UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

Institut de Biologie Moléculaire des Plantes - IBMP

THÈSE présentée par :

Franziska PINKER

soutenue le : 15 septembre 2014

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

Discipline/ Spécialité : Sciences du Vivant /Biochimie, Biologie Moléculaire et Structurale

Structural characterization of proteinaceous RNase P from *Arabidopsis thaliana*

THÈSE dirigée par :

M. Claude SAUTER	Dr, Université de Strasbourg
M. Philippe GIEGE	Dr, Université de Strasbourg

RAPPORTEURS :

M. Walter ROSSMANITH	Dr, Université de Vienne
Mme Emmanuelle SCHMITT	Dr, Ecole Polytechnique

AUTRES MEMBRES DU JURY :

Mme Anita MARCHFELDER	Professeur, Université d'Ulm
Mme Pascale ROMBY	Dr, Université de Strasbourg

Für meine Familie

Acknowledgements

First of all I would like to thank the members of my jury, Dr. Emmanuelle Schmitt, Dr. Pascale Romby, Dr. Anita Marchfelder and Dr. Walter Rossmanith, for having agreed to judge my work.

I would like to offer my special thanks to my Ph.D. advisors Dr. Claude Sauter and Dr. Philippe Giegé for having accepted me in their laboratories and for their infinite patience and support during the last three years.

I would like to express my very great appreciation to Agnès and Anthony who gave me so much advice, good discussions and encouragement at the IBMC and IBMP. Furthermore, I thank all the labmembers: Catherine, Marie, Bernard, Jörn, Loukmane, Ayoub, Thalia and Géraldine.

I would like to offer my thanks to all the members of the IBMC and IBMP for answering all kind of questions and making working so much fun, especially Joëlle, Kamel, Anne-So, Mélodie, Hagen, Anja and Bernadette.

Finally, I wish to thank my family and my friends that supported me during the years, who cheered me up and who most importantly encouraged me to go on.

Heiko und Dieter: danke, dass Ihr mich mit auf die tollsten Berge und die schönsten Trails der Region genommen habt. Ohne euch wäre Badisch noch ein Buch mit sieben Siegeln für mich.

Abbreviations

aaRS	Aminoacyl-tRNA-synthetase
CBD	Chitin binding domain
cv	Column volume
Da	Dalton
DLS	Dynamic light scattering
EMSA	Electromobility shift assay
HPLC	High performance liquid chromatography
ICP-MS	Inductively coupled plasma mass spectrometry
LB	Lysogeny broth
MALS	Multi-angle light scattering
MN	Micrococcal nuclease
mRNA	Messenger RNA
MRPP	Mitochondrial ribonuclease P protein
MST	Microscale thermophoresis
MTS	Mitochondrial targeting signal
MWCO	Molecular weight cut-off
NEP	Nuclear-encoded phage-type RNA polymerases
NiNTA	Nickel nitrilotriacetic acid
nts	Nucleotides
PAA	Polyacrylamide
PNPase	Polynucleotide phosphorylase
pre-tRNA	Precursor tRNA
PRORP	Proteinaceous RNase P
RNase P	Ribonuclease P
rRNA	Ribosomal RNA
SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particles
SRCD	Synchrotron radiation circular dichroism
SV	Sedimentation velocity
TCEP	Tris(2-carboxyethyl)phosphine
TFIII	Transcription factor III

Contents

List of Figures

List of Tables

Introduction

1.1 Transfer RNAs

1.1.1 Transfer RNAs - structure and function

Transfer RNAs (tRNAs) play a major role in protein biogenesis. They are the adaptor molecules that bridge the genetic information, stored as DNA and carried after transcription by the mRNAs, to the primary sequence of proteins (Fig. 1.1). Each tRNA possesses a specific base triplet in the anticodon loop that base pairs with corresponding triplet codons in the sequence coded by the mRNA. In addition isoacceptors may exist (tRNAs with the same anticodon triplet but different body sequences). They were found to be less or not implicated in translation *in vivo* or *in vitro*. They were shown to play a role in the regulation of aminoacyl-tRNA-synthetase (aaRS) mRNAs, viral replication, amino acid biosynthesis, cell wall remodelling or antibiotic resistances (Geslain & Pan, 2011).

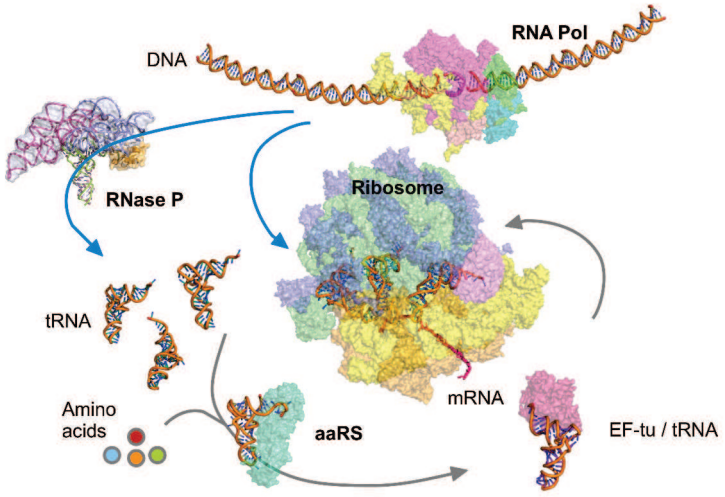


Figure 1.1: tRNA implication in protein translation. DNA coding for tRNAs is transcribed by RNA polymerase III and processed in successive steps to obtain mature tRNAs. These tRNAs are loaded with their specific amino acid and will be directed as tRNA-amino acid/EF-Tu complexes to the ribosomal/mRNA complex. There, mRNAs will serve as a template for the synthesis of the primary sequence of the protein encoded in the DNA.

The genome of *Arabidopsis thaliana* contains around 630 tRNA genes that decode for the full set of 20 amino acids (Lowe & Chan, 2011). tRNAs are loaded with specific amino acids by their cognate aminoacyl-tRNA-synthetases.

In general tRNAs consist of 60-95 nucleotides (nts) that fold into a characteristic cloverleaf secondary

structure by Watson-Crick or unusual base pairs. This cloverleaf structure is characterized by the acceptor arm charging the amino acid and the D arm, rich in dihydrouridine, opposite to the T arm that often contains a conserved TΨC sequence. The anticodon arm is situated between the D and the T arm and is followed by a loop of variable length (variable loop) (Fig. 1.2). Usually nine long distance tertiary interactions help to fold into the three dimensional L-form that is also a characteristic of canonical tRNAs (Rich & RajBhandary, 1976, Giegé, 2008, Giegé *et al.*, 2012). Several maturation steps, described below, are required to produce functional tRNAs (Fig. 1.3).

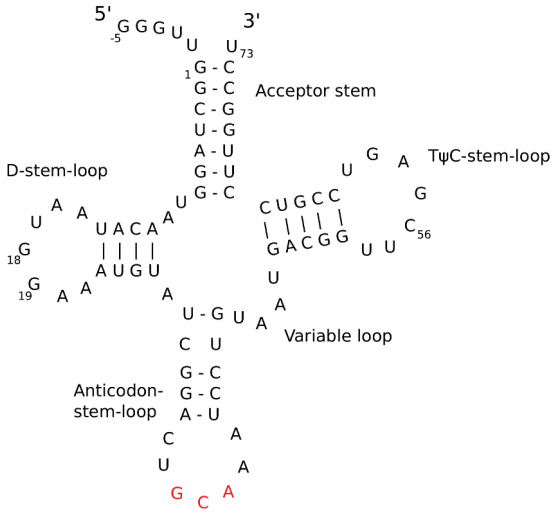


Figure 1.2: Secondary structure of mitochondrial precursor tRNA^{Cys} from *A. thaliana* including the unpaired 5' leader.

1.1.2 Prokaryotic tRNA transcription

In *E. coli* all tRNA genes lack introns and possess a CCA encoded 3' terminus. The 79 tRNA genes are organized in 41 transcriptional units or clusters. 59 out of 79 tRNA genes are located in these clusters and are associated with either rRNA or protein encoding genes. tRNAs encoded in non-ribosomal RNA transcription units are under the control of promoters having conserved -35 and -10 sequences upstream the initiation start (Travers, 1984, Inokuchi & Yamao, 1995).

1.1.3 tRNA transcription in eukaryotic nuclei

Eukaryotic nuclear tRNA transcription is more complex and has been well characterized in yeast. Promoter regions that will be recognized by the RNA polymerase III machinery lie up- and downstream the transcription initiation site. TFIIIC binds within the coding region of the tRNA genes (Fig. 1.4) (White, 2011). These regions are called box A and B, corresponding to bases 8 to 19 and 52 to 62 of a tRNA (D- and T-arm of a tRNA sequence), respectively (Zhang *et al.*, 2011). The 5' upstream region is also important for tRNA gene transcription. TFIIIB complex binds to the TATA-like box at position -29 to -34 (referring to the start of the mature tRNA), a region rich in A and T. In addition to

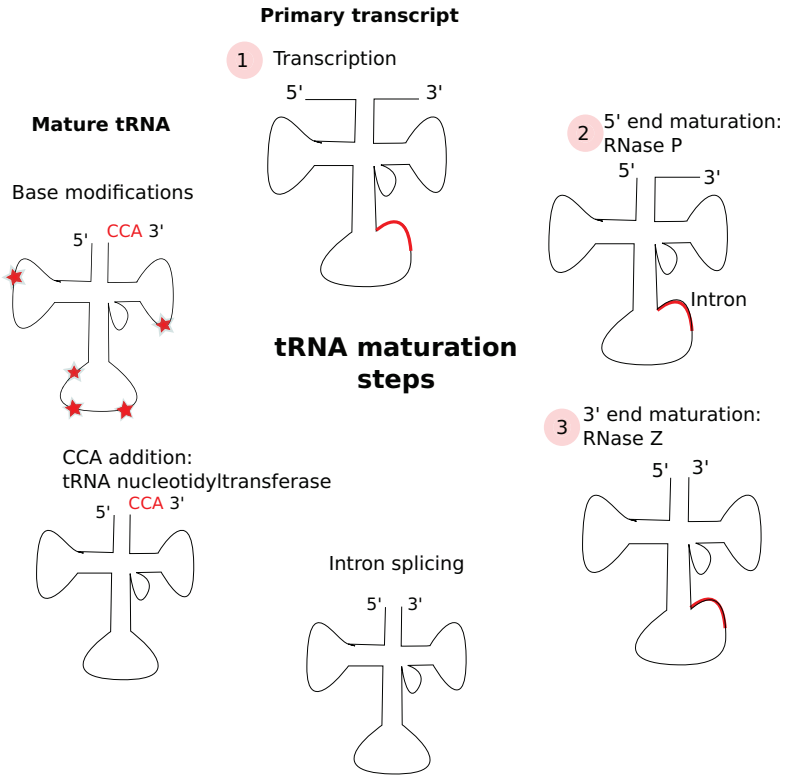


Figure 1.3: tRNA maturation steps in eukaryotic nuclei. The numbering corresponds to the sequential time table of each maturation event. 1) tRNA transcription carried out by RNA polymerase III, 2) 5' end maturation performed by RNase P, 3) 3' end maturation performed by RNase Z. The temporal order of the following steps is not clear but: introns are removed when required, a CCA is added by tRNA nucleotidyltransferases and nucleotides are modified by a multitude of different enzymes.

this a conserved CAA in plant and in yeast is situated at position -7 to -3 which also enhances efficient transcription and serves as transcription initiation site (Zhang *et al.*, 2011, Yukawa *et al.*, 2000, Hasegawa *et al.*, 2003). Transcription termination is promoted by a stretch of T's in the non-coding strand of the DNA that destabilizes the template-PolIII complex (Sprague, 1995).

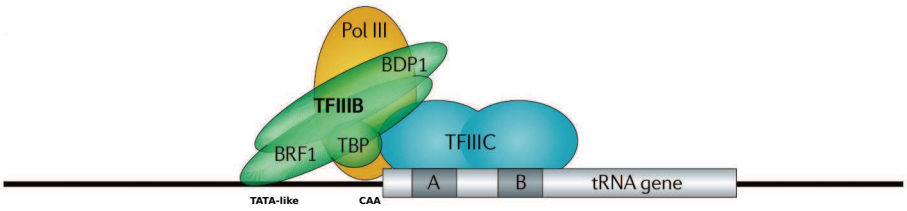


Figure 1.4: Eukaryotic polymerase III transcription unit (adopted from White (2011)). TFIIIB binds to TATA-like and TA-rich 5' upstream regions and PolIII will recognize a conserved CAA motif directly upstream the tRNA gene while TFIIIC specifically binds to conserved tRNA internal boxes A and B.

1.1.4 tRNAs in the model plant Arabidopsis thaliana

The nuclear genome of *Arabidopsis* codes for about 600 tRNAs (manually curated predictions) that are spread over the five chromosomes. *Arabidopsis* chromosome 1 contains two large clusters that

are the result of gene duplication events. The first cluster contains 27 tandem repeats of tRNA^{Pro} and the second 27 tandem repeats of tRNA^{Tyr}-tRNA^{Tyr}-tRNA^{Ser}. Chromosome 2 harbours 75 % of the mitochondrial-like tRNA genes. Nuclear encoded tRNA genes are under a PolIII promotor type 2 and transcribed by polymerase III.

The mitochondrial genome codes for all tRNAs except for tRNA^{Ala,Val,Arg,Thr,Leu,Phe}. Those tRNAs are nuclear encoded and imported into mitochondria from the cytosol (Duchêne *et al.*, 2009, Schneider & Maréchal-Drouard, 2000). tRNAs in mitochondria are transcribed by two nuclear-encoded phage-type RNA polymerases (NEP) (Hedtke *et al.*, 1997). Polycistronic transcription is common in mitochondria and is not tightly controlled (Fig. 1.6) (Holec *et al.*, 2006).

The plastidial genome is very compact and encodes all tRNAs necessary for the protein translation machinery (Michaud *et al.*, 2011). It has been shown in tobacco that a knock-out of plastidial tRNA^{Asn,Cys} genes is deleterious and not compensated by tRNA import from the cytosol (Legen *et al.*, 2007, Michaud *et al.*, 2011). The tRNA transcription is catalyzed by two types of RNA polymerases: plastid-encoded eubacterial-type RNA polymerase and NEP (Hedtke *et al.*, 1997).

The full set of annotated tRNAs in *Arabidopsis* was retrieved from the plantRNA server (Cognat *et al.*, 2012). tRNAs were annotated with 50 nts upstream of the mature tRNA transcript. All nuclear, mitochondrial and plastidial tRNA 5' upstream sequences were aligned separately using the Clustal webserver via the Jalview graphical interface (Troshin *et al.*, 2011, Waterhouse *et al.*, 2009) and nucleotide conservation was highlighted using weblogo (Crooks *et al.*, 2004). No sequence pattern or motifs can be observed in the mitochondrial and plastidial 5' sequences (Fig. 1.5b,c).

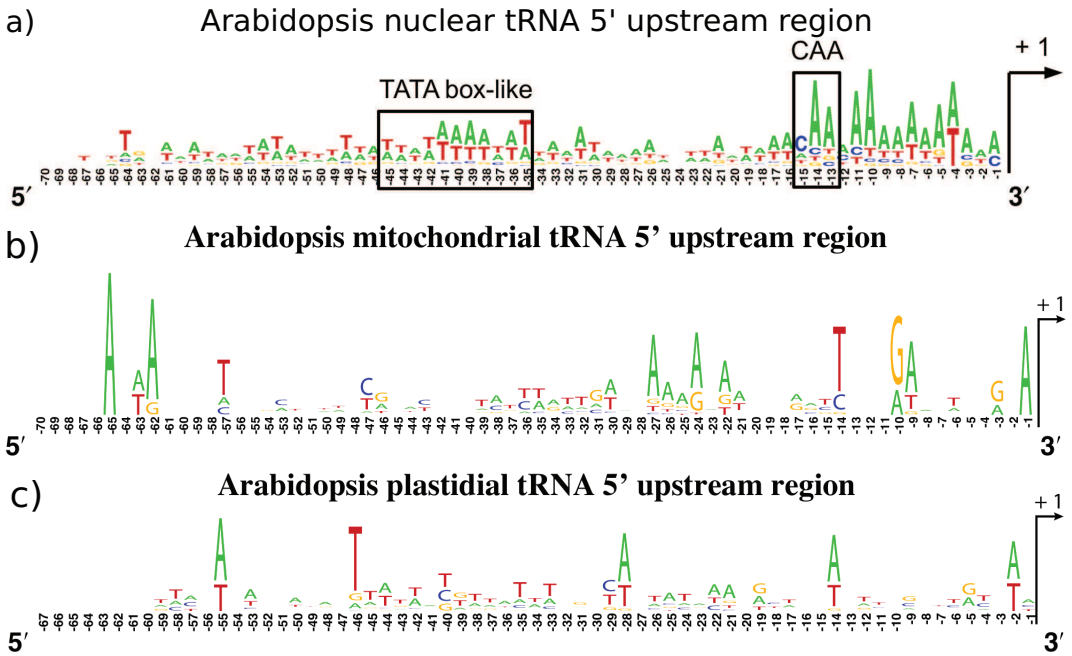


Figure 1.5: 5' upstream sequences of *A.thaliana* tRNA genes. Framed regions correspond to regions recognized by Pol III: TATA-box like elements and a CAA motif. +1 marks the start of the mature tRNA transcript.

Plant mitochondrial genomes passed several events of genomic rearrangements leading to the loss of promotor regions in front of each transcription unit. Thus, the distribution of mitochondrial promo-

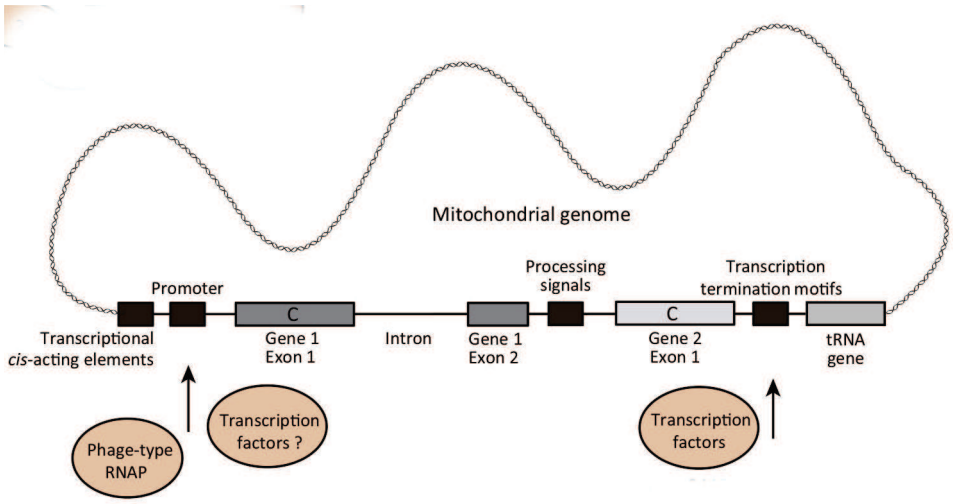


Figure 1.6: A typical *Arabidopsis* mitochondrial transcription unit with polycistronic transcripts and interspersed transcription termination and processing signals (Hammani & Giegé, 2014).

tors is poorly conserved leading to polycistronic transcripts (Hammani & Giegé, 2014). In contrast, the nuclear 5' upstream sequences show a clear pattern as described above, comprising a AT-rich region and a more or less conserved CAA motif some nts upstream the mature tRNA that mark the transcription initiation site. Still the positioning of these motifs differ from those described: the CAA motif found in Arabidopsis at -15 to -13 and the TATA-box like region in a region of -45 to -35 (Fig. 1.5a). All together plant nuclear RNA polymerase III promotors share similar features as their animal counterparts.

Mitochondrial tRNA^{Cys} The precursor tRNA I used during my Ph.D. work is a mitochondrial cysteinyl tRNA consisting of a five nt long leader and no CCA but the discriminator base at the 3' end (Schattner *et al.*, 2005). This precursor tRNA was first chosen in the lab as *in vitro* transcription levels were high and large amounts of pure RNA easily obtained.

1.2 tRNA maturation steps

A primary tRNA transcript will undergo several post-transcriptional maturation events such as 5' and 3' end cleavages, intron splicing and nucleotide modifications as illustrated in Fig. 1.3. Some steps are universally conserved, while others are specific to or absent in certain organisms.

1.2.1 Prokaryotic tRNA maturation steps

The best understood system of tRNA maturation is that from *E.coli*. 5' end maturation is done by a ribonucleoprotein, called RNase P which is described in more detail in section 1.3.1.1. 3' end maturation depends on several exoribonucleases, i.e. RNase PH, RNase T, RNase II, RNase D, RNase BN. This processing step was described to depend on the precursor tRNA as well as on stochastic events (Deutscher, 1984). RNase P acts most efficiently on tRNAs with short 3' trailer sequences. In *B. subtilis* where, in contrast to *E.coli*, one third of the tRNA genes have no CCA, two pathways co-

exist to process the 3' end termini: An exonucleolytic pathway for precursors already containing the CCA motif that resembles that of *E.coli* and an endonucleolytic pathway for tRNA molecules without encoded CCA motif. In the latter case the trailer is cleaved by RNase Z (homologous to *E.coli* RNase BN) (Wolin & Matera, 1999). The most important RNase activities in *E.coli* and *B.subtilis* are illustrated in Fig. 1.7.

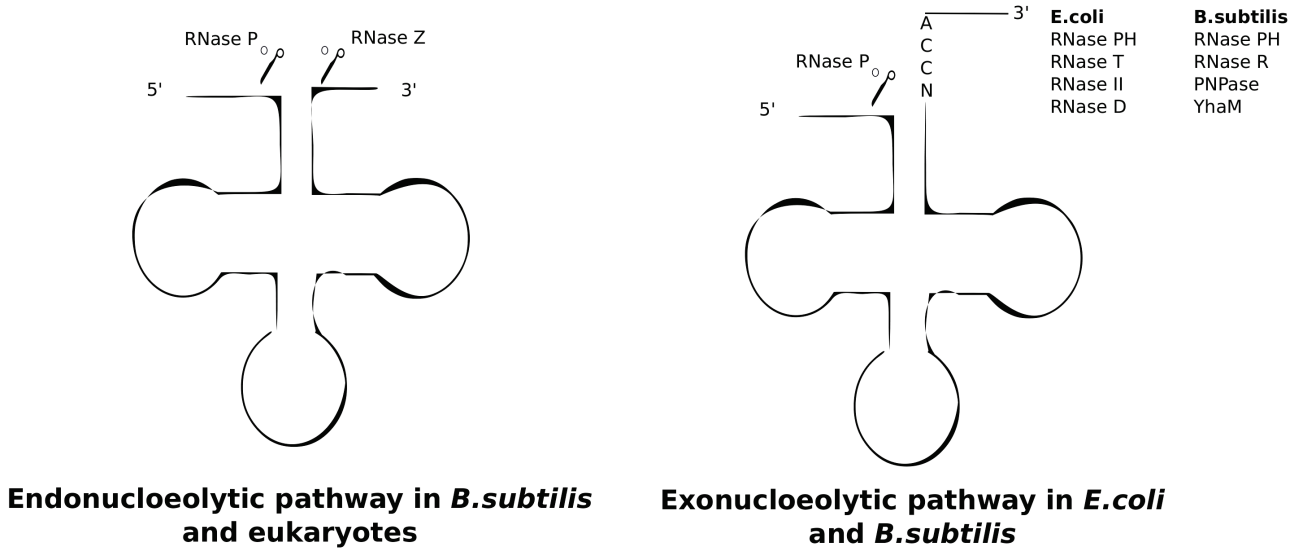


Figure 1.7: Prokaryotic tRNA processing in *E.coli* and *B.subtilis*. Two pathways of tRNA maturation co-exist in *B.subtilis* that are employed depending on the nature of the tRNA transcript. N = discriminator base.

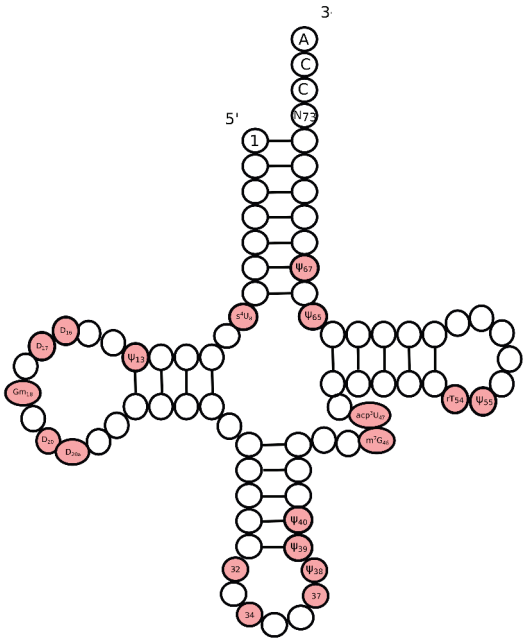


Figure 1.8: tRNA nucleotide modifications in *E.coli*. s^4U - 4-thiouridine, Ψ - pseudouridine, D - dihydrouridine, Gm - 2'-O-methylguanosine, m^7G - 7-methylguanosine, acp^3U - 3-(3-amino-3-carboxypropyl)uridine, rT (m^5U) - 5-methylguanosine.

Almost all post-transcriptional base modifications are known in *E.coli* today. They help folding tRNAs into a correct L-shape or are essential for tRNAs to interact with a variety of other molecules

in translation, degradation or editing (El Yacoubi *et al.*, 2012). Most of the tRNA modifications are found at position 34 and 37 in the anticondon-stem-loop (Fig. 1.8). Specific modifications allow to be stringent enough to discriminate between closely related codons and relaxed enough to accept more than one codon (El Yacoubi *et al.*, 2012). Modifications in base 37, directly next to the anticodon, ensure the stabilization of the first base pair of the anticodon triplet.

1.2.2 Eukaryotic tRNA maturation steps

1.2.2.1 5' end maturation

The first maturation step after tRNA transcription is the removal of the 5' leader of the tRNA transcript, a process that is catalyzed by an enzyme called ribonuclease P (RNase P). In prokaryotes, fungi and animal nuclei this enzymes is composed of a catalytic RNA component supported by one up to ten additional protein subunits. In plants and human mitochondria this endonucleolytic activity is held by an enzyme devoid of RNA, made only of protein. It is called proteinaceous RNase P (PRORP) and is described in more detail below.

1.2.2.2 3' end maturation

Whereas 3' end maturation is catalyzed by exonucleases in *E.coli* and most bacteria, it is cleaved by an endonuclease called RNase Z in eukarya. This gene family was first described in 2002 as being a metallo-hydrolase containing a zinc ion and a fold of two parallel β-sheets flanked by two α-helices. There are two different forms of RNase Z: RNase Z^S of 280-360 amino acids and RNase Z^L of 750-930 amino acids. The long form of RNase Z is only found in eukaryotes (Rossmanith, 2012). In *A. thaliana* there are four isoforms of RNase Z. Two isoforms of RNase Z^S and two of RNase Z^L. RNase Z^{S1} is localized to the cytosol and RNase Z^{S2} is localized to the chloroplasts. RNase Z^{L2} was shown to be targeted to mitochondria. A dual localization to mitochondria and the nucleus was reported for RNase Z^{L1}. Marchfelder and co-workers furthermore illustrated that only the chloroplastidial RNase Z^{S2} deletion mutant was lethal (Canino *et al.*, 2009). They speculated about a nuclear back-up system as a deletion mutant of RNase Z^{L1} was not lethal and showed only little phenotype.

1.2.2.3 CCA addition

CCA addition to tRNA molecules is a prerequisite for tRNA-aminoacylation and is thus a crucial step in tRNA maturation. In eukarya these three nucleotides are not gene encoded and have to be added post-transcriptionally. The enzymes responsible for this activity are called tRNA nucleotidyltransferases, or CCA adding enzymes. They are special as they are able to catalyze the sequential addition of a C-C-A triplet to the 3' end of a tRNA without any DNA or RNA template (Betat *et al.*, 2010).

1.2.2.4 Nucleotide modifications

One of the most important features of tRNAs is their large number of post-transcriptional modifications with a mean value of eight modifications per tRNA species. There are approximately 85 known

modifications in tRNAs (El Yacoubi *et al.*, 2012). Whereas many modifications in the anticondon-stem-loop region affect translation and decoding, modifications in the core body structure are related to stability and structural integrity (Phizicky & Hopper, 2010).

1.2.2.5 Splicing and other factors

Introns in tRNA sequences have been reported in all three domains of life. In bacteria these introns are removed by self-splicing (Reinhold-Hurek B., 1992). In eukarya and archaea these sequences are removed by enzymatic splicing in two steps: endonucleolytic cleavage, removal of the intron and site specific ligation. Several essential snRNAs called U1, U2, U4, U5 and U6, together with distinct proteins form snRNPs. These particles in addition with several other splicing factors form an eukaryotic spliceosome (L. Stryer, 2007). About 70 % of tRNA genes in archaea contain up to three introns. In eukarya introns are less abundant (6 % of tRNA genes contain introns) and shorter. Splicing occurs in the cytosol after the 5' and 3' end maturation (Wolin & Cedervall, 2002). The endonucleolytic excision of the intron is similar to that in archaeal and in eukaryal mRNAs leading to a linear excised intron and a 5' tRNA half with a 2',3' cyclic phosphate at the 3' end and a 3' tRNA half with a 5' hydroxyl. Archaeal and eukaryotic splicing endonucleases are evolutionary related although they evolved different splice site recognition modi (Li *et al.*, 1998). The ligation step that follows the cleavage is done by different enzymes in different organisms using distinct mechnisms. The primary goal is to hydrolyse the cyclic phosphate on the 5' splice half and to phosphorylate the 3' splice half in order to obtain a classical phosphodiester bond. In plants the 2',3' cyclic phosphate is hydrolyzed by a cyclic phosphodiesterase resulting in a 2'-phosphate 3' end of the 5' half. In a next event the 5' hydroxyl is phosphorylized. Subsequently, an ATP is transferred to the protein, then to the 5' phosphate of the 3' exon. In the final ligation step the AMP is released and the phosphodiester linkage is formed (Popow *et al.*, 2012)

1.2.2.6 The La protein

The La protein was first described in human cells and is the first enzyme that recognizes RNA PolIII primary transcripts and other transcripts via interaction of the 3'-poly-U tail. It fulfils the function of a chaperone by stabilizing the 3D structure of pre-tRNAs. In *Arabidopsis* there are two La proteins and deletion of this protein leads to embryolethality (Fleurdepine *et al.*, 2007). Its function is to protect nascent tRNAs and other small RNA molecules from 3' exonuclease degradation and to coordinate pre-tRNA maturation events (Wolin & Cedervall, 2002). Yet in *S. cerevisiae* and *S. pombe* La is not essential and cells lacking the protein can be studied. There are two major pathways of pre-tRNA maturation depending on the presence or absence of La. The sequence of 5' and 3' end maturation is completely altered: in wild type yeast cells La fixes the 3' end of a nascent tRNA transcript and RNase P cleaves the 5' end first followed by the endonucleolytic cleavage of the 3' trailer. In cells lacking the La protein, exonucleases first chop the 3' end and only in a second event RNase P will cleave the 5' leader (Fig. 1.9).

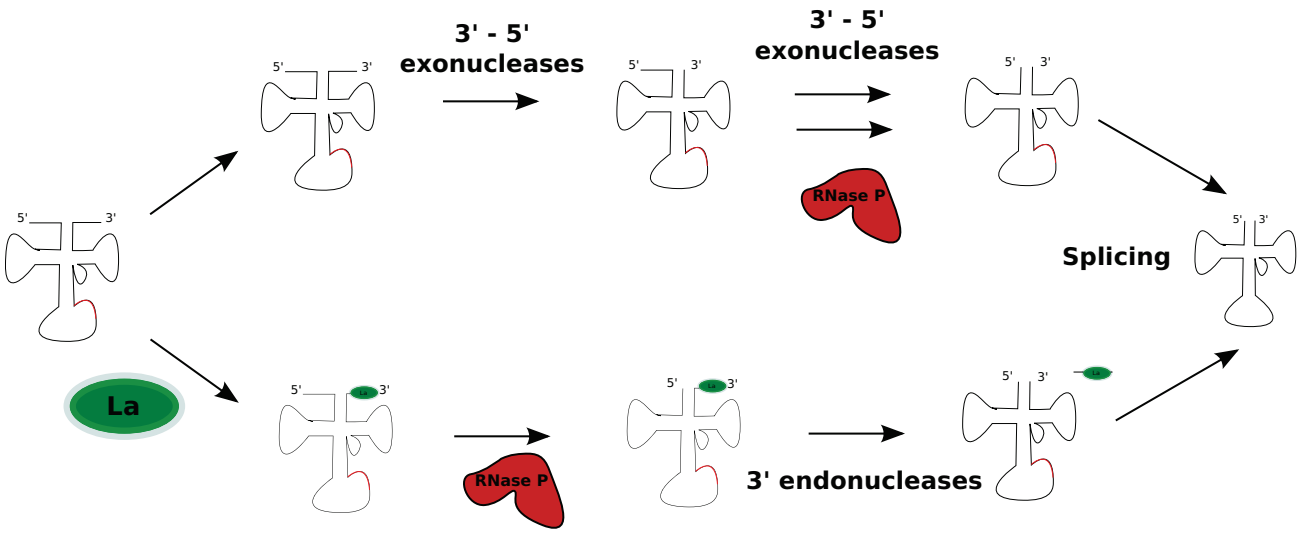


Figure 1.9: The La protein - two pre-tRNA maturation pathways in yeast depending on the presence or absence of La. Its presence on the 3' end leads to initial 5' maturation by RNase P while without La the 3' end will be chopped first by exonucleases and only afterwards the 5' end is processed.

1.3 5' end maturation of tRNAs in different domains of life

The 5' end maturation is an essential step in tRNA biosynthesis. In many organisms the endonucleolytic cleavage of the 5' leader from the tRNA transcript is performed by RNase P. RNase P was first discovered by Sidney Altman in 1969 (Stark *et al.*, 1978). Over more then 15 years Altman and coworkers characterized this enzyme family and unravelled the functional relationship between 5' end maturation, RNA catalysis and RNase P in bacteria (McClain & Lai, 2010). RNase P is an enzyme requiring divalent metal ions such as magnesium. It cleaves the pre-tRNA into the 5'-leader with a 3'-OH and the tRNA with a 5'-monophosphate (Fig. 1.10).

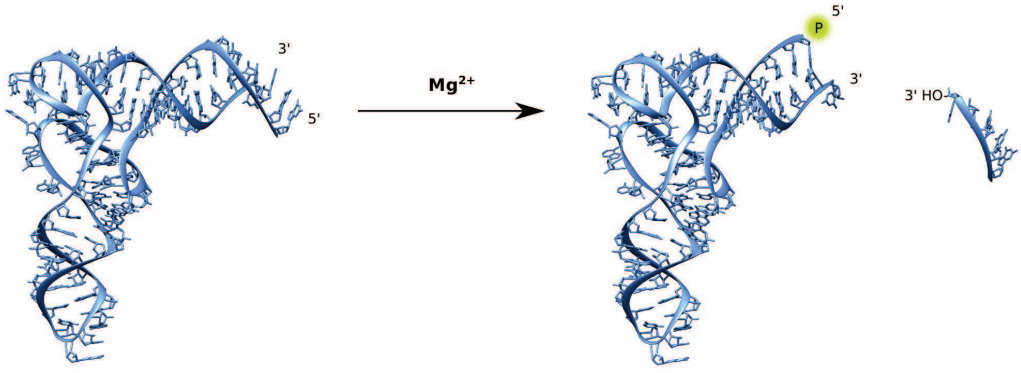


Figure 1.10: Schematic view of RNase P function and 5' end maturation of pre-tRNA^{Cys} in the presence of divalent ions e.g. Mg^{2+}, Mn^{2+}. The cleavage leads to a 5' monophosphate at the tRNA and a hydroxyl group at the 3' end of the leader.

In bacteria and archaea RNase P consists of one large RNA (276 - 400 nts) entity and of one to five protein subunits. In eukaryotes the ribonucleoproteic RNase P contains even more protein subunits and the respective percentages of protein in the holoenzyme masses for bacteria, archaea and eukary-

otes are 10%, 50% and > 70%. The complexity among RNase P proteins reviewed by Jarrous in 2010 is summarized in Table 1.1.

Table 1.1: Examples of RNA and protein contents of RNase P in bacteria, eukaryotes and archaea. Numbers in brackets are mean masses in kDa. Proteins in the same row are homologous. RnpA = protein in bacterial RNase P; *S. cerevisiae* = *Saccharomyces cerevisiae*; *Pfu* = *Pyrococcus furiosus* (after Jarrous & Gopalan (2010), Evans *et al.* (2006)).

Bacteria	**Archaea, *Pfu***	***S. cerevisiae***	***Homo sapiens s.***
RNA (121)	RNA (106)	RNA (118)	H1 RNA (109)
RnpA (13.8)			
	Pop5 (13.8)	Pop5 (19.6)	hPop5 (18.8)
	Rpp30 (24.5)	Rpp1 (32.2)	Rpp30 (29.3)
	Rpp21 (14.3)	Rpr2 (16.3)	Rpp21 (17.6)
	Rpp29 (15)	Pop4 (32.9)	Rpp29 (25.4)
	Rpp38 (13.2)	Pop3 (22.6)	Rpp38 (31.8)
		Pop1 (100.5)	hPop1 (114.7)
		Pop7 (15.8)	Rpp20 (15.7)
		Pop6 (18.2)	hPop6 (Rpp25) (20.6)
		Pop8 (15.5)	hPop8 (Rpp14) (13.7)
			Rpp40 (34.6)

1.3.1 Ribonucleoproteic RNase P

1.3.1.1 Bacterial RNase P

Bacterial RNase P consists of a 300 to 400 nt long RNA (P RNA) and one small protein (Brown & Pace, 1992). It is this RNA entity that holds the catalytic activity of the holoenzyme complex, which is a ribozyme. There are two types of secondary structures of bacterial P RNA (Fig. 1.11). The first group, called ancestral type (A type), is found in *E. coli*. The second type called bacillus type (B type) occurs in *Bacillus subtilis* (Walker & Engelke, 2006). Despite differences in secondary structure elements both types fold into similar tertiary structures. A minimal theoretical P RNA was proposed containing all conserved structural elements of several hundred of sequenced bacterial P RNAs which is only 225 nts long (Fig. 1.11) (Brown *et al.*, 1991). All other structural features, present in native P RNAs would serve to increase thermal stability or decrease the ionic strength needed for catalytic activity. The smallest, so far known P RNA is present in *Mycoplasma fermentas* and consists of only 276 nts (Siegel *et al.*, 1996). All bacterial RNase P RNAs share two independently folded domains:

1. The specificity domain (S-domain) which recognizes the T-stem loop,
2. The catalytic domain that cleaves the pre-tRNA and recognizes the 5' leader, the acceptor stem and the 3' CCA end (Jarrous & Gopalan, 2010).

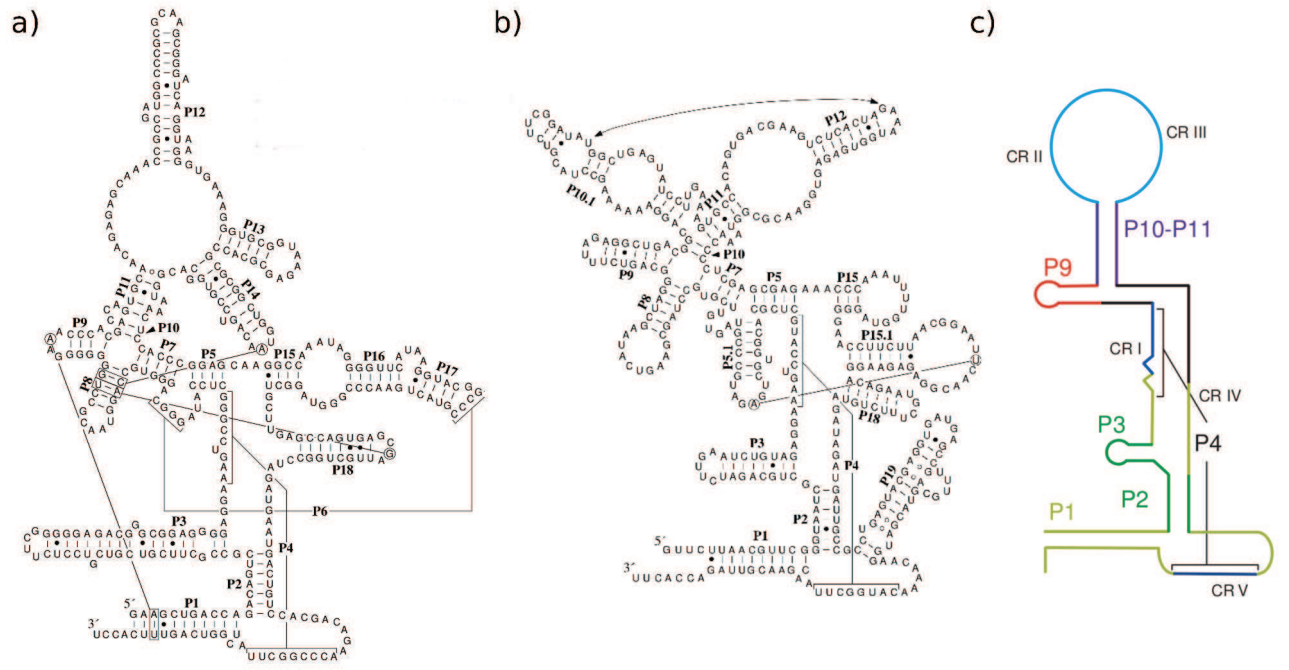


Figure 1.11: Structural elements of prokaryotic RNase P RNA. Secondary structure of a) type A RNase P RNA from *E. coli*, b) type B RNase P RNA from *B. subtilis*, c) minimal RNA elements (Evans *et al.*, 2006, Brown, 1997).

The conserved core of secondary structural elements shown in Fig. 1.11c is shared by (almost) all organisms and is essential for the catalytic function of RNase P. Universally conserved regions are called CR I-V. The core is built by one of the most conserved substructures, regions CR I and CR V, that base pair to form helix P4. Furthermore, it contains CR II and III that are loop regions between helices P12 and P10/P11 (Evans *et al.*, 2006). The core structure is sufficient to cleave pre-tRNAs *in vitro*, but it lacks stabilizing elements that are required for *in vivo* tRNA cleavage. The main differences between type A and type B P RNA are the presence of P16, P17 and P6 and the lack of P5.1, P10.1, P15.1 and P15.2 in type A.

Albeit the P RNA is active alone *in vitro*, *in vivo* RNase P is a ribonucleoprotein that contains at least one protein. In the well-studied bacterium *E.coli* it is called C5 (119 aa, 13.8 kDa) (Tsai *et al.*, 2003). The function of this protein is to enhance the affinity for its substrates, the cleavage rate and the fidelity by stabilizing the catalytic active conformation of the ribozyme (Jarrous & Gopalan, 2010). Furthermore, it helps discriminate between substrate and product and mediates to holoenzyme dimerization (Evans *et al.*, 2006, Fang *et al.*, 2001).

Beside tRNA molecules, bacterial RNase P can cleave other substrates like pre-4.5S RNA, pre-transfer messenger RNA, polycistronic tRNAs, mRNA and riboswitches and some bacteriophage RNA (Alifano *et al.*, 1994, Peck-Miller & Altman, 1991, Hartmann *et al.*, 1995, Altman *et al.*, 2005, Mans *et al.*, 1990, Komine *et al.*, 1994). Crystallographic structures of A- and B-type RNase P show striking similarities in structure and surface charges but low sequence identity (20-30 %) (Reiter *et al.*, 2010, Evans *et al.*, 2006). This indicates an evolutionary pressure to conserve theses features in order to maintain RNA protein interaction. In 2010 the first holoenzyme complex of an A-type RNase P was reported from the thermophilic bacterium *Thermotoga maritima*. The overall fold of this protein

is an α/β-sandwich fold with a conserved core of four β-sheets interacting with the 5' leader of the tRNA but not with the G-1 of the mature tRNA. The protein is in contact with the P RNA through P15 and P3 stem and in CR IV/V regions (Fig. 1.12) (Reiter *et al.*, 2010).

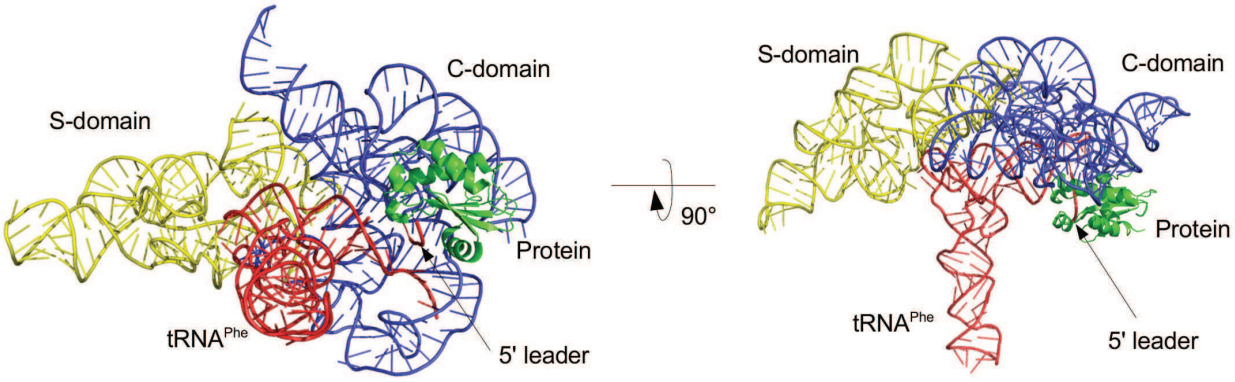


Figure 1.12: Crystal Structure of RNase P holoenzyme of *Thermotoga maritima* with tRNA^{Phe} at a resolution of 3.8 Å(pdb ID: 3q1r) The C-domain is represented in dark blue, the S-domain in yellow, the tRNA^{Phe} with the 5' leader (5 nts) in red and the RNase P protein in green.

1.3.1.2 Archaeal RNase P

Archaea are single-cellular organisms possessing a circular DNA molecule. They were described by Carl Woese and George Fox in the 1960s. Based on the comparison of the small subunits of rRNAs (characterized by a low mutation frequency) they concluded that eukaryotes, bacteria and archaea belong to three different primary kingdoms (Woese & Fox, 1977). Finally, in 1990 archaea were given their own domain (Cavicchioli, 2011). With eukaryotes they share for example the replication, transcription and translation machinery. They often possess extreme properties like the ability to live in high temperatures or in highly acidic environments.

Archaeal RNase P consists of one RNA molecule and four to five protein subunits (Hall & Brown, 2002, Esakova & Krasilnikov, 2010, Jarrous & Gopalan, 2010). The P RNA is classified into two different folds: type A that is similar to bacterial type A and type M that is less complex . Type A (Fig. 1.13a) shows traces of activity *in vitro* under high salt conditions in the absence of proteins whereas type M (Fig. 1.13b) shows no catalytic activity on its own *in vitro* (Walker & Engelke, 2006) and seems to be more related to the eukaryotic P RNA.

All archaeal RNase P proteins have eukaryotic homologues (Esakova & Krasilnikov, 2010). APop4 (hRpp29) and aRpp2 (hRpp21) are in contact with the S-domain of the P RNA and confer to substrate affinity. The other protein pair, aPop5 (hPop5) and aRpp1 (Rpp30), enhance the catalytic activity and are implicated in the cleavage and product release. APop3 (hRpp38) might increase thermal stability. The detection of RNase P RNA in *Nanoarchaeum equitans, Aquifex aeolicus, Pyrobaculum* species and related organisms failed. *N. equitans* turned out to transcribe tRNAs under a mature form whereas tRNAs from *Pyrobaculum* indeed are produced with a 5' leader sequences. The purified extract of *Pyrobaculum* had RNase P activity. Still, a genome-wide detection of RNase P RNA failed. Three out of four P proteins could be detected: Rpp29, Rpp30 and Pop5. Rpp21 that is involved in binding to

the S-domain could not be detected. A thorough search of P RNA led to the discovery of the P RNA lacking the S-domain but still containing the conserved C-domain. This truncated P RNA was called type T P RNA (Fig. 1.13c). This type of reduced RNase P might be an intermediate form towards an organism lacking RNase P activity as is the case in *N. equitans* (Lai *et al.*, 2010).

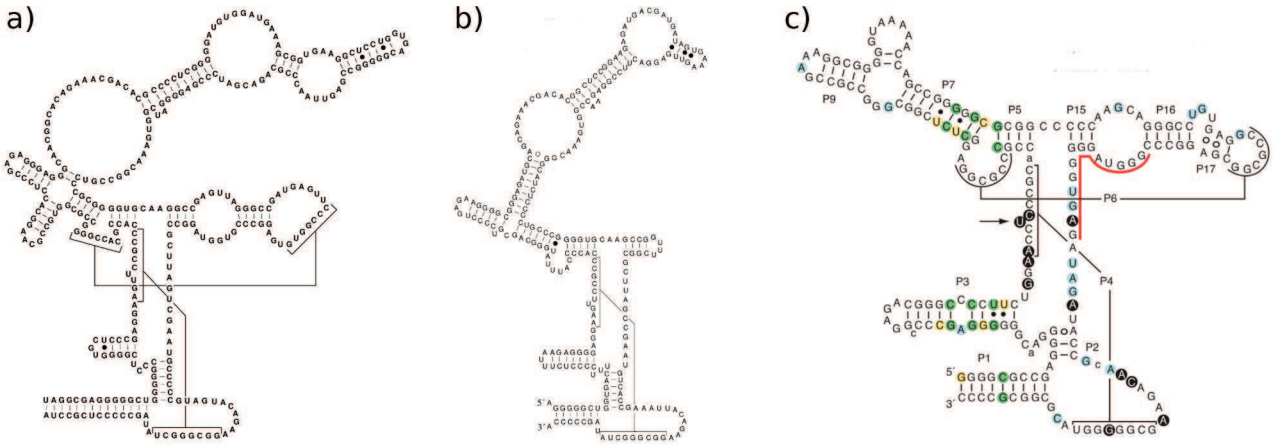


Figure 1.13: Secondary structure of archaeal RNase P RNA. a) P RNA type A from *Pyrococcus horikoshii*, b) P RNA type M from *Methanocaldococcus jannashii*, c) P RNA type T from *Pyrobaculum aeophilum*

1.3.1.3 Eukaryotic RNase P

Nuclear RNase P from yeast RNase P from yeast nuclei consists of a RNA component (RPR1 RNA) (Fig. 1.14a) and nine additional protein subunits with a pI higher than 9 that are essential for yeast viability, (Tab. 1.1) (Xiao *et al.*, 2002). Depletion of one of the RNase P proteins, except Pop3p, leads to a reduction of mature RPR1 RNA which suggests either maturation process of the pre-RPR1 RNA in the holoenzyme complex or its destabilization. RPR1 RNA is transcribed by RNA polymerase III as a 486 nts long transcript. The 5' end is processed with a removal of 84 nts whereas the 3' end is trimmed several times to remove 33 nts to result in a 369 nts long RNA (Xiao *et al.*, 2001).

Eukaryotic RNase P RNA sequences are 30% shorter than bacterial or archaeal ones. The reduction in 2D structural elements is compensated by protein components that stabilize the holoenzyme. Until 2007 it was not clear whether the RNA had a catalytic function by its own in either of these holoenzymes. Willkomm *et al.* demonstrated that RNA retains its ability to specifically cleave its substrate in the absences of proteins *in vitro* (Willkomm & Hartmann, 2007, Kikovska *et al.*, 2007). Yeast RNase P contains nine proteins in total (Esakova & Krasilnikov, 2010). The human nuclear RNase P consists of H1 RNA and 10 proteins of which seven are homologous to yeast RNase P proteins (Tab. 1.1).

RNase P in metazoa nuclei Human nuclear RNase P is composed of H1 RNA (Fig. 1.14b) and at least 10 distinct proteins: Rpp14, Rpp20, Rpp21, Rpp25, Rpp29, Rpp30, Rpp38, Rpp40, hPop1 and hPop5 (Tab. 1.1) (Jarrous & Reiner, 2007, Jarrous, 2002). Eight out of 10 proteins of this complex have been shown to interact with the H1 RNA *in vitro*. While bacterial and archaeal RNase P RNA

can cleave a precursor tRNAs *in vitro* under high salt and Mg^{2+} conditions, human nuclear RNase P RNA lacks this ability. Its activity is restored upon addition of Rpp21 and Rpp29. Protein subunits Rpp20, Rpp21, Rpp25, Rpp29, Rpp30, Rpp38 and hPop1 but not Rpp14 and hPop5 interact with the H1 RNA. The enzymatic activity of hRNase P is regulated by the La protein that binds to the 5' pppG/A and 3' UUU-OH extremities. After phosphorylation of the La protein the latter dissociates from the 5' end opening the space for RNase P processing (Jarrous, 2002).

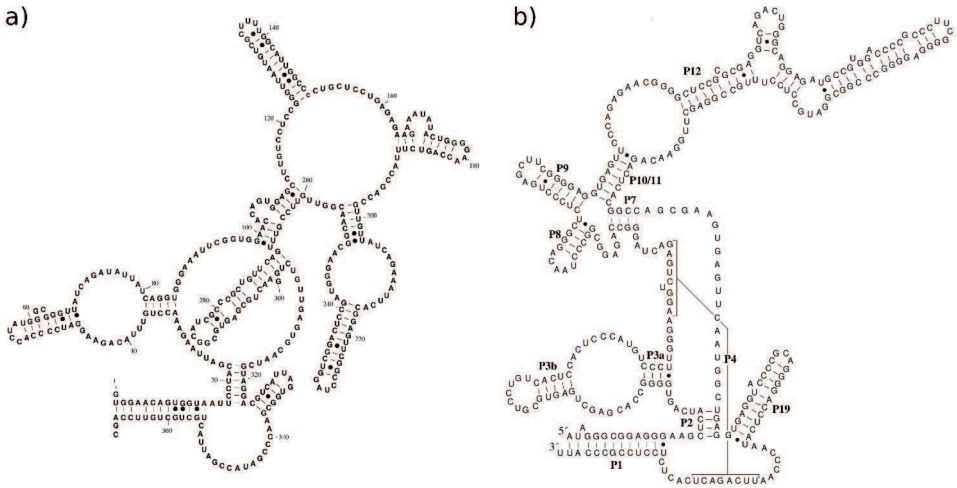


Figure 1.14: Secondary structure of eukaryotic nuclear RNase P RNA. a) P RNA type A from *S. cerevisiae*, b) P RNA from *H. sapiens*.

Mitochondrial RNase P in yeast Purified mtRNase P from *S. cerevisiae* contains the nuclear encoded protein Rpm2p (105 kDa) and an incomplete Rpm1 P RNA that is encoded in the mitochondrial genome. In the active RNase P complex two RNA fragments of the 5' end (90 nts) and of the 3' end (70 nts) original Rpm1 P RNA are present and suffice for pre-tRNA cleavage (Daoud *et al.*, 2012, Vioque, 2010). This RNA contains some conserved structural elements from the bacterial M1 RNA, i.e. P1, P4 and P18. Rpm2 has no sequence homology with any of the bacterial, archaeal or eukaryotic nuclear RNase P proteins. It was reported that mtRNase P from *Aspargillus nidulans* contains seven polypeptides (Vioque, 2010).
As we will see in the next section the existence of a homologous system in human mitochondria has long been an open debate.

1.3.2 Proteinaceous RNase P

Nuclear RNase P of *S. cerevisiae* is the best studied eukaryotic RNase P (Vioque, 2010). Substantial information on organellar RNase P in yeast and other eukaryotes is still missing. In neither higher plants, nor metazoa mitochondrial or plastidial genomes a rnpB gene could be detected. Likewise characterization of the protein content of such a RNP RNase P is still missing (Vioque, 2010). For a long time the 5' end maturation was believed to be catalyzed by RNA-dependent enzymes in all three domains of life. Two interesting scientific debates occurred around this enzyme: More than 40 years ago, Sidney Altman and co-workers had difficulty to find acceptance for the existence of catalytic active RNAs which are responsible for the 5' end maturation of tRNAs. 30 years later other scientists

encountered the same criticism to proof that this endonuclease activity is also held by protein-only enzymes and that this enzyme is not an universally conserved ribozyme. All analyzed RNase P complexes so far consisted of a RNA and one to ten proteins. As shown above, the protein amount increased from bacteria to archaea to eukaryotes, reflecting the growing importance of proteins.

1.3.2.1 Animals: First proof of protein-only RNase P in human mitochondria

The mystery of mitochondrial RNase P from H. sapiens It was reported that human mitochondrial RNase P contains H1 RNA the nuclear P RNA (Puranam & Attardi, 2001, Doersen *et al.*, 1985) and that this activity is responsible for mitochondrial pre-tRNA processing. These findings were later critically discussed as Rossmanith *et. al.* demonstrated that purified human mitochondrial RNase P cannot cleave *E.coli* pre-tRNA$^{Tyrsu3+}$ but mitochondrial pre-tRNA^{Tyr} (Rossmanith *et al.*, 1995). tRNA$^{Tyrsu3+}$ is a suppressor tRNA that can decode the stop codon UAG in *E.coli*.
The controversy is now whether a RNP RNase P and a proteinaceous RNase P co-exist in mitochondria as H1 P RNA was proposed to be imported into human mitochondria by polynucleotide phosphorylase (PNPase) and to have an impact on tRNA processing *in vivo* (Wang *et al.*, 2010, Mercer *et al.*, 2011). The hypothesis of the presence of RNase P RNA in mitochondria was critically analyzed by Rossmannith (Rossmanith, 2012) who outlines several points:

1. The amount of detected H1 RNA in mitochondria seems unlikely to be sufficient for pre-tRNA maturation.

2. None of the nuclear encoded P proteins associated with the H1 RNA contain a mitochondrial targeting sequence or could be detected in human mitochondrial proteomic data.

3. How can unspecific binding of PNPase to short stem-loop structures account for specific import of P and MRP RNA?

4. PNPase is supposed to interact with helix P9 of H1 RNA that was shown to be protected by proteins in footprinting analyses. How would then PNPase have access to this recognition site?

5. The localization of PNPase in the intermembrane space is dubious as its role in mitochondrial RNA metabolism suggest a matrix localization.

6. Is the accumulation of some mitochondrial precursor tRNA in PNPase deficient cells due to deficient RNA degradation or a lack of RNase P processing?

To conclude it seems not likely that RNP RNase P and proteinaceous RNase P co-exist in human mitochondria.

Existence of protein-based RNase P activity was further studied by Walter Rossmanith and co-workers who were the first to overexpress and purify human mitochondrial RNase P and demonstrate its proteinaceous nature *in vitro* (Rossmanith & Karwan, 1998, Holzmann *et al.*, 2008). They characterized the enzyme as a complex of three proteins and named these proteins mitochondrial ribonuclease P

protein 1-3. By sequence homology, functions could be assigned to two components: MRPP1 as a RNA (guanine-9-)methyltransferase containing domain 1 and MRPP2 as a type 2 3-hydroxyacyl-CoA dehydrogenase. MRPP1 or now tRNA methyltransferase 10 C (TRM10C) is the enzyme responsible for the methylation of purines at position 9 of mitochondrial tRNAs. This position is an A or G in 19 out of 22 mitochondrial tRNAs. Interestingly this methylase only functions in a complex with MRPP2, a short-chain dehydrogenase/reductase that uses NAD(H) as a co-factor. This indicates that human mitochondrial RNase P is a multifunctional protein complex. Both, precursor and mature tRNAs serve as substrates for methylation but 5' end maturation and methylation are not coupled processes (Vilardo *et al.*, 2012). Catalytic mutants of MRPP2 proved that dehydrogenase activity is dispensable for methylation and 5' end maturation. Among the three proteins a subcomplex of two is thought to confer to RNA binding, i.e. MRPP1/TRM10C and MRPP2/SDR5C1. The catalytic activity lies in the third protein subunit not characterized at that time: MRPP3 (human PRORP). However, all three proteins are required to form an active RNase P complex. MRPP3 contains a mitochondrial targeting signal and a degenerate N-terminal pentatricopeptide repeat domain (PPR). The C-terminal domain contains four conserved residues (three aspartates and one histidine) that are commonly found in metal-ion based nucleases (Dupureur, 2008). Structural information is lacking due to the low abundance of the three components, to their dynamic association properties and to the difficulty to crystallize human complex proteins (Esakova & Krasilnikov, 2010). Whereas the subcomplex of MRPP1/MRPP2 is stable, the holoenzyme complex with hPRORP seems to be more labile.

1.3.2.2 Plants: Early work on spinach chloroplasts

The role of RNA in eukaryotic organellar RNase P activity has always been controversial. In 1988 Peter Gegenheimer and co-workers were the first to propose a protein-only based RNase P activity in spinach chloroplast. They treated chloroplast extracts with micrococcal nuclease (MN) to test RNase P activity for its RNA content. RNase P activity was not abolished after MN treatment. Thoroughly they examined all possibilities and concluded the absence of a RNA being responsible for pre-tRNA cleavage (Wang *et al.*, 1988, Thomas *et al.*, 2000). This type of RNase P, called PRORP, is now well characterized in *A. thaliana* and described in more detail in section 1.6.

1.4 Life without RNase P

In the archaebacterium *Nanoarchaeum equitans* no RNase P activity could be detected, neither by biochemical nor by bioinformatical means (Randau *et al.*, 2008). It possesses the smallest sequenced genome. It was found that the tRNA genes have a promotor whose position is strictly conserved at -26 nts upstream to the mature tRNA and that will lead to 5' mature transcripts, with three exceptions: $tRNA^{Tyr,His,Meti}$ that have an additional G/A at position -1. This nucleotide is essential for aminoacyl-tRNA-synthetases and could not be produced if a RNase P would be present. It is now proposed that *N. equitans* lost its tRNA 5' leader sequences (supported by the strict conservation of the transcription promotor) and made RNase P unnecessary.

1.5 What about RNase MRP

Eukaryotes possess an enzyme that is related to RNase P, called RNase MRP (mitochondrial RNA processing). Its absence in bacteria and archaea supports the theory of RNase P RNA gene duplication in early eukaryotic ancestors. RNase MRP is localized to two compartments: mitochondria and the nucleolus but doubts remain that the amount detected in the mitochondria would be sufficient for RNA cleavage (Lopez *et al.*, 2009). RNase MRP processes rRNAs and also degrades mRNAs. RNase MRP and RNase P RNA are related, more on a structural level (secondary structure) than on a sequence level (Fig. 1.15). They share similarities in the C-domain that comprises CR I, IV and V and is responsible for RNA cleavage (Fig. 1.15). The lack of similarity in S-domain might explain their distinct substrate specificities (Esakova *et al.*, 2008, Lopez *et al.*, 2009). Eukaryotic RNase MRP contains 10 proteins: Amongst them eight are homologous to the yeast nuclear RNase P proteins (Fig. 1.16) (Hernandez-Cid *et al.*, 2012). In *S. cerevisiae* RNase MRP and RNase P contain 10 and 9 protein subunits, respectively. Eight of theses proteins are shared by both enzymes, i.e. Pop1, Pop3-8 and Rpp1. Snm1 a RNA binding protein and Rmp1 are unique to RNase MRP and their deletion do not alter pre-tRNA or tRNA levels but hamper rRNA processing. Rpr2 (RNase P ribonucleoprotein-2) is the only protein that is unique to RNase P and its depletion resulted in a decrease of both precursor and mature RNase P RNA (Chamberlain *et al.*, 1998). In human RNase MRP and RNase P contain 9 and 10 protein subunits, respectively, from which 9 are shared by both enzymes, i.e. hPop1, Rpp38, Rpp29, hPop5, Rpp25, Rpp20, Rpp14, Rpp30, Rpp40. Rpp21 is unique to RNase P where it is implicated in binding to the S-domain of RNase P RNA and enhancing its substrate specificity (Xu *et al.*, 2012).

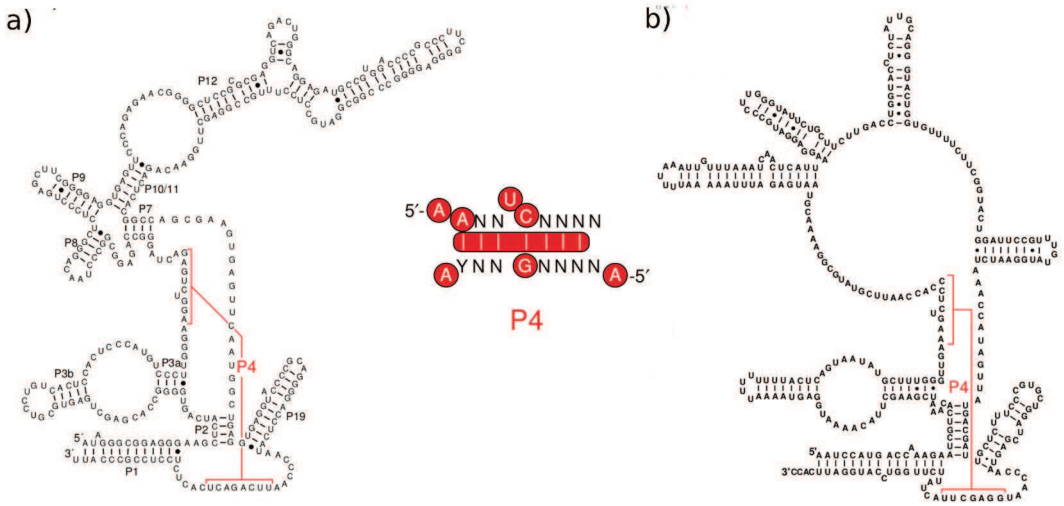


Figure 1.15: Comparison of MRP and RNase P RNA. a) H1 RNase P RNA in *H. sapiens*, b) MRP RNA in *S.cerevisiae*. Conserved residues among all MRP and RNase P RNAs in the P4 helical element are shown in the middle. (Hartmann & Hartmann, 2003)

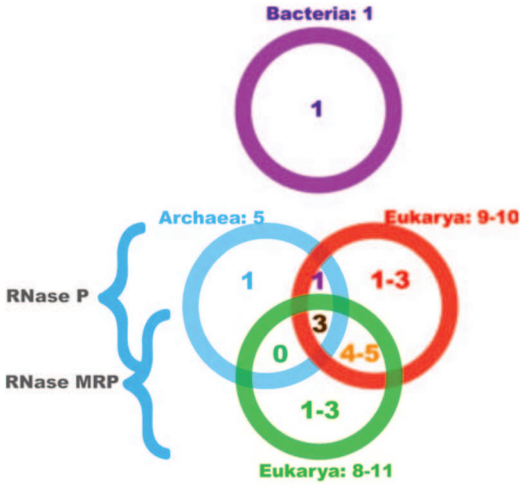


Figure 1.16: RNase MRP and RNase P homology and distribution in all domains of life. The numbers refer to the number of proteins in each enzyme. (Hernandez-Cid *et al.*, 2012)

1.6 Publication 1: PPR proteins shed a new light on RNase P biology

Proteinaceous RNase P (PRORP) belong to the vast family of pentatricopeptide repeat proteins which are major players in all processes of gene expression and translation. PRORPs catalyze the removal of the 5' leader of precursor tRNAs, a maturation step which was long thought to be universally processed by ribonucleoproteins with a RNA component as the catalytic moiety. PRORP enzymes constitute single proteins between a fusion of canonical PPR motifs and an C-terminal nuclease domain.

In *Arbabidopsis* PRORP1 is localized to the organelles and was shown to be essential *in vivo* whereas its paralogues PRORP2-3, both directed to the nucleus (Fig. 1.17), are redundant. Some characteristics of the three *Arabidopsis* homologs are summarized in Tab. 1.2. Downregulation of one of the nuclear PRORPs did not show any phenotype. Experiments using VIGS (virus induced gene silencing) for PRORP1 downregulation resulted in altered structures of chloroplasts and mitochondria. Other substrates than tRNAs were also demonstrated to be recognized by PRORP1. T-elements (tRNA-like structures) in the 3' region of mRNA *orf138* or *nad6* were cleaved by PRORP1 and thus accumulated in mutant *Arabidopsis* plants. Furthermore Giége and co-workers could demonstrate that nuclear RNase MRP is not responsible for tRNA 5' processing. Knock-out of Pop1 and Pop4, essential RNase MRP proteins, did not change mature tRNA levels but as expected increased unprocessed rRNA levels (Gobert *et al.*, 2010, Gutmann *et al.*, 2012).

Table 1.2: General properties of *Arabidopsis* PRORP proteins. Length - full length protein, aa - amino acids, pI - isoelectric point

	PRORP1	**PRORP2**	**PRORP3**
Length	572 aa	528 aa	576 aa
Mass	65 kDa	59 kDa	64 kDa
Function	tRNA, mRNA processing	tRNA, mRNA, snoRNA processing	tRNA, mRNA, snoRNA processing
Localization	Mitochondria, chloroplasts	Nucleus	Nucleus
pI	9.4	6.3	6.0

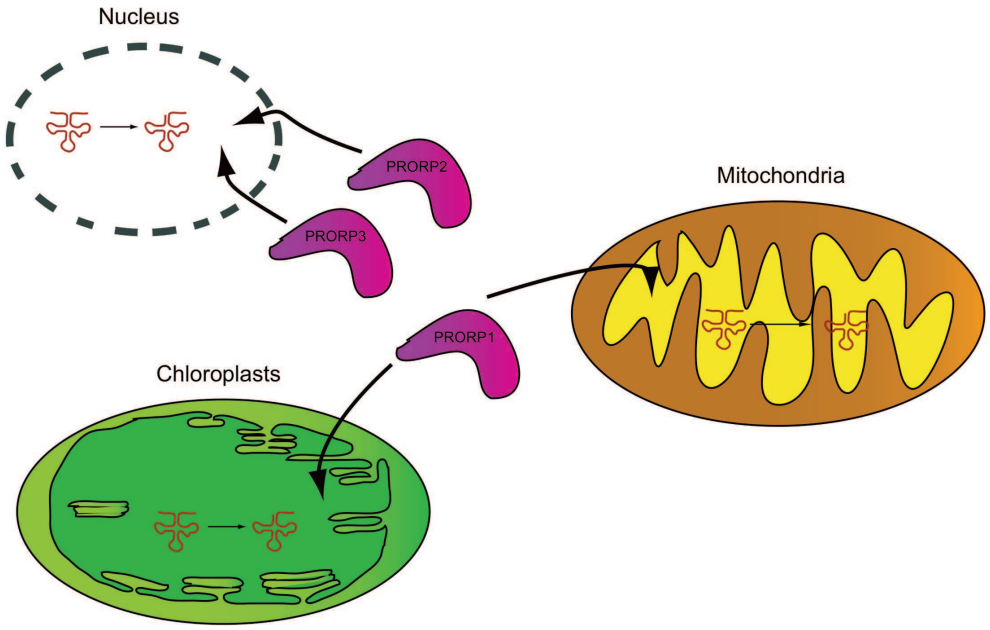


Figure 1.17: PRORP localization in *Arabidopsis*. PRORP1 containing an organellar targeting sequence is localized to mitochondria and chloroplasts while PRORP2-3 are localized to the nucleus.

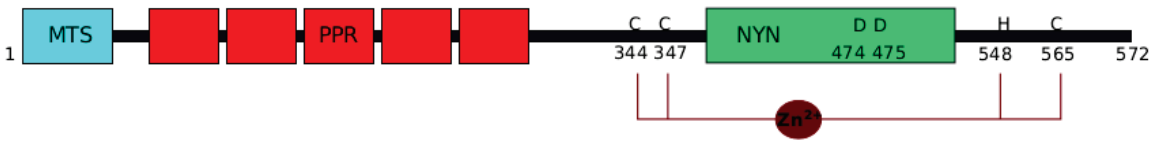


Figure 1.18: Schematic organization of PRORP1 from *A. thaliana*. MTS - mitochondrial targeting signal, PPR - pentatricopeptide repeat, NYN - **N**edd4-BP1, **Y**acP **n**ucleases. A zinc binding motif is structurally connecting the bipartite central domain of PRORP1 which is illustrated by the four zinc chelating residues C344, C347, H548 and C565.

REVIEW

RNA Biology 10:9, 1457–1468; September 2013; © 2013 Landes Bioscience

PPR proteins shed a new light on RNase P biology

Franziska Pinker[1,2], Géraldine Bonnard[1,†], Anthony Gobert[1,†], Bernard Gutmann[1,†], Kamel Hammani[1,†], Claude Sauter[2,†], Peter A Gegenheimer[3,*], and Philippe Giegé[1,*]

[1]Institut de Biologie Moléculaire des Plantes du CNRS; Université de Strasbourg; Strasbourg, France; [2]Institut de Biologie Moléculaire et Cellulaire du CNRS; Architecture et Réactivité de l'ARN; Université de Strasbourg; Strasbourg, France; [3]Department of Molecular Biosciences; University of Kansas; Lawrence, KS USA

[†]These authors contributed equally to this work.

Keywords: RNase P, pentatricopeptide repeat, tRNA maturation, RNA binding protein, evolution

A fast growing number of studies identify pentatricopeptide repeat (PPR) proteins as major players in gene expression processes. Among them, a subset of PPR proteins called PRORP possesses RNase P activity in several eukaryotes, both in nuclei and organelles. RNase P is the endonucleolytic activity that removes 5' leader sequences from tRNA precursors and is thus essential for translation. Before the characterization of PRORP, RNase P enzymes were thought to occur universally as ribonucleoproteins, although some evidence implied that some eukaryotes or cellular compartments did not use RNA for RNase P activity. The characterization of PRORP reveals a two-domain enzyme, with an N-terminal domain containing multiple PPR motifs and assumed to achieve target specificity and a C-terminal domain holding catalytic activity. The nature of PRORP interactions with tRNAs suggests that ribonucleoprotein and protein-only RNase P enzymes share a similar substrate binding process.

Introduction

Pentatricopeptide repeat (PPR) proteins compose a family of RNA binding proteins specific to eukaryotes and mostly involved in gene expression processes in organelles. PPR proteins are particularly numerous in land plants with up to 450 representatives in *Arabidopsis thaliana*.[1] They are composed of tandem arrays of PPR motifs whose primary sequence is very degenerate,[2,3] although their tertiary structure seems to be conserved, with each repeat folding into two antiparallel α helices.[4-6] A succession of PPR motifs would thus make a superhelix that could act as a platform to bind RNA.[2] The combinatorial nature of PPR proteins allows substrate specificity because individual PPR motifs appear to ensure the selection for individual nucleotides.[6,7] Since their discovery over a decade ago, functional studies of PPR proteins have helped to answer many persistent questions regarding organellar gene expression processes.[1] For example, studies are beginning to unravel how sequence specificity is achieved for hundreds of C to U RNA editing sites in transcripts from higher plant organelles.[8] The characterization of PPR proteins has also helped to settle the long-standing debate over the existence of protein-only RNase P enzymes in eukaryotes.[9]

RNase P is a key enzyme of tRNA maturation. It was initially described as the endonuclease activity that removes the 5' leader sequences of tRNA precursors. It is therefore essential for producing functional tRNAs and, hence, indispensable for translation.[10,11] RNase P was first characterized on a molecular level in *Escherichia coli*, where it is composed of an RNA molecule together with a single protein.[12] The discovery that RNase P RNA held the actual catalytic activity of the enzyme[13] won Sidney Altman the Nobel prize in 1989 and helped to establish the "RNA world" theory proposing that one stage in prebiotic evolution consisted of RNA molecules that were able both to catalyze biochemical reactions and to store genetic information.[14] Subsequently, RNase P enzymes were characterized as similar ribonucleoprotein (RNP) enzymes in numerous other organisms and organelles including bacteria, archaea, yeast nuclei and mitochondria and animal cell nuclei.[15,16] Isolated RNA subunits from Bacteria, Archaea, and Eukarya demonstrate catalytic activity only under extreme ionic conditions, whereas the corresponding RNA•protein holoenzymes are maximally active under physiological conditions.[13] Apart from tRNAs, RNP RNase P enzymes are involved in the maturation of a wide array of substrates including rRNAs, protein-coding mRNAs, tmRNA, riboswitches, viral RNA, and snoRNA.[10,16,17] From a mechanistic point of view, RNP RNases P interact with tRNA mainly in the horizontal stacking domain consisting of the T stem-loop and acceptor stem; they utilize two catalytic metal ions and conserved RNA residues for RNA cleavage.[18,19] The structures of RNase P enzymes differ greatly, each containing an RNA molecule (whose structure is considerably reduced in size in some instances[20]) bound by a variable number of protein subunits ranging from one in bacteria to at least nine in eukaryotes.[10,21-23] Still, the central point remained that all RNase Ps contained an RNA moiety responsible for catalytic activity, so that the ribonucleoprotein nature of RNase P became a dogma. RNase P, together with the ribosome, was viewed as one of the ultimate universally conserved vestiges of the RNA world.[15]

Nevertheless, long before the discovery of the PPR protein family, some experimental evidence contradicted the prevailing dogma and suggested that some eukaryotes could use a different kind of enzyme, devoid of RNA, for RNase P activity. Here we review both the early evidence for the existence of protein-only RNase P and the studies describing the actual identification and

*Correspondence to: Peter A Gegenheimer; Email: pgegen@ku.edu; Philippe Giegé; Email: philippe.giege@ibmp-cnrs.unistra.fr
Submitted: 03/28/2013; Revised: 06/03/2013; Accepted: 06/04/2013
http://dx.doi.org/10.4161/rna.25273

characterization on a molecular level of the proteinaceous RNase P enzymes belonging to the PPR family.

