Protein–Precursor tRNA Contact Leads to Sequence-Specific Recognition of 5′ Leaders by Bacterial Ribonuclease P

Kristin S. Koutmou1†, Nathan H. Zahler1†, Jeffrey C. Kurz1, Frank E. Campbell2,3, Michael E. Harris2,3 and Carol A. Fierke1,4⁎

1Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109, USA
2Center for RNA Molecular Biology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4973, USA
3Department of Biochemistry, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4973, USA
4Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Received 22 September 2009; received in revised form 13 November 2009; accepted 13 November 2009
Available online 28 November 2009

Introduction

Throughout phylogeny, the 5′ leaders of precursor tRNA (pre-tRNA) molecules are endonucleolytically removed by ribonuclease P (RNase P) (for a review, see Smith et al.,1 Kirsebom,2 and Christian et al.3). In Bacteria, RNase P is comprised a single ~400-nucleotide tRNA subunit that is catalytic in vitro and a single small (~120 amino acids) protein subunit (P protein) required for activity in vivo. The RNA and protein components function synergistically in both substrate recognition and catalysis, and exhibit a broad substrate...
Recognition of 5’ Leaders by RNase P

In species such as Bacillus subtilis and Escherichia coli, RNase P recognizes and correctly processes the products of more than 80 pre-tRNA genes. Intriguingly, the substrate pools in both of these species include pre-tRNAs that lack multiple known RNase P recognition elements, suggesting the possibility that additional recognition elements remain to be identified.

Previous studies have established that both the PRNA subunit and the P protein subunit of RNase P contribute to molecular recognition of pre-tRNA substrates. PRNA contacts substrates through both base-pairing and backbone interactions, including interactions with 2’ hydroxyls in the acceptor stem and the D stem loop. Proximal to the cleavage site, PRNA base pairs with the first pre-tRNA nucleotide on the 5’ side of the cleavage site (N(−1)) and the 3’ RCCA substrate motif, and recognizes functional groups in the closing base pair of the tRNA acceptor stem. These interactions contribute to efficient catalysis, cleavage site specificity, and high substrate binding affinity. Multiple lines of evidence indicate that the P protein subunit contacts the 5’ leader of pre-tRNA substrates. Furthermore, the interaction between the B. subtilis P protein and the pre-tRNA leader increases the binding affinity for substrates with increasing leader length. Biochemical investigations indicate that the leader contacts side chains located in a large cleft formed by the central β-sheet and an α-helix of P protein (Fig. 1a). Time-resolved fluorescence resonance energy transfer (trFRET) experiments, affinity-cleavage assays, and molecular modeling indicate that this cleft is proximal to the second through seventh positions on the 5’ side of the cleavage site (N(−2) through N(−7)). Importantly, unlike PRNA, no specific interactions between P protein and pre-tRNA have been identified or characterized to date.

RNase P protein also modulates the cation requirements for RNase P function. P protein enhances the apparent affinities of the RNase P holoenzyme for Mg2+ ions that contribute to substrate binding and catalysis and alters the Mn2+ rescue of phosphorothioate substitutions in PRNA helix P4, a catalytically important divalent metal binding site. Biochemical studies suggest that the effects on metal affinity are related to the role of P protein as a structural linchpin in the RNase P holoenzyme–substrate complex, interacting with the pre-tRNA leader, helping to organize the active site, and enhancing reactivity under physiological conditions.

An additional function of P protein is to enhance the ability of RNase P to recognize a wide variety of naturally occurring, noncanonical pre-tRNA substrates. For E. coli RNase P, interactions between P protein and the 5’ leader of pre-tRNA increase both substrate affinity and single-turnover cleavage rate constants in a sequence-dependent manner. The binding and catalytic enhancements specific for a 5’ leader sequence are transferable between substrates, independent of the tRNA sequence. Overall, these observations suggest that interactions between P protein and substrate leaders contribute to substrate specificity; however, the molecular basis for this discrimination remains to be determined.

Here, we investigate the sequence specificity of interactions between bacterial RNase P and the 5’ leaders of pre-tRNA substrates using a combination of biochemical and genomic analyses. These data demonstrate that both the B. subtilis and E. coli RNase P holoenzymes display nucleotide preferences at position N(−4) of pre-tRNA that

---

**Fig. 1.** Interaction of the 5’ leader of pre-tRNA with B. subtilis P protein in RNase P. (a) Modeled structure of the interface between the pre-tRNA leader and P protein based on affinity cleavage showing the backbone of PRNA (blue), P protein (red), the mature tRNA domain of the substrate (brown), and the pre-tRNA 5’ leader (black). The nucleotide at pre-tRNA position N(−4) is shown in green; P protein regions are labeled. (b) B. subtilis P protein crystal structure showing sites where single-cysteine mutations alter the value of Kd,obs for pre-tRNA<sup>Asp</sup>. The side-chain color reflects the magnitude of the effect of the mutation on the value of Kd,obs: yellow, <3-fold increase; blue, 3- to 25-fold increase; green, >25-fold increase. (c) The effect of alanine mutations on the binding selectivity for A(−4) relative G(−4) mapped onto the structure of the P protein. Alanine mutations that do not alter the binding specificity ratio are highlighted in yellow, and those that abolish the sequence preference at N(−4) are shown in green.
leader interact with the P protein.21,24 Consistent with the fourth (N(−4)) nucleotide preferences observed in vitro for B. subtilis and E. coli RNase P are reflected in the sequences of pre-tRNA genes in these species, consistent with nucleotide selectivity at N(−4) contributing to substrate recognition in vivo. Extrapolation from tRNA gene sequence analysis suggests that such preferences are present in the majority of bacterial species, but are likely variable between species.

