



HAL
open science

Détermination du mode d'action et des substrats de RNases P protéiques chez *Arabidopsis thaliana*

Cédric Schelcher

► **To cite this version:**

Cédric Schelcher. Détermination du mode d'action et des substrats de RNases P protéiques chez *Arabidopsis thaliana*. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Strasbourg, 2017. Français. NNT : 2017STRAJ044 . tel-01745310

HAL Id: tel-01745310

<https://theses.hal.science/tel-01745310>

Submitted on 28 Mar 2018

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

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

ÉCOLE DOCTORALE 414

Institut de Biologie Moléculaire des Plantes

THÈSE

présentée par :

Cédric SCHELCHER

soutenue le : 18 septembre 2017

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Science de la Vie et de la Santé Spécialité Aspects moléculaire et cellulaire de la biologie

Détermination du mode d'action et des substrats de RNase P protéiques chez *Arabidopsis thaliana*

THÈSE dirigée par :
M GIEGE Philippe

Docteur, Université de Strasbourg

RAPPORTEURS :
M MIREAU Hakim
Mme MAYER Claudine

Docteur, INRA - AgroParis Tech
Professeure, Université Paris Diderot

AUTRES MEMBRES DU JURY :
Mme DUCHENE Anne-Marie

Professeure, Université de Strasbourg

UNIVERSITE DE STRASBOURG
Ecole doctorale des Sciences de la Vie et de la Santé

THESE DE DOCTORAT
Présentée en vue d'obtenir le grade de
Docteur de l'Université de Strasbourg

Discipline : Sciences de la vie
Spécialité : Biologie moléculaire des plantes

**Détermination du mode d'action
et des substrats de RNases P protéiques
chez *Arabidopsis thaliana***

par

SCHELCHER Cédric

Soutenue le 18 septembre 2017 devant la commission d'examen

- Rapporteur externe :	Prof. MAYER Claudine	Université Paris Diderot
- Rapporteur externe :	Dr. MIREAU Hakim	INRA - AgroParisTech
- Rapporteur interne :	Prof. DUCHENE Anne-Marie	Université de Strasbourg
- Membre invité :	Dr. SAUTER Claude	CNRS - IBMC
- Directeur de thèse :	Dr. GIEGE Philippe	CNRS – IBMP



Remerciements

Je tiens tout d'abord à remercier les professeures Claudine Mayer et Anne-Marie Duchêne, ainsi que le docteur Hakim Mireau d'avoir accepté de juger mon travail de thèse.

Je remercie mon directeur de thèse Philippe Giegé de m'avoir soutenu et supporté durant ces trois ans dans son laboratoire.

J'adresse tous mes remerciements à Anthony Gobert pour son soutien, ses conseils et son savoir qui m'ont été d'une aide inestimable.

Je remercie Claude Sauter de m'avoir fait découvrir le monde des structuralistes.

Je souhaite également remercier tous mes collègues des plateformes de l'IBMP pour leur soutien technique et logistique.

Je souhaite remercier mes amis, particulièrement Emmanuelle, Adrien, Lucie et Kévin pour leur soutien à l'intérieur et à l'extérieur du laboratoire.

Et enfin je remercie ma famille pour leur soutien durant mes 8 ans d'études.

Avant propos

Ce travail de thèse a été réalisé dans l'équipe de Philippe GIEGE à l'Institut de Biologie Moléculaire des Plantes du Centre National de la Recherche Scientifique, soutenu par l'Université de Strasbourg et financé par le Ministère de l'Education Nationale, de l'Enseignement supérieur et de la Recherche.

L'équipe de Philippe GIEGE étudie les protéines PPR (pentatricopeptide repeat), une classe majeure de protéines de liaison à l'ARN universellement conservée chez les eucaryotes. Ces protéines sont impliquées dans des processus d'expression génétique essentiels principalement dans les organites. Les protéines PPR font partie de la superfamille des protéines repliées en solénoïdes composés de motifs répétés hélicoïdaux. Chaque motif PPR d'environ 35 acides aminés se replie en une structure tridimensionnelle conservée en hélice-boucle-hélice dont la superposition forme un solénoïde en rotation dans le sens horaire. Chaque motif PPR ayant la capacité d'interagir spécifiquement avec un nucléotide, la succession de motif PPR permet la reconnaissance spécifique de séquences ARN.

En 2010, l'équipe a identifié et caractérisé les protéines PRORP, des protéines PPR responsables de l'activité RNase P chez les plantes. Cette activité de clivage du fragment 5' supplémentaire des précurseurs d'ARNt est une activité essentielle pour obtenir des ARNt fonctionnels et donc pour la traduction en général. La RNase P est l'unique exemple d'enzyme existante à la fois sous la forme d'une ribonucléoprotéine et d'une protéine seule. Cette seconde forme est spécifique des eucaryotes et assure l'activité RNase P dans les organelles et/ou le noyau.

La caractérisation des protéines PRORP a mis en évidence la présence de trois protéines PRORP chez *Arabidopsis thaliana*. AtPRORP1 est organellaires alors que AtPRORP2 et AtPRORP3 sont toutes deux localisées dans le noyau où elles ont des fonctions redondantes. La caractérisation structurale des protéines PRORP a souligné la segmentation de ces protéines en deux domaines principaux. Dans la partie N-terminale, une succession de 5 motifs PPR que l'on suppose être les domaines d'interaction avec les ARN substrats et dans la partie C-terminale, un domaine NYN PIN like, une métallonuclease responsable du clivage des précurseurs d'ARNt. L'ensemble de la protéine est structuré en forme de Λ , dont l'apex

est formé par un domaine structural de liaison au zinc. *In vivo*, les motifs PPR sont précédés par un signal d'adressage au noyau ou aux organelles.

La caractérisation biochimique des protéines PRORP a démontré leurs affinités supérieures pour les ARNt ayant une séquence 5' supplémentaire courte mais le mode d'interaction entre les protéines PRORP et ses substrats n'a pas encore été caractérisé. Ce travail de thèse s'est divisé en deux axes principaux : (i) la caractérisation de l'interaction entre AtPRORP2 par ses motifs PPR et les pré-ARNt substrats et (ii) l'étude de la diversité des substrats ARN des protéines PRORP. Le premier axe sera développé dans le chapitre 1 des résultats à travers l'étude biochimique et biophysique de mutants PPR et la caractérisation à basse résolution du complexe pré-ARNt/AtPRORP2. Le second axe sera développé au travers des chapitres 2, 3 et 4 où une approche de co-immunoprécipitation de AtPRORP1 et AtPRORP2 est décrite afin de déterminer les substrats non ARNt des protéines PRORP ainsi que leurs partenaires protéiques. Le chapitre 4 sera consacré à la complémentation de mutants de RNase P d'*Arabidopsis thaliana* par la RNase P d'*Aquifex aeolicus*, un nouveau type de RNase P protéique découvert très récemment.

Ce travail de thèse a permis de mieux comprendre l'interaction entre les protéines PRORP et leurs différents substrats. Il contribue également à replacer la fonction des protéines PRORP dans le contexte plus global des mécanismes d'expression génétique cellulaires.

Abréviations

°C : degré Celsius

Aa : acide aminé

AaRP : Aquifex aeolicus RNase P

ADN : acide désoxyribonucléique

ADN-T : ADN de transfert

ADNc : ADN complémentaire

Ala : Alanine

ARN : acide ribonucléique

ARNm : ARN messager

ARNr : ARN ribosomique

ARNsn : petit ARN nucléaire

ARNt : ARN de transfert

Asp : aspartate

ATP : adénosine triphosphate

AU : unité d'absorbance

AUC : ultra centrifugation analytique

BSA : bovine serum albumin

CCase : ARNt nucleotidyl-transférase

col0 : écotype colombia 0

Cys : cystéine

Da : Dalton

DO : densité optique

DTT : dithiothreitol

E. coli : Escherichia coli

EDTA : éthylène diamine tétra-acétate de sodium

EF : elongation factor

eIF : eukaryotic initiation factor

EYFP : enhanced yellow fluorescent protein

FF : fast flow

FPLC : Fast protein Liquid Chromatography

GDP : guanosine diphosphate

GFP : green fluorescent protein

Glu : glutamate

Gly : glycine

His : histidine

Ile : Isoleucine

IP : Immunoprécipitation

IPTG : isopropyl β -D-1-thiogalactopyranoside

IRES : Internal ribosome entry site

ISR : intermediate size repeats

ITC : Isothermal titration calorimetry

ITS : internal transcribed spacer

Kb : kilobase

KD : coefficient de dissociation

kDa : kiloDalton

L : large

LC : chromatographie liquide

LINE : autonomous long interspersed elements

miARN : micro ARN

MNU : mitochondrial nuclease

MRPP3 : Mitochondrial ribonuclease P protein 3

MS : spectrométrie de masse

MST : microscale thermophoresis

MTS : mitochondrial targeting signal

ncARN : ARN non codant

NEP : nuclear encoded polymerase

NLS : nuclear localization signal

NMA : analyse mode normal

NYN : Nedd4-BP1, YacP nucleases

OTS : organellar targeting signal

P : pure

p/v : poids / volume

PABP : poly A binding protein

PAGE : polyacrylamide gel electrophoresis

PCR : polymerase chain reaction

PEP : plastid encoded polymerase

pg : prep grade

PLS : pure large short

PNPase : Polynucleotide Phosphorylase

Pol : polymérase

PPR : pentatricopeptide repeat

Pré-ARNt : ARNt précurseur

Pro : proline

PRORP : protein only RNase P

Rg : rayon de giration

RIN : RNA integrity number

RNP : Ribonucléoprotéine

S : short

S : Svedberg

SAXS : Small angle X-ray scattering

SDS – PAGE : Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SDS : sodium dodecyl sulfate

Sec : selenocystéine

SEC : size exclusion chromatography

Ser : serine

siARN : petit ARN interférent

SINE : short interspersed nuclear elements

snRNP : small nuclear ribonucleoprotein

Tfc : Transcriptional factor

TLS : tRNA-like structure

Tris : trishydroxyméthylaminométhane

U : unité enzymatique

UTR : Untranslated Transcribed Region

Val : valine

Table des matières

I. Introduction.....	1
1. L'expression génétique chez les plantes	1
a. Les différents compartiments cellulaires où l'expression génétique a lieu chez les plantes	1
b. Les différents génomes des plantes	1
Le génome nucléaire.....	1
Le génome mitochondrial	2
Le génome chloroplastique	3
c. Les mécanismes d'expression génétique	3
L'expression génétique nucléaire	3
La transcription nucléaire.....	3
Maturation des extrémités 5' et 3' des transcrits nucléaires	4
Epissage des transcrits nucléaires.....	5
La traduction des transcrits nucléaires.....	6
L'expression génétique chloroplastique	6
La transcription chloroplastique.....	6
Maturation des extrémités 5' et 3' des transcrits chloroplastiques.....	7
Epissage des transcrits chloroplastiques	7
L'édition des transcrits chloroplastiques	8
La traduction des transcrits chloroplastiques.....	8
L'expression génétique mitochondriale.....	10
La transcription mitochondriale	10
Maturation des extrémités 5' et 3' des transcrits mitochondriaux	10
Epissage des transcrits mitochondriaux.....	11
L'édition des transcrits mitochondriaux.....	11
La traduction des transcrits mitochondriaux.....	12
2. Les protéines PPR	13
a. Structure et distribution des protéines PPR.....	13
b. Les différentes sous-familles de protéines PPR.....	14
c. Description des fonctions identifiées.....	15
d. Mécanisme de reconnaissance des nucléotides par les motifs PPR, définition du "code PPR"	16
3. Biogénèse et maturation des ARNt chez les plantes	16
a. Répartition des gènes d'ARNt et des TLS chez les plantes	16
b. Structure et fonction des ARNt	17

c. Biogenèse des ARNt fonctionnels.....	18
Les ARNt nucléaires.....	18
Transcription des ARNt nucléaires.....	18
Maturation des ARNt nucléaires.....	18
Les ARNt mitochondriaux.....	19
Transcription des ARNt mitochondriaux.....	19
Maturation des ARNt mitochondriaux.....	19
Les ARNt chloroplastiques.....	22
Transcription et maturation des ARNt chloroplastiques.....	22
d. Publication 1 : Transfer RNA: From pioneering crystallographic studies to contemporary tRNA biology.....	22
4. Les enzymes à activité RNases P.....	24
a. Les RNase P ribonucléoprotéiques.....	24
b. Les RNases P protéiques.....	25
c. Définition et caractéristiques des RNases P protéiques.....	26
d. PRORP chez Arabidopsis thaliana.....	26
e. Diversité des protéines PPR à activité RNase P.....	27
5. Comparaison de la RNase P RNP et de PRORP.....	28
a. Répartition dans le vivant.....	28
b. Mode de reconnaissance des ARNt.....	29
c. Mécanismes de clivage.....	29
d. Comparaison structurale et mécanistique des RNase P protéiques et ribonucléoprotéiques.....	30

II. Objectifs de ma thèse 32

III. Matériel 33

1. Souches bactériennes.....	33
a. Souches d'Escherichia coli.....	33
TOP 10.....	33
BL21.....	33
b. Souches d'Agrobacterium tumefaciens.....	33
2. Matériel végétal.....	33
a. Plantes p2/p3.....	34
b. Plantes AtPRORP2-HA.....	34
c. Plantes AtPRORP1-HA.....	34
3. Plasmides.....	34

a. pET28b.....	34
b. pUC19.....	34
c. pDONR207.....	34
d. pGWB1.....	35
4. Amorces.....	35
a. Amorces de mutagenèse.....	35
b. Amorces de PCR.....	36
- Clonage de la séquence 5' UTR de PROPR2 et ajout d'une étiquette HA en 3' ainsi que des sites de recombinaison Gateway AttB en 5' et 3'.....	36
c. Amorces de séquençage.....	36
- Séquençage de AtPRORP2 dans pET28b	36
- Séquençage de la séquence 5' UTR de AtPRORP2 dans pDONR207	36
d. Amorces de génotypage des plantes.....	36

IV. Méthodes..... 37

1. Techniques de clonage	37
2. Mutagenèse dirigée de fragments d'ADN.....	38
3. Transformation par choc thermique de bactéries <i>E. coli</i> TOP10	39
4. Préparation de plasmides par lyse alcaline	39
5. Séquençage des plasmides.....	40
6. Expression de protéines dans des cellules de <i>E. coli</i> BL21.....	40
7. Purification des protéines par chromatographie d'affinité	41
8. Purification des protéines par filtration sur gel	41
9. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS- PAGE).....	42
10. Coloration de gel de polyacrylamide	42
11. Transcription <i>in vitro</i>	43
a. Transcription de petites quantités pour clivage.....	43
b. Transcription en grande quantité pour les analyses biophysiques	43
c. Purification de grandes quantités de transcrits	43
12. Extraction des acides nucléiques au phénol / chloroforme	44
13. Précipitation à l'éthanol des acides nucléiques	44
14. Test d'activité de PRORP <i>in vitro</i>.....	44
15. Ultracentrifugation analytique.....	44
16. Thermophorèse à micro échelle	45
17. Titration calorimétrique isotherme	46

18. Diffusion des rayons X aux petits angles	46
19. Agrotransformation d' <i>Arabidopsis thaliana</i> par immersion florale.....	47
20. Extraction d'ADN génomique d' <i>Arabidopsis thaliana</i>	47
21. Génotypage des plantes.....	47
22. Purification des mitochondries à partir de parties aériennes de plantes d' <i>Arabidopsis</i>	48
23. Co-immunoprécipitation de protéines avec des ARN	49
24. Analyse d'ARN par hybridation de type northern.....	50
25. Analyse des ARN par séquençage à haut débit.....	50
a. Analyse qualitative des ARN.....	50
b. Construction des banques d'ADNc	51
c. Séquençage par la technologie Illumina	51
26. Analyse des données de séquençage à haut débit.....	52
a. Contrôle de la qualité du séquençage.....	52
b. Alignement des séquences sur le génome de référence.....	52
c. Analyse comparative des banques d'ADNc	52

V. Résultats chapitre 1 : Etude du rôle des motifs

PPR de AtPRORP2 pour l'interaction avec les pré-ARNt 53

1. Identification des motifs PPR et des positions susceptibles de participer aux interactions avec AtPRORP2.....	53
2. Mutagénèses dirigées des positions 5 et 35 de chaque motif PPR de AtPRORP2	55
3. Optimisation du protocole de purification de AtPRORP2 et de ses mutants ..	56
4. Etude de l'interaction entre AtPRORP2 et les pré-ARNt.....	57
a. Vérification de la conformation des mutants des motifs PPR.....	58
b. Tests d'activité de AtPRORP2 et des mutants de motifs PPR catalytiquement actifs.....	59
c. Test d'affinité entre le mutant catalytiquement inactif AtPRORP2DD et les pré- ARNt.....	60
Etude d'interaction par ultra centrifugation analytique.....	60
Etude d'interaction par Titration calorimétrique isotherme	60
Etude d'interaction par thermophorèse à micro échelle.....	60
Comparaison des différentes méthodes	61

d. Test d'interaction des protéines mutées dans leur motif PPR et catalytiquement inactifs avec les pré-ARNt	62
5. Détermination d'un modèle structural du complexe AtPRORP2 / pré-ARNt..	63
6. Mise en évidence de la flexibilité structurale des protéines PRORP	64

VI. Résultats chapitre 2 : Etude de la diversité des substrats ARN de AtPRORP1 66

1. Co-immunoprécipitation de AtPRORP1 avec ses substrats ARN	66
2. Identification des ARN immunoprécipités par séquençage à haut débit.....	67
a. Choix du type de banque d'ADNc à construire	67
b. Déplétion des ARNr et construction des banques	68
c. Séquençage et analyses de sa qualité.....	68
d. Analyses des résultats de séquençage	69

VII. Résultats chapitre 3 : Etude de la diversité des substrats ARN et des partenaires protéiques de la RNase P nucléaire AtPRORP2 71

1. Construction de plantes complémentées exprimant une version étiquetée de AtPRORP2.....	71
a. Transformation des plantes P2/p2 ; p3/p3 et sélection des plantes avec P2-HA	71
b. Sélection des plantes AtPRORP2-HA/AtPRORP2-HA ; Atprorp2/Atprorp2 ; prorp3/prorp3.....	72
c. Mise au point de la purification de noyaux d'Arabidopsis.....	72

VIII. Résultats chapitre 4 : Complémentation de mutants de PRORP d'Arabidopsis avec une nouvelle RNase P protéique d'*Aquifex aeolicus* 74

1. Fusion de séquences d'adressage d'Arabidopsis à AaRP	74
2. Localisation des différentes versions d'AaRP dans <i>Nicotiana benthamiana</i> ...	75
3. Complémentation de mutants de PRORP d' <i>Arabidopsis thaliana</i> avec AaRP ..	75

IX. Discussion et Perspectives..... 77

1. Mode d'action des protéines PRORP	77
--	----

a. Mode de reconnaissance des ARN par les motifs PPR de PRORP.....	77
b. Nature du complexe formé par PRORP et l'ARNt.....	78
c. Flexibilité des protéines PRORP.....	78
d. Adaptabilité conformationnelle des enzymes de liaison aux ARNt.....	79
2. Diversité de fonctions des protéines PRORP	80
a. Partenaires protéiques des protéines PRORP	81
b. Diversité de substrats ARN des protéines PRORP	82
3. Diversité des enzymes à activité RNase P au cours de l'évolution.....	83
X. Bibliographie	86

Liste des figures

- Figure 1 : Les endosymbioses menant aux plantes supérieures.
- Figure 2 : Compartiments et génomes dans la cellule végétale.
- Figure 3 : Transcription d'un ARNt nucléaire par l'ARN Polymérase III.
- Figure 4 : Epissage des ARNt nucléaires.
- Figure 5 : Initiation de la traduction cytosolique dépendante de la coiffe.
- Figure 6 : Structure des motifs PPR et interaction avec l'ARN
- Figure 7 : Structure des différentes sous familles de protéines PPR.
- Figure 8 : Les différentes étapes de maturation des ARN mitochondriaux de plantes.
- Figure 9 : Mode de reconnaissance des ARN par les motifs PPR définissant un "code PPR".
- Figure 10 : Structure 2D et 3D des ARNt.
- Figure 11 : Modification des ARNt.
- Figure 12 : Structures des RNase P RNP de type A et B.
- Figure 13 : les domaines de PRORP.
- Figure 14 : Localisations et activités connues des protéines PRORP chez Arabidopsis.
- Figure 15 : Répartition des RNase P de type RNP et PRORP chez les eucaryotes.
- Figure 16 : Mécanismes de clivage de l'ARN utilisé par les RNase P de type PRORP et RNP.
- Figure 17 : Structure cristallographique de AtPRORP2
- Figure 18 : Séquences de PRORP végétales et nature des mutations réalisées.
- Figure 19 : Fréquence des acides aminés à toutes les positions des motifs PPR de protéines PRORP du groupe III dans le règne végétal.
- Figure 20 : Protocole de purification optimisé de AtPRORP2.
- Figure 21 : Dimère de AtPRORP2 dans une maille du cristal
- Figure 22 : Purification optimisée de AtPRORP2 et de ses mutants.
- Figure 23 : Comparaison des courbes SAXS des mutants PPR.
- Figure 24 : Activités RNase P *in vitro* des mutants PPR de PRORP2.
- Figure 25 : Tentative de restauration de l'activité RNase P avec des mutants PPR et des pré-ARNt mutés.
- Figure 26 : Profils d'ultracentrifugation analytique (AUC) obtenus avec PRORP2 DD seule et l'ARN seul.
- Figure 27 : Analyse du complexe PRORP / pré-ARNt par titration calorimétrique isotherme (ITC).
- Figure 28 : Analyse du complexe PRORP / pré-ARNt par thermophorèse à micro échelle.
- Figure 29 : Analyse du complexe PRORP / pré-ARNt réalisée avec différents mutants PPR et le pré-ARNt^{cys} mitochondriale L5T0.
- Figure 30 : Structure du complexe AtPRORP2 / pré-ARNt par SAXS
- Figure 31 : Variants conformationnels de PRORP2.
- Figure 32 : Analyse des ARN issus des immuno-précipitations (IP) par hybridation de type northern avec une sonde spécifique de l'ARNt Lysine mitochondrial.
- Figure 33 : Analyse par spectrométrie de masse de la présence de PRORP1 et d'un de ses partenaires connu dans les immuno-précipitations.
- Figure 34 : Déplétion des ARNr avant la construction des banques d'ADNc.
- Figure 35 : Densité de clusters (K) sur la cellule.
- Figure 36 : Nombre et qualité des « reads » obtenus lors du séquençage.
- Figure 37 : Contrôle de la qualité des séquences.
- Figure 38 : Evaluation de la proportion des séquences qui s'alignent sur les génomes de référence.
- Figure 39 : Visualisation de transcrits spécifiquement accumulés dans les fractions immuno-précipitées par rapport au transcriptome total des mitochondries d'Arabidopsis.

Figure 40 : Analyse des données de séquençage à 4 loci représentatifs

Figure 41 : Génotypage des plantes P2/p2 ; p3/p3.

Figure 42 : Génotypage de plantes transformées avec la construction PRORP2-HA.

Figure 43 : Comparaison structurale des RNase P de type RNP et PRORP.

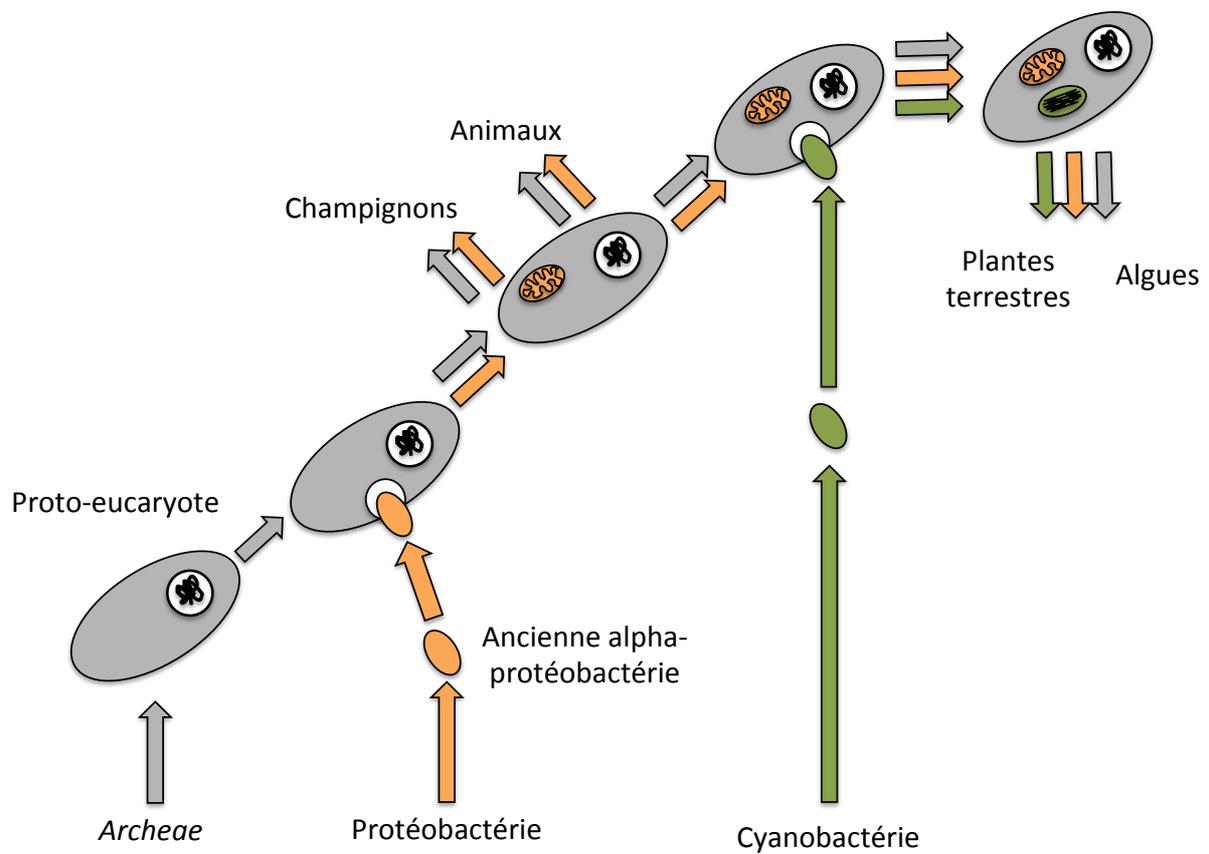


Figure 1 : Les endosymbioses menant aux plantes supérieures.

La mitochondrie est issue d'une première endosymbiose d'une alpha-protéobactérie (orange) avec l'ancêtre des eucaryotes. De cette cellule ont dérivés les animaux et les champignons. Une seconde endosymbiose d'une cyanobactérie (vert) a permis à ces cellules d'obtenir un chloroplaste, ce qui a permis l'essor des plantes terrestres et des algues.

I. Introduction

1. L'expression génétique chez les plantes

a. Les différents compartiments cellulaires où l'expression génétique a lieu chez les plantes

Les plantes sont des organismes eucaryotes photosynthétiques. L'expression génétique a lieu dans le noyau mais aussi dans deux organites semi-autonomes possédant leurs propres génomes : la mitochondrie et le chloroplaste. Ces deux organites sont issus d'événements d'endosymbiose indépendants. L'endosymbiose primaire d'une α -protéobactérie a donné lieu à l'apparition de la mitochondrie et de ce fait au premier organisme eucaryote unicellulaire hétérotrophe (Gray et al., 1999). Une endosymbiose secondaire d'une cyanobactérie a conduit à l'apparition des plastides et à la naissance des premiers représentants du règne végétal (Raven and Allen, 2003) (Figure 1).

Au cours de l'évolution, de nombreux gènes des génomes des organites ont été transférés au noyau les rendant ainsi interdépendants (Timmis et al., 2004). En particulier, 18% des gènes nucléaires d'*A. thaliana* seraient issus d'une cyanobactérie et donc du proto-plaste (Martin et al., 2002). Ce transfert engendre la nécessité d'un mécanisme d'import des protéines vers les organites. Les protéines des organites codées par le génome nucléaire sont traduites dans le cytosol puis importées grâce à des séquences d'adressages spécifiques. Certaines protéines possèdent un double adressage et peuvent être importées dans les deux organites, comme par exemple de nombreuses aminoacyl ARNt synthétases (Duchêne et al., 2005). On peut noter que le transfert génétique inverse (c'est-à-dire du noyau vers les organites) s'est aussi produit. En particulier, 4 % du génome mitochondrial serait issu de séquences dérivant de rétrotransposons nucléaires (Unsel et al., 1997).

b. Les différents génomes des plantes

Le génome nucléaire

Comme pour tous les autres eucaryotes, le génome nucléaire est le plus grand des génomes de la cellule végétale. Il est réparti sur différents chromosomes dont le nombre et la taille varient suivant l'organisme. Celui d'*Arabidopsis thaliana* est comparativement de taille réduite, avec 135 Mb répartis sur 5 chromosomes alors que celui du maïs *Zea mays* a une taille de 2067 Mb répartis sur 20 chromosomes ou 417 Mb sur 19 chromosomes pour le

peuplier *Populus trichocarpa*. La taille des génomes nucléaires est très variable en raison d'une multitude de mécanismes d'expansion du génome. Le génome nucléaire végétal possède un grand nombre de transposons et de rétrotransposons en particulier chez le maïs. Certaines plantes possèdent un génome totalement ou partiellement dupliqué. Il existe également un grand nombre de séquences répétées acquises durant l'évolution (Maumus et al., 2014). Au cours de l'évolution, certains gènes ont également été transférés des organites au noyau, accroissant ainsi la taille du génome nucléaire au détriment de celle du génome d'origine.

Le génome nucléaire code des ARNm (ARN messenger), des ARNt (ARN de transfert), des ARNr (ARN ribosomique), des ARNsn (petit ARN nucléaire ou « small nuclear RNA »), des miARN (micro ARN), des siARN (petit ARN interférents ou « small interfering RNA ») ainsi que de nombreux autres ncARN (« non coding RNA »). Le génome nucléaire code des protéines nucléaires, cytoplasmiques et des organites. Les différents ARNm sont transcrits dans le noyau, maturés puis traduits dans le cytosol. Les protéines obtenues sont alors adressées aux différents compartiments.

Le génome mitochondrial

Suite à un transfert massif d'information génétique vers le noyau, les génomes mitochondriaux ont une taille très réduite par rapport aux génomes procaryotiques dont ils sont issus. Toutefois, en comparaison des autres eucaryotes, les plantes possèdent de très grands génomes mitochondriaux. Par exemple, chez *Arabidopsis thaliana*, le génome mitochondrial a une taille de 366 kb et de 569 kb chez *Zea mays* alors que les génomes mitochondriaux des animaux sont bien plus compacts avec des tailles variant de 16 à 18 kb (Schneider and Ebert, 2004).

Le génome mitochondrial des plantes est composé de molécules d'ADN circulaires et linéaires comportant un grand nombre de séquences répétées dont deux grandes répétitions de 6,5 et 4,2 kb chez *Arabidopsis* et une multitude de répétitions de tailles intermédiaires (ISR) de 50 à 600 pb. La complexité de cette structure génomique engendre une très grande versatilité avec un nombre important de recombinaisons homologues, ce qui entraîne l'hétéroplasmie c'est-à-dire la présence de mitochondries ayant des génomes de structure et de composition différentes dans une même cellule. Cette complexité implique un contrôle très étroit de la recombinaison et de la transmission de l'information génétique mitochondriale.

Le génome mitochondrial d'*Arabidopsis* contient environ 60 gènes dont 32 codent des protéines principalement de la chaîne respiratoire, 3 ARNr et 23 ARNt. Le contenu génétique

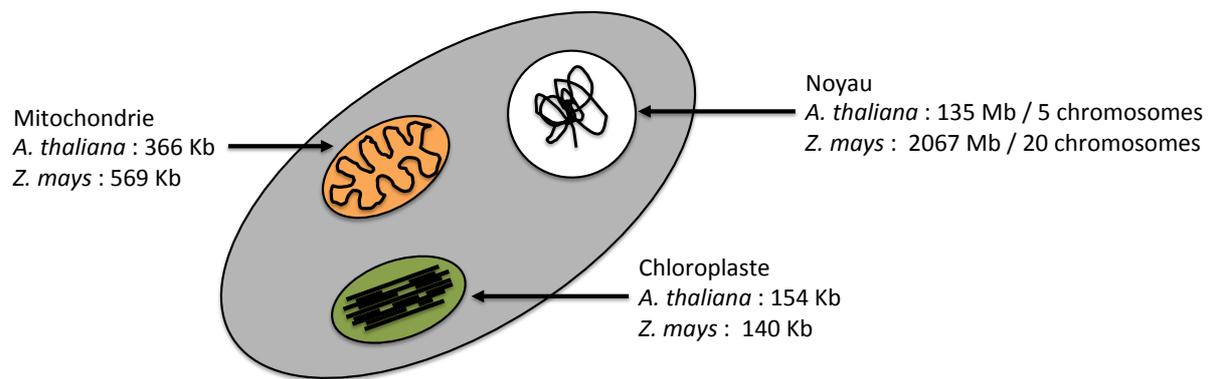


Figure 2 : Compartiments et génomes dans la cellule végétale.

Il existe trois compartiments possédant un génome et son propre système d'expression génétique dans la cellule végétale. Les tailles des génomes variables suivant les espèces sont indiquées ici pour *Arabidopsis thaliana* et le maïs *Zea mays*.

des mitochondries est relativement bien conservé au cours de l'évolution, toutefois, les plantes se distinguent des autres eucaryotes par la présence de gènes codant pour des protéines de maturation des cytochromes de type-c et des protéines ribosomales dans leurs génomes mitochondriaux (Unsold et al., 1997). Les autres protéines et les ARN nécessaires à l'activité mitochondriale non codés par le génome mitochondrial sont importés du cytosol (Gualberto et al., 2014). Ainsi, de nombreux complexes, comme les complexes de la chaîne respiratoire ou le ribosome mitochondrial sont composés de sous-unités codées dans le noyau et d'autres sous-unités codées dans les mitochondries, ce qui nécessite une coordination étroite de l'expression des génomes nucléaires et mitochondriaux (Giegé et al., 2005).

Le génome chloroplastique

Le génome chloroplastique est le plus petit des génomes de la cellule végétale avec seulement 154 kb chez *Arabidopsis*, 140 kb pour *Zea mays* et 157 kb chez *Populus trichocarpa*. Le génome chloroplastique est classiquement représenté sous forme circulaire comportant deux larges parties répétées inversées divisant le génome en une petite et une grande région simple copie.

Le génome chloroplastique d'*Arabidopsis* code 3 ARNr, entre 37 ARNt, 3 des 4 sous-unités de son ARN polymérase de type bactérien, les protéines des photosystèmes I et II, l'ATP synthase et quelques autres protéines chloroplastiques. Les autres ARN et protéines essentielles aux chloroplastes sont, comme pour la mitochondrie, transcrits dans le noyau, les protéines traduites dans le cytosol puis importées dans le chloroplaste (Green, 2011) (Figure 2).

c. Les mécanismes d'expression génétique

L'expression génétique nucléaire

La transcription nucléaire

La première étape de l'expression génétique nucléaire des plantes nécessite cinq ARN polymérase (pol) qui synthétisent chacune des ARN spécifiques comme par exemple les ARNm (pol II), les ARNr (pol I et III), les ARNt (pol III), les miARN (pol II) et siARN (pol IV et pol V). Chacune de ces polymérases requiert la présence de séquence consensus en amont du site d'initiation de la transcription.

L'ARN polymérase I nécessite la présence de deux séquences distinctes appelées promoteur core et une séquence appelée « upstream promoter element » (UPE ou UE)

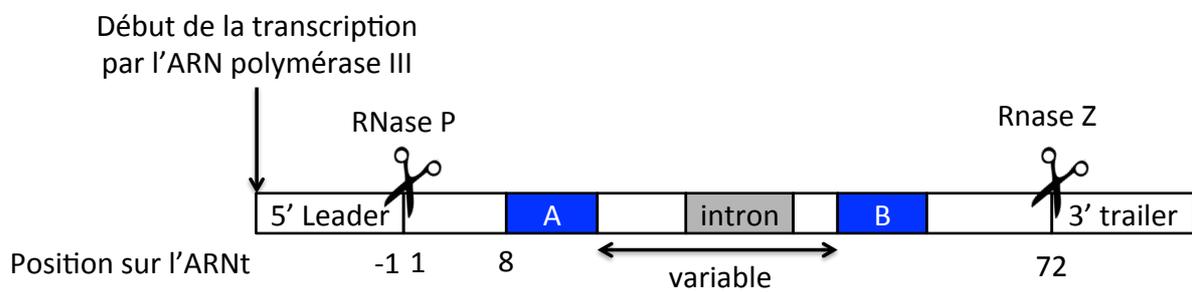


Figure 3 : Transcription d'un ARNt nucléaire par l'ARN Polymérase III.

Les pré-ARNt sont transcrit par l'ARN polymérase III qui commence la transcription en amont du premier nucléotide de l'ARNt et la termine en aval du dernier nucléotide. Ceci nécessite deux clivages endonucléolitiques par la RNase P et RNase Z. Le promoteur est en deux parties appelées boîtes A et B, situées dans les domaines T et D de l'ARNt, qui recrute TfcIII qui lui-même recrute l'ARN polymérase. La distance entre les deux boîtes est variable et peut contenir un intron.

originellement appelée « upstream control sequence » (UCE) chez l'Homme. L'espace entre ces deux éléments est crucial *in vivo* mais le promoteur core est suffisant pour la transcription *in vitro*. Des séquences promotrices courtes de longueur similaire au promoteur core semblent être la norme dans les plantes et certains organismes unicellulaires (Moss et al., 2007).

La majorité des gènes codant pour des protéines possèdent à environ 30 pb en amont du site d'initiation une boîte TATA ayant pour fonction le recrutement de l'ARN pol II. Cette boîte TATA est généralement accompagnée d'une boîte CAT et d'un domaine riche en GC entre les positions -40 et -200 ayant le rôle de régulateur de la transcription. En amont, d'autres éléments peuvent être présents afin de recruter des facteurs de transcription permettant d'augmenter ou diminuer la transcription du gène (Deppmann et al., 2006).

L'ARN pol III nécessaire à la synthèse des ARNt et de l'ARNr 5S reconnaît des séquences promotrices intragéniques appelées boîtes A et B contenues dans la boucle T et D des ARNt. Ces deux boîtes recrutent TfcIII (transcriptional factor III) qui recrute l'ARN polymérase III qui commence le transcrit en amont du site A (Figure 3).

Cette transcription est régulée de différentes manières. Certains promoteurs sont plus actifs que d'autres en fonction de la présence de séquences régulatrices « enhancer » ou « silencer » où se fixent des facteurs de transcriptions qui peuvent être activateur ou inhibiteur de la transcription (Kolovos et al., 2012).

La transcription des gènes est aussi régulée par des facteurs épigénétiques, impliquant en particulier la modification de la queue des histones par l'adjonction post-traductionnelle de groupements e.g. méthyl, acétyl ou phosphate. Ces groupements induisent des changements de compaction de la chromatine, modifiant l'accessibilité de l'ADN autour des histones, permettant ou interdisant l'accès de la polymérase notamment au cours du développement ou de manière tissu-spécifique. Par ailleurs, la méthylation de l'ADN au niveau des îlots CpG permet d'inhiber l'expression génétique notamment au cours du développement ou de manière tissu-spécifique (Saze et al., 2003).

Maturation des extrémités 5' et 3' des transcrits nucléaires

L'ajout d'une coiffe en 5' est spécifique aux ARNm nucléaires. Cet ajout permet la stabilisation du transcrit contre la dégradation de 5' vers 3' (Shimotohno et al., 1977) et l'interaction avec le facteur d'initiation eIF4E qui permet de recruter indirectement le ribosome (Mitchell et al., 2010). La présence de la coiffe est également indispensable à l'exportation du transcrit en dehors du noyau par interaction avec le « cap-binding complex »

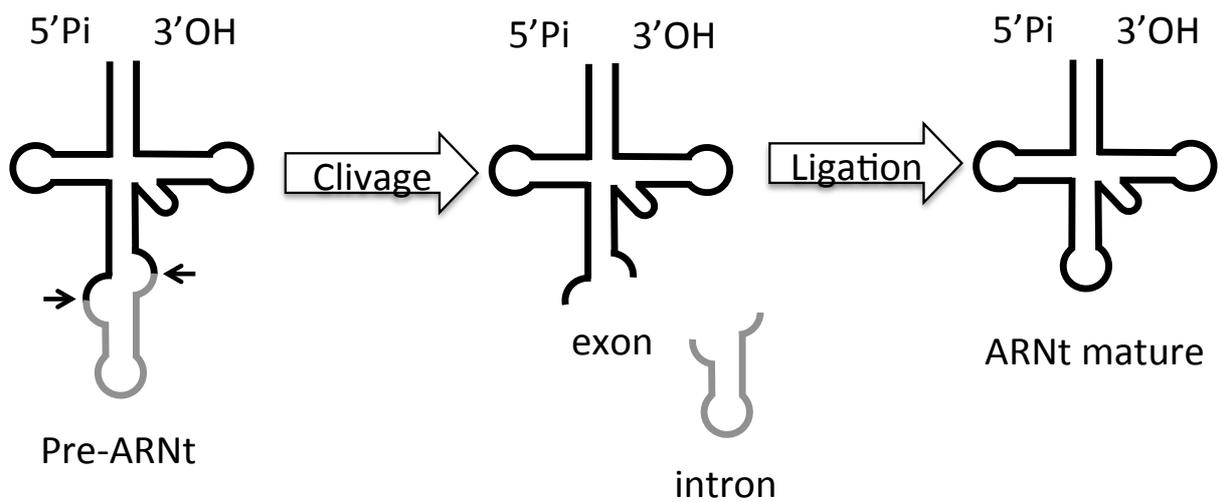


Figure 4 : Epissage des ARNt nucléaires.

La présence d'un intron dans la boucle de l'anti codon nécessite son épissage. Un clivage par une endonucléase permet d'exciser l'intron et une ligase va liguer les deux exon de l'ARNt. Chez *Arabidopsis thaliana* le clivage est réalisé par AtSen1 et AtSen2 et la ligation est assurée par une ARNt ligase.

(Nojima et al., 2007). La coiffe est un résidu guanosine méthylé en position 7 par une méthyltransférase afin d'obtenir une extrémité 5' 7mGpppN (Banerjee, 1980). L'ajout de ce résidu guanosine est possible après suppression du phosphate à l'extrémité 5'.

Chez les eucaryotes, l'ajout d'une longue queue poly A permet de protéger le transcrit de la dégradation de 3' vers 5' (Moore and Proudfoot, 2009). La poly(A) polymérase canonique ajoute au transcrit néo synthétisé des résidus adénosines monophosphates en utilisant comme substrat l'ATP (Chen et al., 2010). Cette queue poly A sert de support d'interaction des PABP (poly A binding protein) jouant un rôle dans l'export, la stabilisation et la traduction des ARNm polyadénylés. Des queues poly A plus courtes trouvées chez les eucaryotes ont un rôle similaire aux queues poly A des procaryotes, où elles servent de signal pour l'adressage vers l'exosome et leur dégradation (Schmid and Jensen, 2008; Lange et al., 2009, 2011).

Chez *A. thaliana*, l'ARNr cytosolique est transcrit sous la forme d'un ARN polycistronique composé des ARNr 18S, 5,8S et 25S séparés par les ITS1 (« internal transcribed spacer ») et ITS2. Les extrémités sont digérées par des exonucléases puis les ITS sont clivés par des endonucléases afin de libérer les différents ARNr (Weis et al., 2015). On notera en particulier, l'implication de la RNase MRP, une RNase reliée évolutivement à la RNase P ribonucléoprotéique, dans le clivage du transcrit ITS1 au niveau du site A3 (Gutmann et al., 2012).

Épissage des transcrits nucléaires

Les ARNm nucléaires sont épissés grâce au spliceosome. Il s'agit d'un grand complexe ribonucléoprotéique constitué de cinq complexes nucléoprotéiques appelés snRNP (« small nuclear ribonucleoprotein ») en association avec plusieurs facteurs protéiques. U1 se lie d'abord du côté 5' de l'intron. U2 se lie alors à un site spécifique sur l'intron. U4 et U6 pontent U1 et U2 afin de réaliser un rapprochement entre le côté 5' de l'intron et la séquence de liaison de U2. U5 s'associe au complexe et rapproche les extrémités 5' et 3' de l'exon et U1 ainsi que U4 quittent le complexe. Le groupement 2'-OH du A du site de fixation de U2 clive la jonction 5' de l'intron et le groupement 3'-OH du nucléotide en 3' de l'exon coupe l'autre jonction ce qui épisse l'ARNm et libère l'intron en forme de lasso (Will and Lührmann, 2011).

Certains ARNt nucléaires sont également épissés grâce à une endonucléase qui libère l'intron puis les exons sont ligués par une ligase (Yoshihisa, 2014) (Figure 4). Chez *A.*

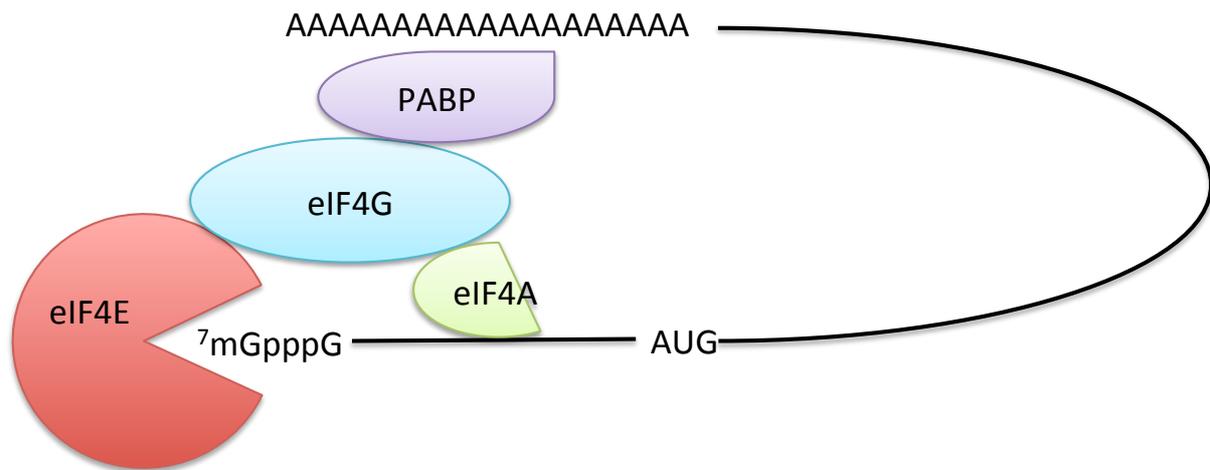


Figure 5 : Initiation de la traduction cytosolique dépendante de la coiffe.

Représentation schématique de la circularisations de l'ARNm par interaction entre les protéines composant le complexe eIF4F (eIF4E, eIF4G et eIF4A) et les protéines de liaison à la queue poly A (PABP).

thaliana, le clivage est assuré par AtSen1 et AtSen2 (Akama et al., 2000) alors que la ligation est réalisée par une ARNt ligase (Englert et al., 2007).

La traduction des transcrits nucléaires

Les ARNm transcrits dans le noyau sont d'abord exportés vers le cytosol pour être traduits. Il existe deux types principaux de mécanisme d'initiation de la traduction décrits pour les ARNm nucléaires.

L'initiation de la traduction dépendant de la coiffe implique l'activation de l'ARNm par l'interaction avec la coiffe en 5' du complexe d'initiation eIF4F composé des facteurs eIF4E, eIF4G et eIF4A. Le facteur eIF4G se lie aux protéines de liaison aux polyA fixées sur la queue poly A (Poly A binding protein : PABP) de l'ARNm, créant un ARNm circulaire (Figure 5). La plupart des ARNm eucaryotes sont traduits par un mécanisme de « balayage », où la petite sous-unité ribosomale (40S) est d'abord chargée de l'ARNt Met initiateur dans un complexe ternaire avec le facteur d'initiation eucaryote 2 (eIF2) lié au GTP dans une réaction promue par eIF1, EIF1A, eIF5 et l'eIF3. Ce complexe de pré-initiation 43S s'attache ensuite à l'extrémité 5' de l'ARNm. Le domaine 5' UTR de l'ARNm est ensuite scanné base par base jusqu'à trouver la séquence complémentaire à l'anticodon de l'ARNt Met initiateur au niveau du site P de la sous-unité 40S du ribosome. Le premier AUG rencontré est généralement privilégié comme codon d'initiation, mais peut être ignoré si son contexte n'est pas conforme au consensus "Kozak", avec une purine à la position -3 et une guanine en position +4 par rapport à l'AUG (+1). La reconnaissance de l'AUG déclenche l'hydrolyse du GTP dans le complexe tertiaire et relâche eIF2-GDP de l'ARNt Met initiateur, ce qui permet la liaison de la grande sous-unité (60S), aidé par eIF5B, pour former un complexe d'initiation 80S.

La traduction peut aussi être initiée de manière indépendante de la coiffe. Certains ARNm peuvent être traduits par des sites d'entrée internes des ribosomes. Ces éléments d'ARN structurés appelés IRES sont trouvés dans des ARNm eucaryotes ainsi que dans des ARN viraux traduits dans leurs hôtes eucaryotes (Cui et al., 2015; Koh et al., 2003).

L'expression génétique chloroplastique

La transcription chloroplastique

Le génome chloroplastique est transcrit par des ARN polymérases de deux origines évolutives différentes.

Les premières d'origines phagiques sont codées par le noyau et sont appelées NEP pour « Nuclear Encoded Polymerase ». Il en existe trois appelées RpoTm adressée à la

mitochondrie, RpoTp adressée au chloroplaste et RpoTmp doublement adressée au chloroplaste et à la mitochondrie.

Par ailleurs, le génome chloroplastique code une ARN polymérase de type bactérienne appelée PEP pour « Plastid Encoded Polymerase ». La reconnaissance des promoteurs par la PEP se fait à la manière des polymérases bactériennes, grâce à des facteurs sigma, qui sont eux codés par le noyau (Nagashima et al., 2004a, 2004b; Ishizaki et al., 2005). Les promoteurs bactériens possèdent deux séquences conservées TATAAT et TTGACA localisées en position -10 et -35 du site d'initiation (Weihe, 2004). La régulation de l'expression du génome chloroplastique est alors tributaire de la régulation des facteurs sigma qui sont eux-mêmes régulés différemment en fonction du cycle circadien (Ichikawa et al., 2008), du développement (Kasai et al., 2004) ou du tissu (Isono et al., 1997b, 1997a).

Maturation des extrémités 5' et 3' des transcrits chloroplastiques

Certains ARN polycistroniques sont clivés par différents événements de maturation endonucléolotique. Les extrémités 5' et 3' des transcrits monocistronique peuvent ensuite être marquées par des protéines telles que PPR10 (voir ci-dessous la section sur les protéines PPR) dans le chloroplaste du maïs qui va protéger le transcrit de la digestion par des exonucléases telles que la PNPase et permettre d'obtenir un transcrit mature des ARNm de *atpH* et *psaJ* (Pfalz et al., 2009). Dans un autre exemple, il a été montré récemment que PPR103 dans le chloroplaste du maïs permet de définir l'extrémité 5' de l'ARNm de *rpl16* en protégeant la partie codante de la dégradation de 5' vers 3' par des exonucléases (Hammani et al., 2016).

Épissage des transcrits chloroplastiques

Les transcrits chloroplastiques possèdent des introns du groupe I et II. Ces deux types d'introns ont initialement été décrits comme des ARN autocatalytiques dont on pouvait prédire que l'épissage ne nécessiterait aucun facteur additionnel. Toutefois, *in vivo*, leur épissage fait appel à des facteurs multiples, i.e. à différentes maturases et à des protéines PPR (Khrouchtchova et al., 2012) (voir ci-dessous).

Dans les introns de groupe I, un *exoG* présent dans l'intron se fixe sur le site de liaison au G présent dans la région P7 ce qui aligne son extrémité 3'-OH avec la partie 5' de l'intron localisée dans la région P1. Cette proximité permet l'attaque de la liaison phosphodiester entre l'intron et l'exon. Il en résulte un groupement 3'-OH libre sur l'exon et l'*exoG* en 5' de l'intron. Le G terminal de l'intron échange sa position avec celle de l'*exoG* sur le site de

fixation ce qui aligne le 3'OH de l'exon avec P1 pour attaquer le 3' de l'intron dans la région P10, ce qui libère l'intron et permet la liaison des deux exons (Hedberg and Johansen, 2013).

Les introns de groupe II comportent six domaines conservés (I à VI) leur conférant une structure secondaire conservée. Le mécanisme d'épissage des introns du groupe II est similaire à celui des introns nucléaires. L'épissage est initié par une trans-estérification impliquant une attaque nucléophile sur le site d'épissage en 5' par le groupe 2'-OH d'un résidu adénosine dans le domaine VI. Il en résulte la formation d'un 3'-OH libre au site d'épissage en 5' de l'intron et d'un intermédiaire en forme de lasso. Dans une seconde étape de trans-estérification, les exons sont joints lorsque le 3'-OH libre au site d'épissage en 5' du second exon forme une liaison phosphodiester 5'-3' avec le nucléotide au site d'épissage en 3' de l'intron. L'intron est finalement sous forme de lasso. Une autre réaction d'épissage implique la "voie hydrolytique", où une molécule d'eau agit comme nucléophile dans la première étape et clive l'exon 5' de l'intron sans former une structure ramifiée (Daniels et al., 1996). La deuxième étape de l'épissage est identique à la réaction du « point ramifié ». Le groupe 3'-OH de l'exon en 5' attaque le phosphate au site d'épissage 3', épissant ainsi les exons et libérant l'intron sous forme linéaire. La présence de formes linéaires et circulaires de molécules d'intron excisées dans les organites végétales souligne toutefois la présence de mécanismes d'épissage alternatifs des introns de groupe II *in vivo* (Li-Pook-Than and Bonen, 2006).

L'édition des transcrits chloroplastiques

L'édition de l'ARN des organites de plantes consiste en des conversions de cytidines en uridines à des positions particulières et conservées dans l'évolution (sites d'édition). En comparaison avec le transcriptome mitochondrial, le transcriptome chloroplastique est peu édité avec seulement 34 sites d'édition de l'ARN (Chateigner-Boutin and Small, 2007). Le mécanisme d'édition est identique à celui retrouvé dans les mitochondries et sera décrit ci-dessous.

La traduction des transcrits chloroplastiques

La traduction de l'information génétique dans les plastes est réalisée par des ribosomes très similaires aux ribosomes bactériens 70S (Yamaguchi and Subramanian, 2000; Yamaguchi et al., 2000; Beligni et al., 2004; Manuell et al., 2007). Presque tous les composants de l'appareil de traduction des plastes ont des homologues proches dans les cyanobactéries, ce qui suggère que les fonctions de base ont été conservées tout au long de l'évolution. L'initiation de la traduction commence par la formation d'un complexe de pré-

initiation composé de la sous-unité ribosomale 30S et de l'ARN de transfert initiateur qui sélectionne le site d'initiation de la traduction dans l'ARNm. Le codon AUG et dans de rares cas, les triplets alternatifs GUG ou UUG, servent de codons d'initiation dans les ARNm plastidiaux (Sugiura et al., 1998). Le mécanisme moléculaire précis par lequel le complexe de pré-initiation du plaste reconnaît le codon initiateur n'est pas entièrement compris. Chez les bactéries, le mécanisme le plus important de l'initiation de la traduction est basé sur la liaison de la séquence anti-Shine-Dalgarno (localisée à l'extrémité 3' de l'ARNr 16S dans la sous-unité ribosomale 30S) à un motif de séquence riche en purines situé 7 ± 2 nucléotides en amont du codon initiateur dans l'ARNm. Ce motif est appelé « Shine-Dalgarno » (ou site de liaison au ribosome, RBS). Alors que certains gènes de plastes possèdent clairement des séquences proches des motifs Shine-Dalgarno de type bactérien et sont susceptibles de les utiliser pour l'initiation de la traduction (Ruf and Kössel, 1988; Drechsel and Bock, 2011), de nombreux autres gènes ne sont pas correctement positionnés par rapport à une séquence Shine-Dalgarno putative. Une étude plus récente a montré que, dans les chloroplastes et les bactéries, les régions d'initiation de la traduction dépourvues de motifs Shine-Dalgarno sont moins structurées que celles qui abritent une séquence canonique Shine-Dalgarno (Scharff et al., 2011). L'absence de formation de structure secondaire autour du codon d'initiation a été proposée comme facilitant la reconnaissance du site et la liaison de la sous-unité ribosomale 30S.

Après la sélection du codon de départ, le recrutement de la sous-unité 50S convertit le complexe de pré-initiation en un complexe d'initiation actif, qui peut ensuite entrer dans la phase d'allongement de la synthèse des protéines. De la même manière que les bactéries, le processus est susceptible d'être assisté par l'initiation de la traduction et les facteurs d'allongement. Des homologues chloroplastiques des facteurs d'initiation et d'allongement bactériens ont été identifiés et, en partie, caractérisés (Beligni et al., 2004; Miura et al., 2007; Ruppel and Hangarter, 2007). Pendant l'élongation, le ribosome glisse le long de la séquence codante de l'ARNm incorporant les acides aminés correspondant à chaque codon dans la chaîne polypeptidique en croissance à l'aide des facteurs d'allongement de la traduction EF-Tu, EF-G et EF-Ts. Les plastes utilisent le code génétique standard, mais il faut garder à l'esprit que des écarts de nucléotides individuels entre la séquence d'ADN et la séquence d'ARNm sont causés par l'édition de l'ARN (Bock, 2000). La traduction est terminée lorsque le ribosome atteint l'un des trois codons de terminaison standard (UAA, UAG ou UGA). Le processus de terminaison nécessite l'assistance de facteurs de libération du ribosome (Meurer et al., 2002). Pour préparer le ribosome pour un nouveau cycle de synthèse protéique, le

complexe ARNm-ribosome est ensuite dissocié par le facteur de recyclage des ribosomes (Brown et al., 2014). RRF dissocie le ribosome 70S en ses sous-unités 30S et 50S, libérant ainsi l'ARNm et rendant la sous-unité 30S libre disponible pour l'initiation de traduction *de novo* (Tiller and Bock, 2014).

L'expression génétique mitochondriale

La transcription mitochondriale

Le génome mitochondrial des plantes est transcrit par des ARN polymerases de type phagiques « NEP » décrites précédemment. La majorité des gènes mitochondriaux sont transcrits par RpoTm (Kühn et al., 2007), alors que RpoTmp semble être spécifique de certains gènes (Kühn et al., 2009). Les promoteurs reconnus par RpoTm sont distribués sur le génome mitochondrial de manière homogène possiblement afin de maintenir l'activité transcriptionnelle malgré des recombinaisons fréquentes du génome mitochondrial. Ainsi le génome mitochondrial est transcrit dans la plupart des cas sous forme polycistronique. Un même gène peut être transcrit par différents promoteurs et beaucoup de régions non fonctionnelles sont transcrites. La transcription mitochondriale apparaît donc comme un processus peu régulé, le contrôle du taux de transcrits à l'équilibre s'effectuant majoritairement au niveau post-transcriptionnel (Giegé et al., 2000; Holec et al., 2006).

Maturation des extrémités 5' et 3' des transcrits mitochondriaux

Contrairement au chloroplaste où les extrémités 5' et 3' des ARNm sont toutes deux bien définies par des mécanismes impliquant des protéines PPR (voir ci-dessus), seules les extrémités 3' des ARNm mitochondriaux semblent bien définies. Leurs extrémités 5' apparaissent comme étant beaucoup plus hétérogènes (Hammani and Giegé, 2014). La définition des extrémités 3' semble utiliser un mécanisme similaire à celui décrit pour PPR10 dans le chloroplaste. Par exemple, dans la mitochondrie d'*Arabidopsis*, MTSF1, une protéine PPR stabilise l'extrémité 3' de l'ARNm de *nad4* en bloquant la dégradation de 3' vers 5' (Haïli et al., 2013). Par ailleurs, la maturation des ARNm mitochondriaux implique des coupures endonucléolytiques qui sont dans certains cas réalisées par les enzymes de maturation des ARNt comme la RNase P et la RNase Z au niveau de structures ARNt-like aussi appelées « t-elements » (Forner et al., 2007; Gobert et al., 2010). Par exemple, chez *Arabidopsis*, la maturation de l'extrémité 3' de *nad6* implique un clivage par la RNase P protéique AtPRORP1 (Gutmann et al., 2012).

Epissage des transcrits mitochondriaux

Comme chez le chloroplaste, le génome mitochondrial possède principalement des introns du groupe II et quelques introns du groupe I. L'épissage en *cis* et en *trans* implique aussi l'activité de différentes maturases et de protéines PPR (voir ci-dessus) (Brown et al., 2014). On notera que *matR*, un gène codant pour une maturase a été retenu dans le génome mitochondrial des Brassicaceae, où il est présent dans le quatrième intron de *nad1*. *MatR* est impliqué dans l'épissage de différents introns mitochondriaux dont celui de *nad1* (Sultan et al., 2016).

L'édition des transcrits mitochondriaux

Avec 400 à 500 sites d'édition, le transcriptome mitochondrial est le plus édité (Unsold et al., 1997; Giegé and Brennicke, 1999). L'édition de l'ARN modifie sa séquence en désaminant des cytidines pour obtenir des uridines. La réaction inverse est également possible mais plus rare. Elle est surtout observée chez les fougères et les mousses et n'est pas observée chez les plantes dicotylédones comme *Arabidopsis* (Gualberto et al., 1989; Pring et al., 1993; Covello and Gray, 1989; Gualberto et al., 1990). Contrairement aux autres processus de maturation post-transcriptionnelle, l'édition de l'ARN modifie l'information codée par les gènes des organites au niveau des transcrits. En conséquence, elle constitue une étape essentielle pour la production de protéines fonctionnelles. En particulier, l'édition de l'ARN restaure des codons conservés dans l'évolution et codant pour des acides aminés essentiels à la fonction des protéines des organites (Giegé and Brennicke, 1999). Alors que les sites d'édition de l'ARN se trouvent principalement dans les régions codantes, certains événements se produisent aussi dans des régions non traduites, comme dans les ARNt et les introns, mais pas dans les ARNr.

Un nombre croissant de facteurs impliqués dans l'édition de l'ARN ont été identifiés. Ils forment probablement un « éditosome », i.e. un complexe fonctionnel d'édition de l'ARN. Comme pour les autres processus d'expression génétique des organites, les protéines PPR jouent un rôle primordial pour l'édition de l'ARN (Kotera et al., 2005). En particulier elles servent pour la reconnaissance spécifique de la séquence adjacente des sites d'édition. Les protéines PPR de la sous-famille DYW (voir ci-dessous) semblent particulièrement importantes dans la mesure où le domaine DYW ressemble à une cytidine déaminase (Salone et al., 2007). Il est donc très probable qu'il porte l'activité catalytique proprement dite de l'édition de l'ARN, mais cela n'a pas encore été démontré expérimentalement. Outre les protéines PPR, d'autres protéines apparaissent comme essentielles à l'édition, comme les

protéines RIP/MORF qui interagissent avec les protéines PPR et d'autres protéines de liaison à l'ARN comme les protéines ORRM ou des protéines OZ à doigt de zinc (pour revue : Sun et al., 2016). Bien que les fonctions de toutes ces protéines formant l'éditosome ne soient pas clairement identifiées, une étude récente suggère que les protéines MORF modifient la structure des protéines PPR pour augmenter leur affinité pour les ARN (Yan et al., 2017).

La traduction des transcrits mitochondriaux

Les ribosomes mitochondriaux (mitoribosomes) ont une origine bactérienne et contiennent de nombreux éléments de structure et de mode d'action hérités des ribosomes procaryotes. Cependant, les mitoribosomes ont aussi considérablement divergé et adopté des caractéristiques spécifiques dans les différentes lignées eucaryotes (Kehrein et al., 2015).

Les mitoribosomes de plantes divergent considérablement des ribosomes bactériens et possèdent une composition spécifique. Les mitoribosomes de plantes supérieures sédimentent à 77-78S et sont donc plus grands que les ribosomes bactériens (Leaver and Harme, 1976). Ils sont composés de trois molécules d'ARNr 26S, 18S et 5S, tous codés dans le génome mitochondrial. Les ARNr 26S et 5S sont présents dans la grande sous unité et l'ARNr 18S est présent dans la petite sous unité ribosomale.

La composition protéique des mitoribosomes de plantes supérieures n'a pas encore été déterminée. Néanmoins, des études anciennes décrivant l'analyse électrophorétique des mitoribosomes de pomme de terre et de haricots ont permis de déterminer qu'ils contiennent environ 68 à 80 protéines (Pinel et al., 1986; Maffey et al., 1997). Les mitoribosomes végétaux contiennent donc plus de protéines que les ribosomes bactériens. C'est une particularité partagée avec les mitoribosomes de mammifères, de levures et de protozoaires bien que ceux-ci soient de plus petites tailles avec des ARNr plus compacts (Ahmed et al., 2016; Woolford and Baserga, 2013; Estrozi et al., 2011). Dans tous ces organismes, on connaît mal le rôle des protéines supplémentaires acquises lors de l'évolution eucaryote (Agrawal and Sharma, 2012). Il est intéressant de noter que deux de ces protéines additionnelles chez l'Homme sont des protéines PPR : MRPS27 et MRPS39 (Greber et al., 2015) dont les fonctions restent inconnues à ce jour, mais dont la position dans la structure suggère qu'ils pourraient participer au recrutement des ARNm sur les ribosomes mitochondriaux.

Contrairement aux chloroplastes où certains ARNm possèdent une séquence de type Shine-Dalgarno (AGGAGG) qui serait impliquée dans le recrutement des ribosomes, aucun

élément de séquence ou de structure pouvant participer au recrutement des mitoribosomes ne semble conservé dans les transcrits de mitochondries de plante (Hazle and Bonen, 2007). De plus, certains ARNm dépourvus de codon de terminaison canonique sont traduits en protéines fonctionnelles (Raczynska et al., 2006). Ces caractéristiques uniques ainsi que des preuves génétiques soutiennent l'idée que des facteurs en *trans* interviennent pour aider le ribosome à traduire les gènes mitochondriaux chez les plantes. Par exemple, la protéine PPR MPPR6 (protéine PPR de maïs 6) se lie à la région 5' non traduite du transcrit de Rps3 (protéine ribosomale S3) chez le maïs et facilite sa traduction (Manavski et al., 2012). Chez *Arabidopsis*, deux protéines PPR supplémentaires, PPR336 et PNM1 (protéine PPR localisée dans le noyau et dans les mitochondries) pourraient faire partie du mitoribosome ou être impliquées dans le contrôle de la traduction, comme suggéré par leur association avec les polysomes mitochondriaux (Uyttewaal et al., 2008; Hammani et al., 2011). L'importance des protéines PPR dans le contrôle de la traduction mitochondriale est encore soulignée par leur rôle de restauration de la fertilité dans les plantes à fleurs. Le réarrangement dynamique du génome mitochondrial des plantes est responsable de l'expression d'ORF chimériques qui provoque un dysfonctionnement mitochondrial conduisant à l'avortement du pollen (Gaborieau et al., 2016). L'expression de ces ORF toxiques est souvent réprimée par des gènes nucléaires appelés restaurateurs de fertilité (Rfs) qui, dans la majorité des cas, codent des protéines PPR (Cabello-Villegas et al., 2002). Par exemple, PPR-B inhibe la traduction de l'une de ces ORF dans les mitochondries de radis (Uyttewaal et al., 2008).

2. Les protéines PPR

a. Structure et distribution des protéines PPR

La plupart des mécanismes d'expression génétiques mitochondriaux et chloroplastiques décrits ci-dessus impliquent des protéines à « pentatricopeptide repeat » (PPR).

Les protéines PPR sont des membres de la superfamille des protéines à alpha-solénoïde composées de motifs répétés hélicoïdaux (Hammani et al., 2014). Les protéines PPR sont des protéines de liaison à l'ARN. Elles sont caractérisées par un assemblage de 2 à environ 30 motifs de 35 acides aminés de séquence dégénérée, répétés en tandem (Small and Peeters, 2000; Lurin et al., 2004). La détermination de la structure cristallographique de protéines PPR a montré que chaque motif PPR est replié en une structure conservée en hélice-boucle-hélice, confirmant les prédictions initiales (Small and Peeters, 2000). Les structures

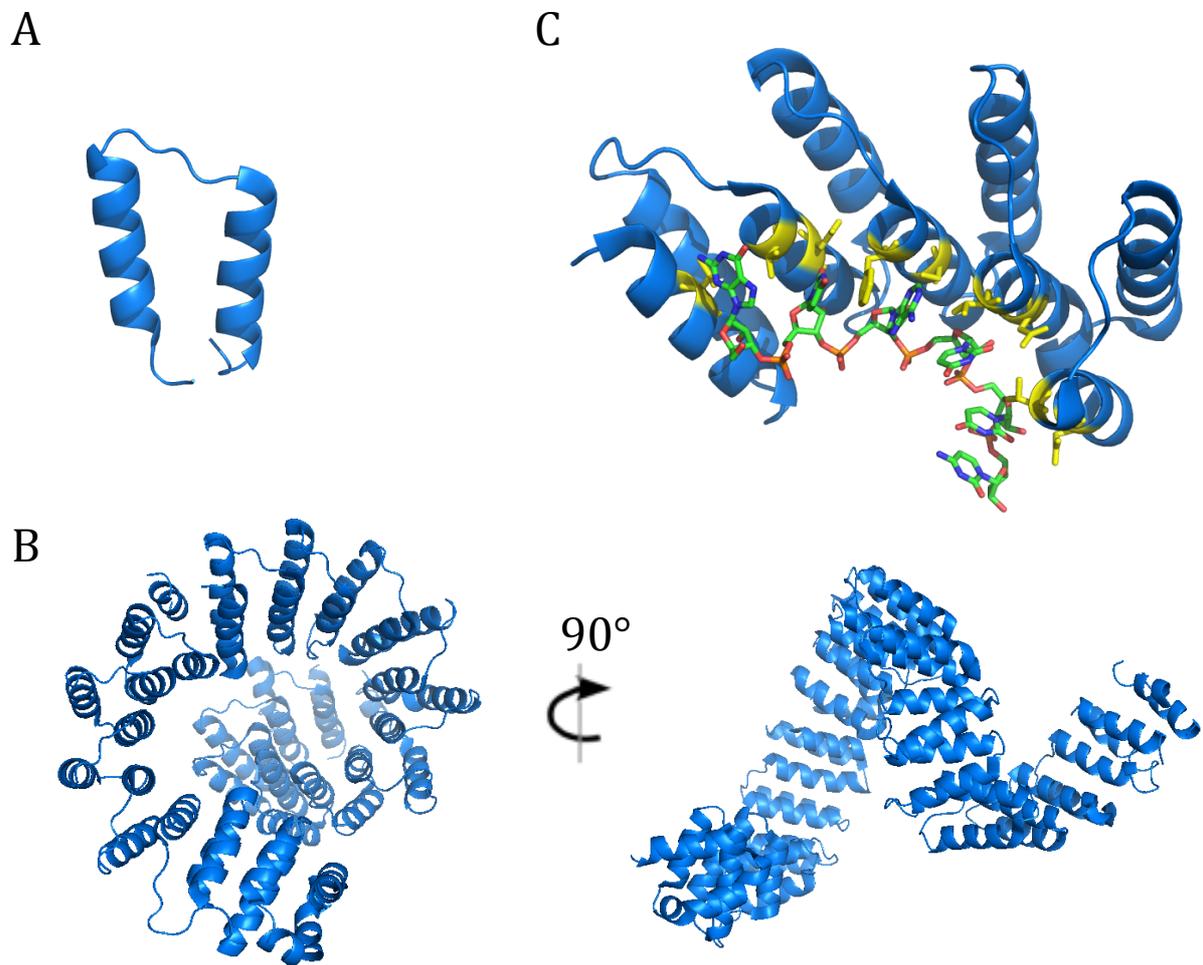


Figure 6 : Structure des motifs PPR et interaction avec l'ARN

Chaque motif PPR est composé d'un enchainement d'une hélice alpha d'une boucle et d'une seconde hélice alpha (A). L'enchainement des motifs PPR forme un solénoïde en rotation dans le sens horaire (B). L'ARN se lie à la surface interne du solénoïde par les positions 5 et 35 de chaque motif PPR ici en jaune (C).

obtenues par cristallographie ou en solution des protéines PPR (e.g. par small angle X-ray scattering, SAXS) révèlent des répétitions en tandem de motifs qui s'assemblent pour former un solénoïde en rotation dans le sens horaire. Les résidus impliqués dans la liaison à l'ARN sont situés sur la surface interne de la super-hélice (Yin et al., 2013) (Figure 6). Bien que la dimérisation des protéines PPR ait été observée dans les structures cristallines de PPR10 (Yin et al., 2013), d'autres études en solution suggèrent que le complexe PPR-ARN implique une forme monomérique de la protéine et que la forme solénoïdale globale est préservée lors de l'interaction avec l'ARN (Li et al., 2014; Gully et al., 2015).

Cette famille de gènes se trouve chez tous les eucaryotes mais les gènes y sont trouvés en nombres très variables. Alors que la famille des gènes codant pour des protéines PPR est représentée par quelques membres chez les animaux et les levures, elle a connu une expansion considérable chez les plantes terrestres, avec près d'un millier de protéines codées chez certaines plantes terrestres telles que les mousses e.g. (*Selaginella moellendorffii*) (Banks et al., 2011) ou certains conifères (Nystedt et al., 2013). Chez les angiospermes (les plantes à fleurs), la famille PPR est l'une des plus grandes familles de gènes dans presque tous les génomes séquencés jusqu'à présent, avec 1 à 2 % du génome codant pour des protéines PPR. Chez *Arabidopsis thaliana* et *Oryza sativa*, 450 et 477 gènes PPR ont été trouvés (Lurin et al., 2004; O'Toole et al., 2008). Les protéines PPR sont codées dans le génome nucléaire, mais les protéines traduites sont presque exclusivement adressées aux plastes et aux mitochondries (Colcombet et al., 2013). L'analyse fonctionnelle d'un nombre rapidement croissant de protéines PPR révèle des fonctions reliées aux mécanismes d'expression génétique et plus particulièrement aux processus post-transcriptionnels (Schmitz-Linneweber and Small, 2008; Nakamura et al., 2012; Barkan and Small, 2014). Cette spécificité d'action est due à leur capacité à reconnaître les séquences d'ARN spécifiques (Barkan et al., 2012; Nakamura et al., 2012; Takenaka et al., 2013; Yagi et al., 2013), chaque protéine ayant une séquence cible spécifique, ou un petit nombre de séquences cibles présentant des similarités de séquences.

b. Les différentes sous-familles de protéines PPR

Les protéines PPR se divisent en deux sous-familles, P et PLS. La sous famille P est caractérisée par la répétition du motif canonique P de 35 acides alors que les protéines de la famille PLS sont composées de la combinaison de différents variants du motif P.

Sous-famille P



Sous-famille PLS

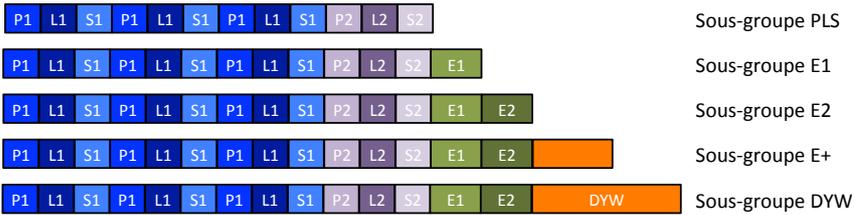


Figure 7 : Structure des différentes sous familles de protéines PPR.

Il existe deux sous familles principales de protéines PPR. La première est la sous famille P comportant uniquement de motifs PPR « purs » (P) de 35 acides aminés et parfois des domaines additionnels (ici en jaune). La seconde est la sous famille PLS composée d'enchaînements de motifs purs (P), longs (L) et court (S). Deux sous-types de motifs P, L et S peuvent être distingués. La famille PLS se caractérise également par la présence de domaines additionnels en C-terminal des motifs PPR. Il s'agit des motifs E1, E2, E+ et DYW impliqués dans des fonctions telles que l'édition de l'ARN.

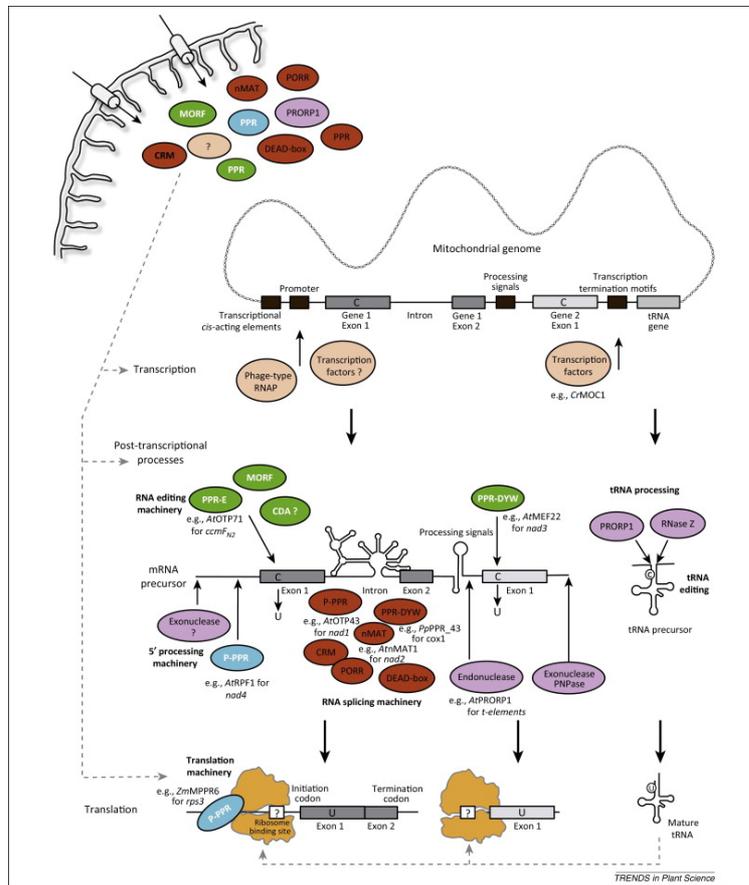


Figure 8 : Les différentes étapes de maturation des ARN mitochondriaux de plantes.

