

# Analysis of the Functional Role of Conserved Residues in the Protein Subunit of Ribonuclease P from *Escherichia coli*

Venkat Gopalan<sup>1</sup>, Andreas D. Baxevanis<sup>2</sup>, David Landsman<sup>2</sup> and Sidney Altman<sup>1\*</sup>

<sup>1</sup>Department of Biology, Yale University, New Haven, CT 06520-8103, USA

<sup>2</sup>National Center for Biotechnology Information National Library of Medicine National Institutes of Health Room 8N-805, Building 38A Bethesda, MD 20894, USA

The processing of precursor tRNAs and some other small cellular RNAs by M1 RNA, the catalytic subunit of *Escherichia coli* ribonuclease P, is accelerated by C5 protein (the protein cofactor) both *in vitro* and *in vivo*. In an effort to understand the mechanism by which the protein cofactor promotes and stabilizes certain conformations of M1 RNA that are most efficient for RNase P catalysis, we have used site-directed mutagenesis to generate mutant derivatives of C5 protein and assessed their ability to promote RNase P catalysis *in vivo* and *in vitro*. Our results indicate that certain conserved hydrophobic and basic residues in C5 protein are important for its function and that single amino acid residue changes in C5 protein can alter the substrate specificity of the RNase P holoenzyme.

© 1997 Academic Press Limited

\*Corresponding author

Keywords: RNase P protein subunit; site-directed mutagenesis

## Introduction

Ribonuclease P (RNase P) is an endoribonuclease that cleaves the 5'-terminal leader sequences of precursor tRNAs (ptRNAs; Figure 1; Altman *et al.*, 1995; Pace & Brown, 1995). In addition to ptRNAs, RNase P from *Escherichia coli* can cleave other endogenous substrates such as the precursors to 4.5S RNA and 10Sa RNA (Bothwell *et al.*, 1976; Komine *et al.*, 1994). RNase P (in conjunction with RNase E) is also involved in the processing of the polycistronic mRNA of the histidine operon in *Salmonella typhimurium* (Alifano *et al.*, 1994). The RNase P holoenzyme of *E. coli* consists of a catalytic RNA subunit (M1 RNA, 377 nucleotides) and a protein subunit (C5, 119 amino acid residues). Under certain conditions *in vitro*, M1 RNA can catalyze the hydrolysis of ptRNAs even in the absence of C5 protein; however, both M1 RNA and C5 protein are essential for the activity of RNase P *in vivo*. In contrast to the RNA subunits of RNase P from Bacteria, those of RNase P from Archaea and Eukarya

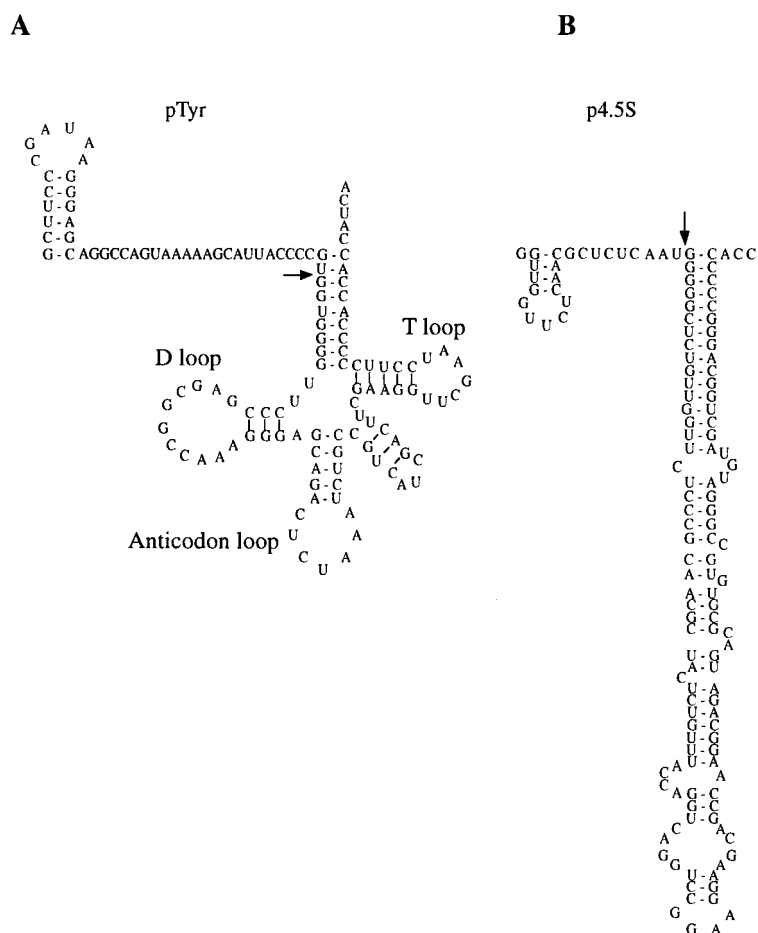
fail to exhibit catalytic activity *in vitro* in the absence of their protein subunits.

Studies performed with different substrates and M1 RNA, in either the presence or absence of C5 protein, have revealed that M1 RNA is a more efficient and versatile enzyme in the presence of the protein cofactor (Guerrier-Takada *et al.*, 1983; Lumelsky & Altman, 1988; Reich *et al.*, 1988; Peck-Miller & Altman, 1991). Determination of kinetic parameters for the hydrolysis of numerous substrates by M1 RNA alone and the RNase P holoenzyme has revealed that M1 RNA (by itself) can achieve the most efficient conformation for recognition of some substrates, while for some others it requires the C5 protein (Kirsebom & Altman, 1989; Peck-Miller & Altman, 1991; Kirsebom & Svård, 1992). Furthermore, the presence of C5 protein can alleviate the deleterious effect of mutations on the activity of M1 RNA in different parts of the M1 RNA molecule (Lumelsky & Altman, 1988). Therefore, C5 protein must engage in specific interactions with the catalytic RNA subunit to stabilize certain conformations of M1 RNA favorable for catalysis and thus play a critical role in recognition/binding of some substrates by the RNase P holoenzyme.

The amino acid residues in C5 protein responsible for the various effects on M1 RNA catalysis have not been identified. Here we have used site-directed mutagenesis to examine the role of conserved residues in C5 protein with regard to RNase P activity *in vivo* and *in vitro*.

Present address: A. D. Baxevanis, Computational Genomics Core, Genome Technology Branch, National Center for Genome Research, National Institutes of Health, Bethesda, MD 20892, USA.

Abbreviations used: RNase P, ribonuclease P; ptRNAs, precursor tRNAs; ts, temperature sensitive; IPTG, isopropyl- $\beta$ -D-thiogalactopyranoside; p4.5S, precursor to 4.5S RNA; ss, single-stranded.



**Figure 1.** Secondary structure of ptRNA<sup>Tyr</sup> and p4.5S RNA, substrates of *E. coli* RNase P. The site of RNase P-mediated cleavage in these substrates is indicated by arrowheads.

