



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Duy Khanh PHUNG

le mercredi 27 septembre 2017

Titre :

Métabolisme de l'ARN chez les archées :
Identification et caractérisation du complexe
ribonucléase β -CASP/ hélicase Ski2-like de *Pyrococcus abyssi*

École doctorale et discipline ou spécialité :

ED BSB : Microbiologie

Unité de recherche :

Laboratoire de Microbiologie et Génétique Moléculaires

Directeur/trice(s) de Thèse :

CLOUET d'ORVAL Béatrice: Directrice de Recherche, CNRS, TOULOUSE

Jury :

BASTA-LE BERRE Tamara: Maître de conférences, Université Paris-Sud, PARIS Rappoteure

LINDER Patrick: Professeur, Université de Genève, GENEVE Rapporteur

CONDON Ciaran: Directeur de Recherche, CNRS, PARIS Examineur

FRANZETTI Bruno: Directeur de Recherche, CNRS, GRENOBLE Examineur

FICHANT Gwennaële: Professeur, Université de Toulouse, TOULOUSE Présidente

RESUME

Les ribonucléases et les hélicases à ARN sont des acteurs clé du métabolisme des ARN et jouent donc des rôles cruciaux pour la régulation de l'expression des gènes. Peu de données sont connues concernant ce métabolisme chez les Archées, le troisième domaine du vivant. L'équipe dans laquelle j'ai effectué mes travaux de thèse s'intéresse au métabolisme de l'ARN chez les archées et plus particulièrement aux ribonucléases β -CASP. Dans ce contexte, nous focalisons nos études sur la compréhension physiologique que pourrait jouer les ribonucléases β -CASP aCPSF1 et aRNase J, orthologue respectivement du facteur de terminaison de la transcription eucaryotes CPSF-73 et RNase J bactérienne. Par analogie avec CPSF-73 et RNase J, qui font partie de complexes multi-protéiques, des indices sur les fonctions des homologues archéens de ces ribonucléases pourraient provenir de l'identification des complexes autour de aCPSF1 et aRNase J.

Utilisant des extraits de *Pyrococcus abyssi* et les protéines recombinantes aCPSF1 et aRNase J comme appâts, nous avons identifié que aRNase J fait partie d'un réseau d'interaction incluant une hélicase de la famille des Ski2-like (ASH-Ski2). En parallèle, des fractionnements d'extrait de *P. abyssi* sur gradient de saccharose par ultracentrifugation indiquent que aRNase J et ASH-Ski2 sont présentes toutes deux dans les fractions de haut poids moléculaires avec les sous-unités du ribosome et ceux de l'exosome. Nous avons aussi démontré une interaction stable entre aRNase J et ASH-Ski2 ainsi que des motifs impliqués dans cette interaction par des expériences de co-purification par chromatographie d'affinité. De plus, les caractérisations biochimiques de ASH-Ski2 indiquent que cette protéine possède une activité d'hydrolyse de l'ATP dépendant de la présence d'acides nucléiques. ASH-Ski2 possède de plus la capacité d'hybridation et de déroulement de deux brins d'acides nucléiques en présence d'ATP.

A notre connaissance, nos résultats sont les premiers à indiquer un complexe contenant une ribonucléase et d'une hélicase à ARN Ski2-like chez les archées. De manière intrigant, aRNase J est orthologue de la RNase J bactérienne et ASH-Ski2 des hélicases Ski2-like des eucaryotes. Cela démontre que les Archées pourraient posséder un système composite impliqué dans le métabolisme des ARN partageant des caractéristiques bactériens et eucaryotes. Ces résultats mettent en lumière l'avantage de l'étude des Archées pour la compréhension des mécanismes moléculaires et évolutives des processus fondamentaux des trois domaines du vivant.

INTRODUCTION.....	4
A. LES ARCHEES.....	5
A-1. MICRO-ORGANISMES UBIQUITAIRES PROCHES DES EUCARYOTES.....	5
A-2. LES MACHINERIES MOLECULAIRES DES ARCHEES.....	7
A-2-1. La machinerie de répllication	7
A-2-2. La machinerie de transcription.....	11
A-2-3. La machinerie de traduction.....	14
B. LA REGULATION POST-TRANSCRIPTIONNELLE DE L'EXPRESSION GENIQUE VIA LE METABOLISME DE L'ARN.....	15
B-1. GENERALITES SUR LA REGULATION DE L'EXPRESSION GENIQUE.....	15
B-2. ACTEURS DU PROCESSING DES ARN	17
B-2-1. Les ribonucléases.....	17
B-2-2. Les hélicases.....	20
B-3. MACHINERIES DE DEGRADATION DES ARN CHEZ LES BACTERIES ET LES EUCARYOTES	22
B-4. VOIES DE PROCESSING DES ARN CHEZ LES EUCARYOTES ET LES BACTERIES.....	23
B-5. VOIES DE PROCESSING DES ARN DE TRANSFERT, ARN RIBOSOMIQUE ET ARN CRISPR CHEZ LES ARCHEES.....	26
B-5-1. Processing des pré-ARN de transfert.....	26
B-5-2. Processing des ARN pré-ribosomiques.....	30
B-5-3. Processing des ARN CRISPR.....	31
B-6. AUTRES ENZYMES ET MACHINERIES DE DEGRADATION DES ARN CHEZ LES ARCHEES	32
B-6-1. aExosome.....	32
B-6-2. RNase R.....	33
B-6-3. Ribonucléases β -CASP.....	34
C. PROJET DE THESE.....	37
MATERIELS ET METHODES.....	39
RESULTATS.....	50
PARTIE I : IDENTIFICATION ET CLASSIFICATION DES RIBONUCLEASES DE LA FAMILLES DES β-CASP CHEZ LES ARCHEES ET CHARACTERISATION DE aCPSF1 DE <i>P. abyssi</i>.....	51
PUBLICATION 1	53
PUBLICATIONS 2/3.....	54
PUBLICATION 4.....	55
PARTIE II : LES RIBONUCLEASES β-CASP aCPSF1 ET aRNase J DE <i>P. abyssi</i> A L'INTERFACE DU METABOLISME DE L'ARN.....	56
II-1. IDENTIFICATION DES PARTENAIRES ET COMPLEXES MULTI-PROTEIQUES DES RIBONUCLEASES β -CASP de <i>P. abyssi</i>	56
II-1-1. Réseaux d'interaction protéine-protéine des ribonucléases β -CASP aCPSF1, aRNase J et de l'hélicase ASH-Ski2 de <i>Pyrococcus abyssi</i> par la technique de « Pull- down ».....	56

II-1-2. Identification de complexes multi-protéiques par fractionnement d'extraits cellulaires de <i>P. abyssi</i> sur gradient de sucrose.....	65
II-1-3. Analyse du contexte génomique des gènes codant pour les.....	70
II-2. IDENTIFICATION D'INTERACTION PROTEINE-PROTEINE <i>IN VITRO</i> ENTRE PARTENAIRES DES RESEAUX DE <i>Pab</i> -aRNase J, <i>Pab</i> -ASH-Ski2 ET <i>Pab</i> -aCPSF1.....	71
II-2-1. Méthode de co-purification par chromatographie d'affinité.....	72
II-2-2. Interaction protéine-protéine observée par co-élution de la protéine-proie <i>Pab</i> -aRNase J (Figure 24).....	73
II-2-3. Interaction protéine-protéine observée par co-élution de la protéine-proie <i>Pab</i> -ASH-Ski2 (Figure 25).....	75
II-3. IDENTIFICATION ET CARACTERISATION DE LA FAMILLE DES HELICASE ASH-Ski2 CHEZ LES ARCHEES.....	76
II-3-1. Caractérisation de la famille Ski2-like chez les Archées: 2 groupes Hel308-Ski2 et ASH-Ski2.....	76
II-3-2. Co-distribution taxonomique de ASH-Ski2 et aRNase J.....	78
II-4. Caractérisation de l'interaction entre aRNase J et ASH-Ski2 de <i>P. abyssi</i>	79
II-4-1. Interaction de <i>Pab</i> -aRNase J avec l'hélicase <i>Pab</i> -(His) ₆ -Hel308-Ski2.....	80
II-4-2. Interaction de <i>Pab</i> -aRNase J avec les variants de <i>Pab</i> -ASH-Ski2.....	80
II-4-3. Interaction de <i>Pab</i> -aRNase J avec le domaine N-terminal de <i>Pab</i> -ASH-Ski2.....	81
II-4-4. Interaction de <i>Pab</i> -ASH-Ski2 avec le variant de <i>Pab</i> -aRNase J.....	82
II-4-5. Le complexe <i>Pab</i> -ASH-Ski2/ <i>Pab</i> -aRNase J identifié <i>in vitro</i> par sédimentation sur gradient de saccharose.....	82

PARTIE III : ACTIVITES ENZYMATIQUES DES HELICASES DE TYPE

Ski2-like DE *P. abyssi*, *Pab*-ASH-Ski2 ET *Pab*-Hel308-Ski2..... 84

III-1. ACTIVITE ATPASIQUE DE <i>Pab</i> -ASH-Ski2 ET <i>Pab</i> -Hel308-Ski2.....	84
III-1-1. <i>Pab</i> -ASH-Ski2 et <i>Pab</i> -Hel308-Ski2 sont des ATPases dépendantes de la présence d'acides nucléiques.....	84
III-1-2. Activité ATPasique des variants de <i>Pab</i> -ASH-Ski2.....	85
III-2. AFFINITE DE <i>Pab</i> -ASH-Ski2 ET <i>Pab</i> -Hel308-Ski2 AUX ACIDES NUCLEIQUES.....	87
III-2-1. Méthode de mesure de l'affinité aux acides nucléiques par la technique de double filtration.....	87
III-2-2. Affinité de <i>Pab</i> -ASH-Ski2 et de <i>Pab</i> -Hel308-Ski2 pour les acides nucléiques.....	88
III-2-3. Affinité des variants de <i>Pab</i> -ASH-Ski2.....	89
III-3. ACTIVITE HELICASE DE <i>Pab</i> -ASH-Ski2 ET <i>Pab</i> -Hel308-Ski2.....	91
III-3-1. Méthodes pour mesurer la capacité d'appariement de brins et de déroulement de duplex d'acides nucléiques.....	91
III-3-2. Activités d'appariement ou « annealing » de <i>Pab</i> -ASH-Ski2, de ces variants et de <i>Pab</i> -Hel308-Ski2.....	92
III-3-3. Activité de déroulement de duplex ou «unwinding» de <i>Pab</i> -ASH-Ski2, des variants et de <i>Pab</i> -Hel308-Ski2.....	93
III-4. CONCLUSIONS SUR LES ACTIVITES DE <i>Pab</i> -ASH-Ski2 ET <i>Pab</i> -Hel308-Ski2.....	96

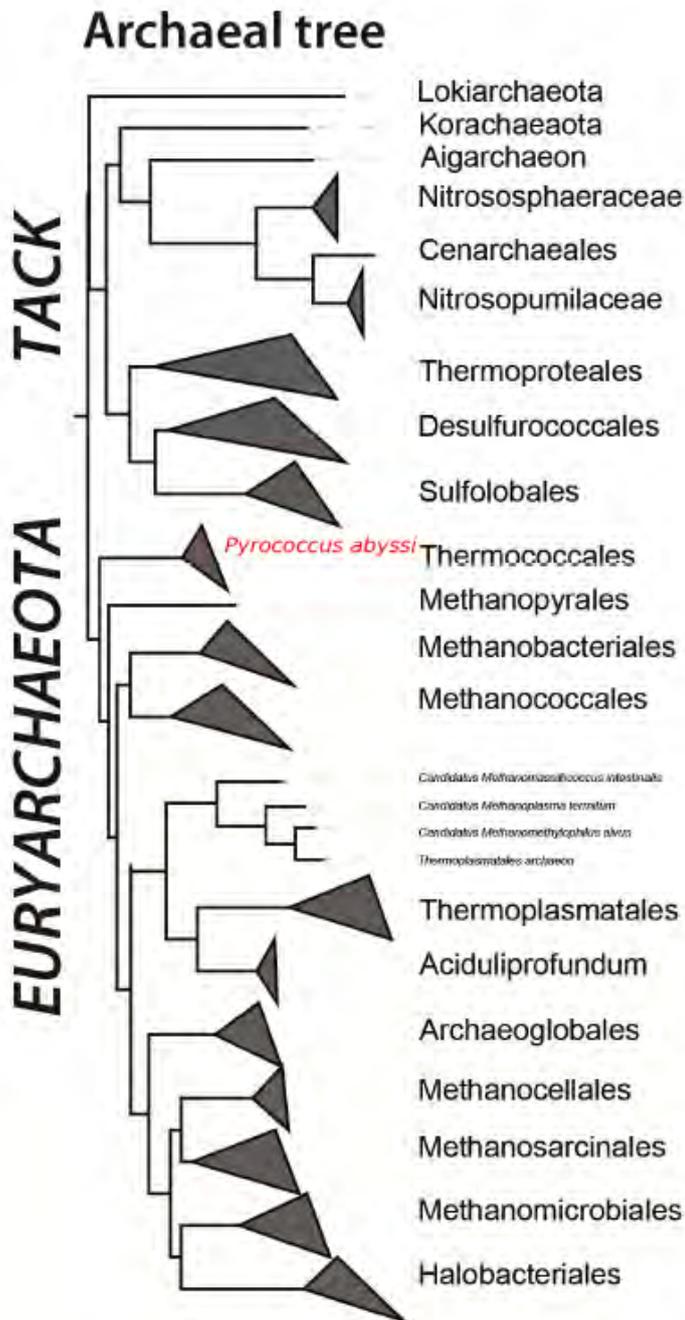
DISCUSSION..... 98

REFERENCES..... 106

ANNEXES..... 128

INTRODUCTION

A



B

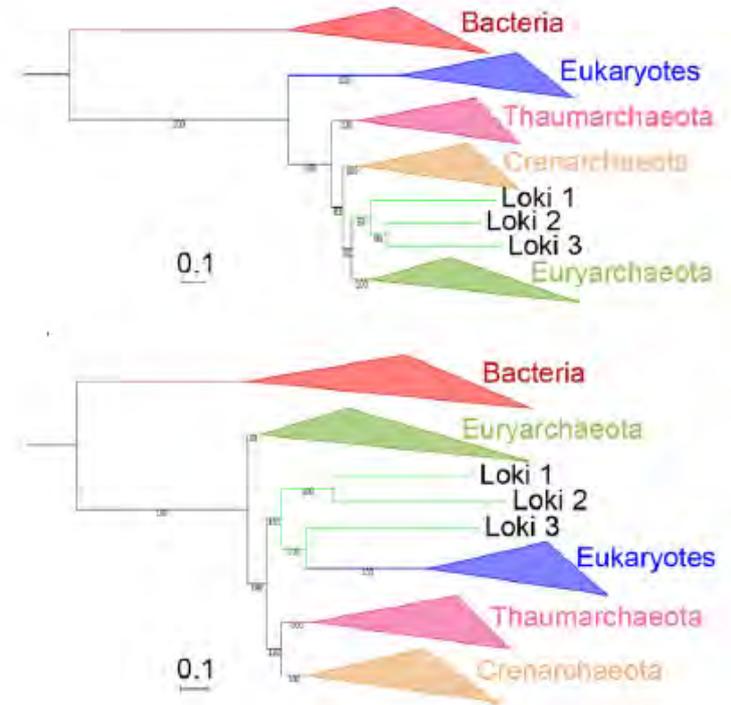


Figure 1: Arbre phylogénétique des archées et deux modèles d'arbre du vivant. A: L'arbre phylogénétique construite par alignement de concaténation de 81 groupes orthologues de protéines conservées issues de 131 génomes d'archée séquencés et annotés (pour méthodes voir Phung *et al*, 2013). Les sous groupes d'un même phylum sont fusionnés. Il est à noter que le groupe des DPANN (nanoarchaea et groupes proches) ne sont pas représentés. *Pyrococcus abyssi* est représenté en rouge au sein des Thermococcales. B: Arbre phylogénétique de la concaténation de 11 protéines de Woese *et al*, 1990 (panel, haut) et de 25 protéines de Spang *et al*, 2016. Figure issue de Da Cunha *et al*, 2017

Au cours de cette partie introductive, les mécanismes moléculaires fondamentaux chez les archées seront présentés succinctement de la réplication de l'ADN, à la traduction des ARN messagers en protéine en passant par la transcription et au «processing» (maturation et dégradation) des ARN. Ces mécanismes seront mis en perspective et seront comparés avec ceux connus chez les bactéries et les eucaryotes. Une grande partie de l'introduction s'attardera sur nos connaissances actuelles sur le métabolisme de l'ARN chez les archées et plus particulièrement aux acteurs moléculaires identifiés et caractérisés à ce jour.

A. LES ARCHEES

A-1. MICRO-ORGANISMES UBIQUITAIRES PROCHES DES EUCARYOTES

Une avancée majeure de la biologie durant la fin du 20^{ème} siècle a été la reclassification des organismes vivants en trois grands groupes : Les Eucaryotes, les Bactéries et les Archées. En se basant sur les séquences des ARN ribosomiques 16S et 18S pour inférer des positions phylogénétiques, les procaryotes, se retrouvent séparés en deux groupes : Bactéries et Archées (Woese and Fox, 1977).

Les archées sont des micro-organismes unicellulaires présentant une morphologie très diversifiée. Initialement, retrouvées dans des conditions extrêmes de pH (proche de 0), de température (de -2.5°C à 113°C), de pression (1200Bar) ou de salinité (jusqu'à 3.9M de NaCl) (OREN *et al.*, 1990; Franzmann *et al.*, 1992; Schleper *et al.*, 1995; Blöchl *et al.*, 1997; Zeng *et al.*, 2009). Les archées sont maintenant reconnues comme des micro-organismes ubiquitaires retrouvés dans les océans, les sols et plus récemment dans les microbiotes humains ((Mihajlovski *et al.*, 2010; Oxley *et al.*, 2010). Leur présence sur notre peau et dans notre système digestif pose la question de leur impact sur notre santé.

A ce jour, les archées sont constituées de quatre phyla majeurs: les DPANN, les Euryarchaea, les TACK et les Asgardarchaea (Lokiarchées) (Zaremba-Niedzwiedzka *et al.*, 2017) (**Figure 1A**). Ces micro-organismes sont des modèles d'intérêt fondamental pour l'étude des mécanismes moléculaires conservés et pour la compréhension de l'histoire évolutive du vivant. En effet les mécanismes moléculaires du système informationnel (réplication, transcription, traduction, recombinaison et réparation de l'ADN) sont proches de ceux décrit chez les eucaryotes (Brochier-Armanet *et al.*, 2011; Lyu and Whitman, 2017). De ce fait des hypothèses sur l'émergence des eucaryotes au sein de la branche Archées de l'arbre du vivant sont proposées et restent très controversées (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*,

2017; Da Cunha *et al.*, 2017) (**Figure 1B**). En effet, depuis la reclassification de Carl Woese du vivant en trois domaines, de nombreuses études essaient de redessiner l'arbre du vivant par différentes approches de comparaison de séquences à partir de données métagénomiques ou de génomes entièrement séquencés et annotés (Hug *et al.*, 2016). Ainsi, à ce jour, deux hypothèses évolutives s'affrontent. L'hypothèse de l'arbre à trois branches propose que les archées et les eucaryotes soient des groupes sœurs partageant un ancêtre commun (**Figure 1B, panel haut**). Cette hypothèse est étayée par des constructions d'arbre à partir de concaténation des protéines de la RNA polymérase et des protéines ribosomiques (Woese and Fox, 1977; Forterre, 2013; Williams *et al.*, 2017). L'autre hypothèse, apparue récemment, propose que les eucaryotes soient apparus au sein des archées et constituent un groupe sœur des Asgardarchées (**Figure B, panel bas**). Les arbres étayant cette hypothèse ont été construits à partir de concaténation de 36 protéines universelles impliquées dans de nombreux processus cellulaires et incluant des protéines de génomes des Asgardarchées nouvellement découvertes, ce qui apporte une information phylogénétique complémentaire à celle des marqueurs informationnels utilisés auparavant tels que les protéines ribosomiques et les sous-unités de l'ARN polymérase (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017).

Les études fonctionnelles chez les archées sont souvent basées sur des études génomiques et protéomiques interprétées en se basant sur nos connaissances chez les bactéries et les eucaryotes. Toutefois, le nombre croissant d'outil de modification des génomes chez certaines souches d'archées permet une approche génétique pour identifier les acteurs impliqués dans les différentes voies métaboliques (Leigh *et al.*, 2011; Atomi *et al.*, 2012; Farkas *et al.*, 2013). Ainsi, des outils génétiques existent chez certaines espèces. C'est le cas chez *Haloferax volcanii* et *Haloferax salinarum*, deux espèces halophiles de la famille des Euryarchées. Il est possible de transformer ces deux souches en introduisant de l'ADN par utilisation du PEG600 après traitement les cellules à l'EDTA afin de débarrasser les surfaces des cellules des glycoprotéines (Cline *et al.*, 1989). Chez *Sulfolobus solfataricus*, une Crénarchées thermophile, par électroporation, il est aussi possible de transformer (Worthington *et al.*, 2003). De plus, il existe des vecteurs d'expression des gènes utilisés dans le but de surexprimer des protéines étiquetées (Albers *et al.*, 2006). Chez les Thermococcales, et notamment chez *Thermococcus barophilus*, des outils permettant de déléter des gènes ont pu être développés. Il est ainsi possible par la méthode du «pop-in/pop-out» d'introduire une

séquence d'intérêt dans le génome en se servant de la simvastatin et du 5-FOA pour sélectionner et contre sélectionner les différentes souches (Thiel *et al.*, 2014).

L'étude des archées est d'un intérêt incontestable. D'une part, d'un point de vue fondamental, elle permet d'apporter des connaissances sur les mécanismes moléculaires archéens en s'appuyant sur des approches phylogénomique et biochimiques pour appréhender l'histoire évolutive entre les trois domaines du vivant et de proposer des modèles d'étude pour clarifier les systèmes eucaryotes et bactériens (Ban *et al.*, 2000). De plus, à partir des archées extrêmophiles, le développement d'outils moléculaires a été possible, c'est l'exemple des polymérase à ADN (Isis, Phusion DNA Polymerase) et plus récemment le marquage spécifique de l'ARN (Tomkuvienė *et al.*, 2012). Le métabolisme cellulaire est aussi un potentiel pour l'industrie chimique avec la production de bio-composants (Blumer-Schuetz *et al.*, 2008; Hess, 2008).

L'archée *Pyrococcus abyssi* a été utilisée comme modèle d'étude dans les travaux présentés dans cette thèse. *P. abyssi* appartenant à l'ordre des Thermococcales du domaine des Euryarchaea est une archée hyperthermophile se développant dans les abysses autour des cheminées hydrothermales à 2000 mètres de profondeur (200 bar) avec une température de croissance optimale de 96°C (Cohen *et al.*, 2003). *P. abyssi* étant un anaérobie strict, sa mise en culture nécessite un équipement en bioréacteur. Le génome de *P. abyssi* est un des premiers génomes d'archées séquencés et annotés (Cohen *et al.*, 2003), c'est pourquoi elle est apparue comme un des modèles de choix pour de nombreuses études en particulier pour comprendre les mécanismes moléculaires de maintenance de l'ADN à une telle température de croissance (Pluchon *et al.*, 2013; Castillo-Lizardo *et al.*, 2014). Son chromosome d'environ 1.7 million de paires de bases est composé de plus de 1800 cadres ouverts de lecture (ORF) avec un GC de 44%. *P. abyssi* possède de plus un plasmide de 3.4 kb constitué de 2 ORF. Il faut cependant noter que les outils génétiques chez cette souche n'ont pas pu être développés contrairement aux Thermococcales du même groupe, *T. kodakarensis* (Santangelo *et al.*, 2008) et plus récemment *T. barophilus* (Thiel *et al.*, 2014).

A-2. LES MACHINERIES MOLECULAIRES DES ARCHEES

A-2-1. La machinerie de réplication

La machinerie de la réplication de l'ADN chez les archées est largement étudiée, en comparaison aux autres systèmes informationnels tels que la machinerie de la transcription et

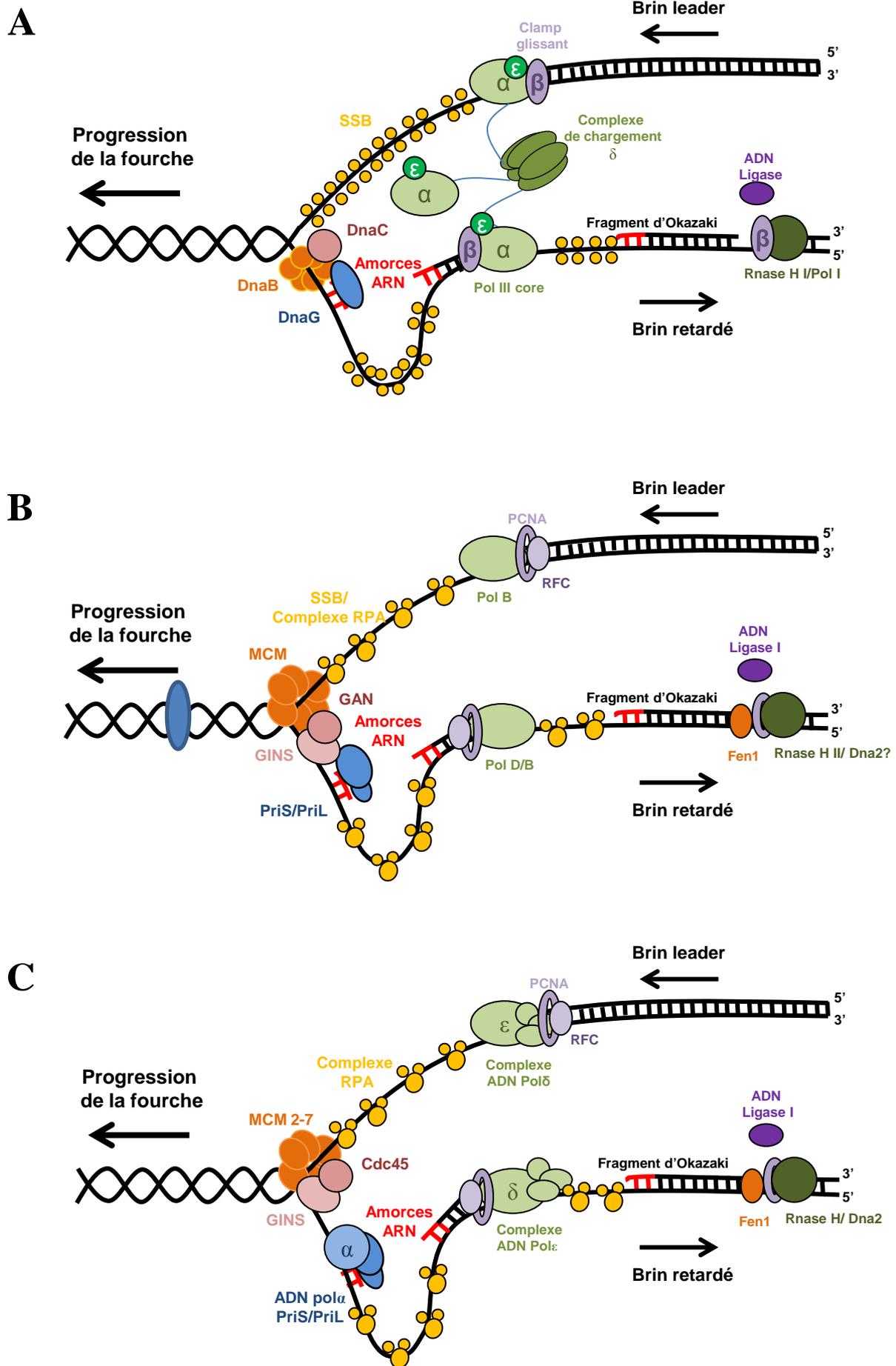


Figure 2: Mécanismes de réplication de l'ADN chez les trois domaines du vivant. L'ADN et l'ARN sont représentés respectivement en noir et en rouge. L'ensemble des protéines décrites ont représentées. Les ADN polymérases et ces cofacteurs sont représentés en vert. Les hélicases sont en orange. Les primases sont en bleu. Les protéines SSB et RPA sont représentées en jaune. La ligase à ADN et les clamps β sont représentés respectivement en violet et violet clair. A : Représentation schématique de la réplication de l'ADN chez les bactéries. B : Représentation schématique de la réplication de l'ADN chez les archées. C : Représentation schématique de la réplication de l'ADN chez les eucaryotes.

de la traduction, qui restent relativement peu caractérisées. En effet, à l'instar des eucaryotes et les bactéries, de nombreux aspects de la réplication sont maintenant bien documentés chez les archées, que ce soit au niveau de la composition, de la structure et de la caractérisation biochimique des complexes multi-protéiques de la machinerie de réplication ou « replisome » (Grabowski and Kelman, 2003; Barry and Bell, 2006; Beattie and Bell, 2011). Le replisome chez les archées est constitué de protéines proches à celles identifiées chez les eucaryotes et constitue un complexe « simplifié » de la machinerie eucaryote fonctionnant sur un petit génome circulaire dont l'architecture et la topologie rappellent celles des bactéries (Grabowski and Kelman, 2003). Il existerait ainsi deux replisomes distincts, celui présent chez les bactéries et celui chez les archées/eucaryotes (**Figure 2**).

Brièvement, chez les bactéries, les chromosomes contiennent une origine de réplication unique constituée d'une région riche en A-T comprenant de multiples copies de boîtes DnaA permettant la fixation de la protéine initiateur de la réplication DnaA. Chez de nombreuses espèces, le gène codant pour DnaA est retrouvé adjacent à l'origine de réplication pour permettre une co-régulation (Marians, 1992). De multiples monomères de DnaA, en se fixant sur les boîtes DnaA, permettent l'ouverture de l'ADN au site *oriC* créant une fourche de réplication. DnaA et une protéine de type AAA+ DnaC sont nécessaires pour le chargement de l'hélicase de réplication DnaB (également de type AAA+) capable de s'assembler sous forme d'un anneau d'homohexamère au site *oriC* (Davey and O'Donnell, 2003). L'interaction entre DnaC en présence d'ATP et l'hexamère de DnaB permettrait d'une part la déstabilisation de l'anneau pour faciliter le chargement de l'hélicase sur le site *oriC* et d'autre part l'augmentation de l'affinité de DnaB pour l'ADN simple brin pour permettre le déroulement de l'ADN double brin au niveau de la fourche (Davey *et al.*, 2002; Davey and O'Donnell, 2003; Arias-Palomo *et al.*, 2013). L'ADN simple brin libéré est protégé par les protéines SSB (Single-Stranded Binding protein). La synthèse des deux brins d'ADN *de novo* par l'ADN polymérase III nécessite la présence de séquences amorces ARN synthétisées par la primase DnaG est qui recrutée au niveau de la fourche de réplication en interagissant avec l'hélicase DnaB. La synthèse d'ADN de 5' vers 3' est continue sur le brin « leader » et discontinue sur le brin retardé (**Figure 2A**). La synthèse du brin « leader » nécessite la fixation d'une des trois sous-unités α catalytique de l'ADN polymérase III par l'intermédiaire de la « β -clamp » à l'ADN matrice. La clamp β permet ainsi la processivité de la réaction de synthèse de l'ADN *de novo*. Chaque sous-unité α de la polymérase possède de plus une sous-unité exonucléase ϵ de « proofreading » permettant l'excision d'un nucléotide incorporé par

erreur à l'extrémité 3' du nouveau brin d'ADN. Lors de la synthèse discontinue le chargement de la clamp β sur la nouvelle amorce permet le recrutement et la synthèse de l'ADN par la sous-unité α . Lorsque la synthèse atteint l'amorce précédente, la sous-unité α se décroche et libère un fragment d'ADN d'environ 1 kb aussi appelé fragment d'Okazaki. Enfin, les fragments d'Okazaki sont reliés entre eux par la ligase à ADN et forment ainsi un ADN continu.

Du fait d'une structure chromosomique bactérienne similaire à celle retrouvée chez les archées, il a été initialement proposé que le chromosome archéen ait une seule origine de réplication (*oriC*). En effet chez *P. abyssi*, une origine unique de réplication avec à proximité les gènes codant pour le complexe de reconnaissance de l'origine ORC (Origin Recognition Ccomplex) a initialement été décrite (Myllykallio *et al.*, 2000). Cependant chez *Sulfolobus solfataricus*, au moins trois origines de réplication ont été identifiées (Robinson *et al.*, 2004; Lundgren *et al.*, 2004). Il semble ainsi que chez les archées, le nombre d'origine de réplication par chromosome soit variable selon les espèces. Dans tous les cas, les protéines en charge de la réplication sont homologues à celles des eucaryotes et possèdent très peu de similarités avec les protéines du réplisome bactérien (Edgell and Doolittle, 1997; Grabowski and Kelman, 2003). Ainsi, la formation des complexes de pré-réplication est initiée par la fixation du complexe de reconnaissance de l'origine ORC (**Figure B et 2C**). Ce complexe est constitué de Orc1 et Cdc6 et permet la reconnaissance de l'origine de réplication et le chargement de l'hélicase réplivative hexamérique MCM (MiniChromosome Maintenance) (Machida *et al.*, 2005). Chez les eucaryotes, l'activation de l'hélicase MCM nécessite le recrutement des protéines GINS et Cdc45 (Makarova *et al.*, 2012). Des homologues du complexe GINS sont aussi retrouvés chez les archées contrairement à Cdc45. Cependant, la fonction de Cdc45 pourrait être assurée par la protéine GAN homologue de la protéine Rec J bactérienne (Li *et al.*, 2011). Après l'ouverture de l'ADN double brin, les protéines du complexe RPA (Replication Protein A) protègent l'ADNsb. Chez les archées, cette fonction est assurée à la fois par les protéines de type SSB chez les crénarchées et les protéines RPA chez les Euryarchées (Komori and Ishino, 2001; Richard *et al.*, 2004). Toutefois, les protéines SSB chez les archées présentent une structures plus proches des protéines RPA eucaryotes que de leur homologues SSB bactériens (Kerr *et al.*, 2003). Une amorce ARN/ADN est synthétisée chez les archées par le complexe primase PriS/PriL et chez les eucaryotes par le complexe primase PriS/PriL/Pol α pour permettre la synthèse d'un pré-fragment d'Okazaki (Lao-Sirieix and Bell, 2004; Lao-Sirieix *et al.*, 2005). La synthèse des brins leader et tardif se

fait simultanément et respectivement par les polymérase PolB et PolB/D chez les archées tandis qu'elle est assurée par les Pol ϵ et δ chez les eucaryotes. La processivité de la synthèse de l'ADN est assurée par la sous-unité catalytique des différentes polymérase qui est reliée à la pince PCNA sous forme d'anneau (Beattie and Bell, 2011; Lindås and Bernander, 2013). Le chargement de PCNA sur l'ADN est aidé par les protéines RFC (Lyu and Whitman, 2017). Lors de la synthèse des brins tardifs, la RNase HIII ainsi que l'endonucléase FEN1 sont recrutées pour de dégrader l'hybride ARN/ADN, permettant à l'ADN ligase I de relier les différents fragments d'Okazaki entre eux (Warbrick, 2000; Vivona and Kelman, 2003).

La RNase H, est une ribonucléase ubiquitaire qui est retrouvée chez les trois domaines du vivant, permet la dégradation des hybrides ARN/ADN formés lors de l'initiation de la réplication de l'ADN. La RNase H est très variable au niveau de la structure et la spécificité de ces substrats (Tadokoro and Kanaya, 2009). Elle possède la capacité de coupure des liaisons phosphodiester par l'intermédiaire de la présence de deux ions métalliques. Basée sur les différences au niveau des séquences des acides aminés, la RNase H chez les bactéries et les archées est classée respectivement parmi deux groupes, RNase HI et RNase HIII (Kochiwa *et al.*, 2007). Sous ces deux formes, elle joue un rôle clé dans la réplication de l'ADN en dégradant les amorces ARN des fragments d'Okazaki et lors des processus de réparation de l'ADN en excisant des erreurs d'incorporation de simple ribonucléotide dans les brins d'ADN (Tadokoro and Kanaya, 2009). La plupart des génomes d'archées possède seulement une seule copie du gène codant pour la RNase HIII. La structure cristallographique résolue chez *Archaeoglobus fulgidus* de la RNase HIII en complexe avec la protéine PCNA montre que la RNase HIII est ancré à PCNA par son motif PIP-box mais reste flexible autour de cette dernière par l'intermédiaire d'un domaine charnière (Bubeck *et al.*, 2011). L'interaction avec PCNA permet l'augmentation à la fois du clivage de l'incorporation par erreur de ribonucléotide et l'hydrolyse des amorces ARN formées durant la maturation des fragments d'Okazaki. Chez *P. furiosus*, la RNase HIII a même été montré capable de dégrader des duplexes ARN/ARN en présence d'ions Mn^{2+} (Kitamura *et al.*, 2010). De plus, il a été observé que la structure 3D de la RNase H de *P. furiosus* est similaire au domaine PIWI de la protéine Ago (connue pour être impliquée dans le « RNA silencing » chez les eucaryotes) alors que leur séquence en acides aminés est très divergente. Les souches *Halobacterium* sp. NRC-1, *Sulfolobus tokodaii* et *Pyrobaculum aerophilum*, en plus de leur copie de RNase HIII, présentent le gène codant pour la RNase HI. Il a été démontré que la RNase HI de *Halobacterium* sp. NRC-1 pouvait cliver une jonction ARN-ADN et celle de *S. tokodaii* présentait une activité de dégradation sur des duplexes ARN/ARN (Ohtani *et al.*, 2004a; Ohtani *et al.*, 2004b).

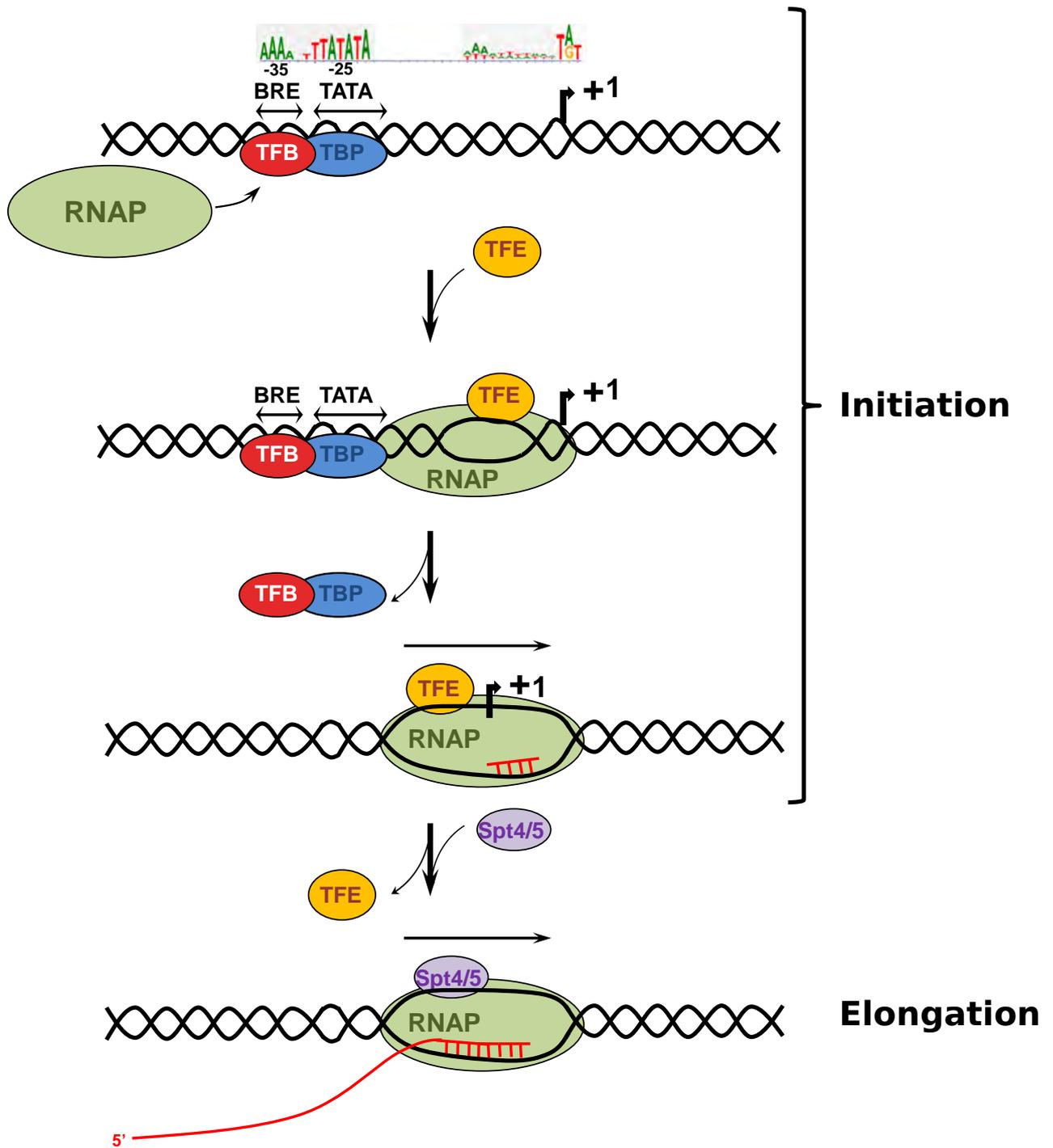


Figure 3: Mécanismes d'initiation et d'élongation de la transcription chez les archées. L'ADN et l'ARN sont représentés respectivement en noir et en rouge. Les positions de l'élément B-reconnaissance (BRE), de la TATA-box et le point +1 de transcription sont représentés. La RNA polymérase est représentée en vert. Les facteurs d'initiation de la transcription TFB (Transcription Facteur B), TBP (TATA-Binding Protein) et TFE sont respectivement représentés en rouge, en bleu et orange. Le facteur d'élongation de la transcription est représenté en violet.

Au cours des cycles cellulaires, l'ADN génomique est susceptible d'être endommagé par différents types de stress. Les facteurs de stress peuvent provenir de l'environnement (agents chimiques, radiations) ou directement issues de processus cellulaires (stress oxydatif, erreur de réplication). Ces dommages à l'ADN peuvent provoquer des cassures doubles brins et entraîner l'arrêt de la fourche de réplication. Les archées possèdent des mécanismes de réparation de l'ADN impliquant la recombinaison homologue ou non homologue (Non homologous end joining ou NHEJ). La recombinaison homologue consiste à la dégradation par le complexe Mre11/Rad50-HerA/NurA d'un des deux brins au niveau de la cassure double brin afin pour former une extrémité simple brin, extrémité qui pourra par l'intermédiaire de la recombinase RadA recombiner avec une séquence d'ADN homologue (Hopkins and Paull, 2008). La recombinaison non homologue implique une ligature des extrémités des brins d'ADN par le complexe NHEJ archée constitué d'une ligase à ADN, d'une polymérase, d'une phosphoestérase et de la protéine Ku (Bartlett *et al.*, 2013).

A-2-2. La machinerie de transcription

La machinerie de transcription chez les archées partage de nombreuses propriétés avec celle des eucaryotes (Bell and Jackson, 1998). Chez les eucaryotes, il existe trois différentes polymérases à ARN : l'ARN polymérase I ou RNAPI pour la synthèse des ARNr ; l'ARN polymérase II ou RNAPII pour celle des ARNm et ARN non codants et l'ARN polymérase III ou RNAPIII pour celle de l'ARNr 5S, des ARNt et certains microARN. La RNAPII est constituée de 12 sous-unités au sein desquelles Rpb4 et Rpb7 sont impliquée dans l'initiation de la transcription. Les autres sous-unités forment le core catalytique de la RNAPII (Armache *et al.*, 2003). Les archées possèdent une unique RNAP dont la séquence en acides aminés et la structure sont très proches de la RNAPII eucaryote (Langer *et al.*, 1995; Hirata *et al.*, 2008). L'ensemble des génomes d'archées possèdent des gènes codant pour des protéines homologues à l'ensemble des 12 sous-unités à l'exception de Rpb9. Rpb8 n'est pas retrouvées chez les Euryarchées, qui possèdent seulement 10 sous-unités de la RNAPII (Korkhin *et al.*, 2009). Par ailleurs, les Crénarchées possèdent une sous-unité Rpo13 spécifique.

La transcription se déroule en trois étapes, l'initiation, l'élongation et la terminaison (**Figure 3**). A l'exception de la terminaison qui reste peu caractérisée, la transcription chez les archées est un mécanisme qui commence à être bien décrit avec l'identification de nombreux facteurs permettant de contrôler les différentes étapes (Werner, 2013; Fouqueau *et al.*, 2017). L'initiation de la transcription requiert trois éléments dans la région promotrice: l'élément B-

recognition (BRE) à la position -35, la séquence TATA box à la position -25 et la séquence initiateur qui correspond au point +1 de la transcription (Wurtzel *et al.*, 2010; Jäger *et al.*, 2014; Babski *et al.*, 2016). Le facteur TFB (Transcription Factor B) et le facteur TBP (TATA-Binding Protein) reconnaissent respectivement la séquence BRE et TATA box pour recruter la RNAP au point +1 de la transcription. Les protéines TBP, TFB et la RNAP forment un complexe de pré-initiation nécessaire et suffisant pour la formation de la bulle de transcription et l'initiation de la synthèse des ARN (Werner and Weinzierl, 2002). Les facteurs TBP et TFB des archées sont respectivement les homologues des facteurs TBP et TFIIB eucaryotes et assurent les mêmes fonctions. TFE (Transcription Factor E), est homologue au facteur THIE eucaryote en se fixant sur la clamp de la RNAP, facilite l'ouverture des brins d'ADN et le chargement du brin matrice par la RNAP (Blombach *et al.*, 2016). La formation de la bulle de transcription par TFE est facilitée par la composition riche en A-T de la région entre l'élément BRE et le point +1 de transcription qui nécessite moins d'énergie pour l'ouverture des deux brins (Smollett *et al.*, 2017).

L'étape d'élongation de la transcription requiert le recrutement du facteur Spt4/5 dont l'homologue bactérien est NusG (Werner and Weinzierl, 2002). La substitution du facteur TFE par le facteur Spt4/5 sur la RNAP par compétition au même site permet le passage de l'initiation à l'élongation de la transcription (Fouqueau *et al.*, 2017).

La RNAP peut parfois, lors de l'élongation, effectuer un « backtracking » pour remédier à un éventuelle « mismatch » et entraîner une pause de la transcription. Ce retour en arrière de la RNAP entraîne la sortie du brin d'ARN néo synthétisée du site actif et forme un complexe non productif (Nudler, 2012). Afin de pallier à cet arrêt et permettre une reprise de la transcription, des homologues de la protéine TFIIS existent chez les archées. La protéine TFIIS eucaryote a pour rôle la stimulation de l'activité endoribonucléolytique la RNAP sur l'ARN « backtracked » pour reformer une extrémité 3'-hydroxyl et permettre une reprise de la synthèse de l'ARN (Schweikhard *et al.*, 2014).

L'étape de terminaison de la réaction de transcription reste énigmatique chez les archées. A ce jour aucun acteur (séquence ADN ou facteur) n'a été clairement identifié comme responsable de l'arrêt de la transcription ou de la dissociation de la RNAP qui permettrait de libérer l'ARN néo-synthétisé. Des candidats peuvent toutefois être proposés au vu des facteurs de terminaison existant chez les bactéries et les eucaryotes (Ciampi, 2006; Clouet-d'Orval *et al.*, 2010; Phung *et al.*, 2013).

Chez les bactéries, deux processus principaux interviennent dans le mécanisme de la terminaison de la transcription: la terminaison Rho-dépendante et Rho indépendante (**Figure 4A**). Lors d'une terminaison Rho-dépendante, l'hélicase Rho (homohexamère), recrutée sur la RNAP lors de l'étape d'élongation, se lie au brin d'ARN sur des séquences spécifiques en amont du terminateur de transcription (**Figure 4A, schéma gauche**). La liaison entre Rho et la RNAP est essentielle pour permettre la séparation de l'ARN néo synthétisé de la matrice ADN par l'activité de translocation de Rho sur l'ARN en présence d'ATP (Epshtein *et al.*, 2010). Lors d'une terminaison Rho-indépendante, l'extrémité 3' de l'ARN forme une structure en épingle à cheveux juste en amont d'une région poly(U) (**Figure 4A, schéma droit**). Cette structure particulière de l'ARN envahit le site catalytique de la RNAP engendrant ainsi des changements de conformation avec une déstabilisation de l'hétéroduplex ARN/ADN (Santangelo and Roberts, 2004; Epshtein *et al.*, 2010).

La terminaison de la transcription par la RNAPoIII chez les eucaryotes se déroule en deux étapes avec une coupure endoribonucléolytique suivie d'une réaction de polyadénylation (**Figure 4B**). Une séquence AAUAAA hautement conservée et une région riche en GU localisée à 30-40 nucléotide en aval du codon stop du pré-ARNm sont les signaux pour le recrutement par la RNAPII du complexe appelé «Cleavage and Polyadenylation Specificity Factor» ou CPSF et du «Cleavage stimulation Factor» ou CstF (Mandel *et al.*, 2008; Proudfoot, 2011; Huimin Zhang *et al.*, 2015). L'affinité du complexe CPSF sur le motif AAUAAA et du complexe CstF sur la région riche en GU a aussi pour conséquence la coupure de l'ARN entre ces deux régions après un dinucléotide CA par l'endoribonucléase CPSF73, un des composants du complexe CPSF. Après coupure, la poly(A) polymérase présente dans le complexe CPSF permet la synthèse de la queue poly(A) stabilisant les ARNm (Kaufmann *et al.*, 2004). Un modèle dit «torpedo» permet de proposer que l'ARN en aval du site de coupure est digéré par les exoribonucléases processives Xrn2 (chez l'humain) ou Rat1 (chez la levure) qui rattrapent la RNAPII pour la libérer de l'ADN matrice (Kim *et al.*, 2004; West *et al.*, 2004). Pour la transcription par la RNAPII des ARN non codants, la terminaison implique l'hélicase Sen1 qui emploie une stratégie similaire à celle de la protéine Rho bactérienne. En présence d'ATP, l'activité de déroulement des acides nucléiques par l'hélicase Sen1 permet la translocation de l'ARN et la déstabilisation de l'hétéroduplex ARN/ADN (Porrua and Libri, 2013).

L'unique RNAP chez les archées est impliquée dans la transcription de l'ensemble des ARN cellulaires (**Figure 6**). Cela englobe les ARN messagers (ARNm), les ARN ribosomiques

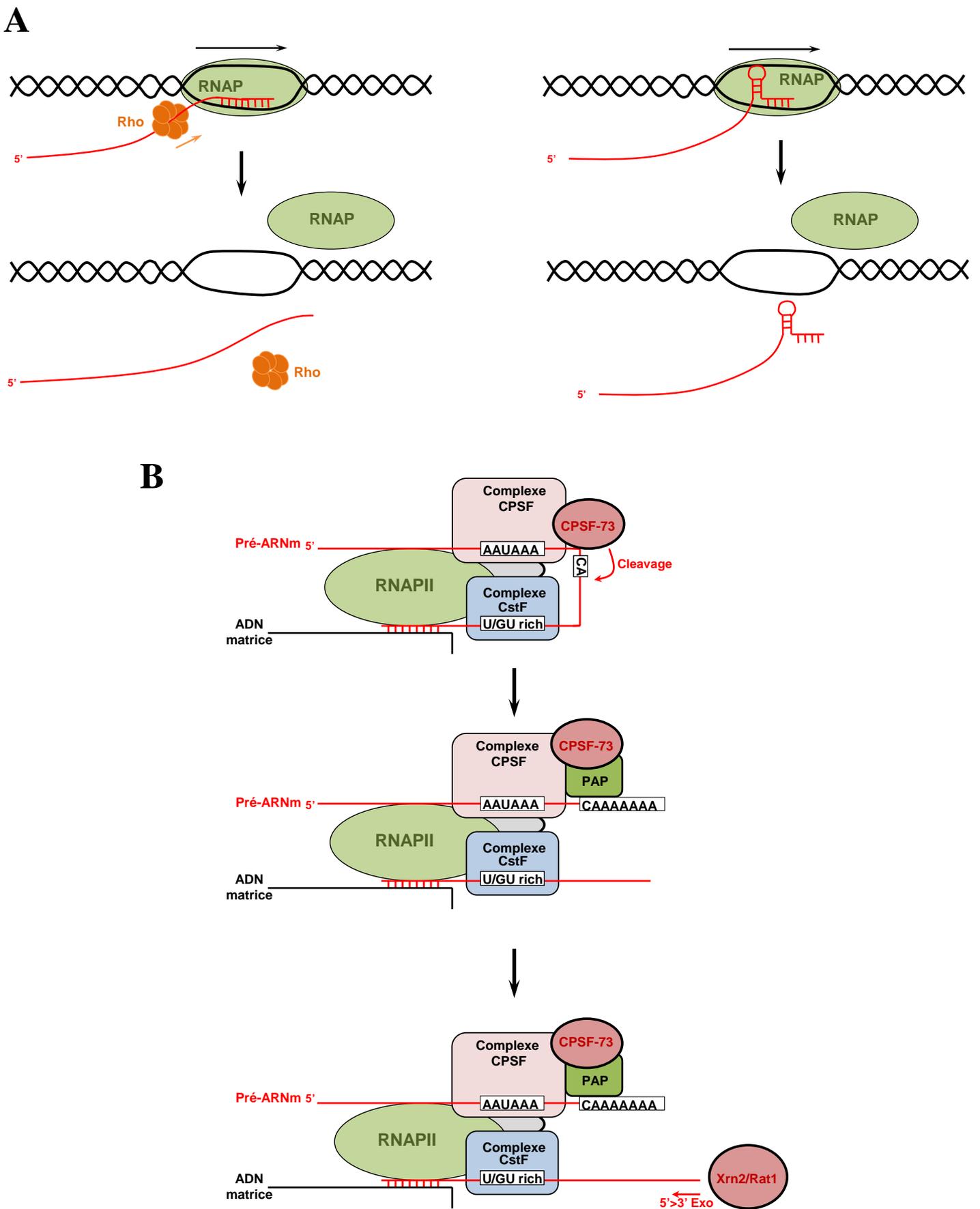


Figure 4: Mécanismes de terminaison de la transcription chez bactéries et les eucaryotes. L'ADN et l'ARN sont représentés respectivement en noir et en rouge. A : Terminaison de la transcription chez les bactéries. Terminaison Rho dépendante (panel gauche) et Rho indépendante (panel droit). L'hélicase hexamérique Rho est représentée en orange. L'ARN polymérase est représenté en vert clair. B : Terminaison de la transcription des ARN messagers par l'ARN polymérase II. La RNAPII est représenté en vert clair. Le complexe CPSF (Cleavage and Polyadenylation Specificity Factor) est représenté en rose. L'endoribonucléase CPSF-73 est représentée en rouge. La Poly(A)polymérase est en vert. Le complexe CstF (Cleavage Stimulation Factor) est représenté en bleu. Les séquences de reconnaissance des complexes CPSF et CstF, respectivement AAUAAA et U/GU rich sont représentées.

(ARNr), les ARN de transfert (ARNt) et divers groupes d'ARN non codant (ARNnc) tels que les ARN guides à boîtes C/D et H/ACA, les ARN qui constitue la RNase P, les ARN CRISPR (ARNcr) et des ARN régulateurs spécifiques. Bien que la transcription soit similaire entre les archées et les eucaryotes, les ARNm des archées ont des propriétés comparables aux ARNm des bactéries. Les transcrits primaires sont tri-phosphorylés sans coiffe, sans introns ni de queue poly(A) et sont souvent polycistroniques. Toutefois, des données suggèrent que les ARNm des archées posséderaient un temps de demi-vie légèrement supérieur à ceux des bactéries qui sont de l'ordre de 40sec à 60 min (Jamie Richards *et al.*, 2008; Evgenieva-Hackenberg and Klug, 2009). En effet, les temps de demi-vie de nombreux transcrits chez *Methanococcus vannielii* varient de 7 à 57 min, de 6,3 à 120 min chez *Sulfolobus solfataricus* et de 4 à 80 min chez *Haloferax mediterranei* (Hennigan and Reeve, 1994; Jäger *et al.*, 2002). Ces études pionnières sur la stabilité des transcrits chez les archées proposent que les transcrits subissent une coupure endoribonucléolytique suivie d'une dégradation exoribonucléolytique 3'-5'. Toutefois, aucun signal de coupure endoribonucléolytique n'a été observé sur ces ARN pouvant supporter ce modèle de dégradation des ARN. Or des études récentes ont montré qu'une sous-unité du facteur d'initiation de la traduction a/eIF2- γ se fixerait à l'extrémité 5' triphosphorylé des transcrits et pourrait les protéger de la voie de dégradation 5'-3' chez *Sulfolobus solfataricus* (Hasenöhr *et al.*, 2008; Arkhipova *et al.*, 2015). Par cette interaction, a/eIF2 pourrait avoir un rôle de protection similaire à la coiffe retrouvée chez les eucaryotes. Ces études suggèrent ainsi l'existence d'une voie de dégradation 5'-3' chez les archées.

A-2-3. La machinerie de traduction

Le ribosome, essentiel à la synthèse protéique, est l'une des machineries macromoléculaires les plus grosses de la cellule. La structure globale du ribosome est conservée chez les trois domaines du vivant. Le ribosome archée a une taille similaire à celui retrouvé chez les bactéries avec un coefficient de Svedberg de 70S composé des ARN ribosomiques 23S et 5S pour la sous-unité 50S et des ARNr 16S pour la sous-unité 30S ainsi que d'une panoplie de petites protéines ribosomiques. Bien que la taille des ARNr du ribosome d'archée soit similaire à celle des ARNr des bactéries, la particule ribosomique d'archée semble plus complexe et est plus proche globalement du ribosome eucaryote composé des sous-unités 40S (ARNr 18S) et 60S (28S, 5S et 5,8S). En effet, le nombre de protéines ribosomiques est de 57, 68 et 78 respectivement chez les bactéries, les archées et les eucaryotes (Londei, 2007). Parmi ces différentes protéines, 34 sont strictement conservées à travers les trois domaines du

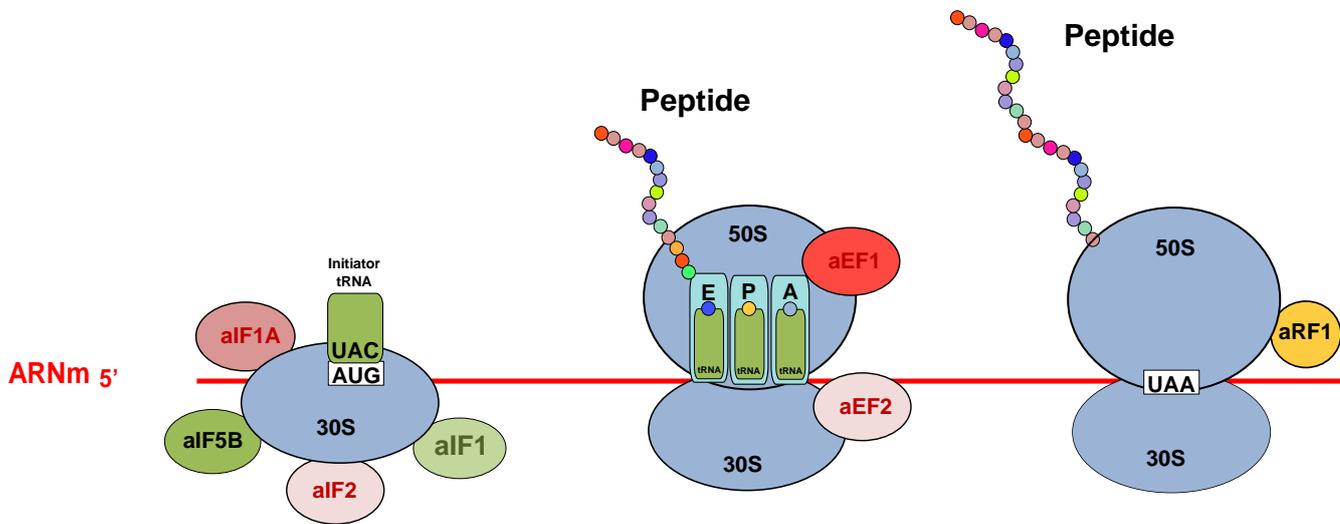


Figure 5: Mécanismes de la traduction chez les archées. Les trois étapes de la traduction sont illustrés. Les différentes sous-unités du ribosome (30S et 50S) et l'ARN sont représentés respectivement en bleu et en rouge. Les ARN de transfert sont représentés par des rectangles verts, les acides aminés et les peptides par des cercles. Les facteurs d'initiation, d'élongation et de terminaison de la traduction sont représentés.

vivant. Celles des archées sont plus proches de celles des eucaryotes par rapport aux bactéries (Yutin *et al.*, 2012). Les archées possèdent une protéine ribosomale unique à eux et partagent exclusivement avec les eucaryotes 33 autres protéines tandis que les bactéries en possèdent 23 qui ne sont pas retrouvées ni chez les archées, ni chez les eucaryotes (Lecompte *et al.*, 2002). Ainsi, cette composition particulière des ribosomes dans les trois domaines du vivant suggère que les archées possèdent une version simplifiée d'un ribosome de type eucaryote. Cette hypothèse est étayée par le comportement similaire des ribosomes archées (*S. solfataricus*, *T. acidophilum* *T. celer*) et eucaryotes face à des antibiotiques (aminoglycosides) ciblant la synthèse des protéines (Cammarano *et al.*, 1985).

Les facteurs de traduction sont bien mieux étudiés chez les archées. Très proches de leurs homologues eucaryotes, ces protéines interviennent dans les mêmes processus lors des différentes étapes de la traduction. Il a été ainsi proposé que le système de traduction des ARN en protéines comprenant les ribosomes et leurs facteurs de traduction des eucaryotes puisse avoir des origines chez les archées (Hartman *et al.*, 2006). Ainsi la traduction est constituée de différentes étapes, l'initiation, l'élongation, la terminaison et le recyclage du ribosome (**Figure 5**). Durant l'initiation avec l'aide des facteurs d'initiation de la traduction, le ribosome identifie le codon «start» sur l'ARN messager et permet le recrutement de l'ARN de transfert initiateur. A ce jour, les protéines connues pour être impliquées dans ce processus sont les facteurs aIF2, aIF1, aIF1A et aIF5B (Hasenöhr *et al.*, 2009; Schmitt *et al.*, 2010). Lors de l'élongation, l'ARNt entre sur le site A de la grande sous-unité du ribosome et reconnaît le codon sur l'ARNm avec l'aide du facteur d'élongation aEF1 (**Figure 5**). S'ensuit une trans-peptidation de l'acide aminé porté par l'ARNt sur la chaîne peptidique sur le site P. La translocation de trois nucléotides par le ribosome permet le passage au codon suivant et est assistée par le facteur aEF2. Enfin, la terminaison de la traduction consiste à la reconnaissance du codon «stop» le relargage de l'ARNm et de la protéine néo-synthétisée et le désassemblage des deux sous-unités du ribosome. Cette étape est assistée par le facteur de terminaison aRF1.

B. LA REGULATION POST-TRANSCRIPTIONNELLE DE L'EXPRESSION GENIQUE VIA LE METABOLISME DE L'ARN

B-1. GENERALITES SUR LA REGULATION DE L'EXPRESSION GENIQUE

La régulation de l'expression génique conditionne précisément et temporellement le développement de chaque cellule. Il existe trois types de régulation de l'expression des

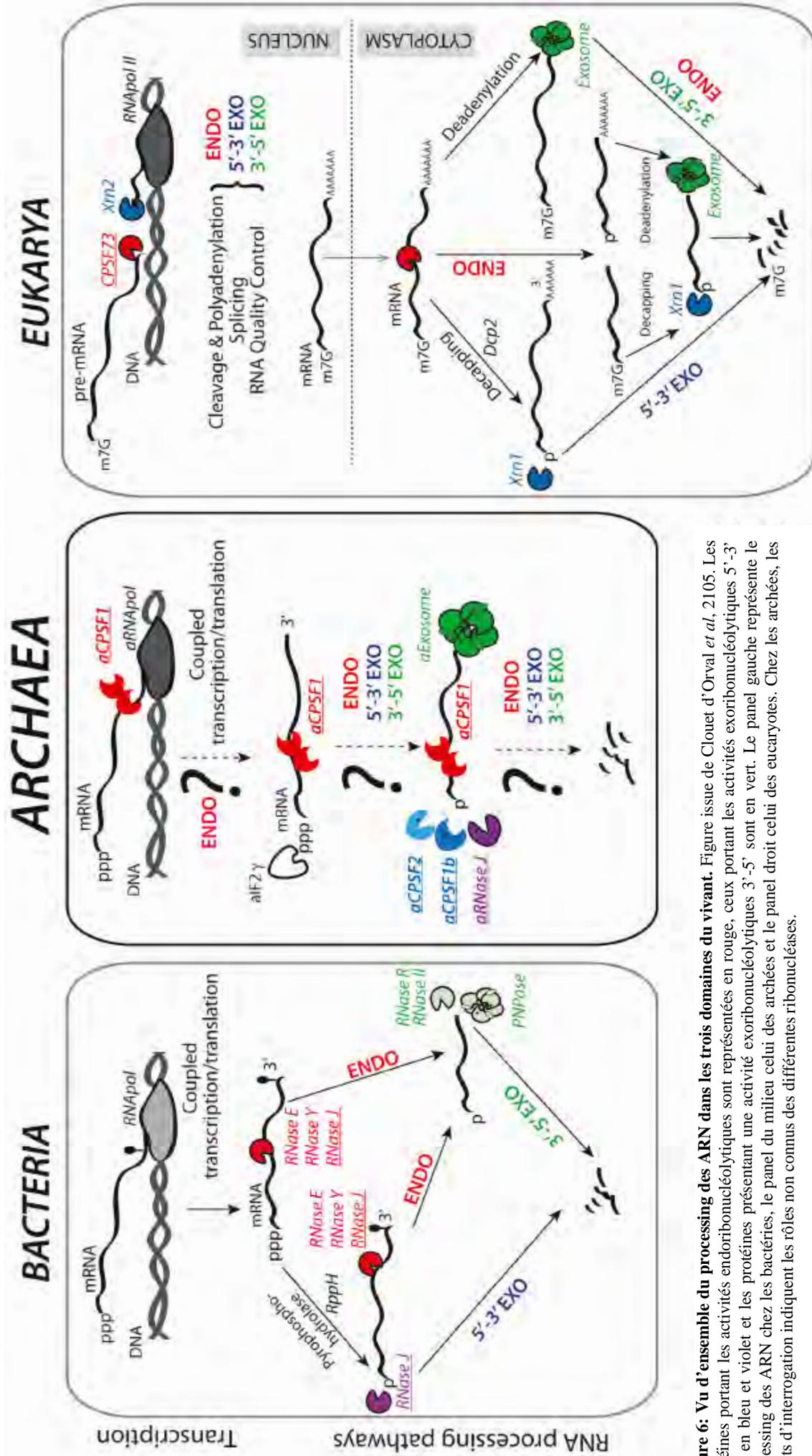


Figure 6: Vu d'ensemble du processing des ARN dans les trois domaines du vivant. Figure issue de Clouet d'Orval *et al.*, 2105. Les protéines portant les activités endoribonucléolytiques sont représentées en rouge, ceux portant les activités exoribonucléolytiques 5'-3' sont en bleu et violet et les protéines présentant une activité exoribonucléolytiques 3'-5' sont en vert. Le panel gauche représente le processing des ARN chez les bactéries, le panel du milieu celui des archées et le panel droit celui des eucaryotes. Chez les archées, les points d'interrogation indiquent les rôles non connus des différentes ribonucléases.

gènes : les régulations transcriptionnelles, les régulations post-transcriptionnelles et les régulations traductionnelles. Ces différents niveaux de régulation chez les eucaryotes et les bactéries sont complexes et ont été largement étudiés. (Halbeisen *et al.*, 2008; Richard J. Jackson *et al.*, 2010; Picard *et al.*, 2013). Cependant, chez les archées, ces processus de régulation commencent juste à être étudiés.

Différents aspects de la régulation transcriptionnelle sont connus chez les archées. L'ensemble des gènes est régulé positivement par les facteurs d'initiation de la transcription (TBP et TFB) et d'élongation (TFE et Spt4/5) (**Figure 3**). La transcription peut aussi être régulée négativement chez les archées, notamment via la protéine TrpY, qui se fixe sur l'ADN et peut changer sa topologie en créant un superenroulement de l'ADN déplaçant ainsi les facteurs d'initiation de la transcription TBP et TFB (Karr *et al.*, 2008). Des facteurs de transcription plus spécifiques existent tels que les protéines Lrp sont impliqués dans la synthèse des composant du métabolisme des acides aminées et peuvent à la fois activer ou inhiber la transcription (Peeters and Charlier, 2010).

Comme mentionné auparavant, la régulation traductionnelle chez les archées commence à être décrite (**Figure 5**). Cette régulation est critique lors de l'étape d'initiation qui est déterminante pour le taux de synthèse des protéines grâce notamment à la présence de nombreux facteurs d'initiation similaire à ceux existant chez les eucaryotes.

La régulation post-transcriptionnelle implique le métabolisme des ARNn un processus clé qui permet après la transcription de former et/ou de dégrader des molécules d'ARN matures et fonctionnelles tout en contrôlant leur qualité dans un contexte cellulaire spatio-temporel précis. Nous nous intéressons dans ce travail de thèse tout particulièrement à la compréhension des mécanismes moléculaires impliqués dans le métabolisme de l'ARN (**Figure 6**).

La synthèse et la dégradation de tous les ARN (ARNm, ARNt, ARNr, ARNnc) incluent des voies de maturation, de modification et de contrôle-qualité appelées plus généralement voies de « processing » des ARN. Ces voies de régulation font intervenir des mécanismes moléculaires complexes faisant appel à une panoplie d'enzymes qui incluent des ribonucléases, des hélicases, des protéines de liaison aux ARN et des enzymes de modification de ribose, de base et de la coiffe des ARN en 5'. La plupart de nos connaissances sur le métabolisme des ARN sont issues de modèles d'études bactériens tels que *E. coli*, *B. subtilis*, *S. aureus* et *H. pylori* (Carpousis *et al.*, 1994; Mathy *et al.*, 2007; Redko *et al.*, 2013)

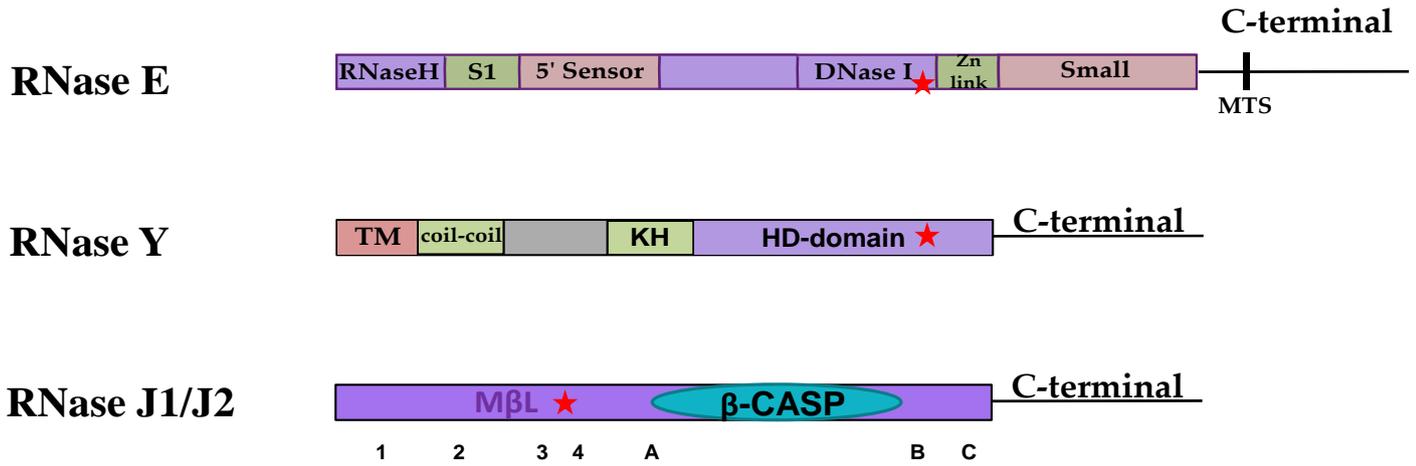
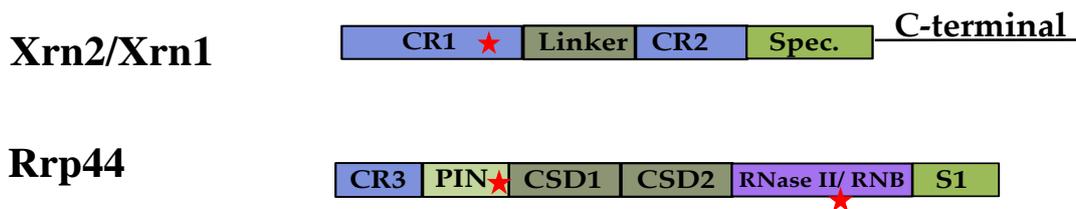
A**B**

Figure 7: Architecture de certaines ribonucléases chez les bactéries et les archées. A: Représentations schématiques de la RNase E, de la RNase Y et de la RNase J bactériennes. Les différents domaines des protéines sont représentés et leur fonction sont indiquées dans le texte (voir A-2-3). Les domaines catalytiques sont indiqués par les étoiles rouges. B: Représentation des ribonucléases Xrn2/Xrn1 et Rrp44 du complexe exosome chez les eucaryotes.

et eucaryotes (levure et cellules humaines) (Mandel *et al.*, 2008; Tomecki *et al.*, 2010) (**Figure 6**). Avant d'aborder l'état de nos connaissances sur les mécanismes moléculaires et les acteurs du métabolisme de l'ARN chez les archées, nous ferons brièvement un état de l'art des acteurs associés et des machineries de dégradation des ARN décrites chez les eucaryotes et les bactéries.

B-2. ACTEURS DU PROCESSING DES ARN

B-2-1. Les ribonucléases

Les RNases sont en général des enzymes ayant une activité catalytique permettant de fragmenter l'ARN de manière plus ou moins spécifique pour les endoribonucléases ou de les rogner à partir des extrémités (5' vers 3' ou 3' vers 5') pour les exoribonucléases. Les RNases peuvent jouer des rôles clés dans les voies de maturation, de contrôle qualité et de la dégradation des molécules d'ARN (Arraiano *et al.*, 2013; Bandyra *et al.*, 2013) (**Figure 6**). Ces protéines sont souvent retrouvées au cœur de complexes macromoléculaires avec d'autres enzymes et partenaires protéines. Les RNases sont ainsi des acteurs essentiels pour la maturation de précurseurs d'ARN (ARNt ou ARNr). Elles jouent des rôles cruciaux au sein des systèmes de contrôle-qualité des ARN défectueux et permettent une protection contre les invasions de génomes viraux et d'éléments transposables. Les RNases sont finalement des effecteurs des voies de signalisation contrôlant l'expression des gènes au niveau post-transcriptionnelle. Ces enzymes peuvent agir soit sur de l'ARN simple brin ou soit sur l'ARN double brin. Les ribonucléases sont classées parmi deux grands groupes, les hydrolases et les phosphorylases.

Les ribonucléases de la famille des hydrolases catalysent les réactions d'hydrolyse et nécessitent la présence de molécules d'eau afin d'attaquer les liaisons phosphodiester entre les bases des ARN. Les protéines de cette famille nécessitent le plus souvent la présence d'ion divalent afin d'être activées. Au sein de cette famille, nous pouvons citer comme exemples la RNase III, une endoribonucléase double brin, impliquée dans la maturation des ARNr (Wu *et al.*, 2000). La RNase H qui est impliquée dans la réplication de l'ADN par hydrolyse des primers des fragments d'Okazaki (Ogawa and Okazaki, 1984). Les RNase J et RNase E chez différentes bactéries sont impliquées dans la dégradation des ARNm, la maturation des ARNr et des ARNt (Carpousis, 2007; Britton *et al.*, 2007; Linder *et al.*, 2014).

Les phosphorylases contrairement aux hydrolases ne requièrent pas de molécules d'eau pour être activées. Les protéines de cette famille catalysent l'ajout d'un groupement phosphate issue de phosphate inorganique entre les bases de l'ARN entraînant la formation de NDP. Nous pouvons citer l'exemple de la PNPase présente dans le complexe dégradosome à ARN chez *E. coli* (Carpousis *et al.*, 1994). La RNase PH est impliquée dans la maturation de l'extrémité 3' des ARNt (Wen *et al.*, 2005) ou encore les sous-unités de l'exosome eucaryotes (Mitchell *et al.*, 1997).

A ce jour il est répertorié plus de 30 familles de ribonucléases dans les trois domaines du vivant avec au moins 20 et 60 ribonucléases différentes retrouvées respectivement chez les bactéries et les eucaryotes (Stoecklin and Mühlemann, 2013). Ce nombre reflète la diversité des fonctions à accomplir par ces enzymes dans la cellule.

Ainsi, chez les bactéries, la RNase E présente chez *E. coli* est une ribonucléase essentielle et permet d'initier la dégradation de l'ensemble des ARNm. En effet, il a été montré que dans un mutant d'*E. coli* portant une mutation thermosensible sur le gène *rne* codant la RNase E, le temps de demi vie de près de 60% des ARNm est affecté (Stead *et al.*, 2011). La RNase E est aussi décrite comme étant impliquée dans la maturation des ARN r (9S et 16S), des ARNt et du ribozyme, ARN M1, composant de la RNase P (Li *et al.*, 1999; Ow and Kushner, 2002; Ko *et al.*, 2008; Mackie, 2013). La RNase E est une endoribonucléase à simple brin avec une préférence pour les régions riches en A/U. Au niveau structural, elle est constituée d'un core catalytique très conservé et essentiel pour la viabilité de la cellule (Kaberdin *et al.*, 1998) (**Figure 7A**). Ce core est constitué d'un domaine S1, d'un RNase H-domain non actif, d'un domaine DNase I et un d'un domaine 'Small domain' responsable de sa tétramérisation (**Figure 7A**). Dans sa partie C-terminale, peu conservée, des micros domaines ont été caractérisés comme étant essentiels pour l'interaction de la RNase E avec d'autres protéines impliquées dans le métabolisme de l'ARN (RhlB, PNPase, Enolase...) et former le complexe dégradosome à ARN (**Figure 9A**) (Carpousis, 2007). La présence d'une hélice amphipatique nommée MTS (membrane targeting sequence) permet son ancrage à la membrane interne des cellules d'*E. coli* (Khemici *et al.*, 2008).

Chez les bactéries gram-positives, qui ne présentent pas de RNase E, c'est la RNase Y qui est la protéine principale de la dégradation des ARN simples brins par des coupures endoribonucléolytique. Il a été montré que la délétion du gène codant pour la RNase Y chez *B. subtilis* entraîne une augmentation du temps de demi-vie des ARNm (Durand *et al.*, 2012).

Comme la RNase E, la RNase Y possède une préférence de coupure au niveau de région riche en A/U et des substrats 5' monophosphate. Cette protéine est constituée d'un petit domaine N-terminal transmembranaire, un domaine coil-coil pouvant être impliqué dans l'oligomérisation, d'un domaine KH impliqué dans l'interaction avec les ARN et un domaine HD portant l'activité catalytique (Lehnik-Habrink *et al.*, 2011) (**Figure 7A**).

Chez les bactéries, l'activité 5' vers 3' exoribonucléolytique est principalement portée par la protéine RNase J1 et dans la moindre mesure par son paralogue RNase J2. La particularité de RNase J1 est sa capacité à présenter aussi une activité endoribonucléolytique avec un même site catalytique (Li de la Sierra-Gallay *et al.*, 2008). Ces protéines, présentes chez les Firmicutes, sont impliquées dans la dégradation de l'ensemble des ARNm et sont impliquées dans la maturation des ARN ribosomique 16S (Britton *et al.*, 2007; Mathy *et al.*, 2007). RNase J fait partie des ribonucléases de la famille des β -CASP qui sont présentes dans les trois domaines du vivant et sont largement présentes chez les archées (**Figure 7A**). Une partie de l'introduction sera dédiée à la description de cette famille de protéines.

Au niveau du noyau des eucaryotes, la protéine Xrn2/Rat1 est connue pour être impliquée dans la dégradation des pré-ARNm et joue un rôle primordial lors de la terminaison de la transcription des ARNm pour dégrader l'ARN en aval de la coupure par CPSF73 et libérer ainsi la RNAPol II dans le « Torpedo model » (Kim *et al.*, 2004; West *et al.*, 2004) (**Figure 4**). Xrn2/Rat1 possède au niveau structural deux régions conservées CR (CR1 et CR2) espacées d'un linker. L'activité catalytique serait portée par le domaine CR1 présentant des résidus conservés de liaison aux ions divalents (**Figure 7B**). La protéine présente aussi un domaine spécifique et un domaine C-terminal non structurés et peu conservés. Dans le cytoplasme, existe un homologue de cette protéine nommé Xrn1 impliquée dans la dégradation des ARNm et présente les mêmes caractéristiques au niveau structural (Long and McNally, 2003).

Dans le cytoplasme des cellules eucaryotes, la protéine Rrp44 joue un rôle concerté dans la dégradation et le control de la qualité des ARN avec le complexe exosome (Chlebowski *et al.*, 2013). Rrp44 présente un domaine RNB portant une activité exoribonucléolytique, deux « Cold Shock Domain », un PIN domain portant une activité endoribonucléolytique juxtaposé d'un domaine CR3 composé de cystéines conservées et un domaine S1 (**Figure 7B**). Rrp44 présente ainsi une double activité ribonucléolytique sur deux sites différents.

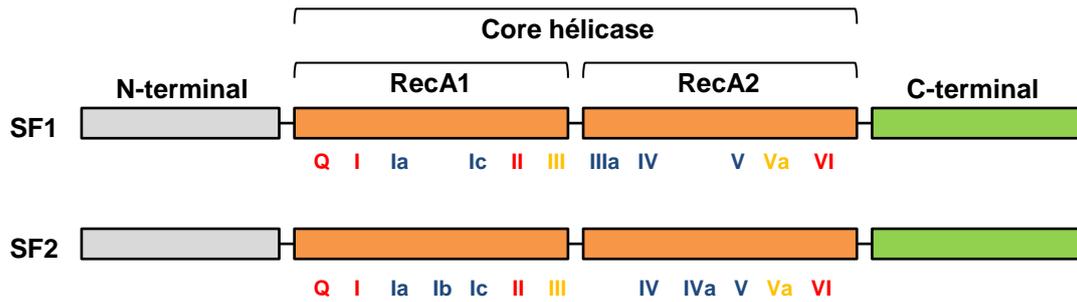
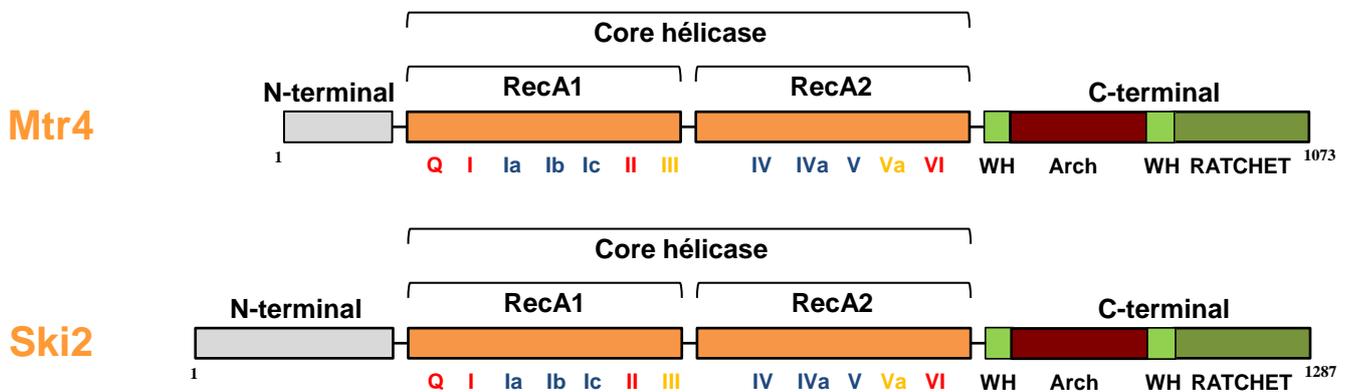
A**B**

Figure 8: Les hélicases des superfamilles 1 (SF1) et 2 (SF2). A : Représentation schématique des différents domaines des hélicases SF1 et SF2. Le core SF2 constitué des domaines RecA1 et RecA2 sont représentés en orange. Les extensions N-terminales sont représentées en gris et C-terminales en vert. Les différents motifs impliqués dans la liaison et l'hydrolyse de l'ATP sont en rouge. Ceux impliqués dans la coordination avec les acides nucléiques et les sites de fixation des NTP sont en jaune. En bleu sont représentés les motifs impliqués dans la liaison avec les acides nucléiques. B : représentation schématique des hélicases de la famille des Ski2. Les différents domaines et motifs sont représentés comme pour A. Mtr4 et Ski2 possèdent, en plus des motifs WH et Ratchet typique des Ski2, un Arch domain représenté en marron.

B-2-2. Les hélicases

Les hélicases constituent une catégorie de protéines motrices consommatrices d'énergie sous forme de nucléotide triphosphates, le plus souvent de l'ATP, et ont pour substrats les acides nucléiques. Toutes les hélicases sont ainsi des P-loop NTPases et possèdent des sites typiques de fixation et d'hydrolyse de NTP, Walker A et B (Fairman-Williams *et al.*, 2010). Ces protéines sont impliquées à la fois dans le métabolisme de l'ADN et de l'ARN et permettent le déroulement des brins d'acide nucléique ainsi que leur translocation, le déplacement des complexes nucléoprotéiques et le remodelage de la chromatine (Singleton *et al.*, 2007; Pyle, 2008; Linder and Jankowsky, 2011). Elles jouent ainsi des rôles cruciaux aussi bien dans la réplication, la recombinaison et la réparation de l'ADN que dans la maturation et la dégradation des ARN et l'initiation de la traduction. L'ensemble des hélicases sont réparties en six superfamilles (SF1 à SF6) selon leur homologie de séquence d'acide aminé, la polarité de leur activité enzymatique 5'-3' ou 3'-5' et leur structure (Gorbalenya and Koonin, 1993; Singleton *et al.*, 2007; Berger, 2008). Il existe ainsi deux types d'hélicase, celles pouvant former des hexamères toroïdaux (SF3 à SF6) et celles n'en formant pas (SF1 et SF2). Toutefois, des protéines classées comme étant des hélicases par homologie de séquences peuvent de manière générale hydrolyser de l'ATP en présence d'acides nucléiques mais ne possèdent pas systématiquement la capacité à dérouler les duplexes d'ADN ou ARN. C'est le cas de la DEAD-box eIF4A-III, une hélicase ATP-dépendante jouant un rôle d'échafaudage et qui est impliquée dans l'épissage des pré-ARNm et la dégradation des ARNm non-sens (Shibuya *et al.*, 2006).

Les hélicases appartenant aux superfamilles SF1 et SF2 possèdent toutes des motifs conservés caractéristiques au sein de deux domaines structuraux (nommés « DEAD » et « Helicase C » dans la base de données Pfam) (**Figure 8A**). Chacun de ces deux domaines peut se replier et former des structures quasi identiques de type RecA-like (RecA1 et RecA2 composés de 5 feuillets β centraux entourés de 4 hélices α) juxtaposés reliés par un linker et constituent le core catalytique hélicase de la protéine. Au sein de ce core, des motifs conservés Q, I à VI sont présents. Les motifs I et II qui correspondent aux motifs Walker A et B sont conservés et sont impliqués, avec les motifs Q et VI, dans la fixation et l'hydrolyse de NTP (Fairman-Williams *et al.*, 2010). Les autres motifs, plus ou moins conservés entre les SF1 et SF2, sont caractéristiques et participent à la formation du site catalytique. Nous les retrouvons aussi impliqués dans la liaison aux acides nucléiques (**Figure 8A**). La plupart des hélicases possèdent des extensions N-terminale et/ou C-terminale de part et d'autre du core hélicase

caractérisant leur fonction par la liaison à des partenaires protéiques ou par la fixation d'acides nucléiques spécifiques. Ainsi ces extensions aux fonctions diverses sont fondamentales pour l'action ciblée des hélicase à ARN (Jankowsky, 2011).

Chez les bactéries et les eucaryotes, les hélicases SF1 et SF2 ont été classées par des analyses phylogénétiques et les sous-familles ont été définies selon leurs séquences et leurs caractéristiques structurales et mécanistiques (Jankowsky, 2011). Ainsi, la superfamille SF1 regroupe 3 sous-familles, UvrD/REP, Pif1 et Upf1. La plupart des hélicases de la superfamille SF1 sont impliquées dans le métabolisme de l'ADN (Dillingham, 2011; Raney *et al.*, 2013). La superfamille SF2 est quant à elle divisée en 9 sous-familles, RecG, RecQ, Rad3/XPD, Ski2, T1R, Swi/Snf, RIG-I, DEAD-box et DEAH/RHA. Les hélicases SF2 possèdent un large spectre d'activité et sont impliquée dans la quasi-totalité des métabolismes de l'ARN et de l'ADN.

Les hélicases de la famille DEAD sont les plus retrouvées chez les eucaryotes avec 37 membres chez les humains et 26 chez *S. cerevisiae* (Fairman-Williams *et al.*, 2010). Chez les bactéries on en dénombre 5 membres chez *E. coli* (Fairman-Williams *et al.*, 2010). Les protéines de cette famille possèdent toutes les résidus Asp-Glu-Ala-Asp (DEAD) caractéristiques du motif 2. Les hélicases de cette famille sont impliquées dans tous les aspects du métabolisme de l'ARN, que ce soit au sein du spliceosome avec la protéine Prp28 chez les eucaryotes, dans l'initiation de la traduction avec eIF4A ou dans des complexes protéines dédiés à la dégradation des ARN avec RhlB, CshA et RphA chez les bactéries (voir dans le chapitre suivant) (Kalman *et al.*, 1991; Redko *et al.*, 2013; Giraud *et al.*, 2015).

Bien que les hélicases de la famille Ski2-like nommée ainsi d'après le premier représentant (Ski2p de *S. cerevisiae*) soient impliquées dans des voies fondamentales du métabolisme des ARN chez les eucaryotes, elles forment une famille relativement petite au sein de la superfamille SF2. Elles interviennent dans la dégradation, la maturation des ARN et dans des voies d'épissage chez les eucaryotes par leur activité de déroulement des duplexes d'acides nucléiques avec une polarité de 3' vers 5' (**Figure 8B**). Elles ont des représentants chez tous les eucaryotes et les archées et sembleraient sensiblement absentes chez les bactéries (Johnson and Jackson, 2013; Chamieh *et al.*, 2016). Les hélicases Mtr4, Ski2 et Suv3 sont souvent au cœur de larges complexes multi-protéiques dédiés à la dégradation des ARN cellulaires (Johnson and Jackson, 2013) (voir partie II-3 ci-dessous). L'hélicase Brr2 fait aussi partie de la famille des Ski2 et a pour particularité d'être formée de deux cores hélicase SF2 juxtaposés

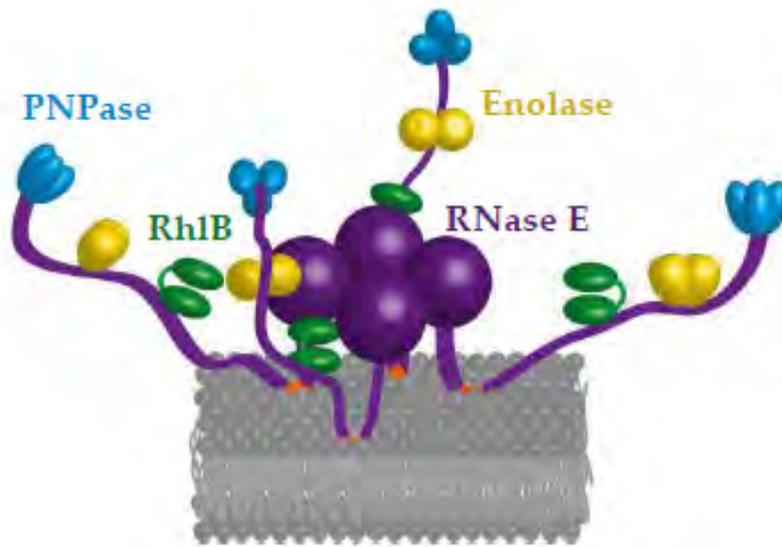
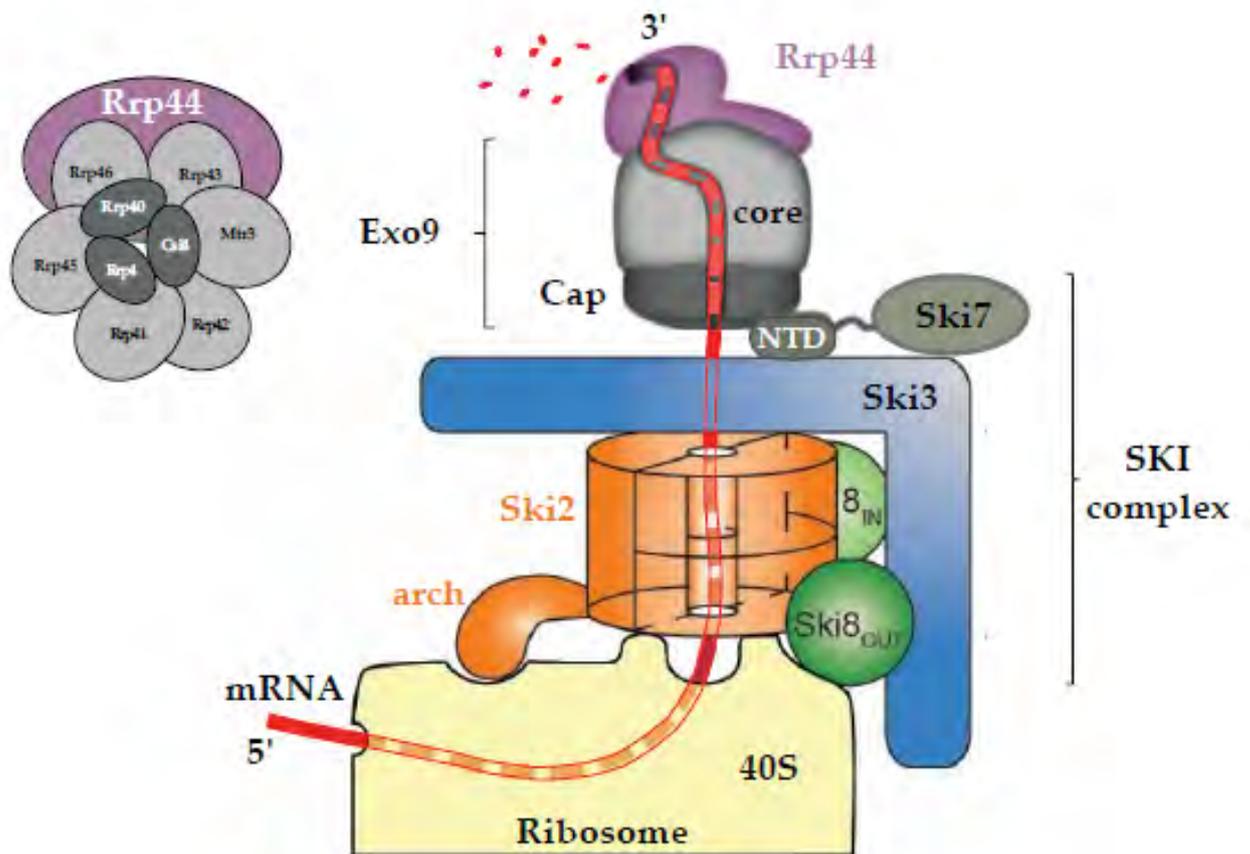
A**B**

Figure 9: Le complexe dégradosome chez *E. coli* et le complexe SKI-exosome chez la levure. A: Représentation schématique du dégradosome. La RNase E est représentée sous forme tétramérique en violet et ancrée à la membrane (gris) par son domaine MTS (orange). L'hélicase Rh1b, l'énolase et la PNPase sont représentées respectivement en vert, jaune et bleu. B: Model d'adressage de l'ARNm du ribosome par le complexe SKI vers l'exosome pour la dégradation. Représentation schématique du mécanisme de dégradation d'un ARNm bloqué sur le ribosome (jaune). Le recrutement du complexe SKI (Ski2 en orange, Ski3 en bleu et Ski8 en vert) permet l'adressage de l'ARNm vers le complexe exosome (gris). Le domaine N-terminal de la protéine Ski7 (gris foncé) permet l'interaction avec la coiffe de l'exosome. La dégradation de l'ARNm de l'extrémité 3' vers 5' est assurée par l'exoribonucléase Rrp44 (violet). La composition de l'exosome vu du dessus de la coiffe est aussi représentée. Schéma adapté de Halbach *et al.*, 2013 et Schmidt *et al.*, 2016.

pour atteindre la masse moléculaire d'environ 250 kDa. Cette protéine est connue pour être impliquée dans le déroulement des petits ARN nucléaires snARNU4/U6 composant du spliceosome pour activer la dissociation de U4 (Lingdi Zhang *et al.*, 2015). De manière sporadique, la protéine HelyY de *M. smegmatis* a été montrée comme étant une ATPase pouvant dérouler des duplexes ARN et ADN/ARN de 3' vers 5' de façon similaire à Suv3 décrite chez les eucaryotes (Shu *et al.*, 2004; Uson *et al.*, 2015).

