Functional interactions of nuclear RNase P in Arabidopsis thaliana

Mathieu Bruggeman

Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire des Plantes – CNRS – UPR2357

THÈSE présentée par :

Mathieu BRUGGEMAN

soutenue le : **10 décembre 2020**

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

Discipline : Sciences de la vie et de la Santé
Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Functional Interactions of Nuclear RNase P in *Arabidopsis thaliana*

-

Interactions fonctionnelles des RNase P nucléaires chez *Arabidopsis thaliana*

THÈSE dirigée par :

M. GIEGÉ Philippe	Docteur, CNRS - Université de Strasbourg
M. GOBERT Anthony	Docteur, CNRS - Université de Strasbourg

RAPPORTEURS :

Mme. BOGUTA Magdalena	Professeur, Académie Polonaise des Sciences
M. SCHMITZ-LINNEWEBER Christian	Docteur, Université Humboldt de Berlin

AUTRES MEMBRES DU JURY :

Mme. BOUSQUET-ANTONELLI Cécile	Docteur, CNRS - Université de Perpignan
Mme. FRUGIER Magali	Docteur, CNRS - Université de Strasbourg

Acknowledgments

The author wishes to heartily thank everyone that deserves to be for his or her help, encouragement or support during the course of this PhD work.

He also deeply apologizes to anyone he might accidentally have forgotten.

Finally, the author wishes to publicly acknowledge the good fortune he had, and that now sadly no longer exists, to be able to obtain his Bachelor of Life science by distance learning at the Université Pierre et Marie Curie (Paris-VI).

Abbreviations

2D : 2 dimensional
3D : 3 dimensional
Å : angstrom
aa : amino acid
aaRS : aminoacyl-tRNA synthetases
AMP : adenosine monophosphate
ATP : adenosine triphosphate
BMV : Brome mosaic virus
CCase : CCA nucleotidyltransferase
cDNA : complementary DNA
CDS : coding sequence
CMV : Cucumber mosaic virus
Co-IP : co-immuno-precipitation
Col-0 : Columbia ecotype
CRISPR : Clustered regularly interspaced short palindromic repeats
C-RT-PCR : circular RT-PCR
Cryo-EM : cryo-electron microscopy
DNA : deoxyribonucleic acid
DNase : Deoxyribonuclease
eIF : eukaryotic initiation factor
GFP : green fluorescent protein
GTP : guanosine triphosphate
HA : Human influenza hemagglutinin
HARP : homolog of Aquifex aeolicus
HCV : hepatitis C virus
HIV : human immunodeficiency viruses
kDA : kilo Dalton
KO : knockout
LB : medium : lysogeny broth medium
LC : liquid chromatography
lncRNA : long non-coding RNA
MCS : multiple cloning site
Mg : magnesium
mRNA : messenger RNA
MS : mass spectrometry
MS medium : Murashige and Skoog medium
ncRNA : non-coding RNA
NGS : next generation sequencing
NLS : nuclear localization signal (or sequence)
NYN : N4BP1 YacP-like Nuclease
ORF : open reading frame
PABP : poly(A) binding protein
PCR : polymerase chain reaction
PDB : protein data bank
PDS : phytoene desaturase
PIC : preinitiation complex
PPR : pentatricopeptide repeat
Pre-tRNA : precursor RNA
PRORP : protein-only RNase P
PTC : peptidyl transferase center
PVDF : polyvinylidene fluoride
qPCR : quantitative PCR
q.s.p : *Quantum Satis* (as much as is sufficient)
RACE : rapid amplification of complementary ends
RNA : ribonucleic acid
RNAP : ribonucleic acid polymerase
RNase : ribonuclease
RNase MRP : ribonuclease mitochondrial RNA processing
RNase P : ribonuclease P
RNase Z : ribnuclease Z
RNP : ribonucleoprotein
RPM : revolutions per minute
RT : reverse transcription
RT : room temperature

RTD pathway : rapid tRNA degradation pathway

rRNA : ribosomal RNA

SAM : S-adenosylmethionine

SAXS : small angle X-ray scattering

snRNA : small nuclear RNA

sncRNA : small non coding RNA

snoRNA : small nucleolar RNA

SRP : signal recognition particle

TAL : transcription activator-like

TAIR : the Arabidopsis information ressource

TC : ternary complex

T-DNA : transfer DNA

TEM : transmission electron microscopy

TLS : tRNA-like structure

TF : transcription factor

TMV : tobacco mosaic virus

TRV : Tobacco rattle virus

TOR : target of rapamycin

tRFs : tRNA fragments

TRM1 : tRNA (N^2,N^2-guanine)-dimethyl transferase

tRNAs : transfer RNA

TYMV : turnip yellow mosaic virus

UV : ultraviolet

VIGS : virus induced gene silencing

WT : wild type

Zn : zinc

List of figures

List of tables

Introduction

Forewords

tRNAs are transcribed by RNA polymerase III in the nucleus of Eukaryotes. They then need to undergo numerous maturation steps by a set of specialised enzymes to become functional. The first of these maturation steps is frequently an endoribonucleotlytic cleavage of the 5' leader sequence of precursor tRNAs (pre-tRNA), performed by an enzyme called RNase P.

My PhD work was on the characterization of the functional interactions of nuclear RNase P in *Arabidopsis thaliana*. In this model plant, RNase P activity is performed in the nucleus by two redundant proteins called PRORP2 and 3. The overall aim of my work was to define the protein partners of AtPRORP2 and 3 in order to identify how RNase P activity is integrated with other cellular processes and to determine if plant nuclear RNase P has other substrates besides precursor tRNAs.

In order to present the current state of knowledge on the biological functions and mechanisms related to my PhD project, I will first present how tRNAs are transcribed and maturated. Then their structure and function will be described (I). Next, I will present the distribution, diversity and functions of tRNA-like structures (II). Since nuclear tRNAs are transcribed by RNA polymerase III (RNA pol III), I will present the specific features of this enzyme, as well as the variety of RNAs that this polymerase transcribes (III). MAF1 is the main know negative regulator of RNA pol III, and our work has uncovered a possible plant specific regulation of its activity by nuclear RNase P. Consequently, I will present MAF1 function and mode of action at the cellular and molecular level (IV). Then I will present in a last part of the introduction, two different steps of tRNA maturation that are specifically related to my PhD. First, I will present in details the diversity of RNase P enzymes that remove the 5' leader sequence of precursor tRNA (V). Then, I will present more briefly the chemical modification of tRNAs nucleotides and more specifically, the reaction catalysed by an enzyme called Trm1, that was shown in this work, to be a protein partner of AtPROP2/3 (VI).

I. tRNAs: biogenesis, structure and functions

First characterized in 1958 as "soluble ribonucleic acid intermediates in protein synthesis" (Hoagland et al., 1958), tRNAs were the first non-coding RNAs (ncRNAs) discovered. tRNAs have a very specific structure (a), that is closely related to their function. The main role of tRNAs is to deliver amino acids, as specified by messenger RNA (mRNA) codons, to ribosomes. However, tRNAs have additional functions that are progressively uncovered (b). Finally, the different post-transcriptional steps involved in tRNA biogenesis and turnover will be presented (c).

a) tRNA structure

tRNAs have a conserved secondary structure often referred to as a "cloverleaf'. Their 3D architecture is an L-shaped fold that is also conserved (Figure 1). However, this 3D structure is rather dynamic and not fixed, as analyses of crystal structure and cryo-EM reconstruction of tRNAs structures in complex with different partners (processing enzymes, ribosomes, tRNA synthetases) show that tRNAs can display a large range of structural arrangements to perform their functions (Westhof and Auffinger, 2012).

tRNAs length vary from around 70 to 100 nucleotides. As noted by Westhof and Auffinger, (2012), they are evolutionary constrained by two opposite forces. They must have unique features in order for their cognate tRNA synthetase to aminoacylate the correct tRNA with the specified amino acid corresponding to its anticodon. And on the other hand, their 3D structure is constrained by the necessity to fit in the ribosomes binding sites, in order for correct protein synthesis to occur. As such, tRNA structure harbours a common architecture inside which molecular diversity exists to ensure the unicity of each tRNA.

The cloverleaf secondary structure is made up of four helices, three loops and a variable domain. First, the acceptor stem (also called acceptor helix) that carries the amino acid at its 3' end is usually made of seven base pairs. The first four base pairs of the acceptor

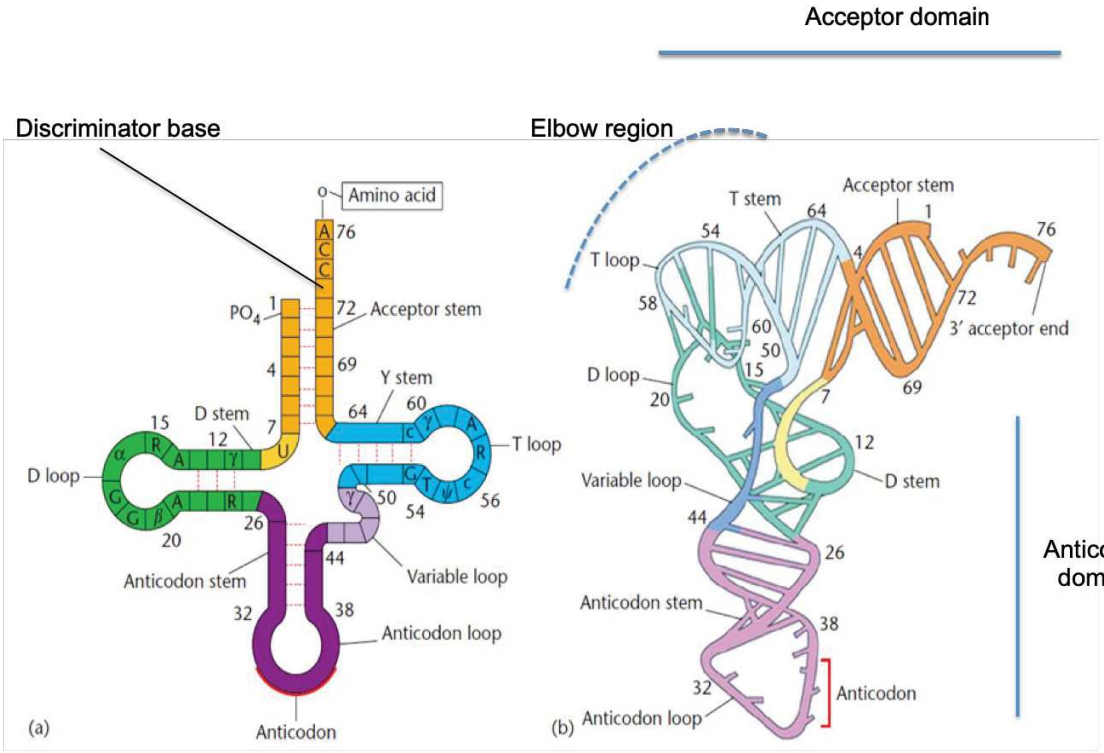
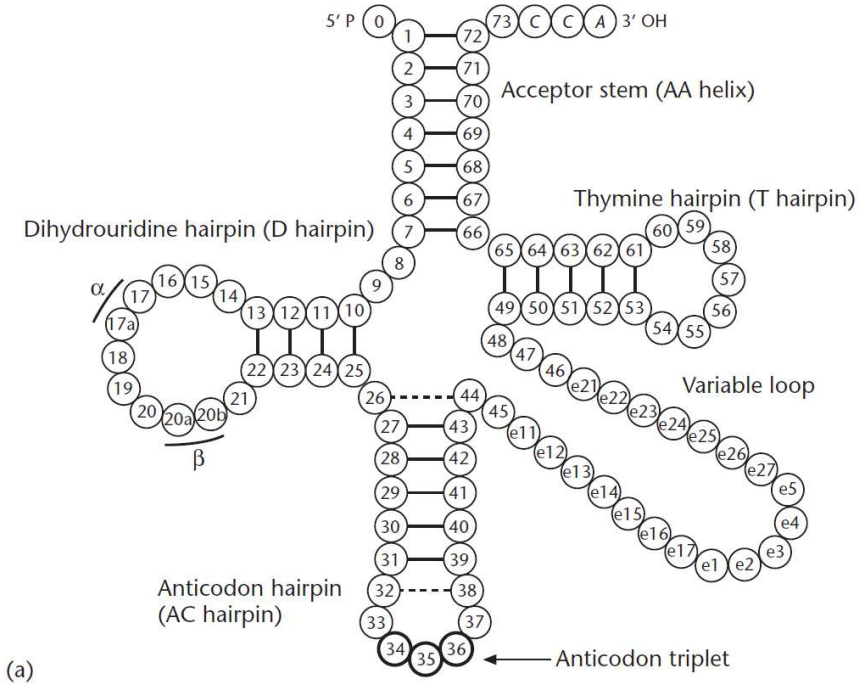


Figure 1: **2D and 3D schematic representation of a tRNA**. a) Cloverleaf diagram of the 2D folding pattern of a tRNA. Numbers indicate the nucleotide base, almost universally found at this position in all tRNA sequences known to date. A = adenosine; G = guanosine; C = cytidine; U = uridine; R = adenosine or guanosine; Y = cytidine or uridine; T = ribothymidine; and Ψ = pseudouridine. b) Folding pattern, in 3D, of yeast phenylalanine tRNA. The coiled tube indicates the sugar-phosphate backbone of the molecule. The cross rungs represent the nucleotide base pairs while the short rungs depict bases not involved in base-base hydrogen bonding. The colours in the 3D diagram refer to the 2D structure. Diagrams are reproduced and adapted from Rich and Kim (1978) and Goldman (2008).

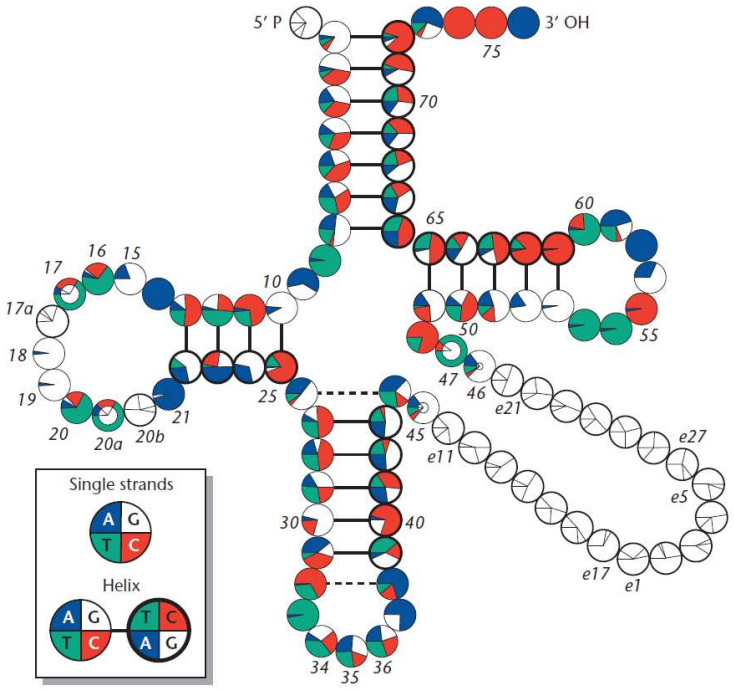
helix are important for tRNA synthetase recognition (Giegé et al. 2012) as noted above. Position 73, located just before the CCA is called the 'discriminator', due to its role in the correct aminoacylation of tRNAs. Then on the 5' side of the cloverleaf, the D (dihydrouridine) arm is named after this modified base that occurs in its loop. The first three base pairs of the D helix also constitute identity elements for aminoacyl synthetases. Then, the anticodon arm contains in its loop the anticodon triplet that will pair with the complementary codon of the mRNA during protein synthesis in the ribosome. The identity of the residues of the anticodon loop are also important for correct aminoacylation. The anticodon arm is followed by the variable region that contains between 4 and 21 residues. This variable loop is the major source of variability in tRNA sizes, hence its name. The last conserved domain defining tRNAs is the T-arm, named after a thymidine present in the conserved TΨC motif of its loop. Finally, the 3'end the mature tRNA is always terminated by the CCA sequence where the amino acid is attached.

tRNA sequences alignment show that in the 2D structure, some positions nearly always consist of the same type of base (called 'invariants'). Some other positions are conserved as either purines or pyrimidines (called 'semi-invariants') (Figure 2). Invariant and semi-invariant positions represent up to 20 nucleotides in tRNAs. The 3D structure of tRNAs strongly relies on the four helices described above. The acceptor helix and the thymidine arm stack together to form the acceptor domain while the dihydrouridine arm stacks together with the anticodon arm to make the anticodon domain. This structure is overall shaped like an L, as the two domains forms an angle of around 90°. This 3D structure is maintained by loop-loop tertiary contacts (between the T and D loops) and by tertiary contacts (involving non-Watson-Crick pairings and the formation of base triplets) between the single stranded junctions linking the helical domains and the deep groove of the D helix. tRNA structures are also rigidified by magnesium cations (Auffinger et al., 2011). Indeed, it was shown that in the absence of Mg^{2+}, some tRNAs could not be amino-acylated (Madore et al., 1999). The role of magnesium cations is also important for, as stated above, the dynamic structural changes that tRNAs can adapt when interacting with their different protein partners (Westhof and Auffinger, 2012).

All these elements point out the importance of tRNAs structure. Indeed, if for mRNA the sequence is considered as the determinant of its biological function, for tRNA it is the



(a)



(b)

Figure 2: **Nomenclature and base distributions in elongator tRNA**. (a) Numbers indicate the accepted nomenclature of tRNA molecules as defined in Sprinzl et al., 1998. Straight black lines indicate secondary base pairing while broken lines indicates unusual base pairing. (b) Distribution of the four common bases at each position as calculated from 932 tRNA sequences by Auffinger and Westhof, 1998. In helices, the colour code for paired residues are arranged so as to follow Watson-Crick pairings. The 5' strand has a thin outer circle while the 3' strand has a thick outer circle (from Westhof and Auffinger, 2012).

structure (i.e the spatial fold of the RNA to form a tertiary structure) that is the basis of its biological activity (Westhof and Auffinger, 2012).

b) tRNAs biological functions

tRNAs canonical function is to deliver amino acids, as specified by messenger RNA (mRNA) codons to the protein synthesis machineries; in the cytoplasm (for Prokaryotes and nuclear eukaryotic mRNA) and in organelles (in Eukaryotes only). To perform this step, tRNAs need to be (correctly) amino-acylated. Indeed, tRNAs can be subdivided into families of isoacceptors. Each defined by the corresponding (usually called "cognate") amino acid it carries. Aminoacylation is performed by a set of specific proteins called aminoacyl-tRNA synthetase. Based on sequence analysis and afterwards on subsequent structure investigations aminoacyl-tRNA synthetases (aaRS) are divided in two distinct classes (class I and II), based on their differences at both the structural and sequence level (Kaiser et al., 2020). Linking of tRNA molecules with their cognate amino acid is one of the most crucial step of protein synthesis, as once the amino acid is charged to the tRNA, it will be incorporated in the elongating peptidic chain in the ribosome, whether or not it is the correct one (Giegé and Eriani, 2012). Aminoacyl-tRNA synthetase recognize their cognate tRNA by specific features defining "tRNA identity". For some tRNAs, the major identity element is the anticodon (Goldman, 2008). However, the main features can also be in the acceptor stem with little to no specificity conferred by the anticodon. Still, in most cases, the major specific feature seems to be the anticodon with additional identity elements residing in the acceptor stem, and/or the D-loop (Goldman, 2008, Westhof and Auffinger, 2012) (Figure 3). Charging of the cognate amino acid to the tRNA is a two-step process. First the amino acid is activated to an aminoacyl adenylate (with a high energy bond). Then this activated molecule is bound to the aminoacyl tRNA synthetase and the amino acid is transferred to the tRNA (the high energy bond is preserved, as it will be necessary for protein synthesis) (Figure 4). Due to the existence of proofreading (also called "editing") mechanisms, that allow aminoacyl tRNA synthetase to reject incorrect aminoacyl adenylates, the aminoacylation process is highly accurate. Indeed, the mischarging frequency has been estimated at less than 10^{-6} (Goldman, 2008).

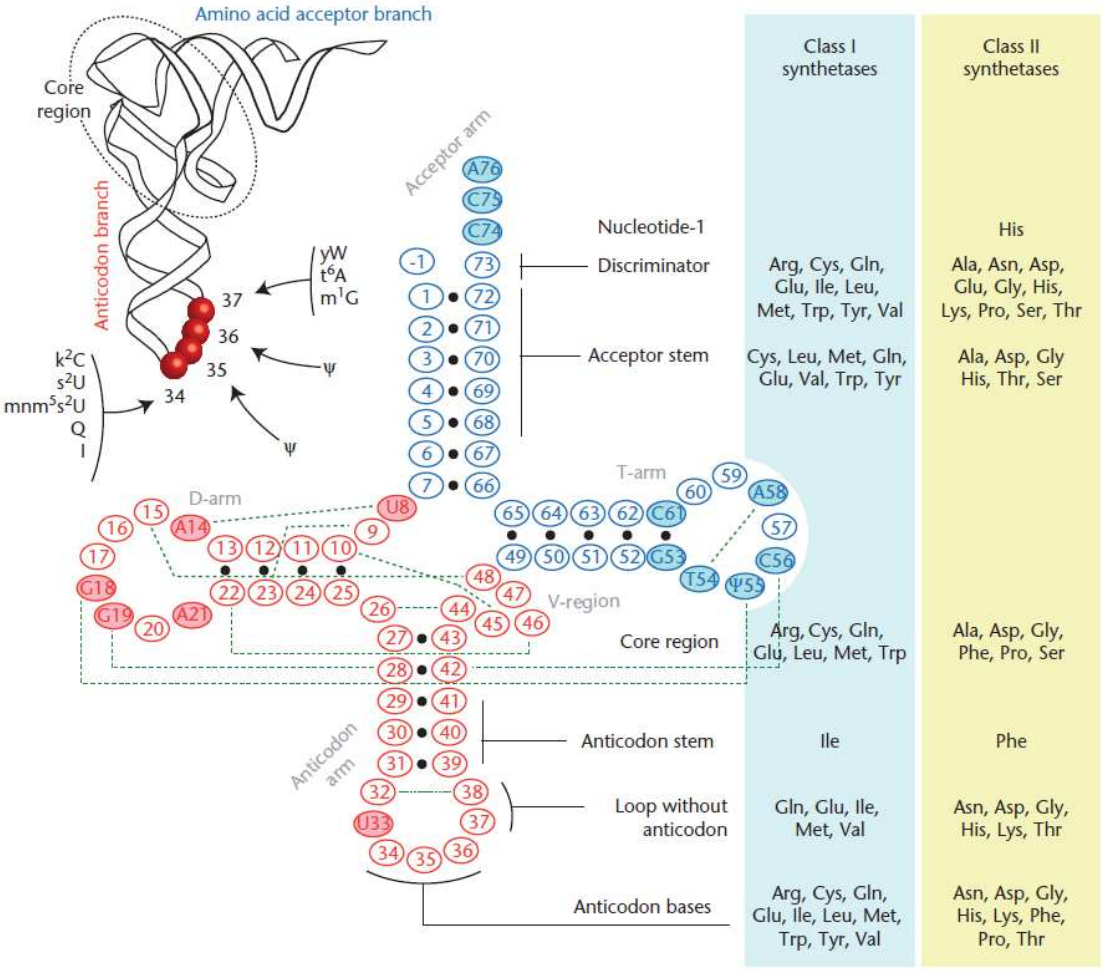


Figure 3: **Cloverleaf and 3-dimensional folding of a tRNA with the location of known identity determinants**. Conventional numbering system is used. The • symbol indicates Watson–Crick base pairings, including G•U pairs; dotted green lines indicate other pairings, mostly between constant and/or semi-constant residues important for tRNA L-shaped architecture. Location of identity elements in the tRNA molecule is shown with a distinction between identities of tRNAs recognised by class I and class II synthetases. Characterised individual modified residues that act as identity determinants in tRNA anticodon loops are shown in the inset (k^2C, lysidine; s^2U, 2-thiouridine; mnm^5s^2U, 5-methylaminomethyl-2-thiouridine; Q, queuosine; I, inosine; Ψ, pseudouridine; m^1G, 1-methylguanosine; t^6A, N6-threonylcarbamoyladenosine; yW, wybutosine). From Giegé and Eriani (2012).

Charged tRNAs will then play a central role in protein synthesis. Even if all the details of protein synthesis are not yet fully understood, numerous studies, including cryo-EM data and X-ray crystallography structures have uncovered the main features of this process. In Eukaryotes, initiation of protein synthesis starts when a 40S ribosome small subunit is combined with an mRNA, an initiator methionyl-tRNA (Met-tRNA^{i}) and initiation factors (Dever and Green, 2012). After the recruitment of the 60S large subunit of ribosomes, the interaction between the codon of the messenger RNA and the anticodon of the initiator tRNA takes place in the peptidyl ('P' site) of the 80S ribosome, and allows the binding of an elongator aminoacyl-tRNA to the aminoacyl ('A') site of the ribosome, this step being the first of the elongation phase of protein synthesis. Although apparently quite simple, translation initiation requires, in Eukaryotes, no less than 12 dedicated proteins called eukaryotic initiation factors (eIFs).

In more details, eIF2-GTP will first bind Met-tRNA^{i} to form the ternary complex (TC). The TC will bind with eIF5, eIF3, eIF1 and eIF1A to the 40S (small) ribosomal subunit; forming the preinitition complex (PIC). On the other hand, the mRNA is activated by binding of eIF4F (a complex formed of eIF4E, eIF4G and eIF4A) at its 5' m^7 (7-methylguanosine) cap and of the PABP (poly(A) binding protein) on the 3' poly(A) tail. After scanning of the PIC on the mRNA, a 'closed' complex will form when base pairing is established between the anticodon of the Met-tRNA^{i} in the P-site and the mRNA AUG codon (Figure 5). And the second codon of the open reading frame (ORF) is in the A site of the ribosome. In stark contrast with the initiation step that requires many initiation factors, the second step of protein synthesis (called "elongation") requires only a minimal set of proteins (called elongation factors). Elongation will start when the cognate aminoacyl-tRNA, bound in a ternary complex with eEF1A and GTP, will be delivered to this A site of the ribosome. GTP is then hydrolysed and the elongation factor released. At this stage, the aa-tRNA is locked in the A site of the large subunit. A peptide bond formation will occur with the P-site peptidyl-tRNA. This reaction will be catalysed by the peptidyl transferase center (PTC) consisting of conserved ribosomal RNA (rRNA) of the large ribosomal subunit. These RNAs will position the substrates for catalysis. The peptydil-tRNA will transfer the growing peptide to the amino acid attached in the A site. The energy for this reaction (called the "peptidyl transferase" reaction or 'peptide bond formation') comes from the high energy bond conserved between the tRNA and its cognate amino acid. Peptide bond formation is the heart of protein synthesis, and will be

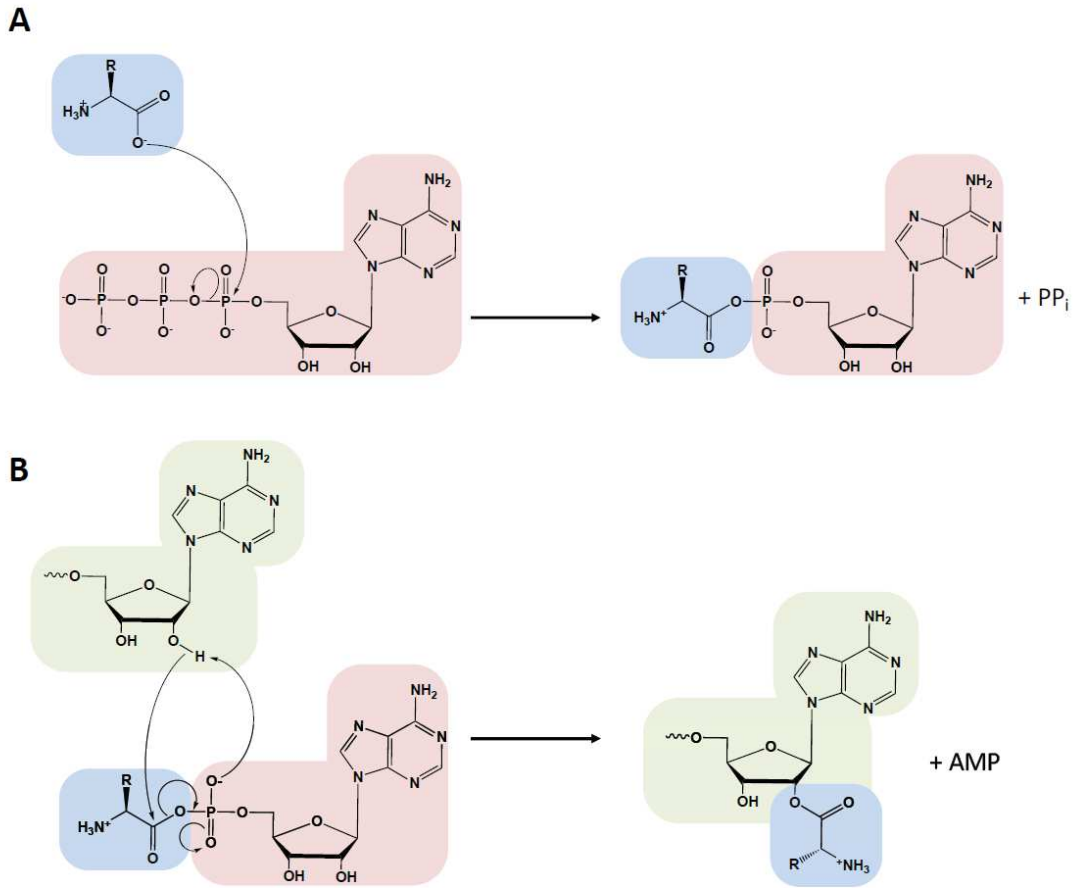


Figure 4: **The aminoacylation reaction**. In the first step, the amino acid (blue) is activated with ATP (red) in the synthetase active site (not depicted), forming the aminoacyl-AMP and releasing PPi. The amino acid is transferred to the tRNA (green) and AMP is released (depicted in the image is the transfer to the 3'-OH characteristic of class I aaRS, while in class II transfer happens at the 2'-OH with a 3'-OH attack in the second step). The diagram is taken from Rubio Gomez and Ibba (2020).

followed by ribosomal conformational changes that will lead the tRNA in the A site to translocate to the P site while the one in the P site will translocate to the exit ('E') site. This translocation is promoted by eEF2. Following translocation, a deacylated tRNA occupies the E site and peptidyl-tRNA is positioned in the P site. And the A site is available again for the next aminoacyl-tRNA. This elongation cycle will then repeats until the nascent protein is completely formed (Dever et al., 2018). It is worth noting that it seems that the deacylated tRNA will be released slowly from the E site (in human ribosomes) following translocation, independently of the binding of A-site aa-tRNA (Ferguson et al., 2015).

Termination of eukaryotic translation will occur when a STOP codon (UAA, UAG, UGA) will enter the A site of the ribosome. It is catalysed by two proteins: eRF1 and eRF3, that are thought to act collaboratively. eRF1 is a tRNA-shaped protein (Song et al., 2000) composed of three domains. A first domain is responsible for codon recognition. A middle one (analogous to a tRNA acceptor stem) that will extend in the PTC, promotes peptide release and a C-terminal domain interacts with eRF3. The binding of eRF3 to eRF1 is thought to increase the peptide release efficiency (Dever and Green, 2012).

It is important to note that, rather than static, the pool of tRNAs is now considered to be dynamic. Different stresses have been shown to alter the global tRNA pool. This is traditionally viewed as a modulation precisely directed to conform to the evolution of the codon usage of stress-related genes (Kirchner and Ignatova, 2014). However, tRNAs also have alternative roles in stress signalling. Under deprivation of amino acids, part of cytoplasmic tRNAs were shown to promptly relocate to the nucleus in Yeast, thus globally repressing translation (Whitney et al., 2007). From a broader perspective, tRNAs are considered to have a role in the general amino acid control pathway where they would act as signalling agents (Dever and Hinnebusch, 2005). In Yeast, during amino acid starvation, non-charged tRNAs activate the GCN2 kinase through binding to its C-terminal domain, which is homologous to histidyl tRNA synthase. GCN2 will in turn phosphorylate eIF2α. This phosphorylation will activate the translation of mRNAs encoding GCN4, which is required for the selective transcription of genes in amino acid biosynthesis pathways to increase amino acid biosynthesis (Murguia et al., 2012 and Kirchner and Ignatova, 2014).

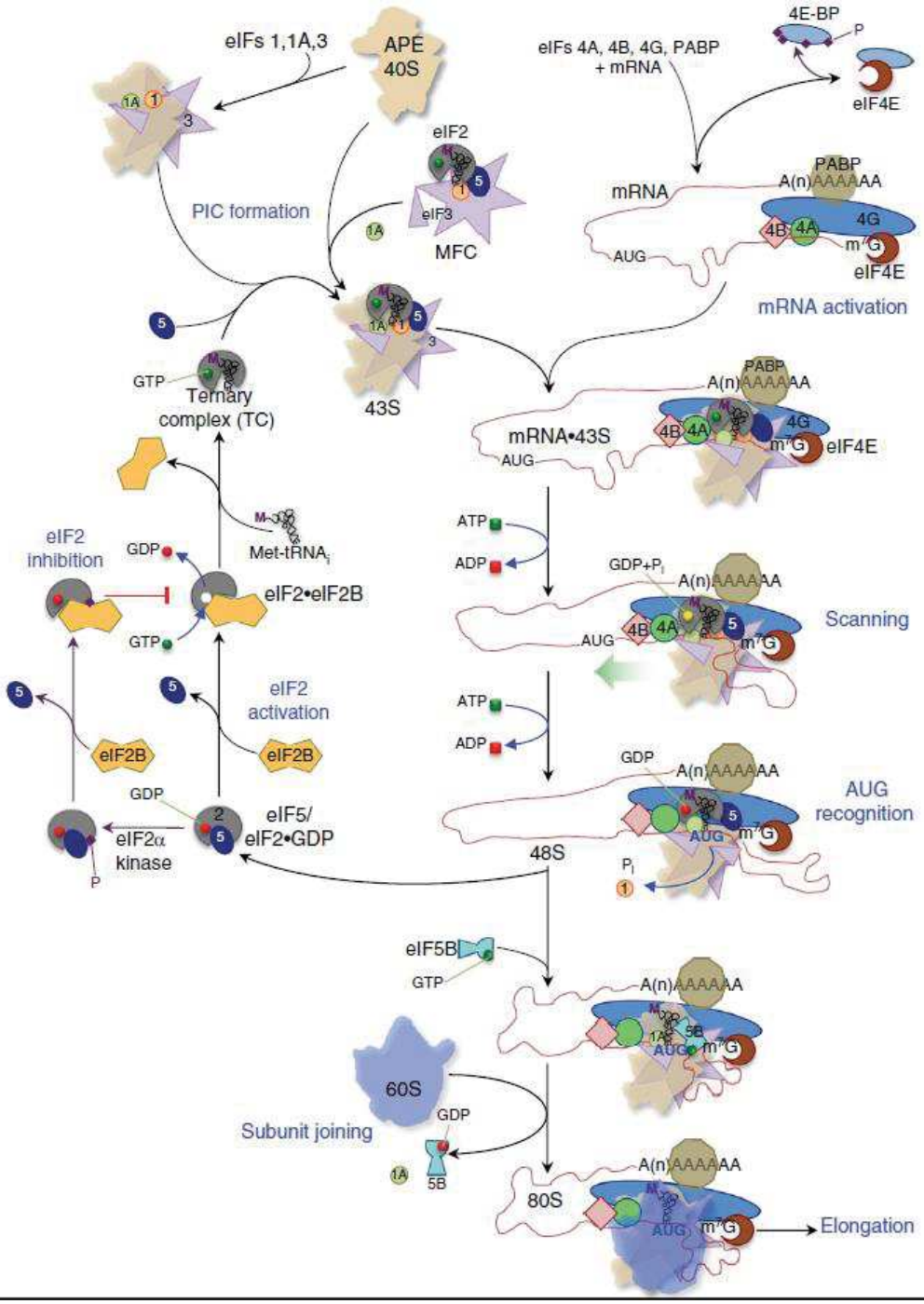


Figure 5: **Overview of the general eukaryotic translation initiation pathway**. The pathway for recruiting the initiator tRNA to the mRNA AUG codon in the context of a 80S ribosome (bottom right) is depicted as a series of major steps, labelled with blue text, linked with black arrows. Individual eukaryotic initiation factor (eIF) cartoons and complexes are labelled with black text and nucleotide hydrolysis/inorganic phosphate release reactions are shown by blue arrows. The broad green arrow indicates the direction of scanning toward the AUG codon. The regulatory reactions leading to eIF2 and eIF4E inhibition are shown with plum and red arrows. The timing of release of some factors from initiating ribosomes/mRNA (eIF4F, eIF4B, or eIF3) is not yet clear, and consequently is not shown. Figure taken from Merrick and Pavitt (2018).

Additional functions of tRNAs beyond protein synthesis are in fact currently constantly extended. For example, tRNAs are now considered as also implicated in targeting proteins for degradation via the N-end rule pathway (Varshavsky, 1997), or considered as important player in the regulation of apoptosis by binding to cytochrome *c* (Mei et al. 2010). It can be noted that tRNAs can also be involved in retroviral replication where they act as reverse transcription primers (Marquet et al., 1995; Piekna-Przybylska et al., 2010). For example, specific tRNAs, such as $tRNA^{Lys}$ for HIV and $tRNA^{Pro}$ for murine leukaemia virus, are packaged in virions and serve as primers (for DNA synthesis) by reverse transcription (Mark et al., 1997 and Kleiman et al., 2010).

Last, but not least, new pathways that generate tRNA fragments (tRFs) are progressively uncovered and the different roles of these small non-coding RNA (sncRNAs) in different cellular processes such as translation regulation and responses to stress are currently investigated. tRNA fragments can be found in all domains of life (Kirchner and Ignatova, 2014). Cleavage of tRNAs to generate tRFs is catalysed by stress inducible endonucleases such as Rny1 (that belong to the RNase T2 family) in Yeast (Thompson et al., 2009) and angiogenin (belonging to the RNase A family) in human (*H. Sapiens*)(Yamasaki et al., 2009). In plants, it was recently shown that the major players in tRFs biogenesis are also RNases T2 (Megel et al., 2018). tRFs can originate from tRNA precursors, but most seem to derive from mature tRNAs. Two main classes of tRFs can be generated. When the cleavage occurs in the anticodon region, a long tRF (30-35 nucleotides) will be produced. On the contrary, when cleavage is located in the D or the T region, a short tRF will be generated (19-25 nucleotides) (Megel et al., 2018). In plants, the most interesting example of tRFs function is, to date, the investigation led by Ren and colleagues that showed that rhizobial tRFs act as signalling molecules to regulate plant nodulation (Ren et al., 2019).

c) tRNAs biogenesis, turnover and intracellular dynamics

Similar to most RNAs, tRNAs are transcribed as precursor molecules (precursor tRNA or pre-tRNA) and need to undergo extensive post-transcriptional maturations in order to become functional. These post-transcriptional steps are performed by a large number of factors. These maturation processes have been extensively studied in the budding yeast

(*Saccharomyces cerevisiae*), including works from the teams of Profs. Hopper, Phizicki and Yoshihisa. This led to the progressive identification of nearly all of the genes encoding factors involved in yeast tRNA maturation and to the progressive elucidation of their functions. Consequently, we will present what is known in yeast and refer, whenever it is possible, to the plant homologs.

tRNAs post-transcriptional maturation steps include the removal of the 5' leader and 3' trailer sequences, the removal of introns (in intron containing pre-tRNAs), the addition of the CCA triplet and numerous chemical modifications of nucleosides. In yeast up to 25 base modifications have been characterized (Hopper, 2013)

Removal of the 5' leader sequence is catalysed by RNase P in all domains of life. In yeast RNase P has interesting features. Similar to the ancestral bacterial ribonucleoprotein RNase P, both the nuclear and mitochondrial enzymes are composed of RNA and proteins. While nuclear RNase P consists of a single catalytic RNA and nine proteins (Hopper, 2013), mitochondrial RNase P is composed of a different RNA and one protein subunit has been identified to date (Daoud, 2012). Interestingly, most of the proteins of yeast (and human) RNase P are shared with RNase MRP, an enzyme involved in pre-rRNA processing (Xiao et al., 2002; Jarrous and Gopalan, 2010). Since RNase P is central to my PhD project, I will present its diversity, i.e. plant RNase P, in more details in a separate section (V).

The 3' trailer sequence of pre-tRNAs is removed in yeast by both exonucleases (LHP1 and REX1) and an endonuclease (TRZ1). The two exonucleases were shown to be non-essential (Hopper, 2013), whereas TRZ1 is encoded by an essential gene (Takaku et al., 2003). In plants, the removal of the 3' trailer is catalysed by the endonuclease RNase Z. RNase Z exists in two forms. Short versions of 200 to 400 amino acids co-exist with long versions of 700 to 900 aa (Canino et al., 2009). Only the short form of RNase Z is found in Prokaryotes whereas Eukaryotes possess the long form and sometime both as found in plants. RNase Z belongs to the superfamily of metallo-β-lactamases (MBLs), characterized by their α-β/β-α fold. Metallo-β-lactamases contain five highly conserved motifs that are involved in the metal coordination in the active site (Aravind, 1999). Work from Canino et al., 2009, showed that *Arabidospsis* genome encodes the 4 short and long versions of RNase Z. According to their data, AtRNaseZ short forms s1 and s2 are located in the cytoplasm and the chloroplast respectively, while the long form L1 is located in mitochondria and the nucleus, and L2 only in mitochondria.

It is worth noting that RNase P's action usually precedes RNase Z. However, there are some known exceptions, as described e.g. by O'Connor and Peebles (1991).

A 3' terminal CCA sequence is required in all tRNAs for aminoacylation. In most cases, this CCA sequence is added post-transcriptionally by a tRNA nucleotidyl transferase. It can be noted that in some bacteria, i.e. *E. coli*, the CCA sequence is encoded within the tRNA genes, however, a tRNA nucleotidyl transferase still exists and was shown to be involved in tRNA 3'end repair (Zhu and Deutscher, 1987). In *S. cerevisiae,* tRNA nucleotidyl transferase is encoded by a single gene (CCA1, Aebi et al. 1990). CCA1 encodes 3 isoforms generated by alternative transcriptional and translational start sites, and located in the cytoplasm, the nucleus and mitochondria (Martin and Hopper, 1994). In Arabidopsis, the tRNA nucleotidyltransferase was also identified (At1g22660). This protein localizes to the cytoplasm, plastids and mitochondria. Its distribution appears to be controlled, at least in part, by the occurrence of alternative translation initiation sites (Von Braun et al., 2007).

Splicing is another important post-transcriptional step for intron containing tRNAs. The number of intron-containing tRNA gene vary considerably between organisms and phyla. *E. coli* tRNA genes do not contain introns. In human (*H. Sapiens*), *C. elegans* or *Drosohila* this percentage is around 5%, and it can be superior to 50% in some archaeal genome (Phizicky and Hopper, 2010). In Arabidopsis, according to the plantRNA database (Cognat et al., 2013), around 12% of nuclear tRNAs possess introns. These introns are all occurring in only two tRNAs: $tRNA^{met\_e}$ et $tRNA^{tyr}$. In some cases, only intron-containing tRNA genes are modified at specific bases (Johnson and Abelson 1983). However, intron deletion was carried out in yeast in a tRNA that did not depend on the presence of an intron for specific base modification and no obvious consequences could be observed from this deletion (Mori et al., 2010). Consequently, the evolutionary reason for retaining introns in some tRNA genes remains an open question. In yeast, pre-tRNA splicing occurs in three steps. In a first step, a tRNA splicing endonuclease removes the intron from the pre-tRNA. This enzyme is in yeast a heterotetramer composed of Sen2, Sen34, sen15 and sen54 (Trotta et al. 1997; Paushkin et al., 2004). Sen2 and Sen34 are conserved from Archaea to human and are the catalytic subunits of the splicing endonuclease responsible for cleavage at, respectively, the 5' and 3' splice sites (Trotta et al., 1997). The second step of the reaction is the ligation of the 5' and 3' exons. It is catalysed by the yeast tRNA ligase Trl1 (Phizicky et al., 1986).

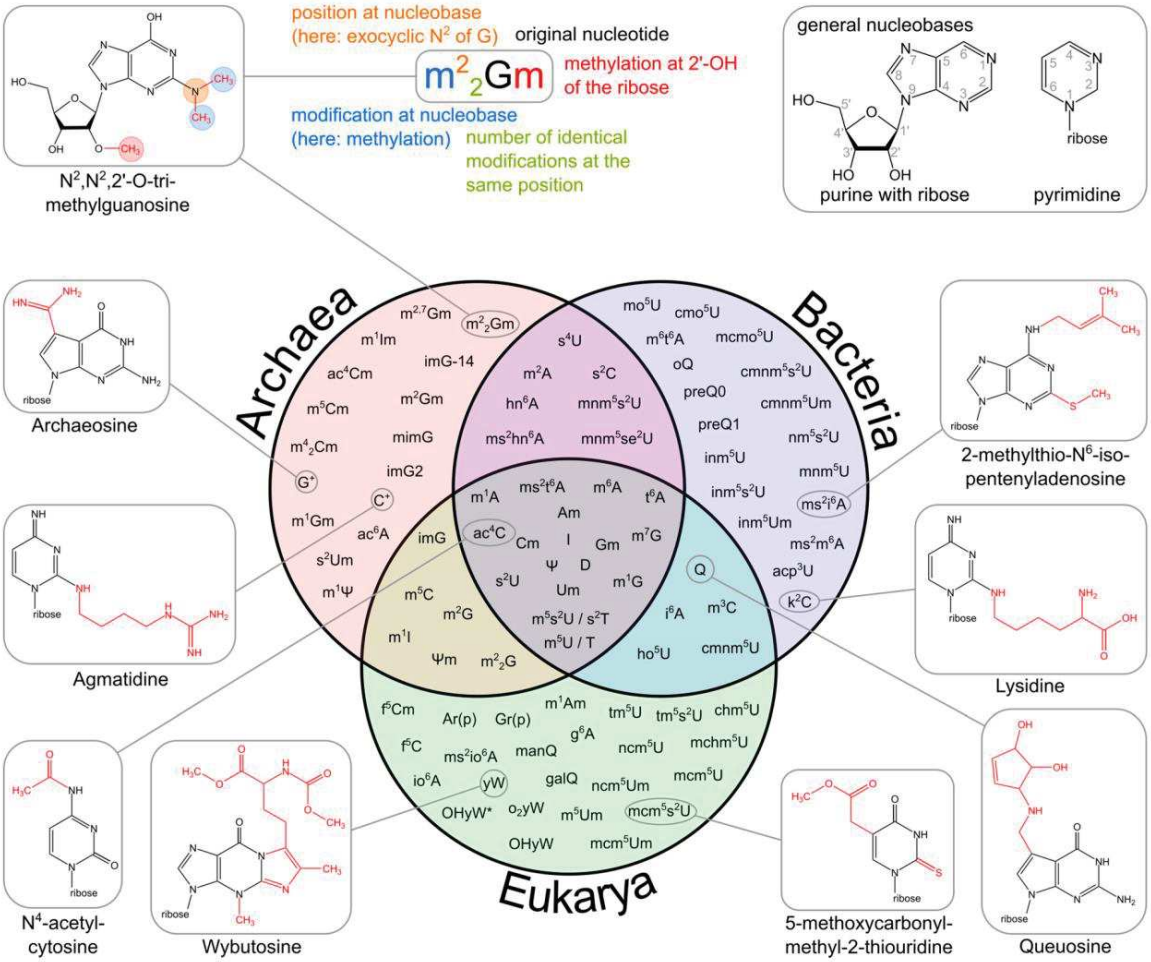


Figure 6: Diversity of tRNA modifications. The upper part of the image explains the abbreviation system used to describe RNA modifications, with N^2,N^2,2′-O-trimethylguanosine ($m^2{}_2Gm$) as an example. It also shows the atom numbering in the purine and pyrimidine rings as well as in the ribose. An abbreviation in front of the base letter describes a base modification, whereas letters after the base stand for ribose alterations. Superscripted numbers specify the position at the base and subscripted numbers indicate the frequency of identical modifications at the same position. Abbreviations are as follows: ac—acetyl, acp—aminocarboxypropyl, chm—carboxyhydroxymethyl, cmo—oxyacetic acid, cmnm—carboxymethylaminomethyl, f—formyl, g—glycinyl, gal—galactosyl, hn—hydroxynorvalylcarbamoyl, ho—hydroxy, i—isopentenyl, inm—isopentenylaminomethyl, io—cis-hydroxyisopentenyl, m—methyl, man—mannosyl, mchm—carboxyhydroxymethyl methyl ester, mcm—methoxycarbonylmethyl, mcmo—oxyacetic acid methyl ester, mnm—methylaminomethyl, mo—methoxy, ncm—carbamoylmethyl, nm—aminomethyl, r(p) —5-O-phosphono-b-d-ribofuranosyl, s—thio, se—seleno, t—threonylcarbamoyl, tm—taurinomethyl. The Venn diagram summarizes data collected from the RNA modification database and contains the 93 post-transcriptional modifications that are found in tRNAs according to the RNA modification database (RNAMDB; https://mods.rna.albany.edu/). Some examples are shown with their chemical structure. Image taken from Lorenz et al., (2017).

This step leaves a residual 2' phosphate at the splice junction that is removed in the third step of splicing, catalysed by the 2' phosphotransferase, encoded by TPT1 (Spinelli et al., 1997). This ligation mechanism is conserved in plants (Gegenheimer et al., 1983; Schwartz et al., 1983; Culver et al., 1994; Englert and Beier, 2005). Finally, occurring at different time during the post-transcriptional maturation of tRNAs, a multitude of nucleoside modifications can happen (Figure 6). More than 85 have been identified in the three kingdoms of life (El yacoubi et al., 2012). And this number keeps increasing as new discoveries are made. Overall, in yeast, because of tRNA modifications, more than 15% of cytoplasmic tRNA nucleotides are not one of the 4 canonical RNA nucleotides (Phizicky and Hopper 2010). Base modifications are considered to serve various functions such as translation fidelity, tRNA stability and tRNA discrimination. The most well known example is the deamination of adenosine (A) to inosine (I) at the wobble position 34. This modification extends codon-anticodon possible interaction as I can base pair with U, C and A (whereas A only pairs with U) (Gerber and Keller, 1999). tRNA modifications are also key for tRNA stability. Indeed, it was shown that tRNAs lacking essential modifications are unstable, and consequently rapidly degraded (Alexandrov et al., 2006). Interestingly even if many tRNA modification enzymes are well conserved and despite the important role of base modifications for tRNAs, most genes coding for tRNA modification enzymes are not essential. In *S. cerevisiae* only 4 genes (TAD2 and TAD3, responsible for adenosine A34 to inosine I34 deamination and TRM6 and TRM61, responsible for methylation of adenosine m^1A_{58}) were shown to be essential (Anderson et al., 1998; Gerber and Keller, 1999). Finally, it is worth noting that tRNA modifications are dynamic. In yeast, it was shown (by Kamenski et al., 2007; Chan et al., 2012) that heat or oxidative stress result in changes in tRNA modifications, leading to translational reprogramming.

Once fully mature, tRNAs are considered very stable with half-lives estimated to be from 9 hours to up to days (Anderson et al., 1998; Phizicky and Hopper, 2010; Gudipati et al., 2012). In yeast, two separate pathways for tRNA turnover have been discovered: the 3'-5' exonucleolytic degradation by the nuclear exosome and the 5' to 3' exonucleolyic degradation by the RTD pathway (for a review see Phizicky and Hopper, 2010). These two tRNA turnover pathways appear to function in tRNA quality control, eliminating tRNAs that are inappropriately processed, modified, or folded (Hopper, 2013).

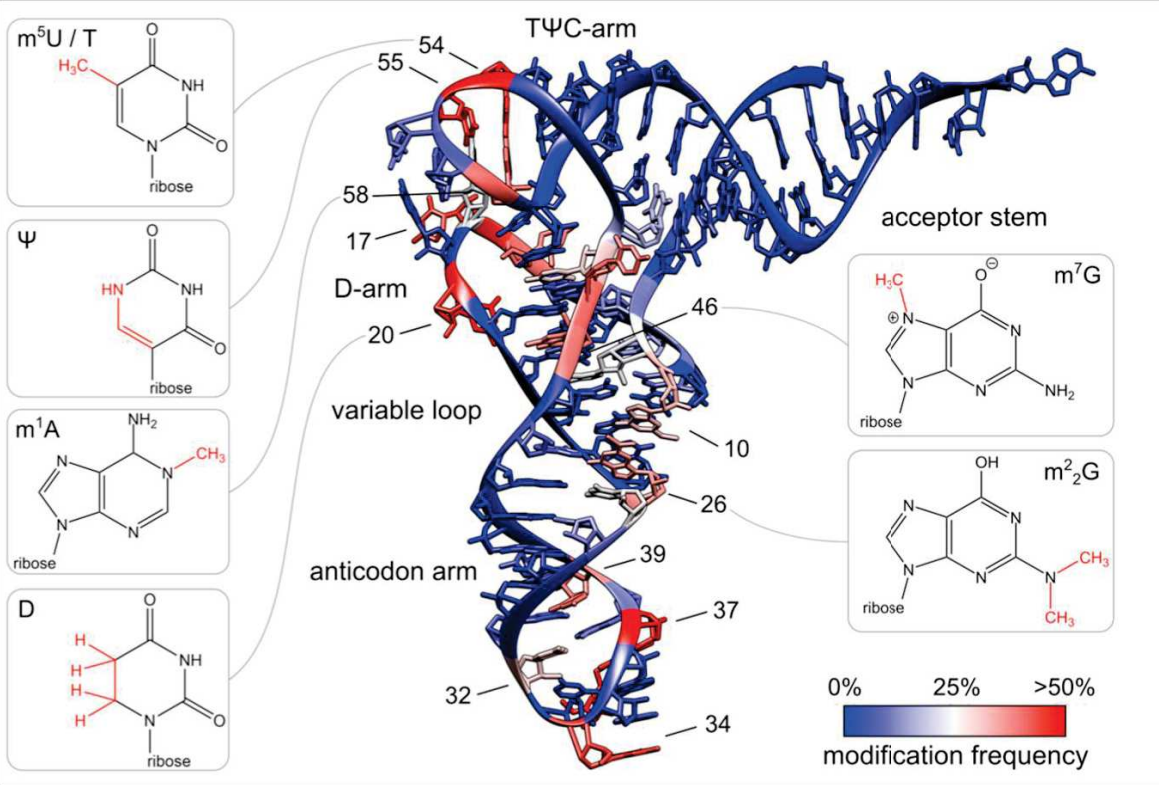


Figure 7: **Post-transcriptional modifications in a tRNA shown in 3D**. The coloured tRNA structure shows the modification frequency of each base. The modification data were taken from the tRNAmodviz database and plotted on the crystal structure of $tRNA^{Phe}$ from S. cerevisiae. Blue coloured bases are rarely modified; red coloured bases are modification hotspots. tRNAs possess two regions with high modification levels—the anticodon loop (especially positions 34 and 37) and the core or elbow region, where D- and T-loop bases interact with each other and stabilize the tertiary fold. For some important positions, the chemical structure of the most frequent modification at this position is shown. Image taken from Lorenz et al. (2017).

II. tRNA like structures: distribution and functions

Beyond the canonical tRNAs described in the previous section, a number of larger RNAs contain in their sequence or structure elements that resemble tRNAs. These so-called "tRNA-like structures" or TLS were first discovered in 1970 in plant viruses. Pinck et al., and Yot et al., (1970) discovered that the 3'end of the Turnip yellow mosaic virus (TYMV) RNA can be aminoacylted with valine. Since that date, TLS have been defined as either an RNA molecule that can react efficiently with one or more tRNA-specific enzyme or as an RNA that has a sequence similar to the one of a tRNA, meaning that its secondary and / or tertiary structure closely resembles to the canonical clover-leaf secondary structure or L-shaped tertiary structure of a tRNA (Mans et al., 1991).

As mentioned above, tRNA like structures were initially discovered 50 years ago in plant viruses. Because of their essential role for viral infection, they are still actively investigated today. More specifically, TLS are found at the 3' end of the genome of some of plant positive strand RNA viruses, such as Tymoviruses (Turnip yellow mosaic virus (TYMV) genus), Tobamoviruses (Tobacco mosaic virus (TMV) genus), Bromoviruses (Brome mosaic virus (BMV) genus) or Cucumoviruses (Cucumber mosaic virus (CMV) genus) (Dreher, 2009).

Interestingly, TLS can be aminoacylated by aminoacyl tRNA synthetases. To date, TLS from plant viruses have been shown to be specifically aminacylated by valine, histidine or tyrosine (Dreher, 2009).