## Results

### Rationale

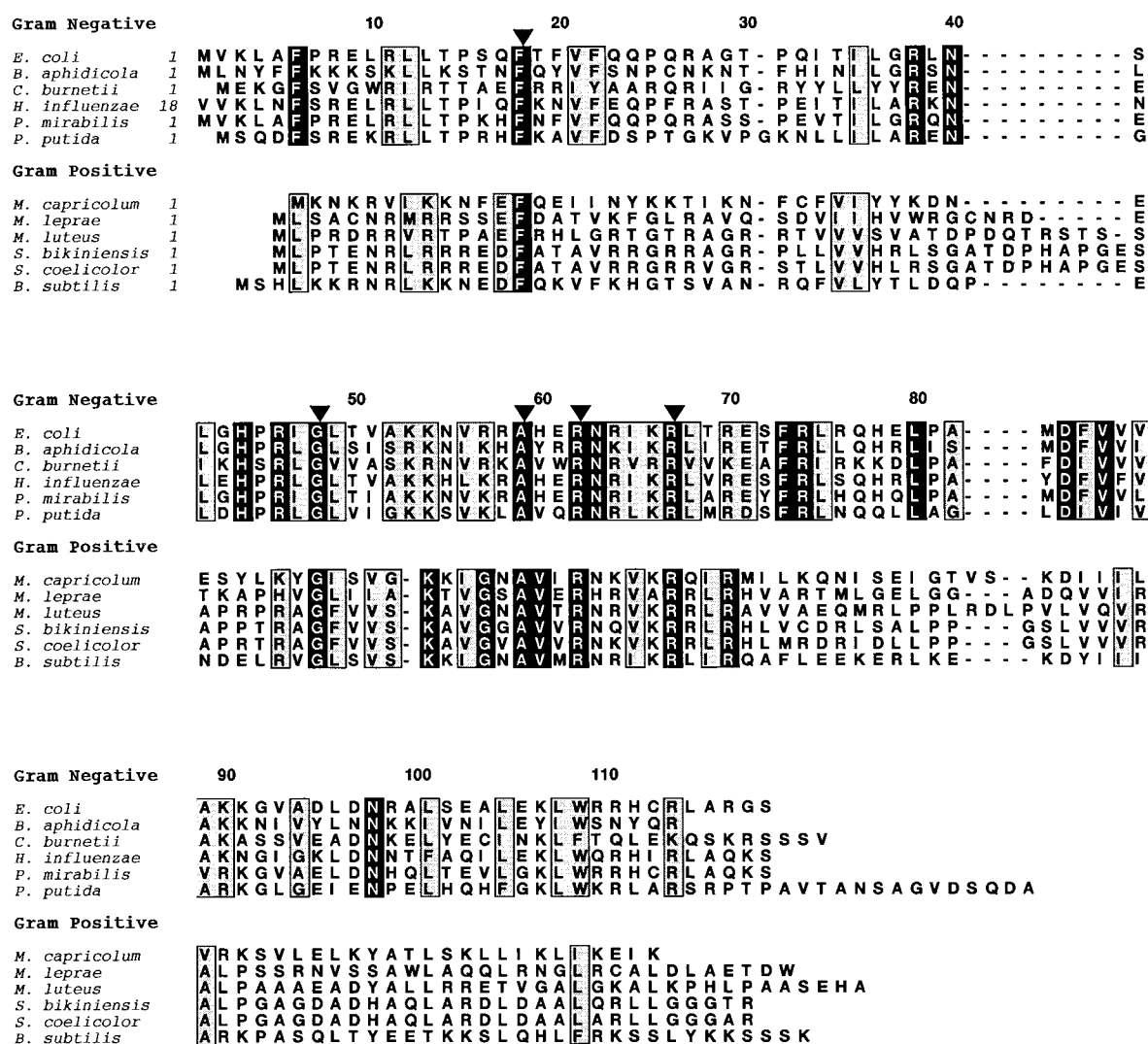
There are a few conserved basic and hydrophobic residues which contribute to the limited identity observed among different prokaryotic RNase P protein subunits (Figure 2). However, there are several positions in the polypeptide chain at which the physicochemical properties of the residues are similar. None of the consensus RNA-binding motifs (Burd & Dreyfuss, 1994) identified in other RNA-binding proteins is present in C5 protein. However, since arginine-rich sequences in bacteriophage, viral, and ribosomal proteins are thought to mediate RNA recognition (Lazinski *et al.*, 1989; Burd & Dreyfus, 1994), it is noteworthy that there is a stretch of ten residues in C5 protein (from position 60 to 70) that is rich in arginine and lysine residues.

The affinity of C5 protein for M1 RNA increases 500-fold as the ionic strength is increased from 0.1 to 1.0 M NH<sub>4</sub>Cl (Talbot & Altman, 1994). The salt dependence of the C5 protein-M1 RNA interaction suggests that hydrophobic interactions play a role in holoenzyme formation. Also, recent crystallo-

graphic studies provide evidence for aromatic amino acid residues in RNA-binding proteins stacking on RNA bases in their RNA ligands (Oubridge *et al.*, 1994; Nagai, 1996). Therefore, in this study, emphasis has been placed on altering aromatic residues in addition to certain basic residues in C5 protein.

### Genetic complementation

The mutation R46H in the chromosomal gene encoding C5 protein results in a temperature-sensitive (ts) phenotype in *E. coli* (Schedl & Primakoff, 1973; Kirsebom *et al.*, 1988). Recently, an *E. coli* strain, T7A49, which contains both the *rnpA49* mutation (i.e. C5 R46H) and the T7 RNA polymerase gene in its chromosome was constructed (Guerrier-Takada *et al.*, 1995). Transformation of these cells with a plasmid bearing the gene encoding wild-type C5 protein can rescue this mutation and abolish the ts phenotype. The various mutant derivatives of C5 protein (subcloned in the same vector as the wild-type C5 protein and under control of the T7 RNA polymerase promoter) were analyzed for their ability to complement the R46H mutation in T7A49 cells. In  $\lambda$  DE3 lysogens (such as T7A49



**Figure 2.** Multiple alignment of sequences of protein subunits of RNase P from Bacteria using Clustal W (Thompson *et al.*, 1994). Sequences are grouped as Gram negative and Gram positive. The numbering scheme at the top of the Figure is based on the sequence of the protein subunit of *E. coli* RNase P, and the numbers preceding each sequence indicate the position of the first shown residue for that sequence. Identical and conserved regions were calculated for each group. Vertical arrowheads identify positions in the alignment where the residues for all the sequences are identical. Positions with absolute conservation have a dark background, while those positions showing only conservative substitutions have a gray background. The sequences of the protein subunits of RNase P from *Escherichia coli*, *Buchnera aphidicola*, *Coxiella burnetii*, *Haemophilus influenzae*, *Proteus mirabilis*, *Pseudomonas putida*, *Mycoplasma capricolum*, *Mycobacterium leprae*, *Micrococcus luteus*, *Streptomyces bikiniensis*, *Streptomyces coelicolor*, and *Bacillus subtilis* were used in this alignment.

cells) the T7 RNA polymerase gene is under the control of an isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG)-inducible *lac* UV5 promoter, which allows a low level of transcription even in the uninduced state. Therefore, our complementation analyses were performed in the absence of IPTG.

