (Dever et al. 2018; Rodnina 2018) l'est aussi chez les archées (Woese et al. 1990; Ito et al. 2014; Narrowe et al. 2018; Vitagliano et al. 2001; Benelli et al. 2017). Les étapes d'élongation sont principalement contrôlées par deux protéines à activité GTPase retrouvées dans les trois domaines du vivant : aEF1A et aEF2 (**Figure 6**). Durant la première étape, un ARNt amino-acylé entre dans le site A du ribosome et reconnaît le codon de l'ARNm avec l'aide du facteur aEF1A. La seconde étape est la trans-peptidation durant laquelle l'acide aminé porté par l'ARNt dans le site A est ajouté à la chaîne peptidique naissante porté par l'ARNt dans le site P. aEF1B permet le recyclage de aEF1A-GDP en aEF1A-GTP. La translocation est la dernière étape et entraîne le mouvement réciproque du ribosome et de l'ARNm qui provoque le décalage du

codon dans le site A. C'est la liaison du facteur aEF2 qui permet la translocation du ribosome et la libération des ARNt du site E. Les structures 3D d'aEF1A lié au GDP (*Pyrococcus horikoshii* et *Sulfolobus Solfataricus*) ou au GTP (*Aeopyrum pernix*) et d'aEF2-GTP (*Pyrococcus horikoshii*) ont été résolues (Ito et al. 2014; Vitagliano et al. 2001; Tanzawa et al. 2018).

#### **4.3 Terminaison et recyclage des ribosomes**

La terminaison de la traduction et le recyclage des ribosomes assurent la libération de la chaîne polypeptidique complète et la dissociation des deux sous unités du ribosome, de l'ARNt et de l'ARNm. Ce sont des étapes qui ont peu été décrites chez les archées. La terminaison a lieu lorsque le ribosome reconnaît un des trois codons stop (UAA, UAG ou UGA) au site A et nécessite l'intervention du facteur RF (pour release factor) aRF1 qui est homologue à eRF1 eucaryote (**Figure 6**). La structure du complexe aEF1A:GTP:aRF1 chez *Aeopyrum pernix* suggère cependant que le facteur aEF1A joue aussi un rôle dans la terminaison de la traduction (Kobayashi et al. 2012).

La dissociation des sous unités du ribosome et leur recyclage sont coordonnés par la protéine ABCE1 qui est homologue à ABCE1 eucaryote (Barthelme et al. 2011). Ce facteur intervient lors d'une terminaison canonique via aRF1/aEF1A ou après un arrêt prématuré du ribosome lors des voies de surveillance des ARN NGD (No Go Day) et NSD (Non Stop Decay) (Gerovac and Vogel 2019).

#### **4.4 Les voies de dégradation des protéines**

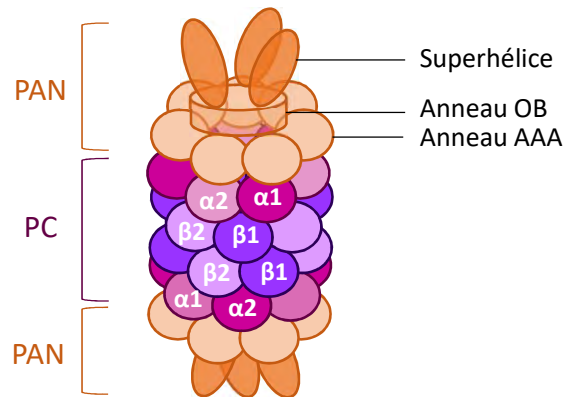
Le protéasome est la principale machine moléculaire pour la dégradation régulée des protéines intracellulaires chez les eucaryotes, les archées et certaines bactéries (actinomycètes). Cet assemblage macromoléculaire compartimenté dégrade de manière sélective les protéines mal repliées, mal traduites ou endommagées et joue un rôle essentiel dans le contrôle qualité des protéines, dans la réponse au stress et dans de nombreux autres processus (Majumder and Baumeister 2019). Alors que le protéasome eucaryote 26S a été caractérisé en détails, le protéasome-ATPase archée reste encore peu décrit.

Chez les eucaryotes, il existe deux voies majeures pour l'élimination sélective des protéines indésirables : le système protéasome-ubiquitine (SPU) et la voie autophagie-lysosome. Le SPU constitue la principale voie de dégradation pour les protéines intracellulaires alors que les

organelles, les protéines de surfaces et les pathogènes sont dégradés par autophagie. L'adressage à un processus de dégradation dépend de la protéine ubiquitine. L'ubiquitine se lie, via une cascade enzymatique, de manière covalente aux protéines par une liaison isopeptidique entre le groupe carboxyle de son résidu glycine en C-terminal, et le groupe amine des lysines des protéines cibles. Les répétitions d'une même réaction génèrent un "code ubiquitine" de signaux de topologies et de longueurs différentes permettant l'adressage spécifique. Chez les bactéries et les archées, l'existence de ces voies n'a pas été signalée, mais de petites protéines de type ubiquitine (SAMP et Pup) ont été identifiées.

La composition du protéasome varie en fonction des domaines du vivant mais la composante centrale, la particule centrale (PC ou protéasome 20S) est hautement conservée d'un point de vue architectural. Le PC est un baril de protéines en forme de tonneau, composé de quatre anneaux heptamériques superposés (De Mot et al. 1998; Baumeister et al. 1998). Les anneaux extérieurs comprennent des sous-unités de type  $\alpha$ , tandis que les anneaux intérieurs possèdent des sous-unités de type  $\beta$  (Figure 7). L'empilement coaxial des quatre anneaux crée trois cavités internes. La cavité centrale est la chambre catalytique, où la dégradation des protéines a lieu. Les sites actifs possèdent trois résidus qui sont conservés chez les eucaryotes, bactéries et archées, soulignant un mécanisme de clivage des liaisons peptidiques universel. Les deux cavités externes servent d'antichambres, où les protéines du substrat peuvent être stockées dans un état non plié avant à la dégradation (Majumder and Baumeister 2019).

La dégradation des protéines par la PC ne nécessite pas d'hydrolyse de l'ATP pour les petits peptides ou les protéines non structurées. Cependant, les plus grands substrats protéiques nécessitent des ATPases AAA+ qui déplient les substrats protéiques et les transfèrent dans la cavité centrale pour les dégrader. Les ATPases de type AAA+ forme un groupe étendu d'ATPase impliqué dans diverses activités cellulaires. Chez les eucaryotes, l'ATPase du protéasome est un assemblage de "particules régulatrices AAA-ATPases" (Rpt1-6) qui constitue le cœur de la particule régulatrice (PR). Chez les actinobactéries, une AAA-ATPase principale a été identifiée : ARC (ou MpA) et récemment un homologue de la protéine Cdc48, Cpa, a montré interagir avec le protéasomes de façon similaire aux autres AAA-ATPase (Ziemski et al. 2018). Chez les archées, deux ATPases de type AAA+ ont été identifiées comme associé au protéasome (Forouzan et al. 2012) : la protéine PAN (pour proteasome activating nucleotidase) (Figure 7), et la protéine "VCP-like ATPase" (VAT). PAN possède une activité

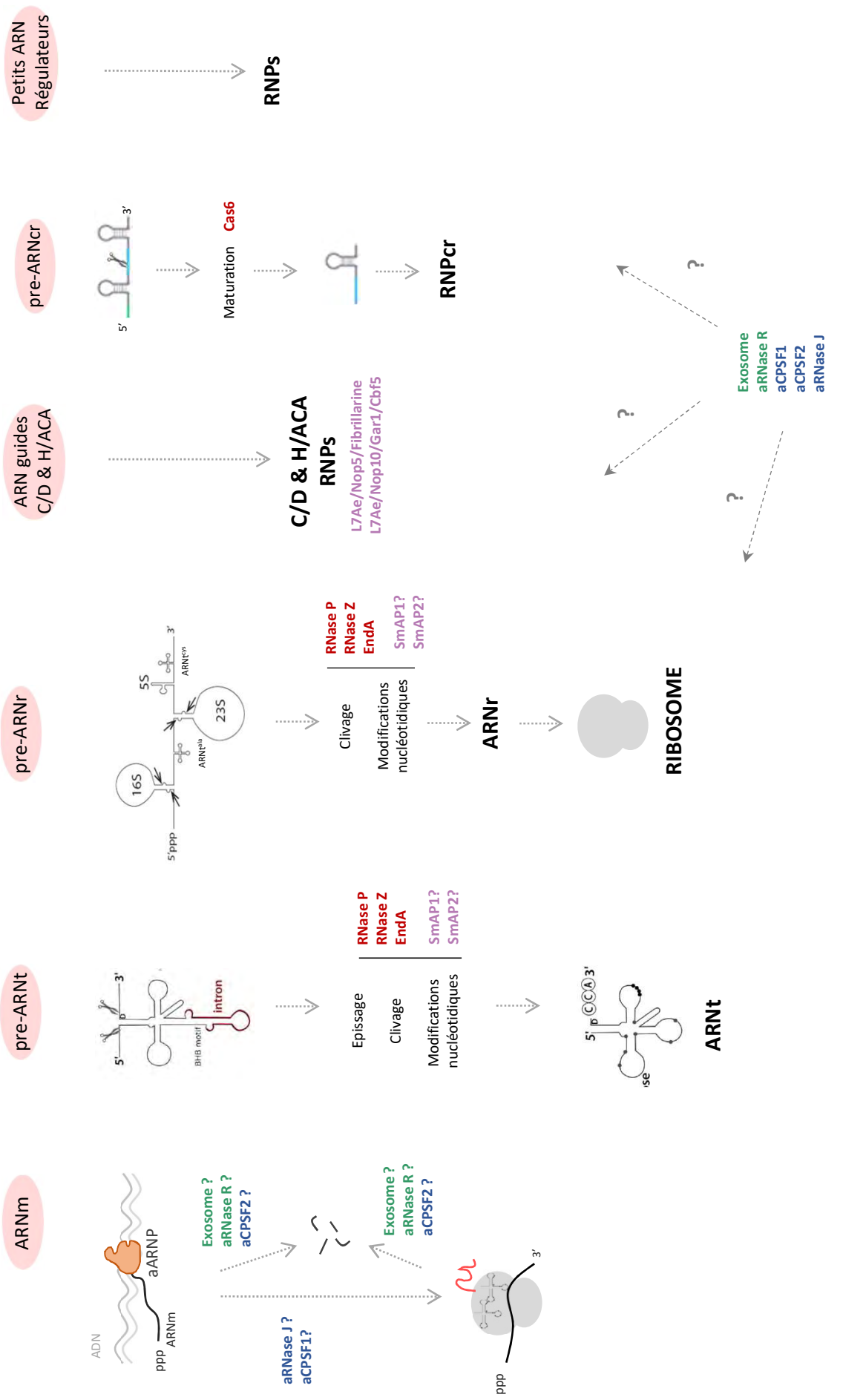


**Figure 7. Schéma architectural du protéasome-PAN d'archée .** La particule centrale du protéasome (PC) forme un tonneau constitué de quatre anneaux heptamériques superposés (violet). Six protomères de PAN s'associent pour former trois superhélices, un anneau hexamérique de domaines OB (Oligonucléotide/oligosaccharide binding) et un anneau hexamérique AAA (Orange). Deux hexamères de PAN se lient de part et d'autre à la PC via leur domaine ATPase (« anneau AAA ») en C-terminal.

chaperonne *in vitro*, en empêchant l'agrégation de protéines dénaturées, et aide à leur repliement (Franzetti et al. 2002; Navon and Goldberg 2001; Benaroudj and Goldberg 2000). Elle possède 41-45% d'homologie avec les protéines Rpts eucaryotes. La structure en cryo-microscopie de PAN et la PC du protéasome de *Archaeoglobus fulgidus* a été résolue récemment (Majumder et al. 2018). VAT est l'homologue de l'ATPase Cdc48 eucaryote (p97/VCP) qui joue un rôle crucial dans un large spectre d'activités cellulaires, y compris le dépliement de l'ubiquitine (Twomey et al. 2019; Barthelme and Sauer 2012; Barthelme et al. 2014).

## II. La biologie des ARN chez les Archées

La biologie des ARN joue un rôle critique dans la capacité des cellules à s'adapter aux changements environnementaux. L'expression de l'information génétique implique trois grands types d'ARN majeurs : les ARN messagers (ARNm) qui définissent le protéome ; les ARN stables qui regroupent les ARN ribosomiques (ARNr) et les ARN de transfert (ARNt) et qui fonctionnent ensemble pour décoder les séquences d'ARNm ; les ARN régulateurs non codants (ARNnc), qui régulent la stabilité et/ou la capacité d'un ARNm à être traduit, et qui guident l'ajout de modifications chimiques (panel d'ARN guides à boîtes C/D et H/ACA). Le processing de l'ARN regroupe tous les processus liés à la vie d'un ARN dans la cellule : sa maturation, ses modifications post-transcriptionnelles et sa dégradation. Les voies de processing de l'ARN sont au centre de la régulation de l'expression génique et sont fondamentales pour définir les caractéristiques phénotypiques d'une cellule. Elles permettent une adaptation rapide aux conditions environnementales. Ainsi les ARNs après leur transcription subissent des étapes de maturations et des modifications chimiques pour devenir fonctionnels dans la cellule. La quantité et qualité de chaque type d'ARN est étroitement contrôlée par le taux relatif de transcription, de maturation et de dégradation (Wagner and Romby 2015; Hourri-Zeevi and Rechavi 2017). Chez les archées, les informations sur les mécanismes de processing des différents types d'ARN et les enzymes associées aux modifications de l'ARN sont fragmentées (Clouet-d'Orval et al. 2018). Les informations sur les acteurs impliqués dans les voies métaboliques de processing de l'ARN obtenues chez les bactéries et les eucaryotes permettent de proposer des modèles chez les archées.



**Figure 8. Les types d'ARN archée.**

L'unique ARN polymérase archée transcrit tous les différents types ARN. Ils sont transcrits en précurseurs et subissent différentes étapes de maturation pour devenir fonctionnel. Seulement quelques étapes de la maturation des ARNt, ARNr et ARNcr ont été décrites à ce jour chez les archées. Les EndoRNases sont en rouge, les ExoRNases 3'-5' en vert et 5'-3' en bleu. Les protéines associées sont en violet.



Chez les Archées tous les ARN cellulaires sont transcrits par une ARN polymérase unique (ARNP) proche de l'ARNPII eucaryote bien que certains ARN aient des caractéristiques retrouvées chez les bactéries (**Figure 8**) (Phung et al. 2017).

## 1. Les ARN messagers

Les ARNm d'archées partagent des caractéristiques avec les ARNm bactériens : ils n'ont pas d'intron, ni de coiffe méthylée à l'extrémité 5', ni de queue poly(A) et possèdent une extrémité 5' triphosphorylée protectrice, des séquences « Shine-Dalgarno » d'initiation de la traduction, des régions 5' non traduites (5'-UTR) et/ou Leaderless. Des études proposent que les extrémité 5'-UTR contribuent à la stabilité de l'ARNm chez *Methanosarina mazei* (Cao et al. 2014) ce qui contraste avec d'autres Euryarchées et Crénarchées dans lesquels les transcripts Leaderless sont majoritaires (Babski et al. 2016; Londei 2005). Ces deux types d'ARNm pourraient être pris en charge par différents mécanisme de dégradation à l'extrémité 5' (Clouet-d'Orval et al. 2018). Plusieurs revues décrivent le temps de demi-vie des ARNm d'archées dans différentes conditions (Evguenieva-Hackenberg and Klug 2009; Evguenieva-Hackenberg and Klug 2011). En général, les ARNm archées ont un temps de demi-vie significativement plus grands que celui des ARNm bactériens. Brièvement, la demi-vie des transcripts varie de 7 à 57 min chez *Methanococcus vannielii* (Hennigan and Reeve 1994) et de 4 à 80 min chez *Haloferax mediterranei* (Jäger et al. 2002). Plus récemment, une image globale de l'expression différentielle chez *Methanosarcina acetivorans* a fournis des évidences à l'échelle du génome pour les demi-vie des ARN modulés par les sources de carbone et conduise à la conclusion que plus de la moitié des gènes du métabolisme sont contrôlés par la dégradation (Peterson et al. 2016).

Comme chez les bactéries et certains eucaryotes (Plantes, Algues et Champignons), les ARNm archées peuvent se structurer, notamment dans les régions 5' non traduites, pour former des riboswitches. Les riboswitches sont des segments génomiques cis-régulateurs non codants qui contrôlent l'expression des gènes en aval en subissant un changement conformationnel lors de la liaison des ligands ou sous l'effet de la température (Gupta and Swati 2019).

Le niveau d'expression de chaque ARN est déterminé par son taux de transcription et son taux de dégradation. Les voies de dégradation de l'ARNm sont essentielles pour maintenir l'homéostasie cellulaire. Elles permettent de passer d'une adaptation transitoire relative à un

taux de croissance à une réponse rapide face aux stimuli externes. La stabilité d'un ARNm peut être déterminée par un lot spécifique de ribonucléases, par la séquence et les structures de la molécule d'ARNm elle-même, ou par des petits ARN régulateurs (ARNnc). Les ARNnc sont souvent impliqués dans la dégradation ou la stabilisation d'ARNm cibles. A l'inverse des bactéries et des eucaryotes, les voies de dégradation et de contrôle qualité de l'ARNm chez les archées sont peu décrites et les facteurs contrôlant sa stabilité restent à identifier (**Figure 8**) (Clouet-d'Orval et al. 2018).

Chez *Sulfolobus solfataricus*, des études *in vitro* et *in vivo* suggèrent que la sous unité du facteur de traduction aIF2 se lie à l'ARNm au niveau de son extrémité 5' triphosphate et protège ainsi le transcrit d'une dégradation exoribonucléolytique de 5' vers 3' (Arkhipova et al. 2015; Hasenöhr et al. 2008). La liaison du facteur de la traduction aIF2 aux extrémités triphosphorylées des ARNm pourrait avoir une fonction similaire à la coiffe retrouvée à l'extrémité 5' des ARNm eucaryotes (m<sup>7</sup>Gppp) et suggère qu'une dégradation 5'-3' pourrait exister chez les archées. Ceci est conforté par la présence de l'activité 5'-3' exoribonucléolytique portée par les membres de la famille de ribonucléases  $\beta$ -CASP (Clouet-d'Orval et al. 2015). Cependant, chez les Euryarchées et Crénarchées, les exoribonucléases aRNase J et aCPSF2 ont montré une préférence pour des substrats ARN avec une extrémité 5' mono-phosphate. Le mécanisme moléculaire de la dégradation 5'-3' des ARNm reste inconnu et une enzyme capable de convertir l'extrémité tri-phosphate en mono-phosphate reste à identifier. La superfamille des phosphohydrolases Nudix, retrouvée dans les trois domaines du vivant, est impliquée dans la déprotection des extrémités 5' des ARNm chez les eucaryotes et les bactéries (Julius and Yuzenkova 2019). Dans les cellules eucaryotes, ce sont les protéines Dcp2 et Nudt16 de la superfamille Nudix qui enlèvent la coiffe 5' des ARNm. Chez les bactéries, l'extrémité tri-phosphorylée est convertie en mono-phosphate par la famille RppH de la superfamille Nudix. Deux de ces protéines ont déjà été clonées et exprimées chez *Sulfolobus solfataricus* mais aucune activité pyrophosphohydrolase n'a été détectée *in vitro*. A ce jour, il n'existe pas de signaux pour un clivage spécifique endoribonucléolytique de l'ARNm qui pourrait initier les voies de dégradation de l'ARN, même si l'endoribonucléase aCPSF1 est conservée dans tous les génomes archées (Clouet-d'Orval et al. 2015).

Des données récentes suggère qu'une surveillance des ARNm pourrait exister chez les Archées. Notamment grâce à l'identification de la protéine Pelota/Dom34-like (Ragan et al.

1996) qui est ubiquitaire dans les génomes d'archées (Clouet-d'Orval et al. 2018). Chez les eucaryotes, Pelota (dont la structure ressemble à eRF1) en complexe avec Hbs1 (qui ressemble à aEF1A), se lie au site A du ribosome et déclenche le début du processus de dissociation du ribosome (Becker et al. 2011). Une base structurale et un mécanisme universel du recyclage des ribosomes pour les archées et les eucaryotes ont été proposés par l'analyse par cryo-microscopie électronique de complexes reconstitués contenant l'ATPase ABCE1 et le facteur de terminaison Pelota (Becker et al. 2012). Beaucoup de questions restent à élucider dans les voies de contrôle qualité chez les Archées, en particulier les rôles de la machinerie conservée exosome et ceux des exo/endoribonucléases 5'-3'  $\beta$ -CASP dans les mécanismes de dégradations des ARNm.

## 2. Les ARN de transfert

Les ARNt d'archées sont transcrits comme de longs précurseurs (pré-ARNt) qui contiennent une séquence en 5' et une séquence terminale en 3'. Les ARNt matures requièrent une maturation aux deux extrémités et souvent l'ajout post-transcriptionnel d'une extrémité CCA en 3' (**Figure 8**). Le nombre de gènes qui codent des ARNt dans les génomes archées varie de 36 (*Methanobrevibacter smithii*) à 61 (*Methanosarcina barkeri*) (Chan and Lowe 2009). Ils peuvent être codés comme une unité seule de transcription en opéron avec d'autres ARNt, ou en unités polycistroniques en combinaison avec des transcrits d'ARNr ou ARNm. En moyenne, 15% des ARNt d'archées contiennent des introns dont la taille varie de 11 à 129 nucléotides (Kanai 2014). Leur épissage est un mécanisme spécifique qui requiert une endonucléase EndA, homologue à l'endonucléase des ARNt eucaryote (Thompson and Daniels 1988). Chez certaines archées, et notamment *Nanoarchaeum equitans*, les ARNt peuvent être codés sur deux gènes séparés, appelés « ARNt split ». Les deux parties de l'ARNt sont alors jointes par EndA pour former un ARNt mature (Randau et al. 2005; Chan et al. 2011).

Les ARNt sont les types d'ARN qui possèdent le plus de modifications post-transcriptionnelles avec 22 nucléosides modifiés universels retrouvés dans tous les organismes (Motorin and Helm 2010; Marck and Grosjean 2002). Certaines de ces modifications facilitent le repliement, la dynamique de structures et conformations des ARNs. Chez les archées, au moins 47 modifications différentes ont été identifiées dont certaines sont spécifiques à ces organismes, par exemple l'archaeosine ou l'agmatidine (Grosjean et al. 2008; Phillips et al. 2008).

La maturation des ARNt est un procédé conservé dans les trois domaines du vivant. La séquence guide en 5' d'un pré-ARNt est clivée par la RNase P dans tous les organismes. La maturation en 3' est catalysée par l'universelle endonucléase RNase Z. RNase Z reconnaît des sous structures sur l'ARNt, notamment le bras T et la tige acceptrice. S'il n'est pas déjà codé par l'ADNt, le motif CCA, qui est essentiel pour la liaison de l'acide aminé, est ajouté par la tRNA nucléotidetransferase (ou CCA-adding-enzyme ou tRNA terminal transferase) après clivage par la RNase Z (Wilusz et al. 2011). L'épissage des introns des ARNt d'archées est un mécanisme spécifique qui requiert une endonucléase EndA, homologue à l'endonucléase des ARNt eucaryote (Thompson and Daniels 1988). EndA reconnaît et clive les motifs Bulge-Helix-Bulge (BHB) des introns (**Figure 8**). Après clivage, les extrémités des exons sont reliées par une ligase à ARN et les introns sont relâchés sous forme circulaire.

### 3. Les ARN ribosomiques

Les ARNr d'archées sont, comme chez les eucaryotes et bactéries, produits à partir d'ARN polycistroniques précurseurs. Ils sont ensuite maturés pour permettre l'assemblage des ribosomes. Les ribosomes d'archées sédimentent à 70S, et sont composés des ARNr 23S, 16S et 5S. La majorité des génomes d'archées ne possèdent qu'un opéron *rrn*, à l'exception de certains qui en possèdent jusqu'à quatre répartis dans le génome comme pour *Methanococcus vannielii* ou *Halolamina sediminis* (Acinas et al. 2004). Deux principales organisations de l'opéron *rrn* sont retrouvées chez les archées. Dans les deux organisations, les ARNr 16S et 23S sont co-transcrits en de larges précurseurs. L'ARNr 5S est lui soit retrouvé en aval du gène de l'ARNr 23S comme chez les Euryarchées, soit il est codé séparément comme dans le génome des TACK. Les gènes *rrn* sont encadrés en 5' et 3' par des ETS (External Transcribed Spacers) et séparés par un ITS (Internal Transcribed Spacer) (Klug et al. 2007). Dans la plupart des Euryarchées l'ITS 1 code un ARNt<sup>ALA</sup> et dans certains cas l'ETS en 3' peut coder pour un ARNt<sup>CYS</sup> (Deppenmeier et al. 2002; Hartmann et al. 2009). Les précurseurs intermédiaires 16S et 23S sont circulaires et sont clivés et dégradés par des endo- et exoribonucléases non identifiées, pour donner des transcrits matures (**Figure 8**) (Tang et al. 2002; Danan et al. 2012; Qi et al. 2020). Le nombre de modifications chimiques sur les ARNr archées varie selon les groupes (Nomura et al. 2002). Chez les archées hyperthermophiles, le nombre de modifications observées est proche de celles des eucaryotes (~100), ce nombre

est réduit chez les halophiles (Grosjean et al. 2010; Grosjean et al. 2008). Chez *Sulfolobus solfataricus* par exemple, 67 méthylation 2'-O-ribose ont été reportés et seulement quatre chez *Haloferax volcanii* (Grosjean et al. 2008; Dennis et al. 2015). Cette disparité n'est pas encore bien expliquée. La plupart des modifications sont des pseudouridines et des méthylations 2'-O-ribose. Chez les archées et eucaryotes, ces modifications sont gouvernées par deux mécanismes indépendants (De Zoysa and Yu 2017). Le premier implique des enzymes appelées pseudouridine synthases et méthyltransférase qui reconnaissent le substrat et catalyse respectivement l'isomérisation de l'uridine en pseudouridine et le transfert du groupement méthyle sur le 2'-O-ribose. Le second fait intervenir les ARN guide à boîte C/D et boîte H/ACA. Chez les archées, la plupart des modifications sont réalisées par des complexes RNP guidés par des ARN. Il est aussi proposé que ces ARN guides soient des « Chaperons » qui facilitent le repliement des ARNr (Dennis et al. 2015).

Les voies de biogénèse des ribosomes chez les eucaryotes et les bactéries sont complexes et comprennent la maturation des ARNr et leur assemblage avec les protéines ribosomales. Ces étapes nécessitent l'action coordonnée d'endo- et exoribonucléases qui agissent avec des facteurs d'assemblage. Seulement quelques étapes des voies de biogénèse des ribosomes archées ont été identifiées (Ferreira-Cerca 2017) et certaines activités endoribonucléolytiques ont été identifiées mais la plupart reste élusives. Les régions qui entourent les séquences de l'ARNr 16S et 23S contiennent des motifs BHB clivés par l'endonucléase EndA, qui intervient aussi dans l'épissage des ARNt (**Figure 8**) (Thompson and Daniels 1988). Après l'action d'une ligase à ARN, les précurseurs circulaires 16S et 23S sont ouverts par un clivage endonucléolytique et dégradé par des endo- et exoribonucléases non identifiées, en transcrits matures (Tang et al. 2002; Danan et al. 2012). Basé sur des expériences *in vitro* chez *Pyrococcus Horikoshii*, il est proposé que aNob1, homologue de l'endonucléase Nob1 eucaryote -responsable du clivage de l'ARNr 18S chez la levure- soit responsable de la maturation en 3' de l'ARNr 16S (Veith et al. 2012). Les ARNt codés dans l'ITS ou dans l'ETS en 3' sont clivés par la RNase P et RNase Z. La RNase Z est aussi impliquée dans la maturation en 5' de l'ARNr 5S en clivant la structure ARNt-like (Danan et al. 2012). Les ARNr 23S et 16S archées peuvent contenir des introns qui contiennent les motifs BHB et sont clivés par EndA durant la maturation (Nomura et al. 2002; Itoh et al. 2003).

#### 4. Les ARN à boîte C/D et H/ACA

Les ARN à boîte C/D et H/ACA s'associent à des protéines et forment des complexes ribonucléoprotéiques (RNP) qui guident respectivement la 2'-O-méthylation du ribose et la pseudouridylation sur les ARNr et ARNt (**Figure 8**).

Les ARN à boîte C/D, qui font entre 50 et 60 nucléotides, sont caractérisés par des motifs (ou boîtes) de petites séquences : C/C' (RUGAUGA) et D/D' (CUGA) qui forment des motifs structuraux appelés Kink-turn (K-turn) et/ou Kink-loop (K-loop) (Gomes-Filho et al. 2018). Ils s'assemblent en complexes RNP actifs avec les protéines L7Ae, Nop5 et Fibrillarine (Omer et al. 2002). L7Ae se lie aux motifs structuraux Kink-turn et Kink-loop et recrute les protéines Nop5 et Fibrillarine. Nop5 permet la liaison entre les deux ARN et la Fibrillarine porte l'activité catalytique méthyltransférase.

Les ARN à boîte H/ACA d'archées sont moins nombreux qu'à boîte C/D et généralement constitués d'une seule tige-boucle d'environ 65-75 nucléotides qui est immédiatement suivie d'une boîte H (ANANNA) ou ACA (ACA) à un seul brin (Yip et al. 2013). À l'intérieur de la tige, deux brins non appariés forment une boucle bipartite qui forme la poche de pseudouridylation. Les protéines L7Ae, Nop10, Gar1 et Cbf5 sont associées aux ARN à boîte H/ACA pour former un RNP actif (Charpentier et al. 2007). La pseudouridine synthase Cbf5 se lie aux séquences H/ACA, L7Ae reconnaît les structures K-turn ou K-loop dans la tige de l'ARN, Nop10 organise les interactions entre L7Ae et Cbf5 et renforce la liaison à l'ARN et Gar1 est nécessaire pour la libération de l'ARN. La pseudouridylation stabiliserait les structures de l'ARN. Il a été démontré que la perte de ces modifications dans l'ARNr chez la levure affecte la traduction (Liang et al. 2009)(Gomes-Filho et al. 2018; Yip et al. 2013).

#### 5. Les ARN CRISPR

Les ARN CRISPR (ARNcr) sont impliqués dans la défense immunitaire de la cellule contre des éléments génétiques infectieux, comme des plasmides ou des virus. Les systèmes d'immunité CRISPR-cas sont trouvés dans la majorité des génomes archées (Makarova et al. 2015; Makarova et al. 2020). Brièvement, les locus CRISPR codent pour des courtes répétitions palindromiques espacés par des séquences uniques nommées espaceurs ou « spacers » qui s'apparient avec des éléments génétiques étrangers. Les gènes *cas* associés codent les protéines nécessaires à la réaction de défense, incluant des protéines de liaison à l'ARN, des

nucléases, des hélicases ou des polymérasés. La transcription du locus CRISPR produit un long pré-ARNcr qui est clivé à chaque répétition en ARNcr individuel (**Figure 8**). Durant la phase d'interférence, les molécules d'ADN ou ARN invasives qui s'apparient avec les ARNcr sont clivés par des complexes RNP effecteurs composés par des protéines Cas associées à des ARNcr individuels (Koonin et al. 2017). La nouvelle classification des systèmes CRISPR-Cas inclut 2 classes, 6 types et 33 sous-types qui diffèrent selon leur composition et fonctionnement (Makarova et al. 2020). Brièvement, la classe 1 possède un complexe effecteur multiprotéique alors que celui de la classe 2 est composée d'une seule protéine. Les archées possèdent généralement le système CRISPR de classe 1.

## **6. Les protéines liant l'ARN**

Dans la biologie des ARN, les protéines des superfamilles Sm/Hfq ou L7Ae, jouent un rôle important dans l'assemblage des complexes RNP.

### **6.1 Les protéines Sm-like/Hfq**

Les protéines Sm-like/Hfq forment une superfamille de protéines liées à l'ARN qui sont conservées à travers les trois domaines du vivant (Mura et al. 2013). Elles possèdent un domaine commun composé de deux motifs Sm connectés par une région variable en longueur ou motifs. Leur séquence primaire n'est pas conservée, mais elles partagent la même structure 3D en anneau/tonneau (Wilusz and Wilusz 2013). Chez les eucaryotes et les bactéries, les protéines Sm/Hfq ont des activités « chaperonnes » permettant l'agencement d'échafaudage moléculaire pour l'assemblage de complexes RNP et qui influencent les structures des ARNs (Vogel and Luisi 2011). La famille de protéines Sm/Hfq chez les archées sont appelées SmAPs. Il existe des paralogues : SmAP1 and SmAP2 et aussi SmAP3 avec un domaine CTD plus long chez certaines Crénarchées comme *Sulfolobus solfataricus* ou *Sulfolobus tokodaii* (Mura et al. 2013). Un homologue bactérien Hfq-like a aussi été identifié dans le génome de *Methanococcus jannaschii* (Nielsen et al. 2007). Des études ont montrés chez *Haloferax volcanii* et *Sulfolobus solfataricus* que le gène *SmAP1* est souvent aux côtés du gène codant la sous unité L37e du ribosome 50S, suggérant un rôle dans la maturation et la stabilisation des ARNr (Mura et al. 2013) et que le gène *SmAP2*, dans la plupart des Crénarchées et Thaumarchées, est en aval du gène codant une méthionine

adénosyltransférase qui pourrait l'impliquer dans la méthylation de l'ARN ou l'ADN. Un rôle dans la maturation ou la biogénèse des ARNt a aussi été suggéré chez *Archaeoglobus fulgidus* et *Sulfolobus solfataricus* puisque les protéines SmAP1 et SmAP2 sont associées à la composante ARN de la RNase P (Törö et al. 2001).

## **6.2 La protéine L7Ae**

L7Ae est un membre de la famille de protéines L7Ae/L30 retrouvée chez les archées et les eucaryotes. Ces protéines partagent la capacité de lier les motifs K-turn ou K-loop (Mao et al. 1999; Ban et al. 2000; Kuhn et al. 2002) (Ref). L7Ae est une protéine de la grande sous unité du ribosome, qui fait également partie des complexes de modification RNP C/D et H/ACA (comme mentionné ci-dessus) et du complexe RNase P. Des données récentes suggèrent que L7Ae peut réguler la traduction en se liant directement à la région 5' non traduite de son propre ARNm (Daume et al. 2017). La présence de L7Ae dans les complexes RNase P, RNP C/D et H/ACA, et le ribosome chez les archées pourrait indiquer une coordination de la régulation des machineries macromoléculaires impliquant différentes facettes de la traduction (modification et maturation des ARNt et ARNr).

## **III. Ribonucléases chez les Archées**

Les acteurs clés du processing et dégradation des ARN sont les ribonucléases (RNases). Les RNases catalysent le clivage exo- ou endoribonucleolytique de ponts phosphodiester et agissent en concertation avec d'autres enzymes telles que les hélicases à ARN, les polymérases poly(A) et les pyrophosphohydrolases. Elles sont classées en deux grands groupes, les hydrolases et les phosphorylases. Les hydrolases nécessitent des molécules d'eau pour catalyser le clivage des liaisons phosphodiesters alors que les phosphorylases utilisent un phosphate inorganique (Pi) et libèrent des nucléosides 5'-diphosphates (NPD). Le nombre de ribonucléases présentes dans les organismes -24 chez *Escherichia coli*, 34 chez *Saccharomyces cerevisiae*, 60 chez l'homme- traduit leur diversité de fonctions. Elles sont essentielles à la maturation et au contrôle qualité des ARN et peuvent être effectrices des voies de signalisation qui contrôlent l'expression génique au niveau post-transcriptionnel. Elles peuvent aussi intervenir contre les invasions de génomes viraux ou contre la propagation des éléments transposables (Stoecklin and Mühlemann 2013). Le renouvellement ou la maturation des ARN



sont souvent initiés par un clivage interne d'une endoribonucléase suivi par des dégradations à partir des extrémités 5' ou 3' sous l'action d'exoribonucléases. Les RNases peuvent être regroupées selon leur spécificité de substrats, la nature des produits ou leurs mécanismes d'action (processive ou distributive).

Une partie de mon travail de thèse a contribué à répertorier les familles de RNases (structures et fonctions) chez les archées dans le cadre d'une revue parue dans le journal FEMS Microbiology (Clouet-d'Orval et al. 2018). La revue est jointe à ce manuscrit ([Publication 1](#)). Une partie des figures du chapitre III paragraphe 1 2, annotées [Figures FEMS](#), font références à celle trouvées dans la revue. Les familles de RNases archées citées dans la revue ont été définies par comparaison avec leurs homologues bactériens et eucaryotes ([Tableau 2 FEMS](#)). Les études phylogénomiques sur chacun des membres des familles de RNases au sein des génomes ont permis d'établir leur occurrence dans la phylogénie des Archées ([Figure 1 FEMS](#)).

## 1. Les EndoRNases

Les RNases universelles RNase P et RNase Z sont ubiquitaires dans les génomes d'archées ([Figure 1 FEMS](#)). Les familles EndA et Nob1 sont présentes chez les eucaryotes et la famille Cas6 chez les bactéries ([Table 2 FEMS](#)).

**RNase P** est une endoribonucléase spécifique de site qui catalyse l'hydrolyse d'un pont phosphodiester à la jonction entre une extrémité 5' simple brin et une région double brin. Dans les trois domaines du vivant, son rôle le plus connu est de cliver la séquence guide en 5' des ARNt précurseurs (pre-ARNt) (Howard et al. 2013). La RNase P est un complexe RNP (RiboNucléoProtéique) assemblé autour d'une sous-unité catalytique ARN (RPR) ([Figure 2A FEMS](#)). Il en existe trois types chez les archées (Type-A, type-M et type-T) caractérisés par deux domaines ARN fonctionnels, le domaine de liaison aux substrats (S-domain) et le domaine catalytique (C-domain) (Gopalan 2007; Howard et al. 2013; Klemm et al. 2016; Samanta et al. 2016). Respectivement, des protéines sous-forme d'hétéro-dimère (RPP21/RPP29) et d'hétéro-tétramère (POP5/RPP30) se lient aux domaines S et C. Il a plus récemment été montré qu'un autre partenaire, RPP38, aussi connu sous le nom de L7Ae, est associé à l'ARN de la RNase P via des motifs K-turn (Fukuhara et al. 2006; Cho et al. 2010; Lai et al. 2014).

**aRNase Z** est une endoribonucléase universelle qui catalyse la maturation en 3' des pré- ARNt. Elle a été classée par similarité de séquences comme appartenant à la famille des  $\beta$ -lactamases métallo-dépendantes (Aravind 1999). La RNase Z est présente dans les trois domaines du vivant et existe sous deux formes : une forme courte (250-350aa) présente dans tous les domaines et une forme longue (750-900aa) retrouvée uniquement chez les eucaryotes provenant probablement d'un évènement de duplication de gène (Vogel et al. 2005). Elle est caractérisée par un domaine  $\beta$ -lactamase interrompu par une petite insertion. Sa structure cristallographique n'a pas encore été résolue mais le modèle que nous avons réalisé révèle un bras flexible qui ressort du classique domaine  $\beta$ -lactamase (**Figure 2C FEMS**).

**EndA** est une endonucléase de l'épissage de l'ARN qui est strictement conservées dans la phylogénie des archées (**Figure 1 FEMS**). C'est une enzyme multimérique qui reconnaît et clive des motifs structuraux spécifiques nommés motif en Bulge-Helix-Bulge (BHB) (**Figure 3 FEMS**) (Thompson and Daniels 1988; Marck and Grosjean 2003). EndA catalyse des étapes critiques du processing des ARNt et ARNr. Il existe quatre formes d'EndA, classées selon leur composition en sous unités, et qui contiennent deux ou quatre sites actifs (**Figure 2B FEMS**).

**aNob1** est une endoribonucléase homologue au facteur d'assemblage du ribosome eucaryote Nob1. Son activité endoribonucléolytique testée *in vitro* montre que chez *Pyrococcus horikoshii* aNob1 clive de façon efficace les substrats ARN contenant le site D des ARN pré-ribosomiques. Sa structure résolue en spectroscopie à résonance magnétique nucléaire (RMN) révèle la présence d'un domaine PIN en N-terminal, typique à de nombreuses nucléases, relié par un linker à un domaine en doigt de zinc de type « Zinc ribon » en C-terminal (**Figure 2C FEMS**) (Veith et al. 2012).

**RNase HII** est une endonucléase ubiquitaire chez les archées (**Figure 1 FEMS**). Chez les bactéries et eucaryotes les RNases H sont impliquée dans la réplication et réparation de l'ADN (Tadokoro and Kanaya 2009; Kochiwa et al. 2007). Les enzymes RNase H clivent les substrats hybrides ARN/ADN avec un mécanisme de catalyse impliquant deux ions métal. Elles ont été classifiées en deux types selon leur séquence d'acides aminés : RNase HI de type 1 et RNase HII de type 2. Il a récemment été montré que RNase HII est impliquée dans la réparation lors de l'incorporation d'un ribonucléotide par erreur chez les archées. (Heider et al. 2017).

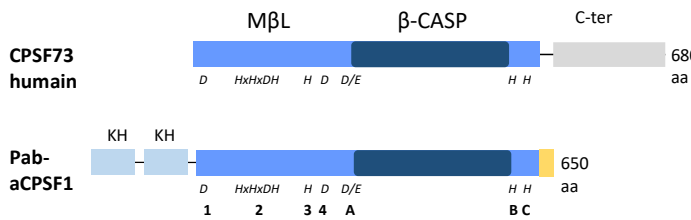
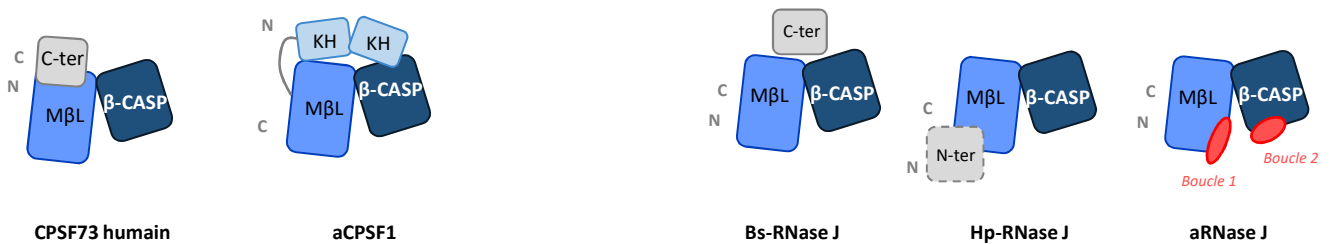
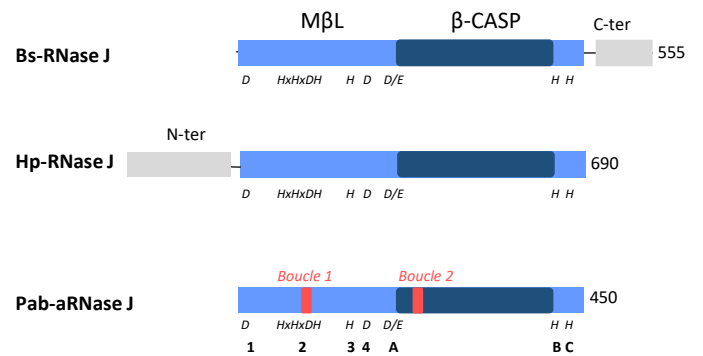
**Cas6** est une des protéines du mécanisme de défense CRISPR-Cas avec Cas5 et Cas7. Ces protéines forment de larges complexes RNP avec l'ARN CRISPR (ARNcr) dédiés à la dégradation

d'ADN et ARN invasifs. L'activité endonucléolytique de Cas6 est nécessaire à la maturation des ARNcr de la classe 1. Cas6 possède deux motifs conservés : un domaine ferredoxin ou RRM (RNA recognition motifs) et un motif riche en glycines en C-terminal (**Figure 2C FEMS**). C'est une nucléase métallo-indépendant dont le site actif prédit de l'enzyme montre des similarités avec le site actif de l'endoribonucléase archée EndA.

## 2. Les ExoRNases 5'-3' : la famille $\beta$ -CASP

La famille  $\beta$ -CASP, qui comprend les enzymes de clivage de l'ARN (CPSF73 et RNase J) et de réparation de l'ADN (Artemis, SNM1 et PSO2), fait partie de la superfamille des métallo- $\beta$ -lactamases, enzymes retrouvées dans les trois domaines du vivant (Dominski 2007). Les membres de cette famille contiennent un domaine métallo- $\beta$ -lactamase (M $\beta$ L) conservé, caractérisé par cinq motifs impliqués dans la liaison de métaux, généralement des ions  $Zn^{2+}$ , et dans la catalyse. Ce domaine est composé de deux feuillets centraux  $\beta$  entourés de cinq hélices  $\alpha$  exposées au solvant. Les enzymes de la famille  $\beta$ -CASP possèdent un domaine  $\beta$ -CASP (un feuillet central de six brins  $\beta$  parallèles entourés de cinq hélices  $\alpha$ ) inséré dans le domaine M $\beta$ L. La poche du site catalytique, qui se forme à l'interface des deux domaines, contient deux cations divalents ( $Zn^{2+}$ ,  $Mn^{2+}$  ou  $Mg^{2+}$ ). Les ions étroitement liés sont coordonnés par des résidus histidines ou acides aspartiques conservés dans les motifs A (Asp ou His), B (His) et C (His) du domaine  $\beta$ -CASP et les acides aminés des motifs 1-4 (principalement Asp) du noyau  $\beta$ -lactamase. Le mécanisme de catalyse proposé correspond à une attaque nucléophile sur le groupe phosphate des liaisons phosphodiester. La majorité des ribonucléases  $\beta$ -CASP partagent une double activité endo- et 5'-3' exoribonucléolytique catalysée par le même site. Cette double fonction enzymatique au sein du même site n'est pas encore bien comprise. Elle pourrait être engendré par la liaison d'un autre facteur ou via l'association dans un complexe multiprotéique.

Parmi les enzymes  $\beta$ -CASP agissant sur l'ARN, deux groupes ont émergés d'après leurs fonctions cellulaires et leur phylogénomie, l'un est lié à la protéine CPSF73 eucaryote et l'autre à la protéine RNase J bactérienne.

**A****B**

**Figure 9. Domaines d'organisation et représentation schématisques des RNases β-CASP.** A- Domaines d'organisation de CSPF73 humaine et aCPSF1 de *Pyrococcus abyssi* (Pab) B- Domaines d'organisation de RNase J de *Bacillus subtilis* (Bs), *Helicobacter pylori* (Hp) et *Pyrococcus abyssi*. Le domaine métallo-β-lactamase (MβL) est coloré en bleu clair et le domaine β-CASP en bleu foncé. Les motifs conservés sont indiqués. Les boucles spécifiques d'aRNase J sont en rose. Le motif responsable de la dimérisation d'aCPSF1 est en jaune. La structure du domaine N-terminal de Hp-RNase J (en tiret) n'a pas été résolue.

## 2.1 Le groupe CPSF

Chez les eucaryotes, le facteur ubiquitaire CPSF73 est une sous unité du complexe multiprotéique CPSF (Cleavage and Polyadénylation specificity factor) impliqué dans la maturation de l'extrémité 3' de tous les transcrits de l'ARNP II et dans la polyadénylation des ARNm. Il catalyse le clivage endoribonucléolitique au niveau du motif CA qui se situe en aval du consensus AAUAAA sur l'ARNm avant la réaction de polyadénylation. Son activité 5'-3' exoribonucléolytique intervient dans la maturation des pré-ARNm des histones chez les métazoaires (Dominski et al. 2005; Dominski and Marzluff 2007). CPSF73 possède un domaine en C-terminal, en plus des domaines M $\beta$ L et  $\beta$ -CASP, dont la structure n'a pas été résolue à ce jour.

Chez les bactéries, il existe des homologues CPSF qui sont distribués de façon sporadique et dont la fonction n'est pas encore décrite (Quentin, Y. & Clouet d'Orval, B., observations non publiées)(Dominski et al. 2013)

Chez les archées, les membres du groupe CPSF sont très nombreux et sont classés en trois groupes d'orthologues distincts (aCPSF1, aCPSF1b et aCPSF2) (Tableau 1 & Fig. 1A FEMS) (Phung et al. 2013; Clouet-d'Orval et al. 2010; Clouet-d'Orval et al. 2015).

**aCPSF1** possède, en plus des domaines M $\beta$ L et  $\beta$ -CASP, un domaine N-terminal accessoire composé de deux motifs KH consécutifs (motifs de liaison aux acides nucléiques) et un motif en C-terminal impliqué dans la dimérisation de l'enzyme (**Figure 9A**). Les structures cristallographiques d'aCPSF1 de *Pyrococcus horikoshii*, *Methanosarcina mazeii* et *Methanlobus thermautotrophicus* présentent une architecture en trois parties dans laquelle les domaines KH sont reliés au noyau central via un long linker. aCPSF1 possède une double activité chez les Thermococcales et Methanococcales : elle est capable de cliver au niveau d'un site CA dans un ARN simple brin et possède une activité 5'-3' exoribonucléolytique restreinte aux substrats ARN 5' monophosphorylés (Phung et al. 2013; Levy et al. 2011). Son activité est comparable avec celle de CPSF73 chez les eucaryotes (Dominski et al. 2013). La présence d'aCPSF1 dans l'ensemble des génomes archées, suggère que son activité est au centre de voies métaboliques conservées et elle a récemment été proposée comme impliquée dans la terminaison de la transcription sous le nom de FttA chez *Thermococcus kodokarensis* (Sanders et al. 2020).

**aCPSF1b** représente un petit groupe de ribonucléases  $\beta$ -CASP, restreint au groupe des Methanococci et Korarchaeota, et est étroitement lié à aCPSF1. Son activité 5'-3' exoribonucléolytique a été montrée chez *Methanococcus janashii*, mais aucune activité endonucléolytique n'a été observée. Les membres de ce groupe ne possèdent pas le domaine N-terminal accessoire (**Figure 4B FEMS**) et aucune structure d'aCPSF1b n'a été résolue à ce jour.

**aCPSF2** est un cluster moins conservé chez les archées comparé au cluster aCPSF1 (**Figure 1 FEMS**). Les membres de ce groupe, retrouvés chez les TACK/Proteoarchaeota et de façon sporadique chez les Euryarchées, sont restreints aux domaines M $\beta$ L et  $\beta$ -CASP excepté chez *Sulfolobus solfataricus* qui possède une extension N-terminale. Actuellement, aucune structure d'aCPSF2 n'a été résolue (un modèle est présenté dans la **Figure 4B FEMS**). aCPSF2 de *Sulfolobus solfataricus* et *Sulfolobus acidocaldarius* possèdent une activité 5'-3' exoribonucléique qui dépend de la présence d'ions  $Mg^{2+}$  (Levy et al. 2011; Märten et al. 2013).

## **2.2 Le groupe RNase J**

Le groupe RNase J est représenté dans les bactéries, les euryarchées et les plantes.

Chez les bactéries, la ribonucléase RNase J qui est présente dans plus de la moitié des espèces, notamment chez les Gram positives, possède un rôle clé dans la maturation des ARNr et le renouvellement des ARNm (Linder et al. 2014; Durand and Condon 2018; Mathy et al. 2007). En plus des domaines M $\beta$ L et  $\beta$ -CASP, RNase J possède une extension généralement en C-terminal (CTD). De nombreuses études enzymatiques et structurales de la RNase J ont permis d'identifier les différences entre les RNase J de différentes espèces. Chez *Bacillus subtilis* le domaine CTD est impliqué notamment dans la dimérisation de l'enzyme (Li de la Sierra-Gallay et al. 2008). RNase J possède une double activité endo- et exoribonucléolytiques 5'-3'. La plupart des Firmicutes possèdent deux paralogues, RNase J1 et RNase J2, dont le rôle *in vivo* n'est pas encore clair (Even et al. 2005). Chez *Staphylococcus aureus* et *Bacillus subtilis*, l'activité exoribonucléolytique de RNase J1 est impliquée dans la dégradation de l'ARNm et dans la maturation de l'extrémité 5' de l'ARNr 16S. En contraste, les RNase J1 chez *Bacillus subtilis* et chez *Staphylococcus aureus* sont capables de dégrader de façon efficace les extrémités trisphosphorylées (Li de la Sierra-Gallay et al. 2008; Even et al. 2005; Hausmann et

al. 2017). Cette différence n'est pour le moment pas expliquée. RNase J1 et RNase J2 ont été proposées comme analogues fonctionnels de la RNase E, endoribonucléase clé chez *Escherichia coli* et chez les bactéries Gram négatives (Even et al. 2005). RNase E permet d'initier la dégradation des ARNm et intervient dans la maturation des ARNr. Elle fait partie du complexe multiprotéique dégradosome à ARN qui fait intervenir l'exoribonucléase 3'-5' PNPase (polynucléotide phosphorylase), l'hélicase à ARN RhlB et l'enzyme glycolytique énoylase (Carpousis et al. 2009).

Chez les plantes, la protéine RNase J semble impliquée dans le contrôle qualité des ARN en éliminant les longs transcrits anti-sens et en dégradant les extrémités 5' des ARN (Hotto et al. 2020; Sharwood et al. 2011). Elle contient une extension en N-terminal (Chloroplast transit peptide) et un domaine de liaison à l'ADN GT-1 en C-terminal (excepté chez les algues) dont le rôle n'est pas encore décrit. Les cyanobactéries qui sont étroitement liées à l'ancêtre évolutif des chloroplastes végétaux contiennent les deux ribonucléases RNase J et RNase E.

**aRNase J** est présente dans la majorité du grand groupe des Euryarchées, à l'exception des Diafoarchées et des Archeoglobi. Contrairement à la RNase J bactérienne qui arbore un domaine CTD, l'aRNase J est restreinte aux domaines M $\beta$ L et  $\beta$ -CASP (**Figure 9B**). Elle possède cependant deux insertions spécifiques, de 9 et 12 acides aminés dans le domaine  $\beta$ -CASP et dans la partie N-terminal du domaine M $\beta$ L respectivement (Loop 1 et Loop 2) (Clouet-d'Orval et al. 2010). Les activités enzymatiques et la résolution de la structure d'aRNase J montrent que ces deux insertions interagissent avec l'ARN et sont importantes pour l'activité ribonucléolytique de l'enzyme (Clouet-d'Orval et al. 2010; Zheng et al. 2017). L'activité 5'-3' exoribonucléolytique d'aRNase J est très processive *in vitro* et a d'abord été identifiée chez les Thermococcales et Methanococci (Clouet-d'Orval et al. 2010; Levy et al. 2011). Par des analyses biochimiques, protéomiques et phylogénomiques, les résultats de l'équipe suggèrent que l'aRNase J participe aux voies de dégradation de l'ARN à proximité du ribosome, en coordination avec l'hélicase ASH-Ski2 et/ou l'exosome à ARN. (Phung et al. 2020)(Phung Thèse 2017).

### 3. Les ExoRNases 3'-5'

La dégradation des ARN via leur extrémité 3' est critique dans les trois domaines du vivant. Dans le domaine des archées, cette activité est assurée par deux exoribonucléases 3'-5' :

l'exosome à ARN et l'aRNase R. La machinerie de l'exosome d'archée, qui est homologue à l'exosome eucaryote, est retrouvée dans la grande majorité des groupes archées excepté les groupes Methanomicrobiales, Haloferacales, Halobactériales et Methanococci. A l'inverse, l'aRNase R, qui est homologue à la RNase R bactérienne, est représentée uniquement chez les Methanomicrobiales, Haloferacales et Halobactériales (**Figure 1 FEMS**) (Clouet-d'Orval et al. 2018). L'absence de l'exosome à ARN 3'-5' dans ces groupes semble être compensée par la présence de la ribonucléase 3'-5' RNase R. Cependant, chez les Methanococci, aucune enzyme avec une activité équivalente n'a été détectée. L'activité 3'-5' pourrait être portée par une enzyme spécifique archée non connue à ce jour ou pourrait être compensée par les autres machineries ribonucléolytiques endo- ou 5'-3' exo.

### **3.1 L'aRNase R**

Les enzymes RNases R-like sont caractérisées par un domaine RNB faisant partie de la famille des RNases II/RNB. La RNase R bactérienne est impliquée dans le contrôle qualité des ARN. La plupart des données sur la RNase R chez les archées proviennent d'études biochimiques *in vitro* à partir d' *Haloferax volcanii* dans lequel il a été montré que le gène codant pour l'aRNase R était essentiel (Portnoy and Schuster 2006; Matos et al. 2012). Les analyses phylogénétiques menées en collaboration avec l'équipe de G. Fichant (LMGM) ont permis d'identifier 3 types de aRNase R nommés arbitrairement a, b et c) qui ne ressemblent pas à la RNase R bactérienne (Clouet-d'Orval et al. 2018). Dans le groupe des Halophiles, aRNase Ra est restreinte au domaine RNB, par contraste avec les deux autres formes RNase Rb et RNase Rc, retrouvées chez quelques méthanococcales, contenant un domaine C terminal accessoire avec des propriétés de liaison à l'ARN (**Figure 4A FEMS**). Nous proposons que ces différentes formes de l'aRNase R sont certainement le résultat de transferts horizontaux de gènes issus des génomes bactériens.

### **3.2 L'Exosome à ARN**

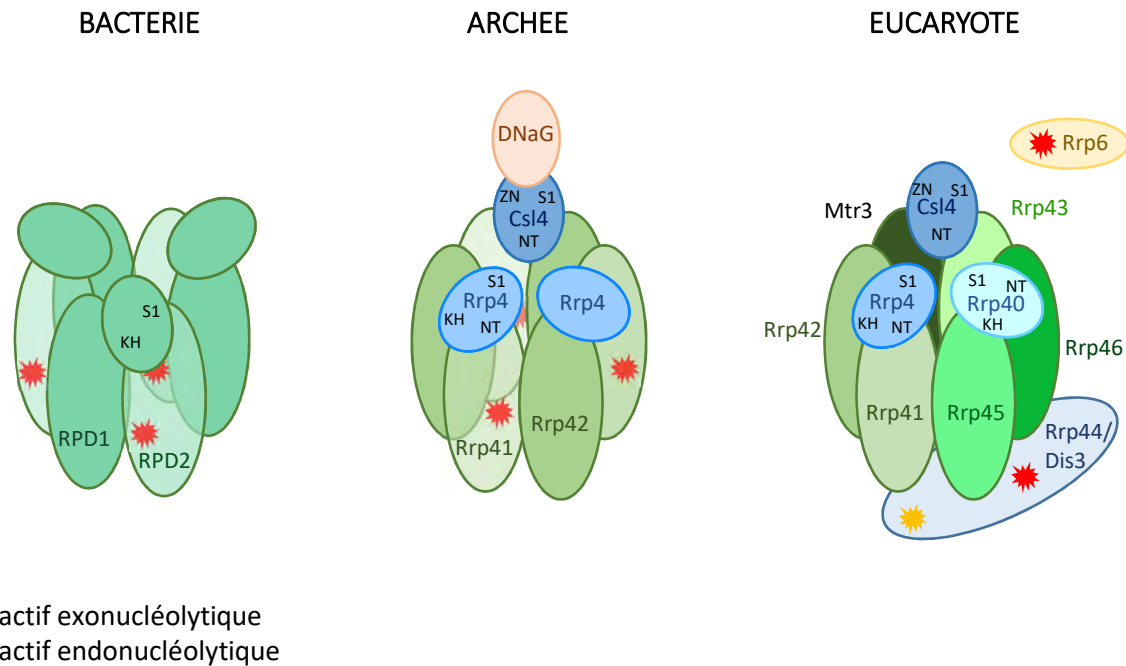
**L'exosome à ARN d'archée** a été isolé pour la première fois chez *Sulfolobus solfataricus* (Evguenieva-Hackenberg et al. 2003) et plusieurs études sur différents organismes, en particulier les Sulfolobales et les Pyrococcales, ont permis sa caractérisation enzymatique et structurale (Evguenieva-Hackenberg et al. 2014; Evguenieva-Hackenberg et al. 2017).



L'exosome d'archée catalyse la dégradation phosphorolytique d'un ARN simple brin à partir de l'extrémité 3' en présence de phosphate inorganique (Pi) et libère un nucléoside 5'-diphosphates (NDPs). Il est aussi capable de faire la réaction reverse en utilisant des NDPs pour synthétiser des hétéro-polymères aux extrémités 3' des ARNs (une queue poly(A)) (Evguenieva-Hackenberg et al. 2014; Evguenieva-Hackenberg et al. 2020). Le core de l'exosome d'archée est composé d'un trimère de dimère Rrp41/Rrp42 qui forme le tonneau et dont le site catalytique est porté par Rrp41 (Figure 10). La coiffe de l'exosome est un hétérotrimère composé des protéines de liaison à l'ARN Rrp4 (qui possèdent un domaine de liaison à l'ARN S1 et KH) et/ou Csl4 (dispose d'un domaine S1 et ZN), dont la stœchiométrie *in vivo* est inconnue (Witharana et al. 2012). *In vitro*, l'exosome-coiffe dégrade les ARN de façon plus efficace que le tonneau hexamérique seul (Büttner et al. 2005; Evguenieva-Hackenberg et al. 2008). Chez *Sulfolobus solfataricus*, la sous-unité Rrp4 confère une forte spécificité poly(A) à l'exosome (Roppelt et al. 2010) alors que la sous-unité Csl4 permet la liaison avec la sous unité aDNaG, composé spécifique de l'exosome archée. aDNaG était annotée comme une primase de type bactérienne, mais les études biochimiques montrent qu'elle fait partie intégrante de l'exosome archée et possède un domaine de liaison à l'ARN (Witharana et al. 2012; Evguenieva-Hackenberg et al. 2008). Le gène codant DNaG est plus conservé que les autres protéines de l'exosome chez les archées, ce qui suggère qu'elle possède aussi un rôle indépendant de l'exosome.

Les connaissances actuelles sur les fonctions de l'exosome reposent uniquement sur des données biochimiques et moléculaires. La manière dont les différentes fonctions associées sont régulées reste inconnue (Ogami et al. 2018). Comme chez les eucaryotes, la sélection et l'accessibilité des différents substrats sont probablement influencés par d'autres protéines qui restent à être identifiées. Pour le moment, une interaction entre l'exosome et le facteur aEF1A a été observée chez *Sulfolobus solfataricus* (Witharana et al. 2012) et une étude montre la liaison entre la protéine Nop5 impliquée dans la méthylation des ARNr et l'exosome –Rrp4 (Gauernack et al. 2017).

Chez les eucaryotes, l'exosome est impliqué dans le processing et la dégradation 3'-5' de tous les types d'ARN (Mitchell et al. 2003). Le core de l'exosome chez l'homme et la levure est catalytiquement inactif et comprend un anneau hexamérique de RNases PH (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46 & Rrp45), et des protéines de liaison à l'ARN Rrp4, Rrp40 et Csl4



**Figure 10. Schéma de la machinerie 3'-5' exoribonucléolytique dans les 3 domaines du vivant.** La PNPase bactérienne (à gauche) forme un tonneau homotrimérique. Chaque monomère contient deux domaines RPD, un domaine S1 et un domaine KH. L'activité catalytique phosphorolytique est portée par la sous unité RPD2. Les domaines KH et S1 forment une coiffe de liaison à l'ARN sur le dessus de la PNPase. L'exosome archée (au milieu) est composé de trois dimères Rrp41-Rrp42 formant un tonneau hexamérique. La coiffe est constituée des protéines Rrp4, Csl4 et DNaG de stœchiométrie variable *in vitro*. Les trois sites catalytiques sont portés par les sous unités Rrp41. L'exosome eucaryote (à droite) est constitué de neuf sous unités formant un tonneau hexamérique. Les protéines Csl4, Rrp4 et Rrp40 constituent la coiffe de l'exosome eucaryote. Son activité dépend des ribonucléases Rrp44/Dis3 et de Rrp6 (noyau). Les protéines homologues sont de la même couleur.

(**Figure 10**). Ces sous unités forment un canal conservé qui achemine les substrats ARN simple brin vers l'exoribonucléase Rrp44/Dis3 processive. Rrp44/Dis3 possède des activités hydrolytiques endo et exoribonucléolytiques (Lebreton and Séraphin 2008; Schaeffer et al. 2009; Halbach et al. 2012). Dans le noyau, l'exosome possède une sous unité catalytique supplémentaire : l'exoribonucléase 3'-5' Rrp6.

Les hélicases à ARN de la famille Ski2, Mtr4 et Ski2, sont des facteurs auxiliaires aidant à l'acheminement du substrat ARN vers à l'exosome nucléaire et cytoplasmique respectivement (Halbach et al. 2012; Johnson and Jackson 2013; Weick et al. 2018; Zinder and Lima 2017). Dans le cytoplasme l'exosome s'associe avec le complexe Ski2-Ski3-Ski8 (Ski) (Halbach et al. 2013) qui est principalement impliqué dans la dégradation des ARNm et la surveillance des ARNm en cours de traduction (voies de dégradation appelées No Go Decay, Non Stop Decay et Non Mediated Decay) (Chlebowski et al. 2013). Récemment, des études en cryo-microscopie ont montré l'interaction directe du complexe Ski avec le ribosome 80S et proposent un modèle de recrutement de l'exosome sur le ribosome bloqué pendant la traduction (Schmidt et al. 2016). L'exosome nucléaire s'associe aux complexes protéiques TRAMP, NEXT et PAXT, formés autour de l'hélicase Mtr4, et participe à la maturation en 3' et à la dégradation d'une grande variété de substrats ARN dont des ARNr (Chlebowski et al. 2013; Ogami et al. 2018; Wyers et al. 2005; Lubas et al. 2015; Meola et al. 2016; Schuller et al. 2018). Chez les bactéries, la PNPase forme un anneau homotrimérique qui ressemble à l'exosome eucaryote et d'archées (**Figure 10**) (Shi et al. 2008). Elle est impliquée dans le contrôle qualité et le renouvellement des ARN et est responsable de la synthèse post-transcriptionnelle de courtes queues déstabilisant l'ARN, qui servent d'armature pour une interaction efficace avec les exoribonucléases (Slomovic and Schuster 2008). Chez *Escherichia coli*, la PNPase est retrouvée soit en complexe avec l'hélicase à ARN RhIB soit liée à l'endoribonucléase RNase E comme faisant partie du dégradosome (Lin-Chao et al. 2007).

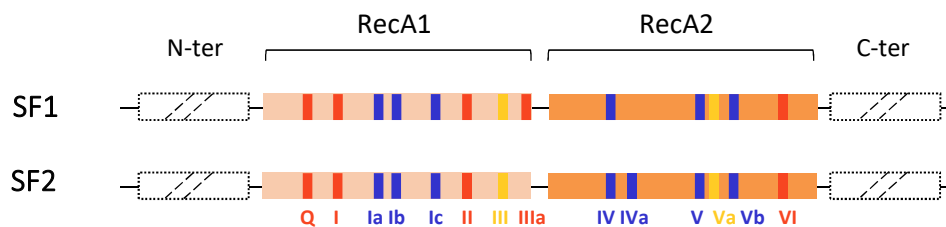
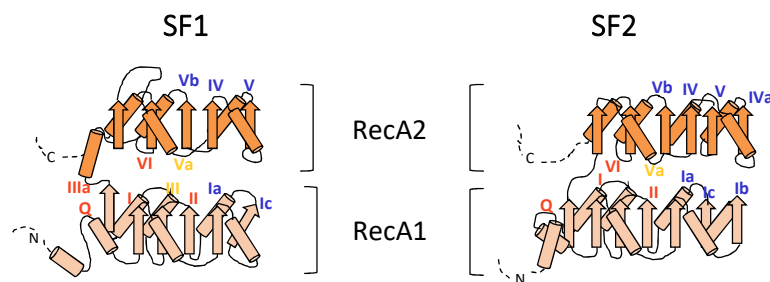
#### **IV. Les Hélicases à ARN chez les Archées**

Les hélicases sont impliquées dans tous les aspects du métabolisme de l'ADN et de l'ARN comprenant la réplication, la réparation, la recombinaison, la transcription et le remodelage de la chromatine (Khemici and Linder 2018; Jarmoskaite and Russell 2014; Cordin and Beggs 2013; Zinder and Lima 2017; Jankowsky and Fairman 2007). Ces enzymes sont des moteurs

moléculaires qui utilisent l'énergie de l'hydrolyse de nucléosides triphosphates (NTP, le plus souvent de l'ATP) pour fixer et remodeler les acides nucléiques ou les complexes protéines-acides nucléiques (Berger 2008) en se déplaçant de manière unidirectionnelle le long des molécules d'ADN ou d'ARN double brin ou simple brin (Gao et al. 2019). D'après leur séquence et caractéristiques structurales, les hélicases ont été répertoriées en six superfamilles (de SF1 à SF6) (Berger 2008). Les membres des familles SF1 et SF2 sont monomériques et partagent un noyau structural commun. Les hélicases SF3 à SF6 forment toutes des anneaux hexamériques et sont divisées en ATPase de type RecA (SF4 et SF5) ou AAA+ (SF3 et SF6). Le groupe SF3 comprend des hélicases de virus. Les groupes SF4 et SF5 sont des hélicases impliquées dans la réplication et la terminaison de la transcription respectivement. Le groupe SF6 inclut notamment l'hélicase MCM (minichromosome maintenance protein complex). Les hélicases sont traditionnellement connues pour dérouler l'ADN ou l'ARN double brin de manière ATP-dépendante (activité dite « unwinding »). De manière remarquable des évidences expérimentales suggèrent que certaines d'entre elles ont aussi la capacité de faciliter l'assemblage de brins complémentaires d'acides nucléiques (activité dite « annealing ») (Wu 2012; Linder and Jankowsky 2011) (**Tableau 2**). Par exemple, l'hélicase à ARN II nucléaire (également connu sous le nom de Gu) favorise le repliement intramoléculaire de l'ARN indépendamment de l'ATP (Valdez et al. 1997). Les protéines de la famille DEAD-box Ded1 et Mss116, qui sont respectivement importantes pour la traduction et le métabolisme de l'ARN mitochondrial, possèdent une activité « annealing » *in vitro* (Yang and Jankowsky 2005). Chez les cyanobactéries l'hélicase CrhR catalyse la réaction d'appariement de brins complémentaires en présence d'ATP (Chamot et al. 2005). Dans la suite, nous nous intéressons plus particulièrement aux superfamilles SF1 et SF2 reconnues pour leur rôle critique dans le métabolisme de l'ARN.

Organismes	Hélicases	Superfamilles	Activité Unwinding	Activité Annealing	Références des activité annealing	
	RNA Helicase II/Gu		✓	✓	Valdez 1997	
Human and Yeast	p68 family (Ddx5,p72,Ddx42p)		✓	✓	Rössler 2001; Uhlmann-Schiffler 2006	
<b>ARN</b>	DED1/DDX3X/LAF-1	SF2	✓	✓	Yang and Jankowsky 2005	
	Cyanobacteria CrhR		✓	✓ (ATP)	Chamot 2005	
	Yeast Mss116p		✓	✓	Halls 2007	
	Human and Yeast RNA Helicase A/DHX9		✓ (3'-5')	✓	Xing 2011	
	Virus NS3	SF3	✓	✓	Gebhard 2012	
Human	RECQ5β		✓ (3'-5')	✓	Garcia 2004	
Drosophila, Human and Yeast	BLM/Sgs1/RECQ2/mus309	SF2	✓ (3'-5')	✓	Machwe 2005; Cheok 2009	
Human	RECQ1		✓ (3'-5')	✓	Sharma 2005; Muzzolini 2007	
Human and Yeast	Dna2	SF1	✓ (5'-3')	✓	Masuda-Sasa 2006	
Human	CSB			✓	Muftuoglu 2006	
Phage	UvsW	SF2	✓ (3'-5)	✓	Nelson and Benkovic 2007	
Archaea	Hel112/Lhr2		✓ (Monomer)	✓	De Felice 2007	
Human	Pif1 family	SF1	✓ (5'-3')	✓	Gu 2008	
<b>ADN</b>	Human and drosophila	HARP/SMARCAL1		✓ (ATP)	Bansbach 2009; Yuang 2009; Yusufzai 2008	
	Archaea	Hjm/Hel308A		✓ (3'-5')	Li 2008	
		WRN/RECQ3	SF2	✓ (3'-5')	✓	Muftuoglu 2008
		RECQ4		✓ (3'-5')	✓	Xu and Liu 2009; Suzuki 2009
	Human	AH2 (ZRANB3)			✓ (ATP)	Yusufzai 2010
	TWINKLE	SF4	✓ (5'-3')	✓	Sen 2012	
Mycobacterium Tuberculosis	XPB		✓ (3'-5')	✓	Balasingham 2012	
Fungi and Yeast	Hrq1	SF2	✓ (3'-5')	✓	Kwon 2012	
Human	RECQL5		✓ (3'-5')	✓	Khadka 2016	

**Tableau 2. Les activités annealing des Hélicases.** Les organismes décrits sont ceux dans lesquels ont été montré l'activité annealing (appariement de brins d'acides nucléiques) des hélicases. La polarité de l'activité unwinding (déroutement de duplex d'acides nucléiques) et la dépendance d'ATP pour l'activité annealing sont indiquées.

**A****B**

**Figure 11. Motifs et organisation structurale des hélicases SF1 & SF2.**

**A-** Séquence d'organisation du noyau hélicase des SF1 et SF2. Le noyau SF1/SF2 est composé de deux domaines RecA. Les motifs caractéristiques sont colorés selon leur fonction prédominante : en rouge les motifs impliqués dans la liaison et l'hydrolyse de l'ATP, en jaune les motifs qui ont un rôle dans la coordination entre les sites de liaisons des NTP et des acides nucléiques, en bleu les motifs qui lient les acides nucléiques. La taille des domaines et la distance entre chaque motif n'est pas à l'échelle.

**B-** Position des motifs caractéristiques du noyau hélicase des SF1 et SF2. Les brins sont représentés par des cylindres et les hélices par des flèches. Les motifs conservés sont colorés comme dans le panel A et B.

## 1. Les Familles SF1-SF2

Les hélicases SF1 et SF2 sont caractérisées par deux domaines centraux qui constituent le noyau catalytique portant l'activité hélicase : les domaines pfam « DEAD » et pfam « Helicase C ». Ces deux domaines RecA1 et RecA2 se structurent en domaine de type « RecA-like » constitués de cinq feuillets  $\beta$  entourés de quatre hélices  $\alpha$ , reliés par un linker avec au moins 12 motifs communs. Les motifs les plus conservés sont les résidus qui coordonnent la liaison et l'hydrolyse de l'ATP (motifs I, II, nommés aussi Walker A et B, et VI) (**Figure 11**). Dans la structure, ces résidus sont situés au niveau du sillon formé entre les deux domaines RecA. La liaison de l'ATP, en présence du substrat ARN ou ADN, entraîne une conformation dite « fermée » qui provoque une rotation de 20° entre les deux domaines RecA. L'hydrolyse de l'ATP et le relargage de l'ADP conduisent au retour de la conformation « ouverte » (Jarmoskaite and Russell 2014). Ces changements de conformations peuvent conférer aux hélicases une activité « unwinding » ou « annealing » (Yang and Jankowsky 2005) Les motifs IIIa et Q-motif coordonnent la base adénine de l'ATP. Les motifs III et Va permettent la coordination entre la liaison du NTP et des acides nucléiques. Les motifs Ia-c, IV, IVa, V et Vb, sont impliqués dans la liaison aux acides nucléiques, principalement au niveau du squelette sucre-phosphate (**Figure 11**).

La majorité des hélicases SF1 et SF2 possèdent des domaines accessoires à leur extrémité N- et/ou C-terminal avec des fonctions spécifiques telles que des domaines portant des activités enzymatiques (nucléases ou protéases) ou ayant des capacités de liaison aux acides nucléiques (ARN ou ADN) (Zinc finger, OB-fold) ou encore engagés dans des interactions protéiques (CARD) (Fairman-Williams et al. 2010).

Les hélicases SF1 et SF2 sont divisées en sous-famille selon leurs motifs conservés, leur structure et leurs activités. Le groupe SF1 contient trois sous familles : UvrD/Rep, Pif1-like et Upf1-like ; et le groupe SF2 en comprend dix : RecG-like, RecQ-like, Rad3/XPD, Ski2-like, T1R, Swi/Snf, RIG-I-like, DEAD-box, DEAH/RHA et NS3/NPH-II (Fairman-Williams et al. 2010). Même si les hélicases SF1 et SF2 partagent des similarités de structures, elles jouent un rôle dans un large éventail de fonctions et agissent sur divers substrats. Certaines sous-familles peuvent inclure des hélicases à ARN et à ADN, d'autres exclusivement des hélicases soit à ADN (RecQ-like), soit à ARN (DEAD-box).

Les hélicases SF1 et SF2 sont des acteurs majeurs des voies de maturation et de dégradation de l'ARN dans les trois domaines du vivant (Jankowsky and Fairman 2007). Elles font notamment partie des machineries moléculaires comme le dégradosome, le spliceosome et le ribosome (Jarmoskaite and Russell 2014; Linder and Jankowsky 2011; Khemici and Linder 2018). Les dégradosomes bactériens, qui sont notamment impliqués dans la maturation et la dégradation des ARNm, sont associés à des hélicases de la famille DEAD-box. Le dégradosome d'*E.coli* est composé de la RNase E (endonucléase), la PNPase (exoribonucléase 3'-5'), l'énolase et l'hélicase RhlB. RhlB facilite la dégradation des ARN par la PNPase. Les Firmicutes telles que *Bacillus subtilis* et *Staphylococcus aureus* ont aussi un complexe de type dégradosome composé de RNase J (une exo- et endonucléase 5'-3'), RNase Y (endonucléase), la PNPase (exoribonucléase 3'-5'), l'énolase, la phosphofructokinase, et l'hélicase à ARN CshA (Khemici and Linder 2018). La biogenèse des ribosomes chez *E. coli* implique aussi des hélicases de la famille DEAD-box : DbpA, SrmB, CsdA et RhlE. DbpA permet le réarrangement des ARN dans le précurseur 50S, SrmB semble empêcher l'appariement de l'ARN 5S et favorise le repliement de l'ARN 23S natif. CsdA fonctionne comme une protéine chaperonne de la biogénèse des ribosomes mais intervient aussi dans la traduction et la dégradation des ARNm (Jarmoskaite and Russell 2014).

Les cellules humaines expriment environ 70 hélicases à ARN et la biogénèse des ribosomes chez *Saccharomyces cerevisiae* impliquent 19 hélicases, dont 17 sont essentielles. Certaines jouent un rôle dans la maturation des pré-ARNr en remodelant les structures aux sites de coupure par des ribonucléases. Par exemple, l'hélicase DEAD-box Dbp3 facilite le clivage de l'ARN par la RNase MRP pour la maturation de l'ARN 25S (Jarmoskaite and Russell 2014). L'épissage des pré-ARNm chez les eucaryotes est un processus qui également requiert un très grand nombre d'hélicases à ARN dans le cadre de l'assemblage du spliceosome (réarrangement de structures d'ARN et dissociation de protéines). C'est le cas de l'hélicase DEAD-box Prp2 et de la Ski2-like Brr2 (Jarmoskaite and Russell 2014; Boneberg et al. 2019; Sloan and Bohnsack 2018).

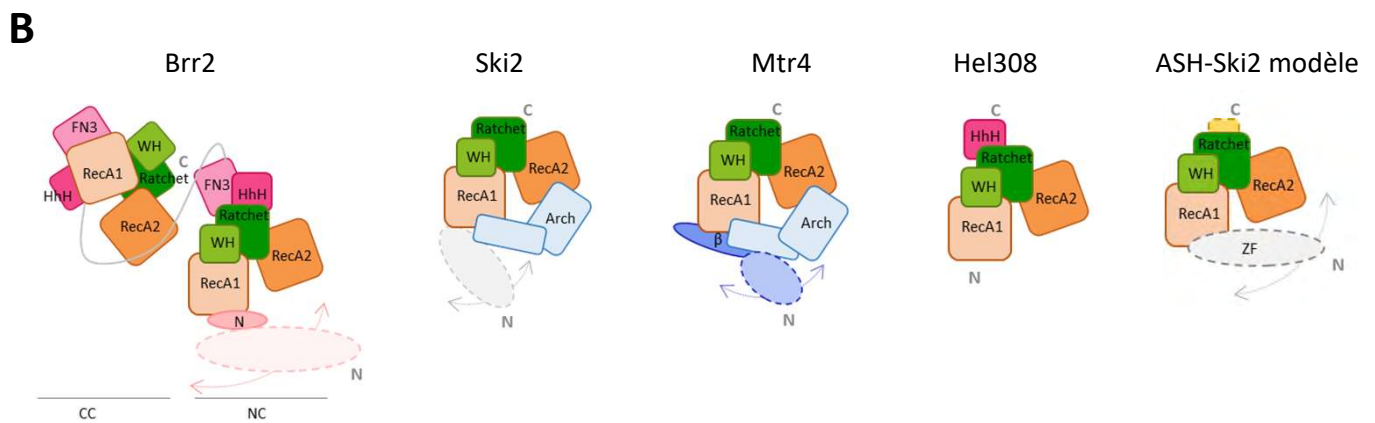
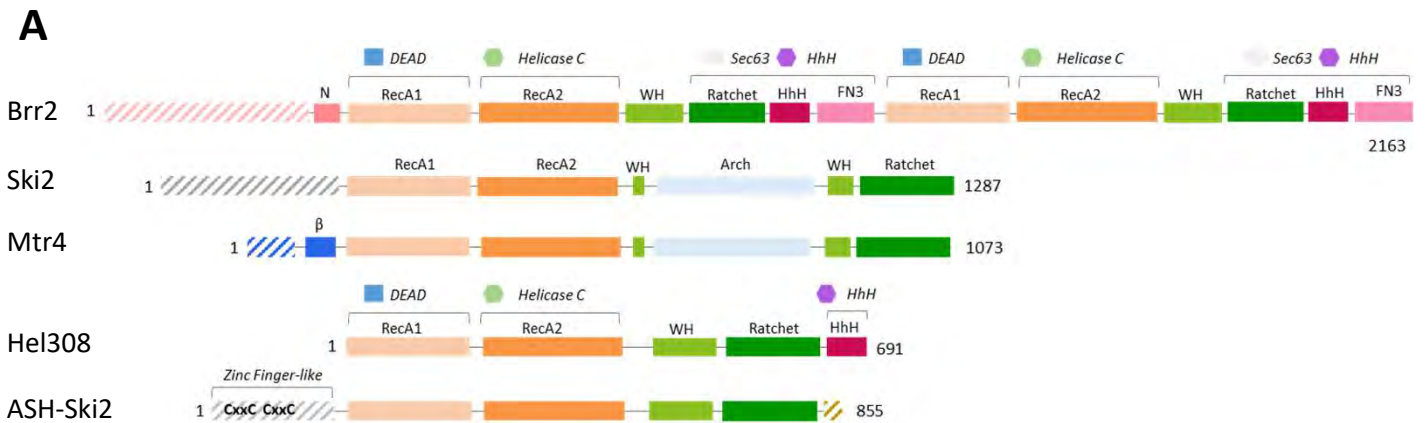
Les différents groupes d'hélicases SF1 et SF2 d'archées ont été mis en évidence par des analyses comparatives de génomes (Chamieh et al. 2016; Hajj et al. 2019) (**Publication Annexe : Hajj M., El-Hamaoui S., Batista M. et al. 2019**). La plupart des familles SF1 et SF2 sont représentées chez les archées à l'exception des familles Pif1-like, RecG-like et DEAH/RHA. La



famille SF1 est représentée chez les archées par les sous familles UvrD-like et Upf1-like qui sont présentes uniquement chez les Euryarchées. Upf1-like est impliquée dans la réplication et la stabilité de l'ADN (Higashibata et al. 2003; Gloor et al. 2012). Sa fonction n'est pas connue chez les archées. La famille UvrD-like est retrouvée dans les trois domaines du vivant, elle est impliquée dans la réparation de l'ADN chez les bactéries et les eucaryotes (Chamieh et al. 2016; Rouillon and White 2011). Son rôle physiologique chez les archées n'est pas déterminé mais les protéines UvrA/B/C chez les halophiles semblent jouer un rôle dans la réparation de l'ADN (Giroux and MacNeill 2016; Crowley et al. 2006). La famille Lhr (initialement classée dans le groupe RecQ-like), constituée des protéines Lhr1 et Lhr2, est ubiquitaire dans les génomes archées. Chez les bactéries, Lhr possède une activité de déroulement des duplexes ADN/ADN et ARN/ADN. Le rôle des hélicases Lhr n'est pas encore décrit chez les archées. Dans l'équipe, Mirna Hajj, étudiante en thèse, étudie cette famille d'hélicase. Les protéines DEAD-box sont des hélicases à ARN présentes dans les trois domaines du vivant. Elles sont impliquées dans de nombreux processus tels que la transcription et traduction des ARNm et la biogénèse des ribosomes. Elles sont présentes chez toutes les Euryarchées à l'exception des Pyrococcales et Halobactéries (Chamieh et al. 2016). Elles ont été peu étudiées chez les archées mais la délétion du gène codant pour l'hélicase DeaD chez *Thermococcus kodakarensis* induit une diminution de la croissance cellulaire (Nagaoka et al. 2013). Finalement la famille Ski2 représentée par les hélicases de type Hel308/Hjm et ASH-Ski2 est ubiquitaire chez les archées. L'équipe s'intéresse plus spécifiquement à l'hélicase ASH-Ski2, objet de cette thèse, puisqu'elle est partenaire de la ribonucléase aRNase J chez les Thermococcales (Phung et al. 2020).

## 2. Les hélicases Ski2-like

La famille Ski2-like, nommée ainsi par rapport à son premier membre identifié (l'hélicase Ski2 eucaryote) constitue une famille restreinte d'hélicase SF2 retrouvée uniquement chez les eucaryotes et les archées. Chez les eucaryotes les hélicase Ski2-like interviennent dans de larges complexes multiprotéiques et jouent un rôle fondamental dans la dégradation, la maturation et l'épissage des ARNs. Les Ski2-like possèdent 12 motifs conservés dans leur domaine central SF2 (RecA1 et RecA2) et sont caractérisées par la présence en C-terminal d'un domaine Winged helix (WH) et d'un domaine Ratchet (**Figure 12**) (Johnson and Jackson 2013).



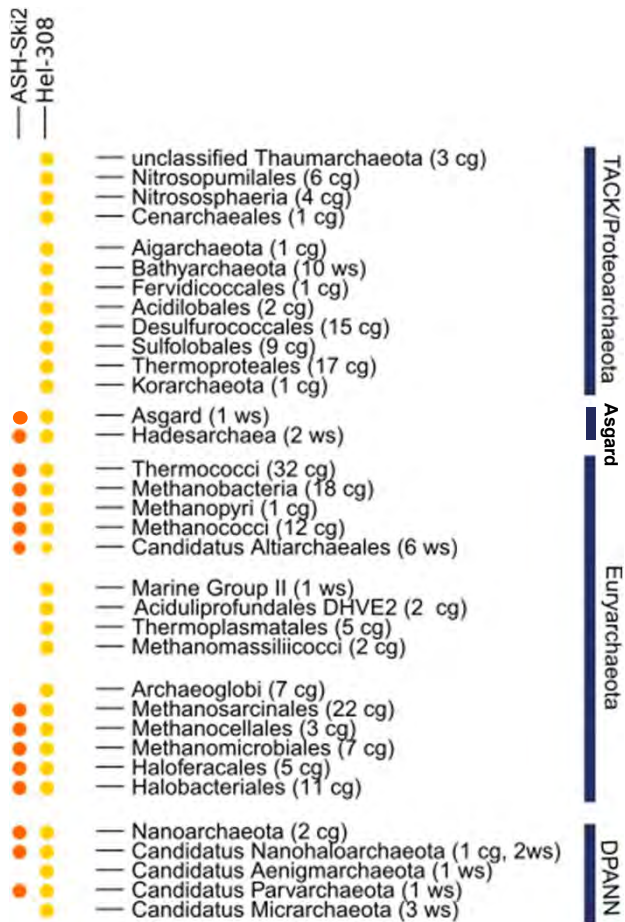
**Figure 12. Architecture de la famille des Ski2.** A- Organisation des domaines des hélicases Ski2-like. Le nombre d'acides aminés provient de Brr2 Humain, Ski2 et Mtr4 de *Saccharomyces cerevisiae*, Hel308 de *Pyrococcus furiosus* et ASH-Ski2 de *Pyrococcus abyssi*. Les Pfam sont indiqués et les domaines sont à l'échelle. Les hélicases de la famille Ski2 possèdent un noyau commun composé des domaines RecA1 (orange clair), RecA2 (orange), Winged Helix (WH, en vert clair) et Ratchet (vert). Elles peuvent aussi présenter des domaines accessoires : Helix loop Helix (HhH) ou Helix hairpin Helix (HhH) (rose foncé) domaine Fibronectine de type III (FN3) (rose clair),  $\beta$  hairpin (bleu). Brr2 contient 2 cassettes. Ses domaines Ratchet, HhH et FN3 constituent le domaine Sec63 (Pfam: PF02889). Mtr4 et Ski2 ont un domaine Arch (bleu clair, PF13234) inséré dans le domaine WH. Les domaines dont la structure n'a pas été résolue sont hachurés. La structure de ASH-Ski2 n'est pas connue. Son domaine N-terminal, qui contient 4 cystéines strictement conservées chez les Euryarchées, pourrait former un domaine Zinc finger (gris). B- Représentation schématique de l'architecture des hélicases Ski2. L'arrangement architectural est basé sur des structures connues (Brr2 pdb: , Mtr4 pdb:4qu4, Ski2 pdb:, Hel308 pdb: 2zj2) à l'exception de ASH-Ski2 qui est basé sur un modèle Phyre2 (25% identité avec Brr2). Les 2 cassettes de Brr2 sont indiquées: cassette N-terminal (NC) et cassette C-terminal (CC). Les domaines non résolus sont en pointillés. Les extrémités N et C-terminal ends sont indiquées.

Ces deux domaines sont impliqués dans la liaison aux acides nucléiques et forment un anneau autour de l'extrémité 3' de l'ARN ou l'ADN. Les Ski2-like peuvent aussi présenter des domaines accessoires. Toutes les structures Ski2-like connues à ce jour, en complexe ou liée à leur substrat, sont listées dans le **tableau 3**.

Les hélicases de type Ski2 identifiées possèdent une activité de déroulement des duplexes avec une polarité 3' vers 5'. L'activité de déroulement et le mécanisme de translocation le long de l'ARN ou ADN a été proposé à partir de différentes structures cristallographiques de l'hélicase Hel308 d'*Archeoglobus fulgidus* (Büttner et al. 2007; Pyle 2008) (Tanner and Linder 2001; Craig et al. 2017; Gao and Yang 2019). L'hélicase Hel308 d'archée est la seule structure entière résolue avec un domaine Helix-hairpin-Helix (HhH) en périphérie, qui interagit avec l'extrémité 3' de l'ADN pour agir comme un « frein moléculaire » auto-inhibiteur pour réprimer l'activité de l'hélicase (Büttner et al. 2007; Pyle 2008; Richards et al. 2008).

Les hélicases Ski2 et Mtr4 eucaryotes contiennent une insertion dans le domaine WH, nommé « Arch » (Halbach et al. 2012; Weir et al. 2010). Ce domaine forme une courbure mobile et se termine en une région globulaire qui est positionnée au-dessus du site d'entrée de l'ARN. L'hélicase à ARN Brr2 eucaryote possède aussi le domaine HhH qui, avec le domaine Fibronectine III (FN3), forme le domaine Sec63 (Vester et al. 2020; Zhang et al. 2009). Le domaine FN3 est généralement associé aux interactions protéines-protéines mais le rôle du domaine Sec63 chez Brr2 reste encore incertain. Elle contient deux copies du noyau hélicase-domaine Sec63. Chez les eucaryotes, Brr2 est impliquée dans l'épissage des pré-ARNm. Elle permet la conversion du complexe spliceosome pré-catalytique (Complexe B) en complexe actif (Complexe Bact) (Vester et al. 2020). La structure du domaine N-terminal de Mtr4, Ski2 et Brr2 n'a pas encore été résolue.

Les hélicases Mtr4 et Ski2, sont associées à l'exosome nucléaire ou cytoplasmique respectivement. Elles facilitent le recrutement et l'adressage des substrats ARN en déroulant les structures secondaires d'ARN et en libérant les complexes RNP pour rendre accessible l'extrémité 3' des substrats à l'exosome à ARN (Falk et al. 2017; Weick et al. 2018; Halbach et al. 2012). Mtr4 interagit directement avec les sous unités Mpp6 et Rrp6 de l'exosome à ARN et intervient dans les complexes TRAMP, PAXT et NEXT impliqués dans la dégradation des ARN nucléaires (Patrick et al. 2017; Bernstein et al. 2008; Weick et al. 2018; Meola et al. 2016). Ski2 est associé à l'exosome à ARN par le complexe Ski composé des protéines Ski3, Ski8 et Ski7 et



**Figure 13.** Distribution taxonomique des hélicases Ski2-like Hel308 et ASH-Ski2 à travers la phylogénie des archées.

sont impliquées dans la dégradation des ARNm dans le cytosol. Il a récemment été montré que le complexe Ski interagit directement avec la sous unité 40S du ribosome (Halbach et al. 2012; Halbach et al. 2013; Schmidt et al. 2016; Zinoviev et al. 2020).

Hel308 (aussi nommée Hjm) est une hélicase à ADN, retrouvée chez les archées et les eucaryotes, qui est impliquée dans la réplication de l'ADN. Même si une étude chez *Pyrococcus horikoshii* montre qu'elle est capable de lier des substrats ARN (Zhang et al. 2008), les données chez *Methanothermobacter thermautotrophicus*, *Sulfolobus tokodaii* et chez l'homme démontrent qu'Hel308 cible préférentiellement les fourches de réplication (Guy and Bolt 2005; Tafel et al. 2011; Li et al. 2008). Son activité « unwinding » de polarité 3' vers 5' chez l'homme est stimulée par la présence de la protéine de réplication RPA (Tafel et al. 2011). Des données montrent que chez *Sulfolobus Tokodaii*, Hel308 interagit avec l'endonucléase spécifique des jonctions de holliday, Hjc (Li et al. 2008) et que chez *Sulfolobus islandicus* Hel308 interagit avec l'ATPase PINA impliquée dans les jonctions de Holliday (Zhai et al. 2018). Chez les archées, des études phylogénomiques, en collaboration avec l'équipe de G. Fichant, permettent de montrer que la famille des hélicases Ski2 chez les archées se divisent au moins en deux grands groupes : le groupe homologue à Hel308 et un groupe spécifique nommé « archaeal specific helicase » (ASH) ou encore ASH-Ski2 qui possède en plus des domaines caractéristiques des Ski2-like, un domaine en N-terminal composé de quatre cystéines strictement conservée (**Figure 13**). Hel308 est codée dans l'ensemble des génomes d'archées contrairement à ASH-Ski2 retrouvée dans la majorité des Euryarchées et Asgards et dans certaines DPANN. L'activité et la fonction d'ASH-Ski2 n'est pas encore connue. Mes travaux de thèse ont en partie consisté à décrire les activités enzymatiques de ASH-Ski2 de *P. abyssi*.

Protéines	Organismes	Pdb	Substrats	Complexes	Fonctions	Références	
Mtr4	S.cerevisiae	2XGJ	RNA/ADP	Trf4-Air2 TRAMP complex	Nuclear RNA quality control	Wier et al 2012	
		4QU4				Taylor et al 2014	
		4U4C	Falk et al 2014				
		5VZJ	Wasmuth et al 2017				
		5OKZ	Falk et al 2017				
		5OOQ	Falk et al 2017				
	Human	6FSZ	Exosome	Schuller et al 2018			
		6D6Q, 6D6R	Exosome bound to RNA	Weick et al 2018			
		6C9O	ZCCHC8-CTD	Puno and Lima 2018			
		6IEG		Wand et al 2019			
		6IEH	with NRDE2	Wand et al 2019			
	6RO1	NVL and ZCCHC8	Lingaraju et al 2019				
Ski2	S.cerevisiae	4A4Z	ADPNP	Ski2-3-8 complex	Cytoplasmic RNA quality control	Halbach et al 2012	
		4A4K					
		4BUJ				Halbach et al 2013	
Brr2	S.cerevisiae	3HIB	ADPNP	sec63 domain	RNA splicing	Zhang et al 2009	
		3IM1,3IM2		sec63 domain		Pena et al 2009	
		5DCA		Prp8 jab1 domain		Absmeier et al 2015	
		5GAN, 5GAO, 5GAP		U4/U6-U5 tri snRNP		Nguyen et al 2016	
		3JCM		U4/U6-U5 tri snRNP		Wan et al 2016	
		5NRL		B complex		Plaschka et al 2016	
		5GM6		B <sub>act</sub> complex		Yan et al 2016	
		5LQW		B <sub>act</sub> complex		Rauhut et al 2016	
		5LJ5		C complex		Galej et al 2016	
		5M52		Prp8		Absmeier et al 2017	
		H.sapiens		4F91			Santos K.F et al 2012
				4KIT		Prp8 Jab1/MPN domain	
				3JCR		U4/U6-U5 tri snRNP	Agagonov et al 2016
	5O9Z		B complex	Bertram et al 2017			
		5XJC	C* complex	Zhang et al 2017			
		5YZG	C complec	Zhan et al 2018			
		5Z57	B <sub>act</sub> complex	Zhang et al 2018			
C.termophilum	6QWS		Absmeier et al 2020				
	6QV4	ATPyS					
	6QV3	ADP					
Hel308	A. fulgidus	2P6R	DNA	ADPNP	DNA replication	Buttner et al 2007	
		2P6U					
	P.horikoshii	2Z41				Zhang et al 2008	
	S.solfataricus	2VA1				Richards et al 2008	
		2ZJA					
		2ZI8, 2ZI2				Oyama et al 2009	
		2ZJ5	ADP				
	S.islandicus	5YWW				Zhai et al 2018	

**Tableau 3. Structures des hélicases Ski2-like.** Les structures résolues de Mtr4, Brr2 et Hel308 en présence de substrats ou en complexes avec d'autres protéines et les numéros d'accension pdb sont listés.

**V. Publication 1 : Clouet d'Orval B., Batista M. et al. 2018**

**Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea  
FEMS Microbiology Reviews**







REVIEW ARTICLE

# Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea

Béatrice Clouet-d'Orval<sup>1,\*</sup>, Manon Batista<sup>1</sup>, Marie Bouvier<sup>1</sup>, Yves Quentin<sup>1</sup>, Gwennaele Fichant<sup>1</sup>, Anita Marchfelder<sup>2</sup> and Lisa-Katharina Maier<sup>2</sup>

<sup>1</sup>Laboratoire de Microbiologie et de Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, 31062 Toulouse, France and <sup>2</sup>Biology II, University of Ulm, 89069 Ulm, Germany

\*Corresponding author: Laboratoire de Microbiologie et de Génétique Moléculaires, (LMGM), Centre de Biologie Intégrative (CBI), Université de Toulouse, 118 route de Narbonne, CNRS, 31062 Toulouse, France. Tel: 33 561 33 58 75; E-mail: [Beatrice.Clouet-Dorval@ibcg.biotoul.fr](mailto:Beatrice.Clouet-Dorval@ibcg.biotoul.fr)

**One sentence summary:** The authors provide a survey of archaeal ribonucleases and RNA-binding proteins and a state of the art on RNA-processing pathways in Archaea.

Editor: Sonja-Verena Albers

## ABSTRACT

RNA-processing pathways are at the centre of regulation of gene expression. All RNA transcripts undergo multiple maturation steps in addition to covalent chemical modifications to become functional in the cell. This includes destroying unnecessary or defective cellular RNAs. In Archaea, information on mechanisms by which RNA species reach their mature forms and associated RNA-modifying enzymes are still fragmentary. To date, most archaeal actors and pathways have been proposed in light of information gathered from Bacteria and Eukarya. In this context, this review provides a state of the art overview of archaeal endoribonucleases and exoribonucleases that cleave and trim RNA species and also of the key small archaeal proteins that bind RNAs. Furthermore, synthetic up-to-date views of processing and biogenesis pathways of archaeal transfer and ribosomal RNAs as well as of maturation of stable small non-coding RNAs such as CRISPR RNAs, small C/D and H/ACA box guide RNAs, and other emerging classes of small RNAs are described. Finally, prospective post-transcriptional mechanisms to control archaeal messenger RNA quality and quantity are discussed.

**Keywords:** Archaea; ribonucleases; RNA processing; RNA decay; RNA-binding proteins; CRISPR-Cas

## INTRODUCTION

RNA biology plays a critical role in accurate gene expression, a prerequisite for all cellular processes from the simplest to most complex organisms. Expression of genetic information depends on messenger RNAs (mRNAs) to define the proteome and on ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) to decode mRNA sequence information as well as on additional regulatory non-coding RNAs (ncRNAs). During protein synthesis, bonds are formed in an active site of the ribosome residing on the large subunit, called the peptidyl transferase centre, which is devoid of any r-protein side chains, thus establishing the ribosome as a ribozyme (Nissen *et al.* 2000). In this framework, RNA process-

ing and decay paths stand at the centre of post-transcriptional regulation of gene expression and, therefore, are fundamental in defining the phenotypic characteristics of a cell. In addition, they permit a rapid adaptation to changes of environmental conditions. The continuing discovery of multiple mechanistically distinct pathways in Bacteria and Eukarya highlights the complexity of these processes. The quality and level of each cellular RNA species are tightly controlled and determined by the relative rates of transcription, maturation and decay (Durand *et al.* 2015; Wagner and Romby 2015; Houry-Zeevi and Rechavi 2017).

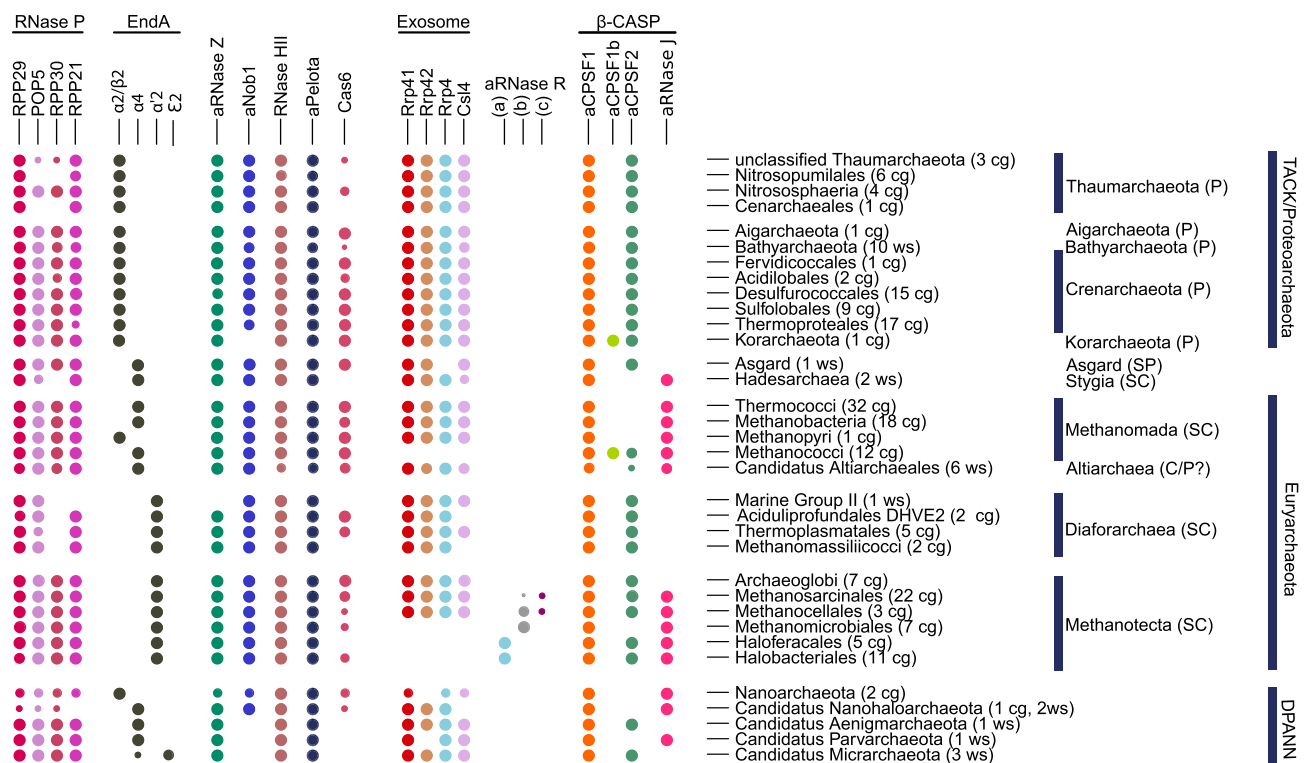
Based on evidence obtained from Eukarya and Bacteria, it is now established that an ordered action of a cell-specific

Received: 15 January 2018; Accepted: 17 April 2018

© FEMS 2018. All rights reserved. For permissions, please e-mail: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

set of RNA-modifying and RNA-processing enzymes is required from RNA synthesis to decay. Enzymes which act directly on RNA molecules are ribonucleases (RNases) that catalyse the exo- or endoribonucleolytic cleavage of phosphodiester bonds and act in concert with other RNA-related ancillary enzymes such as RNA helicases, poly(A) polymerases and pyrophosphohydrolases. RNA-processing enzymes are often found in RNA-degrading complexes and may be recruited to their respective targets by several means: cis-acting RNA sequence elements, trans-acting proteins or ncRNAs. RNA maturation or turnover is often initiated through internal cleavage by the action of endoribonucleases but also through 3'-to-5' and 5'-to-3' exoribonucleolytic decay by the action of the exoribonucleases. The tight teamwork between different RNA-degrading systems defines the outcome of each RNA species within the cell. Ribonucleases can be grouped according to their substrate specificity (single and/or double-stranded RNA or RNA/DNA hybrid), the nature of the product they release or their mechanism of action (processive or distributive). In-depth reviews covering enzymes involved in critical aspects of RNA processing in Bacteria and Eukarya have highlighted the considerable number of these processes from the nascent RNA to its degradation ((Durand et al. 2015; Wagner and Romby 2015; Houry-Zeevi and Rechavi 2017) and chapters in Stoeklin and Mühlemann 2013).

Archaeal microorganisms, omnipresent in Earth's ecosystems, constitute a unique domain of life with features at the frontier of Bacteria and Eukarya that had led to a revolution in the field of evolutionary biology. In this regard, elucidating fundamental cellular processes in Archaea is of a critical importance to understand life on earth. Moreover, the importance of studying archaeal microorganisms recently emerged from their identification in complex microbiomes of plants, animals and humans and opened the question of their impacts on the biosphere and health (Moissl-Eichinger et al. 2018). Finally, the close phylogenetic relationship between Archaea and Eukarya increasingly highlights the advantage of using archaeal systems which use simpler proteins or ribonucleoprotein (RNP) complexes. Indeed, archaeal and eukaryotic cells share many key components of their processing systems, such as replication, transcription and translation machineries (Brochier-Armanet, Forterre and Gribaldo 2011; Adam et al. 2017). This provides valuable insights in understanding associated enzymatic activities and molecular processes. Based on genomic data it has been accepted that archaeal phylogeny includes the TACK (Thaumarchaeota-Aigarchaeota-Crenarchaeota-Korarchaeota)/Proteorarchaeota group, the Euryarchaeota and the DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota) lineages (Fig. 1) (Guy and Ettema 2011). However, in view of the



**Figure 1.** Taxonomic distribution of ribonuclease protein families across the archaeal phylogeny. The size of the filled circles is proportional to the percentage of members of ribonucleases in each family. Each family is represented by a single colour which is also used in Figs 2 and 4. Classification of taxonomic groups is according to the phylogenetic tree (Adam et al. 2017), with the exception of the position of cluster I. The DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota) refers to the superphylum of Extremophile Archaea (Rinke et al. 2013). The taxonomic status (Adam et al. 2017) is shown in parentheses: C = Class; P = Phylum; SC = Super Class; SP = Super Phylum. To retrieve all members of each ribonuclease family, we collected an initial set of proteins using PFAM and/or COG profile annotations from complete archaeal genomes. Subsequently a graph partitioning method (MCL) was used to remove false positives. Hidden Markov model (hmm) profiles were built for each family and used with hmmsearch (HMMER package) to identify homologues in annotated archaeal genomes retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/genome>). In addition, to identify potential unannotated genes, we performed tblastn searches. Finally, the percentage of occurrence of each ribonuclease family gene is computed for each taxonomic group (one strain per species) by taking into account the presence or absence of a candidate gene in a genome, i.e. paralogues are counted only once. For each taxonomic group, if complete annotated genomes (cg) were available, these were used in the analysis, otherwise incomplete genomes (ws) were used. For incomplete genomes (ws), the absence of a protein may be due to the fact that the genomes are partial. For references and further details, see section 'Archaeal ribonuclease families'.

mosaic content of archaeal genomes, an enthusiastic debate is taking place around the topology of the archaeal phylogenetic tree and its connection to the eukaryal branch and questions the veracity of our present picture of the tree of life (Eme et al. 2017; Levasseur et al. 2017; Van der Gulik, Hoff and Speijer 2017).

