

# Determination of the Specificity Landscape for Ribonuclease P Processing of Precursor tRNA 5' Leader Sequences

Courtney N. Niland,<sup>†</sup> Jing Zhao,<sup>†</sup> Hsuan-Chun Lin,<sup>†</sup> David R. Anderson,<sup>‡</sup> Eckhard Jankowsky,<sup>§</sup> and Michael E. Harris<sup>\*†</sup>

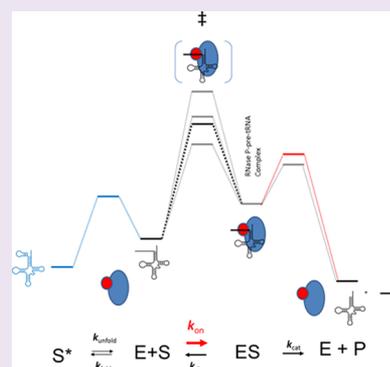
<sup>†</sup>Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, United States

<sup>‡</sup>School of Business, CUNY Baruch College, New York, New York 10010, United States

<sup>§</sup>Center for RNA Molecular Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, United States

## Supporting Information

**ABSTRACT:** Maturation of tRNA depends on a single endonuclease, ribonuclease P (RNase P), to remove highly variable 5' leader sequences from precursor tRNA transcripts. Here, we use high-throughput enzymology to report multiple-turnover and single-turnover kinetics for *Escherichia coli* RNase P processing of all possible 5' leader sequences, including nucleotides contacting both the RNA and protein subunits of RNase P. The results reveal that the identity of N(−2) and N(−3) relative to the cleavage site at N(1) primarily control alternative substrate selection and act at the level of association not the cleavage step. As a consequence, the specificity for N(−1), which contacts the active site and contributes to catalysis, is suppressed. This study demonstrates high-throughput RNA enzymology as a means to globally determine RNA specificity landscapes and reveals the mechanism of substrate discrimination by a widespread and essential RNA-processing enzyme.



To function in the cell, most RNA-processing enzymes and ribonucleases must recognize many alternative RNA substrates. Multiple substrate recognition is an inherent and essential function characteristic of complex ribonucleoprotein machines such as the spliceosome, ribosome, and enzymes involved in processing tRNAs, snRNAs, and siRNAs. A major challenge they all face is to recognize their cognate substrates among the excess of noncognate binding sites in the transcriptome. Therefore, a complete understanding of RNA metabolism and gene expression requires characterizing in detail the substrate RNA sequences and structures that drive enzyme specificity. Achieving a comprehensive understanding of the mechanistic basis for substrate discrimination by RNA-processing enzymes is challenging but important for realizing their potential as drug targets<sup>1–3</sup> and platforms for engineering synthetic biology.<sup>4–6</sup>

Ribonuclease P is a ubiquitous and essential tRNA-processing endonuclease that generates the mature 5' end of tRNA by removal of the 5' leader sequence.<sup>7</sup> In bacteria, RNase P is composed of a large catalytic RNA subunit (P RNA) and a smaller protein subunit, both of which are responsible for substrate recognition. In *Escherichia coli*, a single RNase P enzyme must process all 87 pre-tRNAs, which vary greatly in sequence and length of their 5' leaders with some base conservation observed only at nucleotides proximal to the cleavage site<sup>8,9</sup> (Figure 1A). Despite this lack of 5' leader sequence conservation, experimental approaches such as cross-linking, chemical protection, mutagenesis, as well as X-ray crystallography of the *Thermotoga maritima* RNase P-product

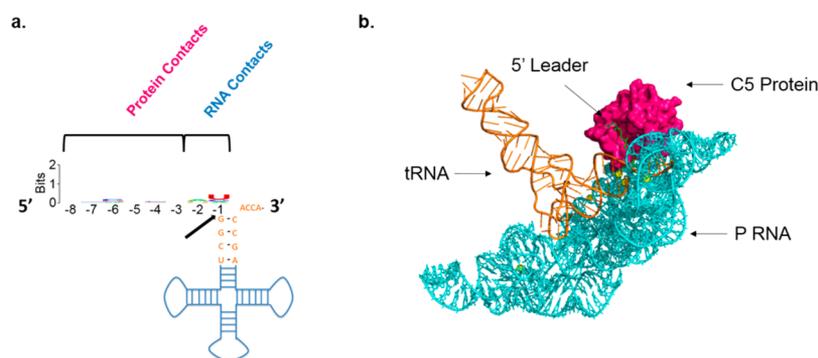
complex, demonstrate both the protein subunit and RNA subunit make direct contact with 5' leader nucleotides N(−8) to N(−1).<sup>10–21</sup> Comparison of the kinetics of genomically encoded *E. coli* pre-tRNAs reveals that they react *in vitro* with essentially equivalent  $k_{\text{cat}}/K_m$  values.<sup>16,21</sup> Similar multiple-turnover reaction kinetics are due, in part, to variation in the strength of 5' leader sequence interactions that may compensate for weaker affinities of different tRNAs.<sup>8,21</sup> Bioinformatic analyses suggest that sequence-specific contacts between the protein subunit and the 5' leader sequences of pre-tRNAs may be common in bacterial RNase P and may lead to species-specific substrate recognition.<sup>9</sup>

Current structure models of pre-tRNA bound to bacterial RNase P indicate that the two nucleotides, N(−2) and N(−1), immediately 5' to the cleavage site interact directly with the catalytic P RNA subunit consistent with biochemical data. The N(−1) nucleobase interacts with a universally conserved adenosine in the J5/15 region of P RNA.<sup>14,15,22</sup> N(−2) is proposed to interact with J18/2 although the chemical details of this interaction are not defined.<sup>14,19,23</sup> More distal 5' leader sequences N(−8) to N(−3) are contacted by the small, but essential protein subunit (termed C5 in *E. coli*).<sup>24–26</sup> Biochemical and biophysical data show that the 5' leader binds to a cleft on the surface of P protein in an extended conformation.<sup>13,26,27</sup> An X-ray crystal structure of *T. maritima*

