Modeling the Thermoproteaceae RNase P RNA

Patricia P. Chan,1 James W. Brown2 and Todd M. Lowe1,*
1Department of Biomolecular Engineering, University of California Santa Cruz; Santa Cruz, CA USA; 2Department of Microbiology, North Carolina State University; Raleigh, NC USA

The RNA component of the RNase P complex is found throughout most branches of the tree of life and is principally responsible for removing the 5’ leader sequence from pre-tRNA transcripts during tRNA maturation. RNase P RNA has a number of universal core features, however variations in sequence and structure found in homologs across the tree of life require multiple Rfam covariance search models to detect accurately. We describe a new Rfam search model to enable efficient detection of the diminutive archaeal Type T RNase P RNAs, which are missed by existing Rfam models. Using the new model, we establish effective score detection thresholds, and detect four new RNase P RNA genes in recently completed genomes from the crenarchaeal family Thermoproteaceae.

Introduction

Ribonuclease P (RNase P) has been studied intensively for its role in removing the 5’-leaders from pre-tRNA transcripts during maturation. This ribonucleoprotein complex includes one or more well-studied proteins which vary by phylogenetic domain, and has one catalytic RNA subunit in most species with the notable exceptions among land-plants, mitochondria, chloroplasts and a small number of thermophilic microbes. The RNase P RNA (RPR) is the most evolutionarily conserved subunit of this complex, with characteristic structural differences among bacteria, archaea and eukaryotes.1 RPRs typically consist of two structural domains with separate functions: the specificity domain involved in substrate binding, and the catalytic domain needed for enzymatic cleavage. The Rfam database classifies all known RPRs into four different families: nuclear RNase P from eukaryotes, types A or B RNase P from bacteria and archaeal RNase P.2 Although grouped together by Rfam, archaeal RPRs can be further divided into the two distinct types A and M.3 The structure of archaeal type A RPR closely resembles that of bacterial type A RPR, and is the most common archaeal form found in currently sequenced genomes. The type M archaeal RPR, by contrast, lacks highly conserved RNA stem-loop structures in both the specificity, and catalytic domains; it has been found within the euryarchaeal genera Archaeoglobus, Methanocaldococcus, Methanococcus and Methanothermococcus. A new, significantly shortened form of archaeal RPR, type T, was recently found in multiple species within the crenarchaeal clade Thermoproteaceae, adding a third distinct form to Archaea.4 Due to the absence of most of the specificity domain in this variant, the existing Rfam archaeal covariance model fails to identify it. Here, we review the features of the archaeal type T RPR, and the development of a covariance model to identify this unusual, newly recognized form of the RNA. Using this Rfam model, we detected additional type T RPR genes in newly available Thermoproteaceae genomes. In the course of our survey of all archaeal genomes, we also unexpectedly identified a novel type M variant in the clade Archaeoglobaceae.

Results and Discussion

Common features of type T RNase P RNAs. The shortened, type T form of
RNase P RNA was recently described\(^4\) in species of the genus *Pyrobaculum* (*P. aerophilum*, *P. arsenaticum*, *P. calidifontis*, *P. islandicum*, *P. oguniense* and *P. neutrophilum*), *Caldivirga maquilinensis*, and *Vulcanisaeta distributa*; all belong to the same phylogenetic family, Thermoproteaceae. In general, all type T RPRs have a catalytic domain closely resembling that of archaeal type A RPRs, but lack most of the specificity domain (Fig. 1). While the universally conserved positions in the P4 stem, the P2/ P4 joining region, and the P15/P2 joining region are present in type T RPRs, we note four specific differences that help to distinguish type T from other forms. First, the P2 stem is only 3 bp in length, which is relatively short compared with the 6 bp or 7 bp stems found in other archaea or bacteria, respectively.\(^5\) Second, the P15 stem in all identified type T RPRs is 1 bp shorter than the typical P15 found in type A RPRs. Third, the 2-nt P5/P15 linker is contracted compared with the typical 3-nt linker usually found. Fourth, the P10 stem that typically extends to P11, and P12 of the specificity domain in type A RPRs is terminated with a small loop (Fig. 1B and D) or is completely missing (Fig. 1C).

