Chapter 2
The Evolution of RNase P and Its RNA

J. Christopher Ellis and James W. Brown

2.1 Introduction

Ribonuclease P (RNase P) is a ribonuclease responsible for the 5' maturation of transfer RNA (tRNA). The RNase P holoenzyme is most commonly comprised of a single RNA and one or several associated proteins. It is the RNA, not any protein, that is the catalytic subunit of the enzyme; RNase P is an RNA enzyme. The reaction it carries out is the hydrolysis of a phosphodiester bond in the pre-tRNA, generating two RNA products: the 5'-mature tRNA and the 5' leader RNA fragment. The RNase P RNA subunit is present in all three Domains of life (Bacteria, Archaea, and Eukarya), and in at least some mitochondria and plastids. RNase P RNA, nuclear splicing RNAs (snRNAs), and the ribosomal RNA are the only catalytic RNAs described to date that can conduct multiple catalytic cycles; other “ribozymes” are self-reactive. The presence of RNase P RNA in all branches of living things and its ability perform catalytically have lead to the hypothesis that RNase P is a relic of the RNA world.

RNase P has been studied primarily in terms of the maturation of tRNA, but the holoenzyme has also been shown to cleave the other substrates and play other important roles within the cell. It is also involved in the maturation of a host of other noncoding RNA such as the 2 S, 4.5 S, tmRNAs, and snoRNAs, and in the leader sequences of mRNAs it has also been shown to recognize and cleave riboswitches such as the coenzyme B₁₅ riboswitch in *E. coli* and *Bacillus subtilis* and the adenine riboswitch of the adenine efflux pump transcript in *Bacillus subtilis*. Interestingly neither the coenzyme B₁₅ riboswitch nor the adenine riboswitch contains a predicted structure that is known to be recognized as an RNase P RNA substrate (Hori et al. 2000a; Gimple and Schon 2001; Tous et al. 2001; Altman et al. 2005; Coughlin et al. 2008; Seif and Altman 2008).

Although the sequences of RNase P RNAs are highly variable and recognize many different substrates, a core of the sequence and secondary structure is conserved among examples in all living things (Fig. 2.1). Sequences and sequence length vary...
considerably among RNase P RNAs, especially in eukaryotes, but in Bacteria and Archaea only five major distinct classes of structures have been described so far: “A” (ancestral) type, “B” (named for *Bacillus*, in which it was first discovered) type, “C” (Chloroflexi) type, “M” (Methanococci) type, and “P” (*Pyrobaculum*) type (Figs. 2.2 and 2.3). RNase P RNAs in eukaryotes (including organelles in addition to the nuclear enzymes) are more variable in both sequence and structure, and have not yet been divided into clear structure classes.

Molecular therapeutics is a field of particular current interest, in particular the engineering of RNA molecules for therapeutic use in humans and other animals. An important approach has been the engineering of External Guide Sequences (EGS) that create RNase P substrates in pathogenic RNAs, thereby directing specific destruction of these RNAs by the innate RNase P. This approach has been applied with some success in viral pathogenesis (Reyes-Darias et al. 2008). Another approach

---

**Fig. 2.1** Consensus secondary structure of bacterial, archaeal, and eukaryotic nuclear RNase P RNAs. Elements of structure absent in only single, small evolutionary groups are included: P12 (absent in *Mycoplasma fermentans*), P2 and everything distal to P10/11 (reduced or absent in *Pyrobaculum*). Grey lines indicate connections that vary in structure between conserved elements. Helices are labeled P1-12 according to Haas and Brown (1998), and conserved sequence regions are labeled CR I-V according to Chen and Pace (1997).

---

**Fig. 2.2** Bacterial RNase P RNA structures classes. Representative RNase P RNA secondary structures of bacterial types A, B, and C. Helices are labeled P1-19 according to Haas and Brown (1998). Type A RNAs are found in most Bacteria, and sometimes lack P13 and P14 (in beta-proteobacteria) or P18 (Chlorobi). Type B RNAs are present in most Firmicutes, and in at least some mollicutes lack both P12 and P10.1. Type C RNAs are found in some Chloroflexi; structural variation in this group has not been investigated. The presence of P19 is variable in all types of RNase P RNAs. Many bacterial RNAs also contain additional peripheral helices in P12 or between P15 and P16, or between P16 and P17. The lengths of P3 and P12 are highly variable. Structures are from the RNase P Database (Brown 1999).
has been to engineer the RNase P RNA itself. One example is the recent engineering of an RNase P ribozyme to target the overlapping coding region of two capsid proteins in the murine cytomegalovirus (MCMV) (Bai et al. 2008). The engineered RNase P ribozyme was found to decrease viral growth by 2,000-fold (Bai et al. 2008).