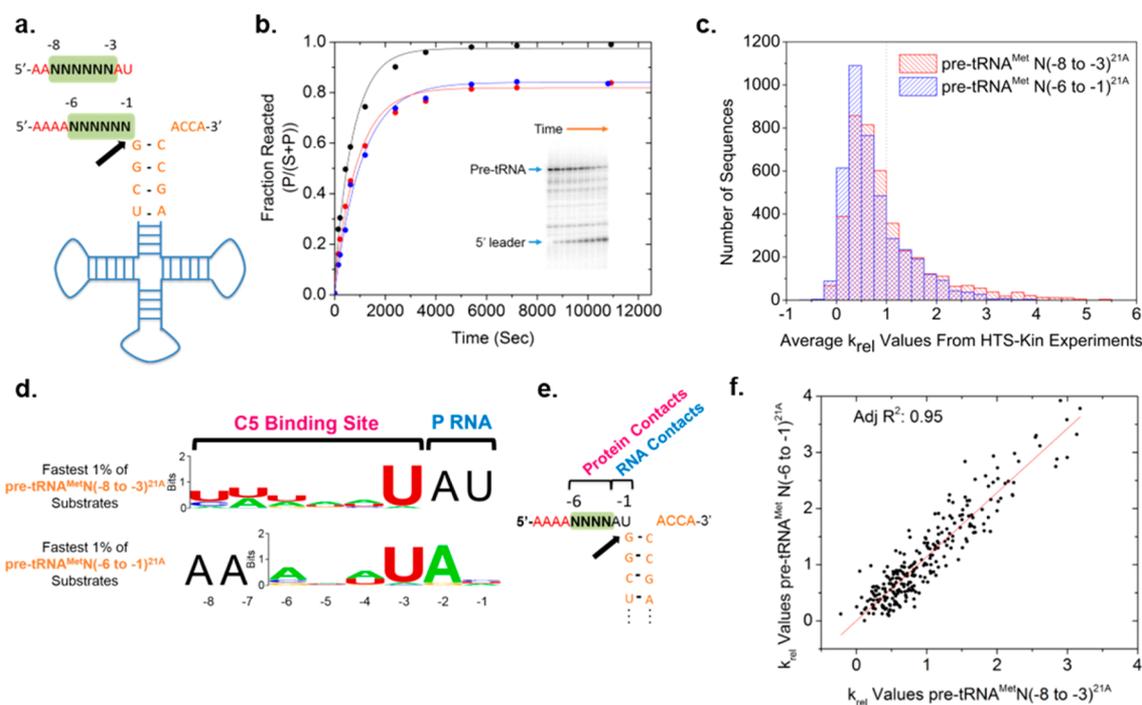
Received: March 23, 2016

Accepted: June 7, 2016

Published: June 23, 2016



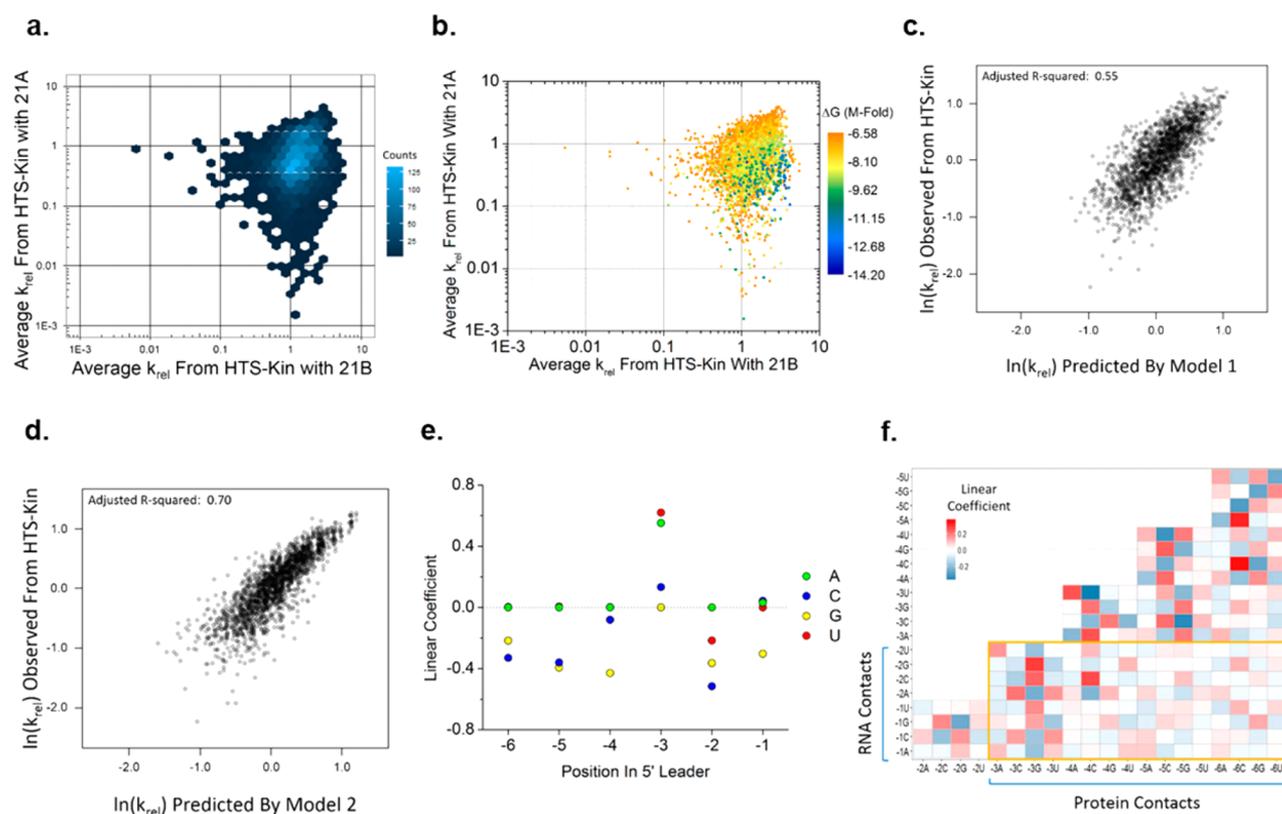
**Figure 1.** Recognition of the pre-tRNA 5' leader sequence by the RNA and protein subunits of ribonuclease P. (a) A sequence alignment of all 5' leaders in *Escherichia coli* pre-tRNAs reveals no identifiable conserved sequence motif. (b) X-ray crystal structure of the *Thermotoga maritima* RNase P-product complex from Reiter et al. P RNA subunit is shown in cyan, protein subunit in pink, tRNA body in orange, and green 5' leader.



**Figure 2.** Determination of the effects of all possible pre-tRNA 5' leader sequence variants on the relative  $k_{\text{cat}}/K_m$  for RNase P processing. (a) The six nucleotides randomized in pre-tRNA<sup>Met85</sup>N(-6 to -1) and pre-tRNA<sup>Met85</sup>N(-8 to -3) substrate pools are indicated by a green box, RNase P cleavage site indicated with arrow. (b) Multiple-turnover kinetics of genomically encoded pre-tRNA<sup>Met</sup>(WT)<sup>21A</sup> substrate (black) to the randomized pre-tRNA<sup>Met</sup>N(-8 to -3)<sup>21A</sup> (red) and pre-tRNA<sup>Met</sup>N(-6 to -1)<sup>21A</sup> (blue) populations. The inset shows typical results from RNase P reactions run on a denaturing polyacrylamide gel, each lane shows a time point that separates substrate and product. (c) Histogram of the number of substrate variants with a particular  $k_{\text{rel}}$  value from HTS-Kin showing the average of three experiments with the pre-tRNA<sup>Met</sup>N(-6 to -1) and two with pre-tRNA<sup>Met</sup>N(-8 to -3). (d) Sequence probability logos identify sequence preference in the fastest 1% of substrates in the HTS-Kin reactions. Black letters show the wildtype sequence and indicate nonrandomized positions. (e) The location of the 256 sequences that are in common between the randomized pre-tRNA<sup>Met</sup>N(-8 to -3)<sup>21A</sup> and pre-tRNA<sup>Met</sup>N(-6 to -1)<sup>21A</sup> populations at positions N(-6) to N(-3) are indicated by a green box. (f) The observed  $k_{\text{rel}}$  values from the two separate independent experiments and distinct randomized pools are plotted and fit to a linear function, red line.

