10-20-2016

Analysis of the RNA Binding Specificity Landscape of C5 Protein Reveals Structure and Sequence Preferences that Direct RNase P Specificity.

Hsuan-Chun Lin

Jing Zhao

Courtney N Niland

Brandon Tran

Eckhard Jankowsky

See next page for additional authors

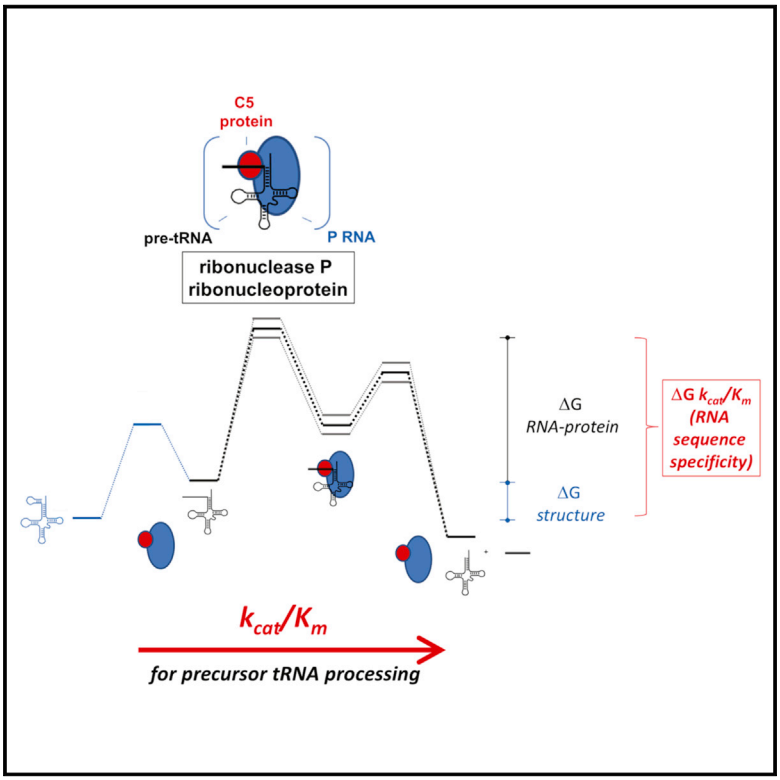
Authors
Hsuan-Chun Lin, Jing Zhao, Courtney N Niland, Brandon Tran, Eckhard Jankowsky, and Michael E Harris

Article

Cell Chemical Biology

Analysis of the RNA Binding Specificity Landscape of C5 Protein Reveals Structure and Sequence Preferences that Direct RNase P Specificity

Graphical Abstract



Authors

Hsuan-Chun Lin, Jing Zhao, Courtney N. Niland, Brandon Tran, Eckhard Jankowsky, Michael E. Harris

Correspondence

meh2@cwru.edu

In Brief

Lin et al. use high-throughput methods to analyze the binding kinetics and equilibrium affinities for all possible sequence combinations in the binding site of bacterial RNase P protein. The results reveal how both RNA structure and sequence direct the specificity of a ribonucleoprotein enzyme.

Highlights

- Kinetic aspects of RNA-protein interactions play a critical biological role
- Reaction kinetics and affinity measured for all possible C5 protein binding sites
- C5 structure and sequence preferences direct ribonucleoprotein enzyme specificity
- A comprehensive approach for analysis of RNA binding specificity is illustrated

Lin et al., 2016, Cell Chemical Biology *23*, 1271–1281
October 20, 2016
http://dx.doi.org/10.1016/j.chembiol.2016.09.002

Please cite this article as: Lin et al., Analysis of the RNA Binding Specificity Landscape of C5 Protein Reveals Structure and Sequence Preferences that Direct RNase P Specificity, Cell Chemical Biology (2016), http://dx.doi.org/10.1016/j.chembiol.2016.09.002

Cell Chemical Biology
Article

Analysis of the RNA Binding Specificity Landscape of C5 Protein Reveals Structure and Sequence Preferences that Direct RNase P Specificity

Hsuan-Chun Lin,[1] Jing Zhao,[1] Courtney N. Niland,[1] Brandon Tran,[1] Eckhard Jankowsky,[2] and Michael E. Harris[1,3,*]
[1]Department of Biochemistry
[2]Department of Biochemistry, Center for RNA Molecular Biology
Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA
[3]Lead Contact
*Correspondence: meh2@cwru.edu
http://dx.doi.org/10.1016/j.chembiol.2016.09.002

SUMMARY

RNA binding proteins (RBPs) are typically involved in non-equilibrium cellular processes, and specificity can arise from differences in ground state, transition state, or product states of the binding reactions for alternative RNAs. Here, we use high-throughput methods to measure and analyze the RNA association kinetics and equilibrium binding affinity for all possible sequence combinations in the precursor tRNA binding site of C5, the essential protein subunit of *Escherichia coli* RNase P. The results show that the RNA sequence specificity of C5 arises due to favorable RNA-protein interactions that stabilize the transition state for association and bound enzyme-substrate complex. Specificity is further impacted by unfavorable RNA structure involving the C5 binding site in the ground state. The results illustrate a comprehensive quantitative approach for analysis of RNA binding specificity, and show how both RNA structure and sequence preferences of an essential protein subunit direct the specificity of a ribonucleoprotein enzyme.

INTRODUCTION

The regulation of gene expression at the RNA level relies on the binding of numerous RNA binding proteins (RBPs) to a variety of functional RNA classes (Iadevaia and Gerber, 2015; Mitchell and Parker, 2014; Shi and Barna, 2015; Singh et al., 2015; Van Asche et al., 2015). To delineate and understand the functions of RBPs it is critical to understand their specificity, that is, how they discriminate between alternative RNA binding sites. Significant progress has been made in defining protein binding sites within the transcriptomes of cells (Ascano et al., 2012; Licatalosi et al., 2008; Zhao et al., 2010). New approaches also allow simultaneous measurements of protein binding to thousands or more different substrate variants in vitro and in vivo (Campbell and Wickens, 2015; Cook et al., 2015; Jankowsky and Harris, 2015).

