

Structure-function analysis in nuclear RNase P RNA

David R. Engelke, Eileen Pagán-Ramos & Anthony J. Tranguch

Department of Biological Chemistry, The University of Michigan, Ann Arbor, MI 48109–0606, USA

Key words: nuclear, RNase P, RNA structure, yeast

Abstract

Eukaryotic ribonuclease P (RNase P) enzymes require both RNA and protein subunits for activity *in vivo* and *in vitro*. We have undertaken an analysis of the complex RNA subunit of the nuclear holoenzyme in an effort to understand its structure and its similarities to and differences from the bacterial ribozymes. Phylogenetic analysis, structure-sensitive RNA footprinting, and directed mutagenesis reveal conserved secondary and tertiary structures with both strong similarities to the bacterial consensus and distinctive features. The effects of mutations in the most highly conserved positions are being used to dissect the functions of individual subdomains.

Abbreviations: RPR1 – ribonuclease P ribonucleoprotein 1 gene from *Saccharomyces cerevisiae*; Pu – purine ribonucleoside.

Eukaryotic vs. prokaryotic RNase P

RNase P is a ribonucleoprotein enzyme present in all living organisms that cleaves precursor tRNA molecules, generating mature 5' termini [1]. The similarity between enzymes from various sources is underscored by the fact that each recognizes pre-tRNAs from a wide variety of sources across kingdom lines. There are also distinct differences, however. The bacterial RNase P RNA subunits retain catalytic activity *in vitro* in the absence of the small protein subunit [2], although the protein is required for function *in vivo*. In contrast, the eukaryotic enzymes have a significantly higher protein content [3–8] and there are no reports of the RNA subunits alone being catalytic. Thus, RNase P represents an interesting problem in macromolecular evolution - the modern coexistence of the bacterial enzyme that is more in the 'RNA World', in that the RNA can independently bind and cleave substrates, with the eukaryotic enzymes, in which more functions depend on protein. Although RNase P activities have been identified in a wide range of eukaryotes and a structure has been proposed for the RNA subunit of the vertebrate enzyme [9], the work described here focuses on obtaining a high-resolution structure for the yeast

nuclear RNase P and begins to explore the function of subdomains.

Yeast nuclear RNase P

Isolation of nuclear RNase P holoenzymes from both *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* allowed identification of the RNA subunits and the genes encoding them [10–12]. Both enzymes appear to have at least one protein subunit, with a mass in excess of 100 kDa [6, 7], similar to the yeast mitochondrial enzyme [5], although it is not yet known whether the *S. cerevisiae* enzyme contains additional protein subunits.

The RNase P RNA gene in *S. cerevisiae* was shown to be single-copy, essential, and transcribed by RNA polymerase III *in vivo* [12, 13]. The 369 nucleotide mature RNA is embedded in an unusual transcription unit containing a 'disposable' internal promoter for RNA polymerase III. The primary transcript has an 84 nucleotide leader that contains a tRNA-like internal promoter with 'A block' and 'B block' elements. The DNA region upstream of the transcription initiation site has several sequence elements in common with the

U6 small nuclear RNA gene, but only influences transcriptional efficiency rather than being essential. The primary transcript also has up to 28 extra 3' nucleotides before the oligo-U terminator. The leader is removed in a single step after assembly of the RNP enzyme (J. Lee, J. Chamberlain, and D. Engelke, unpubl.), whereas the 3' terminus is matured through a series of intermediates.

RNA structure analyses

Phylogenetic analysis of yeast nuclear RNase P RNA was used to obtain a secondary structure model of the RNA within the holoenzyme [14]. Six additional *Saccharomyces* species were identified with an ideal spread of sequence divergence (62% to 92% identity). When combined with sequence data for *Schizosaccharomyces* [15], a well-supported consensus secondary structure emerged that conformed to all the yeast data, and was also consistent with several aspects of the sequences emerging for vertebrates. This structure, after refinement by nuclease and chemical footprinting [16, see below], is shown in Figure 1. After assignment of the secondary structure it was possible to identify positions that are conserved among yeast or that have analogues across kingdom lines in the bacterial consensus structure [17]. In addition, substitution of RNase P RNA coding sequences from heterologous yeast and other organisms allowed us to ask which poorly conserved regions were likely to be required specifically for functioning in *S. cerevisiae* [18].

Structure-sensitive enzymatic and chemical RNA footprinting of the *S. cerevisiae* holoenzyme and deprotonized RNA were used to refine the phylogenetic model. The results strongly supported most aspects of the phylogenetic structure, although minor modifications were made to accommodate both types of data. In addition, the results indicate which regions of the RNA protrude from the holoenzyme and are solvent-accessible [16]. The most exposed regions, those accessible to nuclease attack, are indicated by lack of shading in Figure 1. In general, these exposed regions correlated to those found to be non-essential in the substitution experiments [18]. It is not possible on the basis of this data to distinguish between positions that are protected by direct interaction with the protein component and those positions that are inaccessible in the presence of protein because the correctly folded form of the molecule tucks them into the interior.