It is also interesting to note that plant virus TLS are characterised by their diversity. Alongside the already noted diversity in specific aminoacylation, plant virus TLS are diverse in terms of structure and functions. Indeed, if some tRNA like structures are very similar in term of 2D and 3D structure to *bona fide* tRNAs (the viral TLS of the TYMV for example), others have strongly evolved and now have only conserved "vestigial" tRNA like structural characteristics (Dreher, 2009, Hammond et al., 2009). On a functional point of view, the roles of TLS in plant viruses range from involvement in the encapsidation of viral RNA to the recruitment of the host CCA nucleotidyltransferase (CCase) as a telomerase to maintain intact 3' CCA termini. Finally, they can also be involved in the presentation of minus strand promoter elements for viral replication (Dreher, 2009).

In the best-studied model, the TYMV TLS, it has been shown that aminoacylation of the TLS by a valine residue is necessary for infection. As it plays a major role in translational enhancement and down regulation of minus strand initiation (Dreher, 2009). However, some plant virus TLS have also been shown to be non-essential.

Not restricted to plant viruses, tRNA like structure have also been found in other viruses, for example Hepatitis C virus and related animal pestivirus or in the cricket paralysis virus (Diaz-Toledano and Gomez, 2015). Interestingly, RNase P from HeLa cells or from the cyanobacterium *Synechocystis* sp. were shown to cleave the HCV at this TLS site. However, only in *vitro*, and no evidence of this cleavage was found in *vivo* (Diaz-Toledano and Gomez, 2015).

However TLS are not found exclusively in viruses. Indeed, it was shown that in human, the interferon-alpha 5 (IFNA5) mRNA was cleaved *in vitro* by RNase P and predicted chemically and by bioinformatics to contain an TLS at the RNase P cleavage site (Diaz-Toledano and Gomez, 2015).

More interestingly, two tRNA like molecules have been shown to be generated from the mammalian long non-coding RNAs MALAT1 and MENβ *in vivo*.

Wilusz and al., (2008) showed that the nascent MALAT1 transcript contains a TLS targeted by RNase P. The cleavage products generate a lncRNA known to be misregulated in different human cancers (5' end of the pre-MALAT1 transcripts) and a 61 nucleotides tRNA like structure broadly expressed in human tissues.

The same team further showed that two non-coding RNA isoforms are produced from a single RNA polymerase II transcript at the Men ε/β locus in human (Sunwoo et al. 2009). This single pre-transcript contains a 3' poly-A rich region located upstream of a tRNA-like structure. As for MALAT1, cleavage by RNase P at the TLS site generates a TLS truncated Menβ transcript containing a poly(A) rich 3' end in addition to the Men ε that contains the TLS. The two non-coding RNA isoforms are located in nuclear paraspeckles and directly interact with the p54/nrb complex (a known protein component of the paraspeckles). Knockdown of the Men ε/β transcripts led to the disruption of paraspeckles, highlighting their importance (Sunwoo et al., 2009).

TLS are also found in plant transcripts. They can be found in both nuclear encoded transcripts and mitochondrial encoded transcripts. For nuclear encoded transcripts, Plewka et al., 2018 showed that two Arabidopsis lncRNAs, GUT15 and CR20-1 contain a TLS. However, only the GUT15 transcript was studied in details and the authors showed

that GUT15 is a precursor of two RNAs, a tRNA-like molecule and a tRNA-derived fragment, GUT15-tRF-F5. The authors further showed that this tRNA-like molecule contains a CCA motif and has an inhibiting effect on the splicing of GUT15 intron. However, they did not investigate the mechanism by which this tRNA-like molecule is generated, nor its function.

In plant mitochondria, numerous TLS have been described, i.e. in the 5' or 3' untranslated regions of mitochondrial mRNAs (Hanic-Joyce et al., 1990, Forner et al., 2007, Canino et al., 2009, Gobert et al., 2010, Gutmann et al., 2012, Fujii et al., 2016). These structures serve as maturation signals for endonucleolytic cleavages that can be performed by either RNase P or Z. For instance, my host laboratory could show that the mitochondrial RNase P enzyme AtPRORP1 cleaves both *in vitro* and *in vivo* a conserved TLS in nad6 mRNA to obtain the mature 3' end of this mRNA (Gobert et al., 2010, Gutmann et al., 2012).

Finally, tRNAs sequences have another additional function in plants. Indeed, in *Arabidospsis*, cell-to-cell mobile mRNAs are enriched in tRNA like-motifs or are transcribed from genes located in close proximity to tRNA annoted loci, which form di-cistronic tRNA-mRNA transcripts at high frequency (Zhang et al., 2016). Their result suggests that tRNA derived sequences are, in plants, sufficient to mediate mRNA transport. Moreover, these tRNA-derived sequences were experimentally added to nonmobile transcripts and showed as sufficient to mediate transport of these otherwise nonmobile transcripts (Zhang et al., 2016).

The result section of my PhD work will provide another example of how a conserved tRNA like structure is recognized and cleaved *in vivo* by an RNase P enzyme.

III. RNA polymerase III and its transcripts

Ribonucleic acid polymerases (RNAPs) are the enzymes responsible for the transcription of deoxyribonucleic acid (DNA) into RNA. During evolution, RNAPs have diversified and their number of protein subunit has greatly increased. I will first present the evolution of RNA polymerases and the specific features of RNA pol III (a). Then I will present how RNA pol III synthesizes RNA (b) and the variety of its products (c).

a) Evolution, diversity and structure of RNA polymerases

In Bacteria, a unique RNA polymerase is made up of 5 subunits whereas in Eukaryotes, RNAPs are composed of at least 12 proteins. Moreover, if in Archaea and Bacteria, a single RNAP transcribes all the RNA coding genes, in Eukaryotes three different nuclear polymerases are usually found, and up to five in plants.

In Eukaryotes, RNA pol I transcribes the 18S, the 5.8S and the 28S ribosomal RNA (rRNA) genes. RNA pol II transcribes messenger RNAs of protein-coding genes and numerous non coding (nc)RNAs. While RNA pol III transcribes small non coding RNAs such as transfer RNA (tRNAs), 5S rRNA, U6 RNA, 7SL RNA (Cramer, 2002).

In plants, two additional RNA polymerases exist (IV and V). They were shown to be necessary for the production and function of small interfering RNAs (Pikaard et al., 2008; Wierzbicki et al., 2008). It is worth noting that in plants, the existence of a sixth polymerase is presently debated (Trujillo et al., 2018).

This change from a single polymerase to multiple specialized enzymes was accompanied by an increase in the number of subunits of the RNAPs. Indeed, as noted previously, the bacterial RNAP has 5 subunits (β and β' subunits, an α subunit homodimer, and ω

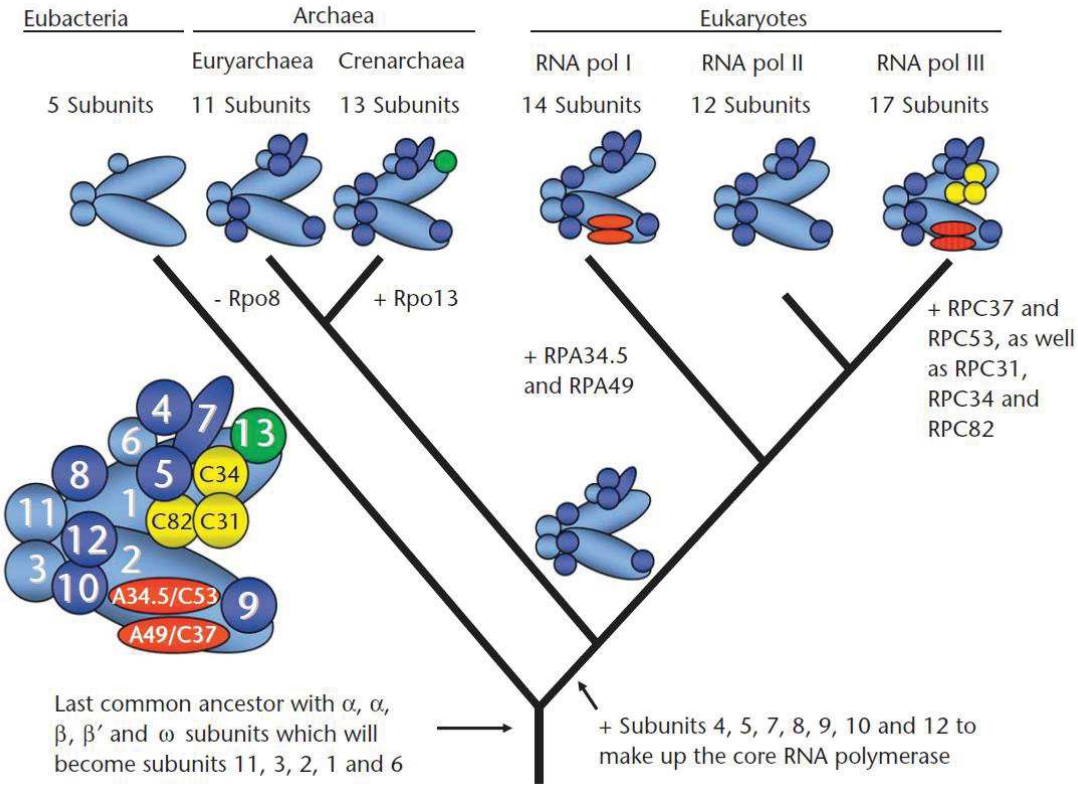


Figure 8: **Schematic representation of the evolution of the main cellular RNA polymerases**. The five subunits present in eubacteria are coloured in light blue, whereas the seven subunits that were added to form the ancestral core RNA polymerase are coloured in dark blue. The other subunits present in some enzymes are coloured in green, red and yellow as indicated. The phylogeny depicted, with eubacteria being the sister group of Archaea and Eukaryotes and with RNAPI being the sister group of RNAPII and III, is generally accepted as being most likely. The different branch lengths of RNAPI, II and III also represent their relative amino acid substitution rates. Figure taken from Drouin et al. (2010).

subunit). It is considered that during evolution, these five subunits were retained and seven others were added to make a total of 12 subunits in the RNAP of the common ancestor of Archaea and Eukaryotes (Korkhin et al., 2007). Drouin and Carter (2009 and 2010) recapitulated the evolution of RNAPs. Their results are shown in Figure 9. As it can be seen in this Figure, eukaryotic RNA pol II has 12 subunits, RNA pol I has 14 and RNA pol III up to 17. Most of these new subunits originate from the recruitment of pre-existing general transcription factors (Drouin and Carter., 2010). These proteins functions are to recruit the polymerase to the DNA promoter where they will contribute to the polymerase binding to DNA and transcription initiation. This stable integration of transcription factors in constitutive subunits of RNA pol I and III leads to a faster and more efficient transcription initiation compared to RNA pol II. However, RNA pol II system that recruits transiently these initiation factors offers a stricter control. This can be seen as a trade-off where RNA pol I and III produce more transcripts, whereas RNA Pol II, which transcripts can be potentially harmful to the cell if unregulated, has a stricter control of its activity (Khatter et al., 2017).

Furthermore, Drouin and Carter showed that RNA polymerases likely evolved towards an increase in the transcription efficiency of the specific type of RNA they transcribe. As most functional differences between RNA pol I, II and III are clustered near their catalytically active site.

From a more structural point of view, eukaryotic RNAPs have a core of 10 proteins subunits (five subunits are the homologs of the bacterial one, plus five additional one). This core encloses a central cleft where the DNA will be located during transcription as well as two channels, one for the RNA product and one for the NTPs substrates. This core is considered, by structural biologists, as having a “crab-claw” shape (Figure 9). Core elements also consist of two “pincers” called “clamp” and “jaw”. Their function is to stabilize the DNA at the downstream end and to allow closing and opening of the cleft. During transcription, the RNA polymerase will maintain a so-called “transcription bubble” with separated DNA strands. This will facilitate nucleotides addition, allow the enzyme to translocate along the template and stabilize the hybrid between the nascent RNA transcript and the DNA template (Khatter et al., 2017). In addition to this core, eukaryotic RNA polymerases share two additional subunits that will form the “stalk”. And additional peripheric subunits exist in pol I and pol III (khatter et al., 2017).

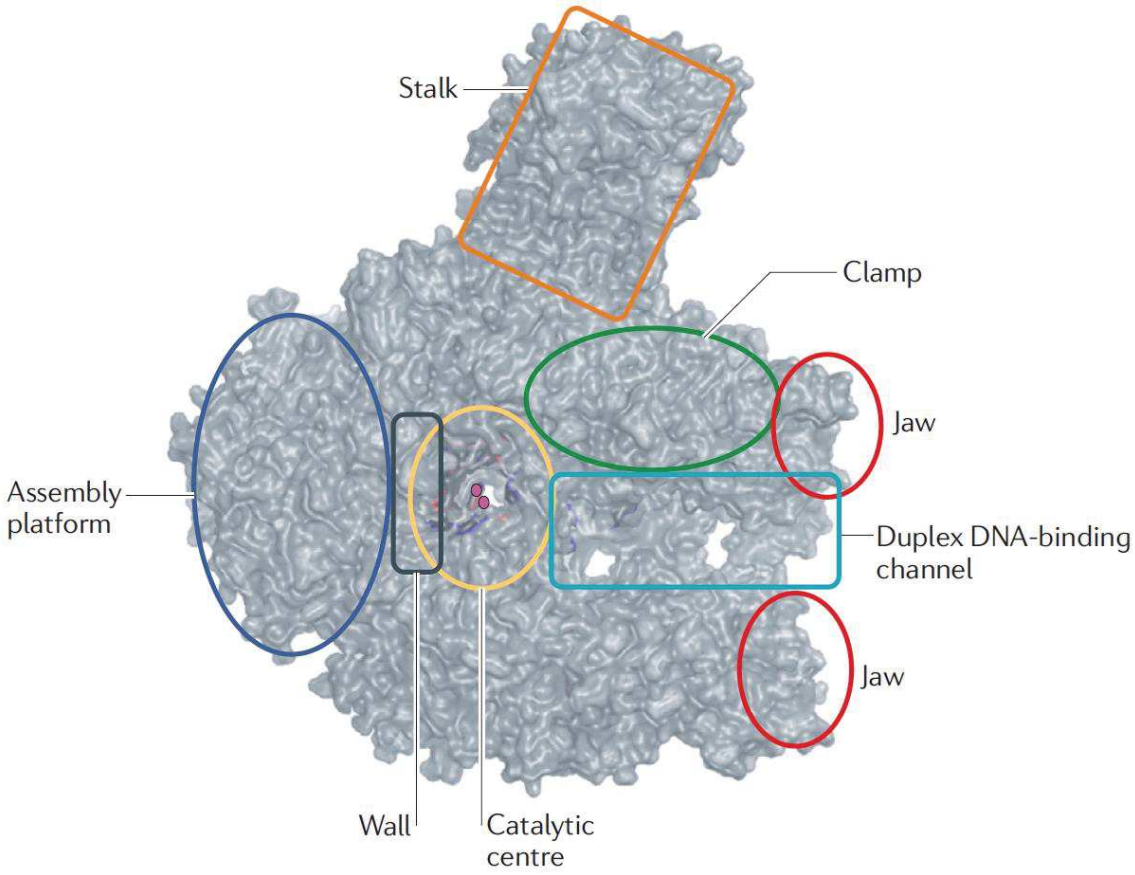


Figure 9: **Overall architecture of an RNA polymerase**. All important structural and functional features of an RNA polymerase are shown. Magenta spheres indicate the two active-site Mg^{2+} ions. The structural information was obtained from eukaryotic RNAPII Protein Data Bank entry 1Y1W. Image taken from Werner and Grohmann (2011)

In contrast to the core elements that are the essential part for transcription elongation, the main role of these peripheric subunits is initiation and termination of transcription (Khatter et al., 2017). For example, in RNA polymerase III, three subunits (C31, C34 and C82) form a complex unique to RNA polymerase III that plays an important role in transcription initiation by interacting with initiation factor TFIIIB (Thuillier et al., 1995).

b) RNA polymerase III transcription

Plant RNA polymerases have been studied in details by the group of C.S Pikaard (at Indiana University, US). Using affinity purification and mass spectrometry, they identified all the subunits of *A. thaliana* RNA polymerases I and III (Ream et al., 2015). Arabidopsis orthologs of all 17 yeast RNAP III and 12 out of 14 RNAP I subunits were identified in their study. This finding reinforces the hypothesis that RNAP I and III are similar in composition throughout Eukaryotes (Ream et al., 2015). However, if in yeast single genes encode each RNA polymerase subunit, in Arabidopsis, it is common that numerous RNA polymerase subunits have numerous alternative forms (Ream et al., 2015, see Table 1 for details).

At the structural level, the mode of action of RNA polymerase III is less well known than that of RNA pol II. However, cryo-EM structures of unbound and elongating yeast RNA pol III have been obtained (Hoffmann et al., 2015). For the transcription initiation step, RNA polymerase III requires TFIIIB and TFIIIC to transcribe tRNA genes. The TFIIIC complex will bind the intragenic promoter elements of tRNA genes and position the TFIIIB complex (composed of TBP, TFII-related factor (Brf1) and Bdp1) on the tRNA flanking promoter (Khatter et al., 2017). It seems that TFIIIB is the main actor in RNA pol III initiation as it was shown that TFIIIC is only necessary to position correctly TFIIIB (Kassavetis et al., 1990). The cryo-EM structures of Yeast RNA pol III ((Hoffmann et al., 2015) show that RNAP III is an interconnected machinery in which the movements of the core are closely concerted with the peripheral subunits, that can sense initiation and termination signals (Khatter et al., 2017).

Termination of transcription by RNA polymerase III is quite specific. It usually terminates when transcribing short sequences of T (T>4 or 5) in the absence of any specific accessory factors (Dieci et al., 2007). Recognition by RNA pol III of this termination signal can be

influenced by the region surrounding the T cluster (Braglia et al., 2005). Also, some examples of RNA pol III transcripts terminating without oligo(dT) terminator region have been reported (White, 1998). The main consequence of this unique termination signal is that the vast majority of RNA pol. III transcripts have a uniform 3'end composed of a stretch of Us (Arimbasseri, 2018). The La proteins will bind to this 3'end, protecting it from exonuclease degradation. For tRNAs this step will direct the beginning of their post-transcriptional maturation. La binding will also facilitate the assembly of snRNA into RNA-protein complexes and it will lead to nuclear retention of some snRNAs (Wolin and Cedervall, 2002).

For tRNAs, the La binding was shown to serve essentially as a chaperone (Bayfield et al., 2007) by protecting the 3'end against exonucleases and to direct it towards the 5'-first processing pathway. Indeed, in yeast, when the La protein is absent, pre-tRNAs are first matured at their 3' end by exonucleases before removal of the 5' leader (Yoo et al., 1997). Concerning La proteins, it is worth noting that they are evolutionary conserved in all Eukaryotes, but contrary to mammals where they are essential genes, their depletion is not lethal in Yeast (Arimbasseri, 2018).

La proteins have also been studied in *Arabidopsis*. The Arabidopsis genome encodes two putative La proteins. However, only one could rescue *S. cerevisiae* La nuclear functions (La protein depletion is not lethal in WT cells, but it is in specific genetic backgrounds) in complementation assays and was able to bind to the 3' poly(U) end of RNA pol III transcripts *in vivo*. This protein was named AtLa1 (At4g32720) and was shown to be an essential gene in *Arabidopsis* (Fleurdépine et al., 2007).

		Pol I coverage (%)		Pol III coverage (%)		Pol II coverage (%)		Pol IV coverage (%)		Pol V coverage (%)	
Subunit	Gene ID	Total	Unique	Total	unique	Total	Unique	Total	Unique	Total	Unique
NRPA1	At3g57660	49	49	0	0	0	0	0	0	0	0
NRPB1	At4g35800	0	0	0	0	59	59	0	0	0	0
NRPC1	At5g60040	0.6	0	22	22	0	0	3	3	4	4
NRPD1	At1g63020	0	0	0	0	0	0	58	58	0	0
NRPE1	At2g40030	0	0	0	0	0	0	0	0	63	63
NRPA2	At1g29940	51	51	0	0	0	0	0	0	0	0
NRPB2	At4g21710	2	2	0	0	63	63	0	0	3	3
NRPC2	At5g45140	0	0	39	39	0	0	0	0	0	0
NRP(D/E)2	At3g23780	0	0	0	0	0	0	18	4	37	13
NRPD2b	At3g18090	0	0	0	0	0	0	15	0	27	0
NRPA3	At1g60850	61	61	0	0	0	0	0	0	0	0
NRP(B/D/E)3a	At2g15430	4	4	0	0	72	57	32	28	58	45
NRPE3b	At2g15400	4	4	0	0	23	4	3	0	53	41
NRPC3	At1g60620	7	3	57	57	0	0	0	0	0	0
NRPA4	?	0	0	0	0	0	0	0	0	0	0
NRPB4	At5g09920	0	0	0	0	61	61	0	0	0	0
NRPC4	At5g62950	0	0	55	41	0	0	0	0	0	0
NRPC4-like	At3g28956	0	0	14	0	0	0	0	0	0	0
NRP(D/E)4	At4g15950	0	0	0	0	0	0	13	13	8	8
NRP(A/B/C/D)5	At3g22320	59	59	44	44	63	63	15	15	0	0
NRPE5	At3g57080	0	0	0	0	0	0	0	0	39	39
NRPB5-like	At5g57980	0	0	0	0	0	0	0	0	0	0
NRPE5-like	At2g41340	0	0	0	0	0	0	0	0	0	0
NRPE5-like	At3g54490	0	0	0	0	0	0	0	0	0	0
NRP(A/B/C/D/E)6a	At5g51940	0*	0*	36	16	40	15	33	15	33	15
NRP(A/B/C/D/E)6b	At2g04630	0*	0*	35	15	40	15	18	0*	18	0*
NRPA7	At1g75670	5	5	0	0	0	0	0	0	0	0
NRPB7	At5g59180	0	0	0	0	51	51	0	0	0	0
NRPB7-like	At4g14520	0	0	0	0	0	0	0	0	0	0
NRPC7	At1g06790	0	0	48	48	0	0	0	0	0	0
NRPD7	At3g22900	0	0	0	0	0	0	52	52	0	0
NRPE7	At4g14660	0	0	0	0	0	0	9	9	33	33
NRP(A/B/C/D/E)8a	At1g54250	65	56	37	28	66	30	0	0	9	0
NRP(A/B/C/D/E)8b	At3g59600	62	37	33	24	66	30	18	18	9	0
NRPA9	At3g25940	60	60	0	0	0	0	0	0	0	0
NRP(B/D/E)9a	At3g16980	0	0	0	0	30	22	0	0	22	22
NRP(B/D/E)9b	At4g16265	0	0	0	0	37	28	22	22	22	22
NRPC9a	At4g07950	0	0	11	0	0	0	0	0	0	0
NRPC9b	At1g01210	0	0	11	0	0	0	0	0	0	0
NRP(A/B/C/D/E)10	At1g11475	70	55	87	72	70	55	54	54	70	55
NRPB10-like	At1g61700	15	0	15	0	15	0	0	0	15	0
NRP(B/D/E)11	At3g52090	0	0	0	0	75	75	56	56	68	68
NRP(A/C)11	At2g29540	29	29	50	50	0	0	0	0	7	7
NRP(A/B/C/D/E)12	At5g41010	0	0	16	16	16	16	16	16	16	16
NRPB12-like	At1g53690	0	0	0	0	0	0	0	0	0	0
NRPA13	At3g13940	53	53	0	0	0	0	0	0	0	0
NRPA14	At5g64680	61	61	0	0	0	0	0	0	0	0
NRPC13	At3g49000	0	0	27	27	0	0	0	0	0	0
NRPC14a	At4g25180	0	0	57	57	0	0	0	0	0	0
NRPC14b	At5g09380	0	0	32	32	0	0	0	0	0	0
NRPC15	At5g49530	0	0	38	38	0	0	0	0	0	0
NRPC16	At5g23710	0	0	25	25	0	0	0	0	0	0
NRPC17	At4g01590	0	0	54	33	0	0	0	0	0	0
NRPC17-like	At4g35680	0	0	8	0	0	0	0	0	0	0

* confirmed by immunoblot analyses but not detected by LC-MS/MS

Table1: **Peptide coverage for Arabidopsis subunits of affinity-purified Pols I, II, III, IV or** V by the Pikaard Group in *Arabidopsis thaliana* and their proposed nomenclature for the different RNA polymerase subunits. As it can be seen, some subunits are shared between the different polymerases. Table originally published by Ream et al., (2015).

c) The variety of RNA pol III transcripts

rRNAs are mainly transcribed by RNA pol I, but the 5S rRNA, that will form with the 5.8S and the 28S rRNAs the large subunit of the cytosolic ribosome is transcribed by RNA pol III. In Eukaryotes, 5S rRNA genes are often organized in repetitive clusters. Transcription of these genes is based on the recruitment of the (gene-specific) transcription factor TFIIIA that will bind the internal box A and box C promoter elements. 5S rRNA gene expression can also be modulated by upstream promoter elements. 5S rRNA genes were studied in details in Arabidopsis. Using Next Generation Sequencing (NGS) data, more than 2000 5S rRNA genes were shown to be located in three major loci on chromosome 3, 4 and 5. By analysing different ecotypes, the authors further showed that these loci exhibited different epigenetics marks and different expression patterns (Simon et al., 2017). This result indicates that differential usage of 5S rDNA loci exist in distinct *Arabidopsis* ecotypes.

RNA polymerase III also transcribes the U6 small nuclear (sn) RNA. U6 is an essential component of the spliceosome, the ribonucleoprotein complex involved in splicing in Eukaryotes. The promoter organization of U6 genes is quite variable between species. U6 genes were also studied at the molecular level in Arabidopsis. Waibel and Filipowicz showed in 1990 that U6 RNA is transcribed by RNA pol III in Arabidopsis and that only an upstream element and a TATA box sequence are necessary for the *in vivo* transcription of the three Arabidopsis U6 genes, thus showing that the determinants of RNA polymerase III between vertebrates and plants are different.

The 7SL RNA, that forms the scaffold of the cytoplasmic signal recognition particle (SRP), a ribonucleoprotein complex found in all organisms is also an RNA pol III transcript. SRP can associate with ribosomes. Binding of the SRP to a nascent protein will target this protein to a membrane-bound protein translocation complex. Similar to U6 genes, 7SL RNA genes are evolutionary quite diverse. Concerning the 7SL RNA, their promoter sequence was characterized in *Arabidopsis thaliana* (Yukawa et al., 2005); They showed that Arabidopsis 7SL-1 transcription requires both an upstream sequence element and a canonical TATA box, but also two internal control regions, similar to the tRNA gene specific A and B boxes.

Vault RNAs, Y RNAs and virus encoded RNAs (like VA-I and VA-II RNAs of adenovirus) are also transcribed by RNA polymerase III (Dieci et al., 2007) as well as RNase P catalytic RNA in all Eukaryotes possessing the RNP form of this enzyme (Tollervey, 2005).

Finally, the main role of RNA polymerase III is to perform the transcription of nuclear tRNA genes in Eukaryotes. Large numbers of tRNAs genes are generally dispersed in the nuclear genome of Eukaryotes, although some tRNAs genes can be arranged into clusters (Tollervey, 2005). The transcription of nuclear tRNA genes is characterized by the requirement for two promoter sequences within the transcribed sequence (i.e. downstream of the transcription start), called Box A and Box B. Box A corresponds to the portion of the tRNA gene that encodes the D arm (nucleotides 8–19), while Box B generally corresponds to the portion of the tRNA gene that encodes the TΨC arm (nucleotides 52–62). Flanking sequences, primarily on the 5' side and most frequently including a TATA element, can also govern the efficiency of the transcription process (Goldman, 2010). A graphical summary of these elements is given in Figure 10. In plants, the -1 to -10 region is particularly rich in A and C residues and CAA triplets were shown to be transcription initiation sites (Michaud et al., 2011).

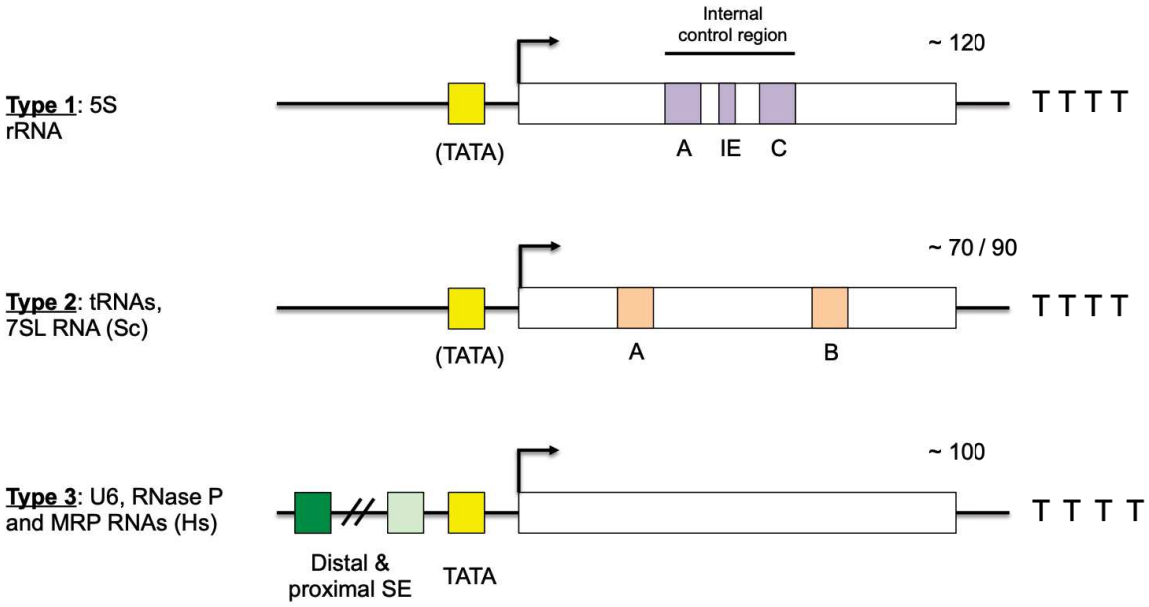


Figure 10: **Schematic representation of the three main RNA polymerase (Pol) III promoter types**. The approximate positions of upstream and internal control elements are indicated relative to the transcription start site (depicted by an arrow). A stands for box A; B for box B; C for box C and IE for intermediate element. TTTT is the transcription termination signal. SE stands for sequence element. TATA indicates TATA box or TATA-like sequence. Several mixed type promoters have also been reported that variously combine upstream and internal control elements. Parentheses indicate the fact that not all type 1 and type 2 promoters contain a TATA element. Examples of RNAs whose genes display the different types of promoters are indicated on the left (Sc = *S. cerevisiae*; Hs = *Homo sapiens*). Approximate size of the gene is indicated at the 3' end of the gene. Figure adapted from Schramm and Hernandez (2002) and Dieci et al. (2007).

IV. MAF1 is the main negative regulator of RNA polymerase III

MAF1 was originally identified in a reverse genetic screen in *Saccharomyces cerevisiae* as a repressor of RNA polymerase (Pol) III transcription (Boguta et al., 1997; Pluta et al., 2001). It is a phosphoprotein conserved in Eukaryotes. Its activity is regulated by phosphorylation and dephosphorylation.

I will first present MAF1 mode of action and its function in regulating RNA polymerase III (a), then the structural details of its mode of action as recently reported by Vörlander et al (2020) will be described (b). Finally, I will focus more specifically on what is known of MAF1 in plants (c).

a) MAF1 role and mode of action

RNA polymerase III is the eukaryotic specific polymerase that transcribes tRNAs, 5S ribosomal RNAs and other small non-coding RNAs (ncRNAs) such as the U6 spliceosomal RNA and the 7SL RNA.

As such its activity is crucial for cell growth and homeostasis. Consequently, it needs to be tightly regulated to avoid insufficient or excessive supply of these essential RNAs. For example, it has been estimated that RNA pol III activity utilizes up to 15% of the nucleotides used by dividing cells for nuclear gene transcription (Moir and Willis, 2013). For these reasons, an uncontrolled increase of RNA pol III transcription activity is currently considered, as playing an important role in cancer biology (see for example Palian, 2014 and Grewal, 2015).

To date, the main known direct regulator of RNA polymerase III activity is a phosphoprotein called MAF1. This gene was discovered by M. Boguta and coworkers in a reverse genetic screen in *Saccharomyces cerevisiae* and was shown to act as a repressor of RNA polymerase III transcription (Boguta et al., 1997; Pluta et al., 2001) in response to stress and growth signals. Since then, MAF1 was shown to be evolutionary conserved in Eukaryotes and studied in a wide variety of organisms including: worms, flies, mammalian system and plants (for a review see Leniewska and Boguta, 2017).

These researches show that MAF1 mode of action is conserved across Eukaryotes. MAF1 is inactive when phosphorylated. It is a substrate of various kinases, including the target of rapamycin (TOR) pathway or other kinases. For instance, in yeast, protein kinase A (PKA), Sch9 (S6K homolog) and casein kinase 2 (CK2) have been shown to phosphorylate MAF1 (Moir et al., 2006; Lee et al., 2009; Wei et al., 2009; Graczyk et al., 2011). This phosphorylation occurs at different sites. It is not yet clear if each of these kinases has, individually, the ability to prevent MAF1 interaction with RNA Pol III. When dephosphorylated, MAF1 becomes active and will localize to the nucleus to repress RNA polymerase III activity. In yeast, two phosphatase proteins have been shown to dephosphorylate MAF1: PP2A (Oficjalska-Pham et al., 2006) and the PP4 phosphatase complex (Oler and Cairns, 2012). It is also worth noting, that while in yeast, the dephosphorylation of MAF1 leads to a re-localization of MAF1 in the nucleus from the cytoplasm, allowing a mechanism of fine-tuning regulation of MAF1 activity, in mammalian cells, MAF1 seems to be constantly localised in the nucleus.

Differences also arise in the effects of MAF1 suppression or over-expression in the different organisms studied to date. If in unicellular organisms (yeast), MAF1 downregulation leads to a growth defective phenotype (Pluta et al., 2001), in multicellular organisms the effects seem to be contrasted. For example, in *C. elegans,* depletion of MAF1 ortholog increases lifespan (Cai et al., 2016). On the contrary, it is the overexpression of MAF1 that increases lifespan in the gut of flies (Filer et al., 2017). Another interesting example can be found in mice, where Bonhoure and coworkers (2015) showed that the loss of MAF1 confers obesity resistance. These findings are currently explained in terms of metabolism and energy expenses. It is thought that futile cycles of tRNA synthesis and degradation leads to inefficient cellular energy use and *in fine*, to central metabolic pathways reprogramming (Willis et al., 2018 and Bonhoure et al., 2019).

At the molecular level, the precise mode of action of MAF1 had remained incompletely understood until recently. A Cryo-EM study published in 2020 now provides mechanistic and molecular details on MAF1 mode of action. It shows that MAF1 binds between the clamp, wall and protrusion domains of RNA polymerase III and represses its activity through direct competition with TFIIIB (Vörlander et al., 2020).

b) MAF1 protein structure and mode of action at the molecular level

Crystal structures of MAF1 have been obtained for yeast (*Saccharomyces cerevisiae*), human (*Homo sapiens*) and Sweet orange (*Citrus sinensis*) (Figure 11 and 12).

Comparison of these three structures shows a conserved globular structure characterized by an alpha-beta sandwich. Where central antiparallel beta sheets are flanked by alpha helices on both sides. MAF1 structure is overall stabilized by a large number of hydrophobic interactions involving primarily the side chain of aromatic residues (Soprano et al., 2017). Beyond the overall conservation of the protein fold, protein sequences alignments show that two motifs, PDYDFS and WSFNFFFYN, are particularly well conserved as well as the RKLKR motif composing MAF1 nuclear localization signal (Soprano et al., 2017) (Figure 11).

Finally, it is noteworthy that these crystal structures as well as different biochemical studies have shown that MAF1 phosphorylation sites are located in a poorly conserved internal region that connects the N- and C- terminal regions that are much more conserved (Vörlander et al., 2020).

Despite its central role in RNA Pol III activity, MAF1 precise mode of action was largely unknown until recently. Based on a low resolution (~20 Å) electron-microscopy reconstruction of MAF1 bound to Pol III performed by Vannini and co-workers (2010) the proposed mechanism of MAF1 repression of RNA pol III activity was a binding of MAF1 to the clamp domain of RNA polymerase III therefore inhibiting the assembly of TFIIIB on DNA and the recruitment of RNA Pol III to the TFIIIB-DNA complex (Vannini et al., 2010). However, the cryo-EM study of Vörlander and al., who obtained a much higher resolution (3.3 Å) of a recombinant MAF1 bound to a purified *S. cerevisiae* RNA pol III presents a very different model. According to this data, the exact mechanism of the inhibition of RNA polymerase III transcription by MAF1 is not the one proposed by Vannini and coworkers. In fact, in the model proposed by Vörlander et al., the elements involved in transcription initiation of RNA polymerase III are sequestered in the MAF1-bound state, as MAF1 binding site overlaps with the binding site of the transcription factor TFIIIB and DNA in the pre-initiation complex (Figure 13).

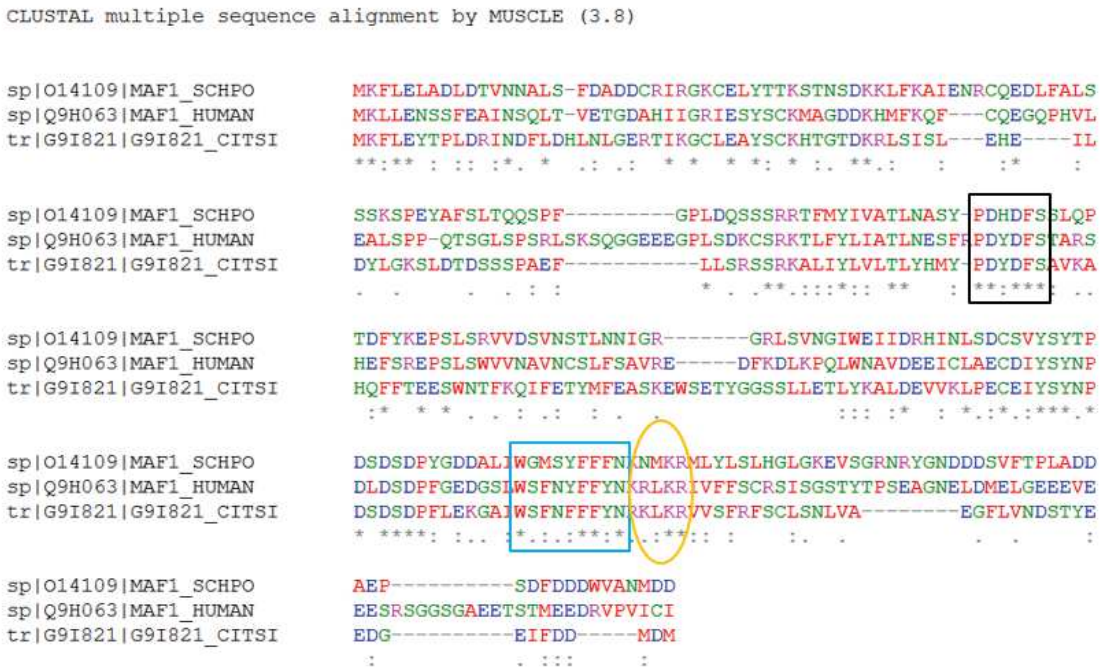


Figure 11: Clustal alignment of the protein sequence of MAF1 in Human (*H. sapiens*), Yeast (*S. pombe*) and sweet orange (*C. sinensis*) summarising the results of Soprano et al., 2013 and 2017. The two conserved peptides called "MAF1 signature" are shown in black (PDYDFS) and blue box (WSFNFFFYN). The yellow circle highlights the predicted NLS. Overall the large conservation of the protein sequence in these three evolutionary distant species can clearly be seen.

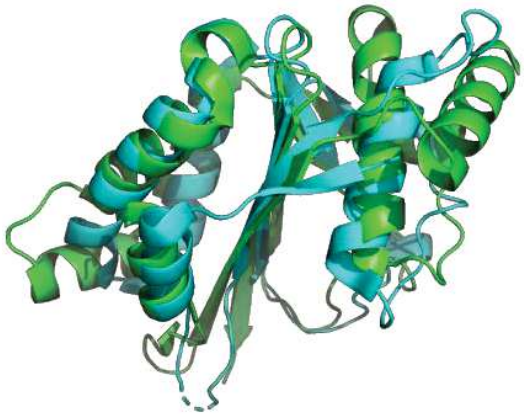


Figure 12: Aligment using Pymol of the crystal structure of MAF1 in human (*H. sapiens*) (PDB 3nr5, in blue) and Sweet orange (*C. sinensis*). (PDB 5U4Z, in green) The overall conservation of MAF1 structure can clearly be seen, highlighting that MAF1 structure is necessary for its function.

Also, these authors found by mapping the electrostatic potential onto the surface of MAF1 that a conserved part of MAF1 structure (called by the authors “interface B”) is composed of an helix and two negatively charged loops and does mimic a turn of B-DNA backbone (Figure 13). It is this interface B that will bind the C34 WH2 domain of RNA pol III.
Then the author overlaid the structure of MAF1 in complex with RNA polymerase III and the structure of RNA pol III in the pre-initiation complex state (TFIIIB-Pol III) showing that the position of TFIIIB and promoter DNA in the pre-initiation complex state overlap with MAF1 binding site (Figure 14). At last, their structure of MAF1 bound to RNA polymerase III show that the phosphosites of MAF1 can likely remain substrates of the different kinases even in complex with RNA pol III, allowing rapid de-repression of transcription.

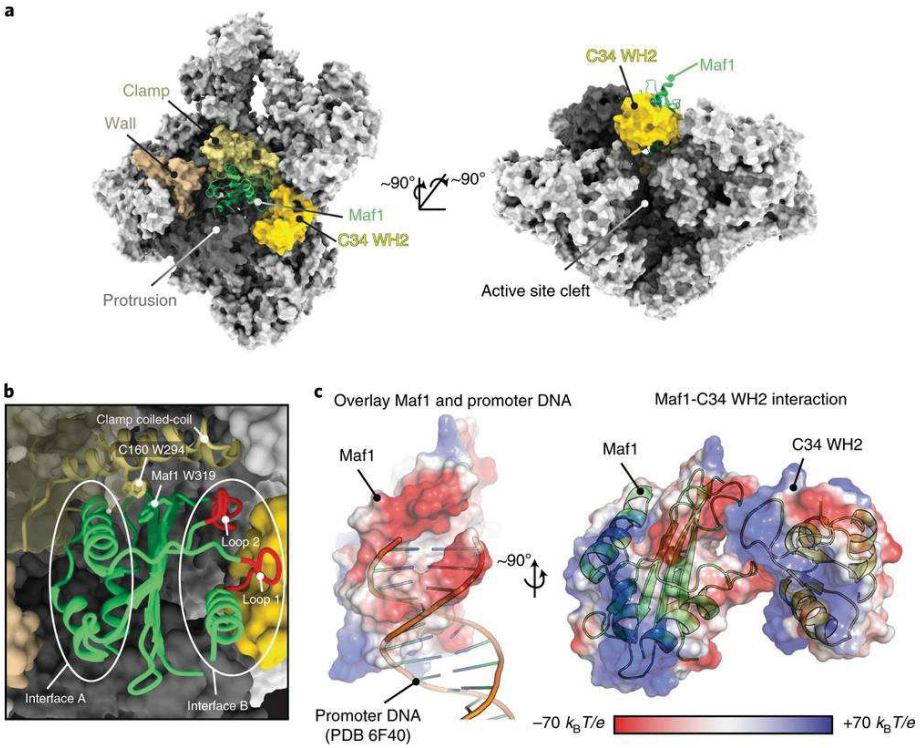


Figure 13: **Cryo-EM images of MAF1 bound to RNA polymerase III**. Pol III is shown as a surface and Maf1 as a ribbon representation. Maf1 (green) is bound between the clamp (ocher), wall (beige), protrusion (dark gray) and C34 WH2 domain (yellow). b, Close-up of the Maf1 binding site, with elements referred to in the text labeled. c, Maf1 loops 1 and 2 mimic DNA and bind the mobile C34 WH2 domain. Image taken from Vörlander et al. (2020).

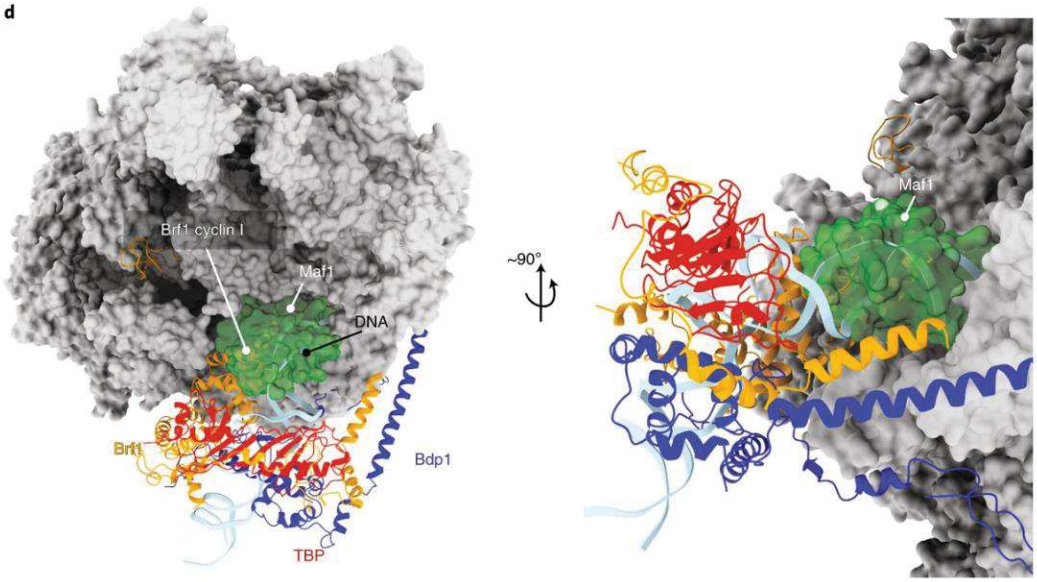


Figure 14: **Maf1 clashes with TFIIIB and DNA in the PIC**. Superposition of the Maf1–Pol III structure reported here and the Pol III preinitiation complex (PDB 6F40) in two views. Maf1 is shown as a green transparent surface and the TFIIIB and promoter DNA in ribbon representation. Image taken from Vörlander et al. (2020).

c) Research on MAF1 in plants

In plants, to date, the published research works on MAF1 are from the Brazilian research group of C.E. Benedetti (Soprano et al., 2013 and 2017) and from a Korean group (Ahn et al., 2018). Published articles show that MAF1 function (the repression of RNA pol III activity) and its mode of action (the regulation of MAF1 activity by phosphorylation / dephosphorylation of its phosphosites and the nuclear localization of MAF1 when in its active state) are all conserved in plants.

The first article was a study of MAF1 effect on canker development upon infection with *Xanthomonas citri* in *C. sinensis* (Soprano et al., 2013). The research group that performed this research was working on Transcription activator-like (TAL) effectors from *Xanthomonas* species. These pathogen's proteins act as transcription factors in plant cells. Their research focused on how these TAL effectors activate the host transcription. In a previous yeast-two hybrid screen, the authors had found that MAF1 interacts with a TAL effector protein called PthA4 (de Souza et al., 2012). Using plants overexpressing CsMAF1 (under the control of the 35S promoter) or down-regulated for CsMAF1 (using lines with RNA interference (RNAi) against CsMAF1), the authors reported that reduced levels of CsMAF1 led to a dramatic increase in the number and size of lesions caused by *X. citri* infection. On the contrary the overexpressing lines were less susceptible to *X. citri* infection. It is worth noting that according to these authors, CsMAF1 overexpression was detrimental to seedling growth. This does not seem to be the case in *Arabidopsis thaliana* (Ahn et al. 2018).

Overall, the author concluded that PthA4 counteracts CsMAF1 activity. As such, TAL effectors from *X. citri* target the regulator of RNA Pol III to coordinately increase the transcription of host genes involved in ribosome biogenesis and cell proliferation (Soprano et al., 2013).

In a second article (Soprano et al., 2017) the same research group solved the crystal structure of CsMAF1 and showed that the regulation process by protein phosphorylation is conserved. They also showed that MAF1 structure is conserved as compared with the yeast and human MAF1. Furthermore, they investigated the effect of auxin as it promotes canker development (Cernadas and Benedetti, 2009) and the effect of mTOR inhibition in WT and CsMAF1-silenced plants. They overall concluded that auxin prevents CsMAF1

from binding RNA Pol III in the nucleolus through the action of a citrus mTOR homolog (Soprano et al., 2017).

Finally, the last article on plant MAF1 investigated the function of MAF1 in *Arabidopsis thaliana*. The main conclusion is that AtMaf1 is regulated, similar to other species, by phosphorylation/dephosphorylation through the activity of TOR and PP4/PP2A phosphatases and that its repressor activity is critical for plant survival during environmental stresses (Ahn et al., 2018). However, when we used for our study the AtMAF1 mutant lines used by Ahn et al. (a SALK T-DNA insertion line), we did not identify the T-DNA insertion sites reported by the authors, but rather an insertion site inconsistent with a knock-out mutation of *maf1* gene. Therefore we decided not to use these lines.

Overall, MAF1 structure and functions seems to be conserved in plants when compared to other groups of Eukaryotes. However, the function of MAF1 in other physiological stresses beyond bacterial infection and its role in plant development are still open questions. Also, we have seen that MAF1 cellular location in Yeast is a fine-tuning regulation mechanism. Our investigation of a conserved non-coding sequence in MAF1 genes from the green lineage led us to propose another fine tuning mechanism for MAF1 expression in plants. This will be presented in details in the Results section.

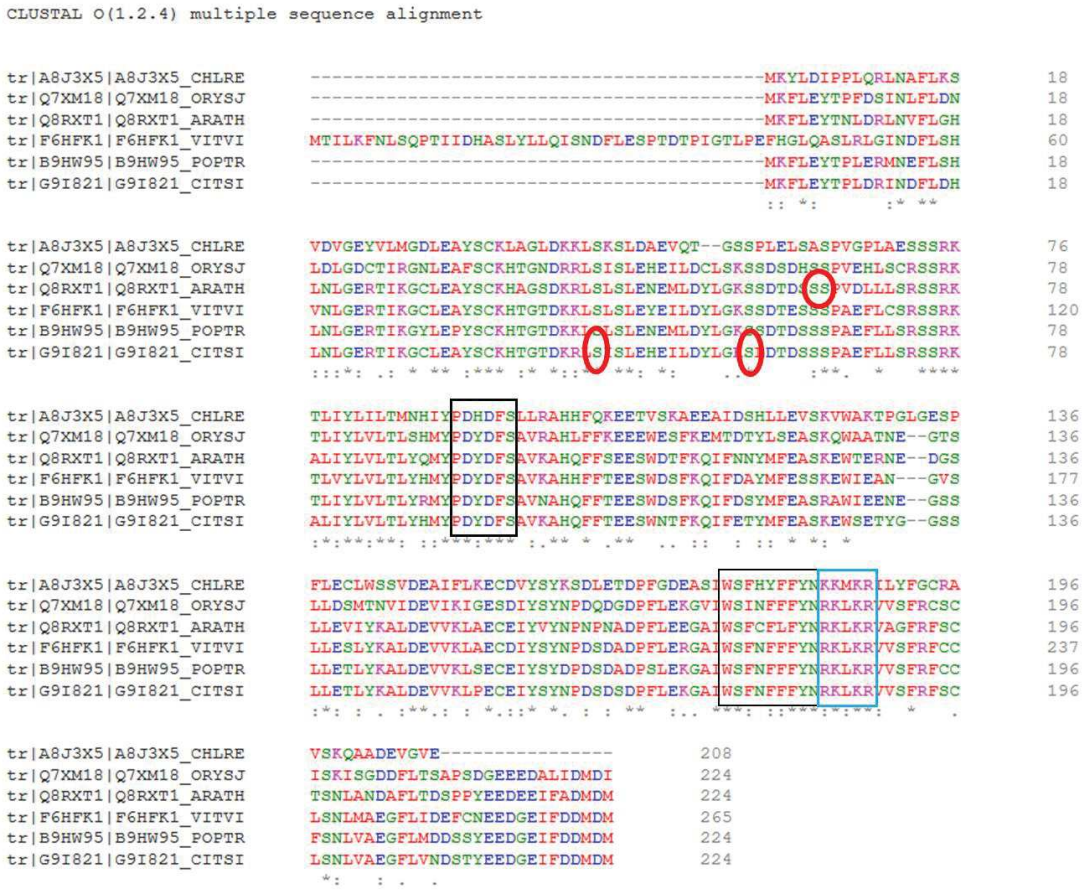


Figure 15: Clustal omega alignment of MAF1 protein sequence in 6 plant model species (Thale cress (*Arabidopsis thaliana*), the green algae *Chlamydomonas reinhardtii*, Rice (*Oryza sativa*), Grape vine (*Vitis vinifera*), the Black cotton wood (*Populus trichocarpa*) and Sweet orange (*Citrus sinensis*). Red circles indicate sites experimentally determined as phosphorylated. Black boxes highlight the two sequences considered as "MAF1 signature" in Soprano et al. Blue box show the putative NLS determined in Soprano et al. 2017.

V. RNase P diversity, function and mode of action

RNase P, the enzyme that performs the 5' maturation of tRNA precursors and of RNAs containing tRNA-like structures has been at the center of my PhD research projects.
Among the many maturation steps that pre-tRNAs undergo in order to be fully functional, the first step consists in the removal of the 5' leader sequence of precursors tRNAs by RNase P. This essential activity is conserved in all domains of life (Hopper et al., 2010). RNase P is, to date, the only known enzyme that exists as both a RiboNucleoProtein enzyme (RNP) and a protein-only enzyme (a). The structure and mode of action of both types of enzymes have been studied in details (b) showing an interesting case of convergent evolution between RNP and protein-only RNase P. Finally, the diversity of substrates and of functions held by the different kinds of RNase P enzymes in both Eukaryotes and prokaryotes will be discussed (c).

a) RNase P is the universal enzyme that catalyses removal of the 5' leader sequence of precursor-tRNAs. This enzyme exists as both a Ribonucleoprotein and a protein-only form

RNase P was first isolated in *E. coli* and defined in 1972 as the endonucleolytic RNase activity that by its cleavage "removes all extra nucleotides present at the 5' terminus of the precursor (..) exposing the 5' end of the mature tRNA" (Robertson et al., 1972). In 1983, RNase P was identified as a ribozyme (which stand for **ribo**nucleic acid en**zyme**). Indeed, Sidney Altman and colleagues showed that in *E. coli* and *B. subtilis*, "the RNA moieties (..) can cleave tRNA precursor molecules (..). The RNA acts as a true catalyst under these conditions whereas the protein moieties of the enzymes alone show no catalytic activities" (Guerrier-Takada et al., 1983). This form of RNase P was widely considered as the sole form of RNase P until a decade ago (Altmann 2007). Indeed, in Archaea, Bacteria and Eukaryotes, all RNase P characterized were, until 2008, ribonucleoprotein (RNP) particles containing a ribozyme. This finding of a conserved RNA holding the catalytic activity of the RNP particles in the three domains of life led to the

hypothesis that RNase P would be one of the sole vestige of the proposed pre-biotic "RNA world" (a term coined by Gilbert in 1986) and more controversially, to the assumption that RNase P would solely and universally occur as an RNP form.

The occurrence of a novel form of protein-only RNase P was first formally demonstrated in 2008 by Holzmann et al. (2008) that characterized human mitochondrial RNase P. This enzyme is composed of 3 proteins (MRPP1, 2 and 3), with MRPP3 (later called PRORP for "PRotein-Only RNase P") holding the catalytic activity. In fact, it was already suspected since a much longer time that RNase P activity would not rely in all organisms or compartments on a conserved catalytic RNase P RNA (see Hartmann et al., 2009). In particular, two early articles dating from 1995 and 2000 by Thomas et al, had already suggested that in Spinach (*Spinacia oleracea*) chloroplast, RNase P activity was putatively held by a single protein (Thomas et al., 1995) and that furthermore, this enzyme did "not utilize the ribozyme-type pre-tRNA cleavage mechanism" (Thomas et al., 2000). However, these results were received with scepticism as the nature of the protein and its coding gene could not be identified. Nonetheless, with the advent of the capacity to sequence massively DNA and the important efforts to sequence complete genomes from a vast variety of species, it appeared clearly that the catalytic RNA from the RNP form of RNase P was not conserved in a wide variety of Eukaryotes. As such, the occurrence of another form of this enzyme became highly plausible (Pinker et al., 2013).