The specificities of RBPs are often viewed as differences in equilibrium binding affinities between RNA ligands. This perspective allows the development of quantitative models of sequence discrimination that often correlate with binding preferences of an RBP in the cell. Application of information theory to quantify the specificity of DNA binding proteins provides a precedence and context for advancing the understanding of RNA specificity (Schneider et al., 1986; Stormo, 2013). However, equilibrium conditions do not always apply in vivo, and it is generally appreciated that the kinetic aspects of RNA-protein interactions play a critical biological role (Jankowsky and Harris, 2015; Mackereth and Sattler, 2012; Mitchell and Parker, 2014; Ray, 1983). Currently, the linkages between the kinetic mechanisms of RBP binding and specificity are poorly understood. Only a single recent study reported the mechanistic analysis of the effects of large numbers of sequence variants in an RNA stem loop recognized by phage MS2 coat protein (Buenrostro et al., 2014). The results showed that the determinants of RBP MS2 RNA discrimination are structure and position specific, and selectivity is achieved in large part through differences in association rate constants.

For no RBP has the link between specificity and binding mechanism been examined for all possible sequence variants. Here, we conducted such a comprehensive study for RBP C5, the protein subunit of *E. coli* RNase P (RNase P), an essential tRNA 5′ end processing endonuclease (Figure 1). C5 binds the 5′ leader sequences of precursor tRNAs (ptRNAs) at a defined region, N-3 to N-8, relative to the RNase P cleavage site at N1. To systematically link specificity to binding kinetics it is necessary to consider the free energy landscape for the RNA-protein association reaction (Figure 1A). Specificity can arise due to differences in the ground state of RNA variants, the ground state of their RNA-protein complexes, and the transition states for RNA-protein complex formation, or through a combination thereof. Previously, we used high-throughput sequencing to measure the effect on k_{cat}/K_m for all possible sequence variations in the binding site of C5. The results demonstrated that, although the genomically encoded C5 binding sites do not show sequence or structure signatures, C5 contributes inherent 5′ leader sequence specificity to the magnitude of k_{cat}/K_m for RNase P (Guenther et al., 2013; Koutmou et al., 2010). A general free energy landscape experimentally determined for processing of a model ptRNA^{Met} with a genomically encoded leader sequence by *Escherichia*

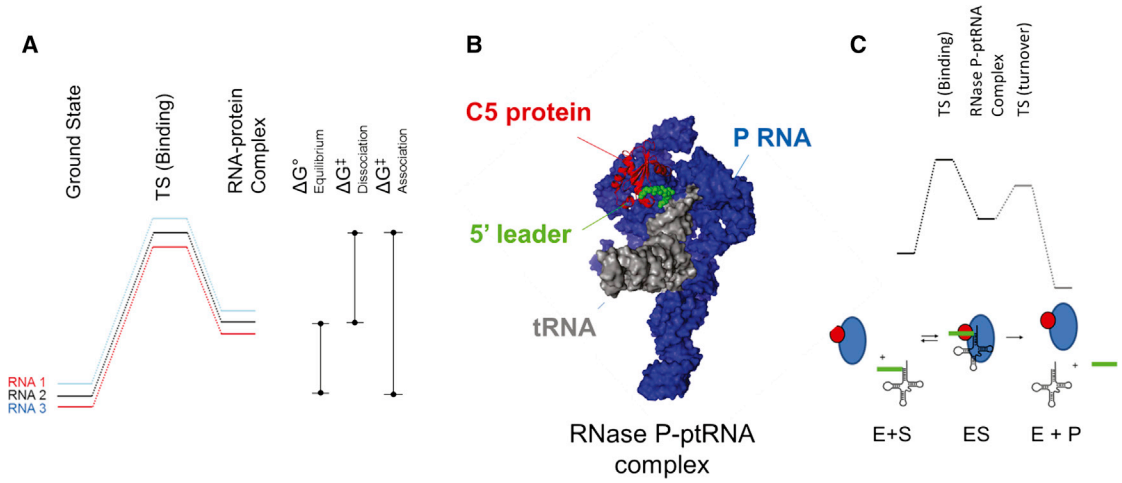


Figure 1. RNA Binding Protein C5 Contributes to *E. coli* RNase P Specificity by Binding ptRNA 5′ Leader Sequences

(A) Free energy landscape for a first-order reversible RBP association reaction. RNA 1–3 illustrate the potential for sequence variation to affect the free energies of the ground state, transition state (TS), and the RNA-protein complex.

(B) Structure model of RNase P composed of P RNA (blue) and C5 protein (red) complexed with a substrate ptRNA (gray) with the 5′ leader (green) bound by C5.

(C) The reaction mechanism for *E. coli* RNase P processing of a model ptRNA^{Met82} (Yandek et al., 2013) involves rapid cleavage relative to dissociation such that the rate constant at limiting substrate concentration (k_{cat}/K_m) reflects the association step.

coli RNase P involves fast cleavage relative to substrate dissociation, and thus k_{cat}/K_m measures the association step (Figure 1C). However, the global effects of sequence variation on this landscape, and therefore the underlying link between C5 specificity and RNase P specificity for ptRNA, are not understood.

We comprehensively determined how the RNA binding protein C5 contributes to *E. coli* RNase P substrate specificity using high-throughput kinetic and equilibrium binding methods. The data show that C5 sequence specificity arises due to favorable RNA-protein interactions that stabilize both the transition state for association and the enzyme-substrate (ES) complex. Unfavorable RNA structures involving the C5 binding site in the free RNA make independent contributions to specificity. Thus, C5 protein specificity contributes to *E. coli* RNase P substrate discrimination for 5′ leader sequences and structure due to effects on both the ground state and transition state for binding.