Les ARN mitochondriaux de plantes sont transcrits sous forme polycistronique. Des modifications post-transcriptionnelles sont nécessaires afin d'obtenir des transcrits matures : les extrémités 5' et 3' des ARNm précurseurs sont clivés, certains sites sont édités, les introns sont épissés et les régions inter-cistroniques sont clivées. Toutes ces étapes sont réalisées par des facteurs codés dans le noyau et importés dans la mitochondrie. Certains de ces facteurs sont représentés dans ce diagramme. Les « ? » indiquent des facteurs encore inconnus. Abréviations : At : *Arabidopsis thaliana*, Pp : *Physcomitrella patens*, Zm : *Zea mays*. On notera le rôle prépondérant des protéines PPR dans les différentes étapes de l'expression génétique mitochondriale. Figure issue de Hammani & Giegé (2014).

L'étude initiale avait permis d'identifier des variants de taille avec des motifs plus longs (L) et plus courts (S). Une étude récente a permis de préciser d'autres variations entre ces motifs. Les motifs P1 et P2 diffèrent du motif P au niveau de la première hélice. Les motifs L1 et L2 (« long PPR 1 and 2 ») sont des variants longs de 35 et 36 acides aminés qui diffèrent l'un de l'autre dans la seconde hélice. Les motifs S1 et S2 (« short PPR 1 and 2 ») sont des variants courts de 32 aa (Cheng et al., 2016). Les motifs P1 et P2 se trouvent uniquement dans la sous famille PLS qui possède un certain nombre de triplets P1-L1-S1 suivis d'un triplet P2-L2-S2.

La sous famille PLS se divise en plusieurs sous-groupes selon les domaines présents à l'extrémité C-terminale. Le sous-groupe « PLS » ne comporte pas de domaine C-terminal additionnel. Les sous-groupes E1, E2, E+ et DYW sont caractérisés par l'addition successive de domaines supplémentaires (Figure 7). Les domaines E1 et E2 sont des motifs PPR-like de 34 acides aminés. Le domaine DYW est un domaine putatif de déamination des cytidines trouvé dans les protéines PPR ayant une activité d'édition de l'ARN. Toutefois, toutes les protéines PPR impliquées dans l'édition de l'ARN ne sont pas des protéines à domaine DYW.

c. Description des fonctions identifiées

De manière générale, la fonction des protéines PPR est le plus souvent associée aux mécanismes post-transcriptionnels dans les organites. Il a été démontré qu'elles sont impliquées dans la transcription (chez l'homme et la levure), la stabilité de l'ARN (Pfalz et al., 2009) et sa dégradation, le clivage endo-nucléolytique de l'ARN, e.g. pour l'activité RNase P (voir ci-dessous) (Jonietz et al., 2010; Schelcher et al., 2016) ainsi que le contrôle de la traduction. Chez les plantes, les protéines PPR sont en outre impliquées dans l'édition de l'ARN (Kotera et al., 2005) ainsi que l'épissage des introns (de Longevialle et al., 2007). De manière générale, il a été montré dans la plupart des cas, que ces fonctions impliquent la reconnaissance spécifique d'éléments ARN *in cis*. Par exemple, pour l'édition de l'ARN, des séquences spécifiques présentes à proximité des sites d'édition de l'ARN sont reconnues par les protéines PPR. De même, l'extrémité des transcrits des organites est définie par la fixation spécifique de protéines PPR dans les régions inter-géniques, servant ainsi de barrière à la dégradation de ces transcrits par des exo-ribonucléases (Schmitz-Linneweber and Small, 2008) (Figure 8).

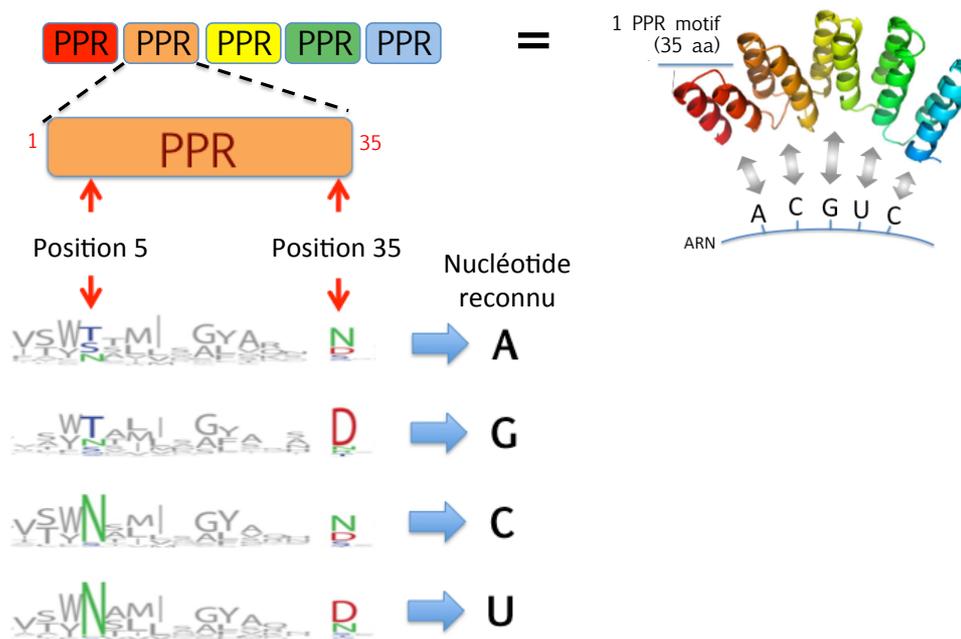


Figure 9 : Mode de reconnaissance des ARN par les motifs PPR définissant un "code PPR".

Chaque motif PPR individuel de 35 acides aminés possède une structure en hélice boucle hélice qui va permettre la reconnaissance spécifique d'un nucléotide grâce aux positions 5 et 35 du motif. Le diagramme du bas illustre que la nature des acides aminés aux positions 5 et 35 induit la reconnaissance d'un nucléotide spécifique, définissant un « code PPR ». Par exemple, un motif ayant deux N aux positions 5 et 35 reconnaitra spécifiquement un C. Figure dérivée de Barkan et al. 2012.

d. Mécanisme de reconnaissance des nucléotides par les motifs PPR, définition du «code PPR»

Les protéines PPR ont la capacité de se lier à une séquence spécifique de l'ARN grâce aux motifs PPR qui reconnaissent chacun un nucléotide spécifique. L'ARN peut se positionner à l'intérieur de la super-hélice formée par l'empilement des hélices α . Les bases se retrouvent alors intercalées entre deux acides aminés à des positions précises permettant la création de ponts hydrogènes entre les bases et certains résidus. Les acides aminés spécifiques présents aux positions requises pour la reconnaissance spécifique des bases sont trouvés aux positions 5 et 35 des motifs PPR (Cheng et al., 2016). Ces mêmes positions ont précédemment été appelées « 4 et 34 » ainsi que « 6 et 1' » par d'autres auteurs. La nature des acides aminés à ces positions impose la spécificité pour la base reconnue par chaque motif (Figure 9). Cette correspondance permet de définir un code de reconnaissance reliant les motifs PPR à leurs substrats ARN (le « code PPR »). Par exemple, la présence de deux asparagines aux positions 5 et 35 d'un motif PPR, lui confère une spécificité pour les cytidines (Barkan et al., 2012; Yin et al., 2013). Ce code a été défini pour des protéines PPR reconnaissant des motifs linéaires d'ARN simple brin. Bien que ce code permette de prédire la cible de nombreuses protéines PPR, il ne permet pas de prédire toutes les cibles, en particulier, celles présentes dans les ARN structurés. Il est donc fort probable que le code PPR connu sous sa forme actuelle soit encore incomplet. Des études en cours visent à mieux appréhender la diversité de mode d'action des protéines PPR.

3. Biogénèse et maturation des ARNt chez les plantes

a. Répartition des gènes d'ARNt et des TLS chez les plantes

Parmi les 658 gènes d'ARNt chez *Arabidopsis*, 598 sont nucléaires, 23 mitochondriaux et 37 chloroplastiques (chiffres issus de "The plantRNA database" (<http://plantrna.ibmp.cnrs.fr/>) (Cognat et al., 2013). De nombreux ARNt doivent être importés dans les mitochondries pour assurer la traduction.

Il existe également d'autres structures dérivant des ARNt comme les structures « ARNt-like » (TLS) notamment dans les génomes d'ARN de virus de plantes ou les « t-element » que l'on retrouve principalement dans des transcrits d'ARNm codés par le génome mitochondrial. Les virus utilisent les TLS de leurs génomes pour se répliquer. Certains TLS comme celui du TYMV, peuvent être amino-acylés. Dans l'ADN mitochondrial du blé, les

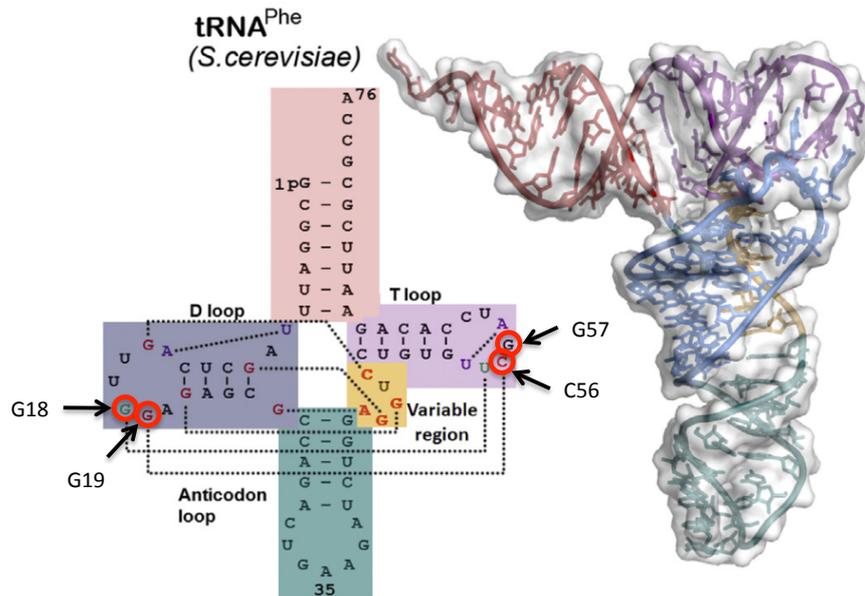


Figure 10 : Structure 2D et 3D des ARNt.

Structure 2D et structure 3D expérimentale de l'ARNt^{Phe} de *S. cerevisiae*. L'ARNt est composé d'un bras accepteur en rose, du domaine D en bleu, du domaine de l'anti-codon en vert, d'une boucle variable en jaune et du domaine T en violet. Les boucles des domaines T et D s'empilent pour former une structure 3D en L. Les pointillés représentent les interactions tertiaires qui stabilisent l'ARNt mature et certains nucléotides conservés dans l'ensemble des ARNt de structure canonique (étudiés lors de ce travail de thèse) sont mis en évidence en rouge. Figure dérivée de Fernandez-Millan et al. (2016).

éléments en ARNt-like sont situés à proximité de gènes actifs (18S-5S, 26S, ARNt^{Pro}) et dans le même sens de transcription. Ces éléments peuvent être traités comme des ARNt et clivés par les enzymes de maturation des ARNt, i.e. la RNase P et la RNase Z ce qui permet de libérer les différents ARN des polycistrons.

Les « short interspersed nuclear elements » (SINE ou éléments nucléaires intercalés courts) sont des éléments mobiles non autonomes qui se trouvent dans une grande variété de génomes eucaryotes et sont transcrits par l'ARN polymérase III. La famille de SINE la plus connue est la famille Alu (Kramerov and Vassetzky, 2005). Les retroéléments SINE utilisent la machine enzymatique des « autonomous long interspersed elements » (LINE) pour leur propagation par retrotransposition. Un modèle général de rétroposition SINE et LINE a été proposé par Sun et al. (Sun et al., 2007) est basé sur des régions 3' communes ou des queues poly (A). Bien que la plupart des SINE soient dérivés de manière ancestrale des gènes d'ARNt (ou dans de rares cas des ARN 7S ou 5S), la structure typique en feuille de trèfle de l'ARNt n'est semble-t-il pas entièrement conservée. Les SINE dérivés des ARNt ont une structure composite comprenant une partie 5' dérivant des ARNt suivie d'une région non apparentée aux ARNt. Dans tous les cas, les boîtes A et B correspondant aux promoteurs internes des ARNt sont présentes dans la partie dérivant des gènes d'ARNt. Dans les espèces *A. thaliana* et *Brassica*, diverses familles de SINE liées à l'ARNt, appelées RathE1, RathE2, RathE3, S1 et S2 ont été identifiées (Pélissier et al., 2004; Michaud et al., 2011).

b. Structure et fonction des ARNt

Les ARNt sont des ARN non codants possédant un repliement secondaire et une structure tertiaire précise. Ils sont composés de 70 à 100 nucléotides en général qui se replient pour former le bras accepteur, le domaine D, le domaine T, la région variable et le domaine de l'anticodon (Giege et al., 2012). Les ARNt ont pour fonction de faire le lien entre le code génétique et les acides aminés correspondants. L'ARNt est chargé spécifiquement avec l'acide aminé correspondant à son anticodon qui correspond lui-même au codon présent sur l'ARNm. Il apporte l'acide aminé au ribosome et sert ainsi d'adaptateur entre l'ARNm et la protéine. La structure secondaire des ARNt en feuille de trèfle permet un repliement tertiaire en L. (Moras et al., 1980) (Figure 10). Le côté long du L portant à son extrémité le domaine anticodon s'insère dans le ribosome et présente l'acide aminé face à la chaîne protéique en formation afin de réaliser la liaison peptidique (Simonović and Steitz, 2009)

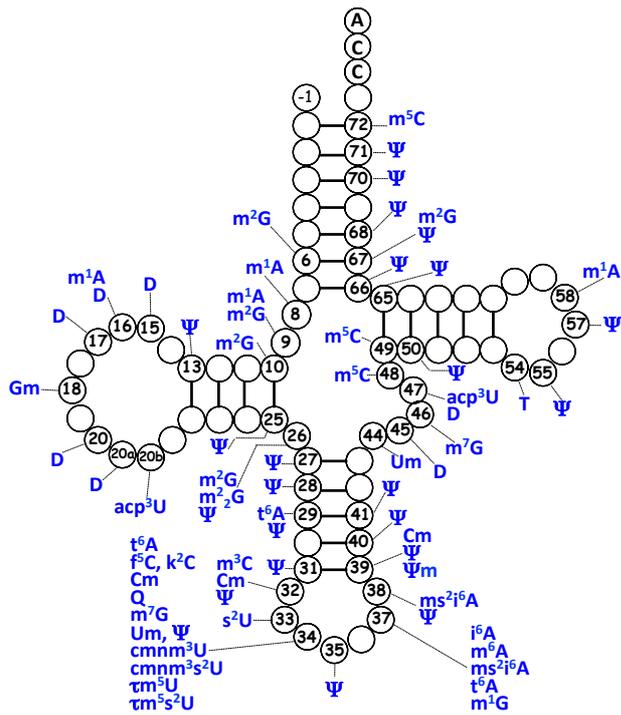


Figure 11 : Modification des ARNt.

Représentation non exhaustive des modifications de bases les plus courantes aux différentes positions des ARNt. On notera en l'occurrence la prépondérance des bases méthylées (m) ou pseudouridylées (Y). Figure issue de Fernandez-Millan et al. (2016).

c. Biogenèse des ARNt fonctionnels

Les ARNt nucléaires

Transcription des ARNt nucléaires

Les ARNt nucléaires sont transcrits par l'ARN polymérase III. Leurs séquences promotrices contiennent une boîte A et une boîte B où se fixe un complexe protéique recrutant la polymérase. La transcription des gènes d'ARNt est initiée au niveau d'une boîte CAA localisée 3 à 8 paires de bases en amont du gène codant l'ARNt. Cette extension 5' de l'ARNt doit être maturée. La terminaison de la transcription s'effectue par l'intermédiaire d'une séquence, transcrite en poly U sur l'ARNt (Michaud et al., 2011).

Maturation des ARNt nucléaires

Les premières étapes de maturation consistent à cliver les séquences additionnelles en 5' et en 3' des ARNt précurseurs. Les séquences en 5' et 3' sont respectivement maturées par deux endonucléases, la RNase P et la RNase Z (Rossmannith, 2012). La RNase P peut être une ribonucléoprotéine, une protéine seule ou un complexe de protéines. Les étapes de maturation varient suivant les ARNt, elles impliquent la modification de nombreuses bases et parfois l'épissage de l'ARN (Salinas-Giegé et al., 2015). On notera que les modifications de nucléotides sont particulièrement représentatives des ARNt, avec un grand nombre de nucléotides non-canoniques, ou de nucléotides modifiés dans leur structure. Ces modifications de nucléotide, e.g. des pseudo-uridylation, méthylations, isomérisations, thiolations ou réductions sont incorporées post-transcriptionnellement par des enzymes spécialisées. Certaines modifications participent à l'établissement de la structure 3D des ARNt ou à sa stabilisation comme e.g. la pseudo-uridylation dans la boucle T. D'autres modifications en particulier dans la boucle de l'anticodon participent directement à l'interaction avec les ARNm et à la lecture des codons (Figure 11).

La dernière étape de la maturation des ARNt est l'ajout d'un groupement CCA du côté 3' de l'ARNt. C'est la ARNt nucleotidyl-transférase aussi appelée « CCCase » qui effectue cet ajout afin de pouvoir charger l'ARNt avec son acide aminé *via* ce groupement par une estérification (Hou, 2010). L'ajout du groupement CCA joue aussi le rôle de contrôle qualité, les ARNt non maturés sont dégradés.

Une fois matures, les ARNt fonctionnels sont chargés avec des acides aminés par des aminoacyl ARNt synthétases spécifiques, pour revue (Pang et al., 2014) et peuvent être utilisés par la machinerie traductionnelle.

Les ARNt mitochondriaux

Transcription des ARNt mitochondriaux

Les gènes mitochondriaux sont transcrits par deux ARN polymerases de type NEP à partir de nombreux promoteurs souvent caractérisés par une séquence consensus CRTA (Fey and Maréchal-Drouard, 1999; Kühn et al., 2005). De nombreux promoteurs sont dispersés le long du génome, et il est fréquent de trouver des promoteurs multiples pour un seul gène. Les ARNt sont souvent codés au sein de polycistrons.

Maturation des ARNt mitochondriaux

La première étape de maturation des ARNt est leur libération du transcrit primaire polycistronique (Michaud et al., 2011). Après cela, de nombreuses modifications post-transcriptionnelles sont utilisées pour former la structure secondaire nécessaire pour obtenir un ARNt stable et fonctionnel (Helm et al., 1999). Il existe une grande diversité chimique dans les modifications de bases apportées aux ARNt, avec des méthylations, isomérisations, thiolations, formylations et ribosylations, ainsi que d'autres, par une multitude de facteurs nucléaires. On notera que certaines modifications comme l'ajout de 5-taurinométhyl sont spécifiques des mitochondries (Salinas-Giegé et al., 2015). Les modifications des ARNt peuvent, là aussi, être largement divisées en deux catégories : celles qui confèrent une stabilité structurale et un repliement correct et celles qui assurent la spécificité fonctionnelle des ARNt en affectant leur interaction avec d'autres facteurs.

Alors que de nombreux ARNt canoniques non modifiés se replient dans des structures de conformation correcte (Sampson and Uhlenbeck, 1988; Harrington et al., 1993), il a été démontré que les transcrits *in vitro* d'ARNt mitochondriaux non modifiés peuvent souvent se replier dans une gamme de structures secondaires non fonctionnelles (Helm et al., 1998). Les modifications vers le centre de l'ARNt jouent généralement un rôle dans sa structure et sa stabilité, souvent en modulant la rigidité de certains domaines. Par exemple, la saturation du cycle pyrimidine de l'uracile pour former de la di-hydro uridine entraîne une plus grande flexibilité conformationnelle (Dalluge et al., 1996), tandis que l'introduction de la pseudo uridine, par l'isomérisation de l'uridine, agit pour augmenter la rigidité structurale de l'ARNt (Davis, 1995). La méthylation de bases peut également influencer la structure de l'ARNt de plusieurs façons. Par exemple, la méthylation de l'adénine en position 9 (A9) dans l'ARNt Lys mitochondrial humain par le sous-complexe RNase P de MRPP1 et MRPP2 (Vilardo et al., 2012) a une incidence sur la structure de l'ARNt. L'ARNt Lys mitochondrial non modifié

se plie dans une structure en épingle à cheveux non fonctionnelle, qui n'est pas reconnue par son aminoacyl ARNt synthétase (Helm et al., 1998; Sissler et al., 2004). L'introduction de m1A9 seule dans l'ARNt Lys mitochondrial est suffisante pour favoriser la formation de la structure en trèfle en perturbant la paire de bases A-U64 Watson-Crick dans l'épingle (Helm et al., 1999). Cependant, les effets stabilisants de la N1-méthylation de l'adénine ont également été identifiés, sur la boucle T à la position 58, catalysée dans les mitochondries humaines par la tRNA methyltransferase TRMT61B (Chujo and Suzuki, 2012).

Les modifications apportées aux positions 34 et 37 des résidus d'ARNt sont importantes pour maintenir l'exactitude et la fidélité de la traduction et représentent la plus grande diversité chimique trouvée dans le profil de modification des ARNt. Un seul ou ces deux résidus sont modifiés dans presque tous les ARNt étudiés jusqu'à présent. La traduction d'un ARNm dépend des interactions précises entre les trois bases des codons (appelées 1, 2 et 3) et les triplets de l'anticodon de l'ARNt correspondant (aux positions 36, 35 et 34). Cependant, en raison de la dégénérescence du code génétique, plusieurs codons peuvent être reconnus par un seul ARNt. Les codons dégénérés contiennent des résidus identiques aux positions 1 et 2 et se différencient par la variabilité de la position 3. Pour ce faire, les interactions entre les résidus 3 et 34 ne sont pas canoniques, permettant une gamme beaucoup plus grande de paires de base possibles, une caractéristique appelée « wobble ». La position 34, ou la base wobble, est souvent occupée par une uridine, capable de s'associer avec l'une des quatre bases en raison d'une flexibilité conformationnelle accrue dans la boucle de l'anticodon. Ce scénario est suffisant pour la majorité des codons dans lesquels le résidu en position 3 est entièrement dégénéré. Cependant, dans un certain nombre de cas, la présence d'une purine ou d'une pyrimidine en position 3 spécifie des codons pour des acides aminés différents. L'augmentation de la discrimination par la base wobble requise pour un décodage précis est obtenue grâce à sa modification post-transcriptionnelle. Par exemple, la formation de la modification τ m5 (s2) U par l'action concertée de la protéine de liaison GTP 3, GTPBP3, (Villarroya et al., 2008), MTO1 (optimisation de traduction d'ARNt mitochondrial 1) (Li and Guan, 2002) et MTU1 (mitochondrial Tari-thiouridylase spécifique de l'ARNt, TRMU), favorise grandement l'appariement des bases avec les purines et empêche la lecture incorrecte des codons (Yokoyama et al., 1985) comme cela a été démontré pour l'ARNt mitochondrial Leu humain (UUR) (Kirino et al., 2004). La reconnaissance des purines à la position 3 est également illustrée par la formylation (substitution d'un groupement par un -CH=O) à la position C34 dans l'ARNt mitochondrial Met, la f5C34 résultante permet à

l'ARNt de lire les codons AUA et AUG (Bilbille et al., 2011). Alors que la sélection rigoureuse de l'ARNt correct est la clé pour assurer la précision, une interaction codon-anticodon stable est essentielle pour l'efficacité de la traduction. Pour cette raison, les ARNt avec anticodons portant un U ou un A en position 36 nécessitent souvent une modification à la position adjacente 37 (Lamichhane et al., 2013). Par exemple, dans la position 37 des mitochondries de mammifère, on retrouve des modifications N6-thréonylcarbamoyladénosine (t6A37) et N6-isopentényladénosine (i6A37) catalysées par TRIT1 (ARNt isopentényltransférase 1) dans des cellules humaines (Yarham et al., 2014) et modifiées ultérieurement par CDK5RAP1 (kinase dépendante de la cycline 5 protéine 1 associée à la sous-unité régulatrice) pour former un groupement ms2i6A37 (Reiter et al., 2012). Un résidu guanosine peut également être présent en position 37 dans des ARNt mitochondriaux humains, qui sont souvent méthylés pour former une N1-méthylguanosine (m1G37) par TRMT5 (Brulé et al., 1998). La modification m1G perturbe la formation de l'appariement de base Watson-Crick, bloquant ainsi le couplage de base entre G37 et l'ARNm pour empêcher un décalage du cadre de lecture (Urbonavicius et al., 2003; Urbonavičius et al., 2001). En interférant avec l'appariement de base, m1G37 peut également empêcher les interactions avec des nucléotides voisins de l'autre côté de la boucle de l'anticodon afin d'aider à la formation de la structure de la boucle canonique requise pour les interactions codon-anticodon (Cabello-Villegas et al., 2002).

Enfin, les modifications post-transcriptionnelles des ARNt peuvent aider à la fidélité de la traduction en assurant une bonne aminoacylation. La reconnaissance d'un ARNt par son aminoacyl ARNt synthétase dépend de l'identification de nucléotides et de motifs structuraux particuliers, principalement dans la boucle anticodon et dans le bras accepteur, tels que la base discriminatrice à la position 73, et le CCA ajouté en 3' post-transcriptionnellement par TRNT1 (Nagaike et al., 2001). On suppose que la majorité des modifications contribue indirectement à l'identité des ARNt grâce à la stabilisation d'une caractéristique structurale reconnue. Cependant, que ce soit en tant que déterminant de la reconnaissance comme dans la modification Ψ 35 de l'ARNt Tyr de levure (Bare and Uhlenbeck, 1986), ou en tant qu'anti-déterminant comme cela a été décrit pour la modification m1G37 dans l'ARNt Asp de levure (Pütz et al., 1994) un rôle d'élément d'identité directe pour un nucléotide modifié ne peut pas être exclu pour les ARNt mitochondriaux (Van Haute et al., 2016).

Les ARNt chloroplastiques

Transcription et maturation des ARNt chloroplastiques

Comme mentionné ci-dessus, la transcription chloroplastique est effectuée par deux types d'ARN polymérase, la "plastid encoded eubacterial-type RNA polymerase" (PEP) d'origine bactérienne et la "nuclear-encoded phage-type RNA polymerases" (NEP) d'origine phagique. Les promoteurs des NEP se déclinent en deux versions, de type I avec un motif YRTA et de type II ne possédant pas de séquence consensus. Certains ARNt sont co-transcrits avec d'autres ARN. Par exemple, l'ARNt_{Ile}, l'ARNt_{Ala} et l'ARNt_{Arg} sont co-transcrits avec un ARNr depuis un promoteur NEP et / ou un promoteur PEP dans les plantes supérieures. Les ARNt_{Glu} et ARNt_{Val} sont transcrits à partir d'un promoteur PEP chez *Arabidopsis thaliana* (Michaud et al., 2011). Les gènes de ARNt sont dispersés sur le génome mais sont aussi parfois arrangés en opérons (Unseld et al., 1997). Un seul gène peut être placé sous le contrôle de plusieurs promoteurs. Comme pour les autres gènes de l'organite, la transcription des ARNt semblent être peu contrôlée ou modulée (Michaud et al., 2011). Contrairement aux ARNt de certaines mitochondries comme les mitochondries animales, les ARNt chloroplastiques ont une structure canonique. Outre les étapes de maturation universellement conservées, leur maturation implique des processus post-transcriptionnels comme la déamination d'adénosine en inosine en position 34. La disruption de *TADA*, le gène codant pour la déaminase responsable de cette édition de l'ARN a un effet majeur sur l'efficacité de la traduction chloroplastique (Delannoy et al., 2009).

d. Publication 1 : Transfer RNA: From pioneering crystallographic studies to contemporary tRNA biology

Cette revue préparée principalement par l'équipe de Claude Sauter et mon équipe d'accueil retrace l'historique des recherches sur les ARNt depuis l'obtention des premiers cristaux jusqu'à la biologie moléculaire et structurale moderne. J'ai participé à la préparation de cette revue en compilant les données bibliographiques et en réalisant une partie des figures.

La majorité des ARNt possèdent une structure 2D en feuille de trèfle et 3D en L. En effet, chez certaines espèces animales, les ARNt mitochondriaux ne possèdent pas de boucles T et D. Le cas le plus extrême est *Romanomermis culicivorax* qui possède dans ses

mitochondries des ARNt fonctionnels de 42 nucléotides contre 70 à 95 habituellement. Les ARNt possèdent également de nombreuses modifications post-transcriptionnelles. Ces modifications peuvent avoir un rôle de stabilisation de la structure de l'ARNt mais aussi dans la reconnaissance des codons des ARNm. En effet la modification en inosine de la base 34 se trouvant être la première base de l'anticodon permet de décoder tous les codons finissant par A, C ou U ce qui permet de limiter le nombre d'ARNt différents dans la cellule.

Après maturation, les ARNt doivent être chargés avec l'acide aminé correspondant à leur anticodon. Pour cela une aminoacyl tRNA synthetase (aaRS) fixe l'acide aminé correspondant en 3' de l'ARNt. Il existe deux classes d'aaRS basées sur la topologie du domaine catalytique. La classe I est trouvée sous forme monomérique et est caractérisée par la présence d'un repliement de type Rossmann alors que la classe II est trouvée généralement sous forme de dimère et possède un feuillet β antiparallèle.

Les ARNt possèdent de nombreux autres partenaires protéiques dans la cellule. Que ce soient les RNase P et Z servant à la maturation des fragments 5' leader ou 3' trailer, les différentes enzymes de modifications post-transcriptionnelles ou les protéines ribosomales. Les structures d'ARNt en complexe avec des facteurs d'élongation ou d'initiation et à différents stades de l'élongation ont permis de mettre en évidence la flexibilité des ARNt.

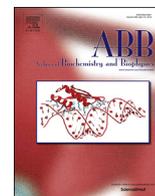
Les ARNt ne possèdent pas qu'un rôle dans la traduction. Par exemple chez les bactéries l'ARNt^{Ser} est utilisé dans la synthèse de la paroi cellulaire. Les virus utilisent également les motifs récurrents des ARNt pour les mimer et utiliser la machinerie cellulaire de l'hôte à son avantage comme les rétrovirus qui utilisent une structure ARNt-like pour initier la transcription inverse de leur génome par leur hôte. Les ARNt peuvent aussi jouer un rôle de régulateur grâce au riboswitch T-box contrôlant l'expression de nombreux gènes reliés à l'amination et au métabolisme des acides aminés.

Aujourd'hui l'étude des ARNt fait face à de nouveaux enjeux. La compréhension de la plasticité des ARNt est un de ces nouveaux défis, de même que la connaissance des ARNt non canoniques responsables de maladies et leur reconnaissance par différents partenaires. Par ailleurs, le mouvement des ARNt dans la cellule est encore très peu connu. On connaît peu de chose sur l'importation des ARNt notamment dans la mitochondrie où le pool d'ARNt est un mélange à proportions différentes selon les organismes d'ARNt codés dans la mitochondrie et d'ARNt codés dans le noyau puis importés. Finalement, une nouvelle branche de la biologie des ARNt a récemment émergé, avec la découverte des tRF (tRNA-derived fragments). Leur accumulation spécifique laisse penser que ces tRF sont une nouvelle classe d'ARN non codants qui peuvent avoir d'importantes fonctions de régulation.



Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Transfer RNA: From pioneering crystallographic studies to contemporary tRNA biology

Pablo Fernández-Millán ^a, Cédric Schelcher ^b, Joseph Chihade ^{a, c}, Benoît Masquida ^d,
Philippe Giegé ^b, Claude Sauter ^{a, *}

^a UPR 9002 du CNRS, Architecture et Réactivité de l'ARN, IBMC, Université de Strasbourg, 15 rue René Descartes, Strasbourg, F-67084, France

^b UPR 2357 du CNRS, IBMP, Université de Strasbourg, 12 rue du général Zimmer, F-67084, Strasbourg, France

^c Department of Chemistry, Carleton College, 1 North College Street, Northfield, MN 55057, USA

^d UMR 7156 du CNRS, Génétique Moléculaire, Génomique, Microbiologie, Université de Strasbourg, 4 Allée Konrad Roentgen, Strasbourg, F-67084, France

ARTICLE INFO

Article history:

Received 3 November 2015

Received in revised form

29 February 2016

Accepted 3 March 2016

Available online xxx

Keywords:

Transfer RNA

Crystallography

Protein synthesis

Genetic code

Translation

RNA:protein recognition

ABSTRACT

Transfer RNAs (tRNAs) play a key role in protein synthesis as adaptor molecules between messenger RNA and protein sequences on the ribosome. Their discovery in the early sixties provoked a worldwide infatuation with the study of their architecture and their function in the decoding of genetic information. tRNAs are also emblematic molecules in crystallography: the determination of the first tRNA crystal structures represented a milestone in structural biology and tRNAs were for a long period the sole source of information on RNA folding, architecture, and post-transcriptional modifications. Crystallographic data on tRNAs in complex with aminoacyl-tRNA synthetases (aaRSs) also provided the first insight into protein:RNA interactions. Beyond the translation process and the history of structural investigations on tRNA, this review also illustrates the renewal of tRNA biology with the discovery of a growing number of tRNA partners in the cell, the involvement of tRNAs in a variety of regulatory and metabolic pathways, and emerging applications in biotechnology and synthetic biology.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Since the mid-1950s crystallography has accompanied the birth and the development of molecular biology, providing biologists with pictures of their favorite biomolecules and helping them describe biological processes at an atomic scale [1]. The determination of the iconic three-dimensional structure of the DNA double helix by Watson and Crick revolutionized the understanding of life [2]. It led Crick to formulate his central dogma explaining the propagation of genetic information by DNA replication, transcription into RNA and translation into proteins, following a genetic code that associates every base triplet or codon in the DNA/RNA sequence with an amino acid in the polypeptide chain [3,4]. Yet an essential piece of the puzzle was missing and Crick proposed the 'adaptor hypothesis' in 1955, postulating the existence of a dedicated class of molecules making the link between nucleic acid and protein sequences [3]. Soon after a family of small RNAs, first called soluble RNAs (sRNAs) and quickly renamed transfer RNAs (tRNAs),

were isolated along with an essential family of enzymes, aminoacyl-tRNA synthetases (aaRSs), which catalyze the specific coupling of amino acids to their tRNA carriers [5].

At this point, it became crucial to gain an insight into the structural properties of these molecules in order to understand the specificity of substrate recognition and of catalytic mechanisms ensuring the fidelity of tRNA aminoacylation reactions and, by extension, of protein synthesis. This triggered a worldwide effort to decipher the crystal structure of tRNAs and related enzymes. In 1966 the sequencing of tRNA^{Ala} from yeast revealed complementary stretches of nucleotides suggesting a 2D folding resembling a cloverleaf [6]. The next three sequences of tRNA^{Tyr}, tRNA^{Phe}, and tRNA^{Ser} [7–9] reinforced the idea of a four-arm organization. They also led to the concept of anticodon base triplet complementary to the codon of the carried amino acid, establishing a physical link between nucleic acid and protein alphabets/sequences. Hence, it became clear that all tRNAs should share a common scaffold to ensure their recognition by the ribosome and by elongation factors, while also expressing a specific identity with regard to their cognate aaRSs. The quest for structural data led to a first important breakthrough in the mid-seventies with the determination of the

* Corresponding author.

E-mail address: c.sauter@ibmc-cnrs.unistra.fr (C. Sauter).

crystal structure of yeast tRNA^{Phe} by two groups in the US and one in the UK [10–12], followed by that of yeast tRNA^{Asp} in France [13]. These crystallographic studies revealed the characteristic L-shape of adaptor molecules. The story continued with the determination of the first crystal structures of tRNA:aaRS complexes in the early nineties reviewed in Ref. [14] and culminated in the 2000s with the description of the adaptors bound to the A, P and E sites of the ribosome [15].

In the context of the celebration of a century of X-ray crystallography, the aim of this review is to illustrate how early studies of the key protagonists in the translational machinery such as tRNAs have contributed to many aspects to structural biology and how vital this tRNA research continues to be. Solving the initial structures represented a considerable challenge and required methodological developments in sample preparation, crystallization and crystallography. For almost 15 years, tRNAs were the sole source of structural information about RNA and provided the basis for the understanding of RNA architecture and function, as well as information about RNA:protein recognition and the role of RNA modifications. Finally, while tRNAs entered textbooks in the nineties as central actors in translation [16], a number of functional questions have recently emerged from the genome-wide examination of tRNA diversity. These include the discovery of new tRNA partners and related pathways and the observation of tRNA mutations related to human pathologies, stimulating novel research directions on tRNA function, structure, dynamics, and biotechnology.

2. From first crystal structures to the -omics era

A first step towards the crystallographic analysis of tRNAs was the separation of different species from natural sources like baker's yeast. RNA is distinguished from DNA by the presence of a hydroxyl group in 2' position of the sugar which confers both flexibility and reactivity, making RNA molecules much more labile and difficult to work with than DNA. In the sixties the method of countercurrent distribution, based on partitioning tRNAs according to their hydrophobicity in aqueous and organic solvents, played a key role. This technique allowed the preparation of pure samples on the hundreds of milligram scale required for biochemical characterization, sequencing and eventually crystallization [17]. The introduction of the vapor diffusion method was also crucial for the success of tRNA crystallization [18]. Other micromethods, such as dialysis or free-interface diffusion, were developed to crystallize aaRSs and aaRS:tRNA complexes because they were better adapted to sensitive biological samples and have since been widely adopted. A new discipline called biocrystallography, aiming at better understanding and mastering the process of biological crystallization, directly emerged from the difficulties encountered in the preparation of crystals suitable for diffraction studies [19]. These efforts eventually yielded tRNA crystals diffracting at resolution of ~3 Å at the turn of the seventies.

Today, when it has become possible to determine X-ray structures on a simple laptop computer, it is difficult to imagine how solving a tRNA structure once represented long and tedious work, collecting diffraction pattern on photographic films and building a model manually in hand-drawn electron density maps (Fig. 1). These pioneering crystallographic studies led to the well-known L-shape model. The building of full atomic models revealed how helical domains of the cloverleaf get stacked pairwise (acceptor and T helices, D and anticodon helices; Fig. 2A) to form the two arms of the L, held together by a network of tertiary interactions predicted from sequence conservations by Levitt in 1969 [20]. Although the angle between the arms in tRNA^{Asp} is more open than in tRNA^{Phe} (100° instead of 90°), a sign of the intrinsic flexibility of tRNA backbone, both tRNAs adopt a similar fold, in agreement with the

boomerang-shape proposed based on early small angle X-ray scattering (SAXS) measurements [21]. This canonical 2D cloverleaf and 3D L structure were later confirmed in a dozen crystal structures of free tRNAs, as well as by tens of thousands of tRNA sequences derived from a growing number of genomes and collected in dedicated databases such as the tRNAdb (<http://trna.bioinf.uni-leipzig.de>) [22]. Exceptions to this general pattern have been found in animal mitochondria, where tRNAs can display size reductions due to shorter loops in the D or T arms, and in some cases, the complete absence of one or two of arms (Fig. 2B). The most extreme case is a functional tRNA consisting of only 42 nucleotides (to be compared with usual sizes of 70–95 nt) recently described in the mitochondria of the worm *Romanomermis culicivorax* [23]. This raises questions about the folding and the stability of such tRNAs and their recognition by cellular factors including the ribosome.

Another structural characteristic of tRNAs is the presence of numerous post-transcriptionally modified nucleosides (Fig. 2C). Besides methylated bases, the first sequence of yeast tRNA^{Ala} also revealed the presence of dihydrouridines in one loop, as well as a ribo-thymidine (T) and a pseudouridine (Ψ) in another, leading to the current names of D- and T- (or TΨC-) loops/stems as these modifications are almost universally conserved. Inosine was also found at position 34, soon identified as the first position of the anticodon, which base pairs with the third position of the codon in the messenger RNA. Following the 'wobble hypothesis' of Crick [3], an inosine at this position in the tRNA extends its decoding capacity to the synonymous codons ending by A, C or U (GCA, GCC and GCU in the case of tRNA^{Ala} with an IGC anticodon). For this reason, the number of tRNA species in a cell is always lower than the 61 codons, and can be as low as 22 in mammalian mitochondria, due to the presence of specific modifications at positions 34 and 37 of the anticodon loop [24]. Modified nucleosides can serve as recognition signals for cognate synthetases (see below) and are known to increase the stability of the tRNA scaffold, as reviewed in Ref. [25]. A classic example of modification that affects tRNA structure is the m1A at position 9 in human mitochondrial tRNA^{Lys} which drives the formation of the functional cloverleaf over an alternative hairpin structure [26]. More than a hundred such RNA modifications have been identified (Fig. 2C) and their structures and positions in tRNAs, together with information on their biosynthetic pathways, have been cataloged in the RNA modification (<http://rna-mdb.cas.albany.edu>) and the Modomics (<http://modomics.genesilico.pl>) databases [27,28].

3. Exploring the mechanisms of tRNA:synthetase recognition

The first structural information on tRNA:synthetase interactions came more than a decade after the solution of tRNA^{Phe} structure. The delay was mainly due to the difficulties in the preparation of homogeneous complexes for crystallization and crystal quality optimization. The first structures of tRNA:aaRS complexes specific for glutamine and aspartate [29,30] highlighted that, although synthetases catalyzed the same type of reaction, they diverge in terms of sequence and structure. The aaRS family was divided into two classes based on the topology of their catalytic domain [31]: a Rossmann fold, with its characteristic parallel β-sheet, is found in monomeric Class I aaRSs, while Class II enzymes are generally dimeric and are built around an antiparallel β-sheet. The determination of a number of crystal structures of tRNA:aaRS complexes has revealed the specificities and dynamics of class I and class II aaRS binding to tRNA.

The observation of close contacts between the partners, together with mutagenesis experiments on both partners, reinforced the concept of a second genetic code. This operational code involves tRNA determinants and tRNA anti-determinants, which

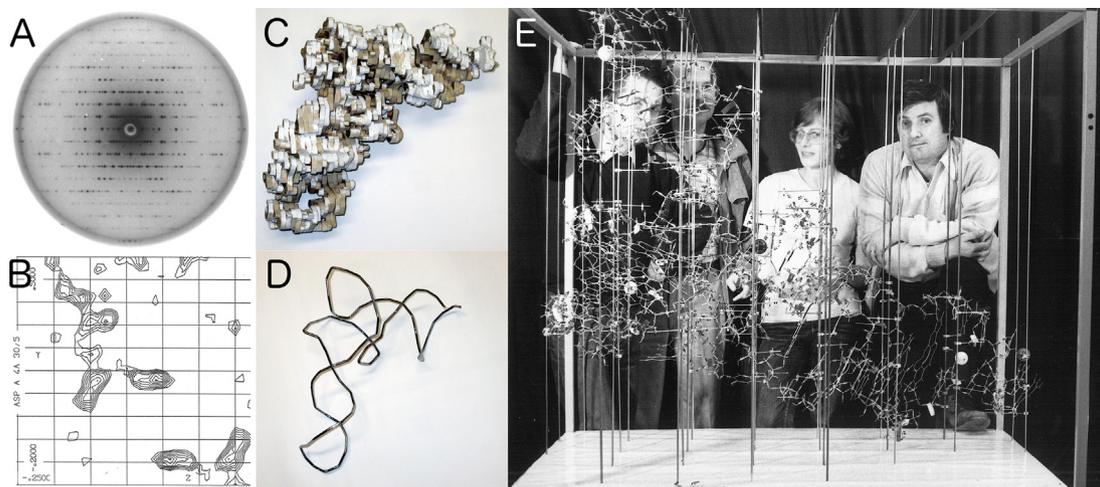


Fig. 1. Determination of yeast tRNA^{ASP} crystal structure [13]. Structural investigations on yeast tRNA^{ASP} started in the seventies and illustrate well how crystallographic studies were performed at the time. (A) Diffraction patterns were collected on photographic films from which reflection intensities were quantified. Multiple isomorphous replacement with heavy atoms was used to determine sets of phases and to compute slices of electron density maps (B). A low resolution model in balsa wood was first assembled (C), then the tRNA backbone was built in metal wire by connecting the positions of phosphates, which were easily identified in the density (D). Using an optical device called Richards box (see https://en.wikipedia.org/wiki/Frederic_M._Richards) and electron density maps manually transferred on glass windows, the French research team at IBMC, Strasbourg (E) could build a wire frame atomic model (scale: 2 cm = 1 Å) showing the characteristic L-shape of yeast tRNA^{ASP}. From the left to the right: J.-C. Thierry, R. Giegé, M.-B. Comarmond, D. Moras (courtesy of R. Giegé).

are chemical signals defining the identity of a family of tRNAs with respect to their cognate aaRS or preventing the interaction and misacylation by a non cognate enzyme, respectively [32]. As might be expected, the anticodon region often constitutes a major identity element. In several complexes, anticodon nucleotides are distorted so as to be bound in specific enzymatic pockets (see, for example, the structure of the GlnRS-tRNA^{Gln} complex [33]).

Identity elements are also frequently found in the acceptor stem, but the interpretation of specific protein contacts with this region of the tRNA is more difficult, since interactions with both the ribose-phosphate backbone and nucleotide bases may play a role. The clearest example may be provided by two recent structures of *A. fulgidus* alanyl-tRNA synthetase [34]. In one, the enzyme is complexed with a cognate tRNA, containing the G3:U70 wobble pair, which is an identity element, while the second is a complex with an A3:U70 tRNA mutant. The structures reveal a multifaceted set of protein-RNA interactions which direct the A76 nucleotide of the cognate tRNA into the aminoacylation active site, but distort the backbone of the mutant tRNA and prevent A76 from attaining the proper configuration for aminoacylation. Other tRNA:aaRS structures have demonstrated idiosyncratic modes of tRNA recognition that are difficult to describe within the framework of identity elements, such as *T. thermophilus* SerRS recognition of the long variable arm of tRNA^{Ser} [35] and “shape-selective” contacts between *E. coli* CysRS and tRNA^{Cys} [36].

These structural studies also shed light on the high degree of tRNA plasticity in tRNA:aaRS interactions [37]. For example, in the case of yeast aspartylation system, the comparison of tRNA^{ASP} structure with the structure of its complex with AspRS revealed important conformational changes altering its structure upon enzyme binding, i.e. with the two arms of the overall L shape of tRNA^{ASP} presenting a reduced angle [30]. In class I aaRSs including an editing domain, the 3' end of the tRNA can swing from the aminoacylation site to the editing site to allow hydrolysis of an incorrect amino acyl ester [38]. Despite the rapidly growing number of crystal structures of tRNA:aaRS complexes, the existing data remain highly biased in favor of bacterial systems, with few archaeal and eukaryotic structures available. Knowledge of organelle aminoacylation systems is even more scarce, with only a few

isolated aaRS structures but no structures of complexes [37]. Recently, mitochondrial systems have drawn substantial interest due to the discovery of mutations in tRNAs and aaRSs associated with a variety of human pathologies [39,40]. Future investigations will be necessary to examine how these mutations affect expression, folding, binding, catalysis or other functions of tRNAs aside from aminoacylation and translation.

4. tRNA as template for RNA modeling

As mentioned above, the first tRNA structures were, for a long period, the only examples of complex RNA architecture (Fig. 3). As such, they were instrumental in elaborating how double-stranded helices can be connected by structured loops in which all nucleotide edges (Watson-Crick, Hoogsteen, sugar) are used in base-pairing. Important roles of the 2' hydroxyl group of nucleotides were noted, such as its capacities to constrain the ribose pucker, forcing A-form RNA helices, and to simultaneously act as a donor and acceptor of hydrogen bond. Despite their relatively limited size and number, as the only available templates for RNA modeling, the tRNA crystal structures have long been a source of inspiration.

The observation of an overall L-shape architecture stabilized by pairwise helix stacking led to the concept of RNA domain assembly. This assembly is further stabilized by loop-loop interactions on one side and base triples on the other (Fig. 4). The stunning beauty of the structures was due to the harmony of the backbone curves resulting from interactions between bases in a layout so natural that the whole appeared to be unconstrained. As masterpieces resulting from millions of years of molecular adaptation, the tRNA crystal structures inspired structural biologists. The observation that the single strands joining the helical segments adopted precise architectures led to the idea that they could constitute individual structural motifs.

This essential RNA structural repertoire (See Fig. 4) could in principle be used as a building set to generate theoretical RNA structural models by pasting together individual building blocks. However, in the eighties the organization of the secondary structures of ribosomal RNAs (rRNA) already indicated that the RNA structural diversity was not exhaustively contained in the tRNA

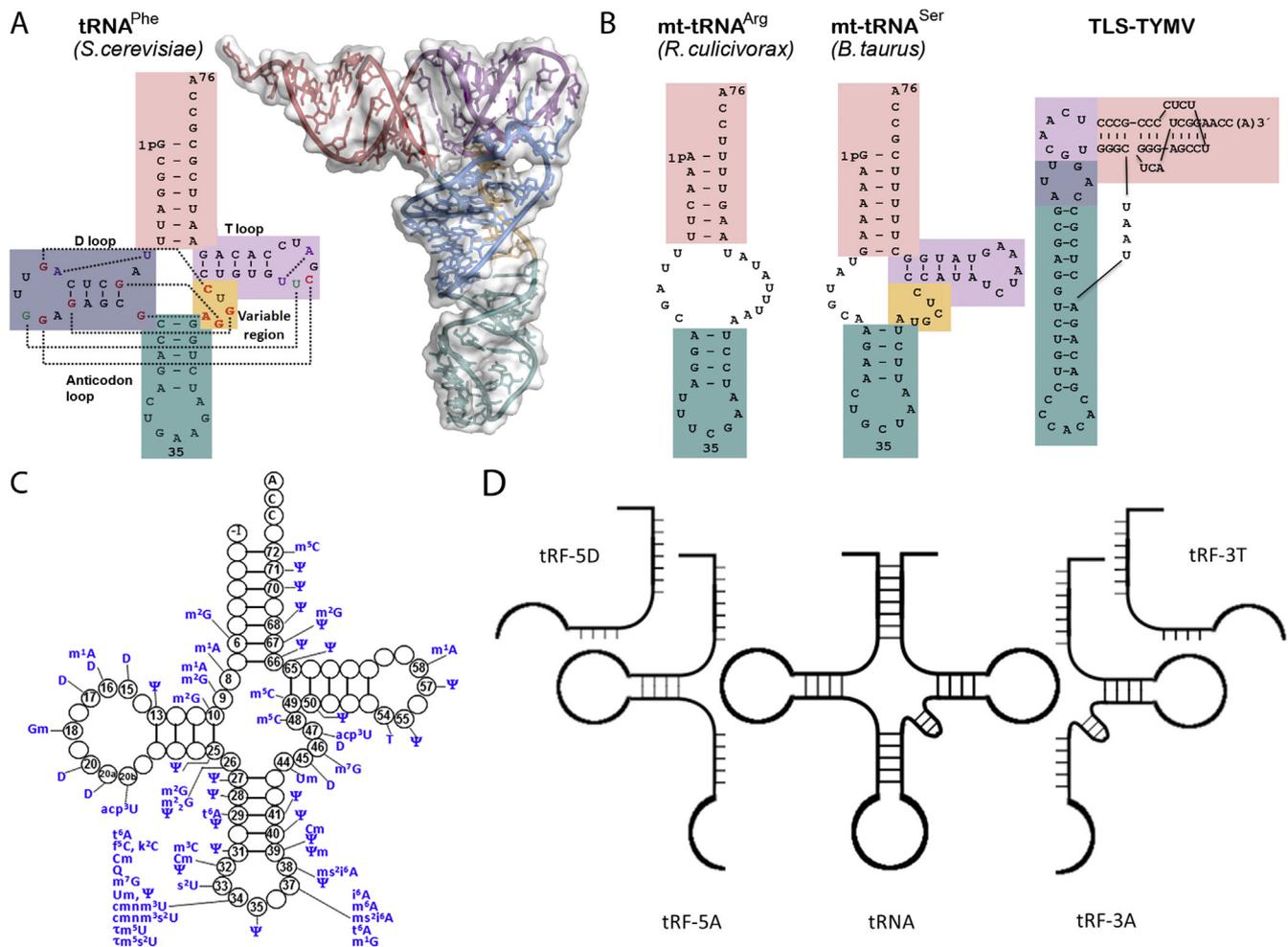


Fig. 2. tRNA structural organization. (A) 2D and 3D structures of tRNA^{Phe} (PDBid 1TN1). Tertiary interactions that stabilize the L-shape are depicted on the cloverleaf as broken lines (more details in Fig. 4). The arms of the cloverleaf are depicted with the same colors throughout the figure. (B) Examples of variations on the classical 2D organization with typical armless mitochondrial tRNAs and the tRNA-like structure (TLS) found in 3' of the genomic RNA of Turnip Yellow Mosaic Virus (TYMV). (C) Distribution of post-transcriptional modifications on a tRNA cloverleaf with conventional nucleotide numbering. (D) General nomenclature of tRNA-derived fragments (tRFs). Mature tRNAs and precursors can be cleaved at different positions to produce various tRFs. The atomic model of tRNA^{Phe} was rendered using PyMOL (Schrödinger), as well as atomic models in Fig. 4 and 5.

structures. The number of loop residues well illustrates this idea. The tRNAs display seven- or eight-membered loops, while rRNAs contain less or more nucleotides in the loops closing hairpins. Moreover, rRNAs present internal loops or complex junctions linking two to five helical segments impossible to map onto tRNA subdomains. At this stage, it was difficult to anticipate how the RNA structural repertoire would expand. However structural biologists could understand that the diversity of base-pairs beyond the Watson-Crick classics would be at play. Additional likely interactions were observed by analysis of tRNA crystal packing. Thus, building unknown motifs from interactions between nucleotides became a field of research. All atom modeling independently added stereochemical restraints on top of the secondary structure constraints, conferring significant predictive power to the method. Structural biologists could use RNA helices, base triples and kissing loops from the tRNA structures to build complex architectures like group I introns [41].

The number of available RNA crystal structures increased gradually (Fig. 3), each time revealing new structural features that helped structural biologists to better understand RNA architecture and folding principles [42–44]. However the mist really cleared upon the appearance of the crystal structures of the ribosomal subunits [45–47], which multiplied the size of the structural

repertoire by ~7 fold and revealed a wealth of magnificent and daring RNA folds, RNA-RNA and RNA-protein interaction motifs. Strikingly, most of these folds are recurrent, like the sarcin or loop E motif, the GNRA and UNCG tetraloops, kink-turns [48,49], C-loops [50], UA handles [51], and A-minor motif [52,53]. For example, the A-minor motif is very frequent due to the versatility of the O2' group interactions, although it cannot be deduced from comparative sequence analysis [54]. Some of these motifs are even reminiscent of tRNA: one can find a fair representation of anticodon loops, U-turns, as in the structure of the lariat-capping ribozyme [55], and T-loops, the latter being crucial for tRNA recognition by RNase P [56], T box regulators [57] and ribosome [15,58].

RNA modeling has never been as necessary as today since massive numbers of RNA sequences are being identified by new generation sequencing (NGS) methods at a rate far faster than conventional structural methods such as crystallography can follow up. The continued efforts of structural biologists to increase the number of RNA crystal structures are thus mandatory to approach completion of the RNA structural repertoire which makes up the modeling building set. The identification of both secondary structures and nucleotides in proximity in RNAs by high-throughput chemical probing methods can then be coupled to automatic modeling methods based on the adjunction of RNA fragments. This

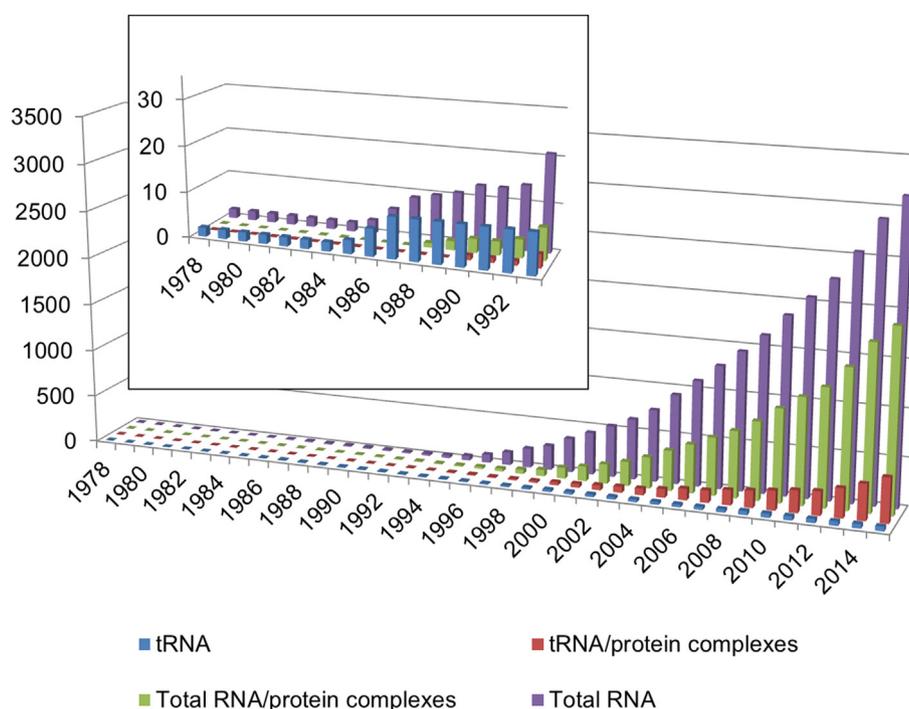


Fig. 3. tRNA and RNA entries in the PDB (as of December 2014). The plot bar shows the number of PDB entries per year for structures of tRNA alone (red), tRNA:protein complexes (blue, this group includes tRNA:ribosome complexes), all RNA:protein complexes (green) and all isolated RNAs including tRNAs (violet). The inset gives a close-up view for the years 1978–1993. Until 1994 and the determination of first ribozyme structures, tRNAs were the exclusive source of information on the folding and architecture of complex RNA molecules. Then the study of the ribosome led to a burst in RNA and tRNA/RNA:protein complex data in the PDB.

type of approach, exemplified by MOHCA-seq coupled to modeling with Rosetta [59] have proven to reach subhelical resolution (10 Å).

5. tRNA partners in the cell

As a central actor of the translation machinery, many of tRNA's partners are involved in the process of protein synthesis, including aaRSs, initiation and elongation factors that deliver aminoacyl-tRNAs to the ribosome. A growing number of functional complexes not directly related to protein synthesis have also been structurally characterized over the years, including maturation enzymes involved in tRNA biogenesis, regulatory T-boxes in mRNA and tRNA-like structures in viral RNAs as shown in Fig. 5, illustrating the diversity of cellular functions associated with tRNAs.

5.1. tRNA biogenesis

Like many other RNA molecules, tRNAs are expressed as precursor transcripts and need to undergo several maturation steps in order to become functional. tRNA precursors are first cleaved by two endonucleases called RNase P and RNase Z, which remove 5' leader and 3' trailer sequences of tRNA precursors respectively. The RNase P function was believed to be ubiquitously performed by ribonucleoproteins containing a ribozyme (P RNA) until the recent discovery of protein-only RNase P enzymes in eukaryotes reviewed in Ref. [69]. The determination of the structure of the bacterial *Thermotoga maritima* RNase P in complex with tRNA has revealed that tRNA-P RNA recognition occurs through shape complementarity, specific intermolecular contacts and base-pairing interactions [56]. Interestingly, although a structure of protein-only RNase P in complex with tRNA has not yet been solved, initial biochemical and biophysical data suggest that protein-only and ribonucleoprotein RNase P share similar processes for specific tRNA recognition [70]. After the initial end-trimming, tRNAs undergo

several additional modification and editing events [71]. A number of structures of tRNA modifying enzymes in complex with complete tRNA have exposed several stages of the tRNA modification process, including initial RNA binding, the formation of covalent intermediates and the release of the modified RNA reviewed in Ref. [72]. The structures of enzyme-tRNA complexes allowed tRNA modifying enzymes to be categorized in distinct classes depending on their requirement for the L-shape structure of tRNA for RNA recognition. Group I enzymes bind either nucleotides in the anticodon loop or residues embedded inside the tRNA structure, whereas group II enzymes require the three dimensional structure of tRNA for recognition [73]. The last step in tRNA maturation is the addition of a CCA 3' group by CCA adding enzymes (CCases). Structural analyses have revealed that CCases use a template independent RNA polymerization mechanism [68]. More recently, the determination of structures of archaeal CCA adding enzymes in complex with tRNA has revealed how the enzymes discriminate between the incorporation of C or A at position 76. The discrimination against incorporation of C at this position arises because protein flexibility results in the improper geometry of CTP in the active site [74].

5.2. tRNAs in the ribosome

Aminoacylated tRNA in complex with initiation and elongation factors reaches the ribosome where the incorporation of amino acids into the emerging protein chain through peptide bond formation is achieved. A rapidly increasing number of high-resolution structures of tRNA in complex with ribosome, ribosomal subunits and/or elongation factors at different stages of translation have been obtained by both X-ray crystallography and cryo-electron microscopy (cryoEM) [75]. These structures together with biochemical and computational approaches have revealed the plasticity of tRNA molecules during their interaction with the

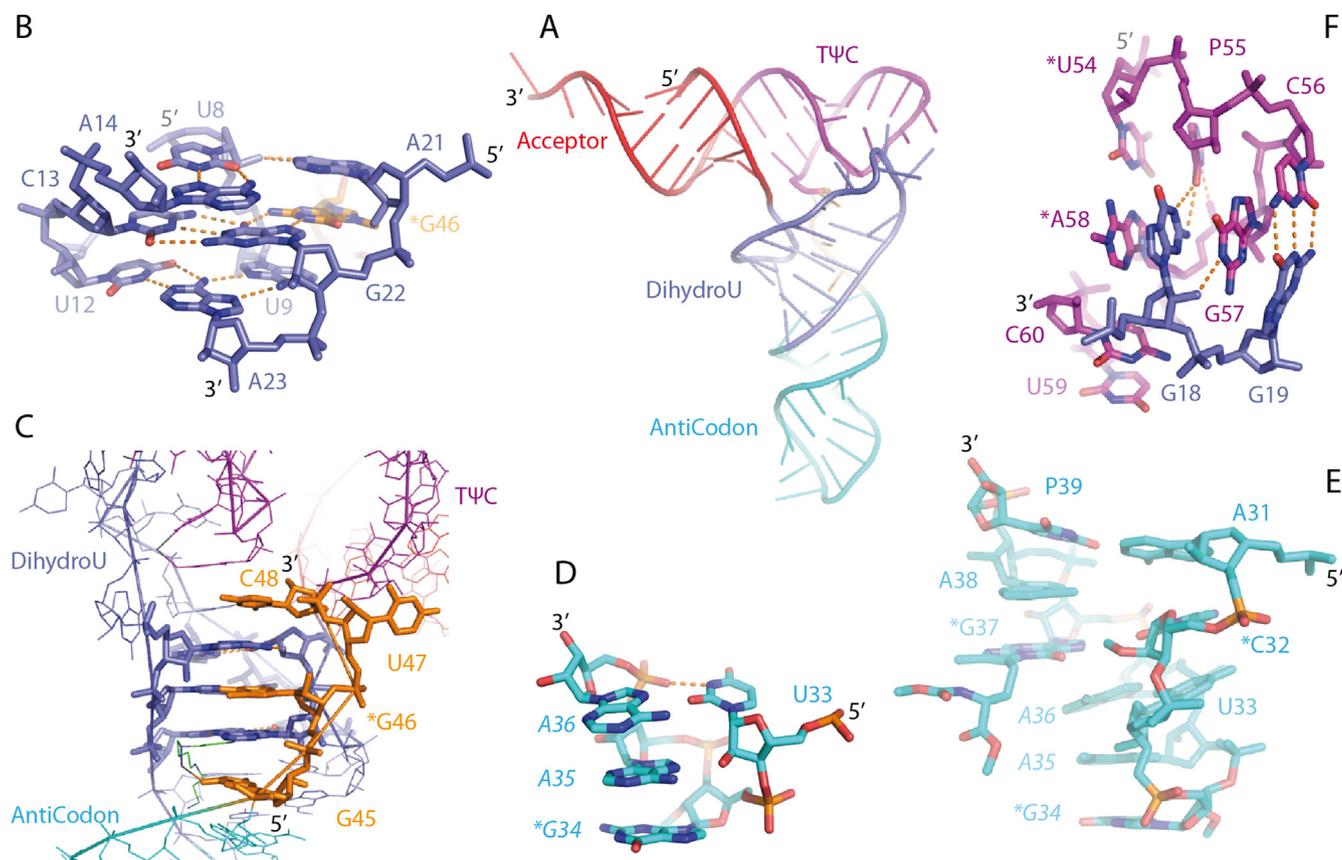


Fig. 4. The early observation of tRNA structures shed light on RNA folding and architecture. (A) Overall ribbon view of the tRNA^{Phe} crystal structure (PDB ID 1TN1). The different domains are indicated and colored individually. The variable loop lies behind the dihydrouridine stem loop (dim orange). (B) The core of the tRNA is stabilized by base triple interactions. In the foreground, the last two base pairs of the D stem are represented, U₁₂-A₂₃, and C₁₃-G₂₂. In the deep groove of the stem the Hoogsteen edge of U₉ and the Watson-Crick edge of m⁷G₄₆ interact with the Hoogsteen edge of A₂₃ and of G₂₂, respectively. The first nucleotide of the D loop, A₁₄, is stabilized by U₈, which directly connects the Acceptor stem to the D domain. The resulting *trans* Watson-Crick/Hoogsteen interaction is further stabilized by A₂₁, the terminal residue of the D loop. m⁷G₄₆ stacks in between A₂₁ and G₂₂. As seen in (C), stacking interactions extend farther than the base triple ensemble, which is wrapped by C₄₈ and G₄₅. The latter interacts with the first base pair of the D stem (not shown), while the former stacks onto U₅₉, the unpaired penultimate nucleotide in the 3' side of the T loop. (D) The U-turn motif consists in the interaction between the N3 imino proton of U₃₃ with the proximal phosphate oxygen atom of A₃₆. This motif is observed recurrently in RNA structures, for example in the T loop of the tRNA (P₅₅ and M1A₅₈ in panel F). (E) A view of the overall anticodon loop shows the arrangement of the three nucleotides 34–36 (italic) involved in mRNA recognition. Hyper modification of nucleotide 37 prevents its interaction with a fourth nucleotide on the mRNA, which would result in mistranslation. (F) A view of the nucleotides involved in the interaction between the T and D loops depicts how stacking interactions promote close contact between nucleotides which result in exquisite hydrogen bonds. Notably, packing contacts in tRNA crystals between anticodon loops of adjacent tRNA molecules dictate the codon-anticodon interaction observed on the ribosome. Hydrogen bonds are indicated by solid dashed lines. "*" indicates modified nucleotides. *G₄₆: 7-methylguanosine; *G₃₄: 2'-O-methylguanosine; *C₃₂: 2'-O-methylcytidine; *U₅₄: 5-methyluridine; P₅₅: pseudouridine; *A₅₈: 1-methyladenosine.

ribosome. It thus becomes increasingly evident that the intrinsic structural flexibility of tRNA is pivotal for protein synthesis [76]. In the ribosome, tRNAs bind the primary aminoacyl (A) site, travel to the peptidyl (P) site and leave the ribosome at the level of the exit (E) site. In this process, tRNA conformational changes are required for the concerted action of the ribosome's small and large subunits during translation initiation, for the decoding steps where aminoacylated tRNA are incorporated and for ribosome translocation [77].

Translation is initiated with the binding of the initiator tRNA^{Met} to the P site of the small ribosomal subunit, where its anticodon basepairs with the mRNA start codon. Initiator tRNAs possess structural features that distinguish them from elongator tRNAs, including the elongator tRNA^{Met}, and enable their specific recognition and shuttling by initiation factors IF2/aIF2/eIF2. In *Escherichia coli*, the initiator tRNA bears a C1-A72 mismatch (instead of a G1-C72 pair in the elongator) which is essential for the formylation of the methionine. This formyl group ensuring a specific binding to IF2, instead of EF1A (or EF-Tu), the elongation factor [78]. In Eukaryotes and Archaea, a weak base A1-U72 is a major identity element for binding to e/aIF2 whereas Ψ54, a major recognition signal for eEF1A, is absent. Met-tRNA^{Met} is shuttled by IF2 or e/aIF2

to the P site of the ribosome. In all kingdoms, its anticodon stem contains 3 consecutive G-C pairs followed by an additional C32-A38 wobble pair [64,79]. This unique sequence leads to a specific anticodon loop conformation which is proposed to facilitate the accommodation of Met-tRNA^{Met} in the P site.

In prokaryotes, the small subunit of ribosomes binds the initiator tRNA, initiation factors and mRNA to form the 30S pre-initiation complex in a sequential process. After the positioning of the mRNA, the anticodon of the initiator tRNA is base-paired with the initiation codon and the sequential release of initiation factors induces association of the large subunit to establish the 70S initiation complex [80]. In contrast, in eukaryotes, the initiation of translation is a more complex and tightly regulated process engaging over ten initiation factors, in which eukaryotic ribosomes use a scanning mechanism resulting in the formation of the initiation complex with the initiator tRNA in its ultimate location [81,82]. In both eukaryotes and prokaryotes, the variety of conformations observed for initiator tRNAs reveals the plasticity of the acceptor arm and suggests that both the assembly of ribosome subunits and the integration of initiator tRNA into the P site might be regulated by the distinct conformations adopted by tRNA during

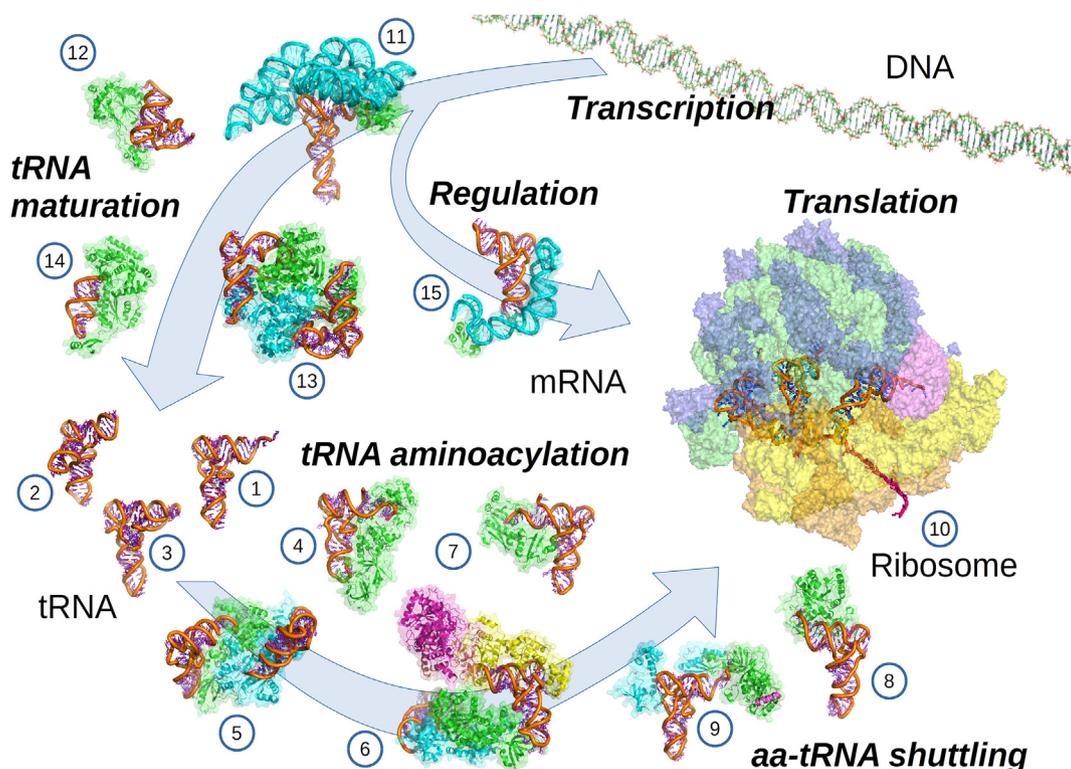


Fig. 5. Structural gallery of tRNAs and their cellular partners. This selection of representative structures shows the diversity of tRNA partners and includes free tRNAs and tRNA-like molecules with (1) the yeast phenylalanyl-tRNA [10,11], (2) the murine selenocysteinyl-tRNA characterized by a large variable region and an acceptor stem with 8 base-pairs [60], (3) the tRNA-like structure present at the 3' of TYMV genomic RNA [61]; enzymatic complexes active in tRNA aminoacylation and protein synthesis, involving (4) a monomeric class I glutaminyl-tRNA synthetase [29], (5) a dimeric class II aspartyl-tRNA synthetase [30], (6) the transamidosome [62], (7) the methionyl-tRNA^{Met}-formyltransferase [78]; bacterial and archeal elongation factors (8) EF-Tu [63] and (9) aIF2 [64], and (10) the bacterial ribosome bound to EF-Tu:tRNA [65]; tRNA maturation enzymes with (11) the ribonucleoprotein RNase P [56] (12) the ribonuclease Z [66], (13) tRNA-guanine transglycosylase associated with tRNA in so-called lambda open conformation [67] and (14) CCA adding-enzyme [68]; regulatory elements represented by (15) the T-box riboswitch [57]. All tRNAs are depicted with an orange backbone and pink bases, free tRNAs are represented, tRNA partners are shown with a transparent surface and different colors are used to distinguish their subunits.

the initial steps of translation [83].

When the ribosome is assembled, decoding of the mRNA can start and the delivery of elongator tRNAs is ensured by the elongation factors. The determination of the structure of ternary complex made of EF-Tu, Phe-tRNA^{Phe} and an analog of GTP represented an important breakthrough in the mid 90s and shed light on the recognition of aminoacylated tRNAs [63]. Recent progress in ribosome crystallography enabled the observation of this ternary complex docked onto the ribosome (Fig. 5) and revealed the structural plasticity of tRNAs [65]. Then, during the decoding process, the ribosome has to discriminate between cognate aminoacylated tRNA-elongation factor complexes and non-cognate complexes. This is achieved through a kinetic discrimination mechanism that allows incorrect tRNA species to be rejected at different stages. Decoding relies on Watson-Crick pairing between mRNA codons and tRNA anticodons together with 30S conformational changes [84]. The ribosome plays an active role in the accuracy of this process as it uses an ensemble of local and global conformational changes to control the fidelity of tRNA selection e.g. Refs. [85,86]. Here again the structural flexibility of tRNA plays a fundamental role to allow the decoding process in both prokaryotic and eukaryotic systems [87].

For the translocation of ribosomes, peptide transfer takes place after the entry of tRNA in the A site guiding the formation of the pre-translocation complex poised to translocate tRNA-mRNA duplexes by one codon e.g. Refs. [87,88]. In this motion, the acceptor stems of tRNA in the A and P sites move 5' to 3', resulting in the hybrid A/P and P/E configurations [89]. This tRNA reorganisation is

associated with the rotation of the 30S relative to the 50S subunit along with the stalk rearrangement as shown by cryoEM studies of prokaryote and eukaryote complexes, which revealed several tRNA intermediate configurations [76,90,91].

Interestingly, in some cases, tRNAs interact with ribosomes in ways that go beyond the primary association of tRNA as part of the mRNA decoding process. A tRNA was also found associated to the large subunit of the mitochondrial ribosome in animals. This tRNA, most likely mitochondrial tRNA^{Phe}, is referred to as CP tRNA. It is an integral component of the central protuberance (CP) of the 39S subunit, where it structurally replaces the 5S ribosomal RNA, which is not encoded in animal mitochondrial genomes, despite being ubiquitous in cytoplasmic ribosomes [92,93].

5.3. tRNA beyond translation

In addition to their canonical function, tRNA molecules participate in a variety of other processes [37], such as cell wall synthesis in bacteria (tRNA^{Ser}) or the porphyrin synthesis pathway (tRNA^{Glu}). Viruses frequently take over some feature of tRNA biology. In particular, retroviruses recruit particular host tRNAs for the priming of reverse transcription of their genomes [94]. Likewise, large DNA viruses such as bacteriophages, phycodnaviruses, and mimiviruses express their own tRNAs, most probably to regulate translation during viral infection [95]. In bacteria, the T-box riboswitch differs from all other identified riboswitches by using a tRNA as its effector. This riboswitch is believed to monitor the aminoacylation state of tRNAs and to control the expression of genes involved in a number

of functions related to aminoacylation and the metabolism of amino acids [96]. The crystal structure of the highly conserved T-box riboswitch in complex with tRNA has revealed how the distal stem I region of the riboswitch interacts with the tRNA corner to anchor it to the riboswitch [57].

