The Design and Catalytic Properties of a Simplified Ribonuclease P RNA

DAVID S. WAUGH, CHRISTOPHER J. GREEN, NORMAN R. PACE

Ribonuclease P (RNase P) RNA is the catalytic moiety of the ribonucleoprotein enzyme that removes precursor sequences from the 5′ ends of pre–transfer RNAs in eubacteria. Phylogenetic variation according to recently proposed secondary structure models was used to identify structural elements of the RNase P RNA that are dispensable for catalysis. A simplified RNase P RNA that consists only of evolutionarily conserved features was designed, synthesized, and characterized. Although the simplified RNA (*Min* 1 RNA) is only 263 nucleotides in length, in contrast to the 354 to 417 nucleotides of naturally occurring RNase P RNAs, its specificity of pre-tRNA cleavage is identical to that of the native enzymes. Moreover, the catalytic efficiencies of the *Min* 1 RNA and the native RNA enzymes are similar. These results focus the search for the catalytic elements of RNase P RNAs to their conserved structure.

THE MATURE 5′ ENDS OF tRNAs ARE formed by the action of ribonuclease P (RNase P), an endonuclease. In vivo, the eubacterial RNase P consists of essential protein (~14 kD) and RNA (~400 nucleotides) components (*1*). However, the RNase P RNA alone is an efficient and accurate catalyst at high salt concentrations in vitro (*2*). The high ionic strength evidently screens anionic repulsion between the enzyme and substrate RNAs (*3*).

Understanding the mechanism of RNase P–mediated catalysis requires the identification of the structural elements of the RNase P RNA that are important for the reaction. One way to identify potentially important structure in the RNA is by deleting sequences that are not required for activity, seeking the minimum functional structure. This approach has been useful in localizing the portion of tRNA that is recognized by RNase P (*4*) and in the study of other catalytic RNAs (*5*).

The RNase P RNAs are too complex for an efficient identification of nonessential sequences by random deletion mutagenesis. We therefore used phylogenetic comparisons of secondary structure models to identify potentially dispensable sequences. Secondary structure models of nine RNase P RNAs from two eubacterial phyla, the "Gram-positive bacteria" (such as *Bacillus megaterium* in Fig. 1) and the "purple bacteria" (such as *Escherichia coli* in Fig. 1), have been derived with the use of a comparative approach to identify base-paired sequences (*6*). Differences between the models of the RNAs from the two phyla are due substantially to the phylum-specific occurrence of discrete helical elements at various positions in a core of homologous primary and secondary structure (*7*). A tenet of phylogenetic theory is that evolutionary conservation of function is accompanied by conservation of essential structure. Since the phylum-specific helical elements are not present in the RNase P RNAs of all organisms, they are not expected to be crucial for the enzymatic activity of the RNA. In order to test this hypothesis, we relied on the structure models to design a simplified RNase P RNA that consists only of phylogenetically conserved features.

The *E. coli* RNase P RNA (M1 RNA) was used as the foundation of the design because its structure model has fewer interruptions in the conserved core than do those of the RNase P RNAs from Gram-positive bacteria. There are four proposed helices in the *E. coli* RNase P RNA model that have no counterparts in the Gram-positive versions and so are not expected to be essential for enzymatic activity. These helices involve nucleotides 28 to 52, 150 to 165, 183 to 227, and 260 to 290 of the folded *E. coli* sequence in Fig. 1. They were excluded from the design of the simplified RNA through the use of two strategies: (i) simple omission from the design, and (ii) replacement of *E. coli* sequences with segments of a Gram-positive RNase P RNA that lacks the *E. coli*–specific helical elements.

Nucleotides 28 to 52 in the *E. coli* RNase P RNA structure model form the apical part of a longer helix that involves nucleotides 20 to 61. Since only the core-proximal portion of this helix (nucleotides 20 to 27 and 53 to 61) has a homolog in the Gram-positive RNA models (such as *B. megaterium* in Fig. 1), nucleotides 28 to 52 were omitted in the design of the simplified RNA. The residual stem was capped with the former loop (UUCG, nucleotides 39 to 42). We were concerned that the shortened helix might be less stable than the original one, so U_{25} in the *E. coli* sequence was replaced with C in the simplified RNA. This change converts a G-U pair to an ostensibly more stable G-C pair in the proposed, residual helix.