RESULTS

Variation of ptRNA Sequence in the C5 Binding Site Affects RNase P Association and Equilibrium Binding but Not the Cleavage Step

To determine how sequence variation in the C5 protein binding site affects the free energy landscape for the RNase P reaction, we first analyzed a population of substrates containing all sequence variants at positions in the ptRNA 5′ leader that interact with C5 using single-turnover kinetics, multiple-turnover kinetics, and equilibrium binding. Randomization was carried out in the background of a well-characterized *E. coli* ptRNA^{Met82} substrate with a genomically encoded sequence that binds to *E. coli* RNase P with nanomolar affinity, and reacts with a k_{cat} that reflects the substrate cleavage step and a k_{cat}/K_m that measures association (Figure 1C) (Sun et al., 2006; Yandek et al., 2013). The binding and kinetics of the ptRNA^{Met82}(N-3-8) substrate population, which is randomized at positions N(-3) to N(-8), were compared with those of the native ptRNA^{Met82} containing the genomically encoded C5 binding site (AAAAAG) (Figure 2A). Additional 5′ leader sequences (21A) were added to ptRNA^{Met82} for subsequent high-throughput studies as described below.

Randomization of the C5 binding site has little effect on single-turnover reaction kinetics performed at saturating RNase P concentration (Figure 2B). Under these conditions the catalytic step is rate limiting for the native ptRNA^{Met82} (Sun et al., 2010; Yandek et al., 2013). Therefore, this result suggests that C5 binding does not contribute significantly to the cleavage step consistent with previous studies (Hsieh et al., 2004; Hsieh and Fierke, 2009; Niranjanakumari et al., 1998), although an effect on RNase P catalysis for a small number of variants cannot be excluded. Because tRNA binds with much higher affinity and dissociates much more slowly than the 5′ leader sequence cleavage product (Kurz et al., 1998; Sun et al., 2006), the results further suggest that turnover of ES (k_{cat}) is unlikely to be affected by leader randomization. Previously, we showed that the multiple-turnover kinetics of the randomized ptRNA^{Met82}(N-3-8) population differ markedly from the time course of ptRNA^{Met82} in multiple-turnover reactions at limiting substrate concentrations (Figure 2C) (Guenther et al., 2013). In competitive reactions containing alternative substrates different ptRNAs react according to their k_{cat}/K_m (Anderson, 2015), which reflects the rate constant for ptRNA^{Met82} association (Guenther et al., 2013; Yandek et al., 2013). Since the same randomization does not affect the cleavage step this result indicates that C5 contributes primarily to association. To test this and gain additional insight into the effects of leader sequence variation on RNase P mechanism we measured the effect of randomizing the C5 binding site on equilibrium binding affinity. Gel mobility shift assays were used to quantify free and bound ptRNA. Figure 2D shows that randomization of the C5 binding site alters the observed binding affinity of the

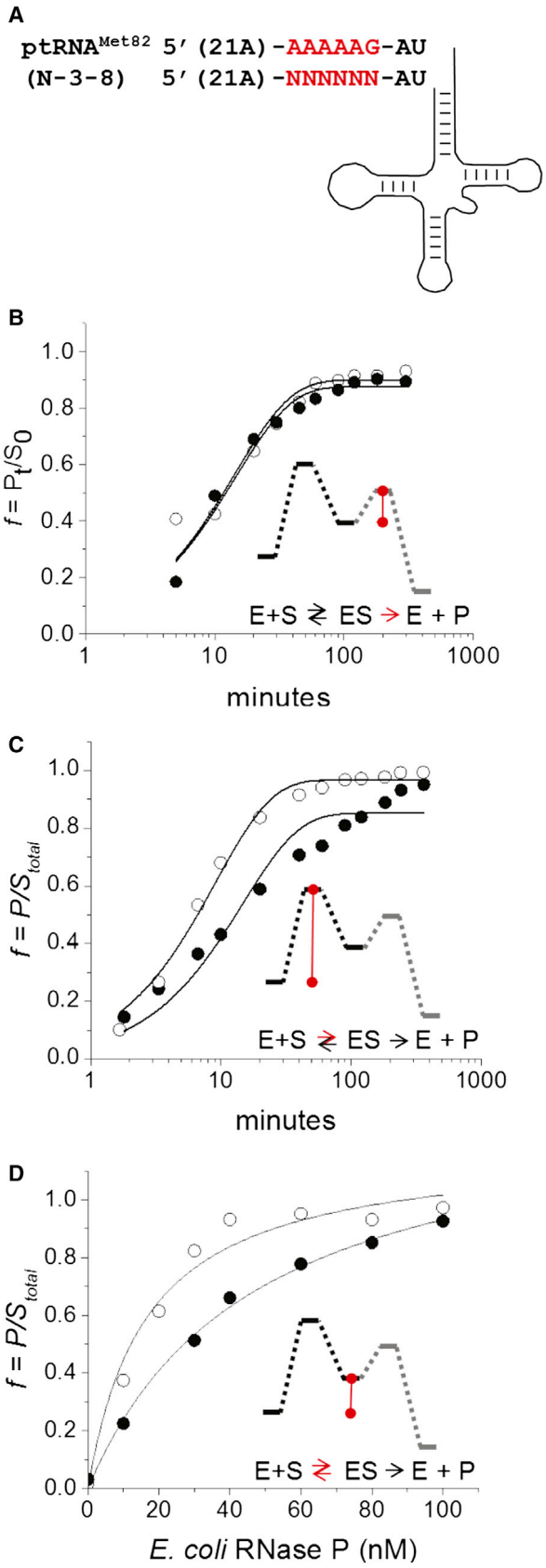


Figure 2. Single-Turnover Kinetics, Steady-State Kinetics, and Equilibrium Binding Affinities of ptRNA^{Met82}21A and the ptRNA^{Met82} 21A(N-3-8) Randomized Population

(A) Proximal leader sequence of ptRNA^{Met82}21A showing the C5 binding site (red) and the nucleotide positions randomized ptRNA^{Met82}21A(N-3-8). The sequence of the 21A 5′ leader is shown in Figure 3A.