The discovery that RNase P exists in a protein-only form in human mitochondria was followed by the characterisation in 2010 and 2012 of *Arabidopsis* RNase P enzymes by my host laboratory (Gobert et al. 2010; Gutmann et al. 2012). In brief, Arabidopsis encodes three PRORP proteins. PRORP1 is responsible for RNase P activity in mitochondria and chloroplasts while PRORP2 and 3 perform nuclear RNase P activity. Accordingly, RNP RNase P is completely absent from Arabidopsis genome. Contrary to human mitochondria, plant PRORP enzymes are active alone and do not require accessory proteins for their activity (for a review see Schelcher et al., 2016). Protein-only forms of RNase P were then further characterised in a larger variety of organisms, in *Trypanosoma brucei* (Täschner and al., 2012) and in other species of the green lineage such as *Ostreococcus tauri* (Laiet al., 2011), one of the plant models for cell biology, *Physcomitrella patens* (Sugita and al., 2014) and the model green algae *Chlamydomonas reinhardtii* (Bonnard et al., 2016). In 2015, a phylogenetic analysis (Lechner et al., 2015) showed that PRORP occurs in 4 out of the 5 main eukaryotic phyla and that PRORP and RNP RNase P

are mutually exclusive, in entire organisms or in compartments / organelles (Figure 16). In land plants that constitute our main models, only the protein-only form of RNase P occurs and RNP RNase P was entirely lost.

Very recently, beyond the occurrence of PRORP in Eukaryotes, the diversity of RNase P in the different domains of life was challenged again. The group of Roland Hartmann characterized another form of protein only RNase P in prokaryotes. They named these proteins "HARP" (Homologs of Aquifex RNase P). HARP was characterised in a hyperthermophilic bacterium *Aquifex aeolicus*. It is the smallest protein-only form of RNase P (Nickel et al., 2017) as it is a 23-kDa polypeptide comprising a metallonuclease domain only (while PRORP are usually around 60kDa). It has then been shown that this family of proteins is distributed in both Bacteria and Archaea. However, this form of protein only RNase P often co-exists with the RNP form of RNase P. As such, it is yet unclear what is the exact function of HARP and whether RNP RNase P and HARP have redundant functions or different substrate spectra (Daniels et al., 2019; Schwarz et al., 2019)(Figure 17).

To end this overview of RNase P distribution in the three domains of life, it is important to stress the prominence of this enzyme. Indeed, depletion of RNase P is lethal, as unprocessed tRNAs conserving their 5' leader sequence are not functional and protein synthesis cannot take place. The only known exception to this rule occurs in an archea (*Nanoarchaeum equitans*) where tRNA transcription starts at position +1. Consequently, RNase P activity is not required and was not evolutionary retained in this organism.

In plant genomes, the number of PRORP enzyme varies. For example *P. patens* has 3 PRORPs (Sugita et al. 2014) similar to *Solanum lycopersicum L*. (tomato), while *C. reinhardtii* expresses different isoforms of PRORP from a single gene (Bonnard et al. 2016). In *Arabidopsis* nuclear genome 3 PRORP proteins are encoded. PRORP1 is targeted to the organelles (mitochondria and chloroplasts) and PRORP2 and 3 (which are a recent duplication) are nuclear (Figure 18 and 19).

More generally, as shown by sequence alignments of PRORP proteins (Lechner et al, 2015), these proteins can be divided into three groups. Groups I and II includes PRORP addressed to organelles while PRORP from group III are nuclear. In *Brassicaceae,* no PRORP from group II can be found.

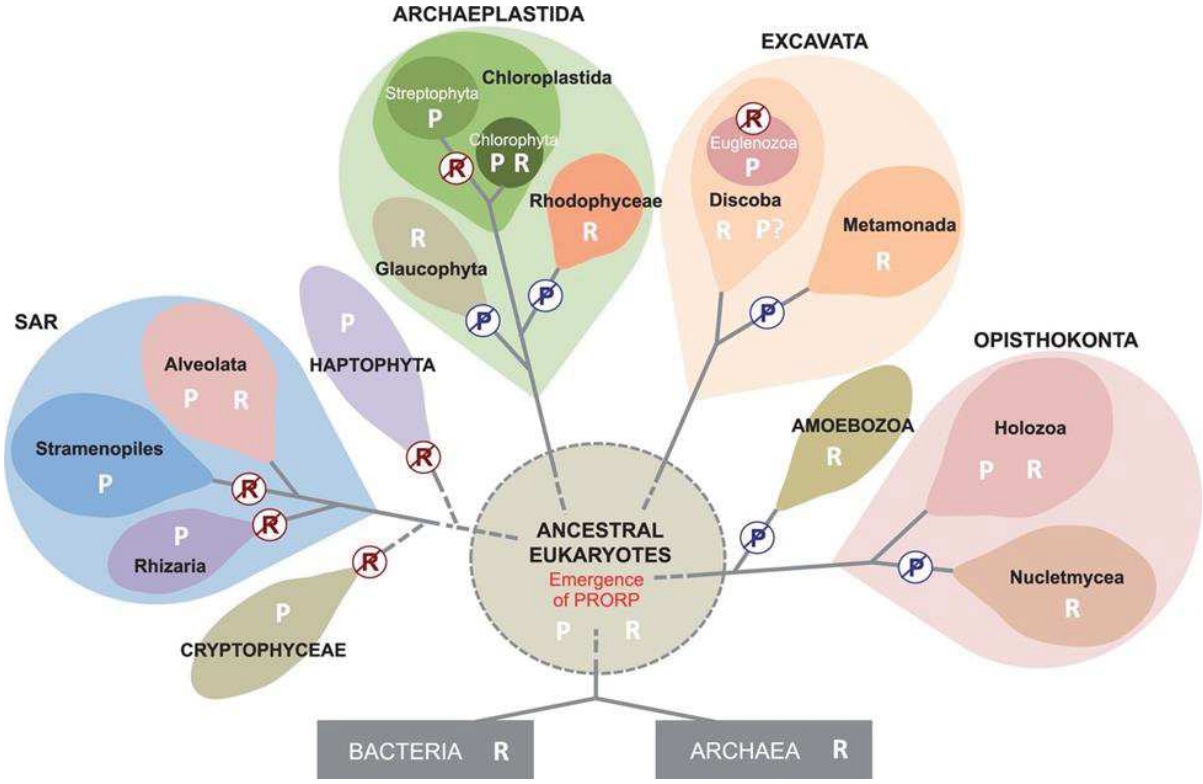


Figure 16: **Distribution of RNP and PRORP RNase P enzymes in the eukaryal domain of life**. R and P indicate the occurrence of RNP and PRORP RNase P enzymes in the respective groups, based on the study from Lechner et al., 2015. Crossing out P or R indicates putative evolutionary events associated with the loss of PRORP or RNP RNase P. The question mark indicates an example where limited genomic data prevented conclusions as to the occurrence of the given enzyme type in the respective group. Image taken from Lechner et al. (2015).

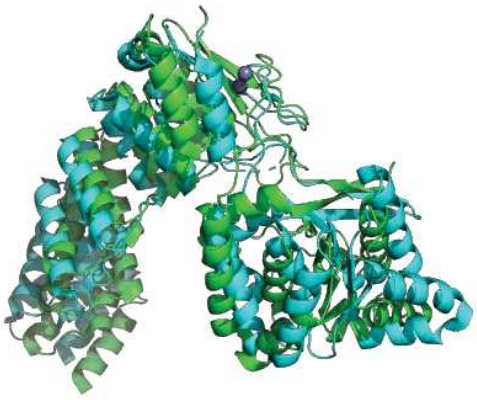


Figure 17: **Alignment of AtPRORP1** (localized in mitochondria and chloroplasts in green) **and the nuclear AtPRORP2** (in blue) using Pymol. Showing that the protein 3D structure of the nuclear and organellear Arabidopsis PRORP enzymes is highly conserved. The Zn^{2+} bound ligand is represented by a grey sphere. PDB structure for AtPRORP1 is 4g23, for AtPRORP2 5ft9.

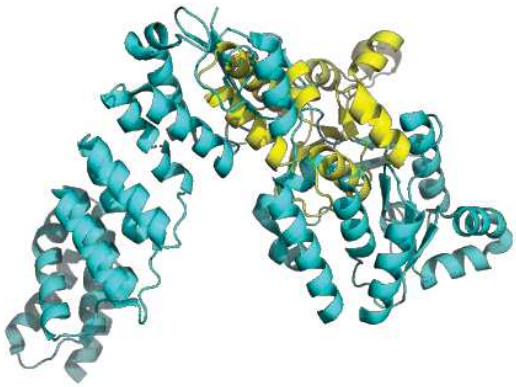


Figure 18: **Alignment, using Pymol, of the predicted structure of HARP and the crystal structure of PRORP2.** Predicted structure of HARP (Aq_880) was generated using the PHYRE 2 webserver and is shown in yellow. PDB for AtPRORP2 crystal structure is 5ft9 (in blue). HARP does not show any obvious RNA binding site. And seems to be formed of a unique catalytic domain. Aspartate residues, experimentally shown as essential for catalysis in AtPRORP are also found in this protein (see Nickel et al., 2017 for details).

b) PRORP structure and mode of action

Since in *Arabidopsis,* and more largely in land plants, RNase P solely exist as protein-only enzymes, I will focus this section on PRORP structure and mode of action at the molecular level and only refer to the RNP form, when necessary, to highlights the similitude or the differences between these two types of enzyme.

Two crystal structures have been obtained for Arabidopsis PRORP2 (Karasik et al., 2016 and Pinker et al., 2015 and 2017). They show that the enzyme is composed of three main parts. The catalytic domain is a metallonuclease domain that holds the endonuclease activity (Gobert et al., 2010). It is located at the C-terminal part of the enzyme. This catalytic domain belongs to the NYN (N4BP1 YacP-like Nuclease) family defined by Anantharaman and Aravind, 2006, (for a review see Gobert et al., 2019) of the PIN-like super-family of nucleases. NYN domains are characterized by a Rossmanoid fold that consists of a central β-sheet, sandwiched with α-helices at both sides (α/β/α sandwich fold). The catalytic domain of PRORP proteins contains four conserved aspartate residues. They are involved in the binding of metal ions during catalysis and were shown to be essential for tRNA maturation (Gobert et al., 2013). The N-terminal part of the enzyme belongs to the RNA binding, eukaryotic specific family of pentatricopeptide repeat (PPR) proteins (Small and Peeters, 2000; Small and Barkan, 2014). PPR proteins are known to be involved in large variety of mostly organellar post-transcriptional processes, such as RNA editing, splicing and the maturation of transcripts termini (Giegé, 2013; Cheng et al., 2016). They are composed of ~35-amino acid tandem repeats, with degenerate primary sequences but a conserved helix-turn-helix structure. The characterization of PPR proteins mode of action showed that two residues (at position 5 and 35) located respectively at the start and the end of PPR motifs are particularly important to achieve RNA binding specificity (Cheng et al., 2016). The PPR domain of PRORP is responsible for the specific binding of PRORP to precursor tRNAs. Structural (X-ray crystallography and SAXS data) and biochemical studies (especially footprinting assays) showed that AtPRORP PPR domain is made of 5 PPR motifs (Karasik et al., 2016; Pinker et al., 2015) and that it is essentially PPR domains 2 and 3 of Arabidopsis PRORP that are necessary to bind pre-tRNAs (Pinker et al., 2017). Indeed, positions 5 and 35 of motifs PPR 2 and 3 are extremely conserved in plants (Lechner et al., 2015) and their mutation lead to an important decrease of PRORP2 *in vitro* activity (Pinker et al., 2017). A recent publication

by Teramoto et al. (2020) described the crystal structure of PRORP1 PPR domain in complex with a tRNA and confirmed the primary role of motifs PPR2 et 3 to bind the tRNA elbow. However, this article also revealed that PRORP PPR motifs bind RNA with a distinct way as compared to canonical PPR proteins.

Finally, the NYN and PPR domains are connected by a structural zinc-binding domain that allows flexibility to the enzyme overall Λ shape. The zinc-binding domain is at the summit of the Λ and the PPR and NYN domains make its two arms. The flexibility between the PPR and the NYN domain is considered to be crucial for the catalytic site to accommodate its substrate and to release its product (Pinker et al., 2017).

These specificities of PRORP structure define PRORP mode of action, which was found to be similar to the one of the RNP RNase P. The PPR domains of PRORP binds the pre-tRNA elbow, at the level of conserved residues in the T and D loops, similar to the specificity domain of the P-RNA and the NYN domain of PRORP contacts the tRNA acceptor arm similar to the catalytic domain of the P-RNA (Pinker et al., 2017).

Overall, the different biochemical and structural data available for AtPRORP2 structure and mode of action led colleagues from the lab to build a model (Figure 20) of AtPRORP2 in complex with a pre-tRNA. This model is in accordance with available SAXS data and shows that PPR motif 3 interacts with base G18 of the pre-tRNA, PPR motif 2 with base C56. This model also shows that major conformational changes take place when AtPRORP2 binds the pre-tRNA. Indeed, the protein has to achieve a more open conformation, in order for the catalytic domain to accommodate its substrate. This proposed mode of action is similar to the one of RNP RNase P as characterised by Reiter et al., 1995 and Pan et al., 2010. Similarly, the catalytic mechanism seems to be shared between PRORP and RNP RNase P. They both use a two-metal-ion mechanism in order to catalyse the phosphodiester bond cleavage of the pre-tRNA.

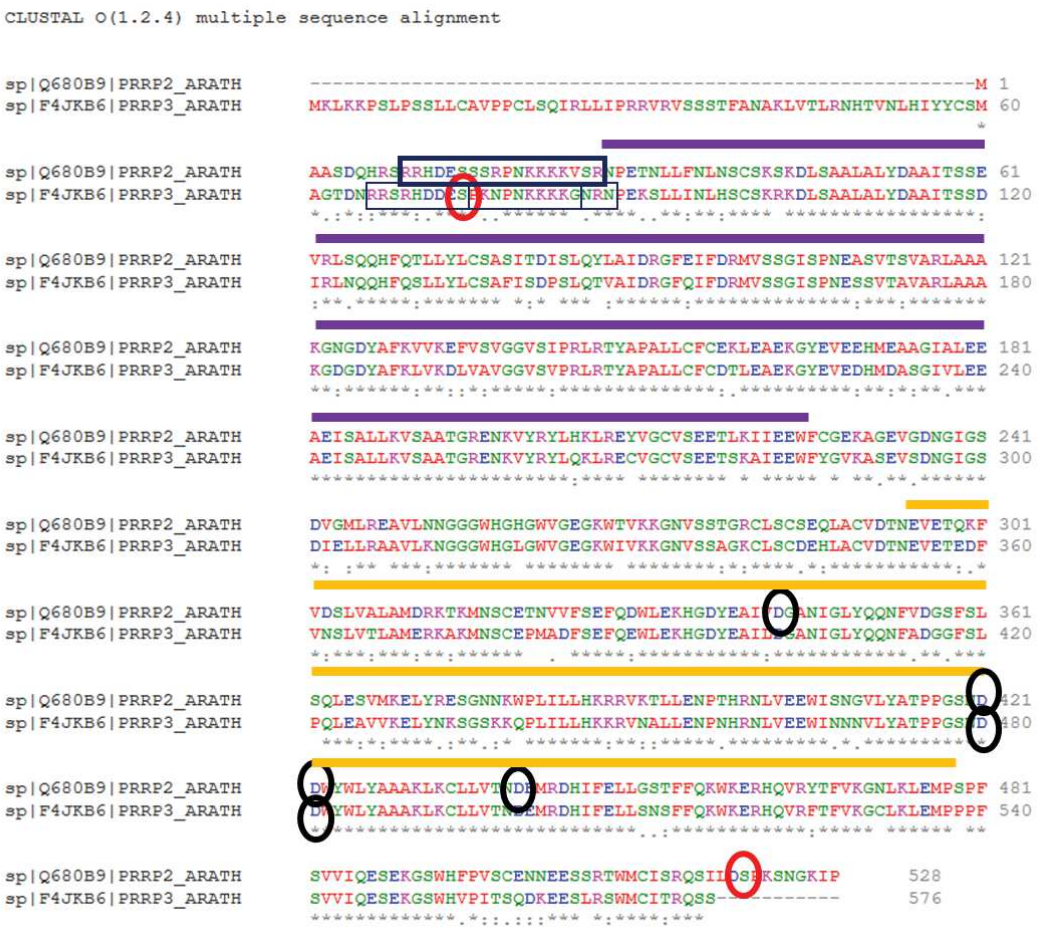


Figure 19: **Clustal omega alignment of AtPRORP2 and 3 protein sequences**. Red circles indicate phosphorylation sites detected on the Athena DB (and for PRORP2 also in our experiments). Black circle indicate Aspartate residues experimentally shown as necessary for PRORP activity (Gobert et al.,2010). Black boxes indicate the Nuclear Localization Signals (NLS) of PRORP2 and 3 predicted using Localizer (http://localizer.csiro.au/). For PRORP3, two overlapping NLS were predicted. Purple bar indicates the RNA binding domain, the orange one the catalytic domain of the protein (Gobert et al., 2013).

Figure 20: **SAXS-based model of PRORP2 in complex with a pre-tRNA substrate**. Best solution model of the complex obtained under distance (PPR2 and -3 interacting with C56 and G18, respectively) and SAXS constraints. This model is confirmed by a recent report from Teramoto et al., 2020. Image taken from Pinker et al., 2017.

However, PRORP and RNP RNase P enzymes do differ on other aspects. For example, kinetic analyses of their enzymatic activities revealed that, surprisingly, the RNP enzymes have superior catalytic efficiencies (Pavlova et al., 2012). This contradicts the hypothesis that evolution could replace RNPs by PRORP enzymes because they are more efficient. However, all these measures were of course performed *in vitro*. Consequently, results do not necessarily reflect what is really happening *in vivo*. Likewise, other evolutionary cues might be in favour of PRORP, for instance, the fact that it is easier to express and regulate a single protein enzyme as compared to RNP enzymes composed of a ribozyme and several protein subunits.

Also, PRORP proteins and RNPs have different pre-tRNA substrate requirements. If 5' leader and 3' trailer sequence were found to be involved in substrate binding in the RNP RNase P, this does not seem to be the case for PRORP (Schelcher et al., 2016). *In vitro* cleavage assays showed that PRORP optimal cleavage efficiencies were obtained with leader sequence of size comprised between 2 and no longer than 10 nucleotides (Howard et al., 2016, Karasik et al., 2016). Finally, PRORP, contrary to bacterial RNP can catalyse pre-tRNA cleavages independently of the residue immediately upstream of the cleavage site. Even if *in vitro*, cleavage efficiency is increased when an adenosine is at this site (Brillante et al., 2016).

c) Diversity of RNase P substrate

Beyond the structural similarities between PRORP and RNP RNase P, the fact that they represent a case of convergent evolution is also supported by the findings of Weber et al., (2013) that showed that yeast nuclear RNP can be functionally replaced by *Arabidopsis* PRORP (the nuclear or the organellar form) or by *T. brucei* PRORP *in vivo*. In all cases, tRNAs were correctly processed and accumulated normally. Furthermore, no obvious phenotypic differences could be observed in the yeast lines expressing PRORP.

In addition, this elegant series of experiments addressed the debated question of the diversity of RNase P substrates in the two types of enzymes. Indeed, based on the analysis of temperature sensitive mutants of RPR1, the gene encoding nuclear RNase P RNA (called *rpr1-ts* strains (Coughlin et al., 2008; Marvin, 2011, Samanta 2006)), it was considered that a vast variety of RNAs could be RNase P substrates (Marvin et Engelke, 2009 (a and b), Jarrous et Gopalan, 2010). However, Weber et al., (2013) showed that these RNAs did

not accumulate in the strains where the catalytic RNA was removed and replaced by PRORP. Moreover, another *rpr1-ts* strain was created, and the previous results could not be repeated.

Consequently, it is wise to consider that PRORP RNA substrates are restricted to pre-tRNAs and tRNA-like structures (TLS). TLS are RNA sequences which fold in a tertiary structure very similar to the one of a tRNA. Many TLS have been discovered in the genomes of plant virus and are important for virus replication (reviewed by Dreher, 2008).

To date, the list of known "non tRNA" substrates of RNase P is limited. In *E. coli*, RNase P processes the 4.5S rRNA and tmRNAs (Hartmann et al., 2009). In human (*H. sapiens*), two long-coding RNAs where shown to be substrates of the nuclear RNase P. "MEN beta" is an approximately 20-kb transcript. It contains a genomically encoded poly(A)-rich tract at its 3'-end. The 3'-end of MEN beta is generated by RNase P cleavage (Sunwoo et al., 2009). MALAT1 is another human long noncoding RNA. A highly conserved small RNA of 61 nucleotides originating from the MALAT1 locus was identified by Wilusz et al., (2008). They showed that RNase P cleaves the nascent MALAT1 transcript downstream of a genomically encoded poly(A)-rich tract to simultaneously generate the 3' end of the mature MALAT1 transcript and the 5' end of the small RNA. In both cases, RNase P cleavage occurs at the start site from the TLS contained in these ncRNA.

In *Arabidopsis*, to date, AtPRORP1 has been shown to be involved in the cleavage of the mitochondrial *nad6* and *orf291* RNAs at the level of predicted TLS (Gobert et al., 2010, Fujii et al., 2016). In the nucleus, AtPRORP2/3 have been shown to be involved in the cleavage of a small nucleolar RNA (snoRNA) precursor (Gutmann et al., 2012) located upstream of a tRNA. Nonetheless, the full extent of PRORP RNA substrates *in vivo* remains to be identified.

VI. tRNA (N^2, N^2-guanine)-dimethyl transferase (Trm1)

Although PRORP enzymes are functional alone, results obtained in my host laboratory showed that they can also functionally interact with other gene expression factors, suggesting a channelling of processes for the maturation of RNAs. For instance Arabidopsis PRORP1 interacts with another mitochondrial nuclease (Bouchoucha et al., 2019). Similarly, the results described hereafter will describe an interaction between the nuclear PRORP2 and the tRNA methyltransferases TRM1A and B *in vivo*. Hence, this section will be concluded by an introduction on tRNA methyltransferases in general (a) and more specifically on tRNA (N^2, N^2-guanine)-dimethyltransferase (Trm1) enzymes (b).

a) tRNA methyltransferases functions and mode of action

Among tRNA modifications, methylation is the posttranscriptional addition of methyl groups to specific residues in a tRNA molecule. It is one of the most common chemical modifications occurring in living organisms. Methylation occurs in a wide range of biomolecules, including lipids, proteins (including histones), metabolites and nucleic acids (Chen et al., 2020). Methylations have implications for a variety of cellular processes; including, but not restricted to: transcription, translation, epigenetics, development, or bacterial host defence. Concerning RNA, the methylation of nucleotides appears in most RNAs, constituting around 40% of the post-transcriptional modifications identified so far (Ovcharenko and Rentmeister, 2018).

One important methylation of RNA is the N^6-methyladenosine (m^6A) modification in mRNA. Indeed, 20 to 40% of mammalian transcripts are m^6A methylated and methylated mRNAs tend to contain multiple m^6A per transcript (Frye et al., 2018). N^6-methyladenosine (m^6A) has been shown to affect the translation and stability of the modified transcripts. It is proposed that this methylation provides a mechanism to facilitate transcriptome switching through the binding of specific protein complexes. In mammalian cells, this mechanism has been shown to be most important in the regulation

of groups of transcripts during cell state maintenance and transition (For a review, see Frye et al., 2018).

As introduced above, nucleotide bases modifications are part of the extensive maturation steps that newly synthetized tRNAs need to undergo in order to be fully functional. Methylation is the most abundant of these chemical modifications of tRNAs bases (Juhling et al., 2009; Machnicka, 2013). These chemical modifications are especially important for tRNA structure. As methylated nucleotides, will stabilize the L-shaped tRNA structure and improve their molecular recognition.

tRNA methylation is an ancient process (Jackman et al., 2013). Indeed, it has been shown that out of the 18 tRNA modifications that occur in tRNA in the 3 domains of life, 13 are methylated nucleotides (Cm, Gm, Um, m^5C, m^5U, m^3U, m^1G, m^7G, m^2G, $m^{2,2}G$, m^1A, m^6A and $m^{6,6}A$) (Jackman et al., 2013).

tRNA methylation is performed by a class of specialized enzymes called tRNA methyltransferases. These enzymes are responsible for the transfer of a methyl group from a methyl donor, most commonly S-adenosylmethionine (SAM or AdoMet), to any of several different locations on a target RNA nucleotide. The most common substrates of tRNA methyltransferases are: adenosine (A), guanosine (G), cytosine (C) and uridine (U); the 4 canonical RNA bases. But methyl groups are also added to modified nucleotides, such as pseudouridine (ψ), inosine (I) and many more complex species, as well as to the ribose 2'-hydroxyl (Swinehart et al., 2014). Interestingly, human PRORP interacts with the tRNA methyltransferase TRMT10C. Contrary to PRORP proteins from other Eukaryote groups, this interaction is required for human PRORP RNase P activity (Holzmann et al., 2008).

tRNA methyltransferases form a large enzyme superfamily. This enzymatic family seems to contain members of various structural classes, which have likely arisen independently during evolution (for a review, see Swinehart et al., 2014). Indeed, as noted by these authors, even between closely evolutionary related enzymes there are examples of unusual substrate specificity and chemistry.

Finally, it can be noted that, as tRNA methyltransferases (and more largely tRNA modification enzyme) are abundant, most of individual deletion of these genes have often no obvious effect on the organism fitness. For example, in yeast (*S. cerevisiae*) only one tRNA methyltransferase (m^1A58 methyltransferase TRM6/TRM61) has been shown

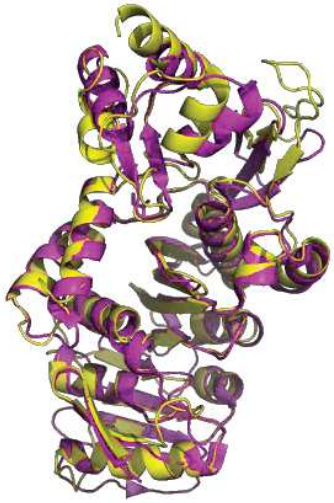


Figure 21: structure alignment of TRM1 from P. horikoshii and A. thaliana. The crystal structure of *P. horikoshii* (magenta) and of the predicted structure (created using PHYRE2) of AtTRM1A (yellow) were aligned using Pymol. The high degree of structural conservation can be seen.

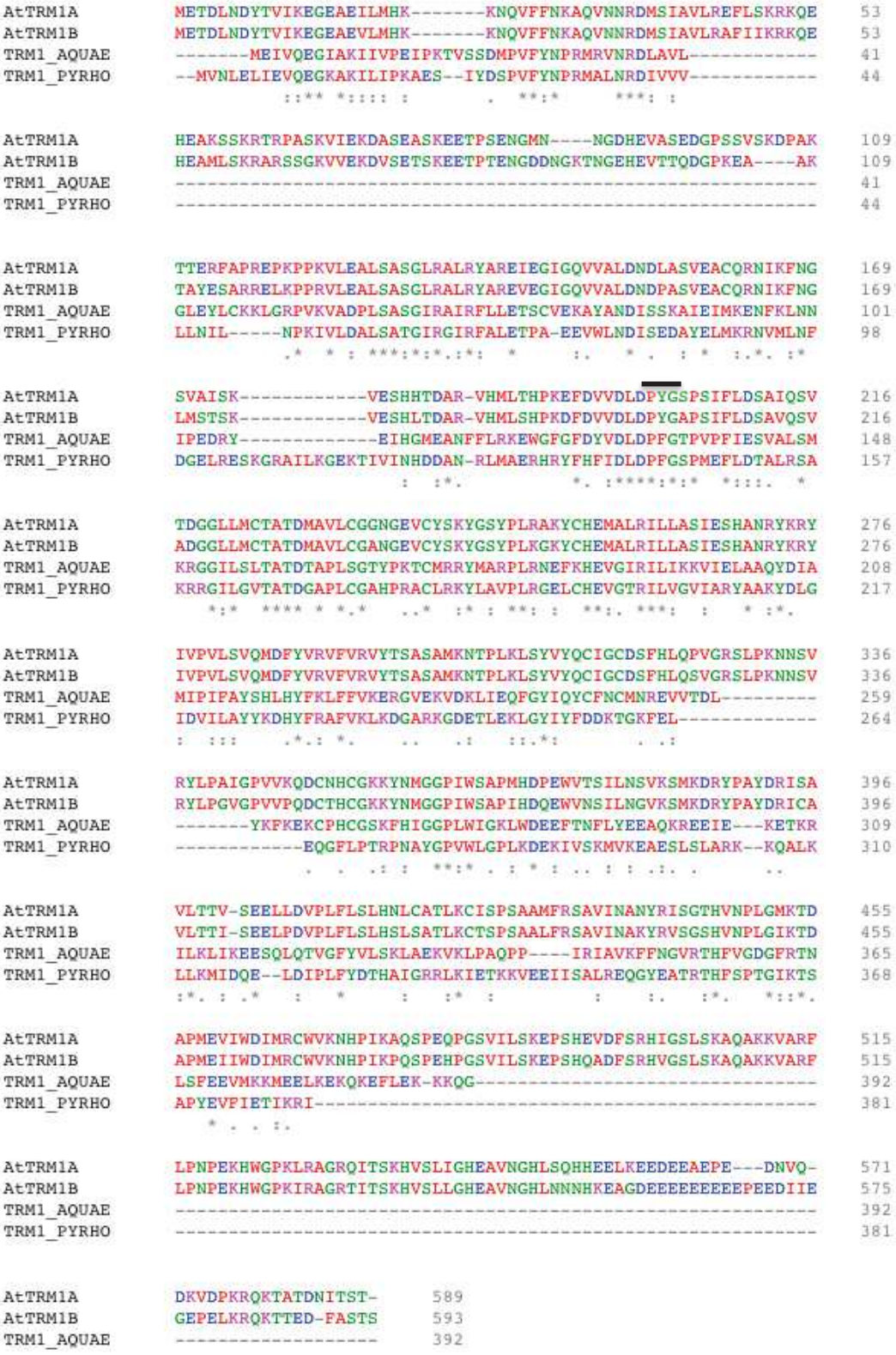


Figure 22: **Clustal Omega alignment of the TRM1 proteins** from *A. aeolicus* (Aquae) and *P. horikoshii* (Pyrho) (both functionally and structurally characterised) with the nuclear AtTRM1A and B. The characteristic DPFG/DPPY sequence motif in the methyltransferase active site (as defined in Swinehart et al, 2014) is indicated by a black line.

to be strictly essential for cell viability (Phyzicki and Hopper, 2010). However, the importance of tRNA methyltransferases can be seen, for example, by the fact that mutations in these enzymes are increasingly associated with human diseases (Blaesius et al., 2018; Davarniya et al., 2015; Najmabadi, 2011).

b) tRNA (N^2,N^2-guanine)-dimethyltransferase (Trm1)

Trm1 is the enzyme that catalyses the methylation of Guanosine at position 26 (and sometime also position 27) in tRNAs. This methylation can consist of either a monomethylation (m^2G) or a dimethylation ($m^{2,2}G$). Whether the reaction is a monomethylation or a demethylation depends on the organism or the tRNA modified (Constantinesco et al., 1999; Ellis, 1986).

Trm1 from both *A. aeolicus* (a Bacteria) and *P. horikoshii* (an Archaea) have been structurally characterized. They appear to share a similar N-terminal domain that contains the methyltransferase active site, including a characteristic DPFG/DPPY sequence motif (Swinehart et al., 2014). This sequence motif contains an aspartate that, according to Ihsanawati et al., 2008, is proposed to act as a general base to remove a proton from the target N2 amino group. It is worth noting that according to another study from the same group, *Aquifex aeolicus* Trm1 catalyses the transfer of methyl groups to guanine 27 of tRNAs in addition to guanine 26 (Awai et al., 2011).

In *Arabidopsis thaliana*, three tRNA (N^2, N^2-guanine)-dimethyltransferases have been predicted: TRM1 A, B and C. Trm1C is predicted to be targeted to organelles (mitochondria and chloroplast), while TRM1A and B are predicted to be nuclear. They have not been functionally or structurally characterized.

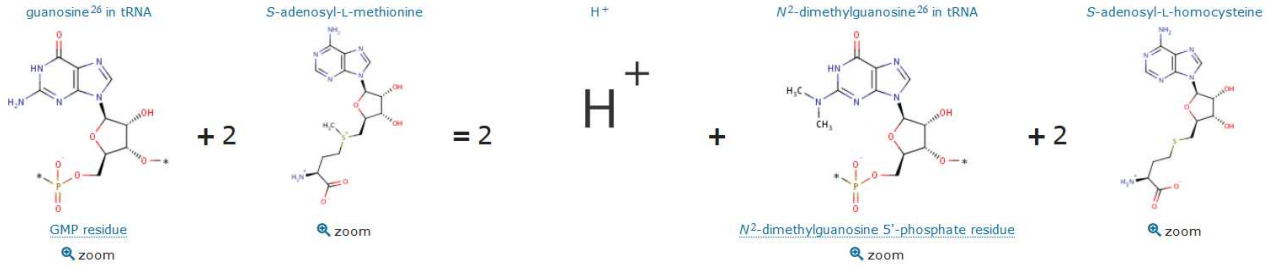


Figure 23: **Schematic of the reaction catalysed by TRM1**. Taken from Uniprot databse, (Q9NXH9 - TRM1_Human).

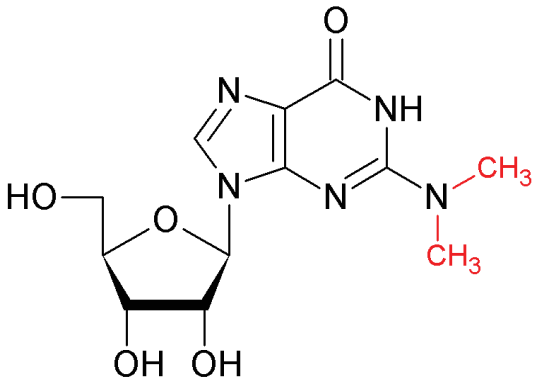


Figure 24: Schematic of the N2,N2-dimethylguanosine (m2,2G) molecule. Image taken from modomics (https://iimcb.genesilico.pl/modomics/).

Material and methods

Material

VII. Plant lines:

All experiments were performed using *Arabidopsis thaliana* as a model organism. The ecotype Columbia (Col-O) was considered as the "Wild Type" (WT) reference in all experiments.
Mutants were obtained from the T-DNA insertion lines: Salk T-DNA, GABI-KAT and WiscDslox. In all cases, T-DNA insertions were confirmed using PCR amplifications and sequencing of the resulting PCR products. To confirm that a line was knockout for the gene of interest, RT-PCR was performed to amplify the coding sequence (CDS) as well as RT-qPCR with primers located after the T-DNA insertion site. Relative expression was calculated and compared to WT.
After germination, plants were grown in soil under short days conditions (12 hours light / 12 hours dark). After transplanting, they were grown in long days conditions (16 hours light / 8 hours dark).
For *in vitro* growth, seeds were sterilized (see Methods) and sown on MS media + 1% sucrose. After stratification in the dark at +4°C for 48h, plants were grown in long days conditions (16h light/8h dark).
For liquid growth, seeds were sterilized and transferred to 100 mL of liquid MS media + 2% sucrose. After stratification for 48 hours at 4°C in the dark, they were left under agitation in longs days conditions (16 h light/8 h dark).

VIII. Bacterial strains

a) Escherichia coli:

E. coli is a Gram-negative Bacteria commonly found in the microbiota of mammals. It has been used since decades in microbiology and molecular biology. Key findings in biology have been made using this model organism like the discovery of operons (Jacob et al., 1960 and 1961), the achievement of genetically modified organisms (Cohen et al, 1973), how DNA replicates (Lehman et al., 1958) and how the genetic code functions (Crick and al., 1961).
E. coli substrains used in biology laboratories all originate from two strains called K-12 and B. For example, the two strains DH5α and DH10B (that was branded under the name TOP10 by ThermoFischer/Invitrogen) both derive from the K-12 strain. The BL21 substrain derives from the B strain.

Strains used for plasmid amplification:

- DH5α strain:

These *E. coli* cells were engineered to increase transformation efficiency. They are KO for endonuclease I (*endA1*) and deficient for homologous recombination (*recA1*). They are considered as useful to clone large plasmids.
Genotype: *F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–*

- TOP10 strain.

TOP10 cells are branded by ThermoFischer. They are probably identical to DH10B (Invitrogen) and NEB 10-beta. They are also recombination and endonuclease I deficient (recA1 / endA1). Chemically competent TOP1O cells were routinely used for plasmid amplification.
Genotype: *str. K-12 F– Δ(ara-leu)7697[Δ(rapA'-cra')] Δ(lac)X74[Δ('yahH-mhpE)] duplication(514341-627601)[nmpC-gltI] galK16 galE15 e14–(icdWT mcrA) φ80dlacZΔM15 recA1 relA1 endA1 Tn10.10 nupG rpsL150(StrR) rph+ spoT1 Δ(mrr-hsdRMS-mcrBC) λ– Missense(dnaA glmS glyQ lpxK mreC murA) Nonsense(chiA gatZ fhuA? yigA ygcG) Frameshift(flhC mglA fruB)*

Strains used for protein production:

- BL21(DE3) strain:

These chemically competent cells contain the T7 RNA polymerase gene in fusion with the IPTG-inducible promoter. They are used to express plasmids with the T7 promoter.
Genotype: *E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*

- Rosetta2(DE3) pLysS strain:

These strains derive from BL21. They were designed to enhance eukaryotic proteins expression. For this purpose, they were supplemented with the rare codons AGG, AGA, AUA, CUA, CCC and GGA. These tRNA genes are carried by a chloramphenicol resistant plasmid.
Genotype: *E. coli str. B F– ompT gal dcm lon? hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL orip15A](CmR)*

b) Agrobacterium tumefaciens:

- GV3101

This strain has a C58 chromosomal background (with Rifampicin resistance). The Ti plasmid pMP90 has gentamycine resistance. T-DNA sequence was removed from the Ti plasmid. As such, only transformation with a binary vector containing the Ti-region will give a T-DNA binary system allowing transfer of DNA into the genome of the transformed plant. The GV3101 strain has Gentamycin and Rifampicin resistance. A third antibiotic is used for the binary vector.

IX. Plasmids:

Plasmids are extrachromosomal DNA molecules that replicates autonomously. The term was coined by J. Lederberg in 1952. Plasmids can be found in bacteria, archaea and some Eukaryotes (Summers, 1996).
The first cloning plasmids were created in the 1970s. Plasmids used in laboratory contain several important features including: selection markers, a replication origin, regulatory elements to control expression and a (relatively) small size to facilitate bacterial transformation.

a) For PCR cloning:

- pGEM®-T easy (Promega):

This pre-linearized vector contains 3'- T overhang at the insertion site. T7 promoter regions flank the multiple cloning site (MCS). The MCS region is inserted in the peptide coding region for β-galactosidase. Allowing blue-white screen selection of transformed clones. Inserted DNA can be removed by a single restriction enzyme digestion with BstZI, EcoRI or NotI.

- pDONOR 207

This plasmid is a Gateway(TM)-adapted entry vector. It is designed to generate attL flanked entry clones containing the gene of interest. Insertion of the desired DNA sequence is made by recombination with an attB expression clone or PCR product. p207 contains the sequence of the ccdB gene between the attP sites for negative selection.
This plasmid has gentamycin resistance.

TRANSFORMATION vector:

- pEarleyGate

pEarleyGate vectors are Gateway®-compatible plant transformation vectors. They were developed by the lab of Craig Pikaard. They allow Agrobacterium-mediated plant transformation that translationally fuse different tags (FLAG, HA, cMyc for example) or tandem affinity purification epitope tags onto target proteins. An adjacent fluorescent protein can also be added.

For PROTEIN Expression

- pET28B for AtPRORP2 and AtTRZ3 expression

This plasmid is a Bacterial expression vector containing the T7 promoter. It possesses kanamycin resistance. DNA insertion is made by restriction enzyme cloning.

X. PRIMERS:

The list of all primers used here is given in Table 2.

To design primers, we used different websites, including but not restricted to primer3plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi),
NCBI primers (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), qprimerDB (https://biodb.swu.edu.cn/qprimerdb/)
To design the genotyping primers of T-DNA insertion lines, the T-DNA primer design software available on the SIGnAL website was used.
Primer specificity was checked using the "BLAST" tool on TAIR.

Primers were all ordered from IDT®.

Primer ID	Gene ID	Primer sequence
RT-qPCR		
qPCR AtMAF e6e7 Fw	At5g13240	CGA AAG GCT TTG ATC TAC TTG G
qPCR AtMAF e6e7 Rv	At5g13240	AAA GGT GTC CCA GCT TTC CT
qPCR AtRNaseZ3 Fw	At1G52160	TTT GCT CAG AGA CAG CAG GA
qPCR AtRNase Z3 Rv	At1G52160	CCC CAT ACA TTC ACC GAA AG
qPCR AtNRPC1 Fw	At5g60040	CGA AAC AGT GAG CTA ATA TCA GGA
qPCR AtNRPC1 Rv	At5g60040	CCT AGA GGA AAA ATA TCG AGT GCT A
qPCR AtNRPC7 Fw	At1g06790	CAA AGC CTT TTG CTC CAA TG
qPCR AtNRPC7 Rv	At1g06790	AAG AAA CAG GAC CCA AAC CA
qPCR 18S rRNA Fw		TGG AGG GCA AGT CTG GTG CC
qPCR 18S rRNA Rv		CGG CCG ACC CAT CCC AAG G
qPCR AtMAF1_Fw (Ahn et al.)	At5g13240	CGA GGA AGG AGC AAT ATG GT
qPCR AtMAF1_Rv (Ahn et al.)	At5g13240	TGC ATC ATT TGC CAG GTT AC
qPCR_UBC Fw	At1g14400	CTG CGA CTC AGG GAA TCT TCT AA
qPCR_UBC Rv	At1g14400	TTG TGC CAT TGA ATT GAA CCC
qPCR TUB4 Fw	At5g44340	AAC GCT GAC GAG TGT ATG GTT
qPCR TUB4 Rv	At5g44340	CCA AAG GTA GGA TTA GCG AGC
qPCR 5S rRNA Fw (Ahn et al.)		GGA TGC GAT CAT ACC AGC
qPCR 5S rRNA Rv (Ahn et al.)		GAG GGA TGC AAC ACG AGG
qPCR 5S rRNA Fw (A. Probst)		GGA TGC GAT CAT ACC AG
qPCR 5S rRNA Rv (A. Probst)		GGG AGG TCA CCC ATC CTA GT
qPCR U6 snRNA Fw		TTG TGG TTT GGC CCT CAC AT
qPCR U6 snRNA Rv		GGG CCA ATT CAC AAG CAA CC
qPCR 7SL SRP_FW_1		TAG TCG CTG CTC TGT GAA GC
qPCR 7SL SRP_Rv 1		GAT CGG TTC TGA TGT GGG CA
qPCR 7SL SRP_FW_2		AGC CTT CGC TCA ACT CAA CA
qPCR 7SL SRP_Rv 2		GGT TCC TCG GAT CGG TTC TG
qPCR AtLa1 Fw	At4g32720	AGA TCT TAA GGC TGT TTT CGG
qPCR AtLa1 Rv	At4g32720	CAC CTT CAT TGG CTA ATA CTG C
qPCR AtLa2 Fw	At1g79880	GGA GAT AAC TCA TGT GCT AGC
qPCR AtLa2 Rv	At1g79880	ACC TCA GAG CTG ACA GAA TTA C
qPCR AtNRPC2 Fw	At5g45140	CCA CGT GTT ATG ATG ACA AGA C
qPCR AtNRPC2 Rv	At5g45140	TCT CTT TCC ATT TCT CCC ACT C
qPCR AtCCase Fw	At1g22660	AGC AAA GTT GCG GAT TTA TGA CCA G
qPCR AtCCase Rv	At1g22660	TGC GGT GCC AAA TTT CAT TGT AGG
Genotyping		
LP GABI_031C10	At5g13240	GAT AGG CAA ACA TGC AGG AAG
RP GABI_031C10	At5g13240	CTT CCT CGA GGA AAG GAT CAG
GABI _031C10 39yr	At5g13240	CTC ATT CCT CTC AGT CCA TTC CTA
T-DNA 8474 Gabi-Kat		ATA ATA ACG CTG CGG ACA TCT ACA TTT T
M13 Fw		GTA AAA CGA CGG CCA GT
M13 Rv		CAGGAAACAGCTATGAC
LP_SALK_018512	At3g02320	TGT GCG TCG ATA GTG TTT GAC
RP_SALK_018512	At3g02320	ACC ATT GAC TGC TTC ATG ACC
LP_SALK_206384	At5g15810	GGC TCT TAT CCA CTG AAA GGG
RP_SALK_206384	At5g15810	ATG ACA TTC AAA GCT TGT GCC
LP_WiscDsLox244G04	At5g15810	TTC GGG CTC TTC TTC TTC TTC
RP_WiscDsLox244G04	At5g15810	CTC TCC CTA AGG TTT TGT GGC
LP GABI_696C04	At1G74700	CTT TCG ACA TTG GTC GT
RP GABI_696C04	At1G74700	TAG AGG TTG AGT TAC CAC C

GABI_696C04_LB		ATATTGACCATCATACTCATTGC
LP_CSHL_GT9807	At1G52160	ATG ATA AAC TCA ATG CCC TAT TTA CAC
RP_CSHL_GT9807	At1G52160	AGT AGG ACT ACT TGT CAC
Ds3		ACC CGA CCG GAT CGT ATC GGT
LbB1.3		ATTTTGCCGATTTCGGAAC
P745 (for WiscDsLox)		AACGTCCGCAATGTGTTATTAAGTTGTC
LP_G74	At2g16650	GGCATCTGTTACTTCAGTGGCTA
RP_G66	At2g16650	CATTTGTTATTACCACTTTCACGG
LP_G39 (P2-HA)	At2g16650	TGCTTGTGACAAATGATGAG
RP_HA tag (Rv)	At2g16650	AGCGTAATCTGGAACATCGTATGG
Flag-LB4		CGTGTGCCAGGTGCCCACGGAATAGT
LP_G14	At4g21900	AGCCGGAGCATAAGTTCTCA
RP_G13	At4g21900	TGCATGTTCCTTAAACGCTATG
LBa1		TGGTTCACGTAGTGGGCCATCG
Gene amplification		
Promoteur AtMAF1 Fw	At5g13240	TCT TTC TGC ATT GCA TGT CC
Promoteur AtMAF 1 Rv	At5g13240	GGT CGT TAA TTT CCA AAA TTT GTA TTT CTT TGA ACC G
CDS AtMAF 1 Fw	At5g13240	ATG AAG TTC TTG GAG TAC ACT AAC C
CDS AtMAF 1 Rv	At5g13240	TCA CAT ATC CAT GTC GGC
AtMAF1 sequencing Fw_1	At5g13240	GAG GAG ACA ACC AAA ATA GAA AAT TAA CCC GAG
AtMAF1 sequencing Rv_1	At5g13240	TCT CTT TCT ATT ACG AGG CTA AGA TCT CCC
AtMAF1 sequencing Fw_2	At5g13240	AGA AAC GAA AAC CCG CAA CG
AtMAF1 – i6 overlapping Fw	At5g13240	ACG ATT TCA GCG CTG TCA AAG CAC ATC AG
AtMAF1 – i6 overlapping Rv	At5g13240	TTG ACA GCG CTG AAA TCG TAG TCC GGA TAC
AtMAF1 – TLS overlapping Fw	At5g13240	ATA TGT CTA AAA TCT AAA ACA AAC TTG
AtMAF1 – TLS overlapping Rv	At5g13240	AGA TTT TAG ACA TAT TTA AAG GGG G
AtTRM1A CDS Fw	At3g02320	ATG GAA ACG GAT CTG AAT G
AtTRM1A CDS Rv	At3g02320	TTA TGT TGA GGT AAT ATT ATC TGT TGC
AtTRM1B CDS Fw	At5g15810	ATGGAAACGGATCTCAATG
AtTRM1B CDS Rv	At5g15810	CTATGATGTTGAGGCAAAATC
Sequencing primer AtTRM1A/B Fw	At3g02320 At5g15810	TTT GAT GTG GTT GAT CTT G
5' and 3' RACE		
5'RACE Rv (RT + External PCR)	At5g13240	TGA CAG CGC TGG AAA CGA GAA AAT AG
5'RACE Rv Internal PCR	At5g13240	TGC GAT GGT GAT AGA GTT ATA TAC
3'RACE Fw External PCR	At5g13240	CCG GAC TAC GAT TTC AGG TAA TTG TG
3'RACE Fw Internal PCR	At5g13240	GTG TTG TAA ACC TGA AGG G
illumina 3'raceMAF internal PCR	At5g13240	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GAC GAT CGT GTT GTA AAC CTG AAG GG
R21 primer		/5rApp/CTGACNNNNNNNNNNNNNNTGGAATTCTCGGGT GCCAAGGC/3ddC/
Rv primer against R21		CCTTGGCACCCGAGAATTCCA
5'RACE adapter primer		GAACACUGCGUUUGCUGGCUUUGAUGAAA
5' RACE fw		GAACACTGCGTTTGCTGGCTTTGATG
Primers for cloning		
AtMAF CDS attb1 Fw	At5g13240	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTA TGA AGT TCT TGG AGT ACA CTA ACC
AtMAF CDS attb2 Rv	At5g13240	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT CAC ATA TCC ATG TCG GC
AtMAF Full Length Fw Attb1	At5g13240	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CTT TCT GCA TTG CAT GTC C

AtMAF Full Length Rv Attb2	At5g13240	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC GTA CAT AAA CAC AAA ATC TCA AC
AtMAF1 CDS attb1 Fw N-ter tag	At5g13240	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GAA GTT CTT GGA GTA CAC TAA CC
AtMAF1 CDS attb2 Rv C-ter tag	At5g13240	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTG CAT ATC CAT GTC GGC
AtPRORP2 CDS attb1 N-ter tag	At2g16650	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTG GAT GGC TGC TTC TGA TCA AC
AtPRORP2 CDS attb2 C-ter tag	At2g16650	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT AGG AAT CTT CCC ATT ACT CTT
AtTRZ3 attb1 CDS N-ter tag	At1G52160	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTG GAT GAT AAA CTC AAT GCC C
AtTRZ3 attb2 CDS C-ter tag	At1G52160	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTG CAA AGC TTC TTC CTT TAA GTC
AtTRM1A attb1 CDS N-ter tag Fw	At3g02320	GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC ATGGAAACGG ATCTGAATG
AtTRM1A attb1 CDS C-ter tag Rv	At3g02320	GGGG AC CAC TTT GTA CAA GAA AGC TGG GTG TGTTGAGGTAATATTAT CTGTTGC
AtTRM1A attb1 CDS Rv (with Stop codon)	At3g02320	GGGG AC CAC TTT GTA CAA GAA AGC TGG GTT TTA TTATGTTGAGGTAATATTAT CTGTTGC
AtTRM1B attb1 CDS N-ter tag Fw	At5g15810	GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC ATGGAAACGGATCTCAATG
AtTRM1B attb1 CDS Rv (with Stop codon)	At5g15810	GGGG AC CAC TTT GTA CAA GAA AGC TGG GTT TTA CTATGATGTTGAGGCAAAATC
In vitro cleavage and c-RT-PCR		
AtMAF1_e6_Fw	At5g13240	TCG AAA GGC TTT GAT CTA CTT GG
AtMAF1_e7_Rv	At5g13240	TCT GCT TAA AGG TGT CCC AGC
AtMAF1_e6_T7_Fw	At5g13240	GAA TTG TAA TAC GAC TCA CTA TAG GGT CGA AAG GCT TTG ATC TAC TTG G
RT et Rv cPCR	At5g13240	TCT ACA GTT CAA TGA ACC G
Fw S_RNA cPCR	At5g13240	TCA ATT CTC TCC ATT TCC C
Fw L_RNA cPCR	At5g13240	GCT GTC AAA GCA CAT CAG

Table 2: list of primers used in this study

XI. Plant T-DNA lines:

T-DNA lines used in this study are the following. They were all verified by PCR using primers designed by the SIGnal website (http://signal.salk.edu/tdnaprimers.2.html).

At2g16650 : *prorp2* : EHQ_F435C08
At4g21900 : *prorp3* : Salk_126397
At5g13240 : *maf1* : Gabi-Kat_031C10
At1g74700 : *trz1* : Gabi-Kat_696C04
At1g52160 : *trz3* : CSHL_GT9807
At3g02320 : *trm1A* : Salk_018512
At5g15810 : *trm1B* : Salk_206384
At5g15810 : *trm1B* : WiscDsLox244G04
At2g06990 : *hen2-2* : Gabi-Kat_774H07

XII. Antibodies:

- Primary antibodies:

The primary antibody against HA peptide was raised in mouse. It was obtained from Sigma-Aldrich®. It is a monoclonal Anti-HA tag antibody produced in mouse. It enables the detection of HA-tagged proteins. Following the manufacturers ' instructions, this antibody was used, for Western Blot, at a dilution of 1/10 000, and a dilution of 1/100 for immunoTEM.

- Secondary antibodies:

GAM (Goat against Mouse) or GAR (Goat against Rabbit) secondary antibodies were used depending on the primary antibodies. These secondary antibodies, conjugated with horseradish peroxidase were used at a 1/10 000 dilution and visualized with enhanced chemi-luminescent reagents (Roche).

Methods

XIII. PLANTS:

a) Crossing:

Arabidopsis crossing was performed according to the following procedure. Flowers were appropriate for crossing when the stigma is just pointing out but the stamens are not yet mature. Surrounding mature flowers are snipped away so that they do not pollinate the chosen flower. The stigma should look crystalline and the surrounding anthers should not have matured. Surrounding siliques are removed so that the seed from the cross is not contaminated. Some of the flower petals can be removed if possible not the sepals. The immature anthers - there should be 6; 4 long and 2 short are also removed.
The female flower is held with tweezers and the flower should splay out. Pollen from chosen anther is applied with other tweezers. A lot of pollen is needed to ensure the maximum production of 40 seeds per silique. The base of the flower treated is marked with ink. Bagging is performed one week later, removing any flowers that have formed subsequently. Crosses should be performed both ways to confirm nuclear inheritance i.e male and female roles swopped.

b) Agro-transformation of Arabidopsis by floral dip:

Floral dip was performed following the method established by Clough et al., 1998. *Agrobacterium* were grown overnight in 5 ml LB medium with the appropriate antibiotics at 28°C under agitation. The following day, 1.5 ml of overnight culture was incubated into 100 ml of LB medium with appropriate antibiotics and incubation continued at 28°C for 20 hours. Agrobacterium were collected by centrifugation at 4000 rpm for 10 min. Then resuspended in equal volume of infiltration medium and incubated for 1 hour at RT. Inflorescences of the plants were soaked in the solution for 30 seconds, drained and blacked out for 24 hours. Then returned to their original conditions.

Infiltration medium :

	Final	*Stock*	*1 L*	*600 ml*	*400 ml*	*200 ml*
MS medium + vit + MES (M0255)	*½*	*1*	*2.165*	*1.299 g*	*0.866 g*	*0.433 g*
Sucrose	*5 %*		*50 g*	*30 g*	*20 g*	*10 g*
Silvet L-77	*1/2000*		*500*	*300 µl*	*200 µl*	*100 µl*
Acetosynringone	*200 µM*	*200 mM*	*1 ml*	*600 µl*	*400 µl*	*200 µl*

c) VIGS

Constructions used were the same as in Gutmann et al. 2012. The method used to perform Virus Induced Gene Silencing is adapted from Burch-Smith et al., 2006.
The TRV virus is composed of 2 genomic RNAs encoded by 2 plasmids pTRV1 and pTRV2 that will be transformed in Arabidopsis plants by *Agrobacterium* infiltration. The sequence of interest corresponding to the gene that should be down-regulated *in vivo* is cloned in pTRV2. Briefly, *Agrobacterium* were grown in 5 mL LB and appropriate antibiotics overnight. Then 1 mL was used to inoculate a 20 mL culture of LB and appropriate antibiotics. This was done for pTRV1 and pTRV2. After centrifugation 4000 g for 10 minutes, bacteria were resuspended in water. After centrifugation they were resuspended in agroinfiltration buffer (MES 10 mM, MgCl2 10 mM and AcetosSyringone 150 µM). DO_{600} was adjusted to 1,5 and Ag*robacterium* containing pTRV1 and pTRV2 were mixed at equal volume. After 2 hours incubation at RT, 3 leaves from young plantlets (15 to 17 days post germination) were infiltrated with a 1 mL syringe. Phenotype was visible from 10 to 12 days post infection. For molecular analyses, samples were taken 3 weeks after infiltration. pTRV2 empty vector (EV) is the negative control. pTRV2 expressing a phytoene desaturase (PDS) sequence is the positive control. The down regulation of PDS gene expression leads to a white leaves phenotype.

d) SEEDS sterilisation

A maximum of 100 µl of Arabidopsis seeds was used per tube. All steps were performed under a sterile flow hood. 1 ml of 70% ethanol was added. After 1 minute, seeds were washed 5 times by adding and removing 1 mL of H_2O. Then 1 mL of bleach (50%) was added. Then, seeds were rotated for 5 minutes on a wheel. After 5 minutes, seeds were washed 5 times by adding and removing 1 mL of H_2O.

XIV. DNA methods:

a) DNA extraction:

DNA was extracted using 3 different methods.