The results of our complementation analyses are summarized in Table 1. In general, altering the identity of the conserved residues exerts a moderate to severe effect on the activity of the protein *in vivo*. While substitution of some of the conserved hydrophobic residues individually with Ala had only a moderate effect on their function *in vivo*,

changing two hydrophobic residues simultaneously to Ala resulted in a very drastic effect. For example, the single mutants C5 F22A, C5 F73A, and C5 W109A can partially rescue the ts phenotype of the T7A49 cells while the double mutants C5 F22A/W109A and C5 F73A/W109A fail to exhibit any activity *in vivo* (Table 1). The alteration of positively charged residues (such as R62A or R67A) had a moderate effect on the ability of C5 protein to rescue the ts phenotype of T7A49 cells. Here we have not constructed any double mutants in which two basic residues were altered simultaneously. Also, the mutants C5 R57P and C5

**Table 1.** Results of RNase P assays using C5 protein or its mutant derivatives

Protein used to reconstitute with M1 RNA	ptRNA <sup>Tyr</sup>		p4.5S RNA	T7A49
	Relative initial velocity (%) 30°C	43°C	Relative activity (%) 43°C	Complementation <sup>a</sup> 43°C
Wild-type C5	100	100	100	++
F18A	36	34	2	ND
F22A	42	41	<1	+
F73A	76	72	64	+
W109A	76	69	44	+
F18W, W109F	64	87	48	+
F18A, W109A	<1	<1	<1	-
F18A, F22A	7	3	<1	ND
F22A, W109A	43	8	<1	-
F18A, F73A	13	2	<1	-
F73A, W109A	2	<1	<1	-
R62A	81	67	<1	+
N63V	71	84	36	++
K66A	55	50	14	+
R67A	49	18	4	+

<sup>a</sup> Mutants have been classified qualitatively into three categories based on their ability to rescue the ts phenotype of T7A49 cells grown in liquid media. If the complementation observed with the mutant protein was comparable to that of the wild-type protein, the mutants were classified as ++; the mutants which behaved like the untransformed T7A49 cells and showed a complete loss of growth after two hours of heat shock were classified as -. There is an intermediate classification defined as +, which refers to mutants that were able to either weakly or moderately rescue the ts phenotype. This classification into three groups, although arbitrary, has enabled us to determine if mutations introduced in C5 protein had a severe, moderate, or nil effect on the activity of the protein *in vivo*. ND, not determined.

N63P were not able to rescue the ts phenotype of T7A49 cells (data not shown).

### RNase P assays using C5 protein or its mutant derivatives

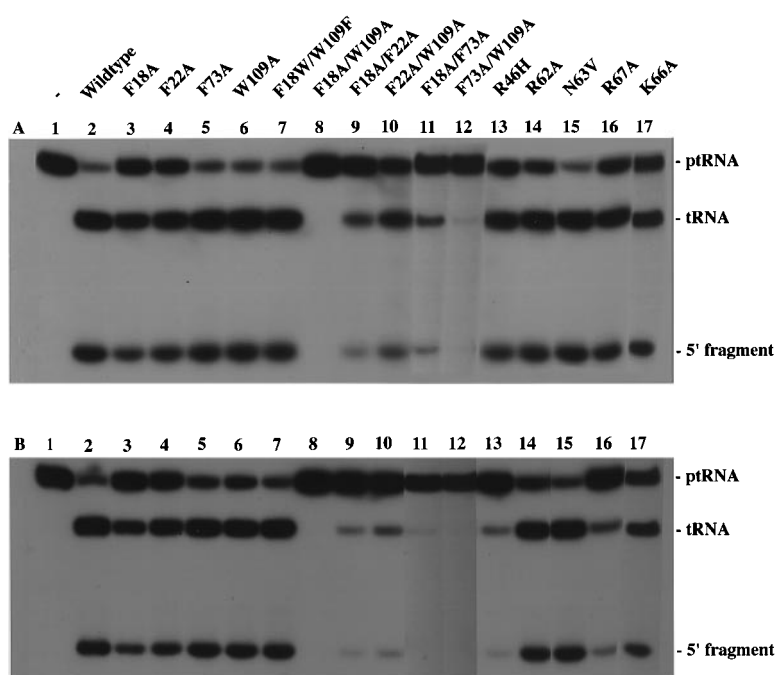
We examined the ability of the various mutants to participate in RNase P catalysis *in vitro*. Wild-type C5 and its mutant derivatives were purified subsequent to their overexpression in an *E. coli* strain, BL21 (DE3), a  $\lambda$  lysogen in which expression of the T7 RNA polymerase gene is regulated by the *lac* UV5 promoter. Fractionation of the crude cell extracts revealed that many of the overexpressed mutant proteins were present (to varying extents) in the P30 (the pellet obtained after centrifugation at 30,000g) unlike the wild-type C5 protein, which fractionates to the S30 (the supernatant obtained after centrifugation at 30,000g). It is possible that these mutant derivatives of C5 protein, when overexpressed in BL21 (DE3) cells at 37°C, aggregate and form inclusion bodies. A purification procedure has been described by Baer *et al.* (1989) for isolating C5 R46H from the P30. We have employed the same procedure (with some modifications) to purify the various mutant derivatives that were constructed in this study (see Materials and Methods). Although the purification procedure involves the use of a strong denaturing agent, we have verified using spectroscopic techniques that several of these mutants thus isolated do regain structure after the step-wise removal of urea (Gopalan *et al.*, 1997). All the mutant derivatives of C5 protein were purified to near homogen-

eity as judged by silver staining of SDS-polyacrylamide gels.

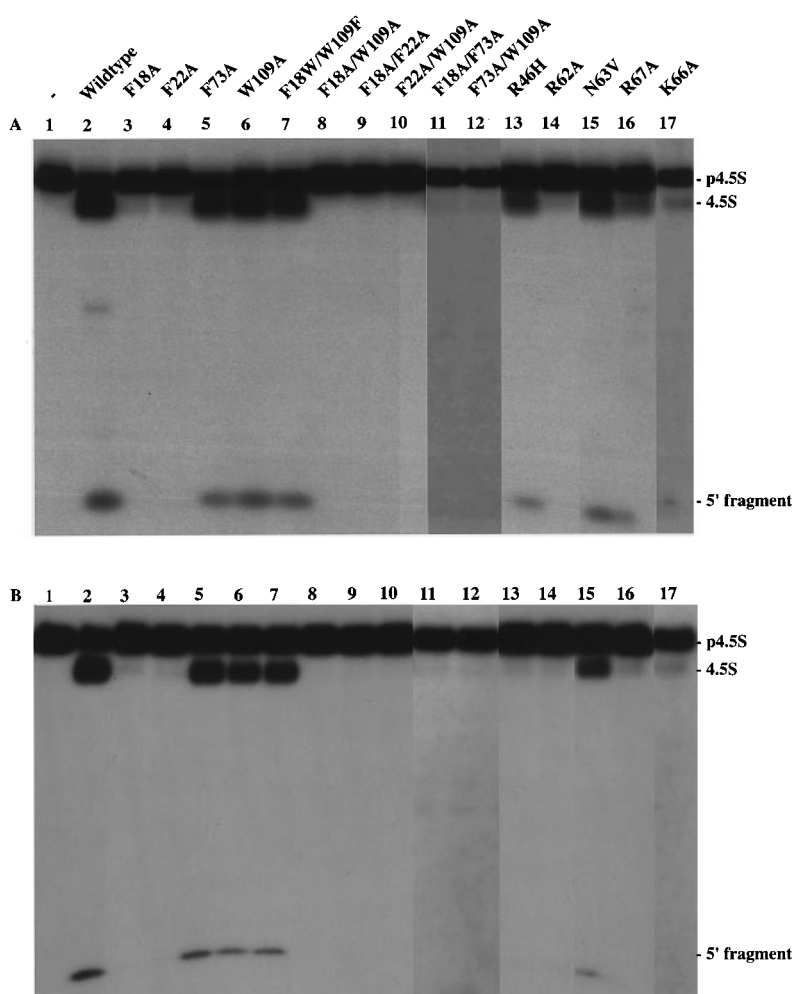
The various mutant derivatives of C5 protein were reconstituted with wild-type M1 RNA and the ability of these holoenzymes to cleave the precursors to tRNA<sup>Tyr</sup> (ptRNA<sup>Tyr</sup>) and 4.5S RNA (p4.5S RNA) was examined. A vast excess (200-fold) of the protein relative to M1 RNA was used to ensure that holoenzyme (M1 RNA + C5 protein) assembly was favored even with those mutant derivatives of C5 protein that might exhibit RNA-binding defects. We performed RNase P assays at 30°C and 43°C to check for thermosensitivity. The results of these assays are depicted in Figure 3 (for cleavage of ptRNA<sup>Tyr</sup>) and Figure 4 (for cleavage of p4.5S RNA).