Early Evidence for the Existence of Protein-only RNase P

Origins and expectations. The earliest reports of protein-only RNase P came from eukaryotic organelles—chloroplasts and mitochondria—that typically encode some or, in plant chloroplasts and vertebrate mitochondria, all of the tRNAs needed for translation of organellar-encoded proteins. In animals, mitochondrial tRNA genes are interspersed among protein-coding genes, such that production of functional mRNA species requires excision of mature tRNAs by precise 5'- and 3'-terminal endonucleolytic cleavages.[24] In chloroplasts, most tRNA genes are transcribed into end-extended precursors bearing 5'- and 3'-terminal extensions that must be removed to yield mature tRNA.[25]

The earliest expectations for the nature of RNase P from these organelles were based on their established bacterial origins:[26] mitochondria descended from the α-proteobacteria[27] and chloroplasts arose from within the cyanobacteria.[28] Members of both bacterial phyla possess "conventional" (*E. coli*-like) ribonucleoprotein forms of RNase P. In particular, bacterial-like RNase P RNA has been identified in all sequenced red algae chloroplasts and in many green algae in the Prasinophyte lineage.[10] For example, the cyanelle of the alga *Cyanophora paradoxa* encodes a homolog of cyanobacterial RNase P RNA.[29] This RNA alone exhibits weak catalytic activity at high salt concentrations, but can be restored to activity under physiological conditions by assembly with a cyanobacterial protein subunit.[30] (The equivalent protein subunit in *C. paradoxa* is presumably nuclear-encoded and imported into the cyanelle.) Likewise, the mitochondrion of the early-branching protozoan *Reclinomonas americana* encodes a proteobacterial-type RNase P RNA[31] that is dependent upon a proteobacterial protein subunit for activity.[32]

Initial evidence for an RNA component. Early support for a bacterial-like composition of mitochondrial RNase P was provided by genetic and biochemical determinations that in mitochondria of budding yeast (*Saccharomyces cerevisiae*), RNase P contained an essential, mitochondrial encoded, RNA distantly related to the RNA subunit of bacterial RNase P[33] and a nuclear-encoded protein unrelated to the bacterial protein subunit.[34]

The earliest characterizations of a putative vertebrate mitochondrial RNase P (from rat liver[35] or human cells[36]), did not directly test for the presence of an RNA component. Further efforts by one group, however, led to a claim that human mitochondrial RNase P activity could be attributed entirely to a small amount of nuclear RNase P imported into mitochondria.[37] Because these investigations employed a precursor to *E. coli* $tRNA^{Tyr}$suIII—a substrate for nuclear but not for vertebrate mitochondrial RNase P[38]—the enzyme described is now thought to be the abundant nuclear RNase P present in the starting cytosolic extracts.[35,39]

A critical assay: The substrate unmasked. Meanwhile, in the plant kingdom, transcription and processing of chloroplast tRNAs had been demonstrated in 1983 by Gruissem and Hallick.[25] Further investigation of chloroplast RNase P was initiated with the expectation that it, too, would resemble the bacterial enzyme. Preliminary evidence accumulated by 1986–87 suggested that crude preparations of RNase P from both spinach and tobacco chloroplasts were sensitive to treatment both with protease and with nuclease, consistent with activity residing in an RNA•protein complex (ref. 40 and Wang et al., poster presentation, 1986 Cold Spring Harbor RNA Processing Meeting). An apparent inhibition of chloroplast RNase P by S. aureus micrococcal nuclease (MN) is shown in **Figure 1A**, lanes 3–6.

At the time, three primary criteria were used to confirm the presence of an essential RNA component in an RNA processing activity: (1) sensitivity to pre-treatment with nucleases having little or no specificity for RNA sequence or structure, (2) buoyant density in Cs salts, and (3) presence of co-fractionating RNA species of appropriate size (150–400 nucleotide length). In the most common nuclease sensitivity protocol, an enzyme fraction is incubated with micrococcal nuclease (MN) in presence of its catalytic cofactor Ca^{2+}. The nuclease is then inactivated by addition of EGTA, which chelates most divalent cations much more strongly than it does Mg^{2+}, a required cofactor for all RNase Ps. Remaining RNase P activity is then assayed by addition of substrate directly to the treated enzyme fraction. All three assays are, however, susceptible to artifacts or misinterpretation. In particular, nuclease treatment is complicated by the fact that most suitable nucleases are difficult to inhibit cleanly, but residual activity will destroy the reaction substrate. EGTA-inactivated MN often displayed some inhibition of RNA processing.[40,41] Moreover, inhibition by active MN of non-RNA-containing enzymes had been observed (e.g., refs. 40 and 42). This was interpreted as resulting from degradation of bulk RNA, present in a crude extract, which was thought to stabilize the RNA processing complex under investigation.[42,43]

In order to conclusively show whether MN treatment was specifically inactivating chloroplast RNase P, Wang et al. asked whether MN-inhibited RNase P activity could be recovered by addition of non-specific RNA. The dramatic result, as shown in **Figure 1A**, lanes 7–10, was that addition of yeast RNA or of synthetic polynucleotides completely reversed the apparent inhibition by MN.[44] Further work[41] showed, not surprisingly, that Ca^{2+}-depleted MN retains substrate binding ability, reversibly binding pre-tRNA with an apparent K_d of 1.35 μM. Polyanions such as heparin or synthetic polynucleotides compete with pre-tRNA for binding MN. The final picture is that binding of catalytically inactive MN to RNA substrate sterically blocks access to the cleavage site. Addition of excess non-specific RNA sequesters the inactive nuclease and frees the pre-tRNA substrate for productive cleavage by the processing enzyme. This phenomenon is referred to as "substrate occlusion" or "substrate masking."[41]

Chloroplasts. With a reliable assay in hand, progress was rapid, and Wang et al. determined that chloroplast RNase P is completely insensitive to digestion with concentrations of micrococcal nuclease 20- to 50-fold greater than those required to inactivate *E. coli* RNase P.[44] Furthermore, as shown in **Figure 1B**, the chloroplast activity has a buoyant density in CsCl (1.28 g/cm^3) that is precisely centered within the density distribution

of bulk protein.[44] In this context, it is essential to note that because observed buoyant densities are a function of the density medium, and for values determined by refractometry, are also influenced by solvent composition, they cannot be directly compared between experiments. (In CsCl gradients, buoyant densities for pure protein, *E. coli* RNase P, and pure RNA are 1.28, 1.7, and ~2.0 g/cm^3; in Cs_2SO_4 these are 1.23, 1.55, and 1.65 g/cm^3.[43]) The most stringent test for presence or absence of an RNA is the extent to which enzyme activity co-fractionates with bulk protein or with a known protein-only enzyme. The coincidence of protein and enzyme densities for plant chloroplast and human mitochondrial RNase P indicates that neither enzyme could possess more than one copy of a 10- to 20-nt long RNA.[44,45]

Mechanistic differences between the chloroplast enzyme and the ribozyme-type RNase P affirmed that the chloroplast enzyme could not have an RNA subunit like that of bacterial or yeast nuclear RNase P[46] (discussed in detail in the section on structural mimicry). Further studies of the 1000–2000-fold purified chloroplast activity indicated that it does not co-purify with any RNAs that can be 3'-end labeled,[47] and that its hydrodynamic size, determined by gel filtration corresponds to a ~70 kDa globular protein.[47]

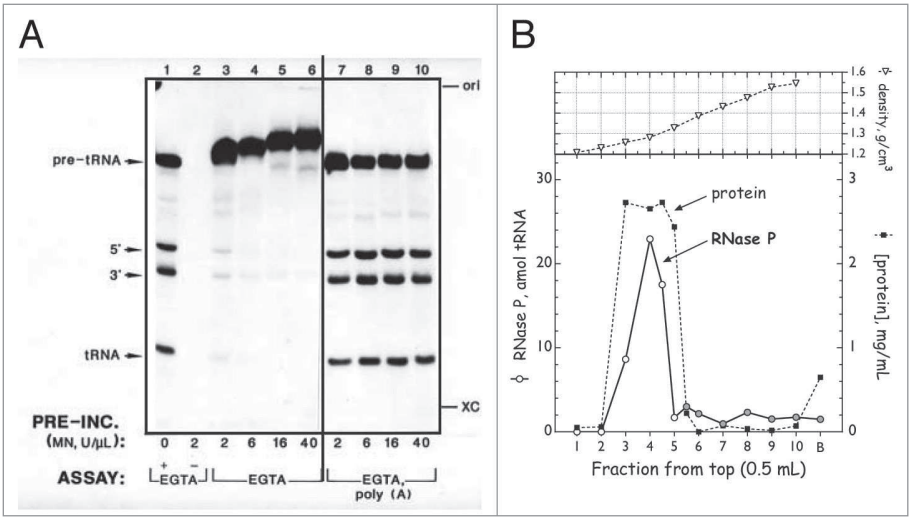


Figure 1. Plant RNase Ps do not contain an RNA component. (**A**) Resistance of spinach chloroplast RNase P to digestion with micrococcal nuclease (MN). Crude enzyme fraction was incubated with the indicated amounts of MN plus 5 mM $CaCl_2$ (30 min at 37 °C) after which excess EGTA was added, followed by substrate and reaction buffer. Lane 1, positive control for RNase P (pre-incubated without MN); lane 2, positive control for MN (as lane 1 but MN not inactivated prior to addition of substrate); lanes 3–6, pre-incubated with 2–40 U MN/μl and treated with EGTA prior to assay; lanes 7–10, as lanes 3–6 with addition of 1 μg poly(A)/μl prior to assay. Modified from reference 44 (Essentially identical results were obtained with wheat nuclear RNase P[48]). (**B**) Buoyant density of spinach chloroplast RNase P.[44] Fraction II chloroplast enzyme (~5 mg) was pretreated with MN (1 U/μl, 20 min; terminated with EGTA), brought up to 1.0 ml with gradient buffer, and layered over 4.0 ml of CsCl solution (1.40 g/ml). After centrifugation to equilibrium, fractions were collected from the top and density was determined by refractometry. CsCl was removed by dialysis and fractions were assayed for RNase P. Lower panel, distribution across the gradient of total protein (filled squares) and RNase P activity (open circles: amol mature tRNA formed; shaded circles: non-tRNA-sized material). Upper panel, observed buoyant density of each fraction.

Plant nuclei. Knowing that most soluble plastid proteins are encoded in the nucleus, translated in the cytoplasm, and imported into the organelle, Wang et al.[44] suggested that chloroplast RNase P or related polypeptides could have been recruited to process pre-tRNAs encoded in the nucleus and mitochondrion. To investigate this possibility, Oommen[48] used the techniques successful for chloroplasts to demonstrate that wheat embryo extracts contained an authentic RNase P activity with properties essentially identical to those of chloroplast RNase P. (The reaction requirements and substrate specificity of this activity [ref. 48 and unpublished observations] suggested that it was localized to the nucleus). This activity is resistant to amounts of micrococcal nuclease at least 5-fold greater than required to fully inactivate *E. coli* RNase P. In CsCl gradients, the distribution of wheat RNase P activity is absolutely coincident with the distribution of bulk protein (1.28–1.29 g/cm^3).[48] Active fractions across the final ion-exchange column contained no RNA molecules whose abundance was correlated with RNase P activity; trace RNAs larger than tRNA present in the active fractions could be removed without reducing RNase P activity. Finally, gel filtration chromatography in the absence of urea indicated a hydrodynamic size corresponding to a ~120 kDa globular protein or protein complex.

Somewhat later, another group presented essentially identical results: a buoyant density identical with bulk protein and complete resistance to MN treatment.[49] On the basis of its reaction requirements, this activity could be identified with nuclear rather than mitochondrial RNase P. At the time, these data were interpreted as consistent with wheat nuclear RNase P containing an RNA subunit associated with a large number of proteins that conferred a protein-like buoyant density and protected the RNA from nuclease attack.[49] Other researchers separated two RNase P activities, possibly nuclear and mitochondrial, from carrot cell suspension culture.[50] Presence or absence of RNA components was not established: buoyant densities were not determined and results of MN treatment were inconclusive because controls for substrate masking were not included and reaction products were not characterized. Of the two activities, one was inhibited only partially by a 10-fold excess of MN; the second was completely inhibited by either active or inactive MN at all concentrations tested, indicative of unresolved substrate masking.

Plant mitochondria. In 1990, two groups reported processing in vitro of plant mitochondrial pre-tRNAs with homologous mitochondrial extracts. Marchfelder et al. showed that RNase P-like activity in Oenothera mitochondrial lysates was completely

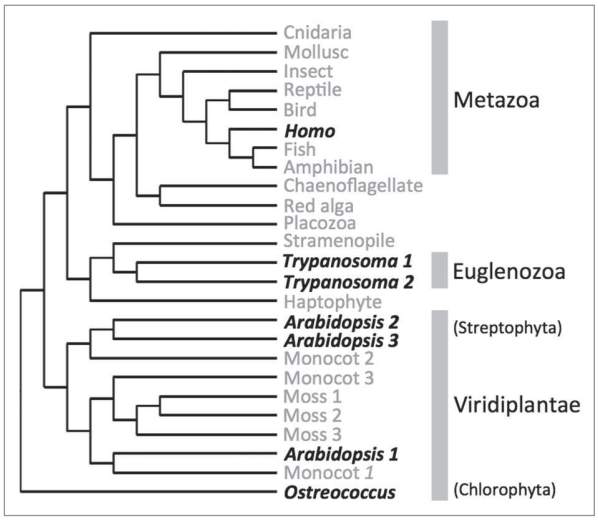


Figure 2. The occurrence of PRORP in eukaryote lineages is represented in an unrooted neighbor-joining phylogenic tree derived from Gobert et al.[61] Representative PRORP protein sequences described by Gobert et al. from evolutionarily distant eukaryotes were used for the phylogenetic analysis. Grey names show the incidence of putative PRORP sequences in the respective subgroups whereas black names indicate species where PRORP proteins were experimentally shown to hold RNase P activity. The demonstration that RNase P activity could be held by PRORP proteins in distantly related eukaryote groups such as Metazoa, Euglenozoa and Viridiplantae strongly suggest that PRORP evolved early in eukaryote history.

inhibited by either inactive or active MN when assayed in the absence of poly(A),[51] consistent with substrate masking. Hanic-Joyce and Gray, on the other hand, stated that the activity in wheat mitochondria was insensitive to MN digestion when assayed in the presence of poly(A).[52] In the absence of further physical characterization, these observations, though intriguing, were not seen as compelling.

Human mitochondria. The first purification of an authentic mitochondrial RNase P from vertebrates was reported by Rossmanith and colleagues[38] in 1995. Using a fully homologous system with a mitochondrial-specific substrate, they achieved a clean separation of human mitochondrial RNase P from the nuclear enzyme, which was by then known to be an RNA•protein complex.[53] Using an approach similar to that of Wang et al.,[44] Rossmanith then made a rigorous finding that the mitochondrial enzyme consisted entirely of protein.[45] First, activity was fully resistant to digestion with a 10-fold excess of MN. Second, in Cs_2SO_4 gradients, the buoyant density of RNase P activity (1.23 g/cm^3) was well within the distribution of bulk protein and was identical with the density of pre-tRNA 3'endonuclease, a known protein enzyme. The mitochondrial activity was cleanly separated from *E. coli* RNase P, which pelleted at the bottom of the gradient (density > 1.45 g/cm^3).[45] Third, the most highly-purified enzyme contained only RNAs of tRNA size and smaller, which could be degraded by MN treatment without affecting enzyme activity. Fourth, the mass of human mitochondrial RNase P, determined by rate zonal sedimentation, was about 170 kDa, substantially smaller than the smallest known RNA-containing RNase P. Additionally, since mitochondrial-specific RNase P could be isolated from mitochondrial mutants completely lacking mtDNA,[45] the mitochondrial enzyme was definitely encoded in the nucleus and imported into the organelle.

Kinetoplastid mitochondria. Mitochondria of the kinetoplastid parasite *Trypanosoma brucei* encode no tRNAs. Instead, all tRNAs are encoded in the nucleus and imported into the mitochondrion. Although it is uncertain whether any tRNAs are imported as 5'-extended precursors, it is known that kinetoplastid mitochondria do possess an active RNase P. In 2001, Salavati used the "masking-free" MN assay[44] to demonstrate that highly-purified *T. brucei* RNase P was unaffected by digestion with a 10-fold excess of MN.[54] Some RNAs larger than tRNA were present in active fractions but could be degraded without effect on RNase P activity. Notably, the hydrodynamic size estimated by gel filtration chromatography was about 70 kDa, the same size as chloroplast RNase P. In the absence of buoyant density or mechanistic data, however, these results were not considered definitive.

Hindsight. In retrospect, the ability to recognize the existence of protein-only RNase Ps was hindered by (1) justifiable expectations that organelles would have bacterial-type RNase P, most likely containing an organelle-encoded RNA subunit and an imported, nuclear-encoded polypeptide; (2) knowledge that yeast mitochondrial RNase P conformed to this model; (3) indications that RNase P in *C. paradoxa* cyanelles and *R. americana* mitochondria would follow the bacterial paradigm; and (4) evidence that yeast and human nuclear RNase Ps contained an RNA subunit related to the bacterial prototype. On the other hand, there was no obvious reason to doubt the validity of experimental work supporting a protein-only composition for RNase P in animal mitochondria, plant chloroplasts, or plant nuclei, nor was there convincing experimental evidence supporting other interpretations. Nevertheless, these conclusions remained controversial until isolated polypeptides, overexpressed from cloned cDNAs corresponding to defined genetic loci, were shown to possess RNase P activity.

Identification at the Molecular Level of Protein-Only RNase P

Characterization of the RNase P enzyme in human mitochondria. The concept of protein-only RNase P was definitely accepted only when the core components responsible for RNase P activity in human mitochondria were identified at the molecular level.[55] In that study, Rossmanith and coworkers confirmed that this RNase P activity did not require any RNA component. Using an elegant approach combining proteomic identification of human mitochondrial RNase P (mtRNase P) complexes, in vitro mtRNase P activity assay and reverse genetics, the authors' work led to the conclusion that only three individual polypeptide subunits were strictly required for the reconstitution of mtRNase P activity and that their mode of action was concerted. These three polypeptides composing the mtRNase P holoenzyme are nuclear-encoded and were named respectively MRPP1, 2, and 3 (for Mitochondrial RNase P Proteins). MRPP1 (or *TRMT10C*) encodes a putative tRNA:^{m1}G9-methyltransferase whereas

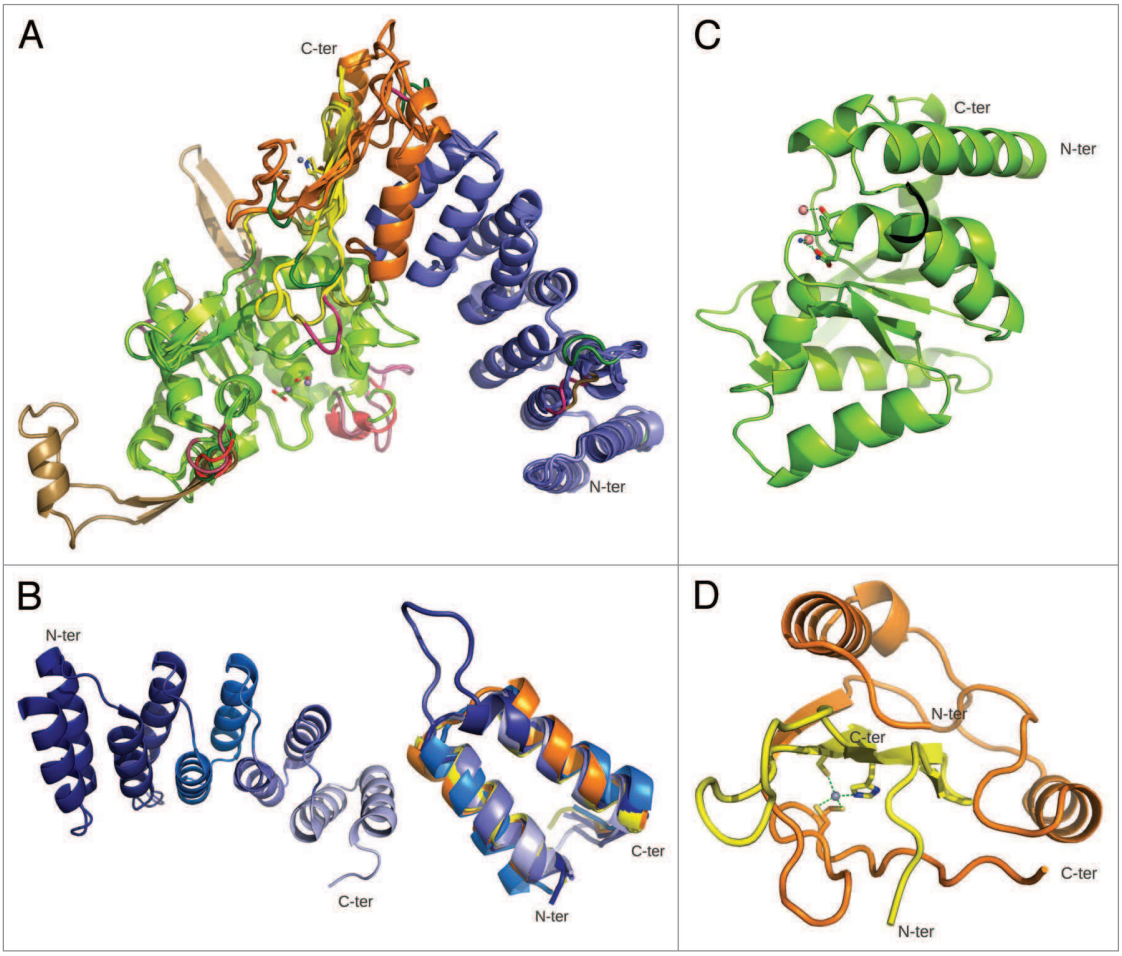


Figure 3. PRORP are two-domain PPR proteins. (**A**) 3D models were built using SwissModel[93] for all characterized members of the PRORP family based on At-PRORP1 crystal structure (PDB ID 4G24). This global view shows superimposed structure and models with PPR domains in blue, N-terminal and C-terminal connecting regions in orange and yellow, respectively, and catalytic NYN domains in green. Insertions/deletions to the reference structure of At-PRORP1 are colored as following: *A. thaliana* PRORP2, PRORP3, *O. tauri* PRORP, Trypanosoma PRORP2, and human mitochondrial MRPP3 indels are in violet, red, dark green, light brown, and pink, respectively. Little structural variations are observed. (**B**) At-PRORP1 PPR domain. (Left) This view of the whole domain highlights individual PPR motifs in light to dark blue from N to C terminus. (Right) Superposition of the five PPR motifs from *A. thaliana* PRORP1 (represented with the same color code as on the left) and the two PPR motifs (in orange and yellow) of human mitochondrial RNA polymerase (PDB ID 3SPA) illustrating the conservation of the PPR fold. (**C**) At-PRORP1 catalytic domain. Manganese ions shown as pink spheres and two water molecules bridging one Mn^{2+} ion to conserved Asp474 in the catalytic site. (**D**) At-PRORP1 connecting region. This region is composed of a N-terminal half (orange) following the PPR domain and a C-terminal half (yellow) following the catalytic domain. It binds a zinc ion (gray) coordinated by C344, C345 (orange) and H548, C565 (yellow).

MRPP2 (or *SDR5C1*) encodes a 3-hydroxyacyl-CoA dehydrogenase and *MRPP3* encodes a protein containing a metallonuclease domain as well as a PPR domain.[9,55]

MRPP1 catalyzes the methylation of specific bases (G9 or A9) in mitochondrial tRNAs and interacts with tRNAs in vitro,[56] although its methyltransferase activity is not required for tRNA cleavage by the mtRNase P holoenzyme.[56] Little is known about the involvement of MRPP2 in mtRNase P activity. Binding to MRPP2 is critical for MRPP1 to perform mitochondrial tRNA methylation, although MRPP2's dehydrogenase activity seems to be dispensable.[56] Reciprocally, although MRPP1 and MRPP2 are essential components of the mtRNase P holoenzyme, neither the methyltransferase nor the dehydrogenase activity, respectively, is required for tRNA processing.[56]

MRPP3 was the only identified subunit of mtRNase P harboring a predicted nuclease domain. Hence, it was hypothesized from the start that the involvement of MRPP3 in mtRNase P activity would be to perform the actual phosphodiester bond hydrolysis.[55] MRPP3 also features PPR motifs. These elements are helical-repeat motifs considered to bind with specificity to single-stranded RNA stretches; they are found in eukaryotic proteins, predominantly those involved in organellar RNA metabolism.[1,2,57] Even though the precise role of MRPP3's PPR motifs in mtRNase P is still unexplored, a tempting proposal is that these repeats contribute to tRNA binding and/or confer base-specific recognition of tRNAs.

Apart from the protein-only RNase P, it was also proposed that RNase P RNA could be imported into human mitochondria,

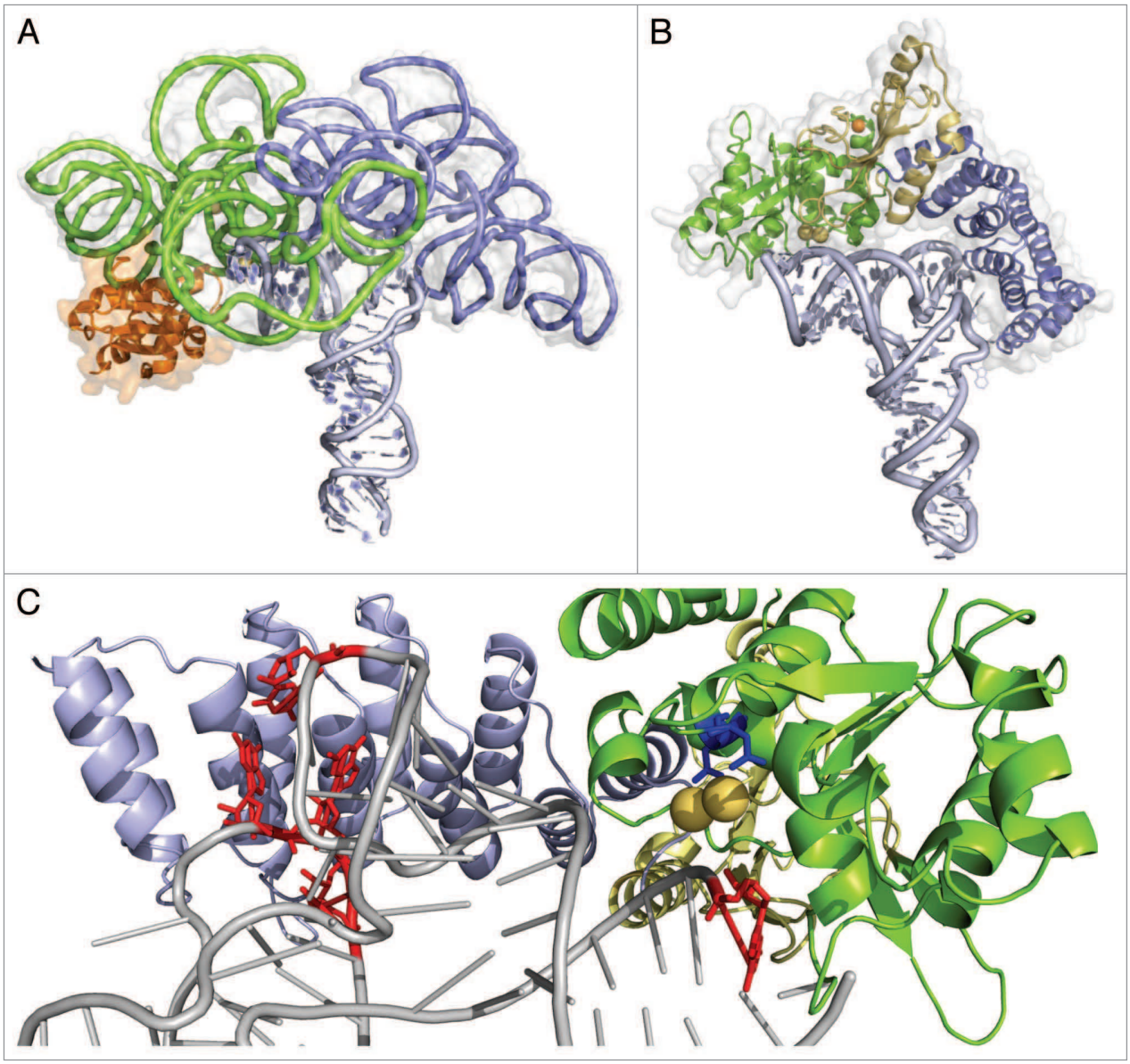


Figure 4. The current model of the PRORP/tRNA complex suggests a common mode of RNA binding in RNP and PRORP RNases P. (**A**) Structure of *Thermotoga maritima* ribozyme (PDBid 3Q1R18) with the catalytic domain in green, the specificity domain in blue, the RNase P protein subunit in orange, the tRNA product in light blue and the molecular surface of the RNP in gray. (**B**) The two-domain architecture of At-PRORP1 structure offers a concave surface that can be docked on the tRNA acceptor arm. The protein shown in the same orientation and same color code as the RNP with the catalytic domain in green, with metal ions bound (yellow spheres) close to the RNA cleavage site and the RNA-binding PPR domain in blue interacting with the region of the D-TψC loops. The central region (yellow) stabilized by a zinc ion (orange sphere) connects the two main PRORP domains. (**C**) A close-up of the PRORP1-tRNA complex model shows conserved catalytic aspartates D474 and D475 (blue) adjacent to tRNA cleavage site at position G+1 (red) as well as U16, G18, G19, and C56 (the nucleotides protected in footprint experiments[77] in red) in contact with PPR motifs. Current functional data indicate that PRORP proteins have evolved an RNA recognition process very similar to that of RNP RNase P.

thus leading to the potential cohabitation of both RNP and protein-only RNase P in this organelle.[58] The occurrence of RNase P RNA in human mitochondria remains controversial and has been discussed in detail by Rossmanith in 2012.[59]

The catalytic subunit of protein-only RNase P is the PPR protein. While the three polypeptides that compose the human mitochondrial RNase P enzyme have some RNA-binding potential, only MRPP3 possesses the features of a metallonuclease that could account for the catalytic activity of RNase P. MRPP3 orthologs could be identified in many eukaryotic organisms and define a new protein family that was named PRORP (for PROteinaceous RNase P). Hence, MRPP3 is now also called human PRORP. These proteins are characterized by the presence of a number of PPR and/or PPR-like motifs, a CXXC Zn finger-like motif and a metallonuclease domain belonging to the NYN family.[60] The function of putative PRORPs identified by sequence similarities

has been explored in depth in Arabidopsis[61,62] and in the protist *T. brucei*.[63] Data are also available for *Ostreococcus tauri*, a primitive unicellular green alga.[64] Arabidopsis expresses three PRORP proteins: At-PRORP1 is a 62 kDa protein with a pI of 9 and is localized to both plastids and mitochondria, whereas At-PRORP2 and At-PRORP3 are 59 kDa proteins, with pI of 6 and are localized in the nucleus.[61,62] RNase P catalytic activity was first assigned unequivocally to the single protein At-PRORP1 in Arabidopsis organelles[61] and later to each of the nuclear proteins on its own.[62] In vitro RNase P activity tests using homologous pre-tRNA substrates were performed with purified recombinant forms of the three Arabidopsis RNase P protein candidates, each carefully verified for the absence of contamination by *E. coli* RNase P.[61,62] Precise mapping of the cleavage site was achieved by high-resolution urea-PAGE or circular RT-PCR, and characterization of the 5' nucleotide of the mature tRNA products showed that each PRORP is a tRNA-specific endonuclease removing 5' extensions from pre-tRNAs and leaving a phosphate group at the 5' end of mature tRNAs. Abolition of the RNase P activity of recombinant PRORPs mutated in two conserved aspartates (predicted to be part of the catalytic site) confirmed that each of the three Arabidopsis PRORPs possessed RNase P activity as a single polypeptide.[61,62]

Two PRORP genes were identified in the fully sequenced trypanosomatid genomes. In *Trypanosoma brucei*, PRORP1 is localized to the nucleus and PRORP2 to the mitochondrion.[63] Using in vitro cleavage assays with purified recombinant proteins, each *T. brucei* PRORP protein appeared to perform the canonical 5' tRNA maturation on its own, similar to Arabidopsis PRORPs.[63] Although studied to a lesser extent, a recombinant PRORP from the green algae *O. tauri* is capable of pre-tRNA 5' processing in vitro.[64] The RNase P activity of these eukaryotic PRORP proteins from distant organisms is most likely shared by other members of this family. The association of a nuclease domain with a PPR domain to create RNase P enzymes represents yet another example of the potential and diversity of functions (i.e., RNA editing, splicing, or translation[1]) acquired by the family of PPR proteins.

Beyond the capacity to perform RNase P activity in vitro, an important testimonial to the generality of PRORP tRNA processing capability came with the observation that Arabidopsis organellar PRORP1 and Trypanosoma nuclear PRORP1 could replace, in vivo, the *E. coli* and yeast nuclear RNase P respectively.[61,63] Wild-type At-PRORP1, but not a protein mutated in the two conserved catalytic aspartates, rescues the lethal knockdown of RNase P RNA in *E. coli*. Similarly, *T. brucei* nuclear PRORP1 can rescue a deletion of the RNA component of yeast nuclear RNase P. These heterologous complementations led to the remarkable result that a single polypeptide can substitute in vivo for a complex ribonucleoprotein structure. Still, PRORP might not be the exact functional equivalent of RNP RNase P as fitness differences were observed between yeast strains non complemented and complemented by PRORP.[63] Similarly, kinetic studies reveal that specificity constants of PRORP are not equivalent, i.e., they are lower than that of RNP RNase P.[62,65]

Further experiments explored the in vivo roles of the three Arabidopsis PRORPs. The lethality of a single-gene knockout of At-PRORP1 and of the double knockout of At-PRORP2 and At-PRORP3 indicate that both the organellar and the nuclear PRORP enzymes fulfill essential functions in vivo, as expected for the authentic RNase P in cellular compartments encoding tRNA.[61,62] The role of At-PRORP in tRNA 5' maturation in both organelles and the nucleus in planta was further explored by downregulation using virus-induced gene silencing.[62] A decrease in PRORP1 specifically affects internal structures of chloroplast and mitochondria and reduces the level of mature organellar tRNAs, while nuclear-encoded tRNA levels are unchanged. Conversely, downregulation of PRORP2 in a prorp3 knockout background has no effect on organellar tRNAs, while the level of nuclear-encoded processed tRNA is reduced compared with control plants. Since downregulation of each PRORP protein causes a reduction of RNase P activity in the cellular compartment where that protein is found, it can be concluded that each PRORP protein is required for processing the tRNA pool in its respective compartment. On the other hand, downregulation of POP1 and POP4, two essential protein components of RNase MRP (a ribonucleoprotein related to the nuclear RNP RNase Ps and involved in cytosolic rRNA maturation) affected rRNA maturation but did not reduce nuclear tRNA levels.[62] Altogether, these results are consistent with PRORP proteins being the sole source of RNase P activity in both organelles and the nucleus of plants.

A report by Krehan, et al. has shown that RNase P activity as well as RNase MRP RNA are present in a wheat embryo extract immune-precipitated with POP1 antibodies.[66] This result has been interpreted as a clue for the presence of an RNP RNase P enzyme in plant nuclei.[67] Since the downregulation of POP1 in planta resulted in decreased RNase MRP activity and did not affect RNase P activity,[62] we believe that the results instead reflect the presence of both PRORP and RNase MRP in the immunoprecipitated fraction, i.e., that the two enzymes might be present in a single complex in planta as also proposed by Krehan, et al.[66]

In Trypanosoma, PRORP activity was analyzed after immunodepletion, with anti-PRORP antibodies, of RNase P activity in a whole-cell extract. Depletion of both nuclear PRORP1 and mitochondrial PRORP2 abolishes all activity, suggesting that *T. brucei* contains no other RNase P.[63] More studies are required, however, to understand the function of *T. brucei* PRORPs in vivo. Since a complete set of tRNAs is imported from the cytosol into mitochondria in Trypanosoma, it will be particularly interesting to identify the substrates of the mitochondrial PRORP2 in vivo.

Collectively, experimental data obtained in distantly-related eukaryotes has clearly established that RNase P activity can reside in a single polypeptide. Moreover, in plants and Trypanosomes, PRORP proteins provide RNase P activity in vivo in both organelles and in the nucleus.

The substrate spectrum of PRORP, like that of RNP RNase Ps, goes beyond tRNAs. RNase P was first defined as the activity performing the 5' maturation of tRNA precursors. Still, extensive analyses of ribonucleoprotein RNase P functions have revealed that RNase P can be involved in the maturation of a much

wider variety of substrates in both prokaryotes and eukaryotes.[10] After finding that PRORP proteins could perform the 5' maturation of tRNA precursors in Arabidopsis, Trypanosoma, and Ostreococcus,[61,63,64] it was logical to investigate whether PRORP proteins are entirely tRNA-specific or whether they, like RNP RNase Ps, are involved in the maturation of other substrates.

The assumption that PRORP enzymes might be involved in the maturation of other RNAs is supported by the fact that numerous tRNA-derived sequences or structures are present in plant genomes. For instance, tRNA-like sequences called "t-elements" are present in transcripts of plant mitochondrial DNA, where they separate individual mRNAs.[68] Similarly, in the nucleus, SINE RNAs are derived from tRNAs, although their canonical cloverleaf structure has apparently been lost.[69] Another argument comes from the observation that Arabidopsis PRORP1 can replace *E. coli* RNP RNase P in vivo.[61] Bacterial RNP RNase P is responsible for the maturation of many non-tRNA substrates, including the precursor to the 4.5S RNA.[70] Two substrates that contain tRNA-like recognition elements are the precursor to C4 antisense RNA of bacteriophage P1 and P7, which possesses a tRNA-like structure with short D- and T-loops;[71] and the precursor to tmRNA, part of whose structure resembles the horizontal stacking domain (acceptor stem plus T-stem and loop) of $tRNA^{Ala}$,[72] a known minimal substrate for *E. coli* RNase P.[73] *E. coli* RNase P is also involved in processing polycistronic mRNAs such as the histidine operon transcript,[74,75] and in cleavage of some riboswitches, including those for the coenzyme B12.[17] It can thus be speculated that Arabidopsis PRORP1 could catalyze the maturation of all these *E. coli* non-tRNA substrates. Alternatively, it is possible that some of these non-tRNA maturation steps are not essential or that they can be rescued by other enzymatic systems in the absence of ribonucleoprotein RNase P.

Preliminary results, both in vitro and in vivo, have confirmed that Arabidopsis PRORPs are indeed involved in the maturation of other RNA substrates. In particular, PRORP1 is able to perform in vitro the endonucleolytic cleavage of tRNA-like t-elements present in the mitochondrial transcripts of Arabidopsis nad6 and *Brassica napus* orf138,[61] and PRORP1 activity is required in vivo to accumulate nad6 mRNA.[62]

Similarly, Arabidopsis PRORP2 and 3 are indirectly involved in the maturation of snoRNA.[62] In Arabidopsis, a dicistronic precursor to $tRNA^{Gly}$ and the snoRNA snoR43.1 is processed by both RNase P and the pre-tRNA 3'-processing endonuclease RNase Z, with RNase P cleavage of the pre-$tRNA^{Gly}$ portion being a prerequisite for the cleavage by RNase Z that separates mature $tRNA^{Gly}$ from mature snoR43.1.[76] In PRORP downregulation mutants, snoR43 failed to accumulate to normal levels whereas tRNA-snoRNA precursor levels increased, showing that nuclear PRORP activity is required for the accumulation of this snoRNA.[62]

An initial investigation of the PRORP/tRNA complex has revealed that minimal tRNA structural features are required for recognition by PRORP alone. For example, and like the bacterial RNP RNase P, the tRNA acceptor stem is essential whereas the anticodon domain is not. Unlike the bacterial enzyme, PRORP cleavage is impaired by the absence of the D domain from tRNAs or t-elements.[61,77] Thus, maturation of mitochondrial mRNAs by cleavage of some t-elements (such as the one from ccmC mRNA that lacks both D and anticodon domains[68]) might require additional proteins acting as PRORP partners to recognize these structures. Similarly, in humans, the requirement for MRPP1 and 2 might reflect an inability of Hs-PRORP alone to bind the non-canonical tRNA structures characteristic of vertebrate mitochondria.[55] This would also explain why the plant PRORP1 can function in both chloroplasts and mitochondria, since tRNAs from plant chloroplasts and mitochondria closely resemble bacterial tRNAs.

The diversity of substrates identified so far for PRORP remains limited. Other potential RNA substrates will have to be investigated at the transcriptome-wide level, for example, through comparative transcriptome analyses of PRORP downregulation mutants or by global sequencing of RNA partners immune-precipitated in complex with PRORP proteins.

Emergence and distribution of PRORP enzymes in eukaryotes. RNase P is a ubiquitous enzyme, found in all organisms with the exceptions of symbiotic Archaea, such as *Nanoarchea equitans*, several species of Pyrobaculum and *Aquifex aeolicus* in which transcription of tRNAs starts at position +1.[78] The RNP form of RNase P is widespread as it is present in Bacteria, in Archaea and in Eukarya with characterized activities in both the nucleus and mitochondria (as, for example, in yeast).[10] On the contrary, identified PRORP RNase Ps are limited to eukaryotes (**Fig. 2**), having been identified in human mitochondria,[55] *Arabidopsis thaliana* mitochondria, chloroplasts, and nuclei[61,62] and in *Trypanosoma brucei* mitochondria and nuclei.[63] In the green alga *Ostreococcus tauri*, a PRORP protein was found to have RNase P activity but its localization was not determined.[64] However, bacterial-type RNase P ribozymes can be found encoded in both mitochondrial and plastidial genomes along with an RNP RNase P protein in the nucleus.[29,64] Nonetheless, although all characterized PRORP proteins are eukaryotic, they are not restricted to endosymbiotic organelles as was previously assumed.[11]

Database analyses confirm that PRORP proteins constitute a eukaryote-specific family of enzymes. Putative PRORP sequences can be found in nearly all major eukaryotic groups (i.e., in Metazoa, Streptophyta, Chlorophyta, Kinetoplastida, Stramenopiles, and Oomycetes) with the notable exceptions of fungi and amoebozoa.[61] The appearance of PRORP can essentially be defined by the event that led to the fusion of a PPR domain with an NYN nuclease domain. The precise timing of this event and the evolutionary history of PRORP remain to be established. Still, its occurrence as experimentally shown for Metazoa, Euglenozoa and for both Streptophyta and Chlorophyta in Viridiplantae (**Fig. 2**), already suggests that PRORP appeared very early in the evolution of eukaryotes.[61]

The emergence of PRORP has been proposed to be related to the acquisition of organelles.[5] Similarly, Howard et al. suggested that the evolutionary drive for RNP replacement by PRORP might have resided in different substrate specificities between nuclear and organelle RNase P enzymes, in the difficulty of importing a large RNA such as that for RNase P into mitochondria, or in the "vulnerability" of organelle RNA toward

RNP RNase P enzymes.[67] All these propositions assume that PRORP initially arose as an organelle-targeted enzyme, which is not established and not necessarily true. Indeed, PRORP clearly emerged as a nuclear gene by fusion of genes encoding a PPR RNA-binding protein and an NYN metallonuclease domain (discussed in the following section). Because the nuclear RNase P activity of PRORP is found in distantly related eukaryotes, PRORP nuclear activity is most likely ancient. It is thus possible that PRORP might have first functioned as a nuclear enzyme. If so, the evolutionary impetus to replace a RNP complex containing one RNA and up to ten proteins by a protein-only enzyme might have resided in the fact that the simpler enzyme assembles faster, is easier to regulate and requires fewer cellular resources for its biogenesis.

PRORP enzymes are two-domain proteins. Initial structural predictions of PRORP based on sequence analyses indicated the presence of PPR modules in the N terminus and of a NYN-like catalytic domain in the C terminus. This organization into two α-helix-rich domains was supported by biophysical characterization (circular dichroism and small angle X-ray scattering) of At-PRORPs in solution and is consistent with the X-ray crystal structure of At-PRORP1.[77,79] Taking together 1-, 2-, and 3-D data available for this enzyme family, comparative models of representative PRORP members are presented in **Figure 3A**. These models pinpoint the general conservation of the PRORP fold from unicellular algae to humans. Small variations are observed, mainly in peripheral loops. Long insertions are present in plasmodial enzymes, as is often observed in proteins from this parasite family.[80]

The N-terminal PPR domain forms a superhelical structure very similar to those described in TPR (TetratricoPeptide Repeat) domains, an evolutionary-related domain involved in protein•protein interactions.[2,81] As illustrated in **Figure 3B**, it contains five PPR and PPR-like motifs: two canonical ones and three displaying remote sequence similarities. Despite their divergent sequences, these PPR modules are structurally similar and superimposable on those found in the only other PPR protein of known three-dimensional structure, i.e., human mitochondrial RNA polymerase.[4] This confirms, as was originally proposed,[2] that the defining feature of PPR family members is a conserved structural fold of PPR motifs rather than of conserved sequence elements.

The catalytic domain of PRORP adopts an α/β/α sandwich fold (**Fig. 3C**) belonging to the PIN-like nuclease family.[82,83] A similar architecture is found in the nuclease domain of T4 RNase H,[84] and of human SMG6 and SMG5, two essential factors in nonsense-mediated mRNA decay,[82] as well as of a recently characterized MCPIP1 RNase (MCP-1 induced protein 1) that participates in the regulation of immune response by degrading the mRNA of inflammatory cytokines.[83] Among the four aspartate residues involved in the binding of metal ions,[79] two are strictly conserved in PRORPs (D474 and D475 in At-PRORP1) and in other nucleases of the PIN/NYN family and are essential for pre-tRNA cleavage.[77]

These two functional domains are connected by a split zinc-binding module derived from the central and the C-terminal regions of PRORP (**Fig. 3D**) and which forms the tip of the overall "V shape" of PRORP, the PPR and catalytic domains being the two arms of the V. The concave surface of the PPR superhelix in one arm thus faces the catalytic groove in the other arm, thereby exposing conserved aspartate residues and metal ions, making the overall architecture look like tweezers.

Is PRORP a structural mimic of ribonucleoprotein RNase P? The bacterial RNP RNase P docks onto the acceptor stem of its pre-tRNA substrate, with an interaction extending from the tRNA corner (T and D loops), which is recognized by the specificity domain (S-domain) to the cleavage point between nucleotides –1 and +1, which is apposed to the catalytic domain (C-domain).[18,85,86] In *E. coli* tRNAs, the 3' terminal CCA interacts specifically with a complementary sequence in a loop of the RNase P RNA,[87] whereas the pre-tRNA leader interacts with the protein subunit of the holoenzyme.[88]

The bipartite organization of PRORPs (**Fig. 3**) is reminiscent of that of RNP RNase P, with the PPR domain playing the role of the S-domain to ensure recognition of the pre-tRNA and its orientation in the catalytic domain. In support of this role, removal of the four N-terminal PPR motifs of At-PRORP1 leads to a 34-fold drop of affinity for the substrate and a > 2000-fold loss of enzymatic activity.[79] Similarly, the deletion of the S-domain in the RNP RNase P resulted in 30- to 13 000-fold loss in catalytic performance, depending upon the substrate used. However, the S-domain deletion, surprisingly, led to more accurate cleavage site selection.[89]

On the substrate side, deletions altering the pre-tRNA structure show that for PRORP, just as for the RNP RNase P, the anticodon stem-loop is dispensable, whereas the D and T loops are required. Footprint experiments confirmed that the corner of the tRNA L-fold interacts with At-PRORP1 to give strong protection of residues U16, G18-19, and C56.[77]

The PRORP/pre-tRNA complex was modeled based on the At-PRORP1 crystal structure using as geometrical restraints the binding of the T/D loops by the PPR domain and the positioning of the cleavage point in the vicinity of conserved aspartate groups constituting the metal-binding site. **Figure 4** illustrates the potential similarity between PRORP and RNP RNase P in the way they bind their pre-tRNA substrates.[77] Another model of PRORP/pre-tRNA complex has been proposed, it shows PRORP interacting on the side rather as on the top of tRNAs.[67] However, the latter does not take in account footprinting and tRNA deletion results that suggested contacts between PRORP and tRNA residues U16, G18, G19 as well as C56, while the anticodon stem is dispensable for recognition.[77]

Despite the overall similarity of their substrate-binding modes, however, the two types of RNase P—employing a protein or an RNA catalytic component—are mechanistically distinct.[79] Both cleave a phosphodiester bond by nucleophilic attack of hydroxide ion apical to O3' of the upstream ribose, generating products with 3'-hydroxyl and 5'-phosphoryl termini. The presence of metal-binding sites in the structure of At-PRORP1 suggests that the proteinaceous enzymes use a two-metal-ion mechanism[90] to deprotonate water and to stabilize the transition state. However, the tolerance of PRORPs

to an *R*p-phosphorothioate modification of the scissile bond in the presence of Mg^{2+} as cofactor is a striking difference from the RNP enzyme,[65,91] indicating that the metal in PRORP does not directly coordinate the *pro-R*p-oxygen of the target phosphodiester. Rather, it appears that, whereas the RNase P RNA subunit employs one hydrated divalent cation to provide the attacking hydroxide and a second metal hydrate to protonate the leaving group,[18,79] the proteinaceous RNase P utilizes a more conventional mechanism akin to that of known protein metallonucleases, in which the metal ions serve primarily to stabilize the charge and structure of the trigonal bipyramidal transition state, and general acid-base chemistry is accomplished by the carboxylate groups of aspartate (and possibly the imidazole nitrogen of histidine). The binding affinities of PRORPs for their pre-tRNA substrate are in the micromolar range.[65,79] These values are one or two orders of magnitude lower than for RNP RNases P and may indicate more transient interaction with substrates. Nevertheless, these proteinaceous enzymes are efficient enough to complement *E. coli* RNP RNase P.[61] So the precise functional advantages of the PRORP and RNP RNase P mechanistic dissimilarities remain to be identified.

Concluding Remarks

Within the PPR family, the characterization of PRORP proteins has finally settled the long-lasting debate over the existence of an alternative system devoid of RNA for RNase P activity in eukaryotes. The discovery of protein-only RNase P indicates that the distribution and evolutionary history of RNase P are more complex than previously thought. The functional and mechanistic comparison of PRORP with RNP RNase P will have important implications for our understanding of the evolution of living systems. Indeed, it will illustrate how convergent evolution has found two independent routes to catalyze the 5' maturation of tRNAs: either with an RNA-based enzyme or a protein-only enzyme. This mechanistic comparison leads to important questions. For instance, the mechanism by which PPR motifs confer PRORP substrate specificity remains to be elucidated. Future work, in particular determination of the crystal structure of PRORP in complex with tRNA, will establish whether PPR motifs indeed bind conserved residues in the single-stranded D and T loops of tRNAs as was previously suggested,[77] and thus whether the PRORP mode of RNA recognition is in conformity with the overall mode of RNA recognition recently proposed for PPR proteins.[7,92]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by the French "Centre National de la Recherche Scientifique" and by the University of Strasbourg. AG and OF were supported by an ANR Blanc research grant "PRO-RNase P, ANR 11 BSV8 008 01" to PG and CS and by the LabEx consortium "MitoCross." FP and BG were supported by PhD grants from the University of Strasbourg.

References

1. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 2008; 13:663-70; PMID:19004664; http://dx.doi.org/10.1016/j.tplants.2008.10.001
2. Small ID, Peeters N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 2000; 25:46-7; PMID:10664580; http://dx.doi.org/10.1016/S0968-0004(99)01520-0
3. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16:2089-103; PMID:15269332; http://dx.doi.org/10.1105/tpc.104.022236
4. Ringel R, Sologub M, Morozov YI, Litonin D, Cramer P, Temiakov D. Structure of human mitochondrial RNA polymerase. Nature 2011; 478:269-73; PMID:21947009; http://dx.doi.org/10.1038/nature10435
5. Goldfarb KC, Borah S, Cech TR. RNase P branches out from RNP to protein: organelle-triggered diversification? Genes Dev 2012; 26:1005-9; PMID:22588715; http://dx.doi.org/10.1101/gad.193581.112
6. Filipovska A, Rackham O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Mol Biosyst 2012; 8:699-708; PMID:22234420; http://dx.doi.org/10.1039/c2mb05392f
7. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 2012; 8:e1002910; PMID:22916040; http://dx.doi.org/10.1371/journal.pgen.1002910
8. Fujii S, Small I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 2011; 191:37-47; PMID:21557747; http://dx.doi.org/10.1111/j.1469-8137.2011.03746.x
9. Rossmanith W, Holzmann J. Processing mitochondrial (t)RNAs: new enzyme, old job. Cell Cycle 2009; 8:1650-3; PMID:19411849; http://dx.doi.org/10.4161/cc.8.11.8502
10. Lai LB, Vioque A, Kirsebom LA, Gopalan V. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. FEBS Lett 2010; 584:287-96; PMID:19931535; http://dx.doi.org/10.1016/j.febslet.2009.11.048
11. Esakova O, Krasilnikov AS. Of proteins and RNA: the RNase P/MRP family. RNA 2010; 16:1725-47; PMID:20627997; http://dx.doi.org/10.1261/rna.2214510
12. Stark BC, Kole R, Bowman EJ, Altman S. Ribonuclease P: an enzyme with an essential RNA component. Proc Natl Acad Sci USA 1978; 75:3717-21; PMID:358197; http://dx.doi.org/10.1073/pnas.75.8.3717
13. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 1983; 35:849-57; PMID:6197186; http://dx.doi.org/10.1016/0092-8674(83)90117-4
14. Cech TR. The efficiency and versatility of catalytic RNA: implications for an RNA world. Gene 1993; 135:33-6; PMID:8276275; http://dx.doi.org/10.1016/0378-1119(93)90046-6
15. Altman S. A view of RNase P. Mol Biosyst 2007; 3:604-7; PMID:17700860; http://dx.doi.org/10.1039/b707850c
16. Evans D, Marquez SM, Pace NR. RNase P: interface of the RNA and protein worlds. Trends Biochem Sci 2006; 31:333-41; PMID:16679018; http://dx.doi.org/10.1016/j.tibs.2006.04.007
17. Altman S, Wesolowski D, Guerrier-Takada C, Li Y. RNase P cleaves transient structures in some riboswitches. Proc Natl Acad Sci USA 2005; 102:11284-9; PMID:16061811; http://dx.doi.org/10.1073/pnas.0505271102
18. 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 2010; 468:784-9; PMID:21076397; http://dx.doi.org/10.1038/nature09516
19. Reiter NJ, Osterman AK, Mondragón A. The bacterial ribonuclease P holoenzyme requires specific, conserved residues for efficient catalysis and substrate positioning. Nucleic Acids Res 2012; 40:10384-93; PMID:22904083; http://dx.doi.org/10.1093/nar/gks744
20. Seif ER, Forget L, Martin NC, Lang BF. Mitochondrial RNase P RNAs in ascomycete fungi: lineage-specific variations in RNA secondary structure. RNA 2003; 9:1073-83; PMID:12923256; http://dx.doi.org/10.1261/rna.5880403
21. Hartmann E, Hartmann RK. The enigma of ribonuclease P evolution. Trends Genet 2003; 19:561-9; PMID:14550630; http://dx.doi.org/10.1016/j.tig.2003.08.007
22. Marvin MC, Engelke DR. RNase P: increased versatility through protein complexity? RNA Biol 2009; 6:40-2; PMID:19106627; http://dx.doi.org/10.4161/rna.6.1.7566
23. Jarrous N, Gopalan V. Archaeal/eukaryal RNase P: subunits, functions and RNA diversification. Nucleic Acids Res 2010; 38:7885-94; PMID:20716516; http://dx.doi.org/10.1093/nar/gkq701
24. Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. Nature 1981; 290:470-4; PMID:7219536; http://dx.doi.org/10.1038/290470a0
25. Gruissem W, Greenberg BM, Zurawski G, Prescott DM, Hallick RB. Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. Cell 1983; 35:815-28; PMID:6652686; http://dx.doi.org/10.1016/0092-8674(83)90114-9

26. Gray MW, Cedergren R. The new age of RNA. FASEB J 1993; 7:4-6; PMID:7678565
27. Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. Mitochondrial origins. Proc Natl Acad Sci USA 1985; 82:4443-7; PMID:3892535; http://dx.doi.org/10.1073/pnas.82.13.4443
28. Giovannoni SJ, Turner S, Olsen GJ, Barns S, Lane DJ, Pace NR. Evolutionary relationships among cyanobacteria and green chloroplasts. J Bacteriol 1988; 170:3584-92; PMID:3136142
29. Baum M, Cordier A, Schön A. RNase P from a photosynthetic organelle contains an RNA homologous to the cyanobacterial counterpart. J Mol Biol 1996; 257:43-52; PMID:8632458; http://dx.doi.org/10.1006/jmbi.1996.0145
30. Pascual A, Vioque A. Functional reconstitution of RNase P activity from a plastid RNA subunit and a cyanobacterial protein subunit. FEBS Lett 1999; 442:7-10; PMID:9923593; http://dx.doi.org/10.1016/S0014-5793(98)01621-4
31. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, et al. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 1997; 387:493-7; PMID:9168110; http://dx.doi.org/10.1038/387493a0
32. Seif E, Cadieux A, Lang BF. Hybrid E. coli--Mitochondrial ribonuclease P RNAs are catalytically active. RNA 2006; 12:1661-70; PMID:16894220; http://dx.doi.org/10.1261/rna.52106
33. Hollingsworth MJ, Martin NC. RNase P activity in the mitochondria of Saccharomyces cerevisiae depends on both mitochondrion and nucleus-encoded components. Mol Cell Biol 1986; 6:1058-64; PMID:3537697
34. Morales MJ, Dang YL, Lou YC, Sulo P, Martin NC. A 105-kDa protein is required for yeast mitochondrial RNase P activity. Proc Natl Acad Sci USA 1992; 89:9875-9; PMID:1409716; http://dx.doi.org/10.1073/pnas.89.20.9875
35. Manam S, Van Tuyle GC. Separation and characterization of 5'- and 3'-tRNA processing nucleases from rat liver mitochondria. J Biol Chem 1987; 262:10272-9; PMID:3301831
36. Doersen CJ, Guerrier-Takada C, Altman S, Attardi G. Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. J Biol Chem 1985; 260:5942-9; PMID:2581945
37. Puranam RS, Attardi G. The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. Mol Cell Biol 2001; 21:548-61; PMID:11134342; http://dx.doi.org/10.1128/MCB.21.2.548-561.2001
38. Rossmanith W, Tullo A, Potuschak T, Karwan R, Sbisà E. Human mitochondrial tRNA processing. J Biol Chem 1995; 270:12885-91; PMID:7759547; http://dx.doi.org/10.1074/jbc.270.21.12885
39. Doersen CJ, Guerrier-Takada C, Altman S, Attardi G. Characterization of an RNaseP activity from HeLa cell mitochondria. J Biol Chem 1985; 260:5942-9; PMID:2581945
40. Yamaguchi-Shinozaki K, Shinozaki K, Sugiura M. Processing of precursor tRNAs in a chloroplast lysate. Processing of the 5' end involves endonucleolytic cleavage by an RNase P-like enzyme and precedes 3' end maturation. FEBS Lett 1987; 215:132-6; http://dx.doi.org/10.1016/0014-5793(87)80127-8
41. Wang MJ, Gegenheimer P. Substrate masking: binding of RNA by EGTA-inactivated micrococcal nuclease results in artifactual inhibition of RNA processing reactions. Nucleic Acids Res 1990; 18:6625-31; PMID:2123540; http://dx.doi.org/10.1093/nar/18.22.6625
42. Ryner LC, Manley JL. Requirements for accurate and efficient mRNA 3' end cleavage and polyadenylation of a simian virus 40 early pre-RNA in vitro. Mol Cell Biol 1987; 7:495-503; PMID:3031477
43. Hall TA, Brown JW. The ribonuclease P family. Methods Enzymol 2001; 341:56-77; PMID:11582805; http://dx.doi.org/10.1016/S0076-6879(01)41145-1
44. Wang MJ, Davis NW, Gegenheimer P. Novel mechanisms for maturation of chloroplast transfer RNA precursors. EMBO J 1988; 7:1567-74; PMID:16453848
45. Rossmanith W, Karwan RM. Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. Biochem Biophys Res Commun 1998; 247:234-41; PMID:9642109; http://dx.doi.org/10.1006/bbrc.1998.8766
46. Thomas BC, Chamberlain J, Engelke DR, Gegenheimer P. Evidence for an RNA-based catalytic mechanism in eukaryotic nuclear ribonuclease P. RNA 2000; 6:554-62; PMID:10786846; http://dx.doi.org/10.1017/S1355838200991477
47. Thomas BC, Gao L, Stomp D, Li X, Gegenheimer PA. Spinach chloroplast RNase P: a putative protein enzyme. Nucleic Acids Symp Ser 1995; 33:95-8; PMID:8643412
48. Oommen A. Characterization of transfer RNA processing enzymes from wheat embryo. PhD Dissertation. Lawrence: University of Kansas, 1991.
49. Arends S, Schon A. Partial purification and characterization of nuclear ribonuclease P from wheat. Eur J Biochem 1997; 244:635-45; PMID:9119034; http://dx.doi.org/10.1111/j.1432-1033.1997.t01-1-00635.x
50. Franklin SE, Zwick MG, Johnson JD. Characterization and partial purification of two pre-tRNA 5'-processing activities from Daucus carrota (carrot) suspension cells. Plant J 1995; 7:553-63; PMID:7742855; http://dx.doi.org/10.1046/j.1365-313X.1995.7040553.x
51. Marchfelder A, Schuster W, Brennicke A. In vitro processing of mitochondrial and plastid derived tRNA precursors in a plant mitochondrial extract. Nucleic Acids Res 1990; 18:1401-6; PMID:2326185; http://dx.doi.org/10.1093/nar/18.6.1401
52. Hanic-Joyce PJ, Gray MW. Processing of transfer RNA precursors in a wheat mitochondrial extract. J Biol Chem 1990; 265:13782-91; PMID:1696257
53. Altman S, Baer MF, Bartkiewicz M, Gold H, Guerrier-Takada C, Kirsebom LA, et al. Catalysis by the RNA subunit of RNase P--a minireview. Gene 1989; 82:63-4; PMID:2479591; http://dx.doi.org/10.1016/0378-1119(89)90030-9
54. Salavati R, Panigrahi AK, Stuart KD. Mitochondrial ribonuclease P activity of Trypanosoma brucei. Mol Biochem Parasitol 2001; 115:109-17; PMID:11377745; http://dx.doi.org/10.1016/S0166-6851(01)00273-0
55. 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 2008; 135:462-74; PMID:18984158; http://dx.doi.org/10.1016/j.cell.2008.09.013
56. Vilardo E, Nachbagauer C, Buzet A, Taschner A, Holzmann J, Rossmanith W. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase--extensive moonlighting in mitochondrial tRNA biogenesis. Nucleic Acids Res 2012; 40:11583-93; PMID:23042678; http://dx.doi.org/10.1093/nar/gks910
57. Williams-Carrier R, Kroeger T, Barkan A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA 2008; 14:1930-41; PMID:18669444; http://dx.doi.org/10.1261/rna.1077708
58. Wang G, Chen HW, Oktay Y, Zhang J, Allen EL, Smith GM, et al. PNPASE regulates RNA import into mitochondria. Cell 2010; 142:456-67; PMID:20691904; http://dx.doi.org/10.1016/j.cell.2010.06.035
59. Rossmanith W. Of P and Z: mitochondrial tRNA processing enzymes. Biochim Biophys Acta 2012; 1819:1017-26; PMID:22137969; http://dx.doi.org/10.1016/j.bbagrm.2011.11.003
60. Anantharaman V, Aravind L. The NYN domains: novel predicted RNAses with a PIN domain-like fold. RNA Biol 2006; 3:18-27; PMID:17114934; http://dx.doi.org/10.4161/rna.3.1.2548
61. 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 2010; 17:740-4; PMID:20473316; http://dx.doi.org/10.1038/nsmb.1812
62. Gutmann B, Gobert A, Giegé P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes Dev 2012; 26:1022-7; PMID:22549728; http://dx.doi.org/10.1101/gad.189514.112
63. Taschner A, Weber C, Buzet A, Hartmann RK, Hartig A, Rossmanith W. Nuclear RNase P of Trypanosoma brucei: a single protein in place of the multicomponent RNA-protein complex. Cell Rep 2012; 2:19-25; PMID:22840392; http://dx.doi.org/10.1016/j.celrep.2012.05.021
64. Lai LB, Bernal-Bayard P, Mohannath G, Lai SM, Gopalan V, Vioque A. A functional RNase P protein subunit of bacterial origin in some eukaryotes. Mol Genet Genomics 2011; 286:359-69; PMID:21987179; http://dx.doi.org/10.1007/s00438-011-0651-y
65. 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 2012; 13:2270-6; PMID:22976545; http://dx.doi.org/10.1002/cbic.201200434
66. Krehan M, Heubeck C, Menzel N, Seibel P, Schön A. RNase MRP RNA and RNase P activity in plants are associated with a Pop1p containing complex. Nucleic Acids Res 2012; 40:7956-66; PMID:22641852; http://dx.doi.org/10.1093/nar/gks476
67. Howard MJ, Liu X, Lim WH, Klemm BP, Koutmos M, Engelke DR, et al. RNase P enzymes: Divergent scaffolds for a conserved biological reaction. RNA Biol 2013; 10: In press; PMID:23595059; http://dx.doi.org/10.4161/rna.24513.
68. Forner J, Weber B, Thuss S, Wildum S, Binder S. Mapping of mitochondrial mRNA termini in Arabidopsis thaliana: t-elements contribute to 5' and 3' end formation. Nucleic Acids Res 2007; 35:3676-92; PMID:17488843; http://dx.doi.org/10.1093/nar/gkm270
69. Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. Int Rev Cytol 2005; 247:165-221; PMID:16344113; http://dx.doi.org/10.1016/S0074-7696(05)47004-7
70. Peck-Miller KA, Altman S. Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from Escherichia coli. J Mol Biol 1991; 221:1-5; PMID:1717693; http://dx.doi.org/10.1016/0022-2836(91)80194-Y
71. Hartmann RK, Heinrich J, Schlegl J, Schuster H. Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of Escherichia coli. Proc Natl Acad Sci USA 1995; 92:5822-6; PMID:7597035; http://dx.doi.org/10.1073/pnas.92.13.5822
72. Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. Proc Natl Acad Sci USA 1994; 91:9223-7; PMID:7524073; http://dx.doi.org/10.1073/pnas.91.20.9223
73. McClain WH, Guerrier-Takada C, Altman S. Model substrates for an RNA enzyme. Science 1987; 238:527-30; PMID:2443980; http://dx.doi.org/10.1126/science.2443980
74. Alifano P, Rivellini F, Piscitelli C, Arraiano CM, Bruni CB, Carlomagno MS. Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA. Genes Dev 1994; 8:3021-31; PMID:8001821; http://dx.doi.org/10.1101/gad.8.24.3021

75. Li Y, Altman S. A specific endoribonuclease, RNase P, affects gene expression of polycistronic operon mRNAs. Proc Natl Acad Sci USA 2003; 100:13213-8; PMID:14585931; http://dx.doi.org/10.1073/pnas.2235589100
76. Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, et al. Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. EMBO J 2003; 22:621-32; PMID:12554662; http://dx.doi.org/10.1093/emboj/cdg040
77. 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 2013; 4:1353; PMID:23322041; http://dx.doi.org/10.1038/ncomms2358
78. Randau L, Schröder I, Söll D. Life without RNase P. Nature 2008; 453:120-3; PMID:18451863; http://dx.doi.org/10.1038/nature06833
79. 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. Proc Natl Acad Sci USA 2012; 109:16149-54; PMID:22991464; http://dx.doi.org/10.1073/pnas.1209062109
80. Frugier M, Bour T, Ayach M, Santos MA, Rudinger-Thirion J, Théobald-Dietrich A, et al. Low Complexity Regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. FEBS Lett 2010; 584:448-54; PMID:19900443; http://dx.doi.org/10.1016/j.febslet.2009.11.004
81. Zeytuni N, Zarivach R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. Structure 2012; 20:397-405; PMID:22404999; http://dx.doi.org/10.1016/j.str.2012.01.006
82. Glavan F, Behm-Ansmant I, Izaurralde E, Conti E. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J 2006; 25:5117-25; PMID:17053788; http://dx.doi.org/10.1038/sj.emboj.7601377
83. Xu J, Peng W, Sun Y, Wang X, Xu Y, Li X, et al. Structural study of MCPIP1 N-terminal conserved domain reveals a PIN-like RNase. Nucleic Acids Res 2012; 40:6957-65; PMID:22561375; http://dx.doi.org/10.1093/nar/gks359
84. Bhagwat M, Meara D, Nossal NG. Identification of residues of T4 RNase H required for catalysis and DNA binding. J Biol Chem 1997; 272:28531-8; PMID:9353315; http://dx.doi.org/10.1074/jbc.272.45.28531
85. LaGrandeur TE, Darr SC, Haas ES, Pace NR. Characterization of the RNase P RNA of Sulfolobus acidocaldarius. J Bacteriol 1993; 175:5043-8; PMID:7688716
86. Pan T, Loria A, Zhong K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. Proc Natl Acad Sci USA 1995; 92:12510-4; PMID:8618931; http://dx.doi.org/10.1073/pnas.92.26.12510
87. Kirsebom LA, Svärd SG. Base pairing between Escherichia coli RNase P RNA and its substrate. EMBO J 1994; 13:4870-6; PMID:7525271
88. Kurz JC, Niranjanakumari S, Fierke CA. Protein component of Bacillus subtilis RNase P specifically enhances the affinity for precursor-tRNAAsp. Biochemistry 1998; 37:2393-400; PMID:9485387; http://dx.doi.org/10.1021/bi972530m
89. Wu S, Kikovska E, Lindell M, Kirsebom LA. Cleavage mediated by the catalytic domain of bacterial RNase P RNA. J Mol Biol 2012; 422:204-14; PMID:22626870; http://dx.doi.org/10.1016/j.jmb.2012.05.020
90. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA 1993; 90:6498-502; PMID:8341661; http://dx.doi.org/10.1073/pnas.90.14.6498
91. Thomas BC, Li X, Gegenheimer P. Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. RNA 2000; 6:545-53; PMID:10786845; http://dx.doi.org/10.1017/S1355838200991465
92. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS One 2013; 8:e57286; PMID:23472078; http://dx.doi.org/10.1371/journal.pone.0057286
93. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006; 22:195-201; PMID:16301204; http://dx.doi.org/10.1093/bioinformatics/bti770

1.7 PRORP: A Pentatricopeptide repeat protein

1.7.1 PPR distribution in all domains of life

Pentatricopeptide repeat (PPR) proteins are absent in most sequenced bacteria expect for *Ralstonia solanacearum* (Salanoubat *et al.*, 2002) and can be found in fungi and metazoa. Eurkaryotes possess an average of 5-50 PPR proteins. This protein family expanded in plant organelles, i.e. land plants (Small & Peeters, 2000) with more than 450 in *Arabidopsis* and about 470 in rice. In human only the small number of seven PPR proteins could be detected (LRPPRC, MRPS27, PTCD1-3, MRPP3, POLRMT), 28 in the protist *Trypanosoma brucei* (Pusnik *et al.*, 2007) and 15 in yeast *S. cerevisiae* (Lipinski *et al.*, 2011).
A hypothesis explaining this explosion in land plants was given by Chrzanowska-Lightowlers in 2013. Given the relaxed organization of plant organellar DNA and its plasticity with a lot of introns and editing sites the nuclear genome would have co-evolved to come up with a family of RNA binding proteins that counteract these developments (Lightowlers & Chrzanowska-Lightowlers, 2013).

1.7.2 PPR classifications

Classical PPR motifs consist of a degenerate sequence of 35 amino acids that fold up into two antiparallel α-helices. In *Arabidopsis* a typical PPR protein contains up to 30 of such repeat modules. PPR proteins can be divided in two subfamilies: P-class and PLS-class. Proteins of the P-class contain the archetype of PPR motifs: 35 amino acids with a degenerate sequence forming two anti-parallel α-helices and no additional domains. PPR proteins in general contain 2-30 PPR motifs in tandem arrays. The PLS-class is characterized by a repetition of P-motifs (35 aa), L-motifs (long motifs with 35-36 aa) and S-motifs (short motifs with 31 aa). In this subfamily there are proteins containing a C-terminal E and DYW domain. E stands for extended domain and DYW represents the characteristic last C-terminal amino acids aspartate-tyrosine-tryptophan (Fig. 1.19). PLS PPR proteins are implicated in RNA editing. The E domain is essential for RNA editing whereas the DYW domain seems to be facultative. Interestingly, although RNA specific binding has been demonstrated RNA editing remains to be demonstrated *in vitro*. PRORPs belong to another group having only P-motifs and an additional C-terminal domain (Schmitz-Linneweber & Small, 2008, Barkan & Small, 2014).

1.7.3 Functions of PPR proteins

The P-class PPR proteins in chloroplasts stabilize for example mRNAs by protecting them against 5'-3' exonucleases and guiding them to their correct cleavage site. Yet, in mitochondria they are mostly implicated in RNA cleavage. The can also act as translation activators as in the case of *atpH* whose translation is stimulated by PPR10. PPR proteins are implicated in preventing plants from cytoplasmic male sterility (producing pollen), called *restorer-of-fertility*-genes. In that way the nuclear genome counteracts the attack of organellar evolution that might harm the plant viability. P-class PPR proteins are also splicing factors (e.g. THA8, PPR4, PPR5) for group II introns in chloroplasts and mitochondria where these introns lost their self-splicing capacity.

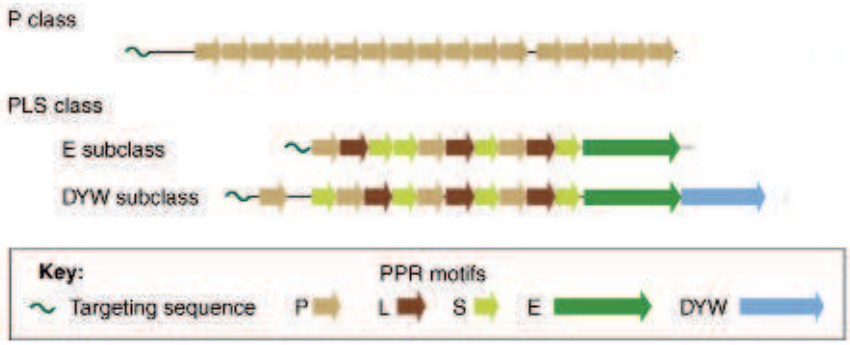


Figure 1.19: Classification of P- and PLS-PPR-proteins. P motifs are the classical 35-mer PPR that may contain additional motifs such as PRORPs. PLS-PPR proteins consist of arrays of 35-mer (P), 35-36-mer (L) and 31-mer (S) motifs that are fused to additional domains (E and DYW). (Schmitz-Linneweber & Small, 2008)

PLS-class PPR proteins are the main actors in RNA editing that is the deamination of cytidine to uridine. In most of the cases these editing factors have one, rarely more, targeting site, reflecting their high specificity.