1) Rapid Arabidopsis DNA extraction (adapted from Edwards et al. 1991)

Approximately 100 µl of sterile glass beads were placed in a screw-cap tube with 400 µl of extraction buffer (200 mM Tris-HCL pH 7,5, 250 mM NaCl, 25 mM EDTA). 1 cm^2 of leaves was added and cells were disrupted by 2 cycles of 30 seconds at 5500 rpm with 30 seconds break in between, in a Precellys® device. After centrifugation at 16 000 g (4°c) for 5 minutes, 200 µl of the supernatant was transferred to a new Eppendorf tube containing 150 µl of cold isopropanol. After vortex and incubation 5 minutes at RT the DNA is pelleted at 16 000 g (4°C) for 10 minutes. Supernatant was removed and the pellet washed with 1 mL of 70% ethanol. After centrifugation at 16 000 g (4°C) for 5 min, the supernatant was removed and the pellet air-dried before being resuspended in 50 µl of sterile water.

2) CTAB Plant genomic DNA extraction (modified by Rahul Patharkar)

300 µl of 2X CTAB buffer (2% (w/v) CTAB, 1,4 M NaCl, 100 mM Tris-HCL pH 8, 20 mM EDTA) was added to a 1.5 mL tube. Two small leaves were added and grinded with a plastic pestles, then incubated at 65°C for 10 minutes (or longer). After cooling, 300 µl of chloroform was added and the tube was vortexed. After centrifugation at 16 000 g (4°C) for 5 minutes, the upper phase was transferred to a new tube containing 180 µl of isopropanol. After vortex, the tube was centrifuged 5 minutes at 16 000 g (4°C). Supernatant was removed and the pellet washed with 500 µl of 70% ethanol. After centrifugation at 16 000 g (4°C) for 5 min, the supernatant was removed. To ensure proper removal of residual ethanol, a short spin of 5 seconds was performed and the remaining liquid was removed. The pellet was air-dried for 5 minutes before being resuspended in 100 µl of sterile water.
Note: 1 µl of linear polyacrylamide (LPA) 20 µg/µl can be added in the buffer at the beginning of the extraction. It acts as a carrier for DNA precipitation and will produce a visible white pellet.

3) Plant DNA extraction method (Method from the lab of Ian Small, UWA, Perth, Australia).

This method was used whenever large quantities of high-quality DNA were required. It was used with different plant species (*Arabidopsis thaliana*, *Solanum lycopersicum*

Brachypodium distachyon) and proved to be the most efficient DNA extraction method out of the three protocols described here. However, this procedure is also the more time consuming.
Glass beads were placed into Precellys tubes containing 500 µl of extraction buffer (0.1 M Tris-HCl pH 8, 0.05 M EDTA, 0.5 M NaCl, 1 % (w/v) PVP40). Two 1 cm^2 leaves were added and tubes were mixed at 5500 g 2x 30 seconds with 1 break of 30 seconds. 66 µl of 10 % SDS was added and mixed with a vortex. After incubation at RT for 15 minutes, tubes were centrifuged at max speed for 10 min. Supernatant (550 µl) was removed to a new 1,5 ml tube containing 166 µl 5 M potassium acetate and mixed. After centrifugation at max speed for 15 minutes (4°C), the supernatant (600 – 700 µl) was carefully pipetted into a new tube containing 0.7 volume of isopropanol and mixed by inverting several times. Tubes were incubated at -20°C for 15 minutes and centrifuged at maximum speed 15 min (4°C). The isopropanol was carefully poured off to retain the DNA pellet at the bottom of the tube. 500 µl of 70 % ethanol was added and mixed gently by tube inversions to wash the pellet. After centrifugation at maximum speed for 5 minutes (4°C), the ethanol was poured off, and the pellet air dried. 50µl of sterile H_20 was used to resuspend the pellet.

b) Phenol / Chloroform extraction:

An equal volume of Tris-HCL saturated phenol/chloroform (pH 8 for DNA and pH 6,8 for RNA) was added to a DNA or RNA sample. After vortexing for 20 seconds, the tube is centrifuged 5 min at full speed. The aqueous layer, containing the nucleic acids is transferred to a new 1,5 mL tube.

c) Ethanol precipitation

To concentrate nucleic acids, 1/10 (volume/volume) of 3M sodium acetate pH 5.2 is added. Then 1 µl of glycogen and three volumes 100% ethanol are added. The mixture is homogenized by pipetting well and inverting and incubated at -80°C for 1 hour or – 20°C overnight. The sample is then centrifuged at 16000 g for 25 min at 4°C. Supernatant is discarded and the pellet washed with 70% ethanol. After centrifugation 16000 g, 5 min at 4°C, the supernatant is discarded and the pellet air dried. Finally, the DNA/RNA is resuspended with H_2O.

d) DNA amplification by polymerase chain reaction (PCR)

PCR was performed using two different polymerases.

a) GOtaq G2 flexi DNA polymerase (Promega)

This polymerase is used for plant genotyping or colony PCR (to check the presence and size of the DNA insert).
Reaction mix was as follow:
5X buffer = 4 µl
MgCl2 (25 mM) = 1,2 µl
dNTPs (10 mM) = 0,4 µl
Primers (FW and RV, 10mM)=0,4 µl each
DNA = 1 µl at 100 ng/µl
Go Taq polymerase = 0,2 µl
H20 = 12,4 µl
V_{total} = 20 µl

PCR program:
(i) initial DNA denaturation step 95°C – 3 min
(ii) DNA denaturation 95°C- 30 sec
(iii) Primers annealing X°C – 30 sec (temperature was set depending on primers TM)
(iv) DNA elongation 72°C – X sec (elongation time depends on the desired length of the PCR products. We usually use 1 min/Kb using GoTaq polymerase)
Steps ii to iv are repeated between 25 and 35 times.
(v) DNA elongation termination 72°C – 3 minutes (can be shortened or increased)
10°C – hold

b) Phusion High Fidelity DNA polymerase (Thermo Fischer):

This polymerase is used for high fidelity amplification of DNA. This is achieved by an exonuclease 3' to 5' "proofreading" activity.
The PCR mix used for this polymerase was usually:
5X buffer = 4 µl
(The High fidelity (HF) buffer is the default buffer used for amplification. However, a "GC" buffer is also provided and should be used for difficult or long DNA templates)
dNTPS (10 mM) = 0,4 µl
Primers (Fw and Rv – 10 mM) = 0,6 µl each
DMSO (if needed) = 0,6 µl
H_2O = 13,2 or 12,6 µl (depending if DMSO is added).
Phusion polymerase = 0,2 µl
V_{total} = 20 µl

The PCR program used was as follow.
Thermal cycler was pre-heated to 98°C
(i) 98°C 30 sec
(ii) 98°C 10 sec
(iii) X°C 30 sec (Annealing temperature was set depending on primers TM)

(iv) 72°C – X sec (elongation time is depending on the desired length of the PCR products. We usually use 30 sec/Kb with Phusion polymerase)
Steps ii to iv are repeated between 25 and 35 times.
(v) 72°C – 3 minutes (can be shortened or increased)
10°C – hold

e) Agarose Gel Electrophoresis

Agarose Gel Electrophoresis allows separation of DNA molecules depending on their molecular weight. Loading buffer is added to the DNA solution (if not already present, as in the GoTaq green buffer (Promega)). This solution is loaded on an 0,75 to 2% (w/v) agarose gel in 0,5X TAE buffer (40 mM Tris-acetate pH 8, 1 mM EDTA). Electrophoresis is carried out at 50-100 V in 0,5X TAE buffer. Ethidium bromide, an intercalating agent added to the gel, allows visualisation of DNA molecules under UV light.

f) Nucleic acids quantification

DNA and RNA were quantified at 260 nm using a NanoDrop 2000c spectrophotometer (ThermoFisher) according to manufacturer's instructions.

g) Sanger DNA Sequencing

Sanger sequencing was used to determine the sequence of PCR products or DNA insertions in plasmids. It was performed at the IBMP "Gene expression Analysis" platform using an Applied Biosystems 3100 apparatus.

XV. RNA methods:

a) Plant total RNA extraction:

Small amounts of plants leaves were flash frozen in a 1,5 mL tube containing 100 µl of sterile glass beads. Samples were grinded 8 seconds with a Silamat® grinder (it is advised to open the tube before grinding to avoid its explosion). Then samples were put on ice and 750 µl of Tri-reagent (MRC) was added. Samples were grinded 12 seconds with the Silamat, then vortexed and incubated at RT for 5 min. 250 µl of chloroform was added and

the sample was vortexed. Tubes were centrifuged 5 minutes at 16 000 g (4°C). The upper aqueous phase containing the RNAs was transferred to a new tube and 250 µl of isopropanol (0,5 ml/1mL trizol) was added. After vortexing and 5' incubation at RT, tubes were centrifuged 8 minutes at 16 000 g (4°C). Supernatant was removed and pellet washed with 500 µl of ethanol 70%. After centrifugation at 16 000 g (4°C) for 5 min, the supernatant was removed. To ensure proper removal of residual ethanol, a short spin of 5 seconds was performed and the remaining liquid was removed. The pellet was air-dried before being resuspended in 20 µl of sterile water.

b) DNase treatment:

DNase treatment was performed using the DNase I enzyme. Reaction was performed in a total volume of 100 µl. 10 µl of DNase I + MgCl2 (10X buffer) was added to 1 to 10 µg of RNA. DNase I (ThermoFisher) was added in a ratio of 1U per 1 µg of RNA with sterile water to reach 100 µl. Reaction was incubated 30 minutes at 37°C and stopped using phenol/chloroform extraction.

c) Agarose gel to ensure RNA integrity:

To check for any RNA degradation, the usual method is to run 100 ng of purified RNA on a formaldehyde denaturing gel. However this process is time consuming and environment non-friendly. For routine check of RNA integrity, a simple non denaturing agarose gel can be used. 100 ng of RNA is simply added to 1 µl of "RNA stop" solution (40%(v/v formamide, 10 mM EDTA, 0.0025% bromophenol blue, 0.0025% xylene cyanol) loading buffer and electrophoresis is carried out at 100 V for 15 min. After visualization under UV light, RNAs are considered as non-degraded if two discrete bands can be seen for the 18S and 28S ribosomal RNAs, with no smear.

d) Gel purification of RNA:

To extract RNA from a denaturing gel, we used the protocol of T.W.Nilsen, 2013. Briefly, the band of interest is removed from the gel using a razor blade and placed in a microcentrifuge tube with 400 µl of gel elution buffer (20 mM Tris-HCl pH 7.5, 0,25 M Sodium acetate, 1 mM EDTA, 0,25% SDS). After freezing for 15 min at -80°C, the RNA is left to diffuse overnight by leaving the tube at room temperature. After centrifugation at room temperature for 10 min at maximum speed, the supernatant is transferred to a new tube and RNA is retrieved using phenol-chloroform followed by ethanol precipitation.

e) Reverse Transcription:

cDNAs synthesis was performed using purified RNAs, after DNA removal using DNase I (ThermoFisher). For all purposes (RT-qPCR, 3' or 5' RACE), an equal quantity of RNA was used for all samples. Reverse Transcription was performed using either Superscript IV RT (ThermoFisher) or GoScript™ Reverse Transcriptase (Promega). In each case the supplier's instructions were followed. To achieve global reverse transcription of RNAs a mix of oligo dT primers and random hexamer primers was used. For reverse transcription of specific RNAs (for example in 3' and 5' race) a specific primer was used.

f) RT-qPCR:

RT-qPCR is a method used to determine the relative level of expression of given genes. The amount of cDNAs is measured by the number of cycle necessary to reach a crossing threshold (Ct), defined as a given amount of fluorescence measured. By comparing to the relative level of "housekeeping" genes (various endogenous controls were used, see primers list for details), the relative expression level of different genes can be compared between samples. Briefly cDNAs were diluted 10 fold and mixed with 250 mM of sense and antisense primers and 5 µl of Eurogentec Takyon™ SYBR® 2X qPCR Mastermix Blue. Reactions were performed with three technical replicates in 384-well plates with a LightCycler 480 Real-Time PCR System (Roche). PCR program was as follow (initial denaturation step of 5 minutes followed by 40 cycles of 10 seconds at 95°C, 15 seconds at 60°C and 15 seconds at 72°C). PCR efficiencies were calculated for all reactions. Only PCR efficiencies superior to 0,9 were retained. After ensuring that all technical replicates have less than 0,5 Ct of difference, the relative expression is calculated using the delta Ct method.

g) cRT-PCR

Cleaved RNA was extracted from gel using the method described by T.W.Nilsen, 2013. Then 1 µg of RNA was ligated using T4 RNA ligase (NEB) according to manufacturers' instructions. Reaction was carried at 37°C for 1 hour and enzyme heat inactivated at 65°C for 15 min. Reverse transcription, using circularised RNA as a matrix was performed using Superscript IV (Thermo fisher) following manufacturers' instructions. A specific primer was designed to match the expected circularised RNA sequence. Then nested PCR was performed using 1µl of cDNA as a template. The final PCR reaction was submitted to PCR clean-up using the Macherey Nagel PCR clean-up kit. A-tailling was performed on the eluted DNA using Go Taq Polymerase (Promega) and DNA was then inserted in pGEM (Promega) according to manufacturers' instructions. Insert was sequenced using M13 primers by Sanger sequencing.

h) 3' and 5' rapid amplification of cDNA ends (RACE):

3'RACE:

3' RACE was performed by adapting the protocol of Scheer et al., 2020. Total RNA extraction and DNase treatment was performed as indicated above. After RNA integrity was checked on an agarose gel, 5 µg of purified RNA was ligated to 1 µl of 10 µM R21 adpator. This primer Primer is 5'adenylated (5'rApp) and as such can act as substrates for T4 ligases in the absence of ATP, preventing from adenylation of unwanted product (rRNA, tRNA and miRNA). Additionally this primer is blocked in 3' by a dideoxycytosine, preventing the formation of primer concatemer. Primer sequence is /5rApp/CTGACNNNNNNNNNNNNNNNTGGAATTCTCGGGTGCCAAGGC/3ddC/
Water was added to a volume of 44 µl and the reaction was incubated at 65°C for 3 minutes and then put on ice for 2 minutes. 5µl of T4 RNA ligase buffer (NEB) was added as well as 1 µl of T4 RNA ligase 1 (ssRNA, NEB). After incubation for 1 hour at 37°C, the ligated RNA products were purified using the Macheley Nagel Nucleospin RNA plant kit
2,5 µg of ligated RNA was used for reverse transcription (RT). RT was performed using SSIV (Thermo) according to the manufacturer's instruction. For this reaction, a RT short primer (A1467) was used, allowing only amplification of adapter-ligated RNAs. Amplification was performed at 56°C, which is the maximal temperature at which the enzyme has 100% efficiency, according to the manufacturer.
After RT, RNase H was added (1µl / reaction). The resulting cDNAs of two RT reaction / sample were pooled, subjected to phenol-chloroform and 1 µl was used as template for nested PCR. Final PCR products were purified using AMPure beads (Agencourt). DNA quantity was determined using Qubit® and profile was checker using a Bioanalyer® chip. After librairy preparation, purified PCR products were sequenced with the Illumina MiSeq system.

5' RACE:

Total RNA extraction, DNase treatment and adaptor ligation was performed as indicated for 3' RACE. As we were investigating the termini of a cleaved RNA, no dephosphorylation or decapping step was needed. RT was performed essentially as for 3' RACE except that the specific primer for RT was designed against the region of interest of the MAF1 gene (our primer was located at the junction of the TLS containing intron and the following exon). Nested PCR was then performed and the resulting PCR products were purified using the Macherey Nagel PCR clean-up kit. After insertion in pGEM-T-easy (Promega), colonies were sequenced in 96 well plates using Sanger Sequencing. Sequencing results were aligned to the sequence of reference using MacVector®.

i) Denaturing acrylamide gel

Denaturing acrylamide gels were used to separate RNAs of small sizes. 250 ng of RNA were usually used. After resuspension in loading buffer (40%(v/v) formamide, 10 mM EDTA, 0.0025% bromophenol blue, 0.0025% xylene cyanol), samples were denatured 5 min at 70°C and loaded on a gel (concentration of the gel depends on the size of the RNA) and migration was performed at 200 V, 25 mA, 10 W, for 1h -1h30. Migration takes place in TBE buffer (90 mM Tris-HCl pH 8, 2 mM EDTA, 90 mM borate).
The table below lists the different elements used depending on the concentration needed.

	6% (50 ml)	*8% (50 ml)*	*12% (50 ml)*	*17% (50 ml)*
Urea 7M	*21 g*	*21 g*	*21 g*	*21 g*
Acryl 19.1 40%	*7.5 ml*	*10 ml*	*15 ml*	*21.3 ml*
TBE 5x	*10 ml*	*10 ml*	*10 ml*	*10 ml*

j) Denaturing Agarose gel

The table below lists the different elements used depending on the concentration needed.

	1.2%	*1.5%*		
	Pour 100 ml	*Pour 100 ml*	*12 wells*	*6 wells*
Agarose (g)	*1.2*	*1.5*	*1.2*	*0.75*
H2O (ml)		*72*	*57.6*	*36*
MOPS 10x (ml)	*10*	*10*	*8*	*5*
Formaldéhyde (ml)	*5*	*18*	*14.4*	*9*
Volume final (ml)	*100*	*100*	*80*	*50*
Tampon de migration				
MOPS	*1x*	*1x*		
Formaldéhyde	*5%*	*2.5%*		

Formaldehyde is added under the fume hood, after agarose was melted. Gel is prepared and run under the hood.

Samples are mixed with RNA loading buffer (50% formamide, 10% formaldehyde, 1x MOPS, 10% glycerol, 25 ng/µl BET, BBP) and denatured 5 min at 70°C before being loaded.

k) T7 RNA In vitro transcription

T7 RNA transcription was performed using the ThermoFisher TranscriptAid T7 High Yield Transcription Kit following the manufacturer instruction. PCR was performed with Fw primer containing the T7 promoter on DNA of interest inserted in a plasmid (usually pGEM-T-easy). PCR products were used as templates for in vitro transcription.

XVI. CLONING & BACTERIAL TRANSFORMATION

a) Competent E. coli preparation

A single colony was grown overnight into 5 ml LB at 37°C. The next day, 1 ml culture was transferred into 500 ml (LB + 5 ml 1 M $MgCl_2$), grown at 18°C until OD_{596} = 0.25-0.7 (This takes 2-3 days (start 500 ml culture 9.00 am day 1; day 3 at 9.00 am OD_{596} = 0.570).
Culture was then placed on ice 30 minutes and harvested at 4500 rpm for 10 min at 4°C.
Bacteria were resuspended in 80 ml TB buffer (ice-cold) and put on ice for 10 min.
Bacteria were harvested all again and resuspended in 20 ml TB (ice-cold) +1.4 ml of DMSO (cold) (final concentration: 7%), then put on ice for 10 min.
Aliquot (100 µl) were made, then flash frozen in liquid nitrogen, and kept at -80°C

TB buffer

	Final concentration	Mw	g/100ml
Pipes	10 mM	302.4	0.302
$MnCl_2$	55 mM	197.92	1.088
$CaCl_2$	15 mM	147.02	0.22
KCl	250 mM	74.55	1.86

Dissolve all components except $MnCl_2$, are dissolved and pH adjusted to 6.7 with KOH. Then $MnCl_2$ is added and solution is sterilised using a pre-rinsed filter.

b) Competent A. tumefaciens preparation

2 *Agrobacterium* colonies (strain GV3101::pMP90) were grown O/N at 28°C, 200 rpm in 4 mL of LB + appropriate antibiotic. This pre-culture was then transferred to 100 mL of LB + appropriate antibiotic, and incubated at 28°C, 200 rpm until OD_{600nm} reached between 0.5 and 1 (usually at least 6). Culture was then chilled on ice for 10 to 30 minutes and centrifuged at 3000 g 5 minutes (4°C). Pellet was resuspended in 20 mL of ice cold 2à mM $CaCL_2$ solution. After 5 minutes centrifugation at 3000 g (4°C), pellets were resuspended in 2 mL of 20 mM ice-cold 20 mM $CaCl_2$. 100 µl aliquots were made and flash frozen in liquid nitrogen, and then stored at -80°C.

c) Cloning

<u>Into pGEM-T-easy</u>:
Cloning was performed according to the manufacturer's instructions. Briefly, a 1:3 vector/insert molar ratio was usually used. The desired quantity of insert (DNA) was added to 1 µ of pGEM, 1 µ of T4 DNA ligase, 5 µ 2X buffer and H_2O qsp 10 µl.
Reaction was left either for 1 hour at RT or overnight at 4°C. And then transformed in *E.coli* .

<u>Into pDONOR207</u>:
pDONOR207 was used as the entry plasmid for Gateway cloning. Gateway® is a branded cloning technology based on the site-specific recombination properties of the lambda bacteriophage. First, attB sites have to be added. attB1 site is added in 5' of the DNA of interest and attb2 site in 3'. Then, using BP clonase II enzyme (Thermo), the attB-PCR product (or a vector containing the attB-PCR product) will recombine with pDONOR207.

The resulting product will contain the DNA sequence of interested flanked by attL sites. Finally, this entry clone with the attL sites will recombine, using LR clonase II (Thermo) with any destination vector containing the attR sites.

d) Bacterial transformation

E.coli – Heat shock transformation

Competent cells were thawed on ice for 10 minutes. Then ligation product was added (≤5 µl) to 50 µl competent cells (*E. coli*). And reaction was incubated on ice for 30 min before heat shock.

Heat shock was performed at 42°C for 30 to 45 seconds and then reaction was placed on ice again for 2 min. After this step, 800 µl to 1 mL of LB was added and bacteria were incubated at 37°C for at least 1 hour.

If needed, Bacteria were collected by centrifugation 10 minutes at 4 000 g (RT) and resuspended in 200 µl of LB. 100 µl was plated on selective media containing the appropriate antibiotic, to select the transformed colonies and incubated O/N at 37°C.

E.coli – Electro transformation

For electro transformation, *E. Coli* M01022 were used. Ligation products were purified by phenol-chloroform and resuspended in a small volume (5 µl) of water. 2 µl of purified ligation products were added to 40 µl of electro-competent bacteria. The cell/DNA mixture was transferred to a chilled electroporation cuvette and electroporation was performed using the following conditions: 2.1 kV, 100 Ω, and 25 µF. The typical time constant is ~2.6 milliseconds. Directly after electroporation, 800 µl of LB (with no antibiotic) was added and the reaction transferred to a new tube and incubated with shaking at 37°C for 1 hour. If needed, Bacteria were collected by centrifugation 10 minutes at 4 000 g (RT) and resuspended in 200 µl of LB. 100 µl was plated on selective media containing the appropriate antibiotic, to select the transformed colonies and incubated O/N at 37°C.

A. tumefaciens – Heat shock transformation

Competent bacteria were thawn on ice. 300 ng to 1 µg of plasmid was added to 100 µL of competent Agrobacterium cells. After 5 minutes on ice, tubes were placed for 5 minutes in liquid nitrogen and then (after opening of the cap, to avoid popping) directly placed at 37°C for 5 minutes. Then 1 ml LB was added and reaction was incubated for 2 hours at 28°C.
200 to 250 µl of the reaction was plated on selective media and colonies were grown at 28°C for 2 days.

XVII. PROTEIN

a) *Protein extraction:*

Three different methods were used for protein extraction. These three protocols can be used to extract proteins from any plant samples (leaves, seedlings, flower buds).

- Method from Evrard et al, 2002

The plant sample was added to a tube containing 100 µl of sterile glass beads and flash frozen. Then the samples were grinded using the Silamat® device. 200 µl of hot (80°C) protein extraction buffer (100 mM Tris-HCL pH 6,8, 1% glycerol, 2%SDS, 100 mM DTT, 4 M urea) was added and samples were vortexed. After centrifugation 5 minutes at 16 000 g (at RT), the supernatant was transferred to a new tube containing 3 volume of cold acetone. Samples were then placed at -20°C for 20 minutes. After 10 minutes of centrifugation at 10 000 g (4°C) the pellet was washed with 80% acetone. Then, after being air dried, the pellet was resuspended in the adequate buffer, depending on further use (SDS-PAGE or MS analysis for example).

- Trizol® / Tri-reagent (MRC)

The beginning of this protocol is identical to the one used for RNA extraction (see above). After the first centrifugation that allows phase separation, the organic phase (at the bottom of the tube) was retrieved. Then DNA was precipitated by adding 100% ethanol (volume / volume) and mixing. After 15 minutes of incubation at RT, tubes were centrifugated 5 minutes at 16 000 g (4°C). The supernatant containing the proteins was transferred to a new tube and proteins were precipitated with acetone as previously described .

- Method adapted from Tsugama et al. (2011)

2 pieces of Arabidopsis leaves (collected using a puncher) were dissolved in 150 µl of protein extraction buffer (10% v/v glycerol, 0.12 M Tris HCL pH 6,8, 4% w/v SDS, 0,005% w/v bromophenol blue, 1,25% v/v β-mercaptoethanol) and briefly grinded with a plastic pestle. After boiling for 10 min at 95°C, tubes were spinned for 5 minute at maximum speed and 100 µl of supernatant was collected and used as protein extract.

Overall, this method proved to be both the most efficient and the fastest to extract *Arabidopsis* total proteins.

b) Protein quantification:

Nanodrop®:

Nanodrop was used for protein quantification at 280 nm using a NanoDrop 2000c™ spectrophotometer (ThermoFisher).

Bradford assay:

Proteins were quantified using the colorimetric Bradford assay. The protein assay solution was purchased from Bio-Rad®. 1 to 20 µl of samples (eventually diluted) was

used. And mixed to 200 µl of the protein assay solution and water qsp 800µl. After incubation for a few minutes at RT, OD_{595} was measured and compared to an established standard range. Using this standard, Protein concentration was determined.
It is worth noting that protein concentration determined using this method is not absolute and Bradford assay reactions tend to be difficult to use when different chemicals, frequently occurring in biological experiments, like SDS are present in the solution.

c) Protein expression:

a) Induction test

This test is performed after cloning and transformation to confirm the correct expression of the protein of interest.
1 colony was used to inoculate 3 mL of LB containing the appropriate antibiotic and 1% glucose (to repress protein expression by inhibiting the T7 RNA polymerase activity) and cultivated at 37°C. This culture was divided in two sub-cultures when OD_{600} had reached 0,7. After centrifugation, Bacteria were resuspended in 3 mL LB complemented with the appropriate antibiotic. In the negative control, sucrose was added to repress protein expression. In the induced culture, 1 mM IPTG was added. After incubation for 18 h at low temperature (usually 17°C), bacteria were centrifuged (10 minutes, 5 000 g). Then, bacteria were resuspended in protein loading buffer (20 µl). This step allows bacterial lysis. Total proteins were then run on an SDS-PAGE and the expression of the desired protein can by visualized either on a stained gel or detected by Western-Blot analysis.

b) Protein purification

PRORP2 coding sequence was cloned into pET28-b(+) (Novagen) to obtain C terminal fusions with histidine affinity tags. The catalytically inactive mutant has NYN active sites Aspartates 421 and 422 mutated to Alanines. Proteins were expressed over night at 18°C in BL21(DE3) E. Coli cells induced with 1 mM IPTG. Bacteria were lysed and centrifuged

30 min at 30 000 rpm (4°C). The cleared bacterial lysates were incubated with the Ni NTA resin (Qiagen). The bound proteins were washed with buffers containing 50 mM imidazole, 20 mM MOPS pH 7,8, 150 mM NaCl and 10 % v/v glycerol and then 75 mM imidazole, 20 mM MOPS pH 7,8, 150 mM NaCl and 10% v/v glycerol. Proteins were eluted from the column using 200 mM imidazole and 500 mM imidazole buffers. Proteins were further purified by size exclusion chromatography, on a Superdex 200 10/300 Increase column, using an automated Äkta pure system in buffer containing 20 mM MOPS pH 7.8, 150 mM NaCl and 10% v/v glycerol.

d) SDS-PAGE (Sodium Docecyl sulfate-polyacrylamide gel electrophoresis)

This method is used to separate proteins according to their molecular weight. Proteins were resuspended in 5X loading buffer (100 mM Tris-HCL pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 3% (v/v) β-mercaptoethanol, 0.01% (w/v) bromophenol blue), denatured at 90°C for 5 minutes and loaded on an SDS-PAGE gel. Protein migration was carried out under constant amperage of 25 mA in Laemmli buffer (25 mM Tris HCL pH 8.3, 200 mM glycine, 0,1%(w/v) SDS). Then, in order to visualize the proteins, the gel can be stained with Coomassie Brilliant Blue or proteins can be transferred to a PVDF membrane for further analysis.

Stacking gel 5%: 5% acrylamide/bisacrylamide 37.5/1, 0.125 M Tris-HCl pH 6,8, 0.1% (w/v) SDS, 0.1% (v/v) APS, 0.01% TEMED

*Separation gel (**X**%)*: **X**% acrylamide/bisacrylamide 37,5/1; 0.375 M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (v/v) APS ; 0.01% TEMED

e) Protein transfer to a PVDF membrane (Wet-method)

This method was performed to electro-transfer proteins from the SDS gel on a PVDF membrane (Immobilon-P, 0,45 µm, Millipore) in order for the proteins to be accessible for antibody detection.

The PVDF membrane was activated using ethanol then equilibrated with the gel in transfer buffer (15% ethanol, 20 mM Tris, 200 mM Glycine) for 15 minutes. Protein transfer was carried out under a 360 mA constant amperage at 4°C for 1 hour in a Mini-Trans-Blot Cell (Bio-Rad®) apparatus. The membrane was then either stained in membrane staining solution (0.1% Coomassie Blue R-250, 50% methanol, 7% acetate) to visualize the correct protein transfer and then washed with the Membrane wash solution (50% methanol, 7% acetate), or directly used for immunodetections.

f) Western Blot – Immunodetection of proteins

This technique is used to detect a protein of interest in a given protein extract.

Membrane was first blocked in a solution of TBS-Tween 0.2% (v/v) (TBS-T) with 5% BSA (or milk) (w/v) for one hour at room temperature. Blocking solution was then discarded and the membrane was incubated in a TBS-T solution and 5% BSA (or milk) containing the primary antibody directed against the protein of interest. This incubation step was preferentially performed overnight at 4°C. But it can also be performed for 1 h at room temperature. The primary antibody solution was then removed (usually it is conserved at -20°C and can be re-used) and the membrane washed 3 times 5 minutes in TBS-T. After the third wash, the membrane was incubated for 40 minutes to 1 hour with the secondary antibody coupled to peroxidase and diluted to 1/ 10 000 in TBS-T. Then the membrane is washed again 3 times in TBS-T and the peroxidase substrate (Lumi-Light Western Blotting Substrate, Roche®) was added. The peroxidase substrate emits light when in contact with the secondary antibody. Light emitted by this reaction will then either impress a photographic film or be detected by an apparatus such as the FusionFX (Vilber),

revealing the presence of the protein of interest (provided that the size of the detected band is identical to the expected size of the protein of interest) and its relative quantity.

g) Protein co-immunoprecipitation (Co-IP)

The aim of this technique is to detect the protein partners of a given protein. Immuno-precipitation assays were performed with *Arabidopsis* protein extracts using the µMACS HA-tagged Protein Isolation Kit (Miltenyi Biotec).
Three different methods were used to perform Co-IP against PRORP2-HA. All of them were efficient to obtain an enriched fraction of our protein of interest. However only the latter was efficient to identify PRORP2 protein partners.
In all cases, the starting material was *Arabidopsis* floral buds that were collected and directly flash frozen. Then, samples were either stored at -80°C or directly processed. Data analysis showed no difference between samples kept at -80°c (even for months) compared to samples directly processed.

Co-IP with 1% Triton:

1 g of flower buds was grinded in liquid nitrogen until obtaining a pale powder. Then, this powder was transferred in lysis buffer (50 mM Tris HCl pH 8, 50 mM NaCl, 1% Triton and proteases inhibitors (Complete EDTA-free Roche®) and rotated on a wheel for 30 minutes at 4°C. After 10 minutes centrifugation at 4°C at 10 000 g, the supernatant was transferred to a new tube and mixed with 50 µl of anti-HA magnetic beads. After 45 minutes rotation at 4°C, the solution was loaded on a magnetic column. After loading, the column was washed 4 times with 500 µl of wash buffer (50 mM Tris pH 8, 50 mM NaCl, 0,1% Triton + proteases inhibitors (Complete EDTA-free Roche®)). Then elution was performed with 120 µl of elution buffer (Miltenyi Biotec) heated at 95°C. The proteins were then analysed either by LC-MS/MS or by Western Blot.

Formaldehyde crosslink method:

1 g of flower buds was grinded in liquid nitrogen for 5 minutes. When a pale powder was obtained, it was transferred to a new mortar and grinded for 10 minutes in 10 mL of Lysis buffer + 0,3% formaldehyde (1% triton, 50 mM Hepes pH 8, NaCl 50 mM and protease inhibitors (Roche complete®)). 192 µl of 16% formaldehyde was added in the buffer directly before use. Then 330 µl of glycine 2M (in tris pH 8) was added to quench the reaction and grinding was continued for 5 minutes. Two rounds of centrifugation were carried out at 10 000 g (4°C) for 15 minutes. Each technical replicate (total n=3) contained 1,2 mL of final supernatant. 13 µl of DTT (0,1 M (made fresh)) was added and after 5 minutes rotation on wheel in cold room, 50 µl of anti-HA beads (Milteny®) were added and rotation at slow speed (4°C) was continued for 45 minutes. This solution was used as input for Co-IP. Each columns was washed 4 times with 400 µl of washing buffer (0.1% triton, 50 mM Hepes pH 8, NaCl 50 mM and protease inhibitor (Roche complete®)). Finally, elution was carried out using 125 µl of heated (at 95°C) elution buffer (Miltenyi Biotech). The proteins were then analysed either by LC-MS/MS or by Western Blot.

Co-IP Igepal buffer:

1,5 g of flower buds were grinded in liquid nitrogen for 5 minutes. When a pale powder was obtained, it was transferred to a new mortar a grinded for 15 minutes in 4,5 mL of protein extraction buffer (50 mM Tis-Hcl pH 7,5; 5 mM MgCl2, 0,1% Igepal, 10% glycerol, 5 mM DTT and protease inhibitor (Roche complete®)) at 4°C. This preparation was centrifuged twice 15 minutes at 4°C, 16 000 g. 2 mL of protein extract was used as input for Co-IP. 50 µl of milteny beads were added and rotated on wheel for 1 hour at 4°C. Each column was washed 4 times with 400 µl of the protein extraction buffer. Finally, elution was carried out using 125 µl of heated (at 95°C) elution buffer (Miltenyi Biotec). The proteins were then analysed either by LC-MS/MS or by Western Blot.

h) LC-MS/MS analysis

All the mass spectrometry analyses were performed at the Strasbourg-Esplanade proteomic platform by the platform staff.

Proteins extracts were precipitated in cold 0.1 M ammonium acetate in 100% methanol and digested with sequencing-grade trypsin. Each sample was analyzed by nanoLC-ESI-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nonLC-1000 (Thermo-Fisher Scientific) with a 160-min gradient. Data were searched against the TAIR database with a decoy strategy (release TAIRv10, 27281 forward protein sequences). Peptides and proteins were identified with Mascot algorithm (version2.5.1, Matrix Science and data were imported into Proline v1.4 software. Total number of MS/MS fragmentation spectra was used to quantify each protein (Spectral Count label-free relative quantification). In the case of Co-IP, to identify significantly enriched proteins, a statistical analysis using the msmsTests R package was carried out using spectral counts (GregoriJ, Sanchez A, 2013). The whole MS dataset was first normalized by the total number of MS/MS spectra (colum-wise normalization). The implemented negative binomial model, based on the solution provided by the edgeR package was used (Robinson et al.,2010). P-values were then adjusted using the Benjamini & Hochberg method. Proteins that were over-represented in Co-IP were visualized as a volcano plot that displays log2-fold-change and -log10-p.value on, respectively, the x and y axes. The graphic was plotted using the Plotly's R graphing library.

XVIII. Electron microscopy:

Transmission electron microscopy (TEM):

Arabidopsis leaves were fixed overnight (4°C) in 2% glutaraldehyde and treated for 2 h with 2% osmium and then in 2% uranyl acetate at 4°C ON. Samples were dehydrated with ethanol and infiltrated with EPON812 medium grade resin. 90 µm sections were cut with an ultracut E microtome (Reichert) and collected on grids coated with formvar (EMS).

Samples were visualized with a Hitachi H-600 transmission electron microscope operating at 75 kV.

Immuno-TEM:

Immuno-Electron microscopy can be used to detect proteins in tissues or cells. Ultrathin sections can be decorated with gold particles (size ranges from 5 to 20 nm) that specifically pinpoint the localization of antibody and thus a protein of interest. This way subcellular, surface, endocytic proteins /markers can be detected to a resolution of 5 nm For immuno-detection, samples were briefly hydrated in water and then saturated in PBS + 1% BSA for 30 minutes. Then, the primary antibody was added (diluted according to the manufacturer's instructions in PBS + 1% BSA) and incubation was carried for 2 hours. After 3 washes with PBS, the secondary antibody coupled on gold particles (diluted in PBS + BSA) was added for 2 hours. Samples were then washed 3 times with PBS and 3 times with water. Finally, samples were visualized with a Hitachi H-600 electron microscope operating at 75 kV.

XIX. SOFTWARE / bioinformatic

The following softwares / websites were used during my PhD.

Phyre2:
A free web-based service for protein 3D structure prediction. Phyre stands for (**P**rotein **H**omology/Analog**Y** **R**ecognition **E**ngine).
It is accessible at the following address:
http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

Pymol:

An open-source model visualization tool to produce 3D images of biological molecules. Pymol is currently distributed by Schrödinger, Inc. But a free version is available to all at: https://github.com/schrodinger/pymol-open-source

Suba4:

SUBA stands for "Subcellular localisation Database for Arabidopsis proteins". It provides a web-based subcellular data query platform for Arabidopsis proteins.
It can be accessed at:
http://suba.live/

BLAST:

Stands for basic local alignment search tool. This algorithm is a web-accessible program to compare primary biological sequences of nucleic acids or amino-acids.
It is accessible at: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Phyologeny . fr

Phylogeny.fr is a web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences..
It is accessible at : https://ngphylogeny.fr/

RESULTS

Thesis objectives

My thesis work comes as a follow up to the functional analysis of the RNase P enzymes called PRORP in Arabidopsis performed by my host laboratory. While significant progresses had been made to understand the functions, the mode of action and the evolution of PRORP enzymes, important questions remained. For instance, the integration of PRORP functions with other gene expression processes remained to be established. Likewise, the comprehensive RNA substrates spectra of PRORP enzymes was not established. My thesis work aimed at contributing to answer these questions.

Results are presented in two sections. In the first one, the question of the diversity of the nuclear PRORP2 RNA substrates is addressed. Initially, we had envisaged to look for PRORP2 novel RNA substrates without *a priori* and through a candidate approach. Because of time constraints, only the latter approach was developed. Most of the thesis work concentrated on PRORP2 cleavage at the level of a conserved tRNA-like structure present in an intron of the pre-mRNA encoding MAF1, the main repressor of RNA polymerase III.

In the second results section PRORP2 protein partners were investigated by a co-immunoprecipitation approach, revealing a complex seemingly involved in the early step of pre-tRNA maturation containing PRORP2 and TRM1A and B modifying enzymes.

Altogether, the results obtained during this doctoral thesis enabled to improve our knowledge on the biogenesis of cytosolic transfer RNAs in plants and more importantly identified a new RNA substrate for PRORP2 that pointed out a possible indirect regulation of RNA polymerase III through the control of its main repressor MAF1.

AtPRORP2 may regulate AtMAF1 level through the cleavage of its pre-mRNA at the level of a conserved tRNA-like structure

a) MAF1 intronic tRNA like structure (TLS) is conserved in Streptophyta

Searching for conserved orthologous introns in five plants, Akkuratov and colleagues (2014) discovered that the most evolutionarily conserved intronic region across these plant genomes was a putative tRNA-like ncRNA located in the sequence of an intron of the MAF1 gene. They hypothesized that this conserved intronic region was "unlikely an alternative splicing exon" as it could not be found in MAF1 mature mRNA in all plant transcripts studied, but a new ncRNA. However, they could not detect it by a northern blot approach. Since MAF1 is the best known negative regulator of RNA polymerase III (Desai et al., 2005) conserved in all Eukaryotes, we decided to investigate the function of this conserved tRNA-like structure (TLS). The TLS present in *Arabidopsis thaliana* MAF1 gene (Figure 25A, B) can be modeled in 2D and 3D using RNAfold software revealing a structure very similar to a canonical tRNA (Figure 25C, D).

Previously, PRORP was shown to catalyse the cleavage of both precursor tRNA and tRNA-like structure in mitochondria (Gobert et al., 2010; Gutmann et al., 2012) we hypothesized that this intronic nuclear TLS could be a new substrate for nuclear RNase P (PRORP2/3) in *Arabidopsis*. Moreover, as MAF1 is the main negative regulator of RNA polymerase III, we further hypothesised that this cleavage might constitute a regulation loop of RNA polymerase III activity.

First, we decided to confirm that this ncRNA sequence is indeed widely conserved in the green lineage. We found that its conservation is not restricted to Angiosperms. This region is also conserved in the most basal Streptophyta (*Klebsormidium* alga) (Figure 26). However, it does not seem to be conserved in the Chlorophyta clade. This conservation across all streptophytes lineage strengthens the hypothesis that MAF1 TLS could be functional and that this function was acquired early in the evolution of the plant lineage.

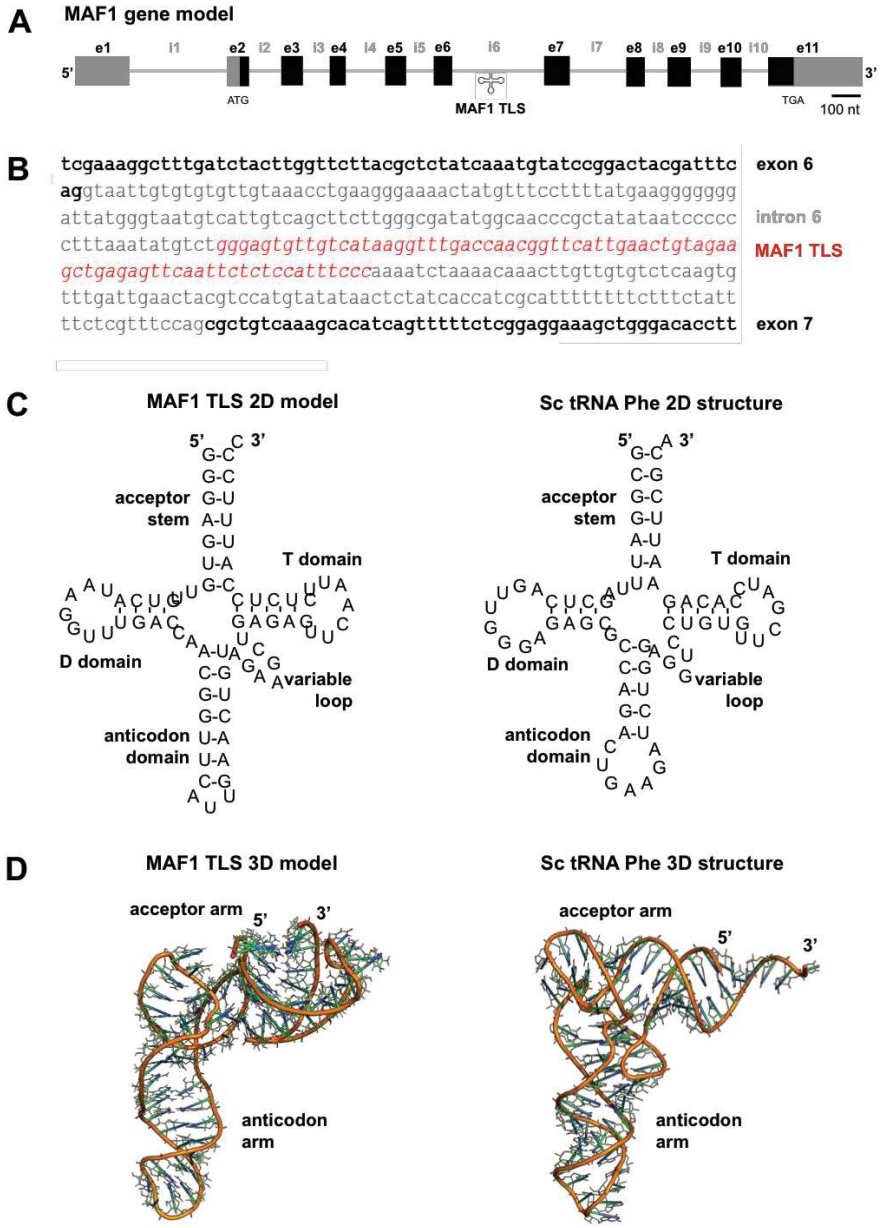


Figure 25: **Genetic context, sequence and structure of Arabidopsis MAF1 TLS**. AtMAF1 (At5g1320) gene is depicted with its 5' and 3' UTR region as grey boxes, exons as black boxes and introns are represented by lines by lines. Scale bar represents 100 nucleotides (A). The 74 nucleotides TLS sequence, located in AtMAF1 6th intron is shown in red (B). The predicted structure of AtMAF1 TLS is shown with the corresponding *bone fide* tRNA domains and loops indicated. The 2D structure of Phenylalanine tRNA from Yeast (S*. cerevisiae* from laboratory strain S288C) is shown to highlight the resemblance between the predicted 2D structure of AtMAF1 TLS and a "canonical" tRNA (C). 3D structures of the conserved intronic tRNA like Structure of Arabidopsis MAF1 (left) and of Phenylalanine tRNA from Yeast (S*. cerevisiae* from laboratory strain S288C) (right) were determined using the RNAcomposer (3D) webserver and the images were drawn with Pymol (D).

This analysis also led us to determine that in addition to the tRNA-like structure, an upstream sequence was also evolutionary conserved. This sequence is located between 13 and 20 bases upstream of the TLS start site. Using RNAfold, we found that this RNA sequence has a hairpin structure. That is to say that the RNA folds on itself in duplex with a loop. Short Hairpin RNA are often found to be involved in RNA silencing, as they can be processed into siRNAs by a Dicer enzyme (Brummelkamp et al., 2002; McManus et al., 2002; Paddison et al., 2002). However, we focused in this work only on the TLS of MAF1 precursor mRNA and not on this conserved hairpin structure. Consequently, further work is required to investigate the possible function of this RNA hairpin.

Knowing that MAF1 is the main negative regulator of RNA polymerase III which transcribes tRNA and other ncRNA (5S ribosomal RNA, U6 spliceosomal RNA and 7SL for example) and that nuclear PRORP2 and 3 in *Arabidopsis thaliana* have been shown to cleave 5' leader sequences of precursors tRNA and of a tRNA-snoRNA precursor *in vivo* (Gobert et al. 2010; Gutmann et al. 2012) we decided to investigate whether AtPRORP2/3 could cleave AtMAF1 TLS *in vitro* and *in vivo*. Also we wanted to determine if this activity would affect AtMAF1 mRNA level and by extension RNA polymerase III activity. Our overall hypothesis being that if Arabidopsis nuclear RNase P would cleave AtMAF1 TLS, this would indirectly impact RNA polymerase III activity.

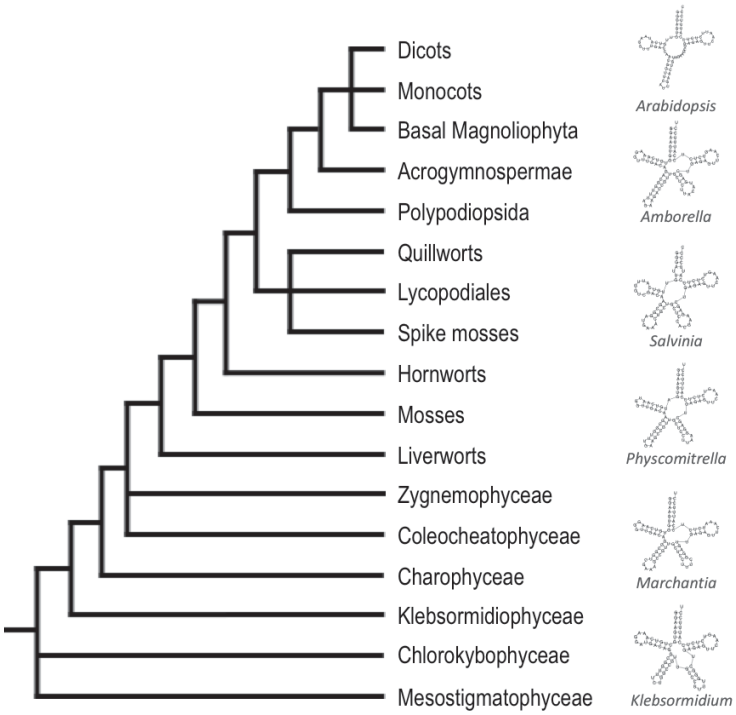


Figure 26: Representative phylogenetic tree of the streptophytes lineage showing the conservation of the TLS present in MAF1 intron in streptophytes through the 2D structure modelisation of exemplary TLS. The TLS 2D structures from *Arabidopsis thaliana* (dicots), *Amborella trichopoda* (basal magnoliophyta), *Salvinia cucullata* (ferns), *Physcomitrella patens* (mosses), *Marchantia polymorpha* (liverworts) and *Klebsormidium nitens* (klebsormidiophyceae) were modelized using RNAfold.

b) Arabidopsis PRORP2 can cleave MAF1 TLS in vitro

As a first step, we investigated whether *Arabidopsis* PRORP2 (referred hereafter as AtPRORP2) could cleave this conserved tRNA-like structure during *in vitro* cleavage assays. For this, two regions containing AtMAF1 TLS were selected. A larger region contains the TLS in its intronic context. It is 438 nucleotides long and covers AtMAF1 intron 6 and parts of exons 6 and 7. A shorter one of 154 nucleotides comprise the TLS surrounded with 51 and 28 nucleotides upstream and downstream respectively and have a size comparable to our pre-tRNA positive control. These two regions were cloned and placed under the control of a T7 RNA polymerase promoter. After *in vitro* transcription, both RNAs were used as substrates in RNase P cleavage assays using a recombinant AtPRORP2 protein. The recombinant AtPRORP2 was produced, following the protocol detailed in the Material and methods section. Briefly, the CDS for AtPRORP2 was cloned into pet28b and expressed from Rosetta DE3 cells. A representative gel of the different fractions obtained at each step can be seen in Figure 3. AtPRORP2 could be purified to near homogeneity as verified on gel and by mass spectrometry analysis. The sample from AtPRORP2 elution 2 was used for activity assays.

As a negative control for activity assays, a catalytic mutant of AtPRORP2 was used. This catalytic mutant has two aspartate amino acids (at position 421 and 422) that have been mutated to alanine. These two aspartates have been shown by Gutmann et al. 2012 to be essential for PRORP2 activity. Consequently, the catalytic mutant is an important control as it shows that the cleavage observed in the experiment is not due to a bacterial RNase contamination that would have been co-purified with PRORP during the expression of the recombinant protein. Then, as a positive control, cleavages were also performed with a canonical pre-tRNA substrate.

Cleavage assays of PRORP2 with both long and short MAF1 transcripts resulted in two distinct bands of apparent sizes of 238 and 200 as well as 103 and 51 nucleotides respectively. Theses sizes are in accordance with canonical RNase P cleavages as observed with the control pre-tRNA. Likewise, no RNA cleavage was observed with PRORP2 catalytic mutant. Altogether, results clearly showed that PRORP2 is able to cleave this TLS in the context of MAF1 intron 6 *in vitro* (Figure 4A).

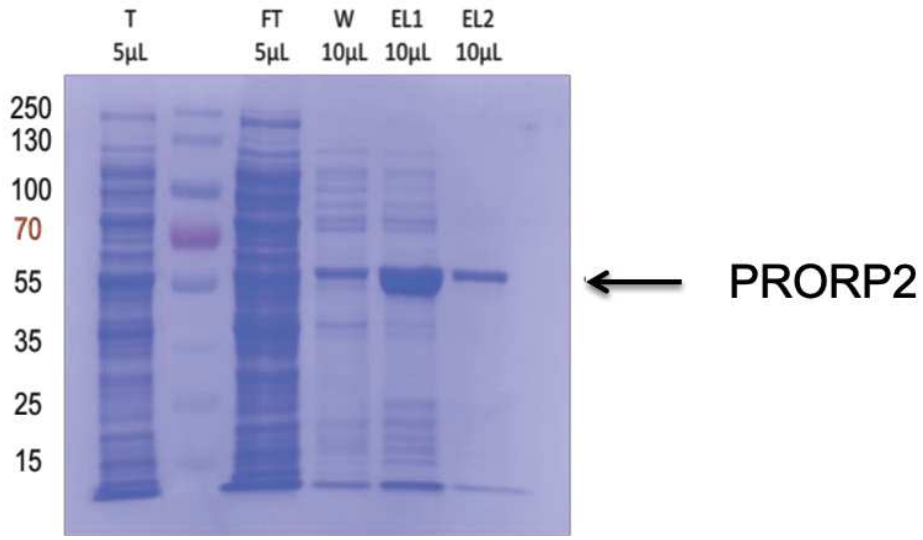


Figure 27: Purification of recombinant PRORP2. Photo of a PVDF, Coomassie stained membrane after transfer of the protein fractions ran on an SDS-PAGE.Fractions correspond to: T = total fraction (diluted bacterial lysate), FT = flow through, W = wash , EL = elution from the IMAC purification. The molecular masses of the protein marker are indicated on the left in kDa. The recombinant PRORP2 protein (59,2 kDa) is indicated by a blue arrow.

In vitro RNase P cleavage assays were usually routinely performed with the reaction occurring for 15 minutes. Still, as we were optimising the RNase P cleavage protocol, reactions were performed with different time points (from 10 to 60 minutes incubation). In these assays no differences could be seen depending on the time the reaction occurred. Similar to precursor tRNAs, the reaction is very rapid and over 10 minutes incubation, no modification of the RNA cleavage pattern could be observed. The remaining uncleaved substrate forming the higher molecular weight bands on the gels very likely correspond to *in vitro* transcripts that are not folded as a tRNAs and as such are not processed. Indeed, after longer incubations, an identical proportion of non-cleaved RNA substrate could still be observed.

Positive controls experiments were performed with a precursor tRNA corresponding to Arabidopsis mitochondrial precursor tRNA Cysteine with a 5' leader sequence of 50 nucleotides and a 3' trailer sequence of 25 nucleotides. Its cleavage results in fragments of 99 and 50 nucleotides, corresponding to a canonical RNase P reaction (Gobert et al., 2010).

c) Mapping of the in vitro cleavage of MAF1 TLS by PRORP2

As a next step, a circular RT-PCR (C-RT-PCR) analysis was performed to precisely map the RNA cleavage sites of the *in vitro* experiments (Figure 28B). For this, the upper band of the cleavage reaction performed with the longest transcripts (exact size = 438 nucleotides) was extracted from gel and retrieved using a protocol from Nilsen TW (cold spring harbour protocol). This RNA was circularized using T4 RNA ligase. A reverse transcription was then performed using a primer designed to bind this ligated RNA. Then, nested PCR was performed. The Forward and Reverse primers being also designed to amplify the cDNA spanning the RNA ligation site. Finally, the resulting PCR products were inserted into pGEM-T-easy vector and then sequenced with Sanger method.