B-3. MACHINERIES DE DEGRADATION DES ARN CHEZ LES BACTERIES ET LES EUCARYOTES

Chez les bactéries et les eucaryotes, les ribonucléases et les hélicases peuvent s'assembler sous forme de complexes multi-protéiques pour agir sur la dégradation et la maturation des ARN. Ces complexes peuvent soit dégrader les ARN par leurs extrémités 5', 3' ou soit induire des coupures endoribonucléolytiques.

Chez *E. coli*, un complexe multi protéique de très grande taille (plus de 2,5 MDa) appelé «dégradosome» se structure autour de la RNase E sous forme tétramérique (Mackie, 2013; Aït-Bara and Carpousis, 2015) (**Figure 9A**). La partie C-terminal de la RNase E, non structurée, contient des motifs d'interaction protéine/protéine et protéine/ARN qui permettent de s'associer à l'hélicase RhlB de la famille DEAD, l'exoribonucléase 3'-5' PNPase et l'énolase (Carpousis, 2007). Le complexe dégradosome de *E coli* permet ainsi la coordination de l'initiation de la dégradation par des coupures endoribonucléolytiques et des attaques exoribonucléolytiques 3' vers 5' des ARN (Mackie, 2013). Néanmoins, la composition du dégradosome diffère selon les conditions de croissance ou de stress (Prud'homme-Généreux *et al.*, 2004; Gao *et al.*, 2006). Chez *Bacillus Subtilis*, un dégradosome-like autour des RNase J1/J2 et de la RNase Y a été proposé qui inclurait une hélicase de type DEAD, CshA (Mathy *et al.*, 2007; Lehnik-Habrink *et al.*, 2010). La RNase J1 est également retrouvée associée aux hélicases CshA et RhpA (DEAD) respectivement chez *Staphylococcus aureus* et *Helicobacter pylori* (Redko *et al.*, 2013; Giraud *et al.*, 2015).

Dans les mitochondries des cellules eucaryotes, un dégradosome-like serait constitué de l'exoribonucléase DSS1 de type RNase II-like avec une activité de 3'-5' et de l'hélicase Suv3 de la famille Ski2-like, capable de dérouler des hétéroduplexes ADN/ARN (Malecki *et al.*, 2009). Ce complexe appelé mtEXO, serait impliqué dans la maturation des ARNr et la dégradation des ARNm (Khidr *et al.*, 2008). En effet, il a été montré que la délétion du gène codant pour Suv3 ou DSS1 entraîne une accumulation des pré-ARNr et des ARNm avec une

queue polyA mitochondriaux qui est associée à un défaut de synthèse des protéines dans ces organites.

Chez les eucaryotes, les hélicases Mtr4 et Ski2 font partie respectivement des complexes TRAMP (noyaux) et SKI (cytoplasme) et jouent des rôles essentiels dans la dégradation des ARN (Anderson and Parker, 1998; LaCava *et al.*, 2005). Ces deux protéines au sein de leurs complexes respectifs sont le plus souvent associées directement ou indirectement à des ribonucléases. Dans le cytoplasme, le complexe SKI (Ski2-Ski3-Ski8) est recruté au niveau du ribosome lors des erreurs de traduction (Schmitt *et al.*, 2016). Ski2 a ainsi été montrée comme interagissant directement avec la sous-unité 40S du ribosome (**Figure 9B**). Cette interaction permet d'une part l'activation de la protéine Ski2 et d'autre part, elle permet la translocation de l'ARNm vers le complexe exosome via la partie N-terminale de la protéine Ski7. Cette translocation de l'ARNm à travers le core de l'exosome vers l'endo/exoribonucléase Rrp44 permet sa dégradation.

Ainsi, l'exosome eucaryote est constitué d'un complexe core de 9 sous-unités inactives (EXO-9) formant un tonneau associé à une coiffe formée de trois protéines (Csl4, Rrp4 et Rrp40) possédant des domaines, S1 et/ou KH, de liaison à l'ARN (**Figure 9B**). Dans le cytoplasme, ce complexe s'associe à la ribonucléase Rrp44 possédant à la fois une activité endo et exoribonucléolytique pour former le complexe EXO-10 (Januszyk and Lima, 2014). Au sein du noyau, le core EXO9 de l'exosome est associé à la ribonucléase Rrp6 par l'intermédiaire de la coiffe (Wasmuth *et al.*, 2014). Cet exosome avec le complexe TRAMP permet la dégradation et la maturation des ARNr 5.8S (Zinder and Lima, 2017).

B-4. VOIES DE PROCESSING DES ARN CHEZ LES EUCARYOTES ET LES BACTERIES

Très brièvement sont décrites dans ce paragraphe les voies majeures de processing des ARN chez les eucaryotes et les bactéries qui nous servent de référence pour essayer d'aborder ces processus chez les archées, objet de nos études.

Chez les eucaryotes, après leur transcription, la majorité des ARN pré-messagers subit dans un premier temps une étape d'épissage des introns, première étape de contrôle-qualité. Ce mécanisme fait intervenir le recrutement de complexes ribonucléoprotéiques ou «spliceosome» composé de petits ARN nucléaires (U1, U2, U4, U5 et U6) associés à un panel de protéines. Après l'excision des introns, le spliceosome dépose des complexes protéiques appelés EJC (Exon junction complex) dont la fonction est le « marquage » de la maturation de

l'ARNm et constituent un des signaux d'export de l'ARNm vers le cytoplasme et qui est nécessaire pour le recrutement des ribosomes (Le Hir and Andersen, 2008). Plusieurs voies de contrôle des ARNm existent. Chez les eucaryotes supérieurs, la voie NMD pour Nonsense-Mediated Decay correspond à un mécanisme de contrôle de la qualité des ARNm qui consiste en la dégradation des ARNm présentant un codon stop précoce pouvant conduire à la traduction de protéine tronquée potentiellement toxique pour la cellule (Silva and Romão, 2009; Karousis *et al.*, 2016). En présence d'un codon stop aberrant en aval d'un complexe EJC, cette voie permet le recrutement du complexe de décoiffage des ARNm (DCP1/DCP2), de l'exoribonucléase 5'-3' XRN1 et l'endoribonucléase SMG6 pour dégrader de l'ARNm aberrant (Silva and Romão, 2009). La voie NGD pour No-Go Decay correspond au mécanisme de dégradation des ARNm bloqués en cours de traduction sur les ribosomes (Harigaya and Parker, 2010). Ce processus fait intervenir les protéines Dom34 et Hsb1 qui déclenchent la dissociation du ribosome des ARN, la coupure endoribonucléolytique de l'ARNm suivie de sa dégradation par Xrn1 en concertation avec le complexe de l'exosome. La voie NSD pour NonStop Decay est déclenchée lorsqu'au cours de la traduction, l'ARNm ne présente pas de codon stop (van Hoof *et al.*, 2002). Cette anomalie peut résulter de l'ajout prématuré d'une queue PolyA lors de la terminaison de la transcription. Cette absence de codon stop entraîne le blocage du ribosome sur l'ARNm et ne permet pas son recyclage pour réenclencher un nouveau cycle de traduction. Afin de palier à ce blocage, le ribosome recrute le complexe SKI et permet l'adressage de l'ARNm vers le complexe exosome afin d'être dégradé de l'extrémité 3'-5' (Schmidt *et al.*, 2016). Et enfin la voie AMD pour AU Rich Element (ARE) Mediated mRNA Decay consiste à la reconnaissance du ARE présent dans la région 3' non codante des ARNm par le complexe exosome afin d'être dégradé (Mukherjee *et al.*, 2002).

Chez les bactéries les transcrits primaires ARNm présentent une extrémité 5' triphosphate. Ces transcrits n'étant pas des substrats préférentiels des exoribonucléases 5'-3', il nécessite de manière générale une activité endoribonucléolytique pour permettre d'initier la dégradation des ARNm et de former des ARN présentant une extrémité 5' monophosphate beaucoup plus susceptible d'être dégradée par des exoribonucléases 5'-3'. Chez *B. subtilis* l'endoribonucléase Y (RNase Y) assure cette fonction de coupure et permet ensuite de déclencher la voie de dégradation 5'-3' par la RNase J1 (Khemici *et al.*, 2015) (**Figure 7**). Chez *E. coli*, l'endoribonucléase RNase E assure cette fonction de coupure initiatrice' parfois avec l'aide de petits ARN tels que les Hfq-binding-sRNA pour une dégradation ciblée de

certain ARNm (Morita and Aiba, 2011). Bien qu'aucune activité exoribonucléolytique 5'-3' n'a à ce jour été découvertes chez *E. coli*, une activité exoribonucléolytique 3'-5' est toutefois présente avec entre autres, la protéine PNPase au sein du dégradosome (Carpousis, 2007) (**Figure 9A**).

Outre ces coupures endoribonucléolytiques, les bactéries et les eucaryotes présentent des protéines impliquées respectivement dans l'attaque du 5' triphosphate et le décoiffage des ARNm. Ainsi, la protéine Rpph, une pyrophosphohydrolase de la famille des Nudix, est capable *in vitro* d'enlever les pyrophosphates de l'extrémité 5' triphosphorylé des ARN pour former des extrémités 5' monophosphorylé beaucoup plus sensible aux activités des exoribonucléases 5'-3' chez *B. subtilis* et permettent aussi d'accélérer la dégradation de nombreux transcrits chez *E. coli* (Deana *et al.*, 2008; Richards and Belasco, 2016). Chez les eucaryotes, qui présentent des ARNm coiffés (7mGpppN), deux voies majeures existent pour la dégradation des ARNm, 5'-3' et 3'-5'. L'initiation de la dégradation des ARNm par ces deux voies est déclenchée en premier lieu par la déanylation de la queue Poly(A) et par le recrutement du complexe de décoiffage Dcp1-Dcp2 au sein duquel Dcp2 est la sous-unité catalytiquement active (van Dijk *et al.*, 2002). L'extrémité 5' exposée est attaquée par la suite par l'exoribonucléase Xrn1 dans le cytoplasme et Xrn2/Rat1 dans le noyau (Nagarajan *et al.*, 2013) (**Figure 7**). La dégradation de l'extrémité 3' vers 5' est quant à elle assurée par le complexe exosome (Januszyk and Lima, 2014).

Les ARN ribosomiques sont synthétisés sous forme de long précurseurs dans toutes les cellules. Les pré-ARNr subissent de nombreuses étapes de maturation et de contrôle de la qualité consistant à des coupures endo et exoribonucléolytiques, des réarrangements et modifications de nucléotides avant être incorporés sous forme mature dans les deux sous-unités du ribosome. Ces étapes de maturation et de modification des ARNr sont très complexes et commencent à être élucidées chez les bactéries et les eucaryotes (Deutscher, 2009; Henras *et al.*, 2015).

Les ARN de transfert sont issus également de transcrits primaires nécessitant une maturation précise en 5' et en 3' pour être fonctionnels. Ces processus de maturation requièrent l'action de la ribonucléase universelle RNase P (Mann *et al.*, 2003; Kazantsev *et al.*, 2005) et de la RNase Z (Dubrovsky *et al.*, 2004; Li de la Sierra-Gallay *et al.*, 2006). De plus, chez les eucaryotes, l'endoribonucléase EndA permet l'excision des introns des pré-ARNt (Bujnicki and Rychlewski, 2000). L'ensemble des ARNt portent un triplé CCA à leur extrémité 3' qui

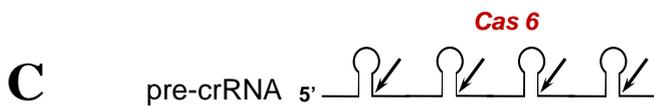
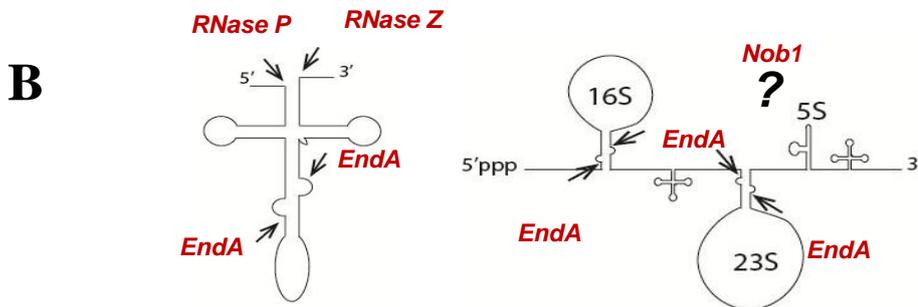
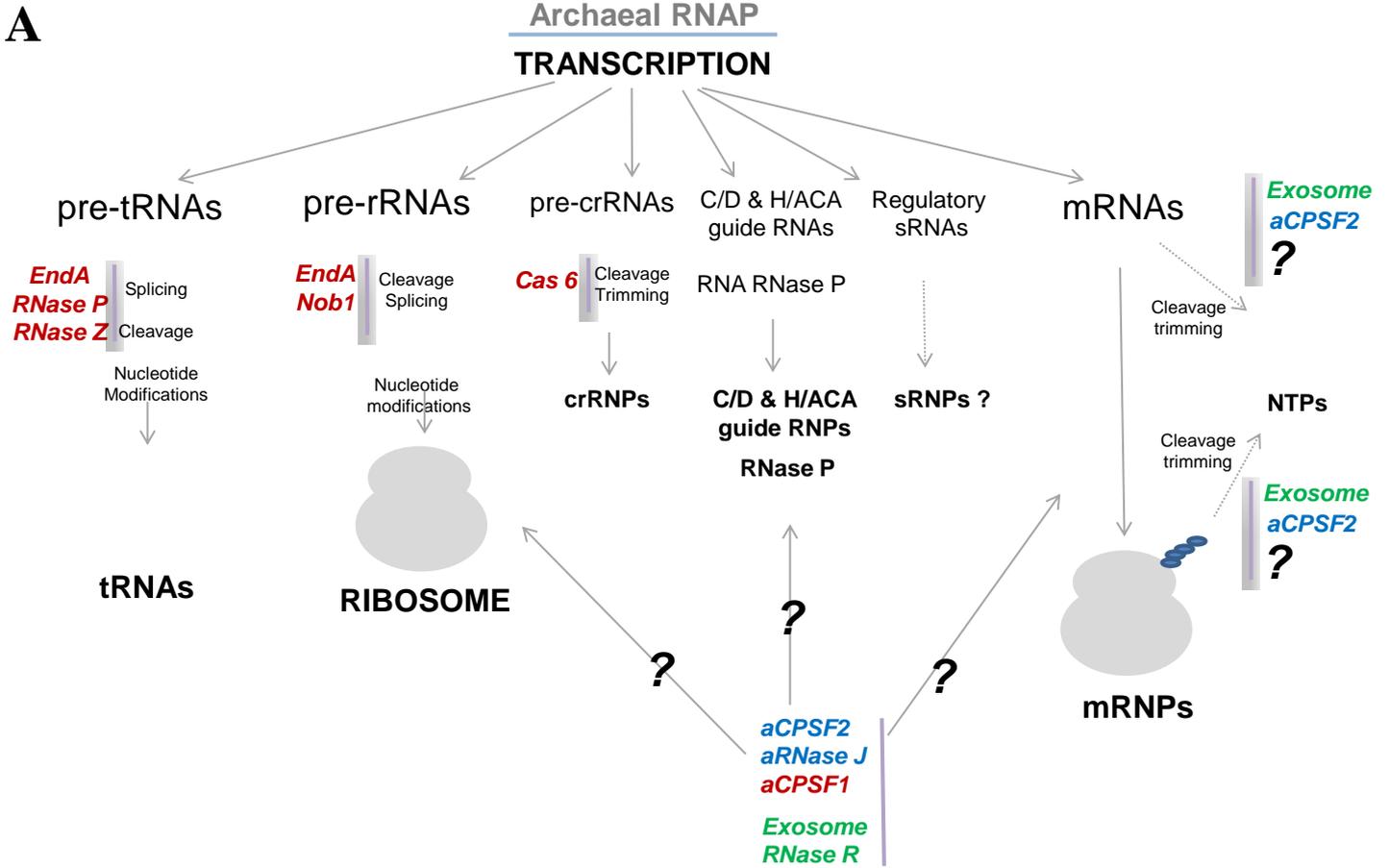


Figure 10: Métabolisme de l'ARN chez les archées. A: Figure issue de Phung *et al*, 2017 (en impression). Toutes les ARN connus et dont la transcription est assurée par l'unique RNA polymérase chez les archées et leur métabolisme sont représentés. Les mécanismes de modifications des ARN de transfert (tRNAs), des ARN ribosomiques (rRNAs), des ARN CRISPR (crRNAs), des ARN à boîtes C/D et H/ACA, des ARN guides, des ARN régulateurs et des ARN messagers (mRNAs) connus sont représentés en bleu et les enzymes correspondants en rouge. Les points d'interrogation correspondent aux ribonucléases non identifiées. Certaines ribonucléases n'ont pas encore de fonction attribuée (aCPSF1, aCPSF2, aRNase J et l'exosome). B: Représentation des sites de coupure de la RNase Z, RNase P, EndA et Nob1 sur les pré-ARN de transfert et les pré-ARN ribosomique. C: Site de coupure de la Ribonucléase Cas 6 sur les pré-ARN CRISPR.

est critique à la fois pour l'interaction avec le site A et P au sein de la grande sous-unité du ribosome (Samaha *et al.*, 1995).

Enfin, il existe de nombreux petits ARN non codants régulateurs (ncRNA ou sRNA) chez les bactéries et les eucaryotes impliqués dans la régulation de l'expression des gènes (Waters and Storz, 2009; Gottesman and Storz, 2011). Ainsi, le ncRNA 6S peut directement séquestrer la sous-unité σ^{70} de la RNA polymérase inhibant ainsi l'initiation de la transcription chez les bactéries (Wassarman, 2007). D'autre comme les «riboswiches» peuvent, par leur liaison au niveau du 5' UTR, empêcher l'initiation de la traduction (Montange and Batey, 2008). Chez les eucaryotes, les siRNA (small interfering RNA) et miRNA (microRNA) régulent l'expression des gènes par appariement spécifique avec un ARNm cible entraînant leur dégradation par l'endoribonucléase Argonaut, ce mécanisme est appelé «silencing» des ARN (Carthew and Sontheimer, 2009).

B-5. VOIES DE PROCESSING DES ARN DE TRANSFERT, ARN RIBOSOMIQUE ET ARN CRISPR CHEZ LES ARCHEES

Cette partie fait l'état de l'art qui est l'objet du chapitre 4 « An overview of ribonuclease repertoire and RNA processing pathways in Archaea» par D.K. Phung, M. Bouvier et B. Clouet-d'Orval du livre «RNA metabolism & gene expression in Archaea» de la série «Nucleic Acids and Molecular Biology» éditrice invitée B. Clouet-d'Orval, éditeur A.W. Nicholson au Edition Springer, sera publiée avant fin 2017.

La description de la voie de maturation universelle des ARN de transfert, les données très fragmentaires sur les voies de processing et d'assemblage du ribosome et le processing des ARN CRISPR chez les archées sont tout d'abord discutées (**Figure 10**). Enfin, l'ensemble des ribonucléases répertoriées chez les archées sont également présentées (**Figure 11**).

B-5-1. Processing des pré-ARN de transfert

Comme chez les bactéries et les eucaryotes, l'ensemble des ARNt chez les archées sont aussi issus de transcrits primaires et nécessite des étapes de maturation menant aux formes fonctionnelles (**Figure 10B**). En effet la majorité des ARNt possède en 5' une séquence leader et une séquence trailer en 3' qui nécessitent d'être excisées précisément afin d'obtenir un ARNt mature et fonctionnel. Ces processus de maturation requièrent la présence des ribonucléases universelles RNase P et la RNase Z présente dans la majorité des génomes

(Pannucci *et al.*, 1999; Schierling *et al.*, 2002). De plus à l'instar des eucaryotes, l'endonucléase spécifique EndA permet l'excision des introns des pré-ARNt (Kleman-Leyer *et al.*, 1997; Tang *et al.*, 2002). Chez certaines archées, le triplet CCA des ARNt est déjà codé dans leurs gènes respectifs et donc présent sur le précurseur (Marck and Grosjean, 2002). Pour d'autres ainsi que pour l'ensemble des eucaryotes, ce triplet est rajouté post-transcriptionnellement par des nucléotidyltransférases (Xiong and Steitz, 2004).

a. RNase P

La ribonucléoprotéine (RNP) RNase P est connue pour jouer un rôle essentiel dans la biogénèse des ARNt. En effet, cet enzyme est responsable de l'excision des séquences leaders en 5' des pré-ARNt (**Figure 10B**). Il est à noter que la RNase P est universellement retrouvée dans l'ensemble des organismes vivants à l'exception de *Nanoarchaeum equitans*. Dans ce cas l'absence de la RNase P est expliquée par l'absence de séquences leader en 5' des pré-ARNt (Randau *et al.*, 2008). La RNase P possède une activité endoribonucléolytique dépendante de la présence d'ion Mg^{2+} et hydrolyse les liaisons phosphodiester du précurseur ARNt et permet de générer une extrémité 5' monophosphate mature (Frank and Pace, 1998). La RNase P est un holoenzyme dont les différentes sous-unités protéiques (RPP) sont organisées autour d'une sous-unité catalytique ARN (RPR) d'environ 400nt en moyenne, ribozyme qui porte l'activité endoribonucléolytique (Hall and Brown, 2002). Le nombre de protéines qui composent la RNase P diffère selon les domaines du vivant, allant d'une seule protéine chez les bactéries, à au moins quatre chez les archées et jusqu'à 10 chez les eucaryotes (Evans *et al.*, 2006).

La sous-unité ARN, au sein des trois domaines, présente différentes structures classées parmi 4 groupes identifiés chez les bactéries et les archées. L'origine de l'activité du ribozyme n'est pas clairement définie puisque que l'ARN possède de grandes différences au niveau des régions connues pour être primordiales pour sa stabilité et son activité (Marquez *et al.*, 2005; Evans *et al.*, 2006; Gopalan, 2007). Récemment, il a même été découvert, une RNase P active uniquement protéique dans les mitochondries humaines ne possédant pas de sous-unité ARN (Howard *et al.*, 2013; Klemm *et al.*, 2016).

La sous-unité ARN de la RNase P bactérienne est active en absence des sous-unités protéiques contrairement aux RNase P eucaryote et archée, qui nécessitent les protéines RPP pour être actives (Pannucci *et al.*, 1999; Torres-Larios *et al.*, 2005). De nombreuses structures ont pu révéler l'importance de certaines parties de l'ARN de la RNase P impliquée dans

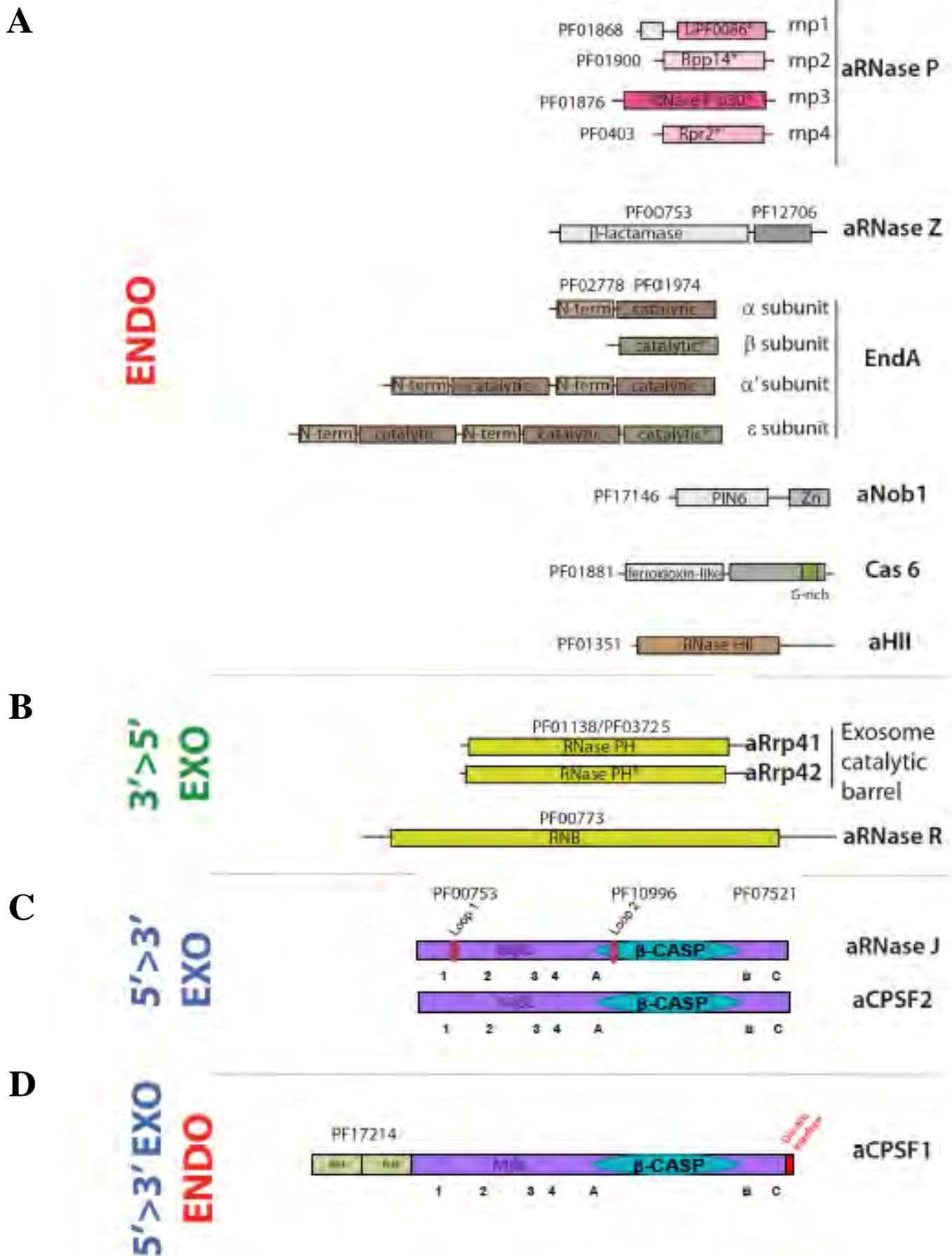


Figure 11: Architecture des ribonucléases chez les archées. Les Pfam des différents domaines sont indiqués. A : Les endoribonucléases RNase P, RNase Z, EndA, Nob1, Cas 6 et RNase HII sont indiquées. B: Représentation des exoribonucléases 3'-5' avec l'exosome et la RNase R. C : Représentation des exoribonucléases 5'-3' avec les β-CASP aRNase J et aCPSF2. D: Représentation de la β-CASP endo- exoribonucléases 5'-3' aCPSF1. Les Pfam des domaines metallo-β-Lactamase et le domaine β-CASP sont représenté respectivement en violet en bleu (<http://pfam.vfam.org>). Les motifs 1 à 4 et A à C impliqués dans la coordination de deux ions zinc sont représentés. aRNase J possèdent de plus deux Loop (1 et 2) spécifique des archées. aCPSF1 possèdent une extension N-terminale, en vert, constituée de deux KH-domain impliqués dans la liaison à l'ARN et un domaine de dimérisation en C-terminal en rouge.

l'activité catalytique et dans la reconnaissance des substrats (Gopalan, 2007). Les sous-unités RPR sont composées de deux domaines d'ARN fonctionnels, le domaine de liaison au substrat (S-domain) et le domaine catalytique (C-domain), qui constitue la sous-unité ancestral de type A retrouvée chez les bactéries (Jarrous and Gopalan, 2010) (**Figure 11A**). Toutefois, la sous-unité RPR de type M est présente chez *Methanococci* et *Archaeoglobulus* (Pannucci *et al.*, 1999). Les propriétés biochimiques des RPR chez les archées sont similaires à celles observées pour les RPR issus des bactéries. Cela suggère que l'ARN archéen contient tous les éléments requis pour la reconnaissance des substrats et l'activité catalytique mais, il est structuralement non fonctionnel en absence de protéine. Il n'y a à ce jour pas de structure atomique résolue de la RNase P dans sa totalité chez les archées, néanmoins des structures des composants protéiques RPP ont été obtenues (Numata *et al.*, 2004; Honda *et al.*, 2008).

b. RNase Z

La protéine universellement conservée RNase Z caractérisée dans l'ensemble des trois domaines du vivant mais aussi présente dans les mitochondries et les chloroplastes est l'enzyme responsable de la maturation de l'extrémité 3' de la majorité des ARNt et de la maturation des ARNr (**Figure 10B**). La RNase Z fait partie de la famille des Elac faisant elle-même partie de la superfamille des metallo- β -lactamase qui possèdent des fonctions très diverses (Aravind, 1999) (**Figure 11A**). Les analyses approfondies des membres de la superfamille metallo- β -lactamase ont permis l'identification de deux sous-familles d'hydrolase d'acides nucléiques, la famille des RNase Z et les β -CASP (Aravind, 1999; Daiyasu *et al.*, 2001) (**Figure 11A**).

La maturation de l'extrémité 3' des ARNt est de loin plus complexe que celle en 5'. La maturation de l'extrémité 3' (aussi appelée 3' trailers) des ARNt nécessite les activités de clivage endoribonucléolytique et de grignotage exoribonucléolytique dépendant de la présence de la séquence CCA des pré-ARNt. Brièvement, les RNase Z contiennent un domaine metallo- β -lactamase suivi d'un domaine de fixation spécifique de substrats ARNt appelé exosite. Ces enzymes existent sous deux formes : une forme courte (280-360 acides aminés) fonctionnant sous forme dimérique et présente dans les trois domaines du vivant. La forme longue (750-930 acides aminés) est présente seulement chez les eucaryotes et serait issue de la duplication de la forme courte (Vogel *et al.*, 2005). La RNase Z possède une activité endoribonucléolytique permettant de générer une extrémité 3' mature des molécules d'ARNt par l'excision de l'élément 3' trailer du pré-ARNt. Ce clivage permet la formation

d'un groupement 3' hydroxyl sur l'extrémité 3' d'ARNt et un 5' phosphate sur l'élément trailer et nécessite la présence de deux ions Zn^{2+} pour catalyser la réaction (Dominski, 2007; Redko *et al.*, 2007).

La protéine recombinante RNase Z a été montrée capable de cliver efficacement et précisément des précurseurs ARNt *in vitro*. La maturation de l'extrémité 3' des ARNt par la RNase Z a été caractérisée chez de nombreuses espèces d'archées, que ce soit chez *Haloferax volcanii*, *Methanococcus jannaschii* ou encore *Pyrococcus furiosus* (Schierling *et al.*, 2002; Schiffer *et al.*, 2002; Späth *et al.*, 2008). Cette ribonucléase nécessite la structure totale des pré-ARNt pour une activité optimale (Schierling *et al.*, 2002). Toutefois, la RNase Z peut agir sur des substrats possédant une structure de type ARNt présente en amont des ARNr 5S chez certains halophiles et thermococcales (Hölzle *et al.*, 2008). Pour cette raison, la RNase Z pourrait être aussi un acteur majeur de la voie de maturation des ARNr chez les archées (Kleman-Leyer *et al.*, 1997). Au vu de son rôle dans la maturation des ARNt, le gène codant pour cette ribonucléase est montré comme étant essentiel chez *H. volcanii* (Hölzle *et al.*, 2008).

c. EndA

L'endoribonucléase (EndA) archée catalyse le clivage au niveau du 5' et 3' des sites d'épissage des pré-ARNt et permet l'excision de l'intron (**Figure 10B**). Chez les eucaryotes, EndA est constituée de 4 sous-unités: Sen2p, Sen15p, Sen34p et Sen54p (Lopes *et al.*, 2015). Les gènes codants pour les sous-unités α , α' , β et ϵ de EndA d'archées sont homologues à ceux retrouvés chez les eucaryotes et sont présents dans tous les génomes d'archées. Toutefois la composition de EndA diffère selon les archées et est classée en 4 groupes (**Figure 11A**). En effet, chez les euryarchées, EndA peut être présente sous forme homotétramérique de type α_4 (*M. jannaschii*) et homodimérique de type α'_2 (*H. volcanii*). Chez les crenarchées et les nanoarchées, EndA est retrouvée sous forme hétérotétramérique de type $\alpha_2\beta_2$ (Lopes *et al.*, 2015). Plus récemment, EndA peut aussi être retrouvée sous forme d'homodimère de type ϵ_2 chez *C. Micrarchaeum*. Parmi toutes ces configurations, le repliement de EndA permet le clivage efficace du motif d'ARN en bulge-helix-bulge (BHB) constitué de deux boucles de trois nucléotides séparées par une hélice de quatre nucléotides. Les introns des ARNt retrouvés à différentes positions chez les archées mais sont toujours reconnus pour être excisés par EndA (Marck and Grosjean, 2003).

d. Enzymes de modification des ARNt

De plus, les ARN de transfert des archées, comme ceux des bactéries et des eucaryotes, subissent une panoplie de modifications sur les riboses et les bases afin de devenir fonctionnels (Machnicka *et al.*, 2015). Nous ne citerons ici que deux exemples. La modification N(6)-threonylcarbamoyladenosine (t(6)A) est présente dans tous les ARNt et permet de décoder des codons commençant par une adénosine. Cette modification fait intervenir le complexe KEOPS (Kinase, endopeptidase and proteins of small size) (Perrochia *et al.*, 2013). Les méthylations en 2'hydroxyl du ribose se retrouvent à de nombreuses positions sur les ARNt, le plus souvent à la position wobble 34 de l'anticodon. Ces modifications font appel à des petits ARN, appelés ARN guide à boîtes C et D qui permettent de positionner la méthyl-transférase, fibrillarine du complexes RNP (Clouet d'Orval *et al.*, 2001; Bortolin *et al.*, 2003; Dennis and Omer, 2005). Ces RNP interviennent aussi sur les modifications des ARNr.

B-5-2. Processing des ARN pré-ribosomiques

Comme chez les eucaryotes et les bactéries, les ARN ribosomiques sont synthétisés à partir de précurseurs. Le nombre d'opérons d'ARNr par génome varie de 1 à 4 selon les espèces. L'opéron de l'ARNr contient l'ARN 16S et 23S chez les crénarchées, l'ARN 5S et l'ARNt Ala et Cys sont présents en plus chez les euryarchées (Klug *et al.*, 2007). Peu de chose sont connues sur les ribonucléases impliquées dans ces voies de maturation des ARNr. Hormis le rôle de EndA dans l'excision des pré-ARNr 16S et 23S des transcrits primaires par clivage des séquences espaceurs contenant des motifs BHB (Kjems and Garrett, 1988; Lykke-Andersen and Garrett, 1997) (**Figure 10**). Comme chez les eucaryotes, les ARNr archées subissent aussi de nombreuses modifications. Ainsi, les petits ARN à boîtes C/D et H/ACA des archées permettent respectivement la 2'O-méthylation du ribose et la pseudouridylation sur des nucléotides spécifiques des ARNr et jouent un rôle central dans les modifications des ARN (Rozhdestvensky *et al.*, 2003; Dennis and Omer, 2005). Il a de plus été montré que les gènes codant pour les petits ARN à boîte C/D ne possèdent pas de promoteur indépendant mais se retrouvent souvent à des positions leur permettant d'être co-transcrits avec les gènes adjacents. Les petits ARN à boîte C/D et H/ACA sont associés à des protéines spécifiques pour former des complexes ribonucléoprotéiques (RNP), aNop5, aL7ae et aFib pour les ARN à boîte C/D et aNop10, aCbf5, aGar1 et aL7ae pour les ARN à boîte H/ACA. La protéine aL7ae dans les deux cas permet la stabilisation des K-turn qui est essentiel à leur fonction de

modification d'ARN (Klein *et al.*, 2001). Les protéines aFib et Cbf5 jouent les rôles de méthyltransferase et de pseudouridylation respectivement dans les deux complexes (Tran *et al.*, 2003; Charpentier *et al.*, 2005). Les autres protéines ont pour rôle la stabilisation des complexes RNP.

Les processus d'assemblage du ribosome chez les archées restent très peu connus. L'identification de l'homologue de l'endoribonucléase Nob1 eucaryote a permis de suggérer que la maturation de l'extrémité 3', sur le site D des pré-ARNr de la petite sous-unité du ribosome soit semblable entre les deux domaines (**Figure 6**). La protéine Nob1 est ainsi retrouvée chez l'ensemble des groupes des archées à l'exception des Korarchaea (Veith *et al.*, 2012). Des tests *in vitro* de Nob1 de *Pyrococcus horikoshii* ont pu montrer une activité endoribonucléolytique efficace de la protéine sur des substrats ARN présentant un site D du pré-ARN 16S en présence de manganèse (Veith *et al.*, 2012). La structure résolue de la protéine Nob1 par spectroscopie RMN a permis de révéler que le domaine PIN (PiIT N-terminus) est commun avec de nombreuses nucléases et la présence d'un domaine zinc ribbon. De plus, la structure et la fonction de Nob1 semblent être conservées parmi les archées et les eucaryotes en dépit de l'absence d'une large insertion dans le domaine PIN et d'une queue C-terminal typiquement retrouvé chez les eucaryotes (Veith *et al.*, 2012).

B-5-3. Processing des ARN CRISPR

Le système de défense basé sur les ARN CRISPR et les protéines associées aux CRISPR (Cas) permettent à ces complexes de détruire les ARN ou ADN invasifs. Les ARNcr issus de long précurseurs sont maturés par des endoribonucléases pour donner des séquences aux alentours de 70-80nt. Les ARNcr matures sont alors intégrés au sein de complexes ribonucléoprotéiques avec leurs protéines Cas et guident ces complexes vers les séquences d'acides nucléiques exogènes (Wiedenheft *et al.*, 2012; van der Oost *et al.*, 2014). Comme chez les bactéries où la RNase III et la protéine Cas6 sont responsables de la production des ARNcr matures, les membres de la famille des Cas6 chez les archées joueraient aussi ce rôle de maturation des ARNcr (**Figure 10C et Figure 11A**). Cas6 de *P. furiosus* fut la première protéine à être caractérisée et montrant une activité endoribonucléolytique sur les séquences répétées (Carte *et al.*, 2008). De plus, la structure résolue de la protéine a permis de révéler deux repliements en de type ferredoxin-like retrouvés dans les protéines interagissant avec les ARN. Cette protéine possède ainsi la capacité d'interagir avec une séquence spécifique présente dans l'extrémité 5' des régions répétées et est capable de cliver sur un site défini au

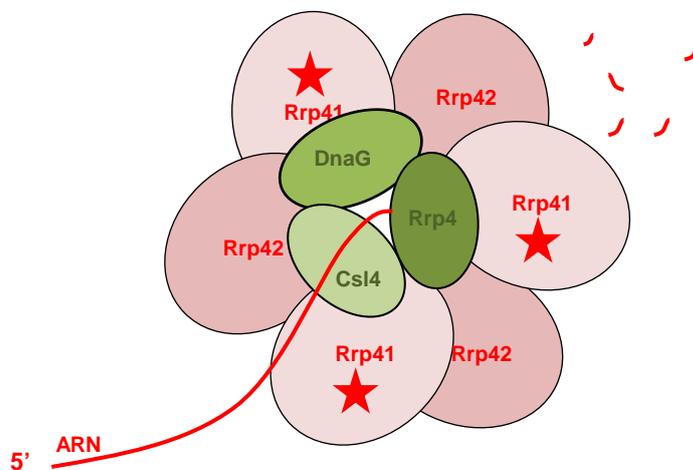


Figure 12: Exosome archée. Représentation schématique du complexe exosome chez les archées. Les protéines Rrp41 et Rrp42 du core catalytique sont représentées respectivement en rose et en rouge. L'étoile rouge indique la sous-unité portant l'activité catalytique. Les protéines de la coiffe, Rrp4, DnaG et Csl4, sont représentées en vert. En trait rouge est représenté un brin d'ARN simple brin.

niveau de l'extrémité 3' (Carte *et al.*, 2008). Le gène codant pour la protéine Cas6 est l'un des gènes codant pour les protéines associées au CRISPR les plus présentes parmi les bactéries et les archées (Hochstrasser and Doudna, 2015). Les séquences en acides aminés des Cas6 chez les archées sont peu conservées et présentent seulement deux motifs conservés : le repliement ferredoxin-like et le motif riche en glycine (Li, 2015). Il est encore difficile de comprendre comment l'endoribonucléase Cas6 présente dans des organismes avec de multiples systèmes CRISPR discrimine les différentes cibles (Hochstrasser and Doudna, 2015).

B-6. AUTRES ENZYMES ET MACHINERIES DE DEGRADATION DES ARN CHEZ LES ARCHEES

B-6-1. aExosome

La prédiction de l'existence d'un exosome chez les archées proche de ce qui existe chez les eucaryotes a été initialement basée sur l'identification de gènes orthologues codant pour les différentes sous-unités de l'exosome eucaryote chez les archées, Rrp4, Rrp41 et Rrp42 (Ribosomal RNA Processing protein). Ces différents composants sont localisés au sein d'un superopéron très conservé et identifié dans la plupart des génomes d'archées à l'exception des Halophiles et les Methanococcales (Koonin *et al.*, 2001). La première évidence expérimental de l'existence d'un exosome archéen a été obtenue par des expériences de co-immunoprécipitation avec des anticorps dirigés contre la protéine Rrp41 de *Sulfolobus solfataricus* et a mené à la purification d'un complexe de 250 kDa. Les protéines orthologues de Rrp4, Rrp41, Rrp42 et Csl4, ainsi que la protéine annotée comme l'homologue de la DnaG sont retrouvées fortement associées dans ce complexe (Evguenieva-Hackenberg *et al.*, 2003). Le core minimal de l'exosome de *S. solfataricus* et de *M. thermautotrophicus* est constitué de six homologues de la RNase PH ainsi que les sous-unités Rrp4 et Csl4 (Evguenieva-Hackenberg *et al.*, 2003; Farhoud *et al.*, 2005). Des études structurales des sous-unités core des exosomes de *S. solfataricus* et de *A. fulgidus* ont révélé un arrangement en un anneau hexamérique de trois hétérodimères de Rrp41-Rrp42 (**Figure 12**) (**Figure 11B**). Il a été montré par des analyses de mutagénèses ciblées que l'activité de l'exosome est portée seulement par la sous-unité Rrp41 avec une activité de type RNase PH commun aux exoribonucléases phosphorolytiques. La sous-unité Rrp42 est inactive mais contribue à la structuration du tonneau pour présenter les sites catalytiques de Rrp41. Le core de l'exosome est associé à une coiffe de trois sous-unités de constitution variable avec les sous-unités Rrp4 et Csl4 possédant chacune un domaine S1 de liaison à l'ARN et la protéine DnaG (Lorentzen *et al.*, 2005; Büttner *et al.*, 2005). Au sein de la structure en neuf sous-unités de l'exosome,

l'ARN se lie à la fois au site catalytique et à l'opposé au niveau le plus étroit du canal central. Le substrat ARN est ainsi dirigé vers la chambre catalytique à travers le pore central qui a la particularité de ne pouvoir laisser passer qu'un simple brin d'ARN. L'exosome d'archée révèle une similarité de séquence avec le core exosome hexamérique eucaryote mais aucune des sous-unités de type RNase PH formant le tonneau ne sont actives (Liu *et al.*, 2006; Dziembowski *et al.*, 2007). De plus, l'agencement des différents sous-unités forme une structure très similaire à celle formée par la PNPase bactérienne et supporte un même mode de dégradation 3'-5' des ARN dans les trois domaines du vivant (Liu *et al.*, 2006; Nurmohamed *et al.*, 2009; Chlebowski *et al.*, 2013).

De manière spécifique, il a été expérimentalement montré chez *Sulfolobus solfataricus*, *Pyrococcus abyssi* et *Methanopyrus kandleri* que l'exosome a aussi la capacité de synthétiser une courte queue d'hétéropolymère à l'extrémité 3' des ARN par son activité de polyadénylation (Ramos *et al.*, 2006; Portnoy and Schuster, 2006). Les deux réactions de phosphorolyse et de polyadénylation nécessitent *in vitro* du phosphate inorganique, du rNDP et des ions Mg^{2+} similaire à l'activité de la PNPase bactérienne (Evguenieva-Hackenberg *et al.*, 2008). Les ions divalent facilitent l'attaque nucléophile du phosphate inorganique (Pi) sur le 3' phosphate de la liaison phosphodiester. Cette réaction de phosphorolyse conduit au relargage de rNDP. De la même manière, lors des réactions de polyadénylation, l'incorporation de rNDP est couplée avec le relargage de Pi. De façon cohérente, l'activité de polyadénylation des ARN n'est observée seulement chez les espèces d'archées possédant un exosome et n'est pas présente chez les Halophiles et certains Méthanogènes qui ne possèdent pas d'homologues Rrp41-Rrp42 (Portnoy and Schuster, 2006).

B-6-2. RNase R

Chez les archées, seuls les Halophiles et certains Méthanogènes possèdent dans leur génome le gène *rnr* codant pour la RNase R (Portnoy and Schuster, 2006). Chez les bactéries, la RNase R a été la première ribonucléase à être décrite comme possédant une activité exoribonucléase hydrolytique 3'-5' avec une capacité inhabituelle à digérer des substrats ARN très structurés sans l'aide d'hélicase (Vincent and Deutscher, 2009). La RNase R bactérienne est constituée de deux domaines CSD (Cold Shock Domain 1 & 2) dans sa partie N-terminale, un core catalytique central RNB avec des motifs impliqués dans la coordination d'ion Mg^{2+} et un domaine S1 d'interaction à l'ARN en C-terminale. Récemment, il a été démontré que la RNase R bactérienne possède sa propre activité hélicase lui permettant de

faciliter la dégradation des substrats ARN en duplexe hautement stables et des duplexes présentant une extrémité 3' sortante (Hossain *et al.*, 2016). Cette activité additionnelle est conférée par les domaines N- et C-terminal qui ont la capacité de former un site de liaison à l'ARN nécessaire au bon fonctionnement de la RNase R *in vivo* et assure aussi l'efficacité du métabolisme de l'ARN chez les bactéries (Hossain *et al.*, 2016). Il a aussi été démontré que le domaine S1, riche en lysine est impliqué dans le recrutement de la ribonucléase par les ribosomes bloqués lors de la traduction dans le but de dégrader les ARNm ne présentant pas de codon stop (Ge *et al.*, 2010).

La aRNase R (RNase R d'archée) a été initialement identifiée par recherche de similarité de séquences dans les génomes d'archées. Chez *H.volcanii*, un gène codant pour une RNase R-like existe et présente 34% d'identité de séquences avec la RNase R d'*E. coli* (Portnoy and Schuster, 2006). Chez les Halophiles et certains Méthanogène, la aRNase R semble être l'unique exoribonucléase qui présente une activité 3' vers 5' mais ne peut dégrader des substrats structurés contrairement à la RNase R bactérienne. En effet, la aRNase R est restreinte au core catalytique central RNB et ne possède pas les extensions en N- (CSD1 et 2) et en C-terminal (S1) spécifiques de la RNase R bactérienne (Matos *et al.*, 2012) (**Figure 11B**).

Des études enzymatiques *in vitro* de la protéine recombinante RNase R de *H. volcanii* sur des substrats ARNt ont pu montrer que l'activité exoribonucléolytique est optimale entre 25°C et 37°C et à une faible concentration en sel (Matos *et al.*, 2012). Ces observations suggèrent que cette enzyme, même provenant d'une espèce halophile ne semble pas être adaptée à un taux élevé en salinité. Toutefois, l'activité résiduelle à haute concentration en sel est suffisante et même essentielle à la survie de *H. volcanii* (Portnoy and Schuster, 2006; Matos *et al.*, 2012). De plus, l'activité de la aRNase R est comparable à la RNase II bactérienne dès lors qu'elles libèrent des produits finaux de 4 à 5nt tandis que la RNase R bactérienne peut digérer complètement les substrats RNA. Ainsi, l'absence de polyadénylation chez les halophiles (absence de l'exosome) n'est pas compensée par une forte activité de la RNase R (Portnoy and Schuster, 2006).

B-6-3. Ribonucléases β -CASP

De manière générale, les ribonucléases de la famille des β -CASP sont retrouvées chez les bactéries et les eucaryotes et sont impliquées à la fois dans le métabolisme des ARN avec pour exemple la RNase J bactérienne et le facteur CPSF73 eucaryote et aussi dans le

métabolisme de l'ADN avec les protéines Artemis, SNM1 et PSO2 (Aravind, 1999; Daiyasu *et al.*, 2001; Callebaut *et al.*, 2002; Dominski *et al.*, 2013). Le nom de cette famille est ainsi issu de l'identification de ces premiers membres (CPSF73, Artemisn SNM1, PSO2.). Les protéines de cette famille sont constituées d'un domaine metallo- β -lactamase (composé de quatre couches $\alpha\beta/\beta\alpha$, commun avec la RNase Z) et d'une insertion d'un domaine β -CASP (six feuillets β centraux entourés de cinq hélices α sur chaque face) caractéristique (**Figure 11C et D**). Ces deux domaines présentent des résidus acides aspartiques, histidines et valines très conservés et essentiels pour former des motifs capables de coordonner deux ions zinc afin de former le site catalytique de l'enzyme (**Figure 11C et D**). L'unique site catalytique de ces protéines peut porter à la fois des activités hydrolytiques endo- et exoribonucléolytique pour certains membres de la famille (Newman *et al.*, 2011). Bien que présentant des extensions N- ou C-terminales, le core catalytique metallo- β -lactamase- β -CASP présente des structures très conservées (Li de la Sierra-Gallay *et al.*, 2008; Mandel *et al.*, 2008).

Au sein des β -CASP agissant sur la dégradation des ARN, deux groupes ont récemment émergés. Un groupe proche du facteur de terminaison de la transcription CPSF73, appartenant au complexe multi-protéique de coupure et de polyadénylation responsable de la coupure endoribonucléolytique définissant les extrémités 3' des ARNm décrit dans la partie I-2-2 (**Figure 4B**). CPSF73 est aussi connue pour être impliquée dans la maturation des pré-ARNm codant des histones (Dominski *et al.*, 2005). Le deuxième groupe des β -CASP est proche de la RNase J bactérienne et est crucial pour la dégradation des ARNm et pour la maturation des ARN ribosomiques 16S (Britton *et al.*, 2007; Mathy *et al.*, 2007). Les deux groupes sont retrouvés chez les trois domaines du vivant. Ainsi, en plus de CPSF73, les eucaryotes possèdent au sein du complexe CPSF, la protéine CPSF100 impliquée dans la stabilisation de ce complexe (Mandel *et al.*, 2008). Chez les eucaryotes supérieurs, la protéine Int, proche de CPSF73, est impliquée dans le processing des ARN médié par les petits ARN au sein de l'«integrator complex» (Baillat and Wagner, 2015). Chez les bactéries, des homologues de CPSF73 sont aussi présents mais restent très sporadiques (Condon and Gilet, 2011). Chez les bactéries et plus particulièrement chez les Firmicutes, RNase J est retrouvée sous deux formes, RNase J1 active et RNase J2 non active (Even *et al.*, 2005). Des homologues de la RNase J sont présents dans les chloroplastes des plantes (Sharwood *et al.*, 2011).

Des gènes codants pour des enzymes de la famille des β -CASP sont également retrouvés dans l'ensemble des génomes d'archées (Nishino *et al.*, 2005; Clouet-d'Orval *et al.*, 2010; Mir-Montazeri *et al.*, 2011; Silva *et al.*, 2011). Ainsi chez les archées, CPSF73 présente trois

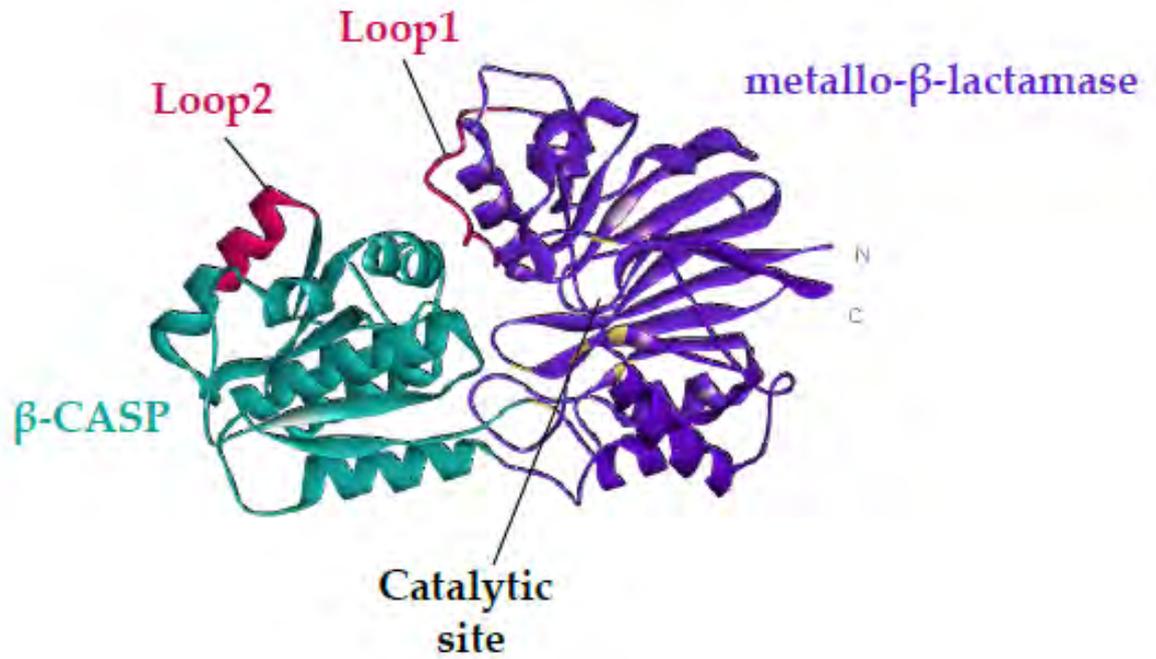
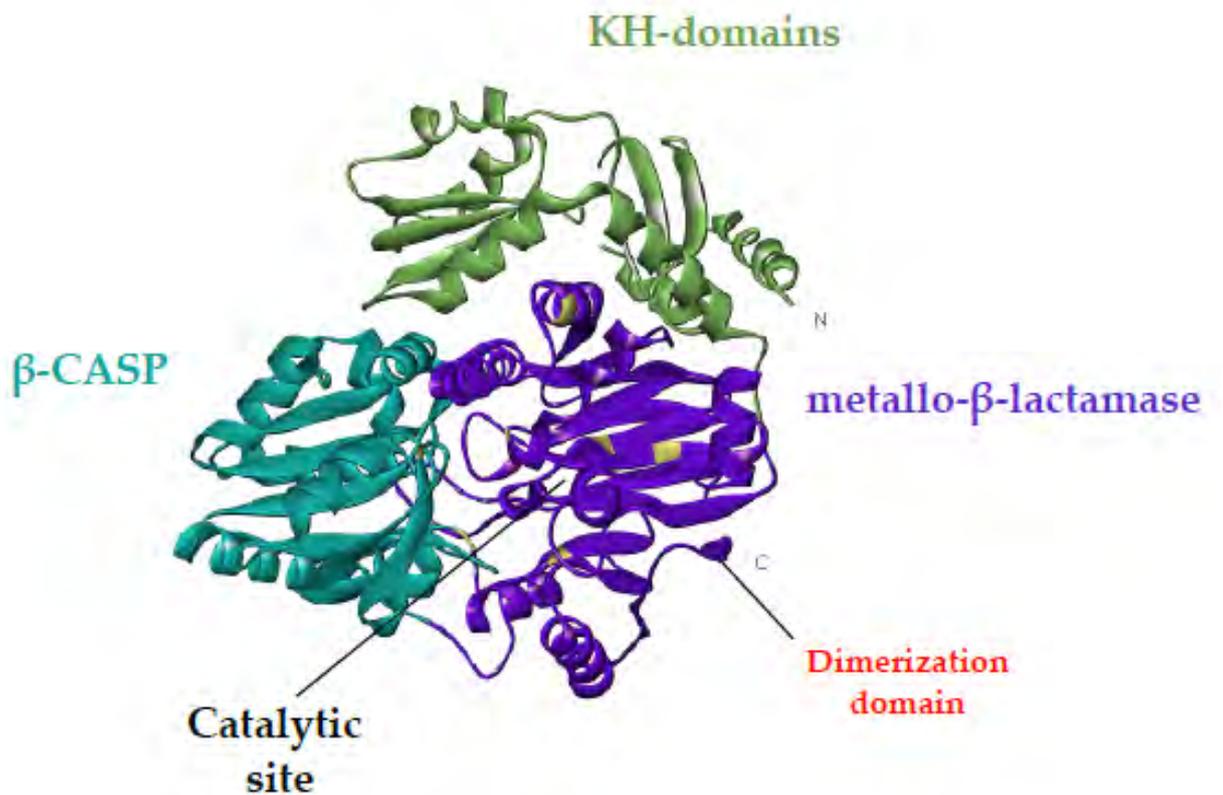
A**B**

Figure 13: Structures résolues de aRNase J et aCPSF1. Les domaines metallo-β-Lactamase et le domaine β-CASP sont représenté respectivement en violet en bleu (<http://pfam.vfam.org>). A: Structure de aRNase J de *Methanobolus psychrophilus* (PDB : 5HAA). Les deux Loop (1 et 2) spécifique des archées sont indiquées. B: Structure aCPSF1 de *Pyrococcus horikoshii* (PDB: 3AF5). Les extensions KH-domain en N-terminale sont représentées en vert. Le domaine de dimérisation en C-terminale est indiqué.

groupes orthologues, aCPSF1, aCPSF2 et aCPSF1b. aCPSF1 est présente dans l'ensemble des archées sans exception. aCPSF2 est retrouvée dans la majorité des archées et aCPSF1b est présente seulement chez les Methanococcales (Phung *et al.*, 2013). Des orthologues de la RNase J bactérienne, nommés aRNase J, sont aussi présents chez la majorité des Euryarchées à l'exception des Thermoplasmatales et des Archeaoglobales et ne sont pas présents chez les Crenarchées (Phung *et al.*, 2013). La congruence de l'arbre phylogénétique des gènes codant pour aRNase J avec celui des Euryarchées reflète un transfert vertical de ce gène au sein des archées et suggère que la protéine aRNase J soit présente au sein de l'ancêtre commun des archées et des bactéries avant leur séparation (Clouet-d'Orval *et al.*, 2010). Les activités de la protéine aRNase J chez *P. abyssi*, *Thermococcus kodakaraensis* and *Methanocaldococcus jannashii* ont été caractérisées (Clouet-d'Orval *et al.*, 2010; Levy *et al.*, 2011). aRNase J a été démontrée comme pouvant dégrader les ARN de 5'-3' de manière très processive. Dans les conditions des tests, aucune activité endoribonucléolytique n'a été observée contrairement à la RNase J bactérienne. Il a aussi été montré que les ARN présentant un 5' monophosphate sont des substrats préférentiels de aRNase J tandis qu'une extrémité 5' triphosphate permet de les protéger contre l'activité de aRNase J (Clouet-d'Orval *et al.*, 2010; Levy *et al.*, 2011). Contrairement à la RNase J bactérienne qui présente une extension C-terminale, la aRNase J archée ne présente pas d'extension N- ou C-terminale et est restreinte au core metallo- β -lactamase et β -CASP. aRNase J présente de manière spécifique et conservée deux Loop (Loop1 et Loop2) jouant un rôle important dans la modulation de l'activité de la protéine (Clouet-d'Orval *et al.*, 2010).

Les structures des protéines aCPSF1 et aRNase J ont été résolues chez différentes archées. Ainsi, la structure obtenue pour aCPSF1 (PDB : 3AF5, 2YCB) est parfaitement superposable avec celle de CPSF-73 eucaryotes (Mir-Montazeri *et al.*, 2011; Silva *et al.*, 2011) (**Figure 13**). Cette conservation est d'autant plus étonnante que les archées dont sont issues les protéines sont pour l'une mésophile et pour l'autre thermophile. Cette conservation stricte des structures peut ainsi suggérer un rôle crucial de ces protéines dans les différentes voies métaboliques quel que soit leur niche écologique.

De manière similaire à aCPSF1, la structure de la protéine aRNase J de *Methanobrevibacter smithii* (PDB : 5HAA), récemment résolue, est aussi superposable à celle de la RNase J bactérienne (**Figure 13**). Du fait de la similitude des structures de la RNase J bactérienne et de la aRNase J archée et leur préférence pour des substrats 5' monophosphate, il a été proposé que aRNase J ait un rôle similaire dans la dégradation des ARNm chez les archées. Les

transcrits primaires chez les archées, comme chez les bactéries, présentent un 5' triphosphate. Les ARN subissent donc une étape nécessaire de modification par des enzymes comparables aux hydrolases Nudix chez les bactéries et les enzymes de décoiffage chez les eucaryotes (Deana *et al.*, 2008; Song *et al.*, 2013) ou une coupure endoribonucléolytique avant de pouvoir entrer dans la voie 5'-3' exoribonucléolytique similaire à celle de la RNase J1 bactérienne. Des enzymes Nudix ont été identifiées chez les archées mais ne présentent à ce jour pas d'activité pyrophosphohydrolase *in vitro* (Hasenöhrl *et al.*, 2011). L'activité endoribonucléolytique, nécessaire pour cette entrée dans la voie de dégradation 5'-3', pourrait être portée par la protéine aCPSF1. En effet, aCPSF1 a présente une activité endoribonucléolytique sur des ARN présentant les dinucléotides CA (Phung *et al.*, 2013). L'ensemble de la caractérisation, biochimique et phylogénomique de la ribonucléase aCPSF1 sera présentée dans la partie I des résultats.

C. PROJET DE THESE

Les rôles physiologiques des protéines β -CASP, aCPSF1 et aRNase J dans le métabolisme de l'ARN ne sont pas encore définis (**Figure 6**). Or, il a été suggéré que aCPSF1, dont le gène est retrouvé en aval du gène codant pour une sous-unité du protéasome, pourrait avoir un lien avec le complexe exosome (Koonin *et al.*, 2001). En effet, la sythénie des gènes codant pour l'appareil de transcription (RNA pol), de traduction (ribosome), de dégradation des protéines (protéasome) et des ARN (exosome) regroupés au sein d'un super-opéron chez la plupart des archées a permis à Koonin et collaborateurs de proposer que la protéine aCPSF1 pourrait jouer un rôle similaire aux ribonucléases Rrp44 et Rrp6, portant l'activité ribonucléolytique, retrouvées associées au core du complexe exosome eucaryote (Koonin *et al.*, 2001).

Dans ce contexte, mes travaux ont permis tout d'abord de proposer que les protéines β -CASP seraient des ribonucléases avec un rôle majeur dans le métabolisme de l'ARN chez les archées (RESULTATS/PARTIE 1- Résultats publiés- Publications 1/2/3/4). Puis, avec l'objectif d'élucider les rôles physiologiques des ribonucléases β -CASP, aCPSF1 et aRNase J, les partenaires protéiques potentielles des ces deux ribonucléases ont été recherchés (RESULTATS/PARTIE 2- Résultats non publiés). Premièrement, nous avons établi les interactomes de aRNase J, aCPSF1 et de la protéine ASH-Ski2 annotée comme une hélicase à ARN appartenant à la famille des hélicase Ski2-like et montré que ces interactomes ont des entités communes. Deuxièmement la co-sédimentation du couple aRNase J/ASH-Ski2 avec les sous-unités du ribosome et de l'exosome permet de proposer une action concertée de ces

acteurs du métabolisme de l'ARN chez les archées. Enfin, la formation d'une interaction stable entre aRNase J/ASH-Ski2 et la co-distribution taxonomique des gènes codant pour aRNase J et ASH-Ski2 chez les euryarchées suggèrent l'existence d'un nouveau complexe, ribonucléase/hélicase, premier de ce type chez les archées (RESULTATS/PARTIE 2- Résultats non publiés). Finalement la caractérisation biochimique de ASH-Ski2 a permis de révéler son activité hélicase et sa capacité de liaison aux d'acides nucléiques (RESULTATS/PARTIE 3- Résultats non publiés). En absence d'ATP, ASH-Ski2 possède la capacité d'appareiller deux brins d'acides nucléiques. En présence d'ATP, la simple liaison de la protéine avec l'ATP est suffisante pour dérouler ces même duplexes. Nous avons aussi montré que les extensions N- et C-terminales de ASH-Ski2 jouent des rôles essentiels dans la modulation de ses propres activités et dans l'interaction avec son partenaire aRNase J. Il semblerait donc qu'un complexe ribonucléase/hélicase mosaïque avec une ribonucléase aRNase J, homologue à la RNase J bactérienne, et une hélicase ASH-Ski2, homologue aux hélicases à ARN Ski2 et Mtr4 eucaryotes, serait présent dans la majorité des euryarchées. Ainsi la mise en évidence de ce nouveau complexe multi-protéique pose la question de l'origine évolutive de ses composants.

MATERIELS

ET

METHODES

Construction des vecteurs d'expression des protéines d'intérêt

ORF	PROTEIN	EXPRESSION VECTORS			
		<i>(His)₆</i> -Nter	Cter- <i>(His)₆</i>		untagged
PAB1751	aRNase J	pET15b-Pab-aRNase J	pET21b-Pab-aRNase J		pET11b-Pab-aRNase J
PAB1868	aCPSF1	pET15b-Pab-aCPSF1			
PAB2313	ASH-Ski2	pET15b-Pab-ASH-Ski2	pET21b-Pab-ASH-Ski2	PJ414-Pab-ASH-Ski2	pET11b-Pab-ASH-Ski2
PAB0592	Hel308-Ski2	pET15b-Pab-Hel308-Ski2			pET11b-Pab-Hel308-Ski2
PAB0316	DnaG	pET15b-Pab-DnaG	pET21b-Pab-DnaG		pET11b-Pab-DnaG
PAB0420	Rrp41	pET15b-Pab-Rrp41			pET11b-Pab-Rrp41
PAB1284	RecJ-SOD	pET15b-Pab-RecJ-SOD			pET11b-Pab-RecJ-SOD

PROTEIN	EXPRESSION VECTORS	
	<i>(His)₆</i> -Nter	untagged
aRNase JΔ1		pET11b-Pab-aRNase JΔ1
ASH-Ski2-ΔN	pET15b-Pab-ASH-Ski2-ΔN	
ASH-Ski2-ΔC	pET15b-Pab-ASH-Ski2-ΔC	
ASH-Ski2-ΔNC	pET15b-Pab-ASH-Ski2-ΔNC	
ASH-Ski2-C124A	pET15b-Pab-ASH-Ski2-C124A	
ASH-Ski2-DomN	pET15b-Pab-ASH-Ski2-DomN	
ASH-Ski2-DomN-C98A	pET15b-Pab-ASH-Ski2-DomN-C98A	

Table M1. Tableaux des différents vecteurs utilisés pour les surexpressions des protéines. Les différents vecteurs sont classés selon qu'ils ajoutent une étiquette histidine ($(His)_6$) en N-terminal, en C-terminal des protéines ou ne présente pas d'étiquette. L'ensemble des constructions des différentes protéines utilisées sont représenté.

Différents plasmides d'expression ont été utilisés au cours de cette étude et sont listés dans la table M1. Succinctement, les vecteurs d'expression des protéines étudiées ont été construits par assemblage de produits de PCR avec le kit de clonage InFusion® (Ozyme). Les différents vecteurs ont été amplifiés par PCR inverse avec l'enzyme PrimeStarmax (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 a été incubée à 37°C en présence de 20 unités de DpnI (ThermoScientific) pendant 1h pour éliminer la matrice. Les gènes codant pour les protéines aRNase J, aCPSF1, ASH-Ski2,

Hel308-Ski2, DnaG, Rpr41 et RecJ-SOD ont été amplifiés par PCR à partir d'ADN génomique de *P. abyssi* avec 1 unité de Phusion (High-Fidelity DNA polymerase de Finnzymes), 0,2mM de dNTPs et 0.5µM des couples d'oligonucléotides appropriés (Programme thermocycleur: [98°C (30'') ; 58°C (30'') ; 72°C (1')*35cycles]. Afin de permettre l'assemblage des produits PCR dans la réaction InFusion®, les oligonucléotides permettant l'amplification des gènes codants les différentes protéines ont été dessinés pour intégrer, à leurs extrémités 5', 15 nucléotides complémentaires aux extrémités des vecteurs. Les réactions InFusion® ont été réalisées comme recommandé par le fournisseur avec des quantités équimolaires de produits de PCR. La totalité des réactions a été utilisée pour transformer 50µL de cellules STELLAR® chimio-compétentes (CLONETECH). Après un choc thermique à 42°C, les cellules ont été reprises avec 450µL de milieu SOC (composition : 4.8mL de milieu SOB, 0.4% glucose, 0.01M MgCl₂ et 0.01M MgSO₄) et incubées à 37°C pendant 1h. Les transformants ont été sélectionnés sur des boîtes de milieu LB agar contenant de l'ampicilline (100µg/mL) après une expression phénotypique d'une heure à 37°C. Les boîtes ont été incubées dans une étuve à 37°C toute la nuit.

L'intégration des inserts dans les différents vecteurs a été vérifiée par PCR sur colonies avec le couple d'oligonucléotides "T7 promoteur"/ "T7 terminateur". Les inserts ont été vérifiés par séquençage en utilisant le même couple d'oligonucléotides (La société MWG Eurofins).

Construction des vecteurs pET15b-ASH-Ski2-C124A et pET15b-ASH-Ski2-DomN-C98A

Les vecteurs pET15b-ASH-Ski2C124A et pET15b-ASH-Ski2-DomNC98A ont été construits par mutagenèse dirigée avec le kit QuikChangeXL® (Agilent) en suivant les recommandations du fournisseur. Brièvement, 10ng du plasmide pET15b-ASH-Ski2 ont été mis en présence de 2,5 unités de *Pfu* Ultra HF et de 125ng du couple d'oligonucléotides For C124A / Inv C124A qui contiennent la mutation à introduire (Programme thermocycleur : [95°C (50'') ; 60°C (50'') ; 68°C (12') *18cycles]. La réaction PCR a été incubée à 37°C en présence de 20 unités de DpnI (ThermoScientific) pendant 1h pour éliminer la matrice. La totalité des réactions a été transformée par choc thermique dans 45µL de cellules XL10-Gold® chimio-compétentes (Agilent) et les transformants ont été sélectionnés en présence d'ampicilline. L'intégration de la mutation C124A a été confirmée par séquençage avec le couple d'oligonucléotides B15-19 /B16-1(La société MWG Eurofins).

Surexpression et purification des protéines recombinantes

Le système hétérologue *E. coli* BL21-CodonPlus (DE3) a été utilisé pour la surexpression des protéines recombinantes à partir des différents vecteurs d'expressions présentés dans la **Table M1**. La souche *E. coli* BL21-CodonPlus (DE3) a été cultivée dans du milieu LB (Luria Berthani) dont la composition est de : 5g/L d'extrait de levure, 10g/L de NaCl, 10g/L de Tryptone, 20mg/L de thymine et dont le pH est ajusté à 7. L'expression des gènes codants les protéines étudiées a été induite avec 0.1mM d'IPTG en phase exponentielle de croissance ($DO_{600nm}=0,6$) pendant 3h à 30°C (Clouet-d'Orval *et al.*, 2010; Phung *et al.*, 2013). Les cellules sont ensuite lavées, culotées à 4500g pendant 15min à 4°C et stockées à -20°C.

Pour purifier les protéines, les culots cellulaires ont été traités au lysozyme (1mg/mL) et lysés par sonication. Après deux étapes de chauffage à 70°C durant 20min et une étape de clarification par centrifugation à 20000g pendant 20min, le surnageant est récupéré. Les protéines étiquetées Histidine sont purifiées sur des colonnes d'affinité HisTrap 1mL (GE-Healthcare) et sur des colonnes échangeuses d'anions Q (GE-Healthcare) ou héparine (GE-Healthcare) par FPLC. L'élution des protéines a été effectuée par différents concentrations (10mM et 50mM) et gradient (50mM à 500mM) d'imidazole. Pour vérifier la surexpression et la purification des protéines, des fractions des protéines éluées ont été séparées sur gel 4-12% SDS-PAGE (Biorad) et colorées au bleu de Coomassie (Instant Blue). Par la suite, les protéines éluées ont été dialysées avec 20mM du tampon Hepes à pH 7.5, 300mM NaCl, 1mM DTT et 10% glycérol et stockées à 16°C. Les protéines sont ensuite séparées sur des gels 4-12% SDS-PAGE avec une gamme BSA. Les gels ont été colorés au Sypro Orange et visualisés au bioimager (Biorad). La concentration des protéines a été déterminée par quantification des bandes avec le logiciel MultiGauge (Fujifilm).

Expérience de Pull-down et analyses protéomiques

L'expérience de Pull-down a été réalisée avec des extraits cellulaires de *P. abyssi* cultivée en continu et en phase exponentielle de croissance dans des bioréacteurs. Cette culture a été effectuée en condition anaérobique stricte et physiologique de croissance (95°C, pH 6.5) au laboratoire d'Ifremer (D. Flament, Brest). Les culots de cellules de *P. abyssi* ont été repris dans du tampon PBS1X (137mM NaCl, 2,7mM KCl, 10mM Na₂HPO₄, 1,76mM KH₂PO₄) (poids/volume 1/3), supplémenté avec 300mM de NaCl et des inhibiteurs de protéases (cOmplete™ de Roche #05892953001). Les cellules ont été lysées par sonication (VibraCell

Bioblock Scientific #4418) et l'extrait brut obtenu est clarifié par centrifugation à 20 000g pendant 30min à 4°C. 15 µL de billes magnétiques couplées au cobalt (Dynabeads His-tag Isolation & Pulldown # Cat no.101.04D) ont été utilisées pour chaque expérience en présence de 20µg de protéines recombinantes appâts étiquetées Histidine. Le mélange Billes-Protéines recombinantes appâts dans un tube eppendorf a été incubé pendant 15min à température ambiante et sous agitation dans un volume final de 300µl. Après trois étapes de lavage avec du tampon PBS1X, les billes ont été incubées avec 1mL de tampon PBS1X et 3mL d'extrait de *P. abyssi* pendant 2h à température ambiante et sous agitation dans des tubes à hémolyse. Après un lavage au tampon PBS1X, les billes ont été transvasées dans des tubes eppendorf 1.5mL et une partie des billes est traitée avec des nucléases (10µl de DNase I à 1U/µl et 10µl de RNase A à 10mg/ml) pendant 30min à température ambiante et sous agitation. Après des étapes de lavage, les billes ont été incubées à 95°C pendant 10min en présence de 10µL de XT sample buffer (4X) (Biorad) et 2µL d'agent réducteur (Biorad) et 38µL de H₂O. 20µL du surnageant ont été récupérés et séparés sur gel 12% SDS-PAGE (Criterion XT 12 puits) et migré sur 5mm. Les bandes ont été découpées en petit cube et envoyées aux plateformes d'analyse protéomique BioGenOuest (BGOP) et analyse protéomique de Paris Sud Ouest (PAPPSO). Pour la plateforme BGOP, les analyses ont été effectuées selon (Pluchon *et al*, 2013), sur la base de données SEQUEST. Pour la plateforme PAPPSO, l'abondance des spectres peptidiques MS/MS et les protéines ont été identifiées dans la base de données X!Tandem. Un test de student a été effectué afin de comparer les spectres présents dans les expériences « témoins sans appât » et celles en « présence d'appât ». Les protéines candidates retenues sont celles présentant un test t inférieur à 0.05. Les réseaux d'interaction ont été construits sur Cytoscape.

Collection et classification des hélicases SF1 et SF2 chez les archées.

Les hélicases appartenant aux superfamilles SF1 et SF2 sont identifiées par Reverse Position Specific BLAST (RPS BLAST) à partir des Pfam DEAD et Helicase C contre la base de données Complete Genomes Data Base (CGDB, version 11). 35 814 séquences sont identifiées et présentent au moins un des deux Pfam DEAD et Helicase C. Parmi ces séquences, 1059 appartiennent aux 110 génomes d'archée séquencés et annotés dans la base de données. Les séquences orthologues sont recherchées à partir des liens d'homologie issus des BLAST. Une partition des séquences orthologues par le Markov Cluster Algorithm (MCL) permet la formation de 52 classes. Les séquences obtenues par le MCL correspondant

aux gènes codant pour Hel308-Ski2 et ASH-Ski2 ont été sélectionnées. Après alignement par MUSCLE (Edgar, 2004) (<http://www.drive5.com/muscle/>), les résidus des régions peu conservés ou mal alignés sont enlevés de l'alignement multiple par trimAl (Capella-Gutiérrez *et al.*, 2009) (<http://trimal.cgenomics.org/>) et résidues conservées dans l'alignement multiple ont été mis en évidence avec WebLogo (Crooks *et al.*, 2004) (<http://weblogo.berkeley.edu/>).

L'arbre phylogénétique des archées a été construit à partir de protéines du ribosome issues de plus de 150 génomes d'archées séquencés utilisant l'approche des supermatrix avec une souche par espèce (Phung *et al.*, 2013). Les séquences ont été alignées par MUSCLE et les alignements ont été vérifiés sur SEAVIEW. Les arbres ont été générés par la méthode des maximums de vraisemblance avec PhyML en utilisant le model LG d'évolution des séquences et sont visualisés avec TreeDyn. Les arbres correspondant aux protéines aRNase J, ASH-Ski2 et Hel308-Ski2 ont été construits de la même manière.

Co-purification par affinité

La souche *E. coli* BL21-CodonPlus (DE3) portant les différents vecteurs d'expressions des protéines étudiées a été cultivée dans dans 200mL de milieu LB. L'expression des gènes codants les protéines étudiées a été induite avec 0.1mM d'IPTG en phase exponentielle de croissance ($DO_{600nm}=0,6$) pendant 3h à 30°C. (Clouet-d'Orval *et al.*, 2010; Phung *et al.*, 2013). Après centrifugation, les culots ont été traités au lysozyme (1mg/mL) et lysés par sonication. Après deux étapes de chauffage à 70°C pendant 20min, les fractions brutes ont été clarifiées par centrifugation à 20 000g à 4°C pendant 20min. Les extraits ont été traités avec de la DNase I (10mg/mL) et RNase A/T1 (10mg/mL) 30 mins à 37°C. Les surnageants ont été récupérés. Les protéines ont été purifiées sur des colonnes d'affinité HisTrap 1mL (GE-Healthcare) par FPLC. L'élution des protéines a été effectuée par différents paliers (10mM et 50mM) et gradient (50mM à 500mM) de concentration en imidazole. Les fractions des protéines éluées ont été séparées sur gel 4-12% SDS-PAGE (Biorad) et colorées au bleu de Coomassie (Instant Blue).

Western blotting

Les fractions des protéines éluées ont été séparées sur gel 4-12% SDS-PAGE et colorées au bleu de Coomassie (Instant Blue). Afin de réaliser des immunodétections de protéines spécifiques, les protéines séparées par SDS-PAGE ont été transférées sur une membrane de

nitrocellulose (Biorad) pendant 3 minutes en utilisant un Trans-Blot Turbo (Biorad). Avant l'incubation avec les anticorps primaires, les membranes ont été bloquées avec une solution PBS1X contenant 5% de lait pendant 2h. Après 3 étapes de lavage de 15 min avec du PBS1X, les membranes ont été incubées 2h à température ambiante avec des anticorps primaires dirigés contre les protéines d'intérêt. Les anticorps primaires ont été dilués dans du PBS1X au 1/10 000°. Après 3 lavages de 5 min au PBS1X, les membranes ont été incubées 1h à température ambiante avec du PBS1X contenant l'anticorps secondaire couplé à la HRP (Horseradish Peroxidase) dilué au 1/5000° (Clontech #631210). Après 3 lavages avec du PBS1X, les protéines ont été détectées par chimiluminescence avec l'appareil ChemiDoc™ (Biorad) après traitement des membranes avec des réactifs ECL (Biorad #170-5061) suivant les recommandations du fabricant.

Entre deux immunodétections, les membranes selon le besoin, ont été incubées avec 15mL de solution Restore™ Plus Western (ThermoScientific) à température ambiante sous faible agitation pour décrocher les anticorps préalablement utilisées. Avant un nouveau blocage au lait, les membranes ont été lavées 3 fois pendant 5 min avec du PBS1X.

Fractionnement sur gradient de sucrose

Les gradients 5-25% ou 10-30% sont réalisés dans des tampons différents : le tampon PBS1x (140mM NaCl) ou le tampon TK (20mM de Tris-HCl, 100mM de KCl, 10mM de MgCl₂, 1mM de DTT) supplémentés ou non avec 20mM d'EDTA. 5mL de solution de saccharose la plus concentrée à 25% ou 30% ont été tout d'abord mis dans un tube d'ultracentrifugation (11mL BECKMAN). Ensuite délicatement, 5mL de solution à 5% ou 10% ont été ajoutés. Le gradient se forme en mettant le tube à l'horizontale durant 3h, puis transféré à la verticale durant toute la nuit à 4°C. 200mg de cellules de *P. abyssi* ont été pesées et reprises dans 3 volumes de tampon PBS1X, PBS1X+ EDTA, TK ou TK + EDTA auxquels aura été ajouté un inhibiteur de protéases. Puis, les cellules ont été lysées par sonication (10x10sec, cycle 50%, puissance 5, VibraCell-Bioblock scientifique), et les extraits ont été clarifiés par centrifugation à 21 000g pendant 30min. Enfin, après avoir récupéré le surnageant, 500µL ont été déposés en haut de chacun des gradients de saccharose pour être ensuite ultracentrifugé (Beckman-Avanti XPN-80, rotor SW41) à 200 000g à 4°C durant 5h. Une solution de saccharose (50% dans H₂O) est introduite par le bas du gradient à 500µL/min. Des fractions de 500µL chacune, ont été collectées. Les fractions de 500µL ont été ensuite aliquotées par 250 µL dans deux tubes, dont une série servira pour l'extraction des ARN et l'autre série sera

utilisée pour la précipitation des protéines. Ainsi pour chaque gradient réalisé, 21 fractions ont été obtenues.

Extraction des ARN

L'ARN est extrait à partir de 250µL de chaque fraction des différents gradients. Pour chaque fraction de 250µL, 400µL de Trizol sont ajoutés ainsi que 300µL de chloroforme. Les tubes sont ensuite agités, incubés (5min à température ambiante), puis centrifugés (15min, à 21 000g). La phase aqueuse contenant les ARN est récupérée, puis ajout d'un volume équivalent au volume de la phase aqueuse d'isopropanol. Les tubes sont ensuite incubés (30min à température ambiante), puis centrifugés (30min à 15-20°C à 20 000g). Le culot est lavé avec 350µL d'éthanol à 75%. Une centrifugation est réalisée (10min à 15-20°C à 20 000g), le surnageant est enlevé, et les tubes sont centrifugés à nouveau à vide (2min à 20 000g). Par la suite, une étape de séchage est réalisée (10 à 15min sous la hotte à température ambiante). Les culots sont re-suspendus dans 10µL d'eau MQ et les échantillons sont chauffés durant 10min à 65°C. Les fractions subissent ensuite un traitement à la DNase I afin d'éliminer entièrement l'ADN des échantillons. Le traitement est effectué à partir des tubes de 10µL des ARN précipités dans l'eau MQ. Dans un premier temps, 78µL d'eau sont ajoutés aux échantillons, puis ajout de 10µL de tampon de DNase I et 2µL de l'enzyme DNase I et incubation des tubes 1h à 37°C. Par la suite, des tubes phase lock gel sont utilisés, dans lesquels sont ajoutés volume/volume de phénol-chloroforme-isoamyl alcool avec la totalité de l'échantillon traitée à la DNase I. Les tubes sont mélangés par retournement, puis incubés (5min à température ambiante) et centrifugés (15min à 15°C à 20 000g). Le surnageant est ensuite récupéré et les ARN sont précipités avec 150µL d'éthanol sodium acétate. Les tubes sont mis 30min à -80°C, puis centrifugés (30min à 4°C à 20 000g). Le culot est lavé avec 150µL d'éthanol 75%. Une centrifugation est réalisée (10min à 15-20°C à 20 000g) puis le surnageant est enlevé, et les tubes sont re-centrifugés à vide (2min à 21 000g). Par la suite, les tubes sont ouverts et sont laissés à sécher (10 à 15min sous la hotte à température ambiante). Les culots sont re-suspendus dans 10µL d'eau MQ et les échantillons sont chauffés durant 10min à 65°C.

Le matériel est entièrement nettoyé avec de l'eau, puis de l'éthanol et enfin de l'inhibiteur de RNase pour obtenir un matériel « RNase-free ». Le gel d'agarose 1% est préparé dans du TBE 1X (89mM TRIS + 89mM Acide Borique + 1mM EDTA). Après ajout de 10µL de LBII (Bleu de bromophénol, Bleu de xylène, EDTA, Formamide) les échantillons sont chauffés (3min à 95°C) puis déposés sur le gel ainsi qu'un marqueur de taille 1kB, la migration s'effectue à

100V pendant 2h30. Les ARN et le marqueur sont révélés par coloration du gel au SYBR Green II 0,01% dans du TBE 1X, pendant 20min, et puis sont visualisés par fluorescence au Typhoon.

Précipitation des protéines.

La précipitation des protéines a été faite sur la deuxième série des fractions de 250 μ L obtenues sur le gradient saccharose 10-30% dans le tampon TK et dans le tampon TK + EDTA.

Dans un premier temps les protéines des fractions de 250 μ L ont été précipitées avec 45 μ L de TCA (Acide Trichloacétique). Les échantillons ont été mélangés par retournement, congelés (30 min à -20°C), puis centrifugés (30min à 4°C à 16000g). Le culot de protéines a été lavé avec 500 μ L d'acétone 100%. Après une étape de centrifugation à 16000g pendant 15 min à 4°C, le surnageant a été enlevé et les tubes ont été séchés 10 à 15 min sous la hotte. Les protéines ont été reprise avec 20 μ L de Tris-HCl à 20mM (pH 8) et 4 μ L de tampon de charge 6X (tampon de charge 1X : 62,5mM Tris-HCl (pH 6,8), 2% de SDS, 5% de β -Mercaptoéthanol, Bleu de bromophénol, 0,01% glycérol (pH 6,2). 1 μ L de NaOH a été ajouté selon le besoin pour avoir un pH approprié. Les fractions de protéines, préalablement dénaturées à 95°C pendant 5 min, ont été séparées à 150V pendant 50 min par SDS-PAGE sur un gel 4%-15% (Biorad) dans un tampon TGX1X (Biorad). Les immunodétections des protéines spécifiques ont été réalisées comme décrit précédemment.

Quantification relative des ARN et des protéines

La quantification des bandes correspondants aux protéines et aux ARN d'intérêt a été réalisée respectivement avec le logiciel ImageLab (Biorad) et MultiGauge (Fujifilm).

Test ATPase de l'hélicase ASH-Ski2 et ses variants

Pour le test ATPase, la réaction finale de 15 μ L été constituée de 250nM de l'hélicase ASH-Ski2 purifiée, d'un tampon constant de 50mM Hepes pH7,5, 50mM KCl, 5mM MgCl₂, 2mM DTT, 2mM ATP, 0,075 μ l [γ 32P] ATP à 6000Ci/mmol et en absence ou en présence de 0,2mg/mL d'ARN totaux de *P. abyssi*. La protéine Pab-ASH-Ski2 a été pré-incubée à 60°C avec les substrats d'acides nucléiques pendant 10 min. La réaction d'hydrolyse de l'ATP a été déclenchée par l'incubation de l'enzyme avec l'ATP [γ 32P] à 65°C. Au temps 0 min, 15 min, 30 min et 60 min, 2 μ L de la réaction sont prélevés, déposés sur une feuille de TLC

(chromatographie sur couche mince) et séparés dans 100mL de KH₂PO₄ à 250mM. Les feuilles de TLC préalablement séchées ont été placées dans des cassettes en présence d'un écran phospho-stimulable (BAS-MS Imaging Plates-Fujifilm) pendant la nuit. Les écrans ont été scannés par un phospho-imager (Typhoon Trio-Amersham Biosciences) avec une résolution de 200 microns. La quantification de l'ATP hydrolysé a été réalisée avec le logiciel MultiGauge (Fujifilm).

Test d'interaction protéines/ acides nucléiques

Les séquences des substrats ADN (S1 & 5RS1) et ARN (US1 & U5RS1) testés ont été listées dans la **Table M2**. L'affinité de l'hélicase *Pab*-ASH-Ski2 pour les différents substrats a été analysée par double filtration sur membrane nitrocellulose et membrane Hybond N+. Les substrats S1, 5RS1 et U5RS1 ont été phosphorylés en 5' avec 50µCi de γ 32P-ATP par 5 unités de T4 polynucléotide kinase (ThermoScientific). Le substrat US1 a été 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 de α 32P-UTP. Les duplexes ADN/ADN (S1/RS1) ou ADN/ARN (S1/U5RS1) ont été obtenus par incubation des différents simples brins à 90°C durant 5 min et une étape de refroidissement lente jusqu'à température ambiante (environ 30 min) dans un tampon SSC 1X (150mM NaCl, 15mM NaCi). Les duplexes ont été dessalés (Micro-bio Spin, Biorad) avant utilisation lors des tests.

Une gamme de dilution de la protéine dans du tampon de dialyse a été réalisée (de 0 à 350nM) et pré-incubée à 60°C pendant 10 minutes. Puis, un mix contenant 0,5nM de substrats radiomarqués et du tampon hélicase (1X) ont été ajoutés. Après une incubation de 15 min à 30°C, la réaction a été arrêtée dans la glace. Les différentes réactions ont été déposées sur un sandwich composé d'une membrane de nitrocellulose (traitée avec 0,5M de KOH et du tampon hélicase 1X) et d'une membrane d'Hybond (traitée avec du tampon hélicase 1X) dans un SlotBlot. Les membranes préalablement séchées ont été placées dans une cassette en présence d'un écran phospho-stimulable (Fujifilm). Les écrans ont été scannés par un phospho-imager (Typhoon Trio) avec une résolution de 200 microns. La quantification des dépôts a été réalisée avec le logiciel MultiGauge. Les courbes et l'analyse des K_d ont été réalisées à l'aide du logiciel GraphPadPrism 7.

Nom	Séquence	Structure
ADN S1	5' ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3'	5' ————— 3'
ARN US1	5' AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU 3'	p* —————
ADN 5RS1	5' GGACATGCTGTCTAGAGACTATCGAT 3'	p* —————
ARN U5RS1	5' GGACAUGCUGUCUAGAGACUAUCGAU 3'	p* —————
Hétéroduplex S1/U5RS1	5' ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3' 3' UAGCUAUCAGAGAUUCUGUCGUACAGG 5'	p* —————
Homoduplex S1/RS1	5' ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3' 3' TAGCTATCAGAGATCTGTCGTACAGGATCGTTCGGTCTTAAGCCGTCGCA 5'	p* —————

Table M2. Tableau des substrats utilisés pour la réalisation des mesures d'affinité. La radioactivité est portée par l'extrémité 5' des substrats, l'hétéroduplex est marqué sur l'ARN U5RS1 et l'homoduplex sur le brin d'ADN S1. Les brins ADN sont représentés en noir et les brins ARN en rouge.

Tests d'appariement et de déroulement de brins

Les séquences des substrats ADN S1 et ARN U5RS1 ont été utilisées et sont listés dans la **Table M2**. La réaction est effectuée dans 45 µl final contenant 5nM d'ADN S1, 5nM de l'ARN U5RS1* radiomarqué et 250nM de ASH-Ski2 dans un tampon Helicase 1X (25mM Tris-HCl pH8, 50 mM NaAc, 5mM MgCl₂ et 2.5mM 12.5mM β-mercaptoéthanol). La réaction est effectuée à 65°C. 6µl de réaction sont prélevés à 0 ; 0.25; 0.5; 1; 2 ; 3; 5 et 10 minutes et les réactions sont arrêtées avec 1,25 µl d'un tampon stop 1X (0.1% SDS, 8mM EDTA, 0.1 mg/mL Protéinase K, 4% glycérol et 0.02% BPB) ont été migrées sur gel acrylamide 8%, 0.1% SDS. Les gels ont été séchés et exposés sur un écran d'autoradiographie. La quantité de radioactivité des différentes bandes est mesurée par autoradiographie (Typhoon) et le pourcentage des différents formes libre ou en duplex sont calculés sur «MultiGauge». Les courbes correspondant sont tracées sur GraphPadPrism 7.

Les tests de l'activité de déroulement de duplexes ADN/ARN (S1/U5RS1*) sont effectués dans les mêmes conditions expérimentales que pour les tests d'appariement de brins. Après 2 minutes de réaction de formation de duplex, 2 µl d'ATP, d'ATP-γ-S, d'ADP à 100mM ou un contrôle avec de l'eau sont ajoutés à la réaction. La révélation et la quantification de duplex formés sont effectuées comme précédemment.

RESULTATS

PARTIE I

IDENTIFICATION ET CLASSIFICATION DES RIBONUCLEASES DE LA FAMILLES DES β -CASP CHEZ LES ARCHEES ET CARACTERISATION DE aCPSF1 DE *P. abyssi*

Dans cette partie, nous présentons les articles publiés en début et au cours de ma thèse définissant les différents groupes d'orthologues des ribonucléases de la famille des β -CASP chez les archées et l'activité enzymatique des ribonucléases du groupe aCPSF1 (PUBLICATIONS 1, 2 et 3, voir ci-après).

Nous avons montré que parmi les 110 génomes d'archées séquencés et annotés, les ribonucléases β -CASP sont regroupées au sein de 4 groupes. Le groupe des aRNase J qui est orthologue des RNase J bactériennes impliquée dans la dégradation des ARN et la maturation des ARNr. Les groupes des aCPSF1, aCPSF2, aCPSF1b proches de la protéine CPSF73 qui est le facteur de terminaison de la transcription eucaryote et fait partie intégrante du complexe CPSF (PUBLICATIONS 1 et 2/3).

Le groupe aRNase J est présent dans la majorité des Euryarchées présente une activité exoribonucléolytique 5'-3' très processive (Clouet et al 2010). Le groupe aCPSF1 présent dans l'ensemble des archées présente deux domaines KH et un motif en C-terminal impliqué dans la dimérisation de la protéine (PUBLICATION 1).

Les tests enzymatiques de aCPSF1 de *Pyrococcus abyssi* (*Pab*-aCPSF1) ont permis de déterminer que la protéine possède une activité endoribonucléolytique préférentiellement sur des dinucléotique CA en simple brin et une activité exoribonucléolytique 5'-3' sur des substrats monophosphate (PUBLICATIONS 1 et 4). Ces activités sont similaires à celles décrites pour l'orthologue eucaryotes, CPSF73, Les domaines KH présents dans la partie N-terminale de aCPSF1 est importante pour la coupure endoribonucléolytique de certains site spécifique et la formation de gros complexes ribonucléique stables. La dimérisation de aCPSF1 est cruciale pour son activité exoribonucléolytique et pour l'interaction avec les substrats ARN. L'ensemble de ces tests enzymatiques a été effectué en tenant en compte du caractère hyperthermophile de notre modèle d'étude, *P. abyssi* (PUBLICATIONS 4). Les conditions expérimentales particulières ont ainsi été mises en place pour caractériser ces

protéines, de la surexpression des protéines, à leur purification en passant par leurs tests *in vitro* de leurs activité endo- et exoribonucléolytique à haute température.

L'ensemble de ces approches suggère un rôle essentiel de aCPSF1 et aRNase J dans le métabolisme de l'ARN chez les archées (PUBLICATIONS 2/3).

**IDENTIFICATION ET CLASSIFICATION DES RIBONUCLEASES DE LA
FAMILLES DES β -CASP CHEZ LES ARCHEES ET CARACTERISATION
DE aCPSF1 DE *P. abyssi***

PUBLICATION 1

**Archaeal β -CASP ribonucleases of the aCPSF1 family are
orthologs of the eukaryal CPSF-73 factor**

Duy Khanh Phung, Dana Rinaldi, Petra S. Langendijk-
Genevaux, Yves Quentin, Agamemnon J. Carpousis, and Béatrice
Clouet-d'Orval.