Altering amino acid residues in C5 protein elicits various effects on RNase P catalysis. There are mutants (such as C5 F18A, C5 F22A, and C5 R62A) which can cleave ptRNA<sup>Tyr</sup> fairly efficiently but not p4.5S RNA (compare lanes 3, 4 and 14 in Figure 3A versus Figure 4A). Some mutants display a decrease in activity with increase in assay temperature from 30°C to 43°C (for example, C5 F22A/W109A; compare lane 10 in Figure 3A versus 3B). There are at least two mutants (C5 F18A/W109A and C5 F73A/W109A) which are severely defective in promoting RNase P activity with either substrate at both 30°C and 43°C (Figure 3A and B, lanes 8 and 12).

The initial velocity for hydrolysis of ptRNA<sup>Tyr</sup> was measured for the mutant holoenzymes and compared with that of the wild-type holoenzyme (Table 1). The initial velocity observed for the wild-



**Figure 3.** Effect of mutations in C5 protein on the activity of the RNase P holoenzyme with ptRNA<sup>Tyr</sup> as the substrate. Holoenzymes, composed of M1 RNA (1 nM) and either wild-type C5 protein or its mutant derivatives (200 nM), were reconstituted and then assayed for activity at either 30°C (A) or 43°C (B). For more details, refer to Materials and Methods.



**Figure 4.** Effect of mutations in C5 protein on the activity of the RNase P holoenzyme with p4.5S RNA as the substrate. Holoenzymes, composed of M1 RNA (1 nM) and either wild-type C5 protein or its mutant derivatives (200 nM), were reconstituted and then assayed for activity at either 30°C (A) or 43°C (B). For more details, refer to Materials and Methods.

type holoenzyme at 30°C and 43°C are 12 min<sup>-1</sup> and 36 min<sup>-1</sup>, respectively. The initial velocities observed with the mutant holoenzymes are reported as the relative activities compared to that of the wild-type holoenzyme (Table 1). While alteration of some of the conserved hydrophobic residues (for example, F18A, F22A, F73A, and W109A) individually resulted in a maximum of threefold decrease in RNase P activity, changing two hydrophobic residues simultaneously to Ala rendered the mutants nearly inactive, especially when the assay temperature was 43°C. This pattern is reminiscent of the results obtained with the complementation assay described above.

For the hydrolysis of p4.5S RNA by the wild-type and mutant RNase P holoenzymes, the initial velocity was not calculated. The relative activity of the mutant holoenzymes, compared to the wild-type holoenzyme, is provided in Table 1 and was calculated on the basis of per cent cleavage observed while assaying for activity with 100 nM substrate at 43°C for ten minutes. There were no qualitative differences in the relative activities when the same assay was performed with 500 nM substrate (data not shown).

## Discussion

### Effect of mutations in C5 protein on RNase P activity *in vitro*

We have demonstrated that certain hydrophobic and basic residues in C5 protein are important for RNase P catalysis *in vivo* and *in vitro* and that some of the single amino acid changes in C5 protein alter the substrate specificity of the RNase P holoenzyme.

The impairment of function (i.e. RNase P catalysis) observed with certain mutant derivatives of C5 protein could result from: (1) the inability of the mutants to fold into a stable tertiary structure, or (2) loss of critical nucleic acid/amino acid residue contacts in the holoenzyme complex. Alteration of side-chains that are fully or partially buried in the wild-type C5 protein structure might result in either local or global destabilization of the tertiary structure depending on: (1) the ability of the protein to adapt to this mutation (*via* structural rearrangements), and (2) the extent to which the cavity, created as a result of the mutation, is deleterious to function (Matthews, 1993; Cordes *et al.*, 1996). The substitution of solvent-exposed amino acid residues (for example, Arg62 or Lys66) with Ala would *a priori* not be expected to perturb the tertiary structure of the protein. Since solvent-exposed residues form the molecular surfaces that mediate binding to ligands, the alteration of such residues in C5 protein might result in loss of contacts in the holoenzyme complex.

The observed increase in affinity of C5 protein for M1 RNA with increasing ionic strength suggests that hydrophobic interactions play a role in holoenzyme assembly (Talbot & Altman, 1994).

The effects observed with the various mutant derivatives of C5 protein in which hydrophobic residues have been altered individually, or in pairs, implicate these residues as being important for holoenzyme assembly and function. The replacement of conserved hydrophobic residues individually was much less detrimental to function than the simultaneous alteration of two hydrophobic residues. For instance, when ptRNA<sup>Tyr</sup> served as the substrate, the holoenzymes reconstituted with mutants C5 F18A, C5 W109A, and C5 F18A/W109A exhibit initial velocities of 34%, 69% and <1%, respectively, relative to that of the wild-type holoenzyme (Table 1). On the basis of fluorescence spectroscopic analyses performed with wild-type C5 protein and its mutant derivatives C5 F18A, C5 F73A, C5 F22A, C5 F18A/F73A, C5 F18A/F22A, and C5 F18W/W109F, we concluded that Phe18 and Phe73 influence the fluorescence emission of Trp109 and that Phe18 and Phe73 are proximal to Trp109 in the tertiary structure as part of an aromatic core (Gopalan *et al.*, 1997). It is noteworthy that the two mutants C5 F18A/W109A and C5 F73A/W109A display negligible RNase P activity at 30°C and 43°C with the two substrates used in this study (Figures 3 and 4) and certainly are the most severely compromised mutants with regard to RNase P catalysis. The fluorescence data taken together with our functional analysis of the various mutants suggest that Phe18, Phe73, Phe22, and W109 play an important role in maintaining the structural core of the protein. Studies are in progress to determine the stability of the various mutant derivatives using circular dichroism spectroscopy.

There are some hydrophobic mutants which display a decrease in activity with increasing temperature. For example, the relative activities (compared to that of the wild-type holoenzyme) observed with the mutants C5 F18A/F73A and C5 F22A/W109A are 13% and 43%, respectively, at 30°C, and 2% and 8%, respectively, at 43°C (Table 1). This result suggests that destabilization caused by certain hydrophobic mutations manifests in a defective phenotype only at the higher temperature, perhaps a reflection of the adaptability of the protein to these alterations at the lower temperature. This is consistent with results from various studies on the structural responses of proteins to replacement of amino acid residues, which have revealed an unexpected degree of tolerance to even seemingly disruptive hydrophobic mutations (Matthews, 1993; Cordes *et al.*, 1996).