The analysis of 23 revealed cleavage sites corresponding to the 3'end of the *in vitro* transcripts concatenated to the TLS 5' start site. This showed that PRORP2 cleavage had precisely occurred at the -1 site of the TLS start site. It is worth noting that either

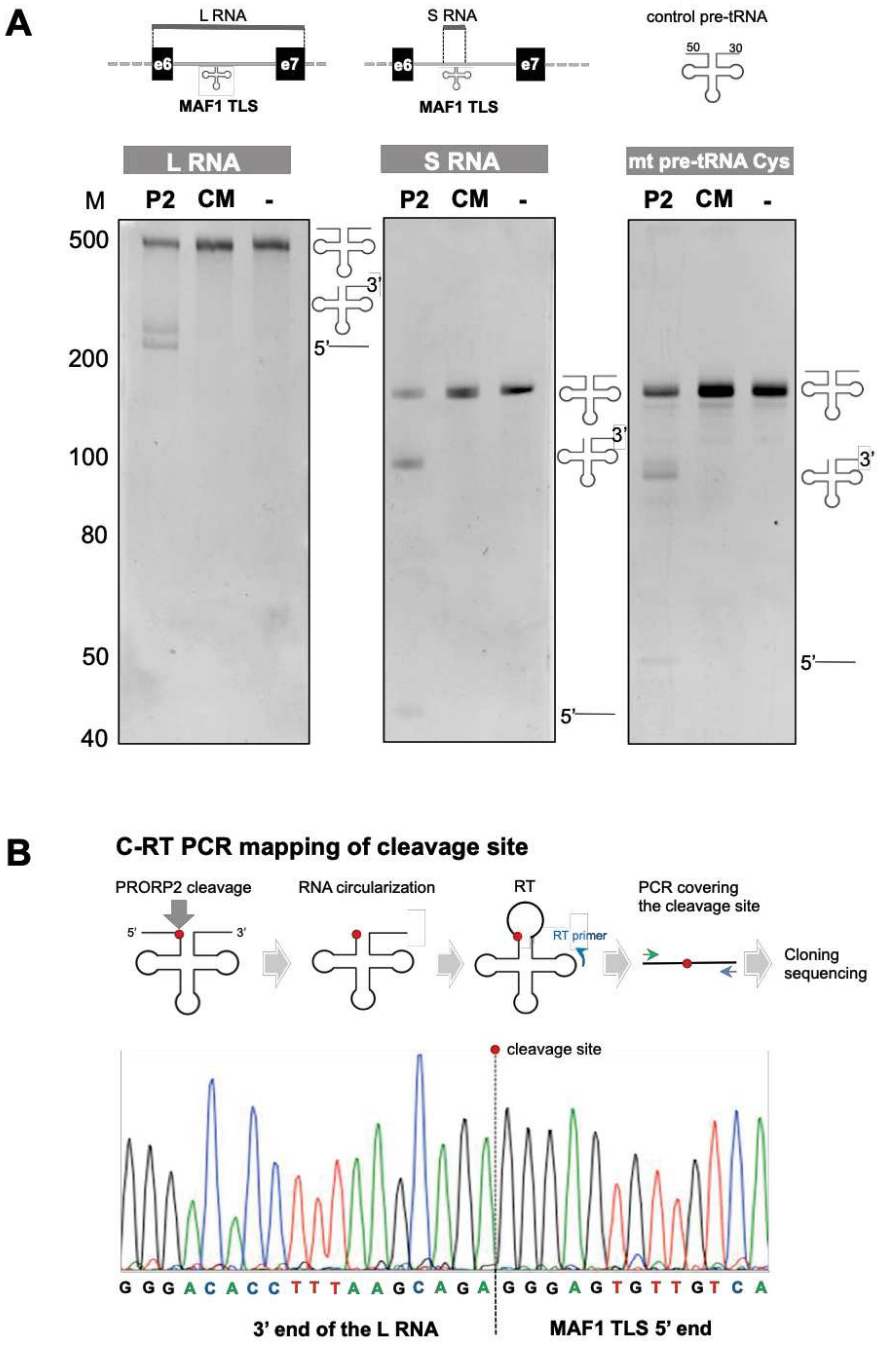


Figure 28: ***In vitro* cleavage assays of MAF1 tRNA like structure by PRORP2**. A long RNA (L RNA), corresponding to a region containing part of exon6, intron6 and part of exon7 of *Arabidopsis* MAF1 and a short region (S RNA) corresponding to a portion around the TLS of intron6 of MAF1 were transcribed *in vitro* and incubated with recombinant PRORP2 (P2). As negative controls, the reaction was performed without protein (-) or with a catalytic mutant of PRORP2 (CM). As a positive control, a precursor tRNA (the mitochondrial precursor tRNA Cysteine, mt pre-tRNA Cys) was used, showing that the recombinant PRORP2 is catalytically active. The reaction products were run on a 10% denaturing acrylamide gel and visualized under UV after ethidium bromide staining. Schematic drawing of 2D tRNA indicate to which part of the transcribed RNAs the observed bands correspond. Molecular weight markers are indicated in nucleotides (A). Schematic drawings of the different steps performed in the C-RT PCR assay are depicted. The red dot indicates the cleavage site determined by C-RT-PCR. A representative result of the consensus sequence obtained after sequencing of the PCR products is shown, highlighting that the 3' end of the transcribed RNA is ligated to the exact start site of the TLS (B).

adenine (A) or cytosine (C) were found to be added between the two extremities of the cleaved transcripts (in 10 and 13 clones respectively). These nucleotides do not come from the initial T7 transcript and were most probably inserted during the RNA ligation step of the C-RT-PCR analysis.

In conclusion, our sequencing results showed that PRORP2 cleavage was, as predicted, taking place at a position immediately upstream of the TLS (Figure 28B), confirming that *Arabidopsis thaliana* MAF1 precursor mRNA TLS is, *in vitro*, a substrate of RNase P.

d) Down-regulation of Arabidopsis PRORP2 and 3 leads to an up-regulation of AtMAF1 at the transcriptional level

RNase P is an essential function and the double mutant *prorp2*/*prorp3* was shown to be lethal (Gutmann et al., 2012). Thus in order to investigate the effect(s) of the putative cleavage of MAF1 TLS *in vivo*, we down-regulated AtPRORP2 by virus induced gene silencing (VIGS) in a *prorp3* knock-out background (as previously performed in Gutmann et al., 2012). This technology is based on the infection of plants with viral vectors expressing the tobacco rattle virus (TRV) RNA as well as a sequence of interest that we wish to down-regulate (Burch-Smith et al., 2006). The down-regulation mutants were then assayed at the transcriptional level to determine the relative expression levels of MAF1 and other tRNA related genes by quantitative RT-PCR (RT-qPCR). The efficiency of the virus infection was monitored by a positive control (+PDS), downregulating the level of the Phytoene desaturase gene. In brief, this control works as follows. Phytoene desaturase is a gene involved in carotenoid biosynthesis. Carotenoids are pigments attached to the membrane within the cell. Pigments are chemical compounds reflecting only certain wavelengths of visible light. In plants (as well as algae and cyanobacteria) they are the mean by which the energy of sunlight is captured for photosynthesis. Carotenoids are called accessory pigments as during photosynthesis they pass their absorbed energy to chlorophyll. In addition to their role as photosynthetic pigments, carotenoids are also photoprotectors. Chloroplasts cannot function properly without carotenoids (Sun et al., 2018). As such, the downregulation of PDS leads to a whitish leaves

phenotype in infected plants probing that the infection was effective and enabling to monitor in time the progression of virus infection in leaves (figure 30).

As a negative control, infection with a viral empty vector (EV) was performed. These plants do not exhibit a strong specific phenotype (it can simply be seen that they had a viral infection) when the samples are harvested (3 weeks after infection). At the molecular level, all experiments performed were done using this negative control as a reference.

The downregulation of AtPRORP2 (in *prorp3* background) was monitored by RT-qPCR in three biologically independent replicate experiments. Our result show that AtPRORP2 expression was only 40% of AtPRORP2 expression level, in experiment plant compared to control plants (*prorp3* ko infected with empty vector). Interestingly this downregulation of PRORP2/3 did not seem to affect, at the transcriptional level, the expression levels of other tRNA related genes that we assayed such as, LA proteins (AtLA1 (AT4G32720) and AtLa2 (AT1G79880)), *Arabidospsis* tRNA nucleotidyltransferase (At1g22660) also called "CCase", and RNA polymerase III subunits NRPC 1 (AT5G60040), NRPC 2 (AT5G45140) and NRPC 7 (AT1G06790)). In contrast, AtMAF1 appears to be upregulated. Indeed, its expression at the transcriptomic level is 3 times higher in the tested plants compared to the control (Figure 29). This led us to propose a model where, the down-regulation of Arabidopsis nuclear RNase P results in an impaired cleavage of AtMAF1 TLS (i.e. to a decreased degradation of AtMAF1 mRNA precursor), leading to an over accumulation of MAF1 mature mRNA, and consequently, the increase of its transcriptomic steady state level.

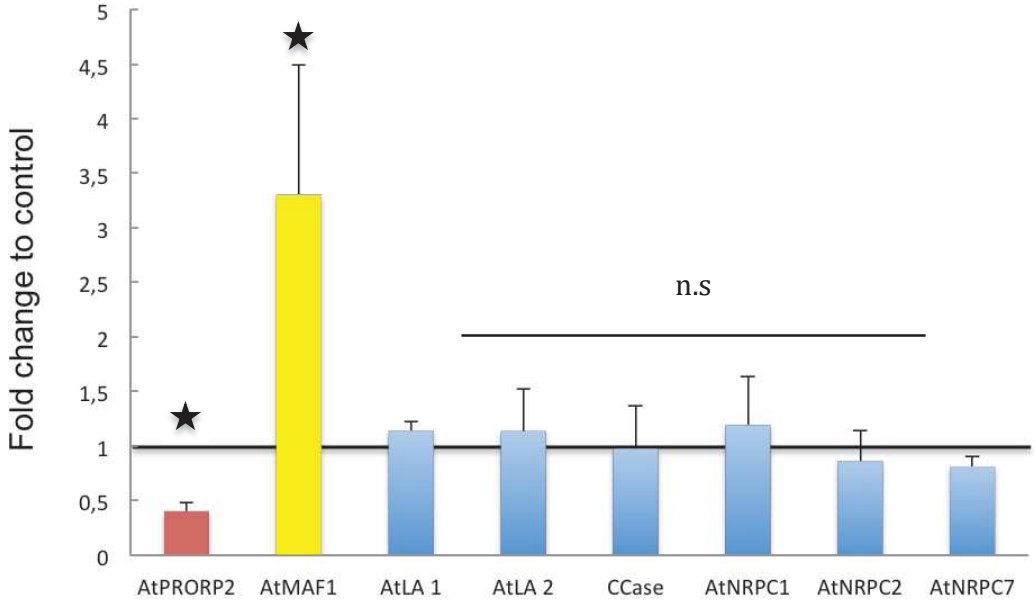


Figure 29: **Relative transcript fold changes in expression levels of AtMAF1 and other RNA polymerase III related genes in plants with VIGS against PRORP2** in *prorp3* background compared to plants with VIGS by an empty vector assayed by RT-qPCR. Standard deviation of 3 independent experiments are indicated by positive error bars. Relative fold change expression levels compared to control, indicated on the Y axis, were calculated using the delta-delta Ct method. AtLA1 and 2 are the two LA proteins found in Arabidopsis. CCase is *Arabidospsis* tRNA CCA nucleotidyltransferase, NRPC stands for Nuclear RNA Polymerase III, the number indicating the sub-units (here 1, 2 and 7). Difference in level of expression between the two conditions (VIGS against PRORP2 compared to VIGS against empty vector) was assayed using student's t-test and considered statistically significant if $p < 0,05$ (indicated by a star). Otherwise, difference was considered not statistically significant (n.s).

Beyond the molecular phenotypes of plants down-regulated by VIGS for AtPROP2 in a *prorp3* knock-out background, we also monitored macroscopic phenotypes and the effects of RNase P down-regulation at the cellular level. PRORP2 down-regulated plants showed clear growth phenotypes with variegated leaves in comparison to control plants (Figure 30 and 31).

In order to monitor cellular phenotypes, we sampled leaves at 3 weeks post-infection (i.e. samples were collected at the same time for RT-qPCR analyses). After resin embedding, samples were analysed by transmission electron microscopy. Results (Figure 31) show that down-regulation of PRORP2/3 clearly affects the structural organization of the cell in the most affected parts of the leaves at this stage: the midrib of petiole-less leaves (see picture in Figure 31). The older the leaf, the stronger the phenotype: whitening of the main vein. White vein cells (top photo, Figure 31) show, in plastids, unstructured thylakoids and large size plastoglobuli accumulation. The cytoplasm of the cell is full of small vesicles. These observations are characteristic of senescing cells. By contrast, younger petiole-less leaves that are not displaying such a strong whitening (middle photo, Figure 31) have a cellular content that appears similar to the control (bottom photo, Figure 31). Consequently, to avoid any bias in our analysis, care was taken when sampling leaves, to collect only the ones displaying a mild macroscopic phenotype in the plants down-regulated for AtPRORP2/3.

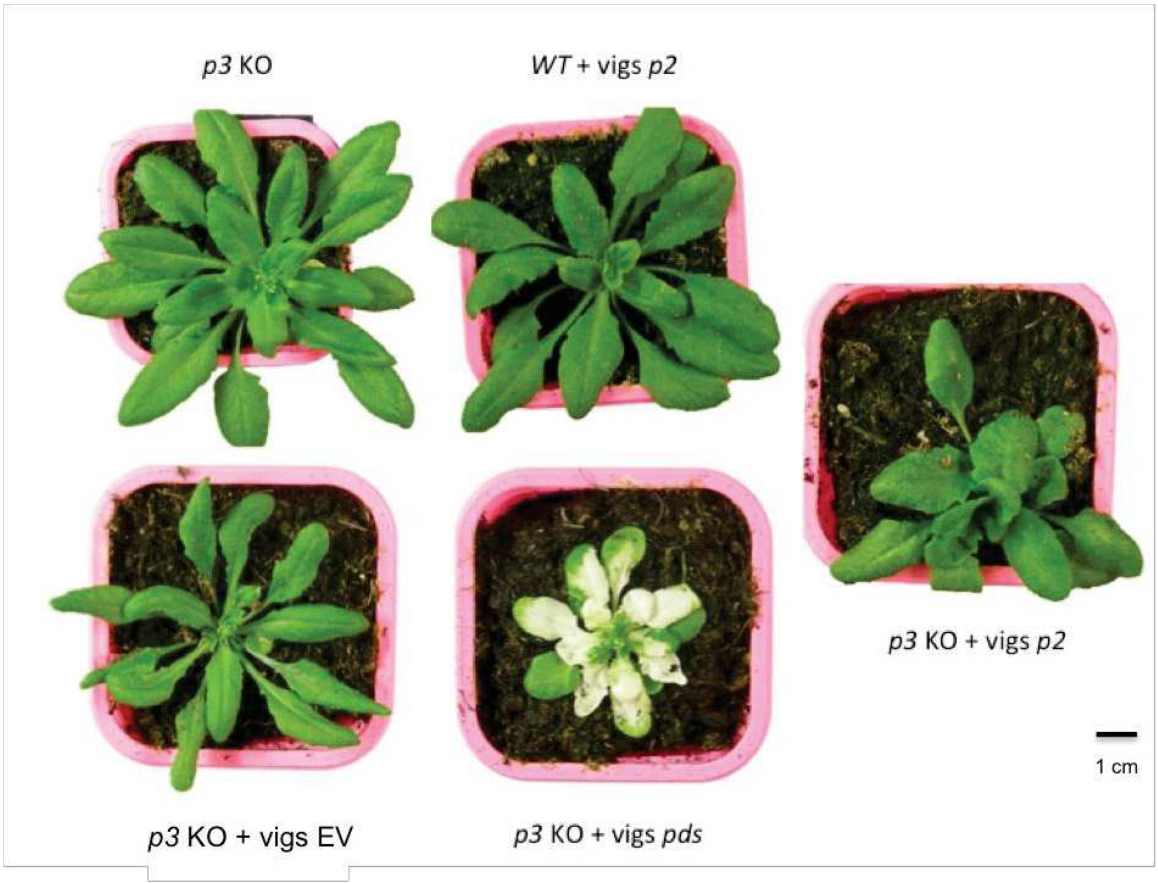


Figure 30: **Representative images of plants after virus induced gene silencing (VIGS) against PRORP2** in WT (Col-0) and *prorp3* knock-out (KO) background. Images were taken 18 days after VIGS infection. Plants were consequently 28 days after germination (Stage 5 from Boyles and al.). All plants were sown at the same time and grown in the same conditions. WT = wild type, p2 = PRORP2, p3 = prorp3, EV = empty vector, pds = phytoene desaturase.

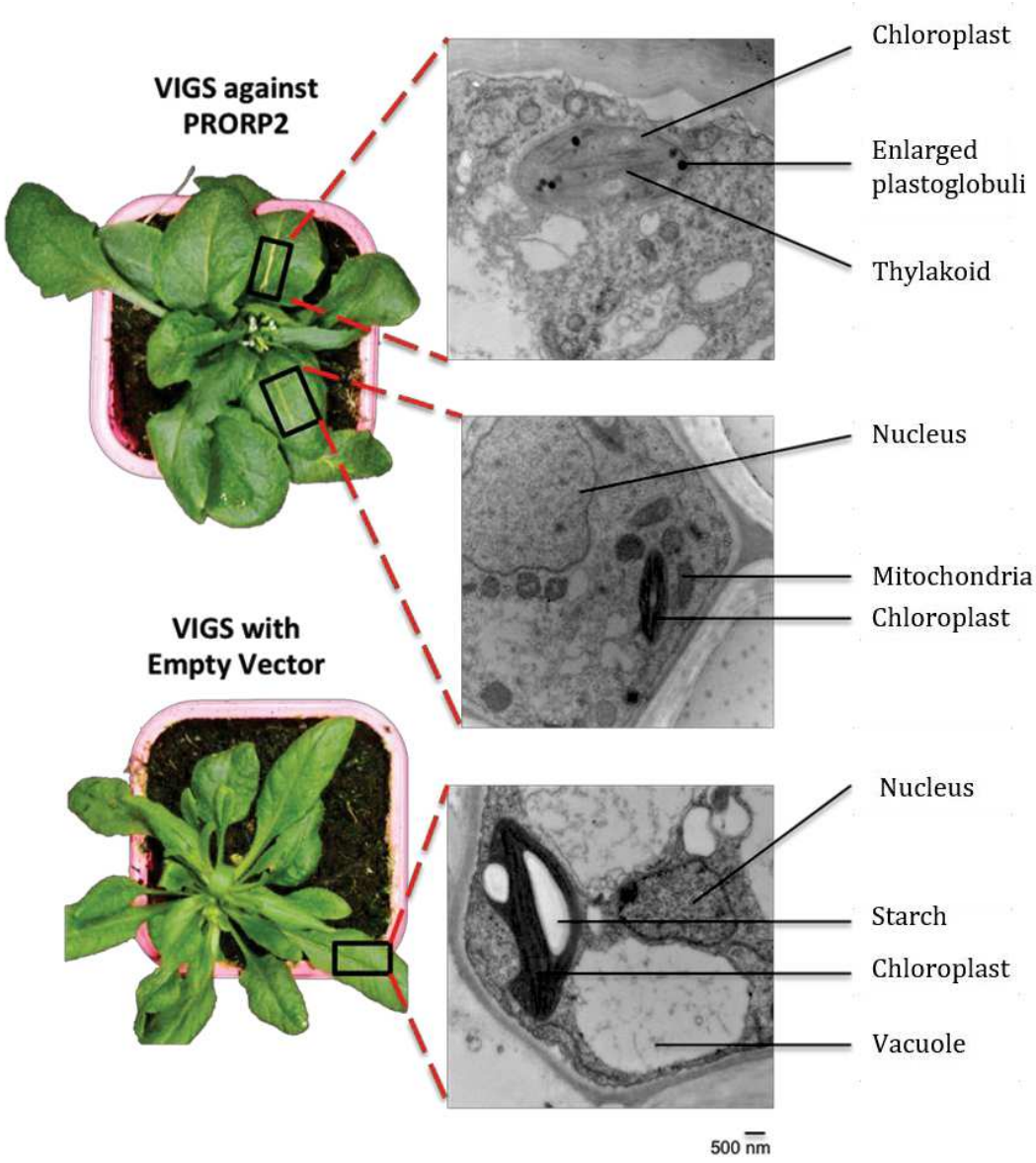


Figure 31: **Representative images of plants 3 weeks after VIGS infection and its effect at the cellular level**. Representative photographs of plants 3 weeks after VIGS infection against PRORP2 (above) or with an empty vector (below) are shown (left). Black boxes on leaves indicates where a 0.5 mm band was dissected in the leaf midrib and further processed to be observed by transmission electronic microscope (TEM). Representative images after visualisation on a Hitachi H 7500 transmission electron microscope are shown for VIGS against PRORP2 (upper and middle) or for the empty vector (lower) on the right. Scale bar is identical for the three TEM images. Cellular compartments are indicated using black arrows.

This result nicely complements the one previously obtained in the research group. Indeed, Gutmann et al. (2012), had shown severe cellular defects in plants infected with a pTRV2 plasmid containing a sequence designed to silence AtPRORP1 (the organellar RNase P in *Arabidopsis*). However, no clear cellular phenotype could be observed in plants with a VIGS against AtPRORP2. According to our result, this is very likely due to a difference in the part of the plant sampled to perform TEM observations.

Indeed, when sampling the most affected part of the leaf, a clear phenotype can be seen at the cellular level, highlighting the importance of nuclear RNase P. However, in parts of the leaves that do not exhibit a severe macroscopic phenotype, the cellular defects are only mild (Figure 31), indicating that in these leaves, the level of nuclear RNase P, albeit decreased is still sufficient for cells to function properly.

Altogether, the RT-qPCR data and the cellular observation of plants down-regulated for nuclear RNase P suggest that AtMAF1 is indeed over-expressed at the transcriptomic level when nuclear RNase P is down-regulated. However, down-regulation of *Arabidopsis* nuclear RNase P leads to an important stress at the cellular level and for the entire organism. Thus, since MAF1 is known to be a stress regulated protein, at this stage, we could not rule out that the increase in AtMAF1 mRNA level is solely due to a lower amount of AtPRORP2/3, but might also be a response to the stress caused by this depletion. Consequently, in order to test the direct effect of AtPRORP2/3 activity on AtMAF1 TLS, we used 3'RACE to test whether AtMAF1 TLS is indeed cleaved *in vivo* at its start site.

e) 3' RACE reveals accumulation of MAF1 transcripts termini at the 5' extremity of the TLS

Before selecting 3' RACE-Seq to analyse AtMAF1 transcript ends, we initially performed 5' RACE. In a pilot assay we had used Sanger sequencing of multiple bacterial colonies containing an insert with the 5' RACE final PCR products. We could not conclude from these first results, as we did not know if the more than one hundred sequences analysed corresponded or not to PCR duplicates. In addition Sanger sequencing did not yield enough depth to correctly assess AtMAF1 mRNA transcript ends. The main problem lies

in that the TLS is found in an intron of MAF1 and in consequence only on pre-mRNA sequences that are not abundant compare to the mature mRNA.

Thus in order to determine if AtMAF1 TLS is processed *in vivo* and might thus regulate AtMAF1 transcriptional level, we decided to analyse the 3' ends of the AtMAF1 transcripts using 3' rapid amplification of complementary ends followed by NGS sequencing (3'RACE-Seq). Total RNA was extracted from WT plants, from plants knocked out for *prorp3* (AT4G21900), and after infection either by a VIGS construct to down-regulate AtPRORP2 (AT2G16650) or with an empty vector (EV). Finally, a *hen2-2* mutant (T-DNA line GK_774H07 (N474299) characterised by Lange et al., 2011) was also used. The activity of the nuclear exosome that mediates 3'-5' RNA degradation activity is reduced in this line. We expect that degradation byproducts of PRORP cleavage of MAF1 TLS are decreased in this line. In consequence, it should improve detection of the true 3' end of the cleavage product of PRORP. In the 3' RACE assays we used a 3' adaptor containing a random sequence of 15 nucleotides allowing to remove PCR duplicates from the analysis after NGS sequencing of the final PCR products.

Results from the 3' RACE-Seq are showed as a close-up for the intron 6 (from position 11, forward primer location, to 316, the end of the intron 6) but percentages are calculated from the total of number of reads found along the 306 nucleotides long amplified sequence analysed by 3'RACE-Seq (Figure 32A, C, D). The TLS is localised at position 135-209 in our snap-shot (blue background in Figure 8). We can visualize peaks from the positions -1 to -3 from the start of the TLS (position 134, 133, 132) in the WT, *prorp3* (p3 KO), *prorp3*-vigsEV (p3 KO + vigs EV) and *hen*2 backgrounds (Figure 32). The close-up view around the TLS 5' extremity (Figure 32B, C, D), shows that the sums of these peaks represent around 40% of total reads in WT (Figure 32A, B), 42% in p3 KO, 68% in p3 KO + vigs EV (Figure 32C) and 71% in *hen*2 backgrounds (Figure 32D). Hence, it is clear that an endoribonucleolytic cleavage occurs around the stated positions. However, the 3'end by-products resulting from the cleavage are not at a precise position varying between -3 to -1 in our results. The 3' extremity in *hen*2 background where exosome activity on the 3' end should be abolished is also not clear between these positions while the percentage of reads is the highest compare to other backgrounds. The blurred 3' extremity could be the result of the nibbling by exoribonucleases after the endoribonucleolytic cleavage or an imprecise cleavage by the endoribonuclease

responsible of this cut. I would like to stress that PRORP enzymes can have imprecise cleavage leading for example to the extra nucleotide found at position -1 in tRNA His (Placido et al. 2010). The downregulation of PRORP2 in *prorp3* (p3 KO + vigs P2 in Figure 32A, B) seems to result in a reduction of the percentage of transcript ends at position 132, 133 and 134. However, the third replicate behave not like the two first one and extra replicates are needed to confirm the result. In any case, the three peaks represent only 25 % of total reads in the two first replicates but 50 % in the last replicate. If we take into account only the two first replicates, it is a 37% decrease compare to the WT, 52% compare to p3 KO and 63% compare to p3 KO + vigs EV. This result seems to support our hypothesis that AtPRORP2/3 are indeed responsible for the cleavage of AtMAF1 TLS *in vivo*. Further replicates are needed for all the genotypes to strengthen the result.

The analysis of 3' ends by-products all along the studied sequence (intron 6 – exon 7) shows that poly-A tail (A-rich + onlyA) is the major addition (in red, Figure 32 and Figure 33A). When the exosome activity is reduced (*hen*2 background), the percentage of poly-A increases above 70% because these poly-A tailed by-products are accumulating (Figure 33A). These are well known substrates of the exosome (Lange et al., 2009). The focus on modifications at positions 132, 133, 134 before the TLS (Figure 33B) clearly shows poly-A tail as the main modification. The cleavage by-products are therefore the target of the exosome. Since the addition of poly-A tails was described as a signal for RNA degradation (Lange et al., 2009), this strengthens the idea that these transcripts are degraded by the RNA decay machinery after endoribonucleolytic cleavage. Overall, the results suggest that nuclear PRORPs are responsible for the endoribonucleolytic cleavage of MAF1 pre-mRNA at the 5' end of a TLS in intron 6. Upon cleavage, the by-product located upstream the cleavage site is degraded by the nuclear exosome associated with the helicase HEN2. We could note that the cleavage occurs in the WT in standard growth conditions.

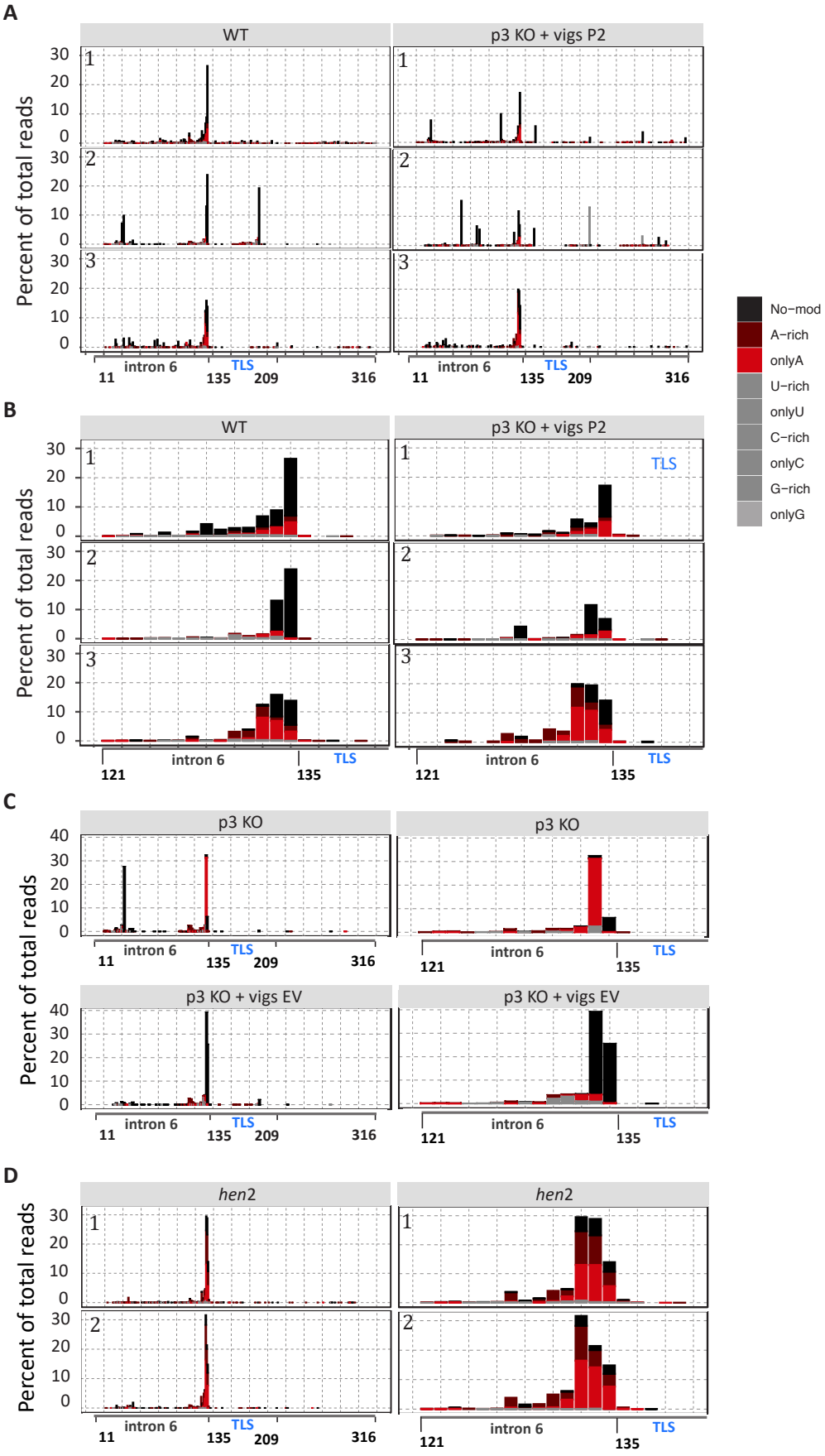

A
WT
p3 KO + vigs P2
Percent of total reads
intron 6
TLS
11
135
209
316
No−mod
A−rich
onlyA
U−rich
onlyU
C−rich
onlyC
G−rich
onlyG
B
121
C
p3 KO
p3 KO + vigs EV
D
hen2

Figure 32: **Analysis of the 3'RACE-seq results showing cleavage by-products 3' ends along the intron 6** (A, and left panels of C and D) and for a close-up around the 5' end of the TLS (B, right panels of C and D) for the different genotypes. Results are expressed as percentage of total reads for the whole sequence analysed in this experiment. The colours express the reads unmodified (black), reads with poly-A untemplated added nucleotides (A-rich or only A in red colours) and reads with other types of untemplated added nucleotides (in grey). Three replicates of the experiment for WT and p3 KO + vigs P2 are presented in A and B, one experiment was done for p3 KO and p3 KO + vigs EV (C) and two replicates were done for *hen*2 (D).

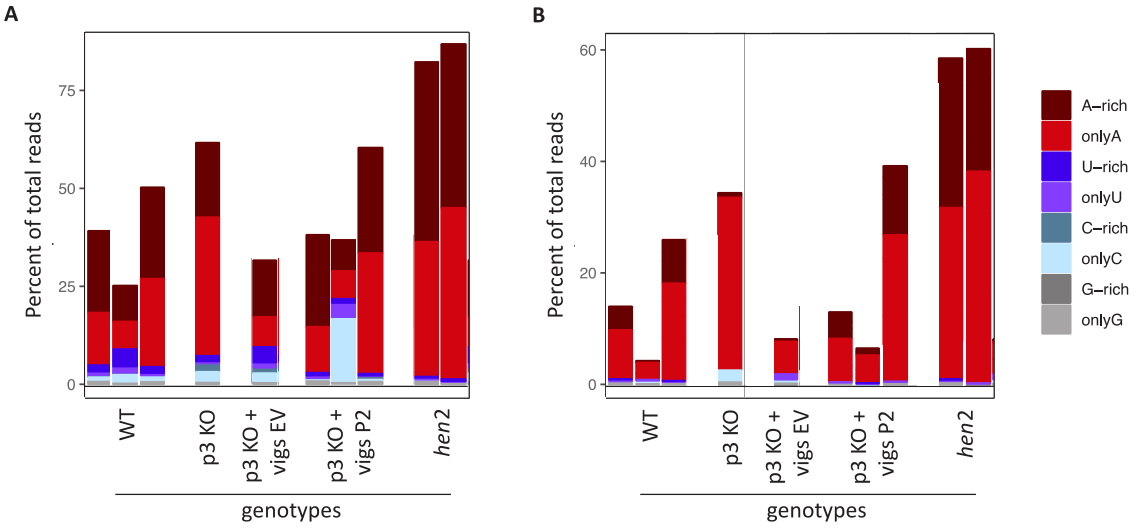


Figure 33: **Addition of extra-nucleotides at the end of cleavage by-products** along the analysed sequence (A) and at position 132,133, 134 corresponding to -1 to -3 upstream the 5'end of the TLS (B) for the different genotypes tested. The colour code is described on the right.

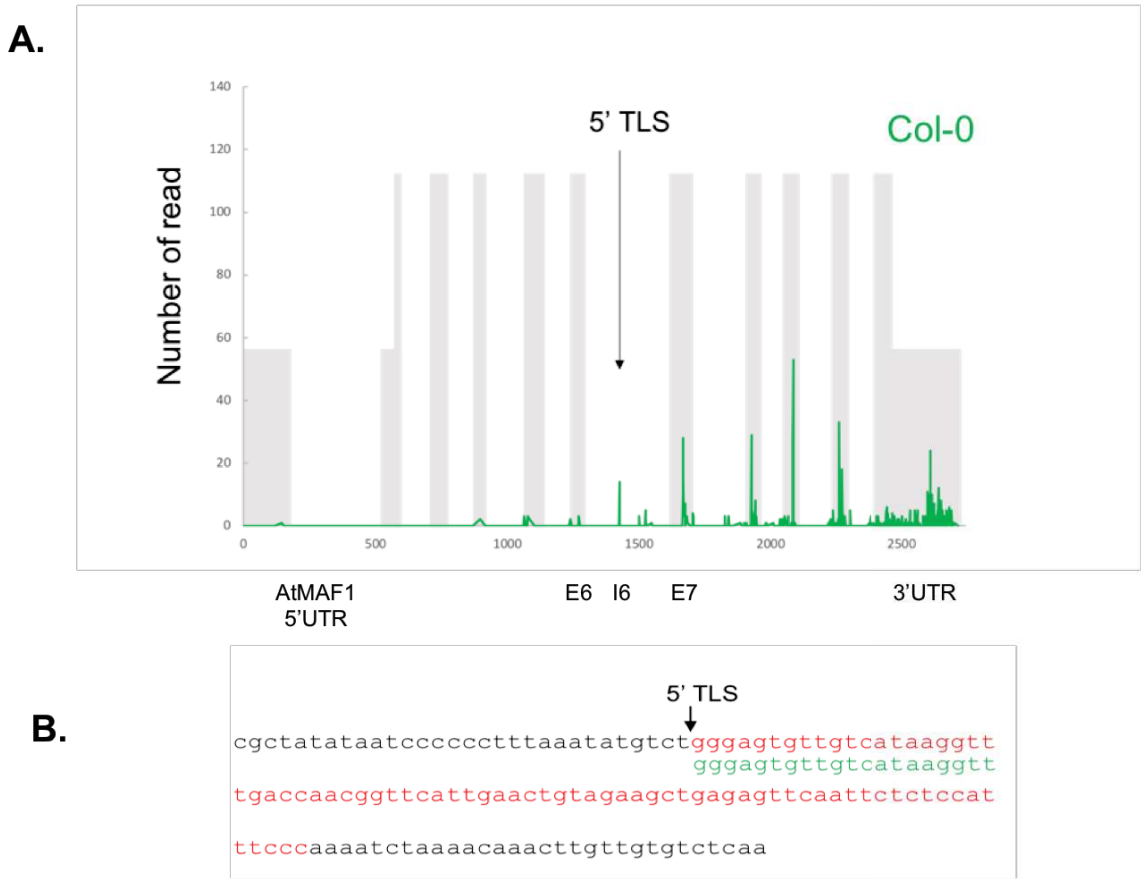


Figure 34: **Graphical representation of PARE-seq data from Arabidopsis Next-Gen database for AtMAF1 gene** (At5g13240). The position of the TLS, in MAF1 intron 6 is indicated by an arrow (A). TLS sequence is indicated in red, and the sequence of the read found in the Next-Gen database is indicated below in green. The start site of the TLS is indicated by an arrow (B).

Beyond our own experimental 3'RACE-Seq results, we also used available PARE-seq (parallel analysis of RNA ends) data produced by the Pamela Green's lab (publicly available at next-gen sequence). PARE-seq is a technique used to study RNA degradation and identify for example miRNA cleavage sites. When focusing on MAF1, the number of transcript ends starts to be above background at the exact 5' start site of the TLS (Figure 34) and this is the only peak localised in an intron. I would like to stress that the TLS sequence is different enough from the tRNA sequences for the reads to be specific. This is an additional proof that a endoribonucleolytic cleavage really occurs in this intron 6 exactly at the 5' end of the transcript.

f) Analysis of RNA polymerase III transcripts when PRORP2/3 is downregulated

As a next step, after showing that AtMAF1 mRNA level was likely also regulated by Arabidopsis nuclear RNase P we decided to investigate if, in our VIGS conditions where AtPRORP3 is absent and AtPRORP2 is down-regulated, RNA transcripts from RNA polymerase III are also down-regulated. Indeed, in our proposed model, AtPRORP2/3 catalyse the cleavage of AtMAF1 precursor mRNA, thereby reducing the level of its mature mRNA and consequently possibly its abundance at the protein level and its repression of RNA pol III. Accordingly, we hypothesised that the downregulation of AtPRORP2/3 by VIGS might lead to an overall decrease of RNA pol III transcripts. It had already been shown by previous work from the team that when AtPRORP2/3 are downregulated, the amount of some mature tRNAs is also decreased (Gutmann et al., 2012), although no transcriptome wide analysis had been performed. So we decided to investigate if this was also the case for other RNA pol III transcripts. Using RT-qPCR we assayed the relative expression level of three other RNA pol III transcripts: 5S ribosomal RNA, 7SL RNA and U6 spliceosomal RNA. The analysis of three independent experiments revealed that 5S rRNA, U6 snRNA and 7SL RNA levels were not significantly decreased in *prorp3 KO* plants with vigs against PRORP2 samples compared to *prorp3* KO plants with vigs by an Empty Vector samples (Figure 35). However, these measures may reveal the steady-state level of these transcripts, and not directly RNA pol III activity.

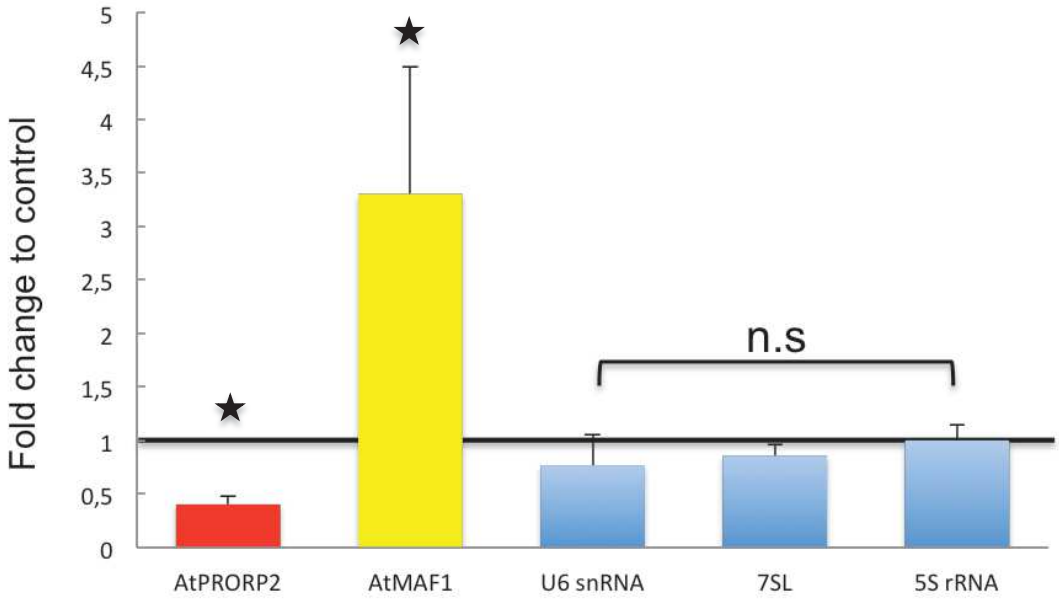


Figure 35: Relative RNA pol III transcripts levels in plants with VIGS against PRORP2 in *prorp3* background compared to plants with VIGS by an empty vector (EV) assayed by RT-qPCR. Standard deviations are indicated by positive error bars. Relative fold change expression level compared to control, indicated on the Y axis, was calculated using the delta-delta Ct method. Difference in level of expression between the two conditions (VIGS against PRORP2 compared to VIGS against empty vector) was assayed using student's t-test and considered statistically significant if $p < 0,05$ (indicated by a star). Otherwise, difference was considered not statistically significant (n.s).

g) Assessment of a possible function of RNase Z for the cleavage of MAF1 TLS

As presented in the introduction, precursor tRNAs are matured in 5' by RNase P and in 3' by RNase Z. While we had identified that RNase P cleaves the conserved AtMAF1 TLS in order to regulate AtMAF1 levels, it was also imaginable that RNase Z would have a similar regulatory function through a cleavage of the 3' extremity of AtMAF1 TLS. We therefore decided to address this hypothesis both *in vitro* and *in vivo*.

As a first step, we investigated if AtMAF1 TLS could be cleaved by a recombinant AtRNase Z3 protein (the nuclear Arabidopsis RNase Z according to Canino et al., 2009) *in vitro*. Recombinant AtRNase Z3 was produced, following the protocol detailed in the Material and methods section, with the CDS for AtRNase Z3 cloned into pet28b and expressed from Rosetta DE3 cells. AtRNase Z3 could be purified to near homogeneity as verified on gel and by immunodetection (Figure 36), and by mass spectrometry analysis. The sample from TRZ3 elution 1 was used for activity assays.

Cleavage assays were performed as described above, with the short transcript representing MAF1 TLS and a pre-tRNA as a positive control. Transcripts were incubated with either AtRNase Z3 alone, AtPRORP2 alone or with both proteins. Result showed that AtMAF1 TLS could indeed be cleaved by AtRNase Z3 but much more effectively in the presence of AtPRORP2 (Figure 37). A similar result was observed with the control tRNA. This suggests that RNase P activity must precede RNase Z cleavage and that only 5' mature tRNAs or TLS are good substrates for RNase Z. Nuclear precursor tRNAs are bound *in vivo* by the LA protein at their 3' extremity. This protein was proposed to prevent RNase Z endonucleolytic cleavage and thus to orient pre-tRNAs towards maturation of their 5' end first (Blewett and Maraia, 2018). Our results obtained *in vitro* suggest that the LA protein is not the main player to drive the preferential cleavage by RNase P first. The nature of the pre-tRNA substrate itself seems to be sufficient to induce preferential cleavage by RNase P first, in accordance with previous literature (Hopper, 2013).

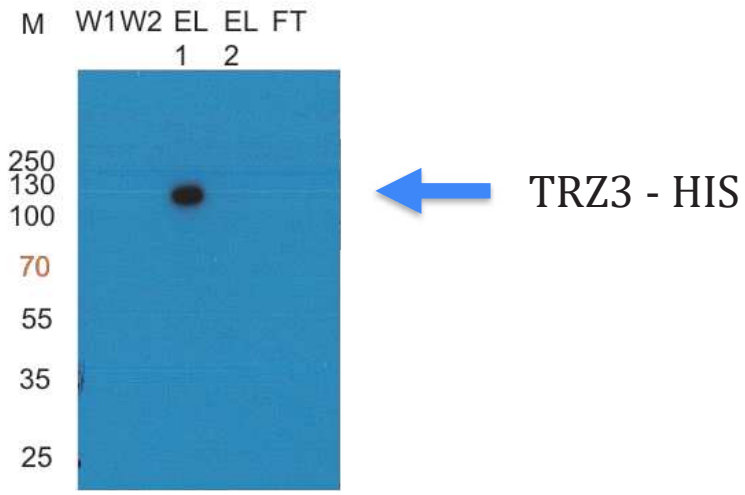


Figure 36: Western blot analysis of AtTRZ3-HIS (calculated MW: 98,6 Kda) after protein expression in bacterial cells and batch purification. Wash (W), Elution (EL) and Flow Through (FT) samples were run on an 8% SDS-PAGE gel. After transfer to a PVDF membrane, western blotting was performed using antibody against HIS. Migration positions of the molecular mass markers are indicated on the left (M).

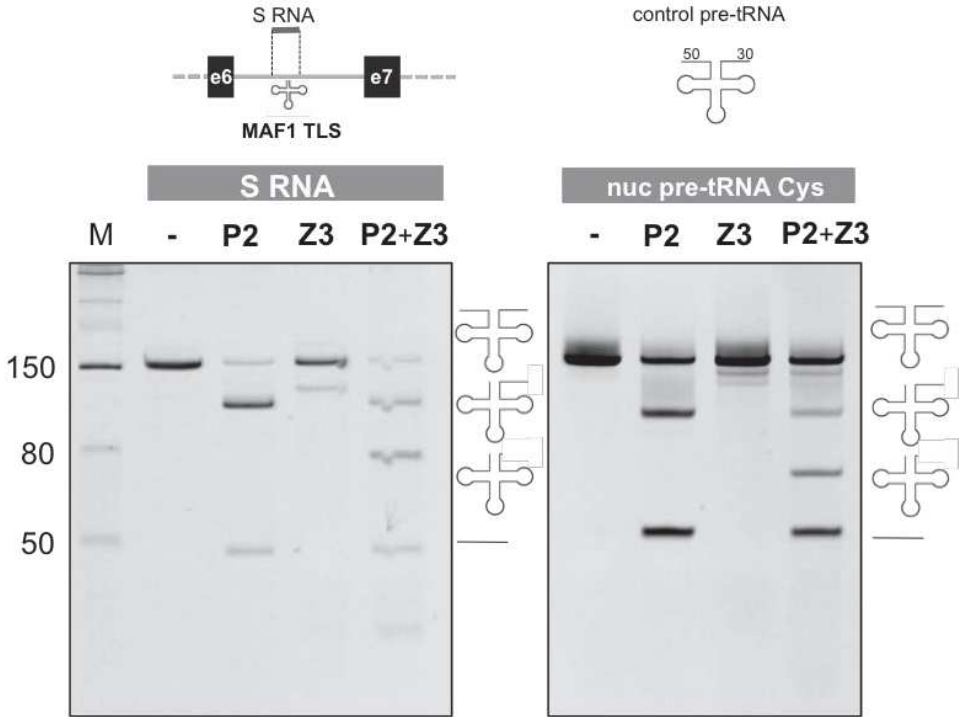


Figure 37: In vitro cleavage assay of MAF1 TLS by AtPRORP2 and AtRNase Z3. A short region called S RNA (corresponding to a portion around the TLS of intron 6 of MAF1) was transcribed *in vitro* and used as a substrate for recombinant PRORP2 (P2) and/or RNase Z3 (Z3). As a negative control, the reaction was performed without protein (-). As a positive control, a precursor tRNA (the nuclear precursor tRNA Cysteine) was used as a substrate, showing that the recombinant PRORP2 and RNase Z3 are catalytically active. The reaction products were run on an acrylamide gel and visualized under UV after ethidium bromide staining.

Then, as a next step, we investigated whether such an RNase Z cleavage could occur at the level of MAF1 TLS *in vivo*. For this, we analysed our 3' RACE Seq assays. This data did not reveal any accumulation of MAF1 transcripts at the specific site at the 3' end of the TLS (position 209) but we can observe a small increase in percentage of reads at position 211 only for p3 KO vigs P2 replicate 2 (Figure 32A).

However, if the results from our previous *in vitro* assays (Figure 37) are correct and that *in vivo*, RNase Z cleavage also mainly follows RNase P cleavage, it is logical that we cannot detect any RNase Z cleavage using 3' RACE. Indeed, the adapter would then ligate itself in 3' of the TLS, but we would not be able to amplify this fragment as our Forward primers are located in 5' of the TLS. Consequently, once an RNase P cleavage has happened, if another cleavage occurs, catalysed by RNase Z, we cannot amplify by PCR until the 3'end of the TLS, as then the RNA is already cleaved. Consequently, with 3'RACE we are only able to detect RNase Z cleavage if it occurs before RNase P cleavage. And we have shown *in vitro* that this is unlikely the case. Consequently, if we want to investigate *in vivo* if AtMAF1 TLS is cleaved by RNase Z, it is necessary to perform a 5' RACE assay. Likewise, we also interrogated public databases. For instance, the analysis of the ncRNA collected and assembled on the Arabidopsis Next-Gen sequence DataBases for the AtMAF1 (At5g13240) showed that no small RNA, i.e. that would correspond to MAF1 TLS, appears to be generated from this gene. Furthermore, Akkuratov et al. (2014) had also searched for a non-coding RNA consisting of the tRNA like region of MAF1 and could not detect such an RNA using RNA blot hybridizations. In conclusion, we did not find any evidence to support the hypothesis that RNase Z would also cleave AtMAF1 TLS, and consequently generate a non-coding RNA. However, we would need to perform additional assays (especially 5' RACE) to be able to convincingly reject this hypothesis.

h) Investigation of alternative hypothesis to explain our findings

Another alternative interpretation of the PARE Seq results could have been that RNA polymerase II would stall at the specific TLS start site, leading to aborted transcription and consequently the abundance in cells of an AtMAF1 pre-mRNA transcripts as observed in our 3'RACE assays. To test this hypothesis, we looked at available Chip-Seq

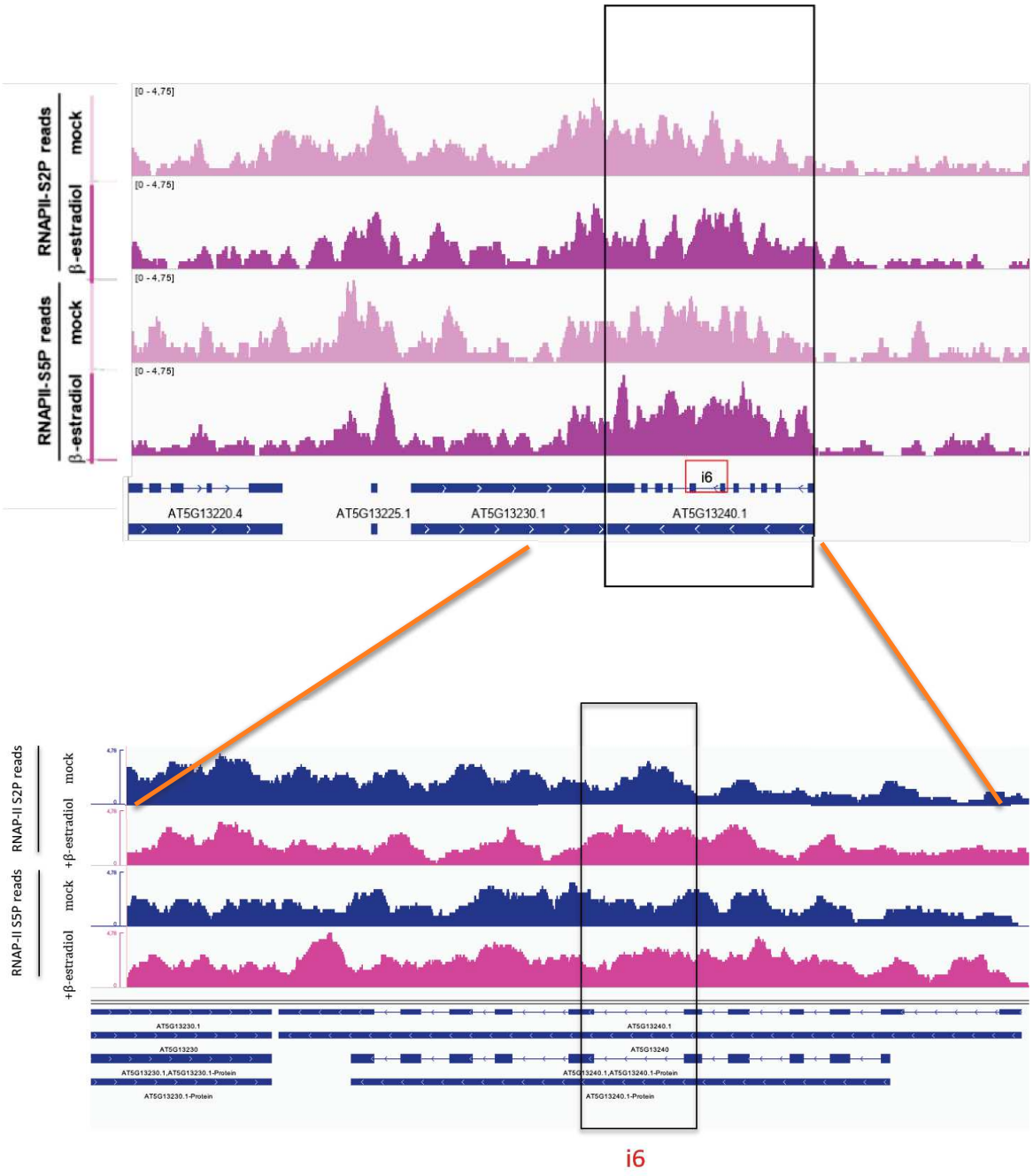


Figure 38: **ChIP-seq profiles of RNAPII-S2P and RNAPII-S5P at the At5g13240** locus upon b-estradiol versus mock treatment. Plots were generated with the Integrative Genomics Browser using representative biological replicates. In gene models, thin bars represent introns and thick bars exons. Position of intron 6, containing the TLS in AtMAF1 is indicated by a red box. A close-up view on AtMAF1 (At5g13240) is provided in the bottom panel.

data for *Arabidopsis* RNA polymerase II (Antosz et al., 2020). Analysis of their data clearly shows that RNA pol II occupancy seems quite homogenous within the whole gene body

(Figure 38). Moreover this result is similar for the two antibody used (these two antibodies, named RNAPII-S2P and RNAPII-S5P, are specific for the phosphorylated forms of RNAPII-CTD (the carboxy-terminal domain of the large subunit of RNAPII)). Additionally, it is worth noting that these data were generated by using a mutated version of the elongation factor TFIIS which is required for efficient rescue of arrested RNA Pol. II by stimulating the intrinsic transcript cleavage activity of the polymerase. Consequently, in their lines, TFIIS activity is blocked without the addition of β-estradiol. Interestingly, for the AtMAF1 gene, RNA pol. II occupancy does not seem to vary dramatically upon treatment with β-estradiol. Supporting the idea that the observed cleavage in AtMAF1 precursor mRNA is not linked to RNA Pol II backtracking / arrest.

Moreover, we also consider, this interpretation as unlikely because the number of transcript termini corresponding to MAF1 TLS 5' extremity were reduced in AtPRORP2 down-regulation mutants, thus pointing a direct role of RNase P for the generation of these termini. Consequently, we rule out the hypothesis or RNA pol. II stalling to explain our findings.

Another possible explanation concerning the biological functions of AtMAF1 TLS is that it would be involved in intron splicing. As this structure is evolutionary conserved in introns and as introns are known to be spliced for gene expression to be correctly achieved, it is a possible explanation that AtMAF1 TLS would be linked in a way to intron splicing. To investigate this hypothesis, we used transgenic *maf1* ko lines complemented with different versions of AtMAF1 (these lines are described below).

We performed RT-PCR on the lines having either a version of the AtMAF1 gene without the TLS or intron 6 of AtMAF1 (that contains the TLS) under the control of its own promoter (as AtMAF1 promoter sequence is unknown we amplified 2 kbp upstream of the 5'UTR) to see if AtMAF1 mRNA size is affected. Our result (Figure 39) shows that the size of the coding sequence (CDS) is similar in WT lines compared to our complemented lines. Consequently, intron 6 is correctly removed even in the absence of AtMAF1 TLS, indicating that this evolutionnary conserved TLS is not involved in intron splicing.