1.8 The RNA recognition code

RNA recognition is modular: One repeat recognizes one nucleotide mode (Kobayashi *et al.*, 2012, Yagi *et al.*, 2013). Taking the numeration of Ian Small and co-workers the amino acid at position 6 and the first amino acid of the following repeat (1') would specifically recognize one nucleotide. They found that $T_6D_{1'} \rightarrow G, T/S_6N_{1'} \rightarrow A, N_6D_{1'} \rightarrow U$and $N_6N/S_{1'} \rightarrow C$. In contrast to other RNA binding motifs such as PUF and TALE proteins, PPR bind their target RNA in a parallel manner (5'-3' to N-C-terminal) being also able to recognize much longer single RNA tracts ranging from 12 to 29 nts (Barkan & Small, 2014, Barkan *et al.*, 2012).
A well-studied example of PPR proteins is the *maize* PPR10 that consists of 19 PPR motifs (Prikryl *et al.*, 2011) and is localized to chloroplasts. Its function is to guide endo- and exonucleases to their correct site on mRNAs by sequestering the mRNA 3' and 5' termini. It is also implicated in transcription control by unmasking the Shine-Delgarno sequence on the mRNA and thereby increasing transcription levels (Prikryl *et al.*, 2011). Using this model PPR protein Barkan and co-workers established a RNA recognition code that was also confirmed using modified RNA sequences. Analyzing footprinting data they found a minimal RNA sequence of 17 nts that is in contact with 19 PPR motifs.

1.9 Structural information on PPR/RNA interactions

The first structural data of PPR proteins was published in 2011 by Temiakov and co-workers from the human mitochondrial RNA polymerase (pdb ID: 3SPA). This structure confirmed the predicted fold of two antiparallel α-helices per PPR motif and shows two PPR motifs in the N-terminal region of the enzyme (residues 263 - 330). Their presence is essential for the formation of the RNA polymerase initiation complex (Ringel *et al.*, 2011). One year later the crystal structure of organellar PRORP1

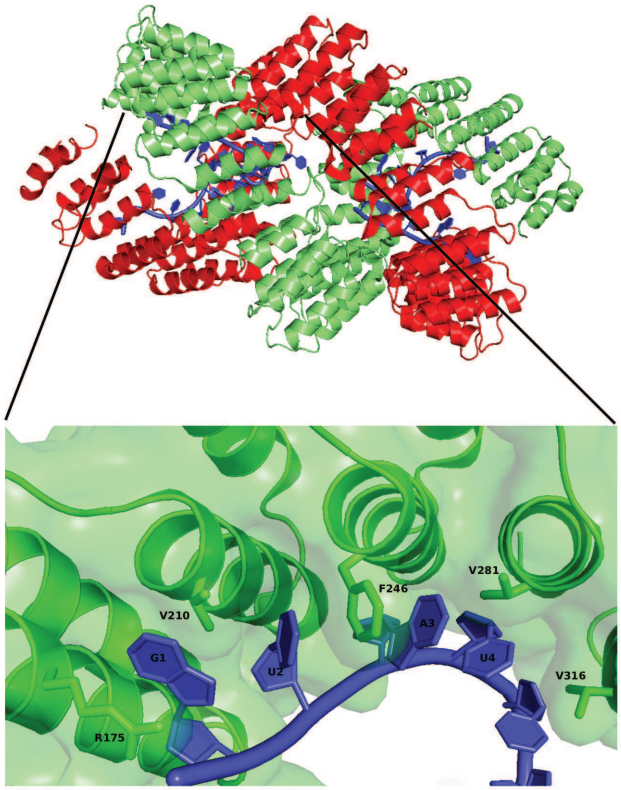


Figure 1.20: *Maize* PPR10 in complex with PSAJ RNA. Overall structure of the dimeric complex of PPR10 (green and red) and PSAJ RNA (blue) as well as a zoom into the region of RNA protein contacts. In the PSAJ sequence 5'-GUAU-3' each nucleotide is surrounded by four amino acids. Hydrogen bonds are made between Thr178, Asn213, Ser249 and Asn284 and 5'-GUAU-3', respectively (not shown). In addition each nucleotide is sandwiched by two hydrophobic amino acids, i.e. G1 by Arg175/Val210, U2 by Val210/Phe246, A3 by Phe246/Val281 and U4 by Val281/Val316.

from *Arabidopsis* was solved at 1.75 Å(pdb ID: 4G23) (Howard *et al.*, 2012). Since then more and more crystallographic data of PPR proteins has become available and in 2013 the first structure of PPR10 from *Arabidopsis* in complex with its substrate RNA, PSAJ, was published (pdb ID: 4M59, Fig. 1.20).

Yan *et al.* reported a dimeric complex of two intertwined PPR10 molecules. The PSAJ recognition seen in the structure is the modular recognition of the first four 5' nts of PSAJ and two nts at the 3' end bound to the C-terminus of the second PPR10 molecule. The data confirmed the concept of modular recognition and the predicted code for 6 out of 18 nts. It is still puzzling that previous data on the oligomeric state in solution of PSAJ and PPR10 showed a monomeric population (Yin *et al.*, 2013, Barkan *et al.*, 2012). It needs to be vaildated whether the crystallized complex reflects an *in vivo* complex or a crystallographic artefact.

1.10 Objectives of my thesis

When I started the project in master little was known about the structural organization of PPR proteins to which PRORP proteins belong. The only information on structure came from bioinformatic

predictions.

In this context, the aim of my project was to characterize PRORP proteins from the model plant *A. thaliana* in solution but also to crystallize them alone and in a complex with tRNA to determine their 3D crystal structure.

Optimizing purification protocols for both proteins and RNA were necessary to obtain pure and homogeneous samples in sufficient amounts for structural studies. To maximize crystallization probability I worked on PRORP1-2-3 in parallel as well as on their catalytic mutants. In addition I carried out biophysical and structural analyses on soluble samples to obtain complementary information on both enzymes alone as well as on their complex with pre-tRNAs. This multidisciplinary approach aimed to bring a broader view and deeper insight into how PRORPs recognize and cleave their substrates. The ultimate goal was to determine the structure of a PRORP/tRNA complex contributing to deciphering to the knowledge of how an enzyme completely replaced an ancient ribozyme like RNase P in plants.

Material and methods

2.1 Materials

2.1.1 Bacterial strains

E.coli strains *BL21 (DE3), DH5α, Top10, Rosetta2* and *B834 (DE3)* strains were used during this work. Strains used for cloning and plasmid purification, *Top10* and *DH5α* contain mutations in genes *recA1* and *endA1* coding for a recombinase and an endonuclease. This prevents plasmid degradation, recombination and multimerization (Matsen, 2014). Strains *BL21(DE3)* and *Rosetta2(DE3)pLys* were used for protein expression. *BL21(DE3)* express T7 RNA polymerase that is under the control of a lacUV5 promoter and induced by IPTG. In the presence of T7 polymerase the protein cDNA in a pET28 vector containing a T7lac promoter will be expressed. The basal T7 transcription can be reduced by adding 1% (w/v) glucose during cell growth (Biolabs, 2014). *Rosetta2(DE3)* cells contain an additional plasmid that encodes rare tRNAs and enhances the expression of proteins containing a lot of rare codons.

Genotypes are:
Top10: F-mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-
DH5α: F-endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG ϕ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ-
BL21(DE3): F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
Rosetta2(DE3): F-ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (Cam^R)

2.1.2 Plasmids

2.1.2.1 Protein expression vectors

pET28b(+)

This vector is used to express His_6-tagged proteins that are purified by metal ion affinity chromatography. It is a high copy number protein expression vector of 5368 bp containing a T7 promoter, a kanamycin resistance gene and an IPTG regulated lactose operon. The cDNAs are cloned in frame with a 3' end His_6-tag.

pTYB1

PTYB1 is used to express and purify proteins without a tag sequence. This low copy number

C-terminal fusion vector of 7477 bp is designed for the insertion of a target gene into a multiple cloning site upstream of the *S. cerevisiae* VMA intein and a chitin binding domain (CBD) construct containing an ampicillin resistance gene (Thomas, 2014).

pTYB12

Same characteristics as pTYB1 but the intein-CBD is N-terminal to the protein of interest and a length of 7417 bp.

2.1.2.2 RNA transcription vectors

pUC19

pUC19 is a high copy number cloning vector for *E.coli* containing an ampicillin resistance gene. It is 2686 base pairs long and is used for pre-tRNA cloning. Pre-tRNA sequences were amplified with primers containing the T7 RNA polymerase promoter sequences and ligated in pUC19.

2.1.3 Protein constructs

The clones for protein expression, pET28b-PRORP1-2-3 as well as their catalytic mutants, were already available in my group at the IBMP.
PRORP2-3 are full length constructs containing a C-terminal His_6-tag after expression. PRORP1 cDNA was cloned without the mitochondrial targeting signal (MTS) predicted by Predotar (Small, 2003) and Target P (Nielsen *et al.*, 1997). In addition double catalytic mutants were used. Two essential aspartates in the catalytic site that are conserved among PRORPs were mutated into alanines, and are referred to as PRORP2mDD for double catalytic mutant PRORP2. Mutated aspartates in PRORP1-2-3 are D474/D475, D421/D422 and D480/D481, respectively (Gobert *et al.*, 2010).
During the optimization of the purification protocol of PRORP1 I observed two bands of similar size that were neither separable with gel filtration nor with ion exchange chromatography. The lower band corresponded to a N-terminal degradation product that was still upstream of the first predicted PPR motif. Thus, we cloned a shorter version of PRORP1-MTS referred to as PRORP1-cris.

2.1.4 RNA substrates

Arabidopsis thaliana mitochondrial $tRNA^{Cys}$ (Accession code NC_001284.2 localization 104885 - 104955) was cloned into RNA transcription vector pUC19 downstream the T7 promoter and endonuclease restriction sites EcoR1 or BmR1 depending on the the 3' trailer sequences. BmR1 was used for mature 3' ends and EcoR1 for substrates containing a trailer sequence. In this case a mature 3' end refers to an end having the discriminator base but no CCA.

Table 2.3: *A. thaliana* mitochondrial tRNA^{Cys} sequences used during the work

Name	Sequence
L0T0	**GGCUAGGUAACAUAAUGGAAAUGUAUCGGACUGCAAAUCCUGUAAUGACGGUUCGACUCCGUCCUUGGCCU**
L5T0	GGGUU**+L0T0**
L51T0	GAGAGGAAGAAAGAACAACCGUUUUACUUUGGCACAUGAGGUGGCGGGUUU**+L0T0**
L5T30	GGUUU**+L0T0+**ACACCUUCAUGACCAGAAAUAACU
L11T30	GUGGCGGGUUU**+L0T0+**ACACCUUCAUGACCAGAAAUAACUG
L21T30	GGCACAUGAGGUGGCGGGUUU**+L0T0+**ACACCUUCAUGACCAGAAAUAACUG
L31T30	GUUUUACUUUGGCACAUGAGGUGGCGGGUUU**+L0T0+**ACACCUUCAUGACCAGAA AUAACUG
L41T30	GAGAACAACCGUUUUACUUUGGCACAUGAGGUGGCGGGUUU**+L0T0+**ACACCUUC AUGACCAGAAAUAACUG
L51T30	GAGAGGAAGAAAGAACAACCGUUUUACUUUGGCACAUGAGGUGGCGGGUUU**+L0T0+**ACACCUUCAUGACCAGAAAU AACUG
L5T0 MAC	GGGUUGGCUAGGUAACAUAAUGGAAAUGUAUAAUGACGGUUCGACUCCGUCCUUGGCCU
L5T0 GUC	GGGUUGGCUAGGUAACAUAAUGGAAAUGUAUCGGACU**GUC**AAUCCUGUAAUGACGGUUCGACUCCGUCCUUGGCCU

2.1.5 Primers

Table 2.4: Primer sequences used for protein and tRNA cloning and sequencing

Name	Sequence
PRORP2	
P2int_f	TAAAGCTAGCTCTGATCAACACCGGTCTC
P2-nlsint_f	TAAAGCTAGCAGAAACCCAGAAACAAATCTCC
InteinP2_f1	TAAAGAATGCTGCTGCTTCTGATCAACACC
InteinP2_f2	TAAAGAATGCTTCTGATCAACACCGGTCTC
InteinP2_r	TAAACTCGAGCTAAGGAATCTTCCCATTACTCTTAGG
tRNA^{Cys}	
ptrnCmT0_rv (G61)	AGGCCAAGGACGGAGTCGAAC
ptrnCm-T7G_f (G59)	TAAAGGATCCGAATTGTAATACGACTCACTATAGGGTTGGCTAGGTAACATAATGG
mut-tRNAcysL5T0_f	GTATCGGACTGCTAATCCTGTAATG
mut-tRNAcysL5T0_r	CATAGCCTGACAGTTAGGACATTAC
306.ptrnC-m-ara-L05_F (II-G25)	TAAAGGATCCGAATTGTAATACGACTCACTATAGGTTTGGCTAGGTAACATAATGG
306.ptrnC-m-ara-T05_R (VI G95)	TAAAGAATTCACTGGGAATGTGGTGTAGGCCAAGGACG
PRORP1 cris	
G76 306.AtP1cris_F	GAAGGAGATATACCATGGCAGCTTCTCCTTCTGAAAAC
G77 306.AtP1cris_R	CAGAAGGAGAAGCTGCCATGGTATATCTCCTTCTTAAAG
Sequencing PRORP2	
19.z0774 (306.N570211-LB)_f	TTTGAATAATGGAGGTGGGTG
P2 rv (20.z0880 (MRPP32-smart1_b))_r	AATACCTTGCCATTGACCGTG
20.z0897 (306.F435C08-RB_f	TGGTTTCGAGATTTTTGATCG
20.z0879 (MRPP32-smart1)_r	TGGTTAGTTCGGGGATAAGTCCT
VI.G39 (306.P2end_F)	TGCTTGTGACAAATGATGAG
VII.G60 (306.P2e2_RV)	TGAGACAAATGTGGTCTTCAGC

2.1.6 Devices

Concentration determination

NanoDrop® ND-1000 Spectrophotometer, ThermoScientific

Centrifugation

Centrifuge SIGMA 3-16PK, Fisher Scientific, or a similar centrifuge

Sorvall Hitachi Discovery M150SE micro-ultracentrifuge

Cell lysis

Ultrasons Annemasse Tech device

Chromatography systems

BioLogic ^{TM} DuoFlow, BioRad

and Äkta pure, GE Healthcare

Chromatography media and other columns

HIS-Select® HF Nickel Affinity Gel, Sigma, No. H0537

Chitin Resin, New England Biolabs, No. S6651

Anion exchange medium: HiTrap Capto DEAE, GE Healthcare, No. 28-9165-37

Size exclusion chromatography columns

Superdex 200, 10/300, GE Healthcare, No. 17-5175-01;

HiLoad 16/30, GE Healthcare

Superdex 75, 10/300, GE Healthcare, No. 17-5174-01

Bio SEC-3, 300 Å, 4.6 $\times$ 300 mm, 3 mm, Agilent

Bio SEC-3, 150 Å, 4.6 $\times$ 300 mm, 3 mm, Agilent

Macromolecule Concentration

Amicon Ultracentrifugal Filter Units, Millipore, 10K and 30K MWCO

DLS devices

Dynapro Nanostar, Wyatt

Zetasizer NanoSeries Nano-S, ZEN1600, Malvern

ITC measurements

ITC200 Isothermal titration Calorimeter, MicroCal, Inc., GE Healthcare

Crystallization screen pipetting robot

Mosquito® Crystal, TPP LabTech

2.2 Methods

2.2.1 Protein production

2.2.1.1 PRORP overexpression in E. coli Rosetta 2 cells

For His-tagged proteins, *E. coli Rosetta 2* cells were transformed with the vector pET-28b(+) carrying the *prorp* cDNA. PRORP constructs include a C-terminal His_6-tag.
The starter culture was grown for 16 h at 37 °C on an orbital shaker with 160 rpm in LB medium containing 1% (m/v) glucose to repress protein expression efficiently (for pET28) and the respective antibiotic (Tab. 2.5) to select the colonies containing the construct. 1 l of LB medium (1% (m/v) glucose and the respective antibiotic were inoculated with 10 ml of the starter culture and grown at 37 °C on an orbital shaker with 160 rpm up to an OD_{600nm} of 0.6. Bacteria were pelleted (5,000 g, 4°C, 20 min) and washed once with LB medium. The pellet was resuspended in 1 l fresh LB medium with antibiotic and protein expression was induced by adding IPTG at a final concentration of 0.5 mM. Protein expression was carried out for 16 h at 18 °C on an orbital shaker with 160 rpm. Cells were centrifuged (5,000 g, 4°C, 20 min) washed with LB medium, centrifuged and the pellet was stored at -20 °C or directly used for purification. One gramm of wet weight of cells were lysed by sonication in 10 ml lysis buffer (100 ml lysis buffer contain one tablet of protease inhibitor) doing ten cycles of 30 sec sonication at 120 V and 60 sec on ice.

Table 2.5: Concentrations of antibiotics used during protein purification depending on plasmid and cell type.

	pET28b [μg/ml]	pTYB [μg/ml]
E.coli Rosetta 2	Kanamycin [50], Chloramphenicol [34]	Chloramphenicol [34], Ampicillin [100]
E.coli BL21 (DE3)	Kanamycin [50]	Ampicillin [100]

LB medium (1 l): 10 g NaCl, 10 g tryptone, 5 g yeast extract
Kanamycin: 50 mg/ml in water
Chloramphenicol: 34 mg/ml in ethanol
Glucose: 20 % (m/v) in water
TCEP: 1 M in water (Sigma, No. C4706)
Lysis buffer: 15% (w/v) glycerol, 50 mM Hepes-Na pH 7.5, 5 mM imidazole, 250 mM NaCl, 1 mM TCEP, 1 anti-protease tablet in 100 ml
Chitin buffer: 15% (w/v) glycerol, 50 mM Hepes-Na pH 7.5, 500 mM NaCl, 1 mM TCEP, 1 anti-protease tablet in 100 ml
Protease inhibitor: SIGMAFAST^{TM} Protease Inhibitor Cocktail Tablets, EDTA-Free (SIGMA, No. S8830), 1 tablet in 100 ml lysis buffer

2.2.1.2 Affinity chromatography of His-tagged proteins on a nickel column

Metal ion affinity chromatography is based on the coordination of metal ions (e.g. nickel) to the imidazole ring of histidines. Artificially His-tagged proteins interact strongly with the nickel-charged nitrilotriacetic acid agarose matrix.
All purification steps were done using a Biologic HPLC system or an Äkta pure system. The protein lysate was loaded on a 5 ml column containing NiNTA matrix equilibrated with buffer A. After protein loading the column was washed with 8 column volumes (cv) of buffer A, then buffer mixtures (A/B) containing 10 mM (8 cv) and 15 mM (8 cv) imidazole to clean the column of non-specifically bound proteins. The His-tagged proteins were eluted with 3 cv of a buffer containing 250 mM imidazole (which displaces the His-tagged proteins) and another 2.5 cv with 500 mM imidazole to strip off all other proteins. Flow rate was 0.5 ml/min for protein loading and 2 ml/min for column washing and protein elution. Protein elution was monitored by absorption at 280 nm. After each purification the matrix was regenerated as recommended by the manufacturer. Protein quality and quantity was assessed on SDS-PAGE gels containing 8% (w/v) polyacrylamide and with a spectrophotometer. To do this, 4 to 16 μl of each fraction were mixed with 4 μl loading buffer and heated at 95°C for 5 min. Five μl of marker proteins (BioRad, No. 161-0363) were used as size reference. The gel was run at 80 V in the stacking gel and at 150 V in the separation gel, then stained with heated Coomassie blue R-250 for 15 min and unstained in unstaining solution. Fractions containing PRORP were pooled, concentrated and the buffer was exchanged for SEC buffer using membranes with a 30K MWCO .

Buffer A: 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 5 mM Imidazole, 1 mM TCEP
Buffer B: 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 500 mM Imidazole, 1 mM TCEP
Polyacrylamide-Protein gel: Rotiphorese® Gel 30 (37.5:1): 30 % acrylamide/bisacrylamide, mixing ratio 37.5:1
Protein ladder: Precision Plus Protein Unstained Standard, BioRad (Cat.No. 161-0363); ThermoFishherScientific (No. 26619)
PAGE-Loading buffer: 90%: 0.1% (m/v) bromphenol blue, 25% (w/v) glycerol, 5% (w/v) SDS, 156.25 mM Tris-HCl pH 6.8 and 10 mM β-mercaptoethanol
PAGE-Running buffer (10X): 144 g glycine, 10 g SDS, 30 g Tris-HCl
PAGE-Staining solution (1 l): 2.5 g Coomassie Brilliant Blue R-250, 100 ml glacial acetic acid, 400 ml H_2O, 500 ml methanol
Unstaining solution: 10 % (v/v) ethanol, 7.5 % (v/v) acetic acid, 600 ml H_2O

2.2.1.3 Affinity chromatography of the intein-tagged proteins on a chitin matrix

Intein purification system

Inteins are protein sequences that cleave autocatalytically of a protein sequence post-translationally. The first described intein was an H^+-ATPase, VMA1, from *S. cerevisiae*. It shows sequence homology to known homing endonucleases (Gimble F., 1992). Self-splicing proteins are now commonly used in protein one-step purification systems in biotechnology. Inteins are bipartite and fused at one side to an affinity tag, i.e. chitin binding domain (CBD), that will specifically bind to a matrix and on the other side to the protein of interest. Mutations in the intein sequence, N- or C-terminal, prevent N- or C-terminal cleavage from the chitin binding domain. Upon addition of reducing agents like β-mercaptoethanol, DTT or cystein the intein will cleave itself from the protein of interest (Fig. 2.21). It will stay attached to the CBD as the cleavage is not complete (Anraku & Satow, 2009).

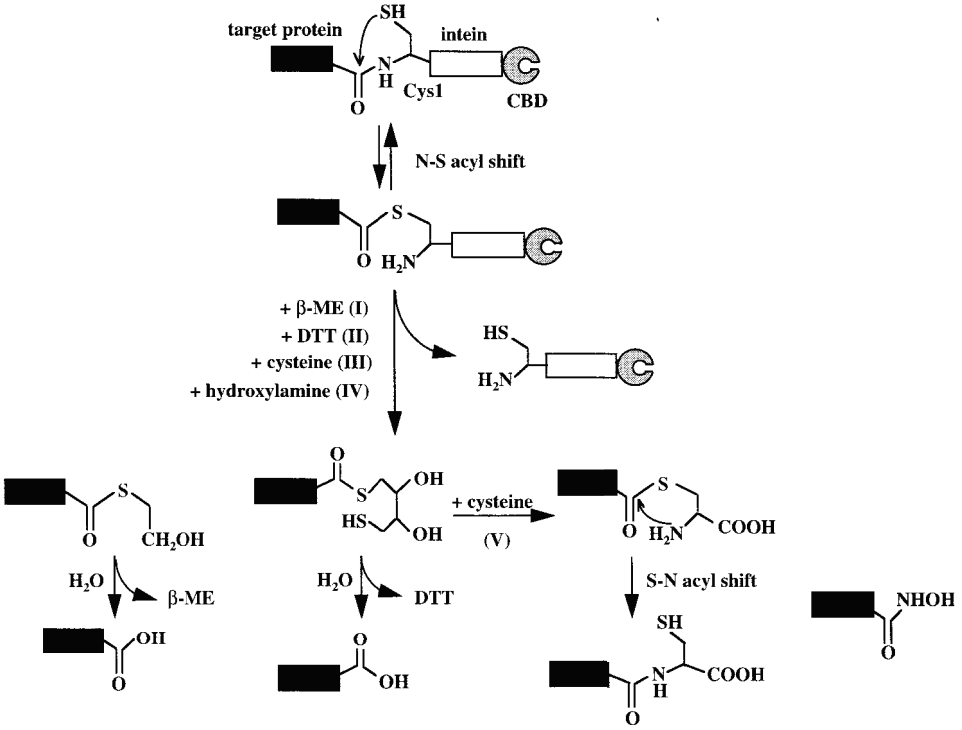


Figure 2.21: The chemical mechanism of the intein N-terminal cleavage reaction. A site specific mutation that is responsible for complete C-terminal intein cleavage was mutated. This construct, when inserted between a target protein and a CBD, leads to release of the protein of interest. The intein remains coupled and fixed to the column matrix via the CBD. The presence of reducing agents like β-mercaptoethanol (I), DTT (II), cysteine (III) or hydroxylamine (IV) induces cleavage of the thioester bond (Chong *et al.*, 1997).

PRORP sequences were cloned into a vector coding for a N- or C-terminal chitin binding domain. After overexpression tests only the C-terminal fused constructs were kept for expression and purification. The active form and the catalytic mutant of PRORP2 with and without the nuclear localization signal were cloned using this approach.

The chitin column was equilibrated with chitin buffer at 2 ml/min and the protein lysate was loaded onto the column at 1 ml/min. The column was washed with at least 10 cv of chitin buffer (without protease inhibitor) at 1.5 ml/min. 5 cv of cleavage buffer were passed at the same flow and the column was incubated 16 h at 4 °C in this buffer.
The next morning the protein was eluted in 4 cv at 1.5 ml/min in fractions of 2 ml and a second cleavage was performed. Fractions containing the protein after affinity chromatography were pooled and concentrated and the buffer was exchanged for SEC buffer using membranes with a 30K MWCO.
The chitin resin can be used five times and needs to be stripped after each use of the CBD:

1. 3 cv 0.3 M NaOH at 1.5 ml/min
2. Incubation 30 min
3. 7 cv 0.3 M NaOH at 1 ml/min
4. 20 cv H_2O at 1.5 ml/min

Cleavage buffer: 15% (w/v) glycerol, 50 mM Hepes-Na pH 7.5, 500 mM NaCl, 50 mM DTT, 1 mM TCEP

2.2.1.4 Size exclusion chromatography

Size exclusion chromatography (SEC) is a method to separate molecules by their hydrodynamic properties, i.e. volume and shape. Proteins were separated on a Superdex 200 10/300 GL column with a separation range of 10 - 600 kDa. The column was equilibrated at 0.6 ml/min in SEC1 buffer. The concentrated protein pool after affinity chromatography was loaded by manual injection. A maximum of 500 μl of the concentrated protein pool were injected in a 1 ml static loop per run and the protein was eluted with a flow rate of 0.5 ml/min (pressure $\leq$ 4 MPa). The loop was rinsed with 3 loop volumes to load a maximum of the protein onto the column. The proteins were separated at 0.6 ml/min and fractions of 0.3 or 0.5 ml were collected. Purest protein fractions according to SDS-PAGE were pooled and concentrated at up to 10 mg/ml and ultracentrifuged at 125,000 g for 1 h before use or storage at 4 °C.

SEC1 buffer : 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 1 mM TCEP

2.2.2 RNA production

2.2.2.1 Template preparation

Plasmid preparation Transformed *E.coli Top10* cells were grown overnight at 37 °C in 250 ml LB/Ampicillin (100 μg/ml) and the plasmid was purified using a kit (NucleoBond Extra Maxi, Machery & Nagel) and eluted in two volumes of 400 μl of 1X TE buffer.

Ampicillin: 100 mg/ml in water
1X TE buffer: 10 mM Tris pH 8, 1 mM EDTA

PCR 800 μl reaction medium contain 160 ng plasmid DNA, 200 μM of each dNTP, 1 U/50 μl PCR reaction of Phusion DNA polymerase (2 U/μl, New England Biolabs, M0530L), 0.5 μM of each primer and 1X HF buffer (NEB). The PCR reaction product was purified using a NucleoSpin Gel and PCR clean-up kit (Machery and Nagel, No. 740609).

PCR protocol:

Initial denaturation: 98°C 30 seconds

30 cycles:

- Denaturation: 98°C 10 seconds
- Annealing: 60°C 30 seconds
- Elongation: 72°C 30 seconds

Final extension: 72°C 10 min

2.2.2.2 In vitro transcription

Large scale transcription *In vitro* transcription was performed in a reaction mix containing 10 mM NaCl, 30 mM $MgCl_2$, 2 mM spermidine, 40 mM Tris-HCl pH 8, 10 mM DTT, 4 mM of each rNTP, 5 mM GMP, 0.01 % Triton, pyrophosphatase 2 ng/μl, 100 nM PCR template and 0.01 mg/ml T7 at 37 °C for 4 - 5 h. The T7 RNA polymerase was produced according to a protocol from Guillaume Bec using *E.coli BL21(DE3)* containing the T7 RNA polymerase cDNA in a RIL/pBH161 vector. To assess the transcription level 10 μl of the reaction mix and 5 μl of RNA loading dye were incubated 5 min at 70 °C and and loaded on a 12 % polyacrylamide/8 M urea gel in 1X TBE buffer. The gel was run at 250 V and stained with ethidium bromide.

RNA loading dye: 95% formamide, 20 mM EDTA, 0.05 % (w/v) bromophenol blue, 0.05 % (w/v) xylen cyanol

Polyacrylamide-RNA: Rotiphorese® Gel 40 (19:1): 40 % (w/v) acrylamide/bisacrylamide, mixing ratio 19:1

Small scale transcription For small scale transcription the RiboMAX^{TM} kit (Promega) was used. 20 μl transcription reaction mix contained 1 μg linearized DNA, 7.5 mM of each rNTP, 2 μl of enzyme Mix, T7 Express (T7 RNA Polymerase, Recombinant RNasin® Ribonuclease Inhibitor and Recombinant Inorganic Pyrophosphatase) and 1X T7 Express buffer. The reaction was incubated 3 h at 37 °C. Plasmid DNA was digested by adding 2 μl DNase I for 15 min at 37 °C. 30 μl of water were added and RNA extracted as described below. The pellet was resuspended in 50 μl of water and purified from excess nucleotides on a G50 resin.

2.2.2.3 Polyacrylamide gel electrophoresis

Proteins used during *in vitro* transcription assays were eliminated using an equal volume of a saturated phenol/chloroform solution pH 5.2, vortexing and transferring the upper aqueous phase into a new tube. Nucleic acids were precipitated with 2.5 volumes of cold ethanol, and 1/10 volume of sodium acetate pH 5 for at least one hour at -20 °C. The nucleic acids were centrifuged at 21.000 g for 25 min at 4 °C. The pellet was dried in a speed vacuum concentrator and stored until usage at -20 °C. A pellet of 5 ml *in vitro* transcribed RNA was dissolved in 600 μl H_2O, 100 μl EDTA, 300 μl RNA loading dye and 200 μl 50 % (w/v) glycerol to fit into the big gel pocket.
A 12 % PAA/8 M urea gel (33 x 40 cm) in 1X TBE buffer was casted. Before use it was preheated for at least 6 hours at constant current of 600 V. The gel was run 16 h at 600 V. The RNA was visualized by UV shadowing, the corresponding gel was cut, sliced in small pieces and RNA eluted.

Passive elution RNA was eluted form the ploy acrylamide gel passively for 16 h at 4 °C in RNA elution buffer. For the elution 2 ml Eppendorff tubes were half filled with gel slices and RNA elution buffer and placed on a test tube rotator. The gel slices were separated from buffer via centrifugation at 1.000 g using syringes clogged with glass wool (Fig. 2.22a). The final RNA solution was concentrated and transferred into the desired buffer using Amicon filter units with 10K MWCO.

Electroelution Electroelution was performed twice for 1 h at 150 V using a Whatman® Elutrap electroelution system. The device was filled with 1X TBE buffer and closed with non-permeable membranes (Fig.2.22b, 1). The chamber was placed in an electric current so that RNA was eluted into 800 μl 1X TBE buffer through a semi-permeable membrane (Fig.2.22b, 2).

TBE, 10X: 1 M Tris base, 1 M boric acid, 20 mM EDTA
RNA elution buffer: 500 mM sodium acetate pH 5, 1 mM EDTA, 0.1 % SDS

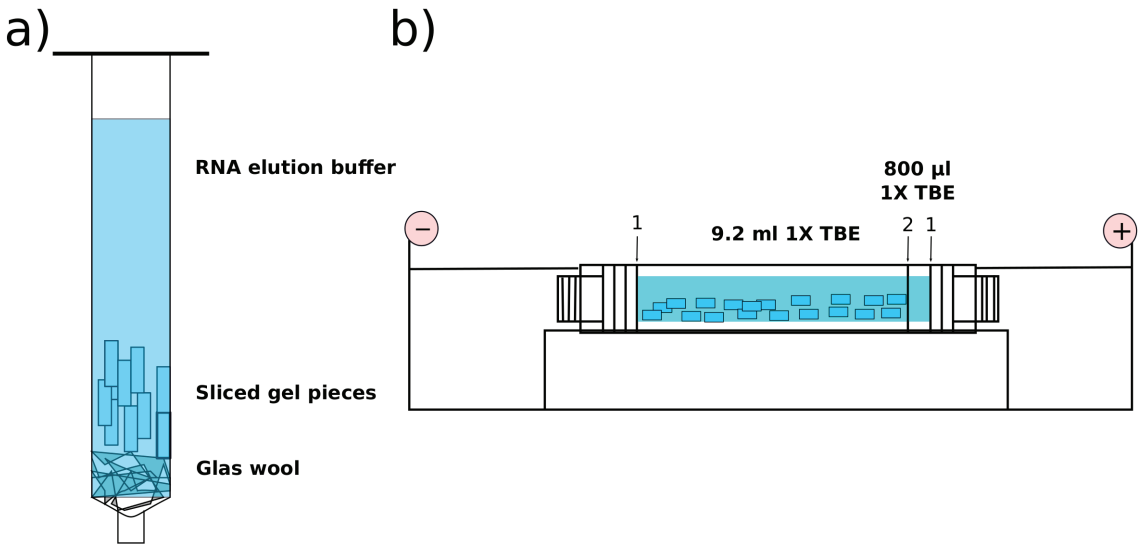


Figure 2.22: RNA gel elution. a) Setup to dry the gel slices after passive elution. A disposable syringe is clogged with sterile glass wool and wet gel sliced stacked on the top of it. To dry the slices syringes are placed in 15 ml Falcon tube and spinned at maximal 2,000 g for one minute, b) schematic representation of a Elutrap electroelution chamber. Gel slices are placed in a solution of 9.2 ml 1X TBE buffer and eluted for 1 h at 150 V two times in 800 μl 1X TBE buffer.

2.2.2.4 Chromatography

The advantage of RNA purification by chromatographic methods is that the RNA remains native during all steps once folded during transcription (Uhlenbeck, 1995).

Anion exchange Diethylaminoethyl (DEAE), a weak anion exchange medium, was selected. One ml DEAE Capto columns were purchased from GE Healthcare. The columns and the system were equilibrated with buffer A. A total volume of 2.5 ml *in vitro* transcription mix was loaded onto the column via a 5 ml loop at 1 ml/min. The loop and the column were then washed with 20 ml buffer A (Fig. 2.23A) and three elution gradients were used. The first linear gradient eluted all free nucleotides (Fig. 2.23B). The second linear gradient separated short abortive transcripts, the correct transcript from longer transcripts at 2 ml/min (Fig. 2.23C). The final gradient of 5 cv at 2 ml/min at 100 % B strips the column from all bound molecules (Fig. 2.23D). The column was equilibrated with buffer A for the next injection until pressure and conductivity remained stable. 1.5 ml fractions were collected during the second gradient and fractions were analysed on a 12 % (w/v) PAA/8M Urea gel. RNA containing fractions were pooled and rinsed once with buffer SEC tRNA on Amicon Ultracentrifugal Filters with a 10K MWCO to a final volume of 400-500 μl.

Buffer A: 100 mM NaCl, 50 mM Bis-Tris pH 6.5, 0.2 mM EDTA
Buffer B: 2 M NaCl, 50 mM Bis-Tris pH 6.5, 0.2 mM EDTA

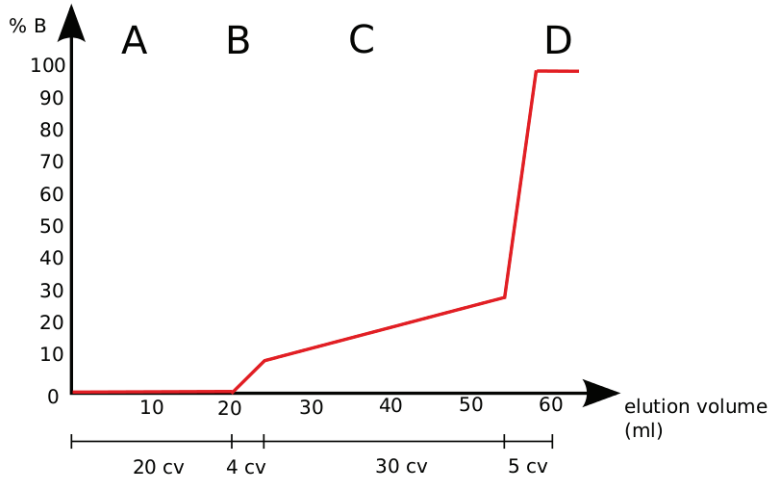


Figure 2.23: RNA elution profile for anion exchange chromatography.Step A) loading of the transcription and binding of tRNA, B) gradient from 0 - 10 % (v/v) B to elute free nucleotides, C) gradient from 10 - 30 % (v/v) B to sequentially elute abortive transcripts from full length transcripts, D) stripping of all strong binders with a high concentration of salt, 100 % (v/v) B.

Size exclusion chromatography A second purification step was necessary to remove aggregates on a Superdex 75, 10/300 column (GE Healthcare). The column was equilibrated in RNA SEC buffer at 0.5 ml/min. A 1 ml loop was loaded with a maximum of 500 μl per run and the loop was washed with 3 ml. Over one cv the RNA was eluted at 0.5 ml/min. Fractions of 0.3 ml were collected and tested for RNA content on a 12 % (w/v) PAA/8M Urea gel.

RNA SEC buffer: 50 mM Bis-Tris pH 6.5, 250 mM NaCl

2.2.3 Protein quality control

2.2.3.1 Dynamic light scattering

Homogeneity and aggregation of purified PRORP proteins were assessed with dynamic light scattering (DLS).

In DLS experiments, monochromatic visible laser light (λ= 630 nm) is scattered by particles in the sample solution. The smaller the particles, the faster they move in solution due to the brownian motion. The translational diffusion coefficient can be determined from the autocorrelation of the scattered light intensity in time (Fig. 2.24). Hence, the hydrodynamic radius of the particles can be derived using the Stokes-Einstein equation:

$$R_H = \frac{k_B T}{6\pi\eta D_T} \tag{2.1}$$

where R_H is the hydrodynamic radius, k_B the Boltzmann constant, η the absolute viscosity of the solvent and D_T the translational diffusion coefficient.
A single population of identical particles gives a monomodal size distribution. The sharpness of the intensities is indicative of the monodispersity of the sample. If all particles are identical the peak will be sharp and the sample is monodisperse.

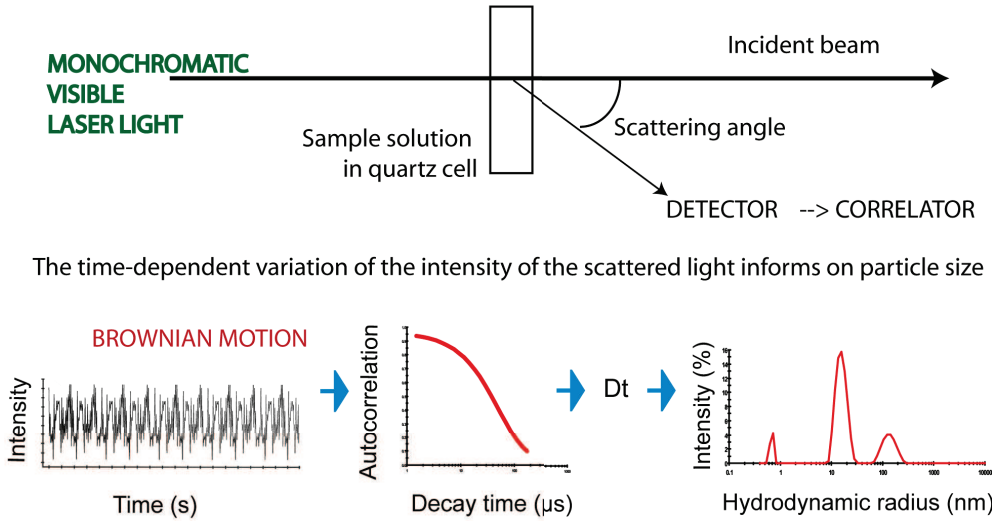


Figure 2.24: Principle of DLS experiments Lorber *et al.* (2012).

Two instruments were used for DLS measurements:
Zetasizer^{TM} NanoS (Malvern Instruments) with a 4 mW He-Ne laser (λ = 633 nm) and detection at θ= 173°
Dynapro NanoStar^{TM} (Wyatt Technologies) with a 100 mW He-Ne laser (λ =633 nm) and detection at θ=90°

All measurements were carried out at 20 °C. 20 μl of protein solution at $\geq$ 1 mg/ml were placed in a quartz cuvette using the Zetasizer NanoSeries Nano-S and 4 μl at lower concentrations using the Nanostar instrument. To check the effective removal of particles larger than 100 nm two measurements were carried out: before ultracentrifugation (c = 0.6 mg/ml) and after an ultracentrifugation of 1 h at 125,000 g (Sorvall Hitachi Discovery M150SE micro-ulracentrifuge). For precise diameter calculations the buffer viscosity, the refractive indices as well as the solvent density were determined using an AMV viscosimeter, an Abbe refractometer and a 5 ml glass pycnometer, respectively (Tab 2.6).
Since the scattered light intensity is measured at only one angle, a conventional DLS experiment can never lead to the exact particle mass when measurements are performed at one concentration except for spheric particles. An alternative to determine the size of a protein is multi-angle light scattering (MALS). MALS experiments are done in-line with a SEC. Three detectors at different angles collect the scattered light intensities simultaneously leading to the determination of the precise mass of the purified proteins (Lorber *et al.*, 2012). The MALS experiments were carried out with Dr. Isabelle Billas at the Institute of Genetics and Molecular and Cellular Biology in Illkirch, France.

Table 2.6: Summary of used buffers for DLS measurements with their dynamic (absolute) viscosity and refractive indices.

Name	Refractive index	Dynamic viscosity ($mPa * s$)
Water	1.333	
SEC1	1.355	1.724
SEC1_2	1.3502	1.286
SEC2	1.3502	1.286

2.2.3.2 Activity assay of PRORP proteins

PRORP activity was assayed using a L5T0 *Arabidopsis thaliana* mitochondrial $tRNA^{Cys}$, i.e. a pre-tRNA with a 5 nt long leader and no trailer sequence. Initial activity assay conditions were those described by Rossmanith *et al.* (1995).
Protein samples were diluted to 0.25 mg/ml in SEC2 buffer and the pre-tRNA to 0.2 mg/ml. In one assay 1 μl of protein, 1 μl tRNA and 8 μl of SEC2 buffer were incubated 20-30 min at room temperature. The reaction was stopped by adding 5 μl of 6 M guanidine-HCl and 35 μl of water. tRNAs were extracted with 50 μl of a saturated phenol/chloroform solution (50:50, pH 6.8). The aqueous phase was transferred into a new tube and tRNAs were precipitated with 0.5 μl of 20 mg/ml glycogen (ThermoFischer Scientific, No. R0551), 5 μl of 3 M sodium acetate pH 5 and 125 μl of absolute ethanol at -20 °C for 1 h. The precipitated RNA was centrifuged (25 min, 21.000 g) and the pellet was dried and dissolved in 10 μl RNA loading dye. The samples were analysed on 12 % denaturing PAA gel containing 8 M urea and 1X TBE buffer. Nucleic acids were stained with ethidium bromide for 5 min.

2.2.4 Inductively coupled plasma mass spectrometry

Inductively coupled plasma mass spectrometry (ICP-MS) is a method to detect and quantify ions in liquid or solid samples, zinc ions for instance in PRORPs.

ICP-MS instruments consist of two distinct parts: One generating the positively charged ions and the other separating them according to their mass-to-charge ratio (Fig. 2.25). In more detail that is a 1) nebulizer/spray chamber, generating an aerosol from the liquid sample and sorting out droplets of suitable size, 2) an ICP torch/HF generator ionizing the plasma argon gas and the passing sample atoms and 3) a mass spectrometer analyzing the ions.
In the ICP torch filled with argon gas a plasma is formed in an inductive field with temperatures of up to 10,000 K. The strong electromagnetic field will ionize the argon atoms leading to some positively charged argon ions and free electrons that will collide and further ionize other argon atoms. The vaporized sample passing the chamber will be dehydrated and broken into single

atoms and further ionized. The mass spectrometer separates the ions according to their mass-to-charge ratio. Using a calibration standard the exact quantity of the atom of interest can be determined (Thomas, 2004).

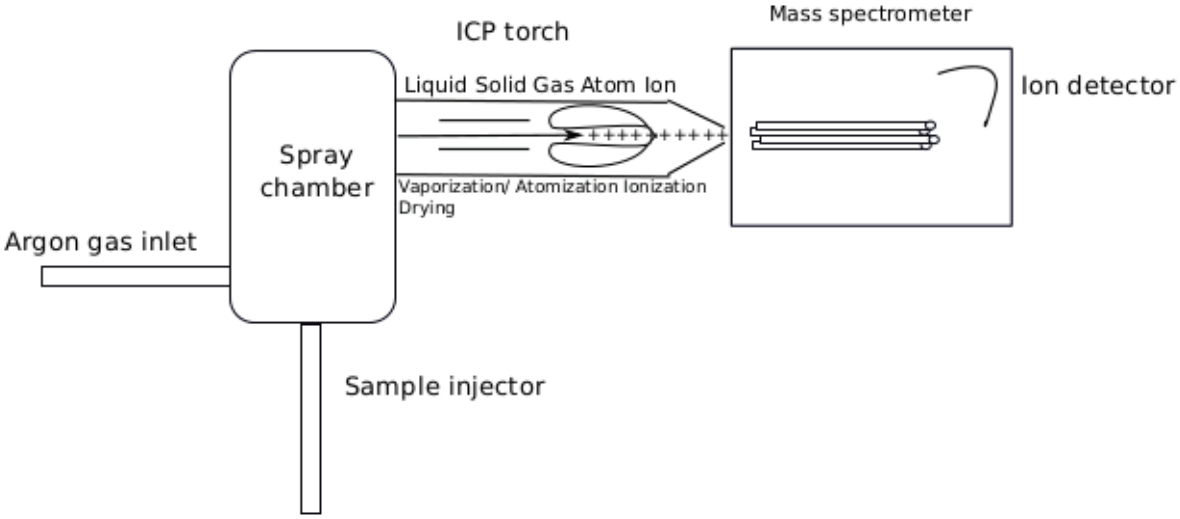


Figure 2.25: Principle of ICP-MS analyses (reproduced after Thomas (2004)). The liquid sample is injected into the spray chamber and mixed with argon gas. In the ICP torch the sample is solidified, decomposed into atoms and further ionized and eventually analyzed with a mass spectrometer. Quantification of ions can be made in comparison to a calibration curve.

Experiments were performed with René Boutin in the Laboratoire d'Hydrologie et de Geochimie, Strasbourg, France. Wild type PRORP2 (14 μg/ml), wild type PRORP1 and various mutants (28 μg/ml) in SEC buffer/Nitric acid were analyzed.

2.2.5 Methods to determine affinity parameters of the PRORP/tRNA interaction

To study protein/RNA interactions I wanted to be sure of the binding parameters of the two partners. There are numerous of techniques to determine affinity constants each one possessing advantages and drawbacks. In the following I will illustrate the methods I used to determine the K_D of PRORP2mDD/L5T0 $tRNA^{Cys}$ interaction.

2.2.5.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a label-free method to determine thermodynamic parameters of substrate ligand binding such as affinity constants or binding stoichiometry.

The change of heat is an universal characteristic of chemical or physical change that happens in most chemical reactions. ITC is a non-destructive, label-free method that detects heat changes over time at constant temperature. It allows a rapid and direct measurement of thermodynamic

parameters and binding properties, as well as a determination of the active concentration of a molecule in solution (Ghai *et al.*, 2012). In a typical ITC experiment (Fig. 2.26a) a series of highly concentrated substrate or protein solution (c $\sim$ 300 μM) is injected (20 injections of 2 μl) into a 300 μl injection vessel containing the interaction partner at $\sim$ 20-30 μM. Injections take place every 2 min to let the system come back to baseline. During the course of injection heat changes are detected under the form of differential power that is needed to keep the temperature constant with the reference cell. Thus, the raw data is a plot of power change versus time. The integration of each peak results in the plot of a binding isotherm of enthalpy versus the molar ratio of ligand-to-substrate (O'Neill & Gaisford, 2011). Using this isotherm, one can calculate the enthalpy of binding (Δ_bH), the equilibrium binding constant (K_b) and the reaction stoichiometry (n) (Fig. 2.26b).

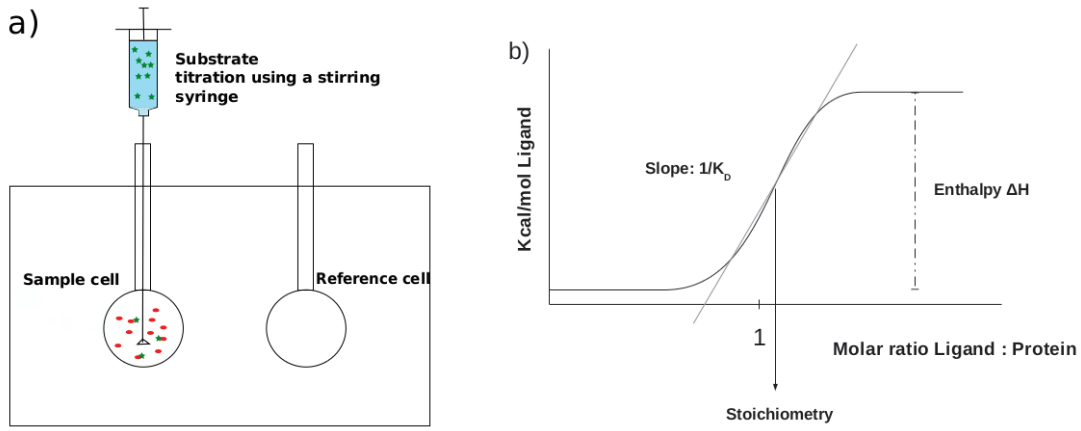


Figure 2.26: ITC setup and data output. a) Scheme of an ITC instrument (reproduced from Zhou *et al.* (2011)). A reference cell and a sample cell are situated in a thermally isolated container. During titration of the ligand and mixing of the sample solution heat is produced or absorbed leading to a temperature change with respect to the reference cell. This signal can be used to calculate thermodynamic parameters of the studied system, b) A scheme of a binding isotherm. The slope of the tangent through the inflection point gives the affinity constant, the inflection point the binding stoichiometry between the ligand and the protein and the difference between the lower and upper plateau allows the calculation of the binding enthalpy.

At first a blank measurement was performed with the temperature set to 20 °C. The syringe was loaded with PRORP2mDD solution at 187 μM in SEC2 buffer with 5 mM $MgCl_2$ and 20 times 2 μl were injected into SEC2 buffer with 5 mM $MgCl_2$ every 120 seconds. Two identical experiments were done at 20 °C and 8 °C. The cell contained 280 μl of 22 μM L5T0 $tRNA^{cys}$ in SEC2 buffer with 5 mM $MgCl_2$. Data analysis was carried out by Eric Ennifar using the microcal Origin7 software.

2.2.5.2 Microscale thermophoresis

Microscale thermophoresis (MST) is a method to determine affinity parameters and diffusion constants using the phenomenon that molecules move in temperature gradients.

Thermophoresis is the characteristic of particles moving in temperature gradients, usually from warm to cold regions. MST instruments make use of this phenomenon and consist of several parts: 1) the temperature-controlled tray for 16 capillaries and 2) the optical system with an infrared (IR) laser that is coupled to the path of the fluorescence laser and focused to the same point in the sample. Sample solutions are transferred inside the capillary by capillary forces. There are two instrument setups available: 1) Excitation (λ = 280 nm) and emission (λ = 360 nm) of fluorescent amino acids (tryptophane, tyrosine, phenylalanine) in the Monolith NT.LabelFree instrument and 2) with three different types of LED combinations in the Monolith NT.115 instrument requiring one of the partners to be labelled with a fluorophore. In a titration experiment the concentration of the fluorophore is kept constant and the concentration of the non-labelled partner varies. The highest concentration of the non-labelled partner should be at least 20 fold of the expected K_D. For each sample the initial fluorescence is recorded (Fig. 2.27A) and variation between the capillaries should not exceed 10 % unless the fluorophore is masked by the interaction with its partner. When the IR laser is switched on the radiation is absorbed by water molecules in the sample leading to a local heating and the formation of a temperature gradient. At the same time the temperature change induces a change in fluorescence of the dye which is an inherent property of each fluorophore (Fig. 2.27B). The second, slower event is thermodiffusion (Fig. 2.27C) of the molecules in the temperature gradient that is maintained by the IR laser. This diffusion occurs as far as a steady state is reached which means that thermodiffusion equals mass diffusion. Once the IR laser is switched off an inverse T-jump in fluorescence signal is observed (Fig. 2.27D), followed by the back diffusion of the molecules with the fluorescent signal coming back to near baseline levels (Fig. 2.27E).

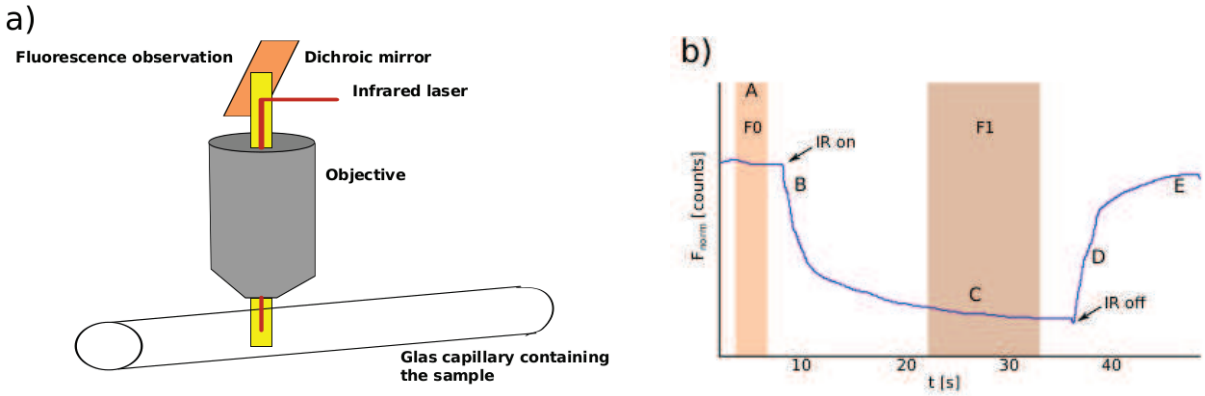


Figure 2.27: Microscale thermophoresis. a) A scheme of the parts of an MST instrument. The fluorescence and the infrared laser are focused to the same point in the sample solution, b) illustrates the change of fluorescence during different steps of MST measurement: A) initial fluorescence (F0), B) IR laser ON with immediate change of fluorescence signal, C) thermophoretic movement of molecules creating a slight decrease of the fluorescence signal, D) IR laser turned OFF, fluorescence coming back to initial values (reproduced after Seidel *et al.* (2013)).

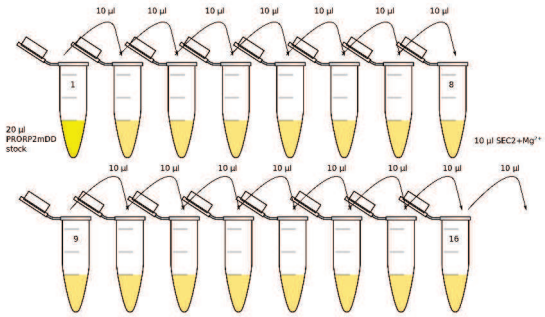


Figure 2.28: Dilution scheme for MST titration points

A Cy5 labelled L5T0 tRNA^{Cys} at the 5' extremity was ordered from IBA GmBH (Göttingen, Germany) and solubilized in water. Cy5 is excited at a wavelength of 640 nm and emits light of a wavelength of 670 nm.
MST protocol:

1. The L5T0 tRNA^{Cys} stock solution (158 μM) was diluted to 2 μM in SEC2 buffer with 5 mM $MgCl_2$.

2. To test the fluorescence signal and capillary effects two standard and two hydrophilic capillaries were filled with a 50 nM Cy5-L5T0-tRNA^{Cys} solution and the laser LED set to 20 % and the IR laser to 40 %. 100 % IR laser power will increase the local temperature of the sample of about 7 °C.

3. Serial dilution of PRORP2mDD: Initial PRORP2mDD concentration 188 μM in SEC2 with 5 mM $MgCl_2$. Dilution scheme in Fig. 2.28.

4. Using the same tip 10 μl of a 100 nM Cy5-L5T0-tRNA^{Cys} solution were added into each PCR tube, beginning with tube 16.

5. Filling of the capillaries and placing them into the capillary holder.

6. MST measurement.

Fluorescence variations due to the dilution of the non-labelled binding partner can occur. This might be due to quenching or enhancement of the fluorescence signal upon binding, an unspecific adsorption of one of the partners to the glass capillary or aggregation of the protein during binding to the labelled tRNA. To answer this question a SDS-denaturation assay was carried out as described in the FAQ protocol of Nanotemper, Fig. 2.29. Experiments showed that Monolith NT.115^{TM} hydrophilic capillaries (NanoTemper Technologies GmbH, Munich, Germany) and a supplement of 0.1 % Tween20 improved signal stability and data quality. Using hydrophilic capillaries a complete dataset was measured (Fig. 6.10).

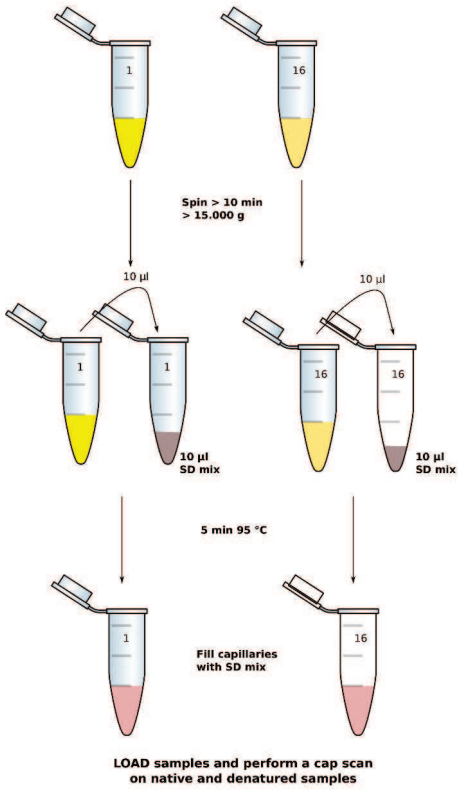


Figure 2.29: Protocol to test if fluorescence changes are due to protein adsorption on the glass capillary (adapted from FAQ Nanotemper). SD-mix: 4 % SDS, 40 mM DTT

2.2.5.3 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is used to determine the molecular mass, oligomeric state as well as affinity parameters of interacting macromolecules by measuring the sedimentation coefficients.

Two types of data can be derived from an AUC experiment using two experimental setups:

1. Sedimentation velocity (SV) experiments give information about hydrodynamic parameters of a molecule such as its size and shape. The rate at which boundaries of molecules move in a gravitational field is recorded over time.

2. Sedimentation equilibrium experiments provide thermodynamic parameters such affinity constants and stoichiometry of binding. They monitor the concentration distribution at the equilibrium between sedimentation and free diffusion (Cole *et al.*, 2008).

An advantage over other techniques is that there are almost no biological constraints on molecular weight and size of the sample, almost all buffers can be used and molecules can be studied under native conditions. Electric neutrality is the only requirement for the particles to sediment in a gravitational field. The combination of three forces gives the Svedberg constant.

1. The force on the sedimenting particle:

$$M_p = \omega^2 \times r \tag{2.2}$$

where M_p is the mass of the particle in gram per mole, ω the rotor speed in radians per second and r the distance of the molecule to the rotor center in centimetres.

2. The counterforce on the particle exerted by the displaced solvent:

$$M_s = \omega^2 \times r \tag{2.3}$$

where M_s is the mass of the solute in gram per mole.

3. The frictional force:

$$F = f \times \nu \tag{2.4}$$

with f being the frictional coefficient in kilogram per second and ν the sedimentation velocity in metre per second.

The rearrangement of all these equations leads to the following relationship:

$$s \equiv \frac{\nu}{\omega^2 \times r} = \frac{M_b}{f} \tag{2.5}$$

where s is the Svedberg constant in time unit and M_b is the effective (or buoyant) mass of the particle in gram per mole. 1 Svedberg is 10^{-13} s.
A second parameter accessible with AUC is the translational diffusion coefficient, D, by recording the motion and the shape of the moving concentration boundaries. Using approximate solutions to the Lamm equation, s and D are available in an AUC experiment. The Svedberg equation puts the ratio of s and D in relationship being proportional to the buoyant particle mass:

$$\frac{s}{D} = \frac{M_b}{RT} \tag{2.6}$$

where R is the universal gas constant in Joule per mole and Kelvin and T the absolute temperature in Kelvin.

Velocity experiments were done in two channel cells of sector shaped compartments (Cole *et al.*, 2008). 450 μl of different ratios of PRORP2mDD/L5T0 tRNA^{Cys} were prepared in SEC2 buffer with 5 mM $MgCl_2$. L5T0 tRNA^{Cys} concentration was kept constant at 1 μM and PRORP2mDD was added to different final concentrations: 0.5 μM, 1 μM, 5 μM and 10 μM, respecting the limit of 1.2 absorption units measured with the UV spectrophotometer at 260 nm. As a reference 1 μM L5T0 tRNA$^{Cys)}$ and 17 μM PRORP2mDD were measured. 400 μl of sample solution were loaded in the sample cell and 404 μl of buffer were placed in the reference cell. Temperature was set to 20 °C. Extinction coefficients according to protparam were used for PRORP2mDD ε_{280nm} = 89840 $M^{-1}cm^{-1}$ with M = 60350 Da (Gasteiger *et al.*, 2005) and L5T0 tRNA^{Cys} ε_{260nm} = 613000 $M^{-1}cm^{-1}$ (according to nanodrop measurements).

2.2.6 Analyzing PRORP/pre-tRNA interactions

In addition to the determination of the affinity constant I wanted to have a visible proof of a formed complex in solution that is stable enough to be analyzed with SAXS and that can be eventually crystallized.

2.2.6.1 Size exclusion chromatography as a tool for studying PRORP-tRNA interactions

For analytical size exclusion experiments an Agilent Bio SEC-3 was used on an Agilent high-performance liquid chromatography system. The setup was the same as used for SAXS experiments at the SWING beamline at synchrotron SOLEIL, France. The flow was set to 0.2 ml/min. The column was calibrated with 15 μl of a BioRad gel filtration standard (no. 151-1901) then the protein was analyzed alone, followed by the RNA alone and mixtures of the two in different molar ratios. About 1.5 - 2 μg of pre-tRNA and 50 μg of PRORP were sufficient for good absorption signals at 260 nm and 280 nm, respectively. Elution times of each species were monitored. Shorter elution times were expected for the complex. Different substrates, wild type PRORP2 and its catalytic mutant as well as different buffer conditions were compared (Tab. 6.2).

2.2.6.2 Electromobility shift assay

For an electromobility shift assay (EMSA) 200 ng of L5T0 tRNA^{Cys} were incubated with increasing amounts of wild type PRORP2 in SEC2 buffer for 30 min at 25 °C. Molecules were separated on a 6 % PAA native gel in Hepes-KCl buffer at 4 °C and 5 V/cm-gel height. Gels were stained with ethidium bromide.

10 X Hepes-KCl: 500 mM Hepes pH 7.6, 500 mM KCl

2.2.6.3 Complex modeling

The X-ray crystallographic structure of PRORP1 (Howard *et al.* (2012); pdb ID: 4G23) was published in 2012 at the same time we published the first model of PRORP2 complexed to a precursor tRNA based on SAXS and biomolecular data. We reused this high resolution data to refine our model. The coordinates of PRORP1 served as a template and were loaded onto the ElNemo webserver that is a tool to predict possible conformational states of macromolecules (Suhre & Sanejouand, 2004). Suitable models were fitted to a tRNA model taking into account contacts revealed during foot print analysis using pymol (Schrödinger, LLC, 2010).

2.2.6.4 Crosslinking

UV crosslinking was used to obtain stable, covalently bound PRORP/tRNA complexes. A first approach, applied during foot printing experiments, used irradiation at 254 nm (Gobert *et al.*, 2013). This produces inter- or intramolecular tRNA/tRNA or tRNA/protein adducts if the molecules are in close proximity. A drawback is that UV irradiation can break RNA strands and may induce protein

degradation. A second approach makes use of ribonucleotides containing a photoreactive group. The most common nucleotides are thio-substituted uridines ($_4$SU) or guanosines ($_6$SG) (Fig. 2.30). If they are close to an amino acid side chain like lysine, then the sulfur radical formed upon UV irradiation at 365 nm can trigger a nucleophilic attack to make a covalent bond (Harris & Christian, 2009). Using modified RNAs requires longer wavelengths that are less harmful to biomolecules.
Two modified pre-tRNA substrates were purchased from IBA, Göttingen, Germany. The first L5T0 tRNA^{Cys} construct (no. 1) with thio-G and thio-U at positions indicated in Fig. 2.30 and the second construct (no. 2) with identical modified bases plus a phosphothioate linkage between position -1 U and +1 G.

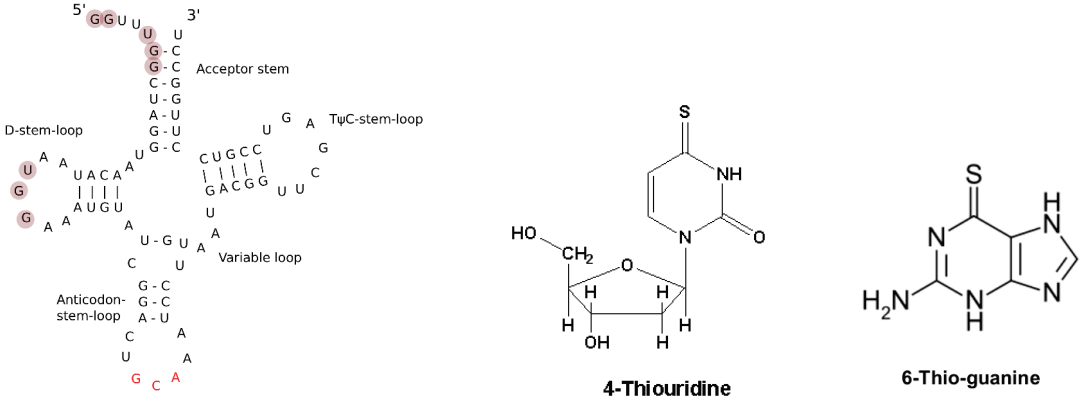


Figure 2.30: Modified tRNA for crosslinking experiments. The coloured circles indicate modified nucleotides whose structures are depicted on the right.

To test the effect on tRNA and protein of various radiation doses 200 ng of L5T0 tRNA in 10 μl SEC2 buffer with or without 5 mM $MgCl_2$ were irradiated with doses ranging from 40 to 640 mJ/cm^2 at 254 nm and 365 nm. The integrity of the tRNA was verified on a 12 % (w/v) PAA/8M Urea gel. The same experiment was repeated with the 1 μg PRORP2 alone and the results were evaluated on a 7.5 % (w/v) TGX gel (Biorad).
For cross-linking experiments 1.3 μM tRNAs no. 1 and increasing amounts of PRORP2mDD in SEC2 buffer with 5 mM $MgCl_2$ in 25 μl reaction mix were incubated 30 min at room temperature ($\sim$ 25 °C) and subsequently irradiated at 365 nm with 640 mJ/cm^2. Ten μl of each reaction were withdrawn and a 7.5 % (w/v) TGX gel was run to detect complexes on protein level and a 12 % (w/v) PAA/8M Urea gel or a native gel for RNA detection.

2.2.7 Synchrotron radiation circular dichroism

Synchrotron radiation circular dichroism (SRCD) is used to determine the secondary structure content of proteins as α-helices and β-sheets absorb polarized light differently.

SRCD spectroscopy is a method that is based on the fact that chiral molecules like proteins absorb left and right polarized light differently. A synchrotron radiation source has the advantage of an increased photon flux resulting in higher signal-to-noise ratios, less material is needed, faster and

more accurate measurements can be made. The exploitable wavelength at synchrotron sources can be as low as 168 nm that result in higher information data compared to conventional CD instruments where a xenon light source is used. The resulting spectrum is the sum of the weighted secondary structure elements present in the protein. α-helices show two distinct negative peaks at 222 and 208 nm and a positive peak at 190 nm. β-sheets have one single negative peak at 212 to 215 nm and a positive peak at 190 to 195 nm. CD spectra of proteins containing helices and sheets are less accurate on structure information as the curves are dominated by the helical components. Still, SRCD can improved the gain of information by collecting useful data in the far UV region ($< 190nm$) that is below the limits of laboratory instruments equipped with deuterium lamps. SRCD can provide structural information on the secondary content of proteins, and be applied to study environmental effects such as temperature, salt and pH conditions. It can also monitor effects of single mutations that have consequences for the secondary structure (Wallace & Janes, 2010).

SRCD experiments were performed on the DISCO beamline at synchrotron SOLEIL, France. The instrument was calibrated for magnitude and polarization with a 6.1 mg/ml D-10-camphorsulfonic acid solution. PRORP proteins (10 mg/ml) in 100 mM potassium phosphate, 50 mM KCl, 10 % (w/v) glycerol and 1 mM TCEP were placed in a SRCD CaF_2 cuvette of 8 μm pathlength. Three spectra from 170 to 280 nm were measured at 10, 20, 30, 40, 50, 60 and 70 °C to assess the thermal stability of wild type PRORP1-2. The data were processed (spectrum averaging, solvent base line subtraction) using CDtools (Lees *et al.*, 2004). The secondary structure content of PRORPs was evaluated using the VARSLC method (Manavalan & Johnson, 1987) in DICHROWEB (Whitmore & Wallace, 2008).

2.2.8 Small angle x-ray scattering

Small angle X-ray scattering (SAXS) is an in solution technique used to determine biophysical parameters of biomolecules such as the radius of gyration, the maximal diameter, the specific volume but also their global shape and domain organization.

SAXS, like X-ray crystallography, uses the phenomenon that photons are scattered by valence electrons in macromolecules. These scattered photons can be detected on 2D detectors and intensity is measured as a function of the scattering angle, θ (Fig. 2.31). As in solution particles are freely moving, the orientation of the molecules is lost in SAXS experiments and the deduced information is the scattered intensity versus the momentum transfer:

$$s = \frac{4\pi \times sin\Theta}{\lambda} \tag{2.7}$$

where s is the scattering vector, θ is the half angle between the incident and the scattered beam and λ the wavelength in nm.

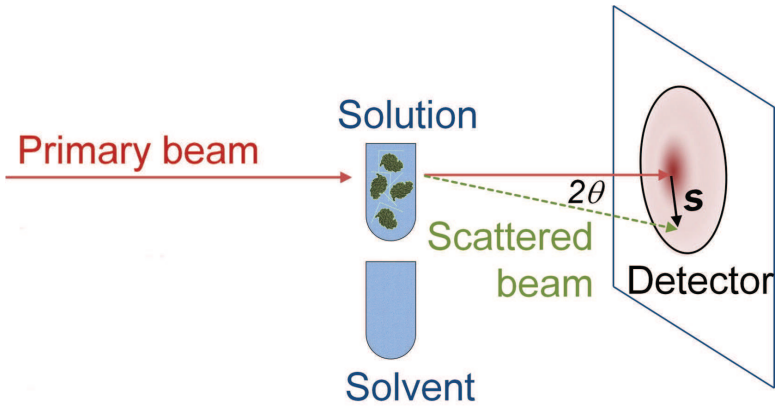


Figure 2.31: SAXS experimentation setup. A synchrotron X-ray beam illuminates the sample and a small fraction of light is deviated by the electrons of the molecules in solution by an angle θ. The scattered intensity is recorded. From this experiment the intensities can be plotted against the scattering vector, s (Petoukhov & Svergun, 2013).

The difference of intensity between the sample and the buffer is proportional to the concentration of the particles and the squared contrast (Fig. 2.32a). A difficulty in measuring biological samples with SAXS is that they contain almost no electron-rich atoms leading to a poor contrast. There are several parameters that can be determined by SAXS. The momentum transfer and the forward scattering intensity, I(0), can be directly retrieved by the SAXS data and the radius of gyration (R_g) is available through the Guinier approximation:

$$I(s) = I(0) \times exp(-\frac{1}{3}R_g^2 \times s^2) \tag{2.8}$$

but only at small angles where $s \times R_g < 1.3$. In practice R_g and I(0) are accessible through the Guinier plot which is a plot of ln[I(s)] versus s^2. Its slope is the R_g and its intercept with the y-axis gives I(0). The Guinier plot is an indicator of good sample quality but a good Guinier plot cannot exclude the presence of aggregates in solution. Complementary methods like DLS are needed to verify sample quality. Another value that can be determined, yet with not such accuracy, is the molecular weight of a particle: I(0) is proportional to the molecular weight of the particle and its accuracy is usually sufficient to determine the oligomeric state of particles in solution.
The second method to determine the radius of gyration and I(0) is the distance distribution function, P(r), which also contains information on the intramolecular atomic distances. It gives information about the global organization of the molecule, e.g. if it is a perfect sphere or made of two domains (Fig. 2.32b). Another information that can be obtained, yet flawed due to the low resolution data, is the reconstruction of *ab initio* 3D shapes of the solute particles. This became possible with the introduction of automated bead modeling a method implemented in several programs of the ATSAS suite (Fig. 2.33) (Petoukhov *et al.*, 2012).

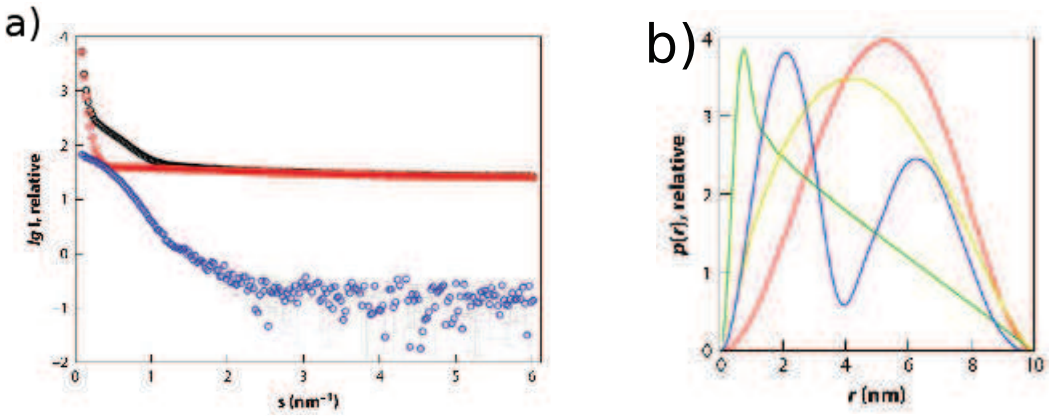


Figure 2.32: SAXS data analysis. a) Relative intensities of the scattered light plotted against the scattering vector, s. SAXS data from the buffer (red), the sample (black) and the resulting curve after subtraction of the background scattering from the sample scattering (blue), b) the distribution function of intra-atomic distances: perfect sphere (red), long rod (green), flat disk (yellow), hollow sphere (blue), dumbbell (pink) (Svergun & Koch, 2003).

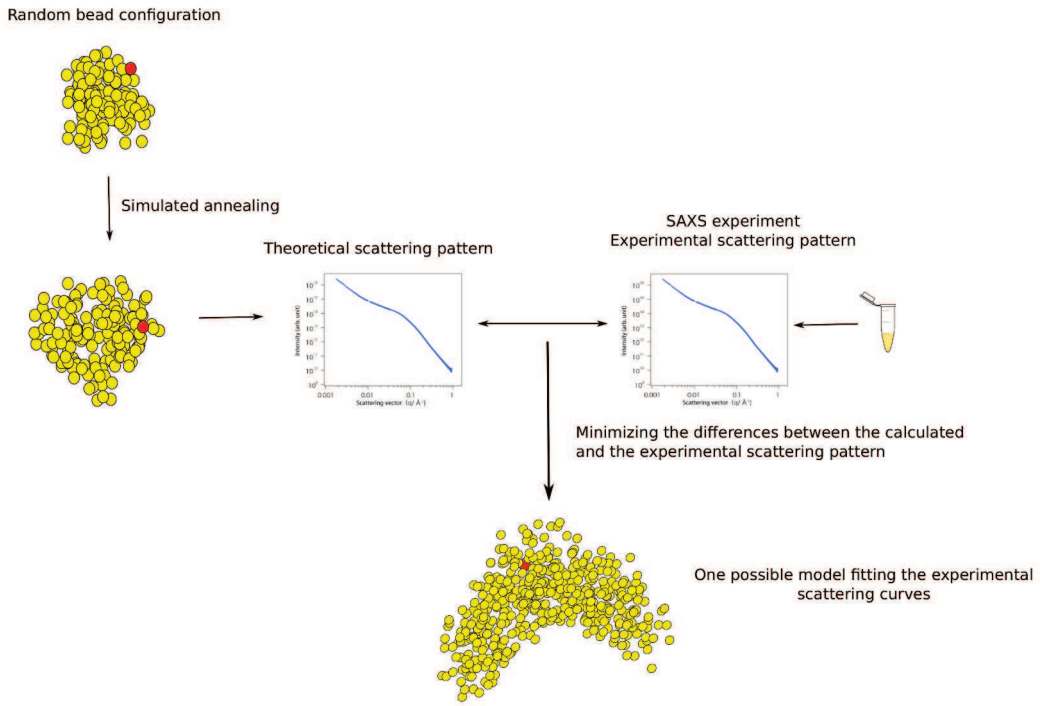


Figure 2.33: *Ab initio* modeling process. During the process beads that represent atoms are moved upon theoretical heating (simulated annealing) and cooled down. From this new bead model a theoretical scattering curve is calculated. This process is repeated such that the differences between the experimental and the calculated scattering curve become minimal.