Centre National de la Recherche Scientifique, LMGM and Université de Toulouse, Toulouse,
France

Archaeal β -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor

Duy Khanh Phung, Dana Rinaldi, Petra S. Langendijk-Genevaux, Yves Quentin, Agamemnon J. Carpousis and Béatrice Clouet-d'Orval*

Centre National de la Recherche Scientifique, LMGM and Université de Toulouse, UPS, F-31062 Toulouse, France

Received July 10, 2012; Revised October 22, 2012; Accepted November 1, 2012

ABSTRACT

Bacterial RNase J and eukaryal cleavage and polyadenylation specificity factor (CPSF-73) are members of the β -CASP family of ribonucleases involved in mRNA processing and degradation. Here we report an in-depth phylogenomic analysis that delineates aRNase J and archaeal CPSF (aCPSF) as distinct orthologous groups and establishes their repartition in 110 archaeal genomes. The aCPSF1 subgroup, which has been inherited vertically and is strictly conserved, is characterized by an N-terminal extension with two K homology (KH) domains and a C-terminal motif involved in dimerization of the holoenzyme. Pab-aCPSF1 (*Pyrococcus abyssi* homolog) has an endoribonucleolytic activity that preferentially cleaves at single-stranded CA dinucleotides and a 5'-3' exoribonucleolytic activity that acts on 5' monophosphate substrates. These activities are the same as described for the eukaryotic cleavage and polyadenylation factor, CPSF-73, when engaged in the CPSF complex. The N-terminal KH domains are important for endoribonucleolytic cleavage at certain specific sites and the formation of stable high molecular weight ribonucleoprotein complexes. Dimerization of Pab-aCPSF is important for exoribonucleolytic activity and RNA binding. Altogether, our results suggest that aCPSF1 performs an essential function and that an enzyme with similar activities was present in the last common ancestor of Archaea and Eukarya.

INTRODUCTION

RNA processing and degradation are critical to the survival of all cells and acknowledged as a means of regulating gene expression. In particular, the nature of

RNA 5' and 3'-ends is known to have major impact because they control the entry and directionality of endo- and exoribonucleases involved in these processes. In Archaea, exploration of RNA processing and degradation pathways is still in its early stages. Because easy genetic approaches are not readily available, functional studies in the Archaea are often based on genomic and proteomic analyses that are interpreted in light of our understanding of RNA metabolism in Bacteria and Eukarya. Unlike its eukaryal counterpart, archaeal mRNA is not capped at its 5'-end nor is it polyadenylated at its 3'-end [(1) for review]. Transcription in the Archaea is performed by a eukaryal-like RNA polymerase that initiates at 'TATA-boxes' [for review see (2)]. However, little is known about mRNA 3'-end maturation and transcription termination.

Processing of the 3'-end is an essential step in converting eukaryotic pre-mRNAs to mature polyadenylated mRNAs [for review see (3)]. This process is executed by the cleavage and polyadenylation macromolecular complex, which is well-described in both yeast and mammals (4–9). Among other proteins, this machinery includes the cleavage and polyadenylation specificity factor (CPSF-73), a 73 kDa subunit that carries out the endonucleolytic cleavage at a CA motif 20–30 nt downstream of the AAUAAA consensus sequence before polyadenylation (10,11). In the maturation of metazoan histone pre-mRNA, CPSF-73 has also been shown to act as a 5'-3' exoribonuclease in the degradation of the transcript downstream of the cleavage site (12). Based on sequence similarity, a gene encoding a homolog of the eukaryotic CPSF-73 has been reported to be prevalent in archaeal genomes (13) raising the question of its role in archaeal RNA metabolism. Recently, CPSF-73 homologs in *Methanocaldococcus jannashii* and *Methanothermobacter thermotrophicus* have been shown to have nuclease activity (14–16). However, the enzymatic properties and specificities of archaeal CPSF(aCPSF)-73 homologs remain to be clearly delineated.

*To whom correspondence should be addressed. Tel: +33 561335816; Fax: +33 561335886; 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.

The eukaryotic CPSF-73 is a member of the β -CASP family of metallo- β -lactamases (17), which includes ribonucleases important in RNA metabolism that are widespread in all three domains of life (18). The β -CASP proteins have the first four signature motifs of the metallo- β -lactamase superfamily followed by a distinct region [β -CASP domain, (19)] that is characterized by three short conserved motifs A (Asp or Glu), B (His) and C (His). These enzymes use a zinc-dependent mechanism in catalysis and act as 5'-3' exonucleases and/or endonucleases (19,20). Examples of β -CASP nucleases include RNase J1, a key player of *Bacillus subtilis* RNA metabolism that functions as an endonuclease and a 5'-3' exonuclease (21-23). We have shown recently that orthologs of bacterial RNase J with 5'-3' exonuclease activity are widespread in the Euryarchaea (24). Similar activity has been described for a closely related β -CASP protein of the crenarchaeon *Sulfolobus solfataricus* (25). Moreover, several crystal structures of aCPSF-73 homologs in *M. thermoautotrophicus*, *Pyrococcus horikoshii* and *Methanosarcina mazei* have been solved. These proteins are dimeric (14,26) and have a tripartite architecture consisting of a N-terminal region with two K homology (KH) RNA-binding motifs, a central β -metallo-lactamase domain and a C-terminal β -CASP domain (14,16,26). Altogether, identification of homologs of the eukaryotic CPSF-73 and of bacterial RNase J in the Archaea raises the question of their role in RNA metabolism and their evolutionary origin.

The metallo β -lactamase protein superfamily is highly represented in Archaea (26). Among them, the archaeal β -CASP family members were proposed to be RNA hydrolases as their sequences are more closely related to bacterial RNase J and to eukaryotic CPSF-73 than to β -CASP proteins involved in DNA repair and recombination (19). Here we report the inventory, classification and phylogenetic analysis of the archaeal β -CASP proteins, allowing us to identify seven β -CASP clusters. Among them, one is clearly related to bacterial RNase J [archaeal RNase (aRNase J)] and three are related to CPSF-73. Members of the aCPSF1 group are present in all Archaea whose genomes have been sequenced. We show that Pab-aCPSF1 from the thermococcal archaeon *Pyrococcus abyssi* has the same activities and specificity as its eukaryotic counterpart, CPSF-73. Thus, aCPSF1 family members are authentic orthologs of CPSF-73.

MATERIALS AND METHODS

Collection of archaeal β -CASP proteins

Genome entries of the complete archaeal and bacterial genomes were retrieved from EMBL (<http://www.ebi.ac.uk/genomes/>) and processed by a set of perl programs into a MySQL database. We used the RPS-Blast program to annotate protein sequences according to the conserved domain database available at the NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and computed pairs of one-to-one ortholog genes with BlastP as follows: two genes *a* and *b* from genomes *A* and *B*, are considered to be orthologs if *a* is the best hit of *b* in

genome *A* and reciprocally, and if *a* (or *b*) has a paralogous gene named *c* then the score of *a* versus *b* should be greater than the score of *a* (or *b*) versus *c*. The proteins of the β -CASP family have the first four signature motifs of the metallo- β -lactamase superfamily followed by a distinct globular domain [named β -CASP domain, (19)] that is used to identify new members of the family. Callebaut *et al.* (19) identified a list of β -CASP family members in eukaryotes, bacteria and archaea. The archaeal list included only 14 species and 40 β -CASP candidate proteins. To update the annotation of the β -CASP proteins in the 110 complete archaeal genomes, we used the β -CASP domain of each candidate sequence as query in a Psi-Blast search against the protein sequences of complete archaeal genomes. To maximize the sensibility of the prediction, we set the maximal *E*-value $< 1e-05$ and the maximal number of iterations to 20 to be able to recover all putative candidates in each individual Psi-Blast search. This resulted in an initial collection of 375 proteins.

Protein classification

For each protein of the initial collection, we retrieved orthologs in each archaeal genome to obtain orthologous proteins pairs. Among these, 13 proteins were not identified by the PsiBlast search. A graph was produced where vertices correspond to proteins and edges to orthologous relationships. This graph included six connected components. The application of a partition algorithm [MCL with an inflation operator setup to 1.2 (27)] revealed nine well-defined groups of proteins (orthologous groups, OGs). The protein members of two groups (GloB, COG0491, Zn-dependent hydrolases including glyoxylases and MtrA, COG4063, tetrahydromethanopterin S-methyltransferase, subunit A) do not have the A, B and C β -CASP signature motifs. Hence, they were false positives and were discarded from further analysis. The other OG proteins possess the signature of the β -CASP family and fall in three related COGs: COG1782 (predicted metal-dependent RNase, consisting of a metallo-beta-lactamase domain and an RNA-binding KH domain), COG1236 (predicted exonuclease with beta-lactamase fold involved in RNA processing) and COG0595 (predicted hydrolase of the metallo-beta-lactamase superfamily). As the different protein groups were not found in all archaeal genomes, we systematically searched for putatively unannotated genes (missed by annotation or pseudogenes) with the TblastN program. The strategy used to compute the clusters of orthologous proteins was validated by the very low level of paralogy observed in each group (less than three paralogs per genome).

Alignments and trees computation

The archaeal phylogeny was inferred from a set of 53 ribosomal proteins using the super matrix approach (28) with only one strain per species. Sequences were aligned using the MUSCLE program (29). The alignments were inspected and manually refined using the SEAVIEW sequence editor (30) and trimmed using trimAl (31).

These parsed alignments were concatenated to produce a single alignment of 6938 residues. The maximum likelihood trees were computed with PhyML (32) using the LG model of sequence evolution. The gamma-distributed substitution rate variation was approximated by four discrete categories with shape parameter and proportion of invariant sites estimated from the data. Non-parametric bootstrap values were computed (100 replications of the original dataset) using the same parameters. Trees were visualized and annotated with TreeDyn (33).

Construction of vectors for the expression of Pab-CPSF1 and variants

The supplementary Table S1 summarizes oligonucleotides used in this study. The plasmid used for expression of HIS-tagged Pab-CPSF1 was pET15b. The coding sequence (*PAB1868*) was amplified by PCR from *P. abyssi* genomic DNA and cloned as an XhoI-BamHI PCR fragment using OLC5 and OLC3 oligonucleotides to give the plasmid pEC-Pab-CPSF1. The Δ KH and Δ Cter variants were constructed using Δ OL Δ KHC5/OLC3 and OLC5/OL Δ C3 oligonucleotide pairs. The H261A and H594A variants were generated by site-directed mutagenesis of pEC-Pab-CPSF1 with the appropriate oligonucleotides (OLIA261/OL2A261; OL1A594/OL2A594, respectively) and the QuikChange II XL Kit (Stratagene).

Overexpression and purification of wild type Pab-aCPSF1 and variants

The BL21-CodonPlus (DE3) *Escherichia coli* strain carrying pEC-Pab-aCPSF1 or plasmids bearing mutations was induced at an OD₆₀₀ of 0.6 by addition of 0.1 mM IPTG, and incubated 3 h at 30°C. A cell extract was heated to 70°C for 10 min and clarified by low-speed centrifugation. TALON Metal resin (Clontech) was used for IMAC purification of the HIS-tagged protein. Proteins (~10 μ g/ μ l) were dialysed against 20 mM Hepes pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT and 1% glycerol and stored at 4°C. The His-tag was removed by treatment with thrombin (0.5 u/ μ l) for 2 h at room temperature, before incubation at 70°C for 15 min and centrifugation to recover the supernatant. An aliquot of the purified protein was analysed by Coomassie stained 10% SDS-PAGE and migrated at 72 kDa in agreement with the predicted molecular mass of 72 kDa for the wild type protein and at 55 kDa for the Δ KH variant.

RNA synthesis and labeling

In vitro transcription with T7 RNA polymerase was performed as described by the manufacturer (Promega) using PCR fragments as templates. The sR47 and sRkB templates were prepared as described previously (34,35) and the sR47MutCG and sR47Mut21U using oligonucleotides OLSr47MutCG5/OLsR47MutCG3 and OLSr47Mut21U5/OLsR47Mut21U3, respectively (Supplementary Table S1). [α -³²P] UTP and [γ -³²P] GTP were added to the *in vitro* transcription mix to synthesize uniformly labeled transcripts and 5'-end triphosphorylated labeled RNAs (p*pp RNA), respectively.

5'-end monophosphate RNA labeling was performed on dephosphorylated RNA or synthetic RNA with T4 polynucleotide kinase in the presence of [γ -³²P] ATP. 3'-end labeling was carried out with T4 RNA ligase in presence of [$5'$ -³²P] pCp and DMSO. All labeled RNAs were purified on denaturing 8 or 10% PAGE.

RNase assay

A typical enzyme excess reaction in a final volume of 15 μ l contained 5 nM ³²P-RNA, 6 μ M wild type Pab-aCPSF1 or variants, 20 mM Hepes, pH 7.5, 100 mM KCl, 1.5 mM MgCl. Reactions were started by addition of the enzyme and incubated at 65°C and repeated in at least three independent experiments. Samples of 4 μ l were withdrawn at the indicated times and the reactions were stopped by incubation with proteinase K (20 u) for 10 min at 37°C before addition of formamide-containing dye supplemented with 10 mM EDTA or spotted directly on thin layer chromatography (TLC) plates (PEI-cellulose, Nagel). The samples and T1/OH ladders were denatured for 1 min at 95°C before separation on 10% PAGE/8 M urea sequencing gels. TLC plates were developed with 0.25 M KH₂PO₄ and gels were dried before analysis using PhosphoImager and MultiGauge software.

Electrophoretic mobility shift assay

EMSA was performed as previously described (34). RNA and ribonucleoprotein (RNP) complexes were separated on a native 5% (19:1) polyacrylamide gel containing 0.5 \times TBE and 5% glycerol. Electrophoresis was performed at room temperature at 250 V in 0.5 \times TBE running buffer containing 5% glycerol. The gels were dried and visualized using a Fuji-Bas 1000 phosphorImager.

Size exclusion chromatography

After IMAC purification, Pab-aCPSF1 and its variants were concentrated by ultrafiltration (Millipore Anicon Ultra 30 K) and loaded onto a Superdex 200 10/300GL gel filtration column (GE Healthcare), pre-equilibrated in 20 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM DTT, 1 mM EDTA and 1% glycerol. The protein standard kit (GE-healthcare) containing ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa) was used to estimate the molecular mass of the native protein. The flow rate was fixed at 0.5 ml/min and elution of the protein was monitored by absorbance at 280 nm.

RESULTS

Groups of orthologous β -CASP proteins in the Archaea.

We undertook a detailed analysis of all archaeal β -CASP members to define OGs and to elucidate their evolutionary relationships. We collected β -CASP sequences from 110 complete archaeal genomes and classified them based on sequence conservation as described in Methods. Of the nine clusters that emerged from this analysis (Supplementary Figure S1A), members of the GloB and MtrA

clusters did not have the conserved A, B and C motifs characteristic of β -CASP proteins; members of the $\alpha\beta$ Cx cluster had non-canonical spacing between the A and B sequence motifs (Supplementary Figure S1B); members of the $\alpha\beta$ Cy and $\alpha\beta$ Cz clusters were not monophyletic, suggesting complex evolution that might include horizontal gene transfers with bacteria. These OGs were not further considered here. Using the remaining four groups of orthologous β -CASP proteins, we constructed a tree that was rooted with eukaryotic CPSF-73 and bacterial RNase J (Figure 1A). These OGs, which we named aRNase J, aCPSF1, aCPSF1b and aCPSF2, correspond to distinct and specific subtrees. This configuration validates our classification procedure. Apart from the members of the major aCPSF1 group harboring an N-terminal extension of 110 amino acids, archaeal β -CASP candidates are commonly restricted to the β -CASP and metallo- β -lactamase core domains (composed of an average of 420 amino acids) with no additional N- or C-terminal extensions (see below). Note that the *S. solfataricus* aCPSF2 member (Sso 0386) with a 40 amino acids N-terminal region, is peculiar among the aCPSF2 group.

Bacterial and aRNase J were separated from the CPSF-like OGs by a long branch (100% bootstrap support), suggesting an early evolutionary separation. Bacterial RNase J is distinguished from the archaeal homologs by a characteristic C-terminal extension (Figure 1B). The 65 aRNase J members were exclusively present in Euryarchaea as described previously (24) (Supplementary Figure S2). Recently, three members of this group have been reported to have 5'-3' exoribonuclease activity (15,24) (Table 1) and a phylogenetic analysis showed that aRNase J has been inherited vertically, suggesting an ancient origin predating the separation of Bacteria and Archaea (24).

Proteins related to eukaryal CPSF-73 clustered into three OGs: aCPSF1 (112 members) and aCPSF1b (11 members) corresponding to COG1782 and aCPSF2 (80 members) corresponding to COG1236. The aCPSF2 OG is distributed among Crenarchaeota, Euryarchaeota and Thaumarchaeota (Supplementary Figure S2). One member of this subgroup, which was misidentified as RNase J orthologs, has been reported to have a 5'-3' exonucleolytic activity (25) (Figure 1A and Table 1). The aCPSF1 OG corresponds to a highly conserved family with an N-terminal extension containing two KH RNA binding motifs specific to this group, and a C-terminal motif that is part of a protein dimer interface (Figures 1B and 2). This OG is notable because of its remarkable conservation in all Archaea with no exception to date (Supplementary Figure S2). Moreover, the congruence between the archaeal and aCPSF1 phylogenetic trees (Figure 3) shows that aCPSF1 has been inherited vertically, suggesting an ancient origin predating the emergence of Archaea. The small aCPSF1b OG branching close to the aCPSF1 OG is restricted to the Methanococcales (Figure 1A and Supplementary Figure S2). Only one member (MJ0162, misidentified as RNase J orthologs) is biochemically characterized and harbors 5'-3' exonucleolytic activity (25) (Figure 1A and Table 1). The aCPSF1b proteins, which appear to have an

undecipherable ancient origin, lack the N-terminal extension that is characteristic of the aCPSF1 family.

The crystal structure of three aCPSF1 members has recently been reported (14,16,26) as well as endoribonuclease activity for one of them (15) (Figure 1A and Table 1). However, little is known about the substrate specificity and enzymatic properties of the aCPSF1 members. To better characterize the ubiquitous aCPSF1 OG, we investigated the properties of the *P. abyssi* member of this group. The *P. abyssi* genome contains three open reading frames with β -CASP protein signatures. *PABI751*, *PABI868* and *PABI035* are members of the aRNase J, the aCPSF1 and the highly divergent $\alpha\beta$ Cy cluster, respectively (Figure 1A and Supplementary Figure S1A). *PABI035* was not further considered here. *PABI751*, denoted as Pab-aRNase J, corresponds to the recently identified ortholog of the bacterial RNase J (Figure 1B) (24,36). This protein has been shown to have a highly processive 5'-end-dependent exonuclease activity with a 5'-3' directionality (24). In the following sections, we analysed the mode of ribonuclease cleavage and substrate specificity of Pab-aCPSF1 as well as the function of the N-terminal KH domains and C-terminal protein dimer interface.

Pab-aCPSF1 has endo- and exoribonuclease activity

We investigated the enzymatic activity of recombinant Pab-aCPSF1 (untagged version) by performing assays in enzyme excess using the well-characterized 64-nt sR47 RNA substrate, corresponding to a *P. abyssi* box C/D guide RNA (24,34) (Figure 4). Incubation of Pab-aCPSF1 with 5'-end-labeled triphosphorylated RNA (5'p*pp RNA) yielded two major products of 21 and 59 nt in length (Figure 4A). These products correspond to cleavages after cytosines C21 and C59, which are located in the only two CA dinucleotides in the sR47 RNA substrate (Figure 4A). Assays with triphosphorylated 3'-end-labeled (5'ppp RNAp*Cp) and uniformly labeled (5'ppp RNA(U)*) substrates yielded two and four RNA products, respectively, corresponding to cleavages at the CA dinucleotides (Figure 4A). The addition of divalent ions to the reaction buffer did not stimulate activity, nor did the addition of EDTA inhibit the reaction. However, the activity of the enzyme was strongly inhibited by addition of 1,10-phenanthroline, a potent Zn²⁺ chelator (Supplementary Figure S3A). Furthermore, the substitution of conserved residues in the β -lactamase and β -CASP motifs (H261A and H594A in motifs 2 and B, respectively, Figure 1B) abolished ribonuclease activity (Supplementary Figure S3B). The recently published crystal structures of several aCPSF1 orthologs show that these histidines are involved in coordinating two zinc ions that are essential for catalysis (14,16,26). Altogether, these results reveal that Pab-aCPSF1 is a *bona fide* β -CASP protein and that the activity reported here is not due to a contaminating ribonuclease. We performed similar assays with sRkB, a 216 nt non-coding RNA recently identified in *P. abyssi* (35) (Figure 5A). sRkB was cleaved at nine positions: five corresponded to CA dinucleotides and the other four to GC, CC, AG or AC

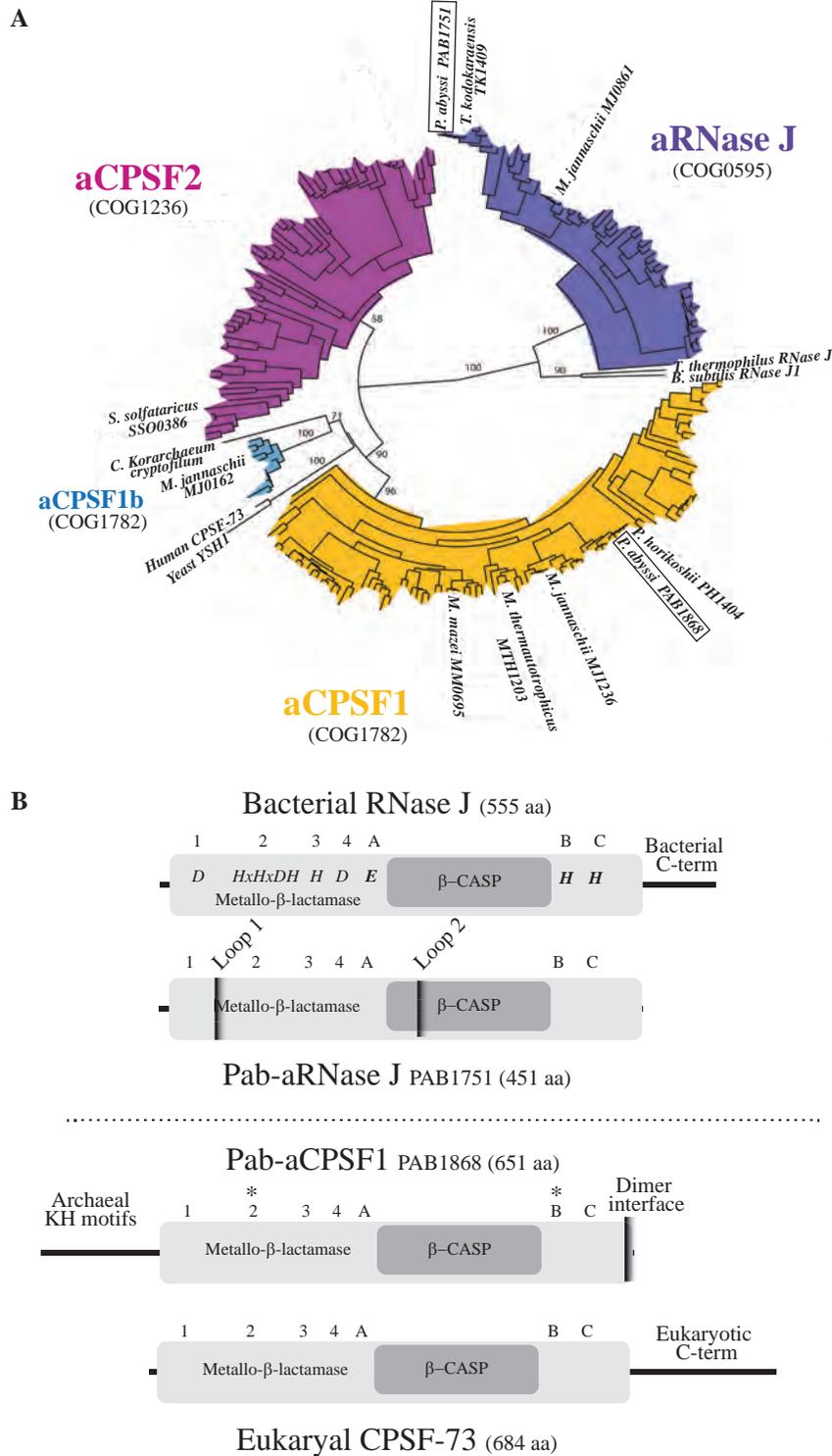


Figure 1. The archaeal metallo-β-lactamase β-CASP protein family. (A) Phylogenetic tree of the four β-CASP OGs. The β-CASP proteins with reported structure or/and activity were added as landmarks: yeast and human CPSF 73 kD (10), *M. thermotrophicus* MTH1203 (14), *P. horikoshii* PH1404 (16), *M. mazei* MM0695 (26), *B. subtilis* RNase J1 (48), *Thermus thermophilus* RNase J (21,49), *P. abyssi* PAB1751 (24), *M. jannaschii* MJ0861, MJ0162 and MJ1236 (15), *S. solfataricus* SSO0386 (25). The *P. abyssi* β-CASP members are framed. It should be mentioned that the landmarked *Candidatus Korarchaeum cryptofilum* β-CASP member is not considered as a CPSF1b fellow in view of its sequence. (B) Schematic representation of Pab-aRNase J (PAB1751) and Pab-aCPSF1 (PAB1868), homologs of bacterial RNase J (Top) and eukaryal CPSF-73 (Bottom), respectively. The β-CASP and metallo-β-lactamase domains are highlighted in dark and light grey, respectively. The three β-CASP motifs (A-C), four metallo-β-lactamase motifs (1–4) and N- and C-terminus are indicated with respective features. Pab-aRNase J loop 1 and 2 insertions are indicated (24). The asterisks (*) indicate the position of conserved amino acids that have been mutated in this study, H261A and H594A.

Table 1. Reported archaeal β -CASP proteins with their 5'-3' exo- and/or endoribonucleolytic activity

β -CASP OG	Protein	Reference	Endo-	5'-3'Exo
aCPSF1	PAB1868	This work	+	+
	MJ1236*	Levy <i>et al.</i> (15)	+	-
aCPSF1b	MJ0162*	Levy <i>et al.</i> (15)	-	+
aCPSF2	Sso0386*	Hasenohrl <i>et al.</i> (25)	-	+
aRNase J	PAB1751	Clouet-d'Orval <i>et al.</i> (24)	-	+
	TK 1409	Clouet-d'Orval <i>et al.</i> (24)	-	+
	MJ0861	Levy <i>et al.</i> (15)	-	+

The proteins marked by asterisks were misidentified as aRNase J homologs.

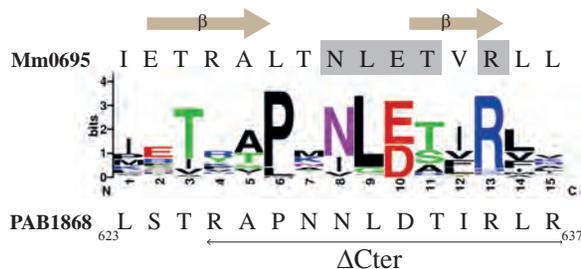


Figure 2. Sequence conservation of the last 16 residues of the 110 members of aCPSF1 is shown using Weblogo representation (<http://weblogo.berkeley.edu>). The amino acid C-terminal sequence of *M. mazei* Mm0695 and of *P. abyssi* PAB1868 (Pab-aCPSF1) are specified on the top and bottom, respectively. The residues establishing interacting hydrogen bonds in the protein-protein interface of the dimeric structure of Mm0695 (26) are highlighted in grey and the β -sheets formed by these residues are indicated. The last 12 residues that were deleted in aCPSF1 Δ Cter variant are indicated by a horizontal arrow under the *P. abyssi* sequence.

dinucleotides (Figure 5A). Two CA dinucleotides located in the highly stable P4 helix of sRkB were not cleaved suggesting a preference for single-stranded CA dinucleotides. This conclusion is supported by results with the sR47MutCG and sR47MutU21, which are not cleaved at position 21 (Figure 5B). C21 and A22 are embedded in an extended RNA helix in the sR47MutCG variant and C21 is replaced by a U in the sR47MutU21 variant (Figure 5B, higher panel). Altogether, these results show that Pab-aCPSF1 has endoribonuclease activity with a preference for cleavage at single-stranded CA dinucleotides.

To test whether the phosphorylation state of the 5'-end of the RNA substrate affects Pab-aCPSF1 activity, we performed assays with 5' monophosphorylated sR47 (Figure 4B). The degradation of 5'p RNA contrasts markedly with that of 5'ppp RNA owing to the production of GMP or UMP (Figure 4B). The 5'p*RNA generated radiolabeled GMP, which corresponds to the 5' terminal base in the sR47 substrate, whereas the uniformly labeled 5'p RNA(U)* generated radiolabeled UMP (Figure 4B). Comparable results were observed with the sRkB RNA substrate (Figure 5A). The 3'-end-labeled 5'p RNA yielded radiolabeled p*Cp, but not the 3'-end-labeled 5'ppp RNA. These results strongly suggest

that Pab-aCPSF1 has a 5' monophosphate-dependent 5'-3' exoribonucleolytic activity. This dependence is strict because neither 5'ppp (Figure 4A and Figure 5A) nor 5'hydroxyl (Figure 4C) transcripts can be degraded exonucleolytically. However, it should be noted that the distal products of endonucleolytic cleavage of the 5'ppp and 5'hydroxyl substrates do not appear to be degraded by the exoribonucleolytic activity as evidenced by the absence of UMP production. This suggests that products of endonucleolytic cleavage are somehow protected from exoribonucleolytic digestion. Interestingly, the exoribonucleolytic activity appears to be slowed or impeded by RNA secondary structure (compare the production of UMP* with the sR47, sR47MutCG and sR47 Mut21U substrates) (Figure 5B, lower panel). Finally, like other exoribonucleases from the β -CASP family [see ref. (24)], Pab-aCPSF1 can partially degrade 5'-end-labeled DNA oligonucleotides to mononucleotides (Supplementary Figure S3C). In conclusion, Pab-aCPSF1 has a dual activity: an endoribonuclease activity that preferentially cleaves at single-stranded CA dinucleotides, and exoribonuclease activity that is restricted to 5'-monophosphorylated RNA substrates.

Pab-aCPSF1 N-terminal and C-terminal extensions are involved in RNA binding and protein dimerization

To investigate further the properties of Pab-aCPSF1, we produced a variant deleted for the last 12 residues of Pab-aCPSF1 (Pab-aCPSF1 Δ Cter) (Figure 2). In the crystal structures of *M. mazei* (Mm) and *M. thermautotrophicus* aCPSF1 (14,26), these residues form a network of hydrogen bonding interactions at the interface of the dimeric holoenzyme. Furthermore, the interacting residues correspond to a sequence motif that is conserved in the aCPSF1 family (Figure 2). Gel filtration shows that the Δ Cter variant is mostly monomeric, whereas the wild type protein is dimeric (Figure 6A, left and middle panels), thus validating the role of the C-terminus in dimerization. Note that the Pab- Δ Cter recombinant protein is highly sensitive to proteolysis in the region linking the N-terminal KH domain and the core β -CASP metallo- β -lactamase domains (Supplementary Figure S4). We assayed the activity of the Δ Cter variant using the 5'p*pp RNA, 5'ppp RNA (U)* and 5'p RNA (U)* substrates. Our data show that the exonucleolytic activity of Pab-aCPSF1 Δ Cter is impaired as evidenced by the absence of UMP production with the 5'p RNA(U)* substrate (Figure 6B, left panel). Although endonucleolytic cleavage of 5'ppp RNA (U)* appears to be weak, cleavages at C21 and C59 are clearly detected with the 5'p*pp RNA and 5'p RNA(U)*.

To test the importance of the Pab-aCPSF1 N-terminus containing two KH domains (Figure 1), we produced a Pab-aCPSF1 Δ KH variant missing the first 179 residues. Gel filtration shows that Pab-aCPSF1 Δ KH is dimeric (Figure 6A, right panel). Thus, the N-terminal extension is not involved in dimerization. We assayed the activity of Pab-aCPSF1 Δ KH using the sR47 substrate (Figure 6B). The pattern of digestion is comparable with wild type (Figure 4A) except that the 59 nt RNA corresponding to

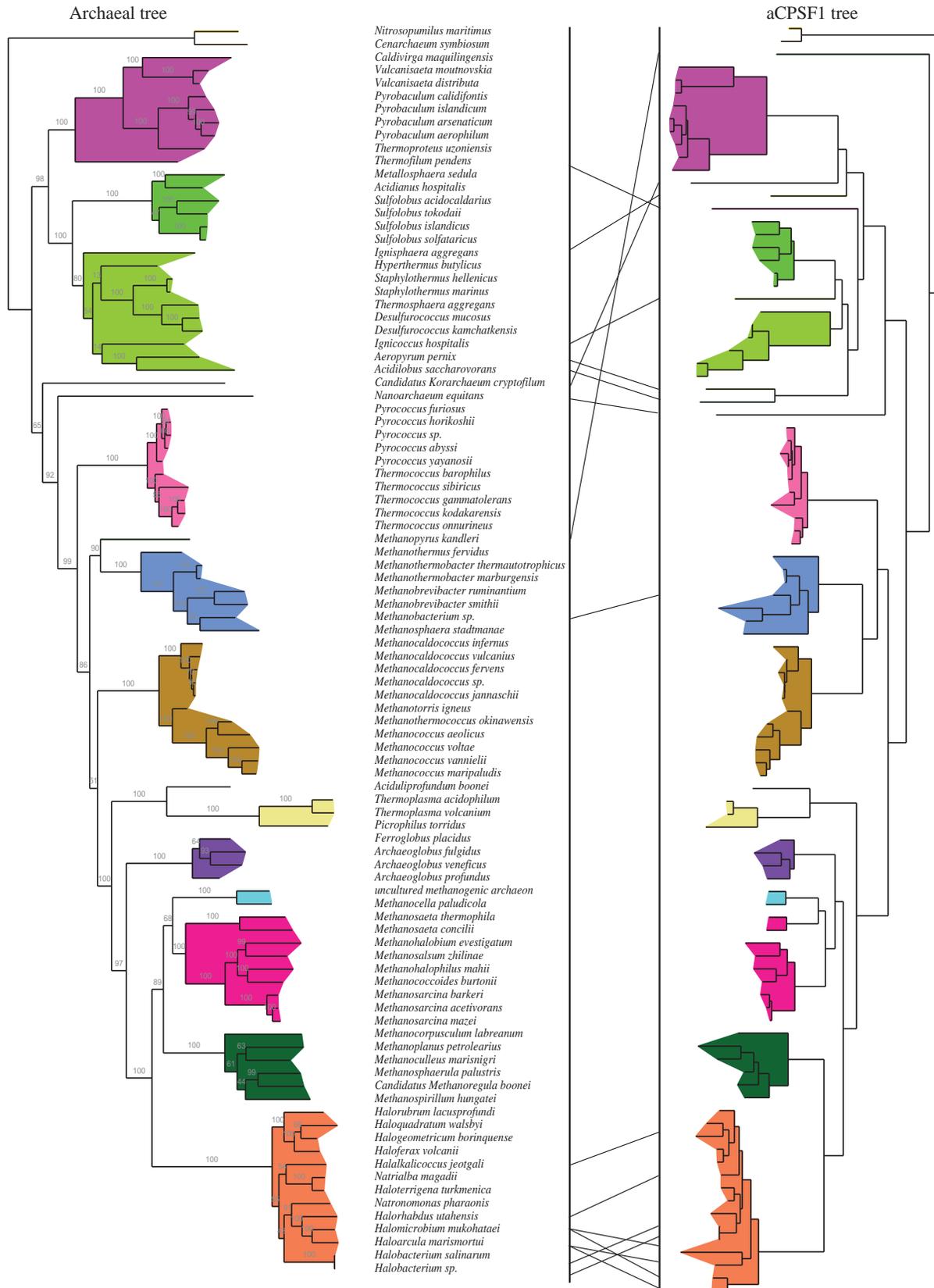


Figure 3. Congruence of the archaeal and aCPSF1 phylogenetic trees. The archaeal tree was constructed from a concatenated sequence of 53 ribosomal proteins. The archaeal aCPSF1 tree was constructed from sequences that were available at the time of analysis. Individual archaeal clades are colored to facilitate the comparison between the two trees.

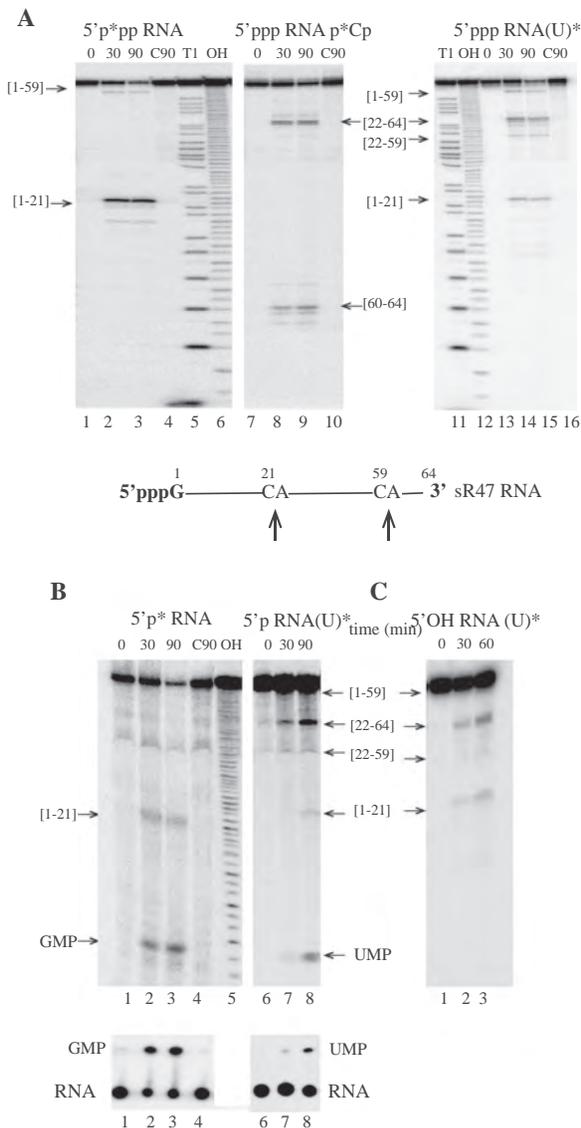


Figure 4. *In vitro* activity of Pab-aCPSF1. The products of the reaction were analysed on a 10% PAGE and in parallel on TLC (B, Bottom). (A) Kinetic analysis of RNA cleavage of 5'-end-triphosphate sR47 (<1 nM) by an excess of Pab-aCPSF1 (6 μM) at 65°C for the indicated times (0, 30 and 90 min). 5'-end-labeled sR47 RNA carrying a 5' γ -[32 P] (5'p*pp RNA), lanes 1–4; 3'-end-labeled with [32 P]Cp (5'ppp RNAp*Cp), lanes 7–10; uniformly labeled with α -[32 P]UTP (5ppp RNA(U)*), lanes 13 to 16. Substrates incubated 90 min without protein (C90) are shown in lanes 4, 10 and 16; RNase T1 ladder (T1) in lanes 5 and 11; hydroxyl ladders (OH) in lanes 6 and 12. The cleavage products are indicated by arrows and their position is indicated in the sR47 schematic drawing. The boundaries of the cleavage products are given in brackets. (B) Kinetic analysis of RNA cleavage of 5'-end-monophosphate sR47 by Pab-aCPSF1 (6 μM) at 65°C for the indicated time. 5'-end-labeled sR47 RNA carrying a 5' α -[32 P] (5'p* RNA) is shown in lanes 1–4; uniformly-labeled sR47 (5'p RNA(U)*) in lanes 6–8 and (C90) and (OH) in lanes 4 and 5, respectively. (C) Kinetic analysis of RNA cleavage of 5'-end hydroxyl uniformly labeled sR47 (α -[32 P] UTP, 5'OH RNA(U)*) by Pab-aCPSF1 (7 μM) at 65°C at the indicated times (lanes 1–3).

a cleavage at C59 was not produced. We conclude that the N-terminal KH domains are not necessary for catalytic activity, but are most likely involved in the recognition of certain specific sites. Because KH domains are

predicted to bind nucleic acids (14,16), we examined the capacity of the Δ KH variant to bind sR47 RNA by electrophoretic mobility shift assay (EMSA) (Figure 6C). On incubation with increasing protein concentration, three major distinct RNP complexes were detected with wild type Pab-aCPSF1. sR47 was fully shifted at a protein concentration of about 1 μM. 5'-triphosphate and 5'-monophosphate RNA bound with similar affinity, suggesting that the nature of the 5'-end is not important for binding. As well, we analysed the RNP patterns of sR47MutCG and sR47Mut21U substrates, which are invalid for endonucleolytic cleavage at position 21. Preliminary data showed similar overall binding affinities and high molecular weight RNP complexes as observed in Figure 6C (data not shown). However, the intensity of each RNP band was somehow different from sR47 EMSA, which did not permit a clear conclusion. This opens the question of whether binding and endonucleolytic activity could be uncoupled in future studies. The affinity for the Δ KH variant is slightly lower and the higher molecular weight RNP complexes are less stable as evidenced by smearing in the gel (Figure 6C). We also analysed the Δ Cter variant by EMSA. This variant is severely impaired in its capacity to bind RNA, suggesting that dimerization of the holoenzyme is important for RNA binding (Figure 6C). Altogether, these results show that the dimerization of Pab-aCPSF1 is important for exoribonuclease and RNA binding activity, whereas the KH domains participate in endoribonucleolytic cleavage at certain sites and are important for the stability of high molecular weight RNP complexes.

DISCUSSION

In this study, we systematically identified β -CASP proteins in Archaea and classified them according to sequence similarities to determine their phylogenetic relationships (Figure 1 and Supplementary Figure S1) and their taxonomic distribution (Supplementary Figure S2 and Figure 7). Among the seven archaeal β -CASP OGs that we identified: one is related to the bacterial RNase J (aRNase J), three are related to the eukaryal CPSF-73 (aCPSF1, aCPSF1b and aCPSF2). aRNase J, which is distributed exclusively in the Euryarchaeota, includes three members known to have 5'-end-dependent exonucleolytic activity (15,24,25) (Table 1). The aCPSF-like proteins are clearly divided into three clusters: aCPSF1 includes extremely well conserved members in all Archaea whose genome has been sequenced (this work); aCPSF2 groups more divergent members that are widespread in the Archaea and includes some that were previously misidentified as RNase J orthologs (15,25); aCPSF1b members are only present in the Methanococcales. aCPSF1b and aCPSF1 are closely related, but aCPSF1b lacks the N-terminal extension containing two KH domains. In summary, our phylogenetic analysis rectifies the misidentification of certain archaeal β -CASP proteins as aRNase J homologs (15,25) and clarifies their evolutionary origin.

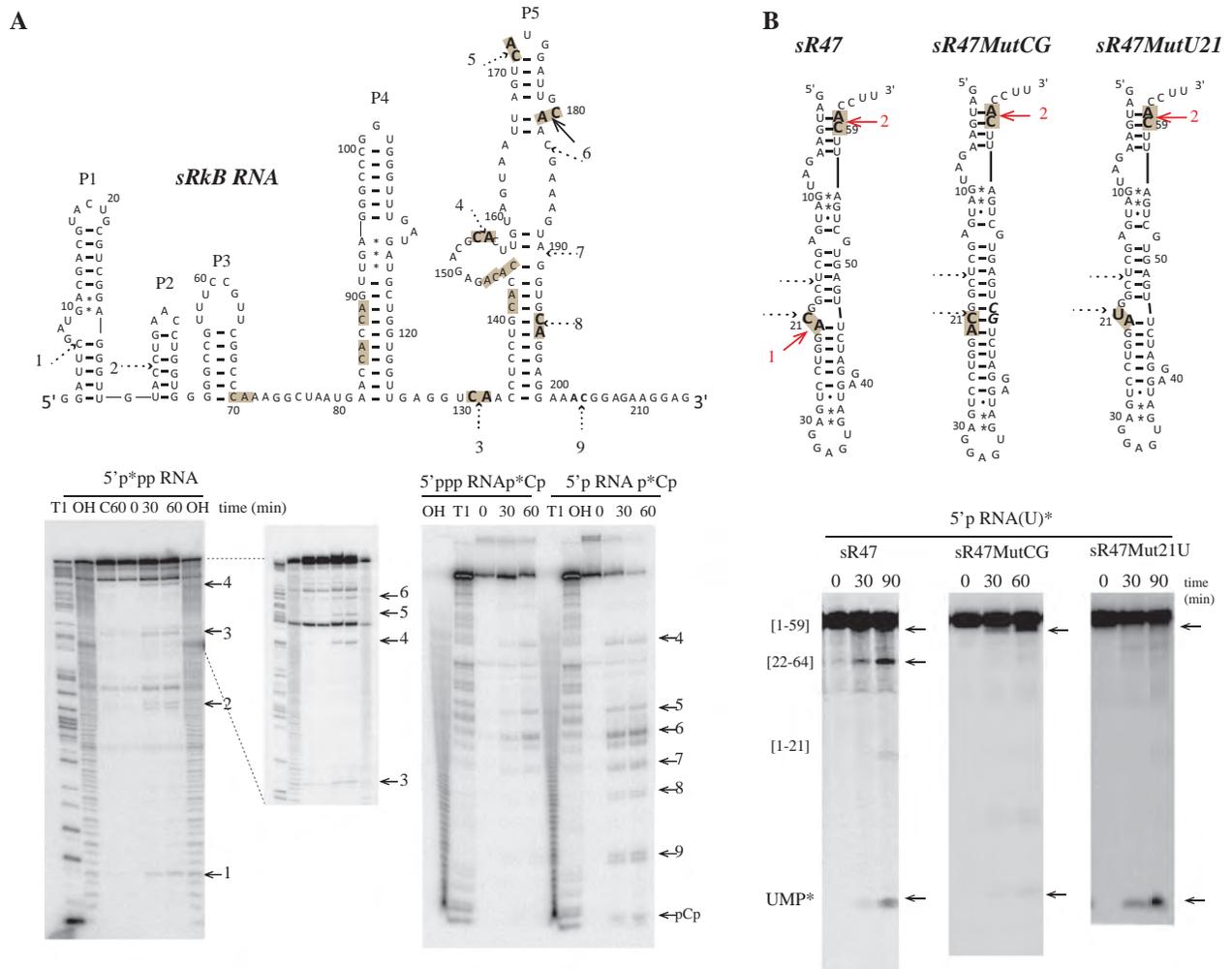


Figure 5. *In vitro* activity of Pab-aCPSF1 on sRkB and sR47 variant RNA substrates. **(A)** Cleavage sites on the sRkB map are indicated by solid and dotted arrows for strong and weak, respectively. The cleavage sites are numbered from 1 to 9 and the 11 CA dinucleotides of sRkB RNA are highlighted in grey (Top). The kinetic analysis of cleavage of the 5'p*pp RNA, 5'ppp RNAp*Cp and 5'p RNAp*Cp sRkB substrates at 65°C is shown (Bottom). **(B)** Cleavage sites induced by Pab-aCPSF1 on the sR47, sR47MutCG and sR47Mut21U maps (Top). The CA dinucleotides are highlighted in grey and the cleavage sites are indicated by red arrows. The sR47MutCG and the sR47Mut21U mutants possess an extra CG dinucleotide (bold italic characters, forming base pairs with G20 and C21) and a U at position 21, respectively. The CG insertion impedes cleavage at C21 and reduced the exonucleolytic degradation pathway as shown by the kinetic analyses of uniformly labeled sR47 and sR47MutCG (5'p RNA(U)*). The replacement of the U by a C impedes cleavage at position 21 but does not affect the exonucleolytic degradation of 5'p sR47Mut21U (U)* (Bottom). See Figure 2 legend for symbols.

Here, based on biochemical studies, we report that Pab-aCPSF1 has both an endonucleolytic and 5'-3' exonucleolytic activity. In addition, the endonucleolytic cleavage occurs in single-stranded RNA with a pronounced preference for CA dinucleotides. We reveal that the C-terminal homodimeric interface, initially identified in the crystal structures of *M. mazei* and *M. thermautotrophicus* members (14,26), is conserved amongst aCPSF1 homologs. Disruption of this interface in Pab-aCPSF1 results in monomeric enzyme that has endoribonuclease activity, but is deficient for exoribonuclease activity. In the same manner, protein interactions were shown to be required for full activity of eukaryotic CPSF-73, which forms a heterodimer with CPSF-100 (an inactive CPSF-73 paralog) (37). Deletion of the N-terminal KH domains of Pab-aCPSF1 abolishes endoribonuclease cleavage at certain specific sites and destabilizes high

molecular weight RNPs, without affecting exoribonucleolytic activity and the dimeric state of the enzyme. Given the general prevalence of KH domains in proteins associated with transcriptional and translational regulation (PNPase, the exosome, NusA and ribosomal proteins) (38), it seems likely that they will have an important role in aCPSF1 specificity. Mj-aCPSF1 has been recently reported to have endonucleolytic but not exonucleolytic activity (15) similar to the activity of the Pab-aCPSF1ΔCter variant studied here. Despite this apparent inconsistency, we believe that most aCPSF1 members are likely to have endo- and exoribonucleolytic activity because this is a property of both CPSF-73 and Pab-aCPSF1 (10-12).

All archaeal β-CASP proteins characterized to date, except for Mj-aCPSF1 (15), display a 5'-3' exoribonucleolytic activity that is dependent on the 5' phosphorylation

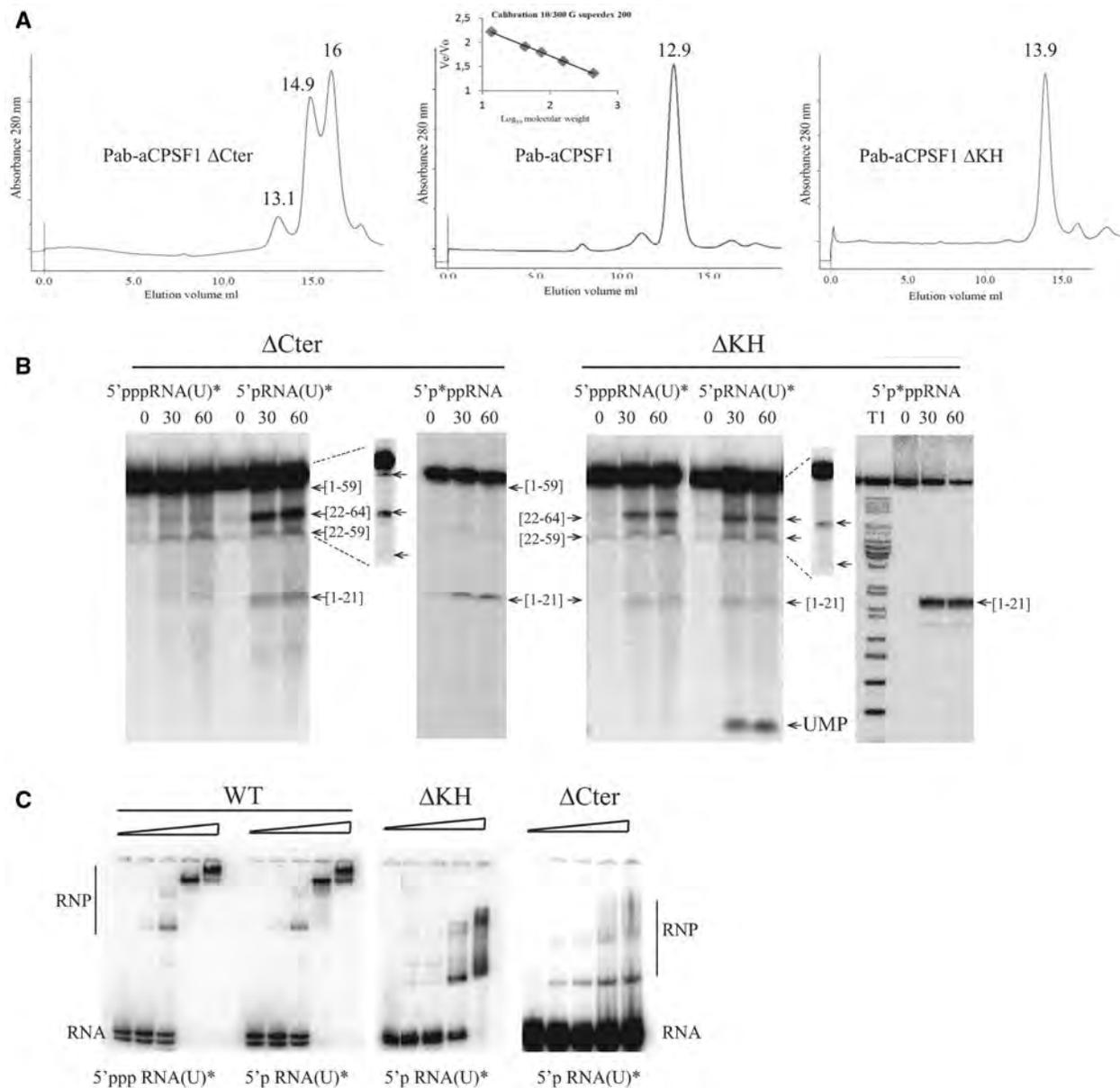


Figure 6. Pab-aCPSF1 Δ Cter (Δ Cter) and Pab-aCPSF1 Δ KH (Δ KH) variants. (A) Size exclusion chromatography of Pab-aCPSF1 and its variants. The proteins were separated on Superdex 200 and the apparent molecular masses defined using the inset calibration curve. Gel filtration yielded a minor peak at 13.1 ml and major peaks at 14.9 and 16 ml for Pab-aCPSF1 Δ Cter corresponding to dimeric protein (148 kDa), monomers (55 kDa) and a proteolysis products (30 kDa) (Supplementary Figure S4), respectively. It should be noted that the monomer elutes at higher volume than expected (55 kDa instead of 72 kDa) (Supplementary Figure S4) (left). Gel filtration yielded a peak at 12.8 ml for Pab-aCPSF1 corresponding to a dimeric protein (159 kDa) (middle). Gel filtration yield a peak at 13.9 ml for Pab-aCPSF1 Δ KH corresponding to dimeric protein (92 kDa) (right). (B) Kinetic analysis of RNA cleavage of 5'p*pp RNA, 5'ppp RNA(U)* and 5'p RNA(U)* substrates by protein variants (6 μ M) at 65°C for the indicated times. Symbols are as in Figure 2. The products of the reaction were analysed on a 10% PAGE. Long migrations are shown in the inset (dotted lines) for 5'p RNA(U)* products. (C) Gel retardation assay with 10 fmol of 5'p RNA(U)* or of 5'ppp RNA(U)* incubated with increasing concentrations (0, 0.3, 0.6, 1.25 and 2.5 μ M) of Pab-aCPSF1 (WT), Δ KH or Δ Cter variants. RNA-protein complexes (RNP) are indicated.

state of the substrate [(15,24,25), this work]. Previous biochemical work showed that the translation initiation factor α eIF2 binds to and protects RNA with 5'-triphosphorylated ends from degradation in the Crenarchaeon *S. solfataricus* (25,39). This observation suggests parallels to the principal mechanisms of 5'-3' RNA decay in Bacteria and Eukarya (1), in which Nudix hydrolases in Bacteria (40,41) and decapping enzymes in Eukarya (42) trigger mRNA degradation by

producing 5'-monophosphate ends. Nevertheless, comparable enzymes remain to be discovered in the Archaea. In conclusion, the nature of the substrate 5'-end (tri- versus mono-phosphorylated) emerges as a major determinant in the activity of β -CASP ribonucleases.

To highlight the prospective archaeal RNA degradation machinery, we have summarized the distribution of the archaeal β -CASP ribonucleases, together with the archaeal exosome and aRNase R, which have 3'-5'

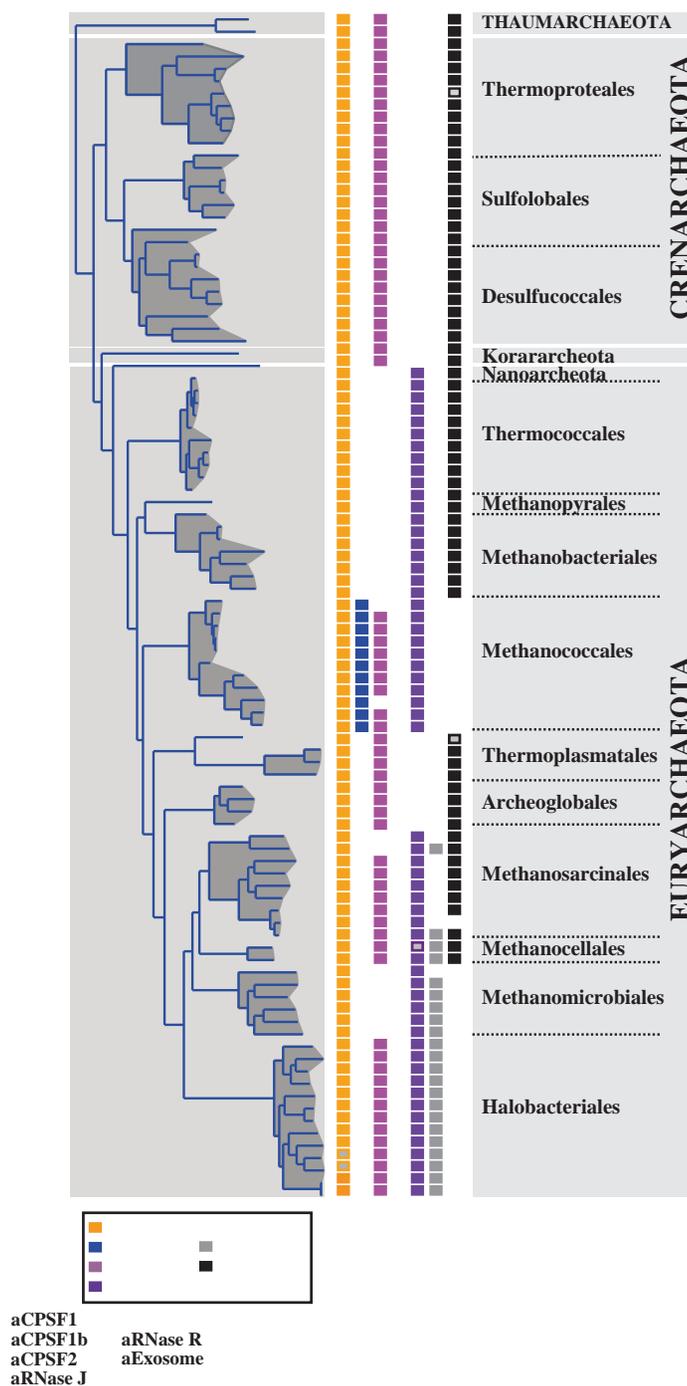


Figure 7. Distribution of the four archaeal β -CASP OGs, archaeal exosome and aRNase R on the archaeal lineage inferred from a set of concatenated conserved proteins (see ‘Materials and Methods’ section). Open rectangles indicate the existence of more than one member.

exonucleolytic activity (1,43,44) (Figure 7). It should be noted that the Crenarchaeota, Thaumarchaeota and Korarchaeota, which were recently described to be part of the ‘TACK’ superphylum and speculated to be at the origin of Eukarya (45), all have the same combination of RNA degrading enzymes (aCPSF1, aCPSF2 and archaeal exosome). In contrast, distribution in the Euryarchaeota is heterogeneous apart from the ubiquitous aCPSF1 homolog. As previously reported (13), the exosome is missing from Methanococcales, Methanomicrobiales and

Halobacteriales, illustrating the divergence in the Euryarchaeota. In the Halobacteriales, the emergence of an RNase R-like protein is believed to compensate this deficiency (46) (Figure 7). In the Methanococcales, the absence of the exosome correlates with the presence of aCPSF1b homologs, suggesting a possible functional link between the exosome and β -CASP proteins.

In conclusion, the enzymatic properties of aCPSF1 members are comparable with eukaryal CPSF-73 including 5'-end-dependent exoribonuclease activity and

an endoribonuclease activity with a preference for single-stranded CA dinucleotides. The strict conservation of these orthologs throughout the Archaea suggests a fundamental role in RNA metabolism. An analogy can be made with the eukaryal CPSF-73, which is a component of the machinery required for mRNA 3'-end maturation and termination of RNA polymerase II transcription (9,47). Our results suggest that a CPSF-like β -CASP protein was present in the last common ancestor of Archaea and Eukarya. We speculate that the highly conserved aCPSF1 might be an active component of an essential RNA-processing complex involved in mRNA degradation and/or 3'-end processing and transcription termination. By analogy to CPSF-73, which is part of a multi-component RNP, clues to the function of the archaeal homolog might come from future studies aimed at identifying archaeal complexes containing aCPSF1 homologs.

SUPPLEMENTARY DATA

Supplementary Data are available on NAR Online: Supplementary Table 1 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

Thanks to members of the Carpousis group for helpful discussions, G. Fichant, M. Bouvier, L. Minvielle-Sébastien and D. Flament for critical reading of the manuscript and F. Anglès and R. Simon for technical assistance. Author contributions: Y.Q., A.J.C. and B.C.O. designed research; D.K.P., D.R., P.L.G., Y.Q. and B.C.O. performed research and analysed data; Y.Q., A.J.C. and B.C.O. wrote the paper.

FUNDING

Centre National de la Recherche Scientifique (CNRS) with additional funding from the Agence Nationale de la Recherche (ANR) [BLAN08-1_329396]; from the Université de Toulouse (UPS) (AO1_2011). Funding for open access charge: Agence Nationale de la Recherche (ANR) [BLAN08-1_329396].

Conflict of interest statement. None declared.

REFERENCES

1. 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**, 275–317.
2. Werner, F. and Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.*, **9**, 85–98.
3. Proudfoot, N.J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev.*, **25**, 1770–1782.
4. Mandel, C.R., Bai, Y. and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol. Life Sci.*, **65**, 1099–1122.
5. Millevoi, S. and Vagner, S. (2010) Molecular mechanisms of eukaryotic pre-mRNA 3'-end processing regulation. *Nucleic Acids Res.*, **38**, 2757–2774.
6. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R. III, Frank, J. and Manley, J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**, 365–376.
7. Dominski, Z. (2010) The hunt for the 3'-endonuclease. *Wiley Interdiscip. Rev. RNA*, **1**, 325–340.
8. Chan, S., Choi, E.A. and Shi, Y. (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip. Rev. RNA*, **2**, 321–335.
9. Dominski, Z., Yang, X.C. and Marzluff, W.F. (2005) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell*, **123**, 37–48.
10. 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**, 953–956.
11. Ryan, K., Calvo, O. and Manley, J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, **10**, 565–573.
12. Yang, X.C., Sullivan, K.D., Marzluff, W.F. and Dominski, Z. (2009) Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol. Cell Biol.*, **29**, 31–42.
13. 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.
14. Silva, A.P., Chechik, M., Byrne, R.T., Waterman, D.G., Ng, C.L., Dodson, E.J., Koonin, E.V., Antson, A.A. and Smits, C. (2011) Structure and activity of a novel archaeal beta-CASP protein with N-terminal KH domains. *Structure*, **19**, 622–632.
15. Levy, S., Portnoy, V., Admon, J. and Schuster, G. (2011) Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biol.*, **8**.
16. Nishida, Y., Ishikawa, H., Baba, S., Nakagawa, N., Kuramitsu, S. and Masui, R. (2010) Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*. *Proteins*, **78**, 2395–2398.
17. Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol.*, **1**, 69–91.
18. Condon, C. and Gilet, L. (2011) *Nucleic Acids and Molecular Biology*, Vol. 26, Springer-verlag, Heidelberg, Berlin, pp. 245–267.
19. Callebaut, I., Moshous, D., Mornon, J.P. and de Villartay, J.P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res.*, **30**, 3592–3601.
20. Dominski, Z. (2007) Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.*, **42**, 67–93.
21. Li de la Sierra-Gallay, L., Zig, L., Jamalli, A. and Putzer, H. (2008) Structural insights into the dual activity of RNase J. *Nat. Struct. Mol. Biol.*, **15**, 206–212.
22. Mathy, N., Benard, 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**, 681–692.
23. Britton, R.A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W.C., Mathy, N., Tobin, C., Daou, R., Szyk, J. and Condon, C. (2007) Maturation of the 5'-end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol. Microbiol.*, **63**, 127–138.
24. 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.
25. Hasenohr, D., Konrat, R. and Blasi, 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.
26. Mir-Montazeri, B., Ammelburg, M., Forouzan, D., Lupas, A.N. and Hartmann, M.D. (2011) Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J. Struct. Biol.*, **173**, 191–195.
27. van Dongen, S. and Abreu-Goodger, C. (2012) Using MCL to extract clusters from networks. *Methods Mol. Biol.*, **804**, 281–295.

28. Brochier-Armanet,C., Boussau,B., Gribaldo,S. and Forterre,P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.*, **6**, 245–252.
29. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
30. Galtier,N., Gouy,M. and Gautier,C. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, **12**, 543–548.
31. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
32. Guindon,S. and Gascuel,O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.
33. Chevenet,F., Brun,C., Banuls,A.L., Jacq,B. and Christen,R. (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, **7**, 439.
34. Nolivos,S., Carpousis,A.J. and Clouet-d’Orval,B. (2005) 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.*, **33**, 6507–6514.
35. Phok,K., Moisan,A., Rinaldi,D., Brucato,N., Carpousis,A.J., Gaspin,C. and Clouet-d’Orval,B. (2011) Identification of CRISPR and riboswitch related RNAs among novel non-coding RNAs of the euryarchaeon *Pyrococcus abyssi*. *BMC Genomics*, **12**, 312.
36. Even,S., Pellegrini,O., Zig,L., Labas,V., Vinh,J., Brechemmier-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 Res.*, **33**, 2141–2152.
37. Dominski,Z., Yang,X.C., Purdy,M., Wagner,E.J. and Marzluff,W.F. (2005) A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol. Cell Biol.*, **25**, 1489–1500.
38. Valverde,R., Edwards,L. and Regan,L. (2008) Structure and function of KH domains. *FEBS J.*, **275**, 2712–2726.
39. Hasenohrl,D., Lombo,T., Kabardin,V., Londei,P. and Blasi,U. (2008) Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. USA*, **105**, 2146–2150.
40. Richards,J., Liu,Q., Pellegrini,O., Celesnik,H., Yao,S., Bechhofer,D.H., Condon,C. and Belasco,J.G. (2011) An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol. Cell*, **43**, 940–949.
41. Deana,A., Celesnik,H. and Belasco,J.G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature*, **451**, 355–358.
42. Song,M.G., Li,Y. and Kiledjian,M. (2010) Multiple mRNA decapping enzymes in mammalian cells. *Mol. Cell*, **40**, 423–432.
43. Evguenieva-Hackenberg,E., Walter,P., Hochleitner,E., Lottspeich,F. and Klug,G. (2003) An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep.*, **4**, 889–893.
44. 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.
45. Guy,L. and Ettema,T.J. (2011) The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol.*, **19**, 580–587.
46. 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.
47. Proudfoot,N. (2004) New perspectives on connecting messenger RNA 3'-end formation to transcription. *Curr. Opin. Cell Biol.*, **16**, 272–278.
48. Newman,J.A., Hewitt,L., Rodrigues,C., Solovyova,A., Harwood,C.R. and Lewis,R.J. (2011) Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure*, **19**, 1241–1251.
49. Dorleans,A., Li de la Sierra-Gallay,I., Piton,J., Zig,L., Gilet,L., Putzer,H. and Condon,C. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure*, **19**, 1252–1261.

**IDENTIFICATION ET CLASSIFICATION DES RIBONUCLEASES DE LA
FAMILLES DES β -CASP CHEZ LES ARCHEES ET CARACTERISATION
DE aCPSF1 DE *P. abyssi***

PUBLICATIONS 2

**Universal RNA-degrading enzymes in Archaea:
Prevalence, activities and functions of β -CASP
ribonucleases.**

Béatrice Clouet-d'Orval , Duy Khanh Phung, Petra S. Langendijk-
Genevaux, Yves Quentin.

Centre National de la Recherche Scientifique, LMGM and Université de Toulouse, Toulouse,
France

PUBLICATIONS 3 (Annexes)

**An overview of ribonuclease repertoire and RNA
processing in archaea**

Duy Khanh Phung, Marie Bouvier and Béatrice Clouet-d'Orval.

Centre National de la Recherche Scientifique, LMGM and Université de Toulouse, Toulouse,
France



Review

Universal RNA-degrading enzymes in *Archaea*: Prevalence, activities and functions of β -CASP ribonucleases



Béatrice Clouet-d'Orval ^{a, b, *}, Duy Khanh Phung ^{a, b}, Petra S. Langendijk-Genevaux ^{a, b}, Yves Quentin ^{a, b}

^a Université de Toulouse; UPS, 118 Route de Narbonne, F-31062 Toulouse, France

^b CNRS; LMGM; F-31062 Toulouse, France

ARTICLE INFO

Article history:

Received 28 February 2015

Accepted 28 May 2015

Available online 6 June 2015

Keywords:

Archaea
RNA processing
 β -CASP ribonucleases
aCPSF1
aCPSF2
aRNase J

ABSTRACT

β -CASP ribonucleases are widespread in all three domains of life. They catalyse both 5'-3' exoribonucleolytic RNA trimming and/or endoribonucleolytic RNA cleavage using a unique active site coordinated by two zinc ions. These fascinating enzymes have a key role in 3' end processing in Eukarya and in RNA decay and ribosomal RNA maturation in Bacteria. The recent recognition of β -CASP ribonucleases as major players in *Archaea* is an important contribution towards identifying RNA-degrading enzymes in the third domain of life. Three β -CASP orthologous groups, aCPSF1, aCPSF2, aCPSF1b, are closely related to the eukaryal CPSF73 termination factor and one, aRNase J, is ortholog of the bacterial RNase J. The endo- and 5'-3' exoribonucleolytic activities carried by archaeal β -CASP enzymes are strictly conserved throughout archaeal phylogeny suggesting essential roles in maturation and/or degradation of RNA. The recent progress in understanding the prevalence, activities and functions of archaeal β -CASP ribonucleases is the focus of this review. The current status of our understanding of RNA processing pathways in *Archaea* is covered in light of this new knowledge on β -CASP ribonucleases.

© 2015 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

1. Introduction

RNA processing and degradation pathways, involving ribonucleases and ancillary RNA-modifying enzymes constitute a crucial aspect in the regulation of gene expression in Bacteria and Eukarya (Fig. 1). They are a key factor in sensing and in providing rapid responses to changing cellular environments. The levels of all cellular RNAs are governed by their rate of synthesis and degradation. Their processing or turnover or both also play a central role by contributing to the regulation of the half-lives and maturation of specific messenger RNAs (mRNA) or non-coding RNAs (ncRNAs). Cells must therefore precisely control these RNA degradation and processing activities to prevent unwanted removal of functionally important molecules and to monitor precisely their gene expression at the post-transcriptional level.

Most of our knowledge of RNA processing and degradation pathways comes from Bacteria and Eukarya (Fig. 1). The panoply of known RNA degrading pathways has been extensively covered in recent reviews [1–4]. A large number of ribonucleases are involved in processing of primary transcripts to mature RNA, quality control of RNA, and regulation of gene expression. More than 30 ribonuclease families have been identified to accurately act during cleaving and trimming RNA. These include both endonucleases and 5'-3' and 3'-5' exoribonucleases [1,2]. In addition to ribonucleases, ancillary RNA-modifying enzymes such as RNA helicases, poly A polymerases and pyrophosphohydrolases are important in mediating RNA processing and/or turnover.

In general, RNA decay occurs both in the 5'-3' and 3'-5' directions. Therefore the nature of the 5' and 3' ends is critical for the fate of the RNA. For example, in eukaryotes, the steps involving removal of the 5' end 7-methylguanosine cap and of the poly(A) tail at the 3' end is considered rate-limiting for mRNA decay (Fig. 1, right panel) [5]. In the same manner, the bacterial decay pathway of 5' triphosphorylated RNA can be initiated by the removal of the 5' pyrophosphate by RppH enzymes (Fig. 1, left panel) [6].

In spite of our extensive understanding of RNA processing enzymes in *Eukarya* and *Bacteria*, very little is known about these

* Corresponding author. Centre National de la Recherche Scientifique, LMGM-Université de Toulouse, UPS, Bat IBCG, 118 route de Narbonne, F-31062 Toulouse Cedex 9. Tel.: +33 561335816.

E-mail address: Beatrice.Clouet-dOrval@ibcg.biotoul.fr (B. Clouet-d'Orval).

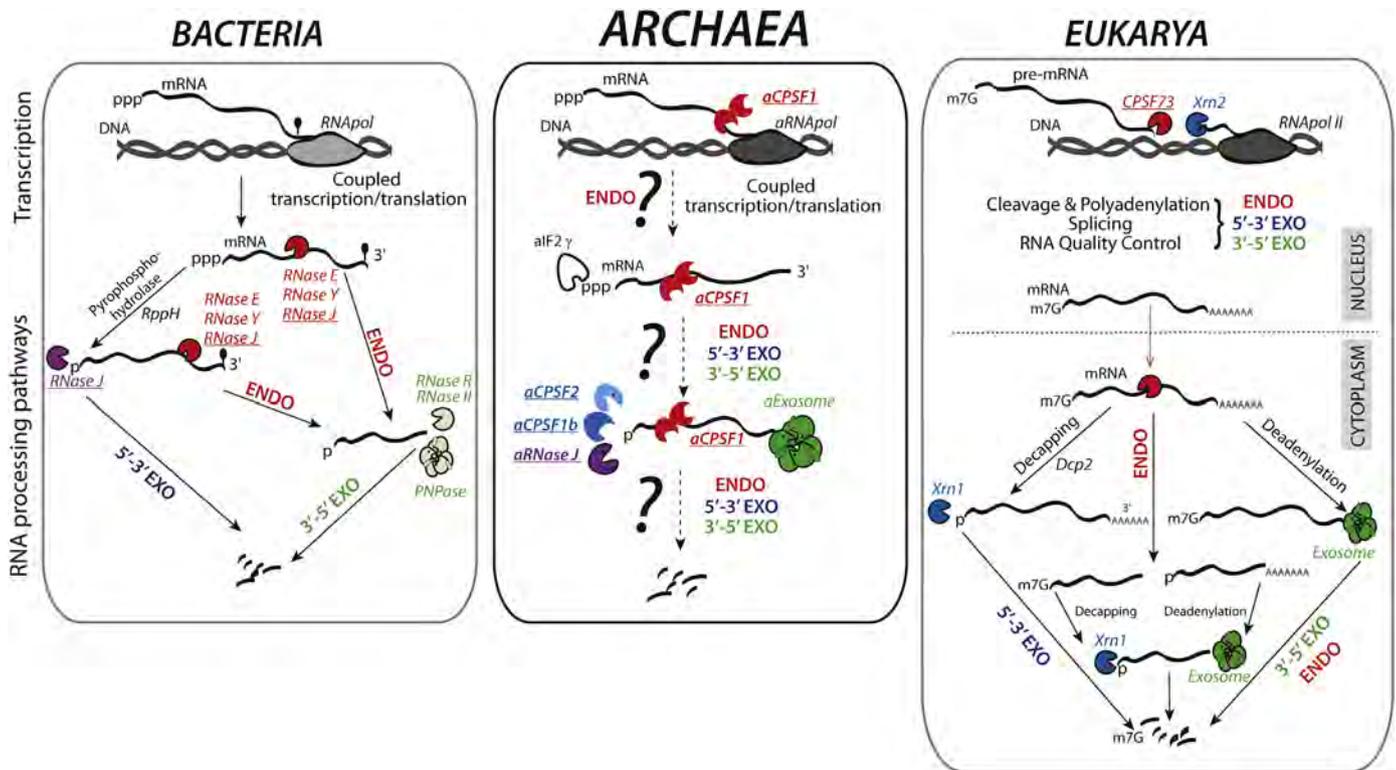


Fig. 1. General schematic overview of RNA processing pathways in the three domains of life. Enzyme names are in italics. Ribonucleases belonging to the β -CASP family are underlined. Endoribonucleolytic activities are in red, 5'-3' exoribonucleolytic in blue and purple, and 3'-5' exoribonucleolytic in green. Only a subset of bacterial (left panel) and eukaryal (right panel) enzymes are depicted in the scheme, note that not all are found in one cell. For more detailed information, please refer to [1,2,26,61]. In *Archaea* (central panel), very few activities have been reported and RNA processing events (?) wait to be described. The four major archaeal β -CASP orthologous groups, aCPSF1, aCPSF2, aCPSF1b and aRNase J which are discussed in this review are depicted. The a/eIF2(- γ) factor reported to protect 5'triphosphorylated RNA-ends in *Sulfolobales* is also indicated.

processes in the *Archaea* (Fig. 1). Even though archaeal RNAs are transcribed by a unique RNA polymerase similar to the eukaryotic RNA polymerase II, archaeal RNA structure is similar to that in *Bacteria*. They have triphosphorylated 5' ends as do bacterial RNAs but no long eukaryal-like poly(A) tails or other specific structures such as bacterial terminators [7]. Indeed, *in vitro* and *in vivo* evidence suggest that the translation initiation factor a/eIF2(- γ) binds to RNA 5'-triphosphorylated ends and protects transcripts from a 5' end-dependent degradation pathway in the crenarchaeota *Sulfolobus solfataricus* (Fig. 1, central panel) [8]. In this case, a/eIF2 (- γ) subunit exhibits an additional function with resemblance to the eukaryotic cap-complex.

As in the other domains of life, processing of tRNA in *Archaea* is performed by the ribonucleoprotein RNase P [9,10] and RNase Z [11,12]. The RNA-splicing endoribonuclease removes introns from archaeal pre-tRNA and participates in rRNA processing [13,14]. Furthermore, a eukaryotic-like exosome with a 3'-5' exoribonucleolytic activity is present in most *Archaea* with the exception of the *Halophiles* and *Methanococcales* [7,15–19]. An important breakthrough towards building basic knowledge of archaeal RNA-degrading enzymes came from the discovery of archaeal β -CASP family proteins. These enzymes which possess dual 5'-3' exo- and endoribonucleolytic activities [20–24] may account for the 5'-3' direction RNA decay and RNA maturation events. Recognition of the status of β -CASP ribonucleases as major enzymes in *Archaea* is certainly an important contribution towards the understanding of archaeal mRNA, regulatory RNA and stable RNA processing pathways. The recent progress in the understanding of the emergence, prevalence, activities and functions of archaeal β -CASP ribonucleases is the focus of this review.

2. The β -CASP ribonuclease family

β -CASP ribonucleases which form a family within the metallo β -lactamase superfamily are widespread in all three domains of life. It emerged over the last decade that β -CASP CPSF73 factor and RNase J ribonuclease are key players in RNA metabolism in *Eukarya* and *Bacteria*, respectively (Fig. 1, right and left panels) [25,26].

In *Eukarya*, the ubiquitous β -CASP CPSF73 factor is part of a large multi-protein complex that is involved in the maturation of the 3' end of all RNA Polymerase II transcripts and in the polyadenylation of the mRNAs [26–28]. In *Bacteria*, the β -CASP RNase J proteins are present in more than half of reported species which includes important pathogens. Bacterial RNase J possesses a dual ribonucleolytic activity and fulfils key functions in bacterial ribosomal RNA maturation and mRNA decay (see Refs. in [26,29]). The remarkable properties and physiological functions of β -CASP enzymes, mostly in *Bacteria* and *Eukarya*, have recently been discussed and reviewed [26,29].

Extensive structural and biochemical studies indicate that enzymes of the β -CASP family contain a bipartite metal dependant catalytic core (for more detailed information refer to [26,29]). Briefly, the β -CASP domain is inserted into the canonical metallo β -lactamase domain (Fig. 2A). The catalytic site pocket which contains two zinc ions is formed at the interface of the two domains. The tightly bound ions are coordinated by conserved histidine and acid aspartic residues which include the signature motifs A (Asp or Glu), B (His) and C (His) of the β -CASP domain (see below). It has been proposed a mechanism for two metal-ion catalysis through a nucleophilic attack on the phosphate group of phosphodiester

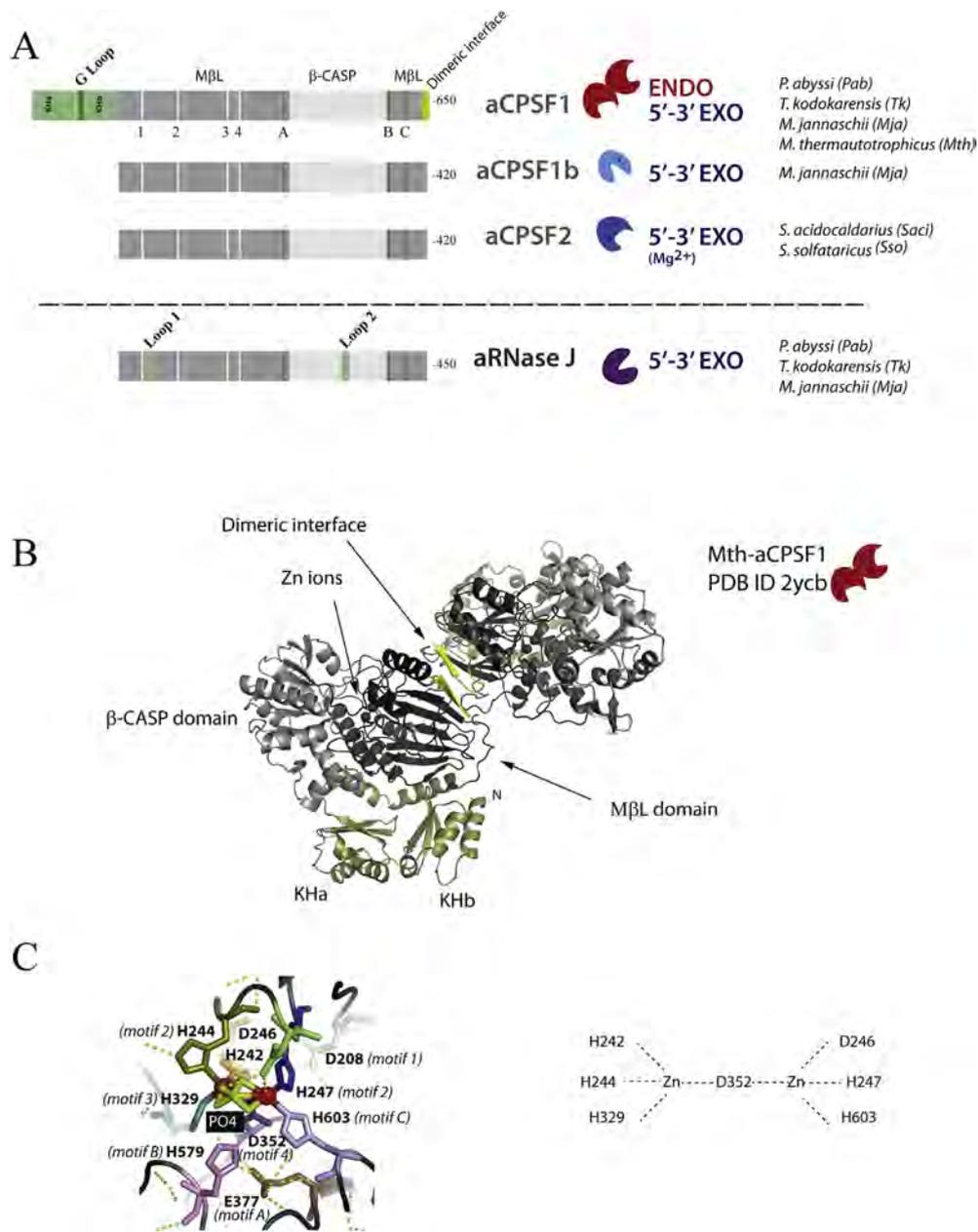


Fig. 2. Archaeal β -CASP ribonuclease features. A- Constituent domains of the four major archaeal β -CASP ribonuclease orthologous groups [24] are depicted as coloured rectangles: grey, β -metallo- β -lactamase (M β L); light grey, β -CASP; green, KH domains of aCPSF1 and specific Loop1 and Loop2 of aRNase J; yellow, dimeric interface. The β -CASP family is characterized by three short conserved motifs characterized by one amino acid: A (Asp or Glu), B (His) and C (His). Subsequence columns indicate the *in vitro* ribonucleolytic activities together with species in which they have been determined. All aCPSF2 except one member are commonly restricted to the β -CASP and metallo- β -lactamase core domains (composed of an average of 420 aa) with no additional N- or C-terminal extensions. Sso-aCPSF2 with a 40 aa N-terminal extension is an exception. Mja-aCPSF1 was shown to display only an endonucleolytic activity [22] and Mth aCPSF1 a nuclease activity [23]. B- A ribbon model of the overall structure of Mth-aCPSF1 dimer (PDB ID 2ycb). Zinc ions are in red. The β -metallo- β -lactamase (M β L) is in grey, β -CASP in light grey, and KH domains in green. C- The detailed structure of the catalytic centre of Mth-aCPSF1 (PDB 2YCB) is shown. Two zinc atoms (in red) are coordinated in an octahedral geometry by the bound phosphate (PO4) and the conserved residues H242, H244, D247, H247 of motif 2, H329 of motif 3, D352 of motif 4, H579 of motif B, and H603 of motif C. Hydrogen bonds are shown as dashed lines.

bonds [30,31]. A hallmark of the β -CASP ribonucleases is their dual endo- and exonucleolytic activities, which are catalysed by the same catalytic site. Recent studies tried to rationalize the 5'-3' exo- and endonucleolytic dual-enzymatic modes of these enzymes by looking at several bacterial RNase J crystallized in different conditions. Remarkably a number of structural changes and movements of the two enzyme domains were observed [32,33]. However these structures do not provide a clear clue how the RNA is positioned to lie across the catalytic site and how the switch from exo-

endonucleolytic activity is promoted [34]. In addition, this switch could involve the binding of another factor or the association in a multi-protein complex.

3. Archaeal β -CASP ribonuclease orthologous groups

Early studies of bacterial RNase J, mentioned that members of the β -CASP family were present in *Euryarchaea* [35]. Subsequent in-depth phylogenomic approach revealed many related proteins

forming a vast family with members encoded by at least two genes per archaeal genome [24]. This familial relationship is based on conserved protein motifs of the β -CASP β -lactamase core domain [24,25]. Nonetheless, based on graph partitioning and phylogenetic relationship (110 annotated archaeal genomes including: 2 thaumarchaea, 25 creanarchaea, 83 euryarchaea) four archaeal β -CASP orthologous groups (OGs) have been defined: aCPSF1, aCPSF1b, aCPSF2 and aRNase J (Fig. 2A) [24]. The topology of the β -CASP OG phylogenetic tree shows a clear dichotomy with aCPSF1, aCPSF1b, aCPSF2 OGs closely related to eukaryal CPSF73 factor and aRNase J OG to bacterial RNase J (Fig. 2A) [20,26]. This suggests that archaeal and bacterial RNase J and, aCPSF1 and eukaryal CPSF73 are orthologous, respectively.

In addition, three other clusters (aC β x, aC β y & aC β z) composed of highly divergent β -CASP proteins were defined [24]. Together, presently sequenced archaeal genomes encode two to five β -CASP proteins. Phylogenomic studies have highlighted the importance of each β -CASP enzyme group within each archaeal clade [24,26]. Note that the initial designation of *S. solfataricus* and *Methanocaldococcus jannaschii* β -CASP members was based on RNase J proteins [21,22], leading to some misclassifications in the latter group. This has now been revised [24]. Features and phylogenetic distribution of the four main groups are briefly described below. The aRNase J, aCPSF2 and aCPSF1b group members are restricted to the β -CASP β -lactamase core domain (Fig. 2A). Notably, aRNase J members are limited to *Euryarchaeota*, aCPSF1b members are restricted to the *Methanococcale* clade and aCPSF2 members are strictly conserved in *Crenarcheota* and widely spread and more divergent in *Euryarchaeota*.

Outstandingly aCPSF1 members are found in all archaea examined and are distinguished by their N-terminal domain being composed of a tandem repeat of two type-II KH domains predicted to bind nucleic acids and a C-terminal dimeric interface (Fig. 2A) [36,37]. The crystal structures of three aCPSF1 members (*Pyrococcus horikoshii*, *Methanosarcina mazei* and *Methanothermobacter thermoautotrophicus*) show an organization with a tripartite architecture consisting of the β -CASP and β -lactamase domains folded apart of the both KHa and KHb motifs (Fig. 2B) [23,38,39]. The common β -lactamase and β -CASP domains are formed by two opposing $\beta\beta\beta\alpha\beta\alpha\beta$ motifs forming seven-stranded β -sheets sandwiched on either side by three α -helices and five stranded parallel β -sheets that is surrounded by five α -helices, respectively (Fig. 2B). Note that the three-dimensional constellation of active-site residues is nearly identical across the family [23,30–33,38–40]. Indeed two zinc atoms are coordinated in an octahedral geometry by the bound phosphate and the conserved residues of motif 1, 2, 3, B and C (Fig. 2C). An N-terminal domain composed of two KH domains (KHa (or N–KH) and KHb (or C–KH)) is packed to the typical central β -metallo-lactamase and the β -CASP domains by a long helical linker [23,38,39]. In addition, two aCPSF1 crystal structures were found to contain two molecules per asymmetric unit suggesting that aCPSF1 is a homodimer. The protein dimeric interface consists mostly of a short C-terminal segment of 12 amino acids [23,24,39] (Fig. 2A and B). In view of their wide phylogenetic distribution, archaeal β -CASP enzymes are likely to perform essential functions in archaeal metabolism that remain to be identified.

4. Ribonucleolytic activities of archaeal β -CASP orthologous group enzymes

Several biochemical studies have established *in vitro* enzymatic activities of archaeal β -CASP enzymes using recombinant proteins expressed in *Escherichia coli* [20–22,41]. Specific protocols were set up to reliably follow *in vitro* RNA cleavage and degradation of archaeal hyperthermophilic enzymes [42]. We present below a

brief overview of *in vitro* ribonucleolytic activities of each of the archaeal β -CASP orthologous group enzymes: aRNase J, aCPSF2, aCPSF1b and aCPSF1 (Fig. 2A).

Archaeal β -CASP enzymes with only a 5'-3' exoribonucleolytic activity are restricted to those which carry the β -CASP β -lactamase core domain alone i.e. aRNase J, aCPSF2 and aCPSF1b group members (Fig. 2A). Indeed archaeal 5'-3' exoribonucleolytic activity was first identified in studies of aRNase J recombinant proteins from *Pyrococcus abyssi* (Pab) and *Thermococcus kodakarensis* (Tk) [20]. The *M. jannaschii* (Mj) aRNase J (denoted mjRNase J3 in Ref. [22]) also possesses similar 5'-3' exonuclease trimming activity [22]. These enzymes are highly processive and have no detectable endonucleolytic activity under the conditions tested [20,22]. In these studies, it has been shown that the residues of the highly conserved motifs that coordinated the two zinc ions are critical for enzyme activity (Fig. 2A and C). In addition, the resistance of the 5'-3' processing reaction to EDTA is explained by the extremely high affinity of the two ions and the activity is inhibited only by addition of zinc-specific chelators [20]. As well, the two small highly conserved euryarchaeal RNase J domains denoted Loop1 and Loop2 (Fig. 2A), have been revealed to be important since their deletion in Pab-RNase J significantly reduced *in vitro* ribonucleolytic activity [20].

Pab-aRNase J, Tk-aRNase J and Mj-aRNase J show a greatly reduced activity in degrading RNA substrates with 5' triphosphate ends *in vitro* implying that the subset of cellular RNA with 5' triphosphate ends would also be refractory to degradation by these exonucleases *in vivo*. Similarly, aCPSF2 proteins from *S. solfataricus* (Sso-aCPSF2) and *Sulfolobus acidocaldarius* (Saci-aCPSF2), initially identified as archaeal RNase J orthologs, both have 5'-3' exonucleolytic activities that are affected by the phosphorylation state of the 5' end of transcript [21,41]. All these enzymes show some residual nucleolytic activity *in vitro*. It is possible that this is due to spontaneous loss of the 5' triphosphate or endogenous cleavage of the RNA leaving a 5' monophosphate. However, in the case of Sso-aCPSF2 and Saci-aCPSF2, this residual activity is likely to involve 5'-3' exonuclease degradation since it is eliminated by binding of the γ -subunit of the translational initiation factor a/eIF2 at the 5' terminus of RNA [21,41] previously shown to impede 5'-3' RNA decay *in vivo* in *S. solfataricus* [8]. In spite of the similar catalytic activities of this group of enzymes, Sso- and Saci-aCPSF2 are dependent on magnesium ions whereas no other known β -CASP enzyme requires this divalent metal cation [21,41]. This unusual magnesium requirement might be a general property of all aCPSF2 group members. In this light, it will be interesting to determine whether the aCPSF2 homolog from *M. jannaschii* (initially denoted mjRNase J [22]) also possesses this requirement. Further understanding of the enzymatic and functional properties of aRNase J and aCPSF2, in particular the role of magnesium, must await structural studies. Finally, aCPSF1b recombinant protein from *M. jannaschii* (initially named mjRNase J1 [22]) was shown to harbor a 5' end dependent 5'-3' exonucleolytic activity, but no endonucleolytic activity. This activity contrasts with the dual activity of the closely phylogenetically related aCPSF1 group members [24] (see below). In view of the taxonomic distribution of each β -CASP orthologous group, all archaeal genomes should therefore encode at least one β -CASP enzyme with a 5' end dependent 5'-3' exoribonucleolytic activity (Fig. 3).

In light of the astounding conservation of aCPSF1 group throughout the archaeal phylogeny, enzymatic activities of aCPSF1 β -CASP enzymes should be critical in archaeal cells (Fig. 3). To date only aCPSF1 recombinant proteins from *P. abyssi* (Pab-aCPSF1) and *T. kodakarensis* (Tk-aCPSF1) have been shown to harbor dual endo- and 5'-3' exoribonucleolytic activities [24]. Further investigations revealed that Pab-aCPSF1 preferentially cleaves RNA at a single-stranded CA dinucleotide and possesses

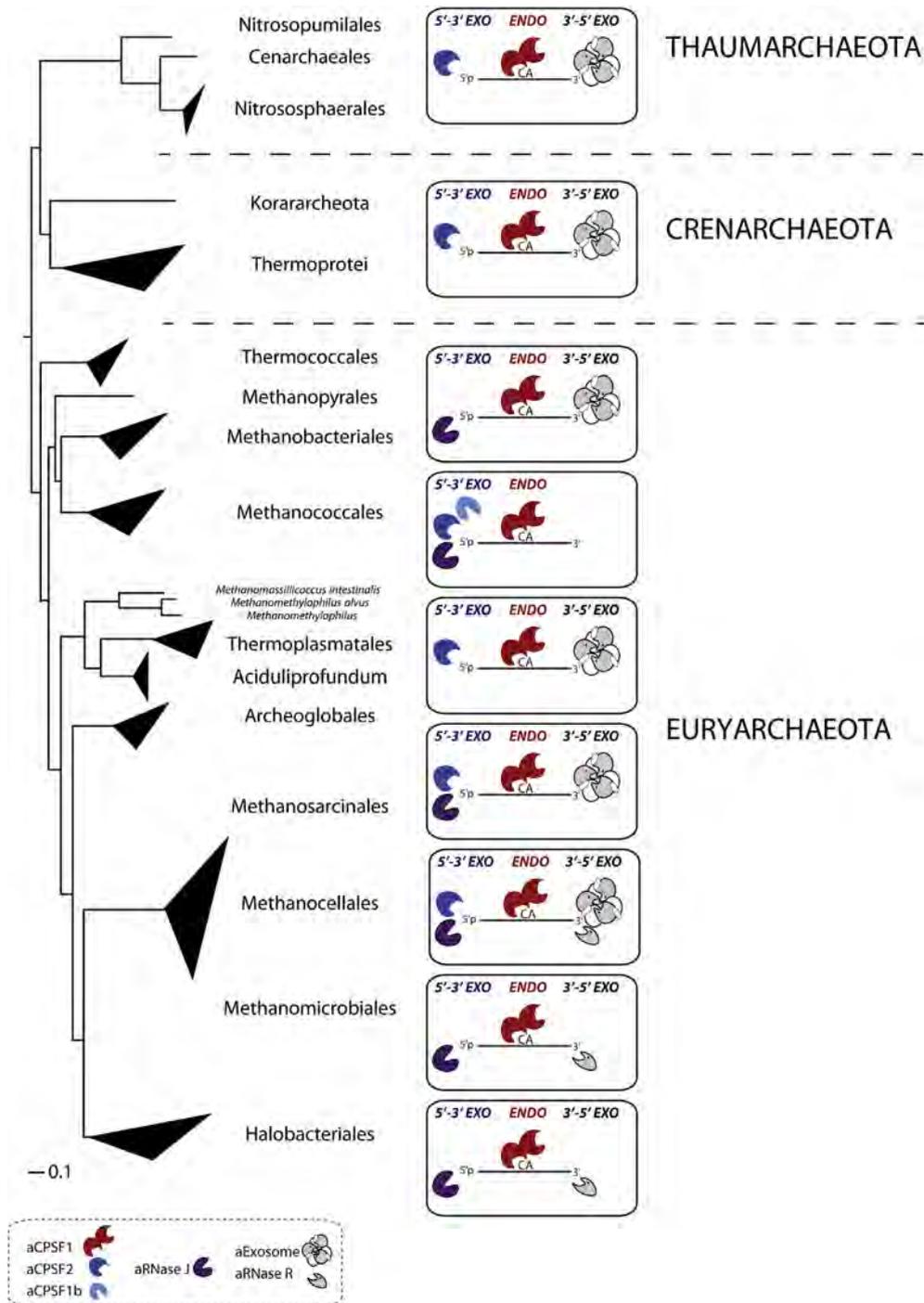


Fig. 3. Schematic phylogenetic distribution of ribonucleolytic activities carried by the four major β -CASP orthologous groups, the exosome and aRNase J in *Archaea*. Colour and symbol for each enzyme are indicated. The archaeal tree was constructed from a concatenated alignment of 81 orthologous groups of well conserved protein sequences in 131 complete archaeal genomes (for methods see Ref. [24]). Subtrees including sequences from the same phyla were collapsed. Note that aCPSF1 may also act as a 5'-3' exoribonuclease.

an 5'-3' exoribonucleolytic activity mainly restricted to 5' monophosphorylated RNA substrates, comparable to eukaryal CPSF73 factor within the machinery required for termination of RNA polymerase II transcription [24]. The cleavage specificity at single-stranded CA dinucleotide is still to be elucidated. The 5' monophosphate requirement of the exoribonucleolytic activity is strict since 5' triphosphorylated transcripts cannot be degraded exonucleolytically by Pab-aCPSF1. The same biochemical properties are shared by aRNase J (see above) and bacterial RNase J

[20,31,43]. In contrast, *M. jannaschii* aCPSF1 (initially named mjRNase J2 [22]), was shown to display only an endonucleolytic activity with no detectable exonucleolytic activity. Recombinant *M. thermautotrophicus* aCPSF1 was also shown to be a nuclease but these experiments did not allow the authors to determine whether it is an endo- or an exonuclease [23]. The apparent differences in the reported activities of *M. jannaschii* and *M. thermautotrophicus* aCPSF1 compared to Pab-aCPSF1 may simply be due to differences in protein purification procedures or

experimental conditions. Indeed significant *in vitro* RNA cleavages by the eukaryal ortholog CPSF73 and *P. abyssi* aCPSF1 required relatively high concentrations of enzymes implying that only a fraction of the CPSF protein preparations are active for RNA cleavage [24,30]. Since it is known that CPSF73 operates in a multi-protein complex, it is possible that the activities of these enzymes might be influenced by interaction with other protein partners. Since both eukaryal CPSF73 and pyrococcal aCPSF1 both have endo- and exoribonucleolytic activities, it seems likely that most aCPSF1 members also possess these biochemical properties [30,44,45].