Beyond their canonical fold, the three-dimensional architecture of tRNAs is also found in a variety of other RNA molecules. These structures, called tRNA-like (TLS), are often found in 5' and 3' translated regions of plant mitochondrial mRNAs. They appear to serve as RNA processing signals for the maturation of mRNAs as they are recognized and cleaved by RNase P and RNase Z enzymes [97,98]. Similarly, a number of positive strand RNA plant viral genomes also possess 3'-tRNA-like structures. Although differing considerably from canonical tRNAs in terms of secondary structure, these TLS exhibit a three-dimensional fold that accurately mimics tRNAs and tRNA properties, including the capacity to be aminoacylated. These viral TLS often appear to be essential for viral replication [95]. For instance, the turnip yellow mosaic virus (TYMV) TLS (i) is recognized by ValRS like a true tRNA to aminoacylate the viral genomic RNA and (ii) binds the promoter region in the 3' untranslated region of the TYMV RNA for negative-strand synthesis. The determination of the TYMV TLS three-dimensional structure has revealed its global tRNA mimicry and has also shown how it exploits a completely different array of intramolecular interactions to resemble tRNA. These specific interactions enable the TLS to switch conformations as required for its function [61].

6. Perspectives and new challenges of tRNA structural biology

6.1. Understanding the conformational dynamics of tRNAs

The gallery of tRNA structures, either as isolated molecules or in complex with cellular partners (Fig. 5), highlights the flexibility of the L-shape scaffold. However, the cloverleaf behind this L-shape is not always the only 2D conformation possible, nor the most stable one. Defects in tRNA modification or mutations in mitochondrial tRNAs related to human pathologies were for instance shown to destabilize the cloverleaf and favor alternate 2D folds for which 3D models still need to be established [26,99]. Also, atypical tRNAs found in animal mitochondria, with sequence alterations ranging from a shortening of D/T loops, to the absence of D-arm, T-arm or both arms, deserve further investigation to understand their folding and their stability and to understand how translation factors can cope with such unusual and presumably flexible substrates. The example of the human mitochondrial aspartyl-tRNA synthetase suggests that the enzyme has relaxed its specificity and shows a higher plasticity than its *E. coli* homolog, an evolutionary response to the structural alteration of its cognate tRNA [100].

6.2. Understanding tRNA trafficking in the cell

In eukaryotes, gene expression is not restricted to the nucleus. It also takes place in organelles, in particular in mitochondria. The pool of tRNAs required for mitochondrial translation can be composed of both tRNAs encoded in mitochondria and in the nucleus, with the latter being expressed in the cytosol and imported to mitochondria by precise import systems. tRNA import is ubiquitous in eukaryotes, although very diverse. In yeast and in some animals, a full set of tRNA is encoded in mitochondria and tRNA import is not required a priori but nevertheless takes place [101]. In contrast, in trypanosomes, not a single tRNA is encoded in the mitochondrial genome and the full set of tRNA is imported. An intermediate situation is found in plants where a number of tRNAs are not encoded in mitochondria and have to be imported [102].

The precise nature of the factors involved in this process and their diversity throughout eukaryotes is only beginning to be unravelled. In yeast, the tRNA^{Lys}₁ is shuttled to the mitochondrial surface by the glycolytic enzyme enolase 2 and taken over by the precursor of the mitochondrial lysyl-tRNA synthetase to be co-imported in mitochondria through the protein import machinery [103]. In trypanosomes, tRNAs interact with the cytosolic elongation factor eEF1a to reach the mitochondrial surface. Then, two mitochondrial membrane proteins, Tb11.01.4590 and Tb09.v1.0420, together with Tim17 and Hsp70, are part of subsequent tRNA import machinery in the inner membrane showing that tRNA and protein import mechanisms share components [104]. Finally, in plants some TOM proteins are believed to act as tRNA receptors on the mitochondrial surface, while the voltage dependent anion channel (VDAC) may be the main translocation channel through the outer mitochondrial membrane [105]. In all of these systems tRNAs interact with β -barrel forming proteins such as VDAC, TOM40 or other components of the protein import machinery, presumably allowing them to cross mitochondrial membranes. In all cases, structural modeling predicts that mature tRNA have to undergo either complete or partial unfolding, or at least major structural rearrangements in order to cross mitochondrial membranes. The determination of structures of these factors in complex with tRNA will reveal the mechanism of tRNA import at atomic resolution and, more generally, will help scientists to understand how tRNA import can be fine-tuned to attain the pool of tRNA required for mitochondrial translation.

6.3. Understanding the biogenesis and functions of tRNA fragments

Transfer RNAs are key players in a variety of molecular processes, including, but not limited to, translation. For these crucial functions, tRNAs have to be quality-controlled and their turn-over regulated. Examples of this regulation include two tRNA surveillance pathways identified in yeast that degrade hypo-modified or mis-processed pre-tRNAs, as well as mature tRNAs lacking modifications. A nuclear tRNA surveillance pathway involving polyadenylation by the TRAMP complex and degradation by the exosome deals primarily with hypo-modified tRNAs, while another pathway termed the "rapid tRNA decay" pathway involves a CCaase and degradation by Xrn1 [106].

Beyond tRNA degradation pathways, a number of other cellular activities generate tRNA-derived fragments (tRFs). The wide accumulation of numerous tRFs suggests that these tRNA cleavage products potentially represent a new class of small non-coding RNAs that may have important regulatory functions. tRFs are evolutionarily widespread and created by cleavage at a number of specific positions most probably by different endonuclease activities (Fig. 2D). The discovery of small non-coding RNAs such as microRNA (miRNA) and small interfering RNA (siRNA) and their functions as major regulators of gene expression has been a major breakthrough [107]. Recent analysis of transcriptomes in all three domains of life has suggested that tRFs constitute another class of small regulatory RNAs with biologically relevant functions because (i) in most cases tRFs are not generated from abundant tRNAs, and the amount of tRFs does not correlate with initial tRNA abundance or with the corresponding tRNA gene copy number. (ii) The cleavage patterns of tRFs seem to be dependent on tRNA anticodons. (iii) Cleavage patterns of tRNAs often vary according to the cell status and/or developmental stages. (iv) Some tRFs have been found in association with key components of the RNA silencing machineries such as Argonaute or Piwi. The functions of tRFs are related to a rapidly growing number of molecular processes, including translation, translation initiation, RNA degradation, reverse transcription, gene silencing, vacuolar ribophagy and cell survival [108]. The

determination of three-dimensional structures of tRFs alone and in complex with protein partners will aid in understanding how tRFs are generated as well as the mode of action and functional diversity of this novel class of non coding RNAs [108,109].

6.4. From tRNA to systems biology

The early observation of tRNAs capable of decoding nonsense UGA, UAA and UAG codons in *E. coli* [110], the reattribution of UGA to Phe in yeast mitochondria [111], and more recently the discovery of the 21st and 22nd amino acids, selenocysteine (Sec) and pyrrolysine (Pyl), coded by UGA and UAG [112,113], have contributed to the idea of utilizing nonsense codons to incorporate non-canonical amino acids (ncAAs) into proteins. The expansion of genetic code in vivo requires the introduction of an orthogonal pair consisting of a suppressor tRNA decoding a nonsense codon and a modified cognate synthetase able to transfer the desired ncAA without interfering with tRNAs of the host organism. This has been achieved in a variety of ways, including the selection of variants of archeal TyrRS [114,115] or PylRS [116,117] activating ncAAs. For optimal incorporation further engineering of host organism factors such as the elongation and release factors, or the peptidyl-transferase site of the ribosome, may be necessary. More than one hundred such compounds have already been successfully incorporated into proteins, introducing new functional groups for cross-linking or fluorescent labeling purposes or to mimic epigenetic modifications [118,119]. In an alternative approach, RNA aptamers known as flexizymes have been designed to aminoacylate any tRNA with ncAAs in cell free translation systems [120]. This system can be used to reprogram the code and introduce several ncAAs in a protein. These emerging strategies of code expansion or reprogramming by reassigning rare codons to ncAAs open up vast perspectives for the production of artificial proteins or more generally for the design of new organisms in the field of synthetic biology [121,122].

7. Concluding remarks

Forty years after the determination of the first tRNA crystal structure, the adaptor molecule has not yet revealed all its secrets. While at the end of the 90s, investigations carried out with model organisms may have given the feeling that the major aspects of tRNA:synthetase interactions had been revealed, the systematic exploration of genomes since the early 2000 and an increasing interest in pathology-related mutations, both in tRNAs and associated proteins, have revitalized the field and led to the discovery of many new aspects of tRNA biology. With the latest progress in cryoEM, one can easily anticipate a burst of new data describing how tRNAs move on the ribosome and how tRNA dynamics contributes to the decoding of mRNA from bacteria to higher eukaryotes, including cytosolic, mitochondrial and plastidial translation machineries. The integration of structural data from high resolution crystallography, cryoEM and correlative microscopy in the cell will provide a much deeper insight into their distribution and functions with applications in human medicine, drug design and synthetic biology, making the field of tRNA structural biology more active and open than ever.

Acknowledgments

The authors wish to acknowledge the contribution of Dr Richard Giegé in sharing his expertise on the history of tRNA biology and personal research artifacts, such as tRNA models, photographs, and printed electron density maps. This work was supported by the French “Centre National de la Recherche Scientifique”, by the

University of Strasbourg, an ANR Blanc research grant “PRO-RNase P” (ANR-11-BSV8-008-01/02) to PG and CS and by the LabEx consortium “MitoCross” in the frame of the French National Program “Investissements d’Avenir” (ANR-11-LABX-0057_MITOCROSS).

References

- [1] R. Giegé, C. Sauter, Biocrystallography: past, present, future, HFSP J. 4 (2010) 109–121.
- [2] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid, Nature 171 (1953) 737–738.
- [3] F.H.C. Crick, L. Barnett, S. Brenner, R.J. Watts-Tobin, General nature of the genetic code for proteins, Nature 192 (1961) 1227–1232.
- [4] J.H. Matthaei, O.W. Jones, R.G. Martin, M.W. Nirenberg, Characteristics and composition of RNA coding units, Proc. Natl. Acad. Sci. U. S. A. 48 (1962) 666–677.
- [5] F. Chapeville, F. Lipmann, G. von Ehrenstein, B. Weisblum, W.J. Ray, S. Benzer, On the role of soluble ribonucleic acid in coding for amino acids, Proc. Natl. Acad. Sci. U. S. A. 48 (1962) 1086–1092.
- [6] R.W. Holley, J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrill, J.R. Penswick, A. Zamir, Structure of a ribonucleic acid, Science 147 (1965) 1462–1465.
- [7] J.T. Madison, G.A. Everett, H. Kung, Nucleotide sequence of a yeast tyrosine transfer RNA, Science 153 (1966) 531–534.
- [8] U.L. Rajbhandary, S.H. Chang, A. Stuart, R.D. Faulkner, R.M. Hoskinson, H.G. Khorana, Studies on polynucleotides, LXVIII: the primary structure of yeast phenylalanine transfer RNA, Proc. Natl. Acad. Sci. U. S. A. 57 (1967) 751.
- [9] H.G. Zachau, D. Dütting, H. Feldmann, The structures of two serine transfer ribonucleic acids. Hoppe-Seyler's Z. Für Physiol. Chem. 347 (1966) 212–235.
- [10] S.H. Kim, F.L. Suddath, G.J. Quigley, A. McPherson, J.L. Sussman, A.H.J. Wang, N.C. Seeman, A. Rich, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, Science 185 (1974) 435–440.
- [11] J.D. Robertus, J.E. Ladner, J.T. Finch, D. Rhodes, R.S. Brown, B.F.C. Clark, A. Klug, Structure of yeast phenylalanine tRNA at 3 Å resolution, Nature 250 (1974) 546–551.
- [12] C.D. Stout, H. Mizuno, J. Rubin, T. Brennan, S.T. Rao, M. Sundaralingam, Atomic coordinates and molecular conformation of yeast phenylalanyl tRNA. An independent investigation, Nucleic Acids Res. 3 (1976) 1111–1123.
- [13] D. Moras, M.B. Comarmond, J. Fischer, R. Weiss, J.C. Thierry, J.P. Ebel, R. Giegé, Crystal structure of yeast tRNA(Asp), Nature 288 (1980) 669–674.
- [14] M. Ibbá, C. Francklyn, S. Cusack, Aminoacyl-tRNA Synthetases (Georgetown, TX: Landes Biosciences), 2005.
- [15] A. Korostelev, S. Trakhanov, M. Laurberg, H.F. Noller, Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements, Cell 126 (2006) 1065–1077.
- [16] M.Z. Barciszewska, P.M. Perriguet, J. Barciszewski, tRNA – the golden standard in molecular biology, Mol. Biosyst. 12 (2016) 12–17.
- [17] B.P. Doctor, J. Apgar, R.W. Holley, Fractionation of yeast amino acid-acceptor ribonucleic acids by countercurrent distribution, J. Biol. Chem. 236 (1961) 1117–1120.
- [18] A. Hampel, M. Labanauskas, P.G. Connors, L. Kirkegard, U.L. Rajbhandary, P.B. Sigler, R.M. Bock, Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's, Science 162 (1968) 1384–1387.
- [19] R. Giegé, E. Touze, B. Lorber, A. Theobald-Dietrich, C. Sauter, Crystallogenic trends of free and liganded aminoacyl-tRNA synthetases, Cryst. Growth Des. 8 (2008) 4297–4306.
- [20] M. Levitt, Detailed molecular model for transfer ribonucleic acid, Nature 224 (1969) 759–763.
- [21] J. Witz, 1964: the first model for the shape of a transfer RNA molecule. An account of an unpublished small-angle X-ray scattering study, Biochimie 85 (2003) 1265–1268.
- [22] F. Jühling, M. Mörl, R.K. Hartmann, M. Sprinzl, P.F. Stadler, J. Pütz, tRNAdb 2009: compilation of tRNA sequences and tRNA genes, Nucleic Acids Res. 37 (2009) D159–D162.
- [23] S. Wende, E.G. Platzer, F. Jühling, J. Pütz, C. Florentz, P.F. Stadler, M. Mörl, Biological evidence for the world's smallest tRNAs, Biochimie 100 (2014) 151–158.
- [24] T. Suzuki, T. Suzuki, A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs, Nucleic Acids Res. 42 (2014) 7346–7357.
- [25] B. El Yacoubi, M. Bailly, V. de Crécy-Lagard, Biosynthesis and function of posttranscriptional modifications of transfer RNAs, Annu. Rev. Genet. 46 (2012) 69–95.
- [26] M. Helm, R. Giegé, C. Florentz, A Watson–Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNA-Lys, Biochemistry 38 (1999) 13338–13346.
- [27] W.A. Cantara, P.F. Crain, J. Rozenski, J.A. McCloskey, K.A. Harris, X. Zhang, F.A.P. Vendeix, D. Fabris, P.F. Agris, The RNA modification database, RNAMDB: 2011 update, Nucleic Acids Res. 39 (2011) D195–D201.
- [28] M.A. Machnicka, K. Milanowska, O.O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K.M. Rother, et al., MODOMICS: a database of RNA modification pathways—2013 update, Nucleic Acids Res. 41 (2013) D262–D267.

- [29] M.A. Rould, J.J. Perona, D. Söll, T.A. Steitz, Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA-Gln and ATP at 2.8 Å resolution, *Science* 246 (1989) 1135–1142.
- [30] M. Ruff, S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J.C. Thierry, D. Moras, Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA (Asp), *Science* 252 (1991) 1682–1689.
- [31] G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs, *Nature* 347 (1990) 203–206.
- [32] R. Giegé, M. Sissler, C. Florentz, Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res.* 26 (1998) 5017–5035.
- [33] M.A. Rould, J.J. Perona, T.A. Steitz, Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase, *Nature* 352 (1991) 213–218.
- [34] M. Naganuma, S. Sekine, Y.E. Chong, M. Guo, X.-L. Yang, H. Gamper, Y.-M. Hou, P. Schimmel, S. Yokoyama, The selective tRNA aminoacylation mechanism based on a single G-U pair, *Nature* 510 (2014) 507–511.
- [35] V. Biou, A. Yaremchuk, M. Tukalo, S. Cusack, The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser), *Science* 263 (1994) 1404–1410.
- [36] S. Hauenstein, C.-M. Zhang, Y.-M. Hou, J.J. Perona, Shape-selective RNA recognition by cysteinyl-tRNA synthetase, *Nat. Struct. Mol. Biol.* 11 (2004) 1134–1141.
- [37] R. Giegé, Toward a more complete view of tRNA biology, *Nat. Struct. Mol. Biol.* 15 (2008) 1007–1014.
- [38] M. Guo, P. Schimmel, Structural analyses clarify the complex control of mistranslation by tRNA synthetases, *Curr. Opin. Struct. Biol.* 22 (2012) 119–126.
- [39] J.A. Abbott, C.S. Francklyn, S.M. Robey-Bond, Transfer RNA and human disease, *Front. Genet.* 5 (2014) 158.
- [40] H. Schwenzler, J. Zoll, C. Florentz, M. Sissler, Pathogenic implications of human mitochondrial aminoacyl-tRNA synthetases, *Top. Curr. Chem.* 344 (2014) 247–292.
- [41] F. Michel, E. Westhof, Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *J. Mol. Biol.* 216 (1990) 585–610.
- [42] J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, et al., Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science* 273 (1996) 1678–1684.
- [43] H.W. Pley, D.S. Lindes, C. DeLuca-Flaherty, D.B. McKay, Crystals of hammerhead ribozyme, *J. Biol. Chem.* 268 (1993) 19656–19658.
- [44] W.G. Scott, J.T. Finch, R. Grenfell, J. Fogg, T. Smith, M.J. Gait, A. Klug, Rapid crystallization of chemically synthesized Hammerhead RNAs using a double screening procedure, *J. Mol. Biol.* 250 (1995) 327–332.
- [45] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* 289 (2000) 905–920.
- [46] B.T. Wimberly, D.E. Brodersen, W.M. Clemons, R.J. Morgan-Warren, A.P. Carter, C. Vonnrhein, T. Hartsch, V. Ramakrishnan, Structure of the 30S ribosomal subunit, *Nature* 407 (2000) 327–339.
- [47] M.M. Yusupov, G.Z. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H. Cate, H.F. Noller, Crystal structure of the ribosome at 5.5 Å resolution, *Science* 292 (2001) 883–896.
- [48] D.J. Klein, T.M. Schmeing, P.B. Moore, T.A. Steitz, The kink-turn: a new RNA secondary structure motif, *EMBO J.* 20 (2001) 4214–4221.
- [49] M. Meyer, E. Westhof, B. Masquida, A structural module in RNase P expands the variety of RNA kinks, *RNA Biol.* 9 (2012) 254–260.
- [50] A. Lescoute, N.B. Leontis, C. Massire, E. Westhof, Recurrent structural RNA motifs, isostericity matrices and sequence alignments, *Nucleic Acids Res.* 33 (2005) 2395–2409.
- [51] L. Jaeger, E.J. Verzemnieks, C. Geary, The UA₂ handle: a versatile submotif in stable RNA architectures, *Nucleic Acids Res.* 37 (2009) 215–230.
- [52] E.A. Doherty, R.T. Batey, B. Masquida, J.A. Doudna, A universal mode of helix packing in RNA, *Nat. Struct. Mol. Biol.* 8 (2001) 339–343.
- [53] P. Nissen, J.A. Ippolito, N. Ban, P.B. Moore, T.A. Steitz, RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 4899–4903.
- [54] B. Masquida, B. Beckert, F. Jossinet, Exploring RNA structure by integrative molecular modelling, *New Biotechnol.* 27 (2010) 170–183.
- [55] M. Meyer, H. Nielsen, V. Oliéric, P. Roblin, S.D. Johansen, E. Westhof, B. Masquida, Speciation of a group I intron into a lariat capping ribozyme, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 7659–7664.
- [56] N.J. Reiter, A. Osterman, A. Torres-Larios, K.K. Swinger, T. Pan, A. Mondragón, Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA, *Nature* 468 (2010) 784–789.
- [57] J. Zhang, A.R. Ferré-D'Amaré, Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA, *Nature* 500 (2013) 363–366.
- [58] J. Lehmann, F. Jossinet, D. Gautheret, A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNase P and T-box leaders, *Nucleic Acids Res.* 41 (2013) 5494–5502.
- [59] C.Y. Cheng, F.-C. Chou, W. Kladwang, S. Tian, P. Cordero, R. Das, Consistent global structures of complex RNA states through multidimensional chemical mapping, *eLife* 4 (2015) e07600.
- [60] O.M. Ganichkin, E.A. Anedchenko, M.C. Wahl, Crystal structure analysis reveals functional flexibility in the selenocysteine-specific tRNA from mouse, *PLoS ONE* 6 (2011) e20032.
- [61] T.M. Colussi, D.A. Costantino, J.A. Hammond, G.M. Ruehle, J.C. Nix, J.S. Kieft, The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA, *Nature* 511 (2014) 366–369.
- [62] M. Blaise, M. Bailly, M. Frechin, M.A. Behrens, F. Fischer, C.L. Oliveira, H.D. Becker, J.S. Pedersen, S. Thirup, D. Kern, Crystal structure of a transfer-ribonucleoprotein particle that promotes asparagine formation, *EMBO J.* 29 (2010) 3118–3129.
- [63] P. Nissen, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B.F. Clark, J. Nyborg, Crystal structure of the ternary complex of Phe-tRNA(Phe), EF-Tu, and a GTP analog, *Science* 270 (1995) 1464–1472.
- [64] E. Schmitt, M. Panvert, C. Lazennec-Schurdevin, P.-D. Coureux, J. Perez, A. Thompson, Y. Mechulam, Structure of the ternary initiation complex alf2-GDPNP-methionylated initiator tRNA, *Nat. Struct. Mol. Biol.* 19 (2012) 450–454.
- [65] T.M. Schmeing, R.M. Voorhees, A.C. Kelley, Y.-G. Gao, F.V. Murphy, J.R. Weir, V. Ramakrishnan, The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA, *Science* 326 (2009) 688–694.
- [66] O. Pellegrini, I. Li de la Sierra-Gallay, J. Pilon, L. Gilet, C. Condon, Activation of tRNA maturation by downstream uracil residues in *B. subtilis*, *Structure* 20 (2012) 1769–1777.
- [67] R. Ishitani, O. Nureki, N. Nameki, N. Okada, S. Nishimura, S. Yokoyama, Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme, *Cell* 113 (2003) 383–394.
- [68] K. Tomita, S. Fukai, R. Ishitani, T. Ueda, N. Takeuchi, D.G. Vassilyev, O. Nureki, Structural basis for template-independent RNA polymerization, *Nature* 430 (2004) 700–704.
- [69] F. Pinker, G. Bonnard, A. Gobert, B. Gutmann, K. Hammani, C. Sauter, P.A. Gegenheimer, P. Giegé, PPR proteins shed a new light on RNase P biology, *RNA Biol.* 10 (2013) 1457–1468.
- [70] A. Gobert, F. Pinker, O. Fuchsbaue, B. Gutmann, R. Boutin, P. Roblin, C. Sauter, P. Giegé, Structural insights into protein-only RNase P complexed with tRNA, *Nat. Commun.* 4 (2013) 1353.
- [71] H. Betat, Y. Long, J.E. Jackman, M. Mörl, From end to end: tRNA editing at 5'- and 3'-terminal positions, *Int. J. Mol. Sci.* 15 (2014) 23975–23998.
- [72] A. Guelorget, B. Golinelli-Pimpaneau, Mechanism-based strategies for trapping and crystallizing complexes of RNA-modifying enzymes, *Structure* 19 (2011) 282–291.
- [73] S. Goto-Ito, T. Ito, M. Kuratani, Y. Bessho, S. Yokoyama, Tertiary structure checkpoint at anticodon loop modification in tRNA functional maturation, *Nat. Struct. Mol. Biol.* 16 (2009) 1109–1115.
- [74] B. Pan, Y. Xiong, T.A. Steitz, How the CCA-adding enzyme selects adenine over cytosine at position 76 of tRNA, *Science* 330 (2010) 937–940.
- [75] V. Ramakrishnan, The ribosome emerges from a black box, *Cell* 159 (2014) 979–984.
- [76] X. Agirrezabala, M. Valle, Structural insights into tRNA dynamics on the ribosome, *Int. J. Mol. Sci.* 16 (2015) 9866–9895.
- [77] S. Uemura, C.E. Aitken, J. Korch, B.A. Flusberg, S.W. Turner, J.D. Puglisi, Real-time tRNA transit on single translating ribosomes at codon resolution, *Nature* 464 (2010) 1012–1017.
- [78] E. Schmitt, M. Panvert, S. Blanquet, Y. Mechulam, Crystal structure of methionyl-tRNA^{Met} transformylase complexed with the initiator formyl-methionyl-tRNA^{Met}, *EMBO J.* 17 (1998) 6819–6826.
- [79] P. Barraud, E. Schmitt, Y. Mechulam, F. Dardel, C. Tisne, A unique conformation of the anticodon stem-loop is associated with the capacity of tRNA^{Met} to initiate protein synthesis, *Nucleic Acids Res.* 36 (2008) 4894–4901.
- [80] A. Tsai, A. Petrov, R.A. Marshall, J. Korch, S. Uemura, J.D. Puglisi, Heterogeneous pathways and timing of factor departure during translation initiation, *Nature* 487 (2012) 390–393.
- [81] N. Sonenberg, A.G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets, *Cell* 136 (2009) 731–745.
- [82] A.G. Hinnebusch, The scanning mechanism of eukaryotic translation initiation, *Annu. Rev. Biochem.* 83 (2014) 779–812.
- [83] P. Julián, P. Milon, X. Agirrezabala, G. Lasso, D. Gil, M.V. Rodnina, M. Valle, The cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*, *PLoS Biol.* 9 (2011) e1001095.
- [84] J.M. Ogle, F.V. Murphy IV, M.J. Tarry, V. Ramakrishnan, Selection of tRNA by the ribosome requires a transition from an open to a closed form, *Cell* 111 (2002) 721–732.
- [85] N. Demeshkina, L. Jenner, E. Westhof, M. Yusupov, G. Yusupova, A new understanding of the decoding principle on the ribosome, *Nature* 484 (2012) 256–259.
- [86] T.M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation, *Nature* 461 (2009) 1234–1242.
- [87] C.E. Aitken, J.D. Puglisi, Following the intersubunit conformation of the ribosome during translation in real time, *Nat. Struct. Mol. Biol.* 17 (2010) 793–800.
- [88] S. Kuhlkoetter, W. Wintermeyer, M.V. Rodnina, Different substrate-dependent transition states in the active site of the ribosome, *Nature* 476 (2011) 351–354.
- [89] D. Moazed, H.F. Noller, Intermediate states in the movement of transfer RNA in the ribosome, *Nature* 342 (1989) 142–148.
- [90] X. Agirrezabala, H.Y. Liao, E. Schreiner, J. Fu, R.F. Ortiz-Meoz, K. Schulten, R. Green, J. Frank, Structural characterization of mRNA-tRNA translocation

- intermediates, *Proc. Natl. Acad. Sci.* 109 (2012) 6094–6099.
- [91] N. Fischer, A.L. Konevega, W. Wintermeyer, M.V. Rodnina, H. Stark, Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy, *Nature* 466 (2010) 329–333.
- [92] A. Amunts, A. Brown, J. Toots, S.H. Scheres, V. Ramakrishnan, The structure of the human mitochondrial ribosome, *Science* 348 (2015) 95–98.
- [93] B.J. Greber, P. Bieri, M. Leibundgut, A. Leitner, R. Aebersold, D. Boehringer, N. Ban, The complete structure of the 55S mammalian mitochondrial ribosome, *Science* 348 (2015) 303–308.
- [94] S.F.J. Le Grice, “In the beginning”: initiation of minus strand DNA synthesis in retroviruses and LTR-containing retrotransposons, *Biochemistry* 42 (2003) 14349–14355.
- [95] T.W. Dreher, *Viral tRNAs and tRNA-like structures*, Wiley Interdiscip. Rev. RNA 1 (2010) 402–414.
- [96] J.C. Grigg, A. Ke, Sequence, structure, and stacking: specifics of tRNA anchoring to the T box riboswitch, *RNA Biol.* 10 (2013) 1761–1764.
- [97] A. Gobert, B. Gutmann, A. Taschner, M. Gößringer, J. Holzmann, R.K. Hartmann, W. Rossmann, P. Giegé, A single *Arabidopsis* organellar protein has RNase P activity, *Nat. Struct. Mol. Biol.* 17 (2010) 740–744.
- [98] B. Gutmann, A. Gobert, P. Giegé, PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*, *Genes Dev.* 26 (2012) 1022–1027.
- [99] M. Sissler, M. Helm, M. Frugier, R. Giegé, C. Florentz, Aminoacylation properties of pathology-related human mitochondrial tRNA-Lys variants, *RNA* 10 (2004) 841–853.
- [100] A. Neuenfeldt, B. Lorber, E. Ennifar, A. Gaudry, C. Sauter, M. Sissler, C. Florentz, Thermodynamic properties distinguish human mitochondrial aspartyl-tRNA synthetase from bacterial homolog with same 3D architecture, *Nucleic Acids Res.* 41 (2013) 2698–2708.
- [101] N.S. Entelis, S. Kieffer, O.A. Kolesnikova, R.P. Martin, I.A. Tarassov, Structural requirements of tRNA-Lys for its import into yeast mitochondria, *Proc. Natl. Acad. Sci.* 95 (1998) 2838–2843.
- [102] T. Salinas-Giegé, R. Giegé, P. Giegé, tRNA biology in mitochondria, *Int. J. Mol. Sci.* 16 (2015) 4518–4559.
- [103] M. Baleva, A. Gowher, P. Kamenski, I. Tarassov, N. Entelis, B. Masquida, A moonlighting human protein is involved in mitochondrial import of tRNA, *Int. J. Mol. Sci.* 16 (2015) 9354–9367.
- [104] D. Seidman, D. Johnson, V. Gerbasí, D. Golden, R. Orlando, S. Hajduk, Mitochondrial membrane complex that contains proteins necessary for tRNA import in *Trypanosoma brucei*, *J. Biol. Chem.* 287 (2012) 8892–8903.
- [105] T. Salinas, A.-M. Duchêne, L. Maréchal-Drouard, Recent advances in tRNA mitochondrial import, *Trends Biochem. Sci.* 33 (2008) 320–329.
- [106] J.E. Wilusz, J.M. Whipple, E.M. Phizicky, P.A. Sharp, tRNAs marked with CCACA are targeted for degradation, *Science* 334 (2011) 817–821.
- [107] S. Jonas, E. Izaurralde, Towards a molecular understanding of microRNA-mediated gene silencing, *Nat. Rev. Genet.* 16 (2015) 421–433.
- [108] C. Megel, G. Morelle, S. Lalande, A.-M. Duchêne, I. Small, L. Maréchal-Drouard, Surveillance and cleavage of eukaryotic tRNAs, *Int. J. Mol. Sci.* 16 (2015) 1873–1893.
- [109] A. Kanai, Disrupted tRNA genes and tRNA fragments: a perspective on tRNA gene evolution, *Life* 5 (2015) 321–331.
- [110] G. Eggertsson, D. Söll, Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*, *Microbiol. Rev.* 52 (1988) 354–374.
- [111] G. Macino, G. Coruzzi, F.G. Nobrega, M. Li, A. Tzagoloff, Use of the UGA terminator as a tryptophan codon in yeast mitochondria, *Proc. Natl. Acad. Sci. U. S. A.* 76 (1979) 3784–3785.
- [112] A. Schön, A. Böck, G. Ott, M. Sprinzl, D. Söll, The selenocysteine-inserting opal suppressor serine tRNA from *E. coli* is highly unusual in structure and modification, *Nucleic Acids Res.* 17 (1989) 7159–7165.
- [113] G. Srinivasan, C.M. James, J.A. Krzycki, Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA, *Science* 296 (2002) 1459–1462.
- [114] K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, et al., Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* 30 (2002) 4692–4699.
- [115] L. Wang, T.J. Magliery, D.R. Liu, P.G. Schultz, A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins, *J. Am. Chem. Soc.* 122 (2000) 5010–5011.
- [116] C.R. Polycarpo, S. Herring, A. Bérubé, J.L. Wood, D. Söll, A. Ambrogelly, Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase, *FEBS Lett.* 580 (2006) 6695–6700.
- [117] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N ϵ -(*o*-azidobenzoyloxycarbonyl) lysine for site-specific protein modification, *Chem. Biol.* 15 (2008) 1187–1197.
- [118] J.A. Johnson, Y.Y. Lu, J.A. Van Deventer, D.A. Tirrell, Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications, *Curr. Opin. Chem. Biol.* 14 (2010) 774–780.
- [119] T. Mukai, T. Yanagisawa, K. Ohtake, M. Wakamori, J. Adachi, N. Hino, A. Sato, T. Kobayashi, A. Hayashi, M. Shirouzu, et al., Genetic-code evolution for protein synthesis with non-natural amino acids, *Biochem. Biophys. Res. Commun.* 411 (2011) 757–761.
- [120] H. Murakami, A. Ohta, H. Ashigai, H. Suga, A highly flexible tRNA acylation method for non-natural polypeptide synthesis, *Nat. Methods* 3 (2006) 357–359.
- [121] D. Söll, A tRNA-guided research journey from synthetic chemistry to synthetic biology, *RNA* 21 (2015) 742–744.
- [122] N. Terasaka, Y. Iwane, A.-S. Geiermann, Y. Goto, H. Suga, Recent developments of engineered translational machineries for the incorporation of non-canonical amino acids into polypeptides, *Int. J. Mol. Sci.* 16 (2015) 6513–6531.

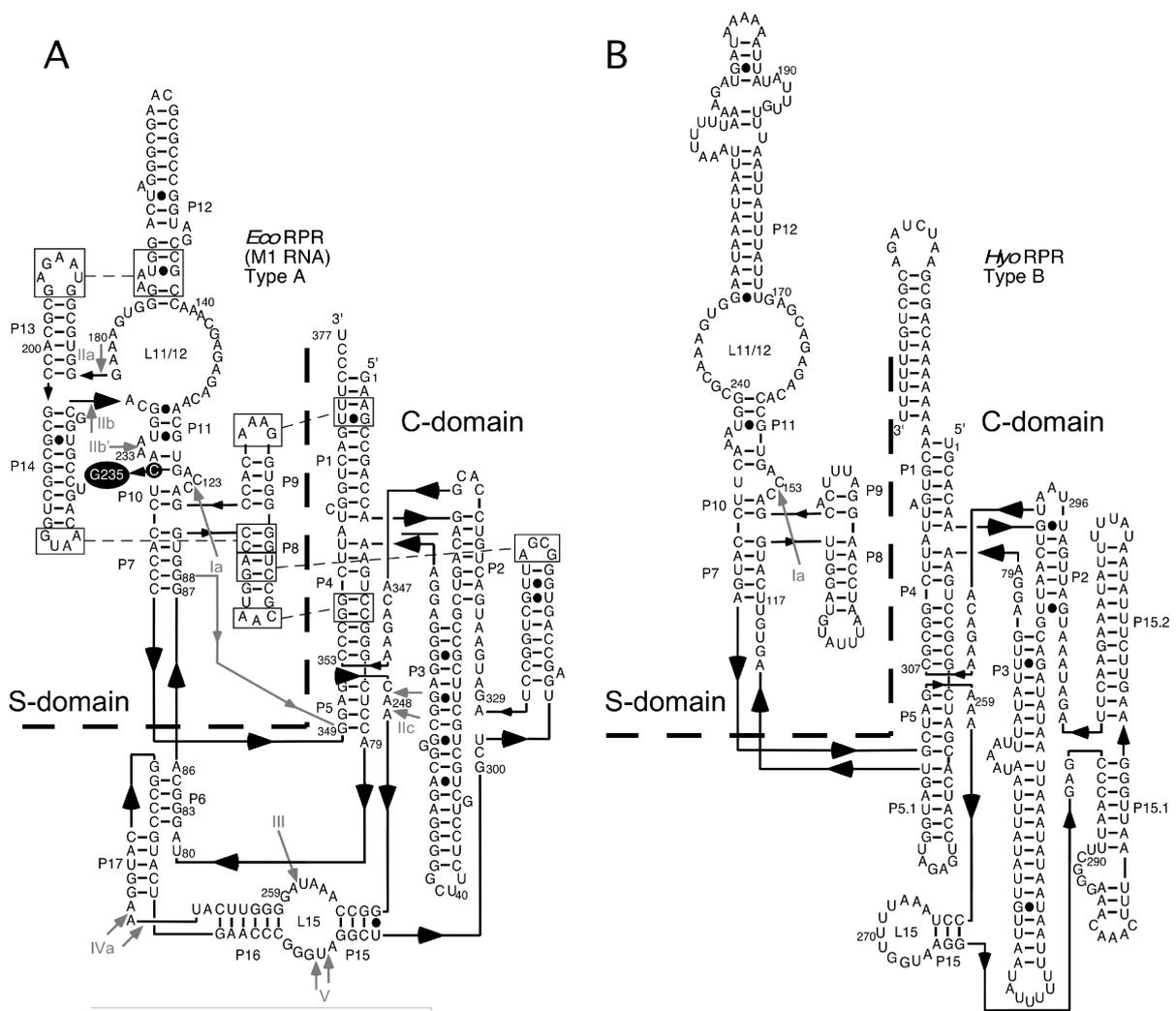


Figure 12 : Structures des RNase P RNP de type A et B.

Représentation des structures secondaires de l'ARN des RNase P RNP de type "Ancestral" (A) comme trouvé chez *E. coli* et de type "Bacillus" comme trouvé chez *Mycoplasma hyopneumoniae* (Hyo) (B), mettant en évidence les domaines de spécificité (S) et les domaines catalytiques (C) ainsi que les domaines conservés P1 à P15 pour le type B et P1 à P17 pour le type A. Figure issue de Wu et al. 2016.

4. Les enzymes à activité RNases P

a. Les RNase P ribonucléoprotéiques

Jusqu'aux années 2000, il était admis que les RNases P, dans tous les domaines du vivant, étaient des particules ribonucléoprotéiques (RNP) composées d'une sous-unité ARN portant l'activité catalytique (c'est-à-dire un ribozyme) et d'une ou plusieurs sous-unités protéiques (Lai et al., 2010; Guerrier-Takada et al., 1983). Cette structure est conservée chez tous les procaryotes (bactéries et archées) avec une sous-unité protéique unique chez les bactéries et un maximum de 3 à 4 sous-unités protéiques chez les archées. Dans le noyau des eucaryotes les RNases P de type RNP sont plus complexes avec la présence d'une dizaine de sous-unités protéiques (Kouzuma et al., 2003; Kazantsev and Pace, 2006; Lai et al., 2010). Dans les organites, les exemples de RNases P RNP identifiées, notamment chez les levures, révèlent souvent une structure d'ARN catalytique très dégénérée et de taille réduite ainsi que des sous-unités protéiques sans homologie avec celles des RNases P RNP nucléaires (Seif et al., 2003).

Chez les bactéries, l'ARN de la RNase P RNP se compose de 300 à 450 nucléotides et peut être divisé en deux types majeurs : A pour « Ancestral » et B pour « Bacillus ». L'ARN P est composé de deux domaines indépendamment structurés. Les deux domaines structuraux de l'ARN P bactérienne sont responsables de différentes activités : le domaine de spécificité (S) reconnaît la boucle T Ψ C du pré-ARNt, tandis que le domaine catalytique (C) reconnaît la tige acceptrice et la séquence 3'CCA et catalyse le clivage de la séquence 5' leader du pré-ARNt. La comparaison des structures cristallines du domaine S des ARN P de *Bacillus subtilis* (type B) et de *Thermus thermophilus* (type A) a révélé que, bien que les structures secondaires et tertiaires soient différentes, une ouverture est conservée dans le domaine S (bordé par les nucléotides connus pour impliquer des interactions directes avec la boucle T Ψ C de l'ARNt). La région conservée dans les deux structures de domaine S est formée par un module structural non hélicoïdal comprenant les connecteurs J11 / 12 et J12 / 11 ainsi que les tiges P9, P10 et P11. L'orientation similaire de ces éléments place les nucléotides interagissant avec le substrat, A108 de P9, A226 de P11 et A172 à partir de J12 / 11 (numération *T. thermophilus*), dans des positions presque identiques. Cette conformation est maintenue par les structures adjacentes. La structure de type A est stabilisée par deux interactions entre P14 et P8 et entre P12 et P13. Les ARN P de type B n'ont pas de tiges P13 et P14, mais un autre élément hélicoïdal, P10.1. La structure de type B est stabilisée par une interaction du récepteur tetraloop-tetraloop entre P10.1 et L12, et une interaction A-mineure entre P10.1 et

P7 (Figure 12). Malgré des différences structurales, le mode de reconnaissance et de catalyse de ces deux types d'ARN sont similaires (Torres-Larios et al., 2006).

Plusieurs formes de RNase P ont été identifiées chez les eucaryotes. On trouve différentes RNase P dans les noyaux, les mitochondries et les chloroplastes. Des RNase P RNP nucléaires ont été détectées dans une large gamme d'eucaryotes et ont été caractérisées en détails chez la levure et chez l'Homme (Eder et al., 1997; Chamberlain et al., 1998). La composition de l'holoenzyme RNase P nucléaire est beaucoup plus complexe que celle de l'enzyme bactérienne, avec une sous-unité ARN essentielle et une dizaine de sous-unités protéiques, dont de nombreuses protéines POP (Processing Of Precursor) ainsi que RPP21, la seule sous unité protéique entièrement spécifique de la RNase P, absente dans la RNase MRP (Lechner et al., 2015).

Des RNase P RNP sont trouvées dans les mitochondries et / ou les plastes de certains eucaryotes. Les mitochondries et les plastes possèdent leur propre génome codant pour un ensemble partiel ou complet d'ARNt. Ils sont issus de l'endosymbiose primaire avec une α -proteo-bactérie ancestrale et secondaire avec une cyanobactérie, respectivement, les plastes pouvant provenir d'événements d'endosymbiose tertiaire dans certains groupes eucaryotes. Il n'est donc pas surprenant de trouver des ARN P de type bactérien encore codés dans certains génomes des organites. C'est le cas chez les champignons et des algues basales incluant des Glaucophytes, des Rhodophytes et des Chlorophytes. Dans toutes ces espèces, les ARN de la RNase P sont très dégénérés et souvent très réduits, en particulier chez les champignons (Seif et al., 2005; Rossmannith, 2012; Lechner et al., 2015). Les sous unités protéiques de ces RNase P des organites sont très peu connues, mais semblent se distinguer fondamentalement des sous-unités trouvées dans les enzymes nucléaires (Lechner et al., 2015).

b. Les RNases P protéiques

L'identification de nombreuses RNases P de type RNP tant chez les procaryotes que chez les eucaryotes, dans le noyau et les organites a mené à un dogme proposant que l'activité RNase P soit universellement portée par des RNP. La RNase P serait donc un vestige universellement conservé du "monde ARN" prébiotique (Altman, 2007). Toutefois des analyses biochimiques réalisées sur des chloroplastes d'épinard et des mitochondries humaines ont contredit ce dogme en proposant qu'un autre type de RNase P uniquement protéique puisse exister (Wang et al., 1988; Pinker et al., 2013; Rossmannith and Karwan, 1998).

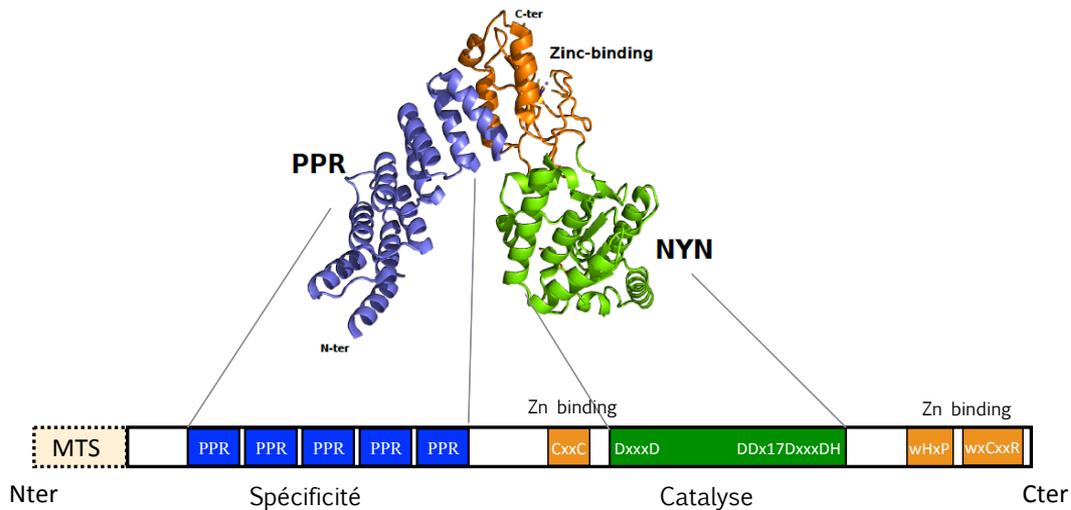


Figure 13 : les domaines de PRORP.

Les protéines PRORP sont composées de deux domaines principaux, le domaine PPR de liaison à l'ARN (en bleu) et le domaine catalytique NYN (en vert) portant l'activité endonucléase. Les deux domaines sont reliés par un domaine structural de liaison au zinc (en orange) qui constitue la charnière de la structure en forme de L des protéines PRORP. Les signaux d'adressage, aux organites ou au noyau sont présents aux extrémités N-terminales des enzymes.

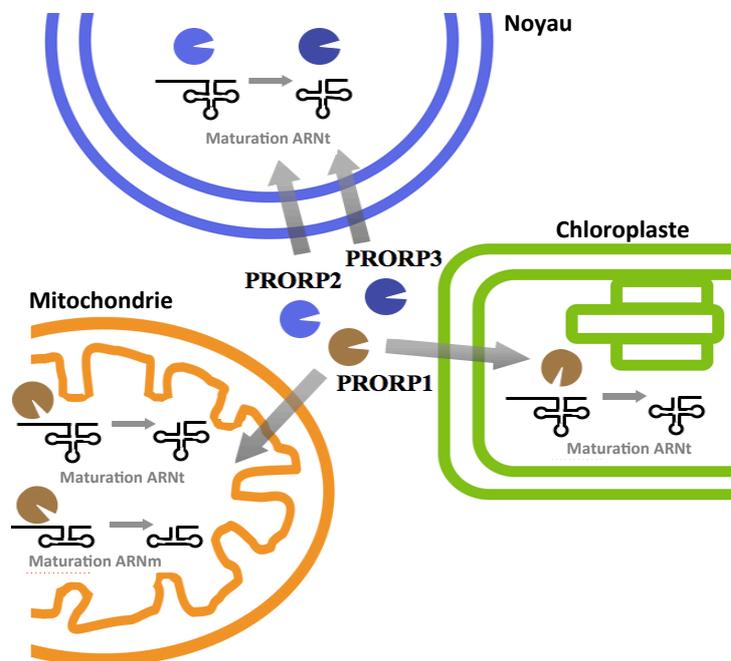


Figure 14 : Localisations et activités connues des protéines PRORP chez Arabidopsis.

Il existe 3 protéines PRORP chez Arabidopsis. PRORP1 est adressée aux mitochondries et aux chloroplastes où elle mature les ARNm mais également certains ARN messagers dans la mitochondrie. PRORP2 et PRORP3 sont toutes deux nucléaires et sont responsables de la maturation en 5' des ARNt.

Des travaux publiés en 2008 et 2010 sur des mitochondries humaines et les plantes ont depuis prouvé l'existence d'enzymes purement protéiques portant l'activité RNase P (Holzmann et al., 2008; Gobert et al., 2010). Ces enzymes appartiennent à la famille des protéines à pentatricopeptide repeat (PPR).

c. Définition et caractéristiques des RNases P protéiques

Les RNase P protéiques ont été nommés PRORP pour « protein-only RNase P ». Ces enzymes, spécifiques des eucaryotes, possèdent deux domaines principaux. Le premier contient un nombre variable de motifs PPR, généralement cinq. Le second contient un motif métallonucléase de type NYN utilisant deux ions magnésium comme cofacteurs et responsable de l'activité catalytique proprement dite de l'enzyme. Les deux domaines principaux, PPR et NYN sont reliés par un domaine structural de liaison au zinc (Gobert et al., 2013) (Figure 13). La fonction des motifs PPR serait de lier l'ARN et surtout de conférer à la protéine la spécificité pour ses substrats ARNt (Gutmann et al., 2012).

d. PRORP chez Arabidopsis thaliana

Il existe trois protéines PRORP différentes chez *Arabidopsis thaliana* appelées AtPRORP1, AtPRORP2 et AtPRORP3. Ces protéines ont été caractérisées dans l'équipe de P. Giegé. La détermination de la localisation par microscopie en utilisant plantes transformées avec les différentes protéines PRORP fusionnées avec l'EYFP ont montré que AtPRORP1 est localisée dans les mitochondries et les chloroplastes alors que AtPRORP2 et AtPRORP3 sont localisées dans le noyau (Gobert et al., 2010) (Figure 14). AtPRORP1 est un gène essentiel, il n'est pas possible d'obtenir un mutant knock-out viable car les embryons homozygotes mutants ne se développent pas dans la graine (Gobert et al., 2010). En revanche, les mutants knock-out individuels de AtPRORP2 et AtPRORP3 sont viables et ne montrent aucun phénotype macroscopique. En revanche, la double mutation de AtPRORP2 et AtPRORP3 résulte dans la létalité gamétophytique. Ainsi, il apparaît que les deux protéines AtPRORP2 et AtPRORP3 possèdent des fonctions redondantes et essentielles (Gutmann et al., 2012). En raison du caractère essentiel des gènes de PRORP, leurs fonctions ont été étudiées *in vivo* en dérégulant leurs niveaux d'expression par « virus induced gene silencing » (VIGS). Au bout d'une dizaine de jour, la dérégulation de AtPRORP1 induit l'apparition de zones de chlorose jaunes alors que la dérégulation de AtPRORP2 dans un fond génétique mutant pour *prop3* conduit à un blanchissement du système vasculaire et une senescence foliaire rapide

(Gutmann et al., 2012). L'analyse de l'accumulation des transcrits dans ces plantes dérégulées a permis de montrer une baisse de la maturation en 5' des ARNt nucléaires et des organites. Par ailleurs, elle a montré une accumulation réduite des ARNt Gly GCC et snoR43 provenant d'un même ARN polycistronique dans le noyau lors de la dérégulation de AtPRORP2 ainsi qu'une baisse de la quantité de l'ARNm *nad6* dans la mitochondrie lors de la dérégulation de AtPRORP1. Le transcrit de *nad6* comporte une structure ARNt-like qui est reconnue et clivée par AtPRORP1, ce qui montre la capacité de PRORP à maturer des transcrits différents des précurseurs d'ARNt canoniques (Gutmann et al., 2012). L'analyse du clivage d'une collection d'ARNt mutés a permis de démontrer l'importance des boucles T et D dans la reconnaissance des ARNt par PRORP. Ceci a été confirmé par des analyses de footprinting qui ont permis de mettre en évidence l'importance, pour l'interaction avec PRORP, de résidus universellement conservés dans les ARNt. Les résidus U16, G18, G19 et C56 sont localisés dans les boucles T et D et sont protégés par PRORP de la digestion par des RNases (Gobert et al., 2013). Ces résultats ont permis de proposer que la reconnaissance des ARNt par PRORP soit similaire à la reconnaissance des ARNt par la RNase P RNP.

e. Diversité des protéines PPR à activité RNase P

Comme c'est le cas pour l'ensemble des protéines PPR, les protéines PRORP sont spécifiques des eucaryotes. Elles ne sont pas universelles ou réparties de manière homogène dans les différents règnes mais sont toutefois très répandues. Elles sont trouvées dans les 4 super-groupes principaux des eucaryotes, i.e chez les Opisthokontes (dont font partie les animaux), chez les Archaeplastida (dont font partie les plantes), chez les Excavata (dont font partie les trypanosomes) et chez les SAR (dont font partie les diatomées ou les apicomplexes comme Plasmodium) (Lechner et al., 2015).

Chez les animaux, une seule protéine PRORP localisée dans les mitochondries est systématiquement présente. Initialement appelée MRPP3 pour « Mitochondrial ribonuclease P protein 3 » cette protéine nécessite la présence de partenaires protéiques pour être fonctionnelle, comme montré chez l'Homme (Holzmann et al., 2008). Par contre, aucune protéine PRORP nucléaire n'est présente chez les animaux où l'activité RNase P nucléaire est portée par des RNP. Chez de nombreux autres eucaryotes, les protéines PRORP sont présentes en plusieurs copies. En particulier, chez les plantes et les trypanosomes, elles sont présentes à la fois dans le noyau et les organites. On note toutefois que les protéines PRORP sont complètement absentes des champignons (Lechner et al., 2015). Chez les plantes

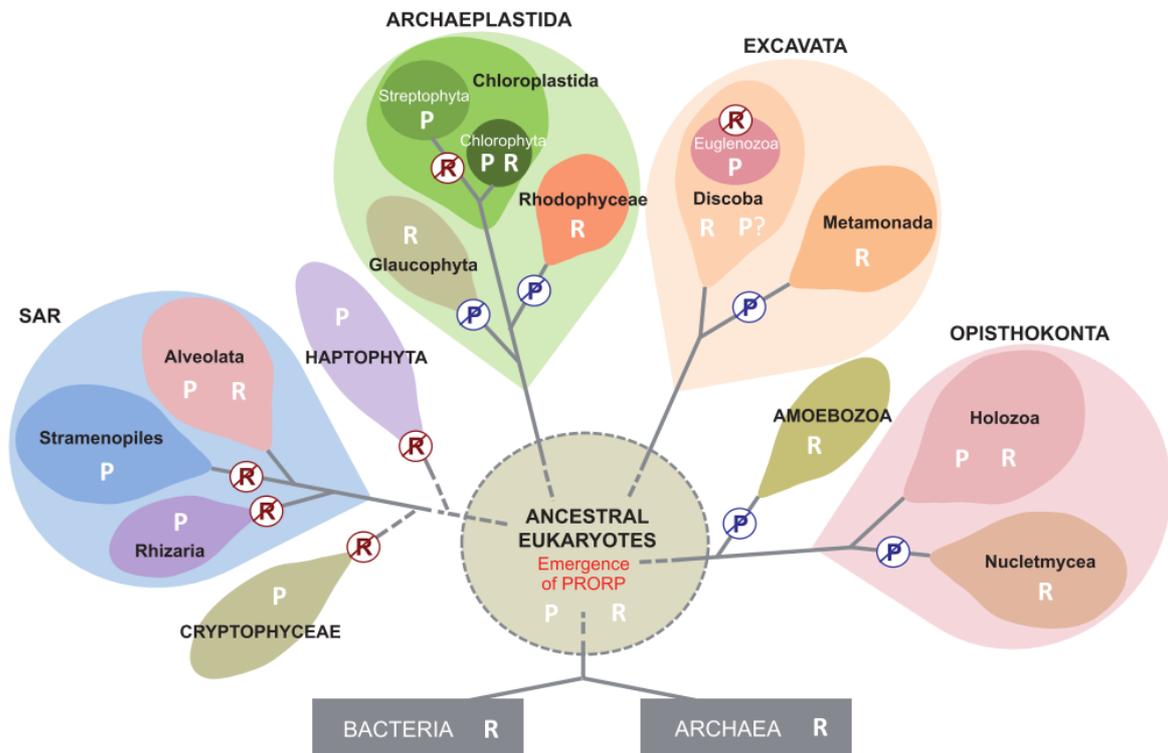


Figure 15 : Répartition des RNase P de type RNP et PRORP chez les eucaryotes.

Représentation schématique de la répartition des RNase P de type RNP (R) et PRORP (P) dans les groupes majeurs d'eucaryotes. Les P ou R barrés indiquent des événements évolutifs putatifs associés à la perte de PRORP ou de la RNase P RNP. Le point d'interrogation indique un exemple où les données génomiques limitées ont empêché de conclure quant à la présence du type d'enzyme donné dans un groupe particulier. Le diagramme illustre la vaste distribution des protéines PRORP chez les eucaryotes et montre comment la distribution des RNase P a apparemment impliqué plusieurs événements de pertes de PRORP ou de la RNase P RNP (Figure issue de Lechner et al., 2015).

supérieures, une étude phylogénétique basée sur l'alignement de plusieurs centaines de protéines PRORP a montré que ces protéines peuvent être classées en trois groupes. Les groupes I et II comportent les protéines PRORP adressées aux organites alors que les protéines du groupe III sont nucléaires (Lechner et al., 2015). Il est intéressant de noter qu'Arabidopsis ainsi que l'ensemble des brassicacées ne possèdent pas de protéine PRORP du groupe II. En effet, AtPRORP1 appartient au groupe I, alors que AtPRORP2 et AtPRORP3 sont issus d'une duplication récente et appartiennent toutes deux au groupe III. De manière très intéressante, la comparaison de la répartition des protéines PRORP et des RNase P RNP a montré que les deux types d'enzymes étaient mutuellement exclusifs dans des organismes entiers ou dans les compartiments d'un même organisme (Lechner et al., 2015). Cette analyse a aussi permis de définir une histoire évolutive de la répartition des RNases P dans les différents groupes d'eucaryotes, en montrant e.g. que les protéines PRORP étaient initialement des enzymes mitochondriaux chez les animaux alors qu'elles étaient initialement uniquement nucléaires chez les plantes (Lechner et al., 2015).

5. Comparaison de la RNase P RNP et de PRORP

a. Répartition dans le vivant

Chez les procaryotes (bactéries et archées), seule la RNase P RNP est présente. L'analyse phylogénétique des gènes de PRORP suggère une origine monophylétique. PRORP serait apparu dans un eucaryote ancestral par la fusion d'un gène PPR avec un gène de nucléase NYN. PRORP se serait ensuite propagé dans les différents groupes d'eucaryotes quelques fois en remplaçant totalement les RNase P RNP comme chez les plantes et les trypanosomes ou en se spécialisant pour l'activité RNase P d'un compartiment particulier, comme c'est le cas chez les animaux où une RNase P RNP est présente dans le noyau et une protéine PRORP dans les mitochondries. D'autres groupes comme les Amoebozoa ou les Metamonada ont uniquement conservé les RNase P RNP (Lechner et al., 2015) (Figure 15).

On notera toutefois que l'activité RNase P n'est pas entièrement universelle. Une archée dépourvue d'activité RNase P a été décrite. En effet, *Nanoarchaeum equitans* possède des gènes d'ARNt exprimés sous le contrôle d'un promoteur conservé 26 nt en amont de la première base de l'ARNt mature ce qui permet de transcrire les ARNt à partir de leur position +1. Ainsi aucune séquence 5' supplémentaire n'est transcrite et l'activité RNase P non nécessaire n'a pas été conservée (Randau et al., 2008).

b. Mode de reconnaissance des ARNt

Des expériences de complémentation fonctionnelle ont permis de montrer que PRORP pouvait se substituer à une RNase P RNP *in vivo* (Gobert et al., 2010). Ceci suggère que les deux types de RNase P peuvent partager un même spectre de substrat et avoir des modes d'action similaires. La RNase P RNP bactérienne se lie à l'ARNt au niveau du bras accepteur afin de cliver la séquence 5' leader. Pour cela elle interagit en particulier au niveau du coude de l'ARNt formé par l'empilement des domaines T et D ainsi qu'au niveau de la tête catalytique entre les nucléotides +1 et -1 de l'ARNt précurseur. (Lagrandeur et al., 1993; Pan et al., 1995; Reiter et al., 2010). La séquence 5' leader peut alors interagir avec la sous-unité protéique de la RNP (Kurz et al., 1998). Chez *E. coli*, le CCA à l'extrémité 3' terminale de l'ARNt se lie à une séquence complémentaire dans une boucle de l'ARN P de la RNase P RNP.

L'analyse du mode d'action de PRORP (Gobert et al., 2013) et son organisation en deux parties laisse penser que le domaine catalytique NYN joue un rôle identique au domaine catalytique (C) de l'ARN P de la RNP et que les motifs PPR jouent un rôle de reconnaissance identique à celui du domaine de spécificité (S) de l'ARN P (Pinker et al., 2013). En particulier, comme les RNase P RNP, PRORP reconnaît le coude de l'ARNt où il se lie avec des résidus conservés tels que les résidus G18, G19, C56 et R57 dans les boucles du D et du T (Gobert et al., 2013).

c. Mécanismes de clivage

Bien qu'ayant des mécanismes de reconnaissance de l'ARNt similaires, les deux types d'enzymes semblent avoir des mécanismes catalytiques différents (Howard et al., 2012; Pavlova et al., 2012; Pinker et al., 2013). Les deux enzymes clivent la liaison phosphodiester par une attaque nucléophile en amont de l'ion hydroxyde apical en O3' du ribose, ce qui génère des produits avec des extrémités 5' phosphoryle et 3' hydroxyle. La présence de deux ions métalliques dans la structure de AtPRORP1 d'*Arabidopsis* et la nécessité d'avoir des ions métalliques (Mn^{2+} ou Mg^{2+}) pour l'activité de clivage *in vitro* par PRORP suggèrent que PRORP utilise un mécanisme à deux ions métalliques pour dé-protoner l'eau et stabiliser l'état de transition (Steitz and Steitz, 1993). Néanmoins, PRORP peut cliver des substrats avec une modification RP-phosphorothioate au niveau du site de clivage canonique en utilisant le Mg^{2+} comme cofacteur. Ceci représente une différence importante avec le ribozyme de la RNase P (Thomas et al., 2000; Pavlova et al., 2012), car cela suggère que l'oxygène pro-RP de la

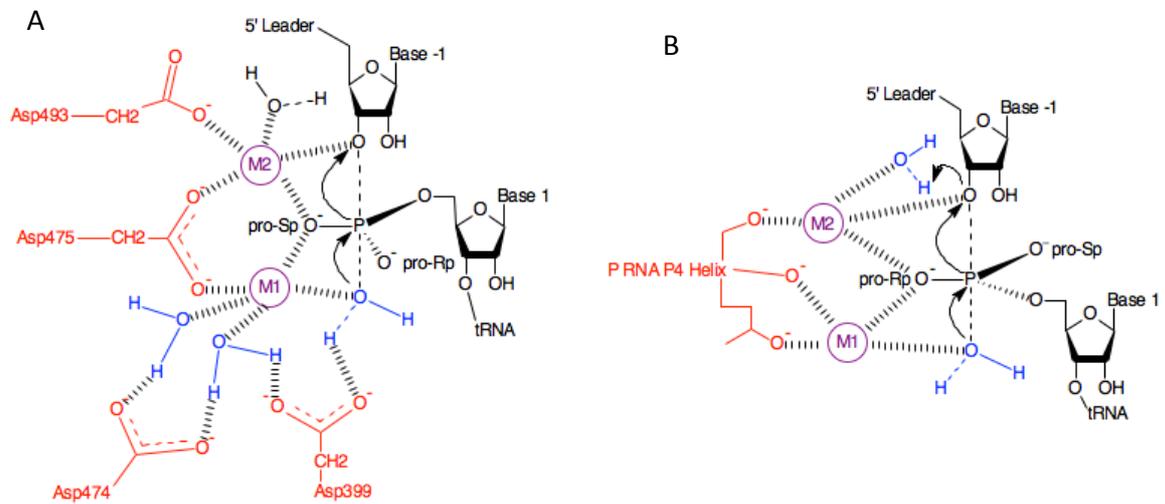


Figure 16 : Mécanismes de clivage de l'ARN utilisé par les RNase P de type PRORP et RNP.

Mécanisme de clivage par PRORP (A) et par la RNase P RNP (B). M1 et M2 représentent deux ions métalliques, coordonnés soit à des aspartates conservées du site catalytique de PRORP, soit à l'hélice P4 du domaine catalytique du ribozyme de la RNase P. Les flèches noires représentent le transfert des électrons. Chez PRORP, M1 et M2 interagissent avec un oxygène « pro-Sp » alors que dans la RNP ils interagissent avec un oxygène « pro-Rp ».

liaison phosphodiester cible n'est pas directement coordonné par l'ion métallique dans la structure de PRORP. Il apparaît plutôt que PRORP utilise un mécanisme plus classique, similaire à celui des autres metallonucleases caractérisées, dans lequel la fonction principale des ions métalliques est de stabiliser la structure et la charge de l'état de transition. Par la suite, une réaction chimique acide-base classique est effectuée par les groupes carboxylates des aspartates conservés dans le site catalytique de PRORP (Howard et al., 2012; Reiter et al., 2010). En revanche, le ribozyme de la RNase P utilise un cation divalent hydraté pour fournir un hydroxyde pour l'attaque nucléophile et un second ion métallique hydraté pour protoner le groupe à l'extrémité 3' de la séquence 5' leader (Donghi and Schnabl, 2011). Dans une autre étude, Howard et al. proposent un modèle dans lequel la liaison d'un ion Mg^{2+} dans le site catalytique augmente l'affinité du deuxième ion métallique (Howard et al., 2015). La comparaison des mécanismes proposés pour les activités RNase P de PRORP et du ribozyme montre que les deux enzymes utilisent des ions métalliques catalytiques, bien que les ligands métalliques soient différents. Les résidus aspartates conservés de PRORP coordonnent les ions métalliques, alors que le ribozyme de la RNase P RNP utilise principalement des atomes d'oxygène pour coordonner les ions métalliques (Schelcher et al., 2016) (Figure 16).

d. Comparaison structurale et mécanistique des RNase P protéiques et ribonucléoprotéiques.

Les RNase P protéiques et RNP réalisent toutes deux la maturation en 5' des pré-ARNt en clivant entre les positions +1 et -1, bien que leur efficacité *in vitro* semble différente (Schelcher et al., 2016). Leur activité est identique bien que les mécanismes utilisés et leurs structures soient différents.

Les protéines PRORP se caractérisent par la présence d'un domaine NYN responsable de l'activité et de domaines PPR responsables de l'interaction avec le pré-ARNt. Le domaine NYN des PRORP est différent des autres domaines NYN, car il possède deux motifs (D/E/T/H/N/P/G)h3D(G/A)xN et DDx15-39(S/T)xDx3DH séparés par 50 à 80 acides aminés. Hormis les positions 5 et 35, les séquences des domaines PPR sont très peu conservées, seule la structure composée d'hélice alpha et de boucle l'est. On peut également noter la présence d'une boucle plus grande entre les deux hélices du motif PPR2 de AtPRORP1. Les RNases P RNP se composent d'un ARN divisé en deux domaines, le domaine C catalytique et le domaine S de spécificité. L'ARN est accompagné d'une ou plusieurs protéines en fonction de l'organisme.

Les deux types de RNase P contactent le pré-ARNt au niveau de l'empilement de la boucle T et D du pré-ARNt mais les mécanismes de clivages sont différents. La RNase P RNP utilise un oxygène pro-Sp alors que PRORP utilise un oxygène pro-Rp.

Les deux types d'enzymes reconnaissent les pré-ARNt mais l'efficacité de clivage est accrue avec des séquences 5' leader courtes, i.e. d'une taille de 2 à 10 nt. Une séquence plus longue peut engendrer la formation de structures secondaires alternatives empêchant la reconnaissance du pré-ARNt par la RNase P. De plus la nature du nucléotide en -1 ne semble pas avoir d'effet majeur pour l'activité de PRORP, alors qu'elle est importante pour les RNase P RNP.

Toutes ces similarités et différences sont compilées dans la publication 2 : Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity. Cette publication a été écrite dans le cadre d'une édition spéciale du journal *Biomolecules* intitulée : *Function and Structure of RNase P in Fungi, Bacteria and Human Cells*. Ma participation à la préparation de la revue a consisté à la rédaction du plan et de la première ébauche du texte. J'ai par ailleurs réalisé la recherche bibliographique exhaustive des différents points abordés dans la revue. J'ai effectué tout le travail de compilation des différentes mesures de cinétiques enzymatiques décrites dans les tables de la revue ainsi que l'analyse de la différence mécanistique entre les deux types de RNase P.

Review

Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity

Cédric Schelcher ¹, Claude Sauter ² and Philippe Giegé ^{1,*}

¹ UPR 2357, Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire des Plantes, Université de Strasbourg, 12 rue du général Zimmer, F-67084 Strasbourg, France; cedric.schelcher@etu.unistra.fr

² UPR 9002, Centre National de la Recherche Scientifique, Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, 15 rue René Descartes, Strasbourg F-67084, France; c.sauter@unistra.fr

* Correspondence: giegé@unistra.fr; Tel.: +33-3-6715-5363

Academic Editor: Denis Drainas

Received: 19 May 2016; Accepted: 17 June 2016; Published: 24 June 2016

Abstract: RNase P, the essential activity that performs the 5' maturation of tRNA precursors, can be achieved either by ribonucleoproteins containing a ribozyme present in the three domains of life or by protein-only enzymes called protein-only RNase P (PRORP) that occur in eukaryote nuclei and organelles. A fast growing list of studies has investigated three-dimensional structures and mode of action of PRORP proteins. Results suggest that similar to ribozymes, PRORP proteins have two main domains. A clear functional analogy can be drawn between the specificity domain of the RNase P ribozyme and PRORP pentatricopeptide repeat domain, and between the ribozyme catalytic domain and PRORP N4BP1, YacP-like Nuclease domain. Moreover, both types of enzymes appear to dock with the acceptor arm of tRNA precursors and make specific contacts with the corner of pre-tRNAs. While some clear differences can still be delineated between PRORP and ribonucleoprotein (RNP) RNase P, the two types of enzymes seem to use, fundamentally, the same catalytic mechanism involving two metal ions. The occurrence of PRORP and RNP RNase P represents a remarkable example of convergent evolution. It might be the unique witness of an ongoing replacement of catalytic RNAs by proteins for enzymatic activities.

Keywords: tRNA biogenesis; RNase P; PRORP; crystal structures; kinetic analyses

1. Introduction

Similar to Janus, the two-faced Roman god of transitions and passages that concomitantly looks to the future and to the past, RNase P has two faces. This essential enzymatic activity that removes the 5' leader sequences of tRNA precursors occurs either as ribonucleoproteins (RNP) involving a ribozyme [1] or as protein-only enzymes [2–5]. The ancestral RNP RNase P, first described in *Escherichia coli* [6–8] is ubiquitous in bacteria and archaea, although it is absent from an archaea species where tRNA transcription starts at position +1 and RNase P is not needed [9]. RNP RNase P also occurs in the nucleus and/or organelles of a variety of eukaryotes, notably in animal nuclei and in both mitochondria and the nucleus in fungi [10]. Nevertheless, a completely distinct form of RNase P devoid of RNA also appeared during eukaryote evolution and is present in the nucleus and/or organelles of several eukaryotes [11].

RNP RNases P are present in all three domains of life, their catalytic RNA (P RNA) structure is relatively well conserved in prokaryotes and eukaryote nuclei [10]. However, it is highly degenerate in several organelles [12]. P RNAs contain two main domains, i.e., a catalytic (C) and a specificity (S)

domain. The protein content of RNP RNase P is comparatively more divergent. While a single protein subunit is found in bacteria, 4 to 5 subunits occur in archaea and up to 10 in eukaryote nuclei [10,13–16].

In contrast, protein-only RNase P enzymes are entirely specific to eukaryotes. These enzymes, termed PRORP (for PROtein-only RNase P or PROteinaceous RNase P), have two main domains; a C-terminal catalytic domain belonging to the N4BP1, YacP-like Nuclease (NYN) family [17] and an N-terminal pentatricopeptide repeat (PPR) domain involved in RNA binding [18]. The two main domains are linked by a connecting domain involving zinc [19,20]. PRORP proteins were initially described in human mitochondria and in the organelles and nucleus of the model plant *Arabidopsis thaliana* [20–23]. They were also characterized in *Trypanosoma brucei* nucleus and mitochondria [24], and in other species of the green lineage, i.e., *Ostreococcus tauri* [25], the moss *Physcomitrella patens* [26], and the model green algae *Chlamydomonas reinhardtii* [27]. Interestingly, while separate RNase P enzymes were always found in the nucleus and organelles of eukaryotes (either multiple RNPs, multiple PRORPs or a combination of both), *Chlamydomonas* utilizes a single PRORP protein for RNase P activity in the nucleus, mitochondria and the chloroplast, thus making the most compact and versatile RNase P machinery described to date in both prokaryotes and eukaryotes [27]. Likewise, PRORP sequences are also found in a diversity of other eukaryotes, in particular in Stramenopiles, Alveolata, and Rhizaria. PRORP thus occurs in four out of five main eukaryote super-groups, but never co-exists with RNPs in the same compartment. Indeed, the occurrence of PRORP and RNPs seems mutually exclusive in cell compartments where gene expression takes place, or even in entire organisms [11]. For instance, although RNase P RNA and PRORP have both been proposed to be present in human mitochondria [28], PRORP appears to be sole responsible for RNase P in this compartment [21,29]. Remarkably, PRORP proteins are capable of functionally replacing RNPs for RNase P activity, as shown by complementation experiments in *Escherichia coli* [22] and *Saccharomyces cerevisiae* [30], thus showing that the two types of enzymes are fully interchangeable, at least for the maturation of essential substrates, such as pre-tRNAs. In some cases, eukaryote RNPs might have been retained because they evolved additional functions that could not be performed by PRORP, e.g., as observed in human nuclei with the requirement of RNP RNase P for the formation of RNA polymerase III initiation complexes [31]. Nonetheless, RNase P activity makes a unique example of an activity that can be performed either by RNA or by proteins. It might be the witness of a still ongoing transition from the ancestral RNA world, where molecular processes were catalyzed by RNA molecules, to the contemporary world, where enzymatic activities are primarily held by proteins.

Here, we review the biochemical, structural, and mechanistic data obtained for protein-only RNase P and discuss how PRORP structure, pre-tRNA recognition, and cleavage differ, but also resemble that of RNP RNase P.

2. Diversity of PRORP Sequence Features

Functionally characterized, as well as putative, PRORP sequences are all typified by an α -superhelical domain containing tandem arrays of PPR motifs and a specific NYN metallonuclease domain, bridged by a bipartite zinc-binding domain. Sequences of the N-terminal PPR domains are highly degenerate, but recognizable, by the 3D fold specific for PPR motifs [11,18]. Recent advances on PPR proteins' mode of action suggest that each PPR motif specifically binds a single nucleotide in RNA substrates, according to a "PPR code" where the nature of residues, in particular at positions five and 35 of the respective motifs, is responsible for the specificity of the motif for a given nucleotide [32–35]. A longer stretch of PPR motifs would, thus, bind a higher number of nucleotides. Among PRORP sequences, the numbers of PPR motifs predicted with TPRpred [11] range from two to four. In addition, structure predictions systematically reveal variable numbers of additional PPR-like motifs [11]. This variability in the size and nature of PPR domains suggests that the respective PRORP enzymes might have different RNA binding properties, e.g., variable affinities for RNA substrates.

The C-terminal regions of PRORP enzymes contain a metallonuclease domain that was shown to be responsible for the endonuclease activity [22]. This domain belongs to the PIN-like super-family of

nucleases and, more precisely, to the NYN family. The actual NYN domain of PRORP appears to be of prokaryote origin as it resembles putative bacterial YacP ribonuclease sequences. *YacP* genes are ubiquitous in prokaryotes and eukaryotes, with the notable exception of fungi [17]. While NYN domain proteins are widespread in eukaryotes, the NYN domains of PRORP proteins have specificities that allow discrimination between bona fide PRORP proteins and other NYN domain proteins. They are characterized by a specific signature involving conserved aspartates and histidines, i.e., composed by a first motif (D/E/T/H/N/P/G)_{h3}D(G/A)_xN and a second motif DD_{x15-39}(S/T)_xDX₃DH separated by 50 to 80 amino acids. This NYN signature of PRORP differs significantly from the basic NYN signature [17], and also differs to some degree from the Pfam RNase_Zc3h12a signature (PF11977). Although conserved in PRORP, the NYN domains have lineage-specific disparities, e.g., the spacer sequence between the two motifs is typically longer in plants as compared to animals [11].

The connecting zinc-binding domain is split into two subparts upstream and downstream of the NYN domain. Its first part involves a highly conserved CxxC motif, where both cysteines were shown to be required for zinc coordination in *Arabidopsis* and human [19,20,36]. However, the CxxC motif is not ubiquitous in PRORP. For instance, in the Chlorophyta *Chlamydomonas* CxxC is replaced by CxxA [27]. The second part of the zinc-binding domain is defined by (W/Y/F)HxPx and (W/F)_xC_{x2-3}(R/K) signatures [11]. Here, as well, lineage-specific variations are observed. Altogether, this variability in the connecting domain suggests that the respective PRORP enzymes might bind zinc and/or other metals with different affinities.

Beyond the PPR, the NYN and the connecting domain, other signatures are present in PRORP sequences of specific phyla. For example, a glycine-rich insertion is characteristic of land plant PRORPs. In some phyla, a PPP(Y/F/C)(S/T) motif is found between the NYN domain and the second zinc-binding subdomain. The occurrence of such specific insertions might indicate additional phylum-specific functions, e.g., through interactions with yet unidentified proteins.

3. Comparison of PRORP Three-Dimensional Structures

First analyses of PRORP sequences along with biochemical and biophysical studies, including synchrotron radiation circular dichroism and small angle X-ray scattering, showed that PRORP enzymes are organized in two main domains and contain a high proportion of α -helices [20,22]. This was confirmed by the determination of atomic resolution structures of *Arabidopsis* organellar PRORP1 and nuclear PRORP2 as well as of human PRORP (previously referred to as mitochondrial ribonuclease P protein 3, MRPP3) by X-ray crystallography [19,36–39]. These structures reveal a largely similar architecture and, thus, highlight the overall conservation of the PRORP fold in distantly related eukaryotic phyla.

3.1. Common Structural Features

As expected, the N-terminal PPR domain forms a superhelical structure closely resembling that of TPR (tetratricopeptide repeat) motifs, a domain evolutionary related to PPR and involved in protein-protein interactions [40,41]. Despite the high primary sequence divergence between *Arabidopsis* and human PRORP sequences, their PPR motifs are, structurally, highly similar (Figure 1). They are also superimposable with other motifs from PPR proteins of established 3D structure [42–45], thus emphasizing the strong structural conservation of PPR motifs among eukaryotes.