Archaeal RNA species and RNA biology-associated processes which finely regulate gene expression appear to be mosaic, with both eukaryal and bacterial features. A unique archaeal RNA polymerase (RNAP) related to the eukaryal RNA polymerase II (RNAPII) is in charge of transcribing all archaeal RNA classes of which mRNAs share bacterial features with no eukaryotic 5'-end capping structure and generally no introns. To date, only a single example of an intron within a gene encoding a CBF5 pseudouridine synthetase has been identified in some Crenarchaeota (Yokobori et al. 2009). In addition, archaeal pre-tRNAs contain introns that occur in eukaryal nuclear tRNAs, and several eukaryotic-like RNA-processing enzymes and RNA regulatory factors have been found to be encoded in archaeal genomes. This includes the 3'-to-5' exosome machinery, the eukaryotic-like sets of ncRNAs namely C/D and H/ACA box small RNAs (sRNAs) guiding 2'-O-ribose-methylation and pseudouridylation modifications and tRNA intron-processing enzymes. However, as in Bacteria, RNA transcription and translation are proposed to be coupled in Archaea. In this mosaic context, pathways dictating the biogenesis and the fate of archaeal RNAs wait for most of them to be precisely characterised.

This review condenses the state of our current knowledge on associated players and pathways mediating maturation and degradation of archaeal RNA species. This focus opens with a comprehensive overview on the major protein families acknowledged to date which are capable of degrading or binding RNA molecules. This includes a comprehensive study of the taxonomic distribution of key ribonucleases setting the functional status of each enzyme across the archaeal phylogeny. Subsequently, maturation pathways leading to functional tRNAs, rRNAs, non-coding guide H/ACA and C/D box sRNAs, clustered regularly interspaced short palindromic repeat RNAs (CRISPR or crRNAs) and novel emerging classes of ncRNAs are described. Finally, a prospective on archaeal mRNA turnover and quality control pathways is discussed to propose mRNA-signalling pathways at regulating gene expression in Archaea.

### Archaeal ribonuclease families

Our purpose is to provide a panoramic view of archaeal enzymes and machines that cleave and degrade the various RNA types in light of bacterial and eukaryotic systems to question their significance. Since 2001, studies on bacterial and eukaryal RNA-processing apparatuses have supplied the foundation for the discovery of RNA-degrading enzymes implicated in fundamental aspects of RNA processing in Archaea (see Koonin, Wolf and Aravind 2001; Klug et al. 2007; Evguenieva-Hackenberg and Klug 2009; de Koning et al. 2010; Clouet-d'Orval et al. 2015; Phung, Bouvier and Clouet-d'Orval 2017 for review). The 10 archaeal RNase families highlighted in Tables 1A and 1B are defined based on their bacterial and eukaryal counterparts (Arraiano et al. 2013; Stoecklin and Mühlemann 2013) (Table 2). We performed a phylogenomic search of members of each family within archaeal genomes to highlight the occurrence of these RNase families across the phylogeny of Archaea. This is shown and described in Fig. 1.

### Ubiquitous archaeal endoribonucleases

To date, archaeal endoribonucleolytic activities were shown to be carried out by the universal RNase P and RNase Z families; the ubiquitous EndA, Nob1, aPelota and aCPSF1 families shared with Eukarya; and finally, the Cas6 family shared with Bacteria (Table 1A). Note that the aCPSF1 subfamily is described in the section 'The universal  $\beta$ -CASP ribonuclease family' together with other members of the  $\beta$ -CASP family.

**RNase P (EC 3.1.26.5)** RNase P is a site-specific endoribonuclease which catalyses the hydrolysis of the phosphodiester bond at the junction between a 5'-end single-stranded and a double-stranded region. This reaction requires  $Mg^{2+}$  ions for the attack of the phosphorous atom in the scissile bond (Guerrier-Takada et al. 1983). In all domains of life, RNase P is best known for its role in removing the 5'-leaders of tRNA precursors (pre-tRNAs) transcribed as mono- or polycistronic transcripts (Frank and Pace 1998; Evans, Marquez and Pace 2006; Esakova and Krasilnikov 2010; Jarrous and Gopalan 2010; Lai et al. 2010b). However, RNase P is also involved in processing the precursor of 4.5S RNA, the transfer-messenger RNA (tmRNA) and riboswitches in bacteria, and of ncRNAs, such as intron-encoded box C/D small nucleolar RNAs, in Eukarya (Esakova and Krasilnikov 2010; Altman 2011). It has been proposed that this virtually ubiquitous enzyme has independently originated at least twice in evolution with different architectures (Lechner et al. 2015). The most ancient type of RNase P holoenzyme is a ribozyme-based RNP complex assembled on a catalytic RNA subunit (RPR) (Guerrier-Takada et al. 1983). This RNP complex adopts divergent scaffolds throughout the three domains of life (see Lai et al. 2010b; Klemm et al. 2016; Samanta et al. 2016 for review). More recently, protein-only RNase P enzymes were discovered in human mitochondria (PRORP) and in the bacterium *Aquifex aeolicus* (HARP), highlighting a more intricate evolution of RNase P than previously thought (Hartmann and Hartmann 2003; Howard et al. 2013; Lechner et al. 2015; Klemm et al. 2016; Nickel et al. 2017). Using the HARP protein sequence from *A. aeolicus*, homologous proteins could also be identified in archaeal genomes (Nickel et al. 2017). However tRNA-processing activity has not yet been shown for these archaeal HARP proteins.

The diversity of RPRs and RPPs, composing RNase P in the three domains of life, has been extensively reviewed (Gopalan 2007; Altman 2011; Howard et al. 2013; Klemm et al. 2016; Samanta et al. 2016) (Table 2). Briefly, RPRs are structurally related and classified according to specific structural features important for catalysis and substrate recognition (Gopalan 2007). In Archaea, three types of RPRs are now well defined: the archaeal A- and M-types characterised by two functional RNA domains, the substrate-binding domain (S-domain) and the catalytic domain (C-domain) (Fig. 2A); the newly identified T-type, a shorter RPR only retaining the conventional C-domain but lacking a recognisable S-domain, identified in Thermoproteaceae and the related Crenarchaea *Caldivirga maquilungensis* and *Vulcanisaeta distributa* (Lai et al. 2010a).

The A-type, related to the bacterial type but lacking the P13/P14 and P18 regions, is the most common in the archaeal phylogeny (Harris et al. 2001) (Figs 1 and 2A). The M-type which is restricted to the Methanococci and *Archaeoglobus* resembles the A-type without the P8 and P16 regions, the P6 pseudoknot and a loop in P15 (Pannucci et al. 1999; Gopalan 2007). While the archaeal A- and M-type RPRs contain all of the elements required for substrate recognition and catalysis, with biochemical properties similar to synthetic minimal bacterial

Table 1A. Acknowledged archaeal endoribonuclease families.

Activity	Family	Name	Organism	PDB accession	Reference	
ENDO	RNase P	POP5/RPP30	<i>P. horikoshii</i>	2czv*	Kawano et al. (2006)	
			<i>T. kodakarensis</i>	3wz0	Suematsu et al. (2015)	
		POP5	<i>P. furiosus</i>	2av5	Wilson et al. (2006)	
		RPP21/RPP29	<i>P. horikoshii</i>	2zae*	Honda et al. (2008)	
			<i>P. furiosus</i>	2ki7	Xu et al. (2009)	
		RPP21	<i>P. furiosus</i>	2k3r	Amero et al. (2008)	
		RPP29	<i>M. thermautrophilus</i>	1oqk	Boomershine et al. (2003)	
			<i>A. fulgidus</i>	1pc0, 1ts9	Sidote and Hoffman (2003); Sidote, Heideker and Hoffman (2004)	
		RPP38/L7Ae	<i>M. jannaschii</i>	1xbi	Suryadi et al. (2005)	
			<i>P. abyssi</i>	1pxw	Charron et al. (2004a)	
	EndA/ Sen	$\alpha 4$	<i>M. jannaschii</i>	1a79*	Li, Trotta and Abelson (1998)	
		$\alpha 2$	<i>A. fulgidus</i>	1rlv*/ 2gfw/ 3p1y/ 1r0v	Li and Abelson (2000); Xue, Calvin and Li (2006); Hirata, Kitajima and Hori (2011); Zhang and Li (2004)	
			<i>N. equitans</i>	3iey	Mitchell et al. (2009)	
			<i>T. acidophilum</i>	2ohc	Kim et al. (2007)	
		$\alpha 2\beta 2$	<i>A. fulgidus</i>	3p1z*/ 3ajv	Hirata, Kitajima and Hori (2011); Okuda et al. (2011)	
			<i>P. aerophilum</i>	2zyz	Yoshinari et al. (2009)	
		$\epsilon 2$	ARMAN	4fz2*	Hirata et al. (2012)	
		rnz	aRNase Z	<i>P. abyssi</i>		model* (39% Id. elac1)
		Pelota/ Dom34	aPelota	<i>T. acidophilum</i>	2qi2*	Lee et al. (2007)
				<i>A. pernix</i>	3wxm	Kobayashi et al. (2010)
	<i>A. fulgidus</i>		3oby	Lee et al. (2010)		
	<i>P. furiosus</i>		3j15	Becker et al. (2012)		
PIN	aNob1	<i>P. horikoshii</i>	2lcq*	Veith et al. (2012)		
RNase H	HII	<i>A. fulgidus</i>	1i39*/ 3p83	Chapados et al. (2001); Bubeck et al. (2011)		
		<i>M. jannaschii</i>	1eke	Lai et al. (2000)		
		<i>P. furiosus</i>	3pkm*/ 3ufc	Wang et al. (2011); Park et al. (2012)		
Ferredoxin like	Cas6	<i>S. solfataricus</i>	3zfv	Reeks et al. (2013)		
		<i>P. horikoshii</i>	3af6*/ 3af5	Nishida et al. (2010)		
$\beta$ -CASP	aCPSF1	<i>M. mazei</i>	2xr1	Mir-montazeri et al. (2011)		
		<i>M. thermotrophicus</i>	2ycb	Silva et al. (2011)		

Structural models shown in Fig. 2 are indicated \*.

Note that aCPSF1 has been classified in here but carries also a 5'-3' exoribonucleolytic activity (see section 'Archaeal ribonuclease families').

Table 1B. Acknowledged archaeal exoribonuclease families.

Activity	Family	Name	Organism	PDB accession	Reference	
3'-5' EXO	PDX/RNase PH	Exosome (Csl4)	<i>A. fulgidus</i>	3m7n*/3m85	Hartung et al. (2010)	
		Exosome (Rrp4)	<i>A. fulgidus</i>	2ba0*/2ba1	Büttner, Wenig and Hopfner et al. (2005)	
		Exosome	<i>S. solfataricus</i>	2je6/2jea/2jeb, 2br2, 3l7z, 4ba1/4ba2, 2c37/2c38/2c39	Lorentzen et al. (2005, 2007); Lorentzen and Conti (2012); Lu, Ding and Ke et al. (2010); Lorentzen and Conti (2005)	
			<i>P. abyssi</i>	2po2/ 2pnz/2po0/ 2po1	Navarro et al. (2008)	
	RNB/RNase II	aRNase R	<i>H. volcanii</i>		model a* (28% Id. dis3l2)	
			<i>M. concili</i>		model b* (28% Id. dis3l2/OB-fold)	
			<i>M. conradii</i>		model c* (28% Id. dis3l2/ dsRBD)	
	5'-3' EXO	$\beta$ -CASP	aCPSF1b	<i>M. jannaschii</i>		model (32% Id. <i>M. mazei</i> aCPSF1)
			aCPSF2	<i>S. acidocaldarius</i>		model* (33% Id. <i>M. mazei</i> aCPSF1)
aRNase J			<i>M. psychrophilus</i>	5haa/5hab*	Zheng et al. (2017)	

Structural models shown in Fig. 4 are indicated \*.



**Table 2.** Distribution of ribonuclease enzymes and subunits across the three domains of life.

Enzymes	Bacteria	Archaea	Eukarya
RNaseP/RPR	Type A	Type A/M/T	H1 RNA
RNaseP/RPP	C5 protein/RnpA	POP5 RPP21/29/30 RPP38/L7Ae	POP5 RPP21/29/30 RPP38 RPP20,RPP25, RPP14,RPP40,POP1
EndA		$\alpha / \alpha' / \beta / \epsilon$	Sen 2/15/34/54
RNase Z	RNase Z	aRNase Z	RNase Z <sup>S</sup> / RNase Z <sup>L</sup>
Pelota/Dom34		aPelota	Pelota/Dom34
Nob1		aNob1	Nob1
RNase H	RNaseHI/HII	aRNase HII	RNase H1/H2
Cas6	Cas6	Cas6	
CPSF-like	Unch. CPSF <sup>*</sup>	aCPSF1/1b/2 aC $\beta$ x/aC $\beta$ y/aC $\beta$ z	CPSF73/ Int11/RC-68 CSPF100 <sup>Δ</sup> /Int9/RC-74 <sup>Δ</sup>
RNase J-like	RNase J	aRNase J	
Exosome/PNPase	RNase PH	Rrp41/42 <sup>Δ</sup> Rrp4 Csl4	Rrp41 <sup>Δ</sup> /42 <sup>Δ</sup> Rrp4 Csl4 Rrp44/Dis3
RNase R-like	RNase R	aRNase R <sup>a</sup> /R <sup>b</sup> /R <sup>c</sup>	

CPSF73/CPSF100 heterodimer (cleavage and polyadenylation specific factor complex) involved in 3' end processing of eukaryal canonical and histone pre-mRNA  
Int11/RC-68/ Int9/RC-74 (integrator complex) involved in 3' end processing of eukaryal pre-snRNA (see Dominski, Carpousis and Clouet-d'Orval 2013 for review)  
Uncharacterised CPSF-like  $\beta$ -CASP are detected in bacteria; <sup>Δ</sup>inactive subunit.

RPRs, they are structurally defective in the absence of protein (Guerrier-Takada et al. 1983; Kikovska, Svård and Kirsebom 2007). Only A-type RPRs show residual *in vitro* endoribonucleolytic activity (Pannucci et al. 1999; Gopalan 2007).

Interestingly, the catalytic RPR is associated to a single protein in Bacteria, at least five in Archaea and up to 10 in eukaryal nuclei (Frank and Pace 1998; Evans, Marquez and Pace 2006; Ellis and Brown 2009; Hartmann et al. 2009; Esakova and Krasilnikov 2010; Jarrous and Gopalan 2010; Lai et al. 2010b; Samanta et al. 2016) (Tables 1A and 2, Fig. 2A). Although archaeal RNase P RNAs are similar in both sequence and structure to those of Bacteria, all archaeal RNase P protein subunits are orthologous to their cousins in eukaryotic RNase P (Hall and Brown 2001, 2002; Samanta et al. 2016) (Table 2). Until recently, the A- and M-types of archaeal RNase P were described as containing at least four proteins (POP5, RPP21, RPP29 and RPP30) (Table 1A). A fifth partner, RPP38, also known as 50S ribosomal protein L7Ae, was recently shown to be associated with the RNA component of both RNase P types (Fukuhara et al. 2006; Cho et al. 2010; Lai et al. 2014, 2017).

Over the years, structural studies and biochemical reconstitutions of the RNase P subunits of *Pyrococcus furiosus*, *P. horikoshii*, *Methanothermobacter thermoautotrophicum*, *Methanocaldococcus jannaschii* and *Methanococcus maripaludis* have led to valuable insights into protein-aided RNA catalysis showing that protein content and divalent metal ions affect cleavage fidelity and catalytic efficiency (Kouzuma et al. 2003; Tsai et al. 2006; Pulukkunat and Gopalan 2008; Li, Willkomm and Hartmann 2009; Chen et al. 2010, 2011, 2012; Sinapah et al. 2011; Gao et al. 2017; Kimura 2017).

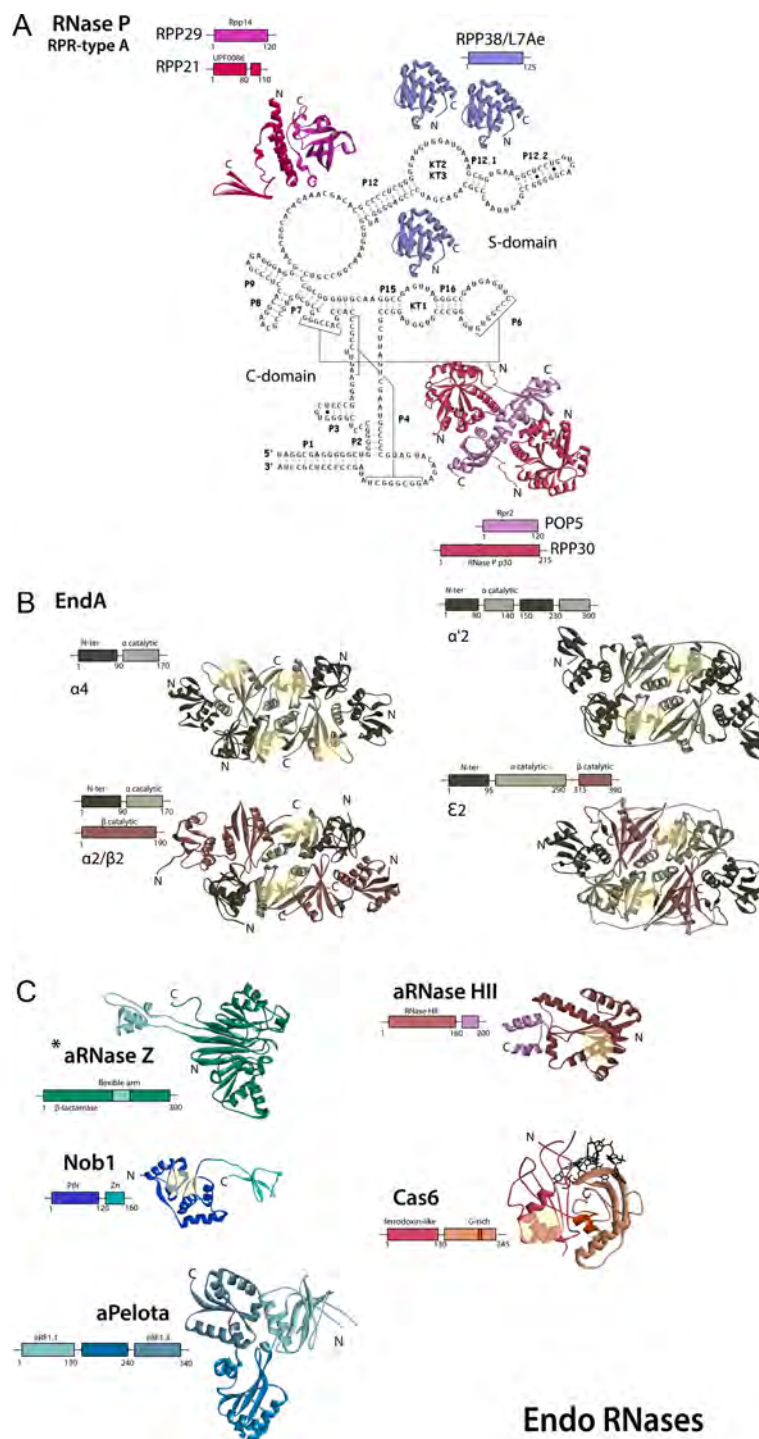
The dominant structural feature of the 11-kDa RPP29 subunit is a sheet of six antiparallel  $\beta$ -strands, wrapped around a core of conserved hydrophobic amino acids (Fig. 2A) (Sidote and Hoffman 2003; Numata et al. 2004; Sidote, Heideker and Hoffman 2004; Amero et al. 2008). The RPP21 subunit comprises an N-terminal domain consisting of two long  $\alpha$ -helices, and a central linker and C-terminal domain (CTD) which folds into a zinc ribbon domain forming an L-shaped structure (Fig. 2A)

(Kakuta et al. 2005; Amero et al. 2008). These two subunits form a 30-kDa complex (RPP21•RPP29) (Honda et al. 2008; Xu et al. 2009) with enhanced substrate affinity (see Samanta et al. 2016; Kimura 2017 for review).

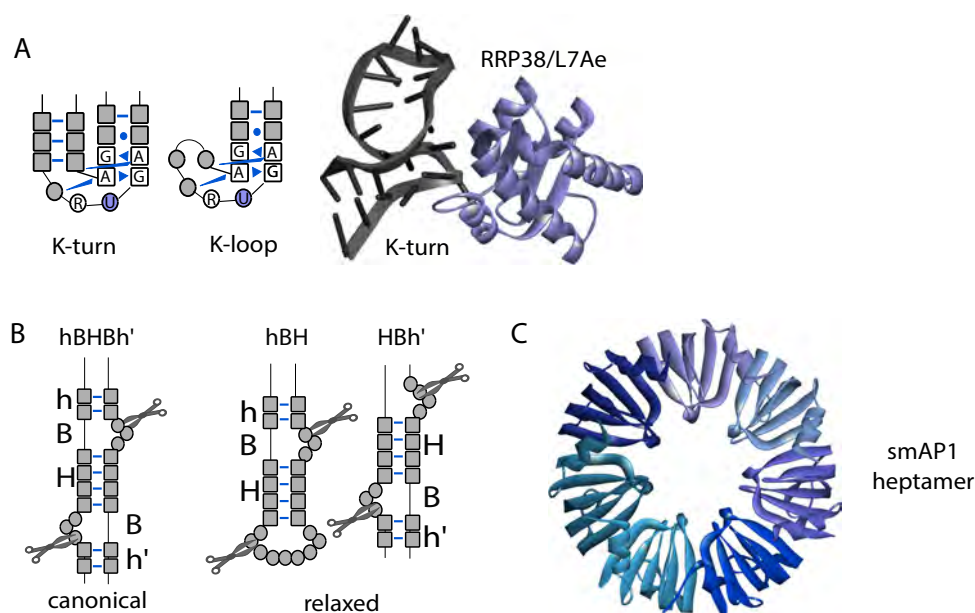
Despite weak sequence similarities of POP5 and RPP30, structural and footprinting studies revealed a heterotetrameric POP5•RPP30 complex, in which a homodimer of POP5 sits between two RPP30 molecules and contacts the C-domains of the RNA moiety by an hydrophobic interface (Kawano et al. 2006; Tsai et al. 2006; Honda et al. 2010; Crowe et al. 2011) (Fig. 2A). The POP5•RPP30 complex enhances the rate of RPR-catalysed self-processing by about 100-fold (Pulukkunat and Gopalan 2008).

Finally, the RPP38/L7Ae protein family (see section 'Archaeal small RNA-binding proteins') is described to specifically recognise a Kink-turn RNA structural motif (also named K-turn or KT), a key architectural motif in RNA structure, characterised by a three-nucleotide bulge flanked by a 5' canonical helix and a 3' non-canonical helix capped with two sheared non-canonical G•A base pairs in tandem that induces a tightly kinked structure with a twist (Huang and Lilley 2013, 2016) (see Lilley 2012; for review) (Fig. 3A). A novel integrated approach supports the formation of 'double KT' modules in A- and M-type archaeal RPR variants which are bound by the RPP38/L7Ae protein and are predicted to facilitate the functions of other RPPs (Lai et al. 2017) (Fig. 2A). More specifically, it has been proposed that the formation of the K-turn in the C-domain of A-type RPR (KT1) could promote the interaction between the RPR and POP5•RPP30, while the double KT in the S-domain of both A- and M-types (KT2 and KT3) may facilitate a long-range tertiary interaction required for proper binding of the pre-tRNA substrate (Lai et al. 2017).

Even though there is no atomic resolution structure of full archaeal RNase P enzymes available to date, a three-dimensional (3D) model (not including RPP38/L7Ae) has been constructed based on crystal structures of the individual protein subunits and of the RPP21•RPP29 and POP5•RPP30 binary complexes (Kimura and Kakuta 2012). This proposed 3D model of



**Figure 2.** Domain organisation and structure of archaeal endoribonucleases. Structure (PDB accession numbers) and corresponding model organism references are reported in Table 1A and marked by an asterisk. When no structure was available (protein name tagged with an asterisk in figure), we used the Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2>) to generate a 3D protein model. Domains correspond to Pfam annotation (Pfam database <http://pfam.xfam.org/>). Catalytic sites are highlighted with yellow areas. (A) Archaeal RNase P is composed of a RNA moiety (ribonucleotide sequence shown in black), structured into two domains: a catalytic domain (C-domain) and a structural domain (S-domain). The secondary structure folding of *P. horikoshii* A-type is represented with typical helical structure (P) numbering, and three Kink-turn motifs (KT). Specific sets of heterodimer (RPP21•RPP29) and heterotetramer (POP5•RPP30) protein subunits bind to the S- and C- domains, respectively. In addition, the small RPP38/L7Ae protein binds and cleave BHB motifs (see Fig. 3B). (B) Structure and domains of the four different forms of EndA found across the archaeal phylogeny are shown. The  $\alpha_4$ -type is a homotetramer containing four active catalytic sites. Each  $\alpha$  subunit is composed of two Pfam domains (PF02778 'N-term' and PF01974 'catalytic', light and dark brown, respectively). The  $\alpha_2/\beta_2$ -type heterotetramer is shown in dark pink. The  $\alpha_2$ -type is a heterodimer composed of two units of fused  $\alpha$ -types leaving only two active catalytic sites. Finally, the  $\epsilon_2$ -type is a version with two active catalytic sites composed of a two units of fused  $\alpha/\beta$  types. The archaeal EndAs recognise and cleave BHB motifs (see Fig. 3B). (C) Five other archaeal endoribonucleases, aPelota, RNase Z, Nob1, RNase HII, Cas6, acting as monomers or dimers are shown. The RNase Z which is characterised by a  $\beta$ -lactamase domain interrupted by a small insertion has no crystal structure reported to date. A Phyre2-built structural model of *P. abyssi* RNase Z shows that the flexible arm (light blue) protrudes from the classical  $\beta$ -lactamase domain. Note that Cas6 protein sequences show limited sequence conservation with only two common motifs: the ferredoxin fold or RRM (RNA recognition motifs) domains and a glycine-rich motif located at the C-terminus. The structure of *P. abyssi* Cas6 with an RNA fragment is shown.



**Figure 3.** RNA structural motifs and small RNA binding proteins. (A) The structure of the Kink-turn motif with and without RRP38/L7Ae. Left panel: The Kink-turn (K-turn) RNA consists of a canonical (grey squares) and a non-canonical (unfilled squares) RNA stem separated by a short asymmetric loop (circles) with a sharp bend between the two stems. The internal loop contains a highly conserved flipped-out U (in blue circle) followed by a third non-Watson Crick base pair and the non-canonical stem usually contains two tandem sheared GA base pairs (non-Watson Crick base pairing is in blue triangle) (Lescoute et al. 2005). Middle panel: The Kink-loop (K-loop) is an RNA structural motif related to the K-turn with a single stem closed by a terminal loop (Nolivos, Carpousis and Clouet-d'Orval 2005). Both RNA motifs are recognised by the L7Ae protein. Right panel: Ribbon structure representation of RRP38/L7Ae from *P. horikoshii* in complex with the K-turn of the SL12 RNA fragment (helix P12 of RPR) (PDB accession number:5DCV). The protein is represented by its ribbon diagram in purple and the RNA is drawn in black/grey stick form. The RNA-protein interaction occurs primarily through nucleotides in the non-canonical stem (sheared GA pairs) and at the 3-nucleotide bulge (G-G-U) (Oshima et al. 2016). (B) Bulge-Helix-Bulge RNA (BHB) motifs. Left panel: The canonical BHB motif or hBHBh' comprises two three-nucleotide bulges (B) enclosing a central main helix (H) of four base pairs. The bulges are flanked by shorter helices (h and h') comprising at least two base pairs (Thompson and Daniels 1990; Marck and Grosjean 2003). The relaxed versions of the BHB motif lack one of the flanking helices and one bulge and are accordingly referred to as hBH and hBh' (Marck and Grosjean 2003). Only the  $\alpha_2\beta_2$  and  $\varepsilon_2$  EndA forms can process relaxed BHB motifs (Tocchini-Valentini, Fruscoloni and Tocchini-Valentini 2005; Fujishima et al. 2011). Grey squares are for Watson Crick base pairs and grey circles for single-stranded nucleotides. Scissors indicate the EndA cleavage sites within the bulges. (C) Structure of the AF-Sm1 heptamer Sm-like protein. The top view of the ribbon representation of the *Archaeoglobulus fulgidus* Sm1 (AF-Sm1) is shown (PDB accession number 1L5L, Törö et al. 2001). For clarity, the monomers are drawn in a colour gradation from blue to violet. AF-Sm1 forms a seven-membered ring. It was shown that no conformational changes are induced by binding of RNA to the interior of the heptamer (Törö et al. 2001).

*P. horikoshii* RNase P highlights that protein-RNA interactions in archaeal RNase P could replace RNA-RNA contacts in bacterial RNase P. This framework will assist further structural and functional studies on archaeal, as well as eukaryotic, RNase P enzymes.

Interestingly, the taxonomic distribution of the RPP subunits and RPR types among the available archaeal sequenced genomes (Fig. 1) confirmed a previous analysis showing that genes coding for RPP21 are missing or are too divergent to be detected in genomes encoding T-type RPRs that lack an S-domain (Lai et al. 2010a). Further investigations are needed to elucidate how RPP29 interacts with T-type RPRs in the absence of RPP21. Collectively, the plasticity in the subunit composition and of T-type RNase P in Thermoproteaceae highlights the rich evolutionary story of RNase P within the archaeal domain (Lai et al. 2010b). Furthermore, it is interesting to note that the only known organism without RNase P is the archaeon *Nanoarchaeum equitans*, which belongs to the fast-evolving DPANN phylogenetic lineage with extremely reduced genome sizes (Brochier et al. 2005; Rinke et al. 2013), presumably because the pre-tRNAs in this organism are transcribed without leader sequences (Randau, Schröder and Söll 2008) (Fig. 1).

Previous observations, in agreement with our phylogenomic analysis (Fig. 1), identified all five RPPs in one or more members of Euryarchaeota, Crenarchaeota, Thaumarchaeota, Nanoarchaeota (except for *N. equitans*), Bathyarchaeota, Nanohaloar-

chaeta, Diapherotrites and Lokiarchaeota (Samanta et al. 2016). The RPP distribution in Thaumarchaeota is similar, with POP5 and RPP30 missing in some genomes whereas RPP21 homologues are absent in some Thermoproteales strains (*Thermoproteus uzoniensis*, *Pyrobaculum arsenaticum* and *V. distribute*) as previously reported by Lai et al. (2010a). In addition, RPP30 homologues are not found in genomes from the Diapherachaea group. However, the occurrence of the other partners suggests that these subunits may have diverged beyond the sensitivity of the similarity search methods. This is supported by the observation that RPP30 is less conserved than the other RNase P subunits (Samanta et al. 2016). Last of all, it has also been proposed that these RPPs arose early in archaeal phylogeny and that putative ancestral archaeal RPP genomic loci include genes encoding other core cellular machineries including open reading frames for several ribosomal, exosomal and proteasomal proteins (Samanta et al. 2016).

**The splicing endonuclease EndA** (EC 3.1.27.9 transferred to EC 4.6.1.16) The archaeal RNA-splicing endonuclease, termed EndA, is a multimeric enzyme with variable subunit compositions (Fig. 2B, Table 1A) that recognises a specific RNA structural motif named Bulge-Helix-Bulge (BHB) which consists in its canonical form of a four-base-pair central helix flanked by two juxtaposed three-nucleotide bulges in a helical context with at least two adjacent base pairs (Fig. 3B) (Thompson and



Daniels 1990; Marck and Grosjean 2003). Asymmetric endoribonucleolytic cleavages occur within the bulges producing two RNA molecules bearing a 3'-end 2', 3'-cyclic phosphate and a 5'-end hydroxyl group. EndA is strictly conserved among the archaeal phylogeny (Fig. 1) and catalyses critical steps of tRNA and rRNA processing (described in section 'Processing pathways of archaeal ncRNAs').

Overall, archaeal EndA enzymes, classified upon their subunit composition, present the same structure (Fig. 2B), suggesting an accommodation over time to diverse canonical and non-canonical BHB splicing motif substrates (Fig. 3B) (Marck and Grosjean 2003). Their different quaternary structures that may coincide with the diversity of substrates observed within different archaeal groups could have arisen by gene duplication and subsequent subfunctionalisation (Tocchini-Valentini, Fruscoloni and Tocchini-Valentini 2005). The homotetrameric  $\alpha_4$  form (Fig. 2B) with potentially four active catalytic sites is predominantly found in Euryarchaea (Fig. 1) but also, interestingly, in the recently discovered Asgard group described to be closely branched to Eukarya (Zaremba-Niedzwiedzka et al. 2017). In contrast, the homodimeric ( $\alpha_2$ ) form, predominantly found in Euryarchaeota, and the heterotetrameric ( $\alpha_2\beta_2$ ) configuration, present mostly in Crenarchaeota and only exceptionally in Nanoarchaeota or the euryarchaeon *Methanopyrus kandleri* (Calvin and Li 2008), possess only two active catalytic sites (Fig. 2B). Briefly, the  $\alpha_2$  homodimeric or  $\alpha_4$  homotetrameric forms are only capable of processing canonical BHB motifs (Lykke-Andersen and Garrett 1997; Li, Trotta and Abelson 1998; Li and Abelson 2000) (Fig. 3B), while the  $\alpha_2\beta_2$  heterotetrameric variant in addition can bind and process non-canonical BHB motifs (Fig. 3B). A fourth type of splicing endonuclease, the  $\varepsilon_2$  homodimer, with a similarly relaxed substrate recognition was described in the deeply branching Archaea ARMAN-1 and ARMAN-2 (Fujishima et al. 2011; Hirata et al. 2012) (Fig. 1).

Finally, comparative sequence analysis of the subunits revealed strong similarity with two protein subunits, Sen34p and Sen2p, of the eukaryal heterotetrameric tRNA splicing endonuclease suggesting an ancient origin of EndA splicing enzymes before the emergence of Eukarya (Trotta et al. 1997; Hirata et al. 2012) (see Lopes et al. 2015 for review) (Table 2).

***aRNase Z*** (EC 3.1.26.11) From respective sequence similarities, the RNase Z enzymes were classified as belonging to the family of metal-dependent  $\beta$ -lactamases (Aravind 1999) (Table 2). RNase Z homologues are present in organisms of all three domains of life (Späth et al. 2008) and ubiquitously in archaeal genomes with the exception of *Nanopusillus acidilobi* 605887 (Fig. 1). RNase Z exists in two forms: a short form (250–350 aa) present in all domains of life and a long form (750–900 aa) restricted to Eukarya that probably arose from a gene duplication event (Vogel et al. 2005; Ma et al. 2017). Being a metallo- $\beta$ -lactamase, RNase Z enzymes exhibit the characteristic  $\alpha\beta/\beta\alpha$  fold of sandwiched  $\beta$ -sheets and in the catalytic centre the characteristic histidine motif (HxHxDH) and additional histidine and aspartate residues (Fig. 2C, Table 1A) (Aravind 1999; Ishii et al. 2005, 2007; Li de la Sierra-Gallay, Pellegrini and Condon 2005). In addition, RNase Z has a flexible arm (coined the exosite) that is inserted between the N- and C-terminal halves and protrudes from the protein body (Ishii et al. 2007). Crystal structures of archaeal RNase Z have not yet been resolved; a model is shown in Fig. 2C. The short RNase Z dimer binds the pre-tRNA with one monomer by clamping the exosite onto the tRNA elbow by hydrogen bonds to the backbone of mainly the T-arm and acceptor stem (Ishii et al. 2005; Li de la Sierra-Gallay et al. 2006; Zhang and

Ferré-D'Amaré 2016). Structures resembling these two stacked helices are in fact the minimal substrate of most RNase Z enzymes with the exception of the *Haloferax volcanii* RNase Z that also requires the D-arm (Schierling et al. 2002; Schiffer, Rösch and Marchfelder 2002; Ishii et al. 2005). Catalysis that depends on the coordination of  $Zn^{2+}$  ions generates a tRNA with a 3'-end hydroxyl group (Späth, Canino and Marchfelder 2007). For the *M. jannaschii*, *P. furiosus*, *H. volcanii*, *P. aerophilum* and *Thermoplasma acidophilum* RNase Z enzymes, an *in vitro* tRNA-processing activity was sought (Schierling et al. 2002; Schiffer, Rösch and Marchfelder 2003; Hölzle et al. 2008; Späth et al. 2008). These enzymes show only small differences in their biochemical requirements, in the cleavage site position and in substrate specificity. It is interesting to note that for RNase Z of *H. volcanii* a high-salt concentration inhibits its activity, contrasting to the organism's halophilic life style (Späth et al. 2008). While the maturation state of the pre-tRNA influences the processing activity of RNase Z, the outcome is species and tRNA specific. Intron-containing precursors and substrates with a long 5'-leader are not processed by *H. volcanii* RNase Z but can be cleaved by *P. furiosus* RNase Z (Späth et al. 2008). The presence of a CCA in the 3'-trailer inhibits cleavage by *T. acidophilum* RNase Z but does not influence *P. aerophilum* and *M. jannaschii* RNase Z activities (Schiffer, Rösch and Marchfelder 2003; Späth et al. 2008).

***aNob1*** (EC 3.1.) Several archaeal homologues of eukaryotic ribosome assembly factors have been identified and are valued as models for functional and structural studies (Veith et al. 2012; Hellmich et al. 2013). This is the case for the eukaryotic endoribonuclease Nob1 (Table 2). Nob1 of *P. horikoshii* (Pho) efficiently cleaves RNA substrates encompassing a D-site, *in vitro* in a manganese-dependent manner. In addition, the structure of Pho-Nob1, solved by nuclear magnetic resonance spectroscopy, revealed the presence of a PIN (PiIT N-terminus) domain typical of many nucleases and of a zinc ribbon domain (Table 1A). These two domains are structurally connected by a flexible linker (Fig. 2C) (Veith et al. 2012). However, Pho-Nob1 lacks a large insertion in the PIN domain and a long C-terminal tail that are both typical of eukaryotic Nob1 (Veith et al. 2012). Further studies showed that the structure and function of Nob1 are conserved in all the major branches of Archaea with the exception of Korarchaea and some DPANN with incomplete genomes (Fig. 1). Nevertheless, it should be noted that shorter CTDs are observed in some Crenarchaeota such as certain Thermoproteales (13 out of 17 genomes), the Acidilobales (2/2), the Fervidicoccales (1/1) and some Desulfirococcales (3/15).

***RNase HII*** (EC 3.1.26.4) RNA-primed replication intermediates in archaeal cells are similar to those formed during eukaryotic DNA replication (Kochiwa, Tomita and Kanai 2007). RNase H enzymes play a crucial role in DNA replication by removing the RNA primer of Okazaki fragments, and in DNA repair by removing the occasional single ribonucleotides incorporated in the DNA (see Tadokoro and Kanaya 2009 for review). It was recently shown that initiation of the misincorporated ribonucleotide repair pathway in Archaea relies on RNase HII (Heider et al. 2017). The RNase HII gene is conserved in all complete archaeal genomes with the exceptions of *Nitrosopumilus sediminis* and *N. koreensis* AR1 (Fig. 1, Table 1A). The RNase HI genes have a lower occurrence and a discontinuous distribution (Kochiwa, Tomita and Kanai 2007). This result is consistent with an essential and non-overlapping function of the RNase HII in archaeal genomes.



RNase H enzymes which cleave the P-O 3'-bond of the RNA strand embedded in RNA/DNA hybrid substrates with a two-metal-ion catalysis mechanism vary greatly in domain structures and substrate specificities (see Tadokoro and Kanaya 2009 for review) (Fig. 2C). RNase H from Bacteria and Archaea have been classified into type 1 RNase H (RNase HI) and type 2 RNase H (RNase HII) based on differences in their amino acid sequences (Kochiwa, Tomita and Kanai 2007) (Table 2). The structures of *M. jannaschii* (Lai et al. 2000), *Thermococcus kodakarensis* (Muroya et al. 2001) and *Archaeoglobus fulgidus* (Bubeck et al. 2011) share a main chain fold, termed the RNase H-fold, and steric configurations of the four acidic active-site residues (Fig. 2C). In addition, it has been reported that the 3D structure of RNase H is similar to that of the PIWI domain of the *P. furiosus* argonaute-like (Ago) protein suggesting an evolutionary relationship between the RNase H and the Ago proteins in prokaryotes. Interestingly, in the presence of Mn<sup>2+</sup> ions, Pf RNase HII digests RNA–RNA duplexes (Kitamura et al. 2010).

**aPelota** (EC 3.1.) *Drosophila melanogaster* Pelota and *Saccharomyces cerevisiae* DOM34, initially identified as important proteins for meiotic cell division, were recently shown to be key players in the no-go decay (NGD) surveillance pathway by which mRNAs containing stalled ribosomes are degraded by endoribonucleolytic cleavage (see Simms, Thomas and Zaher 2017 for a review). Pelota-like proteins (or aPelota), homologues of eukaryal Pelota, were also identified in archaeal genomes (Table 1A). The protein aPelota from *Sulfolobus solfataricus* shares about 20% identity with its eukaryal counterparts (Ragan et al. 1996) (Table 2) and is widely conserved in Archaea (Atkinson, Baldauf and Hauriuk 2008) (Fig. 1). Recently, the crystal structures of archaeal Pelota proteins from *T. acidophilum*, *A. fulgidus* and *S. solfataricus* revealed three structural domains (Lee et al. 2007, 2010). The central and CTDs (Fig. 2C) share similarities with the eukaryotic translational release factor eRF1, while the N-terminal domain has a weak similarity with RNA-binding Sm-fold proteins (described in section 'Archaeal small RNA-binding proteins') (Fig. 3C). In addition, these structural studies reported an inter-domain structural plasticity of aPelota proteins suggesting that large conformational changes are essential for their functions (Lee et al. 2010). It was shown that aPelota from *T. acidophilum* has a ribonucleolytic activity (Lee et al. 2007). In addition, the archaeal translational elongation factor EF1 $\alpha$  binds to archaeal RF1 and aPelota in a GTP-dependent manner (Kobayashi et al. 2010; Kobayashi, Ishitani and Nureki 2013). It was proposed that a positively charged patch of aPelota domain 1 interacts with the decoding centre of the ribosome to specifically recognise the empty site A of the stalled ribosome and thereby induce the destabilisation of the ribosome (Kobayashi et al. 2010; Becker et al. 2012; Kobayashi, Ishitani and Nureki 2013).

**Cas6** The CRISPR-Cas defence mechanism involves repeat-associated mysterious proteins (RAMPs) that are classified into three large groups: Cas5, Cas6 and Cas7. These proteins form large RNP complexes together with CRISPR RNA (crRNA) dedicated to the destruction of invading DNA or RNA (Makarova et al. 2011, 2015). In type I and type III systems, the Cas6 protein processes the crRNA precursor into crRNAs of about 70–80 nucleotides length with a 5'-hydroxyl and a 3'- or cyclic 2'-3'-phosphate group (Hale et al. 2008, 2009; Makarova et al. 2015) (see Charpentier, van der Oost and White 2013 for review). Note that in type I-C systems, this reaction is catalysed by Cas5d (Garside et al. 2012; Nam et al. 2012).

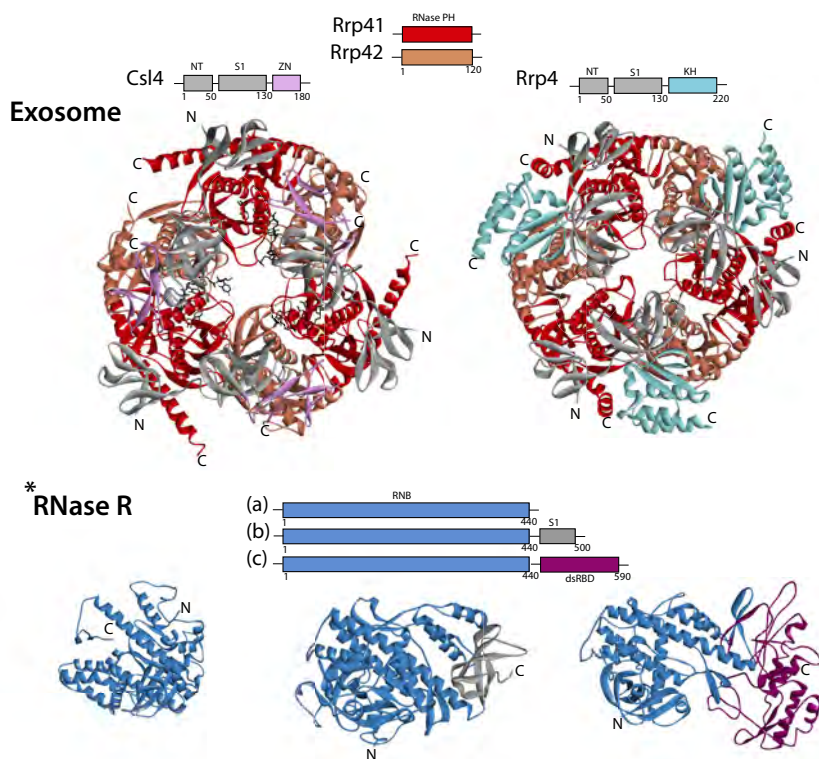
Cas6 protein sequences show limited conservation with only two common motifs: the ferredoxin fold or RRM (RNA recognition motifs) domains and a glycine-rich motif located at the C-terminus (Table 1A, Fig. 2C). Cas6 is a metal-independent nuclease and the predicted active site of the enzyme shows similarities to the active site of the archaeal EndA endoribonuclease. Interestingly, the nature of amino acids located in the active site differs in the Cas6 proteins from the different subtypes. Some Cas6 proteins act as homodimers, whereas others are active as monomers. Several Cas6 crystal structures have been obtained and reveal conserved structural features (see Hochstrasser and Doudna 2015 for a review). Cas6 enzymes contain two RAMP domains that form RRM folds. The C-terminal RRM fold consists of a glycine-rich loop (G-loop), a groove-binding element and a  $\beta$ -hairpin. The G-loop binds and stabilises the RNA substrate, the groove-binding element interacts with major groove of the RNA structure and the  $\beta$ -hairpin is important for positioning the scissile phosphate. In the cleavage reaction, Cas6 interacts with the 5'-region of the CRISPR repeat and cleaves at a specific site within the 3'-region. Substrate recognition of some of the Cas6 proteins evolved together with the CRISPR-repeat sequence, and certain Cas6 proteins can only recognise specific repeat sequences. Depending on the ability of the repeat sequence to form stem-loops, Cas6 proteins either interact with a preformed stable stem-loop structure or fold an unstructured RNA into a stem-loop structure for their cleavage (see Charpentier, van der Oost and White 2013 for a review).

#### Most prevalent archaeal 3'-to-5' exoribonucleases

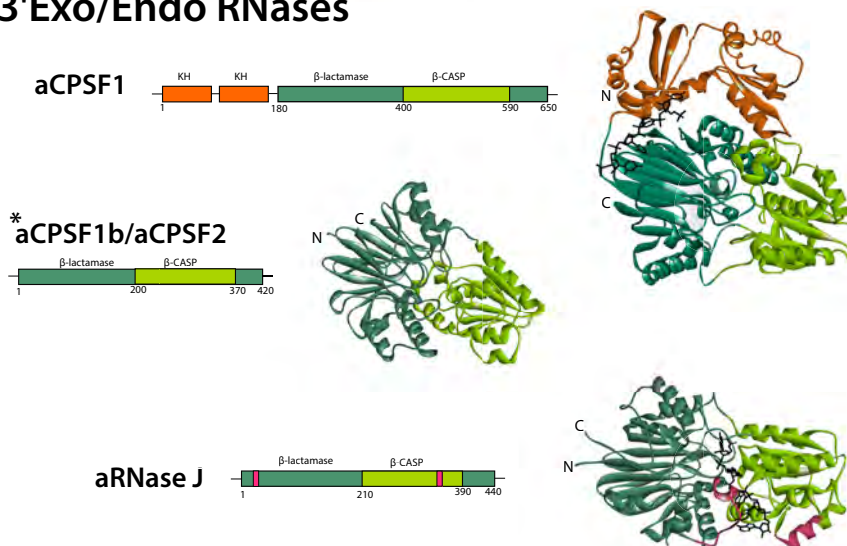
3'-end-RNA trimming activities are critical in all three domains of life. In Archaea, this activity is carried out by two types of 3'-to-5' exoribonucleases (Table 1B). The archaeal exosome machinery, homologous to the eukaryotic exosome, is commonly found in Archaea. In contrast, the archaeal RNase R-like enzyme, homologous to the bacterial RNase R, is only sporadically encountered (Fig. 1, Table 2).

**aExosome** (EC 3.1.13.1) More than a decade ago, the detection of an operon encoding homologues of Rrp4, Rrp41 and Rrp42, eukaryal rRNA-processing proteins that form the core of the eukaryal exosome, in most archaeal genomes strongly suggested the existence of a conserved archaeal 3'-end RNA-degrading machinery (Koonin, Wolf and Aravind 2001) (Table 2). The complete set of exosome protein subunits is conserved in most archaeal genomes with the exception of species from Methanococci, Methanomicrobiales, Haloferacales and Halobacteriales (Fig. 1) (Evguenieva-Hackenberg et al. 2014). Many studies using different archaeal model organisms, in particular Sulfolobales and Pyrococcales, allowed a thorough characterisation of the archaeal exosome by determining its structures and *in vitro* enzymatic activities (see Hartung and Hopfner 2009; Chlebowski et al. 2013; Evguenieva-Hackenberg et al. 2014 for recent reviews). Briefly, the archaeal exosome (aExosome), commonly composed of a nine-subunit core, appears as heterogeneous complexes (Fig. 4A). Consistently, three Rrp41 and Rrp42 subunits, belonging to the PH RNase family, form a hexameric barrel-like structure with a narrow RNA entrance pore to which a variable heterotrimeric RNA-binding cap of Rrp4 and/or Csl4 subunits is tightly bound (Evguenieva-Hackenberg et al. 2003; Ramos et al. 2006; Witharana et al. 2012) (Table 1B). In this context, the Rrp4 subunit confers strong poly(A) specificity to the exosome while Csl4 allows the binding of a aDnaG subunit to the exosome (Witharana et al. 2012; Hou, Klug and Evguenieva-Hackenberg 2013). These cap RNA-binding subunits harbour different types

## A 3'-5' Exo RNases



## B 5'-3'Exo/Endo RNases



**Figure 4.** Domain organisation and structure of archaeal exoribonuclease and  $\beta$ -CASP ribonuclease families. Structure (PDB accession numbers) and corresponding model organism references are reported in Table 1B and marked by an asterisk. Domains correspond to Pfam annotation (Pfam database <http://pfam.xfam.org>). Catalytic sites are highlighted with yellow areas. (A) Structures and domains of the archaeal 3'-to-5' exoribonucleases are shown. Upper panel: archaeal exosomes are usually formed by the association of three modules composed of two subunits Rrp41 and Rrp42 of the PDX protein (only Rrp41 is active). Two different archaeal exosome structures are shown, one with a cap formed by three Csl4 subunits and the other by three Rrp4 subunits. Lower panel: Phyre2-built structural models of the three different forms (annotated a to c) of archaeal RNase R are displayed. The a form (RNase<sup>Ra</sup>) is restricted to the RNB domain and the b and c forms (RNase<sup>Rb</sup> and RNase<sup>Rc</sup>) display 3'-extension RNA binding domains. (B) Structures and domains of the major types of  $\beta$ -CASP ribonucleases are displayed. Only the structures of aCPSF1 and aRNase J have been resolved. Phyre2-built structural model for *S. acidocaldarius* aCPSF2 is shown. aCPSF1 is a unique  $\beta$ -CASP ribonuclease carrying 5'-to-3' exo- and endoribonucleolytic activities with a single catalytic site and composed of a  $\beta$ -lactamase/ $\beta$ -CASP core and an N-terminal extension with two KH motifs. In contrast, aCPSF2 and aRNase J are restricted to the  $\beta$ -lactamase/ $\beta$ -CASP core and carry only a 5'-to-3' exoribonucleolytic activity.

of RNA-binding domains (S1-domain, KH-motif and Zn-ribbon) (Fig. 4A). Three active sites in the lumen of the barrel structure carried by the Rrp41 subunits are responsible for the phosphoryl catalysis of single-stranded RNA from the 3'-end and the release of nucleoside 5'-diphosphates (NDPs) (Lorentzen et al. 2005, 2007; Hartung et al. 2010). The aExosome is also able to perform the reverse reaction by using NDPs to synthesise heteropolymers, such as adenine-rich tails, at the 3'-end of RNAs (Portnoy et al. 2005). The 3'-to-5' exoribonucleolytic and RNA-tailing activities can be modulated with Pi concentration and depend on the ATP:ADP ratio (Evguenieva-Hackenberg et al. 2008).

Despite recent advances in dissecting the aExosome composition and activities through *in vitro* structural and biochemical studies, its contribution in specific biological pathways remains unknown, especially in RNA maturation and RNA decay mechanisms. First clues may come from observations showing an interaction with the splicing enzyme EndA in *M. thermotrophicus* (Farhoud et al. 2005), the S1-domain of the RNA-binding protein TK2227 in *T. kodakarensis* (Li et al. 2010) and the translational elongation factor EF1 $\alpha$  in *S. solfataricus* (Witharana et al. 2012). Moreover, the heterogeneity of the exosomal complexes and its localisation at the cell periphery as observed in *S. solfataricus* may reflect a multilayered regulation of its 3'-tailing or degradation activities upon substrate recognition (Witharana et al. 2012).

The physiological relevance of the composite RNA-binding cap in mechanisms of substrate selection is still lacking. Its understanding may come from the analysis of archaeal exosomes from species in which the set of subunits is incomplete. It will be interesting to understand how aDnaG is recruited in Diaforarchaea species in which the Cls4 subunit appear to be absent. Nevertheless, it should be noted that Csl4 is encoded in a separate operon to Rrp4, Rrp41 and Rrp42 which precludes the identification of potentially highly divergent copies in those genomes. In addition, monitoring how the absence of the Rrp4 subunit in Asgard genomes impacts aExosome activity may help identifying the specific functions of each cap subunit. Finally, the absence of the core subunit Rrp42 in some DPANN may suggest a different aExosome composition in these strains (Fig. 1).

Interestingly, in species from the Methanomicrobiales, Haloferacales and Halobacteriales group, the absence of the exosome seems to be compensated by the presence of other 3'-to-5' exoribonucleases, namely RNase R-like enzymes (see below, Fig. 1). However, in the case of the Methanococci group, no enzyme with an equivalent activity could be detected suggesting that in these strains, a 3'-to-5' exoribonucleolytic activity may not exist and may be compensated by other endo- or 5'-to-3' ribonucleolytic machineries (Clouet-d'Orval et al. 2015).

**aRNase R (EC 3.1.13.1)** As mentioned above, in some cases, the 3'-to-5' exoribonucleolytic activity of an archaeal cell is not dictated by an exosome complex but instead by a bacterial-like RNase R enzyme (Portnoy and Schuster 2006) (Figs 1 and 4A, Table 2). This is the case for halophiles and methanococcales in which the operon encoding the exosomal subunits is missing (Fig. 1). Briefly, RNase R-like enzymes which are characterised by an RNB domain with a unique  $\alpha\beta$ -fold are part of the RNase II/RNB family present in the three domains of life (Fig. 4A, Table 2). It is now well established that bacterial RNase R encoded by the *mnr* gene and composed of the RNB domain flanked by RNA-binding domains is able to degrade structured RNAs from their 3'-ends to their 5'-ends to perform important functions in the cell from RNA quality control to virulence (for

a review, see Arraiano et al. 2013). Most knowledge on archaeal RNase R (aRNase R) comes from *in vitro* biochemical studies carried out with that from *H. volcanii* in which the *mnr* gene has been shown to be essential (Portnoy and Schuster 2006; Matos et al. 2012).

Interestingly, our searches for RNase R-like members in archaeal genomes identified at least three types of domain organisation (arbitrarily named a, b and c here) with none resembling the bacterial enzyme (Figs 1 and 4A and Table 1B). In the halophile group, aRNase R<sup>a</sup> is restricted to the RNB domain with no extra RNA-binding domain. In contrast, two other forms, RNase R<sup>b</sup> and RNase R<sup>c</sup>, found in some methanococcales, contain an extra CTD with RNA-binding properties (Fig. 4A). These different forms of aRNase R have certainly resulted from horizontal gene transfer (HGT) from bacterial genomes. More specifically, we used the archaeal proteins annotated as RNase R as a query to search the NCBI non-redundant bacterial database with blastp. The 200 best hits were retrieved and aligned with the archaeal sequences (mafft program) and the sequence redundancies were removed with Jalview (98% identity threshold). A phylogenetic tree, computed with PhyML, revealed that archaeal sequences were dispersed as several clusters in the bacterial tree, suggesting multiple independent HGTs from bacterial to archaeal genomes. RNase R<sup>a</sup> homologues are found in all analysed Haloferacale and Halobacteriales strains (Fig. 1). All sequences are clustered in a subtree connected to the bacterial sequences with a long branch. This pattern is in agreement with the presence of an RNase R<sup>a</sup> homologue in the last common ancestor of Haloferacales and Halobacteriales. This ancestral gene was probably acquired by lateral gene transfer from Bacteria to Archaea. Furthermore, three RNase R<sup>b</sup> homologue groups (Methanomicrobiales strains, *Methanosaeta concilii* GP6, *Methanocella paludicola* SANAE, *Methanocella arvoryzae* MRE50) are observed (Fig. 1). This is in agreement with three independent HGT events from bacterial strains. Finally, the RNase R<sup>c</sup> homologues are found in one subtree that includes one sequence from Methanocellale strain (*Methanocella conradii* HZ254) and sequences from Methanosarcinales strains (Fig. 1). The location of the *M. conradii* sequence as an outgroup may indicate an initial HGT from Bacteria to Archaea followed by a second HGT between Archaea. However, more data are required to confirm the evolutionary origin of archaeal aRNase R homologues and to understand how they work in RNA processing and RNA decay, and whether they recruit other specific or common factors.

#### The universal $\beta$ -CASP ribonuclease family

In the last few years, a prevalent family of enzymes, namely the  $\beta$ -CASP ribonucleases, has emerged as central in providing additional ribonucleolytic activities in archaeal cells (for a review, see Clouet-d'Orval et al. 2015) (Tables 1A and 1B). Briefly, members of the  $\beta$ -CASP ribonuclease family are of major importance in RNA-processing and RNA decay pathways in Eukarya and Bacteria (for a review, see Dominski, Carpousis and Clouet-d'Orval 2013) (Table 2). Typically,  $\beta$ -CASP members harbour a  $\beta$ -lactamase domain, composed of a four-layered  $\alpha\beta/\beta\alpha$  architecture, in which a so-called  $\beta$ -CASP domain is inserted (Fig. 4B). Several crystal structures of bacterial, eukaryal and archaeal members have reported an identical structural bipartite topology with a catalytic site formed at the interface of the two domains and involving conserved histidine and aspartic acid residues that coordinate two Zn<sup>2+</sup> ions. Importantly,  $\beta$ -CASP ribonucleases have the propensity to carry dual, 5'-to-3' exo- and endo-ribonucleolytic activities with a single active site (Li de la Sierra-Gallay



et al. 2008; Dorléans et al. 2011; Newman et al. 2011). Archaeal  $\beta$ -CASP members have been identified by in-depth phylogenomic analyses that revealed three major archaeal  $\beta$ -CASP clusters designated aCPSF1, aCPSF2 and aRNase J, based upon their close relationship to either bacterial RNase J or eukaryal CPSF73 (cleavage polyadenylation specific factor) (Dominski, Carpousis and Clouet-d'Orval 2013; Phung et al. 2013; Clouet-d'Orval et al. 2015) (Table 2 and Table 1A). Interestingly, the congruence of phylogenetic trees based on aCPSF1, aCPSF2 and aRNase J sequences and a set of concatenated conserved protein sequences reflecting the evolution of the Archaea showed that the three major archaeal  $\beta$ -CASP clusters have been inherited vertically (Clouet-d'Orval et al. 2010; Phung et al. 2013). This indicates that  $\beta$ -CASP ribonucleases are ancient enzymes that existed in the last universal common ancestor before the separation of the actual three domains of life. Features of each  $\beta$ -CASP ribonucleases clusters are briefly introduced below.

**aCPSF1** (CPSF73, EC 3.1.4.1) Members belonging to the aCPSF1 cluster typically contain, in addition to the  $\beta$ -CASP/ $\beta$ -lactamase core domains, a supplementary N-terminal domain composed of two consecutive tandem KH motifs. Three crystal structures of aCPSF1 from *P. horikoshii*, *Methanosarcina mazei* and *M. thermautotrophicus* show a tripartite architecture with the KH-domains link to the central core via a long helical linker and two molecules per asymmetric unit (Nishida et al. 2010; Mir-Montazeri et al. 2011; Silva et al. 2011) (Fig. 4B, Table 1A). It was shown that *P. abyssi* aCPSF1 is dimeric in solution with a highly conserved short C-terminal segment serving as interface (Phung et al. 2013) (see Dominski, Carpousis and Clouet-d'Orval 2013; Clouet-d'Orval et al. 2015 for reviews). The *in vitro* enzymatic activity of thermococcales and methanococcales aCPSF1 revealed a dual activity with an endoribonucleolytic activity at single-stranded CA site and a 5'-to-3' exoribonucleolytic activity restricted to 5'-monophosphorylated RNA substrates (Levy et al. 2011; Phung et al. 2013). These activities are comparable to those displayed by the eukaryal CPSF73 (see Dominski, Carpousis and Clouet-d'Orval 2013 for review). Remarkably, an aCPSF1 member is encoded in each single archaeal genome (Fig. 1). Together, these observations suggest that the dual activities of aCPSF1 are at the centre of critical and conserved metabolic pathways important for the archaeal cell physiologies which remain to be determined. In addition, a small group of  $\beta$ -CASP proteins, aCPSF1b, closely related to aCPSF1 but lacking the N-terminal extension (Phung et al. 2013), is restricted to the Methanococci group in which no 3'-to-5' exoribonucleolytic machineries could be detected by homology searches (Fig. 1). A distantly related sequence is also present in *Korarchaeum cryptofilum* OPF8.

**aCPSF2** (CPSF73, EC 3.1.4.1) The aCPSF2 group was identified as a cluster of less conserved sequences compared to the aCPSF1 cluster (Phung et al. 2013; Clouet-d'Orval et al. 2015) (Table 1B, Fig. 4B). Members of this group, consistently found in the TACK/Proteoarchaeota and more sporadically in Euryarchaeota, are architecturally restricted to the  $\beta$ -CASP/ $\beta$ -lactamase core domains except for the *S. solfataricus* member which possess an N-terminal extension (for reviews, see Dominski, Carpousis and Clouet-d'Orval 2013; Clouet-d'Orval et al. 2015). To date, no aCPSF2 structures have been reported (a model is shown in Fig. 4B), and curiously enough, the 5'-to-3' ribonucleolytic activity detected *in vitro* for Sulfolobales members requires  $Mg^{2+}$  ions in contrast to other  $\beta$ -CASP ribonucleases (Levy et al. 2011). In *S. acidocaldarius*, the gene encoding aCPSF2 is not es-

sential and its deletion is not associated with a modification of the growth rate (Märtens et al. 2013), meaning that, in these conditions, the exoribonuclease aCPSF2 does not affect the overall cellular metabolism.

**aRNase J** (EC 3.1.13.B1) More than a decade ago, homologues of bacterial RNase J were proposed to be encoded in archaeal genomes (Even et al. 2005). It is now clear that archaeal RNase J (aRNase J) exists throughout most Euryarchaea with the exception of the Diafoarchaea and Archeoglobi groups (Fig. 1, Table 1B). In contrast to bacterial RNase J proteins which harbour a CTD, aRNase J enzymes are restricted to the  $\beta$ -CASP/ $\beta$ -lactamase core domains (Fig. 4B) (Clouet-d'Orval et al. 2010). An aRNase J 5'-to-3' exoribonucleolytic activity was first identified for recombinant proteins from Thermococci and Methanococci. *In vitro*, these enzymes show a high processivity, no detectable endoribonucleolytic activity under the conditions tested and a greatly reduced activity in degrading RNA substrates with 5'-triphosphate ends (Clouet-d'Orval et al. 2010; Levy et al. 2011). Recently, the first RNA-free and RNA-bound structures of aRNase J from *Methanobolus psychrophilus* provided detailed mechanistic insights. Furthermore, the unique conserved archaeal loops I and II were found to be involved in the *in vitro* ribonucleolytic activity and in RNA interaction of aRNase J (Clouet-d'Orval et al. 2010; Zheng et al. 2017). Finally, distinct dimerisation/tetramerisation patterns were observed for archaeal and bacterial RNase Js (Zheng et al. 2017).

In conclusion, the roles of archaeal  $\beta$ -CASP ribonucleases in specific biological mRNA decay and RNA quality control mechanisms remain to be elucidated. The challenge will be to identify how  $\beta$ -CASP ribonucleases are specifically recruited and if they are part of multiprotein complexes as are their bacterial and eukaryal counterparts. The expected outcome will clarify the roles in archaeal cell physiology of 5'-to-3' exo- and endo-ribonucleolytic activities carried by this conserved enzyme family.

## Archaeal sRNA-binding proteins

In RNA biology processes, the RNA-associated proteins of the Sm/Hfq and ribosomal L7Ae superfamilies, sometimes called RNA chaperones, have an important role in RNP complex assembly and have emerged as key players in myriad of RNA-processing pathways in Eukarya and Bacteria.

### Archaeal Sm-like proteins

Members of the Sm-like/Hfq proteins form a superfamily of RNA-binding proteins of ancient related origin (Mura et al. 2013). They contain a common bipartite sequence, known as the Sm-domain, consisting of two Sm motifs connected by a region of variable length and motif. And while the proteins do not show a high similarity at the primary structure level, they share a tertiary structure and a multimeric quaternary architecture that supports interactions with partner molecules. Another conserved characteristic of the Sm and Sm-like proteins is their preference for single-stranded uracil-rich RNAs. The ubiquitous eukaryal Sm proteins and their Sm-like paralogues (>18 paralogues) are molecular scaffolds for the assembly of RNP complexes. They were first identified in spliceosomal RNP complexes where they associate with uridine-rich small nuclear RNAs. Hfq, the bacterial branch of the Sm superfamily, was discovered 50 years ago as a host replication factor for the bacteriophage Q $\beta$  (Franze de Fernandez, Eoyang and August 1968). In many bacterial organisms, Hfq is perceived primarily as the core component

of a global post-transcriptional network involving small regulatory RNAs (Nielsen et al. 2007). It is now proposed that homologues exist in the three domains of life and several recent studies try to overcome the gaps in our knowledge on the relevance of this RNA-associated protein family in Archaea.

Archaeal Sm-like/Hfq superfamily proteins were also detected by database searches (Salgado-Garrido et al. 1999). While their biological functions remain to be discovered, they were fundamental in elucidating the biochemical and structural properties of their eukaryal counterparts. To unify the nomenclature, the archaeal homologues have recently been termed SmAPs (Mura et al. 2013). These proteins share the Sm-fold, common to all members of the Sm superfamily even for highly divergent sequences, but in general do not contain any C-terminal extension as observed in some eukaryotic Sm-like and some bacterial Hfq (Mura et al. 2013) (Fig. 3C). Archaea often possess two SmAP paralogues, SmAP1 and SmAP2. In the euryarchaeal Pyrococcales and Halophiles, only one Sm-like protein has been identified. This is orthologous to SmAP1. There is a high-sequence similarity between members of the SmAP1 family, with up to 60% sequence identity (Törö et al. 2001). The identity level drops to 30% or less (Törö et al. 2001) between SmAP1 and SmAP2 proteins from the same organism. Crenarchaeota, like *S. solfataricus* (Sso), *S. tokodaii* (Sto) and *P. aerophilum* (Pae), can possess an additional Sm-like paralogue, SmAP3 (Mura et al. 2003b, 2013). SmAP3 proteins are particular in possessing a longer CTD (Mura et al. 2003b). Finally, *Methanococcus jannaschii* (Mja) contains an Hfq-like protein (Nielsen et al. 2007) rather than a conventional Sm-like protein (Salgado-Garrido et al. 1999).

**SmAP1 proteins** The crystal structures of many archaeal SmAPs have been resolved since the beginning of the 2000s. In particular, the crystal structures of SmAP1 cyclic homoheptamers of crenarchaeota *P. aerophilum* (Pae-SmAP1/PDB code 1ISF) (Mura et al. 2001), and of euryarchaeota *A. fulgidus* (Afu-SmAP1/PDB code 1I4K) (Törö et al. 2001), *P. abyssi* (Pab-SmAP1/PDB code 1H64) (Törö et al. 2002) and *M. thermautotrophicum* (Mth-SmAP1/PDB code 1IS1) (Collins et al. 2001) (Fig. 3C) strengthen the proposed model for human Sm core domain (Kambach, Walke and Nagai 1999) by providing the first atomic resolution of Sm monomers in an intact ring.

SmAP1 monomers adopt a barrel-type structure consisting of an N-terminal  $\alpha$ -helix followed by a strongly bent five-stranded  $\beta$ -sheet (Collins et al. 2001; Mura et al. 2001; Törö et al. 2001, 2002). The highly curved nature of the  $\beta$ -sheet allows the formation of a compact hydrophobic core. The cyclic heptameric ring, composed of seven monomers (7-mer), involves interactions between the  $\beta$ 4 and  $\beta$ 5-strands of adjacent monomers to form a continuous  $\beta$ -sheet in the oligomer. The loop L4 that links the strands  $\beta$ 3 and  $\beta$ 4 and that define the distal face varies more in length and amino acid sequence than any other loops. Most SmAP1s exhibit a short loop L4 when compared to eukaryotic Sm-like proteins and do not include extended N- and C-termini, with the exception of Mth-SmAP1 that contains a slightly longer N-terminal sequence when compared to other archaeal proteins (Collins et al. 2001). The cyclic heptameric complex is extremely stable and can resist denaturing conditions, even in the absence of RNA, as shown for eukaryotic Sm-like but not Sm proteins. The residues in loops L2, L3 and L5 lie towards the central cavity of the ring and residues located in loops L3 and L5 form a conserved uridine-binding pocket (Törö et al. 2001), also called internal binding site (Thore et al. 2003). This central cavity is surrounded by a positively charged region on the loop L4/distal face (Törö et al. 2001; Thore et al. 2003). Beyond

their capacity to form cyclic heptameric rings, SmAP1 proteins can also assemble into two stacked heptameric rings and, like Pae- and Mth-SmAP1s, can also polymerise into fibrillary ultrastructures (Mura et al. 2003a, b). This oligomeric plasticity of SmAPs, which is observed across the entire protein family and may have unsuspected biological significance, will not be further developed in this review (see Mura et al. 2013 for review).

Both Afu- and Pab-SmAP1 proteins were shown to preferentially bind U-rich RNA molecules (Törö et al. 2001; Thore et al. 2003). Binding of U5 or U7 oligonucleotides by Afu-SmAP1 (PDB code 1I5L) (Törö et al. 2001) or Pab-SmAP1 (PDB 1M8V) (Törö et al. 2002; Thore et al. 2003), respectively, does not induce major structural changes and involves the conserved His37, Asn39 and Arg63 residues of the internal binding site of the same monomer. A second binding site, the external binding site, was identified in the crystal structure of the Pab-SmAP1/U7 complex (Thore et al. 2003). This site consists of the residues Arg4 from the N-terminal  $\alpha$ -helix and Tyr34 from the  $\beta$ 2-strand of two monomers of two cyclic heptamers in head-to-head orientation. The Tyr34 residue is absolutely conserved in all the archaeal SmAP1 proteins (Thore et al. 2003). Its substitution by a valine does not affect Pba-SmAP1 binding to an oligo(U), but it strongly reduced binding to the Eukaryotic Sm consensus 'AAUUUUUG' RNA. These results suggest an interplay between the two RNA-binding sites where the external binding site stabilises additional nucleotides after a specific recognition of the U-rich sequence by the internal binding site (Thore et al. 2003).

**SmAP2 proteins** Crystal structures of SmAP2 proteins from *S. solfataricus* (Märtens et al. 2017; PDB code 4XQ3) and *A. fulgidus* (Törö et al. 2002; PDB code 1LJO) have also been resolved. Afu-SmAP2 has essentially the same monomer fold as Afu-SmAP1 and harbours the conserved uridine-binding pocket (with the exception of His37 which is replaced by Tyr), suggesting a common RNA-binding mode for both proteins. Band-shift experiments confirmed that Afu-SmAP2 has a clear preference for U-rich sequence (Achsel, Stark and Lührmann 2001). However, conversely to Afu-SmAP1, the oligomerisation of Afu-SmAP2 is strongly dependent on the pH and the presence of RNA molecules. Indeed, while the electron micrographs of an Afu-SmAP2/U10 complex show that in the presence of an oligo(U), Afu-SmAP2 can assemble into a cyclic heptameric ring (Achsel, Stark and Lührmann 2001), at acidic pH and in the absence of RNA, Afu-SmAP2s form a cyclic hexameric (6-mer) structure in the crystals that is not stable (Törö et al. 2002). The opposite charge distribution of the L4/distal face that is predominantly positive in Afu-SmAP1 and negative in Afu-SmAP2 creates differences in the ionic interactions between monomers that could explain the different oligomerisation behaviour of Afu-SmAP1 and 2 (Törö et al. 2002). Nonetheless, while Sso-SmAP1 and -SmAP2 also exhibit similar opposite charge distribution, Sso-SmAP2 proteins were shown to assemble into cyclic heptameric rings in the crystals (Märtens et al. 2015). Finally, another clear difference between free Afu-SmAP1 heptamers and Afu-SmAP2 hexamers is the diameter of the central cavity of the Afu-SmAP2 ( $\approx 8\text{\AA}$ ) that is smaller than that of SmAP1 proteins (10–15 $\text{\AA}$ ) and that will not allow RNA binding without significant structural rearrangement (Törö et al. 2002). As suggested by the electron micrographs of Afu-SmAP2/U10 complexes, it was proposed that assembly around a U-rich RNA sequence may lead to the formation of an Afu-Sm2 heptamer (Törö et al. 2002).

**SmAP3 proteins** SmAP3 from *P. aerophilum* (Pae-SmAP3) contains the conserved Sm-domain linked to a novel CTD of nearly 70

residues. The crystal structure of Pae-SmAP3 shows that these additional residues form a compact CTD with a novel mixed  $\alpha/\beta$  fold (Mura et al. 2003b). The strong sequence conservation of Pae-, Sto- and Sso-SmAP3 CTDs suggests that they will have a similar structure (Mura et al. 2003b). A comparative structural analysis found weak structural similarity between the CTD of Pae-SmAP3 and a CTD of yeast TATA-box binding protein (Mura et al. 2003b). Although it is unclear how the CTD of SmAP3 would modify its function (biochemical or structural), it suggests that SmAP3s could have novel nucleic acid binding properties and may be used for physiological functions entirely unrelated to RNA metabolism (Mura et al. 2003b). Moreover, Pae-SmAP3 monomers assemble into an unexpected 14-mer structure composed of two heptamers with different conformations (apical or equatorial) that is perforated by a cylindrical pore and is bound to 14 cadmium ions (Mura et al. 2003b).

Interestingly, it was shown that Sso-SmAP1, Sso-SmAP2 and Sso-SmAP3 can co-precipitate, suggesting that *in vivo*, the Sso-SmAPs could assemble into hetero-oligomers (Märtens et al. 2017). Similar behaviour was also previously observed for Afu-SmAP1 and Afu-SmAP2 (Törö et al. 2001). While hetero-oligomeric SmAP assemblies remain to be characterised in Archaea, a differential composition could impact their biological functions as shown for eukaryal Sm-like hetero-heptamers (Mura et al. 2013).

**Hfq-like proteins** A homologue of bacterial Hfq was identified in the genome of *M. jannashii* which does not contain an open reading frame encoding a SmAP protein (Salgado-Garrido et al. 1999; Nielsen et al. 2007; Mura et al. 2013). Like bacterial Hfq, Mja-Hfq contains the conserved Hfq Sm2 motif (YKHA consensus) but possesses a shorter C-terminal tail and N-terminal  $\alpha$ -helix compared to Hfq homologues of *Escherichia coli* (Eco) and *Staphylococcus aureus* (Nielsen et al. 2007; Mura et al. 2013). The charge distribution on loop L4 of Eco-Hfq and Mja-Hfq is also different as Mja-Hfq L4 is predominantly negatively charged (Nielsen et al. 2007; Mura et al. 2013). The resolution of Mja-Hfq crystal structure (PDB code 2QTX) identified a cyclic hexameric complex also characteristic of Hfq proteins (Nielsen et al. 2007). Mja-Hfq preserves the conserved uridine-binding pocket (also named distal binding site) but the distal adenine-binding site and lateral RNA-binding site show considerable structural changes compared to bacterial Hfq (Nikulin et al. 2017). Despite these differences, Mja-Hfq is able to partially complement the pleiotropic phenotypes of Hfq knockout mutants in *E. coli* (Sittka et al. 2009) and *Salmonella enterica* (Nielsen et al. 2007). Moreover, Mja-Hfq forms a ternary complex with the Spot42 sRNA and its *sucC* mRNA target *in vitro* (Nielsen et al. 2007). Finally, Hfq homologues were also discovered on four Thermococcales plasmids and three unrelated Methanococcales plasmids (Krupovic et al. 2013). These archaeal Hfq proteins are particular and contain an additional N-terminal C<sub>2</sub>H<sub>2</sub>-type zinc finger domain suggesting an additional potential role in DNA or RNA binding (Krupovic et al. 2013).

#### Archaeal L7Ae ribosomal protein

L7Ae is a member of the L7Ae/L30 protein family found in both Archaea and Eukarya. The members of the L7Ae/L30 protein family share the capacity to bind K-turn RNA motifs (Mao, White and Williamson 1999; Ban et al. 2000; Kuhn, Tran and Maxwell 2002) (Fig. 3A). Archaeal L7Ae is homologous to the human 15.5kD protein and yeast Snu13 small nucleolar RNP core proteins and shares their capacity to bind the K-turn of C/D and H/ACA-box sRNAs (Kuhn, Tran and Maxwell 2002; Rozhdestvensky et al. 2003). In contrast to 15.5kD/Snu13 proteins, L7Ae also

has the ability to recognise Kink-loop (K-loop) motifs (Vidovic et al. 2000; Hamma and Ferré-D'Amaré 2004; Nolivos, Carpousis and Clouet-d'Orval 2005; Li and Ye 2006; Soss and Flynn 2007; Gagnon et al. 2010). An alignment of several archaeal L7Ae and eukaryotic 15.5 kD proteins uncovered a highly conserved RNA-binding region containing amino acid residues important for the binding of K-turn motifs (Gagnon et al. 2010).

The archaeal L7Ae protein is a component of the large ribosomal subunit of *Haloarcula marismortui* (Ban et al. 2000) and serves as a core protein of the C/D and H/ACA box sRNP nucleotide modification complexes (Kuhn, Tran and Maxwell 2002; Tang et al. 2002; Rozhdestvensky et al. 2003; Charron et al. 2004a; Baker et al. 2005; Gagnon et al. 2010). It is also part of the archaeal RNase P complex (Kouzuma et al. 2003; Fukuhara et al. 2006; Lai et al. 2014; Oshima et al. 2016) (see section 'Ubiquitous endoribonucleases'). Moreover, recent evidence also suggests that L7Ae can regulate translation by direct binding to the 5'-untranslated regions of some messengers, including its own transcript (Daume et al. 2017). Finally, an *in silico* approach in *P. abyssi* identified three unknown small ncRNAs containing K-turn motifs as interacting partners of L7Ae (Phok et al. 2011). Together, the presence of L7Ae in RNase P, C/D and H/ACA box sRNPs and ribosome in Archaea offer fascinating prospects for future integrative studies in understanding coordination in the regulation of macromolecular machineries involved in specific facets of translation.

## Processing pathways of archaeal ncRNAs

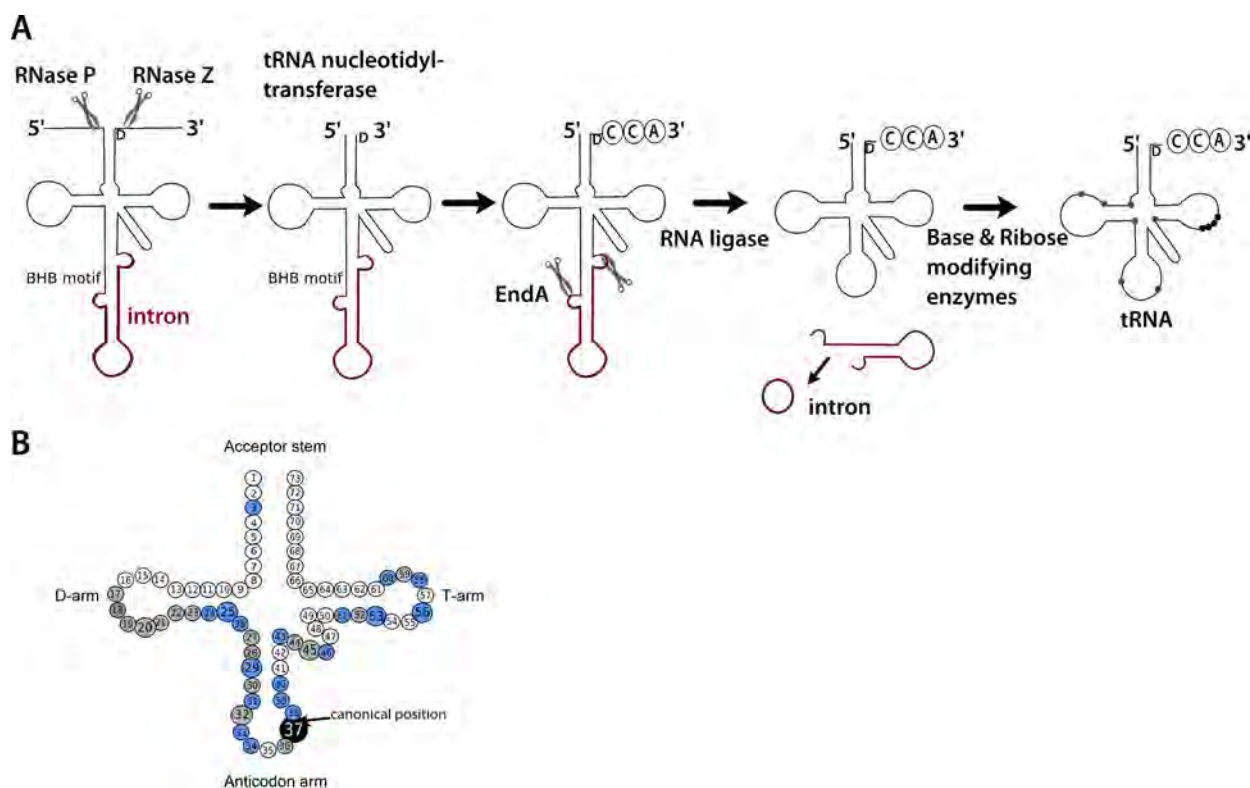
### Maturation of tRNAs

Maturation of tRNAs is a highly conserved process in all three domains of life. The tRNA sequence is transcribed as a longer precursor (pre-tRNA) containing a 5'-leader and a 3'-end trailer sequence (Fig. 5A). Thus release of the mature tRNA requires processing at both transcript ends, and often post-transcriptional addition of a 3'-CCA-end. Furthermore, extensive nucleoside modifications and, in some cases, removal of introns are needed.

The number of tRNA genes in archaeal genomes ranges from 36 in *M. smithii* to 61 in *M. bakeri* (Chan and Lowe 2009, <http://gtrnadb.ucsc.edu/>). The archaeal tRNA genes are either encoded as single transcriptional units in operons together with other tRNAs or in polycistronic units in combination with other transcripts such as ribosomal and messenger RNAs. A recent analysis of transcriptional start sites in *H. volcanii* identified individual sites for 39 of 52 tRNAs (Babski et al. 2016), whereas in *M. thermoautotrophicus* 10 of 39 tRNAs are apparently single-gene transcriptional units and two operons exist with five transcriptionally linked tRNAs (Smith et al. 1997).

Remarkably, archaeal tDNA can adopt a variety of non-canonical arrangements. In *N. equitans* six tRNA genes are split with their 5'-halves containing a sequence stretch at their 3' end that can base pair with the sequence preceding the cognate 3'-halves to form a non-canonical BHB motif (Fig. 3B). After recognition by EndA, both halves are joined in a trans-splicing reaction that releases a mature tRNA (Randau et al. 2005). Disrupted tRNAs are not only found in reduced-genome archaea but also exist in free-living Crenarchaeota such as *Caldivirga maquilgensis*, *Aeropyrum pernix*, *Staphylothermus marinus*, *Staphylothermus hellenicus* and *Thermosphaera aggregans* (Chan, Cozen and Lowe 2011). Interestingly, the location of the disruption corresponds in most cases to the canonical site of tRNA introns between position 37 and 38 (Fig. 5B) although it can also be positioned between nucleotides 30/31 or 59/60 in some *Staphylothermus* species, *N. equitans* and *Thermofilum pendens*, respectively (Chan, Cozen and





**Figure 5.** Processing pathways of transfer RNA precursors. (A) Maturation of a tRNA precursor in Archaea. The tRNA precursor is processed at the 5'-end by RNase P, while the 3'-end is generated by RNase Z cleavage. If not encoded by the tDNA, the CCA triplet is added by the tRNA nucleotidyl-transferase. If an intron (brown) is present within the tRNA, the Bulge-Helix-Bulge RNA (BHB) motif formed at the exon-intron junctions is recognised and cleaved by EndA. Subsequently, the tRNA-halves are joined by the RtcB RNA ligase and the intron is circularised. tRNA maturation is finalised by nucleoside modifications at various positions (black and grey dots). The order of processing events given here is deduced from work in *H. volcanii*, different processing orders have been observed as explained in the text. For references and further details, see section 'Maturation of tRNAs'. (B) Intron positions in archaeal tRNAs. The secondary structure of a tRNA is shown with the tRNA universal numbering (Sprinzl et al. 1987). Filled circles indicate the nucleotide located 5' to the insertion site (including tRNA interruptions found in split-tRNAs). Positions conserved across phyla are indicated in grey, and sites specific for Thermoproteales are given in blue. Larger circles indicate positions with high frequency (more than 10) according to Sugahara et al. (2008), Fujishima et al. (2009, 2011).

Lowe 2011). In *C. maquilingensis*, tRNA genes are even split into three segments (Fujishima et al. 2009). Finally, as an extreme case of fragmented tDNAs, permuted tRNA genes have been found where both tRNA-halves are part of one operon but the 3'-half precedes the 5' portion (Chan, Cozen and Lowe 2011). Generation of the full-length tRNA depends on a BHB motif encoded by the termini of the pre-tRNA. Upon splicing, the two halves are joined into circular intermediates that are subsequently opened by the action of RNase P and RNase Z to form a mature tRNA (Chan, Cozen and Lowe 2011). Altogether, the archaeal tDNA landscape codes for a multitude of tRNA-precursor variants, each requiring specific maturation pathways (Fujishima and Kanai 2014).

**5'-end tRNA processing** The 5'-leader sequence of a pre-tRNA is removed in all organisms by the universal RNase P (Figs 1 and 5A). RNase P substrate recognition relies on distinct structural characteristics of the conserved tRNA 3D structure. Especially critical is the so called 'elbow' where D- and T-loop nucleotides interact via base-stacking to form a hydrophobic surface (Zhang and Ferré-D'Amaré 2016). As shown for bacterial RNase P, the distance along the T-arm-acceptor-stem-axis correctly positions the cleavage site (Reiter et al. 2010). Other determinants are the +1 nucleotide of mature tRNA, the discriminator nucleotide at position 73 and the pre-tRNA-leader (Kirsebom 2007 and references therein).

**3'-end tRNA processing** Processing of the pre-tRNA 3' end is catalysed by the universal endonuclease RNase Z (Figs 1 and 5A). RNase Z recognises key substructures of the tRNA molecule, especially the T-arm and acceptor stem and cleaves directly after the discriminator base in most cases (see section 'Ubiquitous endoribonucleases'). In the majority of Archaea (e.g. *H. volcanii*, *T. acidophilum* or *N. equitans*), the CCA-end essential for amino acid attachment, accurate positioning for peptide-bond formation and tRNA quality control (Wilusz et al. 2011) is not encoded in the tDNA (Hartmann et al. 2009). Therefore, the CCA-motif is added by the nucleotidyl-transferase after RNase Z cleavage. Consequently, tRNA molecules with post-transcriptionally added CCA-ends should not be substrates for RNase Z since this would result in an endless cycle of CCA-addition and cleavage (Minagawa et al. 2004). Indeed, the presence of a CCA in the 3'-trailer inhibits cleavage by *T. acidophilum* RNase Z, whereas RNase Z enzymes from *P. aerophilum* and *M. jannaschii* process CCA-containing trailers *in vitro* (Schiffer, Rösch and Marchfelder 2003; Minagawa et al. 2004; Späth et al. 2008). Both species are examples for the uncommon case of archaeal genomes encoding a mixed population of tDNAs with varying portions of encoded and non-encoded CCA-ends (Marck and Grosjean 2002). Whether CCA removal is prevented by a protein factor or fast amino acetylation *in vivo* waits to be determined. Currently, no information is available on the mode of 3'-processing in the rare cases of archaeal species such as *A. pernix* or *P. furiosus* that

encode only CCA-containing tRNA genes (Marck and Grosjean 2002). *Pyrococcus furiosus* RNase Z has only been studied *in vitro* and in combination with non-CCA substrates (Hölzle et al. 2008; Späth et al. 2008). However, the bacterial RNase Z in *Thermotoga maritima*, where almost all tDNAs encode the CCA triplet, does not process at the discriminator base but after the CCA due to differences in substrate positioning by this bacterial RNase Z (Minagawa et al. 2004; Ishii et al. 2005). In bacterial systems, with encoded CCA-ends, a common 3'-end maturation pathway is an exonucleolytic processing (Li and Deutscher 1996). But what holds true for archaeal systems with CCA-encoding tDNAs still awaits to be determined.

The post-transcriptional addition of the CCA-end is achieved by the enzyme ATP(CTP):tRNA nucleotidyltransferase (EC 2.7.7.25) also termed the CCA-adding-enzyme or tRNA terminal transferase (Deutscher 1990) (Fig. 5A). It catalyses the template and primer-independent addition of nucleotides and is either essential if CCA is not encoded in the tDNA or required for maintenance of CCA integrity (Zhu and Deutscher 1987; Aebi et al. 1990). The terminal transferase is conserved in all domains of life but two classes distinct in sequence and especially structure exist: class I enzymes are present in Archaea and class II enzymes found in Eukarya and Bacteria (Yue, Maizels and Weiner 1996). Both classes share only a homologous N-terminal catalytic core, whereas the other domains (neck, body and tail) differ structurally as well as in their arrangement resulting in a distinct overall architecture: U shaped in the class I of *A. fulgidus* opposed to the sea-horse shape of class II enzymes (Okabe et al. 2003). Co-crystallisation of the *A. fulgidus* class I terminal transferase dimer with different substrates shows that the tRNA T-arm-acceptor-stem-half is recognised by a complementary shaped cleft in the body and tail domain (Tomita et al. 2006). The tRNA elbow is contacted by the tail domain placing the discriminator base between head and neck domain fixing the 3'-end within the active site (Zhang and Ferré-D'Amaré 2016). All interactions except for the base at position 72 are to the phosphate-sugar-backbone of tRNA allowing for sequence-independent recognition (Xiong and Steitz 2004). The specificity of the nucleotide-binding pocket of the *A. fulgidus* terminal transferase is not a function of the protein residues alone but co-catalytically adapted by influence of the growing 3'-terminus during CCA-addition resulting in switching from CTP to ATP specificity for the final nucleotide transfer (Tomita et al. 2006; Pan, Xiong and Steitz 2010). A growing body of evidence suggests that the terminal transferase can add more than one CCA to a tRNA or tRNA-like 3'-end acting as a degradation signal in the course of RNA surveillance (Wilusz et al. 2011). Indeed, as shown for the *A. fulgidus* terminal transferase *in vitro*, the tRNA is twisted and compressed by movement of the head domain after the first CCA-addition and if the tRNA structure is destabilised a bulge will form upon torsion and a second CCA cycle could ensue (Kuhn et al. 2015).

**tRNA intron splicing** On average, 15% of archaeal tRNAs contain introns ranging in size from 11 to 129 nucleotides (Yoshihisa 2014). This ranges from 48% intron-containing tRNAs in Crenarchaeota and almost 70% in Thermoproteales. Multiple introns are encoded in 17% of all intron-containing tRNAs as for example in *T. pendens* or *P. calidifontis* tRNAs (Marck and Grosjean 2003; Sugahara et al. 2008; Yoshihisa 2014). However, tRNA introns are present across all domains of life and their canonical position is between nucleotide 37 and 38 in the anticodon loop. In Archaea, they have been found at more than 30 non-canonical positions along the entire tRNA molecule (Marck and

Grosjean 2003) (Fig. 5B). Introns are mostly located between the D-arm and the anticodon loop in archaeal species but also in the 3'-half in Thermoproteales (Fujishima et al. 2010). Archaea do not possess self-splicing group I introns but use an archaea-specific splicing mechanism that requires an enzyme with homology to parts of the eukaryotic tRNA splicing endonuclease (Table 2) (Lykke-Andersen et al. 1997; Li, Trotta and Abelson 1998). Archaeal introns are characterised by BHB motifs recognised and cleaved by the EndA (Thompson and Daniels 1990) (described in section 'Ubiquitous endoribonucleases'). The BHB motif or a variant thereof is the necessary and sufficient determinant for a splicing reaction in Archaea (Fig. 3B). The arbitrary location of archaeal introns and tRNA splits is only possible because of this 'motif-only' definition of splice sites. Non-canonical BHB motifs recognised by variant forms of EndA are often associated with tRNA introns at non-canonical positions or split-tRNAs (Marck and Grosjean 2003; Hirata et al. 2012) (Fig. 3B).

After cleavage by EndA, the exons are joined by the RNA ligase and the intron is released as a circularised form (Fig. 5A). BHB motifs formed by sequences adjacent to the junction in split-tRNAs trigger a trans splicing reaction used to combine the individually encoded tRNA-halves (see above, Randau et al. 2005; Chan, Cozen and Lowe 2011). The 3'-phosphate RNA splicing ligase was only recently purified from *M. kandleri* and assigned to the RtcB protein family (Englert et al. 2011). A tRNA ligation activity was demonstrated for *P. aerophilum* RtcB *in vitro* and the *P. horikoshii* RtcB structure revealed a novel active centre composition and structural fold (Englert et al. 2012). RtcB homologues are found across all domains even in bacteria although bacterial tRNAs contain self-splicing group I introns. Thus, the bacterial RtcB homologue cannot catalyse a splicing-associated ligation reaction but might be used to generate circular RNAs (see section 'Small non-coding box C/D and box H/ACA guide RNA biogenesis').

**Post-transcriptional tRNA chemical modifications** Natural RNA molecules contain over 100 post-transcriptional modified nucleosides, needed for cellular functions of RNAs in all three domains of life (see Agris 2015; Agris et al. 2017; Våre et al. 2017 for a review). Chemical modifications of the canonical adenosine, guanosine, cytidine and uridine nucleosides are generated enzymatically during the maturation processes of major RNA species (tRNA, rRNA and mRNA). Briefly, modifications include pseudouridine formation, methylations, deamination, carbamoylation, acetylation, carboxylation and addition of amino acids, all of which can be found in two dedicated databases (RNAMDB, <http://rna-mdb.cas.albany.edu/RNAmods/> (Cantara et al. 2011) and Modomics <http://modomics.genesilico.pl/> (Machnicka et al. 2013)). RNA modification biosynthesis requires simple but also sequential and multiple enzymatic activities.

With a large variety of post-transcriptionally modified nucleosides, tRNAs are the most extensively chemically modified RNA entity (17% of the nucleosides in average) with 22 modified nucleosides universally retrieved in all organisms (Motorin and Helm 2011; El Yacoubi, Bailly and de Crécy-Lagard 2012; Grosjean and Westhof 2016) (Fig. 5A). Some of these modifications have been shown to facilitate RNA folding, structure and conformational dynamics. More precisely, in Archaea, at least 47 different modifications have been identified with some of them, such as archaeosine or agmatidine, being specific to this domain of life (Grosjean et al. 2008; Phillips and de Crécy-Lagard 2011). The numerous and diverse chemical moieties in tRNAs affect global 3D conformation by providing unique recognition determinants important for effective tRNAs decoding ability (see Våre et al. 2017



for a review). Indeed, chemical moieties at position 34 (wobble base) and 37 (dangling base) in the tRNAs anticodon loop expand the genetic code by stabilising codon–anticodon pairs (Grosjean, de Crécy-Lagard and Marck 2010; Grosjean and Westhof 2016). As an example, the universal N6-threonylcarbamoyl adenosine (t6A37), which occurs exclusively at the position 31 of all tRNAs decoding ANN codons ( $N \frac{1}{4}$  A, C, U or G), facilitates tRNA binding to the ribosome and prevents frameshifting in the course of translation (see Pichard-Kostuch et al. 2017 for a review).

**Order of tRNA-processing events?** A recent dissection of *M. kandleri* RNA-seq data pin points the identification of tRNA precursors before and after the unique C8-to-U8 editing activity and enables the determination of the order of tRNA-processing events (Su, Tripp and Randau 2013). According to this, pre-tRNA is first processed at its 3'-end followed by a cleavage at the 5'-end, before nucleotide modifications are introduced and C-to-U editing occurs, followed by final removal of introns (Su, Tripp and Randau 2013). In contrast, *in vitro* studies show that a precursor containing a 5'-leader sequence or an intron is only processed by the recombinant RNase Z from *P. furiosus* but not by the *H. volcanii* one (Späth et al. 2008). Furthermore, a yeast tRNA precursor expressed in *H. volcanii* cells undergoes 5'- before 3'-processing (Palmer, Nieuwlandt and Daniels 1994). However, unpublished results from the Marchfelder lab demonstrate that upon depletion of RNase P in *H. volcanii* cells, the 5'-end but not the 3'-end tRNA maturation is impaired. The order of the tRNA-processing events is therefore not clearly defined and a certain degree of flexibility exists. The expanding number of archaeal whole-transcriptome analyses will certainly facilitate our understanding of the degree of this plasticity.

### Ribosomal RNA biogenesis

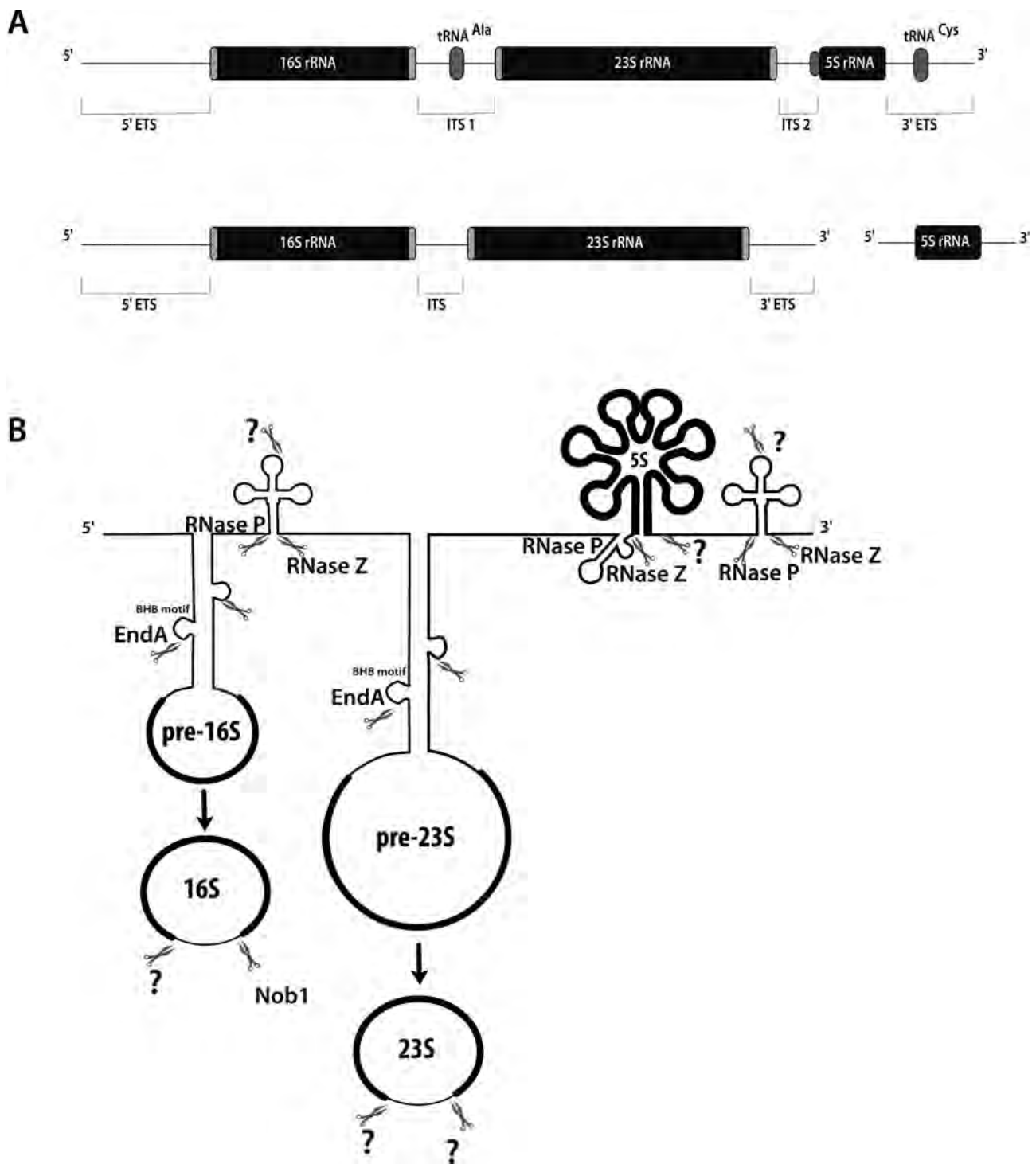
Ribosomes function in protein synthesis in all domains of life and are universally conserved macromolecular machineries composed of rRNAs and specific sets of ribosomal proteins. Eukaryal and bacterial ribosome biogenesis pathways are complex and involve RNA processing and ribosomal protein assembly that require a series of specific enzymes and assembly factors acting in coordinated and multilayered manner. Only a few steps in the archaeal pathways have been currently identified, but the larger part remains to be explored (for a review, see Ferreira-Cerca 2017) (Fig. 6). Despite the universal conserved status of the RNA moieties of the ribosome with a common core structure across all domains of life, domain-specific rRNA expansion segments mostly localised at the ribosome periphery are present in Archaea and Eukarya (Petrov et al. 2014). Indeed analysis of 16S rRNA sequences led to the recognition of Archaea as a separate domain of life (Woese, Kandler and Wheelis 1990; Albers et al. 2013; Gulen et al. 2016). In addition, ribosomal proteins vary in size and number across the three domains with 32 universal and more than 60 domain-specific ribosomal proteins (Márquez et al. 2011) (for a review, see Ferreira-Cerca 2017). Typically archaeal ribosomes sediment at 70S and are composed of the 23S, 16S and 5S rRNAs.

**Diversity of rRNA operons in archaeal genomes** Across all domains of life, rRNA encoding genes are organised in *rrn* operons (Lafontaine and Tollervey 2001). The majority of archaeal genomes possess only one *rrn* operon. However, up to four operons, scattered on the chromosome, are present in some euryarchaeal species such as *M. vannielii* or *Halolamina sediminis* (Acinas et al. 2004; Stoddard et al. 2015). The *rrn* operon organisation differs across archaeal lineages, and disrupted and im-

balanced rDNA arrangements can be found. In general, euryarchaeal rDNA typically comprises the 16S and 23S rRNA genes linked by an internal transcribed spacer (ITS) and the 5S rRNA gene framed by 5'- and 3'-external transcribed spacers (ETS) (Klug et al. 2007). The ITS typically encodes a tRNA<sup>Ala</sup> as found in the model Euryarchaeota *M. mazei* or *H. volcanii* (Deppenmeier et al. 2002; Hartman et al. 2010) (Fig. 6A). The 3'-flanking ETS occasionally encodes an additional tRNA as in *H. volcanii* and *Haloquadratum walsbyi* the tRNA<sup>Cys</sup> (Bolhuis et al. 2006; Hartman et al. 2010) and the 5'-ETS of *M. fervidus* *rrn* operon contains the 7S RNA and tRNA<sup>Ser</sup> genes (Haas, Daniels and Reeve 1989). Often, crenarchaeal *rrn* operons do not encode tRNAs in their intergenic sequences and encode the 5S rRNA in a separate transcriptional unit elsewhere in the genome as in *S. acidocaldarius* or *T. tenax* (Durovic and Dennis 1994; She et al. 2001; Siebers et al. 2011). Analysis of available genomic data shows that this arrangement of coupled 16S-23S rRNA and separate 5S transcription without interspersed tRNAs is also common throughout the TACK superphylum in *Candidatus Korarchaeum cryptofilum* OPF8, *Cenarchaeum symbiosum*, *Candidatus Caldarchaeum subterraneum* and *P. furiosus* (Hallam et al. 2006; Elkins et al. 2008; Nunoura et al. 2011; Ikeda et al. 2017). In some specific cases, archaeal genomes possess a completely different rDNA organisation as for instance in *M. janaschii* where one *rrn* operon encodes all three rRNAs, while the other is missing the 5S rRNA (Bult et al. 1996) or in *N. equitans* and *T. acidophilum* where all three rRNAs are transcribed separately (Ree and Zimmermann 1990; Waters et al. 2003).

**Processing of primary rRNA into mature 16S and 23S rRNAs** Ribosomal RNAs are part of a precursor transcript (pre-rRNA), which is matured into the functional rRNAs in a series of steps. In Bacteria and Eukarya, those pathways involve the coordinated action of endo- and exoribonucleases as well as RNA modification machineries which have been best characterised in *E. coli* and *S. cerevisiae* (Deutscher 2009; Henras et al. 2015). In Archaea, these steps have been investigated by analyses of processing products and intermediates identified by nuclease protection assays, primer extensions and cDNA analyses and *in vitro* processing experiments (Chant and Dennis 1986; Durovic and Dennis 1994; Dennis, Ziesche and Mylvaganam 1998; Ciammaruconi and Londei 2001; Tang et al. 2002). As a general feature in Bacteria, primary polycistronic rRNA precursors contain flanking inverted repeat regions at the 5'-leader and 3'-trailer sequences of the respective 16S and 23S rRNA precursors (16S and 23S pre-rRNAs, respectively) (Deutscher 2009). These repeats form extended helical structures which are cleaved by the bacterial RNase III within the double-stranded region to produce the 16S and 23S pre-rRNAs. In Archaea, these stems contain a BHB motif recognised by the splicing endonuclease EndA (Fig. 6B) as occurs in the tRNA splicing reaction (Chant and Dennis 1986; Thompson and Daniels 1990). Hence, the 16S and 23S pre-rRNAs are circularised and the ETS and ITS sequences adjacent to the BHB motif are ligated as shown in *S. solfataricus*, *S. acidocaldarius*, *H. salinarum* and *A. fulgidus* (Tang et al. 2002; Danan et al. 2012). The circular 16S and 23S pre-rRNAs are quickly opened by endoribonucleolytic cleavages and trimmed by unidentified endo- and exo-ribonucleases, respectively, to yield the mature transcripts (Chant and Dennis 1986; Tang et al. 2002). The 5'-end appears to be matured faster than the 3'-end, which might reflect a sequential assembly of the ribosome or a sequential formation of structural elements needed for processing (Chant and Dennis 1986; Dennis, Ziesche and Mylvaganam 1998).

In addition, a second pathway for the direct generation of the mature 5'-end via an endoribonucleolytic activity was described



**Figure 6.** Processing pathways of ribosomal RNA precursors. (A) Structure of two archaeal *rm* operons. The structure of the *rm* operons of *H. volcanii* (Euryarchaeon) and *S. acidocaldarius* (a representative of the TACK superphylum) are given at the top and the bottom, respectively. In both structures, the 16S and 23S rRNAs are co-transcribed as large precursors. The *rm* genes (black boxes) are flanked by external transcribed spacers (ETS) at the 5'- and 3'-end and separated by an internal transcribed spacer (ITS). While in genomes of the TACK phyla, the 5S rRNA is encoded separately, in euryarchaeal genomes, the 5S rRNA gene is located downstream of the 23S rRNA gene. In most Euryarchaea, the ITS1 encodes tRNA<sup>Ala</sup> and in some cases the 3'ETS codes for tRNA<sup>Cys</sup> (dark grey boxes). In *H. volcanii*, a tRNA-like structure (also in dark grey) is encoded immediately upstream of the 5S rRNA. Sequences flanking the 16S and 23S rRNA sequences (marked by light grey boxes) can fold into Bulge-Helix-Bulge RNA (BHB) motifs. (B) Archaeal rRNA processing steps. The rRNA precursor molecule (*H. volcanii* pre-rRNA is depicted as example) is processed by various nucleases. Some activities remain elusive (indicated by a question mark), while others have been characterised (indicated by scissors). The inverted repeats flanking the large rRNA genes fold into BHB motifs processed by EndA. After ligation by the RNA ligase, the circular 16S and 23S precursors are further matured by unidentified activities yielding the mature rRNAs. Based on *in vitro* experiments, it was proposed that the 16S rRNA 3'-end is processed by Nob1. The tRNAs encoded in the ITS and 3'-ETS are processed by RNase Z and RNase P. The tRNA-like structure at the 5'-end of 5S rRNA is processed by RNase Z to release the mature 5S rRNA 5'-end. However, the 5S rRNA 3'-end maturation pathway is still unsolved. Cleavages have also been observed within tRNAs. For references and further details, see section 'Ribosomal RNA biogenesis'.

in *H. marismortui* for both pre-rRNAs (Dennis, Ziesche and Mylvaganam 1998). This splicing-independent maturation of the 23S rRNA plays a minor role, whereas the release of the 16S rRNA by endoribonucleolytic cleavage is an obligatory outcome event in the context of the *H. marismortui* rRNA operon since EndA-processing stems are missing (Dennis, Ziesche and Mylvaganam 1998).