All aCPSF1 members harbor an N-terminal extension with two KH motifs which are not essential for aCPSF1 ribonucleolytic activities but important for its RNA binding capacities [23,24]. This is consistent with the tripartite architecture of the three reported aCPSF1 crystal structures from *P. horikoshii*, *M. thermautotrophicus* and *M. mazei*. In these, both KHa and KHb motifs are folded into independent domains separate from the core β -CASP and β -lactamase domains (Fig. 2B) [23,38,39].

aCPSF1 members are dimeric in solution with a dimer interface composed of 12 amino acids [22–24,38,39] that is conserved among all aCPSF1 members (Fig. 2B) [24]. Interestingly, disruption of this interface impairs the exonucleolytic activity of Pab-aCPSF1 but not its endonucleolytic activity. For the bacterial β -CASP enzymes, it has been proposed that the exonucleolytic and endonucleolytic activities involve two alternative enzyme configurations [34] (see above section). If this is indeed the case, one explanation for the results obtained with Pab-aCPSF1 is that only the dimeric form of aCPSF1 can adopt an exonucleolytic conformation. The ability to separate the two aCPSF1 activities experimentally now provides a powerful tool to explore the potential conformational changes involved in the exo- versus endonucleolytic modes of RNA trimming of the archaeal enzymes and of β -CASP enzymes from all three domains of life.

5. Taxonomic distribution of archaeal β -CASP ribonuclease activities

All *Crenarchaeota* and *Thaumarchaeota* genomes encode aCPSF1 and aCPSF2 copies. On the other hand, although aCPSF1 is also conserved in *Euryarchaeota*, the distribution of aCPSF2, aRNase J and aCPSF1b is more heterogeneous (Fig. 3) [24]. In light of reported *in vitro* enzymatic activities (see above), β -CASP RNA-degrading activities can now be envisioned for each archaeal clade (Fig. 3). Each archaeal genome encodes at least one 5'-3' β -CASP exoribonuclease. This activity could be carried either by an aCPSF2 and/or an aRNase J group member.

From biochemical studies (see above), primary transcripts with 5' triphosphorylated ends will be protected from aRNase J, aCPSF2, aCPSF1b and aCPSF1 exonucleolytic activity whereas aCPSF1 can potentially act as an endonuclease on primary (5' triphosphorylated) and secondary (5' monophosphorylated) transcripts and as 5'-3' exonuclease exclusively on secondary transcripts (Fig. 1).

Furthermore, the addition of the phylogenetic distribution of the 3'-5' exonucleolytic activity carried either by the archaeal exosome and aRNase R, allow to highlight the prospective archaeal RNA-degrading activities for each archaeal clade (Fig. 3) [7,16,17,24]. In *Eukarya* and *Bacteria*, the conserved eukaryal exosome and bacterial polynucleotide phosphorylase (PNPase) homomultimeric counterpart have been observed to be part of essential RNA decay machines that relies on many factors such as RNA helicases, endoribonucleases and poly(A) polymerase activities [46,47]. These factors help to recruit these RNA-degrading machines to substrates and/or make them eligible for

degradation. To date, archaeal RNA degradation multi-protein complex machinery have not been yet identified. The hypothesis of an interaction between aCPSF1 and several archaeal exosomal components using *M. thermautotrophicus* was recently tested by pull-down experiments but no interactions were detected *in vitro* [23]. Besides it should be noted that the archaeal exosome is missing from *Methanococcales*, *Methanomicrobiales* and *Halobacteriales* [15,24]. In the *Halobacteriales*, the acquisition of a bacterial RNase R protein is supposed to functionally compensate the lack of the archaeal exosome [7,18]. The *Methanococcale* clade is particularly interesting since both archaeal exosome and aRNase R are missing (Fig. 3). Remarkably, this correlates with the presence of aCPSF1b members which are restricted to this clade [24]. Therefore the deficiency of a 3'-5' exonucleolytic activity in *Methanococcales* could be compensated somehow either by an exoribonuclease that is not related to exosome and RNase R and/or by the 5'-3' exonucleolytic activity of the restricted β -CASP aCPSF1b group.

6. Towards β -CASP ribonuclease functions in archaeal cell metabolism

Although the functions of the β -CASP ribonucleases in *Archaea* are not as well understood as in *Eukarya* and *Bacteria*, the available data suggest that these ribonucleases may have related functions in the three domains of life. Experimental evidence shows that both bacterial RNase J and eukaryal CPSF73 are part of large multi-enzyme complexes involved in both the maturation and degradation of mRNA [48,49]. The entry of RNase J into the mRNA degradation pathway is controlled by the phosphorylation state of the 5' end whereas mRNA 3' end formation by CPSF73 apparently depends on signals in the mRNA that are detected by the CPSF complex [50,51]. The bacterial and eukaryal β -CASP ribonucleases appeared as versatile ancient enzymes that act at both the 5' and 3' ends of mRNA to control their stability and maturation. The archaeal translation initiation factor a/eIF2(- γ) was shown to bind and protect RNA with 5'-triphosphorylated ends from degradation in the *Crenarchaeon* *S. solfataricus* [8,21]. This observation suggests parallels to the principal mechanisms of 5'-3' RNA decay in *Bacteria* and *Eukarya* [7], in which Nudix hydrolases in *Bacteria* [52,53] and decapping enzymes in *Eukarya* [54] trigger mRNA degradation by producing 5' monophosphate ends (Fig. 1, right & left panels). Nevertheless, comparable enzymes remain to be discovered in the *Archaea*. In conclusion, the nature of the substrate 5' end (tri- vs monophosphorylated) emerges as a major determinant in the activity of β -CASP ribonucleases. The next challenge is to identify in which, RNA maturation, RNA decay or RNA quality control pathways, aCPSF1, aCPSF2 and aRNase J enzymes are acting.

A first clue towards the biological relevance of aCPSF2 was recently obtained in *Sulfolobus acidocaldarius* (Saci) strain of the thermoprotei crenarchaeal clade [41]. Comparison of the transcriptome of wild type and aCPSF2-deleted strains revealed differential transcript abundance for 560 genes [41]. This global effect suggests that Saci-aCPSF2 enzyme plays a major role in 5'-3' directed mRNA decay in *S. acidocaldarius* mRNA turnover [41]. Although the *S. acidocaldarius* wild type and mutant strains had the same growth rate meaning that Saci-aCPSF2 is obviously not an essential enzyme and that the observed differential transcript abundance does not affect the overall cellular metabolism. We can speculate that the lack of the aCPSF2 5'-3' exoribonucleolytic activity is somehow compensated by other ribonucleolytic activities (as other β -CASP or unknown 5'-3' exoribonuclease(s) or the 3'-5' exoribonucleolytic activity of the exosome) to maintain a proper functional pool of RNAs. To date, no other direct analysis of the role

of 5'-3' exoribonucleolytic β -CASP activities has been reported. The redundancy of these activities in most of euryarchaeal cells may reflect a more versatile function of aCPSF2 and aRNase J. This may contrast with aCPSF1 activity which is likely essential in each archaeal cell in light of its strict phylogenetic conservation. By analogy to the eukaryal CPSF73, which is part of a multi-protein complex, clues to the function of the archaeal homolog might come from future studies aimed at identifying archaeal complexes containing aCPSF1 members. Interestingly, a recent study using *Haloferax volcanii* (Hvo) of the euryarchaeal halobacterial clade as model identified Hvo-aCPSF1 as a protein partner of a tRNA maturation pathway [55]. The *H. volcanii* tRNA thiolase enzyme (Hvo-NcsA) which was shown to be essential for maintaining cellular pools of thiolated tRNA^{Lys}_{UUU} and for *H. volcanii* growth at high temperature formed a complex with Hvo-aCPSF1. In addition, proteins involved in ubiquitin proteasome and translation are also recovered within Hvo-NcsA proteome obtained in pull down assays using *H. volcanii* strains expressing tagged version of Hvo-NcsA [55]. It is not clear yet how all those partners interact and act within the tRNA thiolation pathway. Nevertheless the interaction of aCPSF1 with components of the proteasome suggests a connection between RNA processing and protein degradation. This hypothesis is supported by the genomic context of aCPSF1 in which the gene encoding one proteasome subunit (the one located outside of the exosome/proteasome superoperon) is adjacent to the gene encoding aCPSF1 in most archaeal genomes [15] (Y. Quentin & B. Clouet-d'Orval, unpublished observations). The assumption that genes from the same operon encode proteins of the same complex or involved in the same process suggest that the proteasome and aCPSF1 could be functionally coordinated. Furthermore a superoperon that encodes the predicted exosome components also encoded the catalytic subunits of the archaeal proteasome. Altogether the genomic context analysis suggests interplay between exosome and proteasome and aCPSF1 [7,15]. This is consistent with RNA quality control systems described in *Eukarya* (NMD, NSD, NRD pathways) and in *Bacteria* (trans-translation tmRNAs) where aberrant mRNAs are induced by aberrant translation, elongation and termination (see for review [56]). Moreover in eukaryotes the interdependence of mRNA translation and decay is immediately evident from the functional roles of the 5'-cap and the 3'-poly(A) tail that promote translation initiation and antagonizes mRNA decay, respectively [57].

7. Concluding remarks

β -CASP proteins constitute a large family of enzymes with ribonucleolytic activities in the three domains of life. Archaeal β -CASP members should ensure major roles in archaeal cell metabolism considering their phylogenetic distribution and the functions of their eukaryal and bacterial β -CASP counterparts, CPSF73 and RNase J. First clues towards the understanding of these enzyme functions come from in-depth phylogenomics, biochemical, transcriptomic and proteomic analyses. Initial conclusions suggest a role of archaeal β -CASP enzymes in RNA processing mechanisms, mechanisms which still need to be described in archaeal cells. Next challenges are the understanding of the interplay with the exosome, proteasome and transcription machineries. What gives members of the β -CASP family their specificity *in vivo* is still an open question (Fig. 1, central panel) although clues to this puzzle might emerge from the search of multi-protein complex(es) targeted to RNA processing. Moreover, a deeper knowledge of structural and biochemical characteristics of archaeal β -CASP ribonucleases of which some but not all possess both types of activity (endo- and 5'-3' exo-) will certainly expand our understanding of enzymatic

mechanisms that rationalize the dual-enzymatic modes and the RNA substrate recognition of this enzyme family in the three domains of life.

Finally archaeal β -CASP phylogenetic distribution raises the question of the origin of β -CASP proteins in the *Archaea*. The comparison of the archaeal specie tree and aRNase J, aCPSF1 and aCPSF2 protein sequence trees, respectively, shows that all β -CASP proteins have an evolutionary pattern dominated by vertical inheritance [20,24] (Quentin, Y. unpublished observation). Obviously, this is consistent with an ancient origin of aCPSF1 and aCPSF2 orthologous groups suggesting a duplication of an aCPSF ancestor predating the emergence of *Archaea* [24] and with the presence of aRNase J in the common ancestor of the *Euryarchaeota* [20]. Nevertheless the presence of RNase J orthologs in *Bacteria* and its absence in *Eukarya* suggest evolutionary scenarios that require gene losses or horizontal gene transfer at a fairly early stage in evolution. Alternative scenarios can be proposed based on the phylogenetic relationships between *Archaea*, *Bacteria* and *Eukarya* which are at present a fascinating debate in progress [58–60,62].

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Thanks to G. Fichant and members of the Carpousis group for helpful discussions and M. Chandler and M. Bouvier for critical reading of the manuscript. BCO and YQ are supported by the Centre National de la Recherche Scientifique (CNRS) with additional funding from the Agence Nationale de la Recherche (ANR) (grant BLAN08-1_329396) and from the Université de Toulouse (UPS, AO1_2011 & IDEX « Emergence » program 2014).

References

- [1] G. Stoecklin, O. Muhlemann, RNA decay mechanisms: specificity through diversity, *Biochimica biophysica acta* 1829 (2013) 487–490.
- [2] C.M. Arraiano, F. Mauxion, S.C. Viegas, R.G. Matos, B. Seraphin, Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA, *Biochimica biophysica acta* 1829 (2013) 491–513.
- [3] M.P. Hui, P.L. Foley, J.G. Belasco, Messenger RNA degradation in bacterial cells, *Annu. Rev. Genet.* 48 (2014) 537–559.
- [4] S. Laalami, L. Zig, H. Putzer, Initiation of mRNA decay in bacteria, *Cell. Mol. life Sci.* : CMLS 71 (2014) 1799–1828.
- [5] H. Tourriere, K. Chebli, J. Tazi, mRNA degradation machines in eukaryotic cells, *Biochimie* 84 (2002) 821–837.
- [6] H. Celesnik, A. Deana, J.G. Belasco, Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal, *Mol. Cell.* 27 (2007) 79–90.
- [7] E. Evguenieva-Hackenberg, G. Klug, RNA degradation in *Archaea* and gram-negative bacteria different from *Escherichia coli*, *Prog. Mol. Biol. Transl. Sci.* 85 (2009) 275–317.
- [8] D. Hasenohrl, T. Lombo, V. Kabardin, P. Londei, U. Blasi, 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 (2008) 2146–2150.
- [9] D.N. Frank, N.R. Pace, Ribonuclease P: unity and diversity in a tRNA processing ribozyme, *Annu. Rev. Biochem.* 67 (1998) 153–180.
- [10] W.Y. Chen, D. Singh, L.B. Lai, M.A. Stiffler, H.D. Lai, M.P. Foster, V. Gopalan, 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.* 40 (2012) 4666–4680.
- [11] S. Schiffer, S. Rosch, A. Marchfelder, Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes, *EMBO J.* 21 (2002) 2769–2777.
- [12] Y. Redko, I. Li de Lasierra-Gallay, C. Condon, When all's zed and done: the structure and function of RNase Z in prokaryotes, *Nature reviews, Microbiology* 5 (2007) 278–286.
- [13] H. Li, C.R. Trotta, J. Abelson, Crystal structure and evolution of a transfer RNA splicing enzyme, *Science* 280 (1998) 279–284.
- [14] T.H. Tang, T.S. Rozhdestvensky, B.C. d'Orval, M.L. Bortolin, H. Huber, B. Charpentier, C. Branlant, J.P. Bachellerie, J. Brosius, A. Huttenhofer, RNomics

- in Archaea reveals a further link between splicing of archaeal introns and rRNA processing, *Nucleic acids Res.* 30 (2002) 921–930.
- [15] E.V. Koonin, Y.I. Wolf, L. Aravind, 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 (2001) 240–252.
- [16] E. Lorentzen, P. Walter, S. Fribourg, E. Evguenieva-Hackenberg, G. Klug, E. Conti, The archaeal exosome core is a hexameric ring structure with three catalytic subunits, *Nat. Struct. Mol. Biol.* 12 (2005) 575–581.
- [17] E. Evguenieva-Hackenberg, P. Walter, E. Hochleitner, F. Lottspeich, G. Klug, An exosome-like complex in *Sulfolobus solfataricus*, *EMBO reports* 4 (2003) 889–893.
- [18] V. Portnoy, G. Schuster, RNA polyadenylation and degradation in different Archaea: roles of the exosome and RNase R, *Nucleic acids Res.* 34 (2006) 5923–5931.
- [19] E. Evguenieva-Hackenberg, L. Hou, S. Glaeser, G. Klug, Structure and function of the archaeal exosome, *Wiley interdisciplinary reviews, RNA* 5 (2014) 623–635.
- [20] B. Clouet-d'Orval, D. Rinaldi, Y. Quentin, A.J. Carpousis, Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity, *J. biological Chem.* 285 (2010) 17574–17583.
- [21] D. Hasenohrl, R. Konrat, U. Blasi, 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 (2011) 99–107.
- [22] S. Levy, V. Portnoy, J. Admon, G. Schuster, Distinct activities of several RNase J proteins in methanogenic archaea, *RNA Biol.* 8 (2011) 1073–1083.
- [23] A.P. Silva, M. Chechik, R.T. Byrne, D.G. Waterman, C.L. Ng, E.J. Dodson, E.V. Koonin, A.A. Antson, C. Smits, Structure and activity of a novel archaeal beta-CASP protein with N-terminal KH domains, *Structure* 19 (2011) 622–632.
- [24] D.K. Phung, D. Rinaldi, P.S. Langendijk-Genevaux, Y. Quentin, A.J. Carpousis, B. Clouet-d'Orval, Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor, *Nucleic acids Res.* 41 (2013) 1091–1103.
- [25] I. Callebaut, D. Moshous, J.P. Mornon, J.P. de Villartay, Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family, *Nucleic acids Res.* 30 (2002) 3592–3601.
- [26] Z. Dominski, A.J. Carpousis, B. Clouet-d'Orval, Emergence of the beta-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation, *Biochimica biophysica acta* 1829 (2013) 532–551.
- [27] Z. Dominski, X.C. Yang, M. Purdy, E.J. Wagner, W.F. Marzluff, A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100, *Mol. Cell. Biol.* 25 (2005) 1489–1500.
- [28] D. Baillat, M.A. Hakimi, A.M. Naar, A. Shilatifard, N. Cooch, R. Shiekhattar, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II, *Cell* 123 (2005) 265–276.
- [29] C. Condon, L. Gilet, The Metallo-β-lactamase Family of Ribonucleases, *Nucleic Acids and Molecular Biology*, Springer-verlag, Berlin Heidelberg, 2011, pp. 245–267.
- [30] C.R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J.L. Manley, L. Tong, Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease, *Nature* 444 (2006) 953–956.
- [31] I. Li de la Sierra-Gallay, L. Zig, A. Jamalli, H. Putzer, Structural insights into the dual activity of RNase J, *Nat. Struct. Mol. Biol.* 15 (2008) 206–212.
- [32] A. Dorleans, I. Li de la Sierra-Gallay, J. Piton, L. Zig, L. Gilet, H. Putzer, C. Condon, Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J, *Structure* 19 (2011) 1252–1261.
- [33] J.A. Newman, L. Hewitt, C. Rodrigues, A. Solovyova, C.R. Harwood, R.J. Lewis, Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1, *Structure* 19 (2011) 1241–1251.
- [34] J. Richards, J.G. Belasco, Ribonuclease J: how to lead a double life, *Structure* 19 (2011) 1201–1203.
- [35] S. Even, O. Pellegrini, L. Zig, V. Labas, J. Vinh, D. Brechemmier-Baey, H. Putzer, Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E, *Nucleic acids Res.* 33 (2005) 2141–2152.
- [36] N.V. Grishin, KH domain: one motif, two folds, *Nucleic acids Res.* 29 (2001) 638–643.
- [37] R. Valverde, L. Edwards, L. Regan, Structure and function of KH domains, *FEBS J.* 275 (2008) 2712–2726.
- [38] Y. Nishida, H. Ishikawa, S. Baba, N. Nakagawa, S. Kuramitsu, R. Masui, Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*, *Proteins* 78 (2010) 2395–2398.
- [39] B. Mir-Montazeri, M. Ammelburg, D. Forouzan, A.N. Lupas, M.D. Hartmann, Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor, *J. Struct. Biol.* 173 (2011) 191–195.
- [40] H. Ishikawa, N. Nakagawa, S. Kuramitsu, R. Masui, Crystal structure of TTHA0252 from *Thermus thermophilus* HB8, a RNA degradation protein of the metallo-beta-lactamase superfamily, *J. Biochem.* 140 (2006) 535–542.
- [41] B. Martens, F. Amman, S. Manoharadas, L. Zeichen, A. Orell, S.V. Albers, I. Hofacker, U. Blasi, Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2, *PLoS one* 8 (2013) e76569.
- [42] D.K. Phung, B. Clouet-d'Orval, Tips and Tricks to Probe the RNA-degrading activities of hyperthermophilic archaeal beta-CASP ribonucleases, *Methods Mol. Biol.* 1259 (2015) 453–466.
- [43] N. Mathy, L. Benard, O. Pellegrini, R. Daou, T. Wen, C. Condon, 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA, *Cell* 129 (2007) 681–692.
- [44] K. Ryan, O. Calvo, J.L. Manley, Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease, *RNA* 10 (2004) 565–573.
- [45] X.C. Yang, K.D. Sullivan, W.F. Marzluff, Z. Dominski, Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing, *Mol. Cell. Biol.* 29 (2009) 31–42.
- [46] A. Chlebowski, M. Lubas, T.H. Jensen, A. Dziembowski, RNA decay machines: the exosome, *Biochimica biophysica acta* 1829 (2013) 552–560.
- [47] S. Nurmohamed, B. Vaidialingam, A.J. Callaghan, B.F. Luisi, Crystal structure of *Escherichia coli* polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly, *J. Mol. Biol.* 389 (2009) 17–33.
- [48] M. Lehnik-Habrink, H. Pfortner, L. Rempeters, N. Pietack, C. Herzberg, J. Stulke, The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex, *Mol. Microbiol.* 77 (2010) 958–971.
- [49] Z. Dominski, The hunt for the 3' endonuclease, *Wiley interdisciplinary reviews, RNA* 1 (2010) 325–340.
- [50] C.R. Mandel, Y. Bai, L. Tong, Protein factors in pre-mRNA 3'-end processing, *Cell. Mol. life Sci. : CMLS* 65 (2008) 1099–1122.
- [51] N.J. Proudfoot, Ending the message: poly(A) signals then and now, *Genes Dev.* 25 (2011) 1770–1782.
- [52] J. Richards, Q. Liu, O. Pellegrini, H. Celesnik, S. Yao, D.H. Bechhofer, C. Condon, J.G. Belasco, An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*, *Mol. Cell.* 43 (2011) 940–949.
- [53] A. Deana, H. Celesnik, J.G. Belasco, The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal, *Nature* 451 (2008) 355–358.
- [54] M.G. Song, Y. Li, M. Kiledjian, Multiple mRNA decapping enzymes in mammalian cells, *Mol. Cell.* 40 (2010) 423–432.
- [55] N.E. Chavarria, S. Hwang, S. Cao, X. Fu, M. Holman, D. Elbanna, S. Rodriguez, D. Arrington, M. Englert, S. Uthandi, D. Soll, J.A. Maupin-Furlow, Archaeal Tuc1/Ncs6 homolog required for wobble uridine tRNA thiolation is associated with ubiquitin-proteasome, translation, and RNA processing system homologs, *PLoS one* 9 (2014) e99104.
- [56] T. Inada, Quality control systems for aberrant mRNAs induced by aberrant translation elongation and termination, *Biochimica biophysica acta* 1829 (2013) 634–642.
- [57] B. Roy, A. Jacobson, The intimate relationships of mRNA decay and translation, *Trends Genet. : TIG* 29 (2013) 691–699.
- [58] T.A. Williams, P.G. Foster, C.J. Cox, T.M. Embley, An archaeal origin of eukaryotes supports only two primary domains of life, *Nature* 504 (2013) 231–236.
- [59] T.A. Williams, T.M. Embley, Archaeal “dark matter” and the origin of eukaryotes, *Genome Biol. Evol.* 6 (2014) 474–481.
- [60] A. Spang, J.H. Saw, S.L. Jorgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A.E. Lind, R. van Eijk, C. Schleper, L. Guy, T.J. Ettema, Complex archaea that bridge the gap between prokaryotes and eukaryotes, *Nature* 521 (2015) 173–179.
- [61] K.J. Bandrya, M. Bouvier, A.J. Carpousis, B.F. Luisi, The social fabric of the RNA degradosome, *Biochimica biophysica acta* 1829 (2013) 514–522.
- [62] K. Raymann, C. Brochier-Armanet, S. Gribaldo, The two-domain tree of life is linked to a new root for the Archaea, *Proc Natl Acad Sci U S A* 112 (21) (2015) 6670–6675.

**IDENTIFICATION ET CLASSIFICATION DES RIBONUCLEASES DE LA
FAMILLES DES β -CASP CHEZ LES ARCHEES ET CARACTERISATION
DE aCPSF1 DE *P. abyssi***

PUBLICATION 4

**Tips and Tricks to Probe the RNA-Degrading Activities
of Hyperthermophilic Archaeal β -CASP Ribonucleases**

Duy Khanh Phung and Béatrice Clouet-d'Orval.

Centre National de la Recherche Scientifique, LMGM and Université de Toulouse, Toulouse,
France

Tips and Tricks to Probe the RNA-Degrading Activities of Hyperthermophilic Archaeal β -CASP Ribonucleases

Duy Khanh Phung and Béatrice Clouet-d'Orval

Abstract

The importance of ribonucleases in posttranscriptional control of gene expression has been established in Eukarya and Bacteria for over a decade. However, this process has been overlooked in Archaea, which are of universal importance to elucidate fundamental biological mechanisms and to study the evolution of life on Earth. Very few ribonucleolytic activities have been reported in Archaea, and RNA metabolism pathways wait to be described. Recently we have identified two major groups of archaeal ribonucleases, aCPSF1 and aRNase J, which are members of the β -CASP metallo- β -lactamase family. Here, we describe in vitro methods to characterize the endo- and exoribonucleolytic activities of hyperthermophilic archaeal β -CASP ribonucleases. The use of various labeled RNA substrates allows defining the specificity of RNA cleavage and the directionality of the exoribonucleolytic trimming activity of the archaeal enzymes which work at high temperature. Elucidating in vitro ribonucleolytic activities is one step toward the understanding of the role of β -CASP ribonucleases in RNA metabolism pathways in archaeal cells.

Key words β -CASP ribonucleases, Exoribonucleolytic activity, Endoribonucleolytic activity, Archaea, Thermococcales, Hyperthermophilic enzymes

1 Introduction

RNA-modifying enzymes have been extensively described in Bacteria and in Eukarya. A large number of ribonucleases are essential for the processing of primary transcripts to mature RNAs, for quality control, for host defense, and for regulation of gene expression. Ribonucleases operate as endonucleases and 3'-to-5' or 5'-to-3' exoribonucleases and cooperate in different RNA degradation and maturation systems (for review, *see* [1]). Very few activities have been reported in Archaea [2], and RNA metabolism pathways wait to be described. First hints toward identifying RNA processing and degradation pathways have come recently from biochemical characterizations and phylogenomic studies of two major archaeal β -CASP ribonucleases: aCPSF1 and aRNase J [3–6].

The β -CASP nucleases have emerged as central to RNA metabolism over the last decade [7–10]. These enzymes, which catalyze both endo- and 5'-to-3' exoribonucleolytic degradation [11], are unique among all known ribonucleases (for review, *see* [10]). The β -CASP enzymes acting on RNA form two separate groups, one related to the eukaryal cleavage and polyadenylation factor CPSF73 [12] and the other to bacterial RNase J [13]. The archaeal ribonucleases aCPSF1 and aRNase J, orthologs of CPSF73 and RNase J, respectively, are phylogenetically conserved among archaeal taxonomic groups [4]. Thus, Archaea might possess a composite RNA processing system sharing both eukaryal and bacterial features. This highlights the importance in developing archaeal model to gain insight into molecular mechanisms and evolution of fundamental processes of life. We used the thermococcale *Pyrococcus abyssi* as a model which was identified in an elevated temperature ecological niche, the deep-sea hydrothermal vent [14]. Therefore, we developed specific protocols to decipher the ribonucleolytic activities of hyperthermophilic β -CASP enzymes. This short chapter presents briefly some tips and tricks specific to probe hyperthermophilic ribonucleolytic activities and is a complement to techniques described earlier to characterize other archaeal RNases [15] and to the extensive tutorial for investigating ribonucleolytic activities provided by Tomecki et al. in Chapter 25 of this book. The methods described in the three main sections to characterize the degradation patterns of in vitro transcribed RNA substrates by recombinant archaeal hyperthermophilic ribonucleases is summarized in Fig. 1.

A combination of RNA-degrading assays is performed at 65 °C. Note that the optimal growth temperature of hyperthermophilic archaeon could exceed 100 °C. However, we choose 65 °C to avoid significant in vitro chemical (nonenzymatic) hydrolysis of RNA at higher temperature. RNA substrates labeled in different manners are engineered to determine the type(s) of the ribonucleolytic activities (5' to 3' exo-, 3' to 5' exo- and endo-) involved, as explained extensively in Chapter 25. Various RNA substrates with the same nucleotide sequence but having different 5' or 3' end radiolabels allow the identification of precise internal RNA cleavage sites and/or of the directionality and processivity status of the exonucleolytic activity. Importantly, side-by-side characterization of wild-type and catalytically inactive variants (bearing mutations of the charged residues that coordinate pair of zinc ions in the catalytic site, [11]) of β -CASP enzymes is necessary to certify that the observed in vitro degradation patterns are due to the β -CASP enzymes and not to contaminant nuclease activities.

5. Standard LB broth and LB-agar plates containing 25 µg/mL chloramphenicol and 100 µg/mL ampicillin.
6. Lysis buffer: 300 mM sodium phosphate pH 7.4, 0.5 mM DTT, 10 % (v/v) glycerol, 35,000 U/mL lysozyme, protease inhibitor cocktail tablet EDTA-free (Roche). The tablet is added right before use.
7. 1 M imidazole in water.
8. Sonicator for bacterial cell lysis.
9. Automated chromatography system (e.g., Akta Pure M from GE Healthcare).
10. Prepacked 1 mL cobalt affinity resin, HiTrap™ TALON® crude (GE healthcare).
11. Buffer A: 300 mM sodium phosphate pH 7.4, 0.5 mM DTT, 10 % glycerol.
12. Buffer B: 300 mM sodium phosphate pH 7.4, 0.5 mM DTT, 10 % glycerol, 1 M imidazole.
13. Prepacked 1 mL anion exchange column, HiTrap Q FF (GE healthcare).
14. Buffer AQ: 50 mM Tris-Cl pH 7.4, 0.5 mM DTT, 10 % glycerol.
15. Buffer BQ: 1 M NaCl, 50 mM Tris-Cl pH 7.4, 0.5 mM DTT, 10 % glycerol.
16. Storage buffer: 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT, and 10 % glycerol.
17. Centrifugal filtration device (e.g., Vivaspın, 6 mL concentrator, 30,000 MWCO).
18. 0.5 U/µL thrombin.
19. Precast 4–15 % Biorad SDS-PAGE gels and vertical electrophoresis system.

2.2 RNA Substrate Synthesis, Labeling, and Purification

1. 10× transcription buffer: 400 mM Tris-HCl pH 7.5 measured at 37 °C, 60 mM MgCl₂, 100 mM NaCl, 20 mM spermidine.
2. 10× TAP reaction buffer: 0.5 M sodium acetate (pH 6.0), 10 mM EDTA, 1 % β-mercaptoethanol, and 0.1 % Triton® X-100.
3. 80 U/µL T7 RNA polymerase.
4. 10 mM solution of each rNTP.
5. RNasin inhibitor (Promega).
6. 10 U/µL Tobacco Acid Pyrophosphatase.
7. Phenol/chloroform/isoamyl alcohol (25:24:1) mixture.
8. 3 M sodium acetate solution, adjusted to pH 5.2 with acetic acid.

9. Ethanol.
10. [α - ^{32}P] UTP (800 Ci/mmol).
11. [γ - ^{32}P] GTP (6,000 Ci/mmol).
12. [$5'$ - ^{32}P] pCp (3,000 Ci/mmol).
13. [γ - ^{32}P] ATP (6,000 Ci/mmol).
14. 1 U/ μL CIAP phosphatase.
15. 10 U/ μL T4 polynucleotide kinase.
16. DMSO (dimethyl sulfoxide).
17. 20 U/ μL T4 RNA ligase.
18. Vertical polyacrylamide gel electrophoresis (PAGE) system with power supply and gel casting parts (16 \times 16 cm glass plates, 0.8 mm spacers and comb).
19. 10 \times TBE buffer: 890 mM Tris base, 890 mM boric acid, 20 mM EDTA pH 8.0. Sterilize with a 0.22 μm pore size filter.
20. Denaturing polyacrylamide gel containing 8 M urea and 1 \times TBE (*see Note 3*). Add 60 μL of 10 % (w/v) ammonium persulfate and 5 μL of *N,N,N',N'*-tetramethylethylenediamine per 10 mL of gel solution and quickly pour mixture into gel plates equipped with spacers. Insert the appropriate comb and wait for \sim 30 min until polymerization is complete.
21. RNA elution buffer: 0.5 M ammonium acetate (NH_4OAc), 1 mM EDTA.
22. Speedvac apparatus.
23. Micro Bio-Spin P30 gel columns (Bio-Rad).

2.3 RNA Degradation Assays

1. 5 \times NR buffer: 500 mM KCl, 7.5 mM MgCl_2 , 100 mM HEPES pH 7.5, 10 % glycerol.
2. Proteinase K mix: 20 mM HEPES pH 8, 10 mM EDTA, 0.2 % (w/v) SDS, 0.5 ng/mL proteinase K.
3. 0.5 M EDTA solution. Adjust to pH 8 with NaOH.
4. FUD (formamide urea dye): 20 mL deionized formamide, 5 g urea, 12.5 mg xylene cyanol, 12.5 mg bromophenol blue, 0.25 mL of 0.5 M EDTA pH 8, 0.5 mL of 5 \times TBE.
5. Heating block or thermocycler (*see Note 4*).
6. T1 ladder: Mix 1 μL ^{32}P -RNA (5,000 cpm/ μL), 1 μL *E. coli* tRNA 10 $\mu\text{g}/\mu\text{L}$, and 4 μL of a solution containing 20 mM sodium citrate, 1 mM EDTA, and 7 M urea. Heat 1 min at 95 $^\circ\text{C}$. Cool on ice for 5 min and then add 1 μL of RNase T1 (0.01 U/ μL). Heat 10 min at 55 $^\circ\text{C}$ and add 10 μL of FUD.
7. OH $^-$ ladder: Mix 1 μL of ^{32}P -RNA (5,000 cpm/ μL) and 10 μL of a solution containing 50 mM $\text{Na}_2\text{CO}_3/\text{NaHCO}_2$ and 1 mM EDTA. Heat 5 min at 95 $^\circ\text{C}$ and add 10 μL of FUD.

8. Pre-coated, POLYGRAM® CEL 300 PEI/UV₂₅₄ thin layer chromatography (TLC) sheets (Macherey-Nagel).
9. 0.25 M KH₂PO₄.
10. Phosphorimager with exposure plates and dedicated quantification software.
11. Vertical polyacrylamide gel electrophoresis (PAGE) system with power supply and gel casting parts (38×20 cm glass plates, 0.4 mm spacers, and comb).
12. **Items 19 and 20** of Subheading [2.2](#).

3 Methods

3.1 Preparation of Active Recombinant Archaeal aRNase J and aCPSF1 with an N-Terminal 6xHis-Tag

3.1.1 Production of Cell Paste

1. Transform the pET15b derivative encoding the β-CASP ribonuclease of interest into BL21-CodonPlus (DE3) cells. Select transformants on LB plates supplemented with chloramphenicol and ampicillin.
2. Use a single colony to inoculate 20 mL of LB medium supplemented with chloramphenicol and ampicillin. Incubate overnight at 37 °C.
3. Dilute (1/100) the overnight pre-culture into 800 mL of medium (*see Note 5*). Incubate at 37 °C until optical density at 600 nm (OD₆₀₀) reaches ~0.6.
4. Add IPTG to a final concentration of 0.1 mM and incubate 3 h at 30 °C.
5. Centrifuge for 15 min at 8,000×g.
6. Remove supernatant and store the cell pellet frozen at -20 °C.

3.1.2 Purification of N-Terminally (His)₆-Tagged β-CASP Ribonucleases

1. Resuspend the cell pellet in ice-cold lysis buffer (~5 mL per gram of pellet) containing 10 mM imidazole and incubate at 4 °C for 20 min. Sonicate the mixture on ice using 15 pulses of 10 s each (active cycle 50 %; conditions may vary with the sonicator).
2. Heat the cell extract to 70 °C for 20 min and clarify by centrifugation at 15,000×g for 20 min to eliminate the precipitated, heat-labile *E. coli* proteins (*see Note 6*). Repeat once.
3. Use a cobalt metal resin column for IMAC purification of the (His)₆-tagged protein with an automated chromatography system. Install the column on the system and equilibrate it with buffer A at a flow rate of 1 mL/min. Load the protein extract on the column at a flow rate of 1 mL/min and start collecting 1 mL fractions.
4. Use a linear gradient of 5–100 % of buffer B over 20 column volumes to elute the recombinant protein from the column. The aRNase J protein is usually eluted in a single peak around ~10 % B (Fig. 2). If it is the case, then proceed directly

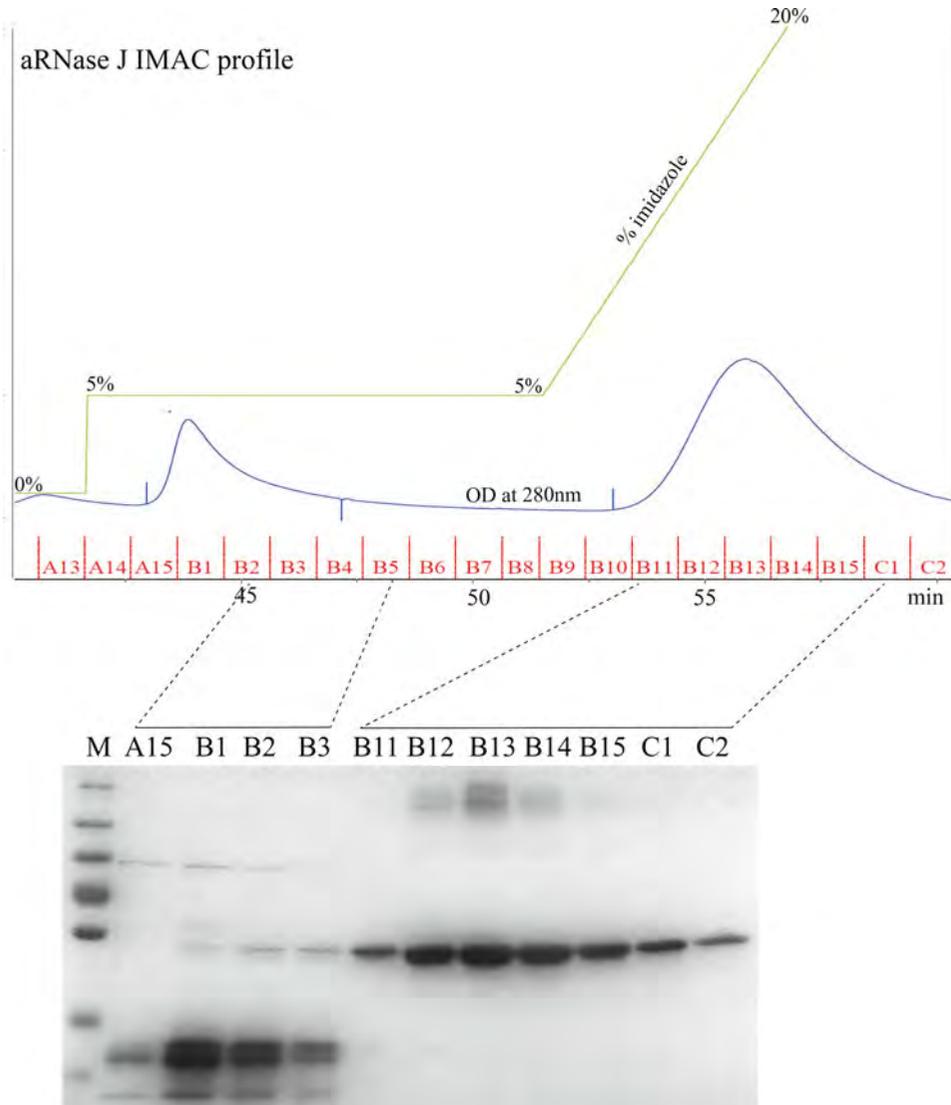


Fig. 2 IMAC purification chromatogram of the (His)₆-tagged aRNase J protein from heated *P. abyssi* cell extract. Collected fractions are loaded onto an 4–12 % SDS-PAGE gel to follow the elution of the tagged protein

to **step 6**. The aCPSF1 protein elutes along with contaminants within several peaks and requires an additional purification step (**step 5**).

5. Pool fractions corresponding to the aCPSF1-containing peaks. Dilute the sample with two volumes of buffer AQ and load onto the HiTrap Q FF column that has been equilibrated with buffer AQ at a flow rate of 1 mL/min. Elute the aCPSF1 protein with a linear gradient of 10–100 % of buffer BQ over 20 column volumes (usually around 20 %). Collect 1 mL fractions during the gradient.

6. Combine fractions corresponding to the most intense peaks of the chromatogram.
7. Dialyze proteins against storage buffer (*see Note 7*) and store at 4 °C at a concentration of ~0.25 to 0.5 µg/µL (concentrations can be estimated with a colorimetric assay such as Bradford assay using serial dilutions of BSA ranging from 0.2 to 10 µg/mL as standards). If possible, concentrate proteins to ~1 µg/µL with a centrifugal filtration device (*see Note 8*).
8. Remove the (His)₆-tag by treatment with thrombin (0.5 U/µL) for 2 h at room temperature (*see Note 9*). Then, incubate at 70 °C for 15 min to precipitate the heat-labile thrombin. Centrifuge to recover the supernatant.
9. Load an aliquot of the purified protein on a 4–15 % SDS-PAGE gel to check its purity. Estimate the concentration of protein and store it as described in **step 7**.

3.2 RNA Substrate Synthesis, Labeling, and Purification

Various labeled RNAs (Fig. 1) are used as substrates in RNA-degrading assays (*see Note 10*). DNA templates for transcription reactions are obtained by PCR amplification using two oligonucleotides encoding the T7 promoter sequence and corresponding RNA substrate sequence. Alternatively, DNA templates are obtained by hybridization of T7 promoter top strand oligonucleotide with the corresponding bottom strand oligonucleotide using a 2:1 ratio [16].

A synthetic RNA oligonucleotide substrate interrupted by a nine-atom spacer (9S) formed by three ethylene glycol units replacing the ninth nucleotide from the 5' end is used to identify the directionality of the exoribonucleolytic activity of aRNase J [3].

3.2.1 Synthesis of High Amount of 5' End Triphosphorylated and Monophosphorylated RNA Transcripts

1. To prepare high amount of 5' end triphosphorylated transcript, perform an in vitro transcription reaction with T7 RNA polymerase. Use PCR fragments as templates (300 ng/µL) in a 100 µL reaction volume in 1× transcription buffer containing 3.3 mM of each rNTP, 10 mM DTT, 8 mM MgCl₂, 20 U of RNasin inhibitor, and 80 U of T7 RNA polymerase.
2. Incubate transcription mixture at 37 °C for 180 min. Add 40 U of T7 polymerase and incubate at 37 °C for 180 min. Gel-purify the RNA transcripts as described in Subheading 3.2.6 (*see Note 11*).
3. To prepare 5' end monophosphorylated transcript, incubate 15 µg of triphosphorylated purified RNA (**step 2**) with 5 U of tobacco acid pyrophosphatase in 20 µL of 1× TAP buffer for 60 min at 37 °C.
4. Add 80 µL of RNase-free water and 100 µL of phenol/chloroform/isoamyl alcohol mixture and vortex for 30 s. Centrifuge at 20,000 ×g for 5 min at room temperature. Collect the upper

aqueous phase (100 μL) and precipitate by mixing with 1/10 volume of 3 M sodium acetate and 3 volumes of cold ethanol. Incubate at $-20\text{ }^{\circ}\text{C}$ for at least 2 h.

5. Centrifuge the sample for 30 min at $10,000\times g$ in a refrigerated centrifuge and discard supernatant. Add 3 volumes of 75 % chilled ethanol. Repeat the centrifugation step and dry the pellet in a Speedvac apparatus.

3.2.2 Preparation of Uniformly Labeled Transcripts (5' ppp RNA(U*))
(See **Note 12**)

1. Perform a transcription reaction with T7 RNA polymerase in a 20 μL volume reaction (*see* Subheading 3.2.1) with the following modifications. Add 3 μL of [α - ^{32}P] UTP (800 Ci/mmol) to the in vitro transcription mix containing 0.1 mM UTP and 2 mM each of ATP, CTP, and GTP.
2. Remove the excess of free [α - ^{32}P] UTP by passage of the sample through a Micro Bio-Spin P30 gel column following manufacturer's instructions.
3. Gel-purify the uniformly labeled RNA transcripts by proceeding as described in Subheading 3.2.6.

3.2.3 Preparation of 5' End-Labeled Triphosphosphate Transcripts RNA Substrates (p*pp RNA)
(See **Note 12**)

1. Perform a transcription reaction with T7 RNA polymerase in a 20 μL volume reaction (*see* Subheading 3.2.1) with the following modifications. Add 3 μL of [γ - ^{32}P] GTP (6,000 Ci/mmol) to the in vitro transcription mix containing GTP at a tenfold lower concentration (0.3 mM).
2. Proceed as described in **steps 2** and **3** of Subheading 3.2.2.

3.2.4 Preparation of 5' End-Labeled Monophosphate RNA Substrates (p*RNA)
(See **Note 12**)

1. Incubate 15 μg of RNA with 20 U of CIAP phosphatase in $1\times$ CIAP phosphatase buffer supplied with the enzyme for 45 min at $37\text{ }^{\circ}\text{C}$ to obtain dephosphorylated RNA.
2. Extract with phenol/chloroform/isoamyl alcohol mixture as described in Subheading 3.2.1, **step 4**.
3. Mix dephosphorylated RNA (from **step 2**) or corresponding amount of synthetic oligoribonucleotide with 10 U of T4 polynucleotide kinase and 2 μL [γ - ^{32}P] ATP (6,000 Ci/mmol) in $1\times$ T4 polynucleotide kinase buffer supplied with the enzyme. Incubate at $37\text{ }^{\circ}\text{C}$ for 90 min.
4. Proceed as described in **steps 2** and **3** of Subheading 3.2.2.

3.2.5 Preparation of 3' End-Labeled Triphosphosphate and Monophosphate Transcripts and Synthetic RNAs (pppRNAp*Cp/pRNAp*Cp)
(See **Note 12**)

1. For a 20 μL reaction volume, use 3 $\mu\text{g}/\mu\text{L}$ of triphosphorylated RNA (Subheading 3.2.1, **step 2**) or dephosphorylated RNA (Subheading 3.2.1, **step 5**) or 20 μM of synthetic oligoribonucleotide. Add 3 μL of [$5'$ - ^{32}P] pCp (3,000 Ci/mmol), 3 μL of DMSO, and 20 U of T4 RNA ligase in $1\times$ T4 RNA ligase buffer. Incubate at $37\text{ }^{\circ}\text{C}$ for 60 min.
2. Proceed as described in **steps 2** and **3** of Subheading 3.2.2.

3.2.6 Purification of RNA Substrates (See Note 13)

1. Purify each RNA substrate on a denaturing 8 % or 10 % or 20 % PAGE gel (see Note 3) depending on the length of RNA. Bands corresponding to RNA are revealed either by autoradiography (for labeled RNA) or UV shadowing (for unlabeled RNA) (see Note 14).
2. Excise the most intense band from the gel using a nuclease-free scalpel blade. Crush it and soak with 300 μ L to 1 mL (depending of the band size) of RNA elution buffer.
3. Incubate the gel slurry at 37 °C for 2 h or overnight at 4 °C.
4. Centrifuge sample for 30 min at 8,000 $\times g$ at 4 °C. Carefully recover the supernatant (without pipetting any gel slurry) and pass it through a disposable 0.45 μ m (pore size) filter unit.
5. Precipitate the resulting solution by adding 3 volumes of cold ethanol. Incubate at -20 °C for at least 2 h.
6. Centrifuge sample for 30 min at 20,000 $\times g$ at 4 °C. Discard the supernatant and wash the pellet with 300 μ L of 70 % ethanol. Dry pellet in a Speedvac apparatus and then resuspend it in RNase-free water.

3.3 RNA Degradation Assays

This section describes the characterization of ribonucleolytic activities of purified archaeal β -CASP proteins (*P. abyssi* aCPSF1 and aRNase J as examples) by kinetic experiments performed with a panel of RNA-labeled substrates under conditions where the enzyme (Subheading 3.3.1) or the substrate (Subheading 3.3.2) is in excess. The aRNase J specimen is a 5' to 3' exoribonuclease with preference for 5' monophosphorylated substrates, while aCPSF1 exhibits an endoribonucleolytic activity that preferentially cleaves single-stranded CA dinucleotides (Fig. 1) [3, 4]. Processivity of exonucleases (see Chapter 25 of this book) can be assessed in experiments with an excess of RNA substrates over enzymes (Subheading 3.3.2). In this condition, a highly processive 5'-to-3' exoribonuclease will produce only labeled mononucleotides from a 3'-end-labeled substrate. This is the case of aRNase J (Fig. 1) [3]. By contrast, a distributive exonuclease will yield RNA intermediate products.

3.3.1 Assay in Excess of Enzymes

Under this condition, the excess of enzyme should be at least fourfold.

1. For a typical 15- μ L assay, mix on ice 3 μ L of 5 \times NR buffer (see Note 4), 1 μ L of 32 P-RNA substrate (10,000 cpm/ μ L corresponding to ~5 fmol) (see Subheading 3.2 and Note 10), 10 μ L of ultrapure RNase-free water in a 0.5 mL microtube to minimize evaporation water. Add 1 μ L of purified β -CASP protein (final concentration from 10 nM to 6 μ M; see Note 15) on ice. Follow steps 2–5 to separate the reaction

products by denaturing gel electrophoresis (*see* **Note 3**) and **steps 6–8** to resolve nucleoside products by TLC.

2. For the zero time point, withdraw 4 μL from the mixture kept on ice and incubate immediately with 4 μL of proteinase K mix at 37 °C during 10 min. Vortex, centrifuge, and stop the reaction by adding 10 μL of FUD solution. Store the 18 μL mixture at -20 °C.
3. Start the reaction by incubating the rest of the mixture prepared in **step 1** at 65 °C in a heating block or thermocycler (*see* **Note 16**).
4. At each time point (from 0 to 90 min), centrifuge the reaction tubes. Withdraw a 4 μL aliquot and process it as described in **step 2**.
5. Heat the 18 μL mixtures prepared in **steps 2** and **4** at 95 °C for 1 min. Load 5 μL of each sample on a denaturing PAGE gel (*see* **Note 3**). Load in parallel RNase T1 and hydroxyl ladders of RNA substrate. Perform electrophoresis using a power of 35 W until the bromophenol blue reaches 5 cm from the bottom edge of the gel. Dry and expose the gel to a phosphorimager plate.
6. For the zero time point, withdraw 2 μL from mixture of **step 1** and spot directly on TLC plate at 2 cm from bottom.
7. Proceed to **step 3**. At each time point, take out 2 μL of reaction and spot directly on TLC plate as in **step 6**.
8. Develop TLC plates with 0.25 M of KH_2PO_4 . Dry the TLC plate and expose it to a phosphorimager plate during 12 h.
9. Following sample exposure (**step 5** or **8**), scan the plate with a phosphorimager and quantify gel band intensities using dedicated software.
10. Assess the appearance of degradation products from gel band signals obtained from at least three distinct experiments.

3.3.2 Assay in Excess of RNA Substrates

Under this condition, the protein/substrate ratio should be 1:20.

1. Mix on ice 3 μL of 5 \times NR buffer, 1 μL of ^{32}P -RNA substrate (10,000 cpm/ μL corresponding to ~5 fmol) (*see* Subheading 3.2 and **Note 10**), 1 μL of cold RNA substrate (20 pmol), and 9 μL of ultrapure RNase-free water in a 0.5 mL microtube to minimize evaporation of water. Add 1 μL of purified β -CASP protein (final concentration 50 nM corresponding to ~1 pmol) on ice.
2. Proceed as described in **steps 2–10** of Subheading 3.3.1.

4 Notes

1. To prepare plasmid vectors expressing β -CASP ribonucleases from *P. abyssi*, genes of interest (PAB1751 and PAB 1868) are amplified by PCR from *P. abyssi* genomic DNA using appropriate oligonucleotides and cloned as XhoI-BamHI PCR fragments in frame with the N-terminal 6xHis tag from the second to the stop codon into the pET15b plasmid vector. Alternatively, the pET21 plasmid (or other appropriate vector) may be used at this stage to clone in frame the open reading frame of the nuclease with a C-terminal 6xHis tag.
2. We use fast and reliable kits such as the QuikChange kit from Stratagene for site-directed mutagenesis (*see* also Chapter 25 of this book). Site-directed mutagenesis primers should be designed to introduce one or two mutations of the conserved amino acids of the catalytic site of the β -CASP enzymes.
3. The percentage of the acrylamide of the denaturing gel is selected according to the size of RNA substrates and reaction products.
4. The buffer reaction presented here is the most appropriate to test β -CASP ribonucleases activities in our hands. β -CASP ribonuclease activities do not require Mg^{2+} ions and are resistant to EDTA. Caution should be taken in analyzing RNA-degrading activities requiring divalent ions at high temperature. Indeed high concentration of Mg^{2+} ions will increase the *in vitro* chemical (nonenzymatic) hydrolysis of RNA at high temperature. Note also that the *P. abyssi* β -CASP ribonucleases are inactive at temperature below 50 °C in general and still active above 90 °C.
5. We prefer to divide the 800 mL in four flasks of 1 L of 200 mL culture. In our hands, this increases the expression of the recombinant protein of interest.
6. The *P. abyssi* recombinant proteins are thermo-resistant. The heating step at 70 °C is considered as the first step of purification and eliminates most of *E. coli* proteins.
7. We prefer to use a HEPES buffer because its pH changes less drastically at high temperature than a Tris buffer.
8. The aCPSF1 protein can be concentrated without any problems to high concentrations. However, aRNase J is more sensitive and can precipitate at high concentrations.
9. We prefer to remove the (His)₆-tag to perform enzymatic characterizations. Nevertheless, similar results are obtained in the presence of the tag proving that (His)₆-tag is stable at high temperature. Other tags (such as the calmodulin-binding peptide tag) can be sensitive to temperature and should not be

used for protein affinity purification and enzymatic assay at high temperature.

10. To probe RNA-degrading activities of *P. abyssi* β -CASP enzymes, we use single- and double-stranded RNA substrates. Their sequences correspond to the well-characterized 67 nt-*P. abyssi* C/D box guide sRNA sR47 [17], 149 nt-*P. abyssi* pre-tRNA^{trp} (with high content of RNA secondary structure), and 215 nt- *P. abyssi* noncoding RNA sRkB [18]. We also use a 21 nt synthetic oligonucleotide (corresponding to the first 21 nt of sR47 RNA) interrupted by a spacer formed by three ethylene glycol units [3] to stop the progression of the processive aRNase J exonuclease as well as circularized sR47 RNA (see Chapter 25 of this book for the preparation of circularized RNA).
11. The pyrophosphates produced during the transcription reaction precipitate. It may be necessary to add extra μL of EDTA (0.5 M) until the white precipitate disappears before proceeding to gel purification.
12. Caution should be taken in handling radiomaterials. Please refer to your radiation safety officer for local regulation and proper equipment.
13. Tutorials of RNA purification by preparative polyacrylamide gel electrophoresis is also described in ref [19] and in Chapter 18 of this book.
14. Stick fluorescent or radiolabeled markers to the gel which will subsequently aid in carefully positioning it over the autoradiography for detection and excision of gel bands.
15. The concentration of enzymes in the reaction mixture should be adapted for each β -CASP ribonuclease. Optimal concentration for aRNase J is 40 nM. In contrast for aCPSF1, high concentration of enzyme (6 μM) is required to observe RNA-degrading activities. This probably means that purified aCPSF1 is mostly in an inactive conformation.
16. A thermocycler with a heated lid is recommended to avoid evaporation at 65 °C during the kinetic reaction.

Acknowledgments

This work is supported by the Centre National de la Recherche Scientifique (CNRS) with additional funding from the Agence Nationale de la Recherche (ANR) [BLAN08-1_329396] and from the Université de Toulouse (UPS). D.K.P. is supported by a Ph.D. scholarship from the French “Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche.”

References

1. Stoecklin G, Muhlemann O (2013) RNA decay mechanisms: specificity through diversity. *Biochim Biophys Acta* 1829:487–490
2. Evguenieva-Hackenberg E, Klug G (2009) RNA degradation in Archaea and Gram-negative bacteria different from *Escherichia coli*. *Prog Mol Biol Transl Sci* 85:275–317
3. Clouet-d'Orval B, Rinaldi D, Quentin Y et al (2010) Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *J Biol Chem* 285:17574–17583
4. Phung DK, Rinaldi D, Langendijk-Genevaux PS et al (2013) Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* 41:1091–1103
5. Hasenohrl D, Konrat R, Blasi 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
6. Martens B, Amman F, Manoharadas S et al (2013) Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *PLoS One* 8:e76569
7. Callebaut I, Moshous D, Mornon JP et al (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* 30:3592–3601
8. Dominski Z (2007) Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* 42:67–93
9. Condon C, Gilet L (2011) The metallo-β-lactamase Family of Ribonucleases, vol 26, *Nucleic Acids and Molecular Biology*. Springer, Berlin Heidelberg, pp 245–267
10. Dominski Z, Carpousis AJ, 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
11. Li de la Sierra-Gallay I, Zig L, Jamalli A et al (2008) Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol* 15:206–212
12. Mandel CR, Kaneko S, Zhang H et al (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444:953–956
13. Even S, Pellegrini O, Zig L et al (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. *Nucleic Acids Res* 33:2141–2152
14. Cohen GN, Barbe V, Flament D et al (2003) An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* 47:1495–1512
15. Evguenieva-Hackenberg E, Wagner S, Klug G (2008) In vivo and in vitro studies of RNA degrading activities in Archaea. *Methods Enzymol* 447:381–416
16. Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* 180:51–62
17. Nolivos S, Carpousis AJ, Clouet-d'Orval B (2005) 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* 33:6507–6514
18. Phok K, Moisan A, Rinaldi D et al (2011) Identification of CRISPR and riboswitch related RNAs among novel non-coding RNAs of the euryarchaeon *Pyrococcus abyssi*. *BMC Genomics* 12:312
19. Petrov A, Wu T, Puglisi EV et al (2013) RNA purification by preparative polyacrylamide gel electrophoresis. *Methods Enzymol* 530:315–330

PARTIE II

LES RIBONUCLEASES β -CASP aCPSF1 ET aRNase J DE *P. abyssi* A L'INTERFACE DU METABOLISME DE L'ARN

II-1. IDENTIFICATION DES PARTENAIRES ET COMPLEXES MULTI-PROTEIQUES DES RIBONUCLEASES β -CASP de *P. abyssi*

Les ribonucléases β -CASP, aCPSF1 et aRNase J, sont respectivement ubiquitaires chez les archées et conservées dans la majorité des Euryarchées. Ces enzymes pourraient donc jouer un rôle clé dans le métabolisme cellulaire et en particulier dans le processing des ARN. Afin d'élucider leurs fonctions physiologiques dans le modèle *P. abyssi*, nous avons mené en parallèle deux types d'approches à partir d'extraits cellulaires obtenus à partir de cellules en phase exponentielle de croissance. La première approche a permis d'identifier les réseaux d'interaction protéine-protéine de aCPSF1 et aRNase J par des expériences de « Pull down » en utilisant les protéines recombinantes *Pab*-aCPSF1 et *Pab*-aRNase J. La deuxième approche a permis d'observer les profils de sédimentation des protéines endogènes aCPSF1, aRNase J et de leurs partenaires protéiques potentiels par des expériences de fractionnement cellulaire sur gradient de saccharose en utilisant des anticorps correspondants.

II-1-1. Réseaux d'interaction protéine-protéine des ribonucléases β -CASP aCPSF1, aRNase J et de l'hélicase ASH-Ski2 de *Pyrococcus abyssi* par la technique de « Pull-down »

a. Principe et Méthode de la technique de « Pull-down »

Afin d'identifier les réseaux d'interactions protéiques des ribonucléases β -CASP de *P. abyssi*, nous avons utilisé la méthode mise au point par l'équipe de D. Flament (Ifremer, Brest), qui a servi à l'identification des réseaux d'interaction des protéines impliquées dans la maintenance de l'ADN chez *P. abyssi* (Pluchon *et al.*, 2013). Cette technique, appelée génériquement « Pull-down », permet à partir d'une protéine appât recombinante purifiée et d'un extrait cellulaire de *P. abyssi*, d'identifier les partenaires protéiques potentiels et donc de définir son « réseau d'interaction protéine-protéine ou interactome ».

En collaboration avec l'équipe de D. Flament nous avons appliqué cette méthode à trois protéines-appâts correspondants aux deux ribonucléases β -CASP, *Pab*-aRNase J et *Pab*-aCPSF1 et l'hélicase *Pab*-ASH-Ski2 de *P. abyssi*, respectivement. Pour cela, les protéines-appâts, *Pab*-aRNase J et *Pab*-aCPSF1, possédant une étiquette (His)₆ placée à l'extrémité N et/ou C-terminal sont surexprimées chez *E. coli* et purifiées à homogénéité par chromatographie d'affinité. Chaque protéine purifiée est alors incubée avec des extraits cellulaires de *P. abyssi*. Les extraits cellulaires sont préparés à partir de culture en bioréacteur en condition anaérobie stricte et récoltée en phase exponentielle de croissance (95°C, pH 6.5) au laboratoire d'Ifremer (Pluchon *et al.*, 2013). Après incubation de la protéine-appât avec l'extrait cellulaire, une chromatographie d'affinité sur des billes magnétiques couplées au cobalt est réalisée pour pêcher les partenaires de la protéine appât. Afin d'identifier si les interactions détectées sont médiées par la présence des acides nucléiques, deux conditions, avec ou sans traitement aux nucléases (RNase A et DNase I), sont menées en parallèle au moment de l'incubation avec l'extrait. Des expériences témoins en absence de la protéine appât sont effectuées afin d'éliminer les interactants non spécifiques. Les protéines de l'extrait de *P. abyssi* co-purifiées avec les protéines recombinantes sont analysées par HPLC couplée à de la spectrométrie de masse. Deux plateformes de protéomique ont été sollicitées pour l'analyse de nos échantillons.

La première plateforme est celle de BioGenOuest (BGOP) qui avait permis à l'équipe de D. Flament à l'Ifremer, d'identifier les réseaux d'interaction de protéines de *P. abyssi* impliquées dans la maintenance de l'ADN (Pluchon *et al.*, 2013). Cette plateforme identifie les peptides du mélange sur la base de données protéique SEQUEST. Un indice de confiance, appelé valeur SR pour 'signal significance ratio' est alors calculé comme décrit dans Pluchon *et al.*, 2013. Ces indices correspondent au rapport entre le score Mascot des protéines de *P. abyssi* détectées dans l'expérience en présence et en absence de la protéine appât. De plus sont intégrés dans nos valeurs les expériences témoins issus de Pluchon *et al.*, 2013. Les valeurs SR inférieures à la valeur de 0.25 sont considérées comme reflétant une interaction significative entre la protéine appât et le candidat. La deuxième plateforme sollicitée pour des raisons de rapidité d'analyse, est celle d'Analyse Protéomique de Paris Sud Ouest (PAPPSO). L'abondance des spectres peptidiques MS/MS et les protéines sont identifiées dans la base de données X!Tandem. Les spectres MS/MS correspondent aux peptides retenus et re-fragmentés et analysés. Le nombre de spectre reflète l'abondance et la quantité de protéines. Un test de student est effectué afin de comparer les spectres présents dans les expériences témoins sans

appât et celles en présence d'appât. Les protéines candidates retenues sont celles présentant un test t inférieur à 0.05.

Les partenaires-candidats de la protéine *Pab*-(His)₆-aCPSF1, analysées par la plateforme BGOP sont classés dans la **Table 1**. Les partenaires-candidats de *Pab*-aRNase J étiquetées histidine en C-terminal (*Pab*-aRNase J-(His)₆), ainsi que *Pab*-ASH-Ski2 étiquetées en N- et C-terminale (*Pab*-(His)₆-ASH-Ski2, *Pab*-ASH-Ski2B-(His)₆) analysées par la plateforme PAPPSON sont classés dans les **Tables 2** et **3**. Il est à noter que lors de nos différents tests, nous avons constaté que la protéine *Pab*-aCPSF1 possède une affinité intrinsèque pour des billes de Nickel (**Table 1**). Elle ne présente donc pas de valeur SR significative dans les réseaux d'interaction quel que soit les protéines recombinantes servant d'appât.

L'ensemble des protéines identifiées de manière significative autour des différentes protéines appâts sont représentée sous forme de réseaux sur la **Figure 14** pour *Pab*-(His)₆-aCPSF1 et la **Figure 15** pour les protéines *Pab*-aRNase J-(His)₆, *Pab*-(His)₆-ASH-Ski2 et *Pab*-ASH-Ski2-(His)₆ avec les données issues respectivement des plateformes BGOP et PAPPSON. Les trois réseaux sont représentés ensemble sur la **Figure 17**.

b. Interactome de *Pab*-aCPSF1

Le réseau d'interaction a été obtenu uniquement avec la protéine appât en présence d'une étiquette en N-terminal, *Pab*-(His)₆-aCPSF1. Les expériences ont été effectuées en triplicata. Cet interactome est constitué de 13 partenaires ayant un indice de confiance « SR » inférieur à 0,25 (**Table 1** et **Figure 14**). De manière remarquable, on retrouve au sein de cet interactome la ribonucléase *Pab*-aRNase J. L'indice de confiance de cette interaction est légèrement plus faible en présence de nucléases. Ceci permet de proposer que les deux ribonucléases β-CASP de *P. abyssi*, *Pab*-aCPSF1 et *Pab*-aRNase J pourraient agir conjointement dans des mêmes voies métaboliques. De plus, on observe parmi les partenaires la présence des sous-unités A' (RpoA1) et A'' (RpoA2) de la RNA polymérase ainsi que les sous-unités RPA32 et RPA41 de la réplication A protéine dont la propriété principale est le maintien de l'ADN sous forme simple brin par des liaisons directes (Kur *et al.*, 2005). Le recrutement des sous-unités RPA sur l'ADN se font lors des événements de réplication et de réparation de l'ADN ainsi que de la transcription des ARN. Récemment, le complexe RPA a été montré capable d'aussi bien se lier à de l'ADN qu'à de l'ARN chez *S. solfataricus* (Morten *et al.*, 2017). Les indices de confiance pour ces partenaires ne semblent pas être affectés par la présence de nucléases. On

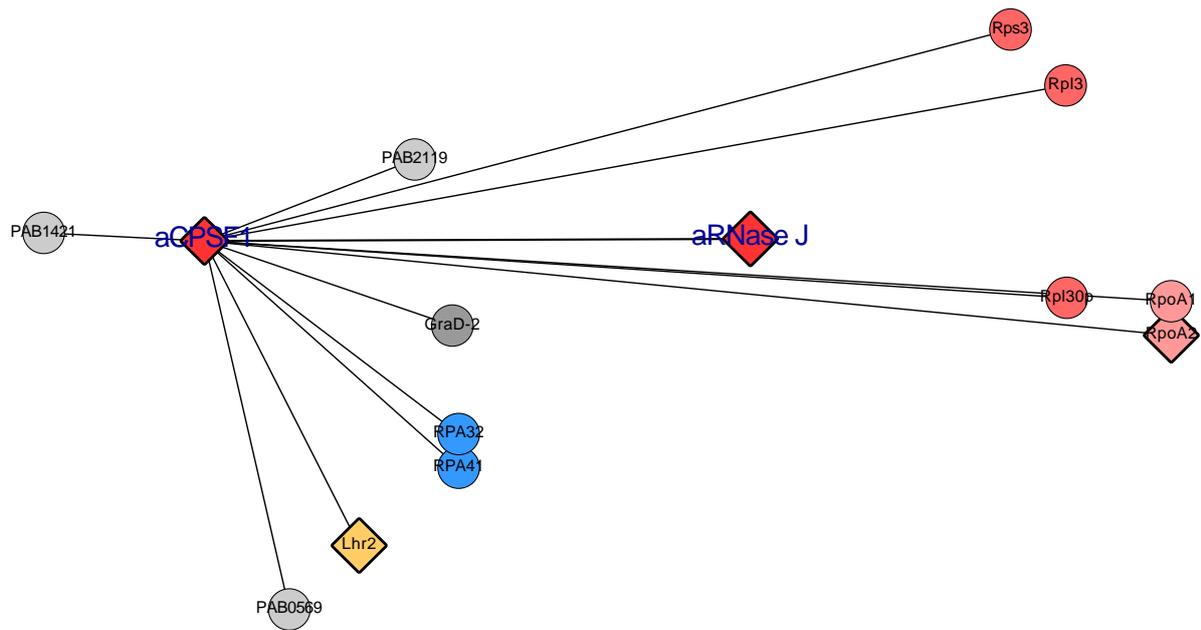


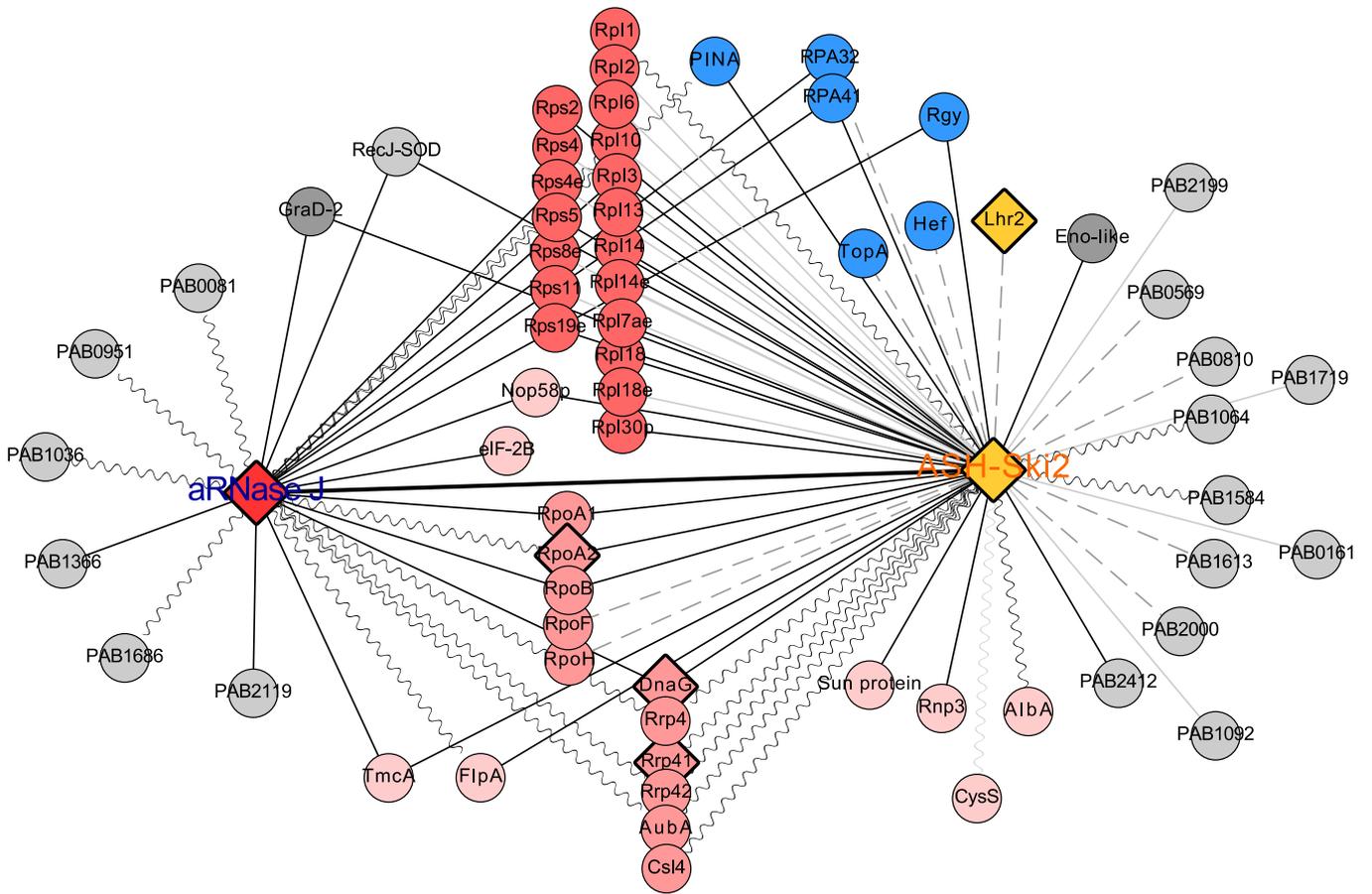
Figure 14: Réseaux d'interaction autour de la protéine aCPSF1 à partir des données de la plateforme BGOP. Protéines présentes dans les réseaux d'interaction obtenus par « pull-down » d'extraits brut de *P. abyssi* avec des protéines étiquetées appâts *Pab*-(his)₆-aCPSF1. Les réseaux de protéines sont analysés par spectrométrie de masse. Seules les protéines ayant une valeur SR inférieure à 0.25, un nombre de peptide au dessus de 5 et ne présentant pas de sensibilité aux nucléases sont montrées. En rouge sont indiquées les protéines impliquées dans le métabolisme de l'ARN, en bleu celui de l'ADN et en gris les protéines aux fonction non définie. L'hélicase Lhr2 est représenté en orange. Les traits droits indique des interaction non sensible aux nucléases.

peut donc proposer dans un premier temps que *Pab*-aCPSF1 serait en réseau au sein de la cellule au niveau de la fourche de transcription et avec des facteurs impliqués dans le métabolisme de l'ARN. Cette hypothèse est aussi cohérente avec la détection de protéines ribosomiques L30p et L3, et S3. Au sein de ce réseau, sont aussi retrouvées les protéines PAB2119 annotée comme étant une méthyl transférase des ARNt et GraD-2 annotée Glucose-1-phosphate thymidyltransferase (**Figure 14**).

On peut noter que l'indice de confiance le plus fort (SR=0) dans le réseau de *Pab*-(His)₆-aCPSF1 est celui de la protéine annotée Lhr2, « Large helicase related ». Cet indice ne varie pas en présence ou absence de nucléases. La protéine Lhr2, de fonction inconnue à ce jour, appartient à la superfamille des hélicases SF2 (voir étude phylogénomique–collaboration Fichant et article de Chamieh *et al*, 2016). Nous avons initié la caractérisation de cette protéine pour essayer de comprendre sa présence dans le réseau de *Pab*-aCPSF1 mais des premiers essais d'expression de la protéine recombinante *Pab*-Lhr2 au cours de mon stage de Master 2 se sont avérés infructueux. En effet, nous avons obtenu des difficultés de clonage et d'expression de la protéine pleine taille qui s'est avérée être toxique pour *E.coli*. Nous avons pu cependant cloner la partie du gène correspondant au core SF2 d'une part et à la partie C-terminal d'autre part. Ce qui nous a permis de produire et purifier ces fragments de protéines Lhr2 et de produire des anticorps polyclonaux de lapin (Eurogentec). La caractérisation enzymatique du core SF2 de la protéine a été initiée mais ne sera pas exposé dans ce travail de thèse.

Pour finir, nous avons ainsi identifié un candidat, PAB0569, annoté 3-isopropylmalate déshydratase possédant un Pfam de la famille des PAC2 (Proteasome Assembly Chaperone 2) qui regroupe des chaperonnes permettant l'assemblage du complexe protéasome. De manière intéressante, une interaction entre cette protéine est à corrélérer avec les données récemment publiés de l'équipe de Maupin-Fulow montrant un lien entre aCPSF1 et le complexe protéasome halophile (Chavarria *et al.*, 2014). En effet, dans cette étude chez *Haloferax volcanii*, aCPSF1 ainsi que la protéine annotée Ubiquitin activating E1, une protéine Ubiquitin-fold et une sous-unité homologue de Rpt1-6 du protéasome 26S ont pu être identifiées dans le réseau d'interaction de la protéine NcsA, impliquée dans la thiolation des ARNt (Chavarria *et al.*, 2014).

A



B

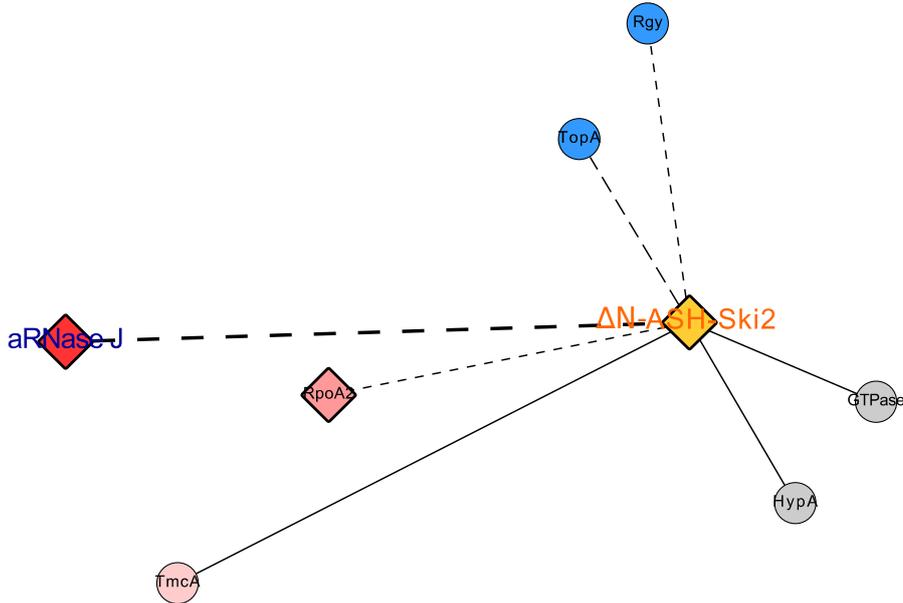


Figure 15: Réseaux d'interaction autour des protéines aRNase J et ASH-Ski2 à partir des données de la plateforme PAPPSO.

A : Protéines présentes dans les réseaux d'interaction obtenus par « pull-down » d'extraits brut de *P. abyssi* avec des protéines étiquetées appâts *Pab*-aRNase J-(His)₆, *Pab*-(His)₆-ASH-Ski2 et *Pab*-ASH-Ski2-(His)₆. B : Réseau obtenu avec la protéine étiquetée appâts *Pab*-(His)₆-ΔN-ASH-Ski2. Les réseaux de protéines sont analysés par spectrométrie de masse. Seules les protéines sensibles ou non aux nucléases et ayant une valeur inférieure à 0.05 au test de *student* sont montrées. En rouge sont indiquées les protéines impliquées dans le métabolisme de l'ARN, en bleu celui de l'ADN et en gris les protéines aux fonction non définie. Les protéines représentées par un losange sont celles dont nous possédons des anticorps spécifiques. Les traits droits indiquent des interactions non sensibles aux nucléases et ceux en forme de vague sont sensibles aux nucléases. ASH-Ski2 et Lhr2 sont indiquées en orange.

c. Interactome de *Pab*-aRNase J

L'interactome de *Pab*-aRNase J a été obtenu en effectuant des Pull-Down à partir de la protéine étiquetée en C-terminal (*Pab*-aRNase J-(His)₆). L'analyse protéomique, menée en duplicata, a été confiée à la plateforme PAPPSO (**Table 2**). Le réseau d'interaction de *Pab*-aRNase J issu de ces analyses est reporté dans la **Figure 15A** et inclut 17 partenaires identifiés de manière significative dont l'interaction ne semble pas être affectée par le traitement aux nucléases (**Table 2**). Parmi les membres du réseau, nous avons identifié une majorité de protéines annotées comme acteurs du métabolisme de l'ARN (**Figure 16A**). En effet l'interaction de *Pab*-aRNase J avec les protéines du ribosome, Rpl3 et Rps4e (résistante au traitement nucléase) et Rpl10 (sensibles aux nucléases) est observée (**Table 2**). Bien que les protéines composant les sous-unités des ribosomes soient majoritaires dans la cellule, elles ne sont pas retrouvées dans les interactomes de Pluchon et al. La présence de ces protéines est donc bien spécifique à l'interactome de aRNase J. Nous identifions également l'ARN polymérase les sous-unités RpoA1, RpoB et la sous-unité RpoA2 ainsi que le facteur d'initiation de la transcription eIF-2B (Kyrpides and Woese, 1998). De plus, la protéine DnaG est aussi présente dans le réseau ainsi que les autres protéines de l'exosome (Rrp4, Rrp42 et AubA) (Hou *et al.*, 2014). Cependant les interactions avec les composants de l'exosome semblent sensibles au traitement nucléases. La présence de protéines de modification des ARN telles que TmcA impliquée dans l'acétylation des ARNt chez les bactéries et les eucaryotes (Ikeuchi *et al.*, 2008; Sharma *et al.*, 2015), la protéine Nop5p du complexe RNP de méthylation des ARNt et ARNr, la fibrillarine (FlpA) impliquée dans les premières étapes de maturation des pré-ribosome et la stabilité des ribosomes (Omer *et al.*, 2002; Bortolin *et al.*, 2003; Rodriguez-Corona *et al.*, 2015) et la protéine méthyl transférase putative des ARNt (PAB2119) authentifie la surreprésentation de protéines impliquées dans le métabolisme des ARN (**Figure 16A**). Au sein de ce réseau, sont également présentes certaines protéines identifiées comme acteurs du métabolisme de l'ADN. C'est le cas des deux sous-unités du complexe RPA, RPA32 et RPA41 (Komori and Ishino, 2001), de la reverse gyrase (Forterre, 2002) ainsi que la protéine PINA, caractérisée récemment dans la résolution des jonctions de Holliday chez *Sulfolobus islandicus* (Zhai *et al.*, 2017). De plus, nous identifions deux protéines présentant des motifs DHH, PAB1366 et PAB1284 (annotée RecJ phosphoesterase) caractéristiques des protéines pouvant intervenir sur les acides nucléiques (Aravind and Koonin, 1998). En plus du motif DHH, PAB1284 présente deux domaines caractéristiques d'interaction avec les acides nucléiques, S1 et OB fold. Nous l'avons nommée RecJ-SOD

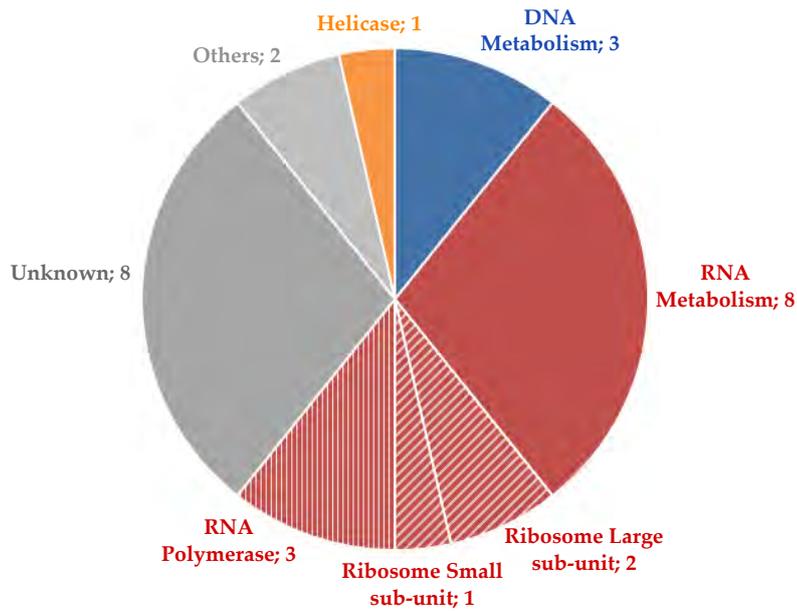
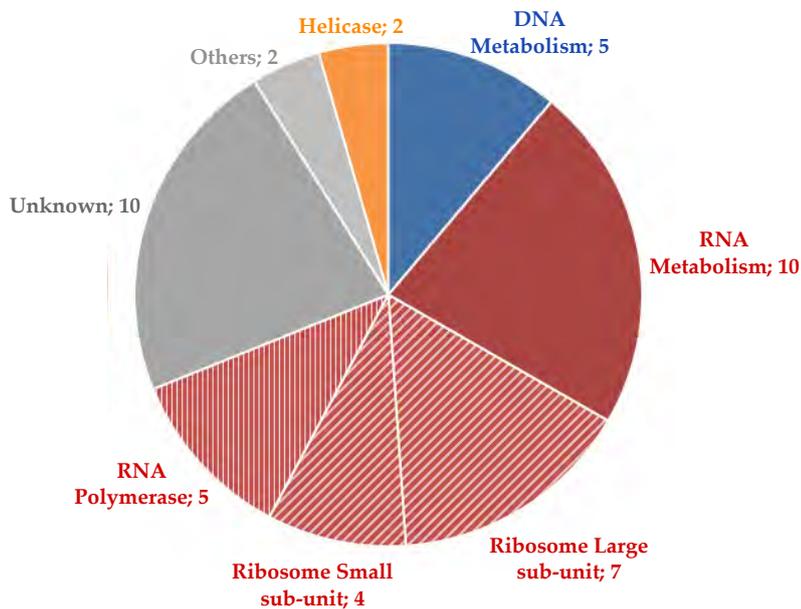
A**aRNaseJ****B****ASH-Ski2**

Figure 16: Rôles des protéines identifiées dans les réseaux d'interaction de aRNase J et ASH-Ski2. Distributions des protéines selon leurs métabolisme ADN, ARN, ou d'autre métabolisme. En rouge sont indiquées les protéines impliquées dans le métabolisme de l'ARN, en bleu celui de l'ADN, en gris sombre les protéines aux fonction non définie et en gris clair d'autre métabolisme que celui des acides nucléiques. A: Distribution des protéines identifiées dans le réseaux d'interaction de aRNase J. B: Distribution des protéines identifiées dans le réseaux d'interaction de ASH-Ski2.

pour la différencier des homologues de la RecJ bactérienne identifiées récemment comme pouvant dégrader à la fois de l'ADN et l'ARN simple brin (Ogino *et al.*, 2017). Pour finir, un des partenaires du réseau de *Pab*-aRNase J qui a retenu notre attention est la protéine annotée comme RNA hélicase que nous avons nommé *Pab*-ASH-Ski2 d'après les études phylogénomique et Chamieh *et al.*, 2016. Les caractéristiques de *Pab*-ASH-Ski2 permettent de l'identifier comme appartenant à la famille des hélicase Ski2, une des familles d'hélicases de la Superfamille SF2. *Pab*-ASH-Ski2 est caractérisée par un domaine spécifique en N-terminal conservé (Voir Chapitre 2-Partie 1).

d. Interactome de *Pab*-ASH-Ski2

A la suite de l'identification de *Pab*-ASH-Ski2 dans le réseau d'interaction de *Pab*-aRNase J, Nous avons décidé de déterminer le réseau de protéines associées à l'hélicase putative *Pab*-ASH-Ski2. Nous avons analysé l'interactome à partir des 2 protéines appâts, l'une étiquetée en N-*(Pab*-(His)₆-ASH-Ski2) et l'autre en C-terminal (*Pab*-ASH-Ski2-(His)₆). Chaque Pull-down a été effectué comme pour *Pab*-aRNase J en duplicata et les interactants ont été identifiés avec la plateforme d'analyse protéomique PAPPISO et sont présentés respectivement dans les **Tables 3A** et **3B**. Les réseaux d'interaction des deux protéines sont représentés dans la **Figure 15A** et **Figure 16B**. Parmi les nombreux interactants identifiés avec la protéine *Pab*-ASH-Ski2 étiquetée en N- ou en C-terminal, nous avons pu constater la présence de l'appareil de traduction avec les protéines ribosomiques Rpl7ae, Rpl3, Rpl13, Rpl18, Rpl19e, Rpl30, Rps2, Rps4 et Rps5. Nous identifions également les sous-unités de la RPA polymérase, RpoA1, RpoA2 et RpoB. Les protéines impliquées dans la dégradation des ARN avec les différentes sous-unités de l'exosome (DnaG, Rrp4, Rrp41, Rrp42, Csl4 et AubA) et la sous-unité 3 de la RNase P impliquée dans la maturation des ARNt sont aussi présentes dans le réseau d'interaction. Nous retrouvons de plus des acteurs impliqués dans la modification des ARN avec les protéines TmcA, FlpA, Nop5p, Sun protein et AlbA impliquée dans la stabilité des ARNr. Concernant le métabolisme de l'ADN, nous retrouvons les sous-unités du complexe RPA (RPA32 et RPA41), la protéine PINA et la reverse gyrase. La protéine RecJ-SOD est aussi retrouvée dans le réseau de *Pab*-ASH-Ski2. Parmi l'ensemble de ces protéines et de manière réciproque, nous retrouvons aussi *Pab*-aRNase J dans le réseau d'interaction de *Pab*-ASH-Ski2 confirmant les résultats obtenus avec le réseau de *Pab*-aRNase J (**Figure 15A**).

De manière intéressante, nous constatons la présence spécifique de certaines protéines en fonction de l'emplacement de l'étiquette histidine (N- ou C-terminal). En effet, les protéines telles que la topoisomerase A, Hef, Lhr2, les sous-unités RpoH et RpoF de la RNA pol ainsi que la sous-unité RPA32 sont seulement retrouvées dans le réseau de *Pab*-(His)₆-ASH-Ski2 (**Table 3A**). Pour *Pab*-ASH-Ski2-(His)₆, les partenaires sont plus nombreux avec la présence de nombreuses protéines du ribosome ainsi que des protéines aux fonctions inconnues telles que la cystéine tRNA ligase (CysS) (**Table 3B**). Ainsi la position de l'étiquette histidine en N- ou en C-terminal sur la protéine imposerait une contrainte spatiale sur l'interaction avec certains partenaires.

La protéine hélicase ASH-Ski2 présente dans la majorité des Euryarchées appartient à la famille des hélicase Ski2 de la superfamille SF2. Elle est caractérisée par un domaine en N-terminal spécifique qui contient un motif en zinc finger avec 4 cystéines conservées (**Figure 26B**). Nous avons réalisé des expériences de Pull Down avec pour appât un variant de la protéine *Pab*-ASH-Ski2 invalidé de son domaine N-terminal afin de tester l'importance de ce domaine dans la formation du réseau de *Pab*-ASH-Ski2. Pour cela, la protéine *Pab*- Δ N-ASH-Ski2 étiquetée en N-terminal surexprimée dans *E. coli* à partir du vecteur pET15b-*Pab*- Δ N-ASH-Ski2 est immobilisée sur la colonne d'affinité His-trap et incubée avec des extraits cellulaires de *P. abyssi*. Après lavage, les échantillons élués ont été analysés par la plateforme PAPPSO et les résultats sont présentés dans la **Table 3C**. De manière intéressante, peu de partenaires sont identifiées dans le contexte de ce mutant de *Pab*-ASH-Ski2 (**Figure 15B**). La protéine TmcA, identifiée dans le réseau de *Pab*-ASH-Ski2 sauvage, est significativement retenue en présence et absence de nucléases ainsi que deux autres candidats, PAB0399 et PAB2398, respectivement annotées sur Uniprot comme étant une hydrogénase et une GTPase, non identifiés dans le pull down de *Pab*-ASH-Ski2 sauvage. De plus, certains partenaires de la version sauvage de *Pab*-ASH-Ski2 sont devenus sensible aux nucléases tels que la reverse gyrase, la Topoisomerase I, la aRNase J et la sous-unité RpoA2. Il semblerait donc que le domaine spécifique N-ter de *Pab*-ASH-Ski2 aurait un rôle prépondérant dans l'interaction de *Pab*-ASH-Ski2 avec les partenaires du réseau. Certains résidus ou régions de ce domaine pourraient établir ou stabiliser des interactions protéine-protéine.