Here we examine our current understanding of the evolutionary variation of the RNase P holoenzyme in the three domains of life (Bacteria, Archaea, and Eukaryotes), especially with respect to evolutionary diversity in RNA structure and protein composition.

2.2 Bacterial Ribonuclease P

In Bacteria, the RNase P holoenzyme is a ribonucleoprotein complex comprised of a single RNA (~400 nt) and a single small protein subunit (~14 kDa). The secondary structure of the RNA was determined by phylogenetic comparative analysis of several hundred sequences over the course of many years (James et al. 1988; Harris et al. 2001). In Bacteria, these secondary structures fall into two structure classes: A-type, the most common, and B-type, found in only in Gram-positive Bacteria (Fig. 2.2) (Brown et al. 1991, 1993, 1996; Haas et al. 1996b). In addition, some of the Chloroflexi (green non-sulfur Bacteria) RNase P RNAs share features intermediate between type A and B RNAs, and these have sometimes been referred to as type C (Haas and Brown 1998).

The tertiary structure of the RNase P RNA is formed by coaxially stacked helical domains. These domains are stabilized and joined together by long range docking interactions and result in a remarkably planar tertiary structure (Kazantsev et al. 2005; Torres-Larios et al. 2005). The most evolutionarily conserved sequences in the RNA are located near the substrate binding surface of the molecule, whereas the most variable sequences (variable in sequence, secondary, and tertiary structure) are located on the periphery of the RNA and are involved in stabilizing the core RNA structure (Kakuta et al. 2005; Kazantsev et al. 2005). Substrate recognition outside of the catalytic site is performed almost entirely by the P7/P8/P9/P10 cruciform and L15 (Christian and Harris 1999). The loop of P8 in the cruciform (L8) interacts with the catalytic center in P4, and this interaction is stabilized by tertiary contacts between P14 and P18 with L8 (Brown et al. 1996; Massire et al. 1998). P8 has been shown to be involved in the recognition of the substrate T-loop, but the details of this interaction are not known (Nolan et al. 1993). The other substrate recognition element L15 is the loop distal of P15; in type A RNAs, this is an internal loop between P15 and P16, in type B RNAs, it is the terminal loop of P15. Although the two helices proximal and distal to L15 are not well conserved in sequence, the sequence GGUU immediately preceding the 3' strand of P15 is conserved in most bacteria. This sequence is directly involved in substrate recognition by basepairing to the 3'-NCCA tail of the pre-tRNA (Kirsebom and Svärd 1994). Some bacterial RNase P RNAs (notably those of some cyanobacteria and Chlamydiae) lack this conserved sequence and yet are functional in the absence of protein; recognition of the 3'-NCCA tail of pre-tRNA by these RNAs is uncharacterized.
2.2.1 Bacterial RNase P RNA Structure Classes

A-type RNase P RNAs are the most common form of the RNA in Bacteria (and Archaea) and Escherichia coli is the model for this structure class. The A-type RNase P RNA molecule have several RNA elements not found in other RNase P RNA types: P6, P13, P14, P16, and P17 (Haas and Brown 1998). Although these elements vary in sequence and in length and are not found in all RNase P RNAs, they play important roles in the tertiary structure of the RNA itself by creating long-range interactions that stabilize the functional core of the RNA. P16 and P17 are an extension of L15, and the loop of P17 basepairs with an asymmetric bulge between P5 and P7 to create helix P6 and a pseudoknot in the global RNA structure. Another pseudoknot is found in all known RNase P RNAs, created by P2 and P4, and some contain yet another pseudoknot created by the basepairing of P8 with the 3'-tail of the RNA. P13 and P14 stack coaxially; the loop of P14 forms a tertiary interaction with P8, whereas P13 interacts with the base of P12; the details of this interaction remain unclear (Brown et al. 1996).

B-type RNase P RNAs differ dramatically from A-type, but not in regions involved directly in substrate recognition (P8 and L15) or in the catalytic core (primarily P4 and surrounding joining regions). B-type RNase P RNAs are found in most Firmicutes (low G+C Gram-positive Bacteria) and the RNase P RNA of Bacillus subtilis is the model example. Type B RNAs lack both the P16/P17/P6 and P13/P14 structures, but these seem to be replaced by type B-specific elements P5.1/P15.1 and P10.1 (Haas et al. 1996b). The sequences and structures shared by both type A and type B RNase P RNAs define a "core" present in the RNase P RNAs that are catalytically proficient in vitro in the absence of protein. That this core contains all of the sequences and structures required for catalytic function has been confirmed by the characterization of "mini" and "micro" RNase P RNAs consisting only of this core (Waugh et al. 1989; Siegel et al. 1996). Although catalytically active, these core RNAs require substantial external stabilization; the implication of this is that although the core contains all of the elements of sequence and structure required for the substrate recognition and catalysis, the phylogenetically variable regions contribute substantially to the stabilization of this functional core (Kazantsev et al. 2005; Torres-Larios et al. 2005).