**Type T RNase P RNA variants.** Closer inspection of the secondary structures among the identified type T RPRs reveals three variants, one for each genus (Fig. 1). The 20-nt P1 stem in *C. maquilinensis* and *V. distributa* RNAs is about twice the length of those in *Pyrobaculum*. Although a long P1 stem has been observed in the predicted type A RPR of *Aeropyrum*

---

**Figure 1.** Predicted secondary structures of type T RNase P RNAs. (A) *Methanothermobacter thermoautotrophicus* RNase P RNA (RPR), a typical archaeal type A RPR, has both catalytic and specificity domains.\(^1\) It is shown for comparison with type T RPRs. Common structural differences between type A and type T RPRs shown in red. Universally conserved nucleotides depicted by black circles—(B–D) Type T RPRs found in *Pyrobaculum aerophilum*, *Caldivirga maquilinensis* and *Vulcanisaeta distributa* have structural differences in P1, P5, P7, P8 and P9 stems, shown in blue.
perranix. The length of these type T members is among the longest in all verified archaeal RPRs. It was found in previous studies that P1 interacts with the terminal loop of P9 (L9) as part of the mechanism for orienting the catalytic, and specificity domains in bacterial RPRs. A longer P1 stem that can contact L9 was found to significantly increase the catalytic activity of RNase P in Methanothermobacter thermoautrophicus. While both Pyrobaculum and V. distributa RPRs have a typical GNRA tetraloop in L9, this tetraloop does not exist in C. maquilingensis extended P7 stem (Fig. 1C), which has taken the place of P9 and P10. Thus, this atypical non-GNRA terminal loop may not serve to enhance catalytic activity in C. maquilingensis.

A typical P8 stem, similar to the one in archaeal type A RPRs, is only observed in V. distributa, but not in the other two variants (Fig. 1BD). P8 was found to be involved in T-loop recognition of pre-RNAs in bacteria, mostly by interacting with L18 which is also absent in all archaeal RPRs. The non-essentiality of P8 may be explained by recent studies demonstrating the replacement of the L18-P8 interaction by a protein-protein association and structural evidence for an indirect role of P8 in recognition of the T-loop.

A few other characteristics distinguish type T variants. The C. maquilingensis and V. distributa RNAs have the shortest P5 stem (2 bp vs. a typical 4 bp) observed in archaea. In addition, the V. distributa variant has a 2-nt joining region between P5 and P7, whereas other archaeal RPRs have no joining region.

Searching with the type T covariance model. A previously developed covariance model built with only the Pyrobaculum RPR sequences does not perform well in searching for the two other type T variants. This lack of generality is most likely due to the subtle differences in secondary structure noted above, as well as large disparity in G/C content between the Pyrobaculum RPRs (74–78%) vs. those found in Caldibirga maquilingensis (61%) and Vulcanisaeata distributa (66%). We therefore structurally aligned the RPR sequences from Caldibirga maquilingensis, Vulcanisaeata distributa and all six Pyrobaculum species to create a type T covariance model using Infernal software.

To establish a false-positive score threshold for this model, we scanned 20 randomly generated genomes at each of 3 different G/C contents (< 40, 50 and > 60%). The maximum false positive scores for these were 0, 16.8 and 13.24 bits respectively. For comparison, we scanned the same randomly generated genomes with the existing Rfam archael RNase P covariance model and obtained scores within similar ranges (Table 1).

By employing this newly expanded model to new genomes, we identified four additional shortened forms of RPR, all within species in the Thermoproteaceae family: Pyrobaculum sp 1860, Vulcanisaeata moutnovskia, Thermoproteus tenax and Thermoproteus uzoniensis (Fig. 2; Data S1). The scores for these new identifications were close to the observed range for RPR sequences in the training set (124.2–167.0 bits; Data S2) and far exceeded the false positive threshold (16.8 bits), indicating that these are reliable new identifications. As expected, the RPRs in P. sp 1860 and V. moutnovskia have the same secondary structure, and over 88% sequence identity when compared with other Pyrobaculum species and V. distributa (Fig. 2A and B). A partial RPR sequence fragment with a score of 33.61 bits was also detected in P. sp 1860, which is not similar to the high-scoring version found, so its origin is uncertain.

Manual structural comparison shows that the RPRs in T. uzoniensis and T. tenax could be considered as Pyrobaculum type T variants, with sequence features highly similar to the Pyrobaculum orthologs (Fig. 2C and D). The 16S rRNA genes of T. tenax and T. uzoniensis place them closer to Pyrobaculum species (96%) than C. maquilingensis and V. distributa (93% and 94% respectively), consistent with the relative similarities of the new RPR genes. We also searched the P. sp 1860, V. moutnovskia, T. tenax and T. uzoniensis genomes with the existing Rfam archael covariance model to ensure there was only one RPR per genome and as expected, did not find any additional matches. A search for the RNase P proteins revealed likely homologs of Pop5, Rpp30 and Rpp29, but not Rpp21, as we previously observed for the other Pyrobaculum and Vulcanisaeata species, further solidifying the genetic association of type T RPR and the conspicuous absence of Rpp21.