Results

Pre-tRNA affinities of B. subtilis and E. coli RNase P depend on the nucleotide at N(−4)

There is growing evidence suggesting that sequence-specific interactions between the 5′ leader of pre-tRNA and bacterial P protein enhance substrate binding affinity and/or cleavage rates, although no specific contacts have yet been identified. Previously, the thermodynamics and kinetics of pre-tRNA association with B. subtilis RNase P suggested that the fourth (N(−4)) and the fifth (N(−5)) nucleotides on the 5′ side of the cleavage site in the pre-tRNA leader interact with the P protein.21,24 Consistent with this, structural models of the B. subtilis RNase P holoenzyme based on affinity cleavage and trFRET measurements place N(−4) and N(−5) near the central cleft of the protein (Fig. 1a). Pre-tRNA affinities of these cations.28,29 Therefore, we examined whether the sequence-specific interaction between P protein and N(−4) is influenced by the divalent metal ion concentration. The affinity of B. subtilis RNase P for A(−4) pre-tRNA increases 65-fold over the same range (Fig. 2c); the dependence of Kd,obs on [Ca2+]f is well-described by a cooperative binding isotherm (Eq. (1)), yielding a Hill coefficient (nH) of 5.5±0.5 and a midpoint for the transition (K1/2, Ca) of 4.5±0.5 mM. The modest increase in the Hill coefficient

Table 1. Effects of altering the nucleobase at N(−4) in B. subtilis pre-tRNAA5p on RNase P substrate affinity

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Leader Sequence</th>
<th>B. subtilis Kd,obs (nM)b</th>
<th>E. coli Kd,obs (nM)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(−4)</td>
<td>5′-ACAU</td>
<td>2 mM [Ca2+]f 9±2 5±2</td>
<td>2 mM [Ca2+]f 35±8</td>
</tr>
<tr>
<td>G(−4)</td>
<td>5′-GCAU</td>
<td>2 mM [Ca2+]f 200±20 11±3</td>
<td>2 mM [Ca2+]f 170±60</td>
</tr>
<tr>
<td>U(−4)</td>
<td>5′-UCAU</td>
<td>2 mM [Ca2+]f 70±10 8±1</td>
<td>2 mM [Ca2+]f 100±30</td>
</tr>
<tr>
<td>C(−4)</td>
<td>5′-GCAU</td>
<td>2 mM [Ca2+]f 150±50 6±1</td>
<td>2 mM [Ca2+]f 54±9</td>
</tr>
<tr>
<td>P(−4)</td>
<td>5′-POCAU</td>
<td>2 mM [Ca2+]f 30±7  ndf</td>
<td>2 mM [Ca2+]f ndf</td>
</tr>
<tr>
<td>DAP(−4)</td>
<td>5′-DAPCAU</td>
<td>2 mM [Ca2+]f 40±7  ndf</td>
<td>2 mM [Ca2+]f ndf</td>
</tr>
<tr>
<td>2AP(−4)</td>
<td>5′-2APCAU</td>
<td>2 mM [Ca2+]f 170±40 ndf</td>
<td>2 mM [Ca2+]f ndf</td>
</tr>
</tbody>
</table>

a Position N(−4) is highlighted.
b Conditions: 50 mM Mes and 50 mM Tris (pH 6.0) at 37 °C and indicated [Ca2+]f. [KCl] was adjusted to maintain ionic strength, as described in Materials and Methods.

c Not determined.
for the metal dependence of substrate affinity compared to previous measurements ($n_{11}=4\pm 1$ for pre-tRNA$^{\text{Asp}}$ with 2- and 3-nucleotide leaders)\textsuperscript{28} most likely results from differences in substrate structure and solution conditions (e.g., $\text{Mg}^{2+}$ versus $\text{Ca}^{2+}$). These results indicate that the affinity of the G$\text{(-4)}$ substrate is coupled to the binding of multiple $\text{Ca}^{2+}$ ions with dissociation constants in the millimolar range. If the same set of metal ions also stabilizes the binding of the A$\text{(-4)}$ substrate, the affinity of the metal ions must be significantly higher. Functionally, this dependence leads to a significant reduction in sequence selectivity at higher divalent cation concentrations (10–25 mM) frequently used to measure RNase P activity (see Smith et al.,\textsuperscript{1} Christian et al.,\textsuperscript{3} and Harris and Christian\textsuperscript{28} and references therein), although the overall pattern of nucleotide preference is retained (Fig. 2b). In Bacteria, the physiological concentration of available $\text{Mg}^{2+}$ has been estimated at 2 mM or lower in vivo.\textsuperscript{37} Thus, the coupling of sequence specificity with divalent metal binding results in sequence selectivity at physiological cation concentrations.

**RNase P recognizes two functional groups in the N$\text{-4)}$ nucleobase**

To define the interactions between *B. subtilis* RNase P and the N$\text{-4)}$ nucleobase that contribute to sequence selectivity, we prepared pre-tRNA with adenosine analogs incorporated at N$\text{-4)}$. These analogs systematically alter the presentation of exocyclic amines along the adenosine Watson–Crick face (for structures, see Fig. S1 in Supplementary Data), including the following: purine (P$\text{-4)}$), 2-aminopurine (2AP$\text{-4)}$), and 2,6-diaminopurine (DAP$\text{-4)}$).\textsuperscript{39} The binding affinity of *B. subtilis* RNase P for pre-tRNA-containing adenosine analogs at N$\text{-4)}$ indicates that recognition of adenosine involves two distinct functional groups (Table 1). Removal of the N6 exocyclic amine by incorporation of purine (P$\text{-4)}$) decreases the binding affinity by 3-fold compared to A$\text{-4)}$ pre-tRNA, indicating a positive interaction likely caused by the formation of a hydrogen bond with the N6 amine. Additionally, the DAP substitution at N$\text{-4)}$ decreases the binding affinity by 4-fold, demonstrating an unfavorable interaction with the N2 amine of this modified base (such as a steric clash) and, by extension, with the N2 exocyclic amine of a guanosine at this position. The 2AP analog combines these two structural changes, removal of the N6 exocyclic amine and addition of the N2 exocyclic amine, leading to a decrease in the affinity of RNase P for 2AP$\text{-4)}$ pre-tRNA by 19-fold. The decrease in binding affinity for the 2AP substitution is consistent with the additive effect of each of the two individual interactions. Furthermore, the affinity of the 2AP$\text{-4)}$ substrate is comparable to that of G$\text{-4)}$ (Table 1), suggesting that the differential binding affinities for the A$\text{-4)}$ and G$\text{-4)}$ substrates can be accounted for by interactions with these two functional groups.