The evolutionary transition between type A and type B RNAs appears to have happened abruptly, with no evolutionary intermediates or alternative descendents described to date (Haas et al. 1996b). However, the RNase P RNAs found in the phylogenetically distant Chloroflexi (green non-sulfur Bacteria) are intermediate between types A and B. This seems to be an example of convergent evolution, rather than horizontal transfer, because although the secondary structures of these RNAs are related, their sequences are not specifically similar.

2.2.2 Dimerization Mediated by the RNA

The RNase P holoenzyme in Bacteria is a heterodimer of one molecule each of the RNA and protein subunits. The holoenzyme heterodimer of both E. coli (type A) and Bacillus subtilis (type B) can in turn dimerize in solution (Buck et al. 2005a). Buck et al. demonstrated that although both enzymes could form these dimers, the E. coli holoenzyme forms a heterogeneous mixture of holoenzyme monomers and dimers, and this mixture shifts almost entirely to monomers in the presence of mature tRNA (Buck et al. 2005a). However, the B. subtilis holoenzyme exists almost exclusively as dimers, and does not shift to monomers in the presence of mature tRNA (Buck et al. 2005a). Buck et al. argue convincingly that it is the attributes of the RNA, not the protein, that are responsible for dimer formation in both E. coli and B. subtilis. Therefore, it is presumably the additional RNA elements (i.e., P5.1, P10.1, P15.1, P15.2) in type B RNase P RNA that are associated with stable dimer formation in B-type RNAs, and the absence of these RNA elements in A-type RNase P RNA that disfavor dimer complex formation in the presence of substrate. The biological relevance (and even existence) of these dimers in vivo is unknown and, perhaps doubtful.

2.2.3 The Role of the Protein Subunit

With so much variation at the level of the ancient catalytic RNA, one might anticipate that the associated protein would be highly variable in sequence and structure, and that perhaps the proteins that associate with type A and B RNAs might be readily distinguishable. This is not the case. The protein is highly conserved not only in the organisms encoding the ancestral A-types RNA but also with the organisms encoding the type B RNase P RNAs. It is approximately 14 kDa and shares an unusual left-handed crossover connection and a large central cleft (Stams et al. 1998). This motif is shared by ribosomal protein S5 and ribosomal translocase elongation factor, which suggests these proteins evolved from a common ancestor of the ancient translational machinery (Stams et al. 1998).

Although the protein sequence and structure is highly conserved in Bacteria, the functional impact of the protein differs somewhat depending on the type of RNase P RNA to which it binds. In B. subtilis, the protein increases the holoenzymes substrate specificity (Crarly et al. 1998; Niranjanakumari 1998), whereas the E. coli protein stabilizes the tertiary structure of the corresponding RNA (Guerrier-Takada et al. 1983; Westhof et al. 1996; Buck et al. 2005a). Furthermore, in E. coli the protein decreases the Mg⁶⁺ dependence of the RNA to fold into the native state from the intermediate state but the protein in B. subtilis does not change the RNase P RNA's Mg⁶⁺ dependence (Loria and Pan 1996; Buck et al. 2005a). Despite some functional differences that are dependent on RNase P RNA type it is clear that the protein plays an important role in both types of RNAs in substrate recognition and tertiary structure of the RNA, and the significance of these biochemical differences in vitro is unclear, given that these proteins have been shown to be interchangeable in vivo wherever tested (Waugh and Pace 1990; Gössinger and Hartmann 2007).
2.2.4 **RnpA is Part of a Conserved Genomic Arrangement**

In bacteria, the gene that encodes the RNase P RNA protein subunit (rnpA) is located immediately downstream and in the same orientation as rpmH, which encodes for the ribosomal protein L34. This gene arrangement and their location near the origin of replication are highly conserved in bacteria with only a couple of known exceptions. The two genes in *E. coli* have been shown to be part of an operon with one minor and two major promoters upstream of rpmH and two transcriptional termination signals downstream of rpmH (upstream of rnpA) (Hansen et al. 1982, 1985; Panagiotidou et al. 1992). Most of the transcripts from the operon do not include the rnpA coding region and therefore the proteins are differentially expressed. Ribosomal protein L34 is abundant in cells but the RNase P protein is expressed at much lower levels, in part to the transcriptional disparity but also because the rnpA transcript is a poor match to the codon bias of *E. coli*, decreasing the translation efficiency further.