RNase P with a 5' leader sequence oligonucleotide soaked into the crystal is consistent with these intimate contacts<sup>28</sup> (Figure 1B). *In vitro* structure–function studies demonstrate that sequence variation at 5' leader nucleotides involved in both RNA–RNA and RNA–protein interactions can have significant functional effects on RNase P processing rates as well as cleavage site specificity.<sup>9,21,29</sup> For example, analysis of 5' leader point mutations and amino acid substitutions in P protein support favorable hydrogen bonding interactions between an adenosine at N(-4) and the protein subunit of *Bacillus subtilis* RNase P; however, the sequence preference of *E. coli* RNase P

at this position is diminished.<sup>9</sup> Additionally, mutation of the 5' leader at N(-1) showed altered processing rates and that are partly rescued by compensatory mutations P RNA in J5/15.<sup>15,30</sup> At present, the 5' leader sequence specificity of RNase P enzymes and the manner in which sequence variation affects the reaction mechanism are not well-defined. As a consequence, we lack the information necessary to predict the effects of variation in the sequences of genomically encoded tRNA precursors *in vivo* and the understanding to identify the range of cognate substrates in the transcriptome.



**Figure 3.** Quantitative modeling of 5' leader sequence specificity of RNase P. (a) Density plot of average  $k_{rel}$  for each substrate variant in three HTS-Kin reactions in the 21A context compared to two reactions with the 21B substrate context. The plot shows the number of substrates in each hexagonal region indicated by the scale at right (less than 1% of data omitted for clarity). (b) Plot of average  $k_{rel}$  for 5' leader variants in the context of 21A or 21B sequence. Individual points are colored by their M-Fold predicted highest folding free energy between the 5' leader sequence and 21A sequence (less than 1% of data omitted for clarity). (c) PWM model used to describe the HTS-Kin data evaluated by plotting the observed  $k_{rel}$  from HTS-Kin against the predicted  $k_{rel}$  from the model. (d) A PWM model with an included coupling term used to describe the data evaluated by plotting the observed  $k_{rel}$  from HTS-Kin against the predicted  $k_{rel}$  of the model. (e) Linear coefficients for each nucleotide, indicated by color, from the PWM portion of model 2 shown for each position in the 5' leader indicated at the bottom. (f) Heatmap of coupling coefficients from the model. Position and nucleotide identity indicated on each axis and the coupling coefficient indicated at the vertex.

As an initial step toward addressing this problem, we determined the sequence specificity of *E. coli* RNase P for 5' leader sequence positions N(−8) to N(−3) in a noninitiator pre-tRNA<sup>Met82</sup>, which includes binding sites for both the RNA and protein subunits.<sup>31</sup> Previously, we measured the relative  $k_{cat}/K_m$  values for RNase P processing of all possible substrate variations in the C5 binding site using high-throughput sequencing kinetics (HTS-Kin). HTS-Kin combines Illumina sequencing and internal competition kinetic analysis to measure the relative second-order rate constant ( $k_{cat}/K_m$ ) for thousands of alternative RNA substrates simultaneously. Quantitative modeling of the resulting rate-constant distribution revealed the functional sequence specificity of C5 protein. However, the specificity for nucleotides proximal to the cleavage site that interact with P RNA, as well as the mechanism by which discrimination between alternative substrates is achieved remain unknown.

Here, we determine the effects of all possible sequence variants at N(−6) to N(−1) in the 5' leader that contact the P RNA and C5 protein subunits on both multiple-turnover and single-turnover reaction kinetics. The observed specificity of RNase P under multiple-turnover conditions reveals N(−2) and N(−3) as the key determinants of alternative substrate discrimination. Moreover, we find that competition for alternative substrates occurs at the level of substrate association,

not catalysis, and that this conceals specificity for N(−1) at the cleavage step. Thus, a combined approach of high-throughput RNA enzymology and quantitative modeling of sequence specificity provides a general means to globally determine RNA specificity landscapes. Taken together, the results reveal the key specificity determinants for an essential RNA-processing enzyme and support a general specificity landscape and mechanism by which discrimination is achieved.

## RESULTS AND DISCUSSION

**RNase P Processing of Pre-tRNA Substrate Pools.** To characterize the specificity of *E. coli* RNase P for pre-tRNA 5' leader sequences, we randomized nucleotides N(−6) to N(−1) that include positions interacting with the RNA subunit (N(−2) and N(−1)), as well as N(−6) to N(−3) that interact with the C5 protein subunit, providing 4096 substrate variants (Figure 2A).<sup>15,17–21,32</sup> Analysis of multiple-turnover kinetics of the pre-tRNA<sup>Met</sup> N(−6 to −1) randomized population demonstrated that this pool of substrates reacts with kinetics that are overall slower than the native pre-tRNA<sup>Met82</sup> with its genomically encoded 5' leader sequence (Figure 2B). Similarly slow reactivity was observed for the pre-tRNA<sup>Met</sup> N(−8 to −3) substrate population randomized at positions N(−8) to N(−3) as reported in Guenther et al.<sup>31</sup> Thus, randomization of the 5' leader from N(−6) to N(−1) or from N(−8) to N(−3) affects

RNase P multiple-turnover processing kinetics similarly, likely due to their commonly randomized region (N(-6) to N(-3)).