The mutants C5 R62A, C5 K66A, and C5 R67A have enabled us to examine the role of some conserved basic residues in RNase P catalysis. Although mutants C5 R62A and C5 K66A are able to efficiently cleave ptRNA<sup>Tyr</sup>, their ability to cleave p4.5S RNA is compromised. Unlike C5 R62A and C5 K66A, which do not display a thermosensitive phenotype with regard to ptRNA<sup>Tyr</sup> cleavage, C5 R67A displays higher activity at 30°C (with both substrates tested here) compared to that

observed at 43°C. This behavior of C5 R67A is reminiscent of that of C5 R46H reported by Baer *et al.* (1989). The catalytic efficiency of the C5 R46H mutant holoenzyme was lower than that of the wild-type holoenzyme and it was demonstrated that assembly of the mutant holoenzyme was defective (Baer *et al.*, 1989). Multiple weak, non-covalent bonds play a crucial role in stabilizing macromolecular complexes such as the RNase P holoenzyme. The absence of such interactions in the mutant RNase P holoenzyme complexes, containing either C5 R67A or C5 R46H, will presumably manifest as a thermosensitive phenotype of functional activity on account of the low stability of the holoenzyme.

The alteration of conserved residues in mutants such as C5 N63V and C5 F73A does not result in drastic changes in the *in vitro* activity of the protein (under conditions examined here). Furthermore, for the purpose of RNA-protein footprinting experiments, we have prepared some single cysteine-substituted mutants (for example, C5 S16C/C113S) that are active both *in vivo* and *in vitro*. All these mutants serve as controls to indicate that not all mutations are detrimental to the function of the protein (data not shown).

#### Altered substrate specificity of certain mutant RNase P holoenzymes

The mutants C5 F18A, C5 F22A, and C5 R62A are fairly efficient in promoting hydrolysis of p<sub>4.5S</sub> RNA but not p<sub>4.5S</sub> RNA (Table 1; Figures 3 and 4). The observation that these mutants can help M1 RNA cleave p<sub>4.5S</sub> RNA indicates that these mutants do bind M1 RNA. However, the holoenzymes that are formed by the interaction of M1 RNA and these mutants (individually) must be different in their structure compared to the wild-type holoenzyme, as reflected by the narrower substrate specificity of the mutant holoenzymes. Kinetic analyses are needed to distinguish whether the inability of these mutant holoenzymes to cleave p<sub>4.5S</sub> RNA is due to weak substrate binding or to a slow rate of cleavage.

#### Function of mutant derivatives of C5 protein *in vivo*

There is a reasonable correlation between the ability of the various mutant derivatives to participate in RNase P catalysis *in vitro* and that observed *in vivo* (Table 1). At 43°C, mutants C5 F18A/W109A, C5 F22A/W109A, and C5 F18A/F73A are severely compromised in their ability to cleave p<sub>4.5S</sub> RNA and p<sub>4.5S</sub> RNA *in vitro* and are also unable to support growth of *T7A49* cells. C5 F22A and C5 R67A, whose p<sub>4.5S</sub> RNA cleavage activity is reduced 2.5- to 5-fold *in vitro*, exhibit only moderate complementation *in vivo*.

However, there are some mutants, such as C5 R62A, which catalyze the cleavage of p<sub>4.5S</sub> RNA quite efficiently but are able to support only moderate complementation *in vivo*. It is noteworthy

that C5 R62A does not catalyze the hydrolysis of p<sub>4.5S</sub> RNA *in vitro*; this raises the possibility that a defect in 4.5S RNA biosynthesis might underlie the inability of these mutants to fully complement the ts phenotype. However, since our complementation assay examines the ability of *T7A49* cells to grow at 43°C, it is possible that heat shock proteins which are induced at 43°C are able to suppress the phenotypic effects of a defect in 4.5S RNA biosynthesis (Wood *et al.*, 1992). Therefore, any defect in 4.5S RNA biosynthesis (as a result of low RNase P activity) might not manifest as a severe growth defect in *T7A49* cells at 43°C.

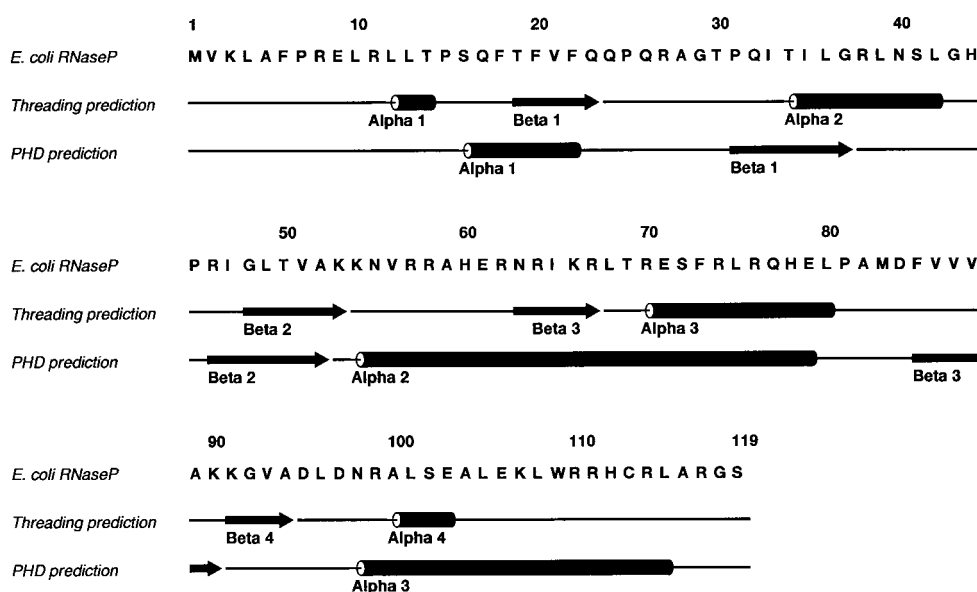
Although an evaluation of *in vitro* and *in vivo* results leads to useful inferences, discrepancies in correlating *in vitro* with *in vivo* results may be the consequence of several factors. In this report, we have tested the activity of certain mutant derivatives of C5 protein *in vitro* with two of the substrates of RNase P. This might not be adequate to extrapolate and draw conclusions about the *in vivo* performance of these mutants considering that RNase P acts on as many as 60 different substrates *in vivo* and that the effects of C5 protein on RNase P catalysis are substrate-identity dependent (Kirsebom & Altman, 1989; Peck-Miller & Altman, 1991; Kirsebom & Svård, 1992).

The failure of some of the mutant derivatives to complement the ts phenotype of *T7A49* cells could be due to lower affinity of the mutants, relative to wild-type C5 protein, for M1 RNA. It is to be expected then that overexpression of these mutants in *T7A49* cells would favor holoenzyme assembly and thus rescue the ts phenotype. However, the IPTG-induced overexpression of C5 wild-type protein (or its mutant derivatives) proved to be toxic to the cell regardless of growth temperature (data not shown).

#### Structural predictions of C5 protein

In the absence of a three-dimensional structure of C5 protein, it is not possible to explain fully the functional phenotypes observed with the mutant derivatives of C5 protein. Until X-ray crystallographic or NMR spectroscopic studies establish the tertiary structure of C5 protein, a working three-dimensional model of C5 protein will help to design rational mutants and dissect the role of various residues in the function of C5 protein.