A handful of variations of this genetic arrangement have been described, that of *Aquifex aeolicus* which we will be discussed below, and that of species of the genus *Thermus*. *Thermus* species are thermophilic, and their RNase P enzymes are composed of the usual type A RNA and rnpA protein. However, in all species of *Thermus* examined, rnpA and rpmH are completely overlapping (Fig. 2.4). The start codons of two genes are separated by only four base pairs and share the same orientation (Feltens et al. 2003). The first start codon initiates the rnpA coding sequence in the −1 register relative to rpmH and the second start codon initiates the coding sequence for rpmH (Feltens et al. 2003). This results in an rnpA protein that is substantially longer protein (163 amino acids) than the usual ~120 amino acids. In the lactic acid bacterium *Lactococcus lactis subsp. cremoris*, the rnpA gene is preceded by an ORF (LACR0128) rather than rpmH, and an rpmH homolog has not be identified in the genome (Wegmann et al. 2007).

2.2.5 **And Then There was One: Aquifex and the Missing Link**

*Aquifex aeolicus* is a deep branching hyperthermophilic bacterium. No candidate RNase P RNA- or protein-encoding genes have been identified in the genome of this species, despite the fact that tRNA gene organization in this genome implies the need for this enzyme; tRNA genes are (as usual in Bacteria) clustered and presumably co-transcribed, and some are located within the rRNA operon (Li and Altman 2004). Biochemical investigation of this surprising finding has been hampered by the fact that this organism is very difficult to cultivate, however some progress has been made. Substitution of Mg²⁺ with other divalent cations such as Co²⁺ completely inhibited the RNase P activity in extracts, consistent with what is seen in other bacteria (Kazakov and Altman 1991; Lombo and Kamberdin 2008; Marszałkowski et al. 2008). Furthermore, depletion of RNA from the cell extracts of *A. aeolicus* eliminates RNase P activity. Both of these are indirect evidence that an RNA is required for RNase P activity, but if this is the case, this RNA must not resemble any previously described RNase P RNA (Lombo and Kamberdin 2008).

Although no RNase P RNA gene has been found in the *A. aeolicus* genome, what about the conserved gene arrangement typically associated with rnpA? In relatives of *A. aeolicus* such as *Sulfuricurvibacterium azorense* and *Pseudosarcina marina* (which contain more-or-less typical bacterial type A RNase P RNAs lacking P18), the usual bacterial arrangement and orientation of rpmH → rnpA is present (Ogasawara et al. 1985; Salazar et al. 1996). However, in *A. aeolicus*, rnpA is not found downstream of rpmH, nor anywhere else in the genome (Li and Altman 2004; Marszałkowski et al. 2008). How, then, is *A. aeolicus* processing pre-tRNAs in the apparent absence of an rnpA protein, which is essential in vitro in other bacteria? Is it possible that the novel variant of an RNase P RNA can process pre-tRNAs in the absence of the usually associated rnpA protein? Or has this species invented an entirely new RNase P enzyme of some type?

2.3 **Archaeal Ribonuclease P**

Like Bacteria, Archaea are single cell non-eukaryotic microorganisms. Archaea are divided into two major phyla: Crenarchaeae and Euryarchaeae. Although Archaea are distinguishable from both eukaryotes and bacteria, they are in many ways like eukaryotes, and in many other ways like bacteria. The RNase P holoenzyme in Archaea is an excellent evolutionary example of this. Most archaeal RNase P RNAs are of the
type A found also in bacteria (Fig. 2.3), and at sufficiently high salt concentrations (4 M ammonium acetate and 300 mM MgCl₂) the RNase P RNAs of some of many of these Archaea are catalytically active in the absence of protein in vitro, and the activity of these RNAs is enhanced at more moderate ionic strength by the addition of the Bacillus subtilis (a bacterium) RNase P protein (Pannucci et al. 1999). However, the archaean RNase P holoenzyme generally contains four proteins, none of which are clearly homologous to the bacterial rpmP RNase P protein, but these are clearly homologous to the core RNase P proteins of the eukaryotic nucleus (Hall and Brown 2004; Hartmann and Hartmann 2003). Because there is not any clear-cut homology between the bacterial RNase P protein on one hand and the archaean and eukaryotic nuclear proteins on the other, it may be that the RNA subunit of RNase P evolved first, perhaps as part of the RNA World, and the proteins evolved independently following the divergence of the bacterial and eukaryotic/archaean lineages.