With the exception of EndA activity, archaeal exo- and endoribonucleolytic activities which generate mature 16S and 23S rRNA remain to be discovered. However, more recently, conserved homologues of the eukaryal endonuclease Nob1 responsible for end cleavage in the yeast 18S rRNA maturation pathway were identified in most archaeal genomes (Fatica, Tollervey and Dlakić 2004) (Fig. 1, Table 2). Analysis of *P. horikoshii* Nob1 revealed not only a conserved domain structure but also a similar *in vitro* cleavage specificity against a pre-16S like substrate resulting in the exact mature 3'-end of 16S rRNA (Veith et al. 2012).

**Processing of resulting 5'EST/ITS/3'EST products** The 5'EST/ITS/3'EST ligation by-product, containing tRNAs and generated after the processing by EndA, is further processed by the enzymes involved in tRNA processing such as RNase Z and RNase P at the tRNA or tRNA-like structures present (Chant and Dennis 1986; Tang et al. 2002) (Fig. 6B). Processing by RNase P is a late event occurring subsequent to pre-16S rRNA excision and after the initiation of tRNA 3'-processing (Chant and Dennis 1986; Dennis, Ziesche and Mylvaganam 1998). Moreover, the tRNA included in the 3'-EST is also released by the tRNA-processing enzymes. In *H. salinarum*, the processing of tRNA<sup>Cys</sup> is unordered, but the processing of the tRNA<sup>Cys</sup> 5'-end precedes 5S rRNA 3'-end maturation (Chant and Dennis 1986). Finally, tRNAs of archaeal rRNA operons are sometimes subject to endoribonucleolytic cleavage within their body as suggested in *H. salinarum* and *H. marismortui* (Chant and Dennis 1986; Dennis, Ziesche and Mylvaganam 1998).

**Processing of mature 5S rRNAs** Processing stems are not involved in 5S pre-rRNA release. In *H. salinarum* and *H. marismortui*, 5S pre-rRNA formation is rapidly followed by maturation into 5S rRNA (Chant and Dennis 1986; Dennis, Ziesche and Mylvaganam 1998). Moreover, analysis of 5S pre-rRNA-processing steps in *H. volcanii* cell extracts led to the identification of RNase Z as the endonuclease activity responsible for 5'-end maturation (Hözlze et al. 2008) (Fig. 6B). The region upstream of the 5S pre-rRNA transcript is able to fold into a tRNA-like structure comprising a D-replacement loop, a shortened acceptor stem as well as a shortened T-arm. *In silico* data show that the D- and T-arm-like regions can pair as in the canonical tRNA structure, giving rise to the tRNA elbow conformation recognised by RNase Z (Hözlze et al. 2008; Zhang and Ferré-D'Amaré 2016). *In vitro*, cleavage of the 5S pre-rRNA tRNA-like structure at the 3'-end releases the *H. volcanii* 5S rRNA mature 5'-end as the cognate trailer (Hözlze et al. 2008). Therefore, RNase Z is not only implicated in rRNA processing to generate mature tRNAs encoded within the ITS and ETS regions, but is also able to mature the 5'-end of 5S rRNA. Finally, the recent circRNA-seq approach of Sorek and collaborators identified 5S rRNA circular intermediate forms in *S. solfataricus* (Danan et al. 2012). However, whether the circular form is associated with the 5S rRNA-processing pathway has not been determined yet (Danan et al. 2012).

A recent study links the protein FAU-1, previously described as a RNA-binding protein, to pre-5S and 16S rRNA processing in *P. furiosus* and *T. kodakarensis* (Ikeda et al. 2017). The FAU-1 protein appears also to be important in the quality control of the mature

23S and 16S transcripts (Mackie 2013; Sulthana, Basturea and Deutscher 2016). FAU-1 proteins of both Thermococcales were shown to process a 5S rRNA precursor *in vitro*, preferentially at UA sites (Ikeda et al. 2017). Moreover, deletion of the gene in *T. kodakarensis* impaired cell growth and appeared to alter degradation patterns of the 23S and 16S rRNAs (Ikeda et al. 2017).

**Homing introns within archaeal 16S and 23S pre-rRNAs** In Crenarcheota and Euryarchaeota, 16S pre-rRNAs may contain introns (Nomura, Sako and Uchida 1998; Morinaga, Nomura and Sako 2002; Itoh, Nomura and Sako 2003; Nunoura et al. 2011). This is also the case for the 23S pre-rRNA of Desulfurococcales and Thermoproteales (Itoh, Nomura and Sako 2003). The most numerous and diverse set of introns is found in the genera *Thermoproteus*, *Pyrobaculum* and *Caldivirga*. However, the intron distribution within species is very variable and exhibits a biogeographic rather than phylogenetic distribution (Jay and Inskeep 2015). The rRNA introns, like tRNA introns, contain BHB motifs that ensure their removal by the EndA splicing endoribonuclease during rRNA processing (Lykke-Andersen et al. 1997).

Despite some short and non-coding introns, archaeal pre-rRNA introns are generally rather long (often >700 nt) and mostly contain one or more genes that encode the homing endonuclease characterised by the LAGLI-DADG motif (Morinaga, Nomura and Sako 2002; Stoddard 2005). The sequence-specific homing endonuclease allows mobility of the intron by reintegration at compatible intron-less homing sites. The recognition sequences for archaeal homing endonucleases are pseudopalindromic or asymmetric sites, mostly found in conserved regions of the rRNA (Stoddard 2005). The rRNA introns are stably maintained within the cells and have also been identified as circularised RNAs *e.g.* in *Desulfurococcus mobilis* and *P. aerophilum* (Kjems and Garrett 1988; Burggraf et al. 1993).

**Post-transcriptional rRNA chemical modifications** An increasing number of post-transcriptional chemical modifications are present on rRNAs from Bacteria to Eukarya. Indeed base modifications, such as adenosine methylation, 2'-hydroxyl group methylation and conversion of uridine residues to pseudouridine by base rotation, occur during ribosome biogenesis (for review, see Ferreira-Cerca 2017). These nucleoside chemical modifications are proposed to fine-tune the ribosome biogenesis and functions (Sloan et al. 2017). In hyperthermophiles archaea, the observed number of modifications is close to those reported of eukaryotic rRNAs (Grosjean, de Crécy-Lagard and Marck 2010; Dennis et al. 2015; Sloan et al. 2017). In contrast, a reduced set is observed in halophilic archaea. As an example, in *S. solfataricus*, 67 2'-hydroxyl group ribose methylations were reported compared to only 4 in *H. volcanii* (Grosjean et al. 2008; Dennis et al. 2015). The functional significance of this disparity is not yet understood.

The most ubiquitous modifications are pseudouridine (isomer of uridine) and 2'-O-ribose methylation. These facilitate RNA folding, structure and conformational dynamics (Motorin and Helm 2011). In Archaea and Eukarya, both modifications are governed by two independent mechanisms, an RNA-independent reaction and an RNA-dependent reaction (De Zoysa and Yu 2017). The first involves single protein enzymes called pseudouridine synthases and methyl-transferases that alone recognise the substrate and catalyse the isomerisation of uridine to pseudouridine and the transfer of a methyl group at the 2'O-ribose, respectively. The second are RNA-guided pseudouridylation and methylation by box H/ACA and box C/D RNP complexes, respectively. Briefly the archaeal C/D box sRNA is



characterised by short sequence motifs, two sets of box C and C' (RUGAUGA) and box D and D' (CUGA) sequences, which fold into the widespread K-turn or K-loop structural motifs that is recognised by L7Ae (Fig. 3A) (Klein et al. 2001; Rozhdestvensky et al. 2003; Nolivos, Carpousis and Clouet-d'Orval 2005). In addition, two other proteins, Nop5 and the methyltransferase fibrillar are recruited to generate an active RNP complex (Omer et al. 2002; Bortolin, Bachelierie and Clouet-d'Orval 2003). The archaeal H/ACA box sRNAs is composed of one, two or three H/ACA motifs. Each archaeal H/ACA stem-loop structure contains K-turn or K-loop structural motifs which binds L7Ae. The complete archaeal H/ACA sRNPs contain three other proteins: the pseudouridine synthetase aCbf5 and aNop10 and aGar1 required to reinforce aCBF5 activity (Charpentier, Muller and Branlant 2005; Muller et al. 2008). The RNA component of the complex serves as a guide that base pairs with the RNA substrate and directs the enzyme (aCbf5 or aNop1) to carry out the pseudouridine formation or methyl transfer at a specific site. In Archaea, ribosomal and tRNA nucleosides can be modified by these RNA-guided pathways.

Interestingly, in Archaea, the RNA-dependent reactions involve more than 100 small box guide RNAs that target specific sites of rRNAs (Muller et al. 2008). In addition, these small-guide RNAs have been suggested to serve as 'RNA chaperones' in facilitating rRNA folding (Bachelierie et al. 1995; Steitz and Tycowski 1995; Dennis et al. 2015). However, in a subset of Archaea, the same modification can either be synthesised from protein-based or RNA-guide systems (Clouet-d'Orval, Gaspin and Mougín 2005). This is the case for the dimethylation of the two universally conserved adenosines located at the 3'-end of the 16S rRNA, generally dependent on the protein-based enzyme, RsmA/KsgA/Dim1, but alternatively can use an sRNA-guide system as in *N. equitans* (Seistrup et al. 2017). Moreover, the widespread methylation at position U2552 of the 23S rRNA is mostly performed by the bacterial-like methyltransferase RlmE/RrmJ in most Euryarchaea with the exception of Thermococcales and Crenarchaea, in which the C/D box sR25 is in charge of guiding methylation at this position in the absence of RlmE/RrmJ homologues (Dennis et al. 2015).

#### CRISPR RNA processing

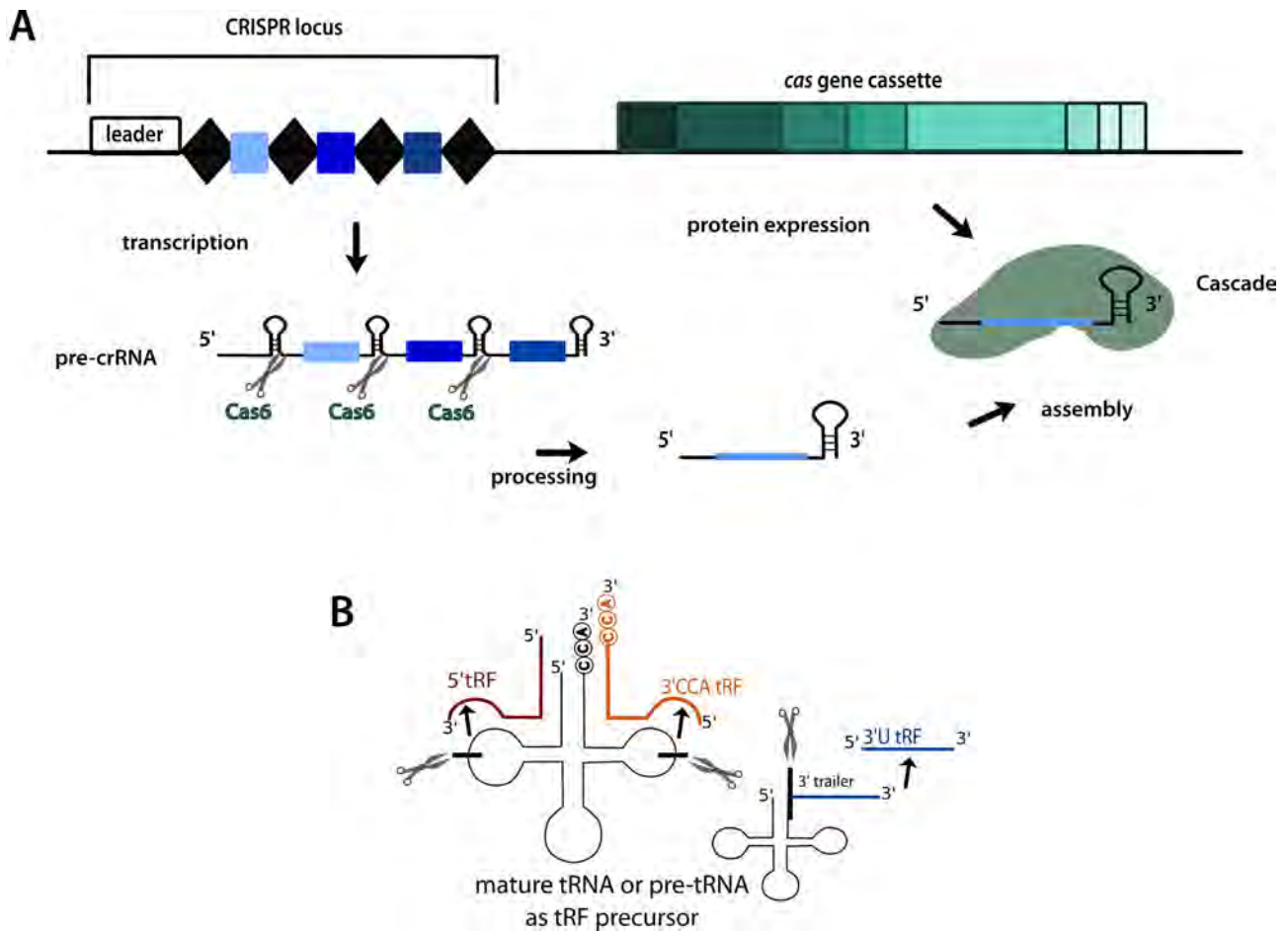
Another type of ncRNA has sparked great interest in recent years. Long thought of as non-coding junk DNA, the long repeat arrays found in most archaeal and bacterial genomes have now been identified as part of a specific and hereditary defence mechanism based on RNPs: the CRISPR-Cas system (for recent reviews, see van der Oost et al. 2014; Makarova et al. 2015; Koonin and Makarova 2017). Key players of this system are the small crRNAs and the Cas proteins (Makarova et al. 2015) (Fig. 7A). CRISPR loci encode recurring sequence elements interspaced by unique spacer sequences matching foreign genetic elements that are transcribed into the crRNA (Bolotin et al. 2005; Mojica et al. 2009). The associated *cas* genes encode the proteins needed for the defence reaction (Makarova et al. 2015), which commences in three functional stages: adaptation, processing, interference (see Rath et al. 2015; Mohanraju et al. 2016 for reviews). During adaptation parts of the nucleic acid of an invading genetic element are integrated into the CRISPR loci as new spacers. In the processing stage, a precursor RNA transcribed from the CRISPR locus is generated and subsequently processed within the repeat sequences separating the individual spacers (Fig. 7A). The resulting crRNA is integrated into the effector complex and used during the third stage to sequence specifically base pair with the re-invading nucleic acid, which is subsequently cleaved and degraded.

Despite their common functional principle, CRISPR-Cas systems are quite diverse. A plethora of individual types (termed I–VI) and an even broader variety of subtypes exhibiting specific differences both in composition and functional aspects exist (for an overview, see van der Oost et al. 2014; Makarova et al. 2015; Koonin and Makarova 2017). Currently, all CRISPR-Cas systems are divided into two classes, the main difference being that class 1 systems have a multiprotein effector complex (Cascade-, Csm-, or Cmr-complex) while the effector complex of class 2 systems consists of a single Cas protein (Cas9, Cas12 and Cas13). Archaea generally possess class 1 CRISPR systems, whereas Bacteria may also contain class 2 systems (Vestergaard, Garrett and Shah 2014; Makarova et al. 2015). Recently, the first, and to date, only example of an archaeal class 2 system (type V) with the single effector Cas protein Cas12a has been described for *Candidatus Methanomethylophilus alvus* (Makarova et al. 2015).

Processing of the pre-crRNA precursor is quite uniform amongst different types of CRISPR-Cas systems and involves Cas proteins (see Charpentier et al. 2015 for a review). Maturation of class 1 system crRNAs depends on the activity of the Cas6 endoribonuclease family (described in section 'Ubiquitous endoribonucleases') (Charpentier et al. 2015), with the exception of the type I-C system which depends on Cas5d (Garside et al. 2012; Nam et al. 2012). Cas6 binds pre-crRNA within the repeat region that often encodes palindromic motifs. If the repeats can stably fold into a hairpin, binding of Cas6 is governed by the structural element as seen in type I-D, I-E and I-F systems (Charpentier et al. 2015). Unstructured repeats with an only weak folding potential as in type I-A, I-B and III rely on sequence elements for Cas6 binding that in turn may transiently stabilise the hairpin-fold (Wang et al. 2011; Shao and Li 2013; Niewoehner, Jinek and Doudna 2014). The mature crRNA contains a single spacer sequence flanked by repeat-derived 5'- and 3'-handles. The 5'-handle is commonly eight nucleotides long whereas the size of the 3'-handle is quite variable as repeat length varies across species (Brouns et al. 2008; Carte et al. 2008; Lange et al. 2013). After Cas6 processing, a second trimming step by a yet unidentified nuclease further reduces the 3'-handle length via a ruler-dependent mechanism in type III systems (Hatoum-Aslan, Maniv and Marraffini 2011; Zhang et al. 2012; Tamulaitis, Venclovas and Siksnys 2017). Similar trimming also occurs for type I-A, I-B and I-D systems in *H. volcanii*, *Clostridium thermocellum*, *M. maripaludis*, *T. tenax* and *Synechocystis* sp. PCC6803 (Richter et al. 2012; Scholz et al. 2013; Plagens et al. 2014; Maier, Dyall-Smith and Marchfelder 2015). Interestingly, systems in which the crRNA seems to be further trimmed after processing by Cas6 possess repeat sequences for which no folding or only weak folding is predicted. Cas6 is only loosely bound to these handles and does not permanently (or not at all) associate with the effector complexes thereby rendering the 3'-handle accessible (Charpentier et al. 2015; Li 2015).

#### Small non-coding box C/D and box H/ACA guide RNA biogenesis

As mentioned above, 2'-O-ribose methylations are abundant in archaeal rRNAs with a prevalence in thermophilic species (Dennis et al. 2015). In contrast, only few pseudouridinylation have been detected in archaeal RNAs so far. As a consequence, C/D box sRNAs outnumber H/ACA sRNAs in archaeal genomes (see Randau 2015 for a review). Indeed *in silico* analysis, immunoprecipitation experiments and more recently high-throughput RNA-seq datasets have revealed a plethora of C/D box sRNA sequences in various archaeal phyla like *Ignicoccus hospitalis*, *M. maripaludis*, *M. kandleri*, *N. equitans*, *P. calidifontis*, *S. acidocaldarius*



**Figure 7.** Processing pathways of CRISPR RNA and tRNA-derived fragment precursors. (A) crRNA maturation pathway. Components of the CRISPR-Cas system are encoded by the CRISPR-locus and the *cas* gene cassette. The CRISPR locus is composed of individual spacer sequences matching previously encountered invaders (blue boxes) interspaced by repeat sequences (black diamonds) and its transcription is governed by the promoter sequence contained in the leader region. The transcribed precursor, the pre-crRNA, may fold into secondary structures within the repeats. These motifs are recognised and processed by Cas6, releasing the mature crRNAs comprising an individual spacer sequence and repeat-derived 5'- and 3'-handles. Certain subtype-specific Cas proteins assemble together with the crRNA into cascade complexes to eliminate the invader DNA. The simplified overview depicted here is based on a type I system, typically found in haloarchaeal species. For references and further details, see section 'CRISPR RNA processing'. (B) Generation of stable tRNA-derived fragments (tRFs). tRFs can be processed from mature tRNAs or precursor tRNAs. Three forms of tRFs have been identified in Archaea (Gebetsberger and Polacek 2013): the 5'-tRF (red) stems from a cut in the D-loop of mature tRNA, whereas 3'-CCA tRFs (orange) are released via processing within the T-loop. In contrast, the 3'-U tRFs (blue) are generated from precursor tRNA processing and comprise the precursor 3'-trailer.

and *T. tenax* (Gaspin et al. 2000; Omer et al. 2000; Phok et al. 2011; Tripp et al. 2017 and references therein).

Comprehensive analyses revealed that only a fraction of archaeal C/D box sRNAs is flanked by an individual promoter while co-transcriptional arrangements with neighbouring genes are predominant (Tripp et al. 2017). But co-transcription with other sequences requires subsequent processing events to generate mature guide RNAs. In the case of C/D box sRNAs co-transcribed with tRNAs as seen for the *T. tenax* tRNA<sup>P<sup>ro</sup></sup> or *N. equitans* tRNA<sup>Val</sup>, the release of C/D box sRNAs occurs by processing via the tRNA-processing enzymes (Randau 2012; Tripp et al. 2017). Nevertheless, the general process releasing mature C/D box RNAs is still not understood. Although relaxed BHB motifs were predicted for some archaeal C/D box sRNAs, processing in *S. acidocaldarius* appeared independent of flanking sequences (Berkemer et al. 2015; Tripp et al. 2017). Maturation of archaeal small-guide RNAs could also involve other archaeal endo- and exoribonucleases such as the archaeal exosome and/or ribonucleases of the  $\beta$ -CASP family (see section 'Archaeal ribonuclease families'), reminiscent of the

eukaryotic guide sRNA maturation pathway (Massenet, Bertrand and Verheggen 2017). Finally, the most intriguing feature of archaeal small-guide RNA biogenesis is the discovery of significant amounts of circularised C/D box sRNAs in *M. kandleri*, *N. equitans*, *P. furiosus* and *S. solfataricus* (Starostina et al. 2004; Danan et al. 2012; Randau 2012; Su, Tripp and Randau 2013). It is proposed that circularisation stabilises RNAs in high-temperature conditions (Starostina et al. 2004; Danan et al. 2012). In *P. furiosus*, the circular C/D box RNAs are shown to be integrated into functional C/D box RNP complexes (Starostina et al. 2004). C/D box sRNAs may also be embedded in tRNA intron sequences as seen in the intron of tRNA<sup>TP</sup> which is conserved across most Euryarchaeota (Clouet d'Orval et al. 2001; Weisel, Wagner and Klug 2010). The C/D box sRNA is released by EndA and circularised by the RtcB ligase as seen in *H. volcanii* and *H. salinarum* (Clouet d'Orval et al. 2001; Weisel, Wagner and Klug 2010). However, in the majority of cases, C/D box sRNAs are not intron-encoded and it remains to be determined how circularisation occurs.

### Emerging classes of ncRNAs

**Transfer RNA-derived fragments** Fragments derived from tRNA sequences (tRNA-derived fragment or tRF), initially thought to be intermediates of tRNA degradation, are now identified as functional small ncRNAs (for reviews, see Keam and Hutvagner 2015; Kumar, Kuscu and Dutta 2016) (Fig. 7B). Their function was extensively studied in Eukarya but they have also been described in Bacteria and Archaea (Bernick et al. 2012; Gebetsberger et al. 2012; Heyer et al. 2012; Kumar et al. 2014).

Transfer RNA-derived fragments can be generated from tRNA precursors or mature tRNAs. The amount of a given tRNA-derived fragment does not correlate with the amount of the tRNA progenitor implying that tRNA biogenesis and tRNA-derived fragment-processing pathways are independently regulated (Gebetsberger and Polacek 2013). The nature of the processing and regulatory factors generating tRNA-derived fragments is not known (Keam and Hutvagner 2015). Upon endoribonucleolytic cleavage, different classes of tRNA-derived fragments can emerge depending on the processing site and the nature of the precursor (Gebetsberger and Polacek 2013). Archaeal tRNA-derived fragments from several *Pyrobaculum* species, *N. equitans* and *H. volcanii* were found to encompass 5'-tRFs (derived from a cleavage within the D-arm of mature tRNA), 3'-CCA tRFs (derived from a cleavage within the T-loop of mature tRNA) and 3'-U tRFs (remainder of 3'-trailer of pre-tRNA) (Bernick et al. 2012; Gebetsberger et al. 2012; Heyer et al. 2012; Randau 2012; Gebetsberger and Polacek 2013). A detailed analysis of tRFs is only available for *H. volcanii* (Gebetsberger et al. 2012). Transfer RNA-derived fragments correspond to 11 of 51 tRNAs of *H. volcanii* with a majority derived from tRNA 5'-ends in response to different stresses, as also seen in eukaryotes (Gebetsberger et al. 2012; Heyer et al. 2012). In an in-depth functional analysis of the alkaline-stress induced tRNA<sup>Val</sup> 5'-tRF (tRF[Val]) (Gebetsberger et al. 2012, 2017), tRF[Val] was shown to co-purify with ribosomes and bind the small ribosomal subunits *in vitro*, inhibiting translation (Gebetsberger et al. 2012). tRF[Val] was also recently demonstrated to bind the ribosome in proximity to the mRNA channel and to affect the overall protein biosynthesis in archaeal, bacterial and eukaryal ribosomes (Gebetsberger et al. 2017). Similar translational control by tRFs occurs in eukaryotic HeLa cells (Sobala and Hutvagner 2013). tRFs are therefore potent regulators in all domains of life but their full gene regulatory potential has yet to be unravelled.

**Circular RNAs** Another emerging class of small ncRNAs with an interesting biogenesis pathway are circular RNAs. They are rare but present in all domains of life. Most archaeal circular RNAs described are the result of the intron ligation step of the splicing reaction. Some are intermediates of rRNA-processing events, such as the circular precursors of 16S and 23S rRNAs (Tang et al. 2002; Danan et al. 2012) (section 'Ribosomal RNA biogenesis'). But those transcripts are further processed and exert their biological function in the form of a linear end product. In contrast stable circular RNAs are generated by excision of the rRNA introns coding for homing endonucleases (Kjems and Garrett 1988; Burggraf et al. 1993) (see section 'Ribosomal RNA biogenesis'). In addition, tRNA precursor splicing events also produce stable circular RNA entities observed for the pre-tRNA<sup>TP</sup> of Euryarchaeota like *H. volcanii* (see section 'Ribosomal RNA biogenesis').

A more systematic search for circular RNAs by the novel sequencing method, circ-RNA-seq, identified a plethora of circular RNAs in *S. solfataricus*, with a great diversity of sequences corresponding to rRNA intermediates, tRNA introns and C/D and H/ACA box RNAs, 7S SRP RNA and RNase P and open read-

ing frame-derived circular RNAs from 37 different genes (Danan et al. 2012). In addition, a circular form of the 7S SRP RNA has also recently been described for *T. tenax* where circularisation restores the permuted sequence by processing via the splicing machinery (Plagens et al. 2015). Nevertheless, the molecular mechanisms that generate circular RNAs await discovery and the biological function and significance of these circular forms need to be experimentally tested. Circular RNAs have been proposed to represent either processing intermediates or to be a means of stabilisation and protection (Danan et al. 2012).

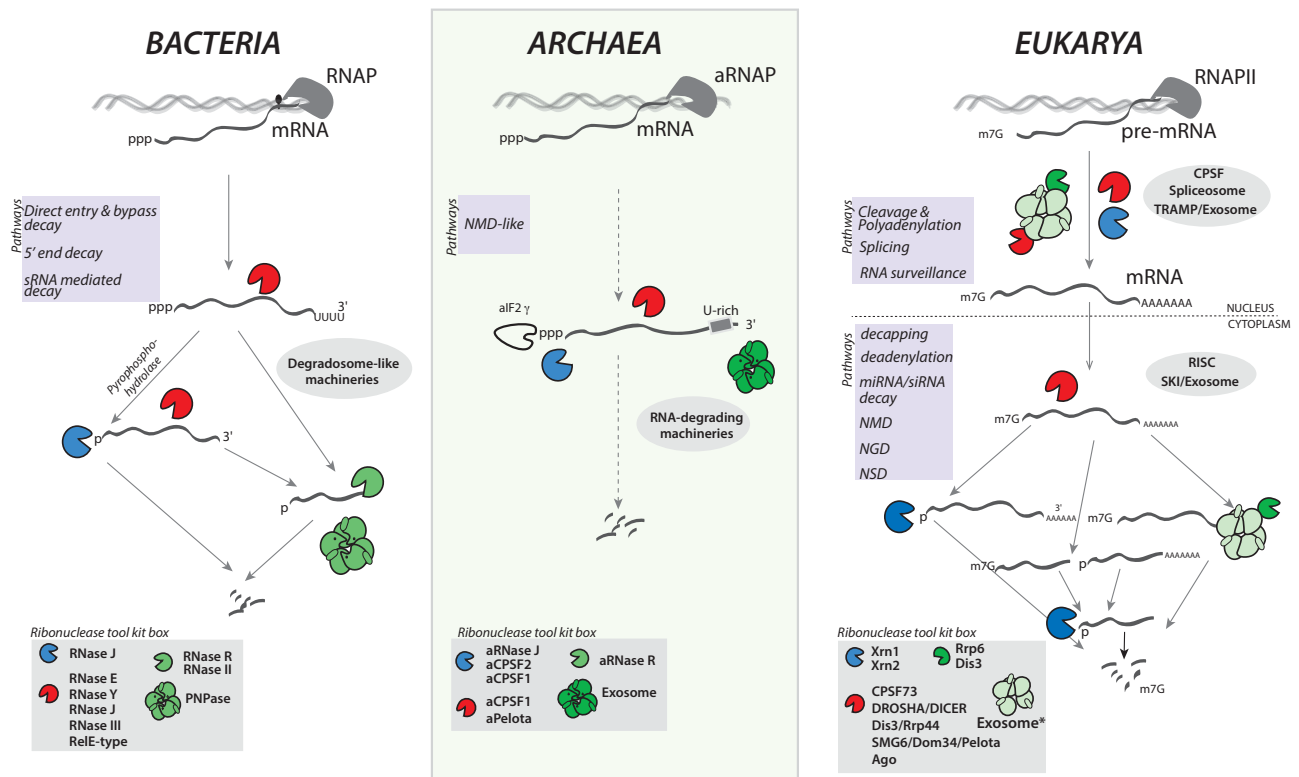
### Messenger RNA turnover in Archaea—lessons from Eukarya and Bacteria

For several decades, the post-transcriptional mechanisms by which the level, the stability and the quality of mRNAs are finely regulated have been an area of intense research interest and have been reported in comprehensive reviews (see chapters in Stoecklin and Mühlemann 2013). In essence, mRNA decay pathways are critical in defining timescales for cellular events by allowing transient adaptation relative to a growth rate and rapid responses to external stimuli to meet cellular requirements for homeostasis. The stability of an mRNA is determined by cell-specific set of ribonucleases, the sequence and structural features of the mRNA molecule itself and related proteins or regulatory RNAs. In this regard, small ncRNAs, such as sRNA, miRNA and siRNA, are critical and often mediate co-degradation, sequestration or stabilisation of target mRNAs in a regulatory hierarchy. In Archaea, in contrast to Bacteria and Eukarya, pathways controlling mRNA stability and quality are still poorly documented, even if several ribonuclease families carrying endo-, 5'-to-3' and 3'-to-5' exo-ribonucleolytic activities have been identified (Fig. 1) (Clouet-d'Orval et al. 2015) (see section 'Archaeal ribonuclease families').

### Overview of mRNA decay pathways in Bacteria and Eukarya

In Bacteria, most mRNAs are polycistronic with concurrent transcription and translation. A key event in their degradation is an initial cleavage by an endonuclease which is followed by the action of exoribonucleases (Fig. 8). This direct entry pathway controls the turnover of most mRNAs and recruits the hydrolytic endoribonucleases, RNase E in  $\gamma$ -proteobacteria and RNase Y in most Firmicutes as well as in some  $\delta$ - and  $\epsilon$ -proteobacteria (Laalami, Zig and Putzer 2014). Remarkably and despite their sequence disparity, both enzymes serve as assembly platforms for other enzymes, such as RNA helicases and exoribonucleases and are membrane bound (Bandyra et al. 2013; Hui, Foley and Belasco 2014; Redder 2016). In addition, 5'-end-dependent access pathways which sense the nature of the mRNA 5'-terminus involve the recruitment of pyrophosphohydrolase enzymes of the Nudix family which trigger deprotection or 'decapping' (as a diphosphate is removed) and leave a 5'-end monophosphate mRNA which is a subsequent substrate for 5'-to-3' exoribonucleases, such as RNase J in most Bacteria with the exception of the  $\gamma$ -proteobacteria (Bandyra et al. 2013; Hui, Foley and Belasco 2014). Their association in multienzyme complexes or degradosome-like machineries and their subcellular localisation allow precise control of the mode of action of these potentially highly destructive enzymes (Kaberdin, Singh and Lin-Chao 2011; Deutscher 2015; Khemicci et al. 2015) (Fig. 8). In the case of mRNA lacking a termination codon, the stalled ribosomes are rescued by a mechanism known as trans-translation that involves a specialised bacterial RNA (tmRNA) with features of both tRNA and





**Figure 8.** Schematic of the major mRNA quality control and decay pathways of the three domains of life. mRNA or precursor mRNA (pre-mRNA) (dark grey wavy line) is transcribed from genomic DNA (double light grey wavy lines) by the RNA polymerase (RNAP for Bacteria, aRNAP for Archaea and RNAPII (RNA polymerase II) for Eukarya). The 5' cap structure (m7G) and poly-A tail (AAAAAA) of eukaryal mRNA are shown. The endo-, 5'-to-3' exo- and 3'-to-5' exoribonucleases are represented as red, blue and green 'pacman', respectively. In addition, the 3'-to-5' exo machineries (archaeal and eukaryal exosome and bacterial PNPase) shown as green complexes are formed of multiple subunits. The ribonuclease tool kit boxes are specified at the bottom of each panel. Some of these machineries in bacterial cell and eukaryotic cytoplasm have been shown to be involved in the ribosome-rescue pathway. For clarity, mRNAs in polysomes are not represented. The mRNA processing and decay pathways (framed within each panel) are specified: for bacteria, it includes the direct entry and bypass decays, 5'-end decay and sRNA mediated decay; for archaea, a non-sense mediated decay (NMD) has been proposed; for Eukarya, in the nucleus, it includes cleavage and polyadenylation, splicing and RNA surveillance whereas in the cytoplasm, it includes in addition to the classical decapping and deadenylation pathways, the micro-RNA and silencing RNA (miRNA/siRNA) decays, nonsense-mediated decay (NMD), no-go decay (NGD) and nonstop decay (NSD).

mRNA and an escort protein (SmpB) which has been shown to associate with bacterial RNase R *in vitro* (Inada 2013; Hui, Foley and Belasco 2014). More recently, in Firmicutes, a family of ribosome-dependent endoribonuclease Rae1/YacP has been proposed to play a role in a translation-related process with overlapping similarities with classical toxin-antitoxin systems involved in mediating bacterial stress responses (Leroy et al. 2017).

In contrast to bacterial mRNAs, eukaryal mRNAs are often monocistronic, must transit the nuclear membrane before being translated and have longer half-lives. In this setting, the maturation of eukaryal RNAPII-mRNA transcripts requires further intricate processes, such as 5'-capping, 3'-addition of extensive polyadenylation tracts, intron removal and chemical modifications. All of these sophisticated steps are orchestrated and controlled by RNA surveillance and decay pathways which recruit a number of multiprotein scaffolds (Chlebowski et al. 2013; Nagarajan et al. 2013; Porrua and Libri 2013; Schweingruber et al. 2013; Inada 2017) (Fig. 8). Nuclear RNA (including cryptic stable transcript and snoRNA) control quality pathways and cytoplasmic mRNA decay pathways rely on the exosome/Rrp6/Dis3 machineries with dual, endo- and 3'-to-5' exoribonucleolytic activities and on the ribonucleases from the XRN family with 5'-to-3' exoribonucleolytic activity (Chlebowski et al. 2013; Nagarajan

et al. 2013). These are associated with specific sets of co-factors to identify and degrade deadenylated and decapped transcripts (Stoecklin and Mühlemann 2013) (Fig. 8). The complexes are also involved in co-translational quality control mechanisms allowing elimination of a variety of aberrant mRNAs (errors in transcription, processing or induced by stresses). In addition, there are numerous quality control systems in place to differentiate between paused ribosomes and terminally stalled ribosomes that need to be rescued. In Eukarya, three cytoplasmic mRNA quality control systems monitor mRNAs for translational errors: nonsense-mediated decay (NMD), non-stop decay (NSD) and no-go decay (NGD) pathways (Inada 2013, 2017) (Fig. 8).

#### Messenger RNA half-lives

Earlier reviews specifically described half-lives of several archaeal mRNAs in different conditions (Evguenieva-Hackenberg and Klug 2009, 2011). In general, it appeared that some archaeal mRNAs have significantly longer average half-lives than most mRNAs in Bacteria. Briefly, half-lives of several transcripts varied from 7 to 57 min in *M. vannielii* (Hennigan and Reeve 1994), from 2 to 20 min in *S. solfataricus* and *S. acidocaldarius* (Bini et al. 2002; Andersson et al. 2006) and from 4 to 80 min in *H. mediterranei* (Jäger et al. 2002). More recently, a first global picture of differential expression and half-lives for *M. acetivorans* provided

evidence of drastic genome-wide shifts in RNA half-lives modulated by carbon source and led to the conclusion that more than half of metabolism genes are controlled by degradation (Peterson *et al.* 2016). Nevertheless, even if a number of cases of mRNA cleavage have been documented in global analysis of mRNA decay at single-gene resolution, available data are still fragmentary. In *M. jannaschii*, evidence of directional 3'-to-5' RNA degradation comes from the accumulation of transcript 5'-ends from upstream of protein-coding genes (Zhang and Olsen 2009). However, the core subunits of the archaeal exosome in Methanococcales are missing (Fig. 1). This is also the case in the halophilic euryarchaeon *Halobacterium salinarum* NRC-1 which instead possesses a bacterial-like RNase R (Fig. 1) with transcript stabilities ranging from 5 min to more than 18 min (Hundt *et al.* 2007). However, to date no specific mRNA-endoribonucleolytic cleavage signal that could initiate RNA decay has been identified even though the aCPSF1 endoribonuclease is absolutely conserved in complete annotated archaeal genomes (Figs 1 and 8) (see Clouet-d'Orval *et al.* 2015 for a review).

### Specific mRNA features

Regardless of the similarities between archaeal and eukaryal transcription machineries, archaeal mRNAs show bacterial features with mostly no introns, no 5'-methyl caps and no long 3'-poly A tails and potentially polycistronic structures (Fig. 8). Nonetheless, *in vitro* and *in vivo* studies using the crenarchaeota *S. solfataricus* showed evidence for cap-like structure in which the archaeal translation initiation factor aIF2- $\gamma$  subunit binds to the RNA 5'-triphosphorylated ends and protects transcripts from a 5'-end-dependent degradation pathway (Hasenöhrl *et al.* 2008; Arkhipova *et al.* 2015). In this instance, the aIF2- $\gamma$  subunit resembles the eukaryotic cap complex strongly suggesting that a directional 5'-to-3' degradation pathway may exist in archaeal cells. This is supported by the ubiquitous occurrence of a 5'-to-3' exoribonucleolytic activity carried by members of the archaeal  $\beta$ -CASP family (Fig. 1). However, in Archaea, while 5'-to-3' exoribonucleases, namely aRNase J and aCPSF2, have been identified to have a preference for monophosphorylated RNA substrates in Euryarchaea and Crenarchaea (Clouet-d'Orval *et al.* 2010; Hasenöhrl, Konrat and Bläsi 2011), the molecular mechanism of 5'-to-3' mRNA decay remains unknown and an enzyme capable of converting the 5'-triphosphorylated ends of their mRNAs to monophosphates remains to be discovered. In eukaryotic cells, at least two mRNA decapping pyrophosphohydrolases of the Nudix superfamily, Dcp2 and Nudt16, remove the 5'-cap structure (Arribas-Layton *et al.* 2013). In Bacteria, the triphosphorylated 5'-termini are converted to monophosphorylated ends by the Nudix superfamily RppH (Celesnik, Deana and Belasco 2007). Although Nudix hydrolases are widely distributed among the three domains of life, archaeal Nudix hydrolases with 'decapping' activity have not yet been identified. Two of these proteins from *S. solfataricus* were cloned and expressed in heterologous system but no pyrophosphohydrolase activity could be detected in *in vitro* assays (Hasenöhrl, Konrat and Bläsi 2011).

It has been shown that large 5'-UTRs contribute to the mRNA stability of methanogenesis genes in *M. mazei*, (Cao *et al.* 2014). This is in contrast to other Euryarchaeota and Crenarchaeota in which leaderless transcripts are dominant (Babski *et al.* 2016), and to eukaryal organisms in which mRNA stability is mainly regulated by elements embedded in the transcript 3'-UTR (see Geissler and Grimson 2016 for a review). Finally, in many cases described above, mRNAs from genes known or predicted to be transcribed in operons exhibited similar half-lives. This was confirmed in a methanogenic archaeon by recent findings allow-

ing the genome-wide mapping of the RNA-processing site (PSSs) obtained by the specific sequencing of the 5'-monophosphate transcripts that arise from processing events (Qi *et al.* 2017). This approach detected overrepresented 5'-untranslated and intergenic regions in the polycistronic operons encoding ribosomal proteins signifying that mechanisms of ribosomal protein synthesis and stoichiometry may primarily consist of mRNA-processing-mediated post-transcriptional regulation (Qi *et al.* 2017).

Furthermore, archaeal regulatory sRNAs are undoubtedly involved in many biological processes, including metabolic regulation, adaptation to extreme conditions and stress response presumably by altering translation and stability of specific target mRNAs as reported in Bacteria and Eukarya (Marchfelder *et al.* 2012; Babski *et al.* 2014). In the last decade, three classes of small ncRNAs, namely intergenic, antisense and sense RNAs, that can target either the 5'-UTRs or the 3'-UTRs of their respective target mRNAs have been detected in several Crenarchaeota and Euryarchaeota (Babski *et al.* 2014). These include *A. fulgidus* (Tang *et al.* 2002), *S. solfataricus* (Tang *et al.* 2005), Halophiles (Soppa *et al.* 2009; Straub *et al.* 2009; Fischer *et al.* 2011), *M. psychrophilus* (Li *et al.* 2015), *T. kodakarensis* (Jäger *et al.* 2014), *M. mazei* (Jäger *et al.* 2009) and *P. abyssi* (Phok *et al.* 2011; Toffano-Nioche *et al.* 2013). Some are differentially expressed with growth conditions (Straub *et al.* 2009; Jaschinski *et al.* 2014). Nevertheless, a mystery remains regarding mRNA decay directly mediated by sRNA since no RNase III-like double-stranded endoribonuclease activity has been detected so far in archaeal cells. It is conceivable that, in Archaea, the EndA activity cleaving within BHB motifs overcomes the lack of double strand-specific RNase III-like activity.

### Role of SmAPs in mRNA stability?

The SmAPs (described in section 'Archaeal Sm-like proteins') could play important physiological functions in archaeal mRNA stability. Indeed, while in Eukaryotes, Sm proteins serve as molecular scaffold for RNP complex assembly, as exemplified by their role in rRNA processing by small nucleolar RNPs, in Bacteria, Hfq acts as an RNA chaperone, facilitating base-pairing interactions between small regulatory ncRNAs (sRNAs) and their mRNA targets, and directly influencing the structures of some RNAs (Mura *et al.* 2013). However, the function of SmAP proteins in archaeal cells is currently mostly unknown but first details have been uncovered in *H. volcanii* and *S. solfataricus*. The *smAP1* gene is very often encoded immediately upstream or even overlapping the gene for 50S ribosomal protein L37e, suggesting a conserved role in processing or stabilisation of rRNAs (Mura *et al.* 2013). For the *SmAP2* paralogue, in most Crenarchaeota and Thaumarchaeota, the gene lies directly upstream of, and on the same strand as, a methionine adenosyl transferase gene that could be involved in DNA or RNA methylation. Additionally, a role in tRNA processing or biogenesis was suggested since *A. fulgidus* Afu-SmAP1 and Afu-SmAP2 (Törö *et al.* 2001) and Sso-SmAP1 and Sso-SmAP2 (Märtens *et al.* 2015) associate with the RNA molecule of RNase P. Moreover, the interacting proteins and RNA substrates of Hvo-SmAP1 (Fischer *et al.* 2010) and Sso-SmAP1 and Sso-SmAP2 (Märtens *et al.* 2015, 2017) identified by co-immunoprecipitation experiments hint at possible SmAP biological functions. The candidate binding partners belong to similar functional classes (rRNA and tRNA modification and processing, RNA decay and translation) like the partners identified for bacterial Hfq and eukaryotic Sm-like proteins (Fischer *et al.* 2010; Märtens *et al.* 2017). While most interactions between SmAPs and candidate partners remain to be confirmed, a direct interaction between the Sso-SmAP1 and



Sso-SmAP2 and the exosomal subunit aDnaG, required for efficient polyadenylation of RNA substrates and that might be required for membrane localisation of the Sso-aExosome, was confirmed by affinity purification (Märtens et al. 2017). A link between the exosome and Sm-like proteins also exists in eukaryotes (Luz et al. 2010). While intron-containing RNAs are enriched with Sso-SmAP1, suggesting possible functions in tRNA/rRNA processing, Sso-SmAP2 seems to bind predominantly to mRNAs and could therefore be involved in mRNA decay or stability (Märtens et al. 2015). Finally, an *H. volcanii* strain deleted for the Sm1 motif of Hvo-SmAP1 ( $\Delta sm1$ ) that does not influence L37e expression or affect cell growth was shown to exhibit enhanced swarming (Maier et al. 2015). This observation is corroborated by differential RNA-seq gene expression analysis. Transcripts encoding proteins required for motility are upregulated in the  $\Delta sm1$  mutant (Maier et al. 2015). Whether Hvo-SmAP1 is directly involved in cell motility regulation or affects a downstream regulator like an sRNA remains to be determined (Maier et al. 2015).

#### Messenger RNA quality control pathway?

Controlling mRNA quality is another important challenge for the cell to prevent the synthesis of harmful proteins derived from aberrant mRNAs. As mentioned earlier, in Eukarya, three different decay pathways stimulate the degradation of aberrant mRNAs and the dissociation of ribosomes by recruiting mRNA surveillance factors at the stalled ribosomes (Fig. 8). Recently, it emerged that the structural basis of these mRNA surveillance pathways could exist in Archaea. This is supported by the identification of an ubiquitous archaeal Pelota/Dom34-like protein (aPelota) (Ragan et al. 1996) (described in section 'Ubiquitous endoribonucleases', Figs. 1 and 2C). Indeed, aPelota, in a complex with the archaeal translation elongator factor EF1 $\alpha$ , could mimic the tRNA•EF-Tu complex bound to the A site of the ribosome similarly to proposed ribosome disassembly pathways in Eukarya (Kobayashi et al. 2010; Saito et al. 2010; Becker et al. 2012). A structural basis and a universal mechanism for archaeal and eukaryal ribosome recycling were proposed from cryo-electron microscopy reconstructions of eukaryotic and archaeal ribosome recycling complexes containing the ATPase ABCE1 and the Pelota termination factor paralogue (Becker et al. 2012). This is in agreement with early findings showing that the molecular mimicry of tRNA in the distorted 'A/T state' conformation by Pelota-like proteins enables the complex to efficiently detect and enter the empty A site of the stalled ribosome which involves the interaction of the positively charged patch of Pelota-like domain A with the decoding centre (Kobayashi et al. 2010; Franckenberg, Becker and Beckmann 2012). Future work should determine if translation arrests induce endoribonucleolytic cleavages of the aberrant mRNA by aPelota bound to the A site of the ribosome and recruit other degradation factors, such as the exosome or  $\beta$ -CASP enzymes with the 3'-to-5' and 5'-to-3' exoribonucleolytic activities, respectively, to eliminate the aberrant mRNA fragments. A first clue towards the biological relevance of the 5'-to-3'  $\beta$ -CASP ribonuclease of the aCPSF2 group comes from transcriptome analyses of wild-type and aCPSF2-deleted *S. acidocaldarius* strains which reveal a global effect with differential transcript abundance for 560 genes (Märtens et al. 2013). Other indications may come from genomic context analysis where the gene encoding aCPSF1 in the majority of archaeal genomes is adjacent to the gene encoding a proteasome subunit suggesting interplay between RNA and protein-degrading machineries (Koonin, Wolf and Aravind 2001). In addition, a study in *H. volcanii* identified Hvo-aCPSF1 as intimately linked to partners

of the formation of thiolated tRNA (Lys)UUU which includes homologues of ubiquitin-proteasome (Chavarría et al. 2014).

In conclusion, many open questions remain and coherent pictures of mRNA decay and RNA quality control pathways are still missing in Archaea (Fig. 8). Future work using improved genetic tools for established model organisms coupled with specific high-throughput RNA sequencing technologies, global measurement of mRNA half-lives and comparative genomics will certainly give an input to the field. The expected outcome will definitely illuminate the roles of the conserved 3'-to-5' exosome machineries and  $\beta$ -CASP 5'-to-3' exo/endoribonucleases in specific biological mRNA decay mechanisms.

## CONCLUDING REMARKS

Despite the wealth of information gathered to date regarding RNA maturation, processing, function and life cycles within archaeal cells, many open questions remain, and RNA maturation and degradation pathways in Archaea are for the most part still a mystery. Several key steps of rRNA maturation, for instance, are still far from being understood, and a comprehensive overview of common principles and species-specific peculiarities is about to emerge. Future in-depth landscape exploration of these pathways in different genomic contexts will certainly provide general principles of ribosomal biogenesis and functional assembly in Archaea and will highlight processing and assembly steps shared between Archaea, Bacteria and Eukarya. In general, many ribonucleases and other RNA-modifying enzymes are still expected to be discovered. Challenges will be to identify the key events and to understand how the actors of RNA metabolism are specifically recruited in pathways controlling archaeal RNA life cycle.

Furthermore, it should not be forgotten that archaeal species have very diverse ecological niches, and may have evolved specific and biochemically unique pathways, in addition to the conserved ones, to maintain the integrity of their RNAs in extreme conditions. Finally, deciphering RNA metabolism pathways at work in the recently discovered archaeal lineages will enlarge evolutionary perspectives in understanding the players acting in RNA biology throughout the three distinct phylogenetic domains, especially in the context of the recent lively debate that oppose the standard three-domain (Woese, Kandler and Wheelis 1990) to the two-domain (Eocyte) topology models of the tree of life (Koonin 2015; Raymann, Brochier-Armanet and Grubaldo 2015; Adam et al. 2017; Da Cunha et al. 2017; Eme et al. 2017; Van der Gulik, Hoff and Speijer 2017)

## ACKNOWLEDGEMENT

We are indebted to Mike Chandler for critical reading of the manuscript and to Christine Gaspin for helpful discussions.

## FUNDING

This work was supported by the Centre National pour la Recherche Scientifique (CNRS); the Idex-emergence program of the 'Université de Toulouse' (UPS) [RNArch to BCO]; the French 'Agence Nationale pour la Recherche' [ANR-16-CE12-0016-01 to BCO] and by funding of the DFG [MA1538/21-1 to AM].

**Conflict of interest.** None declared.

## REFERENCES

- Achsel T, Stark H, Lührmann R. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc Natl Acad Sci USA* 2001;**98**:3685–9.
- Acinas SG, Marcelino LA, Klepac-Ceraj V et al. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrm operons. *J Bacteriol* 2004;**186**:2629–35.
- Adam PS, Borrel G, Brochier-Armanet C et al. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* 2017;**11**:2407–25.
- Aebi M, Kirchner G, Chen JY et al. Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. *J Biol Chem* 1990;**265**:16216–20.
- Agris PF. The importance of being modified: an unrealized code to RNA structure and function. *RNA* 2015;**21**:552–4.
- Agris PF, Eruysal ER, Narendran A et al. Celebrating wobble decoding: half a century and still much is new. *RNA Biol* 2017;**1**–17, DOI: 10.1080/15476286.2017.1356562.
- Albers S-V, Forterre P, Prangishvili D et al. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat Rev Microbiol* 2013;**11**:713–9.
- Altman S. Ribonuclease P. *Philos T Roy Soc B* 2011;**366**:2936–41.
- Amero CD, Boomershine WP, Xu Y et al. Solution structure of *Pyrococcus furiosus* RPP21, a component of the archaeal RNase P holoenzyme, and interactions with its RPP29 protein partner. *Biochemistry* 2008;**47**:11704–10.
- Andersson AF, Lundgren M, Eriksson S et al. Global analysis of mRNA stability in the archaeon *Sulfolobus*. *Genome Biol* 2006;**7**:R99.
- Aravind L. An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol (Gedruckt)* 1999;**1**:69–91.
- Arkhipova V, Stolboushkina E, Kravchenko O et al. Binding of the 5'-triphosphate end of mRNA to the  $\gamma$ -subunit of translation initiation factor 2 of the Crenarchaeon *Sulfolobus solfataricus*. *J Mol Biol* 2015;**427**:3086–95.
- Arraiano CM, Mauxion F, Viegas SC et al. Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA. *Biochim Biophys Acta* 2013;**1829**:491–513.
- Arribas-Layton M, Wu D, Lykke-Andersen J et al. Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim Biophys Acta* 2013;**1829**:580–9.
- Atkinson GC, Baldauf SL, Haurlyuk V. Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol Biol* 2008;**8**:290.
- Babski J, Haas KA, Näther-Schindler D et al. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* 2016;**17**:629.
- Babski J, Maier L-K, Heyer R et al. Small regulatory RNAs in Archaea. *RNA Biol* 2014;**11**:484–93.
- Bachelier JP, Michot B, Nicoloso M et al. Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci* 1995;**20**:261–4.
- Baker DL, Youssef OA, Chastkofsky MIR et al. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Gene Dev* 2005;**19**:1238–48.
- Ban N, Nissen P, Hansen J et al. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 2000;**289**:905–20.
- Bandyra KJ, Bouvier M, Carpousis AJ et al. The social fabric of the RNA degradosome. *Biochim Biophys Acta* 2013;**1829**:514–22.
- Becker T, Franckenberg S, Wickles S et al. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 2012;**482**:501–6.
- Berkemer SJ, Höner ZU, Siederdisen C et al. Processed small rnas in archaea and BHB elements. *Genomics Comput Biol* 2015;**1**:18.
- Bernick DL, Dennis PP, Lui LM et al. Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four pyrobaculum species. *Front Microbiol* 2012;**3**:231.
- Bhuiya MW, Suryadi J, Zhou Z et al. Structure of the Aeropyrum pernix L7Ae multifunctional protein and insight into its extreme thermostability. *Acta Crystallogr F Struct Biol Cryst Commun* 2013;**69**:979–88.
- Bini E, Dikshit V, Dirksen K et al. Stability of mRNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *RNA* 2002;**8**:1129–36.
- Bolhuis H, Palm P, Wende A et al. The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* 2006;**7**:169.
- Bolotin A, Quinquis B, Sorokin A et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, Engl)* 2005;**151**:2551–61.
- Boomershine W.P., McElroy C.A., Tsai H.-Y., Wilson R.C., Gopalan V., Foster M.P. Structure of Mth11/Mth Rpp29, an essential protein subunit of archaeal and eukaryotic RNase P. *Proceedings of the National Academy of Sciences of the United States of America* 2003;**100**:15398–403.
- Bortolin M-L, Bachelier J-P, Clouet-d'Orval B. In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA<sup>Trp</sup>. *Nucleic Acids Res* 2003;**31**:6524–35.
- Brochier C, Gribaldo S, Zivanovic Y et al. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* 2005;**6**:R42.
- Brochier-Armanet C, Forterre P, Gribaldo S. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* 2011;**14**:274–81.
- Brouns SJJ, Jore MM, Lundgren M et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008;**321**:960–4.
- Bubeck D, Reijns MAM, Graham SC et al. PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* 2011;**39**:3652–66.
- Bult CJ, White O, Olsen GJ et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 1996;**273**:1058–73.
- Burggraf S, Larsen N, Woese CR et al. An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci USA* 1993;**90**:2547–50.
- Büttner K, Wenig K, Hopfner K-P. Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol Cell* 2005;**20**:461–71.
- Calvin K, Li H. RNA-splicing endonuclease structure and function. *Cell Mol Life Sci* 2008;**65**:1176–85.
- Cantara WA, Crain PF, Rozenski J et al. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* 2011;**39**:D195–201.
- Cao Y, Li J, Jiang N et al. Mechanism for stabilizing mRNAs involved in methanol-dependent methanogenesis of cold-adaptive *Methanosarcina mazei* zm-15. *Appl Environ Microb* 2014;**80**:1291–8.
- Carte J, Wang R, Li H et al. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Gene Dev* 2008;**22**:3489–96.

- Celesnik H, Deana A, Belasco JG. Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal. *Mol Cell* 2007;27:79–90.
- Chan PP, Cozen AE, Lowe TM. Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol* 2011;12:R38.
- Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 2009;37:D93–7.
- Chant J, Dennis P. Archaeobacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum*. *EMBO J* 1986;5:1091–7.
- Chapados BR, Chai Q, Hosfield DJ et al. Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication. *J Mol Biol* 2001;307:541–56.
- Charpentier B, Muller S, Branlant C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* 2005;33:3133–44.
- Charpentier E, Richter H, van der Oost J et al. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol Rev* 2015;39:428–41.
- Charpentier E, van der Oost J, White MF. crRNA Biogenesis. In: Barrangou R, van der Oost J (eds.) *CRISPR-Cas Systems*, Berlin, Heidelberg: Springer, 2013, 115–44.
- Charron C, Manival X, Cléry A et al. The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol* 2004a;342:757–73.
- Charron C., Manival X., Charpentier B., Branlant C., Aubry A. Purification, crystallization and preliminary X-ray diffraction data of L7Ae sRNP core protein from *Pyrococcus abyssi*. *Acta Crystallographica. Sect. D, Biological Crystallography* 2004b;60:122–4.
- Chavarría NE, Hwang S, Cao S et al. Archaeal Tuc1/Ncs6 homolog required for wobble uridine tRNA thiolation is associated with ubiquitin-proteasome, translation, and RNA processing system homologs. *PLoS One* 2014;9:e99104.
- Chen W-Y, Pulkunat DK, Cho I-M et al. Dissecting functional cooperation among protein subunits in archaeal RNase P, a catalytic ribonucleoprotein complex. *Nucleic Acids Res* 2010;38:8316–27.
- Chen W-Y, Singh D, Lai LB et al. Fidelity of tRNA 5'-maturation: a possible basis for the functional dependence of archaeal and eukaryal RNase P on multiple protein cofactors. *Nucleic Acids Res* 2012;40:4666–80.
- Chen W-Y, Xu Y, Cho I-M et al. Cooperative RNP assembly: complementary rescue of structural defects by protein and RNA subunits of archaeal RNase P. *J Mol Biol* 2011;411:368–83.
- Chlebowski A, Lubas M, Jensen TH et al. RNA decay machines: the exosome. *Biochim Biophys Acta* 2013;1829:552–60.
- Cho I-M, Lai LB, Susanti D et al. Ribosomal protein L7Ae is a subunit of archaeal RNase P. *Proc Natl Acad Sci USA* 2010;107:14573–8.
- Ciammaruconi A, Londei P. In vitro processing of the 16S rRNA of the thermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* 2001;183:3866–74.
- Clouet d'Orval B, Bortolin ML, Gaspin C et al. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* 2001;29:4518–29.
- Clouet-d'Orval B, Gaspin C, Mougin A. Two different mechanisms for tRNA ribose methylation in Archaea: a short survey. *Biochimie* 2005;87:889–95.
- Clouet-d'Orval B, Phung DK, Langendijk-Genevaux PS et al. Universal RNA-degrading enzymes in Archaea: prevalence, activities and functions of  $\beta$ -CASP ribonucleases. *Biochimie* 2015;118:278–85.
- Clouet-d'Orval B, Rinaldi D, Quentin Y et al. Euryarchaeal beta-CASP proteins with homology to bacterial RNase J have 5'- to 3'-exoribonuclease activity. *J Biol Chem* 2010;285:17574–83.
- Collins BM, Harrop SJ, Kornfeld GD et al. Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. *J Mol Biol* 2001;309:915–23.
- Crowe BL, Bohlen CJ, Wilson RC et al. Assembly of the complex between archaeal RNase P proteins RPP30 and Pop5. *Archaea* 2011;2011:1–12.
- Da Cunha V, Gaia M, Gabelle D et al. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet* 2017;13:e1006810.
- Danan M, Schwartz S, Edelheit S et al. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* 2012;40:3131–42.
- Daume M, Uhl M, Backofen R et al. RIP-Seq suggests translational regulation by L7Ae in Archaea. *MBio* 2017;8:e00730-17.
- De Koning B, Blombach F, Brouns SJJ et al. Fidelity in archaeal information processing. *Archaea* 2010;2010, DOI: 10.1155/2010/960298.
- Dennis PP, Tripp V, Lui L et al. C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. *BMC Genomics* 2015;16:632.
- Dennis PP, Ziesche S, Mylvaganam S. Transcription analysis of two disparate rRNA operons in the halophilic archaeon *Haloarcula marismortui*. *J Bacteriol* 1998;180:4804–13.
- Deppenmeier U, Johann A, Hartsch T et al. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biot* 2002;4:453–61.
- Deutscher MP. Ribonucleases, tRNA nucleotidyltransferase, and the 3' processing of tRNA. *Prog Nucleic Acid Re* 1990;39:209–40.
- Deutscher MP. Chapter 9 maturation and degradation of ribosomal RNA in bacteria. In Condon C (ed). *Molecular Biology of RNA Processing and Decay in Prokaryotes*. Elsevier, 2009, 369–91.
- Deutscher MP. How bacterial cells keep ribonucleases under control. *FEMS Microbiol Rev* 2015;39:350–61.
- De Zoysa MD, Yu Y-T. Posttranscriptional RNA pseudouridylation. *Enzymes* 2017;41:151–67.
- Dominski Z, Carpousis AJ, Clouet-d'Orval B. Emergence of the  $\beta$ -CASP ribonucleases: Highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim Biophys Acta* 2013;1829:532–51.
- Dorléans A, Li de la Sierra-Gallay I, Piton J et al. Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' Exo/Endoribonuclease RNase J. *Structure* 2011;19:1252–61.
- Durand S, Tomasini A, Braun F et al. sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol Rev* 2015;39:316–30.
- Durovic P, Dennis PP. Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaeobacterium *Sulfolobus aetdocaldarius*: similarities to eukaryotic rRNA processing. *Mol Microbiol* 1994;13:229–42.
- El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* 2012;46:69–95.



- Elkins JG, Podar M, Graham DE et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci USA* 2008;**105**:8102–7.
- Ellis JC, Brown JW. The RNase P family. *RNA Biol* 2009;**6**:362–9.
- Eme L, Spang A, Lombard J et al. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 2017;**15**:711–23.
- Englert M, Sheppard K, Aslanian A et al. Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation. *Proc Natl Acad Sci USA* 2011;**108**:1290–5.
- Englert M, Xia S, Okada C et al. Structural and mechanistic insights into guanylylation of RNA-splicing ligase RtcB joining RNA between 3'-terminal phosphate and 5'-OH. *Proc Natl Acad Sci* 2012;**109**:15235–40.
- Esakova O, Krasilnikov AS. Of proteins and RNA: the RNase P/MRP family. *RNA* 2010;**16**:1725–47.
- Evans D, Marquez SM, Pace NR. RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci* 2006;**31**:333–41.
- Even S, Pellegrini O, Zig L et al. Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Res* 2005;**33**:2141–52.
- Evguenieva-Hackenberg E, Hou L, Glaeser S et al. Structure and function of the archaeal exosome. *WIRES RNA* 2014;**5**:623–35.
- Evguenieva-Hackenberg E, Klug G. RNA degradation in Archaea and Gram-negative bacteria different from *Escherichia coli*. *Prog Mol Biol Transl Sci* 2009;**85**:275–317.
- Evguenieva-Hackenberg E, Klug G. New aspects of RNA processing in prokaryotes. *Curr Opin Microbiol* 2011;**14**:587–92.
- Evguenieva-Hackenberg E, Roppelt V, Finsterseifer P et al. Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome. *Biochemistry* 2008;**47**:13158–68.
- Evguenieva-Hackenberg E, Walter P, Hochleitner E et al. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep* 2003;**4**:889–93.
- Farhoud MH, Wessels HJCT, Steenbakkers PJM et al. Protein complexes in the Archaeon *Methanothermobacter thermautotrophicus* analyzed by blue native/SDS-PAGE and mass spectrometry. *Mol Cell Proteomics* 2005;**4**:1653–63.
- Fatica A, Tollervey D, Dlakić M. PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA* 2004;**10**:1698–701.
- Ferreira-Cerca S. Life and death of ribosomes in archaea. In Clouet-d'Orval, B (ed.) *RNA Metabolism and Gene Expression in Archaea*. Cham: Springer International Publishing, 2017, 129–58.
- Fischer S, Benz J, Späth B et al. The archaeal Lsm protein binds to small RNAs. *J Biol Chem* 2010;**285**:34429–38.
- Fischer S, Benz J, Späth B et al. Regulatory RNAs in *Haloferax volcanii*. *Biochem Soc Trans* 2011;**39**:159–62.
- Franckenberg S, Becker T, Beckmann R. Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. *Curr Opin Struct Biol* 2012;**22**:786–96.
- Frank DN, Pace NR. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem* 1998;**67**:153–80.
- Franze de Fernandez MT, Eoyang L, August JT. Factor fraction required for the synthesis of bacteriophage Q $\beta$ -RNA. *Nature* 1968;**219**:588–90.
- Fujishima K, Kanai A. tRNA gene diversity in the three domains of life. *Front Genet* 2014;**5**:142.
- Fujishima K, Sugahara J, Kikuta K et al. Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *Proc Natl Acad Sci USA* 2009;**106**:2683–7.
- Fujishima K, Sugahara J, Miller CS et al. A novel three-unit tRNA splicing endonuclease found in ultrasmall Archaea possesses broad substrate specificity. *Nucleic Acids Res* 2011;**39**:9695–704.
- Fujishima K, Sugahara J, Tomita M et al. Large-scale tRNA intron transposition in the archaeal order Thermoproteales represents a novel mechanism of intron gain. *Mol Biol Evol* 2010;**27**:2233–43.
- Fukuhara H, Kifusa M, Watanabe M et al. A fifth protein subunit Ph1496p elevates the optimum temperature for the ribonuclease P activity from *Pyrococcus horikoshii* OT3. *Biochem Biophys Res Commun* 2006;**343**:956–64.
- Gagnon KT, Zhang X, Qu G et al. Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 2010;**16**:79–90.
- Gao X, Oshima K, Ueda T et al. A three-dimensional model of RNase P in the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Biochem Biophys Res Commun* 2017;**493**:1063–68.
- Garside EL, Schellenberg MJ, Gesner EM et al. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR endonucleases. *RNA* 2012;**18**:2020–8.
- Gaspin C, Cavallé J, Erauso G et al. Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol* 2000;**297**:895–906.
- Gebetsberger J, Polacek N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol* 2013;**10**:1798–806.
- Gebetsberger J, Wyss L, Mleczko AM et al. A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol* 2017;**14**:1364–73.
- Gebetsberger J, Zywicki M, Künzi A et al. tRNA-Derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* 2012;**2012**:1–11.
- Geissler R, Grimson A. A position-specific 3'UTR sequence that accelerates mRNA decay. *RNA Biol* 2016;**13**:1075–7.
- Gopalan V. Uniformity amid diversity in RNase P. *Proc Natl Acad Sci USA* 2007;**104**:2031–2.
- Grosjean H, de Crécy-Lagard V, Marck C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett* 2010;**584**:252–64.
- Grosjean H, Gaspin C, Marck C et al. RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* 2008;**9**:470.
- Grosjean H, Westhof E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res* 2016;**44**:8020–40.
- Guerrier-Takada C, Gardiner K, Marsh T et al. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 1983;**35**:849–57.
- Gulen B, Petrov AS, Okafor CD et al. Ribosomal small subunit domains radiate from a central core. *Sci Rep* 2016;**6**:20885.
- Guy L, Ettema TJG. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* 2011;**19**:580–7.
- Haas ES, Daniels CJ, Reeve JN. Genes encoding 5S rRNA and tRNAs in the extremely thermophilic archaeobacterium *Methanothermobacter feravidus*. *Gene* 1989;**77**:253–63.
- Hale C, Kleppe K, Terns RM et al. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 2008;**14**:2572–9.
- Hale CR, Zhao P, Olson S et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 2009;**139**:945–56.
- Hall TA, Brown JW. The ribonuclease P family. *Method Enzymol* 2001;**341**:56–77.

- Hall TA, Brown JW. Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins. *RNA* 2002;**8**:296–306.
- Hallam SJ, Konstantinidis KT, Putnam N et al. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* 2006;**103**:18296–301.
- Hamma T, Ferré-D'Amaré AR. Structure of protein L7Ae bound to a K-Turn derived from an archaeal Box H/ACA sRNA at 1.8 Å resolution. *Structure* 2004;**12**:893–903.
- Harris JK, Haas ES, Williams D et al. New insight into RNase P RNA structure from comparative analysis of the archaeal RNA. *RNA* 2001;**7**:220–32.
- Hartman AL, Norais C, Badger JH et al. The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 2010;**5**:e9605.
- Hartmann E, Hartmann RK. The enigma of ribonuclease P evolution. *Trends Genet* 2003;**19**:561–9.
- Hartmann RK, Gössringer M, Späth B et al. The making of tRNAs and more - RNase P and tRNase Z. *Prog Mol Biol Transl Sci* 2009;**85**:319–68.
- Hartung S, Hopfner K-P. Lessons from structural and biochemical studies on the archaeal exosome. *Biochem Soc Trans* 2009;**37**:83–7.
- Hartung S, Niederberger T, Hartung M et al. Quantitative analysis of processive RNA degradation by the archaeal RNA exosome. *Nucleic Acids Res* 2010;**38**:5166–76.
- Hasenöhrl D, Konrat R, Bläsi U Identification of an RNase J ortholog in *Sulfolobus solfataricus*: implications for 5'-to-3' directional decay and 5'-end protection of mRNA in Crenarchaeota. *RNA* 2011;**17**:99–107.
- Hasenöhrl D, Lombo T, Kaberdin V et al. Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proc Natl Acad Sci USA* 2008;**105**:2146–50.
- Hatoum-Aslan A, Maniv I, Marraffini LA. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci USA* 2011;**108**:21218–22.
- Heider MR, Burkhart BW, Santangelo TJ et al. Defining the RNaseH2 enzyme-initiated ribonucleotide excision repair pathway in Archaea. *J Biol Chem* 2017;**292**:8835–45.
- Hellmich UA, Weis BL, Lioutikov A et al. Essential ribosome assembly factor Fap7 regulates a hierarchy of RNA-protein interactions during small ribosomal subunit biogenesis. *Proc Natl Acad Sci USA* 2013;**110**:15253–8.
- Hennigan AN, Reeve JN. mRNAs in the methanogenic archaeon *Methanococcus vannielii*: numbers, half-lives and processing. *Mol Microbiol* 1994;**11**:655–70.
- Henras AK, Plisson-Chastang C, O'Donohue M-F et al. An overview of pre-ribosomal RNA processing in eukaryotes. *WIREs RNA* 2015;**6**:225–42.
- Heyer R, Dörr M, Jellen-Ritter A et al. High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii*. *RNA Biol* 2012;**9**:1011–8.
- Hirata A, Fujishima K, Yamagami R et al. X-ray structure of the fourth type of archaeal tRNA splicing endonuclease: insights into the evolution of a novel three-unit composition and a unique loop involved in broad substrate specificity. *Nucleic Acids Res* 2012;**40**:10554–66.
- Hirata A, Kitajima T, Hori H. Cleavage of intron from the standard or non-standard position of the precursor tRNA by the splicing endonuclease of *Aeropyrum pernix*, a hyperthermophilic Crenarchaeon, involves a novel RNA recognition site in the Crenarchaea specific loop. *Nucleic Acids Res* 2011;**39**:9376–89.
- Hochstrasser ML, Doudna JA. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem Sci* 2015;**40**:58–66.
- Hözl A, Fischer S, Heyer R et al. Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* 2008;**14**:928–37.
- Honda T, Hara T, Nan J et al. Archaeal homologs of human RNase P protein pairs Pop5 with Rpp30 and Rpp21 with Rpp29 work on distinct functional domains of the RNA subunit. *Biosci Biotechnol Biochem* 2010;**74**:266–73.
- Honda T, Kakuta Y, Kimura K et al. Structure of an archaeal homolog of the human protein complex Rpp21-Rpp29 that is a key core component for the assembly of active ribonuclease P. *J Mol Biol* 2008;**384**:652–62.
- Hou L, Klug G, Evgueniev-Hackenberg E. The archaeal DnaG protein needs Csl4 for binding to the exosome and enhances its interaction with adenine-rich RNAs. *RNA Biol* 2013;**10**:415–24.
- Houri-Zeevi L, Rechavi O A matter of time: small rnas regulate the duration of epigenetic inheritance. *Trends Genet* 2017;**33**:46–57.
- Howard MJ, Liu X, Lim WH et al. RNase P enzymes. *RNA Biol* 2013;**10**:909–14.
- Huang L, Lilley DMJ. The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA* 2013;**19**:1703–10.
- Huang L, Lilley DMJ. The kink turn, a key architectural element in RNA structure. *J Mol Biol* 2016;**428**:790–801.
- Hui MP, Foley PL, Belasco JG. Messenger RNA degradation in bacterial cells. *Annu Rev Genet* 2014;**48**:537–59.
- Hundt S, Zaigler A, Lange C et al. Global analysis of mRNA decay in *Halobacterium salinarum* NRC-1 at single-gene resolution using DNA microarrays. *J Bacteriol* 2007;**189**:6936–44.
- Ikeda Y, Okada Y, Sato A et al. An archaeal RNA binding protein, FAU-1, is a novel ribonuclease related to rRNA stability in *Pyrococcus* and *Thermococcus*. *Sci Rep* 2017;**7**:12674.
- Inada T. Quality control systems for aberrant mRNAs induced by aberrant translation elongation and termination. *Biochim Biophys Acta* 2013;**1829**:634–42.
- Inada T. The ribosome as a platform for mRNA and nascent polypeptide quality control. *Trends Biochem Sci* 2017;**42**:5–15.
- Ishii R, Minagawa A, Takaku H et al. Crystal structure of the tRNA 3' processing endoribonuclease tRNase Z from *Thermotoga maritima*. *J Biol Chem* 2005;**280**:14138–44.
- Ishii R, Minagawa A, Takaku H et al. The structure of the flexible arm of *Thermotoga maritima* tRNase Z differs from those of homologous enzymes. *Acta Crystallogr F* 2007;**63**:637–41.
- Itoh T, Nomura N, Sako Y. Distribution of 16S rRNA introns among the family Thermoproteaceae and their evolutionary implications. *Extremophiles* 2003;**7**:229–33.
- Jäger D, Förstner KU, Sharma CM et al. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* 2014;**15**:684.
- Jäger A, Samorski R, Pfeifer F et al. Individual gyp transcript segments in *Haloferax mediterranei* exhibit varying half-lives, which are differentially affected by salt concentration and growth phase. *Nucleic Acids Res* 2002;**30**:5436–43.
- Jäger D, Sharma CM, Thomsen J et al. Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci USA* 2009;**106**:21878–82.

- Jarrous N, Gopalan V. Archaeal/eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic Acids Res* 2010;**38**:7885–94.
- Jaschinski K, Babski J, Lehr M et al. Generation and phenotyping of a collection of sRNA gene deletion mutants of the haloarchaeon *Haloferax volcanii*. *PLoS One* 2014;**9**:e90763.
- Jay ZJ, Inskeep, WP. The distribution, diversity, and importance of 16S rRNA gene introns in the order Thermoproteales. *Biol Direct* 2015;**10**:35.
- Kaberlin VR, Singh D, Lin-Chao S. Composition and conservation of the mRNA-degrading machinery in bacteria. *J Biomed Sci* 2011;**18**:23.
- Kakuta Y, Ishimatsu I, Numata T et al. Crystal structure of a ribonuclease P protein Ph1601p from *Pyrococcus horikoshii* OT3: an archaeal homologue of human nuclear ribonuclease P protein Rpp21. *Biochemistry* 2005;**44**:12086–93.
- Kambach C, Walke S, Nagai K. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* 1999;**9**:222–30.
- Kawano S, Nakashima T, Kakuta Y et al. Crystal structure of protein Ph1481p in complex with protein Ph1877p of archaeal RNase P from *Pyrococcus horikoshii* OT3: implication of dimer formation of the holoenzyme. *J Mol Biol* 2006;**357**:583–91.
- Keam SP, Hutvagner G. tRNA-derived fragments (tRFs): emerging new roles for an ancient RNA in the regulation of gene expression. *Life (Basel)* 2015;**5**:1638–51.
- Khemici V, Prados J, Linder P et al. Decay-initiating endoribonucleolytic cleavage by RNase Y is kept under tight control via sequence preference and sub-cellular localisation. *PLoS Genet* 2015;**11**:e1005577.
- Kikovska E, Svärd SG, Kirsebom LA. Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proc Natl Acad Sci USA* 2007;**104**:2062–7.
- Kim YK, Mizutani K, Rhee K-H et al. Structural and mutational analysis of tRNA intron-splicing endonuclease from *Thermoplasma acidophilum* DSM 1728: catalytic mechanism of tRNA intron-splicing endonucleases. *J Bacteriol* 2007;**189**:8339–46.
- Kimura M. Structural basis for activation of an archaeal ribonuclease P RNA by protein cofactors. *Biosci Biotechnol Biochem* 2017;**81**:1670–80.
- Kimura M, Kakuta Y. Structural biology of the ribonuclease P in the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. In Prakash, A, Satyanarayana, T, Johri, BN (eds) *Microorganisms in Sustainable Agriculture and Biotechnology*. Dordrecht, Netherlands: Springer, 2012, 487–508.
- Kirsebom LA. RNase P RNA mediated cleavage: substrate recognition and catalysis. *Biochimie* 2007;**89**:1183–94.
- Kitamura S, Fujishima K, Sato A et al. Characterization of RNase HII substrate recognition using RNase HII-argonaute chimaeric enzymes from *Pyrococcus furiosus*. *Biochem J* 2010;**426**:337–44.
- Kjems J, Garrett RA. Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* 1988;**54**:693–703.
- Klein DJ, Schmeing TM, Moore PB et al. The kink-turn: a new RNA secondary structure motif. *EMBO J* 2001;**20**:4214–21.
- Klemm BP, Wu N, Chen Y et al. The diversity of ribonuclease P: protein and RNA catalysts with analogous biological functions. *Biomolecules* 2016;**6**:27, <https://doi.org/10.3390/biom6020027>.
- Klug G, Evgueniev-Hackenberg E, Omer A et al. RNA processing. In: Cavicchioli, R (ed.) *Archaea: Molecular and Cellular Biology*. Washington DC: ASM Press, 2007, 158–74.
- Kobayashi K, Ishitani R, Nureki O. Recent structural studies on Dom34/aPelota and Hbs1/aEF1 $\alpha$ : important factors for solving general problems of ribosomal stall in translation. *Biophysics* 2013;**9**:131–40.
- Kobayashi K, Kikuno I, Kuroha K et al. Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1 $\alpha$  complex. *Proc Natl Acad Sci USA* 2010;**107**:17575–9.
- Kochiwa H, Tomita M, Kanai A. Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evol Biol* 2007;**7**:128.
- Koonin EV. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Philos T Roy Soc B* 2015;**370**:1–15.
- Koonin EV, Makarova KS. Mobile genetic elements and evolution of CRISPR-Cas Systems: all the way there and back. *Genome Biol Evol* 2017;**9**:2812–25.
- Koonin EV, Wolf YI, Aravind L. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res* 2001;**11**:240–52.
- Kouzuma Y, Mizoguchi M, Takagi H et al. Reconstitution of archaeal ribonuclease P from RNA and four protein components. *Biochem Biophys Res Commun* 2003;**306**:666–73.
- Krupovic M, Gonnet M, Hania WB et al. Insights into dynamics of mobile genetic elements in hyperthermophilic environments from five new *Thermococcus* plasmids. *PLoS One* 2013;**8**:e49044.
- Kuhn C-D, Wilusz JE, Zheng Y et al. On-enzyme refolding permits small RNA and tRNA surveillance by the CCA-adding enzyme. *Cell* 2015;**160**:644–58.
- Kuhn JF, Tran EJ, Maxwell ES. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* 2002;**30**:931–41.
- Kumar P, Anaya J, Mudunuri SB et al. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol* 2014;**12**:78.
- Kumar P, Kuscu C, Dutta A. Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends Biochem Sci* 2016;**41**:679–89.
- Li de la Sierra-Gallay I, Mathy N, Pellegrini O et al. Structure of the ubiquitous 3' processing enzyme RNase Z bound to transfer RNA. *Nat Struct Mol Biol* 2006;**13**:376–7.
- Li de la Sierra-Gallay I, Pellegrini O, Condon C. Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* 2005;**433**:657–61.
- Li de la Sierra-Gallay I, Zig L, Jamali A et al. Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol* 2008;**15**:206–12.
- Laalami S, Zig L, Putzer H. Initiation of mRNA decay in bacteria. *Cell Mol Life Sci* 2014;**71**:1799–828.
- Lafontaine DL, Tollervey D. The function and synthesis of ribosomes. *Nat Rev Mol Cell Biol* 2001;**2**:514–20.
- Lai L, Yokota H, Hung LW et al. Crystal structure of archaeal RNase HII: a homologue of human major RNase H. *Structure* 2000;**8**:897–904.
- Lai LB, Chan PP, Cozen AE et al. Discovery of a minimal form of RNase P in pyrobaculum. *Proc Natl Acad Sci USA* 2010a;**107**:22493–8.
- Lai LB, Tanimoto A, Lai SM et al. A novel double kink-turn module in euryarchaeal RNase P RNAs. *Nucleic Acids Res* 2017;**45**:7432–40.
- Lai LB, Vioque A, Kirsebom LA et al. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett* 2010b;**584**:287–96.



- Lai SM, Lai LB, Foster MP et al. The L7Ae protein binds to two kink-turns in the *Pyrococcus furiosus* RNase P RNA. *Nucleic Acids Res* 2014;**42**:13328–38.
- Lange SJ, Alkhnbashi OS, Rose D et al. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res* 2013;**41**:8034–44.
- Lechner M, Rossmannith W, Hartmann RK et al. Distribution of ribonucleoprotein and protein-only RNase P in eukarya. *Mol Biol Evol* 2015;**32**:3186–93.
- Lee HH, Jang JY, Yoon H-J et al. Crystal structures of two archaeal Pelotas reveal inter-domain structural plasticity. *Biochem Biophys Res Commun* 2010;**399**:600–6.
- Lee HH, Kim Y-S, Kim KH et al. Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol Cell* 2007;**27**:938–50.
- Leroy M, Piton J, Gilet L et al. Rae1/YacP, a new endoribonuclease involved in ribosome-dependent mRNA decay in *Bacillus subtilis*. *EMBO J* 2017;**36**:1167–81.
- Lescoute A, Leontis NB, Massire C et al. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res* 2005;**33**:2395–409.
- Levasseur A, Merhej V, Baptiste E et al. The rhizome of Lokiarchaeota illustrates the mosaicism of archaeal genomes. *Genome Biol Evol* 2017;**9**:2635–9.
- Levy S, Portnoy V, Admon J et al. Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biol* 2011;**8**:1073–83.
- Li D, Willkomm DK, Hartmann RK. Minor changes largely restore catalytic activity of archaeal RNase P RNA from *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res* 2009;**37**:231–42.
- Li H. Structural principles of CRISPR RNA processing. *Structure* 2015;**23**:13–20.
- Li H, Abelson J. Crystal structure of a dimeric archaeal splicing endonuclease. *J Mol Biol* 2000;**302**:639–48.
- Li H, Trotta CR, Abelson J. Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* 1998;**280**:279–84.
- Li J, Qi L, Guo Y et al. Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon *Methanobrevibacter psychrophilus*. *Sci Rep* 2015;**5**:9209.
- Li L, Ye K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 2006;**443**:302–7.
- Li Z, Deutscher MP. Maturation pathways for *E. coli* tRNA precursors: a random multienzyme process in vivo. *Cell* 1996;**86**:503–12.
- Li Z, Santangelo TJ, Cuboňová L et al. Affinity purification of an archaeal DNA replication protein network. *mBio* 2010;**1**:e00221–10.
- Lilley DMJ. The structure and folding of kink turns in RNA. *WIREs RNA* 2012;**3**:797–805.
- Lopes RRS, Kessler AC, Polycarpo C et al. Cutting, dicing, healing and sealing: the molecular surgery of tRNA. *WIREs RNA* 2015;**6**:337–49.
- Lorentzen E., Conti E. Crystal structure of a 9-subunit archaeal exosome in pre-catalytic states of the phosphorolytic reaction. *Archaea* 2012;**7**:21869.
- Lorentzen E., Conti E. Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. *Molecular Cell* 2005;**20**:473–81.
- Lorentzen E, Dziembowski A, Lindner D et al. RNA channelling by the archaeal exosome. *EMBO Rep* 2007;**8**:470–6.
- Lorentzen E, Walter P, Fribourg S et al. The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat Struct Mol Biol* 2005;**12**:575–81.
- Lu C, Ding F, Ke A. Crystal structure of the *S. solfataricus* archaeal exosome reveals conformational flexibility in the RNA-binding ring. *PLoS One* 2010;**5**:e8739.
- Luz JS, Ramos CRR, Santos MCT et al. Identification of archaeal proteins that affect the exosome function in vitro. *BMC Biochem* 2010;**11**:22.
- Lykke-Andersen J, Aagaard C, Semionenkova M et al. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci* 1997;**22**:326–31.
- Lykke-Andersen J, Garrett RA. RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO J* 1997;**16**:6290–300.
- Ma M, Li de la Sierra-Gallay I, Lazar N et al. The crystal structure of Trz1, the long form RNase Z from yeast. *Nucleic Acids Res* 2017;**45**:6209–16.
- Machnicka MA, Milanowska K, Osman Oglou O et al. MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res* 2013;**41**:D262–7.
- Mackie GA. Determinants in the rpsT mRNAs recognized by the 5'-sensor domain of RNase E. *Mol Microbiol* 2013;**89**:388–402.
- Maier L-K, Benz J, Fischer S et al. Deletion of the Sm1 encoding motif in the lsm gene results in distinct changes in the transcriptome and enhanced swarming activity of *Haloferax* cells. *Biochimie* 2015;**117**:129–37.
- Maier L-K, Dyll-Smith M, Marchfelder A. The adaptive immune system of *Haloferax volcanii*. *Life (Basel)* 2015;**5**:521–37.
- Makarova KS, Aravind L, Wolf YI et al. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 2011;**6**:38.
- Makarova KS, Wolf YI, Alkhnbashi OS et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 2015;**13**:722–36.
- Mao H, White SA, Williamson JR. A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat Struct Biol* 1999;**6**:1139–47.
- Marchfelder A, Fischer S, Brendel J et al. Small RNAs for defence and regulation in archaea. *Extremophiles* 2012;**16**:685–96.
- Marck C, Grosjean H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 2002;**8**:1189–232.
- Marck C, Grosjean H. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* 2003;**9**:1516–31.
- Márquez V, Fröhlich T, Armache J-P et al. Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J Mol Biol* 2011;**405**:1215–32.
- Märtens B, Amman F, Manoharadas S et al. Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *PLoS One* 2013;**8**:e76569.
- Märtens B, Bezerra GA, Kreuter MJ et al. The heptameric smap1 and smap2 proteins of the crenarchaeon *Sulfolobus solfataricus* bind to common and distinct RNA targets. *Life (Basel)* 2015;**5**:1264–81.
- Märtens B, Hou L, Amman F et al. The SmAP1/2 proteins of the crenarchaeon *Sulfolobus solfataricus* interact with the exosome and stimulate A-rich tailing of transcripts. *Nucleic Acids Res* 2017;**45**:7938–49.

- Massenet S, Bertrand E, Verheggen C. Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* 2017;**14**:680–92.
- Matos RG, López-Viñas E, Gómez-Puertas P et al. The only exoribonuclease present in *Haloferax volcanii* has an unique response to temperature changes. *BBA-Gen Subjects* 2012;**1820**:1543–52.
- Minagawa A, Takaku H, Takagi M et al. A novel endonucleolytic mechanism to generate the CCA 3' termini of tRNA molecules in *Thermotoga maritima*. *J Biol Chem* 2004;**279**:15688–97.
- Mir-Montazeri B, Ammelburg M, Forouzan D et al. Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J Struct Biol* 2011;**173**:191–5.
- Mitchell M, Xue S, Erdman R et al. Crystal structure and assembly of the functional Nanoarchaeum equitans tRNA splicing endonuclease. *Nucleic Acids Res* 2009;**37**:5793–802.
- Mohanraju P, Makarova KS, Zetsche B et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 2016;**353**:aad5147.
- Moissl-Eichinger C, Pausan M, Taffner J et al. Archaea are interactive components of complex microbiomes. *Trends Microbiol* 2018;**26**:70–85.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading, Engl)* 2009;**155**:733–40.
- Morinaga Y, Nomura N, Sako Y. Population dynamics of archaeal mobile introns in natural environments: A shrewd invasion strategy of the latent parasitic DNA. *Microb Environ* 2002;**17**:153–63.
- Motorin Y, Helm M. RNA nucleotide methylation. *WIREs RNA* 2011;**2**:611–31.
- Muller S, Leclerc F, Behm-Ansmant I et al. Combined in silico and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res* 2008;**36**:2459–75.
- Mura C, Cascio D, Sawaya MR et al. The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. *Proc Natl Acad Sci USA* 2001;**98**:5532–7.
- Mura C, Kozhukhovskiy A, Gingery M et al. The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). *Protein Sci* 2003a;**12**:832–47.
- Mura C, Phillips M, Kozhukhovskiy A et al. Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proc Natl Acad Sci USA* 2003b;**100**:4539–44.
- Mura C, Randolph PS, Patterson J et al. Archaeal and eukaryotic homologs of Hfq. *RNA Biol* 2013;**10**:636–51.
- Muroya A, Tsuchiya D, Ishikawa M et al. Catalytic center of an archaeal type 2 ribonuclease H as revealed by X-ray crystallographic and mutational analyses. *Protein Sci* 2001;**10**:707–14.
- Nagarajan VK, Jones CI, Newbury SF et al. XRN 5'→3' exoribonucleases: structure, mechanisms and functions. *Biochim Biophys Acta* 2013;**1829**:590–603.
- Nam KH, Haitjema C, Liu X et al. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* 2012;**20**:1574–84.
- Navarro MVAS, Oliveira CC, Zanchin NIT et al. Insights into the mechanism of progressive RNA degradation by the archaeal exosome. *J Biol Chem* 2008;**283**:14120–31.
- Newman JA, Hewitt L, Rodrigues C et al. Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure* 2011;**19**:1241–51.
- Nickel AI, Wäber NB, Gößringer M et al. Minimal and RNA-free RNase P in *Aquifex aeolicus*. *Proc Natl Acad Sci USA* 2017;**114**:11121–6.
- Nielsen JS, Bøggild A, Andersen CBF et al. An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. *RNA* 2007;**13**:2213–23.
- Niewoehner O, Jinek M, Doudna JA. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res* 2014;**42**:1341–53.
- Nikulina A, Mikhailina A, Lekontseva N et al. Characterization of RNA-binding properties of the archaeal Hfq-like protein from *Methanococcus jannaschii*. *J Biomol Struct Dyn* 2017;**35**:1615–28.
- Nishida Y, Ishikawa H, Baba S et al. Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*. *Proteins* 2010;**78**:2395–8.
- Nissen P, Hansen J, Ban N et al. The structural basis of ribosome activity in peptide bond synthesis. *Science* 2000;**289**:920–30.
- Nolivos S, Carpousis AJ, Clouet-d'Orval B. The K-loop, a general feature of the *Pyrococcus* C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res* 2005;**33**:6507–14.
- Nomura N, Sako Y, Uchida A. Molecular characterization and postsplicing fate of three introns within the single rRNA operon of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *J Bacteriol* 1998;**180**:3635–43.
- Numata T, Ishimatsu I, Kakuta Y et al. Crystal structure of archaeal ribonuclease P protein Ph1771p from *Pyrococcus horikoshii* OT3: an archaeal homolog of eukaryotic ribonuclease P protein Rpp29. *RNA* 2004;**10**:1423–32.
- Nunoura T, Takaki Y, Kakuta J et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* 2011;**39**:3204–23.
- Okabe M, Tomita K, Ishitani R et al. Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure. *EMBO J* 2003;**22**:5918–27.
- Okuda M, Shiba T, Inaoka D-K et al. A conserved lysine residue in the crenarchaea-specific loop is important for the crenarchaeal splicing endonuclease activity. *J Mol Biol* 2011;**405**:92–104.
- Omer AD, Lowe TM, Russell AG et al. Homologs of small nucleolar RNAs in Archaea. *Science* 2000;**288**:517–22.
- Omer AD, Ziesche S, Ehardt H et al. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci USA* 2002;**99**:5289–94.
- Oshima K, Kakiuchi Y, Tanaka Y et al. Structural basis for recognition of a kink-turn motif by an archaeal homologue of human RNase P protein Rpp38. *Biochem Biophys Res Commun* 2016;**474**:541–6.
- Palmer JR, Nieuwlandt DT, Daniels CJ. Expression of a yeast intron-containing tRNA in the archaeon *Haloferax volcanii*. *J Bacteriol* 1994;**176**:3820–3.
- Pan B, Xiong Y, Steitz TA. How the CCA-adding enzyme selects adenine over cytosine at position 76 of tRNA. *Science* 2010;**330**:937–40.
- Pannucci JA, Haas ES, Hall TA et al. RNase P RNAs from some Archaea are catalytically active. *Proc Natl Acad Sci USA* 1999;**96**:7803–8.
- Park H-M, Shin M, Sun J et al. Crystal structure of a Cas6 paralogous protein from *Pyrococcus furiosus*. *Proteins* 2012;**80**:1895–900.
- Peterson JR, Thor S, Kohler L et al. Genome-wide gene expression and RNA half-life measurements allow predictions of regulation and metabolic behavior in *Methanosarcina acetivorans*. *BMC Genomics* 2016;**17**:924.



- Petrov AS, Bernier CR, Gulen B et al. Secondary structures of rRNAs from all three domains of life. *PLoS One* 2014;**9**:e88222.
- Phillips G, de Crécy-Lagard V. Biosynthesis and function of tRNA modifications in Archaea. *Curr Opin Microbiol* 2011;**14**:335–41.
- Phok K, Moisan A, Rinaldi D et al. Identification of CRISPR and riboswitch related RNAs among novel noncoding RNAs of the euryarchaeon *Pyrococcus abyssi*. *BMC Genomics* 2011;**12**:312.
- Phung DK, Bouvier M, Clouet-d'Orval B. An overview of ribonuclease repertoire and RNA processing pathways in archaea. In: Clouet-d'Orval, B (ed.) *RNA Metabolism and Gene Expression in Archaea*. Cham: Springer International Publishing, 2017, 89–114.
- Phung DK, Rinaldi D, Langendijk-Genevaux PS et al. Archaeal  $\beta$ -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* 2013;**41**:1091–103.
- Pichard-Kostuch A, Daugeron M-C, Forterre P et al. Function and biosynthesis of the universal tRNA modification N6-threonylcarbamoyl-adenosine. In: Clouet-d'Orval B (ed.) *RNA Metabolism and Gene Expression in Archaea*. Cham: Springer International Publishing, 2017, 177–200.
- Plagens A, Daume M, Wiegel J et al. Circularization restores signal recognition particle RNA functionality in *Thermoproteus*. *Elife* 2015;**4**:e11623.
- Plagens A, Tripp V, Daume M et al. In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res* 2014;**42**:5125–38.
- Porrua O, Libri D. RNA quality control in the nucleus: The Angels' share of RNA. *Biochim Biophys Acta* 2013;**1829**:604–11.
- Portnoy V, Evgueniya-Hackenberg E, Klein F et al. RNA polyadenylation in Archaea: not observed in *Haloferax* while the exosome polynucleotidylates RNA in *Sulfolobus*. *EMBO Rep* 2005;**6**:1188–93.
- Portnoy V, Schuster G. RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Res* 2006;**34**:5923–31.
- Pulukunat DK, Gopalan V. Studies on *Methanocaldococcus jannaschii* RNase P reveal insights into the roles of RNA and protein cofactors in RNase P catalysis. *Nucleic Acids Res* 2008;**36**:4172–80.
- Qi L, Yue L, Feng D et al. Genome-wide mRNA processing in methanogenic archaea reveals post-transcriptional regulation of ribosomal protein synthesis. *Nucleic Acids Res* 2017;**45**:7285–98.
- Ragan MA, Logsdon JM, Sensen CW et al. An archaeobacterial homolog of pelota, a meiotic cell division protein in eukaryotes. *FEMS Microbiol Lett* 1996;**144**:151–5.
- Ramos CRR, Oliveira CLP, Torriani IL et al. The *Pyrococcus* exosome complex. *J Biol Chem* 2006;**281**:6751–9.
- Randau L. RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol* 2012;**13**:R63.
- Randau L. Evolution of small guide RNA genes in hyperthermophilic archaea. *Ann NY Acad Sci* 2015;**1341**:188–93.
- Randau L, Calvin K, Hall M et al. The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proc Natl Acad Sci USA* 2005;**102**:17934–9.
- Randau L, Schröder I, Söll D. Life without RNase P. *Nature* 2008;**453**:120–3.
- Rath D, Amlinger L, Rath A et al. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* 2015;**117**:119–28.
- Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci USA* 2015;**112**:6670–5.
- Redder P. How does sub-cellular localization affect the fate of bacterial mRNA? *Curr Genet* 2016;**62**:687–90.
- Ree HK, Zimmermann RA. Organization and expression of the 16S, 23S and 5S ribosomal RNA genes from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Res* 1990;**18**:4471–8.
- Reeks J, Sokolowski RD, Graham S et al. Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. *Biochem J* 2013;**452**:223–30.
- Reiter NJ, Osterman A, Torres-Larios A et al. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* 2010;**468**:784–9.
- Richter H, Zoepfel J, Schermuly J et al. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res* 2012;**40**:9887–96.
- Rinke C, Schwientek P, Sczyrba A et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;**499**:431–7.
- Rozhdestvensky TS, Tang TH, Tchirkova IV et al. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* 2003;**31**:869–77.
- Saito K, Kobayashi K, Wada M et al. Omnipotent role of archaeal elongation factor 1 alpha (EF1 $\alpha$ ) in translational elongation and termination, and quality control of protein synthesis. *Proc Natl Acad Sci USA* 2010;**107**:19242–7.
- Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S et al. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J* 1999;**18**:3451–62.
- Samanta MP, Lai SM, Daniels CJ et al. Sequence analysis and comparative study of the protein subunits of archaeal RNase P. *Biomolecules* 2016;**6**:22, DOI: 10.3390/biom6020022.
- Schierling K, Rösch S, Rupprecht R et al. tRNA 3' end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *J Mol Biol* 2002;**316**:895–902.
- Schiffer S, Rösch S, Marchfelder A. Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J* 2002;**21**:2769–77.
- Schiffer S, Rösch S, Marchfelder A. Recombinant RNase Z does not recognize CCA as part of the tRNA and its cleavage efficiency is influenced by acceptor stem length. *Biol Chem* 2003;**384**:333–42.
- Scholz I, Lange SJ, Hein S et al. CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS One* 2013;**8**:e56470.
- Schweingruber C, Rufener SC, Zünd D et al. Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta* 2013;**1829**:612–23.
- Seistrup KH, Rose S, Birkedal U et al. Bypassing rRNA methylation by RsmA/Dim1 during ribosome maturation in the hyperthermophilic archaeon *Nanoarchaeum equitans*. *Nucleic Acids Res* 2017;**45**:2007–15.
- Shao Y, Li H. Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. *Structure* 2013;**21**:385–93.
- She Q, Singh RK, Confalonieri F et al. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* 2001;**98**:7835–40.
- Sidote DJ, Heideker J, Hoffman DW. Crystal structure of archaeal ribonuclease P Protein aRpp29 from *Archaeoglobus fulgidus*. *Biochemistry* 2004;**43**:14128–38.

- Sidote DJ, Hoffman DW. NMR structure of an archaeal homologue of ribonuclease P Protein Rpp29. *Biochemistry* 2003;**42**:13541–50.
- Siebers B, Zaparty M, Raddatz G et al. The complete genome sequence of *Thermoproteus tenax*: a physiologically versatile member of the Crenarchaeota. *PLoS One* 2011;**6**:e24222.
- Silva AP, Chechik M, Byrne RT et al. Structure and activity of a novel archaeal beta-CASP protein with N-terminal KH domains. *Structure* 2011;**19**:622–32.
- Simms CL, Thomas EN, Zaher HS. Ribosome-based quality control of mRNA and nascent peptides. *WIREs RNA* 2017;**8**:e1366, DOI: 10.1002/wrna.1366.
- Sinapah S, Wu S, Chen Y et al. Cleavage of model substrates by archaeal RNase P: role of protein cofactors in cleavage-site selection. *Nucleic Acids Res* 2011;**39**:1105–16.
- Sittka A, Sharma CM, Rolle K et al. Deep sequencing of *Salmonella* RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol* 2009;**6**:266–75.
- Sloan KE, Warda AS, Sharma S et al. Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* 2017;**14**:1138–52.
- Smith DR, Doucette-Stamm LA, Deloughery C et al. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* 1997;**179**:7135–55.
- Sobala A, Hutvagner G. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biol* 2013;**10**:553–63.
- Soppa J, Straub J, Brenneis M et al. Small RNAs of the halophilic archaeon *Haloferax volcanii*. *Biochem Soc Trans* 2009;**37**:133–6.
- Soss SE, Flynn PF. Functional implications for a prototypical K-Turn binding protein from structural and dynamical studies of 15.5K. *Biochemistry* 2007;**46**:14979–86.
- Späth B, Canino G, Marchfelder A. tRNase Z: the end is not in sight. *Cell Mol Life Sci* 2007;**64**:2404–12.
- Späth B, Schubert S, Lieberoth A et al. Two archaeal tRNase Z enzymes: similar but different. *Arch Microbiol* 2008;**190**:301–8.
- Sprinzl M, Hartmann T, Meissner F et al. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* 1987;**15**(suppl):r53–r188.
- Starostina NG, Marshburn S, Johnson LS et al. Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc Natl Acad Sci USA* 2004;**101**:14097–101.
- Steitz JA, Tycowski KT. Enhanced perspective: small RNA chaperones for ribosome biogenesis. *Science* 1995;**270**:1626–.
- Stoddard BL. Homing endonuclease structure and function. *Quart Rev Biophys* 2005;**38**:49–95.
- Stoddard SF, Smith BJ, Hein R et al. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* 2015;**43**:D593–8.
- Stoecklin G, Mühlemann O. RNA decay mechanisms: specificity through diversity. *Biochim Biophys Acta* 2013;**1829**:487–90.
- Straub J, Brenneis M, Jellen-Ritter A et al. Small RNAs in haloarchaea: identification, differential expression and biological function. *RNA Biol* 2009;**6**:281–92.
- Su AAH, Tripp V, Randau L. RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile *Methanopyrus kandleri*. *Nucleic Acids Res* 2013;**41**:6250–8.
- Suematsu K, Ueda T, Nakashima T et al. On archaeal homologs of the human RNase P proteins Pop5 and Rpp30 in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Biosci Biotechnol Biochem* 2015;**79**:952–9.
- Sugahara J, Kikuta K, Fujishima K et al. Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermoproteales. *Mol Biol Evol* 2008;**25**:2709–16.
- Suryadi J, Tran EJ, Maxwell ES et al. The crystal structure of the methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the Box C/D RNAs. *Biochemistry* 2005;**44**:9657–72.
- Sulthana S, Basturea GN, Deutscher MP. Elucidation of pathways of ribosomal RNA degradation: an essential role for RNase E. *RNA* 2016;**22**:1163–71.
- Tadokoro T, Kanaya S. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* 2009;**276**:1482–93.
- Tamulaitis G, Venclovas Č, Siksnys V. Type III CRISPR-Cas immunity: major differences brushed aside. *Trends Microbiol* 2017;**25**:49–61.
- Tang T-H, Polacek N, Zywicki M et al. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 2005;**55**:469–81.
- Tang TH, Rozhdestvensky TS, d Orval BC et al. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* 2002;**30**:921–30.
- Thompson LD, Daniels CJ. Recognition of exon-intron boundaries by the *Halobacterium volcanii* tRNA intron endonuclease. *J Biol Chem* 1990;**265**:18104–11.
- Thore S, Mayer C, Sauter C et al. Crystal structures of the *Pyrococcus abyssi* sm core and its complex with RNA. *J Biol Chem* 2003;**278**:1239–47.
- Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. Co-evolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc Natl Acad Sci USA* 2005;**102**:15418–22.
- Toffano-Nioche C, Ott A, Crozat E et al. RNA at 92°C: the non-coding transcriptome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *RNA Biol* 2013;**10**:1211–20.
- Tomita K, Ishitani R, Fukai S et al. Complete crystallographic analysis of the dynamics of CCA sequence addition. *Nature* 2006;**443**:956–60.
- Törö I, Basquin J, Teo-Dreher H et al. Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. *J Mol Biol* 2002;**320**:129–42.
- Törö I, Thore S, Mayer C et al. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J* 2001;**20**:2293–303.
- Tripp V, Martin R, Orell A et al. Plasticity of archaeal C/D box sRNA biogenesis. *Mol Microbiol* 2017;**103**:151–64.
- Trotta CR, Miao F, Arn EA et al. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* 1997;**89**:849–58.
- Tsai H-Y, Pulukkunat DK, Woznick WK et al. Functional reconstitution and characterization of *Pyrococcus furiosus* RNase P. *Proc Natl Acad Sci USA* 2006;**103**:16147–52.
- Van der Gulik PTS, Hoff WD, Speijer D. In defence of the three-domains of life paradigm. *BMC Evol Biol* 2017;**17**:218.
- Van der Oost J, Westra ER, Jackson RN et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* 2014;**12**:479–92.

- Väre VYP, Eruysal ER, Narendran A et al. Chemical and conformational diversity of modified nucleosides affects tRNA structure and function. *Biomolecules* 2017;7:29, DOI: 10.3390/biom7010029.
- Veith T, Martin R, Wurm JP et al. Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic Acids Res* 2012;40:3259–74.
- Vestergaard G, Garrett RA, Shah SA. CRISPR adaptive immune systems of Archaea. *RNA Biol* 2014;11:156–67.
- Vidovic I, Nottrott S, Hartmuth K et al. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* 2000;6:1331–42.
- Vogel A, Schilling O, Späth B et al. The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biol Chem* 2005;386:1253–64.
- Wagner EGH, Romby P Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet* 2015;90:133–208.
- Wang R, Preamplume G, Terns MP et al. Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 2011;19:257–64.
- Waters E, Hohn MJ, Ahel I et al. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 2003;100:12984–8.
- Weisel J, Wagner S, Klug G. The Nop5-L7A-fibrillarin RNP complex and a novel box C/D containing sRNA of *Halobacterium salinarum* NRC-1. *Biochem Biophys Res Commun* 2010;394:542–7.
- Wilson RC, Bohlen CJ, Foster MP et al. Structure of Pfu Pop5, an archaeal RNase P protein. *Proc Natl Acad Sci* 2006;103:873–8.
- Wilusz JE, Whipple JM, Phizicky EM et al. tRNAs marked with CCACCA are targeted for degradation. *Science* 2011;334:817–21.
- Witharana C, Roppelt V, Lochnit G et al. Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie* 2012;94:1578–87.
- Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990;87:4576–9.
- Xiong Y, Steitz TA. Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. *Nature* 2004;430:640–5.
- Xu Y, Amero CD, Pulukkunat DK et al. Solution structure of an archaeal RNase P binary protein complex: formation of the 30-kDa complex between *Pyrococcus furiosus* RPP21 and RPP29 is accompanied by coupled protein folding and highlights critical features for protein-protein and protein-RNA interactions. *J Mol Biol* 2009;393:1043–55.
- Xue S, Calvin K, Li H. RNA recognition and cleavage by a splicing endonuclease. *Science* 2006;312:906–10.
- Yokobori S, Itoh T, Yoshinari S et al. Gain and loss of an intron in a protein-coding gene in Archaea: the case of an archaeal RNA pseudouridine synthase gene. *BMC Evol Biol* 2009;9:198.
- Yoshinari S, Shiba T, Inaoka D-K et al. Functional importance of crenarchaea-specific extra-loop revealed by an X-ray structure of a heterotetrameric crenarchaeal splicing endonuclease. *Nucleic Acids Res* 2009;37:4787–98.
- Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. *Front Genet* 2014;5:213.
- Yue D, Maizels N, Weiner AM. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA* 1996;2:895–908.
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017;541:353–8.
- Zhang J, Ferré-D'Amaré AR. The tRNA elbow in structure, recognition and evolution. *Life (Basel)* 2016;6:3, DOI: 10.3390/life6010003.
- Zhang J, Olsen GJ. Messenger RNA processing in *Methanocaldococcus (Methanococcus) jannaschii*. *RNA* 2009;15:1909–16.
- Zhang J, Rouillon C, Kerou M et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 2012;45:303–13.
- Zhang Y, Li H. Structure determination of a truncated dimeric splicing endonuclease in pseudo-face-centered space group P2(1)2(1)2. *Acta Crystallographica. Sect. D, Biological Crystallography* 2004;60:447–52.
- Zheng X, Feng N, Li D et al. New molecular insights into an archaeal RNase J reveal a conserved processive exoribonucleolysis mechanism of the RNase J family. *Mol Microbiol* 2017;106:351–66.
- Zhu L, Deutscher MP. tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. *EMBO J* 1987;6:2473–7.