An  $\alpha\beta\alpha\beta\alpha$  core motif (Figure 5) was predicted for all the protein subunits of RNase P from bacteria by the PHD algorithm (Rost & Sander, 1993). Reported accuracy rates are 72% for prediction of protein secondary structure by this algorithm. Any secondary structure prediction is likely to be compatible with various tertiary structure models. However, we entertained the possibility that C5 protein could adopt a fold the core of which consists of an anti-parallel  $\beta$ -sheet flanked on one side by two  $\alpha$ -helices, since this fold has already been observed in other RNA-binding proteins (for example, U1A, a spliceosomal protein and S6, a



**Figure 5.** Predictions of the secondary structure of C5 protein. The predictions were obtained using the PHD algorithm (Rost & Sander, 1993) and based on threading analysis (Bryant & Lawrence, 1993) in which the sequences of C5 protein and its homologs were individually threaded through the X-ray structure of a human U1A mutant protein complexed with an RNA hairpin (Oubridge *et al.*, 1994). The threading analysis used only the protein coordinates from the B/Q complex in this PDB file. For further information on the data obtained from the threading analysis and for coordinates of the tertiary structure model, contact [landsman@nih.gov](mailto:landsman@nih.gov).

ribosomal protein; Lindahl *et al.*, 1994; Nagai *et al.*, 1990, 1995; Nagai, 1996). Although the PHD prediction of  $\alpha\beta\beta\alpha\beta\alpha$  is different from the  $\beta\alpha\beta\alpha\beta\alpha$  motif in U1A, we decided to use the "threading" technique to evaluate the possibility that C5 protein adopts a fold similar to that of U1A.

Threading (or homology model building) methods enable examination of the ability of a given query sequence to adopt a three-dimensional structure already in the protein database (Bryant & Lawrence, 1993) and are capable of revealing structural similarities which are not evident from conventional sequence alignments. Our computations indicate that despite the low degree of sequence identity among the 12 RNase P protein sequences from various bacteria, ten out of 12 sequences can adopt the U1A fold consisting of a four-stranded anti-parallel  $\beta$ -sheet with two  $\alpha$ -helices packed on one side (unpublished observations). Based on the sequence-structure alignment generated using homology model building, we constructed a hypothetical three-dimensional model of C5 protein which identifies roles for various conserved residues. The  $\beta$ -sheet in C5 protein could provide a surface for docking the RNA ligand, as in other well-characterized RNA-binding domains (Oubridge *et al.*, 1994; Nagai *et al.*, 1995). A stretch of seven amino acid residues (RNRIKRL), containing four basic residues, is conserved in nearly all the RNase P protein subunits from bacteria. Interestingly, the NRIRK sequence is in strand  $\beta 3$  (according to the U1A nomenclature) and is the equivalent of the highly conserved RNP1 consen-

sus motif in U1A which plays a crucial role in RNA binding. The surface electrostatic potential map of C5 protein reveals a binding cleft, rich in conserved, positively charged residues (such as Lys54, Arg62, and Lys66), which could anchor a loop present in M1 RNA (data not shown).

Although a number of correct structural predictions have been made with computational methods, it is critical to bear in mind that the secondary and tertiary structure predictions mentioned above serve only as working models of C5 protein. Further experimental data in support of or in contradiction of these predictions are required.

## Materials and Methods

### Materials

The various reagents used in this study were obtained from the indicated commercial sources: restriction enzymes, T4 DNA polymerase, T4 DNA ligase, T4 polynucleotide kinase, and *E. coli* DNA polymerase (Klenow fragment) were from New England Biolabs; T7 RNA polymerase, SP6 RNA polymerase, *Sma*I, RNasin and M13K07 helper phage were from Promega; nucleoside triphosphates and C50 CM Sephadex were from Pharmacia Biotech; carbenicillin was from Gemini Bioproducts; BCA protein assay reagent was from Pierce; QuickSpin Sephadex columns were from Boehringer Mannheim; and [ $\alpha$ - $^{32}$ P]GTP was from Amersham Biochemicals.

**Table 2.** Oligonucleotides used for mutagenesis

Mutation(s)	Sequence of oligonucleotide used for site-directed mutagenesis	Name of plasmid
F18A	5' G GAA AAC GAA GGT <b>AGC</b> CTG GGA CGG GG 3'	pBSC5Sn7
W109A	5' G GCG GCG <b>CGC</b> TAA TTT TTC C 3'	pBSC5Sn8
R62A	5' GT CAG ACG TTT <u>GAT</u> CCG ATT <b>AGC</b> TTC ATG GGC G 3'	pBSC5Sn9
N63P	5' GT CAG ACG TTT <u>GAT</u> CCG <b>CGG</b> GCG TTC ATG GG 3'	pBSC5Sn10
N63V	5' GT CAG ACG TTT <u>GAT</u> CCG <b>AAC</b> GCG TTC ATG GG 3'	pBSC5Sn13
F22A	5' C ACG CTG CGG <u>CTG</u> CTG <b>CGC</b> AAC GAA GG TGA ACT GGG 3'	pBSC5Sn16
F73A	5' C ATG TTG GCG CAG ACG <b>CGC</b> GCT TTC ACG CGT CAG 3'	pBSC5Sn17
K66A	5' G GCG CAG ACG GAA GCT TTC <u>GCG</u> <u>AGT</u> CAG ACG <b>AGC</b> AAT CCG ATT GCG TTC ATG GGC 3'	pBSC5Sn24
R67A	5' G GCG CAG ACG GAA GCT TTC <u>GCG</u> <u>AGT</u> CAG <b>AGC</b> TTT AAT CCG ATT GCG TTC ATG GGC 3'	pBSC5Sn25
F18A, W109A	5' G GAA AAC GAA GGT <b>AGC</b> CTG GGA CGG GG 3' (F18A) 5' G GCG GCG <b>CGC</b> TAA TTT TTC C 3' (W109A)	pBSC5Sn78
F18W, W109F	5' GAA AAC GAA GGT <b>CCA</b> CTG GGA CGG G 3' (F18W) 5' GTG GCG GCG <b>GAA</b> TAA TTT TTC C 3' (W109F)	pBSC5Sn112
F18A, F22A	5' C ACG CTG CGG CTG CTG <b>CGC</b> AAC GAA GGT <b>AGC</b> CTG GGA CGG GG 3'	pBSC5Sn716
F18A, F73A	5' G GAA AAC GAA GGT <b>AGC</b> CTG GGA CGG GG 3' (F18A) 5' C ATG TTG GCG CAG ACG <b>CGC</b> GCT TTC ACG CGT CAG 3' (F73A)	pBSC5Sn717
F22A, W109A	5' C ACG CTG CGG CTG CTG <b>CGC</b> AAC GAA GG TGA ACT GGG 3' (F22A) 5' G GCG GCG <b>CGC</b> TAA TTT TTC C 3' (W109A)	pBSC5Sn816
F73A, W109A	5' C ATG TTG GCG CAG ACG <b>CGC</b> GCT TTC ACG CGT CAG 3' (F73A) 5' G GCG GCG <b>CGC</b> TAA TTT TTC C 3' (W109A)	pBSC5Sn817

Since the single-stranded DNA used as the template for site-directed mutagenesis of C5 gene contains the gene in the sense orientation, the oligonucleotides used are in the antisense orientation. The bold letters indicate the codon that was altered. The underlined nucleotides represent changes in the wobble position (of the respective codon) that was introduced in addition to the desired mutation. These additional modifications (i.e. silent mutagenesis) led to changes in the restriction pattern of the mutant gene and enabled the rapid screening of mutants. In many cases there were no other changes in addition to the desired mutation since the mutation itself resulted in a restriction pattern that is distinct from that of the wild-type C5 gene.