2.3.1 Archaeal RNase P RNA Structure Classes

As in bacteria, most archaean RNase P RNAs are A-type RNAs, largely resembling bacterial RNase P RNAs in terms of conserved sequences and structures. Archael type A RNase P RNAs do differ from those of most bacteria in the absence of P13/P14 (also absent in the type A RNAs of beta-proteobacteria) and P18 (also absent in the type A RNAs of Chlorobi and Aquificae). However, despite this structural similarity between the bacterial and archaean RNase P RNA in general, the first available archaean RNase P RNA sequences were not as bacterial-like as others, and could not be shown to be catalytically active in the absence of protein. It therefore became a thought that, like eukaryotic RNase P RNAs, the archaean RNase P RNA was absolutely dependent on associated protein for catalytic activity. However, it was subsequently shown that many archaean RNase P RNAs are catalytically active in the absence of protein, but require extreme ionic conditions (4 M ammonium acetate, 300 mM MgCl₂) for this to be expressed (Pannucci et al. 1999). As in Bacteria, the protein subunits in the archaean enzymes seem not to contribute directly toward catalysis, but at least predominantly toward the stabilization of the superstructure of the RNA subunit.

One of the subsets of archaean RNase P RNAs that does not display catalytic activity even under high ionic conditions are those with a different structure class, the M-type RNase P RNAs (Fig. 2.3) To date, only five archaean genera (Archaeoglobus, Methanothermus, Methanocorpus, and Methanococcus) are known to have M-type RNase P RNAs; this suggests that Archaeoglobus may be related to the Methanocorpus, despite uncertainty in the placement of this genus in phylogenetic trees based on RNA sequences (Brown 1999). M-type RNAs are essentially similar to archaean A-type RNAs but lack two RNA structural elements that are essential (at least in Bacteria) for substrate recognition: P8 and L15 (along with P16/P17/P6) (Fig. 2.3). P8 is part of the highly conserved cruciform consisting
a β-barrel structure (Numata et al. 2004). The protruding β-strand (β7) forms a β-sheet with β4 which may be involved in protein–protein interaction with the other RNase P subunits (Numata et al. 2004). Two possible RNA binding sites were identified in Ph1771p. The first is a loop region connecting strands β2 and β3 which is composed of hydrophilic residues exposed to solvent and the other is composed of α-helices 1-4 and β-strand β6 forming a cluster of positively charged amino acids (Numata et al. 2004). Ph177p shares structural similarity to the other RNA binding proteins: Staphylococcus aureus translational regulator Hfq and Haloarcula marismortui ribosomal protein L21E (Numata et al. 2004).

RNase P protein Ph1877p is a homolog of human RNase P protein Rpp30 and is composed of ten α-helices and seven β-strands forming a TIM barrel (Takagi et al. 2004; Kawano et al. 2006). Ph1877p interacts with Ph1481p in the RNase P holoenzyme, which is consistent with the interactions determined in other archaeal RNase P protein–protein interactions (Hall and Brown 2004; Kifusa et al. 2005). The structure of the Ph1481p-Ph1877p dimer has been determined in co-crystals.

RNase P protein Ph1601p is a homolog to human RNase P protein Rpp21 and is composed of an N-terminal domain of two α-helices, and the central and C-terminal regions form a zinc ribonuclease domain, giving the protein an L-shape (Kakuta et al. 2005). Several clusters of positively charged amino acids along one face of the L-arms, suggest an RNA binding role, whereas four Cys residues bind a zinc molecule stabilizing the N-terminus and C-terminus domains (Kakuta et al. 2005).

RNase P protein Ph1481p is a homolog to human RNase P protein Pop5 and is composed of five antiparallel β-sheets and five α-helices forming a α/β globular protein (Kawano et al. 2006). Whether or not this structure is meaningfully similar to the bacterial RNase P protein structures remains a matter of disagreement (Stams et al. 1998; Spitzfaden et al. 2000; Kazantsev et al. 2003; Takagi et al. 2004), but it is generally believed that any such similarity is most likely the result of degenerate rather than homology. Certainly, the structure of Ph1481 is similar to the generic ribonuclease protein (RNP) domains found in a number of RNA binding proteins, suggesting a role in the binding of the protein to RNase P RNA or pre-tRNA, Ph1877p and Ph1481p form a heterotrimer in solution, with a homodimer of Ph1481p in the center of two monomers of Ph1877p (Kawano et al. 2006). Whether this heterotrimer is associated with one or two molecules of the RNA subunit of RNase P has not been determined.

