In vitro selection of RNase P RNA reveals optimized catalytic activity in a highly conserved structural domain

D. N. Frank, A. E. Ellington and N. R. Pace

RNA 1996 2: 1179-1188

RNA (1996), 2:1179–1188. Cambridge University Press. Printed in the USA.

In vitro selection of RNase P RNA reveals optimized catalytic activity in a highly conserved structural domain

DANIEL N. FRANK,[1] ANDREW E. ELLINGTON,[2] and NORMAN R. PACE[1]

[1] Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3102, USA
[2] Departments of Biology and Chemistry, Indiana University, Bloomington, Indiana 47405, USA

ABSTRACT

In vitro selection techniques are useful means of dissecting the functions of both natural and artificial ribozymes. Using a self-cleaving conjugate containing the *Escherichia coli* ribonuclease P RNA and its substrate, pre-tRNA (Frank DN, Harris ME, Pace NR, 1994, *Biochemistry 33*:10800–10808), we have devised a method to select for catalytically active variants of the RNase P ribozyme. A selection experiment was performed to probe the structural and sequence constraints that operate on a highly conserved region of RNase P: the J3/4-P4-J2/4 region, which lies within the core of RNase P and is thought to bind catalytically essential magnesium ions (Harris ME et al., 1994, *EMBO J 13*:3953–3963; Hardt WD et al., 1995, *EMBO J 14*:2935–2944; Harris ME, Pace NR, 1995, *RNA 1*:210–218). We sought to determine which, if any, of the nearly invariant nucleotides within J3/4-P4-J2/4 are required for ribozyme-mediated catalysis. Twenty-two residues in the J3/4-P4-J2/4 component of RNase P RNA were randomized and, surprisingly, after only 10 generations, each of the randomized positions returned to the wild-type sequence. This indicates that every position in J3/4-P4-J2/4 contributes to optimal catalytic activity. These results contrast sharply with selections involving other large ribozymes, which evolve improved catalytic function readily in vitro (Chapman KB, Szostak JW, 1994, *Curr Opin Struct Biol 4*:618–622; Joyce GF, 1994, *Curr Opin Struct Biol 4*:331–336; Kumar PKR, Ellington AE, 1995, *FASEB J 9*:1183–1195). The phylogenetic conservation of J3/4-P4-J2/4, coupled with the results reported here, suggests that the contribution of this structure to RNA-mediated catalysis was optimized very early in evolution, before the last common ancestor of all life.

Keywords: phylogeny; ribozyme; RNA processing; tRNA

INTRODUCTION

The endonuclease ribonuclease P (RNase P) is an essential enzyme that cleaves precursor-tRNA (pre-tRNA) molecules to generate the mature 5′ termini of tRNAs. This enzyme is ubiquitous and has been characterized to some extent in the three primary lines of evolutionary descent: the Bacteria, Archaea, and Eucarya (Darr et al., 1992a; Pace & Brown, 1995). Indeed, RNase P activity exists in all organelles and cells that transcribe tRNA. In most instances, RNase P functions as a ribonucleoprotein (RNP) complex. The bacterial version of this enzyme is composed of single protein and RNA species, whereas additional proteins are found in the archaeal and eucaryal RNPs (Altman et al., 1993; Pace & Brown, 1995).

The function of RNase P is of particular interest because, at least in the Bacteria, it is the RNA subunit of the holoenzyme that is catalytically active. Thus, the bacterial RNase P is fundamentally a ribozyme and, by extension, it is likely that the archaeal and eucaryal RNase P RNAs also are the catalytic centers of their respective holoenzymes. Consequently, considerable effort has been invested in structural analyses of the RNA moiety of RNase P (primarily the bacterial RNA) in order to ascertain how this ribozyme functions in tRNA maturation (Altman et al., 1993; Kirsebom, 1995; Pace & Brown, 1995). The secondary structure of the bacterial RNase P RNA has now been established by phylogenetic comparative analysis (Fig. 1; Pace & Brown, 1995; Brown et al., 1996) and the tertiary structure of the RNA is being studied actively by biophysi-

Reprint requests to: Norman R. Pace, Department of Plant and Microbial Biology, 111 Koshland Hall #3102, University of California, Berkeley, California 94720-3102, USA; e-mail: nrpace@nature.berkeley.edu.