### Construction of pBSC5

Vioque *et al.* (1988) had constructed the plasmid pARE7 in which a semisynthetic C5 gene was placed downstream from the promoter and the ribosome binding site of gene 10 of T7 bacteriophage. For objectives unrelated to this study, we constructed a plasmid pVG2, which is analogous to pARE7 except that it lacks the *EcoRI* site in the vector. In order to clone the C5 gene into a vector which possesses the f1 filamentous phage origin of replication, the entire nucleotide sequence containing the T7 RNA polymerase promoter, translational signals, the C5 coding region and the T7 transcription terminator was moved *en bloc* from pVG2 to pBluescript II KS (+) to generate the plasmid, pBSC5. This subcloning involved: (1) digesting pVG2 with *BglII* and *EcoRV*, (2) filling-in the 3' recessed ends created by the *BglII* digest, and (3) ligating this blunt-ended DNA fragment into pBluescript II KS (+) which had been digested with *PvuII*. Upon digesting pBluescript II KS (+) with *PvuII*, the *lacI*, multiple cloning box, and the *lacZ* sequences present in between the two *PvuII* sites were deleted. In pBSC5, the C5 gene is cloned in the same orientation as *lacI* and *lacZ* in the parental vector (pBluescript II KS (+)).

### Site-directed mutagenesis of C5 protein

Mutations were engineered in the gene encoding C5 protein using the "oligonucleotide-directed mutagenesis without phenotypic selection" procedure (Kunkel, 1985). Single-stranded (ss) DNA was prepared by coinfection of C/236 (pBSC5) with the helper phage M13K07. The ss

DNA was isolated according to the manufacturer's instructions in the pALTER kit (Promega). Mutagenesis reactions were performed according to the protocol described by Kunkel (1989). The various mutant derivatives of C5 protein were generated using the DNA oligonucleotides described in Table 2. DNA oligonucleotides were synthesized at the Keck Biotechnology Resource Laboratory at Yale Medical School.

In the T7 promoter-driven system for overexpression of proteins (Studier *et al.*, 1990), it is preferable to clone the target gene initially into a host that does not contain the T7 RNA polymerase gene to ensure that there is no plasmid instability due to toxicity associated with expression of the target gene. Therefore, to obtain clones of the mutant derivatives of C5 protein, DH5 $\alpha$  was transformed with the mutagenesis reactions and the plasmid DNA containing the desired mutation isolated. The presence of the engineered mutation was confirmed by sequencing of the various plasmid DNA samples. Subsequently, these plasmid DNAs were used to transform BL21(DE3) cells and the respective mutant proteins overexpressed.

### Complementation assay *in vivo*

T7A49 cells containing both the *rnpA49* mutation (i.e. C5 R46H) in their chromosomes and the T7 RNA polymerase gene under control of the *lac UV5* promoter (Guerrier-Takada *et al.*, 1995) were transformed with plasmids bearing the wild-type C5 protein or its mutant derivatives under the control of a promoter for T7 RNA polymerase transcription. The permissive temperature for this strain is 30°C while the non-permissive tempera-

ture is 43°C. The transformants bearing the various mutants were then grown overnight at 30°C in LB media supplemented with carbenicillin (100 µg/ml). Cultures were reinoculated the next morning in fresh LB carbenicillin media and grown at 30°C until the  $A_{600} \sim 0.20$ . The cultures were then shifted to 43°C and the cell growth at the non-permissive temperature monitored by measuring  $A_{600}$  at regular time intervals. Cells (without any plasmids) were grown in the absence of carbenicillin and served as the negative control while cells transformed with pBSC5 (wild-type C5 protein) served as the positive control for complementation of the *ts* phenotype.

### Overexpression and purification of mutant derivatives of C5 protein

*BL21(DE3)* cells containing plasmids encoding the various mutant derivatives of C5 protein (see Table 2) were grown to  $A_{600} \sim 0.4$  and then induced with 2 mM IPTG. After establishing in small scale cultures (approximately 4 ml) that there was T7 promoter-driven overexpression of the various proteins, large scale (1 l) cultures were grown to isolate the respective proteins. When the crude cell extract was prepared either by sonication or by grinding the cells with alumina and subsequently centrifuged at 30,000g for 30 minutes, it was discovered that the mutant proteins were present in the P30 (i.e. the pellet obtained after the centrifugation) rather than the S30 (i.e. the supernatant obtained after the centrifugation). The relative amounts in the P30 and the S30 varied for the various mutants, perhaps indicative of the folding properties of the various mutants. The mutant proteins were isolated from the P30 following the protocol described by Baer *et al.* (1989). This purification scheme for the isolation of C5 R46H involves solubilizing the P30 with 4 M urea (in 50 mM Tris-hydrochloride (pH 7.5), 10 mM magnesium acetate, 60 mM ammonium chloride) in order to recover the protein (Baer *et al.*, 1989). The urea-solubilized suspension was then centrifuged for 30 minutes at 30,000g. The P30 wash was dialyzed against 50 mM Tris-hydrochloride (pH 7.5), 100 mM ammonium chloride, 10 mM magnesium chloride, 10 mM dithiothreitol (DTT). The protein precipitates upon removal of urea. The precipitated protein is then resuspended in a buffer containing 7 M urea and further purified using CM Sephadex C50 chromatography. Our modifications from this procedure are: (1) the cells were lysed using a sonicator instead of being crushed in alumina, and (2) the deliberate omission of the reducing agent before loading the protein on a CM Sephadex column. In the absence of DTT, dimer formation is promoted. The dimeric version of the C5 mutants elutes at a higher salt concentration, relative to the monomer, and can help purify the C5 mutant proteins from contaminants that elute at lower salt concentrations.

Aliquots from the various fractions eluted from CM Sephadex columns were electrophoresed on SDS-polyacrylamide gels and stained with silver nitrate. Only those fractions which showed high purity (>95%) were pooled and the stocks were stored at -70°C. Immediately before use in RNase P assays, the various mutant preparations were thawed and treated with 10 mM DTT. The protein concentration of the various preparations was assessed using the BCA test. The standard curve for the BCA test was generated using wild-type C5 protein, the concentration of which was determined by measuring Trp absorbance at 280 nm.

### Assays for RNase P activity *in vitro*

Plasmid encoding M1 RNA (pJA2') was linearized with *FokI* and transcribed by T7 RNA polymerase as described by Vioque *et al.* (1988). The RNA was then isolated using a QuickSpin column procedure (Vioque *et al.*, 1988). Plasmids encoding ptRNA<sup>Tyr</sup> and p4.5S RNA were linearized with *FokI* and *SmaI*, respectively, and these RNA substrates were internally labeled with [ $\alpha$ -<sup>32</sup>P]GTP during *in vitro* transcription with T7 RNA polymerase and then purified on denaturing 8% (w/v) polyacrylamide/7 M urea gels.

Cleavage of RNA substrates was performed with RNase P holoenzymes reconstituted using 1 nM M1 RNA and 200 nM of C5 protein or its mutant derivatives. The assays were performed with 100 nM ptRNA<sup>Tyr</sup> or p4.5S RNA at 30°C or 43°C. The cleavage reactions were performed in 10 mM Hepes (pH 7.5), 10 mM magnesium acetate, 400 mM ammonium acetate, 5% (v/v) glycerol and 0.01% (v/v) NP-40. The ptRNA<sup>Tyr</sup> assays were carried out at 30°C and 43°C for ten minutes and five minutes, respectively. The p4.5S RNA assays were carried out at 30°C and 43°C for 20 minutes and ten minutes, respectively. Autoradiograms were obtained after separating the products of the various reactions on an 8% (for ptRNA<sup>Tyr</sup>) or 7% (for p4.5S RNA) polyacrylamide/7 M urea gel. The autoradiograms were scanned using a Microtek MSF-300Z scanner and the Scan Maker Plug-in software for Adobe Photoshop version 3.0. In order to group the hydrophobic mutants together, the autoradiograms from two different experiments were used to generate the composites shown in Figures 3 and 4.