2.3.2.1 Is L7Ac a Fifth RNase P Protein in Pyrococcus horikoshii OT?

L7Ac protein was first described as a protein associated with the small ribosomal subunit. However, the crystal structure of the complete large ribosomal subunit (50 S) with the associated proteins revealed L7Ac was actually associated with the large subunit (Ban et al. 2000). More recently, L7Ac was found to be a component of C/D and H/ACA small nucleolar RNAs (sRNA) (Kuhn et al. 2002; Rozhdestvensky et al. 2003). L7Ac’s association with sRNAs is at kink-turns (k-turns), found in both ribosomal RNA and sRNAs (Rozhdestvensky et al. 2003).
pre-tRNAs with 1-3 nt leaders, which would require an RNase P activity or a suitable replacement. An unusual RNase P-like RNA has recently been identified in the genomes of several species of the genus *Pyrobaculum*, and this RNA is associated with RNase P activity. This RNA consists of only the catalytic domain of RNase P RNA (Fig. 2.3). No genes encoding proteins specifically related to the RNase P proteins of any kind have been identified in the genomes of *Pyrobaculum* species.

2.4 Eukaryotic Ribonuclease P

RNase P in eukaryotes is much more complex than in either Bacteria or Archaea. Eukaryotes contain a nuclear RNase P, that processes pre-tRNAs in the nucleolus for cytoplasmic translation, a mitochondrial RNase P for processing mitochondrially encoded tRNAs, and in plants and algae a plastid RNase P for plastid-encoded tRNAs. These enzymes are entirely distinct from each other. In addition, the nuclear RNase P is a member of an enzyme family with RNase MRP, which is involved in the rRNA processing also in the nucleolus.

Eukaryotic nuclear RNase P enzymes are more diverse in terms of RNA structure and protein composition than are those of Bacteria or Archaea, and so are less well understood. However, the secondary structures of RNase P RNAs in some phylogenetic groups have been determined in some detail. The structure of the core of the nuclear RNA is common to all and contains many of the core elements of the bacteria and archaeal RNAs. The protein composition of the nuclear enzyme is only well known in the human and yeast systems, in which the enzyme is composed of at least 9 proteins, some quite large. The nuclear RNase P RNA is not catalytically proficient in the absence of protein.

Organellar RNase P enzymes are even less well understood than are those of the nucleus. In some primitive mitochondria and plastids, the RNA subunits of RNase P (Fig. 2.5) are clearly related to those of the proteobacteria and cyanobacteria from which these organelles arose, although the proteins are not. In humans, and presumably metazoans in general, the traditional RNase P has apparently been dispensable with entirely, and replaced by a "Rube Goldberg" amalgamation of proteins that carry out the same function (Holzmann et al. 2008; Walker and Engelke 2008). The nature of chloroplast RNase P in green plants is unclear; there is evidence that this RNase P, like the protein only "RNase P" found in human mitochondria, might also be an all-protein enzyme, but the components of this enzyme have yet to be identified.

2.4.1 *Saccharomyces cerevisiae* a Model for RNase P in Eukaryotes

The best understood eukaryotic RNase P system is that of the yeast *Saccharomyces cerevisiae*. Much is also known about the human system, but our insight is clearer

---

**Fig. 2.5** Primitive organellar RNase P RNAs. The sequences and structures of RNase P RNAs encoded in the primitive mitochondria of the jacobid *R. americana* and the primitive plastid of the rhodophyte *P. purpurea* reflect their bacterial ancestry. Helices are labeled P1-P17 according to Haas and Brown (1998). The structures RNase P RNAs in other, less primitive mitochondria and plastids are more divergent and less well characterized. RNase P in animal mitochondria and perhaps green plant chloroplasts lack an RNA component (see text). Structures are from the RNase P Database (Brown 1999)

---

in the yeast system, and what is clear in the human system is, not surprisingly, generally consistent with what is also found in yeast (at least in the case of the nuclear enzyme). Most other eukaryotic systems are known primarily from the sequences of the RNA and protein subunits identified by their similarity to those of yeast or humans. Here we will examine the nuclear and mitochondrial RNase P RNA holoenzymes and a closely related ribonucleaseoprotein complex, the RNase MRP holoenzyme.

2.4.1 Nuclear RNase P

The secondary structure of the nuclear RNase P in *S. cerevisiae* is consistent with much of the phylogenetically conserved core of the bacterial and archaeal enzymes, including P1, P2, and P3, P4, P7, P10/11, P12, P15, and conserved regions I-V
2 The Evolution of RNase P and Its RNA

this has not been used consistently. These elements in the RNase P RNA of *S. cerevisiae* include eP8, eP9, eP15, and eP19. P3 in most Eukaryotes, but not in bacteria or Archaea, includes a large internal loop that is the binding site for the Pop6/7 protein heterodimer (Perederina et al. 2007). RNase MRP RNA shares both this loop and these proteins. The binding of Pop6/7 to P3 in the RNA may mediate the binding of other RNase P and RNase MRP protein subunits (Perederina et al. 2007). Nuclear RNase P RNAs notably lack L15, conserved in most bacterial and archaeal RNase P RNAs and the binding site for the 3'-NCCA tail of the pre-rRNA substrate. In addition, the P7/P8/P9/P10 cruciform found in most archaeal and bacterial RNase P RNAs is unusual in *S. cerevisiae* because of an extra RNA hairpin proximal of P8. However, other nuclear eukaryotic RNase P RNAs such as those found in *Homo sapiens*, *Schizosaccharomyces pombe*, and *Pichia strastburgensis* maintain the "standard" cruciform structure but lack the obvious sequence similarity to their homologs in Bacteria or Archaea (Frank et al. 2000).