## Projet de thèse

Les ribonucléases de types  $\beta$ -CASP ont été identifiées comme des enzymes ubiquitaires chez les Archées, cependant leurs rôles physiologiques dans le métabolisme de l'ARN ne sont pas encore définis. Dans mon équipe, nous cherchons à comprendre le rôle de la ribonucléase 5'-3' aRNase J dans les voies de processing de l'ARN. Le réseau d'interactions protéine-protéine établis d'aRNase J a permis de confirmer sa place au sein du métabolisme de l'ARN et a notamment mis en évidence la présence d'une protéine annotée comme étant une hélicase à ARN, ASH-Ski2 et la présence des sous unités de l'Exosome.

Dans ce contexte, mes travaux ont consisté tout d'abord à répertorier les familles de RNases (structures et fonction) connues chez les Archées pour la rédaction d'une revue (INTRODUCTION, PARTIE III- [Publication 1](#)). Ils ont permis d'élargir le réseau d'interactions d'aRNase J et ASH-Ski2 en établissant les interactomes des sous unités de l'exosome Rrp41 et Csl4 (RESULTATS, PARTIE II – Résultats non publiés) et de valider et caractériser les interactions aRNase J/ASH-Ski2, aRNase J/Exosome (RESULTATS, PARTIE III – Résultats publiés – [Publication 2](#)) et aRNase J, ASH-Ski2 /Ribosome (RESULTATS, PARTIE III – Résultats non publiés) chez les Thermococcales. Nous avons ensuite décrit l'activité enzymatique de l'hélicase ASH-Ski2 (RESULTATS, PARTIE IV – [Article en préparation](#)) et finalement initier la caractérisation biophysique des protéines aRNase J et ASH-Ski2 dans le but de résoudre leur structure (RESULTATS, PARTIE V).

Les résultats obtenus, par des approches de protéomiques, biochimiques et enzymatiques, montrent que l'aRNase J est engagée spécifiquement avec l'hélicase ASH-Ski2 et l'exosome à ARN à proximité du ribosome (Résultats publiés – [Publication 1](#) - Phung et al 2020) et que ASH-Ski2 est une hélicase à ARN, avec une polarité 3'- 5' dont l'activité prépondérante est l'assemblage de brins (activité annealing) (Rédaction en cours : [Batista et al, en préparation](#)). L'ensemble des données obtenues qui sont confortées par des approches phylogénomiques, offrent de nouvelles perspectives quant à l'identification de nouvelles machineries moléculaires impliquées dans le métabolisme de l'ARN chez les Archées.



# MATERIELS ET METHODES





## I. Construction des vecteurs d'expression

Tous les vecteurs utilisés (construits ou déjà disponibles) dans ces travaux sont listés dans le

**tableau 4**. Les oligos qui ont permis de construire les vecteurs sont listés dans le **tableau 5**.

GENES	PROTEINES	pet15b	pet11b	PetDuet
pab1751	aRNase J		✓	✓
	aRNase J Δ1		✓	
	aRNase J Δ2		✓	
pab2313	ASH-Ski2	✓	✓	✓
	ASH-Ski2 ΔN	✓	✓	
	ASH-Ski2 ΔC	✓	✓	
	ASH-Ski2-C124A	✓	✓	
	ASH-Ski2-C98A	✓		
	ASH-Ski2-C98A-C124A	✓		
	ASH-Ski2 DomN	✓	✓	
	ASH-Ski2 DomN C124A	✓	✓	
	ASH-Ski2 DomN C98A	✓		
pab0592	Hel308	✓	✓	
pab2314	Csl4	✓	✓	
	Csl4 ΔN	✓		
	Csl4 ΔC	✓		
	Csl4-F121A	✓		
	Csl4-F129A	✓		
pab0420	Rrp41	✓	✓	

**Tableau 4. Vecteurs utilisés dans cette étude.** En noir, les vecteurs disponibles dans l'équipe, en rose ceux construits pendant ces travaux.

OLIGONUCLEOTIDES	SEQUENCES
T7 Prom	TAATACGACTCACTATAGGGGA
T7 Ter	GCTAGTTATTGCTCAGCGGT
For pET15b	CATATGGCTGCCGCGCGC
Rev pET15b / pET11b	GGATCCGGCTGCTAACAAAGCC
For pET11b	CATATGTATATCTCTTCTTAAAGTT
For pETDuet	CTTATACTTAATAATACTAAGATGG
Rev pETDuet	GTAAAGAAACCGCTGCTGCGAAATT
For aRNaseJ pET11b	GGAGATATACATATGTGGGAGGAGATAAACATGATCA
Rev aRNaseJ pET11b	TTAGCAGCCGGATCCTCATCCCTCCAATGAGCCAG
For aRNase J petDuet	TTAGTTAAGTATAAGAAGGAGATATACATATGTG
Rev aRNase J petDuet	CAGCGTTTCTTTACTTAGCAGCCGGATCCTCATC
For ASH-Ski2 pET15b	CGCGGCAGCCATATGCTATTCTGTTATTCGCCAGGGAGG
Rev ASH-Ski2 pET15b	AGCCGGATCCTCGAGTTATGGTTTTCTTCTCTCCTCACCTT
For ASH-Ski2 pET11b	GGAGATATACATATGATGCTATTCGTTATTCGCCAG
Rev ASH-Ski2 pET11b	TTAGCAGCCGGATCCTTATGGTTTTCTTCTCTCCTCA
For ASH-Ski2 petDuet	ATCTCGATCCCGCAAATTAAT
Rev ASH-Ski2 petDuet	CTCGAATTCGGATCCTTATGGTTTTCTTCTC
For ASH-Ski2ΔN-ter (194- 855) pET15b	CCGCGCGGCAGCCATATGACGATTGACGAGCTGGAT
Rev ASH-Ski2ΔC-ter (1-573) pET15b	AGCCGGATCCTCGAGTTAAGAGGTTAGCAATTTCAAGGCCAC
Rev DomN (1-193) ASH-Ski2 pET15b	TTAGCAGCCGGATCCTTA TCTCTCAACTTAATACGCTTC

For Hel308 pET15b	CGCGGCAGCCATATGATGAAAGTTGGAGAGCTAAACG
RevHel308 pET15b	TTAGCAGCCGGATCCTTATGGGTTTCAGGAAATAGTCC
For Csl4 pET15b	CGCGGCAGCCATATGTTGGAGGAAGGTGAGGAGAG
Rev Csl4 pET15b	TTAGCAGCCGGATCC TCATAGCTTCACCTTCCTGTA
For Csl4 pET11b	GGAGATATACATATGTTGGAGGAAGGTGAGGAGAG
For Csl4 $\Delta$ N-ter pET15b	AGGATCCTTTGTCGTCAACTGAATC
Rev Csl4 $\Delta$ C-ter pET15b	CCTCCAATACCTAAGAAAGGGGA
For Rrp41 pET15b	CGCGGCAGCCATATGATGATGGAGAAGCCAGAGGG
Rev Rrp41 pET15b	TTAGCAGCCGGATCCTCACTCACTTCCCTCAACCT
For aRNase J $\Delta$ 1 pET15b	GATTAGAACCCTGTCGAGCCTTATTC
Rev aRNase J $\Delta$ 1 pET15b	GACCTTAGGAAGCTCGGTGCAATAC
For aRNase J $\Delta$ 2 pET15b	TAAGACCCTAGATCCCTTCATC
Rev aRNase J $\Delta$ 2 pET15b	CTAAAGGAAGTTTCCCAAGCTA
For ASH-Ski2 C124A	GAGAGTACATAGCCGAGAGATGTG
Rev ASH-Ski2 C124A	CACATCTCTCGGCTATGTACTCTC
For ASH-Ski2 C98A	AGGCCAGTCTGGCCAGGCTCTGT
Rev ASH-Ski2 C98A	ACAGAGCCTGGCCAGACTGGCCT
For Csl4 F121A	GTAAAGGACGGCGCCGTTGAGGATTTAAGGAA
Rev Csl4 F121A	TAAATCCTCAACGGCGCCGTCCTTAACCTGAG

**Tableau 5. Liste des oligonucléotides utilisés pour les clonages des vecteurs d'expression.**

## 1. Clonage avec le kit InFusion® (Ozyme)

Les vecteurs sont amplifiés par PCR inverse avec l'enzyme Prime Star max (Ozyme) en présence de 10ng d'ADN plasmidique et de 0.3µM des couples d'oligonucléotides appropriés (Programme thermocycler : [98°C (15') ; 56°C (15'') ; 72°C (1'30'')]\*35cycles). La réaction PCR est incubée à 37°C en présence de 20 unités de DpnI (ThermoScientific) pendant 1h afin d'éliminer la matrice. Les gènes codant pour les protéines ASH-Ski2, aRNase J, Hel308, Rrp41, et Csl4 sont amplifiés par PCR à partir d'ADN génomique de *Pyrococcus abyssi* avec 1 unité de Phusion (High-Fidelity DNA polymerase de Finnzymes), 0,2mM de dNTPs et 0.5µM des couples d'oligonucléotides (Programme thermocycleur: [98°C (30'') ; 58°C (30'') ; 72°C (1')]\*35cycles). Pour permettre l'assemblage des produits PCR dans la réaction InFusion®, les oligonucléotides permettant l'amplification des séquences codantes sont synthétisés pour intégrer, à leurs extrémités 5', 15 nucléotides complémentaires aux extrémités des vecteurs. Les réactions InFusion® sont réalisées comme recommandé par le fournisseur avec des quantités équimolaires de produits PCR. La totalité de la réaction est transformée par choc thermique dans 50µL de bactéries Stellar® chimio-compétentes (Ozyme) et les transformants sont sélectionnés en présence d'ampicilline (100µg/ml) après 1h d'expression phénotypique à 37°C. L'intégration des inserts dans les vecteurs est vérifiée par PCR sur colonies avec les oligonucléotides T7 Prom et T7 Ter. Les inserts sont vérifiés par séquençage avec le même couple d'oligonucléotides (Société MWG Eurofin).

## 2. Clonage par PCR inverse

La méthode utilisée pour construire les vecteurs d'expression contenant les gènes ASH-Ski2ΔN, ASH-Ski2ΔC, ASH-Ski2ΔNC, ASH-Ski2 DomN, aRNase J Δ1, aRNase J Δ2, Csl4ΔN et Csl4ΔC consiste à exclure une partie du gène WT par PCR inverse en utilisant des oligonucléotides phosphorylés en 5' de part et d'autre du domaine à éliminer. Les vecteurs sont amplifiés par PCR inverse avec l'enzyme Prime Star max (Ozyme) en présence de 10ng d'ADN plasmidique et de 0.3µM des couples d'oligonucléotides phosphorylés appropriés (Programme thermocycler : [98°C (15') ; 56°C (15'') ; 72°C (1'30'')]\*35cycles). La réaction PCR est incubée à 37°C en présence de 20 unités de DpnI (ThermoScientific) pendant 1h afin d'éliminer la matrice. Les vecteurs sont circularisés en utilisant 30 unités de T4 DNA ligase (BioLabs) pendant 1h à 20°C. La totalité de la réaction est transformée par choc électrique

dans 50µL de bactéries DH5α électro-compétentes et les transformants sont sélectionnés en présence d'ampicilline (100µg/ml) après 1h d'expression phénotypique à 37°C. Les inserts mutés sont vérifiés par séquençage avec le couple d'oligonucléotides T7 Prom et T7 Ter (Société MWG Eurofin).

### 3. Mutagenèse dirigée

Les vecteurs pet11b et pet15b exprimant les protéines recombinantes ASH-Ski2 C98A, ASH-Ski2 C124A, ASH-Ski2 C98A-C124A, ASH-Ski2 DomN C98A, ASH-Ski2 DomN C124A, Csl4 F121A et Csl4 F129A sont construits par mutagenèse dirigée avec le kit QuikChangeXL® (Agilent) comme recommandé par le fournisseur. 10ng de plasmide pet11b ou pet15b exprimant ASH-Ski2, ASH-Ski2-DomN ou Csl4 sont mis en présence de 2,5 unités de Pfu Ultra HF et de 125ng des couples d'oligonucléotides contenant la mutation ponctuelle à introduire (Programme thermocycleur : [95°C (50'') ; 60°C (50'') ; 68°C (12')] \*18cycles). La réaction PCR est incubée à 37°C en présence de 20 unités de DpnI (ThermoScientific) pendant 1h afin d'éliminer la matrice. La totalité de la réaction est transformée par choc thermique dans 45µL de bactérie de XL10-Gold® chimio-compétentes (Agilent) et les transformants sont sélectionnés en présence d'ampicilline (100µg/ml). L'intégration des mutations est confirmée par séquençage avec les oligonucléotides T7 Prom et T7 Ter (Société MWG Eurofin).

## II. Production des protéines recombinantes

Les tampons sont listés dans le **tableau 6**.

<b>Tampons</b>	<b>Composition</b>
LB	5g/L yeast extract; 10g/L NaCl; 10g/L Tryptone; 20mg/mL Thymine pH 7
PBS 1X	137mM NaCl; 2,7mM KCl; 10mM Na <sub>2</sub> HPO <sub>4</sub> ; 1,8mM KH <sub>2</sub> PO <sub>4</sub>
Tampon NA	50mM Tris-HCl pH 7.5; 150mM NaCl; 10% Glycerol
Tampon IM	20mM NaP pH 7.4; 300mM NaCl; 10% Glycerol
Tampon GF	HEPES pH 7.8; 300 ou 500mM NaCl; 10% Glycérol
Tampon COP	20mM HEPES 7.4, NaCl 300mM, 5mM MgCl <sub>2</sub> , 0,1mM ZnCl <sub>2</sub> , 0,2mM ATP
Tampon ATPase	50mM Hepes pH 7.5; 50mM KCl; 5mM MgCl <sub>2</sub> ; 2mM DTT
Tampon SSC 1X	150mM NaCl; 15mM NaCl
Tampon Hélicase 1X	25mM Tris-HCl pH 8, 50mM NaAc, 5mM MgCl <sub>2</sub> , 2.5mM β-Mercaptoethanol
Tampon STOP	0.1% SDS, 8mM EDTA, 0.1mg/ml protéinase K, 4% glycérol, 0.02% Bromophénol bleu

**Tableau 6. Liste et composition des tampons utilisés dans cette étude.**

## 1. Surexpression des protéines recombinantes

Les vecteurs pETs sont transformés dans la souche d'expression *E.coli* BL21 (DE3) codon+ par électroporation. Les transformants sont sélectionnés en présence d'ampicilline (100µg/ml) et de chloramphénicol (30µg/ml). Des cultures de nuit, diluées au 1/100<sup>e</sup> dans du milieu LB (Luria Berthani, 5g/L d'extrait de levure, 10g/L NaCl, 10g/L Tryptone, 20mg/mL Thymine pH 7) contenant de l'ampicilline et du chloramphénicol, sont cultivées à 37°C sous agitation (180rpm) jusqu'à atteindre une DO = 0,6 à 600nm. La production de protéines est induite par l'ajout de 0,1mM d'IPTG pendant 3h à 30°C. Les cellules sont culotées par centrifugation (15min, 6000rpm à 4°C (rotor JA-14)). Après un lavage avec du PBS 1X (137mM NaCl, 2.7mM KCl, 10mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8mM KH<sub>2</sub>PO<sub>4</sub>), les culots de cellules obtenus après une nouvelle centrifugation (15min, 5000rpm à 4°C (rotor J.S 5.3)) sont congelés à -20°C pour conservation.

## 2. Lyse et préparation de l'extrait brut

Les culots cellulaires surexprimant les protéines d'intérêt sont resuspendus dans 10mL de tampon NA (50mM Tris HCl pH 7.5, 150mM NaCl, 10% glycérol) ou IA (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol) supplémenté de lysosyme (1mg/ml) et d'une demi pastille d'inhibiteur de protéases (cOmplete™ Roche) et subissent trois cycles de [5x10sec] de sonication (50% cycle) (Bioblock scientific, Vibra Cell). Un traitement RNase A (20ng/mL), RNase T1 (1000 unités/µl) et DNase I (20ng/mL) est réalisé à 37°C pendant 30min. Après un chauffage de 70°C pendant 20min, la clarification de l'extrait brut est effectuée par centrifugation (20 min, 15 000g à 4°C (rotor JA-20)). Le surnageant obtenu est ensuite filtré (0.22µm). La suite de la purification a été réalisée en FPLC (Fast Protein Liquid Chromatography) sur un système Akta à température ambiante.

## 3. Purification des protéines sans étiquette : *Pab-aRNase J*, *Pab-Hel308*, *Pab-ASH-Ski2* et ses variants

L'extrait cellulaire est dilué à une concentration finale de 150mM en NaCl et est injecté sur une colonne Heparin de 1ml (HiTrap™ Heparin HP GE Healthcare) préalablement équilibrée dans le tampon NA (50mM Tris HCl pH 7.5, 150mM NaCl, 10% glycérol) et placée sur un système Akta FPLC (Fast Protein Liquid Chromatography) à température ambiante. Après une étape de lavage à 200mM de NaCl, l'élution est réalisée avec un gradient linéaire de 300mM



à 1M de NaCl. Ces étapes sont communes à la purification de *Pab*-aRNase J, *Pab*-Hel308, *Pab*-ASH-Ski2 et ses variants. Une étape de purification supplémentaire est nécessaire pour *Pab*-Hel308, *Pab*-ASH-Ski2 et ses variants. Les fractions des pics correspondant aux protéines d'intérêts sont diluées (Cf 150mM NaCl) et injectées sur une colonne Q-HP (GE Healthcare) pour Hel308 ou SP-HP (GE Healthcare) pour *Pab*-ASH-Ski2 et ses variants. Le même programme est appliqué. La pureté des fractions protéiques est visualisée après migration sur un gel SDS-PAGE 4-15% (BioRad) et révélations au bleu de coomassie (Quick Coomassie, CliniSciences).

Dans le cas où les protéines purifiées sont destinées à des essais de cristallisation : les fractions protéiques (3 à 5ml d'échantillon) sont déposées sur une colonne HiLoad 16/60 Superdex 200 PG (Limites de fractionnement : 10 000 -600 000Da) préalablement équilibrée dans un tampon GF (50mM HEPES pH 6.8, 300 à 500mM NaCl, 10% glycérol). Les fractions contenant la protéine pure sont concentrées par centrifugation différentielle sur membrane semi perméable (entre 3 et 10min à 5000g) à l'aide du système vivaspin20 dont la membrane (seuil de coupure (MWCO) =20 000 ou 50 000Da) a été préalablement rincée avec du tampon GF. La concentration a été suivie par analyse de l'absorbance à 280nm, le rapport DO260/DO280 a été vérifié pour estimer le degré de pureté (0.6 atteste d'une solution de protéine pure, dépourvue d'acides nucléiques).

#### **4. Purification des protéines avec étiquette Histidine: *Pab*-aRNase J-(His)<sub>6</sub>, (His)<sub>6</sub>-*Pab*-Rrp41, (His)<sub>6</sub>-*Pab*-Csl4, (His)<sub>6</sub>-*Pab*-ASH-Ski2 et ses variants**

L'extrait cellulaire est injecté sur une colonne de Nickel de 1ml (HisTrap™ GE Healthcare) préalablement équilibrée dans le tampon IM (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol) et disposée sur un système Akta FPLC à température ambiante. Après un lavage à 10 et 20mM d'Imidazole, l'élution est réalisée avec un gradient linéaire d'Imidazole (50 à 500mM). Les fractions contenant les protéines d'intérêt sont ensuite diluées à 150mM NaCl et déposées sur une colonne SP-HP (GE Healthcare) pour Csl4 et Heparin (GE Healthcare) pour Rrp41, ASH-Ski2 et ses variants. La pureté des fractions protéiques est visualisée après migration sur un gel SDS-PAGE 4-15% (BioRad) et révélation au bleu de Coomassie (Quick Coomassie, CliniSciences).

## 5. Quantification des protéines

La concentration des protéines est déterminée à partir d'une gamme BSA (0,16 à 2,5mg/ml) sur un gel SDS-PAGE 4-15% (Biorad) coloré au Sypro Orange. Les protéines ont été visualisées sur un Bioimager et quantifiées sur le logiciel MultiGauge.

## 6. Conservation des protéines

Les protéines sont dialysées (Slide-A-Lyser®, Thermo Scientific) dans le tampon GF (50mM HEPES pH 7.8, 300 ou 500mM NaCl, 10% Glycérol) et conservées à 16°C ou aliquotées et congelées à -80°C.

## III. Pull down pour établir les réseaux d'interactions des protéines

### (His)<sub>6</sub>-Pab-Rrp41 et (His)<sub>6</sub>-Pab-Csl4

#### 1. Pull down

Cette méthode consiste à incuber une protéine recombinante avec une étiquette Histidine fixée sur des billes de Nickel avec un extrait cellulaire de *P. abyssi*. Les pulls down sont réalisés en triplicat avec des extraits cellulaires de *P. abyssi* cultivés en phase exponentielle de croissance en bioréacteurs dans des conditions physiologiques (95° C, pH 6.5, anaérobie) au Laboratoire Ifremer (D. Flament ; Brest). Les culots de cellules de *P. abyssi* sont resuspendus dans du PBS 1X (poids/volume 1/3), supplémenté de 300mM de NaCl et d'un mix d'inhibiteurs de protéase (cOmplete™ Roche). Après sonication (VibraCell Bioblock Scientific) l'extrait est clarifié par centrifugation (10 000g, 60min, 4°C) et supplémenté de 10mM MgCl<sub>2</sub>. Les protéines purifiées Csl4 et Rrp41 avec une étiquette Histidine sont utilisées comme appât. 20µg de protéines appâts sont immobilisées dans 0,6mg de billes magnétiques de Cobalt (Dynabeads, Invitrogen). Le complexe billes-appât est incubé avec 2mg d'extrait de *P. abyssi* sous rotation durant une nuit à température ambiante. Pour éliminer les interactions médiées par les molécules d'ADN ou ARN, cette étape est réalisée en présence d'un mixe DNase/RNase (10µg/ml de RNase A et DNase I). Les complexes de protéines formés *in vitro* sont séparés sur un portoir magnétique et lavés avec du PBS 1X (3\*200µl). Un control est aussi réalisé sous les mêmes conditions en utilisant des billes de cobalt sans appât. Les protéines sont éluées dans 25µl de tampon XT (Biorad) contenant 2µl d'agent réducteur 20X pendant 10min à 95°C.

## 2. Analyse protéomique

Les analyses de spectrométrie de masses sont réalisées à Pris Sud-Ouest PAPPSO (<http://papso.inra.fr>). Après une courte migration sur gel SDS-PAGE (Criterion XT - Precast gel – BioRad), les bandes de protéines sont découpées et envoyées à la plateforme PAPPSO. Le traitement des bandes est décrit dans la publication (Branson and Freitas 2016). Les données sont traitées pour identifier les signaux des interactions spécifiques. La globalité des spectres spécifiques est normalisée entre les séries de répliques. Un cut-off de 2 spectres normalisés est utilisé comme signal MS minimum pour la validation des hits du réseau. Les spectres normalisés sont ensuite moyennés entre répliques et références versus control pour calculer le nombre de « spectres référencés ». Le calcul du score de « l'index spécifique » est le ratio des spectres normalisés versus les contrôles. L'«index de spécificité » varie de 0 à 1 (avec un maximum de seuil = 1), plus la valeur est proche de 0 meilleure est la spécificité. Les réseaux d'interactions sont construits sur le logiciel Cytoscape.

## IV. Test d'interactions *in vitro*

### 1. Co-purification de mélange de culots protéiques recombinés sur colonne de Nickel en système FPLC

Deux culots cellulaires de 400ml sont repris dans 10mL de tampon IA (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol) supplémenté de lysosyme (1mg/ml) et d'une demi pastille d'inhibiteur de protéases (cOmplete™ Roche) et sont conservés sur glace pendant toutes les étapes. Trois cycles de [5x10sec] de sonication (50% cycle) sont ensuite réalisés. Après deux cycles de 20min de chauffage à 70°C, la clarification des extraits bruts est effectuée par centrifugation (20 min, 15 000g, 4°C). Dans certains cas, un traitement à la RNase A (20ng/mL), RNase T1 (1000 unités) et DNase I (20ng/mL) a été fait à 37°C pendant 30min entre les 2 étapes de chauffages. Le surnageant obtenu est filtré (0.22µm). La purification par affinité est réalisée en FPLC (Fast Protein Liquid Chromatography) sur un système Akta à 4°C, en déposant l'extrait cellulaire filtré sur une colonne de 1ml de Nickel (HisTrap™ GE Healthcare) préalablement équilibrée dans le tampon IA (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol). Les éluions sont obtenues avec un gradient linéaire d'imidazole (50 à 500mM). Les fractions protéiques

sont séparées sur un gel SDS-PAGE 4-15% (BioRad) puis colorée au bleu de Coomassie (Quick Coomassie, CliniSciences) ou analysées en Western Blot.

## **2. Co-purification de mélange de protéines recombinantes purifiées sur colonne de Nickel en système FPLC**

Un mélange contenant les protéines recombinantes pures *Pab*-aRNase J (9 $\mu$ M) et (His)<sub>6</sub>-*Pab*-ASH-Ski2 DomN (3 $\mu$ M) ont été chauffées 20min à 70°C dans un tampon COP (20mM HEPES 7.4, NaCl 300mM, 5mM MgCl<sub>2</sub>, 0,1mM ZnCl<sub>2</sub>, 0,2mM ATP). Le mélange est injecté sur une colonne de Nickel (HisTrap™ GE Healthcare) préalablement équilibrée dans le tampon IA (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol) et placée sur un système Akta FPLC à température ambiante. Les éluions sont obtenues avec un gradient linaires d'Imidazole (50 à 500mM). Les fractions protéiques sont séparées sur un gel SDS-PAGE 4-15% (BioRad) puis colorée au bleu de Coomassie (Quick Coomassie, CliniSciences) ou analysées en Western Blot.

## **3. Co-purification de mélange de protéines purifiées sur bille de Nickel**

Un mélange de 200 $\mu$ l contenant la protéine recombinante purifiée (His)<sub>6</sub>-*Pab*-ASH-Ski2 (150pmol), la sous unité 30S purifiée du ribosome de *P.abysyi* (30pmol) (La sous unité 30S a été purifié selon la méthode expliquée dans (Coureux et al. 2016) et un ARNm 5'-GGCAAGGAGGUAAAAUGUUCAA-3' (150pmol) dans du tampon NI (10mM HEPES 7.5; 10mM MgAc, 150mM NH<sub>4</sub>Cl, 3mM  $\beta$ -Mercaptoethanol) sont chauffés 5min à 50°C puis incubés dans une colonne en plastique (ThermoScientific™) avec 25 $\mu$ l de résine de Nickel (Ni-NTA ThermoScientific™) 15min à température ambiante. Après lavage un avec 250 $\mu$ l de tampon NI puis 250 $\mu$ l de tampon NI supplémenté de 30mM d'Imidazole, l'éluion 1 et 2 sont réalisées avec respectivement 35 et 50 $\mu$ l de tampon NI supplémenté de 250mM d'Imidazole. Les protéines sont visualisées après séparation dans un gel SDS-PAGE 16% et coloration au bleu de Coomassie.

## **4. Mélange de protéines purifiées en système HPLC**

150pmol de *Pab*-aRNase J ou *Pab*-ASH-Ski2, 30pmol de la sous unité 30S du ribosome de *P.abysyi* et 150pmol d'un ARNm 5'-GGCAAGGAGGUAAAAUGUUCAA-3' sont chauffées 5min à 50°C. Les injections sont réalisées sur une colonne d'exclusion de taille (Agilent Bio SEC- 5)

placée sur un système HPLC (High performance liquid chromatography) à température ambiante et équilibrée dans un tampon (10mM HEPES 7.5, 10mM MgAc, 6mM  $\beta$ -Mercaptoethanol, 150 ou 300mM  $\text{NH}_4\text{Cl}$ ). Les fractions des pics correspondants aux protéines sont récupérées puis concentrées avant d'être déposées sur un gel SDS-PAGE 16% et coloré au bleu de Coomassie.

## V. Tests enzymatiques

Les substrats sont décrits dans le **tableau 7**.

Nom	Séquence	Taille (Nt)
S1	5'-ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT-3'	50
US1	5'-AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU-3'	50
5RS1	5'-GGACATGCTGTCTAGAGACTATCGAT-3'	26
U5RS1	5'-GGACAUGCUGUCUAGAGACUAUCGAU-3'	26
3RS1	5'-ACGCTGCCGAATTCTGGCTTGCTAGG-3'	26
U3RS1	5'-ACGCUGCCGAAUUCUGGCUUGCUAGG-3'	26
S1/5RS1	5'-ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT-3' 3'-TAGCTATCAGAGATCTGTCGTACAGG-5'	
S1/U5RS1	5'-ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT-3' 3'-UAGCUAUCAGAGAUCUGUCGUACAGG-5'	
US1/U5RS1	5'-AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU-3' 3'-UAGCUAUCAGAGAUCUGUCGUACAGG-5'	

**Tableau 7. Substrats utilisés dans ces travaux.** Leur séquence et leur taille sont indiqués.

## 1. Radiomarquage des substrats

Les substrats S1, 5RS1, U5RS1 et U3RS1 sont phosphorylés en 5' avec 50µCi d'ATP [ $\gamma^{32}$ ] en utilisant 5 unités de T4 polynucléotide kinase (ThermoScientific) durant 1h à 37°C. Le substrat radiomarqué US1 est obtenu par transcription *in vitro* de l'ADN S1 additionné d'un promoteur T7 en utilisant le kit MEGAscript (Ambion) selon les recommandations du fournisseur et en présence de 50µCi d'UTP [ $\alpha^{32}$ ]. Les duplexes US1/U5RS1 et US1/U3RS1 sont obtenus par incubation des simples brins à 90°C pendant 5min dans un tampon SSC 1X (150mM NaCl, 15mM NaCl) suivi d'une étape de refroidissement lente à température ambiante. Les duplexes sont dessalés sur une colonne (Micro-bio Spin, Biorad).

## 2. Test ATPase

250nM de la protéine recombinante pure est pré-incubée 10min à 60°C en présence ou non d'un extrait d'ARN totaux de *P.abysyi* dans le tampon ATPase (50mM Hepes pH 7.5, 50mM KCl, 5mM MgCl<sub>2</sub>, 2mM DTT). La réaction est initiée par l'ajout d'ATP (2mM d'ATP et 0.75µCi [ $\gamma^{32}$ ] ATP (6000 Ci/mmol)) et est incubée à 65°C. 2µl de réaction sont collectés à 0, 5, 10, 30 et 60min et déposés sur une membrane TLC (Thin Layer Chromatography). Après migration dans un tampon KH<sub>2</sub>PO<sub>4</sub>, la membrane TLC est séchée et placée toute la nuit dans une cassette en présence d'un écran photo-stimulable (BAS-MS Imaging Plates-Fujifilm). Les écrans sont scannés par un phospho-imager (Typhoon Trio-Amersham Biosciences) avec une résolution de 100microns. La quantification de l'ATP hydrolysé a été réalisée avec le logiciel MultiGauge.

## 3. Test d'affinité pour les acides nucléiques

Une gamme de dilution de la protéine est réalisée (de 0 à 350nM) et est pré-incubée à 60°C pendant 10min. Un mixe composé de 0.5nM de substrats marqués et de tampon hélicase 1X (25mM Tris-HCl pH 8, 50mM NaAc, 5mM MgCl<sub>2</sub>, 2.5mM  $\beta$ -Mercaptoethanol) est ajouté à la solution protéique dans un volume final de 50µl. Après 15min d'incubation à 30°C, la réaction est stoppée dans la glace. Les réactions sont déposées et filtrées à travers une membrane de nitrocellulose (traitée au préalable avec 0.5M de KOH puis dans du tampon hélicase 1X) et une membrane de Nylon (traitée au préalable avec du tampon hélicase 1X) dans un Slot-Blot. Les membranes sont séchées et placées dans une cassette en présence d'un écran phospho-stimulable (BAS-MS Imaging Plates-Fujifilm) pendant 4h. Les écrans sont scannés par un



phospho-imager (Typhoon Trio-Amersham Biosciences) avec une résolution de 100microns. La quantification des dépôts est faite avec le logiciel MultiGauge. Les courbes et l'analyse des Kd sont réalisés à l'aide du logiciel GraphPad Prism 7.

## **4. Test d'appariement et déroulement des brins d'acides nucléiques**

### **4.1 Déroulement de duplexes d'acides nucléiques (Unwinding)**

Une solution Starting contenant de l'ATP (5mM), MgCl<sub>2</sub> (5mM) et un oligo trap froid (U5RS1 ou U3RS1, 200nM) et une solution de Réaction composée du duplexe ARN/ARN radiomarqué (US1/U5RS1\* ou US1/U3RS1\*, 5nM), de la protéine (250nM) et de tampon Hélicase 1X (25mM Tris-HCl pH 8, 50mM NaAc, 5mM MgCl<sub>2</sub>, 2.5mM β-Mercaptoethanol) sont pré-incubés 5min à 65°C. La réaction unwinding est déclenchée par l'ajout de la solution Starting à la solution de Réaction. Une cinétique est réalisée à 65°C en prélevant 5µl du mélange à 0, 2, 5, 10 et 20min puis ajoutée à 1.25µl de tampon STOP 1X (0.1% SDS, 8mM EDTA, 0.1mg/ml protéinase K, 4% glycérol, 0.02% Bromophénol bleu). Un contrôle sans protéine est réalisé pour chaque duplexe. Les substrats simples brins sont séparés des duplexes par électrophorèse dans un gel acrylamide 8% contenant 0.1% de SDS à 160V pendant 1h. Le gel est séché et placé dans une cassette contenant un écran phospho-stimulable (BAS-MS Imaging Plates-Fujifilm) o/n. Les écrans sont scannés par un phospho-imager (Typhoon Trio-Amersham Biosciences) avec une résolution de 200microns. La quantification des bandes des substrats simples brins et des duplexes est faite avec le logiciel MultiGauge. Les courbes sont réalisées avec le logiciel GraphPad Prism 7.

### **4.2 Appariement de brin d'acides nucléiques (Annealing)**

Une solution de Réaction contenant le brin d'ARN (US1) ou ADN (S1) froid (5nM), du MgCl<sub>2</sub> (5mM) et du tampon Hélicase 1X est pré-incubée séparément de la protéine (3µM) pendant 5min à 65°C. Le brin d'ADN (5RS1\*) ou ARN (U5RS1\*) radiomarqué (5nM) est ajouté à la solution réaction avec la protéine. Une cinétique est réalisée à 65°C en prélevant 5µl du mélange à différents temps puis ajoutée à 1.25µl de tampon STOP 1X (0.1% SDS, 8mM EDTA, 0.1mg/ml protéinase K, 4% glycérol, 0.02% Bromophénol bleu). Un contrôle sans protéine est réalisé pour chaque duplexe. Les substrats simples brins sont séparés des duplexes par électrophorèse dans un gel acrylamide 8% contenant 0.1% de SDS à 160V pendant 1h. Le gel

est séché et placé dans une cassette contenant un écran phospho-stimulable (BAS-MS Imaging Plates-Fujifilm) o/n. Les écrans sont scannés par un phospho-imager (Typhoon Trio-Amersham Biosciences) avec une résolution de 200microns. La quantification des bandes des substrats simples brins et des duplexes est faite avec le logiciel MultiGauge. Les courbes sont réalisées avec le logiciel GraphPad Prism 7.

## **VI. Tests Biophysiques et structurales**

### **1. Diffusion dynamique de la lumière en solution (Dynamic Light Scattering)**

Les solutions de protéines pures sont centrifugées (16°C, 5 à 10min à 10000g) afin d'éliminer les agrégats. 8µl du surnageant est déposé dans une cuvette (cuvettes jetables de 4µl – référenceWyatt- ou cuvette Quartz de 1 µl qui permettent de faire des mesures en variant la température de 15°C à 100°C), elles-mêmes disposées ensuite dans l'appareil de DLS (NanoStar, Wyatt Technology, 663nm). Pour chaque protéine, 10 acquisitions de 5sec sont réalisées à 20.5°C après avoir fait une mesure du tampon comme contrôle (absence de signal de diffusion). Les fluctuations d'intensité mesurées et la fonction d'autocorrélation sont analysées par le logiciel DYNAMICS V7 et permet d'extraire différents paramètres : le coefficient de diffusion, le pourcentage de polydispersité et le rayon hydrodynamique de l'espèce en solution.

### **2. SEC-MALS (Size Exclusion Chromatography – Multi Angle Light Scattering)**

Les solutions de protéines pures sont centrifugées (16°C, 5 à 10min à 10000g) afin d'éliminer les agrégats. 10 à 160µl de surnageant (1.5mg/ml à 5mg/ml) sont analysés sur une colonne Superdex Increase S200 5/150 (GE Healthcare) par diffusion statique de la lumière. La colonne est équilibrée dans le tampon GF (filtré 0.1 µm) sur un système chromatographique Infinity 1260 LC (Agilent Technology). Les données sont collectées sur un système MALS (Wyatt Technology) à 8 angles (DAWN HELEOS-II) couplé à un réfractomètre (Optilab T-rEX). Les échantillons sont injectés sur la colonne et la séparation réalisée à un débit de 0.35 ml/min à 18°C. Les résultats sont analysés avec le logiciel ASTRA 6.1 (Wyatt Technology) et permet la mesure de la masse molaire, du rayon hydrodynamique et ainsi de connaître l'état d'oligomérisation de la protéine mais aussi de l'homogénéité et de l'état d'agrégation de l'échantillon.

### 3. Fluorimétrie à balayage différentiel (Differential Scanning Fluorimetry)

La Fluorimétrie à balayage différentiel est utilisée pour caractériser la stabilité thermique des protéines et déterminer les conditions dans lesquelles les protéines sont les plus stables. Un criblé préliminaire est réalisé pour les protéines aRNase J et ASH-Ski2. Les mesures sont réalisées en triplicats en plaque 96 puits (Hard Shell, Bio-Rad) à 3 concentrations en protéines (2, 5 et 10 $\mu$ M), 3 concentrations en NaCl (200, 400 et 800mM) et 3 concentrations en SYPRO® Orange (10 $\times$ ) (Invitrogen) (2X, 5X et 10X) avec un volume final de 20 $\mu$ l par puits. Les transitions thermiques sont monitorées avec un système CFX96 (Real-time PCR, Bio-Rad) et les mélanges soumis à un gradient de température de 20 à 100°C par incrément de 0.3°C. Les courbes de dénaturation des protéines sont obtenues sur le logiciel CFX Manager (v 3.1) et les températures de fusion ( $T_m$ ) identifiées aux points d'inflexion des courbes de fluorescence  $f(T)$ . Un criblé commercial (Fundament, Jena Bioscience) est ensuite réalisé pour aRNase J. 15 $\mu$ l de chaque solution du criblé ont été prélevés et disposés sur une plaque de 96 puits avec 10 $\mu$ M d'aRNase J, 10X de sypro orange dans un tampon avec 400mM de NaCl en volume final de 26 $\mu$ l. La plaque a ensuite subi le même gradient de température et les courbes de dénaturation sont obtenues de la même façon que ci-dessus.

### 4. Essais de cristallisation

Les essais de cristallisation sont réalisés par la méthode de diffusion de vapeur en goutte assise dans des boîtes de 96 puits (MRC 2 Well Crystallization Plate, MDL) à l'aide du robot Mosquito®LCP (TTPLabtech). Différents criblés commerciaux sont réalisés (Qiagen) et listés dans le **tableau 8**. Dans chaque goutte, 200 à 250nl de solution protéique et 150 à 200nl de la solution d'agent précipitant du puits (80 $\mu$ l) correspondant sont déposés. Les boîtes sont scellées et stockées à 12°C dans un automate de visualisation (FORMULATRIX) ; les images des gouttes de cristallisation, régulièrement enregistrées, sont analysées via le logiciel Rockmacker.

Protéine	Cribles de cristallisation
<i>Pab</i> -ASH-Ski2 DomN +/- <i>Pab</i> -aRNase J (Ratio 2 :1 et 3 :1)	JSCG+
	Pro Complex
	Compass
	PEG I
	PEG II
	PhClear I
	Ph Clear II
<i>Pab</i> -ASH-Ski2	JSCG I
	JSCG II
	JSCG III
	Cation
	Anion
	AmSO4
	PhClear I
	PhClear II
	PEG I
	PEG II
	PACT
	MbClassI
	Pro Complex
	Nucleix
	Compass
Plaque 1	
Plaque 2	
<i>Pab</i> -ASH-Ski2 ΔC	JSCG I
	JSCG II
	JSCG III
	JSCG +
<i>Pab</i> -ASH-Ski2/ ADN (5RS1)	Nucleix
	Compass
	JSCG III

**Tableau 8. Cribles de cristallisation testés pour les protéines *Pab*-ASH-Ski2, *Pab*-ASH-Ski2ΔC, *Pab*-ASH-Ski2-DomN et *Pab*-aRNase J.**

## 5. Modélisation structurale des protéines

Les structures connues des protéines sont récupérées dans la banque de données des protéines PDB (Protein Data Bank) et modélisées dans le logiciel Discovery studio. Les structures des protéines non résolues à ce jour sont prédites et modélisées dans le serveur Phyre2 (Protein Homology/analogY Recognition Engine V 2.0). Phyre2 utilise des méthodes de détection d'homologie avancées pour construire des modèles 3D à partir de la séquence en acides aminés.

# RESULTATS





**I. Publication 2 : Phung D.K., Etienne C., Batista M. et al. 2020**

**RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the  $\beta$ -CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic**

**RNA exosome machinery**

**Nucleic Acids Research**



# RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the $\beta$ -CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic RNA exosome machinery

Duy Khanh Phung<sup>1,†</sup>, Clarisse Etienne<sup>1,†</sup>, Manon Batista<sup>1</sup>, Petra Langendijk-Genevaux<sup>1</sup>, Yann Moalic<sup>2</sup>, Sébastien Laurent<sup>2</sup>, Sophie Liuu<sup>3</sup>, Violette Morales<sup>1</sup>, Mohamed Jebbar<sup>2</sup>, Gwennaele Fichant<sup>1</sup>, Marie Bouvier<sup>1</sup>, Didier Flament<sup>2</sup> and Béatrice Clouet-d'Orval<sup>1,\*</sup>

<sup>1</sup>Laboratoire de Microbiologie et de Génétique Moléculaires, UMR5100, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse, France, <sup>2</sup>Ifremer, Univ Brest, CNRS, Laboratoire de Microbiologie des Environnements Extrêmes, F-29280 Plouzané, France and <sup>3</sup>Micalis Institute, PAPPSO, INRA, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France

Received October 24, 2019; Revised January 14, 2020; Editorial Decision January 16, 2020; Accepted January 23, 2020

## ABSTRACT

A network of RNA helicases, endoribonucleases and exoribonucleases regulates the quantity and quality of cellular RNAs. To date, mechanistic studies focussed on bacterial and eukaryal systems due to the challenge of identifying the main drivers of RNA decay and processing in Archaea. Here, our data support that aRNase J, a 5'-3' exoribonuclease of the  $\beta$ -CASP family conserved in Euryarchaeota, engages specifically with a Ski2-like helicase and the RNA exosome to potentially exert control over RNA surveillance, at the vicinity of the ribosome. Proteomic landscapes and direct protein–protein interaction analyses, strengthened by comprehensive phylogenomic studies demonstrated that aRNase J interplay with ASH-Ski2 and a cap exosome subunit. Finally, *Thermococcus barophilus* whole-cell extract fractionation experiments provide evidences that an aRNase J/ASH-Ski2 complex might exist *in vivo* and hint at an association of aRNase J with the ribosome that is emphasised in absence of ASH-Ski2. Whilst aRNase J homologues are found among bacteria, the RNA exosome and the Ski2-like RNA helicase have eukaryotic homologues, underlining the mosaic aspect of archaeal RNA machines. Altogether, these results suggest a fundamental role of  $\beta$ -CASP RNase/helicase complex in archaeal RNA metabolism.

## INTRODUCTION

Post-transcriptional regulation of gene expression demands accurate and timely RNA processing and decay to ensure coordinated cellular behaviours and fate decisions. Therefore, understanding RNA metabolic pathways and identifying RNA processing machineries, composed in general of ribonucleases (RNases) and ancillary enzymes such as RNA helicases, are major challenges in RNA biology. Currently, the best-understood RNA-dedicated pathways at the molecular level are those of Bacteria and Eukarya. In contrast, in Archaea, these molecular processes have been overlooked and are far from being understood.

Archaea, micro-organisms with signature sequences reported in all terrestrial and in human microbiome (1), have attracted considerable attention because of the orthologous relationships existing between their information processing machineries and those of eukaryotes (2–8). Regarding RNA machineries, it is worth mentioning that the archaeal 70S ribosome appears to be a simplified orthologous version of the eukaryal 80S ribosome with a reduced protein number (9,10) and that the archaeal RNA polymerase (RNAP) shares several features with eukaryal RNAPII such as similarities in amino acid sequences and of structures (11,12). Furthermore, most archaeal genomes contain genes encoding an evolutionary-conserved phosphorolytic 3'-5' RNA-degrading machinery, the RNA exosome (13), with the exception of Halophiles and some methanogens that possess homologues of bacterial RNase R (14,15). In addition to its ribonucleolytic activity, the archaeal RNA exosome possesses a 3'-end RNA-tailing activity (14). This machinery,

\*To whom correspondence should be addressed. Tel: +33 561 335 875; Fax: +33 561 335 886; Email: Beatrice.Clouet-dOrval@ibcg.biotoul.fr

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present address: Duy Khanh Phung, University College London, Gower Street, Darwin Building, London WC1E 6BT, UK.

which shares high sequence and structure similarity with its eukaryotic counterpart, is constituted of a central catalytic core of three dimers of Rrp41-Rrp42 forming an hexamer ring with the three Rrp41 subunits carrying the catalytic activity (16–18). On the top, an RNA-binding platform composed of a trimer of Rrp4 and/or Csl4 subunits that has high affinity for A-rich RNA sequences and that is required for effective RNA degradation form the RNA exosome cap (19–22). In addition, a protein annotated as DnaG, composed of a primase domain, is part of the RNA exosome cap through an interaction with Csl4 (23). To date, the contribution of the archaeal RNA exosome to specific biological pathways is still unknown and it remains to determine if archaeal cells harbour dedicated RNA exosomes, with heterogeneous and/or homogenous trimer cap composition *in vivo*.

The  $\beta$ -CASP RNases appear as versatile ancient enzymes with dual endo- and 5'-3' exo-ribonucleolytic activities that act to control RNA maturation and stability in Bacteria and Eukarya (24,25). Our previous studies identified and characterized three major groups of archaeal  $\beta$ -CASP RNases: aCPSF1 and aCPSF2 orthologous to the eukaryal Cleavage & Polyadenylation Specific factor CPSF73 and aRNase J orthologous to bacterial RNase J (26). This composite setting raises the question of the role of each  $\beta$ -CASP group in archaeal RNA homeostasis as well as the evolutionary origin of this family of enzyme among the three domains of life. In recent highlights on this mosaic system, we proposed potential functions within RNA-degradation machineries in Archaea (15,27).

In Eukarya and Bacteria, as a general common thread, both mRNA decay and translation are intimately coordinated. These involve, after decapping/deprotection and deadenylation, the 5'-3' exoribonucleolytic activity supplied by the eukaryal Xrn1 or bacterial RNase J exo-RNases, which is complemented by the 3'-5' exoribonucleolytic activity supplied by the eukaryal RNA exosome or bacterial PNPase, respectively (28,29). Conserved 5'-3' exoribonucleolytic activities are crucial in mRNA homeostasis in all eukaryotic and most bacterial cells. In archaeal cells, the 5'-3' exoribonucleolytic activity is also conserved and is carried by aRNase J and aCPSF2. The role of this activity has been little studied (30,31) and it remains to be clarified in which specific RNA decay and processing pathways it is involved. aRNase J and aCPSF2 exo-RNases have been identified being highly processive and having a preference for mono-phosphorylated RNA substrates in Euryarchaeota and Crenarchaeota, respectively (32,33). In addition, a directional 5'-3' RNA decay pathway was proposed to be at play in the crenarchaeal *Sulfolobus solfataricus* cells. *In vitro* and *in vivo* studies evidenced cap-like structure of translation initiation factor (aIF2- $\gamma$ ), protecting RNA 5'-triphosphorylated ends from a 5'-end-dependent decay (30,33–34).

In this study, we focus on aRNase J as we identified this enzyme to be encoded in most of the Euryarchaeota phylum (26,32,35–36). Deciphering the physiological functions and relevance of 5'-3' ribonucleolytic activity of the aRNase J is a mandatory step towards understanding its impact on cellular RNA homeostasis in euryarchaeal cells. Using combined biochemical, proteomics and genetics cou-

pled with phylogenomic analyses, we show that aRNase J forms a widely-conserved multi-protein complex with the archaeal specific helicase of the Ski2 family, ASH-Ski2 and we present evidences that, in *Thermococcales* cells, the aRNase J cross-talks with the RNA exosome. Moreover, our mutational analyses pinpoint functionally important domains involved in the aRNase J/ASH-Ski2 and aRNase J/Csl4 protein–protein interactions, respectively. Finally, we observed that a potential aRNase J–ribosome interaction exists and that could physically link 5'-3' mRNA decay to translation in Euryarchaeota.

Altogether, our work builds the first blocks of complexes and networks involved in RNA-metabolic pathways in Euryarchaea as we propose that aRNase J participates in RNA decay routes in the vicinity of the ribosome, in coordination with the ASH-Ski2 helicase and/or the RNA exosome, pointing a close relationship between the RNA exosome and Ski2-like helicase in Euryarchaea. The mosaic setting of players around the 5'-3' exo-RNase aRNase J, which is orthologous to bacterial RNase J, give a milestone towards the conservation of general principles of RNA-processing across the three domains of life since the ASH-Ski2 and RNA exosome have homologues found in Eukarya.

## MATERIALS AND METHODS

### Vectors and Oligonucleotides

The supplementary Tables S1 and 2 summarize T7-promotor-driven pET vectors and oligonucleotides used in this study. All constructions were obtained by assembling polymerase chain reaction fragments using InFusion<sup>®</sup> cloning kit (Takara). Using appropriate sets of oligonucleotides, pET vectors were amplified with the PrimeSTAR Max DNA polymerase (Takara) and the coding sequence of *Pyrococcus abyssi* aRNase J (PAB1751), ASH-Ski2 (PAB2313), Hel308 (PAB0592), DnaG (PAB0316), Rrp41 (PAB0420), Rrp4 (PAB0419) and Csl4 (PAB2314) were amplified from genomic DNA using the Phusion High-Fidelity DNA polymerase (ThermoFisherScientific). The pET vectors expressing the ASH-C124A, Csl4-F121A and Csl4-F129A variants were generated by site-directed mutagenesis of their wild-type counterparts with appropriate sets of oligonucleotides using the QuikChange II XL Kit (Stratagene) as recommended.

### *Thermococcus barophilus* strains

The chromosomal copy of the TERMP\_01768 (encoding *Tba*-ASH-Ski2) and the TERMP\_00146 (encoding *Tba*-aRNase J) were deleted by the pop-in/pop-out method to generate the *Thermococcus barophilus*  $\Delta$ ASH-Ski2 and  $\Delta$ aRNase J strains, respectively using published protocol (37,38).

### Production and purification of bait proteins

*Escherichia coli* BL21-CodonPlus (DE3) cells freshly transformed with pET15-aRNase J, pET15-ASH-Ski2, pET21-ASH-Ski2, pET15-Csl4 or pET15-Rrp41 vectors (Supplementary Table S1) were grown in 400 ml of LB medium at 37°C. Protein production was induced in exponential phase

at an  $OD_{600nm}$  of 0.6 by the addition of 0.1 mM IPTG (Isopropyl  $\beta$ -D-1-thiogalactopyranoside). After 3 h of induction at 30°C, the cells were collected, suspended in 10 ml of lysis buffer (50 mM NaPhosphate, 300 mM NaCl, 10 mM Imidazole) supplemented with 1 mg.ml<sup>-1</sup> of lysozyme and a mix of ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (cOmplete™, Roche) and lysed by sonication (4×[5\*10 s], 50% cycle, VibraCell Bioblock Scientific). When mentioned, the cleared extracts, obtained by centrifuging the crude extracts (20 000 g 4°C, 20 min), were treated with a mix of RNase A (20  $\mu$ g.ml<sup>-1</sup>), RNase T1 (1U. $\mu$ l<sup>-1</sup>) and DNase I (20  $\mu$ g.ml<sup>-1</sup>) containing 10 mM of MgCl<sub>2</sub> for 30 min at 37°C. After a heating step at 70°C for 20 min, the extracts were furthered clarified by centrifugation (20 000 g, 4°C, 20 min). The recombinant proteins were two-step purified from the soluble fractions to near homogeneity using FPLC (Fast Protein Liquid Chromatography, Äkta-purifier10, GE-Healthcare): first, by a nickel-attached HiTrap chelating column (HiTrap 1 ml, GE-Healthcare) with a linear gradient of 100–500 mM imidazole; secondly, by a heparin column (HiTrap Heparin 1 ml, GE-Healthcare) with a linear gradient of 300 mM to 1 M NaCl. The eluted fractions containing His-tagged proteins were pooled and dialyzed overnight against 500 ml of dialysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT (Dithiothréitol), 10% glycerol buffer).

### Co-purification by chromatography affinity

Cells expressing His-tagged bait proteins were produced as described above. Cells expressing untagged prey proteins were obtained after transforming *E. coli* BL21-CodonPlus (DE3) cells with pET11-aRNase J, pET11-ASH-Ski2, pET11-Csl4 or pET11-*Pab*-Rrp41 vectors (Supplementary Table S1) and using the same conditions for growth and induction. Two cell pellets from 200 ml of cultures expressing either bait or prey protein were mixed upon lysis (as before). The bait proteins alone or in complex with the prey proteins were purified by nickel affinity chromatography using Ni<sup>2+</sup>-NTA column matrices (HiTrap 1 ml, GE-Healthcare). The elution was obtained with a linear 100–500 mM imidazole gradient after washing steps at 10 and 50 mM of imidazole. For each couple of bait/prey that were tested, the co-purification assays were at least performed twice either in presence or absence of RNase/DNase treatment (as above). The proteins from recovered fractions were analysed by Western blotting (4–15% Mini PROTEAN® TGX Stain-Free™ Gels & Trans-blot® Turbo™ Nitrocellulose Transfer Pack, BioRad). The bait proteins were probed using a His Tag HRP-conjugated antibody (Ozyme) diluted 5000-fold. The prey proteins were probed using polyclonal antibodies against *Pab*-aRNase J, *Pab*-ASH-Ski2, *Pab*-Csl4, *Pab*-Rrp41, *Pab*-DnaG or *Pab*-aCPSF1 (custom polyclonal antibodies, Eurogentec) diluted 10 000-fold and an anti-rabbit IgG HRP conjugate (Promega) diluted 5000-fold.

### Pull-down by affinity purification assays

The pull-down assays were performed in triplicate with clarified whole-cells extract of *P. abyssi* cultivated on exponen-

tial growth phase in bioreactors under physiological conditions (95°C, pH 6.5, anaerobic). The cell extracts were prepared as described in (39). Briefly, *P. abyssi* pelleted cells were re-suspended in 1/3 w/v of 1× phosphate-buffered saline (PBS) buffer (Euromedex), supplemented with 300 mM of NaCl and a mix of EDTA-free protease inhibitor (cOmplete™, Roche). After sonication (VibraCell Bioblock Scientific), the crude extracts were clarified by centrifugation (20 000 g, 30 min, 4°C). Purified His-tagged proteins were used as baits in clarified whole-cell extract of *P. abyssi*. Briefly, 20  $\mu$ g of bait proteins were immobilized on 0.6 mg of cobalt-coated magnetic beads (Dynabeads, Invitrogen). The complex baits-beads were further incubated with 2 mg of *P. abyssi* extract under rotation for 2 h at room temperature. The protein complexes formed *in vitro* were separated on a magnet and washed extensively with PBS buffer (3\*4 ml and 2\*1 ml). A second round of pull-down assays was performed with an additional step to eliminate non-specific interaction via DNA or RNA molecules by incubating the protein complexes with an RNase/DNase mix (10  $\mu$ g.ml<sup>-1</sup> of RNase A and DNase I) for 30 min at room temperature before applying the magnetic force. A control was also performed under identical conditions using cobalt-coated beads alone instead of the baits-beads complexes. Purified protein complexes were eluted in 25  $\mu$ l of XT sample buffer (BioRad) containing 2  $\mu$ l of 20× reducing buffer at 95°C for 10 min and separated by sodium dodecyl sulphate-polyacrylamide gelelectrophoresis (SDS-PAGE) short migration on 12% SDS-PAGE (Criterion XT Precast gels–BioRad). After a short migration, the proteins were extracted by cutting out each protein track into one piece followed by shotgun proteomic analyses. For western blot analyses, protein complexes were separated by SDS-PAGE on 4–20% pre-cast Bis-Tris gel (NuPAGE Life technology).

### Proteomic analysis

The mass spectrometry analyses were performed at Paris Sud Ouest PAPPSO proteomics core facility (<http://papso.inra.fr>). Briefly, after a short run on a SDS-PAGE gel, bands of proteins were excised from the gel. In-gel tryptic digestion was performed 6 h at 37°C with 100 ng of modified trypsin (Promega) dissolved in 50 mM NH<sub>4</sub>CO<sub>3</sub>. Peptides were extracted with 0.1% trifluoroacetic acid (TFA) and 50% acetonitrile (ACN). Extracted peptides were dried and suspended in 30  $\mu$ l of 0.05% HCOOH, and 2% ACN. A total of 4  $\mu$ l of samples were then loaded on a NanoLC-Ultra system (Eksigent). Eluted peptides were analysed on-line with a QExactive mass spectrometer (Thermo Scientific Electron) using a nanoelectrospray interface. Peptide ions were analysed using Xcalibur 2.1 with data-dependent acquisition steps. Peptides were identified with X!TandemPipeline open source software by spectrum matching approach. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015856".

The MS data were processed in order to identify specific interaction signals. Global specific spectra from samples were normalized, as described in (40), between replicate series. We used a cut-off of two normalized spectra as



minimum MS signal for network hit validation. Normalized spectra were then averaged between replicates and referenced versus control to calculate the number of ‘Referenced Spectra’. Calculation of the ‘Specificity Index’ score is the ratio of the averaged normalized spectra in control versus assay. The ‘Specificity Index’ varies from 0 to 1 (with a maximum threshold of 1) as the specificity decreases. For the AP-MS experiments with aRNase J as bait, nuclease and nuclease-free assays were run in triplicate, whereas for ASH-Ski2 the same analyses were performed in 4 replicates. For aRNase J  $n = 6$  for the controls and for ASH-Ski2  $n = 8$  for the controls. For ASH-Ski2, we took only into consideration hit proteins detected independently in both His-Tag orientations. The effect of nuclease has been identified as:

$$\text{Nucl. Effect} = 100 \times \frac{\text{Spectra No Nuclease} - \text{Spectra Nuclease}}{\text{Spectra No Nuclease}}$$

Diagram networks have been created using Cytoscape 3.7.2 software (41).

### Multiple sequence alignments and phylogenetic tree constructions

The complete archaeal and bacterial genome entries were retrieved from the EBI (European Bioinformatics Institute). The collection of helicases from super-families SF1 and SF2 were identified by RPS-Blast of Pfam domains DEAD (PF00270) or Helicase\_C (PF00271) among a set of completely sequenced non-redundant genomes composed of 114 archaeal genomes and 1105 bacterial genomes. A total of 17 661 helicases were identified and further compared to each other using the BlastP program. Partitioning of the all-vs-all Blast results was achieved using the Markov Cluster Algorithm (42) and resulted in 84 clusters of >5 orthologous archaeal helicases. The archaeal sequences were collected and corresponded to occurrences with COG1202 for ASH-Ski2. We also collected an initial set of proteins using COG profile annotations (COG1096 for Csl4 and COG0689 for Rrp41) from complete archaeal genomes. Hidden Markov model (hmm) profiles were built for each family and used with hmmsearch (HMMER package, (43)) to identify homologues in our archaeal proteome sample. In addition, to identify potential unannotated genes and pseudogenes, we performed tblastn searches.

The archaeal phylogenetic tree was inferred from a concatenated dataset of 81 protein families obtained from COG annotation. If multiple copies occurred in a genome, all paralogues were removed. The sequence alignments for each family were created using the MUSCLE program (44) with the default parameters. We used the trimAL program (45) to remove spurious sequences and poorly aligned positions and to analyse the quality of the alignments according to gap numbers and residue conservation in the columns of the alignments. These parsed alignments were concatenated to produce a single alignment of 17 827 residues. When a species did not have a record for a family, the missing sequence was replaced by gaps in the alignment. The maximum-likelihood tree was computed with PhyML (46) and the optimal combinations of parameters was selected using ProtTest3 program (47). The LG model of sequence evolution was used and the gamma-distributed substitution rate variation was approximated by eight discrete categories

with shape parameter and proportion of invariant sites estimated from the data. The statistical branch support was inferred with the parametric bootstrap. The archaeal Ski2-like helicase alignment was built by Mafft incorporating local pairwise refinement (L-INS-i) up to 2000 iteration (maxiterate 2000) (48) and trimmed with trimAL as described above. The archaeal ASH-Ski2 helicase tree was computed with the same approach as that of the species tree except that the gamma-distributed substitution rate variation was approximated by four discrete categories. Both trees (species and Ski2-like helicases) were arbitrarily rooted and were drawn with the online version of iTOL (49).

### Sucrose gradient fractionation of cell extracts

*Pyrococcus abyssi* and *T. barophilus* cells were grown at 92°C and 85°C, respectively, in continuous culture in a 5-l Gazlift bioreactor in MES medium under anaerobic conditions at pH 6.8 (50). Cells were maintained in exponential growth phase between 2 and 4\*10<sup>8</sup> cells per ml. After cell culture harvesting at 8°C, Cells were pelleted by centrifugation (10 000 g, 20 min at 4°C) and washed two times with a sterile sea salt solution at 30 g.l<sup>-1</sup>. Dry cell pellets were stored at -80°C. A total of 200 mg of cell pellet was re-suspended in 1/3 w/v of buffer TK buffer (20 mM Tris-HCl, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) containing a cocktail of EDTA-free protease inhibitor (cOmplete™, Roche) or TK-EDTA buffer (TK buffer supplemented with 20 mM EDTA). Whole-cell extracts were prepared by sonication (10\*10 s, 50% cycle, VibraCell-Bioblock scientific). Lysates were cleared at 21 000 g for 30 min. Approximately 10 mg of whole-cell extract protein were layered on a linear 10–30% sucrose gradient prepared in TK or TK-EDTA buffers and centrifuged for 5 h at 35 000 rpm at 4°C in Beckman-Avanti XPN-80 SW41 rotor. It was not possible to determine the positions of 30S and 50S ribosomal subunits by classical A254 scanning with the ISCO UA-6 gradient fraction collector due to the inherent colorimetry of the *P. abyssi* and *T. barophilus* cellular extract. For each gradient, 21 fractions of 500 µl were collected. A total of 250 µl of each fraction were precipitated with 45 µl of 100% trichloroacetic acid (TCA). After 30min at -20°C, samples were centrifuged 30 min at 16 000 g at 4°C. The protein pellets were washed with 500 µl of glacial acetone (16 000 g, 15 min at 4°C), dried and re-suspended in 20 µl of 20 mM Tris-HCl pH 8. After adding 4 µl of 6× loading buffer (375 mM Tris-HCl pH 6.8, 12% SDS, 30% β-mercaptoethanol, 0.6% Bromophenol Blue, 36% glycerol), samples were heated at 95°C for 10 min before proceeding to SDS-PAGE.

### Relative quantification of ribosomal RNAs by Slot Blot

A total of 20µl of each fraction collected from sucrose gradient, equally diluted beforehand to set the fraction with the highest nucleic acid content at a concentration of 2.5 ng.µl<sup>-1</sup>, was added to 55 µl of denaturation buffer (2.2 M formaldehyde, 50% deionized formamide, 0.5 mM EDTA, 10 mM MOPS, 4 mM NaCl) and incubated at 65°C for 5 min. Each 75 µl-sample was spotted on a nylon membrane (Amersham Hybond-XL; GE Healthcare) by vacuum filtration (PR648-Hoefler™ Slot Blot). After UV-crosslinking, blots were hybridized overnight at 42°C with

5'-end  $P^{32}$ -labelled oligonucleotides in Roti-Hybrid-Quick buffer (Roth) and washed three times at 42°C for 15 min with SSC buffer (SSC 20×: 3M NaCl, 300 mM sodium citrate, pH 7) containing 0.1% of SDS (5×, 1× and 0.1×, respectively). Sequences of antisense oligonucleotides used to detect *T. barophilus* 16S and 23S rRNAs are listed in Supplementary Table S2. Radioactive signals were visualized on a phosphorImager (Typhoon Trio-Amersham-Bioscience) and quantified using MultiGauge software (Fujifilm).

## RESULTS

### The protein–protein interaction network of aRNase J includes key components of RNA metabolism

To explore the protein interaction network of the euryarchaeal aRNase J, we carried out affinity purification coupled to mass spectrometry (AP-MS) analyses in *P. abyssi* cell extracts, as already described (39). Recombinant aRNase J from *P. abyssi*, with C-terminal (His)<sub>6</sub>-tag (aRNase J-His), was produced, purified and used as baits for AP-MS experiments. In order to discriminate specific protein interaction from column background, we performed mock AP-MS analyses with *P. abyssi* cell extracts in absence of protein bait. In addition, to determine if some interactions are mediated through RNA or DNA, we implemented a nuclease treatment after incubating the cell extract with the bait. The co-purified partners were identified by bottom-up proteomic techniques coupled with mass spectrometry. We used an algorithm that prioritizes the specificity of the interaction to classify partners identified in triplicate AP-MS. The list from the most to the least specific is displayed on Supplementary Table S3. Remarkably, we found that aRNase J is at the centre of a protein interaction network that includes proteins with central functions in physiological processes. The pie chart, shown on Figure 1A (left panel), categorizes these processes. A large majority (64%) of the protein partners of aRNase J-His are annotated as involved in RNA metabolism from transcription (RNA polymerase subunits), RNA modification (C/D guide RNP components) and RNA decay (RNA exosome) to translation (ribosomal proteins and translation initiation factors). Amongst protein partners, 14 were detected with a high-specific score in the aRNase J pull-downs whereas no spectra signal were sensed in the controls for these candidates (values of 0 in Supplementary Table S3). Remarkably, the Rrp41 and Rrp42 catalytic core subunits and the Rrp4 cap subunit of the RNA exosome were retrieved among the foremost of the fourteen highly significant protein partners (Supplementary Table S3). Both subunits of the RNA polymerase (RpoA1/A2) were also detected with highly significant scores. Ribosomal proteins and components of the C/D box guide RNP complex are also part of the aRNase J network (51,52). Remarkably, these exploratory analyses reveal PAB2313, a putative ATP-dependent RNA helicase, as a potential partner of aRNase J-His.

In addition, proteins known to be involved in DNA maintenance processes, like RPA41, were also recovered as significant partners (Supplementary Table S3). RPA41 corresponds to the large subunit of the RPA complex that binds to single-stranded DNA. Interestingly, a physical in-

teraction between RPA41 and the RNA polymerase has been previously demonstrated in *P. abyssi* (39). It was revealed that the RPA complex enhance transcription rate *in vitro*, probably via RPA41 interacting with the non-template strand of the elongating complex. In this light, the presence of RPA subunit in the interaction network of aRNase J could be considered as a further evidence for the involvement of aRNase J in RNA metabolism.

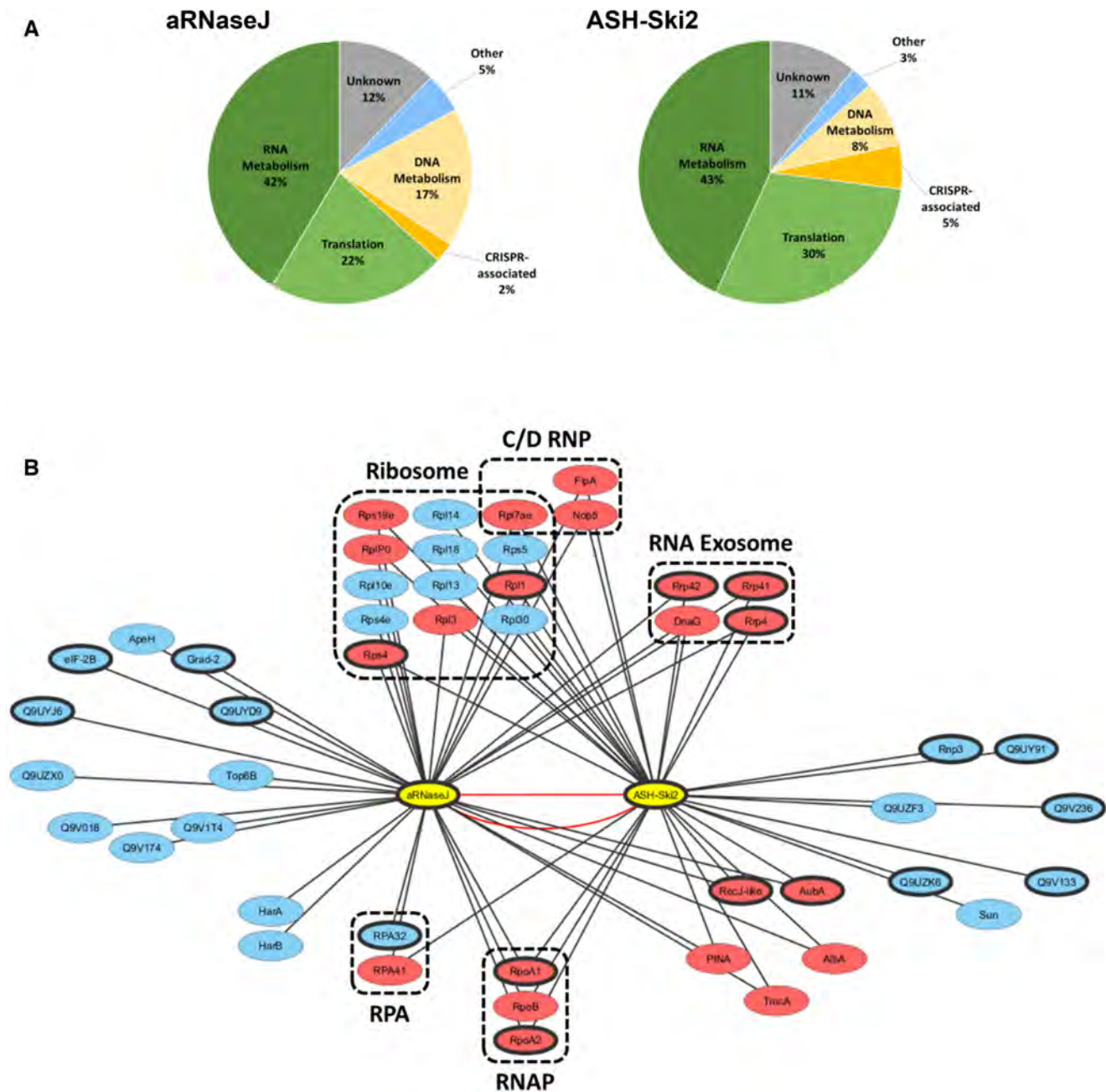
Finally, to spotlight protein complexes of aRNase J network, we built a protein interaction map of aRNase-His based on the spectral index of Supplementary Table S3 (Figure 1B). In light of this network, we propose that the 5'-3' exonuclease aRNase J cross-talks with RNA machineries of the cell.

### The *P. abyssi* ASH-Ski2 helicase shares a common network with *P. abyssi* aRNase J RNase and the RNA exosome

Interestingly, in the network of aRNase J, we retrieved PAB2313, a putative RNA helicase, that appears to be a member of the Ski2-like helicase family (53) renowned for their crucial role in RNA decay in eukaryotic cells (54) and we named it ASH-Ski2. To gain further experimental evidence of a coordinated action of aRNase J and ASH-Ski2, we performed new rounds of AP-MS analyses in duplicate with recombinant C- and N-terminal (His)<sub>6</sub>-tagged versions of *P. abyssi* ASH-Ski2 as baits, respectively, and obtained an exhaustive list of partners (Supplementary Table S4). The ASH-Ski2 partner pie chart shows that a large majority (73%) of identified proteins are annotated as players of RNA metabolism and translation. Indeed, out of 34 in total, 12 ribosomal proteins and four subunits of the RNA exosome are identified as specific partners (Figure 1A, right panel and Supplementary Table S4). Notably, aRNase J is ranked as the first candidate of the ASH-Ski2 partner list of Supplementary Table S4. aRNase J is detected, both with a high-specific score and reference spectra in assays treated or not with nucleases. Furthermore, 21 protein partners are common to both aRNase J and ASH-Ski2 networks (Figure 1B). This highlights a tight cross-talk of aRNase J and ASH-Ski2 in *P. abyssi* cells. It should be noted that several partners of ASH-Ski2 are affected by the nuclease-treatment when compared to aRNase J, which includes the RNA polymerase subunits but also the RNA exosome subunits.

To ensure that aRNase J and ASH-Ski2 are part of the protein interaction network of the RNA exosome, we performed additional pull-down assays, in duplicate, using a purified N-tagged His-Rrp41 of *P. abyssi* as bait. Amongst the pull-downed proteins of the Rrp41 subunit, by western blotting, we specifically detected endogenous Csl4 and DnaG cap subunits, aRNase J and ASH-Ski2, but not the endo-RNase aCPSF1 (Figure 2). Altogether with the limitation that the pull-down was performed with the sole Rrp41 subunit and not with a re-formed complex with the Rrp42 core and Csl4/Rrp4 cap subunits, these results support a potential interplay between aRNase J, ASH-Ski2 and components of the archaeal RNA exosome in *P. abyssi* cells. This is thoroughly consistent with the results of the above AP-MS analysis set.



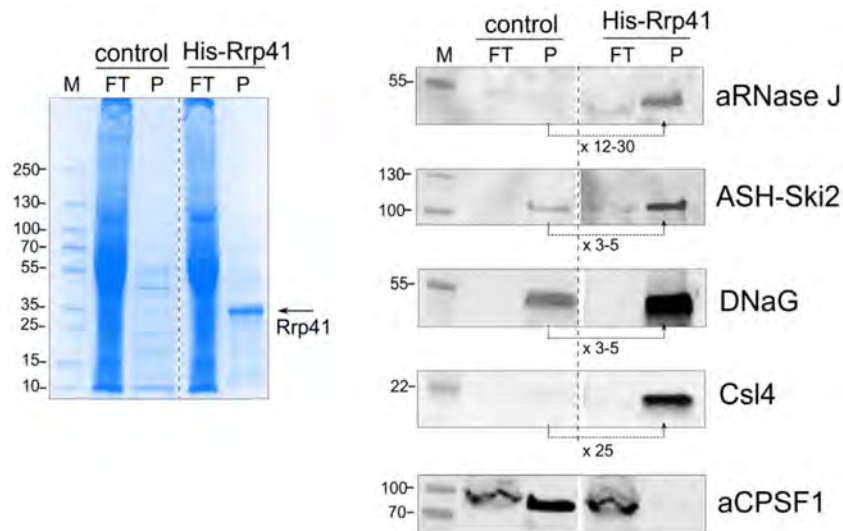


**Figure 1.** (A) Functional classification pie charts of proteins identified by pulldown-MS/MS approach using recombinant aRNase J and ASH-Ski2 as baits, respectively. (B) Network of protein partners for aRNase J and ASH-Ski2. The aRNase J and ASH-Ski2 baits are labelled in yellow. Network backbone is represented in black lines except for ASH-Ski2 and aRNase J for which reciprocal interactions are in red. Common partners of ASH-Ski2 and aRNase J are indicated in pink whereas specific partners are indicated in blue. Bolded circle indicate highly specific hits, absence in all control runs, with a 'spectra index' of 0 (see Supplementary Tables S3 and 4).

### The archaeal ASH-Ski2 members share identical taxonomic distribution as aRNase J members

To gain insights into the relationship that exists between archaeal aRNase J and ASH-Ski2 members as well as with the RNA exosome subunits, we compared their taxonomic distributions amongst the archaeal phylogeny. To do so, we first upgraded a phylogenomic study of the superfamily of the archaeal SF2 helicase (53) in order to accurately identify the Ski2 family of helicases encoded in archaeal

genomes. In contrast to the role of eukaryotic Ski2-like RNA helicases in RNA homeostasis, reports on the functions of archaeal Ski2-like helicase family members are still scarce. As a common theme throughout SF2 helicase families, the unique characteristics of Ski2-like family members are mainly derived from accessory domains that decorate the helicase core and provide additional functionalities (55,56). Ski2-like helicase structures demonstrate that the molecular 'core' of all Ski2-like helicases is a ring-like



**Figure 2.** Pull down assays using the recombinant (His)-Rrp41 protein as bait in whole-cell extracts of *Pyrococcus abyssi*; FT corresponds to the flow through, P to the pull-downed fraction and M to the protein ladder. Each fraction was analysed by Coomassie blue staining of SDS-PAGE (left panel) and by western blotting using specific antibodies for aRNase J, ASH-Ski2, DnaG, Csl4 and aCPSF1 (right panel). To highlight the pull-downed proteins, the relative signals were average from two independent assays as indicated below each panel.

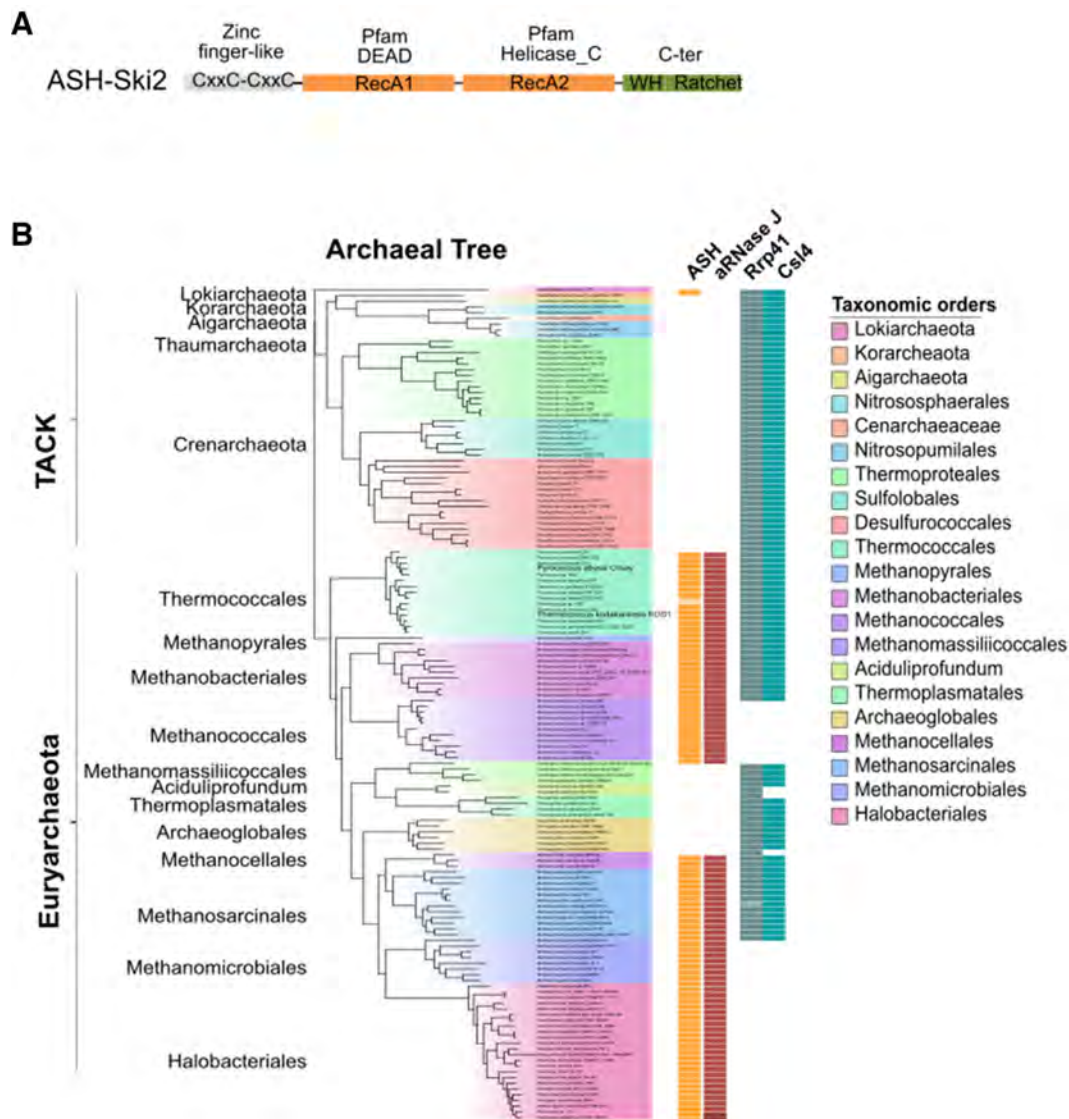
four domain assembly of two RecA domains, a winged helix domain and a ratchet domain (54,57). By performing RPS Blast searches, we retrieved all archaeal Ski2-like helicase member sequences from 147 non-redundant completely sequenced archaeal genomes. In addition to Hel308 members, characterized by the COG 1204 and documented with enzymatic features of DNA helicases (54,58–61), we also retrieved the ASH-Ski2 group, characterized by the COG 1202 (53). Interestingly ASH-Ski2 members hold an additional N-terminal domain in which four strictly conserved cysteines have the propensity to form a zinc-finger-like motif (62) (Figure 3A). ASH-Ski2 members are specific of euryarchaeal phylogenetic groups with the Lokiarchaeum as an exception (Figure 3B). Remarkably aRNase J and ASH-Ski2 members present similar occurrence, which is restricted to most euryarchaeal groups, with the exception of the archeoglobales and thermoplasmatales. Furthermore, the co-evolution of aRNase J and ASH-Ski2 members shown by the comprehensive congruence of aRNase J and ASH-Ski2 phylogenetic trees, with only few observed leaf displacements within taxonomic groups, reinforces their potential functional interplay (Supplementary Figure S1). Altogether, protein networks combined with our phylogenomics analyses argue for a coordinated action of aRNase J and ASH-Ski2 in specific cellular processes in Euryarchaeota.

Subsequently, we compared the taxonomic distribution of aRNase J, ASH-Ski2 and the RNA exosome Rrp41 and Csl4 subunits along the archaeal phylogeny (Figure 3B). Like in our previous studies (15,27), we showed that the RNA exosome is not present throughout the whole archaeal taxonomy and that aRNase J, as well as ASH-Ski2, do not follow the taxonomic distribution of RNA exosome components. Indeed, it is absent from the Halobacteria and Methanomicrobiales members for which the 3'-5' exoribonucleolytic activity is carried by another enzyme, the aRNase R (14,15). In the Methanococci group, no enzyme

with equivalent activity could be, so far, detected by comparative genomics.

#### aRNase J forms a stable complex with the helicase-like ASH-Ski2, *in vitro*

To investigate a direct interaction between aRNase J and ASH-Ski2, we used an *in vitro* co-purification assay on nickel affinity chromatography to test pairwise interactions. A recombinant (His)<sub>6</sub>-tagged bait (His-ASH-Ski2) and an untagged prey (aRNase J) from *P. abyssi* were expressed independently in *E. coli* cells using the pET15b and pET11b expression vectors, respectively (Supplementary Table S1). Cells over-expressing the recombinant (His)<sub>6</sub>-tagged bait or the untagged prey proteins were mixed in a 1:1 ratio prior the lysis step. After nucleic acids removal with a cocktail of nucleases, most of the proteins from *E. coli* were eliminated at 70°C for a specific enrichment of the thermo-resistant proteins from *P. abyssi*. The clarified cellular extracts were then injected on a nickel column to retain the bait His-ASH-Ski2 protein and its potential interacting partners. Washing steps with low imidazole concentrations served at removing weakly bound contaminants. The bait and the retained interacting proteins were eluted with a linear gradient of imidazole. The protein content of each recovered fractions was analysed by SDS-PAGE. The specific presence of aRNase J and His-ASH-Ski2 was revealed by western blotting using specific antibodies as shown in Figure 4A. As expected, aRNase J co-elutes with His-ASH-Ski2 even after nucleic acids removal. This establishes that, at high temperature and ionic strength (300 mM NaCl), aRNase J and His-ASH-Ski2 associate to form a stable complex that is not mediated through RNA or DNA molecules (Figure 4A, top panel). As a control, to demonstrate that protein co-purification is a consequence of complex formation rather than non-specific interactions with the column matrices, we controlled that untagged versions of aRNase J or ASH-Ski2



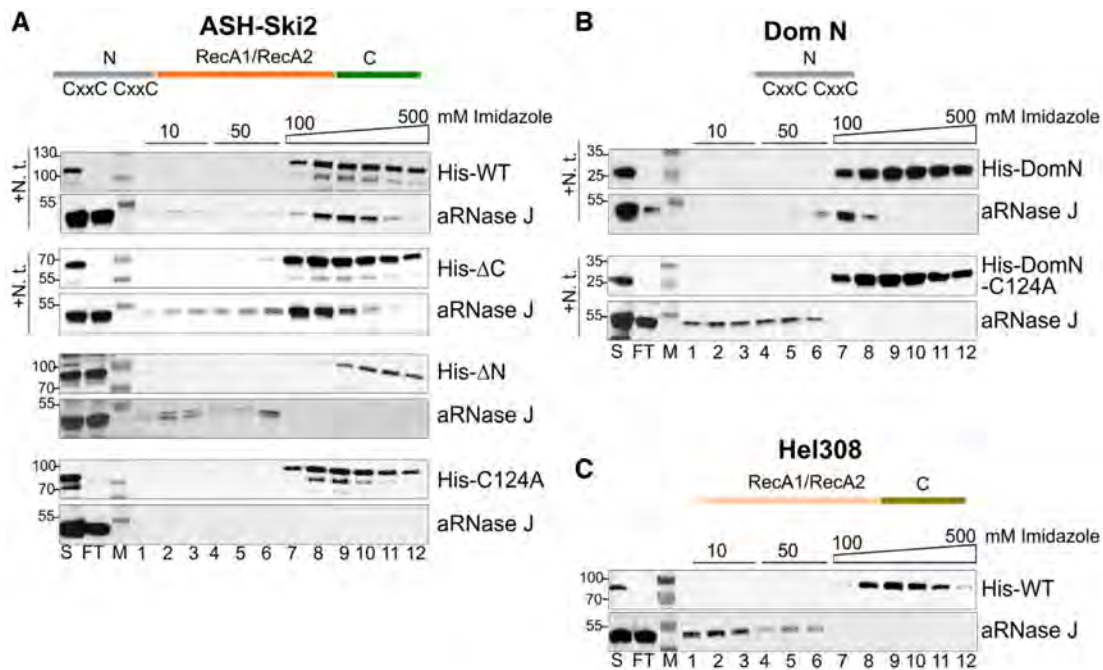
**Figure 3.** (A) Schematic protein domain architecture of the ASH-Ski2 group. The common SF2 core is composed of the RecA1 and RecA2 subdomains (Pfam DEAD & Helicase\_C) in orange, of the C-terminal domain in green and of the N-terminal domain in grey. The strictly conserved cysteine residues of the N-terminal domain are represented. (B) Taxonomic distributions of ASH-Ski2, aRNase J and Rrp41/Csl4 from the RNA exosome among archaeal genomes. Their occurrence is plotted and juxtaposed to the archaeal specie tree. Taxonomic orders and Pfam domains are colour coded. Darker and lighter shades indicate multiple copies and pseudogenes, respectively.

are not intrinsically retained on  $\text{Ni}^{2+}$ -NTA column matrices (Supplementary Figure S2A). Despite the robustness of these data, we cannot completely exclude a stabilization of an ASH-Ski2 /aRNase J complex by remaining *E. coli* proteins like ribosomal proteins that are highly abundant. However, we have additional evidence supporting an interaction between aRNase J and ASH-Ski2. By comparing the sedimentation profiles (5–25% sucrose gradient) of each individual purified protein to the one obtained with a mixture of both (Supplementary Figure S2B), we showed that the sedimentation profile of aRNase J is affected by the presence of ASH-Ski2. Indeed, whilst the sedimentation profile of aRNase J alone most likely corresponds to its homotrimeric form as observed for *Methanobolus psychrophilus* aRNase J in solution (63), in presence of ASH-Ski2, a sig-

nificant proportion of aRNase J is shifted to fractions of different molecular weight. This shift is most likely due to an interaction with ASH-Ski2 that is found in the same fractions as aRNase J.

To uncover which protein domain of ASH-Ski2 is specifically involved in the formation of a complex with aRNase J, variants of His-ASH-Ski2 were used as baits for the untagged prey aRNase J. Pairwise interactions were tested as previously described performing *in vitro* co-purification assays (Figure 4A, lower panels). Our results indicate that the accessory C-terminal domain of ASH-Ski2 (His- $\Delta$ C) is not required in establishing an *in vitro* interaction with aRNase J as the elution profile is comparable to the one with full-length ASH-Ski2. In contrast, the accessory N-terminal domain of ASH-Ski2 (His- $\Delta$ N) seems to be critical for the in-





**Figure 4.** *In vitro* co-purification assay of aRNase J challenged by His-ASH-Ski2 and His-Hel308. (A) Schematic representation of the protein domain architecture of ASH-Ski2 from *Pyrococcus abyssi* and affinity co-purification assays on nickel chromatography column of recombinant untagged aRNase J with His-ASH-Ski2 and its variants. The full-length protein (His-WT), deleted for the C-terminal (His- $\Delta$ C) and N-terminal (His- $\Delta$ N) domains or harbouring a punctual substitution of the conserved cysteine residue into alanine (His-C124A) were used as baits. The fractions were analysed by western blotting using an His-Tag or a specific antibody against aRNase J. S: supernatant loaded on the affinity Nickel column; FT: flow through; 10 mM (lanes 1–3) and 50 mM (lanes 4–6) imidazole washing steps; Elution with an Imidazole gradient from 100 to 500 mM (lanes 7–12). Experiments performed with nuclease treatment (+Nt) are indicated on the left of each panel. (B) Assays as performed in (A) with the wild-type N-terminal domain DomN (His-WT) or containing the C124A substitution (His-C124A) of ASH-Ski2. (C) Assays as carried in (A) but with His-Hel308 as the protein bait.

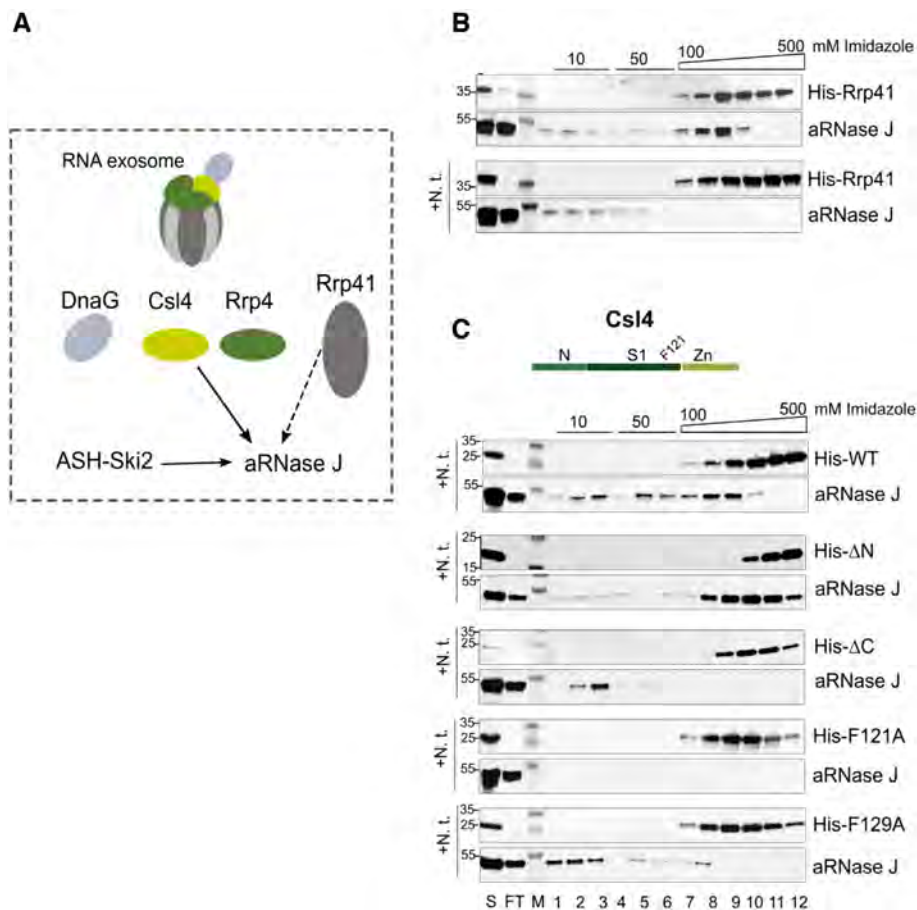
teraction since aRNase J does not co-elute with the  $\Delta$ N variant of ASH-Ski2. Finally, we showed that the accessory N-terminal domain of ASH-Ski2 expressed alone (His-DomN) is sufficient to establish an *in vitro* interaction with aRNase J (Figure 4B, top panel).

The sequence alignment of the N-terminal domains of archaeal ASH-Ski2 highlights the presence of four strictly conserved cysteines with the propensity to form a zinc finger-like motif (Figure 4A). To test whether this motif is critical for the interaction between ASH-Ski2 and aRNase J, we substituted the C124 residue of full-length His-ASH-Ski2 and His-DomN by an alanine (A) and we challenged their capacity to retain aRNase J on  $\text{Ni}^{2+}$ -NTA column as before. Our results show that the conserved C124 residue is critical in establishing or in stabilizing an interaction between the full-length or DomN of ASH-Ski2 and aRNase J (Figure 4A and B, lower panels). Indeed, the single C124A substitution is sufficient in destabilizing this interaction.

Finally, to confirm the specificity of the interaction between aRNase J and ASH-Ski2, we performed similar *in vitro* co-purification assays using aRNase J as a prey and His-Hel308, another Ski2-like helicase encoded in the *P. abyssi* genome, as bait. In this configuration, aRNase J is not retained on the  $\text{Ni}^{2+}$ -NTA column matrices (Figure 4C) meaning that aRNase J specifically interacts with ASH-Ski2. This result is not surprising as members of the Hel308 family do not possess the N-terminal zinc finger-like motif extension.

#### aRNase J forms a complex with the exosome cap-subunit Csl4 *in vitro*

We also tested the pairwise interactions between aRNase J or ASH-Ski2 and RNA exosome subunits that are among the most retrieved proteins with aRNase J and ASH-Ski2 in the AP-MS analyses (Figure 1B; Supplementary Tables S3 and 4). The results of the *in vitro* co-purification assays performed as described above are summarized in Figure 5A. First, we challenged the catalytic subunit His-Rrp41 as bait with aRNase J or ASH-Ski2 as preys. Only aRNase J could be recovered in the elution fractions together with His-Rrp41. The Rrp41/aRNase J complex is nucleic acid-dependent since its formation is sensitive to nuclease treatment (Figure 5B). Then, we challenged the exosome cap subunits His-Rrp4, DnaG-His or His-Csl4 as baits with aRNase J or ASH-Ski2 as preys. In our conditions, neither Rrp4 nor DnaG were found to interact with aRNase J (Supplementary Figure S2C) or Rrp4, DnaG nor Csl4 with ASH-Ski2 (Supplementary Figure S2D). In contrast, the exosome cap-subunit Csl4 is able to form a complex with aRNase J that is not sensitive to nuclease treatment and therefore is not mediated by nucleic acids (Figure 5C). To go further in the characterization of this association, we then searched for the protein domains of Csl4 that are involved in the formation of the Csl4/aRNase J complex. Whilst the N-terminal domain of Csl4 ( $\Delta$ N) is not required, its C-terminal domain ( $\Delta$ C) is critical for the formation of the Csl4/aRNase J complex (Figure 5C).



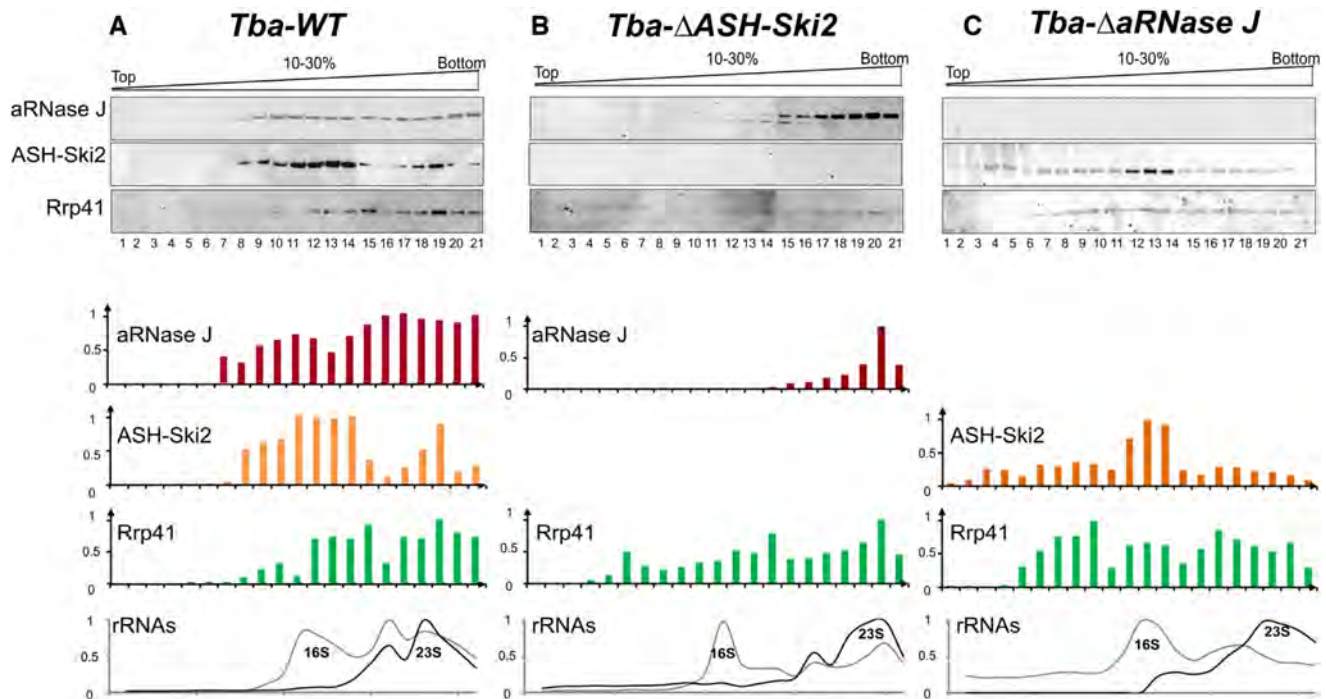
**Figure 5.** *In vitro* interaction of ASH-Ski2 and aRNAse J with the RNA exosome subunits. (A) Schematic representation of the RNA exosome machine and of the pairwise interactions detected in this study. Plain arrows indicate interactions resistant to nuclease treatment. Dotted arrows indicate interactions sensitive to nuclease treatment. The *in vitro* co-purification assays of ASH-Ski2 or aRNAse J challenged by His-Csl4 are shown in (B) and (C), respectively. The schematic representation of the protein domain architecture of Csl4 is shown (see also Supplementary Figure S3B). Wild-type Csl4 (His-WT) and variants with a deletion of the C-terminal (His-ΔC) or N-terminal (His-ΔN) or a punctual substitution of phenylalanine 121 in alanine in the S1 domain (His-F121A) were used as baits. Legend details are as in Figure 4.

To highlight specifically conserved residues in Csl4 sequences that are co-distributed with aRNAse J, we derived weblogos from multiple alignments of Csl4 co-distributed or not with aRNAse J. Several positions (including F/Y121 and F129, numbering according to the *P. abyssi* sequence) in the S1 domain of Csl4 could be identified as highly conserved in Csl4 sequences in the context of the co-distribution with aRNAse J (Supplementary Figure S3A). Strikingly, the phenylalanine residues, F121 and F129, located on the central S1 domain of Csl4 and solvent-exposed in the *P. abyssi* Csl4 structural model (Supplementary Figure S3B) are also key for the association with aRNAse J (Figure 5C).

#### aRNAse J and ASH-Ski2 and the RNA exosome co-sediment with high molecular weight complexes in sucrose gradient fractionation of whole-cell extract from *P. abyssi* and *Thermococcus barophilus*

Altogether, our results propose a tight relationship between aRNAse J, ASH-Ski2 and the RNA exosome. Here, we asked whether these protein partners could form higher or-

der protein complexes in archaeal cells. To this end, we discriminated the sedimentation profile behaviours of endogenous aRNAse J, ASH-Ski2 and Rrp41 in two Thermococcales organisms, *T. barophilus* (*Tba*) and *P. abyssi* (*Pab*) by doing ultracentrifugation of whole-cell extracts on continuous 10–30% sucrose density gradient (Figure 6A and Supplementary Figure S4A, respectively). Since to this date, *P. abyssi* cannot be genetically modified, the *T. barophilus* strain was used as a genetically tractable model (37,38) to assess the sedimentation profile of endogenous aRNAse J in absence of ASH-Ski2 and vice-versa. Moreover, note that the antibodies produced to detect our proteins of interest from *P. abyssi* by immunoblotting cross-react with their counterparts from *T. barophilus* whilst preserving their specificity. Unexpectedly, the vast majority of these proteins co-sediment in heavy fractions with barely any signal in light fractions. This is not the case for the β-CASP endoRNAse *Pab*-aCPSF1, which was not identified in any of the protein interaction networks that we uncovered in this study (Supplementary Figure S4). It is interesting to note that *Pab/Tba*-aRNAse J and *Pab/Tba*-ASH-Ski2 are similarly distributed through the gradient (Figure 6A and Supple-



**Figure 6.** Sedimentation profiles of endogenous aRNase J, ASH-Ski2 and Rrp41 of *Thermococcus barophilus* wild-type (WT) in (A),  $\Delta$ ASH-Ski2 in (B) and  $\Delta$ aRNase J in (C), strains. Clarified whole-cell extracts were fractionated on 10–30% density sucrose gradients by ultracentrifugation. Equal volumes of fraction, precipitated with TCA, were monitored by western blotting using specific antibodies. The digital images and bar charts representing the relative amounts of protein are given. The ribosomal RNAs were also profiled by slot blotting equal volume of denatured fractions with probes complementary to 16S and 23S rRNAs. The relative quantification of 16S and 23S ribosomal RNAs is plotted in grey and black, respectively.

mentary Figure S4). This distribution pattern is also highly similar to the one of the RNA exosome subunit *Pab/Tba-Rrp41* (Figure 6A and Supplementary Figure S4). In conclusion, our data are indicative of a co-migration of the *Pab/Tba-aRNase J*, the *Pab/Tba-ASH-Ski2* and the Rrp41 RNA exosome subunit in high molecular weight fractions.

To go further, we fractionated whole-cell extracts from *T. barophilus* strains enclosing a deletion of the gene encoding *Tba-ASH-Ski2* (*Tba- $\Delta$ ASH-Ski2*, Figure 6B) or *Tba-aRNase J* (*Tba- $\Delta$ aRNase J*, Figure 6C). On one hand, when *Tba-ASH-Ski2* is missing, most of the endogenous *Tba-aRNase J* protein is shifted towards the bottom of the gradient. On the other hand, when *Tba-aRNase J* is not present, the distribution profile of the endogenous *Tba-ASH-Ski2* is also clearly altered; but in this case, *Tba-ASH-Ski2* is more evenly spread out along the gradient. Note that the profile of the core RNA exosome subunit, Rrp41, was barely affected in both genomic contexts. Overall, these results strongly suggest that the aRNase J & ASH-Ski2 are part of higher-order complexes in *T. barophilus* cells. In addition, we observe that the 16S and 23S ribosomal RNA profiles overlap the ones of aRNase J, ASH-Ski2 and Rrp41. In absence of ASH-Ski2, aRNase J is shifted towards the ribosome. From these results and in light of data reported for bacterial RNase J and eukaryal Ski2 helicases (64,65), we propose that these two enzymes cooperate with the ribosome in Thermococcale cells.

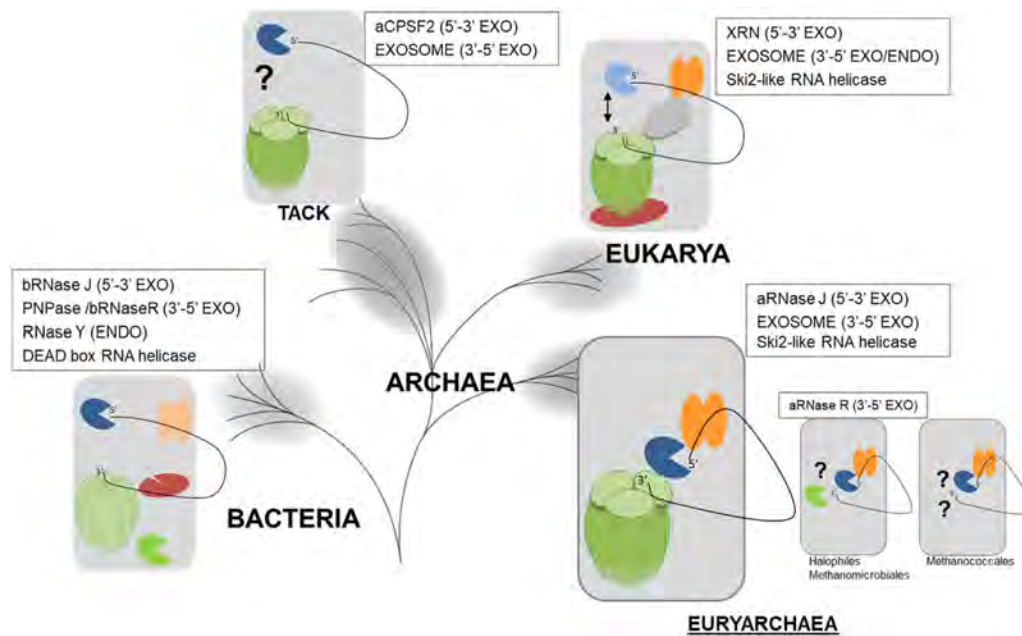
We also challenged the physical association of *Tba-aRNase J* and *Tba-ASH-Ski2* in higher-order complexes, by performing similar fractionation assays but at low mag-

nesium concentration (TK-EDTA buffer) (Supplementary Figure S5). In these conditions, the assembled ribosomes dissociate into individual subunits (66). In the wild-type strain, endogenous *Tba-aRNase J* and *Tba-ASH-Ski2* co-sediment with the 16S and 23S rRNAs from the small and large subunits, respectively (Supplementary Figure S5A). Again, we observe a perfect overlay of the sedimentation profiles of *Tba-aRNase J* and *Tba-ASH-Ski2* as previously seen at a higher concentration of magnesium (Figure 6A and Supplementary Figure S5A). Similar results were obtained with extracts from *P. abyssi* cells (Supplementary Figure S4B). More interestingly, in absence of crosstalk with ASH-Ski2, a flagrant alteration of the sedimentation behaviour of *Tba-aRNase J* was also detected. Indeed, in the context of the *Tba- $\Delta$ ASH-Ski2* strain, at low magnesium concentration, *Tba-aRNase J* is shifted from the bottom of the gradient to fractions overlapping the profiles of the 16S and 23S rRNAs (Supplementary Figure S5B). In absence of *Tba-aRNase J*, we observe overlay sedimentation profiles of *Tba-ASH-Ski2* and *Tba-Rrp41* (Supplementary Figure S5C). From these results, we argue for an association of aRNase J with ribosome/polysome that is emphasised in absence of ASH-Ski2 in addition to an interplay between aRNase J, ASH-Ski2 and Rrp41 RNA exosome subunit.