**B. subtilis** P protein side chains contribute to pre-tRNA affinity and sequence specificity

Structural models of the RNase P pre-tRNA complex place nucleotides N$\text{-2)}$ through N$\text{-7)}$ of the pre-tRNA leader near the P protein central cleft and RNR motif\textsuperscript{19,22,23} (Fig. 1a). To identify P protein...
To test the importance of interactions between side chains in the central cleft and the RNR motif with pre-tRNA, we prepared a series of alanine mutations in P protein and measured the binding affinities of the mutant RNase P enzymes for A(−4) and G(−4) pre-tRNAs (Table 2). Alanine mutations were incorporated into P protein sites identified by cysteine scanning mutagenesis (F16, F20), as well as into additional amino acids in the central cleft (S25 and Y34) and the RNR motif (R60, N61, K64, and R65). Pre-tRNA binding was evaluated in 3.5 mM [Ca\(^{2+}\)] to permit measurement of weak pre-tRNA affinities. However, under these conditions, the A(−4)/G(−4) pre-tRNA selectivity ratio \([K_{d,obs}(G^{−4})/K_{d,obs}(A^{−4})]\) is only 2-fold for wild-type RNase P (Fig. 2b), allowing only a qualitative analysis of mutations that affect selectivity.

All of the alanine point mutations in the P protein decrease the affinity of RNase P for both G(−4) and A(−4) pre-tRNAs (Table 2); the effect of these mutations on pre-tRNA affinity is equal to (F16) or less than (F20, S25) the decreases observed for cysteine substitutions at the same position (Fig. 1b). The affinity of RNase P for A(−4) pre-tRNA is decreased >10-fold by four mutations (F16A, Y34A, R60A, and R65A) and 5- to 8-fold by four other mutations (F20A, N61A, K64A). These data are consistent with the cysteine scanning study and confirm that the central cleft and the RNR motif of the P protein contribute to the pre-tRNA affinity of RNase P. Additionally, several P protein mutations alter the pre-tRNA substrate selectivity observed at N(−4). As expected for mutations that disrupt favorable contacts with adenosine at N(−4), three mutations (F20A, Y34A, and R60A) cause a 2- to 3-fold larger decrease in the affinity of A(−4) pre-tRNA compared to that of G(−4) pre-tRNA, causing the A(−4)/G(−4) selectivity ratio to decrease to ≤1. This loss in selectivity was confirmed by substrate affinity measurements at 2.0 mM [Ca\(^{2+}\)]. Under these conditions, the A(−4)/G(−4) selectivity ratio decreases from a value of 2 for wild-type RNase P to a value of 2 for RNase P reconstituted with F20A and Y34A P protein (Table 3). The selectivity ratio at 3.5 mM calcium also decreases modestly (to 1.5) for two other mutations (F16A and N61A). However, changes in selectivity for these mutations are not confirmed by measurement of either the selectivity ratio of the F16C mutant (Table 2) or the affinity of the N61A mutant for DAP(−4) pre-tRNA in 3.5 mM Ca\(^{2+}\) (134±43 μM). Mapping the magnitude of the

<table>
<thead>
<tr>
<th>Protein</th>
<th>G(−4)</th>
<th>A(−4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>200±20</td>
<td>30±7</td>
</tr>
<tr>
<td>F20A</td>
<td>400±90</td>
<td>200±30</td>
</tr>
<tr>
<td>Y34A</td>
<td>210±70</td>
<td>100±25</td>
</tr>
<tr>
<td>Y34F</td>
<td>110±30</td>
<td>50±5</td>
</tr>
</tbody>
</table>

* Conditions: 50 mM Mes and 50 mM Tris (pH 6.0) at 37 °C and 2 mM [Ca\(^{2+}\)]. About 425 mM [KCl] was used to adjust/maintain ionic strength, as described in Materials and Methods.
selectivity ratios onto the P protein structure (Fig. 1c) reveals that the largest decreases in the selectivity ratio are caused by mutations in hydrophobic side chains located in the central cleft (F20 and Y34), presumably near the pre-tRNA leader, and in a conserved side chain in the RNR motif (R60), a region of the protein distal to the 5′ leader in structural models (Fig. 1a). These side chains may play a role in the nucleobase selectivity of RNase P for N(−4) pre-tRNA.

To identify the mechanism of sequence selectivity, we further characterized the properties of the mutations with the largest effects on N(−4) selectivity (F20A, Y34A, and R60A). Single-turnover assays catalyzed by saturating RNase P [0.5 mM MgCl2 and 2 mM Co(NH3)6Cl3, pH 7] indicate that mutations in the central cleft (F20A, Y34A, and Y34F) of P protein alter the observed cleavage rate constant by ≤2-fold. Therefore, these mutations affect only the pre-tRNA binding affinity. However, the R60A mutation in the RNR motif decreases the observed single-turnover rate constant by 3-fold under these same conditions, suggesting that side chains in the RNR motif may contribute to pre-tRNA sequence specificity differently from side chains in the central cleft.

To examine whether residues in the RNR motif and the central cleft enhance pre-tRNA affinity by interacting with the 5′ leader and/or tRNA, we measured the binding affinity of B. subtilis RNase P mutants for mature tRNA35,36. Wild-type RNase P has a significantly weaker affinity for mature tRNA35,36, compared to pre-tRNA35,36, even at elevated calcium concentrations (Kd,obs=370±80 nM at 3.5 mM Ca2+).21 This binding affinity is largely unperturbed by four mutations in the central cleft of P protein (V32C, S51C, Y34A, and I86C; Kd,obs=250–400 nM), suggesting that the decrease in pre-tRNA affinity is due to disruption of an interaction with the pre-tRNA leader. Conversely, the R60A mutation in the RNR motif decreases the affinity of RNase P for mature tRNA by >8-fold (R60A, Kd,obs=1200 nM; wild-type, Kd,obs=140±100 nM, at 5 mM Ca2+). The fact that this decrease approaches the loss of A(−4) pre-tRNA affinity caused by the R60A mutation (14-fold; Table 2) suggests that the R60 side chain interacts mainly with the tRNA portion of the substrate, and not with the 5′ leader. Overall, these data are consistent with amino acid residues in the central cleft of P protein contributing to affinity and N(−4) specificity through direct contacts with the 5′ leader, while amino acids in the RNR motif likely influence selectivity through positioning of bound tRNA. These differential mechanisms are consistent with the locations of the two regions of the P protein in the structural model of the RNase P pre-tRNA complex.24,40