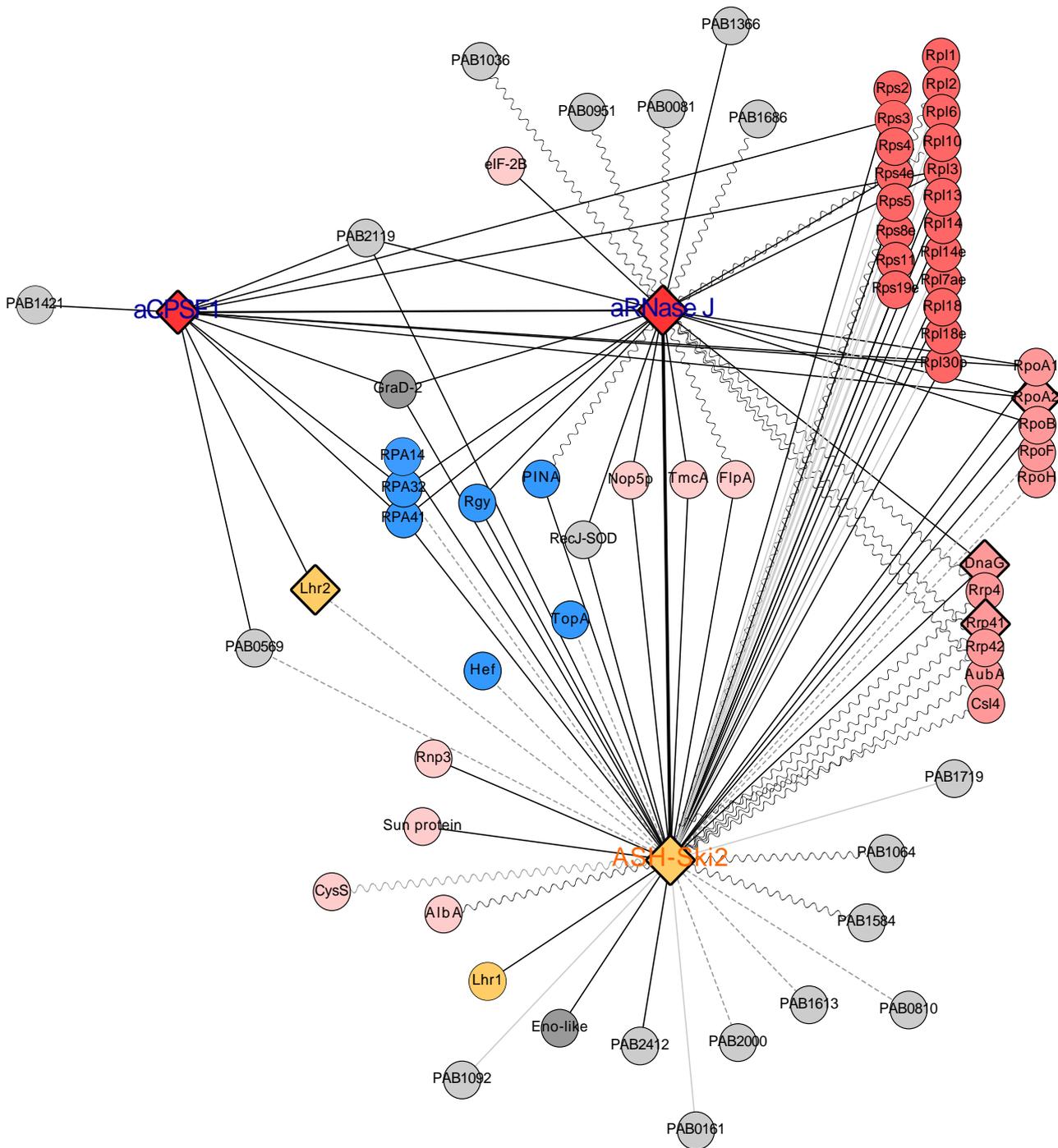


Figure 17: Réseaux d'interaction autour des protéines aRNase J et ASH-Ski2 et aCPSF1 à partir des données de la plateforme PAPPSSO et BGOP. Résumé de l'ensemble des protéines identifiées dans les trois réseaux d'interaction avec des protéines étiquetées appâts *Pab*-aRNase J-(His)₆, *Pab*-(His)₆-ASH-Ski2, *Pab*-ASH-Ski2-(His)₆ et *Pab*-(His)₆-aCPSF1. En rouge sont indiquées les protéines impliquées dans le métabolisme de l'ARN, en bleu celui de l'ADN et en gris les protéines aux fonction non définie. Les protéines représentées par un losange sont celles dont nous possédons des anticorps spécifiques. Les traits droits indique des interaction non sensible aux nucléases et ceux en forme de vague sont sensible aux nucléases. ASH-Ski2 et Lhr2 sont indiquées en orange.

e. Validation de la spécificité des partenaires identifiés dans les réseaux de *Pab*-ASH-Ski2, *Pab*-aRNase J et *Pab*-aCPSF1

Le réseau d'interaction protéine-protéine globale défini autour des acteurs de la maintenance de l'ADN déterminé chez *P. abyssi* avec 23 protéines appâts décrite dans Pluchon et al (Introduction, A-1) nous a permis de valider la spécificité des interactions observées dans les réseaux de *Pab*-aCPSF1, *Pab*-aRNase J et de *Pab*-ASH-Ski2 (Pluchon et al., 2013). En effet, ces expériences de Pull-down ont été effectuées dans les mêmes conditions avec des extraits cellulaires obtenus à partir de même culture de *P. abyssi* en phase exponentielle de croissance lysée dans un tampon PBS supplémenté avec 300mM de NaCl afin de se rapprocher le plus des conditions biologiques et d'éviter des interactions non spécifiques. La **Table 4** résume les quelques partenaires communs entre ces deux lots d'expériences. La protéine *Pab*-ASH-Ski2 est retrouvée uniquement dans l'interactome de la *Pab*-RNase HII impliquée dans la dégradation des amorces ARN servant à la synthèse des fragments d'Okazaki (**Figure 2**). La ribonucléase *Pab*-aRNase J est détectée respectivement dans les interactomes de *Pab*-NucS, de *Pab*-AlkA 3-méthyladenine DNA glycosylase et de *Pab*-RadA. La protéine *Pab*-RecJ-SOD est présente uniquement dans le réseau de l'hélicase *Pab*-DNA2. En ce qui concerne les sous-unités de la *Pab*-RNAP, l'hélicase *Pab*-Lhr2 et la reverse gyrase, ces partenaires se retrouvent de manière récurrente dans les réseaux d'interactions de la plupart des protéines impliquées dans le métabolisme de l'ADN. Cependant *Pab*-aCPSF1 n'est retrouvée dans aucun des réseaux issus des protéines utilisées par Pluchon et al. L'absence de *Pab*-aCPSF1 dans ces interactomes, ainsi que dans ceux de aRNase J et de ASH-Ski2 pourrait être expliquée par l'affinité intrinsèque de aCPSF1 aux billes de cobalt. De ce fait, l'interaction avec aCPSF1 est systématiquement considérée comme non significative par rapport au témoin et donc écartée des listes des protéines établissant une interaction significative avec la protéine appât.

f. Conclusions sur les réseaux d'interaction protéine-protéine des ribonucléases β -CASP et des hélicases like ASH-Ski2 et Lhr2

De manière remarquable, nos résultats montrent que de nombreux candidats appartiennent simultanément aux réseaux de *Pab*-ASH-Ski2, *Pab*-aRNase J et *Pab*-aCPSF1. Ainsi nous pouvons définir un réseau commun (**Figure 17**). Le résultat majeur, marquant et original, est l'identification pour la première fois d'une interaction entre les ribonucléases β -CASP d'archées et des hélicases putatives que nous avons nommées aLhr2 et ASH-Ski2 suite à nos

études phylogénomiques en collaboration avec l'équipe de G. Fichant (LMGM, Toulouse). Ainsi, *Pab*-aRNase J est retrouvée dans le réseau de *Pab*-ASH-Ski2 et réciproquement. De plus, toutes les protéines identifiées dans les réseaux communs de *Pab*-aCPSF1 et de *Pab*-aRNase J sont retrouvées dans celui de *Pab*-ASH-Ski2. Enfin, l'hélicase *Pab*-Lhr2 en lien avec *Pab*-aCPSF1 est identifiée dans le réseau de *Pab*-ASH-Ski2. L'ensemble de ces résultats permet de proposer une interconnexion fonctionnelle entre les 2 ribonucléases β -CASP et de l'hélicase *Pab*-ASH-Ski2 (**Figure 17**). Au vu de la présence systématique des protéines intervenant dans le métabolisme de l'ARN dans les réseaux de *Pab*-aCPSF1, *Pab*-aRNase J et *Pab*-ASH-Ski2, il est possible que ces acteurs aient un rôle conjoint, concerté ou couplé dans les différentes voies métaboliques des ARN (**Figure 16**). En particulier, la présence significative des protéines constituant la grande sous-unité du ribosome et les sous-unités RpoA1, RpoA2, RpoB permet de proposer que *Pab*-aCPSF1, *Pab*-aRNase J et *Pab*-ASH-Ski2 sont à l'interface des machineries de transcription et de traduction. Le lien avec la transcription est confortée par la présence des deux sous-unités du complexe RPA, protéines de recouvrement de l'ADN simple brin recrutées lors de la réplication et/ou de la transcription ainsi que des protéines annotées comme des Methyl-transférase like et Glucose-1-phosphate thymidyltranférase. De plus, le lien avec la traduction en est renforcé avec la présence au sein du réseau commun des enzymes de modification des ARNr et ARNt telle que les methyltransferases FlpA et TmcA. Un lien récurrent avec l'exosome, complexe de dégradation exorinucléolytique de 3'-5' des ARN, est observé. En effet, les composants du core catalytique de l'exosome, Rrp41 et Rrp42, et le composant de la coiffe Csl4 sont détectés dans les réseaux obtenus en absence de nucléases signifiant que ce lien serait médié par des acides nucléiques. De plus, DnaG, autre composant de la coiffe de l'exosome, est également détecté dans les résultats de protéomique de la plateforme (PAPPSO) (**Tables 3A et 3B**).

Pour finir, le partenaire PAB1284, une protéine que nous avons nommé RecJ-SOD, a retenu notre intérêt puisqu'elle est systématiquement identifiée dans nos analyses comme étant un partenaire au centre du réseau commun (**Figure 15A**). PAB1284 est caractérisée par un domaine catalytique DHH (retrouvé chez les protéines RecJ bactériennes) associé à un domaine Zinc-finger, un domaine S1 et un domaine OB-fold suggérant une capacité de reconnaissance de molécules d'ARN. Toutefois, la fonction et l'activité enzymatique de RecJ-SOD restent à être caractérisées.

II-1-2. Identification de complexes multi-protéiques par fractionnement d'extraits cellulaires de *P. abyssi* sur gradient de sucrose

Afin d'obtenir des évidences expérimentales sur l'existence de complexes multi-protéiques incluant des ribonucléases β -CASP, nous avons effectué des fractionnements d'extrait cellulaire de *P. abyssi* sur gradient de sucrose et visualisé la présence des ribonucléases β -CASP aRNase J et aCPSF1 et de certains de leurs partenaires identifiés dans les réseaux d'interaction décrits ci-dessus. Afin de visualiser les protéines endogènes de *P. abyssi*, nous avons utilisé des anticorps de lapin (Eurogentec) dirigés contre les protéines recombinantes de *Pab*-aCPSF1, *Pab*-aRNase J, *Pab*-ASH-Ski2, *Pab*-aLhr2-SF2, *Pab*-Rrp41 et *Pab*-DnaG et utilisé les anticorps de lapin dirigés contre les sous-unités RpoD et RpoA'' de *P. furiosus* (généreusement donnés par Dr. W. Hausner, Université de Regensburg, Allemagne).

a. Méthode

La préparation des extraits cellulaires de *P. abyssi* a été identique à celle effectuée dans les expériences décrites précédemment de Pull Down. Des culots de cellules ont été repris soit dans du PBS1X (137mM NaCl, 2.7 mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄) (Tampon PBS), soit dans un tampon contenant 20mM Tris-HCl, 100mM de KCl, 10mM MgCl₂ et 1mM DTT (tampon TK). Après la lyse des cellules par sonication, les extraits clarifiés sont déposés sur un gradient linéaire de sucrose 5-25% ou 10-30% et séparés par ultracentrifugation à 35 000 rpm à 4°C respectivement soit pendant 20h ou 5h. Les composants cellulaires sont ainsi séparés selon leur vitesse de sédimentation. Des fractions de 0,5 mL sont récoltées. Les protéines des fractions sont précipitées au TCA et visualisées par Western Blot. Les ARN des fractions sont extraits au Trizol puis précipités. Différents fractionnements cellulaires (**Figures 18 à 21**) ont permis de localiser nos protéines d'intérêt et les ARN ribosomiques. De plus nous avons déterminé leur quantité relative au sein de chaque fraction.

b. Fractionnement cellulaire sur gradient de sucrose 5-25% dans le Tampon PBS

L'étude des profils de sédimentation des ribonucléases β -CASP d'un extrait de *P. abyssi* a été initiée sur un gradient de sucrose 5-25% (Pourcentage généralement utilisé pour identifier le profil de sédimentation des sous-unités de l'exosome ou des complexes de poids moléculaire intermédiaire (Evguenieva-Hackenberg *et al.*, 2003) dans un tampon PBS identique aux expériences de « pull-down » (voir Partie II-1-1). Ces premières expériences permettent

d'observer que aCPSF1 endogène est essentiellement retrouvée sous sa forme dimérique (fractions 8 à 12 dans l'extrait) dans ces conditions d'expérience (**Figure 18A, panel haut**). A l'opposé, aRNase J endogène sédimente dans des fractions de haut poids moléculaire (fraction 20 et 21). Avec l'extraction des ARN des différentes fractions séparés sur gel agarose 1% et colorés au SybrGreen, ces fractions 20 et 21 correspondent à la sédimentation des sous-unités du ribosomique 30S et 50S et des ribosomes/polysomes (**Figure 18A, panel milieu**). De manière remarquable, la protéine ASH-Ski2 endogène se retrouve en partie dans ces mêmes fractions 20 et 21 bien que la majorité se retrouve dans les fractions 7 et 8 correspondant à la sédimentation de protéine d'environ 100kDa déterminé à partir des protéines étalons et de la protéine *Pab*-ASH-Ski2 recombinante (fractions 7 à 8-**Figure 18B**). On observe donc une co-sédimentation (fractions 20 et 21) de la ribonucléase aRNase J et du ribosome ainsi que d'une fraction de ASH-Ski2 endogène. Lorsque la concentration de sel est augmentée à 500 mM, ASH-Ski2 semble se dissocier des complexes de haut poids moléculaire pour se retrouver exclusivement sous forme monomérique contrairement à aRNase J qui reste associé dans des fractions 18 et 19 (**Figure 18A, panel bas**). Dans les deux cas en présence de 500mM NaCl, les fractions semblent décalées vers le haut du gradient. Ce décalage pourrait être expliqué par une plus grande densité du gradient dû à la présence de 500mM de NaCl. On peut noter que la protéine recombinante *Pab*-aRNase J sédimente de manière diffuse formant certainement des oligomères avec un pic dans les fractions 11 et 12 correspondant à un poids moléculaire approximatif de 200 kD qui pourraient correspondre à des formes trimère ou tétramère au vu du poids moléculaire théorique de aRNase J de 50 kD (**Figure 18B**). Cette observation a également été faite sur des expériences de gel filtration S-200 (Donnée non montré).

En conclusion, dans ces conditions expérimentales, les deux ribonucléases β -CASP, aCPSF1 et aRNase J ne semblent pas associées aux mêmes fractions de sédimentation. De manière intéressante, la ribonucléase aRNase J est associée aux fractions de haut poids moléculaire contenant l'hélicase ASH-Ski2 et le ribosome. Cependant la résolution des fractions de haut poids moléculaire dans un gradient de sucrose 5-25% est limitée.

c. Fractionnement cellulaire sur gradient de sucrose 10-30% dans le Tampon PBS

Afin d'augmenter la résolution des fractions de très haut poids moléculaire qui contiennent les sous-unités du ribosome, nous avons effectué des gradients de sucrose 10-30 % tout d'abord dans un tampon PBS (**Figure 19**). Les profils de sédimentation de aRNase J, ASH-Ski2 et des

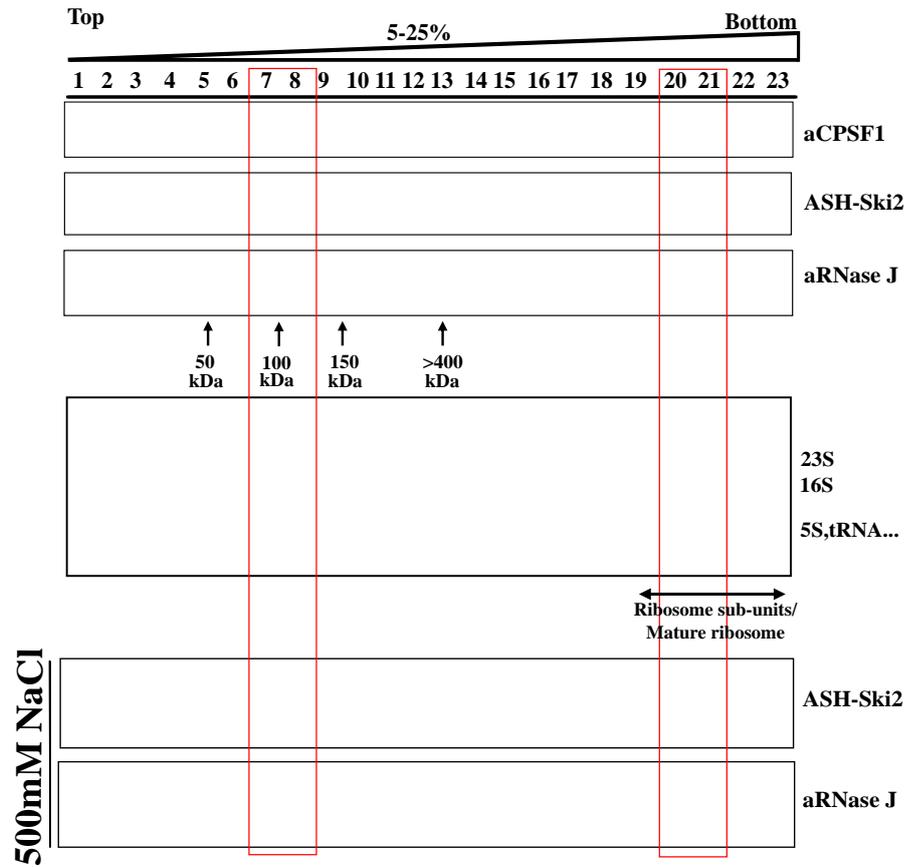
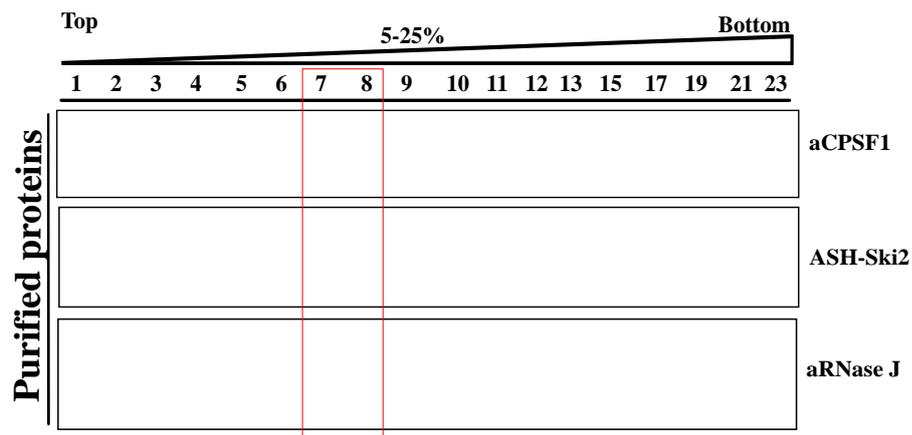
A**B**

Figure 18: Localisation des protéines aCPSF1, aRNase J et ASH-Ski2 sur gradient de sucrose 5-25%. A. Fractionnement d'extrait de *P. abyssi* sur gradient 5-25% sucrose PBS 1X et ultra-centrifugation à 35000 rpm durant 20h. Les protéines sont révélées par western blot avec des anti-corps anti-aCPSF1, ASH-Ski2 et aRNase J (panel haut). Gel agarose 1% TBE1X des extractions ARN (panel milieu). Expérience en présence de 500mM de NaCl (panel bas). Les poids moléculaires des fractions sont indiqués et sont issues de fractionnement de protéines étalon sur le type de gradient. B : Fractionnement des protéines recombinantes purifiées.

sous-unités 30S, 50S et du ribosome peuvent ainsi être identifiés. La quantification relative des signaux obtenus par Western blot permet de visualiser clairement la co-sédimentation de ASH-Ski2 et d'une fraction de aRNase J endogène avec les ARN ribosomiques (Fractions 18-21, **Figure 19A**). On observe également d'une proportion importante de aRNase J endogène qui co-sédimente avec la fraction ribosome/polysome (voir encadré fractions 18 à 22 (**Figure 19A**)).

L'ajout d'EDTA à 20 mM au tampon PBS lors des préparations de l'extrait cellulaire et du gradient entraîne comme attendu une dissociation des polysomes et du ribosome (**Figure 19B**). En effet les ARN ribosomiques 16S et 23S disparaissent dans ces conditions des fractions de très haut coefficient de sédimentation et se retrouvent majoritairement dans les fractions 12-13 pour le 16S et la fraction 15 pour le 23S. Cependant on observe une quantité résiduelle d'ARN ribosomique dans les fractions 18 à 23 montrant que la dissociation n'est peut-être pas totale en présence de 20 mM EDTA (**Figure 19B**). Néanmoins, dans ces conditions, on observe clairement une co-sédimentation d'aRNase J et d'ASH-Ski2 avec les ARN ribosomiques (Fractions 10 à 16-**Figure 19B**). De plus la proportion de ribonucléase aRNase J disparaît des fractions 18 à 21 et se retrouve condensée dans les fractions 9 à 17 avec ASH-Ski2 endogène.

L'augmentation de la force ionique par addition de 1M NaCl au tampon PBS entraîne une dissociation de ASH-Ski2 (Fraction 1 et 2-**Figure 19C**) de complexe de haut poids moléculaire qui se retrouve alors sous sa forme monomérique. La ribonucléase aRNase J est également déplacé vers des fractions de vitesse de sédimentation plus basse (Fractions 11 et 12-**Figure 19C**), cependant sa vitesse ne correspond pas à la protéine sous sa forme monomérique. Ces résultats sont en accord avec ceux du gradient 5-25% en présence de 500mM de NaCl (**Figure 18A**)

d. Fractionnement cellulaire sur gradient de sucrose 10-30% dans le Tampon TK

Nous avons favorisé dans ces expériences l'utilisation d'un tampon TK dans lequel le KCl remplace le NaCl afin de se rapprocher des conditions intracellulaires déterminées par la mesure des niveaux des ions à l'intérieur de certaines cellules eucaryotes et bactériennes (Roe *et al.*, 1998; van Eunen *et al.*, 2010). Ainsi les profils de sédimentation des ribonucléases aCPSF1 et aRNase J et des hélicases aLhr2 et ASH-Ski2 et d'autres partenaires du réseau tels que les sous-unités de la RNA polymérase et de l'exosome de *P.abysyi* ont été analysés sur des

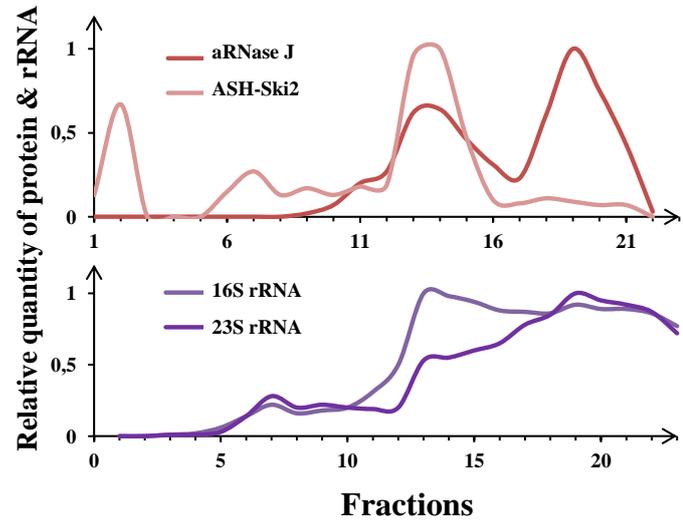
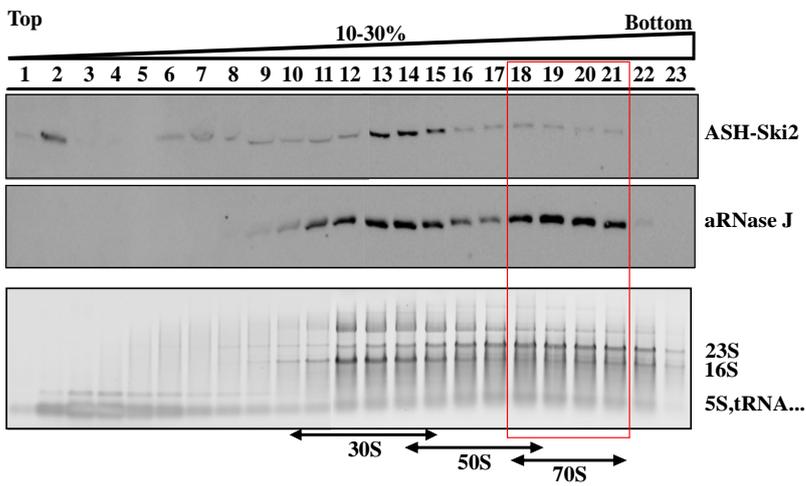
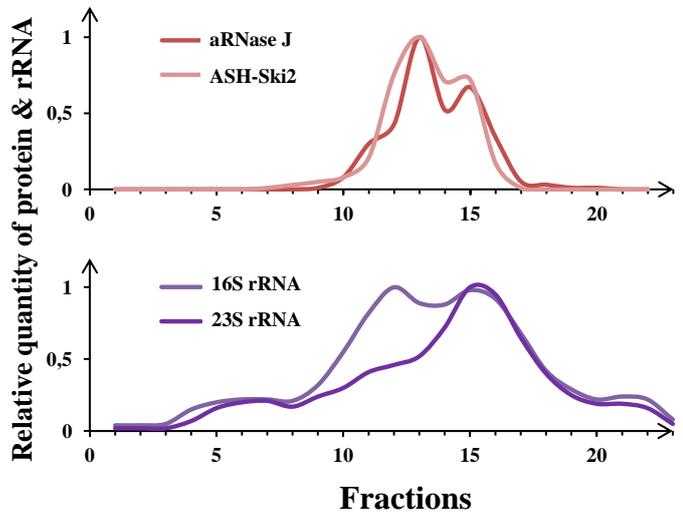
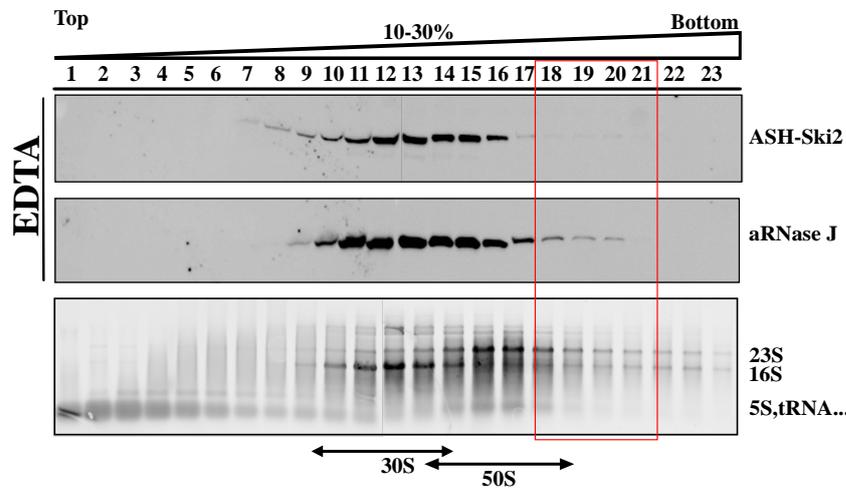
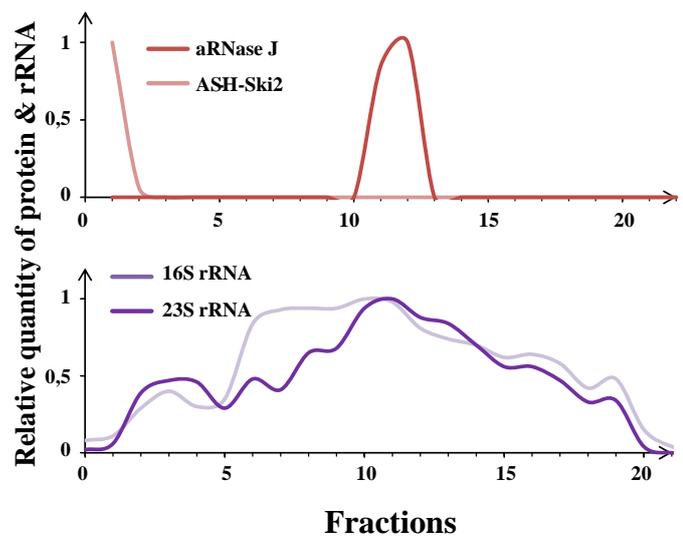
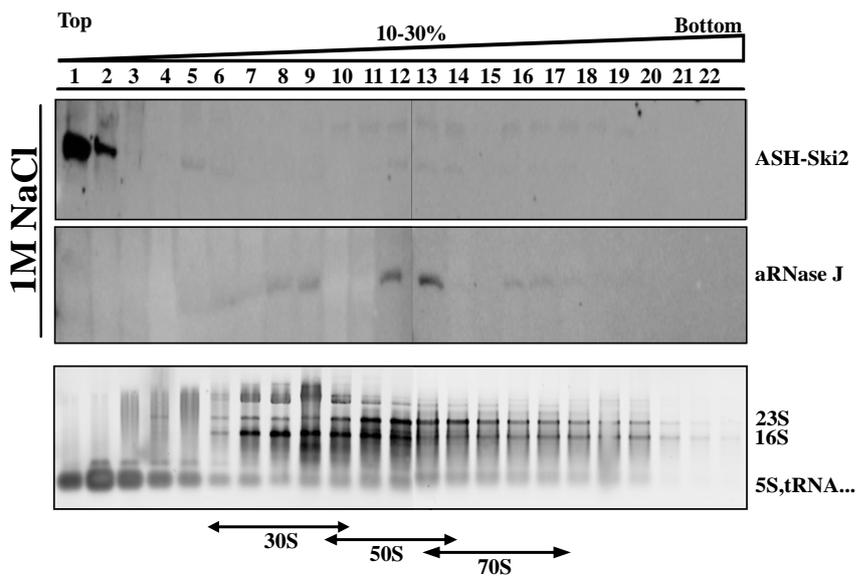
A**B****C**

Figure 19: Localisation des protéines aRNase J et ASH-Ski2 sur gradient de sucrose 10-30%. A: Fractionnement d'extrait de *P. abyssi* sur gradient 10-30% sucrose PBS 1X et ultra-centrifugation à 35000 rpm durant 5h. Western blotting avec anti-corps anti ASH-Ski2 et anti-aRNase J (panel haut). Gel agarose 1% TBE1X des extractions ARN (panel bas). B: Fractionnement d'extrait de *P. abyssi* sur gradient 10-30% sucrose PBS 1X supplémenté de 20mM EDTA. C: Fractionnement d'extrait de *P. abyssi* sur gradient 10-30% sucrose supplémenté avec 1M de NaCl. Les panels droits de chaque figure indique les quantification relative effectuée sur ImageLab (Biorad). Les différents fractions dans lesquelles sédimentent les sous-unités du ribosome sont indiquées.

gradients 10-30 % dans un tampon TK en absence et présence de 20mM EDTA (**Figure 20**). Certaines différences dans les profils de sédimentation sont observées selon le tampon utilisé.

Le profil de sédimentation de aCPSF1 présente 3 pics et varie légèrement en présence d'EDTA (**Figure 20A et B**). Dans le tampon TK, les fractions 1-2 correspondraient à sa forme dimérique et 3-4 pourraient correspondre au dimère avec d'autres partenaires protéiques à identifier (**Figure 20A**). Ces pics se restreignent à la fraction 2 en présence d'EDTA (**Figure 20B**). Les fractions 5 et 6 contiennent la majorité de la protéine endogène aCPSF1 dans les deux conditions suggérant qu'aCPSF1 pourrait être associée de manière significative à d'autres facteurs cellulaires. On observe également un profil étalé au-delà de la fraction 9 suggérant un recrutement plus large de aCPSF1 dans différents complexes de la cellule.

À l'opposé de aCPSF1, les profils de l'hélicase aLhr2 varient clairement en fonction de la présence d'EDTA. Lhr2 sédimente dans les fractions 1 et 2 (forme mono ou dimérique ?) et dans les fractions de 7 à 21 dans un tampon TK suggérant que aLhr2 appartiendrait à plusieurs complexes multi-protéines de composition variable. Ces complexes seraient sensibles à la présence d'EDTA puisque les 2 profils des fractions 10 à 21 (**Figure 20A et B**) ne sont pas superposables. En TK-EDTA, on observe une accumulation de aLhr2 dans la fraction 5 avec aCPSF1 (**Figure 20B**). Cette superposition pourrait avoir une signification physiologique sachant que aLhr2 est une protéine retrouvée dans le réseau d'interaction de aCPSF1 (**Figure 14-Table 1**).

De manière remarquable et telle observés dans le tampon PBS, les profils de sédimentation d'aRNase J et d'ASH-Ski2 se superposent parfaitement en absence et en présence d'EDTA (**Figure 20A et B**). Les pics des profils en présence d'EDTA sont moins larges et se superposent au pic de l'ARN ribosomique 23S d'après nos quantifications relatives qui restent indicatives. Ainsi en absence de ribosome/polysomes (dissociés par la présence d'EDTA), aRNase J et ASH-Ski2 disparaissent des fractions 17 à 21.

Afin de comparer les profils de sédimentation d'aCPSF1, aLhr2, aRNase J et ASH-Ski2 endogènes, avec ceux des composants de l'exosome de *P. abyssi*, des anticorps dirigés contre la coiffe de l'exosome DnaG d'une part et de la sous-unité catalytique Rrp41 du core, d'autre part, ont été utilisés pour révéler leur profil en absence et en présence d'EDTA (**Figure 20**). Le profil de sédimentation de la protéine de la coiffe DnaG recouvre ceux de aRNase J et ASH-Ski2. DnaG s'accumule dans les fractions de très haut poids moléculaire avec un

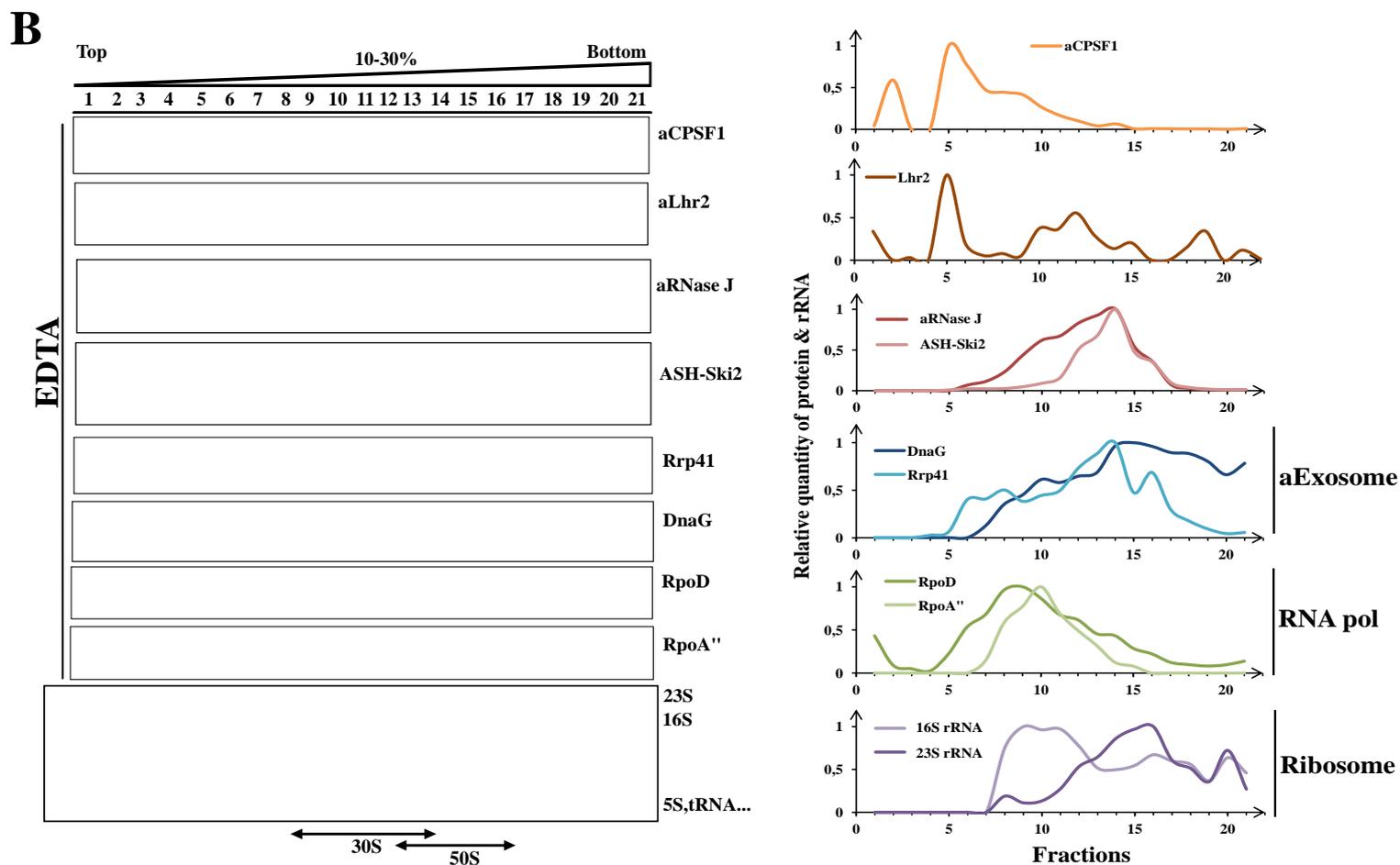
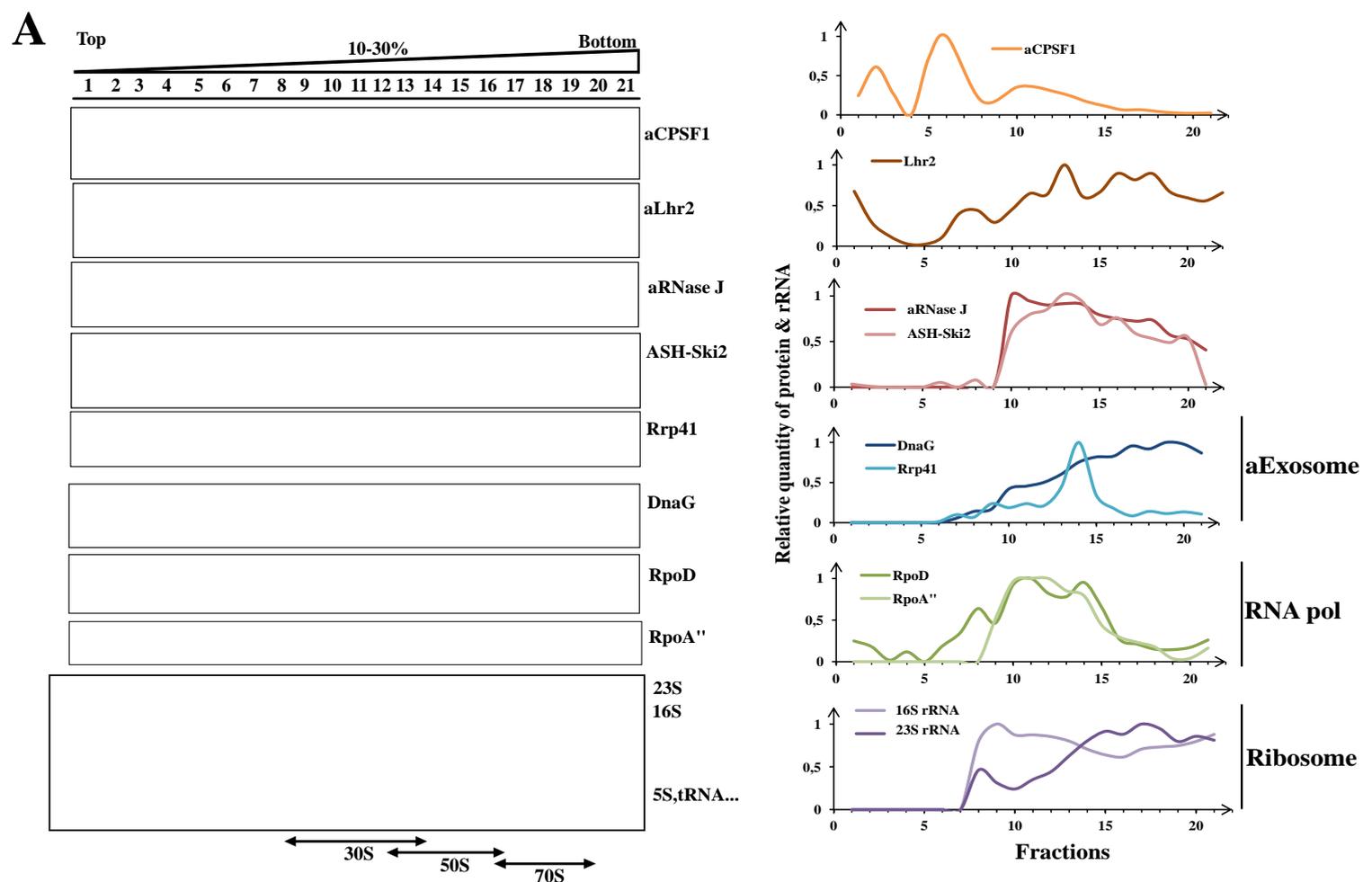


Figure 20: Localisation des protéines aCPSF1, Lhr2, aRNase J, ASH-Ski2, des sous-unités de l'exosome et celui de la RNA pol sur gradient de sucrose dans un tampon TK. A: Fractionnement d'extrait de *P. abyssi* sur gradient 10-30% sucrose tampon Tris-HCl, KCL et ultracentrifugation à 35000 rpm durant 5h. Western blotting avec anti-corps anti-aCPSF1, -Lhr2, -ASH-Ski2, -aRNase J, -Rrp41, -DnaG, -RpoD et -RpoA''. Gel agarose 1% TBE1X des extractions ARN traitées à la DNase I visualisés au SybrGreenII. Quantification normalisées des bandes issues des Western Blotting et extractions des ARN (panels droit). B: Fractionnement d'extrait de *P. abyssi* dans les mêmes conditions que A dans un tampon Tris-HCl, KCL supplémenté avec 20mM d'EDTA et ultracentrifugation à 35000 rpm durant 5h.

maximal dans les fractions 19 à 20. Ce profil est altéré en présence d'EDTA et le pic est déplacé dans les fractions 14 à 16 avec une nette diminution dans les fractions suivantes. A l'opposé, Rrp41 a un profil marqué avec un pic majoritaire dans un tampon TK qui suit la sous-unité 50S du ribosome (fraction 14- **Figure 20**). En présence d'EDTA, ce pic s'élargit (faible diminution au profit de fractions majoritairement moins lourdes) et étonnement recouvre celui de ASH-Ski2 (Fractions 5 à 14). Ainsi en présence d'EDTA, la fraction 14 contient à la fois aRNase J, ASH-Ski2, DnaG, Rrp41 et l'ARN ribosomique 23S.

Dans ce contexte, les profils de sédimentation des sous-unités RpoD et RpoA2 se recouvrent comme attendu et sont légèrement affecté par la présence d'EDTA avec un déplacement des pics majoritaires des fractions 11 et 14 vers fraction 10). Dans ce cas, les sous-unités de la RNA polymérase ne co-sédimentent ni avec aCPSF1, ni ASH-Ski2, ni aRNase J mais présentent un profil de sédimentation qui recouvre partiellement la sous-unité 30S du ribosome (**Figure 20**). Cette sédimentation pourrait être expliquée par les tailles proches de la RNAPol avec ces facteurs de transcription et de la petite sous-unité du ribosome (autour de 700 kDa).

Nous avons réitéré les expériences de fractionnement cellulaire afin de vérifier la reproductibilité des profils observés pour aRNase J et ASH-Ski2. Ainsi, le **Figure 21** montre les profils de aRNase J et ASH-Ski2 se superposent comme montré précédemment (**Figure 20**). Cette superposition est d'autant plus marquée en présence d'EDTA. De plus, un fractionnement avec des extraits cellulaires ayant subi un traitement aux ribonucléases RNase A/T1 donne des profils de sédimentation altérés pour aRNase J et ASH-Ski2. Dans ces conditions, le pic de aRNase J se restreint aux fractions 8 à 12 (**Figure 21A**), tandis que ASH-Ski2 s'accumule dans les fractions 1 à 3 (**Figure 21B**) correspondant certainement à son état monomérique et des fractions de très haut poids moléculaire (fractions 20 à 22-**Figure 21B**). Les ARN cellulaires semblent donc déterminant pour la localisation de la sédimentation ASH-Ski2 et sa co-localisation avec aRNase J. Cette localisation de ASH-Ski2 dans les fractions de très haut poids moléculaire est similaire à ce qui est observée pour la protéine SKI3 du complexe SKI chez les eucaryotes lorsque l'extrait de levure sont traité avec de la RNase T1 (Schmidt *et al.*, 2016). D'après les auteurs, cette accumulation pourrait correspondre au recrutement du complexe SKI au niveau des ribosomes matures bloqués sur des ARNm non fonctionnels afin de dégrader les transcrits et permettre la libération des ribosomes.

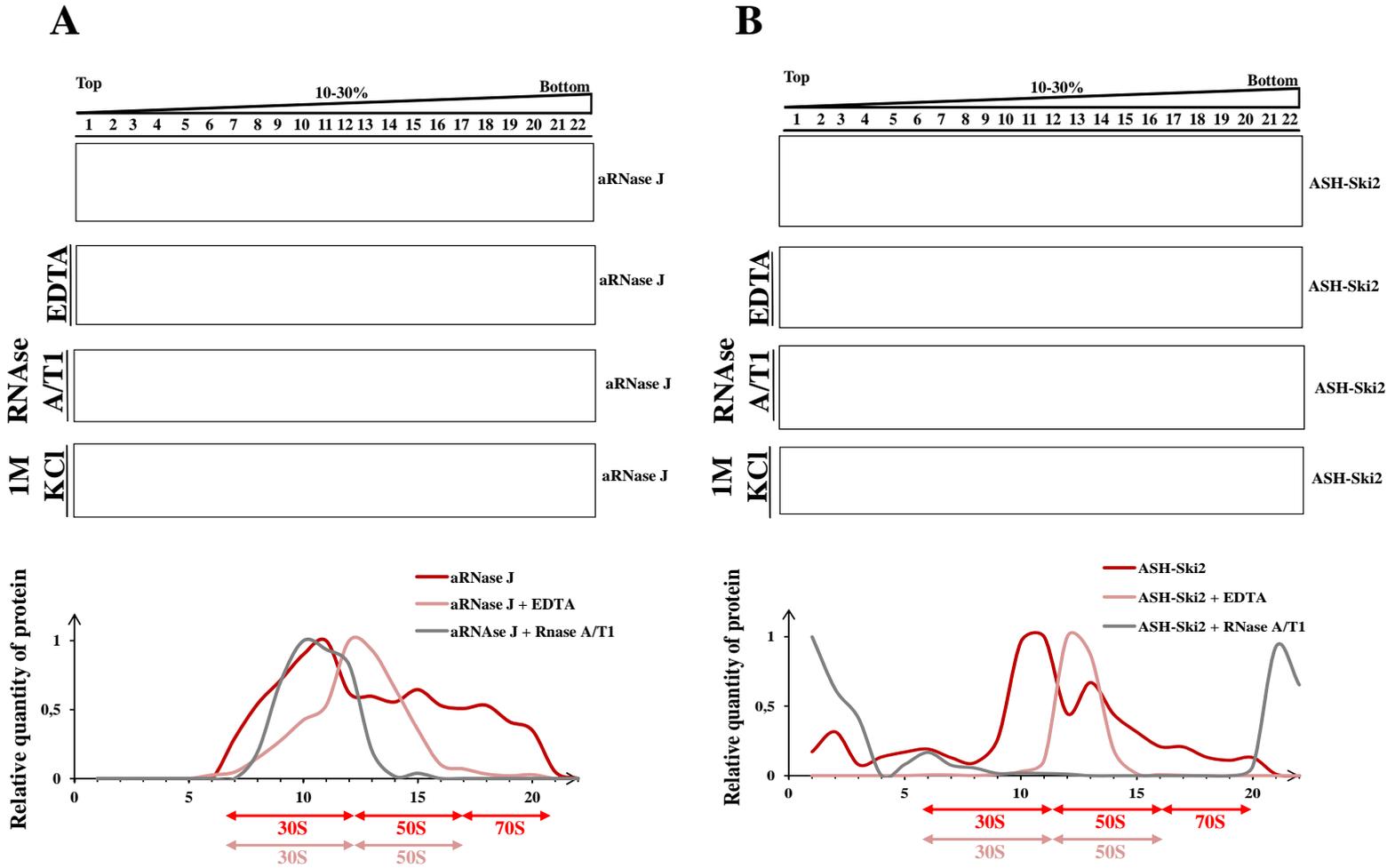


Figure 21. Rôle de l'EDTA, d'un traitement aux nucléases et 1M de KCL sur localisation des protéines aRNase J et ASH-Ski2 dans un tampon TK . Fractionnement d'extrait de *P. abyssii* sur gradient 10-30% sucrose tampon Tris-HCl, KCL et ultracentrifugation à 35000 rpm durant 5h. Western blotting avec anti-corps anti-aRNase J et ASH-Ski2 (panel haut). Quantification normalisée des bandes issues des Western Blotting et extractions des ARN (panels bas). A: Localisation de aRNase J. B: Localisation de ASH-Ski2. Les différentes fractions dans lesquelles sédentent les sous-unités du ribosome sont indiquées avec les différentes conditions.

e. Conclusions sur les profils de sédimentation

L'ensemble des fractionnements à partir d'extraits de *P. abyssi* sur différents types de gradient montre de manière récurrente une co-sédimentation de aRNase J et ASH-Ski2. Les profils observés recouvrent ceux des sous-unités du ribosome et de l'exosome. aRNase J se retrouve dans les fractions correspondantes au ribosome. La dissociation du ribosome en présence d'EDTA engendre la disparition de aRNase J dans les fractions de très haut coefficient de sédimentation. Dans les différentes fractions où la présence de aRNase J et ASH-Ski2 est détectée, les sous-unités de l'exosome, DnaG et Rrp41, sont également remarquées. De manière intéressante, aCPSF1 et Lhr2 présentent des profils de sédimentation qui permet de proposer que aLhr2 et aCPSF1 pourraient interagir en concertation. Ces résultats sont cohérents avec les résultats de protéomique identifiant aLhr2 dans le réseau d'interaction de aCPSF1.

II-1-3. Analyse du contexte génomique des gènes codant pour les ribonucléases aCPSF1 et aRNase J et leurs partenaires potentiels.

Un article de Koonin en 2001 proposait une concertation entre les machineries de l'exosome, celui du protéasome, de celle de la transcription et de la traduction chez les archées. En effet, par des analyses bio-informatiques, l'existence d'un super-opéron unique ou morcelé dans l'ensemble des génomes d'archées a été démontré (Koonin *et al.*, 2001). Au sein de cet opéron, les gènes codant pour les sous-unités du core de l'exosome (Rrp41 et Rrp42) ainsi que la coiffe (Rrp4) sont présentes juste en aval de la sous-unité alpha active du protéasome. Deux protéines du ribosome sont aussi présente (L15E and L37AE) ainsi que la sous-unité RpoP de l'ARN polymérase (**Figure 22**).

De manière intéressante, l'analyse du contexte génomique du gène codant pour aCPSF1, gène conservé dans tous les génomes séquencés à ce jour, nous indique la présence du gène codant pour la sous-unité bêta du protéasome en amont de celle-ci chez les thermococcales. Cette localisation de aCPSF1 ainsi que la présence de la protéine PAB0569 pouvant intervenir dans l'assemblage du protéasome identifié dans le réseau d'interaction de aCPSF1 (partie II-1-1) permet de suggérer un rôle concerté de aCPSF1 avec la machinerie de dégradation des protéines, hypothèse émise par Chavarria *et al* en 2014.

L'analyse du contexte génomique du gène codant pour aRNase J montre la présence du gène *idsA* codant pour une protéine synthétisant une chaîne courte d'isoprényl diphosphate

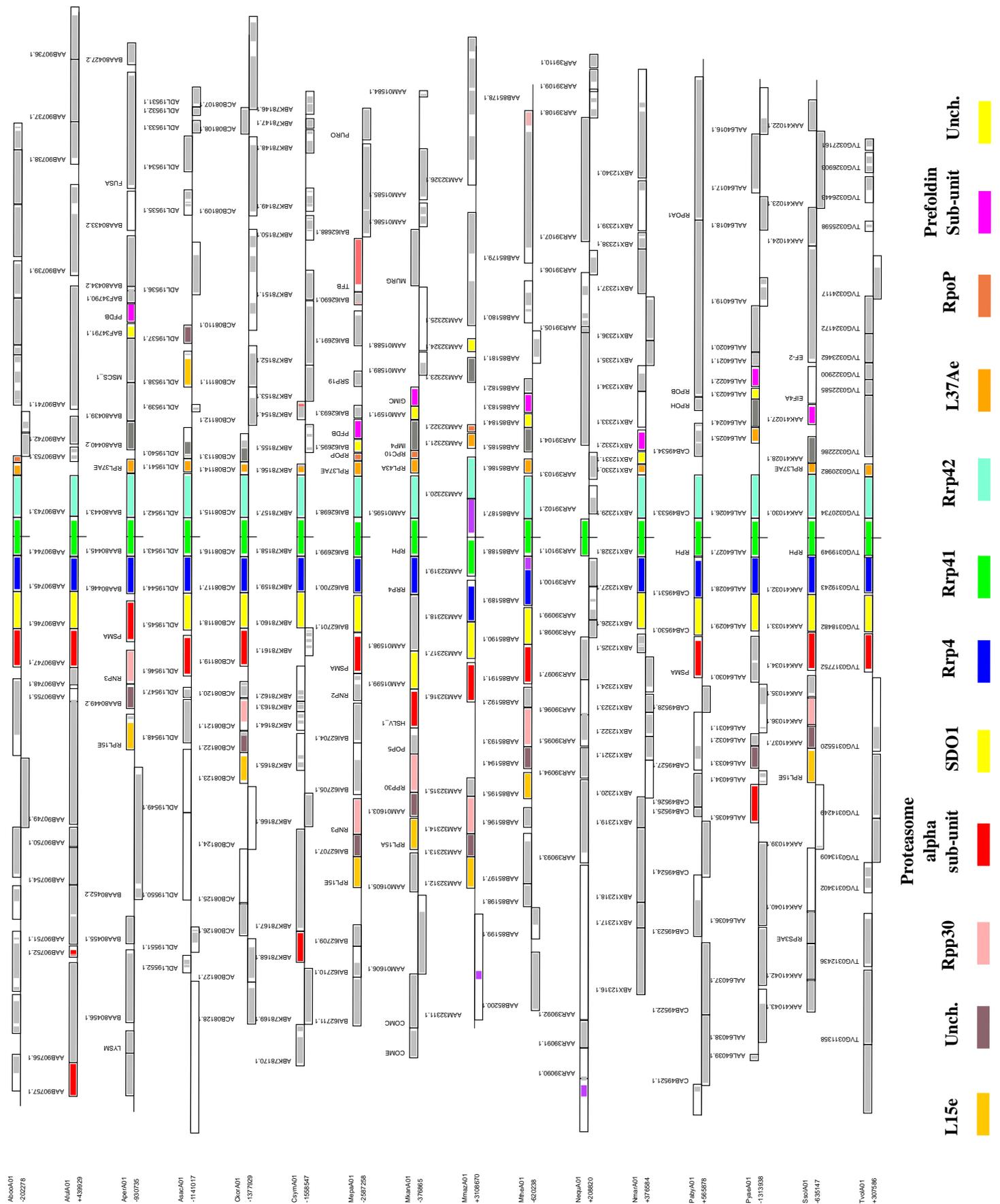


Figure 22: Contextes génomiques autour des gènes codant pour le core de l'exosome. Contexte génomique centré sur les gènes codant pour les sous-unités de l'exosome chez les représentants de chaque groupes d'archées. Les COG de Rrp4, Rrp41 et Rrp42 ont été recherché dans l'ensemble des génomes d'archée séquencé. Les gènes les plus conservés sont indiqués.

certainement importante dans la composition membranaire dans l'ensemble des euryarchées. La fonction cellulaire de IdsA reste cependant inconnue mais ce contexte révèle peut-être un lien entre aRNase J et la membrane. Un tel lien est observée chez les bactéries entre respectivement les ribonucléases RNase E et RNase Y et les membranes internes de *E. coli* et *B. subtilis* (Khemici *et al.*, 2008; Lehnik-Habrink *et al.*, 2011).

Toujours au sein des Thermococcales, lorsque nous regardons le contexte génomique du gène codant pour ASH-Ski2, nous observons, en aval de celle-ci, la présence du gène codant pour la protéine Csl4, sous-unité de la coiffe de l'exosome (**Figure 23**). Ces deux gènes sont de plus chevauchants comme ceux codant pour les sous-unités Rrp4, Rrp41 et Rrp42 de l'exosome. Cette organisation d'une sous-unité de l'exosome avec ASH-Ski2 renforce les liens établis entre la protéine ASH-Ski2 et le complexe exosome montrés à partir des réseaux d'interaction et des profils de sédimentation. De plus, d'autres gènes codant pour des protéines de modification des ARN retrouvées au sein des réseaux d'interaction de ASH-Ski2 et de aRNase J ont aussi pu être observés dans cette région. C'est le cas des gènes codant pour les protéines FlpA et Nop5p et un gène codant pour un régulateur de la transcription Lrp (**Figure 23**). De manière intéressante, nous identifions aussi la présence du gène codant pour la sous-unité RpoL de la RNA polymérase dans la région de ASH-Ski2.

II-2. IDENTIFICATION D'INTERACTION PROTEINE-PROTEINE *IN VITRO* ENTRE PARTENAIRES DES RESEAUX DE *Pab*-aRNase J, *Pab*-ASH-Ski2 ET *Pab*-aCPSF1

Les interactomes de *Pab*-aRNase J et de *Pab*-ASH-Ski2 réalisés à partir d'extrait cellulaire de *P. abyssi* (Partie II-1, **Figure 17**), incluent les sous unités catalytiques Rrp41 et de la coiffe DnaG de l'aExosome (Figure réseau aCPSF1 aRNase J ASH-Ski2) et un nouveau facteur RecJ-SOD, aux propriétés catalytiques et fonctionnelles à définir. De plus, les résultats de fractionnement cellulaire sur gradient de sucrose montrent une co-sédimentation des protéines endogènes aRNase J, ASH-Ski2 avec les composants de la partie catalytique et de la coiffe de l'exosome (Partie II-2- **Figure 20**). Dans ce contexte, nous avons testé, par co-purification par chromatographie d'affinité, les interactions protéine-protéine établies *in vitro* entre ces différents partenaires à l'aide des protéines recombinantes correspondantes. Nous avons choisi de tester l'interaction *in vitro* de *Pab*-aRNase J et de *Pab*-ASH-Ski2 respectivement avec la ribonucléase β -CASP *Pab*-aCPSF1, les sous-unités de l'exosome *Pab*-DNaG et *Pab*-

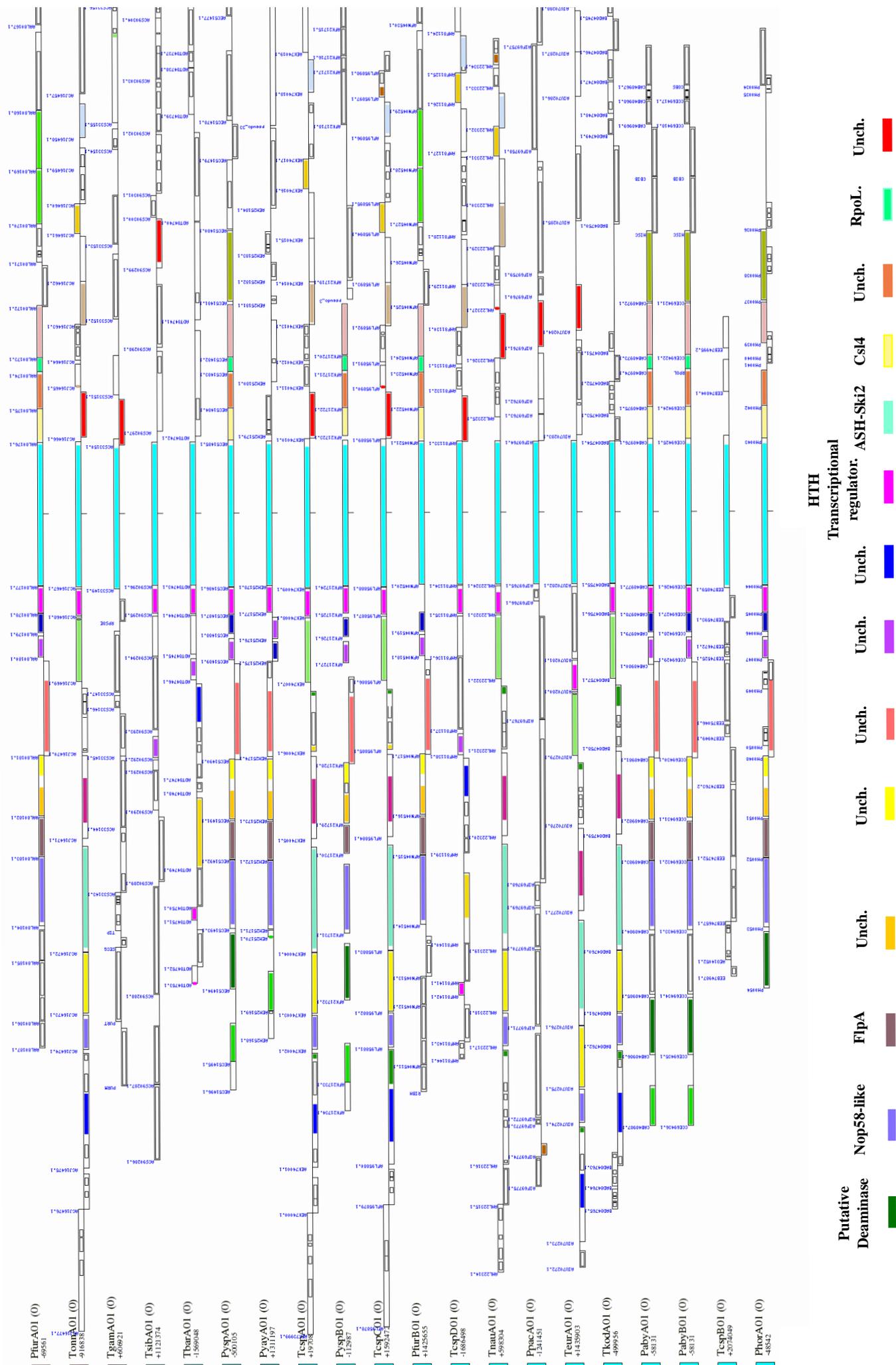


Figure 23: Contextes génomiques autour du gène codant pour ASH-Ski2 chez les Thermococcales. Contexte génomique centré sur le gène codant pour ASH-Ski2. . Le COG de ASH-Ski2 a été recherché dans l'ensemble des génomes d'archée séquencé. Les gènes les plus conservés sont indiqués.

Rrp41. Chaque protéine est surexprimée dans *E. coli* à partir d'un vecteur d'expression répertorié dans la **Table M1**. Si une protéine sans étiquette (protéine proie) est retenue sur la colonne d'affinité en présence d'une protéine avec l'étiquette 6XHis (protéine appât) dans nos conditions expérimentales (20mM NaPhosphate, 300mM NaCl, 0,5mM DTT, 10% glycérol), nous proposons que ces 2 protéines établissent une interaction *in vitro* qui peut être directe ou médiée par des acides nucléiques en fonction de la dépendance de l'interaction observée à un traitement aux nucléases (RNase A/DNase I) lors de la lyse des culots cellulaires.

II-2-1. Méthode de co-purification par chromatographie d'affinité

Des vecteurs d'expression pET spécifiquement construits pour cette étude (**Table M1**) sont utilisés pour surexprimer d'une part les protéines appâts avec une étiquette 6XHis et d'autre part les protéines proies dans la souche *E. coli* BL21 Codon+ contenant le système d'expression T7. Le choix de l'étiquette histidine a été retenu du fait de sa petite taille, de sa résistance à un traitement thermique (70°C) et la possibilité de purifier les protéines par FPLC. La surexpression des protéines proies et appâts est induite par l'ajout d'IPTG à des cultures en phase exponentielle de croissance. Pour étudier la capacité de deux protéines à établir une interaction, deux culots de cellules ayant surexprimés soit la protéine proie soit la protéine appât, sont mélangés pour l'étape de lyse qui inclut deux étapes de chauffage à 70°C. Le surnageant est déposé sur une colonne d'affinité au Nickel via FPLC (Fast Protein Liquid Chromatography). La protéine appât portant l'étiquette 6XHis est retenue par affinité grâce à son étiquette de six résidus histidines sur les billes de Nickel. La protéine proie est également retenue que si elle est capable d'interagir avec la protéine appât. Un traitement aux nucléases (RNase A-DNase I) des extraits après mélange des culots lors de la lyse permet de détecter l'interaction appât-proie médiée par des acides nucléiques. Des lavages à faible concentration en Imidazole (10 à 50mM) éliminent la protéine proie si elle n'interagit pas avec la protéine appât. Les éluions à concentration croissante d'Imidazole (50 à 500mM) permettent de décrocher par compétition la protéine appât ainsi que la protéine proie associée. Les différentes fractions sont récoltées et les protéines de chaque fraction sont séparées par sur un gel SDS-PAGE puis visualisées par coloration au bleu de Coomassie et révélées par Western blot avec les anticorps polyclonaux dirigés contre chaque protéine d'intérêt (disponibles dans l'équipe).

Pour notre étude, seules les protéines *Pab*-aRNase J et *Pab*-ASH-Ski2 sont utilisées en tant que protéines proies et exprimées à partir d'un vecteur pET11b. Nous avons vérifié que ces

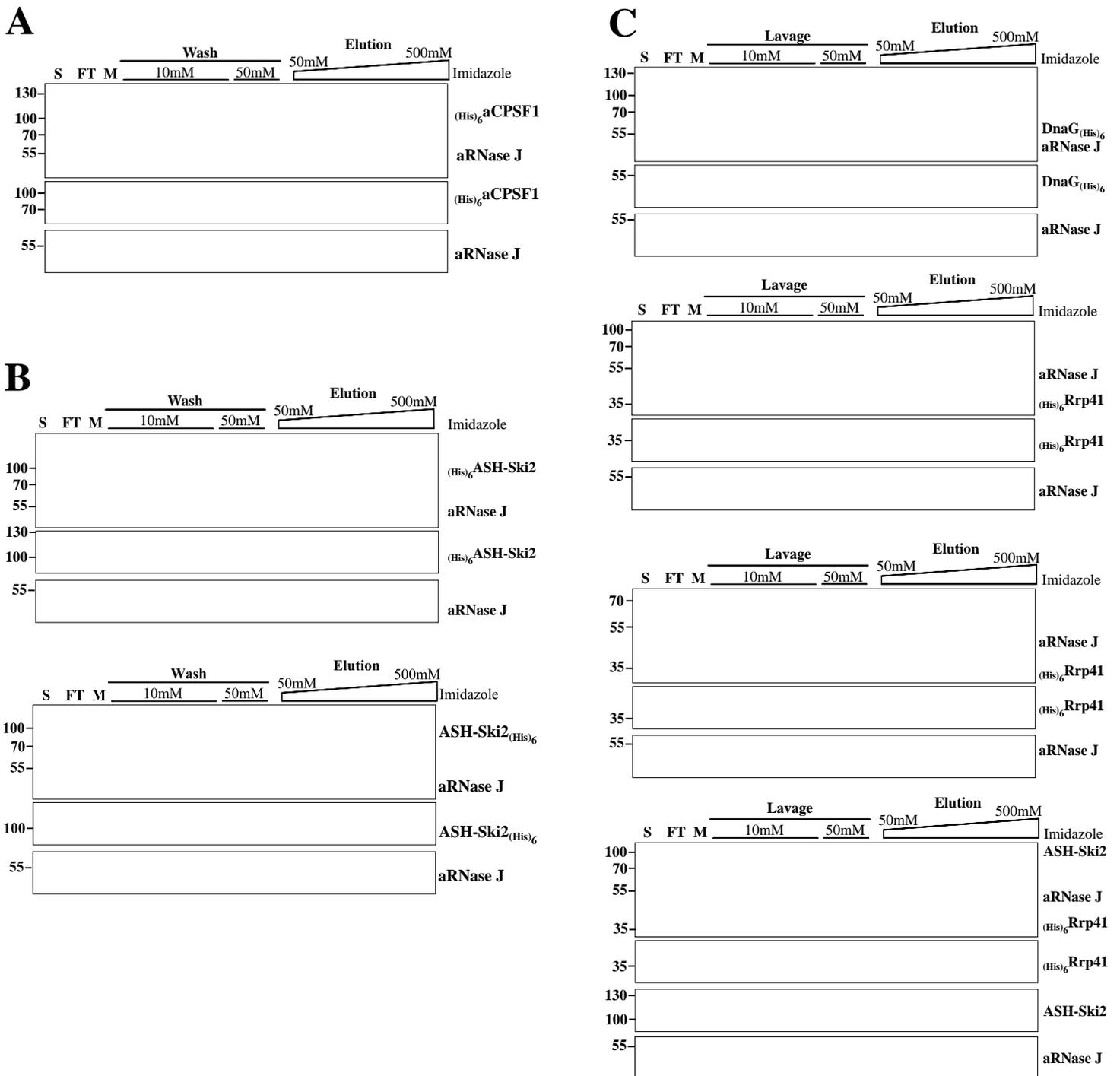


Figure 24: Co-purification des protéines des réseaux d'interaction par chromatographie d'affinité avec aRNase J comme proie.

A: Les vecteurs pET15b-(His)₆-aCPSF1 et pET11b-aRNase J ont été transformés séparément dans la souche *E. Coli* BL21 DE3 Codon+. Les deux protéines ont été surexprimées par induction à l'IPTG. Les deux culots de cellules ont été mélangés et lysés par sonication. L'extrait après chauffages à 70°C sont traité ou non aux nucléases (Rnase A/Rnase T1 et Dnase I). Après clarification, l'extrait est chargé sur une colonne de nickel est purifié par chromatographie d'affinité. Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (panel haut) ou par Western blotting (panel bas). B : Co-purification avec ASH-Ski2 étiquetée en N- ou C-terminal comme appât. C: Co-purification de aRNase J avec DnaG et Rrp41.

deux protéines n'ont aucune affinité intrinsèque pour la colonne de Nickel (**Figure Annexe1**). A l'opposé les protéines *Pab-aCPSF1*, *Pab-DnaG* et *Pab-Rrp41* ayant une affinité intrinsèque pour la colonne de Nickel ne peuvent être utilisées en tant que proies et sont donc exclusivement utilisés comme appâts. La position de l'étiquette à l'extrémité N ou C-terminale de la protéine appât peut être critique dans certaines configurations pour détecter des interactions protéine-protéine. En effet nous avons observé que dans certains cas l'emplacement de l'étiquette 6XHis en N-ter ou C-ter de la protéine appât compromet l'interaction. Ainsi selon les cas, les protéines appâts sont exprimées soient à partir d'un vecteur pET15b ou d'un vecteur pET21b qui étiquettent les appâts en N- ou C-terminal, respectivement. Ici, les protéines dont les noms sont précédés de *Pab*-(His)₆ sont étiquetées en N-ter tandis que les protéines dont les noms sont suivis de -(His)₆ sont étiquetées en C-ter.

Les résultats des différentes expériences de co-purification pour chaque couple proie-appât sont montrés dans les **Figure 24 et 25** et sont résumés dans la **Table 5 et 6**.

II-2-2. Interaction protéine-protéine observée par co-élution de la protéine-proie *Pab-aRNase J* (Figure 24)

a. Interaction de *Pab-aRNase J* avec l'appât *Pab*-(His)₆-aCPSF1

La protéine aRNase J étant retrouvée dans l'interactome de aCPSF1, l'interaction potentielle entre ces deux protéines a été testée (**Figure 24A**). L'appât *Pab-aCPSF1* ne permet pas de retenir la proie *Pab-aRNase J*. En effet comme attendu la majorité de *Pab*-(His)₆-aCPSF1 est retrouvée dans les fractions d'élution (230mM d'imidazole environ). Cependant *Pab-aRNase J* est principalement retrouvée dans les fractions non retenues FT (« *Flow Through*») et dans les premières fractions de lavage et totalement absente des fractions d'élution. La proie *Pab-aRNase J* n'est donc pas co-purifiée avec l'appât *Pab*-(His)₆-aCPSF1. Ces résultats indiquent que les deux ribonucléases de la famille des β -CASP *Pab-aRNase J* et *Pab*-(His)₆-aCPSF1 ne peuvent interagir *in vitro* dans ces conditions.

b. Interaction de *Pab-aRNase J* avec les appâts *Pab*-(His)₆-ASH-Ski2 ou *Pab*-ASH-Ski2-(His)₆

Avec le couple proie/appât, *Pab-aRNase J*/*Pab*-(His)₆-ASH-Ski2, nous observons qu'une fraction de *Pab-aRNase J* est co-éluee avec *Pab*-(His)₆-ASH-Ski2 (**Figure 24B, panel haut**). Cette co-élution est insensible au traitement RNase/DNase. Ainsi, nous proposons que ce

couple de protéines interagisse de manière indépendante de la présence d'acides nucléiques. Cependant la co-élution des deux protéines est dépendante des deux étapes de chauffage à 70°C lors de la lyse concomitante des deux culots de cellules surexprimant chacune des deux protéines. La co-élution des deux protéines n'est pas observée lorsque l'expérience est menée à 4°C ou à température ambiante. Il semblerait donc que le passage à 70°C est un pré requis pour l'interaction entre ces deux protéines hyperthermophiles certainement en privilégiant des conformations spécifiques et optimales pour l'interaction.

Avec le couple proie/appât, *Pab*-aRNase J/ *Pab*-ASH-Ski2-(His)₆, la protéine proie *Pab*-aRNase J traîne dans les fractions de lavages (**Figure 24B, panel bas**). La co-élution n'est clairement observée que dans la première fraction d'élution au début du gradient d'imidazole entre 50 et 100mM. Il semblerait donc que dans la configuration de l'étiquette en C-terminal de *Pab*-ASH-Ski2, l'interaction entre les deux protéines soit beaucoup plus labile.

Ces expériences de co-purification par affinité permettent tout de même de proposer qu'une interaction protéine-protéine entre *Pab*-ASH-Ski2 et *Pab*-aRNase J est possible *in vitro* et ne semble pas médiée par des acides nucléiques contrairement à d'autres interactions observés dans cette étude et décrites ci-dessous. La position de l'étiquette sur l'appât *Pab*-ASH-Ski2 est critique et indique que la présence de l'étiquette en C-terminal compromet l'interaction protéine-protéine qui peut s'établir dans nos conditions d'expérience.

La co-élution de *Pab*-aRNase J/*Pab*-(His)₆-ASH-Ski2 dans nos expériences de co-purification est cohérente avec les résultats de sédimentation sur gradient de sucrose et de chromatographie d'exclusion sur colonne S-200 des deux protéines purifiées (Résultat Manon Batista dans l'équipe), obtenus en parallèle, qui montrent respectivement leur co-sédimentation et leur assemblage sous forme de complexe. L'ensemble de nos résultats permettent de proposer l'existence d'un nouveau complexe multi-protéique composé de la ribonucléase *Pab*-aRNase J et de l'hélicase *Pab*-ASH-Ski2.

c. Interaction de *Pab*-aRNase J avec les appâts *Pab*-DnaG-(His)₆ et *Pab*-(His)₆-Rrp41

Pour tester l'interaction entre *Pab*-aRNase J et *Pab*-DnaG (composant de la coiffe de l'exosome), l'appât *Pab*-DnaG-(His)₆ est utilisé. Les expériences de co-purification par affinité ne détectent aucune interaction entre les deux protéines puisque *Pab*-aRNase J est exclusivement retrouvée dans les fractions du FT et des premiers lavages (**Figure 24C, panel haut**).

De manière intéressante, nos résultats de co-purification avec l'appât *Pab*-(His)₆-Rrp41 permettent de mettre en évidence une co-élution de *Pab*-aRNase J et de *Pab*-(His)₆-Rrp41 (**Figure 24C, panel centre**). Cependant, cette co-élution est sensible au traitement nucléases (RNase A-DNase I) réalisé après la mise en contact des culots cellulaires lors de la lyse (Voir méthode 1-2-1). En effet, *Pab*-aRNase J est absente des fractions d'élution dans ce cas (**Figure 24C, panel bas**). L'interaction entre *Pab*-aRNase J et *Pab*-(His)₆-Rrp41 serait donc labile, médiée ou stabilisée par des acides nucléiques. De façon intéressante, lorsqu'une expérience de co-purification est réalisée en présence de *Pab*-(His)₆-Rrp41 comme appât et avec les deux protéines proies *Pab*-aRNase J et *Pab*-ASH-Ski2 (exprimées indépendamment), aucune des deux protéines proies n'est retrouvée dans les fractions d'élution (**Figure 24C, panel bas**). Ainsi la co-élution observée entre *Pab*-aRNase J et *Pab*-(His)₆-Rrp41 n'est plus détectée en présence de *Pab*-ASH-Ski2. Il semblerait donc bien qu'il y ait formation d'un complexe *Pab*-ASH-Ski2/*Pab*-aRNase J qui se fait au détriment de l'interaction médiée par les acides nucléiques entre *Pab*-aRNase J et *Pab*-(His)₆-Rrp41.

II-2-3. Interaction protéine-protéine observée par co-élution de la protéine-proie *Pab*-ASH-Ski2 (Figure 25)

a. Interaction de *Pab*-ASH-Ski2 avec l'appât *Pab*-aRNase J-(His)₆

Il faut mentionner que la co-purification de *Pab*-aRNase J avec *Pab*-ASH-Ski2 (**Figure 24B**) n'est pas observée dans le contexte où *Pab*-(His)₆-aRNase J ou *Pab*-aRNase J-(His)₆ sont utilisés comme appâts et *Pab*-ASH-Ski2 comme proie (**Figure 25A**). La présence de l'étiquette (His)₆ en N-terminal ou C-terminal de *Pab*-aRNase J empêche l'interaction avec *Pab*-ASH-Ski2. L'étiquette pourrait ainsi altérer la structure de la protéine ou masquer les motifs ou résidus impliqués dans l'interaction protéine-protéine observée dans la configuration *Pab*-aRNase J/*Pab*-(His)₆-ASH-Ski2 (voir ci-dessus II-2-2-b).

b. Interaction de *Pab*-ASH-Ski2 avec les composants de l'exosome *Pab*-DnaG-(His)₆ et *Pab*-(His)₆-Rrp41

Les expériences de co-purification par affinité ne détectent aucune co-élution entre la proie *Pab*-ASH-Ski2 et respectivement les appâts *Pab*-DnaG-(His)₆ et *Pab*-(His)₆-Rrp41 dans nos conditions (**Figure 25B**). En effet, dans les deux expériences, nous observons *Pab*-ASH-Ski2 est exclusivement retrouvée dans les fractions de lavage. Ces résultats indiquent donc que

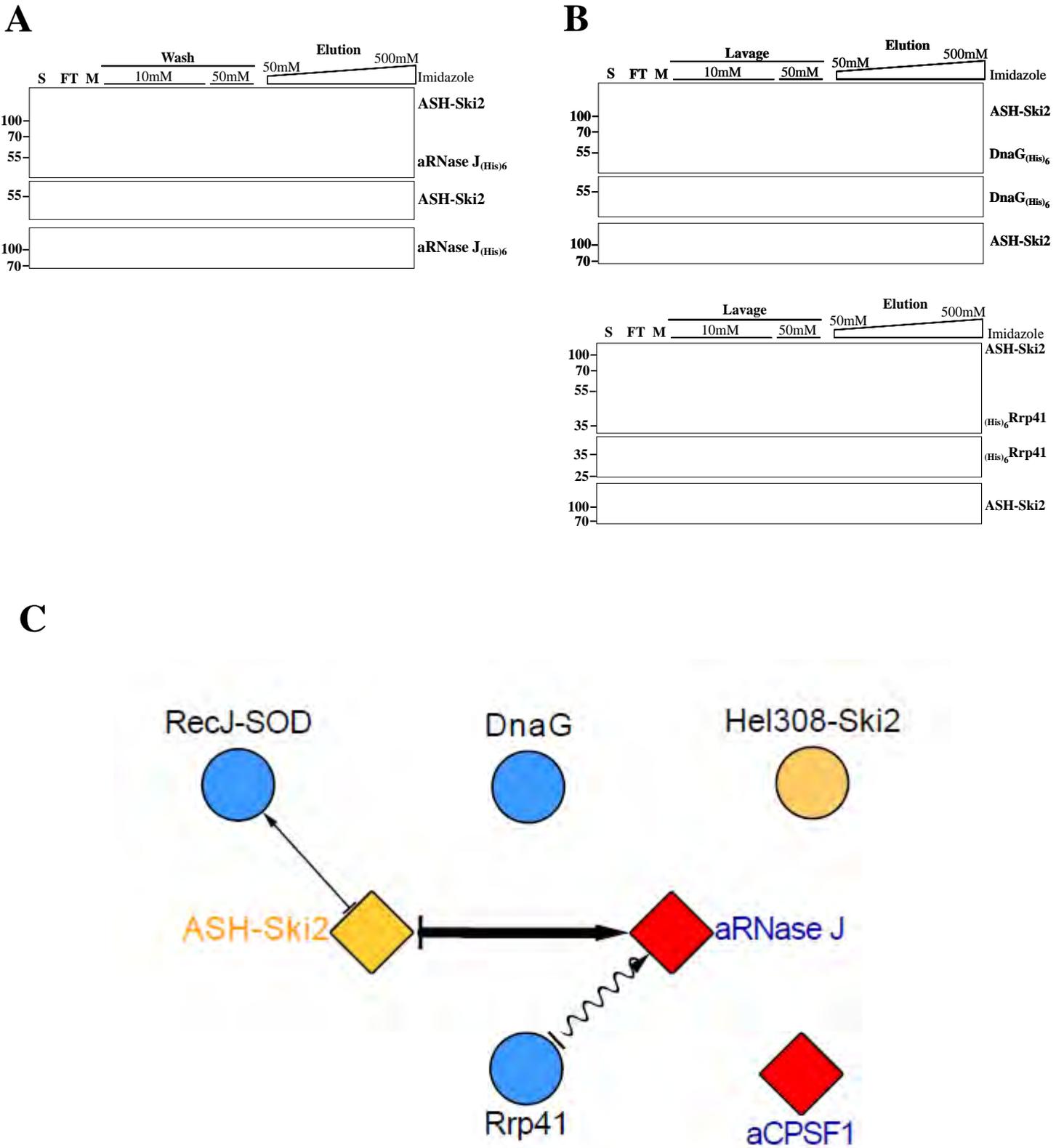


Figure 25: Co-purification des protéines des réseaux d'interaction par chromatographie d'affinité avec ASH-Ski2 comme proie.
 A: Les extraits ont été préparé comme pour la **Figure 24** avec aRNase J étiquetée . Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (panel haut) ou par Western blotting (panel bas). B : Co-purification avec DnaG et Rrp41 étiquetées. C: Résumé des tests d'interaction de aRNase J et ASH-Ski2. Les flèches droites indiquent des interactions non sensibles à des traitement aux nucléases. Celui en forme de vague représente une interaction acide nucléique dépendant.

l'hélicase *Pab*-ASH-Ski2 ne peut établir d'interaction ni avec *Pab*-DnaG-(His)₆ ni avec *Pab*-(His)₆-Rrp41 dans ces conditions.

Ainsi, l'ensemble des résultats de co-purification des protéines identifiées dans les réseaux d'interaction de *Pab*-aRNase et *Pab*-ASH-Ski2 sont résumés sur la **Figure 25C**.

II-3. IDENTIFICATION ET CARACTERISATION DU GROUPE HELICASE ASH-Ski2 CHEZ LES ARCHEES

Les réseaux d'interaction, les profils de sédimentation ainsi que des tests d'interaction de aRNase J et ASH-Ski2 nous ont permis de suggérer que ASH-Ski2 pourrait être un partenaire du métabolisme des ARN chez *P. abyssi* conjointement avec la ribonucléase aRNase J. Nous avons donc effectué des analyses phylogénomiques pour identifier les membres appartenant au groupe Hélicase ASH-Ski2 appartenant à la famille des hélicases de Ski2-like dans les génomes d'archées séquencés et annotés disponible dans notre banque de données (en collaboration avec l'équipe de G. Fichant, LMGM, Toulouse).

II-3-1. Caractérisation de la famille Ski2-like chez les Archées: 2 groupes Hel308-Ski2 et ASH-Ski2

Le groupe des protéines ASH-Ski2 d'archées appartient à la famille des hélicases Ski2-like retrouvée à la fois chez les eucaryotes et les archées (voir Introduction). Suite à l'identification de la protéine ASH-Ski2 dans le réseau d'interaction de aRNase J, nous avons initié une étude phylogénomique qui a consisté à identifier l'ensemble des protéines de type hélicases dans les 110 génomes d'archées séquencés et annotés (2015) afin de définir leur distribution et leur conservation. Une « Reverse Position Specific BLAST » (RPS BLAST) sur les Pfam correspondant aux domaines RecA1 et RecA2 (DEAD domain : PF00270 & Helicase C : PF00271) du core hélicase dont les scores de chaque résidus ont été calculé. Ainsi une collection de 1059 séquences est obtenue. Les orthologues de chacune des protéines sont recherchés par les liens d'homologie. Les clusters ou classes ont ainsi été définis par le « Markov Cluster Algorithm » (MCL). Cette étude nous a permis d'obtenir une classification en groupes d'orthologues des membres de la superfamille SF2. Parmi ces groupes sont retrouvés le groupe des hélicases RgiI connues chez les eucaryotes pour être impliquées dans la reconnaissance des ARNdb viral (Kowalinski *et al.*, 2011). Le groupe des hélicases Cas3 qui jouent des rôles cruciaux dans les systèmes CRISPR (Jackson *et al.*, 2014). Le groupe des

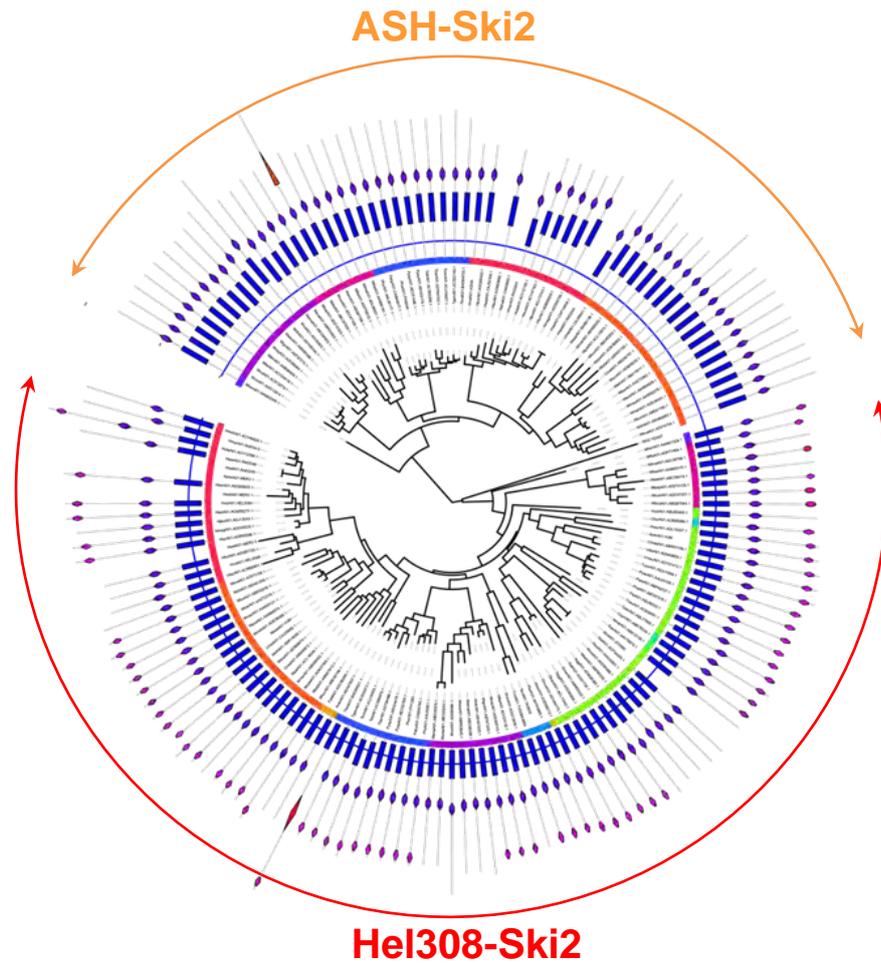
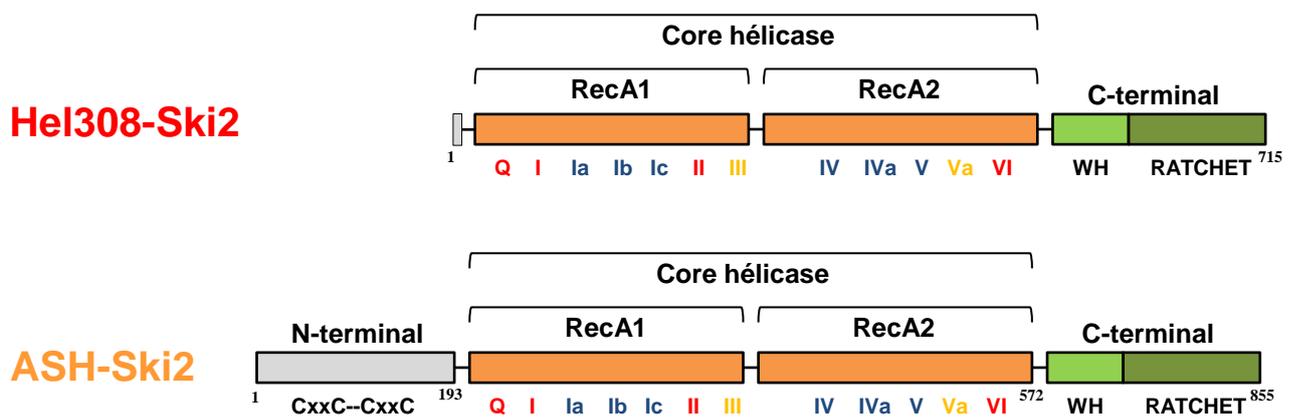
A**B**

Figure 26: Hélicases de la famille Ski2 chez les archées. Etudes phylogénétique autour de ASH-Ski2 et Hel308-Ski2. A: Classification des deux groupes d'hélicase Ski2 chez les archées. Le groupe Hel308-Ski2 est représenté en rouge et ASH-Ski2 en orange. Les Pfam DEAD et Helicase C sont représentés respectivement par des rectangles bleu et par des losanges violets. B: Architecture des hélicases Hel308-Ski2 et ASH-Ski2. Le core SF2 constitué des domaines RecA1 et RecA 2 est représenté en orange. Les extensions N-terminales sont représentées en gris et C-terminales en vert. Les différents motifs impliqués dans la liaison et l'hydrolyse de l'ATP sont en rouge. Ceux impliqués dans la coordination avec les acides nucléiques et les sites de fixation des NTP sont en jaune. En bleu sont représentés les motifs impliqués dans la liaison avec les acides nucléiques. ASH-Ski2 possède en plus des motifs WH et Ratchet typique des Ski2, une extension N-terminale avec un motif zinc finger.

protéines HRQ ainsi que le groupe Lhr divisé en deux sous-groupes aLhr1 et aLhr2, ont des homologues chez les bactéries et n'ont à ce jour pas de fonctions définies. Les groupes des hélicases DEAD-box et RecQ avec certains membres caractérisés chez les archées n'ont aucun représentant codé par le génome de *P. abyssi* (Shimada *et al.*, 2009). Enfin le groupe Ski2-like qui retient notre attention ici est divisé en deux sous-groupes nommés, Hel308-Ski2 et ASH-Ski2. Le génome de *P. abyssi* code pour un représentant de chaque sous-groupe. Ces résultats sont en accord avec l'étude publiée par Chamieh *et al.*, 2016.

Pour caractériser le groupe Ski2-like chez les archées, nous avons construit un arbre phylogénétique avec les membres de ce groupe identifiés (**Figure 26A**) à partir des données issues de notre classification. Le sous-groupe Hel308-Ski2 a été nommé d'après la protéine Hel308 identifiée chez l'humain qui elle-même est l'homologue de la protéine Mus308 retrouvée chez la drosophile. Hel308 humaine et Mus308 font toute deux parties des hélicases SF2 mais ne sont pas classées parmi les 9 sous-familles (Boyd *et al.*, 1990; Takata *et al.*, 2013). L'ensemble de ces protéines est décrit comme impliqué dans le métabolisme de l'ADN et plus particulièrement dans la réparation des cassures double brin. Les membres du groupe ASH-Ski2 sont annotés comme des hélicase à ARN.

Une analyse comparative des séquences des membres des deux sous-groupes nous permet de mettre en évidence les motifs spécifiques de chaque sous-groupe qui pourraient expliquer leur rôle respectif dans la cellule. Pour cela, les séquences obtenues par le MCL issues de la **Figure 26A** correspondant aux gènes codant pour Hel308-Ski2 et ASH-Ski2 ont été sélectionnées. Après alignement par MUSCLE, les résidus des régions peu conservés ou mal alignés ont été enlevés de l'alignement multiple par trimAl. Les résidus conservés sont visualisés avec WebLogo. Les motifs caractéristiques, Motifs Q, I à VI, qui forment le core catalytique des membres de la superfamille SF2, sont visibles (**Figure 27 et 28**). Le domaine C-terminal caractérisant le groupe des hélicases Ski2-like eucaryotes et peu conservé en séquence est retrouvé dans les deux groupes (Johnson and Jackson, 2013). Cette extension C-terminal constituée d'un motif en Winged Helix et d'un domaine Ratchet adopte une structure tridimensionnelle en anneau qui recouvre le core catalytique (**Figure 26B**). Des études ont pu montrer le rôle inhibiteur du domaine C-terminal sur l'activité ATPasique et sur l'activité de déroulement des duplexes d'acides nucléiques de Hel308-Ski2 de *Sulfolobus solfataricus* (Richards *et al.*, 2008). Néanmoins il existe des différences visibles dans l'analyse par Weblogo dans les parties C-terminales des deux sous-groupes. Par exemple le sous-groupe

Core SF2 helicase motifs

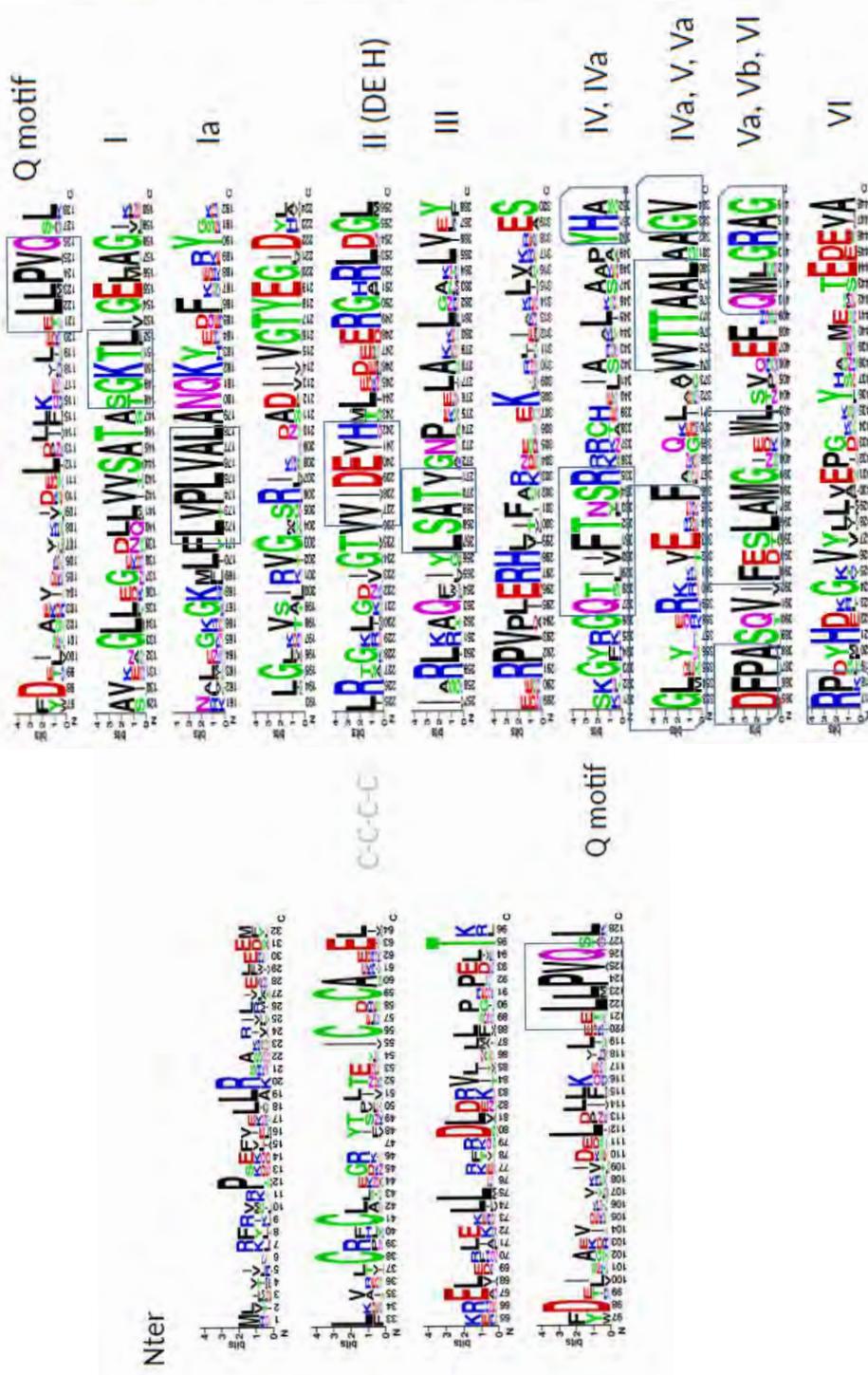
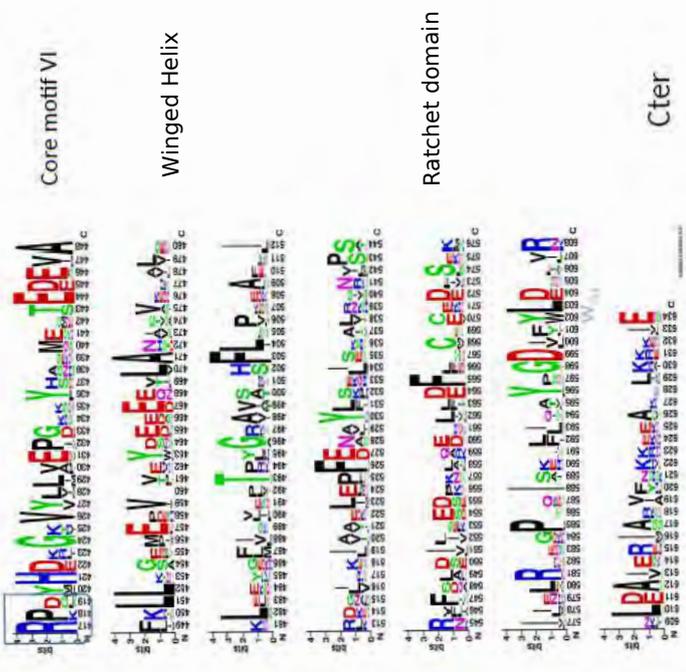


Figure 27: Weblogo de ASH-Ski2. Représentation par WebLogo des différents motifs conservés de la protéine ASH-Ski2. Les séquences obtenues par le MCL correspondant aux gènes codant pour ASH-Ski2 ont été sélectionnées. Après alignement par MUSCLE, les résidus des régions peu conservés ou mal alignés sont enlevés de l'alignement multiple par trimAl et ceux conservés sont visualisés avec WebLogo (<http://weblogo.berkeley.edu/>).