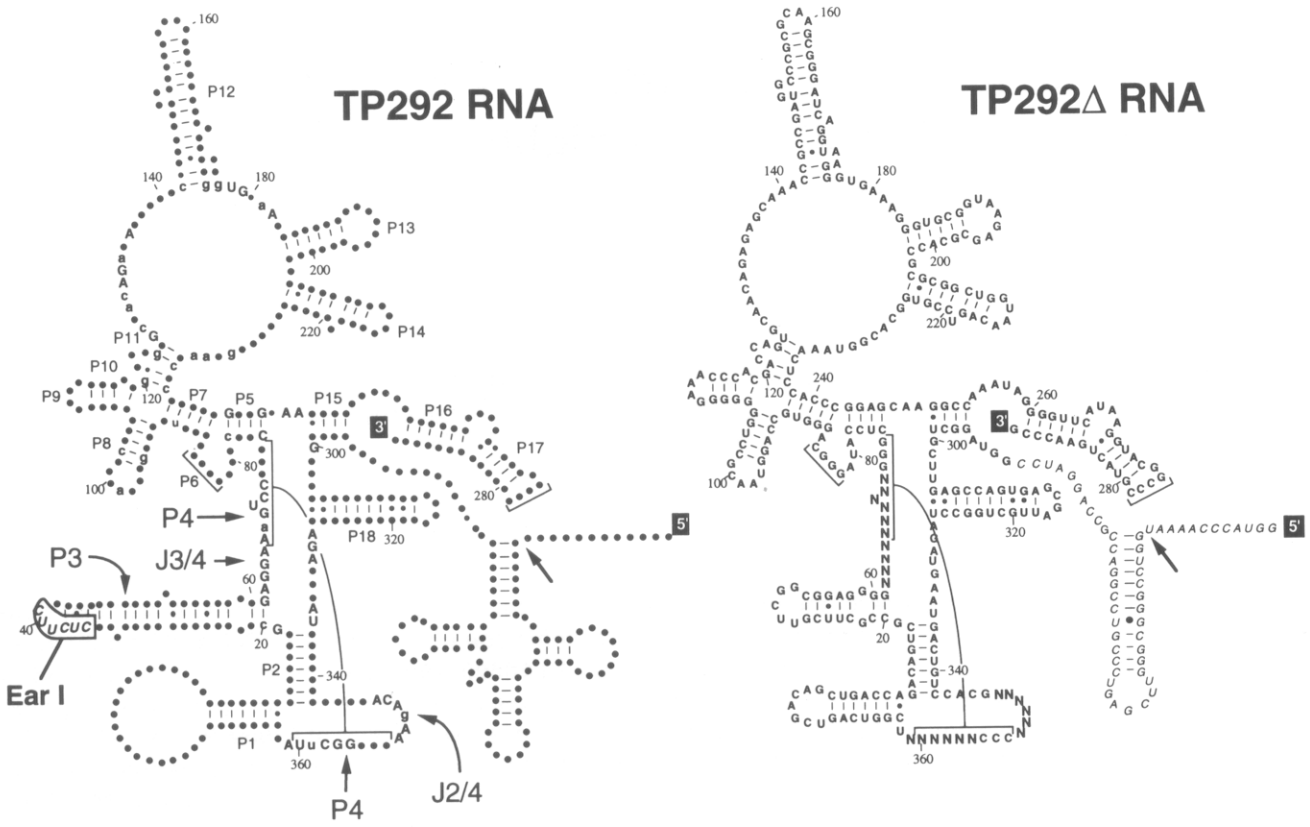


FIGURE 1. Phylogenetic conservation and secondary structures of native (TP292) and randomized (TP292Δ) pre-tRNA/RNase P conjugate RNAs. The nucleotide numbering scheme is based on the *E. coli* RNA (Haas et al., 1994). The boxed region highlights a unique *Ear* I restriction site, which was deleted in construct TP292Δ. Upper case, nucleotide that are at least 80% conserved among bacterial RNase P RNAs. Lower case, nucleotide that are at least 80% conserved among bacterial RNase P RNAs. P, Paired (helical) region. J, sequence Joining two helices. Brackets connected by a line denote helix P4. Adapted from Haas et al. (1994). Nucleotides comprising the 5′ leader, substrate, and linker portions of TP292Δ are italicized, and the RNase P moiety is in bold type. The site of cleavage is denoted by an arrow.

cal and comparative approaches (Harris et al., 1994; Westhof & Altman, 1994; Pan, 1995; Brown et al., 1996).

The underlying assumption of phylogenetic comparative analysis is that evolution conserves sequences and structures that are critical for the overall function of an RNA. For instance, the helical element P4 of the RNase P RNA, along with its flanking "single-stranded" sequence elements (J3/4 and the 3′ portion of J2/4) are nearly invariant in sequence throughout the bacterial domain (Fig. 1; Pace & Brown, 1995); archaeal and eucaryal homologues of this region are also clearly identifiable (Forster & Altman, 1990; Nieuwlandt et al., 1991; LaGrandeur et al., 1993; Tranguch & Engelke, 1993). Such extremely strict conservation implies strongly that this domain is central to RNase P structure and/or function, but does not indicate whether the conserved nucleotides function directly in catalysis or are required for other aspects of RNase P RNA function, such as RNA biosynthesis, stability, or RNase P protein binding. These various functional roles must be deconvoluted experimentally. In fact, recent phosphorothioate modification–interference studies suggest that catalytically essential Mg^{2+} ions bind to phosphates at both ends of P4 (Hardt et al., 1995; Harris & Pace, 1995). We would thus like to understand what roles, if any, the base moieties within the conserved J3/4-P4-J2/4 region play during catalysis. One possibility is that some or all of the conserved nucleotides participate in tertiary structure interactions that form the core of RNase P. However, because most of the P4 region nucleotides are invariant, the comparative approach cannot be used to search for such tertiary interactions. A more direct method of investigating the structure of the J3/4-P4-J2/4 region and its possible influence on catalysis is thus needed.

In vitro genetic selection has proven to be the most efficient means of assaying large, comprehensive sets of mutations that have been introduced into macromolecules (Gold et al., 1993; Szostak & Ellington, 1993). By performing genetic screens on ribozymes, such as RNase P, in vitro, it is possible to focus solely on the determinants of RNA-mediated catalysis, rather than RNA biosynthesis. In this paper, we report the results

of a selection experiment in which 22 nt of the J3/4-P4-J2/4 region of the *Escherichia coli* RNase P RNA were completely randomized. The goal of this selection experiment was to determine the constellation of sequences that can substitute for the native sequence within the J3/4-P4-J2/4 region and maintain catalytic activity. By analyzing the range of permissible sequence variation within this region, we hoped to assess the constraints that RNase P-mediated catalysis places on the identity of each nucleotide. As with comparative sequence analysis, our underlying assumption was that critically important nucleotides would remain invariant or change in a stereotypic pattern over the course of the selection, whereas functionally dispensable nucleotides would vary randomly.

Because selection requires an ability to separate functional from nonfunctional populations of molecules and RNase P is normally recycled during a reaction, this selection experiment utilized a conjugate of pre-tRNA and *E. coli* RNase P RNA that cleaves itself intramolecularly. Cleavage in *cis* produces a shortened RNA molecule that can be separated from uncleaved RNAs by a variety of physical techniques. The pre-tRNA/RNase P RNA construct used in this study, TP292, undergoes self-cleavage with a rate equivalent to the rate of chemistry measured in the native RNase P reaction (Frank et al., 1994). Selection based on self-cleavage should thus be an extremely sensitive means of differentiating between the catalytic activities of variant RNase P RNAs.

RESULTS

Overview of in vitro selection system

RNase P normally cleaves in *trans*, so in vitro selection based on catalytic activity has not been possible previously. Our scheme for selecting RNase P RNA variants in vitro (Fig. 2A) uses the self-cleaving pre-tRNA-RNase P RNA conjugate TP292, which has been described previously (Fig. 1; Frank et al., 1994). In the TP292 configuration, self-cleavage releases the 5′ leader sequence from the conjugate. Self-cleavage absolutely requires divalent metal ions (preferably Mg^{2+}), so the reaction can be initiated by the addition of Mg^{2+} and quenched rapidly with EDTA. Furthermore, the rate of self-cleavage can be modulated by pH, species of monovalent ion, or temperature, thus affording precise control over the stringency of a particular selection experiment.