## DISCUSSION

Elucidating the RNA processing machineries is of a critical importance to understand the molecular basis of RNA decay and processing in Euryarchaeota. In this study, to





**Figure 7.** Schematic representation of the interplay between enzymes and machineries in charge of 5'-3' and 3'-5' RNA decay and surveillance in the three domains of life. The life tree model is as in (84). The 3' end RNA trimming activity is carried by the RNA exosome machinery (barrel in green and cap in light green) in most Archaea (15) and Eukarya (76) that is similar to bacterial PNPase (barrel in light green). In Halobacteria and Methanomicrobiales the aRNase R, orthologous to bRNase R, (green packman) carries the 3'-5' exo activity. In Methanococci, no 3'-5' exo activity has yet been described (15). The 5'-3' exoribonucleolytic activities (blue packman) are carried by XRN RNases in Eukarya (85), bRNase J and aRNase J in Bacteria and Euryarchaea and aCPSF2 in the TACK group. RNA helicase activities shown to be involved in RNA surveillance belong to the Ski2-like family in Eukarya (42) and Euryarchaea (this study) and to the DEAD box family in Bacteria (86). The interplay between 5'-3' and 3'-5' exoribonucleolytic activities are reported to take place in the RNA-degradosome-like complex in Bacteria that also involves endoRNases (red packman) (87) and through the 5'-3' communication by a 'closed-loop' architecture of mRNA in Eukarya (88). In most Euryarchaea, we propose that aRNase J/ASH-Ski2 complex interplays with the RNA exosome. In the TACK group and the Halophiles, Methanomicrobiales and Methanococcales, such an interplay remains to be characterized (question marks).

our knowledge, we provide the first experimental evidences hinting at the existence of high-order  $\beta$ -CASP protein complex(es) in Thermococcales that would carry 5'-3' & 3'-5' exoribonucleolytic and helicase-like activities and that could play pivotal roles in archaeal RNA surveillance. First, we identified the interaction landscape of aRNase J from *P. abyssi*, a 5' to 3'  $\beta$ -CASP exoRNase (32). Candidate partners include, notably, a Ski2-like helicase ASH-Ski2, the archaeal RNA exosome that exhibits 3' to 5' exoribonucleolytic and 3' end-tailing activities (17) and ribosomal proteins. Of high significance, we also found these proteins to be candidate partners in the interaction landscape of ASH-Ski2 and in the pull-down of the core exosome subunit Rrp41 (Figures 1B and 2). In addition, we showed that aRNase J is able to interact, *in vitro*, with ASH-Ski2 and the exosome cap subunit Csl4 (Figures 4 and 5). Our *in vitro* data are strengthened by our phylogenomic studies showing that, with only one exception, ASH-Ski2 is only encoded in archaeal genomes that carry a gene for aRNase J and that amino acids of Csl4, involved in the *in vitro* interaction with aRNase J, are only conserved in archaeal genomes where their encoding genes are co-distributed (Figure 3). Finally, our *T. barophilus* whole-cell extract fractionation experiments provide evidences that an aRNase J/ASH-Ski2 complex might exist *in vivo* as their sequential deletions modify the sedimentation profile of their protein partner (Figure 6). Individually, aRNase J and ASH-Ski2 do not appear to

be essential in *T. barophilus* in optimal growth conditions. However, we could not obtain a strain that combined both gene deletions. This synthetic lethality as demonstrated in yeast for genes involved in same RNA degradation pathways (67,68) strongly hints at a functional relationship between aRNase J and ASH-Ski2 in Thermococcales.

In conjunction with our previous demonstration that  $\beta$ -CASP RNases, ubiquitous in Archaea, must be key players in archaeal RNA metabolism (26,27), our current data support a model whereby the  $\beta$ -CASP RNase, aRNase J, the ASH-Ski2-like helicase and the RNA exosome interplay in most euryarchaeal cells (Figure 7). It remains to determine how this interplay could impact the archaeal RNA metabolism. Since our data imply that ASH-Ski2 has an impact on aRNase J and *vice versa*, we favour a scenario where, at least, ASH-Ski2 and aRNase J have a functional connection. This is all the more so given that aRNase J and ASH-Ski2 show a taxonomic co-distribution amongst the archaeal phylogeny and nearly perfect phylogenetic tree congruence suggesting a co-evolution and potentially an ancient origin (Figure 3B and Supplementary Figure S1, respectively). Based on what is known about the role of these factors in Eukarya and Bacteria, we discuss what we believe to be the most probable scenarios. Evidently, further work will be needed to determine how these proteins alone or in complex shape the euryarchaeal RNA metabolism.



First, recent published work proposes that ASH-Ski2, also named Eta for euryarchaeal termination activity, influences transcription termination by disrupting archaeal transcription elongation complex (69). However, as its helicase activity is rather slow and cannot keep up with wild-type RNAP elongation rates, Eta was recently proposed to be more likely involved in the rescue of transcriptional arrest (70). In our study, the presence of RPA and RpoB subunits in the interaction network of ASH-Ski2 could be a further evidence for the involvement of ASH-Ski2 in the transcription. In addition, transcriptional termination and rescue pathways also involve proteins with ribonucleolytic activities. In Eukarya, the 5'-3' exoribonucleolytic activity of Rat1p/Xrn2p was proposed to be associated in a transcription termination pathway called 'torpedo model' (71). In chloroplast, RNase J was shown to play a critical role in removing antisense RNAs that are generated by read-through transcription (72). Nevertheless, chloroplast RNase J exhibits a DNA binding domain found in transcription factors of plants that is not present in archaeal RNase J and that could be essential to fulfil this function (73). Finally, co-factors of the nuclear RNA exosome were shown to connect transcription termination to RNA processing by guiding terminated transcripts to the appropriate exonuclease (74). In Archaea, it is still not known if a ribonucleolytic activity is critical since, until now, the complete picture of transcription termination is far from being understood (70). It remains to identify if aRNase J, ASH-Ski2 and the RNA exosome are involved together or alone in transcription termination or rescue pathways which have been shown to be key in influencing the fate of a transcribed RNA.

Moreover, based on the roles of the Ski2-like helicases in Eukarya (75), the idea that the euryarchaeal ASH-Ski2 group is closely connected to RNA-related pathways emerges (Figure 7). Indeed, eukaryal Ski2, Mtr4 and Brr2 which are described as RNA helicases that process RNAs in a 3'-5' direction were shown to play critical roles in RNA decay, surveillance and splicing, respectively (54,75). Notably, Ski2 and Mtr4 are essential co-factors of the cytoplasmic and nuclear RNA exosomes, respectively, by being involved in the recruitment of RNA substrates (13). Likewise, our work provides several evidences of interplay between the archaeal RNA exosome and ASH-Ski2. Our data also suggest that this interplay is done through aRNase J that interacts *in vitro* with Csl4 from the RNA exosome cap and with ASH-Ski2. Altogether, our results converge towards a cross-talk between the 5'-3' and 3'-5' exo-ribonucleolytic activities carried by aRNase J and the RNA exosome, respectively. This is consistent with what is observed in Eukarya, where mounting evidences show a functional interplay between surveillance pathways taking place at the 5' and 3'-ends of RNAs (76) (Figure 7).

Finally, it must be remembered that aRNase J is the homologue of bacterial RNase J (bRNase J) which was shown to be critical in mRNA decay and rRNA maturation pathways and to be part of a degradosome-like complex (29,77-78). bRNase J which can carry 5'-3' exo- and endo-RNase activities was shown to associate with proteins identified as, among others, helicase-like proteins and a polynucleotide phosphorylase (PNPase). Bacterial PNPase form an RNA-degradation machine with 3'-5' exori-

bonucleolytic and polyadenylation activities reminiscent to the archaeal RNA exosome (79) (Figure 7). Although, in some organisms, the bRNase J degradosome-like complex appears to be transient and dynamic, limiting its validation *in vitro*, this is consistent with interplay between 5'-3' and 3'-5' RNA surveillance decay in Bacteria. Therefore, it emerges that, in the three domains of life, RNA surveillance could be taken care of by evolutionary conserved RNase-based RNA-degradation machines with RNA helicase and with 5'-3' and 3'-5' exoribonucleolytic activities (Figure 7).

From cellular fractionation assays, we observed that both aRNase J and ASH-Ski2 could potentially associate with the ribosome. Interestingly, it was also shown that bRNase J can associate with translating ribosomes (64) and the crystal structure of the eukaryotic 80S ribosome-Xrn1 complex showed that the 5'-3' Xrn1 nuclease binds at the mRNA exit site of the ribosome (80). Since the sedimentation profiles of *T. barophilus* and *P. abyssi* aRNase J resemble to the one reported for the eukaryal 5'-3' exo-RNase Xrn1, we speculate that the archaeal aRNase J is also associated with polysomes in Thermococcales cells. This is reminiscent of what is described in eukaryal mRNA surveillance pathways (28) in which mRNA decay is initiated by an endonucleolytic cleavage of ribosome-associated mRNAs (81,82), followed by rapid degradation in the 3'-5' direction by the RNA exosome or in the 5'-3' direction by Xrn1 (83). Recently, direct experimental evidence showed that the cytoplasmic RNA exosome is able to interact with the ribosome through a direct interaction mediated by the SKI complex (composed of Ski2-Ski3-Ski8 factors). More precisely, it was shown that ribosome binding displaces the auto-inhibitory domain of the Ski2 RNA helicase (65).

On a side note, it is important to notice that the RNA exosome is not present throughout the whole archaeal taxonomy (15,27) and that aRNase J does not accurately follow the taxonomic distribution of RNA exosome components. In the Halophiles that do not encode the exosome subunits, the 3'-5' exo-ribonucleolytic activity is carried by another enzyme, the aRNase R (14,15). It would be interesting to know whether the halophilic and methanomicrobiale aRNase J, alone or in complex with ASH-Ski2 complex, interplays with aRNase R. In the Methanococci group, no 3'-5' exo-RNase could be detected by comparative genomics. In this case, aRNase J might connect to specific archaeal RNases that remain to be identified (Figure 7). In conclusion, this work establishes a solid base in the largely overlooked but extremely important field of archaeal RNA metabolism. We identified two phylogenetical conserved factors in Euryarchaea, the 5'-3' exo-RNase aRNase J and the Ski2-like helicase ASH-Ski2, which have homologues in Bacteria and Eukarya, respectively (Figure 7). Therefore, a mosaic system in Euryarchaeota could be at the centre of networks interplaying with the archaeal RNA exosome and the ribosome. This is an example showing that Archaea use composite systems and thereby elucidating mechanism in archaeal organisms could give valuable insights into understanding function and evolutionary routes of protein networks or/and high-order complexes. Analogous to eukaryal and bacterial systems, we propose that these factors act in partnership in major archaeal RNA decay or surveillance pathways. Deeper insights need to be

gained into the function and importance of these enzymes in mRNA decay, ribosomal RNA maturation, mRNA 3'-end formation and transcription termination. It remains, among other aspects, to be established how these factors might be involved in shaping the RNA landscape in euryarchaeal cells. Finally, the settlement of specific enzymatic properties of ASH-Ski2 and the determination of structure-function of ASH-Ski2/aRNase J and Csl4/aRNase J complexes will ascertain their physiological functions in the context of the interplay with aRNase J.

## DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015856.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are indebted to L. Correia of the Paris Sud Ouest PAPPSO proteomics core facility (<http://papsu.inra.fr>) which is supported by INRA (<http://www.inra.fr>), the Ile-de-France regional council (<https://www.iledefrance.fr/education-recherche>), IBI SA (<https://www.ibisa.net>) and CNRS (<http://www.cnrs.fr>) for LC/MS analyses and to L. Correia. We thank Y. Quentin for its expertise in taxonomic identification of archaeal Csl4 and Rrp41 members, L. Plassart and J. Caumes for technical help and P. Vitali and M. Kwapisz for helpful discussions.

## FUNDING

CNRS; University Paul Sabatier; University of Western Brittany; Ifremer; University of Toulouse, Idex-emergence program (to B.C.O.); French Ministère de l'Enseignement Supérieure et de la Recherche PhD Fellowship (to D.K.P and C.E.); French 'Agence Nationale pour la Recherche' [ANR-16-CE12-0016 to B.C.O]; Funding for open access charge: ANR [ANR-16-CE12-0016-01].

*Conflict of interest statement.* None declared

## REFERENCES

- Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C. and Schmitz, R.A. (2018) Archaea are interactive components of complex microbiomes. *Trends Microbiol.*, **26**, 70–85.
- Adam, P.S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. (2017) The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.*, **11**, 2407–2425.
- Brochier-Armanet, C., Forterre, P. and Gribaldo, S. (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.*, **14**, 274–281.
- Da Cunha, V., Gaia, M., Gabelle, D., Nasir, A. and Forterre, P. (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.*, **13**, e1006810.
- Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. (2018) Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.*, **16**, 120.
- Koonin, E.V. (2015) Archaeal ancestors of eukaryotes: not so elusive any more. *BMC Biol.*, **13**, 84.
- Van der Gulik, P.T.S., Hoff, W.D. and Speijer, D. (2017) In defence of the three-domains of life paradigm. *BMC Evol. Biol.*, **17**, 218.
- Raymann, K., Brochier-Armanet, C. and Gribaldo, S. (2015) The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6670–6675.
- Yutin, N., Puigbò, P., Koonin, E.V. and Wolf, Y.I. (2012) Phylogenomics of prokaryotic ribosomal proteins. *PLoS One*, **7**, e36972.
- Ferreira-Cerca, S. (2017) Life and death of ribosomes in archaea. In: Clouet-d'Orval, B. (ed). *RNA Metabolism and Gene Expression in Archaea, Nucleic Acids And Molecular Biology*. Springer International Publishing, Cham, Switzerland, Vol. **32**, pp. 129–158.
- Werner, F. (2013) Molecular mechanisms of transcription elongation in archaea. *Chem. Rev.*, **113**, 8331–8349.
- Fouqueau, T. and Werner, F. (2017) The architecture of transcription elongation. *Science*, **357**, 871–872.
- Chlebowski, A., Lubas, M., Jensen, T.H. and Dziembowski, A. (2013) RNA decay machines: the exosome. *Biochim. Biophys. Acta*, **1829**, 552–560.
- Portnoy, V. and Schuster, G. (2006) RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Res.*, **34**, 5923–5931.
- Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A. and Maier, L.-K. (2018) Insights into RNA processing pathways and associated-RNA degrading enzymes in Archaea. *FEMS Microbiol. Rev.*, **42**, 579–613.
- Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.*, **12**, 575–581.
- Evguenieva-Hackenberg, E., Hou, L., Glaeser, S. and Klug, G. (2014) Structure and function of the archaeal exosome. *Wiley Interdiscip. Rev. RNA*, **5**, 623–635.
- Koonin, E.V., Wolf, Y.I. and Aravind, L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.*, **11**, 240–252.
- Hou, L., Klug, G. and Evguenieva-Hackenberg, E. (2014) Archaeal DnaG contains a conserved N-terminal RNA-binding domain and enables tailing of rRNA by the exosome. *Nucleic Acids Res.*, **42**, 12691–12706.
- Hou, L., Klug, G. and Evguenieva-Hackenberg, E. (2013) The archaeal DnaG protein needs Csl4 for binding to the exosome and enhances its interaction with adenine-rich RNAs. *RNA Biol.*, **10**, 415–424.
- Märtens, B., Hou, L., Amman, F., Wolfinger, M.T., Evguenieva-Hackenberg, E. and Bläsi, U. (2017) The SmAP1/2 proteins of the crenarchaeon *Sulfolobus solfataricus* interact with the exosome and stimulate A-rich tailing of transcripts. *Nucleic Acids Res.*, **45**, 7938–7949.
- Evguenieva-Hackenberg, E., Roppelt, V., Finsterseifer, P. and Klug, G. (2008) Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome. *Biochemistry*, **47**, 13158–13168.
- Witharana, C., Roppelt, V., Lochnit, G., Klug, G. and Evguenieva-Hackenberg, E. (2012) Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie*, **94**, 1578–1587.
- Dominski, Z., Carpousis, A.J. and Clouet-d'Orval, B. (2013) Emergence of the  $\beta$ -CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim. Biophys. Acta*, **1829**, 532–551.
- Condon, C. and Gilet, L. (2011) The metallo. In: *Nucleic Acids and Molecular Biology*. Springer-verlag, Berlin, Heidelberg, Vol. **26**, pp. 245–267.
- Phung, D.K., Rinaldi, D., Langendijk-Genevaux, P.S., Quentin, Y., Carpousis, A.J. and Clouet-d'Orval, B. (2013) Archaeal  $\beta$ -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res.*, **41**, 1091–1103.
- Clouet-d'Orval, B., Phung, D.K., Langendijk-Genevaux, P.S. and Quentin, Y. (2015) Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of  $\beta$ -CASP ribonucleases. *Biochimie*, **118**, 278–285.
- Roy, B. and Jacobson, A. (2013) The intimate relationships of mRNA decay and translation. *Trends Genet.*, **29**, 691–699.

29. Lehnik-Habrink, M., Lewis, R.J., Mäder, U. and Stülke, J. (2012) RNA degradation in *Bacillus subtilis*: an interplay of essential endo- and exoribonucleases. *Mol. Microbiol.*, **84**, 1005–1017.
30. Hasenöhr, D., Lombo, T., Kaberdin, V., Londei, P. and Bläsi, U. (2008) Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 2146–2150.
31. Märten, B., Amman, F., Manoharadas, S., Zeichen, L., Orell, A., Albers, S.-V., Hofacker, I. and Bläsi, U. (2013) Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *PLoS One*, **8**, e76569.
32. Clouet-d'Orval, B., Rinaldi, D., Quentin, Y. and Carpousis, A.J. (2010) Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *J. Biol. Chem.*, **285**, 17574–17583.
33. Hasenöhr, D., Konrat, R. and Bläsi, U. (2011) Identification of an RNase J ortholog in *Sulfolobus solfataricus*: implications for 5'-to-3' directional decay and 5'-end protection of mRNA in Crenarchaeota. *RNA*, **17**, 99–107.
34. Arkhipova, V., Stolboushkina, E., Kravchenko, O., Kljashtorny, V., Gabdulkhakov, A., Garber, M., Nikonov, S., Märten, B., Bläsi, U. and Nikonov, O. (2015) Binding of the 5'-Triphosphate End of mRNA to the  $\gamma$ -Subunit of Translation Initiation Factor 2 of the Crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol.*, **427**, 3086–3095.
35. Condon, C. and Putzer, H. (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res.*, **30**, 5339–5346.
36. Condon, C. (2010) What is the role of RNase J in mRNA turnover? *RNA Biol.*, **7**, 316–321.
37. Birien, T., Thiel, A., Henneke, G., Flament, D., Moalic, Y. and Jebbar, M. (2018) Development of an effective 6-methylpurine counterselection marker for genetic manipulation in *thermococcus barophilus*. *Genes (Basel)*, **9**, E77.
38. Thiel, A., Michoud, G., Moalic, Y., Flament, D. and Jebbar, M. (2014) Genetic manipulations of the hyperthermophilic piezophilic archaeon *Thermococcus barophilus*. *Appl. Environ. Microbiol.*, **80**, 2299–2306.
39. Pluchon, P.-F., Fouqueau, T., Crezéc, C., Laurent, S., Briffotiaux, J., Hogrel, G., Palud, A., Henneke, G., Godfroy, A., Hausner, W. *et al.* (2013) An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs. *PLoS One*, **8**, e79707.
40. Branson, O.E. and Freitas, M.A. (2016) Tag-count analysis of large-scale proteomic data. *J. Proteome Res.*, **15**, 4742–4746.
41. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
42. Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.
43. Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.
44. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
45. Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
46. Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.
47. Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, **27**, 1164–1165.
48. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
49. Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*, **44**, W242–W255.
50. Godfroy, A., Raven, N.D. and Sharp, R.J. (2000) Physiology and continuous culture of the hyperthermophilic deep-sea vent archaeon *Pyrococcus abyssi* ST549. *FEMS Microbiol. Lett.*, **186**, 127–132.
51. Omer, A.D., Ziesche, S., Ebhardt, H. and Dennis, P.P. (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5289–5294.
52. Bortolin, M.-L., Bachelier, J.-P. and Clouet-d'Orval, B. (2003) In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA(Trp). *Nucleic Acids Res.*, **31**, 6524–6535.
53. Chamieh, H., Ibrahim, H. and Kozah, J. (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene*, **576**, 214–228.
54. Johnson, S.J. and Jackson, R.N. (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol.*, **10**, 33–43.
55. Fairman-Williams, M.E., Guenther, U.-P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
56. Pyle, A.M. (2011) RNA helicases and remodeling proteins. *Curr. Opin. Chem. Biol.*, **15**, 636–642.
57. Weir, J.R., Bonneau, F., Hentschel, J. and Conti, E. (2010) Structural analysis reveals the characteristic features of Mtr4, a DEXH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12139–12144.
58. Guy, C.P. and Bolt, E.L. (2005) Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res.*, **33**, 3678–3690.
59. Fujikane, R., Komori, K., Shinagawa, H. and Ishino, Y. (2005) Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J. Biol. Chem.*, **280**, 12351–12358.
60. Li, Z., Lu, S., Hou, G., Ma, X., Sheng, D., Ni, J. and Shen, Y. (2008) Hjm/Hel308A DNA helicase from *Sulfolobus tokodaii* promotes replication fork regression and interacts with Hjc endonuclease in vitro. *J. Bacteriol.*, **190**, 3006–3017.
61. De Felice, M., Aria, V., Esposito, L., De Falco, M., Pucci, B., Rossi, M. and Pisani, F.M. (2007) A novel DNA helicase with strand-annealing activity from the crenarchaeon *Sulfolobus solfataricus*. *Biochem. J.*, **408**, 87–95.
62. Krishna, S.S., Majumdar, I. and Grishin, N.V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.*, **31**, 532–550.
63. Zheng, X., Feng, N., Li, D., Dong, X. and Li, J. (2017) New molecular insights into an archaeal RNase J reveal a conserved processive exoribonucleolytic mechanism of the RNase J family. *Mol. Microbiol.*, **106**, 351–366.
64. Redko, Y., Aubert, S., Stachowicz, A., Lenormand, P., Namane, A., Darfeuille, F., Thibonnier, M. and De Reuse, H. (2013) A minimal bacterial RNase J-based degradosome is associated with translating ribosomes. *Nucleic Acids Res.*, **41**, 288–301.
65. Schmidt, C., Kowalinski, E., Shanmuganathan, V., Defenouillère, Q., Braunger, K., Heuer, A., Pech, M., Namane, A., Berninghausen, O., Fromont-Racine, M. *et al.* (2016) The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. *Science*, **354**, 1431–1433.
66. Sharrock, W.J. and Rabinowitz, J.C. (1979) Fractionation of ribosomal particles from *Bacillus subtilis*. *Meth. Enzymol.*, **59**, 371–382.
67. Van Hoof, A., Staples, R.R., Baker, R.E. and Parker, R. (2000) Function of the ski4p (Csl4p) and ski7p proteins in 3'-to-5' degradation of mRNA. *Mol. Cell Biol.*, **20**, 8230–8243.
68. Johnson, A.W. and Kolodner, R.D. (1995) Synthetic lethality of sep1 (xrn1) ski2 and sep1 (xrn1) ski3 mutants of *Saccharomyces cerevisiae* is independent of killer virus and suggests a general role for these genes in translation control. *Mol. Cell Biol.*, **15**, 2719–2727.
69. Walker, J.E., Luyties, O. and Santangelo, T.J. (2017) Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E6767–E6773.
70. Maier, L.-K. and Marchfelder, A. (2019) It's all about the T: transcription termination in archaea. *Biochem. Soc. Trans.*, **47**, 461–468.
71. Rosonina, E., Kaneko, S. and Manley, J.L. (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev.*, **20**, 1050–1056.
72. Sharwood, R.E., Halpert, M., Luro, S., Schuster, G. and Stern, D.B. (2011) Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA*, **17**, 2165–2176.
73. Halpert, M., Liveanu, V., Glaser, F. and Schuster, G. (2019) The Arabidopsis chloroplast RNase J displays both exo- and robust endonucleolytic activities. *Plant Mol. Biol.*, **99**, 17–29.



74. Kim, K., Heo, D.-H., Kim, I., Suh, J.-Y. and Kim, M. (2016) Exosome Cofactors connect transcription termination to RNA processing by guiding terminated transcripts to the appropriate exonuclease within the nuclear exosome. *J. Biol. Chem.*, **291**, 13229–13242.
75. Jarmoskaite, I. and Russell, R. (2014) RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.*, **83**, 697–725.
76. Łabno, A., Tomecki, R. and Dziembowski, A. (2016) Cytoplasmic RNA decay pathways—enzymes and mechanisms. *Biochim. Biophys. Acta*, **1863**, 3125–3147.
77. Roux, C.M., DeMuth, J.P. and Dunman, P.M. (2011) Characterization of components of the *Staphylococcus aureus* mRNA degradosome holoenzyme-like complex. *J. Bacteriol.*, **193**, 5520–5526.
78. Redko, Y., Galtier, E., Arnion, H., Darfeuille, F., Sismeiro, O., Coppée, J.-Y., Médigue, C., Weiman, M., Cruveiller, S. and De Reuse, H. (2016) RNase J depletion leads to massive changes in mRNA abundance in *Helicobacter pylori*. *RNA Biol.*, **13**, 243–253.
79. Raj, R., Mitra, S. and Gopal, B. (2018) Characterization of *Staphylococcus epidermidis* Polynucleotide phosphorylase and its interactions with ribonucleases RNase J1 and RNase J2. *Biochem. Biophys. Res. Commun.*, **495**, 2078–2084.
80. Tesina, P., Heckel, E., Cheng, J., Fromont-Racine, M., Buschauer, R., Kater, L., Beatrix, B., Berninghausen, O., Jacquier, A., Becker, T. *et al.* (2019) Structure of the 80S ribosome-Xrn1 nuclease complex. *Nat. Struct. Mol. Biol.*, **26**, 275–280.
81. Shoemaker, C.J. and Green, R. (2012) Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.*, **19**, 594–601.
82. Doma, M.K. and Parker, R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561–564.
83. Parker, R. (2012) RNA degradation in *Saccharomyces cerevisiae*. *Genetics*, **191**, 671–702.
84. Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. (2017) Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.*, **15**, 711–723.
85. Nagarajan, V.K., Jones, C.I., Newbury, S.F. and Green, P.J. (2013) XRN 5'→3' exoribonucleases: structure, mechanisms and functions. *Biochim. Biophys. Acta*, **1829**, 590–603.
86. Khemici, V. and Linder, P. (2018) RNA helicases in RNA decay. *Biochem. Soc. Trans.*, **46**, 163–172.
87. Redder, P. (2018) Molecular and genetic interactions of the RNA degradation machineries in Firmicute bacteria. *Wiley Interdiscip. Rev. RNA*, **9**, doi:10.1002/wrna.1460.
88. Vicens, Q., Kieft, J.S. and Rissland, O.S. (2018) Revisiting the closed-loop model and the nature of mRNA 5'-3' communication. *Mol. Cell*, **72**, 805–812.

## SUPPLEMENTARY DATA

# RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the $\beta$ -CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic RNA exosome machinery

Duy Khanh Phung<sup>1§</sup>, Clarisse Etienne<sup>1§</sup>, Manon Batista<sup>1</sup>, Petra Langendijk-Genevaux<sup>1</sup>, Yann Moalic<sup>2</sup>, Sébastien Laurent<sup>2</sup>, Sophie Liuu<sup>3</sup>, Violette Morales<sup>1</sup>, Mohamed Jebbar<sup>2</sup>, Gwennaele Fichant<sup>1</sup>, Marie Bouvier<sup>1</sup>, Didier Flament<sup>2</sup> and Béatrice Clouet-d'Orval<sup>1\*</sup>

<sup>1</sup> Laboratoire de Microbiologie et de Génétique Moléculaires, UMR5100, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse, France

<sup>2</sup> Ifremer, Univ Brest, CNRS, Laboratoire de Microbiologie des Environnements Extrêmes, F-29280 Plouzané, France

<sup>3</sup> Micalis Institute, PAPPSO, INRA, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France

<sup>§</sup>These authors have participated equally to this work

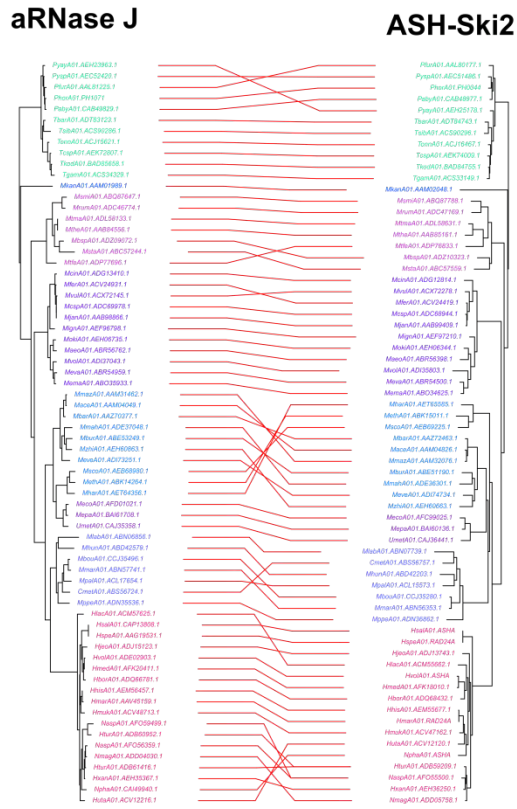
\*To whom correspondence should be addressed. Tel: (33)561335875 Email: [Beatrice.Clouet-dOrval@ibcg.biotoul.fr](mailto:Beatrice.Clouet-dOrval@ibcg.biotoul.fr)

Present Address [Duy Khanh Phung] University College London, Gower Street, Darwin Building, London WC1E 6BT United Kingdom

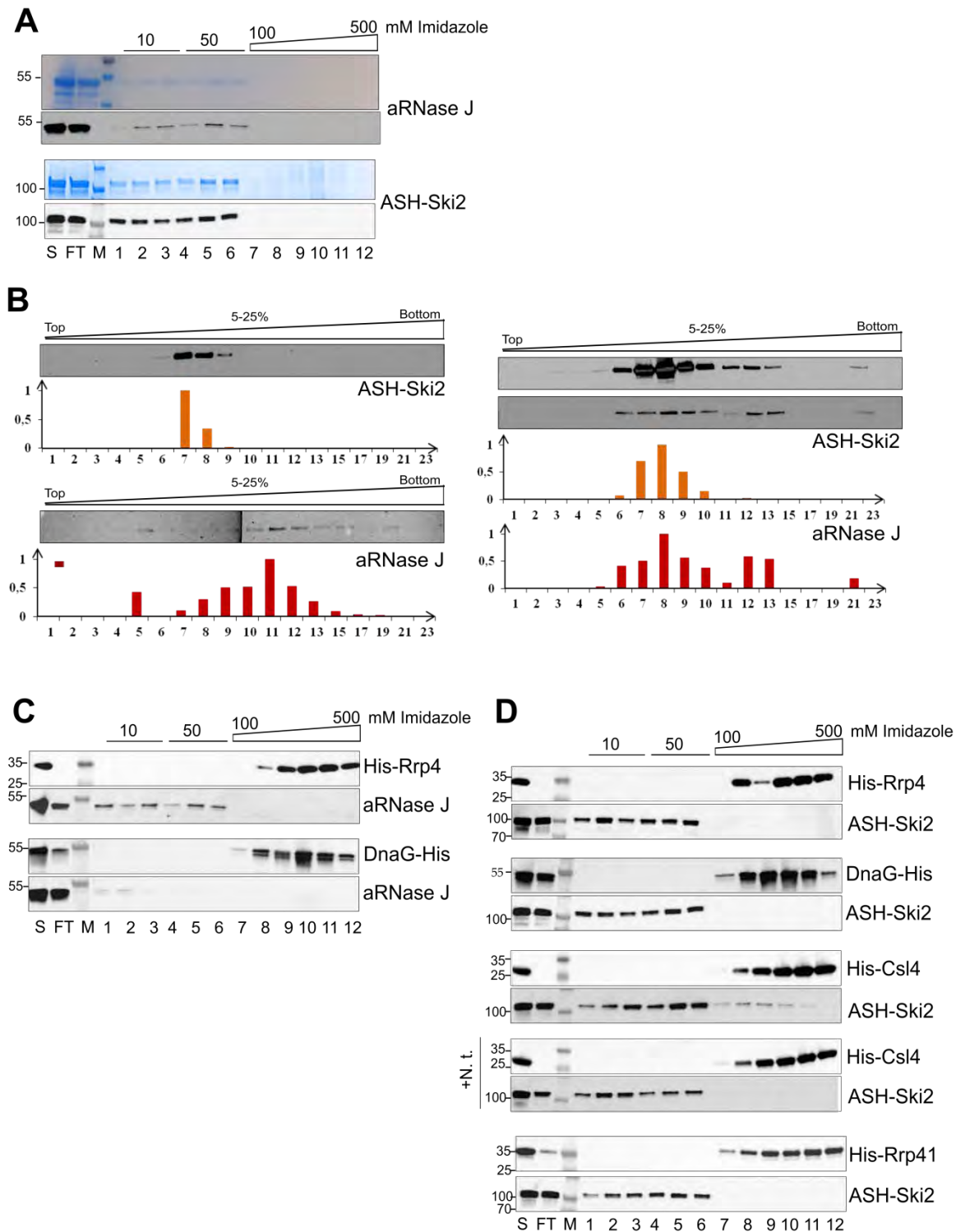
## CONTENT:

Supplementary Figure S1-S5 and legends for Figure S1-S5

Supplementary Table S1-S4 and legends for Table S1-S4



**FIGURE S1-** The aRNase J and ASH-Ski2 phylogenetic trees are congruent. Tips from the same genome are linked by a red line. Colour code for taxonomic order as in Figure 4

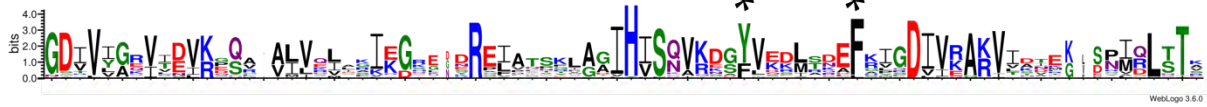


**FIGURE S2- *In vitro* co-purification assays. (A)** Affinity purification controls. Untagged aRNase J (top) and *Pab*-ASH-Ski2 (bottom) are not intrinsically retained on nickel column matrices. **(B)** Sedimentation profiles of *Pab*-aRNase J-(His)<sub>6</sub> (2μM) and *Pab*-(His)<sub>6</sub>-ASH-Ski2 (1μM) on linear PBS1X 5-25% sucrose gradients (200 000g during 20h). Upper panel: purified protein alone. Lower panel: mixture of both proteins. Proteins were separated on a 4-15% SDS-polyacrylamide gel and monitored by Western blot using specific antibodies. **(C)** Co-purification assays in which *P. abyssi* aRNase J-(prey) was challenged with *Pab*-(His)<sub>6</sub>-Rrp4 and *Pab*-DnaG-(His)<sub>6</sub> (baits) respectively. **(D)** Co-purification assays in which *P. abyssi* ASH-Ski2 (prey) was challenged with *Pab*-(His)<sub>6</sub>-Rrp4, *Pab*-DnaG-(His)<sub>6</sub>, *Pab*-(His)<sub>6</sub>-Csl4, *Pab*-(His)<sub>6</sub>-Rrp41 and *Pab*-DnaG-(His)<sub>6</sub> (baits) respectively. Legend as in Figure 6

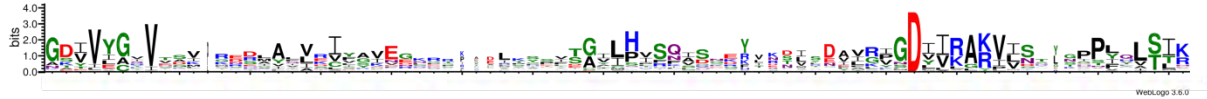
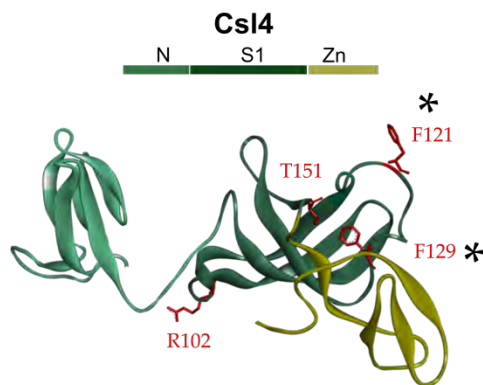


**A****Csl4 S1 domain motifs**

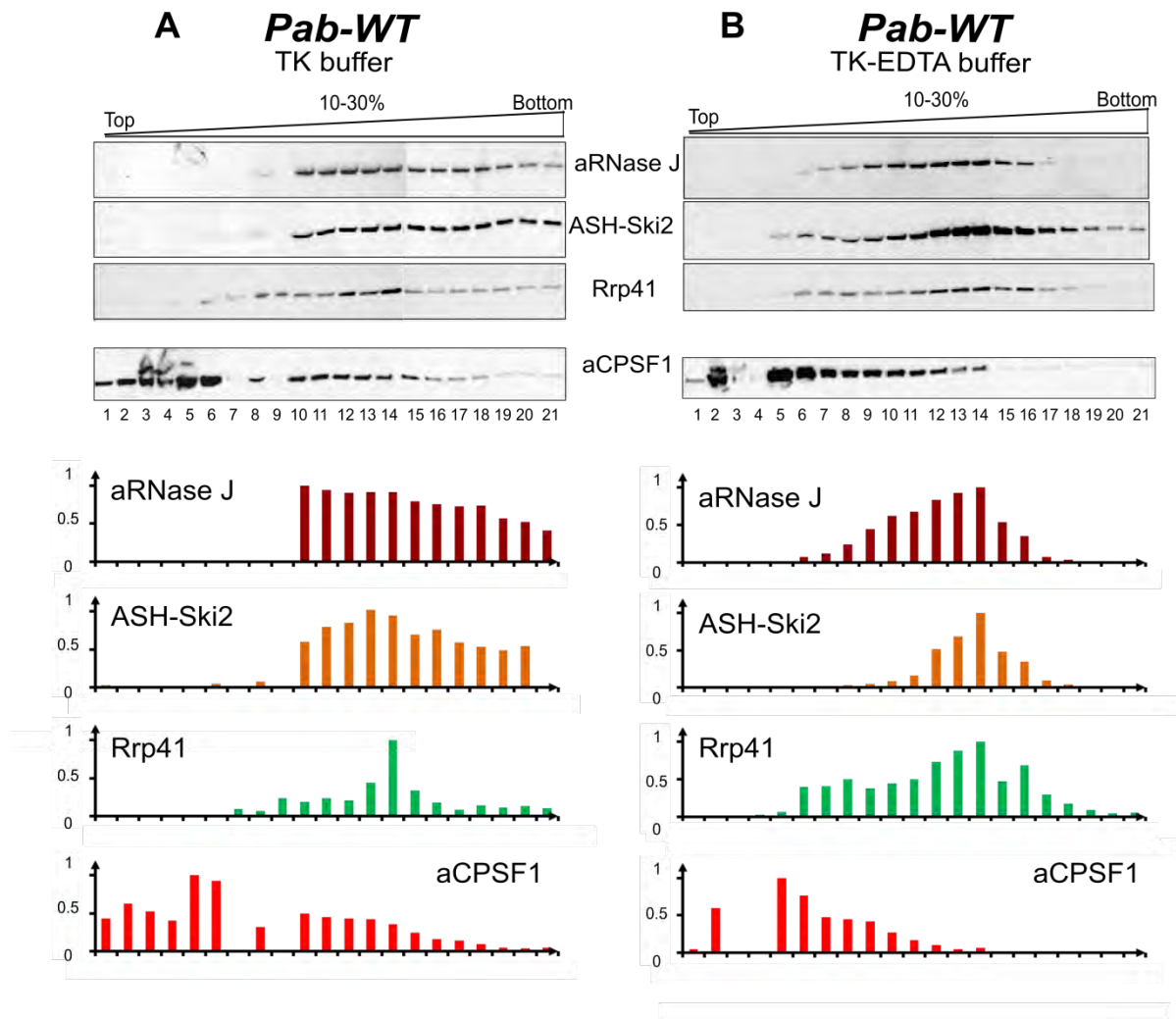
co-distributed with aRNase J



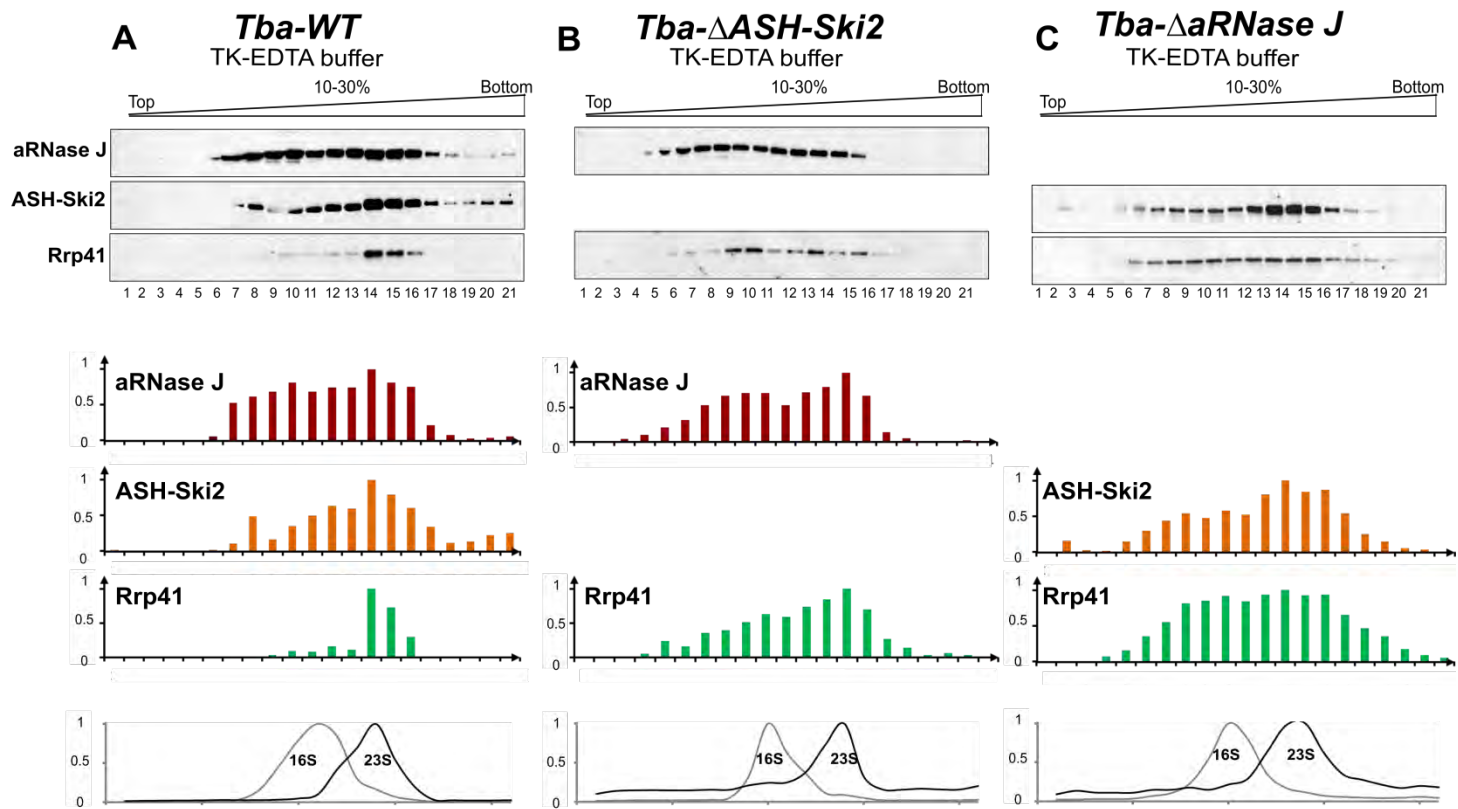
not distributed with aRNase J

**B**

**FIGURE S3-** Csl4 structural domains. **(A)** Weblogo sequences derived from multiple alignments of Csl4 sequences co-distributed or not with aRNase J (referred to distribution of aRNase J & Csl4 displayed on Fig. 4). Several residues are specifically conserved in Csl4 sequences co-distributed with aRNase J. **(B)** The structural model for *P. abyssi* Csl4 is built using Phyre2 software. The solvent-exposed phenylalanine residues F121 and F129 are indicated by stars. S1 is for S1 domain and Zn for Zinc Finger motif containing domain.



**FIGURE S4-** Sedimentation profiles of *Pab*-aRNase J, *Pab*-ASH-Ski2 and *Pab*-Rrp41 from WT strain *P. abyssi* cell extract performed in TK in (A) and TK-EDTA in (B) buffers, as for Figure 7. In here, the endogenous *Pab*-aCPSF1  $\beta$ -CASP endo-RNase is also monitored as control.



**FIGURE S5-** Sedimentation profiles of *Tba*-aRNase J, *Tba*-ASH-Ski2 and *Tba*-Rrp41 from the *Tba*-WT in (A), *Tba*- $\Delta$ ASH-Ski2 in (B), and *Tba*- $\Delta$ aRNase J in (C), strain cell extracts prepared in TK-EDTA buffer. Legend as for Figure 7.

GENES	PROTEINS	N-His pET15b	C-His pET21b	pET11b
<i>pab1751</i>	aRNase J	√	√	√
<i>pab2313</i>	ASH-Ski2	√		√
	ΔC-ASH-Ski2	√		
	ΔN-ASH-Ski2	√		
	ASH-Ski2- C124A	√		
	DomN	√		
	DomN C124A	√		
<i>pab0592</i>	Hel308	√		
<i>pab2314</i>	Csl4	√		√
	ΔC-Csl4	√		
	ΔN-Csl4	√		
	Csl4-F121A	√		
	Csl4-F129A	√		
<i>pab0316</i>	DnaG		√	
<i>pab0419</i>	Rrp4	√		
<i>pab0420</i>	Rrp41	√		

**TABLE S1-** pET vector constructions used to express recombinant proteins in *E.coli*.

	<b>OLIGONUCLEOTIDES</b>	<b>SEQUENCES (5'-3')</b>
<b>Reverse PCR</b>	For pET15b	CATATGGCTGCCGCGCGGC
	Rev pET15b / pET11b	GGATCCGGCTGCTAACAAAGCC
	For pET11b	CATATGTATATCTCCTTCTTAAAGTT
<b>Insert PCR</b>	For aRNaseJ pET11b	GGAGATACATATGTGGGAGGAGATAAACATGATCA
	Rev aRNaseJ pET11b	TTAGCAGCCGGATCCTCATCCCTCCAATGAGCCAG
	For ASH-Ski2 pET15b	CGCGGCAGCCATATGCTATTTCGTTATTCGCCAGGGAGG
	Rev ASH-Ski2 pET15b	AGCCGGATCCTCGAGTTATGGTTTTCTTCTCTCCTCACCTT
	For ASH-Ski2 pET11b	GGAGATACATATGATGCTATTTCGTTATTCGCCAG
	Rev ASH-Ski2 pET11b	TTAGCAGCCGGATCCTTATGGTTTTCTTCTCTCCTCA
	For ASH-Ski2ΔN-ter (194- 855) pET15b	CCGCGCGGCAGCCATATGACGATTGACGAGCTGGAT
	Rev ASH-Ski2ΔC-ter (1-573) pET15b	AGCCGGATCCTCGAGTTAAGAGGTTAGCAATTTCAAGGCCA C
	Rev DomN (1-193) ASH-Ski2 pET15b	TTAGCAGCCGGATCCTTA TCTCTCAACTTTAATACGCTTC
	For Hel308 pET15b	CGCGGCAGCCATATGATGAAAGTTGGAGAGCTAAACG
	RevHel308 pET15b	TTAGCAGCCGGATCCTTATGGGTTTCAGGAAATAGTCC
	For Csl4 pET15b / pET11b	CGCGGCAGCCATATGTTGGAGGAAGGTGAGGAGAG
	Rev Csl4 pET15b	TTAGCAGCCGGATCC TCATAGCTTACCTTCTGTGA
	For Csl4 pET11b	GGAGATACATATGTTGGAGGAAGGTGAGGAGAG
	For Csl4ΔN-ter pET15b	AGGATCCTTTGTCGTCAACTGAATC
Rev Csl4ΔC-ter pET15b	CCTCCAATACCTAAGAAAGGGGA	
For DnaG pET21b	GGAGATACATATGATGAAAAGAAAGAGGGCGATAAT	
Rev DnaG pET21b	GTGGTGCTCGAGTGCCTCGGCGAAGGTTATTACCTT	
For Rrp4 pET15b	CGCGGCAGCCATATG ATGAAGAGGATTTTTGTTCAAAT	
Rev Rrp4 pET15b	TTAGCAGCCGGATCC TTAAGCCCCTCTGGCTTCTC	
For Rrp41 pET15b	CGCGGCAGCCATATGATGATGGAGAAGCCAGAGGG	
Rev Rrp41 pET15b	TTAGCAGCCGGATCCTCACTCACTTCCCTCAACCT	
<b>Directed mutagenesis</b>	For aRNase J Δ1 pET15b	GATTAGAACCCTGTCGAGCCTTATTC
	Rev aRNase J Δ1 pET15b	GACCTTAGGAAGCTCGGTGCAATAC
	For ASH-Ski2 C124A	GAGAGTACATAGCCGAGAGATGTG
	Rev ASH-Ski2 C124A	CACATCTCTCGGCTATGTACTCTC
<b>Slot Blot</b>	For Csl4 F121A	GTAAAGGACGGCGCCGTTGAGGATTTAAGGAA
	Rev Csl4 F121A	TAAATCCTCAACGGCGCCGTCCTTAACTTGAG
<b>Slot Blot</b>	<i>Tba</i> 16S rRNA	CTCCACCCCTTGTAGTGCTC
	<i>Tba</i> 23S rRNA	CCACGGCTGACGAACATTGCC

**TABLE S2-** List of oligonucleotides used in this study

ID	ORF	Name	Description	Ref. Spectra	Spec. Index	Nuclease Effect (%)
Q9V119	PAB0420	Rrp41	RNA Exosome core subunit Rrp41	7.64	0	76.78
G8ZHS0	PAB2163	RPA41	RPA41 subunit	6.69	0	8.94
Q9V2F9	PAB0071	AubA	RNA-binding protein AU-1	6.03	0	100.00
Q9V118	PAB0421	Rrp42	RNA Exosome core subunit Rrp42	5.37	0	59.81
Q9UYB6	PAB1306	eIF-2B	Translation initiation factor eIF-2B subunit 2-like	4.67	0	10.53
Q9V120	PAB0419	Rrp4	RNA Exosome cap subunit Rrp4	4.65	0	55.20
Q9V2M1	PAB2313	ASH-Ski2	ATP-dependant RNA Helicase	4.61	0	57.80
Q9V113	PAB0425	RpoA2	DNA-directed RNA polymerase subunit A''	4.28	0	85.48
Q9V114	PAB0424	RpoA1	DNA-directed RNA polymerase A'	3.97	0	58.80
Q9UY85	PAB1284	RecJ-like	RecJ-like phosphoesterase	3.68	0	0.00
Q9UY07	PAB1115	Grad-2	Glucose-1-phosphate thymidyltransferase	3.67	0	0.00
Q9UYD9	PAB1036	<i>Unch,</i>	Uncharacterized protein	3.33	0	30.22
Q9UYJ6	PAB1366	<i>Unch,</i>	DHH family phosphohydrolase	2.34	0	43.59
Q9V1Z1	PAB2165	RPA32	RPA32 subunit	2.34	0	0.00
Q9V1F2	PAB0316	DnaG	RNA exosome cap subunit DnaG	30.88	0.02456	8.14
Q9V115	PAB0423	RpoB	DNA-directed RNA polymerase subunit beta	10.32	0.03366	46.07
Q9V1T4	PAB2119	<i>Unch,</i>	Uncharacterized RNA methyl transferase-like	6.65	0.05610	1.43
Q9V1U8	PAB2397	Rps4e	30S ribosomal protein S4e	2.51	0.06917	14.00
P61881	PAB3089	HarA	Archaeal histone A	4.31	0.07685	0.00
Q9V1F5	PAB3117	HarB	Archaeal histone B	4.25	0.09445	0.00
Q9UWR8	PAB1166	Rpl1	50S ribosomal protein L1	3.36	0.09646	91.01
G8ZJV0	PAB1300	ApeH	ApeH acylamino-acid-releasing enzyme	1.83	0.12459	0.00
Q9UZX0	PAB1686	<i>Unch,</i>	Uncharacterized Cas7/DevR-like protein	2.29	0.15261	57.73
Q9V2L5	PAB2306	FlpA	Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	13.21	0.16112	47.16
Q9UZ78	PAB2428	TmcA	tRNA(Met) cytidine acetyltransferase	18.53	0.24122	0.00
P61992	PAB0361	Rps4	30S ribosomal protein S4	1.69	0.38540	78.78
Q9V1T5	PAB2120	Rpl3	50S ribosomal protein L3	2.57	0.40940	0.00
Q9V174	PAB1994	<i>Unch,</i>	Uncharacterized DUF536 containing protein	1.58	0.48199	100.00
Q9V2L4	PAB2305	Nop5p	Nop5 C/D snoRNP	4.21	0.48716	0.00
Q9UZN6	PAB1633	PINA	PIN domain ATPase	2.89	0.49010	0.00
Q9V0G8	PAB1813	Rps19e	30S ribosomal protein S19e	1.98	0.52322	100.00
Q9UXS5	PAB1167	Rpl10	50S ribosomal protein L10	2.31	0.57916	26.95
Q9V018	PAB1721	<i>Unch,</i>	Uncharacterized hydrolase, metallo-beta-lactamase family protein	1.65	0.61968	24.91
Q9V1N3	PAB7094	AlbA	DNA/RNA-binding protein	3.98	0.76378	100.00
Q9UYU9	PAB1444	Rpl10e	50S ribosomal protein L10e	1.50	0.77055	0.00
Q9V135	PAB0407	Top6B	Type 2 DNA topoisomerase 6 subunit B	4.23	0.87969	0.00
P62008	PAB0460	Rpl7Ae	50S ribosomal protein L7Ae	1.41	0.89414	24.30

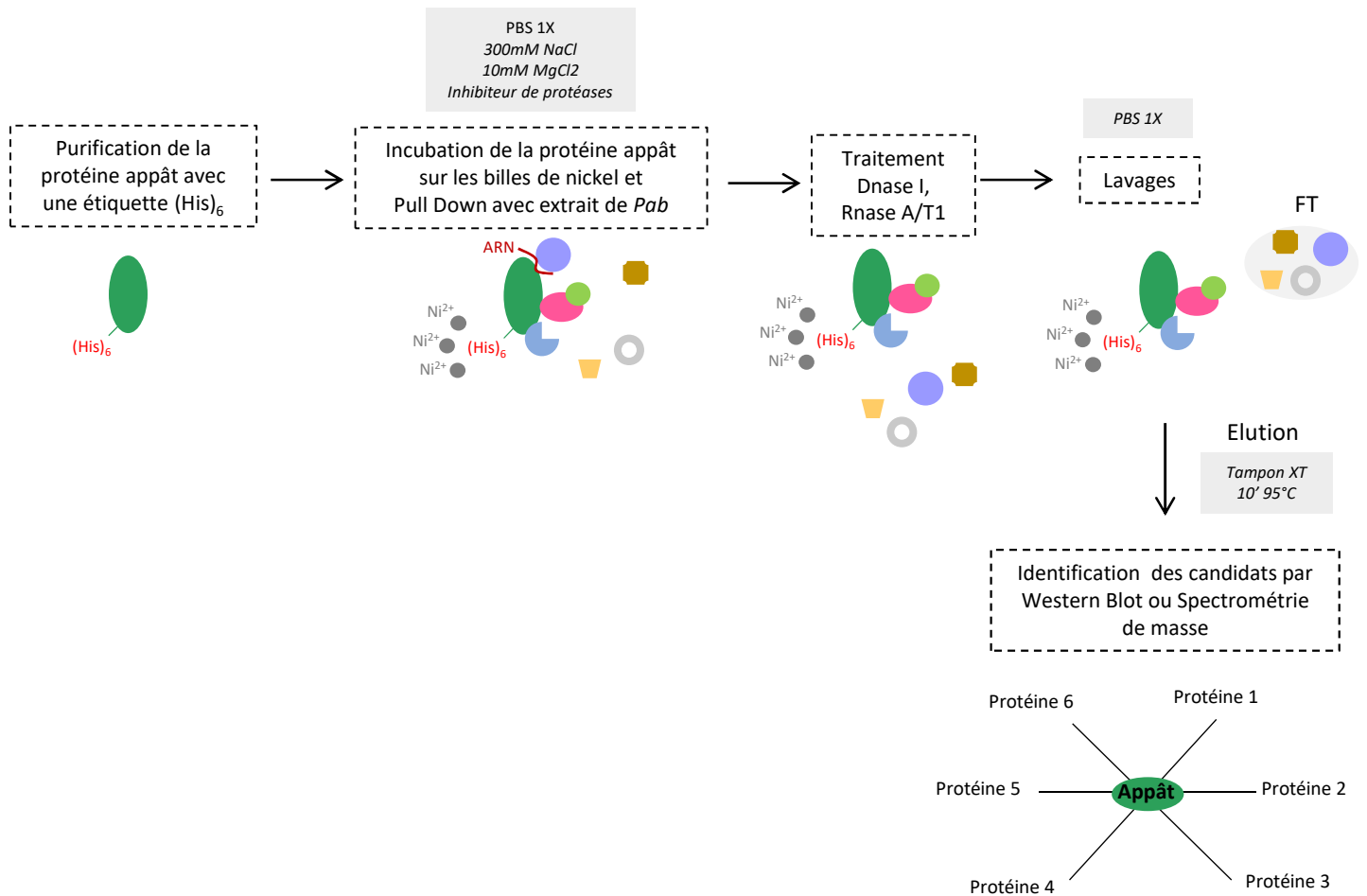
**TABLE S3.** List of potential proteins partners identified by pulldown-MS/MS with *Pab*-aRNase J-(His)<sub>6</sub> as bait protein. Controls were run 6 times whereas “no nuclease” and “nuclease” assays were run in triplicate.

ID	ORF	Name	Description	Ref, Spectra	Spec, Index	Nucl, Effect (%)
G8ZI82	PAB1751	aRNaseJ	5'-3' Exoribonuclease	10.73	0	24.14
Q9V120	PAB0419	Rrp4	RNA Exosome cap subunit Rrp4	9.95	0	100
Q9V133	PAB2412	Unch,	Uncharacterized DUF530 containing protein	8.87	0	0
Q9V2F9	PAB0071	AubA	RNA-binding protein AU-1	8.08	0	97.28
Q9UY85	PAB1284	RecJ-like	RecJ-like phosphoesterase	8.00	0	36.25
Q9V114	PAB0424	RpoA1	DNA-directed RNA polymerase A'	7.30	0	40.17
Q9V118	PAB0421	Rrp42	RNA Exosome core subunit Rrp42	6.40	0	100
Q9V119	PAB0420	Rrp41	RNA Exosome core subunit Rrp41	5.28	0	100
Q9UY91	PAB1064	Unch,	Uncharacterized Cas6-like protein	4.26	0	100
Q9V113	PAB0425	RpoA2	DNA-directed RNA polymerase subunit A''	4.00	0	39.13
Q9UWR8	PAB1166	Rpl1	50S ribosomal protein L1	3.49	0	100
Q9V236	PAB0161	Unch,	Uncharacterized translation initiation factor IF2/IF5-like	3.49	0	54.41
Q9UXX7	PAB1136	Rnp3	Ribonuclease P protein component 3	3.47	0	21.67
Q9UZK6	PAB1613	Unch,	CRISPR-associated endoribonuclease	2.91	0	89.50
P61992	PAB0361	Rps4	30S ribosomal protein S4	2.75	0	0
G8ZHS0	PAB2163	RPA41	RPA41 subunit	14.21	0.00957	5.48
Q9V1V6	PAB2137	Rpl30	50S ribosomal protein L30	4.93	0.01919	10.52
Q9UZF3	PAB1584	Unch,	Uncharacterized DUF2184 containing protein	4.34	0.02184	100
Q9V0G8	PAB1813	Rps19e	30S ribosomal protein S19e	2.86	0.04021	20.37
Q9UXS5	PAB1167	RplP0	Acidic ribosomal protein P0 homolog	2.39	0.05071	4.54
Q9V089	PAB2390	Sun protein	Sun protein (Fmu protein)	20.47	0.05175	22.35
P62008	PAB0460	Rpl7Ae	50S ribosomal protein L7Ae	8.33	0.10606	70.18
Q9V196	PAB0365	Rpl13	50S ribosomal protein L13	3.07	0.13018	1.21
Q9V1V5	PAB2136	Rps5	30S ribosomal protein S5	6.13	0.14079	10.91
Q9V115	PAB0423	RpoB	DNA-directed RNA polymerase subunit beta	20.87	0.14375	39.93
Q9UZN6	PAB1633	PINA	PIN domain ATPase	2.74	0.18422	0
Q9V1U6	PAB2436	Rpl14	50S ribosomal protein L14	2.72	0.18868	42.52
Q9V2L4	PAB2305	Nop5p	Nop5 C/D snoRNP	11.04	0.27236	2.64
Q9V1F2	PAB0316	DnaG	RNA exosome cap subunit DnaG	11.42	0.30179	21.99
Q9V1N3	PAB7094	AlbA	DNA/RNA-binding protein	19.59	0.40021	95.47
Q9V1V4	PAB2135	Rpl18	50S ribosomal protein L18	2.16	0.47113	89.28
Q9UZ78	PAB2428	TmcA	tRNA(Met) cytidine acetyltransferase	24.85	0.51726	0
Q9V2L5	PAB2306	FlpA	Fibrillar-like rRNA/tRNA 2'-O-methyltransferase	8.84	0.52605	22.83
Q9V1T5	PAB2120	Rpl3	50S ribosomal protein L3P	3.63	0.57003	29.48

**TABLE S4.** List of potential proteins partners identified by pulldown-MS/MS with *Pab*-(His)<sub>6</sub>-ASH-Ski2 and *Pab*-ASH-Ski2-(His)<sub>6</sub> as bait proteins, respectively. Controls were run 8 times whereas “no nuclease” and “nuclease” assays were run in quadruplicate.







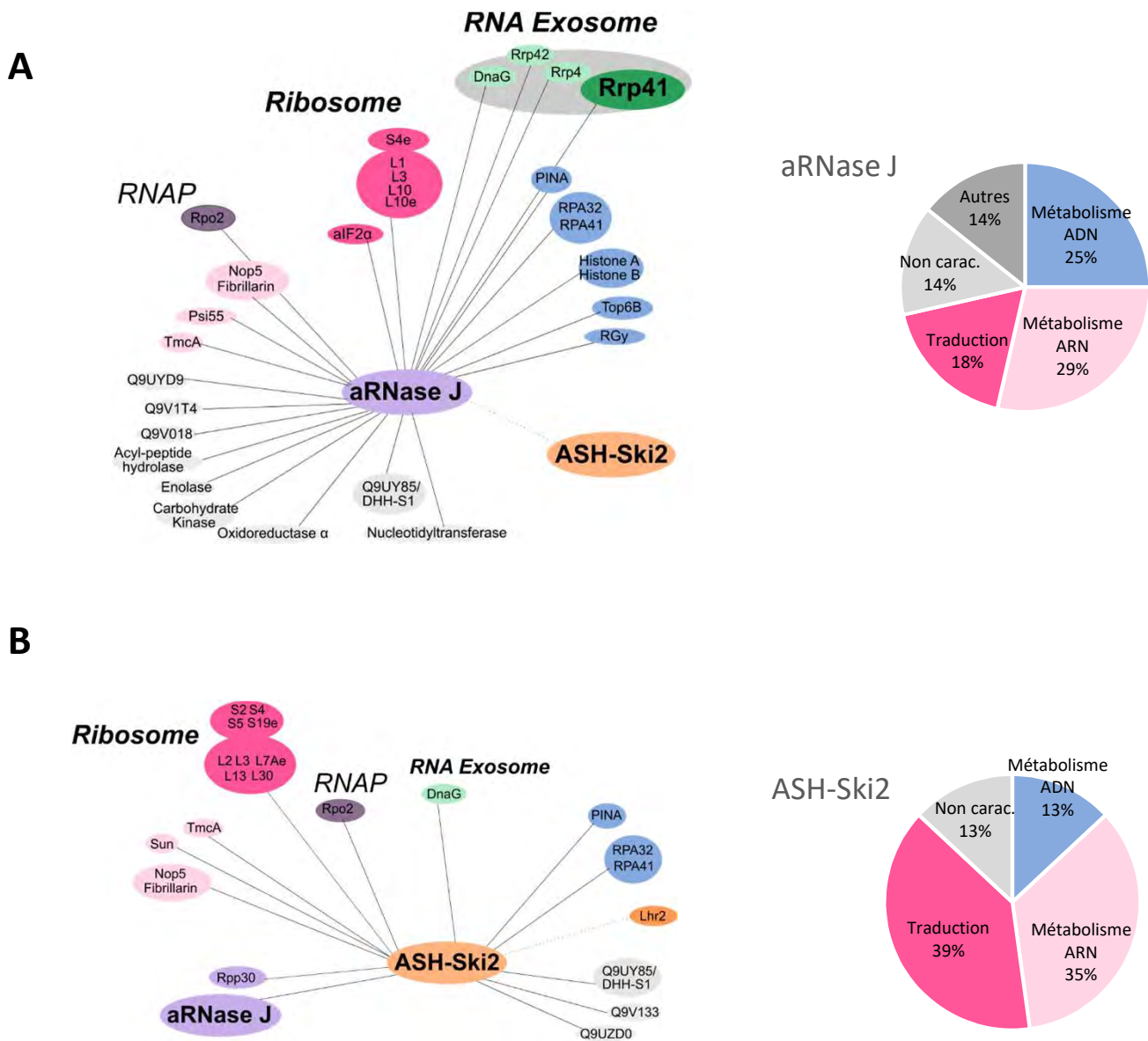
**Figure 14.** Schéma expérimental de la technique dite de « Pull Down » à partir d'une protéine appât (en vert) et d'un extrait cellulaire (protéines colorées). Les différents tampons sont indiqués et décrits dans la partie Matériel et Méthodes. Cette méthode permet d'obtenir une liste de candidats.

## II. Etudes des réseaux d'interactions d'aRNase J, ASH-Ski2 et des sous unités Rrp41 et Csl4 de l'Exosome à ARN

Les réseaux d'interactions (ou interactomes) de la ribonucléase 5'-3' aRNase J et de l'hélicase ASH-Ski2 ont été publiés dans l'article Phung et al 2020 ([Publication 2](#)). Dans cette partie du manuscrit, nous avons repris les données de ces interactomes après avoir établi les réseaux d'interactions des sous unités Rrp41 et Csl4 avec traitement aux nucléases (non publiés) de façon à pouvoir comparer et analyser l'ensemble des interactomes.

### 1. Technique dite de « pull down »

Les réseaux d'interaction ont été établis pour l'exoribonucléase aRNase J, l'hélicase ASH-Ski2 et les sous unités de l'Exosome Rrp41 et Csl4 de *P. abyssi* (*Pab*-aRNaseJ-(His)<sub>6</sub>, *Pab*-ASH-Ski2-(His)<sub>6</sub>, (His)<sub>6</sub>-*Pab*-ASH-Ski2, (His)<sub>6</sub>-*Pab*-Rrp41 et (His)<sub>6</sub>-*Pab*-Csl4) par la technique dite de « pull down » mise au point par l'équipe de D. Flament (Ifremer, Brest). Cette technique a permis d'établir les réseaux d'interaction de protéines impliquées dans la maintenance de l'ADN (Pluchon et al. 2013). La méthode consiste à fixer sur des billes de nickel des protéines recombinantes portant une étiquette Histidine en N- ou C-terminal préalablement purifiée à homogénéité et de les incuber avec un extrait cellulaire de culture de *P. abyssi* en fin de phase exponentielle. Les espèces protéiques de l'extrait interagissant avec la protéine appât d'intérêt sont éluées et identifiées par spectrométrie de masse ([Figure 14](#)). La protéine recombinante *Pab*-aRNase J-(His)<sub>6</sub> a été exprimée avec une étiquette Histidine en C-terminal à partir d'un vecteur d'expression pET21b chez *E. coli*. Les protéines recombinantes (His)<sub>6</sub>-*Pab*-Rrp41 et (His)<sub>6</sub>-*Pab*-Csl4 ont été exprimées avec une étiquette Histidine en N-terminal à partir d'un vecteur d'expression pET15b chez *E. coli*. Les protéines recombinantes (His)<sub>6</sub>-*Pab*-ASH-Ski2 et *Pab*-ASH-Ski2-(His)<sub>6</sub> ont été exprimées avec une étiquette Histidine en N-terminal et en C-terminal respectivement avec les des vecteurs d'expression pETb21b et pET15b. Après purification à homogénéité sur colonne HisTrap et Heparin, 20µg de protéines *Pab*-aRNase J-(His)<sub>6</sub>, (His)<sub>6</sub>-*Pab*-ASH-Ski2, *Pab*-ASH-Ski2-(His)<sub>6</sub>, (His)<sub>6</sub>-*Pab*-Rrp41 ou (His)<sub>6</sub>-*Pab*-Csl4 sont fixées sur des billes de nickel pour être utilisées comme appât. L'extrait cellulaire de *P. abyssi* est ensuite ajouté sur les billes et incubé pendant une nuit à température ambiante ou à 4°C suivi d'un traitement à température ambiante avec les enzymes DNase I et RNase A. Ce traitement aux nucléases permet d'éliminer un certain nombre d'espèces protéiques



**Figure 15.** Réseaux d'interactions obtenus avec les protéines appâts aRNase J-(His<sub>6</sub>) (**A**) et ASH-Ski2-(His<sub>6</sub>) / (His<sub>6</sub>)-ASH-Ski2 (**B**). Les schémas des réseaux d'interactions (panels de gauche) indiquent le lien avec chaque partenaire avec des couleurs selon la fonction. Dans le réseau de ASH-Ski2, le candidat Lhr2 est indiquée en pointillée car détecté uniquement dans le réseau de (His<sub>6</sub>)-ASH-Ski2. La répartition (panels de droite) selon la fonction des candidats d'après l'annotation actuelle du génome de *P. abyssi* : métabolisme de l'ARN en rose clair, traduction en rose foncé (incluant les protéines ribosomiques), métabolisme de l'ADN en violet clair et les protéines de fonction inconnue en gris.

interagissant avec la protéine appât par l'intermédiaire d'acides nucléiques. Après lavage dans du tampon PBS 1X, l'éluât est déposé sur un gel SDS-PAGE (migration courte) puis découpé et envoyé à la plateforme d'Analyses Protéomiques de Paris Sud Ouest (PAPPSO) pour identification des espèces protéiques associées par spectrométrie de masse. Afin d'éliminer les interactions non spécifiques, des expériences contrôles sont réalisées en parallèle par incubation de l'extrait cellulaire avec les billes de nickel en absence de protéine appât. L'ensemble de ces expériences ont été réalisées en triplicata. Les données de spectrométrie MS/MS ont été analysées selon un algorithme (établi par S. Laurent, Ifremer) utilisé pour déterminer les partenaires interagissant spécifiquement avec la protéine appât. La valeur « Spec, Index » prend en compte la spécificité (normalisation par rapport aux échantillons témoins), plus cette valeur est proche de zéro meilleure est la spécificité. La valeur « Ref, Spectra » reflète la quantité de protéines dans chaque échantillon c'est-à-dire l'abondance des spectres d'un peptide donné, plus le chiffre est élevé plus la quantité de peptide est importante. Les deux valeurs « Spec, Index » et « Ref, Spectra » permettent d'établir un classement des interactions les plus spécifiques vers les moins spécifiques (Voir la méthode dans la [Publication 2](#) (Phung et al. 2020)) .

## 2. Réseaux d'interactions d'aRNase J et ASH-Ski2

Les données de la [Publication 2](#) (Phung et al. 2020) obtenues pour les appâts (His)<sub>6</sub>-*Pab*-ASH-Ski2, *Pab*-ASH-Ski2-(His)<sub>6</sub> et *Pab*-aRNase J-(His)<sub>6</sub> dans la condition avec un traitement aux nucléases ont été sélectionnées afin de pouvoir les comparer avec les données récentes non publiées obtenues avec les appâts (His)<sub>6</sub>-*Pab*-Rrp41 ou (His)<sub>6</sub>-*Pab*-Csl4. Les partenaires protéiques de *Pab*-aRNase J-(His)<sub>6</sub> sélectionnés sont classés dans le [tableau 9](#) et ceux de *Pab*-ASH-Ski2-(His)<sub>6</sub> et (His)<sub>6</sub>-*Pab*-ASH-Ski2 sont classés dans le [tableau 10](#).

- **Réseau d'interactions de *Pab*-aRNase J**

La majorité des espèces protéiques du réseau de *Pab*-aRNase J sont annotées comme impliquées dans le métabolisme de l'ARN (33%) ([Figure 15A](#), [Tableau 9](#)). On retrouve la sous unité Rpo2 de l'ARNP, les enzymes de modifications des ARN telles que, la Fibrillarin et Nop5 qui font partie du complexe RNP à boîte C/D responsable de la méthylation du 2'-O-ribose des ARNr et ARNt et les sous unités Rrp41/Rrp42/Rrp4/DNaG de l'Exosome à ARN qui possède une activité 3'-5' exoribonucléolytique. On retrouve également des protéines de la petite et

grande sous-unité du ribosome (S4e, L3, L10 et L10e) ainsi que le facteur d'initiation de la traduction aIF2. Cependant, 29% des protéines partenaires sont annotées comme jouant un rôle dans la maintenance de l'ADN. En effet les facteurs de réplication, RPA41 et RPA32 qui sont des protéines capables de lier de l'ADN simple brin dans le processus de réplication pour prévenir la formation du duplexe et protéger l'ADN simple brin contre les attaques des nucléases sont détectés. De plus les histones A et B qui permettent la compaction de l'ADN ou encore l'ATPase PINA qui semble impliquée dans la réparation de l'ADN au niveau des jonctions de Holliday sont présentes de manière significative. De manière remarquable comme montré dans la Publication 2, le réseau d'aRNase J comporte une protéine annotée comme étant une hélicase de la famille Ski2 (PAB2313), renommée ASH-Ski2 (Archaeal Specific Helicase) dont l'activité enzymatique fait l'objet de cette thèse dans la partie IV. Chez les eucaryotes, les hélicases de type Ski2 ont un rôle crucial dans la dégradation et le contrôle qualité des ARN et font partie de complexe multiprotéique (SKI ou TRAMP complexe) qui s'associe à l'Exosome à ARN cytoplasmique ou nucléaire (Halbach et al. 2012; Wasmuth et al. 2017; Meola et al. 2016; Schmidt et al. 2016).



ID	ORF	NAME	Description	Ref, Spectra	Spec, Index
G8ZHS0	PAB2163	RPA41	Replication factor A (RPA41)	6,09	0
Q9UY07	PAB1115	Nucleotidyltransferase	Glucose-1-phosphate thymidyltransferase	4,59	0
Q9UYB6	PAB1306	aIF2	Translation initiation factor (ARNt binding)	4,18	0
Q9UY85	PAB1284	DHH-S1	DHH family	3,73	0
Q9V1Z1	PAB2165	RPA32	Replication factor A (RPA32)	3,10	0
Q9UYD9	PAB1036		Uncharacterized protein	2,32	0
Q9V118	PAB0421	Rrp42	Exosome complex component Rrp42	2,16	0
Q9V120	PAB0419	Rrp4	Exosome complex component Rrp4	2,08	0
Q9V1F2	PAB0316	DNaG	Exosome component	28,36	0,02674
Q9V1T4	PAB2119		Uncharacterized protein	6,55	0,05692
Q9V115	PAB0423	Rpo2	DNA-directed RNA polymerase subunit 2/ catalytic subunit	5,57	0,06242
Q9V1F5	PAB3117	Histone B	Archaeal histone B	6,33	0,06338
P61881	PAB3089	Histone A	Archaeal histone A	4,46	0,07430
Q9V1U8	PAB2397	S4e	30S ribosomal protein S4e	2,16	0,08043
Q9UZ86	PAB2423	Rgy	Reverse gyrase	2,08	0,08342
G8ZJV0	PAB1300	Acyl-peptide hydrolase	Acyl-peptide hydrolase /peptide modification	2,66	0,08556
Q9V1T5	PAB2120	L3	50S ribosomal protein L3	6,87	0,15324
Q9UZ78	PAB2428	TmcA	tRNA(Met) cytidine acetyltransferase	21,79	0,20513
Q9UZN6	PAB1633	PINA	ATPase / Holliday junction	5,14	0,27581
Q9V2L5	PAB2306	Fibrillarin	Fibrillarin-like rRNA/tRNA 2'-O- methyltransferase	6,98	0,30492
Q9V2L4	PAB2305	Nop5	Component of RNA methylating RNP complex (C/D box)	4,94	0,41602
Q9UYU9	PAB1444	L10e	50S ribosomal protein L10e	1,73	0,66868
Q9V135	PAB0407	Topo-6B	Type 2 DNA topoisomerase 6 subunit B	4,93	0,75582
Q9UXS5	PAB1167	L10	50S ribosomal protein L10	1,69	0,79280
Q9V018	PAB1721		UPF0282 protein PYRAB09800	1,24	0,82529
G8ZGL4	PAB2359	oxidoreductase $\alpha$	2-oxoglutarate ferredoxin oxidoreductase subunit alpha	1,06	0,90562
Q9V1A5	PAB0356	Psi55 synthase	Probable tRNA pseudouridine synthase B	2,61	0,90804
Q9V192	PAB0367	Enolase	Eno-like enolase related/glycolyse pathway	2,30	0,99450

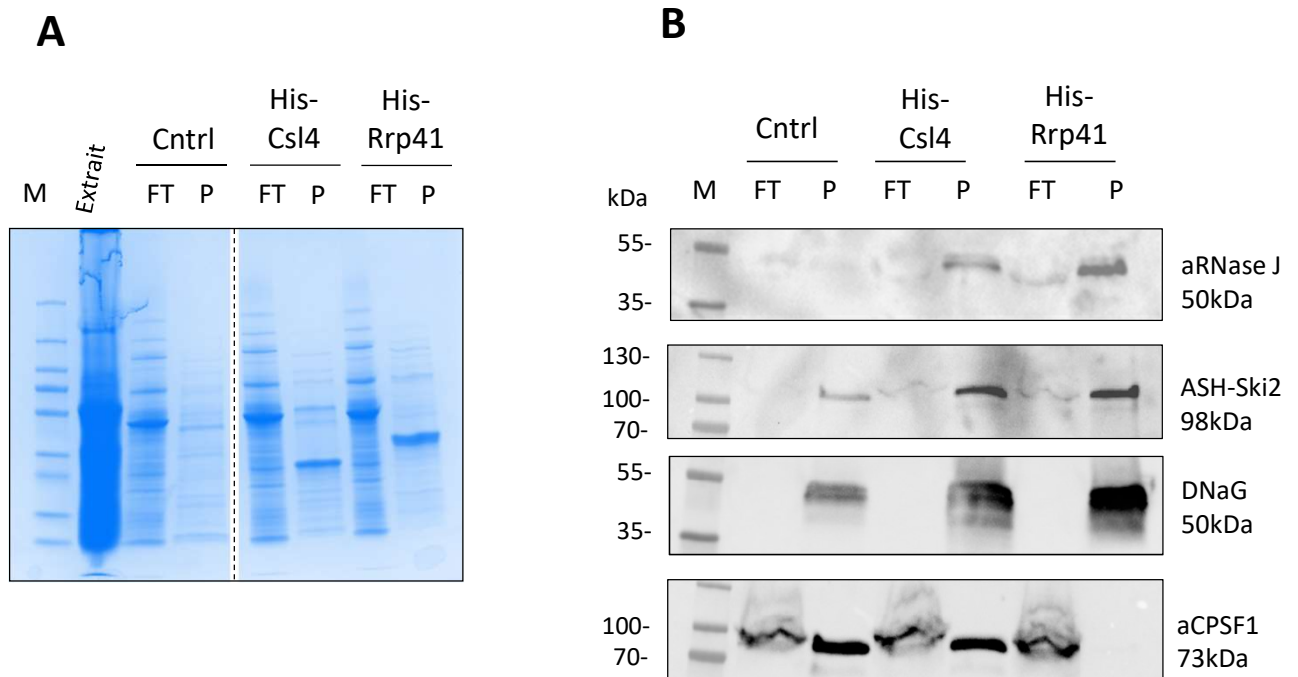
**Tableau 9 : Liste des partenaires protéiques identifiés par pull down-MS/MS de *Pab*-aRNase J-(His)<sub>6</sub> utilisée comme appât.** L'identifiant et la description des protéines sont issus de la banque de données uniprot (UP000009139). Le nombre de spectres référencés « Ref, Spectra » reflète la quantité de protéine et l'indice « Spec, Index » indique la spécificité de l'interaction. Les protéines de l'exosome (vert) et de la traduction (rose) sont colorés.

- **Réseau d'interactions de *Pab*-ASH-ski2**

Seuls les partenaires retrouvés à la fois avec la protéine ASH-Ski2-(His)<sub>6</sub> et (His)<sub>6</sub>-*Pab*-ASH-Ski2 sont sélectionnés et classés dans le **Tableau 10**. De manière remarquable on retrouve aRNase J dans le haut du classement avec un score reflétant une spécificité d'interaction d'après nos critères. De plus, une grande majorité d'espèces identifiées sont annotées comme jouant un rôle dans le métabolisme de l'ARN (35%) et dans la traduction (39%) (**Figure 15B, Tableau 10**). Les partenaires interagissant de manière spécifique d'après nos analyses incluent des enzymes de modification de l'ARN telle que TmcA et les protéines formant le complexe RNP guide à boîtes C/D composées de la Fibrillarin (2'-O-méthyltransférases) et de la protéine Nop5. De plus on observe la sous unité Rpp30 de l'endonucléase RNase P responsable de la maturation en 3' des pré-ARNt, la sous unité Rpo2 de l'ARNP, la sous unité de l'exosome DnaG et les protéines ribosomales S2, S4, S19e, S5 de la petite sous-unité, et L2, L30, L13, L7Ae et L3 de la grande sous-unité. Dans ces réseaux d'interaction, de nombreux partenaires restent non caractérisés. Bien que l'annotation du génome de *P. abyssi* (UP000009139) qui sert de base à notre analyse ait été révisé (Gao and Wang 2012), il reste dans ce génome un grand nombre de phases ouvertes de lecture codant pour des facteurs inconnus à ce jour. De manière intéressante, comme souligné dans la **Publication 2**, les réseaux d'interactions d'aRNase J et d'ASH-Ski2 partagent un certain nombre de partenaires communs qui incluent entre autre le complexe RNP à boîtes C/D (Fibrillarine, Nop5), l'ATPase PINA et le facteur DnaG décrit comme associé à Csl4, sous-unité de la coiffe de l'exosome à ARN (Hou et al. 2013; Clouet-d'Orval et al. 2018).

ID	ORF	NAME	Description	Ref, Spectra	Spec, Index
Q9V133	PAB2412		Putative uncharacterized protein	9,82	0
Q9V076	PAB1751	aRNase J	Ribonuclease J	8,14	0
Q9UY85	PAB1284	DHH-S1	DHH family	5,56	0
Q9V191	PAB0368	S2	30S ribosomal protein S2	3,70	0
P61992	PAB0361	S4	30S ribosomal protein S4	3,26	0
Q9V1T8	PAB2122	L2	50S ribosomal protein L2	2,99	0
Q9UXX7	PAB1136	Rpp30	Ribonuclease P protein component 3	2,50	0
G8ZHS0	PAB2163	RPA41	Replication factor A (RPA41)	13,45	0,01661
Q9V1V6	PAB2137	L30	50S ribosomal protein L30	4,09	0,02320
Q9V0G8	PAB1813	S19e	30S ribosomal protein S19e	2,17	0,05353
Q9UZN6	PAB1633	PINA	ATPase/ Holliday junction	3,78	0,13434
Q9V196	PAB0365	L13	50S ribosomal protein L13	2,90	0,13820
Q9V089	PAB2390	Sun	Sun protein (Fmu) S-adenosyl methionine/RNA binding domain	15,71	0,15367
Q9V1V5	PAB2136	S5	30S ribosomal protein S5	5,62	0,17536
Q9V1Z1	PAB2165	RPA32	Replication factor A (RPA32)	8,64	0,25612
Q9V2L4	PAB2305	Nop5	Component of RNA methylating RNP complex (C/D box)	10,56	0,28459
P62008	PAB0460	L7Ae	50S ribosomal protein L7Ae/ c/d box RNA	2,32	0,38664
Q9V1F2	PAB0316	DNaG		8,84	0,39313
Q9UZD0	PAB0810		Predicted ATPase	3,36	0,41686
Q9V115	PAB0423	Rpo2	DNA-directed RNA polymerase subunit 2/ catalytic subunit	13,09	0,42611
Q9UZ78	PAB2428	TmcA	tRNA(Met) cytidine acetyltransferase TmcA	25,64	0,51075
Q9V2L5	PAB2306	Fibrillarin	Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	6,54	0,71192
Q9V1T5	PAB2120	L3	50S ribosomal protein L3	2,55	0,81252

**Tableau 10: Liste des partenaires protéiques identifiés par pull down-MS/MS de *Pab-ASH-Ski2-(His)<sub>6</sub>* et de *(His)<sub>6</sub>-Pab-ASH-Ski2*. Voir la légende du tableau 9. Les protéines ribosomales sont en rose et la ribonucléase en violet.**



**Figure 16.** Pull down à partir des protéines appâts  $(\text{His})_6\text{-Pab-Rrp41}$  &  $(\text{His})_6\text{-Pab-Csl4}$  avec un extrait de *Pyrococcus abyssi*. FT correspond au flow through, P aux fractions éluées et M au marqueur protéique. **A.** Gel SDS PAGE coloré au bleu de comassie. **B.** Western Blot. Les protéines aRNase J, ASH-Ski2, DNaG et aCPSF1, sont révélées grâce à un anticorps spécifique. L'expérience témoin a été réalisée en parallèle sans protéine appât.

### 3. Réseaux d'interactions des sous unités de l'exosome Rrp41 et Csl4

Afin de clarifier la présence des sous-unité de l'exosome à ARN dans les réseaux d'aRNase J et ASH-Ski2, nous avons entrepris d'établir les réseaux de la sous unité catalytique Rrp41 du tonneau de l'Exosome à ARN (appât (His)<sub>6</sub>-Pab-Rrp41) et la sous unité Csl4 de la coiffe (appât (His)<sub>6</sub>-Pab-Csl4). La même méthodologie expérimentale et les mêmes critères d'analyse sont utilisés afin d'homogénéiser et comparer l'ensemble des réseaux obtenus par les appâts (His)<sub>6</sub>-Pab-Rrp41, (His)<sub>6</sub>-Pab-Csl4 avec ceux obtenus pour (His)<sub>6</sub>-aRNase J et (His)<sub>6</sub>-ASH-Ski2/ASH-Ski2-(His)<sub>6</sub>.

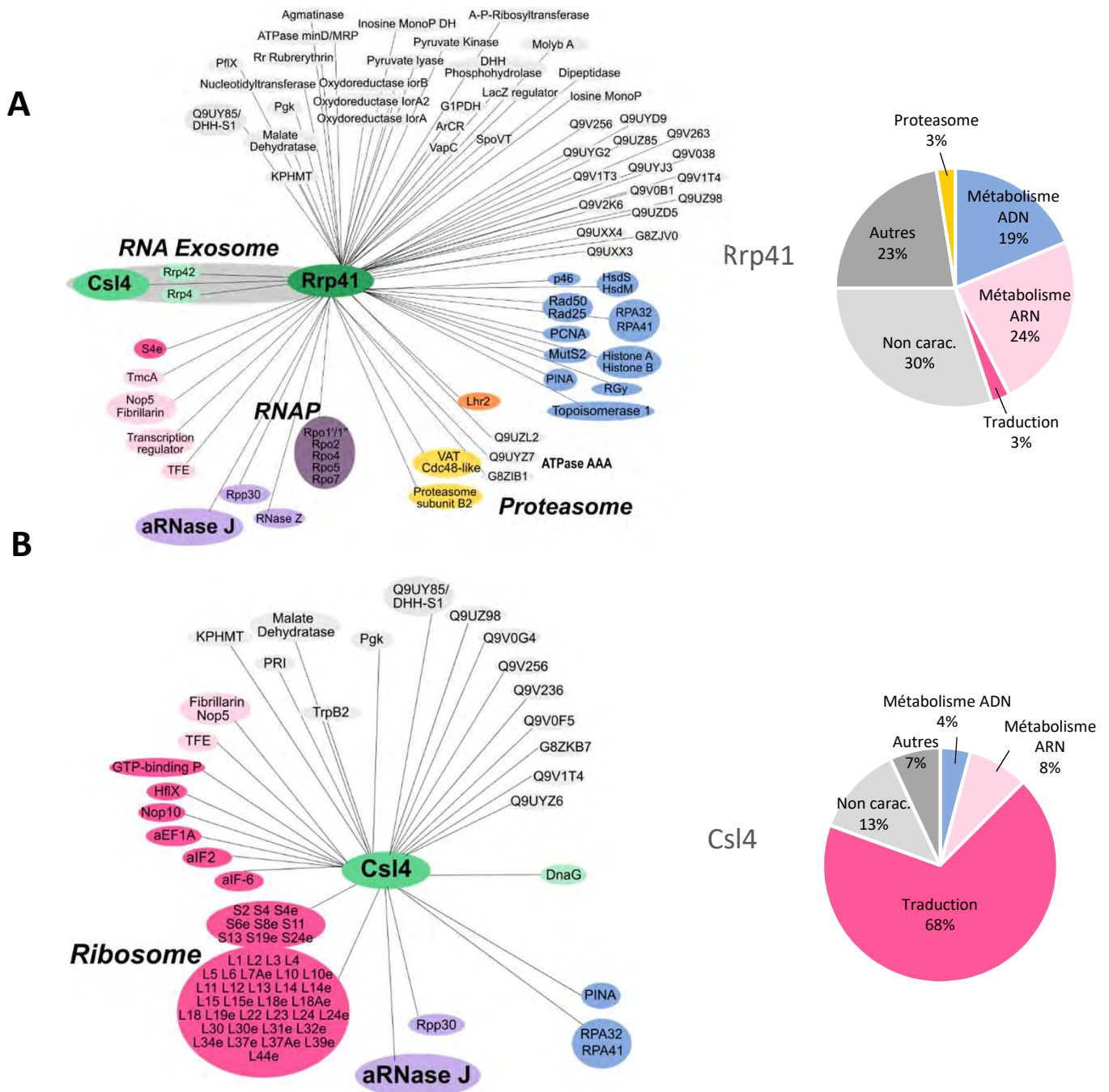
Notre première approche a été d'effectuer une analyse par Western Blot des éluats obtenus avec les appâts (His)<sub>6</sub>-Pab-Rrp41, (His)<sub>6</sub>-Pab-Csl4 afin de détecter les espèces protéiques d'intérêt avec les anticorps spécifiques dirigés contre aRNase J, ASH-Ski2 et DNaG de *P. abyssi* (disponible dans l'équipe). Cette analyse révèle la présence d'aRNase J, ASH-Ski2 et DNaG dans les éluats des Pull down de (His)<sub>6</sub>-Pab-Rrp41 et (His)<sub>6</sub>-Pab-Csl4 (**Figure 16A et 16B**).

Nous avons également utilisé un anticorps dirigé contre la ribonucléase aCPSF1 de *P. abyssi* qui n'est pas retrouvé dans les réseaux d'ASH Ski2 et aRNase J. aCPSF1 est détectée dans l'expérience témoin sans protéine appât ainsi que dans les fractions d'élution (P) de l'expérience avec l'appât (His)<sub>6</sub>-Pab-Csl4. aCPSF1 interagit donc de manière non spécifique sur les billes. L'absence de détection dans les fractions d'élution avec la protéine appât (His)<sub>6</sub>-Pab-Rrp41 reflète peut-être une saturation des billes (**Figure 16B**).

Ces résultats révélés par Western blot nous ont conforté dans notre hypothèse d'un lien spécifique entre l'exosome à ARN, ASH-Ski2 et aRNase J. Nous avons donc entrepris de réitérer les expériences de Pull down avec (His)<sub>6</sub>-Pab-Rrp41 et (His)<sub>6</sub>-Pab-Csl4 afin d'identifier l'ensemble des partenaires par spectrométrie MS/MS via la plateforme PPAPSO comme décrit précédemment.

- **Réseau d'interactions de Pab-Rrp41**

Un nombre de protéines candidates (80) sont identifiées avec un score significatif dans l'expérience réalisée avec l'appât (His)<sub>6</sub>-Pab-Rrp41. Parmi elles, on retrouve 24% de protéines annotées comme appartenant au métabolisme de l'ARN. En particulier, les autres composants de l'exosome à ARN (deux sous-unité de la coiffe, Csl4 et Rrp4 et la sous-unité du tonneau Rrp42) sont détectés de manière significative. De plus des facteurs impliqués dans la



**Figure 17.** Réseaux d'interactions obtenues avec les protéines appâts  $(His)_6$ -*Pab*-Rrp41 (**A**) et  $(His)_6$ -*Pab*-Csl4 (**B**). Les schémas des réseaux d'interactions (panels de gauche) indiquent le lien avec chaque partenaire avec des couleurs selon la fonction. La répartition (panels de droite) selon la fonction des candidats d'après l'annotation actuelle du génome de *P. abyssi* : métabolisme de l'ARN en rose clair, traduction en rose foncé (incluant les protéines ribosomiques), métabolisme de l'ADN en violet clair, le protéasome et les protéines associées en jaune et les protéines de fonction inconnue ou impliquées dans d'autres processus en gris.



transcription (sous unités de l'ARNP ; facteur d'initiation de la transcription, facteur de régulation) ainsi que dans la maturation et modification des ARN (Rpp30, Nop5, Fibrillarin, TmcA) sont présents (**Figure 17A, Tableau 11**). De manière cohérente avec les résultats précédents, on retrouve réciproquement aRNase J avec un score reflétant une interaction spécifique (**Figure 17A, Tableau 11**). Cependant une majorité des candidats partenaires sont soit des espèces non annotées (30%) (non caractérisées) dans le génome de *P. abyssi*, soit des enzymes impliquées dans diverses voies métaboliques (23%) comme par exemple des déshydrogénases (PAB1138, PAB113), une Pyruvate kinase (PAB1441) ou la protéine de la glycolyse P<sub>gk</sub> (PAB1679). Les données disponibles sur ces enzymes du métabolisme chez les archées restent limitées. On trouve également 19% de protéines impliquées dans le métabolisme de l'ADN, notamment dans la réplication (PCNA, la sous unité de la primase p46 et les facteurs de réplication RPA41 et RPA32), dans la réparation (rad25, rad50 et la reverse gyrase), dans les modifications de l'ADN (HsdM et HsdS) et les protéines de liaison à l'ADN (Histone A et Histone B). Deux des partenaires du réseau de Rrp41 sont des facteurs du protéasome (la sous unité  $\beta$ 2 du tonneau) ou associés au protéasome (l'ATPase VAT ou Cdc48-like).

ID	ORF	NAME	Description	Ref, Spectra	Spec, Index
Q9UXX4	PAB1138		NADH-dep. dyhydrogenase related protein	7,67	0
Q9UXX3	PAB1139		NADH-dep. dyhydrogenase related protein	7,48	0
Q9UZP0	PAB0732	Rpo4	DNA-directed RNA polymerase, subunit 4	7,48	0
Q9UZS4	PAB0713	Pyruvate lyase	Pyruvate formate-lyase	5,22	0
G8ZJS6	PAB0744	Lhr2	ATP-dependent helicase	4,21	0
Q9V2F8	PAB2271	VapC	Ribonuclease VapC/putative toxin VapC	4,18	0
Q9UYC3	PAB1046	ArCR	Uncharacterized ArCR	4,16	0
G8ZJV0	PAB1300		Acyl-peptide hydrolase	3,49	0
G8ZFR3	PAB0041	Molyb A	Molybdopterin-guanine dinucleotide	3,42	0
Q9UYD9	PAB1036		Uncharacterized protein	3,28	0
Q9UYB7	PAB1307	MutS2	DNA-binding protein MutS2	3,26	0
Q9UYC4	PAB1312	Rad25	DNA repair protein RAD25	3,26	0
Q9V0B1	PAB0590		Uncharacterized protein	2,63	0
Q9V2K6	PAB2299		HTH-type transcriptional regulator	2,47	0
Q9UYG2	PAB1020		Uncharacterized protein	2,11	0
Q9UYJ6	PAB1366	DHH Phosphohydrolase	DHH family phosphohydrolase	2,02	0
Q9V2F7	PAB3032	SpoVT	Regulator protein, putative	1,96	0
Q9V291	PAB2235	p46	DNA primase large subunit PriL	1,73	0
Q9V038	PAB0644		Uncharacterized protein	1,55	0
Q9UZC8	PAB0812	Rad50	DNA double-strand break repair ATPase	1,40	0
Q9UZX5	PAB1691		Transcription regulator	1,35	0
Q9V263	PAB0144		Phosphate regulatory protein, putative	1,31	0
Q9V2I5	PAB0050	Agmatinase	Agmatinase	1,08	0
Q9V1X9	PAB2150	HsdS	restriction-modification enzyme	12,34	0,02371
Q9UYS8	PAB1430	Topo1	DNA topoisomerase 1	11,25	0,02599
Q9UZN6	PAB1633	PINA	ATPase	18,50	0,03220
Q9UY23	PAB1105	Rpo1	RNA polymerase subunit 1	6,38	0,04754
Q9V278	PAB0128	Rad25	DNA repair protein	6,32	0,04792
G8ZIB1	PAB0633		AAA domain-containing protein	3,21	0,09447
Q9V2G8	PAB2278	LacZ regulator	LacZ expression regulatory protein	3,87	0,11329
G8ZHT1	PAB2149	HsdM	restriction modification enzyme	12,59	0,11696
Q9V113	PAB0425	RpoA''	DNA-directed RNA polymerase subunit A''	19,52	0,12842
Q9V0A0	PAB0598	Rr rubrerythrin	Rr rubrerythrin	3,12	0,19117
Q9V041	PAB2388	Dipeptidase	Dipeptidase	1,44	0,20355
Q9UZ86	PAB2423	Rgy	Reverse gyrase	25,66	0,20922
Q9V2M2	PAB2314	Csl4	Exosome complex component Csl4	17,61	0,23488
Q9V114	PAB0424	Rpo1'	RNA polymerase subunit 1'	48,03	0,23704
Q9V1T4	PAB2119		Uncharacterized protein	1,20	0,24313
Q9UYI6	PAB1005	aIF-6	Translation initiation factor 6	1,23	0,24705
Q9UY85	PAB1284	DHH-S1	RecJ-like phosphoesterase	19,71	0,24860
Q9UXX7	PAB1136	Rpp30	Ribonuclease P protein component 3	2,25	0,26443
Q9UYZ7	PAB1478	ATPase AAA	cell division control protein	74,14	0,27204
Q9V1U8	PAB2397	S4e	30S ribosomal protein S4e	1,80	0,33078
Q9V1W5	PAB0211	Inosine monoP DH	Inosine-5'-monophosphate DH	1,27	0,34466
G8ZHL8	PAB1441	Pyruvate kinase	Pyruvate kinase	54,01	0,35121
Q9UZD5	PAB1578		Probable L-aspartate decarboxylase	3,52	0,37937
Q9V0E2	PAB0569	Malade dehydratase	3-isopropylmalate dehydratase	10,00	0,38336

Q9UYS6	PAB0950	TFE	Transcription factor E	5,36	0,38819
G8ZHS0	PAB2163	RPA41	Replication factor A (RPA41)	35,31	0,39365
Q9UYJ3	PAB1363		Uncharacterized protein	16,17	0,40180
Q9V0E1	PAB0570	KPHMT	cofactor synthesis	5,87	0,40603
Q9UY07	PAB1115	Nucleotidyltransferase	Glucose-1-P thymidyltransferase	28,68	0,41816
Q9UZ98	PAB0830		Proliferating-cell nucleolar antigen P120	1,40	0,41870
Q9V1Z1	PAB2165	RPA32	Replication factor A (RPA32)	10,35	0,44177
Q9V120	PAB0419	Rrp4	Exosome complex component Rrp4	11,29	0,44679
Q9V076	PAB1751	aRNase J	5'-3' exoribonuclease	13,06	0,45277
Q9UZ78	PAB2428	TmcA	tRNA(Met) cytidine acetyltransferase TmcA	84,20	0,45752
Q9UZ60	PAB0852	ATPase minD/MRP	ATPase MRP superfamily	9,33	0,49006
Q9UYX8	PAB1465	PCNA	DNA polymerase sliding clamp	8,39	0,49424
Q9UZP4	PAB0728	RNase Z	Ribonuclease Z	1,13	0,52538
Q9V265	PAB0143	Oxidoreductase IorB	Indolepyruvate oxidoreductase subunit B	5,57	0,53109
Q9V116	PAB7151	Rpo5	RNA polymerase subunit 5/ dsDNA binding	1,85	0,56017
Q9V1N2	PAB2086	VAT	Cdc48-like ATPase AAA family	18,71	0,57478
Q9V1F5	PAB3117	Histone B	Archaeal histone B	12,81	0,57644
Q9UZW0	PAB1679	Pgk	Phosphoglycerate kinase	17,26	0,60693
Q9V1T3	PAB2118		Uncharacterized protein	27,86	0,61114
Q9V1E4	PAB2035	A-P-Ribosyltransferase	Adenine phosphoribosyltransferase	5,24	0,62041
Q9V118	PAB0421	Rrp42	Exosome core complex component	5,27	0,64417
Q9UZL2	PAB1618		ATPase of the AAA superfamily, putative	3,43	0,64869
Q9UZ57	PAB0855	Oxidoreductase IorA	Indolepyruvate oxidoreductase subunit	9,65	0,68716
Q9V2L4	PAB2305	Nop5	RNA methylating RNP complex (C/D box)	10,23	0,70876
Q9V256	PAB0150		Uncharacterized protein	2,67	0,71958
Q9UZR8	PAB0718	Oxidoreductase IorA2	Indolepyruvate oxidoreductase subunit	6,13	0,74764
Q9V0V0	PAB1907	G1PDH	Glycerol-1-phosphate dehydrogenase	19,81	0,76866
Q9UZ85	PAB0838		Uncharacterized protein	8,03	0,77382
Q9UZM6	PAB0742	PflX	pyruvate formate-lyase	1,45	0,81442
P61881	PAB3089	Histone A	Archaeal histone A	7,90	0,84127
Q9V2L5	PAB2306	Fibrillarin	rRNA/tRNA 2'-O-methyltransferase	12,19	0,90062
Q9V0N9	PAB1867	Proteasome subunit $\beta$ 2	Proteasome subunit $\beta$ 2	6,84	0,95122
Q9V115	PAB0423	Rpo2	RNA polymerase subunit 2	61,63	0,95575

**Tableau 11 : Liste des partenaires protéiques identifiés par pull down-MS/MS de *Pab-Rrp41*.**

Voir la légende du tableau 9 et 10.

- **Réseau de *Pab*-Csl4**

Le réseau obtenu à partir de l'appât (His)<sub>6</sub>-*Pab*-Csl4 comprend également un nombre important (74) de candidats. De façon remarquable, parmi eux, on détecte 43 protéines ribosomiques (sur un total de 64 protéines ribosomiques chez *P. abyssi*) dont 33 des 39 protéines de la grande sous-unité 50S. D'autres facteurs de la traduction tels que le facteur d'initiation aIF2, le facteur d'élongation aEF1A, le facteur aIF-6 de fonction inconnue sont présents en plus de la protéine Nop10 importante dans la biogénèse des ribosomes (**Figure 17B, Tableau 12**). On peut donc proposer que le co-facteur Csl4 de la coiffe de l'exosome à ARN de *P. abyssi* est fortement connecté au ribosome dans les conditions de l'expérience. D'autres partenaires (8%) sont annotées comme intervenant dans le métabolisme de l'ARN telles que la transcription (facteur d'élongation TFE) et la modification des ARN (Nop 5, Fibrillarin). Finalement, aRNase J apparait de manière significative dans ce réseau, ce qui est en adéquation avec les résultats précédents. Bien que l'on retrouve le co-facteur DnaG, on note que les protéines du tonneau de l'exosome Rrp41 et Rrp42, et la sous unité de la coiffe Rrp4 n'apparaissent pas dans cette analyse. Il est connu que DnaG interagit avec l'exosome à ARN via la sous-unité de la coiffe Csl4 (Hou et al. 2013; Evguenieva-Hackenberg et al. 2014). L'absence de la globalité des sous-unités de l'exosome dans le réseau de Csl4 peut laisser supposer que le complexe Csl4/DnaG pourrait agir indépendamment de la machinerie de l'exosome à ARN.

ID	ORF	NAME	Description	Ref, Spectra	Spec, Index
Q9V197	PAB0364	L18e	50S ribosomal protein L18e	16,75	0
Q9UYI5	PAB7388	L18Ae	50S ribosomal protein L18Ae	12,41	0
GTP-binding				10,21	0
Q9UYI3	PAB1357	Protein	GTPase/ Translation associated GTPase		
Q9V150	PAB1970	TrpB2	Tryptophan synthase beta chain	7,38	0
Q9UY20	PAB3419	S24e	30S ribosomal protein S24e	6,98	0
Q9V2S3	PAB2353	L11	50S ribosomal protein L11	6,58	0
Q9V112	PAB7152	L30e	50S ribosomal protein L30e	5,76	0
Q9UWR8	PAB1166	L1	50S ribosomal protein L1	5,66	0
Q9V236	PAB0161		Uncharacterized protein	4,30	0
Q9V1T7	PAB7083	L23	50S ribosomal protein L23	3,81	0
Q9UXZ2	PAB1221	L44e	50S ribosomal protein L44e	3,56	0
Q9V1A0	PAB0360	S13	30S ribosomal protein S13	3,41	0
Q9V0F5	PAB1804		Uncharacterized protein	2,60	0
Q9V1V7	PAB2138	L15	50S ribosomal protein L15	2,51	0
Q9V1V2	PAB2133	L32e	50S ribosomal protein L32e	2,35	0
Q9V202	PAB7067	L37Ae	50S ribosomal protein L37Ae	2,35	0
P62004	PAB7160	L37e	50S ribosomal protein L37e	2,11	0
Q9V0W3	PAB8165	L24e	50S ribosomal protein L24e	1,96	0
G8ZKB7	PAB0758		Uncharacterized protein	1,94	0
Q9UYS3	PAB1427	S6e	30S ribosomal protein S6e	1,70	0
Q9UZR3	PAB0722	HflX	GTPase HflX associated with 50S	1,70	0
Q9V1V3	PAB2134	L19e	50S ribosomal protein L19e	1,69	0
Q9UZJ7	PAB7289	L34e	50S ribosomal protein L34e	1,31	0
Q9V0J6	PAB0538	PdxT	Pyridoxal 5'-phosphate synthase subunit PdxT	1,06	0
Q9UYI6	PAB1005	aIF-6	Translation initiation factor 6	13,24	0,02289
Q9V196	PAB0365	L13	50S ribosomal protein L13	20,25	0,03610
Q9UZN6	PAB1633	PINA	ATPase/ Holliday junction	12,24	0,04864
Q9V1V1	PAB2132	L6	50S ribosomal protein L6	17,22	0,05091
Q9V0L6	PAB0522	PRI	Ribose-5-phosphate isomerase A	5,39	0,05623
Q9V0D2	PAB0575	L15e	50S ribosomal protein L15e	7,34	0,05976
Q9V0G4	PAB0558		GTP-binding protein, putative	7,22	0,08245
Q9UYI8	PAB3384	L39e	50S ribosomal protein L39e	3,27	0,08950
Q9UYI7	PAB7387	L31e	50S ribosomal protein L31e	7,18	0,10321
Q9V1V6	PAB2137	L30	50S ribosomal protein L30	16,58	0,12472
Q9V1U7	PAB2128	L24	50S ribosomal protein L24	9,85	0,13567
Q9UZJ5	PAB7290	L14e	50S ribosomal protein L14e	12,59	0,13932
Q9V1U0	PAB2396	L22	50S ribosomal protein L22	14,49	0,14352
Q9V1U6	PAB2436	L14	50S ribosomal protein L14	13,83	0,17139
Q9V1U8	PAB2397	S4e	30S ribosomal protein S4e	3,36	0,17707
Q9UZL4	PAB0749	S8e	30S ribosomal protein S8e	1,65	0,18335
Q9UYU9	PAB1444	L10e	50S ribosomal protein L10e	7,27	0,20385
Q9UZ98	PAB0830		Proliferating-cell nucleolar antigen P120	2,67	0,21882
Q9V1T5	PAB2120	L3	50S ribosomal protein L3	26,80	0,22726
Q9V1T4	PAB2119		Uncharacterized protein	1,25	0,23420
Q9V1U9	PAB2130	L5	50S ribosomal protein L5	6,80	0,23962
Q9UY85	PAB1284	DHH-S1	RecJ-like phosphoesterase	19,80	0,24743
Q9V1V4	PAB2135	L18	50S ribosomal protein L18	9,50	0,26505

Q9V0E3	PAB7213	Nop10	Ribosome biogenesis protein Nop10	1,01	0,28982
Q9V1T6	PAB2121	L4	50S ribosomal protein L4	26,46	0,29564
Q9V1T8	PAB2122	L2	50S ribosomal protein L2	14,28	0,31161
P62008	PAB0460	L7Ae	50S ribosomal protein L7Ae	9,97	0,34135
Q9V0G8	PAB1813	S19e	30S ribosomal protein S19e	6,82	0,36887
Q9UYZ6	PAB1477		Uncharacterized protein	1,11	0,39465
Q9UXX7	PAB1136	Rpp30	Ribonuclease P protein component 3	1,36	0,43767
Q9V256	PAB0150		Uncharacterized protein	4,24	0,45317
Q9V0V7	PAB0465	aEF1A	Elongation factor 1A Translation	78,93	0,46498
P62010	PAB0362	S11	30S ribosomal protein S11	4,34	0,47616
Q9V2L4	PAB2305	Nop5	RNA methylating RNP complex (C/D box)	15,14	0,47890
Q9V1U1	PAB2125	S3	30S ribosomal protein S3	2,74	0,48456
Q9UXS5	PAB1167	L10	50S ribosomal protein L10	24,56	0,50056
Q9V0E1	PAB0570	KPHMT	cofactor synthesis	4,68	0,50953
Q9V2L5	PAB2306	Fibrillarin	rRNA/tRNA 2'-O-methyltransferase	21,30	0,51551
Q9UZW0	PAB1679	Pgk	Phosphoglycerate kinase	19,82	0,52856
		Malate		6,81	0,56287
Q9V0E2	PAB0569	dehydratase	3-isopropylmalate dehydratase		
Q9V076	PAB1751	aRNase J	5'-3' exoribonuclease	10,50	0,56326
Q9UYS6	PAB0950	TFE	Transcription factor E	3,66	0,56795
Q9UXS6	PAB1168	L12	50S ribosomal protein L12	2,57	0,57647
Q9V1F2	PAB0316	DNaG	Exosome component	66,34	0,58191
P61992	PAB0361	S4	30S ribosomal protein S4	3,34	0,70645
G8ZHS0	PAB2163	RPA41	Replication factor A	19,03	0,73054
Q9V191	PAB0368	S2	30S ribosomal protein S2	2,39	0,86352
Q9UYB6	PAB1306	aIF2	translation initiation factor eIF-2B subunit 2	21,10	0,86991
Q9V1Z1	PAB2165	RPA32	Replication factor A	4,76	0,96135

**Tableau 12 : Liste des partenaires protéiques identifiées par pull down-MS/MS de *Pab-Csl4*.**

Voir la légende du tableau 9 et 10.