Next, the relative  $k_{\text{cat}}/K_m$  for all 4096 sequences in the pre-tRNA<sup>Met</sup> N(-6 to -1) randomized population were measured using HTS-Kin. HTS-Kin measures changes in the relative abundance of alternative RNA substrates over reaction time using next-generation sequencing, and these data are used to calculate relative rate constants using internal competition kinetics.<sup>33–36</sup> A relative second-order rate constant,  $k_{\text{rel}}$ , was calculated for each substrate that is normalized to the rate constant for the genomically encoded sequence (i.e.,  $k_{\text{rel}} = (k_{\text{cat}}/K_m(\text{variant})) / (k_{\text{cat}}/K_m(\text{AAAGAU}))$ ). A histogram showing the distribution of average relative rate constants for all substrate variants from three replicate experiments is shown in Figure 2C. This histogram defines a rate-constant distribution that describes the entire range of effects of sequence variation on  $k_{\text{cat}}/K_m$ . The rate-constant distribution for the pre-tRNA<sup>Met</sup> N(-6 to -1) population is similar in shape to that previously observed for pre-tRNA<sup>Met</sup> N(-8 to -3) randomized population.<sup>31</sup> A significant number of sequences (ca. 200) in the pre-tRNA<sup>Met</sup> N(-6 to -1) population reacted with  $k_{\text{rel}}$  greater than 2-fold faster than the reference, indicating that the genomically encoded 5' leader sequence at N(-6) to N(-1) is not optimal for  $k_{\text{cat}}/K_m$ .

**Modeling of RNase P Substrate Specificity.** To identify the sequence determinants optimal for  $k_{\text{cat}}/K_m$ , we calculated sequence probability logos from 5' leader substrate variants with the top 1% of  $k_{\text{rel}}$  values<sup>37</sup> (Figure 2D). Surprisingly, the results indicated little sequence preference at N(-1) and N(-4) which were previously indicated to contribute to specificity.<sup>9,15,30</sup> In contrast, N(-2) and N(-3) showed strong preferences for adenosine and uridine, respectively. Optimal sequence determinants in the protein binding site (-6 to -3) of the 5' leader are similar to those previously observed for these positions in HTS-Kin results obtained using the pre-tRNA<sup>Met</sup> N(-8 to -3) population.<sup>31</sup> A subset of 256 sequence variants is common to both the pre-tRNA<sup>Met</sup> N(-6 to -1) and pre-tRNA<sup>Met</sup> N(-8 to -3) randomized populations (Figure 2E). As an internal check on accuracy, we compared the  $k_{\text{rel}}$  values measured for these sequences in the two independent randomized populations (Figure 2F). The results show that the measured  $k_{\text{rel}}$  in both HTS-Kin experiments are highly correlative, demonstrating the reproducibility of the technique.

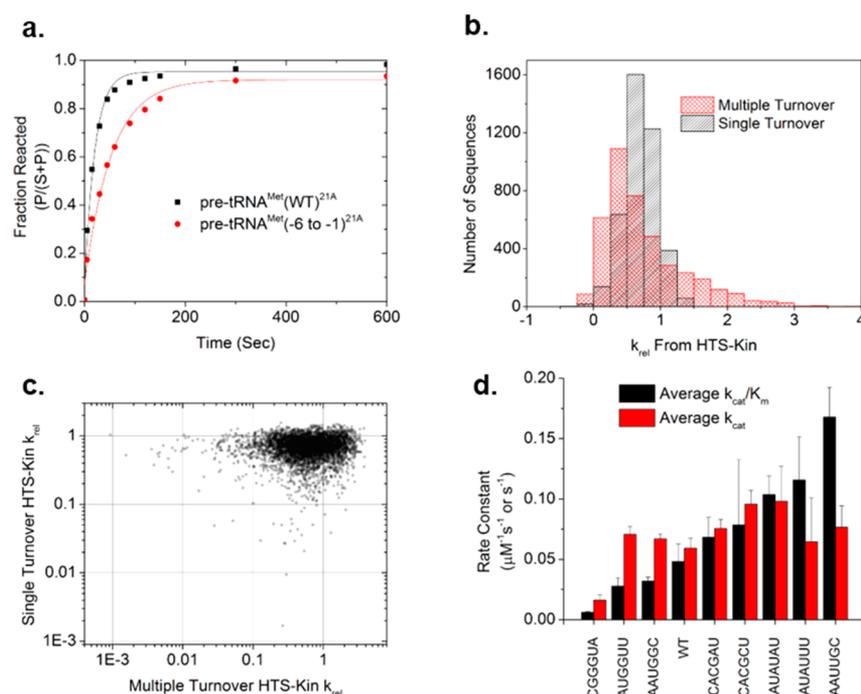
To amplify pre-tRNAs for high-throughput sequencing without ligation bias, an additional 21 nucleotides were added to the terminal 5' end of pre-tRNA<sup>Met</sup> N(-6 to -1) substrates. Cross-linking, X-ray crystallography, and FRET studies demonstrate that the 5' leader sequence is bound in an extended single-stranded conformation,<sup>13,26,27</sup> therefore these additional 21nt (termed 21A) could result in formation of unfavorable secondary structure in the ground state that would complicate identification of intrinsic RNase P sequence specificity. We tested for these effects by comparing the  $k_{\text{rel}}$  values determined for a second pre-tRNA<sup>Met</sup> N(-6 to -1) population in which each nucleotide in the 21nt upstream sequence was changed to its Watson–Crick complement (termed 21B). Differences in  $k_{\text{rel}}$  value for the same 5' leader substrate variant in the two different contexts identify potential instances of inhibitory secondary structure. Comparison of the rate-constant distributions for the pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21A</sup> and pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21B</sup> populations (Figure 3A) shows that the majority of substrate processing rates correlate between the 21A and 21B substrate backgrounds. However, a

subset of the population of N(-6) to N(-1) substrate variants reacts with greater than 2-fold slower kinetics in the 21A leader sequence context compared to its complement 21B.

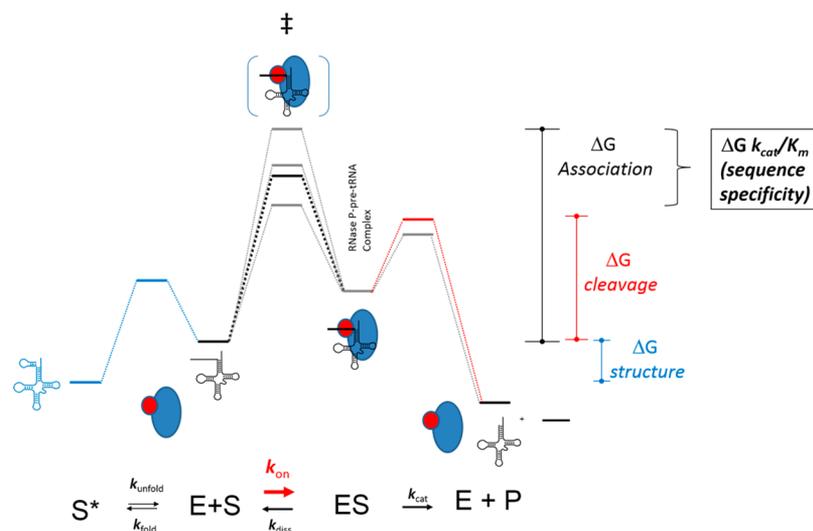
To test whether inhibitory secondary structure explains the deviation of rate constants of some substrates between the 21A and 21B populations, M-Fold was used to calculate the predicted folding free energies between the 21A sequence and 5' leaders.<sup>38</sup> An overlay of these predicted folding free energies on a plot comparing the rate-constant distributions for the pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21A</sup> and pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21B</sup> populations is shown in Figure 3B. The 5' leader sequences predicted to form the most stable secondary structure in the presence of the 21A context react with slower kinetics and show the greatest degree of displacement from the rate constant measured in the 21B context. Thus, a portion of substrates are affected in the ground state in a manner independent of the intrinsic sequence specificity of RNase P for N(-6) to N(-1). To isolate effects of 5' leader sequence variation on RNase P processing independent of upstream sequence context, we selected and averaged the data for those substrate variants with  $k_{\text{rel}}$  values that varied by less than 2-fold between the 21A and 21B contexts for further analysis of intrinsic sequence specificity.