We were reluctant simply to omit from the design the three remaining phylum-specific helices present in the *E. coli* RNase P RNA because of concern that, in the absence of detailed knowledge of their structures, we might not properly restore the remnant secondary and tertiary structure. Consequently, the regions of the *E. coli* RNA sequence containing those helices were replaced with blocks of Gram-positive sequence that occupy homologous positions in the proposed structure models, yet contain fewer nucleotides. Nucleotides 137 to 225 and 257 to 291 in the *E. coli* RNase P

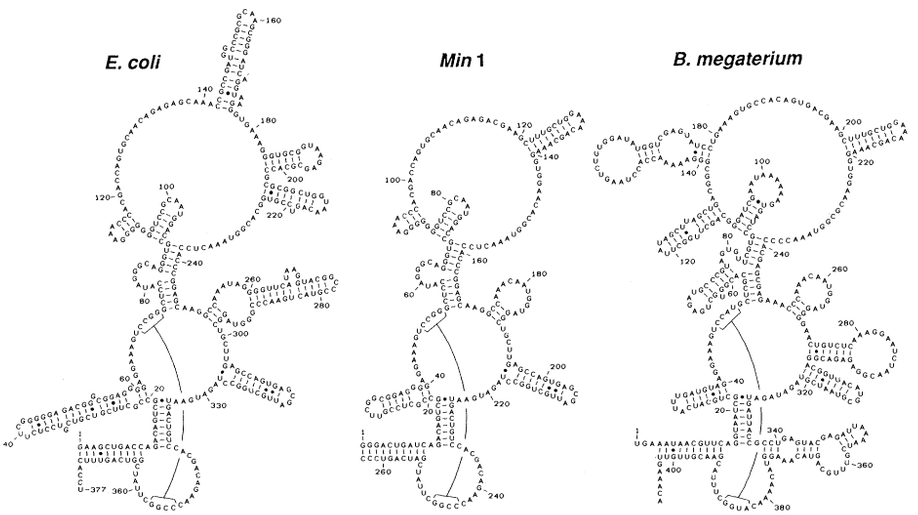

Fig. 1. Proposed secondary structures of the *B. megaterium* and *E. coli* RNase P RNAs based on phylogenetic comparisons (*6*), and the proposed secondary structure of the *Min* 1 RNase P RNA. The 5′ terminal nucleotide in each sequence corresponds to position 1. Base-paired residues are indicated by hyphens (canonical Watson-Crick pairs) or by solid dots (noncanonical pairs, such as G·U). The brackets denote a potential pseudoknot interaction.

D. S. Waugh and N. R. Pace, Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, IN 47405.
C. J. Green, Department of Biomedical Research, SRI International, Menlo Park, CA 94025.

RNA were respectively replaced in the designed sequence with nucleotides 196 to 226 and 258 to 261 from the *B. megaterium* RNA (*8*). Thus the simplified RNA design is a composite of RNase P RNA sequences from two different organisms.

A few additional changes from the *E. coli* sequence were made near the ends of the designed RNA in order to create useful restriction sites and to accommodate a bacteriophage T7 RNA polymerase promoter (*9*). These latter alterations maintained the potential for base pairing between the termini of the synthetic RNA. The proposed secondary structure of the simplified RNase P RNA, called *Min* 1 RNA, is shown in Fig. 1. The double-stranded *Min* 1 RNA gene adjacent to a T7 promoter was assembled from synthetic oligodeoxyribonucleotides and cloned into a plasmid vector (*10*). The *Min* 1 RNA was produced by run-off transcription in vitro and its catalytic properties were examined as described (*11*).