(B) Single-turnover kinetic analyses of ptRNA^{Met82}21A (open circles) and ptRNA^{Met82}21A(N-3-8) (filled circles). Assays were performed at saturating (>1 μM) RNase P concentration with limiting (10 nM) concentrations of substrate ptRNA in order to observe the effect of -3-8 randomization on the catalytic step. The data are fit to a single exponential function.

(C) Multiple-turnover kinetic analyses of ptRNA^{Met82}21A (open circles) and ptRNA^{Met82}21A(N-3-8) (filled circles). Assays were performed using excess substrate (1 μM) with limiting RNase P concentration (5 nM) in order to estimate the effect of -3-8 randomization on k_{cat}/K_m. The data are fit to a single exponential function to illustrate the difference in kinetics of the genomically encoded sequence and the randomized population.

(D) Equilibrium binding analyses of ptRNA^{Met82}21A (open circles) and ptRNA^{Met82}21A(N-3-8) (filled circles). Assays were performed with limiting ptRNA (1 nM) and a range of RNase P concentrations, and the formation of RNase P-ptRNA complex was quantified using EMSA in order to measure the effect of N-3-8 randomization on equilibrium binding affinity. The data are fit to a single-site equilibrium binding model.

ptRNA^{Met82}(N-3-8) population relative to ptRNA^{Met82}. Thus, these data indicate that C5 binding specificity does not contribute significantly to RNase P catalysis, but alters both association kinetics and equilibrium binding affinity.

Comprehensive, High-Throughput Determination of the Effects of Sequence Variation in the C5 Binding Site on Association Kinetics, k_{rel}, and Equilibrium Binding Affinity, $K_{A,rel}$

The degree of correlation between association kinetics and equilibrium binding affinity necessarily depends on how sequence variation alters the free energy landscape for the binding reaction (Figure 1A). Linear free energy relationships between rate and equilibrium constants for the same reaction can provide information on how changes in molecular structure affect the reaction mechanism. Linear free energy relationships are used extensively in mechanistic studies of physical organic chemistry (Kirby and Nome, 2015; Lassila et al., 2011) and protein folding (Fersht et al., 1992; Matouschek and Fersht, 1993; Sosnick, 2008). There are limits to interpreting free energy relationships arising from differences in ground state, multiple reaction channels, and changes in mechanism (Farcasiu, 1975; Jencks, 1985). Nonetheless, correlating the effects of C5 binding site sequence variation on ptRNA kinetics and equilibrium binding affinity can reveal potential differences in the free energy of the transition state for association versus the ES complex. Accordingly, we measured the relative rate constants and equilibrium association constants for all possible sequence variants in the C5 binding site. As described in the next section, we used the resulting rate and equilibrium constant distributions to globally analyze the effect of C5 binding site sequence variation on RNase P mechanism.

The relative k_{cat}/K_m (k_{rel}) for RNase P processing of ptRNA^{Met82} randomized at positions N-3-8 were determined previously using high-throughput sequencing kinetics (HiTS-KIN) (Guenther et al., 2013). In brief, the time-dependent changes in distribution of RNA species in the substrate population due to RNase P processing were analyzed by Illumina sequencing, and relative rate