## 4. Conclusions

De manière remarquable, nous observons un certain nombre de partenaires communs aux quatre réseaux obtenus dans cette étude. Le schéma de la **figure 18** permet de visualiser les partenaires appartenant à plusieurs réseaux. L'ensemble de ces données permet de proposer un lien entre les facteurs ASH-Ski2, aRNase J et Csl4 avec certaines protéines constitutives du ribosome, en particulier avec les protéines de la grande sous unité L2, L3, L7Ae, L10, L10e, L13 et L30 présentes dans le réseau partagé de Csl4 et aRNase J ou Csl4 et ASH-Ski2. La protéine S4e, seul lien avec des protéines ribosomiques du réseau Rrp41, semble commune aux réseaux de aRNase J et de Csl4. Il faut noter que la protéine L7Ae, en plus de sa fonction en tant que protéine ribosomale, est aussi impliquée dans la modification des ARN dans les complexes RNP à boîte C/D.

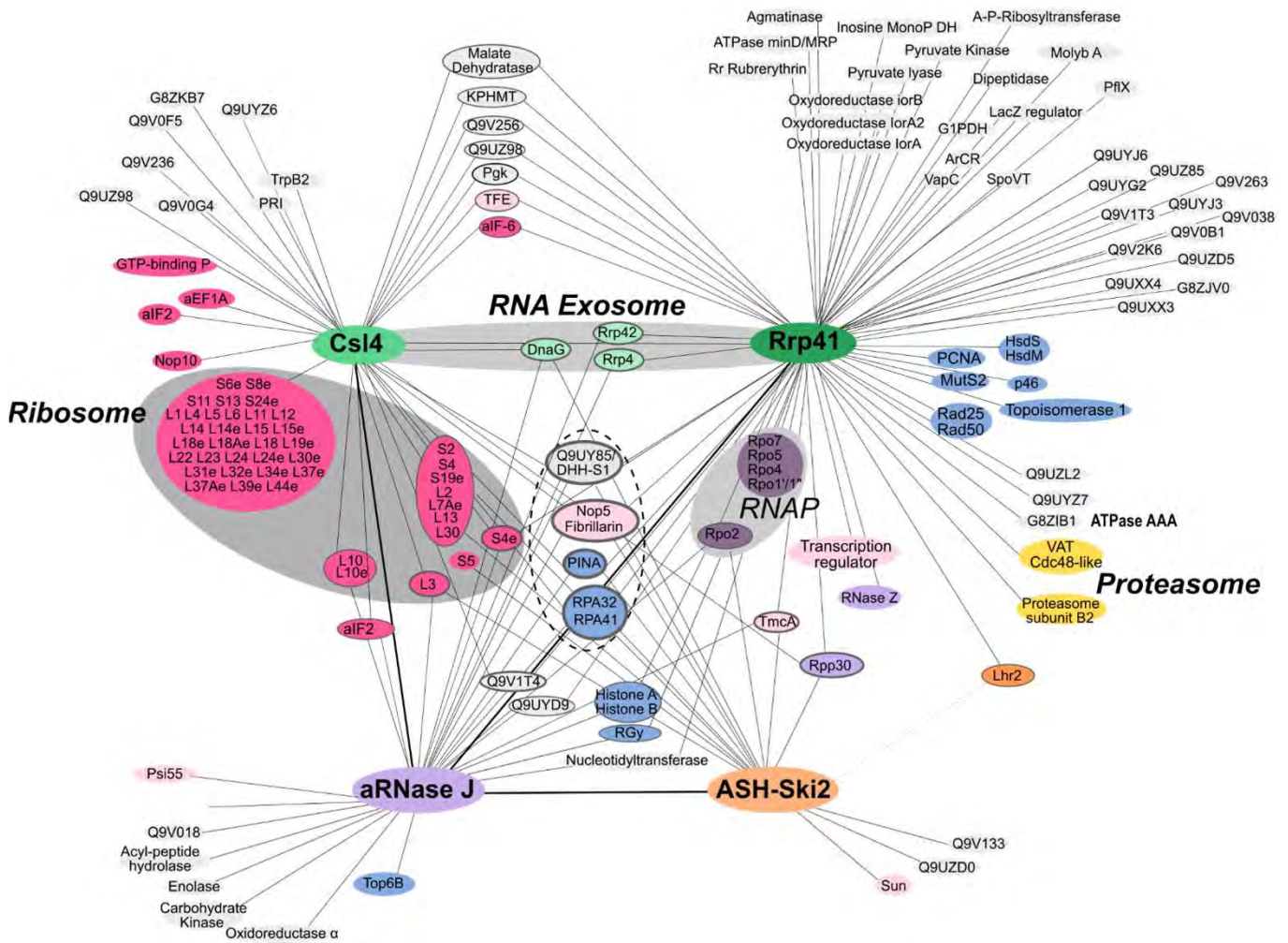
De manière remarquable, au centre, se situent six candidats communs: l'ATPase PINA, montrée comme impliquée dans les jonctions de Holliday, les facteurs de réplication RPA31 et RPA41 liant l'ADN simple brin, les composants de la particule RNP à boîtes C/D (Nop5, Fibrillarin) et une protéine annotée comme une RecJ-like que nous avons nommée DHH-S1 d'après sa composition en domaines protéiques.

On peut noter la présence d'un partenaire commun dans les réseaux de Rrp41 et ASH-Ski2 annoté comme hélicase de type Lhr2 pour « Long Helicase Related » encore non caractérisée chez les Archées. Ce partenaire est en cours de caractérisation par l'équipe et fait l'objet d'étude de thèse de Mirna Hajj.

## III. Interaction entre partenaires du réseau d'interaction de *Pab*-aRNase J

### 1. Co-purification par affinité de protéines recombinantes Proie/Appât

Afin de mettre en évidence et valider des interactions entre les protéines du réseau (**Figure 18**), nous avons établi un protocole de co-purification sur chromatographie d'affinité sur billes de nickel à partir d'un système appât/proie de protéines recombinantes de *P. abyssi* (**Figure 19A**). La protéine appât est exprimée à partir de vecteurs d'expression (pET15b ou pET21b) afin d'obtenir une protéine avec une étiquette (His)<sub>6</sub> en N- ou C-terminal. La protéine proie sans étiquette est exprimée à partir du vecteur d'expression pET11b. Les protéines proie/appât sont exprimées indépendamment. Les extraits cellulaires de culture de *E. coli*



**Figure 18.** Superposition des réseaux d'interaction de aRNase J-(His<sub>6</sub>) (violet) et ASH-Ski2-(His<sub>6</sub>) / (His<sub>6</sub>)-ASH-Ski2 (orange) , (His<sub>6</sub>)-Pab-Rrp41 (vert foncé) et (His<sub>6</sub>)-Pab-Csl4 (vert). Au centre, les facteurs communs aux quatre réseaux sont entourés en pointillé. Les machineries cellulaires sont indiquées en gras.

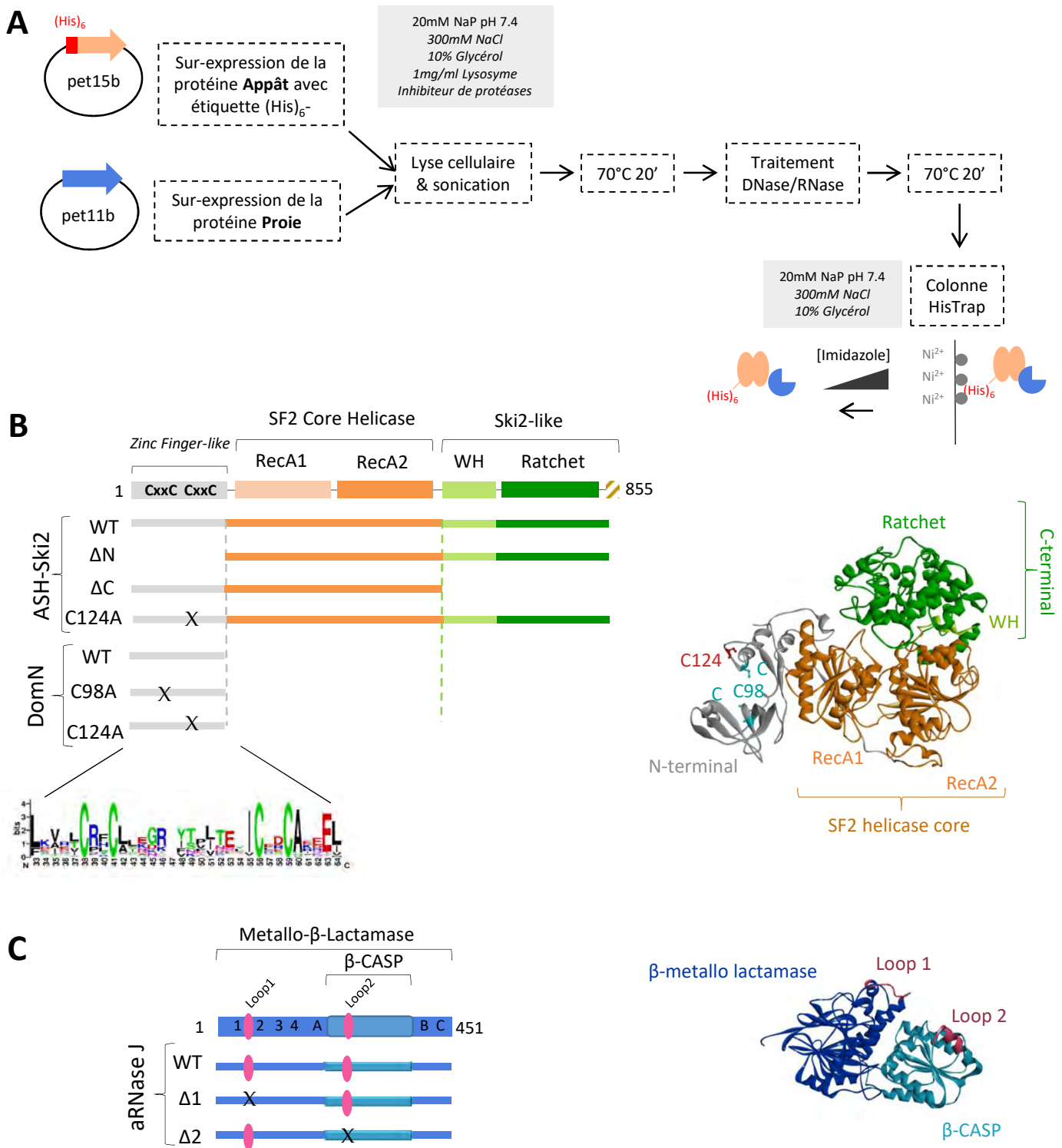
ayant exprimé soit la protéine appât, soit la protéine proie sont mélangées pendant l'étape de lyse. Les protéines thermosensibles de *E. coli* sont éliminées par chauffage à 70°C pour un enrichissement spécifique des protéines thermorésistantes de *P. abyssi*. Les acides nucléiques sont éliminés par un traitement avec un cocktail de nucléases. Après une deuxième étape de chauffage, les extraits clarifiés sont ensuite incubés sur une colonne de Nickel HisTrap qui retient la protéine appât possédant l'étiquette (His)<sub>6</sub> et potentiellement la protéine proie si les deux protéines établissent une interaction dans ces conditions expérimentales. Les étapes de lavage à faible concentration d'Imidazole permettent d'éliminer les contaminants se liant de manière non-spécifique. La protéine appât et ses interactants sont éluées grâce à un gradient linéaire d'imidazole. Le contenu en protéine de chaque fraction collectée est analysé sur un gel SDS-PAGE et coloré au Bleu de Coomassie. La présence de l'appât et de la proie sont révélées par Western Blot avec des anticorps spécifiques de chaque protéine d'intérêt.

## 2. *Pab*-aRNase J et *Pab*-ASH-Ski2 forment un complexe stable *in vitro*

Pour identifier si une interaction directe peut s'établir entre les deux protéines d'intérêt aRNase J et ASH-Ski2 de *P. abyssi*, nous avons tout d'abord envisagé de co-exprimer dans un vecteur d'expression de type pETDuet le couple de protéines recombinantes appât/proie. Après différentes stratégies et tentatives de clonage, nous n'avons pas réussi à obtenir un vecteur contenant les séquences codantes pour *Pab*-ASH-Ski2 et *Pab*-aRNase J, suggérant une toxicité pour *E. coli* dans ce contexte.

- **Vecteurs d'expression utilisés**

Nous avons alors effectué des expériences de co-purification à partir d'un couple proie *Pab*-aRNase J et appât (His)<sub>6</sub>-*Pab*-ASH-Ski2 exprimé indépendamment dans *E. coli*. Nous avons construit des vecteurs exprimant les protéines sauvages ainsi que des vecteurs exprimant des protéines tronquées de certains domaines. Pour *Pab*-ASH-Ski2, des constructions exprimant la protéine sauvage (WT), la protéine tronquée du domaine N-terminal ( $\Delta$ N) (domaine spécifique des hélicases-like de type ASH-Ski2 d'euryarchées), la protéine tronquée du domaine C-terminal (domaine spécifique de la famille des hélicases de type Ski2) ( $\Delta$ C) et également des protéines possédant des résidus substitués par des Alanine au niveau des Cystéines (C98A et C124A) du motif en doigt de zinc (zinc finger) du domaine N-terminal ont été utilisées (**Figure 19B**). En plus de *Pab*-aRNase J, les vecteurs exprimant la protéine sauvage



**Figure 19. A.** Schéma expérimental de la technique de co-purification sur chromatographie d'affinité sur billes de nickel à partir d'un système appât/proie de protéines recombinantes de *P. abyssi*. Les différents tampons sont indiqués et décrits dans la partie Matériel et Méthodes. **B.** Domaines d'organisation (panel de gauche) et modèle structurale (Panel de droite) réalisé avec le logiciel Phyre2 de la protéine *Pab*-ASH-Ski2. Le Weblogo des acides aminés conservés du domaine N-terminal (en gris) de *Pab*-ASH-Ski2 est montré. Le core SF2 est composé des domaines RecA1 et RecA2 (orange) et le domaine C-terminal est composé des domaines WH et Ratchet (vert). **C.** Domaines d'organisation (panel de gauche) et modèle structurale (panel de droite) obtenu avec le logiciel Phyre2 de *Pab*-aRNase J. Le domaine metallo  $\beta$ -Lactamase est en bleu foncé et le domaine  $\beta$ -CASP en bleu ciel. Les motifs Loop1 et 2 en rose.

et les protéines ayant une délétion au niveau des domaines spécifiques appelés « Loop1 » ( $\Delta 1$ ) ou « Loop2 » ( $\Delta 2$ ) ont été utilisés (**Figure 19C**). Les domaines « Loop1 » et « Loop2 » ont été montrés comme importants pour l'activité exoribonucléolytique de *Pab*-aRNase J par l'équipe (Clouet-d'Orval et al. 2010).

- ***Pab*-aRNase J et (His)<sub>6</sub>-*Pab*-ASH-Ski2 forment un complexe stable**

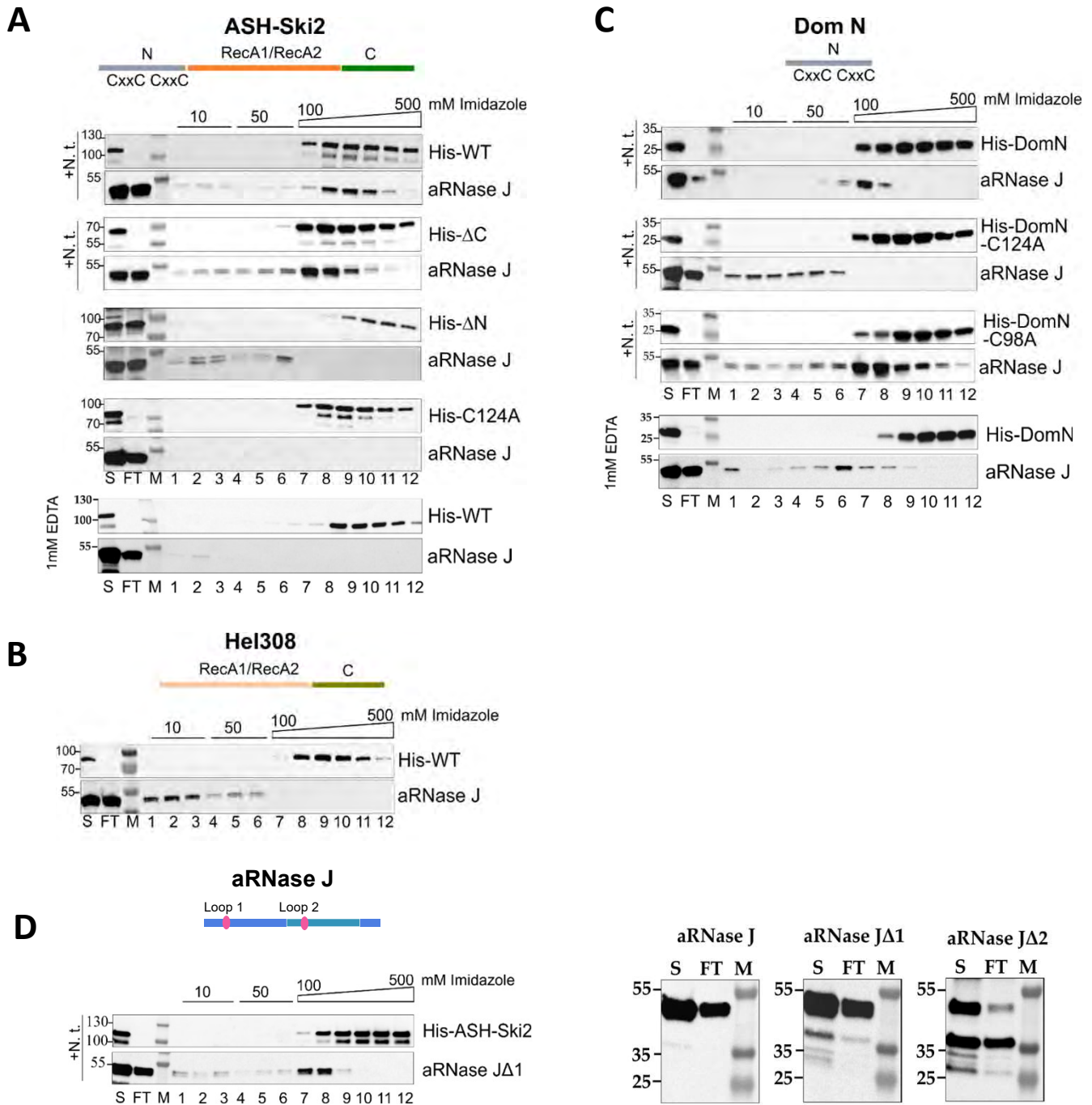
L'ensemble des résultats obtenus à partir des expériences de co-purification sont montrées dans la **Figure 19**. Les protéines d'intérêt *Pab*-aRNase J et (His)<sub>6</sub>-*Pab*-ASH-Ski2 et leurs variants sont révélées par Western Blot. Nous observons qu'avec les versions sauvages des protéines recombinantes, *Pab*-aRNase J est co-éluée avec (His)<sub>6</sub>-*Pab*-ASH-Ski2. En effet les deux espèces sont détectées dans les fractions correspondant au gradient d'imidazole (fractions 8 à 10, panel du haut, **Figure 20A**). Nous proposons que *Pab*-aRNase J et (His)<sub>6</sub>-*Pab*-ASH-Ski2 forme un complexe protéique stable à haute force ionique (300mM NaCl, condition de l'expérience) et indépendamment de la présence d'acides nucléiques. En effet le traitement nucléase n'altère pas le profil observé.

L'expérience témoin a été réalisée avec la protéine proie *Pab*-aRNase J seule, sans protéine appât (Voir les données supplémentaires **Figure S2A** dans la **publication 2** (Phung et al. 2020)). Dans ce cas *Pab*-aRNase J est éliminée dans les lavages. De même lorsque la *Pab*-aRNase J est mise en présence de *Pab*-Hel308 (autre hélicase de la famille Ski2 phylogénétiquement distincte de ASH-Ski2, voir Partie IV de l'introduction), la proie n'est pas retenue sur la colonne (**Figure 20B**).

- **Le domaine N-terminal (DomN) de (His)<sub>6</sub>-*Pab*-ASH-Ski2 est essentiel et suffisant pour la formation d'un complexe stable**

Pour établir quels sont les domaines de la protéine appât et de la protéine proie respectivement impliqués dans l'interaction protéique observée, nous avons réalisé ce même type d'expérience avec des protéines tronquées ou altérées (**Figure 19B**). L'absence du domaine C-terminal de (His)<sub>6</sub>-*Pab*-ASH-Ski2 n'altère pas le profil de co-purification observée. Le même profil que dans le cas de la version sauvage de (His)<sub>6</sub>-*Pab*-ASH-Ski2 est observé (**Figure 20A**). A l'opposé, la co-purification avec la version tronquée pour le domaine N-terminal ( $\Delta N$ ) montre un profil altéré indiquant que ce domaine est important dans la formation du complexe. Nous proposons que le domaine N-terminal soit le domaine qui permet d'établir une interaction avec *Pab*-aRNase J.





**Figure 20.** Co-purification *in vitro* de *Pab*-aRNase J avec *Pab*-(His)<sub>6</sub>-ASH-Ski2 et *Pab*-(His)<sub>6</sub>-Hel308. **A.** Représentation schématique des domaines de *Pab*-ASH-Ski2 et co-purification *in vitro* de *Pab*-aRNase J avec *Pab*-(His)<sub>6</sub>-ASH-Ski2 et ses variants. La protéine *Pab*-ASH-Ski2 sauvage (His-WT), délétée du domaine N-terminal (His-ΔN), du domaine C-terminal (His-ΔC), avec une mutation ponctuel (His-C124A) sont utilisées comme appât. Les fractions ont été analysées par Western Blot en utilisant un anticorps His-Tag ou spécifique dirigé contre aRNase J. S: surnageant déposé sur la colonne HisTrap, FT: Flow through; les fractions 1-3 et 4-6 sont les lavages avec 10mM et 50mM d'imidazole respectivement; l'éluion est faite avec un gradient d'imidazole de 100 à 500mM (fractions 7 à 12). Les expériences réalisées avec un traitement aux nucléases sont notées +N.t. **B.** Co-purification avec *Pab*-(His)<sub>6</sub>-Hel308 comme appât. **C.** Co-purification avec le domaine N-terminal de *Pab*-ASH-Ski2 sauvage (His-DomN) ou avec une mutation ponctuelle (His-DomN-C98A et His-DomN-C124A). **D.** Copurification de *Pab*-(His)<sub>6</sub>-ASH-Ski2 avec *Pab*-aRNase J délété du motif Loop1 (aRNase JΔ1).

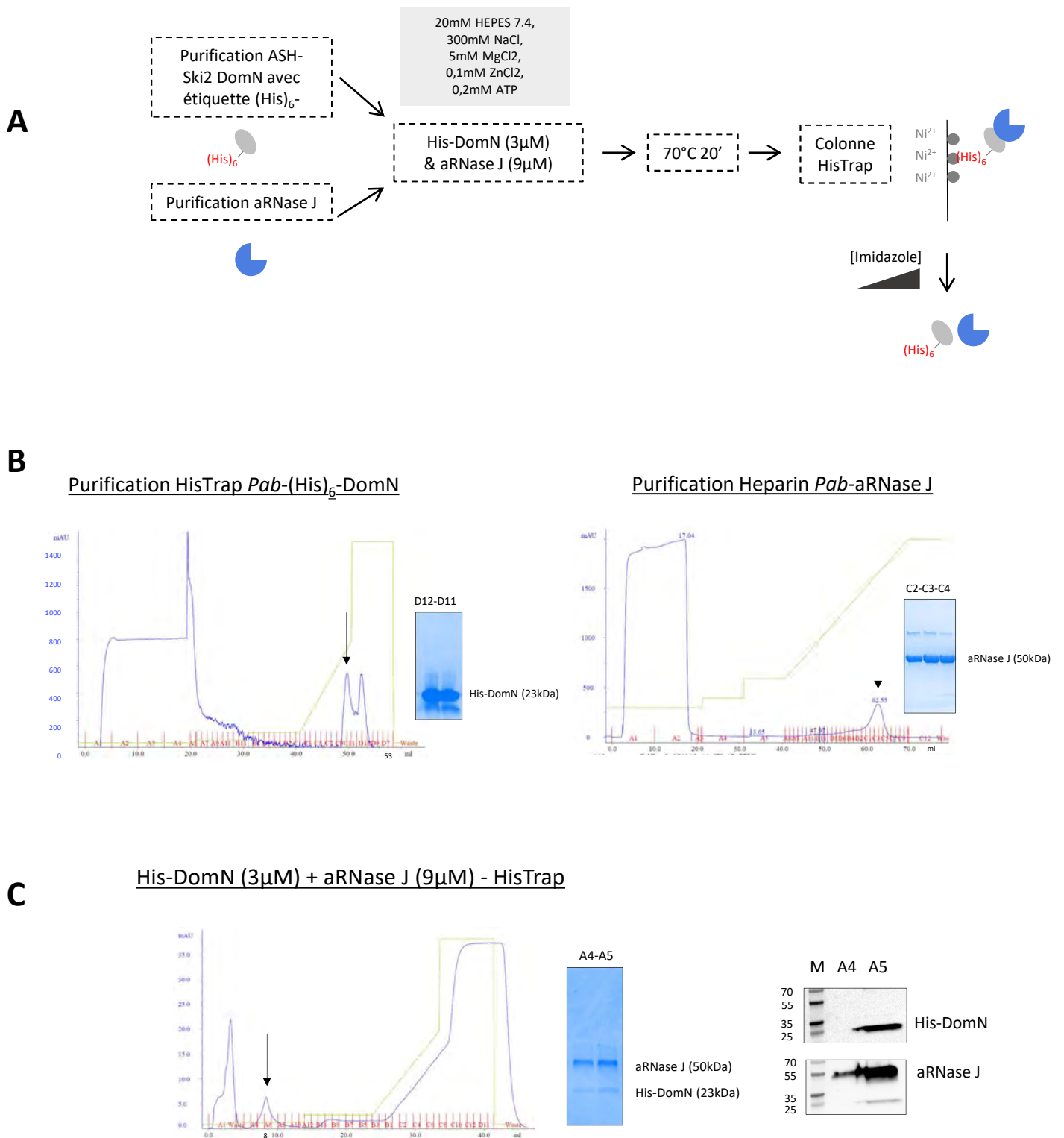


Des analyses comparatives de séquences ont permis de définir un WebLogo du domaine N-terminal qui montre que quatre cystéines dans le domaine N-terminal de ASH-Ski2 sont strictement conservées et seraient typiques d'un motif en doigt de zinc (voir **Figure 19B & Publication 2** (Phung et al. 2020)). La protéine proie (His)<sub>6</sub>-*Pab*-ASH-Ski2 avec un résidu cystéine 124 substitué par un résidu en Alanine (C124A) n'est pas en mesure de retenir *Pab*-aRNase J sur la colonne. Nos résultats montrent donc que le résidu C124 est important pour l'interaction protéine-protéine entre ASH-Ski2 et aRNase J.

Pour montrer que le domaine N-terminal de *Pab*-ASH-Ski2 est le domaine d'interaction avec *Pab*-aRNase J, le domaine (DomN) a été exprimé indépendamment du reste de la protéine et a été testé pour sa capacité à retenir *Pab*-aRNase J sur la colonne d'affinité (**Figure 20C**). On observe que *Pab*-aRNase J est co-purifié avec le domaine (His)<sub>6</sub>-*Pab*-DomN indépendamment d'un traitement préalable aux nucléases. De même que pour la protéine pleine taille, la substitution C124A induit une altération du profil d'élution de *Pab*-aRNase J (**Figure 20C**). On note cependant que la substitution de la cystéine 98 n'affecte pas l'interaction dans ce contexte.

Au vue de ces derniers résultats, nous proposons que le domaine accessoire N-terminal de *Pab*-ASH-Ski2 puisse se structurer en un motif en doigt de zinc qui serait impliqué dans l'interaction avec *Pab*-aRNase J. Pour tester cette hypothèse, nous avons réalisé les mêmes expériences de co-purification *in vitro* en présence d'EDTA (1mM) pour chélater les ions divalents pouvant participer à la structuration du motif en doigt de zinc. Dans cette condition, l'élution de *Pab*-aRNase J est altérée et s'effectue dans les fractions de lavages à 10 mM et 50 mM d'imidazole (fractions 1 à 7) que ce soit avec les appâts (His)<sub>6</sub>-*Pab*-ASH-Ski2 ou (His)<sub>6</sub>-*Pab*-DomN (**Figure 20A et 20C**, dernier panel).

La co-purification est également observée entre les protéines recombinantes purifiées à homogénéité (**Figure 21A**). *Pab*-aRNase J est purifiée sur une colonne héparine (**Figure 21B** panel de droite). (His)<sub>6</sub>-*Pab*-Dom N est purifié sur une colonne de Nickel. Après dialyse dans le tampon 20mM HEPES 7.4, NaCl 300mM, 5mM MgCl<sub>2</sub>, 0,1mM ZnCl<sub>2</sub>, 0,2mM ATP et concentration, les deux espèces purifiées sont mélangées dans un ratio 1 pour 3, incubées et injectées sur une colonne HisTrap selon le protocole de la **Figure 21A**. On observe alors que la protéine appât (His)<sub>6</sub>-*Pab*-DomN n'est plus retenue sur la colonne d'affinité en présence de



**Figure 21.** Interaction en FPLC du domaine N-terminal de ASH-Ski2 (DomN) et aRNase J. **A.** Méthode expérimentale de la co-purification des protéines purifiées à homogénéités *Pab*-(His)<sub>6</sub>-ASH-Ski2-DomN (His-DomN) et *Pab*-aRNase J préalablement purifiées (expliqué dans le Matériel et méthodes, partie II). **B.** Chromatogramme de la purification de *Pab*-(His)<sub>6</sub>-ASH-Ski2-DomN sur colonne de Nickel (HisTrap) (Panel de gauche) et chromatogramme de la purification de *Pab*-aRNase J sur colonne Heparin (Panel de droite). **C.** Chromatogramme du mélange *Pab*-(His)<sub>6</sub>-ASH-Ski2-DomN / *Pab*-aRNase J injecté sur colonne de Nickel et Western Blot du pic (Fractions A4-A5) contenant aRNase J et His-DomN.

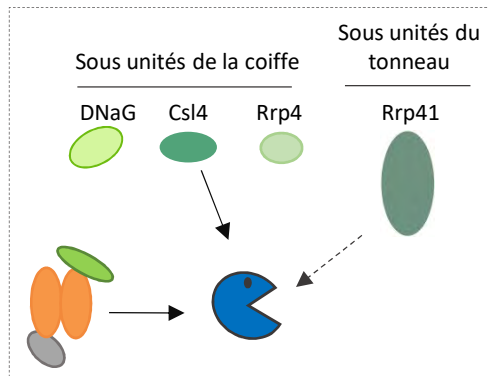
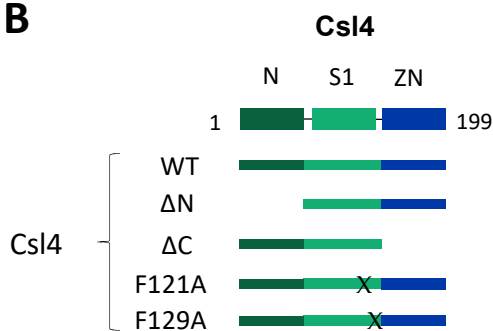
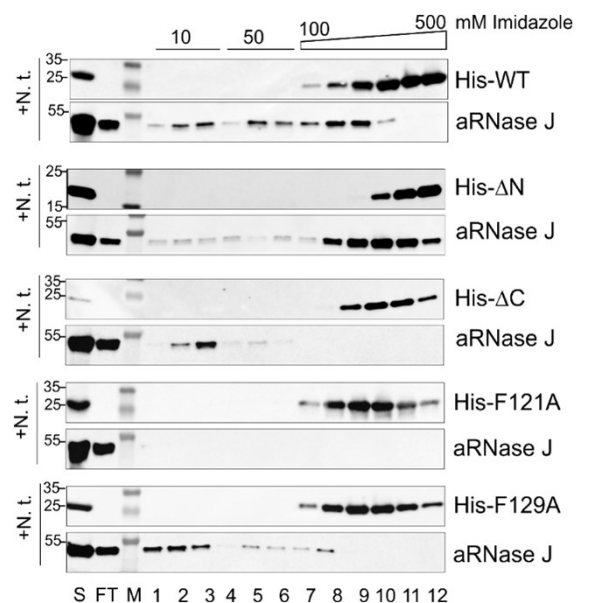
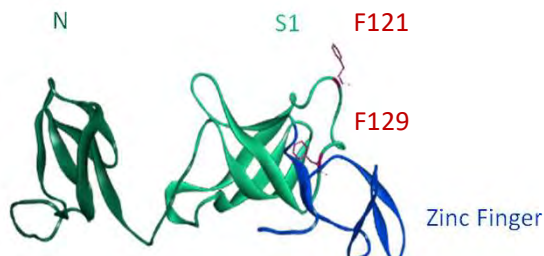
*Pab*-aRNase J et se retrouve dans les mêmes fractions de lavage. Ce résultat témoigne de la formation d'un complexe entre *Pab*-aRNase J et (His)<sub>6</sub>-*Pab*-Dom N de ASH-Ski2.

- **Le domaine Loop1 de *Pab*-aRNase J n'est pas requis pour la formation du complexe (His)<sub>6</sub>-*Pab*-ASH-Ski2/ *Pab*-aRNase J.**

Nous avons essayé réciproquement d'identifier les domaines de *Pab*-aRNase J importants pour la formation du complexe *Pab*-aRNase J/*Pab*-ASH-Ski2. Nous avons pour cela réalisé une série d'expériences avec les protéines recombinantes proies *Pab*-aRNase JΔ1 et *Pab*-aRNase JΔ2, délétées des motifs Loop 1 et 2 respectivement (**Figure 19C**). La co-purification de la proie *Pab*-aRNase JΔ1 avec l'appât (His)<sub>6</sub>-ASH-Ski2 est observée (**Figure 20D**). Le motif Loop1 n'est donc pas important pour la formation du complexe *Pab*-aRNase J/*Pab*-ASH-Ski2. En ce qui concerne la proie *Pab*-aRNase JΔ2, cette espèce est instable puisque l'on observe l'apparition d'une espèce protéique au-dessus de 35kD qui ne correspond pas à la protéine de pleine taille de 50kD environ (**Figure 20D**, panel de droite). La protéine *Pab*-aRNase JΔ2 est probablement plus sensible à la protéolyse au cours des étapes de co-purification bien qu'un cocktail d'anti-protéases soit ajouté au cours de l'étape de lyse cellulaire (voir dans « matériels et méthodes »). On ne peut donc pas conclure sur l'importance du domaine Loop2 dans la formation du complexe *Pab*-aRNase J/*Pab*-ASH-Ski2.

### **3. *Pab*-aRNase J et la sous-unité de la coiffe *Pab*-Csl4 de l'exosome à ARN forment un complexe stable *in vitro***

Les sous-unités de l'exosome à ARN, l'hélicase ASH-Ski2 et la ribonucléase aRNase J partagent un réseau d'interaction commun (**Figure 18**). Afin de caractériser si certaines sous-unités de l'exosome sont susceptibles d'interagir avec aRNase J et ASH-Ski2, nous avons entrepris des expériences de co-purification par chromatographie d'affinité sur colonne HisTrap comme décrit précédemment (**Figure 19A**). Les protéines appât (His)<sub>6</sub>-*Pab*-Rrp41, (His)<sub>6</sub>-*Pab*-Rrp4, (His)<sub>6</sub>-*Pab*-Csl4 et (His)<sub>6</sub>-*Pab*-DnaG exprimées respectivement avec le système d'expression pET dans *E. coli* ont été utilisées pour identifier si des interactions protéiques stable pouvaient s'établir avec *Pab*-aRNase J et *Pab*-ASH-Ski2 (Voir les **Figure 5** & **Figure S2C** de la **Publication 2** (Phung et al. 2020)). Les interactions *in vitro* mises en évidence sont résumées sur la **Figure 22A**. Comme montré dans la **publication 2**, les appâts (His)<sub>6</sub>-*Pab*-Rrp4 et (His)<sub>6</sub>-*Pab*-DnaG ne

**A****B****C****D**

Csl4 S1 domain motifs

co-distributed with aRNase J



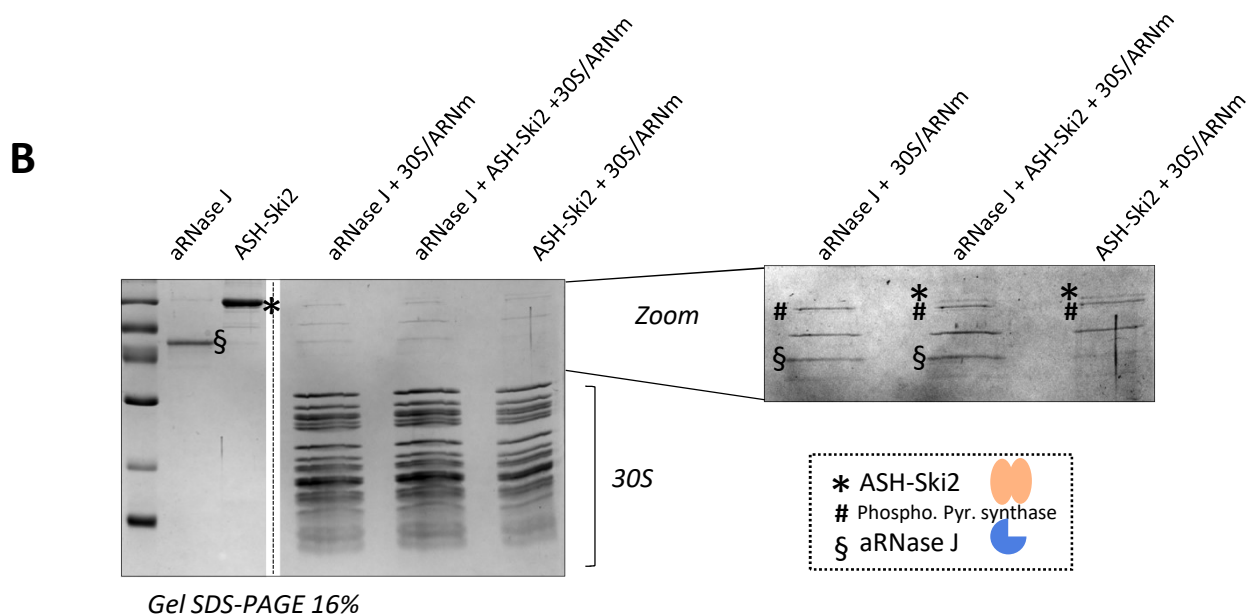
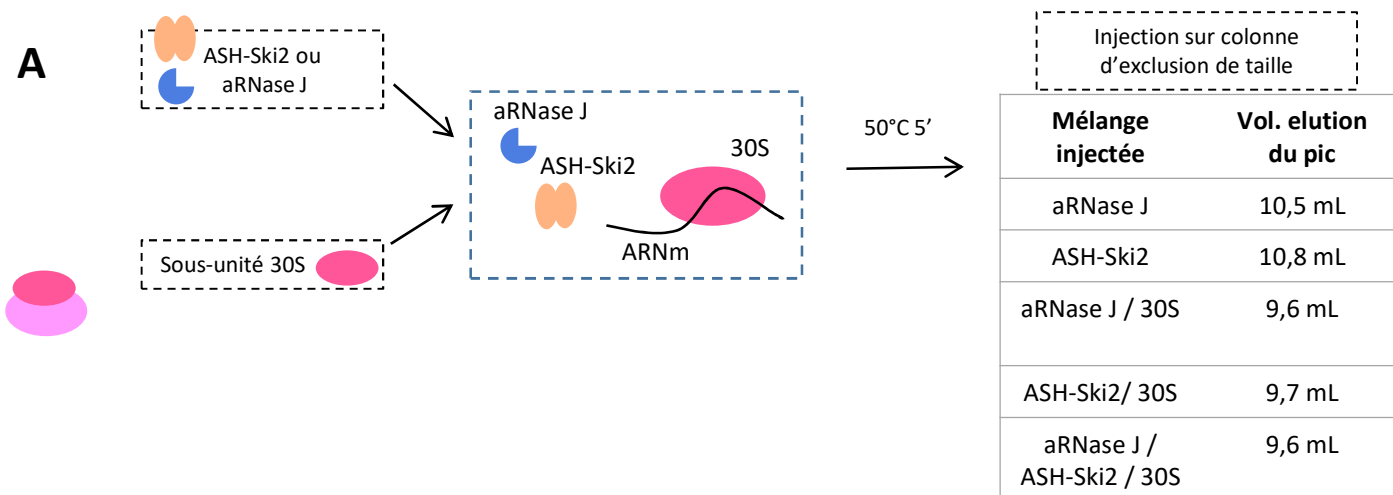
not distributed with aRNase J



**Figure 22.** Co-purification *in vitro* de *Pab*-aRNase J avec les sous unités de l'exosome à ARN. **A.** Représentation schématique de l'exosome à ARN et des interactions détectées dans ces expériences. Les flèches en traits pleins signifie une interaction résistante aux traitements nucléases alors que les flèches en pointillé correspond à une interaction sensible aux traitements de nucléases. **B.** Représentation schématique de *Pab*-Csl4. **C.** Co-purification de *Pab*-aRNase J avec les protéines *Pab*-Csl4 sauvage (His-WT), délété du domaine N-terminal (His-ΔN) ou C-terminal (His-ΔC) ou présentant une mutation ponctuelle (His-F121A et His-F129A) utilisées comme appâts. **D.** Modèle structurelle de *Pab*-Csl4 effectué avec le logiciel Phyre 2. Les Weblogo des acides aminés conservés lorsque Csl4 et aRNase J sont co-distribués à travers l'arbre des archées ou non sont indiqués.

permettent de co-purifier ni *Pab*-aRNase J, ni *Pab*-ASH-Ski2. L'appât (His)<sub>6</sub>-*Pab*-Rrp41 (sous-unité catalytique de l'exosome) permet de co-purifier *Pab*-aRNase J en absence de traitement aux nucléases suggérant que l'interaction observée est médiée par des acides nucléiques (Voir la Figure 5B de la Publication 2 (Phung et al. 2020)).

Dans cette partie, nous détaillerons uniquement les résultats obtenus avec l'appât (His)<sub>6</sub>-*Pab*-Csl4, une des sous-unités constitutives de la coiffe de l'exosome à ARN (voir la partie III de l'introduction). La sous-unité de la coiffe Csl4 de l'exosome est constituée de trois domaines distincts avec un domaine central de type S1 et un domaine en C-terminal de type Zinc Finger (Figure 22B). Nos résultats montrent que l'appât (His)<sub>6</sub>-*Pab*-Csl4 interagit avec *Pab*-aRNase J pour former un complexe stable qui est résistant au traitement aux nucléases. On observe la co-occurrence de (His)<sub>6</sub>-*Pab*-Csl4 et *Pab*-aRNase J dans les fractions d'élution 7, 8, 9 et 10 (Figure 22C, panel du haut). Afin de caractériser les domaines de *Pab*-Csl4 important pour la formation de ce complexe, nous avons réalisé des expériences de co-purification avec des versions de *Pab*-Csl4 tronquées en N-terminal ( $\Delta$ N), en C-terminal ( $\Delta$ C) et invalidées pour les résidus Phénylalanines 121 (F121A) et 129 (F129A) (Figure 22B). Seul l'appât, (His)<sub>6</sub>-*Pab*-Csl4 $\Delta$ N est capable de retenir *Pab*-aRNase J sur la colonne (Figure 22C, 2ème panel). On peut conclure que le domaine N-terminal de Csl4 n'est pas requis pour l'interaction avec *Pab*-aRNase J. Au contraire, le domaine C-terminal, qui correspond à un motif en doigt de zinc (Figure 22D), est critique pour la formation du complexe *Pab*-aRNase J/ *Pab*-Csl4. Pour identifier les résidus de *Pab*-Csl4 importants pour la formation de ce complexe, nous avons recherché en collaboration avec l'équipe de G. Fichant, les résidus conservés dans la séquence de Csl4 dans le contexte d'une co-distribution taxonomique de aRNase J et Csl4. Nous avons obtenu des séquences Weblogo dérivées d'alignements multiples de séquences de Csl4 co-distribuées ou pas avec aRNase J (Figure 22D). En effet les gènes codant pour les sous-unités de l'exosome à ARN sont présents dans la majorité des génomes archées contrairement au gène codant pour aRNase J qui est présent uniquement chez les Euryarchées (Voir la Figure 1 de la Publication 1 (Clouet-d'Orval et al. 2018)). Plusieurs résidus du domaine S1 de Csl4 ont été identifiés comme conservés lors de la co-distribution Csl4/aRNase J, dont les résidus F121 et F129 qui semblent exposés au solvant dans le modèle structural de Csl4 de *P. abyssi* (Figure 22D). On observe que la substitution des résidus Phénylalanine F121 et F129 en résidus Alanine invalident la formation du complexe, respectivement (Figure 22C). Le profil d'élution



**Figure 23.** Chromatographie d'exclusion de taille en HPLC des protéines *Pab*-ASH-Ski2 et *Pab*-aRNase J avec la sous unité 30S du ribosome. **A.** Schéma expérimental de l'expérience d'interaction en chromatographie d'exclusion de taille des protéines ASH-Ski2 et aRNase J avec la sous unité du ribosome 30S de *P. abyssi* en HPLC. La sous unité 30S a été purifiée à partir d'un culot de cellules de *P. abyssi* par une centrifugation sur un coussin de saccharose, injection sur une colonne Sephacryl S2400 et gradient de saccharose 10-30% (Pour la méthode, voir Coureux et al. 2016). Un contaminant reste cependant présent de façon prépondérante et correspond à la phosphoenol pyruvate synthase (91kDa) (indiqué ici par un #). Les protéines *Pab*-aRNase J et *Pab*-ASH-Ski2 ont été purifiées comme expliqué dans le matériel et méthode, partie II. 130pmol de *Pab*-aRNase J et/ou *Pab*-ASH-Ski2 sont mis en présence de 30pmol de *Pab*-30S dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol et chauffées 5' à 50°C puis injectée sur une colonne Agilent Bio SEC- 5. Les fractions des pics majoritaires ont été concentrées et déposées sur un gel SDS PAGE coloré au bleu de Coomassie présenté en ci-dessus (**B**). Le pic correspondant à *Pab*-aRNase J est indiqué par un § et celui de *Pab*-ASH-Ski2 par un \*.



de *Pab*-aRNase J est altéré par rapport au profil observé avec la version sauvage. Nous proposons que les résidus F121 et F129 du domaine S1 sont critiques pour la formation d'une conformation de *Pab*-Csl4 capable d'interagir avec *Pab*-aRNase J *in vitro*.

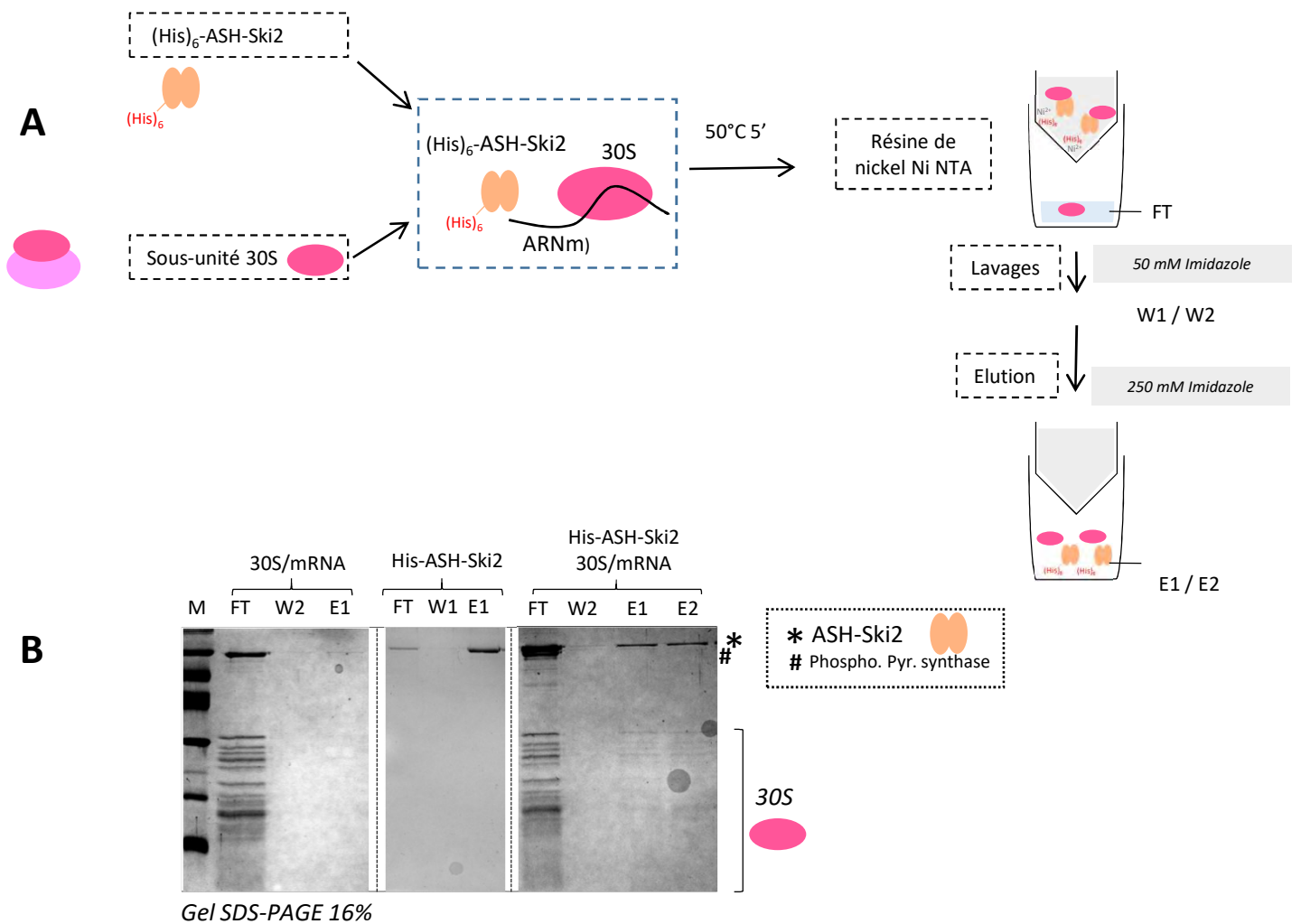
#### 4. Etudes préliminaires de co-purification de *Pab*-aRNase J et *Pab*-ASH-Ski2 avec la sous-unité 30S purifiée du ribosome de *P. abyssi*

Les expériences présentées dans cette partie ont été effectuées en collaboration avec l'équipe de E. Schmitt (Ecole Polytechnique, Palaiseau).

Les réseaux d'interactions obtenus dans la partie I des résultats mettent en évidence une étroite relation entre *Pab*-aRNase J et *Pab*-ASH-Ski2 et certaines protéines du ribosome. En effet le réseau de *Pab*-aRNase J et de *Pab*-ASH-Ski2, respectivement, révèle des protéines ribosomiques de la particule 30S et 50S (S4e, L3, L10 et L10e pour *Pab*-aRNase J / S2, S4, S19e, S5, L2, L30, L13, L7Ae et L3 pour *Pab*-ASH-Ski2).

Afin d'établir si des interactions spécifiques existent entre *Pab*-aRNase J, *Pab*-ASH-Ski2 et le ribosome, nous avons initié des expériences de co-purification à partir de la sous-unité 30S de ribosome de *P. abyssi* purifiée à homogénéité disponible dans l'équipe de E. Schmitt (Ecole Polytechnique, Palaiseau) (voir pour le protocole de purification de la sous unité 30S (Coureux et al. 2016)). Nous avons réalisé deux types d'expériences de co-purification avec une séparation sur chromatographie d'exclusion de taille en HPLC et une sur chromatographie d'affinité sur bille de nickel, respectivement. Ces techniques ont été élaborées par l'équipe de E. Schmitt et sont utilisées classiquement pour identifier les interactions entre la particule 30S et les facteurs d'initiation de la traduction chez *P. abyssi* (Coureux et al. 2016).

Pour l'analyse par chromatographie d'exclusion de taille, 150pmol de protéines recombinantes *Pab*-aRNase J et *Pab*-ASH-Ski2 purifiées à homogénéité sont incubées séparément ou ensemble avec 30pmol de particule 30S et 150pmol d'ARNm pendant 5min à 50°C avant injection sur une colonne d'exclusion de taille (Agilent Bio SEC- 5) placée sur un système HPLC à température ambiante (tampon à 150 ou 300mM NH<sub>4</sub>Cl) (**Figure 23A**). Les volumes d'élution du pic majoritaire dans les différentes conditions testées sont indiqués sur le tableau de la **Figure 22A** (Les chromatogrammes sont en **Annexe-Figures 1 à 5**). Après concentration des fractions correspondantes au pic d'élution, une analyse des échantillons sur un gel SDS-PAGE coloré au bleu de Coomassie permet de visualiser les espèces protéiques de



**Figure 24.** Chromatographie d'affinité *in vitro* d'ASH-Ski2 avec la particule 30S du ribosome de *P. abyssi*. **A.** Schéma expérimental de la chromatographie d'affinité sur billes de nickel de *Pab*-(His)<sub>6</sub>-ASH-Ski2 utilisée comme appât pour capturer *Pab*-30S. 150pmol de (His)<sub>6</sub>-*Pab*-ASH-Ski2 (Appât), 30pmol de *Pab*-30S (Proie) et 150pmol d'un ARNm synthétique ont été mélangés et chauffés 5min à 50°C dans un tampon 10mM HEPES 7.5, 10mM MgAc, 150mM NH<sub>4</sub>Cl, 3mM β-Mercaptoéthanol puis incubés 15min dans une colonne en plastique contenant la résine de nickel. Après lavage avec une faible concentration en imidazole (30mM), les composés retenus sur la résine ont été élués avec 250mM d'Imidazole. **B.** Gel SDS PAGE coloré au bleu de Coomassie. FT correspond au flow through, W aux lavages, E1 et E2 aux éluions et M au marqueur protéique. La protéine *Pab*-ASH-Ski2 est indiquée par un \* et le contaminant retrouvé de façon général lors de la purification de *Pab*-30S indiqué par un #.

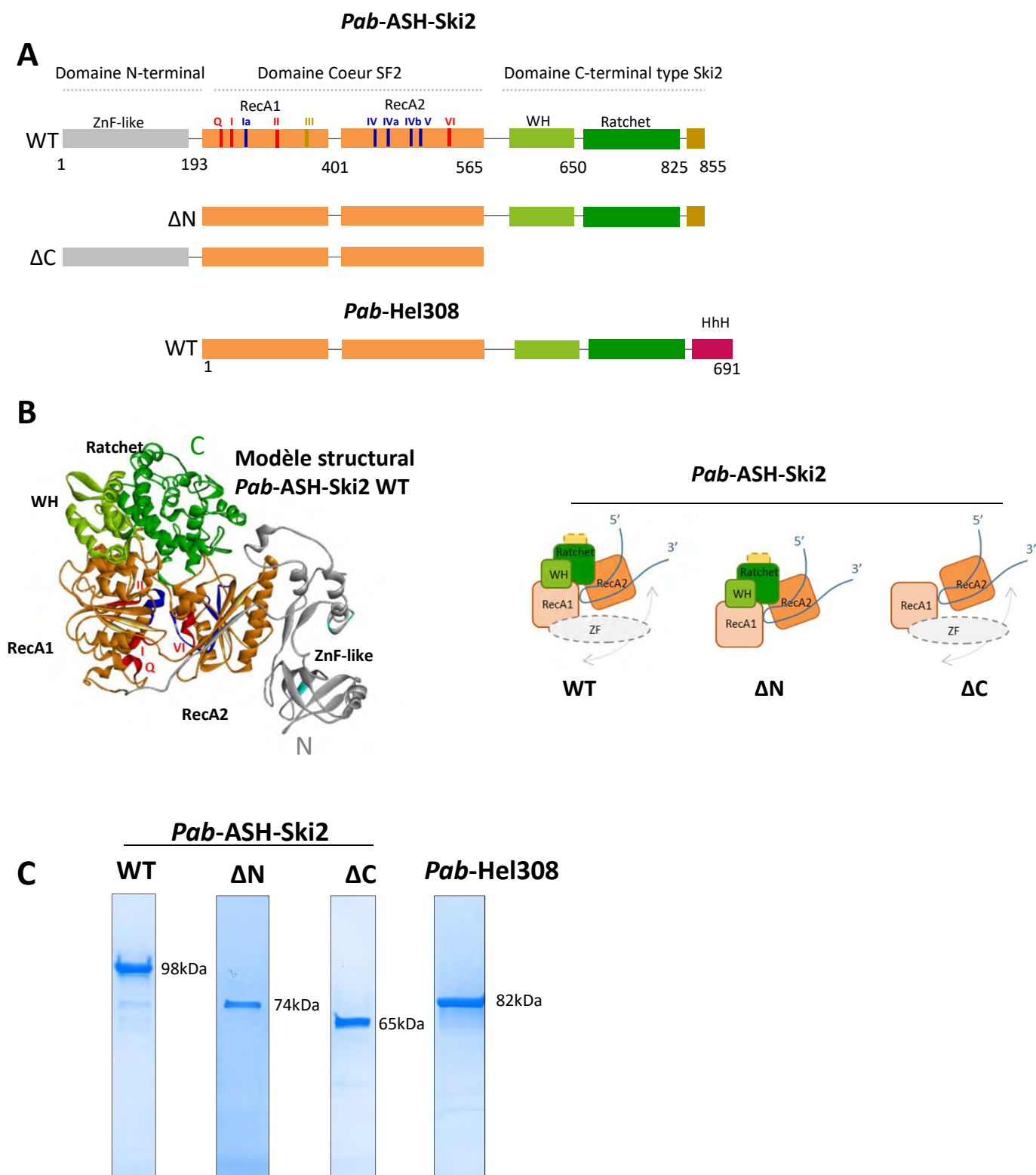
l'échantillon selon leur taille. Nous avons choisi un tampon à 300mM de NH<sub>4</sub>Cl pour éviter la précipitation de *Pab*-aRNase J, protéine stable uniquement à haute force ionique (voir Partie IV des résultats). On observe un décalage vers les hauts poids moléculaires des pics d'élution de *Pab*-aRNase J et *Pab*-ASH-Ski2 dans les deux cas, c'est-à-dire en présence d'une ou des deux protéines simultanément dans le mélange contenant la particule 30S (**Figure 23B**). Ces premiers résultats sont encourageants et permettent de proposer qu'une interaction entre la particule 30S et les protéines *Pab*-aRNase J et *Pab*-ASH-Ski2 pourraient s'établir dans ces conditions expérimentales. Ces expériences devront être répétées et réitérées avec des variants de *Pab*-aRNase J et *Pab*-ASH-Ski2 afin d'identifier l'importance des différents domaines protéiques dans cette interaction.

Pour l'analyse par chromatographie d'affinité, la protéine recombinante (His)<sub>6</sub>-*Pab*-ASH-Ski2 sert d'appât pour capturer la particule 30S. 150pmol de (His)<sub>6</sub>-*Pab*-ASH-Ski2 sont incubées avec 30pmol de particule 30S purifiée et 150 pmol d'ARNm pendant 5min à 50°C avant incubation sur les billes de nickel (**Figure 24A**). Après des lavages à 30 mM d'imidazole et une élution à 250mM d'imidazole, les différentes fractions sont analysées sur gel SDS-PAGE (**Figure 24B**). Comme attendu, la protéine appât seule est retenue sur les billes grâce à son étiquette Histidine en N-terminal. La particule 30S est retenue sur les billes seulement en présence de la protéine appât (His)<sub>6</sub>-*Pab*-ASH-Ski2. Nous proposons que la particule 30S peut donc co-purifiée avec (His)<sub>6</sub>-*Pab*-ASH-Ski2 dans ces conditions expérimentales. Ces résultats ne sont que préliminaires et devront être confirmés par répétition de cette expérience. On pourra également envisager d'utiliser *Pab*-aRNase J-(His)<sub>6</sub> comme appât.

L'ensemble des données obtenues par chromatographies d'exclusion de taille et d'affinité sont en adéquation et sont cohérents avec nos résultats de fractionnement cellulaire montrant que aRNase J et ASH-Ski2 de *P. abyssi* sont présents dans des fractions de haut poids moléculaires contenant les particules ribosomiques (voir la **Figure 6A** de la **Publication 2** (Phung et al. 2020)).

## **IV. Caractérisation biochimique de *Pab*-ASH-Ski2**

L'hélicase Hel308 qui appartient également à la famille des hélicases de type-Ski2 chez les archées est décrite comme une hélicase à ADN impliquée dans la résolution des jonctions de Holliday chez *M. thermautotrophicus*, *P. furiosus* et *S. tokodaii* (Voir Partie IV de



**Figure 25. Les hélicases de la famille Ski2 de *P. abyssi*** **A.** Architecture en domaines de *Pab*-ASH-Ski2 et *Pab*-Hel 308, voir la légende Figure 12 de l'introduction. Les motifs impliqués dans la liaison et l'hydrolyse de l'ATP sont en rouge, dans la coordination avec les acides nucléiques et les sites de fixation des NTP en jaune et dans la liaison aux acides nucléiques en bleu **B.** Modèle structural de *Pab*-ASH-Ski2 construit avec l'outil Phyre2 et les schémas des domaines des variants de *Pab*-ASH-Ski2. **C.** Les protéines recombinantes purifiées à homogénéité sont visualisées sur un gel SDS-PAGE (4-15%) après coloration au bleu de Coomassie. La taille de chaque protéine est indiquée.

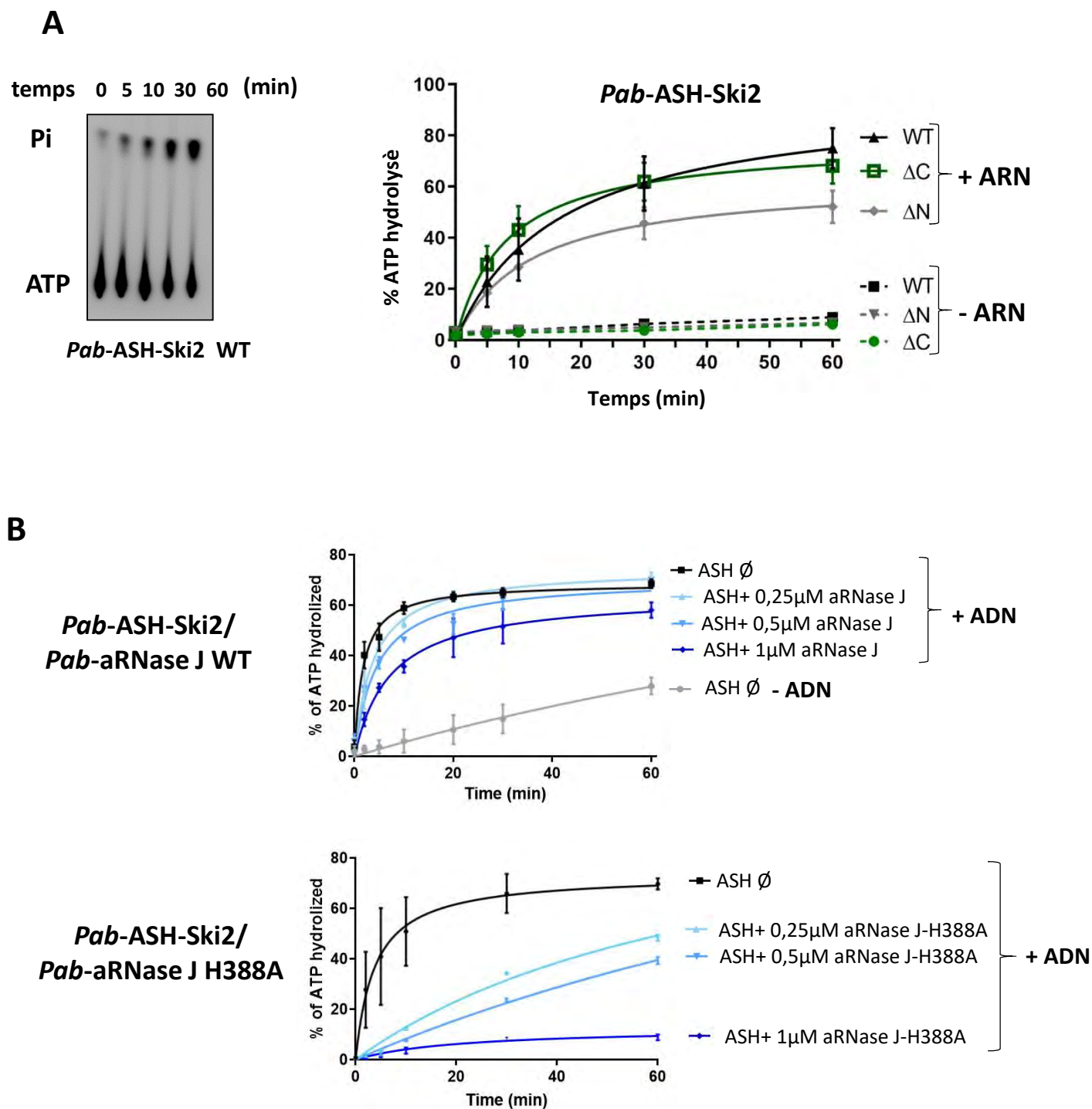
l'introduction) (Guy and Bolt 2005; Li et al. 2008; Fujikane et al. 2005). A ce jour l'activité de ASH-Ski2 chez les Euryarchées reste à déterminer. Dans ce contexte, nous avons analysé les propriétés enzymatiques de ASH-Ski2 de *P. abyssi* (*Pab*-ASH-Ski2) en parallèle de celles de *Pab*-Hel308 (**Figure 25A**).

Nous avons mis au point un panel de tests enzymatiques pour observer la capacité d'hydrolyser de l'ATP et de liaison aux acides nucléiques (ARN et ADN) ainsi que la capacité à former ou à dérouler des molécules d'ADN et/ou ARN appariées. Ce travail avait été initié au cours de la thèse de D.K. Phung (Phung D.K. 2017 Thesis) dans l'équipe. Les résultats préliminaires alors obtenus ont servi de base pour la mise au point des essais enzymatiques décrits dans cette étude. Afin de mener l'étude de la structure fonction des hélicases de type ASH-Ski2 chez les archées, nous avons également comparé les activités de la version sauvage de *Pab*-ASH-Ski2 à celles de variants de *Pab*-ASH-Ski2 (*Pab*-ASH-Ski2 $\Delta$ N et *Pab*-ASH-Ski2 $\Delta$ C). Le but est de comprendre l'importance du domaine accessoire en N-terminal contenant quatre cystéines conservées et du domaine typique des hélicases de type Ski2 en C-terminal sur l'activité et la spécificité de *Pab*-ASH-Ski2 (**Figure 25B**).

Les protéines recombinantes utilisées dans ces tests enzymatiques sont surexprimées dans *E. coli* grâce au système d'expression pET avec le vecteur pET11 qui permet une sur-expression des protéines dans une version sans étiquette. Trois étapes successives ont permis de purifier à homogénéité les protéines d'intérêt avec une première étape de chauffage à 70°C suivie d'une purification sur une colonne Heparin puis sur une colonne échangeuse de cations (SP-HP) pour *Pab*-ASH-Ski2 ou d'anions (Q-HP) pour *Pab*-Hel308 en système FPLC (**Figure 25C**). Cette partie fait l'objet **d'un article en cours de rédaction: The Archaeal Specific Helicase of the Ski2-like family, ASH-Ski2, a novel RNA helicase with strand-annealing activity. Manon Batista, Petra Langendijk-Genevaux, Mirna Hajj, Duy Khanh Phung, Laura Plassart, Marie Bouvier, Gwennaële Fichant, and Béatrice Clouet-d'Orval.**

## 1. Activité ATPasique

Une cinétique de l'hydrolyse de l'ATP a été réalisée avec 250nM de *Pab*-ASH-Ski2 en présence ou non d'ARN totaux de *P. abyssi* à 65°C dans un mélange réactionnel contenant un ratio d'ATP froid et radiomarqué [ $\gamma$ 32P] et des ions Mg<sup>2+</sup>. La température de 65°C a été choisie sachant que *Pab*-ASH-Ski2 est une enzyme issue d'un organisme hyperthermophile qui a une



**Figure 26. A.** Activité ATPasique de *Pab*-ASH-Ski2 WT, ΔN et ΔC. Un mélange d'ATP froid/radiomarqué [ $\gamma^{32}$ ] est incubé à 65°C avec 200nM de protéine recombinante purifiée et des ARN totaux de *P. abyssi*. Des aliquots sont prélevés au cours du temps d'incubation. Les produits de réactions sont analysés par TLC (Panel du haut). Le pourcentage d'ATP hydrolysé est représenté en fonction du temps (Panel du bas). Chaque donnée  $\pm$  est une moyenne de trois expériences. Les expériences ont été réalisées avec (+ARN) ou sans (-ARN) ajout d'ARN totaux de *P. abyssi*. **B.** Activité ATPasique de *Pab*-ASH-Ski2 WT en présence d'ADN phiX174 et avec différentes concentration (0.25, 0.5 et 1μM) de *Pab*-aRNase J ou le mutant catalytique de *Pab*-aRNase J (aRNase J-H388A).



température optimale de croissance proche de 100°C. L'apparition des phosphates inorganiques (Pi) est suivie au cours du temps et quantifiée après séparation sur une chromatographie sur couche mince (TLC) dans un tampon  $\text{KH}_2\text{PO}_4$  (Figure 26A, panel de gauche). Toutes les expériences ont été réalisées en triplicata.

- ***Pab*-ASH-Ski2 WT**

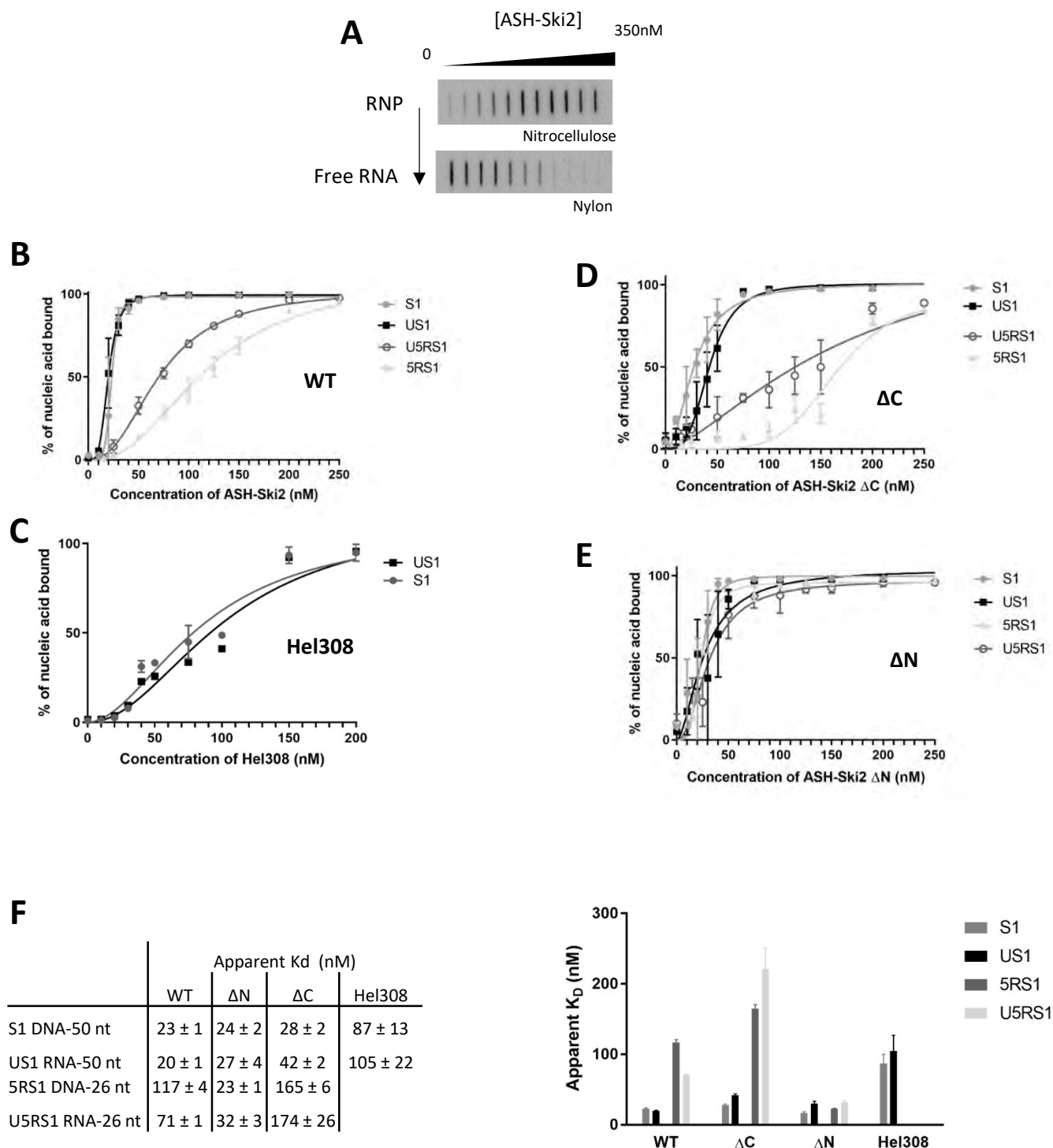
Les résultats montrent que *Pab*-ASH-Ski2 est capable d'hydrolyser l'ATP en présence d'ARN totaux ou d'ADN (Figure 26A et B). En absence d'ARN ou d'ADN, aucune activité n'est détectée. Ces résultats sont cohérents avec les propriétés générales des hélicases à ARN qui adoptent un état actif par un changement de conformation médiée par la liaison aux acides nucléiques (Jankowsky and Fairman 2007). Des résultats préliminaires obtenus lors de la thèse de D.K. Phung montrent que le mutant ponctuel *Pab*-ASH-Ski2-K245A (mutation dans le motif I connu pour être impliqué dans la liaison et l'hydrolyse de l'ATP) n'est pas capable d'hydrolyser l'ATP en présence ou en absence d'acides nucléiques (Phung D.K. 2017 Thesis). D'autres données de l'équipe montrent que  $(\text{His})_6$ -*Pab*-Hel308 a une activité ATPase sans spécificité vis-à-vis de l'ARN ou de l'ADN (Phung D.K. 2017 Thesis). De plus les données de la littérature pour les protéines de *M. thermotrophicus* et *S. tokodaii* montrent une activité d'hydrolyse de l'ATP stimulée en présence d'ADN simple brin pour *Mth*-Hel308 (Guy and Bolt 2005) et en présence d'ADN double brin pour *Sto*-Hel308 (Li et al. 2008).

L'ensemble de ces données semble suggérer que les hélicases de type Ski2-like chez les archées possède une activité ATPasique dépendante des acides nucléiques sans spécificité ni vis-à-vis de l'ARN ni de l'ADN.

- ***Pab*-ASH-Ski2  $\Delta\text{N}$  et  $\Delta\text{C}$**

Seule l'absence du domaine N-terminal ( $\Delta\text{N}$ ) affecte légèrement l'activité ATPasique de l'enzyme en présence d'ARN totaux de *P. abyssi* (Figure 26A). Dans le modèle structural (Figure 25B) les motifs Q et I, impliqués dans la liaison et l'hydrolyse de l'ATP (Voir Partie IV de l'introduction) (Figure 25A), semblent être à proximité du domaine accessoire en N-terminal. On peut proposer que le domaine N-terminal stabilise une configuration optimale du site actif.

En ce qui concerne  $\Delta\text{C}$ , la cinétique d'hydrolyse est similaire à celle de la protéine WT. Le domaine C-terminal contient les domaines WH et Ratchet (Figure 25A) qui selon la structure des hélicases de type Ski2 connues (Johnson and Jackson 2013), n'influenceraient pas la



**Figure 27.** Interaction des protéines *Pab*-ASH-Ski2 (WT, ΔN et ΔC) et *Pab*-Hel308 avec les substrats ARN et ADN présentés dans le Tableau 7 . **A-** Tests d'affinité par double filtration sur membrane de nitrocellulose DEAE et HydondN avec 5nM de substrats ARN\*(US1 et U5RS1) ou ADN\*(S1 et 5RS1) incubés avec des concentrations croissantes (0, 10, 20, 30, 40, 50, 75, 100, 150, et 200nM) de *Pab*-ASH-Ski2 WT, ΔN, ΔC ou *Pab*-Hel308 à 30°C pendant 15min. **B à E-** Le pourcentage de formation de complexe RNP à chaque concentration en protéine est reporté en fonction du temps pour calculer la constante d'affinité apparente ( $K_D$ ) pour chaque protéine. Les valeurs sont la moyenne de trois expériences indépendantes. **F-** Les valeurs obtenues sont reportées dans un tableau et comparées sur un graphe.

formation du site actif. De même, de manière cohérente avec ces résultats, l'activité ATPasique de Hel308 de *Archeoglobus fulgidus* est indépendante des domaines WH et Ratchet (Büttner et al. 2007).

## 2. Liaison aux acides nucléiques

La méthodologie adoptée pour mesurer l'affinité de *Pab*-ASH-Ski2 pour les acides nucléiques a été mise au point et initiée dans le cadre de la thèse de D.K Phung dans l'équipe. Pour cette étude, nous avons réitéré ces expériences en triplicata en utilisant des versions de protéines recombinantes sans étiquette.

Un panel de substrats listés dans le **tableau 7** se compose de différents types de molécules d'ADN ou d'ARN simple brins de 50 ou 26 nucléotides (nt) de long, nommés respectivement S1, US1, 5RS1 et U5RS1. Ces substrats se réfèrent à ceux utilisés dans l'étude des propriétés enzymatiques de *Mther*-Hel308 (Guy and Bolt 2005). Les substrats ADN S1, 5RS1 et ARN U5RS1 sont des oligonucléotides synthétiques et le substrat ARN US1 est une molécule obtenue par transcription *in vitro*.

Les substrats marqués à l'extrémité 5' par addition d'un phosphate radioactifs sont incubés avec une concentration croissante en protéine (0 à 350nM) en présence de 1mM d'ATP et 5mM de MgCl<sub>2</sub> pendant 15min à 30°C (des résultats similaires ont été obtenus à 65°C). Les complexes ribonucléoprotéiques (RNP) formés sont séparés des molécules d'ADN et ARN libres par un système de double filtration en utilisant des membranes aux propriétés différentes. Il faut noter que le système de double filtration (Wong and Lohman 1993) a été choisi suite à des essais infructueux de migration sur gel acrylamide natif à 5% où les complexes RNP restent bloqués dans les puits du gel (Phung D.K. 2017 Thesis). Dans le système de double filtration, les complexes RNP (fractions liées) sont retenus sur la membrane de nitrocellulose/ Hybond N+ alors que les molécules d'ADN et d'ARN libres sont retenues sur la membrane de nylon chargées positivement (**Figure 27A**). Le pourcentage de fractions de molécules substrats liées versus libres est quantifié à l'aide du logiciel Multigauge pour chaque concentration en protéine. Les valeurs des constantes apparentes d'affinité reportées dans le tableau de la **Figure 27F** sont calculées à l'aide du logiciel Prism et sont générées à partir d'expériences réalisées en triplicata. On observe, dans tous les cas de figures, des courbes de liaison aux acides nucléiques de type sigmoïdale avec un coefficient de Hill supérieur à 1. Ceci

reflète soit une liaison coopérative de la protéine à son substrat avec au moins deux sites de liaison sur la molécule de substrat et/ou un changement conformation de la protéine lors de la fixation du substrat (**Figure 27**). La coopérativité semble être indépendante de la présence des domaines accessoires N et C-terminal, montrant une propriété intrinsèque du domaine cœur SF2 (RecA1/RecA2) qui porte les sites principaux de liaison aux acides nucléiques (**Figure 25A**) (Fairman-Williams et al. 2010).

- ***Pab-ASH-Ski2* WT et *Pab-Hel308***

*Pab-ASH-Ski2* a une affinité identique pour les substrats de 50 nt que ce soit de l'ADN (S1) ou de l'ARN (US1) avec une constante autour de 20nM (**Figure 27B & 27F**). Pour les substrats courts de 20 nt de long, une différence notable est observée avec une plus grande affinité pour l'ARN (U5RS1) que pour l'ADN (5RS1). Dans l'ensemble, *Pab-Hel308* semble avoir une affinité plus faible pour les acides nucléiques que *Pab-ASH-Ski2* sans spécificité de liaison vis à vis de la nature des substrats ADN ou ARN (**Figure 27C**). Il faut noter que les courbes obtenues pour *Pab-Hel308* sont préliminaires, avec une seul réplikat pour le substrat US1 et deux pour le substrat S1.

- ***Pab-ASH-Ski2*  $\Delta$ N et  $\Delta$ C**

En ce qui concerne *Pab-ASH-Ski2- $\Delta$ C*, les constantes apparentes d'affinité reflètent une affinité deux à trois fois plus faible que pour *Pab-ASH-Ski2-WT* (**Figure 27D & 27F**). On peut proposer que le domaine C-terminal, contenant les motifs WH et Ratchet typiques de la famille des hélicases de type Ski2, participe à la stabilisation de la conformation active du domaine cœur SF2 pour la liaison aux acides nucléiques (RecA1/RecA2) (**Figure 25B**).

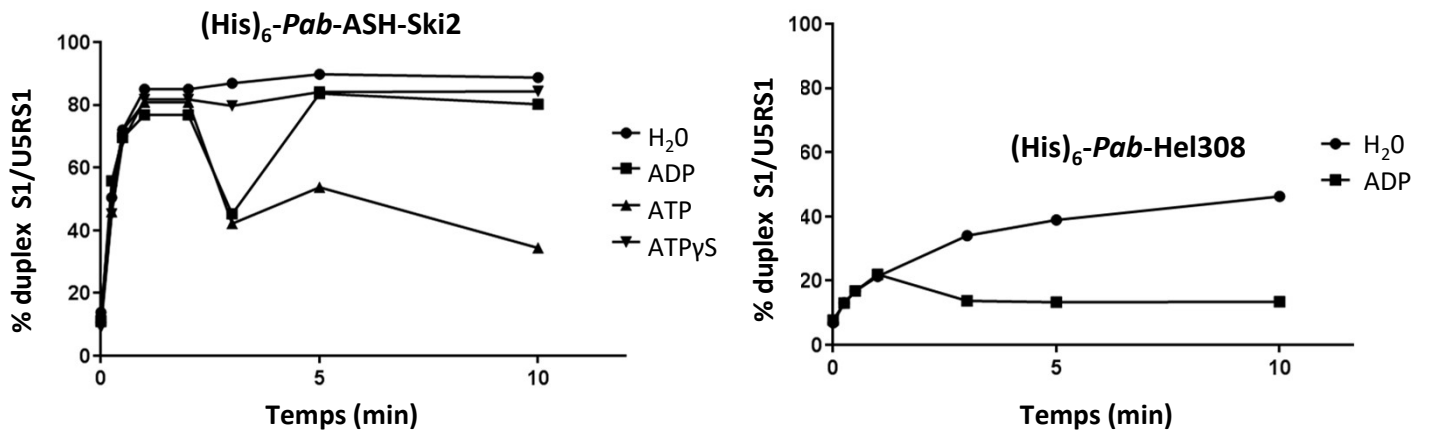
En absence du domaine N-terminal (*Pab-ASH-Ski2- $\Delta$ N*), on observe des constantes apparentes d'affinité similaires pour les substrats de 50 nt et de 26 nt (**Figure 27E & 27F**). D'après ces données, le domaine N-terminal ne posséderait donc pas de site de fixation propre aux acides nucléiques puisque les affinités sont semblables à celles observés pour *Pab-ASH-Ski2-WT*. Cependant la différence d'affinité selon la longueur du substrat observée pour *Pab-ASH-Ski2-WT* est atténuée dans cette configuration de la protéine, ce qui semble indiquer une importance relative du domaine N-terminal dans la reconnaissance des substrats.

### 3. Activité hélicase

Les résultats préliminaires obtenus par D.K. Phung (Phung D.K. 2017 Thesis) montrés sur la **Figure 28** dans l'équipe ont servi de base pour la mise au point des réactions de déroulement et de formation de brins décrite dans cette étude. Sur la **Figure 28**, on observe une cinétique rapide d'appariement de brins ADN/ARN de la protéine recombinante (His)<sub>6</sub>-*Pab*-ASH-Ski2 qui est indépendante de la présence d'ATP. L'ajout d'ATP au temps 2 minute engendre une diminution du pourcentage de duplex formé. Ceci n'est pas observé en présence d'ATPγS (ATP non hydrolysable). En présence d'ADP, la réaction de déroulement de brin n'est que transitoire. On propose que (His)<sub>6</sub>-*Pab*-ASH-Ski2 adopte une conformation dite « de déroulement » lors de la liaison à l'ATP ou à l'ADP. Cette conformation est maintenue tant qu'il y a de l'ATP à hydrolyser. Pour (His)<sub>6</sub>-*Pab*-Hel308, on observe une activité d'appariement beaucoup plus faible. En présence d'ADP, la protéine semble être figée dans une conformation favorisant la dissociation de brins. L'ensemble de ces données nous permet de proposer que la protéine ASH-Ski2 peut adopter une conformation dite « d'appariement » et une conformation dite de « déroulement ou dissociation de brins » en présence d'ATP. Le changement de conformation nécessite l'hydrolyse de l'ATP et le relargage du produit ADP. Nous avons choisi d'analyser indépendamment ces deux activités correspondantes à différents états ou conformations de la protéine selon la présence ou l'absence d'ATP et ceci avec des protéines recombinantes sans étiquette pour s'affranchir de l'effet non spécifique que pourrait générer la présence d'étiquette en N- ou C-terminal. Les substrats choisis pour cette étude sont les substrats décrits dans le **Tableau 7** qui résume les différents types de duplex qui peuvent se former avec des extrémités 3' ou 5' sortantes selon la nature des oligonucléotides de 26 nt utilisés.

- **Activité d'appariement de brins (dite « annealing »)**

Les cinétiques de la formation de duplex ARN/ARN (US1/U5RS1\*), ADN/ARN (S1/U5RS1\*) et ADN/ADN (S1/5RS1\*) ont été effectuées, en triplicata, en présence de 250nM de *Pab*-ASH-Ski2 ou *Pab*-Hel308 à 65°C en absence d'ATP. Les produits de la réaction sont visualisés après migration sur un gel acrylamide natif 8% (**Figure 29A et B**, panel du haut). Le pourcentage de duplex formé est quantifié au cours du temps (**Figure 29A et B**, panel du bas).



**Figure 28.** Données préliminaires (figure 39 A et B de la thèse de DK Phung (2017)) correspondants à l'activité d'appariement et de déroulement de brins des protéines recombinantes (His)<sub>6</sub>-Pab-ASH-Ski2 et (His)<sub>6</sub>-Pab-Hel308. 5 nM de substrat ADN S1 et de substrat ARN U5RS1 sont mélangés en présence de 250 nM de protéines à 65°C. L'apparition de la formation de l'hétéroduplex avec une extrémité 3' sortante visualisée sur un gel acrylamide natif 8% est quantifiée en fonction du temps. Dans ces expériences 5mM d'ATP , ADP ou ATPγS sont ajoutés 2 minutes après le début de la réaction. Ces expériences préliminaires ont servi de base pour la mise au point des cinétiques de dissociation et d'appariement de brins.



### **Pab-ASH-Ski2 WT et Pab-Hel308**

En présence de *Pab-ASH-Ski2* WT, on observe une cinétique très rapide d'appariement de brin du duplex US1/U5RS1\* jusqu'à plus de 80% de U5RS1\* hybridé à US1 après 5 minutes d'incubation. La cinétique de formation de duplex est sensiblement réduite en présence d'ADN S1 avec moins de 40% d'hétéroduplex S1/U5RS1 observé. De plus, la formation de homoduplex S1/5RS1 avec les deux brins ADN est très faible, semblable à l'expérience témoin sans protéine (**Figure 29A**). La protéine *Pab-ASH-Ski2* WT présente donc une préférence marquée pour les brins ARN en stabilisant la formation d'homoduplex ARN/ARN dans ces conditions expérimentales.

A l'opposé des résultats observés pour *Pab-ASH-Ski2*, *Pab-Hel308* a une capacité d'appariement de brin très faible proche de l'expérience témoin dans ces conditions expérimentales quel que soit les substrats (**Figure 29B**).

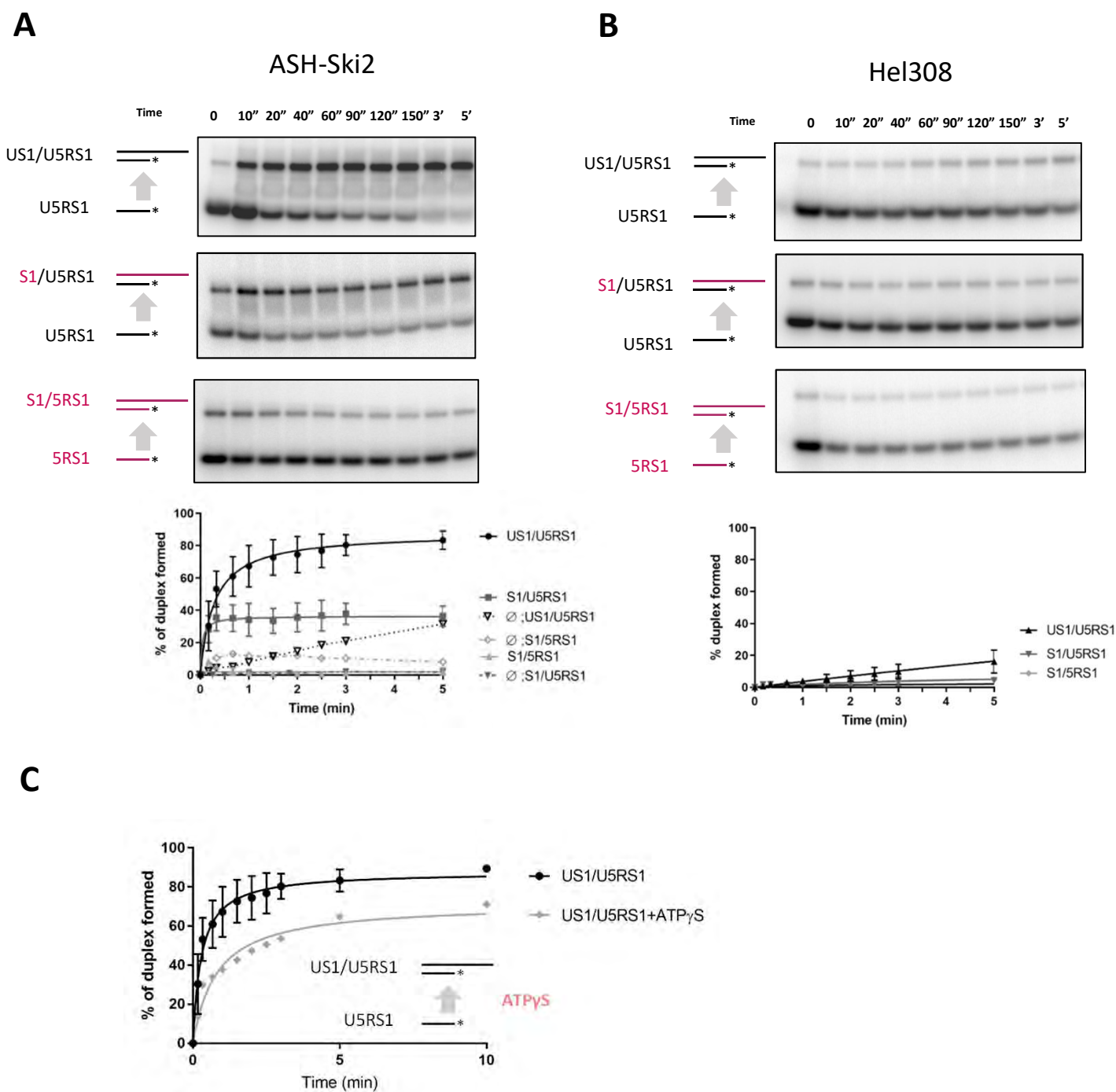
Nous avons également regardé l'effet de la présence d'analogues d'ATP dans le milieu réactionnel pour *Pab-ASH-Ski2*. La présence d'ATPyS non hydrolysable affecte sensiblement la cinétique d'appariement de brins. L'ATPyS se fixe au site catalytique sans être hydrolysé et pourrait engendrer un changement de conformation de la protéine ne favorisant plus l'appariement de brins. (**Figure 29C**).

### **Pab-ASH-Ski2 $\Delta$ N et $\Delta$ C**

La cinétique d'appariement de brin avec les substrats ARN US1/U5RS1\* est sensiblement plus rapide pour *Pab-ASH-Ski2 $\Delta$ C* et *Pab-ASH-Ski2 $\Delta$ N* comparée avec celle de *Pab-ASH-Ski2* WT (**Figure 30A**). Il semblerait que les domaines N- et C-terminal soient importants pour la spécificité de substrats. Les cinétiques d'appariement d'hétéroduplex (**Figure 30B**) ou d'homoduplex ADN (**Figure 30C**) observées pour *Pab-ASH-Ski2 $\Delta$ C* et *Pab-ASH-Ski2 $\Delta$ N* semblent rapides mais un plateau est atteint rapidement avec un pourcentage de duplex formé très inférieur à celui observé dans le cas d'un homoduplex ARN.

- **Activité de déroulement de duplex ou dissociation de brins (dite « unwinding »)**

Nous avons testé la capacité de *Pab-ASH-Ski2* et de *Pab-Hel308* à dérouler (activité unwinding) le duplex d'ARN US1/U3RS1 avec une extrémité 5' sortante ou le duplex d'ARN US1/U5RS1 avec une extrémité 3' sortante. Cette capacité est observée en présence d'ATP (**Figure 31A**). Les duplex sont formés au préalable *in vitro* par hybridation en incubant de l'ARN simple brin US1 (50 nt) et de l'ARN simple brin U3RS1 ou U5RS1 (26 nt) complémentaires respectivement



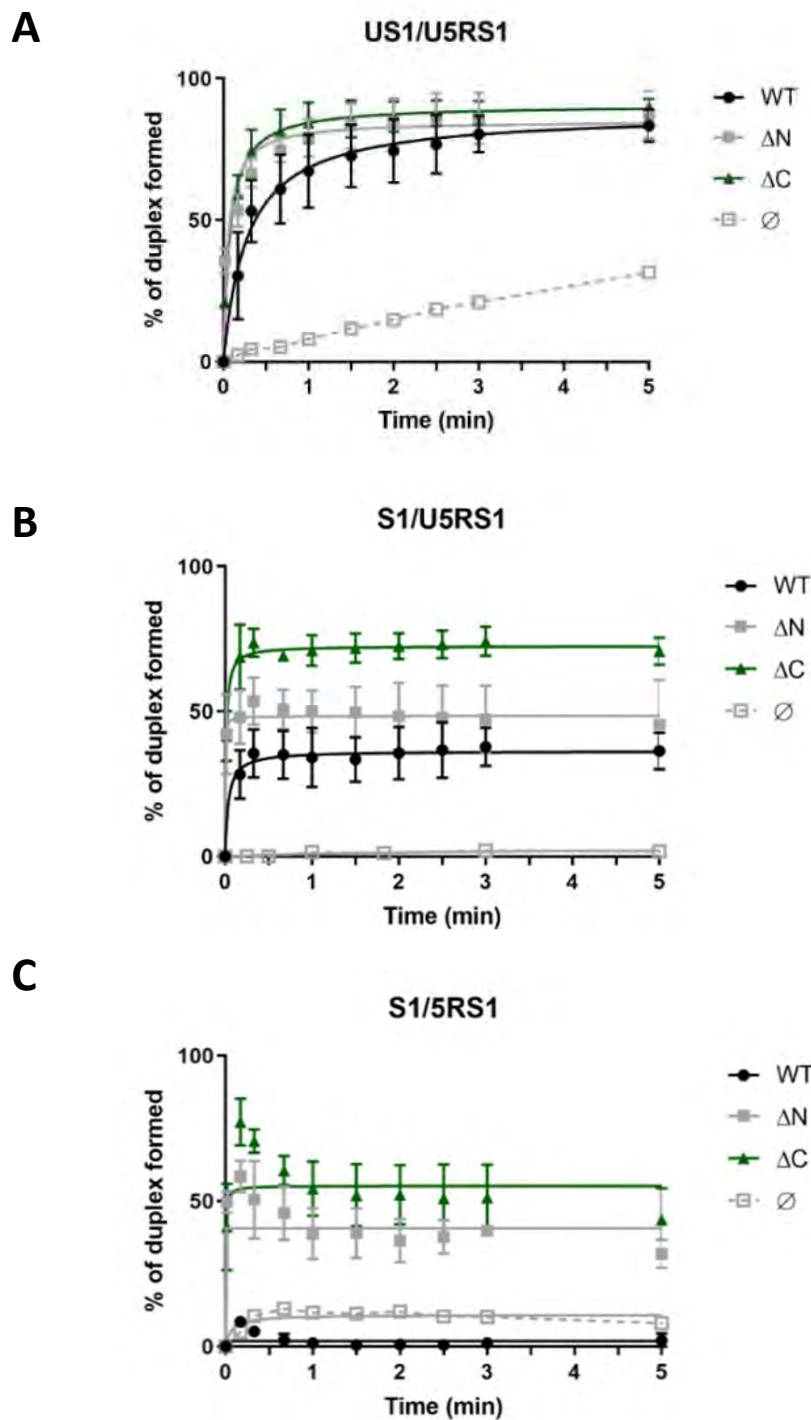
**Figure 29.** Cinétiques d'appariement de brins de *Pab*-ASH-Ski2 WT et *Pab*-Hel308 La formation de duplex ARN/ARN (US1/U5RS1), ADN/ARN (S1/U5RS1) ou ADN/ADN (S1/5RS1) (Tableau 7) en absence (∅) ou en présence de 250nM de protéine est suivie à 65°C. Les molécules d'ARN sont représentées en noir et d'ADN en violet dans les schéma de droite. La réaction a été analysée après migration en gel natif acrylamide à 8% après traitement à la protéinase K (Panel du haut). Les graphes montrent le pourcentage de duplex formé en fonction du temps. Des triplicatas ont été réalisés pour chaque condition expérimentale. **A.** Cinétiques obtenues en présence de *Pab*-ASH-Ski2 WT **B.** Cinétiques obtenues en présence de *Pab*-Hel308. **C.** Cinétique d'appariement de US1/U5RS1 en présence de *Pab*-ASH-Ski2 WT et d'analogue d'ATP non hydrolysable ATPγS (en rose).

des extrémités 3' et 5' de US1 à 90°C pendant 5min dans un tampon 150mM NaCl et 15mM NaCl suivi d'une étape de refroidissement lente à température ambiante. Pour visualiser la dissociation des brins appariés au cours du temps, l'ARN U3RS1 ou U5RS1 est préalablement marqué en 5' par une réaction de phosphorylation en présence d'ATP [ $\gamma P^{32}$ ]. Une cinétique de dissociation de brin est suivie entre 0 et 40 min après incubation de 5 nM de duplex pré-formés (US1/U3RS1\* ou US1/U5RS1\*) à 65°C en présence de 250nM de protéine, d'ATP (5mM) et de 200 nM oligo Trap non marqué (U3RS1 ou U5RS1). L'excès d'oligo Trap capture l'oligo US1 au fur et à mesure du déroulement du duplex (US1/U3RS1\*) ou (US1/U5RS1\*) pour empêcher la réaction réverse d'appariement de s'effectuer au cours du temps. Les produits de la réaction sont séparés par électrophorèse sur un gel acrylamide natif 8% et l'apparition de molécules d'ARN simple brin U3RS1\* ou U5RS1\* est quantifiée à l'aide du logiciel MultiGauge au cours du temps pour obtenir le pourcentage de duplex déroulés. Un témoin sans ajout de protéine permet de suivre la stabilité intrinsèque des duplex au cours du temps de la réaction. Chaque cinétique est effectuée en triplicata. Ces expériences montrées ci-dessous ont été principalement menées avec des substrats ARN.

#### **Pab-ASH-Ski2 WT et Pab-Hel308**

*Pab-ASH-Ski2* est capable de dérouler 75 % de duplex US1/U5RS1\* après 40 min d'incubation contre seulement 35% avec le duplex US1/U3RS1\* (**Figure 31A**, panel du haut). Ces données suggèrent que l'hélicase *Pab-ASH-Ski2* possède une polarité de déroulement de duplex de l'extrémité 3' vers l'extrémité 5', conformément à la polarité décrite pour les hélicases de la famille Ski2 (Johnson and Jackson 2013). Contrairement à *Pab-ASH-Ski2*, *Pab-Hel308* ne semble pas avoir la capacité de dérouler le duplex composé de molécules d'ARN US1/U5RS1\* dans ces conditions (**Figure 31A**, panel du bas). Ce résultat semble indiquer une différence de reconnaissance de substrats entre les deux hélicases de type Ski2 de *P. abyssi*. De manière cohérente, dans la littérature *Mth-Hel308* est décrite comme une hélicase à ADN avec une capacité à dérouler des duplex d'ADN (Guy and Bolt 2005).

La comparaison des cinétiques de déroulement du duplex US1/U5RS1\* en présence et absence d'ATP et ou de MgCl<sub>2</sub> indique une cinétique optimale en présence d'ATP et ions Mg<sup>2+</sup>. L'ajout d'analogue de l'ATP tel que ADP ou ATP $\gamma$ S non hydrolysable a un effet significatif avec une diminution de l'activité de dissociation de brins (**Figure 31B**). La présence d'ions Mg<sup>2+</sup> est également importante pour une activité optimale de dissociation de brins.



**Figure 30.** Comparaison des cinétiques d'appariement de brins de *Pab-ASH-Ski2* WT, *Pab-ASH-Ski2*ΔN et *Pab-ASH-Ski2*ΔC dans les mêmes conditions que celles décrites dans la Figure 29A. **A.** Cinétique de formation de duplex US1/U5RS1 **B.** Cinétique de formation de duplex S1/U5RS1 **C.** Cinétique de formation de duplex S1/5RS1.

### **Pab-ASH-Ski2 ΔN et ΔC**

Nous avons ensuite comparé la cinétique de déroulement du duplex US1/U5RS1\* avec une extrémité 3' sortante de *Pab-ASH-Ski2* WT avec celle de *Pab-ASH-Ski2ΔN* et *Pab-ASH-Ski2ΔC*. Le domaine N-terminal n'affecte pas la cinétique de dissociation des brins (**Figure 32**). Cependant on observe une cinétique plus lente en absence du domaine C-terminal. Il est possible que la conformation du domaine cœur RecA1/RecA2 requise pour dissocier les brins soit stabilisée par la présence du domaine C-terminal. Ces résultats sont cohérents avec l'activité de Hel308 d'*Archeoglobus fulgidus* qui montre également que le domaine C-terminal est requis pour une activité optimale de déroulement de duplex d'ADN (Büttner et al. 2007).

## **4. Conclusions**

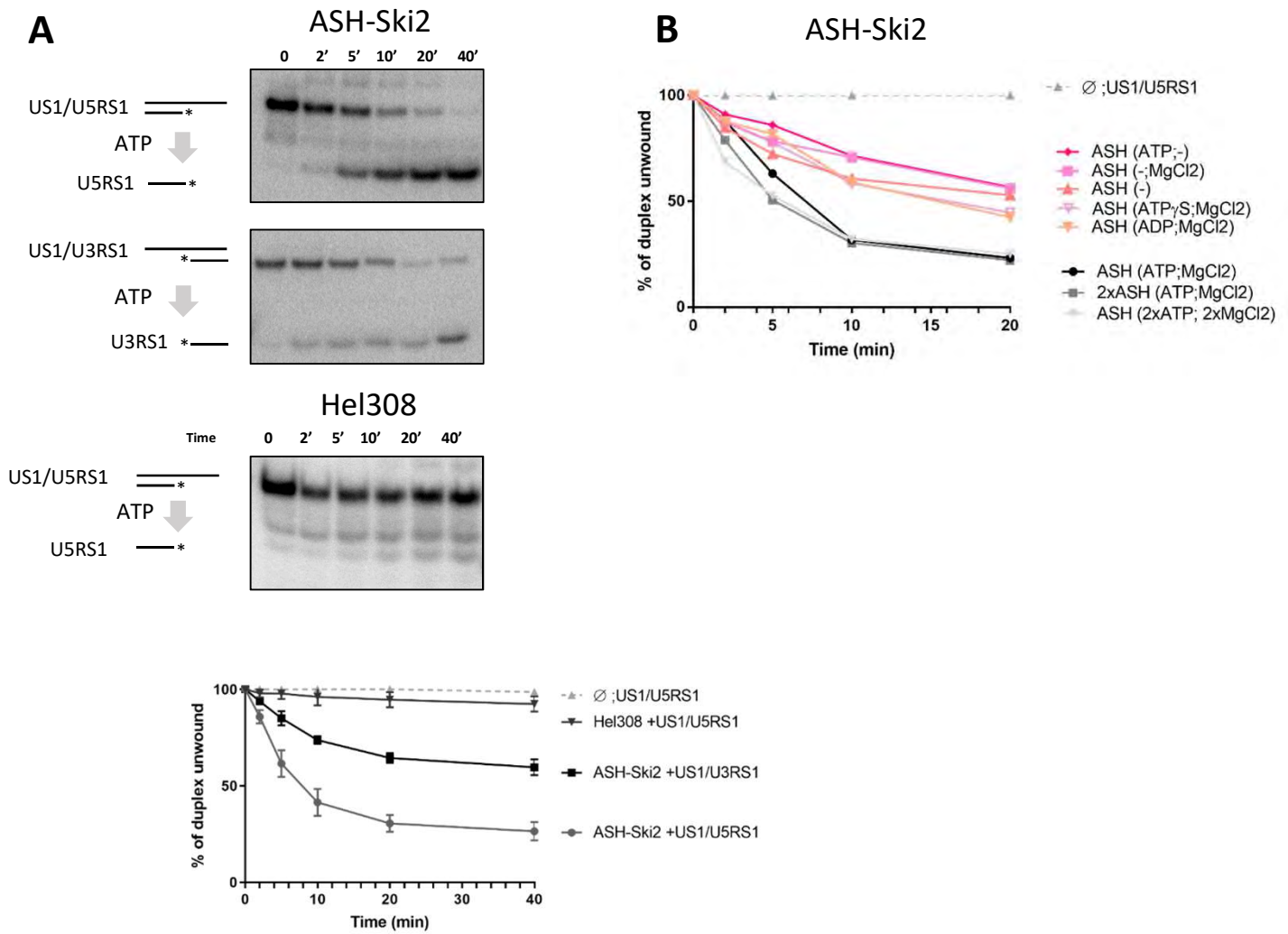
- ***Pab-ASH-Ski2* possède une activité hélicase à ARN avec une polarité de 3' vers 5'**

L'étude des différentes propriétés permet de proposer que *Pab-ASH-Ski2* possède une activité hélicase à ARN. En effet, l'activité d'appariement de brins observée dans cette étude apparait bien plus efficace en présence de substrats ARN que de substrat ADN bien que l'activité ATPasique est sans spécificité vis à vis de la nature des acides nucléiques, ARN ou ADN. Il est intéressant de noter que l'activité d'appariement ne nécessite pas la présence d'ATP, ainsi la spécificité vis-à-vis de l'ARN semble être indépendante de l'activité ATPasique portée par le cœur SF2. Des données préliminaires ont montré que le mutant ponctuel de l'activité ATPasique, *Pab-ASH-Ski2-K245A* possède une activité d'appariement réduite à 30% comparée à celle de la protéine WT (Phung D.K. 2017 Thesis). Ces données doivent être réitérées dans ces conditions d'expériences et le mutant sera testé pour l'activité d'appariement et de déroulement des substrats.