The *Min* 1 RNA is catalytically active and has a cleavage specificity for pre-tRNA that is indistinguishable from that of the naturally occurring RNase P RNAs. However, its kinetic properties (Table 1) and optimum reaction conditions (Fig. 2) differ from those of the native (*E. coli*) RNA (*12*). The Michaelis constant (K_m) of the *Min* 1 RNA is approximately two orders of magnitude greater than that of the native RNA, indicating that the *Min* 1 RNA has less affinity for the substrate than does the native RNA. However, at their respective optimal ionic strengths, the maximum velocity (k_{cat}) of the *Min* 1 RNA is 20-fold greater than that of the native RNA and about the same as that of the holoenzyme under physiological conditions (*13*). The enhanced k_{cat} of the *Min* 1 RNA probably reflects a more rapid release of product than occurs with the native RNA, for which product release is rate limiting (*3*). Since RNase P RNA binds to the mature domain of the pre-tRNA, the reduced affinity (higher K_m) of the *Min* 1 RNA for the substrate would result in more rapid dissociation of enzyme and product following cleavage. Overall, the catalytic efficiencies, or specificity constants (k_{cat}/K_m) (*14*), of the *Min* 1 and *E. coli* RNase P RNAs are similiar.

The *Min* 1 RNase P RNA activity is significantly more sensitive to temperature and requires higher concentrations of monovalent (but not divalent) salt than the native RNA activity (Fig. 2). The temperature sensitivity of the *Min* 1 RNA activity suggests that the intramolecular forces responsible for its folding may be weaker than in the native RNA, possibly because of the absence of the phylum-specific structures or a defect in the design of the RNA. The resulting instability could also be responsible for the increased salt dependence of the *Min* 1 activity. If the intramolecular packing forces of the *Min* 1 RNA are weaker than those of the native RNA, then electrostatic repulsion between phosphates could distort the structure into an inactive form. The high salt concentration required for activity of the *Min* 1 RNA could provide counterions to titrate such destabilizing repulsion and allow the RNA to assume its active conformation.

Table 1. Kinetic parameters for processing of pre-tRNA^{Asp} in vitro by the *E. coli* and the *Min* 1 RNase P RNAs. The activity of both RNase P RNAs was assayed at 37°C in 50 m*M* magnesium acetate, 50 m*M* tris-acetate (*p*H 8.0), and 0.05% Nonidet P40 (Sigma). Reactions containing the *E. coli* RNase P RNA (10^{-9} *M*) also included 0.5*M* ammonium acetate, and reactions containing the *Min* 1 RNase P RNA (10^{-8} *M*) also included 3*M* ammonium acetate. For each RNA enzyme, assays were conducted at various substrate concentrations above and below K_m. Aliquots were withdrawn from the reactions at regular intervals and processed as described in the legend to Fig. 2.

RNase P RNA	K_m (*M*)	k_{cat} (min^{-1})	K_{cat}/K_m (M^{-1} min^{-1})
Min 1	5×10^{-6}	10	2×10^6
E coli.	4×10^{-8}	0.4	1×10^7

The addition of the RNase P protein from either *E. coli* or *Bacillius subtilis* does not stimulate the activity of the *Min* 1 RNase P RNA at physiological ionic strength. This observation might suggest that one or more of the omitted structural elements are required for interaction with the proteins. However, the RNAs from Gram-positive organisms also lack these structures, and yet are able to form active holoenzymes with the *E. coli* RNase P protein (*2*). Consequently, we believe that the omitted structures are not directly involved in the binding of the proteins. The inability of the proteins to stimulate catalysis by the *Min* 1 RNA at physiological ionic strength may be a consequence of the requirement by the synthetic RNA for particularly high salt concentrations in order to achieve appreciable activity. The conditions required to activate the *Min* 1 RNA may be incompatible with the function or assembly of the holoenzyme. The native holoenzyme also is inactive at high salt concentrations (*15*).