Then, the catalytic domain of PRORP implements an $\alpha/\beta/\alpha$ sandwich fold typical of all structurally-characterized PIN-like and flap nucleases [17,46,47]. A comparable fold is also present in the ribonuclease MCP-1 induced protein 1 (MCPIP1) that takes part in the regulation of the immune response through the degradation of inflammatory cytokines mRNAs [47]. Similarly, the nuclease domains of T4 RNase H [48] and of SMG6 and SMG5, two essential proteins involved in nonsense-mediated mRNA decay in humans, also adopt a similar fold [46]. The crystal structure of PRORP1 revealed that four aspartate residues are involved in the binding of metal ions [19] contrary to flap nucleases that utilize six conserved aspartates [17]. Among the PRORP NYN domain, four

positions (D399, D474, D475, and D493 in Arabidopsis PRORP1) are strictly conserved in all sequences and were shown to be essential for pre-tRNA maturation [20]. In other proteins containing a NYN domain and not involved in RNase P activity, the positions equivalent to D399, D475, and D493 are also conserved, but position 474 is often an alanine [17].

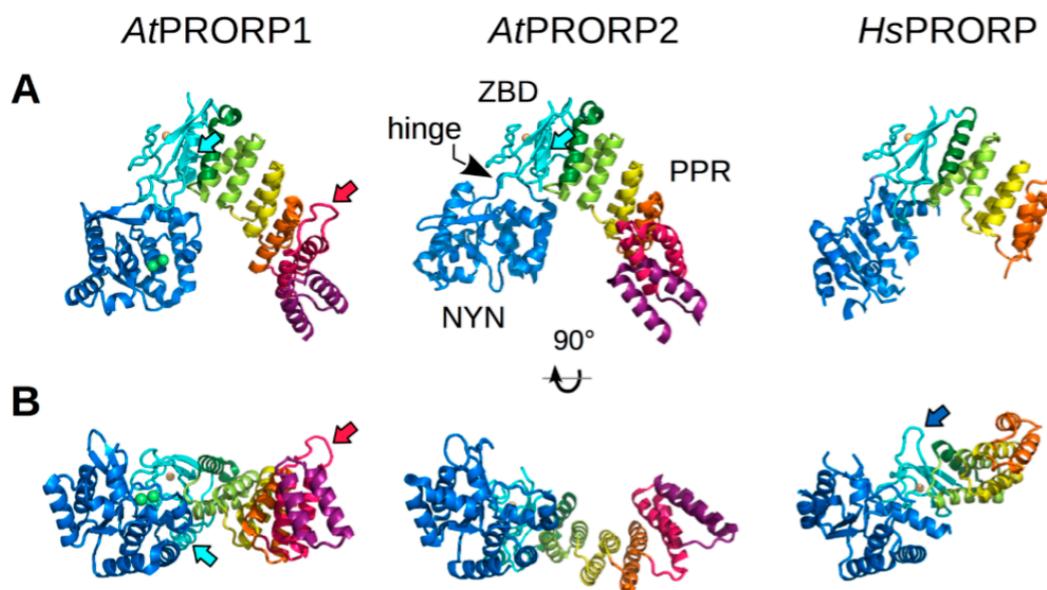


Figure 1. Protein-only RNase P (PRORP) architecture. As of May 2016 nine crystal structures of PRORP enzymes are available in the Protein Data Bank, four for Arabidopsis PRORP1 in the presence of different metal ions (PDB IDs: 4G24, 4G25, 4G26), two for PRORP2 (5DIZ, 5FT9) and three for truncated forms of human PRORP lacking PPR1-4 or PPR1-2 modules (4ROU, 4XGL, 4XGM). Selected structures correspond to *At*PRORP1 with two bound manganese ions shown in green (4G24, 1.95 Å resolution, [19]), and highest resolution models of *At*PRORP2 (5FT9, 3.05 Å resolution, Pinker et al. in preparation) and of *Hs*PRORP (4XGL, 1.8 Å resolution, [36]). All PRORP enzymes contain an N-terminal pentatricopeptide repeat (PPR) domain made of five PPR modules (PPR 1-5 depicted in violet, red, orange, yellow, and green, respectively) followed by a single helix (half PPR or PPR5B, shown in dark green), and a C-terminal N4BP1, YacP-like Nuclease (NYN) catalytic domain (NYN, depicted in dark blue). These two functional domains are linked by a bipartite zinc binding domain (ZBD, depicted in cyan) coordinating a Zn^{2+} ion (shown in gold). PRORP structures are either superimposed according to their PPR3-5B modules and ZBD (A) or to their catalytic domain (B). These two orthogonal views highlight the possible reorientation of PPR and NYN domains with respect to each other, owing to a flexible hinge present between the NYN and ZBD. Red, cyan, and blue arrows indicate the long loop of PPR2 motif present in *At*PRORP1, the 30-residue-long linker inserted in the ZBD next to the PPR domain in Arabidopsis enzymes, and the extended lariat loop in the central region of human PRORP, respectively. Molecular representations of PRORPs were prepared with PyMOL (Schrödinger, Portland, OR, USA).

The two main domains are bridged by the zinc-binding domain, which appears to have a structural role, i.e., to connect the two functional domains and, thus, to give PRORP its overall structural design (Figure 1). The zinc-binding domain forms the apex of PRORP characteristic Λ shape. In this configuration, the PPR and NYN domains form the two arms of the Λ . On one side, the concave part of the PPR superhelix, that provides a platform for interaction with RNA if PRORP conforms to the overall mode of action of PPR proteins [43,49], faces the catalytic groove on the other side of the Λ , thus exposing conserved aspartate residues and metal ions toward the predicted RNA binding side of the protein.

3.2. Specific Structural Features

The comparison of Arabidopsis organellar PRORP1 with nuclear PRORP2 structures revealed that PPR domains in both enzymes are very similar and composed of five and a half PPR and PPR-like motifs corresponding to 11 consecutive α -helices. However, differences are found: for instance, motif PPR2 of PRORP1 has an extension between its two α -helices (Figure 1). The function of this loop is unknown, but might be involved in protein-protein interactions that specifically take place in Arabidopsis organelles. While PRORP1 and 2 share the same overall Λ -shaped structure, the angle of the Λ differs significantly (Figure 1A), the angle in PRORP2 being more open than that in PRORP1 [19,39]. This difference could reflect specificities of organellar and nuclear PRORP modes of action. This is, however, unlikely because both PRORP1 and 2 appear to be able to cleave any pre-tRNA of canonical fold, as reviewed below, and pre-tRNA substrates have a canonical fold in both the Arabidopsis nucleus and organelles (contrary to animals, where mitochondrial tRNA structures often differ significantly from nuclear encoded tRNA structures [50]). A more likely hypothesis is that the two crystal structures have captured two alternate conformations that can be adopted by PRORP proteins. PRORP1 structure shows structural domains organized in a plane (Figure 1B), whereas PRORP2 and human PRORP structures display a left or right rotation of PPR with respect to NYN domains. This twist is allowed by the presence of a hinge between the catalytic and zinc-binding domains [36,39]. Human PRORP, as PRORP2, adopts a more relaxed Λ shape. Such conformational flexibility/adaptability, which can be detected in solution by small-angle X-ray scattering (SAXS) (Pinker et al., in preparation) might be required for PRORP function, e.g., to select and bind precursor substrates, and release products.

Similarly, the comparison of human PRORP structure with plant PRORP structures revealed some discrepancies. Plant PRORP proteins feature a 30-residue-long loop-helix-loop linker inserted between the PPR and the zinc-binding domains which is replaced by a short loop in the human enzyme. In turn the latter has a longer lariat loop in its zinc-binding connecting domain as compared to Arabidopsis PRORPs. The human structure is stabilized by an extended hydrogen bond network and makes a finger between the PPR and NYN domain (Figure 1B), whereas it adopts a more compact horseshoe-like bend in plants. The longer lariat in humans does not seem to be involved in binding TRMT10C and SDR5C1, the two additional subunits of human mitochondrial RNase P [21,51], as its exchange with a shorter linker did not result in enzyme activity loss in the presence of the two partners. The loop substitution did not make PRORP work independently from its partners either [36]. This discrepancy might, alternatively, be related to intrinsic differences in PRORP potential for flexibility or structural changes in plants and animals. Still, the reason why human PRORP requires two protein cofactors is not clear yet. One hypothesis is that the NYN domain contains structural features, like the lariat loop, that inhibit its activity in the absence of partners. PRORP chimeras, in which the whole human NYN domain was substituted by that of PRORP1, were used in an attempt to create a cofactor-independent PRORP [36,37], but only the latter group succeeded in partially restoring an activity. It was also proposed that the alternate conformation triggered by a salt bridge with a neighboring arginine in the loop containing the two conserved aspartates could hamper binding of one metal ion required for catalysis [37]. However, the absence of metal ions in the catalytic site is not specific to human PRORP. Two (independent) crystallographic studies of PRORP2 led to the same observation of a metal-free NYN domain. Overall, these data are rather in favor of a role of TRMT10C and SDR5C1 as structural chaperones to select and stabilize an active conformation of human PRORP. The fact that only mutants with truncated PPR domains could be crystallized, not a full-length construct, suggests that PRORP is, indeed, a dynamic protein.

4. Mechanistic Analyses of Protein-Only RNase P Activity

The capacity of PRORP enzymes to functionally replace RNP enzymes *in vivo* suggested that both types of RNase P might share a similar mode of action. For its activity, bacterial RNP RNase P docks onto the acceptor arm of the pre-tRNA substrate. In particular, the P RNA specificity (S) domain

interacts with the tRNA corner formed by the stacked T and D loops and the catalytic (C) domain contacts the cleavage site between nucleotides -1 and $+1$ of pre-tRNAs [52–54]. Then, the pre-tRNA leader sequences interact with the protein subunit of the RNP [55]. Finally, in *E. coli*, the 3' terminal CCA of tRNAs binds a complementary sequence in a P RNA loop of RNP RNase P. The initial analysis of PRORP bipartite organization suggested that the PPR domain of PRORP might play a role similar to that of the P RNA S domain, while the NYN domain would have a function similar to that of the C domain in RNP RNase P [5]. A growing body of evidence now suggests that PRORP indeed conforms to that model, although its mode of action also differs in some aspects to that of RNP RNase P.

4.1. Kinetic Analyses of PRORP Activity

A number of studies have investigated the activity and kinetic constants of PRORP cleavage of pre-tRNAs. Activity assays were performed with the three Arabidopsis PRORP enzymes, with plant organellar and nuclear pre-tRNAs, as well as with bacterial pre-tRNA substrates. Reactions were performed as single or multiple turnover kinetics experiments. Results summarized in Table 1 reveal major kinetic parameter disparities between individual studies, depending on the protein used and/or the pre-tRNA investigated. For instance, k_{obs} vary from 1 to 7 min^{-1} , K_D values determined with a same protein but different substrates range from 60 to 2300 nM, while K_M values determined in multiple turnover conditions vary from 140 to 2000 nM (Table 1) and can be as low as 3 nM with a bacterial substrate cleaved by PRORP3 in single turnover conditions [56] (Table 2). The detailed comparison of results suggests that the three Arabidopsis PRORP enzymes perform RNase P activity with comparable catalytic efficiencies. However, some pre-tRNA substrates appear to be cleaved more efficiently by a given PRORP paralogue. This might reflect true, although subtle, differences in the mode of action of organellar versus nuclear PRORP enzymes. The two types of PRORP enzymes might have slightly different substrate recognition processes [57] which might reflect the co-evolution of a given enzyme with its substrate population in vivo. In other cases, the observed kinetic parameters discrepancies most probably also reflect differences in experimental setups and the intrinsic nature of pre-tRNAs used in the respective studies. For instance, tRNAs with a highly stable acceptor stem, such as *Thermus thermophilus* pre-tRNA^{Gly}, are most probably inherently better RNase P substrates than other pre-tRNAs with less stable structures, such as Arabidopsis mitochondrial pre-tRNA^{Cys}.

4.2. Involvement of PRORP cis-Elements for RNase P Activity

Beyond the analysis of wild-type PRORP proteins activity, some studies have investigated the variations of PRORP activity and kinetic parameters in PRORP mutants. Results summarized in Table 1 confirm the essential role of the four conserved aspartates (D399, D474, D475, and D493 in Arabidopsis PRORP1) of the NYN domain for RNase P cleavage. Results also show that different deletions of N-terminal parts of PRORP, especially deletions of PPR motifs strongly affect RNase P activity. In the PPR domain of plant PRORPs, positions 5 and 35 of motif PPR3 are the most conserved and point mutations at these positions strongly affect RNase P activity [56,58]. However, in the latter studies, mutations in the PPR domain were analyzed by RNase P cleavage assays and direct RNA binding was not analyzed. Results thus represent indirect evidence to show the importance of the PPR domain for RNA recognition and do not conclusively show if and/or how PPR motifs confer specificity to PRORP enzymes. In another study, lysine residues positioned at both tips of PRORP Λ shape were found in contact with the pre-tRNA substrate [59], thus, in agreement with the general substrate binding process proposed for PRORP proteins [5]. Mutated residues that were shown to interact with pre-tRNA and/or to be essential for RNase P activity are shown in Figure 2.

Table 1. Kinetic analyses of the importance of PRORP cis-elements for RNase P activity. At stands for *Arabidopsis thaliana*, Tt for *Thermus thermophilus*, “mito” for mitochondrial, “chloro” for chloroplastic, “nuc” for nuclear and “NA” for not applicable in the case of a leaderless RNA substrate. Grey boxes indicate results of single turnover kinetic experiments (with k_{obs} , the observed first order constant obtained by the fit of a single exponential equation), whereas green boxes indicate results of RNase P cleavage assays performed in multiple turnover conditions (with k_{cat} and K_M , the Michaelis-Menten parameters). K_D were derived from fluorescence polarization binding assays.

RNase P	Domain Mutated	Position Mutated	pre-tRNA	Cleavage (+/- or %)	k_{obs} (min^{-1})	k_{cat} (min^{-1})	K_M (nM)	K_D (nM)	References
A ^{PRORP1}	-	D474A/D475A		+	1.4 ± 0.1			700 ± 100	[22]
	Nt deletion	Δ76		-	<0.001			24,000 ± 10,000	
	Nt deletion	Δ245			<0.001			1100 ± 100	[19]
	NYN	D399A			<0.001			1400 ± 600	
	NYN	D474A			<0.001			1400 ± 100	
	NYN	D475A			<0.001			300 ± 100	
	NYN	D493A			2.22 ± 0.12	3.72 ± 0.3	670 ± 230	510 ± 120	[57]
	-	-	At mito pre-tRNA ^{Cys}		1.2 ± 0.3			250 ± 34	
	-	-			1.2 ± 0.4			859 ± 159	
	PPR1	K101A			1.3 ± 0.3			389 ± 35	[59]
	PPR1	K109A			0.4 ± 0.1			609 ± 70	
	NYN	K439A			1.3 ± 0.1			377 ± 23	
NYN	K460A			2.5 ± 0.1			1624 ± 344		
PPR1	K101A/K109A			0.5 ± 0.1			888 ± 127	[60]	
PPR1/NYN	K101A/K439A			0.5 ± 0.1			763 ± 92	[22]	
PPR1/NYN	K109A/K439A		Tt pre-tRNA ^{Gly}	+	2.3 ± 0.6				
A ^{PRORP1}	-	D474A/D475A		+					
	-	-		-					
	NYN	D474A/D475A		100					
	PPR2	N136T		60					
	PPR3	T180N		26					
	PPR4	G215N		85					
	-	-	At chloro pre-tRNA ^{Phe}		100				
	Nt deletion	Δ89		35					[58]
	Domain deletion	ΔPPR1		3					
	Domain deletion	ΔPPR1-2		0					
	Domain deletion	ΔPPR1-3		0					
	Domain deletion	Δ89 PPR2		0					
Domain deletion	Δ89 PPR3		0						
-	-	At nuc pre-tRNA ^{Cys}		2.1 ± 0.12	2.52 ± 0.3	140 ± 50	60 ± 10	[57]	
-	-	At nuc pre-tRNA ^{Phe}		2.22 ± 0.18	2.4 ± 0.12	550 ± 50	2300 ± 300		
-	-	At nuc pre-tRNA ^{Phe}		4.68 ± 0.18	2.1 ± 0.18	160 ± 50	330 ± 60		

Table 1. Cont.

RNase P	Domain Mutated	Position Mutated	pre-tRNA	Cleavage (+/− or %)	k_{obs} (min^{-1})	k_{cat} (min^{-1})	K_M (mM)	K_D (mM)	References
APPRORP2	-		Tt pre-tRNA ^{Gly}		5.0 ± 1.2				[60]
	-	D393A			1.1 ± 0.1				
	-	D421A			<0.001				
	-	D422A	At nuc pre-tRNA ^{Gly} 8:1		<0.001				
	-	D440A			<0.001				
APPRORP3	-	H445A			0.02 ± 0.004				[39]
	-	Δ141			<0.001				
	-		At mito pre-tRNA ^{Cys}		0.78 ± 0.18	0.78 ± 0.12	340 ± 60	350 ± 70	
	-		At chloro tRNA ^{Phe}		1.08 ± 0.18	0.9 ± 0.12	340 ± 100	140 ± 10	
	-		At nuc pre-tRNA ^{Cys}		1.62 ± 0.12	1.8 ± 0.12	940 ± 130	6100 ± 2100	
APPRORP2	-		At nuc pre-tRNA ^{Phe}		2.1 ± 0.12	1.38 ± 0.12	250 ± 50	350 ± 40	[57]
	-		At pre-tRNA ^{Gln}		7.7 ± 2.7		300 ± 90		
	-		Tt pre-tRNA ^{Gly}		1.8 ± 0.1				
	-	T113S			2.0 ± 0.1				
	-	R145N			2.0 ± 0.1				
	-	R145D	Tt pre-tRNA ^{Gly}		1.15 ± 0.02				
	-	T113N			1.56 ± 0.04				
	-	T113N-R145N			0.38 ± 0.02				
	-	T113N-R145D			0.047 ± 0.002				
	-		At mito pre-tRNA ^{Cys}		1.38 ± 0.12	1.32 ± 0.12	430 ± 30	300 ± 70	
APPRORP3	-		At chloro pre-tRNA ^{Phe}		1.38 ± 0.12	0.78 ± 0.12	440 ± 50	220 ± 30	[57]
	-		At nuc pre-tRNA ^{Cys}		1.80 ± 0.12	0.48 ± 0.12	420 ± 100	1500 ± 200	
	-		At nuc pre-tRNA ^{Phe}		4.32 ± 0.18	3.72 ± 1.38	2000 ± 850	380 ± 50	
	-								

Table 2. Kinetic analyses of the importance of pre-tRNA *cis*-elements for RNase P cleavage by PRORP enzymes. At stands for *Arabidopsis thaliana*, Tt for *Thermus thermophilus*, Bs for *Bacillus subtilis*, “mito” for mitochondrial and “nuc” for nuclear. L and T indicate lengths of leader and trailer sequences respectively. “Aa” substrates represent minimal substrates lacking the tRNA anticodon and D domains as described by Brillante et al. [56]. Values for the results published by Imai et al. [58] are graphical estimates as numbers were not provided in the article. Grey boxes indicate results of single turnover kinetic experiments (with either k_{obs} , the observed first order constant, or k_{react} , the maximum rate constant, and its associated K_M), whereas green boxes indicate results of RNase P cleavage assays performed in multiple turnover conditions.

pre-tRNA	Type of Mutation on pre-tRNA	RNase P	% of Cleavage	k_{obs} or k_{react} (min^{-1})	K_M (nM)	K_D (nM)	References
At mito pre-tRNA ^{Cys}	-	AtPRORP1	100 ± 7				[20]
	ΔAC		75 ± 9				
	ΔDAC		0 ± 0				
	G18A		15 ± 2				
	G18C		10 ± 1				
	G19A		85 ± 3				
	G19C		90 ± 5				
	C56A		0 ± 0				
	C56G		0 ± 0				
	G57A		90 ± 6				
	G57C		10 ± 1				
	1CG72		100 ± 10				
Δ3'	95 ± 1						
3' CCA	5 ± 2						
At chloro pre-tRNA ^{Phe}	-		100				[58]
	C56G		18				
	C56A		30				
	C56U		22				
	A57G		60				
	A57C		25				
	A57U		26				
	A58G		15				
	A58C		36				
	A58U		35				
	A59G		100				
A59C	75						
A59U	85						
At nuc pre-tRNA ^{Gly}	L23:T10	AtPRORP2		0.7 ± 0.1		118 ± 26	[39]
	L23:T05			1.0 ± 0.1		52 ± 12	
	L23:T01			0.7 ± 0.1		17 ± 5	
	L13:T01			0.7 ± 0.1		6 ± 1	
	L08:T01			1.1 ± 0.1		3 ± 1	
Tt pre-tRNA ^{Gly}	-(14)	AtPRORP3		1.67 ± 0.03	4.8 ± 0.4		[56]
	L7			1.7 ± 0.1	3.1 ± 0.7		
	L4			1.7 ± 0.1	3.4 ± 0.7		
	L2			1.6 ± 0.1	3.4 ± 0.8		
	L1			0.17 ± 0.02	5.4 ± 2.2		
	mature (CCA)			1.6 ± 0.1	4.9 ± 1.0		
	no trailer			1.5 ± 0.1	4.6 ± 0.7		
	40-nt trailer			1.5 ± 0.1	5.3 ± 1.1		
	-			1.67 ± 0.03	4.8 ± 0.4		
	U1-A72			2.2 ± 0.1	5.3 ± 0.9		
	U-1			2.9 ± 0.1	8.1 ± 1.4		
	G-1, A73			2.3 ± 0.1	6.5 ± 1.0		
	A-1, A73			5.1 ± 0.2	7.8 ± 1.4		
	A73			1.67 ± 0.04	4.5 ± 0.6		
	ΔAC			1.48 ± 0.04	1.7 ± 0.3		
	ΔD			0.36 ± 0.02	86 ± 16		
	AaT			0.066 ± 0.002	1839 ± 168		
	Aab1T			0.33 ± 0.01	1685 ± 218		
	Aab4T			0.26 ± 0.01	1151 ± 125		
	Aab9T			0.42 ± 0.01	40 ± 6		
G18->A18		1.87 ± 0.07	22 ± 3				
G19->A19/C56 > U56		1.78 ± 0.06	7.7 ± 1.2				
C56->U56		1.81 ± 0.05	6.4 ± 0.9				
A57->C57		1.56 ± 0.05	6.7 ± 0.9				

Table 2. Cont.

pre-tRNA	Type of Mutation on pre-tRNA	RNase P	% of Cleavage	k_{obs} or k_{react} (min^{-1})	K_M (nM)	K_D (nM)	References
Bs pre-tRNA ^{Asp}	L0	AtPRORP1		NA		3400 ± 400	[57]
	L1		4.68 ± 0.18	150 ± 60			
	L2		9 ± 1.2	310 ± 20			
	L3		1.92 ± 0.06	140 ± 40			
	L4		1.5 ± 0.06	150 ± 40			
	L5		1.5 ± 0.06	190 ± 60			
	L10 L14		1.5 ± 0.06 1.2 ± 0.06	100 ± 50 100 ± 50			

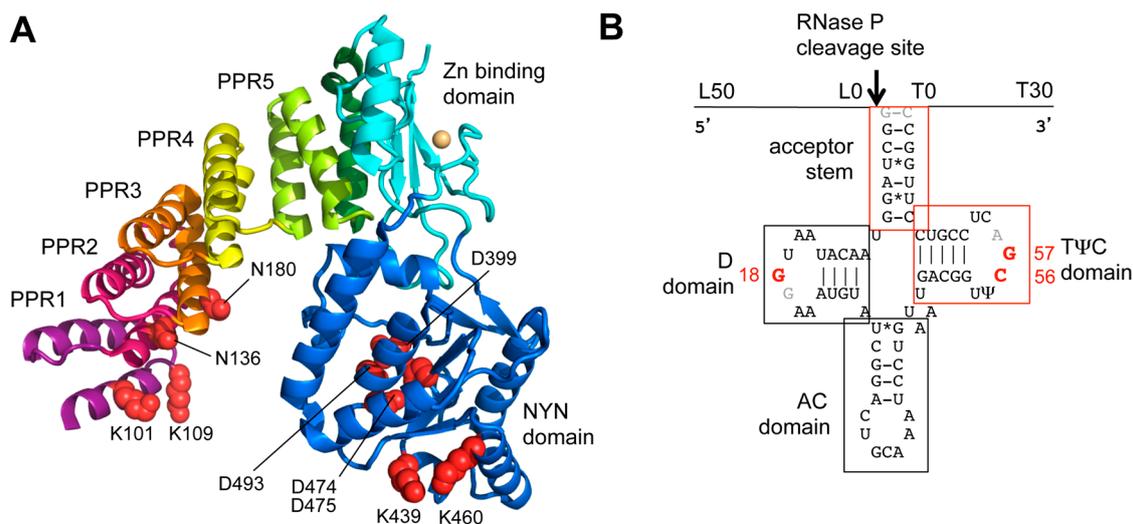


Figure 2. PRORP and pre-tRNA *cis*-elements required for RNase P activity. **(A)** Arabidopsis PRORP1 crystal structure shows the three domain organization of PRORP, with the PPR and NYN domains connected by a zinc-binding domain. Side chains of residues that were shown to interact with pre-tRNA and/or to be essential for RNase P activity by mutagenesis analyses (see Table 1) are shown in space-filling mode (red spheres); and **(B)** a secondary structure model of Arabidopsis mitochondrial pre-tRNA^{Cys}. L and T indicate the leader and trailer sequences and their respective length spans in the different substrates used in the analyses summarized in Table 2. tRNA domains and residues that were found to be essential for RNase P activity are indicated in red. Other positions where mutations had no effect on activity are shown in grey.

4.3. Requirements of Pre-tRNA Substrate *cis*-elements for PRORP Activity

Similarly, a number of studies have investigated the nature of pre-tRNA *cis*-elements required for PRORP mediated RNase P cleavage. For this, activity assays were performed with collections of either plant or bacterial pre-tRNA substrates, with varying leader and trailer sequences lengths with domain deletions as well as point mutations. Reactions were performed with the three Arabidopsis PRORP paralogues during single or multiple turnover kinetics experiments. Results summarized in Table 2 suggest that contrary to bacterial RNP RNase P, the leader and trailer sequences are not involved in substrate binding by PRORP. Optimal cleavage efficiencies could be obtained with leader sequences of at least two nucleotides and no longer than 10 nucleotides (Table 2). Transcripts with longer leaders might form alternative secondary structures *in vitro* that prevent the binding of PRORP, thus decreasing the apparent cleavage efficiency. Similarly, contrary to bacteria, the nature of the residue immediately upstream of the pre-tRNA cleavage site does not seem to be of primary importance for PRORP activity. Catalysis takes place with the four possible residues, although its efficiency increased two-fold when an A was immediately upstream of the cleavage site [56].

Deletions of tRNA domains revealed their respective importance for PRORP activity. While the deletion of the anticodon domain of pre-tRNAs did not affect RNase P activity, deletions of the D and T domains dramatically affected RNase P activity. Still, highly reduced RNase P activity could be obtained with minimal substrates composed of the acceptor and T domains, suggesting that interaction with PRORP is somehow possible in the absence of the tRNA D domain [56]. However, other minimal substrates containing the acceptor and D domains only were not investigated, thus preventing to find out whether minimal activity could also take place in the absence of the tRNA T domain. Nonetheless, results suggest that both the D and especially the T domain of pre-tRNAs play a critical role in RNase P cleavage by PRORP.

Nucleotide substitutions in pre-tRNA substrates helped map more precisely the residues required for PRORP activity. Results suggest that individual residues are necessary for RNase P activity in both the D and T loop of pre-tRNAs, in particular, G18 in the D loop and both C56 and a purine at position 57 (Table 2). Moreover, footprint experiments showed protection by PRORP for residues belonging to the D and T loops of pre-tRNAs, thus confirming PRORP contact with the corner of pre-tRNAs [20]. Altogether, results suggest that PRORP, similar to bacterial RNP RNase P docks onto the acceptor arm of pre-tRNAs, with no contact with the anticodon arm and appears to specifically interact with residues from the D and T loop at the corner of tRNAs.

4.4. PRORP Cleavage Mechanism

Beyond the common architectural features and substrate specificities of PRORP and RNP RNase P for pre-tRNA binding, the two types of biocatalyst appear to share a similar catalytic mechanism [5,19,60]. The RNP enzyme was shown to use a two-metal-ion mechanism to cleave the phosphodiester bond [54,61]. It proceeds through a nucleophilic attack of the hydroxide ion apical to the O3' of the upstream ribose, which generates products with 5'-phosphoryl and 3'-hydroxyl termini. The presence of two Mn²⁺ ions in the structure of Arabidopsis PRORP1 and metal dependency (Mn²⁺ or Mg²⁺) for PRORP cleavage suggest that protein-only enzymes also use a two-metal-ion mechanism to produce the attacking hydroxide ion and stabilize the transition state [19]. Nonetheless, PRORP1 is able to cleave substrates with an R_P-phosphorothioate modification at the canonical cleavage site using Mg²⁺ as a cofactor, which is not the case of ribozyme RNase P [60,62]. This suggests that the *pro*-R_P-oxygen of the scissile phosphate is not directly coordinated by a metal in PRORP active site, and suggests that the whole active site may adopt a different configuration. PRORP are proposed to use one hydrated divalent cation to provide the attacking hydroxide and a second hydrated metal ion to protonate the leaving group like the RNase P ribozyme [63,64].

In the latter study, Howard et al. also propose a model where the binding of one Mg²⁺ ion in the catalytic site increases the affinity of the second metal ion resulting in the cooperativity observed at low concentrations of pre-tRNAs. A comparison of the proposed mechanisms for PRORP and ribozyme RNase P activities shows that both enzymes require catalytic metal ions, although the metal ligands are different. Conserved aspartate residues coordinate the metal ions in PRORP, whereas ribozyme RNase P mainly uses non-bridging phosphodiester oxygens for metal ion coordination.

5. Comparison of PRORP and RNP RNase P Modes of Action

Altogether, biochemical studies published since the first descriptions of PRORP enzymes reveal that they share common features with RNP RNase P, but also diverge in many aspects. Alongside the discovery of protein-only RNase P, initial appreciations anticipated that the single subunit protein-only RNase P might be a much more efficient enzyme than RNP RNase P and that higher catalytic efficiency might have been the driving force for the replacement of ribozyme RNase P by PRORP during evolution. Pre-tRNA binding and cleavage studies of PRORP now suggest that it is not the case. For example, binding affinities of PRORP proteins for their pre-tRNA substrate are in the micromolar range [19,60]. These values are one or two orders of magnitude lower than for RNP RNases P. This might indicate that PRORP interaction with pre-tRNA substrates is more transient than that of RNP

RNase P. Still, kinetic parameters of PRORP activity were monitored *in vitro*, and it cannot be ruled out that PRORP RNA binding affinity or catalytic efficiency might be higher *in vivo*, e.g., in a specific biochemical environment where the respective PRORP proteins accumulate and/or in complex with other factors that might modulate PRORP activity *in vivo*.

RNA binding assays revealed that the two types of enzymes are clearly distinct at the level of substrate recognition in the vicinity of the cleavage point. Hence, PRORP binding to pre-tRNA does not seem to involve specific interactions with pre-tRNA leader and trailer sequences contrary to RNP RNase P. In contrast, the two types of enzyme have a similar organization with the specificity and catalytic domains of P RNAs being analogous to the PPR and NYN domains of PRORP proteins. Moreover, the two types of enzymes appear to dock on the acceptor arm of pre-tRNAs and to make contact with specific residues at the corner of tRNAs formed by the stacking of T and D loops of tRNAs. However, despite these important similarities, PRORP cannot be considered as a true structural mimic of the RNase P ribozyme, at least not in the sense where the exact three-dimensional shape of a nucleic acid is mimicked by a protein for functional reasons, e.g., pathogen proteins shaped as DNA to avoid host defenses or ribosome recycling factors shaped as a tRNA to dissociate ribosomal subunits [65]. Similarities in substrate recognition processes of the two types of RNase P enzymes rather point to a remarkable case of convergent evolution, where a protein-only enzyme that appeared in eukaryotes independently evolved a substrate binding process reminiscent from that of the ancestral RNA-based RNase P.

6. Concluding Remarks

While major progress has been achieved to understand the biological functions [5] and evolution of protein-only RNase P [11], as well as to determine its three-dimensional architecture and mode of action (reviewed here), major questions remain. One of the key issues will be to precisely understand how substrate specificity is achieved and what the exact involvement of the individual PPR motifs in this process is. It was predicted that separate PPR motifs of PRORP might specifically interact with individual residues in the T and/or D loops of pre-tRNAs [5,56,58]. If PPR motifs of PRORP conform to the overall mode of action of PPR proteins [32–35] this would imply that PPR motifs of PRORP probably interact with the Watson-Crick side of residues, such as C56, R57, or G18. This would, in turn, imply that pre-tRNA substrates recognized by PRORP are not stabilized by Watson-Crick interactions that occur in mature tRNAs, i.e., the C56-G19 interaction or the G18-Ψ55 interaction. This would mean that these interactions that are required for the overall structure and stability of mature tRNAs [66] are only established upon release of the 5' mature tRNA products from PRORP. The precise understanding of PRORP substrate specificity and the definition of its minimal substrate might give clues to identify transcriptome wide spectra of PRORP substrates *in vivo*. Similar to RNP RNase P that are involved in the maturation of a wide array of transcripts beyond pre-tRNAs [14], PRORP might as well be involved in the cleavage of non-tRNA transcripts. For instance, in *Arabidopsis*, published examples already show that PRORP1 is involved in the cleavage of the mitochondrial *nad6* and *orf291* RNAs at the level of predicted tRNA-like structures [22,67] and that PRORP2/3 are involved in the cleavage of a small nucleolar RNA (snoRNA) precursor in the nucleus [23]. Other challenges will be to understand the precise functions and mechanism for PRORP interactions with specific protein partners, i.e., TRMT10C and SDR5C1 in human mitochondria.

Answering the open questions regarding PRORP mode of action will rely on the determination of three-dimensional structures of PRORP either in complex with pre-tRNA substrates and/or with 5' mature tRNA products. Such structures obtained either in solution or at atomic resolution will reveal the exact RNA binding and cleavage properties of PRORP and uncover the plasticity and dynamics of PRORP maturation of tRNA precursors.

Acknowledgments: This work was supported by the French “Centre National de la Recherche Scientifique”, by the University of Strasbourg, an ANR Blanc research grant “PRO-RNase P” (ANR-11-BSV8-008-01/02) to P.G. and C.S. and by the LabEx consortium “MitoCross” in the frame of the French National Program “Investissement d’Avenir” (ANR-11-LABX-0057_MITOCROSS).

Author Contributions: All authors contributed to the writing of the article and the preparation of the figures.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NYN	N4BP1, YacP-like Nuclease
PPR	pentatricopeptide repeat
P-RNA	ribonucleoprotein RNase P RNA subunit
Pre-tRNA	precursor tRNA
PRORP	protein-only RNase P
RNP	ribonucleoprotein
tRNA	transfer RNA

References

- Altman, S. A view of RNase P. *Mol. Biosyst.* **2007**, *3*, 604–607. [[CrossRef](#)] [[PubMed](#)]
- Wang, M.J.; Davis, N.W.; Gegenheimer, P. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J.* **1988**, *7*, 1567–1574. [[PubMed](#)]
- Rossmannith, W.; Karwan, R.M. Characterization of human mitochondrial RNase P: Novel aspects in tRNA processing. *Biochem. Biophys. Res. Commun.* **1998**, *247*, 234–241. [[CrossRef](#)] [[PubMed](#)]
- Salavati, R.; Panigrahi, A.K.; Stuart, K.D. Mitochondrial ribonuclease P activity of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **2001**, *115*, 109–117. [[CrossRef](#)]
- Pinker, F.; Bonnard, G.; Gobert, A.; Gutmann, B.; Hammani, K.; Sauter, C.; Gegenheimer, P.A.; Giegé, P. PPR proteins shed a new light on RNase P biology. *RNA Biol.* **2013**, *10*, 1457–1468. [[CrossRef](#)] [[PubMed](#)]
- Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849–857. [[CrossRef](#)]
- Guerrier-Takada, C.; Altman, S. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science* **1984**, *223*, 285–286. [[CrossRef](#)] [[PubMed](#)]
- Guerrier-Takada, C.; Lumelsky, N.; Altman, S. Specific interactions in RNA enzyme-substrate complexes. *Science* **1989**, *246*, 1578–1584. [[CrossRef](#)] [[PubMed](#)]
- Randau, L.; Schroder, I.; Söll, D. Life without RNase P. *Nature* **2008**, *453*, 120–123. [[CrossRef](#)] [[PubMed](#)]
- Hartmann, E.; Hartmann, R.K. The enigma of ribonuclease P evolution. *Trends Genet.* **2003**, *19*, 561–569. [[CrossRef](#)] [[PubMed](#)]
- Lechner, M.; Rossmannith, W.; Hartmann, R.K.; Tholken, C.; Gutmann, B.; Giegé, P.; Gobert, A. Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol. Biol. Evol.* **2015**, *32*, 3186–3193. [[CrossRef](#)] [[PubMed](#)]
- Seif, E.; Leigh, J.; Liu, Y.; Roewer, I.; Forget, L.; Lang, B.F. Comparative mitochondrial genomics in zygomycetes: Bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Res.* **2005**, *33*, 734–744. [[CrossRef](#)] [[PubMed](#)]
- Ellis, F.C.; Brown, J.W. The evolution of RNase P and its RNA. In *Ribonuclease P*; Liu, F., Altman, S., Eds.; Springer: New York, NY, USA, 2010; pp. 17–40.
- Lai, L.B.; Vioque, A.; Kirsebom, L.A.; Gopalan, V. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: Challenges and prospects. *FEBS Lett.* **2010**, *584*, 287–296. [[CrossRef](#)] [[PubMed](#)]
- Lai, L.B.; Cho, I.-M.; Chen, W.-Y.; Gopalan, V. Archaeal RNase P: A Mosaic of Its Bacterial and Eukaryal Relatives. In *Ribonuclease P*; Liu, F., Altman, S., Eds.; Springer: New York, NY, USA, 2010; pp. 153–172.
- Walker, S.C.; Marvin, M.C.; Engelke, D.R. Eukaryote RNase P and RNase MRP. In *Ribonuclease P*; Liu, F., Altman, S., Eds.; Springer: New York, NY, USA, 2010; pp. 173–202.
- Anantharaman, V.; Aravind, L. The NYN domains: Novel predicted mases with a pin domain-like fold. *RNA Biol.* **2006**, *3*, 18–27. [[CrossRef](#)] [[PubMed](#)]

18. Giegé, P. Pentatricopeptide repeat proteins: A set of modular rna-specific binders massively used for organelle gene expression. *RNA Biol.* **2013**, *10*, 1417–1418. [[CrossRef](#)] [[PubMed](#)]
19. Howard, M.J.; Lim, W.H.; Fierke, C.A.; Koutmos, M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16149–16154. [[CrossRef](#)] [[PubMed](#)]
20. Gobert, A.; Pinker, F.; Fuchsbaauer, O.; Gutmann, B.; Boutin, R.; Roblin, P.; Sauter, C.; Giegé, P. Structural insights into protein-only RNase P complexed with trna. *Nat. Commun.* **2013**, *4*, 1353. [[CrossRef](#)] [[PubMed](#)]
21. Holzmann, J.; Frank, P.; Löffler, E.; Bennett, K.L.; Gerner, C.; Rossmannith, W. RNase P without RNA: Identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **2008**, *135*, 462–474. [[CrossRef](#)] [[PubMed](#)]
22. Gobert, A.; Gutmann, B.; Taschner, A.; Gößringer, M.; Holzmann, J.; Hartmann, R.K.; Rossmannith, W.; Giegé, P. A single Arabidopsis organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.* **2010**, *17*, 740–744. [[CrossRef](#)] [[PubMed](#)]
23. Gutmann, B.; Gobert, A.; Giegé, P. Prorp proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* **2012**, *26*, 1022–1027. [[CrossRef](#)] [[PubMed](#)]
24. Täschner, A.; Weber, C.; Buzet, A.; Hartmann, R.K.; Hartig, A.; Rossmannith, W. Nuclear RNase P of *Trypanosoma brucei*: A single protein in place of the multicomponent RNA-protein complex. *Cell Rep.* **2012**, *2*, 19–25. [[CrossRef](#)] [[PubMed](#)]
25. Lai, L.B.; Bernal-Bayard, P.; Mohannath, G.; Lai, S.M.; Gopalan, V.; Vioque, A. A functional RNase P protein subunit of bacterial origin in some eukaryotes. *Mol. Genet. Genom.* **2011**, *286*, 359–369. [[CrossRef](#)] [[PubMed](#)]
26. Sugita, C.; Komura, Y.; Tanaka, K.; Kometani, K.; Satoh, H.; Sugita, M. Molecular characterization of three PRORP proteins in the moss *Physcomitrella patens*: Nuclear prorp protein is not essential for moss viability. *PLoS ONE* **2014**, *9*, e108962. [[CrossRef](#)] [[PubMed](#)]
27. Bonnard, G.; Gobert, A.; Pinker, F.; Arrivé, M.; Salinas, T.; Giegé, P. A single gene encodes both organelles and nuclear RNase P enzymes in *Chlamydomonas reinhardtii*. *Plant J.* **2016**. in press.
28. Wang, G.; Chen, H.W.; Oktay, Y.; Zhang, J.; Allen, E.L.; Smith, G.M.; Fan, K.C.; Hong, J.S.; French, S.W.; McCaffery, J.M.; et al. PNPase regulates RNA import into mitochondria. *Cell* **2010**, *142*, 456–467. [[CrossRef](#)] [[PubMed](#)]
29. Rossmannith, W. Of P and Z: Mitochondrial tRNA processing enzymes. *Biochim. Biophys. Acta* **2012**, *1819*, 1017–1026. [[CrossRef](#)] [[PubMed](#)]
30. Weber, C.; Hartig, A.; Hartmann, R.K.; Rossmannith, W. Playing RNase P evolution: Swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet.* **2014**, *10*, e1004506. [[CrossRef](#)] [[PubMed](#)]
31. Serruya, R.; Orlovetskie, N.; Reiner, R.; Dehtiar-Zilber, Y.; Wesolowski, D.; Altman, S.; Jarrous, N. Human RNase P ribonucleoprotein is required for formation of initiation complexes of RNA polymerase iii. *Nucleic Acids Res.* **2015**, *43*, 5442–5450. [[CrossRef](#)] [[PubMed](#)]
32. Barkan, A.; Rojas, M.; Fujii, S.; Yap, A.; Chong, Y.S.; Bond, C.S.; Small, I. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **2012**, *8*, e1002910. [[CrossRef](#)] [[PubMed](#)]
33. Takenaka, M.; Zehrmann, A.; Brennicke, A.; Graichen, K. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* **2013**, *8*, e65343. [[CrossRef](#)] [[PubMed](#)]
34. Yagi, Y.; Hayashi, S.; Kobayashi, K.; Hirayama, T.; Nakamura, T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **2013**, *8*, e57286. [[CrossRef](#)] [[PubMed](#)]
35. Cheng, S.; Gutmann, B.; Zhong, X.; Ye, Y.; Fisher, M.F.; Bai, F.; Castleden, I.; Song, Y.; Song, B.; Huang, J.; et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* **2016**. in press. [[CrossRef](#)] [[PubMed](#)]
36. Reinhard, L.; Sridhara, S.; Hallberg, B.M. Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res.* **2015**, *43*, 5664–5672. [[CrossRef](#)] [[PubMed](#)]
37. Li, F.; Liu, X.; Zhou, W.; Yang, X.; Shen, Y. Auto-inhibitory mechanism of the human mitochondrial RNase P protein complex. *Sci. Rep.* **2015**, *5*, 9878. [[CrossRef](#)] [[PubMed](#)]
38. Pinker, F.; Giegé, P.; Sauter, C. Crystallization and crystallographic analysis of an Arabidopsis nuclear proteinaceous RNase P. *Acta Crystallogr. F Struct. Biol. Commun.* **2015**, *71*, 1372–1377. [[CrossRef](#)] [[PubMed](#)]

39. Karasik, A.; Shanmuganathan, A.; Howard, M.J.; Fierke, C.A.; Koutmos, M. Nuclear protein-only ribonuclease P2 structure and biochemical characterization provide insight into the conserved properties of tRNA 5' end processing enzymes. *J. Mol. Biol.* **2016**, *428*, 26–40. [[CrossRef](#)] [[PubMed](#)]
40. Small, I.D.; Peeters, N. The PPR motif—A TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **2000**, *25*, 46–47. [[CrossRef](#)]
41. Zeytuni, N.; Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **2012**, *20*, 397–405. [[CrossRef](#)] [[PubMed](#)]
42. Ringel, R.; Sologub, M.; Morozov, Y.I.; Litonin, D.; Cramer, P.; Temiakov, D. Structure of human mitochondrial RNA polymerase. *Nature* **2011**, *478*, 269–273. [[CrossRef](#)] [[PubMed](#)]
43. Yin, P.; Li, Q.; Yan, C.; Liu, Y.; Liu, J.; Yu, F.; Wang, Z.; Long, J.; He, J.; Wang, H.W.; et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **2013**, *504*, 168–171. [[CrossRef](#)] [[PubMed](#)]
44. Ke, J.; Chen, R.Z.; Ban, T.; Zhou, X.E.; Gu, X.; Tan, M.H.; Chen, C.; Kang, Y.; Brunzelle, J.S.; Zhu, J.K.; et al. Structural basis for RNA recognition by a dimeric PPR protein complex. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1377–1382. [[CrossRef](#)] [[PubMed](#)]
45. Ban, T.; Ke, J.; Chen, R.; Gu, X.; Tan, M.H.; Zhou, X.E.; Kang, Y.; Melcher, K.; Zhu, J.K.; Xu, H.E. Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.* **2013**, *288*, 31540–31548. [[CrossRef](#)] [[PubMed](#)]
46. Glavan, F.; Behm-Ansmant, I.; Izaurralde, E.; Conti, E. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J.* **2006**, *25*, 5117–5125. [[CrossRef](#)] [[PubMed](#)]
47. Xu, J.; Peng, W.; Sun, Y.; Wang, X.; Xu, Y.; Li, X.; Gao, G.; Rao, Z. Structural study of mcpip1 N-terminal conserved domain reveals a PIN-like RNase. *Nucleic Acids Res.* **2012**, *40*, 6957–6965. [[CrossRef](#)] [[PubMed](#)]
48. Bhagwat, M.; Meara, D.; Nossal, N.G. Identification of residues of T4 RNase H required for catalysis and DNA binding. *J. Biol. Chem.* **1997**, *272*, 28531–28538. [[CrossRef](#)] [[PubMed](#)]
49. Gully, B.S.; Cowieson, N.; Stanley, W.A.; Shearston, K.; Small, I.D.; Barkan, A.; Bond, C.S. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atph RNA. *Nucleic Acids Res* **2015**, *43*, 1918–1926. [[CrossRef](#)] [[PubMed](#)]
50. Salinas-Giegé, T.; Giegé, R.; Giegé, P. tRNA biology in mitochondria. *Int. J. Mol. Sci.* **2015**, *16*, 4518–4559. [[CrossRef](#)] [[PubMed](#)]
51. Vilardo, E.; Nachbagauer, C.; Buzet, A.; Taschner, A.; Holzmann, J.; Rossmannith, W. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase-extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* **2012**, *40*, 11583–11593. [[CrossRef](#)] [[PubMed](#)]
52. LaGrandeur, T.E.; Darr, S.C.; Haas, E.S.; Pace, N.R. Characterization of the RNase P RNA of *Sulfolobus acidocaldarius*. *J. Bacteriol.* **1993**, *175*, 5043–5048. [[PubMed](#)]
53. Pan, T.; Loria, A.; Zhong, K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12510–12514. [[CrossRef](#)] [[PubMed](#)]
54. Reiter, N.J.; Osterman, A.; Torres-Larios, A.; Swinger, K.K.; Pan, T.; Mondragon, A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **2010**, *468*, 784–789. [[CrossRef](#)] [[PubMed](#)]
55. Kurz, J.C.; Niranjanakumari, S.; Fierke, C.A. Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA^{Asp}. *Biochemistry* **1998**, *37*, 2393–2400. [[CrossRef](#)] [[PubMed](#)]
56. Brillante, N.; Gossringer, M.; Lindenhofer, D.; Toth, U.; Rossmannith, W.; Hartmann, R.K. Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.* **2016**, *44*, 2323–2336. [[CrossRef](#)] [[PubMed](#)]
57. Howard, M.J.; Karasik, A.; Klemm, B.P.; Mei, C.; Shanmuganathan, A.; Fierke, C.A.; Koutmos, M. Differential substrate recognition by isozymes of plant protein-only ribonuclease P. *RNA* **2016**, *22*, 782–792. [[CrossRef](#)] [[PubMed](#)]
58. Imai, T.; Nakamura, T.; Maeda, T.; Nakayama, K.; Gao, X.; Nakashima, T.; Kakuta, Y.; Kimura, M. Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in *Arabidopsis thaliana* play a crucial role in recognition of nucleotide bases at TYC loop in precursor tRNAs. *Biochem. Biophys. Res. Commun.* **2014**, *450*, 1541–1546. [[CrossRef](#)] [[PubMed](#)]
59. Chen, T.H.; Tanimoto, A.; Shkriabai, N.; Kvaratskhelia, M.; Wysocki, V.; Gopalan, V. Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme. *Nucleic Acids Res.* **2016**. in press. [[CrossRef](#)] [[PubMed](#)]

60. Pavlova, L.V.; Gossringer, M.; Weber, C.; Buzet, A.; Rossmann, W.; Hartmann, R.K. tRNA processing by protein-only versus RNA-based RNase P: Kinetic analysis reveals mechanistic differences. *Chembiochem* **2012**, *13*, 2270–2276. [[CrossRef](#)] [[PubMed](#)]
61. Steitz, T.A.; Steitz, J.A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6498–6502. [[CrossRef](#)] [[PubMed](#)]
62. Thomas, B.C.; Li, X.; Gegenheimer, P. Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. *RNA* **2000**, *6*, 545–553. [[CrossRef](#)] [[PubMed](#)]
63. Donghi, D.; Schnabl, J. Multiple roles of metal ions in large ribozymes. *Met. Ions Life Sci.* **2011**, *9*, 197–234. [[PubMed](#)]
64. Howard, M.J.; Klemm, B.P.; Fierke, C.A. Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.* **2015**, *290*, 13454–13464. [[CrossRef](#)] [[PubMed](#)]
65. Tsonis, P.A.; Dwivedi, B. Molecular mimicry: Structural camouflage of proteins and nucleic acids. *Biochim. Biophys. Acta* **2008**, *1783*, 177–187. [[CrossRef](#)] [[PubMed](#)]
66. Giegé, R. Toward a more complete view of tRNA biology. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1007–1014. [[CrossRef](#)] [[PubMed](#)]
67. Fujii, S.; Suzuki, T.; Giegé, P.; Higashiyama, T.; Koizuka, N.; Shikanai, T. The restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. *Plant J.* **2016**. in press. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

II. Objectifs de ma thèse

Bien que des progrès significatifs aient été réalisés ces dernières années pour comprendre la fonction et le mode d'action des RNase P protéiques de type PRORP, de nombreuses questions restaient ouvertes. *Arabidopsis thaliana* comporte trois protéines PRORP. AtPRORP1 est localisée dans la mitochondrie et le chloroplaste alors que AtPRORP2 et AtPRORP3 sont toutes deux localisées dans le noyau. Il est connu que ces protéines ont d'autres substrats que les pré-ARNt. En effet, AtPRORP1 mature l'ARNm de NAD6 dans la mitochondrie et AtPRORP2 et AtPRORP3 maturent le transcrit précurseur snoR43-ARNt^{Gly}. Une première question essentielle était de comprendre comment les motifs PPR de PRORP peuvent assurer la spécificité des substrats de PRORP. En effet, alors que les motifs PPR sont caractérisés comme des motifs d'interaction avec l'ARN, l'importance respective de chaque motif pour la reconnaissance des pré-ARNt par PRORP n'a pas été étudiée. Une autre question fondamentale était de déterminer l'ensemble des substrats des protéines PRORP au niveau des transcriptomes entiers, afin de mieux appréhender les fonctions des protéines PRORP dans le cadre global de la cellule végétale.

Ainsi, le premier objectif de ma thèse a été de déterminer le rôle respectif des motifs PPR de PRORP. Sur la base d'analyses biochimiques et structurales nous avons proposé un modèle d'interaction entre AtPRORP2 et les ARNt et pu en déduire quel motif pouvait être en interaction avec quelle base. Les interactions PPR / ARN ont été étudiées biochimiquement en mutant les positions 5 et 35 de chaque motif PPR de AtPRORP2 et en analysant l'affinité des protéines mutées pour les ARN, i.e. par ultracentrifugation analytique et en observant l'effet des mutations sur l'activité catalytique.

Le second objectif de la thèse a été de déterminer le spectre de substrats des protéines PRORP. Une étude pilote a été réalisée avec AtPRORP1 pour déterminer ses substrats ARN dans la mitochondrie d'*Arabidopsis*. Pour cela, des plantes complémentées avec une version étiquetée de AtPRORP1 déjà disponibles au laboratoire, ont été utilisées pour co-immunoprécipiter AtPRORP1 avec ses substrats ARN. Les ARN ont été utilisés pour construire des banques d'ADNc et identifier les substrats par séquençage à haut débit. Par ailleurs, des plantes complémentées avec une version étiquetée de AtPRORP2 ont été construites afin de déterminer l'ensemble des substrats des protéines PRORP nucléaires chez *Arabidopsis*.

L'ensemble de ces travaux avait pour but de faire progresser à la fois la compréhension du mode d'action des protéines PRORP et de mieux connaître leur intégration dans l'ensemble des mécanismes d'expression génétique chez les plantes.

III. Matériel

1. Souches bactériennes

a. Souches d'*Escherichia coli*

Les souches bactériennes BL21 et TOP10 d'*E. coli* ont été utilisées.

TOP 10

La souche TOP10 est une souche optimisée permettant une production importante de plasmide grâce à la mutation des gènes *endA1* et *recA* supprimant respectivement l'activité endonucléase intracellulaire et les recombinaisons homologues.

BL21

La souche BL21 est une souche d'expression comportant le gène de l'ARN polymérase T7 sous le contrôle du promoteur *lacUV5*. L'induction du promoteur par l'IPTG permet l'expression de la polymérase T7 qui à son tour permet l'expression de gènes placés sous le contrôle du promoteur T7.

Génotypes :

- BL21(DE3) : F⁻ ompT hsdSB(rB⁻, mB⁻) gal dcm (DE3)
- TOP10 : F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG

b. Souches d'*Agrobacterium tumefaciens*

La souche bactérienne GV3101 :: pMP90 d'*Agrobacterium tumefaciens* est utilisée pour transformer les plantes par immersion des hampes florales. La souche GV3101 a été développée à l'Université de Ghent dans le département Genetische Virologie (GV) par insertion du plasmide helper pMP90, un variant sans la région T du pTi C58.

2. Matériel végétal

Toutes les plantes utilisées sont des *Arabidopsis thaliana* écotype Col 0 sauf le mutant d'insertion d'ADN-T pour le gène *AtPRORP2* qui est de l'écotype Ws4.

a. Plantes p2/p3

Les plantes doubles mutantes AtPRORP2/PRORP3 dont la mutation knock-out *prorp2* est à l'état homozygote et la mutation knock-out *prorp3* à l'état hétérozygote ont été obtenues par croisement de deux mutants d'insertion ADN-T (EHQ_F435C08 et SALK_126397), suivi d'une autofécondation de la génération F1.

b. Plantes AtPRORP2-HA

Les plantes exprimant AtPRORP2-HA sont obtenues par agro-transformation de plantes *prorp2prorp3/prorp2PRORP3* avec une construction exprimant AtPRORP2 en fusion avec une étiquette HA, placée sous le contrôle du promoteur endogène de AtPRORP2 cloné dans le vecteur pGWB1 (voir ci-dessous).

c. Plantes AtPRORP1-HA

Les plantes AtPRORP1-HA ont été obtenues par Ayoub Bouchoucha par complémentation du mutant ADN-T GK_385G09 avec une construction exprimant une fusion AtPRORP1-HA placée sous le contrôle du promoteur endogène de AtPRORP1.

3. Plasmides

a. pET28b

Ce plasmide est utilisé pour l'expression de protéines recombinantes en système bactérien. Les séquences placées sous le contrôle d'un promoteur T7 sont exprimées pour produire une fusion avec une étiquette 6-Histidines à l'extrémité N-terminale de la protéine.

b. pUC19

Ce plasmide qui se réplique en grande quantité est utilisé ici pour cloner les différents substrats ARN utilisés. L'intégration d'un promoteur T7 lors du clonage du gène d'intérêt permet la transcription *in vitro*.

c. pDONR207

Ce plasmide est un vecteur d'entrée (ou « donneur ») de la technologie Gateway. Il comporte le gène de résistance pour la gentamicine.

d. pGWB1

Ce plasmide est un vecteur de destination de la technologie Gateway et un plasmide binaire. Dans les bactéries, il exprime une résistance pour la kanamycine et dans les plantes pour l'hygromicine. Aucun promoteur n'est présent en amont de la cassette de clonage et aucune fusion à une étiquette n'est codée dans le plasmide.

4. Amorces

a. Amorces de mutagenèse

Tableau 1 : Amorces utilisées lors de la mutagenèse dirigée de AtPRORP2
F : Forward, R : Reverse

Mutation	Sens	Amorce
L34T	F	5' ggagcatgaattgagattgaagtgagattgtttctgggttctg 3'
	R	5' cagaaaccagaaacaaatctcacctcaatctcaattcatgctcc 3'
S65D	F	5' ctccagtgaagtacgtctcgaccagcaacattttcaaacgc 3'
	R	5' gcgttgaaaatgttgctggctcgagacgtacttactggaag 3'
Q70T	F	5' gtgaagtacgtctcagccagcaacattttacgacgttctgtaccta 3'
	R	5' taggtacagaagcgtcgtaaaatgttgctggctgagacgtacttcac 3'
N108D	F	5' gttcggggataagtctctgatgagcatctgttact 3'
	R	5' agtaacagatgcctcatcaggacttatccccgaac 3'
T113N	F	5' ccagtctagccactgaattaacagatgcctcattag 3'
	R	5' ctaatgaggcatctgttaattcagtggttagactgg 3'
R145N	F	5' tctgttgagggtgtgtcgattcctaactgaggacatag 3'
	R	5' catatgtcctcagattaggaatcgacacacctccaacaga 3'
A150N	F	5' gtgtcgattcctcgtctgaggacatataatccggctttgctctg 3'
	R	5' cagagcaaagccggattatatgtcctcagacgaggaatcgacac 3'
E180N	F	5' gctgcaggtattgcattgaatgaggcagagatatcggtct 3'
	R	5' agccgatatctctgcctcattcaatgcaatacctgcagc 3'
S185T	F	5' gcattggaggaggcagagataacggctttgttg 3'
	R	5' caacaaagccgttatctctgcctcctccaatgc 3'
S215D	F	5' agttgagggaaatgttggtgcgttgatgaagaaactttgaaaattattg 3'
	R	5' caataattttcaaagtttctcatcaacgcagccaacatattccctcaact 3'

b. Amorces de PCR

- Clonage de la séquence 5' UTR de PRORP2 et ajout d'une étiquette HA en 3' ainsi que des sites de recombinaison Gateway AttB en 5' et 3'

Attb-5'UTR-PRORP2 :

5' GGGGACAAGTTTGTACAAAAAAGCAGGCTTTCATAGTCATGGCATTAG
C 3'

PRORP2-HA-STOP-AttB :

5' GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAAGCGTAATCTGGAACA
TCGTATGGGTAAGGAATCTTCCCATTACTCTT 3'

c. Amorces de séquençage

- Séquençage de AtPRORP2 dans pET28b

Promoteur T7 : 5' TAATACGACTCACTATAGGG 3'

Terminateur T7 : 5' GCTAGTTATTGCTCAGCGG 3'

IIIG74 : 5' GGCATCTGTTACTTCAGTGGCTA 3'

19Z0774 : 5' TTTGAATAATGGAGGTGGGTG 3'

20Z0896 : 5' GACCCGTCTCTTGTGCAATAG 3'

- Séquençage de la séquence 5' UTR de AtPRORP2 dans pDONR207

pDONR207 Seq cassette F : 5' TCGCGTTAACGCTAGCATGGATCTC 3'

pDONR207 Seq cassette R : 5' GTAACATCAGAGATTTTGAGACAC 3'

IIIG74 : 5' GGCATCTGTTACTTCAGTGGCTA 3'

19Z0774 : 5' TTTGAATAATGGAGGTGGGTG 3'

20Z0896 : 5' GACCCGTCTCTTGTGCAATAG 3'

IIIG49 : 5' TTCATAGTCATGGCATTAGCTT 3'

VG63 : 5' GTCGAATTCGTGGTCTCTCC 3'

d. Amorces de génotypage des plantes

PRORP2 WT/HA : IIIG74 : 5' GGCATCTGTTACTTCAGTGGCTA 3'

VIIG66 : 5' CATTGTATTACCACTTTCACGG 3'

PRORP2 ADN-T : 21Z0908 : 5' CGTGTGCCAGGTGCCACGGAATAGT 3'

VIIG66 : 5' CATTGTATTACCACTTTCACGG 3'

PRORP3 WT : IIIG14 : 5' AGCCGGAGCATAAGTTCTCA 3'^e

IIIG13 : 5' TGCATGTTTCCTTAAACGCTATG 3'
PRORP3 ADN-T : 20Z0885 : 5' TGGTTCACGTAGTGGGCCATCG 3'
IIIG13 : 5' TGCATGTTTCCTTAAACGCTATG 3'

IV. Méthodes

1. Techniques de clonage

Amplification de fragments d'ADN par PCR

La réaction de polymérisation en chaîne d'ADN (PCR pour « polymerase chain reaction ») permet de dupliquer en grand nombre une séquence d'ADN connue grâce à des amorces spécifiques. La réaction se passe en 3 étapes principales. La première consiste à dénaturer l'ADN à 95°C, la seconde permet d'hybrider les amorces en se plaçant à une température appropriée. La troisième est l'étape d'élongation de l'ADN, réalisée à 72°C, température optimale de polymérisation pour les Taq ADN polymérases classiques. Ces étapes sont répétées 30 fois en général (le nombre de cycles peut varier selon les applications), et suivies par une étape d'élongation finale de 5 min permettant de terminer la synthèse des brins d'ADN. Trois enzymes de PCR ont été utilisées ici : la GoTaq ADN polymérase, la Pfu ADN polymérase (Promega) et la Phusion DNA polymérase (NEB). Ces deux dernières enzymes possèdent une activité de relecture, exonucléase de 3' vers 5'. Elles sont donc plus fidèles et sont utilisées pour les clonages. Les programmes de PCR suivent les recommandations des fournisseurs. Les réactions sont réalisées dans une machine T3000 Thermocycler (Biometra).

Clonage par la technologie Gateway

La technologie Gateway est une technologie de clonage très versatile basée sur la recombinaison homologue qui permet d'insérer une séquence d'intérêt dans un plasmide donneur pour ensuite pouvoir la transférer facilement dans de nombreux plasmides accepteurs.

Un fragment d'ADN est amplifié par PCR grâce à des amorces spécifiques contenant à leurs extrémités les sites attB1 et attB2. Les sites attB recombinent avec les sites attP des plasmides donneurs par l'intermédiaire de la « BP clonase ». Le fragment peut ensuite être

intégré par d'autres recombinaisons utilisant la « LR clonase » dans divers plasmides accepteurs de la technologie Gateway possédant des sites attR.

Clonage utilisant des enzymes de restriction

Les fragments d'ADN amplifiés avec des amorces contenant des sites de restriction ainsi que les vecteurs sont digérés par des enzymes de restriction selon les conditions recommandées par le fournisseur. Les fragments de petite taille tels que ceux représentant les précurseurs d'ARNt sont insérés dans pUC19 avec un ratio molaire de 1 : 1 dans le milieu réactionnel. 1 pmol de vecteur et 1 pmol d'insert sont mis en présence de 3U de T4 DNA ligase dans un volume réactionnel de 10 μ L contenant 5 μ L de tampon ligase 2x. La réaction est incubée 16 h à 20°C.

2. Mutagenèse dirigée de fragments d'ADN

La mutagenèse dirigée est réalisée selon deux méthodes différentes.

La première est basée sur le protocole du kit "QuikChange Site-Directed Mutagenesis" de Stratagene. Deux amorces complémentaires comportant la mutation à introduire sont conçues grâce à l'outil disponible publiquement (https://www.genomics.agilent.com/primerDesignProgram.jsp?toggle=uploadTrans&mutate=true&_requestid=420884). 25 ng du plasmide pET28b(+) contenant l'ADNc de AtPRORP2 sont utilisés dans une réaction de PCR avec 125 ng de chaque amorce, 1 μ L du mélange de dNTP fourni et 5 μ L de tampon de réaction 10X et 2,5 U de Pfu polymérase, pour un volume total de réaction de 50 μ L. La Pfu polymérase possède une activité de correction de 3' vers 5' ce qui la rend très fidèle. Le produit d'amplification par PCR n'est pas méthylé contrairement au vecteur utilisé comme matrice qui a été isolé à partir de bactéries *E. coli*. Cette différence permet de dégrader le vecteur non muté servant de matrice en digérant les produits de réaction avec l'enzyme Dpn I, une enzyme de restriction clivant seulement les fragments d'ADN méthylés. 1/10 de la réaction de digestion est utilisé pour la transformation de bactéries compétentes.

La seconde méthode est basée sur un protocole décrit par Edelheit et al (Edelheit et al., 2009). Deux réactions de polymérisation sont réalisées en parallèle chacune avec une seule des deux amorces utilisées dans la méthode précédente. Le brin néo-synthétisé n'est alors pas utilisé comme matrice ce qui empêche la propagation et l'accumulation des mutations. La présence d'une seule amorce empêche leur appariement et leur insertion dans le plasmide

néo-synthétisé. Chaque PCR est réalisée avec 400 ng de plasmide portant le gène d'intérêt, 40 pmol d'amorces sens ou antisens dans un milieu réactionnel contenant le Master Mix 2x Phusion Hot Start Flex de Biolabs dans un volume total de 25 μ L par PCR. Après 35 cycles de PCR, les deux réactions de PCR sont réunies afin d'apparier les deux brins du plasmide néo-synthétisé. Pour cela, le mélange est dénaturé pendant 5 min à 95°C puis la température est abaissée progressivement avec des étapes de 1 min à 90°C, de 1 min à 80°C puis des étapes de 30 sec tous les 10°C jusqu'à arriver à une température à 37°C. Le mélange est alors digéré pendant 2 h avec 20 U de DpnI et 1/10 du volume est utilisé pour la transformation de bactéries compétentes.

3. Transformation par choc thermique de bactéries *E. coli* TOP10

Les bactéries utilisées sont rendues « thermo-compétentes », c'est-à-dire que leur transformation est réalisée par choc thermique. Les bactéries sont mises en présence de 5 ng de plasmide ou 1/10 de la réaction de mutagenèse digéré par DpnI et incubées sur la glace pendant 30 min. Le mélange subit ensuite un choc thermique de 30 sec à 42°C permettant l'entrée du plasmide dans la cellule puis les bactéries sont placées 2 min sur la glace. Du milieu LB est ajouté aux bactéries qui sont incubées 1h à 37°C. Cette étape permet de régénérer la paroi des bactéries, d'obtenir une première division cellulaire et l'expression du gène de résistance à l'antibiotique porté par le plasmide. Les bactéries sont ensuite étalées sur du milieu LB contenant l'antibiotique approprié.

Les bactéries *E. coli* BL21 sont transformées de la même manière avec 100ng de plasmide.

4. Préparation de plasmides par lyse alcaline

Les plasmides sont purifiés avec le kit "NucleoSpin Plasmid QuickPure" de Macherey-Nagel. Ce kit repose sur le principe de la lyse alcaline. Un culot de bactéries provenant de 3 ml de culture est remis en suspension dans du tampon TE additionné de RNase A. Une seconde solution contenant du SDS et du NaOH permet de lyser les bactéries et de dénaturer l'ADN. Une troisième solution contenant de l'acétate de potassium neutralise le lysat et permet de renaturer l'ADN plasmidique alors que l'ADN génomique plus long précipite avec les débris cellulaires et les protéines. Le précipité est séparé du surnageant contenant l'ADN plasmidique par centrifugation. Le surnageant est déposé sur une colonne de silice possédant une grande affinité pour l'ADN. La colonne est lavée avec de l'éthanol à 70% qui dissout la majorité des composants autres que l'ADN adsorbé à la silice, notamment les

sels. L'élution de l'ADN plasmidique est réalisée avec de l'eau miliQ. Les différents tampons sont fournis par le fabricant.

5. Séquençage des plasmides

Le séquençage est réalisé par Abdelmalek ALIOUA et Sandrine KOECHLER de la plateforme d'Analyse de l'expression génique (<http://www.ibmp.cnrs.fr/plateformes/analyse-de-l'expression-genique/>) de l'IBMP équipée d'un séquenceur Hitachi 3130x/ Genetic Analyzer. La technique de séquençage utilisée est la méthode de Sanger basée sur l'ajout de ddNTP qui bloquent l'élongation de l'ADN polymérase de manière statistique à toutes les positions. Une PCR unidirectionnelle est réalisée avec une Taq polymérase à haute-fidélité : une étape d'hybridation de 15 sec à 50°C est suivie d'une étape d'élongation de 4 min à 60°C. La PCR est réalisée pendant 40 cycles avec 100 fois plus de dNTP que de ddNTP. Les différents fragments sont ensuite séparés par électrophorèse capillaire. Les capillaires de 45 µM de diamètre et de 50 cm de long sont au nombre de 16 et permettent une résolution de 900 à 1000 bases. A la sortie des capillaires, un laser excite les fluorochromes fixés aux ddNTPs. Chaque base possède un fluorochrome spécifique, la détection de la longueur d'onde d'émission de chaque fluorochrome permet d'identifier la base incorporée et ainsi de déterminer la séquence d'ADN.

6. Expression de protéines dans des cellules de *E. coli* BL21

Les différents plasmides contenant des séquences d'ADNc de AtPRORP2 sauvage (WT) ou mutées sont transformés dans la souche BL21 de *E. coli*. Une pré-culture de LB contenant 1 % (m/v) de glucose et l'antibiotique approprié est incubée pendant 16 h à 37°C. Le glucose permet de réprimer de l'expression de l'ARN polymérase T7 et donc indirectement de AtPRORP2 inséré dans pET28. Une culture de 1 L de LB contenant 1 % (m/v) glucose est inoculée avec 10 ml de pré-culture et incubée à 37°C sous agitation jusqu'à atteindre une DO_{600nm} de 0,6. Le milieu de culture est alors centrifugé à 5000 g, à 4°C pendant 20 min. Le culot est lavé avec du milieu LB afin d'éliminer le glucose, puis remis en suspension dans 1 L de LB avec antibiotique. L'expression est induite par addition de 0,5 mM d'IPTG. Les protéines sont produites pendant 16 h à 18°C sous agitation. Les bactéries sont centrifugées à 5000 g, à 4°C pendant 20min et lavées avec du LB. Le culot est alors immédiatement lysé ou stocké à -20°C. 1 g de bactérie est lysé dans 10 mL de tampon de lyse (15 % glycérol, 50 mM HEPES-Na pH 7,5, 5 mM imidazole, 250 mM NaCl, 0,2 mM DTT, inhibiteurs de protéases) par trois cycles de décompression à 18000 psi avec une presse de

type « French ». Le lysat est centrifugé 30 min à 4°C à 30.000 g afin de retirer les débris cellulaires puis filtré à 0,22 μ m pour retirer les particules les plus fines.

7. Purification des protéines par chromatographie d'affinité

Les protéines recombinantes possédant une étiquette poly histidine sont purifiées sur une colonne d'affinité dont la matrice porte des ions nickel immobilisés. Toutes les étapes de purification sont réalisées par FPLC (Fast Protein Liquid Chromatography) avec un appareil ÄKTA Pure. Le lysat est déposé sur une colonne "HisTrap FF crude" de 1 ml contenant une matrice Ni-NTA équilibrée avec le tampon A. Après chargement, la colonne est lavée avec 10 volumes de colonne de tampon A, puis 20 volumes avec le mélange de tampons A et B requis pour obtenir 25 mM d'imidazole. Ces différents lavages permettent de retirer les protéines non spécifiques accrochées à la colonne. Les protéines d'intérêts sont éluées avec 5 volumes de tampon B contenant 500 mM d'imidazole afin de décrocher la totalité des protéines fixées à la matrice. Le débit de tampon utilisé est de 0,5 ml/min lors du chargement et de 2 ml/min lors du lavage et de l'éluion. L'éluion des protéines est contrôlée par l'absorbance à 280 nm à la sortie de la colonne. Après chaque purification, la matrice est régénérée suivant les recommandations du constructeur.

	Hepes-Na pH7,5	NaCl	glycérol	imidazole pH7,5	DTT
Tampon A	50 mM	250 mM	15 % (m/v)	5 mM	0,2 mM
Tampon B	50 mM	250 mM	15 % (m/v)	500 mM	0,2 mM

8. Purification des protéines par filtration sur gel

La filtration sur gel est une méthode de chromatographie permettant de séparer les protéines en fonction de leur taille. Les fractions contenant les protéines purifiées par affinité sont soit injectées directement en tandem sur une colonne HiLoad Superdex 75 PG, soit concentrées jusqu'à atteindre un volume de 500 μ L pour être injectées manuellement sur une colonne Superdex 200 10/300 increase grâce à une boucle de 1 ml lavée avec 3 ml de tampon SEC1. Les débits et les pressions limites utilisés sont ceux recommandés par le constructeur. Si les protéines sont utilisées pour des tests d'activité, elles sont éluées avec le tampon SEC1. Le tampon d'éluion SEC2 est lui utilisé pour des expériences d'ultra centrifugation analytique.

	Hepes-Na pH7,5	NaCl	glycérol	DTT	MgCl ₂
Tampon SEC1	50 mM	250 mM	15 % (m/v)	0,2 mM	/
Tampon SEC2	30 mM	150 mM	5 % (m/v)	0,2 mM	5 mM

9. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE)

Les protéines sont analysées par électrophorèse de type SDS-PAGE. Cette technique permet leur séparation en condition dénaturante selon leur taille. Le gel est composé de deux parties, un gel de concentration contenant 5 % d'acrylamide et un gel de séparation contenant un pourcentage variable d'acrylamide (8 % d'acrylamide le plus souvent). La migration est effectuée dans du tampon Laemmli 1 X (25 mM Tris-HCl pH 8,3, 200 mM glycine, 0,1 % (p/v) SDS). Les protéines sont dénaturées dans du tampon de charge 1 x (50 mM Tris-HCl pH 6,8, 5 % (v/v) glycérol, 2 % (p/v) SDS, 0,1 % (p/v) bleu de bromophénol, 3 % (v/v) beta-mercaptoéthanol) avant d'être déposées dans les puits. La migration est réalisée à un ampérage constant de 25 mA jusqu'à arriver à la résolution souhaitée.

	Acrylamide/ bisacrylamide 29/1	Tris-HCl pH 6,8	Tris-HCl pH 8,8	SDS	APS	TEMED
Gel de concentration 5%	5 %	125 mM	/	0,1 % (p/v)	0,1 % (v/v)	0,01 % (v/v)
Gel de séparation 8%	8 %	/	375mM	0,1 % (p/v)	0,1 % (v/v)	0,01 % (v/v)

10. Coloration de gel de polyacrylamide

Les gels de polyacrylamide sont colorés pendant 1 heure sous agitation dans une solution contenant 0,25 % (m/v) de bleu de Coomassie R250, 40 % (v/v) de méthanol et 10 % (v/v) d'acide acétique, puis décolorés 2 heures dans une solution contenant 20 % (v/v) d'éthanol et 10 % (v/v) d'acide acétique.

11. Transcription *in vitro*

La séquence à transcrire est clonée dans le vecteur pUC19 possédant en amont du site de clonage la séquence du promoteur de l'ARN polymérase T7. Le plasmide est linéarisé en aval de la séquence de l'ARN à transcrire avec une enzyme de restriction appropriée.

a. Transcription de petites quantités pour clivage

La transcription *in vitro* d'ARN est réalisée avec un kit 'TranscriptAid T7 High Yield Transcription' de Thermo Scientific. L'ARN polymérase T7 se fixe sur son promoteur et transcrit jusqu'à l'extrémité du plasmide linéarisé. On parle de transcription « run off ». La transcription est réalisée à 37°C pendant 4 à 8 heures puis la réaction est stoppée par un traitement avec 2 U de DNase I afin d'hydrolyser l'ADN. Les nucléotides sont éliminés par filtration sur gel GE Healthcare illustra ProbeQuant G50. Les autres composants du milieu de transcription sont éliminés par une extraction au phénol / chloroforme et les ARN produits sont précipités à l'éthanol.

b. Transcription en grande quantité pour les analyses biophysiques

Les analyses biophysiques demandent de grandes quantités d'ARN que l'on ne peut pas obtenir avec le kit décrit plus haut. L'ARN est donc produit avec une T7 ARN polymérase recombinante (produite par la plateforme « Production et Purification des Protéines » de l'institut (www.ibmp.cnrs.fr/plateformes/production-et-purification-des-proteines/)). La transcription est réalisée dans un tampon contenant 30 mM Tris HCl pH 8, 10 mM DTT, 0,1 % (v/v) Triton X100, 4 mM MgCl₂, 2 mM spermidine, 4 mM de chaque rNTP, 0,4 U/ml RNase OUT, 0,1 mg/ml BSA, 0,01 µg/µL d'ADN linéarisé et 0,03 µg/µL T7 RNA polymérase pendant 16 h à 37°C.

c. Purification de grandes quantités de transcrits

Les transcrits obtenus sont purifiés par extraction au phénol / chloroforme suivie d'une précipitation avec un volume égal d'isopropanol. Le culot est repris dans un tampon dénaturant contenant 95 % (v/v) de formamide, 20 mM EDTA, 0,05 % (p/v) bleu de bromophénol, 0,05 % (p/v) xylène cyanol, puis les ARN sont déposés sur un gel de polyacrylamide dénaturant contenant 7M d'urée afin de séparer les différents transcrits et la matrice ADN. La bande contenant le transcrit est découpée, broyée, additionnée de TBE 1x

incubée sous agitation pendant 16 h puis le mélange est filtré. Les transcrits contenus dans le filtrat sont précipités à l'éthanol.