The *S. cerevisiae* nuclear RNase P has no fewer than nine proteins associated with the holoenzyme complex (POP1, POP3, POP4, POP5, POP6, POP7, POP8, RPP1, and RPR2) (Fig. 2.7). Only POP4, POP1, and the heterodimer POP6/7 appear to bind directly to the RNase P RNA (Houser-Scott et al. 2002; Srisawat et al. 2002; Perederina et al. 2007). Four of these proteins (POP4, RPR2, RPP1, and POP5) are homologous to the RNase P protein subunits found in Archaea (Ph1771p, Ph1601p, Ph1877p, Ph1481p in *P. horikoshii*, respectively). However, no similarity can be identified among any of the nuclear RNase P proteins subunits and the bacterial
\textit{mpA} protein. The specific roles of the nuclear proteins are not well-defined, but although these proteins are essential for catalysis, the catalytic center of the enzyme resides in the RNA rather than protein.

### 2.4.1.2 Mitochondrial RNase P

The RNase P of \textit{S. cerevisiae} mitochondrion is composed of an RNA subunit (Rpm1p) that is 490 nucleotides long and a single 105 kDa protein subunit (Rpm2p). The yeast mitochondrial RNase P RNA is so A+T-rich that the structure cannot be resolved by comparative analysis, but seems perhaps to contain only the core elements of the catalytic domain (Wise and Martin 1991). The Rpm2 protein is encoded in the nucleus and transported to the mitochondria and shares no obvious similarity to the RNase P protein subunit of bacteria or the nucleus (Morales et al. 1992). The protein is essential for catalytic activity (Morales et al. 1992). Surprisingly, the protein component (Rpm2p) is responsible for a host of functions in the nucleus and cytoplasm of the cell unrelated to the RNase P. For example, the C-terminal domain of Rpm2p is directly involved in the maturation of the RNase P RNA but even without this region of the protein the mitochondrial RNase P holoenzyme retains activity (Stribinskis et al. 2001b). Furthermore, Rpm2p is a transcriptional activator in the nucleus and interacts with the cytoplasmic processing bodies (P-bodies) along with Dcp2p (Stribinskis and Ramos 2007). Rpm2p activates transcription of many nuclear-encoded mitochondrial proteins, such as those involved in the import apparatus (TOM40, TOM6, TOM20, TOM22, and TOM37), and mitochondrial chaperons (HSP60 and HSP10) (Stribinskis et al. 2005). These multifunctional roles within the nucleus/cytoplasm and the lack of any conserved similarity with bacterial RNase P protein suggest that this was a pre-existing protein recruited to mitochondrial RNase P, perhaps to replace the original bacterial RNase P protein.

### 2.4.1.3 RNase MRP

RNase “mitochondrial RNA processing” (MRP) is also an endoribonuclease, but is found only in eukaryotes. RNase MRP in \textit{S. cerevisiae} has composed a single RNA molecule and at least 10 protein subunits (Pop1p, Pop3p, Pop4p, Pop5p, Pop6p, Pop7p, Pop8p, Rmp1p, Rmp1p, and Smn1p) (Fig. 2.8). Eight of these protein subunits are also subunits of RNase P. Although RNase MRP lacks the Rpr2p of RNase P, it contains two proteins unique to RNase MRP: Smn1p and Rmp1p. The RNA component of RNase MRP’s catalytic domain (Domain 1) shares obvious similarity with the nuclear RNase P RNA in both sequence and secondary structure (Forster and Altman 1990; Li et al. 2002). These similarities in associated proteins and conserved RNA sequences/structures imply that the RNase MRP evolved from RNase P shortly after the three domains (Bacteria, Archaea, and Eukaryotes) diverged.

\textbf{Fig. 2.8} \textit{S. cerevisiae} RNase MRP secondary structure, protein:protein and protein:RNA interactions (Walker et al. 2005). Helices with clear homologs in RNase P RNA from the same organism are labeled P1–P19. Direct RNA-binding proteins are shaded grey. Pop1p binds to the P3 loop (between P3a and P3b). The Pop6p/Pop7b heterodimer also binds P3. The location of Pop4p binding to the RNA is not known. Smn1p and Rmp1p, the only RNase MRP proteins not shared with RNase P, are shaded black.