The near-native catalytic efficiency of the *Min* 1 RNA (under the appropriate conditions) demonstrates that the phylum-specific structures that were excluded from the design are not necessary for the specific cleavage of pre-tRNA. This result raises the question of their role or roles in vivo. The salt dependence and temperature sensitivity of the *Min* 1 RNA may indicate that one or more of the variable elements contribute to the global stability of the RNA. It is also possible that they relate to unknown, phylum-specific functions of the RNase P RNAs in vivo. In any case, the catalytic function of the naturally occurring RNase P RNAs resides in their phylogenetically conserved structures. The fact that the structure models successfully predict the dispensability of extensive sequences that are scat-

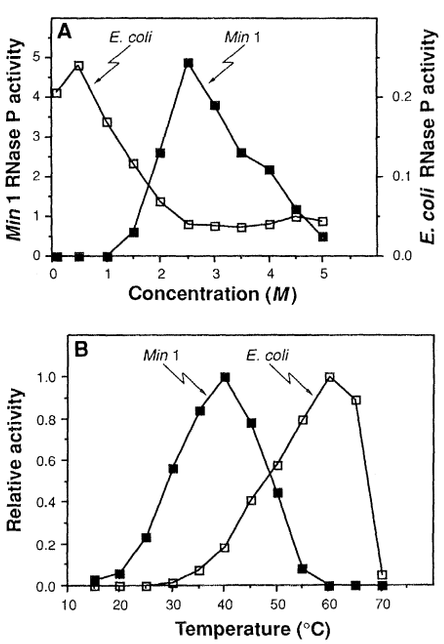

Fig. 2. (**A**) Influence of ammonium acetate concentration on the in vitro tRNA processing activity of the *E. coli* and *Min* 1 RNase P RNAs. All of the reactions (20 μl) contained 50 m*M* tris-acetate (*p*H 8.0), 50 m*M* magnesium acetate, 0.05% Nonidet P40 (Sigma), [^{32}P]pre-tRNA^{Asp} (10^{-7} and 10^{-5} *M* for the *E. coli* and *Min* 1 RNase P RNA reactions, respectively), and either the *E. coli* or the *Min* 1 RNase P RNA at 10^{-8} *M*. The reactions were incubated at 37°C for 20 min and stopped by the addition of 3 volumes of ethanol. The products were recovered as ethanol precipitates, resuspended in 10 μl of 8*M* urea, 20 m*M* EDTA, and 0.1% SDS, and resolved by electrophoresis in 8% polyacrylamide gels containing 8*M* urea. After fixing and drying the gels, the bands were located by autoradiography, excised, and counted. Activity is expressed as moles of substrate cleaved per mole of enzyme per minute. (**B**) Influence of temperature on the in vitro tRNA processing activity of the *E. coli* and *Min* 1 RNase P RNAs. All reactions (20 μl) contained 50 m*M* Hepes (*p*H 8.0), 50 m*M* magnesium acetate, and 0.05% Nonidet P40 (Sigma). Reactions containing *E. coli* RNase P RNA (10^{-8} *M*) also included 0.5*M* ammonium acetate and 10^{-7} *M* [^{32}P]pre-tRNA^{Asp}. Reactions containing the *Min* 1 RNase P RNA (10^{-8} *M*) also included 3*M* ammonium acetate and 10^{-5} *M* [^{32}P]pre-tRNA^{Asp}. Incubations were carried out at 37°C, and the reactions were processed as described above. For each RNA enzyme, "relative activity" denotes the fraction of the maximum activity observed over the indicated temperature range.

tered throughout the interior of the natural RNAs supports the accuracy of the models.

The *Min* 1 RNA may prove more useful than the natural RNase P RNAs for some studies. Because it is smaller than the natural molecules, the *Min* 1 RNA may be more amenable to spectroscopic studies or to crystallization for diffraction analyses. The reduced affinity of the *Min* 1 RNA for the substrate, coupled with its holoenzyme-like maximum velocity, makes *Min* 1 RNA more useful than the natural RNAs for steady-state kinetic analyses that are not complicated by the rate-limiting step of product release. Finally, the *Min* 1 RNA sequence is a new starting point for the elimination of additional sequence elements. We anticipate that further reductions in the size of the active molecule are possible, since many random deletion mutants of the natural RNAs lack conserved features and yet retain some (usually very low) activity (*16, 17*). The inspection of RNase P RNAs from more diverse eubacteria should provide further perspective on the minimal active structure and allow the rational design of RNase P RNAs that are even simpler than the *Min* 1 RNA.