Y34 and F20 contribute to the recognition of adenosine functional groups at N(−4)

To further probe the possibility of direct interactions between the side chains of P protein residues F20 and Y34 and the N(−4) nucleobase, we measured the binding affinities for RNase P reconstituted with F20A or Y34A P protein mutants for pre-tRNA-containing adenosine analogs at N(−4) (Table 3). As described above, the affinity of wild-type RNase P for A(−4) pre-tRNA is 3-, 4-, and 22-fold higher than that for P(−4), DAP(−4), and G(−4) pre-tRNA, respectively. Relative to the wild-type enzyme, the affinity of F20A and Y34A RNase P enzymes for P(−4) and DAP(−4) pre-tRNA decreases by 3- to 6-fold. The resulting affinities are comparable to that of A(−4) pre-tRNA (Table 3), indicating a loss of both favorable interactions with the N6 position and unfavorable interactions with an N2 exocyclic amine of an adenosine at N(−4)4. The F20 and Y34 side chains may contribute to these interactions either through a direct contact with the N(−4) nucleobase or by contributing to the correct positioning of the 5′ leaders in the enzyme-substrate complex. To distinguish between these possibilities, we examined the role of the hydroxyl of Y34, a possible hydrogen-bonding partner for the N6 amine of an adenosine at N(−4). The Y34F mutation in P protein also decreases A(−4) pre-tRNA binding affinity, although to a smaller extent than the Y34A mutation (Table 3). Furthermore, removal of the Y34 hydroxyl group results in a loss of binding selectivity for A(−4) pre-tRNA compared to the P(−4), DAP(−4), and G(−4) pre-tRNA substrates (Table 3). Overall, the A(−4)/G(−4) selectivity ratio is decreased from 22- to 2-fold for this mutant (Fig. 3). This result suggests that Y34 may form a hydrogen bond with A(−4), perhaps as an acceptor for the N6 exocyclic amine, and that the energetic contribution of this contact may supersede the importance of any stacking interaction between the Y34 phenyl group and the A(−4) purine ring in conferring substrate specificity. Although RNase P is able to bind substrates with any nucleotide at position N(−4), the contact with Y34 confers an energetic preference for A over G at this position, which, in turn, is transduced into the recognition of certain pre-tRNAs. This contact is one of many...
interactions that contribute to the specificity of RNase P for pre-tRNAs and a limited set of other substrates. However, if this sequence preference for A relative to G at N(−4) is also observed in pre-tRNA genes, then this interaction may be crucial for the specific recognition of pre-tRNA substrates by B. subtilis RNase P in vivo.

Analysis of nucleotide enrichments in pre-tRNA gene sequences

In the absence of conflicting evolutionary constraints, sequence-specific interactions between RNase P and pre-tRNA leaders could influence the nucleotide composition of pre-tRNA genes at contacted positions. Similarly, in the presence of multiple factors that influence the selection of pre-tRNA leader sequences, RNase P and pre-tRNA substrate populations may evolve in parallel. In this case, RNase P and pre-tRNA substrates from a given species may form a “matched set” with the molecular recognition properties of the enzyme tuned to a specific substrate population, and these nucleotide preferences may be observable in the sequences of pre-tRNA genes. Evidence indicating that this type of mechanism links RNase P substrate recognition and pre-tRNA gene sequences comes from the interaction between a conserved nucleotide in the J5/15 region of PRNA and the N(−1) position of pre-tRNA. In vitro, this interaction favors a U at N(−1), and the pre-tRNA genes from diverse Bacteria and Archaea show significant enrichment of U at this position.14

To explore the possibility that RNase P selectivity at N(−4) contributes to recognition of pre-tRNA substrates in vivo, we examined whether the nucleotide composition of pre-tRNA genes at N(−4) from B. subtilis and E. coli correlates with binding affinity preferences observed in vitro. Positions in pre-tRNA 5′ leaders from B. subtilis and E. coli, where one or more nucleotides are significantly enriched relative to the background nucleotide composition of the genome, were identified using background-corrected information analysis.41–43 This method quantifies nucleotide enrichments in terms of information (expressed in bits). Results of this type of analysis are commonly represented as “sequence logo” plots, where the total bar height indicates the degree to which a position differs from background and where the relative heights of letters represent the nucleotide composition at the position.41–43 Predicted tRNA genes were identified using tRNAscan SE,10 and leader nucleotides N(−15) through N(−1) were determined. The background nucleotide composition in regions of the genome proximal to pre-tRNA genes was estimated from the combined nucleotide frequencies at positions N(−1) through N(−15) for all tRNA genes. Uncertainties in information content were calculated based on this background distribution,42 and the statistical significance of observed nucleotide enrichments was determined using chi-square analysis.

As shown in Fig. 4, pre-tRNA genes in B. subtilis and E. coli show statistically significant nucleotide enrichments at several positions, including N(−4). In B. subtilis, the base composition at position N(−4) is enriched in A and U with reduced C content and low G content [N(−4): 37% A, 5% G, 13% C, and 45% U; background: 33% A, 11% G, 22% C, and 34% U]. These changes are significant (p = 0.02), even when compared to the high A/U background nucleotide content characteristic of the B. subtilis genome. Qualitatively, increased A and U content and decreased G content are consistent with the relative binding affinities of B. subtilis RNase P for pre-tRNA
substrates with varying sequences at N(−4) (A > U > C > G; Table 1). In *E. coli*, a statistically significant \( p = 10^{-3} \) nucleotide enrichment that results primarily from an increase in C content to 44% compared to the background level of 26% is also observed in the pre-tRNA leader at N(−4). This increase leads to a combined composition of 64% A and C at N(−4), the two nucleotides that confer the highest pre-tRNA binding affinity for *E. coli* RNase P (Table 1). The information content at N(−4) for *E. coli* pre-tRNA \( R_{c,-4} = 0.1 \pm 0.04 \) bit is less than that observed at N(−4) for *B. subtilis* \( R_{c,-4} = 0.22 \pm 0.06 \) bit. This difference indicates a reduced level of sequence conservation at N(−4) in *E. coli*, relative to *B. subtilis*, and is consistent with the overall weaker preference observed at N(−4) for *E. coli* RNase P in *vitro*.