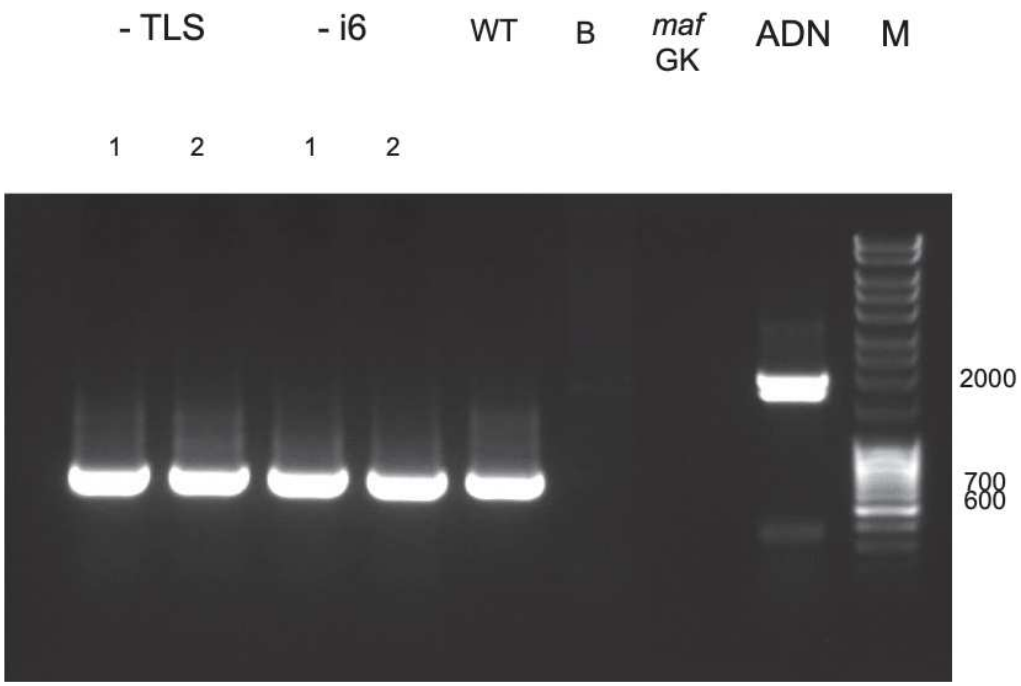


Figure 39: RT-PCR performed on AtMAF1 CDS in WT (Col-0) lines, *maf1* KO lines (*maf* GK) and maf1 KO lines complemented with a version of AtMAF1 that does not contain the TLS sequence (-TLS) or the TLS containing intron (-i6). AtMAF1 CDS size is 675 base pair. On DNA the expected size is 1891 base pair. PCR product deposited in the line "DNA" was conducted on a pool of DNA containg DNA extracted from WT, -TLS and –i6 lines. The –i6 lines do not have AtMAF1 intron n°6, explaining the presence of two distinct bands due to the size difference.

i) Development of genetic tools to investigate a possible interplay between RNase P and RNase Z activities in Arabidopsis

To investigate a possible role of the nuclear RNase Z in intron 6 cleavage at the 3'end of the TLS, we first genotyped plant insertion lines characterised by Canino et al. (2009), that have a T-DNA insertion in AtRNaseZ1 (GABI Kat 696C04) and in AtRNase Z3 (CSHL_GT9807), as well as a cross that had been made between these lines by S. binder (Ulm university). No double homozygous knock out for AtRNaseZ1 and Z3 could be identified, confirming the results of Canino et al. (2009). Then, we crossed *Atprorp3* KO lines (described by Gobert el al., 2010 and Gutmann et al., 2012) with *AtRNase z3* lines (described by Canino et al., 2009). After harvesting seeds from the homozygous F2 generation, we analysed plants having the *p3p3trz3trz3* genotype. These plants did not exhibit any obvious macroscopic phenotype. The result was expected as both AtPRORP2 (for RNase P) and AtTRZ1 (for RNase Z) have been shown to complement AtPRORP3 and AtTRZ3, respectively (Gobert el al., 2010; Gutmann et al., 2012 and Canino et al., 2009).

These plants will be used to further investigate if cleavage of AtMAF1 TLS by RNase P is followed by RNase Z, to generate a new non-coding RNA. As double Knock out of nuclear RNase P or RNase Z in *Arabidopsis* is lethal, these lines are necessary in order to down regulate additionally both AtPRORP2 and AtTRZ1, for example by VIGS.

j) Development of genetic tools to investigate the function of AtMAF1 TLS in vivo

In order to further investigate the biological functions of AtMAF1, we ordered and verified a T-DNA line from the Gabi-Kat collection. The line 031C10 was a verified Gaki-Kat T-DNA insertion line and we confirmed by genotyping and RT-PCR that this line had indeed a T-DNA insertion in AtMAF1 (Figure 40). It is interesting to note that this line does not have any specific macroscopic phenotype. Finally, after sequencing of the region where the T-DNA was inserted we confirmed the T-DNA insertion (Figure 40). Then to investigate in more details the biological function(s) of AtMAF1 evolutionary conserved TLS we created new genetically modified AtMAF1 lines. For this, a 2 kbp

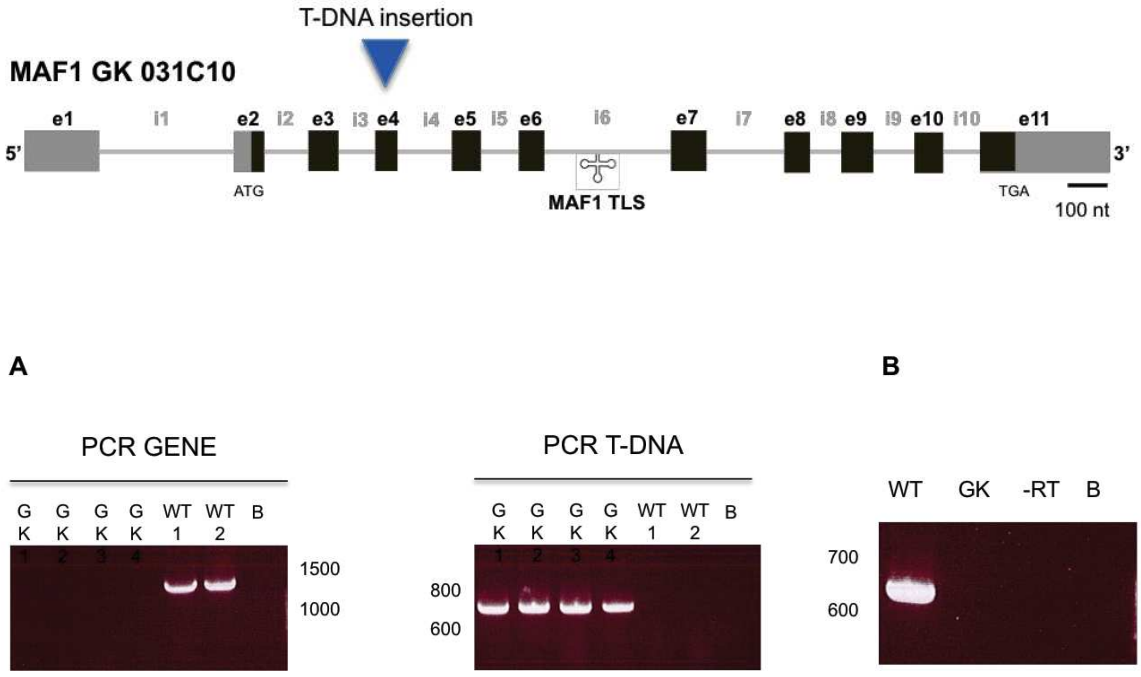


Figure 40: Characterisation of Atmaf1 GabiKAT 031C10 T-DNA insertion line. Schematic insertion of the T-DNA in *Atmaf1* gene is shown (upper part). T-DNA insertion exact localization was verified by sequencing. Insertion of the T-DNA line was verified by PCR. GK stand for Gabi Kat, WT for Wild Type and B for blank. DNA molecular weight markers are indicated by numbers (A). *Atmaf1* Knock Out was also verified by RT-PCR. After reverse transcription on total RNA, AtMAF1 CDS was amplified using primers at both extremities. MAF1 CDS could be amplified in the WT (Col-O) but not in the Gabi-Kat 031C10 line. In the "–RT" lane, the reverse transcription mix + RNA but without the reverse transcriptase enzyme was used as a template for PCR. In the blank lane "B", only water was used as PCR template (B).

region upstream of the MAF1 gene was first amplified (as the exact promoter sequence of AtMAF1 is unknown) with the full length genomic version of the gene and introduced in an entry vector (here pDonor207) using the Gateway ® technology. Then mutagenesis was performed to remove either the TLS (Δ TLS) or the entire intron containing the TLS (Δ i6). A version with only AtMAF1 coding sequence (CDS) was also created under the control of a strong promoter (35S AtMAF1) to investigate the effect of AtMAF1 over-expression (Figure 41). Finally, these three versions (and an unmodified version of AtMAF1 with 2 kbp of the upstream genomic region (that will be used as a positive control to complement the *maf1* KO lines)) were cloned in destination vectors and, after transformation in *Agrobacterium tumefaciens* (GV 3101), *maf1* KO plants (Gabi 031C10) were transformed using the floral dip method. Plants effectively transformed were selected using the Basta herbicide. Plants containing the Basta herbicide resistance BAR gene (and consequently transformed) were resistant to this treatment and further genotype by PCR to monitor the insertion of the sequence of interest (Figure 41). These plants will need to be led to homozygosity to obtain stable lines.

Likewise, the *maf1* GabiKat mutant line described above was also crossed with the *prorp3* KO line. This is still ongoing work as plant homozygous for both the MAF1 insertion and the PRORP3 insertion stills needs to be identified by genotyping the F2 generation.

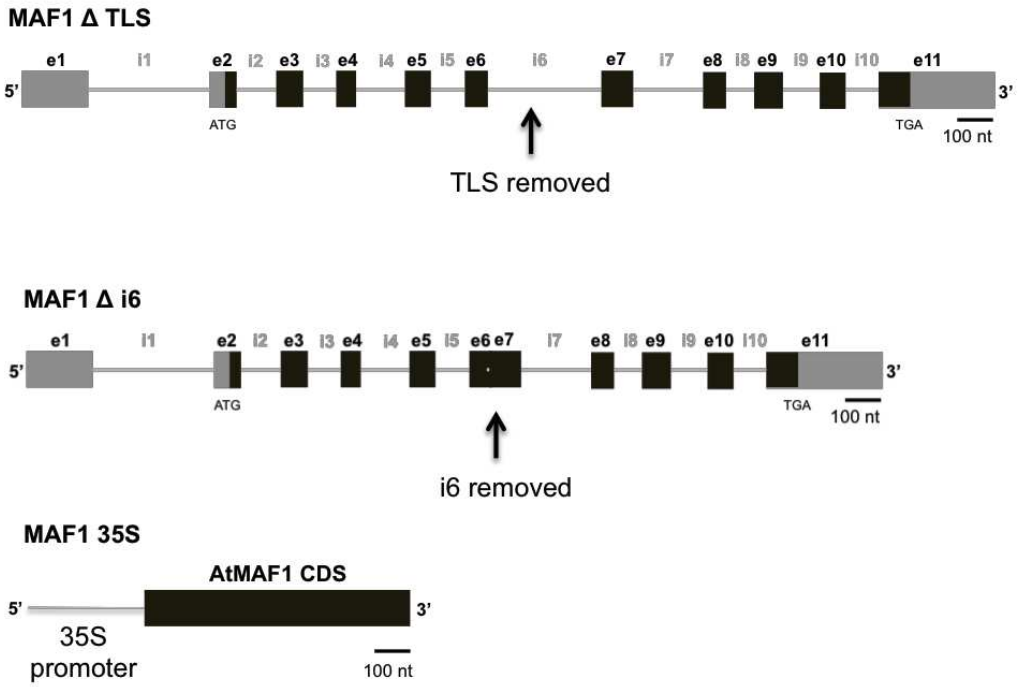


Figure 41: **schematic representation of the different AtMAF1 versions created to complement *maf1* KO lines** and investigate, *in vivo*, the biological function of the evolutionary conserved TLS. In MAF1 Δ TLS, the TLS was removed. In MAF1 Δ I6 it is the intron containing the TLS that was removed. Finally, a version containing a strong promoter (35S) fused to AtMAF1 CDS was generated (MAF1 35S).

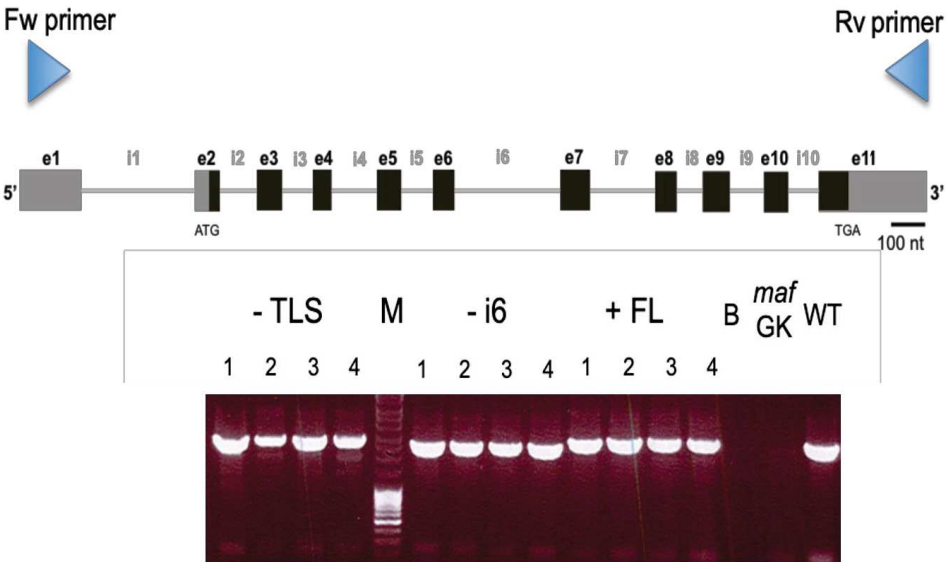


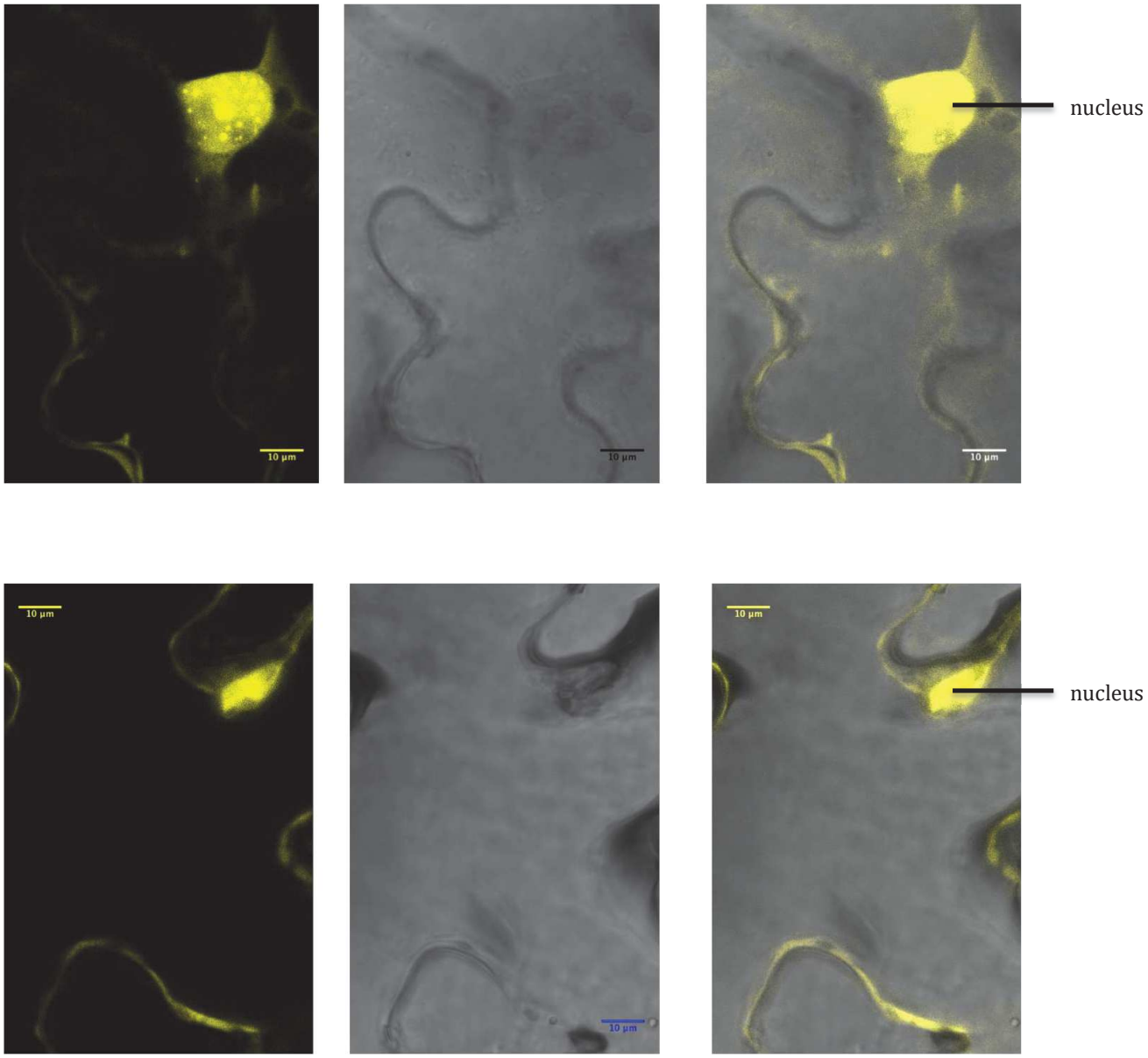
Figure 42: **Genotyping by PCR of AtMAF1 complementation lines.** To ensure correct insertion of the modified versions of AtMAF1 in the *maf1* ko lines, PCR was performed on genomic DNA. AtMAF1 gene could not be amplified in the ko lines (*maf* GK), whereas it is amplified in the ko lines complemented with At MAF1 without the TLS sequence (-TLS), without the TLS containing intron (-i6) or with the full length gene sequence as a positive control (+FL). WT DNA (Col-0) is used as a positive control and the PCR mix without the DNA polymerase (B) as a negative control.

k) MAF1 subcellular localization in Arabidopsis

Finally, we investigated Arabidopsis MAF1 sub-cellular localization because MAF1 has been shown to be either cytoplasmic and nuclear or only nuclear depending on which organism its localisation was studied (Wei and Zheng, 2010). For this we used transient expression of AtMAF1 in *Nicotiana benthamiana*. In brief, AtMAF1 CDS was cloned in pEarleygate 102 (35S :: MAF CDS :: CFP) and transiently expressed in *N. benthamiana* by agro-infiltration. Subcellular localization of the recombinant protein was determined using confocal microscopy. This experiment clearly showed that AtMAF1 is both cytosolic and nuclear as in *S. cerevisiae* (Figure 43). This could suggest that contrary to humans (*Homo sapiens*) where localization of MAF1 seems to be exclusively nuclear, the co-localization of MAF1 in both the cytosol and in the nucleus seems to be a widely distributed pattern throughout evolution.

But this result is still to take with caution. As for the moment, this sub-cellular colocalization of AtMAF1 was determined only using a heterologous system and using confocal microscope. Investigating in a stable Arabidopsis line under the control of its own promoter AtMAF1 subcellular localization is consequently necessary in order to conclude in a more affirmative manner.

Nevertheless, it is worth noting that this result is in accordance with the published data on MAF1 sub-cellular localisation in plants. Indeed, Soprano et al. (2013) had shown that in Citrus MAF1 localized to the cytosol and to the nucleus. Likewise, Ahn et al (2018) had found (after transient expression of AtMAF1 fused with GFP under the control of AtMAF1 promoter) that AtMAF1 was mainly localised to the nucleus but was also found in the cytosol.



<u>Figure 43</u>: AtMAF1 CDS was transiently expressed in *N. benthamiana* and subcellular localization was investigated using a confocal microscope. Images are shown with (left) or without laser excitation (middle) and merged (right). Scale bar was set using the image J software. AtMAF1 localizes dually to the nucleus and the cytosol of the plant cells. Localisation of the fluorescent signal only at the border of the cell in the cytoplasm is due to the vacuole.

Determination of PRORP2 protein partners in Arabidopsis

In addition to determining a new RNA substrate for PRORP2 in Arabidopsis, my PhD project work also focused on determining PRORP proteins partners in the nucleus of *Arabidopsis*.

Indeed, as exposed previously, RNase P is the sole known enzyme that exists as both a Ribonucleoprotein (i.e a protein-RNA complex where the RNA holds the catalytic activity) and in a protein only form. As the RNP form is the major form in Prokaryotic organisms, it is very likely the ancestral form of RNase P. As such, understanding the evolutionary advantages that led to the protein only form to be retained in different eukaryotic organisms and in few Prokaryotes is an interesting question. In addition, if in accordance with the theory of an RNA world prior to the existence of our protein-based world, life relied in its early form on RNA to perform enzymatic reactions, the study of this RNP to a protein-based enzyme transition is of broader interest.

As also exposed previously, *in vitro* studied have ruled out the possibility of a change from an RNP RNase P to a protein only RNase P by a gain in catalytic efficiency. As the RNP form studied, were (*in vitro*), more efficient than PRORP (Pavlova et al., 2012) in catalysing the RNase P activity.

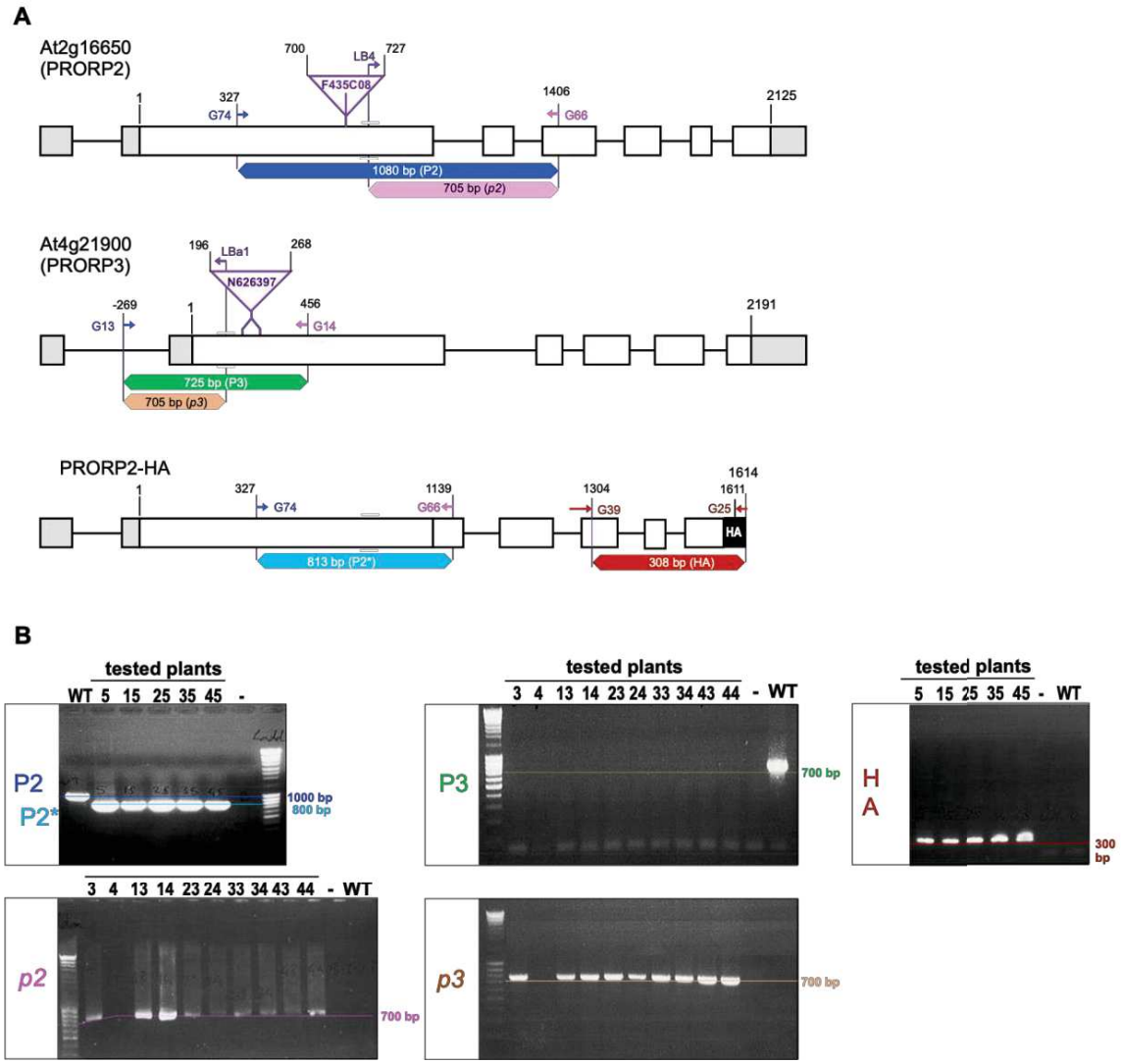
Consequently, an alternative hypothesis is that RNase P transition from an RNP form to a protein only form was favoured by implication of PRORP in new functions. To discover the new function(s) in which PRORP2 might be involved in Arabidopsis, we decided to investigate the protein partners of AtPRORP2 using a hypothesis-free approach. The rationale being that AtPRORP2 is likely involved in the same function as its protein partners. Alternatively, our hypothesis was that AtPRORP2 might be part of a tRNA maturation complex. For example, in complex with the LA protein and / or RNase Z.

l) Transgenic prorp2prorp3 ko lines complemented with PRORP2-HA

Arabidospsis thaliana lines knock-out for *prorp2* and *prorp3* complemented by AtPRORP2 with a C-terminal HA tag were generated by Cedric Schelcher (former PhD student and Post-doctoral researcher in the lab) and Anthony Gobert. These lines are knocked out, by T-DNA insertion, for *prorp3* (At4g21900) and *prorp2* (At2g16650) and they contain the genomic sequence of AtPRORP2, including its 5'UTR region, but not the 3'UTR. At the C-terminal part of the protein, a single HA tag has been inserted.

In order to ensure the right genotype of the starting material, I genotyped plants issued from this line to ensure that they were knocked-out for *PRORP2* and *PRORP3* and complemented by AtPRORP2-HA. The primers used are stated in Figure 44(A) above the gene structures (G74. LB4. G66. G13. LBa1. G14) and their sequences are found in Table 2. The sizes of each PCR product for WT *PRORP2* allele (P2), mutated *PRORP2* allele (p2), WT *PRORP3* allele (P3) and mutated *PRORP3* allele (p3) are stated underneath the gene structures. These sizes correspond to the sizes found for the PCR products of the corresponding PCR after migration through electrophoresis on agarose gels (Figure 44 B). Except for plant 4 (no amplification: genomic DNA probably lost), the plants tested are all homozygous for *prorp*2 *prorp*3 PRORP2-HA. We know that they are homozygous for AtPRORP2-HA because they all contain the AtPRORP2-HA and if the parent was heterozygous for this T-DNA it should segregate and 25% of the plants would be WT.

In conclusion, the selected plants are proper starting material to investigate AtPRORP2 protein partners' using a hypothesis-free approach based on co-immunoprecipitation.



<u>Figure 44</u>: Identification of double knock-out plants for prorp2 and prorp3 complemented by the insertion of prorp2 with an HA tag. A) Gene sequence for AtPRORP2, AtPRORP3 and P2-HA. Color arrows and their name indicate primers. T-DNA insertion, determined in Gutmann et al. (2012) is indicated by triangles. Expected size of the PCR product is indicated by the colour line under the gene structure. B) 1% agarose gel of the PCR products. Colour in each image refers to the product indicated in gene model (A). WT plants are Columbia (Col-0) ecotype. PRORP2-HA gene has one intron that was deleted to help in genotyping. Explaining the difference in size in the PRORP2 gene PCR between WT plants and PRORP2-HA tested plants.

m) PRORP-2 HA can be successfully Co-Immunoprecipitated using these lines

Co-immunoprecipitation was performed as presented in the material and methods section using Milteny® anti-HA beads.

In the lab, a commonly used protocol for Co-IP is to grind frozen flower buds in a buffer containing 50mM of Tris-HCL pH 8, 150 mM of Salt, 1% Triton and anti-proteases. After centrifugation, the supernatant, containing the protein extract is incubated with anti-HA beads and the manufacturer's protocol is followed.

I started performing Co-IP using this protocol and after elution performed Western-Blot to check that PRORP2-HA was effectively enriched. For this, different fractions were retrieved: plant lysate (lysate), flow through from the columns (flow through), washing steps (wash) and elution. These fractions were loaded on an SDS-PAGE to immune-detect PRORP2-HA by Western Blot.

The Western Blot result presented in Figure 45 shows a signal specific of the HA-tag only on the elution lane (EL. on the right). Extra signals are found on the WT and HA-line elution lanes but they certainly correspond to the ECL reacting to the light and heavy chains of the HA-antibody. This result clearly shows that PRORP2-HA is enriched in the elution solution compared to the other fractions. In conclusion, the selected HA-line is useful in order to immunoprecipitate PRORP2-HA and therefore can be used to identify AtPRORP2 protein partners.

However, after analysis by mass spectrometry of the proteins enriched by Co-immunoprecipitation, the results showed that PRORP2-HA was enriched but only at relatively low level (around 150 spectra by IP) and more importantly, all the possible interactants proteins were also detected at very low level (less than 10 spectra).

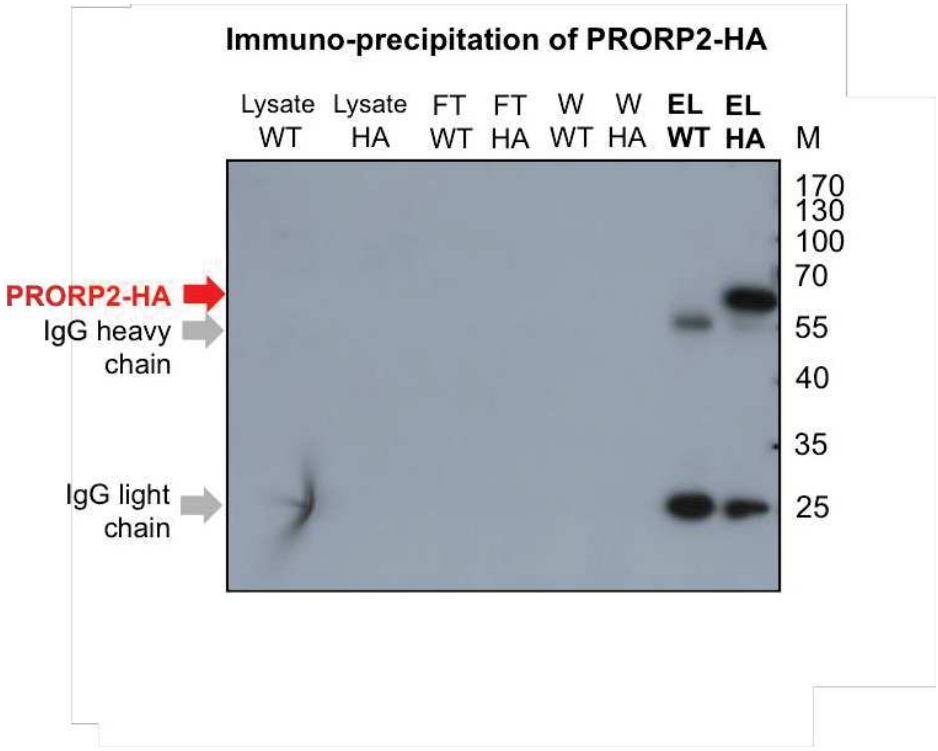


Figure 45: Western blot result after Co-immunoprecipitation with anti-HA beads. Lysate, Flow through (FT), Wash (W) and elution (EL) samples were run on an 8% SDS-PAGE gel. Proteins were transferred to a PVDF membrane and processed for western blotting using antibody against HA. Migration positions of the molecular mass markers are indicated on the right (M). Position of the IgG heavy and light chain are indicated in grey (left). A red arrow indicates the band corresponding to AtPRORP2-HA.

A statistical analysis was performed with the package IPinquiry developed at IBMP. This analysis identified 99 proteins with a number of spectra above 3 that are statistically significantly enriched in HA plants compared to WT (padj=0.05). Out of these, 11 were RNA binding proteins. Interesting candidates were proteins belonging to the class of the DEA(D/H)-box RNA helicase, the U2 small ribonucleoprotein or subunits of the mediator complex. However, none of these possible interactants had a number of spectra above 10 and most importantly, none was always retrieved in all co-IP experiments performed (n=9). The first 25 proteins classified upon their significance (padj) are presented in Table 3.

From these results, it can be noted, that one gene is of particular interest. AT3G49490 encodes an unknown protein. But using a protein blast approach, it appears that this protein contains a WW protein domain similar to the one of MNU2, a mitochondrial protein shown to be a protein partner of PRORP1 (Bouchoucha et al.).

However, the other significant proteins are mostly contaminations because of their localisations and their functions. The representation of the data on a volcano plot (fig.46) shows that PRORP2-HA does not have any protein partner of interest detected by this method. Meaning these experimental conditions are not adequate to identify AtPRORP2 protein partners.

Consequently, to determine AtPRORP2 interactome, it was necessary to improve the Co-IP protocol.

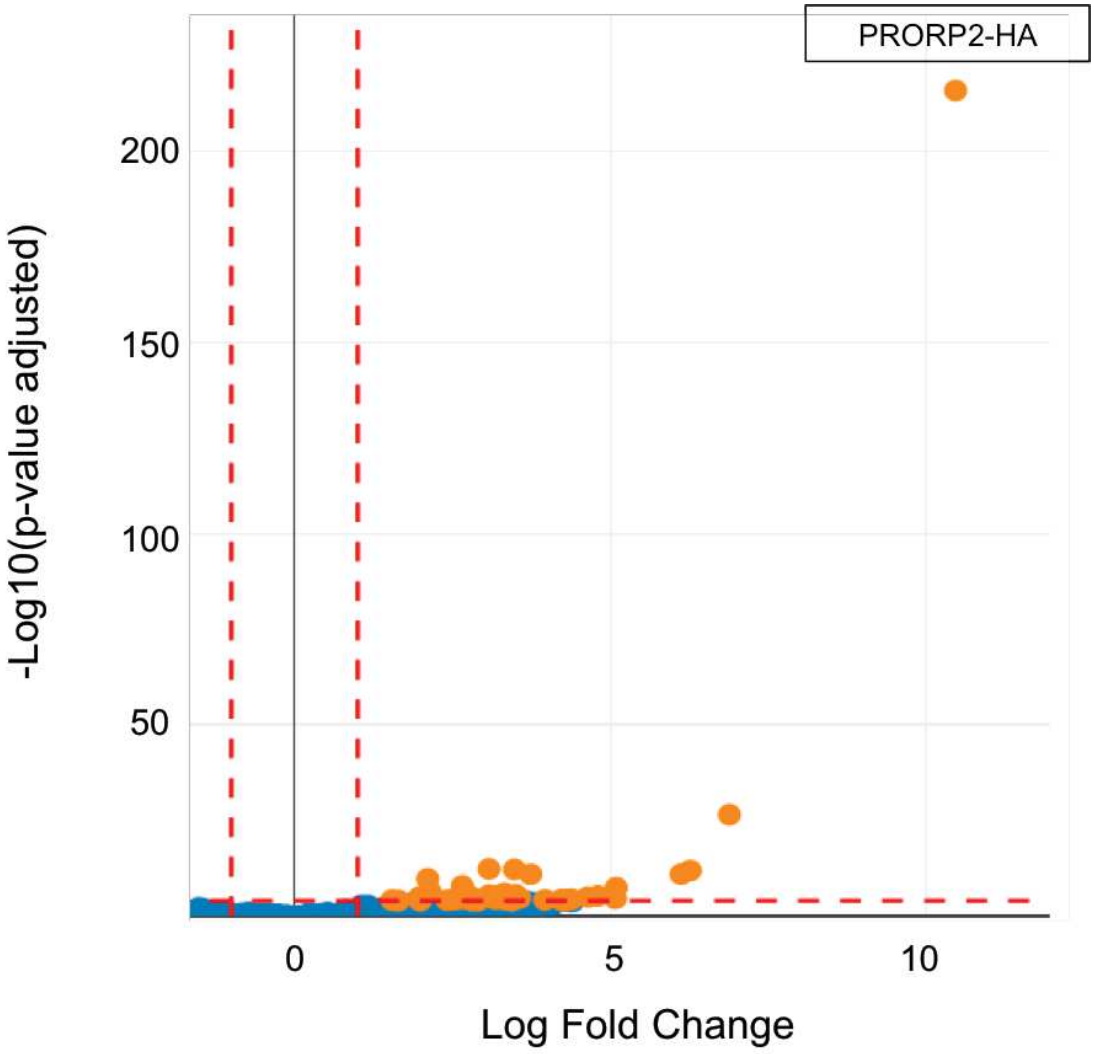


Figure 46: semi-Volcano plot of Co-IP experiments (n=9) of AtPRORP-HA co-immunoprecipitation assays using the common Co-IP method (grinding of flower buds in 50mm Tris pH = 8, 50 mM Nacl and 1% Triton). AtPRORP2-HA, our bait, was successfully enriched in our Co-IP experiments using this method. However, no interactants of interest could be detected. The adjusted p-value (y axis) is the p-value, adjusted for multiple testing. The log fold change (x-axis) is the log2 of the measure describing the change in protein quantity. The vertical dotted line indicates the threshold for a fold change of 2. The horizontal dotted line indicates the threshold for an adjusted *P*-value of 0.05. Orange dots display proteins significantly enriched in our experiments. Blue dots proteins not significantly enriched.

Rank	Gene name	AGI	adjusted p-value	Description
1	PRORP2	AT2G16650.1	1,8E-216	Proteinaceous RNase P 2
2	ESP	AT1G54040.2	3,3E-27	Epithiospecifier protein
3	ONERIC1	AT3G21140.1	5,2E-13	Pyridoxamine 5'-phosphate oxidase family protein
4	UPF1	AT5G47010.1	6,9E-13	Regulator of nonsense transcripts 1 homolog
5	-	AT3G49490.1	1,4E-12	Uncharacterized protein T9C5.90
6	ATP1	ATMG01190.1	1,2E-11	ATPase subunit 1
7	-	AT1G51560.1	1,3E-11	Pyridoxamine 5'-phosphate oxidase family protein
8	NAIP1	AT4G15545.1	2,0E-10	NAI1 interacting protein, involved in ER body formation.
9	PRS5	AT2G44530.1	1,3E-08	Ribose-phosphate pyrophosphokinase 5, chloroplastic
10	-	AT1G07310.1	3,8E-08	Calcium-dependent lipid-binding (CaLB domain) family protein
11	DRP1C	AT1G14830.1	5,2E-07	DRP1C
12	DBP	AT2G45820.1	1,2E-06	Remorin
13	PRS2	AT1G32380.1	2,8E-06	Ribose-phosphate pyrophosphokinase 2, chloroplastic
14	-	AT5G14720.1	3,8E-06	Protein kinase superfamily protein
15	PRS1	AT2G35390.2	5,2E-06	Ribose-phosphate pyrophosphokinase 1, chloroplastic
16	TADA	AT1G68720.1	5,7E-06	tRNA(adenine(34)) deaminase, chloroplastic
17	PME5	AT2G47040.1	5,7E-06	Pectinesterase 5
18	ABCC14	AT3G62700.1	7,7E-06	ABC transporter C family member 14
19	DRP2B	AT1G59610.1	9,5E-06	Dynamin-2B
20	-	AT1G26850.1	1,1E-05	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
21	SCPL34	AT5G23210.1	1,1E-05	Serine carboxypeptidase-like 34
22	TPLATE	AT3G01780.1	2,0E-05	Protein TPLATE
23	-	AT3G23300.1	2,6E-05	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
24	VHA-E3	AT1G64200.1	3,0E-05	vacuolar H+-ATPase subunit E isoform 3
25	CHUP1	AT3G25690.1	3,1E-05	Hydroxyproline-rich glycoprotein family protein

Table 3: List of the 25 proteins with the best adjusted p-value in our co-iP experiments using the lab classic method. Genes are ranked according to their adjusted p-value. Gene name and described or predicted gene function are indicated when available. AGI stands for Arabidopsis Genome Initiative locus code and is the unique identifier assigned to each locus in *Arabidopsis thaliana*. AtPRORP2 is highlighted in yellow, and in blue the unknown protein At3g49490 – see text for details.

n) Improving PRORP2 Co-IP protocol

A) Arabidopsis nuclei purification

To improve my co-IP results, a first step was to purify nuclei to perform co-IP directly on the organelle where AtPRORP2 is localized, as it is done for mitochondrial or chloroplastic proteins.

For this purpose, different protocols were tried and their efficiency assayed by confocal microscope observation after staining with DAPI. Based on this observation, one protocol (Calikowski TT and Meier I. Isolation of nuclear proteins. Methods Mol Biol. 2006;323:393-402. doi:10.1385/1-59745-003-0:393) was the most efficient in our hands.

However, when performed on purified nuclei, Co-IP did not prove efficient. Indeed, AtPRORP-HA was enriched at very low level (average number of spectra < 10).

We explain this result by the fact that a much higher amount of purified nuclei should have been used and by the possible loss of proteins during the nuclei purification steps. Indeed as protein in the nucleus are less abundant than in the cytosol, and as most protocol to purify nuclei include numerous purification steps; it is likely that nucleus membrane breaks done during purification and consequently releases its proteins. This is possibly not an issue when the goal is to retrieve nuclear RNA or DNA. However, in our case, we were interested in proteins, where an important amount (1 mg of protein is commonly considered as the minimal quantity needed to perform Co-IP) is desired, as no amplification steps exist to date for these molecules.

Consequently, we decided to concentrate our effort on Co-IP protocols using total proteins extracts to directly immunoprecipitated our bait (AtPRORP2-HA).

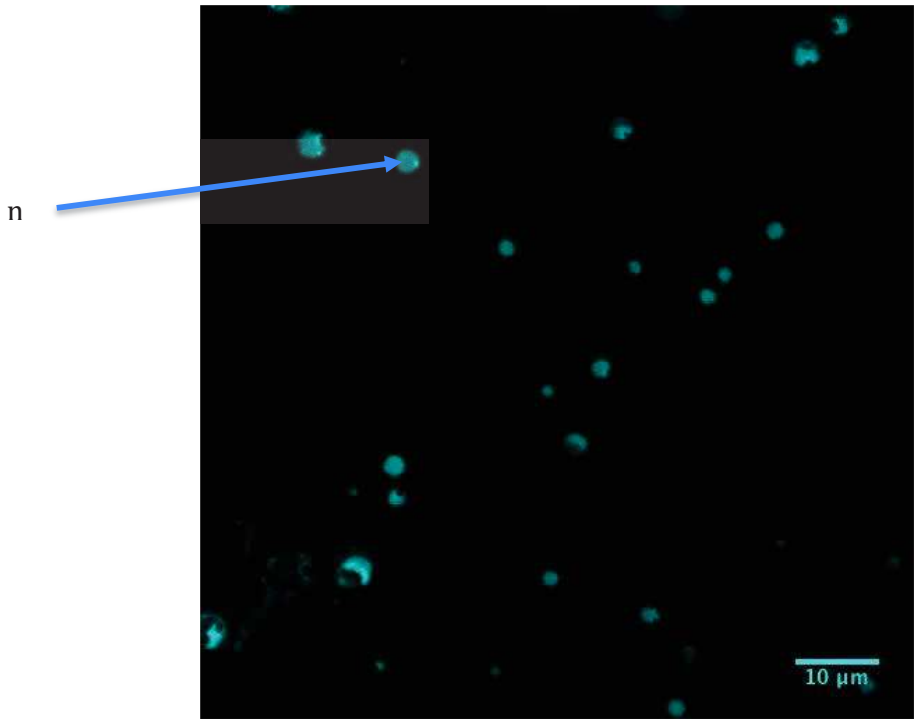


Figure 47: Representative image of Arabidopsis nuclei enriched with the protocol from Calikowski and Meier (2006), observed with a Zeiss LSM 780 confocal microscope, after staining with 4′,6-diamidino-2-phenylindole (DAPI). "n" is for nucleus.

B) Formaldehyde crosslink Immuno-precipitation

The next protocol used has been developed by the IBMP-IBMC Proteomic platform. The detailed protocol used is given in the Material and method section. This protocol proved very efficient to immuno-precipitate AtPRORP2-HA. Indeed, using 0,7g of *Arabidopsis*' flower buds we could commonly detect more than 200 spectra of AtPRORP2-HA after MS/MS analysis.

However, only few possible interactants were obtained as visualised in the volcano plant (Figure 48). Moreover, as for the first method, the identity of these interactants was highly variable between the different biological replicates. In more details, 52 proteins were detected with a number of spectra superior to 3. Including different RNA binding proteins, transcriptional regulator and DEA(D/H)-box RNA helicase. As it can be seen in the volcano plot, for only 9 additional proteins to AtPRORP2, the statistical confidence of their enrichment (adjusted p-value) is inferior to 0,05.

However, even if these proteins were potentially interesting AtPRORP2 partners, they were detected at very low level and not in all experiments. Indicating that our approach was not successful to determine AtPRORP2-HA protein partners.

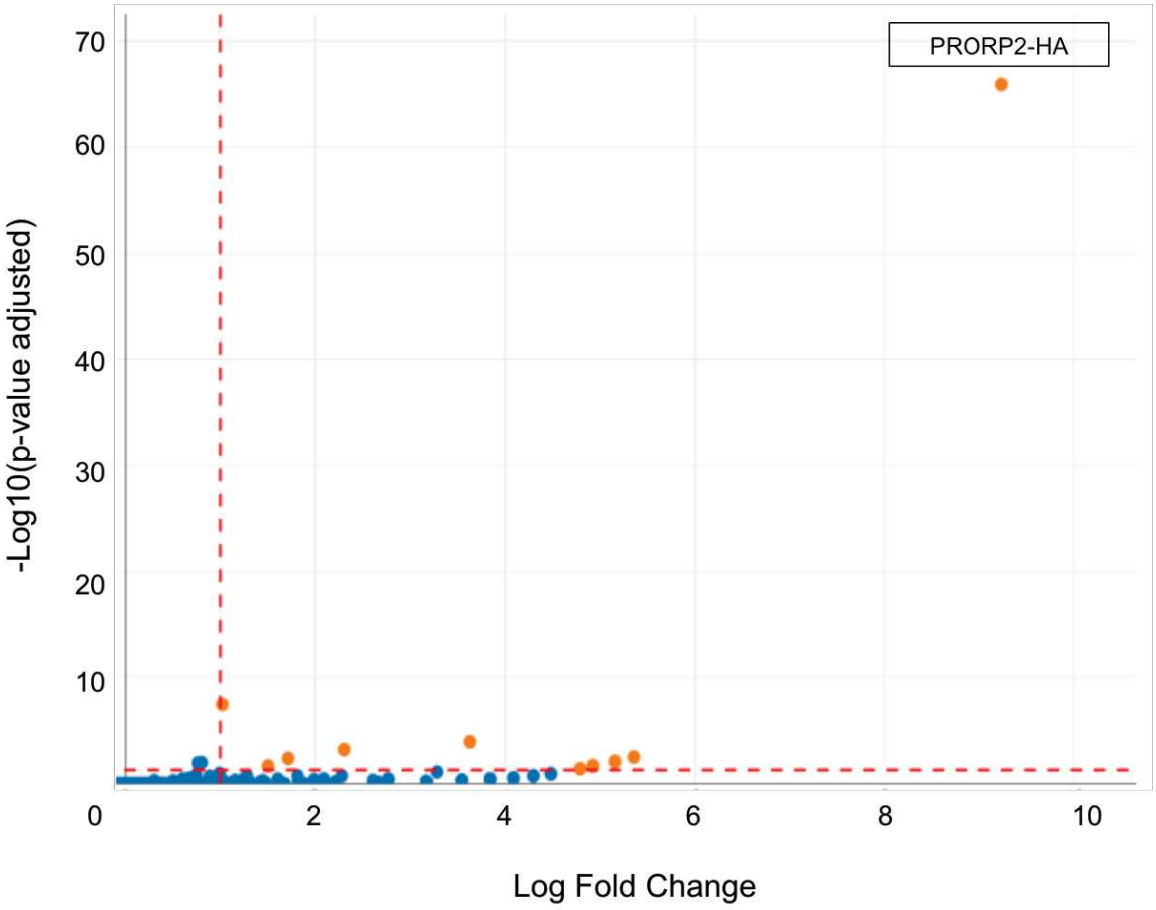


Figure 48: semi-volcano plot of Co-IP experiments (n=3) of PRORP-HA co-immunoprecipitation using the Formaldehyde crosslink method. PRORP2-HA, our bait, was successfully enriched in our Co-IP experiments using this method. However, no interactants of interest could be detected. The adjusted p-value (y axis) is the p-value, adjusted for multiple testing. The log fold change (x-axis) is the log2 of the measure describing the change in protein quantity. The vertical dotted line indicates the threshold for a fold change of 2. The horizontal dotted line indicates the threshold for an adjusted *P*-value of 0.05 Orange dots display proteins significantly enriched in our experiments. Blue dots proteins not significantly enriched. Red dotted line indicate the threshold chosen for statistical significance in –Log10(p-value adjusted) and Log Fold Change.

o) Co-IP with Igepal / NP-40 and determination of AtPRORP2 protein partners

To improve my results, I decided to use a protocol developed in the laboratory of Dr. Pascal Genschik, at IBMP. As members of his team had successfully used it to co-Immunoprecipitate nuclear proteins in Arabidopsis. A detailed protocol for this experiment is given in the Material and Method section.

Moreover, as our previous Co-IP results had shown that it was unlikely that PRORP2-HA was engaged in strong interactions with proteins partners, I decided not to add salt to the protein extraction buffer, in order to reduce stringency, as salt concentration can affect protein-protein interactions. Using this method I hypothesized that I would be more likely to co-IP proteins partners involved in transient interactions with AtPRORP2.

This method proved successful to co-immunoprecipitate AtPRORP2-HA and identify two potential protein partners. Analysis of the proteins co-immunoprecipitated with PRORP2-HA show that two N2,N2-dimethylguanosine tRNA methytransferase 1 named TRM1A (AT5G15810.1) and TRM1B (AT3G02320.1) were significantly enriched as it can be visualized on the volcano plot in Figure 49.

This result is interesting. Indeed, as previously presented, after transcription by RNA pol III, tRNAs are bound at their extremity by LA protein, that bind their poly-U tails and are considered as necessary for the maintenance of tRNAs structure. However, as AtLA1 protein is not found in our Co-IP experiments, our result might indicate that AtTRM1A and B, that catalyse an extremely conserved base modification, considered as playing a role in tRNA structure (Lorenz et al., 2017), are interacting with AtPRORP2, in order for the tRNA to fold correctly for efficient RNase P activity.

Consequently, the next step was to assay if AtPRORP2 and AtTRM1A and B interaction was a direct interaction or if their interaction was mediated by tRNAs.

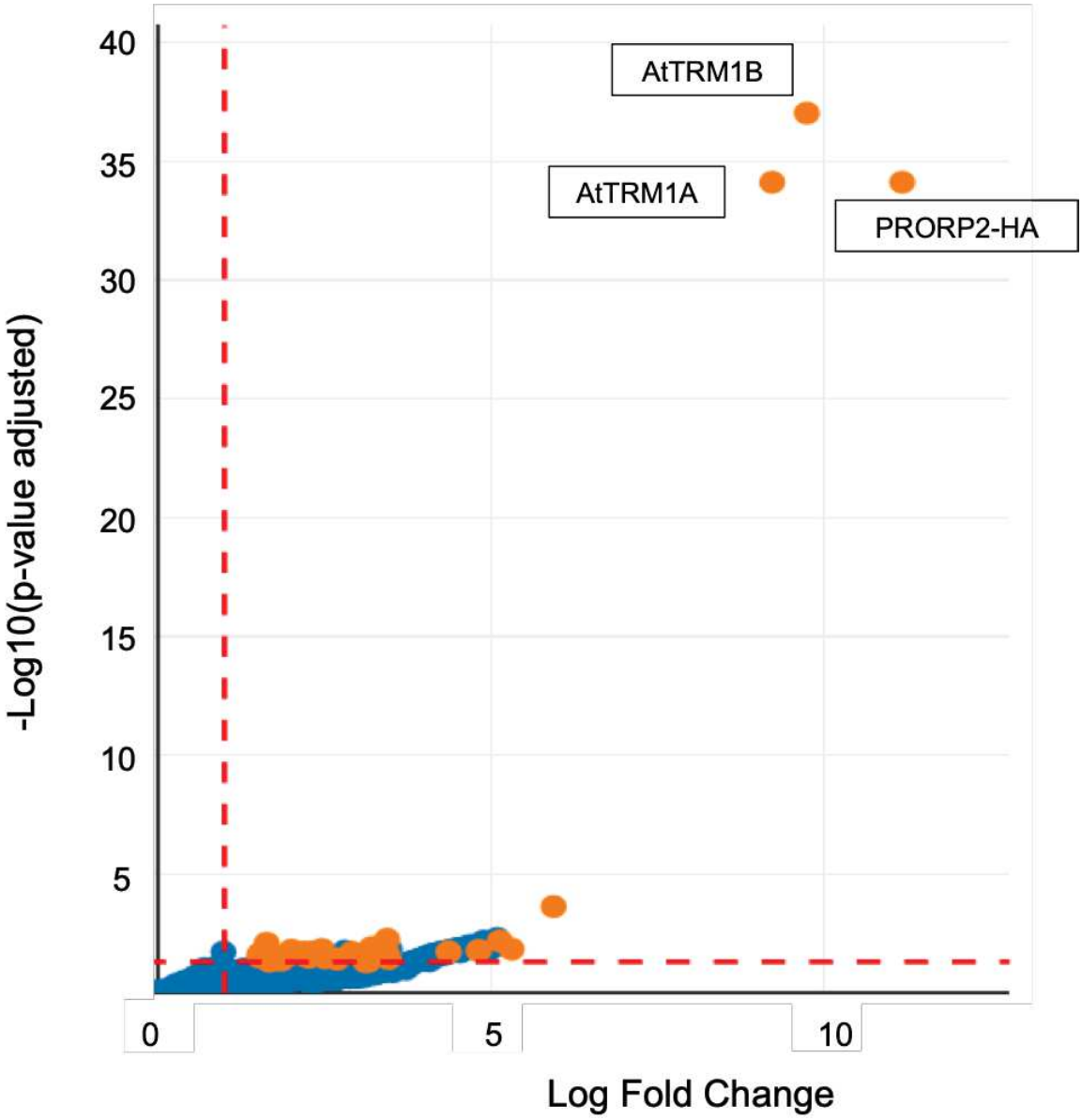


Figure 49: semi-volcano plot of Co-IP experiments (n=5) of PRORP-HA co-immunoprecipitation using the +IGEPAL / - Salt method. PRORP2-HA, our bait, was successfully enriched in our Co-IP experiments using this method. Two proteins were also statistically enriched in these experiments; AtTRM1A (AT3G02320) and AtTRM1B (AT5G15810); and were considered as possible interactants, to be tested in additional assays. TRM1 stands for N2,N2-dimethylguanosine tRNA methyltransferase. The adjusted p-value (y axis) is the p-value, adjusted for multiple testing. The log fold change (x-axis) is the log2 of the measure describing the change in protein quantity. The vertical dotted line indicates the threshold for a fold change of 2. The horizontal dotted line indicates the threshold for an adjusted *P*-value of 0.05. Orange dots display proteins significantly enriched in our experiments. Blue dots proteins not significantly enriched. Red dotted line indicate the threshold chosen for statistical significance in –Log10(p-value adjusted) and Log Fold Change.

p) Co-IP with addition of a ribonuclease enzyme (Benzonase®) or of NaCl (150 mM) to determine if the interaction between AtPRORP2 and AtTRM1 A/B is direct or indirect.

Co-immunoprecipitation assay was repeated with two different methods to investigate whether the interaction between AtPRORP2 and AtTRM1A/B is direct or indirect.

In a first assay, 150mM of salt (NaCl) was simply added to the protein extraction buffer to assay if salt concentration affects the interaction between AtPRORP2 and AtTRM1A andB.

Our result (Figure 51) shows that addition of salt to the buffer clearly decreased the number of spectra corresponding to the AtTRM1A and B proteins. (These experiments were carried out twice; each time with two technical replicates).

This result both tends to confirm that the interaction between PRORP2 and TRM1A/B is a true protein-protein interaction. As if TRM1A/B was only precipitated as a contaminant, addition of salt should not affect its presence or absence. Additionally, this result indicates that the interaction between TRM1A/B and PRORP2 is sensitive to salt concentration, meaning that it is a weak interaction.

In a second assay, we investigated if the addition of RNase would affect the interaction between AtPRORP2 and AtTRM1A and B. For this purpose, Co-immunoprecipitations were performed with the addition of an engineered nuclease enzyme (branded under the name « Benzonase » by Merck®), or with addition of RNase A/T mix (Thermo Fischer). Different methods were tried, either by adding the enzyme directly in the input fraction (#1). Or by adding the enzyme directly on the Co-IP Milteny® magnetic column (#2), or by adding the enzyme after binding of the bait protein (AtPRORP2-HA) on the column, extensive washes and elution in a new tube containing the RNase (#3). However, none of these methods proved successful.