REFERENCES AND NOTES

1. B. C. Stark, R. Kole, E. J. Bowman, S. Altman, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3717 (1978); K. Gardiner and N. R. Pace, *J. Biol. Chem.* **255**, 7507 (1980).
2. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. R. Pace, S. Altman, *Cell* **35**, 849 (1983).
3. C. Reich, G. J. Olsen, B. Pace, N. R. Pace, *Science* **239**, 178 (1988).
4. W. H. McClain, C. Guerrier-Takada, S. Altman, *ibid.* **238**, 527 (1987).
5. J. W. Szostak, *Nature* **322**, 83 (1986); O. C. Uhlenbeck, *ibid.* **328**, 596 (1987).
6. B. D. James *et al.*, *Cell* **52**, 19 (1988).
7. We use the term "homologous" in its strictest sense: homologous sequences have common ancestry and function. Homologous sequences are not necessarily identical; identical sequences are not necessarily homologous.
8. Portions of the *B. megaterium* RNase P RNA sequence were used instead of another Gram-positive RNase P RNA sequence because this would result in the presence of a convenient restriction site in the synthetic gene.
9. Nucleotides 2 to 4 in the *E. coli* sequence were changed from AAG to GGA in the *Min* 1 RNase P RNA in order to accommodate a bacteriophage T7 promoter. Nucleotides 371 to 372 in the *E. coli* sequence were therefore changed from UU to CC in the *Min* 1 RNA in order to maintain Watson-Crick complementarity between the termini. The proposed base pair involving nucleotides 9 and 365 in the *E. coli* sequence was replaced by a different canonical base pair in order to create useful restriction sites (Bcl I and Cla I) near the termini of the synthetic RNase P RNA gene.
10. The two halves of the synthetic gene, defined by the Hind III site in the middle of the designed RNase P RNA, initially were assembled and cloned separately into plasmid vectors. Six synthetic oligodeoxyribonucleotides (40 to 60 nucleotides in length) were annealed to construct each half of the gene. The two halves subsequently were joined at the Hind III site and cloned in a plasmid vector to create pDW133 (*17*). The *Min* 1 RNase P RNA is produced by T7 RNA polymerase-directed transcription of Sma I–cut pDW133 DNA in vitro.
11. C. Reich *et al.*, *J. Biol. Chem.* **261**, 7888 (1986).
12. The *E. coli* RNase P RNA (M1 RNA) was produced by transcription in vitro. An Alu I restriction fragment of *E. coli* DNA that includes all but the first four nucleotides of the RNase P RNA coding sequence (*18*) was joined to a T7 promoter and cloned in pUC19 [C. Yamisch-Perron, J. Vieira, J. Messing, *Gene* **33**, 103 (1985)] to create a template for in vitro transcription (pDW27) (*17*). The T7 RNA polymerase–directed run-off transcript of Sna BI–cut pDW27 DNA has two nucleotides at its 5′ end (GG) and four nucleotides at its 3′ end (UUAC) that are not present in the natural RNase P RNA (Fig. 1). In addition, nucleotides 3 and 4 in the *E. coli* RNase P RNA sequence were changed from AG to UC in the run-off transcript as a consequence of the fusion of the T7 promoter and the truncated RNase P RNA gene.
13. The k_{cat} of the *E. coli* and *Bacillus subtilis* holoenzymes depends upon reaction conditions and the character of the substrate. Examples of values for k_{cat} of the *E. coli* holoenzyme with pre-tRNA substrates are 0.35 min^{-1} [N. Lumelsky and S. Altman, *J. Mol. Biol.* **202**, 443 (1988)], 2 min^{-1} (*2*) 6.6 min^{-1} [C. Guerrier-Takada, A. van Belkum, C. W. A. Pleij, S. Altman, *Cell* **53**, 267 (1988)], and 18.3 min^{-1} (*4*). With the substrate and reaction conditions used in this work, k_{cat} for the holoenzymes of both *E. coli* and *B. subtilis* (*3*) is about 10 min^{-1}.
14. A. Fersht, *Enzyme Structure and Mechanism* (Freeman, New York, 1985), p. 105.
15. K. J. Gardiner, T. L. Marsh, N. R. Pace, *J. Biol. Chem.* **260**, 5415 (1985).
16. C. Guerrier-Takada and S. Altman, *Cell* **45**, 177 (1986); N. P. Lawrence and S. Altman, *J. Mol. Biol.* **191**, 163 (1986).
17. D. S. Waugh, thesis, Indiana University, Bloomington (1989).
18. R. E. Reed, M. F. Baer, C. Guerrier-Takada, H. Donis-Keller, S. Altman, *Cell* **30**, 627 (1982); R. E. Reed and S. Altman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5359 (1983); H. Sakamoto, N. Kimura, F. Nagawa, Y. Shimura, *Nucleic Acids Res.* **11**, 8237 (1983).