In this scheme, a pool of DNA encoding variant TP292 RNAs (the library used in this study is described below) is first transcribed in vitro by T7 RNA polymerase in a reaction that includes an excess of the nucleotide analogue 5′ guanosine monophosphorothioate (GMPS). GMPS serves to initiate transcription, thereby producing a population of RNAs with thiol groups incorporated exclusively at their 5′ ends. Modified RNAs are then covalently linked to a solid support (Sulfolink® coupling gel; Pierce) via reaction of the RNA-thiols with iodoacetyl-derivatized agarose. Unreacted RNAs are removed by extensive washing at 50 °C with a high-salt buffer (3 M NaCl or NH_4OAc). Typically, 80–90% of input RNAs are covalently bound to the agarose under these conditions (Table 1). Agarose-linked RNAs are subsequently resuspended in reaction buffer (minus magnesium) and self-cleavage is initiated by adding divalent metal ions. Because the tethered RNAs are covalently coupled to the agarose beads through the leader sequence, active molecules cleave themselves from the solid support, whereas unreacted RNAs remain bound. Unreacted RNAs can then be removed rapidly from solution by centrifugation or filtration. A representative round of selection (round-10, described below) is presented in Figure 2B. Following separation of reactive from unreactive molecules, reverse transcription and PCR are used to synthesize a pool of DNA, complete with a regenerated leader sequence, that can be used for a subsequent round of selection.

A solid-phase strategy was chosen rather than gel-purification of products for two reasons: (1) sample recovery is much faster and more efficient; and (2) the ability to separate precursor from product RNAs by solid-support coupling is independent of the relative lengths of the RNAs; the use of a short (12-nt) leader sequence greatly simplified the process of regenerating precursor genes by PCR.

TABLE 1. Summary of selection.

Round	RNA[a]	pH	Reaction time (s)	Half-lives[b]	% Bound[c]	% Eluted[d]
1	15.0	8	500	2,500	87	0.8
2	10.0	8	500	2,500	89	0.6
3	10.0	8	500	2,500	78	0.5
4	5.3	8	500	2,500	85	0.5
5	7.0	8	500	2,500	90	1.0
6	12.5	8	500	2,500	87	0.6
7	6.4	8	500	2,500	88	1.1
8	12.1	8	500	2,500	84	1.8
9	4.3	8	100	500	82	10.0
10	8.4	7	100	50	80	15.0

[a] Micrograms of RNA applied to agarose solid-support. Round-1 contained three fully representative libraries of randomized sequence-containing RNA.

[b] Estimated number of half-lives a TP292 cleavage reaction would undergo given identical conditions.

[c] Percentage of input RNA bound to agarose after extensive washing with high-salt buffers at 50 °C. Measured by Cerenkov counting of radiolabeled RNA.

[d] Percentage of bound RNA eluted after addition of Mg^{2+} at 50 °C. Measured by Cerenkov counting of radiolabeled RNA. The values reported for rounds 1–6 represent a measure of the background of uncleaved RNA that was recovered in each elution step. The remaining values have not been corrected for this background.

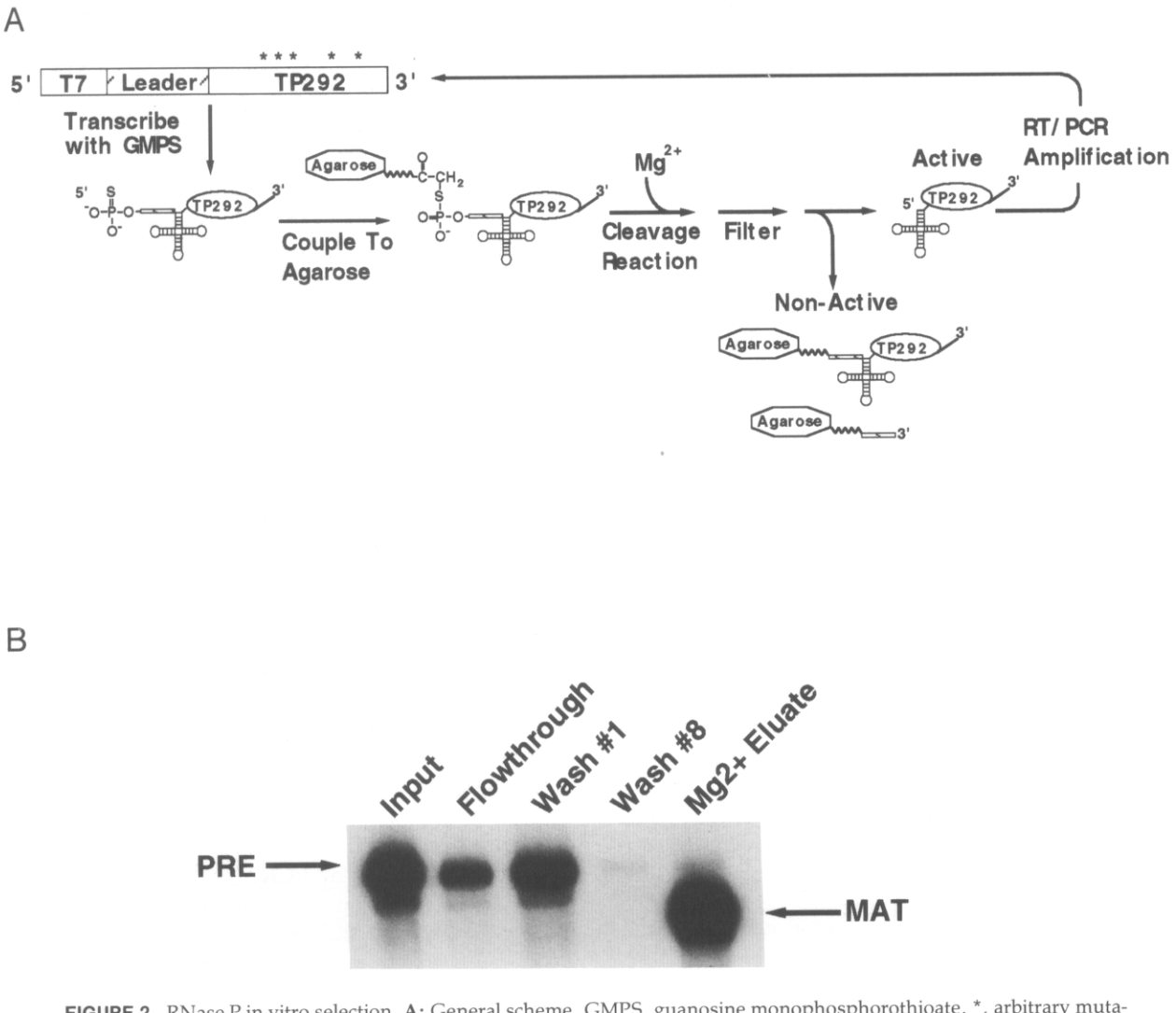


FIGURE 2. RNase P in vitro selection. **A:** General scheme. GMPS, guanosine monophosphorothioate. *, arbitrary mutations in TP292 gene. **B:** Results from round-10 of selection. Internally radiolabeled TP292Δ RNAs recovered from various steps of the selection procedure were separated on a 5% polyacrylamide, 7 M urea gel. Input: RNA applied to agarose-beads (0.05% of material was loaded on the gel). Flowthrough: RNA in solution after binding reaction (4% loaded on gel). Wash #1: RNA eluted by first 3 M NaCl wash (1% loaded on gel). Wash #8: RNA eluted by eighth high-salt wash (4% loaded on gel). Mg2+ Eluate: RNA eluted by addition of 25 mM $MgCl_2$ (1% loaded on gel).