Hel308-Ski2 présente une séquence atypique conservée d'arginine (RxRxR) tandis que le sous-groupe ASH-Ski2 présente une séquence riche en lysine (**Figure 27 et 28**). Ces deux motifs chargés positivement pourraient potentiellement jouer un rôle dans la spécificité d'action des membres respectifs des deux sous-groupes. Toutefois les membres des deux sous-groupes Hel308-Ski2 et ASH-Ski2 ne présentent pas de domaine Arch caractéristique des hélicases Ski2-like, Mtr4 et Ski2, eucaryotes dont les structures sont résolues (Ryan N Jackson *et al.*, 2010; Halbach *et al.*, 2013).

Plus précisément, l'alignement de toutes les séquences des deux sous-groupes Hel308-Ski2 et ASH-Ski2 chez les thermococcales, en utilisant l'outil MUSCLE, permet d'observer des résidus très conservés en dehors des motifs répertoriés (**Figure 27 et 28**). En particulier, les membres du sous-groupe ASH-Ski2 possède une extension N-terminal d'environ 190 acides aminés avec un motif de type « doigt de zinc », CxxC espacés de 22 résidus d'acides aminés strictement conservés (**Figure 26B**). Ce motif est décrit dans d'autres protéines pour être impliqué dans l'interaction avec des acides nucléiques, des protéines ou des lipides (Krishna *et al.*, 2003a). Ce motif pourrait donner la spécificité des membres du sous-groupe ASH-Ski2, reconnaissance des substrats ou d'interaction avec des partenaires protéiques.

II-3-2. Co-distribution taxonomique de ASH-Ski2 et aRNase J

Afin d'observer la distribution des gènes codant pour les membres Hel308-Ski2 et ASH-Ski2 au sein des archées, les Cluster Orthologous Groups (COG) des deux sous-groupes ont été recherchés. Les COG1204 et COG1202 correspondent respectivement aux sous-groupes Hel308-Ski2 et ASH-Ski2. Le sous-groupe Hel308-Ski2 est présent dans la quasi-totalité des génomes d'archées tandis que le sous-groupe ASH-Ski2 est exclusivement présent dans la majorité des Euryarchées à l'exception des Thermoplasmatales et les Archaeoglobales (**Figure 29**). De manière remarquable, nous observons une co-distribution stricte chez les euryarchées du COG0595 correspondant au groupe des ribonucléases aRNase J et du COG 1202 correspondant au sous-groupe ASH-Ski2 avec une absence de représentants de ces deux COG chez les Thermoplasmatales et Archaeoglobales (**Figure 29**). De plus, on peut observer une évolution verticale des membres de ces deux COG. En effet la congruence de l'arbre phylogénétique des archées construits à partir de core génome et respectivement des arbres construits à partir des séquences des groupes ASH-Ski2 et aRNase J permet de proposer une origine ancestral de ces deux gènes (**Figure 30**). La co-distribution et la congruence des arbres sont des éléments qui renforcent le lien observé entre *Pab*-ASH-Ski2 et *Pab*-aRNase J

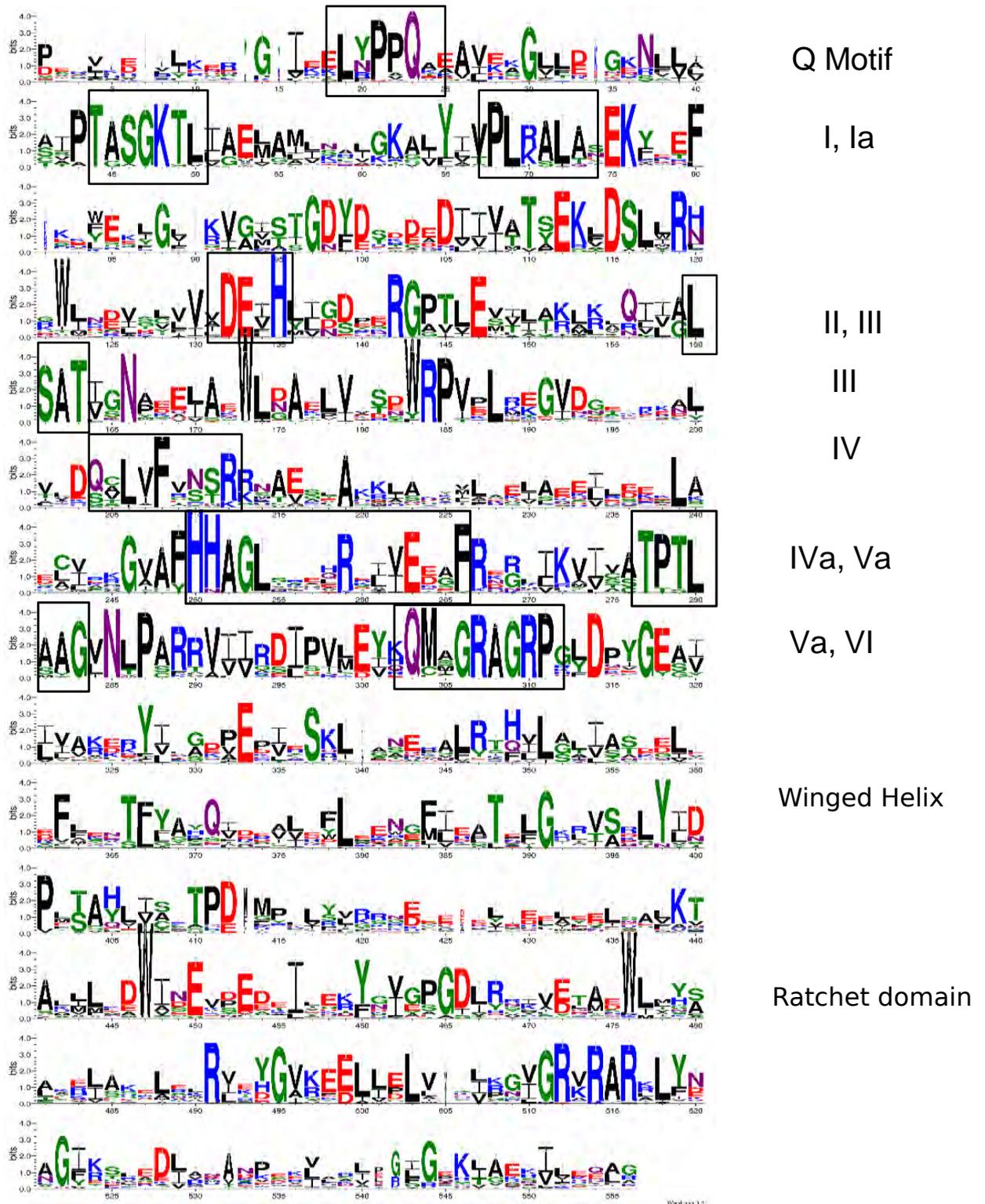


Figure 28: Weblogo de Hel308-Ski2. Représentation par WebLogo des différents motifs conservés de la protéine Hel-308-Ski2. Les séquences obtenues par le MCL correspondant aux gènes codant pour Hel308-Ski2 ont été sélectionnées. Après alignement par MUSCLE, les résidus des régions peu conservés ou mal alignés sont enlevés de l'alignement multiple par trimAl et ceux conservés sont visualisés avec WebLogo (<http://weblogo.berkeley.edu/>).

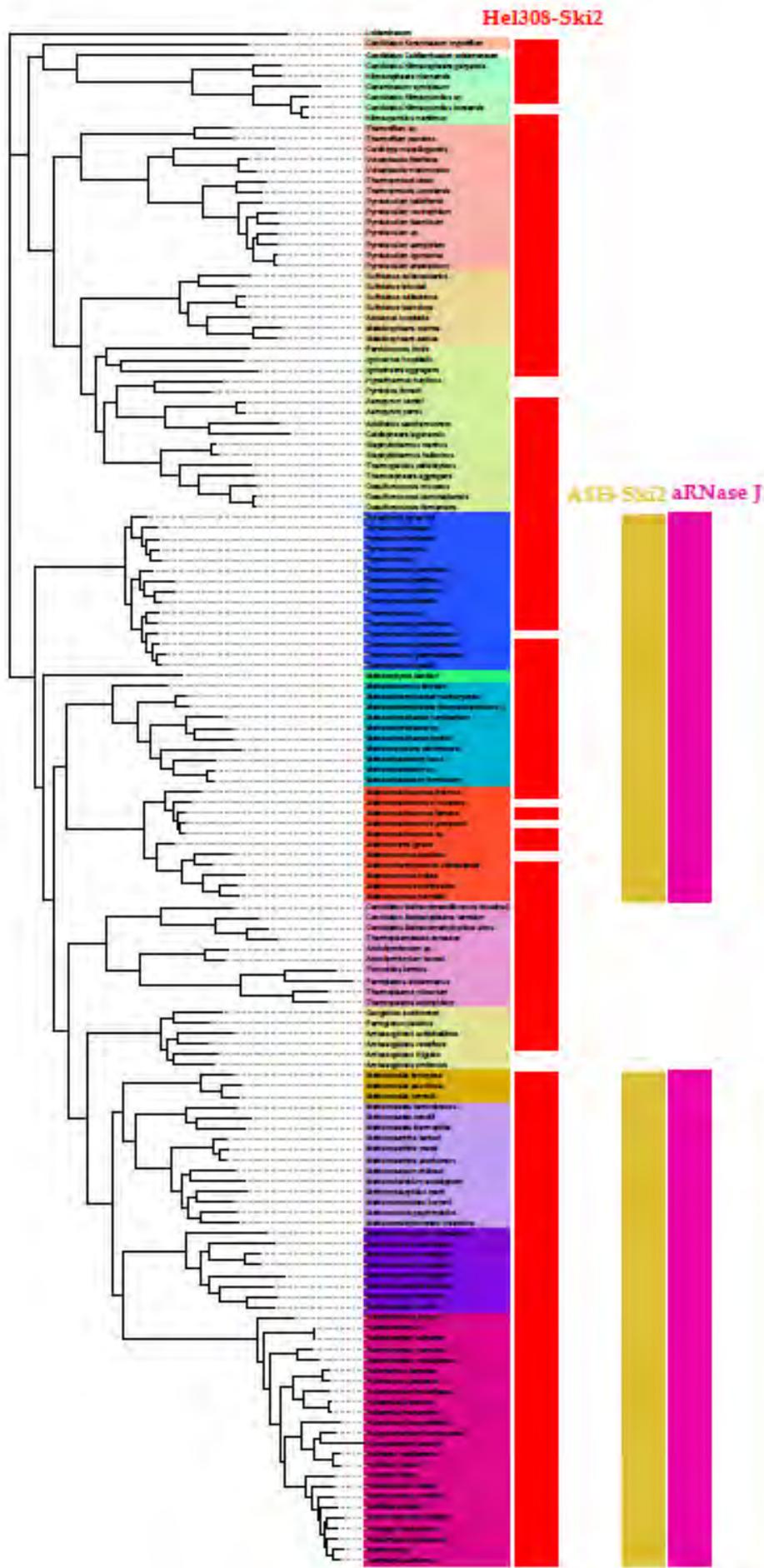
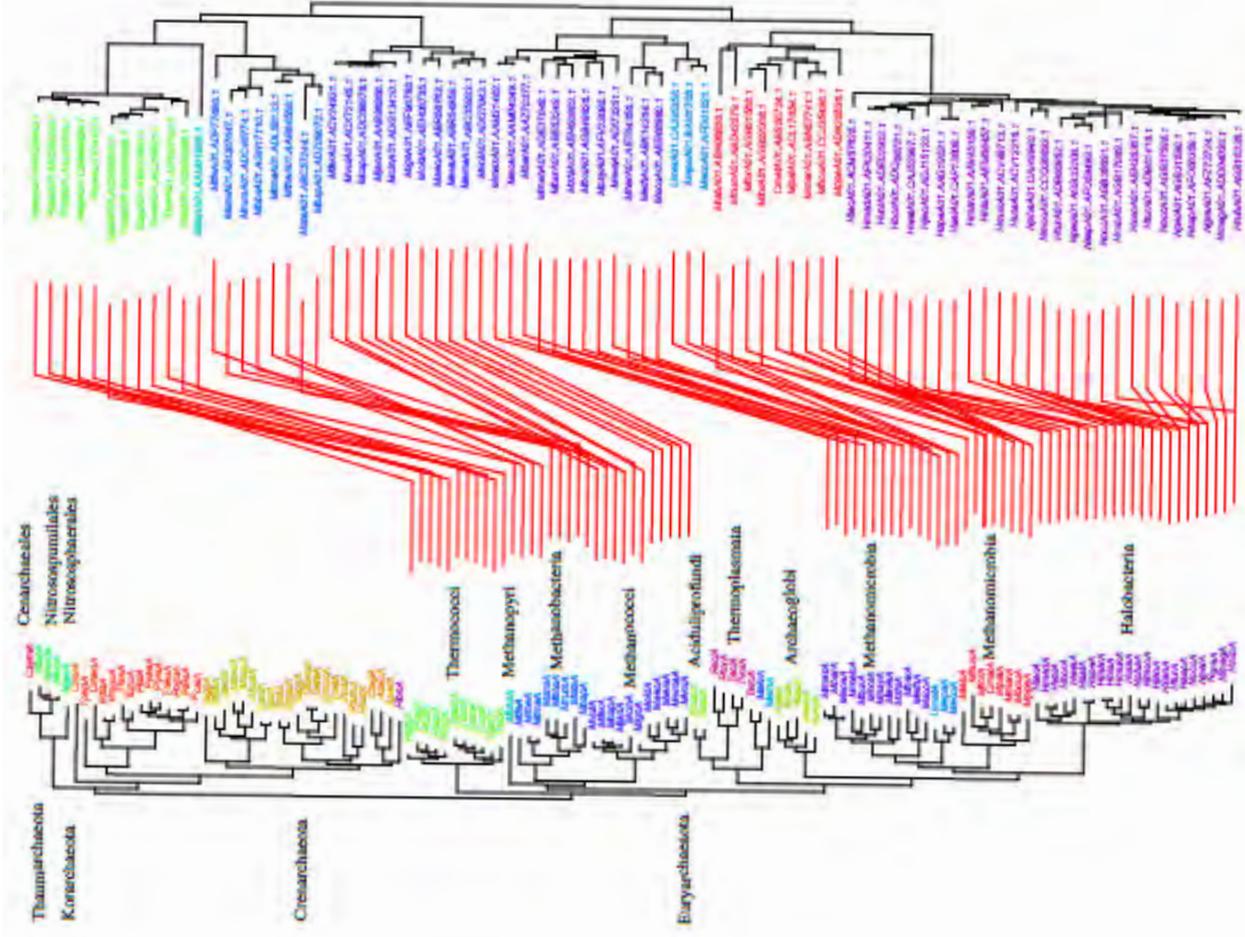


Figure 29: Distributions taxonomiques des gènes codant pour Hel308-Ski2, ASH-Ski2 et aRNase J chez les archées. Les séquences de Hel308-Ski2, ASH-Ski2 et aRNase J sont représenté sur l'arbre phylogénétique des archées construite par alignement de concaténation de 81 groupes orthologues de protéines conservées issues de 131 génomes d'archée séquencés et annotés.

A Archaeal species

aRNase J



B

Archaeal species

ASH-Ski2

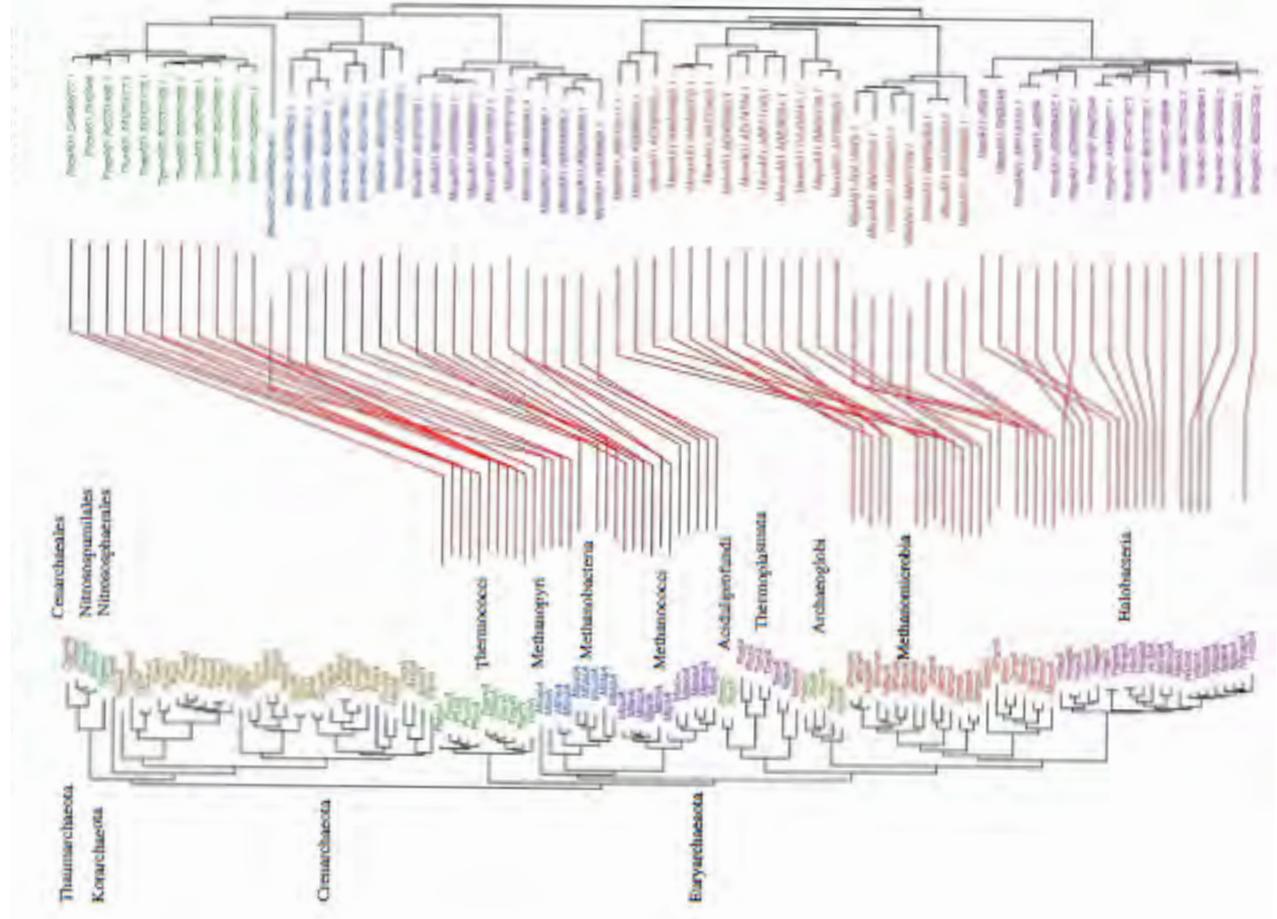


Figure 30: Transfert vertical des gènes codant pour aRNase J et ASH-Ski2. L'ensemble des séquences de ASH-Ski2 et aRNase J ont été alignées et sont représentées sur l'arbre phylogénétique des archées construite par alignement de concaténation de 81 groupes orthologues de protéines conservées issues de 131 génomes d'archée séquencés et annotés. A: Alignement des séquences de aRNase J. B: Alignement des séquences de ASH-Ski2.

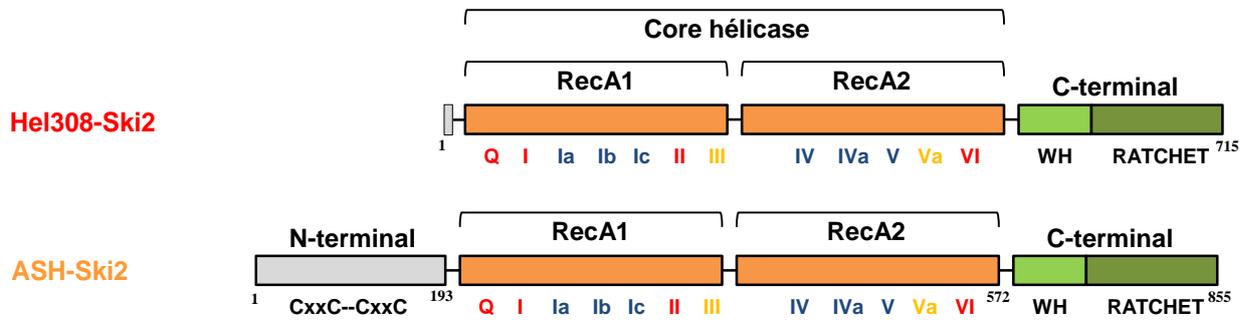
par l'analyse des interactomes, des tests de co-élution et des profils de sédimentation décrits dans les parties ci-dessus (Parties II-1 et II-2)

II-4. Caractérisation de l'interaction entre aRNase J et ASH-Ski2 de *P.abysyi*

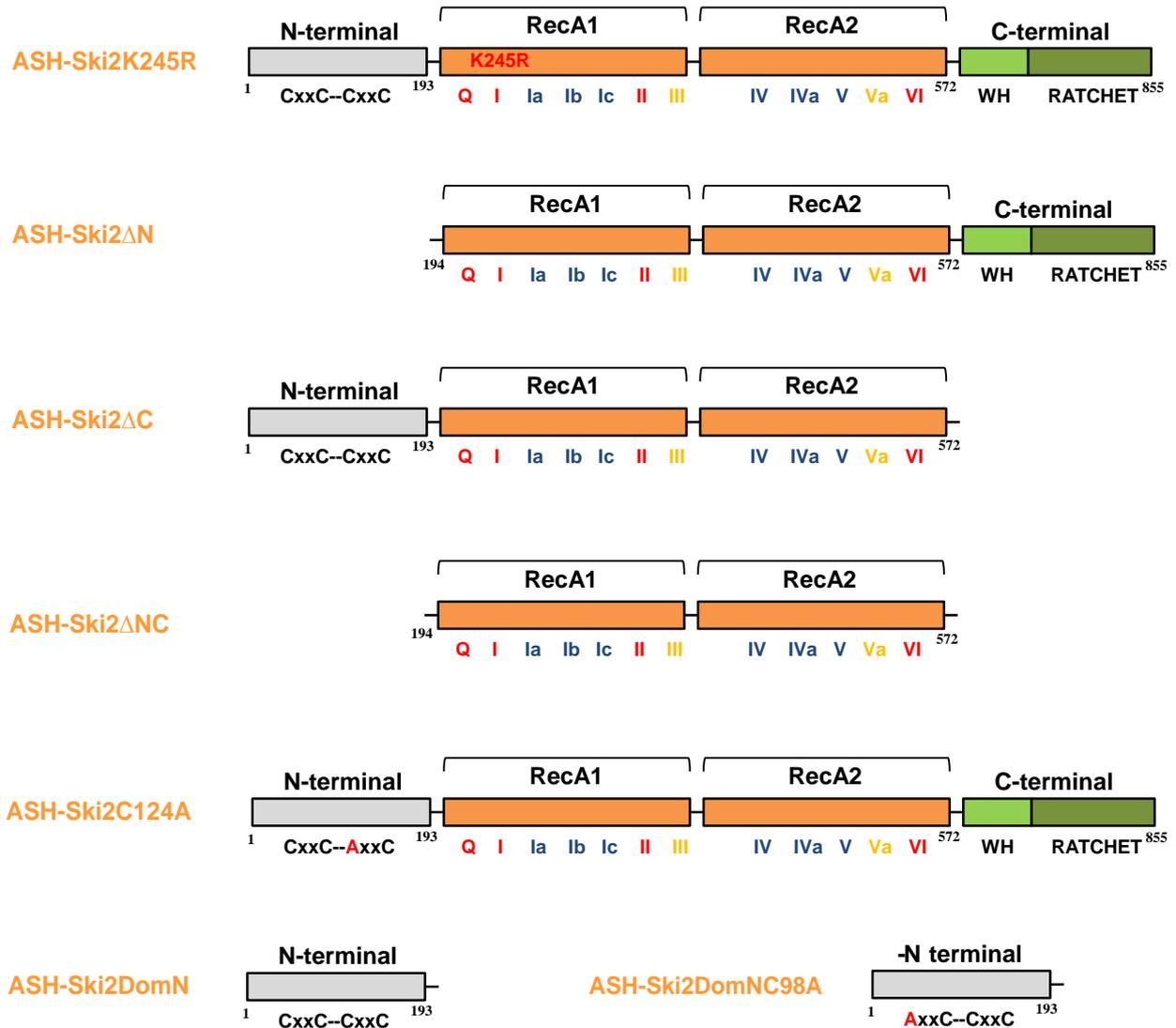
Les expériences de co-purification exposées dans la partie II-2 mettent en évidence une interaction stable entre *Pab*-ASH-Ski2 et *Pab*-aRNase J. Nous nous sommes donc attachés à définir les domaines protéiques impliqués dans cette interaction. Dans un premier temps, nous avons voulu définir les domaines de *Pab*-ASH-Ski2 impliqués dans l'interaction. L'hélicase *Pab*-ASH-Ski2 appartient à la famille d'hélicase Ski2-like (voir Partie II-3). Cette famille est caractérisée par un core catalytique typique SF2, et d'une extension C-terminale constituée d'un domaine Ratchet et d'une région Wing Hélix qui joue le rôle de domaine régulateur (Johnson and Jackson, 2013) (voir ci-dessus Partie II-3- **Figure 26B**). Le groupe ASH-Ski2, conservé chez les Euryarchées, se distingue du groupe Hel308-Ski2 par une extension N-terminale possédant un motif de type « Doigt de Zinc » ou « Zinc Finger » composé de 4 Cystéines (CxxC (x)₂₂ CxxC) strictement conservé dans l'ensemble des hélicases de ce groupe (**Figure 26B**). De manière générale, les motifs de type « Zinc Finger » se structurent autour d'un ion Zinc et sont décrits pour établir des interactions protéine-protéine, protéine-ARN ou -ADN et protéine-lipide (Grishin, 2001; Krishna *et al.*, 2003b).

Pour déterminer l'implication des différents domaines de *Pab*-ASH-Ski2 dans l'interaction stable avec *Pab*-aRNase J, des expériences de co-purification par affinité avec la proie *Pab*-aRNase J sont réitérées avec *Pab*-Hel308-Ski2 et avec des protéines variantes de *Pab*-ASH-Ski2 (**Figure 31, Table 5**). Des vecteurs permettant d'exprimer la protéine Hel308-Ski2 ou des protéines recombinantes variantes de *Pab*-ASH-Ski2 ayant respectivement une délétion du domaine N-terminal (Δ N), du domaine C-terminal (Δ C), ou des domaines N-terminal et C-terminal (Δ NC) ainsi qu'un vecteur exprimant le domaine N-terminal de *Pab*-ASH-Ski2 ont été construits (**Table M1**). De plus, par mutagenèse dirigée, nous avons obtenu un mutant ponctuel permettant de substituer la cystéine C124 du motif de type « Zinc Finger » en Alanine (*Pab*-C124A). Des modélisations structurales montrent que cette mutation ponctuelle C124A suffit à déstructurer le domaine N-terminal de *Pab*-ASH-Ski2 (donnée préliminaire de Manon Bastista, étudiante en thèse)

A



B



aRNase JΔLoop1

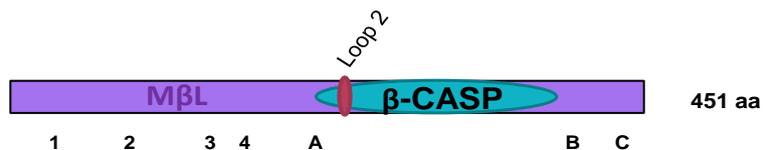


Figure 31: Représentation des protéines sauvages et des variants de ASH-Ski2 et aRNase J. A: Architecture de Hel308-Ski2 et ASH-Ski2 sauvages avec les différents domaines et motifs conservés. B: Représentation des différents variants de ASH-Ski2. La protéine ASH-Ski2 substituée sur la lysine 245 en arginine est nommée ASH-Ski2K245R. La protéine déletée de sa partie N-terminale: ASH-Ski2ΔN, C-terminale: ASH-Ski2ΔC, des deux domaines: ASH-Ski2ΔNC, mutée sur la cystéine C124 en alanine: ASH-Ski2C124A, le domaine N-terminale seul: ASH-Ski2DomN et mutée sur la cystéine 98 en alanine du domaine N-terminale seul: ASH-Ski2DomNC98A. La protéine aRNase J déletée de sa Loop1 est aussi représentée: aRNase JΔL1.

II-4-1. Interaction de *Pab*-aRNase J avec l'hélicase *Pab*-(His)₆-Hel308-Ski2

Nous avons mené des expériences de co-purification en présence du membre du groupe Hel308-Ski2 (**Figure 32A**). La protéine proie *Pab*-aRNase J ne co-élué pas avec *Pab*-(His)₆-Hel308-Ski2 et se retrouve exclusivement dans la fraction FT avec un profil d'élution identique au témoin *Pab*-aRNase J seule (**Figure Annexe1**). Nous proposons donc que *Pab*-aRNase J est capable uniquement d'interagir avec le membre *Pab*-ASH-Ski2 du groupe des hélicases Ski2-like.

II-4-2. Interaction de *Pab*-aRNase J avec les variants de *Pab*-ASH-Ski2

a. Interaction de *Pab*-aRNase J avec le variant *Pab*-(His)₆-ΔC

La protéine appât *Pab*-ASH-Ski2 déléte de sa partie C-terminale est nommée *Pab*-(His)₆-ΔC. Les expériences de co-purification menées avec la proie *Pab*-aRNase J (**Figure 32B**) permettent de visualiser une co-élution indépendante du traitement nucléase comme observée pour la version sauvage de *Pab*-ASH-Ski2 (Figure 3B). Ainsi le domaine C-terminal ne semble pas impliqué dans l'interaction de *Pab*-ASH-Ski2 avec *Pab*-aRNase J dans nos conditions expérimentales.

b. Interaction de *Pab*-aRNase J avec le variant *Pab*-(His)₆-ΔN

La protéine de *Pab*-ASH-Ski2 déléte de sa partie N-terminale est nommée *Pab*-(His)₆-ΔN. A l'opposé du résultat ci-dessus, les profils d'élution de l'appât *Pab*-(His)₆-ΔN et la proie *Pab*-aRNase J ne se superposent pas (**Figure 32C**). Le domaine N-terminal est donc essentiel dans ce contexte à la co-élution des deux protéines. Cela permet de proposer l'existence d'un site d'interaction spécifique dans le domaine N-terminal de *Pab*-(His)₆-ASH-Ski2 permettant de retenir *Pab*-aRNase J sur la colonne d'affinité.

c. Interaction de *Pab*-aRNase J avec le variant *Pab*-(His)₆-ΔNC

Pour tester l'implication du core SF2 de ASH-Ski2 dans l'interaction avec *Pab*-aRNase J, une délétion des deux extensions N- et C-terminales de ASH-Ski2 a été effectuée (*Pab*-(His)₆-ΔNC). De manière surprenante, *Pab*-aRNase J co-élué avec *Pab*-(His)₆-ΔNC indépendamment du traitement aux nucléases (**Figure 32D**). Ce résultat suggère qu'il existerait un deuxième site d'interaction dans le domaine SF2 core, inaccessible en présence du domaine C-terminal, puisque *Pab*-(His)₆-ΔN appât ne co-élué pas *Pab*-aRNase J (**Figure**

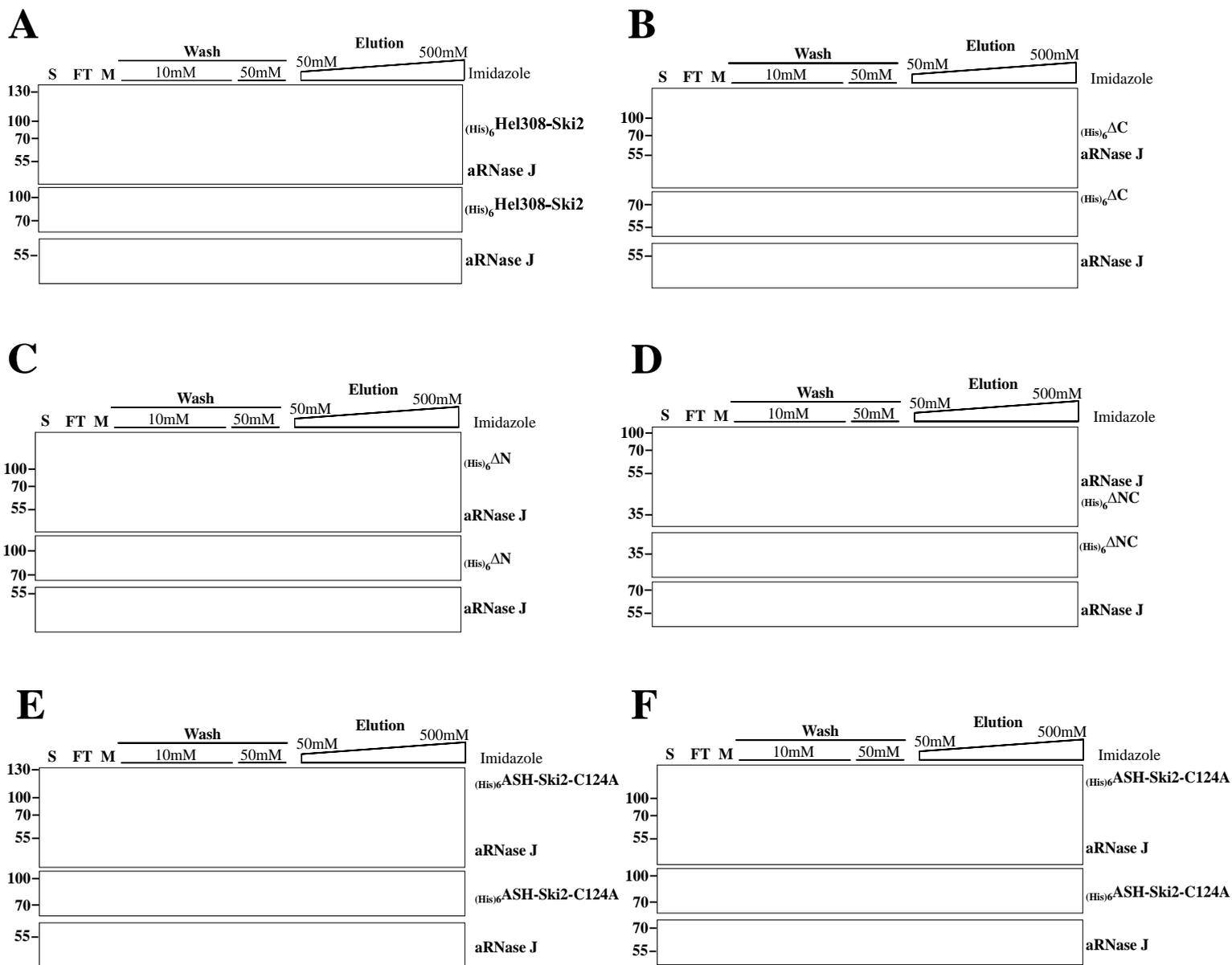


Figure 32: Tests d'interaction de aRNase J avec les différents variants de ASH-Ski2. A: Les extraits ont été préparé comme pour la **Figure 24** avec aRNase J non étiquetée comme proie et Hel308-Ski2 comme appât avec une étiquette en présence de traitement aux nucléases. Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (panel haut) ou par Western blotting (panel bas). B : Co-purification de la protéine sans sa partie C-terminale : ASH-Ski2ΔC étiquetée. C: Co-purification de la protéine sans sa partie N-terminale : ASH-Ski2ΔN étiquetée. D: Co-purification de la protéine sans ces deux extensions N- et C-terminales : ASH-Ski2ΔNC étiquetée. E: Co-purification de la protéine mutée sur la cystéine C124 en alanine: ASH-Ski2C124A étiquetée avec les extraits non traités aux nucléases. F: Co-purification de la protéine mutée sur la cystéine C124 en alanine: ASH-Ski2C124A étiquetée avec les extraits traités aux nucléases.

32C). Le domaine C-terminal pourrait jouer un rôle d'effecteur en modulant la capacité d'interaction de l'hélicase ASH-Ski2 avec ses partenaires. De manière intéressante, comme montré dans la partie III de ce travail, le domaine C-terminal est également capable de moduler l'activité enzymatique de *Pab*-ASH-Ski2.

d. Interaction de *Pab*-aRNase J avec le variant *Pab*-(His)₆-ASH-C124A

Pour finir, nous avons testé l'interaction de *Pab*-aRNase J avec la protéine mutée *Pab*-(His)₆-C124A. Dans ce contexte, *Pab*-aRNase J est retrouvée en faible quantité dans les éluions sans traitement RNase/DNase et est absent après traitement (**Figure 32E et F**). La cystéine C124 de *Pab*-ASH-Ski2 est donc importante pour la co-élution avec *Pab*-aRNase J. On peut donc proposer au vu de la conservation du résidu C124 du motif (CxxC (x)₁₄ CxxC) que la formation d'un motif de type « Zinc Finger » serait critique pour l'interaction de *Pab*-ASH-Ski2 avec la ribonucléase *Pab*-aRNase J

II-4-3. Interaction de *Pab*-aRNase J avec le domaine N-terminal de *Pab*-ASH-Ski2

Nos résultats précédents suggèrent fortement que le domaine N-terminal de *Pab*-ASH-Ski2 est important pour l'interaction avec *Pab*-aRNase J. Afin de confirmer ce résultat, nous avons construit un vecteur d'expression permettant d'exprimer le domaine N-terminal de *Pab*-ASH-Ski2 seul (*Pab*-DomN). Dans le but de déterminer si le domaine N-terminal (*Pab*-(His)₆-DomN) seul est suffisant pour établir une interaction avec la protéine proie *Pab*-aRNase J, nous avons réalisé une co-purification par affinité comme décrit précédemment (**Figure 33A**). De manière remarquable, nous observons que *Pab*-aRNase J co-élue avec *Pab*-(His)₆-DomN et ceci même après traitement aux RNase/DNase. Ce résultat indique que le domaine N-terminal de *Pab*-ASH-Ski2 seul est suffisant pour établir une interaction stable avec *Pab*-aRNase J. Ce résultat est cohérent avec nos résultats précédents.

Nous avons également obtenu un mutant ponctuel C98A dans *Pab*-DomN. Cette position est la première cystéine conservée du motif type Zinc Finger (**Figure 31B**). Dans ce cas nous observons que *Pab*-aRNase J co-élue avec *Pab*-(His)₆-DomN-C98A que ce soit en présence ou en absence de traitement aux nucléases (**Figure 33B**). Ces données indiquent que la substitution de la cystéine dans le contexte du domaine N-terminale seul n'est pas suffisante pour altérer l'interaction avec *Pab*-aRNase J.

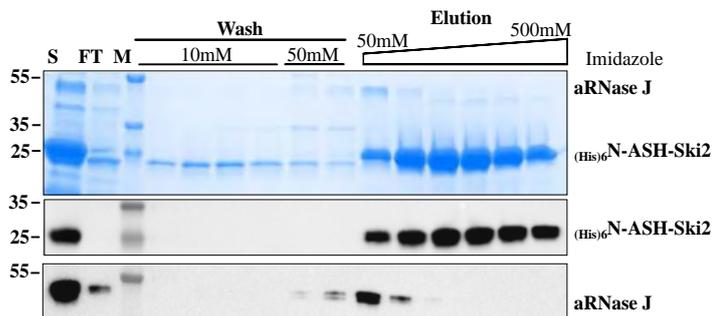
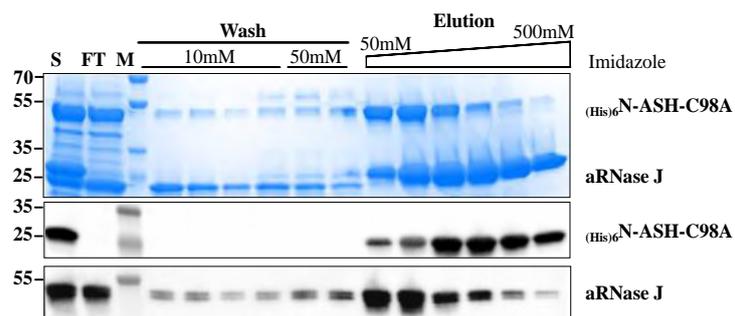
A**B**

Figure 33: Tests d'interaction de aRNase J avec le domaine N-terminal de ASH-Ski2. A: Les extraits ont été préparé comme pour la **Figure 24** avec aRNase J non étiquetée comme proie et le domaine N-terminal de ASH-Ski2 comme appât avec une étiquette en présence de traitement aux nucléases. Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (panel haut) ou par Western blotting (panel bas). B : Co-purification de la protéine mutée sur la cystéine 98 en alanine du domaine N-terminale seul: ASH-Ski2DomNC98A étiquetée.

II-4-4. Interaction de *Pab*-ASH-Ski2 avec le variant de *Pab*-aRNase J

Les analyses phylogénomiques qui ont permis de caractériser le groupe aRNase J dans l'ensemble des Euryarchées couplées à l'étude de l'activité enzymatique de *Pab*-aRNase J (voir Intro et Clouet d'Orval *et al.*, 2010) ont pu montrer que deux insertions nommées Loop1 et Loop2 respectivement dans le core β -lactamase et dans le core β -CASP sont l'originalité de ce groupe de ribonucléases de la famille β -CASP. Afin d'identifier l'importance de ces Loop dans l'interaction avec ASH-Ski2 montrée dans les expériences ci-dessus, nous avons testé la capacité de co-élution de la proie *Pab*-aRNase J Δ loop1 (*Pab*-aRNase J Δ 1) avec *Pab*-(His)₆-ASH-Ski2 (**Figure 31B et 34**). Ainsi, lorsque nous purifions les deux protéines ensemble en absence de traitement aux nucléases, *Pab*-aRNase J Δ 1 est retrouvée co-éluee avec *Pab*-(His)₆-ASH-Ski2 (**Figure 34A**). De manière intéressante, lorsque nous traitons l'extrait aux nucléases, *Pab*-aRNase J Δ 1 n'est seulement présente dans les fractions FT et n'est pas co-éluee avec *Pab*-(His)₆-ASH-Ski2 (**Figure 34B**). Ces résultats indiquent qu'en absence de la Loop1, aRNase J ne semble être capable d'interagir avec ASH-Ski2 que seulement lorsque les acides nucléiques sont encore présents dans les extraits. La Loop1 de aRNase J pourrait jouer ainsi un rôle crucial dans l'interaction stable avec ASH-Ski2.

II-4-5. Le complexe *Pab*-ASH-Ski2/*Pab*-aRNase J identifié *in vitro* par sédimentation sur gradient de saccharose

L'ensemble des résultats décrits ci-dessus permettent de proposer que *Pab*-aRNase J et *Pab*-ASH-Ski2 établissent une interaction protéine-protéine spécifique qui implique au moins deux sites d'interaction de la protéine *Pab*-ASH-Ski2, l'un présent dans le domaine core SF2 masqué par le domaine C-terminal, et le second présent dans le domaine N-terminal qui contient une série de cystéines conservées dont la cystéine C124 qui formeraient un motif de type « Zinc Finger ». L'interaction observée semble ne pas être médiée par des acides nucléiques. Nous avons pu montrer également que dans certains cas la présence de l'étiquette 6X HIS interfère dans l'établissement de l'interaction protéine-protéine. De plus, une étape d'incubation à 70°C est requise pour observer une co-élution des deux protéines. La haute température pourrait favoriser une ou des conformations protéiques spécifiques. Ce type de changement de conformation a déjà été observé et décrit pour la protéine Ski2 du complexe SKI chez les eucaryotes dont le recrutement sur la sous-unité 40S du ribosome nécessite le passage d'un état fermé à celui d'ouvert afin de permettre son activation (Schmidt *et al.*, 2016).

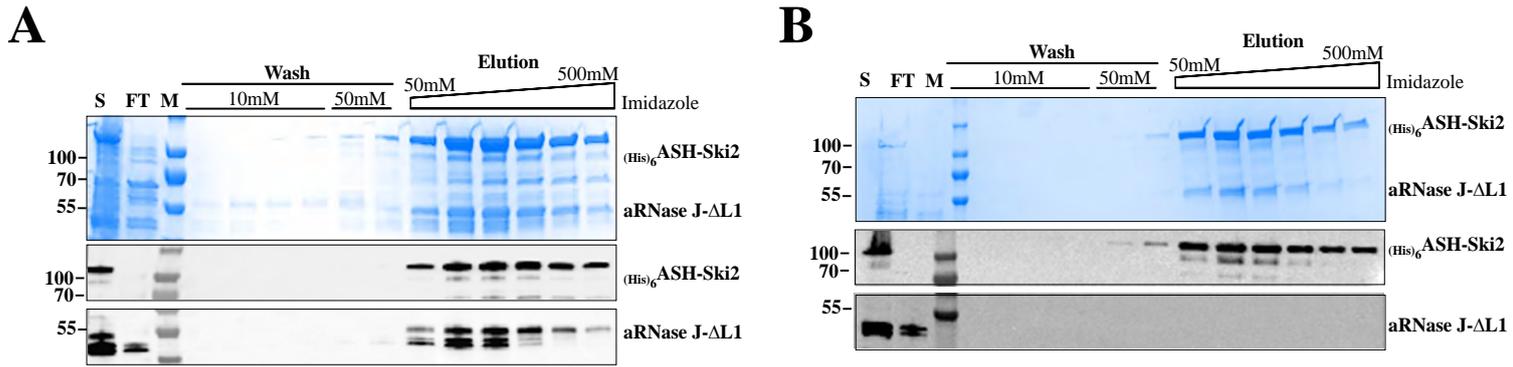


Figure 34: Rôle de la Loop1 de aRNase J dans l'interaction avec ASH-Ski2. A: Les extraits ont été préparé comme pour la **Figure 24** avec aRNase J déléetée de sa Loop1 (aRNase JΔL1) non étiquetée comme proie et ASH-Ski2 comme appât avec une étiquette en absence de traitement aux nucléases. Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (panel haut) ou par Western blotting (panel bas). B : Co-purification de la protéine aRNase J déléetée de sa Loop1 (aRNase JΔL1) non étiquetée comme proie et ASH-Ski2 comme appât avec une étiquette en présence de traitement aux nucléases.

En parallèle de ces expériences de co-purification des protéines proie-appât, les profils de sédimentation des protéines recombinantes *Pab*-aRNase J et *Pab*-ASH-Ski2 purifiées seules ont été déterminés sur gradient 5-25% sucrose (**Figure 35A**). Le profil de *Pab*-ASH-Ski2 correspond à sa forme monomérique et celui de aRNase J correspond à des formes multimériques difficiles à identifier (**Figure 35A**). Lorsque les deux protéines recombinantes purifiées *Pab*-aRNase J et *Pab*-ASH-Ski2 (ratio 2 pour 1) sont mélangées et déposées sur le gradient après une étape de chauffage 70°C, nous observons que le profil du mélange diffère de celui des profils des protéines seules (**Figure 35B**). En particulier la présence de *Pab*-aRNase J affecte le profil de sédimentation de *Pab*-ASH-Ski2 qui se retrouve déplacé vers la droite du gradient dans les fractions 6 à 13 avec des vitesses de sédimentation plus rapide et un pic dans la fraction 8. Ces résultats confortent nos résultats de co-purification des protéines proie-appât et de phylogénomique montrant une co-distribution taxonomique des deux protéines chez les euryarchées (**Figure 25C et 29**).

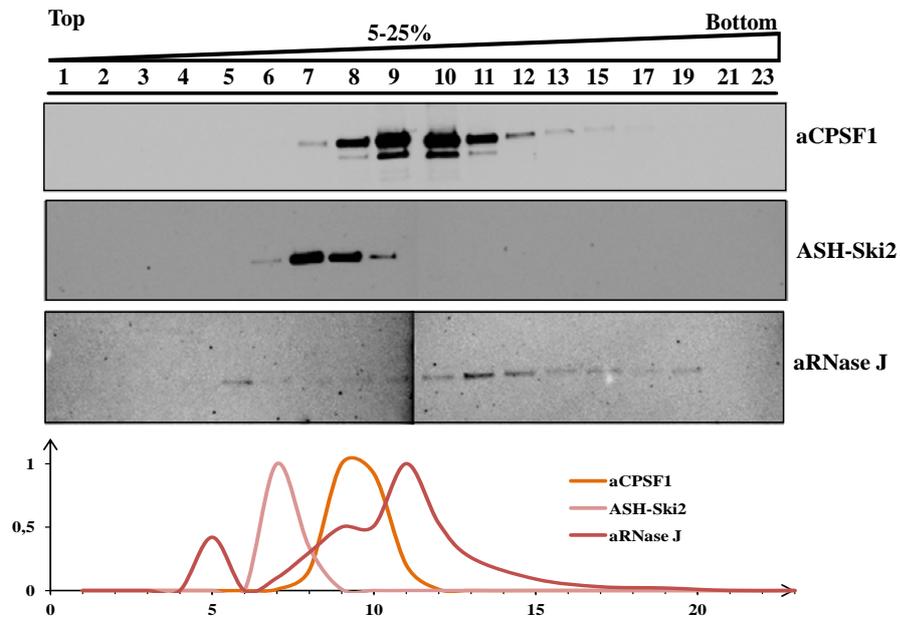
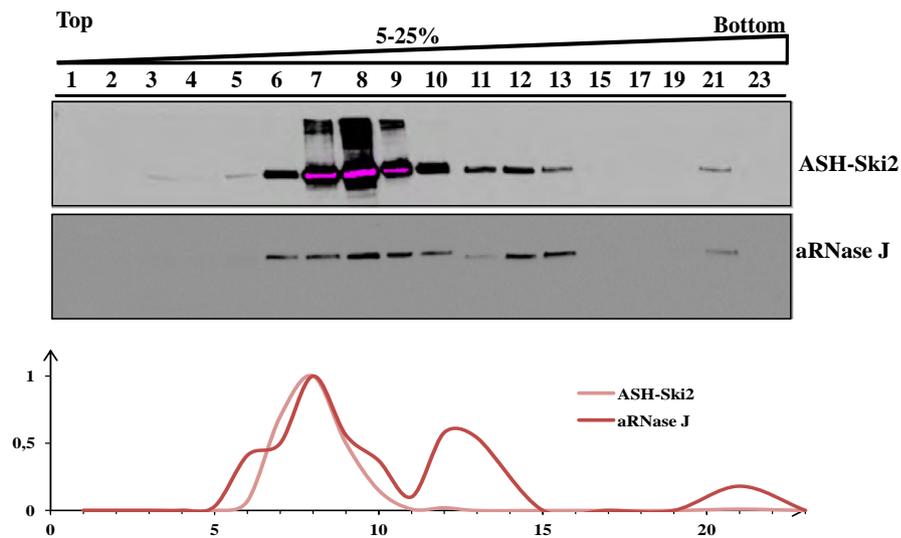
A**B**

Figure 35: Localisation des protéines aCPSF1, aRNase J et ASH-Ski2 recombinantes purifiées seules ou mélangées sur gradient de sucrose 5-25% A: Fractionnement séparé des protéines aCPSF1, aRNase J et ASH-Ski2 sur des gradients de sucrose 5-25% PBS 1X et ultra-centrifugation à 35000 rpm durant 20h. Western blotting avec anti-corps anti-aCPSF1, ASH-Ski2 et aRNase J (panel haut). Gel agarose 1% TBE1X des extractions ARN (panel bas). B: Fractionnement de protéines recombinantes ASH-Ski2 mélangées avec aRNase J (ratio 1/2) dans les mêmes conditions que A.

PARTIE III

ACTIVITES ENZYMATIQUES DES HELICASES DE TYPE Ski2-like DE *P. abyssi*, *Pab-ASH-Ski2* ET *Pab-Hel308-Ski2*

Pour définir les activités enzymatiques du nouveau partenaire de *Pab*-aRNase J, *Pab-ASH-Ski2*, nous avons entrepris des études enzymatiques pour juger de son activité ATPasique (hydrolyse d'ATP en ADP et phosphate inorganique (Pi)), sa capacité de liaison aux acides nucléiques (ARN ou ADN sous forme simple brin ou en duplex) et d'appariement de simple brin et déroulement de duplex d'acides nucléiques. De plus, afin de comparer les caractéristiques enzymatiques des deux hélicases de type Ski2-like (Voir Partie 1-3) présentes chez *P. abyssi*, nous avons testé en parallèle les activités enzymatiques *in vitro* de la protéine *Pab-Hel308-Ski2*. Cette comparaison permet d'initier une étude structure-fonction de la famille Ski2-like chez les archées et d'appréhender leur fonction cellulaire respective. Dans cette étude nous avons également utilisé la panoplie de variant protéiques de *Pab-ASH-Ski2* déjà mentionnés dans l'étude de l'interaction entre *Pab-ASH-Ski2* et *Pab*-aRNase J (voir Partie II-4 de ce manuscrit) pour identifier l'importance du domaine core hélicase SF2, du domaine N-terminal spécifique du groupe ASH-Ski2 et du domaine C-terminal typique de la famille Ski2-like et commun à *Pab-ASH-Ski2* et *Pab-Hel308-Ski2*. Les protocoles que nous avons utilisés, sont basés sur ceux décrits dans la caractérisation enzymatique de Hel308-Ski2 de *Methanothermobacter thermotrophicus* (*Mther-Hel308-Ski2*) (Guy and Bolt, 2005). Nous avons donc choisi des substrats et conditions identiques pour pouvoir comparer nos résultats avec les données publiées. Pour la suite, chaque protéine testée a été surexprimée à partir d'un vecteur pET15b dans des cellules de *E. coli* et purifiée à homogénéité après deux chromatographies d'affinité successives, HisTrap et Héparine. Les activités enzymatiques sont donc déterminées pour des protéines possédant une étiquette (His)₆ en N-terminale.

III-1. ACTIVITE ATPASIQUE DE *Pab-ASH-Ski2* ET *Pab-Hel308-Ski2*

III-1-1. *Pab-ASH-Ski2* et *Pab-Hel308-Ski2* sont des ATPases dépendantes de la présence d'acides nucléiques

a. *Pab-ASH-Ski2*

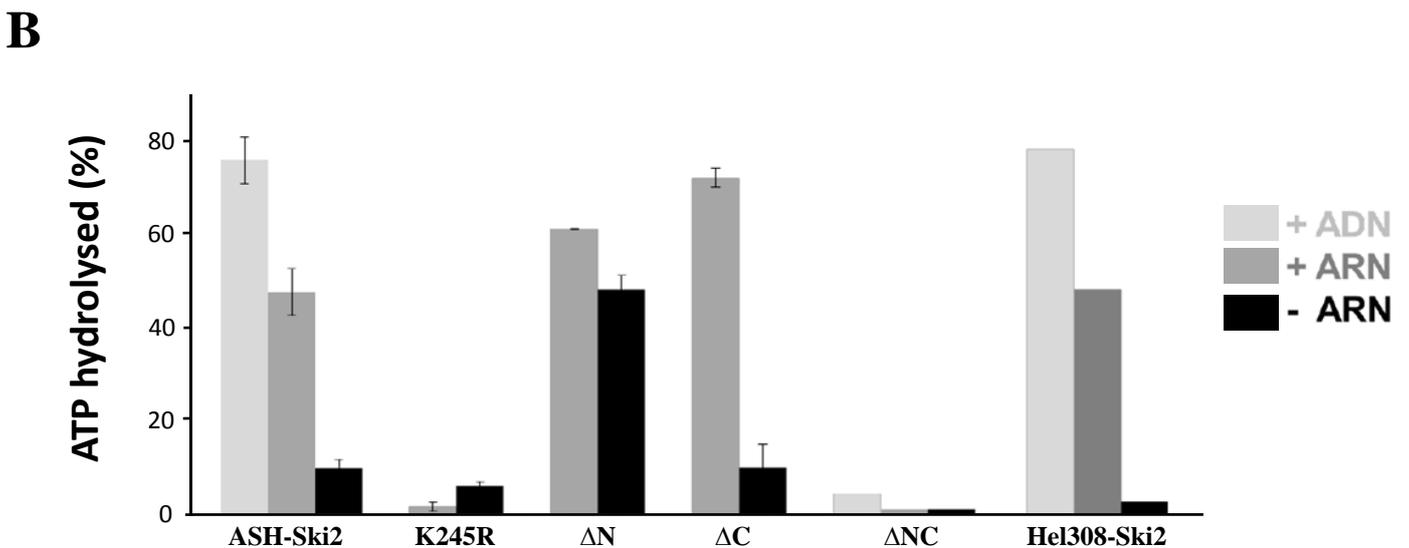
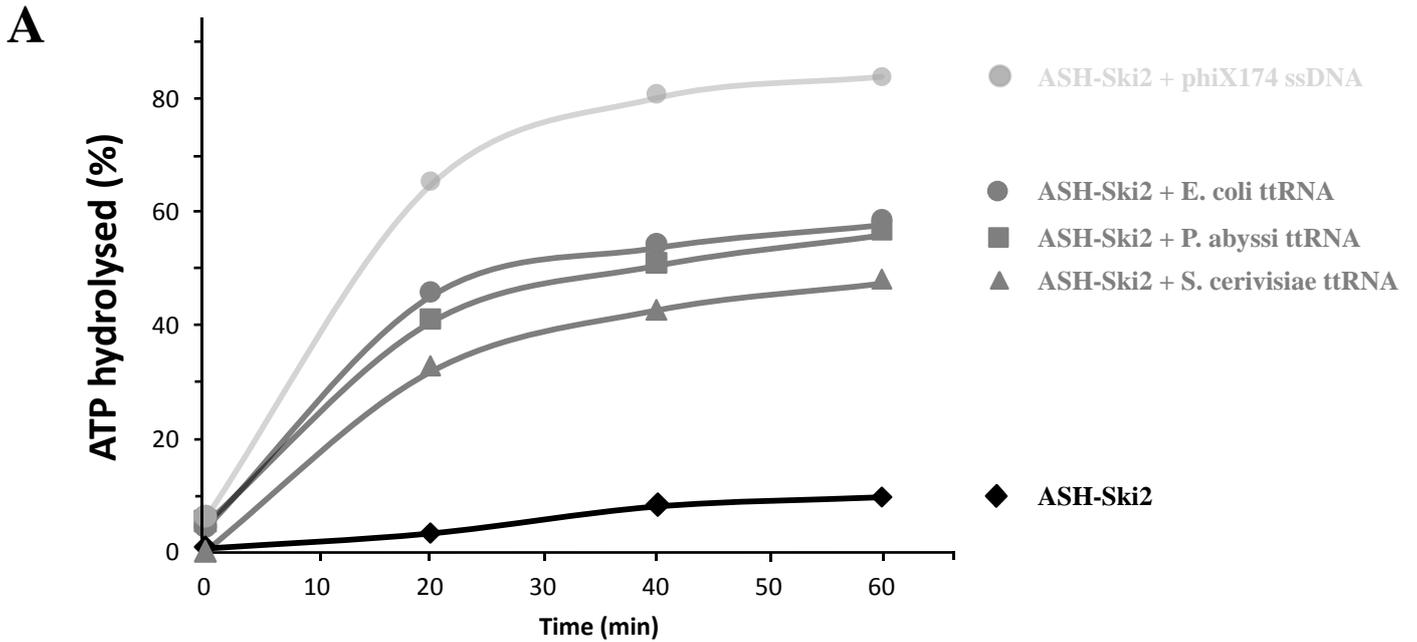


Figure 36: Tests ATPasique de ASH-Ski2, ces variants et Hel308-Ski2. A: Activité ATPasique de ASH-Ski2 sauvage testée avec 250nM de protéines avec de l'ADN simple brin de phiX174, de l'ARN totaux de *P. abyssi*, de *S. cerevisiae* et d'*E. coli* à 65°C et migré sur TLC. Quantification par MultiGauge. B: Activité testées avec différents variants de la protéine ASH-Ski2 et Hel308-Ski2. ASH-Ski2 sauvage (WT), mutée sur le site catalytique par substitution de la Lysine en position 245 en Arginine (K245R), délétée du domaine N-terminal (ΔN), délétée du domaine C-terminal (ΔC) et délétée des deux domaines N et C-terminal (ΔNC).

Dans un premier temps, une cinétique d'hydrolyse de l'ATP [$\gamma^{32}\text{P}$] radio marqué en présence de *Pab*-ASH-Ski2 à une concentration de 250nM avec ou sans ARN a été réalisée à 65°C dans un tampon Hepes-KCL (50mM Hepes pH7.5, 50mM KCl, 5mM MgCl₂, 2mM DTT, 2mM ATP, 0,075 μl [$\gamma^{32}\text{P}$] ATP à 6000Ci/mmol en absence ou en présence de 0,2mg/mL d'acides nucléiques). L'apparition de phosphate inorganique au cours du temps est quantifiée après migration sur chromatographie sur couche mince (TLC). On observe dans ce test préliminaire qu'en absence d'ARN ou ADN, *Pab*-ASH-Ski2 a une activité ATPasique résiduelle faible. La très faible proportion de phosphate inorganique apparaissant au cours du temps peut être liée à une dégradation intrinsèque de l'ATP à haute température 65°C. La capacité d'hydrolyse de l'ATP de *Pab*-ASH-Ski2 apparait clairement en présence d'ADN simple brin viral de phiX174 ou d'ARN totaux quel que soit son origine *E. coli* ou de *S. cerevisiae* ou de *P. abyssi* (**Figure 36A**). On observe l'hydrolyse de 50% de l'ATP au bout de 60 min en présence d'une même concentration de 0.2mg/mL d'ARN totaux et de 70% en présence de l'ADN simple brin (**Figure 36A**). Une différence d'activité ATPasique en présence d'ARN et d'ADN est observée. Cette différence est difficilement interprétable au vu de la nature des acides nucléiques totaux utilisé lors de ces tests et l'absence de réplica. Ainsi, la liaison aux acides nucléiques doit favoriser la formation de la conformation active de l'enzyme pour l'hydrolyse de l'ATP.

b. *Pab*-Hel308-Ski2

Nous avons testé l'activité de la protéine *Pab*-Hel308-Ski2 à 250nM dans les mêmes conditions (**Figure 36C**). De manière homogène à *Pab*-ASH-Ski2, *Pab*-Hel308-Ski2 possède une activité ATPasique dépendante de la présence d'acides nucléiques. Ainsi, les hélicases Ski2-like de *P. abyssi* ne semble pas présenter de spécificité vis à vis de l'ARN ou ADN simple brin *in vitro*. L'interaction avec les acides nucléiques qui active l'activité ATPasique s'effectue indépendamment de la présence de ribose ou désoxyribose. Ces résultats rejoignent certaines études montrant une interaction directe entre l'hélicase et le squelette phosphate des acides nucléiques (Kammel *et al.*, 2013; Tuteja *et al.*, 2014).

III-1-2. Activité ATPasique des variants de *Pab*-ASH-Ski2

Les extensions N et C-terminales des hélicases de la superfamille SF2 sont importantes pour l'interaction avec leurs partenaires, la modulation de leurs activités enzymatiques et pour leur spécificité de substrats (Matsuno *et al.*, 2006; He *et al.*, 2010; Gao *et al.*, n.d.). Nous avons donc obtenu des protéines variantes de *Pab*-ASH-Ski2 mutée sur le site catalytique (K245R),

tronquée en N-terminale (ΔN), en C-terminale (ΔC) ou en N- et C-terminales (ΔNC) (**Figure 31B**) que nous avons purifié à homogénéité et utilisé dans nos tests enzymatiques. La comparaison des activités est réalisée avec les ARN totaux de *P. abyssi*. Chaque test est réitéré trois fois pour obtenir un écart-type visible sur le graphique (**Figure 36C**).

a. *Pab-ASH-Ski2K245R*

Nous avons construit un mutant ponctuel dans le motif I typique du core hélicase SF2 à la position 245 (**Figure 31B**) avec une substitution d'une lysine en arginine (K245R). Cette mutation est connue pour être essentielle à l'activité d'hydrolyse de l'ATP pour les hélicases SF2 RAD3, eIF4A ou encore Snf2 eucaryotes (Sung *et al.*, 1988; Pause and Sonenberg, 1992; Laurent *et al.*, 1993). Ce variant protéique ne présente pas d'activité d'hydrolyse d'ATP que ce soit en absence ou en présence d'acides nucléiques (**Figure 36C**). L'analyse de ce variant permet de confirmer l'importance du motif I caractéristique des hélicases de la superfamille SF2 dans l'activité enzymatique de *Pab-ASH-Ski2* et la spécificité de la cinétique d'hydrolyse de l'ATP observé pour *Pab-ASH-Ski2* sauvage.

b. *Pab-ASH-Ski2- ΔN*

Dans le cas de la protéine *Pab-ASH-Ski2 ΔN* où le domaine N-terminale (spécifique du groupe ASH-Ski2 voir Partie II-3) est absent, on observe une légère augmentation de l'hydrolyse de l'ATP en présence d'ARN après 20 min d'incubation par rapport à la protéine sauvage (**Figure 36C**). De manière surprenante, l'activité ATPasique de ce variant est indépendante de la présence d'acides nucléiques. En absence du domaine N-terminal, l'enzyme est déjà dans une conformation active pour l'hydrolyse de l'ATP sans avoir besoin de liaison aux acides nucléiques.

c. *Pab-ASH-Ski2- ΔC*

En absence du domaine C-terminal une augmentation significative de l'hydrolyse de l'ATP en présence d'ARN totaux de *P. abyssi* est observée. Le domaine C-terminal de *Pab-ASH-Ski2* a donc un effet inhibiteur sur l'hydrolyse de l'ATP dans ces conditions (**Figure 36C**). Ce résultat est cohérent avec l'étude enzymatique de Hel308-Ski2 de *Sulfolobus solfataricus* (*Sso-Hel308-Ski2*) où le domaine C-terminal est montré avoir également un effet inhibiteur sur l'activité ATPasique (Woodman *et al.*, 2011).

d. *Pab-ASH-Ski2-ΔNC*

En présence de la protéine *Pab-ASH-Ski2 ΔNC* tronquée à la fois pour les extensions N- et C-terminal, aucune activité ATPasique n'est observée (**Figure 36C**). Ainsi le domaine core hélicase SF2 par lui-même est incapable d'hydrolyser l'ATP bien qu'il porte les motifs de reconnaissance et d'hydrolyse de l'ATP. En tant que tel *Pab-ASH-Ski2 ΔNC* reste dans une conformation inactive.

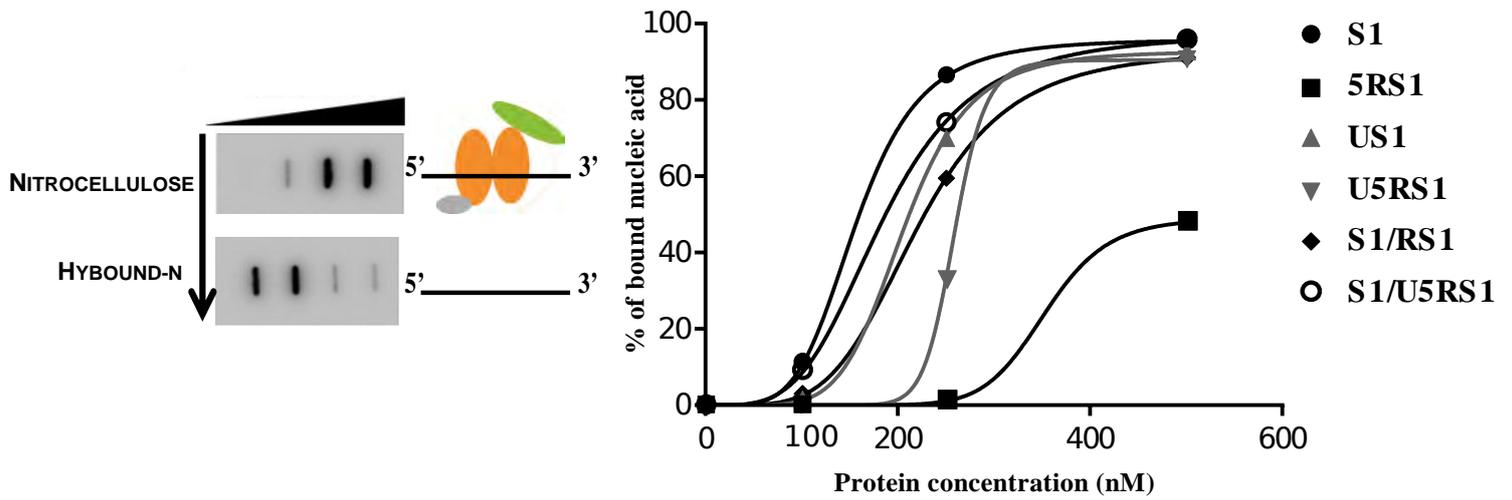
En conclusion, les résultats ci-dessus montrent que *Pab-ASH-Ski2* possède une activité ATPasique dépendante de la liaison aux acides nucléiques. L'intégrité du motif I (liaison et hydrolyse de l'ATP) caractéristique des hélicases SF2 est requise pour l'hydrolyse de l'ATP. L'extension N-terminale ne semble pas impliquer directement dans l'activité catalytique mais serait importante pour moduler l'activité de *Pab-ASH-Ski2* en fonction de la présence d'acides nucléiques. Cependant de manière intéressante *Pab-Hel308-Ski2* qui ne possède pas d'extension N-terminale est inactive en absence d'acides nucléiques contrairement à *Pab-ASH-Ski2ΔN*. Ainsi, *Pab-ASH-Ski2ΔN* et *Pab-Hel308-Ski2* présente des comportements divergents. L'extension C-terminale de *Pab-ASH-Ski2* quant à elle semble jouer un rôle modulateur ou de frein pour l'activité d'hydrolyse de l'ATP en cohérence avec l'étude de *Sso-Hel308-Ski2* qui révèle le rôle auto-inhibiteur ou de frein moléculaire du domaine C-terminale adjacent au domaine core hélicase SF2 en le recouvrant dans la structure tridimensionnelle (Jodi D. Richards *et al.*, 2008) (**Figure 26B**).

III-2. AFFINITE DE *Pab-ASH-Ski2* ET *Pab-Hel308-Ski2* AUX ACIDES NUCLEIQUES

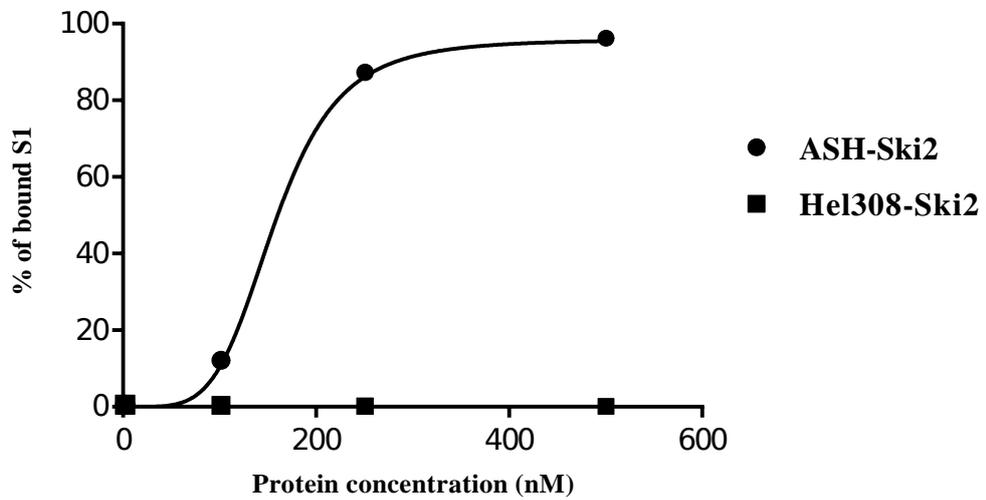
III-2-1. Méthode de mesure de l'affinité aux acides nucléiques par la technique de double filtration

Différents types de substrats ADN, ARN simple brin de 25 et 50 nt de long et des duplex ADN/ADN utilisés dans cette étude pour mesurer la capacité de liaison aux acides nucléiques de *Pab-ASH-Ski2* et *Pab-Hel308-Ski2* sont répertoriés dans la **Table M2**. Nous avons choisi des substrats identiques à ceux décrits dans la caractérisation enzymatique de l'hélicase *Mther-Hel308-Ski2* (Guy and Bolt, 2005). Les oligonucléotides ADN S1 et RS1 (complémentaire en séquence de S1) et l'ARN US1 obtenu par transcription *in vitro* sont des substrats simples brins de 50 nucléotides. L'oligonucléotide ADN 5RS1 et l'oligonucléotide

A



B



C

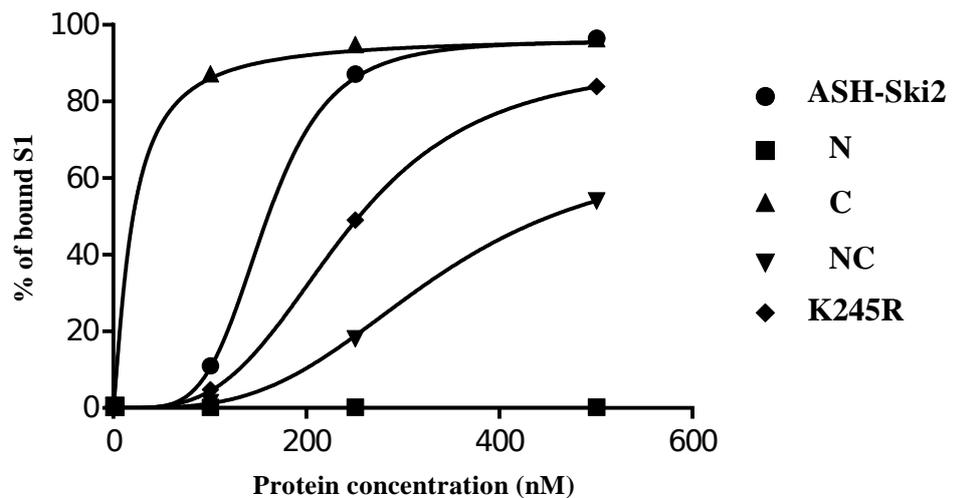


Figure 37: Tests d'interaction de ASH-Ski2, ces variants et Hel308-Ski2. A: Test d'affinité par double filtration sur membrane de nitrocelluloses et Hybond N (panel gauche) et courbes des tests d'affinité pour les différents substrats ADN simple brin (S1 et 5RS1), ARN simple brin (US1 et U5RS1) et en duplexe (S1/RS1 et S1/U5RS1) en présence de 250nM de ASH-Ski2 incubé 15min à 30°C. B: Tests d'interaction de Hel308 pour l'ADN simple brin S1 comparés avec ASH-Ski2 sauvage. C: Test affinité des différents variants de ASH-Ski2 pour l'ADN simple brin S1. La protéine déléetée de sa partie N-terminale: ASH-Ski2 Δ N, déléetée de sa partie C-terminale: Δ C, déléetée des deux domaines: Δ NC et la protéine ASH-Ski2 substituée sur la lysine 245 en arginine est nommée K245R.

ARN U5RS1 ont des séquences complémentaires respectivement aux 26 nt de l'extrémité 5' de S1 et nous permet de tester les substrats partiellement en double brin avec une extrémité 3' sortante. Les duplexes ADN/ADN (S1/RS1) ou ADN/ARN (S1/U5RS1) sont obtenus par incubation des simples brins à 90°C durant 5min et un refroidissement lent jusqu'à température ambiante (environ 30min) dans un tampon SSC 1X (150mM NaCl, 15mM NaCi). Les duplexes sont dessalés avant utilisation lors des tests.

De manière générale, 5nM de substrat ARN ou ADN radiomarqué est incubé 15 min à 30°C avec une concentration croissante de protéine d'intérêt en présence ou absence de 5mM d'ATP cofacteur hydrolysable ou d'ATP- γ S, cofacteur non hydrolysable. Nous avons également mené ces expériences à 65°C et n'avons pas observé de différences significatives dans ces deux conditions. Nous avons donc comparé les affinités des différentes protéines pour les acides nucléiques à 30°C.

Les complexes nucléoprotéiques formés après incubation sont séparés des molécules d'ARN ou ADN non liés par une méthode de double filtration sur membrane nitrocellulose/Hybond-N. Cette méthode a été choisie car la migration des complexes nucléoprotéiques sur gel d'acrylamide natif 5% est impossible avec des complexes formés restant dans les puits. Les complexes nucléoprotéiques sont retenus sur la membrane de nitrocellulose, les molécules ADN ou ARN non liés sont retenues sur la membrane Hybond-N (**Figure 37A, panel haut**). Après révélation par autoradiographie, une quantification sur le logiciel «MultiGauge» permet de déterminer la quantité d'acides nucléiques liés et non liés pour chaque concentration de protéine. Les pourcentages d'acides nucléiques liés aux protéines reportés sur un graphe en fonction de la concentration en protéine permettent de calculer les constantes de dissociation (K_d) pour chaque protéine testée à l'aide du logiciel «Prism» (**Figure 37**). Les résultats présentés sont des résultats préliminaires, ces tests devront être répétés au moins trois fois pour s'assurer des valeurs des constantes d'affinité reportées. Les valeurs de K_D présentées sont donc informatives et permettent juste de faire des comparaisons entre types d'acides nucléiques et différentes protéines. L'ensemble de ces résultats restent à confirmer.

III-2-2. Affinité de *Pab-ASH-Ski2* et de *Pab-Hel308-Ski2* pour les acides nucléiques

La protéine *Pab-ASH-Ski2* a une affinité relative (environ 160nM) très similaire pour les oligonucléotides ARN ou ADN de 50 nt (S1 et US1). De manière intéressante, les courbes ont un aspect sigmoïde (**Figure 37A**) qui reflète la présence de plusieurs sites de liaison aux acides nucléiques et/ou une liaison coopérative reflétant un phénomène d'allostérie par

changement de conformation de la protéine *Pab*-ASH-Ski2. L'ajustement (préliminaire) à une courbe sigmoïde permet d'observer un nombre de Hill supérieure à 1 suggérant la présence de plusieurs sites de liaison aux acides nucléiques de la protéine *Pab*-ASH-Ski2. Pour les substrats plus courts de 26 nucléotides, ADN (5RS1) ou ARN (U5RS1), l'affinité semble diminuée (K_D supérieur à 200 nM) par rapport aux substrats de 50nt. De plus, dans ce cas, la protéine *Pab*-ASH-Ski2 semble avoir une préférence pour le substrat ARN U5RS1 par rapport à l'ADN 5RS1 de même taille. Pour finir, les courbes de liaison aux substrats en duplex ADN/ADN (S1/RS1) ou ADN/ARN (S1/U5RS1) semblent identiques à celles avec les substrats de 50nt (**Figure 37A**).

De manière remarquable et inattendue, la protéine *Pab*-Hel308-Ski2 ne se fixe à aucun des substrats utilisés dans cette étude. L'absence de formation de complexes nucléoprotéiques stables suggèrent que la conformation adoptée par la protéine *Pab*-Hel308-Ski2 dans ces conditions expérimentales ne permet pas la fixation aux acides nucléiques (**Figure 37B**).

III-2-3. Affinité des variants de *Pab*-ASH-Ski2

Nous avons comparé l'affinité relative des différents variants de *Pab*-ASH-Ski2 avec le substrat S1 (ADN 50nt). Nous observons des différences dans les % de formation de complexe en fonction de la concentration en protéine pour tous les variants par rapport à la version sauvage de la protéine (**Figure 37C**).

a. *Pab*-ASH-Ski2K245R

Pab-ASH-Ski2K245R est un variant altéré dans le motif I du core hélicase SF2. Nous observons que la mutation ponctuelle K245R affecte la capacité de liaison au substrat ADN S1 avec 50% de formation de complexe à 250 nM comparé au 100% pour *Pab*-ASH-Ski2 sauvage (**Figure 37C**). Cependant l'aspect sigmoïde de la courbe de liaison semble conservé. L'intégrité du motif I qui est requis pour la structuration du core hélicase SF2 semble importante pour stabiliser une conformation protéique optimale de liaison aux acides nucléiques.

b. *Pab*-ASH-Ski2 Δ C

En ce qui concerne *Pab*-ASH-Ski2 Δ C, ce variant semble avoir une meilleure affinité pour le substrat S1 (ADN de 50nt) que *Pab*-ASH-Ski2 avec un pourcentage de formation maximale de complexe (plateau) atteint à 100 nM de concentration en protéine (**Figure 37C**). Les

concentrations en protéine *Pab*-ASH-Ski2- Δ C utilisées dans ce test ne permettent pas de déterminer la forme de la courbe, sigmoïde ou Michaelis-Menten ni de proposer une valeur pour le K_D . Des expériences avec des concentrations plus faibles en protéine devront être répétées afin de déterminer avec précision le K_D de l'interaction. Cependant cette observation préliminaire permet de proposer un effet modulateur ou effecteur du domaine C-terminal dans la fixation aux acides nucléiques. En absence de ce domaine, la liaison aux acides nucléiques semble être plus forte que pour la protéine pleine taille.

c. *Pab*-ASH-Ski2 Δ N

De manière surprenante, *Pab*-ASH-Ski2 Δ N ne se fixe à aucun des substrats utilisés dans cette étude comme observé pour *Pab*-Hel308-Ski2 (**Figure 37C**). L'absence du domaine N-terminal spécifique du groupe ASH-Ski2 est rédhibitoire pour la liaison stable aux acides nucléiques dans ces conditions. Il est à noter que *Pab*-ASH-Ski2 Δ N et *Pab*-Hel308-Ski2 sont, toutes deux, constituées exclusivement d'un domaine core des hélicases SF2 associé à un domaine C-terminal typique de la famille Ski2-like (voir Partie II-3). Nous pouvons donc conclure que la conformation adoptée par une protéine constituée seulement de ces deux domaines structuraux ne permet pas d'établir une liaison stable aux acides nucléiques dans nos conditions. Les sites de liaison aux acides nucléiques ne seraient pas accessibles dans ce contexte.

d. *Pab*-ASH-Ski2 Δ NC

Cependant, le core SF2 hélicase par lui-même est capable de se lier au substrat ADN S1. Ceci est montré par l'étude de la protéine tronquée à la fois de sa partie N et C-terminales (Δ NC). *Pab*-ASH-Ski2 Δ NC montre une affinité beaucoup plus faible pour l'ADN S1 que la protéine sauvage avec seulement 50 % de liaison à une concentration de 500 nM (**Figure 37C**). Ceci révèle l'existence d'un site de liaison aux acides nucléiques de faible affinité dans le core hélicase SF2 de la protéine comme décrit dans la littérature. En effet les domaines RecA1 et RecA2 possèdent des motifs (Ia, Ib, Ic, IV, Iva et V) connues pour être impliqués dans les liaisons aux acides nucléiques (**Figure 26B**) (Fairman-Williams *et al.*, 2010). Le site de liaison du core SF2 de *Pab*-ASH-Ski2 semble donc être masqué par le domaine C-terminal puisque la protéine *Pab*-ASH-Ski2 Δ N ne peut se lier aux acides nucléiques. Cette observation permet à nouveau de proposer un effet modulateur ou effecteur du domaine C-terminal comme proposé ci-dessus pour expliquer les propriétés de liaison de protéine *Pab*-ASH-Ski2 Δ C.

L'ensemble de cette étude préliminaire de liaison aux acides nucléiques permet de proposer que *Pab-ASH-Ski2* aurait au moins deux sites de liaison aux acides nucléiques (sans spécificité entre ADN ou ARN mais avec une préférence pour les substrats de 50 nt versus 26 nt). Un site fort de liaison existerait au sein du domaine N-terminal spécifique du groupe ASH-Ski2 puisqu'une délétion de ce domaine entraîne une perte totale de liaison (voir *Pab-ASH-Ski2ΔN*). Un site de plus faible affinité serait porté par le core SF2 hélicase (voir *Pab-ASH-Ski2ΔNC*) qui serait masqué en présence du domaine C-terminal (voir *Pab-ASH-Ski2ΔN* et *Pab-Hel308-Ski2*). Il serait important de déterminer si le motif de type Zinc Finger conservé dans le domaine N-terminal (voir partie II-3) est critique dans la liaison aux acides nucléiques.

III-3. ACTIVITE HELICASE DE *Pab-ASH-Ski2* ET *Pab-Hel308-Ski2*

III-3-1. Méthodes pour mesurer la capacité d'appariement de brins et de déroulement de duplex d'acides nucléiques

Afin de tester l'activité d'appariement des brins d'acides nucléiques de ASH-Ski2, de ces variants et de Hel308-Ski2. Nous avons choisi d'utiliser comme substrats l'ADN simple brin S1 (50nt) et l'ARN simple brin U5RS1 (26nt) complémentaire de S1. Ces deux substrats, hybridés, peuvent former des hétéroduplexes ADN/ARN avec une extrémité 3' sortante. L'ARN U5RS1 est marqué en 5' phosphorylation en présence d'ATP-[γ -32P] (U5RS1*). Nous avons réalisé une cinétique (0 à 10 minutes) d'appariement des brins en présence de 5nM d'ADN S1, 5nM de l'ARN U5RS1* radiomarké et 250nM de ASH-Ski2 dans un tampon Helicase 1X (25mM Tris-HCl pH8, 50 mM NaAc, 5mM MgCl₂ et 2.5mM 12.5mM β -mercaptoéthanol). La réaction est effectuée à 65°C afin d'éviter l'hybridation non spécifique des duplexes et de respecter la propriété thermorésistante de ASH-Ski2. Les réactions sont arrêtées dans un tampon stop 1X (0.1% SDS, 8mM EDTA, 0.1 mg/mL Protéinase K, 4% glycérol et 0.02% BPB) et sont migrées sur gel acrylamide 8%, 0.1% SDS. La quantité de radioactivité des différentes bandes observées est mesurée par autoradiographie (Typhoon) et le pourcentage des différents formes libre ou en duplex sont calculés sur «MultiGauge» (**Figure 38A**). Les courbes correspondant sont tracées sur «Prism».

Pour tester l'activité de déroulement de duplexes ADN/ARN (S1/U5RS1*), nous sommes restés dans les mêmes conditions expérimentales que lors des tests d'appariement de brins. Ainsi, après 2 minutes de réaction de formation de duplex, nous ajoutons soit de l'ATP, de

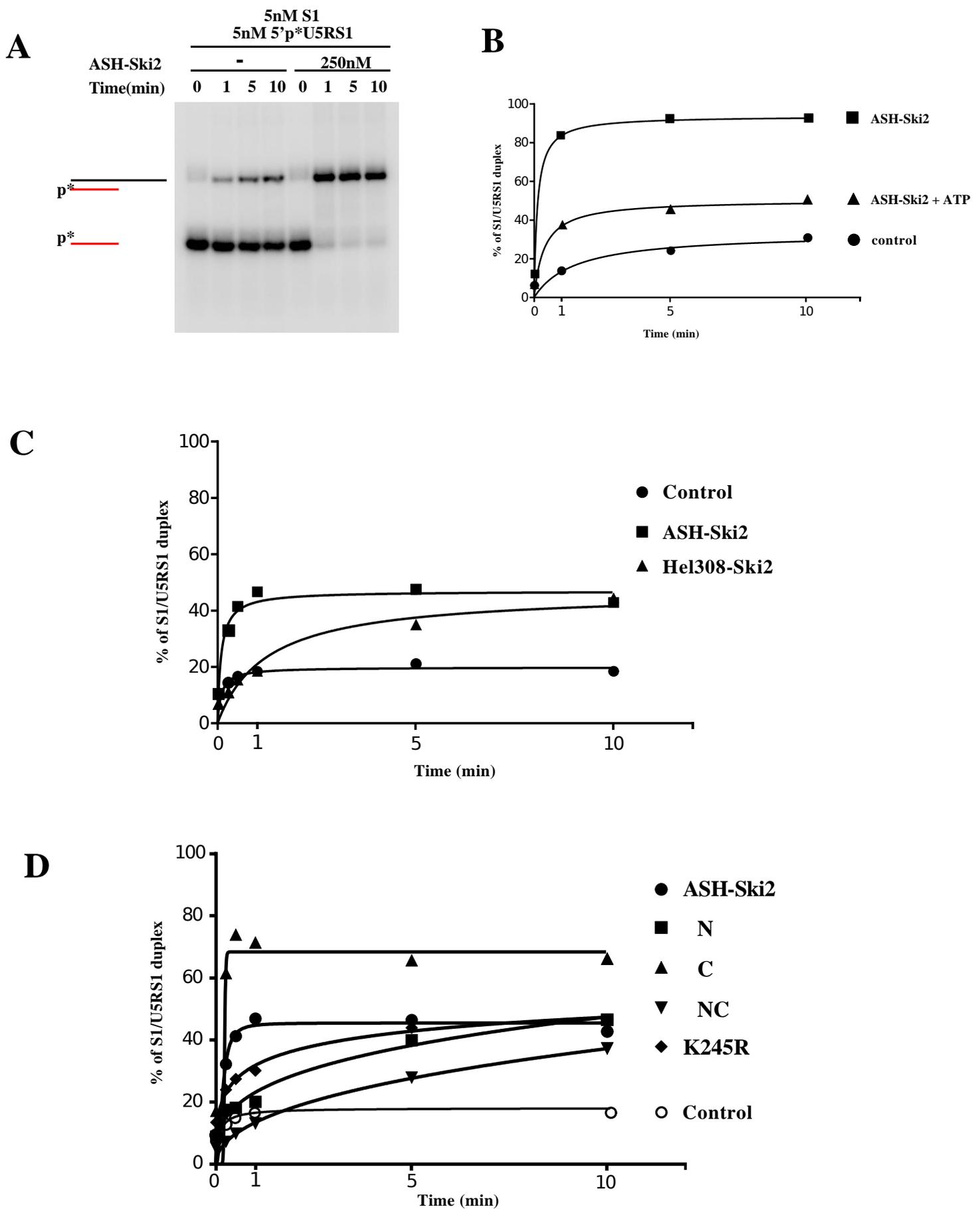


Figure 38: Tests de formation de duplex de ASH-Ski2, ces variants et Hel308-Ski2. A: Tests de formation des duplexes (S1 avec U5RS1 radiomarqué) en présence de 250 nM de ASH-Ski2 en absence ou en présence de 5mM d'ATP au cours du temps à 65°C. Migration sur gel acrylamide 8%, 1X TBE, 0.1% SDS durant 1h.30 à 160V. Les gels sont séchés et révélés par autoradiographie. Les bandes sont quantifiées par Multi Gauge. B: Tests d'appariement des brins d'acide nucléiques avec Hel308-Ski2 comparé à ASH-Ski2 sauvage. C: Tests d'appariement des brins d'acide nucléiques radiomarqués des différents variants de la protéine ASH-Ski2. La protéine détéelée de sa partie N-terminale: ASH-Ski2 Δ N, délétée de sa partie C-terminale: Δ C, délétée des deux domaines: Δ NC et la protéine ASH-Ski2 substituée sur la lysine 245 en arginine est nommée K245R.

l'ATP- γ -S, de l'ADP ou un contrôle avec de l'eau. La révélation et la quantification de duplex formé ont été les mêmes que pour précédemment.

III-3-2. Activités d'appariement de *Pab-ASH-Ski2*, de ces variants et de *Pab-Hel308-Ski2*

a. *Pab-ASH-Ski2* et *Pab-Hel308-Ski2*

En absence de protéine, nous observons moins de 20% de duplex formé alors qu'en présence de 250nM de *Pab-ASH-Ski2*, 90 % de duplex sont formés après 10min de réaction (**Figure 38A et B**). L'expérience en présence de 5mM d'ATP ne permet d'obtenir que 40% de duplex (**Figure 38B**). Ces résultats permettent de visualiser une activité d'appariement de brin de *Pab-ASH-Ski2* qui ne requiert pas l'activité ATPasique de l'enzyme. Au contraire, la présence d'ATP hydrolysable en ADP diminue la capacité de *Pab-ASH-Ski2* à favoriser l'appariement de brins.

L'étape suivante a été d'observer la capacité d'appariement de brin de *Pab-Hel308-Ski2* et des protéines variantes de *Pab-ASH-Ski2* en absence d'ATP. Il faut noter que les tests sont alors effectués en condition sub-optimale où la protéine *Pab-ASH-Ski2* ne forme en absence d'ATP que 45% de duplex après 10 min d'incubation. Nous avons dû adapter le protocole à la concentration la plus faible des protéines purifiées, ce qui nous oblige à tester les différentes réactions avec un volume plus élevé en protéine. De ce fait, la quantité en sel (tampon hélicase 1X + tampon de conservation des protéines) dans les tubes de réaction est plus élevée que dans l'expérience montrée sur la **Figure 37**. Cette hausse de concentration en sel réduit ainsi la formation de duplex stable. Ainsi, *Pab-Hel308-Ski2* montre une aptitude réduite à stabiliser la formation de duplex dans ces conditions par rapport à *Pab-ASH-Ski2* (**Figure 38C**). Après 1 min d'incubation, seulement 20% de duplex est formé comparé 45% de la protéine *Pab-ASH-Ski2*.

b. *Pab-ASH-Ski2K245R*

Lorsque nous testons les différents variant de *Pab-ASH-Ski2*, nous observons que *Pab-ASH-Ski2K245R* montre également une capacité réduite avec 30% de duplex formé sensiblement plus importante que *Pab-Hel308-Ski2* (**Figure 38D**).

c. *Pab-ASH-Ski2ΔN* et *Pab-ASH-Ski2ΔNC*

Pour les protéines *Pab-ASH-Ski2ΔN* et *Pab-ASH-Ski2ΔNC*, nous observons une réduction de la capacité à stabiliser le duplex avec un effet plus prononcé en absence des deux domaines (**Figure 38D**). Cette observation pour *Pab-ASH-Ski2ΔNC* semble corrélée avec une diminution de l'affinité pour les acides nucléiques. Cependant, l'absence de liaison stable aux acides nucléiques de *Pab-ASH-Ski2ΔN* et *Pab-Hel308-Ski2* (**Figure 37B**) et l'observation d'une faible capacité d'appariement de brins semblent contradictoire. Il se pourrait que dans ces deux cas, des liaisons entre les protéines et les acides nucléiques soit possibles permettant l'appariement des deux brins. Ces interactions ne semblent toutefois pas suffisantes pour maintenir des complexes ribonucléoprotéiques stables pour être observés lors des tests de liaisons (**Figure 37B**).

d. *Pab-ASH-Ski2ΔC*

De manière remarquable, la protéine *Pab-ASH-Ski2ΔC* montre une cinétique d'appariement de brins extrêmement rapide comparé à *Pab-ASH-Ski2* (**Figure 38D**). Il semblerait donc que l'absence du domaine C-terminal met en exergue la capacité de la protéine à stabiliser le duplex ADN/ARN. Ainsi le domaine C-terminal serait capable de moduler négativement à la fois l'activité ATPasique (voir Partie II-1) et la capacité d'appariement de brins.

L'ensemble de ces résultats indique que la protéine *Pab-ASH-Ski2* a une capacité d'appariement de brins en absence d'ATP et dans le contexte du mutant catalytique K245R. De manière intéressante, le core hélicase SF2 a la capacité intrinsèque d'apparier deux brins pour former un duplex stable. Le domaine N terminal augmente cette capacité tandis que le domaine C-terminal a un effet modulateur inhibiteur.