12. Extraction des acides nucléiques au phénol / chloroforme

L'extraction des acides nucléiques est réalisée par l'ajout d'un volume d'un mélange de phénol / chloroforme (v/v) saturé avec un tampon Tris-HCl pH 6,8. L'extraction est réalisée à l'aide d'un vortex de paillasse pendant 1 min à température ambiante. L'extrait est centrifugé 15 min à 16 000 g afin de séparer la phase organique de la phase aqueuse contenant les acides nucléiques, les protéines se trouvant à l'interphase.

13. Précipitation à l'éthanol des acides nucléiques

La solution contenant les acides nucléiques est supplémentée de $1/10^{\text{ème}}$ de volume d'acétate d'ammonium 3 M pH 5,3 et de $1/50^{\text{ème}}$ de volume de glycogène. Les sels permettent de précipiter l'acide nucléique, le glycogène est un entraîneur, facilitant la précipitation. 2,5 volumes d'éthanol 100 % sont ajoutés, le mélange est homogénéisé puis incubé à -20°C pendant au moins 30 min. Le mélange est ensuite centrifugé 15 min à 4°C à 16 000 g afin de sédimenter les acides nucléiques. Le culot est lavé avec 1 ml d'éthanol 70 % afin d'éliminer les sels. Finalement, le culot est séché avant d'être solubilisé dans l'eau.

14. Test d'activité de PRORP *in vitro*

Les tests d'activité RNase P *in vitro* sont réalisés dans un tampon contenant 30 mM Tris-HCl pH 8, 40 mM de NaCl, 4,5 mM MgCl_2 , 20 $\mu\text{g/ml}$ BSA, 2,5 nM DTT et 20 U RNase OUT. Une solution contenant 2 μM de substrat représentant un précurseur d'ARNt est mise en présence de 170 nM d'enzyme à 25°C pendant 15 à 30 min. dans un volume réactionnel de 10 μL . La réaction est arrêtée par ajout de phénol / chloroforme 1:1). Après cette extraction, l'ARN est précipité à l'éthanol et re-suspendu dans un tampon de charge contenant 95 % (v/v) de formamide, 20 mM EDTA, 0,05 % (p/v) bleu de bromophénol, 0,05 % (p/v) xylène cyanol. Les ARN sont séparés sur gel dénaturant de polyacrylamide-urée 8 % et visualisés par coloration du gel dans une solution de bromure d'éthidium (2 $\mu\text{L}/100\text{ mL}$) et illumination aux UV à 260 nm grâce à un trans-illuminateur Vilber Lourmat Ebox VX2.

15. Ultracentrifugation analytique

L'UltraCentrifugation Analytique (AUC) est une méthode biophysique permettant de déterminer des paramètres d'affinité tels que le K_D de deux macromolécules grâce à la

comparaison des coefficients de sédimentation des deux macromolécules et du complexe formé par les deux entités. Ici l'AUC est utilisée pour mesurer l'affinité d'un mutant catalytique de PRORP pour des précurseurs d'ARNt.

Les expériences d'AUC ont été réalisées à la plateforme « Biologie et génomique structurale » (<http://www.igbmc.fr/technologies/5/team/53/>) à l'IGBMC à Strasbourg, avec Mme Catherine Birck. La centrifugeuse Beckman Coulter proteomeLab XL-I utilisée comporte un système de mesure d'absorbance en fonction de la distance à l'axe de rotation. Les échantillons ont été centrifugés à 220000g à 20°C jusqu'à sédimentation totale de toutes les espèces du mélange. Différents mélanges d'ARNt et de protéines ont été préparés dans 400 µl de tampon SEC2 contenant 5 mM de MgCl₂. La concentration en ARNt a été maintenue constante à 0,7 µM et la protéine a été ajoutée à des concentrations finales variant de 0,25 µM à 10 µM. Les données ont été analysées à l'aide du logiciel SEDFIT (www.analyticalultracentrifugation.com/) et d'un modèle de distribution de coefficients de sédimentation continue. Les distributions obtenues pour chaque échantillon ont été intégrées pour déterminer les coefficients de sédimentation moyen en fonction des concentrations de protéines et pour générer des isothermes « swfast ». Les isothermes ont été analysés avec le logiciel SEDPHAT pour obtenir une estimation du K_D. Le logiciel GUSSE a été utilisé pour tracer et intégrer les distributions des coefficients de sédimentation et pour générer les isothermes (Brautigam, 2015).

16. Thermophorèse à micro échelle

La thermophorèse à micro-échelle ou « microscale thermophoresis » (MST) est aussi utilisée pour mesurer l'affinité entre deux macromolécules. Les expériences de MST ont été réalisées à la plateforme « Biologie et génomique structurale » (<http://www.igbmc.fr/technologies/5/team/53/>) à l'IGBMC à Strasbourg, avec Mme Catherine Birck. Cette technique nécessite toutefois que l'une des deux entités testées soit fluorescente. Pour cela un pré-ARNt fusionné en 5' au fluorochrome cy5 est utilisé. Une solution de protéine à une concentration de 300 µM est diluée en cascade au demi puis supplémentée avec 50 nM ou 100 nM de pré-ARNt. Les expériences de MST sont réalisées avec un instrument Monolith NT LabelFree (NanoTemper Technologies GmbH, Munich, Allemagne). Trois types de capillaires ont été testés (Standard, Hydrophobic, Premium) dans du tampon SEC2 supplémenté ou non avec 0,1 % (v/v) Tween 20 et / ou 0,3 mg/ml de BSA (afin de stabiliser le signal du fluorochrome qui a tendance à augmenter avec la charge protéique),

avec une puissance de laser de 20 %, 40 % et 80 %. Les données sont analysées avec le logiciel Monolith MO.

17. Titration calorimétrique isotherme

La titration calorimétrique isotherme (ITC) permet de mesurer l'affinité et la stœchiométrie de deux molécules en interaction. L'expérience a été réalisée par F. Pinker en collaboration avec E. Ennifar à l'Institut de Biologie Moléculaire et Cellulaire de Strasbourg. La cellule de référence est remplie d'eau et la cellule échantillon est remplie avec 280 μ l d'une solution de pré-ARNt^{Cys} mitochondrial L5T0 à 22 μ M dans du tampon SEC2. La seringue est remplie avec une solution de AtPRORP2 DD à 187 μ M dans du tampon SEC2. 20 injections de 2 μ L sont réalisées en mesurant la variation de température entre les cellules.

18. Diffusion des rayons X aux petits angles

Les mesures de SAXS ont été réalisées sur la ligne de lumière SWING du synchrotron SOLEIL (Saint-Aubin, France). Les différentes espèces des échantillons ont été séparées par injection sur une colonne d'exclusion de taille Bio SEC-3 équilibrée avec le tampon SEC2. Pour déterminer la structure du complexe AtPRORP2 / préARNt, un mélange de 30 μ l de PRORP2DD catalytiquement inactif d'une concentration de 320 μ M et de 25 μ l de pré-ARNt^{Cys} mitochondrial L5T0 à 185 μ M a été injecté. Afin de prévenir la dissociation du complexe, le tampon SEC2 a été supplémenté avec 1,5 μ M de AtPRORP2DD. AtPRORP2WT et AtPRORP2DD ont également été caractérisées seules dans le tampon SEC2. Le traitement des données, l'interprétation et l'évaluation des R_g sur les profils d'élution ont été réalisés avec Foxtrot et l'analyse des données (détermination de R_g , d_{max}) avec l'ATSAS. En tenant compte d'une analyse en mode normal (NMA) réalisée avec ElNeMo, les structures cristallographiques de AtPRORP2 ont été perturbées pour sélectionner des modèles qui représentaient mieux le profil SAXS de PRORP2, en raison d'une qualité d'ajustement calculée avec CRY SOL. Des modèles complets de PRORP2 et PRORP2DD (y compris leurs extensions N- et C-terminales respectives non décrites dans la structure cristallographique) ont été générés sous des contraintes SAXS en utilisant DADIMODO, un programme d'affinement basé sur un algorithme génétique. Les modèles du complexe PRORP2/préARNt ont été construits avec SASREF, en utilisant les données SAXS comme contraintes, en particulier avec une charnière flexible entre les domaines de liaison au zinc et le domaine catalytique dans la structure en solution et en tenant compte des contraintes de distance entre les résidus mentionnés dans les motifs PPR2 et 3, le domaine catalytique et les

nucléotides dans l'ARNt. Les modèles SASREF ont été inspectés manuellement pour supprimer les collisions stériques et ajustés avec PyMOL à l'aide du plugin SAXS SASpy.

19. Agrotransformation d'*Arabidopsis thaliana* par immersion florale

Arabidopsis thaliana est transformé par la technique de « floral dip ». Une pré-culture d'*Agrobacterium tumefaciens* transformée avec un vecteur binaire comportant la construction à insérer dans la plante est incubée 48 h à 28°C, puis utilisée pour inoculer un plus grand volume de milieu (1/2 000ème de volume) incubée 24 h à 28°C. Les bactéries sont récoltées par centrifugation à 5000 g pendant 10 min à température ambiante puis remises en suspension dans du milieu d'Agroinfiltration composé à partir de MS ½ supplémenté de 5 % (m/v) de sucrose, 0,025 % (v/v) de Silvet L-77 et 200 µM d'acétytyingone. Le tout est incubé 1 h à température ambiante pour activer les gènes de virulence. Les inflorescences sont trempées 30 sec dans la solution puis égouttées à l'horizontale pendant 15 min avant d'être placées au noir dans des mini serres. Après 24 h, les plantes sont placées en condition normale de culture et sorties de mini serres après 3 jours.

20. Extraction d'ADN génomique d'*Arabidopsis thaliana*

Une jeune feuille de rosette d'*A. thaliana* est prélevée puis broyée dans 500 µL de tampon d'extraction (0,1 Tris-HCl pH 8, 50 mM EDTA, 0,5 M NaCl, 1 % (v/v) PVP40) à l'aide d'un micro-pilon. Ensuite, 66 µL de SDS 10 % (m/v) sont ajoutés et l'extraction est réalisée sous agitation à température ambiante pendant 15 min. L'extrait est centrifugé 10 min à 16 000 g. Le surnageant est prélevé. 166 µL d'acétate de potassium 5 M sont ajoutés. Après agitation et centrifugation 15 min à 16 000 g, le surnageant est récupéré. 0,7 volume d'isopropanol est ajouté. Après incubation sur la glace pendant 15 min, l'extrait est centrifugé 15 min à 16 000 g. Le culot d'ADN génomique obtenu est lavé avec 500 µL d'éthanol 70 % (v/v) puis solubilisé dans du tampon R40 composé de TE 1x et de 40 µg/ml de RNase A en incubant 2 h à température ambiante. L'ADN génomique est à nouveau précipité à l'éthanol et finalement solubilisé dans de l'eau miliQ.

21. Génotypage des plantes

Les génotypages des plantes sont réalisés par PCR après extraction de l'ADN génomique. Les PCR sont réalisées avec de la Pfu ADN polymérase selon les conditions

recommandées par le fournisseur avec une Tm de 58°C, un temps d'élongation de 20 sec et les amorces décrites ci-dessus.

22. Purification des mitochondries à partir de parties aériennes de plantes d'*Arabidopsis*.

L'équivalent de 30 g de parties aériennes de jeunes plantes d'*Arabidopsis thaliana* (fleurs, bourgeons, jeunes feuilles et parties fines de la tige) sont broyés dans 150 ml de tampon d'extraction dans un homogénéisateur pendant 1 fois 5 sec à vitesse maximum et 2 fois 5 sec à vitesse minimum avec 30 sec entre chaque broyage. Le broyat est filtré sur deux couches de Miracloth. Le filtrat est centrifugé à 1500 g pendant 5 min à 4°C afin d'éliminer les débris cellulaires et les plastes. Le surnageant est centrifugé à 20000 g pendant 10 min à 4°C. Le culot enrichi en mitochondrie est remis en suspension à l'aide d'un pinceau dans 5 ml de tampon de lavage. Afin d'éliminer la présence d'agrégats la suspension est homogénéisée 3 fois avec un Potter puis diluée 6 fois. On répète ensuite les deux centrifugations différentielles à 2000 g 5 min puis à 20000 g pendant 10 min. Le culot final est repris au pinceau dans 500 μ L de tampon de lavage puis homogénéisé 3 fois au Potter avant d'être déposé sur un gradient de Percoll composé d'une couche de 5 ml de Percoll 50 % (v/v), une couche de 25 ml à 25 % et une couche de 5 ml à 18 %. Le gradient est centrifugé à 40000 g 45 min à 4°C sans frein. Le haut du gradient est éliminé et les mitochondries purifiées présentes à l'interphase entre les couches à 25 et 50 % de Percoll sont récupérées, diluées 10 fois et sédimentées à 20000 g pendant 10 min à 4°C. Le surnageant est retiré en laissant 5 ml, dilué 6 fois et centrifugé à 10000 g, 10 min à 4°C. Ce lavage est réalisé une seconde fois. Le culot est alors repris dans un volume total de 2 ml et un aliquot utilisé pour doser la quantité de protéines mitochondriales par la méthode de Bradford.

	Tampon d'extraction	Tampon de lavage	Tampon gradient Percoll 5X
Sucrose	0,3 M	0,3 M	1,5 M
tetrasodiumpyrophosphate	15 mM	/	/
MOPS	/	1 mM	50 mM
EDTA	2 mM	1 mM	/
KH ₂ PO ₄	10 mM	/	/
PVP-40	1 % (m/v)	/	/
Ajuster au pH	7,5	7,2	7,2
Avant utilisation ajouter			
BSA	1 %	/	/
Sodium L-ascorbate	20 mM	/	/
Cysteine hydrochloride monohydrate	5 mM	/	/

23. Co-immunoprécipitation de protéines avec des ARN

Les co-immunoprécipitations sont effectuées par chromatographie d'affinité pour des étiquettes (i.e. ici en utilisant des anticorps anti-HA) en utilisant le système μ MACS de Miltenyi Biotec (<http://www.miltenyibiotec.com/en/>).

Un culot correspondant à 1 mg de protéines mitochondriales est repris dans 1 mL de tampon IP (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 % (v/v) Triton X-100, Roche proteases inhibitors complete EDTA free) contenant un détergent non-ionique. La solubilisation des mitochondries est réalisée pendant 30 min en chambre froide sur un agitateur rotatif et par pipetage au travers d'une pointe de cône de pipette automatique P10 après 15 min et 30 min. 50 μ L de billes magnétiques μ MACS anti HA sont ajoutés par mg de protéines mitochondriales puis incubés 30 min à 4°C avec une rotation de 5 rpm. Les colonnes μ MACS sont conditionnées avec 200 μ L de tampon IP, placées sur leur support magnétique puis chargées avec le lysat et les billes à 4°C. Les billes sont lavées 4 fois avec 1 mL de Tampon IP. L'éluion est réalisée en incubant 100 μ L de TRIzol Reagent chauffé à 85°C sur les colonnes pendant 5 min puis en rajoutant 900 μ L de TRIzol Reagent. Le volume mort est

récupéré en soufflant avec une poire dans la colonne. Une seconde élution est réalisée hors du système magnétique avec 1 mL de TRIzol Reagent à 85°C. Les ARN et les protéines sont séparés suivant les recommandations du fournisseur. Les ARN sont précipités avec 0,5 volume d'isopropanol et lavés avec de l'éthanol 70 % (v/v). Les protéines sont précipitées avec 3 volumes d'acétone et lavées 2 fois à l'acétone 80 % (v/v).

24. Analyse d'ARN par hybridation de type northern

Les ARN sont séparés sur un gel dénaturant de polyacrylamide contenant 7 M d'urée. Après la migration, les ARN sont transférés sur une membrane Hybond N+ (Amersham) pendant 45 min à 500 mA, 100 V dans du TAE 0,25 x puis pontés de manière covalente par un traitement aux UV à 120 000 $\mu\text{J}/\text{cm}^2$. Une sonde radioactive (i.e. ici un oligonucléotide représentant un ARNt radiomarqué en 5') est hybridée pendant 16 h à 45°C dans une solution de 3 ml de 6 x SSC 0,5 % (m/v) SDS. Des lavages de stringence croissante sont réalisés à 45°C, 2 lavages de 10 min avec du SSC 2 x puis un lavage de 30 min avec du SSC 2 x, 0,1 % (m/v) SDS. Les signaux sont révélés par autoradiographie, i.e. avec un Phosphorimager (Fujifilm).

25. Analyse des ARN par séquençage à haut débit

a. Analyse qualitative des ARN

Les ARN sont quantifiés avec le spectromètre Qubit (Invitrogen), et le kit « RNA High Sensibility » puis dilués dans la gamme de détection d'une puce « RNA 6000 Pico » pour Bioanalyzer (Agilent) afin de vérifier la qualité des ARN.

Le QuBit utilise un fluorophore dont l'émission n'a lieu qu'en présence d'acides nucléiques. Ce principe permet au Qubit d'être un outil de mesure à la fois sensible (10 pg/ μL) et précis.

Le Bioanalyzer utilise des puces comportant des microcanaux permettant une séparation des ARN suivant leur taille. La puce RNA 6000 pico permet d'analyser jusqu'à 11 échantillons concentrés entre 50 et 5000 pg/ μl . La qualité des ARN est définie par le « RIN » (RNA Integrity Number) qui ne peut pas être pleinement considéré ici, car l'algorithme n'est pas conçu pour de l'ARN issu uniquement de mitochondries de plantes. Le calcul du RIN est basé sur différentes régions et pics du chromatogramme notamment les pics du marqueur ajouté à l'ARN analysé et le pic des ARN ribosomiques 18S et 28S alors que les

mitoribosomes de plantes comportent des ARNr 26S à la place du 28S. (Mueller and Schroeder, 2004).

b. Construction des banques d'ADNc

Les banques d'ADNc sont réalisées à partir de 200 ng d'ARN issus des immunoprécipitations (IP) ou 500 ng d'ARN mitochondrial total (Total) avec le kit « ScriptSeq complete kit (plant leaf) » d'Illumina en suivant les recommandations du constructeur. Après retrait des ARNr grâce à des sondes spécifiques de chaque ARNr, les ARN sont précipités à l'éthanol. Le retrait des ARNr est vérifié au Bioanalyzer avec une puce RNA 6000 pico. Les différentes banques d'ADNc sont construites par reverse transcription et par PCR avec des « index » différents c'est-à-dire des petites séquences spécifiques pour chaque banque :

- IP1 index 6 : GCCAAT
- IP2 index 12 : CTTGTA
- IP3 index 1 : ATCACG
- Total 1 index 4 : TGACCA
- Total 2 index 8 : ACTTGA
- Total 3 index 10 : TAGCTT

La distribution de la taille des banques est vérifiée au Bioanalyzer puis quantifiée au Qubit avec le kit « Quanti-iT HS DNA assays ». Les banques sont ensuite normalisées à 4 nM puis dénaturées avec 1 nM de NaOH avant d'être diluées à 10 pM pour le séquençage.

c. Séquençage par la technologie Illumina

Les banques sont séquencées en présence de 5 % de « PhiX control », une banque contrôle réalisée à partir du génome du phage PhiX. Un séquenceur Illumina MiSeq est utilisé avec une puce d'un kit de réactifs « MiSeq Reagent Kit v3 » (150 cycles) afin de séquencer en « pair end » de 2 x 75 nucléotides, c'est-à-dire en générant des séquences de 75 nucléotides à partir des deux extrémités des clones d'ADNc. Grâce aux adaptateurs ajoutés à l'ADNc, ceux-ci se fixent aléatoirement sur la cellule qui porte les séquences complémentaires aux adaptateurs. Une PCR permet l'amplification de ces brins autour du brin initial créant un « cluster » d'ADNc simple brin. Lors du séquençage, chaque nucléotide est couplé à un fluorochrome différent, ce qui permet à chaque cycle de séquençage de lire le nucléotide intégré par excitation du fluorochrome avec un laser. La fluorescence du cluster est donc

représentative du nucléotide incorporé. Le fluorochrome est alors retiré afin de permettre l'ajout du nucléotide suivant.

26. Analyse des données de séquençage à haut débit

a. Contrôle de la qualité du séquençage

La qualité des séquences est déterminée par le logiciel FastQC utilisant les données brutes de séquençage afin de déterminer la qualité de lecture des bases à chaque position pour toutes les séquences présentes dans une banque et cela sur les deux brins générés par le séquençage pair end.

b. Alignement des séquences sur le génome de référence

Les séquences obtenues sont alignées sur le génome de référence avec le logiciel HiSAT2 (V2.0.5) (Kim et al., 2015). Deux alignements sont réalisés, l'un avec seulement le génome mitochondrial d'Arabidopsis et l'autre avec les trois génomes nucléaire, mitochondrial et chloroplastique d'Arabidopsis tous disponibles dans TAIR10. Ces alignements permettent de vérifier la proportion d'ARN mitochondrial par rapport à l'ARN total mais également de révéler la présence d'ARN contaminants provenant d'autres espèces dans les banques de ADNc. Les fichiers alignés sont utilisés pour réaliser l'analyse comparative des ARN provenant des immunoprécipitations par rapport aux ARN totaux du transcriptome mitochondrial.

c. Analyse comparative des banques d'ADNc

L'analyse comparative des banques alignées est réalisée avec le logiciel ASPeaks (V2.0.1) (Kucukural et al., 2013). Ce logiciel permet une analyse différentielle de plusieurs immunoprécipitations avec plusieurs ARN totaux et d'identifier des pics correspondant à des ARN surreprésentés dans les immunoprécipitations.

V. Résultats chapitre 1 : Etude du rôle des motifs PPR de AtPRORP2 pour l'interaction avec les pré-ARNt

Les protéines PRORP sont caractérisées par la présence d'un domaine NYN qui porte l'activité catalytique de l'enzyme ainsi que des domaines PPR d'interaction avec l'ARN dont le rôle prédit serait de conférer aux protéines PRORP leur spécificité de substrats ARN. Jusqu'à ce jour, aucune protéine PPR interagissant avec des ARN structurés tels que les ARNt n'avait été étudiée. Les études mécanistiques des protéines PPR, en particulier la détermination du « code PPR » a été réalisée uniquement avec des protéines PPR dont les cibles sont des ARNm simples brins non structurés.

L'équipe d'accueil a démontré, par des tests d'activité RNase P *in vitro* utilisant PRORP et une collection de pré-ARNt mutants ainsi que par une approche de type « footprinting », que PRORP interagit avec des résidus conservés sur les boucles T et D des ARNt, i.e. dans le coude formé par l'empilement de ces deux boucles dans la structure 3D des ARNt (Gobert, 2013).

Le but de l'étude présente a été de démontrer l'importance des motifs PPR dans l'interaction avec les ARNt et de comprendre le mécanisme conférant la spécificité de reconnaissance des pré-ARNt par PRORP. Cette étude devait aussi montrer dans quelle mesure le code PPR défini pour les ARN linéaires pouvait s'appliquer ou non à des ARN structurés comme les ARNt. Pour cela, les cinq motifs PPR de AtPRORP2, une des deux protéines PRORP nucléaires d'*Arabidopsis*, ont été individuellement mutés aux positions 5 et 35 afin d'abolir leurs interactions avec les nucléotides de leur cible ARN. L'interaction entre ces différents mutants PPR et le pré-ARNt a été mesurée indirectement par des tests d'activité RNase P *in vitro* mais également directement en déterminant les K_D correspondant aux interactions de chaque mutant PPR avec les pré-ARNt. Le résultat de ces analyses combiné avec des analyses structurales en solution par « Small angle X-ray scattering » (SAXS) du complexe PRORP / pré-ARNt a permis de construire un modèle 3D expliquant le mode d'interaction des motifs PPR de PRORP avec les nucléotides des ARNt.

1. Identification des motifs PPR et des positions susceptibles de participer aux interactions avec AtPRORP2

Les positions des motifs PPR de AtPRORP2 initialement prédites en utilisant le logiciel TPRpred (Karpenhalli et al. 2007) (<https://toolkit.tuebingen.mpg.de/#/tools/tprpred>). La structure cristallographique (Pinker F. et al, 2015) a permis de confirmer la position des

différentes hélices alpha caractéristiques des motifs PPR ainsi que la taille des boucles qui les séparent (Figure 17). Les motifs PPR sont des domaines d'interaction spécifique avec l'ARN supposés être responsables de la reconnaissance des substrats par l'enzyme. Cette spécificité est médiée par différents résidus, en particulier deux acides aminés aux positions 5 et 35 des motifs PPR. La conservation des résidus aux positions 5 et 35 de AtPRORP2 a été déterminée par l'alignement de 59 séquences de protéines PRORP du groupe III. En effet, AtPRORP2 est au même titre que AtPRORP3, une protéine PRORP du groupe III comme décrit par Lechner et al. (Lechner M., 2015). Bien que la séquence d'un motif PPR soit dégénérée et que seule sa structure en hélice-boucle-hélice soit bien conservée, les acides aminés permettant l'interaction spécifique avec l'ARN ont dû co-évoluer avec les bases nécessaires à la reconnaissance des ARNt. Ainsi, comme tous les ARNt dans les différentes espèces doivent être maturés de la même manière par les RNase P, les bases nécessaires à la reconnaissance doivent être également conservées. Les positions 5 et 35 des motifs PPR importantes pour la reconnaissance des ARNt devraient donc être conservées dans l'évolution. Cette analyse a donc permis d'identifier les positions 5 et 35 des 5 motifs PPR dans la séquence AtPRORP2. Il s'agit pour le motif PPR1 des positions L34 et S65, pour le motif PPR2 de Q70 et N108, pour le motif PPR3 de T113 et R145, pour le motif PPR4 de A150 et E180 et pour le motif PPR5 de S185 et S215 (Figure 18). L'alignement des protéines PRORP de 14 monocotylédones et de 45 dicotylédones a permis de déterminer la probabilité avec laquelle chaque acide aminé est trouvé aux différentes positions dans les protéines PRORP du cluster III (Figure 19).

On observe en l'occurrence qu'à la position 35 des motifs PPR1 et PPR2 un N peut être substitué par un S ou un T surtout chez les monocotylédones. Une analyse plus poussée permettrait de montrer si cette variabilité reflète que (i) certaines clades utilisent spécifiquement certaines combinaisons d'acides aminés et ont donc des codes PPR spécifiques ou que (ii) les différents acides aminés (ici N, S ou T) sont fonctionnellement équivalents et ont été utilisés indifféremment durant l'évolution. De même, l'analyse a montré que des résidus R et K sont conservés à la position 35 du motif PPR3. La position 5 du motif PPR4 semble peu conservée contrairement à sa position 35. Le contraire est observé pour le motif PPR5 avec une conservation de la position 5 contrairement à la position 35. Ainsi, les résultats suggèrent que le code PPR, sous sa forme actuelle, pourrait être incomplet et que certains acides aminés pourraient être fonctionnellement équivalents comme N, S et T ou R et K. Alors que certains résidus, i.e. N, S et T sont des acides aminés déjà décrits dans le code PPR, d'autres acides aminés comme R et K n'ont encore jamais été utilisés dans les

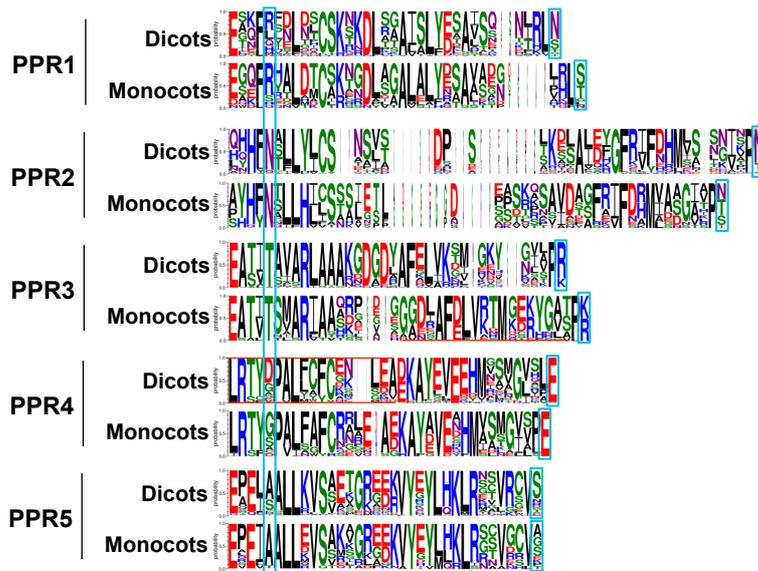


Figure 19 : Fréquence des acides aminés à toutes les positions des motifs PPR de protéines PRORP du groupe III dans le règne végétal.

Représentation de la fréquence des acides aminés à toutes les positions des motifs PPR obtenue avec le logiciel weblogo à partir d'un alignement de 14 séquences PRORP de plantes monocotylédones (Monocots) et 45 séquences PRORP de plantes dicotylédones (Dicots) avec jusqu'à deux séquences par ordre. Les cadres bleus représentent les positions 5 et 35 des motifs PPR respectifs.

différentes versions publiées du code PPR. La faible conservation d'une des positions dans les motifs PPR 4 et 5 laisse penser que ces motifs pourraient (i) ne pas jouer un rôle dans l'interaction avec l'ARN, ou (ii) utiliser des positions différentes des positions 5 et 35. De même, le motif PPR1 possède une position 35 conservée entre les acides aminés N, S et T alors que la position 5 est plus variable avec des acides aminés R, L, Q et H qui n'ont jamais été référencés dans le code PPR. Ainsi, l'implication du motif PPR1 dans l'interaction avec l'ARN est également hypothétique.

2. Mutagénèses dirigées des positions 5 et 35 de chaque motif PPR de AtPRORP2

Afin d'étudier l'implication des différents motifs PPR de AtPRORP2 pour la reconnaissance des ARNt, toutes les positions 5 et 35 de tous les motifs PPR ont été mutées ; ces mutations ont été choisies selon le code PPR afin de changer la spécificité de chaque motif PPR ou d'introduire une spécificité pour les motifs dont le code PPR ne prévoit pas de reconnaissance spécifique. Ces mutagénèses ont été appliquées sur la séquence codante de AtPRORP2 active c'est-à-dire une version sauvage de la protéine, ainsi que sur un mutant catalytiquement inactif de cette même protéine par mutation des aspartates catalytiques 421 et 422 en alanine. Ainsi les positions L34 et S65 du motif PPR1 ont été modifiées en T et D induisant en théorie la reconnaissance d'une guanine par le motif PPR1. Pour le motif 2, les mutations Q70T et N108D doivent théoriquement changer la spécificité de PPR2 d'une cytosine pour une guanine. Pour le motif 3, les mutations T113N et R145N doivent théoriquement changer la spécificité de PPR3 d'une purine pour une cytosine. Pour le motif 4, les mutations A150N et E180N doivent en théorie induire la reconnaissance d'une cytosine par PPR4. Finalement, pour le motif 5, les mutations S185T et S215 D doivent en théorie induire la reconnaissance d'une guanine par PPR5. Deux méthodes de mutagenèse décrites en IV.2 ont été utilisées en parallèle. La première étant la méthode recommandée par le fournisseur du système QuickChange Site-Directed Mutagenesis de Agilent et la seconde décrite par Edelheit, O et al, 2009. Des clones mutants pour les 10 positions dans les motifs PPR ont pu être obtenus et analysés par séquençage. Ensuite des doubles mutants pour les positions 5 et 35 des motifs PPR respectifs ont été construits par la même méthode. Les plasmides portant les deux versions de la protéine sauvage (catalytiquement active et inactive) ainsi que les 10 doubles mutants 5 et 35 des différents motifs PPR dans les protéines catalytiquement active et inactive ont été transformés dans des cellules d'*E. coli* BL21 afin d'exprimer, de purifier et de caractériser les différentes versions d'AtPRORP2.

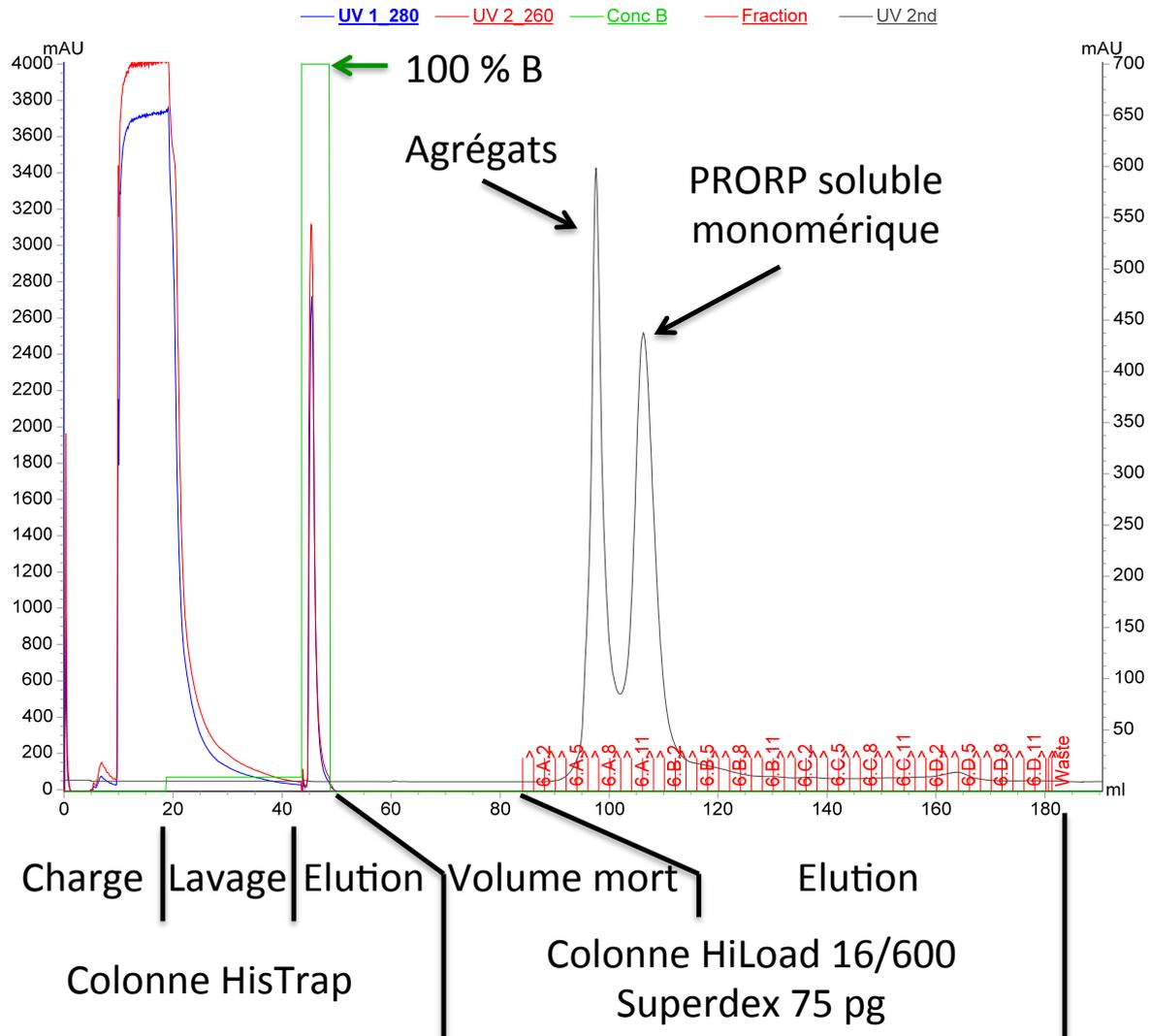


Figure 20 : Protocole de purification optimisé de AtPRORP2.

La purification optimisée ici est basée sur l'utilisation d'une valve versatile dans le système FPLC. L'intégralité du pic d'éluion de la colonne d'affinité est alors chargée directement sur la colonne d'exclusion de taille limitant ainsi les pertes. L'axe Y à gauche correspond à l'absorbance pour UV1 et UV2 à 280 nm et 260 nm mesurée à la sortie de la colonne d'affinité. L'axe Y de droite correspond à l'absorbance UV 2nd 280 nm mesurée à la sortie de la colonne d'exclusion de taille. « Conc B » représente la concentration du tampon B utilisé lors de l'étape de chromatographie d'affinité.

Tableau 2 : Résumé des différentes mutations sur AtPRORP2

Les différentes mutations des positions 5 et 35 de chaque motif PPR ont été réalisées sur la version active de la protéine et sur la version inactive.

Motif PPR	AtPRORP2 active	AtPRORP2 inactive
WT	/	D421A, D422A
PPR1	L34D, S65D	L34D, S65D, D421A, D422A
PPR2	Q70T, N108D	Q70T, N108D, D421A, D422A
PPR3	T113D, R145N	T113D, R145N, D421A, D422A
PPR4	A150N, E180N	A150N, E180N, D421A, D422A
PPR5	S185T, S215D	S185T, S215D, D421A, D422A

3. Optimisation du protocole de purification de AtPRORP2 et de ses mutants

Le protocole initial de purification décrit par Gobert et al, 2010 a été amélioré pour obtenir des rendements accrus, nécessaires aux analyses biophysiques réalisées ici. Le protocole optimisé repose sur le couplage direct d'une étape de chromatographie d'affinité et d'une étape de chromatographie par filtration sur gel. Le protocole optimisé permet d'obtenir 2,5 à 3 mg de protéine par litre de culture alors qu'avec le protocole initial on obtenait un maximum de 2 mg de protéine par litre de culture. L'extrait protéique produit dans les cellules d'*E. coli* est injecté sur une colonne HisTrap FF Crude de 1 mL dans un tampon contenant 1 M NaCl afin de détacher les ARN associés et 5 mM imidazole afin de prévenir les interactions aspécifiques entre la résine et d'autres protéines du lysat bactérien. La colonne est alors lavée avec un tampon contenant 50 mM d'imidazole jusqu'à ce que la DO_{280nm} soit stable durant 2 min ou en lavant avec un volume maximum de 30 volumes de colonne. L'élution est alors réalisée avec 5 volumes de colonne d'un tampon contenant 500 mM d'imidazole. La chromatographie est suivie en visualisant l'absorbance sur le logiciel Unicorn 7. La valeur d'absorbance à 280 nm est remise à zéro lors du début de la phase d'élution. Lorsque la valeur d'absorbance à 280 nm atteint 50 mAU, la valve versatile change de position et permet l'envoi du pic d'élution (dans un volume maximal de 5 ml) sur une colonne de filtration sur gel HiLoad 16/600 Superdex 75 pg (Figure 20). Cette injection directe permet de limiter les pertes en gardant la protéine au sein du système et en évitant d'avoir à la concentrer. Cette étape de filtration sur gel permet également de retirer les agrégats, les protéines dégradées, tronquées et les quelques contaminants restants ainsi que de changer de

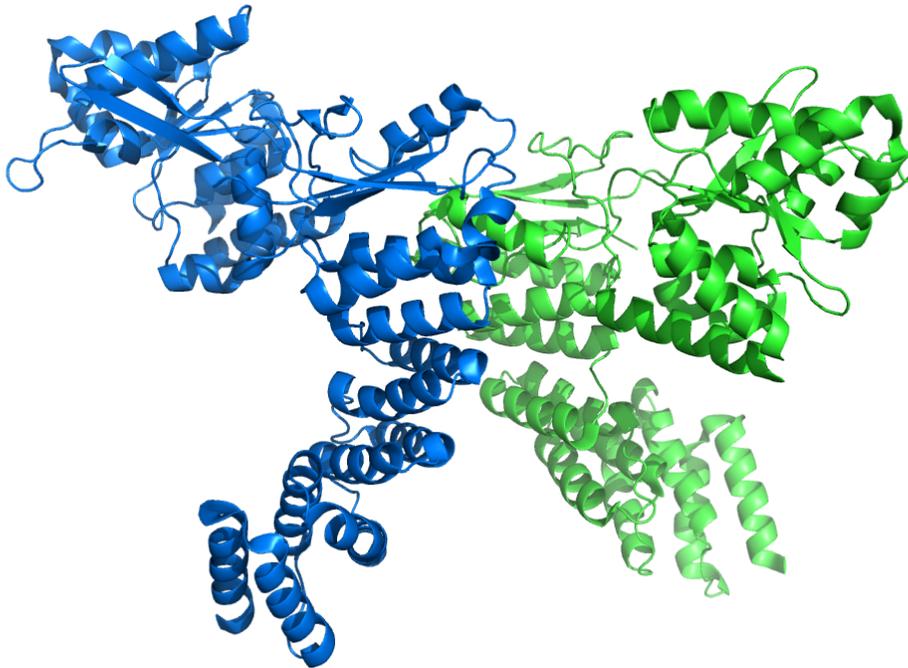


Figure 21 : Dimère de AtPRORP2 dans une maille du cristal
AtPRORP2 forme des dimères à forte concentration.

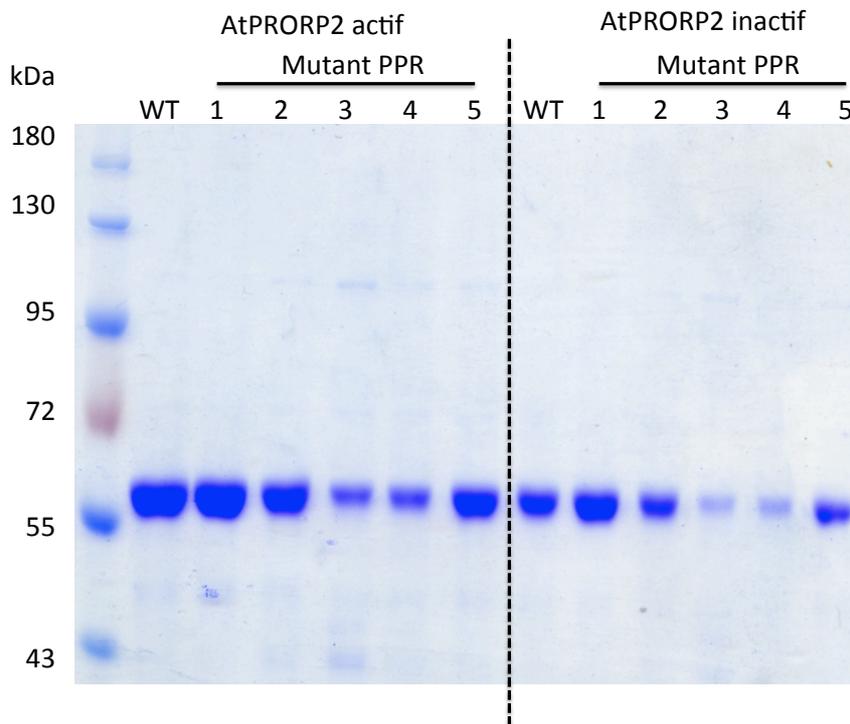


Figure 22 : Purification optimisée de AtPRORP2 et de ses mutants.

AtPRORP2 sauvage (WT) et ses différents mutants des motifs PPR 1 à 5 sont purifiés et analysés par SDS PAGE. Les 12 protéines d'intérêt produites ici sont majoritaires et les contaminants semblent identiques entre les différents mutants. La mutation des motifs PPR 3 et 4 semble engendrer une baisse de la stabilité des protéines qui entraîne une baisse de leurs rendements de purification.

tampon pour ne plus être en présence d'imidazole, ce qui permet l'utilisation directe de la protéine.

L'expression des protéines est réalisée durant 16 h à 17°C sous agitation en induisant l'expression avec 0,5 mM d'IPTG. Une expression trop rapide à 37°C entraîne une baisse de la solubilité de la protéine et une baisse de rendement durant la clarification du lysat lors de la purification. Toutes les versions des protéines AtPRORP2 sont purifiées avec une faible quantité de contaminants et un ensemble de contaminations qui semble homogène entre les différentes protéines. Toutefois, les mutants PPR3 et PPR4 font exception. Les mutants PPR4 sont moins stables et précipitent en absence d'ARN. En effet un lavage à 1 M de NaCl lors de la purification, un traitement RNase ou une chromatographie échangeuse d'ion qui retire l'ARN entraînent tous la précipitation des protéines. La protéine mutée dans le motif PPR4 et catalytiquement active a donc été purifiée en omettant le lavage à 1 M de NaCl. La même protéine catalytiquement inactive n'a pas pu être utilisée pour les analyses biophysiques en raison de la présence de trop grandes quantités de contaminants ARN qui fausse les mesures d'affinité avec les substrats de la protéine. Par ailleurs, les mutations dans le motif PPR3 entraînent également une légère baisse de la stabilité des protéines. En effet les mutants PPR3 précipitent à une concentration inférieure à celle des autres mutants. Afin de minimiser les pertes dues à la précipitation, la concentration des protéines sur des unités de filtration est effectuée en alternant des cycles courts de centrifugation et d'homogénéisation de la solution de protéines afin d'éviter une concentration locale trop importante qui pourrait entraîner la précipitation des protéines. Finalement, il apparaît que les protéines AtPRORP2 ont tendance à former des dimères à haute concentration, e.g. une maille du cristal de AtPRORP2 contient deux protéines tête-bêche (Pinker 2015) (Figure 21). Les analyses biophysique révèlent parfois la formation de dimères dès 3 mg/ml.

Les rendements finaux des différentes protéines sont de 2,5 à 3 mg par litre de culture bactérienne pour la protéine WT et les mutants PPR1, 2 et 5 et d'environ 2 mg par litre de culture pour les mutants PPR3 et 4 (Figure 22).

4. Etude de l'interaction entre AtPRORP2 et les pré-ARNt

Deux approches complémentaires sont utilisées afin d'identifier les motifs PPR nécessaires à la reconnaissance des pré-ARNt par AtPRORP2. La première approche est basée sur la mesure de l'activité RNase P des mutants. C'est une méthode indirecte, en effet si le motif PPR muté est impliqué dans la reconnaissance des pré-ARNt, l'activité catalytique de la protéine devrait être réduite. La seconde approche est directe en mesurant l'affinité des

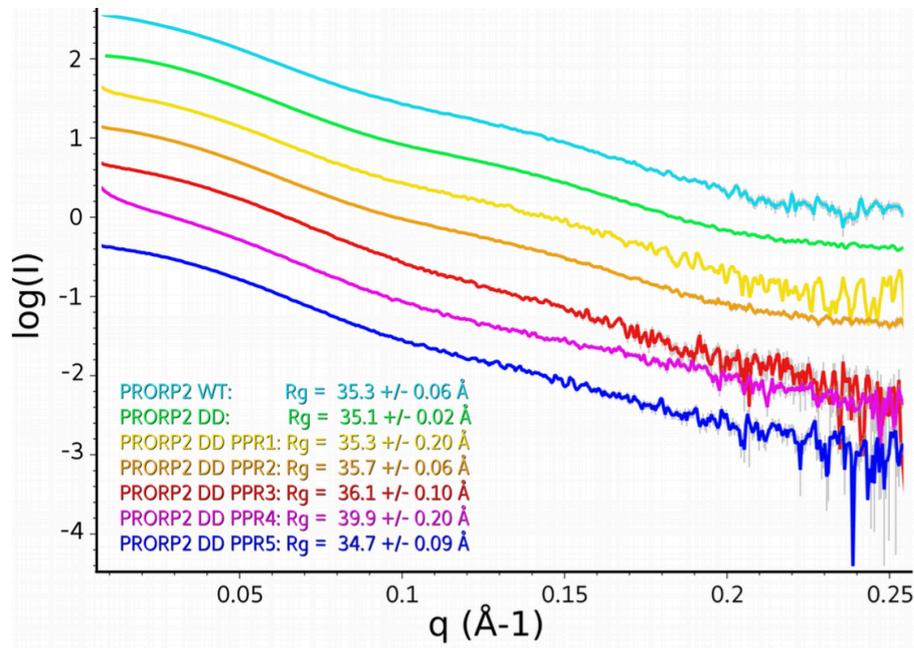


Figure 23 : Comparaison des courbes SAXS des mutants PPR.

Les cinq mutants PPR ont été analysés par SAXS en utilisant le même protocole que pour la protéine WT et son mutant catalytique. Les différentes courbes SAXS ont été séparées pour plus de lisibilité. La similarité des courbes et donc des structures dans l'intervalle de q de 0 à $0,15 \text{ \AA}^{-1}$ confirme l'intégralité structurale des mutants PPR et catalytiques. Les valeurs de R_g dérivées des plots de Guinier sont également très proches. Seul le mutant PPR4 semble se comporter différemment et à tendance à s'agréger comme l'indique la pente élevée de la courbe à petits angles ($q < 0,015 \text{ \AA}^{-1}$) et une augmentation du R_g .

différents mutants de AtPRORP2 pour les ARNt. Afin de vérifier que la perte d'affinité ou d'activité est bien due à une perte de la reconnaissance des substrats ARN causée par les mutations et non à un changement structural causé par les mutations, les enveloppes structurales en solution des différents mutants ont été analysées par diffusion des rayons X aux petits angles (SAXS).

a. Vérification de la conformation des mutants des motifs PPR

Les motifs PPR ayant un repliement en hélice boucle hélice bien définie, une mutation dans leurs séquences pourrait déstabiliser une des hélices et désordonner la protéine. Afin de vérifier la bonne conformation des mutants PPR, leur enveloppe structurale est déterminée par SAXS et comparée avec celle de la protéine sauvage. Les expériences de SAXS sont réalisées au synchrotron SOLEIL sur la ligne SWING (<https://www.synchrotron-soleil.fr/fr/lignes-de-lumiere/swing>). Des solutions de 50 μ L de protéines à une concentration de 10 mg/ml sont injectées sur une colonne de filtration sur gel Bio-SEC3 (Agilent) placée en amont du détecteur SAXS afin de séparer les protéines sous forme monomérique des agrégats. Les données acquises tout au long de l'élution sont traitées avec le logiciel Foxtrot pour obtenir la courbe SAXS. La courbe est ensuite traitée avec le logiciel Primus (<https://www.embl-hamburg.de/biosaxs/primus.html>). Les fenêtres d'acquisitions utilisées sont celles qui correspondent à un rayon de giration (R_g) stable et un fort signal I_0 commun pour tous les mutants afin d'analyser toutes les protéines dans la même conformation. Pour chaque mutant, la courbe SAXS mais aussi le R_g et la D_{max} (distance maximale entre deux atomes dans la molécule en solution) sont calculés et comparés. Les R_g de toutes les protéines sont de 35 ou 36 Å sauf pour le mutant PPR4 où le R_g est de 40 Å. Le mutant PPR4 qui semble moins stable semble également légèrement plus grand et présente une conformation différente. Cette différence est vraisemblablement causée par les contaminants ARN qui co-purifient avec le mutant PPR4.

Ainsi, les résultats suggèrent que les doubles mutations des motifs PPR ne semblent pas affecter la structure globale des protéines (Figure 23). Les protéines mutantes produites ici peuvent donc être utilisées pour comparer leurs activités catalytiques et leurs affinités pour l'ARN avec les protéines sauvages.

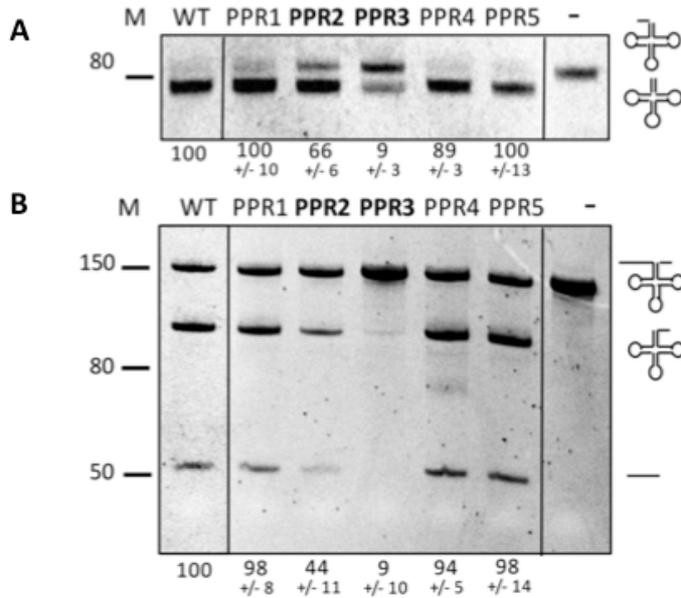


Figure 24 : Activités RNase P *in vitro* des mutants PPR de PRORP2.

Les tests d'activité *in vitro* ont été réalisés avec AtPRORP2 sauvage (WT) et les différents mutants PPR (PPR1 à PPR5) avec des pré-ARNt^{Cys} possédant un 5' Leader de 5 nucléotides (A) ou de 51 nucléotides (B), - indique la réaction contrôle négative avec le pré-ARNt seul. Les ARN ont été séparés sur des gels dénaturants PAGE 17% et les signaux quantifiés avec ImageJ. Les nombres indiquent le pourcentage de clivage, estimé par rapport au clivage de la protéine WT arbitrairement fixé à 100 % dans trois expériences indépendantes. Le marqueur de poids moléculaire (M) est donné en ribonucléotides.

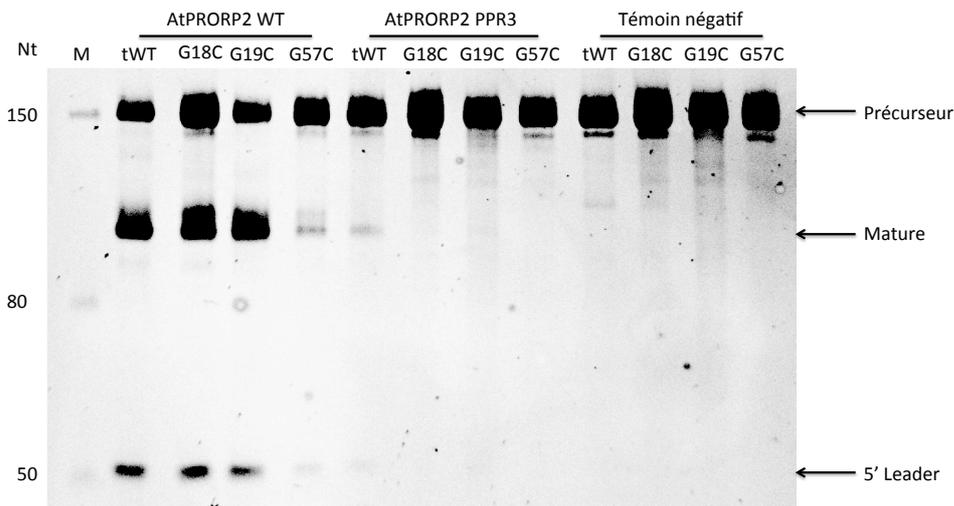


Figure 25 : Tentative de restauration de l'activité RNase P avec des mutants PPR et des pré-ARNt mutés.

Tests d'activité RNase P réalisés avec AtPRORP2 sauvage (WT), le mutant PPR3 et sans enzyme (Témoin négatif) avec l'ARNt^{Cys} L51T30 sauvage (tWT) ou avec les G en position 18, 19 ou 57 mutés en C. M représente le marqueur en nombre de nucléotides (Nt).

b. Tests d'activité de AtPRORP2 et des mutants de motifs PPR catalytiquement actifs

Cette approche permet d'évaluer en quoi les mutations dans les motifs PPR ont affecté l'activité RNase P des protéines. Ce qui permet d'évaluer indirectement dans quelle mesure les mutations dans les motifs PPR ont affecté la capacité de liaison des protéines aux précurseurs d'ARNt. En effet, en présence d'AtPRORP2, un pré-ARNt sera mûri par clivage endonucléolique de la séquence 5' leader. Une baisse de l'affinité pour l'ARNt pourrait engendrer une baisse de l'activité RNase P. Les tests ont été réalisés avec les ARNt Cys mitochondriaux L5T0 et L50T30 (c'est-à-dire avec des séquences 5' leader (L) de 5 ou 50 nucléotides et des séquences 3' trailer (T) de 0 ou 30 nucléotides). Les résultats ont montré que les activités RNase P des mutants PPR1, 4 et 5 sont similaires à celles de la protéine sauvage. Par contre la mutation du motif PPR2 entraîne une baisse d'activité de 56 % avec le transcrit L50T30 et de 34 % avec le transcrit L5T0. Le mutant PPR3 est encore plus affecté, son activité décroissant d'environ 90 % avec les deux pré-ARNt utilisés (Figure 24). Ces résultats ont montré que les motifs PPR3 et dans une moindre mesure PPR2 sont les motifs les plus importants pour l'activité catalytique de AtPRORP2 et suggèrent donc qu'ils soient les plus importants pour la reconnaissance des substrats ARNt de l'enzyme.

Les mutations apportées aux motifs PPR doivent, en théorie, changer leur spécificité en entraînant la reconnaissance d'autres ribonucléotides. En particulier, le motif PPR3 pour lequel le code PPR prédit la reconnaissance d'une purine a été muté pour reconnaître une cytidine. Ainsi des expériences ont été réalisées pour tenter de restaurer l'activité RNase P du mutant PPR3 en utilisant des pré-ARNt portant des mutations de purines en cytidines. Il s'agissait en particulier des mutations G18C, G19C et G57C du pré-ARNt^{cys} mitochondrial L50T30. Ces positions conservées dans les boucles du D et du T des ARNt sont impliquées dans l'interaction avec PRORP (Gobert Pinker 2013) et donc potentiellement reconnues par le motif PPR3. Toutefois, ces expériences n'ont à ce jour pas permis de restaurer l'activité RNase P du mutant PPR3. En effet, l'efficacité de clivage des pré-ARNt mutés qui est affectée, surtout pour la position G57, n'est pas restaurée avec le mutant PPR3 (Figure 25). Les nucléotides mutés étant impliqués dans le repliement 3D du pré-ARNt, il est possible que leur mutation empêche le repliement du pré-ARNt et donc sa reconnaissance par PRORP. D'autres expériences utilisant des ARNt portant une mutation mais aussi une mutation compensatoire maintenant la structure de l'ARNt pourraient ainsi être envisagées. Il est aussi possible que les deux mutations du motif PPR3 aux positions 5 et 35 bien qu'étant suffisantes

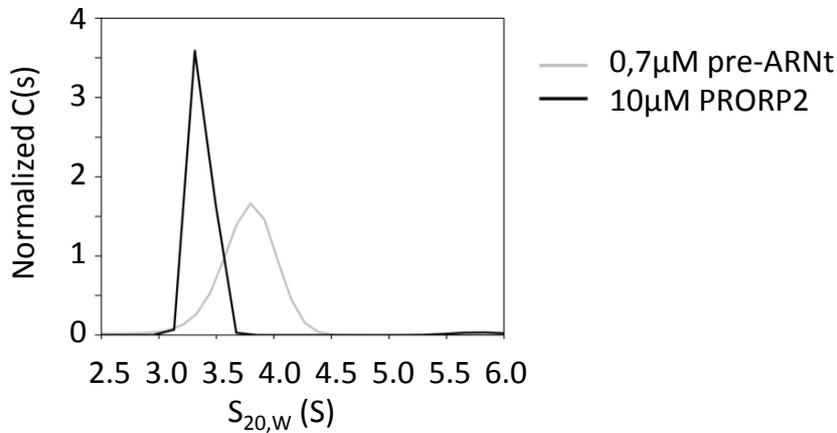


Figure 26 : Profils d’ultracentrifugation analytique (AUC) obtenus avec PRORP2 DD seule et l’ARN seul.

Sédimentation normalisée de la protéine PRORP2 catalytiquement inactive en noir et du pré ARNt^{Cys} mitochondrial L5T0 en gris.

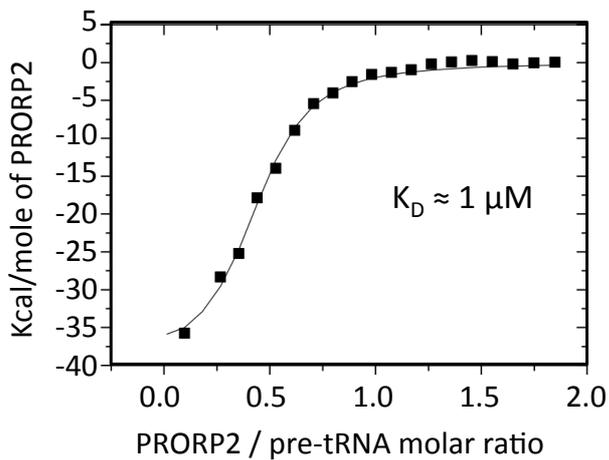


Figure 27 : Analyse du complexe PRORP / pré-ARNt par titration calorimétrique isotherme (ITC).

Expériences d’ITC réalisées par F. Pinker en injectant une solution d’AtPRORP2 DD (catalytiquement inactive) dans la solution de pré-ARNt. La courbe représente la variation de température du système par mole de protéine en fonction du ratio protéine / pré-ARNt. Le K_D obtenu est d’environ $1 \mu\text{M}$.

pour abolir l'activité RNase P ne sont pas suffisantes pour induire la reconnaissance d'une autre base. Des mutations additionnelles à d'autres positions, i.e. à la position 2 (Cheng et al 2016) sont peut-être nécessaires pour restaurer l'activité RNase P.

c. Test d'affinité entre le mutant catalytiquement inactif AtPRORP2DD et les pré-ARNt

Trois méthodes biophysiques différentes ont été utilisées pour mesurer directement l'affinité de PRORP pour les pré-ARNt. La mesure du K_D du complexe PRORP / pré-ARNt requiert l'utilisation d'un mutant catalytiquement inactif. Il s'agit du mutant « AtPRORP2DD » dont 2 aspartates conservées du site catalytique sont mutées en alanines (Gutmann 2012).

Etude d'interaction par ultra centrifugation analytique

La formation du complexe PRORP / pre-ARNt est observée par ultra centrifugation analytique en utilisant pour chaque analyse, six concentrations différentes de AtPRORP2DD (10 μM , 5 μM , 2,5 μM , 1,25 μM , 0,625 μM et 0,312 μM) ainsi que 0,7 μM de pré-ARNt Cys mitochondrial L5T0. La protéine seule à 10 μM est également analysée afin de s'assurer de sa non agrégation au cours de l'expérience. Après centrifugation, on obtient un S_{w20} de 3,364 pour la protéine seule, de 3,777 pour le pré-ARNt seul (Figure 26) et environ de 5,5 pour le complexe. Le K_D calculé est de 1 μM avec un intervalle de confiance à 95 % de 0,3 μM à 3,6 μM .

Etude d'interaction par Titration calorimétrique isotherme

La seconde méthode utilisée pour mesurer l'affinité de PRORP pour les ARNt, la titration calorimétrique isotherme, a été réalisée par F. Pinker et analysée par E. Ennifar. Ces mesures ont été faites en injectant 20 fois 2 μl d'une solution de AtPRORP2 DD à 187 μM dans du tampon SEC2 dans 280 μl d'une solution de pré-ARNt^{Cys} mitochondrial L5T0 à 22 μM dans du tampon SEC2. Le K_D obtenu est d'environ 1 μM (Figure 27).

Etude d'interaction par thermophorèse à micro échelle

Finalement, une troisième méthode, la thermophorèse à micro échelle, a été utilisée pour confirmer les résultats précédents. Cette technique permet de suivre le comportement des molécules dans un gradient de température constitué dans un tube capillaire, la formation de complexe changeant le mouvement des molécules dans le gradient. Cette technique qui utilise

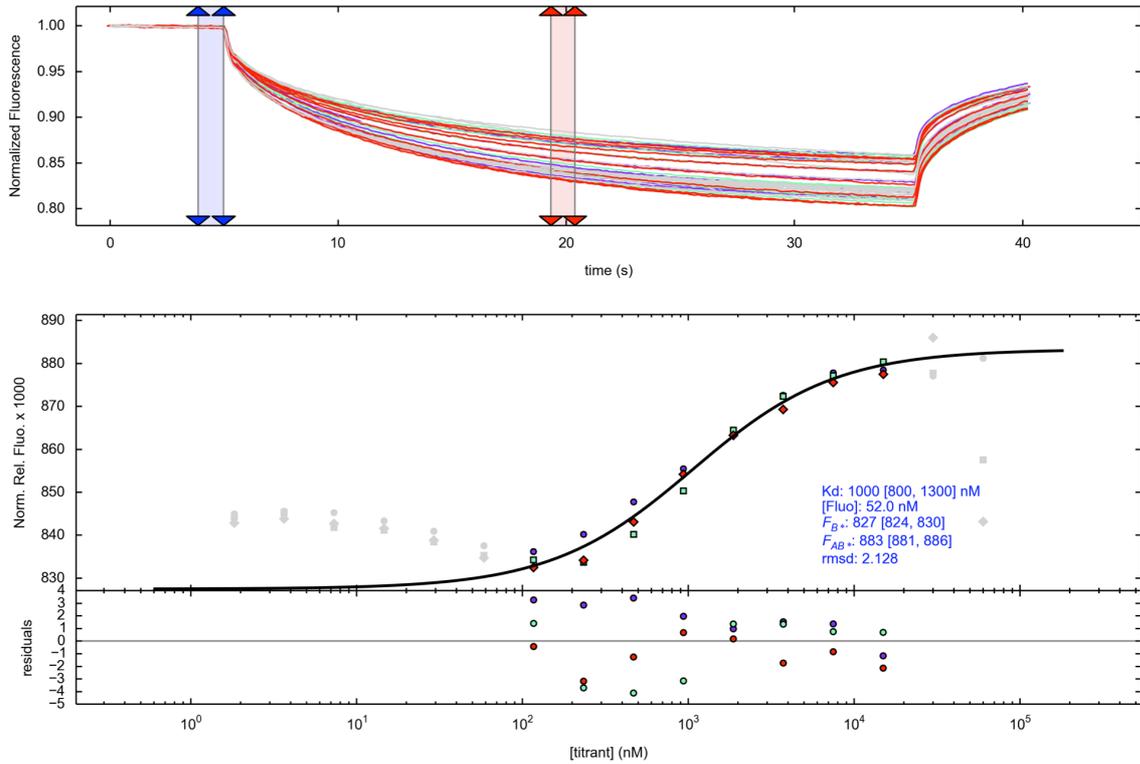


Figure 28 : Analyse du complexe PRORP / pré-ARNt par thermophorèse à micro échelle.

Expériences de thermophorèse a micro échelle réalisées en triplicats avec 52 nM de pré-ARNt^{Cys} mitochondrial L5T0 marqués par un résidu fluorescent Cy5. Le premier point contient 185 μ M de protéine AtPRORP2 DD (catalytiquement inactive) et chaque point suivant est issu d'une dilution en cascade au demi. Le K_D obtenu est de 1 μ M [0,8-1,3 μ M]. Les intervalles entre les deux flèches bleues et les deux flèches rouges représentent les intervalles de temps utilisés pour l'analyse de la stabilité du signal fluorescent.

un ARN fluorescent nécessite ici une protéine fraîchement purifiée et concentrée juste avant la préparation des dilutions. Les expériences ont montré que les protéines PRORP moins fraîches ont tendance à s'agréger dans le capillaire et/ou s'adsorber sur la paroi. Trois types de capillaires disponibles ont été testés. Les capillaires hydrophobes entraînent une adsorption de la protéine et les capillaires « normaux » engendrent une variabilité dans les mesures. Les capillaires « premiums » sont donc utilisés pour les mesures. Ces capillaires sont recouverts d'une couche de polymère empêchant l'adsorption de molécules à leurs surfaces.

Les premières mesures ont été réalisées dans le tampon SEC2 mais l'augmentation de la concentration en protéines entraîne une augmentation de la fluorescence du cys5 greffé sur le pré-ARNt. Afin de saturer l'augmentation du signal en fonction du taux de protéines sans risquer de modifier le K_D , le tampon SEC2 a été supplémenté avec 0,3 mg/ml de BSA. Le signal fluorescent est alors stabilisé à son maximum et homogène entre les capillaires. Deux séries de dilution en cascade au 1 : 2 et 1 : 3 ont été testées. Les dilutions 1 : 2 donnant un écart type moins important, les mesures sont réalisées avec cette dilution. Plusieurs dilutions d'ARNt ont été testées. Une concentration trop importante de pré-ARNt engendre un signal plus variable. Des tests utilisant 200 μM , 100 μM et 50 μM de protéines ont été réalisés. Le signal devant être compris entre 200 et 1000 unités, une concentration de 50 nM d'ARNt engendre un signal d'environ 400 unités. Après une vingtaine d'essais, des concentrations de 371 μM de protéines et de 52 nM d'ARNt ont été sélectionnées. Le K_D mesuré est de 1 μM avec un intervalle de confiance de 0,8 à 1,3 μM pour 3 expériences indépendantes (Figure 28).

Comparaison des différentes méthodes

L'interaction directe entre AtPRORP2DD et les pré-ARNt a été mesurée par trois techniques indépendantes qui ont toutes donné des résultats comparables, avec des K_D mesurés de l'ordre de 1 μM . Cette valeur relativement élevée suggère une interaction très transitoire entre PRORP et ses substrats, en accord avec la fonction de PRORP en tant qu'enzyme de maturation de l'ARN. Bien que les différentes techniques aient donné des résultats comparables, leur mise en œuvre a été plus ou moins aisée. L'ITC demande beaucoup plus de matériel, à des concentrations plus élevées que pour les autres méthodes. L'ITC demande pour des mesures optimales 150 μL à 300 μM par réplica or les protéines PRORP commencent à précipiter vers 200 μM . La MST nécessite aussi une concentration élevée, d'au moins 200 μM et un effet de la concentration est visible lors des mesures. La valeur de thermophorèse chute brutalement à trop forte concentration, ce qui est probablement

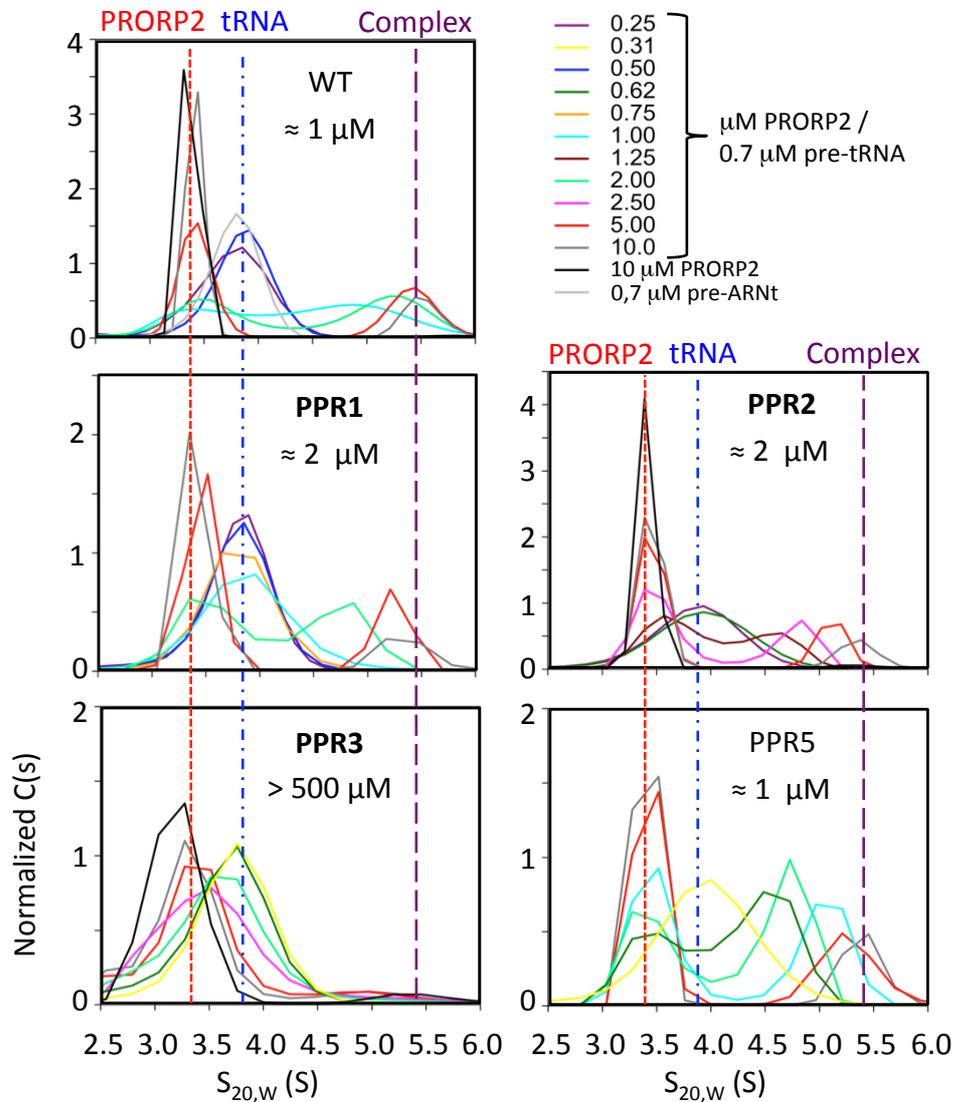


Figure 29 : Analyse du complexe PRORP / pré-ARNT réalisée avec différents mutants PPR et le pré-ARNT^{cys} mitochondriale L5T0.

Les différentes couleurs représentent les différentes concentrations des mutants PPR de PRORP2. La ligne pointillée rouge représente la sédimentation de la protéine PRORP2 seule, la bleue la sédimentation du pré-ARNT seul et la ligne violette la sédimentation du complexe PRORP / pré-ARNT. Le K_D obtenu est indiqué dans chaque cadre correspondant au complexe utilisant la protéine sauvage (WT) ou différents mutants PPR (PPR1 à PPR5).

dû à l'adsorption de la protéine sur le capillaire et ne permet pas d'obtenir un plateau supérieur très clair. L'AUC s'est avérée être la technique la plus facile à mettre en œuvre et la plus reproductible. Elle nécessite des solutions de protéine à 45 μM , concentration à laquelle les protéines PRORP sont stables pendant une longue période.

d. Test d'interaction des protéines mutées dans leur motif PPR et catalytiquement inactifs avec les pré-ARNt

Après avoir étudié l'impact des mutations des motifs PPR sur l'activité RNase P des protéines PRORP, l'effet direct de ces mutations sur la liaison aux ARNt a été étudié. Fort de l'expérience acquise précédemment pour la mesure de K_D , il a été décidé d'utiliser l'ultra centrifugation analytique (AUC) pour mesurer et quantifier l'interaction des mutants PPR avec les pré-ARNt.

Les protéines AtPRORP2DD portant les mutations dans les motifs PPR ont été exprimées dans *E. coli*, purifiées et utilisées pour des expériences d'AUC avec des ARNt Cys L5T0 dans les conditions décrites ci-dessus. Aucune mesure n'a pu être réalisée avec la protéine mutée dans le motif PPR4, car elle s'est avérée être instable dans les conditions expérimentales utilisées ici. La mutation du motif PPR5 semble ne pas affecter la liaison à l'ARN puisque le K_D obtenu est de 0,9 μM ; c'est-à-dire très proche du K_D mesuré avec la protéine sauvage. Avec les mutants PPR1 et PPR2, les K_D mesurés sont de 2,2 μM et 2 μM donc légèrement supérieurs à celui obtenu avec la protéine sauvage. Finalement, les mutations dans le motif PPR3 ont eu l'impact le plus fort, avec un K_D en dehors de la limite de détection des conditions expérimentales utilisées, c'est-à-dire un K_D supérieur à 500 μM et aucun complexe PRORP / ARNt détecté pour le mutant PPR3.

En conclusion, les résultats des deux approches mesurant l'activité RNase P et la liaison directe à l'ARN sont cohérents. Il apparaît que le motif PPR3 est le plus important pour la liaison avec les substrats ARNt. Les motifs PPR1 et PPR2 semblent aussi impliqués dans la liaison à l'ARN, mais de manière beaucoup plus modérée, alors que les motifs PPR4 et PPR5 pourraient ne pas être impliqués du tout dans la liaison aux ARNt (Figure 29).

Mutation		Activité (%)		K_D (μ M) [intervalles de confiance à 95 %]
	Pré-ARNt	L5T0	L50T30	L5T0
WT		100	100	1 μ M [0,6 μ M ; 1,9 μ M]
PPR1		100 \pm 10	98 \pm 8	2,2 μ M [1,7 μ M ; 3,0 μ M]
PPR2		66 \pm 6	44 \pm 11	2 μ M [1,4 μ M ; 2,7 μ M]
PPR3		9 \pm 3	9 \pm 10	>500 μ M
PPR4		89 \pm 3	94 \pm 5	Non mesuré
PPR5		100 \pm 13	98 \pm 14	0,9 μ M [0,7 μ M ; 1,1 μ M]

5. Détermination d'un modèle structural du complexe AtPRORP2 / pré-ARNt

La structure cristallographique de la protéine AtPRORP2 seule (Structure PDB 5FT9) a été déterminée par une doctorante précédente du laboratoire Franziska Pinker et un chercheur post doctorant Pablo Fernandez-Millan en collaboration avec l'équipe de Claude Sauter (IBMC, ARN, UPR9002) (Gobert, 2013 ; Pinker, 2015). La cristallisation a été effectuée par diffusion de vapeur. Le criblage initial a été effectué avec AtPRORP2 à des concentrations de 5 et 10 mg/mL dans le tampon SEC1. Cette concentration s'est révélée excessive, car la protéine a précipité dans de nombreuses gouttes. D'autres expériences ont été réalisées à des concentrations entre 2 et 5 mg/mL et un tampon SEC2 ne contenant que 5 % (p/v) de glycérol au lieu de 15 % (p/v). Différents essais ont été menés en mélangeant 150 nL de solution de protéines avec 150 nL de solution d'agents cristallisants et équilibrants à 277 et 293 K dans un réservoir contenant 35 μ L d'agent cristallisant. Diverses concentrations cristallines et protéiques, ainsi que des additifs et des températures différentes (4°C et 20°C) ont été testés. La meilleure condition de cristallisation a été obtenue avec une solution de cristallisation contenant 200 nM de malonate de sodium pH 6 et 20 % de PEG3350 avec des gouttes comportant 1 μ L de AtPRORP2 à 2,5 mg/mL mélangées à 1 μ L de la solution de cristallisation à 4°C. Chaque maille du cristal forme un parallélépipède dont des arrêtes ont une dimension de 71 Å, 73 Å et 80 Å avec des angles de 63°, 72° et 78°. Le sommet de chaque angle est composé de deux protéines AtPRORP2 tête-bêche. Le cristal a permis de résoudre une structure à une résolution de 3,05 Å, confirmant que la protéine comporte 5 motifs PPR et un domaine NYN reliés par un domaine structural de liaison au zinc.