Design and construction of a randomized TP292 library

In order to test the sequence constraints on the J3/4-P4-J2/4 region of RNase P RNA, we randomized 22 nearly invariant positions within this region and selected those variants that are capable of self-cleavage. The locations of the randomized bases as well as the native sequences are shown in Figure 1. The three base pairs that form one end of P4 (72-74/353-355) were not randomized for three reasons: (1) the phylogenetic support for these pairings is quite strong; numerous examples of covariations exist for each pair; (2) covariations between these base pairs and other elements of RNase P, which might indicate tertiary contacts, are not evident; and, importantly, (3) limiting the randomized region to 22 nt (rather than 28 nt) permitted us to construct multiple complete libraries of all possible sequence variants, thus sampling all of the sequence space accessible to the randomized sequence.

To detect and prevent contamination of the selected libraries by native TP292 RNA, three structural variations were introduced into the initial library, creating construct TP292Δ. First, the tRNA D and Anticodon hairpins were deleted in order to inhibit binding of the forward PCR primer to native TP292. This "minihelix" (McClain et al., 1987) variant of TP292 self-cleaves with a reaction rate comparable to native TP292 (data not shown). Second, the distal end of helix P3, including a unique *Ear* I restriction site, was truncated. Native contaminating RNAs could therefore be selected against by *Ear* I digestion following RT/PCR amplification of se-

lected RNAs. Although the effect of this deletion on cleavage activity was not tested in this context, an identical deletion introduced into the native RNase P RNA produced no discernible phenotype under a variety of conditions (Darr et al., 1992b). Third, the artificial loop (Loop 1) connecting the native ends of the RNase P RNA in the circularly permuted RNA was truncated. All three deletions were introduced into TP292Δ during the construction of the randomized library, so a TP292Δ construct bearing native J3/4-P4-J2/4 sequences did not exist prior to this experiment.

The initial, randomized TP292Δ library was constructed by ligation of PCR products synthesized with primers completely randomized at the appropriate positions (Fig. 1; see the Materials and methods for details). In this manner, a randomized library consisting of 39 μg of full-length TP292Δ DNA was constructed initially. For 22 randomized nucleotides within a 450-nt gene, this represents 4.4 complete sequence libraries of TP292Δ DNA; a given variant has a >98% ($1 - e^{-4}$) probability of being present in a sample of this size. The TP292Δ DNA was further amplified by large-scale PCR in order to generate a starting pool for selection experiments. Batch sequencing of this library revealed no bias in the distribution of nucleotides at each position. RNA transcribed from this pool exhibited no detectable self-cleavage activity (data not shown); the randomized region is crucial to RNase P activity.

In vitro selection of functional P4 variants

Functional variants of TP292Δ were selected from the randomized library over 10 cycles of selection (Table 1). Three independent libraries of RNA were selected upon in the first round, giving a >95% probability that a given variant was present in the initial pool of randomized RNA. In each round, self-cleavage reactions were initiated by the addition of 25 mM $MgCl_2$ to agarose bead-coupled RNA (see the Materials and methods). The stringency of selection was controlled by varying both reaction time and pH (the rate of RNase P cleavage is linearly proportional to pH in the range 5–8 [Smith & Pace, 1993]). As an indication of the stringency of selection, the estimated number of half-lives that a native TP292 cleavage reaction would undergo given identical conditions is presented in Table 1.

Self-cleavage of randomized RNA was monitored in each round by measuring the amount of radiolabeled RNA eluted upon addition of Mg^{2+}, by either Cerenkov counting (Table 1; % Bound versus % Eluted) or gel electrophoresis (Fig. 2B). Cleaved product RNAs were first detected above the background of uncleaved RNA recovered in the Mg^{2+} eluate (0.5–1.0% of bound material) following round-7. By round-10, approximately 15% of bound material was eluted specifically upon addition of Mg^{2+} (Table 1; Fig. 2B). Product RNAs from this round were cloned and the sequences of the randomized regions were determined. Surprisingly, the sequences of 33 of 34 clones were identical to the native sequence within the randomized region, whereas the remaining clone, D10.5, differed from native at only one position, A62G (Table 2). All clones contained the three unique TP292Δ deletions, indicating that the population of active RNAs had not been contaminated by exogenous TP292 RNA. Thus, of 1.8×10^{13} possible variants in our starting pool, the native sequence represents the optimum for self-cleavage activity under the in vitro conditions.

Clones from rounds-7 and -8, in which product RNAs were first detected above background, were also sequenced. These variants define two classes of sequence elements (Table 2). The first class, which predominates in the round-7 pool (14/15 clones), comprises sequences with little, if any, similarity to the native sequence; these elements cannot be folded into a structure similar to helix P4. However, by round-8, only 1 of 27 clones fell into this class. The second class of sequences, which predominates the round-8 pool (26/27 clones), consists of sequences that are either identical or closely related to the native sequence (Table 2). Again, all clones encoded the TP292Δ markers, precluding the possibility of contamination by native TP292.

To determine whether any of the clones with non-wild-type sequences were catalytically active, a subset of these clones were transcribed and RNAs were assayed for self-cleavage activity (Table 2). None of these variants exhibited self-cleavage activity, even after extensive incubation under conditions optimal for TP292 self-cleavage. In contrast, the one round-7 clone with a native sequence within the randomized region (7.96) was highly active. Of 61 round-8 and round-10 clones analyzed, only 4 harbored mutations within the randomized region. None of the round-8 variants (clones 8.50, 8.73, and 8.86) were present among the round-10 clones, suggesting that these mutations were disadvantageous relative to the native sequence. Indeed, RNAs transcribed from clones 8.50, 8.73, and 8.86 exhibited little or no detectable self-cleavage activity when assayed individually (Table 2). In contrast, the sole variant identified among the round-10 clones (10.5) was active, as were the six native-structured clones assayed. These results confirm that, by round-7, a small population of active sequences had emerged from a background of nonfunctional molecules. Following the subsequent round of selection (round-8), the active class of sequence elements became fixed within the population. Thus, the wild-type sequence is apparently not only the global optimum, but there are few, if any, sequence variants that have remotely similar activities.