On peut proposer que *Pab-ASH-Ski2* peut adopter deux conformations l'une favorable à l'appariement de brin en absence d'ATP et l'autre favorable à la dissociation de brins nécessitant la fixation d'ATP et d'ions Mg<sup>2+</sup>. De plus, la dissociation de brins est favorisée avec un duplex possédant une extrémité 3' sortante montrant une polarité de 3' vers 5'.

- **Le domaine C-terminal module sensiblement l'activité hélicase de *Pab-ASH-Ski2***

L'ensemble de ces résultats montre que les domaines accessoires *Pab-ASH-Ski2* n'affectent pas l'activité d'hydrolyse de l'ATP qui est portée par le domaine cœur SF2 constitué des domaines RecA1 et RecA2 et qu'ils n'ont qu'un effet modéré sur la capacité de liaison aux



**Figure 31.** Cinétiques de dissociation de brins de *Pab*-ASH-SKi2 WT et *Pab*-Hel308. **A.** Le duplexe substrat avec une extrémité 3' sortante (US1/U5RS1) ou 5' sortante (US1/U3RS1) formé un ARN U5RS1 ou U3RS1 radiomarqué  $\gamma^{32}$  (Tableau 7) est incubé 40' à 65°C avec 5 mM d'ATP sans protéine ( $\emptyset$ ) ou en présence de 250nM de protéine à 65°C. La réaction a été analysée après migration dans un gel natif acrylamide 8% (Panel du haut). Les graphes montrent le pourcentage de duplex déroulé en fonction du temps (Panel du bas). Chaque donnée  $\pm$  correspond à une moyenne de trois expériences. L'activité de déroulement du duplex US1/U5RS1 de *Pab*-ASH-SKi2 WT est comparée avec celle de *Pab*-Hel308. **B.** L'activité de déroulement du duplex US1/U5RS1 de *Pab*-ASH-SKi2 WT en absence ou en présence d'ATP ou de MgCl<sub>2</sub> et en présence d'ADP et de ATP $\gamma$ S.



acides nucléiques. Les cinétiques d'appariement de brins en absence du domaine C-terminal semble indiquer une accélération de l'association de brin par rapport à la version WT. Egalement la spécificité vis-à-vis de l'appariement de brins ARN/ARN semble moins stricte en absence du domaine C-terminal (**Figure 32**). De plus on observe une réduction de la cinétique de déroulement de duplex en absence du domaine C-terminal (**Figure 32**). L'absence du domaine C-terminal pourrait favoriser la formation d'une conformation de la protéine favorable à l'appariement de brin. Le domaine C-terminal, typique de la famille Ski2-like, serait important à la fois dans la spécificité de substrat et dans l'inflexion de l'activité de dissociation de brins.

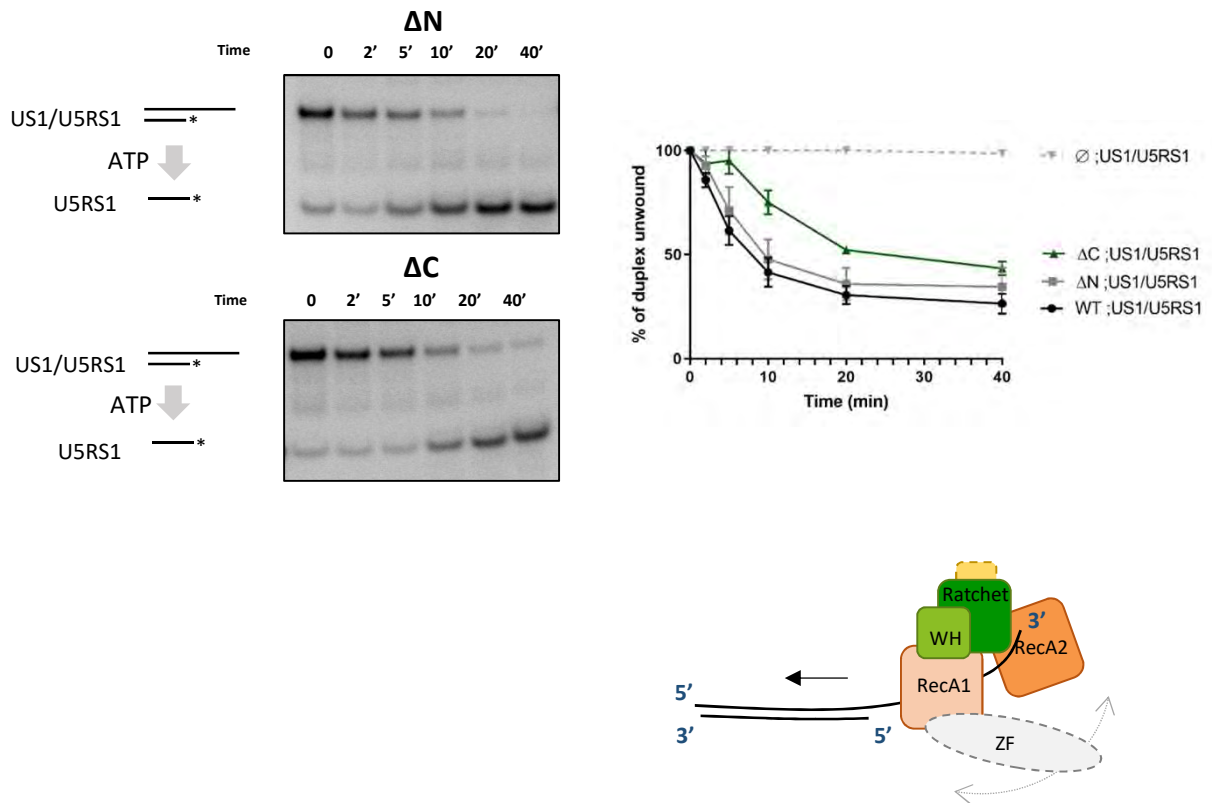
- ***Pab*-aRNase J module l'activité ATPasique de *Pab*-ASH-Ski2**

Des expériences préliminaires de l'activité d'hydrolyse de l'ATP de *Pab*-ASH-Ski2 ont été réalisées en présence d'ADN et de différentes concentrations en protéine *Pab*-aRNase J et son mutant catalytique *Pab*-aRNaseJ-H388A. Ces premiers résultats montrent que la présence de *Pab*-aRNase J ou *Pab*-aRNase J-H388A entraîne une réduction de l'activité ATPasique de *Pab*-ASH-Ski2 (**Figure 26B**). Ces données sont à réitérées en triplicata et à tester en présence d'ARN.

## V. Etudes structurales d'ASH-Ski2 et aRNase J

### 1. Purification de *Pab*-aRNase J et *Pab*-ASH-Ski2

La structure de l'hélicase ASH-Ski2 n'est pas connue actuellement mais celle de l'aRNase J, avec ou sans substrat ARN (PDB :5HAB et 5HAA), a récemment été résolue par cristallographie chez *Méthanolobus psychrophilus* (*Mpsy*-aRNase J) (Zheng et al. 2017). Les structures de *Mpsy*-aRNase J révèlent un réseau de liaisons hydrogène, qui permet le transfert de protons dans le centre catalytique et aide à la catalyse. Une poche de liaison au monophosphate coordonne l'extrémité 5' de l'ARN et explique la préférence pour les substrats 5' monophosphate de l'activité exoribonucléolytique de l'aRNase J. *Mpsy*-aRNase J est sous forme homotétramérique et les motifs spécifiques Loop 1 et Loop 2, préalablement identifiés par mon équipe (Clouet-d'Orval et al. 2010), semblent impliqués dans la liaison aux acides nucléiques. Nous avons initié une étude structurale en cristallographie des protéines ASH-Ski2 et aRNase J de *P. abyssi* en collaboration avec V. Guillet (IPBS, Toulouse), dans le but de résoudre la structure de *Pab*-ASH-Ski2 et de caractériser l'interaction entre *Pab*-ASH-Ski2 et

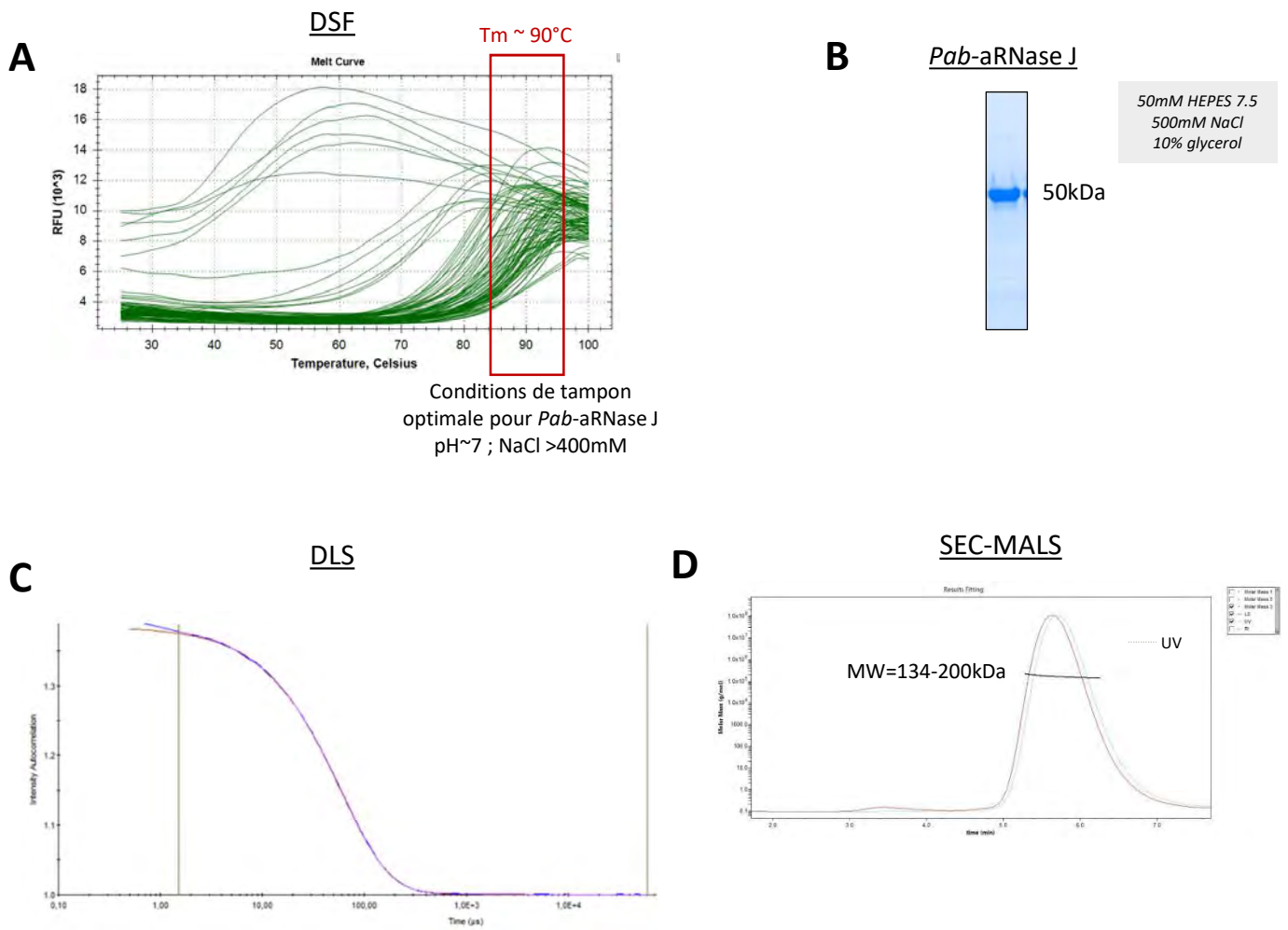


**Figure 32.** Cinétique de dissociation de brin en présence d'ATP des variants *Pab-ASH-Ski2ΔN* et *Pab-ASH-Ski2ΔC* sont comparés à celle de *Pab-ASH-Ski2 WT*. La réaction observé est illustrée par un schéma de *Pab-ASH-Ski2 WT* sur le substrat . La polarité 3' vers 5' est indiquée.

*Pab*-aRNase J. Pour se faire, les protéines *Pab*-aRNase J, *Pab*-ASH-Ski2 et ses variants *Pab*-ASH-Ski2ΔC et *Pab*-ASH-Ski2-DomN ont été purifiées à homogénéité et des cribles de cristallisation ont été réalisées. Il faut noter que les limites de température des appareillages disponibles actuellement est un frein dans les mesures biophysiques des protéines hyperthermophiles.

- **Purification de *Pab*-aRNase J**

La protéine recombinante *Pab*-aRNase J a été purifiée sans étiquette après avoir été surexprimée à partir d'un vecteur d'expression pet11b dans la souche *E. coli*. Les culots bactériens exprimant la protéine d'intérêt ont été lysés et ont subi une étape de chauffage à 70°C pour éliminer les protéines thermosensibles. Les fractions contenant *Pab*-aRNase J pure, obtenues après le passage sur colonne HiTrap Heparin à l'aide d'un système FPLC (Fast Protein Liquid Chromatography), ont été rassemblées puis dialysées dans un tampon déterminé au préalable par fluorimétrie différentielle à balayage (DSF, pour Differential Scanning Fluorimetry). La DSF est une méthode permettant de caractériser la stabilité thermique des protéines et déterminer les conditions dans lesquelles les protéines sont les plus stables. Le fluorophore utilisé est le SYPRO® Orange dont la fluorescence augmente dans un environnement hydrophobe. Lorsque les protéines se dénaturent, les surfaces hydrophobes sont exposées ce qui déclenche une augmentation de la fluorescence. Un crible préliminaire, faisant varier les concentrations en protéine et en fluorophore, a été réalisé afin de déterminer les conditions idéales en termes de rapport signal/bruit à utiliser pour les cribles commerciaux qui seront testés. Les transitions thermiques sont monitorées avec un système CFX96 (Real-time PCR, Bio-Rad) et les mélanges ont été soumis à un gradient de température de 20 à 100°C. Les courbes de dénaturation des protéines ont été obtenues sur le logiciel CFX Manager (v 3.1) (non montrées ici). Le T<sub>m</sub> d'aRNase J observé est de 90°C dans les conditions de 400mM NaCl, 10μM de protéine et 10X de fluorophore. Un crible commercial (Fundament, Jena Bioscience) permettant de tester une gamme large de pH et concentrations salines a ensuite été réalisé pour *Pab*-aRNase J dans les conditions déterminées ci-dessus. La plaque a ensuite subi le même gradient de température et les courbes de dénaturation sont obtenues de la même façon que ci-dessus (**Figure 33A**). Les résultats ont permis de déterminer des conditions optimales pour la stabilité d'aRNase J, à savoir : un pH autour de 7 et une concentration en NaCl supérieure à 400mM. Les fractions contenant la protéine pure *Pab*-aRNase J ont donc été dialysées dans un tampon HEPES 50mM pH 7, 500mM de NaCl et 10%



**Figure 33. Purification à Homogénéité de *Pab*-aRNase J.** **A.** Mesures des  $T_m$  de la protéine *Pab*-aRNase J dans différentes conditions de tampon (crible commercial Fundament, Jena Bioscience) en Fluorimétrie différentielle à balayage (DSF). **B.** Un échantillon de la protéine purifiée à homogénéité de *Pab*-aRNase J est déposé sur un Gel SDS PAGE (4-15%) coloré au bleu de Coomassie. La taille théorique et le tampon de conservation de *Pab*-aRNase J sont indiqués. **C.** Courbe d'autocorrélation de *Pab*-aRNase J mesurée par diffusion dynamique de la lumière (DLS). **D.** Chromatogramme obtenu en SEC-MALS de *Pab*-aRNase J (Colonne S200 I 5/150).

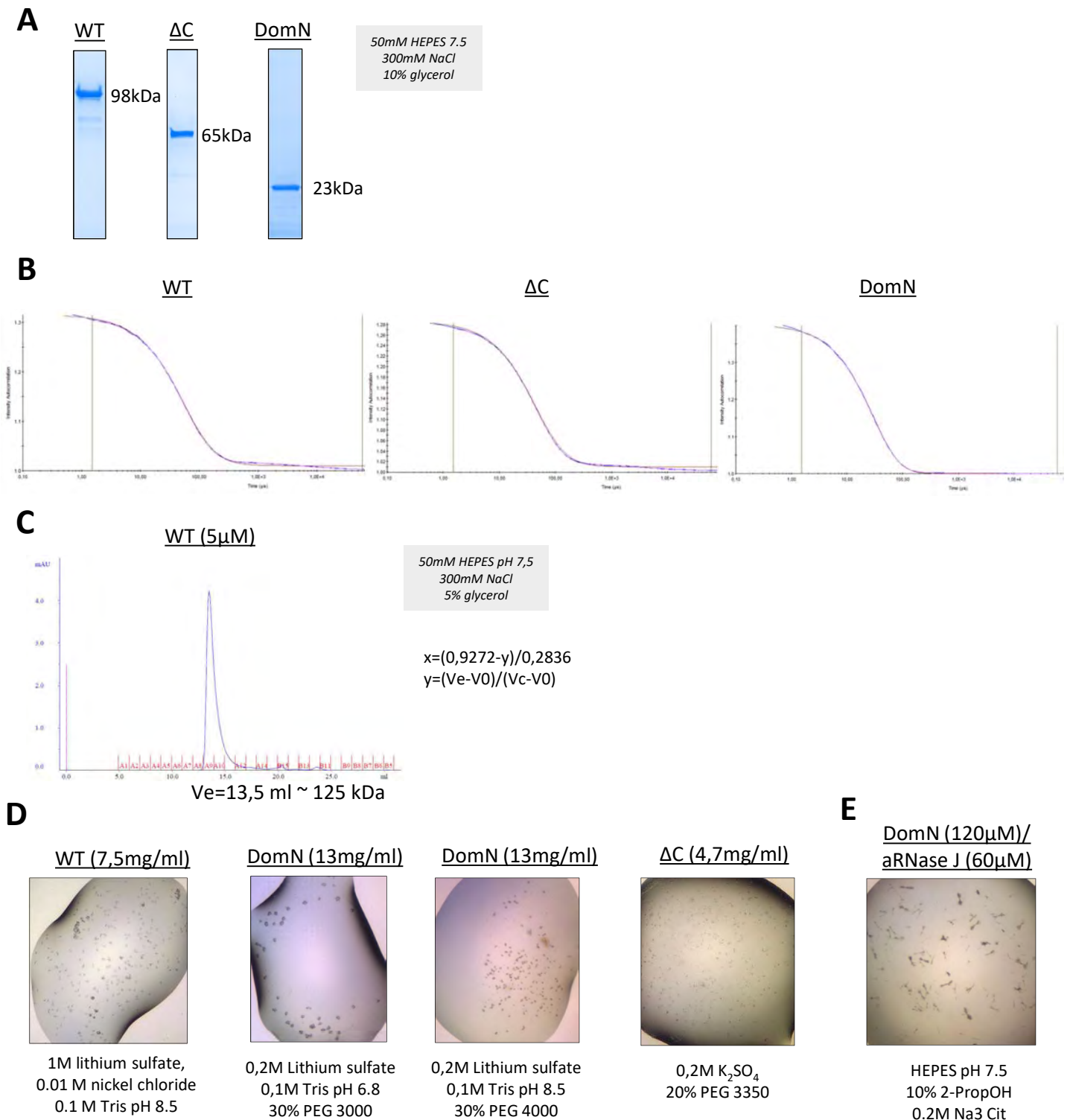
de glycérol et ont été concentrées par centrifugation différentielle sur membrane semi perméable (PES : polyethersulfone) à l'aide de système vivaspin20 (seuil de coupure (MWCO) = 30 000 Da). La concentration a été suivie par analyse de l'absorbance à 280nm et le rapport DO260/DO280 a été vérifié à 0.6 (absence d'acides nucléiques). La pureté de *Pab*-aRNase J a été analysée par SDS-PAGE après coloration au bleu de Coomassie (**Figure 33B**).

- **Purification de *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2-DomN et *Pab*-ASH-Ski2ΔC**

Les protéines recombinantes *Pab*-ASH-Ski2-DomN, *Pab*-ASH-Ski2ΔC et *Pab*-ASH-Ski2 WT ont été exprimées dans un vecteur d'expression pET11b chez *E. coli* puis purifiées séquentiellement sur une colonne HiTrap Heparin, une colonne échangeuse de cations (SP HP) et sur une colonne d'exclusion de taille (Hiload S200 16/60 PG) dans un tampon HEPES 50mM pH 7.5, NaCl 300mM et 10% glycérol. Les fractions contenant les protéines pures ont été rassemblées et concentrées par centrifugation différentielle sur membrane semi perméable à l'aide de système vivaspin20 (MWCO =30 000 ou 50 000Da). La concentration a été suivie par analyse de l'absorbance à 280nm et le rapport DO260/DO280 a été vérifié à 0.6 (absence d'acides nucléiques). La pureté des protéines a été analysée par SDS-PAGE après coloration au bleu de Coomassie (**Figure 34A**). Des mesures de thermostabilité par DSF ont été réalisées pour la protéine *Pab*-ASH-Ski2 cependant, le Tm n'a pu être obtenu, étant au-dessus de la limite de température du système (100°C).

## **2. Homogénéité des échantillons protéiques *Pab*-aRNase J et *Pab*-ASH-Ski2**

L'homogénéité et la qualité des protéines recombinantes purifiées ont été vérifiées en Diffusion dynamique de la lumière en solution (DLS). La DLS est une technique qui exploite le mouvement brownien des molécules en solution et permet d'évaluer le rayon hydrodynamique (Rh) de la protéine ainsi que la polydispersité de la solution (Pd). Ces informations permettent notamment de vérifier l'absence d'agrégats, ce qui est indispensable dans l'optique d'effectuer des essais de cristallisation. Nous avons testé les protéines *Pab*-aRNase J, *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2-DomN et *Pab*-ASH-Ski2ΔC à des concentrations respectives de 1.2, 4.6, 8.5, et 9.8mg/ml. Les fonctions d'autocorrélation, calculées à partir de la mesure des fluctuations d'intensité diffusée par la solution protéique en fonction du temps, permettent d'évaluer la distribution d'intensité en fonction du Rh, la polydispersité et l'éventuelle présence d'agrégat. Les fonctions d'autocorrélation expérimentales des protéines



**Figure 34. Purification et Essais de cristallisation de *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2 DomN et *Pab*-ASH-Ski2 $\Delta$ C.** **A.** Un échantillon de protéine purifiée à homogénéité de *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2-DomN, *Pab*-ASH-Ski2 $\Delta$ N et *Pab*-ASH-Ski2 $\Delta$ C sont déposés sur un Gel SDS PAGE (4-15%) colorés au bleu de Coomassie. La taille théorique et le tampon de conservation des protéines sont indiqués. **B.** Courbes d'autocorrélation de *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2-DomN et *Pab*-ASH-Ski2 $\Delta$ C réalisées en diffusion dynamique de la lumière (DLS). **C.** Chromatographie d'exclusion de taille (Superdex 200 10/300) de *Pab*-ASH-Ski2 WT. **D.** Images des cristaux obtenus pour les protéines *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2 DomN et *Pab*-ASH-Ski2 $\Delta$ C. Les conditions de cristallisations sont décrites. **E.** Images des cristaux obtenus pour le complexe *Pab*-ASH-Ski2-DomN/*Pab*-aRNase J.



testées en DLS montrent un retour rapide à la ligne de base (**Figure 33C et 34B**). La polydispersité évaluée pour chaque protéine semble indiquer qu'il y ait une espèce protéique dominante en solution (~polydispersité <25%).

Protéine	C (mg/ml)	C (μM)	MW th (kDa)	Rh th (nm)	Rh (nm)	%Pd
aRNase J	1.2	24	50.5	3.4	8.3	20.8
ASH-Ski2 DomN	8.5	366	23.1	2.6	4.1	25.9
ASH-Ski2	4.6	47.7	97.7	4.2	6.6	20.5
ASH-Ski2 ΔC	9.8	150	65.5	3.7	5.4	24.5

**Tableau 13. Résumé des résultats de DLS.** Les concentrations protéiques (C) sont exprimées en mg/ml et molarité. Les poids moléculaire (MW) et rayon hydrodynamique théorique (Rh théorique calculé pour une protéine de poids moléculaire identique, monomérique et globulaire et observés (Rh) sont indiqués. Le coefficient de polydispersité (Pd) est exprimé en pourcentage.

*Pab*-aRNase J, *Pab*-ASH-Ski2 WT, *Pab*-ASH-Ski2-DomN et *Pab*-ASH-Ski2ΔC présentent des rayons hydrodynamiques (Rh) évalués plus grands que ceux qui pourraient être prédits pour des protéines globulaires ( $Rh=0.794.MW^{(1/3)}$ ) (**Tableau 13**). Ces valeurs peuvent s'expliquer soit par la présence d'une espèce multimérique en solution soit par la forme de la protéine qui serait plus allongée. Les Rh peuvent aussi être légèrement affectés par la présence de 10% de glycérol dans le tampon.

L'état d'oligomérisation de *Pab*-aRNase J a été analysé par chromatographie d'exclusion de taille couplée à la diffusion statique de lumière multi-angle (SEC-MALS) qui est une méthode qui permet de déterminer la masse molaire et l'état d'oligomérisation mais aussi l'homogénéité et la présence d'agrégation. 20μg de *Pab*-aRNase J ont été injectés sur une colonne Superdex S200 I 5/150 préalablement équilibrée dans un tampon HEPES pH7.5, NaCl 300mM et 5% de glycérol à température ambiante. Les données récoltées par diffusion de lumière multi-angle (**Figure 33D**) nous montrent l'absence d'agrégat et une masse molaire mesurée entre 130 et 200kDa qui suggère la présence de plusieurs états d'oligomérisation de *Pab*-aRNase J et probablement un équilibre dimère-tétramère (masse moléculaire théorique de *Pab*-aRNase J de 50kDa).

Pour estimer l'état d'oligomérisation de *Pab*-ASH-Ski2 WT, nous avons réalisé une chromatographie d'exclusion de taille à température ambiante (S200 increase 10/300) dans un tampon HEPES pH 7.5, 300mM de NaCl et 5% de glycérol après avoir chauffé la protéine purifiée 10min à 60°C. Le chromatogramme (**Figure 34C**) nous révèle la présence d'un seul pic à un volume d'élution de 13,5ml qui correspond d'après la calibration de la colonne à une masse molaire apparente de 127kDa, suggérant que la protéine *Pab*-ASH-Ski2 soit monomérique dans ces conditions (masse moléculaire calculée de *Pab*-ASH-Ski2 de 100kDa). La qualité des solutions protéiques obtenues nous a permis d'initier des essais de cristallisation dans le but de résoudre la structure d'ASH-Ski2 seule, en complexe avec aRNase J ou avec des substrats nucléotidiques.

### **3. Etudes structurales préliminaires par cristallographie**

Des premiers essais de cristallisation ont été mis en place avec la protéine *Pab*-ASH-Ski2 WT en présence ou non d'un substrat ADN de 26Nt (Protéine : ADN ratio 2 : 1 et 3 : 1) puis avec *Pab*-ASH-Ski2ΔC et *Pab*-ASH-Ski2-DomN. Les cribles commerciaux : JSCG+, JSCG I, II et III,

Cation, Anion, AmSO<sub>4</sub>, Compass, PEG I, PEG II, PhClear I, PhClear II et PACT (Voir **tableau 8** dans le Matériel et Méthodes) ont été réalisés avec un système automatisé (robot Mosquito®LCP). De petits cristaux, visibles en UV, se sont formés dans différentes conditions (**Figure 34D**) mais n'ont pas diffracté les rayons X lors des tests de diffraction réalisés au synchrotron de l'ESRF (European Synchrotron Radiation Facility) à Grenoble. Pour améliorer les conditions de cristallisation, nous avons réalisé deux cribles manuels plus spécifiques afin de tester ces premières conditions de cristallisation observées (Voir **Annexe-Figure 6**). A ce jour, aucune condition n'a permis la formation de cristaux de *Pab*-ASH-Ski2 capables de diffracter les rayons X. D'autres essais de cristallisation ont été faits avec un mélange *Pab*-aRNase J/*Pab*-ASH-Ski2-DomN. Sept cribles commerciaux (JSCG +, Pro Complex, Compas, PEG I, PEG II, PhClear I et PhClear II) ont été réalisés avec le même système automatisé (robot Mosquito®LCP) avec deux ratios *Pab*-ASH-Ski2-DomN/*Pab*-aRNase J différents (2 :1 et 3 :1). Des aiguilles cristallines se sont formées dans trois conditions (**Figure 34E**) et les expériences de diffraction n'ont pas conduit à l'observation de taches de diffraction sur les lignes de faisceau du synchrotron de l'ESRF à Grenoble. Ces études structurales sont toujours en cours d'optimisation.



# DISCUSSION

Les travaux de thèse présentés dans ce manuscrit s'articulent autour de l'activité et du rôle de l'hélicase ASH-Ski2 de *P. abyssi* en lien avec l'exoribonucléase 5'-3' aRNase J et de l'exosome à ARN ayant une activité 3'-5'.

Au travers des réseaux d'interactions d'ASH-Ski2, aRNase J et des sous-unité Rrp41 et Csl4 de l'exosome à ARN, nous avons pu dégager des liens réciproques entre ces acteurs du métabolisme de l'ARN (RESULTATS- Partie II). Ces liens sont consolidés par nos résultats montrant une interaction *in vitro* entre ASH-Ski2 et aRNase J qui est médiée par le domaine N-terminal de ASH-Ski2 (RESULTATS- Partie III). Ce domaine contenant un motif de type doigt de zinc semble nécessaire et suffisant pour établir une interaction directe avec aRNase J. Nous avons également identifié une interaction *in vitro* entre la sous unité de la coiffe Csl4 de l'exosome à ARN et aRNase J (RESULTATS- Partie III) qui ouvre des perspectives sur la fonction de l'exosome à ARN chez les archées.

La caractérisation de l'activité biochimique de l'hélicase ASH-Ski2 (RESULTATS- Partie IV) a permis d'identifier ASH-Ski2 comme une hélicase à ARN qui dissocie les duplexes ARN/ARN avec une polarité 3'-5' en présence d'ATP et de Mg<sup>2+</sup>. Cette activité est en adéquation avec les activités montrées pour les hélicases de type Ski2 retrouvés chez les eucaryotes. De manière remarquable ASH-Ski2 possède également une capacité intrinsèque d'appariement de brins ARN/ARN en absence d'ATP. Nous proposons que les activités de déroulement de d'association de brins soient associées à deux conformations adoptées par la protéine ASH-Ski2 en présence ou absence d'ATP. Pour comprendre la structure-fonction de *Pab*-ASH-Ski2, nous avons initié une étude structurale (RESULTATS- Partie V) qui ne nous a pas permis à ce jour de résoudre la structure tridimensionnelle de ASH-Ski2 ni en présence ni en absence d'acides nucléiques.

- **Distribution de ASH-Ski2 dans l'arbre du vivant**

Chez les archées, la famille des hélicases de type Ski2 se compose de deux grands groupes, Hel308 et ASH-Ski2 (INTRODUCTION, Partie IV travaux de P. Genevaux/G. Fichant). Hel308 semble ubiquitaire dans l'ensemble de la phylogénie des Archées et est décrite comme une hélicase à ADN chez *P. furiosus*, *M. thermotrophicus* et *S. tokodaii* (Guy and Bolt 2005; Li et al. 2008; Fujikane et al. 2005). La distribution taxonomique de ASH-Ski2 est restreinte aux Euryarchées et, de manière remarquable, est identique à celle de aRNase J à l'exception des Asgard (Publication 2, travaux de P. Genevaux/G. Fichant) (Figure 35). Ces travaux de thèse

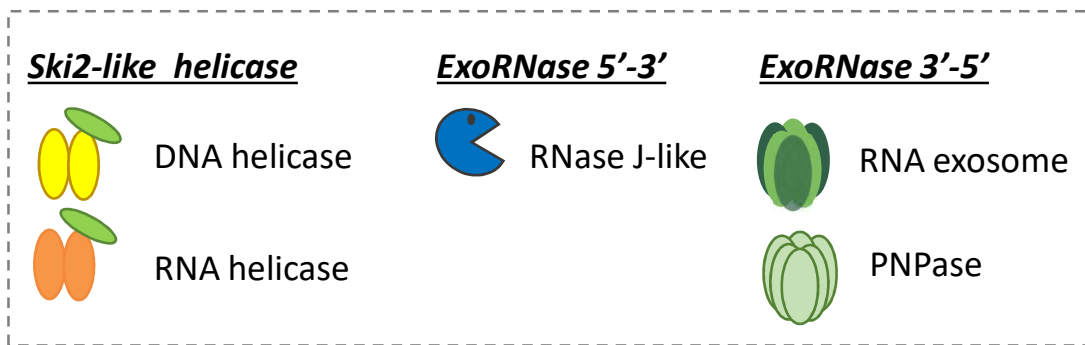
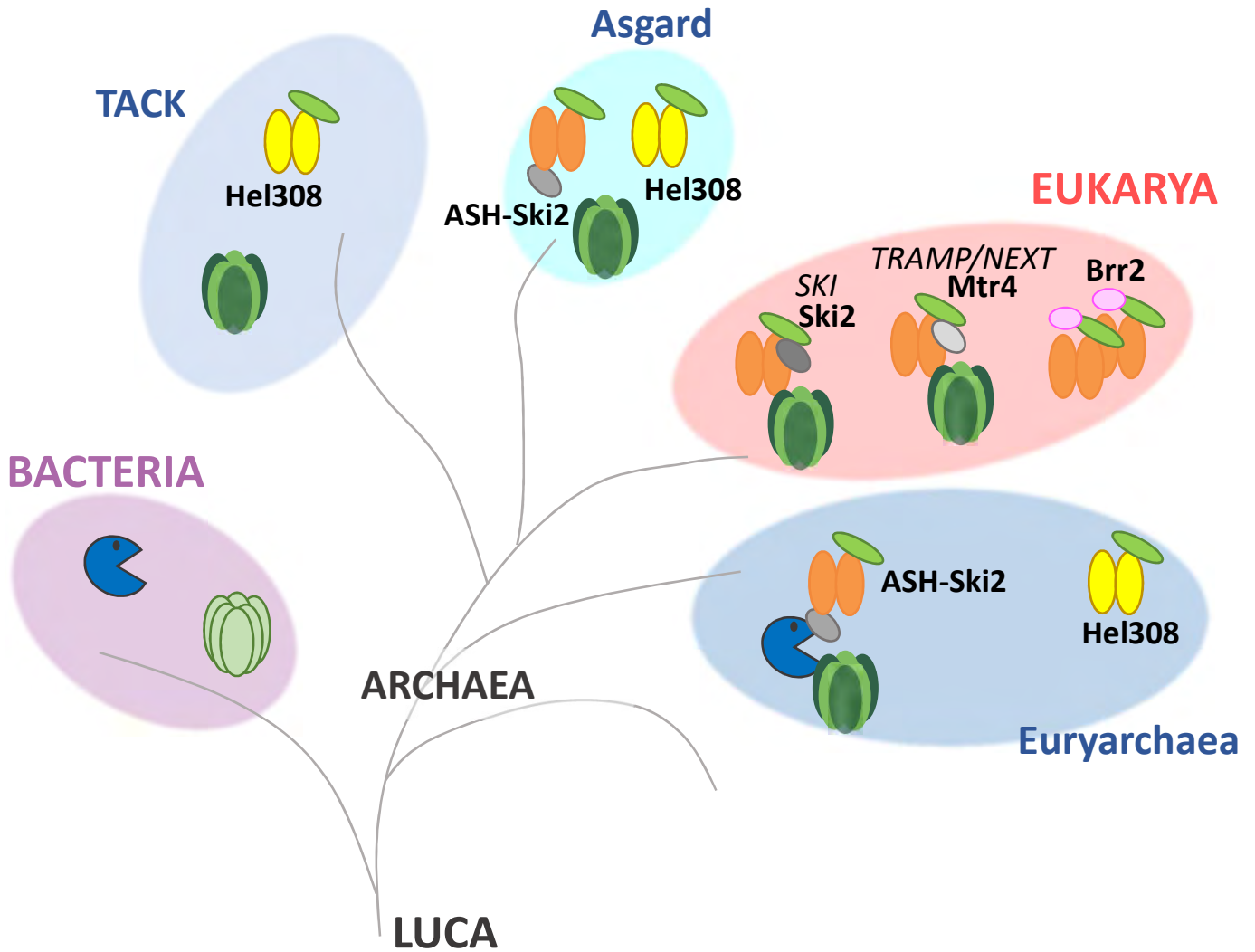


nous permettent d'établir que ASH-Ski2 est une hélicase à ARN (RESULTATS- Partie IV). La distribution des hélicases de la famille de type Ski2 (**Figure 35**) indiquerait une conservation de l'activité hélicase à ARN de type Ski2 (ASH-Ski2) chez les Euryarchées, les Asgards et les eucaryotes et une conservation de l'activité hélicase à ADN (Hel308) chez les TACK et les Euryarchées. On note l'association de l'activité hélicase à ARN de type Ski2-like (Ski2) en lien avec l'exosome à ARN chez les eucaryotes comme proposé par nos travaux chez les Euryarchées. On peut proposer une divergence des hélicases de type Ski2 au cours de l'évolution menant à la reconnaissance de substrats ARN ou ADN, associées respectivement au métabolisme de l'ARN et au métabolisme de l'ADN. Cette évolution pourrait être liée à la présence avec aRNase J. On note l'absence de membres de cette famille chez les bactéries bien que l'exoribonucléase RNase J, homologue à l'aRNase J soit présente. La distribution phylogénétique de ces acteurs pose la question de l'évolution des machineries de dégradation des ARN retrouvées dans les trois domaines du vivant (**Publication 1**) (Clouet-d'Orval et al. 2018; Clouet-d'Orval et al. 2015).

- **Activité enzymatique de ASH-Ski2**

*Pab*-ASH-Ski2 a une activité d'hydrolyse de l'ATP qui est portée par le domaine cœur SF2 constitué des domaines RecA1 et RecA2. *Pab*-ASH-Ski2 semble adopter deux conformations l'une favorable à l'appariement de brin en absence d'ATP et l'autre favorable à la dissociation de brins nécessitant la fixation d'ATP et d'ions  $Mg^{2+}$ . De plus, la dissociation de brins est favorisée avec un duplex possédant une extrémité 3' sortante montrant une polarité de 3' vers 5' (**Figure 36**). Nos résultats indiquent que les domaines accessoires N- et C-terminal n'affectent pas l'activité ATPasique. Cependant la présence d'aRNase J induit une diminution de cette activité. L'interaction avec aRNase J au niveau du domaine N-terminal pourrait engendrer un encombrement allostérique et/ou un changement de conformation du domaine SF2 cœur réduisant son activité enzymatique d'hydrolyse de l'ATP. Il serait intéressant d'examiner à présent les activités de dissociation et associations de brins en présence de aRNase J. Inversement, nous pourrions mesurer l'activité exoribonucléolytique de aRNase J en présence d'ASH-Ski2.

Le domaine C-terminal typique de la famille Ski2-like, contenant le Ratchet domain, est décrit comme un domaine important pour l'activité de déroulement (dissociation de brin) de Hel308 et la délétion de ce domaine est montré induire une létalité dans le cas de Mtr4 (Büttner et



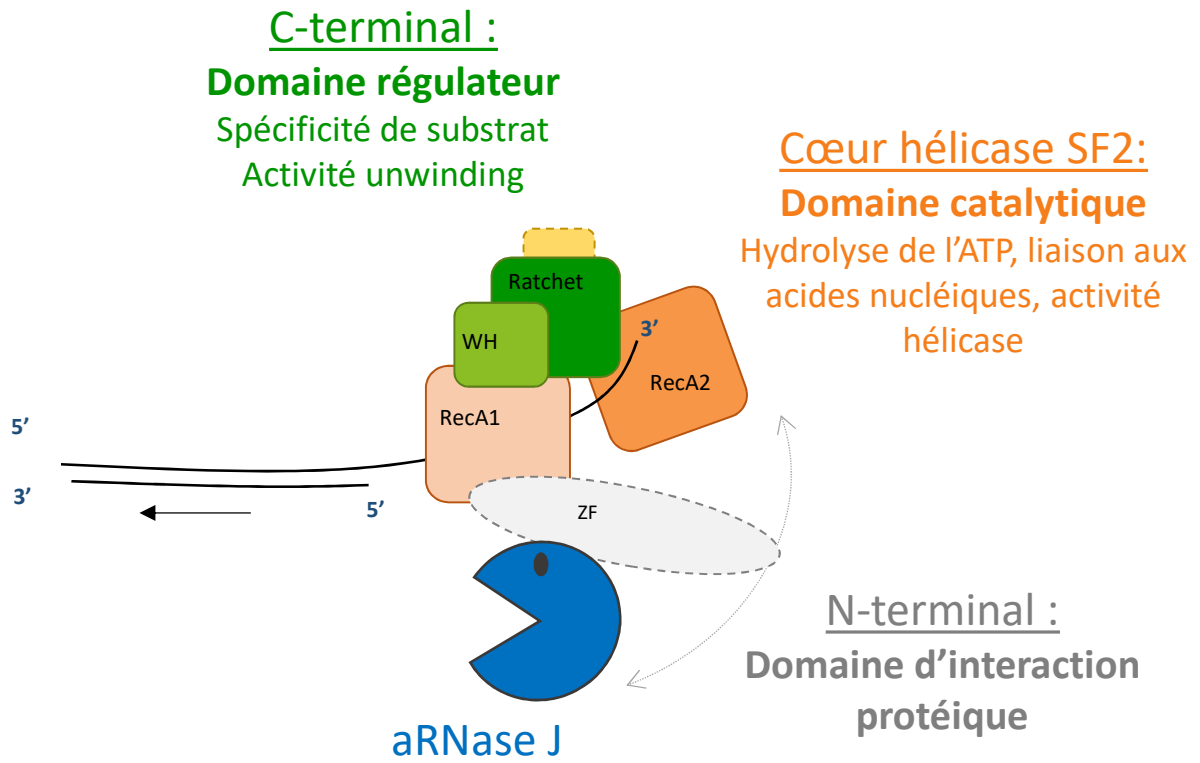
**Figure 35.** Distribution des protéines aRNase J , ASH-Ski2, Hel308 et l'exosome à ARN dans l'arbre du vivant. La branche des bactéries est représentée en violet, celles des archées en bleu et celle des eucaryotes en rouge. Les complexes associés à l'exosome à ARN et aux hélicases Ski2 et Mtr4 sont indiqués.

al. 2007; Holub et al. 2012). Notre étude montre que Le domaine C-terminal, serait important à la fois dans la spécificité de substrat et dans l'inflexion de l'activité vers la dissociation de brins (**Figure 36**). Des études plus détaillées par des approches en molécule unique par exemple permettraient de disséquer les éléments mécanistiques des hélicases de type Ski2 (Hodeib et al. 2017). Il reste à élucider quels sont les substrats ARN de ASH-Ski2. Cette hélicase à ARN est-elle impliquée dans le déplacement de complexe RNP, dans le déroulement de structure d'ARN ou dans la dissociation de brins ?

- **Rôle(s) physiologique(s) de ASH-Ski2 ?**

Nous cherchons à comprendre le rôle d'ASH-Ski2 dans les voies du métabolisme de l'ARN. Récemment, Walker et ses collègues proposent que l'hélicase ASH-Ski2 chez *T. kodakarensis* soit un facteur de la terminaison de transcription qui stimule l'arrêt du complexe d'élongation et le décrochage de l'ARNP (Walker et al. 2017). Cette hypothèse peut être confortée par nos résultats révélant la présence des protéines de liaisons à l'ADN RPA31, RPA42 et la sous unité Rpo2 de l'ARNP dans le réseau d'interactions de *Pab*-ASH-Ski2 (**Publication 2**, RESULTAT-Partie II). Cependant, ce réseau révèle également un lien significatif avec des protéines des sous unités du ribosome et des ribonucléases (aRNase J et exosome à ARN). De plus nos données de fractionnement à partir de culture de cellules de *P. abyssi* montrent que l'hélicase ASH-Ski2 migre dans les hauts poids moléculaires conjointement aux sous-unités du ribosome, de l'exosome à ARN et de l'aRNase J (**Publication 2**). Des données de l'équipe (voir Thèse de D.K. Phung **Figure 20**) montrent que les profils de sédimentation de sous-unité de la RNAP (Rpo3 et Rpo1'') ne se superposent pas à ceux de ASH-Ski2. Ainsi nous proposons en parallèle de l'hypothèse proposée par Walker et ses collègues, d'autres hypothèses sur le rôle physiologique de ASH-Ski2.

Très clairement le réseau de *Pab*-ASH-Ski2 (sans traitement nucléase) (**Publication 2**, RESULTATS- Partie II) révèle de manière significative la présence des sous unités Rrp41, Rrp42, Rrp4 et DnaG de l'exosome à ARN. Il faut noter que chez les eucaryotes, les hélicases de la famille Ski2 ont une fonction fondamentale dans la dégradation, le contrôle qualité et l'épissage des ARN (Johnson and Jackson 2013). Notamment les hélicases Mtr4 et Ski2 appartenant aux complexes multiprotéiques TRAMP, NEXT et SKI sont associés à aux exosomes à ARN nucléaire ou cytoplasmique pour cibler des substrats à ARN spécifiques (ARNm ou ARN non codants) (Falk et al. 2017; Weick et al. 2018; Halbach et al. 2012). De plus



**Figure 36.** Rôles des domaines de l'hélicase ASH-Ski2. Le domaine N-terminal est en gris, le cœur hélicase SF2 en orange et le domaine C-terminal en vert. L'interaction avec aRNase J via le domaine N-terminal d'ASH-Ski2 est schématisée.

nos résultats préliminaires proposent, *in vitro*, une interaction directe entre la sous unité 30S purifiée du ribosome et *Pab*-ASH-Ski2 (RESULTATS, Partie III). Chez les eucaryotes, des structures du complexe SKI, contenant l'hélicase Ski2, interagissant avec le ribosome 80S au cours de la traduction ont été résolues et montrent que l'hélicase Ski2 se lie à la sous unité 40S au niveau de l'entrée de l'ARNm pour faciliter la dégradation de l'ARNm de 3' en 5' par l'exosome à ARN suggérant un rôle dans le contrôle qualité des ARNm (Schmidt et al. 2016; Zinoviev et al. 2020). Dans le noyau, des études ont montré que l'exosome nucléaire et l'hélicase Mtr4 sont liés à la sous unité pré-60S lors de la biogénèse des ribosomes pour permettre la maturation de l'ARNr 5.8S (Schuller et al. 2018).

Au vu du lien étroit entre les hélicases de type Ski2 et le ribosome chez les eucaryotes et de l'ensemble de nos résultats expérimentaux nous proposons que, chez les Euryarchées, l'hélicase ASH-Ski2 pourrait intervenir lors de contrôle qualité des ARNm en cours de traduction ou lors de la biogénèse du ribosome en concertation avec l'exosome à ARN. Cette hypothèse est confortée par les analyses transcriptomiques initiées dans l'équipe. Les transcriptomes d'une souche sauvage de *Thermococcus barophilus* (*Tbar-WT*) (modèle d'étude génétique chez les Thermococcales, proche de *P. abyssi*) et d'une souche invalidée pour le gène codant pour ASH-Ski2 (*Tbar-ΔASH-Ski2*) sont comparées. Les premières analyses révèlent que 171 gènes sont surexprimés dans la souche *Tbar-ΔASH-Ski2* par rapport à la *Tbar-WT* et 139 gènes sont sous-exprimés (données de Marie Bouvier dans l'équipe). Parmi les gènes surexprimés, les gènes codant pour les sous unités du ribosome (L23Ae, L34e, S8e, L18Ae et S27Ae), et les sous unités Rrp4 et Rrp41 de l'exosome à ARN sont identifiés. Ces données préliminaires sont en cours de validation.

- **Mise évidence du complexe AHS-Ski2/aRNase J**

Les travaux menés durant cette thèse ont permis de montrer que l'hélicase à ARN ASH-Ski2 forme un complexe multiprotéique via son domaine N-terminal avec aRNase J qui porte une activité exoribonucléolytique de 5' vers 3' (RESULTATS- Partie III, [Publication 2](#)) (Clouet-d'Orval et al. 2010). Ainsi pour la première fois un complexe associant une hélicase et une ribonucléase est mis en évidence chez les Archées. Nous avons montré que le domaine N-terminal est le domaine d'interaction avec l'aRNase J et avons identifié que les résidus cystéines du motif en doigt de zinc sont importants pour la co-purification du complexe aRNase J/ ASH-Ski2. Nous avons pu définir qu'ASH-Ski2 est sous forme monomérique en

solution tandis que aRNase J semble en équilibre entre une forme dimérique et tétramérique. Il est à noter que la structure résolue pour aRNase J de *M. psychrophilus* montre une forme tétramérique (Zheng et al. 2017). Nous avons initié des études biophysiques pour essayer de définir la stœchiométrie du complexe aRNase J/ASH-Ski2 ainsi que des études structurales du complexe aRNase J/domaine N-terminal de ASH-Ski2. Malheureusement, à ce jour, nous n'avons pu obtenir de données ni sur la stœchiométrie ni sur la structure du complexe aRNase J/ASH-Ski2 afin d'élucider la structure-fonction de ce complexe (RESULTATS- Partie V). Cependant nous avons des évidences de la formation de ce complexe en solution en particulier nous avons montré que l'activité ATPasique de ASH-Ski2 est altérée en présence de aRNase J révélant que la formation du complexe modulerait l'activité ATPasique et possiblement donc l'activité de dissociation de brins qui y est associée (RESULTATS- Partie IV). Il reste également à définir quels sont les motifs de aRNase J critiques pour la formation du complexe aRNase J/ASH-Ski2. La co-occurrence de aRNase J et de ASH-Ski2 au sein de la taxonomie des Euryarchées suggérant une co-évolution de ces deux partenaires conforte nos résultats expérimentaux proposant l'existence du complexe aRNase J/ASH-Ski2 chez les Euryarchées (**Publication 2**). Cependant contrairement à aRNase J, le gène codant pour ASH-Ski2 est présent dans le groupe des Asgards dans lequel aucune activité 5'-3' n'a encore été identifiée (travaux de P. Genevaux/G. Fichant). Est-ce que chez les Asgards, ASH-Ski2 interagirait avec une autre exoribonucléase 5'-3' spécifique de ce groupe taxonomique? D'après nos résultats, ASH-Ski2 est capable de dissocier préférentiellement les duplex d'ARN de 3' en 5' (RESULTATS- Partie IV) et aRNase J possède une activité exoribonucléolytique de des extrémités 5' vers 3' (Clouet-d'Orval et al. 2010).

- **Rôle(s) physiologique(s) du complexe aRNase J/ASH-Ski2?**

Comme indique dans le premier paragraphe de cette discussion, il est proposé que ASH-Ski2 soit impliquée dans la terminaison de la transcription (Walker et al. 2017). La terminaison de la transcription implique également chez les eucaryotes des ribonucléases. Il est proposé que l'activité 5'-3' exoribonucléolytique de Xrn2 (ou Rat1 chez la levure) soit associée à la voie de terminaison de la transcription dans un modèle nommé « Torpedo » (Mandel et al. 2006). Après une coupure endonucléolytique par CPSF73 qui définit l'extrémité 3' des transcrits, l'exoribonucléase 5'-3' processive Xrn2 dégrade l'ARN en aval du site de coupure et rattrape l'ARNPII pour la libérer de l'ADN matrice (Kim et al. 2004; West et al. 2004). On peut imaginer



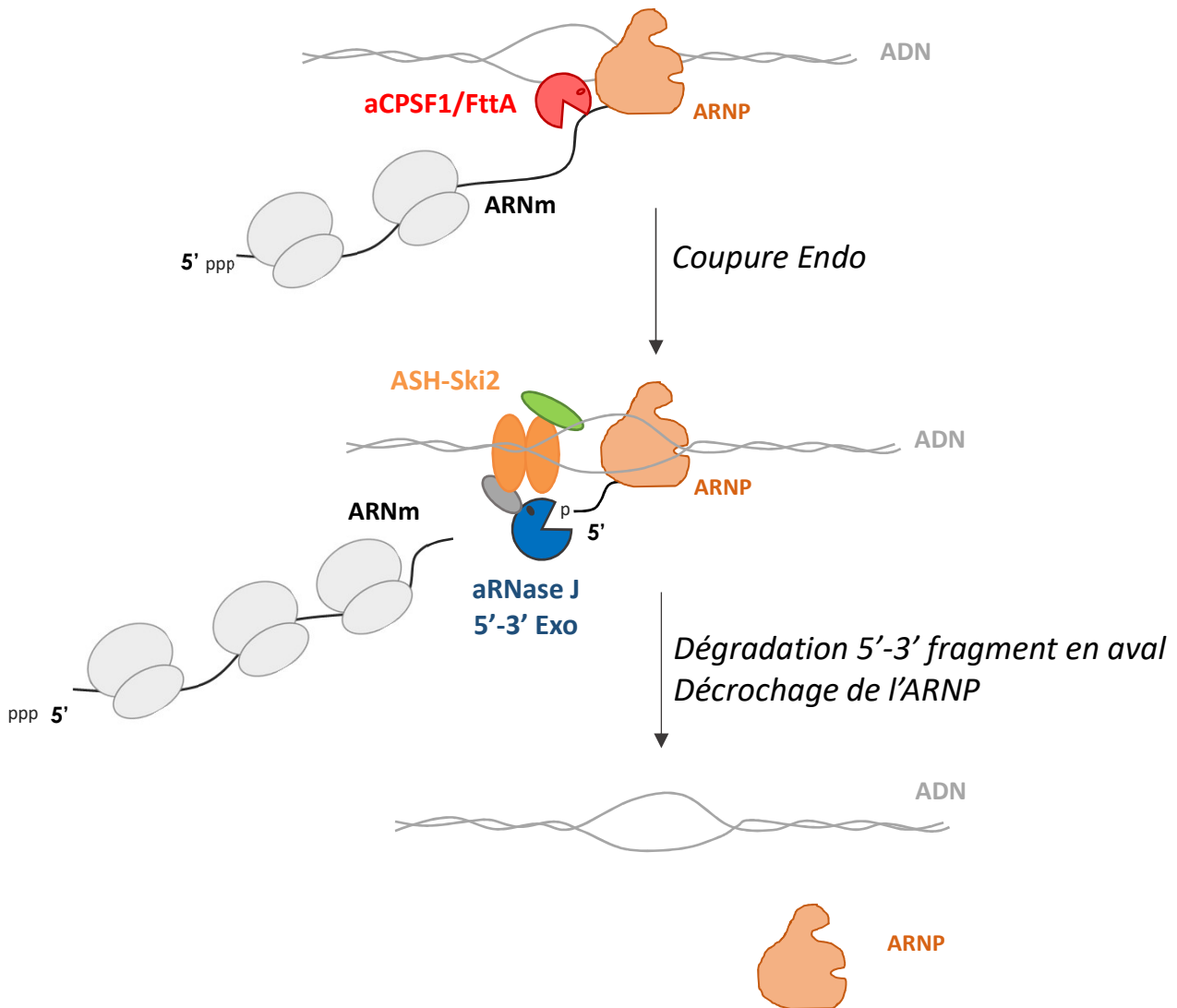
que le complexe ASH-Ski2/aRNase J participerait à ce mécanisme. aRNase J agirait en tant que 5'-3' ExoRNase pour digérer le transcrit restant en cours de transcription après la coupure endonucléolytique qui pourrait être réalisée par aCPSF1/FTTA comme proposé récemment (Sanders et al. 2020) et ASH-Ski en tant que hélicase dissocierait le transcrit naissant de la matrice ADN (**Figure 37**).

Une autre hypothèse en adéquation avec nos données de fractionnement cellulaires (**Publication 2**) comme discuté dans la première partie de la discussion est l'implication du complexe ASH-Ski2/aRNase J dans le contrôle qualité des ARN messagers en cours de traduction (**Figure 38**) (Phung et al. 2020). En effet chez les eucaryotes, des données structurales récentes montrent que l'exoribonucléase 5'-3' Xrn1, impliquée dans divers processus tels que la régulation de la transcription ou le processing des petits ARN, interagit avec le ribosome 80S suggérant un rôle dans le contrôle qualité des ARNm (Tesina et al. 2019). En cohérence nos données préliminaires (RESULTATS- Partie III) montrent une interaction possible entre la sous unité 30S du ribosome purifiée et aRNase J.

- **L'exosome à ARN peut-il exister sous plusieurs formes ?**

Nous avons obtenu plusieurs évidences d'un lien entre la sous-unité catalytique Rrp41 et la sous-unité de la coiffe Csl4 composants de l'exosome à ARN, aRNase J et ASH-Ski2 (**Publication 2**, RESULTATS- Parties II et III). En particulier nous observons uniquement une interaction entre la sous-unité Csl4 de la coiffe et aRNase J. La sous-unité Rrp4 ne semble pas interagir avec aRNase J dans nos conditions expérimentales. Ces deux sous-unités ont en commun un domaine S1 de liaison aux ARN et diffèrent par leur domaine C-terminal avec un domaine en doigt de Zinc pour Csl4 et un domaine de type KH pour Rrp4. Nous souhaitons à présent de reconstituer de complexes d'exosome à ARN *in vitro* pour confirmer les interactions directes de Csl4 avec aRNase J. Nous proposons de reconstituer trois types d'exosome à ARN, avec des coiffes d'homo-trimère de Csl4 ou d'homo-trimère de Rrp4 ou d'heterotrimère de Csl4/Rrp4 afin de tester une spécificité d'interaction avec aRNase J et le complexe ASH-Ski2/aRNase J. Il faut noter que chez les archées, la composition stœchiométrique des sous-unités de la coiffe et la fonction dans le métabolisme de l'ARN de l'exosome à ARN *in vivo* reste à ce jour non élucidées. On peut imaginer qu'il existe plusieurs exosomes à ARN selon la composition de la coiffe dans la cellule qui permettrait de cibler des substrats ARN spécifiques selon les voies métaboliques. Ces différents types d'exosome à ARN pourraient être associés à des

## Modèle « Torpedo »



**Figure 37.** Proposition du modèle dit « Torpedo » de la terminaison de la transcription chez les Euryarchées. Le transcrit pourrait être clivé par une endoribonucléase, aCPSF1/FttA, qui génère une extrémité 5' monophosphate pouvant être dégradé par la ribonucléase 5'-3' aRNase J. ASH-Ski2 dissocierait le transcrit naissant de la matrice ADN.

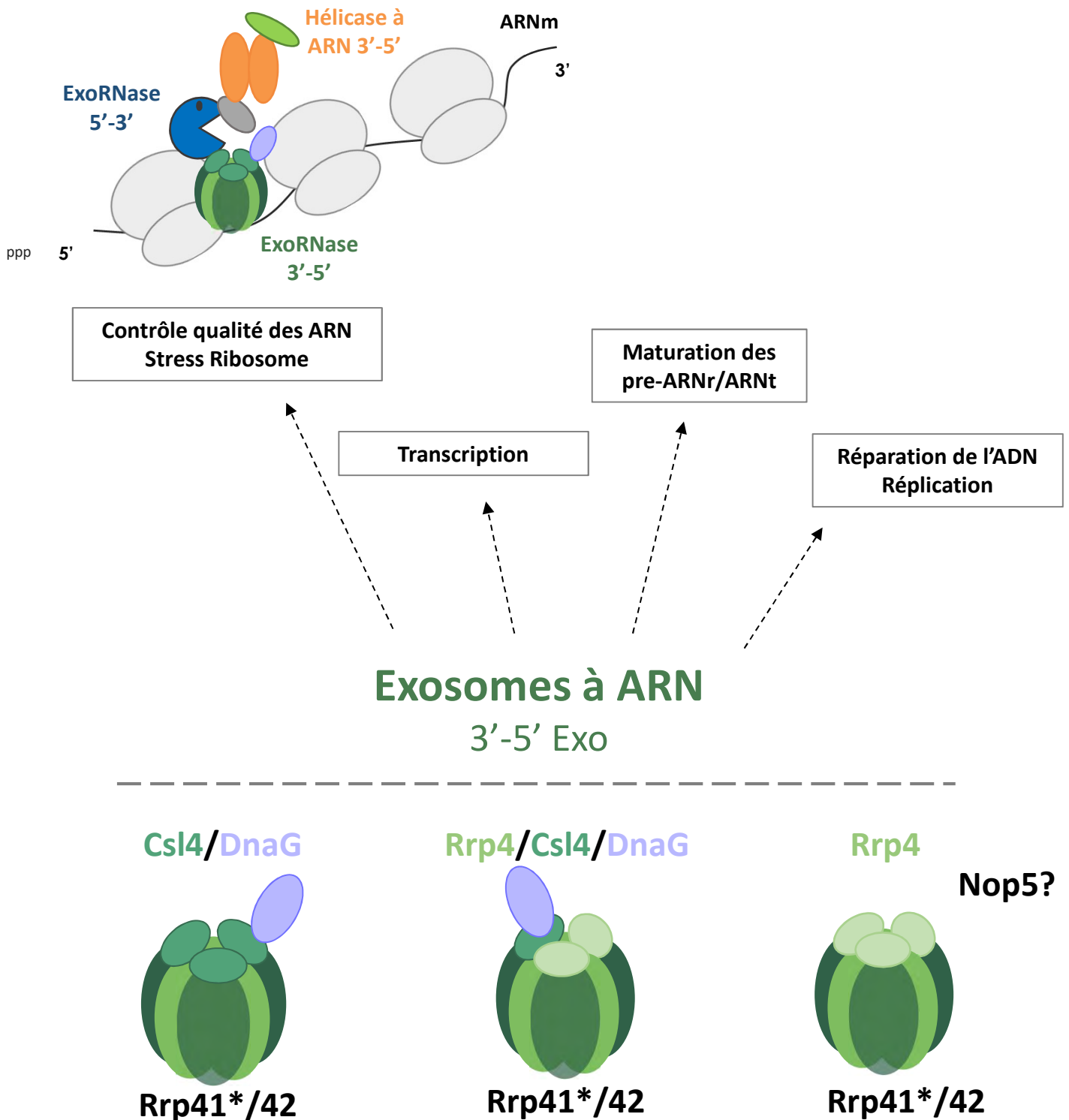
partenaires spécifiques (**Figure 38**). Avec cet objectif nous avons également entrepris de déterminer le réseau d'interaction de Rrp4 pour comparer les partenaires communs et spécifiques entre les réseaux de Csl4 et de Rrp4. Il a été montré que DnaG dont la fonction reste méconnue au sein de l'exosome à ARN est associée à Csl4. Csl4/DnaG pourrait jouer un rôle dans la spécificité de l'activité de l'exosome à ARN. Nous observons également que les partenaires de Csl4 sont majoritairement des protéines du ribosome. On se pose donc la question de la relation qu'il existe entre l'exosome à ARN et le ribosome. Dans ce contexte, nous avons initié à partir de fractions cellulaires de haut poids moléculaires (voir gradient **Figure S4- Publication 2**) contenant à la fois les sous-unités du ribosome et l'exosome, l'analyse de ces échantillons en cryo-microscopie électronique (Travaux de L. Plassart/C. Plisson). Des premières images en coloration négative permettent d'observer des éléments pouvant correspondre à une structure d'exosome à ARN qui parfois semblent associée à un élément de type ribosome.

Dans la littérature il est montré *in vitro* que la sous-unité Rrp4 confère une forte spécificité poly(A) à l'exosome (Roppelt et al. 2010) alors que la sous-unité Csl4 permet la liaison avec la sous unité aDNaG (Evguenieva-Hackenberg et al. 2008) Une étude récente montre une interaction entre Nop5, une protéine appartenant à la RNP guide à boîtes C/D impliquée dans la méthylation des ARNr et ARNt, et l'exosome possédant une coiffe composée d'un trimère de Rrp4 chez *S. solfataricus* (Gauernack et al. 2017).

Chez les eucaryotes, il existe également plusieurs exosomes à ARN, un type cytoplasmique et un type nucléaire avec des compositions en sous-unités différentes (Puno et al. 2019). Ces différents complexes chez les eucaryotes portant à la fois une activité 3'-5' exo-et endo-nucléolytique semblent impliqués dans un large panel de voies métaboliques allant de dégradation, à la maturation au contrôle qualité des ARN (Puno et al. 2019). Notamment, comme mentionné précédemment, les exosomes à ARN sont associés à des hélicases de la famille Ski2 telles que Mtr4 et Ski2 pour faciliter l'adressage des substrats ARN (Falk et al. 2017; Weick et al. 2018; Halbach et al. 2012).

- **Conclusion générale**

D'après nos résultats et les données bibliographiques, nous proposons que chez les euryarchées, un complexe aRNase J/ASH-Ski2/Exosome-Csl4 serait en lien dans le ribosome pour le contrôle qualité des ARNm en cours de traduction (**Figure 38**). De plus, le contexte



**Figure 38.** Le réseau d'interactions de la sous unité catalytique du tonneau de l'exosome, Rrp41, révèle une diversité de partenaires protéiques (RESULTATS- Partie II) ce qui nous laisse penser que le tonneau de l'exosome seul n'est pas spécifique à un seul processus mais peut certainement être requis dans diverses fonctions et voies métaboliques.

génomique observé chez les Thermococcales montre que les gènes codant pour ASH-Ski2 et Csl4 se retrouvent dans un même opéron, ceci renforce un lien fonctionnel entre l'hélicase ASH-Ski2 et la sous-unité de la coiffe Csl4. Les gènes codants pour les sous-unités de l'exosome à ARN sont présents dans la grande majorité des génomes d'archées à l'exception des Halobactéries, Haloferacales et Méthanomicrobiale dans lesquels l'activité 3'-5' exo semble porter par l'exo RNase de type aRNase R (Clouet-d'Orval et al. 2018). On se pose donc la question de la relation entre aRNase R et le complexe aRNase J/ASH-Ski2 qui semble présent dans ces groupes taxonomiques. Pour conclure, ce travail ouvre des perspectives fonctionnelles et évolutives autour de l'exoribonucléase 5'-3' de type  $\beta$ -CASP chez les Euryarchées. aRNase J forme un complexe stable avec l'hélicase à ARN ASH-Ski2 et est associée avec la sous-unité Csl4 de la coiffe de l'exosome à ARN. L'ensemble de ces acteurs pourrait agir ensemble à proximité du ribosome dans des voies métaboliques de contrôle qualité des ARN en cours de traduction.



## REMERCIEMENTS

*Je tiens à remercier en premier les membres du Jury, Emmanuelle Schmitt, Tamara Basta Le Berre, Fabien Darfeuille & Gwennaële Fichant qui ont consacré du temps à la lecture de ce manuscrit et à l'évaluation de mes travaux.*

*Je tiens tout particulièrement à remercier ma directrice de thèse Béatrice Clouet d'Orval, qui m'a associée à cette équipe en tant que doctorante, qui m'a financée, m'a encadrée, m'a faite confiance tout au long de ces quatre années & m'a permise de m'épanouir dans la Science.*

*Merci Duy Khanh, de m'avoir intégrée à ce labo, de m'avoir formée sur tout ce que tu as mis au point dans l'équipe et d'avoir toujours répondu présent même après avoir quitté l'IBCG.*

*Je remercie également toute l'équipe de Béatrice & AJ Carpousis, Marie, Mirna, Julien, Marta, Léo, Isa et Lydia pour leur disponibilité et leur bonne humeur, et tous les stagiaires que j'ai croisés/encadrés et qui ont contribué au projet de l'équipe et tout particulièrement Laura. Un grand merci à Marie, Marta, DK & Laura pour la relecture de ce manuscrit !*

*Je remercie aussi Valérie Guillet de m'avoir encadré sur toutes les expériences de biophysiques et cristallographie, merci à l'équipe de Lionel Mourey de m'avoir permis de revenir manipuler dans cette équipe.*

*Je remercie l'équipe de Gwennaële Fichant pour leur collaboration sur les études phylogénomiques.*

*Je remercie également Emmanuelle Schmitt de m'avoir accueilli une semaine dans son équipe et merci à Gabrielle de m'avoir accompagné sur les différentes manipulations.*

*Je tiens à remercier Sébastien Laurent qui a eu la patience de m'expliquer les analyses MS/MS des Pull down.*

*Merci à mon comité de thèse, Odile Humbert et Ghislaine Henneke, d'avoir pris le temps de lire mes rapports et pour leurs conseils.*

*Merci à mes maîtres de stage de Licence et Master, Pierre Sourdille et Didier Zerbib, d'avoir participé à ma formation de Scientifique.*

*Je souhaite remercier l'ensemble des membres de l'IBCG pour cette ambiance de travail chaleureuse et en particulier l'équipe Upsi/Tarot : Steph, Ramtin, Clément, Lise, Cécilia, Sylvain, Tommy, DK, Elise et Alix mais aussi l'axe Aperçu orchestré par Alex, Laura, David, Moïse, Cyril, Jade, Alix, Côme...avec une mention spéciale pour mes amis et alliés Luis (Filleul) et Mickaël (Ams) !*

*Merci à mon binôme depuis la L1, Cornélia pour ces années d'études inoubliables et ton soutien perpétuel.*

*Merci à mes amis hors-Labo qui ne comprennent pas un mot du titre de ma thèse mais qui ont contribué à la rendre moins stressante pendant ces quatre années : Aurore, Fouko, Antho, Rulian, Viala, Thibault, Allard, Hugo & Camille.*

*Je remercie Tommy pour son inlassable patience et son soutien particulier depuis mon Master 2.*

*Enfin je terminerai en remerciant ma famille : Maman, Papa & Fanny sans qui tout ça n'aurait jamais été possible, merci pour la confiance aveugle que vous m'accordez depuis toutes ces années, pour votre soutien inestimable et vos encouragements précieux. Cette thèse est autant la mienne que la vôtre.*





## Bibliographie

- Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V. and Polz, M.F. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *Journal of Bacteriology* 186(9), pp. 2629–2635.
- Adam, P.S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *The ISME Journal* 11(11), pp. 2407–2425.
- Ahdash, Z., Lau, A.M., Byrne, R.T., Lammens, K., Stüetzer, A., Urlaub, H., Booth, P.J., Reading, E., Hopfner, K.-P. and Politis, A. 2017. Mechanistic insight into the assembly of the HerA-NurA helicase-nuclease DNA end resection complex. *Nucleic Acids Research* 45(20), pp. 12025–12038.
- Allers, T. and Mevarech, M. 2005. Archaeal genetics - the third way. *Nature Reviews. Genetics* 6(1), pp. 58–73.
- Aravind, L. 1999. An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biology* 1(2), pp. 69–91.
- Aravind, L., Iyer, L.M. and Anantharaman, V. 2003. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biology* 4(10), p. R64.
- Arkhipova, V., Stolboushkina, E., Kravchenko, O., Kljashtorny, V., Gabdulkhakov, A., Garber, M., Nikonov, S., Märten, B., Bläsi, U. and Nikonov, O. 2015. Binding of the 5'-Triphosphate End of mRNA to the  $\gamma$ -Subunit of Translation Initiation Factor 2 of the Crenarchaeon *Sulfolobus solfataricus*. *Journal of Molecular Biology* 427(19), pp. 3086–3095.
- Armache, J.-P., Anger, A.M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G.J., Beckmann, R. and Wilson, D.N. 2013. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Research* 41(2), pp. 1284–1293.
- Atomi, H., Imanaka, T. and Fukui, T. 2012. Overview of the genetic tools in the Archaea. *Frontiers in microbiology* 3, p. 337.
- Babski, J., Haas, K.A., Näther-Schindler, D., Pfeiffer, F., Förstner, K.U., Hammelmann, M., Hilker, R., Becker, A., Sharma, C.M., Marchfelder, A. and Soppa, J. 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* 17(1), p. 629.
- Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E. and Lloyd, K.G. 2020. Diversity, ecology and evolution of Archaea. *Nature Microbiology*.
- Balakrishnan, L. and Bambara, R.A. 2013. Flap endonuclease 1. *Annual Review of Biochemistry* 82, pp. 119–138.
- Ban, N., Beckmann, R., Cate, J.H.D., Dinman, J.D., Dragon, F., Ellis, S.R., Lafontaine, D.L.J., Lindahl, L., Liljas, A., Lipton, J.M., McAlear, M.A., Moore, P.B., Noller, H.F., Ortega, J., Panse, V.G., Ramakrishnan, V., Spahn, C.M.T., Steitz, T.A., Tchorzewski, M., Tollervey, D., Warren, A.J.,

- Williamson, J.R., Wilson, D., Yonath, A. and Yusupov, M. 2014. A new system for naming ribosomal proteins. *Current Opinion in Structural Biology* 24, pp. 165–169.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289(5481), pp. 905–920.
- Barthelme, D., Chen, J.Z., Grabenstatter, J., Baker, T.A. and Sauer, R.T. 2014. Architecture and assembly of the archaeal Cdc48•20S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* 111(17), pp. E1687–94.
- Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U. and Tampé, R. 2011. Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proceedings of the National Academy of Sciences of the United States of America* 108(8), pp. 3228–3233.
- Barthelme, D. and Sauer, R.T. 2012. Identification of the Cdc48•20S proteasome as an ancient AAA+ proteolytic machine. *Science* 337(6096), pp. 843–846.
- Bartlett, E.J., Brissett, N.C. and Doherty, A.J. 2013. Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proceedings of the National Academy of Sciences of the United States of America* 110(22), pp. E1984–91.
- Baumeister, W., Walz, J., Zühl, F. and Seemüller, E. 1998. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* 92(3), pp. 367–380.
- Becker, T., Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Sieber, H., Motaal, B.A., Mielke, T., Berninghausen, O. and Beckmann, R. 2011. Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nature Structural & Molecular Biology* 18(6), pp. 715–720.
- Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.-P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., Karcher, A., Thomm, M., Hopfner, K.-P., Green, R. and Beckmann, R. 2012. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 482(7386), pp. 501–506.
- Bell, S.D., Botting, C.H., Wardleworth, B.N., Jackson, S.P. and White, M.F. 2002. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296(5565), pp. 148–151.
- Bell, S.D., Jaxel, C., Nadal, M., Kosa, P.F. and Jackson, S.P. 1998. Temperature, template topology, and factor requirements of archaeal transcription. *Proceedings of the National Academy of Sciences of the United States of America* 95(26), pp. 15218–15222.
- Benaroudj, N. and Goldberg, A.L. 2000. PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nature Cell Biology* 2(11), pp. 833–839.
- Benelli, D. and Londei, P. 2009. Begin at the beginning: evolution of translational initiation. *Research in Microbiology* 160(7), pp. 493–501.
- Benelli, D. and Londei, P. 2011. Translation initiation in Archaea: conserved and domain-specific features. *Biochemical Society Transactions* 39(1), pp. 89–93.

- Benelli, D., La Teana, A. and Londei, P. 2017. Translation Regulation: The Archaea-Eukaryal Connection. In: Clouet-d'Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 71–88.
- Berger, J.M. 2008. SnapShot: nucleic acid helicases and translocases. *Cell* 134(5), pp. 888–888.e1.
- Berkemer, S.J., Maier, L.-K., Amman, F., Bernhart, S.H., Wörtz, J., Märkle, P., Pfeiffer, F., Stadler, P.F. and Marchfelder, A. 2020. Identification of RNA 3' ends and termination sites in *Haloferax volcanii*. *RNA Biology* 17(5), pp. 663–676.
- Bernstein, J., Patterson, D.N., Wilson, G.M. and Toth, E.A. 2008. Characterization of the essential activities of *Saccharomyces cerevisiae* Mtr4p, a 3'->5' helicase partner of the nuclear exosome. *The Journal of Biological Chemistry* 283(8), pp. 4930–4942.
- Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. and Werner, F. 2019. Key concepts and challenges in archaeal transcription. *Journal of Molecular Biology* 431(20), pp. 4184–4201.
- Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C., Smollett, K.L., Albers, S.V., Kay, C.W.M., Thalassinou, K. and Werner, F. 2015. Archaeal TFE $\alpha/\beta$  is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife* 4, p. e08378.
- Bocquier, A.A., Liu, L., Cann, I.K., Komori, K., Kohda, D. and Ishino, Y. 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Current Biology* 11(6), pp. 452–456.
- Boneberg, F.M., Brandmann, T., Kobel, L., van den Heuvel, J., Bargsten, K., Bammert, L., Kutay, U. and Jinek, M. 2019. Corrigendum: Molecular mechanism of the RNA helicase DHX37 and its activation by UTP14A in ribosome biogenesis. *RNA (New York)* 25(11), p. 1576.
- Branson, O.E. and Freitas, M.A. 2016. Tag-Count Analysis of Large-Scale Proteomic Data. *Journal of Proteome Research* 15(12), pp. 4742–4746.
- Brinkman, A.B., Ettema, T.J.G., de Vos, W.M. and van der Oost, J. 2003. The Lrp family of transcriptional regulators. *Molecular Microbiology* 48(2), pp. 287–294.
- Brown, J.W., Daniels, C.J. and Reeve, J.N. 1989. Gene structure, organization, and expression in archaeobacteria. *Critical reviews in microbiology* 16(4), pp. 287–338.
- Brugère, J.-F., Borrel, G., Gaci, N., Tottey, W., O'Toole, P.W. and Malpuech-Brugère, C. 2014. Archaeobiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut microbes* 5(1), pp. 5–10.
- Büttner, K., Nehring, S. and Hopfner, K.-P. 2007. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nature Structural & Molecular Biology* 14(7), pp. 647–652.
- Büttner, K., Wenig, K. and Hopfner, K.-P. 2005. Structural framework for the mechanism of archaeal exosomes in RNA processing. *Molecular Cell* 20(3), pp. 461–471.
- Cao, Y., Li, J., Jiang, N. and Dong, X. 2014. Mechanism for stabilizing mRNAs involved in methanol-dependent methanogenesis of cold-adaptive *Methanosarcina mazei* zm-15. *Applied and Environmental Microbiology* 80(4), pp. 1291–1298.

- Carpousis, A.J., Luisi, B.F. and McDowall, K.J. 2009. Chapter 3 Endonucleolytic Initiation of mRNA Decay in *Escherichia coli*. In: *Molecular biology of RNA processing and decay in prokaryotes*. Progress in molecular biology and translational science. Elsevier, pp. 91–135.
- Chamieh, H., Ibrahim, H. and Kozah, J. 2016. Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* 576(1 Pt 2), pp. 214–228.
- Chamot, D., Colvin, K.R., Kujat-Choy, S.L. and Owtrim, G.W. 2005. RNA structural rearrangement via unwinding and annealing by the cyanobacterial RNA helicase, CrhR. *The Journal of Biological Chemistry* 280(3), pp. 2036–2044.
- Chan, P.P., Cozen, A.E. and Lowe, T.M. 2011. Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biology* 12(4), p. R38.
- Chan, P.P. and Lowe, T.M. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* 37(Database issue), pp. D93–7.
- Charpentier, B., Fourmann, J. and Branlant, C. 2007. Reconstitution of Archaeal H/ACA sRNPs and Test of their Activity. In: *RNA Modification*. Methods in Enzymology. Elsevier, pp. 389–405.
- Chemnitz Galal, W., Pan, M., Kelman, Z. and Hurwitz, J. 2012. Characterization of DNA primase complex isolated from the archaeon, *Thermococcus kodakaraensis*. *The Journal of Biological Chemistry* 287(20), pp. 16209–16219.
- Chlebowski, A., Lubas, M., Jensen, T.H. and Dziembowski, A. 2013. RNA decay machines: the exosome. *Biochimica et Biophysica Acta* 1829(6-7), pp. 552–560.
- Cho, I.-M., Lai, L.B., Susanti, D., Mukhopadhyay, B. and Gopalan, V. 2010. Ribosomal protein L7Ae is a subunit of archaeal RNase P. *Proceedings of the National Academy of Sciences of the United States of America* 107(33), pp. 14573–14578.
- Clouet-d'Orval, B., Batista, M., Bouvier, M., Quentin, Y., Fichant, G., Marchfelder, A. and Maier, L.-K. 2018. Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiology Reviews* 42(5), pp. 579–613.
- Clouet-d'Orval, B., Phung, D.K., Langendijk-Genevaux, P.S. and Quentin, Y. 2015. Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of  $\beta$ -CASP ribonucleases. *Biochimie* 118, pp. 278–285.
- Clouet-d'Orval, B., Rinaldi, D., Quentin, Y. and Carpousis, A.J. 2010. Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *The Journal of Biological Chemistry* 285(23), pp. 17574–17583.
- Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Quérellou, J., Ripp, R., Thierry, J.-C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. and Forterre, P. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Molecular Microbiology* 47(6), pp. 1495–1512.
- Cordin, O. and Beggs, J.D. 2013. RNA helicases in splicing. *RNA Biology* 10(1), pp. 83–95.

- Cossu, M., Da Cunha, V., Toffano-Nioche, C., Forterre, P. and Oberto, J. 2015. Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged Thermococcales chromosomes. *Biochimie* 118, pp. 313–321.
- Costa, A. and Onesti, S. 2009. Structural biology of MCM helicases. *Critical Reviews in Biochemistry and Molecular Biology* 44(5), pp. 326–342.
- Coureux, P.-D., Lazennec-Schurdevin, C., Bourcier, S., Mechulam, Y. and Schmitt, E. 2020. Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Communications Biology* 3(1), p. 58.
- Coureux, P.-D., Lazennec-Schurdevin, C., Monestier, A., Larquet, E., Cladière, L., Klaholz, B.P., Schmitt, E. and Mechulam, Y. 2016. Cryo-EM study of start codon selection during archaeal translation initiation. *Nature Communications* 7, p. 13366.
- Craig, J.M., Laszlo, A.H., Brinkerhoff, H., Derrington, I.M., Noakes, M.T., Nova, I.C., Tickman, B.I., Doering, K., de Leeuw, N.F. and Gundlach, J.H. 2017. Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proceedings of the National Academy of Sciences of the United States of America* 114(45), pp. 11932–11937.
- Crowley, D.J., Boubriak, I., Berquist, B.R., Clark, M., Richard, E., Sullivan, L., DasSarma, S. and McCready, S. 2006. The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Systems* 2, p. 11.
- Da Cunha, V., Gaia, M., Nasir, A. and Forterre, P. 2018. Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genetics* 14(3), p. e1007215.
- Dacks, J.B., Field, M.C., Buick, R., Eme, L., Gribaldo, S., Roger, A.J., Brochier-Armanet, C. and Devos, D.P. 2016. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *Journal of Cell Science* 129(20), pp. 3695–3703.
- Danan, M., Schwartz, S., Edelheit, S. and Sorek, R. 2012. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Research* 40(7), pp. 3131–3142.
- Dar, D., Prasse, D., Schmitz, R.A. and Sorek, R. 2016. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nature Microbiology* 1(10), p. 16143.
- Daume, M., Uhl, M., Backofen, R. and Randau, L. 2017. RIP-Seq Suggests Translational Regulation by L7Ae in Archaea. *mBio* 8(4).
- Delmas, S., Duggin, I.G. and Allers, T. 2013. DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon *Haloferax volcanii*. *Molecular Microbiology* 87(1), pp. 168–179.
- DeLong, E.F. and Pace, N.R. 2001. Environmental diversity of bacteria and archaea. *Systematic Biology* 50(4), pp. 470–478.
- Dennis, P.P., Tripp, V., Lui, L., Lowe, T. and Randau, L. 2015. C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. *BMC Genomics* 16, p. 632.
- Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Bäumer, S., Jacobi, C., Brüggemann, H., Lienard, T., Christmann, A., Bömeke, M.,

- Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *Journal of Molecular Microbiology and Biotechnology* 4(4), pp. 453–461.
- Dever, T.E., Dinman, J.D. and Green, R. 2018. Translation elongation and recoding in eukaryotes. *Cold Spring Harbor Perspectives in Biology* 10(8).
- Dickey, T.H., Altschuler, S.E. and Wuttke, D.S. 2013. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* 21(7), pp. 1074–1084.
- Dominski, Z. 2007. Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Critical Reviews in Biochemistry and Molecular Biology* 42(2), pp. 67–93.
- Dominski, Z., Carpousis, A.J. and Clouet-d'Orval, B. 2013. Emergence of the  $\beta$ -CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochimica et Biophysica Acta* 1829(6-7), pp. 532–551.
- Dominski, Z. and Marzluff, W.F. 2007. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* 396(2), pp. 373–390.
- Dominski, Z., Yang, X. and Marzluff, W.F. 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* 123(1), pp. 37–48.
- Doolittle, W.F. 2020. Evolution: two domains of life or three? *Current Biology* 30(4), pp. R177–R179.
- Drancourt, M., Nkamga, V.D., Lakhe, N.A., Régis, J.-M., Dufour, H., Fournier, P.-E., Bechah, Y., Scheld, W.M. and Raoult, D. 2017. Evidence of archaeal methanogens in brain abscess. *Clinical Infectious Diseases* 65(1), pp. 1–5.
- Driessen, R.P.C. and Dame, R.T. 2013. Structure and dynamics of the crenarchaeal nucleoid. *Biochemical Society Transactions* 41(1), pp. 321–325.
- Duggin, I.G., Dubarry, N. and Bell, S.D. 2011. Replication termination and chromosome dimer resolution in the archaeon *Sulfolobus solfataricus*. *The EMBO Journal* 30(1), pp. 145–153.
- Durand, S. and Condon, C. 2018. RNases and Helicases in Gram-Positive Bacteria. *Microbiology spectrum* 6(2).
- Eme, L. and Ettema, T.J.G. 2018. The eukaryotic ancestor shapes up. *Nature* 562(7727), pp. 352–353.
- Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. 2018. Archaea and the origin of eukaryotes. *Nature Reviews. Microbiology* 16(2), p. 120.
- Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. 2017. Archaea and the origin of eukaryotes. *Nature Reviews. Microbiology* 15(12), pp. 711–723.
- Erauso, G., Reysenbach, A.-L., Godfroy, A., Meunier, J.-R., Crump, B., Partensky, F., Baross, J., Marteinson, V., Barbier, G., Pace, N. and Prieur, D. 1993. *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Archives of Microbiology* 160(5).



- Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Bréchemmier-Baey, D. and Putzer, H. 2005. Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. *Nucleic Acids Research* 33(7), pp. 2141–2152.
- Evguenieva-Hackenberg, E., Gauernack, A.S., Hou, L. and Klug, G. 2020. Enzymatic analysis of reconstituted archaeal exosomes. *Methods in Molecular Biology* 2062, pp. 63–79.
- Evguenieva-Hackenberg, E., Gauernack, S. and Klug, G. 2017. The Archaeal Exosome: Degradation and Tailing at the 3'-End of RNA. In: Clouet-d'Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 115–128.
- Evguenieva-Hackenberg, E., Hou, L., Glaeser, S. and Klug, G. 2014. Structure and function of the archaeal exosome. *Wiley interdisciplinary reviews. RNA* 5(5), pp. 623–635.
- Evguenieva-Hackenberg, E. and Klug, G. 2011. New aspects of RNA processing in prokaryotes. *Current Opinion in Microbiology* 14(5), pp. 587–592.
- Evguenieva-Hackenberg, E. and Klug, G. 2009. RNA degradation in Archaea and Gram-negative bacteria different from *Escherichia coli*. *Prog Mol Biol Transl Sci* 85, pp. 275–317.
- Evguenieva-Hackenberg, E., Roppelt, V., Finsterseifer, P. and Klug, G. 2008. Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome. *Biochemistry* 47(50), pp. 13158–13168.
- Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F. and Klug, G. 2003. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Reports* 4(9), pp. 889–893.
- Fairman-Williams, M.E., Guenther, U.-P. and Jankowsky, E. 2010. SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* 20(3), pp. 313–324.
- Falk, S., Tants, J.-N., Basquin, J., Thoms, M., Hurt, E., Sattler, M. and Conti, E. 2017. Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA (New York)* 23(12), pp. 1780–1787.
- Farkas, J.A., Picking, J.W. and Santangelo, T.J. 2013. Genetic techniques for the archaea. *Annual Review of Genetics* 47, pp. 539–561.
- Ferreira-Cerca, S. 2017. Life and death of ribosomes in archaea. In: Clouet-d'Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 129–158.
- Forouzan, D., Ammelburg, M., Hobel, C.F., Ströh, L.J., Sessler, N., Martin, J. and Lupas, A.N. 2012. The archaeal proteasome is regulated by a network of AAA ATPases. *The Journal of Biological Chemistry* 287(46), pp. 39254–39262.
- Forterre, P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends in Genetics* 18(5), pp. 236–237.
- Forterre, P. 2013. The common ancestor of archaea and eukarya was not an archaeon. *Archaea* 2013, p. 372396.

- Fouqueau, T., Blombach, F. and Werner, F. 2017. Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annual Review of Microbiology* 71, pp. 331–348.
- Fouqueau, T. and Werner, F. 2017. The architecture of transcription elongation. *Science* 357(6354), pp. 871–872.
- Franzetti, B., Schoehn, G., Garcia, D., Ruigrok, R.W.H. and Zaccai, G. 2002. Characterization of the proteasome from the extremely halophilic archaeon Haloarcula marismortui. *Archaea* 1(1), pp. 53–61.
- French, S.L., Santangelo, T.J., Beyer, A.L. and Reeve, J.N. 2007. Transcription and translation are coupled in Archaea. *Molecular Biology and Evolution* 24(4), pp. 893–895.
- Frick, D.N. and Richardson, C.C. 2001. DNA primases. *Annual Review of Biochemistry* 70, pp. 39–80.
- Fujikane, R., Ishino, S., Ishino, Y. and Forterre, P. 2010. Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis. *Genes & Genetic Systems* 85(4), pp. 243–257.
- Fujikane, R., Komori, K., Shinagawa, H. and Ishino, Y. 2005. Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus. *The Journal of Biological Chemistry* 280(13), pp. 12351–12358.
- Fukuhara, H., Kifusa, M., Watanabe, M., Terada, A., Honda, T., Numata, T., Kakuta, Y. and Kimura, M. 2006. A fifth protein subunit Ph1496p elevates the optimum temperature for the ribonuclease P activity from Pyrococcus horikoshii OT3. *Biochemical and Biophysical Research Communications* 343(3), pp. 956–964.
- Gaci, N., Borrel, G., Tottey, W., O'Toole, P.W. and Brugère, J.-F. 2014. Archaea and the human gut: new beginning of an old story. *World Journal of Gastroenterology* 20(43), pp. 16062–16078.
- Gao, J. and Wang, J. 2012. Re-annotation of two hyperthermophilic archaea Pyrococcus abyssi GE5 and Pyrococcus furiosus DSM 3638. *Current Microbiology* 64(2), pp. 118–129.
- Gao, Y., Cui, Y., Fox, T., Lin, S., Wang, H., de Val, N., Zhou, Z.H. and Yang, W. 2019. Structures and operating principles of the replisome. *Science* 363(6429).
- Gao, Y. and Yang, W. 2019. Different mechanisms for translocation by monomeric and hexameric helicases. *Current Opinion in Structural Biology* 61, pp. 25–32.
- Gauernack, A.S., Lassek, C., Hou, L., Dzieciolowski, J., Evguenieva-Hackenberg, E. and Klug, G. 2017. Nop5 interacts with the archaeal RNA exosome. *FEBS Letters* 591(24), pp. 4039–4048.
- Gehring, A.M., Astling, D.P., Matsumi, R., Burkhart, B.W., Kelman, Z., Reeve, J.N., Jones, K.L. and Santangelo, T.J. 2017. Genome Replication in Thermococcus kodakarensis Independent of Cdc6 and an Origin of Replication. *Frontiers in microbiology* 8, p. 2084.
- Gerovac, M. and Vogel, J. 2019. An RNA surprise in bacterial effector mechanisms. *Cell Host & Microbe* 26(6), pp. 709–711.

- Giroux, X. and MacNeill, S.A. 2016. A novel archaeal DNA repair factor that acts with the UvrABC system to repair mitomycin C-induced DNA damage in a PCNA-dependent manner. *Molecular Microbiology* 99(1), pp. 1–14.
- Gloor, J.W., Balakrishnan, L., Campbell, J.L. and Bambara, R.A. 2012. Biochemical analyses indicate that binding and cleavage specificities define the ordered processing of human Okazaki fragments by Dna2 and FEN1. *Nucleic Acids Research* 40(14), pp. 6774–6786.
- Gomes-Filho, J.V., Daume, M. and Randau, L. 2018. Unique archaeal small rnas. *Annual Review of Genetics* 52, pp. 465–487.
- Gopalan, V. 2007. Uniformity amid diversity in RNase P. *Proceedings of the National Academy of Sciences of the United States of America* 104(7), pp. 2031–2032.
- Gorlas, A., Alain, K., Bienvenu, N. and Geslin, C. 2013. *Thermococcus prieurii* sp. nov., a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* 63(Pt 8), pp. 2920–2926.
- Grasso, S. and Tell, G. 2014. Base excision repair in Archaea: back to the future in DNA repair. *DNA Repair* 21, pp. 148–157.
- Greenough, L., Kelman, Z. and Gardner, A.F. 2015. The roles of family B and D DNA polymerases in *Thermococcus* species 9°N Okazaki fragment maturation. *The Journal of Biological Chemistry* 290(20), pp. 12514–12522.
- Gribaldo, S. and Brochier-Armanet, C. 2020. Evolutionary relationships between archaea and eukaryotes. *Nature Ecology & Evolution* 4(1), pp. 20–21.
- Grogan, D.W. 2015. Understanding DNA repair in hyperthermophilic archaea: persistent gaps and other reasons to focus on the fork. *Archaea* 2015, p. 942605.
- Grohmann, D. and Werner, F. 2011. Recent advances in the understanding of archaeal transcription. *Current Opinion in Microbiology* 14(3), pp. 328–334.
- Grosjean, H., de Crécy-Lagard, V. and Marck, C. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Letters* 584(2), pp. 252–264.
- Grosjean, H., Gaspin, C., Marck, C., Decatur, W.A. and de Crécy-Lagard, V. 2008. RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* 9, p. 470.
- Gupta, A. and Swati, D. 2019. Riboswitches in Archaea. *Combinatorial Chemistry & High Throughput Screening* 22(2), pp. 135–149.
- Guy, C.P. and Bolt, E.L. 2005. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Research* 33(11), pp. 3678–3690.
- Hajj, M., El-Hamaoui, S., Batista, M., Bouvier, M., Abdel-Razzak, Z., Clouet-d’Orval, B. and Chanieh, H. 2019. Archaeal helicases: unwinding in the extreme. In: Tutejac, R. ed. *Helicases from All Domains of Life*. Elsevier, pp. 1–13.

- Halbach, F., Reichelt, P., Rode, M. and Conti, E. 2013. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* 154(4), pp. 814–826.
- Halbach, F., Rode, M. and Conti, E. 2012. The crystal structure of *S. cerevisiae* Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA (New York)* 18(1), pp. 124–134.
- Hartmann, R.K., Gössringer, M., Späth, B., Fischer, S. and Marchfelder, A. 2009. The making of tRNAs and more - RNase P and tRNase Z. *Progress in molecular biology and translational science* 85, pp. 319–368.
- Hasenöhrl, D., Lombo, T., Kaberdin, V., Londei, P. and Bläsi, U. 2008. Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proceedings of the National Academy of Sciences of the United States of America* 105(6), pp. 2146–2150.
- Hausmann, S., Guimarães, V.A., Garcin, D., Baumann, N., Linder, P. and Redder, P. 2017. Both exo- and endo-nucleolytic activities of RNase J1 from *Staphylococcus aureus* are manganese dependent and active on triphosphorylated 5'-ends. *RNA Biology* 14(10), pp. 1431–1443.
- Heider, M.R., Burkhart, B.W., Santangelo, T.J. and Gardner, A.F. 2017. Defining the RNaseH2 enzyme-initiated ribonucleotide excision repair pathway in Archaea. *The Journal of Biological Chemistry* 292(21), pp. 8835–8845.
- Henneman, B., van Emmerik, C., van Ingen, H. and Dame, R.T. 2018. Structure and function of archaeal histones. *PLoS Genetics* 14(9), p. e1007582.
- Hennigan, A.N. and Reeve, J.N. 1994. mRNAs in the methanogenic archaeon *Methanococcus vannielii*: numbers, half-lives and processing. *Molecular Microbiology* 11(4), pp. 655–670.
- Higashibata, H., Siddiqui, M.A., Takagi, M., Imanaka, T. and Fujiwara, S. 2003. Surface histidine residue of archaeal histone affects DNA compaction and thermostability. *FEMS Microbiology Letters* 224(1), pp. 17–22.
- Hirata, A., Fujishima, K., Yamagami, R., Kawamura, T., Banfield, J.F., Kanai, A. and Hori, H. 2012. X-ray structure of the fourth type of archaeal tRNA splicing endonuclease: insights into the evolution of a novel three-unit composition and a unique loop involved in broad substrate specificity. *Nucleic Acids Research* 40(20), pp. 10554–10566.
- Hirata, A., Klein, B.J. and Murakami, K.S. 2008. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451(7180), pp. 851–854.
- Hirtreiter, A., Grohmann, D. and Werner, F. 2010. Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Research* 38(2), pp. 585–596.
- Hodeib, S., Raj, S., Manosas, M., Zhang, W., Bagchi, D., Ducos, B., Fiorini, F., Kanaan, J., Le Hir, H., Allemand, J.-F., Bensimon, D. and Croquette, V. 2017. A mechanistic study of helicases with magnetic traps. *Protein Science* 26(7), pp. 1314–1336.
- Holub, P., Lalakova, J., Cerna, H., Pasulka, J., Sarazova, M., Hrazdilova, K., Arce, M.S., Hobor, F., Stefl, R. and Vanacova, S. 2012. Air2p is critical for the assembly and RNA-binding of the TRAMP

complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Research* 40(12), pp. 5679–5693.

Hotto, A.M., Stern, D.B. and Schuster, G. 2020. Plant ribonuclease J: an essential player in maintaining chloroplast RNA quality control for gene expression. *Plants* 9(3).

Hou, L., Klug, G. and Evguenieva-Hackenberg, E. 2013. The archaeal DnaG protein needs Csl4 for binding to the exosome and enhances its interaction with adenine-rich RNAs. *RNA Biology* 10(3), pp. 415–424.

Houri-Zeevi, L. and Rechavi, O. 2017. A matter of time: small rnas regulate the duration of epigenetic inheritance. *Trends in Genetics* 33(1), pp. 46–57.

Howard, M.J., Liu, X., Lim, W.H., Klemm, B.P., Fierke, C.A., Koutmos, M. and Engelke, D.R. 2013. RNase P enzymes: divergent scaffolds for a conserved biological reaction. *RNA Biology* 10(6), pp. 909–914.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., HERNSDORF, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C. and Banfield, J.F. 2016. A new view of the tree of life. *Nature Microbiology* 1, p. 16048.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H. and Takai, K. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577(7791), pp. 519–525.

Ito, K., Honda, T., Suzuki, T., Miyoshi, T., Murakami, R., Yao, M. and Uchiumi, T. 2014. Molecular insights into the interaction of the ribosomal stalk protein with elongation factor 1 $\alpha$ . *Nucleic Acids Research* 42(22), pp. 14042–14052.

Itoh, T., Nomura, N. and Sako, Y. 2003. Distribution of 16S rRNA introns among the family Thermoproteaceae and their evolutionary implications. *Extremophiles: Life Under Extreme Conditions* 7(3), pp. 229–233.

Jäger, A., Samorski, R., Pfeifer, F. and Klug, G. 2002. Individual gvp transcript segments in *Haloferax mediterranei* exhibit varying half-lives, which are differentially affected by salt concentration and growth phase. *Nucleic Acids Research* 30(24), pp. 5436–5443.

Jankowsky, E. and Fairman, M.E. 2007. RNA helicases--one fold for many functions. *Current Opinion in Structural Biology* 17(3), pp. 316–324.

Jarmoskaite, I. and Russell, R. 2014. RNA helicase proteins as chaperones and remodelers. *Annual Review of Biochemistry* 83, pp. 697–725.

Johnson, S.J. and Jackson, R.N. 2013. Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biology* 10(1), pp. 33–43.

Julius, C. and Yuzenkova, Y. 2019. Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. *Wiley interdisciplinary reviews. RNA* 10(2), p. e1512.

- Kampmann, M. and Stock, D. 2004. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucleic Acids Research* 32(12), pp. 3537–3545.
- Kanai, A. 2014. Welcome to the new tRNA world! *Frontiers in genetics* 5, p. 336.
- Kandler, O. and Hippe, H. 1977. Lack of peptidoglycan in the cell walls of *Methanosarcina barkeri*. *Archives of Microbiology* 113(1-2), pp. 57–60.
- Karr, E.A., Isom, C.E., Trinh, V. and Peeters, E. 2017. Transcription Factor-Mediated Gene Regulation in Archaea. In: Clouet-d’Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 27–69.
- Kellner, S., Spang, A., Offre, P., Szöllösi, G.J., Petitjean, C. and Williams, T.A. 2018. Genome size evolution in the Archaea. *Emerging Topics in Life Sciences* 2(4), p. ETLS20180021.
- Kelly, T.J., Simancek, P. and Brush, G.S. 1998. Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*. *Proceedings of the National Academy of Sciences of the United States of America* 95(25), pp. 14634–14639.
- Kelman, L.M. and Kelman, Z. 2014. Archaeal DNA replication. *Annual Review of Genetics* 48(1), pp. 71–97.
- Kelman, L.M. and Kelman, Z. 2018. Do archaea need an origin of replication? *Trends in Microbiology* 26(3), pp. 172–174.
- Kelman, L.M., O’Dell, W.B. and Kelman, Z. 2020. Unwinding 20 years of the archaeal minichromosome maintenance helicase. *Journal of Bacteriology* 202(6).
- Kelman, Z., Lee, J.K. and Hurwitz, J. 1999. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DeltaH contains DNA helicase activity. *Proceedings of the National Academy of Sciences of the United States of America* 96(26), pp. 14783–14788.
- Khemic, V. and Linder, P. 2018. RNA helicases in RNA decay. *Biochemical Society Transactions* 46(1), pp. 163–172.
- Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeá, E., Greenblatt, J.F. and Buratowski, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432(7016), pp. 517–522.
- Kiyonari, S., Tahara, S., Shirai, T., Iwai, S., Ishino, S. and Ishino, Y. 2009. Biochemical properties and base excision repair complex formation of apurinic/apyrimidinic endonuclease from *Pyrococcus furiosus*. *Nucleic Acids Research* 37(19), pp. 6439–6453.
- Kiyonari, S., Uchimura, M., Shirai, T. and Ishino, Y. 2008. Physical and functional interactions between uracil-DNA glycosylase and proliferating cell nuclear antigen from the euryarchaeon *Pyrococcus furiosus*. *The Journal of Biological Chemistry* 283(35), pp. 24185–24193.
- Klemm, B.P., Wu, N., Chen, Y., Liu, X., Kaitany, K.J., Howard, M.J. and Fierke, C.A. 2016. The Diversity of Ribonuclease P: Protein and RNA Catalysts with Analogous Biological Functions. *Biomolecules* 6(2).

- Klinger, C.M., Spang, A., Dacks, J.B. and Ettema, T.J.G. 2016. Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Molecular Biology and Evolution* 33(6), pp. 1528–1541.
- Klug, G., Evguenieva-Hackenberg, E., Omer, A., Dennis, P.P. and Marchfelder, A. 2007. RNA processing. In: Cavicchioli, R. ed. *Archaea: Molecular and Cellular Biology*. ASM Press, Washington D.C. pp. 158–174.
- Kobayashi, K., Saito, K., Ishitani, R., Ito, K. and Nureki, O. 2012. Structural basis for translation termination by archaeal RF1 and GTP-bound EF1 $\alpha$  complex. *Nucleic Acids Research* 40(18), pp. 9319–9328.
- Kochiwa, H., Tomita, M. and Kanai, A. 2007. Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evolutionary Biology* 7, p. 128.
- Kohler, P.R.A. and Metcalf, W.W. 2012. Genetic manipulation of *Methanosarcina* spp. *Frontiers in microbiology* 3, p. 259.
- Komori, K. and Ishino, Y. 2001. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *The Journal of Biological Chemistry* 276(28), pp. 25654–25660.
- Koonin, E.V., Makarova, K.S. and Wolf, Y.I. 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annual Review of Microbiology* 71, pp. 233–261.
- Koonin, E.V. and Wolf, Y.I. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Research* 36(21), pp. 6688–6719.
- Koskinen, K., Pausan, M.R., Perras, A.K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R. and Moissl-Eichinger, C. 2017. First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. *mBio* 8(6).
- Kuehner, J.N., Pearson, E.L. and Moore, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews. Molecular Cell Biology* 12(5), pp. 283–294.
- Kuhn, J.F., Tran, E.J. and Maxwell, E.S. 2002. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Research* 30(4), pp. 931–941.
- Kyrpides, N.C. and Woese, C.R. 1998. Universally conserved translation initiation factors. *Proceedings of the National Academy of Sciences of the United States of America* 95(1), pp. 224–228.
- Li de la Sierra-Gallay, I., Zig, L., Jamalli, A. and Putzer, H. 2008. Structural insights into the dual activity of RNase J. *Nature Structural & Molecular Biology* 15(2), pp. 206–212.
- Lai, S.M., Lai, L.B., Foster, M.P. and Gopalan, V. 2014. The L7Ae protein binds to two kink-turns in the *Pyrococcus furiosus* RNase P RNA. *Nucleic Acids Research* 42(21), pp. 13328–13338.
- Langworthy, T.A., Smith, P.F. and Mayberry, W.R. 1972. Lipids of *Thermoplasma acidophilum*. *Journal of Bacteriology* 112(3), pp. 1193–1200.



- Lebreton, A. and Séraphin, B. 2008. Exosome-mediated quality control: substrate recruitment and molecular activity. *Biochimica et Biophysica Acta* 1779(9), pp. 558–565.
- Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D. and Poch, O. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Research* 30(24), pp. 5382–5390.
- Leigh, J.A., Albers, S.-V., Atomi, H. and Allers, T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiology Reviews* 35(4), pp. 577–608.
- Lemmens, L., Maklad, H.R., Bervoets, I. and Peeters, E. 2019. Transcription Regulators in Archaea: Homologies and Differences with Bacterial Regulators. *Journal of Molecular Biology* 431(20), pp. 4132–4146.
- Levy, S., Portnoy, V., Admon, J. and Schuster, G. 2011. Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biology* 8(6), pp. 1073–1083.
- Li, Z., Lu, S., Hou, G., Ma, X., Sheng, D., Ni, J. and Shen, Y. 2008. Hjm/Hel308A DNA helicase from *Sulfolobus tokodaii* promotes replication fork regression and interacts with Hjc endonuclease in vitro. *Journal of Bacteriology* 190(8), pp. 3006–3017.
- Liang, X.-H., Liu, Q. and Fournier, M.J. 2009. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA (New York)* 15(9), pp. 1716–1728.
- Lin-Chao, S., Chiou, N.-T. and Schuster, G. 2007. The PNPase, exosome and RNA helicases as the building components of evolutionarily-conserved RNA degradation machines. *Journal of Biomedical Science* 14(4), pp. 523–532.
- Linder, P. and Jankowsky, E. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nature Reviews. Molecular Cell Biology* 12(8), pp. 505–516.
- Linder, P., Lemeille, S. and Redder, P. 2014. Transcriptome-wide analyses of 5'-ends in RNase J mutants of a gram-positive pathogen reveal a role in RNA maturation, regulation and degradation. *PLoS Genetics* 10(2), p. e1004207.
- Londei, P. 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiology Reviews* 29(2), pp. 185–200.
- Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G. and Jensen, T.H. 2015. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell reports* 10(2), pp. 178–192.
- Lundgren, M., Andersson, A., Chen, L., Nilsson, P. and Bernander, R. 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proceedings of the National Academy of Sciences of the United States of America* 101(18), pp. 7046–7051.
- Lurie-Weinberger, M.N. and Gophna, U. 2015. Archaea in and on the human body: health implications and future directions. *PLoS Pathogens* 11(6), p. e1004833.