Armed with a secondary structure that is strongly similar to the bacterial consensus and a knowledge of the solvent-exposed regions in the holoenzyme we were able to proceed with tertiary structure modeling. The results [19, 20] show that the yeast consensus conforms closely to the bacterial tertiary structure models [21], especially in the regions where the most conserved positions are in proximity to the tRNA [21]. The most highly conserved patches of sequence shared by the eukaryotic and bacterial consensus structures are indicated by arrowheads in Figure 1.

Directed mutagenesis

Recently we have initiated mutagenesis studies where in these sequence patches are randomized (4–6 nucleotides at a time) to produce a library of 254 to 4000 possible variants, and then select or screen for function in yeast (Pagán-Ramos et al., in press). As predicted, a relatively small percentage of the variants (1–4%) are viable. However, almost all of these invariant positions appear to tolerate at least some types of substitutions [20]. This was especially pronounced in the region from 206–211, where every position gave multiple viable substitution mutants. This is a particularly interesting position for two reasons. First, the corresponding bacterial consensus at this position is ACAGAPuA, which strongly resembles the catalytic heart of the eukaryotic spliceosome in U6 RNA. Second, it is predicted to lie adjacent to the cleavage site, but on the opposite side from the P4 stem residues that are postulated to be part of the active site [22]. The phenotypes conferred by mutations in this region are consistent with an indirect role in catalysis. Many of these variants were temperature-sensitive or required elevated magnesium *in vivo* and *in vitro* (Pagán-Ramos et al., submitted). Since multiple magnesium ions are required cofactors in the cleavage reaction, it is possible that this domain is involved in precise magnesium coordination, and that the viable mutations represent a selection of sequences that only interfere with the coordination of that site enough to retain permissible coordination at physiological levels of magnesium.

While many of the mutations in other regions of the RNase P RNA are without obvious phenotype or have relatively uninteresting folding defects, kinetic analyses of the viable holoenzyme variants from randomization libraries at these highly conserved positions is expected to shed light on the functions of the subdomains.