The pre-tRNA genes from both *B. subtilis* and *E. coli* also show additional positions in the leader with significant nucleotide enrichment, suggesting the possibility of additional sequence-specific contacts (Fig. 4). As expected, both species show strong U enrichments at position N(−1) of pre-tRNA genes, which manifests as an information content of \( R_{c,-1} = 0.61 \pm 0.06 \) bit \( (p = 10^{-10}) \) for *B. subtilis* and \( R_{c,-1} = 0.40 \pm 0.03 \) bit \( (p = 10^{-13}) \) for *E. coli*. In *B. subtilis*, an additional enrichment is observed at position N(−2) in the leader, primarily due to an increase in A content to 65%. This enrichment leads to information content and statistical significance similar to those observed at N(−1) \( (R_{c,-2} = 0.58 \pm 0.06 \) bit, \( p = 10^{-15}) \), suggesting that selective pressure, such as recognition by RNase P, influences pre-tRNA sequences at this position. In addition to N(−1) and N(−4), *E. coli* shows statistically significant nucleotide enrichments at N(−2) and N(−6). The increased G and A content at N(−2) of *E. coli* pre-tRNA genes suggests a purine preference at this position, while the increased C content at N(−6) (47% compared to the 26% C content of the genome) suggests recognition at this position as well. Thus, while both *E. coli* and *B. subtilis* pre-tRNA genes display nucleotide enrichments at multiple positions, the observed patterns are markedly different, both with respect to the positions at which enrichments are observed and with respect to the specific nucleotide preferred. Like the preferences determined for substrate binding affinities, this result suggests the possibility of significant diversity in leader sequence recognition by RNase P from different species of Bacteria.

The observation of statistically significant nucleotide enrichments at several locations in the pre-tRNA leader sequences from *E. coli* and *B. subtilis* raises the intriguing possibility that sequence specificity in interactions with 5′ leaders may be a common element of substrate recognition for bacterial RNase P. To explore this possibility, we extended the information content analysis to 161 species of Bacteria with completely sequenced genomes available from GenBank. Overall, 96% of the species show one or more enriched nucleotides between N(−1) and N(−4) with \( p < 0.05 \), and 90% of the species contain at least one enrichment with \( p < 0.01 \). These results suggest that sequence-specific interactions between RNase P and pre-tRNA leaders are likely to be widespread in Bacteria as a whole. Furthermore, the frequency with which nucleotide enrichments are observed at a given leader position increases with proximity to the RNase P cleavage site (Fig. 4c), increasing from 35% at N(−4) to 80% at N(−1). For positions farther from the cleavage site, the number of species with a significant enrichment plateaus at approximately 15%, likely representing a statistical background in this analysis. The increasing prevalence of nucleotide enrichments for positions near the RNase P cleavage site suggests that recognition and cleavage site selection by RNase P may play a role as a driving force in the selection of pre-tRNA leader sequences.

Unexpectedly, the analysis of bacterial tRNA genes suggests that sequence preferences for positions N(−1) through N(−6) in the 5′ leaders of RNase P are diverse, varying in both the position and the identity of the enriched nucleotides. Specifically with regard to N(−4), 52 of the 161 species exhibit statistically significant nucleotide enrichments at this position \( (p < 0.05) \). The nucleotide compositions at N(−4) for pre-tRNA genes from 25 representative species with statistically significant enrichments at this position are shown in Fig. 5 and include species...
that exhibit strong U (e.g., Mycoplasma mycoides), A (e.g., Lactobacillus johnsonii), C (e.g., Salmonella typhimurium), and G (e.g., Thermus thermophilus) enrichments. This diversity suggests that the interaction between N(−4) and RNase P is variable between species and is therefore likely to involve one or more nonconserved enzyme residues, such as Y34 of B. subtilis P protein. In contrast, the composition of N(−1), which interacts with a conserved nucleotide in PRNA, retains an enrichment of U (>25%) in 94% of the species examined. Overall, the diversity of nucleotide enrichments observed for pre-tRNA genes from diverse species suggests that sequence-specific interactions between RNase P and 5′ leaders likely vary markedly between species.

Discussion

Both B. subtilis and E. coli RNase P demonstrate selectivity for the nucleobase at the N(−4) position of the leader that contributes to molecular recognition of pre-tRNA substrates. For B. subtilis RNase P, this interaction results in a 1.9 kcal/mol binding preference for A(−4) over G(−4) pre-tRNA

\[ A_{\text{pre-tRNA}} \] at 2 mM Ca²⁺ in vitro, while the E. coli enzyme shows a moderate preference (0.6–0.9 kcal/mol) for A and C at this position (Table 1). In B. subtilis RNase P, sequence selectivity at N(−4) is consistent with previous analyses suggesting that nucleotides N(−2) to N(−7) interact with the P protein subunit. Furthermore, comparison of the Kₐ,obs values for G(−4) and A(−4) pre-tRNAs (Table 1) reveals that the previously observed enhanced affinity of pre-tRNA substrates with a 5-nucleotide leader (5′-GACAU; G(−5)) relative to that of pre-tRNA substrates with a 4-nucleotide leader (5′-GCAU) can be accounted for by the G-to-A change at N(−4), rather than by an increase in leader length. In fact, comparison of affinities measured for A(−4) and G(−5) pre-tRNA revealed that the addition of a G at N(−5) has a mild destabilizing—rather than a strong stabilizing—effect. Recognition of the nucleobase at N(−4) pre-tRNA by B. subtilis RNase P is therefore consistent with the results of previous studies.

Interactions with P protein subunit lead to specificity at N(−4)

Structure-probing studies have previously suggested that nucleotides N(−3) through N(−7) of the pre-tRNA leader interact with the central cleft of the P protein subunit in the B. subtilis RNase P–substrate complex. Consistent with this, mutagenesis studies indicate that protein residues in the central cleft interact with the nucleotide at N(−4) of pre-tRNA (Fig. 1). In particular, three amino acids play an essential role in the preferential binding of pre-tRNA bearing an adenosine at N(−4): F20, Y34, and R60 (Tables 2 and 3; Fig. 1).