In a typical SAXS experiment at SOLEIL synchrotron the SAXS capillary is downstream of an analytical gel filtration column that will separate the macromolecules depending on their size. Buffer data is collected in the exclusion volume of the column which is later used to determine the background

signal to be subtracted from the sample signal. A first injection is used to determine the volume where the protein or RNA elutes, followed by the injection of the full sample and collection of the SAXS data in the corresponding volume. The higher the sample concentration the better. 60 μl of sample (10 mg/ml) are injected. At the SWING beamline for biological SAXS two columns can be connected which reduces time loss between experiments where for example different buffers or different column types are needed. A second mode is direct injection of the sample which dramatically shortens the time for one experiment but data quality can be poor if aggregates are present in the sample.

2.2.9 Macromolecule crystallization

Crystallization is the process of the formation of an ordered phase of molecules in solution.

A crystal is an arrangement of molecules in the highest possible order. A macromolecular crystal is made of unit cells that are made of asymmetric units. An asymmetric unit contains all components that, by applying all kind of allowed symmetry operations, makes up a unit cell. The unit cell makes up the whole crystal by simple translational operations (Rupp, 2009).
Requirements for protein samples are purity (at least 95 % pure on SDS-PAGE) and conformational homogeneity which can be determined with DLS. Crystallization variations can also occur within different purification batches and they can (but not necessarily) be influenced by protein tags and surface modifications or amino acid mutations. There are different methods to grow crystals. The most common methods are vapor diffusion, batch and counter diffusion (Fig. 2.34).

Vapor diffusion For screening crystal growth conditions the most convenient method is vapor diffusion using sitting nanodrops (Fig. 2.34a). Its advantage is the possibility of automation and using robotics. Two liquid compartments (the drop and the reservoir) are placed separately in a sealed vapor chamber. The nanodrop contains equal volumes of the protein and the reservoir solution creating a condition where the solution is ideally undersaturated. A solution is undersaturated when the concentration of the solute is below the concentration that can be solubilized. The bottom of the chamber is filled with the crystallant solution ($\geq$ 100 fold of drop volume). As crystallant concentration in the drop is half as high as in the reservoir vapor diffuses from the drop to the reservoir equalizing the concentration in the the two compartments. This leads to an increase of protein and crystallant concentration in the drop. Ideally, during this process the protein concentration will reach supersaturation and start to nucleate (Fig. 2.35A). This results in either nucleation or precipitation. When a nucleus reaches a critical size, it grows forming a crystal until the concentration of the surrounding liquid drops below the solubility curve. The system is then back to dynamic equilibrium. The supersaturation zone is a zone where the concentration of solutes is superior to what can be solubilized under normal circumstances. It can be subdivided into the metastable, the nucleation and the precipitation zone. In the metastable zone crystals can grow but nucleation will not occur. In the nucleation zone the energy to form crystals is high enough and nucleation occurs. In the precipitation zone the solution is so saturated that

molecules precipitate (formation of non-ordered solids).

Batch methods In a batch experiment protein and crystallant solutions are mixed at a high concentration under water-(im)permeable oil layer (Fig. 2.34b). The protein will only crystallize if the initial conditions bring it into the nucleation zone because after mixing the system varies only slightly (Fig. 2.35B).

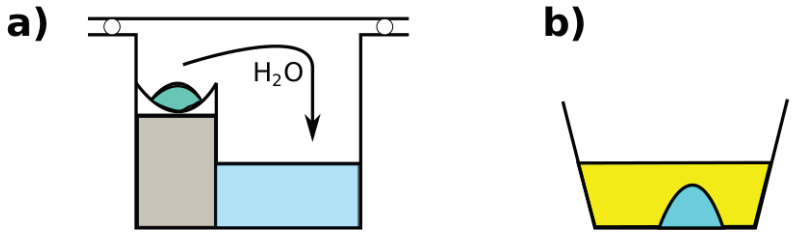


Figure 2.34: Crystallization methods. a) Vapor diffusion in sitting drops. Vapor diffuses from the drop to the reservoir, increasing the protein concentration in the drop, leading to crystallization, b) batch setup where mixing protein an crystallant solution immediately lead to protein supersaturation.

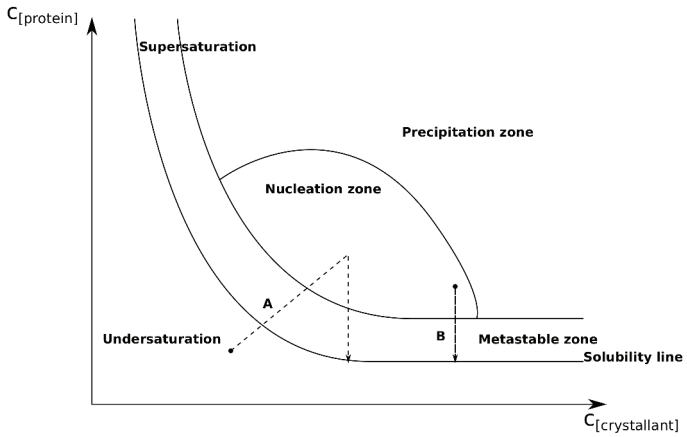


Figure 2.35: Crystallization diagram. The crystallant concentration is plotted against the protein concentration. (A) Vapor diffusion: protein and crystallant mixing leads to an undersaturated solution. Nucleation occurs upon vapor diffusion and increasing protein concentration in the shrinking drop, (B) Batch method: mixing the protein and crystallant solution leads to immediate supersaturation and protein concentration may decrease due to nucleation and crystal formation.

Counter diffusion In a counter diffusion experiment a long thin capillary containing the protein solution is brought in direct contact with the crystallant solution that has to be highly concentrated. Due to the different diffusion constants and a lack of convection the crystallant can diffuse freely into the capillary creating a gradient. The protein concentration meanwhile stays quasi constant as the protein cannot diffuse because it will either precipitate at high crystallant concentration (Fig. 2.36A) or crystallize (Fig. 2.36B, C). Counter diffusion experiments allow screening infinite

concentrations of crystallant components depending on the length of the capillary by creating a supersaturation wave along the length of the capillary (García-Ruiz, 2003, Biertümpfel *et al.*, 2002)

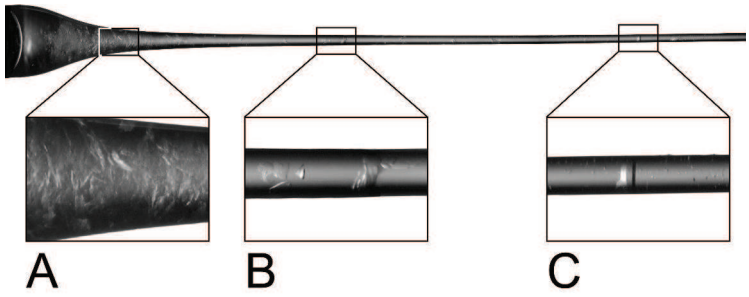


Figure 2.36: Counter diffusion experiment. Crystallant solution can diffuse freely through a capillary filled with protein solution. (A) protein precipitation at high crystallant concentration, (B) nucleation and small crystal formation, (C) growth of a single crystal (Biertümpfel *et al.*, 2002).

2.2.9.1 Initial screening

The conditions for formation of nuclei are not predictable and are searched by screening a maximum of possible solutions combining various chemicals. Screening strategies fall into different categories:

1. Sparce-matrix screens containing a collection of crystallant compositions combined such that a maximum of different conditions are covered,

2. Incomplete factorial screens as a less biased version of a sparce screen where several ingredients are varied in a statistical weighted manner depending on the searched parameters and the experiments the user wants to perform (Carter & Carter, 1979) ,

3. Grid screens where one or some parameters are changed in a systematic manner which is the strategy of choice to optimize crystallization conditions.

Initial screens were carried out using a Mosquito pipetting robot, TPP LabTech and 96-well crystallization plates (CrystalEx microplate, conical flat bottom, 5 sub-wells). Usually 150 nl of protein solution were mixed with 150 nl of crystallant solution. The drop was equilibrated against 35 μl crystallant solution at 4 °C and 20 °C. At least two different protein concentrations and the buffer alone were tested.

2.2.9.2 Crystal optimization

Initial hits, that can be microcystalline or single crystals were first assayed to reproduce under the same conditions or using a different crystallization method such as batch experiments. To do so usually 1 μl of protein and 1 μl of crystallant solution were mixed and equilibrated in crystal tape-sealed Terazaki plates. Tape can be used instead of oil as crystals can be recovered more easily. Once

crystals could be reproduced, salt, PEG, protein concentrations were varied systematically around initial conditions, as well as mixing volume ratios, i.e vol(protein): vol(crystallant) = 1:1, 2:1, 1:2.

2.2.10 X-ray diffraction data collection

To obtain crystallographic data crystals are analyzed with X-rays. This can be done at synchrotrons or using laboratory X-ray diffractometers.

Crystal fishing and mounting

To collect diffraction data crystals need to be fished with small loops (Fig. 2.37a) and cryocooled. Cryocooling serves to reduce the radiation damage during the X-ray diffraction experiment. If the mother liquor does not prevent ice crystal formation upon cryocooling, cryoprotectants such as PEG400, glycerol or ethylene glycol are introduced by soaking. Crystals are flash-cooled in liquid nitrogen directly after fishing and kept and transported to the synchrotron in a dewar filled with liquid nitrogen (Fig. 2.37b).

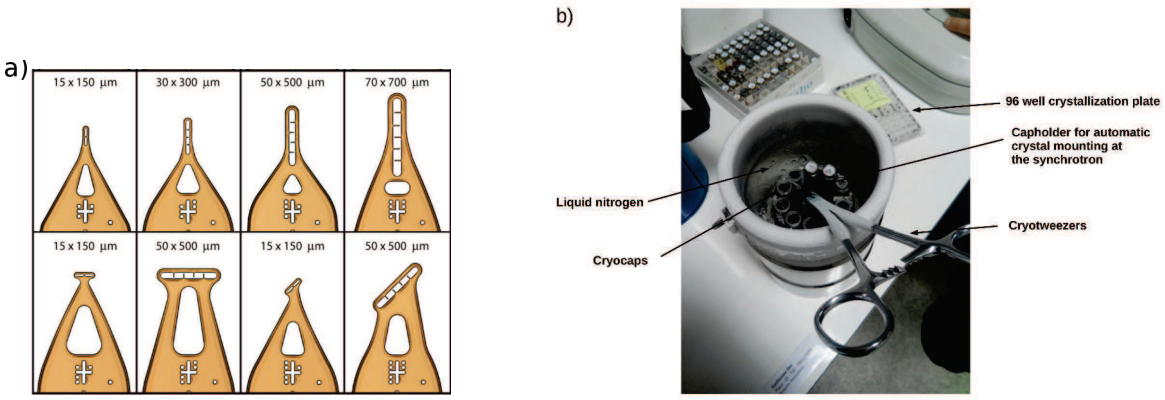
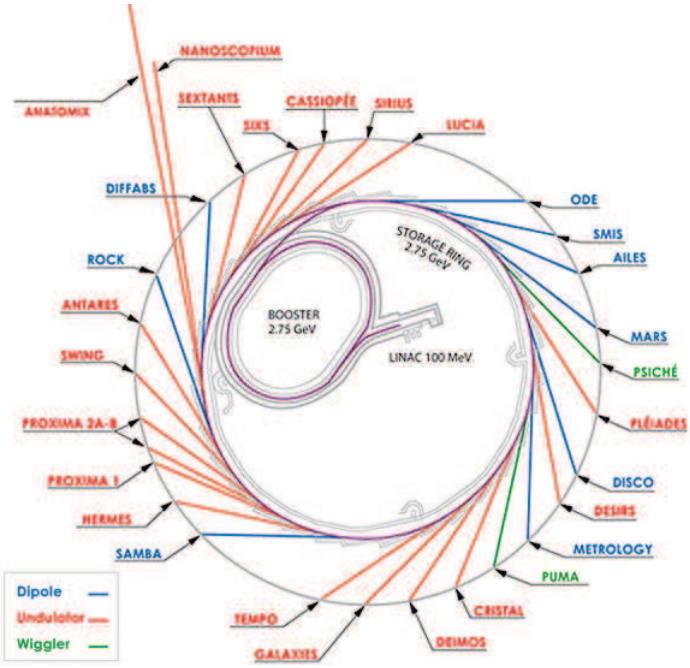


Figure 2.37: Crystal fishing and handling. a) An example of available litholoops that are used for fishing the crystals from their mother liquor (MiTeGen, 2014). b) Crystal handling and storage in liquid nitrogen directly after fishing. The fished crystals that are placed in a loop are transferred into vials that are placed in a capholder. This capholder can be put in an automatic sample exchanger at the synchrotron.

Synchrotron

A synchrotron is an electron accelerator producing different types of very bright light in the region from infrared through to X-rays. X-rays produced by a synchrotron are 10^8 times brighter than that from a laboratory diffractometer. The experimental facilities of a synchrotron beam center are called beamlines (Fig. 2.2.10). These consist of a group of three successive cabins where the beam is captured, selected, focused, and directed toward the samples. Each line can be used for one or more analytical techniques: chemistry, physics, materials, biology, medicine, environment, astrophysics applying diffraction/scattering, spectroscopic, polarimetric or imaging techniques (Candle, 2014). Undulators and wigglers are devices to generate synchrotron radiation, so-called insertion devices. They consist of a series of dipolmagnets causing the passing electrons to accelerate and emitting radiation of high intensity and focus.



In a typical X-ray diffraction experiment the user disposes the crystals already fished and flash-frozen in a sample holder that can be loaded in an automatic sample exchanger. As synchrotron radiation is very intense samples will inevitably suffer from radiation damage.

PRORP2 X-ray data collection and analysis Initial and optimized crystals were analysed at PX1, SLS, Switzerland using a Pilatus 2M detector. PRORP crystals were mounted manually in litholoops and flash-frozen in liquid nitrogen. Data collection was done at 100 K. The X-ray wavelength was 1 Åand 2.07 Åwith a crystal-to-detector distance of 300 mm. Three complete datasets were obtained by collecting 1440 images with 0.25 ° of oscillation. Integration and scaling of diffraction intensities were performed with XDS (Kabsch, 2010), ccp4 software suite (Winn *et al.*, 2011), self-rotation function analysis with GLRF ((Tong & Rossmann, 1997) and molecular replacement using AMoRe (Trapani & Navaza, 2008) and phaser in the phenix package (McCoy *et al.*, 2007, Adams *et al.*, 2010).

3 Protein and RNA production

One important challenge in structural biology is the production of high quality marcomolecules. In particular, high purity and homogeneity are required in crystallography to ensure the growth of well ordered crystals leading to high resolution diffraction data. In my work, these parameters were systematically assessed by standard gel electrophoresis (SDS-PAGE), enzymatic activity assays, as well as size exclusion chromatography (SEC), often used as final purification step and dynamic light scattering (DLS). The samples also need to be purified quickly (to avoid degradation), to be stable for storage during several days without freezing. Finally, they must be available in sufficient amounts to carry out a global structural characterization using methods such as isothermal titration calorimetry (ITC), microscale thermophoresis (MST), analytical ultracentrifugation (AUC) or small angle X-ray scattering (SAXS).
For these reasons I had to scale up and optimize the purification protocols. In order to maximize crystallization probability and to exclude influence of the His-tag, several variants of PRORP1-2-3 were tested during my thesis (Fig 3.2, Aln. 3.1). An advantage of His-tagged protein purification is its rapidity. Combining NiNTA with SEC chromatography the protein is ready for further use after two days and tow steps.
Another strategy is a tag-free protein preparation using a protein in fusion with a self-cleavable sequence like an intein [1] fusing the protein sequence to a self-cleavable protein sequence, i.e. an intein. This purification method involves two steps of on-column cleavage over night which takes longer and yields are comparatively low.

3.1 PRORP1-2-3 constructs

Sequences of wild type PRORP1-2-3 are shown in alignment 3.1 and constructs used during this work are summarized in Fig. 3.2. Two aspartates that are conserved in all PRORP sequences and that are crucial for PRORP activity were mutated into alanines and are referred to as PRORPmDD. The positions are D474/D475 and D421/D422 in PRORP1-2, respectively. In my work all of the biophysical characterizations of the complex were carried out with PRORP2mDD and different tRNA substrates.

[1] An intein is a self-cleavable peptide sequence and is implicated in a process called protein splicing (Anraku *et al.*, 2005).

```
AT|PRORP1|full/1-572  MLRLTCFTPSFSRACCPLFAMMLKVPSVHLHHPRFSPFRF    40
AT|PRORP2|full/1-528  ........................................     0
AT|PRORP3|full/1-517  ........................................     0

AT|PRORP1|full/1-572  YHTSLLVKGTRDRRLILVERSRHLCTLPLAAAKQSAASPS    80
AT|PRORP2|full/1-528  ......MAASDQH......RSRR..............HD    13
AT|PRORP3|full/1-517  ......MAGTDNR......RSRH..............DD    13

AT|PRORP1|full/1-572  ENLSRKAKKKAIQQSPEALLKQKLDMCSKKGDVLEALRLY   120
AT|PRORP2|full/1-528  ESSSRPNKKKKVSRNPETNLLFNLNSCSKSKDLSAALALY    53
AT|PRORP3|full/1-517  ESPKNPNKKKKGNRNPEKSLLINLHSCSKRKDLSAALALY    53

AT|PRORP1|full/1-572  DEARRN.GVQLSQYHYNVLLYVCSLAEAATESSPNPGLSR   159
AT|PRORP2|full/1-528  DAAITSSEVRLSQQHFQTLLYLCSASITDI.SLQYLAIDR    92
AT|PRORP3|full/1-517  DAAITSSDIRLNQQHFQSLLYLCSAFISDP.SLQTVAIDR    92

AT|PRORP1|full/1-572  GFDIFKQMIVDKVVPNEATFTNGARLAVAKDDPEMAFDMV   199
AT|PRORP2|full/1-528  GFEIFDRMVSSGISPNEASVTSVARLAAAKGNGDYAFKVV   132
AT|PRORP3|full/1-517  GFQIFDRMVSSGISPNESSVTAVARLAAAKGDGDYAFKLV   132

AT|PRORP1|full/1-572  KQMKAFG..IQPRLRSYGPALFGFCRKGDADKAYEVDAHM   237
AT|PRORP2|full/1-528  KEFVSVGGVSIPRLRTYAPALLCFCEKLEAEKGYEVEEHM   172
AT|PRORP3|full/1-517  KDLVAVGGVSVPRLRTYAPALLCFCDTLEAEKGYEVEDHM   172

AT|PRORP1|full/1-572  VESEVVPEEPELAALLKVSMDTKNADKVYKTLQRLRDLVR   277
AT|PRORP2|full/1-528  EAAGIALEEAEISALLKVSAATGRENKVYRYLHKLREYVG   212
AT|PRORP3|full/1-517  DASGIVLEEAEISALLKVSAATGRENKVYRYLQKLRECVG   212

AT|PRORP1|full/1-572  QVSKSTFDMIEEWFKSEVATKTG..VKKWDVKKIRDAVVS   315
AT|PRORP2|full/1-528  CVSEETLKIIEEWFCGEKAGEVGDNGIGSDVGMLREAVLN   252
AT|PRORP3|full/1-517  CVSEETSKAIEEWFYGVKASEVSDNGIGSDIELLRAAVLK   252

AT|PRORP1|full/1-572  GGGGWHGQGWLGTGKWNVKRTEMDENGVCKCCKEKLVCID   355
AT|PRORP2|full/1-528  NGGGWHGHGWVGEGKWTVKKGNVSSTGRCLSCSEQLACVD   292
AT|PRORP3|full/1-517  NGGGWHGLGWVGEGKWIVKKGNVSSAGKCLSCDEHLACVD   292

AT|PRORP1|full/1-572  INPVETETFAASLTRLACERFVKAN.......FNQFQEWL   388
AT|PRORP2|full/1-528  TNEVETQKFVDSLVALAMDRKTKMNSCETNVVFSEFQDWL   332
AT|PRORP3|full/1-517  TNEVETEDFVNSLVTLAMERKAKMNSCEPMADFSEFQEWL   332

AT|PRORP1|full/1-572  ERHGPFDAVIDGANMGLVNQ....RSFSFFQLNNTVQRCQ   424
AT|PRORP2|full/1-528  EKHGDYEAIVDGANIGLYQQNFVDGSFSLSQLESVMKELY   372
AT|PRORP3|full/1-517  EKHGDYEAILDGANIGLYQQNFADGGFSLPQLEAVVKELY   372

AT|PRORP1|full/1-572  QISPSKRLPLVILHKSRVNGGPATYPKNRALLEKWKNAGA   464
AT|PRORP2|full/1-528  RESGNNKWPLILLHKRRVK.TLLENPTHRNLVEEWISNGV   411
AT|PRORP3|full/1-517  NKSGSKKQPLILLHKKRVN.ALLENPNHRNLVEEWINNNV   411
```

```
AT|PRORP1|full/1-572   LYATPPGSNDDWYWLYAAVSCKCLLVTNDEMRDHLFQLLG   504
AT|PRORP2|full/1-528   LYATPPGSNDDWYWLYAAAKLKCLLVTNDEMRDHIFELLG   451
AT|PRORP3|full/1-517   LYATPPGSNDDWYWLYAAAKLKCLLVTNDEMRDHIFELLS   451

AT|PRORP1|full/1-572   NSFFPRWKEKHQVRISVTREDGLKLNMPPPYSIVIQESED   544
AT|PRORP2|full/1-528   STFFQKWKERHQVRYTFVKG.NLKLEMPSPFSVVIQESEK   490
AT|PRORP3|full/1-517   NSFFQKWKERHQVRFTFVKG.CLKLEMPPPFSVVIQESEK   490

AT|PRORP1|full/1-572   GTWHVPMSVEDDLQTSRQWLCAKRSKTP.........   572
AT|PRORP2|full/1-528   GSWHFPVSCENNEESSRTWMCISRQSILDSPKSNGKIP   528
AT|PRORP3|full/1-517   GSWHVPITSQDKEESLRSWMCITRQSS.........   517
```

Figure 3.1: Alignment of amino acid sequences of *A. thaliana* wild type PRORP1-2-3. Amino acids are coloured by similarity. Purple corresponds to 100 % sequence identity, pink represents a sequence identity $\geq$ 50 % of amino acids with the similar properties, blue is sequence identity $\geq$ 50 %.

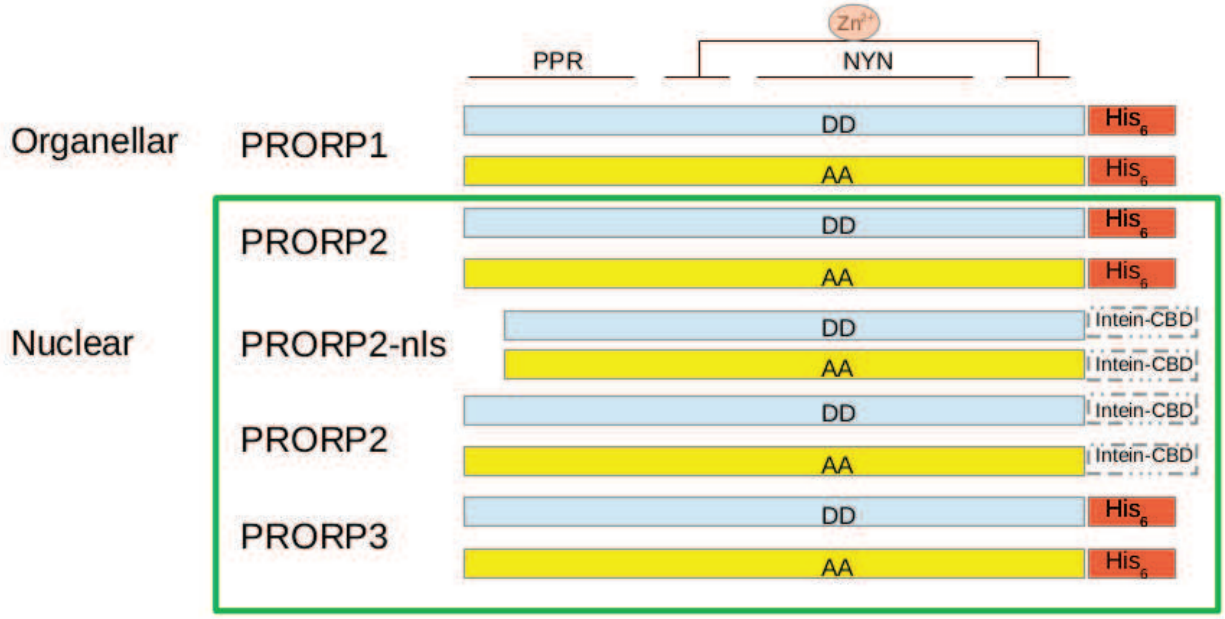


Figure 3.2: Scheme of PRORP constructs used during my work: -nls without nuclear localization signal, DD - wild type with conserved aspartates in active site , AA - double catalytic mutant with aspartates mutated into alanines, NYN - catalytic domain, PPR - pentatricopeptide repeat, Zn - bipartite zinc binding domain, intein-CBD - cleaved chitin binding domain during purification.

3.2 Recloning of PRORP2 in pTYB-vectors

A disadvantage of His_6-tags on proteins is that they introduce a floppy tail that can prevent crystallization in some cases. Therefore, in order to increase crystallization probability, we cloned the cDNA of PRORP2 and variants into pTYB1 and pTYB12 vectors. These vectors express proteins fused to an intein-chitin binding domain. When putting the protein then in reducing conditions the intein will excise itself out of the protein. The excision will be imperfect due to a mutation in the intein sequence

and elution of the target protein from then matrix. Starting from the His-tagged cDNA in the expression vector pET28b(+), forward and reverse primers were designed to amplify the wild type PRORP2 and catalytic mutant with restrictions sites appropriate for the cloning in pTYB1 and pTYB12 (Tab. 2.4) between Nhe1/Xho1 and Bsm1/Xho1, respectively. After intein cleavage PRORP2 has with five additional C-terminal amino acids (LEGSS) or two N-terminal amino acids (AA) for pTYB1 and pTYB12, respectively. In order to test if these proteins are well expressed and efficiently cleaved they were coupled in N- and C- terminal to the intein domain. Indeed, PRORP2 fused with the C-terminal intein sequence was well expressed and processed, but expression and cleavage efficiency of the N-terminal fusion constructs was weak (Fig. 3.3a). In addition, a shorter version of PRORP2, lacking the nuclear localization signal was also cloned into these vectors.

The drawback of this purification method is that pTYB1 is a low copy number vector and protein expression is lower compared to pET28 protein expression. The affinity chromatography step takes at least two days as DTT induced intein cleavage is not very efficient and has to be repeated twice. In addition to this the chitin matrix can only be regenerated five times. Still, proteins purified without tag are of high purity and homogeneity (Fig. 3.3b-d). The first expression tests and purifications of tag-free PRORPs were carried out by Olivier Fuchsbauer, a research engineer in the team.

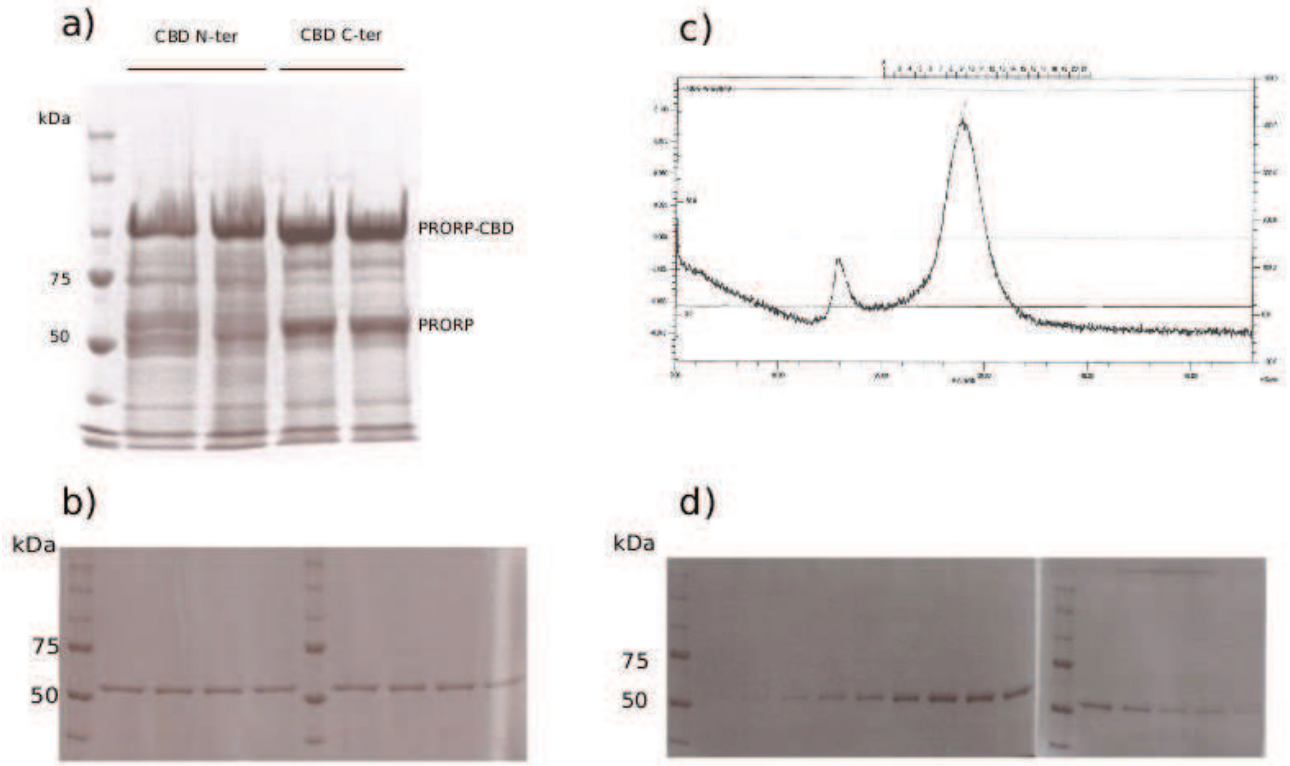


Figure 3.3: Expression and purification of PRORP with intein domain. a) PRORP2 expression test. N-terminal CBD is not cleaved and PRORP not released whereas the CBD C-terminal is efficiently cleaved and PRORP released. Lane 1, 3 correspond to PRORP with nls and lanes 2, 4 without nls signal peptide sequence. b) SDS-PAGE after the first elution of cleaved PRORPs. c)+d) Size exclusion chromatogram and a corresponding 8 % SDS-PAGE gel.

3.3 Optimizing PRORP purification containing a His-tag

The affinity chromatographic matrix enables the separation and a fast binding of His_6-tagged proteins from large amounts of *E.coli* proteins. Then during SEC, aggregates and remaining impurities can be eliminated. Fig. 3.4 shows an example of the purification results before and after optimization. The optimized purification protocol replaced batch affinity purification with multiple different buffers by a HPLC-assisted purification using three washing steps to eliminate the maximum of contaminating proteins. This protocol can be read in detail in Gobert *et al.* (2010). A short step at 250 mM imidazole elutes PRORPs at high purity in a small volume (Fig. 3.4). A buffer containing high salt and glycerol concentration was chosen to increase stability and shelf life in the fridge to avoid freezing of the proteins.

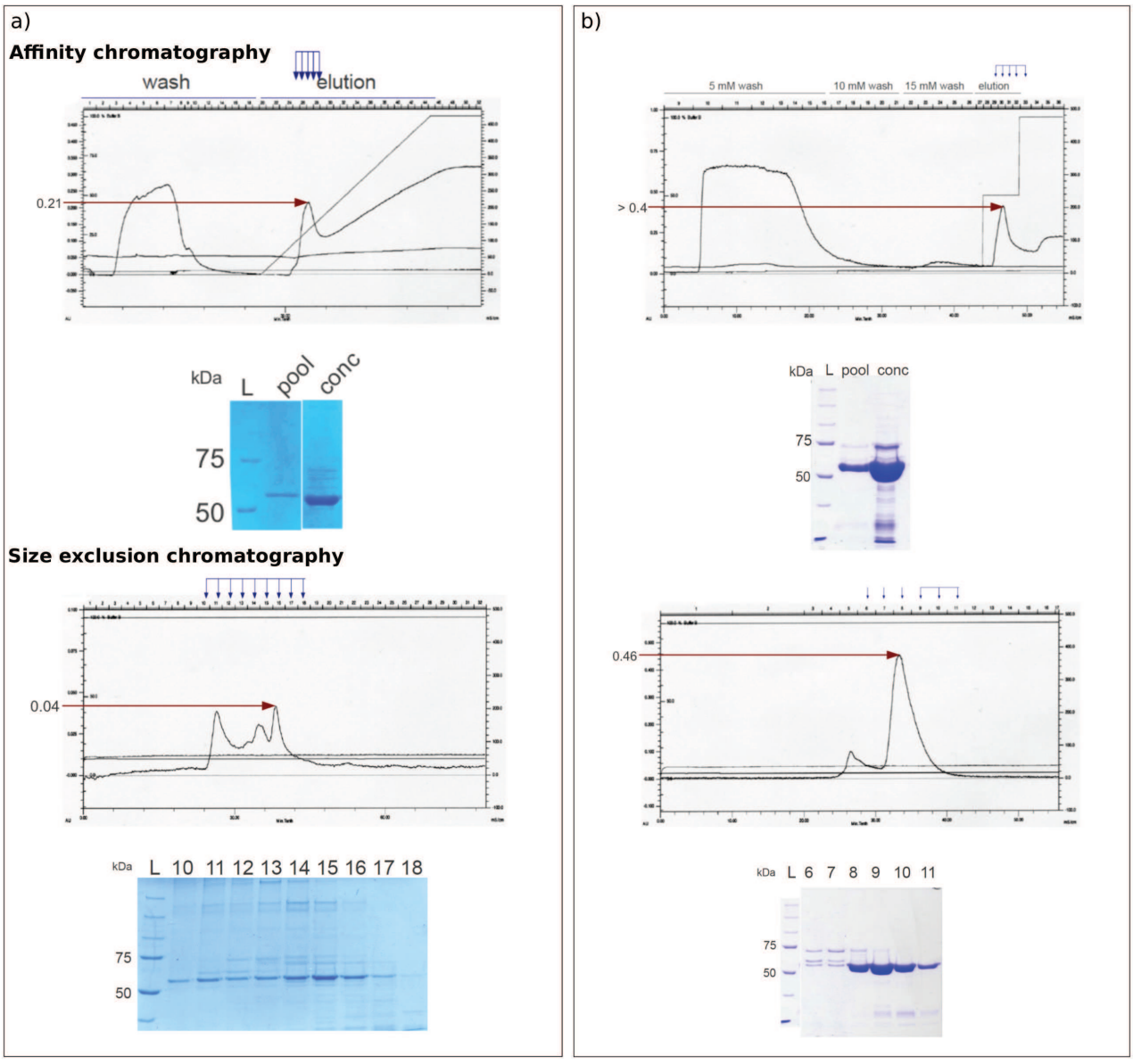


Figure 3.4: PRORP-His_6 purification optimization. Purification results a) before and b) after optimization. Overall, after optimization, protein expression is higher, the proteins elute in a sharper peak and are pure after gel filtration.

3.4 tRNA purification

tRNA purification needs to be efficient to prepare milligram-amounts required for structural studies. Initially, *in vitro* transcribed tRNAs were purified using a classical denaturing polyacrylamide gel purification approach. This has some disadvantages:

1. Large amounts of polyacrylamide is needed, that is toxic,
2. RNA are purified under non-native conditions and structural integrity is not assured,
3. Poor elution efficiency of RNA from PAA gel slices,
4. Long purification procedure.

One advantage however is that RNAs can be purified with a high resolution of a single nucleotide.
A second method is purification using chromatography. As an initial step a weak anion exchanger, like a DEAE matrix, is used to separate the transcript of interest from the nucleotides, abortive transcripts and the T7 polymerase. The second polishing step of gel filtration allows buffer change and separation of transcripts of different sizes. I found that the most efficient way to purify the tRNA was using the purification profile described in section 2.2.2.4. tRNAs were separated on DEAE matrix after extensive washing with a gradient ranging from 290 mM to 670 mM NaCl. Although slight contaminations are present after purification revealed with ethidium bromide, samples are pure and homogeneous enough for SAXS, ITC and DLS experiments.

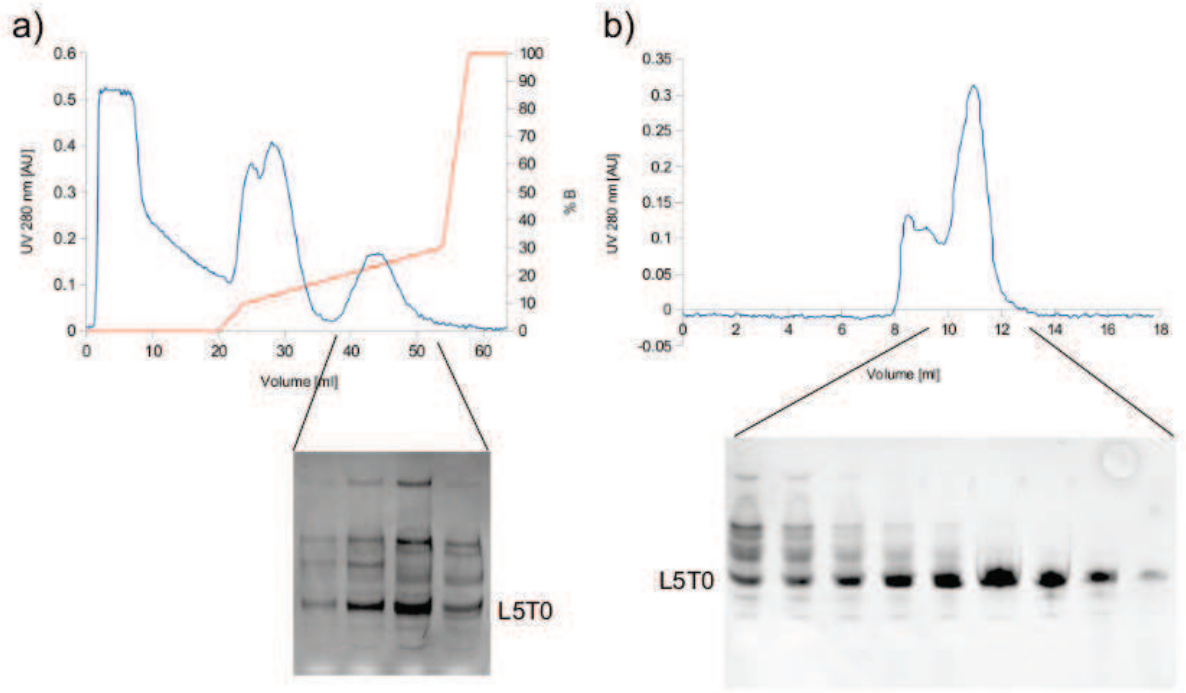


Figure 3.5: Chromatographic purification of tRNA. a) Anion exchange chromatography with corresponding 12 % PAA/8M Urea gel, b) Size exclusion chromatography with corresponding 12 % PAA/8M Urea gel. Gels were stained with ethidium bromide.

3.5 Summary

To summarize after optimization of both, protein and RNA purification protocols, macromolecules of high purity and homogeneity can be obtained within 2-4 days. Even if RNA purity is higher after PAGE purification, yield is lower compared to using chromatography where sample purity is still sufficient for structural studies. Typical yields are indicated in Fig. 3.6.

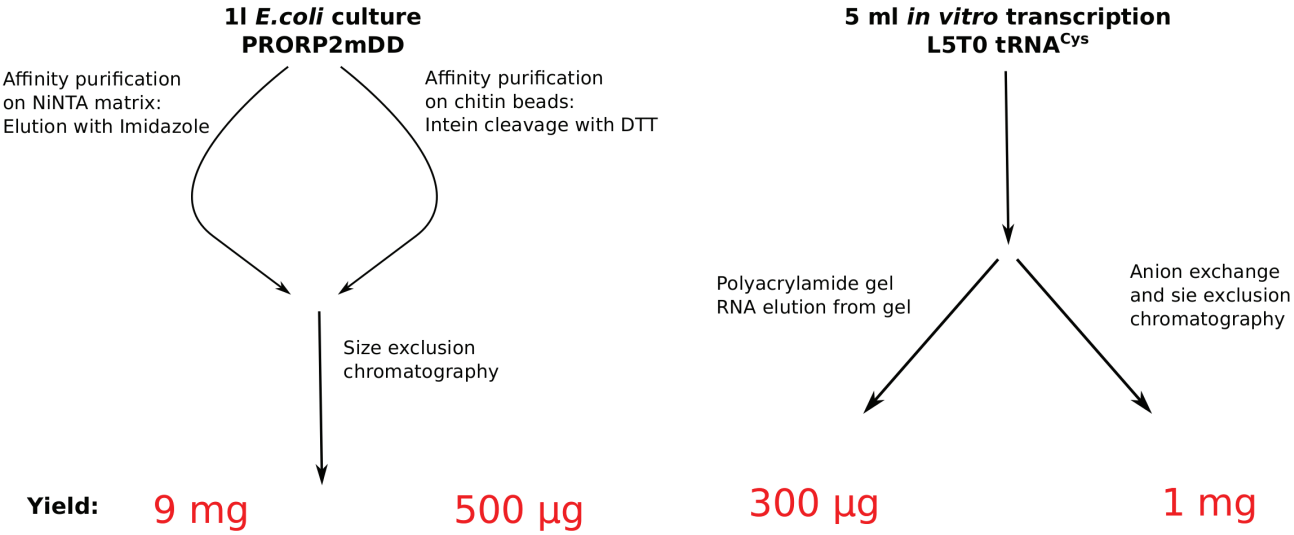


Figure 3.6: Yield of tRNA and PRORP production using different purification protocols. Yields may also vary depending on the constructs.

4 First biophysical and biochemical characterization of PRORP/tRNA interactions

4.1 A multidisciplinary approach

At the time this study was carried out and the article below was submitted no structural data of PRORP was available. This work is the result of a multidisciplinary approach of biophysical and functional studies to decipher how tRNAs are recognized and cleaved by proteinaceous RNase P. It was known that the mode of action must be different to RNP RNase P as tRNAs with a phosphothioate backbone are well processed by PRORPs whereas RNP RNase P dramatically looses activity. We showed that there are several conserved nucleotides of a tRNA that are indispensable of PRORP activity, i.e. G18 in the D-loop, C56 in the TΨC-loop, a R57, while others are anti-determinants like a CCA at the 3' end. Certain positions in the D/TΨC-loop are in contact with the protein but interestingly no interaction were shown in the foot print assays with the 5' leader. A combination of DLS, SEC, MALS and SAXS experiments confirmed that PRORP1-2 are present as monomers with a molecular weight of 60 kDa and SRCD experiments could show that PRORP1-2 contain a high content of α-helices as predicted with bioinformatic tools. Furthermore, using SAXS we could show that PRORPs are composed of two distinct domains with a rather long extension. Based on the SAXS envelope and using homology models of the PPR and the NYN domain we could construct a first model of PRORP in complex with a tRNA.
In conclusion, taken all the results together we propose that PRORPs recognize canonical tRNA rather by structure than by sequence and that this system is another example of structural mimicry.

4.2 Demonstrating the presence of zinc ions in PRORP1 and PRORP2

Preliminary work suggested the presence of zinc ions in the metallonuclease domain of human mitochondrial PRORP (MRPP3). All three *A. thaliana* PRORPs contain a putative zinc binding motif CxxC. Using ICP-MS the presence of zinc ions could be validated in wild type PRORP1 and PRORP2 (Tab. 4.1). In neither purification buffer zinc was added. So with the atom mass of zinc of 65 g/mol, a stoichiometry of one zinc ion per PRORP monomer could be calculated (Tab. 4.1). A study of PRORP1 and its zinc fixation mutants is included in the article.

Table 4.1: Stoichiometry and zinc concentration in wild type PRORP2. Concentration of wild type PRORP1-2 was 29 and 14 mg/ml, respectively.

	PRORP [mg/ml]	**Zn^{66} (p.p.b.)**	**Stoichiometry**
PRORP1 wild type	29	29.49	1
PRORP2 wild type	14	14.5	1
Buffer	0.44	0	

4.3 Publication 2: Structural insights into protein-only RNase P complexed with tRNA

nature COMMUNICATIONS

ARTICLE

Received 6 Jul 2012 | Accepted 5 Dec 2012 | Published 15 Jan 2013

DOI: 10.1038/ncomms2358 OPEN

Structural insights into protein-only RNase P complexed with tRNA

Anthony Gobert[1,*], Franziska Pinker[1,2,*], Olivier Fuchsbauer[2], Bernard Gutmann[1], René Boutin[3], Pierre Roblin[4,5], Claude Sauter[2] & Philippe Giegé[1]

RNase P is the essential activity removing 5′-leader sequences from transfer RNA precursors. RNase P was always associated with ribonucleoprotein complexes before the discovery of protein-only RNase P enzymes called PRORPs (PROteinaceous RNase P) in eukaryotes. Here we provide biophysical and functional data to understand the mode of action of PRORP enzymes. Activity assays and footprinting experiments show that the anticodon domain of transfer RNA is dispensable, whereas individual residues in D and TψC loops are essential for PRORP function. PRORP proteins are characterized in solution and a molecular envelope is derived from small-angle X-ray scattering. Conserved residues are shown to be involved in the binding of one zinc atom to PRORP. These results facilitate the elaboration of a model of the PRORP/transfer RNA interaction. The comparison with the ribonucleoprotein RNase P/transfer RNA complex suggests that transfer RNA recognition by PRORP proteins is similar to that by ribonucleoprotein RNase P.

[1] Institut de Biologie Moléculaire des Plantes du CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France. [2] Institut de Biologie Moléculaire et Cellulaire du CNRS, Architecture et Réactivité de l'ARN, Université de Strasbourg, 15 rue René Descartes, 67084 Strasbourg, France. [3] Laboratoire d'Hydrologie et de Géochimie du CNRS, 1, rue Blessig, 67084 Strasbourg, France. [4] Synchrotron SOLEIL, l'Orme des Merisiers Saint-Aubin, 91410 Gif-sur-Yvette, France. [5] URBIA-Nantes, INRA Centre de Nantes, 60 rue de la Géraudière, 44316 Nantes, France. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.G. (email: philippe.giege@ibmp-cnrs.unistra.fr) or to C.S. (email: c.sauter@ibmc-cnrs.unistra.fr).

RNase P is the ubiquitous activity that catalyses the 5′-maturation of transfer RNAs (tRNAs), as well as of a number of other substrates such as ribosomal RNA, messenger RNA, transfer-messenger RNA or riboswitches[1–3]. RNase P was first described in bacteria where it is composed by a ribonucleoprotein (RNP) complex whose RNA component (P RNA) holds the catalytic activity[4]. RNP RNase P was later found in all three main branches of life, that is, Bacteria, Archaea and Eukarya, and was thus believed to occur universally as a RNP complex[5]. This concept was challenged by early experiments in human mitochondria and spinach chloroplasts that suggested that another type of RNase P devoid of RNA component existed in these organelles[6,7]. Still the dogma of the universality of RNP RNase P remained until the recent characterization of a novel type of RNase P in human mitochondria and plant organelles[8,9]. This novel variant is composed of a single protein that we called PRORP (for PROteinaceous RNase P) and occurs in nearly all major phyla of eukaryotes[9]. Furthermore, RNP RNase P has not been retained in all organisms because, in both *Arabidopsis* and *Trypanosoma,* PRORP enzymes were found to support RNase P activity in both organelles and the nucleus[10,11].

The discovery of PRORP enzymes leads to the question of the respective mode of action of RNP and protein enzymes catalysing the same reaction. RNP RNase P activity is well characterised[2,12], in particular, recent advances such as the determination of the three-dimentional structure of a bacterial RNase P in complex with tRNA have been very important developments[13]. Substrate recognition by RNP RNase P involves the binding to regions distant from the actual cleavage site. It includes stacking interactions between bases in the D and TψC loops of tRNAs and the P RNA specificity (S) domain, an A-minor interaction at the acceptor stem and the formation of canonical base pairs at the 3′-end of tRNA. In particular, key interactions take place between the unstacked bases G19 and C56 of tRNA and the S domain of RNase P. It is also notable that no interaction takes place with the anticodon arm of tRNA. The catalytic active site of RNP RNase P is composed of phosphate backbone moieties, a conserved uridine and at least two catalytically important metal ions[13].

In contrast, the mode of action of PRORP proteins is unknown. These protein-only RNase P enzymes are characterized by the presence of pentatricopeptide (PPR) repeats[14] in their N-terminal part that are believed to be involved in RNA binding and possess an upstream zinc-finger-like motif. The most conserved part of PRORP enzymes lies in their C-terminal part. This region was predicted to be a metallonuclease domain[8] and consigns PRORP to the large family of PIN-like/NYN (N4BP1, YacP-like Nuclease) domain putative ribonucleases[15].

Initial comparison of PRORP and RNP RNase P has suggested that the two classes of enzymes share common features. They both appear to require Mg^{2+} for phosphodiester hydrolysis and both generate 5′-phosphate and 3′-hydroxyl products[8]. However, studies using spinach chloroplast extracts and recombinant *Arabidopsis* PRORP have shown that PRORP is a fundamentally different catalyst than RNP RNase P. The replacement of the phosphodiester backbone of a precursor tRNA by a phosphorothioate moiety at the level of the 5′-maturation site resulted in a strong inhibition of bacterial RNase P activity, while PRORP activity was unaltered[16,17].

To gain functional insight into this novel type of RNase P activity, we investigated how *Arabidopsis* PRORP1 binds tRNA substrates and we performed a biophysical characterization of PRORP1 and 2. This enabled us to define initial mechanistic data on PRORP mode of action. The proposed mode of RNA recognition by PRORP shows striking similarity with that of RNP RNase P, which suggests that protein-only RNase P might have converged to the same tRNA-binding strategy as RNP RNase P.

Results

tRNA cis elements required for PRORP activity. To get mechanistic insights into the mode of action, in particular of RNA recognition of PRORP enzymes, we performed RNase P cleavage assays with recombinant PRORP1 and different mutants of mitochondrial $tRNA^{Cys}$ precursor, a known substrate of PRORP1 *in vivo* (Fig. 1). We first removed the anticodon domain from the tRNA precursor, which did not result in significant decrease of cleavage by PRORP. Then, the removal of both the D and anticodon domains was tested. The resulting mini helix was not cleavable by PRORP. As PRORP enzymes are able to cleave the 5′-leader sequence of any tRNA of canonical structure *in vitro*[8–11], we postulated that the determinants for tRNA recognition by PRORP must reside among positions universally conserved in tRNAs[18]. We thus applied point mutations to such positions in tRNAs to investigate their effect on PRORP activity. The mutation of G18 in the D-loop to A and C, respectively, resulted in severe impairment and total loss of RNase P activity, whereas the mutation of G19 to A or C did not affect RNase P activity. However, mutations of C56 in the TψC loop to A or G resulted in total loss of RNase P cleavage. In the same loop, mutations of G57 to A and C resulted in unaffected and total loss of PRORP cleavage, respectively, consistent with the conservation of a purine at position 57 in tRNA[18]. Next, the exchange of G-C, the first base pair of the acceptor stem by C-G did not result in decreased cleavage efficiency, although we cannot exclude that mis-cleavage did not occur. Finally, we investigated the nature of the 3′-end of tRNA precursors. The absence of a 3′-trailer sequence did not affect 5′-cleavage, whereas the occurrence of a 3′-CCA group strongly reduced RNase P activity (Fig. 1). Further analyses will be necessary to determine if 3′-CCA groups act as PRORP binding antideterminants for all tRNAs, and to uncover the precise involvement of the length and the nature of residues in 3′-trailer sequences for PRORP activity.

Taken together, our results show that the anticodon domain is not involved in RNA recognition by PRORP. The nature of residues at positions 1 and 72 is not discriminant for the activity, while the 3′-CCA seems to act as an antideterminant for PRORP binding. As residues at positions 18 and 57 are involved in interactions between loops D and TψC of tRNA[18], an interaction between loops D and TψC appears to be strictly required for PRORP function. However, the conservation of the G19:C56 interaction does not appear to be critical, although the presence of a cytidine at position 56 seems to be indispensable. Thus, precise residues in loops D and TψC seem to be essential for substrate recognition by PRORP.

tRNA residues in interaction with PRORP. We performed a footprinting analysis in order to map precisely contact points between a PRORP protein and its tRNA precursor substrate (Fig. 2). To determine the tRNA regions that are in interaction with PRORP1, a mitochondrial $tRNA^{Cys}$ precursor containing a leader sequence of five nucleotides was incubated either alone or in complex with PRORP1 and subjected to digestion by nucleases. As the PRORP/tRNA complex used here was UV-crosslinked, we verified with an *in vitro* activity assay that substrate binding had resulted in a catalytically active complex (Supplementary Fig. S1). Three different nucleases were used for this analysis, RNase V1 that only cleaves base-paired RNA regions, RNase T1 that cuts single-stranded RNA only after guanosines and RNase A that cleaves single-stranded RNA after cytidines and uridines. Positions of residues protected from RNase digestion by PRORP

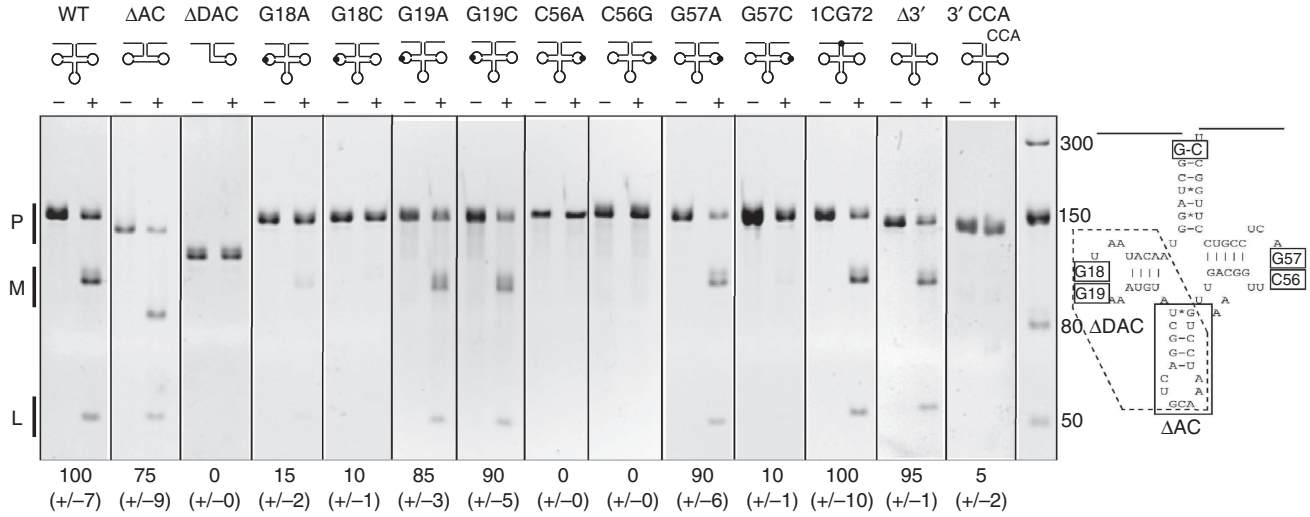


Figure 1 | RNase P *in vitro* cleavage assays performed with *Arabidopsis* recombinant PRORP1 and variants of mitochondrial tRNA^{Cys} precursors. + and − indicate the absence and presence of PRORP proteins in the reactions. WT is the wild-type tRNA^{Cys}, ΔAC the tRNA without the anticodon domain, and ΔDAC without both anticodon and D domains. G18A, for example, shows a tRNA where the guanosine at position 18 was mutated into an adenine. 1CG72 is a tRNA where the G-C base pair at positions 1 and 72 was swapped to a C-G. Δ3′-shows a tRNA precursor without 3′-trailer sequence and 3′-CCA, the precursor with a mature 3′-end containing a CCA. P stands for tRNA precursors, M the 5′-mature products and L the cleaved 5′-leader fragments. The molecular weights of markers are given in nucleotides. PRORP cleavage products were quantified with ImageGauge (Fujifilm). Values were normalized so that 100 corresponds to the cleavage efficiency observed for wild-type tRNA^{Cys} precursor. Cleavage efficiencies are given below the respective panels together with s.d.'s for three representative experiments.

could be mapped down to individual nucleotides through the comparison of RNase digestion profiles with an RNase T1 ladder and the alkaline hydrolysis profile of the mitochondrial tRNA^{Cys} precursor. We observed that discrete positions in the tRNA D-loop, namely U16, G18 and G19 were protected from nuclease digestion by PRORP. Similarly, C56 in the TψC loop was protected from nuclease digestion by PRORP. No other position, in particular close to the actual cleavage site of the tRNA precursor could be reproducibly identified as a site of nuclease protection by PRORP. Altogether, this indicates that individual residues that are in close spatial vicinity in loops D and TψC of tRNA are binding sites for protein-only RNase P enzymes (Fig. 2).

Structural properties of PRORP 1 and 2 in solution. We characterized in parallel both organellar and nuclear enzymes and focussed our study on PRORP1 and PRORP2 (which displays 80% sequence identity with PRORP3). *Arabidopsis* PRORP proteins are active as single-protein enzymes[9]. Their hydrodynamic properties in size exclusion chromatography and in dynamic light scattering confirmed that they are monomers in solution (Fig. 3). The molecular mass determined for PRORP2 in multi-angle light scattering is 62 kDa in good agreement with that calculated from the sequence (60 kDa).

These monodisperse PRORP samples led to sharp synchrotron radiation circular dichroism (SRCD) spectra (Fig. 3d) indicating that both PRORP1 and PRORP2 have a high content in α-helices. The evaluation of PRORP secondary structure content indicates 36/39% of α-helices, 15/16% of β-strands in PRORP1 / PRORP2, respectively. This observation is consistent with structure predictions based on sequence analysis[9](Supplementary Fig. S2).

PRORP samples were further studied by small-angle X-ray scattering (SAXS), a method of structural characterization providing information on the size and shape of biological macromolecules in solution[19–21]. In these experiments, the two enzymes produced very similar scattering curves at small angles, and their estimated gyration radius is $R_g = 33$ Å corresponding to a monomer in solution (Fig. 4). An experimental setup that allows the SAXS analysis downstream of a gel-filtration separation enabled the acquisition of scattering data for PRORP2 with lower noise at higher angles (Fig. 4, blue plot). The derived P(R) function that evaluates the distribution of distances inside the molecular object, confirms the value of R_g and is compatible with an object made of two structured domains. The tail of the distribution ($80 < r < 110$ Å) suggests the presence of extension(s), which may correspond to either N-terminal or C-terminal regions. PRORP proteins are slightly more compact in solution than archaeal and bacterial RNase P RNAs, which display R_g and d_{max} of 38–48 Å and 120–190 Å in SAXS, respectively[22,23]. Overall, they appear as monomeric enzymes with two-domains essentially made of α-helices.

PRORP proteins are zinc-binding enzymes. The analysis of PRORP sequence conservation across eukaryotes revealed that a certain number of residues are highly conserved throughout evolution and might thus be of functional importance[9]. Among them, a putative zinc-finger-like structure is split in two separate motifs. The first motif (CxxC) contains two conserved cysteines upstream of the NYN domain at positions 344 and 347 for PRORP1 (281 and 284 for PRORP2), whereas the second motif involves a conserved histidine and a cysteine, downstream of the NYN domain, at positions 548 and 565, respectively, (Fig. 3a) (494 and 511 for PRORP2). These four particular residues were chosen as best candidates to form a zinc-binding pocket as no other cysteine or histidine outside the catalytic NYN domain is highly conserved among PRORP sequences. We used inductively coupled plasma mass spectrometry to investigate the association of metal cofactors to PRORP proteins. Zinc (^{66}Zn) was present at 29.49 +/− 1.53 p.p.b. in a 30-μg ml^{-1} PRORP1 solution. This corresponds to the occurrence of one zinc atom per PRORP molecule. Other metals were only found as traces. To investigate the importance of the conserved residues for zinc binding, we mutated the four residues to alanines, expressed and purified to

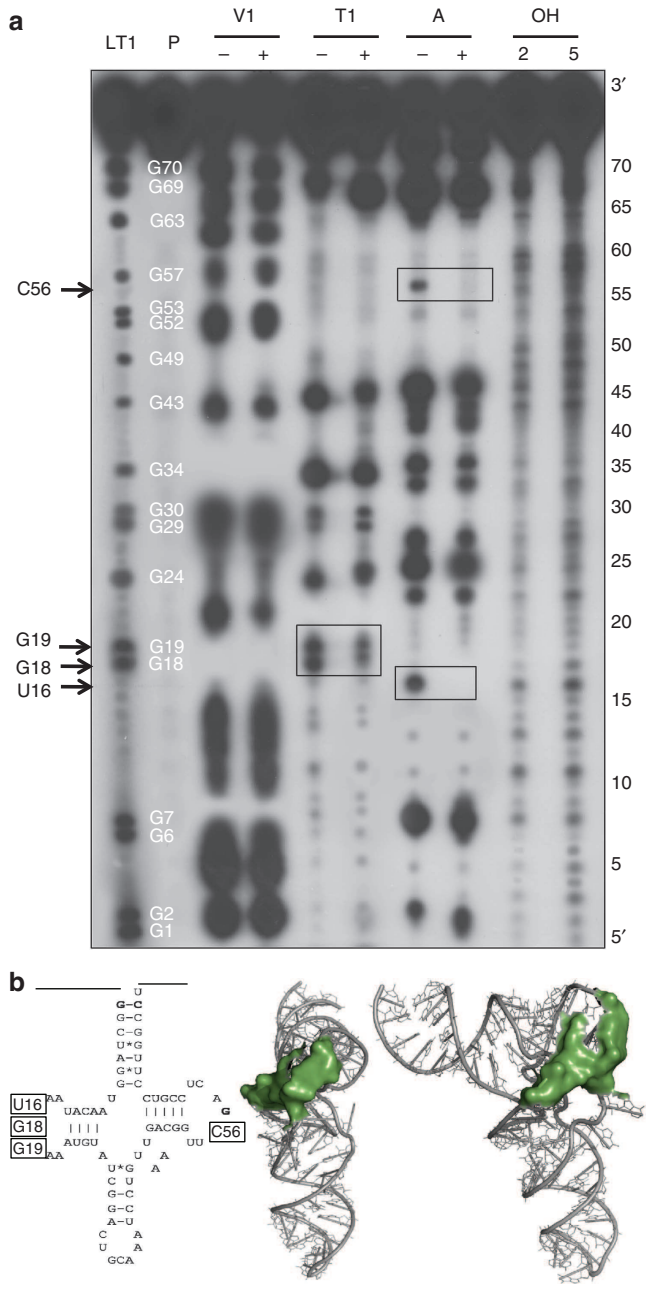


Figure 2 | Footprinting analysis of mitochondrial tRNA^{Cys} precursor in complex with PRORP1. (**a**) Samples were subjected to partial RNase V1, RNase T1 and RNase A digestions. + and − mean that PRORP proteins were present or absent in the reactions. P represents the tRNA precursor probe. LT1 shows an RNase T1 ladder with the corresponding positions of Gs in the tRNA sequence indicated in white. OH show alkaline hydrolysis of the tRNA probe performed for 2 and 5 min to generate an RNA ladder with single-nucleotide increments. RNA samples were separated by high resolution denaturing PAGE. tRNA positions were precisely mapped with the T1 and alkaline ladders. Boxed positions, also indicated on the left by arrows, correspond to tRNA positions reproducibly found protected from nuclease treatment by PRORP interaction in three replicate experiments. (**b**) Secondary and tertiary structural model of mitochondrial tRNA^{Cys} with boxes and green surfaces indicating residues protected by PRORP in footprinting experiments.

homogeneity the respective PRORP1 mutants. The analysis of zinc content in the mutants revealed that in the C344 and C347 mutants zinc levels were reduced by 19% and 29%, respectively, whereas the H548 and the C565 mutants zinc levels decreased by 60% and 75%, respectively (Table 1). As a control we analysed zinc content in the DD474-475 catalytic mutant[9] and found that it was similar to that of wild-type PRORP. The increased lability of zinc in the cysteine and histidine PRORP mutants suggests that the four residues are involved in the stable binding of zinc and that the downstream conserved motif has a stronger affinity for the metal than the upstream CxxC coordination element. We also analysed the capacity of the mutant proteins to perform RNase P activity. The single C565 mutant protein had impaired RNase P activity (Supplementary Fig. S3). The highest lability of zinc in this mutant might have resulted in an unstable protein fold, thus affecting its activity.

Model of the PRORP/tRNA complex. Structural models of PRORP were generated by homology modelling as implemented on the Phyre server[24]. The structure prediction was limited to the two main domains (PPR and NYN) for which templates were identified with Phyre2. Best hits for the three PRORP sequences were a TPR domain (PDBid: 2ooe) and the catalytic domain of an RNase (PDB id: 3v32), respectively (see Supplementary Fig. S2). We also established the structure model of the *Arabidopsis* tRNA^{Cys} used to illustrate footprinting experiments and activity assays. The latter data were combined with the SAXS envelope in order to position domains of PRORP2 (for which we had best SAXS data) with respect to the tRNA substrate. In the docking process, the N-terminal RNA recognition module of PRORP2 containing PPR repeats[9] was placed next to loops D and TψC of the folded tRNA, in particular in contact with positions U16, G18, G19 and C56. For the C-terminal part of the protein, the two conserved aspartates at positions 474 and 475 (Fig. 3a) that were shown to be part of the catalytic active site of PRORP[9] were placed in close vicinity of the tRNA +1 position, where RNase P cleavage takes place (Fig. 5). Our model highlights notable similarities in tRNA-binding mode with the complex of bacterial RNP RNase P where the specificity domain of RNase P RNA interacts with the residues G19 and C56 of the tRNA (Fig. 5).

Discussion

PRORP enzymes were identified as members of the PPR family, a huge class of RNA-binding proteins ubiquitous in eukaryotes[25]. These proteins can be divided into two main super-groups (P and PLS) according to the occurrence of specific classes of PPR domains and of additional C-terminal domains. Several lines of evidence suggest that both types of P and PLS proteins recognize primary sequences of RNA[14]. Interestingly PRORP enzymes do not belong to the two established super-groups. With only very few canonical PPR domains, and the presence of non-canonical putative PPR repeats, they rather define a new subfamily of PPR proteins. As PRORP enzymes bind any tRNA of canonical structure, it is possible that PRORP proteins recognize structured elements of RNA and thus have a mode of RNA recognition distinct from other PPR proteins. Alternatively, our favoured hypothesis is that PPR repeats in PRORP might specifically recognize individual nucleotides in tRNA loops, in particular unstacked bases or residues not involved in Watson–Crick interactions, which are highly conserved among tRNA sequences.

The biophysical characterization of PRORP enzymes has validated bioinformatic predictions and enabled to build a model of the active enzymatic complex. The predominance of SRCD signal for α-helices (Fig. 3d) is in agreement with fold recognition predictions, PPR repeats and NYN domains being mostly composed of α-helices[15,26]. Although the N-terminal region of PRORP does not contain >2–3 canonical PPR motifs, the presence of non-canonical putative PPR domains suggests

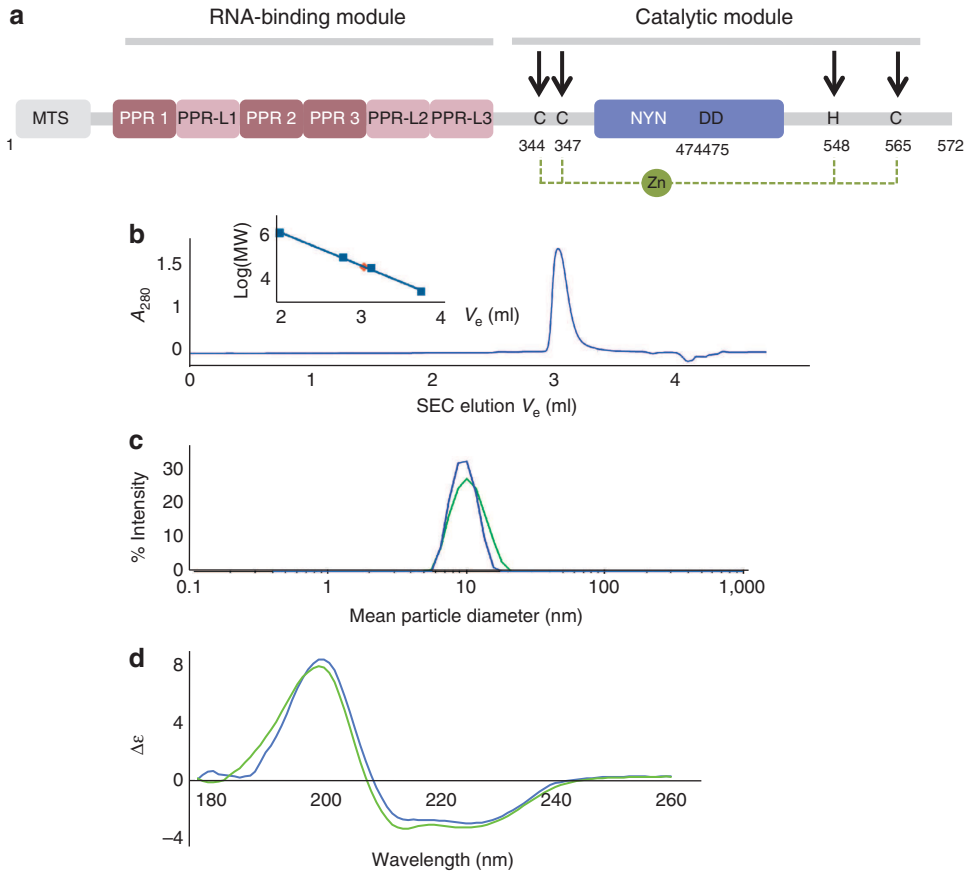


Figure 3 | Characterization of PRORP proteins in solution (**a**) Organization along the sequence of PRORP proteins. In the RNA-binding domain PPR and PPR-L show canonical PPR repeats and putative PPR-like motifs, respectively. For PRORP1, aspartates at positions 474 and 475 are in the catalytic pocket of the enzyme[9], whereas cysteines and a histidine at positions 344, 347, 548 and 565 (indicated by black arrows) are proposed to form a zinc-binding pocket (dashed line) that could stabilize the catalytic domain of PRORP. (**b**) Analysis of PRORP2 oligomeric state in solution by analytical gel-filtration (BioSEC3 column) leading to a MW estimation of 72 kDa (red diamond: Log(MW) = 4,85) by comparison with the elution of model proteins (thyroglobulin: 660 kDa; BSA monomer and dimer: 66, 132 kDa; ribonuclease A: 14 kDa; see inset). (**c**) Hydrodynamic radius distribution for PRORP1 (green) and PRORP2 (blue) in dynamic light scattering, confirming the monodispersity of PRORP samples. (**d**) SRCD analysis of PRORP1 (green) and PRORP2 (blue). SRCD spectra show a dominant peak at 190–200 nm characteristic of α-helices. The evaluation of two-dimentional structure content indicates 36/39% of α-helices, 15/16% of β-strands in PRORP1/PRORP2, respectively.

that this region is arranged in a super-helix as described in structurally related TPR proteins[27], which give the highest score in structure prediction with Phyre2. The C-terminal NYN domain could be modelled based on the structure of the MCPIP1 RNase that adopts an α–β PIN-like/NYN architecture. The Asp residues that are conserved in PRORP sequences are essential to the activity of MCPIP1 (see Supplementary Fig. S2) and their mutation abolished PRORP activity[9,28]. This observation validates the proposed fold. The region that connects the N- and C-terminal domains was identified as a potential zinc-binding motif. Our mutational analysis confirmed that the two conserved Cys residues are involved in metal binding, together with another Cys and a His residue at the C-terminal end of PRORP. Overall, this results in a compact two-domain enzyme, as confirmed by the SAXS analysis in solution, with a zinc ion bridging the central and the C-terminal region (Fig. 5). Very recently, Howard and colleagues published a crystal structure of PRORP1 from *A. thaliana*[29]. It confirms our structural predictions, in particular the superhelical fold of the PPR domain made of 5–6 PPR and PPR-like elements. SAXS data collected in solution on PRORP1 and PRORP2 show a good agreement (experimental and theoretical curves fit with Chi of 4.9 and 2.8, respectively, for data in the range $0.02 < q < 0.2$ Å^{-1}) with the atomic model (PDB id: 4g26). This validates the overall PRORP architecture with two functional domains: a N-terminal RNA-binding PPR domain and a C-terminal PIN-like catalytic domain, bridged together by a bipartite zinc-binding module. The four residues identified by mutagenesis and inductively coupled plasma mass spectrometry as zinc binders are also confirmed by the crystal structure.

Both footprint data and activity assays indicated that the tRNA precursor is essentially recognized by its acceptor arm, whereas the anticodon domain is dispensable. Results suggest that PRORP substrate recognition might be mediated by a limited number of determinants. As PRORP is able to recognize any tRNA of canonical structure, these determinants should be found among highly conserved residues such as G18 in loop D, C56 and R57 in loop TψC[30], which is corroborated by our results. Considering the length of the acceptor arm (45 Å) and the estimation of PRORP dimensions in SAXS (30 Å × 70 Å × 110 Å), the PPR domain is very likely to interact specifically with the D-TψC region at the corner of the tRNA, while the NYN catalytic domain must be located in the vicinity of the 5′-cleavage point. Thus, the proposed model (Fig. 5) shows the two-domains of PRORP2 that sandwich the substrate, their respective position acting as a ruler to determine the correct position of maturation, independently

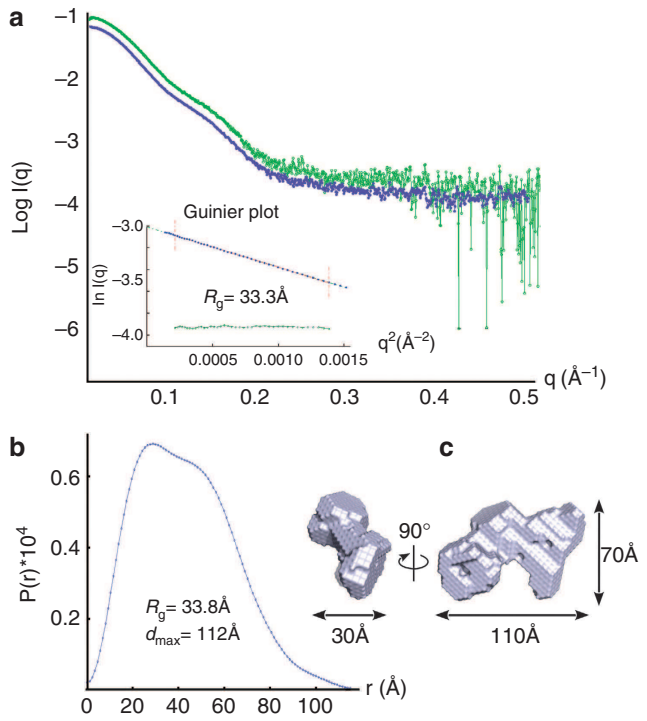


Figure 4 | SAXS analysis (**a**) PRORP1 and PRORP2 produce very similar intensity curves shown in green and blue, respectively. The inset Guinier plot indicates a gyration radius of 33.3 ± 0.1 Å. (**b**) This value is confirmed by the distance distribution function P(r), which also suggests that PRORP proteins are composed of two distinct domains and an extended tail (d_{max} > 100). (**c**) Molecular envelope of PRORP2 derived from SAXS data analysis.

Table 1 | Inductively coupled plasma mass spectrometry identifies zinc in association with PRORP.