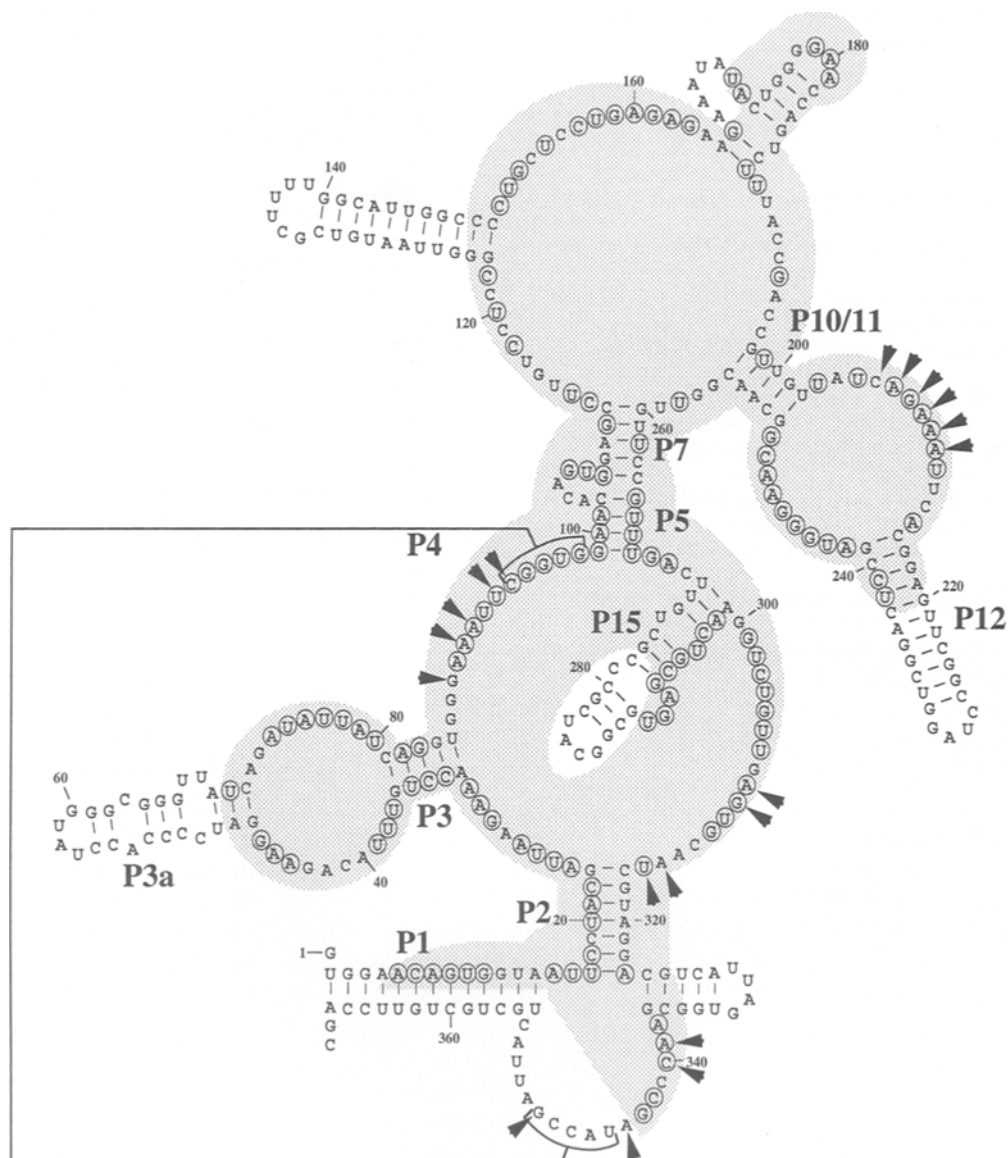


Figure 1. Secondary structure model of *Saccharomyces cerevisiae* RNase P RNA. A secondary structure model of yeast RNase P RNA as it exists in the holoenzyme was derived by a combination of phylogenetic structure analysis [14] and RNA footprinting in solution [16]. Paired regions shared between yeast and eubacterial RNase P RNAs are labeled P1-P15. Clustered positions that have tightly conserved identities even in bacteria are indicated by arrowheads. Positions that are conserved only within yeast are indicated by circled letters. The regions of the RNA that are protected from nuclease attack in the holoenzyme, but not the deproteinized enzyme, are shaded.

Acknowledgements

This work was supported by National Institutes of Health grant GM34869 to DRE, predoctoral fellowships to EP-R from Merck and from the University of Michigan H. H. Rackham Graduate School, and a pre-

doctoral fellowship to AJT from the Life and Health Insurance Medical Research Fund. Computer work was facilitated by National Institutes of Health grant M01 RR00042 to the University of Michigan Clinical Research Center. Oligonucleotide synthesis was facil-

itated by NIH grant P30 CA46592 to the University of Michigan Cancer Center.

References

1. Altman S (1989) *Adv. Enzymol. Relat. Areas Mol Biol.* 62: 1–36
2. Guerrier-Takada C, Gardiner K, Marsh T, Pace N & Altman S (1983) *Cell* 35: 849–857
3. Akaboshi E, Guerrier-Takada C & Altman S (1980) *Biochem. Biophys. Res. Comm.* 96: 831–837
4. Liu M-H, Yuan Y & Reddy R (1994) *Mol. Cell. Biochem.* 130: 75–82
5. Morales MJ, Dang YL, Lou YC, Sulo P & Martin NC (1992) *Proc. Natl. Acad. Sci. USA* 89: 9875–9879
6. Zimmerly S, Drinas D, Sylvers LA & Söll D (1993) *Eur. J. Biochem.* 217: 501–507
7. Lygerou Z, Mitchell P, Petfalski E, Séraphin B & Tollervey D (1994) *Genes & Devel.* 8: 1423–1433
8. Stathopoulos C, Kalpaxis DL & Drinas D (1995) *Eur. J. Biochem.* 228: 976–980
9. Altman S, Wesolowski D & Puranam RS (1993) *Genomics* 18: 418–422
10. Krupp G, Cherayil B, Frendewey D, Nishikawa S & Söll D (1986) *EMBO J.* 5: 1697–1703
11. Lee J-Y & Engelke DR (1989) *Mol. Cell. Biol.* 9: 2536–2543
12. Lee J-Y, Rohlman CE, Molony LA & Engelke DR (1991) *Mol. Cell. Biol.* 11: 721–730
13. Lee J-Y, Evans CF & Engelke DR (1991) *Proc. Natl. Acad. Sci. USA* 88: 6986–6990
14. Tranguch AJ & Engelke DR (1993) *J. Biol. Chem.* 268: 14045–14053
15. Zimmerly S, Gamulin V, Burkard U & Söll D (1990) *FEBS Letters* 271: 189–193
16. Tranguch AJ, Kindelberger DW, Rohlman CE, Lee J-Y & Engelke DR (1994) *Biochemistry* 33: 1778–1787
17. Haas ES, Morse DP, Brown JW, Schmidt FJ & Pace NR (1991) *Science* 254: 853–856
18. Pagán-Ramos E, Tranguch AJ, Kindelberger DW & Engelke DR (1994) *Nucleic Acids Res.* 22: 200–207
19. Tranguch AJ (1994) Ph.D. Dissertation, The University of Michigan, pp. 102–111
20. Pagán-Ramos E, Tranguch AJ, Nolan JM, Pace NR & Engelke DR (1995) In: Agris PF (Ed) *Symposium on RNA Biology, Nucleic Acids Symposium Series* (33: 89–91) Oxford University Press
21. Harris ME, Nolan JM, Malhotra A, Brown JW, Harvey SC & Pace NR (1994) *EMBO J.* 13: 3953–3963
22. Harris ME & Pace NR (1995) *RNA* 1: 210–218