R60 is one of only two invariant residues in bacterial P proteins and is located in the highly conserved region of the protein, the RNR motif (Table 2; Fig. 1). The identification of R60 as important for the sequence-specific recognition of the N(−4) position is unexpected; structural models of the B. subtilis RNase P·pre-tRNA complex place R60 near the tRNA moiety, distal to the leader sequence, and near the P₄ helix of PRNA (Fig. 1). Consistent with this proposal, the R60A mutation decreases the affinity of RNase P for mature tRNA by more than 1.3 kcal/mol, accounting for the majority of the 1.6-kcal/mol decrease in pre-tRNA affinity. These biochemical data argue that R60 has an indirect influence on selectivity at N(−4). The side chain of R60 could either be important for aligning the 5′ leader in the RNase P·pre-tRNA complex such that the nucleobase at N(−4) is positioned to interact correctly with the P protein cleft or be important for stabilizing local PRNA or P protein structure that enhances recognition of pre-tRNA.

The amino acids F20 and Y34 are located on the second and third strands of the central β-sheet, with side chains oriented towards the central cleft that forms the binding pocket for the leader in the RNase P·pre-tRNA structural model (Fig. 1). Alanine substitutions at these residues decrease the binding affinity of A(−4) pre-tRNA by ~1.7 kcal/mol while having a small effect on the affinity of the G(−4) substrate (<0.4 kcal/mol), thereby eliminating selectivity for the nucleobase at N(−4). These data suggest that the side chains of F20 and/or Y34 may form a direct contact with the N(−4) base (Fig. 1) either by base stacking (Phe and Tyr) or by hydrogen bonding (Tyr). Substitution of Phe for Tyr34 has effects similar to those of the Y34A mutation, decreasing the binding affinity of A(−4) pre-tRNA substantially (1.1 kcal/mol) while slightly increasing the affinity of the G(−4) substrate. These data highlight the importance of the hydrogen-bonding potential of Y34 for sequence specificity at N(−4). Additionally, while F20 is conserved in most species, variability is observed at position 34, including a Leu substitution in E. coli RNase P, perhaps contributing to the observed alterations in N(−4) specificity for this enzyme (Table 1).

Model of P protein–adenosine contacts

Substitution of adenosine analogs at N(−4) of pre-tRNA indicates that the A/G selectivity of B. subtilis RNase P involves interactions with two distinct functional groups (Table 1): a favorable interaction with the N6 exocyclic amine of adenosine (−0.7 kcal/mol), possibly due to the formation of a hydrogen bond with the hydroxyl group of Y34, and an unfavorable contact with the N2 exocyclic amine in guanosine (0.9 kcal/mol) (Table 1). Furthermore, RNase P containing the Y34F mutation binds A(−4) and P(−4) pre-tRNAs with comparable affinities, consistent with the loss of a hydrogen bond between the N6 amine and the hydroxyl of Y34. However, this mutation also abrogates the unfavorable interaction with the N2 exocyclic amine (Table 3), leading to an almost complete lack of specificity for the
nucleobase at this position. Therefore, removal of the hydroxyl group of Y34 presumably leads to repositioning of the N(−4) nucleobase and/or protein side chains in the RNase P·pre-tRNA complex, likely mediated by loss of a hydrogen-bond interaction. The F20 mutation also significantly disrupts recognition of the N(−4) nucleobase, since there is little difference between the binding affinities of RNase P containing this mutation for pre-tRNAs with adenosine, purine, or DAP at N(−4). These data indicate that the side chains of both F20 and Y34 are crucial for nucleobase recognition and suggest that both base stacking and hydrogen bonding contribute to selectivity.

One model that is consistent with these data is that the N6 amine of the base at N(−4) forms a hydrogen bond with the hydroxyl group of Y34, while the side chain of F20 orients the N(−4) base for maximal interaction with Y34, perhaps through stacking interactions. Alternatively, hydrogen bonding and hydrophobic interactions may stabilize the P protein fold to optimize interactions with the pre-tRNA leader. In other RNP systems, such as the RNA recognition motif bound to the splicing protein Srp20, sequence-specific recognition has been linked to a direct contact between adenosine and a protein side chain, similar to the model proposed for RNase P. In this case, the exocyclic amine of adenosine forms a hydrogen bond with the side chain of serine, and the base forms a stacking interaction with phenylalanine, leading to a preference for recognizing adenosine relative to guanosine.

**Metal dependence of N(−4) interaction with RNase P**

The magnitude of the selectivity of RNase P for the nucleotide at N(−4) is highly dependent on the concentration of divalent metal ions (Fig. 2), indicating that metal binding and sequence selectivity are coupled. Previous studies have demonstrated that the affinity of the RNase P holoenzyme for pre-tRNA is cooperatively coupled to the binding of divalent cations with a Hill coefficient of ~4. Furthermore, the midpoint for this cooperativity has been shown to depend on the length of the leader sequence; pre-tRNA with a 2-nucleotide leader requires higher Mg2+ concentrations than a substrate with a 5-nucleotide leader. Here the data indicate that the midpoint for activation by divalent cations also depends on the formation of favorable contacts between the nucleobase at N(−4) and P protein. Pre-tRNA association with RNase P occurs in two steps: diffusion-controlled association, followed by a conformational change stabilized by the leader-protein contact. The current data suggest that this conformation change is also stabilized by sequence-specific interactions with the leader sequence. In this context, the steep dependence of the G(−4) pre-tRNA binding affinity on calcium could be explained by stabilization of a conformational change by divalent cations Hsieh et al., manuscript in preparation).

In practical terms, the calcium dependence of substrate selectivity reduces sequence preference under conditions commonly employed in the study of the RNase P holoenzyme in vitro (i.e., 10–20 mM divalent metal ions). However, the physiological concentration of available Mg2+ in Bacteria has been estimated at 2 mM or lower in vitro, concentrations where sequence selectivity is significant (Table 1; Fig. 2). Thus, the coupling of sequence specificity with divalent metal binding results in sequence selectivity at physiological cation concentrations. This coupling leads to the possibility that the sequence selectivity of RNase P in vivo could be modulated by changes in the magnesium concentration. Currently, the regulation of cellular Mg2+ is not well understood. However, intracellular concentrations of Mg2+ are affected by such factors as transport, sequestration in organelles, and binding, particularly to ATP. Therefore, it is possible that the specificity of RNase P is coupled to cellular metabolic activity via Mg2+ concentration.