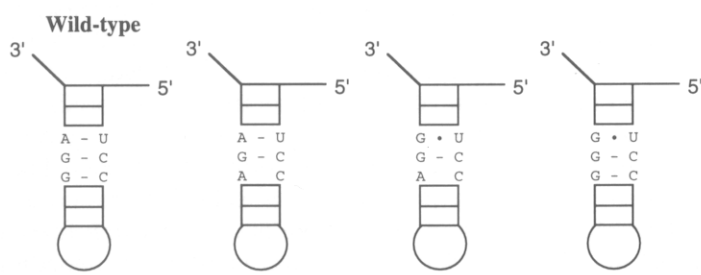
Outside the randomized region, many of the round-7–10 clones contained mutations, which undoubtedly accumulated during RT/PCR amplification. These mutations were found primarily within poorly conserved regions of the RNA and mutations within helices often

TABLE 2. Sequences of randomized J3/4-P4-J2/4 region.

Round	Isolates[a]	Sequence[b]	Clone	Activity[c]
Native	na	A G G A A A G U C C G G G.....A C A G A A C C C G G C U U A		
0	na	N N N N N N N N N N G G G.....N N N N N N C C C N N N N N N		
10	33	A G G A A A G U C C G G G.....A C A G A A C C C G G C U U A		59–89%[d]
	1	g G G A A A G U C C G G G.....A C A G A A C C C G G C U U A	10.5	69
8	24	A G G A A A G U C C G G G.....A C A G A A C C C G G C U U A		nd
	1	A G G A A A G U C C G G G.....A A A G A A C C C G G C U U A	8.50	12
	1	u G G A A A a U C C G G G.....A C U G A A C C C G G C U U A	8.73	–
	1	u a c u A A a c g a G G G.....- C C C A C C C C A A U A A U	8.86	–
7	1	g u u c c A c g C a G G G.....C U C A A A C C C A C A C C G	7.A	–
	1	u u u g g C a U u u G G G.....G U A C C G C C C G C A C C A	7.B	–
	1	c u G u u A G c a C G G G.....A U C A U U C C C C C A C A C	7.C	–
	1	g G u A A G a U a g G G G.....C A G U C C C C C C U C A U C C	7.D	nd
	1	u u u A g g c c u g G G G.....A U U C C C A C C C U U U C U -	7.26	nd
	1	c c a g A A c c C u G G G.....U U U C U C C C C C U U U C C	7.59	nd
	1	A G G A A A G U C C G G G.....G C C C U A C C C A C C A U A	7.61	–
	1	g u G u A c c U C a G G G.....A G U U C G C C C A U U C U C	7.89	nd
	1	u c u A g u a a C u G G G.....- U G C A U C C C C U A G A A C	7.91	–
	1	A c c A u c G c g a G G G.....C G A U A A C C C A A C U U G	7.92	nd
	1	A G G A A A G U C C G G G.....A C A G A A C C C G G C U U A	7.96	83
	1	c G G g A u a a C a G G G.....C C C C U A C C C A C U U C G	7.97	nd
	1	- c u c A A a U g C G G G.....C A C C U C C C C C C U C U - -	7.99	nd
	1	u u G c g u c U a g G G G.....C C C G U C C C C C U A A A C C	7.100	nd
	1	c c c A c u u g a u G G G.....C A A C U G C C C C G U G U A C	7.104	nd

[a] Number of independent clones sequenced.
[b] Nucleotides that differ from the native sequence are in lowercase. N, randomized position. Nucleotides that comprise helix P4 are underlined in the first row.
[c] Activity was measured as the percentage of precursor that self-cleaved after 10 min under standard reaction conditions, pH 8.0. –, no detectable activity; nd, not determined.
[d] Six clones were assayed: 10.1, 10.2, 10.3, 10.8, 10.11, and 10.12.

maintained base pairing (e.g., A-U ↔ G·U ↔ G-C). For the most part, no trend could be discerned in the identities of these mutations. However, an intriguing pattern of mutations was identified within the 3′ half of the minihelix acceptor stem, specifically involving two normally base paired nucleotides. The sequence identities of the nucleotides found within these base pairs in the various clone pools are summarized in Figure 3. Rather than the native sequence, 5′GGA, most round-8 and round-10 clones contained either the sequence 5′AGA (7/55) or 5′AGG (45/55). The one example of 5′AGG found in round-7 was in the clone (7.96) that encoded native sequence within the randomized region (i.e., the one catalytically active clone obtained in this round). Conversely, one of the 5′GGA encoding round-8 clones (8.86) contained a randomized helix P4 region. Thus, the



Round				
7	12	0	1	1
8	2	2	22	0
10	1	5	23	0

FIGURE 3. Sequence motifs identified within the acceptor stem of the pre-tRNA substrate. The number of independent clones with each sequence is summarized for rounds-7, -8, and -10.

vast majority of RNAs that were selected in rounds-8 and -10 (i.e., functional RNAs encoding native sequences in the randomized regions) carried mutations within the acceptor stem of the minihelix substrate. Because the sequence of the 5′ half of the minihelix acceptor stem was held constant by the forward PCR primer, the mutations in the 3′ portion of the stem must introduce mismatches into this helix. Structurally, the C·A and U·G appositions formed by these mutations should disrupt the normal A-form geometry of the acceptor stem (Wyatt & Tinoco, 1993). These perturbed helices may function as better substrates under the conditions of this selection, thus conferring a selective advantage over the native structure. Indeed, introduction of a bulge into a minihelix facilitates its recognition by both EF-Tu and *Xenopus laevis* RNase P (Carrara et al., 1995; Nazarenko & Uhlenbeck, 1995).

DISCUSSION

Helix P4 of the RNase P RNA, along with its flanking "single-stranded" sequences, J3/4 and J2/4, is nearly invariant throughout the phylogenetic domain of Bacteria (Fig. 1; Brown et al., 1994; Haas et al., 1994; Pace & Brown, 1995). Archaeal and eucaryal RNase P RNAs also contain P4 homologues, despite differing substantially from bacterial ribozymes in other structural features (Forster & Altman, 1990; Nieuwlandt et al., 1991; LaGrandeur et al., 1993; Tranguch & Engelke, 1993). The conservation of this structure across all three phylogenetic domains of life, together with the results of biochemical (Nolan et al., 1993; Harris et al., 1994; Hardt et al., 1995; Nazarenko & Uhlenbeck, 1995; Pan, 1995) and mutational (M.A.T. Rubio & N.R. Pace, unpubl.) analyses, clearly identify the P4 region as the center of RNase P structure and function.