PRORP1	Zn^{66} (p.p.b.)	2σ	Stoichiometry
WT	29.49	1.53	1
C344A	23.82	0.46	0.8
C347A	20.89	0.24	0.7
H548A	11.80	0.50	0.4
C565A	7.44	0.21	0.2
DD474-475AA	28.52	1.49	1
Buffer	0.44	0.12	0

Measurements were performed on wild-type PRORP1 (WT), as well as on proteins with point mutations applied to positions predicted to form the zinc-binding pocket. 2σ indicates the s.e. in four replicate measurements.

from the internal sequence of the acceptor arm. Our data suggests an intriguing similarity in the mode of binding of the tRNA with the RNP RNase P. Indeed, earlier work has shown that bacterial RNase P interaction with the D-TψC region influences substrate binding and cleavage[31]. In the same line and similar to PRORP, bacterial and human RNP RNase P did not require the anticodon domain of tRNA for substrate recognition[32]. However, *E. coli* RNase P, contrary to PRORP, still allowed RNase P activity on a tRNA lacking its D domain[32]. The mechanistic model of the novel protein-only RNase P represents a good basis for further investigations of PRORP mode of action by complementary approaches, the ultimate step being the determination of a crystal structure of an active complex of PRORP and tRNA at atomic resolution.

The concept of structural mimicry of nucleic acids by proteins is well established, it has already been observed over 15 years ago[33–35]. The specific case of PRORP is particularly interesting because both a single eukaryotic protein and a considerably more ancient bacterial ribozyme share the same catalytic function and appear to share similar RNA recognition processes. This implies that PRORP could represent an example of convergent evolution, with proteins that have evolved a mechanism of RNA recognition similar to that of catalytic RNA. This opens appealing perspectives for our understanding of the transition between the envisaged pre-biotic RNA world and the modern world dominated by proteins.

Methods

PRORP purification and characterization. *Arabidopsis* recombinant PRORP1 and PRORP2 proteins were expressed in *E. coli* and purified to homogeneity using affinity chromatography as described previously[9]. Before biophysical analyses (see below), a second step of size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) was introduced to improve the quality of PRORP enzymes and to elute them in appropriate buffers. Proteins were concentrated by ultrafiltration to about 10 mg ml^{-1}, ultracentrifuged and stored at 4 °C until use in 50 mM HEPES-Na pH 7.5, 250 mM NaCl, 15% glycerol (w/v), 1 mM *tris*(2-carboxyethyl)phosphin. Sample homogeneity and particle size were systematically verified using dynamic light scattering (Malvern Zetasizer) at 20 °C. Mass determination was performed by multi-angle light scattering using a SEC Superdex 200 column coupled to a Treos instrument (Wyatt technologies) in the storage buffer with 2% glycerol (w/v) only.

RNase P activity assays. cDNAs representing variants of *Arabidopsis* mitochondrial $tRNA^{Cys}$ precursors were designed with leader and trailer sequences of 50 and 30 nucleotides, respectively, cloned in pUC19, transcribed *in vitro* by T7 RNA polymerase. $tRNA^{Cys}$ precursor mutants included tRNAs with the anticodon domain removed (ΔAC), without both anticodon and D domain (ΔDAC), with point mutations at position 1, 18, 19, 56, 57 and 72. Sequences of oligonucleotide used to generate these mutants are available in Supplementary Table. For RNase P cleavage assays, reactions were always performed with three replicates using 0.5 µM transcript and 0.15 µM protein for 15 min at 25 °C as previously described[9]. RNA fragments were separated by denaturing PAGE and visualized by ethidium bromide staining. Quantifications were performed as described[10].

Footprinting analyses. Recombinant PRORP1 was put in presence of equimolar amounts of 5′-^{32}P-γATP radiolabeled mitochondrial $tRNA^{Cys}$ precursors to form a PRORP/tRNA complex. As PRORP and tRNA only interact in a transient manner, the complex obtained was UV-crosslinked for 15 min at 260 nm. Samples were submitted to partial RNase V1 (0.1 U µl^{-1}), RNase T1 (1 U µl^{-1}) and RNase A (1 µg µl^{-1}) digestions in the presence of competitor yeast RNA according to the manufacturer's instructions (Ambion, USA). The radiolabeled tRNA probe was also subjected to partial RNase T1 digestion in denaturing condition and to partial alkaline hydrolysis to generate RNA ladders. RNA samples were recovered by phenol/chloroform extractions, separated by high resolution 8% denaturing polyacrylamide gel electrophoresis and signal was acquired with a FLA-7000 phosphorimager (Fujifilm).

SRCD analysis. SRCD experiments were performed on the DISCO beamline at synchrotron SOLEIL (Saint-Aubin, France). The instrument was calibrated for magnitude and polarization with a 6.1-mg ml^{-1} D-10-camphorsulfonic acid solution. PRORP proteins (10 mg ml^{-1}) in 100 mM potassium phosphate, 50 mM KCl, 10% (v/v) glycerol and 1 mM *tris*(2-carboxyethyl)phosphine were placed in a SRCD CaF_2 cuvette of 8 µm pathlength. Three spectra between 170 and 280 nm were measured at 10, 20, 30, 40, 50, 60 and 70 °C to assess the thermal stability of PRORP1 and PRORP2. Data were processed (spectrum averaging, solvent base line subtraction) using CDtools[36]. The secondary structure content of PRORPs was evaluated using the VARSLC method in DICHROWEB[37].

Small-angle X-ray scattering analysis. Small-angle X-ray scattering (SAXS) experiments were conducted on the SWING beamline at Synchrotron SOLEIL, Saint-Aubin, France. The beam wavelength was set to $\lambda = 1.033$ Å. The 17 × 17 cm^2 low-noise Aviex CCD detector was positioned at a distance of 2107 mm from the sample, with the direct beam off-centred. The resulting exploitable q-range was 0.005–0.5 Å^{-1}, where $q = 4\pi \sin\theta/\lambda$, and 2θ is the scattering angle. PRORP samples at 10 mg ml^{-1} in 100 mM Hepes-Na (pH 7.5), 250 mM NaCl, 5% glycerol and 1 mM *tris*(2-carboxyethyl)phosphine were analysed by direct injection or high-performance liquid chromatography mode. In the first case, they were transferred into the SAXS flow-through capillary cell and a series of 50 frames was recorded. In the second case, they were loaded into a size exclusion column (Agilent Bio SEC-3,

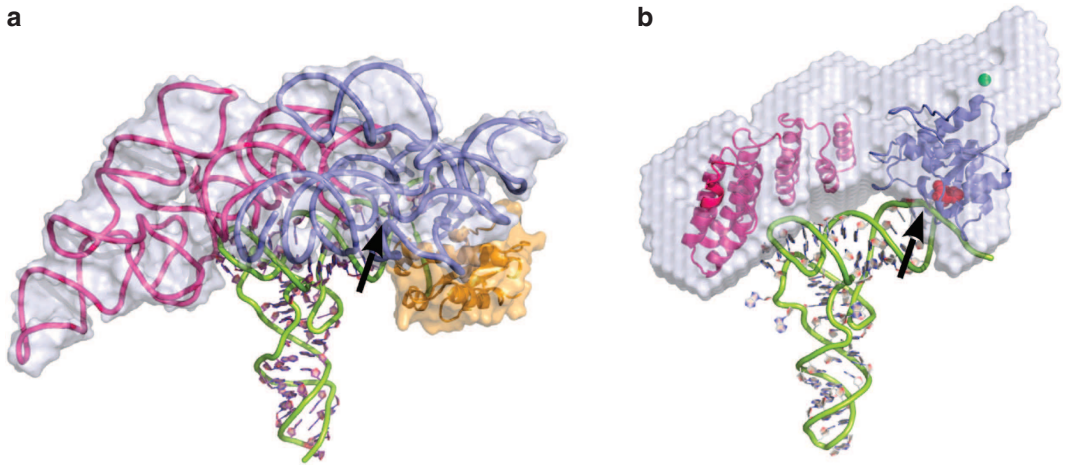


Figure 5 | A common mode of RNA binding in RNP and PRORP RNase P. (**a**) Structure of *Thermotoga maritima* ribozyme (PDBid 3Q1R[13]) with the catalytic domain in blue, the specificity domain in violet, the RNase P protein subunit in orange, the tRNA product in green and the molecular surface of the RNP in grey. (**b**) Model of *A. thaliana* PRORP2 protein shown in the same orientation and same colour code: the catalytic domain (NYN domain) in blue, with conserved Asp residues shown as red spheres adjacent to the 5′-cleavage site (indicated by an arrow), and the RNA-binding domain (PPR domain) in violet, interacting with the region of the D-TψC loops, were positioned in the SAXS envelope. The proposed zinc site is indicated by a green sphere in the central region, which connects the two main PRORP domains. This two-domains architecture offers a concave surface, which can be docked on the tRNA acceptor arm. Our data indicate that PRORP proteins have evolved an RNA recognition process very similar to that of RNP RNase P.

300 Å, 4.6 × 300 mm, 3 µm) using an Agilent high-performance liquid chromatography system and eluted into the SAXS flow-through capillary cell at a flow rate of 0.2 ml min^{-1}. SAXS measurements were collected throughout the whole protein elution time, with a frame duration of 1000 ms and a dead time between frames of 500 ms. Data processing, analysis and modelling steps were carried out with PRIMUS[38], and other programs of the ATSAS suite[39]. The radius of gyration R_g was derived from Guinier approximation[40] and calculated from entire scattering pattern using the indirect transform package GNOM[41], which provides the distance distribution function P(r) of the particle. Based on this distribution, *ab initio* modelling was carried out with DAMMIF[39]. A series of 11 dummy atom models was generated that were compared using the DAMAVER suite[42] to determine the most typical/probable one (that is, showing the lowest averaged normalized spatial discrepancy). The molecular envelope corresponding to this model was used to spatially restrain the positions of PRORP domains and of the tRNA substrate in the model.

Inductively coupled plasma mass spectrometry. For the (w/v) analysis of metal cofactors, PRORP solutions resuspended in 0.5 N nitric acid were analysed with a ThermoElectron X Series II inductively coupled plasma mass spectrometry mass spectrometer operated at 1450 W, with argon carrier gas flow rate of 0.85 l min^{-1}, argon auxiliary gas flow rate of 0.40 l min^{-1}, using a Meinarht quarz nebulizer, a quarz spray chamber with impact bead chilled to 3 °C and sample flow rate set to 0.1 l min^{-1}. Four replicate measurements were performed and values were corrected by an internal ^{115}In standard.

Structure modelling. The overall architecture of PRORP domains was predicted by homology modelling based on the alignment of 181 PRORP ortholog sequences[9] and fold recognition to find remotely related candidates with known structure as implemented on the Phyre2 server[24]. PRORP RNA partner (pre-tRNA^{Cys} from *A. thaliana*) was modelled using S2S[43] based on a sequence alignment with the tRNA^{Cys} from *E. coli* (PDB-id 1B23[44]). PRORP domain models were fit in the SAXS envelope and the tRNA substrate was docked on its concave surface in a way bringing the PPR and catalytic domains in close vicinity of the D-TψC corner and of the cleavage point, respectively. Molecular docking and related figures were performed with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0, Schrödinger, LLC).

References

1. Hartmann, E. & Hartmann, R. K. The enigma of ribonuclease P evolution. *Trends Genet.* **19,** 561–569 (2003).
2. Lai, L. B., Vioque, A., Kirsebom, L. A. & Gopalan, V. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett.* **584,** 287–296 (2010).
3. Jarrous, N. & Gopalan, V. Archaeal/eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic Acids Res.* **38,** 7885–7894 (2010).
4. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857 (1983).
5. Altman, S. A view of RNase P. *Mol. Biosyst.* **3,** 604–607 (2007).
6. Wang, M. J., Davis, N. W. & Gegenheimer, P. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J.* **7,** 1567–1574 (1988).
7. Rossmanith, W. & Karwan, R. M. Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. *Biochem. Biophys. Res. Commun.* **247,** 234–241 (1998).
8. Holzmann, J. *et al.* RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135,** 462–474 (2008).
9. Gobert, A. *et al.* A single Arabidopsis organellar protein has RNase P activity. *Nature Struct. & Molec. Biol.* **17,** 740–744 (2010).
10. Gutmann, B., Gobert, A. & Giegé, P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* **26,** 1022–1027 (2012).
11. Taschner, A. *et al.* Nuclear RNase P of *Trypanosoma brucei*: A single protein in place of the multicomponent RNA-protein complex. *Cell Reports* **2,** 19–25 (2012).
12. Westhof, E. & Altman, S. Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **91,** 5133–5137 (1994).
13. Reiter, N. J. *et al.* Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468,** 784–789 (2010).
14. Schmitz-Linneweber, C. & Small, I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13,** 663–670 (2008).
15. Anantharaman, V. & Aravind, L. The NYN domains: novel predicted RNAses with a PIN domain-like fold. *RNA Biol.* **3,** 18–27 (2006).
16. Thomas, B. C., Li, X. & Gegenheimer, P. Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. *RNA* **6,** 545–553 (2000).
17. Pavlova, L. V. *et al.* tRNA Processing by protein-only versus RNA-based RNase P: kinetic analysis reveals mechanistic differences. *Chembiochem.* (2012).
18. Giegé, R. *et al.* Structure of transfer RNAs: similarity and variability. *Wiley Interdiscip. Rev. RNA* **3,** 37–61 (2012).
19. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285 (2007).
20. Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 307–327 (2007).
21. Mertens, H. D. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–141 (2010).
22. Zwieb, C. *et al.* Structural modeling of RNase P RNA of the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Biochem. Biophys. Res. Commun.* **414,** 517–522 (2011).

23. Kazantsev, A. V. *et al.* Solution structure of RNase P RNA. *RNA* **17,** 1159–1171 (2011).
24. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4,** 363–371 (2009).
25. Lurin, C. *et al.* Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16,** 2089–2103 (2004).
26. Small, I. D. & Peeters, N. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25,** 46–47 (2000).
27. Blatch, G. L. & Lassle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21,** 932–939 (1999).
28. Xu, J. *et al.* Structural study of MCPIP1 N-terminal conserved domain reveals a PIN-like RNase. *Nucleic Acids Res.* **40,** 6957–6965 (2012).
29. Howard, M. J., Lim, W. H., Fierke, C. A. & Koutmos, M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl Acad. Sci. USA* **109,** 16149–16154 (2012).
30. Westhof, E. & Auffinger, P. *Transfer RNA structure eLS* (John Wiley & Sons, Ltd, 2012).
31. Brannvall, M., Kikovska, E., Wu, S. & Kirsebom, L. A. Evidence for induced fit in bacterial RNase P RNA-mediated cleavage. *J. Mol. Biol.* **372,** 1149–1164 (2007).
32. Hardt, W. D., Schlegl, J., Erdmann, V. A. & Hartmann, R. K. Role of the D arm and the anticodon arm in tRNA recognition by eubacterial and eukaryotic RNase P enzymes. *Biochemistry* **32,** 13046–13053 (1993).
33. Putnam, C. D. *et al.* Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with *Escherichia coli* uracil-DNA glycosylase. *J. Mol. Biol.* **287,** 331–346 (1999).
34. Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82,** 701–708 (1995).
35. Nissen, P., Kjeldgaard, M. & Nyborg, J. Macromolecular mimicry. *EMBO J.* **19,** 489–495 (2000).
36. Lees, J. G., Smith, B. R., Wien, F., Miles, A. J. & Wallace, B. A. CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal. Biochem.* **332,** 285–289 (2004).
37. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32** (2004).
38. Konarev, P., Volkov, V., Sokolova, A., Koch, M. & Svergun, D. PRIMUS : a windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277–1282 (2003).
39. Konarev, P., Petoukhov, M., Volkov, V. & Svergun, D. ATSAS 2.1 : a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39,** 277–286 (2006).
40. Guinier, A. La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys. (Paris)* **12,** 161–237 (1939).
41. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25,** 495–503 (1992).
42. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab-initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **36,** 860–864 (2003).
43. Jossinet, F. & Westhof, E. Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21,** 3320–3321 (2005).
44. Nissen, P., Thirup, S., Kjeldgaard, M. & Nyborg, J. The crystal structure of Cys-tRNA(Cys)-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. *Struct. Fold Des.* **7,** 143–156 (1999).

Acknowledgements

We thank Bernard Lorber for critical reading of the manuscript and assistance in dynamic light scattering experiments, as well as Isabelle Billas Massobrio for multi-angle light scattering analysis. We thank Elodie Ubrig and Laurie-Anne Roeckel for technical assistance and Pierre Fechter for advice with footprinting experiments. We also acknowledge the staff of SOLEIL synchrotron (Saint-Aubin, France) for the beamtime allocated to the project, and more particularly Frank Wien for assistance during SRCD data collection on DISCO. This work was supported by the French 'Centre National de la Recherche Scientifique' and by the University of Strasbourg. AG and OF were supported by an ANR Blanc research grant 'PRO-RNase P, ANR 11 BSV8 008 01' to PG and CS and by the LabEx consortium 'MitoCross'. FP and BG were supported by PhD grants from the University of Strasbourg.

Author contributions

A.G., C.S. and P.G. designed and coordinated the experiments. CS directed the biophysical characterization of PRORP and PG directed the biochemical study of the PRORP/tRNA complex. A.G., F.P., O.F., B.G., R.B., P.R., C.S. and P.G. performed the experiments and analysed the results. A.G., F.P., C.S. and P.G. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Gobert, A. *et al.* Structural insights into protein-only RNase P complexed with tRNA. *Nat. Commun.* 4:1353 doi: 10.1038/ncomms2358 (2013).

5 Structure determination of wild type PRORP2

I initially started this structural work on the three *A. thaliana* PRORPs to maximize the chance of success to crystallize one of them. After the publication of the PRORP1 crystal structure ((Howard *et al.*, 2012) I essentially focussed on PRORP2 and its variants to obtain insights into the nuclear form of PRORP.

5.1 Crystallization and crystal analysis of PRORP proteins

In order to obtain accurate 3D structural data with X-ray crystallography proteins have to be crystallized. As at present no tools are available that predict crystallization conditions, commercial screens are usually tested to search for initial conditions that can be further optimized (see section 2.2.9). To set up one screen with drop sizes of 300 nl (150 nl of that are protein solution) 40 to 200 μg of protein at 2 to 10 mg/ml are required. I tested nine different screens at different temperatures, protein constructs and mixing ratios which explains the need of rapid and simple protein purification (Fig. 5.1).
I observed that PRORPs easily precipitate at concentrations higher than 5 mg/ml in most of the screens at room temperature, making working in the cold room a necessity for crystallization, crystal observation and harvesting for cryocooling.

5.1.1 PRORP initial screening

Screens tested for wild type PRORP2-His_6 are summarized in Tab. 5.1. Prior to crystallization protein samples were ultra-centrifuged (1h, 125.000g, 4°C) to remove aggregates. Conformational purity and activity were assayed with DLS and cleavage tests except in the case of the catalytic mutants for which no activity tests were done. Sitting nanodrops of 300 nl were placed with 35 μl reservoir solution in a sealed vapor chamber. 150 nl protein solution (2 - 15 mg/ml) were mixed with 150 nl of the reservoir solution using a pipetting robot (Mosquito^{TM} pipetting robot, TTP LabTech). Because PRORP2 and 3 are highly homologous we tested different screens on the two proteins assuming that one hit for one of the homologue would suit the other. All tested screens are summarized in Tab. 5.1. No other hits than for wild type PRORP2 were obtained as detailed in the next section.

Table 5.1: Summary of initial crystallization screening of PRORP. Screen leading to interesting

Protein construction	**Concentration [mg/ml]**	**Temperature [°C]**	**Screen name/manufacturer**	**Buffer**
P2-His$_6$wt	7.7, 6, 3.4	20	JCSG+ (Molecular Dimensions)	SEC1
	11.4, 7.4	20	Morpheus (Molecular Dimensions)	
	10	20	Index (Hampton Research)	
	6	4, 20	**Magic1**, Magic2	
	3	20	Clear Strategy Screen 1 (Molecular Dimensions), Wizard (EmeraldBio)	
	14, 7, 2.5	4, 20	**PEG/Ion** (Hampton Research)	SEC1 with 15
	8, 6, 4	4	JCSG+ (Molecular Dimensions), Morpheus (Molecular Dimensions)	SEC1 with 5 %
	6, 2.25	4, 20	PEG/Ion (Hampton Research), JCSG+ (Molecular Dimensions)	
P2mDD-His$_6$	8, 6, 4	4	JCSG+ (Molecular Dimensions), Morpheus (Molecular Dimensions)	SEC1 with 5 %
	3	20	Clear Strategy Screen 1 (Molecular Dimensions), Wizard (EmeraldBio)	
	5, 3	4, 20	Magic1, Magic2	
	4	4	PEG/Ion (Hampton Research	a) 50 mM Hep glycerol, b)50 5 % (w/v) gly (NH_4SO_2, 5 %
P1-His$_6$wt cris	2.8	4, 20	Magic1, Magic2, JCSG+ (Molecular Dimensions)	
P2-intein-wt	4, 2, 1	20	PEG/Ion (Hampton Research), JCSG+ (Molecular Dimensions)	
P3-His$_6$	15, 7.5	20	Morpheus (Molecular Dimensions)	
	10	20	Magic1, Magic2	
	3, 2	4, 20	PEG/Ion (Hampton Research), Clear Strategy Screen 1 & 2 (Molecular Dimensions), SaltX (Hampton Research), Index (Hampton Research)	SEC1, SEC2 +

5.1.2 Wild type PRORP2 crystal optimization

An initial hit in the Magic1 screen

An initial hit was obtained at 20 °C in a condition containing 50 mM MES pH 6, 200 mM sodium malonate, 34 % PEG 8000 (D2 in Magic1). We ordered a grid-matrix screen from the Magic1 screen around condition D2 where I obtained hits in conditions B4 and C4 (Fig. 5.1a). Further optimization failed because we could never figure out the actual PEG concentrations, probably due to a failure in the production of the grid-screen. As a result I was not able to reproduce identical solutions and crystals.

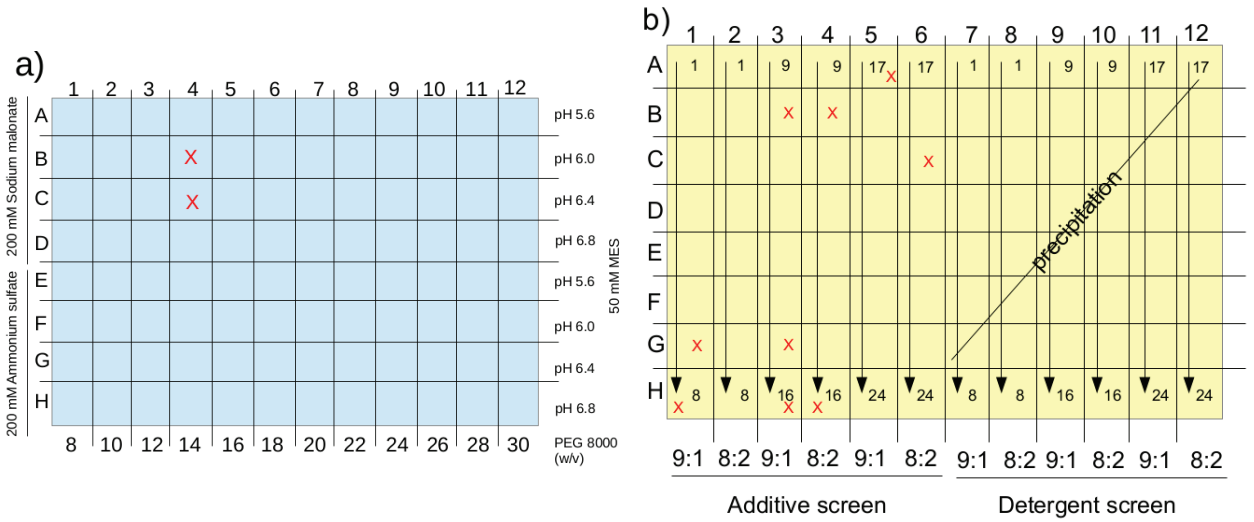


Figure 5.1: Optimization screens for wild type PRORP2. a) Grid-screen matrix for D2, screen Magic1 (MPI, Martinsried, Germany), b) Optimisation of PEG/Ion, E6 with the additive and the detergent screen (Hampton Research). Red crosses mark crystal or crystal-like structures.

PEG/Ion screening hits

In the PEG/Ion screen from Hampton Research several conditions were identified and optimized. During optimization protein concentration, salt and PEG concentration were varied, additives (salts, small organic molecules) or cations were tested. A condition containing 200 mM sodium-malonate pH 6, 20 % PEG3350 was further used to optimize and test additives and detergents. Stock solutions of this condition were prepared that were 1.1x and 1.2x more concentrated than the original solution. Ratios of 9:1 and 8:2 were prepared with the additive and the detergent screen (Fig. 5.1b). Wild type PRORP2 was used at 1 and 4 mg/ml and equilibration took place at 4 °C.
It turned out that best crystallization conditions for wild type PRORP2 are 200 mM Sodium-malonate pH 6, 20 % PEG3350 with drops set up in batch by mixing 1 μl 2.5 mg/ml PRORP2 with 1 μl of crystallant at 4 °C (Fig. 5.2).

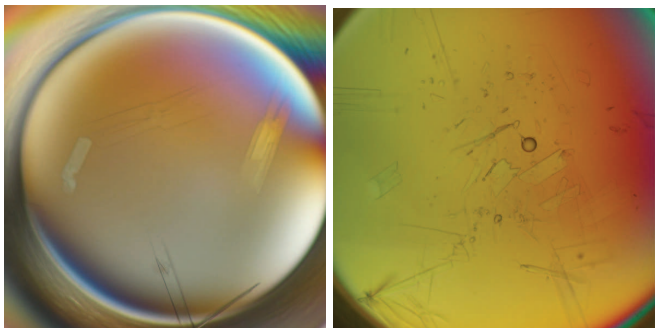


Figure 5.2: Wild type PRORP2 crystals after optimization.

5.1.3 Crystallization screens for PRORP/pre-tRNA complex

Several transcripts were used for co-crystallization assays with PRORP2. In particular, I designed a L5T0 tRNA$^{Cys-GUC}$ substrate which has the body sequence of the habitual L5T0 tRNA^{Cys} with the anticodon mutated to GUC. This creates complementary base pairing between two tRNA molecules and could help crystallization via mediating crystal contacts as found in the crystal packing of yeast tRNA^{Asp} (Ruff *et al.*, 1991).
PRORP2mDD and L5T0 tRNA$^{Cys-GUC}$ and L5T0 MAC tRNA^{Cys} were dialized in SEC2 buffer with concentrations after ultracentrifugation (1h, 4 °C, 125.000 g) of 78 μM, 184 μM and 122 μM, respectively. Complexes were constituted by mixing equal volumes of PRORP2mDD with each substrate resulting in molar ratios of L5T0 tRNA$^{Cys-GUC}$:PRORP2mDD 2.3:1 and L5T0 MAC tRNA^{Cys}:PRORP2mDD 1.5:1. Hanging nanodrops were made of 150 nl of sample solution and 150 nl of reservoir solution. Tested screens were: Natrix and Index (Hampton Research) at 4 °C. Prior DLS measurements of the complex mixtures showed no aggregation.
In another experiment, the PRORP2mDD/L5T0 tRNA^{Cys} complex after ITC was concentrated on 10 K membranes and the absence of aggregates verified with DLS. Index (Hampton Research), JCSG+ (Molecular Dimensions) and Crystal(Hampton Research) Screen were set up at 4 and 20 °C. Neither of the screens showed any hit.

5.2 X-ray crystallographic data are difficult to interpret

Two full datasets of wild type PRORP2 crystals were collected at a resolution of 3 Åand data was processed using XDS and the CCP4 suite (Kabsch, 2010, Evans, 2011). A rotation function analysis using GLRF between 4 and 8 Åconfirmed the presence of two molecules in the asymmetric unit but molecular replacement was particularly tricky and placing the second monomer failed. We think that placement is difficult due to the high amount of α-helices in the protein which makes the definition of one PPR repeat rather difficult at that medium resolution.

5.3 Publication 3: Crystallization of nuclear proteinaceous RNase P 2 from Arabidopis thaliana

Crystallization of nuclear proteinaceous RNase P 2 from *Arabidopis thaliana*

Franziska Pinker[ab]*, Claude Sauter[a*] and Philippe Giegé[b]

[a]UPR9002 du CNRS, Centre national de la recherche scientifique, 15, rue Rene Descarte, Strasbourg, 67084, France

[b]UPR2357 du CNRS, Centre national de la recherche scientifique, 12, rue du général Zimmer, Strasbourg, 67084, France

Correspondence email: c.sauter@ibmc-cnrs.unistra.fr

ABSTRACT

The RNase P activity is ubiquitous and consists of the 5' end maturation of precursor tRNAs. For a long time it has been thought that all RNase P were ribonucleoproteic enzymes. However, the characterization of the human mitochondrial RNase P revealed a novel kind of RNase P composed of proteins only, called PRORP for 'Proteinaceous RNase P' [1]. Whereas in human mitochondria PRORP is formed by a complex of three subunits, RNase P activity in *Arabidopsis thaliana* is held by single proteins. Three homologues exist in *Arabidopsis*. Each protein possesses a pentatricopeptide repeat domain and a metallonuclease domain [2]. Here we a report the purification and crystallization of *Arabidopsis* PRORP proteins. Optimization of initial crystallization conditions led to crystals diffracting at 3 Å.

INTRODUCTION

RNase P is the universal activity catalyzing the maturation of precursor 5'-ends of transfer RNAs (tRNAs), as well as many other substrates such as ribosomal RNA, messenger RNA, transfer-messenger RNA or riboswitches [3–5]. It was first described in bacteria as being a ribonucleoprotein (RNP) complex whose RNA component (P RNA) holds the catalytic activity [6]. RNP RNase P is prevalent in all three domains of life, i.e. bacteria, archaea and eukarya, and was believed to be a conserved RNP complex [7]. Still, early experiments in spinach chloroplasts and later in human mitochondria suggested a proteinaceous type of RNase P in these organelles lacking an essential RNA component. Rossmanith and co-workers demonstrated by recombinant expression of human mitochondrial RNase P in *E.coli* that it is a complex of three protein subunits and no RNA [8,9] . In *Arabidobis* this novel variant is composed of a single protein that was named PRORP (for

PROteinaceous RNase P) [2].

The discovery of PRORP enzymes led to the question whether there is a difference in the mode of action of RNP and protein enzymes catalyzing the same reaction. RNP RNase P activity is well characterized [2,10] and the determination of the X-ray crystallographic structure of a bacterial RNase P in complex with tRNA have been very important advances [11]. Protein-only RNase P enzymes comprise two distinct domains: N-terminal pentatricopeptide (PPR) repeats and a C-terminal catalytic metallonuclease domain. [1,12] The PPR domain confers to RNA binding and substrate specificity [10]. A PPR repeat is composed of modules of 34-36 amino acids folding into a pair of antiparallel α-helices [10]. The most conserved part of PRORP enzymes is in their C-terminal metallonuclease domain which makes PRORPs being a member of the large family of PIN-like/NYN (N4BP1, YacP-like Nuclease) ribonucleases [13]. For a long time high-resolution structural data of PPR proteins was scarce. The first structural data on PPR proteins became available in 2011 by Temiakov and co-workers from the PPR containing human mitochondrial RNA polymerase (pdb ID: 3SPA). The two PPR motifs in the N-terminal region of this enzyme confirmed the predicted fold of two antiparallel α-helices per PPR motif [14]. Quickly, more structures emerged. One of them was the X-ray structure of organellar PRORP1 from *Arabidopsis thaliana* [15] and soon afterwards complex structures between the plant transcription factor PPR10 and its target mRNA, PSAJ [16]. Still, the biological significance of the latter complex remains to be validated.

In order to obtain further insights into the structural organization of PRORP from *Arabidopsis* different crystallization screens and buffers were tested on wild type PRORP2, crystallization conditions optimized and diffraction data collected.

EXPERIMENTAL SECTION

Protein preparation

The open reading frame of *Arabidopsis* PRORP2 (NCBI reference sequence NP_179256) was cloned into a pET-28b(+) vector (Novagen) coding for a C-terminal His_6 tag. Due to cloning one additional N-terminal amino acid (MG instead of M) and two additional amino acids upstream the His-tag (LE) were introduced resulting in a protein of 537 aa.

E.coli BL21 (DE3) transformed with pET28-PRORP2 were grown at 37 °C until $OD_{600\ nm}$ in LB medium containing 50 µg/ml kanamycin. Proteins were expressed over night at 18 °C in LB medium with a final concentration of 0.5 mM IPTG. *E.coli* cells were harvested and the pellet stored at -20 °C

until purification. One gram of cells was lysed in 10 ml of NiNTA buffer A, containing 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 5 % (w/v) glycerol, 5 mM imidazole, 10 mM β-mercaptoethanol and protease inhibitor cocktail (Sigma, S8830, 1 tablet/100 ml) with ten cycles of 30 seconds of sonication (followed by 60 seconds on ice). The NiNTA resin (Sigma, No. H0537) was equilibrated with NiNTA buffer A. The protein was loaded at 1 ml/min and sequentially washed with 10 cv at 2 ml/min with increasing imidazole concentrations (steps at 5, 10 and 15 mM). The protein was eluted with 2.5 cv in 250 mM imidazole at 2 ml/min. The peak fractions were subjected to SDS-PAGE analysis. Protein containing fractions were pooled, concentrated and the buffer was exchanged for SEC1 buffer containing 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol, 1 mM TCEP on a cellulose membrane with a 30K MWCO (Amicon Ultra, Millipore). The concentrated fractions from affinity chromatography were then purified using size exclusion chromatography on a Superdex 200 10/300 GL (GE Healthcare) column at 0.5 ml/min equilibrated with SEC1 buffer. Proteins were concentrated, ultracentrifuged and stored at 4 °C until use. Sample purity and homogeneity was assessed by Coomassie-stained SDS-PAGE (Fig. 1a) and dynamic light scattering at 20 °C (Fig. 1b) using a Zetasizer NanoSeries Nano-S instrument (Malvern).

.

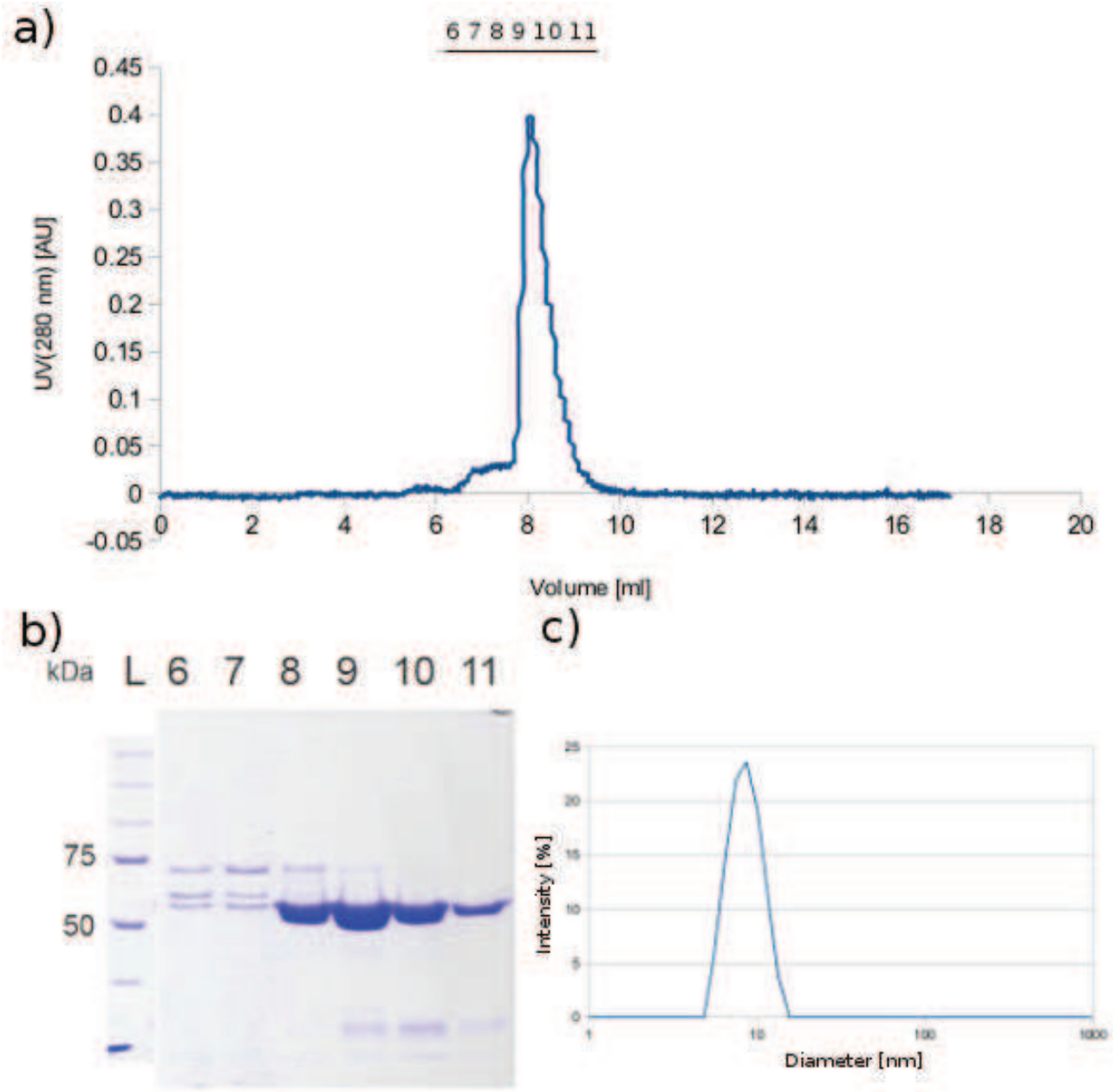


Figure 1: PRORP2 purification. a) Elution profile of wild type PRORP2 during size exclusion chromatography, b) Coomassie-stained 8 % SDS-PAGE of wild type PRORP2. L – molecular weight ladder, 6 – 11 fractions of SEC. c) DLS measurement of purified and ultracentrifuged wild type PRORP2.

Crystallization screening

Initial screening was performed with wild type PRORP2 at 5 – 10 mg/ml in SEC1 buffer. This concentration turned out to be excessive as indicated by many precipitated drops and further screening was done with 2 – 5 mg/ml in a buffer containing only 5 % (w/v) glycerol instead of 15 % (w/v). The screening was carried out by vapor diffusion using a Mosquito crystal pipetting robot (TTP Labtech, UK) and 96-well sitting drop plates (CrystalEx microplate, conical flat bottom, 5 sub-wells, Corning). Different screens were tested by mixing 150 nl of protein solution with 150 nl of crystallant solution and equilibrated at 4 °C and 20 °C over a reservoir containing 35 µl of crystallant solution. Tested screens were: JCSG+, Clear Strategy Screen 1, Morpheus (Molecular Dimensions), Magic1 and Magic2 (curitosy of Jérôme Basquin (MPI, Martinsried, Germany)), Peg/Ion (Hampton Research) and Wizard (EmeraldBio).

Crystallization optimization

The experiments for optimization were set up manually as microbatch experiments in Terazaki microplates (Greiner BioOne, No. 653 102) or hanging nanodrops in 24 well Linbro plates (Hampton Research) by mixing the protein and crystallant solutions in ratios 1:1 or 2:1 with a total volume of 2 µl and 3 µl, respectively. During the optimization various crystallant and protein concentrations, as well as additives and different temperatures (4 °C and 20 °C) were tested. The composition of the protein crystals was assessed on SDS-PAGE and mass spectrometry extensively washed in mother liquor to remove excess mother liquor.

Diffraction data collection and processing

For the diffraction experiments, crystals were mounted in nylon loops with a diameter of 200 µM (Hampton Research, USA), flash-frozen and stored in liquid nitrogen. Three diffraction data sets were collected at 100 K with a wavelength of 1 Å and 2.07 Å (to collect the anomalous signal of zinc) and a Pilatus 2M detector with a crystal-to-detector distance of 300 mm at the PXIII beamline at the synchrotron SLS (Villigen, Swizerland). Crystal data collection statistics are summarized in Tab. 1. All diffraction datasets were processed using the XDS package [17].

RESULTS AND DISCUSSION

The full length wild type PRORP2 was expressed in *E.coli* (referred to as P2wt) and purified to homogeneity using affinity and size exclusion chromatography (Fig. 1). Recombinant P2wt contains 528 residues of its native coding sequence and one N-terminal amino acid (G) and eight C-terminal amino acids (LEHHHHHH) that were introduced during cloning and for affinity purification. The initial crystallization screening was carried out with P2wt in the SEC1 buffer. As we obtained a majority of precipitated drops and no crystal-like hits during screening we changed the protein storage buffer and reduced the glycerol concentration from 15 % to 5 % to favor crystallization. Furthermore we set up all screens at two temperatures: 4 °C and 20 °C. Plate-like crystals (like in Fig. 3) were obtained after several days at 4 °C in SEC1 with 5 % glycerol in 13 out of 96 conditions in the PEG/Ion screen (Hampton Research). Condition E6 containing 200 mM sodium-malonate pH 6 and 20 % PEG 3350 from the PEG/ion Screen was chosen for further optimization.

In the optimization experiments, a wide range of approaches was used to lower the number of crystals and to increase crystal dimensions: (1) protein concentration was varied (2 – 10 mg/ml) as well as (2) PEG concentration (16 – 25 % PEG3350); (3) salt concentration (100 – 200 mM) and (4) various additives were added using the additive screen (Hampton Research); (5) using the batch crystallization method. This yielded in crystals of the same shape with average dimensions of 200 X 100 X 10 µm (Fig. 3) and a diffraction limit of 3 Å resolution. Despite all optimization efforts little of the initial hits could be improved except that crystals could be set up with batch method by mixing 1 µl of protein solution (2.5 mg/ml) with 1 µl of crystallant solution. Initial crystals diffracted to 7 Å but data could not be integrated using XDS. These crystals belonged to space group C2. Two complete data sets at 3 Å were collected from the optimized crystals that belong to space group P1 (Fig. 3). The calculated Mathews coefficient as well as a self-rotation function analysis using GLRF suggested the presence of two molecules in the asymmetric unit and a 2-fold symmetry [18,19]. Molecular replacement with the search model derived from the crystal structure of *A. thaliana* PRORP1 (pdb ID: 4g23) is in progress.

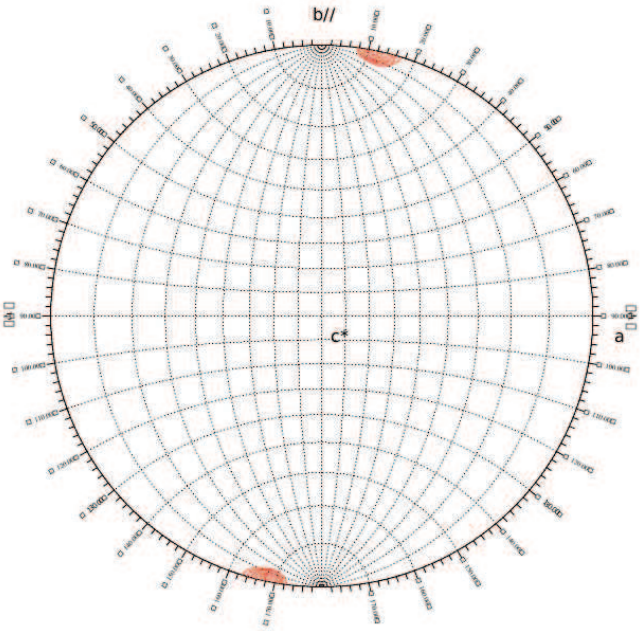


Fig. 2: Self-rotation function determined with GLRF. Red lines define the intensity level of the 2-fold rotation axis.

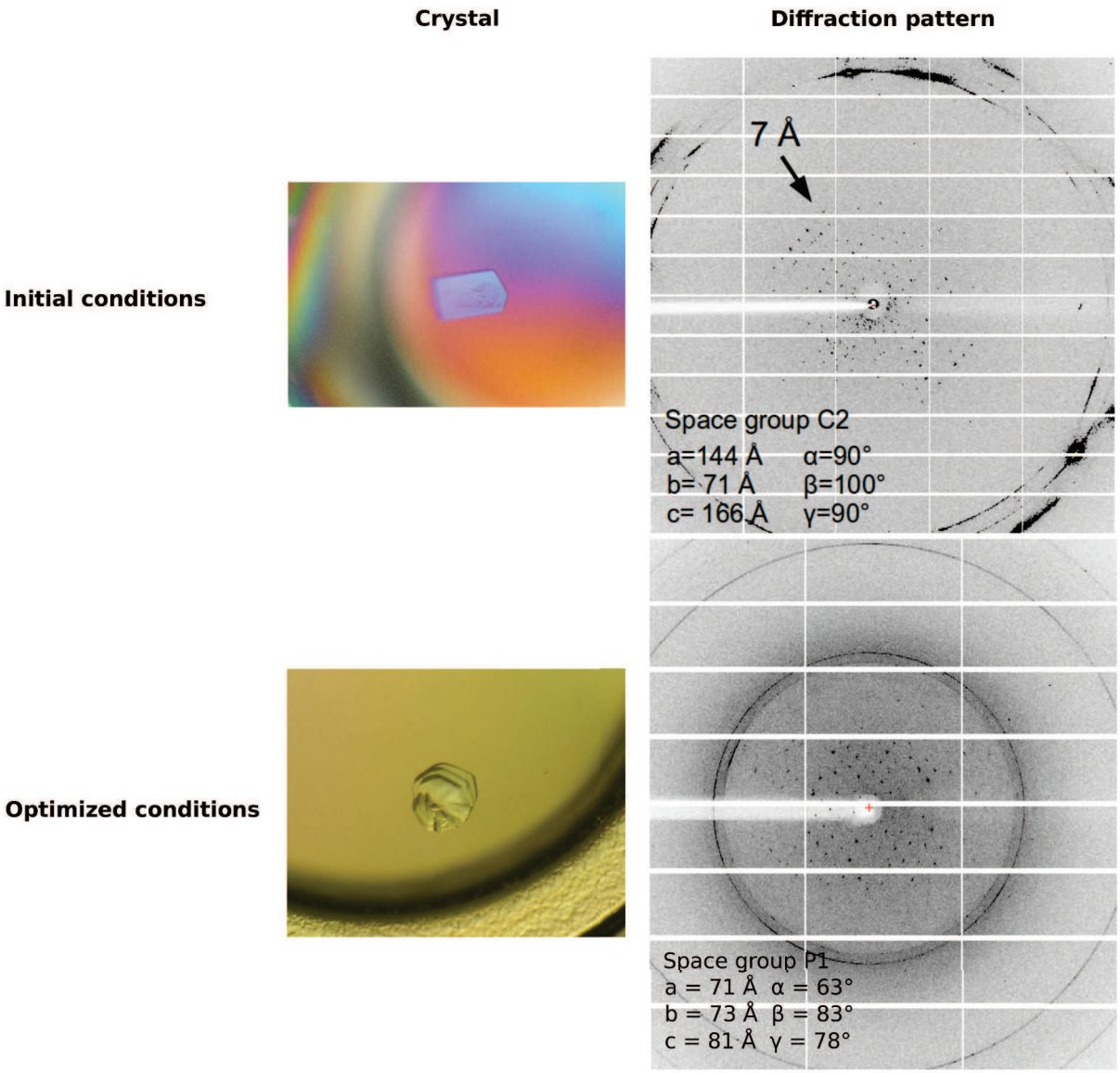


Fig. 3: Crystals and corresponding X-ray diffraction images of wild type PRORP2 in 200 mM sodium-malonate pH 6 and 20 % PEG3350.

Table 1. Data collection statistics for wild type PRORP2 crystals

Crystal	PRORP2 wild type 1	PRORP2 wild type 2	PRORP2 wild type 2_6keV
Data collection statistics			
Beamline	SLS/X06DA	SLS/X06DA	SLS/X06DA
Wavelength (Å)	1.0	1.0	2.07
Detector	Pilatus 2M	Pilatus 2M	Pilatus 2M
Number of images	1440	1440	1440
Space group	P1	P1	P1
Unit cell parameters a, b, c (Å)	70.6 72.9 80.5	68.1 72.6 80.2	68.1 72.6 80.4
Unit cell parameters α, β, γ (°)	63 83 78.4	62.2 83.4 77.3	63.3 83.3 77.3
Resolution range (Å)	50 – 3 (3.19 – 3.0)	50 – 3 (3.19 – 3.0)	50 – 4.48 (4.74 – 4.48)
No. of unique reflections	22575 (4198)	22823 (4031)	14672 (2222)
Completeness (%)	83.1 (95.6)	87.7 (96.1)	93.1 (88.6)
Multiplicity	3.5 (3.6)	3.2 (3.3)	1.8 (1.8)
R_{merge} (%)†	13.5 (216.2)	16.0 (240.3)	8.3 (24.5)
R_{meas} (%)†	16.0 (253.9)	19.3 (287.2)	11.8 (34.6)
CC(1/2) [x]	99.5 (34.2)	99.5 (16.9)	99.4 (88.5)
<I/sigma(I)>	7.8 (0.7)	6.3 (0.5)	6.69 (3.04)
Wilson B factor (Å^2)	88.2	92.6	128.6
no. of molecules in asymmetric unit	2	2	2
solvent content (%)	58	56	56

†$R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$ and redundancy-independent $R_{meas} = ©_{hkl} (n/n-1)^{1/2} ©_i |I_i(hkl) - \langle I(hkl)\rangle| / ©_{hkl} ©_i I_i(hkl)$.

[x]CC(1/2) = percentage of correlation between intensities from random half-datasets (Karplus & Diederichs (2012), Science 336, 1030-33)

[xx] Resolution limited since minimal detector distance possible in *is*Xds mode of beamline X10SA reached

Values in parentheses are for the highest resolution shell.

AUTHOR INFORMATION

Corresponding Author

*Tel: +33 (0)388 417 102; email: c.sauter@ibmc-cnrs.unistra.fr

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This work was supported by the French “Centre National de la Recherche Scientifique”, by the University of Strasbourg, an ANR Blanc research grant ‘PRO-RNase P, ANR 11 BSV8 008 01’ to PG and CS and by the LabEx consortium ‘MitoCross’. FP was supported by a PhD grant from the University of Strasbourg.

ABBREVIATIONS

aa - amino acids; cv - column volume; TECP - Tris(2-carboxyethyl)phosphine hydrochloride;

REFERENCES

1. Holzmann, J. *et al.* RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. *Cell* **135,** 462 –474 (2008).
2. Gobert, A. *et al.* A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol* **17,** 740 –744 (2010).
3. Hartmann, E. & Hartmann, R. K. The enigma of ribonuclease P evolution. *Trends Genet.* **19,** 561–569 (2003).
4. Lai, L. B., Vioque, A., Kirsebom, L. A. & Gopalan, V. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett.* **584,** 287–296 (2010).
5. Jarrous, N. & Gopalan, V. Archaeal/Eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic Acids Res.* **38,** 7885 –7894 (2010).
6. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857 (1983).
7. Altman, S. A view of RNase P. *Mol. Biosyst.* **3,** 604 (2007).
8. Wang, M. J., Davis, N. W. & Gegenheimer, P. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J* **7,** 1567–74 (1988).
9. Rossmanith, W. & Karwan, R. M. Characterization of Human Mitochondrial RNase P: Novel Aspects in tRNA Processing,. *Biochem. Biophys. Res. Commun.* **247,** 234–241 (1998).
10. Barkan, A. & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annu. Rev. Plant Biol.* **65,** null (2014).
11. Reiter, N. J. *et al.* Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468,** 784 –789 (2010).
12. Schmitz-Linneweber, C. & Small, I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13,** 663 –670 (2008).
13. Anantharaman, V. & Aravind, L. The NYN Domains - Novel Predicted RNases with a PIN Domain-Like Fold. *RNA Biol.* **3,** 18–27 (2006).
14. Ringel, R. *et al.* Structure of human mitochondrial RNA polymerase. *Nature* **478,** 269–273 (2011).
15. Howard, M. J., Lim, W. H., Fierke, C. A. & Koutmos, M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5’ processing. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16149–16154 (2012).
16. Yin, P. *et al.* Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **advance online publication,** (2013).
17. Kabsch, W. Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* **21,** 67–72 (1988).
18. Rossmann, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15,** 24–31 (1962).
19. Tong, L. & Rossmann, M. G. in *Methods Enzymol.* (ed. Charles W. Carter, J.) **Volume 276,** 594–611 (Academic Press, 1997).

6 Towards a stable complex of PRORP and a precursor tRNA

My ultimate goal was the determination of the 3D structure of PRORP in complex with a precursor tRNA. In order to study and to crystallize this complex I had to determine a key parameter which is the affinity constant. To stay as close as possible to the conditions compatible with structural biology experiments, I chose ITC among available methods. The first reason was that the instrument was available in the institute and the second that the complex formed during ITC experiments could be used directly for crystallization. I also used complementary methods: MST and AUC (in collaboration with C. Birck, FRISBI platform, IGBMC, Illkirch). An advantage of all these methods is that they are almost insensitive to buffer conditions or temperatures. Another crucial point was to find conditions where both partners, ideally in a complex are stable to be crystallized or to be analyzed by SAXS.

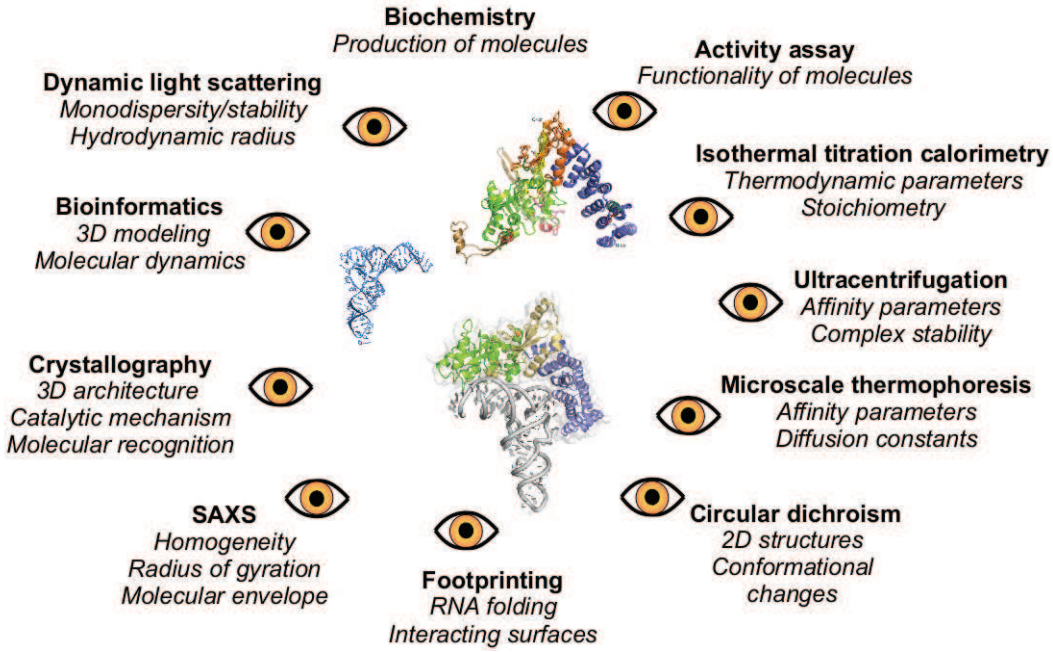


Figure 6.1: Biochemical and biological approaches used and during my thesis and information that could be obtained with each method.

6.1 Optimizing PRORP activity

Activity assays were performed to test the integrity of both wild type enzymes and purified tRNA substrates. Especially after storing the protein for a long period in the fridge ($\geq$4 weeks) an activity assay is a rapid way to check protein quality, in addition to a SDS-PAGE gel or DLS measurement.

These tests were carried out without using radioactivity. As described in detail in section 2.2.3.2 PRORPs were incubated for 20 min with a pre-tRNA substrate and reactions were stopped by adding guanidinium hydrochloride which denatures the protein. RNA were extracted and separated on a 8 % PPA/8M Urea gel (Fig. 6.2).

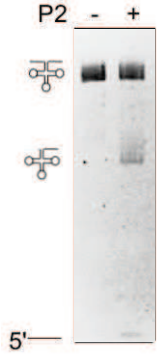


Figure 6.2: Activity assay of wild type PRORP2 (P2) with L51T30 $tRNA^{Cys}$ separated on a 8 % PAA/8M Urea gel, stained with ethidium bromide.

Initially, all routine activity tests were carried out in a buffer with a final concentration of 20 mM Tris-Cl pH 8, 40 mM NaCl, 4.5 mM $MgCl_2$, 20 μg/ml BSA and 2 mM DTT (= MTP buffer). To see the leader sequence on a 8 % PAA/8M Urea gel a tRNA substrate with a 51 nt long leader was chosen. The proteins were purified in a final buffer of 50 mM Hepes-Na pH 7.5, 250 mM NaCl, 15 % (w/v) glycerol and 1 mM TCEP for better solubility and long-term stability. However, in the first ITC experiments no interaction between the substrate and the enzyme was detectable. It turned out that wild type PRORP2 was not active in its purification buffer (Fig. 6.3a). Furthermore, subsequent DLS measurements revealed the presence of aggregates with a diameter greater than 1 μm.
To find a compromise between stability and activity, 13 buffer conditions with increasing salt concentrations were prepared ranging from 59 to 260 mM NaCl in 30 mM Hepes-Na pH 7.5 and 5 % (w/v) glycerol (Fig. 6.3b). A mastermix (MM) containing 22 mM $MgCl_2$, 11 μM DTT, 0.2 mg/ml L5T0 $tRNA^{Cys}$, 0.1 mg/ml BSA and 1.7 U/μl RNase OUT (inhibits RNase A-C, Invitrogen, No. 10777-019, 40 U/μl) was prepared. One reaction volume contained 2.3 μl MM, 6.7 μl of the respective buffer and 1 μl PRORP2wt at 0.1 mg/ml in SEC1 buffer. Reactions and analysis were done as described for the activity assay. Fig. 6.3b shows that protein activity decreases with increasing salt concentrations. These tests showed that the best compromise between PRORP stability and activity is obtained with a buffer containing a maximum 150 mM NaCl. Still, for storage the protein showed best stability in its purification buffer and was extemporaneously dialysed into its appropriate activity buffer.

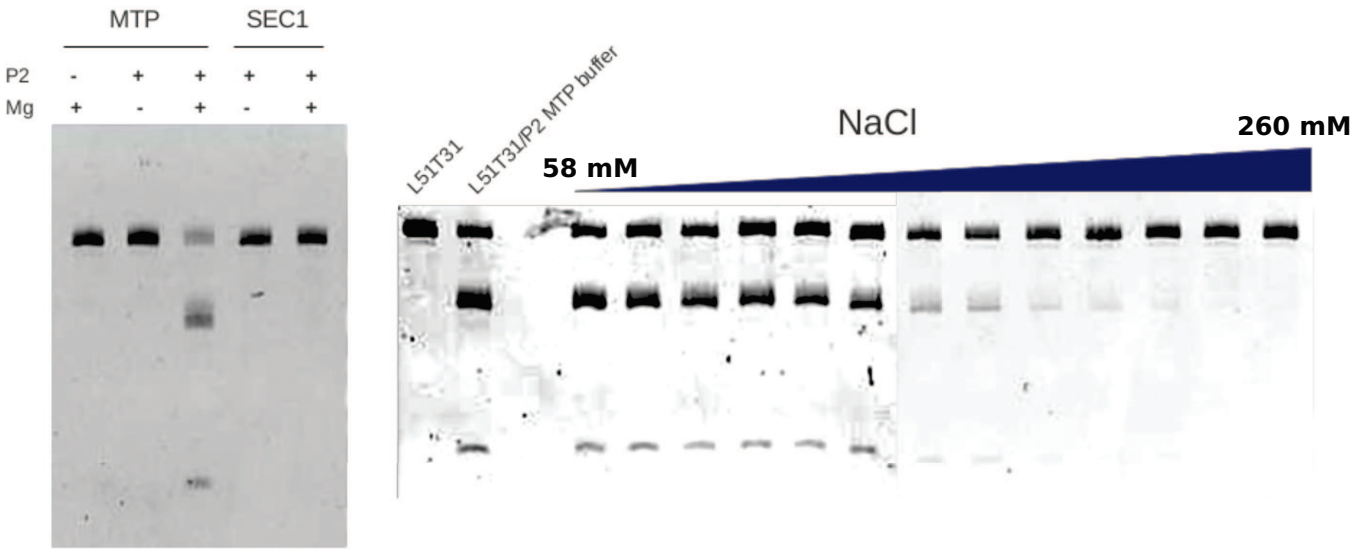


Figure 6.3: Optimization of activity buffer conditions. a) Cleavage activity of wild type PRORP2 using L51T30 tRNA^{Cys} depends on the salt concentration. MTP - low salt buffer, SEC1 - high salt, purification buffer, b) Cleavage efficiency depends on the salt concentration. Concentration on the left is 58 mM with increments of 17 mM. The highest concentration is 260 mM NaCl on the right.

In a second set of experiments, the activity was assayed in the presence of different divalent ions: 5 mM magnesium, manganese, calcium or mixes of 2.5 mM (each) Mn^{2+}/Mg^{2+}, Ca^{2+}/Mg^{2+} (Fig. 6.4).

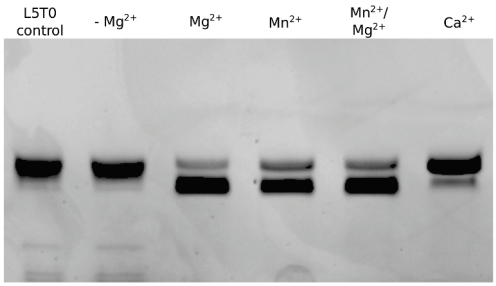


Figure 6.4: PRORP activity depends on divalent metal ions. Incubation of 250 ng PRORP2 with 200 ng L5T0 tRNA^{Cys} in a buffer containing 5 mM divalent ions (when one ion) or 2.5 mM of each ion in a mix.

PRORP is active in the presence of magnesium, manganese at 2.5 mM and 5 mM but shows only trace activity in the presence of 5 mM calcium. Therefore involving PRORP2mDD and L5T0 tRNA^{Cys} SEC2 buffer containing 5 mM $MgCl_2$ was used for further studies. As Ca^{2+} does not abolish cleavage completely and a divalent ion is required for correct tRNA folding, I did not continue working with the wild type protein during complex studies.

6.2 Finding a suitable tRNA substrate

The usual model substrate used in the laboratory for PRORP activity assays was L51T30 tRNA^{Cys}. It was chosen to easily distinguish the precursor from the mature tRNA and the leader sequence on a

8 % PAA (w/v) gel. A drawback is that these long single stranded extensions might cause intra- or intermolecular secondary structure formation, preventing efficient binding and subsequent cleavage. Furthermore, long and flexible structural elements are usually avoided in crystallography projects as they can hinder crystallization. Mitochondrial tRNA^{Cys} with different leader and trailer sequences was already cloned and available in pUC19 in the laboratory at the IBMP.

In order to find a more suitable substrate for PRORP I did *in vitro* cleavage assays using the following tRNAs: L0T30, L5T30, L11T30, L21T30, L31T30, L41T30, L5T0 and L51T0 where L indicates the length in nucleotides of the leader sequence and T the length of the trailer sequence. Plasmids containing tRNA constructs with a trailer sequence were linearized with EcoR1 (Fast digest, Fermentas) and those without trailer sequence were linearized with BmR1 according to the manufactures protocols (NEB, but incubation 4 h at 37 °C). Digested plasmids were extracted with Phenol/Chloroform and precipitated with ethanol. *In vitro* transcription was done using the Ribomax kit. Transcription levels, cleavage efficiency and conformational state, were tested on a denaturing and native 8 % PAA , respectively. Only the substrate L5T0 tRNA^{Cys} was cleaved 100 % and was most homogeneous on a native PAA gel (Fig. 6.5a+b).

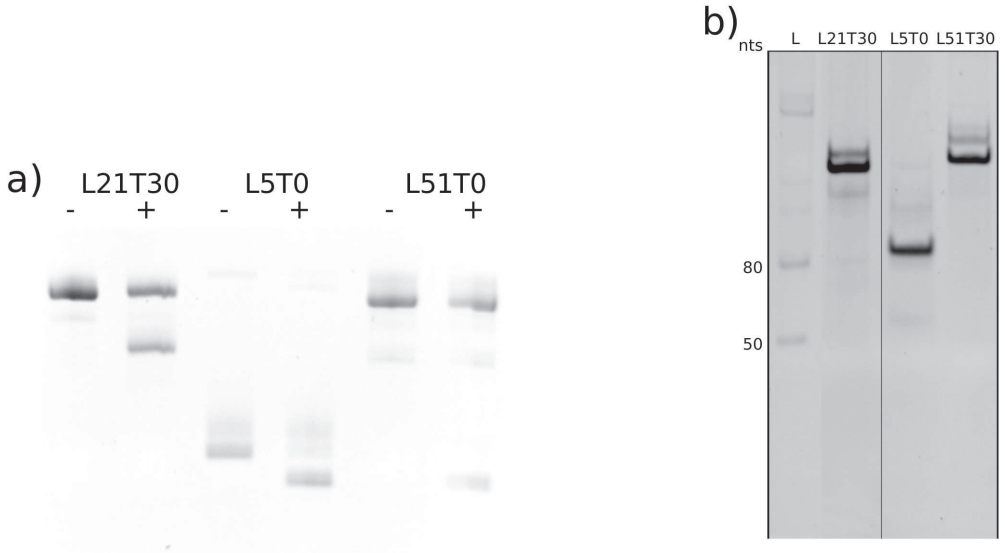


Figure 6.5: Testing cleavage efficiency and conformational homogeneity of different tRNA^{Cys} substrates. a) Cleavage assay using wild type PRORP2 with L21T30, L5T0 and L51T0 tRNA^{Cys}. (-) control without protein, (+) incubation with 250 ng PRORP2. b) Native gel of L21T30, L5T0 and L51T0 tRNA^{Cys}. 250 ng of each tRNA was loaded on a 8 % PAA gel, run at 250 V and stained with ethidium bromide.

Therefore, I decided to take L5T0 tRNA^{Cys} as a working substrate for further structural complex studies as it showed best transcription level compared to the other tested substrates. It is predominantly present in one conformation and is 100 % cleaved by PRORP2 after 30 min.

6.3 Determination of binding affinity between PRORP2 and L5T0 tRNA^{Cys}

6.3.1 Isothermal titration calorimetry

I used ITC to determine binding constants, stoichiometry and thermodynamic parameters. ITC also gives information about the percentage of active molecules in solution. While it can not distinguish between opposite effects such as dilution or aggregation, subsequent DLS measurements provide this complementary information. A main issue is the high amounts of material that are needed for one experiment, the production of RNA being the limiting parameter. Still, RNA can be recovered after the experiment by phenolic extraction or in the case of a successful experiment the sample can be directly used for crystallization trials.

The first ITC experiments did not show any interaction and data could not be analyzed. Subsequent DLS measurements showed complete sample aggregation after the experiments (Fig. 6.6). Each figure shows in a) a control DLS measurement that proves near homogeneity of the protein sample **before** ITC, in b) the raw data of the ITC measurements and in c) the control DLS measurement of the PRORP/tRNA mixture **after** ITC. In these experiments a L5T5 tRNA^{Cys} was used as well as active proteins in SEC1 buffer without $MgCl_2$. ITC showed endothermic and exothermic values for wild type PRORP1/L5T5 tRNA^{Cys} titration and after the seventh injection no thermodynamic changes at all.

In the case of wild type PRORP2/L5T5 tRNA^{Cys} endothermic values were measured which could not be analyzed neither. The control DLS measurements of the samples after ITC experiments also revealed complete aggregation (Fig 6.7c). This led to the question of a suitable sample buffer and an optimal tRNA substrate as the used buffer in these experiments contained high salt in comparison to the buffer used for the activity assays, 250 mM and 30 mM NaCl, respectively (Section 6.1). At that time I worked with the wild type enzymes that in presence of Mg^{2+} are active and would cleave the pre-tRNA. Therefore Mg^{2+} was omitted in these experiments.

Titration of the inactive catalytic mutant PRORP2mDD to L5T0 tRNA^{Cys} was carried out in SEC2 buffer including 5 mM $MgCl_2$. An inactive enzyme was combined with a buffer supplemented with Mg^{2+} which is important for correct RNA folding. Data analysis led to a K_D of 1 μM and a stoichiometry of 0.4 (Fig. 6.8a). This would mean that 50 % of either of the partners is inactive or PRORP is partially dimerized in the given buffer as analytical ultracentrifugation (AUC) experiments proposed later. Subsequent DLS measurements confirmed that the sample did not aggregate.

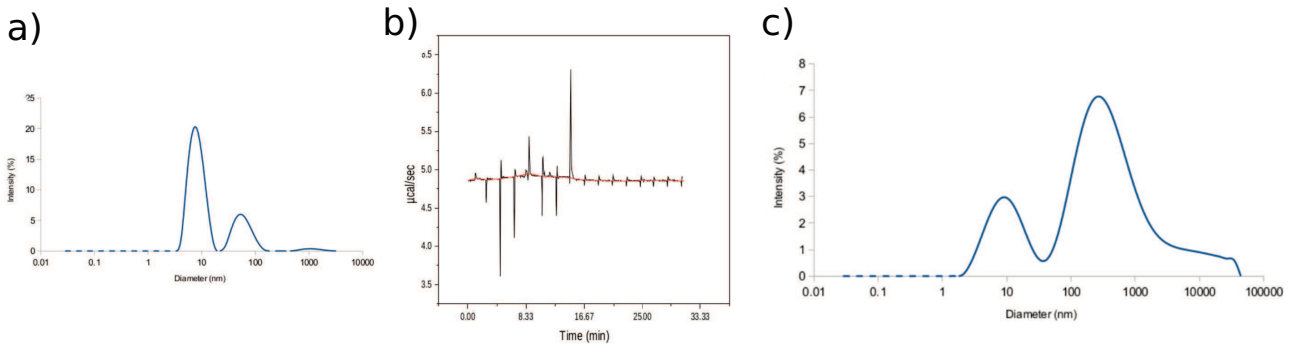


Figure 6.6: ITC and DLS control measurements of wild type PRORP1cris with L5T5 tRNA^{Cys}. a) DLS of wild type PRORP1cris before ITC experiment, b) Titration of wild type PRORP1cris to L5T5 tRNA^{Cys}, c) DLS of the PRORP/tRNA mix after ITC experiment showing aggregation.

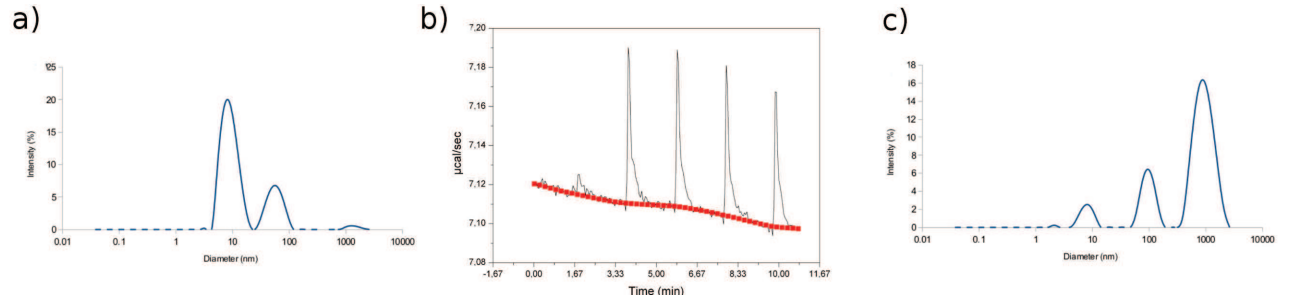


Figure 6.7: ITC and DLS control measurements of wild type PRORP2 with L5T5 tRNA^{Cys}. a) DLS of wild type PRORP2 before ITC experiment, b) Titration of wild type PRORP2 to L5T5 tRNA^{Cys}, c) DLS of the PRORP/tRNA mix after ITC experiment showing aggregation.

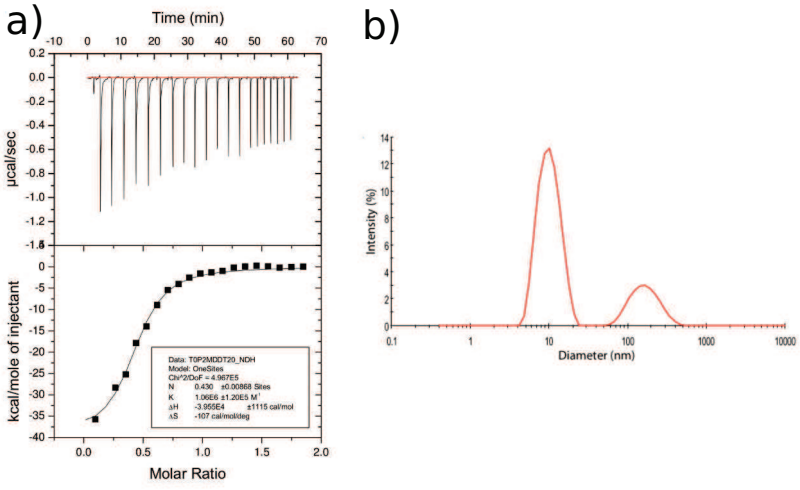


Figure 6.8: ITC and DLS control measurements of PRORP2mDD with L5T0 tRNA^{Cys}. a) Titration of PRORP2mDD to L5T0 tRNA^{Cys}, b) DLS of the PRORP/tRNA mixture after ITC experiment.

Analytical size exclusion chromatography after ITC As DLS measurements showed no aggregation after the ITC experiments a 10 μl sample of the complex solution was loaded on an Agilent Bio SEC-3, 150 Åcolumn which was concentrated beforehand on an amicon ultracentrifugation unit

with a 10K MWCO. The L5T0 tRNA^{Cys} alone elutes at 12.5 min (24 kDa), PRORP2mDD at 11.9 min (60 kDa) and the complex as expected earlier at 11.5 min (Fig. 6.9). The proximity of the tRNA and PRORP2mDD can be explained by the L-shaped form of the tRNA making its apparent mass looking more like a spheric molecule of larger diameter. Despite repetitive attempts I could not reproduce these results by incubating PRORP2mDD and L5T0 tRNA^{Cys} in the final ratio of 32 μM and 18 μM (referring to final concentrations in the ITC expriment but before concentration), respectively.

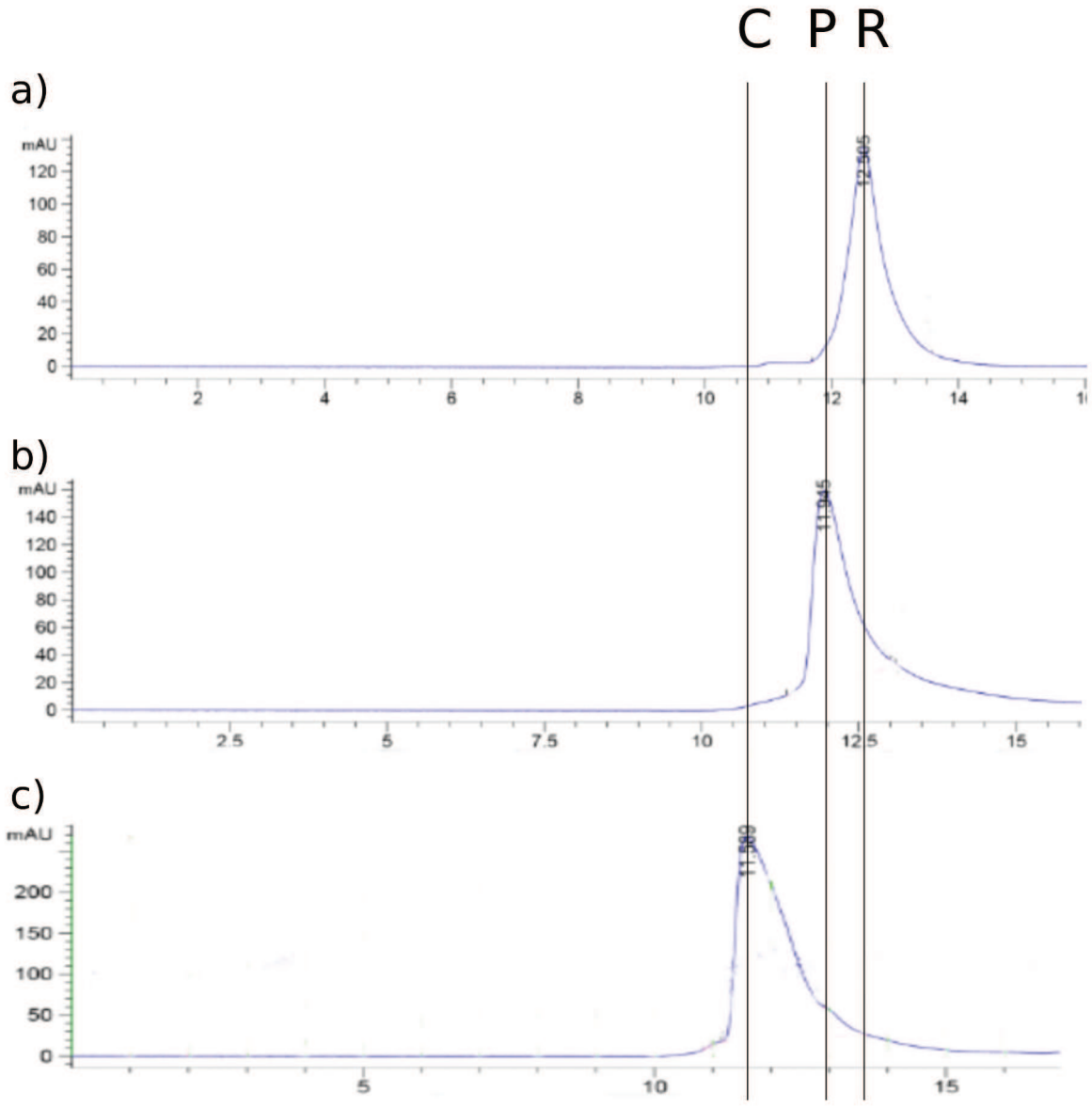


Figure 6.9: Analytical size exclusion of ITC sample. a) tRNA and b) PRORP2mDD are indicated as size references in a plot of arbitrary absorption units (mAU) against the time (min) at 260 nm. c) ITC mixture of tRNA/PRORP2mDD . Elution times of tRNA (R) alone, PRORP2mDD (P) and the complex after ITC (C) are 12.5 min, 11.9 min and 11.5 min, respectively.