III-3-3. Activité de déroulement de duplex de *Pab-ASH-Ski2*, des variants et de *Pab-Hel308-Ski2*

a. *Pab-ASH-Ski2*

A partir d'une cinétique d'appariement de brins, telle décrite ci-dessus, nous avons pu également observer la capacité de *Pab-ASH-Ski2* à dérouler les duplex formés en ajoutant de l'ATP. Ainsi après deux minutes d'incubation, un volume équivalent d'eau, d'ATP, d'ATP- γ -S ou d'ADP est ajouté à la réaction. Nous observons clairement que, pour *Pab-ASH-Ski2*, seul l'ajout d'ATP ou de sa forme non hydrolysable (ATP- γ -S) entraîne rapidement (au temps

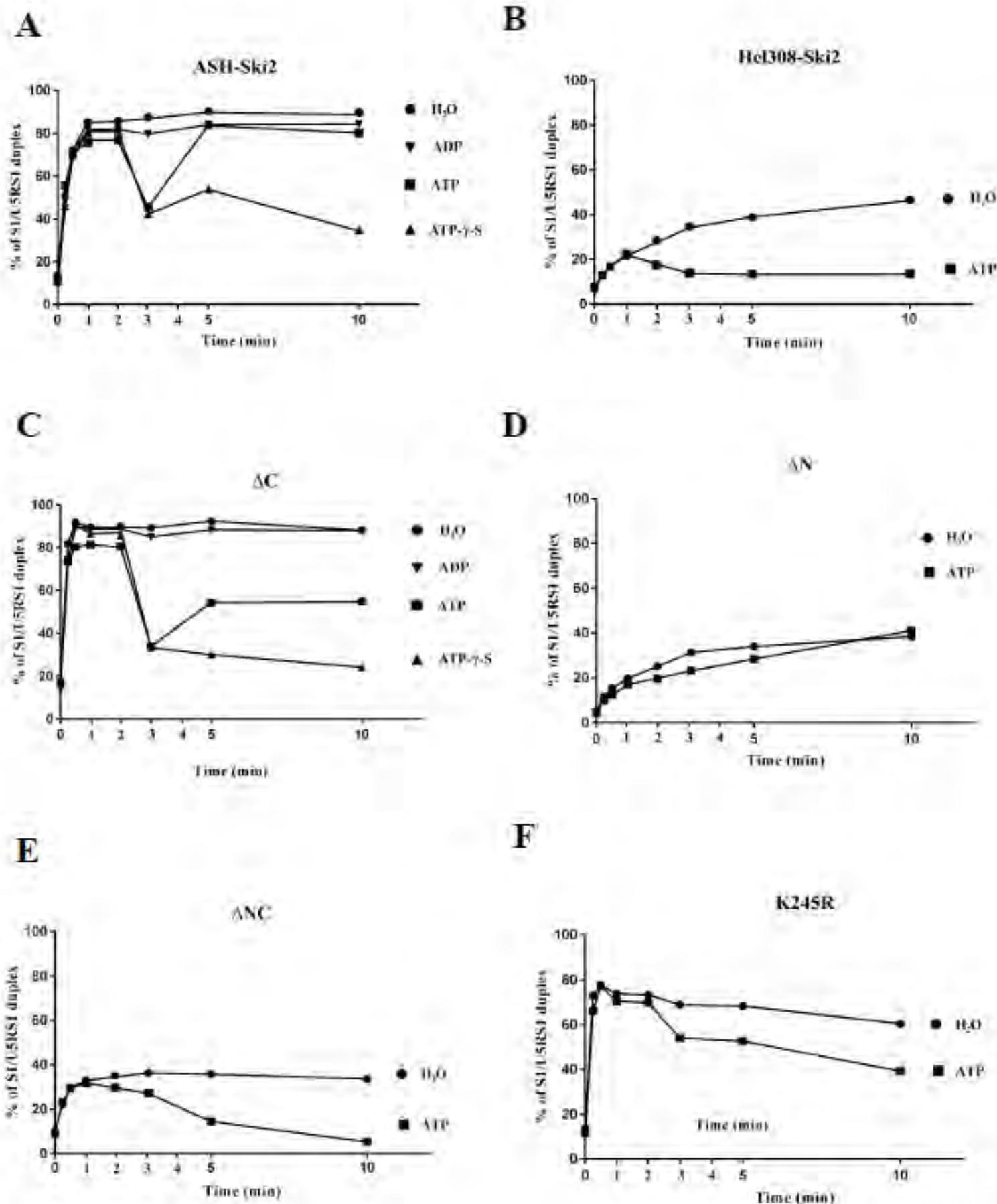


Figure 39: Tests de formation et de déroulement de duplex de ASH-Ski2, ces variants et Hel308-Ski2. A: Tests de déroulement des duplexes (S1 avec U5RS1 radiomarqué) en présence de 250 nM de ASH-Ski2 en absence d'ATP au début de la réaction et ajout de 5mM d'ATP, d'ATP- γ -S ou d'ADP après deux minutes de réaction. Tests effectués à 65°C. Migration sur gel acrylamide 8%, 1X TBE, 0.1% SDS durant 1h.30 à 160V. Les gels sont séchés et révélés par autoradiographie. Les bandes sont quantifiées par Multi Gauge. B: Tests de déroulement des duplexes avec Hel308-Ski2. C: Tests de déroulement des duplexes avec la ASH-Ski2 déletée de sa partie C-terminale: ΔC . D: Tests de déroulement des duplexes avec ASH-Ski2 déletée de sa partie N-terminale: ΔN . E: Tests de déroulement des duplexes avec ASH-Ski2 déletée de ces deux extensions N- et C terminales: ΔNC . F: Tests de déroulement des duplexes avec ASH-Ski2 substituée sur la lysine 245 en arginine est nommée K245R.

3 min) le déroulement de duplex préformé (décrochage avec plus de 80% de duplex à environ 40%) (**Figure 39A**). Les molécules simples brins ainsi formés ne peuvent s'apparier de novo en présence d'ATP- γ -S. Il semblerait que *Pab*-ASH-Ski2 en présence d'ATP non hydrolysable reste figer dans une conformation dite «unwinding» (H^U) ne permettant plus l'appariement de brin (**Figure 39A**). Cependant, en présence d'ATP, la capacité d'appariement de *Pab*-ASH-Ski2 est restaurée après 5 min, pouvant refléter le temps nécessaire pour hydrolyser de l'ATP en ADP. De manière intéressante, ni l'ajout d'ADP ou d'eau ne perturbe la cinétique de formation de duplex (stabilisée à plus de 80%). Cependant il semblerait que nous n'observons pas de turn-over de la réaction d'hydrolyse de l'ATP dans ces conditions *in vitro*. L'enzyme *Pab*-ASH-Ski2 semble bloquée dans une conformation dite «annealing» (H^A) après un seul turn-over (dû par exemple à l'absence de relarguage des produits de l'hydrolyse (ADP ou Pi) qui empêcherait la liaison à une nouvelle molécule d'ATP).

b. *Pab*-Hel308-Ski2

Pab-Hel308-Ski2 a un comportement différent de *Pab*-ASH-Ski2 avec une faible proportion de duplex formée au cours du temps en absence d'ATP qui atteint 40% comparée à près de 90% avec *Pab*-ASH-Ski2 (**Figure 39B**). Ainsi, *Pab*-Hel308-Ski2 serait majoritairement dans une conformation favorisant peu la formation du duplex. L'addition d'ATP à deux minutes de la cinétique engendre le déroulement presque total des duplex d'acides nucléiques et favorise une conformation (H^U).

c. *Pab*-ASH-Ski2 Δ C

Dans les mêmes conditions, le profil de la cinétique d'appariement de brin et déroulement de *Pab*-ASH-Ski2 Δ C est légèrement plus rapide que *Pab*-ASH-Ski2 (**Figure 39C**). Cependant une différence notable est la stabilisation de formation de 50% de duplex à partir de 5 minutes en présence d'ATP. Dans ce cas il semblerait que l'absence du domaine C-terminal permet un turn-over de l'activité ATPasique et un équilibre s'établit entre la formation des conformations (H^A) et (H^U) au cours du temps.

d. *Pab*-ASH-Ski2 Δ N

Le profil de *Pab*-ASH-Ski2 Δ N indique une faible capacité à former et une incapacité à dérouler des duplex en absence ou présence d'ATP (**Figure 39D**). De plus, ce profil se rapproche de celui du témoin sans protéine montré sur la **Figure 38C**. Ainsi, la faible activité

de formation de duplex et l'absence de déroulement de duplex de *Pab*-ASH-Ski2 Δ N pourrait être dû à son incapacité à lier les acides nucléiques (voir Partie III-2-3).

e. *Pab*-ASH-Ski2 Δ NC

Pour *Pab*-ASH-Ski2 Δ NC, le profil est similaire à celui observé pour *Pab*- Hel308-Ski2 (**Figure 39E**). Dans ce contexte, l'absence des deux extensions N- et C-terminal dévoile une activité du domaine core hélicase SF2 qui est capable de former des duplexes en absence d'ATP (30% à 1min) et de dérouler des duplexes en présence d'ATP (presque 0% à 10min).

f. *Pab*-ASH-Ski2K245R

De manière intéressante le profil de *Pab*-ASH-Ski2K245R (**Figure 39F**), mutant catalytique d'hydrolyse de l'ATP n'est pas complètement insensible à l'addition de l'ATP, nous observons une diminution de la quantité de duplex formé de 70% à 50% après l'addition de l'ATP à 3min. Cela pourrait indiquer que la protéine subirait un changement de conformation dû à la présence d'ATP malgré la mutation ponctuelle K245R du motif I.

L'ensemble de ces résultats de test de formation et de déroulement de duplex a permis de révéler que *Pab*-ASH-Ski2 possède à la fois la capacité à appairer deux brins d'acides nucléiques afin de former des duplexes à haute température et à dérouler ces mêmes duplexes en présence d'ATP. Cette activité de déroulement de duplex ne nécessite toutefois pas l'hydrolyse de l'ATP mais seulement sa fixation à la protéine puisque l'activité est visible en présence d'ATP- γ -S non hydrolysable. Ces deux activités ont aussi pu être observées pour la protéine Hel308-Ski2 mais dans une moindre mesure. En effet, Hel308-Ski2 ne présente qu'une faible activité d'appariement de duplex et mais est parfaitement capable de les dérouler.

Par cette approche, nous avons aussi pu montrer que la mutation ponctuelle K245R n'a que très peu d'effet sur son activité d'appariement et de déroulement des duplexes. La partie N-terminal de *Pab*-ASH-Ski2 est critique pour son activité de formation de duplex et est essentielle pour l'activité de déroulement. La partie C-terminale module négativement ces deux activités car en son absence, la protéine semble former plus rapidement les duplex et les déroule plus efficacement en présence d'ATP. Étonnamment, le core SF2 seule de la protéine est capable de faiblement former et de dérouler les duplexes.

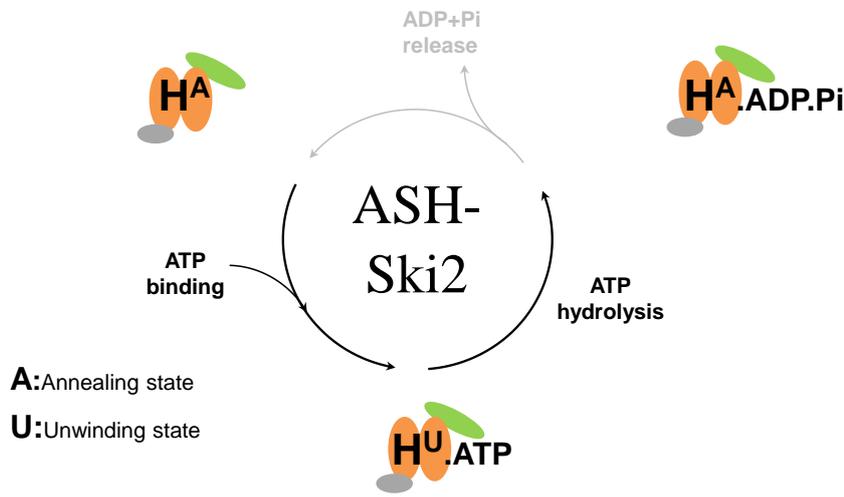
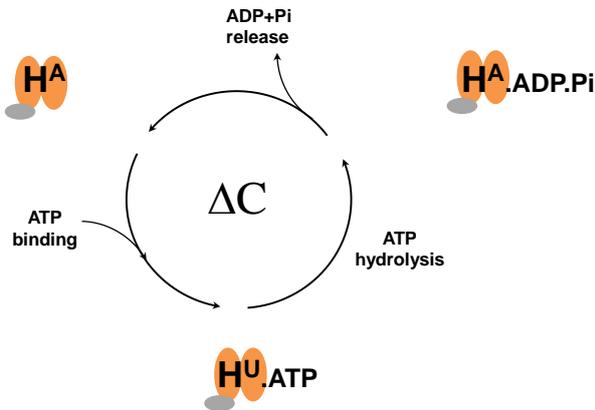
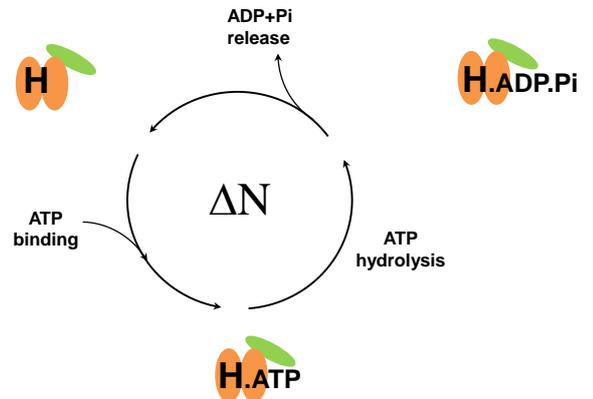
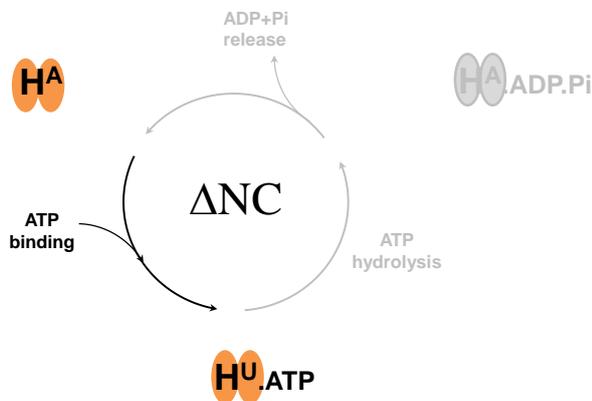
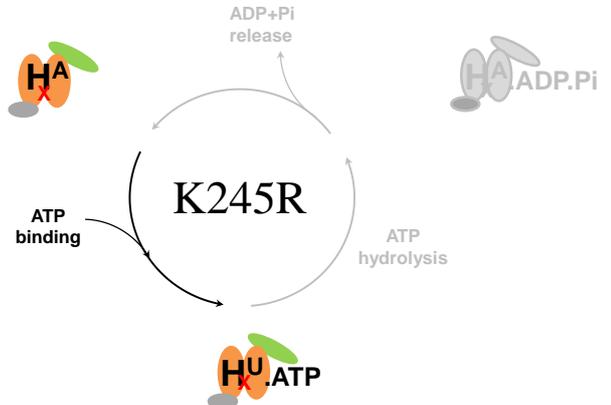
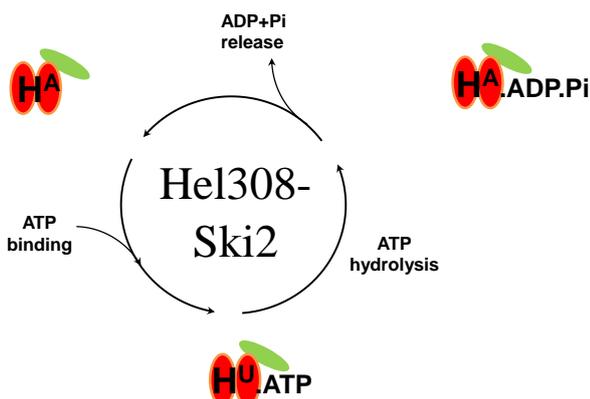
A**B****C****D****E****F**

Figure 40: Récapitulatif des différentes activités de ASH-Ski2, ces variants et Hel308-Ski2. A: Représentation schématique des différentes étapes des activités d'appariement et de déroulement de brins d'acide nucléique de ASH-Ski2 sauvage. La protéine schématisée avec ces différents domaines est représentée avec l'état dans laquelle elle se trouve (H^A pour annealing et H^U pour unwinding). B: Représentation schématique des activités de ASH-Ski2 délétaée de sa partie C-terminale: ΔC . C: Représentation schématique des activités de ASH-Ski2 délétaée de sa partie N-terminale: ΔN . D: Représentation schématique des activités de ASH-Ski2 délétaée de ses deux extensions N- et C-terminales : ΔNC . E: Représentation schématique des activités de ASH-Ski2 substituée sur la lysine 245 en arginine est nommée K245R. F: Représentation schématique des activités de Hel308-Ski2.

III-4. CONCLUSIONS SUR LES ACTIVITES DE *Pab-ASH-Ski2* ET *Pab-Hel308-Ski2*

L'ensemble de ces résultats portant sur les activités d'hydrolyse de l'ATP, de liaison aux acides nucléiques, d'appariement et de déroulement d'hétéroduplex de *Pab-ASH-Ski2*, ses variants et de *Pab-Hel308-Ski2* sont résumés sur la **Figure 40**. Ces tests ont ainsi montré que les protéines *Pab-ASH-Ski2* et *Pab-Hel308-Ski2* sont des ATPases dépendant des acides nucléiques. Il est étonnant de noter que ni *Pab-ASH-Ski2* et ni *Pab-Hel308-Ski2*, dans nos conditions *in vitro*, ne semblent avoir de spécificité de substrats (ADN ou ARN) pour activer l'hydrolyse de l'ATP. En effet, la protéine Hel308-Ski2 de *Sulfolobus solfataricus* est supposée être impliquée dans le métabolisme de réparation de l'ADN (Jodi D. Richards *et al.*, 2008). La spécificité pour les substrats pourrait ainsi provenir des protéines partenaires présentes dans la cellule.

La **Figure 40A** récapitule ainsi l'activité de ASH-Ski2 sauvage. En absence d'ATP, ASH-Ski2 qui est dans un état « annealing » (H^A), peut se lier aux acides nucléiques simples brins (ADN S1/ARN U5RS1) et est capable de former des hétéroduplexes ADN/ARN présentant une extrémité 3' sortante. En présence d'ATP, la liaison de l'ATP sur la protéine permet un changement de conformation de ASH-Ski2 vers un état « unwinding » (H^U). L'hydrolyse de l'ATP n'est pas nécessaire pour le passage de l'état (H^A) vers (H^U). De nombreuses DEAD-box hélicases partagent cette même activité de déroulement sans hydrolyse de l'ATP. C'est le cas pour la protéine CYT-19 impliquée dans la surveillance de structures non natives des ARN à l'intérieur des mitochondries ; la protéine Ded1p impliquée dans de nombreux processus moléculaire (initiation de la transcription, l'épissage des ARN, remodelage de complexe RNP...) (Chen *et al.*, 2008; Liu *et al.*, 2008). Une fois l'ATP hydrolysé, ASH-Ski2 retourne dans l'état (H^A) et est bloquée dans cet état. ASH-Ski2 ne semble pas être capable d'entamer un nouvel cycle. Ce blocage dans cet état (H^A) pourrait provenir d'une faible vitesse ou absence de relargage de l'ADP. *In vitro*, ASH-Ski2 seule ne présente pas un « turn over » observable.

La protéine ASH-Ski2 déléetée de sa partie C-terminale est capable de former des duplex et de les dérouler en présence d'ATP plus rapidement que la protéine sauvage (**Figure 40B**). Immédiatement après hydrolyse de l'ATP, ASH-Ski2 Δ C semble pouvoir entamer un nouveau cycle d'appariement et de déroulement de brins menant vers un état d'équilibre entre les deux conformations. Le domaine C-terminale de ASH-Ski2 jouerait donc un rôle auto

inhibiteur sur les différentes activités de la protéine en limitant son « turn over ». La délétion de la partie N-terminale de ASH-Ski2 montre un effet beaucoup plus drastique sur l'activité de la protéine (**Figure 40C**). En absence de cette extension, la protéine est incapable de former un complexe stable avec les acides nucléiques. Cette faible interaction entraîne la possibilité de former peu de duplex. Néanmoins, ASH-Ski2 Δ N est capable de se lier à l'ATP et de l'hydrolyser indépendamment de la présence des acides nucléiques. Toutefois, cette liaison ne permet pas à ASH-Ski2 Δ N d'avoir une activité de déroulement duplex. Cette absence d'activité de déroulement pourrait être expliquée par sa faible interaction avec les acides nucléiques. La partie N-terminale de ASH-Ski2 jouerait un rôle de régulation de l'hydrolyse de l'ATP en imposant une dépendance à la fixation avec les acides nucléiques. Lorsque les deux domaines N- et C-terminaux sont délétés, la protéine ASH-Ski2 Δ NC est toujours capable de se lier aux acides nucléiques et former des duplexes. Elle est aussi capable de dérouler ces mêmes duplexes en présence d'ATP. ASH-Ski2 Δ NC peut donc à la fois lier les acides nucléiques et l'ATP mais n'est pas capable d'hydrolyser ce dernier résultant du blocage de la protéine dans l'état (H^U) (**Figure 40D**). La délétion de ces deux domaines semble donc affecter la conformation générale du site catalytique de l'enzyme.

Ces états dans lesquelles les protéines sont figées dans la conformation (H^U) en absence d'hydrolyse de l'ATP sont confirmés par les activités de la protéine mutée sur son site catalytique ASH-Ski2K245R (**Figure 40E**). Cette protéine présente une capacité à former des duplex de manière similaire à la protéine sauvage mais est incapable d'hydrolyser l'ATP qu'elle a fixée. La lysine K245 n'est pas essentielle à la fixation de l'ATP mais est cruciale pour son hydrolyse. De manière intéressante, Hel308-Ski2 n'est capable de former que très peu de duplex et ne forme pas de complexes stables avec les acides nucléiques à l'instar de ASH-Ski2 Δ N. Or, cette protéine peut tout à fait dérouler les duplexes en présence d'ATP (**Figure 40F**). Hel308-Ski2 présente donc un « turn over » rapide avec peu d'activité d'appariement de brins.

Nous avons ainsi montré par ces approches l'importance des domaines N- et C-terminaux de ASH-Ski2 dans la régulation de l'hydrolyse de l'ATP, de liaison aux acides nucléiques, de formation et de déroulement d'hétéroduplex ADN/ARN. Il reste maintenant à déterminer la polarité (5'-3- ou 3'-5') de cette activité ainsi que des duplexes ADN/ADN et ARN/ARN. La régulation de l'activité de ASH-Ski2 par des partenaires protéiques sera aussi amenée à être testée, notamment le rôle que pourrait jouer la présence de la ribonucléase aRNase J.

DISCUSSION

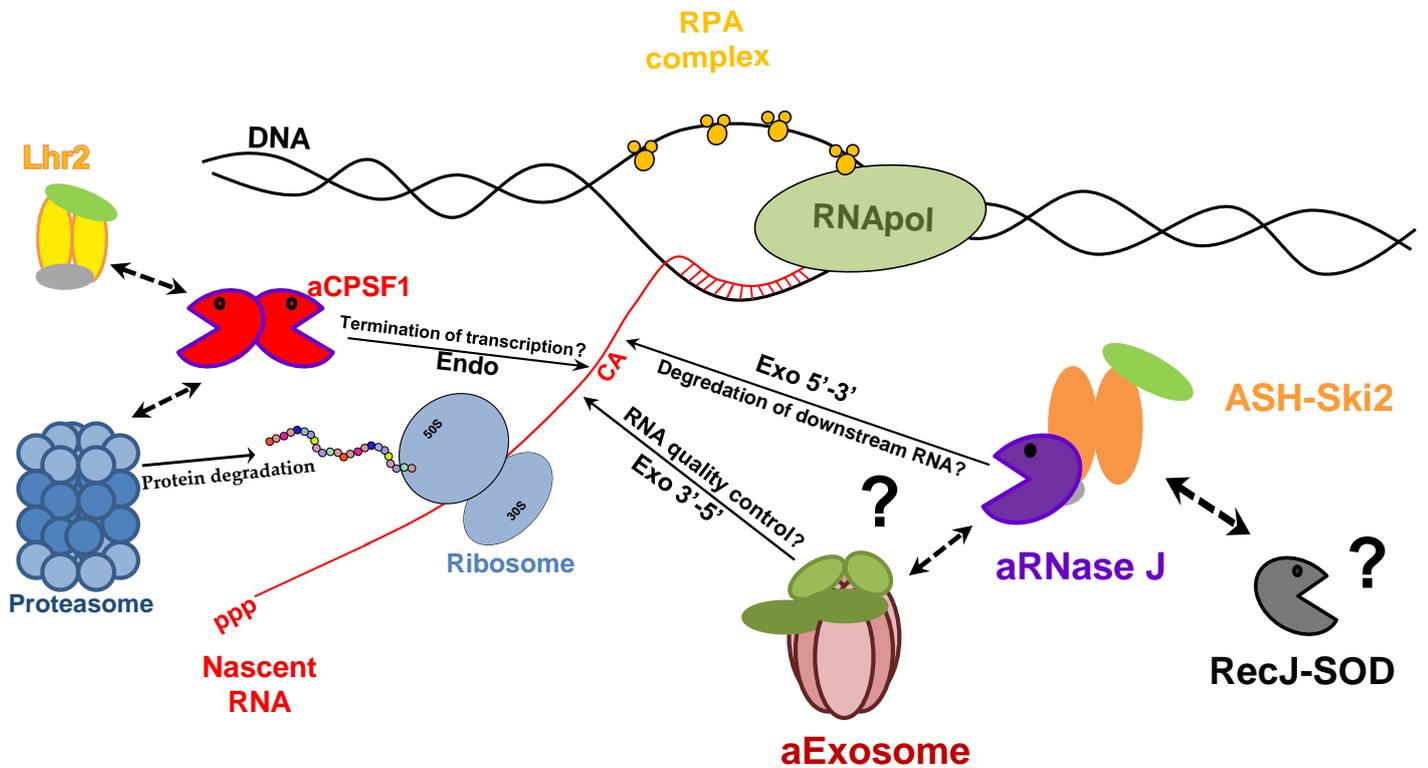


Figure 41: Rôles possibles de aCPSF1, aRNase J et ASH-Ski2. Représentation schématique des rôles que peuvent jouer aCPSF1 et le complexe aRNase J/ASH-Ski2 en concertation respectivement avec le protéasome et l'exosome lors du couplage de la transcription et la traduction. La RNA polymérase et le ribosome sont représentés en cours de synthèse respectivement d'un ARN et d'une protéine. Les flèches continues indiquent les activités et les cibles potentielles des ribonucléases. Les flèches en pointillé indiquent les interactions potentielles entre les différentes protéines. Le dinucléotide CA, cible de aCPSF1 est indiqué sur l'ARN.

L'objectif de cette thèse est d'apporter des éléments sur les rôles physiologiques des ribonucléases β -CASP dans le métabolisme des ARN chez les archées, en particulier des ribonucléases des groupes aCPSF1 et aRNase J. Nous proposons, basées sur des approches biochimiques avec pour modèle d'étude *P. abyssi* et des analyses phylogénomiques, une action coordonnée de la ribonucléase aRNase J et de l'hélicase ASH-Ski2 appartenant à la famille des hélicase de type Ski2-like avec le ribosome et l'exosome ainsi qu'avec la ribonucléase β -CASP aCPSF1 et l'hélicase Lhr2 de la famille des hélicases Lhr, non caractérisée à ce jour (**Figure 41**).

La ribonucléase β -CASP aCPSF1 ubiquitaire et de fonction inconnue

Nos travaux ont permis de montrer la conservation taxonomique stricte de la ribonucléase β -CASP aCPSF1 (voir RESULTAT-Partie I-Publication 1) et la présence de aCPSF1 dans l'interactomes de aRNase J (voir RESULTAT-Partie II). Ces résultats sont cohérents avec les hypothèses proposées par Koonin et collaborateurs en 2001 basées sur des études phylogénétiques qui suggèrent que la ribonucléase β -CASP aCPSF1 jouerait un rôle fondamental dans le processing des ARN (Koonin *et al.*, 2001) (**Figure 9B**). En effet l'analogie entre les activités enzymatiques de *Pab*-aCPSF1 et de son orthologue CPSF73 appartenant au complexe de terminaison et polyadénylation CPSF eucaryote laisse supposer que la ribonucléase aCPSF1 universellement conservée pourrait également être impliquée dans la terminaison de la transcription chez les archées, étape de la transcription où les acteurs restent inconnus à ce jour (voir Partie INTRODUCTION). Cependant nos résultats d'analyses phylogénomiques en collaboration avec Yves Quentin (équipe G. Fichant) ne permettent pas d'identifier des gènes codant pour les homologues des autres composants du complexe CPSF eucaryote dans les génomes d'archées mis à part le gène codant pour la ribonucléase aCPSF1 (voir RESULTAT-Partie I-Publication 2). Néanmoins de manière remarquable l'hélicase Lhr2 (également ubiquitaire chez les archées) a pu être détectée de manière significative dans l'interactome de *Pab*-aCPSF1 ainsi que certains composants de l'exosome (voir RESULTAT-Partie II). Ces résultats convergent vers l'hypothèse proposée par Koonin et collaborateurs en 2001 avec un rôle concerté de aCPSF1 et Lhr2 avec le complexe exosome chez les archées à l'instar des ribonucléases Rrp44/Rrp6 et les hélicases Ski2/Mtr4 dans le métabolisme des ARN chez les eucaryotes. Cependant nous n'avons pu dans ce travail approfondir le lien proposé entre ces composants. En effet des obstacles techniques ne nous ont pas permis d'une part d'étudier l'activité enzymatique de *Pab*-Lhr2 et d'autre part de définir l'interaction entre

Pab-aCPSF1 et *Pab*-Lhr2. En effet la surexpression de la protéine recombinante *Pab*-Lhr2 est toxique pour *E. coli*. De plus l'affinité intrinsèque d'aCPSF1 pour la colonne de nickel ne permet pas d'observer de manière significative la présence d'aCPSF1 endogène dans nos expériences de protéomiques définissant les interactomes de *Pab*-aRNase J et *Pab*-ASH-Ski2.

Cependant, de manière intrigante, une étude récente utilisant comme modèle *Haloferax volcanii* (Hvo) a identifié Hvo-aCPSF1 comme partenaire protéique d'une voie de maturation de l'ARN de transfert (Chavarria *et al.*, 2014). En effet l'étude montre que l'enzyme tRNA thiolase de *H. volcanii* (Hvo-NcsA) qui s'est révélée essentielle pour le maintien du pool cellulaire de l'ARNt Lys UUU thiolés et pour la croissance de *H. volcanii* à haute température formerait un complexe avec Hvo-aCPSF1. En outre, les protéines impliquées du protéasome et de la traduction sont également retrouvées dans le protéome de Hvo-NcsA (Chavarria *et al.*, 2014). Il n'est pas encore clair comment tous ces partenaires interagissent ou agissent dans la voie de thiolation de l'ARNt. Néanmoins, l'interaction de aCPSF1 avec des composants du protéasome suggère une connexion entre le processing de l'ARN et la dégradation des protéines (**Figure 41**). Cette hypothèse est soutenue par le contexte génomique d'aCPSF1 dans lequel le gène codant pour une sous-unité de protéasome est adjacent au gène codant pour aCPSF1 dans la plupart des génomes d'archées (voir RESULTAT-Partie II) (Koonin *et al.*, 2001). L'hypothèse selon laquelle les gènes du même opéron codent pour des protéines du même complexe ou impliquées dans le même processus suggèrent que le protéasome et aCPSF1 pourraient être fonctionnellement coordonnés. En outre, un super-opérateur codant pour les composants exosome code également pour les sous-unités catalytiques du protéasome chez les archées (voir Partie INTRODUCTION). Ainsi, l'analyse du contexte génomique suggère une interaction entre l'exosome et le protéasome et aCPSF1 (Koonin *et al.*, 2001; Evguenieva-Hackenberg and Klug, 2009). Ceci est cohérent avec les systèmes de contrôle de qualité de l'ARN décrits chez les eucaryotes (NMD, NSD, voies NRD) (voir Partie INTRODUCTION) et dans les bactéries (trans-traduction tmRNAs) où les ARNm aberrants sont produits suite à une élongation ou terminaison ou traduction aberrantes (Inada, 2013). De plus, l'interdépendance de la traduction et de la dégradation des ARNm chez les eucaryotes est évidente vu les rôles fonctionnels de la coiffe en 5' et de la queue 3'-poly (A) qui respectivement favorisent l'initiation de la traduction et protègent les ARNm (Roy and Jacobson, 2013).

La ribonucléase β -CASP aRNase J forme un complexe multiprotéique avec l'hélicase ASH-Ski2 chez *P. abyssi*

L'analyse de l'interactome de *Pab*-aRNase J nous a permis d'identifier une hélicase de la famille des hélicases de type Ski2-like retrouvées chez les eucaryotes que nous avons nommée ASH-Ski2 d'après nos données phylogénomiques et l'analyse de Chamieh et al 2016. La formation d'une interaction stable entre les protéines *Pab*-aRNase J et *Pab*-ASH-Ski2 permet de proposer l'existence d'un nouveau complexe, ribonucléase/hélicase, premier de ce type chez les archées. Cette hypothèse est confortée par la co-distribution taxonomique stricte des gènes codant pour aRNase J et ASH-Ski2 chez les euryarchées. Cependant la formation de ce complexe *in vitro* est seulement observée après des étapes de chauffage à haute température indiquant que seules certaines conformations protéiques sont susceptibles d'interagir. Cette interaction spécifique détectée *in vitro* fait intervenir le domaine N-terminal de *Pab*-ASH-Ski2 et plus particulièrement le motif de type « zinc finger » strictement conservé dans l'ensemble des hélicases du groupe ASH-Ski2 et le motif Loop1 typique des ribonucléases du groupe aRNase J chez les archées (voir RESULTAT-Partie II). De manière intéressante la structure cristalline de l'aRNase J de *Methanobolus psychrophilus* disponible sur la banque de données PDB indique une interaction du motif Loop1 avec un substrat de l'ARN (PDB : 5HAB). Le deuxième motif remarquable des aRNase J d'archées (Clouet-d'Orval *et al.*, 2010) est le motif « Loop2 » qui ne semble pas interagir avec les acides nucléiques dans la structure déterminée. Les motifs « Loop 1 et 2 » ont été montrés importants pour l'activité exoribonucléolytique de la ribonucléase avec une diminution de près de 75% et de plus de 90% respectivement (Clouet-d'Orval *et al.*, 2010). Il reste à tester l'importance du motif « Loop2 » dans l'interaction avec *Pab*-ASH-Ski2. Finalement la caractérisation biochimique de *Pab*-ASH-Ski2 a permis de révéler son activité hélicase et sa capacité de liaison aux d'acides nucléiques (voir RESULTAT-Partie III). En absence d'ATP, *Pab*-ASH-Ski2 possède la capacité d'appareiller deux brins d'acides nucléiques. En présence d'ATP, la simple liaison de la protéine avec l'ATP est suffisante pour dérouler ces mêmes brins. Nous avons ainsi montré que les extensions N- et C-terminales de *Pab*-ASH-Ski2 jouent des rôles essentiels dans la modulation de ses propres activités et dans l'interaction avec son partenaire aRNase J. Dans la continuité de ce travail l'équipe par la thèse de Manon Batista a à présent engagé l'étude structure-fonction du complexe *Pab*-aRNase J/*Pab*-ASH-Ski2.

Action concertée de la ribonucléase β -CASP aRNase J, de l'hélicase ASH-Ski2, de l'exosome et du ribosome chez *P. abyssi*

La co-sédimentation des protéines endogènes aRNase J et ASH-Ski2 observée dans des expériences de fractionnement cellulaire à partir de culots de cellules de *P. abyssi* et la position de *Pab*-aRNase J et *Pab*-ASH-Ski2 à l'interface des machineries de transcription, de traduction et de l'exosome au sein de leur interactome permettent de proposer une action concertée de ces deux protéines avec le ribosome et l'exosome (voir RESULTAT-Partie II). De manière intrigante, ASH-Ski2 endogène se retrouve dans des fractions de très haut poids moléculaire après traitement des extraits aux ribonucléases RNase T1 et RNase A (voir RESULTAT-Partie II) qui nous permet de proposer une interaction entre le ribosome et l'hélicase ASH-Ski2. En effet, chez la levure une accumulation de la protéine Ski3 du complexe SKI est observée au niveau du ribosome mature lorsque les extraits sont traités à la RNase I (Schmidt *et al.*, 2016). Les auteurs ont émis l'hypothèse d'un recrutement du complexe SKI au niveau du ribosome dans ces conditions entraînant la dégradation de l'ARNm partiellement clivé par la RNase I par le processus de NonStop Decay, processus au cours duquel le complexe SKI permet la translocation de l'ARNm du ribosome vers le complexe exosome par l'intermédiaire de l'interaction de la partie N-terminale de la protéine Ski7 afin d'être dégradé par la ribonucléase Rrp44 (Halbach *et al.*, 2013; Schmidt *et al.*, 2016). Chez les archées, aucun homologue des protéines Ski3, Ski8 ou Ski7 du complexe SKI n'ont à ce jour été identifiés permettant de faire la jonction entre ASH-Ski2 et l'exosome. De plus, *Pab*-ASH-Ski2 ne semble pas interagir directement avec la sous-unité Rrp41 du core de l'exosome dans nos conditions expérimentales *in vitro*. Cependant *Pab*-aRNase J peut interagir avec *Pab*-Rrp41 *in vitro* dans des conditions non astringentes (absence de traitement aux nucléases) (voir RESULTAT-Partie II). Cependant, il est possible que les protéines de la coiffe de l'exosome d'archées, Rrp4 ou Csl4, soient importantes pour la concertation entre l'exosome et le complexe aRNase J/ASH-Ski2 (**Figure41**). De plus il est intéressant de constater que le gène codant pour Csl4 se situe juste en aval du gène codant pour ASH-Ski2 chez les Thermococcales. De plus les phases ouvertes de lecture sont retrouvés chevauchantes comme dans le cas des sous-unités Rrp4, Rrp41 et Rrp42 de l'exosome (voir RESULTAT-Partie II) (Koonin *et al.*, 2001). Cette organisation particulière pourrait suggérer une co-régulation de leur expression. Enfin, la protéine Csl4 est retrouvée dans le réseau d'interaction de *Pab*-ASH-Ski2 en absence de traitement aux nucléases et au niveau structurale présente comme *Pab*-ASH-Ski2 un motif de type « zinc finger ». L'ensemble de ces observations

pourrait suggérer un lien fonctionnel entre la sous-unité Csl4 de la coiffe de l'exosome et l'hélicase ASH-Ski2 et le ribosome (**Figure 41**). Pour conclure, *Pab*-RNase J possédant une activité 5'-3' exoribonucléolytique pourrait ainsi faire le lien entre l'exosome possédant une activité 3'-5' exoribonucléolytique et *Pab*-ASH-Ski2 avec une activité hélicase. Les activités de grignotage des extrémités 5' et 3' des ARN serait ainsi connectées pour une concertation dans des voies de dégradation des ARN qui restent à définir.

Répartitions taxonomiques des ribonucléases aRNase J de la famille des β -CASP et des hélicases ASH-Ski2 de type Ski2-like

Il semblerait donc qu'un complexe mosaïque avec une ribonucléase aRNase J, homologue à la RNase J bactérienne, et une hélicase ASH-Ski2, homologue aux hélicases à ARN Ski2 et Mtr4 eucaryotes, serait présent dans la majorité des euryarchées. Ainsi la mise en évidence de ce nouveau complexe multiprotéique pose la question de l'origine évolutive de ses composants. Nos analyses phylogénomiques proposent une origine ancienne de ces deux enzymes avec une évolution verticale sans transfert latéral de gènes observables avec nos outils d'analyse. Il semblerait donc que ces gènes auraient été présents dans l'ancêtre commun des euryarchées avec une perte dans les branches donnant naissance aux Thermoplasmatales et Archéoglobales. Des analyses phylogénomiques plus avancées permettront d'émettre des hypothèses sur la coévolution des deux partenaires du complexe ASH-Ski2/aRNase J au sein des euryarchées.

De manière intéressante la concertation entre le complexe ASH-Ski2/aRNase J et l'exosome pourrait peut-être révélée par l'analyse du complexe au sein des Halophiles et de certaines méthanogènes qui ne possèdent pas les gènes codant pour les sous-unités de l'exosome. En contrepartie, ces archées possèdent une ribonucléase homologue à la RNase R bactérienne qui porte également une activité 3'-5' exonucléolytique (voir RESULTAT-Partie I-Publication 2). Il serait alors intéressant d'explorer les caractéristiques du complexe aRNase J/ASH-Ski2 dans ce contexte.

La protéine RecJ-SOD, un nouveau partenaire du métabolisme des ARN chez les euryarchées ?

Les réseaux d'interaction de *Pab*-aRNase J et *Pab*-ASH-Ski2 ont permis d'identifier une protéine annotée comme appartenant à la famille des RecJ-like et annotée comme RecJ-like phosphoestérase de 83 kDa mais différant des nucléases Rec J-like typiquement de 53 kDa.

En effet au vu des domaines structuraux de ce nouveau facteur, nous proposons de nommer cette protéine, RecJ-SOD. En effet elle est constituée d'un domaine S1, d'un motif de type OB-fold d'interaction avec les ARN et des motifs catalytiques de type DHH (Pfam : PF01368).

Les motifs DHH (Pfam: PF01368) et DHHA1 (Pfam: PF02272) sont retrouvés chez les nucléases de type RecJ qui sont des nucléases importantes pour la recombinaison homologue et chez les nano ribonucléases tels que NrnA et B qui sont impliquées dans la dégradation de petits oligonucléotides d'ARN (Fang *et al.*, 2009; Ogino *et al.*, 2017). On distingue donc dans les euryarchées deux gènes différents, l'un codant pour une nucléase RecJ classique qui est retrouvée dans les réseaux d'interaction des enzymes impliquées dans la maintenance de l'ADN chez *P. abyssi* (Pluchon *et al.*, 2013) et l'autre codant pour une protéine de fonction inconnue RecJ-SOD retrouvée dans les réseaux d'interaction de *Pab*-aRNase J et *Pab*-ASH-Ski2. L'organisation structurale en domaines de RecJ-SOD est conservée dans l'ensemble des génomes d'euryarchées avec une distribution taxonomique similaire aux gènes codant pour aRNase J et ASH-Ski2. De plus, des expériences préliminaires de co-purification par affinité avec *Pab*-RecJ-SOD comme appât ont permis d'observer une co-élution avec *Pab*-ASH-Ski2 qui semble résistante au traitement nucléases. L'ensemble de ces données indique que RecJ-SOD est un candidat de choix pour être une nouvelle ribonucléase pouvant intervenir dans le métabolisme de l'ARN chez les Euryarchées en concertation avec le complexe aRNase J/ASH-Ski2 et l'exosome (**Figure 41**).

Conclusion générale

Les travaux menés au cours de cette thèse apportent de nouveaux éléments autour des ribonucléases de type β -CASP chez les archées. Les résultats présentés, exclusivement basés sur des approches *in vitro* et *in silico*, laissent entrevoir le rôle primordial de ces enzymes dans le métabolisme des ARN chez les archées en concertation avec des hélicases et à l'interface de la machinerie de traduction et l'exosome. Actuellement des approches génétiques sont initiées dans l'équipe en collaboration avec M. Jebbar (Ifremer, Brest) et A. Marchfelder (Université d'Ulm, Allemagne) pour définir l'impact des ribonucléases de type β -CASP et de l'hélicase de type Ski2 sur le transcriptome de cellules de *T. barophilus* d'une part et de *H. volcanii* d'autre part. Le gène codant pour l'hélicase ASH-Ski2 a pu être délété dans les deux souches et une analyse transcriptomique préliminaire semble indiquer un impact sur la maturation ou la surveillance des ARN ribosomiques. De manière générale, la

description des complexes multi-protéiques impliqués dans les différentes voies du métabolisme des ARN chez les archées est primordiale pour appréhender l'évolution des complexes de dégradation et de processing des ARN dans le monde du vivant.

REFERENCES

- Aït-Bara, S., and Carpousis, A.J. (2015) RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs. *Mol Microbiol* **97**: 1021–1135.
- Albers, S.-V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampé, R., *et al.* (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl Environ Microbiol* **72**: 102–111.
- Anderson, J.S., and Parker, R.P. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* **17**: 1497–1506.
- Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol* **1**: 69–91.
- Aravind, L., and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* **23**: 469–472.
- Arias-Palomo, E., O'Shea, V.L., Hood, I.V., and Berger, J.M. (2013) The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153**: 438–448.
- Arkhipova, V., Stolboushkina, E., Kravchenko, O., Kljashtorny, V., Gabdulkhakov, A., Garber, M., *et al.* (2015) Binding of the 5'-Triphosphate End of mRNA to the γ -Subunit of Translation Initiation Factor 2 of the Crenarchaeon *Sulfolobus solfataricus*. *J Mol Biol* **427**: 3086–3095.
- Armache, K.-J., Kettenberger, H., and Cramer, P. (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**: 6964–6968.
- Arraiano, C.M., Mauxion, F., Viegas, S.C., Matos, R.G., and Séraphin, B. (2013) Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA. *Biochim Biophys Acta* **1829**: 491–513.
- Atomi, H., Imanaka, T., and Fukui, T. (2012) Overview of the genetic tools in the Archaea. *Front Microbiol* **3** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3462420/>. Accessed July 21, 2017.
- Babski, J., Haas, K.A., Näther-Schindler, D., Pfeiffer, F., Förstner, K.U., Hammelmann, M., *et al.* (2016) Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* **17** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4983044/>. Accessed July 21, 2017.
- Baillat, D., and Wagner, E.J. (2015) Integrator: surprisingly diverse functions in gene expression. *Trends Biochem Sci* **40**: 257–264.
- Bandyra, K.J., Bouvier, M., Carpousis, A.J., and Luisi, B.F. (2013) The social fabric of the RNA degradosome. *Biochim Biophys Acta* **1829**: 514–522.
- 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**: 905–920.
- Barry, E.R., and Bell, S.D. (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev MMBR* **70**: 876–887.

- Bartlett, E.J., Brissett, N.C., and Doherty, A.J. (2013) Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proc Natl Acad Sci U S A* **110**: E1984–1991.
- Beattie, T.R., and Bell, S.D. (2011) Molecular machines in archaeal DNA replication. *Curr Opin Chem Biol* **15**: 614–619.
- Bell, S.D., and Jackson, S.P. (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**: 222–228.
- Berger, J.M. (2008) SnapShot: nucleic acid helicases and translocases. *Cell* **134**: 888–888.e1.
- Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W., and Stetter, K.O. (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extrem Life Extreme Cond* **1**: 14–21.
- Blombach, F., Smollett, K.L., Grohmann, D., and Werner, F. (2016) Molecular Mechanisms of Transcription Initiation—Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol* **428**: 2592–2606.
- Blumer-Schuetz, S.E., Kataeva, I., Westpheling, J., Adams, M.W., and Kelly, R.M. (2008) Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr Opin Biotechnol* **19**: 210–217.
- Bortolin, M.-L., Bachellerie, 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.
- Boyd, J.B., Sakaguchi, K., and Harris, P.V. (1990) mus308 mutants of *Drosophila* exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. *Genetics* **125**: 813–819.
- Britton, R.A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W.C., Mathy, N., et al. (2007) Maturation of the 5′ end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol* **63**: 127–138.
- 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.
- Bubeck, D., Reijns, M.A.M., Graham, S.C., Astell, K.R., Jones, E.Y., and Jackson, A.P. (2011) PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* **39**: 3652–3666.
- Bujnicki, J.M., and Rychlewski, L. (2000) Prediction of a common fold for all four subunits of the yeast tRNA splicing endonuclease: implications for the evolution of the EndA/Sen family. *FEBS Lett* **486**: 328–329.
- Büttner, K., Wenig, K., and Hopfner, K.-P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol Cell* **20**: 461–471.

- Callebaut, I., Moshous, D., Mornon, J.-P., and Villartay, J.-P. de (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30**: 3592–3601.
- Cammarano, P., Teichner, A., Londei, P., Acca, M., Nicolaus, B., Sanz, J.L., and Amils, R. (1985) Insensitivity of archaeobacterial ribosomes to protein synthesis inhibitors. Evolutionary implications. *EMBO J* **4**: 811–816.
- 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.
- Carpousis, A.J. (2007) The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* **61**: 71–87.
- Carpousis, A.J., Van Houwe, G., Ehretsmann, C., and Krisch, H.M. (1994) Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* **76**: 889–900.
- Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**: 3489–3496.
- Carthew, R.W., and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642–655.
- Castillo-Lizardo, M., Henneke, G., and Viguera, E. (2014) Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota Pyrococcus abyssi. *Front Microbiol* **5**
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4134008/>. Accessed July 21, 2017.
- Chamieh, H., Ibrahim, H., and Kozah, J. (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* **576**: 214–228.
- Charpentier, B., Muller, S., and Branlant, C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33**: 3133–3144.
- Chavarria, N.E., Hwang, S., Cao, S., Fu, X., Holman, M., Elbanna, D., et al. (2014) Archaeal Tuc1/Ncs6 Homolog Required for Wobble Uridine tRNA Thiolation Is Associated with Ubiquitin-Proteasome, Translation, and RNA Processing System Homologs. *PLoS ONE* **9**
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4048286/>. Accessed July 22, 2017.
- Chen, Y., Potratz, J.P., Tijerina, P., Campo, M.D., Lambowitz, A.M., and Russell, R. (2008) DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci* **105**: 20203–20208.
- Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013) RNA decay machines: The exosome. *Biochim Biophys Acta* .
- Ciampi, M.S. (2006) Rho-dependent terminators and transcription termination. *Microbiol Read Engl* **152**: 2515–2528.
- Cline, S.W., Lam, W.L., Charlebois, R.L., Schalkwyk, L.C., and Doolittle, W.F. (1989) Transformation methods for halophilic archaeobacteria. *Can J Microbiol* **35**: 148–152.

- Clouet d'Orval, B., Bortolin, M.L., Gaspin, C., and Bachellerie, J.P. (2001) 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* **29**: 4518–4529.
- 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.
- Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., *et al.* (2003) An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* **47**: 1495–1512.
- Condon, C., and Gilet, L. (2011) The Metallo- β -Lactamase Family of Ribonucleases. In *Ribonucleases*. Springer, Berlin, Heidelberg, pp. 245–267 https://link.springer.com/chapter/10.1007/978-3-642-21078-5_10. Accessed July 22, 2017.
- Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.
- Da Cunha, V., Gaia, M., Gadelle, 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.
- Daiyasu, H., Osaka, K., Ishino, Y., and Toh, H. (2001) Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Lett* **503**: 1–6.
- Davey, M.J., Jeruzalmi, D., Kuriyan, J., and O'Donnell, M. (2002) Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol* **3**: 826–835.
- Davey, M.J., and O'Donnell, M. (2003) Replicative helicase loaders: ring breakers and ring makers. *Curr Biol CB* **13**: R594–596.
- Deana, A., Celesnik, H., and Belasco, J.G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**: 355–358.
- Dennis, P.P., and Omer, A. (2005) Small non-coding RNAs in Archaea. *Curr Opin Microbiol* **8**: 685–694.
- Deutscher, M.P. (2009) Maturation and degradation of ribosomal RNA in bacteria. *Prog Mol Biol Transl Sci* **85**: 369–391.
- Dijk, E. van, Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Séraphin, B. (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* **21**: 6915–6924.
- Dillingham, M.S. (2011) Superfamily I helicases as modular components of DNA-processing machines. *Biochem Soc Trans* **39**: 413–423.
- Dominski, Z. (2007) Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* **42**: 67–93.

- Dominski, Z., Carpousis, A.J., and Clouet-d'Orval, B. (2013) Emergence of the β -CASP ribonucleases: Highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim Biophys Acta* .
- Dominski, Z., Yang, X., and Marzluff, W.F. (2005) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**: 37–48.
- Dubrovsky, E.B., Dubrovskaya, V.A., Levinger, L., Schiffer, S., and Marchfelder, A. (2004) Drosophila RNase Z processes mitochondrial and nuclear pre-tRNA 3' ends in vivo. *Nucleic Acids Res* **32**: 255–262.
- Durand, S., Gilet, L., Bessières, P., Nicolas, P., and Condon, C. (2012) Three Essential Ribonucleases—RNase Y, J1, and III—Control the Abundance of a Majority of Bacillus subtilis mRNAs. *PLoS Genet* **8** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3297567/>. Accessed July 21, 2017.
- Dziembowski, A., Lorentzen, E., Conti, E., and Séraphin, B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol* **14**: 15–22.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Edgell, D.R., and Doolittle, W.F. (1997) Archaea and the Origin(s) of DNA Replication Proteins. *Cell* **89**: 995–998.
- Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**: 245–249.
- Eunen, K. van, Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A.B., Mensonides, F.I.C., *et al.* (2010) Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS J* **277**: 749–760.
- Evans, D., Marquez, S.M., and Pace, N.R. (2006) RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci* **31**: 333–341.
- 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 Res* **33**: 2141–2152.
- 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**: 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 (Mosc)* **47**: 13158–13168.
- Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F., and Klug, G. (2003) An exosome-like complex in Sulfolobus solfataricus. *EMBO Rep* **4**: 889–893.
- Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20**: 313–324.

- Fang, M., Zeisberg, W.-M., Condon, C., Ogryzko, V., Danchin, A., and Mechold, U. (2009) Degradation of nanoRNA is performed by multiple redundant RNases in *Bacillus subtilis*. *Nucleic Acids Res* **37**: 5114–5125.
- Farhoud, M.H., Wessels, H.J.C.T., Steenbakkens, P.J.M., Mattijssen, S., Wevers, R.A., Engelen, B.G. van, *et al.* (2005) Protein complexes in the archaeon *Methanothermobacter thermautotrophicus* analyzed by blue native/SDS-PAGE and mass spectrometry. *Mol Cell Proteomics MCP* **4**: 1653–1663.
- Farkas, J.A., Picking, J.W., and Santangelo, T.J. (2013) Genetic Techniques for the Archaea. *Annu Rev Genet* **47**: 539–561.
- Forterre, P. (2002) A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet TIG* **18**: 236–237.
- Forterre, P. (2013) The Common Ancestor of Archaea and Eukarya Was Not an Archaeon. *Archaea* <https://www.hindawi.com/journals/archaea/2013/372396/>. Accessed July 21, 2017.
- Fouqueau, T., Blombach, F., and Werner, F. (2017) Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annu Rev Microbiol* **71**: null.
- Frank, D.N., and Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem* **67**: 153–180.
- Franzmann, P.D., Springer, N., Ludwig, W., Conway De Macario, E., and Rohde, M. (1992) A Methanogenic Archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst Appl Microbiol* **15**: 573–581.
- Gao, J., Lee, K., Zhao, M., Qiu, J., Zhan, X., Saxena, A., *et al.* (2006) Differential modulation of *E. coli* mRNA abundance by inhibitory proteins that alter the composition of the degradosome. *Mol Microbiol* **61**: 394–406.
- Gao, Z., Putnam, A.A., Bowers, H.A., Guenther, U.-P., Ye, X., Kindsfather, A., *et al.* Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *eLife* **5** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4990422/>. Accessed July 22, 2017.
- Ge, Z., Mehta, P., Richards, J., and Karzai, A.W. (2010) Non-stop mRNA decay initiates at the ribosome. *Mol Microbiol* **78**: 1159–1170.
- Giraud, C., Hausmann, S., Lemeille, S., Prados, J., Redder, P., and Linder, P. (2015) The C-terminal region of the RNA helicase CshA is required for the interaction with the degradosome and turnover of bulk RNA in the opportunistic pathogen *Staphylococcus aureus*. *RNA Biol* **12**: 658–674.
- Gopalan, V. (2007) Uniformity amid diversity in RNase P. *Proc Natl Acad Sci U S A* **104**: 2031–2032.
- Gorbalenya, A.E., and Koonin, E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* **3**: 419–429.
- Gottesman, S., and Storz, G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**.

- Grabowski, B., and Kelman, Z. (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* **57**: 487–516.
- Grishin, N.V. (2001) Treble clef finger--a functionally diverse zinc-binding structural motif. *Nucleic Acids Res* **29**: 1703–1714.
- 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.
- 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**: 814–826.
- Halbeisen, R.E., Galgano, A., Scherrer, T., and Gerber, A.P. (2008) Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell Mol Life Sci CMLS* **65**: 798–813.
- Hall, T.A., and Brown, J.W. (2002) Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins. *RNA N Y N* **8**: 296–306.
- Harigaya, Y., and Parker, R. (2010) No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdiscip Rev RNA* **1**: 132–141.
- Hartman, H., Favaretto, P., and Smith, T.F. (2006) The archaeal origins of the eukaryotic translational system. *Archaea* **2**: 1–9.
- Hasenöhrl, D., Fabbretti, A., Londei, P., Gualerzi, C.O., and Bläsi, U. (2009) Translation initiation complex formation in the crenarchaeon *Sulfolobus solfataricus*. *RNA* **15**: 2288–2298.
- Hasenöhrl, 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.
- 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*. *Proc Natl Acad Sci U S A* **105**: 2146–2150.
- Hennigan, A.N., and Reeve, J.N. (1994) mRNAs in the methanogenic archaeon *Methanococcus vannielii*: numbers, half-lives and processing. *Mol Microbiol* **11**: 655–670.
- Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* **6**: 225–242.
- Hess, M. (2008) Thermoacidophilic proteins for biofuel production. *Trends Microbiol* **16**: 414–419.
- He, Y., Andersen, G.R., and Nielsen, K.H. (2010) Structural basis for the function of DEAH helicases. *EMBO Rep* **11**: 180–186.
- Hirata, A., Kanai, T., Santangelo, T.J., Tajiri, M., Manabe, K., Reeve, J.N., *et al.* (2008) Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol Microbiol* **70**: 623–633.

- Hochstrasser, M.L., and Doudna, J.A. (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem Sci* **40**: 58–66.
- Hölzle, A., Fischer, S., Heyer, R., Schütz, S., Zacharias, M., Walther, P., *et al.* (2008) Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* **14**: 928–937.
- Honda, T., Kakuta, Y., Kimura, K., Saho, J., and Kimura, M. (2008) 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* **384**: 652–662.
- Hoof, A. van, Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**: 2262–2264.
- Hopkins, B.B., and Paull, T.T. (2008) The *P. furiosus* mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break. *Cell* **135**: 250–260.
- Hossain, S.T., Malhotra, A., and Deutscher, M.P. (2016) How RNase R Degrades Structured RNA: Role of the Helicase Activity and the S1 Domain. *J Biol Chem jbc.M116.717991*.
- 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.
- 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 Biol* **10**: 909–914.
- Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., *et al.* (2016) A new view of the tree of life. *Nat Microbiol* **1**: nmicrobiol201648.
- Ikeuchi, Y., Kitahara, K., and Suzuki, T. (2008) The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. *EMBO J* **27**: 2194–2203.
- Inada, T. (2013) Quality control systems for aberrant mRNAs induced by aberrant translation elongation and termination. *Biochim Biophys Acta BBA - Gene Regul Mech* **1829**: 634–642.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**: 113–127.
- Jackson, R.N., Klauer, A.A., Hintze, B.J., Robinson, H., Hoof, A. van, and Johnson, S.J. (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J* **29**: 2205–2216.
- Jackson, R.N., Lavin, M., Carter, J., and Wiedenheft, B. (2014) Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr Opin Struct Biol* **24**: 106–114.
- 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 Res* **30**: 5436–5443.

- Jäger, D., Förstner, K.U., Sharma, C.M., Santangelo, T.J., and Reeve, J.N. (2014) Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15**: 684.
- Jankowsky, E. (2011) RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36**: 19–29.
- Januszyk, K., and Lima, C.D. (2014) The eukaryotic RNA exosome. *Curr Opin Struct Biol* **24**: 132–140.
- Jarrous, N., and Gopalan, V. (2010) Archaeal/eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic Acids Res* **38**: 7885–7894.
- Johnson, S.J., and Jackson, R.N. (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* **10**: 33–43.
- Kaberdin, V.R., Miczak, A., Jakobsen, J.S., Lin-Chao, S., McDowall, K.J., and Gabain, A. von (1998) The endoribonucleolytic N-terminal half of *Escherichia coli* RNase E is evolutionarily conserved in *Synechocystis* sp. and other bacteria but not the C-terminal half, which is sufficient for degradosome assembly. *Proc Natl Acad Sci U S A* **95**: 11637–11642.
- Kalman, M., Murphy, H., and Cashel, M. (1991) rhIB, a new *Escherichia coli* K-12 gene with an RNA helicase-like protein sequence motif, one of at least five such possible genes in a prokaryote. *New Biol* **3**: 886–895.
- Kammel, C., Thomaier, M., Sørensen, B.B., Schubert, T., Längst, G., Grasser, M., and Grasser, K.D. (2013) Arabidopsis DEAD-Box RNA Helicase UAP56 Interacts with Both RNA and DNA as well as with mRNA Export Factors. *PLOS ONE* **8**: e60644.
- Karousis, E.D., Nasif, S., and Mühlemann, O. (2016) Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip Rev RNA* **7**: 661–682.
- Karr, E.A., Sandman, K., Lurz, R., and Reeve, J.N. (2008) TrpY regulation of trpB2 transcription in *Methanothermobacter thermautotrophicus*. *J Bacteriol* **190**: 2637–2641.
- Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* **23**: 616–626.
- Kazantsev, A.V., Krivenko, A.A., Harrington, D.J., Holbrook, S.R., Adams, P.D., and Pace, N.R. (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc Natl Acad Sci U S A* **102**: 13392–13397.
- Kerr, I.D., Wadsworth, R.I.M., Cubeddu, L., Blankenfeldt, W., Naismith, J.H., and White, M.F. (2003) Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *EMBO J* **22**: 2561–2570.
- Khemici, V., Poljak, L., Luisi, B.F., and Carpousis, A.J. (2008) The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol Microbiol* **70**: 799–813.
- Khemici, V., Prados, J., Linder, P., and Redder, P. (2015) Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. *PLoS Genet* **11**: e1005577.

- Khidr, L., Wu, G., Davila, A., Procaccio, V., Wallace, D., and Lee, W.-H. (2008) Role of SUV3 helicase in maintaining mitochondrial homeostasis in human cells. *J Biol Chem* **283**: 27064–27073.
- 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**: 517–522.
- Kitamura, S., Fujishima, K., Sato, A., Tsuchiya, D., Tomita, M., and Kanai, A. (2010) Characterization of RNase HII substrate recognition using RNase HII-argonaute chimaeric enzymes from *Pyrococcus furiosus*. *Biochem J* **426**: 337–344.
- Kjems, J., and Garrett, R.A. (1988) Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* **54**: 693–703.
- Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J* **20**: 4214–4221.
- Kleman-Leyer, K., Armbruster, D.W., and Daniels, C.J. (1997) Properties of *H. volcanii* tRNA Intron Endonuclease Reveal a Relationship between the Archaeal and Eucaryal tRNA Intron Processing Systems. *Cell* **89**: 839–847.
- 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**.
- Klug, G., Evguenieva-Hackenberg, E., Omer, A.D., Dennis, P.P., and Marchfelder, A. (2007) RNA Processing. 158–174.
- Kochiwa, H., Tomita, M., and Kanai, A. (2007) Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evol Biol* **7**: 128.
- Ko, J., Izadjoo, M., and Altman, S. (2008) Inhibition of expression of virulence genes of *Yersinia pestis* in *Escherichia coli* by external guide sequences and RNase P. *RNA* **14**: 1656–1662.
- Komori, K., and Ishino, Y. (2001) Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem* **276**: 25654–25660.
- 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.
- Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., *et al.* (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol* **7**: e1000102.
- Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., *et al.* (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**: 423–435.
- Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003a) SURVEY AND SUMMARY: Structural classification of zinc fingers. *Nucleic Acids Res* **31**: 532–550.

- Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003b) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* **31**: 532–550.
- Kur, J., Olszewski, M., Długołęcka, A., and Filipkowski, P. (2005) Single-stranded DNA-binding proteins (SSBs) -- sources and applications in molecular biology. *Acta Biochim Pol* **52**: 569–574.
- Kyrpides, N.C., and Woese, C.R. (1998) Archaeal translation initiation revisited: The initiation factor 2 and eukaryotic initiation factor 2B α - β - δ subunit families. *Proc Natl Acad Sci U S A* **95**: 3726–3730.
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**: 713–724.
- Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* **92**: 5768–5772.
- Lao-Sirieix, S., and Bell, S.D. (2004) The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *J Mol Biol* **344**: 1251–1263.
- Lao-Sirieix, S.-H., Nookala, R.K., Roversi, P., Bell, S.D., and Pellegrini, L. (2005) Structure of the heterodimeric core primase. *Nat Struct Mol Biol* **12**: 1137–1144.
- Laurent, B.C., Treich, I., and Carlson, M. (1993) The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev* **7**: 583–591.
- 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 Res* **30**: 5382–5390.
- Le Hir, H., and Andersen, G.R. (2008) Structural insights into the exon junction complex. *Curr Opin Struct Biol* **18**: 112–119.
- Lehnik-Habrink, M., Newman, J., Rothe, F.M., Solovyova, A.S., Rodrigues, C., Herzberg, C., *et al.* (2011) RNase Y in *Bacillus subtilis*: a Natively Disordered Protein That Is the Functional Equivalent of RNase E from *Escherichia coli*. *J Bacteriol* **193**: 5431–5441.
- Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C., and Stülke, J. (2010) The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. *Mol Microbiol* **77**: 958–971.
- 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 Microbiol Rev* **35**: 577–608.
- Levy, S., Portnoy, V., Admon, J., and Schuster, G. (2011) Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biol* **8**: 1073–1083.
- Li de la Sierra-Gallay, I., Mathy, N., Pellegrini, O., and Condon, C. (2006) Structure of the ubiquitous 3' processing enzyme RNase Z bound to transfer RNA. *Nat Struct Mol Biol* **13**: 376–377.

- Li de la Sierra-Gallay, I., Zig, L., Jamalli, A., and Putzer, H. (2008) Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol* **15**: 206–212.
- Li, H. (2015) Structural Principles of CRISPR RNA Processing. *Struct Lond Engl* **1993** **23**: 13–20.
- Lindås, A.-C., and Bernander, R. (2013) The cell cycle of archaea. *Nat Rev Microbiol* **11**: 627–638.
- Linder, P., and Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* **12**: 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 Genet* **10**: e1004207.
- Liu, F., Putnam, A., and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci* **105**: 20209–20214.
- Liu, Q., Greimann, J.C., and Lima, C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**: 1223–1237.
- Li, Z., Pandit, S., and Deutscher, M.P. (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO J* **18**: 2878–2885.
- Li, Z., Pan, M., Santangelo, T.J., Chemnitz, W., Yuan, W., Edwards, J.L., *et al.* (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res* **39**: 6114–6123.
- Londei, P. (2007) Translational Mechanisms and Protein Synthesis.
<https://sapienza.pure.elsevier.com/it/publications/translational-mechanisms-and-protein-synthesis>. Accessed July 21, 2017.
- Long, R.M., and McNally, M.T. (2003) mRNA decay: x (XRN1) marks the spot. *Mol Cell* **11**: 1126–1128.
- Lopes, R.R.S., Kessler, A.C., Polycarpo, C., and Alfonzo, J.D. (2015) Cutting, dicing, healing and sealing: the molecular surgery of tRNA. *Wiley Interdiscip Rev RNA* **6**: 337–349.
- 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.
- 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. *Proc Natl Acad Sci U S A* **101**: 7046–7051.
- Lykke-Andersen, J., and Garrett, R.A. (1997) RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO J* **16**: 6290–6300.
- Lyu, Z., and Whitman, W.B. (2017) Evolution of the archaeal and mammalian information processing systems: towards an archaeal model for human disease. *Cell Mol Life Sci* **74**: 183–212.

- Machida, Y.J., Teer, J.K., and Dutta, A. (2005) Acute reduction of an origin recognition complex (ORC) subunit in human cells reveals a requirement of ORC for Cdk2 activation. *J Biol Chem* **280**: 27624–27630.
- Machnicka, M.A., Olchowik, A., Grosjean, H., and Bujnicki, J.M. (2015) Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol* **11**: 1619–1629.
- Mackie, G.A. (2013) RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Microbiol* **11**: 45–57.
- 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. *Biol Direct* **7**: 7.
- Malecki, M., Stepien, P.P., and Golik, P. (2009) Assays of the Helicase, ATPase, and Exoribonuclease Activities of the Yeast Mitochondrial Degradosome. In *Helicases*. Humana Press, Totowa, NJ, pp. 339–358 https://link.springer.com/protocol/10.1007/978-1-60327-355-8_24. Accessed July 22, 2017.
- Mandel, C.R., Bai, Y., and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci CMLS* **65**: 1099–1122.
- Mann, H., Ben-Asouli, Y., Schein, A., Moussa, S., and Jarrous, N. (2003) Eukaryotic RNase P: role of RNA and protein subunits of a primordial catalytic ribonucleoprotein in RNA-based catalysis. *Mol Cell* **12**: 925–935.
- 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 N Y N* **8**: 1189–1232.
- Marck, C., and Grosjean, H. (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA N Y N* **9**: 1516–1531.
- Marians, K.J. (1992) Prokaryotic DNA replication. *Annu Rev Biochem* **61**: 673–719.
- Marquez, S.M., Harris, J.K., Kelley, S.T., Brown, J.W., Dawson, S.C., Roberts, E.C., and Pace, N.R. (2005) Structural implications of novel diversity in eucaryal RNase P RNA. *RNA N Y N* **11**: 739–751.
- 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**: 681–692.
- Matos, R.G., Bária, C., Pobre, V., Andrade, J.M., and Arraiano, C.M. (2012) Exoribonucleases as Modulators of Virulence in Pathogenic Bacteria. *Front Cell Infect Microbiol* **2** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417396/>. Accessed July 22, 2017.
- Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y., and Takisawa, H. (2006) The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* **26**: 4843–4852.

- 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. *Environ Microbiol Rep* **2**: 272–280.
- Mir-Montazeri, B., Ammelburg, M., Forouzan, D., Lupas, A.N., and Hartmann, M.D. (2011) Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J Struct Biol* **173**: 191–195.
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* **91**: 457–466.
- Montange, R.K., and Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37**: 117–133.
- Morita, T., and Aiba, H. (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev* **25**: 294–298.
- Morten, M.J., Gamsjaeger, R., Cubeddu, L., Kariawasam, R., Peregrina, J., Penedo, J.C., and White, M.F. (2017) High-affinity RNA binding by a hyperthermophilic single-stranded DNA-binding protein. *Extremophiles* **21**: 369–379.
- Mukherjee, D., Gao, M., O'Connor, J.P., Rajmakers, R., Pruijn, G., Lutz, C.S., and Wilusz, J. (2002) The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J* **21**: 165–174.
- Myllykallio, H., Lopez, P., López-García, P., Heilig, R., Saurin, W., Zivanovic, Y., *et al.* (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**: 2212–2215.
- 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.
- Newman, J.A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C.R., and Lewis, R.J. (2011) Unusual, Dual Endo- and Exonuclease Activity in the Degradosome Explained by Crystal Structure Analysis of RNase J1. *Structure* **19**: 1241–1251.
- Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2005) Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing. *Struct Lond Engl* **13**: 143–153.
- Nudler, E. (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**: 1438–1445.
- NUMATA, T., ISHIMATSU, I., KAKUTA, Y., TANAKA, I., and KIMURA, M. (2004) Crystal structure of archaeal ribonuclease P protein Ph1771p from *Pyrococcus horikoshii* OT3: An archaeal homolog of eukaryotic ribonuclease P protein Rpp29. *RNA* **10**: 1423–1432.

- Nurmohamed, S., Vaidialingam, B., Callaghan, A.J., and Luisi, B.F. (2009) Crystal structure of Escherichia coli polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly. *J Mol Biol* **389**: 17–33.
- Ogawa, T., and Okazaki, T. (1984) Function of RNase H in DNA replication revealed by RNase H defective mutants of Escherichia coli. *Mol Gen Genet MGG* **193**: 231–237.
- Ogino, H., Ishino, S., Kohda, D., and Ishino, Y. (2017) The RecJ2 Protein in the Thermophilic Archaeon Thermoplasma acidophilum Is a 3' 5' Exonuclease and Associates with a DNA Replication Complex. *J Biol Chem jbc.M116.767921*.
- Ohtani, N., Yanagawa, H., Tomita, M., and Itaya, M. (2004a) Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, Sulfolobus tokodaii 7. *Nucleic Acids Res* **32**: 5809–5819.
- Ohtani, N., Yanagawa, H., Tomita, M., and Itaya, M. (2004b) Identification of the first archaeal Type 1 RNase H gene from Halobacterium sp. NRC-1: archaeal RNase HI can cleave an RNA–DNA junction. *Biochem J* **381**: 795–802.
- Omer, A.D., Ziesche, S., Ehardt, 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.
- Oost, J. van der, Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**: 479–492.
- OREN, A., GINZBURG, M., GINZBURG, B.Z., HOCHSTEIN, L.I., and VOLCANI, B.E. (1990) Haloarcula marismortui (Volcani) sp. nov., nom. rev., an Extremely Halophilic Bacterium from the Dead Sea. *Int J Syst Evol Microbiol* **40**: 209–210.
- Ow, M.C., and Kushner, S.R. (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in E. coli. *Genes Dev* **16**: 1102–1115.
- Oxley, A.P.A., Lanfranconi, M.P., Würdemann, D., Ott, S., Schreiber, S., McGenity, T.J., et al. (2010) Halophilic archaea in the human intestinal mucosa. *Environ Microbiol* **12**: 2398–2410.
- Pannucci, J.A., Haas, E.S., Hall, T.A., Harris, J.K., and Brown, J.W. (1999) RNase P RNAs from some Archaea are catalytically active. *Proc Natl Acad Sci U S A* **96**: 7803–7808.
- Pause, A., and Sonenberg, N. (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* **11**: 2643–2654.
- Peeters, E., and Charlier, D. (2010) The Lrp family of transcription regulators in archaea. *Archaea Vanc BC* **2010**: 750457.
- Perrochia, L., Guetta, D., Hecker, A., Forterre, P., and Basta, T. (2013) Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification. *Nucleic Acids Res* **41**: 9484–9499.
- Phung, D.K., Rinaldi, D., Langendijk-Genevaux, P.S., Quentin, Y., Carpousis, A.J., and Clouet-d'Orval, B. (2013) Archaeal β -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* **41**: 1091–1103.

- Picard, F., Loubière, P., Girbal, L., and Coccagn-Bousquet, M. (2013) The significance of translation regulation in the stress response. *BMC Genomics* **14**: 588.
- Pluchon, P.-F., Fouqueau, T., Crezé, C., Laurent, S., Briffotiaux, J., Hogrel, G., *et al.* (2013) An Extended Network of Genomic Maintenance in the Archaeon *Pyrococcus abyssi* Highlights Unexpected Associations between Eucaryotic Homologs. *PLoS One* **8**: e79707.
- Porrúa, O., and Libri, D. (2013) A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* **20**: 884–891.
- 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.
- Proudfoot, N.J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev* **25**: 1770–1782.
- Prud'homme-Généreux, A., Beran, R.K., Iost, I., Ramey, C.S., Mackie, G.A., and Simons, R.W. (2004) Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a “cold shock degradosome.” *Mol Microbiol* **54**: 1409–1421.
- Pyle, A.M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* **37**: 317–336.
- Ramos, C.R.R., Oliveira, C.L.P., Torriani, I.L., and Oliveira, C.C. (2006) The *Pyrococcus* exosome complex: structural and functional characterization. *J Biol Chem* **281**: 6751–6759.
- Randau, L., Schröder, I., and Söll, D. (2008) Life without RNase P. *Nature* **453**: 120–123.
- Raney, K.D., Byrd, A.K., and Aarattuthodiyil, S. (2013) Structure and Mechanisms of SF1 DNA Helicases. *Adv Exp Med Biol* **767** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806203/>. Accessed July 21, 2017.
- Redko, Y., Aubert, S., Stachowicz, A., Lenormand, P., Namane, A., Darfeuille, F., *et al.* (2013) A minimal bacterial RNase J-based degradosome is associated with translating ribosomes. *Nucleic Acids Res* **41**: 288–301.
- Redko, Y., Li de la Sierra-Gallay, I., and Condon, C. (2007) When all's zed and done: the structure and function of RNase Z in prokaryotes. *Nat Rev Microbiol* **5**: 278–286.
- Richard, D.J., Bell, S.D., and White, M.F. (2004) Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Res* **32**: 1065–1074.
- Richards, J., and Belasco, J.G. (2016) Distinct Requirements for 5'-Monophosphate-assisted RNA Cleavage by *Escherichia coli* RNase E and RNase G. *J Biol Chem* **291**: 5038–5048.
- Richards, J.D., Johnson, K.A., Liu, H., McRobbie, A.-M., McMahon, S., Oke, M., *et al.* (2008) Structure of the DNA Repair Helicase Hel308 Reveals DNA Binding and Autoinhibitory Domains. *J Biol Chem* **283**: 5118–5126.
- Richards, J., Sundermeier, T., Svetlanov, A., and Karzai, A.W. (2008) Quality Control of Bacterial mRNA Decoding and Decay. *Biochim Biophys Acta* **1779**: 574–582.

- 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**: 25–38.
- Rodriguez-Corona, U., Sobol, M., Rodriguez-Zapata, L.C., Hozak, P., and Castano, E. (2015) Fibrillarin from Archaea to human. *Biol Cell* **107**: 159–174.
- Roe, A.J., McLaggan, D., Davidson, I., O’Byrne, C., and Booth, I.R. (1998) Perturbation of Anion Balance during Inhibition of Growth of *Escherichia coli* by Weak Acids. *J Bacteriol* **180**: 767–772.
- Roy, B., and Jacobson, A. (2013) The intimate relationships of mRNA decay and translation. *Trends Genet TIG* **29**: 691–699.
- Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.-P., and Hüttenhofer, A. (2003) 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* **31**: 869–877.
- Samaha, R.R., Green, R., and Noller, H.F. (1995) A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* **377**: 309–314.
- Santangelo, T.J., Cubonová, L., Ůubomíra, and Reeve, J.N. (2008) Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl Environ Microbiol* **74**: 3099–3104.
- Santangelo, T.J., and Roberts, J.W. (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell* **14**: 117–126.
- Schierling, K., Rösch, S., Rupprecht, R., Schiffer, S., and Marchfelder, A. (2002) tRNA 3’ end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *J Mol Biol* **316**: 895–902.
- Schiffer, S., Rösch, S., and Marchfelder, A. (2002) Assigning a function to a conserved group of proteins: the tRNA 3’-processing enzymes. *EMBO J* **21**: 2769–2777.
- Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., *et al.* (1995) *Picrophilus* gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J Bacteriol* **177**: 7050–7059.
- Schmidt, C., Kowalinski, E., Shanmuganathan, V., Defenouillère, Q., Braunger, K., Heuer, A., *et al.* (2016) The cryo-EM structure of a ribosome–Ski2–Ski3–Ski8 helicase complex. *Science* **354**: 1431–1433.
- Schmitt, E., Naveau, M., and Mechulam, Y. (2010) Eukaryotic and archaeal translation initiation factor 2: A heterotrimeric tRNA carrier. *FEBS Lett* **584**: 405–412.
- Schweikhard, V., Meng, C., Murakami, K., Kaplan, C.D., Kornberg, R.D., and Block, S.M. (2014) Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc Natl Acad Sci U S A* **111**: 6642–6647.

- Sharma, S., Langhendries, J.-L., Watzinger, P., Kötter, P., Entian, K.-D., and Lafontaine, D.L.J. (2015) Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMP1. *Nucleic Acids Res* **43**: 2242–2258.
- 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.
- Shibuya, T., Tange, T., Stroupe, M.E., and Moore, M.J. (2006) Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA N Y N* **12**: 360–374.
- Shimada, Y., Fukuda, W., Akada, Y., Ishida, M., Nakayama, J., Imanaka, T., and Fujiwara, S. (2009) Property of cold inducible DEAD-box RNA helicase in hyperthermophilic archaea. *Biochem Biophys Res Commun* **389**: 622–627.
- Shu, Z., Vijayakumar, S., Chen, C.-F., Chen, P.-L., and Lee, W.-H. (2004) Purified Human SUV3p Exhibits Multiple-Substrate Unwinding Activity upon Conformational Change. *Biochemistry (Mosc)* **43**: 4781–4790.
- Silva, A.L., and Romão, L. (2009) The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision? *FEBS Lett* **583**: 499–505.
- Silva, A.P.G., Chechik, M., Byrne, R.T., Waterman, D.G., Ng, C.L., Dodson, E.J., *et al.* (2011) Structure and Activity of a Novel Archaeal β -CASP Protein with N-Terminal KH Domains. *Struct England* **19**: 622–632.
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**: 23–50.
- Smollett, K., Blombach, F., Reichelt, R., Thomm, M., and Werner, F. (2017) A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat Microbiol* **2**: nmicrobiol201721.
- Song, M.-G., Bail, S., and Kiledjian, M. (2013) Multiple Nudix family proteins possess mRNA decapping activity. *RNA* **19**: 390–399.
- Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., *et al.* (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173–179.
- Späth, B., Schubert, S., Lieberoth, A., Settele, F., Schütz, S., Fischer, S., and Marchfelder, A. (2008) Two archaeal tRNase Z enzymes: similar but different. *Arch Microbiol* **190**: 301.
- Stead, M.B., Marshburn, S., Mohanty, B.K., Mitra, J., Castillo, L.P., Ray, D., *et al.* (2011) Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res* **39**: 3188–3203.
- Stoecklin, G., and Mühlemann, O. (2013) RNA decay mechanisms: Specificity through diversity. *Biochim Biophys Acta BBA - Gene Regul Mech* **1829**: 487–490.

- Sung, P., Higgins, D., Prakash, L., and Prakash, S. (1988) Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP. *EMBO J* **7**: 3263–3269.
- Tadokoro, T., and Kanaya, S. (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* **276**: 1482–1493.
- Takata, K., Reh, S., Tomida, J., Person, M.D., and Wood, R.D. (2013) Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat Commun* **4**: 2338.
- Tang, T.H., Rozhdestvensky, T.S., Orval, B.C. d', Bortolin, M.-L., Huber, H., Charpentier, B., *et al.* (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* **30**: 921–930.
- 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.
- Tomecki, R., Drazkowska, K., and Dziembowski, A. (2010) Mechanisms of RNA Degradation by the Eukaryotic Exosome. *ChemBioChem* **11**: 938–945.
- Tomkuvienė, M., Clouet-d'Orval, B., Cerniauskas, I., Weinhold, E., and Klimasauskas, S. (2012) Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. *Nucleic Acids Res* **40**: 6765–6773.
- Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T., and Mondragón, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* **437**: 584–587.
- Tran, E.J., Zhang, X., and Maxwell, E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J* **22**: 3930–3940.
- Tuteja, N., Tarique, M., and Tuteja, R. (2014) Rice SUV3 is a bidirectional helicase that binds both DNA and RNA. *BMC Plant Biol* **14** <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4207899/>. Accessed July 22, 2017.
- Uson, M.L., Ordonez, H., and Shuman, S. (2015) Mycobacterium smegmatis HelY Is an RNA-Activated ATPase/dATPase and 3'-to-5' Helicase That Unwinds 3'-Tailed RNA Duplexes and RNA:DNA Hybrids. *J Bacteriol* **197**: 3057–3065.
- Veith, T., Martin, R., Wurm, J.P., Weis, B.L., Duchardt-Ferner, E., Safferthal, C., *et al.* (2012) Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic Acids Res* **40**: 3259–3274.
- Vincent, H.A., and Deutscher, M.P. (2009) Insights into how RNase R degrades structured RNA: analysis of the nuclease domain. *J Mol Biol* **387**: 570–583.
- Vivona, J.B., and Kelman, Z. (2003) The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett* **546**: 167–172.

- Vogel, A., Schilling, O., Späth, B., and Marchfelder, A. (2005) The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biol Chem* **386**: 1253–1264.
- Warbrick, E. (2000) The puzzle of PCNA's many partners. *BioEssays News Rev Mol Cell Dev Biol* **22**: 997–1006.
- Wasmuth, E.V., Januszyk, K., and Lima, C.D. (2014) Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **511**: 435–439.
- Wassarman, K.M. (2007) 6S RNA: a regulator of transcription. *Mol Microbiol* **65**: 1425–1431.
- Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615–628.
- Wen, T., Oussenko, I.A., Pellegrini, O., Bechhofer, D.H., and Condon, C. (2005) Ribonuclease PH plays a major role in the exonucleolytic maturation of CCA-containing tRNA precursors in *Bacillus subtilis*. *Nucleic Acids Res* **33**: 3636–3643.
- Werner, F. (2013) Molecular mechanisms of transcription elongation in archaea. *Chem Rev* **113**: 8331–8349.
- Werner, F., and Weinzierl, R.O.J. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* **10**: 635–646.
- West, S., Gromak, N., and Proudfoot, N.J. (2004) Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**: 522–525.
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331–338.
- Williams, T.A., Szöllösi, G.J., Spang, A., Foster, P.G., Heaps, S.E., Boussau, B., *et al.* (2017) Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci* **114**: E4602–E4611.
- Woese, C.R., and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088–5090.
- Woodman, I.L., Brammer, K., and Bolt, E.L. (2011) Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA). *DNA Repair* **10**: 306–313.
- Worthington, P., Hoang, V., Perez-Pomares, F., and Blum, P. (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* **185**: 482–488.
- Wu, H., Xu, H., Miraglia, L.J., and Crooke, S.T. (2000) Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *J Biol Chem* **275**: 36957–36965.
- Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A., and Sorek, R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res* **20**: 133–141.
- Xiong, Y., and Steitz, T.A. (2004) Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. *Nature* **430**: 640–645.

- Yutin, N., Puigbò, P., Koonin, E.V., and Wolf, Y.I. (2012) Phylogenomics of Prokaryotic Ribosomal Proteins. *PLOS ONE* **7**: e36972.
- Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**: 353–358.
- Zeng, X., Birrien, J.-L., Fouquet, Y., Cherkashov, G., Jebbar, M., Querellou, J., *et al.* (2009) *Pyrococcus* CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *ISME J* **3**: 873–876.
- Zhai, B., DuPrez, K., Doukov, T.I., Li, H., Huang, M., Shang, G., *et al.* (2017) Structure and Function of a Novel ATPase that Interacts with Holliday Junction Resolvase Hjc and Promotes Branch Migration. *J Mol Biol* **429**: 1009–1029.
- Zhang, H., Rigo, F., and Martinson, H.G. (2015) Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol Cell* **59**: 437–448.
- Zhang, L., Li, X., Hill, R.C., Qiu, Y., Zhang, W., Hansen, K.C., and Zhao, R. (2015) Brr2 plays a role in spliceosomal activation in addition to U4/U6 unwinding. *Nucleic Acids Res* **43**: 3286–3297.
- Zinder, J.C., and Lima, C.D. (2017) Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**: 88–100.

ANNEXES

Accession	Gene	Name	kDa	AA	SR-	SR+	#	Description
Q9UZM4	PAB0744	Lhr2	100	867	0	0	24	Lhr-2 large helicase-related
Q9V1T4	PAB2119	Unch.	29	258	0	0.01	7	Methyl-transferase like
Q9V0E2	PAB0569	Unch.	30	267	0.01	0.02	9	Proteasome chaperone
Q9V1Z1	PAB2165	RPA32	32	272	0.01	0.02	11	RPA32 subunit
Q9UYR2	PAB1421	Unch.	30	264	0.01	0.01	5	Uncharacterized
Q9V113	PAB0425	RpoA2	44	397	0.07	0.06	6	DNA-directed RNA polymerase subunit A''
Q9UY07	PAB1115	GraD-2	46	419	0.08	0.10	17	Glucose-1-phosphate thymidyltransferase
Q9V076	PAB1751	aRNase J	50	451	0.06	0.13	13	Ribonuclease J (5'-3' Exo)
Q9V1Y9	PAB2163	RPA41	43	375	0,10	0.13	23	RPA41 subunit
Q9V1U1	PAB2125	Rps3	23	210	0.10	0.16	9	30S ribosomal protein S3
Q9V1V6	PAB2137	Rpl30p	18	155	0.11	0.17	7	50S ribosomal protein L30P
Q9V1T5	PAB2120	Rpl3	41	361	0.15	0.18	8	50S ribosomal protein L3
Q9V114	PAB0424	RpoA1	103	907	0.22	0.25	29	DNA-directed RNA polymerase subunit A'

Table 1: Protéines les plus significativement retrouvées avec **aCPSF1** déterminées par «Pull-down» de **Pab-(His)₆-aCPSF1** (BioGenOuest Plateforme)

Les protéines ont été classées par comparaison et identification par spectrométrie de masse effectuée par la plateforme de protéomique BioGenOuest. Les accessions des protéines et la description des gènes sont issus de la base de données d'Uniprot (<http://www.uniprot.org/uniprot>). Seules les protéines ayant une valeur SR en dessous de 0.25 et un nombre de spectre (#) dans le «Pull-down» au-dessus de 5 dans chaque réplica sont montrées. Pour chaque protéine, la masse molaire en kDa (M) et le cluster des groupes orthologues (COG) correspondant sont indiqués. *Unch* pour les protéines non caractérisées.

Accession	Gene	Name	kDa	AA	ENuc-	ENuc+	Description
Q9V1F2	PAB0316	DnaG	50	447	32±7	29±5	DnaG
Q9UZ78	PAB2428	TmcA	94	817	23±12	28±17	tRNA(Met) cytidine acetyltransferase
Q9V115	PAB0423	RpoB	127	1117	11±2	6±2	DNA-directed RNA polymerase subunit beta
Q9V1T4	PAB2119	<i>Unch.</i>	29	258	7±1	7±2	Methyl-transferase like
Q9V1Y9	PAB2163	RPA41	43	375	7±2	7±4	RPA41 subunit
Q9V2L4	PAB2305	Nop5p	47	404	7±4	7±2	Nop58p-like pre mRNA splicing
Q9V2M1	PAB2313	ASH-Ski2	97	855	5±3	2±1	ATP-dependent RNA helicase
Q9UYB6	PAB1306	eIF-2B	31	276	5±4	4±2	Putative translation initiation factor eIF-2B subunit 2-like
Q9UY07	PAB1115	GraD-2	46	419	4±1	5±2	Glucose-1-phosphate thymidyltransferase
Q9V1T5	PAB2120	Rpl3	40	361	4±2	8±3	50S ribosomal protein L3
Q9UY85	PAB1284	RecJ-SOD	83	740	4±2	4±2	RecJ-like phosphoesterase
Q9V114	PAB0424	RpoA1	103	907	4±2	2±1	DNA-directed RNA polymerase subunit A'
Q9V1Z1	PAB2165	RPA32	31	272	3±1	3±1	RPA32 subunit
Q9UYJ6	PAB1366	DHH phosphohydrolase	36	330	3±0	2±1	DHH family phosphohydrolase
Q9V1U8	PAB2397	Rps4e	28	243	3±2	3±1	30S ribosomal protein S4e
Q9UZ86	PAB2423	Rgy	140	1214	2±1	3±2	Reverse gyrase
<i>Q9V2L5</i>	<i>PAB2306</i>	<i>F1pA</i>	<i>25</i>	<i>227</i>	<i>16±1</i>	<i>10±4</i>	<i>Fibrillar-like rRNA/tRNA 2'-O-methyltransferase</i>
<i>Q9UZN6</i>	<i>PAB1633</i>	<i>PINA</i>	<i>68</i>	<i>608</i>	<i>5±3</i>	<i>7±6</i>	<i>PIN domain ATPase</i>
<i>Q9UXS5</i>	<i>PAB1167</i>	<i>Rpl10</i>	<i>37</i>	<i>341</i>	<i>4±1</i>	<i>3±2</i>	<i>50S ribosomal protein L10</i>
<i>Q9V118</i>	<i>PAB0421</i>	<i>Rrp42</i>	<i>29</i>	<i>274</i>	<i>6±1</i>	<i>2±2</i>	<i>Rrp42p</i>
<i>Q9V2F9</i>	<i>PAB0071</i>	<i>AubA</i>	<i>53</i>	<i>469</i>	<i>6±1</i>	<i>0±0</i>	<i>RNA-binding protein AU-1</i>
<i>Q9V120</i>	<i>PAB0419</i>	<i>Rrp4</i>	<i>29</i>	<i>265</i>	<i>5±1</i>	<i>2±2</i>	<i>Rrp4p</i>
<i>Q9V113</i>	<i>PAB0425</i>	<i>RpoA2</i>	<i>44</i>	<i>397</i>	<i>5±2</i>	<i>1±1</i>	<i>DNA-directed RNA polymerase subunit A''</i>
<i>Q9UYD9</i>	<i>PAB1036</i>	<i>Unch.</i>	<i>40</i>	<i>340</i>	<i>4±1</i>	<i>3±3</i>	<i>Uncharacterized protein</i>
<i>Q9UY55</i>	<i>PAB0951</i>	<i>Unch.</i>	<i>28</i>	<i>247</i>	<i>2±1</i>	<i>0±0</i>	<i>Uncharacterized protein</i>
<i>Q9V2E3</i>	<i>PAB0081</i>	<i>Unch.</i>	<i>37</i>	<i>300</i>	<i>2±1</i>	<i>0±0</i>	<i>Uncharacterized protein</i>
<i>Q9UZX0</i>	<i>PAB1686</i>	<i>Unch.</i>	<i>36</i>	<i>316</i>	<i>3±1</i>	<i>2±2</i>	<i>Uncharacterized protein</i>

Table 2: Protéines les plus significativement retrouvées avec **aRNase J** déterminées par «Pull-down» de **Pab-aRNase J-(His)₆**

(Plateforme d'Analyse Protéomique de Paris Sud Ouest)

Les protéines ont été classées par comparaison et identification par spectrométrie de masse effectuée par la Plateforme d'Analyse Protéomique de Paris Sud Ouest (PAPPSO). Les accessions des protéines et la description des gènes sont issus de la base de données d'Uniprot (<http://www.uniprot.org/uniprot>). Seules les protéines ayant une valeur inférieure à 0.01 (haut du tableau) au test de student, 0.05 (bas du tableau) et un nombre de spectre (#) dans le «Pull-down» au-dessus de 5 sont montrées. Les protéines sensibles au traitement de nucléase sont représentées en italique. Pour chaque protéine, la masse molaire en kDa (M) et le cluster des groupes orthologues (COG) correspondant sont indiqués. *Unch* pour les protéines non caractérisées.

Accession	Gene	Name	kDa	AA	ENuc-	ENuc+	Description
Q9UZ78	PAB2428	TmcA	94	817	35±1	35±5	tRNA(Met) cytidine acetyltransferase
Q9V115	PAB0423	RpoB	127	1117	27±1	25±3	DNA-directed RNA polymerase subunit beta
Q9V192	PAB0367	Eno-like	37	342	25±3	21±4	Eno-like enolase related
G8ZHS0	PAB2163	RPA41	43	375	23±0	22±0	RPA41 subunit
Q9V1Z1	PAB2165	RPA32	31	272	17±3	16±3	RPA32 subunit
Q9UZ86	PAB2423	Rgy	140	1214	17±4	14±5	Reverse gyrase
Q9V133	PAB2412	Unch.	48	415	15±1	16±4	DUF530
Q9V1F2	PAB0316	DnaG	50	447	13±1	12±1	DnaG
Q9V2L4	PAB2305	Nop5p	47	404	13±1	14±2	Nop58p-like pre mRNA splicing
Q9V114	PAB0424	RpoA1	103	907	12±0	7±2	DNA-directed RNA polymerase subunit A'
Q9V2L5	PAB2306	FlpA	25	227	12±1	11±1	Fibrillar-like rRNA/tRNA 2'-O-methyltransferase
Q9V076	PAB1751	aRNase J	49	451	11±2	8±0	Ribonuclease J (5'-3' Exo)
Q9UY85	PAB1284	RecJ-SOD	83	740	10±2	9±1	RecJ-like phosphoesterase
Q9UY07	PAB1115	GraD-2	46	419	8±7	8±6	Glucose-1-phosphate thymidyltransferase
P62008	PAB0460	Rpl7ae	13	123	8±1	4±1	50S ribosomal protein L7Ae
Q9UZD0	PAB0810	Predicted ATPase	61	548	7±1	6±1	Predicted ATPase
G8ZJS6	PAB0744	Lhr2	100	867	6±1	7±1	Lhr-2 large helicase-related
Q9V089	PAB2390	Sun protein	52	450	6±0	5±1	Sun protein (Fmu protein)
Q9V1T5	PAB2120	Rpl3	40	361	6±2	5±2	50S ribosomal protein L3
Q9V113	PAB0425	RpoA2	44	397	6±1	4±0	DNA-directed RNA polymerase subunit A''
Q9V1V5	PAB2136	Rps5	26	236	6±1	5±1	30S ribosomal protein S5
Q9UYS8	PAB1430	TopA	78	685	4±1	4±1	DNA topoisomerase 1
Q9UZN6	PAB1633	PINA	68	608	4±1	3±0	PIN domain ATPase
Q9V0E2	PAB0569	Unch.	30	267	4±0	3±1	Proteasome chaperone
Q9V1V4	PAB2135	Rpl18	23	203	4±2	1±0	50S ribosomal protein L18
Q9V1V6	PAB2137	Rpl30p	17	155	4±1	4±1	50S ribosomal protein L30P
Q9V196	PAB0365	Rpl13	16	142	3±1	4±1	50S ribosomal protein L13
G8ZHR3	PAB0190	Hef	85	749	3±3	2±2	Hef nuclease
Q9V1U6	PAB2436	Rpl14	15	141	3±1	2±0	50S ribosomal protein L14
Q9V0G8	PAB1813	Rps19e	17	150	3±1	3±1	30S ribosomal protein S19e
Q9V116	PAB7151	RpoH	9	82	3±1	2±1	DNA-directed RNA polymerase H
Q9UZP0	PAB0732	RpoF	14	121	3±1	2±1	DNA-directed RNA polymerase, F
Q9UXX7	PAB1136	Rnp3	24	212	3±1	3±1	Ribonuclease P protein component 3
Q9UZK6	PAB1613	Unch.	30	262	3±2	1±1	CRISPR-associated. endo
Q9V181	PAB2000	Unch.	42	370	2±0	3±1	Uncharacterized protein
Q9V191	PAB0368	Rps2	23	202	2±1	4±2	30S ribosomal protein S2
P61992	PAB0361	Rps4	23	180	2±0	3±0	30S ribosomal protein S4
<i>Q9VIN3</i>	<i>PAB7094</i>	<i>AlbA</i>	<i>10</i>	<i>93</i>	<i>24±3</i>	<i>10±1</i>	<i>DNA/RNA- binding protein Alba</i>

<i>Q9V120</i>	<i>PAB0419</i>	<i>Rrp4</i>	29	265	9±1	0±0	<i>Rrp4p</i>
<i>Q9V2F9</i>	<i>PAB0071</i>	<i>AubA</i>	53	469	9±1	1±1	<i>RNA-binding protein AU-1</i>
<i>Q9V118</i>	<i>PAB0421</i>	<i>Rrp42</i>	29	274	4±2	0±0	<i>Rrp42p</i>
<i>Q9V119</i>	<i>PAB0420</i>	<i>Rrp41</i>	27	249	4±1	0±0	<i>Rrp41p</i>
<i>Q9UY91</i>	<i>PAB1064</i>	<i>Unch.</i>	27	239	4±1	0±0	<i>Uncharacterized protein</i>
<i>Q9UZF3</i>	<i>PAB1584</i>	<i>Unch.</i>	19	164	4±2	0±0	<i>uncharacterized protein</i>
<i>Q9UWR8</i>	<i>PAB1166</i>	<i>Rpl1</i>	24	219	3±1	0±0	<i>50S ribosomal protein L1P</i>

Table 3A: Protéines les plus significativement retrouvées avec **ASH-Ski2** déterminées par «Pull-down» de **Pab-(His)₆-ASH-Ski2**

(Plateforme d'Analyse Protéomique de Paris Sud Ouest)

Les protéines ont été classées par comparaison et identification par spectrométrie de masse effectuée par la Plateforme d'Analyse Protéomique de Paris Sud Ouest (PAPPSO). Les accessions des protéines et la description des gènes sont issus de la base de données d'Uniprot (<http://www.uniprot.org/uniprot>). Seules les protéines ayant une valeur inférieure à 0.01 (haut du tableau) au test de student, 0.05 (bas du tableau) et un nombre de spectre (#) dans le «Pull-down» au-dessus de 2 sont montrées. Les protéines sensibles au traitement de nucléase sont représentées en italique. Pour chaque protéine, la masse molaire en kDa (M) et le cluster des groupes orthologues (COG) correspondant sont indiqués. *Unch* pour les protéines non caractérisées.