19. Supported by NIH grant GM34527 to N.R.P. C.J.G. was supported by NIH grant GM29231 to B. Vold (SRI International). We are grateful to M. Baer and S. Altman for providing *E. coli* RNase P protein, to B. Pace for providing *B. subtilis* RNase P protein, and to G. Olsen for assistance with artwork and many fruitful discussions regarding the design of the *Min* 1 RNase P RNA. We are also indebted to J. Brown, A. Burgin, S. Darr, E. DeLong, D. Hunt, T. Schmidt, D. Smith, and J. Thomas for helpful comments on the manuscript and to S. Irvine for typing.

6 March 1989; accepted 21 April 1989

Repeat-Induced G-C to A-T Mutations in *Neurospora*

EDWARD B. CAMBARERI, BRYAN C. JENSEN, ERIC SCHABTACH, ERIC U. SELKER*

In the *Neurospora* genome duplicate sequences are detected and altered in the sexual phase. Both copies of duplicate genes are inactivated at high frequency, whether or not they are linked. Restriction sites change, and affected sequences typically become heavily methylated. To characterize the alterations of the DNA, duplicated sequences were isolated before and after one or more sexual cycles. DNA sequencing and heteroduplex analyses demonstrated that the process (termed RIP) produces exclusively G-C to A-T mutations. Changes occur principally at sites where adenine is 3′ of the changed cytosine. A sequence duplicated at a distant site in the genome lost approximately 10 percent of its G-C pairs in one passage through a cross. A closely linked duplication of the same sequence that was passed twice through a cross lost about half of its G-C pairs. The results suggest a mechanism for the RIP process.

DNA SEQUENCE DUPLICATIONS provide the critical first step for gene amplification and provide the raw material for evolution of new genes. While vital for evolution, sequence duplications can also have negative consequences. Dispersed repeated genes can mediate exchanges resulting in deletions, inversions, or translocations. In addition, altered gene dosage can result in a detrimental imbalance of gene products. In many organisms, such as fungi and bacteria, virtually all genes are present in one copy per haploid genome.

In the multicellular fungus *Neurospora crassa*, the paucity of duplicated genes may not simply be due to natural selection. Duplications are efficiently detected and altered in a specialized (dikaryotic) tissue formed by fertilization (*1, 2*). The process affects both copies of a duplicated sequence as revealed by gene inactivation, changes in the position of restriction sites, and de novo methylation of cytosines in the repeated DNA. Because the process is limited to the stage between fertilization and nuclear fusion, susceptible cells have a nucleus from each parent. Thus a cell should survive a duplication and inactivation, even of an essential gene, so long as the duplication were only in one of the parents. Both nuclei should deliver their genetic material to meiosis. Standard recombination processes would produce meiotic products having different combinations of altered and unaltered copies of the duplicated sequences (Fig. 1). Just the class of cells

Institute of Molecular Biology and Department of Biology, University of Oregon, Eugene, OR 97403.

*To whom correspondence should be addressed.