RNase MRP was originally described as an endoribonuclease that cleaves primers for mitochondrial DNA replication (Chang and Clayton 1987). Later, RNase MRP was shown to be predominantly in the nucleus rather than the mitochondrion, and found to play a role in the processing of the 5.8 S ribosomal RNA precursor. It does this by cleaving the pre-rRNA at site A3 in the first internal transcribed spacer (ITS1). RNase MRP also cleaves the mRNA of \textit{CLB2} in its 5’-UTR (Gill et al. 2004). The cleavage of the 5’-UTR initiates the degradation of the transcript by Xrn1 nuclease. Cells with a defective RNase MRP suffer as a result from a late anaphase delay (Gill et al. 2004). Although both the RNA and protein subunits of RNase MRP are essential, neither the cleavage of pre-rRNA nor the cleavage of mRNA primers associated with DNA replication in the mitochondria are essential for cell viability. Therefore, it seems likely that RNase MRP has other as yet unknown essential function(s).

### 2.5 Conclusion

Since the RNA component of RNase P is conserved among the three Domains, but the proteins are not, it is likely that the last common ancestor of these three Domains contained an RNA-only RNase P, and the associated proteins evolved later
after the evolutionary divergence of bacteria on one hand and Archaea/eukaryotes on the other. The A type RNase P RNA common to most bacteria and Archaea gave rise to the other forms: type B in bacteria (and the organellar RNAs), type M and P in Archaea, and the various forms of the RNA (including that of RNase MRP), in the nucleus of eukaryotes. In the common ancestor of bacteria, the enzyme acquired the mpaA protein, which remained essentially unchanged throughout this Domain. The common ancestor of Archaea and eukaryotes acquired a set of 4 RNase P proteins that remain (perhaps with some additions) in Archaea. Those four proteins along with the RNA serve as the core of the RNases P and MRP of the eukaryotic nucleus, to which many other proteins have been added. Early in eukaryotic evolution (exactly where is unclear, but after the acquisition of most proteins), the RNase P RNA gene was duplicated, and one copy became specialized (including by some protein substitutions) for rRNA processing (at least) whereas the other retained the traditional RNase P function. In at least one archaeon, the need for RNase P seems to have been removed and the enzyme lost. In mitochondria and plastids generally, the RNA remains from their bacterial ancestry, but the mpaA protein has been replaced by other, nuclear-encoded proteins. In the animal mitochondrion, and perhaps green chloroplasts, RNase P has been replaced wholesale by an independent protein-only enzyme.

Acknowledgments This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

 Proc Natl Acad Sci U S A 102(32):12,864–12,889

2 The Evolution of RNase P and its RNA

2 The Evolution of RNA P and its RNA

Chapter 3
Over a Decade of Bacterial Ribonuclease P Modeling

Benoit Masquida, Fabrice Jossinet, and Eric Westhof

3.1 Introduction

The prokaryotic ribonuclease P maturates precursor tRNAs (ptRNAs) by catalytically removing the 5' leader sequence. The RNA subunit from this RNP carries out the catalytic activity. Hence, the RNase P is a ribozyme. The RNP builds up from a ~300-nt RNA and a ~100-aa protein. Two different scaffolds of the bacterial RNase P can be found. They consist of a common catalytic core surrounded by peripheral domains variable in size and structure. The A type (A stems for ancestor) is characteristic of Escherichia coli and cyanobacteria. The B type (B stems for Bacillus) is characteristic of Bacillus subtilis. tRNAs are substrates to many different biological processes including maturation, chemical modification, aminoclaylation, protein translation and more that can be specific of a family of organisms (Sce (Gige 2008) for review), e.g., tRNA-dependent cell wall synthesis in firmicutes (Matsushashi et al. 1965). This implies that selection pressures specific to all the pathways in which tRNAs intervene are simultaneously applied on the same molecular scaffold in a way characteristic to each organism. Conversely, it means that molecules acting on (p)ptRNAs indirectly undergo selection pressures connected to all the pathways involving tRNAs. It is therefore not too surprising that the set of RNase P natural substrates differs slightly among bacteria (Altman et al. 1993; Hartmann et al. 1995; Peck-Miller and Altman 1991) and that RNase P RNAs have different secondary and tertiary structures.

From the molecular modeling point of view, one should thus distinguish the architectural features of the RNase P holoenzyme and substrates that are directly linked to the catalytic reaction from those linked to the other tRNA-processing pathways. Among the tRNA pathways, some are specific to a given tRNA or to a tRNA family (aminoclaylation), and some use the whole pool of different tRNAs (tRNA processing, translation). For example, the anticodon loop is not recognized