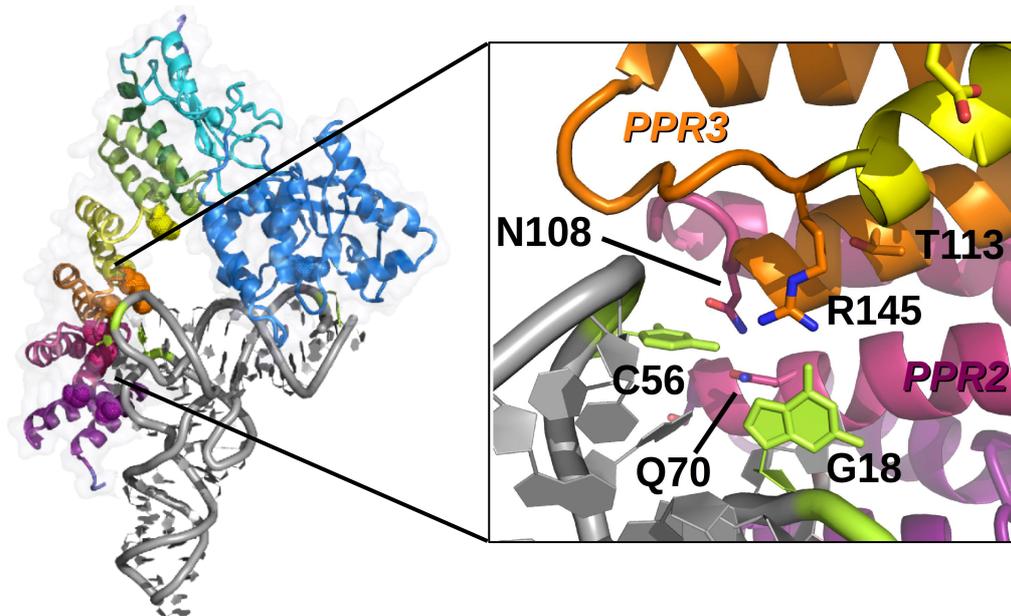


Figure 30 : Structure du complexe AtPRORP2 / pré-ARNt par SAXS

Les différents motifs PPR de PRORP2 sont présentés en pourpre, rose, orange, jaune et vert. Le domaine structural en doigt de zinc est en bleu clair alors que le domaine NYN est en bleu foncé. Les sphères sur les motifs PPR représentent les positions 5 et 35 de chaque motifs. Le pré-ARNt est présenté en gris avec en jaune les bases C56 et G18 en interaction avec les motifs PPR2 et PPR3. Le cadre représente un agrandissement de la zone d'interaction entre AtPRORP2 et le pré-ARNt.

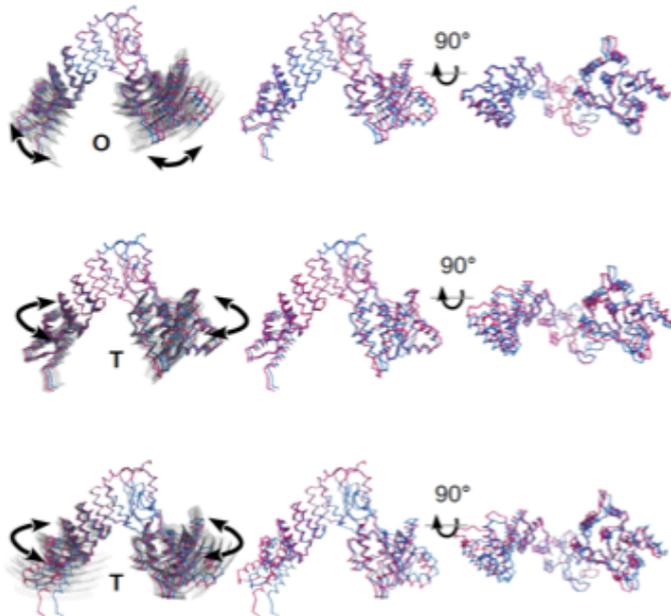


Figure 31 : Variants conformationnels de PRORP2.

L'analyse de différentes fenêtres de SAXS montre une série de conformations de la protéine. L'angle entre le domaine PPR et le domaine NYN peut être plus ou moins ouvert (O) ou les deux domaines peuvent subir différentes torsions (T) autour de l'axe formé par la pointe du Λ de la structure des protéines PRORP.

Toutefois, la compréhension fine du mode d'action des protéines PRORP requiert la résolution de la structure du complexe formé par PRORP et un pré-ARNt. Cela devrait permettre de comprendre comment les motifs PPR, en particulier PPR1, PPR2 et PPR3 interagissent avec l'ARN. Une première étape a consisté à déterminer la structure en solution du complexe PRORP / pré-ARNt par diffusion des rayons X aux petits angles (SAXS). Ce travail a aussi été réalisé en majeure partie par F. Pinker et C. Sauter, il est décrit en détail dans la publication ci-jointe Pinker & Schelcher, et al. 2017. Brièvement, les données biophysiques et biochimiques ont été combinées pour établir un modèle d'interaction entre PRORP et l'ARNt. La structure cristallographique expérimentale de AtPRORP2 et un modèle structural de l'ARNt Cys ont été utilisés pour créer différents modèles possibles du complexe. Les courbes SAXS théoriques des différents modèles sont comparées avec les données expérimentales de SAXS obtenues pour le complexe. Les analyses biochimiques préalables ont démontré que l'interaction de AtPRORP2 impliquait surtout les motifs PPR3 et dans une moindre mesure PPR2 (décrit ici) et que PRORP interagit avec des nucléotides conservés des boucles T et D des ARNt, en particulier les bases G18, G19, C56 et G57 (Gobert 2013). Le code PPR prédit que le motif PPR2 interagit avec un C et le motif PPR3 avec un G ou un A. Un des modèles possibles du complexe implique des interactions entre PPR2 et C56 ainsi qu'entre PPR3 et G18 (Figure 30). Ce modèle correspond le mieux aux données SAXS expérimentales et a donc été retenu. Le Chi mesurant l'accord entre les données théoriques des modèles et les données expérimentales est de 2,4 pour le modèle retenu contre des valeurs supérieures à 3,8 pour tous les autres modèles.

6. Mise en évidence de la flexibilité structurale des protéines PRORP

La structure de PRORP en solution a permis de démontrer la flexibilité de AtPRORP2 et vraisemblablement de toutes les protéines PRORP. Lors de l'analyse par SAXS, la protéine est en solution ce qui lui permet de bouger et d'adopter de différentes conformations. La colonne de chromatographie par exclusion de taille permet de séparer les différentes conformations. En analysant différentes fenêtres de SAXS, il est possible d'obtenir des données structurales pour ces différentes conformations de la protéine. La série de conformations observées montre que le domaine PPR et le domaine NYN de PRORP forment un angle plus ou moins ouvert (Figure 31). Ces deux domaines peuvent aussi subir une rotation au niveau du domaine de liaison au zinc. Ces caractéristiques pourraient être nécessaires afin de s'adapter à différents types de substrats ARN. Il est également possible

que les changements de conformation soient requis pour la liaison avec les substrats et / ou le relargage des ARN matures.

Mon travail a permis de démontrer l'intérêt des motifs PPR2 et surtout PPR3 dans la reconnaissance des pré-ARNt par AtPRORP2. Il est vraisemblable que ces résultats puissent être extrapolés à l'ensemble des protéines PRORP. Les motifs 4 et 5 ne semblent pas impliqués dans la reconnaissance des pré-ARNt, il est possible que leur rôle soit uniquement structural afin de déterminer la distance entre le site de clivage et le site de reconnaissance des pré-ARNt, servant ainsi de règle moléculaire. Il est également possible que les motifs PPR1, 4 et 5 puissent être impliqués dans la reconnaissance d'autres substrats de PRORP encore non identifiés. Cette hypothèse pourra être testée en identifiant les autres substrats ARN des protéines PRORP, à l'échelle des transcriptomes entiers, étude qui fait l'objet de la deuxième partie de mes travaux de thèse.

L'intégralité des résultats de ce chapitre a permis la publication d'un article de recherche en co-premier auteur avec F. Pinker, étudiante précédente, ayant préparé son doctorat en co-tutelle entre mon équipe d'accueil et l'équipe de C. Sauter. L'analyse structurale a été réalisée par F. Pinker et Clauder Sauter (IBMC). Cette collaboration a permis de déterminer la structure cristallographique de AtPRORP2 ainsi que d'analyser par SAXS le complexe AtPRORP2/pré-ARNt (paragraphe 5 et 6). En parallèle, j'ai construit les mutants PPR, réalisé leurs purifications et les analyses biochimiques et biophysiques des protéines (paragraphe 1 à 4). C'est la combinaison des analyses biophysiques, biochimiques et structurales qui a permis de proposer le modèle du complexe AtPRORP2/pré-ARNt présenté dans l'article.

Publication : Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding.

Cet article fait l'objet de la couverture du numéro du 25/08/17 de JBC.

(<http://www.jbc.org/content/292/34.cover-expansion>)

Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding

Franziska Pinker^{1,2,+}, Cédric Schelcher^{2,+}, Pablo Fernandez-Millan¹, Anthony Gobert², Catherine Birck³, Aurélien Thureau⁴, Pierre Roblin^{4,5}, Philippe Giegé^{2,§,*} and Claude Sauter^{1,§,*}

¹Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France

²Université de Strasbourg, CNRS, IBMP, UPR 2357, F-67084 Strasbourg, France

³Université de Strasbourg, CNRS, IGBMC, UMR 7104, F-67404 Illkirch, France

⁴Synchrotron SOLEIL, l'Orme des Merisiers Saint Aubin, F-91410, France

⁵URBIA-Nantes, INRA Centre de Nantes, 60 rue de la Géraudière, Nantes, F-44316, France

^{+,§}contributed equally

*To whom correspondence should be addressed: **Dr Philippe Giegé**, tel: +33 367 155 363, email: philippe.giege@ibmp-cnrs.unistra.fr and **Dr Claude Sauter**, tel: + 33 388 417 102, fax: +33 388 602 218, email : c.sauter@ibmc-cnrs.unistra.fr

Running title: Biophysical analysis of PRORP2-tRNA complex

JBC keywords: Ribonuclease P (RNase P) / precursor tRNA (pre-tRNA) / RNA processing / small-angle X-ray scattering (SAXS) / X-ray crystallography

Other keywords: PRORP / pentatricopeptide repeat / tRNA maturation

ABSTRACT

RNase P is a universal enzyme that removes 5' leader sequences from tRNA precursors. The enzyme is therefore essential for maturation of functional tRNAs and mRNA translation. RNase P represents a unique example of an enzyme that can occur either as ribonucleoprotein or as protein alone. The latter form of the enzyme called PRORP (PRotein-Only RNase P) is widespread in eukaryotes, in which it can provide organellar or nuclear RNase P activities. Here, we have focused on Arabidopsis nuclear PRORP2 and its interaction with tRNA substrates. Affinity measurements helped assess the respective importance of individual pentatricopeptide repeat motifs in PRORP2 for RNA binding. We characterized the PRORP2 structure by X-ray crystallography and by small-angle X-ray scattering (SAXS) in solution, as well as that of its complex with a tRNA precursor by SAXS. Of note, our study reports the first structural data of a PRORP-tRNA complex. Combined with complementary biochemical and biophysical analyses, our structural data suggest that PRORP2

undergoes conformational changes to accommodate its substrate. In particular, the catalytic domain and the RNA binding domain can move around a central hinge. Altogether, this work provides a refined model of the PRORP-tRNA complex that illustrates how protein-only RNase P enzymes specifically bind tRNA and highlights the contribution of protein dynamics to achieve this specific interaction.

INTRODUCTION

Transfer RNA maturation involves many post-transcriptional steps. Among them, 5' leader sequences are removed by an endonuclease activity called RNase P (1). This processing step was assumed to be universally performed by ribonucleoprotein (RNP) particles, containing a ribozyme and up to 10 protein subunits, until the discovery of protein-only RNase P in eukaryotes (2).

These enzymes called PRORP for PRotein-Only RNase P were first identified in human mitochondria and plants (3,4). Contrary to initial

appreciations, PRORP enzymes are not restricted to organelles in few species (5). They are actually widespread in eukaryotes, occurring in mitochondria and / or chloroplasts as well as in nuclei in 4 out of 5 eukaryote supergroups (6). In some groups, i.e. in most Chloroplastida (Viridiplantae), Stramenopiles and Trypanosomatida, PRORP enzymes have seemingly entirely replaced ribonucleoproteins for RNase P activity, as experimentally shown for *Arabidopsis thaliana* (7), *Chlamydomonas reinhardtii* (8) and *Trypanosoma brucei* (9). PRORP enzymes were also characterised in the moss *Physcomitrella patens* (10). Intriguingly, in this species, a nuclear PRORP is not essential although no recognizable ribonucleoprotein RNase P is present (6). Remarkably, the occurrence of PRORP and RNP RNase P appears mutually exclusive in compartments or in entire organisms (6). RNPs or of PRORP might have been retained in specific cell compartments or entire organisms because they have different substrate spectra *in vivo*. Alternatively, in specific clades, RNPs or PRORP might have evolved additional functions that cannot be held by the other type of enzyme. (6).

The biochemical characterization of PRORP enzymes, their structure in solution (11), as well as the crystal structure of mitochondrial *Arabidopsis* and human PRORP enzymes (12-14) have revealed two-domains enzymes. They contain a C-terminal nuclease domain belonging to the NYN (N4BP1, YacP-like Nuclease) family (15) and a N-terminal pentatricopeptide repeat (PPR) domain (16,17) believed to be responsible for RNA binding and substrate specificity. These two main domains are bridged by a central zinc-binding domain (2). A comparative kinetic analysis of RNase P activities by PRORP and RNPs has suggested that the two types of enzymes use different catalytic mechanisms (18). Remarkably, this comparative analysis also revealed that RNPs are better catalysts than PRORP enzymes, at least *in vitro*. However, the identification of tRNA residues in contact with PRORP (11) and RNP RNase P (19) has suggested that RNPs and PRORP use a similar strategy to recognize their substrates (11). Still, how this is achieved at the protein level remained unknown and the dynamics

required for PRORP mode of action was unexplored (20).

In order to tackle these questions, we determined the crystal structure of *Arabidopsis* nuclear PRORP2 together with a SAXS model in solution, and explored the role of its PPR domain for substrate binding. We found that two particular PPR motifs are required for specific tRNA recognition by PRORP2. Furthermore, we studied the PRORP2-tRNA complex in solution by SAXS to establish an interaction model. Taken together, our results identify structural features in PRORP2 important for RNA binding process and suggest that the enzyme is flexible and undergoes conformational changes to perform its activity.

RESULTS

Crystal and solution structure of *Arabidopsis* nuclear PRORP2

Crystallization conditions were extensively screened for different PRORP2 constructs, including catalytically impaired and shortened forms. Crystals were only obtained for the full-length wild-type PRORP2 (21). Their analysis led to the determination of the enzyme structure at a resolution of 3.05 Å (Figure 1A and Supplemental Table 1). As proposed based on sequence similarities with the organellar PRORP1 (2,11), PRORP2 adopts the same characteristic Λ -shape. The N-terminal arm consists of a PPR domain made of five PPR motifs, the C-terminal arm consists of the NYN metallonuclease domain, and a bipartite zinc-binding domain (ZBD) bridges the two arms together. The 27 first residues including the nuclear localization signal (NLS), as well as the 20 last amino acids including the 6-His affinity tag, were probably floppy in the crystal packing and could not be visualized in the electron density map. The two copies of PRORP2 present in the triclinic unit cell are very similar (rmsd = 0.85 Å, Supplemental Figure 1). They also superimpose well with PRORP2 monomers (average rmsd = 1.32 Å) recently described by Karasik et al. (22). As in the latter crystal structure determined concomitantly to ours in the same triclinic environment and in fairly similar crystallization conditions, we do not observe any cation bound to conserved aspartate residues in the catalytic site,

although we and Karasik et al. (22) show that, similar to PRORP1, Mg^{2+} ions are required for catalysis (Supplemental Figure 2).

The full-length PRORP2 and a more compact form PRORP2xs were characterized in solution by SAXS as well (Figure 1C). Their gyration radii (R_g) of respectively 35 and 32 Å derived from the Guinier analysis are in good agreement with the value calculated from the crystal structure (29.5 Å) deprived of N- and C-terminal extensions (Supplemental Figure 3). Possible conformations of PRORP2 were explored by a normal mode analysis (NMA) and the model fitting best the SAXS data (Supplemental Figure 4) was extended to missing N- and C-regions and refined under SAXS constraints (Supplemental Figure 5). Figure 1C represents the resulting model showing the core of nuclear PRORP2 in solution. SAXS data on both PRORP1 and PRORP2 indicate that the structural organization of the enzyme with its central hinge between the catalytic domain and the ZBD introduces flexibility in its backbone and allows more open conformations than those observed in crystal structures (see Discussion and Supplemental Figure 4).

Defining conditions for a stable PRORP2-tRNA complex

A prerequisite for the characterization of the PRORP2-tRNA complex was to define buffer conditions compatible with various biochemical and biophysical analyses, and ensuring both protein and RNA stability. The effect of the salt concentration was tested on PRORP2 catalytic activity. Initial buffer conditions contained 250 mM NaCl to increase the solubility of the enzyme but it appeared to be detrimental to both crystallization (21) and activity. As shown in Supplemental Figure 6 the cleavage of pre-tRNA sharply decreases beyond 150 mM NaCl. The latter concentration was chosen in combination with 5 % glycerol (w/v) as a good compromise to maintain both PRORP2 monodispersity in solution and its activity.

Defining a compact but well-processed RNA substrate was another prerequisite, especially for SAXS or crystallography to avoid long and flexible extensions. We therefore gradually trimmed the leader sequence of our Arabidopsis

pre-tRNA^{Cys(GCA)} constructs from 51 down to 5 nucleotides (Supplemental Figure 7). The best compromise was obtained with the L5T0-tRNA^{Cys} (i.e. with a leader sequence of 5 nucleotides, no trailer sequence and no added CCA) which proved to be fully cleaved and showed up as a single band on native gel. This is in agreement with the study by (22) showing that PRORP2 preferentially binds substrates with short 5' leaders and 3' trailers. L5T0 construct was selected for all subsequent analyses. In the same line, the short PRORP2xs construct was preferred over the full-length catalytic mutant to avoid long and flexible regions, and to facilitate SAXS modelling (Supplemental Figure 5).

PRORP2 affinity for pre-tRNA substrates

The experimental setup defined above was used to determine the affinity of Arabidopsis PRORP2 for pre-tRNA by several biophysical methods. Hence, a catalytically inactive mutant of PRORP2xs containing two alanines in place of two aspartates at positions 421 and 422 in the NYN domain (7) was used in combination with the L5T0 tRNA substrate. Isothermal titration calorimetry (ITC) revealed a first K_D value of 1 μ M for the interaction of PRORP2 with tRNA, with a ΔH of $-3.9 \cdot 10^4$ cal.mol⁻¹, a ΔS of -107 cal.mol⁻¹.degree⁻¹ and a stoichiometry of 0.4 (Figure 2A), thus suggesting that a proportion of either of the partners would not form a complex or that one of the partner (tRNA) partially oligomerizes as seen by SEC in similar conditions (data not shown). Subsequent dynamic light scattering (DLS) and size exclusion chromatography (SEC) measurements confirmed that the sample did not aggregate during the experiment. In order to confirm these results, the PRORP2-tRNA complex formation was also monitored by microscale thermophoresis (MST), using a tRNA^{Cys} precursor labelled in 5' with a fluorescent dye (Cy5). This revealed a K_D of 1 μ M with 95 % confidence limits, between 0.8 and 1.3 μ M in triplicate experiments (Figure 2B). Finally, the PRORP-tRNA complex was subjected to sedimentation velocity analytical ultracentrifugation (SV-AUC). Replicate analyses of titration series with a constant concentration of pre-tRNA^{Cys} and increasing concentrations of PRORP2 yielded an estimated K_D value of 1 μ M

[95% confidence limits, 0.4 μ M, 2.3 μ M] (Figure 2C), in good accordance with the values measured with MST and ITC. Altogether, results indicate a K_D value of the PRORP-tRNA complex in the micromolar range. This relatively low affinity suggests a transient interaction of PRORP enzymes with tRNA precursors, as expected for a maturation enzyme.

Relative importance of PPR motifs for interaction with tRNA

PRORP proteins belong to the huge family of pentatricopeptide repeat (PPR) proteins. These eukaryote specific RNA binding proteins are involved in a wide variety of post-transcriptional processes such as RNA editing, splicing or the maturation of transcript ends (16,23). They are composed of *circa* 35 amino-acids tandem repeats of degenerate primary sequences although bearing a conserved helix-turn-helix structure. The investigation of PPR protein mode of action has revealed that each PPR motif specifically interacts with a defined ribonucleotide. Some residues are particularly important to achieve this specificity, i.e. two residues located toward the start and at the end of PPR motifs, termed here positions 5 and 35, according to Cheng et al. (23). The nature of amino-acids at these positions defines a combinatorial code for RNA recognition by PPR motifs (24-26). The PPR domain of Arabidopsis PRORP2 is composed of 5 such motifs (PPR1-PPR5) (Supplemental Figure 8). Positions 5 and 35 of motifs PPR2 and PPR3 are particularly well conserved in plants, with the occurrences of a Q or a N at PPR2 position 5 and a N or a S at PPR2 position 35 as well as a T at PPR3 position 5 and a R at PPR3 position 35 in an alignment of 100 plant PRORP sequences (6). Shorter alignments are exemplarily shown in Supplemental Figures 8 and 9. Although the “PPR code” is not yet fully understood, (i.e. target RNAs cannot be predicted for many PPR proteins or motifs and the exact involvement of other residues besides positions 5 and 35 for specific RNA binding is not entirely known), the latest resources and knowledge on PPR RNA recognition (23) enable to predict that PRORP2 motif PPR2 would recognize a cytidine while motif PPR3 would recognize a purine. In

contrast, no clear prediction can be made for PPR motifs 1, 4 and 5.

Accordingly, in order to test the relative importance of individual PPR motifs for PRORP function, presumably for substrate recognition, PRORP2 was mutated at positions 5 and 35 of its five PPR motifs. The nature of amino acids and mutations at these positions is indicated on Supplemental Figure 8. The five double mutants were expressed and purified to homogeneity. Solubility, stability and structural integrity of mutants were verified by DLS and SAXS. All mutants behaved similarly to the wild-type enzyme and its catalytically inactive version in solution (Supplemental Figure 10), except PPR4 variant which is unstable and prone to aggregation, especially upon RNA removal. Therefore, the latter mutant could only be used for cleavage assays at low concentration.

Hence the five mutant proteins were assayed for *in vitro* RNase P activity with two tRNA^{Cys} precursors with leader sequences of either 5 or 51 nucleotides. The quantification of relative cleavages revealed that PPR mutants 1, 4 and 5 were as active as wild type PRORP2. However, activity decreased by 45 % (on average for the two tRNA substrates) in PPR2 mutant and by 91 % in PPR3 mutant (Figure 3A). This suggests that motifs PPR2 and 3 are the most important for substrate recognition by PRORP2 and in accordance with comparable results obtained for the organellar PRORP1 (27).

Still, RNase P cleavage experiments represent indirect evidence for PRORP binding to tRNA. We therefore analysed direct interaction between PPR mutants and RNA using SV-AUC to determine which PPR motifs are indeed involved in tRNA binding. For this, the PRORP2 PPR double mutants were further mutated in the catalytic site as described above to obtain catalytically inactive proteins enabling K_D measurements. The observed broad distributions of sedimenting species (between the faster tRNA component and the complex species) are characteristic of rapidly reversible systems, due to dynamically associating and dissociating events during sedimentation. The swfast isotherms were generated and analyzed to get the binding constants. The K_D of PPR5 mutant interaction with tRNA^{Cys} was estimated to 0.9 μ M

[95% confidence limits, 0.7 μ M, 1.1 μ M] similar to wild type PRORP2 [95% confidence limits, 0.6 μ M, 1.9 μ M], whereas the K_D of PPR1 and PPR2 mutants were estimated to 2.2 μ M [95% confidence limits, 1.7 μ M, 3.0 μ M] and 2 μ M [95% confidence limits, 1.4 μ M, 2.7 μ M] respectively. For PPR3, no estimate of the K_D was possible due to the absence of complex formation in the experimental conditions (Figure 3B). The analysis of AUC data showed that K_D variations were not due to the aggregation or misfolding of PPR mutants. No data could be obtained for PPR4 mutant due to its instability at the concentrations required for AUC (see above).

Altogether, results indicate that motif PPR3 is the most important to form the PRORP2-tRNA complex. Mutations on PPR2 motif affects both catalysis and RNA binding, although to a lesser extent. PPR1 mutant is only moderately affected for RNA binding. This is probably because the mutation at position 35 introduces a negative charge that can interfere with RNA binding. Still, even though PPR1 mutant does not display a loss of RNase P activity, our data suggest that PPR1 motif is in the vicinity of the RNA, in accordance with previous work showing contacts between the N-terminal extremity of PRORP and tRNA (28). In contrast, PPR motifs 4 and 5 do not seem to be involved in RNA interaction since their mutations do not alter RNA binding and cleavage.

Modelling of PRORP2-tRNA complex based on SAXS data

The conditions defined to favour a stable PRORP2-tRNA interaction were used to study the PRORP2xs-L5T0 pre-tRNA complex by SAXS. For this purpose, PRORP2xs, pre-tRNAs and their complex were separated by analytical SEC just upstream of the SAXS cell. Because the three entities could hardly be resolved by SEC (Figure 4A) and because of the μ M affinity of PRORP for tRNA, the SEC separation was carried out in the presence of 1.5 μ M of free enzyme in the mobile phase to avoid complex dissociation. The complex was detected in the second part of the main scattering peak where the estimated R_g along data collection makes a plateau at ~ 34 Å (Figure 4B), i.e. a value 6 % larger than that observed for

PRORP2xs alone (Supplemental Figure 3). The SAXS profile and the $P(r)$ distribution are also clearly distinct from those obtained with the isolated enzyme (Supplemental Figure 3).

To build a model of the PRORP2-tRNA complex, an ensemble of data previously available and determined here were considered:

- Our previous footprinting analysis identified tRNA residues protected from RNase digestion and thus in contact with PRORP in the tRNA D and T loops (U16, G18, G19 and C56) (11).
- These residues and others that are universally conserved among canonical tRNAs (29) were mutated. For some of them, mutations resulted in a complete loss of RNase P activity, i.e. G18, C56 and R57 in the full length tRNA (11) as well as C56 and R57 in a minihelix corresponding to the tRNA acceptor domain (30).
- Mutations also showed that the anticodon domain of tRNAs is completely dispensable for RNase P activity (11,30).
- The PRORP2 catalytic pocket that includes D421 and 422 (and binds catalytic Mg (12)) must be next to tRNA positions -1 and +1.
- PPR mutagenesis and AUC data determined here suggest that PPR motifs 2 and 3 are involved in tRNA binding, accordingly at the opposite side of the acceptor arm, i.e. in the D and T loops.
- The PPR code predicts that motifs PPR2 and 3 should bind a cytidine and a purine, respectively. Best candidates are positions G18, G19, C56 and R57.

Taking in account these spatial criteria, a series of models of the complex was generated from the crystal structure of PRORP2 and the model of L5T0 pre-tRNA^{Cys} (prepared by homology modelling using Assemble (31) and based on *E. coli* tRNA^{Cys} (GCA) crystal structure, PDBid: 1U0B (32)). These models test different combinations of PPR / nucleotide interactions. Their theoretical scattering profiles were compared with experimental SAXS data to select the best model (i.e. with best goodness-of-fit or lowest Chi

value). Since PPR3 was found to be the most important for RNA interaction, models were first built with PPR3 interacting with either G18, G19 or R57. In the models involving G19 and R57, the orientation of the tRNA does not allow an additional interaction of PPR2 with a tRNA residue. In contrast, in the model involving PPR3 interaction with G18, PPR2 is close to C56. This model of the complex displayed in Figure 4 is the only one satisfying all distance criteria: (i) Motifs PPR2 and PPR3 binding 2 residues in the D/T loops of tRNA, (ii) nucleotides -1 and +1 in the vicinity of metal binding aspartates. Most importantly, this model also gives the best fit with SAXS experimental data (Chi of 2.4). Other combinations involving a single PPR interaction to bases G19, C56 or G57 lead to a reorientation of the tRNA with respect to PRORP2 and higher Chi values (ranging from 3.8 to 5.0), indicating that the shape of these models are less representative of the objects present in solution (Supplemental Figure 11).

In our model, the backbone conformation of D/T loops was essentially kept as in the *E. coli* template tRNA, but the nucleobases C56 and G18 were flipped in syn conformation to point toward the protein. Indeed, in the unprocessed pre-tRNA transcript, these bases are certainly unpaired to keep D/T loops flexible and accessible to modification enzymes and thus to PPR modules as well. For instance, tertiary interactions that stabilize mature tRNAs (29), i.e. the C56-G19 or the G18-Ψ55 interactions may hardly take place before PRORP binds pre-tRNA, since RNase P cleavage is one of the first step of pre-tRNA maturation after transcription (33). Among PRORP1 lysines shown in a recent mass spectrometry study to cross-link with the tRNA (28), the only two that are conserved in PRORP2 (K42 and K387, which are equivalent to K109 and K439 in PRORP1), are close to the tRNA backbone in our SAXS model of the complex. The comparison of the crystal structure and SAXS models of PRORP2 alone or complex with pre-tRNA suggests that the enzyme core is quite flexible with a main hinge between the zinc-binding and the catalytic domain. In particular, structural differences between free and complexed form of PRORP show that major conformational

changes may take place upon binding of pre-tRNA. It appears that the plasticity of PRORP2 backbone allows opening of its Λ-shape and a rotation of the catalytic domain to accommodate tRNA substrates (Figure 5).

DISCUSSION

Since the first descriptions of protein-only RNase P (3,4), crystal structures and biophysical studies of Arabidopsis and human PRORP revealed the functional organization and structural diversity of these enzymes (reviewed by Schelcher et al. (20)). Initial investigations of PRORP mode of action showed which tRNA residues are in contact with PRORP (11), but the way tRNA recognition is achieved by PRORP was unknown and no data was available on the structure of the PRORP-tRNA complex and on the dynamics of the enzyme.

Mutagenesis of PPR motifs and the model derived from SAXS data acquired on the PRORP2-tRNA complex suggest that the sole PPR2 and 3 motifs of PRORP are involved in tRNA recognition. Besides PRORP, all characterized PPR proteins bind single stranded RNA (16) and the current trend is that tandem arrays of PPR motifs specifically bind unpaired RNA bases in a sequence specific manner (24). Since PRORP binds any tRNA of canonical structure, it was proposed that PRORP diverged from the general mode of action of PPR proteins and developed a specific RNA recognition mode based on structural features rather than on linear sequences (2,30). Our results suggest that this is not completely the case, at least for motifs PPR2 and 3 that appear to follow the canonical base recognition by PPR motifs. Indeed, PPR2 would recognise C56 and PPR3 a purine at the corner of the tRNA. This purine could be either G18 or G57, although our SAXS model indicates that G18 is closer from motif PPR3 and is thus more likely recognized by this motif when PRORP interacts with a full tRNA. However, in the context of a mini-helix mimicking the acceptor arm with a less constrained single-stranded T-loop region, as used by Brillante et al. (30), PPR3 might also be able to bind G57 while PPR2 interacts with C56.

In contrast motifs PPR4 and 5 do not seem to participate in tRNA binding and their function, beside maintaining the PPR architecture, remains elusive. They might be involved in the recognition of yet unidentified non-tRNA substrates of PRORP2. Alternatively, the role of these motifs might be to position PPR2/3 at the right distance from the active site of the NYN domain. The overall PRORP fold would then act as a ruler that selects the cleavage site with respect to tRNA corner held by PPR2/3. However, although positions 5 and 35 of PPR motifs are recognised as the most important for specific RNA binding, the functional relevance of other positions such as position 2, that was also proposed to be involved in RNA interaction, for PRORP / tRNA recognition remains unexplored.

In addition to PRORP RNA recognition process, we determined and compared the crystal and different solution models of PRORP2 in order to capture various conformations of PRORP and look at the conformational landscape of PRORP proteins. To explore possible domain reorientations, we performed a normal mode analysis (NMA) known to provide a good description of low-frequency collective motions in proteins (34). The comparison of NMA perturbed models with our SAXS data in solution suggests that both PRORP1 and PRORP2 can adopt a variety of conformations owing to the presence of a hinge between the catalytic NYN domain and the central zinc-binding domain. Related movements correspond to the opening of the Λ -shape of PRORP and the rotation of its catalytic domain around the central hinge (Supplemental Figure 3). Models were further refined against SAXS data with the genetic algorithm DADIMODO and led to an excellent fit for PRORP2xs, showing again multiple compatible conformations with a reorientation of the catalytic domain (Supplemental Figure 5). The presence of long and floppy extensions in the full-length enzyme led to slightly higher goodness-of-fit (Chi) but the trend remained the same. The superposition of PRORP2 crystal structure and solution models alone or in complex suggest the existence of a continuum of conformations that allow the accommodation of RNA substrates, and possibly with a longer helical domain than a classical tRNA acceptor arm (Figure

5). This adaptability is exemplified by the capacity of PRORP to cleave substrates such as tRNA^{His} that has an 8 base pairs acceptor stem and can be processed by PRORP at position -1 and +1 (35) or by the cleavage of tRNA-like substrates as observed for mitochondrial *nad6* and *orf291* mRNAs (7,36). Overall, the intrinsic plasticity of PRORP2, which likely applies to all members of the PRORP family may be crucial to bind RNA substrates and to release matured RNA products. Such type of conformational adaptation has been observed in other monomeric tRNA binding enzymes. In the case of the human mitochondrial phenylalanyl-tRNA synthetase, the crystal structure of the isolated enzyme revealed a very compact conformation. SAXS data on the complex with tRNA indicated a large rigid-body motion of the anticodon binding domain upon tRNA binding (37). Other examples concerned the bacterial elongation factor Ef-Tu and the conformational transition following GTP hydrolysis that leads to ribosome translocation and Ef-Tu release (38), the bacterial TruB which catalyzes pseudouridine formation at U55 in tRNA and undergoes a large rigid body displacement of its C-terminal region upon tRNA binding (39), or the archeal *O*-Phosphoseryl-tRNA kinase (PSTK), the enzyme that phosphorylates Ser-tRNA^{Sec} to produce *O*-phosphoseryl-tRNA^{Sec} which displays a >60 Å movement of its C-terminal domain to bind the variable region of tRNA^{Sec} (40). Large conformational movements thus seem to be a general feature to modulate the recognition, binding or release of tRNAs.

EXPERIMENTAL PROCEDURES

PRORP2 and tRNA purification

Expression of the full-length enzyme in the wild-type (PRORP2wt) and catalytically impaired (PRORP2mut) versions as well as of PPR mutants, and their purification by IMAC and SEC were performed as described by Pinker et al. (21) A more compact and catalytically inactive form (PRORP2xs) was designed and the region of PRORP2 cDNA coding for residues 25-516 was introduced in a pTYB1 plasmid (New England

Biolabs), to produce the protein fused in C-terminal to self-cleavable intein and chitin-binding domain. In both PRORP2mut and PRORP2xs two essential aspartate residues (421, 422) were mutated in alanines to inhibit the nucleolytic activity and facilitate the formation of stable complexes and PPR motifs were mutated to abolish RNA interaction. Mutations were introduced with the QuikChange (R) Site-Directed Mutagenesis kit (Stratagene) using PCR primers designed with the manufacturers qcprimerdesign online tool.

The expression of PRORP2xs was performed in *Escherichia coli* BL21 (DE3) strain at 17°C in 2 l LB medium containing ampicillin and induced with IPTG. The cells harvested by centrifugation, resuspended in a buffer containing protease inhibitors (Sigma, catalogue No. S8830; one tablet per 100 ml), 50 mM HEPES–Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 1 mM TCEP, and were disrupted by ultrasonication. Cell debris were removed by centrifugation and the supernatant applied onto a 50 ml chitin affinity column (New England Biolabs) equilibrated with the same buffer including 500 mM NaCl. To trigger intein cleavage and PRORP2xs release, 50 mM DTT was added to the buffer and the protein was incubated on the column for 16 h at 4°C. After elution, PRORP2xs was further purified by SEC on a Superdex 200 10/300 GL column (GE Healthcare) and stored in 50 mM HEPES–Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 1 mM TCEP. To reduce NaCl concentration prior to complex formation this storage buffer was exchanged by ultrafiltration or dialysis with a buffer containing 150 mM NaCl, 30 mM HEPES–Na pH 7.5, 5 % (w/v) glycerol, 1 mM TCEP (buffer P). Sample quality was systematically assessed by SDS-PAGE and DLS as described previously (11,21).

Arabidopsis mt-tRNA^{Cys(GCA)} L5T0 and L51T30 pre-tRNA cDNAs (with leader sequences of 5 or 51 nucleotides as well as trailer sequences of 0 or 30 nucleotides) were amplified from *Arabidopsis* seedlings total cDNA, cloned in pUC19 and transcribed *in vitro* by T7 RNA polymerase as described by Gobert et al. (11).

Determination of PRORP2 crystal structure

Crystallisation and crystallographic analysis were performed as described by Pinker et al. (21). In brief, PRORP2 in buffer P was crystallised in 2 µl batch drops at 4°C by mixing (1:1 ratio) with a crystallant solution containing 200 mM sodium malonate pH 6, 20 % (w/v) PEG 3350. Diffraction data were collected at 100 K with a wavelength of 1 Å using a PILATUS 2M detector on the X06DA beamline at the Swiss Light Source (SLS synchrotron, Paul Scherrer Institute, Villigen, Switzerland). Search models for molecular replacement (MR) were derived from the structure of organellar PRORP1 (PDBid: 4G23) using MODELLER (41). MR search was performed with PHASER and the PHENIX package (42,43). Models of single PRORP2 monomer were split in three domains (PPR, ZBD, NYN) and one copy of each was placed in the P1 unit cell. As the self-correlation function indicated the presence of two-fold non-crystallographic symmetry (NCS) the second molecule was localized using a second copy of the catalytic domain. Two complete copies of PRORP2 were then refined at 3.05 Å resolution using a maximum-likelihood target as implemented in PHENIX (42). NCS constraints between the two monomers were applied during refinement, as well as TLS model for atomic displacement parameters. Manual inspection and rebuilding were performed with COOT (44). The final structure (PDBid 5FT9) describes two monomers of PRORP2 including residues 28-515. N-terminal (1-27) and C-terminal (516-536) tails, and two loops (312-321, 499-514) are not visible in the electron density map. Data collection and refinement statistics are given in Supplemental Table 1.

Isothermal titration calorimetry

ITC was performed on a MicroCal ITC instrument (Malvern Instruments, Malvern, UK), at 20°C, with 187 µM PRORP2 in buffer P with 5 mM MgCl₂. 19 samples of 3 µl protein solution were injected with a 120 sec interval in the cell containing 280 µl of 22 µM L5T0 tRNA^{Cys} precursor dialysed in the same buffer. Data analysis was performed with the microcal Origin7 software.

Microscale thermophoresis

MST was performed with a Monolith NT 115 instrument (NanoTemper Technologies GmbH, Munich, Germany). 10 μ l aliquots of a serial dilution of PRORP2 (188 μ M to 6 nM) in buffer P with 5 mM MgCl₂ and 0.3 mg/ml BSA were added to 10 μ l of a 100 nM solution of 5' Cy5 labelled L5T0 tRNA^{Cys} (IBA GmbH, Göttingen, Germany). The LED laser power was set to 20 % and the IR laser power to 40 % and the variation of fluorescence related to molecules and complexes mobility was monitored at 20°C according to manufacturer instructions. Data analysis was performed with the Monolith MO. Affinity analysis software.

Analytical ultracentrifugation

Sedimentation velocity experiments were performed in a Beckman Coulter proteomeLab XL-I instrument at 20°C and 50 000 RPM with absorbance detection. 400 μ l of different ratios of PRORP2 and L5T0 tRNA^{Cys} were prepared in buffer P with 5 mM MgCl₂. tRNA concentration was kept constant at 0.7 μ M and protein was added to final concentrations varying from 0.25 μ M to 10 μ M. Samples of 0.7 μ M tRNA alone and 10 μ M PRORP2 alone were also measured. The fitting of data was performed using SEDFIT software (www.analyticalultracentrifugation.com/) and continuous sedimentation coefficient distribution model. The distributions obtained for each sample were integrated to determine the weight-average sedimentation coefficients EPT swfast (Effective Particle Theory with weight-average s-values of the fast component of the reaction boundary) as a function of protein concentrations and to generate swfast isotherms (isotherms of the fast boundary component of rapidly interacting systems based on Gilbert-Jenkins theory (45)). The isotherms were loaded into SEDPHAT for fitting with the hetero-association model A+B \leftrightarrow AB to obtain an estimate of the K_D. In the analysis, s_A and s_B were fixed at the experimentally determined values for PRORP2 and tRNA respectively, whereas K_{AB} and s_{AB} were subject to optimization through nonlinear regression. The error surface projection analysis was exploited to determine the error intervals of the best-fit K_D values at a 95% confidence level. Buffer density, buffer viscosity and protein partial specific volumes were calculated using

SEDNTERP software. The software GUSSEI was used to plot and to integrate the sedimentation coefficient distributions, and to generate the isotherms (46).

RNase P activity assays

RNase P cleavage assays were performed with three replicates using 0.5 μ M transcript and 0.15 μ M PRORP2 protein in buffer containing 30 mM Tris-HCl pH 8, 40 mM NaCl, 4.5 mM MgCl₂, 20 μ g.ml⁻¹ BSA and 100 nM DTT for 15 min at 25°C as described previously (4). RNA fragments were separated by denaturing PAGE, visualized by ethidium bromide staining and quantified as described (4).

Small angle X-ray scattering analysis

SAXS experiments were performed on the SWING beamline at Synchrotron SOLEIL (Saint-Aubin, France) as described previously (11). In brief, samples were loaded onto a size exclusion column (Bio SEC-3 with 150 Å pore size, Agilent Technologies) and SAXS measurements were performed throughout elution. A mix of 30 μ l of 320 μ M catalytically inactive PRORP2xs and 25 μ l of 185 μ M L5T0 pre-tRNA^{Cys} in buffer P with 5 mM MgCl₂ was injected in the column equilibrated with buffer P, 5 mM MgCl₂ and supplemented with 1.5 μ M PRORP2xs in order to prevent dissociation of partners during elution. PRORP2 and PRORP2xs were also characterized alone in buffer P. Data processing, interpretation, R_g evaluation over elution profiles were performed using Foxtrot (47) and data analysis (determination of R_g, d_{max}) with the ATSAS package (48). Based on a normal mode analysis (NMA) with EINEMo (49) crystal structures of PRORP1 and 2 were perturbed to select models that better represented the SAXS profile of PRORP2 owing to a goodness-of-fit calculated with CRY SOL (50,51). Complete models of PRORP2 and PRORP2xs (including their respective N- and C-terminal extensions not described in the crystal structure) were generated under SAXS constraints using DADIMODO (52), a genetic algorithm-based refinement program. Models of PRORP2-tRNA complex were built with SASREF (53) under

SAXS constraints using this PRORP2xs solution structure with a flexible hinge between the zinc-binding and the catalytic domains, and taking into account distance restraints between above mentioned residues in PPR2-3 modules, the

catalytic domain and nucleotides in the tRNA. SASREF models were manually inspected to remove steric clashes and adjusted in PyMOL using the SAXS plugin SASpy (54).

ACKNOWLEDGEMENTS

The authors thank Dr E. Ennifar and C. Da Veiga for assistance and advice during ITC analysis, teams of SWING beamline (SOLEIL synchrotron, Saint-Aubin, France) and X06DA beamline (Swiss Light Source, Villigen, Switzerland) for beamtime allocation and support during data collection, as well as Dr M. Meyer and B. Masquida for the collection of PRORP2 diffraction data. This work was supported by the French “Centre National de la Recherche Scientifique”, by the University of Strasbourg, an ANR Blanc research grant “PRO-RNase P” (ANR-11-BSV8-008-01/02) to PG and CS, by the LabEx consortium “MitoCross” in the frame of the French National Program “Investissement d’Avenir” (ANR-11-LABX-0057_MITOCROSS). The authors also acknowledge the support and the use of resources of Instruct, a Landmark ESFRI project, and of the French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INSB-05).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

PG and CS designed and coordinated the experiments. FP, CeS, PFM, AG, CB, PR, PG and CS performed the experiments and the analysed results. PG and CS wrote the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available online.

REFERENCES

1. Altman, S. (2007) A view of RNase P. *Mol Biosyst* **3**, 604-607
2. Pinker, F., Bonnard, G., Gobert, A., Gutmann, B., Hammani, K., Sauter, C., Gegenheimer, P. A., and Giegé, P. (2013) PPR proteins shed a new light on RNase P biology. *RNA Biol* **10**, 1457-1468
3. Holzmann, J., Frank, P., Löffler, E., Bennett, K. L., Gerner, C., and Rossmann, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462-474
4. Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R. K., Rossmann, W., and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Molec Biol* **17**, 740-744
5. Lai, L. B., Vioque, A., Kirsebom, L. A., and Gopalan, V. (2010) Unexpected diversity of RNase P, an ancient tRNA processing enzyme: Challenges and prospects. *FEBS Lett* **584**, 287-296
6. Lechner, M., Rossmann, W., Hartmann, R. K., Tholken, C., Gutmann, B., Giegé, P., and Gobert, A. (2015) Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol Biol Evol* **32**, 3186-3193
7. Gutmann, B., Gobert, A., and Giegé, P. (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev* **26**, 1022-1027

8. Bonnard, G., Gobert, A., Pinker, F., Arrivé, M., Salinas, T., and Giegé, P. (2016) A single gene encodes both organelles and nuclear RNase P enzymes in *Chlamydomonas reinhardtii*. *Plant J* **87**, 270-280
9. Täschner, A., Weber, C., Buzet, A., Hartmann, R. K., Hartig, A., and Rossmannith, W. (2012) Nuclear RNase P of *Trypanosoma brucei*: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Reports* **2**, 19-25
10. Sugita, C., Komura, Y., Tanaka, K., Kometani, K., Satoh, H., and Sugita, M. (2014) Molecular characterization of three PRORP proteins in the moss *Physcomitrella patens*: nuclear PRORP protein is not essential for moss viability. *PLoS One* **9**, e108962
11. Gobert, A., Pinker, F., Fuchsbaue, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giegé, P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun* **4**, 1353
12. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A* **109**, 16149-16154
13. Reinhard, L., Sridhara, S., and Hallberg, B. M. (2015) Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res* **43**, 5664-5672
14. Li, F., Liu, X., Zhou, W., Yang, X., and Shen, Y. (2015) Auto-inhibitory Mechanism of the Human Mitochondrial RNase P Protein Complex. *Sci Rep* **5**, 9878
15. Anantharaman, V., and Aravind, L. (2006) The NYN domains: novel predicted RNases with a PIN domain-like fold. *RNA Biol* **3**, 18-27
16. Giegé, P. (2013) Pentatricopeptide repeat proteins: a set of modular RNA-specific binders massively used for organelle gene expression. *RNA Biol* **10**, 1417-1418
17. Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T., and Giegé, P. (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie* **100**, 141-150
18. Pavlova, L. V., Gossringer, M., Weber, C., Buzet, A., Rossmannith, W., and Hartmann, R. K. (2012) tRNA Processing by Protein-Only versus RNA-Based RNase P: Kinetic Analysis Reveals Mechanistic Differences. *ChemBiochem* **13**, 2270-2276
19. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**, 784-789
20. Schelcher, C., Sauter, C., and Giegé, P. (2016) Mechanistic and structural studies of protein-only RNase P compared to ribonucleoproteins reveal the two Faces of the same enzymatic activity. *Biomolecules* **6**, 30
21. Pinker, F., Giegé, P., and Sauter, C. (2015) Crystallization and crystallographic analysis of an Arabidopsis nuclear proteinaceous RNase P. *Acta Crystallogr F Struct Biol Commun* **71**, 1372-1377
22. Karasik, A., Shanmuganathan, A., Howard, M. J., Fierke, C. A., and Koutmos, M. (2016) Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5' End Processing Enzymes. *J Mol Biol* **428**, 26-40
23. Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J., Liu, X., Xu, X., Lim, B. L., Bond, C. S., Yiu, S. M., and Small, I. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J* **85**, 532-547
24. Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* **8**, e1002910
25. Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* **8**, e65343

26. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* **8**, e57286
27. Imai, T., Nakamura, T., Maeda, T., Nakayama, K., Gao, X., Nakashima, T., Kakuta, Y., and Kimura, M. (2014) Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in *Arabidopsis thaliana* play a crucial role in recognition of nucleotide bases at TpsiC loop in precursor tRNAs. *Biochem Biophys Res Commun* **450**, 1541-1546
28. Chen, T. H., Tanimoto, A., Shkriabai, N., Kvaratskhelia, M., Wysocki, V., and Gopalan, V. (2016) Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme. *Nucleic Acids Res* **44**, 5344-5355
29. Giegé, R. (2008) Toward a more complete view of tRNA biology. *Nat Struct Mol Biol* **15**, 1007-1014
30. Brillante, N., Gossringer, M., Lindenhofer, D., Toth, U., Rossmann, W., and Hartmann, R. K. (2016) Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res* **44**, 2323-2336
31. Jossinet, F., Ludwig, T. E., and Westhof, E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057-2059
32. Hauenstein, S., Zhang, C. M., Hou, Y. M., and Perona, J. J. (2004) Shape-selective RNA recognition by cysteinyl-tRNA synthetase. *Nat Struct Mol Biol* **11**, 1134-1141
33. Hopper, A. K., Pai, D. A., and Engelke, D. R. (2010) Cellular dynamics of tRNAs and their genes. *FEBS Lett* **584**, 310-317
34. Tama, F., and Sanejouand, Y. H. (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* **14**, 1-6
35. Placido, A., Sieber, F., Gobert, A., Gallerani, R., Giegé, P., and Marechal-Drouard, L. (2010) Plant mitochondria use two pathways for the biogenesis of tRNA^{His}. *Nucleic Acids Res* **38**, 7711-7717
36. Fujii, S., Suzuki, T., Giegé, P., Higashiyama, T., Koizuka, N., and Shikanai, T. (2016) The Restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in *Arabidopsis*. *Plant J* **86**, 504-513
37. Klipcan, L., Moor, N., Finarov, I., Kessler, N., Sukhanova, M., and Safro, M. G. (2012) Crystal structure of human mitochondrial PheRS complexed with tRNA(Phe) in the active "open" state. *J Mol Biol* **415**, 527-537
38. Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C., and Nyborg, J. (1996) Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure* **4**, 1141-1151
39. Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J., and Stroud, R. M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc Natl Acad Sci U S A* **100**, 12648-12653
40. Sherrer, R. L., Araiso, Y., Aldag, C., Ishitani, R., Ho, J. M., Soll, D., and Nureki, O. (2011) C-terminal domain of archaeal O-phosphoserine-tRNA kinase displays large-scale motion to bind the 7-bp D-stem of archaeal tRNA(Sec). *Nucleic Acids Res* **39**, 1034-1041
41. Fiser, A., and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461-491
42. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
43. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221

44. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. . *Acta Cryst. D* **60**, 2126–2132
45. Schuck, P. (2010) Sedimentation patterns of rapidly reversible protein interactions. *Biophys J* **98**, 2005-2013
46. Brautigam, C. A. (2015) Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol* **562**, 109-133
47. David, G., and Perez, J. (2009) Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Cryst.* **42**, 892-900
48. Konarev, P., Petoukhov, M., Volkov, V., and Svergun, D. (2006) ATSAS 2.1 : a program package for small-angle scattering data analysis. *J Appl Crystallogr* **39**, 277-286
49. Suhre, K., and Sanejouand, Y. H. (2004) Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res* **32**, W610-614
50. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**, 495-503
51. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768-773
52. Evrard, G., Mareuil, F., Bontems, F., Sizun, C., and Pérez, J. (2011) DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J. Appl. Cryst.* **44**, 1264-1271
53. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**, 1237-1250
54. Panjkovich, A., and Svergun, D. I. (2016) Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys Chem Chem Phys* **18**, 5707-5719

FIGURE LEGENDS

Figure 1. Crystal and solution structures of *A. thaliana* nuclear PRORP2.

(A) Side view in cartoon representation illustrating the overall organization of PRORP2. The enzyme is composed of two major domains, the N-terminal PPR domain made of five PPR motifs (PPR 1-5 shown in violet, red, orange, yellow and green, respectively) and the C-terminal metallonuclease domain (in blue), linked together by a zinc-binding module (in cyan).

(B) Bottom view highlighting the residues characterized by mutagenesis in this work. Corresponding side chains in PPR motifs (positions 5 and 35) are depicted as sticks, as well as the two aspartates converted in alanines in the catalytically inactive mutant.

(C) Solution model including flexible N- and C-terminal regions invisible by crystallography obtained under SAXS constraints. The theoretical SAXS profile is overlaid on experimental data. The Guinier plot is given in the inset. All molecular representations were prepared with PyMOL (version 1.7, Schrödinger, LLC).

Figure 2. Titration of PRORP2 interaction with tRNA.

(A-C) The interaction of a catalytically inactive PRORP2 with a L5T0 tRNA^{Cys} precursor having 5 nucleotides long leader sequence was monitored by (A) isothermal titration calorimetry (ITC), (B) microscale thermophoresis (MST) and (C) analytical ultracentrifugation (AUC). In the ITC plot, the slope of the tangent indicates the affinity constant. MST experiments were performed with 52 nM of Cy5 fluorescent RNA. For the MST plot, the normalized fluorescence (Thermophoresis) of all MST traces are plotted against the concentration of PRORP2 in nM shown on a log scale. Red, blue and grey indicate three replicate experiments. In the AUC plot, different colours indicate the particular concentrations of PRORP2 and / or RNA in the respective AUC experiments. The star indicates signals for PRORP2 alone, the cross for RNA alone and the arrow for PRORP2-pre-tRNA^{Cys} complexes. K_D values that could be derived from the three biophysical approaches are indicated in the respective panels.

Figure 3. PRORP2 PPR motifs 2 and 3 are required for pre-tRNA substrate binding.

(A) RNase P *in vitro* cleavage assays were performed with Arabidopsis wild-type PRORP2 (WT) as well as with double mutants modified at positions 5 and 35 of the respective PPR motifs (PPR1-PPR5) and tRNA^{Cys} precursors with either 5 nucleotides leader sequences (upper panel) or with 51 nucleotides leader sequences (lower panel), – indicate reactions with RNA alone. PRORP cleavage products were separated on 17 % denaturing PAGE and quantified with ImageJ. Numbers indicate percentages of cleavage as defined by Gobert et al. 2013. Values were normalized so that 100 correspond to cleavage observed for wild-type PRORP2. The molecular weights of markers (M) are given in ribonucleotides.

(B) Analytical ultracentrifugation sedimentation plots of catalytically inactive WT PRORP2 and PPR mutants in complex with tRNA^{Cys} precursors having 5 nucleotides long leader sequences. Different colours indicate the particular concentrations of PRORP2 isoforms and / or RNA in the respective experiments. Red dotted lines indicate signals for PRORP2 alone, blue dotted lines for RNA alone and purple dotted lines for PRORP2-tRNA^{Cys} complexes. K_D values that could be derived for the different complexes are indicated in the respective panels.

Figure 4. SAXS based model of PRORP2 in complex with a pre-tRNA substrate.

(A) SEC elution profiles (Bio SEC-3/150Å HPLC column) obtained with L5T0 pre-tRNA (green chromatogram, injection: 0.2 nmole, peak elution time: 12.5 min), PRORP2xs (blue, injection: 0.9 nmole, peak elution time: 11.9 min), the complex formed during an ITC titration (orange, injection: 20 µl of ITC mix, peak elution time: 12.0 min) or a 2:1 PRORP2xs:L5T0 mix (red, injection: 0.6:0.3 nmoles, peak elution times: 11.9, 12.3 min). This illustrates the difficulty to stabilize the complex and to get it separated from individual partners.

(B) Evolution of X-ray scattering intensity $I(0)$ (blue plot) and gyration radius R_g (red plot) along the SEC elution of the complex in a mobile phase supplemented with PRORP2xs to minimize the dissociation of the complex. A dataset of 240 SAXS images was collected in the 10-18 min window symbolized by the blue arrow. SAXS images (frames 112-130) were averaged in the region indicated by the green bar where $R_g \sim 34$ Å, leading to the profile shown in (C).

(C) Best solution model of the complex obtained under distance (PPR2 and 3 interacting with C56 and G18, respectively) and SAXS constraints. The theoretical SAXS profile (green curve) is overlaid on experimental data (blue dots). The Guinier plot is given in the inset. Alternative possibilities leading to higher Chi values are shown in Supplemental Figure 11. The zoom on the right highlights the proximity of residues 5 and 35 of PPR motifs 2 and 3 (analysed by mutagenesis) with tRNA residues C56 and G18 at the corner of the tRNA (see open blue arrow).

Figure 5. Domain reorientation in PRORP2.

Comparison of structures of PRORP2 in solution, in the crystal and in solution in complex with a pre-tRNA substrate (Left, Middle, Right) highlights the movement of the catalytic domain with respect to the PPR domain and the zinc-binding domain. The three models have been superimposed according to the latter two domains and are shown in two perpendicular views: from the side and from the inner region of the Λ shape facing the RNA substrate. In solution the catalytic domain in the free and complexed with PRORP2 undergoes a 52° and 23° rotation compared to its position in the crystal structure, illustrating the flexibility of the hinge region and the structural plasticity of PRORP2. The characterization of PRORP1 in solution (see Supplemental Figure 4) shows a similar behaviour and suggests that this may be a general property of PRORP enzymes.

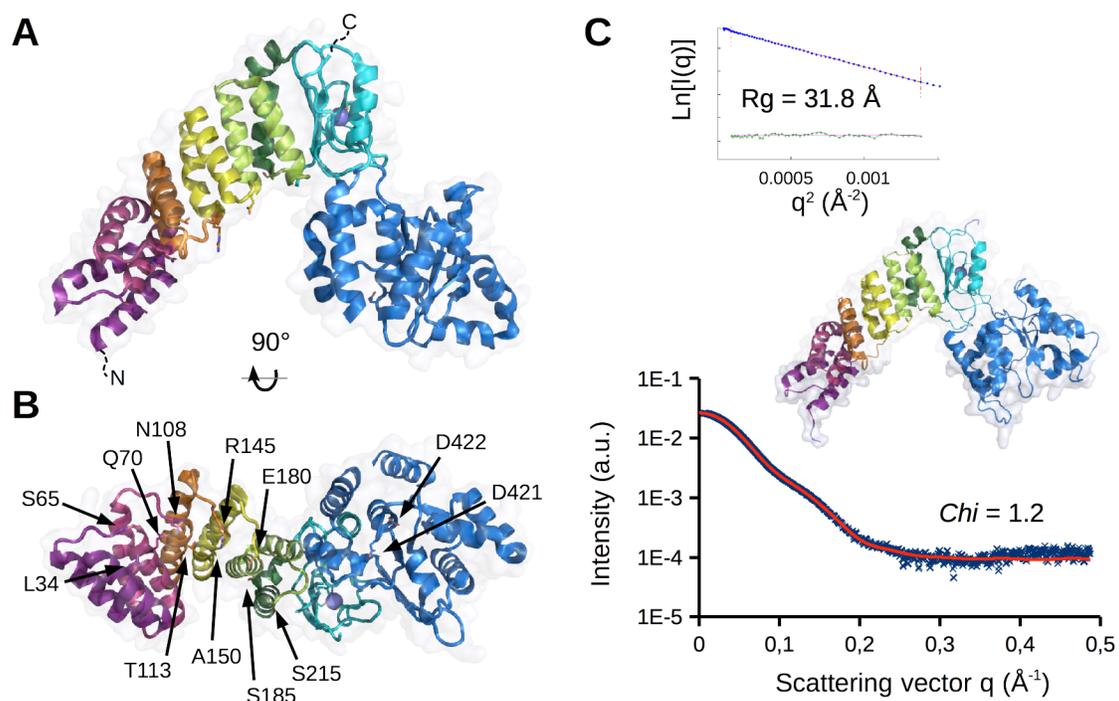


Figure 1

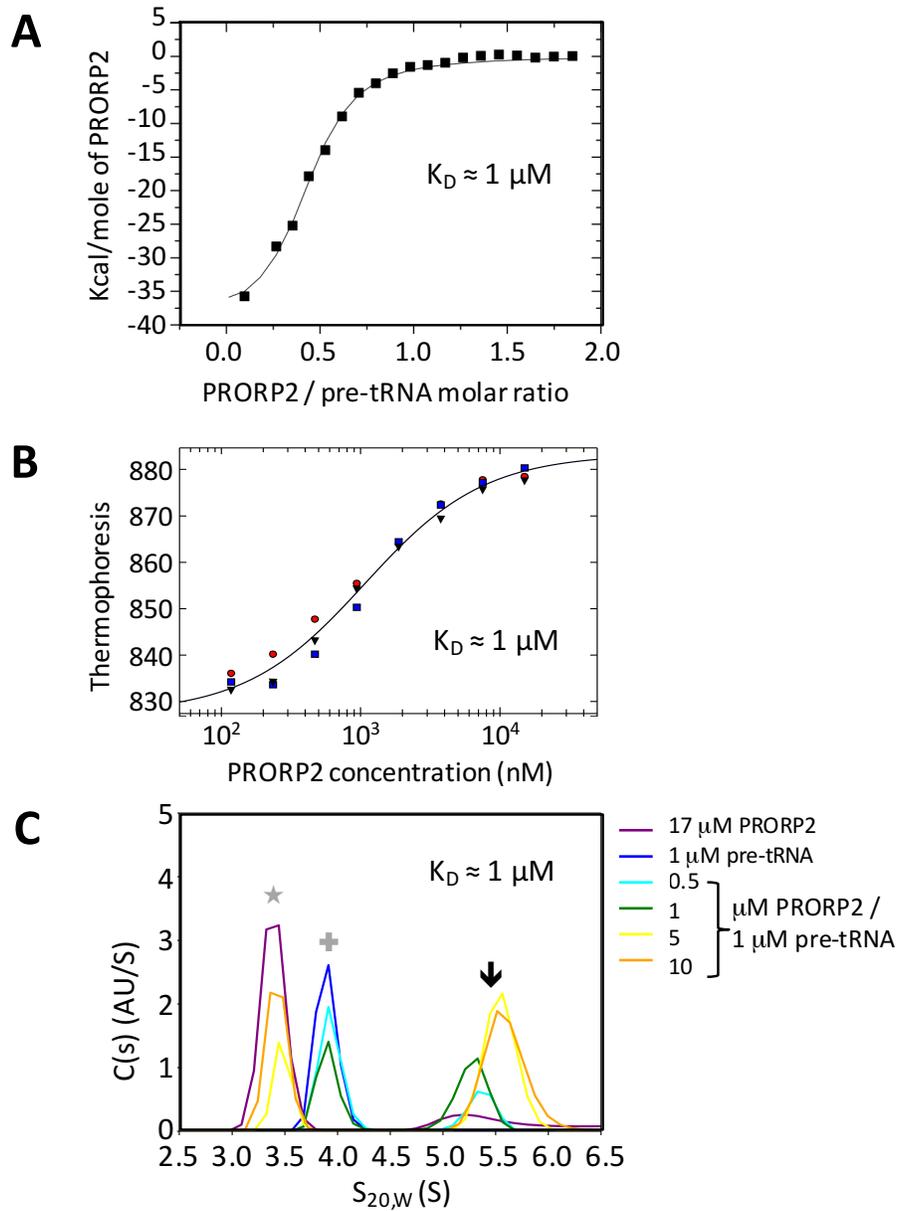


Figure 2

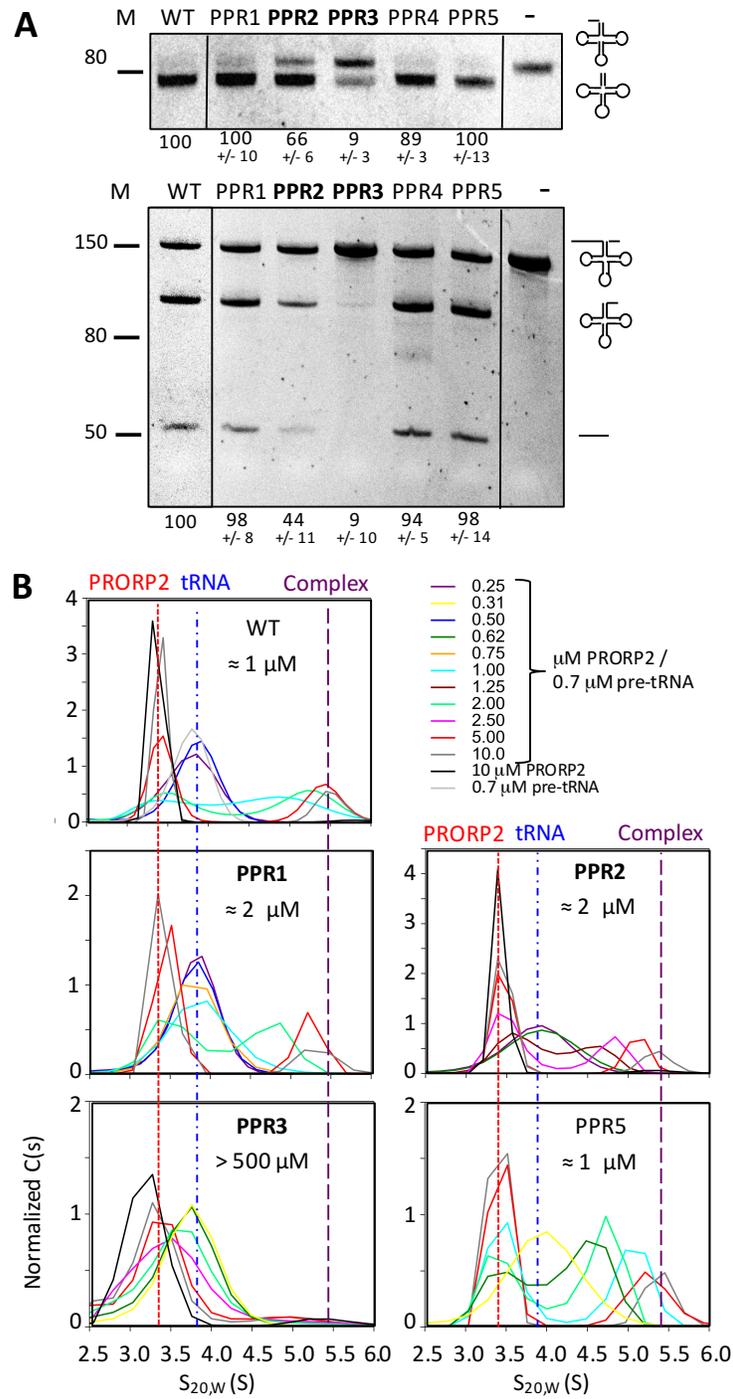


Figure 3

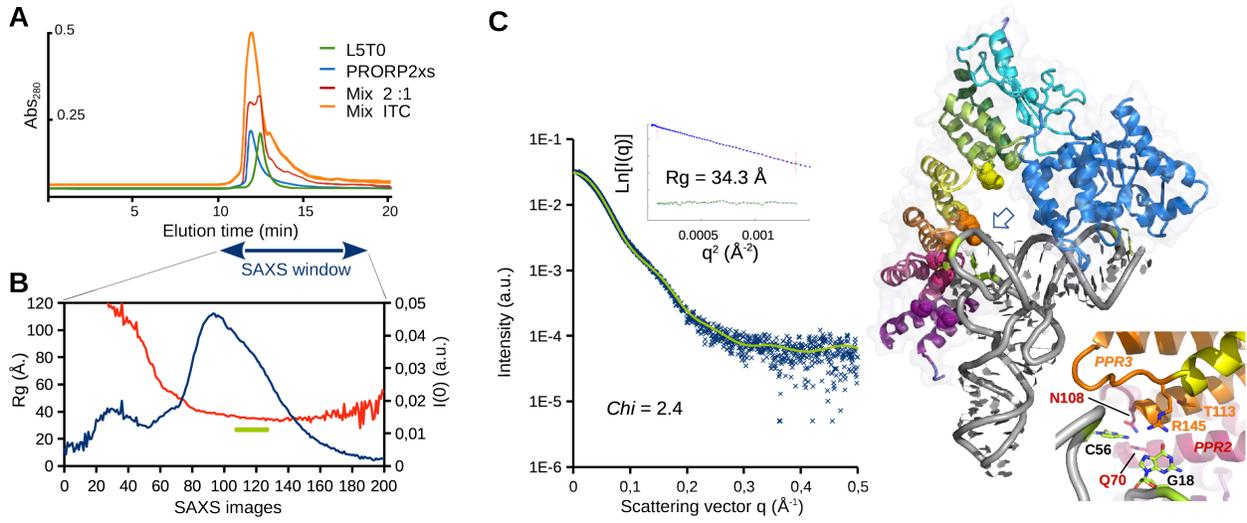


Figure 4

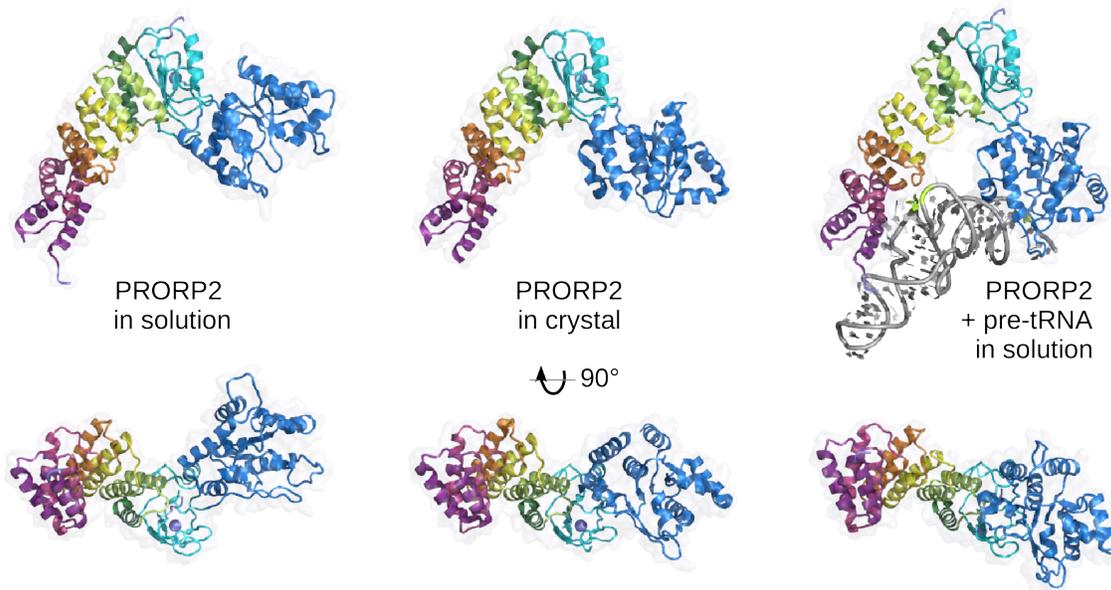


Figure 5

Supplementary information

Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding

Franziska Pinker^{1,2,+}, Cédric Schelcher^{2,+}, Pablo Fernandez-Millan¹, Anthony Gobert², Catherine Birck³, Aurélien Thureau⁴, Pierre Roblin^{4,5}, Philippe Giegé^{2,*} and Claude Sauter^{1,*}

¹ UPR 9002 du CNRS, Architecture et Réactivité de l'ARN, IBMC, Université de Strasbourg, 15 rue René Descartes, Strasbourg, F-67084, France

² UPR 2357 du CNRS, IBMP, Université de Strasbourg, 12 rue du général Zimmer, F-67084 Strasbourg, France

³ UMR 7104 du CNRS, IGBMC, 1 rue Laurent Fries, Illkirch, F-67404, France

⁴ Synchrotron SOLEIL, l'Orme des Merisiers Saint Aubin, F-91410, France

⁵ URBIANantes, INRA Centre de Nantes, 60 rue de la Géraudière, Nantes, F-44316, France

+ contributed equally

* authors for correspondence: philippe.giege@ibmp-cnrs.unistra.fr - c.sauter@ibmc-cnrs.unistra.fr

Table S1. Crystallographic analysis of PRORP2.

Fig. S1. Crystal structure of PRORP2 in triclinic environment.

Fig. S2. PRORP2 activity in the presence of divalent cations.

Fig. S3. SAXS analysis of full-length (PRORP2) and short (PRORP2xs) enzyme alone and in complex with pre-tRNA L5T0.

Fig. S4. PRORP crystal vs solution structures.

Fig. S5. PRORP2 modeling under SAXS restraints.

Fig. S6. Effect of salt concentration on PRORP2 activity.

Fig. S7. Arabidopsis pre-tRNA substrates.

Fig. S8. Positions of residues 5 and 35 of PPR motifs in Arabidopsis PRORP2 and in other plant group III PRORP proteins.

Fig. S9. Alignement of Arabidopsis PRORP2 with 34 members of plant group III PRORP.

Fig. S10. Structural characterization of full-length PRORP2 constructs in solution.

Fig. S11. Modelling of PRORP2:tRNA complex under SAXS and geometrical constraints.

Table S1. Crystallographic analysis of PRORP2

<i>Data collection statistics</i>	
Beamline	X06DA, SLS
Space group	<i>P1</i>
Unit cell lengths a, b, c (Å)	70.5, 72.8, 80.3
Unit cell angles alpha, beta, gamma (°)	63.1, 72.3, 78.4
Resolution range (Å)	35 – 3.05
Highest resolution shell (Å)	3.23 – 3.05
No. of observations	84332 (23923)
No. of unique reflections	14920 (4065)
Completeness (%)	92.4 (98.2)
Multiplicity	3.5 (3.7)
Rmerge / Rmeas (%)†	11.6 (163) / 13.7 (191)
CC1/2 (%)	99.8 (43.2)
$\langle I/\sigma(I) \rangle$	8.7 (0.8)
Unit cell content	2 monomers
Solvent content (%)	57.4
 <i>Refined atomic structure</i>	
Resolution range (Å)	35 – 3.05
R-factor / R-free (%)	24.2 / 28.9
Number of protein atoms, zinc ions	7442 / 2
Average atomic displacement parameters for proteins atoms, zinc ions (Å ²)	145.9 / 109.6
Rmsd and bonds (Å) and angles (°)	0.005 / 1.15
Ramachandran plot £	
residues in most favored, unfavored regions (%)	95.8 / 0.1
PDBid	5TF9

† $R_{\text{merge}} = \frac{\sum hkl \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum_i I_i(hkl)}$ and redundancy-independent

$R_{\text{meas}} = \frac{\sum hkl (n/n-1)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum_i I_i(hkl)}$.

£ Analysis carried out with Molprobit in Phenix package.

Values in parentheses are for the highest resolution shell.