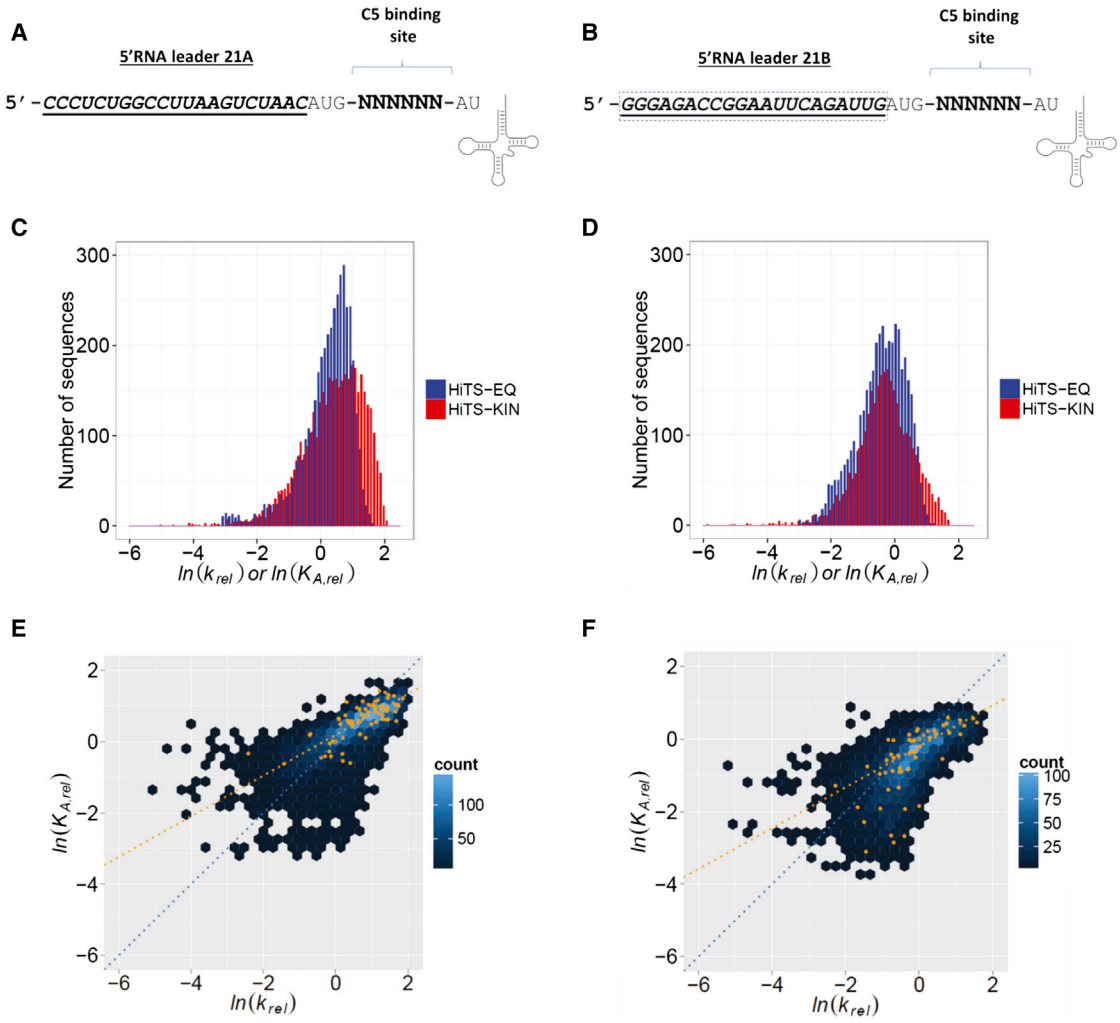


Figure 3. Correlation of Observed Association Rate Constants, k_{rel}, and Equilibrium Association Constants, $K_{A,rel}$, for ptRNA^{Met82}21A(-3-8) and ptRNA^{Met82}21B(-3-8)

(A) Sequence of ptRNA^{Met82}21A. The inverted leader sequences relative to 21B are underlined and the randomized positions N-3-8 are bold.

(B) Sequence of ptRNA^{Met82}21B. The inverted leader sequences relative to 21A are underlined and the randomized positions N-3-8 are bold.

(C) Distributions of lnk_{rel} and ln$K_{A,rel}$ values determined for ptRNA^{Met82}21A(N-3-8). The position of the genomically encoded reference sequence (AAAAAAG) at $k_{rel} = K_{A,rel} = 0$ is indicated by a solid vertical line.

(D) Distributions of lnk_{rel} and ln$K_{A,rel}$ values determined for ptRNA^{Met82}21B(N-3-8). The position of the genomically encoded reference sequence (AAAAAAG) at $k_{rel} = K_{A,rel} = 0$ is indicated by a solid vertical line.

(E) Density plot of lnk_{rel} versus ln$K_{A,rel}$. The scale is shown on the right. Fitting the data to a linear equation yields a slope of 0.52 shown as an orange dotted line. A theoretical reference line with a slope of 1 is shown as a dotted blue line.

(F) Density plot of lnk_{rel} versus ln$K_{A,rel}$. The scale is shown on the right. A fit of the data with ln$k_{rel} > -1$ has a slope of approximately 0.5 is shown as an orange dotted line. A reference line with a slope of 1 is shown as a dotted blue line. The orange dots in panels (E) and (F) indicate genomically encoded leader sequences.

constants calculated using internal competition kinetics (Anderson, 2015). The HiTS-KIN procedure provides the relative k_{cat}/K_m calibrated to the genomically encoded leader sequence (k_{rel} = (k_{cat}/K_m(NNNNNN))/(k_{cat}/K_m(AAAAAG))). For sequences that react faster the reference k_{rel} is greater than 1, while sequences with slower kinetics react with k_{rel} values less than 1. The resulting rate constant distribution reflects the full range of effects of C5 binding site sequence variation on RNase P processing kinetics. As described in more detail below, in this study we also measured the k_{rel} distribution of a second ptRNA^{Met82}(N-3-8) population termed 21B in which the constant distal leader sequence (5′ to the C5 binding site) was changed to its Watson-Crick (WC) complement (Figure 3B) in order to investigate effects of sequence context on specificity.

To measure the equilibrium association constant (K_A) for the same randomized populations of ptRNA^{Met82} substrates, we developed an approach similar to HiTS-KIN that we term high-throughput sequencing equilibrium (HiTS-EQ) binding. The overall HiTS-EQ workflow is shown in Figure S1. HiTS-EQ involves separation of the free and bound fractions of the randomized populations of ptRNA over a range of RNase P concentrations using an electrophoretic mobility shift assay (EMSA). The

populations thus fractionated are gel purified and the distribution of individual RNA variants determined by Illumina sequencing. Quantitative analysis of the changes in the mole fractions of each RNA variant was carried out using a simple competitive equilibrium binding model. This operation allows the calculation of relative equilibrium association constants (K_A) values calibrated to the genomically encoded leader sequence ($K_{A,rel}$ = K_A(NNNNNN)/K_A(AAAAAG)). Assays using single ptRNA substrates were used to test the accuracy of the relative $K_{A,rel}$ values determined by HiTS-EQ, and comparison of the $K_{A,rel}$ values measured from independent experiments demonstrated the reproducibility of the method (Figure S2).

Sequence Variation in the C5 Protein Binding Site Has Equivalent Effects on k_{rel} and $K_{A,rel}$

The ability to measure distributions of k_{rel} and $K_{A,rel}$ allowed us to systematically analyze how variation differentially affects the rate constant versus the equilibrium constant for RNase P binding. The application of a simple linear free energy approach to reveal mechanistic detail is powerful, but assumes a simple one-step reaction mechanism and a common rate-limiting step as described in more detail in Supplemental Information (correlation analysis of k_{rel} and $K_{A,rel}$ values). Nonetheless, quantitatively analyzing the degree to which changes in k_{rel} and $K_{A,rel}$ are correlated provides information on the effects of sequence variation on the binding reaction mechanism (see Figure S3).