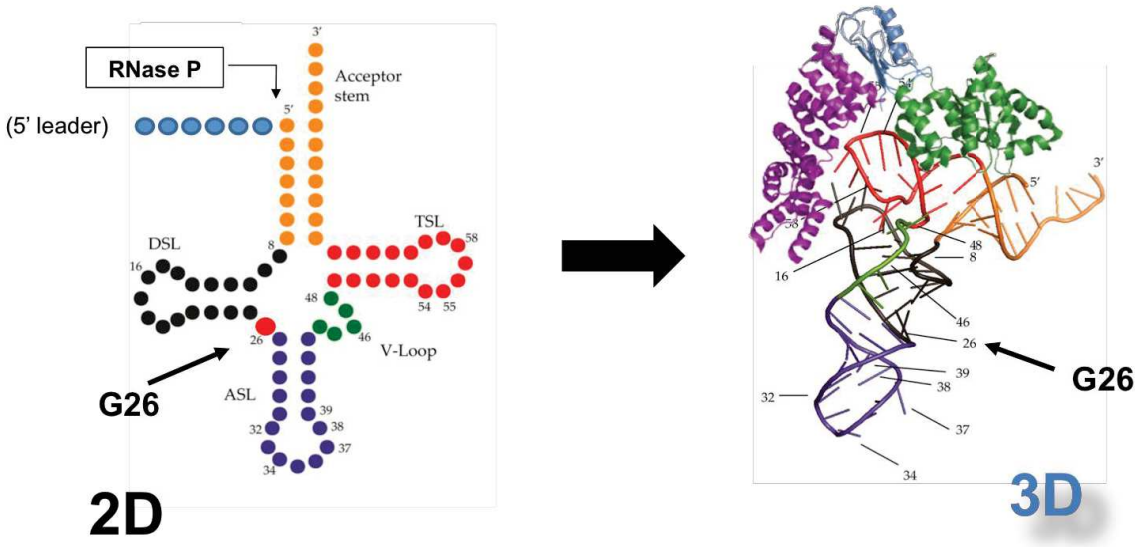


Figure 50: 2D and 3D representation of tRNAS. The 5' leader sequence and base G26 (modified by TRM1) are highlighted. In the cloverleaf (2D) representation of tRNAs, the 5' leader sequence and the conserved base G26 are in close vicinity. However, in 3D (on the right), PRORP2 is located as in the model developed by Pinker et al. (2017), at the T and D loop junction. And base G26 appears to be located on the opposite side of the tRNA.

Accession	Gene name	Ctrl 1	Ctrl 2	P2-HA 1	P2-HA 2	P2-HA 3	P2-HA 4
At2G16650	AtPRORP2			231	251	177	244
At5G15810	AtTRM1B			1	2	8	31
At3G02320	AtTRM1A					7	19

Figure 51: number of spectra obtained in separate Co-IP assays (n=2) using the method from P. Genschik's lab and 150 mM of salt. Only the number of spectra for AtPRORP2 and AtTRM1A/B is shown.

AGI	Gene name	WT	WT + Rnase A/ T	WT + Benzonase	PRORP2-HA	PRORP2-HA + Rnase A/T	PRORP2-HA + Benzonase #1	PRORP2-HA + Benzonase #2	PRORP2-HA + Benzonase #3
AT2G16650.1	PRORP2	0	0	0	221	120	154	152	75
AT5G15810.1	TRM1B	0	0	0	122	111	34	83	33
AT3G02320.1	TRM1A	0	0	0	68	76	24	69	20

Figure 52: number of spectra obtained in separate Co-IP assays with different methods to perform RNase treatment on the input solution. Only the number of spectra for AtPRORP2 and AtTRM1A/B is shown. Three different methods were tried when adding Benzonase (Merck) to the experimental protein preparation (see text for details). AGI stands for Arabidopsis Genome Initiative locus code and is the unique identifier assigned to each locus in *Arabidopsis thaliana*.

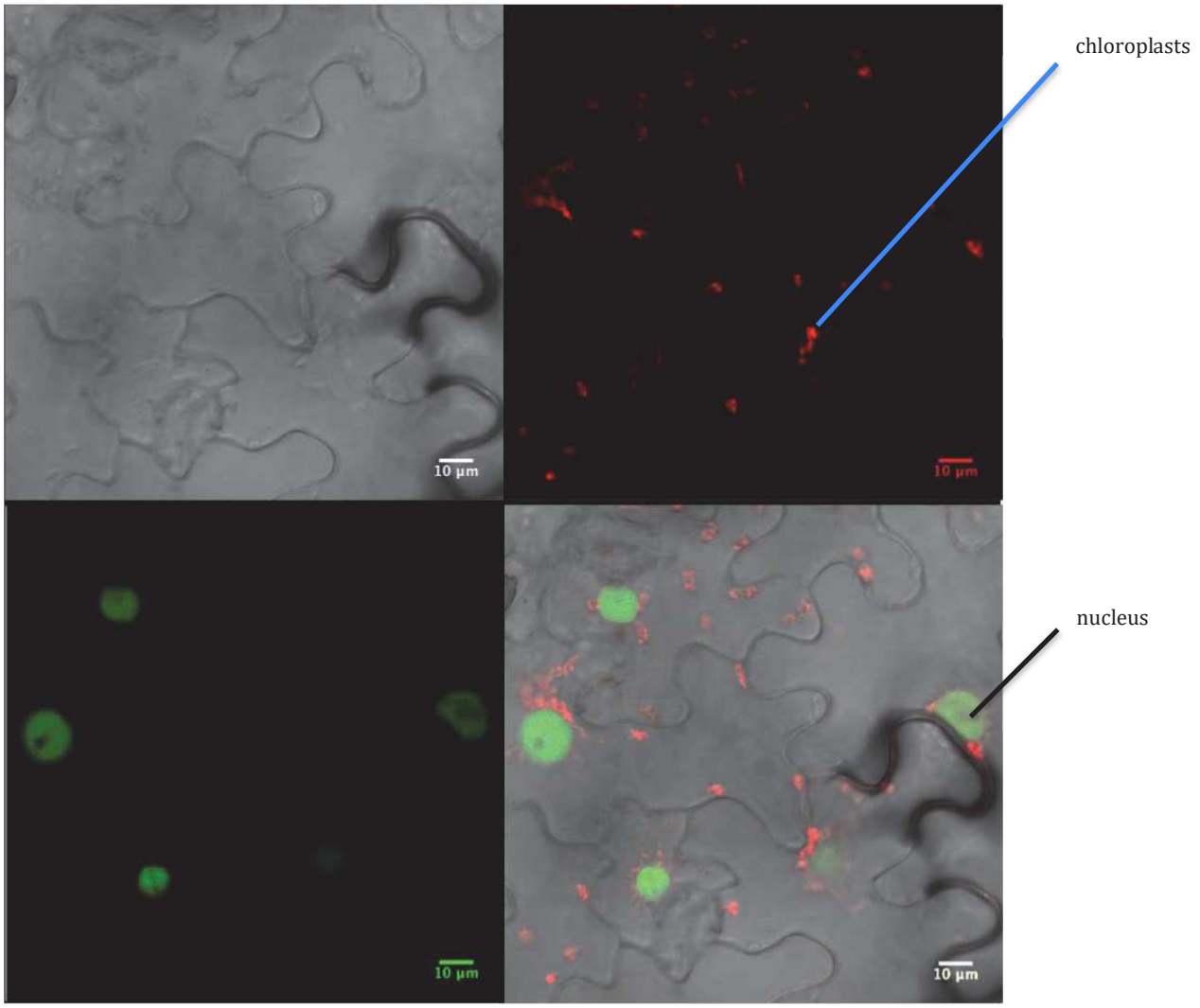
Results showed that PRORP-HA was still co-immunoprecipitated but in a less successful manner (Figure 52). Moreover, addition of Benzonase leads to an increase in the total amount of proteins detected as RNA is degraded. As a consequence, ribosomal proteins are no longer bound together with rRNAs and, dramatically increase the number of proteins eluted with PRORP2-HA. Consequently, it is impossible to decide if the lower number of spectra obtained for TRM1A/B obtained in the different CO-IPs performed with addition of Benzonase are due to the lower number of spectra for PRORP2-HA and the elevated number of additional contaminant proteins detected, or if they are the direct effect of RNA degradation.

Consequently, an efficient protocol to perform RNA degradation during co-immunoprecipitation must be first developed in order to investigate efficiently if the interaction between PRORP2 and TRM1A/B is direct or indirect. For the moment it is still an open question as no definitive evidence can be found to support one or the other hypothesis. Alternatively, a yeast to hybrid assay might be the most efficient way to find the correct answer to this question.

q) TRM1A/B nuclear localization was confirmed by transient expression in Nicotiana benthamiana

To ensure the robustness of our Co-IP results, we decided to confirm the nuclear localization of Arabidopsis TRM1A and B. AtTRM1A and AtTRM1A were cloned in pEarleyGate 102 (35S::TRM1A CDS::CFP:: HA) and pEarleyGate104 (35S:: YFP :: TRM1B CDS) respectively.

After agro-infiltration of *N. benthamiana* leaves, protein sub-cellular localization was determined using confocal imaging. Results clearly show that AtTRM1A and B are both localized in the nucleus (Figure 53 and 54). No other localization could be detected. This result matches with the bio-informatic prediction that were made using “localizer” webserver that indicate the presence of a nuclear localization signal (NLS) in the protein sequence of AtTRM1A and B (Table 2)



<u>Figure 53</u>: AtTRM1A is localized in the nucleus. Localization of AtTRM1A was determined after transient expression of the Arabidopsis protein in *N. benthamiana* leaves and imaged using a Zeiss LSM 780 confocal microscope after 48 hours. Autofluorescence of the chloroplasts can be seen in red in the image of *N. benthamiana* expression TRM1A – CFP (in green)

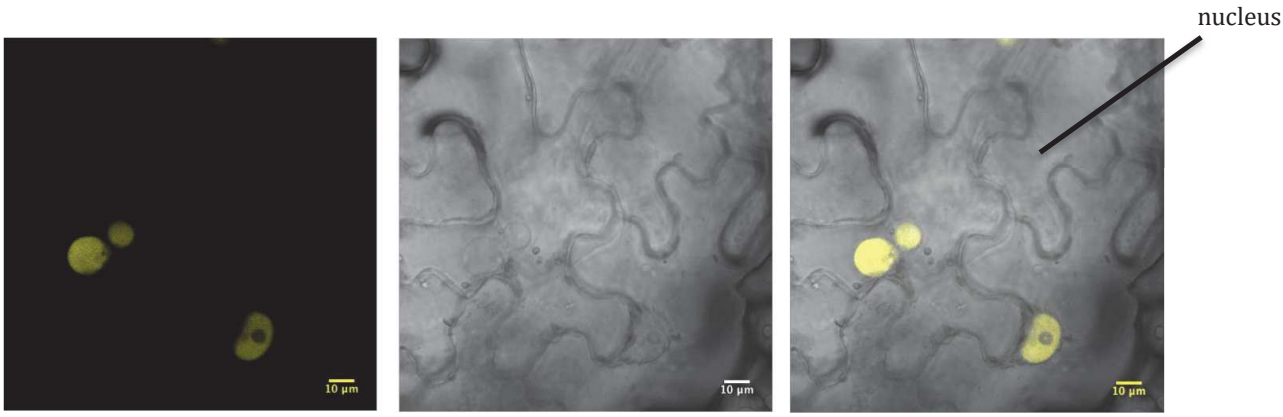


Figure 54: AtTRM1B is localized in the nucleus. Localization of AtTRM1B was determined after transient expression of the Arabidopsis protein in *N. benthaniama* leaves for 48 hours and imaged using a Zeiss LSM 780 confocal microscope.

	Chloroplast	Mitochondria	Nucleus
AtTRM1A	-	-	KRQK, KRKQEHEAKSSKRTRP
AtTRM1B	-	-	PELKRQK, KRKQEHEAMLSKRARS

Table 4: Bio-informatic prediction using the Localizer website (http://localizer.csiro.au/) of the sub-cellular localisation of AtTRM1A (AT3g02320**)** and AtTRM1B (AT5G15810). Amino acid predicted to be Nuclear Localization Signals (NLS) are indicated.

r) Reverse – Co-IP using a transient expression of TRM1A in N. benthamiana

To further investigate the interaction of AtPRORP2 with AtTRMI1A and B, the construct containing the CDS of AtTRM1A cloned into pEarleyGate102 (with the gene of interest under the control of a 35S promoter and with a C-ter CFP-HA tag) used for sub-cellular localization of the protein was used for reverse co-immunoprecipitation. This reverse-co-IP was performed after transitory expression in *N.benthamiana*. Agro-infiltrated leaves were collected 48 hours after infection and Co-IP was performed as previously with Milteny anti-HA beads.

Results show that the homolog of Arabidopsis PRORP2 in *N. benthamiana* is detected in the proteins that co-precipitate with AtTRMIA-HA (Table 3). The number of spectra is however quite low. But this is quite classical in reverse-IP performed in an heterologous system.

This result confirms that AtPRORP2 and AtTRM1A do interact *in vivo*, very likely in the first steps of tRNAs maturation.

Interestingly, the results also showed that RNase Z from *N.benthamiana* was very enriched in the proteins detected in the AtTRM1A elution after Co-IP. However, Arabidopsis closest homolog of the two *N.benthamiana* RNaseZ enriched in this IP, was found to be a plastid localized RNase Z (AtTRZ2 - AT2G04530). As AtPRORP2 and AtTRM1A and B are nuclear proteins we first though that the detected RNase Z proteins were contaminant proteins. However, the Localizer website predicts both a chloroplastic localisation signal and a nuclear localization signal (Table 4) for these proteins. Consequently, these two proteins might be addressed to both cellular compartments. To decide if AtRNAse Z was indeed interacting with AtTRM1A/B we decided to use the same strategy and expressed the nuclear AtRNase Z3 transiently in *N. benthamiana*.

To perform these experiments, the coding sequence of AtPRORP2, AtTRZ3 and AtTRM1B was cloned into pEarley 100, 202 and 203 respectively. And Co-IP will then be again performed using *N. Benthamiana* as a transient expression system.

Accesssion	Gene name	Ctrl	#1	#2	#3
AT3G02320.1	AtTRM1A	0	430	1753	1680
Niben101Scf05698g01017.1	PRORP2 homolog	0	2	4	1
Niben101Scf02732g00013.1	TRM1	0	70	480	480
Niben101Scf05618g04003.1	TRM1	0	70	477	481
Niben101Scf18268g00003.1	TRM1	0	36	226	220
Niben101Scf07152g04036.1	RNaseZ	0	15	45	69
Niben101Scf10557g00002.1	RNaseZ	0	14	36	57

Table 5: number of spectra obtained in separate co-IP assays after transient expression of TRM1A-CFP-HA in *Nicotiana benthamiana*. Results are shown only for AtPRORP2, AtTRM1 and AtTRZ homolog in *N. benthamiana* detected in these experiments. As no spectra was detected in the control, the result is shown in a single column, for simplicity.

	Chloroplast	Mitochondria	Nucleus
NbRNase_Z4036	Yes	-	KKQYIHLQGKQIEKLKKS
NbRNaseZ_0002	Yes	-	KKQYIHLQGKQIEKLKKS

Table 6: predicted subcellular localization of Niben101 Scf07152g04036.1 and Niben101 Scf10557 g00002.1 proteins. These two RNase Z proteins are predicted to be dual localized to the chloroplast and the nucleus by the localizer webserver. Amino acid sequence forming a predicted NLS is indicated.

At last, it is worth noting that *Nicotiana benthamiana* guanine (26)-N(2))-dimethyltransferase (TRM1) was also found enriched in the AtTRM1A-HA sample. This might be an artefact due to the over-expression of AtTRM1A but it might also be that TRM1A and B from a multimer *in vivo*. Furthermore, a phosphorylation site was detected in all experiments on a Serine residue (number 478 and 503) on AtTRM1A. However, this site was mostly detected unphosporylated in the MS/MS experiments.

s) Investigation of the effect of trm1A and B knock out in Arabidopsis thaliana

To gain more insights on the biological functions of AtTRM1A and B we decided to investigate the effect of TRM1A and B depletion using available T-DNA lines. An interesting question was to know if AtTRM1A and B are essential genes and if *Attrm1a* and *Attrm1b* single or double knockout plant lines have a visible phenotype.Three T-DNA lines were ordered and verified. After checking by PCR that the lines effectively contained a T-DNA inserted in the TRM1A/B gene, the exact T-DNA insertion site was mapped by sequencing. The Salk_018512 T-DNA line contains a T-DNA insertion in the 10th intron of the AtTRM1A gene. The Salk_206384 line contains a T-DNA insertion in the 9th exon of the AtTRM1B gene and the WiscDsLox244G04 line contains a T-DNA insertion in the 9th intron of the AtTRM1B gene.

Knock-out of the desired gene was verified by RT-PCR using forward and reverse primers designed to amplify the full length coding sequence (CDS). RT-PCR results show that these lines can effectively be considered as *Attrm1a* and *Attrm1b* knockout lines (Figure 55 and 56) .

It is interesting to note that *trm1a* and *trm1b* knock out plants did not exhibit any obvious phenotype.

The Salk_018512 line was crossed with the WiscDsLox244G04 line and with the Salk_206384 to generate *trm1a – trm1b* double mutants. Genotyping of the F2 generation is ongoing for the WiscDsLox244G04 x Salk_018512. But no double mutants could be identified so far (n=25). Concerning the cross between the Salk_018512 and Salk_206384 lines, the F2 generation will soon be genotyped.

Figure 55: RT-PCR on AtTRM1A gene using primers at the 5' and 3' end of the coding sequence (CDS). The result confirms that TRM1A full-length transcript is not expressed in the Salk_0185212 line.

Figure 56: RT-PCR on AtTRM1B gene using primers at the 5' and 3' end of the coding sequence (CDS). The result confirms that TRM1B full-length transcript is not expressed in the Salk_26384 and WiscdsLox244GO4 line.

Discussion

MAF1, the main negative regulator of RNA polymerase III, contains in plants, a tRNA like Structure (TLS) in its genetic sequence conserved in all the *Streptophyta* genome available to date. Since *Streptophyta* and *Chlorophyta* are considered to have splitted in two evolutionary lineages from the *Viridiplantae* between 725 and 1200 million years ago (Becker and Marin, 2009), we hypothesized that this non coding RNA sequence should play an important role in plant biology that would explain its evolutionary conservation. Because RNase P is an enzyme known to cleave both precursor tRNAs and TLS in vivo (Gobert et al.,2010 and Gutmann et al. 2012) we decided to investigate if MAF1 evolutionary conserved TLS could be, in *Arabidopsis*, a new substrate of the nuclear RNase P enzymes PRORP2/3.

We proved by *in vitro* cleavage assays that AtPRORP2 can cleave AtMAF1 intronic TLS, at its expected start site, as evidenced by our cRT-PCR results. Moreover, we investigated the effect of AtPRORP2/3 down regulation on the level of AtMAF1 at the transcriptomic level and showed that depletion of At*prorp2/3* leads to an accumulation of AtMAF1 transcripts. At last, we showed that *in vivo*, AtMAF1 transcripts termini do accumulate at the exact start site of the TLS and that these specific transcripts have additional nucleotides added in 3', which suggests that they will be degraded.

Possible alternative functions for MAF1 TLS

Our main hypothesis proposes that the conserved MAF1 TLS serves as a signal to cleave MAF1 mRNA and thus to control its level. Still, *ab initio*, alternative hypotheses could be envisaged to explain the function of MAF1 TLS. In a first hypothesis, RNase P cleavage of MAF1 TLS could be followed by RNase Z cleavage to generate a functional non-coding RNA. Alternatively, the observed cleavage of MAF1 TLS could have been due to transcription elongation factors as described for *Arabidopsis* in Antosz et al. 2020, if RNA polymerase II had had a tendency to stall at the specific TLS locus. Finally, it could also be hypothesized that, since AtMAF1 TLS is located in an intron, this conserved structure

would in fact be related to splicing (or alternative splicing). However, the results of investigations presented in this thesis do not provide evidences that would allow to support one of these alternative hypotheses.

Indeed, our investigation on plant knocked out for A*tmaf1* and complemented by MAF1 gene without the TLS sequence or without the intron containing AtMAF1 TLS, showed that AtMAF1 mRNA is spliced correctly because our RT-PCR experiment performed on RNA extracted from these transgenic lines showed that the RT-PCR products obtained had the same size as the WT (Col-0). Moreover, AtMAF1 TLS appears to be the only intronic ncRNA conserved widely in land plants (Akkuratov et al., 2014). If the intronic TLS function had been related to splicing, it would have been likely that similar structures would occur in more than one of the 25 000 genes encoded in the *Arabidopsis* genome, many of which having multiple introns.

Also, the available Chip-Seq data obtained with antibodies against the C-terminal domain of Arabidopsis RNA polymerase II, that can be considered as representing RNA Pol. II occupancy *in vivo*, do not show any obvious site where the polymerase would be stalling on the MAF1 gene. Indeed, the distribution of the reads in the data from Antosz et al. 2020 is homogeneous at the AtMAF1 locus (At5g13240), indicating that RNA polymerase II stalling at the TLS site as a regulation of AtMAF1 level of expression is very unlikely. Consequently, it is thus unlikely that the cleaved transcripts detected in our 3'RACE assays (or in the publicly available Arabidopsis PARE-seq Data) are generated by RNA polymerase II rather than by cleavage by AtPRORP2/3.

Could RNase Z be involved in the production of a novel non coding RNA?

As mentioned above, an alternative hypothesis could be that after cleavage by RNase P, AtMAF1 TLS is also cleaved by RNase Z (the enzyme catalysing the removal of the 3' trailer sequence from tRNA) to generate a functional small RNA. This hypothesis is certainly the most interesting alternative hypothesis to the one where cleavage by RNase P only leads to AtMAF1 pre-mRNA degradation and as such would be a post-transcriptional down-regulation mechanism to fine tune MAF1 expression in land plants.

It must be stated first that the occurrence of a cleavage in 3' of AtMAF1 TLS by RNase Z is not necessarily contradictory with our main hypothesis because the cleavage by both

RNase P and RNase Z to generate a new functional ncRNA could at the same time also lead to impaired splicing and the degradation of AtMAF1 pre-mRNA. Similar cases of cleavage of a TLS to generate a non-coding RNA have already been described in the literature. The most famous examples in Eukaryotes are those of the tRNA like ncRNAs mascRNA and Men β. These two ncRNA are processed from, respectively the lncRNA MALAT1 and MEN β. According to the published articles, MEN β tRNA like small RNA has a CCACCA motif added at its 3' end that will serve as a signal for its rapid degradation. This would explain why this ncRNA is undetectable by a Northern blot approach (Sunwoo et al, 2009; Wilusz et al, 2011; Kuhn et al,2015). In contrast, the mascRNA generated from the MALAT1 lncRNA is an interesting case of a functional tRNA-like non coding RNA as it has been shown to undergo CCA addition (Wilusz et a., 2008) and very recently, to promote global protein translation and cell proliferation by positively regulating the protein level of glutaminyl-tRNA synthetase (QARS) (Lu et al.,2020).

However, even if the hypothesis of a cleavage by RNase Z following RNase P cleavage (as our *in vitro* cleavage assays and the current state of knowledge indicate that this order is the most likely) is interesting and cannot yet be ruled out, we did not find any experimental evidence to support it. Indeed, there is for the moment, no proof that AtMAF1 clevage by RNase P leads to the creation of a new small non-coding RNA (sncRNA) in Arabidopsis *in vivo*. In particular, our *in vitro* assays shows that AtRNase Z3 (TRZ3) likely cleaves only after RNase P cleavage, but we could not identify specific AtRNaseZ3 cleavages in our 3'Race assays nor in the publicly available PARE-seq data. Furthermore, no small RNA corresponding to the TLS sequence could be identified by Northern-blot by Akkuratov et al. or in the different public databases of small RNA that we analysed. Nevertheless, absence of proof is definitively not a proof of absence. Consequently, we will need to further investigate this possibility of an RNase Z cleavage *in vivo*. One way to investigate this question is to perform comprehensive 5'RACE analyses. The 5'RACE that I performed at the beginning of my PhD identified a few sequences that could correspond to a cleavage of AtMAF1 TLS by RNase Z, but as explained above in the Results section, this 5'RACE assay was performed with an RNA adapter that did not allow to differentiate between PCR duplicates and unique RNA reads. Consequently, performing 5'RACE with an RNA adapter containing a random region to identify unique reads from PCR duplicates seems to be the most adequate method to investigate thoroughly if an RNase Z cleavage of AtMAF1 TLS does occur *in vivo* together

with the cleavage by RNase P. At this stage we thus cannot rule out that MAF1 TLS migth be released from its intron and that it could accumulate even at low levels for a yet unidentified function.

Indirect effect of MAF1 cleavage on RNA pol III activity

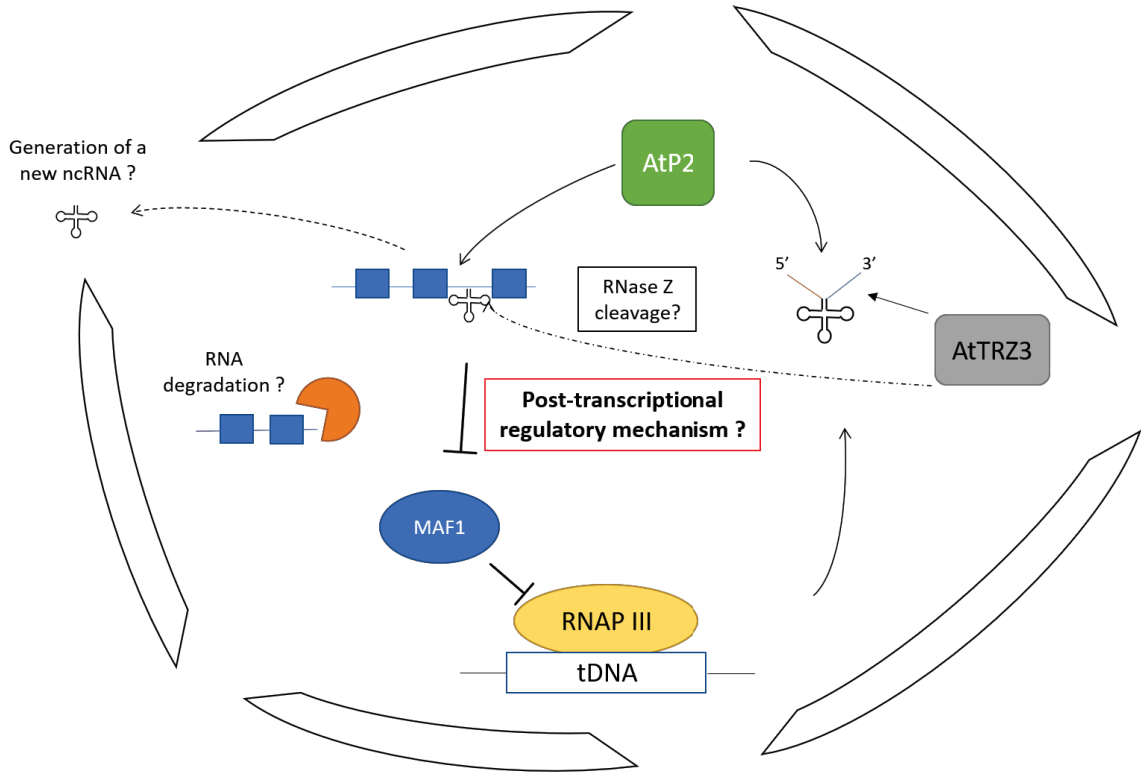
In our model, where PRORP down-regulation results in MAF1 increased level, we predicted that the transcription of all RNAs transcribed by RNA pol. III could be impaired. However, for now on, we could not detect a significant decrease of the levels of RNA pol. III transcripts such as U6 snRNA or 7SL RNAs while previous work had shown that the level of mature nuclear tRNA is decreased in *prorp3* KO mutants where PRORP2 is downregulated (Gutmann et al. 2012). This effect on tRNA levels could be directly due to a decrease in precursor-tRNA maturation (due to the lower amount of RNase P enzyme) and / or to our proposed regulation loop. In all cases, RNA pol. III transcripts levels are not all affected by PRORP down-regulation since 5S rRNA levels variations were never observed, neither in the present study, nor in previously published work (Gutmann et al. 2012). Nonetheless, the absence of RNA steady state level variations, as observed here for U6 snRNA or 7SL RNAs does not prove that RNA pol. III transcription was not regulated. It is imaginable that the turnover or degradation rate of specific transcripts might be regulated as well if RNA pol. III transcription is decreased, i.e. that transcripts such as 5S rRNAs have a longer half-life when their transcription is reduced.

Nonetheless, at this stage, the direct or indirect effects of PRORP down-regulation on RNA polymerase III activity remain to be fully understood. Up to now, we have used virus induced gene silencing (VIGS) to down regulate PRORP enzymes. The effects of this down regulation has been monitored by RT-qPCR (Figure 5 and 10). We also initiated a transcriptome wide RNA-Seq analysis of *prorp3* KO mutants where PRORP2 is downregulated. Preliminary results (not shown here) obtained with three biological replicates do not find statistically significative differences between the down-regulation mutants and the controls. This appears to be due the variability between individual VIGS experiments. Likewise minor transcript level differences might be missed because of the inherent variability of VIGS experiments. In this light, in order to investigate the important question of RNA polymerase III activity during PRORP down-regulation, and

consequently, the effect of the proposed regulation loop of RNA polymerase III activity via RNase P cleavage of AtMAF1 evolutionary conserved TLS, different alternative experimental procedures should be considered. For example, we could set up a different method to investigate the effect of nuclear RNase P depletion in Arabidopsis that would be more homogeneous, for example using the expression of artificial miRNA or based on CRISPR-Cas systems, e.g. using Cas13 that targets RNA (Terns et al., 2018). Alternatively, a complementary approach could be to create reporter lines for RNA polymerase III activity. Using these lines, we could then measure, in a more direct manner, the activity of RNA polymerase III in Arabidopsis and if it is affected or not by nuclear RNase P (*Atprorp2/3*) downregulation. However, such tools are not yet available for *in vivo* assays. For instance, fluorogenic dye to image RNAs have been developed, as described e.g. in a recent article (Bouhedda et al., 2019), but this technique developed is not appropriate for *in vivo* assays. The development of specific tools in the laboratory will this be required to directly monitor RNA pol. III transcription.

To conclude on this part it is worth noting that MAF1 is an important gene, regulating the activity of RNA polymerase III and conserved in all Eukaryotes. *maf1* KO in yeast leads to an accumulation of tRNAs in the cells and a slower growth (Pluta et al., 2001). However, depletion of MAF1 in animal cells affects cell morphology and body size (Johnson et al., 2007; Shor et al., 2010; Rideout et al., 2012) in contrasted ways. Moreover, since RNA pol. III synthetize ncRNA crucial for cell activity, its upregulation is currently increasingly considered as linked to oncogenic transformation in mammal cells (Grewal, S. S. 2015, Gouge, J. et al. 2015, Palian, B. M. et al. 2014). This example highlights the need for RNA pol. III regulation, that is achieved by MAF1 in case of stress or limitation in nutrient availability (Willis, I. M. & Moir, R. D., 2018). The fundamental mechanism of MAF1 action appears to be conserved in Eukaryotes. Usually phosphorylated, MAF1 has to be dephosphorylated to be in its active form. It will then localize to the nucleus and bind RNA pol III, preventing the formation of the pre-initiation complex (Vörlander et al., 2020). In plants, previous studies (Soprano et al. 2013 and 2017; Ahn et al.2018) have shown that this mechanism of MAF1 regulation by phosphorylation/dephosphorylation is conserved. Consequently our study sheds a new light on a possible additional regulation of MAF1 expression in plants. In our model, Arabidopsis nuclear RNase P (AtPRORP2/3) can cleave AtMAF1 precursor mRNA at an evolutionary conserved ncRNA site, forming a tRNA like structure. This cleavage would lead to the degradation of AtMAF1 precursor mRNA and

consequently to a lower accumulation of AtMAF1 transcripts. Since MAF1 main regulation pathway by phosphorylation/dephosphorylation is conserved in plants, our proposed model describes an additional layer of regulation of MAF1 level that would allow a fine tuning of RNA polymerase III activity.



<u>Figure 57</u>: Proposed model of AtMAF1 TLS cleavage by AtPRORP2 and its possible effect(s) at the molecular level. RNAP III stands for RNA polymerase III. tDNA represents tRNA genes. AtP2 and AtTRZ3 are Arabidopsis nuclear RNase P and RNase Z, respectively. Question marks indicate open questions that still need to be investigated. All these steps occur in the nucleus.

Function of MAF1 in stress response

In order to better understand the biological function of Arabidopsis MAF1 at the whole plant level and to understand the biological significance of the regulation pathways involving MAF1, it would also be important to investigate how MAF1 is regulated in response to stress. This might give clues to understand in which conditions a fine tuning of RNA pol. III activity involving MAF1 levels might be required. For this, we could monitor the effects a series of stresses on plants knocked-out for AtMAF1. During my PhD work, preliminary experiments have been performed using an Arabidopsis *maf1* KO strain (Gabi-kat_696C04). Experiments were only performed once and are thus very preliminary. Nonetheless, I observed that AtMAF1 does not seem to be involved in increasing the plant fitness when facing an abiotic stress. Indeed, in all the conditions tested (drought stress and heat stress), we could not detect any obvious difference between *maf1 ko* lines and WT (Col-0) lines. On the contrary, interestingly, in all the biotic stress assays that we performed, with TUMV or *P. syringae* infections, the WT lines always outperformed the *Atmaf*1 KO lines. Indeed, more pathogens were found in the infected *maf1* KO lines when compared to the WT (Col-0). Likewise, in WT plants, the RNA expression level of AtMAF1 was different between infected and non-infected plants. Altogether, this might point out that MAF1 regulation might be related to biotic stress response. Of course, these very preliminary results will need to be repeated to assess in depth the biological significance of MAF1 regulation pathway.

Altogether, functional data on MAF1 stress response should help to understand the importance of this regulatory protein conserved throughout Eukaryotes.

Determination of a model for the early steps of plant nuclear pre-tRNA maturation

For the second research project of my PhD work, I investigated the protein partners of AtPRORP2. For this, AtPRORP2 lines tagged in C-terminal with a unique HA tag were generated in a *prorp2/prorp3* knockout background. Indeed, even if PRORP proteins were shown to catalyse RNase P activity *in vitro* alone, it cannot be ruled out that *in vivo*, PRORP proteins do, in Arabidopsis, associate with one or more accessory proteins to perform RNase P activity, or to perform additional biological functions.

For instance, previous work conducted in the laboratory had focused on the organellar AtPRORP1 (At2g32230). Results showed that in mitochondria, AtPRORP1 is likely involved in interactions with other proteins such as the nuclease MNU2 (At5g09840), the mitochondrial PNPase (At5g14580) or mTERF30 (At1g61980) and in chloroplast with RH3 (At5g26742) (Bouchoucha et al., 2019) . However, only the interaction with AtMNU2 seems to be direct, as evidenced by Yeast 2 hybrids assays. For the nuclear AtPRORP2, we hypothesised that it could interact with other proteins involved in gene expression or, alternatively, that it could constitute a tRNA maturation complex with other tRNA maturating enzyme, like the nuclear RNase Z (AtTRZ1 or AtTRZ3 according to Canino et al., 2009) or with the LA protein (AtLA1, Fleurdépine et al., 2007) as the latter protein is known to bind precursors tRNAs directly after RNA pol. III transcription and because RNase P was shown to catalyse 5' leader removal in the early steps of tRNA maturation (Hooper, 2013).

Our result showed that contrary to organelles where AtPRORP1 likely interacts with proteins involved in gene expression processes that are apparently not related to tRNA biogenesis, the nuclear AtPRORP2 appears to only make interactions with proteins involved in tRNA maturation, i.e. with the tRNA methyltransferases, TRM1A and B.

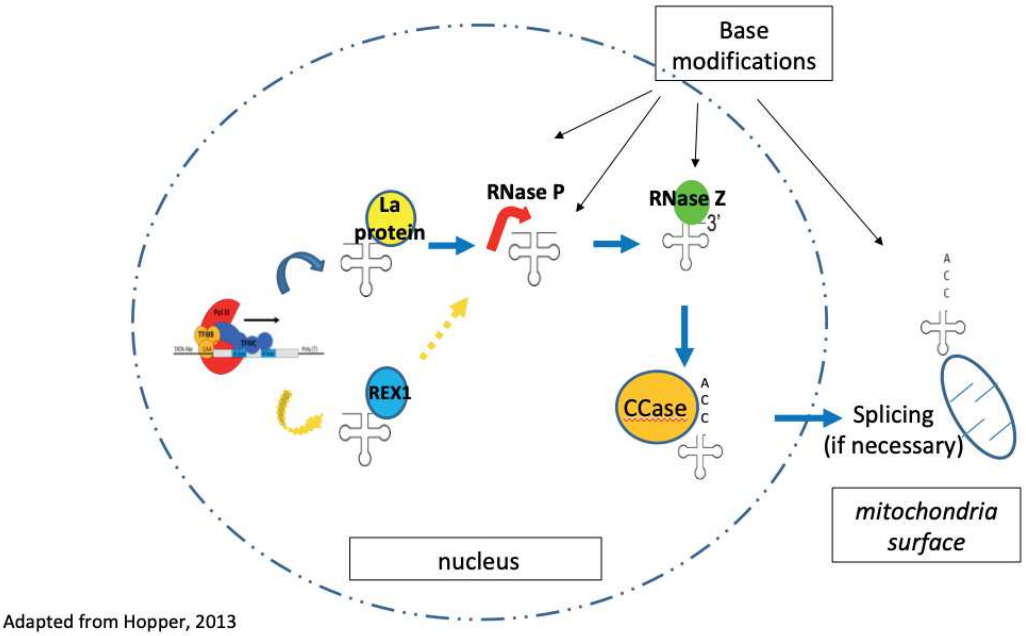
TRM1A and B catalyse the addition of two methyl groups on a very conserved guanine residue located at base 26 on most tRNAs. This base modification is considered as most important for tRNA 3D structure (Lorenz et al., 2017). We confirmed by reverse co-immunoprecipitation that this interaction is indeed occurring *in vivo*, as the homolog of AtPRORP2 in *Nicotiana benthamiana* could be found in co-ip experiments after transient expression of AtTRM1A in *N. benthamiana*.

We tried to further investigate if this interaction between AtPRORP2 and AtTRM1A and B is a direct protein-protein interaction or if it is mediated by tRNAs. For this we performed our co-ip assays with different nucleases and using different protocols. But in our hands none proved efficient to answer our questions, as the number of proteins of interest (our bait PRORP2-HA and its interactants TRM1A and B) were always much lower in these co-ip assays. Consequently, we cannot yet decide if the interaction between AtPRORP2 and AtTRM1A and B is direct or indirect. To further investigate this question, we think the most adequate method will be to use Yeast two hybrid assays. Alternatively, we could also use in planta interaction methods such as FLIM FRET.

Finally, in our reverse co-ip experiments (performed with AtTRM1A tagged with HA and transiently expressed in *N. Benthamiana*) we also found the homolog of Arabidopsis RNase Z. As in *N. benthamiana*, the protein interactant of AtTRM1A is predicted (by bioinformatics) to be dual localized to both chloroplast and nucleus, we decided to further investigate if AtTRM1A and AtTRZ3 can interact together and possibly form a single complex with AtPRORP2.

For this, the coding sequence of AtPROPR2 (with a unique HA tag in C-terminal), AtTRZ3 (with a N-terminal myc tag) and AtTRM1A (with a N-terminal flag tag) were cloned under the control of a strong promoter (35S) in respectively, pEarleyGate 100, 202 and 203.

These constructs will be used to perform transient expression co-immunoprecipitation assays in *N. Benthamina* to test if AtTRM1A can interact with both AtPRORP2 and AtTRZ3 and also if AtPRORP2 and AtTRZ3 can interact indirectly. The latter interaction is unlikely as no spectra for AtTRZ3 was never obtained in our co-immunoprecipitation assays performed with Arabidopsis stable PRORP2-HA lines). Altogether, results already obtained suggest that PRORP2, RNase Z and the methyl transferases TRM1A and B may act in a complex, or in channelling (the reaction catalysed by one enzyme being immediately followed by the activity of the next enzyme) to perform the early steps of nuclear pre-tRNA maturation.



<u>Figure 58</u>: Schematic of the current model for the tRNA maturation pathway in yeast. This model was created using the information presented in Hooper, 2013

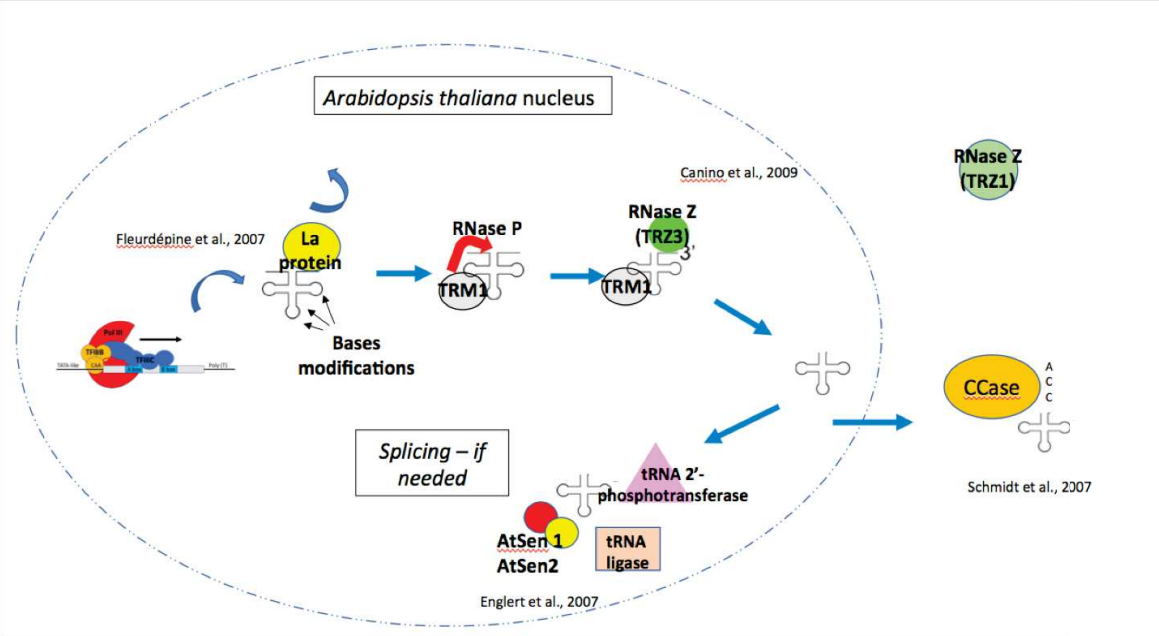


Figure 59: Proposed model for the tRNA maturation pathway in *Arabidopsis thaliana*. Enzyme subcellular localisation is based either on published work (name of first author and year of publication are indicated) or on our results (see text for details)

Discussion

Version Française

MAF1, le principal régulateur négatif de l'ARN polymérase III, contient chez les plantes, une structure ressemblant à un ARN de transfert, appelée en anglais « tRNA-like structure » (TLS). Cette structure est conservée dans sa séquence génétique dans tous les génomes séquencés à ce jour dans le clade des *Streptophyta*. Etant donné que les *Streptophyta* et les *Chlorophyta* sont considérés comme ayant évolués en deux lignées évolutives distinctes des Viridiplantae entre 725 et 1200 millions d'années (Becker et Marin, 2009), nous avons émis l'hypothèse que cette séquence d'ARN non codant devait jouer un rôle important dans la biologie végétale qui expliquerait sa conservation évolutive. Étant donné que la RNase P est une enzyme connue pour cliver à la fois les ARNt précurseurs et les « TLS » *in vivo* (Gobert et al., 2010 et Gutmann et al., 2012), nous avons décidé d'étudier si ce TLS évolutivement conservé dans le génome de MAF1 pouvait être, chez *Arabidopsis*, un nouveau substrat des enzymes RNase P nucléaires PRORP2 et PRORP3.

Nous avons prouvé par des tests de clivage *in vitro* qu'AtPRORP2 peut cliver le TLS de AtMAF1, à son site de départ attendu, comme en témoignent nos résultats de cRT-PCR. De plus, nous avons étudié l'effet de la dérégulation d'AtPRORP2 / 3 sur le niveau d'AtMAF1 au niveau transcriptomique et montré que la dérégulation d'AtPRORP2 / 3 conduit à une accumulation des transcrits AtMAF1. Enfin, nous avons montré qu'*in vivo*, les transcrits AtMAF1 se terminant au site de départ exact du TLS s’accumulent et que ces transcrits spécifiques ont des nucléotides supplémentaires ajoutés en 3 ', ce qui suggère qu'ils seront dégradés.

Discussion relatives aux hypothèses alternatives permettant d'expliquer nos résultats :

Notre hypothèse principale propose que le TLS de AtMAF1 évolutivement conservé sert de signal pour cliver l'ARNm de MAF1 et ainsi contrôler son niveau d'expression. Pourtant, *ab initio*, des hypothèses alternatives pourraient être envisagées pour expliquer la fonction du TLS de MAF1.

Selon une première hypothèse, le clivage RNase P du TLS de MAF1 pourrait être suivi d'un clivage par la RNase Z pour générer un ARN non codant fonctionnel. Une autre hypothése est que le clivage observé du TLS de MAF1 pourrait être dû à des facteurs d'élongation de la transcription comme décrit pour *Arabidopsis* dans l'article d'Antosz et al. 2020, si l'ARN polymérase II avait eu tendance à se bloquer au locus spécifique où se trouve ce TLS. Enfin, on pourrait également émettre l'hypothèse que, puisque le TLS de AtMAF1 est situé dans un intron, cette structure conservée serait en fait liée à l'épissage (ou à l'épissage alternatif).

Cependant, les résultats des recherches présentées dans cette thèse ne fournissent pas de preuves qui permettraient de soutenir l'une de ces hypothèses alternatives.

En effet, nos expériences sur des lignées n'exprimant plus le gène *Atmaf1* et complémentée par le gène MAF1 sans la séquence TLS ou sans l'intron contenant le TLS de AtMAF1, ont montrées que l'ARNm d'AtMAF1 est épissé correctement. En effet, nos expériences de RT-PCR réalisées sur l'ARN extrait de ces lignées transgéniques a montré que les produits de RT-PCR obtenus avaient la même taille que les plantes dites « sauvages » (Col-O). De plus, le TLS de AtMAF1 semble être le seul ARN non codant situé dans un intron largement conservé dans les plantes terrestres (Akkuratov et al., 2014). Ainsi, si la

fonction du TLS avait été liée à l'épissage, il aurait été probable que des structures similaires se produisent dans plus d'un des 25 000 gènes codés dans le génome d'*Arabidopsis*, dont beaucoup ont plusieurs introns.

En outre, les données de Chip-Seq disponibles obtenues avec des anticorps contre le domaine C-terminal de l'ARN polymérase II d'*Arabidopsis*, qui peuvent être considérées comme représentant l'occupation, *in vivo*, de la chromatine par l'ARN Pol. II, ne montrent aucun site évident où la Polymérase ARN II se bloquerait sur le gène de MAF1. En effet, la distribution des séquences obtenues dans les données d'Antosz et al. 2020 est homogène au locus d'AtMAF1 (At5g13240), ce qui indique que le blocage de l'ARN polymérase II au site du TLS en tant que mécanisme de régulation du niveau d'expression d'AtMAF1 est très improbable. Par conséquent, il est donc peu probable que les transcrits clivés détectés dans nos expériences de 3'RACE (ou dans les données de PARE-seq réalisées sur Arabidopsis et accessibles à tous en ligne) soient générés par l'ARN polymérase II plutôt que par clivage par AtPRORP2 / 3.

Discussion relative à un clivage par la RNase Z du TLS de AtMAF1 :

Comme mentionné ci-dessus, une autre hypothèse pourrait être qu'après le clivage par la RNase P, AtMAF1 TLS est également clivé par la RNase Z (l'enzyme catalysant l'élimination de la séquence additionnelle en 3 ' des précurseurs d'ARN de transfert (ARNt)) pour générer un petit ARN fonctionnel. Cette hypothèse est certainement l'hypothèse alternative la plus intéressante à celle où le clivage par la RNase P ne conduit qu'à une dégradation pré-ARNm AtMAF1 et en tant que tel serait un mécanisme de régulation post-transcriptionnel pour affiner l'expression de MAF1 dans les plantes terrestres.

Il faut tout d'abord préciser que l'apparition d'un clivage en 3 ' du TLS de AtMAF1 par la RNase Z n'est pas forcément contradictoire avec notre hypothèse principale car le clivage à la fois par RNase P et RNase Z pour générer un nouvel ARNnc fonctionnel pourrait en même temps également conduire à un épissage altéré et à la dégradation du pré-ARNm AtMAF1. Des cas similaires de clivage d'un TLS pour générer un ARN non codant ont déjà été décrits dans la littérature. Les exemples les plus connus chez les eucaryotes sont ceux des ARN ayant une structure ressemblant à un ARNt (TLS) tels que les ARNnc « mascRNA » et « Men β ». Ces deux ARN non codant sont générés à partir, respectivement, des ARN non codant long : MALAT1 et MEN β. Selon les articles publiés sur le sujet, le TLS de MEN β est un petit ARN possedant un motif CCACCA ajouté à son extrémité 3 ' qui servira de signal pour sa dégradation rapide. Cela expliquerait pourquoi cet ARNnc est indétectable par une approche Northern blot (Sunwoo et al, 2009; Wilusz et al, 2011; Kuhn et al, 2015). En revanche, le mascRNA généré à partir de l'ARNnc MALAT1 est un cas intéressant d'ARN non codant de type ARNt fonctionnel car il a été démontré

qu'il subit une addition du motif CCA (Wilusz et al., 2008) et très récemment, que sa fonction principale serait de promouvoir la traduction globale des protéines et de contrôler la prolifération cellulaire en régulant positivement le niveau protéique de la glutaminyl-ARNt synthétase (QARS) (Lu et al., 2020).

Cependant, même si l'hypothèse d'un clivage par la RNase Z suite au clivage par la RNase P (car nos tests de clivage *in vitro* et l'état actuel des connaissances indiquent que l'ordre le est le plus probable est celui d'un clivage par la RNase Z postérieur à celui par la RNase P) est intéressante et ne peut pas encore être écartée, nous n'avons pas trouvé des preuves expérimentales pour la soutenir. En effet, il n'y a pour l'instant aucune preuve que le clivage d'AtMAF1 par la RNase P conduise à la création d'un nouveau petit ARN non codant (sncRNA) chez *Arabidopsis*, *in vivo*. En particulier, nos tests *in vitro* montrent qu'AtRNase Z3 (TRZ3) ne clive probablement qu'après le clivage de la RNase P, mais nous n'avons pas pu identifier de clivages spécifiques d'AtRNaseZ3 dans nos expériences de 3' Race ni dans les données PARE-seq accessibles en ligne. De plus, aucun petit ARN correspondant à la séquence du TLS de MAF1 n'a pu être identifié par Northern-blot par Akkuratov et al. ou dans les différentes bases de données publiques de petits ARNs que nous avons analysés. Néanmoins, l'absence de preuve n'est définitivement pas une preuve d'absence. Par conséquent, nous devrons approfondir cette possibilité d'un clivage de la RNase Z *in vivo*. L'une des façons d'étudier cette question consiste à effectuer des analyses de 5'RACE. La 5'RACE que j'ai réalisée au début de ma thèse a identifié quelques séquences qui pourraient correspondre à un clivage du TLS d'AtMAF1 par la RNase Z, mais comme expliqué ci-dessus dans la section Résultats, ce test 5'RACE a été réalisé avec un adaptateur ARN qui ne permettait pas de différencier les doublons de PCR des séquences d'ARN

uniques. Par conséquent, la réalisation d'expérience de 5'RACE avec un adaptateur d'ARN contenant une région aléatoire pour identifier les séquences uniques et les différencier de doublons de PCR semble être la méthode la plus adéquate pour étudier de manière approfondie si un clivage par la RNase Z de AtMAF1 TLS se produit *in vivo* aprés le clivage par la RNase P. Ainsi, à ce stade, nous ne pouvons donc pas exclure que le TLS de AtMAF1 soit libéré de son intron et qu'il pourrait s'accumuler même à de faibles niveaux pour une fonction encore non identifiée.

Discussion relative à l'effet éventuel sur l'activité de la Polymérase à ARN III, du clivage du TLS de AtMAF1 par AtPRORP2 et 3

Dans notre modèle, où la régulation à la baisse de PRORP2 et 3 entraîne une augmentation du niveau de l'expression de AtMAF1, nous avons prédit que la transcription de tous les ARN transcrits par l'ARN pol. III pourrait être altéré. Cependant, pour l'instant, nous n'avons pas pu détecter une diminution significative des niveaux d'expression des transcrits de l'ARN polymérase III tels que « U6 snRNA » ou encore le « 7SL ARN » alors que des travaux antérieurs avaient montré que le niveau d'ARNt nucléaire mature est diminué chez les mutants n'exprimant plus AtPRORP3 (« *prorp3 KO* ») et où AtPRORP2 est régulé à la baisse (Gutmann et al.2012). Cet effet sur les niveaux d'ARNt pourrait être directement dû à une diminution de la maturation du précurseur des ARNt (en raison de la quantité inférieure d'enzyme RNase P) et / ou à notre boucle de régulation proposée dans la partie résultat de cette thése. Dans tous les cas, les niveaux de transcription des principaux ARN transcrits par l'ARN pol. III ne sont pas tous affectés par la régulation à la baisse des PRORP nucléaires d'*Arabidopsis* puisque les variations des niveaux d'ARNr 5S n'ont jamais été observées, ni dans la présente étude, ni dans des travaux publiés antérieurement (Gutmann et al. 2012). Néanmoins, l'absence de variations des niveau d'expression de ces ARN à l'état d'équilibre, comme observé ici pour U6 snRNA ou pour le 7SL RNA ne prouve pas que l'activité de transcription de l'ARN pol. III n'est pas dérégulée. Il est en effet imaginable que le taux de renouvellement ou de dégradation de transcrits spécifiques puisse également

être régulé quand l'activité de l'ARN pol. III est diminuée, c'est-à-dire que les transcrits tels que les ARNr 5S auraient ainsi une demi-vie plus longue lorsque leur transcription est réduite.

Néanmoins, à ce stade, les effets directs ou indirects de la régulation négative de PRORP sur l'activité de l'ARN polymérase III restent à comprendre. Jusqu'à présent, nous avons utilisé la méthode de la réduction de l'expression génique induite par l'utilisation d'un virus (VIGS) pour réguler à la baisse les enzymes nucléaires de PRORP chez *Arabidopsis*. Les effets de cette régulation à la baisse ont été surveillés par RT-qPCR. Nous avons également lancé une analyse RNA-Seq à l'échelle du transcriptome des mutants *prorp3 KO* où PRORP2 est régulé à la baisse. Les résultats préliminaires (non présentés ici) obtenus avec trois réplicats biologiques ne trouvent pas de différences statistiquement significatives entre les mutants à régulation négative et les témoins. Cela semble être dû à la variabilité entre les expériences VIGS individuelles. De même, des différences faibles du niveau de transcription de certains transcrits pourraient être manquées en raison de la variabilité inhérente aux expériences de VIGS. Dans cette optique, afin d'étudier la question importante de l'activité de l'ARN polymérase III au cours de la régulation négative des PRORP nucléaires d'*Arabidopsis*, et par conséquent, l'effet de la boucle de régulation proposée de l'activité de l'ARN polymérase III via le clivage par la RNase P du TLS conservé d'AtMAF1 ; différentes alternatives expérimentales doivent être envisagées. Par exemple, nous pourrions mettre en place une méthode différente pour étudier l'effet de l'appauvrissement de la RNase P nucléaire chez *Arabidopsis* qui serait plus homogène, par exemple en utilisant l'expression de miARN artificiel ou basée sur des systèmes CRISPR-Cas, par ex. en utilisant Cas13 qui cible l'ARN (Terns et al., 2018). Une approche complémentaire pourrait être de créer des lignées rapporteuses pour l'activité

de l'ARN polymérase III. En utilisant ces lignées, nous pourrions alors mesurer, de manière plus directe, l'activité de l'ARN polymérase III chez *Arabidopsis* et si elle est affectée ou non par la régulation à la baisse de la RNase P (Atprorp2 / 3) nucléaire. Cependant, ces outils ne sont pas encore disponibles pour réaliser des expériences *in vivo*. Par exemple, un colorant fluorogène pour visualiser des ARN a été développé, comme décrit dans un article récent (Bouhedda et al., 2019). Mais cette technique n'est pas adaptée aux dosages *in vivo*. Le développement d'outils spécifiques sera donc nécessaire pour suivre directement l'activité de transcription de l'ARN pol. III et l'effet éventuel de la dérégulation de l'expression de AtPROP2/3 sur cette dernière.