To comprehensively and quantitatively determine the specificity of RNase P for 5' leader sequences N(-6) to N(-1), we first applied a simple position weight matrix (PWM) model. PWM models are commonly used to describe DNA-binding proteins<sup>39</sup> and consider each position in the 5' leader as independent and therefore uncoupled with other positions in the binding site. Fitting the HTS-Kin data to a PWM model and comparing the observed  $k_{\text{rel}}$  to that predicted by the model indicates this model only explains about half of the correlation between observed rate constant and 5' leader sequence (Figure 3C). Additionally, the data were fit to a PWM model including an additional term that quantifies coupling between positions in the binding site of the 5' leader. Including quantitative coupling coefficients ( $\beta$  values) in the model, which allows the sequence identity at each position to modulate the effects of sequence variation at other positions in the binding site, provides a significantly better fit to the experimental data (Figure 3D). Importantly, similar results were obtained from fitting the complete pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21A</sup> or pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21B</sup> data sets (Figure S1). This result indicates that the inaccuracies introduced by inhibitory secondary structure are relatively minor compared to the dominant contribution from intrinsic RNase P sequence specificity and reflects the overdetermination of these parameters by HTS-Kin. Contributions to specificity are greatest at positions N(-2) and N(-3), consistent with the high probabilities of uridine and adenosine at these positions in the sequence probability logo (Figure 3E). The  $\beta$  values expressing the degree of coupling between positions are shown in a heatmap in Figure 3F. The strongest coupling occurs between adjacent nucleotides. Interestingly, relatively strong coupling coefficients are observed between the RNA binding site and the proximal protein binding site, in particular between the N(-2) and N(-3) positions. This result indicates coupling at the single nucleotide level between the energetic contributions of nucleobases in the pre-tRNA 5' leader that contact P RNA and those that contact C5 protein.

**Specificity Landscape of Substrate Processing.** The magnitude of  $k_{\text{cat}}/K_m$  reflects the first irreversible step, which is substrate association for the kinetic mechanism of pre-



**Figure 4.** Effect of sequence variation in the 5' leader is primarily on substrate affinity and not on catalysis. (a) Single-turnover reactions performed at saturating enzyme conditions with pre-tRNA<sup>Met</sup>(WT)<sup>21A</sup> or pre-tRNA<sup>Met</sup> N(-6 to -1)<sup>21A</sup> from two and three experiments, respectively. (b) Histogram of the distribution of  $k_{rel}$  of substrates in a single-turnover HTS-Kin reaction compared to averaged multiple-turnover HTS-Kin reactions. (c) Comparison of  $k_{rel}$  for each substrate variant (individual points on the graph) in averaged multiple-turnover and a single-turnover HTS-Kin. (d) Single substrate reactions were used to determine the effect of 5' leader sequence variation on the later steps in the reaction, (i.e., E-S → E + P) or the second-order rate constant (i.e., E + S → E + P). The  $k_{cat}$  ( $\text{s}^{-1}$ ) from multiple-turnover reactions for each 5' leader sequence listed on the x-axis from at least three experiments (red). The absolute  $k_{cat}/K_m$  ( $\mu\text{M}^{-1}\text{s}^{-1}$ ) measured for these sequences is shown from at least three experiments (black). Standard deviation is indicated by the error bars.



**Figure 5.** Reaction free-energy profile for RNase P processing of pre-tRNA<sup>Met82</sup> and effects of sequence variation defining the specificity landscape. In this mechanism, the pre-tRNA substrate combined with RNase P (P RNA is shown as a blue oval and P protein a smaller red sphere) to form the RNase P-pre-tRNA complex which undergoes turnover to form the tRNA and 5' leader sequence products and free RNase P. For the genomically encoded pre-tRNA<sup>Met82</sup> leader sequence,  $k_{cat}$  is fast relative to dissociation, and the magnitude of  $k_{cat}/K_m$  reflects association. Differences in the magnitude of  $k_{cat}/K_m$  for alternative pre-tRNAs arise from possible effects of unfavorable structure in the ground state ( $\Delta G$  structure), differences in interactions with the 5' leader sequence that affect the transition state for association ( $\Delta G$  RNA-protein), or large changes in the cleavage rate constant that also affect the magnitude of  $k_{cat}$  ( $\Delta G$  cleavage).

tRNA<sup>Met82</sup>. Processing of pre-tRNA<sup>Met</sup> with its genomically encoded 5' leader sequence involves slow dissociation relative to the RNA cleavage step.<sup>8,21</sup> However, the pre-tRNA<sup>Met</sup> N(-6 to -1) randomized substrate pool contains all sequence

combinations at positions N(-2) and N(-1) that are sites of interaction with the catalytic P RNA subunit of RNase P and could affect the substrate cleavage step. Although the protein subunit of RNase P contributes little to catalysis for optimal

pre-tRNA substrates, the extent to which RNA–protein interactions broadly contribute to catalysis is not known. Additionally, it is not known whether catalytically important contacts between the 5′ leader and the P RNA subunit are stabilized or destabilized by the more distal protein contacts.

To determine the extent to which randomization of N(−6) to N(−1) affects catalysis, we performed single-turnover reactions under saturating enzyme conditions in which the observed rate constant reflects the substrate cleavage step. Similar to the multiple-turnover kinetics, the pre-tRNA<sup>Met</sup> N(−6 to −1)<sup>21A</sup> randomized pool under single-turnover conditions reacted with slower overall kinetics compared to the genomically encoded reference substrate; however, the substrate population reacted to completion (Figure 4A). This result is consistent with smaller overall effects of sequence variation on catalysis, or a smaller number of substrates for which this parameter is affected.