The $K_{A,rel}$ and k_{rel} for the ptRNA^{Met82}21A(N-3-8) substrate population were compared as their natural log in order to correlate these values on a common scale. A histogram illustrating the numbers of sequences with a particular ln$K_{A,rel}$ and lnk_{rel} value is shown in Figure 3C. A similar range of rate and equilibrium constant relative to the genomically encoded leader sequence at ln$K_{A,rel}$ = lnk_{rel} = 0 was observed for both datasets, with a somewhat narrower distribution of ln$K_{A,rel}$ values. Most variants were within a few fold of the native reference, although a significant fraction bound with higher affinity and reacted with faster rate constants.

To correlate effects of leader sequence variation on association kinetics and equilibrium binding, we plotted ln$K_{A,rel}$ versus lnk_{rel} for the individual variants in the ptRNA^{Met82}21A(N-3-8) population (Figure 3E). The majority of sequences follow a linear trend in which ln$K_{A,rel}$ and lnk_{rel} values are correlated with a positive slope of approximately 0.5 (orange dashed line). This correlation between the effects of sequence variation on ln$K_{A,rel}$ and lnk_{rel} values includes the leader sequences for the 87 ptRNAs encoded in the *E. coli* genome (orange dots). Greater displacement from linear correlation is observed at lower values of ln$K_{A,rel}$ and lnk_{rel}. This effect is attributable, in part, to greater experimental error in measuring small rate and equilibrium constants by HiTS-KIN and HiTS-EQ.

The observation of a linear correlation between association kinetics and equilibrium binding affinity indicates that variation in the sequence of the C5 binding site results in a similar degree of stabilization or destabilization of the transition state for association and the bound complex. That is, sequences that are optimal for association and therefore lower the activation energy are also optimal for binding affinity and similarly stabilize the bound state. A slope of less than 1 suggests that there is a greater effect of sequence variation on the transition state for association versus the bound complex, and thus a proportionality greater effect on the observed reaction kinetics.

Contribution of Distal Leader Sequences to Observed C5 Binding Specificity

To facilitate Illumina sequencing, additional 5′ leader sequences were added to ptRNA^{Met82} for amplification by RT-PCR to avoid bias introduced by a primer-ligation step (see Guenther et al., 2013 and Supplemental Information). This nonetheless introduces the potential for structural effects due to intramolecular pairing interactions. In order to identify such contributions to specificity we repeated the HiTS-KIN and HiTS-EQ measurements of k_{rel} and $K_{A,rel}$ using two different 5′ leader extensions, termed 21A and 21B, in which the first 21 nucleotides are switched to their WC complement. We reasoned that effects of sequence variation due to intrinsic C5 specificity will be common between the 21A and 21B datasets, while differences can identify attenuation or enhancement due to sequence context.

For the ptRNA^{Met82}21B(N-3-8) population (Figure 3B) the rate constant and equilibrium constant distributions are similar to those observed in the ptRNA^{Met82}21A context (compare Figures 3C and 3D). Figure 3F shows the observed ln$K_{A,rel}$ versus lnk_{rel} values plotted for each individual sequence variant in the ptRNA-^{Met82}21B population. Similar to ptRNA^{Met82}21A, most variants in the ptRNA^{Met82}21A(N-3-8) population follow a linear relationship in which the association rate constant and equilibrium binding affinity are correlated. In both cases the data for optimal variants (lnk_{rel} > −1) fit a linear trend with a slope of approximately 0.5 (orange dashed line). This result suggests that the ability of C5 protein to discriminate between leader sequences at the transition state for substrate binding is an intrinsic aspect of its mechanism, and that this property is independent of the sequences 5′ of the binding site.

However, the plot of ln$K_{A,rel}$ versus lnk_{rel} for the 21B randomized population is curved relative to the data obtained using the 21A context (compare Figures 3E and 3F). The observation of the curvature in the plot of ln$K_{A,rel}$ versus lnk_{rel} suggests that there is likely to be a second, independent effect of sequence variation on the binding mechanism relative to the ptRNA^{Met82}21A population that may have a greater effect on k_{off} (see Supplemental Information: correlation analysis of k_{rel} and $K_{A,rel}$ values) (Figure S3). We noted that the greatest curvature occurs at the lowest values of k_{rel} and $K_{A,rel}$. The unfavorable nature of the effect combined with the dependence of this effect on the identity of the 5′ distal leader sequence suggests that the observed curvature in plots of lnk_{rel} versus ln$K_{A,rel}$ for ptRNA^{Met82}21B may reflect contributions to specificity from RNA structure.

Proximal Leader Sequence Secondary Structure Interferes with C5 Binding

To identify C5 binding site sequences with the greatest sensitivity to the distal 5′ leader sequence context, we plotted the observed k_{rel} values for the ptRNA^{Met82}(N-3-8)21A versus the 21B population (Figure 4A). For the majority of sequences, the $K_{A,rel}$ (Figure 4A) and k_{rel} values (Figure S4) are consistent between the 21A and 21B contexts. However, a subset of sequences has a significantly lower binding affinity (>2-fold) and a slower rate constant in the 21A context compared with same