**tRNA gene analysis**

The correlation between genomic and biochemical data at the N(−4) and N(−1) positions in the leader of pre-tRNA for both B. subtilis and E. coli is consistent with binding affinity preferences observed in vitro playing a functional role in vivo. The interesting observation that a number of pre-tRNAs possess nucleotides at N(−4), conferring reduced affinity (U, C, and G), while adenosine is preferred at N(−4) in the B. subtilis genome is consistent with previous observations and models of substrate recognition in this system. In bacterial RNase P, known interactions with the N(−1) nucleobase and 2’ hydroxyl, the terminal base pair of the acceptor stem, and the 3’-terminal substrate RCCA motif contribute to, but are not required for, recognition and processing. Sun et al. suggested a model of thermodynamic compensation between the 5’ leader sequences and tRNA bodies of diverse substrates. Recently, Sun et al. suggested a model of thermodynamic compensation between the 5’ leader sequences and tRNA bodies of diverse substrates. These investigations revealed that the observed uniform affinity of RNase P for an array of pre-tRNA substrates is maintained by variations in the energetic contributions between RNase P and the 5’ leader that compensate for alterations in the affinity of RNase P for varied tRNAs. Thus, the importance of nucleotide specificity at N(−4) may be variable for pre-tRNA substrates within a given species.

Agreement between in vitro selectivity and pre-tRNA sequence preferences also suggests that examination of pre-tRNA leader sequences is a facile method for developing testable hypotheses regarding the recognition requirements of RNase P from diverse species. In this regard, the 5’ leaders of tRNA genes from both B. subtilis and E. coli exhibit significant nucleotide enrichments at positions other than N(−1) and N(−4) (Fig. 4), consistent with the possibility of sequence-specific contacts involving additional leader nucleotides. Intriguingly, aside from the increased U content at N(−1), B. subtilis and E. coli differ with respect to all other statistically
significant leader nucleotide enrichments, suggesting that, like N(−4), molecular recognition of nucleotides at additional leader positions by RNase P differs between these species. Analysis of pre-tRNA leader sequences from 161 species further demonstrates that functionally important sequence-specific interactions between RNase P and pre-tRNA leaders may exist at positions other than N(−4) and N(−1). The possibility of widespread sequence preferences in 5′ leader interactions represents a novel paradigm for understanding RNase P specificity across multiple species. Systematic examination of the impact of potential interactions between RNase P and additional 5′ leader positions therefore remains a key goal for a better understanding of substrate recognition by RNase P in vivo.

When combined with previous results for naturally occurring E. coli substrates, the correlation of biochemical and genetic data also suggests a model in which RNase P and its substrate population evolve in parallel. In this model, the enzyme and substrates from a given species form a matched set, optimized for uniform recognition and processing of a potentially diverse set of pre-tRNA and other substrates. The observation that species are more likely to have nucleotide enrichments in 5′ leader sequences near the RNase P cleavage site suggests that recognition by RNase P may be a driving force in leader sequence selection. However, this need not be the only driving force in parallel evolution, and additional constraints on leader sequences, such as selection to prevent recognition by other ribonucleases, could influence the composition of the pre-tRNA substrate set and thereby RNase P recognition requirements. An important implication of this model is that, in the absence of a single overarching leader recognition sequence for all Bacteria, useful information regarding enzyme–substrate interactions is most likely to be gained from investigations of cognate RNase P pre-tRNA pairings using naturally occurring 5′ leader sequences.

Conclusion

Sequence preferences observed at N(−4) for substrate affinity are reflected in the composition of pre-tRNA genes from E. coli and B. subtilis, indicating that specificity at N(−4) likely contributes to substrate recognition in vivo. We propose a model for the interaction between B. subtilis RNase P and N(−4) in which the hydroxyl group of the B. subtilis protein residue Y34 hydrogen bonds with the N6 exocyclic amine of adenosine at N(−4). This is the first demonstration of a sequence-specific interaction between the P protein and pre-tRNA that contributes to molecular recognition. The correlation of genomic and biochemical data presented here suggests that analysis of pre-tRNA genes is a viable method for formulating testable hypotheses regarding RNase P substrate recognition in diverse species.

Materials and Methods

RNA and protein preparation

B. subtilis PRNA, E. coli PRNA, and pre-tRNA substrates with 5′-terminal guanosine residues were prepared by in vitro transcription using linearized DNA templates and T7 RNA polymerase in accordance with standard procedures. Substrates were 5′-end-labeled using [γ-32P]ATP (MP Biomedical) and T4 polynucleotide kinase (New England Biolabs). Pre-tRNA substrates with 5′-terminal nucleotides other than guanosine were prepared by RNA ligation. Synthetic 5′ oligoribonucleotides (Thermo/Dharmacon) comprising the 5′ leader and positions +1 through +5 of pre-tRNA were 5′-end-labeled using [γ-32P]ATP and T4 polynucleotide kinase, and ligated to a 3′ tRNA fragment encoding positions +6 through +77 using splinted ligation with T4 DNA ligase (Fermentas), in accordance with standard protocols. Experiments were carried out using B. subtilis pre-tRNA, a widely used substrate that is canonical with respect to all known sequence-specific RNase P recognition elements. P protein was expressed in E. coli (BL21(DE3) pLysS) and purified by CM-Sepharose ion-exchange chromatography in the presence of urea. E. coli P protein was prepared using the Impact expression system (New England Biolabs), as described previously. Variants of the B. subtilis P protein with site-specific cysteine, alanine, and phenylalanine mutations were prepared as described previously. Prior to use, P proteins were dialyzed overnight against the appropriate reaction buffer, and the final concentrations were determined by absorbance (B. subtilis P protein: ε230 = 5120 M−1 cm−1; E. coli C5 protein ε230 = 5500 M−1 cm−1).

PRNA and pre-tRNA were renatured by heating to 95 °C for 3 min in a reaction buffer lacking divalent metal ions, then incubated at 37 °C for 10 min. Divalent metals were added to the appropriate final concentration, and RNAs were incubated for 30 min. To obtain correct [Ca2+] in the presence of high concentrations of PRNA, after renaturation, we diluted PRNA into the reaction buffer with the desired [Ca2+] and recomagnified it using Amicon ultracentrifugal filter units (10,000 molecular weight cutoff; Millipore) for a minimum of three times. RNA concentrations were determined by absorbance. RNase P holoenzyme (0.5–10 μM) was formed by mixing equimolar concentrations of PRNA and P protein and incubating them at 37 °C for a minimum of 1 h.