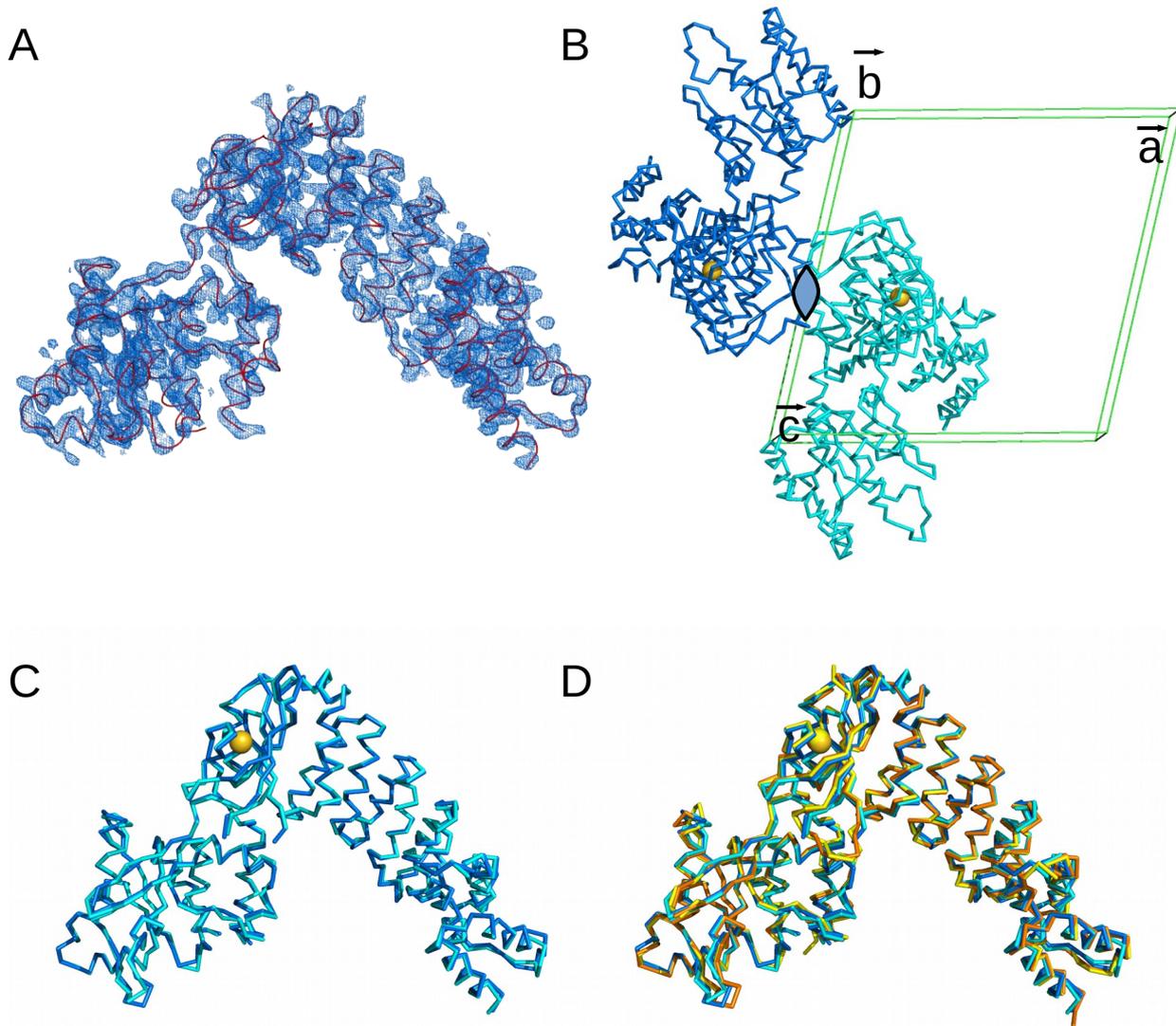


Figure S1. Crystal structure of PRORP2 in triclinic environment. A) Model of PRORP2 monomer with its characteristic Λ shape in the electron density map (resolution: 3.05 Å; contour level: 1.5 sigma). B) Independent monomers of PRORP2 in the triclinic unit cell. The two-fold non-crystallographic relating the two copies is in the direction of b-axis, and perpendicular to the figure and to the (a,c) plane, as detected in the self-correlation function (Pinker et al. 2015). C) Superposition of monomers A and B (cyan, marine, respectively; rmsd = 0.85 Å) of the current structure, and D) superposition with monomers A and B (yellow and orange, respectively; rmsd = 0.62 Å) of the triclinic structure independently determined by Karzic et al (2016). The monomers are very similar with an average rmsd between monomers from either structure of 1.32 +/- 0.09 Å. All rmsd were calculated for a total of 484 superimposed residues.

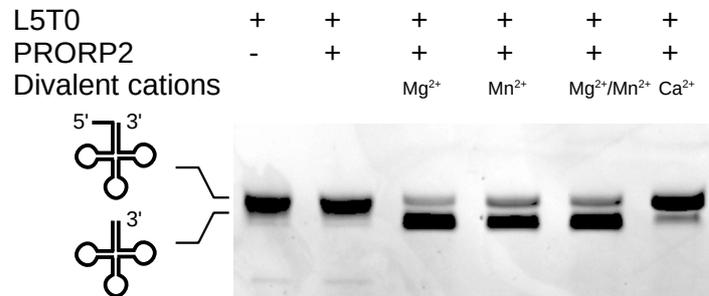
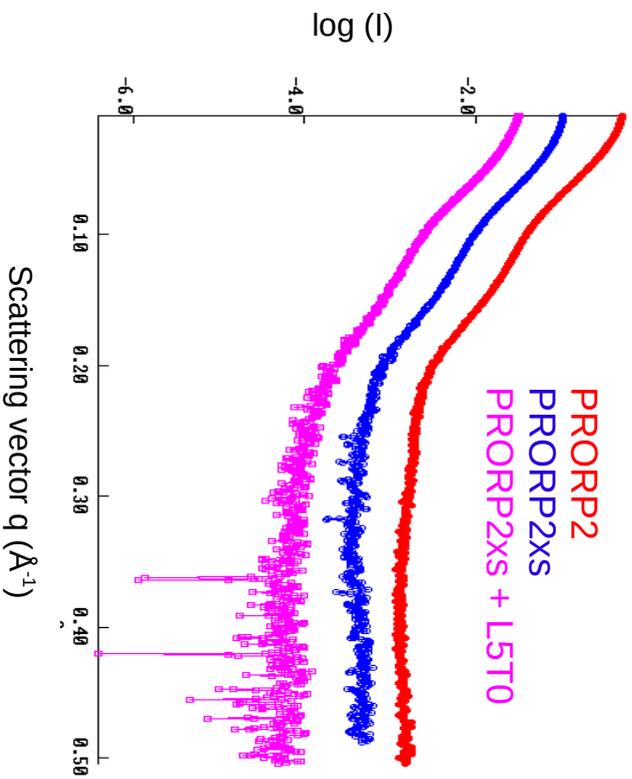
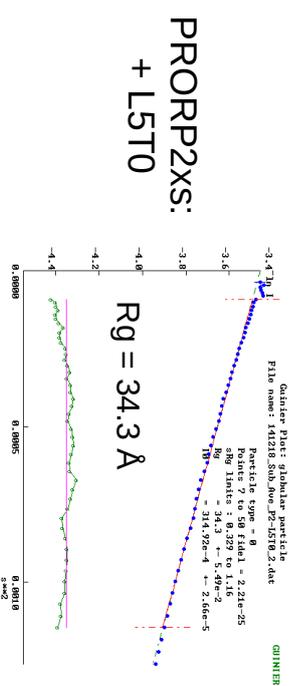
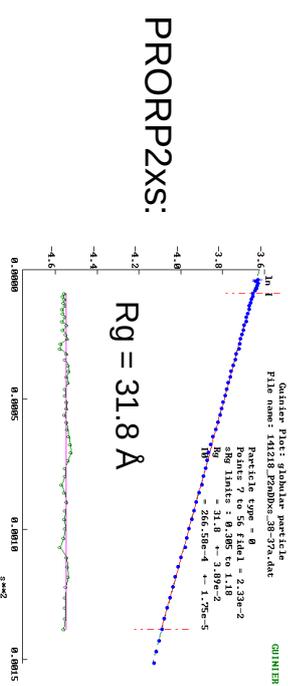
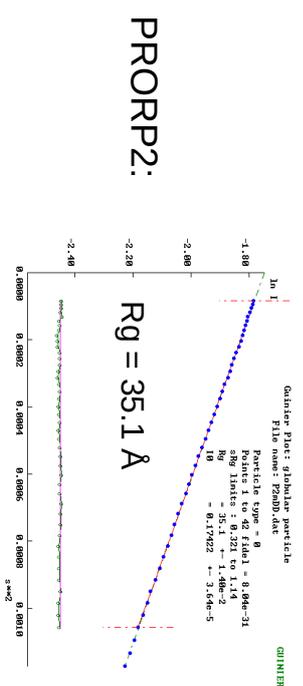


Fig. S2. PRORP2 activity in the presence of divalent cations. A short L5T0 substrate used for biophysical analyses was processed by WT PRORP2 in the presence of different metal ions. WT PRORP2 was extensively dialyzed against a buffer without divalent ions. The first and second lanes show an intact substrate in the absence or in the presence of PRORP2. Each assay contained 200 ng of L5T0 (10 pmoles) and was processed in the presence of 250 ng of enzyme (4 pmoles) and 5 mM of divalent ions (2.5 mM each in the case of a Mg²⁺/Mn²⁺ mix). Both magnesium and manganese can be used by PRORP2 for cleavage, while calcium is much less efficient, although a residual activity is observed.

SAXS profiles



Guinier plots



P(r)

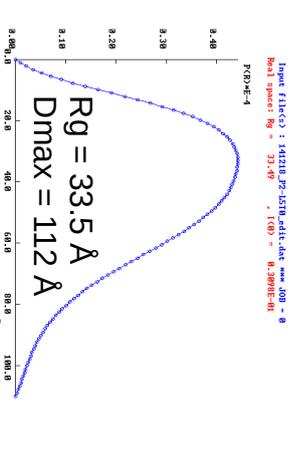
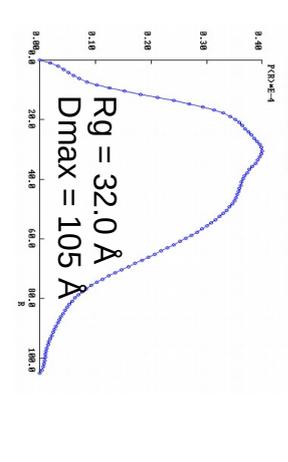
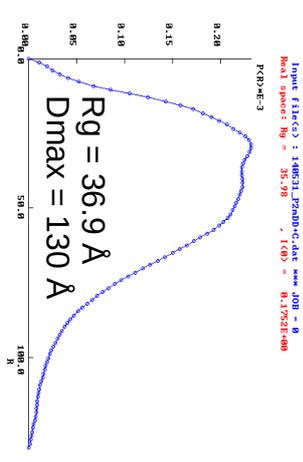


Fig. S3. SAXS analysis of full-length (PRORP2) and short (PRORP2xs) enzyme alone and in complex with pre-tRNA L5T0. Experimental SAXS profiles (Right), Guinier plots (Center) and $P(r)$ distance distribution functions (Left) for each entry isolated by SEC.

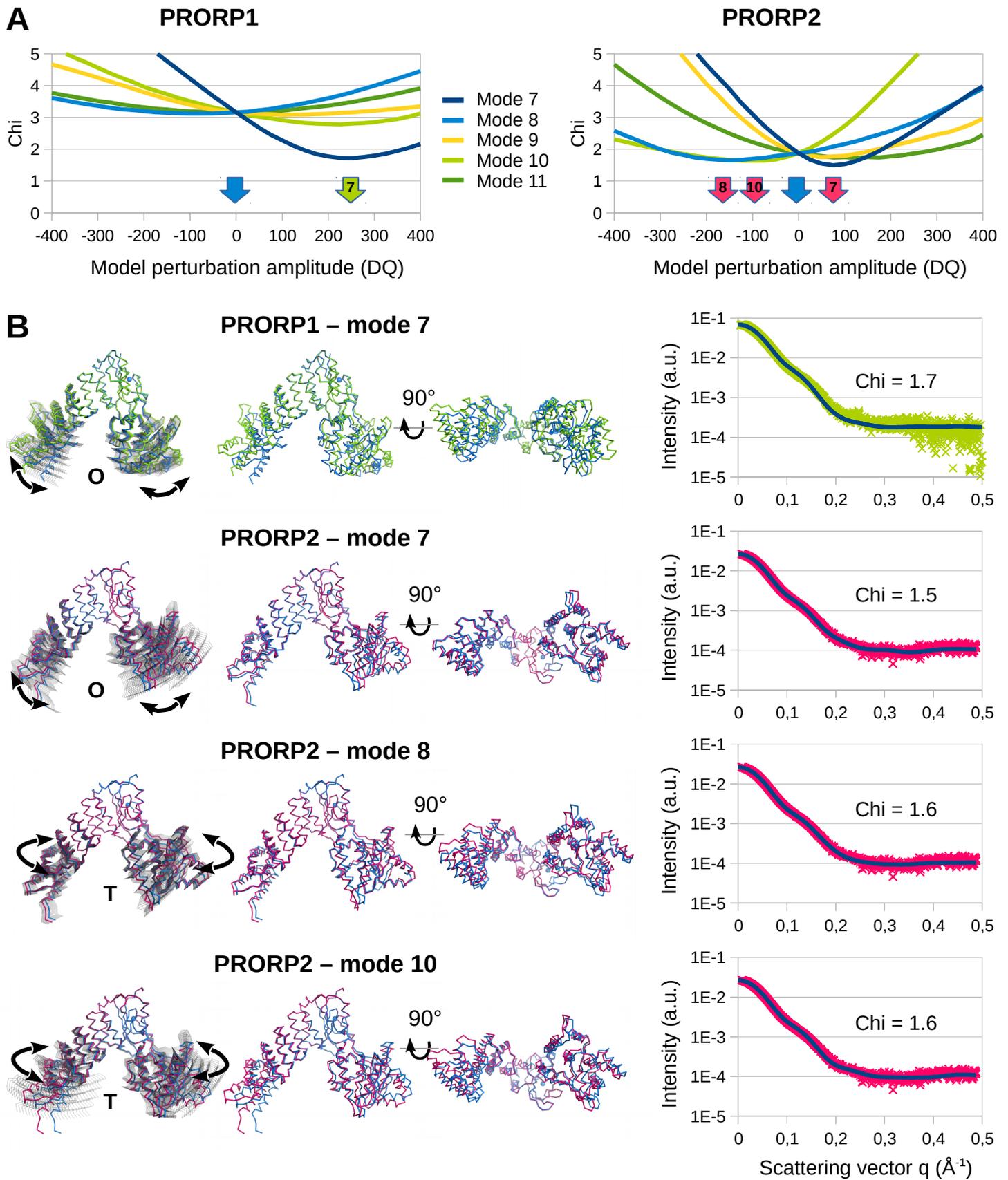


Fig. S4. PRORP crystal vs solution structures. (A) Normal mode analyses were performed with Elnemo (Suhré & Sanejouand, 2004) to produce models fitting SAXS data. Graphs at the top indicate the goodness-of-fit (Chi) determined with Crysol (Svergun et al., 1995) for models generated by the first low-frequency modes (7-11). Blue arrows and models correspond to the starting crystal structure, green/red ones to the best model in each mode for PRORP1/PRORP2, respectively. (B) Ensemble of normal mode perturbed models are shown in grey (Left), as well as superimposed best model and crystal structure in two orientations (Middle) and corresponding experimental (green/red crosses) and theoretical (blue curve) SAXS profiles (Right). All models are superimposed according to their central zinc-binding module. Experimental data for PRORP1 are from Gobert et al. (2013). Mode 7 corresponding to the opening of PRORP1 Λ -shape gives the best fit. In the case of PRORP2, which shows already a more open crystal structure, the perturbation (DQ) is less important and consists in a further opening (O) and twisting (T) of the enzyme backbone along modes 7, 8 and 10.

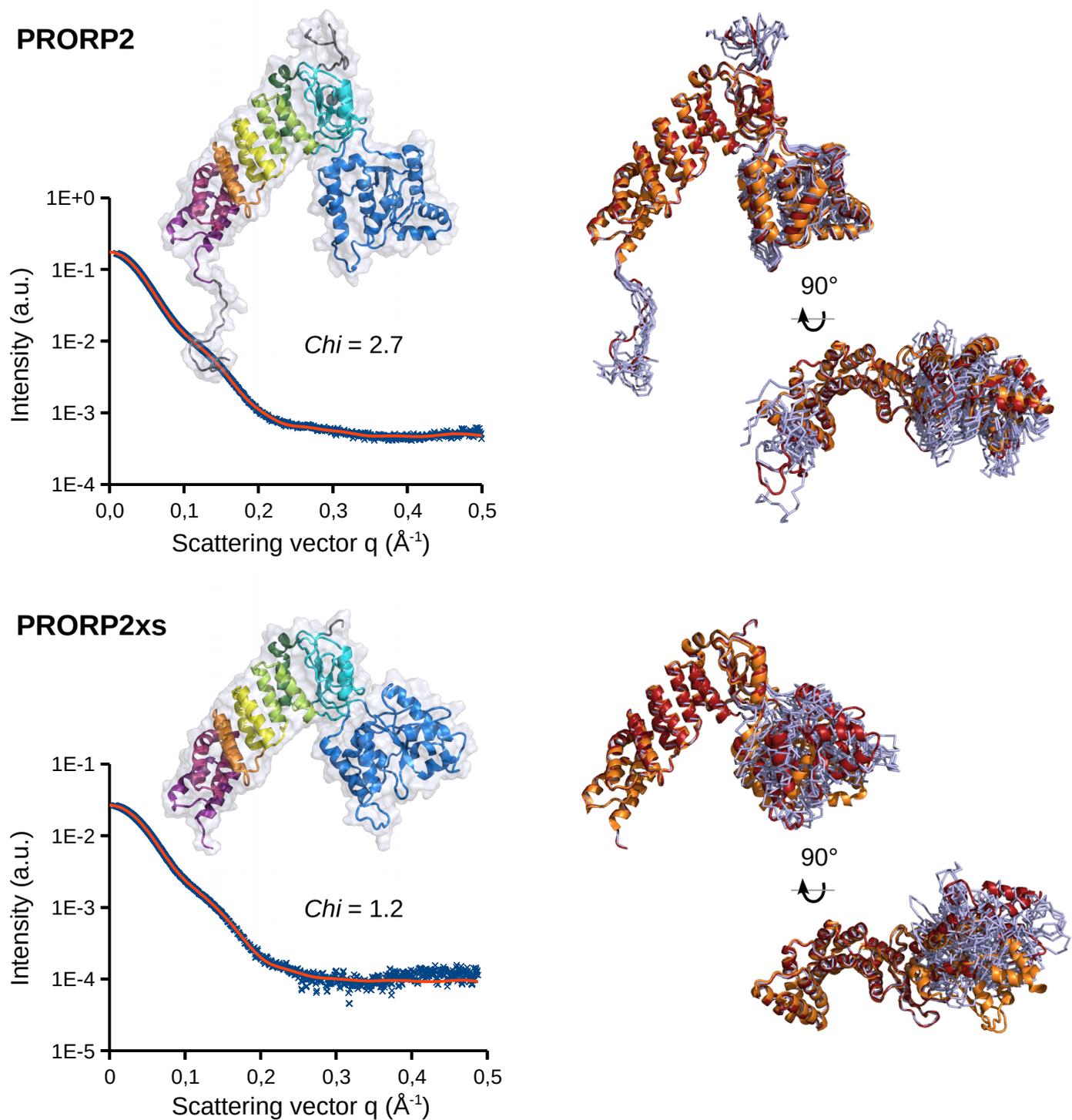


Fig. S5. PRORP2 modeling under SAXS restraints. Full atom models were generated using Modeller (Webb & Sali, 2014) for the full-length (PRORP2 – Top) and short enzyme form (PRORP2xs – Bottom) based on the crystal structure. DADIMODO (Evrard et al., 2011) was used to refine the models under SAXS restraints. A hinge was defined between the connecting zinc-binding domain and the catalytic domain (residues 290-291, 480-481) for rigid-body sampling followed by simulated annealing on the N- and C-terminal tails and hinge region. (Left) Best models and corresponding theoretical profiles superimposed to experimental data. (Right) Ensemble of seven best models selected by the genetic algorithm superimposed to the crystal structure (orange). The best model is shown in red. Goodness-of-fit (Chi) to SAXS data of these models vary from 1.13 to 1.25 and 2.71 to 2.94 and the rotation of the catalytic domain with respect to its position in the X-ray structure ranges from 39° to 156° and 8° to 36° for PRORP2xs and PRORP2, respectively. This suggests an extreme flexibility of the hinge region allowing the reorientation of the catalytic with respect to the PPR domain.

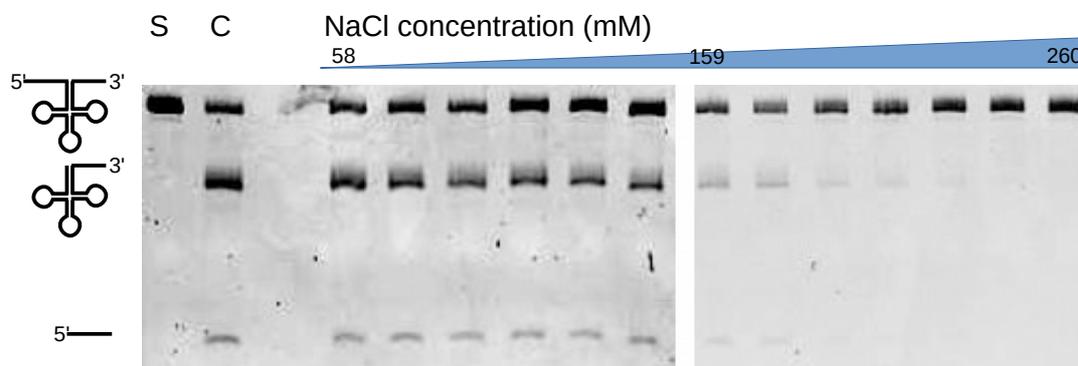


Fig. S6. Effect of salt concentration on PRORP2 activity. The L51T30 substrate was processed by WT PRORP2 in various buffer conditions. The first line corresponds to the unprocessed substrate (S), the second to a positive control (C) performed in a standard buffer containing 20 mM Tris-HCl pH 8, 40 mM NaCl, 4.5 mM MgCl₂, 2 mM DTT and 20 µg/ml BSA. Increasing concentrations of monovalent ions (left to right, 58 to 260 mM NaCl with an increment of 17 mM) were tested in a buffer containing 50 mM HEPES-Na pH 7.5, 5 mM MgCl₂, 2.5 mM DTT, 5 % (w/v) glycerol, 20 µg/ml BSA and 4 U RNase OUT (RNase A-C inhibitor, Invitrogen). Each assay (V = 10 µl) contained 1 µM L51T30, 0.16 µM WT-PRORP2 (except S) and was incubated 30 min at 20°C. The products were analyzed on a denaturing gel (8 M urea / 8 % polyacrylamide gel). The activity of PRORP2 drops above circ. 150 mM NaCl. This concentration was chosen for biophysical analyses as a good compromise to ensure PRORP2 stability and activity.

A

At-tRNA-Cys-L5T0 (76 nuc.)
GGGUU
 GGCUAGGUAACAUAUAUGGAAAUGUAUCGGACUGCAA
 AUCCUGUAAUGACGGUUCGACUCCGUCCUUGGCCU

At-tRNA-Cys-L51T30 (147 nuc.)
GAGAGGAAGAAAGAACAACCGUUUUACUUUGGCACA
UGAGGUGGCGGGUUU
 GGCUAGGUAACAUAUAUGGAAAUGUAUCGGACUGCAA
 AUCCUGUAAUGACGGUUCGACUCCGUCCUUGGCCU
ACACCUUCAUGACCAGAAAUAACUG

B

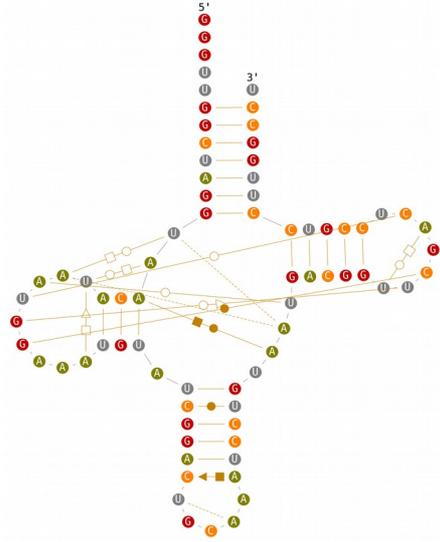


Fig. S7. Arabidopsis pre-tRNA substrates. A) Sequences of PRORP2 substrates derive from Arabidopsis mitochondrial tRNA-Cys(GCA). LxTy stands for a tRNA with a leader (L) of 'x' nucleotides and a trailer (T) of 'y' nucleotides. L and T sequences are shown in italic. B) Cloverleaf 2D structure of L5T0 substrate. Potential tertiary interactions are deduced from a 3D modelling performed with Assemble2 based on the *E. coli* tRNA-Cys crystal structure (PDBID: 1U0B) and are depicted according to the Leontis-Westhof nomenclature (Leontis et al, NAR, 2002).

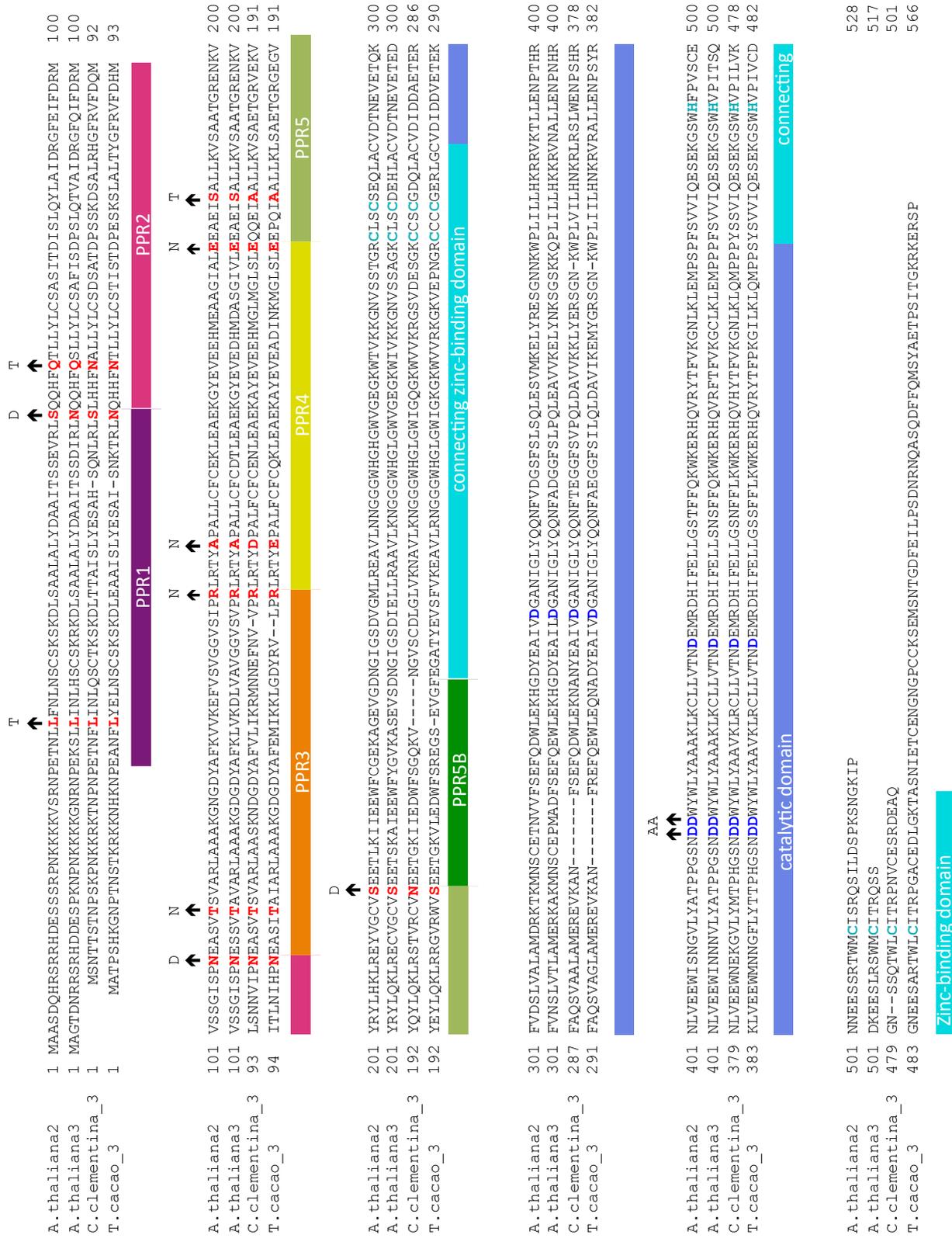


Fig. S8. Positions of residues 5 and 35 of PPR motifs in Arabidopsis PRORP2 and in other plant group III PRORP proteins. The positions of PPR motifs are highlighted with coloured rectangles. Residues at positions 5 and 35 are indicated in red. Black arrows indicate the nature of mutations performed. The catalytic domain is underlined with a blue bar and four conserved aspartates participating in the NVN signature of the domain are indicated in dark blue. Aspartates 421 and 422 were mutated to obtain catalytically inactive PRORP2 proteins. The connecting domain is represented by cyan bars and three conserved cysteines and one histidine responsible for Zn coordination are indicated in cyan.

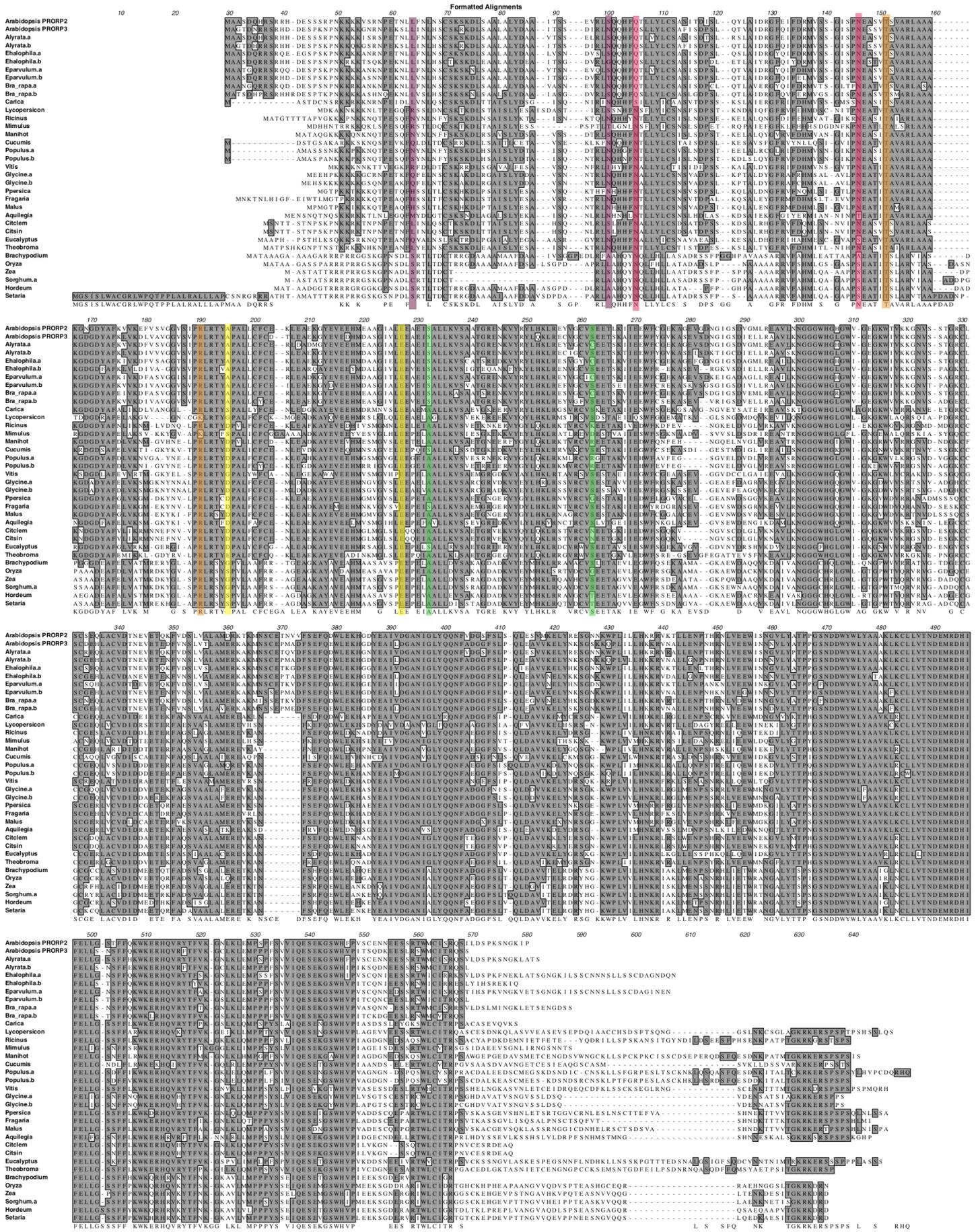


Fig. S9. Alignment of Arabidopsis PRORP2 with 34 members of plant group III PRORP. Positions 5 and 35 of motifs PPR 1, 2, 3, 4, 5 are highlighted in violet, red, orange, yellow and green, respectively.

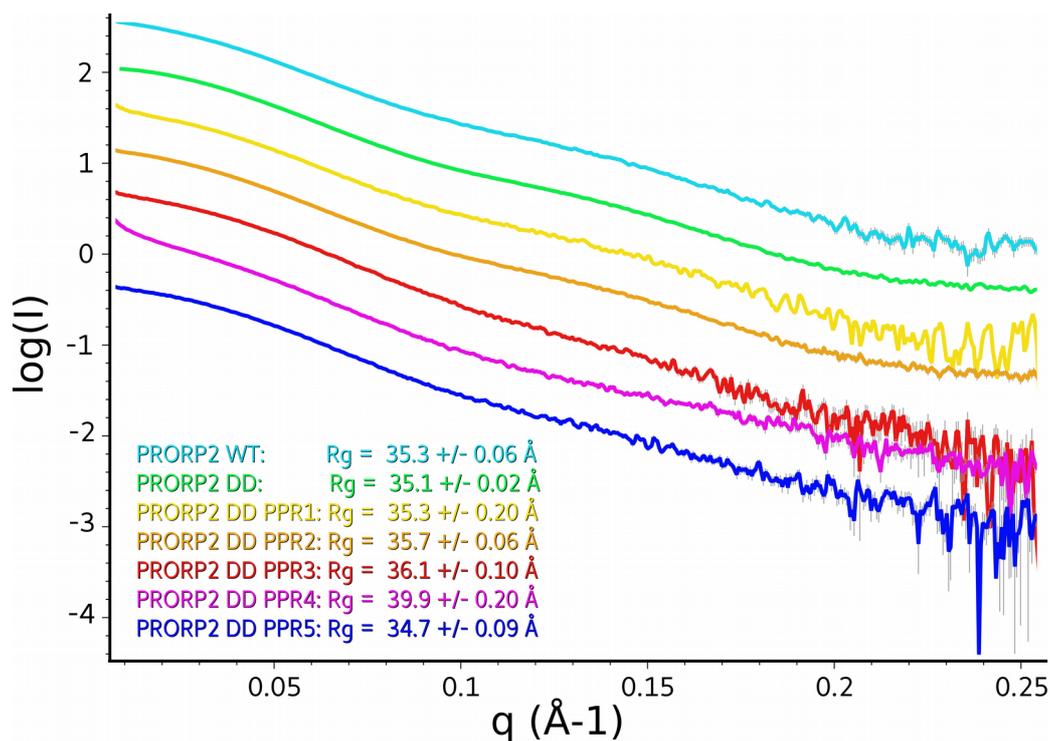


Fig. S10. Structural characterization of full-length PRORP2 constructs (wild-type and mutants) in solution. The five variants mutated in PPR motifs were analyzed by SAXS using the same protocols as for the wild-type enzyme and its catalytically inactive version. SAXS curves are displayed with an offset for clarity. The similarity of their shapes in the q range 0-0.15 \AA^{-1} confirms the structural integrity of the PPR variants and that mutant and wild-type enzymes fold alike. R_g values derived from Guinier plots are also very close. The only outlier is the PPR4 mutant who showed by a strong tendency to aggregate as highlighted by the rising slope of the curve at short angles ($q < 0.015 \text{\AA}^{-1}$) and a increased R_g . This mutant was therefore used in activity assays only, but excluded from AUC analyses.

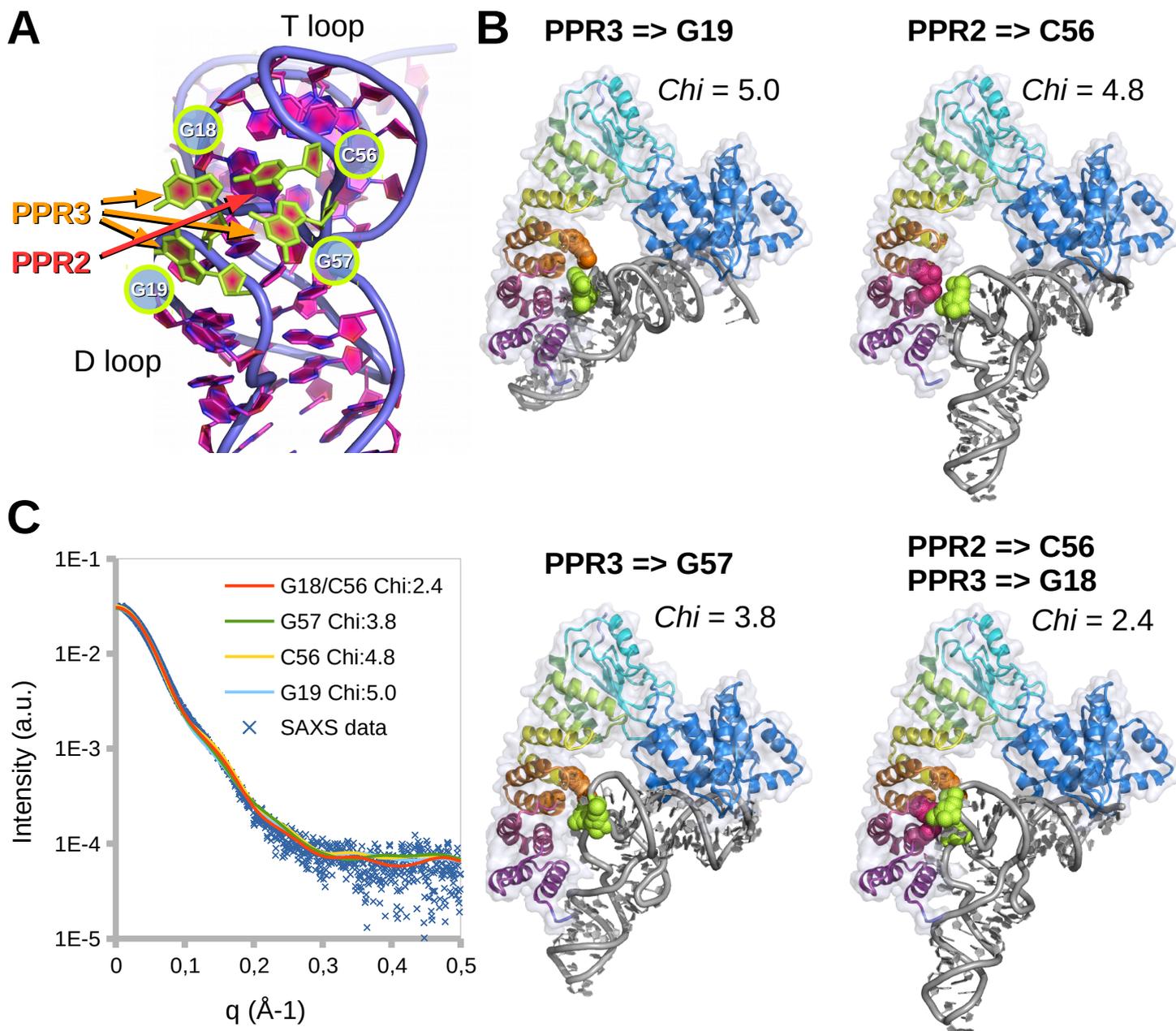


Fig. S11. Modelling of PRORP2:tRNA complex under SAXS and geometrical constraints. A) In this work we showed by mutagenesis and AUC that PPR motifs 2 and 3 are the most important for substrate recognition. Mutagenesis and footprint experiments on the precursor already identified the corner of the tRNA L shape as the main contact point with PRORP together with the end of the acceptor stem (Gobert et al, 2013). The D and T loops contain conserved residues (G18, G19, C56, R57) that can serve as recognition elements for PPR motifs. B) Accordingly, distance constraints between these residues and positions 5 and 35 in PPR 2 (Q70, N108) and 3 (T113, R145), and between the phosphate group of residue 1 and catalytic aspartates in the NYN domain (420, 421) were used to dock the tRNA onto PRORP using SASREF. Resulting models were then manually adjusted to avoid steric clashes using the SASpy (Panjkovich & Svergun, 2016) extension in PyMOL (Schrödinger, LLC). Four possibilities derived from the PPR recognition code were tested: PPR2 interacting with C56, PPR3 with G19 or G57, and a combination of PPR2:C56, PPR3:G18 interactions. The molecular volume of involved residues in PRORP and pre-tRNA are highlighted and the goodness-of-fit (Chi) to SAXS data is indicated next to each model. C) Corresponding theoretical SAXS are superimposed to experimental data (blue plot). The model that best fits the data (bottom right) involves PPR2 and 3 recognising elements in both D and T loops (see details in Fig. 4).

References

- Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giegé, P. (2013). Structural insights into protein-only RNase P complexed with tRNA. *Nature Communications* 4, 1353.
- Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA* 7, 499.
- Panjikovich, A., and Svergun, D.I. (2016). SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* btw 071.
- Pinker, F., Giegé, P., and Sauter, C. (2015). Crystallization and crystallographic analysis of an *Arabidopsis* nuclear proteinaceous RNase P. *Acta Crystallographica Section F Structural Biology Communications* 71, 1372–1377.
- Suhre, K., and Sanejouand, Y.-H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research* 32, W610–W614.
- Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* 28, 768–773.
- Webb, B., and Sali, A. (2014). Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* 47, 5.6.1-32.

Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding

Franziska Pinker, Cédric Schelcher, Pablo Fernández-Millán, Anthony Gobert, Catherine Birck, Aurélien Thureau, Pierre Roblin, Philippe Giegé and Claude Sauter

J. Biol. Chem. published online July 10, 2017

Access the most updated version of this article at doi: [10.1074/jbc.M117.782078](https://doi.org/10.1074/jbc.M117.782078)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2017/07/10/M117.782078.DC1>

This article cites 0 references, 0 of which can be accessed free at

<http://www.jbc.org/content/early/2017/07/10/jbc.M117.782078.full.html#ref-list-1>

VI. Résultats chapitre 2 : Etude de la diversité des substrats ARN de AtPRORP1

Chez *Arabidopsis*, AtPRORP1 est la RNase P présente dans la mitochondrie et le chloroplaste. Il a été démontré par l'équipe qu'elle est capable de cliver une structure ARN-like (TLS) présente dans la partie 3' de l'ARNm de Nad6 (Gutmann, 2012). Une autre étude suggère que AtPRORP1 participe au clivage de l'orf291 dans les mitochondries d'*Arabidopsis* (Fujii et al 2016). PRORP a donc la capacité de maturer d'autres ARN que les ARNt. Mon travail de thèse a donc eu pour but d'étudier la diversité de substrat de AtPRORP1 à l'échelle du transcriptome mitochondrial entier. Cette étude a aussi eu pour but de prouver la faisabilité d'une telle approche afin d'identifier dans un second temps l'ensemble des substrats des protéines PRORP nucléaires, AtPRORP2 et AtPRORP3.

Pour ce faire, une approche basée sur l'immunoprécipitation de complexes ARN / protéines suivie par le séquençage à haut débit des ARN a été utilisée (approche « RIP-Seq »). Des plantes exprimant une version de AtPRORP1 portant une étiquette HA étaient déjà disponibles au laboratoire. Les protéines AtPRORP1 ont été co-immunoprécipitées avec leurs substrats ARN en utilisant des anticorps reconnaissant l'étiquette HA. Le génome mitochondrial étant moins complexe que le génome nucléaire, avec une centaine de gènes codés dans 366 kb, on s'attend à un nombre de substrats de AtPRORP1 relativement limité. Ainsi le séquençage a pu être réalisé avec un séquenceur à « moyen débit » Illumina MiSeq présent à l'institut. Cette étude pilote à petite échelle permet aussi la mise en place des protocoles bioinformatiques pour l'analyse des données de RIP-Seq. L'analyse a été réalisée en priorité pour le transcriptome mitochondrial (i) en raison de l'expertise de notre équipe dans l'étude des processus d'expression génétique dans les mitochondries et (ii), car les substrats chloroplastiques de AtPRORP1 ont déjà été partiellement identifiés dans une étude utilisant des mutants de dé-régulation de AtPRORP1 (Zhou, 2015).

1. Co-immunoprécipitation de AtPRORP1 avec ses substrats ARN

Les plantes AtPRORP1-HA ont été construites par Ayoub Bouchoucha (Bouchoucha et al. en préparation) en transformant des mutants d'insertion AtPRORP1 hétérozygotes avec un plasmide binaire comportant AtPRORP1 sous le contrôle du promoteur endogène de AtPRORP1 et fusionné avec une étiquette HA dans sa partie C-terminale. Les plantes sélectionnées comportant le transgène AtPRORP1-HA et hétérozygotes pour la mutation du gène *prop1* endogène sont autofécondées pour obtenir des plantes *prop1/prop1* possédant

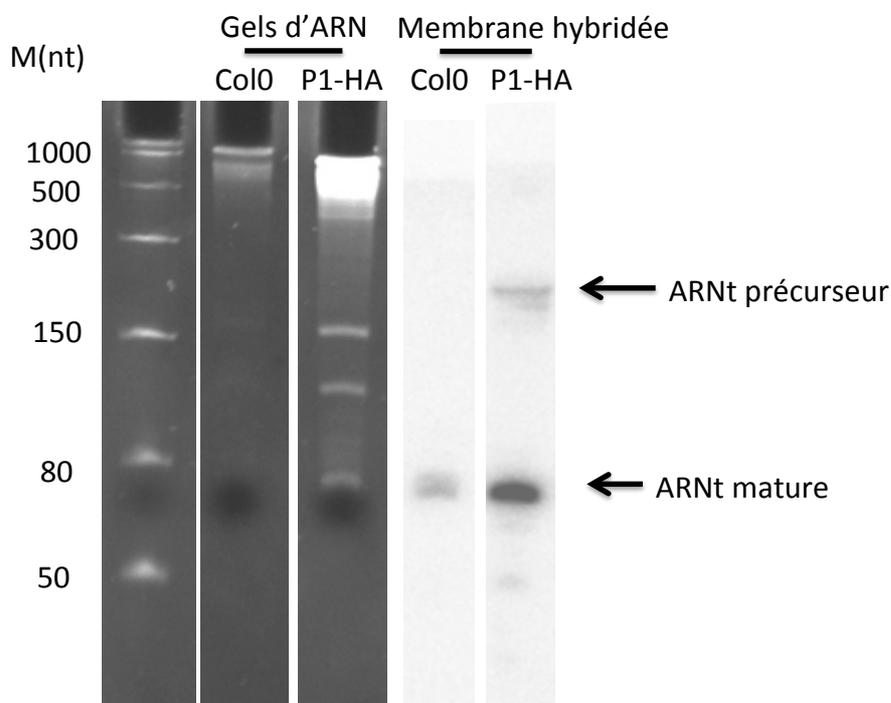


Figure 32 : Analyse des ARN issus des immuno-précipitations (IP) par hybridation de type northern avec une sonde spécifique de l'ARNt Lysine mitochondrial.

Les ARN issus des IP utilisant du matériel de plantes sauvages (Col0) et provenant de lignées PRORP1-HA sont séparés sur un gel polyacrylamide dénaturant, transférés sur membrane de nylon Hybond N+ et hybridés avec une sonde spécifique de l'ARNt lysine mitochondrial. Le marqueur de poids moléculaires (M) est indiqué en ribonucléotides.

Protéine	Col0		PRORP1 Rep1		PRORP1 Rep2		PRORP1 Rep3	
	Rang	Nb spec	Rang	Nb spec	Rang	Nb spec	Rang	Nb spec
PRORP1	/	0	5	194	39	109	10	151
MNU2	/	0	48	40	113	50	70	30

Figure 33 : Analyse par spectrométrie de masse de la présence de PRORP1 et d'un de ses partenaires connu dans les immuno-précipitations.

Les protéines issues des complexes immuno-précipités à partir de plantes sauvages (Col0) et provenant de lignées PRORP1-HA sont analysées par LC-MSMS. La qualité des immuno-précipitations est estimée par l'identification de l'appât d'IP, ici PRORP1 ainsi que de MNU2, un partenaire protéique connu de PRORP1. Le rang représente la place des protéines respectives dans la liste des protéines identifiées classées en fonction du nombre de spectres (Nb spec) détectés.

au moins une fois le gène AtPRORP1-HA (c'est-à-dire homozygotes pour la mutation knock out de *prop1* et possédant au moins une fois l'insertion AtPRORP1-HA), appelées ci-après plantes « AtPRORP1-HA ». Des co-immunoprécipitations utilisant des mitochondries purifiées à partir de ces plantes ont déjà permis d'identifier des partenaires protéiques de AtPRORP1. En particulier, cette étude a permis de mettre en évidence l'interaction directe de AtPRORP1 avec MNU2, une autre nucléase mitochondriale impliquée dans la maturation en 5' d'ARNm mitochondriaux (Stoll and Binder, 2016) mais dont la fonction précise et le mode d'action restent à préciser (Bouchoucha, en préparation)

Les co-immunoprécipitations sont réalisées à partir de mitochondries extraites des plantes AtPRORP1-HA et de plantes contrôle *Arabidopsis col0*. Le protocole de co-immunoprécipitation est basé sur celui utilisé par Ayoub Bouchoucha pour identifier les partenaires protéiques de AtPRORP1 mais en diminuant la concentration de sel de 150 à 50mM de NaCl afin de préserver les interactions ARN / protéine. En effet, en utilisant 150 mM de NaCl, aucun ARN n'était détectable dans les fractions immunoprécipitées. Après immunoprécipitation, les protéines et les ARN sont séparées par une extraction au trizol. Les ARN sont quantifiés avec un appareil QBit puis dilués et analysés avec un Bioanalyzer sur une puce « RNA pico plant ». L'expérience est réalisée trois fois de manière indépendante, c'est-à-dire à partir de trois extractions indépendantes de mitochondries provenant des plantes AtPRORP1-HA qui constituent des répliques biologiques. L'expérience 1 a permis d'immunoprécipiter 250 ng d'ARN et 800 ng environ pour les expériences 2 et 3.

La qualité des immunoprécipitations est vérifiée en détectant la présence d'un substrat connu de AtPRORP1 dans les fractions immunoprécipitées, ici en vérifiant la présence de l'ARNt mitochondriale lysine UUU par hybridation de type northern (Figure 32). Les protéines immunoprécipitées sont aussi analysées par LC-MSMS afin de vérifier la présence de AtPRORP1 et de ses interactants déjà caractérisés. En l'occurrence, les peptides issus de AtPRORP1 font partie des peptides les plus abondants dans les trois répliques biologiques et la protéine MNU2 est aussi systématiquement identifiée (Figure 33).

2. Identification des ARN immunoprécipités par séquençage à haut débit

a. Choix du type de banque d'ADNc à construire

Afin de pouvoir identifier un maximum de substrat de AtPRORP1, la banque choisie doit avoir la plus large fourchette de taille d'ARN possible. Le kit « ScriptSeq™ Complete »

Réplicat ARN IP avant déplétion

ARN IP après déplétion

ARN total avant déplétion

ARN total après déplétion

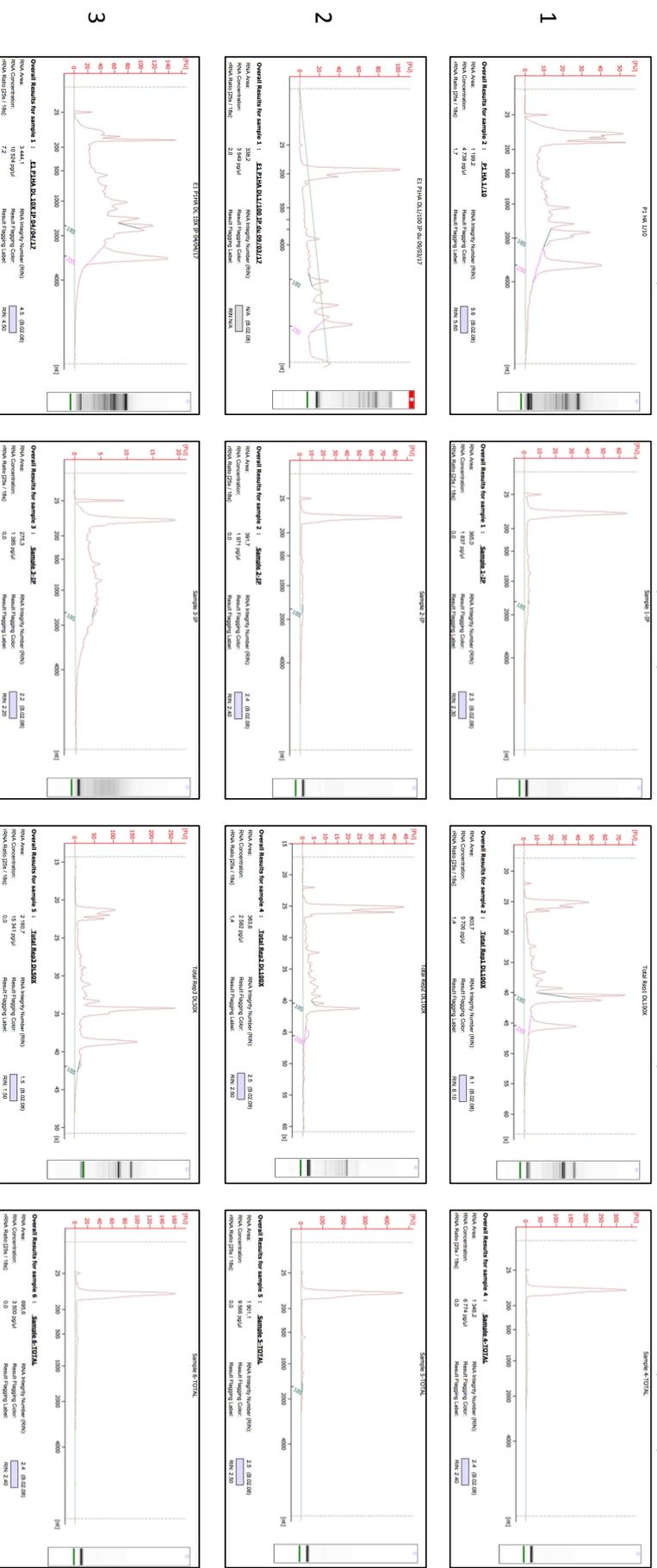


Figure 34 : Déplétion des ARNr avant la construction des banques d'ADNC.

Les ARN issus d'immuno-précipitations de PRORP1 ou les ARN mitochondriaux totaux sont traités pour dépléter leurs ARNr et analysés avec un Bioanalyser. La déplétion est visualisée en particulier par la disparition des pics correspondants aux ARNr mitochondriaux 18S et 26S.

permet de séquencer les ARN à partir de 100 nucléotides (ce qui inclut les ARNt après ajout des adaptateurs ARN nécessaires à la construction des banques), permet une déplétion des ARN ribosomiques, qui même s'ils sont substrats de AtPRORP1, *in vivo* devraient rester en quantité suffisante dans l'échantillon après déplétion. Le kit est disponible en version Seed/Root pour les échantillons avec peu de chloroplastes et la version Leaf pour ceux avec beaucoup de chloroplastes avec une optimisation du retrait des ARNr du chloroplaste. Ici le kit leaf est utilisé. De plus la version low-input du protocole nécessite seulement 100 ng d'ARN. Seuls les petits ARN ne sont pas séquencés avec ce type de banque et nécessiteraient la construction de banques spécifiquement dédiées aux petits ARN. De telles banques n'ont pas été construites, car à ce jour aucun petit ARN fonctionnel n'a été caractérisé dans les mitochondries de plantes supérieures.

b. Déplétion des ARNr et construction des banques

Six banques sont construites, trois avec les ARN immunoprécipités et trois avec les ARN totaux provenant des mitochondries utilisées pour les immunoprécipitations. La quantité d'ARN utilisée pour préparer les banques est la quantité la plus faible obtenue pour une immunoprécipitation soit 200 ng. Les mêmes étapes sont réalisées sur l'ARN total de mitochondrie mais à partir de 500 ng d'ARN. L'ARNr est d'abord retiré suivant les recommandations du fournisseur du kit. 1 μ L de la réaction est analysé avec un Bioanalyzer pour contrôler la déplétion des ARNr (Figure 34). Le reste est utilisé pour construire les banques qui sont purifiées avec le système Agencourt AMPXPure. La distribution des tailles des ADNc est vérifiée à nouveau avec le Bioanalyzer et dosée avec un QBit. Les banques sont alors normalisées à 4 nM et dénaturées avec 1 nM de NaOH puis diluées à 10 pM. 5 % de « PhiX Control v3 » (banque d'ADN du bactériophage Φ X174) sont ajoutés comme contrôle positif du séquençage.

c. Séquençage et analyses de sa qualité

Le séquençage est réalisé avec Sandrine Koechler à la plateforme Analyse de l'expression génique de l'IBMP sur un séquenceur Illumina MiSeq. Avant l'analyse des séquences obtenues, il faut vérifier la qualité du séquençage. Le premier indicateur est la densité des clusters (K) sur la cellule. Une densité trop faible indique une mauvaise accroche des adaptateurs sur la cellule et une densité trop forte entraîne un chevauchement des clusters et l'inexploitabilité des données de séquençage. Avec 1100K/mm² la densité est correcte bien

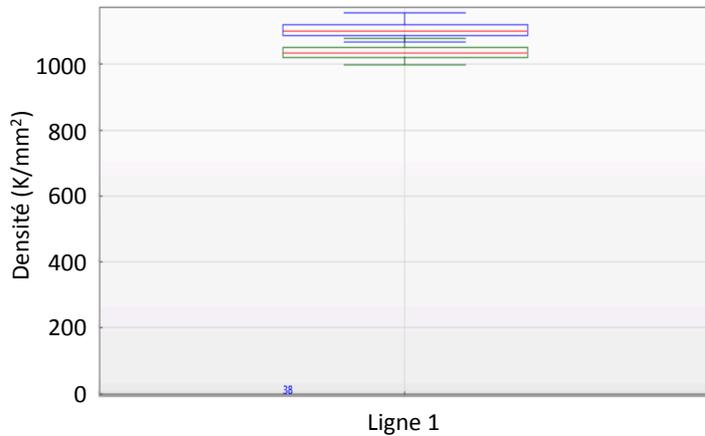


Figure 35 : Densité de clusters (K) sur la cellule.

La qualité des données de séquençage est tout d'abord évaluée en mesurant la densité de clusters générés sur la ligne du séquenceur Illumina MiSeq. La densité est mesurée en K/mm^2 . La boîte bleu représente le nombre de clusters bruts et en vert le nombre de cluster passant le contrôle qualité. Une densité de 1100 K/mm^2 a été obtenue pour le séquençage réalisé ici.

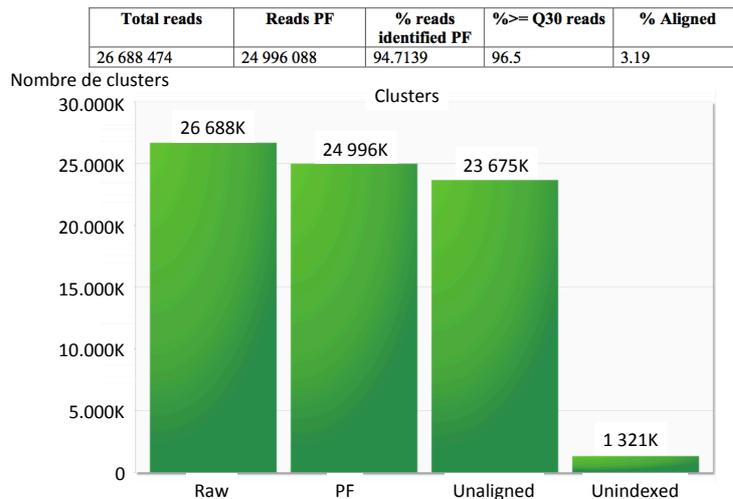


Figure 36 : Nombre et qualité des « reads » obtenus lors du séquençage.

Les données de séquençage sont analysées en évaluant le nombre et la qualité des séquences obtenues. « Total reads » représente le nombre total de séquences (ou reads) lus lors du séquençage. « Reads PF » représente le nombre de reads passant le contrôle qualité de séquençage. « % reads identifiées PF » représente le nombre de reads identifiées qui peuvent être classés dans une des banques (dont l'index a été identifié). « %>=Q30 reads » correspond au pourcentage de reads dont le taux d'erreur de lecture est inférieur ou égal à 1 base pour 1000 bases séquencées. « % aligned » correspond au pourcentage de reads alignés sur le génome de référence de la banque PhiX Control ajoutée à 5% dans la réaction de séquençage. Dans le graphique, « Unaligned » représente le nombre de reads non alignés au génome de référence contrôle et « unindexed » représente le nombre de reads ne comportant pas d'index identifiable.

que légèrement sous la valeur optimale comprise entre 1200 et 1400 K/mm² pour une analyse MiSeq avec la puce MiSeq Reagent V3 150 cycles (Figure 35).

Un total de 26 688 474 séquences (reads) ont été obtenues dont 24 996 088 reads sont conformes au test de contrôle qualité du séquenceur (Figure 36). 94,7119 % de ces reads ont pu être classés dans une des banques séquencées et parmi ces 94,7119 %, 96,5 % possèdent un indice de qualité supérieure ou égale à Q30, ce qui représente statistiquement 1 erreur toutes les 1000 bases. La distribution de la qualité des reads est représentée sur la Figure 37.

d. Analyses des résultats de séquençage

L'analyse bioinformatique a été réalisée par Valérie Cognat à la plateforme de bioimage et bioinformatique de l'institut (<http://www.ibmp.cnrs.fr/> plateformes/bioimage-et-bioinformatique/). Elle a pour but d'identifier les ARN qui sont surreprésentés dans les banques provenant des immunoprécipitations par rapport aux banques représentant l'ARN total des mitochondries d'*Arabidopsis*. Pour se faire, les duplicata de PCR sont tout d'abord retirés avec l'outil FastUniq (V1.1) (Xu, 2012). Entre 63 et 81 % de chaque banque est conservé. Les reads conservés sont alignés avec HiSAT2 (V2.0.5) (Kim, 2015) sur le génome mitochondrial TAIR10 d'*Arabidopsis thaliana* mais également sur le génome entier d'*Arabidopsis*. Environ 90 % de toutes les séquences des banques s'alignent sur le génome complet d'*Arabidopsis*. Les ARN séquencés sont donc bien issus d'*Arabidopsis* et non de contaminants potentiels. Entre 54 % et 78 % des ARN de chaque banque s'alignent sur le génome mitochondrial ce qui reflète la qualité variable des extractions de mitochondries. La purification de mitochondries du réplica 3 semble être de moins bonne qualité avec 54 % de l'ARN total provenant de mitochondries. Les réplicas 1 et 2 sont de meilleure qualité avec 78 % et 72 % d'ARN provenant des mitochondries (Figure 38).

L'analyse RIP-Seq a été réalisée avec ASPeaks (V2.0.1) (Kucukural, 2013). Cet outil permet de comparer l'ARN présent dans les immunoprécipitations avec l'ARN total. Beaucoup de pics correspondant à des enrichissements d'ARN dans les immunoprécipitations sont identifiés (Figure 39). Une analyse statistique est maintenant nécessaire afin d'identifier les loci les plus intéressants, i.e. ceux présentant un enrichissement d'ARN statistiquement significatif dans les 3 réplicas biologiques. Cette analyse statistique est en cours. Toutefois, il apparaît déjà qu'une quinzaine de loci paraissent fortement surreprésentés dans les immunoprécipitations. Certains se trouvent à côté d'un ARNt comme dans les ARNm ATP6-1 et ATP6-2 qui sont tous deux à côté d'un ARNt^{Ser}. D'autres comme dans les ARNm de *cox1*

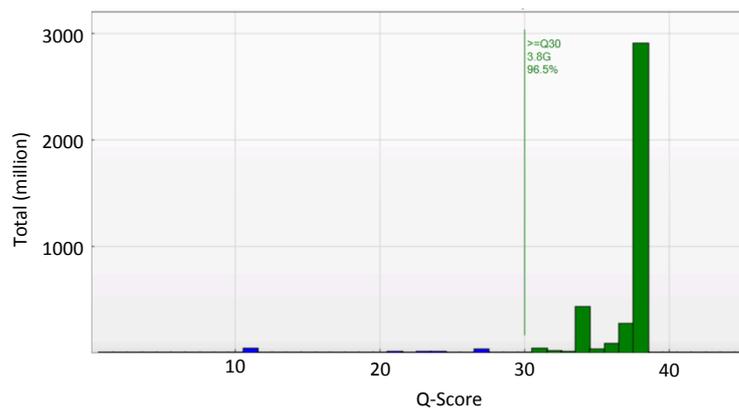


Figure 37 : Contrôle de la qualité des séquences.

Graphique représentant la distribution du score de qualité (Q Score) des séquences. Le score de qualité Q est calculé selon la formule $Q = -10\log_{10}P$, P étant la probabilité d'erreur de séquençage d'une base. Par exemple un score de Q30, qui définit le seuil à partir duquel les séquences sont considérées comme de bonne qualité, représente un taux d'erreur statistique de 1 base sur 1000.

Banque	% d'alignement de la banque sur :	
	Génome total	Génome mitochondrial
IP1	89,26	71,73
IP2	90,28	78,62
IP3	88,21	72,87
Total 1	91,63	61,55
Total 2	90,83	72,99
Total 3	91,59	54,12

Figure 38 : Evaluation de la proportion des séquences qui s'alignent sur les génomes de référence.

La proportion des séquences des différentes banques qui s'aligne sur le génome complet d'*Arabidopsis thaliana* et sur son génome mitochondrial est exprimée en pourcentages.

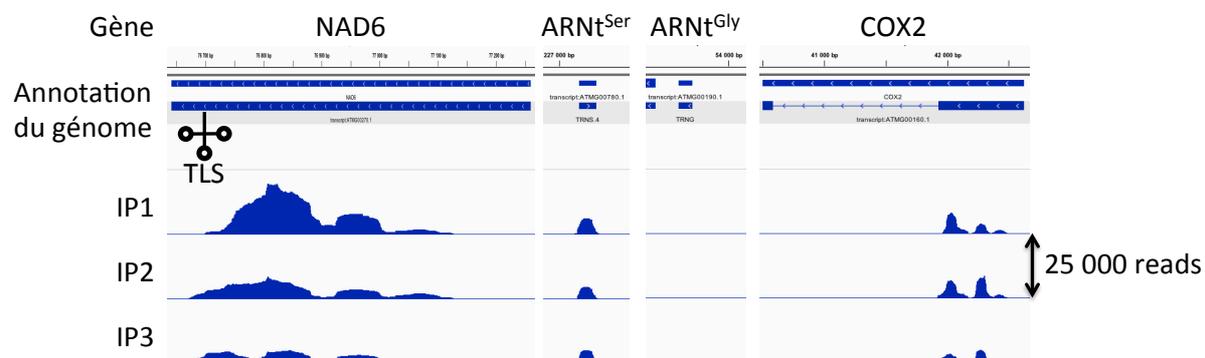


Figure 40 : Analyse des données de séquençage à 4 loci représentatifs.

Représentation graphique de la différence du nombre de reads dans chaque IP. Une échelle représentant 25000 reads est utilisée. Les annotations sont celles fournis par TAIR10. « TLS » représente la position d'une structure ARNt-like trouvée dans NAD6 qui est un substrat caractérisé de PRORP1. Toutefois, certains ARN substrats de PRORP1 comme l'ARNt Gly ne sont pas surreprésentés dans les fractions immunoprécipités.

ou *nad7* se trouvent à côté d'une structure ARNt-like (ou « t-element ») (Forner et al., 2007). Le transcrit de *nad6* qui avait déjà été identifié comme étant un substrat de AtPRORP1 *in vivo* est également fortement surreprésenté dans les immunoprécipitations. Par contre, d'autres ARN qui sont des substrats connus de AtPRORP1, e.g. l'ARNt^{Gly} ne sont pas surreprésentés dans les immunoprécipitations. (Figure 40). L'analyse en cours recherchera des corrélations entre les pics de surreprésentation des ARN et la présence de structures ARNt-like ou d'autres structures d'ARN qui pourraient être reconnues par PRORP.

A terme, les ARN considérés comme les meilleurs substrats candidats de AtPRORP1 seront clonés, transcrits *in vitro* et utilisés dans des tests d'activité afin de vérifier si AtPRORP1 est bien capable de les cliver, démontrant ainsi qu'il s'agit de nouveaux substrats de cette enzyme.

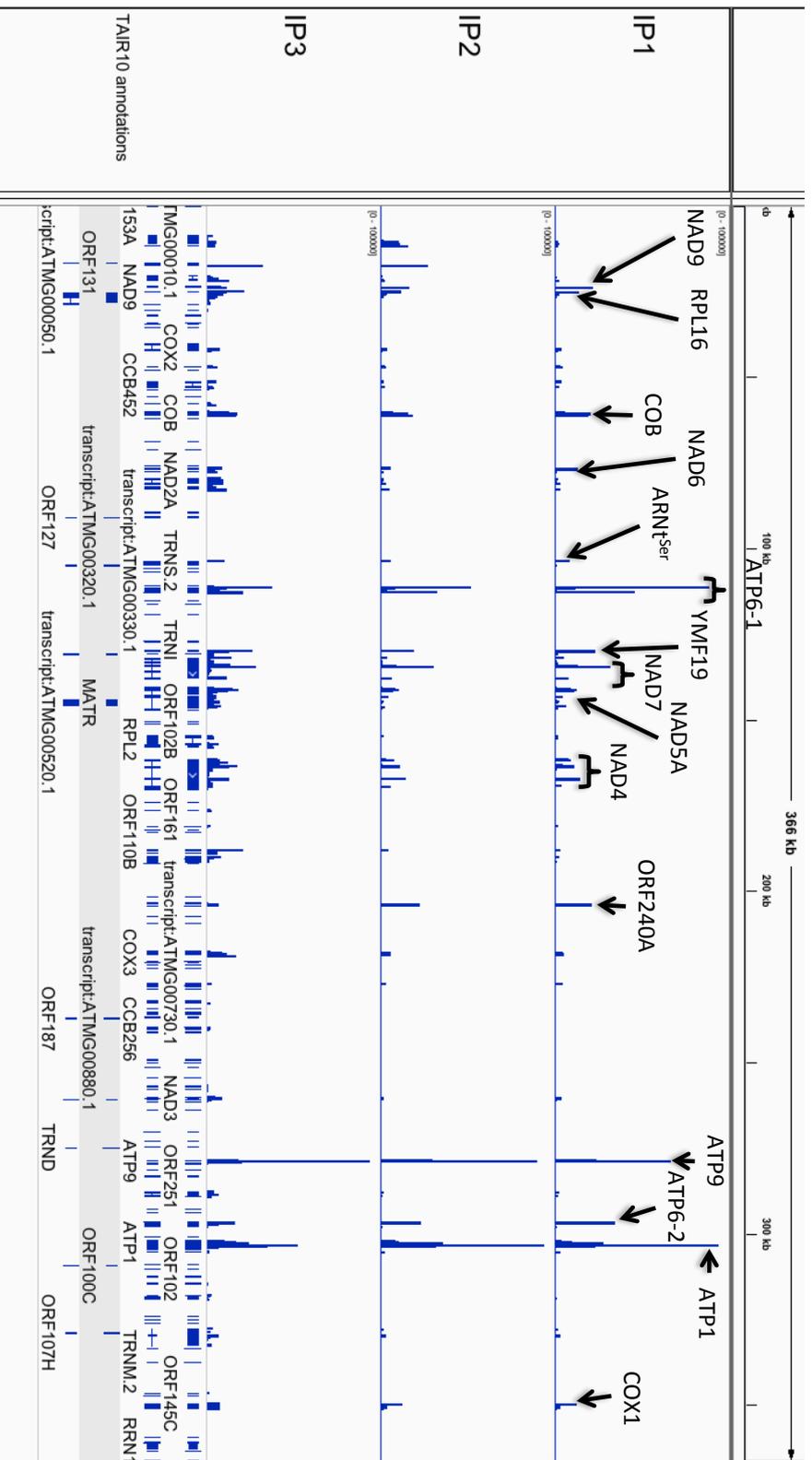


Figure 39 : Visualisation de transcrits spécifiquement accumulés dans les fractions immuno-précipitées par rapport au transcriptome total des mitochondries d'Arabidopsis.

Les trois premiers panneaux représentent des pics détectés après analyse différentielle par ASPEaks, c'est à dire, des ARN séquencés surreprésentés dans les immuno-précipitations par rapport au transcriptome mitochondrial. Le dernier panneau représente les annotations du génome mitochondrial issues de TAIR10. Les flèches indiquent des pics détectés et le nom des transcrits correspondants. On note que certains ARN, e.g. ATP1, ATP6 ou ATP9 s'accumulent fortement dans les trois expériences indépendantes analysées ici.

VII. Résultats chapitre 3 : Etude de la diversité des substrats ARN et des partenaires protéiques de la RNase P nucléaire

AtPRORP2

De la même manière que AtPRORP1 peut maturer l'ARNm de Nad6, il a été démontré que la RNase P nucléaire AtPRORP2 peut maturer le transcrite di-cistronique snoR43-ARNt^{Gly} (Gutmann 2012). L'accumulation de snoR43 est donc dépendante de l'activité de AtPRORP2. L'étude des substrats et des partenaires de AtPRORP1 dans la mitochondrie qui sert ici de test de faisabilité n'est pas achevée mais les premiers résultats semblent prometteurs. La même stratégie sera donc utilisée pour identifier les substrats ARN et les partenaires protéiques de AtPRORP2. Les co-immunoprécipitations seront effectuées à partir de noyaux purifiés et l'ARN obtenu sera analysé par séquençage à haut débit. Le génome nucléaire comporte un grand nombre de séquences dérivant de gènes d'ARNt, comme par exemple les SINE (voir introduction partie 3.a). Ces différents transcrits pourraient être des substrats des protéines PRORP nucléaires.

1. Construction de plantes complémentées exprimant une version étiquetée de AtPRORP2

a. Transformation des plantes P2/p2 ; p3/p3 et sélection des plantes avec P2-

HA

Les plantes exprimant une fusion de AtPRORP2 avec une étiquette d'affinité ont été construites de la même manière que les plantes AtPRORP1-HA. L'étiquette HA a été privilégiée à l'étiquette MYC, car les vecteurs à notre disposition qui possèdent des étiquettes 4 MYC de 40 acides aminés ou 3HA de 27 acides aminés avaient déjà été évalués lors de l'étude précédente. L'étiquette la plus petite a été privilégiée, car on pouvait prédire qu'elle ne nuirait pas à la fonction de AtPRORP2. En effet, l'expérience précédente acquise avec AtPRORP1 a montré que les plantes mutantes de AtPRORP1 pouvaient être complémentées avec des fusions de l'étiquette 3 HA et pas avec l'étiquette 4 MYC. Le transgène AtPRORP2-HA a été placé sous le contrôle du promoteur endogène de AtPRORP2 afin d'obtenir une expression comparable à celle de la protéine sauvage. Toutefois, il n'a pas été possible d'obtenir de plante complémentée avec l'étiquette 3 HA. Une construction alternative où un seul motif HA est fusionné à AtPRORP2 a donc été choisie. Cette construction a permis de compléter avec succès les plantes mutantes.

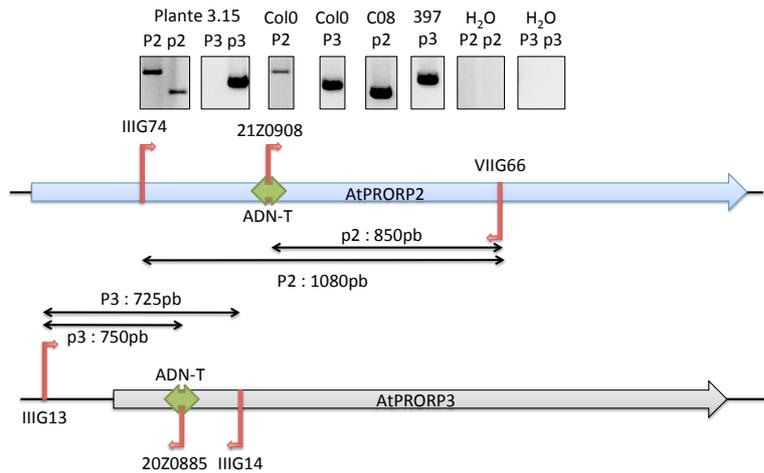


Figure 41 : Génotypage des plantes P2/p2 ; p3/p3.

Génotypages par PCR de la plante « 3.15 » P2/p2 ; p3/p3 avec comme contrôles les mêmes PCR effectués avec de l'ADN de plantes sauvages « Col0 » pour obtenir les version WT des gènes, de « C08 » qui est une plante prop2 homozygote et « 397 » un plante prop3 homozygote. « H₂O » correspond à un contrôle de PCR négatif sans ADN. P2, p2, P3 et p3 correspondent à différentes réactions de PCR visant à amplifier respectivement « PRORP2 » (une séquence endogène de PRORP2), « prop2 » (le gène de PRORP2 interrompu par l'ADN-T), « PRORP3 » (une séquence endogène de PRORP3), et « prop3 » (le gène de PRORP3 interrompu par l'ADN-T). Les flèches rouges représentent les positions des amorces sur AtPRORP2 (en bleu) et AtPRORP3 (en gris). Les zones d'insertion des ADN-T sont représentées par les doubles flèches vertes. Les amorces sont décrites dans la partie III. 4. d. Amorces de génotypage des plantes.

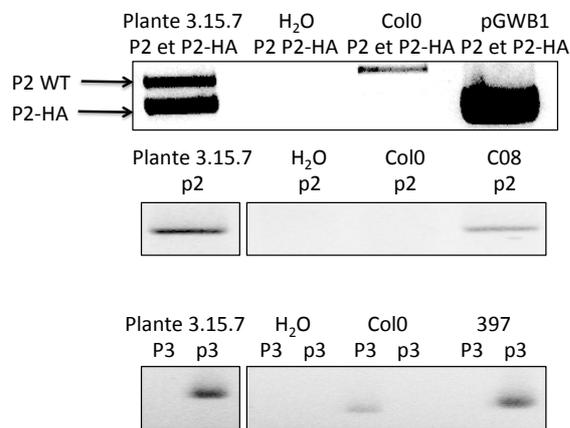


Figure 42 : Génotypage de plantes transformées avec la construction PRORP2-HA.