Variations of P8. Loss of the P8 stem in type M RPRs has been noted as one of the key structural differences distinguishing them from type A archael RPRs (Fig. 3A). However, limited representation of RPRs from some archael clades necessarily allowed a limited assessment of the consistency of this feature among type M RPRs. While conducting structural comparisons between the type T and type M RPR genes, we identified a novel type M variant that includes a typical P8 stem.

Table 1. Summary of RNase P RNA search results

| Genome                | Range of Covariance Model Search Score (bits) |
|-----------------------|---------------------------------------------|
|                       | Archaeal RNase P RNA Model (RF00373)         | Archaeal Type T RNase P RNA Model |
| Thermoproteaceae      | Not Detected – 10.20                         | 117.04 – 168.79                   |
| Other archaea         | 53.55 – 228.58                               | Not Detected – 13.04               |
| Virtual genomes with < 40% GC | Not Detected                                  | Not Detected                      |
| Virtual genomes with about 50% GC | Not Detected – 14.34                        | Not Detected – 16.80               |
| Virtual genomes with > 60% GC | Not Detected                                  | Not Detected – 13.24               |

Infernal v1.0 vsearch was used with the archael type T and existing Rfam archaeal RPR covariance models to search archael genomes (Table S1) and 60 virtual genomes representing G/C content of < 40%, 50% and > 60%. Ranges of bit scores were reported. “Not Detected” indicates that no hits were identified when using default Infernal final score cutoff (0.0).
in three recently sequenced euryarchaea: 
Archaeoglobus profundus, Archaeoglobus 
veneficus and Ferroglobus placidus. This 
was not expected given that Archaeoglobus 
fulgidus, also belonging to the same phy-
logenetic family (Archaeoglobaceae), was 
previously found to lack the P8 stem and 
have a “typical” type M RPR. Like the 
other type M RPRs, the genes in A. pro-
fundus, A. veneficus and F. placidus do 
not have L15, P16, P17 and P6 in their 
predicted structures. Yet, the presence 
of P8 in these species represents a novel 
combination of structural traits (Fig. 3B 
and C). The well-studied A. fulgidus now 
appears to be more similar in terms of 
RPR features to those found in metha-
nogens and not as representative of RPRs 
in the currently available members of the 
Archaeoglobaceae.

Materials and Methods

Genomic data. Complete genomic 
sequences and annotated ORFs for all 
archaeal genomes were obtained from 
NCBI RefSeq. The programs cmbuild and 
cmcalibrate (Infernal v1.0 software package) 
took this file as input to build and cali-
brate the type T covariance model.

Archaeal RNase P RNA sequence 
search. The Infernal v1.0 program 
cmsearch was used to scan for RPR can-
idates in archaeal genomes using both 
the type T RPR covariance model and 
the existing Rfam archaeal RPR covari-
ance model (RF00373). Randomly gen-
erated genomes were scanned with the 
covariance models to determine the false 
positive score threshold. Six genomes 
(Methanococcus maripaludis S2, Sulfolobus 
solfataricus, Pyrobaculum aerophilum, 
Methanobacterium thermoautrophi-
cus, Halogeometricum borinquense and 

Figure 2. Predicted secondary structures of (A) Pyrobaculum sp 1860, (B) Vulcanisaeta moutnovskia, (C) Thermoproteus tenax and (D) Thermoproteus 
uzoniensis RNase P RNAs (RPRs). (A and B) Black circles indicate universally conserved nucleotides. Other highlighted bases in P. sp 1860 and V. mout-
ovskia are relative to other species in the same genus, P. aerophilum (Fig. 1B) and V. distributa (Fig. 1D), respectively. Annotated nucleotides show base 
pairing covariation (green), conservative G-C to G-U changes (yellow) and differences in unpaired regions (blue). Lower case red nucleotides show 
insertions or deletions between RPRs. (C and D) Predicted secondary structures of RPRs in T. tenax and T. uzoniensis resemble the Pyrobaculum type T 
RPR variant (Fig. 1B).
level of complexity to the architectural diversity of RNase P enzymes. The presence and absence of the P8 stem in different, closely related species suggests recent genetic swapping of RPR in *Archaeoglobus fulgidus* by lateral transfer. With the increasing availability of sequenced genomes, we anticipate that the new type T RPR model will help identify new variants for study and thus enable a more complete understanding of this dynamic RNA gene family.