Substrate affinity measurements

Substrate affinity measurements were carried out by centrifuge gel filtration using Sephadex G-75 resin (Sigma-Aldrich), with total substrate concentrations ([S]tot) maintained at <20% of the total enzyme concentration. Binding affinity was measured in the presence of Ca2+, which supports RNase P substrate binding but decreases the rate at which substrate is converted into product. Enzyme and substrate were renatured separately, mixed, and incubated at 37 °C for 5 min prior to loading onto preequilibrated columns. Columns were centrifuged at 6000 rpm for 30 s, and counts in the filtrate and retentate were measured. Dissociation constants were determined by fitting a single binding isotherm to the data. For [Ca2+] titration experiments, the combined contribution of CaCl2 and KCl to ionic strength was held constant at 410 mM: at
2 mM [Ca\(^{2+}\)]_o [KCl] = 405 mM; at 5 mM [Ca\(^{2+}\)]_o [KCl] = 396 mM. Reactions also included 50 mM 4-morpholino-
neethanesulfonic acid (Mes) and 50 mM tris-(hydroxy-
methyl)aminomethane (Tris) (pH 6.0). For the G(−4) 
substrate, midpoints of transition (K\(_{1/2, Ca}\)) and Hill 
coefficients (n\(_{Ca}\)) were determined by fitting K\(_{d,obs}\) values with Eq. (1), where K\(_{d, bkgd}\) is the dissociation constant at saturating [Ca\(^{2+}\)]:

\[
K_{d,obs} = K_{d, bkgd} \left(1 + \frac{K_{1/2, Ca}}{[Ca(I)]^{n_{Ca}}}\right)
\]

Single-turnover kinetic measurements

Single-turnover experiments were performed in 2 mM 
Co[Cl\(_4\)].Cl\(_2\). 0.5 mM MgCl\(_2\), 0.02 mM NH\(_4\)Cl and 
50 mM Mes–Tris buffer (pH 7.0) at 37 °C using saturating RNase P holoenzyme (2 μM), as previously described. In 2 mM cobalt hexamine, PRNA folds, binds pre-tRNA and P protein, but does not catalyze pre-tRNA cleavage, allowing for the measurement of the single-
turnover rate constant for pre-tRNA cleavage at low 
[Mg\(^{2+}\)]. Substrate and product were separated by PAGE 
and quantified by PhosphorImager (Amer sham Bioscience Corp., Piscataway, NJ); observed rate constants 
(k\(_{obs}\)) were determined by fitting a single-exponential 
equation to the data.

5' Leader sequence analysis of tRNA genes

Completely sequenced bacterial genomes were obtained from GenBank. Predicted tRNA genes were located using tRNAscan SE, and the first 15 nucleotides upstream of each tRNA were identified and compiled using custom AWK scripts. (Scripts available upon request.) Results for species shown in Fig. 4 were adjusted to account for the 8- 
nucleotide acceptor stem of pre-tRNA\(^{His}\). For each 
region, background nucleotide composition specific for the region of pre-tRNA genes was estimated from the 
total nucleotide composition of leader positions N(−15) to N(−1). The information content of individual leader positions relative to background, shown as sequence logos in Fig. 4, was calculated according to Eq. (2),

\[
R_{N(i)} = H_{bkgd} - H_{N(i)} - c_i
\]

where R\(_{N(i)}\) is the information content at position \(i\), H\(_{bkgd}\) is the background entropy from positions N(−15) to N(−15), H\(_{N(i)}\) is the entropy of position \(i\), and \(c_i\) is a sample size correction factor. \(c_i\) was calculated as a function of the number of sequences \(n\) using Eq. (3).

\[
H = - \sum_{i=A,C,G,U} f_i \log_2(f_i)
\]

Error bars in Fig. 4, representing standard deviations for information content, were determined empirically using the CALHNB algorithm of Schneider et al. The probabilities \(p\) that observed nucleotide frequencies appear due to random fluctuations in background 
nucleotide composition were determined for each position using standard chi-square analysis.

Acknowledgements

We would like to thank Dr. David Engelke and members of the Fierke group, including Dr. John Hsieh, Dr. James Hougl and, Dr. Terry Watt, for careful reading of this manuscript and helpful discussion. We also thank Dr. Hsieh for providing B. subtilis P protein. This work was funded by National Institutes of Health (NIH) grant GM55387 to C.A.F. and grant GM56742 to M.E.H. N.H.Z. was partially supported by NIH postdoctoral fellowship ES013881, and K.S.K. was partially supported by NIH training grant GM08353.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.11.039

References

Importance of RNA–protein interactions in bacterial 
ribonuclease P structure and catalysis. Bio polymers, 87, 
329–338.
1183–1194.
3. Christian, E. L., Zahler, N. H., Kaye, N. M. & Harris, 
M. E. (2002). Analysis of substrate recognition by the 
ribonucleoprotein endonuclease RNase P. Methods, 28, 
307–322.
mechanism for cleavage of precursor tRNA(Asp) 
catalyzed by the RNA component of Bacillus subtilis 
ribonuclease P. Biochemistry, 33, 10294–10304.
5. Sun, L., Campbell, F. E., Zahler, N. H. & Harris, M. E. 
protein lead to uniformity in binding and catalysis by 
RNase P. EMBO J., 25, 3998–4007.
the processing of the precursor to 4.5 S RNA, a naturally 
occurring substrate for RNase P from Escherichia coli. 
specificity of cleavage by RNase P is mainly depend-
ent on the structure of the amino acid acceptor stem. 
substrates for an RNA enzyme in the presence and 
function of a ribonucleoprotein enzyme: the catalytic 
domain of Bacillus subtilis RNase P complexed with B. 
subtilis RNase P protein. Nucleic Acids Res., 29, 
1892–1897.
Recognition of 5′ Leaders by RNase P

207


P is the catalytic subunit of the enzyme. *Cell*, 35, 849–857.