In order to determine whether the identities of any of the J3/4-P4-J2/4 nucleotides are critical for catalytic activity, as opposed to playing some other role in the function of the RNase P RNA, such as in biosynthesis, stability, or protein binding, we have subjected a randomized pool of RNase P RNAs to in vitro selection. Twenty-two nucleotides were fully randomized within the J3/4-P4-J2/4 region of construct TP292Δ, a self-cleaving pre-tRNA-RNase P RNA conjugate, and catalytically active molecules were isolated by Mg^{2+} elution of RNAs from agarose beads. Active molecules were first detected above a background of inactive RNAs that was invariably present in the Mg^{2+} eluate (ca. 0.5–1.0% of bound material; Table 1) following selection round-7. Following the subsequent round, the majority of material isolated in the elution step consisted of catalytically active molecules. The rapidity with which the active population of RNAs progressed from constituting a minority to the majority of molecules in the overall population (i.e., in the course of one round of selection) is indicative that a strong selective pressure for catalytic activity operated during this experiment. That at least eight rounds of selection were required for the active population of molecules to become fixed, despite the strong selective pressure, further suggests that the founding population of functional variants was quite small.

Surprisingly, we found that within the 22-nt randomized J3/4-P4-J2/4 region, the native sequence was overwhelmingly preferred. Recovery of primarily wild-type variants clearly raises the possibility that exogenous, native TP292 molecules contaminated the experimental population during the course of the selection. Although contaminants cannot be ruled out conclusively, several precautions were taken to minimize their occurrence. First, all reagents were prepared fresh, sterilized, and used exclusively in this experiment. Control PCR reactions were conducted following each round of selection to determine whether the buffers and equipment (e.g., pipettors, eppendorf tubes, etc.) used in RT-PCR were contaminated with TP292 DNA; these tests were invariably negative. Second, the randomized TP292Δ construct was designed in such a way that contaminating molecules could be detected and eliminated. As outlined above, three deletions were introduced into the randomized TP292Δ library during its construction. One of these deletions (e.g., truncation of helix P3) removed a unique *Ear* I restriction site from the construct, thus permitting a direct selection against wild-type contaminating RNase P genes. Consequently, the population of TP292Δ genes was subjected to *Ear* I restriction endonuclease digestion and gel-purification following each round of selection.

Because all of the variants sequenced possessed the three structural markers introduced during construction of the randomized TP292Δ pool, entry of contaminating native sequences would have required recombination with members of the randomized population. However, because replacement of the P3 and Loop 1 sequences would require the occurrence of two precise recombination events, each occurring within a window of less than 10 nt (i.e., the lengths of the sequences separating the randomized P4 regions from the sequences unique to TP292Δ), we feel that the probability of such a scenario is vanishingly small. Even if this were the route by which the native J3/4-P4-J2/4 sequences were introduced into the randomized population, our basic conclusion still stands: the vast majority of the sequence variants present in the randomized population are poor substitutes for the native structure, and consequently, could not compete against what was undoubtedly a small population of active, founder molecules. Finally, many of the active variants contained several distinct point mutations outside the randomized region, indicating that these RNAs had undergone many rounds of selection. In contrast, any contaminating, native-like TP292 RNA would have overtaken the population in

only a few generations, precluding substantial mutational drift (D.N. Frank, unpubl.).

The clear implication of recovering primarily native J3/4-P4-J2/4 sequences in this experiment is that each nucleotide within the randomized region functions in some manner during catalysis and that optimal catalytic activity obtains only when all of these nucleotides assume their native identities: there is only one general solution to the problem of J3/4-P4-J2/4 structure and function. We cannot distinguish direct catalytic roles from indirect structural roles for these nucleotides, of course. Although it is possible that optimality of the native J3/4-P4-J2/4 sequence is the result of idiosyncratic interactions with other portions of the RNase P holoenzyme, several lines of evidence argue that this region is instead optimized for an intrinsic, catalytic function. If the J3/4-P4-J2/4 structure were fixed for historical reasons (e.g., constrained by interactions with the RNase P protein or other domains in the RNA), then removing or lessening these constraints would be expected to invoke compensatory alterations in the P4 region. For instance, bases participating in tertiary interactions would co-vary to maintain complementarity. It is likely that the constraints placed on J3/4-P4-J2/4 were relaxed in this selection, in at least three ways: (1) selection was performed on a modified species of the RNA (i.e., a ribozyme–substrate conjugate harboring three deletions); (2) the RNA, rather than the holoenzyme, was subjected to selection, so RNA–protein interactions essential for cellular activity were not demanded; and (3) RT/PCR inexorably introduces random mutations, which allow further exploration of sequence space. Evidence that the constraints operating on RNase P structure and function were indeed relaxed in this experiment was observed in the case of the tRNA "minihelix" portion of the TP292Δ tether, in which novel structures, catalytically superior to the original one, were evolved over the course of the experiment (Fig. 3). However, despite the fact that numerous point mutations accumulated in the selected RNAs outside of the randomized region, the native J3/4-P4-J2/4 sequence was still recovered in the selection, even in the absence of the protein subunit. Catalytically superior variants of J3/4-P4-J2/4 did *not* arise, despite the relaxed constraints, suggesting that this region is itself optimized for function. This result is in striking contrast to those observed with other large catalytic RNAs, such as Group I self-splicing introns, which readily evolve improved catalytic function under the conditions imposed by selection in vitro (Chapman & Szostak, 1994; Joyce, 1994; Kumar & Ellington, 1995).

The phylogenetic record further supports the contention that J3/4-P4-J2/4 is optimized functionally. The J3/4-P4-J2/4 region is the most highly conserved domain of RNase P RNA (Haas et al., 1994). It includes most of the invariant nucleotides in the RNA: clear homologues of this structure, nearly identical in sequence, occur in the RNase P RNAs of Archaea and Eucarya as well (Forster & Altman, 1990; Nieuwlandt et al., 1991; LaGrandeur et al., 1993; Tranguch & Engelke, 1993). In contrast to the basic J3/4-P4-J2/4 structure, the remainder of the RNase P RNA is much more volatile, both in sequence and structure. If J3/4-P4-J2/4 structure were constrained solely by interactions with the rest of RNase P, rather than being optimized for function per se, then other domains within the RNA might be expected to be as well-conserved as the P4 region.