To calculate the initial velocity, the enzyme assays were performed at the indicated temperature and aliquots withdrawn at regular time intervals and quenched with 9 M urea. The products of the reaction were electrophoresed on polyacrylamide/7 M urea gels and the extent of cleavage was calculated by quantifying the intensity of the various bands using a phosphorimager (Fuji). The extent of cleavage was always restricted to the linear range for product formation.

### Database searches

The GenPept release 91.0 (Benson *et al.*, 1996), EMBL release 43.0 (Rodriguez-Tome *et al.*, 1996), PIR version 45.0 (George *et al.*, 1996), and Swiss-Prot version 31.0 (Bairoch & Apweiler, 1996) databases were searched using the BLASTP algorithm (Altschul *et al.*, 1990), with the *E. coli* RNase P sequences used as the basis for comparison. BLAST search cutoffs used to identify homologs were a Karlin/Altschul score for two aligned sequence segments >70 with a probability of <10<sup>-3</sup>. These database searches identified 12 full-length RNase P sequences.

### Acknowledgements

We are grateful to members of our laboratory, especially Drs Cecilia Guerrier-Takada, Paul Eder and Urban Lundberg, for helpful discussions. Research in the laboratory of S.A. is supported by National Institutes of Health grant GM19422. V.G. was funded by postdoctoral fellowships from the Patrick and Catherine Donaghue Foundation for Medical Research and the Anna Fuller Fund for Cancer Research.

## References

- Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C. M., Bruni, C. B. & Carlomagno, M. S. (1994). RNase E provides substrates for ribonuclease P-dependent processing of polycistronic RNA. *Genes Dev.* **8**, 3021–3031.
- Altman, S., Kirsebom, L. & Talbot, S. J. (1995). Recent studies of RNase P. In *tRNA: Structure, Biosynthesis and Function* (Soll, D. & RajBhandary, U., eds), pp. 67–78. American Society for Microbiology, Washington, DC.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Baer, M. F., Wesolowski, D. & Altman, S. (1989). Characterization *in vitro* of the defect in a temperature-sensitive mutant of the protein subunit of RNase P from *Escherichia coli*. *J. Bacteriol.* **171**, 6862–6866.
- Bairoch, A. & Apweiler, R. (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucl. Acids Res.* **24**, 21–25.
- Benson, D. A., Boguski, M., Lipman, D. J. & Ostell, J. (1996). GenBank. *Nucl. Acids Res.* **24**, 1–5.
- Bothwell, A. L. M., Garber, R. L. & Altman, S. (1976). Nucleotide sequence and *in vitro* processing of a precursor molecule to *Escherichia coli* 4.5S RNA. *J. Biol. Chem.* **251**, 7709–7716.
- Bryant, S. H. & Lawrence, C. E. (1993). An empirical energy function for threading protein sequence through a folding motif. *Proteins: Struct. Funct. Genet.* **16**, 92–113.
- Burd, C. G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
- Cordes, M. H. J., Davidson, A. R. & Sauer, R. T. (1996). Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **6**, 3–10.
- George, D. G., Barker, W. C., Mewes, H. W., Pfeiffer, F. & Tsugita, A. (1996). The PIR-international protein sequence database. *Nucl. Acids Res.* **24**, 17–20.
- Gopalan, V., Golbik, R., Schreiber, G., Fersht, A. R. & Altman, S. (1997). Fluorescence properties of a tryptophan residue in an aromatic core of the protein subunit of ribonuclease P from *Escherichia coli*. *J. Mol. Biol.* **267**, 765–769.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. R. & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
- Guerrier-Takada, C., Li, Y. & Altman, S. (1995). Artificial regulation of gene expression in *Escherichia coli* by RNase P. *Proc. Natl Acad. Sci. USA*, **92**, 11115–11119.
- Kirsebom, L. A. & Altman, S. (1989). Reaction *in vitro* of some mutants of RNase P with wild-type and temperature-sensitive substrates. *J. Mol. Biol.* **207**, 837–840.
- Kirsebom, L. A. & Svård, S. G. (1992). The kinetics and specificity of cleavage by RNase P is mainly dependent on the structure of the amino acid acceptor stem. *Nucl. Acids Res.* **20**, 425–432.
- Kirsebom, L. A., Baer, M. F. & Altman, S. (1988). Differential effects of the mutations in the protein and RNA moieties of RNase P on the efficiency of suppression by various tRNA suppressors. *J. Mol. Biol.* **204**, 879–888.
- Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. (1994). A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 9223–9227.
- Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.
- Kunkel, T. A. (1989). Mutagenesis of cloned DNA. In *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds), pp. 8.1.1–8.1.6. John Wiley, New York.
- Lazinski, D., Grzadielska, E. & Das, A. (1989). Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell*, **59**, 207–218.
- Lindahl, M., Svensson, L. A., Liljas, A., Sedelnikova, S. E., Eliseikina, I. A., Fomenkova, N. P., Nevskaya, N., Nikonov, S. V., Garber, M. B., Muranova, T. A., Rykonova, A. I. & Amons, R. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249–1254.
- Lumelsky, N. & Altman, S. (1988). Selection and characterization of randomly produced mutants in the gene coding for M1 RNA. *J. Mol. Biol.* **202**, 443–454.
- Matthews, B. W. (1993). Structural and genetic analysis of protein folding and stability. *Curr. Opin. Struct. Biol.* **3**, 589–593.
- Nagai, K. (1996). RNA-protein complexes. *Curr. Opin. Struct. Biol.* **6**, 53–61.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Nagai, K., Oubridge, C., Ito, N., Avis, J. & Evans, P. R. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.* **20**, 235–240.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
- Pace, N. R. & Brown, J. W. (1995). Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. *J. Bacteriol.* **177**, 1919–1928.
- Peck-Miller, K. & Altman, S. (1991). Kinetics of the processing of the precursor to 4.5S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* **221**, 1–5.
- Reich, C., Olsen, G. J., Pace, B. & Pace, N. R. (1988). Role of the protein moiety of RNase P, a ribonucleoprotein enzyme. *Science*, **239**, 179–182.
- Rodrigues-Tome, P., Stoehr, P. J., Cameron, G. N. & Flores, T. P. (1996). The European Bioinformatics Institute (EBI) databases. *Nucl. Acids Res.* **24**, 6–12.
- Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70 accuracy. *J. Mol. Biol.* **232**, 584–596.
- Schedl, P. & Primakoff, P. (1973). Mutants of *Escherichia coli* thermosensitive for the synthesis of transfer RNA. *Proc. Natl Acad. Sci. USA*, **70**, 2091–2095.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.
- Talbot, S. J. & Altman, S. (1994). Kinetic and thermodynamic analysis of RNA-protein interactions in the

- ribonuclease P holoenzyme from *Escherichia coli*. *Biochemistry*, **33**, 1406–1412.
- Thompson, J. D., Higgins, D. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
- Vioque, A., Arnez, J. & Altman, S. (1988). Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *J. Mol. Biol.* **202**, 835–848.
- Wood, H., Lurink, J. & Tollervey, D. (1992). Evolutionarily conserved nucleotides within *E. coli* 4.5S RNA required for association with P48 *in vitro* and for optimal function *in vivo*. *Nucl. Acids Res.* **20**, 5919–5925.

*Edited by K. Nagai*

(Received 21 August 1996; received in revised form 13 December 1996; accepted 6 January 1997)