Génotypages par PCR de la plante « 3.15.7 » issue de la transformation de la plante 3.15 avec la construction « PRORP2-HA ». Comme contrôles, les mêmes PCR sont effectuées avec de l'ADN de plantes sauvages « Col0 », de plantes « C08 » qui est une plante prop2 homozygote et de l'ADN de plantes « 397 » un plante prop3 homozygote et le vecteur pGWB1::PRORP2-HA comme contrôle positif de PCR. H₂O correspond à un contrôle négatif sans ADN. P2, p2, P2-HA P3 et p3 correspondent à différentes réactions de PCR visant à identifier respectivement les fragments spécifiques de PRORP2, prop2, PRORP2-HA, PRORP3 et prop3. La différence de taille entre les amplicons « P2 WT » et « P2-HA » correspond à la présence d'un intron absent dans le transgène portant l'étiquette HA.

Pour cela, une lignée de plante d'*Arabidopsis thaliana* homozygote pour la mutation de AtPRORP3 et hétérozygote pour la mutation de AtPRORP2 a été utilisée. Cette lignée avait déjà été mise en place précédemment par croisement et sélection de deux lignés d'insertion d'ADN-T de AtPRORP2 et AtPRORP3 suivie d'une autofécondation afin d'obtenir des plantes homozygotes pour l'un des deux gènes et hétérozygotes pour le second. Le double mutant homozygote de AtPRORP2 et AtPRORP3 ne peut pas être obtenu, car la double mutation est létale (Gobert 2010). Le génotype de ces plantes a été vérifié par PCR (Figure 41). Les plantes sélectionnées sont transformées par immersion des boutons floraux dans une solution d'Agrobactérium transformée avec le plasmide binaire pGWB1 codant pour la fusion AtPRORP2-HA. Les transformants sont sélectionnés par culture *in vitro* sur milieu MS supplémenté en hygromicine. Les plantes résistantes sont alors génotypées par PCR de la même manière que les plantes transformées (Figure 42). Les plantes possédant le transgène AtPRORP2-HA et étant hétérozygotes pour AtPRORP2 et homozygotes pour prop3 sont autofécondées pour obtenir des plantes possédant au moins une version de la fusion AtPRORP2-HA et homozygotes pour les deux mutations knock-out prop2/prop2 et prop3/prop3 de la même manière que précédemment.

b. Sélection des plantes AtPRORP2-HA/AtPRORP2-HA ; Atprop2/Atprop2 ; prop3/prop3

Le double mutant AtPRORP2 et AtPRORP3 présente une létalité gamétophytique (Gobert 2010) sans formation de graines même avortées. Pour obtenir des plantes complémentées homozygotes pour AtPRORP2-HA, il faut donc disséquer les siliques afin de vérifier si la silique est pleine ou s'il manque $\frac{1}{4}$ des graines dans le cas d'une plante hétérozygote pour l'insertion AtPRORP2-HA. Des plantes AtPRORP2-HA/AtPRORP2-HA, prop2/prop2, prop3/prop3 ont pu être obtenues et sont nommées plantes « AtPRORP2-HA » ci-après.

c. Mise au point de la purification de noyaux d'Arabidopsis

Afin de limiter les contaminants, les expériences de co-immuno-précipitation de AtPRORP2 en complexe avec des substrats ARN ou des partenaires protéiques seront réalisées à partir de noyaux purifiés des plantes AtPRORP2-HA.

La mise au point de la purification des noyaux est en cours. Des premiers essais ont été réalisés en utilisant un protocole d'extraction « CellLytic™ PN Plant Nuclei

Isolation/Extraction Kit» de Sigma mais sans les tampons commerciaux. Ce protocole comprend des étapes de lyse cellulaire au Triton X100, des centrifugations différentielles et un coussin de sucrose. Ce protocole n'a en l'état pas donné satisfaction, car il s'est avéré inadapté pour une grande quantité de feuilles et n'a pas permis d'obtenir une quantité de noyaux isolés suffisante pour les immunoprécipitations. Un autre protocole compatible avec une grande quantité de matériel de départ et devant aboutir à de meilleurs rendements de noyaux purifiés est en cours de développement au laboratoire.

A ce stade, la recherche des substrats de AtPRORP1 et AtPRORP2 au niveau des transcriptomes entiers n'est pas achevée. Nous n'avons en l'occurrence pas terminé l'analyse bioinformatique permettant d'identifier des ARN spécifiquement immunoprécipités par AtPRORP1 et commençons seulement l'exploitation des plantes AtPRORP2-HA construites lors de ce travail de thèse. Toutefois la stratégie utilisant la co-immunoprécipitation semble adaptée, car nous avons déjà pu confirmer que des ARN tel que l'ARNm de *nad6* qui est un substrat de AtPRORP1 *in vivo* est enrichi dans les fractions immunoprécipitées. L'approche utilisée ici a aussi déjà permis d'identifier des partenaires protéiques de AtPRORP1 et donne donc des indices quant à l'intégration de l'activité RNase P parmi les autres processus d'expression génétique de la cellule.

Ainsi par cette approche, nous pensons qu'il sera possible de montrer que les protéines PRORP sont intégrées dans des complexes d'expression génétique fonctionnels et qu'elles sont impliquées dans le clivage de nombreux substrats non-ARNt, comme c'est le cas pour les RNases P ribonucléoprotéiques.

VIII. Résultats chapitre 4 : Complémentation de mutants de PRORP d'Arabidopsis avec une nouvelle RNase P protéique d'*Aquifex aeolicus*

Une nouvelle forme de RNase P a récemment été découverte par l'équipe de Roland Hartmann (Université de Marburg) (communication personnelle, article soumis). Jusqu'à présent il était admis que les bactéries et les archées utilisent exclusivement des RNases P ribonucléoprotéiques et que les RNases P protéiques (i.e. les protéines PRORP) sont uniquement trouvées chez les eucaryotes. Or, l'équipe de R. Hartmann a mis en évidence la RNase P utilisée par la bactérie *Aquifex aeolicus*. Ce bacille hyperthermophile de 2 à 6 μm vivant à proximité des volcans sous-marins utilise une petite protéine d'environ 200 acides aminés et d'un poids moléculaire de 23,6 kDa pour assurer l'activité RNase P. Cette protéine nommée « AaRP » pour *Aquifex aeolicus* RNase P a des similarités de séquence avec les nucléases de type NYN. Toutefois, elle ne possède pas de domaine de liaison à l'ARN et son mode de reconnaissance des pré-ARNt est encore inconnu. L'équipe de Roland Hartmann a pu montrer que cette enzyme possédait seule l'activité RNase P *in vitro* et qu'elle pouvait remplacer la RNase P ribonucléoprotéique nucléaire de levure *in vivo*.

Nous avons initié un projet collaboratif avec l'équipe de Roland Hartmann pour déterminer si les protéines AaRP peuvent aussi remplacer fonctionnellement les protéines PRORP *in vivo*. Pour cela, nous tentons de compléter des plantes mutées pour les protéines PRORP nucléaires avec une construction fusionnant AaRP avec un signal de localisation nucléaire d'Arabidopsis (NLS pour « Nuclear Localization Signal ») et nous cherchons à compléter des plantes mutées pour la RNase P mitochondriale et chloroplastique avec une construction fusionnant AaRP avec une séquence d'adressage aux organites d'Arabidopsis (OTS pour « Organelle Targeting Signal »). Ce travail a pour but de déterminer si ce nouveau type de RNase P peut être un équivalent fonctionnel des protéines PRORP.

1. Fusion de séquences d'adressage d'Arabidopsis à AaRP

Trois constructions différentes de AaRP ont été amplifiées et clonées dans pDONR207. La première ne contient aucune séquence d'adressage, seul AaRP flanqué des sites de clonage AttL y est présent. La seconde contient une séquence d'adressage NLS (PKKKRKV ou CCC AAG AAG AAG AGA AAA GUU) déjà utilisée précédemment dans

l'institut en fusion avec AaRP. La dernière construction comporte l'OTS de la glutamate ARNt synthétase d'Arabidopsis en fusion avec AaRP. Les sites AttB1 et AttB2 de même que le NLS ont été intégrés à chaque construction par PCR. En revanche, l'OTS a été intégré par digestion et ligation. L'OTS et AaRP ont été amplifiés séparément en ajoutant un site NcoI dans les amorces afin de pouvoir liguer les deux fragments et de les intégrer ensemble dans le vecteur pDONR207. Les différentes constructions ont ensuite été insérées dans différents vecteurs de destination. En l'occurrence, dans les vecteurs pK7FWG2 et pK7WGF2 qui permettent de fusionner la GFP aux extrémités N-terminales et C-terminales de AaRP afin d'effectuer des tests de localisation par expression transitoire dans *Nicotiana benthamiana*. Les différentes versions de AaRP ont aussi été clonées dans les vecteurs pGWB14 et pGWB15 qui permettent de fusionner une étiquette 3HA. Ces constructions seront utilisées pour tenter de compléter les mutants PRORP d'*Arabidopsis thaliana*.

2. Localisation des différentes versions d'AaRP dans *Nicotiana benthamiana*

L'étude de la localisation des différentes versions de AaRP chez les plantes n'a pas encore débuté. Les constructions de AaRP fusionné à la GFP dans les vecteurs pK7FWG2 et pK7WGF2 seront transfectées dans *Nicotiana benthamiana* de manière transitoire afin de vérifier la localisation des différentes constructions. La construction OTS-AaRP-GFP sera co-transfectée avec une sonde « MitoTracker » afin de pouvoir confirmer sa localisation mitochondriale. La localisation chloroplastique sera confirmée grâce à l'autofluorescence de la chlorophylle des chloroplastes. Les constructions NLS-AaRP-GFP et AaRP-GFP seront transfectées seules et la localisation nucléaire sera vérifiée par coloration au DAPI. Cette étude de localisation devrait permettre de valider les constructions en démontrant que les OTS et NLS d'Arabidopsis sont bien capables d'adresser AaRP aux organites et au noyau *in vivo*.

3. Complémentation de mutants de PRORP d'*Arabidopsis thaliana* avec AaRP

Les différentes constructions seront transformées dans différentes plantes mutantes pour les gènes PRORP d'Arabidopsis décrites ci-dessus. En l'occurrence, la construction OTS-AaRP sera transformée dans une plante hétérozygote AtPRORP1/atprorp1 afin de compléter l'activité RNase P mitochondriale et chloroplastique après autofécondation des plantes transformées exprimant AaRP et portant la mutation pour AtPRORP1. Les constructions sans signal d'adressage ou avec un NLS seront transformées dans des plantes

mutantes homozygotes AtPRORP3 et hétérozygotes pour la mutation de AtPRORP2 afin de compléter l'activité RNase P nucléaire. AaRP-HA sans signal d'adressage pourrait être assez petite pour diffuser passivement du cytosol vers le noyau et ainsi permettre de compléter l'activité RNase P nucléaire. Une telle version de AaRP sans séquence d'adressage a permis de compléter avec succès un mutant de RNase P nucléaire de levure (communication personnelle R. Hartmann).

Ces analyses ne sont encore qu'au stade préliminaire. Le projet commencé récemment comme un sujet annexe à mon travail de thèse revêt toutefois un intérêt particulier. La découverte de ce nouveau type de RNase P protéique ouvre des perspectives très intéressantes quant à la compréhension de la diversité et de l'histoire évolutive des enzymes à activité RNase P. La comparaison fonctionnelle des différents types de RNase P, ribonucléoprotéiques, PRORP et maintenant de type AaRP devrait aider à comprendre la distribution des différents types d'enzyme dans différents groupes phylogénétiques au cours de l'évolution.

IX. Discussion et Perspectives

1. Mode d'action des protéines PRORP

Depuis les premières descriptions des RNase P protéiques dans la mitochondrie humaine et chez les plantes (Holzmann et al., 2008; Gobert et al., 2010), des structures cristallographiques de la protéine PRORP humaine et des protéines PRORP d'Arabidopsis ont été obtenues. Ces études ainsi que d'autres études biophysiques en solution ont révélé l'organisation fonctionnelle et la diversité structurale de ces enzymes (travaux résumés par Schelcher et al. (Schelcher et al., 2016)). Les premières recherches sur le mode d'action des protéines PRORP ont montré quels résidus d'ARNt sont en contact avec l'enzyme (Gobert et al., 2013), mais la façon dont la reconnaissance de l'ARNt est obtenue par PRORP était inconnue et aucune donnée n'était disponible sur la structure du complexe PRORP / ARNt et sur la dynamique de l'enzyme.

a. Mode de reconnaissance des ARN par les motifs PPR de PRORP

L'analyse de la mutagenèse des motifs PPR de AtPRORP2 décrite ici et le modèle dérivé des données SAXS acquises sur le complexe AtPRORP2 / ARNt suggèrent que les seuls motifs PPR2 et PPR3 de PRORP sont impliqués dans la reconnaissance de l'ARNt.

Mis à part l'étude de PRORP décrite ici, toutes les protéines PPR caractérisées ont été décrites comme des protéines de liaison à l'ARN simple brin (Giegé, 2013). L'étude mécanistique des protéines PPR canoniques suggère que les motifs PPR se lient spécifiquement aux bases d'ARN non appariées d'une manière séquence spécifique (Barkan et al., 2012). Puisque PRORP lie n'importe quel ARNt de structure canonique, il a été proposé que PRORP puisse diverger du mode d'action général des protéines PPR et puisse posséder un mode de reconnaissance de l'ARN spécifique, basé sur la reconnaissance de caractéristiques structurales plutôt que sur la reconnaissance de séquences linéaires (Pinker et al., 2013; Brillante et al., 2016). Nos résultats suggèrent que ce n'est pas complètement le cas, du moins pour les motifs PPR2 et PPR3 qui semblent suivre le mode de reconnaissance canonique de bases individuelles par les motifs PPR. En effet, PPR2 reconnaît C56 et PPR3 une purine dans le coude de l'ARNt. Cette purine pourrait être soit G18 soit G57, bien que notre modèle SAXS indique que G18 est plus proche du motif PPR3 et est donc plus susceptible d'être reconnu par ce motif lorsque PRORP interagit avec un ARNt complet. Cependant, dans le contexte d'une mini-hélice (comme utilisée par Brillante et al. 2016) imitant le bras accepteur

avec une boucle du T moins contrainte, (Brillante et al., 2016), PPR3 pourrait également lier le résidu G57 alors que le PPR2 interagit avec le C56.

En revanche, les motifs PPR4 et PPR5 ne semblent pas participer à la liaison à l'ARNt et leur fonction reste spéculative, bien que ces motifs possèdent le repliement caractéristique des motifs PPR. Ils pourraient être impliqués dans la reconnaissance d'autres substrats de AtPRORP2 (qui ne sont pas des ARNt) pas encore identifiés. Alternativement, il est plus probable que la fonction de ces motifs soit essentiellement structurale, qu'ils permettent de positionner les motifs PPR2 et PPR3 à la bonne distance du site actif du domaine NYN. Le repliement global de PRORP agirait alors comme une « règle moléculaire » qui sélectionne le site de clivage à une certaine distance du coude de l'ARNt contacté par PPR2 et PPR3.

b. Nature du complexe formé par PRORP et l'ARNt

Dans l'étude présentée ici, seules les positions 5 et 35 des motifs PPR qui sont reconnues comme les plus importantes pour la liaison spécifique à l'ARN ont été mutées. A terme, l'analyse par mutagenèse devra être étendue. En particulier, la pertinence fonctionnelle d'autres positions telles que la position 2, qui a également été proposée pour être impliquée dans l'interaction avec l'ARN devra être analysée. Cette étude révélera si PRORP utilise spécifiquement d'autres résidus pour l'interaction avec l'ARN. Par ailleurs, l'étude de l'affinité d'interaction entre PRORP et les ARNt a révélé des K_D de l'ordre de 1 μ M, suggérant une interaction transitoire entre PRORP et l'ARNt. Cela suggère aussi un mode de reconnaissance de l'ARN spécifique à PRORP. Par exemple, les protéines PPR canoniques telles que PPR10 ont une affinité bien supérieure pour leur cible ARN (Barkan et al., 2012). La compréhension fine de l'interaction de PRORP avec l'ARN et de son mode d'action passera par la détermination d'une structure à l'échelle atomique du complexe formé par PRORP et l'ARNt. La détermination d'une telle structure devrait confirmer le rôle des motifs PPR2 et PPR3 proposés ici pour l'interaction avec l'ARN, devrait aider à identifier la fonctionnalité des autres motifs PPR. A plus long terme, cela permettra de comprendre (i) la diversité de mode d'action des protéines PPR, et (ii) la diversité mécanistique des enzymes à activité RNase P.

c. Flexibilité des protéines PRORP

Outre la détermination du processus de reconnaissance de l'ARN par PRORP, nous avons pu comparer sa structure cristallographique avec différents modèles de PRORP en

solution. Cela a permis de visualiser différentes conformations de PRORP. Pour explorer la flexibilité des protéines PRORP et observer d'éventuelles réorientations de domaines, nous avons effectué une analyse en modes normaux (NMA) connue pour fournir une bonne description des mouvements collectifs de basse fréquence dans les protéines (Tama and Sanejouand, 2001). La comparaison des modèles perturbés en NMA avec nos données SAXS en solution suggère que AtPRORP1 et AtPRORP2 peuvent adopter une variété de conformations grâce à une charnière entre le domaine catalytique NYN et le domaine central de liaison au zinc. Les mouvements associés correspondent à l'ouverture de la forme en Λ de PRORP et à la rotation de son domaine catalytique autour de la charnière centrale. Les modèles ont été affinés par rapport aux données SAXS avec l'algorithme génétique DADIMODO et ont conduit à l'observation de conformations multiples compatibles avec une réorientation du domaine catalytique. La superposition de la structure cristallographique de AtPRORP2 et des modèles de PRORP en solution seule ou en complexe avec l'ARNt suggère l'existence d'un continuum de conformations qui permettent d'accommoder les substrats ARN. Par exemple, la flexibilité de l'enzyme pourrait permettre de reconnaître des ARN avec un domaine hélicoïdal plus long qu'un bras accepteur d'ARNt canonique. Cette adaptabilité est e.g. illustrée par la capacité de PRORP à cliver des substrats tels que l'ARNt^{His} qui a une tige acceptrice de 8 paires de bases et peut être clivé par PRORP aux positions -1 et +1 (Placido et al., 2010) ou par le clivage de structures ARNt-like comme observé pour les ARNm mitochondriaux *nad6* et *orf291* (Gutmann et al., 2012; Fujii et al., 2016). Dans l'ensemble, la plasticité intrinsèque de AtPRORP2 observée ici, qui s'applique probablement à tous les membres de la famille PRORP, est peut-être cruciale pour lier ses substrats ARN et pour libérer les produits ARN après maturation.

d. Adaptabilité conformationnelle des enzymes de liaison aux ARNt

Un tel type d'adaptation conformationnelle a été observé dans d'autres enzymes monomériques de liaison aux ARNt. Par exemple, dans le cas de la phénylalanine-ARNt synthetase mitochondriale humaine, la structure cristallographique de l'enzyme isolée a révélé une conformation très compacte. Les données SAXS sur le complexe avec l'ARNt ont suggéré un grand mouvement du domaine de liaison à l'anticodon lors de la liaison à l'ARNt (Klipcan et al., 2012). Dans d'autres exemples, le facteur d'élongation bactérien EfTu opère une transition conformationnelle majeure suite à l'hydrolyse du GTP qui conduit à la translocation du ribosome et à la libération du facteur EfTu (Polekhina et al., 1996). La

protéine TruB bactérienne qui catalyse la formation de pseudo-uridine à la position U55 de l'ARNt subit une grande réorientation de sa région C-terminale lors de la liaison de l'ARNt (Pan et al., 2003). Dans un autre exemple, l'O-Phosphoseryl-ARNt kinase (PSTK), l'enzyme qui phosphoryle le Ser-ARNt^{Sec} pour produire l'O-phosphoséryl-ARNt^{Sec} opère un mouvement de plus de 60 Å de son domaine C-terminal pour lier la région variable de l'ARNt^{Sec} (Sherrer et al., 2011). De tels mouvements conformationnels pourraient donc être une caractéristique générale des protéines de liaison aux ARNt. Ces mouvements pourraient être requis pour moduler la reconnaissance, la liaison ou la libération d'ARNt, comme il a été proposé pour PRORP.

2. Diversité de fonctions des protéines PRORP

L'analyse d'une variété de RNases P ribonucléoprotéiques dans de nombreux organismes procaryotes et eucaryotes a montré la diversité fonctionnelle de ces enzymes. Par exemple, il est maintenant évident que les RNases P ribonucléoprotéiques sont impliquées dans la maturation d'un grand nombre d'ARN différents, en plus des ARNt, comme par exemple des ARNr (Peck-Miller and Altman, 1991), des ARNm (Alifano et al., 1994) ou encore des ARN viraux (Hartmann et al., 1995). Par ailleurs les RNases P ribonucléoprotéiques participent à divers processus d'expression génétique dans le cadre de complexes, comme c'est le cas pour la RNase P nucléaire humaine qui est requise pour la formation de complexes d'initiation de l'ARN polymérase III (Serruya et al., 2015).

Ainsi, des études ont été initiées au laboratoire pour déterminer si les RNases P protéiques sont elles aussi impliquées dans la maturation de différents ARN en plus des ARNt et si leurs fonctions impliquent l'interaction avec d'autres acteurs de l'expression génétique. Des premiers résultats ont montré chez *Arabidopsis* qu'au-delà de la maturation de l'ARNt, AtPRORP1 est impliquée dans la maturation des ARNm mitochondriaux au niveau des structures ARNt-like comme observé *in vivo* pour la maturation en 3' de nad6 (Gutmann et al., 2012) ou la maturation de l'orf291 (Fujii et al., 2016). Des études exhaustives visant à identifier l'ensemble des partenaires protéiques et de substrats ARN de AtPRORP1 dans les mitochondries et de AtPRORP2 dans le noyau d'*Arabidopsis* ont maintenant été initiées, en particulier durant ce travail de thèse. Pour cela des approches basées sur l'immunoprécipitation de complexes impliquant AtPRORP1 et AtPRORP2 ont été utilisées.

a. Partenaires protéiques des protéines PRORP

L'analyse des complexes protéiques impliquant AtPRORP1 par Ayoub Bouchoucha (doctorant au laboratoire), a révélé un certain nombre de partenaires protéiques potentiels de AtPRORP1. En particulier des acteurs de l'expression génétique comme RBP6, MTERF30, PNPase et MNU2 ont été identifiés. Les résultats ont montré à ce jour que seule MNU2, une autre nucléase mitochondriale interagit directement avec AtPRORP1. La caractérisation de l'interaction AtPRORP1 / MNU2 a révélé qu'un domaine spécifique de MNU2 et un motif riche en proline de AtPRORP1 sont directement responsables de l'interaction des protéines (Bouchoucha et al. en préparation). Ce motif riche en proline est très bien conservé chez les streptophytes, laissant supposer que les interactions PRORP / MNU sont spécifiques aux organites des plantes terrestres. La principale question est de comprendre la raison fonctionnelle de l'interaction entre AtPRORP1 et MNU2 dans les mitochondries d'Arabidopsis. Une étude a montré que la fonction des protéines MNU est nécessaire pour la maturation de certains ARNm mitochondriaux d'Arabidopsis (Stoll and Binder, 2016) mais leur participation à la maturation d'autres types de transcrits, en particulier les précurseurs d'ARNt, n'a pas été étudiée et le mode d'action de MNU2 est inconnu. Par exemple, il n'est pas clair si MNU2 est un endo ou une exonucléase. Dans le complexe AtPRORP1 / MNU2, une fonction possible de MNU2 pourrait être de dégrader les séquences 5' leader de l'ARNt après le clivage par AtPRORP1. Alternativement, MNU2 pourrait participer directement à la maturation des ARNt et / ou d'ARNm. Par exemple, il pourrait couper de longues séquences leader, tandis que AtPRORP1 effectuerait la maturation finale des séquences leader raccourcies. Ceci serait en accord avec les observations montrant que les pré-ARNt sont souvent transcrits avec de longues séquences leaders dans les mitochondries végétales (Hammani and Giegé, 2014), tandis que *in vitro* PRORP clive préférentiellement les pré-ARNt avec des séquences leaders courtes (Brillante et al., 2016). Une telle coopération de deux nucléases pour la maturation de l'ARN, si elle est avérée, rappellerait l'action concertée de la RNase II et de la PNPase pour la maturation de l'extrémité 3' des ARNm dans les mitochondries de plante (Perrin et al., 2004; Hammani and Giegé, 2014).

La quantité rapidement croissante de données sur le mode d'action des protéines PRORP, c'est-à-dire la détermination de leurs constantes catalytiques a révélé de manière surprenante que, *in vitro*, les enzymes PRORP ne sont pas des catalyseurs aussi bons que les RNase P ribonucléoprotéiques (Schelcher et al., 2016). L'identification de partenaires d'interaction tels que MNU2 permet de proposer que les fonctions des partenaires protéiques

pourraient améliorer ou réguler l'activité des protéines PRORP *in vivo*. En présence de facteurs supplémentaires, les protéines PRORP pourraient s'avérer être de meilleurs catalyseurs. De telles hypothèses devront être étudiées à travers la caractérisation fonctionnelle et mécanistique des complexes impliquant les protéines PRORP. À plus long terme, les réseaux d'interaction complets des enzymes PRORP dans les chloroplastes et le noyau révéleront l'intégralité de leur intégration dans les mécanismes d'expression génétique et parmi les autres processus cellulaires.

b. Diversité de substrats ARN des protéines PRORP

Les données obtenues préalablement montrent que AtPRORP1 est capable de maturer d'autres ARN que les ARNt dans la mitochondrie d'Arabidopsis. Les données de séquençage obtenues visant à identifier les ARN spécifiquement immunoprécipités avec AtPRORP1 sont en cours d'analyse. Du fait de la faible complexité du transcriptome mitochondrial d'Arabidopsis, avec moins d'une centaine de gènes exprimés, (Unsel et al., 1997), la profondeur de séquençage (de 25 à 30 millions de reads obtenus avec un séquenceur MiSeq) devrait être largement suffisante pour l'identification des substrats. L'alignement de ces reads sur le génome de référence a été rendu plus difficile par la présence de nombreux sites d'édition. En effet le transcriptome mitochondrial d'Arabidopsis est modifié par plus de 450 modifications de C en U (Giegé and Brennicke, 1999). Ainsi, une liste des positions des sites d'édition a été ajoutée et utilisée comme des SNP lors de l'analyse afin de pouvoir aligner les séquences d'ADNc éditées sur le génome de référence. Les résultats de séquençage en cours d'analyse suggèrent déjà que certains ARN, en particulier des ARNm sont surreprésentés dans les fractions immunoprécipitées avec AtPRORP1. Toutefois, seule une analyse statistique robuste permettra d'identifier les meilleurs substrats putatifs de AtPRORP1. Les loci identifiés seront analysés afin de vérifier si une structure en TLS (ou autre) est présente, puis clonés et transcrits *in vitro* afin de vérifier si AtPRORP1 est effectivement capable de maturer ces transcrits.

Cette analyse avec AtPRORP1 et le transcriptome mitochondrial sert essentiellement d'analyse pilote pour l'identification à plus grande échelle des substrats de AtPRORP2 dans le noyau. En particulier, il est important d'établir un protocole bio-informatique robuste qui pourra être utilisé avec un transcriptome beaucoup plus complexe comme le transcriptome nucléaire. Cette complexité accrue nécessitera toutefois d'obtenir une profondeur de séquençage bien plus importante. Pour cela, les banques d'ADNc construites à partir d'ARN immunoprécipités avec AtPRORP2 seront plutôt analysées avec un séquenceur de capacité

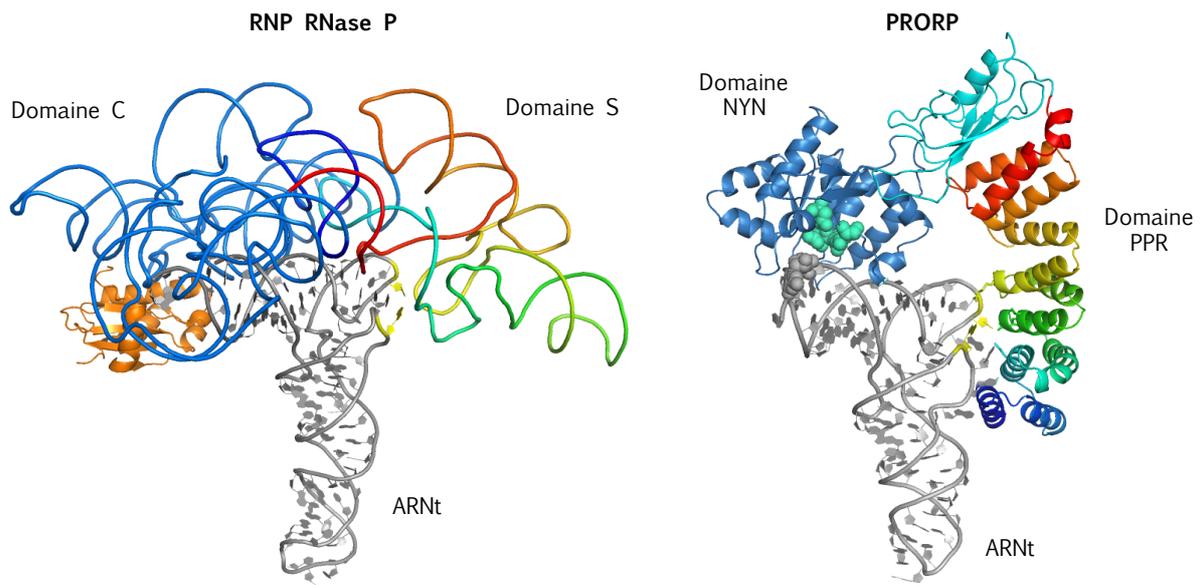


Figure 43 : Comparaison structurale des RNase P de type RNP et PRORP.

La comparaison de la structure expérimentale du complexe de la RNase P de *Thermotoga maritima* avec l'ARNt (à gauche) et du modèle du complexe de PRORP avec l'ARNt (à droite), comme déterminé par Pinker & Schelcher et al. 2017 révèle des similarités dans le mode de reconnaissance de l'ARN. Les deux types d'enzymes reconnaissent des nucléotides conservés dans le coude de l'ARNt (en jaune). Une analogie peut être faite entre les domaines catalytiques des enzymes (en bleu) et leurs domaines de spécificité (en arc en ciel).

supérieure comme le séquenceur Illumina HiSeq (disponible e.g. à l'IGBMC à Strasbourg). Le projet visant à identifier les substrats des protéines PRORP nucléaires n'en est qu'à ses débuts. Le travail de thèse a essentiellement consisté ici à développer les outils nécessaires au projet, en particulier les plantes exprimant une fusion de AtPRORP2 avec un seul motif HA. L'absence de complémentation des mutants AtPRORP2/AtPRORP3 avec une construction exprimant une fusion de AtPRORP2 avec 3 étiquettes HA suggère une différence de mode d'action entre AtPRORP1 et AtPRORP2. En effet, le mutant knock-out de AtPRORP1 a pu être complété avec une fusion comportant 3 étiquettes HA, alors que cela s'est avéré impossible avec AtPRORP2. Cela démontre que l'étiquette 3 HA interfère avec la fonction de AtPRORP2 mais pas avec celle de AtPRORP1. Ceci pourrait indiquer un mode de reconnaissance de l'ARN différent. Alternativement, l'étiquette 3 HA pourrait interférer avec la liaison d'un partenaire de AtPRORP2 pour une fonction essentielle. A court terme, des extraits nucléaires préparés à partir des plantes obtenues lors du travail de thèse seront utilisés dans des essais de co-immunoprécipitation avec AtPRORP2, ce qui devrait révéler à la fois l'ensemble des substrats de cette RNase P nucléaire et permettre d'identifier ses partenaires protéiques.

L'ensemble de ces travaux en cours devra faire progresser considérablement la compréhension de la diversité des rôles biologiques des RNases P protéiques et permettra aussi une comparaison avec la diversité fonctionnelle des RNases P ribonucléoprotéiques, ce qui donnera des indices pour comprendre la diversification des enzymes à activité RNase P au cours de l'évolution.

3. Diversité des enzymes à activité RNase P au cours de l'évolution

L'étude fonctionnelle et mécanistique des enzymes à activité RNase P, de type ribonucléoprotéique et de type PRORP comme réalisée lors de cette thèse suggère que les deux types d'enzyme ont des modes de reconnaissance de l'ARN similaires. Toutes deux reconnaissent des résidus conservés dans les domaines du T et du D des ARNt et une analogie claire peut être faite entre les domaines PPR de PRORP et le domaine de spécificité des RNP ainsi qu'entre le domaine NYN de PRORP et le domaine catalytique des RNP (Schelcher et al., 2016) (Figure 43). Cette analyse pose aussi la question de l'histoire évolutive des RNase P et interroge sur les raisons fonctionnelles de leur répartition et de leur diversification. Les analyses phylogénétiques ont révélé une corrélation entre la présence d'un type donné d'enzyme (RNP RNase P ou PRORP) et l'absence de l'autre dans une lignée et / ou un compartiment spécifique. Les exemples les plus extrêmes sont les champignons, où les

RNases P RNP sont trouvées à la fois dans les mitochondries et dans le noyau alors que PRORP est absente, et les Streptophytes ainsi que les trypanosomes, où les protéines PRORP sont présentes dans les organites et le noyau, alors que les gènes RNP sont absents. Des corrélations similaires sont observables pour tous les principaux groupes d'eucaryotes (Lechner et al., 2015). La présence de protéines PRORP dans quatre des cinq super-groupes d'eucaryotes suggère fortement qu'elles ont évolué très tôt pendant l'évolution des eucaryotes, dans un organisme à la base des groupes eucaryotes modernes. Il semble probable que la fusion de domaines PPR et NYN définissant les protéines PRORP ait eu lieu une seule fois lors de l'évolution (Lechner et al., 2015). Avec l'apparition de PRORP, les deux types d'enzymes ont dû être en compétition. La RNP a prévalu dans certains organismes comme les champignons ou dans certains compartiments comme le noyau des animaux et les enzymes protéiques n'ont pas été retenues. Dans ces organismes, les RNP pourraient avoir gagné des fonctions supplémentaires qui ne pouvaient être fournies par la RNase P protéique. En revanche, PRORP a gagné la compétition dans d'autres organismes, i.e. chez certains chlorophytes, chez les streptophytes, et les trypanosomes et les RNP de type bactérien ont été perdues. Il apparaît que les lignées animales et végétales ont utilisé deux voies très différentes pour distribuer leurs activités RNase P. Les organismes unicellulaires à la base du groupe des animaux semblent avoir retenu les protéines PRORP pour la fonction RNase P mitochondriale et ce statut a été conservé dans toutes les espèces métazoaires. En revanche, les organismes unicellulaires à la base de la lignée végétale semblent avoir initialement conservé les enzymes PRORP uniquement pour leur activité RNase P nucléaire. Ensuite, dans les espèces plus récentes, PRORP a également pris en charge la fonction RNase P des organites (Lechner et al., 2015).

En conclusion, l'analyse globale de la répartition des RNase P suggère que depuis son origine PRORP semble avoir été une enzyme invasive, remplaçant la RNase P RNP ancestrale dans plusieurs groupes d'eucaryotes, dans des organismes entiers ou dans des compartiments cellulaires particuliers. La tendance évolutive à remplacer la RNase P RNP par PRORP est crédible si l'on considère la capacité de PRORP à remplacer instantanément les fonctions des enzymes RNP pour la biogenèse de l'ARNt, comme il a été montré pour la RNase P bactérienne (Gobert et al., 2010) et pour la RNase P nucléaire de levure (Taschner et al., 2012; Weber et al., 2014). La découverte des protéines AaRP, une nouvelle classe de RNase P protéique chez *Aquifex aeolicus* complique encore le tableau. Initialement cantonnées aux eucaryotes, des RNases P protéiques semblent aussi avoir évolué chez les procaryotes, indépendamment des protéines PRORP. Le travail initié lors de cette thèse devra contribuer à

déterminer si ces enzymes sont les équivalents fonctionnels des protéines PRORP. Cette évolution où des RNases P ribonucléoprotéiques sont remplacées par des protéines pourrait témoigner d'un processus de transition continu du « monde des ARN » pré-biotique au monde actuel qui utilise essentiellement les protéines pour les activités catalytiques.

X. Bibliographie

- Agrawal, R.K. and Sharma, M.R.** (2012). Structural aspects of mitochondrial translational apparatus. *Curr. Opin. Struct. Biol.* **22**: 797–803.
- Ahmed, T., Yin, Z., and Bhushan, S.** (2016). Cryo-EM structure of the large subunit of the spinach chloroplast ribosome. *Sci. Rep.* **6**: 35793.
- Akama, K., Junker, V., and Beier, H.** (2000). Identification of two catalytic subunits of tRNA splicing endonuclease from *Arabidopsis thaliana*. *Gene* **257**: 177–185.
- Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C.M., Bruni, C.B., and Carlomagno, M.S.** (1994). Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA. *Genes Dev.* **8**: 3021–3031.
- Altman, S.** (2007). A view of RNase P. *Mol. Biosyst.* **3**: 604–607.
- Banerjee, A.K.** (1980). 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiol. Rev.* **44**: 175–205.
- Banks, J.A. et al.** (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* **332**: 960–963.
- Bare, L.A. and Uhlenbeck, O.C.** (1986). Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry* **25**: 5825–5830.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I.** (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet.* **8**: e1002910.
- Barkan, A. and Small, I.** (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **65**: 415–442.
- Beligni, M.V., Yamaguchi, K., and Mayfield, S.P.** (2004). Chloroplast elongation factor ts pro-protein is an evolutionarily conserved fusion with the s1 domain-containing plastid-specific ribosomal protein-7. *Plant Cell* **16**: 3357–3369.
- Bilbille, Y., Gustilo, E.M., Harris, K.A., Jones, C.N., Lusic, H., Kaiser, R.J., Delaney, M.O., Spremulli, L.L., Deiters, A., and Agris, P.F.** (2011). The human mitochondrial tRNA^{Met}: Structure/function relationship of a unique modification in the decoding of unconventional codons. *J. Mol. Biol.* **406**: 257–274.
- Bock, R.** (2000). Sense from nonsense: How the genetic information of chloroplasts is altered by RNA editing. *Biochimie* **82**: 549–557.
- Brautigam, C.A.** (2015). Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. In *Methods in Enzymology*, pp. 109–133.

- Brillante, N., Gößringer, M., Lindenhofer, D., Toth, U., Rossmanith, W., and Hartmann, R.K.** (2016). Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.* **44**: 2323–2336.
- Brown, G.G., Colas des Francs-Small, C., and Ostersetzer-Biran, O.** (2014). Group II intron splicing factors in plant mitochondria. *Front. Plant Sci.* **5**: 35.
- Brulé, H., Holmes, W.M., Keith, G., Giege, R., and Florentz, C.** (1998). Effect of a mutation in the anticodon of human mitochondrial tRNA(Pro) on its post-transcriptional modification pattern. *Nucleic Acids Res.* **26**: 537–543.
- Cabello-Villegas, J., Winkler, M.E., and Nikonowicz, E.P.** (2002). Solution conformations of unmodified and A37N6-dimethylallyl modified anticodon stem-loops of *Escherichia coli* tRNA^{phe}. *J. Mol. Biol.* **319**: 1015–1034.
- Chamberlain, J.R., Lee, Y., Lane, W.S., and Engelke, D.R.** (1998). Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev.* **12**: 1678–1690.
- Chateigner-Boutin, A.-L. and Small, I.** (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* **35**: e114.
- Chen, L.S., Du-Cuny, L., Vethantham, V., Hawke, D.H., Manley, J.L., Zhang, S., and Gandhi, V.** (2010). Chain termination and inhibition of mammalian poly(A) polymerase by modified ATP analogues. *Biochem. Pharmacol.* **79**: 669–677.
- Cheng, S. et al.** (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* **85**: 532–547.
- Chujo, T. and Suzuki, T.** (2012). Trmt61B is a methyltransferase responsible for 1-methyladenosine at position 58 of human mitochondrial tRNAs. *RNA* **18**: 2269–2276.
- Cognat, V., Pawlak, G., Duchêne, A.M., Daujat, M., Gigant, A., Salinas, T., Michaud, M., Gutmann, B., Giegé, P., Gobert, A., and Maréchal-Drouard, L.** (2013). PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Res.* **41**: D273-9.
- Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthomé, R., and Lurin, C.** (2013). Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol.* **10**: 1557–1575.
- Covello, P.S. and Gray, M.W.** (1989). RNA editing in plant mitochondria. *Nature* **341**: 662–666.
- Cui, Y., Rao, S., Chang, B., Wang, X., Zhang, K., Hou, X., Zhu, X., Wu, H., Tian, Z.,**

- Zhao, Z., Yang, C., and Huang, T.** (2015). AtLa1 protein initiates IRES-dependent translation of *WUSCHEL* mRNA and regulates the stem cell homeostasis of *A. thaliana* in response to environmental hazards. *Plant. Cell Environ.* **38**: 2098–2114.
- Dalluge, J.J., Hashizume, T., Sopchik, A.E., McCloskey, J.A., and Davis, D.R.** (1996). Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Res.* **24**: 1073–1079.
- Davis, D.R.** (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**: 5020–5026.
- Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G.M., Bergdoll, M., Taylor, N.L., Pogson, B.J., Small, I., Imbault, P., and Gualberto, J.M.** (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA^{Arg}(ACG) and is essential for efficient chloroplast translation. *Plant Cell* **21**: 2058–2071.
- Deppmann, C.D., Alvania, R.S., and Taparowsky, E.J.** (2006). Cross-Species Annotation of Basic Leucine Zipper Factor Interactions: Insight into the Evolution of Closed Interaction Networks. *Mol. Biol. Evol.* **23**: 1480–1492.
- Donghi, D. and Schnabl, J.** (2011). Multiple roles of metal ions in large ribozymes. *Met Ions Life Sci* **9**: 197–234.
- Drechsel, O. and Bock, R.** (2011). Selection of Shine-Dalgarno sequences in plastids. *Nucleic Acids Res.* **39**: 1427–1438.
- Duchêne, A.-M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Maréchal-Drouard, L., and Small, I.D.** (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 16484–16489.
- Edelheit, O., Hanukoglu, A., and Hanukoglu, I.** (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* **9**: 61–69.
- Eder, P.S., Kekuda, R., Stolc, V., and Altman, S.** (1997). Characterization of two scleroderma autoimmune antigens that copurify with human ribonuclease P. *Proc Natl Acad Sci U S A* **94**: 1101–1106.
- Englert, M., Latz, A., Becker, D., Gimple, O., Beier, H., and Akama, K.** (2007). Plant pre-tRNA splicing enzymes are targeted to multiple cellular compartments. *Biochimie* **89**: 1351–1365.
- Estrozi, L.F., Boehringer, D., Shan, S.-O., Ban, N., and Schaffitzel, C.** (2011). Cryo-EM

- structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat. Struct. Mol. Biol.* **18**: 88–90.
- Fey, J. and Maréchal-Drouard, L.** (1999). Compilation and Analysis of Plant Mitochondrial Promoter Sequences: An Illustration of a Divergent Evolution between Monocot and Dicot Mitochondria. *Biochem. Biophys. Res. Commun.* **256**: 409–414.
- Forner, J., Weber, B., Thuss, S., Wildum, S., and Binder, S.** (2007). Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: T-elements contribute to 5' and 3' end formation. *Nucleic Acids Res.* **35**: 3676–3692.
- Fujii, S., Suzuki, T., Giegé, P., Higashiyama, T., Koizuka, N., and Shikanai, T.** (2016). The Restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in *Arabidopsis*. *Plant J.* **86**: 504–513.
- Gaborieau, L., Brown, G.G., and Mireau, H.** (2016). The Propensity of Pentatricopeptide Repeat Genes to Evolve into Restorers of Cytoplasmic Male Sterility. *Front. Plant Sci.* **7**: 771–779.
- Giegé, P.** (2013). Pentatricopeptide repeat proteins: a set of modular RNA-specific binders massively used for organelle gene expression. *RNA Biol* **10**: 1417–1418.
- Giegé, P. and Brennicke, A.** (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 15324–15329.
- Giegé, P., Hoffmann, M., Binder, S., and Brennicke, A.** (2000). RNA degradation buffers asymmetries of transcription in *Arabidopsis* mitochondria. *EMBO Rep.* **1**: 164–170.
- Giegé, P., Sweetlove, L.J., Cognat, V., and Leaver, C.J.** (2005). Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in *Arabidopsis*. *Plant Cell* **17**: 1497–1512.
- Giege, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., and Florentz, C.** (2012). Structure of transfer RNAs: Similarity and variability. *Wiley Interdiscip. Rev. RNA* **3**: 37–61.
- Gobert, A., Gutmann, B., Taschner, A., Gössringer, M., Holzmann, J., Hartmann, R.K., Rossmannith, W., and Giegé, P.** (2010). A single *Arabidopsis* organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.* **17**: 740–744.
- Gobert, A., Pinker, F., Fuchsbaauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giegé, P.** (2013). Structural insights into protein-only RNase P complexed with tRNA. *Nat. Commun.* **4**: 1353–1361.
- Gray, M.W. et al.** (1999). Mitochondrial evolution. *Science* **283**: 1476–1481.
- Greber, B.J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D., and Ban, N.** (2015). The complete structure of the 55S mammalian mitochondrial ribosome.

- Science **348**: 303–308.
- Green, B.R.** (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**: 34–44.
- Gualberto, J.M., Domon, C., Weil, J.H., and Grienenberger, J.M.** (1989). Nucleotide sequence of the wheat mitochondrial tRNAGlu(UUC) gene. *Nucleic Acids Res.* **17**: 3586.
- Gualberto, J.M., Milesina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., and Dietrich, A.** (2014). The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **100**: 107–120.
- Gualberto, J.M., Weil, J.H., and Grienenberger, J.M.** (1990). Editing of the wheat coxIII transcript: Evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. *Nucleic Acids Res.* **18**: 3771–3776.
- Guerrier-Takada, C. et al.** (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**: 849–857.
- Gully, B.S., Cowieson, N., Stanley, W.A., Shearston, K., Small, I.D., Barkan, A., and Bond, C.S.** (2015). The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res.* **43**: 1918–1926.
- Gutmann, B., Gobert, A., and Giegé, P.** (2012). PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* **26**: 1022–1027.
- Haïli, N., Arnal, N., Quadrado, M., Amiar, S., Tcherkez, G., Dahan, J., Briozzo, P., Colas Des Francs-Small, C., Vrielynck, N., and Mireau, H.** (2013). The pentatricopeptide repeat MTSF1 protein stabilizes the nad4 mRNA in Arabidopsis mitochondria. *Nucleic Acids Res.* **41**: 6650–6663.
- Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T., and Giegé, P.** (2014). Helical repeats modular proteins are major players for organelle gene expression. *Biochimie* **100**: 141–150.
- Hammani, K. and Giegé, P.** (2014). RNA metabolism in plant mitochondria. *Trends Plant Sci.* **19**: 380–389.
- Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giegé, P.** (2011). An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* **23**: 730–740.
- Hammani, K., Takenaka, M., Miranda, R., and Barkan, A.** (2016). A PPR protein in the PLS subfamily stabilizes the 5'-end of processed rpl16 mRNAs in maize chloroplasts. *Nucleic Acids Res.* **44**: 4278–4288.
- Harrington, K.M., Nazarenko, I.A., Dix, D.B., Thompson, R.C., and Uhlenbeck, O.C.**

- (1993). In vitro analysis of translational rate and accuracy with an unmodified tRNA. *Biochemistry* **32**: 7617–7622.
- Hartmann, R.K., Heinrich, J., Schlegl, J., and Schuster, H.** (1995). Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 5822–5826.
- Van Haute, L. et al.** (2016). Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nat. Commun.* **7**: 12039–12049.
- Hazle, T. and Bonen, L.** (2007). Comparative Analysis of Sequences Preceding Protein-Coding Mitochondrial Genes in Flowering Plants. *Mol. Biol. Evol.* **24**: 1101–1112.
- Hedberg, A. and Johansen, S.D.** (2013). Nuclear group I introns in self-splicing and beyond. *Mob. DNA* **4**: 17–27.
- Helm, M., Brulé, H., Degoul, F., Capanec, C., Leroux, J.P., Giege, R., and Florentz, C.** (1998). The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res.* **26**: 1636–1643.
- Helm, M., Giege, R., and Florentz, C.** (1999). A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNA(Lys). *Biochemistry* **38**: 13338–13346.
- Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., and Gagliardi, D.** (2006). Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase. *Mol. Cell. Biol.* **26**: 2869–2876.
- Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmann, W.** (2008). RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. *Cell* **135**: 462–474.
- Hou, Y.M.** (2010). CCA addition to tRNA: Implications for tRNA quality control. *IUBMB Life* **62**: 251–260.
- Howard, M.J., Klemm, B.P., and Fierke, C.A.** (2015). Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.* **290**: 13454–13464.
- Howard, M.J., Lim, W.H., Fierke, C.A., and Koutmos, M.** (2012). Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci* **109**: 16149–16154.
- Ichikawa, K., Shimizu, A., Okada, R., Satbhai, S.B., and Aoki, S.** (2008). The plastid

- sigma factor SIG5 is involved in the diurnal regulation of the chloroplast gene *psbD* in the moss *Physcomitrella patens*. *FEBS Lett.* **582**: 405–409.
- Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., Kobori, M., Takeba, G., Nakahira, Y., and Shiina, T.** (2005). A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *Plant J.* **42**: 133–144.
- Isono, K., Shimizu, M., Yoshimoto, K., Niwa, Y., Satoh, K., Yokota, A., and Kobayashi, H.** (1997a). Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of sigma70 factors of bacterial RNA polymerases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 14948–14953.
- Isono, M., Wakabayashi, Y., Fujiki, M., Inoue, R., Asakuno, K., and Hori, S.** (1997b). Evidence of apoptotic procedure in deafferented striatum after cortical injury in young adult rats. *Brain Res.* **765**: 319–323.
- Jonietz, C., Forner, J., Hölzle, A., Thuss, S., and Binder, S.** (2010). RNA PROCESSING FACTOR2 is required for 5' end processing of nad9 and cox3 mRNAs in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **22**: 443–453.
- Kasai, K., Kawagishi-Kobayashi, M., Teraishi, M., Ito, Y., Ochi, K., Wakasa, K., and Tozawa, Y.** (2004). Differential expression of three plastidial sigma factors, OsSIG1, OsSIG2A, and OsSIG2B, during leaf development in rice. *Biosci. Biotechnol. Biochem.* **68**: 973–977.
- Kazantsev, A. V and Pace, N.R.** (2006). Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.* **4**: 729–740.
- Kehrein, K., Vargas Möller-Hergt, B., and Ott, M.** (2015). The MIOREX complex - lean management of mitochondrial gene expression. *Oncotarget* **6**: 16806–16807.
- Khrouchtchova, A., Monde, R.-A., and Barkan, A.** (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* **18**: 1197–1209.
- Kim, D., Langmead, B., and Salzberg, S.L.** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**: 357–360.
- Kirino, Y., Yasukawa, T., Ohta, S., Akira, S., Ishihara, K., Watanabe, K., and Suzuki, T.** (2004). Codon-specific translational defect caused by a wobble modification deficiency in mutant tRNA from a human mitochondrial disease. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 15070–15075.
- Klipcan, L., Moor, N., Finarov, I., Kessler, N., Sukhanova, M., and Safro, M.G.** (2012).

- Crystal Structure of Human Mitochondrial PheRS Complexed with tRNAPhe in the Active “Open” State. *J. Mol. Biol.* **415**: 527–537.
- Koh, D.C.-Y., Wong, S.-M., and Liu, D.X.** (2003). Synergism of the 3'-untranslated region and an internal ribosome entry site differentially enhances the translation of a plant virus coat protein. *J. Biol. Chem.* **278**: 20565–20573.
- Kolovos, P., Knoch, T. a, Grosveld, F.G., Cook, P.R., and Papantonis, A.** (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**: 1–8.
- Kotera, E., Tasaka, M., and Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**: 326–330.
- Kouzuma, Y., Mizoguchi, M., Takagi, H., Fukuhara, H., Tsukamoto, M., Numata, T., and Kimura, M.** (2003). Reconstitution of archaeal ribonuclease P from RNA and four protein components. *Biochem. Biophys. Res. Commun.* **306**: 666–673.
- Kramerov, D.A. and Vassetzky, N.S.** (2005). Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* **247**: 165–221.
- Kucukural, A., Özadam, H., Singh, G., Moore, M.J., and Cenik, C.** (2013). ASPeak: An abundance sensitive peak detection algorithm for RIP-Seq. *Bioinformatics* **29**: 2485–2486.
- Kühn, K., Bohne, A.-V., Liere, K., Weihe, A., and Börner, T.** (2007). Arabidopsis phage-type RNA polymerases: accurate in vitro transcription of organellar genes. *Plant Cell* **19**: 959–971.
- Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Börner, T., Millar, A.H., Small, I.D., and Whelan, J.** (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **21**: 2762–2779.
- Kühn, K., Weihe, A., and Börner, T.** (2005). Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Res.* **33**: 337–346.
- Kurz, J.C., Niranjanakumari, S., and Fierke, C.A.** (1998). Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA(Asp). *Biochemistry* **37**: 2393–2400.
- Lagrandeur, T.E., Darr, S.C., Haas, E.S., and Pace, N.R.** (1993). Characterization of the RNase P RNA of *Sulfolobus acidocaldarius*. *J. Bacteriol.* **175**: 5043–5048.
- Lai, L.B., Vioque, A., Kirsebom, L.A., and Gopalan, V.** (2010). Unexpected diversity of RNase P, an ancient tRNA processing enzyme: Challenges and prospects. *FEBS Lett.*

584: 287–296.

- Lamichhane, T.N., Blewett, N.H., Crawford, A.K., Cherkasova, V.A., Iben, J.R., Begley, T.J., Farabaugh, P.J., and Maraia, R.J.** (2013). Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast. *Mol. Cell. Biol.* **33**: 2918–2929.
- Lange, H., Sement, F.M., Canaday, J., and Gagliardi, D.** (2009). Polyadenylation-assisted RNA degradation processes in plants. *Trends Plant Sci.* **14**: 497–504.
- Lange, H., Sement, F.M., and Gagliardi, D.** (2011). MTR4, a putative RNA helicase and exosome co-factor, is required for proper rRNA biogenesis and development in *Arabidopsis thaliana*. *Plant J.* **68**: 51–63.
- Leaver, C.J. and Harmey, M.A.** (1976). Higher-plant mitochondrial ribosomes contain a 5S ribosomal ribonucleic acid component. *Biochem. J.* **157**: 275–277.
- Lechner, M., Rossmanith, W., Hartmann, R.K., Thölken, C., Gutmann, B., Giegé, P., and Gobert, A.** (2015). Distribution of ribonucleoprotein and protein-only RNase P in Eukarya. *Mol. Biol. Evol.* **32**: 3186–3193.
- Li-Pook-Than, J. and Bonen, L.** (2006). Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res.* **34**: 2782–2790.
- Li, Q., Yan, C., Xu, H., Wang, Z., Long, J., Li, W., Wu, J., Yin, P., and Yan, N.** (2014). Examination of the Dimerization States of the Single-stranded RNA Recognition Protein Pentatricopeptide Repeat 10 (PPR10). *J. Biol. Chem.* **289**: 31503–31512.
- Li, X. and Guan, M.-X.** (2002). A human mitochondrial GTP binding protein related to tRNA modification may modulate phenotypic expression of the deafness-associated mitochondrial 12S rRNA mutation. *Mol. Cell. Biol.* **22**: 7701–7711.
- de Longevialle, A.F., Meyer, E.H., Andrés, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D.** (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana*. *Plant Cell* **19**: 3256–3265.
- Lurin, C. et al.** (2004). Genome-Wide Analysis of *Arabidopsis* Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. *Plant Cell Online* **16**: 2089–2103.
- Maffey, L., Degand, H., and Boutry, M.** (1997). Partial purification of mitochondrial ribosomes from broad bean and identification of proteins encoded by the mitochondrial genome. *Mol. Gen. Genet.* **254**: 365–371.
- Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R.** (2012). An

- Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria. *Plant Cell* **24**: 3087–3105.
- Manuell, A.L., Quispe, J., and Mayfield, S.P.** (2007). Structure of the chloroplast ribosome: Novel domains for translation regulation. *PLoS Biol.* **5**: 1785–1797.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D.** (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 12246–12251.
- Maurus, F., Quesneville, H., Bleasby, A., Ware, D., and Leitch, I.J.** (2014). Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. *Nat. Commun.* **5**: 276–277.
- Meurer, J., Lezhneva, L., Amann, K., Godel, M., Bezhani, S., Sherameti, I., and Oelmüller, R.** (2002). A peptide chain release factor 2 affects the stability of UGA-containing transcripts in Arabidopsis chloroplasts. *Plant Cell* **14**: 3255–3269.
- Michaud, M., Cognat, V., Duchêne, A.M., and Maréchal-Drouard, L.** (2011). A global picture of tRNA genes in plant genomes. *Plant J.* **66**: 80–93.
- Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.-H., Hinnebusch, A.G., and Lorsch, J.R.** (2010). The 5'-7-Methylguanosine Cap on Eukaryotic mRNAs Serves Both to Stimulate Canonical Translation Initiation and to Block an Alternative Pathway. *Mol. Cell* **39**: 950–962.
- Miura, E., Kato, Y., Matsushima, R., Albrecht, V., Laalami, S., and Sakamoto, W.** (2007). The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in Arabidopsis yellow variegated mutants. *Plant Cell* **19**: 1313–1328.
- Moore, M.J. and Proudfoot, N.J.** (2009). Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell* **136**: 688–700.
- Moras, D., Comarmond, M.B., Fischer, J., Weiss, R., Thierry, J.C., Ebel, J.P., and Giege, R.** (1980). Crystal structure of yeast tRNA^{Asp}. *Nature* **288**: 669–674.
- Moss, T., Langlois, F., Gagnon-Kugler, T., and Stefanovsky, V.** (2007). A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* **64**: 29–49.
- Mueller, O. and Schroeder, A.** (2004). RNA Integrity Number (RIN) – Standardization of RNA Quality Control Application. *Nano*: 1–8.
- Nagaike, T., Suzuki, T., Tomari, Y., Takemoto-Hori, C., Negayama, F., Watanabe, K.,**

- and Ueda, T.** (2001). Identification and characterization of mammalian mitochondrial tRNA nucleotidyltransferases. *J. Biol. Chem.* **276**: 40041–40049.
- Nagashima, A., Hanaoka, M., Motohashi, R., Seki, M., Shinozaki, K., Kanamaru, K., Takahashi, H., and Tanaka, K.** (2004a). DNA microarray analysis of plastid gene expression in an Arabidopsis mutant deficient in a plastid transcription factor sigma, SIG2. *Biosci. Biotechnol. Biochem.* **68**: 694–704.
- Nagashima, A., Hanaoka, M., Shikanai, T., Fujiwara, M., Kanamaru, K., Takahashi, H., and Tanaka, K.** (2004b). The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the psbD Blue Light-Responsive Promoter (BLRP) in Arabidopsis thaliana. *Plant Cell Physiol.* **45**: 357–368.
- Nakamura, T., Yagi, Y., and Kobayashi, K.** (2012). Mechanistic Insight into Pentatricopeptide Repeat Proteins as Sequence-Specific RNA-Binding Proteins for Organellar RNAs in Plants. *Plant Cell Physiol.* **53**: 1171–1179.
- Nojima, T., Hirose, T., Kimura, H., and Hagiwara, M.** (2007). The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J. Biol. Chem.* **282**: 15645–15651.
- Nystedt, B. et al.** (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**: 579–584.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I.** (2008). On the Expansion of the Pentatricopeptide Repeat Gene Family in Plants. *Mol. Biol. Evol.* **25**: 1120–1128.
- Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J., and Stroud, R.M.** (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 12648–12653.
- Pan, T., Loria, A., and Zhong, K.** (1995). Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 12510–12514.
- Pang, Y.L.J., Poruri, K., and Martinis, S.A.** (2014). tRNA synthetase: TRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **5**: 461–480.
- Pavlova, L. V., Göbringer, M., Weber, C., Buzet, A., Rossmanith, W., and Hartmann, R.K.** (2012). tRNA Processing by Protein-Only versus RNA-Based RNase P: Kinetic Analysis Reveals Mechanistic Differences. *ChemBioChem* **13**: 2270–2276.
- Peck-Miller, K.A. and Altman, S.** (1991). Kinetics of the processing of the precursor to 4-5

- S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* **221**: 1–5.
- Pélissier, T., Bousquet-Antonelli, C., Lavie, L., and Deragon, J.M.** (2004). Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* **32**: 3957–3966.
- Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M., and Gagliardi, D.** (2004). Two exoribonucleases act sequentially to process mature 3'-ends of *atp9* mRNAs in *Arabidopsis* mitochondria. *J. Biol. Chem.* **279**: 25440–25446.
- Pfalz, J., Ali Bayraktar, O., Prikryl, J., and Barkan, A.** (2009). Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* **28**: 2042–2052.
- Pinel, C., Douce, R., and Mache, R.** (1986). A study of mitochondrial ribosomes from the higher plant *Solanum tuberosum* L. *Mol. Biol. Rep.* **11**: 93–97.
- Pinker, F., Bonnard, G., Gobert, A., Gutmann, B., Hammani, K., Sauter, C., Gegenheimer, P.A., and Giegé, P.** (2013). PPR proteins shed a new light on RNase P biology. *RNA Biol.* **10**: 1457–1468.
- Placido, A., Sieber, F., Gobert, A., Gallerani, R., Giegé, P., and Maréchal-Drouard, L.** (2010). Plant mitochondria use two pathways for the biogenesis of tRNA His. *Nucleic Acids Res.* **38**: 7711–7717.
- Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C., and Nyborg, J.** (1996). Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure* **4**: 1141–1151.
- Pring, D., Brennicke, A., and Schuster, W.** (1993). RNA editing gives a new meaning to the genetic information in mitochondria and chloroplasts. *Plant Mol. Biol.* **21**: 1163–1170.
- Pütz, J., Florentz, C., Benseler, F., and Giege, R.** (1994). A single methyl group prevents the mischarging of a tRNA. *Nat. Struct. Mol. Biol.* **1**: 580–582.
- Raczynska, K.D., Le Ret, M., Rurek, M., Bonnard, G., Augustyniak, H., and Gualberto, J.M.** (2006). Plant mitochondrial genes can be expressed from mRNAs lacking stop codons. *FEBS Lett.* **580**: 5641–5646.
- Randau, L., Schröder, I., and Söll, D.** (2008). Life without RNase P. *Nature* **453**: 120–123.
- Raven, J.A. and Allen, J.F.** (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol.* **4**: 209–214.

- Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragón, A.** (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**: 784–789.
- Reiter, V., Matschkal, D.M.S., Wagner, M., Globisch, D., Kneuttinger, A.C., Müller, M., and Carell, T.** (2012). The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA. *Nucleic Acids Res.* **40**: 6235–6240.
- Rossmannith, W.** (2012). Of P and Z: Mitochondrial tRNA processing enzymes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**: 1017–1026.
- Rossmannith, W. and Karwan, R.M.** (1998). Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. *Biochem. Biophys. Res. Commun.* **247**: 234–241.
- Ruf, M. and Kössel, H.** (1988). Structure and expression of the gene coding for the alpha-subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. *Nucleic Acids Res.* **16**: 5741–5754.
- Ruppel, N.J. and Hangarter, R.P.** (2007). Mutations in a plastid-localized elongation factor G alter early stages of plastid development in *Arabidopsis thaliana*. *BMC Plant Biol* **7**: 37–47.
- Salinas-Giegé, T., Giege, R., and Giegé, P.** (2015). TRNA biology in mitochondria. *Int. J. Mol. Sci.* **16**: 4518–4559.
- Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C.** (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* **581**: 4132–4138.
- Sampson, J.R. and Uhlenbeck, O.C.** (1988). Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **85**: 1033–1037.
- Saze, H., Scheid, O.M., and Paszkowski, J.** (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**: 65–69.
- Scharff, L.B., Childs, L., Walther, D., and Bock, R.** (2011). Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genet.* **7**: e1002155.
- Schelcher, C., Sauter, C., and Giegé, P.** (2016). Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity. *Biomolecules* **6**: 30.
- Schmid, M. and Jensen, T.H.** (2008). The exosome: a multipurpose RNA-decay machine.

Trends Biochem. Sci. **33**: 501–510.

Schmitz-Linneweber, C. and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. **13**: 663–670.

Schneider, A. and Ebert, D. (2004). Covariation of Mitochondrial Genome Size with Gene Lengths: Evidence for Gene Length Reduction During Mitochondrial Evolution. J. Mol. Evol. **59**: 90–96.

Seif, E., Leigh, J., Liu, Y., Roewer, I., Forget, L., and Lang, B.F. (2005). Comparative mitochondrial genomics in zygomycetes: Bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. Nucleic Acids Res. **33**: 734–744.

Seif, E.R., Forget, L., Martin, N.C., and Lang, B.F. (2003). Mitochondrial RNase P RNAs in ascomycete fungi: lineage-specific variations in RNA secondary structure. RNA **9**: 1073–1083.

Serruya, R., Orlovetskie, N., Reiner, R., Dehtiar-Zilber, Y., Wesolowski, D., Altman, S., and Jarrous, N. (2015). Human RNase P ribonucleoprotein is required for formation of initiation complexes of RNA polymerase III. Nucleic Acids Res. **43**: 5442–5450.

Sherrer, R.L., Araiso, Y., Aldag, C., Ishitani, R., Ho, J.M.L., Söll, D., and Nureki, O. (2011). C-terminal domain of archaeal O-phosphoserine-tRNA kinase displays large-scale motion to bind the 7-bp D-stem of archaeal tRNA^{Sec}. Nucleic Acids Res. **39**: 1034–1041.

Shimotohno, K., Kodama, Y., Hashimoto, J., and Miura, K.I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. Proc. Natl. Acad. Sci. U. S. A. **74**: 2734–2738.

Simonović, M. and Steitz, T.A. (2009). A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. Biochim. Biophys. Acta - Gene Regul. Mech. **1789**: 612–623.

Sissler, M., Helm, M., Frugier, M., Giege, R., and Florentz, C. (2004). Aminoacylation properties of pathology-related human mitochondrial tRNA(Lys) variants. RNA **10**: 841–853.

Small, I.D. and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. **25**: 46–47.

Steitz, T. a and Steitz, J. a (1993). A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. U. S. A. **90**: 6498–6502.

Stoll, B. and Binder, S. (2016). Two NYN domain containing putative nucleases are

- involved in transcript maturation in Arabidopsis mitochondria. *Plant J.* **85**: 278–288.
- Sugiura, M., Hirose, T., and Sugita, M.** (1998). Evolution and mechanism of translation in chloroplasts. *Annu. Rev. Genet.* **32**: 437–459.
- Sultan, L.D. et al.** (2016). The Reverse Transcriptase/RNA Maturase Protein MatR Is Required for the Splicing of Various Group II Introns in Brassicaceae Mitochondria. *Plant Cell* **28**: 2805–2829.
- Sun, F.-J., Fleurdépine, S., Bousquet-Antonelli, C., Caetano-Anollés, G., and Deragon, J.-M.** (2007). Common evolutionary trends for SINE RNA structures. *Trends Genet.* **23**: 26–33.
- Sun, T., Bentolila, S., and Hanson, M.R.** (2016). The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci.* **21**: 962–973.
- Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K.** (2013). Improved Computational Target Site Prediction for Pentatricopeptide Repeat RNA Editing Factors. *PLoS One* **8**: e65343.
- Tama, F. and Sanejouand, Y.-H.** (2001). Conformational change of proteins arising from normal mode calculations. *Protein Eng. Des. Sel.* **14**: 1–6.
- Taschner, A., Weber, C., Buzet, A., Hartmann, R.K., Hartig, A., and Rossmannith, W.** (2012). Nuclear RNase P of *Trypanosoma brucei*: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Rep.* **2**: 19–25.
- Thomas, B.C., Li, X., and Gegenheimer, P.** (2000). Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. *RNA* **6**: 545–553.
- Tiller, N. and Bock, R.** (2014). The translational apparatus of plastids and its role in plant development. *Mol. Plant* **7**: 1105–1120.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W.** (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**: 123–135.
- Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragón, A.** (2006). Structure of ribonuclease P - a universal ribozyme. *Curr. Opin. Struct. Biol.* **16**: 327–335.
- Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A.** (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15**: 57–61.
- Urbonavičius, J., Qian, Q., Durand, J.M.B., Hagervall, T.G., and Björk, G.R.** (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **20**: 4863–4873.

- Urbonavicius, J., Stahl, G., Durand, J.M.B., Ben Salem, S.N., Qian, Q., Farabaugh, P.J., and Björk, G.R.** (2003). Transfer RNA modifications that alter +1 frameshifting in general fail to affect -1 frameshifting. *RNA* **9**: 760–768.
- Uyttewaal, M., Arnal, N., Quadrado, M., Martin-Canadell, A., Vrielynck, N., Hiard, S., Gherbi, H., Bendahmane, A., Budar, F., and Mireau, H.** (2008). Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* **20**: 3331–3345.
- Vilardo, E., Nachbagauer, C., Buzet, A., Taschner, A., Holzmann, J., and Rossmannith, W.** (2012). A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase-extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* **40**: 11583–11593.
- Villarroya, M. et al.** (2008). Characterization of human GTPBP3, a GTP-binding protein involved in mitochondrial tRNA modification. *Mol. Cell. Biol.* **28**: 7514–7531.
- Wang, M., Davis, N., and Gegenheimer, P.** (1988). Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J* **7**: 1567–1574.
- Weber, C., Hartig, A., Hartmann, R.K., and Rossmannith, W.** (2014). Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet.* **10**: e1004506.
- Weihe, A.** (2004). The transcription of plant organelle genomes. In *Molecular Biology and Biotechnology of Plant Organelles* (Springer Netherlands: Dordrecht), pp. 213–237.
- Weis, B.L., Kovacevic, J., Missbach, S., and Schleiff, E.** (2015). Plant-Specific Features of Ribosome Biogenesis. *Trends Plant Sci.* **20**: 729–740.
- Will, C.L. and Lührmann, R.** (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**: a003707.
- Woolford, J.L. and Baserga, S.J.** (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195**: 643–681.
- Wu, S., Mao, G., and Kirsebom, L.** (2016). Inhibition of Bacterial RNase P RNA by Phenothiazine Derivatives. *Biomolecules* **6**: 38–54.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T.** (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* **8**: e57286.
- Yamaguchi, K., Von Knoblauch, K., and Subramanian, A.R.** (2000). The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**: 28455–28465.

- Yamaguchi, K. and Subramanian, A.R.** (2000). The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**: 28466–28482.
- Yan, J., Zhang, Q., Guan, Z., Wang, Q., Li, L., Ruan, F., Lin, R., Zou, T., and Yin, P.** (2017). MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat. Plants* **3**: 17037.
- Yarham, J.W. et al.** (2014). Defective i6A37 Modification of Mitochondrial and Cytosolic tRNAs Results from Pathogenic Mutations in TRIT1 and Its Substrate tRNA. *PLoS Genet.* **10**: e1004424.
- Yin, P. et al.** (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**: 168–171.
- Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., and Miyazawa, T.** (1985). Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc. Natl. Acad. Sci. U. S. A.* **82**: 4905–4909.
- Yoshihisa, T.** (2014). Handling tRNA introns, archaeal way and eukaryotic way. *Front. Genet.* **5**: 213–229.

Résumé

La publication de la séquence du génome d'*Arabidopsis thaliana* en 2000 a mis en évidence une nouvelle famille de protéines : les protéines à pentatricopeptide repeat (PPR). Ces protéines sont présentes chez tous les eucaryotes, mais particulièrement nombreuses chez les plantes. On en compte par exemple 470 chez *Arabidopsis*. Un sous-groupe de ces protéines PPR portant une activité RNase P a été identifié. Ces protéines appelées PRORP pour "Protein only RNase P" ont été initialement décrites chez les mitochondries humaines et chez *Arabidopsis*. Avant l'identification des protéines PRORP, il était admis que l'activité RNase P qui consiste en la maturation en 5' des pre-ARNt soit universellement portée par des particules ribonucléoprotéiques (RNP) composées d'une sous-unité ARN portant l'activité catalytique (un ribozyme) et d'une ou plusieurs sous-unités protéiques. Ce dogme a été remis en question par la découverte des protéines PRORP. Les protéines PRORP possèdent deux domaines principaux. Le premier contient les motifs PPR, dont la fonction serait de lier l'ARNt et de conférer la spécificité, et le second contient un motif métallonucléase de type NYN responsable de l'activité catalytique de l'enzyme. Des études mécanistiques réalisés au laboratoire ont suggéré que PRORP reconnaît des bases conservées dans les boucles du D et du T de l'ARNt. Mais la façon dont cette reconnaissance est réalisée par les motifs PPR était inconnue.

Le premier objectif de ma thèse a été de contribuer à démontrer que les motifs PPR de PRORP2, une RNase P nucléaire d'*Arabidopsis*, sont bien responsables de la spécificité de substrat de l'enzyme. Pour cela, les positions 5 et 35 des cinq motifs PPR de PRORP2 ont été mutées. Ces mutations ont été introduites dans des séquences de PRORP2 sauvages, afin de tester leurs effets sur l'activité catalytique RNase P, ainsi que dans des séquences d'un mutant catalytique inactif de PRORP2 afin de pouvoir analyser directement l'effet des mutations sur la liaison à l'ARN.

Un second objectif de la thèse est de participer à l'étude fonctionnelle des protéines PRORP chez *Arabidopsis* en étudiant leur réseau d'interaction, plus particulièrement en déterminant les substrats ARN de PRORP1 dans les mitochondries et en identifiant les partenaires protéiques et ARN de PRORP2 dans le noyau.

L'ensemble de ces travaux ayant pour but de préciser le mode d'action des protéines PRORP et de mieux comprendre la diversité de leurs fonctions et leur intégration avec les autres mécanismes d'expression génétique chez *Arabidopsis*.

Ce travail de thèse a permis de faire avancer les connaissances quant au mode d'action des enzymes PRORP, en proposant un modèle de reconnaissance spécifique des ARNt. En particulier, le travail a mis en évidence l'interaction entre les nucléotides C56 et G18 des pré-ARNt et les motifs PPR 2 et 3 de PRORP2. Ce modèle d'interaction PRORP/ARNt devra être confirmé par la détermination d'une structure à l'échelle atomique du complexe PRORP / ARNt. Ce travail permet aussi d'élargir les connaissances sur le mode d'action des protéines PPR, en particulier pour la reconnaissance d'ARN structurés.

Le travail de thèse a aussi participé à l'élucidation des réseaux d'interactions des protéines PRORP dans les mitochondries et le noyau d'*Arabidopsis*. Ces données permettront de voir comment les activités RNases P des organelles et du noyau des plantes sont intégrées avec les autres processus d'expression génétique et ainsi de révéler la diversité de fonctions des RNases P uniquement protéiques.

Cédric SCHELCHER

Détermination du mode d'action et des substrats de RNases P protéiques chez *Arabidopsis thaliana*

Résumé

L'activité RNase P est l'activité essentielle qui élimine les séquences 5' supplémentaires des précurseurs d'ARN de transfert. "PRORP" (PROteinaceous RNase P) définit une nouvelle catégorie de RNase P uniquement protéique. Avant la caractérisation de PRORP, on pensait que les enzymes RNase P étaient universellement conservées sous forme de ribonucléoprotéines (RNP). La caractérisation de PRORP a révélé une enzyme avec deux domaines principaux, un domaine N-terminal contenant plusieurs motifs PPR et un domaine NYN C-terminal portant l'activité catalytique. Nous avons utilisé une combinaison d'approches biochimiques et biophysiques pour caractériser le complexe PRORP / ARNt. La structure du complexe en solution a été déterminée par diffusion des rayons X aux petits angles (SAXS) et les Kd des interactions de différents mutants de PRORP avec l'ARNt ont été déterminées par ultracentrifugation analytique. Notre analyse révèle un cas intéressant d'évolution convergente. Il suggère que PRORP a développé un processus de reconnaissance de l'ARN similaire à celui des RNase P RNP. Par ailleurs, nous avons mis en place une approche de co-immunoprécipitation de PRORP avec l'ARN afin de définir le spectre de substrats des RNase P protéiques.

Mots clés :

RNase P, ARNt, maturation de l'ARN, biophysique, interactions protéines-ARN, PPR, plantes

Résumé en anglais

RNase P is the essential activity that removes 5'-leader sequences from transfer RNA precursors. "PRORP" (PROteinaceous RNase P) defines a novel category of protein only RNase P. Before the characterization of PRORP, RNase P enzymes were thought to occur universally as ribonucleoproteins (RNP). The characterization of PRORP revealed an enzyme with two main domains, an N-terminal domain containing multiple PPR motifs and a C-terminal NYN domain holding catalytic activity.

We used a combination of biochemical and biophysical approaches to characterize the PRORP / tRNA complex. The structure of the complex in solution was determined by small angle X-ray scattering and Kd values of the PRORP / tRNA interaction were determined by analytical ultracentrifugation. We also analyzed direct interaction of a collection of PPR mutants with tRNA in order to determine the relative importance of individual PPR motifs for RNA binding. This reveals to what extent PRORP target recognition process conforms to the mode of action of PPR proteins interacting with linear RNA. Altogether, our analysis reveals an interesting case of convergent evolution. It suggests that PRORP has evolved an RNA recognition process similar to that of RNP RNase P. Moreover, we also implemented a PRORP-RNA co-immunoprecipitation approach to determine the full extent of PRORP substrates.

Keywords :

RNase P, tRNA, RNA maturation, biophysics, protein-RNA interactions, PPR, plants