This experiment was the first time I could observe a complex. It is not clear why sequential titration of PRORP into the substrate solution was important for complex formation in this instance. A direct

mixing of both partners in the same ratio does not lead to the same result. However, the complex is present and stable in solution as AUC experiments showed (Fig. 6.11). Whether the complex dynamic properties in a gel filtration system have an influence on complex stability remains to be determined.

6.3.2 Microscale thermophoresis (MST): PRORP2mDD/L5T0 tRNA^{Cys} interaction

An inherent requirement of MST is a fluorescent labelled molecule. The fluorophore is either introduced in the protein which can cause differences in solubility or it is introduced in the ligand.
In order to obtain reliable data of the binding properties of PRORP2 to the model substrate L5T0 tRNA^{Cys} we decided to work with the catalytic mutant PRORP2mDD. The advantage is that we could add Mg^{2+} which is the divalent ion in the active center of PRORP, but which, more importantly, is necessary for the correct tRNA folding and its structural integrity.
Data were analyzed using the Nanotemper Analysis software. Only the eight points at highest PRORP concentration were taken into account to calculate the affinity constant by resolving the quadratic equation of mass law and an affinity constant of 1.3 μM could be determined (Fig. 6.10c). At lower protein concentrations the MST data was difficult to interpret (Fig. 6.10b). A hypothesis could be that PRORP dimers (not detectable with DLS) are present in mixture and a first slow binding takes place only after dissociation of the dimers. Dimers were detected studying the same system with AUC.
The experiment should also be repeated in hydrophilic capillaries using 0.1 % Tween as detergent to reduce the effect of a potential protein adsorption to the glass surface.

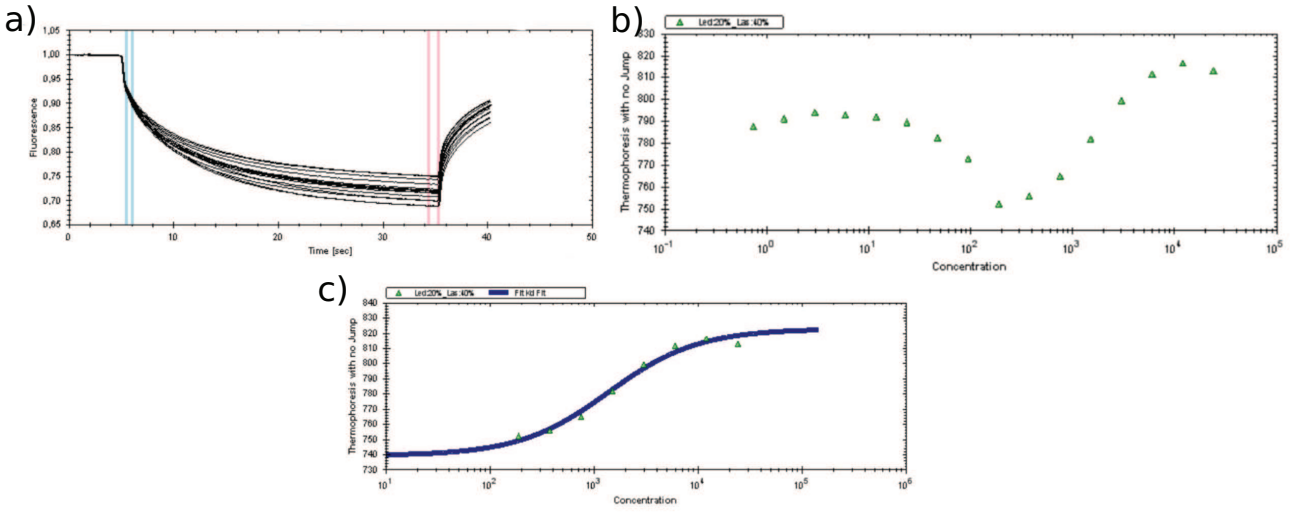


Figure 6.10: Microscale thermophoresis using PRORP2mDD and L5T0 tRNA^{Cys}. a) Normalized fluorescence signal of the titration points during the MST experiment is plotted against time. b) Thermophoresis without temperature jump, c) Plot of selected data points to evaluate the affinity constant including the fit depicted in blue.

6.3.3 Analytical ultracentrifugation: PRORP2mDD/L5T0 tRNA^{Cys} interaction

In order to validate the K_D values obtained with MST and ITC I chose AUC as another label-free technique in solution. It requires only low amounts of RNA (1 μM in 400 μl) but important amounts of protein to make several titration points (0.5 - 18 μM in 400 μl). The most important parameter to obtain quality data are precise sample concentrations and extinction coefficients. The RNA concentration was determined with a nanodrop spectrophotometer using a theoretical extinction coefficient for RNA of 6.1×10^5 M^{-1} cm^{-1} or 40 ng cm^{-1}/μl. This value is about one third lower than the extinction coefficient calculated with OligoCalc, ε_{260nm} = 9.8×10^5 $M^{-1}cm^{-1}$ and M = 24.5 kDa (Kibbe, 2007), leading to an overestimation of the RNA concentration of one third, i.e. 0.7 μM instead of 1 μM. The binding affinities were calculated with the theoretical ε_{260nm}. As most RNA concentrations are determined with a nanodrop spectrophotometer, this choice seemed to be justified.

The control measurements of each binding partner alone showed that at 17 μM about 16 % of PRORP2mDD exists as dimers (s = 2.76 S and s_w = 3.54 S) and 78 % as monomers. tRNA was monomeric in one peak sedimenting at s = 3.17 S and s_w = 4.06 S. This difference can be explained by different form factors that directly influence the frictional force. Figure 6.11a shows that with increasing protein concentrations a new peak at $\sim$ 4.4 S appears that increases in size with the concentration of protein, so with the amount of formed complex (Fig. 6.11a). The calculated K_D was 1.2 μM and is in good accordance to the values measured with MST and ITC.

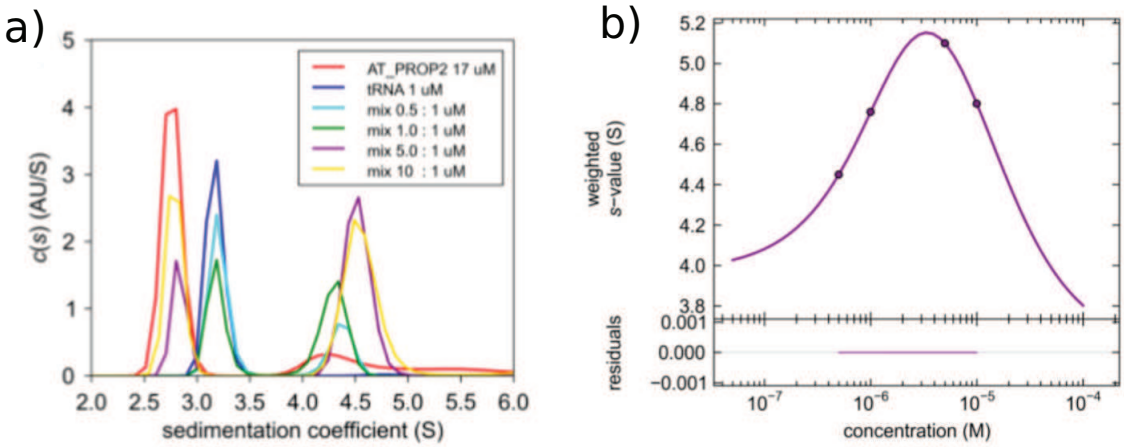


Figure 6.11: Analytical ultracentrifugation PRORP2mDD and L5T0 tRNA^{Cys}. a) Distribution of sedimentation coefficients of the different titration points, b) Isotherm of the weighted-average sedimentation coefficients against the protein concentration.

6.3.4 Summing up affinity constant measurements

I used various methods to determine the binding affinities of my system, including a catalytic inactive mutant of PRORP2 and a L5T0 tRNA^{Cys} in SEC2 buffer, that all have advantages and disadvantages. AUC is the most convenient with rather low sample requirements and label-free like ITC. All methods gave comparable affinity constants of about 1 μM which is comparable to values published in Howard *et al.* (2012).

During gel shift experiments I could not observe a shift under the given conditions. I did not want to use radioactive-labelled RNA and large excesses of protein to see a shift as these conditions are not transposable afterwards into structural experiments.

Table 6.1: K_D values obtained using different methods. The system used for all analyses was PRORP2mDD and L5T0 $tRNA^{Cys}$.

Method	K_D [μM]
MST	1.3
ITC	1
AUC	1.2

6.4 Identifying reliable binding conditions

6.4.1 Analytical size exclusion chromatography

The laboratory is equipped with the same HPLC system used at synchrotron SOLEIL for Bio-SAXS experiments, i.e. an Agilent HPLC system where analytical gel filtration columns can be connected upstream the SAXS capillary. This setup separates aggregates, which can make data analysis completely impossible, from the molecules of interest. Samples that are susceptible to aggregation during transport to the synchrotron can still be analyzed. Having the same setup in the laboratory enabled us to test experiments before going to the synchrotron.
In this sense I wanted to verify the existence of a stable PRORP/tRNA complex analyzable in a SAXS experiment. I tested several PRORPs, wild type as well as catalytic mutants, tRNA substrates with and without trailer sequence, with and without anticodon stem loop, as well as a mature yeast $tRNA^{Phe}$. To increase RNA protein interactions I also tested buffers containing $(NH_4)_2SO_4$ known to stabilize tRNA/amino acyl-tRNA-synthetase complexes (Florentz *et al.*, 1990).

Example of detecting a pre-tRNA/PRORP complex A summary of of my tests using analytical SEC is given in Tab. 6.2, with different mixing ratios of tRNA and PRORP, different incubation times, testing different buffers. An example of tested conditions is illustrated in Fig. 6.12. In this example I tested a buffer containing 50 mM Hepes-Na pH 7.5, 500 mM $(NH_4)_2SO_4$, 5 % (w/v) glycerol, 5 mM $CaCl_2$. It showed that both wild type PRORP2 and a substrate L5T5 $tRNA^{Cys}$ MAC (minus anticodon) elute at the same time. Mixing both partners in equimolar ratios did not change the elution time. Injection of a mixture of tRNA/PRORP2 = 2:1 changed the picture and an upstream peak appeared (15.7 min). This peak increased in height when injecting ratios of tRNA/PRORP2 = 3:1. Testing these high salt concentrations was based on the observation that ammonium sulphate could play a supportive role in tRNA-protein complex crystallization and thus in RNA-protein interaction (Florentz *et al.*, 1990).

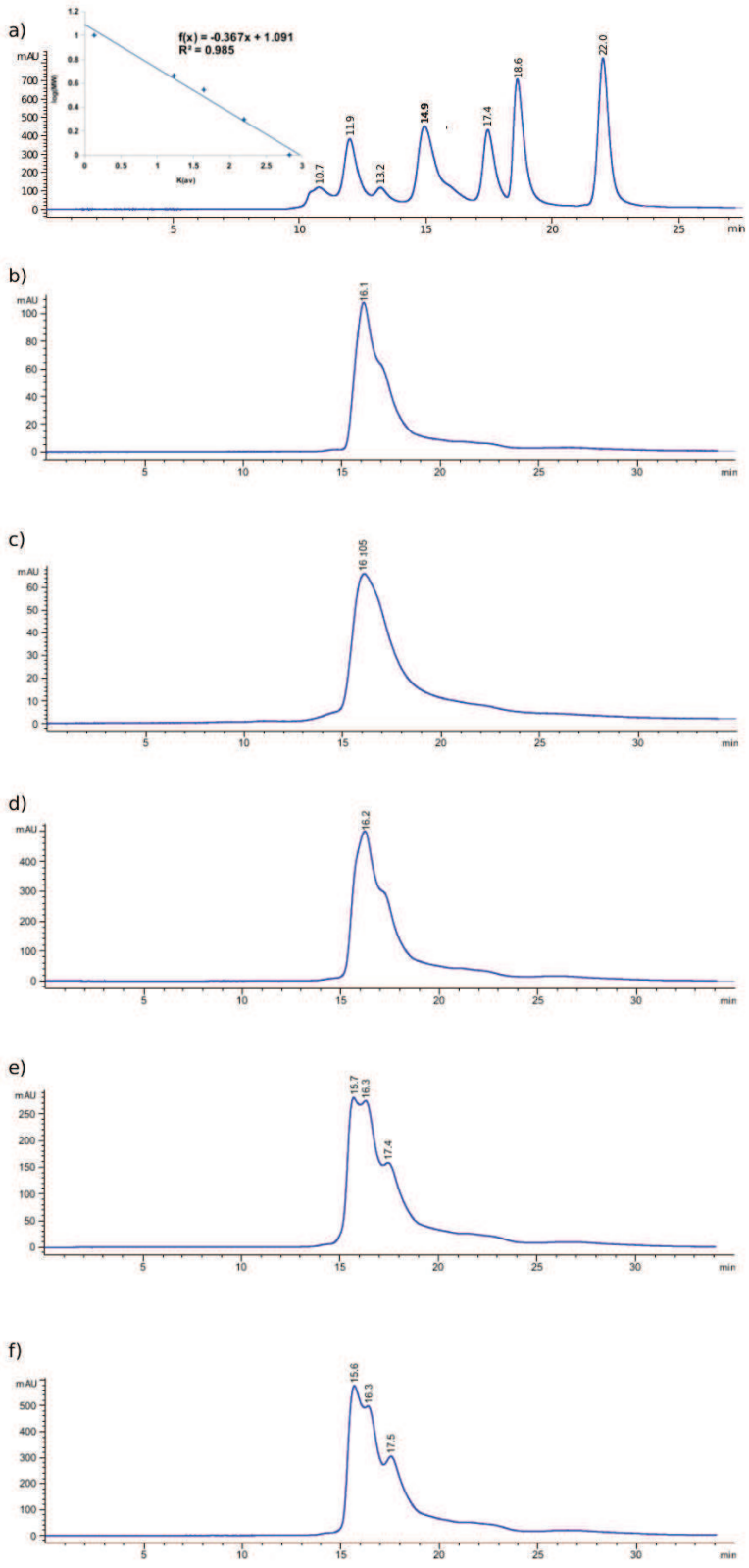


Figure 6.12: Analytical gel filtration: L5T5 tRNA^{Cys} MAC and wild type PRORP2. a) Calibration of Agilent BioSEC3 column with Biorad gel filtration standard (15 μl) in 50 mM Hepes-Na pH 7.5, 100 mM $(NH_4)_2SO_4$, 5 % (w/v) glycerol buffer. The inlet shows the derived regression curve to determine the molecular weight from injected samples. b) L5T5 tRNA^{Cys} MAC, c) wild type PRORP2, d) equimolar ratio of tRNA/PRORP2, e) 2:1 molar ratio of tRNA/PRORP2, f) 3:1 molar ratio of tRNA/PRORP2. Column flow 0.2 ml/min.

Table 6.2: Analytical gel filtration experiments. * These analyses were performed with an Agilent Bio SEC-3 column, with a pore size of 300 Å, an inner column diameter of 4.6 mm and a length of 300 mm and a bead size of 3 μm, useful molecular weight range 5-1250 kDa, ** Agilent Bio SEC-3, with a pore size of 150 Å, an inner column diameter of 4.6 mm and a length of 300 mm and a bead size of 3 μm, useful molecular weight range 0.5-150 kDa, $^{+}$ final concentration after ITC measurement; afterwards samples were concentrated on Amicon filtration units with a 10K MWCO.

Buffer	**tRNA substrate (quantities)**	**PRORP (quantities)**	**Remarks**
50 mM Hepes-Na pH 7.5, 5% (w/v) glycerol, 100 mM $(NH_4)_3SO_4$, 5 mM $MgCl_2$	L5T5 0.08 nmol		*
		P2mDD 0.3 nmol	
	L5T5 0.08 nmol	P2mDD 0.02 nmol	
	L5T5 0.08 nmol	P2mDD 0.15 nmol	
50 mM Hepes-Na pH 7.5, 5% (w/v) glycerol, 500 mM $(NH_4)_3SO_4$, 5 mM $MgCl_2$	L5T5 0.05 nmol		
	L5T5 0.3 nmol	P2mDD 0.2 nmol	t_{inc} = 5 min; 30 min
20 mM Hepes-Na pH 7.5, 5% (w/v) glycerol, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP	MAC L5T5 0.07 nmol	P2mDD 0.5 nmol	t_{inc} = 5 min
SEC1	$tRNA^{Phe}$ (*S. c.*) 0.1 nmol		
		P2wt 0.8 nmol	
	$tRNA^{Phe}$ 1.6 nmol	P2wt 1.8 nmol	
		P3wt 0.7 nmol	
	$tRNA^{Phe}$ 0.2 nmol	P3wt 0.2 nmol	
		P1wt 1.7 nmol	
	$tRNA^{Phe}$ 0.4 nmol	P1wt 1.1 nmol	
		P2wt 0.7 nmol	**
	L5T5		
	MAC L5T5		
50 mM Hepes-Na pH 7.5, 5% (w/v) glycerol, 500 mM $(NH_4)_3SO_4$, 5 mM $CaCl_2$		P2wt 0.7 nmol	
	MAC L5T5 0.3 nmol		
	MAC L5T5 0.7 nmol	P2wt 0.7 nmol	
	MAC L5T5 0.7 nmol	P2wt 1.4 nmol	
	MAC L5T5 0.7 nmol	P2wt 2.1 nmol	
	L5T5 0.6 nmol		
	L5T5 1.6 nmol	P2mDD 0.8 nmol	
SEC2		P2wt 0.8 nmol	
	MAC L5T5 0.08 nmol		
	MAC L5T5 0.8 nmol	P2wt 0.8 nmol	t_{inc} = 1h (ice) and 15 min (RT)
	MAC L5T5 0.8 nmol	P2wt 1.6 nmol	t_{inc} = 15 min (RT)
	MAC L5T5 0.8 nmol	P2wt 2.4 nmol	t_{inc} = 15 min (ice) and 15 min (RT)
	L5T0 18 μM^{+}	P2mDD 32 μM^{+}	samples post-ITC
	L5T0 18 μM	P2wt 32 μM	
	L5T0 54 μM	P2wt 96 μM	
SEC2 + 5 mM $MgCl_2$	L5T0 2.8 μg		
		P2mDD 0.9 nmol	
	L5T0 18 μM^{+}	P2mDD 32 μM^{+}	10 μl inj. vol.; samples post-ITC
	L5T0 0.1 nmol		
	L0T0 0.1 nmol		
	L5T5 0.1 nmol		
	L5T0 18 μM^{+}	P2mDD 32 μM^{+}	samples 7 d post-ITC
	MAC L5T0 0.1 nmol		

To conclude, I was not able to detect a complex using analytical SEC by mixing PRORP and the tRNA. Results were ambiguous and the the case of a putative complex after ITC experiments never reproducible.

6.4.2 Crosslink

I also tested crosslinking as a way to artificially stabilize an interaction, e.g. protein and RNA. The main issue with this approach is the specific and efficient binding of the two partners.
The first crosslinking experiments between the tRNA and PRORPs were performed with non-modified tRNAs. Tests confirmed that L5T0 tRNA^{Cys} is stable at least up to an UV dose of 640 mJ/cm^2 at 254 nm and 365 nm whereas PRORP2 begins to degrade at a dose of 200 mJ/^{2} (Fig. 6.13). Under the same conditions non-modified L5T0 tRNA^{Cys} was stable and showed neither degradation nor intermolecular cross-linking (gels not shown). Nevertheless, experiments to crosslink non-modified tRNA to PRORP failed.
Further, I tested a thio-modified tRNA substrate (Fig. 2.30). Longer (less energetic) UV wavelengths can be used, i.e. 365 nm but intermolecular crosslinking is also more likely if tRNAs are in close proximity (Fig. 6.14). It seems that tRNA adducts formed upon irradiation (Fig. 6.14a , lane 2, without protein). This species decreased with increasing protein concentrations and a species that did not enter the gel (red bands in the wells) formed. Even using 4 % gels these adducts do not enter the gel. Maybe aggregation is induced in the mixture of protein and tRNA upon UV irradiation.

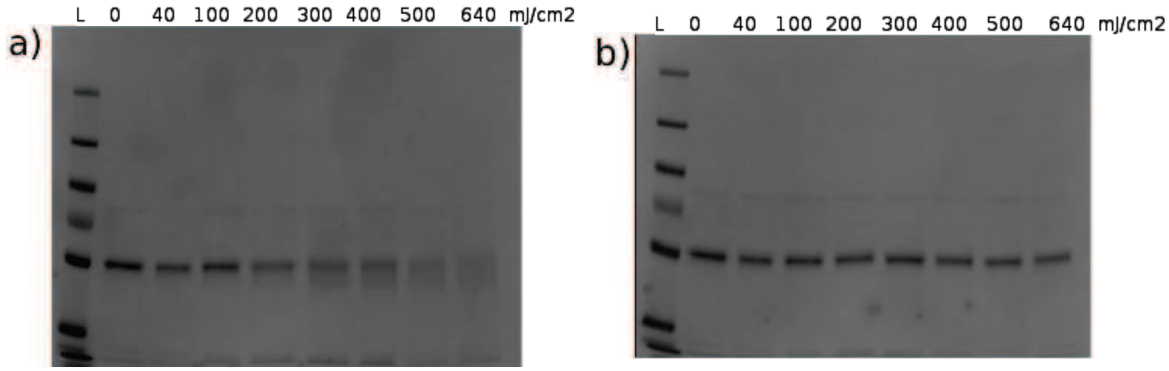


Figure 6.13: Crosslink of wild type PRORP2 at different wavelengths and increasing doses. Each lane contains 1 μg of PRORP2. The 7.5 % TGX SDS-PAA gel is stained with Coomassie. Irradiation at a) 254 nm, b) 365 nm and increasing doses. L - protein ladder.

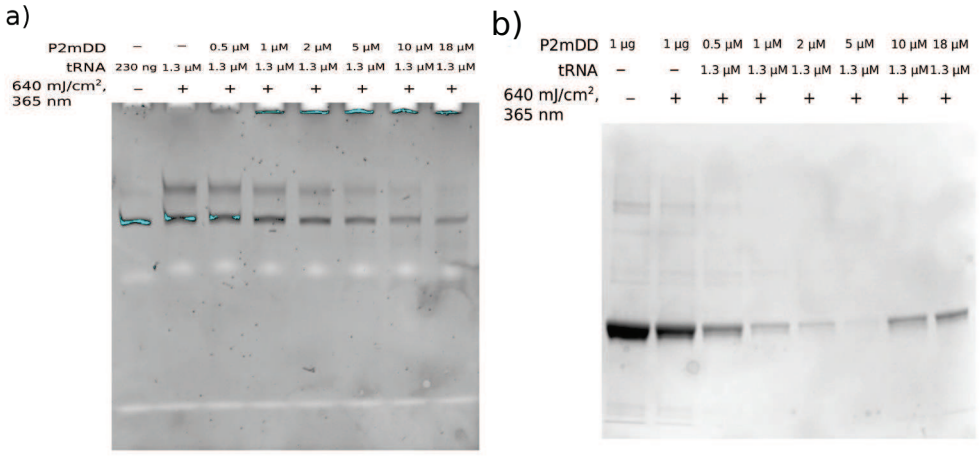


Figure 6.14: Crosslink of PRORP2mDD to L5T0 tRNA^{Cys} at 365 nm, 640 mJ/cm^2. a) RNA gel, 12 % PAA/8M urea, ethidium bromide staining. Irradiation at 365 nm, 640 mJ/cm^2, 1.3 μM tRNA and increasing concentration of PRORP2mDD, b) 7.5 % TGX gel, Coomassie staining. Irradiation at 365 nm, 640 mJ/cm^2, 1.3 μM tRNA and increasing concentration of PRORP2mDD.

6.4.3 SAXS experiments on PRORP2mDD and L5T0 tRNA^{Cys}

SAXS analyses were carried out at the SWING beamline (SOLEIL, Saint-Aubin) to obtain information on the interaction of PRORP2 with its substrate. In practice, the most important challenge was to form a PRORP/tRNA complex stable enough to record a SAXS profile. Due to their weight difference, the two partners, as well as the complex, should easily be separated on a SEC column used upstream the SAXS cell. However, the elongated shape of tRNAs made them elute close to PRORP enzymes on a Bio-SEC3 column with 300 Åpore size (Agilent). In order to improve the resolution we purchased a column with smaller pore size (150 Å) which is more resolutive for medium size biomolecules (molecular weight range 0.5 - 150 kDa) but it was still not possible to clearly identify the three populations, tRNA - PRORP - PRORP/tRNA, based on UV absorption signals. Nevertheless, the complex should be detectable by SAXS along the SEC profile according to its larger gyration radius.

In preliminary experiments performed in the laboratory I also observed that different pre-tRNA substrates eluted according to expected sizes on the analytical SEC column, but only in the presence of Mg^{2+}. Adding magnesium to the buffer dramatically changed elution volumes of tRNAs which was not clearly visible on a native PAA gel. This suggests that Mg^{2+} is required for correct and compact tRNA folding leading to longer elution times, i.e. with smaller apparent molecular weights. This excludes the possibility to work with the active enzymes which would cleave the leader.

In the first trials I prepared mixes of the inactive enzyme and L5T0 tRNA^{Cys} substrate in 1:1, 2:1, 3:1 molar ratios but I got no clear evidence either in the SEC separation, nor in SAXS data, for the existence of such a stable complex. Thus, knowing from three independent analyses (ITC, MST and ITC) that the binding affinity of the catalytic mutant to a L5T0 substrate was about 1 μM, we eventually adopted the following strategy: the SEC column was equilibrated with a buffer containing the enzyme at a concentration close to the K_D to prevent the dissociation of the partners during elution

and the SAXS data collection (Fig. 6.15).

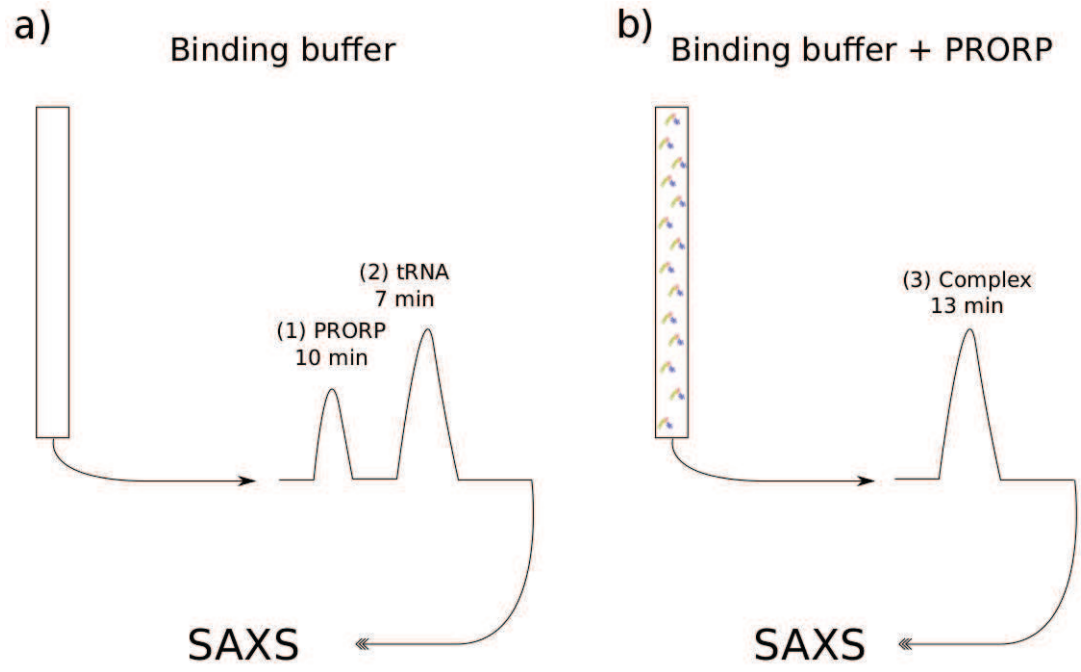


Figure 6.15: SAXS strategy for PRORP/pre-tRNA complex experiments. a) Measurments of individual partners that should elute at different times from the SEC column (arbitrary elution times). b) The SEC system is equilibrated with a buffer supplented with PRORP2mDD which should help maintain the injected preformed complex during SEC separation and SAXS measurements.

The catalytic mutant PRORP2mDD was purified as described before in SEC2 buffer supplemented with 5 mM $MgCl_2$ and concentrated to 19 mg/ml (320 μM). L5T0 $tRNA^{Cys}$ was dialysed into the same buffer and concentrated to 4.6 mg/ml (185 μM). SAXS measurements were first collected on the single molecules, i.e. PRORP2mDD (Fig. 6.16) and L5T0 $tRNA^{Cys}$, as a control.

Then, the column was equilibrated with the same buffer containing 1.5 μM PRORP2mDD. The complex was preformed just before the analysis by mixing 25 μl L5T0 $tRNA^{Cys}$ and 30 μl of PRORP2mDD which corresponds to a solution with a PRORP2mDD : L5T0 $tRNA^{Cys}$ molar ratio of 2.1 : 1. The SAXS profile in the main elution peak, interatomic distance distributions, or P(r) functions, and R_g values determined from the Guinier plot were clearly different from those of individual partners (Tab. 6.3). Characteristic SAXS curves for the two partners and the complex, as well as the P(r) of PRORP2mDD and the complex are shown in Fig. 6.17.

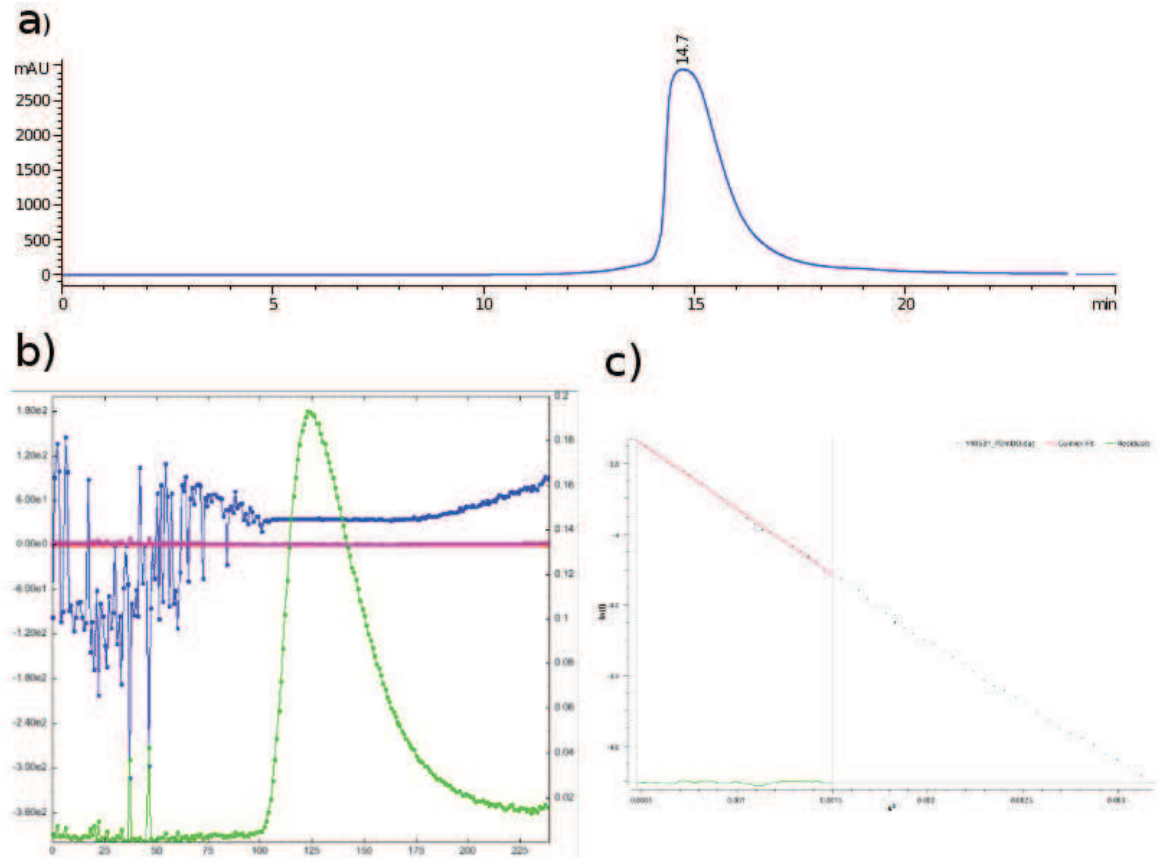


Figure 6.16: Saxs analysis of PRORP2mDD. a) Elution profile from the Bio-SEC3 colum showing a monomeric sample. b) Green curves shows the scattering signal of PRORP2mDD along the elution peak. Blue dots represent estimated R_g as a function of collected frames (x axis) and show that the particles in the peak are homogeneous in size. c) Guinier plot determined with PRIMUS (ATSAS suite) showing the zone in red that was taken into account for R_g calculation and the corresponding residuals in green.

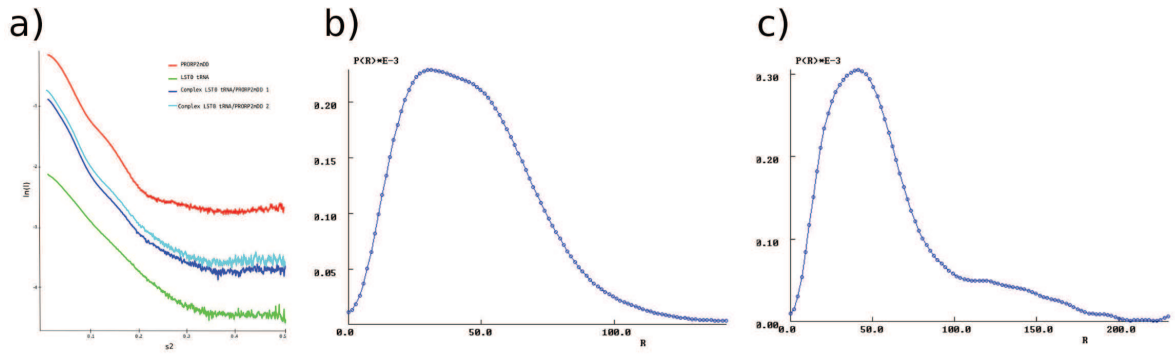


Figure 6.17: Comparison of PRORP2mDD, pre-tRNA and PRORP2mDD/pre-tRNA by SAXS. a) Scattering profiles of PRORP2mDD, L5T0 $tRNA^{Cys}$, and of PRORP2mDD/L5T0 $tRNA^{Cys}$ from two independent measurements. b) P(r) of PRORP2mDD showing its two-domain organization and its elongated shape with a maximal distance of 123 Å. c) P(r) of PRORP2mDD/L5T0 $tRNA^{Cys}$ complex showing its overall globular shape with long extensions and a maximal distance of 230 Å.

Table 6.3: SAXS data analysis. Radius of gyration, R_g, and maximal interatomic distance, D_{max}, of PRORP2, L5T0 tRNA and PRORP2/tRNA complex determined from the Guinier approximation as well as from the distribution function of interatomic distances.

	Guinier $\rightarrow R_g$ [Å]	P(r) $\rightarrow R_g$ [Å]	D_{max} [Å]
L5T0 tRNA^{Cys}	31	35	140
PRORP2mDD	35	35	123
PRORP2mDD/L5T0 tRNA^{Cys}	49	50	230

Model building and refinement is in progress and will benefit from on going mutational analyses on residues critical for RNA binding in PPR motives. These results represent the first structural data of nuclear PRORP2 with a model tRNA precursor and constitute a step towards the crystallization of this complex. On a practical point of view, I will now exploit the SEC protocol validated by SAXS to prepare samples for crystallization.

Discussion and Perspectives

This thesis presents a multidisciplinary approach to investigate the structural and biophysical properties of PRORP enzymes from *Arabidopsis thaliana* as single molecules as well as in complex with tRNA precursors. My aim was to contribute and to expand the knowledge of RNA/protein complexes in general and of tRNA/PRORP interactions in particular, but also of the organization and structure of PPR proteins.
I thus investigated PRORP enzymes in solution using an ensemble of biochemical and biophysical methods, and confirmed that both organellar and nuclear PRORPs are mainly composed of α-helices, as suggested by bioinformatic analysis. They fold in two distinct domains joined by a zinc binding motif. Homology modelling and SAXS data helped building the first model complex of PRORP with a precursor tRNA substrate.

4.1 Technical analysis

4.1.1 RNA purification

In structural biology the production of macromolecules is of critical importance because milligram-quantities of pure and homogeneous samples are required. The most delicate and time consuming step in my project was undoubtedly the purification of tRNAs. To minimize material losses during gel elution and to maintain the RNA in a native state I set up a strategy combining anion exchange and size exclusion chromatography. The main challenge was to find the appropriate salt gradient to obtain best resolution and purity. This led to a rapid and efficient protocol that provided mg-amounts of RNA required for my experiments in SAXS, AUC or ITC.

4.1.2 Structure determination and crystallization of wild type PRORP2

Whereas the purification of PRORP proteins was fairly simple and straightforward its crystallization was less obvious. Initial conditions were searched following a classical trial and error process. Then the challenge consisted in 1) reproducing the crystals and 2) optimizing the conditions to decrease nucleation and to obtain fewer, bigger single crystals if necessary. Still, having single, good-looking crystals does not guarantee that they will diffract.
Reproducibility of PRORP2 crystals was clearly an issue, likely due to minor variations in the purification buffer and batch quality, or to temperature fluctuations during crystallization. As a result of optimization, three X-ray diffraction datasets were collected on wild type PRORP2 crystals. Their low symmetry (space group P1) and medium resolution (3-3.5 Å), as well as the presence of two

enzyme entities with high helical contents in the unit cell complicated the determination of the structure. Indeed, molecular replacement has remained unsuccessful so far despite the use of a variety of softwares.
My next objective will be a massive production of PRORP2 crystals to search for samples diffracting at higher resolution. To do so, I also plan to take profit from a new instrument, the Xtal controller 900 (Xtal Concepts GmbH, Hamburg, Germany), that enables a more rational optimization and a better control over the crystallization conditions, making use of DLS, video and an accurate micro-balance. The crystallization chamber is equipped with two pumps, one for water and a second for the crystallant solution, to control the mother liquor composition in time and therefore nucleation and crystal growth events. In parallel, I will further employ the chipX microfluidic chip (Pinker *et al.*, 2013) to produce wild type PRORP2 crystals for *in situ* X-ray diffraction analyses at room temperature.

4.1.3 Towards a PRORP/tRNA complex

In order to obtain a stable complex I used different methods to determine the binding affinities of my model system. To this end I used a catalytically inactive PRORP2 mutant and a L5T0 $tRNA^{Cys}$ substrate. I found comparable K_D values of about 1 μM with three methods. This affinity seems relatively weak, although PRORP affinity to RNA is in principle not expected to be high since RNase P activity involves the transient interaction of the enzyme with its pre-tRNA substrate. Still, the relatively low affinity could also be explained by the fact that 1) we are working in an *in vitro* system and reaction conditions are not optimal, 2) there may be additional factors *in vivo* that stabilize the complex and increase affinity, 3) PRORP proteins have to recognize and process all kind of canonical tRNAs. Hence this recognition is likely based on structures rather than on sequences since PRORPs can also cleave tRNA-like structures (Gutmann *et al.*, 2012).
A simple mean to test RNA/protein interactions is electromobility shift assay where by titrating the protein against a constant amount of radioactive labelled RNA a complex between the two partners forms which can be detected on a native PAA gel. However, in order to maintain a stable complex during gel migration a huge molar excess of protein over RNA must be used which cannot be transposed into neither SAXS nor crystallization experiments.
In SAXS the challenge is to obtain a clean scattering signal that originates from a single homogeneous population of particles. As the affinity of PRORP for its substrate is low we had to find a strategy to stabilize this complex, especially during the step of size exclusion chromatography used upstream the SAXS analysis to separate the molecules of interest from aggregates. After complex dissociation the two partners will never see each other again because they migrate differently through the gel filtration matrix. To prevent this we supplemented the mobile phase with 1.5 μM of PRORP2mDD. This prevented complex separation and helped collect SAXS data on a PRORP2/L5T0 particle.
Nuclear PRORP2-3 contain a long nuclear localization signal (NLS) that is not cleaved *in vivo*. Fig. 4.18 shows the model of PRORP2 including its long N- and C-terminal extensions that are the NLS and the His-tag, respectively. These unstructured regions are not only detectable in the SAXS data but they may also interfere with crystallization. It will be interesting to test PRORP2-3 without NLS and the His-tag for crystallization purpose. In addition, SAXS may be used to compare a complex

of nuclear PRORP2-3 with L5T0 tRNA^{Cys} and with the same substrate lacking the anticodon arm in order to better define the position of the RNA and to build a more accurate model. Such a compact, less floppy RNA would be also more suitable for crystallization.

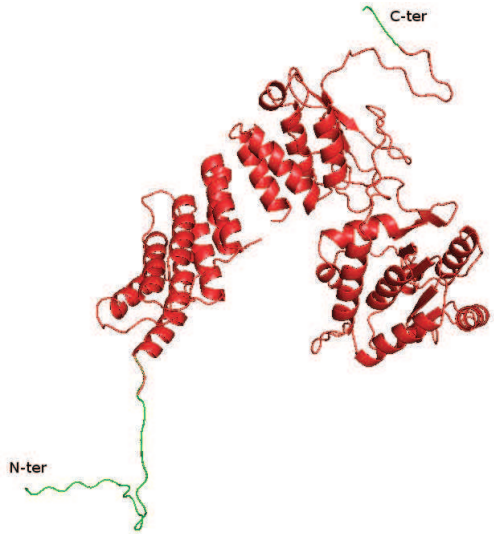


Figure 4.18: PRORP2 model. A model of PRORP2 was built using Modeller and the PRORP1 template (Eswar *et al.*, 2006, Howard *et al.*, 2012). The amino acids lacking in the intein version of PRORP2 are highlighted in green. PRORP2-His and PRORP2-intein contain 537 and 508 aa, respectively.

4.2 Biological discussion

4.2.1 Life without ribonucleoproteic RNase P

4.2.1.1 PRORP proteins have RNase P activity

RNase P is the endonucleolytic activity resonsible for the 5' end maturation of tRNAs. More than 30 years ago Sidney Altman described *E.coli* RNase P as a ribonucleoprotein (RNP) RNase P with an RNA having the catalytic activity (Guerrier-Takada *et al.*, 1983). The discovery of catalytic RNA, also called ribozymes, brought him and Tom Cech the Nobel prize in chemistry in 1989. RNase P was quickly characterized as a complex of proteins and RNA in different prokaryotes and numerous eukaryotes except for animal mitochondria and plants. In 2008, human mitochondrial RNase P was characterized as a complex composed of three proteins devoid of RNA, that are essential *in vitro* (Holzmann *et al.*, 2008).

In plants no protein subunit specific for RNase P and RNase P RNA could ever be detected in the nuclear or the organellar genome. However, three orthologues of one of the human RNase P subunits are encoded in *A. thaliana* genome and are localized to mitochondria and chloroplasts (PRORP1) and to the nucleus (PRORP2-3) (Gobert *et al.*, 2010).

4.2.1.2 PRORP proteins are essential

If PRORP1 is the only RNase P activity in *Arabidopsis* organelles its functionality must be essential. This hypothesis was confirmed by *in vivo* experiments on plants lacking the prorp1 gene. Plants were embryo-lethal i.e. the embryo died at the globular stage (Gobert *et al.*, 2010), whereas only homozygous double mutants of prorp2-3 were not viable supporting the idea that they are redundant enzymes (Gutmann *et al.*, 2012).

4.2.1.3 PRORP proteins are the only RNase P enzymes in Arabidopsis

RNase P has some common protein subunits with RNase MRP in yeast and animals. It was thus believed that RNase P activity was redundant and that RNase MRP could also have acquired the capacity of cleaving precursor tRNAs (Krehan *et al.*, 2012). It was shown that downregulating RNase MRP proteins POP1 and POP4 only changed rRNA processing while tRNA processing remained untouched (Gutmann *et al.*, 2012).

4.2.2 Do PRORP proteins hold an original mode of action among PPR proteins?

4.2.2.1 A novel category of PPR proteins

PPR proteins are subdivided into two main families: the P type present in all eukaryotes and the PLS type present only in plants. The PLS family distinguishes itself from the P family by the occurrence of additional C-terminal domains potentially holding enzyme activities (Lurin *et al.*, 2004). PRORP defines yet another subfamily of PPR proteins having only P type PPR motifs and being fused to a catalytic domain, i.e. a nuclease. Another protein that does not fit to the classical P and PLS types of PPR proteins is the human mitochondrial RNA polymerase.

4.2.2.2 Recognition mode of RNA by PRORP and a minimal substrate

It has been shown that PPR proteins bind single stranded RNA and that they recognize the target RNA in a modular fashion with one repeat specifically binding one nucleotide (Barkan *et al.*, 2012, Barkan & Small, 2014). Specific binding of *maize* PPR10 to PSAJ RNA has been shown in a crystallographic structure (Yin *et al.*, 2013). Contrary to other PPR proteins, it seems that PRORP proteins bind structured RNA. They do not seem to bind a contiguous sequence of unpaired nucleotides but rather nucleotides that are distant in the sequence but in close proximity in the RNA 3D structure. The low binding affinity of PRORP proteins could be explained by the fact that they must bind all tRNAs in the same manner and that they appear to interact with only few unpaired nucleotides in the D/T loop of the tRNA (Gobert *et al.*, 2013). They might also act as a molecular ruler that recognizes overall tRNA-like structures (Gutmann *et al.*, 2012). A crystallographic structure of the PRORP/tRNA complex will reveal at atomic resolution how PRORP proteins recognize their RNA substrates.
Studying PRORP/tRNA interactions is also interesting from an evolutionary point of view, i.e. to determine if PRORP represents a structural mimicry of RNP RNase P and thus recognizes its substrates

the same way as RNP RNase P. In this sense it is interesting to determine the minimal substrate recognized and cleaved by PRORPs. *In vitro* cleavage assays show that PRORPs cleave substrates lacking the anticodon arm, possessing only the D/T-arm which contain conserved nucleotides that have been shown to be protected by the enzyme (Gobert *et al.*, 2010). More precisely, conserved nucleotides G18/G19 and C56 that turned out to be protected in the footprint assays were also shown to be in contact with the RNP RNase P RNA (Reiter *et al.*, 2010).

4.2.3 RNase P evolution

4.2.3.1 The loss of the catalytic RNA

RNase P activity in eukaryotes is performed either by a conventional RNP enzyme or a protein-only complex. RNP RNase P emerged long before the first eukaryote appeared. The prokaryotic genome encodes for the RNase P RNA as well as for its small protein subunit. Early mitochondria that evolved from the acquisition of an α-proteobacterium, likewise encoded a RNase P RNA as still observed in some eukaryotes such as yeast. Still, in eukaryotic nuclei the RNase P composition is far more complex with up to ten protein subunits completing the RNP complex.
Proteinaceous RNase P is present in distantly related eukaryotes and should therefore have appeared in early history of this kingdom.
Yeast, animals and plants could represent three different evolutionary steps of the transition from an organism only using RNP RNase P to an organism only using proteinaceous RNase P. Several hypotheses exist: 1) In organisms where the RNase P RNA gene is absent from mitochondrial genomes, it is possible that this gene was lost during genome recombination or transferred to the nucleus which is a common phenomenon. Then a nuclear encoded nuclease already present in the mitochondrial proteome might have been recruited for RNase P activity. 2) It is also possible that the two types of RNase P co-existed in one compartment albeit a evolutionary proof is missing.
Proteinaceous RNase P composition is also diverse, composed of single proteins in plants and three subunits in human mitochondria. This could be due to the presence of non-canonical tRNAs in human mitochondria that would need special recognition patterns. It was shown recently that by Rossmanith *et al.* (article in press) that human PRORP could not complement the yeast nuclear RNP RNase P whereas all homologues of *Arabidopsis* and *Trypanosoma* PRORP proteins gave rise to viable cells remarkably with unaltered changes in tRNA maturation levels. That human PRORP lacks this capability could indeed be due to its adaptation to the non-canonical form of mitochondrial tRNA and the requirement of additional partners to perform RNase P activity.
Still, even though *Arabidopsis* PRORPs function as single proteins, contrary to human PRORP, it appears that *Arabidopsis* PRORPs are also associated to larger complexes as suggested by ongoing research performed in our laboratory. The functional relevance of these complexes and their precise composition remain to be identified.

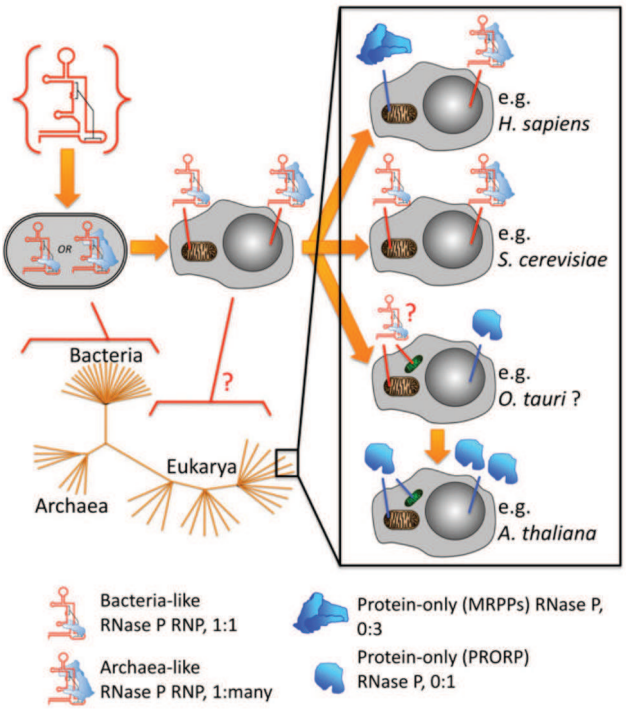


Figure 4.19: RNase P diversity. Phylogenetic repartition of RNase P in bacteria, archeae and eukaryotes. P RNA is represented in red and protein subunits, either for RNP RNase P or PRORP are colored in blue. Numerical ratios are printed below the figure (Goldfarb *et al.*, 2012).

4.2.3.2 Evolutionary diversity of RNase P

A diversity of developments maintained RNase P activity in eukaryotes. In plants and e.g. *Trypanosoma* and animal mitochondria the classical RNP RNase P was replaced by PRORP which likely recognizes its substrates with its PPR domain and potentially with additional proteins (Holzmann *et al.*, 2008, Gobert *et al.*, 2010, Taschner *et al.*, 2012).
In archaea, yeast and animal nuclei the P RNA lost essential capacities to cleave its substrates *in vivo* and is completed with up to 10 additional proteins to stabilize the active complex. In yeast mitochondria, even though RNase P consists of a ribozyme and one additional protein, RNase P composition deviates considerably from the bacterial counterparts. The P RNA does not resemble the bacterial one and the sole protein subunit is nuclear encoded and unrelated to bacterial RNase P proteins (Daoud *et al.*, 2012).
In many fungi mitochondria the RNase P activity has not been identified and it remains unclear whether a third type, still unknown, of RNase P exists in these species.

4.2.4 Integration of PRORP enzymes in a network of RNA expression and regulation

RNase P activity is indispensable for tRNA maturation and cell viability in all organisms except for the archaeon *Nanoarchaeum equitans* where tRNAs are transcribed with a mature 5' end (Randau *et al.*, 2008). Other enzymes like RNase Z, CCAse, enzymes implicated in nucleotide modifications

and splicing are also necessary to obtain functional tRNAs. It is tempting to think that all these factors form a complex for tRNA maturation and processing.

Such a complex has already been proposed in yeast mitochondria containing the RNase P, RNase Z, RNA degradosome and rRNAs from the small and big ribosomal subunits (Daoud *et al.*, 2012).

In plant nuclei PRORPs seem to be associated with POP1, an RNase MRP protein, thus suggesting that RNase MRP and RNase P are present in a same complex (Krehan *et al.*, 2012). PRORP1 was identified concentrated in different zones of the chloroplasts (Gobert *et al.*, 2010) serving as an RNase P reservoir or representing a zone where gene expression takes place.

Future and ongoing research will reveal the integration of PRORP functions among other cellular process in both organelles and the nucleus and reveal its precise mode of action. This should give clues to understand how living systems evolved to replace ribonucleoproteins by protein-only enzymes.

Annexes

E.1 Publication 3: Helical repeats modular proteins are major players for organelle gene expression

Biochimie 100 (2014) 141–150

Contents lists available at ScienceDirect

Biochimie

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

Review

Helical repeats modular proteins are major players for organelle gene expression

Kamel Hammani, Géraldine Bonnard [1], Ayoub Bouchoucha [1], Anthony Gobert [1], Franziska Pinker [1], Thalia Salinas [1], Philippe Giegé*

IBMP-CNRS, UPR2357, Strasbourg F-67084, France

ARTICLE INFO

Article history:
Received 11 July 2013
Accepted 30 August 2013
Available online 8 September 2013

Keywords:
Pentatricopeptide repeat
Octotricopeptide repeat
Half a tetratricopeptide
Mitochondrial transcription termination factor
Modular proteins

ABSTRACT

Mitochondria and chloroplasts are often described as semi-autonomous organelles because they have retained a genome. They thus require fully functional gene expression machineries. Many of the required processes going all the way from transcription to translation have specificities in organelles and arose during eukaryote history. Most factors involved in these RNA maturation steps have remained elusive for a long time. The recent identification of a number of novel protein families including pentatricopeptide repeat proteins, half-a-tetratricopeptide proteins, octotricopeptide repeat proteins and mitochondrial transcription termination factors has helped to settle long-standing questions regarding organelle gene expression. In particular, their functions have been related to replication, transcription, RNA processing, RNA editing, splicing, the control of RNA turnover and translation throughout eukaryotes. These families of proteins, although evolutionary independent, seem to share a common overall architecture. For all of them, proteins contain tandem arrays of repeated motifs. Each module is composed of two to three α-helices and their succession forms a super-helix. Here, we review the features characterising these protein families, in particular, their distribution, the identified functions and mode of action and propose that they might share similar substrate recognition mechanisms.

1. Introduction

Mitochondria, as well as chloroplasts in photosynthetic organisms, are considered as semi-autonomous organelles because they contain a genome inherited from their prokaryotic ancestor. Organelle genomes only encode a tiny fraction of mitochondrial and plastidial proteins, e.g. 13 proteins in human mitochondria, 32 and 79 in Arabidopsis mitochondria and chloroplasts respectively [1–3], thus representing merely an estimated circa 1% of the respective proteomes [4]. However these proteins are essential as they have critical functions for fundamental cellular processes such as respiration or photosynthesis. For their biogenesis, organelles require complete gene expression machineries comprising a transcription apparatus, enzymes responsible for posttranscriptional maturations, modifications, and processing of RNA as well as a fully functional translation system. Organelle gene expression has been extensively studied because of its specific features when compared to nuclear gene expression [5–7]. In some instances, enzymes and processes were inherited from the bacterial ancestor of organelles, e.g. as observed with the bacterial-type PEP RNA polymerase in chloroplasts [8]. However, in most cases, RNA maturation processes as well as the enzymes involved seem to have evolved specifically during eukaryote history, thus making them both organelle specific and eukaryote specific [9]. For example, among these processes, RNA editing restores coding sequences through the insertion and deletion of uridines in trypanosome mitochondria [10] and through cytidine to uridine modifications at hundreds of sites in plant organelles [11,12]. The nature of most factors involved in organelle specific gene expression processes has remained enigmatic for a long time. Given the limited content of organelle genomes, it was obvious that nearly all these factors were encoded in the nucleus and imported into organelles [13].

Putative organelle specific gene expression factors have been searched among recognized RNA binding protein families. Common RNA binding domains include RRM, KH, OB, zinc fingers or Rossmann folds. Their interactions with RNA involve very diverse strategies such as stacking or electrostatic interactions, hydrogen or van der Waals bonding [14]. For instance, RRM proteins bind RNA through β-sheet surface interactions. KH proteins use hydrophobic

* Corresponding author. Institut de Biologie Moléculaire des Plantes du CNRS, 12 rue du général Zimmer, 67084 Strasbourg, France. Tel.: +33 3 67 15 53 63.
E-mail address: giege@unistra.fr (P. Giegé).
[1] The authors are listed in alphabetical order.

0300-9084/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.biochi.2013.08.031

clefts formed by their structure whereas Zinc fingers domains can bind RNA by means of precise residues in α-helices [14]. The involvement of proteins belonging to these families has indeed been identified for organelle gene expression processes. Just to mention a few examples, a family of RRM containing RNA binding proteins was found to be specific to plant mitochondria [15]. MP42, a component of the RNA editing complex in trypanosomes contains both zinc fingers and an OB fold [16] and APO1, a chloroplast splicing factor, contains a zinc-finger-like RNA binding domain [17]. Still, most RNA related processes in organelles could not be related to the function of classical RNA binding proteins.

Apart from the aforementioned protein families, a growing list of nucleic acid binding protein families based on tandem arrays of repeated motifs folded into α-helices is being described. Among them, PUF and TALE protein families have attracted considerable attention [18,19]. PUF proteins are specific from eukaryotes. Their function is often related to developmental control through activation or repression of translation [20]. They contain tandem arrays of 36 amino acids repeats folded into three α-helices. The succession of repeats forms a solenoid that makes a platform for interaction with RNA [21]. Interestingly, each repeat specifically binds a single nucleotide of the RNA target according to a precise recognition code [22,23]. On the other hand, TALE proteins are DNA binding proteins found in bacterial pathogens of plants such as Xanthomonas. They act as transcription factors in plant nuclei to hijack their host cell gene expression [24]. These proteins contain repeats of 34 amino acids folded into two α-helices. Here as well, a precise recognition code could be established between single bases of DNA and individual TALE repeats [25]. The modular nature of both TALE and PUF proteins and the understanding of their mode of action have enabled to engineer recombinant proteins binding RNA or DNA targets of interest [23,26].

It has become increasingly evident that other families of proteins with similar modular architectures of repeated helical motifs are most of the times involved in organelle gene expression. These families include pentatricopeptide repeat proteins (PPR), half a tetratricopeptide proteins (HAT), octotricopeptide repeat proteins (OPR) and mitochondrial transcription termination factors (mTERF). Here, we review identified functions for these families of proteins found in plastids and mitochondria across the entire eukaryote lineage and discuss potential common modes of target recognition processes.

2. PPR proteins

2.1. Discovery and distribution of the PPR gene family

The identification of the PPR protein family has been directly associated with the Arabidopsis Genome Initiative [27]. The release of the first plant genome revealed the existence of a novel gene family whose large size and diversity appeared to be unique to flowering plants. Aubourg et al. identified a family of over 200 proteins with amino-terminal domains characterized by repeated sequence motifs organized in a specific pattern. The family could also be distributed in three subfamilies based on carboxy-terminal domains [28]. The prevalence of these repeated motifs in proteins screened for their predicted organellar localization appeared to be a hallmark of this family [29]. These motifs are composed of a degenerated sequence of 35 amino acids (Fig. 1A) and were called "pentatricopeptide" because of their similarity with tetratricopeptide repeat (TPR) motifs, characterised by 34 amino acids domains and known to be involved in protein–protein interactions [29]. Each PPR motif is composed of two antiparallel α-helices. The succession of motifs (up to 26 in plant PPR protein described so far) is predicted to form a super-helix that includes a central groove containing residues with prevalent positive charges (Fig. 1), consistent with the RNA binding capacity of PPR proteins. The identification of PPR proteins was hindered by sequence degeneracy, by motif lengths variations as well as by the possible occurrence of gaps between motifs. However, the initial descriptions made by two groups in the early 2000's were merged to propose a unified terminology for the modular organisation of the Arabidopsis PPR family [30]. In addition to the canonical 35 amino acid motifs called PPR-P motifs (P for pure), two additional PPR-like motifs were identified: the PPR-S (for short) and the PPR-L (for long) motifs, as well as additional variants termed L2 and S2, that nonetheless all share the same overall helix-turn-helix fold characteristic of PPR motifs. The estimated 450 PPR proteins of Arabidopsis thus fall into two major subfamilies according to the nature of motifs present in the respective proteins. The P subfamily only contains P motifs usually arranged in tandem whereas the PLS subfamily is composed of proteins containing repetitions of P-L-S motifs triplets [30]. The PLS subfamily is also defined by the occurrence of non-PPR extensions at their C-terminal ends. According to the nature of these domains, PLS proteins were further divided into four sub-classes [30,31]. Three motifs, only present in PPR proteins, were identified in these C-terminal extensions, two glutamic acid rich domains were called E (91 amino acids) and E+ (33 amino acids) and a domain of 106 amino acids was named DYW in reference to its three conserved terminal residues. Proteins containing the largest extension comprising all three E, E+ and DYW domains define the DYW subgroup, whereas the sequential loss of one or two motifs defines the E+ and E subgroups respectively [31]. More recently the identification of protein-only RNase P called PRORP in Arabidopsis exemplified another type of organisation for PPR proteins as PRORP proteins contain a few PPR motifs in their N-terminal part and a NYN nuclease domain in their C-terminal half [32]. It is thus possible that other orphan proteins with different organisations, e.g. with few canonical PPR motifs and/or with highly degenerated PPR motifs associated to non-PPR domains have not yet been assigned to the huge PPR protein family in plants.

PPR proteins are typical from eukaryote. They are universally present in eukaryote genomes and are completely absent from prokaryotes with the exception of a few plant pathogens such as Ralstonia that most likely acquired PPR genes during horizontal gene transfer [30]. Among eukaryotes, Streptophyta as well as a number of protists [33] are the only group where the PLS subfamily occurs. In all other organisms, PPR proteins are limited to the P subfamily. Algorithms used to identify PPR motifs have been initially designed according to the Arabidopsis PPR family. This has introduced a bias that impedes the proper identification of all PPR proteins in non-plant organisms. However, new bioinformatics tool now begin to unravel previously unassigned PPR proteins, e.g. in yeast [34]. Still, the number of PPR proteins remains reduced in non-plant organisms. They are particularly limited in fungi and Metazoa with e.g. 7 PPR proteins in human [35], 9 in *Schizosaccharomyces pombe* [36] and 15 in *Saccharomyces cerevisiae* [34]. In all these organisms PPR proteins are mitochondria-localized. The parasitic protozoan *Trypanosoma brucei* constitutes an exception with 28 PPR proteins [37]. Most of them contain a predicted mitochondrial targeting sequence except three. The specificities of kinetoplastids mitochondrial gene expression might explain this relative expansion of the PPR family in this organism as found in plants.

2.2. Functional diversity identified for PPR proteins

Since their discovery, the list of functions attributed to PPR proteins has grown rapidly. These functions are nearly systematically related to gene expression in mitochondria (and in chloroplasts in

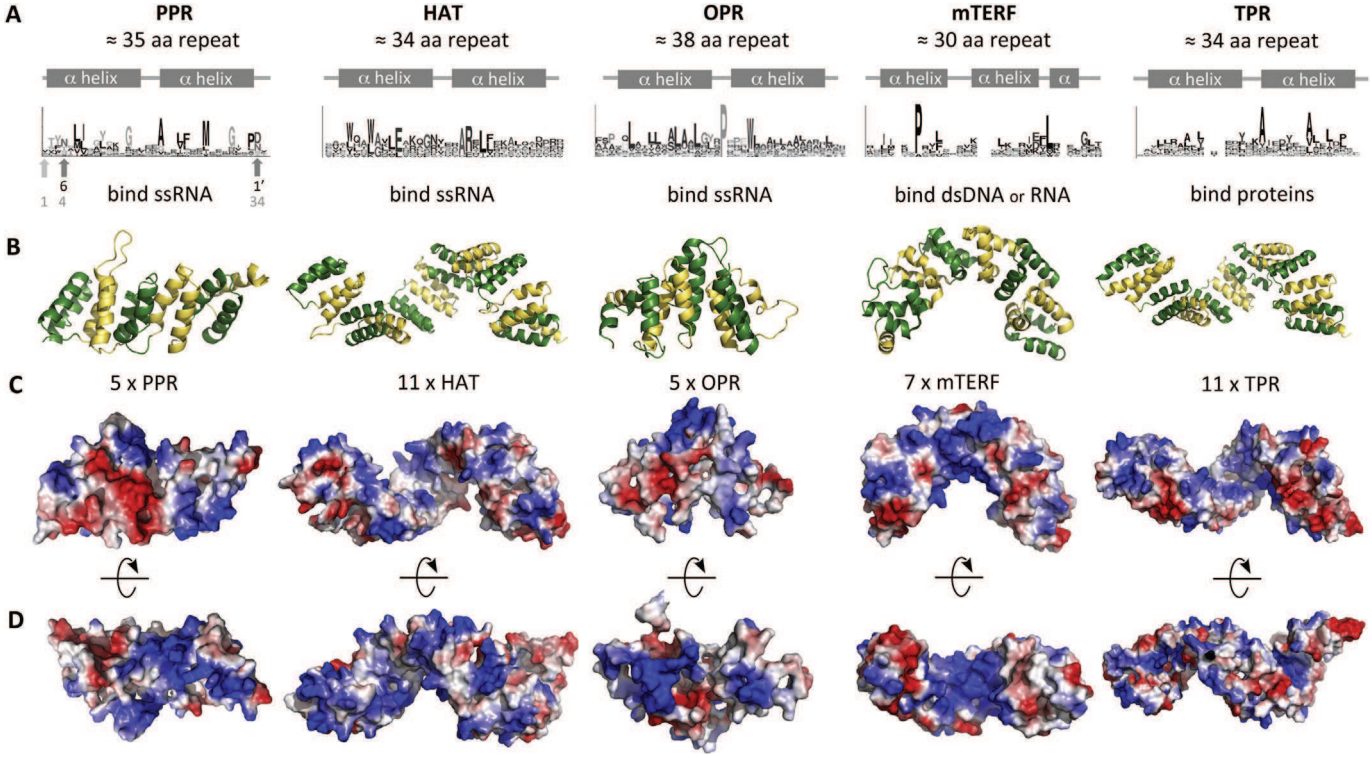


Fig. 1. Conserved organization and structural features of nucleic acid binding PPR, HAT, OPR and mTERF proteins compared to protein binding TPR proteins. (A) In all the respective families of proteins, individual repeats are folded into two to three α-helices, with individual motif sizes ranging from about 30 to 38 amino-acids. Weblogos representing conserved positions in the respective motifs were obtained with (http://weblogo.berkeley.edu/) [116]. The logo for PPR motifs was designed according to Filipovska and Rackham [18]. Black numbers represent positions that were shown to be important for RNA specificity according to the recognition code proposed by Barkan et al. [74]. Grey numbers show the same positions but with another nomenclature as described by Yagi et al. [75]. The logo for HAT motifs was derived from the 11 HAT motifs present in HCF107 [86]. The logo for OPR proteins was designed according to Rahire et al. [91]. The logo for mTERF motifs was derived from 16 motifs present in mTERF3 [99]. The logo for TPR proteins was designed according to Ref. [117]. (B) For each family of proteins, 3D structural models of different numbers of repeats were generated with the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/) [118] for domains of representative proteins (i.e. PPR protein PNM1, HAT protein HCF107, OPR protein TDA1, mTERF3 and TPR protein O-linked GlcNac transferase). Individual repeats are shown alternatively in green or in yellow. (C) Surface charges are displayed for the respective structural models with PyMOL [119] in the same orientation as in B. Blue indicates positive charges whereas red shows negative charges. (D) Models were rotated 90° around the horizontal axis, thus revealing extended positive charge patches in the inner groove of the super-helix for PPR, HAT, OPR and mTERF proteins in contrast with the TPR protein.

the case of plants). These functions go from transcription to translation and include most post-transcriptional RNA maturations steps (Table 1). Overall, the function of PPR proteins (or of PPR domains in multi-domains proteins) is connected with specific RNA sequence recognition and binding [38].

For mitochondrial transcription, some eukaryotes such as fungi or animals use an RNA polymerase containing PPR motifs [39,40], although the precise involvement of PPR motifs is not fully understood here [41]. After transcription, precursor RNA molecules undergo several maturation steps such as RNA splicing. Many plant PPR proteins were found to be essential for RNA splicing. The majority of these proteins belong to the P-subfamily, e.g. in *Arabidopsis* mitochondria, 3 PPR proteins were found to be involved in the splicing of *nad1* intron 1 (OTP43) [42], of *nad2* intron 3 (AB05) [43] and of *nad7* intron 1 (BIR6) [44]. Concerning chloroplasts, OTP70 [45] and OTP51 [46] are essential for the splicing of *rpoc1* and of *ycf3* intron 2, respectively. In yeast as well, DMR1, a PPR protein is required for cox1 and cob splicing [47]. In all cases the involvement of PPR proteins in splicing correlates with the increase of mitochondrial intron numbers in the respective species [9].

In contrast to splicing, the vast majority of PPR proteins involved in plant-type C to U RNA editing [12,48] were found to belong to the PLS subfamily. *Arabidopsis* mitochondrial and chloroplastic transcriptomes undergo 488 and 34 editing reactions respectively [11]. Since these sites do not share any conserved sequence or structure, it was predicted that their recognition would require a high number of specificity factors [12]. It has now become evident that PPR proteins are these specificity factors [38]. The first PPR protein for which a function in RNA editing was established is CRR4, a protein containing 11 PPR motifs and belonging to the E subfamily [49]. Since then, a long list of editing factors has been identified. Many of them belong to the DYW subgroup [9]. Interestingly, the DYW motif has sequence similarity with the active site of cytidine deaminases. Thus, this observation suggests that the DYW domain might act as the catalytic domain of the enzyme [50]. In addition, the number of editing sites seems to correlate well with the number of genes coding for PLS proteins [51]. This correlation is even stricter with DYW proteins [50]. Moreover, PPR proteins such as CRR4 that do not contain a DYW domain were shown to interact *in vivo* with DYW1, a protein that contains a DYW domain and no PPR repeats, thus making a protein complex resembling a DYW PPR protein [52]. Still, other non-PPR proteins of unidentified function were also found to be essential editing factors [53,54]. The precise involvement of DYW domains in the editing reaction thus remains to be established.

RNA maturation also involves processing steps of transcript ends. Among these processes, pre-tRNAs undergo 5′ maturation by an endonuclease activity called RNase P. In many eukaryotes such as human, plants and trypanosomes, this activity that was long though to be universally held by ribonucleoproteins, is performed by protein-only enzymes, i.e. by PPR proteins called PRORP [32,55–57]. Incidentally, in yeast mitochondria, ribonucleoprotein RNase P

Table 1
Assigned functions for PPR, OPR, HAT and mTERF proteins in different groups of eukaryotes. A subset of representative PPR, HAT and mTERF proteins are exemplarily presented for all identified functional classes known to date in the respective eukaryote groups. An exhaustive inventory of PPR proteins with identified functions has been presented by Gutmann et al. [9]. For OPR proteins, the complete list of proteins with identified functions is presented here. "mt" stands for mitochondrial.

	Functions	Organism	Protein	Target	Reference
PPR	Transcription	Fungi	Rpo41p	mt promoters	[40]
		Animals	POLRMT	mt promoters	[39]
		Plants	DG1	PEP	[120]
	RNA cleavage	Fungi	Rpm2p	tRNA and tRNA-like structures	[58]
		Animals	PRORP/MRPP3		[55]
		Plants	PRORP1, 2, 3		[32]
		Trypanosomids	PRORP1, 2		[57]
	RNA editing	Plants	OTP87 OTP81	nad7(Mt)/atp1(Mt) RNAs	[121,122]
	RNA processing	Fungi	Cbp1p	cob RNA	[123]
		Animals	LRPPRC/LRP130	cox1 and 3 RNA	[63,124]
	RNA splicing	Plants	OTP43	nad1 RNA	[42]
	RNA stability	Animals	BSF	bcd RNA	[125]
		Plants	PPR10	atpH/psaJ RNA	[59]
	RNA turnover	Fungi	Pet309p	cox1 RNA	[126]
		Mammals	PTCD1	Leu-tRNA	[127]
	Translation	Fungi	CY A-5	cox1 RNA	[128]
		Animals	MRPS27	mt ribosome	[129]
		Plants	CRP1	*petA* and *psaC* RNA 5′ UTR	[65]
		Trypanosomids	TbPPR5	rRNA	[37]
HAT	3′-end pre-mRNA cleavage, polyadenylation	Fungi	Rna14p	Nuclear mRNA	[130]
		Animals	CstF-77		[78]
		Plants	AtCstF-77		[131]
	Pre-mRNA splicing	Fungi	Clf1p	snRNPs	[132]
		Animals	Prp6	20S U5 snRNPs	[133]
		Plants	Sta1	pre-mRNA	[134]
	Pre-rRNA maturation	Fungi	Utp6p	rRNA	[85]
	mRNA stabilisation	Plants	HCF107	psbH RNA	[86,88]
OPR	Trans-splicing	Green algae	Rat2 Raa1	psaA RNA	[94,95]
	Translation		Tab 1 Tbc2 Tda1	psaB RNA psbC RNA atpA RNA	[91] [90] [93]
mTERF	Transcription termination	Animals	MTERF1	mt DNA	[101]
		Green algae	MOC1		[112]
	DNA replication	Animals	MTERF1, MTERFD1, MTERFD3	mt genome replication pause sites	[135] [136]
	RNA binding	Animals	MTERF4	rRNA	[104]

also involves a PPR protein called Rpm2 [58]. This protein is however completely unrelated to PRORP.

Other functions of PPR proteins do not appear to involve a catalytic reaction but rather use the ability of PPR proteins to stably bind precise RNA targets. Indeed, PPR proteins have been found to be required for the stabilization of transcripts. The molecular process by which this is achieved has first been proposed for PPR10 in chloroplasts. This protein specifically binds transcript termini and serves as a barrier from both 5′ and 3′ directions to RNA decay by exonucleases [59,60]. This process has also been described for plant mitochondria [61]. It might be present as well in other organisms such as *T. brucei*, where six PPR proteins are found to be required for the stabilization of mitochondrial rRNA [37]. Similarly, in *Chlamydomonas reinhardtii*, MCA1 regulates the stability of chloroplast *petA* mRNA [62].

Further studies exemplify as well the diversity of functions adopted by PPR proteins. For instance, in animals, LRPPRC is important for polyadenylation despite its inability to bind polyA tails and is also required for translation [63,64]. Similarly, the chloroplast PPR protein CRP1 is essential for the translation of *petA* and *psaC* mRNA. CRP1 binding site in the 5′ UTR region suggests that its function might be related to translation initiation [65]. In trypanosomes, PPR proteins stimulate mRNA adenylation and uridylation to activate mitochondrial translation [66]. Other PPR proteins were found to be associated to polysomes and might thus be involved in translation as well [67]. Finally, PNM1 was found to be dual localized to mitochondria and the nucleus and was proposed to be involved in the coordination of gene expression between mitochondria and the nucleus [68,69].