To distinguish these possibilities, we measured the relative single-turnover rate constants for all substrate variants in the pre-tRNA<sup>Met</sup> N(−6 to −1)<sup>21A</sup> population by performing HTS-Kin under saturating enzyme conditions. An overlay of histograms showing the observed rate-constant distributions from multiple-turnover versus single-turnover reactions with pre-tRNA<sup>Met</sup> N(−6 to −1)<sup>21A</sup> is shown in Figure 4B. A much narrower distribution of  $k_{rel}$  values is observed for single-turnover conditions demonstrating that few substrate variants in the 5′ leader alter the cleavage step, while the same substrates result in a much greater range of observed  $k_{cat}/K_m$  values. For those substrates in which the single-turnover rate constant was significantly affected, there are two possible interpretations: the catalytic rate constant may be reduced due to variation in the 5′ leader or these substrates may bind more weakly to RNase P and thus have a  $K_m$  above the concentration used. A plot of observed  $k_{rel}$  values for substrates in multiple-turnover reactions versus single-turnover reactions reveals little correlation between the effects of 5′ leader variation and further illustrates the much smaller range of effects on the cleavage step (Figure 4C). The sequence specificity for cleavage under single-turnover conditions can also be modeled. Fitting the single-turnover HTS-Kin data to the PWM model with coupling coefficients described above, identifies N(−1) as the primary contributor to specificity (Figure S2). The magnitude of the calculated PWM scores shows that uridine is optimal at N(−1) consistent with previous studies of RNase P catalysis and the current model of the ES complex.<sup>15</sup>

Thus, comparison of the multiple-turnover and single-turnover rate constants suggests a specificity landscape for 5′ leader sequence discrimination by *E. coli* RNase P (Figure 5) that globally describes the effects of sequence variation on the reaction mechanism. In this model, variation in proximal 5′ leader sequences have relatively small effects on the cleavage step and therefore are likely to also have small effects on  $k_{cat}$ . In contrast, sequence variation at N(−2) and N(−3) alters  $k_{cat}/K_m$ , which controls the competition between alternative substrates. We tested the predicted effects on the  $k_{cat}/K_m$  and  $k_{cat}$  values for a series of single pre-tRNA substrates selected to span the nearly 100-fold range in  $k_{rel}$  as measured by HTS-Kin. As expected, the observed  $k_{cat}/K_m$  values measured using individual substrate reactions showed significant variation based on 5′ leader sequence identity and correlate well with the  $k_{rel}$  from HTS-Kin (Adj.  $R^2 = 0.81$ ) (Figure 4D and Figure S3). The results further demonstrate HTS-Kin as an accurate and reliable method for rapid and comprehensive determination of

relative rate constants. In contrast, the  $k_{cat}$  values measured individually for the same substrates were very similar and varied less than 2-fold. An exception is the slowest reacting 5′ leader variant that is predicted to have inhibitory secondary structure between the 5′ leader and 21A and resulted in decreases in both  $k_{cat}$  and  $k_{cat}/K_m$ . For the majority of substrates, the primary effect of 5′ leader sequence variation is on  $k_{cat}/K_m$  with minimal effects on the substrate cleavage step and consequently minimal effects on  $k_{cat}$ .<sup>8,21</sup>

Previous mutagenesis and cross-linking experiments clearly show a U(−1) makes an optimal direct contact to a conserved adenosine residue in J5/15 of P RNA. However, the insensitivity of  $k_{cat}/K_m$  to the identity of the N(−1) nucleobase is apparent from both the HTS-Kin data and confirmed by single-substrate assays.<sup>15,22,30,40</sup> Insensitivity to sequence variation at N(−1) observed in HTS-Kin indicates formation of contacts with this position occur at a step subsequent to the first irreversible step, which for the genomically encoded 5′ leader sequence is association. As a consequence, variation at N(−1), which contributes primarily to the cleavage step, does not significantly affect the observed  $k_{cat}/K_m$ . Conversely, N(−2) and N(−3) contribute significantly to RNase P specificity as demonstrated by optimal sequence probability logos and quantitative modeling of the HTS-Kin data.

The resulting sequence specificity model is likely to have important implications for substrate processing in the cell. The distribution of  $k_{rel}$  values predicted for the 87 different 5′ leader sequences encoded in the *E. coli* genome show significant variation in magnitude occurring throughout the rate-constant distribution (Figure S4). The fastest reacting 5′ leader sequences matched those in proline, glycine, and alanine pre-tRNAs, which have adenosine or uridine at N(−2) and uridine at N(−3) and thus are predicted to be optimized for  $k_{cat}/K_m$  based on the HTS-Kin results. The slowest reacting 5′ leaders predicted by these results are found in tyrosine, methionine, and arginine which lack at least one of these specificity determinants. Importantly, there are contacts between the P RNA subunit of RNase P and the tRNA body that may also modulate the overall enzyme specificity for these substrates *in vivo*.<sup>23,41–43</sup>

The general model for tRNA biosynthesis in *E. coli* suggests that the rate-limiting step occurs at the level of transcription. For some substrates the presence of specificity determinants in the 5′ leader sequence may act to offset favorable or unfavorable interactions with the tRNA portion of the substrate to maintain uniform rates of 5′ end maturation. However, more recent molecular genetic analyses reveal that RNase P processes multiple polycistronic transcripts in which the order of processing occurs in a strict 3′ to 5′ direction.<sup>29</sup> Thus, for some substrates, the presence or absence of specificity determinants may modulate cleavage relative to other competing cognate binding sites in the cell. The availability of a comprehensive model for 5′ leader sequence specificity allows a more accurate evaluation of the potential differences in molecular recognition *in vivo*.

This study demonstrates the use of high-throughput RNA enzymology methods using common molecular biology techniques and instrumentation as a general means to globally determine the specificity landscape for molecular recognition by an essential RNA-processing enzyme. The application of HTS-Kin to measure both multiple-turnover and single-turnover kinetic parameters for thousands of RNA substrate variants reveals the full range of effects and the intrinsic

sequence specificity for these two kinetic parameters. The ability to measure different kinetic parameters for the same sequence variants allows the construction of simple kinetic schemes for each substrate and insight into how variation alters the free-energy landscape of the reaction. Analysis of the high-throughput biochemical data using quantitative models of sequence specificity provides a way to identify key specificity determinants and quantitatively compare results obtained between experiments. Therefore, the general approach described here for characterizing the specificity of RNase P presents a rigorous and comprehensive way to investigate RNA specificity that is likely to be applicable to a range of experimental systems.