Accession	Gene	Name	kDa	AA	ENuc-	ENuc+	Description
Q9UZ78	PAB2428	TmcA	94	817	40±2	41±1	tRNA(Met) cytidine acetyltransferase
Q9V089	PAB2390	Sun protein	52	450	37±0	28±1	Sun protein (Fmu protein)
Q9V1F2	PAB0316	DnaG	50	447	17±1	13±2	DNA primase DnaG
Q9V2L4	PAB2305	Nop5p	47	404	16±1	13±1	Nop58p-like pre mRNA splicing
Q9V2L5	PAB2306	FlpA	25	227	16±5	11±0	Fibrillar-like rRNA/tRNA 2'-O-methyltransferase
Q9V192	PAB0367	Eno-like	37	342	12±1	12±3	Eno-like enolase related
Q9V076	PAB1751	aRNase J	49	451	11±0	8±0	Ribonuclease J (5'-3' Exo)
Q9V115	PAB0423	RpoB	127	1117	11±1	7±3	DNA-directed RNA polymerase subunit beta
P62008	PAB0460	Rpl7ae	13	123	11±5	3±1	50S ribosomal protein L7Ae
Q9V1V5	PAB2136	Rps5	26	236	9±2	8±1	30S ribosomal protein S5
Q9UY85	PAB1284	RecJ-SOD	83	740	7±4	3±3	RecJ-like phosphoesterase
Q9V1V6	PAB2137	rpl30p	17	155	7±2	5±2	50S ribosomal protein L30P
Q9V1T5	PAB2120	Rpl3	40	361	6±0	5±1	50S ribosomal protein L3
G8ZHS0	PAB2163	RPA41	40	358	6±2	5±0	RPA41 subunit
Q9V1U8	PAB2397	Rps4e	28	243	5±3	3±0	30S ribosomal protein S4e
Q9V1T8	PAB2122	Rpl2	26	239	5±2	3±1	50S ribosomal protein L2
Q9V236	PAB0161	Unch.	23	204	5±2	2±0	Uncharacterized protein
Q9V196	PAB0365	Rpl13	16	142	4±0	3±0	50S ribosomal protein L13
Q9V1U6	PAB2436	Rpl14	15	141	4±0	2±1	50S ribosomal protein L14
Q9UY42	PAB1092	ATP pyrophosphatase	35	303	4±1	1±1	N-type ATP pyrophosphatase superfamily
P61992	PAB0361	Rps4	23	180	4±2	4±1	30S ribosomal protein S4
Q9V0G8	PAB1813	Rps19e	17	150	4±2	2±0	30S ribosomal protein S19e
Q9UZN6	PAB1633	PINA	68	608	3±0	5±1	PIN domain ATPase
Q9UXS5	PAB1167	Rpl10	37	341	3±0	4±1	50S ribosomal protein L10
Q9V133	PAB2412	Unch.	48	415	3±1	4±0	DUF530
Q9V014	PAB1719	Unch.	43	373	3±1	4±1	Methylase
Q9V191	PAB0368	Rps2	23	202	3±1	4±2	30S ribosomal protein S2
Q9V114	PAB0424	RpoA1	103	907	3±1	2±1	DNA-directed RNA polymerase subunit A'
Q9UXX7	PAB1136	Rnp3	24	212	3±2	3±1	Ribonuclease P component 3
P62010	PAB0362	Rps11	14	137	3±2	3±1	30S ribosomal protein S11
Q9V1V1	PAB2132	Rpl6	20	184	3±2	2±1	50S ribosomal protein L6
Q9UZJ5	PAB7290	Rpl14e	8	82	3±1	2±1	50S ribosomal protein L14e
Q9UZL4	PAB0749	Rps8e	14	127	2±1	2±1	30S ribosomal protein S8e
Q9V1T4	PAB2119	Unch.	29	258	2±1	1±1	Methyl-transferase like
Q9V197	PAB0364	Rpl18e	13	120	2±1	1±1	50S ribosomal protein L18e
<i>Q9VIN3</i>	<i>PAB7094</i>	<i>AlbA</i>	<i>10</i>	<i>93</i>	<i>40±8</i>	<i>8±2</i>	<i>DNA/RNA-binding protein Alba</i>
<i>Q9UYV2</i>	<i>PAB0931</i>	<i>CysS</i>	<i>56</i>	<i>477</i>	<i>12±4</i>	<i>1±1</i>	<i>Cysteine tRNA ligase</i>
<i>Q9V120</i>	<i>PAB0419</i>	<i>Rrp4</i>	<i>29</i>	<i>265</i>	<i>12±3</i>	<i>0±0</i>	<i>Rrp4</i>

<i>Q9V118</i>	<i>PAB0421</i>	<i>Rrp42</i>	29	274	9±2	0±0	<i>Rrp42</i>
<i>Q9V2F9</i>	<i>PAB0071</i>	<i>AubA</i>	53	469	7±0	1±0	<i>RNA-binding protein AU-1</i>
<i>Q9V119</i>	<i>PAB0420</i>	<i>Rrp41</i>	27	249	7±1	0±0	<i>Rrp41</i>
<i>Q9UZF3</i>	<i>PAB1584</i>	<i>Unch.</i>	19	164	6±2	0±0	<i>uncharacterized protein</i>
<i>Q9V1A5</i>	<i>PAB05</i>	<i>TruB</i>	38	334	5±1	3±0	<i>tRNA pseudo uridine synthetase</i>
<i>Q9UY91</i>	<i>PAB1064</i>	<i>Unch</i>	27	239	5±3	0±0	<i>Uncharacterized protein</i>
<i>Q9V2E3</i>	<i>PAB0081</i>	<i>Unch</i>	37	300	4±1	0±0	<i>Uncharacterized protein</i>
<i>Q9V113</i>	<i>PAB0425</i>	<i>RpoA2</i>	44	397	3±0	1±1	<i>DNA-directed RNA polymerase subunit A''</i>
<i>Q9V1V4</i>	<i>PAB2135</i>	<i>Rpl18</i>	23	203	3±1	2±1	<i>50S ribosomal protein L18</i>
<i>Q9UYB6</i>	<i>PAB1306</i>	<i>eIF-2B</i>	31	276	3±1	1±1	<i>Putative translation initiation factor eIF-2B subunit 2-like</i>
<i>Q9V1T7</i>	<i>PAB7083</i>	<i>Rpl23p</i>	9	86	2±1	0±0	<i>50S ribosomal protein L23P</i>
<i>Q9V2M2</i>	<i>PAB2314</i>	<i>Csl4</i>	22	199	2±2	0±0	<i>Csl4</i>

Table 3B: Protéines les plus significativement retrouvées avec **ASH-Ski2** déterminées par «Pull-down» de **Pab-ASH-Ski2-(His)₆** (Plateforme d'Analyse Protéomique de Paris Sud Ouest)

Les protéines ont été classées par comparaison et identification par spectrométrie de masse effectuée par la Plateforme d'Analyse Protéomique de Paris Sud Ouest (PAPPSO). Les accessions des protéines et la description des gènes sont issus de la base de données d'Uniprot (<http://www.uniprot.org/uniprot>). Seules les protéines ayant une valeur inférieure à 0.01 (haut du tableau) au test de student, 0.05 (bas du tableau) et un nombre de spectre (#) dans le «Pull-down» au-dessus de 5 sont montrées. Les protéines sensibles au traitement de nucléase sont représentées en italique. Pour chaque protéine, la masse molaire en kDa (M) et le cluster des groupes orthologues (COG) correspondant sont indiqués. *Unch* pour les protéines non caractérisées.

Accession	Gene	Name	kDa	AA	ENuc-	ENuc+	Description
Q9UZ78	PAB2428	TmcA	94	817	44±4	45±6	tRNA(Met) cytidine acetyltransferase
Q9V148	PAB0399	Unch.	15,7	139	9±1	9±2	Hydrogenase nickel incorporation protein HypA
Q9V225	PAB2398	Unch.	35,7	324	3±2	4±1	Putative GTPase
<i>Q9UZ86</i>	<i>PAB2423</i>	<i>Rgy</i>	<i>140</i>	<i>1214</i>	<i>47±10</i>	<i>7±1</i>	<i>Reverse gyrase</i>
<i>Q9UYS8</i>	<i>PAB1430</i>	<i>TopA</i>	<i>78</i>	<i>685</i>	<i>32±7</i>	<i>12±7</i>	<i>DNA topoisomerase 1</i>
<i>Q9V076</i>	<i>PAB1751</i>	<i>aRNase J</i>	<i>49</i>	<i>451</i>	<i>11±5</i>	<i>4±2</i>	<i>Ribonuclease J (5'-3' Exo)</i>
<i>Q9V113</i>	<i>PAB0425</i>	<i>RpoA2</i>	<i>44</i>	<i>397</i>	<i>12±2</i>	<i>4±1</i>	<i>DNA-directed RNA polymerase subunit A''</i>

Table 3C: Protéines les plus significativement retrouvées avec **ASH-Ski2** délétée de son domaine N-terminal déterminées par pull down de *Pab-(His)₆-ΔN-ASH-Ski2* (*Plateforme* d'Analyse Protéomique de Paris Sud Ouest)

Les protéines ont été classées par comparaison et identification par spectrométrie de masse effectuée par la *Plateforme* d'Analyse Protéomique de Paris Sud Ouest (PAPPSO). Les accessions des protéines et la description des gènes sont issus de la base de données d'Uniprot (<http://www.uniprot.org/uniprot>). Seules les protéines ayant une valeur inférieure à 0.01 (haut du tableau) au test de student, 0.05 (bas du tableau) et un nombre de spectre (#) dans le «Pull-down» au-dessus de 5 sont montrées. Les protéines sensibles au traitement de nucléase sont représentées en italique. Pour chaque protéine, la masse molaire en kDa (M) et le cluster des groupes orthologues (COG) correspondant sont indiqués. *Unch* pour les protéines non caractérisées.

	Prey of ASH-Ski2/aRNase J network	SR-	SR+	Bait proteins
Q9V2M1	ASH-Ski2	0,16	0,18	Q9V1A9 Ribonuclease HII
Q9UY85	RecJ-SOD (PAB1284)	0,04	0,02	Q9V2G5 DNA2 homolog
Q9UZM4	Lhr-2 large helicase-related protein	0,02	0,01	Q9UZY0 N-glycosylase/DNA lyase/OGG1
		0,02	0,01	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,06	0,05	RPA
		0,17	0,04	Q9V278 DNA repair protein RAD25
Q9V076	aRNase J	0,05	0,08	Q9V2E8 NucS
		0,06	0,07	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,09	0,14	Q9V233 radA
Q9V114	RNAP rpoA1	0,05	0,05	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,06	0,10	RPA
		0,07	0,07	Q9V278 DNA repair protein RAD25
		0,12	0,15	Q9UYR7 MCM
		0,12	0,15	P0CL74 DNA Ligase
		0,13	0,10	Q9UZY0 N-glycosylase/DNA lyase/OGG1
Q9V113	RNAP rpoA2	0,01	0,01	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,03	0,03	Q9UZY0 N-glycosylase/DNA lyase/OGG1
		0,02	0,01	Q9V278 DNA repair protein RAD25
		0,05	0,08	RPA
		0,07	0,12	Q9V2E8 NucS
Q9V115	RNAP rpoB	0,09	0,08	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,16	0,17	Q9V278 DNA repair protein RAD25
Q9V1Z1	RPA32	0,01	0,01	Q9UZY0 N-glycosylase/DNA lyase/OGG1
		0,02	0,07	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,04	0,09	Q9V2F4 DNA polymerase D Is
		0,04	0,02	Q9V2G5 DNA2 homolog
		0,05	0,13	Q9V278 DNA repair protein RAD25
Q9UZ86	Reverse gyrase	0,02	0,02	Q9UZY0 N-glycosylase/DNA lyase/OGG1
		0,05	0,05	Q9UZ73 AlkA 3-methyladenine DNA glycosylase
		0,06	0,14	RPA
		0,09	0,12	P0CL74 DNA Ligase
		0,14	0,12	Q9V2G5 DNA2 homolog
		0,16	0,12	Q9V278 DNA repair protein RAD25

Table 4 Tableau comparatif des protéines présentes dans les réseaux d'interaction de aRNase J et AH-Ski2 et dans les réseaux autour des protéines impliquées dans le métabolisme de l'ADN issus de Pluchon *et al*, 2013.

			PREY PROTEINS	
			aRNase J	ASH-Ski2
BAIT PROTEINS	β-CASP ribonuclease	(His) ₆ -aRNase J		<i>n.d.</i>
		aRNase J-(His) ₆		-
		(His) ₆ -aCPSF1	-	<i>n.d.</i>
	Ski-like helicase	(His) ₆ -ASH-Ski2	✓	
		ASH-Ski2-(His) ₆	✓	
		(His) ₆ -Hel308-Ski2	-	<i>n.d.</i>
	Exosome sub-unit	DnaG-(His) ₆	-	-
		(His) ₆ -Rrp41	(✓)	-
(His) ₆ -RecJ-SOD		-	✓	

		PREY PROTEINS		
		aRNase J	aRNase J- ΔL1	
BAIT PROTEINS (His) ₆ -Nter	ASH-Ski2	WT	✓	✓
		ΔN	-	<i>n.d.</i>
		ΔC	✓	<i>n.d.</i>
		ΔNC	✓	<i>n.d.</i>
		C124A	(✓)	<i>n.d.</i>
	Nter Domain	WT	✓	<i>n.d.</i>
C98A		✓	<i>n.d.</i>	

Table 5 Tableau récapitulatif des tests de co-purifications. Récapitulatif des tests d'interaction entre *Pab*-aRNase J et *Pab*-ASH-Ski2 avec les protéines des réseaux d'interaction (panel haut).

Tests d'interaction entre *Pab*-aRNase J comme proie avec *Pab*-ASH-Ski2 et ces variants comme appâts.

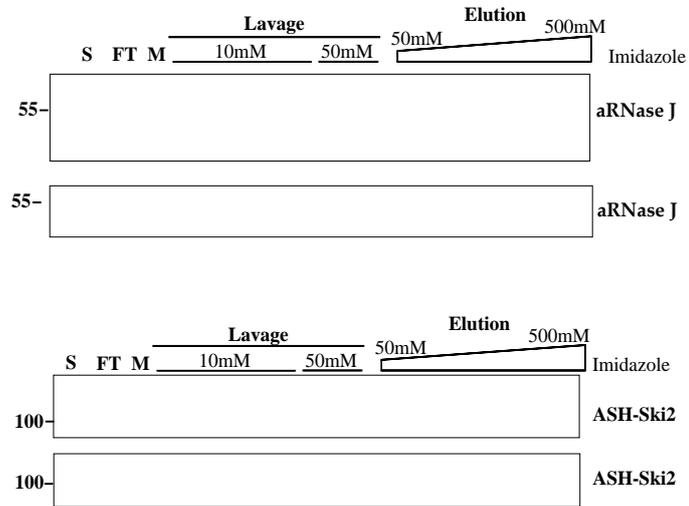


Figure Annexe: Tests contrôles d'affinité des protéines aRNase J et ASH-Ski2 pour la colonne de nickel. Les extraits ont été préparés comme pour **Figure 24** avec aRNase J et ASH-Ski2 non étiquetée. Résultats de la séparation des différentes fractions sur gel 4-20% SDS-PAGE révélés par Instant Blue (Panel haut) et par Western Blotting (panel bas).

AN OVERVIEW OF RIBONUCLEASE REPERTOIRE AND RNA PROCESSING PATHWAYS IN ARCHAEA

Duy Khanh Phung, Marie Bouvier and Béatrice Clouet-d'Orval*

Laboratoire de Microbiologie et de Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, F-31062 Toulouse, France;

*Corresponding author

Abstract

RNA processing plays a crucial role in post-transcriptional regulation of gene expression. Work conducted in Bacteria and Eukarya has defined the predominant mRNA maturation and decay pathways, as well as enzymes and cofactors responsible for these processes. In contrast, our knowledge of the mechanisms controlling RNA quality and processing in Archaea is more fragmentary. In essence, the major actors of RNA processing are ribonucleases acting in cleaving or trimming RNA molecules according to their nature and fate, making these enzymes fascinating and important players to study. More than 30 families of ribonucleases have been described in Bacteria and Eukarya, while only few have been identified in Archaea. This chapter is focused on the major ribonucleases in Archaea. After an overview of archaeal cellular RNA biotypes, we present synthetic up-to-date repertoire of the archaeal ribonuclease families as well as our state of knowledge on their roles in dedicated RNA processing pathways. In addition to this general description of archaeal RNA processing actors, specific pathways involved in processing of rRNAs, tRNAs, crRNAs and C/D sRNAs are detailed in other chapters.

1. Introduction

The survival of a cell depends on its ability to rapidly adapt to changing environmental conditions in which RNA biology plays a critical role as it directs multiple mechanistically distinct processes. Indeed, expression of genetic information relies on three major types of RNAs: messenger RNAs (mRNAs) that are intermediates in defining the proteome; ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) that are effector molecules acting together to decode mRNA sequence information; and finally a plethora of noncoding regulatory RNAs (ncRNAs) that can regulate mRNA (sRNA/microRNA/siRNA), have key roles in mRNA splicing (snRNAs) or guide RNA modifications (box C/D and H/ACA guide RNAs).

Regulation of gene expression occurs at multiple levels including post-transcriptional processes, referred to as RNA processing, which are fundamental in defining the phenotypic characteristic of a cell. In Archaea, this level of gene regulation remains to be largely explored. Identification and characterization of actors involved in archaeal RNA maturation and decay pathways are still in their early stages. While it is now well established that archaeal and eukaryotic cells share many key components of their replication, transcription, and translation machineries (Brochier-Armanet et al, 2011; Spang et al, 2015), it remains to assess in which extent this feature apply to RNA processing and RNA regulatory factors. Providing answers to these points will bring perspectives in understanding the evolutionary route of the players acting in the biology of functional archaeal RNAs.

The level of which any RNA is expressed is determined to the same extent by its rates of transcription and by its decay. In addition, primary transcripts undergo multiple covalent modifications to reach their functional state. These processes involve the orderly action of a battery of cellular enzymes from RNA synthesis and maturation to decay to precisely monitor their quality and level. Therefore, cells depend on RNA processing systems to produce mature RNA from immature precursors as well as to discard any transcriptional byproducts and malformed transcripts. The enzymes directly responsible of these processes are ribonucleases (RNases) that catalyze the exo- or endoribonucleolytic cleavage of a phosphodiester bond from an RNA molecule. They are essential both for nonspecific RNA degradation and for numerous forms of RNA processing. Indeed, they control the fate of cellular mRNAs and make mature rRNAs, tRNAs and regulatory ncRNAs from precursors which undergo complex processes that require a high number of modifications in addition to 5'- and 3'-end

processing and, often, intron splicing. Most of the known RNases are protein enzymes, but, in several cases, the catalytic moiety is an RNA molecule.

While the action of endoribonucleases (endoRNases) initiates RNA maturation and decay through internal cleavage within the transcript sequence, the exoribonucleases (exoRNases) digest RNA from either end. Consequently, cleavage by endoRNases generates substrate for 3'-5' and 5'-3' exoribonucleolytic decay, resulting in a tight cooperation between different RNA degrading systems. In this framework, the nature of the 5' and 3' ends of an RNA molecule defining its accessibility to ribonucleases is critical in its outcome. Further characteristics of ribonucleases included their ability to cleave single and/or double stranded RNA or RNA/DNA hybrid, the nature of their byproduct, their specificity for substrates of defined shape and sequence, their ability to digest DNA besides RNA, and their processive or distributive action. Systematically, exoRNases function by one of two mechanisms, either hydrolytically yielding nucleoside monophosphate products or phosphorolytically using orthophosphate as a nucleophile and producing nucleoside diphosphates in a reversible reaction.

The stability of RNA is among other things determined by a cell-specific set of RNases that can be specific or shared by several RNA processing pathways. In addition, other RNA-related proteins or auxiliary enzymes including RNA helicases, poly(A) polymerases and pyrophosphohydrolases can form RNA degrading complexes or machines to enhance and specify their activity. Moreover, the action of RNases, are tightly controlled through several means to prevent incorrect processing. For instance, the recruitment of RNA-modifying enzymes to their respective targets is specified by number of cis-acting RNA sequence elements, as well as a large repertoire of trans-acting proteins and ncRNAs.

In this context, this is not surprising that RNases, acknowledged as precision tools for RNA, are often essential in cell viability (for review see, (Arraiano et al, 2013)). Understanding the unique and shared roles of these enzymes in the cell is a foremost challenge to decipher post-transcriptional regulation pathways of gene expression. Excellent in-depth reviews covering enzymes involved in critical aspects of RNA processing in Bacteria and Eukarya were highlighted in a special issue of "Biochimica and Biophysica Acta" edited by Stoecklin and Muhlemann (Stoecklin & Muhlemann, 2013) dedicated to RNA Decay mechanisms. In light of the crucial role of RNases in bacterial and eukaryal cellular metabolism, it is of major

importance to identify their archaeal counterparts to deeply understand post-transcriptional gene expression regulation in Archaea.

As mentioned above, archaeal and eukaryotic cells share many key components of their informational machineries (Brochier-Armanet et al, 2011). It is interesting to note that this affiliation extends to several of RNA processing enzymes and RNA regulatory factors which include the exosome and other translation-related systems, as snoRNP guide-dependent modifications and tRNA intron processing. However, homologues to bacterial-specific enzymes have also been identified in archaeal genomes. To this respect, archaeal cells possess original RNA processing, turnover and quality control settings with mosaic features which remain to be explored. This highlights the advantage of developing archaeal models to gain further mechanistic and evolutionary information of fundamental processes across the three domains of life.

2. Archaeal RNA biotypes

The common archaeal RNA biotype that is represented in Figure 1 includes mRNAs, rRNAs, tRNAs and diverse group of ncRNAs such as box C/D and H/ACA guide RNAs, the RNA moiety of RNase P, CRISPR RNAs and diverse specific regulatory ncRNAs. Archaea undoubtedly use a variety of enzymatic pathways to mature and degrade all the cellular types of RNA (**Figure 1**).

In contrast to Eukarya which uses several classes of RNA polymerases (RNAPs) that specifically transcribe distinct and non-overlapping subsets of genes, unique transcription machinery is in charge of the transcription of all types of RNAs in Archaea (**Figure 1**). Nevertheless, the archaeal RNAP is closely related to the eukaryotic RNA polymerase II (RNAPII) in terms of subunit composition, structure, use of general factors and mechanism of action (Werner, 2007; Werner & Grohmann, 2011). The core of the archaeal and eukaryotic transcription machineries is conserved and the minimal promoter (TATA box & B-recognition element) and site-specific transcription start sites are identical. Briefly, binding of TATA-binding protein (TBP) to the promoter induces a cascade that leads of protein recruitment to transcription initiation; transcription factor B (TFB) binds to the TBP-DNA complex, and the DNA–TBP–TFB complex subsequently recruits RNAP and transcription factor E (TFE) (discussed in chapter 1-Werner). In contrast to the basal transcription machinery, gene-specific signal-responsive transcription regulation is mediated by small bacterial-like transcription factors (Aravind, 1999; Peeters et al, 2013). Recently, principles

underlying how these factors cooperate in nucleoid structuring and gene regulation emerged. Accumulating evidence suggests that nucleosomal organization in Archaea is interlinked with transcription processes and is unique in many aspects (Peeters et al, 2015) (discussed in chapter 2-Peeters).

Less is understood on transcription termination and its regulation remains a mystery in Archaea. Whether the formation of RNA 3' end depends on transcription termination factors as in eukaryotic cells (Dominski et al, 2013; Proudfoot, 2011) needs to be determined. The first experimental evidence for intrinsic termination of archaeal transcription show that archaeal transcription termination is stimulated by oligo(T) sequences (Santangelo et al, 2009). More recently a broader sequence diversity of termination motifs was revealed by the term-seq sequencing method to comprehensively map RNA 3' termini of hundreds of genes suggesting an unexpected complexity of archaeal mRNA 3' ends (Dar et al, 2016; Garrett, 2016).

Despite the similarities between archaeal and eukaryal transcription, archaeal mRNAs share bacterial mRNA characteristics with no introns, no 5' methyl caps and no long 3' poly A tails but with 5' triphosphorylated ends and often polycistronic structure (**Figure 1**). However, early data suggest that some archaeal mRNAs have significantly longer half-life than most mRNAs in bacteria (Evguenieva-Hackenberg & Klug, 2009). It was reported that transcripts half-lives vary from 7 to 57 min in *Methanococcus vannielii* (Hennigan & Reeve, 1994), from 6 to 120 min in *Sulfolobus solfataricus* (Bini et al, 2002) and from 4 to 80 min in *Haloferax mediterranei* (Jager et al, 2002). These initial studies proposed that the transcript 3' ends are generated by an endonucleolytic cleavage followed by a directional 3' to 5' degradation. However, no specific endonucleolytic cleavage signal has been characterized so far that could support this early model of mRNA decay. In addition, more recently, *in vitro* and *in vivo* evidence suggested that the translation initiation factor a/eIF2- γ subunit binds to RNA 5'-triphosphorylated ends and protects transcripts from a 5' end-dependent degradation pathway in the crenarchaeota *S. solfataricus* (Arkhipova et al, 2015; Hasenohrl et al, 2008). In this case, a/eIF2- γ subunit exhibits an additional function with resemblance to the eukaryotic cap-complex. This strongly suggests that a directional 5' to 3' degradation pathway may exist in archaeal cells. Nevertheless, pathways and factors controlling mRNA stability in Archaea are still hardly documented.

More generally, in Archaea, all tRNAs are issued from primary transcripts which are extensively processed to yield the mature and functional forms (**Figure 1**). This process requires the universal ribonucleoprotein RNase P (Frank & Pace, 1998) (discussed in chapter 7-Kimura) and RNase Z (Holzle et al, 2008; Redko et al, 2007; Schiffer et al, 2002). In addition an RNA-splicing endoRNase (EndA) removes introns from archaeal pre-tRNA and also participates in rRNA processing (Li et al, 1998; Tang et al, 2002; Thompson & Daniels, 1988) (**Figure 1**). All mature tRNAs harbored at their 3' end a CCA triplet that is essential for both aminoacylation and recognition of the tRNA by the ribosome. For many bacterial and some archaeal tRNAs, the CCA is encoded by their respective genes and therefore is already present in the precursor (Marck & Grosjean, 2002). In other cases, including all eukaryotic and most archaeal tRNAs, the CCA is added post-transcriptionally as a part of the maturation process (Cho et al, 2005; Weiner, 2004).

As in Eukarya and Bacteria, rRNAs are produced from polycistronic RNA precursor (**Figure 1**) (discussed in chapter 6-Ferreira-Cerca). The number of rRNA operons per genome varies from one to four depending on the archaeal specie. rRNA operons contain the 16S and 23S genes in crenarchaea with the addition of, and 5S genes in euryarchaea with a tRNA Ala gene located in the internal transcribed spacer (ITS) and a tRNA Cys gene in the distal portion of the precursor (Klug et al, 2007).

Modifying guide RNAs, abundant in Archaea and named C/D box and H/ACA box sRNAs (**Figure 1**), guide 2'O-methylation at the ribose moiety and pseudouridylation at specific rRNA and tRNA nucleotides, respectively (Dennis & Omer, 2005; Dennis et al, 2001). They are central in the landscape of RNA modifications in the eukaryal and archaeal cells. Both archaeal C/D box and H/ACA box sRNAs associate with a set of proteins to form active ribonucleoproteins (RNPs) in which an intact Kink-turn or Kink-loop RNA motif is critical (Charpentier et al, 2005; Lapinaite et al, 2013; Li & Ye, 2006; Nolivos et al, 2005; Yip et al, 2016; Zago et al, 2005) (discussed in chapter 9-Randau).

Finally, crRNAs called CRISPR RNAs are, so far, exclusively involved in cellular defense against infectious genetic elements, such as plasmids and viruses (**Figure 1**). Several types of CRISPR-cas immunity systems have been described and are found in the majority of archaeal genomes (Makarova et al, 2015). A hallmark of defense mechanisms based on clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins (Cas) are the crRNAs that guide these complexes in the destruction of invading DNA or RNA (for

review, (Barrangou et al, 2013)). The proteins encoded by the *cas* genes include predicted RNA-binding proteins, nucleases, helicases and polymerases. The transcription of the CRISPR arrays produces long pre-crRNA, which are further cleaved at each repeat into individual crRNA (**Figure 1**). Finally, during the interference phase, invading nucleic acids that match a crRNA are cleaved by effector RNP complexes consisting of dedicated Cas proteins associated to individual crRNAs (discussed in chapter 11-Marchfelder)

Altogether only a subset of archaeal ribonucleolytic enzymes has been assigned to specific pathways dictating the biogenesis and the fate of archaeal RNA biotypes. Many pathways and associated actors, especially those controlling mRNA processing and decay wait to be characterized (**Figure 1**).

3. Ribonuclease Families

3.1 RNA processing Actors in Bacteria and Eukarya

One of the main approaches used to decipher the main actors of RNA processing in Archaea is based on experience and knowledge accumulated in Eukarya and Bacteria. In Eukarya, it is now well established that multiple parallel, and partially redundant, mRNA turnover and surveillance pathways take place in the nucleus and cytoplasm of the cells. These processes are mainly 3'-to-5' and 5'-to-3' pathways that are respectively mediated by the 3'-5' Exosome complex and the 5'-3' XRN exonuclease family (for review see (Brooks, 2010; Henras et al, 2015; Nagarajan et al, 2013; Stoecklin & Muhlemann, 2013)). Ribosomal RNA maturation and ribosome biogenesis in Eukarya are also oriented processes starting in the nucleolus and ending in the cytoplasm that involve exo- and endoribonucleolytic processing events coupled with nucleotide modifications. The accuracy of pre-rRNA processing is tightly monitored by quality control mechanism (for review see (Fernandez-Pevida et al, 2015; Henras et al, 2015)).

Although RNA processing pathways including rRNA maturation and mRNA decay have been identified in several bacterial species, RNases and RNA degrading complexes are best studied in the model organisms *E. coli* and *B. subtilis* (Deutscher, 2009; Hui et al, 2014). In *E. coli*, the canonical RNA degradosome is composed of the essential endoRNase RNase E which serves as a scaffold of the complex by interacting with the exoRNase PNPase, the RNA helicase RhlB and the glycolytic enzyme enolase (Bandyra et al, 2013; Carpousis, 2007) (for review, (Mackie, 2013)). In *B. subtilis*, a central endoRNase, RNase Y, has been shown to

interact with the exoRNases PNPase and RNase J1, the DEAD box RNA helicase CshA and the two glycolytic enzymes enolase and frutokinase.

Throughout all domains of life, biogenesis and post-transcriptional modification processes involved in tRNA biology are the most conserved ones as they are fundamental in translating the genetic code but are also involved in regulating gene expression and in modulating apoptosis and several other biological processes (for review see (Huang & Hopper, 2016)). Finally many processes involving endo- and exo-RNases have been discovered in the late years, including for example RNases involved in bacterial CRISPR guide RNA maturation (Hochstrasser & Doudna, 2015) or those of the eukaryotic mi/siRNA pathways (Filipowicz et al, 2008).

Recently the identification of an ever-increasing number of ribonucleolytic enzymes combined with the availability of protein sequence of unknown function in databases has allowed the grouping of RNases in defined families. According to the presence of signature motifs in their amino acid sequence for protein enzymes, bacterial and eukaryal RNases involved in numerous physiological processes have been grouped into more than 30 families (for review, (Stoecklin & Muhlemann, 2013)).

3.2 Identified ribonuclease families in Archaea

Performing systematic comparative analysis and searching for gene sequences and gene-order conservation in the archaeal genomes opened the prospecting of RNA biology actors in Archaea (Koonin et al, 2001). One of the main examples of RNase prediction was the detection of an archaeal exosome by comparative genomic approaches using databases of protein domains (Koonin et al, 2001). Within the reported RNase families (Arraiano et al, 2013; Stoecklin & Muhlemann, 2013), only 9 includes archaeal counterparts (**Table 1**). It should be noted that genes encoding homologs of the key bacterial endoribonucleases- RNase E/G, RNase III and RNase Y families, and of the eukaryal XRN family, that are essential to maintain the integrity of the RNA pool in the cells, could not be identified in any up-to-date annotated archaeal genome. While earlier an RNase E-like activity has been previously described in haloarchaea (Franzetti et al, 1997), this assertion has been called into question since no such homologs could be identified by phylogenomic analysis (Quentin Y. & Clouet-d'Orval B., personal communication). Nonetheless a putative archaeal protein of the thermophilic archaeon *Pyrococcus furiosus* with short sequence similarity to the AU-binding

domain of RNase G, so-called FAU-1, has been shown to have RNA binding properties (Kanai et al, 2003).

The enzymatic activities of the archaeal ribonucleases identified have been characterized *in vitro* and are described one by one in the following sections (**Table 1 & Figure 2**). While tRNA and rRNA processing pathways are supported by strictly conserved endoRNases, the cellular RNA targets of most archaeal exoRNases remained to be identified (**Table 1**).

4. EndoRNases identified in rRNA, tRNA, crRNA processing and DNA maintenance pathways in Archaea

4.1. RNase P

In the different lifeforms, the majority of tRNA transcripts are synthesized as precursors containing 5' leader and 3' trailer sequences that should be accurately excised. The ribonucleoprotein RNase P is an endoRNase responsible for the removal of 5' leader sequences from tRNA precursors and, therefore, has a key role in cellular homeostasis and survival in most organisms. It is interesting to mention that the only known organism without RNase P in the archaeon *Nanoarchaeum equitans* presumably because the pre-tRNAs in this organism are transcribed without leader sequence (Randau et al, 2008) (**Figure 3**).

RNase P which has divergent scaffolds across the three domains of life is an Mg^{2+} -dependent endonucleolytic ribonucleoprotein (RNP) complex which catalyzes phosphodiester bond hydrolysis of tRNA precursors generating tRNAs with mature 5' phosphate-ends and 5' leaders with a 3' hydroxyl group. RNase P holoenzyme is shaped around an essential catalytic RNA subunit (termed RPR) with an average length of 400-nt, and respectively, one, at least four or 10 protein subunits (termed RPPs) in Bacteria, Archaea and Eukarya (for review see (Klemm et al, 2016; Lai et al, 2010; Samanta et al, 2016)).

Unlike the bacterial RPPs that are capable of efficient catalysis in the absence of protein, the archaeal and eukaryal RPRs depend fundamentally on protein for activity. Many structures have provided significant insight into structural features that are key for catalysis and substrate recognition (Gopalan, 2007). Archaeal RPR subunits are composed of two functional RNA domains, the substrate-binding domain (S-domain) and the catalytic domain (C-domain), which commonly resemble to the ancestral bacterial type A (**Figure 2A**) (Jarrous & Gopalan, 2010). However, a type M RPR is encoded by *Methanococci* and *Archaeoglobulus* (Pannucci et al, 1999) (**Figure 2A**). The biochemical properties of the

archaeal RPRs are similar to those of synthetic minimal bacterial RPRs, suggesting that the archaeal RNAs contain all of the elements required for substrate recognition and catalysis but are structurally defective in the absence of protein.

Archaeal RNase P contains at least 4 proteins – RPP21, RPP29, RPP30 and POP5 – (**Figure 2A**), with a fifth – L7Ae – associating with type M RNA, all of which are homologous to yeast and human nuclear RNase P proteins (Samanta et al, 2016). While there is no atomic resolution of full-archaeal RNase P enzymes, structures have been solved for each of the RPP subunits (Numata et al, 2004) (discussed in chapter 7-Kimura). Each subunit is briefly described in the following text.

The RPP21 subunit has a zinc ribbon in the C-terminal domain, a motif found in other nucleic-acid-binding proteins with four invariant cysteines. The RPP29 is characterized by a twisted barrel of seven antiparallel β -strands (Samanta et al, 2016). These two subunits associate to form the RPP21: RPP29 binary complex that functions in enhancing substrate specificity (Sinapah et al, 2011). This sub complex whose structure has been solved contacts the S-domain of the RNA moiety. The POP5 subunit is structurally similar to the unique bacterial RPP with β -strands forming a central cleft and near-universally conserved Arg/Lys and Phe or Tyr residues. The RPP30 shows the least sequence conservation but mutations at specific Arg residues exhibit decreases in activity (Samanta et al, 2016). The POP5:RPP30 binary complex with a hydrophobic interface contacts with the C-domain of the RNA moiety. As previously described for RPP21:RPP29, the POP5:RPP30 binary complex has been shown to affect cleavage site selection (Sinapah et al, 2011).

More recently, it has been shown that the 50S ribosomal protein L7Ae is a subunit of the type M archaeal RNase P (Cho et al, 2010). L7Ae belongs to the PF01248 family of proteins that specifically recognizes and binds Kink-turn structural motif. Interestingly this protein is also part of other macromolecular machines like box C/D and H/ACA guide ncRNAs that coordinate some aspect of translation.

4.2. RNase Z

The universally conserved RNase Z with representatives in all three domains of life as well as in mitochondria and chloroplasts, is the major enzyme participating in the 3' end processing of most tRNAs and potentially in rRNA maturation pathways. RNase Z members belong to the so-called Elac family which is part of the metal dependent β -lactamase superfamily, a group

of metalloproteins which perform a variety of functions (Aravind, 1999) (**Table 1**). Extensive sequence analysis of the members of the metallo- β -lactamase superfamily allowed the identification of several sub-families among which are, RNase Z and β -CASP, two nucleic acid hydrolases (described hereafter in section 5.2) (Aravind, 1999; Daiyasu et al, 2001).

The 3' end maturation of tRNA is far more complex than its 5' end processing. Removal of 3' extensions, also referred as 3' trailers, from pre-tRNA requires an endonucleolytic cleavage that can be followed by an exonucleolytic trimming depending on the presence or absence of the terminal CCA sequence. Briefly RNase Z enzymes contain a metallo- β -lactamase domain followed by a specific pre-tRNA binding domain called exosite (**Figure 2A**). In Archaea these enzymes come only in short-length versions (280 to 360 amino acids) that act as a dimer. RNase Z carries an endoribonucleolytic activity that generates the mature 3'-end of tRNA molecules by removal of the 3'-trailer elements of precursor tRNAs (**Figure 1**). This cleavage, which comes immediately after the discriminator nucleotide (first unpaired nucleotide located 3' of the acceptor stem) and leaves a 3' hydroxyl group on the tRNA 3' end and a 5' phosphate on the trailer element requires two Zn^{2+} ions (for review see (Dominski, 2007; Redko et al, 2007; Spath et al, 2007; Vogel et al, 2005)).

Recombinant RNase Z from archaeal species *Haloferax volcanii*, *Methanococcus janaschii*, and *Pyrococcus furiosus* has been shown to efficiently and accurately cleave tRNA precursors *in vitro* (Schierling et al, 2002; Schiffer et al, 2002; Spath et al, 2008). Archaeal RNase Z enzymes require the entire structure of pre-tRNA substrates for optimal activity (Schierling et al, 2002). However, the substrate specificity of haloarchaeal and thermococcal RNase Z extend to tRNA-like structures which are present 5' of the 5S rRNA in ribosomal operons (Holzle et al, 2008). Therefore RNase Z could also be a major actor of archaeal ribosomal RNA maturation pathway. By all these means it is not surprising that in *H. volcanii*, RNase Z gene is an essential gene (Holzle et al, 2008).

4.3. End A

Archaeal tRNA-splicing endoRNases (EndA) catalyze the endonucleolytic cleavage of pre-tRNAs at the 5' and 3' splice sites to release the intron and produces two half RNA molecules bearing 5' hydroxyl and 2', 3'-cyclic phosphate termini. In yeast, plants, and mammals, EndA is formed by the four non-identical subunits Sen15p, Sen34p, Sen2p, and Sen54p (for review see in (Lopes et al, 2015)). The genes encoding the archaeal α , α' , β and ϵ subunits of EndA which are homologous to their eukaryotic counterparts are ubiquitous in Archaea (**Figure 3**).

Classically archaeal EndA enzymes were classified into three types according to subunit composition: a homotetramer (α_4) in some Euryarchaea like- *M. jannaschii*, a homodimer (α_2) in other Euryarchaea-like *H. volcanii*, and a heterotetramer ($\alpha_2\beta_2$) in the Crenarchaea and the Nanoarchaea (**Figure 2A**) (for review see (Lopes et al, 2015)). Recently, a fourth type from an uncultivated archaeon *Candidatus Micrarchaeum acidiphilum*, deeply branched within Euryarchaea and referred to as ARMAN-2, which is composed of ϵ_2 homodimer and has broad substrate specificity like the crenarchaeal and nanoarchaeal $\alpha_2\beta_2$ type (Hirata et al, 2012) (**Figure 2A**).

In all configurations, the overall folding of EndA that resembles the homotetramer one permits an efficient cleavage of the bulge-helix-bulge (BHB) motif, where two three-nucleotide bulges are separated by a four-nucleotide helix. Archaeal introns are found in different positions in the tRNA in addition to the canonical position found in Eukarya (Marck & Grosjean, 2003). As long as the BHB motif is formed, the archaeal endoRNase will recognize and cleave the intron.

Finally, it was shown that various Archaea use EndA to excise pre-rRNA spacers containing BHB motifs, which remove pre-16S and pre-23S rRNA from primary transcripts (Kjems & Garrett, 1988; Lykke-Andersen & Garrett, 1997) (discussed in chapter 6-Ferreira-Cerca).

4.4. Nob1

In Archaea, very few are known on ribosomal assembly factors which transiently bind to and act on the nascent ribosome in a temporally and spatially well-defined and highly regulated manner. However, several archaeal homologs of eukaryotic ribosome assembly factors served as valuable models for structural, functional, and biophysical studies (Hellmich et al, 2013; Veith et al, 2012). The recent identification of eukaryotic endoRNase Nob1 homologs in archaeal genomes suggests that the 3' terminal maturation of the archaeal pre-rRNA of the small ribosomal subunit is also processed at the D-cleavage site as in eukaryotes (discussed in chapter 6-Ferreira-Cerca). The Nob1-like proteins have been identified in the all the major branches of Archaea with the exception of Korarchaea (Veith et al, 2012) (**Figure 3**). *In vitro*, the full-length Nob1 protein from the archaeon *Pyrococcus horikoshii* (PhNob1), efficiently cleaves RNA-substrates containing the D-site of the pre-16S RNA in a manganese-dependent manner. The structure of PhNob1 solved by nuclear magnetic resonance spectroscopy revealed a PIN (PiIT N-terminus) domain common in many nucleases and a zinc ribbon domain, which are structurally connected by a flexible linker (Veith et al, 2012)(**Figure 2A**).

4.5. Cas6

A hallmark of defense mechanisms based on clustered regularly interspaced short palindromic repeats (CRISPR) and associated sequences (Cas) are the crRNAs that guide these complexes in the destruction of invading DNA or RNA (discussed in chapter 11-Marchefelder). The repeat-spacer array of CRISPR gives rise to a long precursor named pre-crRNA that is, in Type I and III systems, processed by an endoRNase into crRNA intermediates (70–80 nt). Mature crRNAs are integrated into large ribonucleoprotein complexes with their cognate Cas proteins to guide them to the invading foreign RNA or DNA sequences (for review (Sorek et al, 2013; van der Oost et al, 2014; Westra et al, 2012; Wiedenheft et al, 2012)).

Together with the host bacterial RNase III endoRNases, the Cas6 members are responsible for producing crRNAs from the CRISPR precursor (**Figure 1**).

Cas6 from *P. furiosus* (PfCas6) was the first metal-independent endoRNase to be characterized as involved in CRISPR RNA processing within the repeat sequences (Carte et al, 2008). This enzyme was shown to interact with a specific sequence motif in the 5' region of the CRISPR repeat element and to cleave at a defined site within the 3' region of the repeat (Carte et al, 2008). Remarkably *cas6* is one of the most widely distributed CRISPR-associated genes. The Cas6 protein superfamily that has representative in Archaea and Bacteria are known as primary endoRNases in CRISPR system subtypes I-A, I-B, I-E, I-F and Type III (for review see (Hochstrasser & Doudna, 2015)). The amino acid sequence of archaeal Cas6 sequence show limited conservation with only two common motifs: the ferredoxin fold and a glycine rich motif, that are also found in other RNA-binding proteins (Li, 2015) (**Table 1, Figure 2A**). It is only poorly understood how different Cas6 endoRNases, present in organisms with multiple CRISPR systems, differentiate between their targets (for review see (Hochstrasser & Doudna, 2015)).

4.6. RNase HII

RNase H is a ubiquitous enzyme found in all domains of life that cleaves the RNA strand embedded in RNA/DNA hybrid. In archaeal cells, RNA-primed replication intermediates, similar to those formed during eukaryotic DNA replication, have been observed in Euryarchaea and Crenarchaea, arguing for either RNA or RNA–DNA priming at the archaeal replication fork. RNase H enzymes vary greatly in domain structures and substrate specificities (for review see (Tadokoro & Kanaya, 2009)). RNase H cleaves the P-O3' bond of the substrates with a two-metal-ion catalysis mechanism, in which two metal ions are directly

involved in the catalytic function. RNases H from Bacteria and Archaea have been classified into type 1 (RNase HI) and type 2 RNase H (RNase HII) based on differences in their amino acid sequences (Kochiwa et al, 2007). RNase H in its two types is a key component for the growth and survival of all organisms by playing a crucial role in DNA replication by removing the RNA primer of Okazaki fragments and in DNA repair by removing the single ribonucleotides incorporated in the DNA (for review see (Tadokoro & Kanaya, 2009). Most of the archaeal genomes only contain RNase HII genes (**Figure 2A & Figure 3**). The crystal structure of RNase HII from *Archaeoglobus fulgidus* in complex with PCNA revealed three binding modes as the enzyme rotates around a flexible hinge while anchored to PCNA by its PIP-box motif (Bubeck et al, 2011). PCNA binding was shown to promote RNase HII activity in a hinge-dependent manner by enhancing both cleavage of misincorporated ribonucleotides in DNA duplexes, and the comprehensive hydrolysis of RNA primers formed during Okazaki fragment maturation. These findings provide insights into how RNase HII activity is directed during genome replication and repair (Bubeck et al, 2011). Interestingly, *P. furiosus* Pf-RNase HII has been shown to also be able to digest RNA-RNA duplexes in the presence of Mn^{2+} ions (Kitamura et al, 2010). Remarkably, the three-dimensional structure of Pf-RNase H is similar to that of the PIWI domain of the Pf-Ago, an argonaute protein known to act in RNA-induced silencing complex in eukaryotes, although the two enzymes share almost no similarity in their amino acid sequences. In contrast to eukaryotic Ago proteins, archaeal Ago proteins show greater affinity for RNA-DNA hybrids than for RNA-RNA hybrids (Ma et al, 2005).

As an exception, the Halobacterium sp. NRC-1, *Sulfolobus tokodaii* and *Pyrobaculum aerophilum* genomes contains additional RNase HI encoding genes. The Halobacterium sp. NRC-1 RNase HI, whose folding is induced by the binding of divalent metal ions can cleave an RNA-DNA junction (Ohtani et al, 2004b; Tannous & Kanaya, 2014). Interestingly, *S. tokodaii* RNase HI enables degradation from RNA/RNA duplex (Ohtani et al, 2004a).

5-Ribonucleases with uncharacterized cellular RNA substrates

5.1. 3'-5' ExoRNases: The archaeal Exosome & RNase R

3' to 5' exoribonucleolytic activities have been reported in the three domains of life. These activities are performed by PNPase, RNase II and RNase R in Bacteria and by the exosome complex in Eukarya. In Archaea, only two archaeal 3' to 5' exoribonucleolytic activities have been described so far.

5.1.1. Archaeal Exosome

Prediction of the existence of a eukaryotic-like archaeal exosome was initially based on the identification of genes encoding orthologs of eukaryotic exosome subunits, the ribosomal-RNA-processing proteins Rrp4, Rrp41 and Rrp42. These components are organized in conserved super operon that were identified in most Archaea with the exception of the Halophiles and some Methanococcales (Evguenieva-Hackenberg et al, 2014; Evguenieva-Hackenberg & Klug, 2009; Koonin et al, 2001) (discussed in chapter 11-Evguenieva-Hackenberg).

The first experimental evidence for the existence of an exosome-like protein complex in Archaea was obtained by the purification of a 250-kDa protein complex that co-immunoprecipitate with Rrp41 from *Sulfolobus solfataricus* cellular extract. Half of this complex was also shown to co-sediment with ribosomal subunits. Orthologues to Rrp4, Rrp41, Rrp42 and Csl4, as well as a component annotated as a DnaG homologue, were found to tightly associate in a complex. A minimal core of the *S. solfataricus* and *M. thermautotrophicus* exosome was shown to consist of at least six phosphate-dependent ribonuclease PH homologues, along with Rrp4- and Csl4-like subunits (Evguenieva-Hackenberg et al, 2003; Farhoud et al, 2005) (**Figure 2B**).

Subsequently, crystal structures of the *S. solfataricus* and *A. fulgidus* exosome core revealed a hexameric ring-like arrangement of three Rrp41-Rrp42 heterodimers in which both subunits adopt the RNase PH fold common to phosphorolytic exoRNases. Structure-guided mutagenesis revealed that the activity of the complex resides within the active sites of the Rrp41 subunits, whereas the Rrp42 subunits are inactive but contribute to the structuring of the three Rrp41 active sites (discussed in chapter 11-Evguenieva-Hackenberg).

The archaeal exosome subunits share high sequence similarity with their eukaryotic counterparts, even if none of the eukaryotic subunits are active, and assemble in a structure highly similar to the bacterial trimeric PNPase ring-like structure (Dziembowski et al, 2007; Liu et al, 2006) These features support a common basis for a 3' to 5' RNA-degrading machineries in all three domains of life (Chlebowski et al, 2013). Nevertheless a specific distinctive of the archaeal exosome, experimentally shown for the exosome of *Sulfolobus solfataricus*, *Pyrococcus abyssi* and *Methanopyrus kandleri*, is the ability to work as an RNA-

tailing complex with a heteropolymeric polyadenylation activity. Consistently heteropolymeric polyadenylation was only observed in archaeal species containing an exosome complex but not in Halophiles and the methanogens which do not encode Rrp41–Rrp42 homologues (Portnoy et al, 2005; Portnoy & Schuster, 2006; Ramos et al, 2006) (**Figure 3**).

5.1.2. RNase R

Only halophiles and some methanogens, devoid of exosome, possess in their genome an *rnr* gene encoding an RNase R-like enzyme (Portnoy & Schuster, 2006) (**Figure 3**). Shortly, bacterial RNase R has been described as the first ribonuclease possessing a 3' to 5' hydrolytic exoRNase activity with an unusual ability to digest highly structured RNA without the aid of helicase activity (Vincent & Deutscher, 2009).

RNase R homologs were initially identified by amino acid sequence comparison search. Only one gene encoding RNase R-like was identified in *H. volcanii* with 34% sequence identity with *E. coli* RNase R (Portnoy & Schuster, 2006). In Halophiles and in methanogens with no exosome, archaeal RNase R seems to be the unique 3' to 5' exoRNase. Conversely to bacterial RNase R, archaeal RNase R are not able to degrade structured RNA by themselves as they are restricted to the RNB central core domain and lack the N- (CSD1&2) and C-terminal (S1) domains specific (Matos et al, 2012) (**Figure 2D**). *In vitro* enzymatic studies showed that recombinant HvrRNase R demonstrates an optimal exoribonucleolytic activity on tRNA substrates at low salt concentration either at 25°C and 37°C suggesting that HvrRNase R is not adapted to high salinity. However, the residual activity at high salt concentration is sufficient since *rnr* is known to be essential for *H. volcanii* viability (Matos et al, 2012). Overall the studies carried out with HvrRNase R suggest that archaeal RNase R plays a critical role in RNA decay in archaeal species devoid of an exosome complex (Matos et al, 2012; Portnoy & Schuster, 2006).

5.2. 5'-3' Exo and EndoRNases: Archaeal β -CASP ribonucleases

The β -CASP family members that are widespread in Bacteria and Eukarya, encompass RNA cleaving (CPSF73 and RNase J) and DNA repair (Artemis, SNM1 and PSO2) enzymes. In common with RNase Z belonging to the Elac family, members of the β -CASP family are characterized by a metallo- β -lactamase domain additionally with a specific β -CASP domain. Highly conserved aspartic acid, histidine and valine residues form a signature motif that plays key roles in the enzymatic function of the β -CASP enzymes by coordinating directly or

indirectly coordination with two metal ions (usually Zn^{2+}) (for review see (Dominski et al, 2013)). Notably β -CASP enzymes have been shown or suggested to share dual endonucleolytic and 5'-3' exonucleolytic activities catalyzed by a unique catalytic site (for review see (Newman et al, 2011)). These fascinating enzymes are key players in 3' end processing in Eukarya and in RNA decay and ribosomal RNA maturation in Bacteria.

Recently β -CASP ribonucleases have also been identified in Archaea. As their bacterial and eukaryal counterparts, some also exhibit a dual (5'-3' exo- and endo-) ribonucleolytic activity (for review see (Clouet-d'Orval et al, 2015; Dominski et al, 2013)). The recent recognition of β -CASP ribonucleases as major players in Archaea is an important contribution towards identifying RNA-degrading mechanisms in the third domain of life (Clouet-d'Orval et al, 2015) (**Table 1, Figure 2C & 2D**). In-depth phylogenomic analyses and structural studies revealed three major β -CASP orthologous groups in Archaea, aCPSF1 and aCPSF2 which are closely related to eukaryal CPSF73 (Cleavage Polyadenylation Specific Factor) termination factor and, aRNase J which is orthologous to bacterial RNase J (Phung et al, 2013) (for review see (Dominski et al, 2013)). The strictly conservation of archaeal β -CASP enzymes throughout archaeal phylogeny strongly suggests their essential roles in maturation and/or degradation of RNA (Clouet-d'Orval et al, 2015) (**Figure 3**). Features of archaeal β -CASP ribonucleases are described hereafter by their enzymatic activities.

5.2.1. 5'-3' Exo β -CASP ribonucleases: aRNase J & aCPSF2

-aRNase J

Phylogenetic analysis showed that the members of the aRNase J group which are widespread in Euryarchaea (**Figure 3**) have been inherited vertically, suggesting an ancient origin of the RNase J group of enzymes predating the separation of the Bacteria and the Archaea (Clouet-d'Orval et al, 2010). In contrast to bacterial RNase J which harbor an additional N- or C-terminal domain, archaeal aRNase J are restricted to the β -lactamase β -CASP core domain (**Figure 2C**) (Clouet-d'Orval et al, 2010; Dominski et al, 2013). However, euryarchaeal aRNase J contains two small highly conserved domains, denoted Loop1 and Loop2, that are absent in bacterial RNase J (Clouet-d'Orval et al, 2010).

The exoribonucleolytic activity with a 5' to 3' directionality of aRNase J was first identified for recombinant proteins from Thermococcales (Clouet-d'Orval et al, 2010). Subsequent work with aRNase J of *M. jannaschii* also identified a similar 5' to 3' trimming activity (Levy et al, 2011). The euryarchaeal aRNase J which are highly processive enzymes do not have shown any detectable endoribonucleolytic activity under the *in vitro* conditions tested. However, they

exhibit reduced *in vitro* activity on 5' triphosphate-end RNA substrates, indicating that aRNase J exonucleases are 5' end-dependent exoRNases and suggesting that a subset of cellular RNAs with 5' triphosphate ends could be refractory to degradation by aRNase J *in vivo* (Clouet-d'Orval et al, 2010; Levy et al, 2011). Mutational analysis showed that residues of the highly conserved motifs implicated in the coordination of the two zinc ions and the specific Loop1 and Loop2 are critical for *in vitro* ribonucleolytic activity (Clouet-d'Orval et al, 2010). In addition, it is interesting to note that except for the difference in temperature optima, the characteristics of the exoRNase activity of euryarchaeal aRNase J are very similar to the exoRNase activity of RNase J1 from Bacteria (Li de la Sierra-Gallay et al, 2008; Mathy et al, 2007) strongly suggesting that euryarchaeal aRNase J might have similar roles in rRNA processing and mRNA decay. However the capacity of some euryarchaeal aRNase J to degrade single-stranded DNA *in vitro* suggests versatile roles of aRNase J *in vivo*. At least two other exonucleases from the β -CASP family, hSNM1 and Artemis, are known to degrade DNA (for review see (Dominski et al, 2013)). Whether these observations are informative in terms of aRNase J physiological properties and substrate specificity and, bring knowledge on aRNase J singularities remain to be seen.

-aCPSF2

aCPSF2 members are strictly conserved in Crenarcheota and widely spread but more divergent in Euryarchaeota (Clouet-d'Orval et al, 2015; Phung et al, 2013) (**Figure 3**). Much is still not known on the activity and structure of members of aCPSF2 group which is phylogenetically closely related to aCPSF1 group (Dominski et al, 2013; Phung et al, 2013). Like aRNase J, aCPSF2 enzymes are commonly restricted to the β -CASP and metallo- β -lactamase core domains with no additional N- or C-terminal extensions (**Figure 2C**). One study reported '3' exonucleolytic activity *in vitro* activity for the crenarchaeon *S. solfataricus* (Sso) aCPSF2 that is modulated by the phosphorylation state of the 5' end of transcript as seen for aRNase J members (Hasenohrl et al, 2011). However, in contrast to the other archaeal β -CASP proteins (Phung et al, 2013), the activity of Sso-aCPSF2 was only observed in the presence of Mg^{2+} ions (Hasenohrl et al, 2011). Furthermore, the 5'-3' exoribonucleolytic activity of Sso-aCPSF2 has been shown to be blocked by the binding of the translation initiation factor a/eIF2 δ that protects the 5' end of mRNA from degradation *in vivo* and *in vitro* (Hasenohrl et al, 2008). Further understanding of the enzymatic and functional properties of aCPSF2, in particular the role of magnesium, must await structural studies.

A first clue towards understanding the biological relevance of aCPSF2 was obtained in *Sulfolobus acidocaldarius* (Martens et al, 2013). Comparison of transcriptomes from wild type and aCPSF2-deleted strains revealed differential transcript abundance for 560 genes. This global effect suggests that aCPSF2 plays a major role in 5'-3' directed mRNA decay. Since an aCPSF2-deleted strain exhibits no growth defect, the observed differences in transcripts abundance does not affect the overall cellular fitness. At this point, the processing pathways recruiting aCPSF2 remain to be identified.

To conclude, each archaeal genome encodes at least one β -CASP enzyme with a 5'-3' exoribonucleolytic activity which could be carried out either by an aRNase J and/or an aCPSF2 group member (Clouet-d'Orval et al, 2015) (**Figure 3**). This strongly suggests that 5'-3' processing pathways exist in Archaea and are somehow important for the fate of cellular RNA biotypes.

5.3. Dual 5'-3' Exo/Endo- β -CASP ribonuclease: aCPSF1

Phylogenomic studies searching for β -CASP proteins among archaeal genomes have identified an outstanding group, called aCPSF1, that is strictly conserved and orthologous to the eukaryal cleavage and polyadenylation specific factor group defined by human CPSF73 and yeast CPSF100 (Clouet-d'Orval et al, 2015; Phung et al, 2013) (**Figure 3**). Thermococcales aCPSF1 have been shown to harbor *in vitro* 5'-3' exo- and endoribonucleolytic activities with a preference for single-stranded CA dinucleotide as described for eukaryal CPSF73 factor within the machinery required for termination of RNAP II transcription (for review see (Clouet-d'Orval et al, 2015; Dominski et al, 2013)). In Archaea, the relevance of the cleavage specificity at single-stranded CA dinucleotide remains to be elucidated (Phung et al, 2013). In contrast, *M. jannaschii* aCPSF1 (initially named mjRNase J2 in (Levy et al, 2011)) was shown to only display an endoribonucleolytic activity with no detectable exoribonucleolytic activity. aCPSF1 is dimeric in solution with a dimer interface composed of 12 amino acids that is conserved among all aCPSF1 members. Interestingly, disruption of this interface impairs the exoribonucleolytic activity of *P. abyssi* aCPSF1 but not its endoribonucleolytic activity (Phung et al, 2013).

The aCPSF1 members are distinguished by their N-terminal domain composed of a tandem repeat of two type-II KH motifs predicted to bind nucleic acids that extends the β -CASP and metallo- β -lactamase core domains (**Figure 2D**). Crystal structures of dimeric aCPSF1 from *Pyrococcus horikoshii*, *Methanosarcina mazei* and *Methanothermobacter thermoautotrophicus* show a tripartite architecture consisting of the β -CASP and β -lactamase

domains folded apart from both KH motifs (Mir-Montazeri et al, 2011; Nishida et al, 2010; Silva et al, 2011). In view of their wide phylogenetic distribution, aCPSF1 dual ribonuclease likely performs essential functions in archaeal metabolism. Interestingly, a study identified *Haloferax volcanii* aCPSF1 as an interacting components of the proteasome, suggesting a connection between RNA processing and protein degradation (Chavarria et al, 2014).

6-Concluding remarks

Although we highlighted different ribonuclease families potentially involved in modulating RNA pools in Archaea cells, an integrated view of how these ribonucleases regulate RNA dynamics, levels and modification is still missing. The tight cooperation between different RNA processing systems needs also to be discovered. Furthermore ancillary RNA-modifying enzymes such as RNA helicases, poly(A) polymerases, pyrophosphohydrolases and RNA binding proteins such as Sm-like proteins, are undoubtedly important in mediating archaeal RNA processing and/or turnover. Some of these factors are beginning to be described. First, as mentioned above in section 5.1.1, the archaeal exosome can work as an RNA-tailing complex of polyadenylation. More recently the Sm-like proteins in *Sulfolobus*, SmAP1 and SmAP2, with affinity for different classes of ncRNAs and mRNAs have been shown to co-purify with proteins involved in RNA processing, RNA modification, translation and RNA turnover as well as with components of the exosome (Martens et al, 2017).

Besides, the many archaeal proteins of unknown function, for which a ribonucleolytic or related activity cannot be predicted from their sequence, makes us foresee that the discovery of archaeal ribonucleases will remain a dynamic area of research for the years to come.

In essence, deciphering RNA biology processes in Archaea will give us key paradigms for understanding fundamentally conserved processes including RNA maturation and decay which dictate post-transcriptional regulation of gene expression. In addition, examination of the distribution of the individual ribonuclease families among diverse organisms will provide a phylogenetic framework with interesting evolutionary implications. The identification of conserved pathways will retrace the history of RNA metabolism across the domains of life.

Acknowledgements

BCO and MB are supported by the Centre National de la Recherche Scientifique (CNRS) with additional funding from the Agence Nationale de la Recherche (ANR-16-CE12-0016-01) and from the Université de Toulouse (IDEX « Emergence » call 2014).

Table 1: Archaeal RNase families.

	FAMILY*	NAME	ROLE	BACT.*	EUK.*
3'>5' EXO	RNB	RNase R	<i>n.d.</i>	✓	✓
	PDX	Rrp41/Rrp42	<i>n.d.</i>	✓	✓
ENDO/ 5'>3' EXO	β-lactamase	aRNase J	<i>n.d.</i>	✓	
		β-CASP	aCPSF2	<i>n.d.</i>	✓
			aCPSF1	<i>n.d.</i>	
ENDO	Elac	RNase Z	tRNA processing	✓	✓
	RNase P	RNase P	tRNA processing	✓	✓
	SEN	End A	tRNA & rRNA processing		✓
	PIN	Nob1	rRNA processing	✓	✓
	Ferredoxin-like	Cas6	CRISPR RNA processing	✓	
	RNase H	HI/HII		✓	✓

*Occurrence of ribonuclease families in Eukarya (Euk.) and Bacteria (Bact.) as reported in (Aravind & Koonin, 2001; Arraiano et al, 2013; Condon & Putzer, 2002)

Legend to Figures

Figure 1: Archaeal RNA biotypes. The unique archaeal RNAP transcribes all the different biotypes of RNAs. Only few tRNA, rRNA and CRISPR RNA maturation steps have been identified of up-to-date in Archaea. The few known RNA-modifying reactions (Cleavage/Splicing/Nucleotide modifications) are in blue and corresponding enzymes in red. Question marks indicate that some RNases remained to be identified. Note that many nucleotide modifications such as 2'-O-methylations and pseudouridylations are guided by C/D and H/ACA sRNPs, respectively. The stars* indicate that the catalytic moiety of the RNP is an RNA molecule. The exosome, aCPSF1, aCPSF2 and aRNase J have not being yet attributed to any specific RNA processing pathways.

Figure 2: Domain architecture of archaeal RNases. Pfam annotation (<http://pfam.xfam.org/>) is indicated for each enzyme domain. A- EndoRNases with identified RNA targets. The two types of structure for RNA of RNase P (RPR) are shown. Archaeal RNase P RNA subunits are classified as type M (*Methanocaldococcus jannaschii*) and type A (*P. furiosus*) (Pannucci et al, 1999)(The RNase P database, <http://www.mbio.ncsu.edu/RNaseP/seqs&structures.html>). B- 3'-5' ExoRNases with uncharacterized RNA target. C- 5'-3' ExoRNases with uncharacterized RNA target. D- RNase with dual activity.

Figure 3: Phylogenetic distribution of reported RNases. Archaeal phylogenetic was constructed as in (Clouet-d'Orval et al, 2015). Note that the fast evolving parasitic DPANN clan hood (comprising Nanoarchaea & its relatives) is not included in this phylogeny. Grey rectangle indicates that at least one member of the clade possess in its genome a gene encoding the RNases mentioned in red, blue and green upon their activity.

References

- Aravind L (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins. *In silico biology* **1**: 69-91
- Aravind l, Koonin EV (2001) A natural classification of Ribonucleases. *Methods in enzymology* **341**: 1-28
- Arkhipova V, Stolboushkina E, Kravchenko O, Kljashtorny V, Gabdulkhakov A, Garber M, Nikonov S, Martens B, Blasi U, 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**: 3086-3095
- Arraiano CM, Mauxion F, Viegas SC, Matos RG, Seraphin B (2013) Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA. *Biochimica et biophysica acta* **1829**: 491-513

- Bandyra KJ, Bouvier M, Carpousis AJ, Luisi BF (2013) The social fabric of the RNA degradosome. *Biochimica et biophysica acta* **1829**: 514-522
- Barrangou R, Coute-Monvoisin AC, Stahl B, Chavichvily I, Damange F, Romero DA, Boyaval P, Fremaux C, Horvath P (2013) Genomic impact of CRISPR immunization against bacteriophages. *Biochemical Society transactions* **41**: 1383-1391
- Bini E, Dikshit V, Dirksen K, Drozda M, Blum P (2002) Stability of mRNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *RNA* **8**: 1129-1136
- Brochier-Armanet C, Forterre P, Gribaldo S (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Current opinion in microbiology*
- Brooks SA (2010) Functional interactions between mRNA turnover and surveillance and the ubiquitin proteasome system. *Wiley interdisciplinary reviews RNA* **1**: 240-252
- Bubeck D, Reijns MA, Graham SC, Astell KR, Jones EY, Jackson AP (2011) PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic acids research* **39**: 3652-3666
- Carpousis AJ (2007) The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annual review of microbiology* **61**: 71-87
- Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* **22**: 3489-3496
- Charpentier B, Muller S, Branlant C (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic acids research* **33**: 3133-3144
- Chavarria NE, Hwang S, Cao S, Fu X, Holman M, Elbanna D, Rodriguez S, Arrington D, Englert M, Uthandi S, Soll D, Maupin-Furlow JA (2014) Archaeal Tbc1/Ncs6 homolog required for wobble uridine tRNA thiolation is associated with ubiquitin-proteasome, translation, and RNA processing system homologs. *PLoS one* **9**: e99104
- Chlebowski A, Lubas M, Jensen TH, Dziembowski A (2013) RNA decay machines: the exosome. *Biochimica et biophysica acta* **1829**: 552-560
- Cho HD, Verlinde CL, Weiner AM (2005) Archaeal CCA-adding enzymes: central role of a highly conserved beta-turn motif in RNA polymerization without translocation. *The Journal of biological chemistry* **280**: 9555-9566
- Cho IM, Lai LB, Susanti D, Mukhopadhyay B, 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**: 14573-14578
- Clouet-d'Orval B, Phung DK, Langendijk-Genevaux PS, Quentin Y (2015) Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of beta-CASP ribonucleases. *Biochimie*
- Clouet-d'Orval B, Rinaldi D, Quentin Y, Carpousis AJ (2010) Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *The Journal of biological chemistry* **285**: 17574-17583
- Condon C, Putzer H (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic acids research* **30**: 5339-5346
- Daiyasu H, Osaka K, Ishino Y, Toh H (2001) Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS letters* **503**: 1-6

- Dar D, Prasse D, Schmitz RA, Sorek R (2016) Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nature microbiology* **1**: 16143
- Dennis PP, Omer A (2005) Small non-coding RNAs in Archaea. *Current opinion in microbiology* **8**: 685-694
- Dennis PP, Omer A, Lowe T (2001) A guided tour: small RNA function in Archaea. *Molecular microbiology* **40**: 509-519
- Deutscher MP (2009) Maturation and degradation of ribosomal RNA in bacteria. *Progress in molecular biology and translational science* **85**: 369-391
- Dominski Z (2007) Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* **42**: 67-93
- Dominski Z, Carpousis AJ, 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**: 532-551
- Dziembowski A, Lorentzen E, Conti E, Seraphin B (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature structural & molecular biology* **14**: 15-22
- Evguenieva-Hackenberg E, Hou L, Glaeser S, Klug G (2014) Structure and function of the archaeal exosome. *Wiley interdisciplinary reviews RNA* **5**: 623-635
- Evguenieva-Hackenberg E, Klug G (2009) RNA degradation in Archaea and Gram-negative bacteria different from Escherichia coli. *Progress in molecular biology and translational science* **85**: 275-317
- Evguenieva-Hackenberg E, Walter P, Hochleitner E, Lottspeich F, Klug G (2003) An exosome-like complex in Sulfolobus solfataricus. *EMBO reports* **4**: 889-893
- Farhoud MH, Wessels HJ, Steenbakkens PJ, Mattijssen S, Wevers RA, van Engelen BG, Jetten MS, Smeitink JA, van den Heuvel LP, Keltjens JT (2005) Protein complexes in the archaeon Methanothermobacter thermoautotrophicus analyzed by blue native/SDS-PAGE and mass spectrometry. *Molecular & cellular proteomics : MCP* **4**: 1653-1663
- Fernandez-Pevida A, Kressler D, de la Cruz J (2015) Processing of preribosomal RNA in Saccharomyces cerevisiae. *Wiley interdisciplinary reviews RNA* **6**: 191-209
- Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews Genetics* **9**: 102-114
- Frank DN, Pace NR (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annual review of biochemistry* **67**: 153-180
- Franzetti B, Sohlberg B, Zaccari G, von Gabain A (1997) Biochemical and serological evidence for an RNase E-like activity in halophilic Archaea. *Journal of bacteriology* **179**: 1180-1185
- Garrett RA (2016) Archaeal physiology: The secrets of termination. *Nature microbiology* **1**: 16159
- Gopalan V (2007) Uniformity amid diversity in RNase P. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 2031-2032
- Hasenohrl D, Konrat R, Blasi 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

- Hasenohrl D, Lombo T, Kaberdin V, Londei P, Blasi 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**: 2146-2150
- Hellmich UA, Weis BL, Lioutikov A, Wurm JP, Kaiser M, Christ NA, Hantke K, Kotter P, Entian KD, Schleiff E, Wohnert J (2013) Essential ribosome assembly factor Fap7 regulates a hierarchy of RNA-protein interactions during small ribosomal subunit biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15253-15258
- Hennigan AN, Reeve JN (1994) mRNAs in the methanogenic archaeon *Methanococcus vannielii*: numbers, half-lives and processing. *Molecular microbiology* **11**: 655-670
- Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley interdisciplinary reviews RNA* **6**: 225-242
- Hirata A, Fujishima K, Yamagami R, Kawamura T, Banfield JF, Kanai A, 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**: 10554-10566
- Hochstrasser ML, Doudna JA (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in biochemical sciences* **40**: 58-66
- Holzle A, Fischer S, Heyer R, Schutz S, Zacharias M, Walther P, Allers T, Marchfelder A (2008) Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* **14**: 928-937
- Huang HY, Hopper AK (2016) Multiple Layers of Stress-Induced Regulation in tRNA Biology. *Life (Basel)* **6**
- Hui MP, Foley PL, Belasco JG (2014) Messenger RNA Degradation in Bacterial Cells. *Annual review of genetics* **48**: 537-559
- Jager A, Samorski R, Pfeifer F, 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**: 5436-5443
- Jarrous N, Gopalan V (2010) Archaeal/eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic acids research* **38**: 7885-7894
- Kanai A, Oida H, Matsuura N, Doi H (2003) Expression cloning and characterization of a novel gene that encodes the RNA-binding protein FAU-1 from *Pyrococcus furiosus*. *The Biochemical journal* **372**: 253-261
- Kitamura S, Fujishima K, Sato A, Tsuchiya D, Tomita M, Kanai A (2010) Characterization of RNase HIII substrate recognition using RNase HIII-argonaute chimaeric enzymes from *Pyrococcus furiosus*. *The Biochemical journal* **426**: 337-344
- Kjems J, Garrett RA (1988) Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* **54**: 693-703
- Klemm BP, Wu N, Chen Y, Liu X, Kaitany KJ, Howard MJ, Fierke CA (2016) The Diversity of Ribonuclease P: Protein and RNA Catalysts with Analogous Biological Functions. *Biomolecules* **6**
- Klug G, Evguenieva-Hackenberg E, Omer A, Dennis PP, Marchfelder A (2007) RNA processing. In *Archaea: Molecular and Cellular Biology*, Cavicchioli R (ed), ASM Press, Washington D.C. edn, pp 158-174.

- Kochiwa H, Tomita M, Kanai A (2007) Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC evolutionary biology* **7**: 128
- Koonin EV, Wolf YI, 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 research* **11**: 240-252
- Lai LB, Vioque A, Kirsebom LA, Gopalan V (2010) Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS letters* **584**: 287-296
- Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlomagno T (2013) The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **502**: 519-523
- Levy S, Portnoy V, Admon J, Schuster G (2011) Distinct activities of several RNase J proteins in methanogenic archaea. *RNA biology* **8**: 1073-1083
- Li de la Sierra-Gallay I, Zig L, Jamalli A, Putzer H (2008) Structural insights into the dual activity of RNase J. *Nature structural & molecular biology* **15**: 206-212
- Li H (2015) Structural Principles of CRISPR RNA Processing. *Structure* **23**: 13-20
- Li H, Trotta CR, Abelson J (1998) Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* **280**: 279-284
- Li L, Ye K (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**: 302-307
- Liu Q, Greimann JC, Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**: 1223-1237
- Lopes RR, Kessler AC, Polycarpo C, Alfonzo JD (2015) Cutting, dicing, healing and sealing: the molecular surgery of tRNA. *Wiley interdisciplinary reviews RNA* **6**: 337-349
- Lykke-Andersen J, Garrett RA (1997) RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *The EMBO journal* **16**: 6290-6300
- Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**: 666-670
- Mackie GA (2013) RNase E: at the interface of bacterial RNA processing and decay. *Nature reviews Microbiology* **11**: 45-57
- Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews Microbiology* **13**: 722-736
- Marck C, 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* **8**: 1189-1232
- Marck C, Grosjean H (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* **9**: 1516-1531
- Martens B, Amman F, Manoharadas S, Zeichen L, Orell A, Albers SV, Hofacker I, Blasi U (2013) Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *PloS one* **8**: e76569

- Martens B, Hou L, Amman F, Wolfinger MT, Evguenieva-Hackenberg E, Blasi U (2017) The SmAP1/2 proteins of the crenarchaeon *Sulfolobus solfataricus* interact with the exosome and stimulate A-rich tailing of transcripts. *Nucleic acids research*
- Mathy N, Benard L, Pellegrini O, Daou R, Wen T, Condon C (2007) 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* **129**: 681-692
- Matos RG, Lopez-Vinas E, Gomez-Puertas P, Arraiano CM (2012) The only exoribonuclease present in *Haloferax volcanii* has an unique response to temperature changes. *Biochimica et biophysica acta* **1820**: 1543-1552
- Mir-Montazeri B, Ammelburg M, Forouzan D, Lupas AN, Hartmann MD (2011) Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *Journal of structural biology* **173**: 191-195
- Nagarajan VK, Jones CI, Newbury SF, Green PJ (2013) XRN 5'→3' exoribonucleases: structure, mechanisms and functions. *Biochimica et biophysica acta* **1829**: 590-603
- Newman JA, Hewitt L, Rodrigues C, Solovyova A, Harwood CR, Lewis RJ (2011) Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure* **19**: 1241-1251
- Nishida Y, Ishikawa H, Baba S, Nakagawa N, Kuramitsu S, Masui R (2010) Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*. *Proteins* **78**: 2395-2398
- Nolivos S, Carpousis AJ, Clouet-d'Orval B (2005) 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 research* **33**: 6507-6514
- Numata T, Ishimatsu I, Kakuta Y, Tanaka I, Kimura M (2004) Crystal structure of archaeal ribonuclease P protein Ph1771p from *Pyrococcus horikoshii* OT3: an archaeal homolog of eukaryotic ribonuclease P protein Rpp29. *RNA* **10**: 1423-1432
- Ohtani N, Yanagawa H, Tomita M, Itaya M (2004a) Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, *Sulfolobus tokodaii* 7. *Nucleic acids research* **32**: 5809-5819
- Ohtani N, Yanagawa H, Tomita M, Itaya M (2004b) Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA-DNA junction. *The Biochemical journal* **381**: 795-802
- Pannucci JA, Haas ES, Hall TA, Harris JK, Brown JW (1999) RNase P RNAs from some Archaea are catalytically active. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 7803-7808
- Peeters E, Driessen RP, Werner F, Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nature reviews Microbiology* **13**: 333-341
- Peeters E, Peixeiro N, Sezonov G (2013) Cis-regulatory logic in archaeal transcription. *Biochemical Society transactions* **41**: 326-331
- Phung DK, Rinaldi D, Langendijk-Genevaux PS, Quentin Y, Carpousis AJ, 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**: 1091-1103
- Portnoy V, Evguenieva-Hackenberg E, Klein F, Walter P, Lorentzen E, Klug G, Schuster G (2005) RNA polyadenylation in Archaea: not observed in *Haloferax* while the exosome polynucleotidylates RNA in *Sulfolobus*. *EMBO reports* **6**: 1188-1193

- Portnoy V, Schuster G (2006) RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic acids research* **34**: 5923-5931
- Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. *Genes & development* **25**: 1770-1782
- Ramos CR, Oliveira CL, Torriani IL, Oliveira CC (2006) The Pyrococcus exosome complex: structural and functional characterization. *The Journal of biological chemistry* **281**: 6751-6759
- Randau L, Schroder I, Soll D (2008) Life without RNase P. *Nature* **453**: 120-123
- Redko Y, Li de Lasierra-Gallay I, Condon C (2007) When all's zed and done: the structure and function of RNase Z in prokaryotes. *Nature reviews Microbiology* **5**: 278-286
- Samanta MP, Lai SM, Daniels CJ, Gopalan V (2016) Sequence Analysis and Comparative Study of the Protein Subunits of Archaeal RNase P. *Biomolecules* **6**
- Santangelo TJ, Cubonova L, Skinner KM, Reeve JN (2009) Archaeal intrinsic transcription termination in vivo. *Journal of bacteriology* **191**: 7102-7108
- Schierling K, Rosch S, Rupprecht R, Schiffer S, Marchfelder A (2002) tRNA 3' end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *Journal of molecular biology* **316**: 895-902
- Schiffer S, Rosch S, Marchfelder A (2002) Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *The EMBO journal* **21**: 2769-2777
- Silva AP, Chechik M, Byrne RT, Waterman DG, Ng CL, Dodson EJ, Koonin EV, Antson AA, Smits C (2011) Structure and activity of a novel archaeal beta-CASP protein with N-terminal KH domains. *Structure* **19**: 622-632
- Sinapah S, Wu S, Chen Y, Pettersson BM, Gopalan V, Kirsebom LA (2011) Cleavage of model substrates by archaeal RNase P: role of protein cofactors in cleavage-site selection. *Nucleic acids research* **39**: 1105-1116
- Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry* **82**: 237-266
- Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173-179
- Spath B, Canino G, Marchfelder A (2007) tRNase Z: the end is not in sight. *Cellular and molecular life sciences : CMLS* **64**: 2404-2412
- Spath B, Schubert S, Lieberoth A, Settele F, Schutz S, Fischer S, Marchfelder A (2008) Two archaeal tRNase Z enzymes: similar but different. *Archives of microbiology* **190**: 301-308
- Stoecklin G, Muhlemann O (2013) RNA decay mechanisms: specificity through diversity. *Biochimica et biophysica acta* **1829**: 487-490
- Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *The FEBS journal* **276**: 1482-1493
- Tang TH, Rozhdestvensky TS, d'Orval BC, Bortolin ML, Huber H, Charpentier B, Branlant C, Bachelier JP, Brosius J, Huttenhofer A (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic acids research* **30**: 921-930

- Tannous E, Kanaya S (2014) Divalent metal ion-induced folding mechanism of RNase H1 from extreme halophilic archaeon Halobacterium sp. NRC-1. *PloS one* **9**: e109016
- Thompson LD, Daniels CJ (1988) A tRNA(Trp) intron endonuclease from Halobacterium volcanii. Unique substrate recognition properties. *The Journal of biological chemistry* **263**: 17951-17959
- van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature reviews Microbiology* **12**: 479-492
- Veith T, Martin R, Wurm JP, Weis BL, Duchardt-Ferner E, Saffertal C, Hennig R, Mirus O, Bohnsack MT, Wohnert J, Schleiff E (2012) Structural and functional analysis of the archaeal endonuclease Nob1. *Nucleic acids research* **40**: 3259-3274
- Vincent HA, Deutscher MP (2009) The roles of individual domains of RNase R in substrate binding and exoribonuclease activity. The nuclease domain is sufficient for digestion of structured RNA. *The Journal of biological chemistry* **284**: 486-494
- Vogel A, Schilling O, Spath B, Marchfelder A (2005) The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biological chemistry* **386**: 1253-1264
- Weiner AM (2004) tRNA maturation: RNA polymerization without a nucleic acid template. *Current biology : CB* **14**: R883-885
- Werner F (2007) Structure and function of archaeal RNA polymerases. *Molecular microbiology* **65**: 1395-1404
- Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nature reviews Microbiology* **9**: 85-98
- Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annual review of genetics* **46**: 311-339
- Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338
- Yip WS, Shigematsu H, Taylor DW, Baserga SJ (2016) Box C/D sRNA stem ends act as stabilizing anchors for box C/D di-sRNPs. *Nucleic acids research* **44**: 8976-8989
- Zago MA, Dennis PP, Omer AD (2005) The expanding world of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. *Molecular microbiology* **55**: 1812-1828

Figure 1

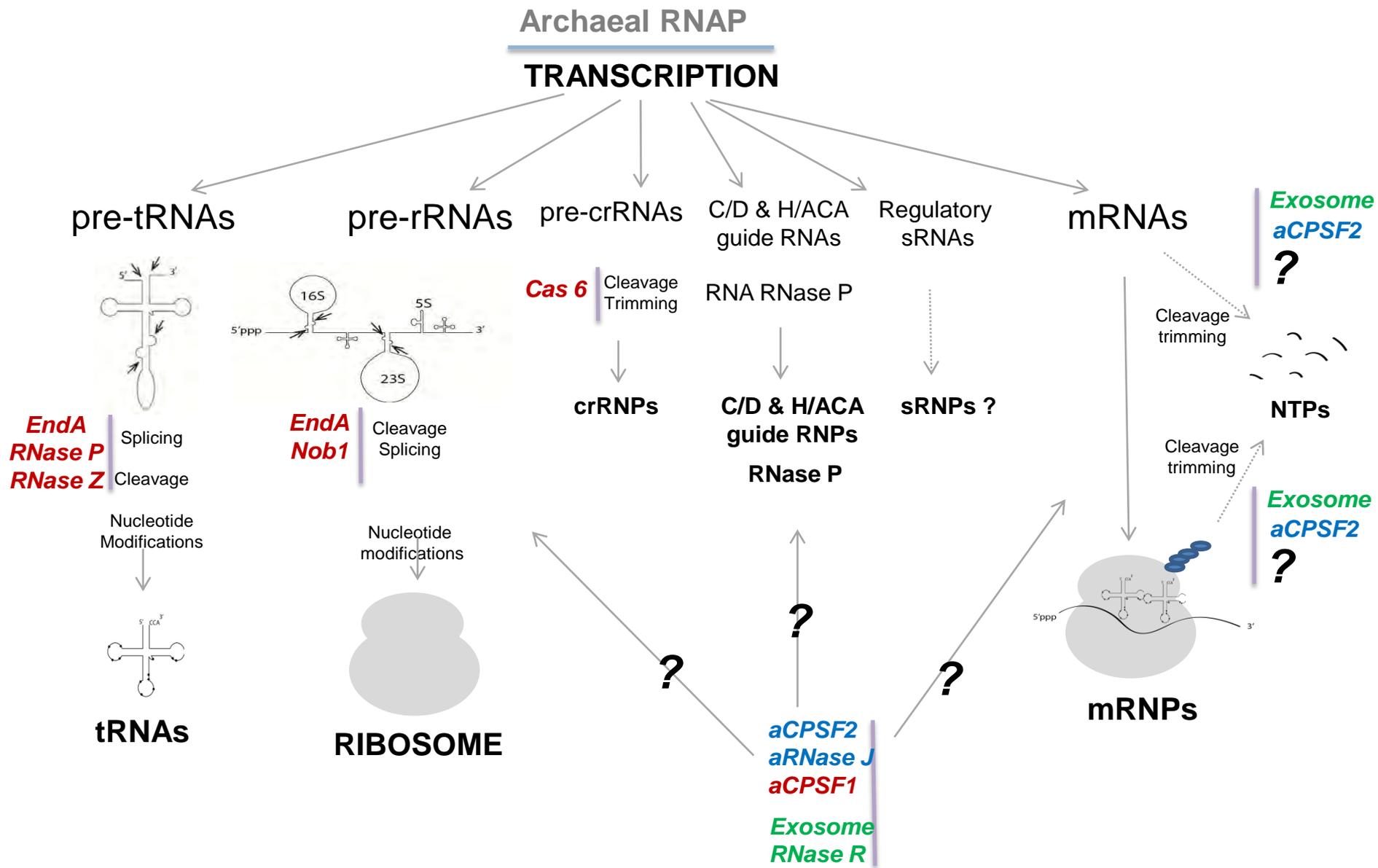
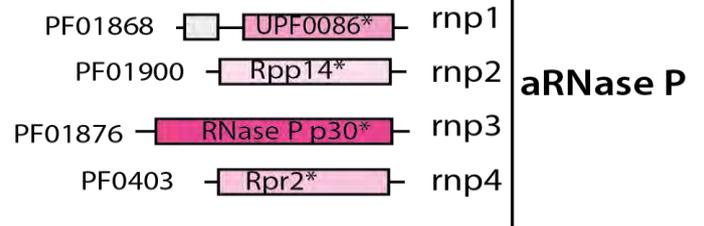
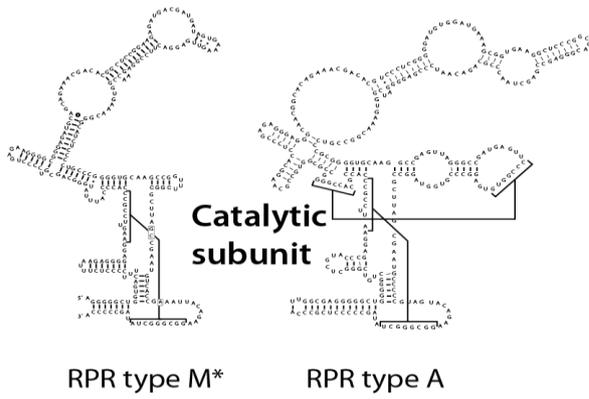
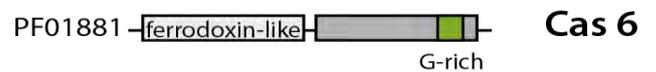
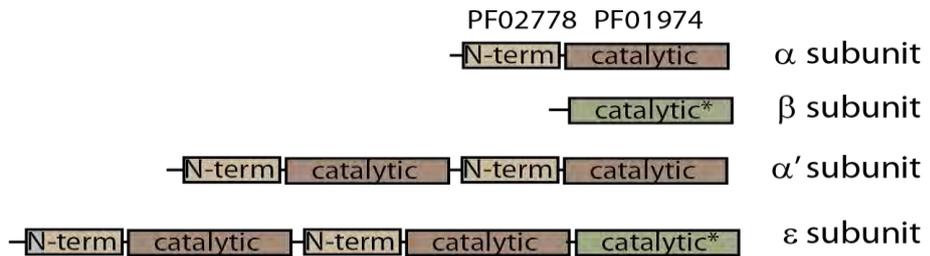


Figure 2

A



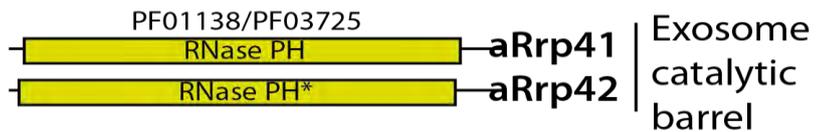
ENDO



B

3'>5'

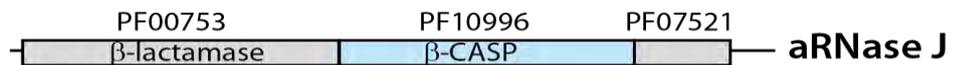
EXO



C

5'>3'

EXO



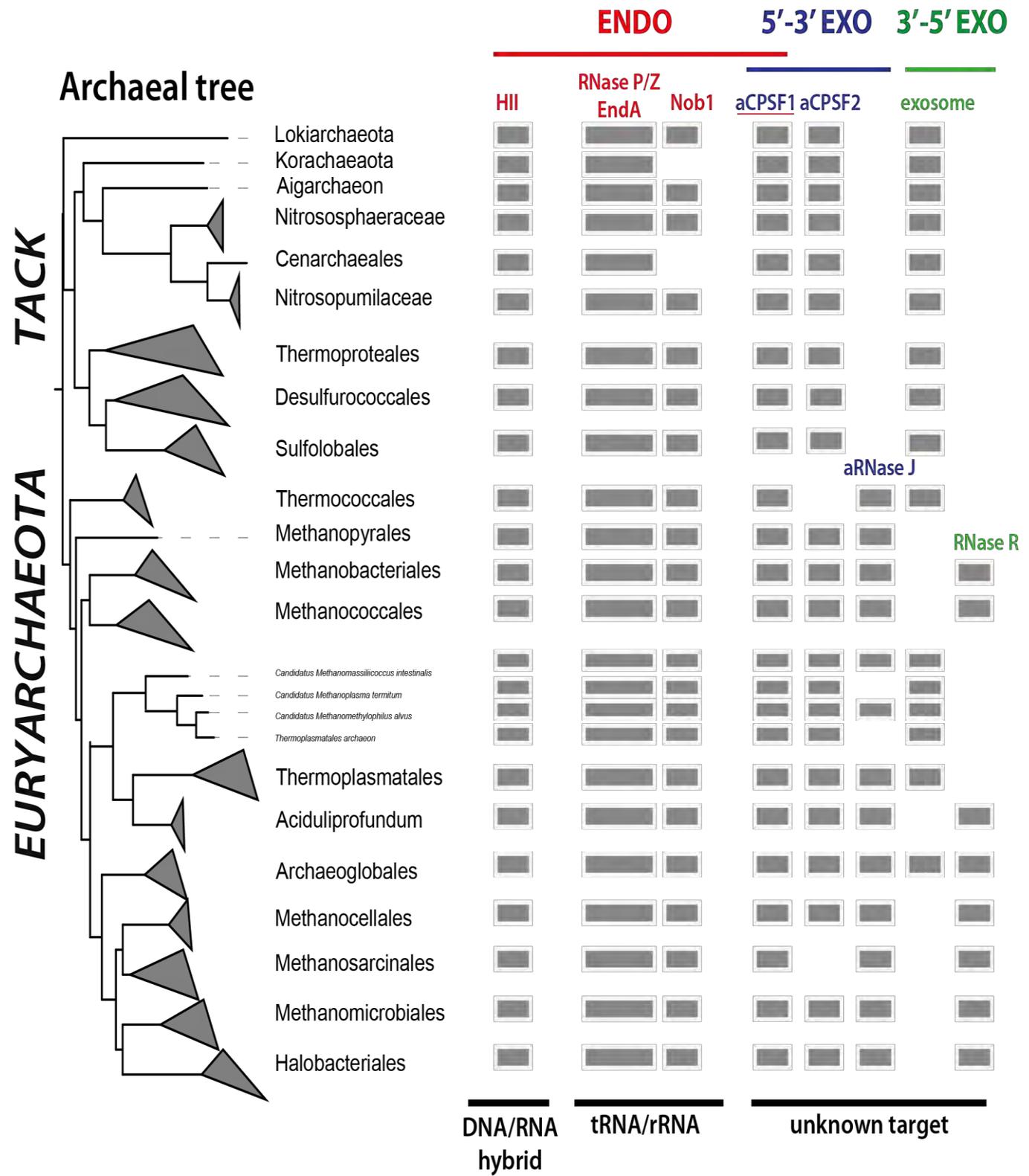
D

5'>3' EXO

ENDO



Figure 3



REMERCIEMENTS

Mes remerciements s'adressent en premier lieu à l'ensemble des membres jury qui ont consacré leur temps pour évaluer mon travail, Tamara Basta le Berre et Patrick Linder, Ciaran Condon, Gwennaële Fichant et Bruno Franzetti.

Merci à Dana par qui tout a commencé.

Je souhaite également remercier l'équipe de Gwennaële Fichant, Yves et Petra pour leurs aides dans les analyse bioinformatiques.

Merci à Anthony Henras et Didier Flament qui ont fait partie de mon comité de thèse et qui m'ont conseillé et appris à faire les gradients et les pull-down et de m'avoir accueillie dans son laboratoire à Brest.

Merci à Violette pour son temps et l'immense aide qu'elle a donné pour purifier toutes ces protéines.

Je veux aussi remercier Odile pour ces conseils sur les tests des hélicases.

J'exprime également toute ma reconnaissance à l'égard de l'ensemble de l'équipe Carpousis qui a toujours répondu présente, Marie, Leo, Isa, Lydia, Manon et A.J.

Un immense merci à tous les stagiaires qui m'ont aidé dans cette entreprise, Clarisse, Laura, Gladys, Jean-Marie, Shouan...et Mannon.

Merci à Lydia et Manon pour la relecture de ce manuscrit.

Je souhaite remercier l'ensemble des gens de l'IBCG, thésard, chercheurs, techniciens... et piliers de bar. Les compagnons de l'upsi, Elise, Alix, Clément, Natacha, Ramtin, Steph, Yves, Sylvain, Cécilia, Thomas, Lise, les copains d'enfance, Xavier, Emma, Vincent, Poupi, Sophie.

Merci à Papa et Maman qui m'ont toujours soutenu dans mes études. Je saurais un jour vous expliquer ce que je fais dans mon travail.

Je veux enfin remercier Bea pour toutes ces années au labo. Merci de ta confiance depuis le début, de ta patience lors de toutes ces échéances, de la liberté que tu m'as offerte pour tout ce que j'ai entrepris et de m'avoir fait découvrir les archées (oui, je te tutoie enfin).