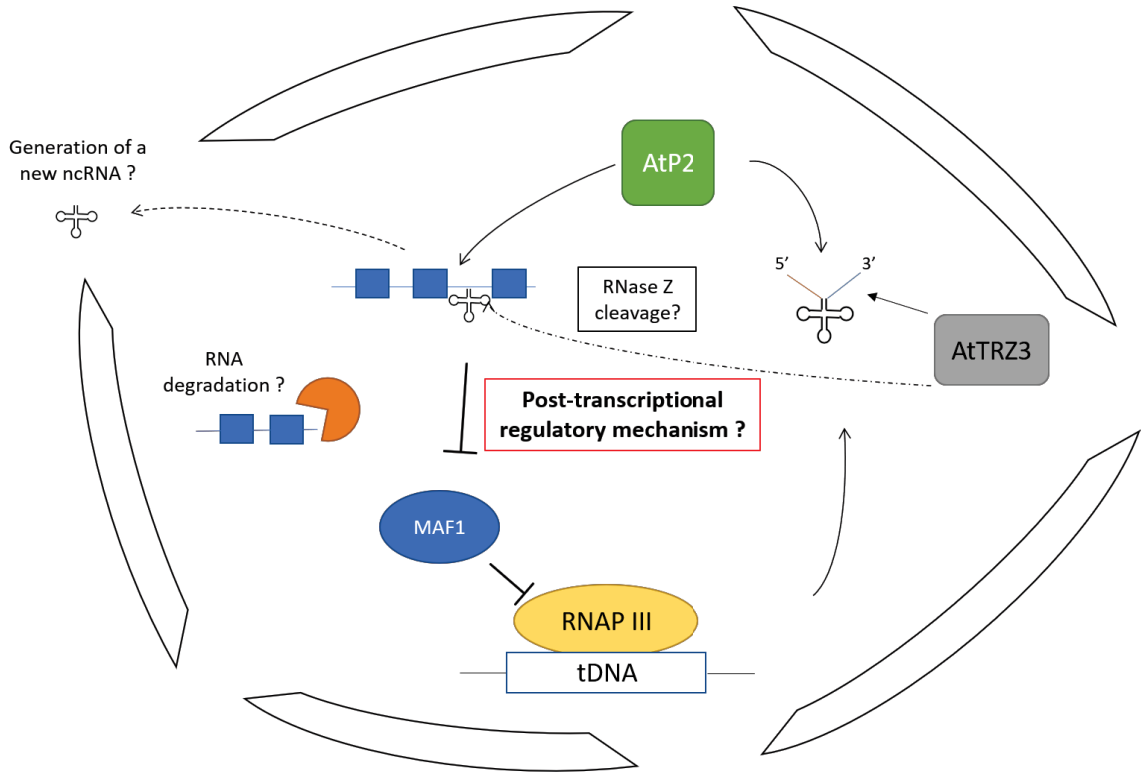
Pour conclure sur cette partie, il convient de noter que MAF1 est un gène important, régulant l'activité de l'ARN polymérase III, et conservé chez tous les Eucaryotes. L'absence d'expression de MAF1 dans la levure conduit à une accumulation d'ARN de transfert dans les cellules et à une croissance plus lente (Pluta et al., 2001). Cependant, l'absence d'expression de MAF1 dans les cellules animales affecte la morphologie cellulaire et la taille corporelle (Johnson et al., 2007; Shor et al., 2010; Rideout et al., 2012) de manière contrastée.

De plus, étant donné que l'ARN pol. III synthétise des ARN non codants cruciaux pour l'activité cellulaire, sa régulation positive est actuellement de plus en plus considérée comme liée à la transformation oncogène dans les cellules de mammifères (Grewal, S. S. 2015, Gouge, J. et al.2015, Palian, B. M. et al.2014). Cet exemple met en évidence la nécessaire régulation de l'ARN pol. III, *in vivo*. Cette régulation est réalisée par MAF1 en cas de stress ou de limitation de la disponibilité des nutriments (Willis, I. M. & Moir, R. D., 2018). Le mécanisme fondamental de l'action de MAF1 semble être conservé chez les Eucaryotes. Habituellement phosphorylé, MAF1 doit être dé-phosphorylé pour être sous sa forme active. Il se localisera ensuite au niveau du noyau et se liera à l'ARN pol III, empêchant la formation du complexe de pré-initiation (Vörlander et al., 2020). Chez les plantes, des études antérieures (Soprano et al.,2013 et 2017; Ahn et al.,2018) ont montrées que ce mécanisme de régulation de MAF1 par phosphorylation / déphosphorylation est conservé. Par conséquent, notre étude jette un éclairage nouveau sur une éventuelle régulation supplémentaire de l'expression de MAF1 chez les plantes. Dans notre modèle, la RNase P nucléaire d'Arabidopsis (AtPRORP2 / 3) peut cliver l'ARNm précurseur

d'AtMAF1 à un site spécifique situé sur l'ARN et conservé évolutivement, qui forme une structure semblable à un ARN de transfert (TLS).

Ce clivage conduirait à la dégradation de l'ARNm précurseur d'AtMAF1 et par conséquent à une accumulation plus faible des transcrits d'AtMAF1.

Étant donné que la voie de régulation principale de MAF1 par phosphorylation / déphosphorylation est conservée dans les plantes, le modèle que nous proposons décrit une couche supplémentaire de régulation du niveau de MAF1 qui permettrait un réglage fin de l'activité de l'ARN polymérase III.



<u>Figure 57</u>: Schéma présentant notre modèle de la possible régulation de l'expression de MAF1 par la RNase P nucléaire chez les plantes terrestres.

RNAP III désigne l'ARN polymerase III.

tDNA les gènes codant pour les ARN de transfert.

AtP2 and AtTRZ3 sont, respectivement, les gènes de la RNase P et de la RNase Z chez *Arabidopsis*.

Les points d'interrogation indiquent au lecteur les sujets sur lesquels des investigations complémentaires sont nécessaires.

Toutes les étapes indiquées dans ce schéma se déroulent dans le noyau.

Discussion relative aux fonctions biologiques de MAF1 chez les plantes terrestre, notamment en cas de stress

Afin de mieux comprendre la fonction biologique de MAF1 chez *Arabidopsis* au niveau de la plante entière et de comprendre la signification biologique des voies de régulation impliquant MAF1, il serait également important d'étudier comment MAF1 est régulée en réponse aux stress.

Cela pourrait en effet nous donner des indices pour comprendre dans quelles conditions un réglage fin de l'activité de transcription de l'ARN pol. III impliquant MAF1 pourrait être nécessaire.

Afin d'étudier cette question, nous avons pu suivre les effets d'une série de stress sur des lignées de plantes n'exprimant pas AtMAF1. Au cours de mon travail de thèse, des expériences préliminaires ont ainsi été réalisées en utilisant une lignée d'*Arabidopsis* n'exprimant pas MAF1 (« *maf1 KO* » (Gabi-kat_696C04)). Les expériences n'ont été effectuées qu'une seule fois et sont donc très préliminaires. Néanmoins, j'ai observé qu'AtMAF1 ne semble pas être impliqué dans la réponse des plantes face à un stress abiotique. En effet, dans toutes les conditions testées (sécheresse et stress thermique), nous n'avons pas pu déceler de différence évidente entre les lignées « maf1 ko » et les lignées sauvage (Col-O).

Au l'inverse, il est intéressant de noter que dans tous les tests de stress biotique que nous avons effectués, avec des infections avec le virus « TUMV » ou avec la bactérie « *P. syringae* », les lignées sauvages ont toujours surpassées les lignées « Atmaf1 KO ». Ainsi, plus d'agents pathogènes ont été trouvés dans les lignées « maf1 KO » infectées par rapport au sauvage (Col-O). De même,

dans les plantes sauvage, le niveau d'expression d'ARN de AtMAF1 était différent entre les plantes infectées et non infectées. Dans l'ensemble, cela pourrait indiquer que la régulation MAF1 pourrait être liée à la réponse au stress biotique. Bien entendu, ces résultats très préliminaires devront être répétés pour évaluer en profondeur la signification biologique de la voie de régulation de MAF1.

Dans l'ensemble, les données fonctionnelles sur la réponse au stress MAF1 devraient aider à comprendre l'importance de cette protéine régulatrice conservée chez tous les Eucaryotes.

Discussion relative aux partenaires protéiques éventuels des RNase P nucléaires chez *Arabidopsis*

Pour le deuxième projet de recherche de mon travail de doctorat, j'ai étudié les partenaires protéiques d'AtPRORP2. Pour cela, des lignées AtPRORP2 étiquetées en C-terminal avec une étiquette HA unique ont été générées dans un fond génétique n'exprimant pas PRORP2 et PRORP3 (« knockout prorp2 / prorp3 »). En effet, même s'il a été démontré que les protéines PRORP catalysaient l'activité RNase P *in vitro* par elle-même, il ne peut être exclu que *in vivo*, les protéines PRORP s'associent, chez *Arabidopsis*, à une ou plusieurs protéines accessoires pour réaliser leur activité RNase P, ou pour participer à d'autres fonctions biologiques.

Par exemple, des travaux antérieurs menés au laboratoire d'accueil de ma thése ont porté sur la RNase P mitochondriale et chloroplastique AtPRORP1 (At2g32230). Les résultats ont montré que dans les mitochondries, AtPRORP1 est probablement impliqué dans des interactions avec d'autres protéines telles que la nucléase MNU2 (At5g09840), la PNPase mitochondriale (At5g14580) ou mTERF30 (At1g61980) et dans le chloroplaste avec RH3 (At5g26742) (Bouchoucha et al., 2019). Cependant, seule l'interaction avec AtMNU2 semble être directe, comme en témoignent les expériences de double hybride réalisé dans la levure. Pour la protéine AtPRORP2 nucléaire, nous avons émis l'hypothèse qu'elle pourrait interagir avec d'autres protéines impliquées dans l'expression génique ou, alternativement, qu'elle pourrait constituer un complexe de maturation des ARN de transfert avec d'autres enzymes de maturation des ARN de transfert, comme la RNase Z nucléaire (AtTRZ1 ou

AtTRZ3 selon Canino et al. ., 2009) ou avec la protéine LA (AtLA1, Fleurdépine et al., 2007) car cette dernière protéine est connue pour se lier aux précurseurs d'ARNt directement après leur synthése par l'ARN pol. III et parce qu'il a été démontré que la RNase P catalyse l'élimination de l'extrémité additionnelle en 5' des pré-ARNt lors des premières étapes de la maturation des ARNt (Hooper, 2013).

Nos résultats montrent que contrairement aux organites où AtPRORP1 interagit vraisemblablement avec des protéines impliquées dans des processus d'expression génique qui ne sont apparemment pas liés à la biogenèse des ARNt, la RNase P nucléaire semble n'interagir qu'avec des protéines impliquées dans la maturation des ARNt : les méthyltransférases d'ARNt, TRM1A et B.

TRM1A et B catalysent l'addition de deux groupements méthyle sur un résidu guanine très conservé situé sur la base 26 sur la plupart des ARNt. Cette modification de base est considérée comme la plus importante pour la structure 3D de l'ARNt (Lorenz et al., 2017). Nous avons confirmé par co-immunoprécipitation inverse que cette interaction se produit bien *in vivo*, car l'homologue d'AtPRORP2 dans *Nicotiana benthamiana* a été trouvé dans des expériences de co-ip après expression transitoire d'AtTRM1A dans *N. benthamiana*.

Nous avons essayé d'étudier plus en détail si cette interaction entre AtPRORP2 et AtTRM1A et B est une interaction directe protéine-protéine ou si elle est médiée par les ARNt. Pour cela, nous avons effectué nos tests de co-immunoprécipitation avec différentes nucléases et en utilisant différents protocoles. Mais, aucune ne s'est avérée efficace pour répondre à nos questions, car le nombre de protéines d'intérêt (notre appât PRORP2-HA et ses interactants TRM1A et B) était toujours beaucoup plus faible dans ces expériences. Par conséquent, nous ne pouvons pas encore décider si

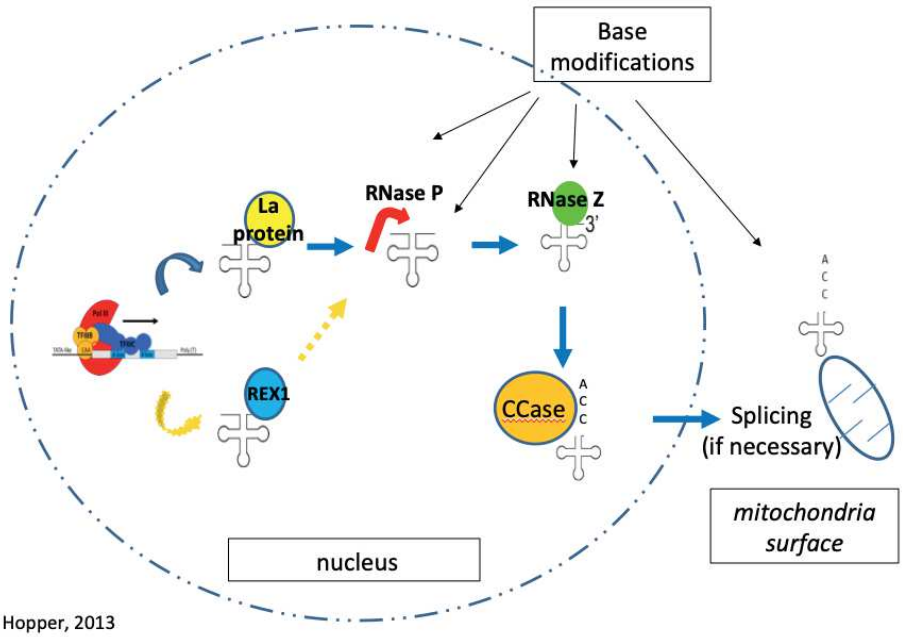
l'interaction entre AtPRORP2 et AtTRM1A et B est directe ou indirecte. Pour approfondir cette question, nous pensons que la méthode la plus adéquate serait d'utiliser des expériences de double hybride en levure. Alternativement, nous pourrions également utiliser des méthodes d'interaction *in planta* telles que la technique du FLIM FRET.

Enfin, dans nos expériences de co-immunoprécipitaiton inverse (réalisées avec la protéine AtTRM1A étiquetées avec une étiquette HA qui ont été exprimée de manière transitoire dans *N. Benthamiana*), nous avons également trouvé l'homologue de la RNase Z d'*Arabidopsis*. Etant donné que chez N. *benthamiana*, ce possible interactant de TRM1A est prédit (par bio-informatique) pour être localisé à la fois dans le chloroplaste et dans le noyau, nous avons décidé d'étudier plus finement si AtTRM1A et AtTRZ3 peuvent interagir ensemble et éventuellement former un seul complexe avec AtPRORP2.

Pour cela, la séquence codante d'AtPROPR2 (avec un tag HA unique en C-terminal), AtTRZ3 (avec un tag MYC situé en N-terminal de la protéine) et AtTRM1A (avec un tag FLAG en N-terminal) ont été clonées sous le contrôle d'un promoteur fort (35S) dans respectivement, pEarleyGate 100, 202 et 203.

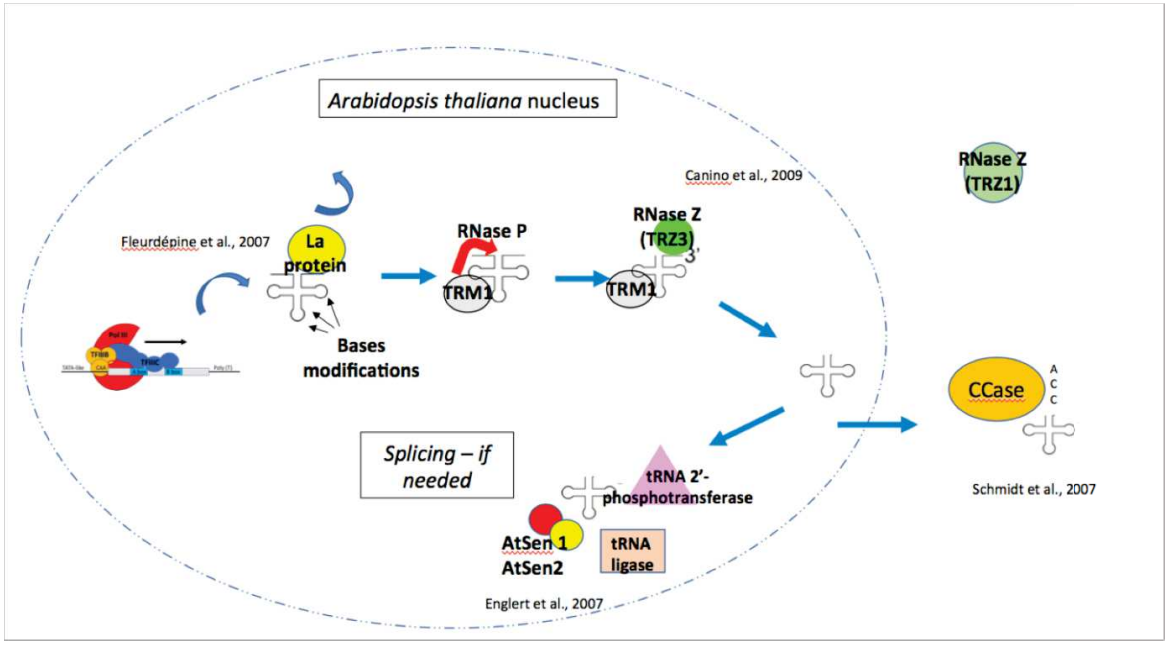
Ces constructions seront utilisées pour effectuer des tests de co-immunoprécipitation en expression transitoire chez *N. Benthamiana* pour tester si AtTRM1A peut interagir à la fois avec AtPRORP2 et AtTRZ3 et également si AtPRORP2 et AtTRZ3 peuvent interagir indirectement. Cette dernière interaction est peu probable car aucun spectre pour AtTRZ3 n'a jamais été obtenu dans nos expériences de co-immunoprécipitation réalisés avec des lignées PRORP2-HA chez *Arabidopsis*.

Au final, ces résultats suggèrent que PRORP2, la RNase Z et les méthyltransférases TRM1A et B peuvent agir dans un complexe, ou de manière coordonnée où la réaction catalysée par une enzyme est immédiatement suivie par l'activité de l'enzyme suivante (« channeling ») pour effectuer les premières étapes de la maturation nucléaire pré-ARNt.



Adapted from Hopper, 2013

Figure 58: Schéma représentant la voie de maturation des ARN de transfert nucléaires chez la levure. Ce modèle reprend les informations présentées dans l'article de Hooper, 2013.



<u>figure 59:</u> Schéma présentant notre modèle de maturation des ARN de transfert nucléaire chez *Arabidopsis thaliana*.

La localisation subcellulaire des enzymes est basée soit sur des travaux déjà publiés (le nom du premier auteur et l'année de publication est indiqué sur le schéma) soit sur nos propres résultats (voir le texte pour plus de détail).

Concluding remarks

Altogether, with this work, I contributed to the functional characterization of protein-only RNase P enzymes that were discovered a decade ago and have been investigated by my host team in plants ever since. Beyond their primary function for pre-tRNA maturation in all the compartments where gene expression takes place, i.e. nucleus, mitochondria and chloroplasts, growing evidence now suggest that they also participate in other gene expression processes, e.g. the maturation of mRNA as found in mitochondria and possibly the indirect regulation of RNA polymerase III as proposed here.

The comparison of the diversity of functions and of substrate spectra between protein-only RNase P enzymes and the ancestral ribonucleoprotein (RNP) RNase P enzymes should give clues to understand why RNP RNase P was retained in some compartments or organisms and why they were replaced by protein-only RNase P in other compartments or organisms. In a more general context, this comparative analysis should give clues to understand how and why a transition took place from a prebiotic “RNA world” to the modern world that predominantly uses proteins for catalytic activities.

BIBLIOGRAPHY

1. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, et al. Rapid tRNA Decay Can Result from Lack of Nonessential Modifications. Molecular Cell. janv 2006;21(1):87- 96.

2. Anderson J, Phan L, Cuesta R, Carlson BA, Pak M, Asano K, et al. The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. Genes & Development. 1 déc 1998;12(23):3650- 62.

3. Antosz W, Deforges J, Begcy K, Bruckmann A, Poirier Y, Dresselhaus T, et al. Critical Role of Transcript Cleavage in Arabidopsis RNA Polymerase II Transcriptional Elongation. Plant Cell. mai 2020;32(5):1449- 63.

4. Antosz W, Deforges J, Begcy K, Bruckmann A, Poirier Y, Dresselhaus T, et al. Critical Role of Transcript Cleavage in Arabidopsis RNA Polymerase II Transcriptional Elongation. Plant Cell. mai 2020;32(5):1449- 63.

5. Arimbasseri GA. Interactions between RNAP III transcription machinery and tRNA processing factors. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. avr 2018;1861(4):354- 60.

6. Arimbasseri GA. Interactions between RNAP III transcription machinery and tRNA processing factors. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. avr 2018;1861(4):354- 60.

7. Arimbasseri GA. Interactions between RNAP III transcription machinery and tRNA processing factors. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. avr 2018;1861(4):354- 60.

8. Awai T, Ochi A, Ihsanawati, Sengoku T, Hirata A, Bessho Y, et al. Substrate tRNA Recognition Mechanism of a Multisite-specific tRNA Methyltransferase, *Aquifex aeolicus* Trm1, Based on the X-ray Crystal Structure. J Biol Chem. 7 oct 2011;286(40):35236- 46.

9. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol. 29 avr 2014;65(1):415- 42.

10. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol. 29 avr 2014;65(1):415- 42.

11. Becker B, Marin B. Streptophyte algae and the origin of embryophytes. Annals of Botany. mai 2009;103(7):999- 1004.

12. Becker B, Marin B. Streptophyte algae and the origin of embryophytes. Annals of Botany. mai 2009;103(7):999- 1004.

13. Bezanilla M, Perroud P - F., Pan A, Klueh P, Quatrano RS. An RNAi System in *Physcomitrella patens* with an Internal Marker for Silencing Allows for Rapid Identification of Loss of Function Phenotypes. Plant Biology. mai 2005;7(3):251- 7.

14. Blewett NH, Maraia RJ. La involvement in tRNA and other RNA processing events including differences among yeast and other eukaryotes. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. avr 2018;1861(4):361- 72.

15. Blewett NH, Maraia RJ. La involvement in tRNA and other RNA processing events including differences among yeast and other eukaryotes. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. avr 2018;1861(4):361- 72.

16. Boguta M, Czerska K, Żołądek T. Mutation in a new gene MAF1 affects tRNA suppressor efficiency in Saccharomyces cerevisiae. Gene. févr 1997;185(2):291- 6.

17. Bonhoure N, Byrnes A, Moir RD, Hodroj W, Preitner F, Praz V, et al. Loss of the RNA polymerase III repressor MAF1 confers obesity resistance. Genes Dev. 1 mai 2015;29(9):934- 47.

18. Bonhoure N, Byrnes A, Moir RD, Hodroj W, Preitner F, Praz V, et al. Loss of the RNA polymerase III repressor MAF1 confers obesity resistance. Genes Dev. 1 mai 2015;29(9):934- 47.

19. Bonhoure N, Praz V, Moir RD, Willemin G, Mange F, Moret C, et al. MAF1 is a chronic repressor of RNA polymerase III transcription in the mouse. Sci Rep. déc 2020;10(1):11956.

20. Bouhedda F, Fam KT, Collot M, Autour A, Marzi S, Klymchenko A, et al. A dimerization-based fluorogenic dye-aptamer module for RNA imaging in live cells. Nat Chem Biol. janv 2020;16(1):69- 76.

21. Bouhedda F, Fam KT, Collot M, Autour A, Marzi S, Klymchenko A, et al. A dimerization-based fluorogenic dye-aptamer module for RNA imaging in live cells. Nat Chem Biol. janv 2020;16(1):69- 76.

22. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, et al. Growth Stage–Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. :13.

23. Braglia P, Percudani R, Dieci G. Sequence Context Effects on Oligo(dT) Termination Signal Recognition by *Saccharomyces cerevisiae* RNA Polymerase III. J Biol Chem. 20 mai 2005;280(20):19551- 62.

24. Braglia P, Percudani R, Dieci G. Sequence Context Effects on Oligo(dT) Termination Signal Recognition by *Saccharomyces cerevisiae* RNA Polymerase III. J Biol Chem. 20 mai 2005;280(20):19551- 62.

25. Brillante N, Gößringer M, Lindenhofer D, Toth U, Rossmanith W, Hartmann RK.

Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. Nucleic Acids Res. 18 mars 2016;44(5):2323- 36.

26. Burch-Smith TM, Schiff M, Liu Y, Dinesh-Kumar SP. Efficient Virus-Induced Gene Silencing in Arabidopsis. 2020;142:7.

27. Burch-Smith TM, Schiff M, Liu Y, Dinesh-Kumar SP. Efficient Virus-Induced Gene Silencing in Arabidopsis. Plant Physiol. sept 2006;142(1):21- 7.

28. Cai Y, Wei Y-H. Stress resistance and lifespan are increased in *C. elegans* but decreased in *S. cerevisiae* by *mafr-1/maf1* deletion. Oncotarget. 8 mars 2016;7(10):10812- 26.

29. Canino G, Bocian E, Barbezier N, Echeverría M, Forner J, Binder S, et al. Arabidopsis Encodes Four tRNase Z Enzymes. Plant Physiol. juill 2009;150(3):1494- 502.

30. Carter R, Drouin G. The Evolutionary Rates of Eukaryotic RNA Polymerases and of Their Transcription Factors Are Affected by the Level of Concerted Evolution of the Genes They Transcribe. Molecular Biology and Evolution. 1 nov 2009;26(11):2515- 20.

31. Carter R, Drouin G. Structural differentiation of the three eukaryotic RNA polymerases. Genomics. déc 2009;94(6):388- 96.

32. Chan CTY, Pang YLJ, Deng W, Babu IR, Dyavaiah M, Begley TJ, et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun. janv 2012;3(1):937.

33. Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J. févr 2016;85(4):532- 47.

34. Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J. févr 2016;85(4):532- 47.

35. Clough SJ, Bent AF. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana: Floral dip transformation of Arabidopsis. The Plant Journal. déc 1998;16(6):735- 43.

36. Clough SJ, Bent AF. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana: Floral dip transformation of Arabidopsis. The Plant Journal. déc 1998;16(6):735- 43.

37. Clough et Bent - 1998 - Floral dip a simplified method forAgrobacterium-m.pdf.

38. Cognat V, Pawlak G, Duchêne A-M, Daujat M, Gigant A, Salinas T, et al. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Research. 1 janv 2013;41(D1):D273- 9.

39. Constantinesco F, Motorin Y, Grosjean H. Characterisation and Enzymatic Properties of tRNA(guanine 26, N2,N2)-dimethyltransferase (Trm1p) from Pyrococcus furiosus. Journal of Molecular Biology. août 1999;291(2):375- 92.

40. Coughlin DJ, Pleiss JA, Walker SC, Whitworth GB, Engelke DR. Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. Proceedings of the National Academy of Sciences. 26 août 2008;105(34):12218- 23.

41. Cramer P. Multisubunit RNA polymerases. Current Opinion in Structural Biology. févr 2002;12(1):89- 97.

42. Daniels CJ, Lai LB, Chen T-H, Gopalan V. Both kinds of RNase P in all domains of life: surprises galore. RNA. mars 2019;25(3):286- 91.

43. Daniels CJ, Lai LB, Chen T-H, Gopalan V. Both kinds of RNase P in all domains of life: surprises galore. RNA. mars 2019;25(3):286- 91.

44. de Souza TA, Soprano AS, Lira NPV de, Quaresma AJC, Pauletti BA, Leme AFP, et al. The TAL Effector PthA4 Interacts with Nuclear Factors Involved in RNA-Dependent Processes Including a HMG Protein That Selectively Binds Poly(U) RNA. Qiu J, éditeur. PLoS ONE. 22 févr 2012;7(2):e32305.

45. Desai N, Lee J, Upadhya R, Chu Y, Moir RD, Willis IM. Two Steps in Maf1-dependent Repression of Transcription by RNA Polymerase III. J Biol Chem. 25 févr 2005;280(8):6455- 62.

46. Desai N, Lee J, Upadhya R, Chu Y, Moir RD, Willis IM. Two Steps in Maf1-dependent Repression of Transcription by RNA Polymerase III. J Biol Chem. 25 févr 2005;280(8):6455- 62.

47. Dever TE, Green R. The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. Cold Spring Harbor Perspectives in Biology. 1 juill 2012;4(7):a013706- a013706.

48. Dever TE, Dinman JD, Green R. Translation Elongation and Recoding in Eukaryotes. Cold Spring Harb Perspect Biol. août 2018;10(8):a032649.

49. Dever TE, Hinnebusch AG. GCN2 Whets the Appetite for Amino Acids. Molecular Cell. avr 2005;18(2):141- 2.

50. Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. Trends in Genetics. déc 2007;23(12):614- 22.

51. Dreher TW. Role of tRNA-like structures in controlling plant virus replication. Virus Research. févr 2009;139(2):217- 29.

52. Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucl Acids Res. 1991;19(6):1349- 1349.

53. Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucl Acids Res. 1991;19(6):1349- 1349.

54. El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. Annu Rev Genet. 15 déc 2012;46(1):69- 95.

55. Englert M. Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. Nucleic Acids Research. 7 janv 2005;33(1):388- 99.

56. Evrard J-L, Nguyen I, Bergdoll M, Mutterer J, Steinmetz A, Lambert A-M. A novel pollen-specific α-tubulin in sunflower: structure and characterization. :10.

57. Evrard J-L, Nguyen I, Bergdoll M, Mutterer J, Steinmetz A, Lambert A-M. A novel pollen-specific α-tubulin in sunflower: structure and characterization. :10.

58. Ferguson A, Wang L, Altman RB, Terry DS, Juette MF, Burnett BJ, et al. Functional Dynamics within the Human Ribosome Regulate the Rate of Active Protein Synthesis. Molecular Cell. nov 2015;60(3):475- 86.

59. Filer D, Thompson MA, Takhaveev V, Dobson AJ, Kotronaki I, Green JWM, et al. RNA polymerase III limits longevity downstream of TORC1. Nature. déc 2017;552(7684):263- 7.

60. Fleurdépine S, Deragon J-M, Devic M, Guilleminot J, Bousquet-Antonelli C. A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. Nucleic Acids Research. mai 2007;35(10):3306- 21.

61. Fleurdépine S, Deragon J-M, Devic M, Guilleminot J, Bousquet-Antonelli C. A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. Nucleic Acids Research. mai 2007;35(10):3306- 21.

62. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. Science. 28 sept 2018;361(6409):1346- 9.

63. Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, et al. ePlant and the 3D Data Display Initiative: Integrative Systems Biology on the World Wide Web. Shiu S-H, éditeur. PLoS ONE. 10 janv 2011;6(1):e15237.

64. GegenheimerS P, GabiusQ H-J, Peeblesl CL, AbelsonII J. An RNA Ligase from Wheat Germ Which Participates in Transfer RNA Splicing in Vitro. :10.

65. GegenheimerS P, GabiusQ H-J, Peeblesl CL, AbelsonII J. An RNA Ligase from Wheat Germ Which Participates in Transfer RNA Splicing in Vitro. :10.

66. Gerber AP. An Adenosine Deaminase that Generates Inosine at the Wobble Position of tRNAs. Science. 5 nov 1999;286(5442):1146- 9.

67. Giegé P. Pentatricopeptide repeat proteins: A set of modular RNA-specific binders massively used for organelle gene expression. RNA Biology. sept

2013;10(9):1417- 8.

68. Giegé R, Jühling F, Pütz J, Stadler P, Sauter C, Florentz C. Structure of transfer RNAs: similarity and variability: Structure of transfer RNAs. WIREs RNA. janv 2012;3(1):37- 61.

69. Gobert A, Gutmann B, Taschner A, Gößringer M, Holzmann J, Hartmann RK, et al. A single Arabidopsis organellar protein has RNase P activity. Nat Struct Mol Biol. juin 2010;17(6):740- 4.

70. Gobert A, Pinker F, Fuchsbauer O, Gutmann B, Boutin R, Roblin P, et al. Structural insights into protein-only RNase P complexed with tRNA. Nat Commun. juin 2013;4(1):1353.

71. Goldman E. Transfer RNA. In: John Wiley & Sons, Ltd, éditeur. Encyclopedia of Life Sciences [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2008 [cité 5 nov 2020]. p. a0000878.pub2. Disponible sur: http://doi.wiley.com/10.1002/9780470015902.a0000878.pub2

72. Graczyk D, Debski J, Muszynska G, Bretner M, Lefebvre O, Boguta M. Casein kinase II-mediated phosphorylation of general repressor Maf1 triggers RNA polymerase III activation. Proceedings of the National Academy of Sciences. 22 mars 2011;108(12):4926- 31.

73. Gregori J, Sanchez A, Villanueva J. msmsTests package Blocks design to compensate batch effects. :8.

74. Gregori J, Sanchez A, Villanueva J. msmsTests package Blocks design to compensate batch effects. :8.

75. Grewal SS. Why should cancer biologists care about tRNAs? tRNA synthesis, mRNA translation and the control of growth. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. juill 2015;1849(7):898- 907.

76. Gutmann B, Gobert A, Giege P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes & Development. 15 mai 2012;26(10):1022- 7.

77. Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJH, et al. Molecular structures of unbound and transcribing RNA polymerase III. Nature. déc 2015;528(7581):231- 6.

78. Holzmann J, Frank P, Löffler E, Bennett KL, Gerner C, Rossmanith W. RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. Cell. oct 2008;135(3):462- 74.

79. Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. Nucleic Acids Res. 4 janv 2017;45(D1):D1064- 74.

80. Hopper AK. Transfer RNA Post-Transcriptional Processing, Turnover, and Subcellular Dynamics in the Yeast *Saccharomyces cerevisiae*. Genetics. mai

2013;194(1):43- 67.

81. Howard MJ, Lim WH, Fierke CA, Koutmos M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. Proceedings of the National Academy of Sciences. 2 oct 2012;109(40):16149- 54.

82. Howard MJ, Karasik A, Klemm BP, Mei C, Shanmuganathan A, Fierke CA, et al. Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. RNA. mai 2016;22(5):782- 92.

83. Ihsanawati, Nishimoto M, Higashijima K, Shirouzu M, Grosjean H, Bessho Y, et al. Crystal Structure of tRNA N2,N2-Guanosine Dimethyltransferase Trm1 from Pyrococcus horikoshii. Journal of Molecular Biology. nov 2008;383(4):871- 84.

84. Imai T, Nakamura T, Maeda T, Nakayama K, Gao X, Nakashima T, et al. Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TψC loop in precursor tRNAs. Biochemical and Biophysical Research Communications. août 2014;450(4):1541- 6.

85. Jackman JE, Alfonzo JD. Transfer RNA modifications: nature's combinatorial chemistry playground: Transfer RNA modifications. WIREs RNA. janv 2013;4(1):35- 48.

86. Jarrous N, Gopalan V. Archaeal/Eukaryal RNase P: subunits, functions and RNA diversification. Nucleic Acids Research. 1 déc 2010;38(22):7885- 94.

87. Jarrous N, Gopalan V. Archaeal/Eukaryal RNase P: subunits, functions and RNA diversification. Nucleic Acids Research. 1 déc 2010;38(22):7885- 94.

88. Johnson PF, Abelson J. The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. Nature. avr 1983;302(5910):681- 7.

89. Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Research. 1 janv 2009;37(Database):D159- 62.

90. Kaiser F, Krautwurst S, Salentin S, Haupt VJ, Leberecht C, Bittrich S, et al. The structural basis of the genetic code: amino acid recognition by aminoacyl-tRNA synthetases. Sci Rep. déc 2020;10(1):12647.

91. Kamenski P, Kolesnikova O, Jubenot V, Entelis N, Krasheninnikov IA, Martin RP, et al. Evidence for an Adaptation Mechanism of Mitochondrial Translation via tRNA Import from the Cytosol. Molecular Cell. juin 2007;26(5):625- 37.

92. Karasik A, Shanmuganathan A, Howard MJ, Fierke CA, Koutmos M. Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5′ End Processing Enzymes. Journal of Molecular Biology. janv 2016;428(1):26- 40.

93. Kirchner G, Vijayraghavan U, Jacobsonll A, Martins NC, Abelson J. Isolation of

a Temperature-sensitive Mutant with an Altered tRNA Nucleotidyltransferase and Cloning of the Gene Encoding tRNA Nucleotidyltransferase in the Yeast Saccharomyces cereuisiae. :6.

94. Kirchner S, Ignatova Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nat Rev Genet. févr 2015;16(2):98- 112.

95. Kotelawala L, Grayhack EJ, Phizicky EM. Identification of yeast tRNA Um44 2'-O-methyltransferase (Trm44) and demonstration of a Trm44 role in sustaining levels of specific tRNASer species. RNA. 12 nov 2007;14(1):158- 69.

96. Kuhn C-D. On-Enzyme Refolding Permits Small RNA and tRNA Surveillance by the CCA-Adding Enzyme. :16.

97. Kuhn C-D, Wilusz JE, Zheng Y, Beal PA, Joshua-Tor L. On-Enzyme Refolding Permits Small RNA and tRNA Surveillance by the CCA-Adding Enzyme. Cell. févr 2015;160(4):644- 58.

98. Lange H, Sement FM, Canaday J, Gagliardi D. Polyadenylation-assisted RNA degradation processes in plants. :8.

99. Lange et al. - Polyadenylation-assisted RNA degradation processes.pdf.

100. Lange H, Sement FM, Canaday J, Gagliardi D. Polyadenylation-assisted RNA degradation processes in plants. Trends in Plant Science. sept 2009;14(9):497- 504.

101. Lee J, Moir RD, Willis IM. Regulation of RNA Polymerase III Transcription Involves *SCH9* -dependent and *SCH9* -independent Branches of the Target of Rapamycin (TOR) Pathway. J Biol Chem. 8 mai 2009;284(19):12604- 8.

102. Leśniewska E, Boguta M. Novel layers of RNA polymerase III control affecting tRNA gene transcription in eukaryotes. Open Biol. févr 2017;7(2):170001.

103. Lorenz C, Lünse C, Mörl M. tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomolecules. 4 avr 2017;7(4):35.

104. Lu X, Huang J, Wu S, Zheng Q, Liu P, Feng H, et al. The tRNA - like small noncoding RNA masc RNA promotes global protein translation. EMBO Rep [Internet]. 19 oct 2020 [cité 5 nov 2020]; Disponible sur: https://onlinelibrary.wiley.com/doi/10.15252/embr.201949684

105. Lu X, Huang J, Wu S, Zheng Q, Liu P, Feng H, et al. The tRNA - like small noncoding RNA masc RNA promotes global protein translation. EMBO Rep [Internet]. 19 oct 2020 [cité 12 nov 2020]; Disponible sur: https://onlinelibrary.wiley.com/doi/10.15252/embr.201949684

106. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways—2013 update. Nucleic Acids Research. 30 oct 2012;41(D1):D262- 7.

107. Marquet R. tRNAs as prirner of reverse transcriptases. :12.

108. Martin NC, Hopper AK. How single genes provide tRNA processing enzymes to mitochondria, nuclei and the cytosol. Biochimie. janv 1994;76(12):1161- 7.

109. Marvin MC, Clauder-Munster S, Walker SC, Sarkeshik A, Yates JR, Steinmetz LM, et al. Accumulation of noncoding RNA due to an RNase P defect in Saccharomyces cerevisiae. RNA. 1 août 2011;17(8):1441- 50.

110. Megel C, Hummel G, Lalande S, Ubrig E, Cognat V, Morelle G, et al. Plant RNases T2, but not Dicer-like proteins, are major players of tRNA-derived fragments biogenesis. Nucleic Acids Research. 25 janv 2019;47(2):941- 52.

111. Mei Y, Yong J, Liu H, Shi Y, Meinkoth J, Dreyfuss G, et al. tRNA Binds to Cytochrome c and Inhibits Caspase Activation. Molecular Cell. mars 2010;37(5):668- 78.

112. Mergner J, Frejno M, List M, Papacek M, Chen X, Chaudhary A, et al. Mass-spectrometry-based draft of the Arabidopsis proteome. Nature. 19 mars 2020;579(7799):409- 14.

113. Mergner et al. - 2020 - Mass-spectrometry-based draft of the Arabidopsis p.pdf.

114. Merrick WC, Pavitt GD. Protein Synthesis Initiation in Eukaryotic Cells. Cold Spring Harb Perspect Biol. déc 2018;10(12):a033092.

115. Moir RD, Lee J, Haeusler RA, Desai N, Engelke DR, Willis IM. Protein kinase A regulates RNA polymerase III transcription through the nuclear localization of Maf1. Proceedings of the National Academy of Sciences. 10 oct 2006;103(41):15044- 9.

116. Mori S, Kajita T, Endo T, Yoshihisa T. The intron of tRNA-TrpCCA is dispensable for growth and translation of Saccharomyces cerevisiae. RNA. 1 sept 2011;17(9):1760- 9.

117. Murguía JR, Serrano R. New functions of protein kinase Gcn2 in yeast and mammals. IUBMB Life. déc 2012;64(12):971- 4.

118. Nickel AI, Wäber NB, Gößringer M, Lechner M, Linne U, Toth U, et al. Minimal and RNA-free RNase P in *Aquifex aeolicus*. Proc Natl Acad Sci USA. 17 oct 2017;114(42):11121- 6.

119. Nilsen TW. Gel Purification of RNA. Cold Spring Harbor Protocols. 1 févr 2013;2013(2):pdb.prot072942-pdb.prot072942.

120. Nilsen TW. Gel Purification of RNA. Cold Spring Harbor Protocols. 1 févr 2013;2013(2):pdb.prot072942-pdb.prot072942.

121. O'Connor JP, Peebles CL. In vivo pre-tRNA processing in Saccharomyces cerevisiae. Mol Cell Biol. janv 1991;11(1):425- 39.

122. O'Connor JP, Peebles CL. In vivo pre-tRNA processing in Saccharomyces cerevisiae. Mol Cell Biol. janv 1991;11(1):425- 39.

123. Oficjalska-Pham D, Harismendy O, Smagowicz WJ, Gonzalez de Peredo A, Boguta M, Sentenac A, et al. General Repression of RNA Polymerase III Transcription Is Triggered by Protein Phosphatase Type 2A-Mediated Dephosphorylation of Maf1. Molecular Cell. juin 2006;22(5):623- 32.

124. Oler AJ, Cairns BR. PP4 dephosphorylates Maf1 to couple multiple stress conditions to RNA polymerase III repression: PP4 dephosphorylates Maf1 to repress Pol III. The EMBO Journal. 21 mars 2012;31(6):1440- 52.

125. Ovcharenko A, Rentmeister A. Emerging approaches for detection of methylation sites in RNA. :8.

126. Ovcharenko et Rentmeister - Emerging approaches for detection of methylation s.pdf.

127. Palian BM, Rohira AD, Johnson SAS, He L, Zheng N, Dubeau L, et al. Maf1 Is a Novel Target of PTEN and PI3K Signaling That Negatively Regulates Oncogenesis and Lipid Metabolism. Kim SK, éditeur. PLoS Genet. 11 déc 2014;10(12):e1004789.

128. Pan T, Loria A, Zhong K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. Proceedings of the National Academy of Sciences. 19 déc 1995;92(26):12510- 4.

129. Pan T, Loria A, Zhong K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. Proceedings of the National Academy of Sciences. 19 déc 1995;92(26):12510- 4.

130. Paule MR. SURVEY AND SUMMARY Transcription by RNA polymerases I and III. Nucleic Acids Research. 15 mars 2000;28(6):1283- 98.

131. Paushkin SV, Patel M, Furia BS, Peltz SW, Trotta CR. Identification of a Human Endonuclease Complex Reveals a Link between tRNA Splicing and Pre-mRNA 3′ End Formation. Cell. avr 2004;117(3):311- 21.

132. Pavlova LV, Gößringer M, Weber C, Buzet A, Rossmanith W, Hartmann RK. tRNA Processing by Protein-Only versus RNA-Based RNase P: Kinetic Analysis Reveals Mechanistic Differences. ChemBioChem. 15 oct 2012;13(15):2270- 6.

133. Phizicky EM, Hopper AK. tRNA biology charges to the front. Genes & Development. 1 sept 2010;24(17):1832- 60.

134. Phizicky EM, Hopper AK. tRNA biology charges to the front. Genes & Development. 1 sept 2010;24(17):1832- 60.

135. Piekna-Przybylska D, DiChiacchio L, Mathews DH, Bambara RA. A sequence similar to tRNA3Lys gene is embedded in HIV-1 U3–R and promotes minus-strand transfer. Nat Struct Mol Biol. janv 2010;17(1):83- 9.

136. Pikaard CS, Haag JR, Ream T, Wierzbicki AT. Roles of RNA polymerase IV in gene silencing. Trends in Plant Science. juill 2008;13(7):390- 7.

137. Pinker F, Giegé P, Sauter C. Crystallization and crystallographic analysis of an *Arabidopsis* nuclear proteinaceous RNase P. Acta Crystallogr F Struct Biol Commun. 1 nov 2015;71(11):1372- 7.

138. Pinker F, Schelcher C, Fernandez-Millan P, Gobert A, Birck C, Thureau A, et al. Biophysical analysis of *Arabidopsis* protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding. J Biol Chem. 25 août 2017;292(34):13904- 13.

139. Placido A, Sieber F, Gobert A, Gallerani R, Giegé P, Maréchal-Drouard L. Plant mitochondria use two pathways for the biogenesis of tRNA His. Nucleic Acids Research. nov 2010;38(21):7711- 7.

140. Pluta K, Lefebvre O, Martin NC, Smagowicz WJ, Stanford DR, Ellis SR, et al. Maf1p, a Negative Effector of RNA Polymerase III inSaccharomyces cerevisiae. Mol Cell Biol. 1 août 2001;21(15):5031- 40.

141. Ream TS, Haag JR, Pontvianne F, Nicora CD, Norbeck AD, Paša-Tolić L, et al. Subunit compositions of Arabidopsis RNA polymerases I and III reveal Pol I- and Pol III-specific forms of the AC40 subunit and alternative forms of the C53 subunit. Nucleic Acids Research. 30 avr 2015;43(8):4163- 78.

142. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Mondragón A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. 2011;20.

143. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragón A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature. déc 2010;468(7325):784- 9.

144. Ren B, Wang X, Duan J, Ma J. Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. Science. 30 août 2019;365(6456):919- 22.

145. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 1 janv 2010;26(1):139- 40.

146. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 1 janv 2010;26(1):139- 40.

147. Samanta MP, Tongprasit W, Sethi H, Chin C-S, Stolc V. Global identification of noncoding RNAs in Saccharomyces cerevisiae by modulating an essential RNA processing pathway. Proceedings of the National Academy of Sciences. 14 mars 2006;103(11):4192- 7.

148. Schmidt von Braun S, Sabetti A, Hanic-Joyce PJ, Gu J, Schleiff E, Joyce PBM. Dual targeting of the tRNA nucleotidyltransferase in plants: not just the signal. Journal of Experimental Botany. 26 nov 2007;58(15- 16):4083- 93.

149. Schramm L. Recruitment of RNA polymerase III to its target promoters. Genes & Development. 15 oct 2002;16(20):2593- 620.

150. Schramm - 2002 - Recruitment of RNA polymerase III to its target pr.pdf.

151. Schwarz TS, Wäber NB, Feyh R, Weidenbach K, Schmitz RA, Marchfelder A, et al. Homologs of *aquifex aeolicus* protein- only RNase P are not the major RNase P activities in the archaea *haloferax volcanii* and *methanosarcina mazei*. IUBMB Life. 8 juill 2019;iub.2122.

152. Schwarz TS, Wäber NB, Feyh R, Weidenbach K, Schmitz RA, Marchfelder A, et al. Homologs of *aquifex aeolicus* protein- only RNase P are not the major RNase P activities in the archaea *haloferax volcanii* and *methanosarcina mazei*. IUBMB Life. 8 juill 2019;iub.2122.

153. Simon L, Rabanal FA, Dubos T, Oliver C, Lauber D, Poulet A, et al. Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in Arabidopsis thaliana. Nucleic Acids Research. 6 avr 2018;46(6):3019- 33.

154. Small ID, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends in Biochemical Sciences. févr 2000;25(2):45- 7.

155. Small ID, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends in Biochemical Sciences. févr 2000;25(2):45- 7.

156. Song H, Mugnier P, Das AK, Webb HM, Evans DR, Tuite MF, et al. The Crystal Structure of Human Eukaryotic Release Factor eRF1—Mechanism of Stop Codon Recognition and Peptidyl-tRNA Hydrolysis. :11.

157. Soprano AS, Abe VY, Smetana JHC, Benedetti CE. Citrus MAF1, a Repressor of RNA Polymerase III, Binds the *Xanthomonas citri* Canker Elicitor PthA4 and Suppresses Citrus Canker Development. Plant Physiol. sept 2013;163(1):232- 42.

158. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep. avr 2017;7(1):44598.

159. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. Compilation of tRNA sequences and sequences of tRNA genes. :6.

160. Sun T. Carotenoid Metabolism in Plants: The Role of Plastids. Molecular Plant. :17.

161. Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, Li L. Carotenoid Metabolism in Plants: The Role of Plastids. Molecular Plant. janv 2018;11(1):58- 74.

162. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN / nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Research. 22 déc 2008;19(3):347- 59.

163. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN / nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome

Research. 22 déc 2008;19(3):347- 59.

164. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN / nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Research. 22 déc 2008;19(3):347- 59.

165. Swinehart WE, Jackman JE. Diversity in mechanism and function of tRNA methyltransferases. RNA Biology. 3 avr 2015;12(4):398- 411.

166. Takaku H. A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. Nucleic Acids Research. 1 mai 2003;31(9):2272- 8.

167. Thomas BC, Li X, Gegenheimer P. Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. RNA. avr 2000;6(4):545- 53.

168. Thompson DM, Parker R. The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae. Journal of Cell Biology. 6 avr 2009;185(1):43- 50.

169. Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, et al. The Yeast tRNA Splicing Endonuclease: A Tetrameric Enzyme with Two Active Site Subunits Homologous to the Archaeal tRNA Endonucleases. Cell. juin 1997;89(6):849- 58.

170. Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, et al. The Yeast tRNA Splicing Endonuclease: A Tetrameric Enzyme with Two Active Site Subunits Homologous to the Archaeal tRNA Endonucleases. Cell. juin 1997;89(6):849- 58.

171. Trujillo JT, Seetharam AS, Hufford MB, Beilstein MA, Mosher RA. Evidence for a Unique DNA-Dependent RNA Polymerase in Cereal Crops. de Meaux J, éditeur. Molecular Biology and Evolution. 1 oct 2018;35(10):2454- 62.

172. Tsugama D, Liu S, Takano T. A rapid chemical method for lysing Arabidopsis cells for protein analysis. Plant Methods. 2011;7(1):22.

173. Tsugama D, Liu S, Takano T. A rapid chemical method for lysing Arabidopsis cells for protein analysis. Plant Methods. 2011;7(1):22.

174. Vannini A, Ringel R, Kusser AG, Berninghausen O, Kassavetis GA, Cramer P. Molecular Basis of RNA Polymerase III Transcription Repression by Maf1. Cell. oct 2010;143(1):59- 70.

175. Waibel F, Filipowicz W, Miescher-Institut F. U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase Ill but contain the same two upstream promoter elements as RNA polymerase 11-transcribed U-snRNA genes. :8.

176. Wei Y, Tsang CK, Zheng XFS. Mechanisms of regulation of RNA polymerase III-dependent transcription by TORC1. EMBO J. 5 août 2009;28(15):2220- 30.

177. Werner M. Vincent Thuillier, Sophie Stettlerl, Andre Sentenac, Pierre Thuriaux and. :9.

178. Whitney ML, Hurto RL, Shaheen HH, Hopper AK. Rapid and Reversible Nuclear Accumulation of Cytoplasmic tRNA in Response to Nutrient Availability□D. Molecular Biology of the Cell. 2007;18:9.

179. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell. nov 2008;135(4):635- 48.

180. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell. nov 2008;135(4):635- 48.

181. Willis IM, Moir RD, Hernandez N. Metabolic programming a lean phenotype by deregulation of RNA polymerase III. Proc Natl Acad Sci USA. 27 nov 2018;115(48):12182- 7.

182. Wilusz JE, Whipple JM, Phizicky EM, Sharp PA. tRNAs Marked with CCACCA Are Targeted for Degradation. Science. 11 nov 2011;334(6057):817- 21.

183. Wilusz JE, Whipple JM, Phizicky EM, Sharp PA. tRNAs Marked with CCACCA Are Targeted for Degradation. Science. 11 nov 2011;334(6057):817- 21.

184. Wilusz JE, Freier SM, Spector DL. 3′ End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. :14.

185. Wilusz JE, Freier SM, Spector DL. 3′ End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. Cell. nov 2008;135(5):919- 32.

186. Wilusz JE, Freier SM, Spector DL. 3′ End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. Cell. nov 2008;135(5):919- 32.

187. Wu J, Niu S, Tan M, Huang C, Li M, Song Y, et al. Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. Cell. nov 2018;175(5):1393-1404.e11.

188. Wu J, Niu S, Tan M, Huang C, Li M, Song Y, et al. Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. Cell. nov 2018;175(5):1393-1404.e11.

189. Xiao S, Scott F, Fierke CA, Engelke DR. Eukaryotic Ribonuclease P: A Plurality of Ribonucleoprotein Enzymes. Annu Rev Biochem. juin 2002;71(1):165- 89.

190. Yamasaki S, Ivanov P, Hu G, Anderson P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. Journal of Cell Biology. 6 avr 2009;185(1):35- 42.

191. Yoo CJ, Wolin SL. The Yeast La Protein Is Required for the 3′ Endonucleolytic Cleavage That Matures tRNA Precursors. Cell. mai 1997;89(3):393- 402.

192. Yukawa Y, Felis M, Englert M, Stojanov M, Matoušek J, Beier H, et al. Plant 7SL RNA genes belong to type 4 of RNA polymerase III- dependent genes that are composed of mixed promoters: 7SL RNA gene promoters. The Plant Journal. 2 juin 2005;43(1):97- 106.

193. Zhu L, Deutscher MP. tRNA nucleotidyltransferase is not essential for Escherichia coli viability. The EMBO Journal. août 1987;6(8):2473- 7.

194. Aravind L. An Evolutionary Classification of the Metallo-ß-Lactamase Fold Proteins. In Silico Biology. 1999;1(2):69- 91.

195. Chen S, Cai X, Luo M. 4.18 - Mechanistic Aspects of Methyltransferases with Small-Molecule Metabolites and Natural Products as Substrates. In: Liu H-W (Ben), Begley TP, éditeurs. Comprehensive Natural Products III [Internet]. Oxford: Elsevier; 2020. p. 474- 96. Disponible sur: http://www.sciencedirect.com/science/article/pii/B9780124095472147821

196. Scheer H, De Almeida C, Sikorska N, Koechler S, Gagliardi D, Zuber H. High-Resolution Mapping of 3' Extremities of RNA Exosome Substrates by 3' RACE-Seq. Methods Mol Biol. 2020; 2062:147- 67.

Université de Strasbourg

Mathieu BRUGGEMAN

Interaction fonctionnelles des RNase P nucléaires chez *Arabidopsis thaliana*

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Résumé

La RNase P est une enzyme essentielle et virtuellement universelle, qui mature l'extrémité 5' des précurseurs d'ARN de transfert (ARNt). Mon travail de thèse a visé à caractériser les interactions fonctionnelles de la RNase P nucléaire (PRORP2) chez la plante modèle en biologie végétale *Arabidopsis thaliana*. L'objectif de ma thèse était de caractériser le réseau d'interaction protéine / protéine de PRORP2 par la technique de co-immunoprécipitation (Co-IP). Ainsi que d'identifier de nouveaux substrats ARN (autres que les ARNt) de PRORP2 au niveau du transcriptome nucléaire. Nos résultats de Co-immunoprécipitation montrent que PRORP2 interagit avec deux enzymes catalysant une modification des bases nucléiques des ARNt : les tRNA methyltransferase (TRM) 1A et 1B. Ce résultat permet de mieux comprendre les étapes précoces de la maturation des ARNt chez les plantes. De plus, des lignées portant une insertion ADN-T dans ces deux gènes ont été génotypées et étudiées phénotypiquement afin de caractériser plus en détail la fonction de ces deux gènes. Enfin, le résultat principal de ma thèse est la mise en évidence d'un nouveau substrat ARN de PRORP2. En effet, le transcrit d'un gène appelé MAF1 (dont la fonction est de réguler l'activité de l'ARN polymerase III qui transcrit les ARNt) contient une structure ressemblant à un ARN de transfert. Mon travail de thèse a permis de montrer que cette structure est conservée dans l'évolution chez les plantes terrestres et qu'elle peut être clivée *in vitro* par la RNase P. Enfin une approche par 3'RACE a permis de montrer que des transcrits de MAF1 finissant juste avant cette structure s'accumulent *in vivo*. Ce résultat permet d'appuyer l'hypothèse que PRORP2 clive l'ARNm de MAF1 soit pour générer un petit ARN non-codant, soit pour réguler le niveau d'expression de MAF1 et par extension l'activité de l'ARN polymerase III. Globalement les résultats obtenus au cours de cette thèse de doctorat ont permis d'améliorer notre connaissance des modalités de biogénèse des ARN de transfert cytosoliques chez les plantes ainsi que de découvrir un nouveau substrat ARN de PRORP2. Ce dernier résultat permet de proposer que PRORP2 pourrait réguler indirectement l'activité de la RNA polymerase III chez *Arabidopsis thaliana*.

Mots-clés : RNase P, ARN de transfert (ARNt), TLS, MAF1, *Arabidopsis thaliana*