## METHODS

**Multiple-Turnover and Single-Turnover Reactions.** *E. coli* C5 protein was overexpressed and purified as described.<sup>44</sup> P RNA and pre-tRNA<sup>Met82</sup> were synthesized by *in vitro* transcription using T7 RNA polymerase and PCR or cloned DNA templates. Pre-tRNA substrates were 5' end labeled with  $\gamma$ -<sup>32</sup>P using polynucleotide kinase. Multiple-turnover reactions were performed in 50 mM Tris-HCl pH 8, 100 mM NaCl, 0.005% Triton X-100, and 17.5 mM MgCl<sub>2</sub>. RNase P was assembled by heating P RNA to 95 °C for 3 min then 37 °C for 10 min. MgCl<sub>2</sub> was added and incubated at 37 °C for 10 min before adding equivalent concentrations of C5 protein. Substrate was prepared separately by combining pre-tRNA with trace amounts of <sup>32</sup>P labeled pre-tRNA, heating to 95 °C for 3 min, incubating at 37 °C for 10 min before adding MgCl<sub>2</sub>. Equal volumes of the enzyme and substrate were mixed to achieve final concentrations of 5 or 10 nM RNase P and 1  $\mu$ M pre-tRNA for randomized pools. Aliquots were taken at selected time points and quenched with an equal volume of formamide loading dye containing 100 mM EDTA. Products were resolved on 15% denaturing polyacrylamide gels and quantified using phosphorimager analysis. Individual substrate assays used 3  $\mu$ M pre-tRNA<sup>Met</sup> and 5 nM RNase P, and the kinetic data were fit to the integrated Michaelis–Menten equation. Single-turnover reactions were performed in 50 mM MES pH 6.0, 100 mM NaCl, 0.005% Triton X-100, and 17.5 mM MgCl<sub>2</sub>. Reactions were performed as described above except the final concentration of holoenzyme was 100 nM and substrate concentrations were <10 nM.

**High-Throughput Sequencing Kinetics (HTS-Kin).** Multiple-turnover HTS-Kin measurements were made as described in Guenther et al. Briefly, reactions were performed as described above, except scaled up 10-fold to provide sufficient material for subsequent analysis. Aliquots were taken during the time course, RNAs were resolved by polyacrylamide gel electrophoresis, and the residual substrate population was eluted and purified. Equal amounts of RNA from individual time points were used as templates for first-strand cDNA synthesis. Products were diluted 1:300, and 1  $\mu$ L of this dilution was used for PCR amplification followed by multiplexed Illumina sequencing in 75 bp single end reads on Hi-Seq 2500 instrument. Single-turnover HTS-Kin reactions were set up as described above for single-turnover reactions and were not scaled up. Primer sequences are included in [Supporting Information](#).

Relative rate-constant ( $k_{rel}$ ) values were calculated using

$$k_{rel} = \frac{\ln \frac{(1-f)}{\frac{R_{i,0}}{R_0} \left( \sum_i \frac{R}{R_0} X \right)}}{\ln \frac{(1-f)}{\sum_i \frac{R}{R_0} X}}$$

where  $R$  is ratio of each substrate variant to the reference,  $R_0$  this same ratio at the start of the reaction, and  $X$  is the mole fraction for each substrate variant. The ratio of substrates is quantified by the number of raw sequence reads from high-throughput sequencing and the total fraction of reaction was determined by quantification of polyacrylamide gels using ImageQuant software as described above.

**Quantitative Modeling of Sequence Specificity.** The HTS-Kin data were fit to a simple position weight matrix model treating each nucleotide in the randomized substrate region as independent and noninteracting using

$$\ln(k_{rel}) = \sum_{i=1}^6 (a_i A_i + c_i C_i + g_i G_i + u_i U_i)$$

where  $a_i$ ,  $c_i$ ,  $g_i$ , and  $u_i$  are integer values (0 or 1) signifying nucleotide identity and  $A_i$ ,  $C_i$ ,  $G_i$ , and  $U_i$  represent the linear coefficients for that nucleotide at position  $i$ . The PWM+IC model considered not only nucleotide identity and position in the randomized region but also the position and identity of other nucleotides in the binding site using the following equation:

$$\ln(k_{rel}) = \sum_{i=1}^6 (a_i A_i + c_i C_i + g_i G_i + u_i U_i) + \beta_j I_j$$

where the second summed terms are pairwise interaction terms. For each of these couplings,  $\beta_j$  is the linear coefficient for interaction  $j$ , and  $I_j$  is an indicator variable.  $I_j$  is 1 for all substrates with that specific pair of nucleotides, and 0 otherwise. Each interaction term which had an absolute  $t$ -value greater than 3.5 ( $p < 0.005$ ) was recorded as significant. A final model was built using stepwise regression, starting with all of the significant pairwise interactions identified in the first step.

## ASSOCIATED CONTENT

### Supporting Information

This material is available free of charge via the Internet. The Supporting Information is available free of charge on the [ACS Publications website](#) at DOI: 10.1021/acschembio.6b00275.

Additional data and information as noted in the text ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [meh2@cwru.edu](mailto:meh2@cwru.edu).

### Funding

This work was supported by US National Institutes of Health (NIH) grant GM056740 and GM096000 to M.E.H.; T32 GM008056 to C.N.N. GM118088 to E.J.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank C. Paz, M. Pedraza, and K. Simmons for assistance with kinetic experiments and N. Molyneaux for assistance processing high-throughput sequencing data.

## REFERENCES

- (1) Hsu, T. Y.; Simon, L. M.; Neill, N. J.; Marcotte, R.; Sayad, A.; Bland, C. S.; Echeverria, G. V.; Sun, T.; Kurley, S. J.; Tyagi, S.; Karlin, K. L.; Dominguez-Vidana, R.; Hartman, J. D.; Renwick, A.; Scorsone, K.; Bernardi, R. J.; Skinner, S. O.; Jain, A.; Orellana, M.; Lagisetty, C.; Golding, I.; Jung, S. Y.; Neilson, J. R.; Zhang, X. H.; Cooper, T. A.; Webb, T. R.; Neel, B. G.; Shaw, C. A.; and Westbrook, T. F. (2015) The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* 525, 384–388.
- (2) Shi, Y.; Joyner, A. S.; Shadrack, W.; Palacios, G.; Lagisetty, C.; Potter, P. M.; Sambucetti, L. C.; Stamm, S.; and Webb, T. R. (2015) Pharmacodynamic assays to facilitate preclinical and clinical development of pre-mRNA splicing modulatory drug candidates. *Pharmacol. Res. Perspect.* 3, e00158.
- (3) Guan, L., and Disney, M. D. (2012) Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.* 7, 73–86.