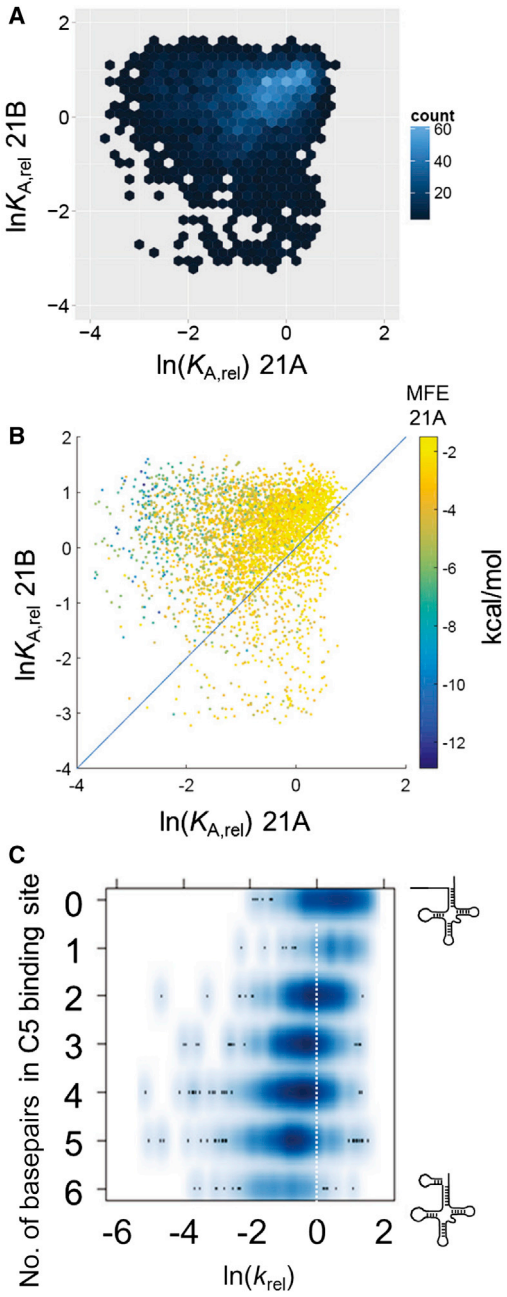


Figure 4. Stable RNA Secondary Structure Makes Unfavorable Contributions to C5 Binding Specificity

(A) Density plot of $\ln(K_{A,rel})$ measured for ptRNA^{Met82}(N-3-8)21A versus 21B.

(B) Dot plot of the same data with an overlay of mean folding free energy (MFE) calculated for each leader sequence in the 21A population. The color code corresponding to the range of MFE values is shown on the right.

(C) Density plot of $\ln k_{rel}$ observed for the ptRNA^{Met82}(N-3-8)21A population as a function of the number of WC pairing interactions involving the C5 binding site in the calculated lowest-free energy structures. The number of WC pairs from 0 to complete pairing (6) is shown on the left.

sequences in the 21B context (note upper left quadrant of Figure 4A). A smaller number of sequences follow the opposite trend, and have faster reactivity and bind with higher affinity in the 21A distribution. This observation further suggests that the sequences that are sensitive to the distal 5′ leader sequence context are subject to a second, unfavorable contribution to specificity that is likely to involve secondary structure.

To test this notion we calculated a mean folding free energy (MFE) for each variant. The results showed that a greater number of sequences in the 21B population have high folding stabilities (MFE < −4 kcal/mol) compared with the 21A population (Figure S4). This result correlates with the shift in the k_{rel} and $K_{A,rel}$ distributions to overall lower values in the 21B population as noted above (compare Figures 3B and 3E). The calculated MFE values for the 21B sequence context are overlaid on the plot of the natural log of the $K_{A,rel}$ values for the 21A versus the 21B population in Figure 4B. The correspondence between the two substrate contexts is greatest for sequences with a low potential to form a stable structure. Conversely, sequences with low predicted folding free energies (<−7 kcal/mol) in the 21B context deviate from this trend and bind with lower affinity than expected. Similarly, the subset of sequences with the most negative predicted MFE in the 21A population bind significantly weaker than those observed for the same sequences in the 21B leader context (Figure S5).

We also calculated for each sequence the number of potential pairing interactions involving the C5 binding site at N(-3-8) contained in the lowest free energy structure model. We used this information to evaluate whether the magnitudes of $K_{A,rel}$ and k_{rel} correlated with increasing potential to form WC pairing interactions in the C5 binding site. Relative to the reference substrate at $\ln k_{rel} = 0$, a clear correlation is observed between increasing numbers of pairing interactions in the C5 binding site and decreasing $\ln k_{rel}$ for ptRNA^{Met82}21A(N-3-8) (Figure 4C). The same trend toward lower affinity and slower association with increasing numbers of predicted pairing interactions in the C5 binding site is observed for both $\ln k_{rel}$ and $\ln K_{A,rel}$ in both the 21A and 21B sequence contexts (Figure S5). Importantly, ptRNA substrates that lack the 21A/B 5′ extension, but nonetheless contain a 6 base pair stem loop that includes the C5 binding site, bind with 20-fold lower affinity in single-substrate EMSA assays (Figure S6).

Thus, the data are consistent with two independent effects of sequence variation on C5 binding: a primary effect governing the behavior of most variants is independent of sequence context representing RNA-protein interactions. A second, independent, contribution to specificity involves the formation of an unfavorable secondary structure in the single-stranded RNA binding site of C5.

Quantitative Analysis of Intrinsic C5 Sequence Specificity

The ability to identify effects due to unfavorable secondary structure allowed us to isolate the population of sequences in which C5 binding specificity is least influenced by these contributions. The subsets of sequences in the 21A and 21B populations that primarily reflect intrinsic C5 sequence specificity for its single-stranded RNA binding site were selected by binning the data points spanning values of ±0.5 reflective to a line with a slope of m = 0.5 drawn through the reference sequence (at $\ln k_{rel} = \ln K_{A,rel} = 0$) in plots of $\ln k_{rel}$ versus $\ln K_{A,rel}$. A value of m = 0.5 was selected because data with $\ln k_{rel} > 0$ in both datasets,