Taken together, our experimental results and the phylogenetic evidence indicate that J3/4-P4-J2/4 has evolved optimal catalytic function. Modification–interference experiments indicate that the J3/4-P4-J2/4 region is involved in binding magnesium ions involved in the hydrolysis of pre-tRNA (Hardt et al., 1995; Harris & Pace, 1995). We therefore propose that it is the metal-binding capacity of the J3/4-P4-J2/4 structure that has been optimized functionally. Given the phylogenetic record, it is likely that this structure achieved evolutionary optimization prior to the ancient radiation of the Bacteria from the Archaea and Eucarya, prior to the origin of life as we know it. It is tempting to speculate that at least this portion of the RNase P ribozyme is truly a remnant of the RNA world.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotides were obtained from either the Indiana University Institute for Molecular and Cellular Biology or Amgen (California) and purified by acrylamide gel electrophoresis. The following oligonucleotides were used (N represents a completely randomized position):

P4B: 5′GCGTGTCGACTGACCGANNNNNGGGN
NNNNNCGTGGACAGTCATTCATC
P4ASal: 5′GGCCGCGGGTCGACAGCTGACCAGACAG
TCGCCGCTTCGTTCGGCGGAGGGGNNNN
NNNNNNGGGCTCCATAGGGCAG
T7MH: 5′GGGAATTCTAATACGACTCACTATAGGTAC
CCAAAATGGTCCGGGCGGGTTC
CP292R: 5′CGGGTTCAGTACGGGCCGT
TP292R: 5′GAAGATCTCGGGTTCAGTACGGGCCGT
142R: 5′TGCGCGGGCCATCGGCGG

Construction of TP292Δ DNA library

The initial randomized library of TP292Δ DNA was constructed by amplifying the gene in two pieces by PCR (Mullis & Faloona, 1987) as modified by Reysenbach et al. (1992). The 5′ half of the library was amplified with the primers T7MH (forward) and P4B (reverse), whereas the 3′ half was amplified with P4ASal (forward) and CP292R (reverse). Amplification products were extracted with phenol–chloroform and chloroform, ethanol precipitated, digested overnight with the restriction enzyme *Sal* I, and isopropanol-precipitated to remove small oligonucleotides. The two samples were then

mixed and ligated together overnight at 16 °C. The products of the ligation reaction were separated on a preparative 2% agarose gel and full-length DNA was excised and passively eluted overnight into a solution of 5 mM Tris, pH 7.4, 0.5 mM EDTA, 0.05% SDS. The eluate was lyophilized to 400 μL and the DNA ethanol precipitated, resuspended in TE buffer (10 mM Tris, pH 7.4, 1 mM EDTA) and quantitated by measuring its absorbance (A_{260}). A total of 39 μg of TP292Δ DNA, representing 4.4 complete libraries, was synthesized by this means. The entire sample was then amplified to 240 μg by 15 cycles of PCR using oligonucleotides T7MH and CP292R.

RNA synthesis and purification

RNAs were synthesized in vitro by run-off transcription using T7 RNA polymerase (gift of B. Pace). The T7 transcription protocol of Milligan and Uhlenbeck (1989) was modified as follows: (1) reactions were performed in 7.5% glycerol; (2) reactions were incubated overnight at 12 °C, rather than 37 °C, to minimize self-cleavage; and (3) 7.5 mM GMPS was added to the reaction (7.5::1 ratio of GMPS/GTP). GMPS was synthesized by the protocol of Burgin and Pace (1990). RNAs were radiolabeled internally by adding 40–50 μCi of [α-^{32}P]GTP to the transcription reactions. Transcription products were purified by electrophoresis on 5% polyacrylamide/7 M urea gels, excised by UV shadowing, and passively eluted (overnight, 4 °C) into a buffer consisting of 0.3 M sodium acetate, 10 mM Tris, pH 7.4, 1 mM EDTA, and 0.5% SDS. Eluted samples were ethanol precipitated, resuspended in TE buffer, and quantitated by specific radioactivity.

RNA selection

In each round of selection, a 150-μL slurry of Sulfolink® gel (Pierce) was washed three times in 250–300 μL of buffer BP8.9 (40% methanol, 20 mM sodium phosphate, pH 8.9, 0.1% SDS, 5 mM EDTA; Burgin and Pace [1990]) and then incubated with 16 μg of *Bacillus subtilis* total rRNA (in 250–300 μL BP8.9) at 25 °C. After 45 min, the agarose beads were pelleted by microcentrifugation and resuspended in 400 μL BP8.9 plus GMPS-initiated RNA (4–15 μg). Before being applied to the agarose beads, the RNA samples were first refolded by incubation in BP8.9 at 65 °C for 3 min, followed by 50 °C for 1 min. RNA and beads were mixed in batch at room temperature for 1–2 h. Unreacted RNA was removed by four washes with 3 M NaCl, 50 mM Tris, pH 8, 5 mM EDTA (15 bed-volumes total) and four washes with RNase P reaction buffer lacking magnesium (3 M NH_4OAc, 44.5 mM Tris, 16 mM Pipes, 0.05% SDS, pH 8.0 or 7.0, as indicated in Table 1; 16 bed-volumes total). In each wash step, the RNA/bead slurry was incubated for 30–60 s at 50 °C in the appropriate buffer, then pelleted by brief microcentrifugation; eluates were removed and quantitated as described below. Following the last wash, the RNA and beads were resuspended in 225 μL of RNase P reaction buffer lacking magnesium, and equilibrated at 50 °C for 1 min. Reactions were initiated by adding 9.3 μL of 1 M $MgCl_2$ (25 mM final) to the RNA and quenched by adding 30 μL of 0.5 M EDTA on ice. Beads were removed from the supernatant by filtration through 0.22-μM spin-filters (Amicon). RNA was precipitated from the supernatant and wash fractions by microcentrifugation following the addition of 5 μg glycogen, 5 μL 0.5 M EDTA, and 3 volumes of ice-cold ethanol; RNA pellets were resuspended in 20 μL TE buffer.

The amount of RNA applied to the beads and removed by washing or Mg^{2+} elution was monitored by both Cerenkov counting and gel electrophoresis of radiolabeled samples. The percentage of RNA bound in each round of selection was calculated by subtracting the total amount of RNA released by the wash steps from that initially applied to the beads. The percentage of RNA eluted was calculated as the amount of RNA eluted by addition of Mg^{2+} divided by the amount of RNA that remained on the agarose beads after the last wash step.

Reverse transcription/PCR amplification of RNAs

Reverse transcription followed the protocol of Burgin and Pace (1990) using the reverse primer CP292R. A portion of the cDNA was then amplified by PCR with the oligonucleotides T7MH and CP292R (oligo TP292R was used to amplify DNA for cloning). Following each round of selection, the resultant PCR products were digested with the restriction enzyme *Ear* I for 2 h at 37 °C. Digestion products were resolved on 1.8% agarose/TBE gels, excised, eluted into TE (6–18 h, 25 °C), filtered through 0.22-μM spin-filters (Amicon), and ethanol precipitated.