- (4) Dominguez, A. A., Lim, W. A., and Qi, L. S. (2016) Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* 17, 5–15.
- (5) Chappell, J., Watters, K. E., Takahashi, M. K., and Lucks, J. B. (2015) A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol.* 28, 47–56.
- (6) Peters, G., Coussement, P., Maertens, J., Lammertyn, J., and De Mey, M. (2015) Putting RNA to work: Translating RNA fundamentals into biotechnological engineering practice. *Biotechnol. Adv.* 33, 1829–1844.
- (7) Esakova, O., and Krasilnikov, A. S. (2010) Of proteins and RNA: the RNase P/MRP family. *RNA* 16, 1725–1747.
- (8) Yandek, L. E., Lin, H. C., and Harris, M. E. (2013) Alternative substrate kinetics of *Escherichia coli* ribonuclease P: determination of relative rate constants by internal competition. *J. Biol. Chem.* 288, 8342–8354.
- (9) Koutmou, K. S., Zahler, N. H., Kurz, J. C., Campbell, F. E., Harris, M. E., and Fierke, C. A. (2010) Protein-precursor tRNA contact leads to sequence-specific recognition of 5' leaders by bacterial ribonuclease P. *J. Mol. Biol.* 396, 195–208.
- (10) Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857.
- (11) Guerrier-Takada, C., McClain, W. H., and Altman, S. (1984) Cleavage of tRNA precursors by the RNA subunit of *E. coli* ribonuclease P (M1 RNA) is influenced by 3'-proximal CCA in the substrates. *Cell* 38, 219–224.
- (12) Frank, D. N., and Pace, N. R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* 67, 153–180.
- (13) Kurz, J. C., and Fierke, C. A. (2000) Ribonuclease P: a ribonucleoprotein enzyme. *Curr. Opin. Chem. Biol.* 4, 553–558.
- (14) Christian, E. L., Zahler, N. H., Kaye, N. M., and Harris, M. E. (2002) Analysis of substrate recognition by the ribonucleoprotein endonuclease RNase P. *Methods* 28, 307–322.
- (15) Zahler, N. H., Christian, E. L., and Harris, M. E. (2003) Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P. *RNA* 9, 734–745.
- (16) Sun, L., Campbell, F. E., Yandek, L. E., and Harris, M. E. (2010) Binding of C5 protein to P RNA enhances the rate constant for catalysis for P RNA processing of pre-tRNAs lacking a consensus (+1)/C(+72) pair. *J. Mol. Biol.* 395, 1019–1037.
- (17) Harris, M. E., Nolan, J. M., Malhotra, A., Brown, J. W., Harvey, S. C., and Pace, N. R. (1994) Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J.* 13, 3953–3963.
- (18) Harris, M. E., Kazantsev, A. V., Chen, J. L., and Pace, N. R. (1997) Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity crosslinking. *RNA* 3, 561–576.
- (19) Christian, E. L., McPheeters, D. S., and Harris, M. E. (1998) Identification of individual nucleotides in the bacterial ribonuclease P ribozyme adjacent to the pre-tRNA cleavage site by short-range photo-cross-linking. *Biochemistry* 37, 17618–17628.
- (20) Christian, E. L., and Harris, M. E. (1999) The track of the pre-tRNA 5' leader in the ribonuclease P ribozyme-substrate complex. *Biochemistry* 38, 12629–12638.
- (21) Sun, L., Campbell, F. E., Zahler, N. H., and Harris, M. E. (2006) Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. *EMBO J.* 25, 3998–4007.
- (22) Brannvall, M., Fredrik Pettersson, B. M., and Kirsebom, L. A. (2002) The residue immediately upstream of the RNase P cleavage site is a positive determinant. *Biochimie* 84, 693–703.
- (23) LaGrandeur, T. E., Huttenhofer, A., Noller, H. F., and Pace, N. R. (1994) Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* 13, 3945–3952.
- (24) Crary, S. M., Niranjanakumari, S., and Fierke, C. A. (1998) The protein component of *Bacillus subtilis* ribonuclease P increases catalytic efficiency by enhancing interactions with the 5' leader sequence of pre-tRNA<sup>Asp</sup>. *Biochemistry* 37, 9409–9416.
- (25) Kurz, J. C., Niranjanakumari, S., and Fierke, C. A. (1998) Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA<sup>Asp</sup>. *Biochemistry* 37, 2393–2400.
- (26) Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W., and Fierke, C. A. (1998) Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15212–15217.
- (27) Rueda, D., Hsieh, J., Day-Storms, J. J., Fierke, C. A., and Walter, N. G. (2005) The 5' leader of precursor tRNA<sup>Asp</sup> bound to the *Bacillus subtilis* RNase P holoenzyme has an extended conformation. *Biochemistry* 44, 16130–16139.
- (28) Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* 468, 784–789.
- (29) Agrawal, A., Mohanty, B. K., and Kushner, S. R. (2014) Processing of the seven valine tRNAs in *Escherichia coli* involves novel features of RNase P. *Nucleic Acids Res.* 42, 11166–11179.
- (30) Zahler, N. H., Sun, L., Christian, E. L., and Harris, M. E. (2005) The pre-tRNA nucleotide base and 2'-hydroxyl at N(-1) contribute to fidelity in tRNA processing by RNase P. *J. Mol. Biol.* 345, 969–985 Epub 2004 Dec 2008.
- (31) Guenther, U. P., Yandek, L. E., Niland, C. N., Campbell, F. E., Anderson, D., Anderson, V. E., Harris, M. E., and Jankowsky, E. (2013) Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature* 502, 385–388.
- (32) Loria, A., and Pan, T. (1998) Recognition of the 5' leader and the acceptor stem of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* RNase P. *Biochemistry* 37, 10126–10133.
- (33) Cornish-Bowden, A. (1984) Enzyme specificity: Its meaning in the general case. *J. Theor. Biol.* 108, 451–457.
- (34) Fersht, A. (1985) *Enzyme structure and mechanism*, 2nd ed., Freeman and Co., New York.
- (35) Herschlag, D. (1988) The role of induced fit and conformational changes of enzymes in specificity and catalysis. *Bioorg. Chem.* 16, 62–96.
- (36) Cleland, W. W., and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev.* 106, 3252–3278.
- (37) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- (38) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.
- (39) Stormo, G. D. (2013) Modeling the specificity of protein-DNA interactions. *Quant. Biol.* 1, 115–130.
- (40) Brannvall, M., and Kirsebom, L. A. (2005) Complexity in orchestration of chemical groups near different cleavage sites in RNase P RNA mediated cleavage. *J. Mol. Biol.* 351, 251–257.
- (41) Loria, A., Niranjanakumari, S., Fierke, C. A., and Pan, T. (1998) Recognition of a pre-tRNA substrate by the *Bacillus subtilis* RNase P holoenzyme. *Biochemistry* 37, 15466–15473.
- (42) Svard, S. G., and Kirsebom, L. A. (1993) Determinants of *Escherichia coli* RNase P cleavage site selection: a detailed in vitro and in vivo analysis. *Nucleic Acids Res.* 21, 427–434.
- (43) Oh, B. K., and Pace, N. R. (1994) Interaction of the 3'-end of tRNA with ribonuclease P RNA. *Nucleic Acids Res.* 22, 4087–4094.
- (44) Guo, X., Campbell, F. E., Sun, L., Christian, E. L., Anderson, V. E., and Harris, M. E. (2006) RNA-dependent folding and stabilization of C5 protein during assembly of the *E. coli* RNase P holoenzyme. *J. Mol. Biol.* 360, 190–203.