2.3. Mode of action of PPR proteins

PPR proteins are RNA binding proteins involved in processes that all require the accurate recognition of RNA targets. As described for PUF and TALE nucleic acid binding proteins, it had been predicted that PPR proteins would bind RNA target in a modular fashion, with each PPR repeat interacting via a few key residues with single nucleotides [38], thus making a recognition code between proteins residues and RNA moieties. The linear PPR array, alone or in association with other proteins, would then determine the sequence binding specificity. Both computational and experimental studies were performed to identify PPR residues that could be responsible for the specificity of interaction with the cognate RNA and to elucidate the code for RNA recognition by PPR proteins.

Based on the coevolution analysis of PPR proteins and their RNA targets, Fuji et al. studied the variability of individual positions in a subset of PPR proteins present in a locus encoding fertility restorer (Rf) proteins [70]. Cytoplasmic male sterility (CMS) originates from a defect in the expression of the mitochondrial genome that impedes the development of a functional male gametophyte. Specific nuclear fertility restorer genes prevent the accumulation of CMS-specific gene products. Rf genes encode mitochondrial PPR proteins, in most case of the class P, holding 15 to 20 PPR motifs. These genes show characteristic features that distinguish them from other PPR genes. In particular, the survey of 212 Rf-like genes from 13 different flowering plants revealed their rapid evolution [70]. The highest probability of diversifying selection (evaluated by non synonymous versus synonymous nucleotide substitution ratio) was found for residues 1, 3 and 6 of the PPR motif (PS51375 prosite

numbering which will be used thereafter). These residues were therefore proposed to be involved in RNA-binding specificity [70].

In another study, Kobayashi et al. used HCF152 a well-studied P-type chloroplastic PPR protein [71,72] to identify the RNA binding mechanism of PPR proteins [73]. Kobayashi et al. examined the RNA binding affinity of a series of mini PPR proteins containing two successive PPR motifs by mobility shift assays. Mutagenesis experiments and structural modelling suggested the involvement of amino acids located in two consecutive motifs: positions 3, 6, 10, 14 in helix A of PPR motif (n) and position 1′, the first position in helix A of PPR motif ($n + 1$). These were originally described as positions 1, 4, 8, 12 and 34 based on pfam numbering. However in the absence of a well-characterized RNA target Kobayashi et al. could not derive a conclusive connection between PPR motifs and RNA [73].

The first code proposed for PPR/RNA interaction emerged from a collaborative work that combined experimental and computational work based on well-characterized PPR/RNA partners [74]. Similar to HCF152, PPR10 is a P-type PPR protein involved in chloroplast transcript processing. PPR10 comprises 19 PPR repeats that recognize a 17-nt minimal RNA ligand presumably in a one to one mode. Among all arrangements of PPR10 motifs in contact with its RNA footprint, a single arrangement emerged because it showed strong correlations between the RNA base and the residues found at positions 6 and 1′. This combination between the two amino acids and RNA residues defines a code. For instance, ND, NN or NS specify pyrimidines while TD and TN specify purines. These rules could be applied to other P-type PPR proteins such as HCF152, CRP1 and their corresponding RNA partners. Furthermore Barkan et al. achieved the recoding of PPR10 (PPR motif 6 (ND) and 7 (NN), aligned to UC nucleotides) and validated the proposed code by mobility shift assay with diverse substituted RNA sequences. The contiguous P-type PPR/RNA duplexes evidenced were limited to nine motifs and eight nucleotides indicating that gaps can be tolerated or might result from structural conformation constraints. Actually, both RNA or protein gaps could interrupt the PPR and RNA duplex as described for P-PPR proteins. The simple two-amino acid code does not explain the diversity of amino acid observed at positions 6 and 1′ in canonical PPR motifs, indeed a third of the existing combinations could not be decoded. However when this code is applied to PLS arrays of RNA editing factors and their known RNA targets, an alignment can be found (with P- and S-PPR motif only) which positioned the last matching motif 6 nucleotides before the edited cytidine residue. It has been proposed that the non-matching L-PPR motifs could reduce the structural constraint between PLS-PPR and their targets for which no gaps in the alignment could be found.

In another attempt to understand how PPR proteins achieve RNA specificity, Yagi et al. searched low variability (conserved associations) between PPR proteins residues and nucleotides upstream of editing sites. A collection of 327 PLS motifs from 24 PPR proteins involved in RNA editing described in *Arabidopsis thaliana* were examined in alignment with their target RNA. Significant low variability was observed at residue 6 and 1′ (as proposed by Barkan et al. [74]) and also at residue 3 in the alignment in which the last PPR motif is aligned with the 4th nucleotide before the edited cytidine [75]. The combination of these three amino acids determines the recognition of a specific RNA sequence in a 1-motif to 1-nucleotide direct and contiguous way as expected in the case of PLS-PPR. According to Yagi et al., only a subset of L-type motifs participates in nucleotide recognition. Similarly, Takenaka et al. showed that the inclusion of L, L2 and S2 motifs to the recognition code improved the prediction of RNA editing target sites [76]. The proposed 3 amino-acids code appear to be efficient to predict editing sites from uncharacterized PPR proteins or for accurate assignment of targets sites from the whole organelle transcriptome. The codes described by the different groups although established with either P-type or PLS PPR proteins are almost identical which strongly support the existence of a universal code for PPR proteins.

3. HAT proteins

The existence of the so-called Half-a-Tetratricopeptide repeat (HAT) proteins has long been overshadowed because of their strong similarity with their relatives Tetratricopeptide repeat (TPR) proteins. It is only in 1998 that Preker and Keller identified these proteins as belonging to a distinctive family (IPR003107) [77]. Members of the HAT family harbour an HAT domain signature generally arranged in tandem repetitions of variable numbers. The HAT motif is similar to the Tetratricopeptide repeat and consists of a 34 amino-acid degenerate sequence folding into two anti-parallel α-helices [78] The HAT repeats stack to form an elongated structure similar to that of TPR proteins (Fig. 1). However, the HAT domain shows several conserved residues that are absent in the TPR motif.

The HAT family in eukaryotes comprises ~10–15 members that are well conserved among species. They localize in both organelles and the nuclear/cytoplasmic compartment where they play various functions related to RNA metabolism including pre-rRNA maturation, pre-mRNA splicing and cleavage and polyadenylation [77]. Prominent examples include mammalian CstF-77 involved in 3′-end pre-mRNA cleavage and polyadenylation [79–81], the *Drosophila* crooked-neck protein Crn involved in pre-mRNA splicing [82,83] and yeast Utp6 involved in pre-rRNA processing [84]. Despite the elucidation of the physiological roles for many HAT proteins, in most of cases, the function of the HAT repeat is unknown. Based on its similarity to the TPR domain, the HAT domain has been proposed to play an analogous role in protein scaffolding. The facts that CstF-77 can form a homodimer [78] and that a peptide ligand has been found for Utp6 have comforted this notion [85]. Nevertheless, the functional specialisation of the HAT family in RNA metabolism and the observation that HAT repeats are often found in ribonucleoprotein complexes raised the possibility that HAT domains may bind RNA. The first clues for the biochemical function of HAT domains came from a study on a member of the plant HAT family, HCF107 [86]. In the green lineage, few HAT members were co-opted to regulate gene expression in the chloroplasts. The HAT motif is best known as R-TPR in the plant field, in regards to its similarity to the TPR domain and its functional specialization in RNA metabolism [87]. *HCF107* encodes a protein solely made of 11 HAT repeats and is the genetic determinant for the accumulation of processed 5′-end *psbH* mRNAs in the chloroplast of higher plants [88]. Using recombinant HCF107 protein and *in vitro* assays, the authors showed that HCF107 binds single stranded RNA with specificity towards its genetically defined mRNA target. As a consequence, the specific binding of HCF107 to the *psbH* pre-mRNA promotes the formation of processed 5′-end transcripts by protecting the mRNA from 5′-3′ exoribonucleases trimming. Furthermore, the RNA binding of HCF107 influences the local mRNA structure in a way that enhances its translation. Analogous activities have been reported for some members of the PPR protein family [60,72] and it is worth noticing that these diverse effects on gene expression are simple ramifications of the passive, but potent RNA binding activity of a helical-repeat protein of sufficient length. In the green algae Chlamydomonas, *NAC2*, a gene that encodes a chloroplast-localized HAT protein has been genetically linked to the stabilisation of a specific processed mRNA in the chloroplast [89]. In light of Hammani et al. work, it is likely that NAC2 plays a function analogous to that of HCF107 and binds the 5′ end of the mRNA target it stabilizes *in vivo*. Although, direct evidence that the HAT domain binds RNA has only been given to HCF107, the RNA binding

and RNA remodelling are two key properties that could contribute to the many physiological functions that were associated to the HAT domain outside of the organelles in non-plant species.

Incremental work on members of the HAT family will certainly contribute to understand the mechanism used by these proteins to influence RNA metabolism. The HAT domain shows strong functional and structural analogies to the PPR domain. In fact, it has been observed as for PPR proteins that the minimal RNA sequence bound by HCF107 matches the number of helical-repeats in the protein [86]. This observation suggests that each repeat targets a specific RNA base and thus, implies that a 1 HAT/1-nt code could be deciphered. To reach this goal, further mechanistic work on various members of the HAT family is needed.

4. OPR proteins

The Octatricopeptide repeat (OPR) proteins are defined by the presence of a set of related degenerate motifs of approximately 38–40 amino acids occurring as tandem arrays of 2–24 motifs per protein. The motif shows a typical degenerated consensus of five residues i.e. PPPEW at position 20–24 of the motif. The most conserved residues are the first Proline and the Tryptophan of the consensus but there is a Leucine at position 6 of the motif that is also well conserved (Fig. 1). Individual motifs are predicted to fold into two α-helices, thus assigning these proteins to the superfamily of proteins forming α-solenoids similar to PPR and HAT proteins.

The first OPR protein was identified in chloroplasts of the green algae *C. reinhardtii* ten years ago [90] and since then, only few OPR proteins have been studied. *In silico* analysis showed that OPR motifs are also found in the closely related alga, *Volvox carterii* but is mainly found in protozoans, such as apicomplexans, and in a small family of proteins in the parasitic alpha-proteobacterium *Coxiella burnetii*. In contrast, only few OPR genes were found in the genome of *A. thaliana* [91]. Analysis of the Chlamydomonas genome in addition to the previously reported OPR proteins allowed the identification of a total of 44 OPR proteins in this organism [92]. However, this number is probably underestimated because of the sequence degeneracy of the motifs that makes OPR proteins difficult to identify. Almost all the OPR proteins identified are predicted to localize to organelles. It is noteworthy that 16 out of the 44 identified OPR proteins contain one or more Fas-activated serine/threonine (FAST) kinase-like domains [93]. The protein family containing such domains is known to interact with proteins involved in RNA processing and translation [92]. In these proteins, the FAST domain is usually followed by a RAP domain (for RNA-binding domain abundant in Apicomplexans), which has been described as a putative RNA binding domain [92]. The RAP domain is found in 4 OPR proteins and in 3 cases the RAP domain follows the FAST kinase-like domains. Interestingly, the C-terminal part of RAP domains (where conservation is the highest) is homologous to the OPR motifs suggesting a common evolutionary origin [93].

In *C. reinhardtii*, 5 OPR proteins have been characterized, namely Rat2 with 2 OPR motifs [94], TAB1 with 10 OPR motifs [91], RAA1 with 14 OPR motifs [95], TBC2 with 16 OPR motifs [90] and TDA1 with 24 OPR motifs [93]. All these proteins have been shown to be involved in post-transcriptional steps of chloroplast gene expression. The Rat2 and Raa1 proteins are involved in the trans-splicing of *psaA* transcript. The Tab1, Tbc2 and Tda1 proteins are required for the translation of *psaB*, *psbC* and *atpA* transcripts, respectively. The RNA-binding capacity of OPR proteins was proposed for all of them. But the direct proof of this RNA-binding activity has been only demonstrated with the Tab1 protein [91]. Indeed, mobility shift assays between the maltose binding protein fused to the 9 OPR motifs of Tab1 and the *psaB* 5′ UTR showed an efficient binding of the protein to the RNA and with a relative binding specificity. The molecular mechanisms underlying RNA target recognition by the OPR motif are not yet known. However, similar to other modular proteins e.g. PPR proteins, it is tempting to speculate that a precise connection exists between individual OPR motifs and RNA moieties. In the future, an in depth biochemical and/or structural analysis of OPR proteins in chlorophyta but also in other eukaryotic lineages should determine if a code, such as the one described for PPR proteins, indeed also exists for OPR proteins.

5. mTERF proteins

The mTERF family is found in metazoan and plants and is characterized by the presence of a degenerated ~30 amino acids motif [96]. mTERF proteins have a modular architecture with variable numbers of mTERF repeats (Fig. 1). Each mTERF motif forms two antiparallel α-helices followed by a 3.10 helix [19,97]. The tandem mTERF repeats stack to form a super-helical protein adopting a “croissant” shape [97,98]. The mTERF family is apparently specific to organelles. Most of their members harbour an N-terminal sequence predicted to address them to mitochondria or the chloroplasts. GFP fusions and genetic analysis have indeed confirmed them to act in organelles [96,99,100]. The mTERF family was named after its founding member in mammals, MTERF1 which mediates mitochondrial transcription termination *in vitro* by binding to a specific DNA sequence between the 16S rRNA gene and the $tRNA^{Leu}$ gene [101]. The mTERF family in mammalian comprises 4 members that have been predominantly implicated in mitochondrial transcription and DNA replication. DNA binding activity has been reported both *in vivo* and *in vitro* for most mTERF proteins in animals (reviewed in Ref. [102]). In addition, the resolution of MTERF1 structure in complex with its DNA target has provided mechanistic insights into its termination effect on mitochondrial transcription. MTERF1 makes specific contacts with its DNA target sequence and allows base flipping from the DNA duplex, which in turn terminates transcription [97].

Altogether, these studies emphasize a global role for mTERFs in the fine-tuning of organellar gene expression and in DNA related functions [102,103]. Interestingly, recent studies have shaded this restrictive idea by adding RNA-related functions to the functional repertoire of this family. MTERF4 is a mitochondrial translation regulator in metazoan. It interacts with a mitochondrial rRNA methyltransferase and targets it to the large subunit of the mitochondrial ribosome. MTERF4 is found in association with rRNA in mitochondria [104]. The solved crystal structure of MTERF4/NSUN4 complex proposes a model where the C-terminal end of MTERF4 interacts with NSUN4 and the other free end would form a positively charged platform that binds rRNA [105,106]. The notion that the mTERF motif can bind RNA substrate is further supported by *in vitro* work showing that the metazoan mitochondrial translation regulator, MTERF3, when expressed as recombinant protein interacts with RNA [107].

The number of mTERF members has increased during the evolution of land plants, with ~30 members in higher plants but the function of very few of them has been studied and even then, not in details [96,99,100]. It is anticipated that these proteins like their animal homologs contribute to the maintenance of gene expression in mitochondria or chloroplast and thus, would play essential roles in respiration and photosynthesis [108]. Indeed, reverse genetics screen for Arabidopsis mTERF genes revealed a high proportion of embryolethal phenotypes [99,109,110]. Most of the *mterf* mutants studied in green organisms were recovered during forward genetic screen aiming at the identification of specific nuclear genes involved in the response to abiotic stress or involved in the embryo-development [100].

In the green algae Chlamydomonas, the mTERF member MOC1 is a mitochondrial protein and the loss of MOC1 enhances the light sensitivity of the mutant [111]. MOC1 binds specifically to a sequence in the mitochondrial genome and acts as a transcription terminator by altering transcription read-through [112].

In Arabidopsis, mutant alleles have been recovered and studied for only 5 mTERF genes. A mutation in the Arabidopsis gene *SOLDAT10*, encoding a plastid-localized mTERF protein, abolishes $^1O_2$-mediated-cell death in the Arabidopsis flu mutant, which usually accumulates $^1O_2$ during a dark-to-light shift [109]. The mutant plant *soldat10* shows pigmentation defect and chloroplast translation impairment. Similarly, a mutation in another mTERF-related gene *RUGOSA2*/*BSM*, whose product is plastid and mitochondria localized, causes pigment and morphological defects with a translation defect in chloroplasts [99,113]. MDA1, another Arabidopsis chloroplast-localized mTERF is involved in the response to abiotic stress and chloroplast development [110]. Like others, *mda1* suffers morphological and pigment defects. *SHOT1* is the only Arabidopsis gene encoding a mitochondrial mTERF product that has been characterized [114]. Loss of SHOT1 suppresses the heat sensitivity in other mutants and confers thermotolerance.

In all these mutants, changes in the steady state level of mitochondrial and/or chloroplastic transcripts have been reported but the basis for these changes was unknown. A role in chloroplastic mRNA splicing for BSM has been suggested [99]. However, a clear link between the phenotype observed in the mutants and the molecular defect has not been firmly established.

The study of a maize gene, Zm-mTERF4 which is orthologous to *BSM*/*RUGOSA2,* provides the first details about the precise molecular function of an mTERF protein in higher plants (Hammani and Barkan, under-review). Zm-mTERF4 localizes to the chloroplasts and Zm-*mterf3* seedlings display typical non-photosynthetic phenotypes and die after the development of three leaves. Zm-mTERF3 binds chloroplast group II introns and promotes the splicing of many of them *in vivo*. The loss of splicing for some tRNAs and ribosomal protein encoding genes lead to the impairment of plastid translation in Zm-*mterf3*. A global loss of plastid translation was similarly observed in a *bsm* mutant [99]. Conservation of function for orthologous factors involved in organellar gene expression, between monocots and dicots, is well documented and it is likely that BSM in Arabidopsis influences the splicing of chloroplastic tRNAs and that this defect is implicated in the loss of translation observed in *bsm* chloroplasts. The loss of plastid translation further leads to embryolethal embryolethality in dicotyledon species like Arabidopsis as observed for *bsm*.

The discovery that metazoan and plant mTERFs are involved in RNA or DNA metabolism in the organelle suggests that the other ~30 mTERF proteins in plants could be involved in any steps of gene expression related to RNA or DNA in the chloroplast or mitochondria. The versatility of the mTERF motif to accommodate DNA and RNA substrates is very interesting in term of functional evolution. Further functional and mechanistic studies will enable to discover the molecular process by which mTERF proteins achieve substrate specificity and how they discriminate DNA and RNA targets.

6. Concluding remarks

Genomic and functional investigations have revealed that many organelle specific gene expression processes are performed by proteins encoded by recently recognized gene families such as the PPR, HAT, OPR and mTERF families. The evolution of these families does not appear to be connected. For instance, consensus motifs derived from the respective protein families do not seem to share common sequence signatures (Fig. 1A). However, all these proteins have a common modular organization, with a tandem arrangement of individual repeats that all share a similar secondary structure based on antiparallel α-helices (Fig. 1B). For all these families, the succession of repeats forms a solenoid structure, thus assigning these proteins to a super-family of proteins of common structural organization also including PUF and TALE proteins [18,19]. Interestingly, the comparison of structural models for representative proteins from all these families suggest that they might all possess nucleic acid binding platforms composed of positively charged amino acids in the concave surface of the super-helix (Fig. 1D). This feature has already been proposed for PPR [115] and mTERF proteins [105] and has been shown for PUF proteins [21]. In this light, it is tempting to speculate that all these families might share related target recognition processes. Up to now, among the aforementioned organellar proteins, the PPR family has clearly attracted the most attention. Its mechanism for specific target selection has been identified with the description of a PPR code connecting precise residues in PPR motifs and individual nucleotides [74–76]. Here again, an appealing possibility is that similar codes might also exist for HAT, OPR and mTERF proteins. Future investigations will clarify this and determine whether general rules for substrate recognition can be derived for the entire super-family of helical repeats proteins.

Acknowledgements

This work was supported by the French Centre National de la Recherche Scientifique and by the University of Strasbourg. Authors were supported by an ANR Blanc research grant “PRO-RNase P, ANR 11 BSV8 008 01” and by the LabEx consortium “MitoCross”.

References

[1] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, Nature 290 (1981) 457–465.
[2] M. Unseld, J.R. Marienfeld, P. Brandt, A. Brennicke, The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides, Nat. Genet. 15 (1997) 57–61.
[3] S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, S. Tabata, Complete structure of the chloroplast genome of *Arabidopsis thaliana*, DNA Res. 6 (1999) 283–290.
[4] S. Da Cruz, P.A. Parone, J.C. Martinou, Building the mitochondrial proteome, Exp. Rev. Proteomics 2 (2005) 541–551.
[5] O. Rackham, T.R. Mercer, A. Filipovska, The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression, Wiley Interdiscip. Rev. RNA 3 (2012) 675–695.
[6] D.B. Stern, M. Goldschmidt-Clermont, M.R. Hanson, Chloroplast RNA metabolism, Annu. Rev. Plant Biol. 61 (2010) 125–155.
[7] A.C. Barbrook, C.J. Howe, D.P. Kurniawan, S.J. Tarr, Organization and expression of organellar genomes, Philos. Trans. R Soc. Lond B Biol. Sci. 365 (2010) 785–797.
[8] K. Liere, A. Weihe, T. Borner, The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation, J. Plant Physiol. 168 (2011) 1345–1360.
[9] B. Gutmann, A. Gobert, P. Giegé, Mitochondrial genome evolution and the emergence of PPR proteins., advances in botanical research, in: L. Drouard (Ed.), Vol. MItochondrial Genome Evolution, Elsevier, 2012, pp. 253–313.
[10] L. Simpson, S. Sbicego, R. Aphasizhev, Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business, RNA 9 (2003) 265–276.
[11] B. Castandet, A. Araya, RNA editing in plant organelles. Why make it easy? Biochemistry (Mosc) 76 (2011) 924–931.
[12] P. Giege, A. Brennicke, RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 15324–15329.
[13] J. Dudek, P. Rehling, M. van der Laan, Mitochondrial protein import: common principles and physiological networks, Biochim. Biophys. Acta 1833 (2013) 274–285.
[14] B.M. Lunde, C. Moore, G. Varani, RNA-binding proteins: modular design for efficient function, Nat. Rev. Mol. Cell Biol. 8 (2007) 479–490.
[15] M. Vermel, B. Guermann, L. Delage, J.M. Grienenberger, L. Marechal-Drouard, J.M. Gualberto, A family of RRM-type RNA-binding proteins specific to plant mitochondria, Proc. Natl. Acad. Sci. U. S. A. 23 (2002) 23.
[16] M. Brecht, M. Niemann, E. Schluter, U.F. Muller, K. Stuart, H.U. Goringer, TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo–exoribonuclease activity, Mol. Cell 17 (2005) 621–630.

[17] K.P. Watkins, M. Rojas, G. Friso, K.J. van Wijk, J. Meurer, A. Barkan, APO1 promotes the splicing of chloroplast group II introns and harbors a plant-specific zinc-dependent RNA binding domain, Plant Cell 23 (2011) 1082–1092.
[18] A. Filipovska, O. Rackham, Modular recognition of nucleic acids by PUF, TALE and PPR proteins, Mol. Biosyst. 8 (2012) 699–708.
[19] E.H. Rubinson, B.F. Eichman, Nucleic acid recognition by tandem helical repeats, Curr. Opin. Struct. Biol. 22 (2012) 101–109.
[20] M. Wickens, D.S. Bernstein, J. Kimble, R. Parker, A PUF family portrait: 3′UTR regulation as a way of life, Trends Genet. 18 (2002) 150–157.
[21] T.A. Edwards, S.E. Pyle, R.P. Wharton, A.K. Aggarwal, Structure of Pumilio reveals similarity between RNA and peptide binding motifs, Cell 105 (2001) 281–289.
[22] X. Wang, J. McLachlan, P.D. Zamore, T.M. Hall, Modular recognition of RNA by a human pumilio-homology domain, Cell 110 (2002) 501–512.
[23] A. Filipovska, O. Rackham, Designer RNA-binding proteins: new tools for manipulating the transcriptome, RNA Biol. 8 (2011) 978–983.
[24] J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function, Annu. Rev. Phytopathol. 48 (2010) 419–436.
[25] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, Breaking the code of DNA binding specificity of TAL-type III effectors, Science 326 (2009) 1509–1512.
[26] D. Hockemeyer, H. Wang, S. Kiani, C.S. Lai, Q. Gao, J.P. Cassady, G.J. Cost, L. Zhang, Y. Santiago, J.C. Miller, B. Zeitler, J.M. Cherone, X. Meng, S.J. Hinkley, E.J. Rebar, P.D. Gregory, F.D. Urnov, R. Jaenisch, Genetic engineering of human pluripotent cells using TALE nucleases, Nat. Biotechnol.. 29 (2011) 731–734.
[27] T.A.G. Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 (2000) 796–815.
[28] S. Aubourg, N. Boudet, M. Kreis, A. Lecharny, *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants, Plant Mol. Biol. 42 (2000) 603–613.
[29] I.D. Small, N. Peeters, The PPR motif – a TPR-related motif prevalent in plant organellar proteins, Trends Biochem. Sci. 25 (2000) 46–47.
[30] C. Lurin, C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyere, M. Caboche, C. Debast, J. Gualberto, B. Hoffmann, A. Lecharny, M. Le Ret, M.L. Martin-Magniette, H. Mireau, N. Peeters, J.P. Renou, B. Szurek, L. Taconnat, I. Small, Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, Plant Cell 16 (2004) 2089–2103.
[31] E. Rivals, C. Bruyere, C. Toffano-Nioche, A. Lecharny, Formation of the Arabidopsis pentatricopeptide repeat family, Plant Physiol. 141 (2006) 825–839.
[32] A. Gobert, B. Gutmann, A. Taschner, M. Gößringer, J. Holzmann, R.K. Hartmann, W. Rossmanith, P. Giegé, A single Arabidopsis organellar protein has RNase P activity, Nat. Struct. Molec. Biol. 17 (2010) 740–744.
[33] M. Rudinger, L. Fritz-Laylin, M. Polsakiewicz, V. Knoop, Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*, RNA (2011).
[34] K.A. Lipinski, O. Puchta, V. Surandranath, M. Kudla, P. Golik, Revisiting the yeast PPR proteins – application of an Iterative Hidden Markov model algorithm reveals new members of the rapidly evolving family, Mol. Biol. Evol. 28 (2011) 2935–2948.
[35] R.N. Lightowlers, Z.M. Chrzanowska-Lightowlers, Human pentatricopeptide proteins: only a few and what do they do? RNA Biol. 10 (2013) 1250–1255.
[36] I. Kuhl, L. Dujeancourt, M. Gaisne, C.J. Herbert, N. Bonnefoy, A genome wide study in fission yeast reveals nine PPR proteins that regulate mitochondrial gene expression, Nucl. Acids Res. 39 (2011) 8029–8041.
[37] M. Pusnik, I. Small, L.K. Read, T. Fabbro, A. Schneider, Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes, Mol. Cell Biol. 27 (2007) 6876–6888.
[38] C. Schmitz-Linneweber, I. Small, Pentatricopeptide repeat proteins: a socket set for organelle gene expression, Trends Plant Sci. 13 (2008) 663–670.
[39] V. Tiranti, A. Savoia, F. Forti, M.F. D'Apolito, M. Centra, M. Rocchi, M. Zeviani, Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database, Hum. Mol. Genet. 6 (1997) 615–625.
[40] A.L. Greenleaf, J.L. Kelly, I.R. Lehman, Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome, Proc. Natl. Acad. Sci. U. S. A. 83 (1986) 3391–3394.
[41] R. Ringel, M. Sologub, Y.I. Morozov, D. Litonin, P. Cramer, D. Temiakov, Structure of human mitochondrial RNA polymerase, Nature 478 (2011) 269–273.
[42] A.F. de Longevialle, E.H. Meyer, C. Andres, N.L. Taylor, C. Lurin, A.H. Millar, I.D. Small, The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana*, Plant Cell 19 (2007) 3256–3265.
[43] Y. Liu, J. He, Z. Chen, X. Ren, X. Hong, Z. Gong, ABA overly-sensitive 5 (ABO5), encoding a pentatricopeptide repeat protein required for cis-splicing of mitochondrial nad2 intron 3, is involved in the abscisic acid response in Arabidopsis, Plant J. 63 (2010) 749–765.
[44] A. Koprivova, C.C. des Francs-Small, G. Calder, S.T. Mugford, S. Tanz, B.R. Lee, B. Zechmann, I. Small, S. Kopriva, Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial nad7 transcripts, J. Biol. Chem. 285 (2010) 32192–32199.
[45] A.L. Chateigner-Boutin, C.C. des Francs-Small, E. Delannoy, S. Kahlau, S.K. Tanz, A.F. de Longevialle, S. Fujii, I. Small, OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1, Plant J. 65 (2011) 532–542.
[46] A.F. de Longevialle, L. Hendrickson, N.L. Taylor, E. Delannoy, C. Lurin, M. Badger, A.H. Millar, I. Small, The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in *Arabidopsis thaliana*, Plant J. 56 (2008) 157–168.
[47] J.I. Moreno, K.S. Buie, R.E. Price, M.A. Piva, Ccm1p/Ygr150cp, a pentatricopeptide repeat protein, is essential to remove the fourth intron of both COB and COX1 pre-mRNAs in *Saccharomyces cerevisiae*, Curr. Genet. 55 (2009) 475–484.
[48] J.M. Gualberto, L. Lamattina, G. Bonnard, J.H. Weil, J.M. Grienenberger, RNA editing in wheat mitochondria results in the conservation of protein sequences, Nature 341 (1989) 660–662.
[49] E. Kotera, M. Tasaka, T. Shikanai, A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts, Nature 433 (2005) 326–330.
[50] V. Salone, M. Rudinger, M. Polsakiewicz, B. Hoffmann, M. Groth-Malonek, B. Szurek, I. Small, V. Knoop, C. Lurin, A hypothesis on the identification of the editing enzyme in plant organelles, FEBS Lett. 581 (2007) 4132–4138.
[51] N. O'Toole, M. Hattori, C. Andres, K. Iida, C. Lurin, C. Schmitz-Linneweber, M. Sugita, I. Small, On the expansion of the pentatricopeptide repeat gene family in plants, Mol. Biol. Evol. 25 (2008) 1120–1128.
[52] C. Boussardon, V. Salone, A. Avon, R. Berthome, K. Hammani, K. Okuda, T. Shikanai, I. Small, C. Lurin, Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids, Plant Cell 24 (2012) 3684–3694.
[53] S. Bentolila, W.P. Heller, T. Sun, A.M. Babina, G. Friso, K.J. van Wijk, M.R. Hanson, RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) E1453–E1461.
[54] M. Takenaka, A. Zehrmann, D. Verbitskiy, M. Kugelmann, B. Hartel, A. Brennicke, Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5104–5109.
[55] J. Holzmann, P. Frank, E. Loffler, K.L. Bennett, C. Gerner, W. Rossmanith, RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme, Cell 135 (2008) 462–474.
[56] B. Gutmann, A. Gobert, P. Giegé, PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis, Genes Dev. 26 (2012) 1022–1027.
[57] A. Taschner, C. Weber, A. Buzet, R.K. Hartmann, A. Hartig, W. Rossmanith, Nuclear RNase P of *Trypanosoma brucei*: a single protein in place of the multicomponent RNA-protein complex, Cell Rep. 2 (2012) 19–25.
[58] M.J. Morales, Y.L. Dang, Y.C. Lou, P. Sulo, N.C. Martin, A 105-kDa protein is required for yeast mitochondrial RNase P activity, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 9875–9879.
[59] J. Pfalz, O.A. Bayraktar, J. Prikryl, A. Barkan, Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts, EMBO J. 28 (2009) 2042–2052.
[60] J. Prikryl, M. Rojas, G. Schuster, A. Barkan, Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 415–420.
[61] N. Haili, N. Arnal, M. Quadrado, S. Amiar, G. Tcherkez, J. Dahan, P. Briozzo, C. Colas des Francs-Small, N. Vrielynck, H. Mireau, The pentatricopeptide repeat MTSF1 protein stabilizes the nad4 mRNA in Arabidopsis mitochondria, Nucl. Acids Res. 41 (2013) 6650–6653.
[62] C. Loiselay, N.J. Gumpel, J. Girard-Bascou, A.T. Watson, S. Purton, F.A. Wollman, Y. Choquet, Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in *Chlamydomonas reinhardtii* chloroplasts, Mol. Cell Biol. 28 (2008) 5529–5542.
[63] F. Sasarman, C. Brunel-Guitton, H. Antonicka, T. Wai, E.A. Shoubridge, LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates post-transcriptional gene expression in mitochondria, Mol. Biol. Cell 21 (2010) 1315–1323.
[64] T. Chujo, T. Ohira, Y. Sakaguchi, N. Goshima, N. Nomura, A. Nagao, T. Suzuki, LRPPRC/SLIRP suppresses PNPase-mediated mRNA decay and promotes polyadenylation in human mitochondria, Nucl. Acids Res. 40 (2012) 8033–8047.
[65] C. Schmitz-Linneweber, R. Williams-Carrier, A. Barkan, RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5′ region of mRNAs whose translation it activates, Plant Cell 17 (2005) 2791–2804.
[66] I. Aphasizheva, D. Maslov, X. Wang, L. Huang, R. Aphasizhev, Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, Mol. Cell 42 (2011) 106–117.
[67] M. Uyttewaal, H. Mireau, M. Rurek, K. Hammani, N. Arnal, M. Quadrado, P. Giege, PPR336 is associated with polysomes in plant mitochondria, J. Mol. Biol. 375 (2008) 626–636.
[68] K. Hammani, A. Gobert, K. Hleibieh, L. Choulier, I. Small, P. Giegé, An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation, Plant Cell 23 (2011) 730–740.
[69] K. Hammani, A. Gobert, I. Small, P. Giegé, A PPR protein involved in regulating nuclear genes encoding mitochondrial proteins? Plant Signal. Behav. 6 (2011) 748–750.

[70] S. Fujii, C.S. Bond, I.D. Small, Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 1723–1728.
[71] T. Nakamura, K. Meierhoff, P. Westhoff, G. Schuster, RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA, Eur. J. Biochem. 270 (2003) 4070–4081.
[72] P. Zhelyazkova, K. Hammani, M. Rojas, R. Voelker, M. Vargas-Suarez, T. Borner, A. Barkan, Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts, Nucl. Acids Res. 40 (2012) 3092–3105.
[73] K. Kobayashi, M. Kawabata, K. Hisano, T. Kazama, K. Matsuoka, M. Sugita, T. Nakamura, Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein, Nucl. Acids Res. 40 (2012) 2712–2723.
[74] A. Barkan, M. Rojas, S. Fujii, A. Yap, Y.S. Chong, C.S. Bond, I. Small, A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins, PLoS Genet. 8 (2012) e1002910.
[75] Y. Yagi, S. Hayashi, K. Kobayashi, T. Hirayama, T. Nakamura, Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants, PLoS One 8 (2013) e57286.
[76] M. Takenaka, A. Zehrmann, A. Brennicke, K. Graichen, Improved computational target site prediction for pentatricopeptide repeat RNA editing factors, PLoS One 8 (2013) e65343.
[77] P.J. Preker, W. Keller, The HAT helix, a repetitive motif implicated in RNA processing, Trends Biochem. Sci. 23 (1998) 15–16.
[78] Y. Bai, T.C. Auperin, C.Y. Chou, G.G. Chang, J.L. Manley, L. Tong, Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors, Mol. Cell 25 (2007) 863–875.
[79] K.G. Murthy, J.L. Manley, The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3′-end formation, Genes Dev. 9 (1995) 2672–2683.
[80] Y. Takagaki, J.L. Manley, A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein, Nature 372 (1994) 471–474.
[81] Y. Takagaki, J.L. Manley, Complex protein interactions within the human polyadenylation machinery identify a novel component, Mol. Cell Biol. 20 (2000) 1515–1525.
[82] K. Zhang, D. Smouse, N. Perrimon, The crooked neck gene of Drosophila contains a motif found in a family of yeast cell cycle genes, Genes Dev. 5 (1991) 1080–1091.
[83] J.M. Burnette, A.R. Hatton, A.J. Lopez, Trans-acting factors required for inclusion of regulated exons in the Ultrabithorax mRNAs of Drosophila melanogaster, Genetics 151 (1999) 1517–1529.
[84] J.E. Gallagher, D.A. Dunbar, S. Granneman, B.M. Mitchell, Y. Osheim, A.L. Beyer, S.J. Baserga, RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components, Genes Dev. 18 (2004) 2506–2517.
[85] E.A. Champion, B.H. Lane, M.E. Jackrel, L. Regan, S.J. Baserga, A direct interaction between the Utp6 half-a-tetratricopeptide repeat domain and a specific peptide in Utp21 is essential for efficient pre-rRNA processing, Mol. Cell Biol. 28 (2008) 6547–6556.
[86] K. Hammani, W.B. Cook, A. Barkan, RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5651–5656.
[87] S. Ben-Yehuda, I. Dix, C.S. Russell, M. McGarvey, J.D. Beggs, M. Kupiec, Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*, Genetics 156 (2000) 1503–1517.
[88] S. Felder, K. Meierhoff, A.P. Sane, J. Meurer, C. Driemel, H. Plucken, P. Klaff, B. Stein, N. Bechtold, P. Westhoff, The nucleus-encoded HCF107 gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts, Plant Cell 13 (2001) 2127–2141.
[89] E. Boudreau, J. Nickelsen, S.D. Lemaire, F. Ossenbuhl, J.D. Rochaix, The Nac2 gene of Chlamydomonas encodes a chloroplast TPR-like protein involved in psbD mRNA stability, EMBO J. 19 (2000) 3366–3376.
[90] A.H. Auchincloss, W. Zerges, K. Perron, J. Girard-Bascou, J.D. Rochaix, Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in *Chlamydomonas reinhardtii*, J. Cell Biol. 157 (2002) 953–962.
[91] M. Rahire, F. Laroche, L. Cerutti, J.D. Rochaix, Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I, Plant J. 72 (2012) 652–661.
[92] M. Simarro, A. Gimenez-Cassina, N. Kedersha, J.B. Lazaro, G.O. Adelmant, J.A. Marto, K. Rhee, S. Tisdale, N. Danial, C. Benarafa, A. Orduna, P. Anderson, Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration, Biochem. Biophys. Res. Commun. 401 (2010) 440–446.
[93] S. Eberhard, C. Loiselay, D. Drapier, S. Bujaldon, J. Girard-Bascou, R. Kuras, Y. Choquet, F.A. Wollman, Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in *Chlamydomonas reinhardtii* chloroplasts, Plant J. 67 (2011) 1055–1066.
[94] C. Balczun, A. Bunse, D. Hahn, P. Bennoun, J. Nickelsen, U. Kuck, Two adjacent nuclear genes are required for functional complementation of a chloroplast trans-splicing mutant from *Chlamydomonas reinhardtii*, Plant J. 43 (2005) 636–648.
[95] L. Merendino, K. Perron, M. Rahire, I. Howald, J.D. Rochaix, M. Goldschmidt-Clermont, A novel multifunctional factor involved in trans-splicing of chloroplast introns in Chlamydomonas, Nucl. Acids Res. 34 (2006) 262–274.
[96] T. Linder, C.B. Park, J. Asin-Cayuela, M. Pellegrini, N.G. Larsson, M. Falkenberg, T. Samuelsson, C.M. Gustafsson, A family of putative transcription termination factors shared amongst metazoans and plants, Curr. Genet. 48 (2005) 265–269.
[97] E. Yakubovskaya, E. Mejia, J. Byrnes, E. Hambardjieva, M. Garcia-Diaz, Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription, Cell 141 (2010) 982–993.
[98] H. Spahr, T. Samuelsson, B.M. Hallberg, C.M. Gustafsson, Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain, Biochem. Biophys. Res. Commun. 397 (2010) 386–390.
[99] E. Babiychuk, K. Vandepoele, J. Wissing, M. Garcia-Diaz, R. De Rycke, H. Akbari, J. Joubes, T. Beeckman, L. Jansch, M. Frentzen, M.C. Van Montagu, S. Kushnir, Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 6674–6679.
[100] T. Kleine, Arabidopsis thaliana mTERF proteins: evolution and functional classification, Front. Plant Sci. 3 (2012) 233.
[101] B. Kruse, N. Narasimhan, G. Attardi, Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination, Cell 58 (1989) 391–397.
[102] S. Peralta, X. Wang, C.T. Moraes, Mitochondrial transcription: lessons from mouse models, Biochim. Biophys. Acta 1819 (2012) 961–969.
[103] M. Roberti, P.L. Polosa, F. Bruni, C. Manzari, S. Deceglie, M.N. Gadaleta, P. Cantatore, The MTERF family proteins: mitochondrial transcription regulators and beyond, Biochim. Biophys. Acta 1787 (2009) 303–311.
[104] Y. Camara, J. Asin-Cayuela, C.B. Park, M.D. Metodiev, Y. Shi, B. Ruzzenente, C. Kukat, B. Habermann, R. Wibom, K. Hultenby, T. Franz, H. Erdjument-Bromage, P. Tempst, B.M. Hallberg, C.M. Gustafsson, N.G. Larsson, MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome, Cell Metab. 13 (2011) 527–539.
[105] E. Yakubovskaya, K.E. Guja, E. Mejia, S. Castano, E. Hambardjieva, W.S. Choi, M. Garcia-Diaz, Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification, Structure 20 (2012) 1940–1947.
[106] H. Spahr, B. Habermann, C.M. Gustafsson, N.G. Larsson, B.M. Hallberg, Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 15253–15258.
[107] A. Wredenberg, M. Lagouge, A. Bratic, M.D. Metodiev, H. Spahr, A. Mourier, C. Freyer, B. Ruzzenente, L. Tain, S. Gronke, F. Baggio, C. Kukat, E. Kremmer, R. Wibom, P.L. Polosa, B. Habermann, L. Partridge, C.B. Park, N.G. Larsson, MTERF3 regulates mitochondrial ribosome biogenesis in invertebrates and mammals, PLoS Genet. 9 (2013) e1003178.
[108] P. Robles, J.L. Micol, V. Quesada, Unveiling plant mTERF functions, Mol. Plant 5 (2012) 294–296.
[109] R. Meskauskiene, M. Wursch, C. Laloi, P.A. Vidi, N.S. Coll, F. Kessler, A. Baruah, C. Kim, K. Apel, A mutation in the Arabidopsis mTERF-related plastid protein SOLDAT10 activates retrograde signaling and suppresses (1)O(2)-induced cell death, Plant J. 60 (2009) 399–410.
[110] P. Robles, J.L. Micol, V. Quesada, Arabidopsis MDA1, a nuclear-encoded protein, functions in chloroplast development and abiotic stress responses, PLoS One 7 (2012) e42924.
[111] C. Schonfeld, L. Wobbe, R. Borgstadt, A. Kienast, P.J. Nixon, O. Kruse, The nucleus-encoded protein MOC1 is essential for mitochondrial light acclimation in *Chlamydomonas reinhardtii*, J. Biol. Chem. 279 (2004) 50366–50374.
[112] L. Wobbe, P.J. Nixon, The mTERF protein MOC1 terminates mitochondrial DNA transcription in the unicellular green alga *Chlamydomonas reinhardtii*, Nucl. Acids Res. 41 (2013) 6553–6567.
[113] V. Quesada, R. Sarmiento-Manus, R. Gonzalez-Bayon, A. Hricova, R. Perez-Marcos, E. Gracia-Martinez, L. Medina-Ruiz, E. Leyva-Diaz, M.R. Ponce, J.L. Micol, Arabidopsis RUGOSA2 encodes an mTERF family member required for mitochondrion, chloroplast and leaf development, Plant J. 68 (2011) 738–753.
[114] M. Kim, U. Lee, I. Small, C.C. des Francs-Small, E. Vierling, Mutations in an Arabidopsis mitochondrial transcription termination factor-related protein enhance thermotolerance in the absence of the major molecular chaperone HSP101, Plant Cell 24 (2012) 3349–3365.
[115] E. Delannoy, W.A. Stanley, C.S. Bond, I.D. Small, Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles, Biochem. Soc. Trans. 35 (2007) 1643–1647.
[116] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[117] G.L. Blatch, M. Lassle, The tetratricopeptide repeat: a structural motif mediating protein-protein interactions, Bioessays 21 (1999) 932–939.
[118] L.A. Kelley, M.J.E. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server, Nat. Protoc. 4 (2009) 363–371.
[119] L. Grell, C. Parkin, L. Slatest, P.A. Craig, EZ-Viz, a tool for simplifying molecular viewing in PyMOL, Biochem. Mol. Biol. Educ. 34 (2006) 402–407.
[120] W. Chi, J. Mao, Q. Li, D. Ji, M. Zou, C. Lu, L. Zhang, Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor

SIG6 in the regulation of chloroplast gene expression in *Arabidopsis cotyledons*, Plant J. 64 (2010) 14–25.
[121] K. Hammani, C. Colas des Francs-Small, M. Takenaka, S.K. Tanz, K. Okuda, T. Shikanai, A. Brennicke, I. Small, The pentatricopeptide repeat protein OTP87 is essential for RNA editing of nad7 and atp1 transcripts in Arabidopsis mitochondria, J. Biol. Chem. 286 (2011) 21361–21371.
[122] K. Hammani, K. Okuda, S.K. Tanz, A.L. Chateigner-Boutin, T. Shikanai, I. Small, A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites, Plant Cell 21 (2009) 3686–3699.
[123] C.L. Dieckmann, T.J. Koerner, A. Tzagoloff, Assembly of the mitochondrial membrane system. CBP1, a yeast nuclear gene involved in 5′ end processing of cytochrome b pre-mRNA, J. Biol. Chem. 259 (1984) 4722–4731.
[124] F. Xu, C. Morin, G. Mitchell, C. Ackerley, B.H. Robinson, The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA, Biochem. J. 382 (2004) 331–336.
[125] R. Mancebo, X. Zhou, W. Shillinglaw, W. Henzel, P.M. Macdonald, BSF binds specifically to the bicoid mRNA 3′ untranslated region and contributes to stabilization of bicoid mRNA, Mol. Cell Biol. 21 (2001) 3462–3471.
[126] F. Tavares-Carreon, Y. Camacho-Villasana, A. Zamudio-Ochoa, M. Shingu-Vazquez, A. Torres-Larios, X. Perez-Martinez, The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial COX1 mRNA in yeast, J. Biol. Chem. 283 (2008) 1472–1479.
[127] O. Rackham, S.M. Davies, A.M. Shearwood, K.L. Hamilton, J. Whelan, A. Filipovska, Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells, Nucl. Acids Res. 37 (2009) 5859–5867.
[128] J.W. Coffin, R. Dhillon, R.G. Ritzel, F.E. Nargang, The *Neurospora crassa* cya-5 nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein, Curr. Genet. 32 (1997) 273–280.
[129] E.C. Koc, L.L. Spremulli, RNA-binding proteins of mammalian mitochondria, Mitochondrion 2 (2003) 277–291.
[130] C.G. Noble, P.A. Walker, L.J. Calder, I.A. Taylor, Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA, Nucl. Acids Res. 32 (2004) 3364–3375.
[131] S.A. Bell, A.G. Hunt, The Arabidopsis ortholog of the 77 kDa subunit of the cleavage stimulatory factor (AtCstF-77) involved in mRNA polyadenylation is an RNA-binding protein, FEBS Lett. 584 (2010) 1449–1454.
[132] K. Vincent, Q. Wang, S. Jay, K. Hobbs, B.C. Rymond, Genetic interactions with CLF1 identify additional pre-mRNA splicing factors and a link between activators of yeast vesicular transport and splicing, Genetics 164 (2003) 895–907.
[133] E.M. Makarov, O.V. Makarova, T. Achsel, R. Luhrmann, The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions, J. Mol. Biol. 298 (2000) 567–575.
[134] B.H. Lee, A. Kapoor, J. Zhu, J.K. Zhu, STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in Arabidopsis, Plant Cell 18 (2006) 1736–1749.
[135] A.K. Hyvarinen, J.L. Pohjoismaki, A. Reyes, S. Wanrooij, T. Yasukawa, P.J. Karhunen, J.N. Spelbrink, I.J. Holt, H.T. Jacobs, The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA, Nucl. Acids Res. 35 (2007) 6458–6474.
[136] A.K. Hyvarinen, J.L. Pohjoismaki, I.J. Holt, H.T. Jacobs, Overexpression of MTERFD1 or MTERFD3 impairs the completion of mitochondrial DNA replication, Mol. Biol. Rep. 38 (2011) 1321–1328.

E.2 Publication 4: ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature

CRYSTAL GROWTH & DESIGN

Article

pubs.acs.org/crystal

ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature

Published as part of the Crystal Growth & Design virtual special issue on the 14th International Conference on the Crystallization of Biological Macromolecules (ICCBM14)

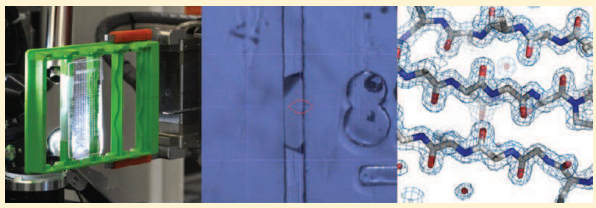
Franziska Pinker,[†] Mathieu Brun,[‡] Pierre Morin,[‡] Anne-Laure Deman,[‡] Jean-François Chateaux,[‡] Vincent Oliéric,[#] Christian Stirnimann,[#] Bernard Lorber,[†] Nicolas Terrier,[‡] Rosaria Ferrigno,[‡] and Claude Sauter*[,†]

[†]Architecture et Réactivité de l'ARN, Université de Strasbourg et CNRS, IBMC, 15 rue René Descartes, F-67084 Strasbourg, France

[‡]Université de Lyon F-69003 LYON, Université Claude Bernard Lyon 1, Institut des Nanotechnologies de Lyon, CNRS UMR5270, F-69622 Villeurbanne, France

[#]Swiss Light Source, Paul Scherrer Institute, Villigen, CH-5232, Switzerland

ABSTRACT: Microfluidic technology has opened new possibilities for the crystallization of biological macromolecules during the past decade. Microfluidic systems offer numerous advantages over conventional crystal growth methods. They enable easy handling of nanovolumes of solutions, extreme miniaturization, and parallelization of crystallization assays, especially for high-throughput screening applications. Our goal was to design a versatile, low cost, and easy-to-use crystallization chip based on counter-diffusion that is compatible with on-chip crystallographic characterization. The ChipX is a microfluidic chip made of cyclic olefin copolymer. It was used to grow crystals of biomolecules and perform complete X-ray diffraction analyses on synchrotron sources. Our results demonstrate that accurate crystallographic data can be collected at room temperature directly from ChipX microfluidic devices for both experimental single-wavelength anomalous dispersion phasing and structure refinement.



INTRODUCTION

Crystallography is a major investigation tool in structural biology. It provides three-dimensional (3D) information on biomolecules (proteins, nucleic acids, viruses, etc.) that is essential for understanding biological processes and designing new pharmaceuticals.[1] Crystallographic analyses, however, rely on the reproducible growth of ordered crystals under well-defined crystallization conditions. Generally, a wide range of chemicals is tested to find an adequate crystallant (a salt, an alcohol, a polymer, or a mixture of them) and the right physical-chemical parameters (e.g., temperature, pH).[2] Therefore, screening experiments can be time-consuming and may need quantities of pure biomolecules that are not easily accessible. Long ago, this cumbersome search of crystallization conditions triggered the interest of crystal growers for miniaturized setups and led to the development of the first pipetting stations known as crystallization robots.[3] Nowadays, such technologies are popular and massively used for high-throughput applications in structural genomics projects.[4]

During the past decade, microfluidic technology has opened new possibilities. Microfluidic systems offer many advantages for crystal growth including extreme miniaturization enabling easy handling of nanovolumes of solutions and parallelization of crystallization assays. They also provide a convectionless environment favorable to the growth of high quality crystals.[5] Pioneering examples are the implementation of free interface diffusion[6] and of nanobatch crystallization[7] in microfluidic chips. They have demonstrated the value of this technology for high-throughput screening. These sophisticated systems require, however, extra pieces of equipment to load samples into the crystallization chips. As an alternative, simpler devices have been developed that (i) can be setup by hand, (ii) use diffusion-based crystallization, and (iii) are compatible with *in situ* crystallographic analysis.[5,8,9]

In a long-term project intended to develop a crystallization chip using the method of counter-diffusion, we identified cyclic olefin copolymer (COC) as a promising material for the fabrication of chips suitable for X-ray diffraction.[10] In the present work, we report the optimization of the dimensions and configuration of our microfluidic device ChipX for screening assays and for *in situ* X-ray diffraction analyses. Using the latest X-ray detector technology and synchrotron data collection strategies, high resolution structures of model biomolecules were determined at room temperature by either molecular

Received: November 30, 2012
Revised: June 19, 2013
Published: June 25, 2013

replacement or experimental single wavelength anomalous (SAD) phasing. Our results demonstrate that microfluidic devices can be used (i) to grow high quality crystals and (ii) to perform *in situ* characterization by X-ray diffraction at room temperature without direct crystal handling. The implemented crystal growth and analysis strategy provides a promising alternative to current practice in biocrystallography, in particular, for fragile crystals that are difficult to handle or to flash cool.

EXPERIMENTAL SECTION

Materials, Chemicals, and Biomolecules. Cyclic olefin copolymer (COC Zeonor 1020 R) was purchased from Zeon Corp. Photolithography was performed using SU8-2100 (CTS) and propylene glycol methyl ether acetate (Sigma Aldrich) as the developer. Chemicals for the electrodeposition of Ni were obtained from Acros Organic (H_2SO_4 and sodium dodecyl sulfate, SDS), Goodfellow (Ni foils, $NiCl_2$, and $NiSO_4$), and Sigma Aldrich (H_3BO_3, H_2O_2). The stripper of the SU8 photoresist was from CTS. Plant thaumatin (207 amino acids, 22 kDa), hen egg-white lysozyme (129 amino acids, 14.5 kDa), bovine insulin (21 + 30 amino acids, 5.7 kDa), *N*-(2-acetamido)-2-iminodiacetic acid (ADA), 2-(*N*-morpholino) ethanesulfonic acid (MES), and DL-tartaric acid were purchased from Sigma Aldrich. The 500 mM aqueous solution of 10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecan-1,4,7-triacetic acid complexed to ytterbium (HPDO3A-Yb; MW 574.1 g) used for SAD phasing was from NatX-ray (Grenoble, France).[11] All other chemicals were of ACS grade and used without further purification. All crystallization solutions were prepared with distilled water and filtered on membranes with a 0.22 μm porosity. The detergent *N*-octyl-β-D-glucopyranoside (βOG, Bachem, Cat. No. P-1110) was added to protein solutions at a concentration of 0.3% (m/v) to facilitate their entry in microchannels by capillarity, as described previously.[10]

Fabrication of Microfluidic Structures. Microdevices were fabricated with either two or three COC layers of different thicknesses. The layer containing the microfluidic channels (from 3 mm down to 600 μm thick) was microstructured by hot-embossing at 38 bar and 145 °C using a mold obtained by a UV-LIGA process as described elsewhere.[12] A stainless steel substrate was first polished mechanically and then electrochemically. After an exposure to O_2 plasma (5 min, 75 W), the steel was covered with a 80-μm-thick layer of SU8-2100 using spin-coating (2300 rpm for 30 s). Then, the substrate was soft-baked at 95 °C for 25 min and irradiated with a Hg lamp (150 mJ/cm^2). After a second bake (95° for 10 min), the substrate was rinsed with PGMEA for 10 min to achieve development of the microstructures. Before the electrodeposition step, the substrate was cleaned by O_2 plasma treatment (20 min, 100 W) to remove photoresist left after development into the trenches that needed to be filled with metal. The electroplating step was carried out in two different electroplating baths: first, in a solution containing 1.3 M $NiCl_2$, 2 M HCl, for 5 min at 50 °C and with a current density of 2 A/dm^2, then in a bath containing 1.1 M $NiSO_4$, 0.25 M $NiCl_2$, 0.7 M H_3BO_3, 25 mM SDS at 55 °C and 5 A/dm^2 (for these conditions, we reached an electroplating speed of 1 μm min^{-1}). This last electrodeposition step was carried out until the metal deposited lightly overflowed the trenches made in the photoresist layer. Finally, the mold was mechanically polished to obtain the required channel thickness (checked by profilometer measurements), and the remaining photoresist was stripped away. A second COC layer (3-mm to 600-μm-thick) consisting of a plain layer of polymer (with holes were drilled to act as inlets) was assembled with the first layer to close the microchannels by solvent-assisted bonding (20 s under methylcyclohexane vapor followed by application of a pressure of 30 bar at 87 °C for 5 min). Finally, in the thinner devices, a third layer (3 mm thick) was added to act as a frame (Figure 1). It was aligned by hand with the previous layers and bonded. Holes were drilled at the extremities of the channels to obtain wells used to load the various crystallization solutions.

Crystallization Experiments. The standard procedure for growing crystals in chips at 20 °C consisted of three steps. First, 3 μL of macromolecular solution was injected (or deposited) with a Hamilton microliter syringe in the sample well to fill the entire channel arborescence. The addition of 0.3% (m/v) of detergent (βOG) made sample loading by capillarity very easy. Second, 2 μL of crystallization solution was introduced in each of the eight wells with a micropipet. After each filling step, the wells were immediately sealed with CrystalClear tape (Hampton Research) to prevent evaporation and displacement of solutions inside channels. The following solutions were used to produce crystals: 30 mg/mL thaumatin and 1.5 M sodium tartrate containing 0.1 M ADA pH 6.5; 50 mg/mL hen lysozyme and 1 M NaCl containing 30% (m/v) PEG-3350, 0.1 M sodium acetate pH 4.5, or 2 M NaCl, 0.1 M sodium acetate pH 4.5, or 1.8 M NaCl, 0.1 M sodium acetate pH 4.5, 50 mM HPDO3-Yb (for SAD experiments); 20 mg/mL insulin and 0.275 M $NaHPO_4/Na_3PO_4$ pH 10.2, 0.01 M Na_3EDTA. Insulin crystals for sulfur-SAD measurements were prepared at 20 °C by vapor diffusion (sitting drop) in CrystalQuick X microplates (Greiner Bio-One) made of 300-μm-thick COC.[13]

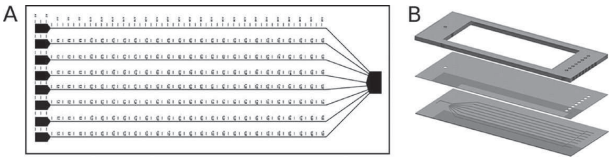


Figure 1. Design of the microfluidic chip optimized for crystallization by counter-diffusion. (A) The fluidic layer measures 35 × 65 mm^2. The eight crystallization channels with a section of 75 × 75 μm^2 and a volume of 253 nL are connected at one side to a single common inlet, in which the protein sample is loaded. After sealing the sample inlet with tape, the crystallant solutions are deposited in the wells at the opposite end of the channels, and crystallant inlets are sealed with tape. Labels along the channels help locate crystals and facilitate their positioning in the X-ray beam during diffraction analyses. Channel length (i.e., the diffusion path) is greater (45 mm) than in previous chip versions (15 mm) to spread the supersaturation gradient, while reducing the dead volume of the system. At the same time, the dead volume of the system was reduced. (B) Chip assembly includes three layers. The top layer is a 3-mm-thick frame which contains the inlets for the biomolecule sample and crystallant solutions. The fluidic layer (0.6-mm-thick) is at the bottom, and the channels are closed by the bonding of a 0.6-mm-thick COC layer producing only low X-ray scattering.

In Situ X-ray Diffraction Analyses. The attenuation of the direct X-ray beam (absorption) by COC sheets with different thicknesses was evaluated with a conventional laboratory source (NONIUS rotating anode operating at 90 mA and 45 kV). The scattering background generated by COC sheets was measured on the FIP-BM30A with exposure times equivalent to those used for actual data collection (60 s). All X-ray diffraction analyses were performed at room temperature (T = 20−25 °C) on automated synchrotron beamlines: (i) FIP-BM30A[14] at the European Synchrotron Radiation Facility (ESRF, Grenoble France) equipped with an ADSC Quantum 315r CCD detector, (ii) X10SA and X06DA[15] at the Swiss Light Source (SLS, Villigen, Switzerland) equipped with a MAR MX225 CCD detector, or DECTRIS PILATUS 2M/6M pixel detectors.

In situ crystal analysis was carried out with microfluidic chips attached onto a chip holder that was maintained in the beam by the arm of a CATS robot. Crystals were characterized inside the chips a week after setting up the crystallization assays. Diffraction-based crystal centering was necessary at extreme positions (−36°/+36°) to circumvent the refraction affecting the alignment due to variable polymer thickness. Exposure time and oscillation range were adjusted to minimize radiation damage at room temperature. Diffraction data were processed using the XDS package.[16] Crystallographic statistics are given in Table 1.

SAD phasing experiments were carried on lysozyme crystals grown by counter-diffusion from a crystallant solution containing an ytterbium derivative (HPDO3-Yb). The chip was directly attached with a magnetic base to the multiaxis PRIGo goniometer at beamline X06DA at SLS, and two sets of 60° were collected on a single crystal

dx.doi.org/10.1021/cg301757g | *Cryst. Growth Des.* 2013, 13, 3333−3340

Table 1. Statistics of On-Chip Data Collections on Lysozyme and Thaumatin Crystals

crystal	lysozyme1	lysozyme2	lysozyme3	lysozyme4	lysozyme5	lysozyme6	lysozyme7	lysozyme8	lysozyme9
Synchrotron	ESRF	ESRF	SLS	SLS	SLS	SLS	SLS	SLS	SLS
Beamline	FIP-BM30A	FIP-BM30A	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA
wavelength (Å)	0.98	0.98	0.98	0.83	0.83	1.00	1.00	1.00	1.00
detector	ADSC Q315r	ADSC Q315r	MAR225	MAR225	MAR225	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 2M
number of images	15	21	29	36	36	1200	1200	1200	1200
oscillation (deg/s)	2/60	1/60	2/1	2/2	2/1	0.05/0.05	0.05/0.05	0.05/0.05	0.05/0.025
acquisition time (min)	17	25	5	5	5	1	1	1	0.5
space group	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2	*P*4(3)2(1)2
a, *b* (Å)	79.17	79.05	79.17	79.21	79.14	79.30	79.36	79.36	79.32
c (Å)	38.34	38.23	38.00	37.87	37.79	37.93	37.94	37.95	38.02
resolution range (Å)	30−2.0	30−1.8	30−1.6	30−1.66	30−1.55	30−1.43	30−1.43	30−1.43	30−1.43
completeness (%)	79.2	82.0	90.9	98.8	83.9	92	96	93.8	96.4
multiplicity	2.9	1.9	2.3	3.0	6.7	4.2	4.1	4.2	3.4
R_{merge}/R_{meas} (%)[a]	9.8/11.9	5.8/7.7	6.0/7.6	9.6/11.7	7.5/8.2	8/9.1	5.0/5.7	5.3/6.1	8.7/10.1
$\langle I/\sigma(I)\rangle$	8.0	10.3	12.4	11.3	14.9	11.1	15.7	15.0	9.2
high resolution shell (Å)	2.12−2.0	1.91−1.8	1.69−1.6	1.76−1.66	1.64−1.55	1.52−1.43	1.52−1.43	1.52−1.43	1.52−1.43
completeness (%)	83.2	82.6	74.5	97.8	85.4	90.6	82.7	77.4	92.8
multiplicity	2.8	1.9	1.5	3.00	6.50	2.8	3.1	3.4	2.4
R_{merge}/R_{meas} (%)[a]	37.1/45.7	15.6/20.4	55.0/74.6	55.3/67.4	61.0/66.2	44.0/53.0	35.8/43.1	36.2/42.2	33.2/40.9
$\langle I/\sigma(I)\rangle$	2.7	4.5	1.8	2.9	3.1	2.4	3.4	2.9	2.8
mosaicity (deg)	0.09	0.11	0.12	0.11	0.09	0.02	0.02	0.02	0.02
Wilson plot *B*-factor (Å^2)	27.3	23.6	26.4	23.3	24.9	24.6	26.2	25.5	25.3
crystal	thaumatin1	thaumatin2	thaumatin3	thaumatin4	thaumatin5	thaumatin6	thaumatin7	thaumatin8	thaumatin9
Synchrotron	SLS	SLS	SLS	SLS	SLS	SLS	SLS	SLS	SLS
Beamline	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA	X06DA
wavelength (Å)	0.98	1.6	1.6	0.83	0.83	1	1	1	1
detector	MAR225	MAR225	MAR225	MAR225	MAR225	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 2M
number of images	30	33	33	36	36	240	240	600	1200
oscillation (deg/s)	1/1	2/1	2/1	2/1	2/1	0.25/0.25	0.25/0.25	0.1/0.1	0.05/0.05
acquisition time (min)	6	5	5	5	5	1	1	1	1
space group	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2	*P*4(1)2(1)2
a, *b* (Å)	58.59	58.56	58.58	58.59	58.59	58.70	58.77	58.70	58.64
c (Å)	151.59	151.62	151.66	151.70	151.70	151.87	151.86	152.09	152.02
resolution range (Å)	50−1.7	50−2.0	50−2.0	50−2.0	50−1.83	60−1.65	60−1.6	60−1.55	60−1.55
completeness (%)	54.2	90.3	90.7	98.9	90.4	95.9	94.6	83.3	98.8
multiplicity	2.2	2.5	2.5	3.1	2.6	4.0	3.8	5.1	4.3
R_{merge}/R_{meas} (%)[a]	5.8/7.3	7.9/9.7	10.0/12.4	16.1/19.4	11.4/14.0	12.5/14.4	10.2/11.8	7.4/8.1	8.5/9.7
$\langle I/\sigma(I)\rangle$	11.4	9.7	7.6	9	7.9	9.7	8.5	12.2	9.8
high resolution shell (Å)	1.8−1.7	2.12−2.0	2.12−2.0	2.12−2.0	1.94−1.83	1.75−1.65	1.69−1.6	1.64−1.55	1.64−1.55
completeness (%)	51.6	61.8	62.3	98.9	84.7	95.7	98.6	78.2	99.2
multiplicity	1.9	1.4	1.4	3.1	2.3	4.0	3.7	5.3	4.2
R_{merge}/R_{meas} (%)[a]	33.3/44.9	21.3/28.0	31.0/41.0	39.4/47.3	46.9/58.2	75.1/86.2	73.3/85.2	46.3/50.4	59.9/68.9
$\langle I/\sigma(I)\rangle$	2.9	2.7	1.7	3.3	2.3	2.0	1.7	2.8	2.0
mosaicity (deg)	0.07	0.10	0.09	0.08	0.08	0.14	0.03	0.05	0.02
Wilson plot *B*-factor (Å^2)	24.5	25.2	26.4	26.8	22.8	26.1	25.3	25.1	25.5

[a] $R_{merge} = \Sigma_{hkl}\ \Sigma_i\ |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\ \Sigma_i\ I_i(hkl)$ and redundancy-independent $R_{meas} = \Sigma_{hkl}\ (n/n-1)^{1/2}\ \Sigma_i\ |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\ \Sigma_i\ I_i(hkl)$.

dx.doi.org/10.1021/cg301757g | Cryst. Growth Des. 2013, 13, 3333−3340

Table 2. Lysozyme, Thaumatin, and Insulin Structure Determination by in Situ X-ray Diffraction[a]

crystal	lysozyme9	thaumatin 9	lysozyme-Yb	insulin
		Data Collection Statistics		
Beamline	SLS/X06DA	SLS/X06DA	SLS/X06DA	SLS/X10SA
wavelength (Å)	1.0	1.0	1.38	1.70
detector	PILATUS 2M	PILATUS 2M	PILATUS 2M	PILATUS 6M
number of images	3 × 1200	2 × 1200	2 × 600	5 × 600
space group	$P4_32_12$	$P4_12_12$	$P4_32_12$	$I2_13$
unit cell parameters *a*, *b*, *c* (Å)	79.3, 79.3, 38.0	58.7, 58.7, 152.1	79.1, 79.1, 38.4	79.0, 79.0, 79.0
resolution range (Å)[a]	30−1.43 (1.52−1.43)	60−1.55 (1.64−1.55)	30−1.75 (1.85−1.75)	50−2.3 (2.4−2.3)[d]
no. of unique reflections	22727 (3680)	39090 (5995)	21245 (3382)	6911 (698)
completeness (%)	98.9 (97.4)	99.1 (99.8)	90.4 (93.9)	97.4 (83.5)
multiplicity	9.4 (5.0)	7.0 (4.2)	7.7 (5.9)	15.0 (9.9)
R_{meas} (%)[b]	9.5 (45.7)	11.5 (67.6)	6.8 (20.8)	7.2 (13.3)
CC(1/2)[c]	99.5 (88.3)	99.7 (78.6)	99.6 (91.7)	99.8 (99.1)
$\langle I/\sigma(I)\rangle$	17.4 (3.7)	11.0 (2.3)	13.7 (4.3)	33.1 (16.0)
		SAD Phasing Statistics (SHELX-2013)		
number of anomalous sites			2 Yb	6 S (3 disulfide bonds)
figure of merit (FOM)			0.62	0.54
pseudofree CC (%)			62.4	58.2
number of CA traced			115 (out of 129)	30 (out of 51)
		Refined Atomic Structure		
resolution range (Å)	30−1.43	60−1.55	30−1.75	30−2.3
R-factor/*R*-free (%)[e]	17.0/19.0	16.4/18.9	18.7/19.3	15.6/16.9
number of protein, solvent	1001/106	1550/180	1001/57	402/15
protein, solvent, ADPs (Å^2)[f]	13.5/27.2	26.0/23.3	29.9/34.5	25.9/29.1
r.m.s.d. on bonds (Å) and angles (°)	0.029/1.49	0.006/1.17	0.006/1.09	0.003/0.59
PDB id	3ZEK	3ZEJ	4BS7	4BS3

[a]Values in parentheses are for high resolution shells. [b]Redundancy-independent $R_{meas} = \Sigma_{hkl}\ (n/n-1)^{1/2}\ \Sigma_i\ |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\ \Sigma_i\ I_i(hkl)$. [c]CC(1/2) the percentage of correlation between intensities from random half-data sets.[27] [d]Resolution limited because minimal detector distance at beamline X10SA was reached. [e]The cross-validation (*R*-free) was calculated with 5% of the data. [f]ADPs: atomic displacement parameters.

with Chi = 0° and 15° at a wavelength of 1.385 Å (LIII absorption edge of Yb). The data sets were processed individually and then merged and scaled using the XDS package. Two Yb atoms were identified in 25 search trials with the HKL2MAP[17] interface for SHELXD.[18]

For sulfur-SAD measurements, five data sets with a total oscillation of 60° were collected at 1.7 Å wavelength from a single crystal of insulin in a CrystalQuick X plate. The beam position on the crystal was shifted from data set to data set. Continuous data collection was performed on a PILATUS 6 M detector at X10SA beamline. The data sets were processed individually and then merged and scaled using the XDS package. Six sulfur atoms in three disulfide bonds were correctly identified with HKL2MAP.

SAD-phasing, density modification, and initial polyalanine model building were performed with SHELXE in HKL2MAP. Crystal structures of thaumatin, lysozyme, and insulin were built and refined with Coot and Phenix.[19,20] Corresponding coordinates and structure factors were deposited with the Protein Databank (Table 2).

RESULTS AND DISCUSSION

A Chip Optimized for Crystallization by Counter-diffusion. Our goal was to design a chip implementing counter-diffusion, a crystallization method known to be very efficient in defining and optimizing crystal growth conditions. It consists of creating a broad concentration gradient of crystallant that propagates along the microchannel. The formation of such a gradient generates a supersaturation wave[21] which allows screening of a great number of potential crystallization conditions in a single experiment. For this reason, the geometry of our microstructures was optimized to take advantage of all benefits of real counter-diffusion. In particular, the effective length of the channels was increased from 15 mm (in the previous design[10]) to 45 mm in order to spread the gradient and improve the screening (Figure 1A). At the same time, the channel section was reduced from 100 × 100 μm^2 to 75 × 75 μm^2 to keep the sample volume to a minimum (i.e., < 300 nL per channel).

As shown in Figure 1A, the chip design features eight parallel microchannels that are connected on one side to a single injection inlet. Each channel has an independent crystallant well at the opposite end. The biomolecule solution was filled in all microchannels simultaneously by capillarity and, then, eight crystallant solutions were loaded into the corresponding wells. Their diffusion through the microchannels led to the formation of gradients of crystallant concentration.

A Chip Optimized for *in Situ* Crystal Analysis. On-chip crystal characterization requires the material of the chip to be transparent enough to X-rays. According to our previous study,[10] thermoplastics such COC or polydimethyl metacrylate (PMMA) have much better characteristics in terms of X-ray absorption/scattering and rigidity than PDMS, which is the popular material in microfluidics. For this reason, we used COC and searched for the best compromise between thickness, absorption on a weak laboratory X-ray source, and scattering signal during X-ray analysis using a strong synchrotron beam. We first investigated the X-ray absorption and the scattering of COC thicknesses from 190 μm to 2.8 mm. The plot in Figure 2 illustrates how the absorption/scattering varies with material thickness. Preliminary tests on a laboratory based diffractometer source indicated that less than 40% of the incident beam is absorbed by a 1.5-mm-thick COC layer. The absorption is even lower at shorter wavelengths on a synchrotron radiation source.

dx.doi.org/10.1021/cg301757g | Cryst. Growth Des. 2013, 13, 3333−3340

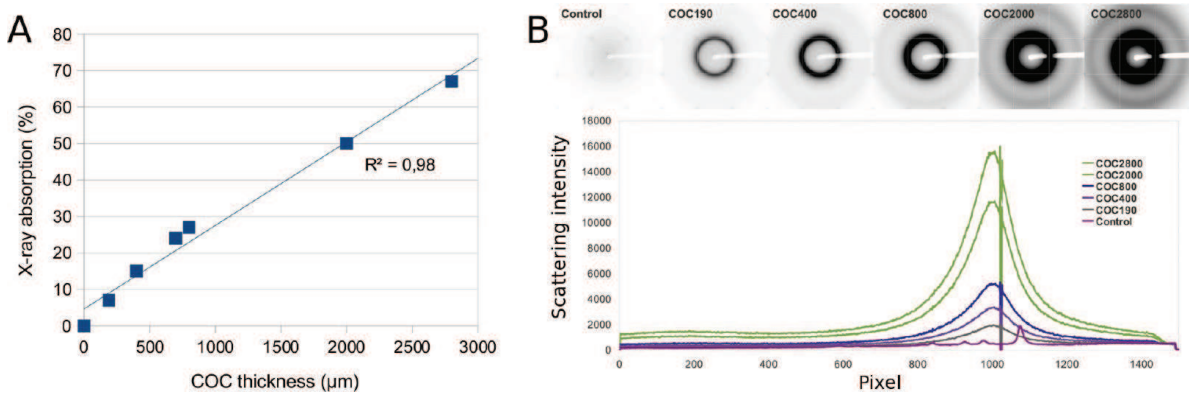


Figure 2. X-ray absorption and scattering by COC. (A) Plot of X-ray beam absorption versus COC thickness. COC sheets were placed in the beam generated by a laboratory source. The intensity of the beam was measured with an X-ray counter, and the relative absorption was calculated by the ratio of counts-per-second in the presence of COC in the beam over that in the absence. Owing to the linear relationship between material thickness and beam attenuation, the absorption was predicted to be less than 40% for a total thickness of 1.5 mm. (B) Comparison of background images (60 s exposure, 300 mm sample-to-detector distance, wavelength 0.98 Å) produced in the absence (control) and in the presence of the same COC sheets in the synchrotron X-ray beam (FIP-BM30A). (Top) Scattering images collected with an ADSC Quantum 315r CCD detector (3072 × 3072 pixels). The intensity of the X-ray signal was measured at each pixel site with a dynamic range of 16 bits (a value of 65 536 corresponds to pixel saturation). The grayscale is the same for all images. (Bottom) Corresponding radial profiles (pixels 1−1536 along the image *x* axis) showing the increase of background intensity with material thickness. The intensity corresponding to an overall thickness of 1200 μm is predicted to be ≤7000 or about 1/10 of the pixel saturation intensity. This is an acceptable background to perform a crystal analysis on a synchrotron source.

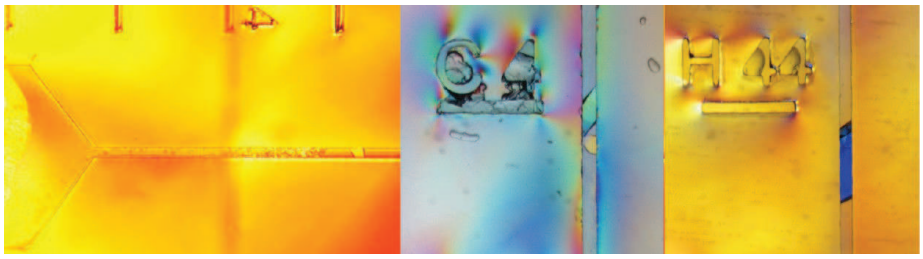


Figure 3. Examples of protein crystallization in microchannels. (Left) Thaumatin crystal growth in a microchannel exhibiting typical counter-diffusion features. The concentrated crystallant diffusing from the well on the left-hand side into the channel generates a concentration gradient. As the supersaturation wave travels through the protein solution, the biomolecule successively precipitates at very high supersaturation, forms microcrystals at intermediate supersaturation and single crystals at moderate to low supersaturation. (Center) Thaumatin bipyramids and (right) single lysozyme crystal filling the entire channel (section: 75 × 75 μm^2). Labels are visible near the crystals.