**Disclosure of Potential Conflicts of Interest**
No potential conflicts of interest were disclosed.

**Acknowledgments**
This work was supported by National Science Foundation Grant EF-082277055.

**Supplemental Materials**
Supplemental materials may be found here: www.landesbioscience.com/journals/rnabiology/article/21502

**Figure 3.** Predicted secondary structures of type M RNase P RNA variants. (A) *Archaeoglobus fulgidus* has a typical archaeal type M RNase P RNA (RPR) and is shown for comparison. (B and C) Newly identified type M RPR variants in *Archaeoglobus veneficus* and *Ferroglobus placidus* have a P8 stem (red) that is missing in other type M RPRs. Other colored nucleotides are annotated as in Figure 2, indicating changes in (B and C) relative to (A).
References

1. Brown JW. The Ribonuclease P Database. Nucleic Acids Res 1999; 27:314; PMID:9847214; http://dx.doi.org/10.1093/nar/27.1.314.

2. Gardner PP, Daub J, Tate J, Moore BL, Osuch HI, Griffiths-Jones S, et al. Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res 2011; 39(Database issue):D141-5; PMID:21062808; http://dx.doi.org/10.1093/nar/gkrq129.

3. Harris JK, Haas ES, Williams D, Frank DN, Brown JW. New insight into Rnase P RNA structure from comparative analysis of the archaeal RNA. RNA 2001; 7:220-32; PMID:11233979; http://dx.doi.org/10.1017/S1355838201001777.

4. Lai LB, Chan PP, Cozen AE, Bernick DL, Brown JW, Gopalan V, et al. Discovery of a minimal form of Rnase P in Pyrobaculum. Proc Natl Acad Sci USA 2010; 107:22493-8; PMID:21135215; http://dx.doi.org/10.1073/pnas.1013961107.

5. Haas ES, Armbruster DW, Vuczon BM, Daniels CJ, Brown JW. Comparative analysis of ribonuclease P RNA structure in Archaea. Nucleic Acids Res 1996; 24:1252-9; PMID:8614627; http://dx.doi.org/10.1093/nar/24.7.1252.

6. Massire C, Jaeger L, Westhof E. Phylogenetic evidence for a new tertiary interaction in bacterial Rnase P RNA. RNA 1997; 3:553-6; PMID:9174090.

7. Massire C, Jaeger L, Westhof E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. J Mol Biol 1998; 279:773-93; PMID:9642060; http://dx.doi.org/10.1006/jmbi.1998.1797.

8. Li D, Willkomm DK, Hartmann RK. Minor changes largely restore catalytic activity of archael Rnase P RNA from Methanothermobacter thermoautotrophicus. Nucleic Acids Res 2009; 37:231-42; PMID:19036794; http://dx.doi.org/10.1093/nar/gkn095.

9. Nolan JM, Burke DH, Pace NR. Circularly permuted tRNAs as specific photofinity probes of ribonuclese P RNA structure. Science 1993; 261:762-5; PMID:7688143; http://dx.doi.org/10.1126/science.7688143.

10. Harris ME, Nolan JM, Malhotra A, Brown JW, Harvey SC, Pace NR. Use of photofinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. EMBO J 1994; 13:3953-63; PMID:7521297.

11. Li D, Gossringer M, Hartmann RK. Archaeal-bacterial chimeric Rnase P RNAs: towards understanding Rnase P's architecture, function and evolution. Chembiochem 2011; 12:1536-43; PMID:21574237; http://dx.doi.org/10.1002/cbic.201100054.

12. Reiter NJ, Onerman A, Torres-Larios A, Swinger KK, Pan T, Mondragon 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.

13. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. Bioinformatics 2009; 25:1335-7; PMID:19367242; http://dx.doi.org/10.1093/bioinformatics/btp157.

14. Pruitt KD, Tarasova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007; 35(Database issue):D61-5; PMID:17130148; http://dx.doi.org/10.1093/nar/gkl842.

15. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res 2010; 38(Database issue):D211-22; PMID:19920124; http://dx.doi.org/10.1093/nar/gkp985.

16. Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. J Comput Biol 2004; 11:413-28; PMID:15285899; http://dx.doi.org/10.1089/1066527041410472.

17. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC Archeal Genome Browser: 2012 update. Nucleic Acids Res 2012; 40(Database issue):D646-52; PMID:22080555; http://dx.doi.org/10.1093/nar/gkr990.