Cloning and sequence analysis

PCR products were digested with restriction enzymes *Eco*R I and *Bgl* II for 2 h at 37 °C, whereas plasmid p153Bsttan (gift of J. Nolan, Tulane Univ.) was digested with *Eco*R I and *Bam*H I. Digested DNAs were separated on 1.5% agarose/TBE gels and the appropriate bands excised, eluted into TE (overnight, 25 °C), filtered through 0.22-μM spin-filters (Amicon), and ethanol precipitated. Ligation and transformation of inserts and vectors followed standard procedures (Sambrook et al., 1989). Positive clones were identified by restriction analysis and sequenced by the dideoxy method (using the Sequenase 2.0 enzyme and protocol from U.S. Biochemical; Sanger et al., 1977) using the oligonucleotide 142R. Sequences were aligned manually with the program SeqApp (Don Gilbert, Indiana University).

Activity assays

RNAs were transcribed from templates generated by PCR amplification of individual TP292Δ clones and purified as described above. The RNA samples were refolded immediately prior to the activity assays by heating in reaction buffer minus magnesium (3 M NH_4OAc, 44.5 mM Tris, 16 mM Pipes, 0.05% SDS, pH 8.0) for 5 min at 65 °C, followed by 2 min at 50 °C, the temperature at which the assays were performed. Reactions were initiated by adding $MgCl_2$ to 25 mM and quenched after 10 min incubation by adding EDTA to 50 mM along with 3 volumes of ice-cold ethanol. Following precipitation, precursor and product RNAs were separated by electrophoresis on 5% denaturing polyacrylamide gels and the extents of reaction quantitated by phosphorimaging (Molecular Dynamics). The TP292Δ RNA concentration in all self-cleavage reactions was 3 nM.

ACKNOWLEDGMENTS

We thank Dr. B. Pace for the gift of T7 RNA polymerase and *Thermus aquaticus* DNA polymerase, Dr. M. Harris for guidance in synthesizing GMPS, Dr. J. Nolan for plasmid p153Bsttan, Drs. E. Haas and J. Brown for providing the Book of P, and members of the Ellington and Pace laboratories for useful discussions. This work was supported by American Cancer Society Postdoctoral Fellowship PF-3933 (D.N.F.) and NIH grant GM 34527 (N.R.P.).

Received July 16, 1996; returned for revision August 20, 1996; revised manuscript received September 30, 1996

REFERENCES

Altman S, Kirsebom L, Talbot S. 1993. Recent studies of ribonuclease P. *FASEB J 7*:7–14.

Brown JW, Haas ES, Gilbert DG, Pace NR. 1994. The ribonuclease P database. *Nucleic Acids Res 22*:3660–3662.

Brown JW, Nolan JM, Haas ES, Rubio MAT, Major F, Pace NR. 1996. Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc Natl Acad Sci USA 93*:3001–3006.

Burgin AB, Pace NR. 1990. Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. *EMBO J 9*:4111–4118.

Carrara G, Calandra P, Fruscoloni P, Tocchini-Valentini GP. 1995. Two helices plus a linker: A small model substrate for eukaryotic RNase P. *Proc Natl Acad Sci USA 92*:2627–2631.

Chapman KB, Szostak JW. 1994. In vitro selection of catalytic RNAs. *Curr Opin Struct Biol 4*:618–622.

Darr SC, Brown JW, Pace NR. 1992a. The varieties of ribonuclease P. *Trends Biochem Sci 17*:178–182.

Darr SC, Zito K, Smith D, Pace NR. 1992b. Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P. *Biochemistry 31*:328–333.

Forster AC, Altman S. 1990. Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. *Cell 62*:407–409.

Frank DN, Harris ME, Pace NR. 1994. Rational design of self-cleaving pre-tRNA ribonuclease P RNA conjugates. *Biochemistry 33*:10800–10808.

Gold L, Tuerk C, Allen P, Binkley J, Brown D, Green L, MacDougal S, Schneider D, Tasset D, Eddy SR. 1993. RNA: The shape of things to come. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 497–510.

Haas ES, Brown JW, Pitulle C, Pace NR. 1994. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc Natl Acad Sci USA 91*:2527–2531.

Hardt WD, Warnecke JM, Erdmann VA, Hartmann RK. 1995. Rp-phosphorothioate modifications in RNase P RNA that interfere with tRNA binding. *EMBO J 14*:2935–2944.

Harris ME, Nolan JM, Malhotra A, Brown JW, Harvey SC, Pace NR. 1994. Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J 13*:3953–3963.

Harris ME, Pace NR. 1995. Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA 1*:210–218.

Joyce GF. 1994. In vitro evolution of nucleic acids. *Curr Opin Struct Biol 4*:331–336.

Kirsebom LA. 1995. RNase P—A scarlet pimpernel. *Mol Microbiol 17*:411–420.

Kumar PKR, Ellington AE. 1995. Artificial evolution and natural ribozymes. *FASEB J 9*:1183–1195.

LaGrandeur TE, Darr SC, Haas ES, Pace NR. 1993. Characterization of the RNase P RNA of *Sulfolobus acidocaldarius*. *J Bacteriol 175*: 5043–5048.

McClain WH, Guerrier-Takada C, Altman S. 1987. Model substrates for an RNA enzyme. *Science 238*:527–530.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol 164*:51–62.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a-polymerase-catalyzed chain reaction. *Methods Enzymol 155*:335–350.

Nazarenko IA, Uhlenbeck OC. 1995. Defining a smaller RNA substrate for elongation factor Tu. *Biochemistry 34*:2545–2552.

Nieuwlandt DT, Haas ES, Daniels CJ. 1991. The RNA component of RNase P from the archaebacterium *Haloferax volcanii*. *J Biol Chem 266*:5689–5695.

Nolan JM, Burke DH, Pace NR. 1993. Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science 261*:762–765.

Pace NR, Brown JW. 1995. Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. *J Bacteriol 177*:1919–1928.

Pan T. 1995. Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry 34*:902–909.

Reysenbach AL, Giver LJ, Wickham GS, Pace NR. 1992. Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol 58*:3417–3418.

Sambrook JE, Fritsch F, Maniatis T. 1989. *Molecular cloning: A laboratory manual, 2nd ed*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Smith D, Pace NR. 1993. Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry 32*:5273–5281.

Szostak JW, Ellington AD. 1993. In vitro selection of functional RNA sequences. In: Gesteland RF, Atkins JF eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 511–534.

Tranguch AJ, Engelke DR. 1993. Comparative structural analysis of nuclear RNase P RNAs from yeast. *J Biol Chem 268*:14045–14055.

Westhof E, Altman S. 1994. Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli*. *Proc Natl Acad Sci USA 91*:5133–5137.

Wyatt JR, Tinoco I. 1993. RNA structural elements and RNA function. In: Gesteland RF, Atkins JF eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 465–496.