We found that a total COC thickness of 1.0−1.5 mm is acceptable in terms of X-ray scattering while warranting a good rigidity of the chip. We added graduations along each channel to facilitate the identification of crystals during growth monitoring or crystallographic characterization.

The first version of the device was made of two 3-mm-thick COC layers. One hosted the channels and the second served as the cover containing the inlets. The thickness of this assembly was not compatible with the requirement of *in situ* diffraction analyses (Figure 2). A second version consisted of microchannels hot-embossed in a 600-μm-thick COC layer and a 3-mm-thick COC cover. In a third attempt to reduce the amount of COC interacting with the X-ray beam, we fabricated a device made of three layers. It was composed of a 600-μm-thick fluidic layer closed by another 600-μm-thick COC layer bonded to a 3-mm-thick COC frame containing the inlets (Figure 1B). This final chip design was rigid enough to be handled by the robotic arm which positions the device in the X-ray beam for crystal analysis. It produced a sufficiently low background that did not interfere with the collection of high quality diffraction data (see below).

Biomolecule Crystallization and on-Chip Crystal Analysis. The novel ChipX design was used to crystallize several proteins. Figure 3 displays a typical counter-diffusion experiment with thaumatin in a microchannel. At variance with former porous PDMS chips, the multilayer COC device was gastight and suitable for long duration crystallization experiments. Neither solution evaporation nor crystal alteration was observed within a period of six months at 20 °C. Crystals could be grown that reached the size of the channels (typical size of 75 × 75 × 100 μm^3) as shown in Figure 3. They could be safely stored inside the chips until their analysis on synchrotron beamlines.

On-chip crystal analyses were conducted on beamlines equipped with CATS robots, i.e., the FIP-BM30 beamline at ESRF and X06DA beamline at SLS. The robotic arm was used to position the chip in the incident X-ray beam (Figure 4A). The total oscillation angle was, however, limited to 72° to avoid the collision of the microplate holder (the green frame with SBS format in Figure 4A,B) with the surrounding equipment. For the X-ray diffraction tests, we have chosen model proteins crystallizing in high symmetry space groups to maximize data completeness. The major difficulty encountered during data collection was the drift of crystals outside of the beam during the oscillation of the robotic arm. With CCD detectors, the

dx.doi.org/10.1021/cg301757g | *Cryst. Growth Des.* 2013, 13, 3333−3340

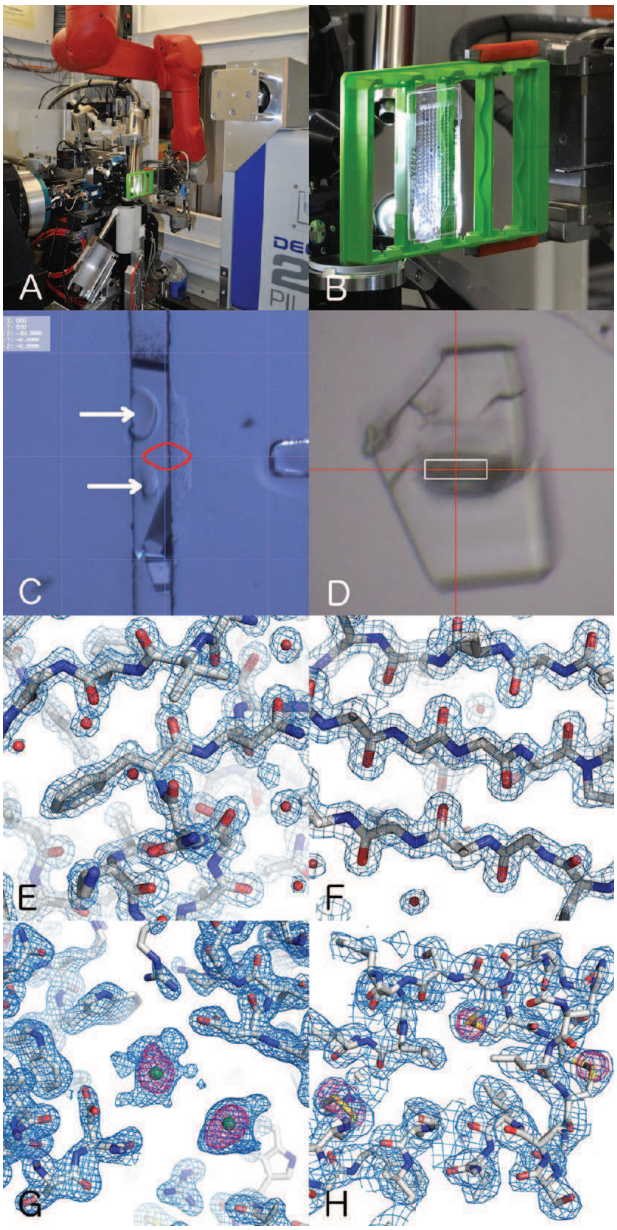


Figure 4. In situ crystal characterization. (A) Experimental setup on beamline X06DA at SLS with the CATS robotic arm (red) holding the chip holder (green). The PILATUS 2 M detector is on the right, and the beam comes from the left. (B) Close-up view of the chip on the SBS holder maintained in the beam by the robot grip. (C) Lysozyme crystal filling the entire microfluidic channel as seen by the alignment camera. The red oval symbolizes the position of the beam (80 × 45 μm^2). After two data collections of 1 min, beam impacts became visible on the crystal (indicated by white arrows), and gas bubbles formed in the mother liquor. (D) Radiation-damaged insulin crystal after a 60° data collection at room temperature in a microplate. The small imprint left by the beam (0.08 mm × 0.04 mm) indicates the very accurate sphere of confusion of CATS robots during data collection. (E) and (F) Refined thaumatin and lysozyme structures at 1.55 Å and 1.44 Å resolution, respectively, in final $2F_o - F_c$ electron density maps contoured at 1.2σ (see statistics in Table 2). (G) Structures of lysozyme solved by Yb-SAD and (H) insulin solved by sulfur-SAD. The initial $2F_o - F_c$ map (after density modification with SHELXE) and anomalous difference map contoured at 1σ and 5σ, respectively. Ytterbium and sulfur signals are clearly visible in anomalous difference maps (magenta). Images of protein structures were generated using PyMOL (http://www.pymol.org).

robot stopped its rotation after each image, moved backward, and accelerated again to reach the right speed and position for the next exposure in stop-start mode. These back and forth movements led to slight misalignments of the crystals which were reflected in statistics (e.g., overall *R*-merge and completeness), but the quality of the crystals was fine as judged from resolution, mosaicity, and overall *B*-factor (Table 1). Misalignments due to crystal movements could be excluded because analyzed samples filled the chip channels as illustrated in Figure 4.

This difficulty was overcome by using PILATUS pixel detectors which enable very fast shutterless data acquisition during the continuous rotation of the sample.[22] Instead of collecting 20−60 individual oscillations of 1−2°, fine sliced data sets were recorded at up to 40 Hz. There was a significant gain in R_{meas}, signal-to-noise ratio, and in completeness as seen in Figure 5 and Table 1. The improvement in mosaic spread confirmed that values observed in stop-start mode were overestimated and accounted for partial overlap of or gap between contiguous images.

Fine φ-slicing was recently reported to improve scaling statistics,[23] and our tests on thaumatin crystals indicated that thin images at low dose gave best results (for instance, 0.05° oscillation range, 0.05 s exposure, beam size 85 × 45 μm^2 and 75% attenuation corresponding to ∼3 × 10^{12} ph/s). The same oscillation range was applied for lysozyme crystals. With such parameters, two to three successive paths were recorded in a couple of minutes on single crystals that were translated to minimize radiation damage (traces of the beam are visible in Figure 4C). This increased the multiplicity and provided high quality data for the refinement of thaumatin and lysozyme structures (Table 2; Figure 4D,E).

SAD Data Collection in ChipX and Microplates. We also tested the feasibility of experimental phasing using PILATUS detectors on crystals grown in microfluidic chips setups as well as in microplates. First, we made an attempt of sulfur-SAD phasing in the ChipX, but the anomalous signal was too weak and not exploitable, probably due to the small size of crystals. Second, we performed a similar experiment on larger insulin crystals (∼250 × 250 × 100 μm^3) grown in a CrystalQuick X microplate held in the beam by the CATS robot. In this case, the S-SAD measurements yielded anomalous data which led to a successful structure determination (Table 2, Figure 4H). This showed that the positioning of the sample during the rotation with CATS robots is accurate enough for SAD complete data set collection (Figure 4D,H). Third, we exploited a strategy developed by Gavira et al.[24] and grew lysozyme crystals in ChipX by counter-diffusion in the presence of ytterbium derivative[11] to increase the anomalous signal. Exploitable Yb-SAD data were collected at the peak wavelength (1.38 Å) in ChipX attached on a multiaxis PRIGo goniometer (Table 2, Figure 4G). These results confirmed that data of sufficient quality for experimental phasing can be collected at room temperature from single crystals in dedicated crystallization setups such as microplates or chips.

■ CONCLUSION

Our novel microfluidic chip is compatible with the crystallization of biological macromolecules and the *in situ* crystal analysis by X-ray diffraction. It is distinguished from previous devices by microchannels that are longer and have at smaller sections. The advantages of the ChipX are 4-fold. First, the complete counter-diffusion process can occur for an optimal crystal growth on minimal sample volumes. Second, a compromise

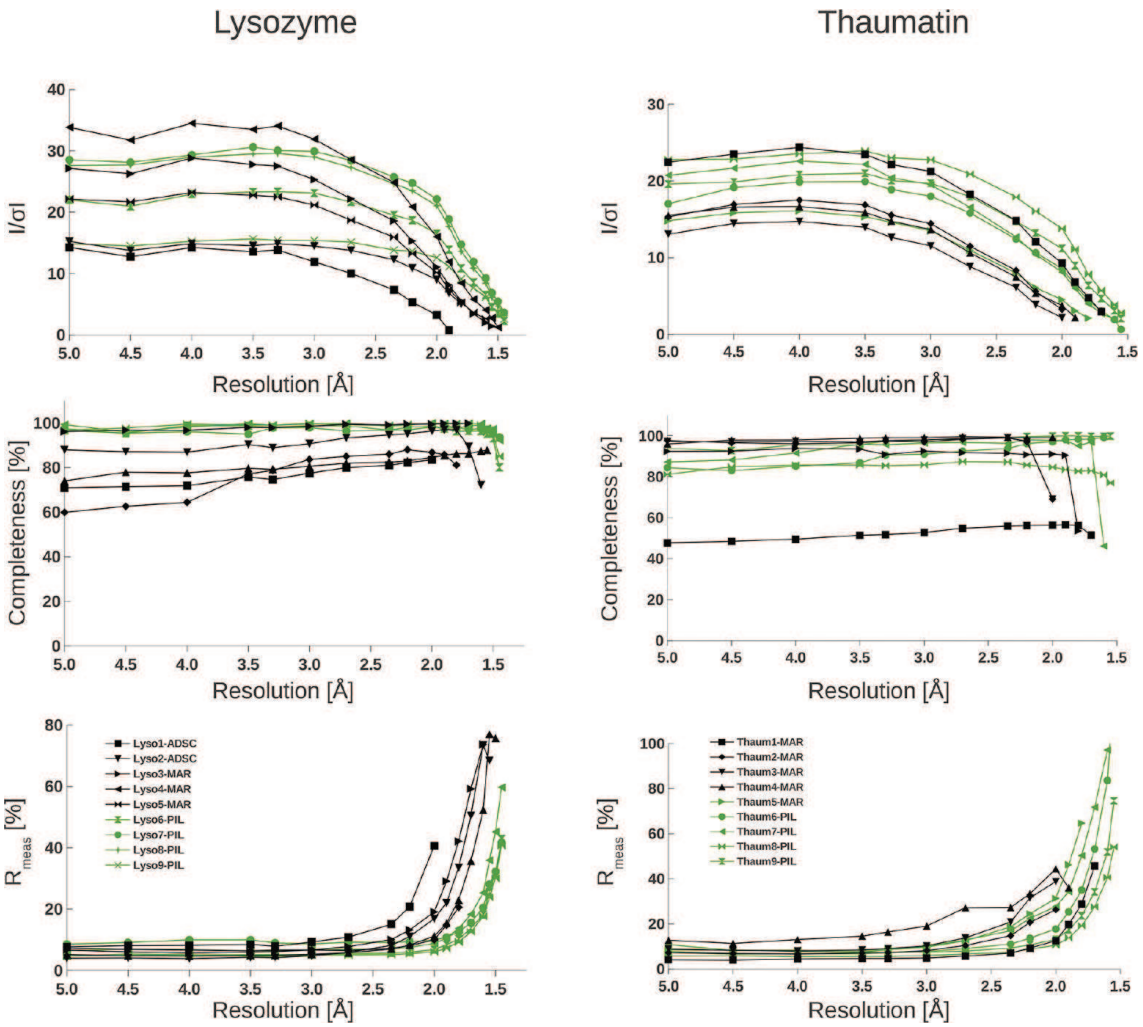


Figure 5. Quality of data collected on lysozyme (left) and thaumatin (right) crystals. Three indicators are given for the data sets detailed in Table 1: completeness, R_{meas}, and signal-to-noise ratio as a function of resolution. Black and green plots correspond to data measured either with CCD detectors (ADSC Q315r, MAR225) or with a PILATUS 2 M pixel detector (PIL). The latter allowed crystal analyses in continuous mode with fine slicing. Resulting data sets showed a higher homogeneity and better statistics.

was found between the overall thickness of the chip material — with lowest background scattering and highest signal-to-noise ratio during crystal analysis — and the rigidity of the device, facilitating fabrication and handling. All steps from crystal growth by counter-diffusion to crystal analysis on-chip using synchrotron radiation sources have been validated. Hence, this new chip design opens new opportunities for fast, efficient, and cost-effective production of high quality crystals in miniaturized systems. Third, the ChipX device offers the possibility of collecting full X-ray diffraction data sets at room temperature. This is particularly promising in the case of fragile biological samples that are difficult to handle or to flash cool, such as crystals of membrane proteins, viruses, or other large assemblies. A comparable in situ characterization strategy was recently used to solve the structure of a bovine enterovirus by merging data collected from dozens of crystals that could not be cryocooled.[25] Finally, with the development of serial crystallography both at synchrotron and X-ray free-electron laser,[26] processing software will soon enable the combination of partial data sets on a routine basis. The increasing number of fast pixel detectors on synchrotron beamlines coupled to robotic sample holders and automated crystal analysis protocols, as illustrated in this work, will make room-temperature data collection much more user-friendly and accessible. In this context, crystallization chips like the ChipX will provide a versatile platform to grow calibrated crystals. Inside these chips, the latter are protected against mechanical stress and dehydration and are ready for safe shipping to beamlines for native or anomalous phasing data collection, without any loss of quality due to unnecessary handling.

AUTHOR INFORMATION

Corresponding Author

*E-mail: c.sauter@ibmc-cnrs.unistra.fr. Tel: +33 388 417 102. Fax: +33 388 602 218.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the teams of FIP-BM30A beamline at ESRF (Grenoble, France) and of X06DA beamline at SLS (Villigen, Switzerland) for the beamtime allocated to this project.

This research was supported by the PNANO Programme from the Agence Nationale pour la Recherche (ANR-07-NANO-060) and by the French National Program "Investissement d'Avenir" (Labex MitCross; ANR-10-IDEX-002-02). Nano-Lyon facilities were used for microfabrication steps. F.P. benefited from a doctoral grant from the University of Strasbourg. M.B. was the recipient of a grant from the French Ministry of Research.

REFERENCES

(1) Giegé, R.; Sauter, C. Biocrystallography: past, present, future. *HFSP J.* **2010**, *4*, 109.
(2) Sauter, C.; Lorber, B.; McPherson, A.; Giegé, R. Crystallization - General Methods. In *International Tables of Crystallography, Vol. F, Crystallography of Biological Macromolecules*, 2nd ed.; Arnold, E., Himmel, D. M., Rossmann, M. G., Eds; John Wiley and Sons: Chichester, 2012; pp 99−120.
(3) Chayen, N. E.; Shaw Stewart, P. D.; Baldock, P. New developments of the IMPAX small-volume automated crystallization system. *Acta Crystallogr. D: Biol. Crystallogr.* **1994**, *50*, 456−458.
(4) Pusey, M. L.; Liu, Z.-J.; Tempel, W.; Praissman, J.; Lin, D.; Wang, B.-C.; Gavira, J. A.; Ng, J. D. Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 359−386.
(5) Sauter, C.; Dhouib, K.; Lorber, B. From Macrofluidics to Microfluidics for the Crystallization of Biological Macromolecules. *Cryst. Growth Des.* **2007**, *7*, 2247−2250.
(6) Hansen, C. L.; Skordalakes, E.; Berger, J. M.; Quake, S. R. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16531−16536.
(7) Zheng, B.; Roach, L. S.; Ismagilov, R. F. Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets. *J. Am. Chem. Soc.* **2003**, *125*, 11170−11171.
(8) Ng, J. D.; Clark, P. J.; Stevens, R. C.; Kuhn, P. In situ X-ray analysis of protein crystals in low-birefringent and X-ray transmissive plastic microchannels. *Acta Crystallogr. D: Biol. Crystallogr.* **2008**, *64*, 189−197.
(9) Stojanoff, V.; Jakoncic, J.; Oren, D. A.; Nagarajan, V.; Poulsen, J.-C. N.; Adams-Cioaba, M. A.; Bergfors, T.; Sommer, M. O. A. From screen to structure with a harvestable microfluidic device. *Acta Crystallograph. Sect. F: Struct. Biol. Cryst. Commun.* **2011**, *67*, 971−975.
(10) Dhouib, K.; Khan Malek, C.; Pfleging, W.; Gauthier-Manuel, B.; Duffait, R.; Thuillier, G.; Ferrigno, R.; Jacquamet, L.; Ohana, J.; Ferrer, J.-L.; Théobald-Dietrich, A.; Giegé, R.; Lorber, B.; Sauter, C. Microfluidic chips for the crystallization of biomacromolecules by counter-diffusion and on-chip crystal X-ray analysis. *Lab. Chip* **2009**, *9*, 1412−1421.
(11) Girard, E.; Anelli, P. L.; Vicat, J.; Kahn, R. High-phasing-power lanthanide derivatives: taking advantage of ytterbium and lutetium for optimized anomalous diffraction experiments using synchrotron radiation. *Acta Crystallogr. D: Biol. Crystallogr.* **2003**, *59*, 1877−1880.
(12) Williams, J. D.; Wang, W. Microfabrication of an electromagnetic power relay using SU-8 based UV-LIGA technology. *Microsyst. Technol.* **2004**, *10*, 699−705.
(13) Le Maire, A.; Gelin, M.; Pochet, S.; Hoh, F.; Pirocchi, M.; Guichou, J. F.; Ferrer, J. L.; Labesse, G. In-plate protein crystallization, in situ ligand soaking and X-ray diffraction. *Acta Crystallogr. D: Biol. Crystallogr.* **2011**, *67*, 747−755.
(14) Jacquamet, L.; Ohana, J.; Joly, J.; Borel, F.; Pirocchi, M.; Charrault, P.; Bertoni, A.; Israel-Gouy, P.; Carpentier, P.; Kozielski, F.; Blot, D.; Ferrer, J.-L. Automated analysis of vapor diffusion crystallization drops with an X-ray beam. *Structure* **2004**, *12*, 1219−1225.
(15) Bingel-Erlenmeyer, R.; Olieric, V.; Grimshaw, J. P. A.; Gabadinho, J.; Wang, X.; Ebner, S. G.; Isenegger, A.; Schneider, R.; Schneider, J.; Glettig, W.; Pradervand, C.; Panepucci, E. H.; Tomizaki, T.; Wang, M.; Schulze-Briese, C. SLS Crystallization Platform at Beamline X06DA—A Fully Automated Pipeline Enabling in Situ X-ray Diffraction Screening. *Cryst. Growth Des.* **2011**, *11*, 916−923.
(16) Kabsch, W. XDS. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(17) Pape, T.; Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **2004**, *37*, 843−844.
(18) Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **2007**, *64*, 112−122.
(19) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(20) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 213−221.
(21) Otálora, F.; Gavira, J. A.; Ng, J. D.; García-Ruiz, J. M. Counterdiffusion methods applied to protein crystallization. *Prog. Biophys. Mol. Biol.* **2009**, *101*, 26−37.
(22) Rajendran, C.; Dworkowski, F. S. N.; Wang, M.; Schulze-Briese, C. Radiation damage in room-temperature data acquisition with the PILATUS 6M pixel detector. *J. Synchrotron Radiat.* **2011**, *18*, 318−328.
(23) Mueller, M.; Wang, M.; Schulze-Briese, C. Optimal fine φ-slicing for single-photon-counting pixel detectors. *Acta Crystallogr. D: Biol. Crystallogr.* **2012**, *68*, 42−56.
(24) Gavira, J. A.; Toh, D.; Lopéz-Jaramillo, J.; García-Ruíz, J. M.; Ng, J. D. Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling. *Acta Crystallogr. D: Biol. Crystallogr.* **2002**, *58*, 1147−1154.
(25) Axford, D.; Owen, R. L.; Aishima, J.; Foadi, J.; Morgan, A. W.; Robinson, J. I.; Nettleship, J. E.; Owens, R. J.; Moraes, I.; Fry, E. E.; Grimes, J. M.; Harlos, K.; Kotecha, A.; Ren, J.; Sutton, G.; Walter, T. S.; Stuart, D. I.; Evans, G. In situ macromolecular crystallography using microbeams. *Acta Crystallogr. D: Biol. Crystallogr.* **2012**, *68*, 592−600.
(26) Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U.; Doak, R. B.; Maia, F. R. N. C.; Martin, A. V.; Schlichting, I.; Lomb, L.; Coppola, N.; Shoeman, R. L.; Epp, S. W.; Hartmann, R.; Rolles, D.; Rudenko, A.; Foucar, L.; Kimmel, N.; Weidenspointner, G.; Holl, P.; Liang, M.; Barthelmess, M.; Caleman, C.; Boutet, S.; Bogan, M. J.; Krzywinski, J.; Bostedt, C.; Bajt, S.; Gumprecht, L.; Rudek, B.; Erk, B.; Schmidt, C.; Hömke, A.; Reich, C.; Pietschner, D.; Strüder, L.; Hauser, G.; Gorke, H.; Ullrich, J.; Herrmann, S.; Schaller, G.; Schopper, F.; Soltau, H.; Kühnel, K.-U.; Messerschmidt, M.; Bozek, J. D.; Hau-Riege, S. P.; Frank, M.; Hampton, C. Y.; Sierra, R. G.; Starodub, D.; Williams, G. J.; Hajdu, J.; Timneanu, N.; Seibert, M. M.; Andreasson, J.; Rocker, A.; Jönsson, O.; Svenda, M.; Stern, S.; Nass, K.; Andritschke, R.; Schröter, C.-D.; Krasniqi, F.; Bott, M.; Schmidt, K. E.; Wang, X.; Grotjohann, I.; Holton, J. M.; Barends, T. R. M.; Neutze, R.; Marchesini, S.; Fromme, R.; Schorb, S.; Rupp, D.; Adolph, M.; Gorkhover, T.; Andersson, I.; Hirsemann, H.; Potdevin, G.; Graafsma, H.; Nilsson, B.; Spence, J. C. H. Femtosecond X-ray protein nanocrystallography. *Nature* **2011**, *470*, 73−77.
(27) Karplus, P. A.; Diederichs, K. Linking crystallographic model and data quality. *Science* **2012**, *336*, 1030−1033.

E.3 Publication 5: The novel RNase P in action

The novel RNase P in action

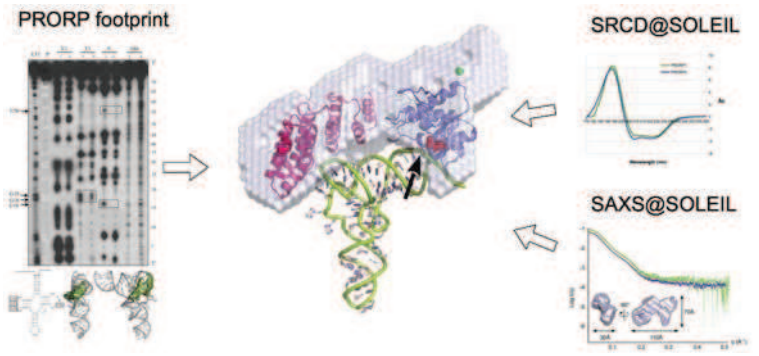
A novel type of RNase P was recently identified which is totally deprived of catalytic RNA. This proteinaceous RNase P (or PRORP) is found in the organelles of many eukaryotes and in the nucleus of some eukaryotes including plants. In order to characterize the architecture of PRORP enzymes and to determine how they bind to pre-tRNAs to perform their 5' maturation we combined biochemical and biophysical approaches. The resulting model of a functional PRORP:substrate complex suggests a tRNA recognition mode similar to that of the ribonucleoproteic RNase P.

5' tRNA maturation in mitochondria

Transfer RNAs (tRNAs) are key actors of protein synthesis: they play the role of adapter molecules during the translation of messenger RNAs by the ribosome into protein sequences. They are produced as precursors with leading and trailing sequences that need to be processed. Their 5' maturation is catalyzed by a ubiquitous enzyme called RNase P. Until recently all known RNase P were ribonucleoproteins, the catalytic activity of the enzyme being held by an RNA molecule. In 2008, a new type of RNase P only composed of proteins was identified in human mitochondria [1] that corresponds to a novel family of nucleases called PRORP for "Proteinaceous RNase P". The group of Philippe Giegé (Institut de Biologie Moléculaire des Plantes, IBMP, Strasbourg) demonstrated that the model plant *Arabidopsis thaliana* possesses three PRORP proteins. PRORP1 is localised in both mitochondria and chloroplasts whereas PRORP2 and PRORP3 are active in the nucleus [2,3]. A collaboration was initiated between two neighbouring institutes in Strasbourg (IBMP and IBMC) to examine these enzymes from *A. thaliana* and a combination of biochemical and biophysical approaches was used to gain a first structural and functional insight into tRNA recognition and maturation by PRORPs.

PRORP: an integrated structural study

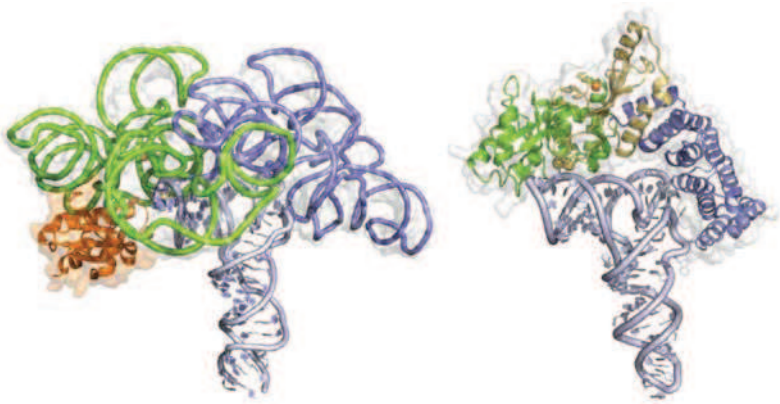
PRORP sequences are characterized by the presence of pentatricopeptide repeat (PPR) motifs and a metallonuclease domain proposed to hold the catalytic center. Because no structure of a close homologue was known at the time we started this study, comparative modeling was carried out on separate domains. We then performed synchrotron radiation circular dichroism (SRCD) on the DISCO beamline to validate the models based on their 2D structure content and to test the conformational stability of PRORP samples prior to further investigations. The presence of a zinc binding motif between the two main domains was demonstrated by site directed mutagenesis of putative zinc chelating residues in association with inductively coupled plasma mass spectrometry. Small angle X-ray scattering (SAXS) data collected on the SWING beamline confirmed the two domain organization of PRORPs and helped place them with respect to each other (Figure ❶).



❶ A first glance at a PRORP:tRNA complex. PRORP2 of *A. thaliana* was built by comparative modelling guided by SRCD and SAXS data, and the resulting model was docked onto a pre-tRNA substrate based on footprint analysis. The RNA cleavage position is indicated by an arrow.

Probing PRORP:tRNA interface

To position PRORP on its RNA substrate, the latter was subjected to RNase digestion in the presence of the enzyme. The protection footprint (Figure ❶) defined the interaction interface and the PRORP enzyme was docked accordingly onto the 3D structure of a pre-tRNA. This model of the maturation complex reveals that eukaryotes have evolved PPR proteins to recognize pre-tRNAs in a similar way as the ribonucleoproteic RNase P reminiscent from the ancient *RNA world* (Figure ❷). Although the scenario of this convergent evolution remains to be established, as well as the precise catalytic mechanism of tRNA maturation, this study is a first step towards the detailed characterization of the PRORP family.



❷ Classical ribonucleoproteic RNase P (left, PDB id: 3Q1R) and PRORP2 (right, model based on PDB id: 4G26) share the same pre-tRNA binding mode [4].

SWING & DISCO beamlines

ASSOCIATED PUBLICATION

Structural insights into protein-only RNase P complexed with tRNA

A. Gobert, F. Pinker, O. Fuchsbauer, B. Gutmann, R. Boutin, P. Roblin, C. Sauter* and P. Giegé.

Nature Communications 4 (2013) art.1353

CORRESPONDING AUTHOR

*Laboratoire "Architecture et Réactivité de l'ARN", UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire (IBMC), Université de Strasbourg, 15 rue René Descartes, 67084 Strasbourg, France

c.sauter@ibmc-cnrs.unistra.fr

REFERENCES

[1] Holzmann et al. Cell 135 (2008), 462
[2] Gobert et al. Nat. Struct. Mol. Biol. 17 (2010), 740
[3] Gutmann et al. Genes Dev. 26 (2012), 1022
[4] Pinker et al. RNA biology 10 (2013), 1457

E.4 Résumé de thèse

RESUME

La production d'ARNt mature et fonctionnel nécessite plusieurs étapes de clivage et modifications des bases. Parmi elles, la maturation de l'extrémité 5' est réalisée par la RNase P. C'est la première enzyme à ARN catalytique, ou ribozyme, assisté d'une ou plusieurs protéines auxiliaires, qui a été découverte dans les années 70 par Sidney Altman. Par la suite, sa caractérisation dans divers organismes a fait germer l'idée d'une RNase P ribonucléoprotéique universellement conservée comme une relique d'un monde à ARN. Cependant dès la fin des années 80, cette universalité a été remise en question par des observations indiquant que ni la RNase P de chloroplastes d'épinard, ni celle des mitochondries humaines ne semblaient contenir de sous-unité ARN. En outre, aucun gène de l'ARN de la RNase P n'était identifiable dans les génomes des plantes.

En 2008, le groupe de Walter Rossmanith à Vienne (Autriche) a été le premier à démontrer que la RNase P des mitochondries humaines est un complexe de trois protéines (MRPP 1-3 pour **M**itochondrial **R**Nase **P** **P**roteins 1-3) et que ce type de RNase P ne nécessite pas d'ARN catalytique pour son activité. Cette étude a proposé que c'est la sous-unité MRPP3, que l'on nomme désormais PRORP pour **PRO**teinaceous **R**Nase **P**, qui porte la fonction catalytique.

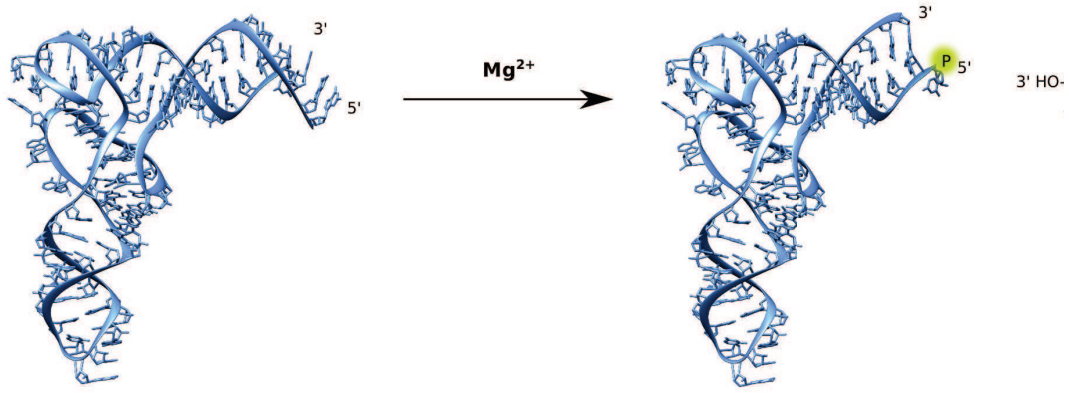


Illustration 1: Activité catalytique de la RNase P. Clivage de la séquence 5' leader d'un ARNt précurseur en présence de ions divalent comme le magnésium ou le manganèse.

Deux ans plus tard, en 2010, le groupe dirigé par Philippe Giegé à Strasbourg a montré qu'il existe trois orthologues de PRORP humaine chez la plante *Arabidopsis thaliana* et que ces protéines sont actives sous forme monomérique isolée. PRORP1 est située dans les

organites (chloroplastes et mitochondries), alors que PRORP2 et PRORP3 sont localisées dans le noyau.

Les enzymes PRORP sont constituées de deux domaines : un domaine qui reconnaît spécifiquement des séquences d'ARN non appariés, le domaine PPR, et un domaine nucléase de type NYN qui assure la coupure endonucléolytique 5' des précurseurs d'ARNt. Les protéines PRORP appartiennent donc à la famille PPR ou protéines à **r**épétitions de **p**entatrico**p**eptides. Un motif PPR classique se compose de 35 acides aminés et forme deux hélices alpha antiparallèles, plusieurs motifs pouvant s'associer en super-hélice. Les protéines à PPR possèdent de 2 à 27 motifs PPR et éventuellement d'autres domaines en C-terminal. Par exemple, le domaine catalytique NYN de PRORP présente un site actif qui fixe des ions métalliques (magnésium). Les deux domaines sont reliés par un domaine bipartite qui implique une région placée entre les domaines PPR et NYN, et l'extrémité C-terminale de PRORP. Ce domaine est caractérisé par un motif de liaison d'un ion zinc qui stabilise la structure de cette jonction.

L'objectif de ma thèse était de contribuer à la caractérisation de cette nouvelle famille de protéines PRORP par des approches biochimiques, biophysiques et structurales. De façon plus générale, rien n'était connu lorsque j'ai débuté ce travail sur l'organisation structurale des protéines PPR, ni sur la façon dont elles assurent une reconnaissance spécifique des séquences d'ARN. De ce fait, l'étude des protéines PRORP était aussi l'occasion de mieux connaître la famille des protéines PPR, particulièrement vaste chez les plantes et impliquée dans de nombreuses étapes du métabolisme de l'ARN.

J'ai commencé par développer et optimiser des protocoles de purification des trois protéines PRORP sauvages d'*A. thaliana* dotées d'une étiquette à 6-His afin d'en obtenir en quantité et pureté suffisantes en vue d'études structurales. Le premier but était de résoudre leur structure 3D à haute résolution par cristallographie, ce qui nécessitait la préparation de molécules pures et homogènes en vue de les cristalliser. L'homogénéité des échantillons a systématiquement été vérifiée par la méthode de diffusion dynamique de la lumière (DLS). Il était également crucial de s'assurer de l'activité des enzymes et donc de leur capacité à réaliser un clivage endonucléolytique en 5' d'un ARNt précurseur.

L'étape de cristallisation s'est, comme souvent, révélée la plus délicate. J'ai recherché des conditions propices en utilisant différents cribles commerciaux en microplaques à l'aide d'un robot de cristallisation qui permet de réaliser des expériences utilisant des volumes de l'ordre

de 100 nl. Au total, une dizaines de cribles (96 conditions chacun) ont été testés pour PRORP2, puis PRORP1 et PRORP3 sauvages, ainsi que pour des mutants catalytiques. Les résultats les plus prometteurs ont systématiquement été obtenus en présence de polyéthylène glycol, et une optimisation m'a permis d'obtenir des données de diffraction de cristaux de PRORP2 sauvage à 3 Å de résolution. Ces cristaux tricliniques contiennent deux exemplaires de l'enzyme dans l'unité asymétrique, mais ont jusqu'ici résisté à toutes tentatives de détermination de la structure de PRORP2 par remplacement moléculaire. La poursuite de l'amélioration de cette forme cristalline, toujours en cours, devrait apporter des données à meilleure résolution et une solution alternative de phasage.

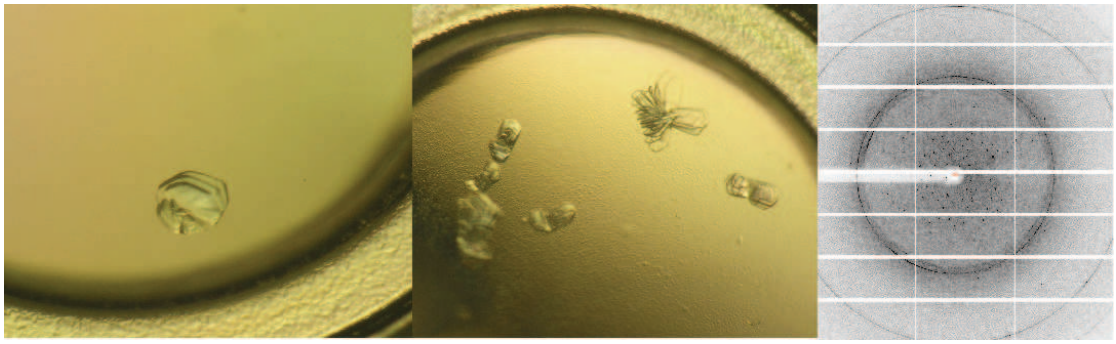


Illustration 2: Cristaux de PRORP2 sauvage et clichés de diffraction à 3 A de resolution sur la ligne PXIII au synchrotron suisse (SLS) grâce à un détecteur de type Pilatus 2M. La maille cristalline triclinique a pour paramètres a = 70.5 A, b = 72.4 A, c = 80 A, alpha = 63 °, beta = 72.3 °, gamma = 78.5 °. Diffraction jusqu'à 3.2 A.

En parallèle de l'approche cristallographique, j'ai mené des expériences de dichroïsme circulaire sous rayonnement synchrotron (SRCD) au synchrotron SOLEIL (St. Aubin, France) pour obtenir des informations sur la composition en structures secondaires des enzymes PRORP. Les données SRCD indiquent que PRORP1 contient 36 % d'hélices alpha et 15 % des feuillets beta, ces valeurs étant de 39 % et 16 % pour PRORP2, en accord avec les prédictions de structure basées sur des analyses de séquences.

En combinant la mutagenèse dirigée et la méthode de spectrométrie de masse par torche de plasma (ICP-MS), nous avons pu déterminer quels acides aminés étaient responsables de la fixation du zinc dans le domaine intermédiaire des PRORP.

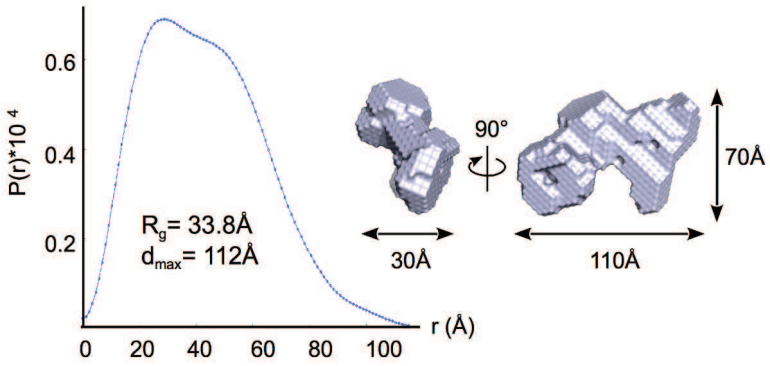


Illustration 3: Analyse de PRORP par SAXS. A gauche fonction de P(r) qui montre que PRORP2 contient deux domaines distincts et une dimensions maximale de 112 A. A droite un modèle ab initio dérivé d'envelope moléculaire de PRORP2.

Illustration 3: Modèle de PRORP2 nucléaire associé à un de ses substrats, le précurseur de l'ARNt cystéine.

Pour compléter cette analyse structurale en solution, j'ai utilisé la diffusion des rayons X aux petits angles (SAXS), également à SOLEIL. Cette méthode permet déterminer le rayon de giration de la molécule ou du complexe étudiés, d'en évaluer la globularité et la présence éventuelle de plusieurs domaines, voir d'obtenir une idée de leur forme à basse résolution. Ces expériences ont confirmé que PRORP 1 et 2 contiennent bien deux domaines distincts, avec un rayon de giration de 33 Å et une dimension maximale de 110 Å. Nous avons pu en tirer un premier modèle des PRORP en solution et de leur organisation globale, qui ont été confirmés par la structure 3D de PRORP1 publiée au même moment par un groupe concurrent. Ces résultats ont fait l'objet d'un l'article intitulé « Structural insights into protein-only RNase P complexed with tRNA » dans Nature Communications début 2013.

Le second objectif de mon travail était la détermination de la structure cristallographique du complexe entre un précurseur d'ARNt et PRORP. Pour faire face à différentes difficultés rencontrées dans cette projet, j'ai été amenée à me former à plusieurs méthodes biophysiques comme la microcalorimétrique (ITC pour isothermal tritration calorimetry), la thermophorèse (MST pour microscale thermophoresis), le retard de migration sur gel, la spectroscopie de fluorescence et l'ultracentrifugation analytique (AUC). Je les ai mises à profit pour tester différents précurseurs de l'ARNt cystéine mitochondrial *d'Arabidopsis* et déterminer leur constante d'affinité vis à vis du mutant catalytique de PRORP2. Les

premières tentatives d'ITC ne montraient aucun signe d'interaction entre le substrat et l'enzyme. Des mesures de DLS ont alors révélé une agrégation des échantillons dans la cellule du calorimètre, ce qui m'a poussée à définir plus précisément un tampon permettant l'activité de l'enzyme, d'une part, et un substrat ARN minimal mais bien cliv*able, d'autre part.*

Ces expériences m'ont en particulier permis de sélectionner un précurseur contenant cinq nucléotides en 5' et aucun en 3' (L5T0) pour les essais de formation du complexe stable. Avec un tampon optimisé à la fois pour la stabilité et l'activité de la protéine, j'ai pu déterminé un K_D de 1 µM par ITC pour un mutant catalytique de PRORP2 vis à vis du précurseur $ARNt^{Cys}$ *L5T0,* valeur confirmée par les méthodes MST et AUC. Un avantage de l'ITC est qu'on peut réutiliser l'échantillon en fin d'analyse, lorsque le complexe a été formé, pour réaliser des essais de cristallisation. Toutefois les premiers cribles réalisés n'ont pas donné des cristaux pour l'instant. En parallèle de cette recherche de conditions de cristallisation, je concentre mes efforts sur la détermination d'une enveloppe moléculaire du complexe par diffusion des rayons X aux petits angles, en exploitant le travail que j'ai mené en amont pour définir les conditions de formation d'un complexe stable.

En conclusion, mon travail de thèse a conduit à un premier modèle structural de cette nouvelle famille de RNase P protéique et à une description de l'interaction entre ces enzymes PRORP et leurs substrats pré-ARNt. La cristallisation et la caractérisation cristallographique des PRORP d'*A. thaliana* que j'ai initiées ont apporté des résultats très prometteurs et sont sur le point d'aboutir. Par ailleurs, mon étude de la stabilité et l'activité du complexe en solution m'a permis de me former à un ensemble de méthodes biophysiques très complémentaires de l'approche cristallographique. Elle fournit en outre une base solide pour la poursuite d'une étude structurale approfondie du mode d'action des PRORP et, plus généralement, de la reconnaissance des ARN par les protéines PPR.

Liste des publications

Gobert A*, **Pinker F***, Fuchsbauer O, Gutmann B, Boutin R, Roblin P, Sauter C, Giegé P. 2013. Structural insights into protein-only RNase P complexed with tRNA. Nature Communications, **4**, 1353 (*equal contribution).

Pinker F, Brun M, Morin P, Deman A, Chateaux JF, Oliéric V, Stirnimann C, Lorber B, Terrier N, Ferrigno R, Sauter C. 2013. ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature. Crystal Growth and Design, **13** (8), 3333–3340.

Pinker F, Bonnard G, Gobert A, Gutmann B, Hammani K, Sauter C, Gegenheimer PA, Giegé P. 2013. PPR proteins shed a new light on RNase P biology. RNA Biology, **10**(9):1457-68.

Hammani K, Bonnard G, Bouchoucha A, Gobert A, **Pinker F**, Salinas T, Giegé P. 2014. Helical repeats modular proteins are major players for organelle gene expression. Biochimie, **100,** 141-50.

Communications orales

Natural antibiotic Microcin C is a tighter binder than synthetic inhibitors. *Mitomot Consortium ; Strasbourg, France* (26 mars 2012).

Structural characterization of a novel proteinaceous RNase P family in *Arabidopsis thaliana.* 5th Japan-China-Korea Graduate Student Forum ; Tsukuba, Japan (21 septembre – 24 septembre 2012).

Looking for PRORP, Perspectives in tRNA biology, Strasbourg, France (28-30 octobre, 2012).

Structural studies of proteinaceous RNase P, RNA workshop, Hirschegg, Autriche (19-23 marc, 2013)

Structural charcterization of a novel proteinaceous RNase P family from *Arabidopsis*. MitoCross Summer School ; Strasbourg, France (14 – 15 juin, 2013)

Communications par affiche

Pinker F, Gobert A, Fuchsbauer O, Giegé P & Sauter C**.** Structural characterization of a novel eukaryotic family of proteinaceous RNase P. FEBS advanced course "Advanced methods in protein crystallization" *;* Nove Hrady, République Tchèque (22 – 29 juin 2012).

Pinker F, Gobert A, Fuchsbauer O, Gutmann B, Boutin R, Roblin P, Giegé P & Sauter C**.** Structural characterization of a novel eukaryotic family of proteinaceous RNase P. *MitoStrass ;* Strasbourg, France (18 décembre 2012).

F. Pinker, A. Gobert, O. Fuchsbauer, B. Gutmann, B. Boutin, P. Roblin, P. Giegé & C. Sauter**.** Structural insights into proteinaceous RNase P complexed with tRNA. *International crystallography school ;* Erice, Italy (30 mai – 6 juin 2013).

Gobert A, **Pinker F**, Bonnard G, Hammani K, Bouchoucha A, Sauter C & Giegé P**.** PPR proteins shed a new light on RNase P biology. *Mitocross Labex meeting;* Strasbourg, France (25 novembre 2013).

Pinker F, Gobert A, Fuchsbauer O, Gutmann B, Boutin R, Roblin P, Giegé P & Sauter C**.** Structural insights into proteinaceous RNase P complexed with tRNA. *RNA meeting 2014 ;* Quebec, Canada (03 juin – 08 juin 2014).

E.5 Validated trainings at the doctoral school ED414

Formations validées par PINKER Franziska

Formations du type Socio-Professionnel

Intitulé de la formation	Session	Durée validée
Author writing workshop - ELSEVIER	2011/2012	02h30
Cours de français B2	2011/2012	48h00
Effective writing March 2012	2011/2012	18h00
Intellectual property rights - 25-26 Nov 2011	2011/2012	16h00
Poster presentation - PhD Day 2014	2013/2014	05h00
	Total des formations	**89h00**

Formations du type Scientifique

Intitulé de la formation	Session	Durée validée
5th Japan-China-Korea Graduate students Forum	2011/2012	13h00
Biomolecular visualization with Pymol - 5,6 Jan 2012	2011/2012	12h00
Mito@Strass 2012	2012/2013	08h00
Mitochodria in health, disease and death of the cell	2012/2013	14h00
Journée des Doctorants ED414 - 19/02/2014	2013/2014	08h00
	Total des formations	**55h00**

Formations du type Congrès

Intitulé de la formation	Session	Durée validée
26th Rhine-Knee Regional Meeting on Biochrystallography	2011/2012	12h00
	Total des formations	**12h00**

Fait à Strasbourg le 20 juin 2014

Ecole Doctorale des Sciences de la Vie et de la Santé
CDE, 46 Boulevard de la Victoire - 67000 STRASBOURG
Tél. : 03 68 85 16 95 - Courriel : melanie.muser@unistra.fr
http://www-ed-sdvs.u-strasbg.fr

Bibliography

Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T., & Zwart, P. 2010. *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, **66**(2), 213–221.

Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C. M., Bruni, C. B., & Carlomagno, M. S. 1994. Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA. *Genes & Development*, **8**(24), 3021–3031.

Altman, S., Wesolowski, D., Guerrier-Takada, C., & Li, Y. 2005. RNase P cleaves transient structures in some riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(32), 11284–11289.

Anraku, Y., & Satow, Y. 2009. Reflections on protein splicing: structures, functions and mechanisms. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, **85**(9), 409–421.

Anraku, Y., Mizutani, R., & Satow, Y. 2005. Protein Splicing: Its Discovery and Structural Insight into Novel Chemical Mechanisms. *IUBMB Life*, **57**(8), 563–574.

Barkan, A., & Small, I. 2014. Pentatricopeptide Repeat Proteins in Plants. *Annual Review of Plant Biology*, **65**(1), 415–442.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y., Bond, Ch., & Small, I. 2012. A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet*, **8**(8), e1002910.

Betat, H., Rammelt, C., & Mörl, M. 2010. tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. *Cellular and molecular life sciences: CMLS*, **67**(9), 1447–1463.

Biertümpfel, C., Basquin, J., Suck, D., & Sauter, C. 2002. Crystallization of biological macromolecules using agarose gel. *Acta Crystallographica Section D Biological Crystallography*, **58**(10), 1657–1659.

Biolabs, New England. 2014. *Bypassing Common Obstacles in Protein Expression*. `https://www.neb.com/tools-and-resources/feature-articles/`

bypassing-common-obstacles-in-protein-expression. [Online; accessed 2014-04-12].

Brown, J. 1997. The ribonuclease P database. *Nucleic Acids Research*, **25**(1), 263–264.

Brown, J., Haas, E., James, B., Hunt, D., Liu, J., & Pace, N. 1991. Phylogenetic analysis and evolution of RNase P RNA in proteobacteria. *Journal of Bacteriology*, **173**(12), 3855–3863.

Brown, J. W., & Pace, N. R. 1992. Ribonuclease P RNA and protein subunits from bacteria. *Nucleic Acids Research*, **20**(7), 1451–1456.

Candle. 2014. *Synchrotron radiation and applications*. http://www.candle.am.

Canino, G., Bocian, E., Barbezier, N., Echeverria, M., Forner, J., Binder, S., & Marchfelder, A. 2009. Arabidopsis Encodes Four tRNase Z Enzymes. *Plant Physiol.*, **150**(3), 1494–1502.

Carter, C., & Carter, C. 1979. Protein crystallization using incomplete factorial experiments. *Journal of Biological Chemistry*, **254**(23), 12219–12223.

Cavicchioli, R. 2011. Archaea - timeline of the third domain. *Nat Rev Micro*, **9**, 51–61.

Chamberlain, J., Lee, Y., Lane, W., & Engelke, D. 1998. Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes & Development*, **12**(11), 1678–1690.

Chong, S., Mersha, F., Comb, D., Scott, M., Landry, D., Vence, L., Perler, F., Benner, J., Kucera, R., Hirvonen, C., Pelletier, J., Paulus, H., & Xu, M. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**(2), 271–281.

Cognat, V., Pawlak, G., Duchene, A.-M., Daujat, M., Gigant, A., Salinas, T., Michaud, M., Gutmann, B., Giege, P., Gobert, A., & Marechal-Drouard, L. 2012. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Research*, **41**(D1), D273–D279.

Cole, J., Lary, J., Moody, T., & Laue, T. 2008. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods in cell biology*, **84**, 143–179.

Crooks, G., Hon, G., Chandonia, J., & Brenner, S. 2004. WebLogo: A Sequence Logo Generator. *Genome Research*, **14**(6), 1188–1190.

Daoud, R., Forget, L., & Lang, F. 2012. Yeast mitochondrial RNase P, RNase Z and the RNA degradosome are part of a stable supercomplex. *Nucleic Acids Research*, **40**(4), 1728–1736.

Deutscher, M. 1984. Processing of tRNA in Prokaryotes and Eukaryote. *Critical Reviews in Biochemistry and Molecular Biology*, **17**(1), 45–71.

Doersen, C., Guerrier-Takada, C., Altman, S., & Attardi, G. 1985. Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. *Journal of Biological Chemistry*, **260**(10), 5942–5949.

Duchêne, A., Pujol, C., & Maréchal-Drouard, L. 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current Genetics*, **55**(1), 1–18.

Dupureur, C. 2008. Roles of metal ions in nucleases. *Biocatalysis and biotransformation/Bioinorganic chemistry*, **12**(2), 250–255.

El Yacoubi, B., Bailly, M., & de Crécy-Lagard, V. 2012. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics*, **46**(1), 69–95.

Esakova, O., & Krasilnikov, A. 2010. Of proteins and RNA: The RNase P/MRP family. *RNA*, **16**(9), 1725 –1747.

Esakova, O., Perederina, A., Quan, C., Schmitt, M. E., & Krasilnikov, A. S. 2008. Footprinting analysis demonstrates extensive similarity between eukaryotic RNase P and RNase MRP holoenzymes. *RNA*, **14**(8), 1558–1567.

Eswar, N., Webb, B., Marti-Renom, M., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U., & Sali, A. 2006. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*, 5–6.

Evans, D., Marquez, S., & Pace, N. 2006. RNase P: interface of the RNA and protein worlds. *Trends in Biochemical Sciences*, **31**(6), 333–341.

Evans, P. 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, **67**(4), 282–292.

Fang, X., Yang, X., Littrell, K., Niranjanakumari, S., Thiyagarajan, P., Fierke, C., Sosnick, T., & Pan, T. 2001. The Bacillus subtilis RNase P holoenzyme contains two RNase P RNA and two RNase P protein subunits. *RNA*, **7**(2), 233–241.

Fleurdepine, S., Deragon, J., Devic, M., Guilleminot, J., & Bousquet-Antonelli, C. 2007. A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. *Nucleic Acids Research*, **35**(10), 3306–3321.

Florentz, C., Kern, D., & Giege, R. 1990. Stimulatory effect of ammonium sulfate at high concentrations on the aminoacylation of tRNA and tRNA-like molecules. *FEBS letters*, **261**(2), 335–338.

García-Ruiz, J. 2003. Counterdiffusion Methods for Macromolecular Crystallization. *Pages 130–154 of:* Charles W. Carter, Jr. and Robert M. Sweet (ed), *Methods in Enzymology*, vol. 368. Academic Press.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M., Appel, R., & Bairoch, A. 2005. Protein identification and analysis tools on the ExPASy server. *Pages 571—-607 of: The proteomics protocols handbook*. Springer.

Geslain, R., & Pan, T. 2011. tRNA: Vast reservoir of RNA molecules with unexpected regulatory function. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(40), 16489–16490.

Ghai, R., Falconer, R., & Collins, B. 2012. Applications of isothermal titration calorimetry in pure and applied research—survey of the literature from 2010. *Journal of Molecular Recognition*, **25**(1), 32–52.

Giegé, R. 2008. Toward a more complete view of tRNA biology. *Nature Structural & Molecular Biology*, **15**(10), 1007–1014.

Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., & Florentz, C. 2012. Structure of transfer RNAs: similarity and variability. *Wiley interdisciplinary reviews. RNA*, **3**(1), 37–61.

Gimble F., Thorner J. 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae. *Nature*, **357**(6376), 301–306.

Gobert, A., Gutmann, B., Taschner, A., Gossringer, M., Holzmann, J., Hartmann, R., Rossmanith, W., & Giege, P. 2010. A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*, **17**(6), 740 –744.

Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., & Giegé, P. 2013. Structural insights into protein-only RNase P complexed with tRNA. *Nature Communications*, **4**, 1353.

Goldfarb, Katherine C., Borah, Sumit, & Cech, Thomas R. 2012. RNase P branches out from RNP to protein: organelle-triggered diversification? *Genes & Development*, **26**(10), 1005–1009.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**(3), 849–857.

Gutmann, B., Gobert, A., & Giege, P. 2012. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes & Development*, **26**(10), 1022–1027.

Hall, T., & Brown, J. 2002. Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins. *RNA*, **8**(3), 296–306.

Hammani, K., & Giegé, P. 2014. RNA metabolism in plant mitochondria. *Trends in Plant Science*.

Harris, M., & Christian, E. 2009. RNA Crosslinking Methods. *Methods in enzymology*, **468**, 127–146.

Hartmann, E., & Hartmann, R. 2003. The enigma of ribonuclease P evolution. *Trends in Genetics*, **19**(10), 561–569.

Hartmann, R., Heinrich, J., Schlegl, J., & Schuster, H. 1995. Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(13), 5822–5826.

Hasegawa, K., Yukawa, Y., Obokata, J., & Sugiura, M. 2003. A tRNALeu-like sequence located immediately upstream of an Arabidopsis clock-regulated gene is transcriptionally active: efficient transcription by an RNA polymerase III-dependent in vitro transcription system. *Gene*, **307**, 133–139.

Hedtke, B., Börner, T., & Weihe, A. 1997. Mitochondrial and Chloroplast Phage-Type RNA Polymerases in Arabidopsis. *Science*, **277**(5327), 809–811.

Hernandez-Cid, Aaron, Aguirre-Sampieri, Sergio, Diaz-Vilchis, Adelaida, & Torres-Larios, Alfredo. 2012. Ribonucleases P/MRP and the Expanding Ribonucleoprotein World. *IUBMB Life*, **64**(6), 521–528.

Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., & Gagliardi, D. 2006. Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase. *Molecular and Cellular Biology*, **26**(7), 2869–2876.

Holzmann, J., Frank, P., Loeffler, E., Bennett, K. L., Gerner, C., & Rossmanith, W. 2008. RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. *Cell*, **135**(3), 462 –474.

Howard, M., Lim, W., Fierke, C., & Koutmos, M. 2012. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(40), 16149–16154.

Inokuchi, H., & Yamao, F. 1995. Structure and Expression of Prokaryotic tRNA Genes. *In:* Söll, Dieter, & RajBhandary, Uttam L. (eds), *tRNA Structure, Biosynthesis, and Function*. Washington, D.C.: AMS Press.

Jarrous, N. 2002. Human ribonuclease P: subunits, function, and intranuclear localization. *RNA*, **8**(1), 1–7.

Jarrous, N., & Gopalan, V. 2010. Archaeal/Eukaryal RNase P: subunits, functions and RNA diversification. *Nucleic Acids Research*, **38**(22), 7885 –7894.

Jarrous, N., & Reiner, R. 2007. Human RNase P: a tRNA-processing enzyme and transcription factor. *Nucleic Acids Research*, **35**(11), 3519–3524.

Kabsch, W. 2010. XDS. *Acta crystallographica. Section D, Biological crystallography*, **66**(2), 125–132.

Kibbe, W. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, **35**, W43–W46.

Kikovska, E., Svärd, S., & Kirsebom, L. 2007. Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proceedings of the National Academy of Sciences*, **104**(7), 2062–2067.

Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., & Nakamura, T. 2012. Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Research*, **40**(6), 2712–2723.

Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K., & Inokuchi, H. 1994. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(20), 9223–9227.

Krehan, M., Heubeck, C., Menzel, N., Seibel, P., & Schön, A. 2012. RNase MRP RNA and RNase P activity in plants are associated with a Pop1p containing complex. *Nucleic Acids Research*, **40**(16), 7956–7966.

L. Stryer, J. Berg, J. Tymoczko. 2007. Biochemistry. *In: Biochemistry*. New York: W. H. Freeman.

Lai, L., Chan, P., Cozen, A., Bernick, D., Brown, J., & Gopalan, V.and Lowe, T. 2010. Discovery of a minimal form of RNase P in Pyrobaculum. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(52), 22493–22498.

Lees, J., Smith, B., Wien, F., Miles, A., & Wallace, B. 2004. CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Analytical biochemistry*, **332**(2), 285–289.

Legen, J., Wanner, G., Herrmann, R., Small, I., & Schmitz-Linneweber, C. 2007. Plastid tRNA Genes trnC-GCA and trnN-GUU are essential for plant cell development. *The Plant Journal*, **51**(5), 751–762.

Li, H., Trotta, C., & Abelson, J. 1998. Crystal Structure and Evolution of a Transfer RNA Splicing Enzyme. *Science*, **280**(5361), 279–284.

Lightowlers, R., & Chrzanowska-Lightowlers, Z. 2013. Human pentatricopeptide proteins: Only a few and what do they do? *RNA Biology*, **10**(9), 1250–1255.

Lipinski, K., Puchta, O., Surendranath, V., Kudla, M., & Golik, P. 2011. Revisiting the Yeast PPR Proteins—Application of an Iterative Hidden Markov Model Algorithm Reveals New Members of the Rapidly Evolving Family. *Molecular Biology and Evolution*, **28**(10), 2935–2948.

Lopez, M., Rosenblad, M., & Samuelsson, T. 2009. Conserved and variable domains of RNase MRP RNA. *RNA Biology*, **6**, 208–220.

Lorber, B., Fischer, F., Bailly, M., Roy, H., & Kern, D. 2012. Protein analysis by dynamic light scattering: methods and techniques for students. *Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology*, **40**(6), 372–382.

Lowe, T., & Chan, P. 2011. *Genomic tRNA Database*. http://lowelab.ucsc.edu/GtRNAdb/. [Online; accessed 2014-03-31].

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Ret, M. Le, Martin-Magniette, M.-L., Mireau, H., Peeters, N., Renou, J.-P., Szurek, B., Taconnat, L., & Small, I. 2004. Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. *Plant Cell*, **16**(8), 2089–2103.

Manavalan, P., & Johnson, W. 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Analytical biochemistry*, **167**(1), 76–85.

Mans, R., Guerrier-Takada, C., Altman, S., & Pleij, C. 1990. Interaction of RNase P from Escherichia coli with pseudoknotted structures in viral RNAs. *Nucleic Acids Research*, **18**(12), 3479–3487.

Matsen, J. 2014. *E. coli genotypes - OpenWetWare*. http://openwetware.org/wiki/E._coli_genotypes#BL21.28DE3.29. [Online; accessed 2014-04-12].

McClain, W., & Lai, L.and Gopalan, V. 2010. Trials, Travails and Triumphs: An Account of RNA Catalysis in RNase P. *Journal of Molecular Biology*, **397**(3), 627–646.

McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., & Read, R. 2007. *Phaser* crystallographic software. *Journal of Applied Crystallography*, **40**(4), 658–674.

Mercer, T., Neph, S., Dinger, M., Crawford, J., Smith, M., Shearwood, A., Haugen, E., Bracken, C.., Rackham, O., Stamatoyannopoulos, J., Filipovska, A., & Mattick, J. 2011. The Human Mitochondrial Transcriptome. *Cell*, **146**(4), 645–658.

Michaud, M., Cognat, V., Duchêne, A., & Maréchal-Drouard, L. 2011. A global picture of tRNA genes in plant genomes. *The Plant Journal*, **66**(1), 80–93.

MiTeGen, LLC. 2014. *MiTeGen*. http://www.mitegen.com/products/microloops/microloops.shtml.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein engineering*, **10**(1), 1–6.

O'Neill, M., & Gaisford, S. 2011. Application and use of isothermal calorimetry in pharmaceutical development. *International Journal of Pharmaceutics*, **417**(1–2), 83–93.

Peck-Miller, K., & Altman, S. 1991. Kinetics of the processing of the precursor to 4·5 S RNA, a naturally occurring substrate for RNase P from Escherichia coli. *Journal of Molecular Biology*, **221**(1), 1–5.

Petoukhov, M., & Svergun, D. 2013. Applications of small-angle X-ray scattering to biomacromolecular solutions. *The International Journal of Biochemistry & Cell Biology*, **45**(2), 429–437.

Petoukhov, M., Franke, D., Shkumatov, A., Tria, G., Kikhney, A., Gajda, M., Gorba, C., Mertens, H., Konarev, P., & Svergun, D.fphizicky. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, **45**(2), 342–350.

Phizicky, E., & Hopper, A. 2010. tRNA biology charges to the front. *Genes & Development*, **24**(17), 1832–1860.

Pinker, F., Brun, M., Morin, P., Deman, A., Chateaux, J., Oliéric, V., Stirnimann, C., Lorber, B., Terrier, N., Ferrigno, R., & Sauter, C. 2013. ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature. *Crystal Growth & Design*, **13**(8), 3333–3340.

Popow, J., Schleiffer, A., & Martinez, J. 2012. Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences*, **69**(16), 2657–2670.

Prikryl, J., Rojas, M., Schuster, G., & Barkan, A. 2011. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(1), 415–420.

Puranam, R., & Attardi, G. 2001. The RNase P Associated with HeLa Cell Mitochondria Contains an Essential RNA Component Identical in Sequence to That of the Nuclear RNase P. *Molecular and Cellular Biology*, **21**(2), 548–561.

Pusnik, M., Small, I., Read, L., Fabbro, T., & Schneider, A. 2007. Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes. *Molecular and cellular biology*, **27**(19), 6876–6888.

Randau, L., Schröder, I., & Söll, D. 2008. Life without RNase P. *Nature*, **453**(7191), 120–123.

Reinhold-Hurek B., Shub D. 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature*, **357**(6374), 173–176.

Reiter, N., Osterman, A., Torres-Larios, A., Swinger, K., Pan, T., & Mondragon, A. 2010. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**(7325), 784 –789.

Rich, A., & RajBhandary, U. 1976. Transfer RNA: Molecular Structure, Sequence, and Properties. *Annual Review of Biochemistry*, **45**(1), 805–860.

Ringel, R., Sologub, M., Morozov, Y., Litonin, D., Cramer, P., & Temiakov, D. 2011. Structure of human mitochondrial RNA polymerase. *Nature*, **478**(7368), 269–273.

Rossmanith, W. 2012. Of P and Z: mitochondrial tRNA processing enzymes. *Biochimica et biophysica acta*, **1819**(9-10), 1017–1026.

Rossmanith, W., & Karwan, R. 1998. Characterization of Human Mitochondrial RNase P: Novel Aspects in tRNA Processing. *Biochemical and Biophysical Research Communications*, **247**(2), 234–241.

Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R., & Sbis, E. 1995. Human Mitochondrial tRNA Processing. *Journal of Biological Chemistry*, **270**(21), 12885–12891.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J., & Moras, D. 1991. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA (Asp). *Science*, **252**(5013), 1682–1689.

Rupp, B. 2009. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 1st edn. New York: Garland Science.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., & Boucher, C. 2002. Genome sequence of the plant pathogen Ralstonia solanacearum. *Nature*, **415**(6871), 497–502.

Schattner, P., Brooks, A., & Lowe, T. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**(2), W686–W689.

Schmitz-Linneweber, C., & Small, I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science*, **13**(12), 663 –670.

Schneider, A., & Maréchal-Drouard, L. 2000. Mitochondrial tRNA import: are there distinct mechanisms? *Trends in Cell Biology*, **10**(12), 509–513.

Schrödinger, LLC. 2010. *The PyMOL Molecular Graphics System, Version 1.3r1*.

Seidel, S., Dijkman, P., Lea, W., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F., Ladbury, J., Schreiber, G., Watts, A., Braun, D., & Duhr, S. 2013. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, **59**(3), 301–315.

Siegel, R., Banta, A., Haas, E., Brown, J., & Pace, N. 1996. Mycoplasma fermentans simplifies our view of the catalytic core of ribonuclease P RNA. *RNA*, **2**(5), 452–462.

Small, I., & Peeters, N. 2000. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences*, **25**(2), 45–47.

Small, Ian. 2003. *Predotar v. 1.03*. https://urgi.versailles.inra.fr/predotar/predotar.html.

Sprague, K . 1995. Transcription of Eukaryotic tRNA Genes. *In:* Söll, Dieter, & RajBhandary, Uttam L. (eds), *tRNA Structure, Biosynthesis, and Function*. Washington, D.C.: AMS Press.

Stark, B., Kole, R., Bowman, E., & Altman, S. 1978. Ribonuclease P: an enzyme with an essential RNA component. *Proceedings of the National Academy of Sciences of the United States of America*, **75**(8), 3717–3721.

Suhre, K., & Sanejouand, Y. 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic acids research*, **32**(July), W610–614.

Svergun, D., & Koch, M. 2003. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, **66**(10), 1735.

Taschner, A., Weber, C., Buzet, A., Hartmann, R., & Hartig, A.and Rossmanith, W. 2012. Nuclear RNase P of Trypanosoma brucei: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Reports*, **2**(1), 19–25.

Thomas, B., Li, X., & Gegenheimer, P. 2000. Chloroplast Ribonuclease P Does Not Utilize the Ribozyme-Type Pre-tRNA Cleavage Mechanism. *RNA*, **6**(4), 545–553.

Thomas, P. 2014. *EcoliWiki*. `http://ecoliwiki.net/colipedia/index.php/Welcome_to_EcoliWiki`. [Online; accessed 2014-04-12].

Thomas, R. 2004. *Practical guide to ICP-MS*. New York, NY: M. Dekker.

Tong, L., & Rossmann, M. 1997. Rotation function calculations with GLRF program. *Pages 594–611 of:* Charles W. Carter, Jr. (ed), *Methods in Enzymology*. Macromolecular Crystallography Part A, vol. Volume 276. Academic Press.

Trapani, S., & Navaza, J. 2008. AMoRe: classical and modern. *Acta Crystallographica Section D: Biological Crystallography*, **64**(1), 11–16.

Travers, A. 1984. Conserved features of coordinately regulated E. coli promoters. *Nucleic Acids Research*, **12**(6), 2605–2618.

Troshin, P., Procter, J., & Barton, G. 2011. Java bioinformatics analysis web services for multiple sequence alignment—JABAWS:MSA. *Bioinformatics*, **27**(14), 2001–2002.

Tsai, H., Masquida, B., Biswas, R., Gopalan, V., & Westhof, E. 2003. Molecular Modeling of the Three-dimensional Structure of the Bacterial RNase P Holoenzyme. *JMB*, **325**(4), 661–675.

Uhlenbeck, O. 1995. Keeping RNA happy. *RNA*, **1**(1), 4–6.

Vilardo, E., Nachbagauer, C., Buzet, A., Taschner, A., Holzmann, J., & Rossmanith, W. 2012. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase–extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic acids research*, **40**(22), 11583–11593.

Vioque, A. 2010. RNase P from Organelles. *Pages 203–222 of:* Liu, Fenyong, & Altman, Sidney (eds), *Ribonuclease P*. New York, NY: Springer New York.

Walker, S., & Engelke, D. 2006. Ribonuclease P: The Evolution of an Ancient RNA Enzyme. *Critical Reviews in Biochemistry and Molecular Biology*, **41**(2), 77 –102.

Wallace, B., & Janes, R. 2010. Synchrotron radiation circular dichroism (SRCD) spectroscopy: an enhanced method for examining protein conformations and protein interactions. *Biochemical Society Transactions*, **38**(4), 861.

Wang, G., Chen, H., Oktay, Y., Zhang, J., Allen, E., Smith, G., Fan, K., Hong, J., French, S., McCaffery, M., Lightowlers, R., Morse III, H., Koehler, C., & Teitell, M. 2010. PNPASE Regulates RNA Import into Mitochondria. *Cell*, **142**(3), 456–467.

Wang, M., Davis, N., & Gegenheimer, P. 1988. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J*, **7**(6), 1567–74.

Waterhouse, A., Procter, J., Martin, D., Clamp, M., & Barton, G. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**(9), 1189–1191.

White, R. 2011. Transcription by RNA polymerase III: more complex than we thought. *Nature Reviews Genetics*, **12**(7), 459–463.

Whitmore, L., & Wallace, B. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**(5), 392–400.

Willkomm, D., & Hartmann, R. 2007. An important piece of the RNase P jigsaw solved. *Trends in Biochemical Sciences*, **32**(6), 247–250.

Winn, M., Ballard, C., Cowtan, K., Dodson, E., Emsley, P., Evans, P., Keegan, R., Krissinel, E., Leslie, A., McCoy, A., McNicholas, S., Murshudov, G., Pannu, N., Potterton, E., Powell, H., Read, R., Vagin, A., & Wilson, K. 2011. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*, **67**(4), 235–242.

Woese, C., & Fox, G. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, **74**(11), 5088–5090.

Wolin, S., & Cedervall, T. 2002. The La Protein. *Annual Review of Biochemistry*, **71**(1), 375–403.

Wolin, S., & Matera, G. 1999. The trials and travels of tRNA. *Genes & Development*, **13**(1), 1 –10.

Xiao, S., Houser-Scott, F., & Engelke, D. 2001. Eukaryotic ribonuclease P: Increased complexity to cope with the nuclear pre-tRNA pathway. *Journal of Cellular Physiology*, **187**(1), 11–20.

Xiao, S., Scott, F., Fierke, C., & Engelke, D. 2002. EUKARYOTIC RIBONUCLEASE P: A Plurality of Ribonucleoprotein Enzymes. *Annual review of biochemistry*, **71**, 165–189.

Xu, Y., Oruganti, S., Gopalan, V., & Foster, M. 2012. Thermodynamics of Coupled Folding in the Interaction of Archaeal RNase P Proteins RPP21 and RPP29. *Biochemistry*, **51**(4), 926–935.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., & Nakamura, T. 2013. Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. *PLoS ONE*, **8**(3), e57286.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H., Wang, J., Zhu, J., Shi, Y., & Yan, N. 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, 168–171.

Yukawa, Y., Sugita, M., Choisne, N., Small, I., & Sugiura, M. 2000. The TATA motif, the CAA motif and the poly(T) transcription termination motif are all important for transcription re-initiation on plant tRNA genes. *The Plant journal: for cell and molecular biology*, **22**(5), 439–447.

Zhang, G., Lukoszek, R., Mueller-Roeber, B., & Ignatova, Z. 2011. Different sequence signatures in the upstream regions of plant and animal tRNA genes shape distinct modes of regulation. *Nucleic Acids Research*, **39**(8), 3331–3339.

Zhou, Xingding, Kini, R Manjunatha, & Sivaraman, J. 2011. Application of isothermal titration calorimetry and column chromatography for identification of biomolecular targets. *Nature Protocols*, **6**(2), 158–165.

Franziska PINKER

Structural characterization of proteinaceous RNase P from *Arabidopsis thaliana*

Résumé en français

La maturation des ARNt en 5' est réalisée par RNase P. C'est un ribozyme chez les bactéries, les fungi et les nulei des mammifères et un enzyme protéique dans les plantes ou des organelles des mammifères qui s'appelle PRORP. Il y a trois PRORP dans A. thaliana. PRORP contiennent deux domaines: un domaine PPR qui reconnaît spécifiquement des séquences d'ARN et un domaine nucléase qui assure la coupure endonucléolytique 5' des précurseurs d'ARNt. Pendant ma thèse j'ai pu montré par des méthodes biophisiques et structurales comme SRCD et SAXS que PRORP1 et 2 sont composées en majorité des helices alpha Elles ont un rayon de giration de 33 Å et contiennent deux domaines distincts avec et une dimension maximale de 110 Å. Pour le complex entre un substrat d'ARNt et PRORP une constante de dissociation de 1 uM a pu être confirmé par la microcalorimétrie, la thermophorèse et l'ultracentrifugation analytique. Ces analyses nous ont permis de construir un modèle PRORP et un substrat d'ARNt.

Mots-clefs : RNase P, protéines PPR, maturation d'ARNt, PRORP, SAXS, Cristallisation

Résumé en anglais

RNase P cleaves 5' leadersof precursor tRNAs. RNase P is a ribozyme in bacteria, fungi and animal nuclei and a protein in animal organelles, plants and many other organism. There are three PRORPs in A. thaliana. MALS, SRCD and SAXS provided first structural information: 1) PRORPs are monomers in solution. 2) PRORP 1-2 have a high alpha-helical content. 3) PRORPs are composed of two distinct domains with a radius of gyration of 33 A. These results together with homology modelling enabled us to build a first model of PRORPs in complex with tRNA. Using three different methods, isothermal titration calorimetry, microscale thermophoresis and analytical ultracentrifugation, a binding constant of about 1 µM could be determined for the system PRORP2mDD and L5T0 tRNA. This helped us conducting a SAXS experiment taking into account the low resolution affinity and designed to provide the direct structural data of a complex of proteinaceous RNase P with a substrate tRNA.

Keywords : RNase P, PPR proteins, tRNA maturation, PRORP, SAXS, Crystallization