- MacNeill, S.A. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *The Biochemical Journal* 425(3), pp. 489–500.
- Maier, L.-K. and Marchfelder, A. 2019. It's all about the T: transcription termination in archaea. *Biochemical Society Transactions* 47(1), pp. 461–468.
- Majumder, P. and Baumeister, W. 2019. Proteasomes: unfoldase-assisted protein degradation machines. *Biological Chemistry* 401(1), pp. 183–199.
- Majumder, P., Rudack, T., Beck, F., Danev, R., Pfeifer, G., Nagy, I. and Baumeister, W. 2018. Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proceedings of the National Academy of Sciences of the United States of America*.
- Makarova, K.S., Koonin, E.V. and Kelman, Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biology Direct* 7, p. 7.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R. and Koonin, E.V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews. Microbiology* 13(11), pp. 722–736.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksny, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R. and Koonin, E.V. 2020. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews. Microbiology* 18(2), pp. 67–83.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L. and Tong, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444(7121), pp. 953–956.
- Mao, H., White, S.A. and Williamson, J.R. 1999. A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nature Structural Biology* 6(12), pp. 1139–1147.
- Marck, C. and Grosjean, H. 2003. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA (New York)* 9(12), pp. 1516–1531.
- Marck, C. and Grosjean, H. 2002. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA (New York)* 8(10), pp. 1189–1232.
- Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V. and Bell, S.D. 2006. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Reports* 7(5), pp. 539–545.
- Märtens, B., Amman, F., Manoharadas, S., Zeichen, L., Orell, A., Albers, S.-V., Hofacker, I. and Bläsi, U. 2013. Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *Plos One* 8(10), p. e76569.

Maruyama, H., Harwood, J.C., Moore, K.M., Paszkiewicz, K., Durley, S.C., Fukushima, H., Atomi, H., Takeyasu, K. and Kent, N.A. 2013. An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Reports* 14(8), pp. 711–717.

Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T. and Condon, C. 2007. 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 129(4), pp. 681–692.

Matos, R.G., López-Viñas, E., Gómez-Puertas, P. and Arraiano, C.M. 2012. The only exoribonuclease present in *Haloferax volcanii* has an unique response to temperature changes. *Biochimica et Biophysica Acta* 1820(10), pp. 1543–1552.

Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. 2003. Identification of short “eukaryotic” Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Reports* 4(2), pp. 154–158.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A. and Jensen, T.H. 2016. Identification of a nuclear exosome decay pathway for processed transcripts. *Molecular Cell* 64(3), pp. 520–533.

Mihajlovski, A., Doré, J., Levenez, F., Alric, M. and Brugère, J.-F. 2010. Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-associated increase of the diversity. *Environmental microbiology reports* 2(2), pp. 272–280.

Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M. and Tollervey, D. 2003. Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Molecular and Cellular Biology* 23(19), pp. 6982–6992.

Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C. and Schmitz, R.A. 2018. Archaea are interactive components of complex microbiomes. *Trends in Microbiology* 26(1), pp. 70–85.

Morriscal, S.W. 2015. DNA-pairing and annealing processes in homologous recombination and homology-directed repair. *Cold Spring Harbor Perspectives in Biology* 7(2), p. a016444.

De Mot, R., Nagy, I. and Baumeister, W. 1998. A self-compartmentalizing protease in *Rhodococcus*: the 20S proteasome. *Antonie Van Leeuwenhoek* 74(1-3), pp. 83–87.

Motorin, Y. and Helm, M. 2010. tRNA stabilization by modified nucleotides. *Biochemistry* 49(24), pp. 4934–4944.

Mura, C., Randolph, P.S., Patterson, J. and Cozen, A.E. 2013. Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biology* 10(4), pp. 636–651.

Nagaoka, E., Hidese, R., Imanaka, T. and Fujiwara, S. 2013. Importance and determinants of induction of cold-induced DEAD RNA helicase in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Journal of Bacteriology* 195(15), pp. 3442–3450.

Nagata, M., Ishino, S., Yamagami, T., Simons, J.-R., Kanai, T., Atomi, H. and Ishino, Y. 2017. Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef. *Scientific Reports* 7(1), p. 16949.

- Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.-P., Reeve, J.N., Santangelo, T.J. and Fondufe-Mittendorf, Y.N. 2013. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics* 14, p. 391.
- Narrowe, A.B., Spang, A., Stairs, C.W., Caceres, E.F., Baker, B.J., Miller, C.S. and Ettema, T.J.G. 2018. Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in archaea and parabasalids. *Genome Biology and Evolution* 10(9), pp. 2380–2393.
- Navon, A. and Goldberg, A.L. 2001. Proteins are unfolded on the surface of the ATPase ring before transport into the proteasome. *Molecular Cell* 8(6), pp. 1339–1349.
- Nielsen, J.S., Bøggild, A., Andersen, C.B.F., Nielsen, G., Boysen, A., Brodersen, D.E. and Valentin-Hansen, P. 2007. An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. *RNA (New York)* 13(12), pp. 2213–2223.
- Nomura, N., Morinaga, Y., Kogishi, T., Kim, E.-J., Sako, Y. and Uchida, A. 2002. Heterogeneous yet similar introns reside in identical positions of the rRNA genes in natural isolates of the archaeon *Aeropyrum pernix*. *Gene* 295(1), pp. 43–50.
- Noon, K.R., Bruenger, E. and McCloskey, J.A. 1998. Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *Journal of Bacteriology* 180(11), pp. 2883–2888.
- Ogami, K., Chen, Y. and Manley, J.L. 2018. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Non-coding RNA* 4(1).
- Olsen, G.J. and Woese, C.R. 1993. Ribosomal RNA: a key to phylogeny. *The FASEB Journal* 7(1), pp. 113–123.
- Omer, A.D., Ziesche, S., Ebhardt, H. and Dennis, P.P. 2002. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proceedings of the National Academy of Sciences of the United States of America* 99(8), pp. 5289–5294.
- Oxley, A.P.A., Lanfranconi, M.P., Würdemann, D., Ott, S., Schreiber, S., McGenity, T.J., Timmis, K.N. and Nogales, B. 2010. Halophilic archaea in the human intestinal mucosa. *Environmental Microbiology* 12(9), pp. 2398–2410.
- Oyama, T., Ishino, S., Shirai, T., Yamagami, T., Nagata, M., Ogino, H., Kusunoki, M. and Ishino, Y. 2016. Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Research* 44(19), pp. 9505–9517.
- Pan, M., Kelman, L.M. and Kelman, Z. 2011. The archaeal PCNA proteins. *Biochemical Society Transactions* 39(1), pp. 20–24.
- Patrick, E.M., Srinivasan, S., Jankowsky, E. and Comstock, M.J. 2017. The RNA helicase Mtr4p is a duplex-sensing translocase. *Nature Chemical Biology* 13(1), pp. 99–104.
- Peeters, E. and Charlier, D. 2010. The Lrp family of transcription regulators in archaea. *Archaea* 2010, p. 750457.

- Peeters, E., Driessen, R.P.C., Werner, F. and Dame, R.T. 2015. The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews. Microbiology* 13(6), pp. 333–341.
- Peterson, J.R., Thor, S., Kohler, L., Kohler, P.R.A., Metcalf, W.W. and Luthey-Schulten, Z. 2016. Genome-wide gene expression and RNA half-life measurements allow predictions of regulation and metabolic behavior in *Methanosarcina acetivorans*. *BMC Genomics* 17(1), p. 924.
- Phillips, G., El Yacoubi, B., Lyons, B., Alvarez, S., Iwata-Reuyl, D. and de Crécy-Lagard, V. 2008. Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP cyclohydrolase I. *Journal of Bacteriology* 190(24), pp. 7876–7884.
- Phung, D.K., Bouvier, M. and Clouet-d'Orval, B. 2017. An overview of ribonuclease repertoire and RNA processing pathways in archaea. In: Clouet-d'Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 89–114.
- Phung, D.K., Etienne, C., Batista, M., Langendijk-Genevaux, P., Moalic, Y., Laurent, S., Liuu, S., Morales, V., Jebbar, M., Fichant, G., Bouvier, M., Flament, D. and Clouet-d'Orval, B. 2020. RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the  $\beta$ -CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic RNA exosome machinery. *Nucleic Acids Research* 48(7), pp. 3832–3847.
- Phung, D.K., Rinaldi, D., Langendijk-Genevaux, P.S., Quentin, Y., Carpousis, A.J. and Clouet-d'Orval, B. 2013. Archaeal  $\beta$ -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Research* 41(2), pp. 1091–1103.
- Phung, D.K. Thèse : Métabolisme de l'ARN chez les Archées : Identification et caractérisation du complexe Ribonucléase  $\beta$ -CASP/hélicase Ski2-like de *Pyrococcus abyssi*. 2017
- Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. 2008. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Research* 36(Database issue), pp. D178–83.
- Pluchon, P.-F., Fouqueau, T., Crezé, C., Laurent, S., Briffotiaux, J., Hogrel, G., Palud, A., Henneke, G., Godfroy, A., Hausner, W., Thomm, M., Nicolas, J. and Flament, D. 2013. An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs. *Plos One* 8(11), p. e79707.
- Polikanov, Y.S., Melnikov, S.V., Söll, D. and Steitz, T.A. 2015. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nature Structural & Molecular Biology* 22(4), pp. 342–344.
- Porrua, O., Boudvillain, M. and Libri, D. 2016. Transcription termination: variations on common themes. *Trends in Genetics* 32(8), pp. 508–522.
- Portnoy, V. and Schuster, G. 2006. RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Research* 34(20), pp. 5923–5931.
- Puno, M.R., Weick, E.-M., Das, M. and Lima, C.D. 2019. Snapshot: the RNA exosome. *Cell* 179(1), pp. 282–282.e1.

- Pyle, A.M. 2008. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual review of biophysics* 37, pp. 317–336.
- Qi, L., Li, J., Jia, J., Yue, L. and Dong, X. 2020. Comprehensive analysis of the pre-ribosomal RNA maturation pathway in a methanoarchaeon exposes the conserved circularization and linearization mode in archaea. *RNA Biology*.
- Ragan, M.A., Logsdon, J.M., Sensen, C.W., Charlebois, R.L. and Doolittle, W.F. 1996. An archaeobacterial homolog of pelota, a meiotic cell division protein in eukaryotes. *FEMS Microbiology Letters* 144(2-3), pp. 151–155.
- Randau, L., Calvin, K., Hall, M., Yuan, J., Podar, M., Li, H. and Söll, D. 2005. The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proceedings of the National Academy of Sciences of the United States of America* 102(50), pp. 17934–17939.
- Raymann, K., Forterre, P., Brochier-Armanet, C. and Gribaldo, S. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biology and Evolution* 6(1), pp. 192–212.
- Richards, J.D., Johnson, K.A., Liu, H., McRobbie, A.-M., McMahon, S., Oke, M., Carter, L., Naismith, J.H. and White, M.F. 2008. Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *The Journal of Biological Chemistry* 283(8), pp. 5118–5126.
- Roberts, J.A., Bell, S.D. and White, M.F. 2003. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Molecular Microbiology* 48(2), pp. 361–371.
- Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. and Bell, S.D. 2004. Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. *Cell* 116(1), pp. 25–38.
- Rodnina, M.V. 2018. Translation in Prokaryotes. *Cold Spring Harbor Perspectives in Biology* 10(9).
- Roppelt, V., Klug, G. and Evguenieva-Hackenberg, E. 2010. The evolutionarily conserved subunits Rrp4 and Csl4 confer different substrate specificities to the archaeal exosome. *FEBS Letters* 584(13), pp. 2931–2936.
- Rouillon, C. and White, M.F. 2011. The evolution and mechanisms of nucleotide excision repair proteins. *Research in Microbiology* 162(1), pp. 19–26.
- Sakakibara, N., Kelman, L.M. and Kelman, Z. 2009. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochemical Society Transactions* 37(Pt 1), pp. 7–11.
- Samanta, M.P., Lai, S.M., Daniels, C.J. and Gopalan, V. 2016. Sequence analysis and comparative study of the protein subunits of archaeal rnaase P. *Biomolecules* 6(2).
- Sanders, T.J., Wenck, B.R., Selan, J.N., Barker, M.P., Trimmer, S.A., Walker, J.E. and Santangelo, T.J. 2020. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nature Microbiology* 5(4), pp. 545–553.
- Santangelo, T.J., Cubonová, L., Skinner, K.M. and Reeve, J.N. 2009. Archaeal intrinsic transcription termination in vivo. *Journal of Bacteriology* 191(22), pp. 7102–7108.

- Santangelo, T.J. and Reeve, J.N. 2006. Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *Journal of Molecular Biology* 355(2), pp. 196–210.
- Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M. and van Hoof, A. 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature Structural & Molecular Biology* 16(1), pp. 56–62.
- Schleper, C. and Sousa, F.L. 2020. Meet the relatives of our cellular ancestor. *Nature* 577(7791), pp. 478–479.
- Schmidt, C., Kowalinski, E., Shanmuganathan, V., Defenouillère, Q., Braunger, K., Heuer, A., Pech, M., Namane, A., Berninghausen, O., Fromont-Racine, M., Jacquier, A., Conti, E., Becker, T. and Beckmann, R. 2016. The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. *Science* 354(6318), pp. 1431–1433.
- Schmitt, E., Coureux, P.-D., Monestier, A., Dubiez, E. and Mechulam, Y. 2019. Start codon recognition in eukaryotic and archaeal translation initiation: A common structural core. *International Journal of Molecular Sciences* 20(4).
- Schuller, J.M., Falk, S., Fromm, L., Hurt, E. and Conti, E. 2018. Structure of the nuclear exosome captured on a maturing preribosome. *Science* 360(6385), pp. 219–222.
- Sharwood, R.E., Halpert, M., Luro, S., Schuster, G. and Stern, D.B. 2011. Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA (New York)* 17(12), pp. 2165–2176.
- Shi, Z., Yang, W.-Z., Lin-Chao, S., Chak, K.-F. and Yuan, H.S. 2008. Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA (New York)* 14(11), pp. 2361–2371.
- Sloan, K.E. and Bohnsack, M.T. 2018. Unravelling the mechanisms of RNA helicase regulation. *Trends in Biochemical Sciences* 43(4), pp. 237–250.
- Slomovic, S. and Schuster, G. 2008. Stable PNPase RNAi silencing: its effect on the processing and adenylation of human mitochondrial RNA. *RNA (New York)* 14(2), pp. 310–323.
- Smollett, K., Blombach, F., Fouqueau, T. and Werner, F. 2017. A global characterisation of the archaeal transcription machinery. In: Clouet-d'Orval, B. ed. *RNA metabolism and gene expression in archaea*. Nucleic acids and molecular biology. Cham: Springer International Publishing, pp. 1–26.
- Spang, A., Caceres, E.F. and Ettema, T.J.G. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 357(6351).
- Spang, A., Eme, L., Saw, J.H., Caceres, E.F., Zaremba-Niedzwiedzka, K., Lombard, J., Guy, L. and Ettema, T.J.G. 2018. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genetics* 14(3), p. e1007080.



- Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T.J.G. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551), pp. 173–179.
- Spitalny, P. and Thomm, M. 2008. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Molecular Microbiology* 67(5), pp. 958–970.
- Stoecklin, G. and Mühlemann, O. 2013. RNA decay mechanisms: specificity through diversity. *Biochimica et Biophysica Acta* 1829(6-7), pp. 487–490.
- Tadokoro, T. and Kanaya, S. 2009. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *The FEBS Journal* 276(6), pp. 1482–1493.
- Tafel, A.A., Wu, L. and McHugh, P.J. 2011. Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *The Journal of Biological Chemistry* 286(18), pp. 15832–15840.
- Takemata, N. and Bell, S.D. 2020. Emerging views of genome organization in Archaea. *Journal of Cell Science* 133(10).
- Takemata, N., Samson, R.Y. and Bell, S.D. 2019. Physical and functional compartmentalization of archaeal chromosomes. *Cell* 179(1), pp. 165–179.e18.
- Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Hüttenhofer, A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proceedings of the National Academy of Sciences of the United States of America* 99(11), pp. 7536–7541.
- Tanner, N.K. and Linder, P. 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Molecular Cell* 8(2), pp. 251–262.
- Tanzawa, T., Kato, K., Girodat, D., Ose, T., Kumakura, Y., Wieden, H.-J., Uchiumi, T., Tanaka, I. and Yao, M. 2018. The C-terminal helix of ribosomal P stalk recognizes a hydrophobic groove of elongation factor 2 in a novel fashion. *Nucleic Acids Research* 46(6), pp. 3232–3244.
- Tesina, P., Heckel, E., Cheng, J., Fromont-Racine, M., Buschauer, R., Kater, L., Beatrix, B., Berninghausen, O., Jacquier, A., Becker, T. and Beckmann, R. 2019. Structure of the 80S ribosome-Xrn1 nuclease complex. *Nature Structural & Molecular Biology* 26(4), pp. 275–280.
- Thiel, A., Michoud, G., Moalic, Y., Flament, D. and Jebbar, M. 2014. Genetic manipulations of the hyperthermophilic piezophilic archaeon *Thermococcus barophilus*. *Applied and Environmental Microbiology* 80(7), pp. 2299–2306.
- Thomm, M., Hausner, W. and Hethke, C. 1993. Transcription factors and termination of transcription in methanococcus. *Systematic and Applied Microbiology* 16(4), pp. 648–655.
- Thompson, L.D. and Daniels, C.J. 1988. A tRNA(Trp) intron endonuclease from *Halobacterium volcanii*. Unique substrate recognition properties. *The Journal of Biological Chemistry* 263(34), pp. 17951–17959.

- Törö, I., Thore, S., Mayer, C., Basquin, J., Séraphin, B. and Suck, D. 2001. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *The EMBO Journal* 20(9), pp. 2293–2303.
- Tourte, M., Schaeffer, P., Grossi, V. and Oger, P.M. 2020. Functionalized membrane domains: an ancestral feature of archaea? *Frontiers in microbiology* 11, p. 526.
- Twomey, E.C., Ji, Z., Wales, T.E., Bodnar, N.O., Ficarro, S.B., Marto, J.A., Engen, J.R. and Rapoport, T.A. 2019. Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science* 365(6452).
- Valdez, B.C., Henning, D., Perumal, K. and Busch, H. 1997. RNA-unwinding and RNA-folding activities of RNA helicase II/Gu--two activities in separate domains of the same protein. *European Journal of Biochemistry / FEBS* 250(3), pp. 800–807.
- Veith, T., Martin, R., Wurm, J.P., Weis, B.L., Duchardt-Ferner, E., Saffertal, C., Hennig, R., Mirus, O., Bohnsack, M.T., Wöhnert, J. and Schleiff, E. 2012. Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic Acids Research* 40(7), pp. 3259–3274.
- Vester, K., Santos, K.F., Kuroпка, B., Weise, C. and Wahl, M.C. 2020. The inactive C-terminal cassette of the dual-cassette RNA helicase BRR2 both stimulates and inhibits the activity of the N-terminal helicase unit. *The Journal of Biological Chemistry* 295(7), pp. 2097–2112.
- Vitagliano, L., Masullo, M., Sica, F., Zagari, A. and Bocchini, V. 2001. The crystal structure of *Sulfolobus solfataricus* elongation factor 1alpha in complex with GDP reveals novel features in nucleotide binding and exchange. *The EMBO Journal* 20(19), pp. 5305–5311.
- Vogel, A., Schilling, O., Späth, B. and Marchfelder, A. 2005. The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biological Chemistry* 386(12), pp. 1253–1264.
- Vogel, J. and Luisi, B.F. 2011. Hfq and its constellation of RNA. *Nature Reviews. Microbiology* 9(8), pp. 578–589.
- Wadsworth, R.I. and White, M.F. 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Research* 29(4), pp. 914–920.
- Wagner, E.G.H. and Romby, P. 2015. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Advances in genetics* 90, pp. 133–208.
- Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B.H., Reimann, J. and Albers, S.-V. 2012. Versatile Genetic Tool Box for the Crenarchaeote *Sulfolobus acidocaldarius*. *Frontiers in microbiology* 3, p. 214.
- Walker, J.E., Luyties, O. and Santangelo, T.J. 2017. Factor-dependent archaeal transcription termination. *Proceedings of the National Academy of Sciences of the United States of America* 114(33), pp. E6767–E6773.
- Wang, H., Peng, N., Shah, S.A., Huang, L. and She, Q. 2015. Archaeal extrachromosomal genetic elements. *Microbiology and Molecular Biology Reviews* 79(1), pp. 117–152.

- Wasmuth, E.V., Zinder, J.C., Zattas, D., Das, M. and Lima, C.D. 2017. Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *eLife* 6.
- Weick, E.-M., Puno, M.R., Januszyk, K., Zinder, J.C., DiMattia, M.A. and Lima, C.D. 2018. Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. *Cell* 173(7), pp. 1663–1677.e21.
- Weir, J.R., Bonneau, F., Hentschel, J. and Conti, E. 2010. Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proceedings of the National Academy of Sciences of the United States of America* 107(27), pp. 12139–12144.
- Werner, F. 2013. Molecular mechanisms of transcription elongation in archaea. *Chemical Reviews* 113(11), pp. 8331–8349.
- Werner, F. 2007. Structure and function of archaeal RNA polymerases. *Molecular Microbiology* 65(6), pp. 1395–1404.
- Werner, F. and Grohmann, D. 2011. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews. Microbiology* 9(2), pp. 85–98.
- West, S., Gromak, N. and Proudfoot, N.J. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432(7016), pp. 522–525.
- White, M.F. and Allers, T. 2018. DNA repair in the archaea-an emerging picture. *FEMS Microbiology Reviews* 42(4), pp. 514–526.
- Williams, T.A., Cox, C.J., Foster, P.G., Szöllősi, G.J. and Embley, T.M. 2020. Phylogenomics provides robust support for a two-domains tree of life. *Nature Ecology & Evolution* 4(1), pp. 138–147.
- Wilusz, C.J. and Wilusz, J. 2013. Lsm proteins and Hfq: Life at the 3' end. *RNA Biology* 10(4), pp. 592–601.
- Wilusz, J.E., Whipple, J.M., Phizicky, E.M. and Sharp, P.A. 2011. tRNAs marked with CCACCA are targeted for degradation. *Science* 334(6057), pp. 817–821.
- Witharana, C., Roppelt, V., Lochnit, G., Klug, G. and Evguenieva-Hackenberg, E. 2012. Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. *Biochimie* 94(7), pp. 1578–1587.
- Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74(11), pp. 5088–5090.
- Woese, C.R., Kandler, O. and Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* 87(12), pp. 4576–4579.

- Wong, I. and Lohman, T.M. 1993. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proceedings of the National Academy of Sciences of the United States of America* 90(12), pp. 5428–5432.
- Wu, Y. 2012. Unwinding and rewinding: double faces of helicase? *Journal of nucleic acids* 2012, p. 140601.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D. and Jacquier, A. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121(5), pp. 725–737.
- Yang, Q. and Jankowsky, E. 2005. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* 44(41), pp. 13591–13601.
- Yang, X., Garnier, F., Débat, H., Strick, T.R. and Nadal, M. 2020. Direct observation of helicase-topoisomerase coupling within reverse gyrase. *Proceedings of the National Academy of Sciences of the United States of America* 117(20), pp. 10856–10864.
- Yao, N.Y. and O'Donnell, M.E. 2016. Evolution of replication machines. *Critical Reviews in Biochemistry and Molecular Biology* 51(3), pp. 135–149.
- Yip, W.S.V., Vincent, N.G. and Baserga, S.J. 2013. Ribonucleoproteins in archaeal pre-rRNA processing and modification. *Archaea* 2013, p. 614735.
- Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A. and Ettema, T.J.G. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637), pp. 353–358.
- Zatopek, K.M., Gardner, A.F. and Kelman, Z. 2018. Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms. *FEMS Microbiology Reviews* 42(4), pp. 477–488.
- Zhai, B., DuPrez, K., Han, X., Yuan, Z., Ahmad, S., Xu, C., Gu, L., Ni, J., Fan, L. and Shen, Y. 2018. The archaeal ATPase PINA interacts with the helicase Hjm via its carboxyl terminal KH domain remodeling and processing replication fork and Holliday junction. *Nucleic Acids Research* 46(13), pp. 6627–6641.
- Zhang, L., Xu, T., Maeder, C., Bud, L.-O., Shanks, J., Nix, J., Guthrie, C., Pleiss, J.A. and Zhao, R. 2009. Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nature Structural & Molecular Biology* 16(7), pp. 731–739.
- Zhang, X., Nakashima, T., Kakuta, Y., Yao, M., Tanaka, I. and Kimura, M. 2008. Crystal structure of an archaeal Ski2p-like protein from *Pyrococcus horikoshii* OT3. *Protein Science* 17(1), pp. 136–145.
- Zheng, X., Feng, N., Li, D., Dong, X. and Li, J. 2017. New molecular insights into an archaeal RNase J reveal a conserved processive exoribonucleolysis mechanism of the RNase J family. *Molecular Microbiology* 106(3), pp. 351–366.

Zhu, Q., Mai, U., Pfeiffer, W., Janssen, S., Asnicar, F., Sanders, J.G., Belda-Ferre, P., Al-Ghalith, G.A., Kopylova, E., McDonald, D., Kosciolk, T., Yin, J.B., Huang, S., Salam, N., Jiao, J.-Y., Wu, Z., Xu, Z.Z., Cantrell, K., Yang, Y., Sayyari, E., Rabiee, M., Morton, J.T., Podell, S., Knights, D., Li, W.-J., Huttenhower, C., Segata, N., Smarr, L., Mirarab, S. and Knight, R. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nature Communications* 10(1), p. 5477.

Ziemski, M., Jomaa, A., Mayer, D., Rutz, S., Giese, C., Veprintsev, D. and Weber-Ban, E. 2018. Cdc48-like protein of actinobacteria (Cpa) is a novel proteasome interactor in mycobacteria and related organisms. *eLife* 7.

Zinder, J.C. and Lima, C.D. 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes & Development* 31(2), pp. 88–100.

Zinoviev, A., Ayupov, R.K., Abaeva, I.S., Hellen, C.U.T. and Pestova, T.V. 2020. Extraction of mRNA from Stalled Ribosomes by the Ski Complex. *Molecular Cell*.

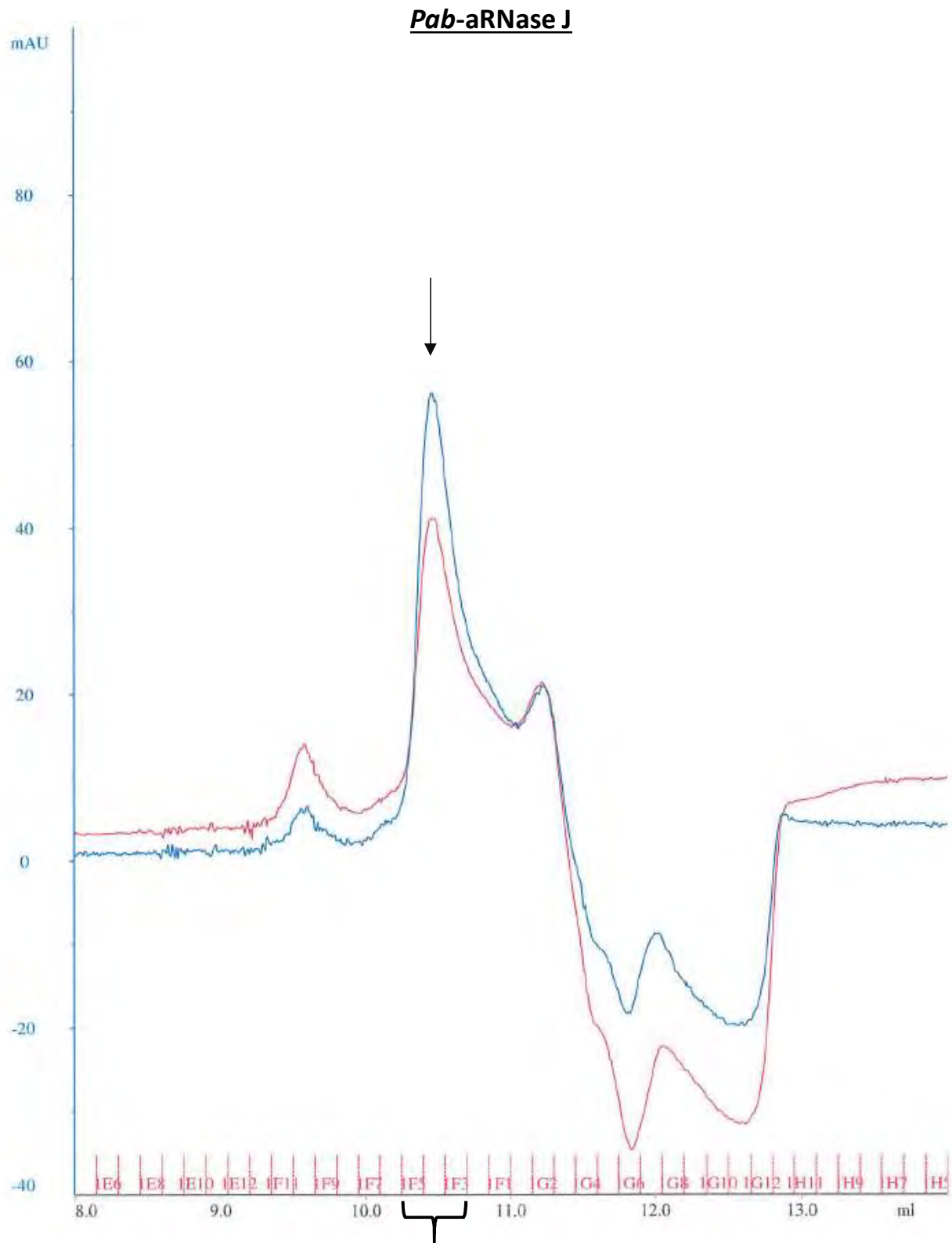
De Zoysa, M.D. and Yu, Y.-T. 2017. Posttranscriptional RNA Pseudouridylation. *The Enzymes* 41, pp. 151–167.



# ANNEXE

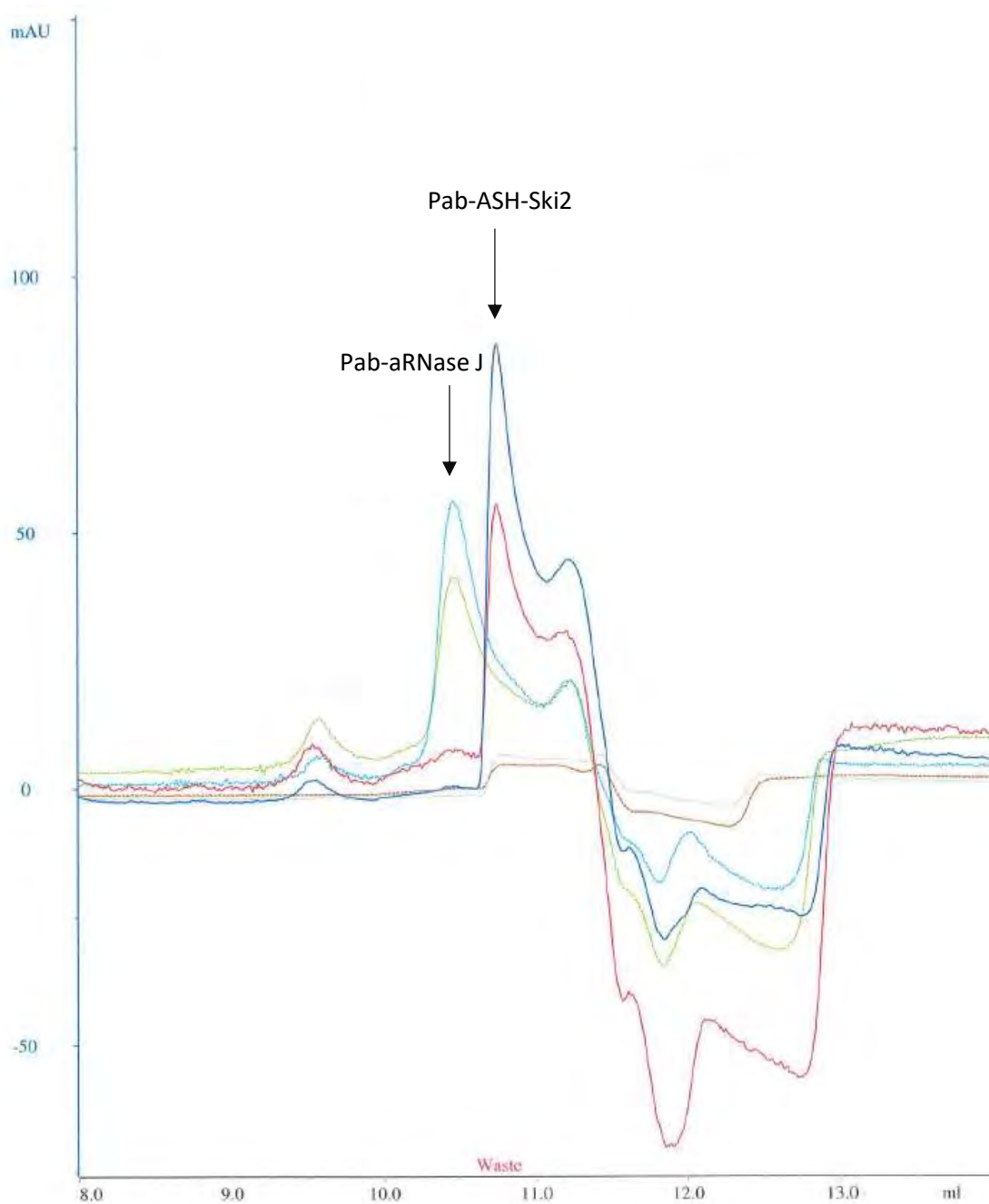






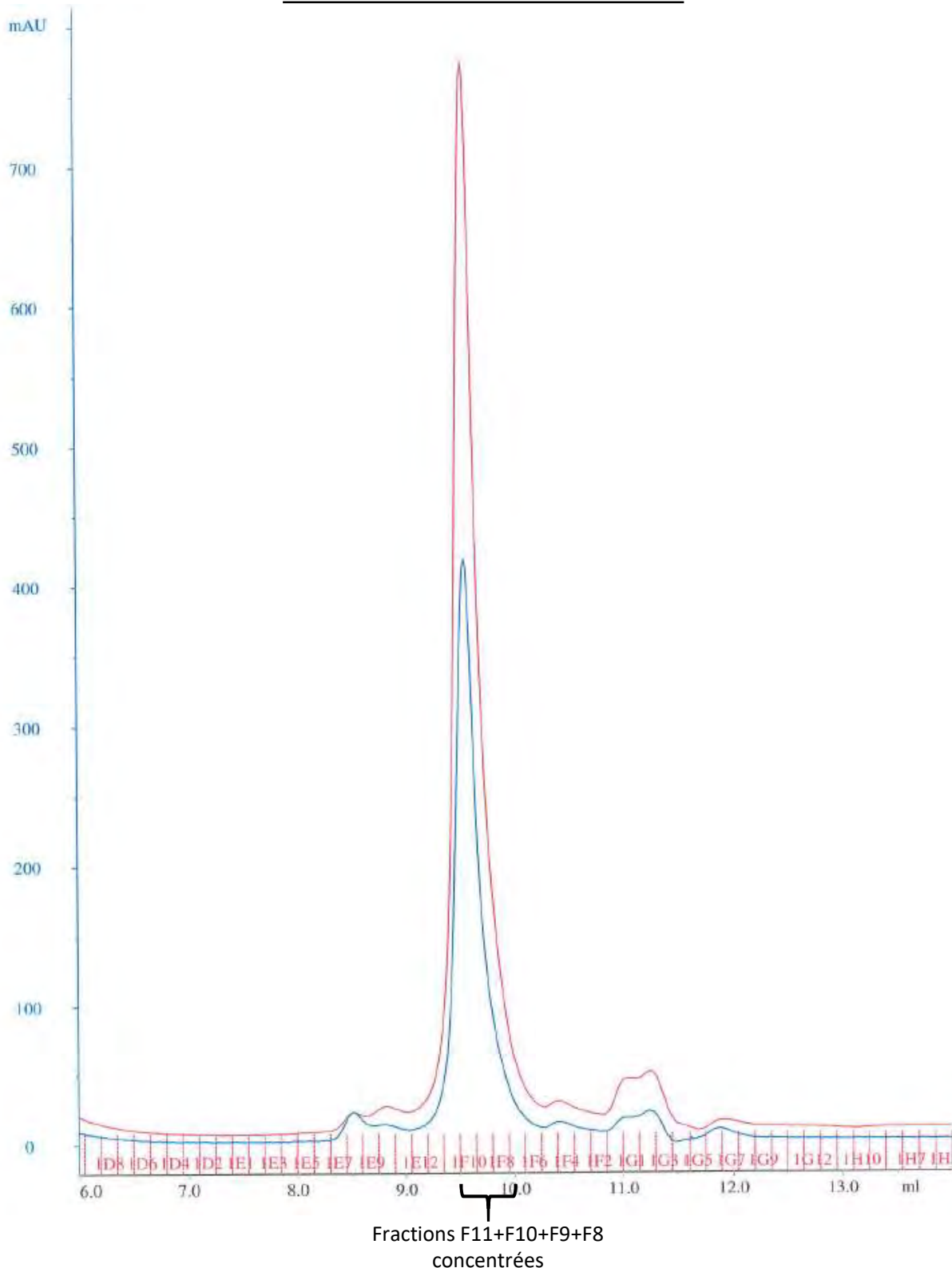
**ANNEXE- Figure 1.** Chromatogramme en HPLC de *Pab-aRNase J*. 1200pmol de *Pab-aRNase J* sont chauffées 5' à 50°C puis injectée sur une colonne Agilent Bio SEC- 5 préalablement équilibrée dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol .

## *Pab*-ASH-Ski2 et *Pab*-aRNase J superposés



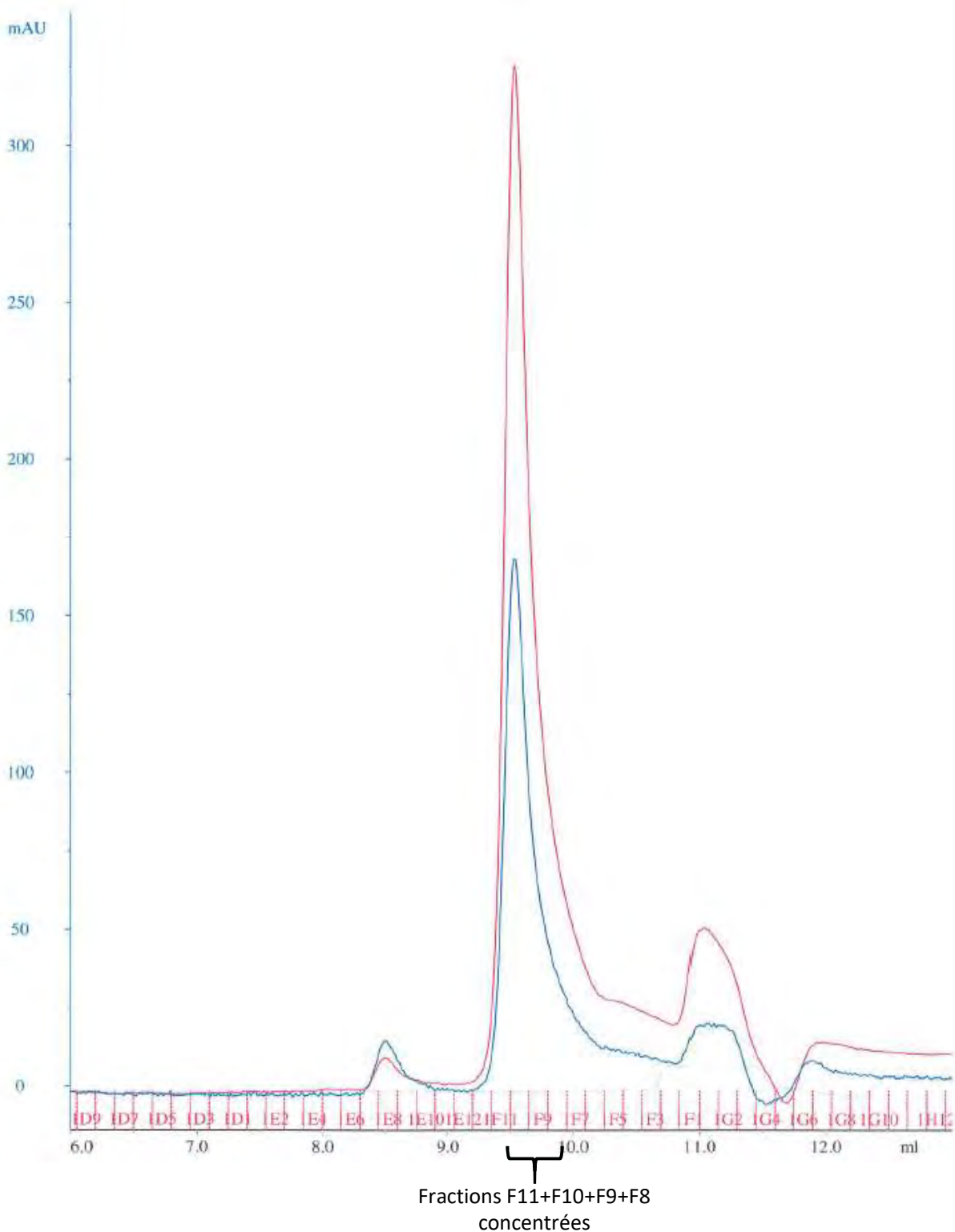
**ANNEXE- Figure 2.** Chromatogramme en HPLC de *Pab*-aRNase J et *Pab*-ASH-Ski2 superposés. 1200pmol de protéines sont chauffées 5' à 50°C puis injectée sur une colonne Agilent Bio SEC- 5 préalablement équilibrée dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol .

**Pab-aRNase J + Pab-30S + ARNm**



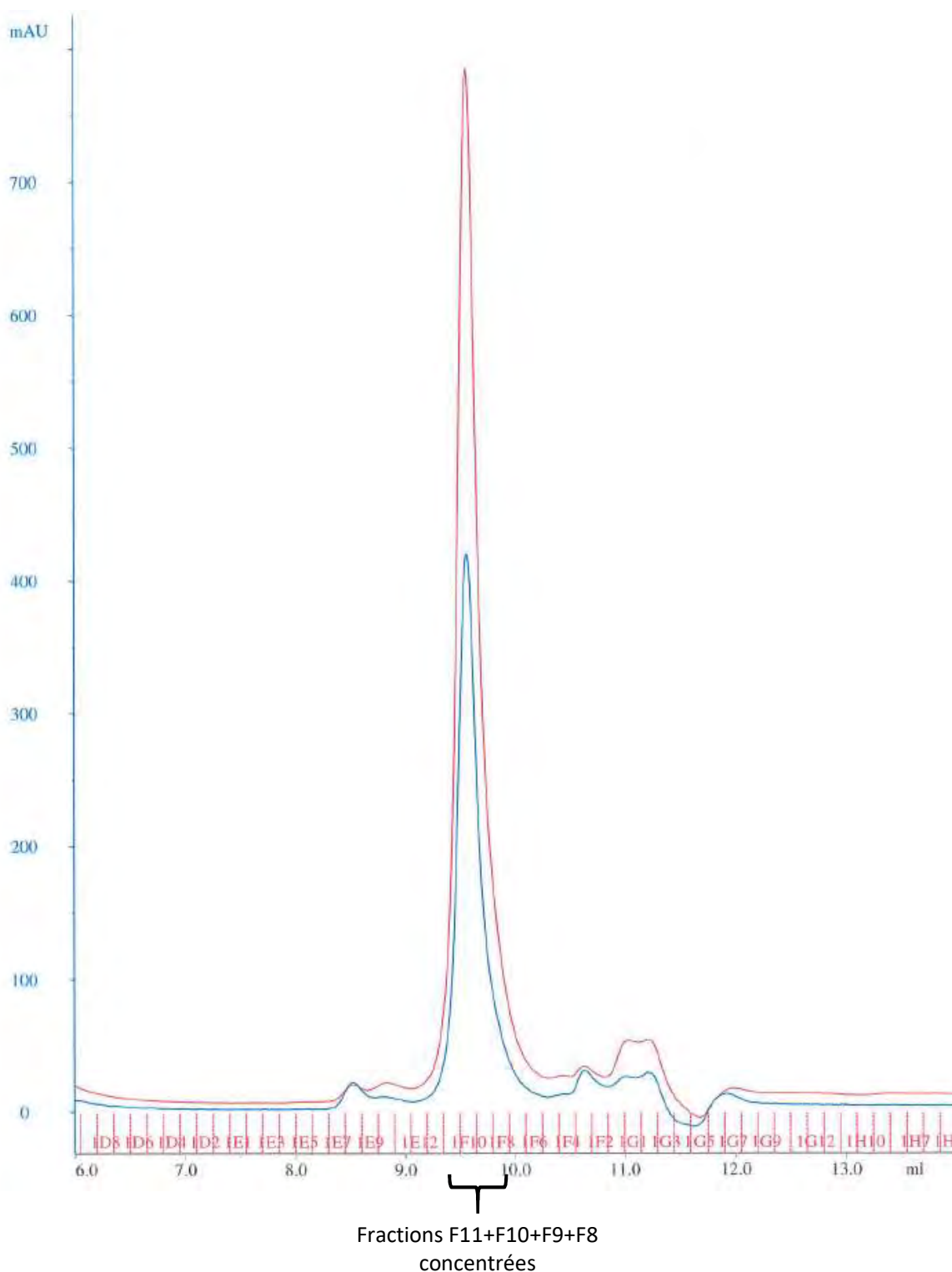
**ANNEXE- Figure 3.** Chromatogramme en HPLC du mélange *Pab*-aRNase J+*Pab*-30S+ARNm. 130pmol de *Pab*-aRNase J sont mis en présence de 30pmol de *Pab*-30S dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol et chauffées 5' à 50°C puis injectées sur une colonne Agilent Bio SEC- 5. Les fractions du pic majoritaire ont été concentrées et déposées sur un gel SDS PAGE coloré au bleu de Coomassie présenté dans la **figure 22** Partie III-RESULTATS.

**Pab-ASH-Ski2 + Pab-30S + ARNm**



**ANNEXE- Figure 4.** Chromatogramme en HPLC du mélange *Pab-ASH-Ski2+Pab-30S+ARNm*. 130pmol de *Pab-ASH-Ski2* sont mis en présence de 30pmol de *Pab-30S* dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol et chauffées 5' à 50°C puis injectées sur une colonne Agilent Bio SEC- 5. Les fractions du pic majoritaire ont été concentrées et déposées sur un gel SDS PAGE coloré au bleu de Coomassie présenté dans la **figure 22** Partie III-RESULTATS.

***Pab*-aRNase J + *Pab*-ASH-Ski2 + *Pab*-30S + ARNm**



**ANNEXE- Figure 5.** Chromatogramme en HPLC du mélange *Pab*-aRNase J+*Pab*-ASH-Ski2+*Pab*-30S+ARNm. 130pmol de *Pab*-ASH-Ski2 et *Pab*-aRNase J sont mis en présence de 30pmol de *Pab*-30S dans un tampon 10mM HEPES pH 7.5, 10mM MgAc, 300mM NH<sub>4</sub>Cl et 6mM β-mercaptoéthanol et chauffées 5' à 50°C puis injectées sur une colonne Agilent Bio SEC- 5. Les fractions du pic majoritaire ont été concentrées et déposées sur un gel SDS PAGE coloré au bleu de Coomassie présenté dans la **figure 22** Partie III-RESULTATS.

		LiSo4			LiSo4			LiSo4			(NH4)4SO4		
		0,7 M	1 M	1,3 M	0,7 M	1 M	1,3 M	0,7 M	1 M	1,3 M	0,7 M	1 M	1,3 M
NiCl2	0,01 M	MES pH 6			Cacod pH 6.8			HEPES pH 7.5			Cacod pH 6.8		
NiCl2	0,05 M												
CaCl2	0,02 M												
ZnCl2	0,02 M												
NiCl2	0,01 M	Tris pH 8			Tris pH 8.5			Bicine pH 9			Tris pH 8.5		
NiCl2	0,05 M												
CaCl2	0,02 M												
ZnCl2	0,02 M												

		pH						pH							
		6	6,8	7,5	8	8,5	9	6	6,8	7,5	8	8,5	9		
NaF	0,1 M							PEG 3350 15%						0,1 M	MgAc
	0,2 M													0,2 M	MgAc
	0,3 M													0,02 M	ZnCl2
	0,4 M													0,02 M	ZnAc
NaF	0,1 M	ZnCl2 0,01 M						PEG 3350 20%						0,1 M	MgAc
	0,2 M													0,2 M	MgAc
	0,3 M													0,02 M	ZnCl2
	0,4 M													0,02 M	ZnAc

**ANNEXE- Figure 6.** Schéma des cribles manuels de cristallisation. Les cribles ont été réalisés sur des plaques de 96 puits.



# ARCHAEAL SF1 AND SF2 HELICASES: UNWINDING IN THE EXTREME

# 1

Mirna Hajj<sup>1,3</sup>, Samar El-Hamaoui<sup>2</sup>, Manon Batista<sup>3</sup>, Marie Bouvier<sup>3</sup>, Ziad Abdel-Razzak<sup>2</sup>,  
Béatrice Clouet d'Orval<sup>3</sup> and Hala Chamieh<sup>1,2</sup>

<sup>1</sup>Laboratory of Applied Biotechnology, Azm Center for Research in Biotechnology and Its Applications, Lebanese University, Tripoli, Lebanon <sup>2</sup>Department of Life and Earth Sciences, Faculty of Science, Lebanese University, Tripoli, Lebanon <sup>3</sup>Laboratoire de Microbiologie et Génétique Moléculaires, Centre de Biologie Intégrative (CBI), Centre National de la Recherche Scientifique (CNRS), Université de Toulouse, UPS, Toulouse, France

## INTRODUCTION

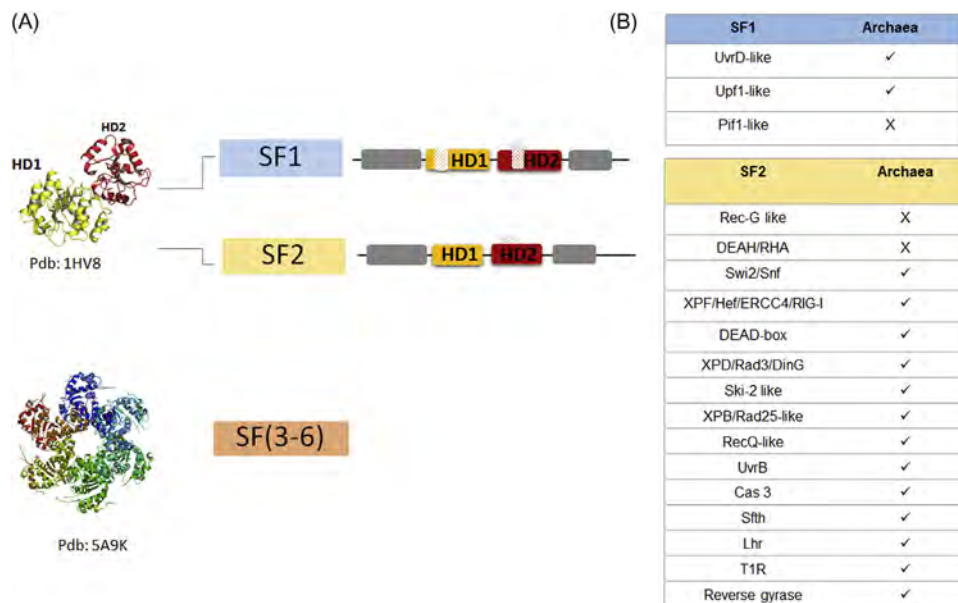
In 1977 Carl Woese and collaborators identified Archaea as a separate domain of life. Since then, Archaea have been considered as valuable study models to understand the diversity of life styles on earth [1]. The first-discovered Archaea were distinguished by their ability to thrive in challenging habitats such as high salinity, pH, temperature, and high pressure. However, novel high-throughput sequencing methods permitted the identification that Archaea constitute a considerable fraction of the Earth's ecosystems with astonishing diversity and omnipresence. Remarkably, archaeal microorganisms are found to play pivotal roles in geochemical cycles as well as being part of human gut microbiota [2,3].

Original classification based on 16S rRNA showed that the archaeal phylogeny embraces two major phylogenetic groups, named Euryarchaeota and Crenarchaeota [4]. Subsequently, phylogenomic analyses using an increasing number of sequenced archaeal genomes led to the characterization of several phyla, including Euryarchaeota and two main "superphyla," namely the TACK superphylum (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota) and the DPANN superphylum (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and NanohaloArchaea) [5,6]. Metagenomics analyses allowed the identification of a novel archaeal clade, named Asgard, with an expanded repertoire of eukaryotic signatures. These findings provide novel hypothesis on the origin of Eukarya within the archaeal domain [7].

At first glance, archaeal cells look like bacterial cells, however unusual composition of membrane lipids and cell envelope made irrevocably clear the existence of profound differences between Archaea and Bacteria. Archaeal membranes are composed of ether lipids instead of ester lipids and the cell envelope does not contain peptidoglycan [8]. While Archaea share some bacterial essential functioning systems of energy metabolism, the informational processing system which includes DNA replication, transcription, and translation, are closely related to Eukarya [9–12]. Moreover, the archaeal genome is organized by either eukaryotic-like histone proteins or

bacterial-like nucleoid-associated proteins [13]. In this mosaic setting, we are interested in deciphering the panel of helicases existing in Archaea that would drive numerous fundamental metabolic pathways.

Helicases are molecular motors that couple the use of energy to countless biological processes. By catalyzing the separation of double stranded nucleic acids into a single stranded one and the dissociation of nucleic-acid associated proteins, helicases participate in all aspects of DNA and RNA metabolism, and help in chromatin remodeling [14–17]. Helicases are grouped into six superfamilies (SF1–6) based on amino acid sequence similarity, oligomeric state (monomeric or hexameric), activity (substrate as single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), or RNA, translocating activity), and polarity ( $5' \rightarrow 3'$ ,  $3' \rightarrow 5'$ ) [16] (Fig. 1.1). The two largest acknowledged superfamilies, SF1 and SF2, which group nonhexameric helicases, perform diverse cellular functions in DNA replication, repair, recombination, RNA metabolism, and protein translation. SF3–6 form hexameric toroid structures (Fig. 1.1) [18–20]. SF3 helicases comprise viral helicases.



**FIGURE 1.1**

(A) Domain organization of SF superfamilies and (B) occurrence of SF1 and SF2 helicases in Archaea. Helicases are classified into six superfamilies which include the hexameric helicases SF3 to SF6 and the non hexameric helicases SF1 and SF2. SF1 and SF2 helicases share a conserved helicase core composed of two RecA-like domain folds (HD1 in yellow and HD2 in red). SF1 helicases present multiple insertions within the helicase core. SF1 and SF2 helicases are divergent in their N- and C-terminus which are attributed to the diversity of helicase functions in vivo (represented by gray boxes). Crystal structure of SF2 DEAD-box helicase from *Methanococcus janaschii*, Pdb: 1HV8. Crystal structure of hexameric helicase E1 from papillomavirus, Pdb: 5A9K.

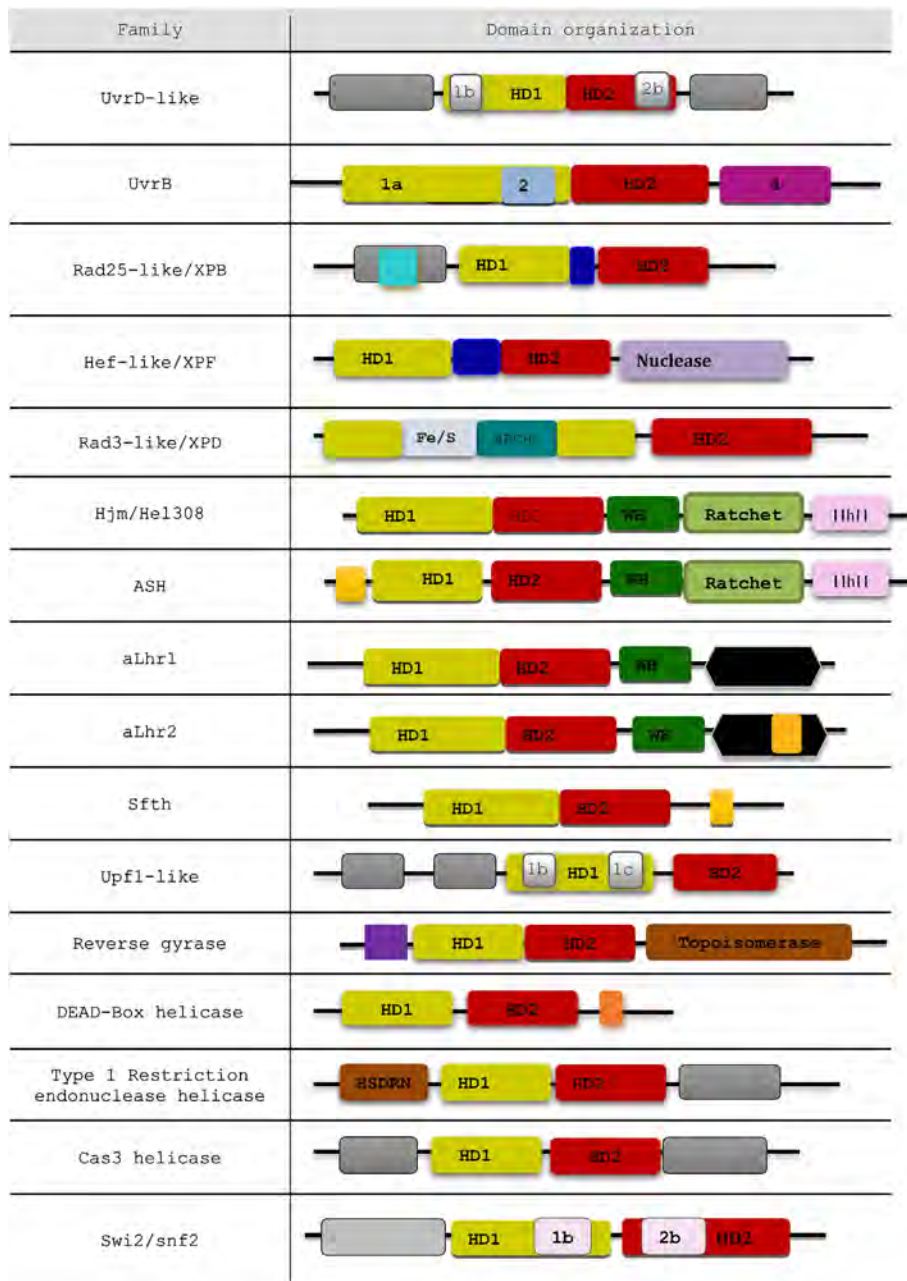
SF4 and SF5 function as replicative and transcription termination factors, respectively. SF6 include the minichromosome maintenance helicases and the RuvB helicase [20,21].

The SF1 and SF2 helicases are characterized by a conserved helicase core formed by two RecA domains which consist of nine characteristic sequence motifs named Q, I, Ia, Ib, and II to VI. These motifs slightly differ between SF1 and SF2 and among those Walker A (motif I) and Walker B (motif II) bind NTPs (Fig. 1.1) [22,23]. The specificity of action of SF1 and SF2 helicases has been mostly attributed to their accessory domains present at their N- and C-terminus in addition to their conserved helicase core. SF1 and SF2 were also classified based on their translocation polarity. Two groups emerged: the SF1/2A and SF1/2B with 3'–5' or 5'–3' translocation polarity, respectively [16]. Another classification was further refined based on the sequence conservation of bacterial and eukaryal-like helicases and on structural and mechanistic features allowing the identification of twelve distinct families (Fig. 1.1) [24]. The SF1 includes UvrD-like/Rep, Pif-1-like, and Upf1-like families, whereas the SF2 accounts for the Rec-G like, RecQ-like, XPD/Rad3/DinG, Ski-2 like, type 1 restriction enzyme helicase subunit (T1R), Swi2/Snf, XPF/Hef/ERCC4/RIG-I nuclease helicase, DEAD-box, and the DEAH/RHA families [18,25,26].

To overcome the gap of knowledge on archaeal helicase families, a comprehensive *in silico* analysis allowed retrieving the first exhaustive list of SF1 and SF2 helicases in Archaea. Each family was named based on knowledge of the function of their bacterial and eukaryotic counterparts (Fig. 1.1) [27]. Most of the already-known SF1 and SF2 families are represented in Archaea. Two major archaeal SF1 families could be defined, the UvrD-like and Upf1-like. Remarkably, the SF1 helicases are mostly restricted to the Euryarchaeota phylum [27]. In addition, none of the sequenced archaeal genomes encode the Pif-1-like helicases which have been shown in Eukarya to be involved in the maintenance of both the mitochondrial and nuclear genomes [28]. Finally, all the major SF2 families were retrieved in the archaeal classification with the exception of the bacterial Rec G-like involved in DNA replication, recombination and repair [29], and the DEAH/RHA families that functions in premessenger RNA splicing and ribosome biogenesis [30] (Fig. 1.1). Structural and mechanistic studies were reported for some members of these families [31–37]. Nevertheless, associated cellular processes have not been yet addressed.

## THE Uvr AND XP HELICASE FAMILIES

Helicases play essential roles at different steps of the DNA repair pathways. Nucleotide excision repair (NER), base excision repair (BER), and DNA mismatch repair (MMR) are the main known pathways, and all act on one of the two strands of the damaged DNA. Among these pathways, NER is the most versatile one and seems to be conserved in all domains of life [38–40]. Briefly, an initial recognition of DNA damage is followed by excision of the damaged base of the ssDNA. These steps require the coordinated action of DNA repair helicases and nuclease enzymes to detect the damaged DNA. The existence of NER pathway in Archaea has not been formally demonstrated, but it is suggested based on the presence of NER pathway like-enzymes. Most archaeal genomes encode for a homologue of the eukaryotic XP (*Xeroderma pigmentosum*) helicases. In addition, few encode for a homologue of the prokaryotic Ultraviolet repair (Uvr) helicase [27,40]. It has been suggested that Archaea use a simplified version of the eukaryal NER pathway and that the bacterial NER systems come from horizontal gene transfer from mesophilic archaeal species [40].



**FIGURE 1.2**

Structural domain organization of archaeal SF1 and SF2 families.

The Helicase domain 1 (HD1) and 2 (HD2) are respectively represented in yellow and red in all helicase families. The C- and N-terminal additional domains are colored as follows: UvrB: domain 1a corresponds to HD1 and contains two insertion domains 1b and 2 that are represented in green and blue, respectively, HD2 represents domain 3; XP helicases: the thumb motif located between HD1 and HD2, the C-terminal damage recognition domain (DRD) and the Fe/S cluster and ARCH domain inserted in HD1 are respectively represented in blue, cyan, blue-white and aquamarine; DEAD-box helicase: the C-terminal DbpA-domain is in orange; Type 1 Restriction

*(Continued)*

Briefly, the Uvr A/B/C/D proteins are known to be the key actors of the NER pathway in Bacteria [41]. Helicase activities are required in two steps of the NER pathway: the first one carried out by the UvrB helicase verifies the damage, whereas the second is part of the post-incision complex and involves the UvrD or PcrA activity which removes the excised fragment containing the damage.

### **The Archaeal UvrD-Like Helicase**

The UvrD-like helicases are described as SF1 helicase with a 3′–5′ polarity. These helicases consist of two RecA helicase core domains HD (named 1a and 2a) and two inserted auxiliary domains 1b and 2b within both HD domains, respectively (Fig. 1.2). These auxiliary domains possess DNA binding motifs [42]. Genes encoding UvrD-like helicases are particularly abundant in genomes of the Euryarchaeota, in particular in the Halobacteriales and Methanosarcinales clades [27]. However, euryarchaeal UvrD-like helicases show an extreme divergence of their N- and C-terminal unstructured regions that are proposed to interact with UvrB. Intriguingly, the UvrD encoding genes do not show the same taxonomic distribution as the other Uvr-like proteins across the archaeal phylogeny. This questions the physiological role(s) of the UvrD-like helicases in Archaea [27]. In fact, UvrD from *Escherichia coli* has been shown to play other roles in DNA replication by acting on Okazaki fragments and in methyl-directed mismatch repair (MMS) [43–46].

### **The Archaeal UvrB-Like Helicase**

The UvrB-like helicases have been described as SF2 helicase with a 3′–5′ polarity and, based on *E. coli* UvrB, a weak ATPase and unwinding activities. Further findings on *Mycobacterium* UvrB suggest that UvrB-like proteins could also display strong ATPase and helicase activities [47,48]. The crystal structure of UvrB showed four domains: domains 1 to 4. Domain 1 was further divided into 1a and 1b. Domains 1a and 3 are conserved regions with the SF2 family members and contain their characteristic motifs (Fig. 1.2), The three auxiliary domains are domain 1b (which is supposed to favor additional interactions with the DNA), domain 2 (which provides interaction with UvrA protein), and domain 4 (which interacts with UvrA and UvrC proteins) [49,50]. Archaeal UvrB-like members are abundant in Euryarchaeota and possess a conserved C-terminal domain which is supposed to interact with UvrC [27]. However, the *in silico* detection of these helicases in the archaeal phylogeny are not yet supported by *in vivo* and *in vitro* experimental evidences and the existence of archaeal UvrA/B/C/D complexes or subcomplexes have not yet been demonstrated. Genetic studies showed that, in halophilic Archaea, the deletion of the genes encoding the UvrA/B/C repair proteins render the cells hypersensitive to ultraviolet (UV) radiation (Table 1.1) [51]. Other studies performed in *Haloferax volcanii* showed that UvrA/B/C proteins work together with the PCNA factor named NerA in repairing DNA damage resulting from exposure to mitomycin [52,53].

---

◀ endonuclease helicase: the N-terminal nuclease domain is in brown; Swi2/Snf: the domain 1B inserted in HD1 and the domain 2B inserted in HD2 are in light pink; reverse gyrase: the additional N-terminal domain and the C-terminal topoisomerase domain are respectively in purple and in brown; Hjm and Lhr: WH domain is colored in dark green; Hjm and ASH: the ratchet is in olive green, the helix-loop-helix (HLH) is in pink; ASH, Lhr and Sfh: cysteine-rich region are in yellow. All grey boxes indicate divergent regions within protein families. Black boxes indicate regions of unidentified functions.

**Table 1.1 Phenotypical Analyses of Archaeal Strains Knocked-out for Gene Encoding SF2 Helicases**

Gene	ORF ID	Organism	Predicted Function	Phenotype	Ref.
XPB	XPB1: SiRe_1128	<i>Sulfolobus islandicus</i>	NER	None observed	[75]
	XPB2 SiRe_1526		NER	None observed	[75]
	TK0928	<i>Thermococcus kodakarensis</i>	NER	Negligible DNA-damage sensitivity to (UV) irradiation, methyl methanesulfonate (MMS) and mitomycin C (MMC)	[68]
Hef/XPF	TK1021	<i>Thermococcus kodakarensis</i>	Helicase-endonuclease	General sensitivity to UV irradiation and MMC. Particular hypersensitivity to MMS	[68]
	HVO_RS19270	<i>Haloferax volcanii</i>	Helicase endonuclease	Moderate sensitivity to DNA crosslinking agents	[69]
XPD	TK0784	<i>Thermococcus kodakarensis</i>	NER	Slight sensitivity to UV irradiation, MMS, and MMC	[68]
	SiRe_1685	<i>Sulfolobus islandicus</i>	NER	None observed	[75]
Hjm/ Hel308	TK1332	<i>Thermococcus kodakarensis</i>	DNA helicase	Sensitivity to MMS	[68]
	PF_RS03450	<i>Pyrococcus furiosus</i>	DNA helicase	None observed	[78]
	SiRe_0250	<i>Sulfolobus islandicus</i>	DNA helicase	Lethal	[75,81]
DEAD	TK0306	<i>Thermococcus kodakarensis</i>	RNA helicase	Decreased cell growth	[110]
UvrA/B/ C	AE004437.1	<i>Halobacterium sp.</i>	NER	Hypersensitivity to UV irradiation	[51]
Reverse Gyrase	AB117612	<i>Thermococcus kodakarensis</i>	Topoisomerase	Decreased growth	[97]
Lhr2	SiRe_1605	<i>Sulfolobus islandicus</i>	DNA/RNA helicase	Sensitivity to MMS	[90]
ASH	TK0566	<i>Thermococcus kodakarensis</i>	Unknown	Slow cell growth, sensitivity to DNA-damaging agents	[86]

Archaeal XP helicases were extensively studied at the structural level and have greatly contributed to our understanding of the eukaryal and bacterial SF2 helicases functions and mode of actions [32,36,37,40,54–58]. In eukaryotes, two types of NER pathways exist: firstly, the TCR (transcription-coupled Repair) pathway which acts to detect DNA damage and which is linked to the transcription machinery [59]; and secondly a global genome repair (GGR) pathway that involves the XPC-hr23B heterodimer to detect DNA damage [60]. In both pathways, a multi-subunit complex



composed of the Transcription Factor II-H complex and of two main helicases XPB and XPD is required to bind and extend the ssDNA around the damage site. The activity of these two helicases allows other NER factors such as the helicase-nuclease XPF-ERCC1 and XPG endonuclease to cleave the 5' and 3' sides of the lesion, respectively. In Archaea, a pathway similar to NER has not yet been demonstrated. The recognition of the lesion has been suggested to occur during transcription by the action of SSB proteins or the RNAP, or through an unidentified damage recognition protein [40]. Most of the archaeal organisms encode for XP-like helicases XPB and XPD and the nuclease-helicase XPF, but they do not follow the same overall distribution across archaeal members raising the question of how the NER pathway will operate in the absence of one of the XP helicase partners. An endonuclease named Bax1 (Binds archaeal XPB) was found to form a stable complex with XPB in vitro. It was suggested that Bax1 may be the archaeal counterpart of the eukaryal nuclease XPG by acting together with XPB in cleaving 3' of the DNA damaged site [58].

### ***XPB/rad25 Helicase***

XPB/Rad25-like helicases have been described as SF2 helicases with a 3'–5' polarity that function in eukaryotic NER pathways, but are also part of the basal transcription machinery in eukaryotes [61]. The crystal structure of XPB from *Archaeoglobus fulgidus* (AfuXPB) gave the first insight on the organization of an archaeal helicase from the SF2 family [54]. The resolved structure of AfuXPB showed, in addition to the two conserved helicase core domains (HD1 and HD2), a 100 amino acid N-terminal domain named the DRD domain (damage recognition domain) resembling the mismatch recognition domain (MRD) of the DNA repair MutS protein and is responsible for recognition of distorted and damaged DNA (Fig. 1.2). The DRD domain does not seem to be conserved across archaeal XPBs as the XPB sequences from Archaea are extremely divergent at their N- and C-termini [27]. Another domain found in XPB is the thumb motif located between HD1 and HD2 (Fig. 1.2). This motif is commonly found in DNA-dependent DNA polymerases and thought to be involved in branched DNA binding. In addition, the HD1 domain contains a RED motif that has a key role in DNA unwinding [54].

### ***Hef-Like Helicase***

Hef-like (*Helicase-associated Endonuclease for Fork-structured DNA*) helicases have been described as SF2 helicases with a 3'–5' polarity that in some cases retain, in addition to the helicase domain, a nuclease domain similar to the eukaryotic XPF endonuclease. The archaeal Hef was first identified from the hyperthermophilic archaeon *Pyrococcus furiosus* as a protein factor that can stimulate holiday junction resolution by the Hjc resolvase [62,63]. Most archaeal genomes, except for Thermoplasmatales, encode a Hef-like helicase. The Hef-like helicases exist in two forms: the long form consists of an N-terminal helicase domain carrying a conserved helicase core domain fused to a C-terminal nuclease and specific to the Euryarchaeota; the simpler version lacks the helicase domain in the Crenarchaeota and Thaumarchaeota (Fig. 1.2) [64]. The crystal structures from *Pyrococcus furiosus* (Pfu) and *Aeropyrum pernix* (Aep) Hef-like helicases provided the first structural information on the XPF family [65,66]. The C-terminal endonuclease domain is structurally related to the one found in the eukaryotic nucleases Mus81 and XPF. In addition to these domains, a conserved “helix-hairpin-helix” (HhH2) motif involved in protein dimerization and DNA binding was also discovered [56,62]. Evidence suggests that PfuHef plays a role in processing stalled



replication forks as it produces splayed duplexes from DNA forks and four-way junctions [35,56,67]. The recombinant *Thermococcus kodakarensis* Hef protein (TkoHef) exhibits a similar activity on fork-structured DNA in vitro [68]. Moreover, the strain knocked-out for *Tkohef* is highly sensitive to mitomycin C, suggesting that Hef is involved in numerous repair processes and is critical for DNA interstrand cross-link repair (Table 1.1) [68]. Similar observations were made on a *H. volcanii* strain that was deleted for the gene-encoding Hef. This strain exhibits end-joining defects, as well as homologous recombination and cross-link repair deficiencies. Conversely, *hef* deletion does not render the cells sensitive to UV radiation, meaning that the NER pathway is not impaired [69].

### **XPD/Rad3-like**

The XPD/Rad3-like family is the only SF2 helicase family with a 5′–3′ polarity with a weak DNA unwinding activity in vitro [70]. Archaeal XPDs were extensively studied at the structural and biochemical levels with four crystal structures published from three different archaeal organisms: *Thermoplasma acidophilum* (*TaaXPD*), *Sulfolobus acidocaldarius* (*SsAXPD*), and *Sulfolobus tokodaii* (*SstXPD*) [55,71–73]. These structures showed that archaeal XPD-like proteins consist of the helicase core domain (HD1 and HD2) and two additional domains that are inserted into HD1: an iron–sulfur 4FeS domain and an Arch domain defined by its arch-shaped structure (Fig. 1.2). The 4FeS domain is essential for the helicase activity and functions as a wedge structure involved in duplex separation. The Arch domain function is not characterized for archaeal XPD-like protein, however it has been shown to be essential for interaction with the CAK—cyclin-dependent kinase (CDK)-activating kinase—in eukaryotic XPD helicase. The published data using Archaea as models provided insights into the mechanism of XPD unwinding where, in Eukaryotes, it is thought that ssDNA passes first through the groove between the HD2 and arch domains, moving then through a hole encircled by the Arch, FeS-cluster, and HD1 domains [55]. Finally, it was shown that *SsaXPD* can efficiently unwind in vitro 5′ overhang, Y fork, or bubble substrates [74].

Altogether, it is not yet fully understood how the NER pathway is operating in Archaea; as deletion of either XPB or XPD did not affect DNA repair in *Sulfolobus islandicus* and only slight sensitivity to DNA damaging agents was observed in the null mutants of *T. kodakarensis* (Table 1.1) [68,75]. One possible explanation is the presence of numerous XPB/Rad25 and XPD/Rad3 encoding genes. These genes could be functionally redundant, and in this case no clear phenotype would be observed [68,75]. Further genetic studies are required to fully elucidate the exact role of XP proteins in DNA repair in Archaea.

### **THE Ski2-LIKE FAMILY**

The Ski2-like family was originally named after the Ski2 RNA helicase that works with the exosome in the eukaryotic turnover and quality control of mRNAs. These SF2 enzymes are mostly RNA helicases, however, only one member—the archaeal Hel308 member—is described as a DNA helicase [76]. In Eukarya, three major helicase groups exist, the cytoplasmic Ski2 and the nuclear Mtr4 associated with the eukaryal exosome machineries, and the Brr2 helicases involved in RNA quality control and RNA splicing, respectively. Initially, the archaeal DNA helicase homolog of the human HelQ was named Hel308 [77]. Subsequently, it appeared that Hel308 members belong to the Ski2-like family [76]. The protein Mth810 was called Hel308a since it resembled the human

Hel308 helicases implicated in DNA repair [77]. The *P. furiosus* was named Hjm (Holliday junction migration) as it was detected through its activity on model Holliday junctions in vitro [78]. Importantly, Hel308 members are commonly found through all the archaeal phylogeny with few exceptions [27]. These enzymes display a ssDNA stimulated ATPase activity and have in vitro activities on DNA forks and holiday junctions. They translocate along both DNA strands in a 3'→5' direction and a better efficiency on forked structures was observed [77]. Deciphering the biochemical activities of *S. tokodaii* Hjm identified, in contrast to other Hjm, unwinding activities in both directions [79]. Hjm was shown to interact with the endonuclease/resolvase Hjc in vitro, therefore it has been proposed that archaeal Hjm/Hel308 promotes replication fork regression through its interaction with Hjc [79,80]. However, in *T. kodakaraensis*, a knockout mutant of the encoding gene is viable in contrast to *Sulfolobus* where the Hjm helicase has been shown to be an essential protein (Table 1.1) [68,75,81].

High resolution structure of Hjm/Hel308 from *A. fulgidus* in an apo form and in complex with DNA captured the duplex-unwinding reaction and showed that initial strand separation is not dependent on ATP [82]. Hel308 consists of the two tandem RecA core domains (HD domains 1 and 2) (Fig. 1.2). The resolved structure showed a protruding region of  $\beta$ -hairpin secondary structure between motifs Va and VI in HD2 [31,36]. This hairpin was proposed to guide nucleic acid strand separation and thus plays a crucial role in coupling translocation to helicase activity. A third domain found in the C-terminal region is a “winged helix” domain (WH domain) identified as a sequence-specific DNA-binding domain in transcription factors, particularly those of the forkhead family (Fig. 1.2) [83]. The WH domain makes extensive contacts with branched structures and may play a role as a “hinge,” permitting Hel308 to bind and then close tightly around its DNA substrate. Mutations in the WH domain abolished DNA binding and unwinding activity [83]. Domain 4 (named as “ratchet”) was proposed to position nucleic acids for translocation linked to ATPase functions. At the C-terminal, a helix-loop-helix (HLH) caps the Hel308 helicase and exhibits typical DNA-binding activity with strong binding to ssDNA. The HLH domain possibly functions to limit the enzyme processivity on branched substrates [36]. Subsequently, the resolution of structures of the eukaryal Mtr4, Brr2, and Ski2 RNA helicases allowed to define the hallmarks for the Ski2-like family based on structural signature domains the helicase SF2 core and the typical C-terminal domain consisting of WH and a ratchet domain (Fig. 1.2). In addition, each family member will possess its own specific accessory domains [76].

Nevertheless, the exact role of archaeal Hel308/Hjm is yet to be uncovered. Since most Archaea lack RecQ, with the exception of Methanosarcinales, Hel308 is proposed to function in DNA recombination and repair as an alternative to the RecQ helicase commonly found in Bacteria and Eukarya [84]. When Hel308 was introduced into an *E. coli* replication-deficient strain, the RecQ-like phenotype was restored [77,84]. Hel308 was shown to physically interact with the Replication Protein A (RPA) that binds to ssDNA and helps loading Hel308 at stalled replication forks [85]

In addition to the Hel308 from the Ski2-like family, phylogenomic analysis demonstrated the presence of a helicase group specific to Archaea and mainly to Euryarchaeota named ASH for Archaea Specific Helicase. ASH has been shown to be evolutionary related to the Ski2-like family [27]. On the sequence level and as shown by secondary structure predictions, ASH is more similar to Hel308 than the other RNA helicases of the Ski2-like family. At the N-terminal, the ASH helicase differ by the presence of a cysteine-rich region which is predicted to be a metal binding motif (Fig. 1.2) [27]. Further, ASH was proposed to act as a transcription termination factor—also named

Eta for euryarchaeal termination activity [86]. Indeed, ASH was shown to disrupt the transcription elongation complex in vivo in *T. kodokarensis*. In addition, deletion of TK0566 (encoding ASH/Eta) results in slow growth and renders cells sensitive to DNA damaging agents (Table 1.1) [86].

## THE Lhr-LIKE AND Sfth-LIKE HELICASE FAMILIES

Lhr (Longest Helicase Related protein) and Sfth, two SF2 helicases with unidentified functions, are the most widespread in archaeal genomes [27]. Indeed, all archaeal genomes encode at least for one Lhr helicase and Lhr gene duplication events are commonly observed. Sfth is found in all Archaea except for the thermoplasmatales and some methanogens. The role of these two helicases have not yet been studied in spite of their large abundance across sequenced genomes.

Lhr that was first discovered in *E. coli* in 1995 is a large protein (1507 and 1538 amino acids residues in *Mycobacterium smegmatis* and in *E. coli*, respectively) [87]. Lhr is a prokaryote helicase since members of this family are only found in Bacteria and Archaea. They are particularly abundant in actinomycetes and proteobacteria [88]. Genetic studies in *E. coli* showed that the cellular growth is not impaired by the deletion of *lhr* [87]. Upon damage by UV irradiation or by treatment with mitomycin C, the expression of Lhr is upregulated in mycobacteria [88]. The first crystal structure of the first half of *M. smegmatis* Lhr (1–856) in complex with AMP-PNP and ssDNA just became available and exemplifies Lhr as a new SF2 helicase family in terms of domain organization [89]. The structure shows the conserved helicase core (HD1 and HD2), followed by a WH domain that displays a similar fold to the one observed in Hjm and RecQ helicases (Fig. 1.2). A novel structural module whose function is unknown is annotated as domain 4. Full length bacterial Lhr-like also possesses an additional C-terminal DEAD-associated helicase domain whose function is still unknown [87]. In Archaea, two Lhrs are commonly found, aLhr1 and aLhr2. A third distant phylogenetically group named Lhr-like is also commonly found in Sulfolobales. In terms of domain organization, archaeal Lhr resembles bacterial Lhr, with the aLhr2 proteins differing from other Lhr by the presence of a Cysteine-rich region which occurs in some metal binding proteins (Fig. 1.2) [27]. A deletion of the SiRe\_1605 gene (corresponding to *lhr2* gene) in *S. islandicus* render the strains more sensitive to MMS (methyl methanesulfonate) and leads to the deregulation of many nucleotide metabolism and DNA repair enzymes as well as the silencing of over 80 genes located at a specific region (Table 1.1) [90]

Sfth is widespread in Bacteria and in Archaea, and is limited to plants and fungi in eukaryotes [91]. Sfth is distinct from other helicases by the presence at its N-terminus of a DUF1998 domain containing a putative zinc-finger. Archaeal Sfth proteins display the same conserved helicase core and the characteristic DUF1998 domain (Fig. 1.2) [27]. The yeast homolog, named Hrq1, is a robust 3'–5' helicase with remarkable biochemical properties as it forms a homo-heptameric ring [91–93]. A deletion of the *hrq1* gene is dispensable for vegetative growth in *Saccharomyces cerevisiae* [94]. In the fission yeast *Saccharomyces pombe*, evidence showed that Hrq1 plays a role in the repair of DNA Interstrand cross-links and in genome maintenance [92]. In addition, the human RecQ4 was proposed to be an orthologous member of Sfth helicases, however it lacks the characteristic DUF1998 domain [92]. The physiological function of Sfth in prokaryotes is yet to be demonstrated. However, a novel defense mechanism against phages was found associated with the DUF1998 domain characteristics of the Sfth helicases. This defense mechanism was named DISARM for defense island systems associated with restriction-modification (RM) [95]. The

DISARM system is widespread in Bacteria and Archaea and consists of a DNA methylase and four other genes annotated as a helicase domain, a DUF1998 domain, a phospholipase D domain, and a gene of unknown function. When this system was introduced into *Bacillus subtilis*, the engineered Bacteria became protected against all three major families of tailed dsDNA phages [95].

## THE REVERSE GYRASE HELICASE FAMILY

Reverse gyrase enzymes are DNA topoisomerases that introduce positive supercoils into DNA and are specific to hyperthermophilic Archaea and Bacteria [96]. Reverse gyrase is one of the enzymes exclusively found in hyperthermophiles. Originally it was thought that this enzyme is essential for life at high temperature. However, *Thermococcus* cells knocked-out for the gene encoding the reverse gyrase are viable, meaning that this gene is not essential [97]. Nevertheless, the mutant strain shows a decrease in cellular growth rate and is thermosensitive, supporting its role at high temperature (Table 1.1). Reverse gyrases consist of a helicase core (HD1 and HD2) and a C-terminal type IA topoisomerase domain (Fig. 1.2). Crystal structures from *Thermotogata maritima* and *A. fulgidus* revealed domain arrangements including the functionally important zinc-finger motif [98,99]. A latch domain is inserted into HD2 and is supposed to play a role in the introduction of positive supercoiling by reverse gyrase. This domain connects the helicase and the topoisomerase domains and is required for their functional cooperation. An insert in HD1 adopts a helix-loop structure in the *T. maritima* enzyme and a  $\beta$ -hairpin structure in *A. fulgidus* reverse gyrase. Both inserts protrude from the same side of the helicase domain. The latch domain, the HD1 insert and the zinc-finger motif are required to bind and guide DNA during strand passage. The topoisomerase domain of the reverse gyrase is similar to the bacterial topoisomerase IA. The physiological function of the reverse gyrase was addressed in *Sulfolobus solfataricus*. A link to DNA repair was made by the identification of the translesion DNA polymerase SSoPolY/Dpo4 and of SSB as protein partner of the reverse gyrase [100,101].

## THE DEAD-BOX LIKE HELICASE

DEAD-box like proteins are SF2 RNA helicases that participate in many processes involving RNA such as transcription, translation, editing, mRNA degradation, and ribosomal assembly [20,102,103]. These proteins exhibit RNA-dependent ATPase activity and helicase activity [104]. The crystal structure of a DEAD-box like helicase from *Methanococcus janaschii* was the first published prototype of helicase core organization into two RecA domains [57]. DEAD-box like helicases are the shortest helicase in length and in many cases the DEAD-box like proteins lack a characteristic C-terminal domain (Fig. 1.2). Some of these helicases have a C-terminal DbpA RNA-binding domain (RBD) known to mediate the recognition of hairpin 92 of the 23S ribosomal RNA [105–107]. DEAD-box helicases are found in most of the Euryarchaeota with the exception of Pyrococcales and Halobacteriales, and only in the Sulfolobales from the Crenarchaeota. The physiological function of DEAD-box like helicases in Archaea is not yet reported, but some studies performed in *Methanococcoides burtonii* and *T. kodakarensis* showed that these DEAD-box helicases are induced upon cold stress and showed reduced growth upon deletion (Table 1.1) [108–110].

## THE Upf1-LIKE/Dna2 FAMILY

Upf1-like/Dna2 SF1 members have unwinding activity with a 5' → 3' polarity. Archaeal Upf1-like proteins possess similarities with the helicase/nuclease Dna2. In eukaryotes, Dna2 plays an important role in DNA replication by Okazaki fragment processing and in DNA repair pathways [105,111]. These proteins are sporadically widespread across archaeal genomes and their presence is restricted to Thermococcales, Archaeoglobales, and Halobacteriales. In addition, their N-terminal and C-terminal regions are extremely divergent (Fig. 1.2). The physiological function of archaeal Dna2 is uncertain. Dna2 from *Pyrococcus horikoshii* was characterized in vitro at the biochemical level and was shown to display 5'–3' DNA helicase and nuclease activities [112].

## THE Cas3, Snf2 AND T1R HELICASES FAMILIES

Cas3 helicase is an SF2 helicase that has been recognized as a novel SF2 clade, but it displays very weak conservation of the helicase core motifs [113]. T1R restriction helicase and cas3 helicases are part of prokaryotic defense mechanism [114]. Cas3 plays an essential role in prokaryotic adaptive immunity against foreign nucleic acids when invading viruses and plasmids [113]. The Type I restriction enzyme is a prokaryotic helicase and functions as part of a large multiprotein complex forming a restriction-modification system to protect Bacteria against foreign DNA [115,116].

SWI2/SNF2 like proteins have been documented in all domains of life [27]. However, in Bacteria and Archaea, the similarity is restricted to the helicase motifs and their function in archaeal physiology is not yet understood [117]. Whether they are involved in recombination—like Rad54 protein, or in transcription like Snf2 or Mot1, or in other chromatin remodeling processes—remains to be elucidated [117]. Crystal structures of SSoSwi2 in the absence or presence of its dsDNA substrate have been resolved [118]. The enzyme consists of two RecA-like domains (HD1 and HD2) and two specific subdomains inserted in each HD domain with a deep cleft separating the two domains (Fig. 1.2). The structure provided insights into how this enzyme translocates along the dsDNA where it was shown that SWI2/SNF2 ATPase moves along the minor groove of a dsDNA substrate without strand separation [119,120]. Studies on SWI2/SNF2 from *S. solfataricus* showed that the enzyme lacks helicase activity, but displays a dsDNA stimulated ATPase activity and translocation activity. It has been suggested that this translocation activity might be responsible for remodeling like activity of Sac7d/Alba:DNA complexes [119,120].

---

## CONCLUSIONS AND FUTURE PERSPECTIVES

Enzymes belonging to the SF1 and SF2 helicase families are structurally similar, but functionally diverse due to the N- or C-terminal elements appended to the conserved helicase core. In this chapter, we have highlighted archaeal families and associated features that characterized the overall archaeal SF1 and SF2 enzymes in which, for the most, cellular functions are still to be discovered. Despite the subtle differences in sequence and structure, it is clear that these protein families participate in fundamental ways that involve substrate recognition and conformational rearrangements that could induce association or dissociation of nucleic acid, but also nucleic acid-binding protein displacement. The question remains to understand their impact on physiology of archaeal cells and

their basic mechanistic features in the context of the extreme conditions encountered by archaeal microorganisms. In addition, further structure—function studies will provide a framework for understanding more complex eukaryotic counterparts often associated in multiprotein complexes.

---

## ACKNOWLEDGMENTS

This work is financed by the Lebanese University (UL) and National Center for research in Lebanon (CNRS-L). MH is a recipient of the AZM-UL excellency fellowship.

---

## REFERENCES

- [1] C.R. Woese, G.E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *Proc. Natl. Acad. Sci. U. S. A.* 74 (11) (1977) 5088–5090.
- [2] L. Eme, et al., Archaea and the origin of eukaryotes, *Nat. Rev. Microbiol.* 16 (2) (2018) 120.
- [3] K. Koskinen, et al., First insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin, *mBio* 8 (6) (2017).
- [4] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. Natl. Acad. Sci. U. S. A.* 87 (12) (1990) 4576–4579.
- [5] M. Pester, C. Schleper, M. Wagner, The Thaumarchaeota: an emerging view of their phylogeny and eco-physiology, *Curr. Opin. Microbiol.* 14 (3) (2011) 300–306.
- [6] L. Guy, T.J. Ettema, The archaeal ‘TACK’ superphylum and the origin of eukaryotes, *Trends Microbiol.* 19 (12) (2011) 580–587.
- [7] A. Spang, et al., Complex archaea that bridge the gap between prokaryotes and eukaryotes, *Nature* 521 (2015) 173.
- [8] S.V. Albers, B.H. Meyer, The archaeal cell envelope, *Nat. Rev. Microbiol.* 9 (6) (2011) 414–426.
- [9] K. Raymann, et al., Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea, *Genome Biol. Evol.* 6 (1) (2014) 192–212.
- [10] K.S. Makarova, E.V. Koonin, Archaeology of eukaryotic DNA replication, *Cold Spring Harb. Perspect. Biol.* 5 (11) (2013) a012963.
- [11] M. Martinez-Pastor, et al., Transcriptional regulation in archaea: from individual genes to global regulatory networks, *Annu. Rev. Genet.* 51 (2017) 143–170.
- [12] V. Anantharaman, E.V. Koonin, L. Aravind, Comparative genomics and evolution of proteins involved in RNA metabolism, *Nucleic Acids Res.* 30 (7) (2002) 1427–1464.
- [13] E. Peeters, et al., The interplay between nucleoid organization and transcription in archaeal genomes, *Nat. Rev. Microbiol.* 13 (6) (2015) 333–341.
- [14] A.M. Pyle, Translocation and unwinding mechanisms of RNA and DNA helicases, *Annu. Rev. Biophys.* 37 (1) (2008) 317–336.
- [15] P. Linder, E. Jankowsky, From unwinding to clamping —the DEAD box RNA helicase family, *Nat. Rev. Mol. Cell Biol.* 12 (2011) 505.
- [16] M.R. Singleton, M.S. Dillingham, D.B. Wigley, Structure and mechanism of helicases and nucleic acid translocases, *Annu. Rev. Biochem.* 76 (2007) 23–50.
- [17] E. Jankowsky, M.E. Fairman, RNA helicases – one fold for many functions, *Curr. Opin. Struct. Biol.* 17 (3) (2007) 316–324.



- [18] Mark S. Dillingham, Superfamily I helicases as modular components of DNA-processing machines, *Biochem. Soc. Trans.* 39 (2) (2011) 413–423.
- [19] T.M. Lohman, E.J. Tomko, C.G. Wu, Non-hexameric DNA helicases and translocases: mechanisms and regulation, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 391.
- [20] S.S. Patel, K.M. Picha, Structure and function of hexameric helicases, *Annu. Rev. Biochem.* 69 (2000) 651–697.
- [21] B. Medagli, S. Onesti, Structure and mechanism of hexameric helicases, *Adv. Exp. Med. Biol.* 767 (2013) 75–95.
- [22] M. Abdel-Monem, H. Durwald, H. Hoffmann-Berling, Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme, *Eur. J. Biochem.* 65 (2) (1976) 441–449.
- [23] D.D. Leipe, et al., Classification and evolution of P-loop GTPases and related ATPases, *J. Mol. Biol.* 317 (1) (2002) 41–72.
- [24] M.E. Fairman-Williams, U.-P. Guenther, E. Jankowsky, SF1 and SF2 helicases: family matters, *Curr. Opin. Struct. Biol.* 20 (3) (2010) 313–324.
- [25] D.C. Beyer, M.K. Ghoneim, M. Spies, Structure and mechanisms of SF2 DNA helicases, in: M. Spies (Ed.), *DNA Helicases and DNA Motor Proteins*, Springer New York, New York, NY, 2013, pp. 47–73.
- [26] K.D. Raney, A.K. Byrd, S. Aarattuthodiyil, Structure and mechanisms of SF1 DNA helicases, in: M. Spies (Ed.), *DNA Helicases and DNA Motor Proteins*, Springer New York, New York, NY, 2013, pp. 17–46.
- [27] H. Chamieh, H. Ibrahim, J. Kozah, Genome-wide identification of SF1 and SF2 helicases from archaea, *Gene* 576 (1, Part 2) (2016) 214–228.
- [28] A.K. Byrd, K.D. Raney, Structure and function of Pif1 helicase, *Biochem. Soc. Trans.* 45 (5) (2017) 1159–1171.
- [29] C.J. Rudolph, et al., Is RecG a general guardian of the bacterial genome? *DNA Repair (Amst.)* 9 (3) (2010) 210–223.
- [30] Y. He, G.R. Andersen, K.H. Nielsen, The function and architecture of DEAH/RHA helicases, *Biomol. Concepts* 2 (4) (2011) 315–326.
- [31] K. Büttner, S. Nehring, K.-P. Hopfner, Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* 14 (2007) 647.
- [32] H. Dürr, et al., X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA, *Cell* 121 (3) (2005) 363–373.
- [33] L. Fan, et al., Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, *Mol. Cell.* 22 (1) (2006) 27–37.
- [34] L. Fan, et al., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, *Cell* 133 (5) (2008) 789–800.
- [35] T. Nishino, et al., Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms, *Structure* 13 (8) (2005) 1183–1192.
- [36] T. Oyama, et al., Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm, *BMC Struct. Biol.* 9 (1) (2009) 2.
- [37] X. Zhang, et al., Crystal structure of an archaeal Ski2p-like protein from *Pyrococcus horikoshii* OT3, *Protein Sci.* 17 (1) (2008) 136–145.
- [38] J.E. Cleaver, E.T. Lam, I. Revet, Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity, *Nat. Rev. Genet.* 10 (11) (2009) 756–768.
- [39] J. Kuper, C. Kisker, Damage recognition in nucleotide excision DNA repair, *Curr. Opin. Struct. Biol.* 22 (1) (2012) 88–93.
- [40] C. Rouillon, M.F. White, The evolution and mechanisms of nucleotide excision repair proteins, *Res. Microbiol.* 162 (1) (2011) 19–26.
- [41] A. Sancar, DNA excision repair, *Annu. Rev. Biochem.* 65 (1996) 43–81.



- [42] W. Yang, Lessons learned from UvrD helicase: mechanism for directional movement, *Annu. Rev. Biophys.* 39 (2010) 367–385.
- [43] H. Boubakri, et al., The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo, *EMBO J.* 29 (1) (2010) 145–157.
- [44] C. Bruand, S.D. Ehrlich, UvrD-dependent replication of rolling-circle plasmids in *Escherichia coli*, *Mol. Microbiol.* 35 (1) (2000) 204–210.
- [45] G.F. Moolenaar, C. Moorman, N. Goosen, Role of the *Escherichia coli* nucleotide excision repair proteins in DNA replication, *J. Bacteriol.* 182 (20) (2000) 5706–5714.
- [46] A.B. Robertson, et al., MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair, *J. Biol. Chem.* 281 (29) (2006) 19949–19959.
- [47] R. Thakur, et al., Phospholipase D activity couples plasma membrane endocytosis with retromer dependent recycling, *eLife* 5 (2016).
- [48] K. Theis, et al., The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch, *Mutat. Res.* 460 (3-4) (2000) 277–300.
- [49] J. Eryilmaz, et al., Structural insights into the cryptic DNA-dependent ATPase activity of UvrB, *J. Mol. Biol.* 357 (1) (2006) 62–72.
- [50] T.R. Waters, et al., Damage detection by the UvrABC pathway: crystal structure of UvrB bound to fluorescein-adducted DNA, *FEBS Lett.* 580 (27) (2006) 6423–6427.
- [51] D.J. Crowley, et al., The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1, *Saline Syst.* 2 (2006) 11.
- [52] D.L. Jones, B.K. Baxter, DNA repair and photoprotection: mechanisms of overcoming environmental ultraviolet radiation exposure in halophilic archaea, *Front. Microbiol.* 8 (2017) 1882.
- [53] X. Giroux, S.A. MacNeill, A novel archaeal DNA repair factor that acts with the UvrABC system to repair mitomycin C-induced DNA damage in a PCNA-dependent manner, *Mol. Microbiol.* 99 (1) (2016) 1–14.
- [54] L. Fan, et al., Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, *Mol. Cell.* 22 (1) (2006) 27–37.
- [55] L. Fan, et al., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, *Cell* 133 (5) (2008) 789–800.
- [56] T. Nishino, et al., Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing, *Structure* 13 (1) (2005) 143–153.
- [57] R.M. Story, H. Li, J.N. Abelson, Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*, *Proc. Natl. Acad. Sci. U. S. A.* 98 (4) (2001) 1465–1470.
- [58] C. Rouillon, M.F. White, The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair, *J. Biol. Chem.* 285 (14) (2010) 11013–11022.
- [59] P.C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: two decades of progress and surprises, *Nat. Rev. Mol. Cell Biol.* 9 (12) (2008) 958–970.
- [60] J.H. Min, N.P. Pavletich, Recognition of DNA damage by the Rad4 nucleotide excision repair protein, *Nature* 449 (7162) (2007) 570–575.
- [61] F. Coin, V. Oksenyshyn, J.M. Egly, Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair, *Mol. Cell.* 26 (2) (2007) 245–256.
- [62] M. Newman, et al., Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition, *EMBO J.* 24 (5) (2005) 895–905.
- [63] K. Komori, et al., Novel endonuclease in Archaea cleaving DNA with various branched structure, *Genes Genet. Syst.* 77 (4) (2002) 227–241.
- [64] J.A. Roberts, S.D. Bell, M.F. White, An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA, *Mol. Microbiol.* 48 (2) (2003) 361–371.
- [65] T. Nishino, et al., Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing, *Structure* 13 (1) (2005) 143–153.

- [66] M. Newman, et al., Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition, *EMBO J.* 24 (5) (2005) 895–905.
- [67] K. Komori, et al., Cooperation of the N-terminal helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks, *J. Biol. Chem.* 279 (51) (2004) 53175–53185.
- [68] R. Fujikane, et al., Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*, *Genes Genet. Syst.* 85 (4) (2010) 243–257.
- [69] R. Lestini, Z. Duan, T. Allers, The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii*, *DNA Repair* 9 (9) (2010) 994–1002.
- [70] O.N. Voloshin, R.D. Camerini-Otero, The DinG protein from *Escherichia coli* is a structure-specific helicase, *J. Biol. Chem.* 282 (25) (2007) 18437–18447.
- [71] J. Kuper, et al., Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation, *EMBO J.* 31 (2) (2012) 494–502.
- [72] S.C. Wolski, et al., Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD, *PLoS Biol.* 6 (6) (2008) e149.
- [73] H. Liu, et al., Structure of the DNA repair helicase XPD, *Cell* 133 (5) (2008) 801–812.
- [74] J. Rudolf, et al., The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway, *Nucleic Acids Res.* 38 (3) (2010) 931–941.
- [75] C. Zhang, et al., Genetic manipulation in *Sulfolobus islandicus* and functional analysis of DNA repair genes, *Biochem. Soc. Trans.* 41 (1) (2013) 405–410.
- [76] S.J. Johnson, R.N. Jackson, Ski2-like RNA helicase structures, *RNA Biol.* 10 (1) (2013) 33–43.
- [77] C.P. Guy, E.L. Bolt, Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands, *Nucleic Acids Res.* 33 (11) (2005) 3678–3690.
- [78] R. Fujikane, et al., Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, *Pyrococcus furiosus*, *J. Biol. Chem.* 280 (13) (2005) 12351–12358.
- [79] Z. Li, et al., Hjm/Hel308A DNA helicase from *Sulfolobus tokodaii* promotes replication fork regression and interacts with Hjc endonuclease in vitro, *J. Bacteriol.* 190 (8) (2008) 3006–3017.
- [80] X. Song, J. Ni, Y. Shen, Structure-based genetic analysis of Hel308a in the hyperthermophilic archaeon *Sulfolobus islandicus*, *J. Genet. Genomics* 43 (6) (2016) 405–413.
- [81] Y. Hong, et al., Dissection of the functional domains of an archaeal Holliday junction helicase, *DNA Repair* 11 (2) (2012) 102–111.
- [82] K. Buttner, S. Nehring, K.P. Hopfner, Structural basis for DNA duplex separation by a superfamily-2 helicase, *Nat. Struct. Mol. Biol.* 14 (7) (2007) 647–652.
- [83] S.J. Northall, et al., DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain, *DNA Repair (Amst.)* 57 (2017) 125–132.
- [84] R. Fujikane, H. Shinagawa, Y. Ishino, The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork, *Genes Cells* 11 (2) (2006) 99–110.
- [85] I.L. Woodman, K. Brammer, E.L. Bolt, Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA), *DNA Repair (Amst.)* 10 (3) (2011) 306–313.
- [86] J.E. Walker, O. Luyties, T.J. Santangelo, Factor-dependent archaeal transcription termination, *Proc. Natl. Acad. Sci. U. S. A.* 114 (33) (2017) E6767–E6773.
- [87] N.B. Reuven, et al., The gene for the longest known *Escherichia coli* protein is a member of helicase superfamily II, *J. Bacteriol.* 177 (19) (1995) 5393–5400.
- [88] H. Ordóñez, S. Shuman, *Mycobacterium smegmatis* Lhr Is a DNA-dependent ATPase and a 3'-to-5' DNA translocase and helicase that prefers to unwind 3'-tailed RNA:DNA hybrids, *J. Biol. Chem.* 288 (20) (2013) 14125–14134.
- [89] A. Ejaz, et al., Structure of mycobacterial 3'-to-5' RNA:DNA helicase Lhr bound to a ssDNA tracking strand highlights distinctive features of a novel family of bacterial helicases, *Nucleic Acids Res.* 46 (1) (2018) 442–455.

- [90] X. Song, et al., Knockout and functional analysis of two DExD/H-box family helicase genes in *Sulfolobus islandicus* REY15A, *Extremophiles* 20 (4) (2016) 537–546.
- [91] L. Yakovleva, S. Shuman, *Mycobacterium smegmatis* SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3' to 5' translocase and helicase activities, *Nucleic Acids Res.* 40 (15) (2012) 7465–7475.
- [92] Matthew L. Bochman, et al., Hrq1, a homolog of the human RecQ4 helicase, acts catalytically and structurally to promote genome integrity, *Cell. Rep.* 6 (2) (2014) 346–356.
- [93] S.-H. Kwon, et al., *Saccharomyces cerevisiae* HrqI requires a long 3'-tailed DNA substrate for helicase activity, *Biochem. Biophys. Res. Commun.* 427 (3) (2012) 623–628.
- [94] A. Shiratori, et al., Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in *Saccharomyces cerevisiae* by gene disruption and Northern analysis, *Yeast* 15 (3) (1999) 219–253.
- [95] G. Ofir, et al., DISARM is a widespread bacterial defence system with broad anti-phage activities, *Nat. Microbiol.* 3 (1) (2018) 90–98.
- [96] P. Lulchev, D. Klostermeier, Reverse gyrase—recent advances and current mechanistic understanding of positive DNA supercoiling, *Nucleic Acids Res.* 42 (13) (2014) 8200–8213.
- [97] H. Atomi, R. Matsumi, T. Imanaka, Reverse gyrase is not a prerequisite for hyperthermophilic life, *J. Bacteriol.* 186 (14) (2004) 4829–4833.
- [98] A.C. Rodríguez, D. Stock, Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA, *EMBO J.* 21 (3) (2002) 418–426.
- [99] M.G. Rudolph, et al., Crystal structures of *Thermotoga maritima* reverse gyrase: inferences for the mechanism of positive DNA supercoiling, *Nucleic Acids Res.* 41 (2) (2013) 1058–1070.
- [100] A. Valenti, et al., Inhibition of translesion DNA polymerase by archaeal reverse gyrase, *Nucleic Acids Res.* 37 (13) (2009) 4287–4295.
- [101] A. Napoli, et al., Functional interaction of reverse gyrase with single-strand binding protein of the archaeon *Sulfolobus*, *Nucleic Acids Res.* 33 (2) (2005) 564–576.
- [102] P. Linder, M.C. Daugeron, Are DEAD-box proteins becoming respectable helicases?, *Nat. Struct. Biol.* 7 (2) (2000) 97–99.
- [103] S.S. Patel, I. Donmez, Mechanisms of helicases, *J. Biol. Chem.* 281 (27) (2006) 18265–18268.
- [104] O. Cordin, et al., The DEAD-box protein family of RNA helicases, *Gene* 367 (2006) 17–37.
- [105] F.V. Karginov, et al., YxiN is a modular protein combining a DExD/H core and a specific RNA-binding domain, *J. Biol. Chem.* 280 (42) (2005) 35499–35505.
- [106] M.G. Rudolph, D. Klostermeier, The *Thermus thermophilus* DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core, *RNA* 15 (11) (2009) 1993–2001.
- [107] K. Kossen, F.V. Karginov, O.C. Uhlenbeck, The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA, *J. Mol. Biol.* 324 (4) (2002) 625–636.
- [108] J. Lim, T. Thomas, R. Cavicchioli, Low temperature regulated DEAD-box RNA helicase from the antarctic archaeon, *Methanococcus burtonii* 11 Edited by J. H. Miller, *J. Mol. Biol.* 297 (3) (2000) 553–567.
- [109] Y. Shimada, et al., Property of cold inducible DEAD-box RNA helicase in hyperthermophilic archaea, *Biochem. Biophys. Res. Commun.* 389 (4) (2009) 622–627.
- [110] E. Nagaoka, et al., Importance and determinants of induction of cold-induced DEAD RNA helicase in the hyperthermophilic archaeon *Thermococcus kodakarensis*, *J. Bacteriol.* 195 (15) (2013) 3442–3450.
- [111] J.W. Gloor, et al., Biochemical analyses indicate that binding and cleavage specificities define the ordered processing of human Okazaki fragments by Dna2 and FEN1, *Nucleic Acids Res.* 40 (14) (2012) 6774–6786.
- [112] H. Higashibata, et al., Helicase and nuclease activities of hyperthermophile *Pyrococcus horikoshii* Dna2 inhibited by substrates with RNA segments at 5'-end, *J. Biol. Chem.* 278 (18) (2003) 15983–15990.

- [113] R.N. Jackson, et al., Fitting CRISPR-associated Cas3 into the helicase family tree, *Curr. Opin. Struct. Biol.* 24 (2014) 106–114.
- [114] K.S. Makarova, Y.I. Wolf, E.V. Koonin, Comparative genomics of defense systems in archaea and bacteria, *Nucleic Acids Res.* 41 (8) (2013) 4360–4377.
- [115] P.R. Bianco, C. Xu, M. Chi, Type I restriction endonucleases are true catalytic enzymes, *Nucleic Acids Res.* 37 (10) (2009) 3377–3390.
- [116] S. Sistla, D.N. Rao, S-Adenosyl-L-methionine-dependent restriction enzymes, *Crit. Rev. Biochem. Mol. Biol.* 39 (1) (2004) 1–19.
- [117] J. Yodh, ATP-dependent chromatin remodeling, *Adv. Exp. Med. Biol.* 767 (2013) 263–295.
- [118] H. Durr, et al., X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA, *Cell* 121 (3) (2005) 363–373.
- [119] L. Aravind, L.M. Iyer, V. Anantharaman, The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism, *Genome Biol.* 4 (10) (2003) R64.
- [120] H. Robinson, et al., The hyperthermophile chromosomal protein Sac7d sharply kinks DNA, *Nature* 392 (6672) (1998) 202–205.