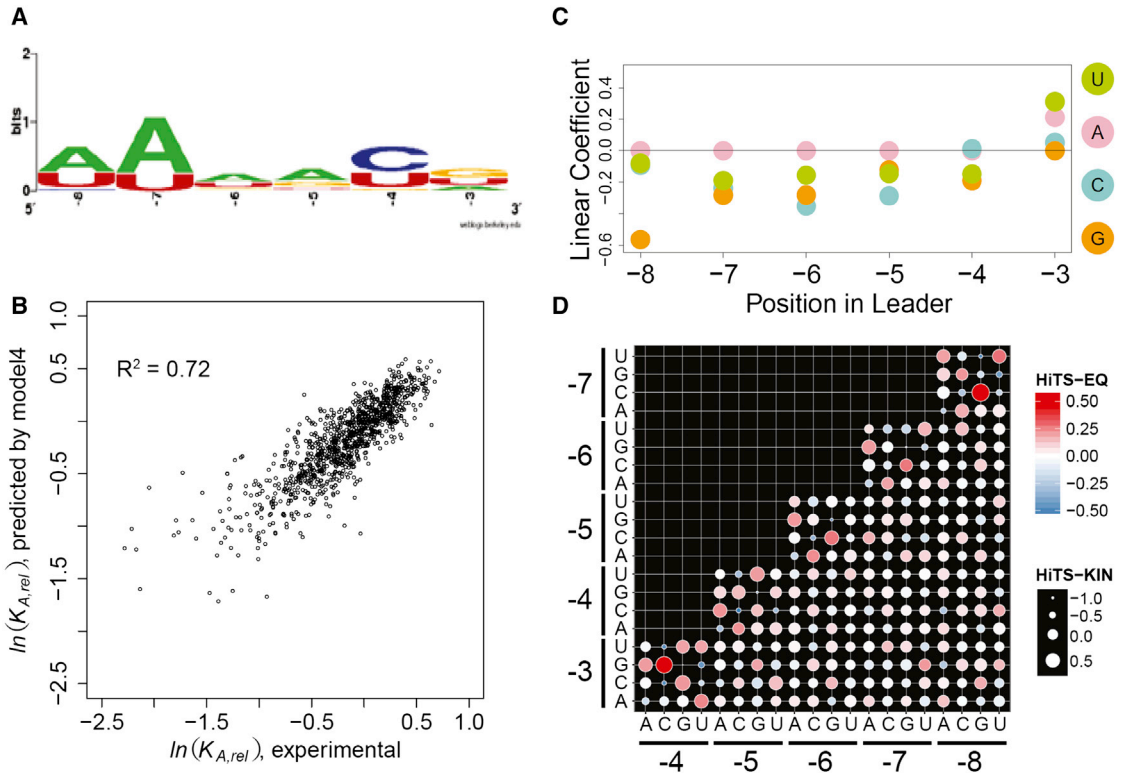


Figure 5. Quantitative Modeling of the RNA Sequence Specificity of C5 Protein

Sequences minimally affected by unfavorable secondary structure were binned as described in the text and in Supplemental Information.

(A) Probability sequence logo calculated from sequences in the top 1% of k_{rel} values.

(B) Comparison of the $K_{A,rel}$ values predicted by linear regression fitting to a specificity model, including position weight matrix (PWM) and interaction coefficient (α) values as determinants to the observed $K_{A,rel}$ values.

(C) PWM values derived from fitting the high-throughput biochemical data. Scoring is relative to the genomically encoded leader (AAAAAG), which contributes a linear coefficient of 0; positive and negative values represent contributions relative to the reference sequences.

(D) Comparison of interaction coefficient (α) values derived from independently fitting the $K_{A,rel}$ and k_{rel} datasets. The magnitude of the α values from fitting $K_{A,rel}$ are indicated by differences in color (with red indicating greater magnitude), while the magnitude of k_{rel} values are indicated by differences in size.

including the majority of endogenous genomically encoded leader sequences, correlate with a slope of 0.5 (see Figures 3C and 3F). This operation captured the majority of sequences in both populations (Figure S7), reflecting the fact that relatively few sequences in the population are subjected to large effects due to secondary structure.

A probability sequence logo of the optimal 1% of sequences reveals nucleobase discrimination throughout the binding site (Figure 5A). However, this analysis does not take into account the entire range of effects of sequence variation of reaction rate and affinity. Therefore, we fit the k_{rel} and $K_{A,rel}$ datasets independently to a quantitative model for sequence specificity that incudes parameters for nucleobase identity and position (position weight matrix [PWM] scores) as well as parameters that quantify the coupling of the contributions between different positions within the binding site (interaction coefficient [α] terms) (Guenther et al., 2013) (see Supplemental Information). As shown in Figure 5B, this PWM binding model that includes α terms explained 74% of the variance in the data. The PWM and α values for both the kinetic and equilibrium binding datasets allows the effects of sequence variation on k_{rel} and K_{rel} to be quantitatively compared. The nucleobase specificity as indicated by the PWM values as well as the α values are essentially identical for both datasets. The PWM scores for each nucleotide at individual positions in the binding site documents the intrinsic sequence preference of C5 (Figure 5C). A heatmap showing the correlation between the α values calculated for the k_{rel} and K_{rel} distributions reveals that both datasets demonstrate coupling between adjacent nucleotides in the binding site (Figure 5D). These results are consistent with the correlation between k_{rel} and $K_{A,rel}$ (see above, Figure 3) and the interpretation that sequence variation in the C5 binding site similarly affects the free energy of the transition state for association and the RNase P-ptRNA complex.

According to this model, the PWM and α parameters derived from quantitative analysis of sequence specificity represent the intrinsic RNA discrimination properties of C5 binding that are independent of the surrounding leader sequence context. To test this we measured the kinetics and equilibrium binding affinity of selected individual variants in the context of ptRNAs with short (10 nt) leader sequences lacking the 21A/B 5′ extension and containing only the C5 binding site (Figure 6A). We focused the experimental analysis on N(-3) and N(-4) because these positions show clear coupling (high α values) and PWM scores (Figure 5), and also lie within the P protein binding site as

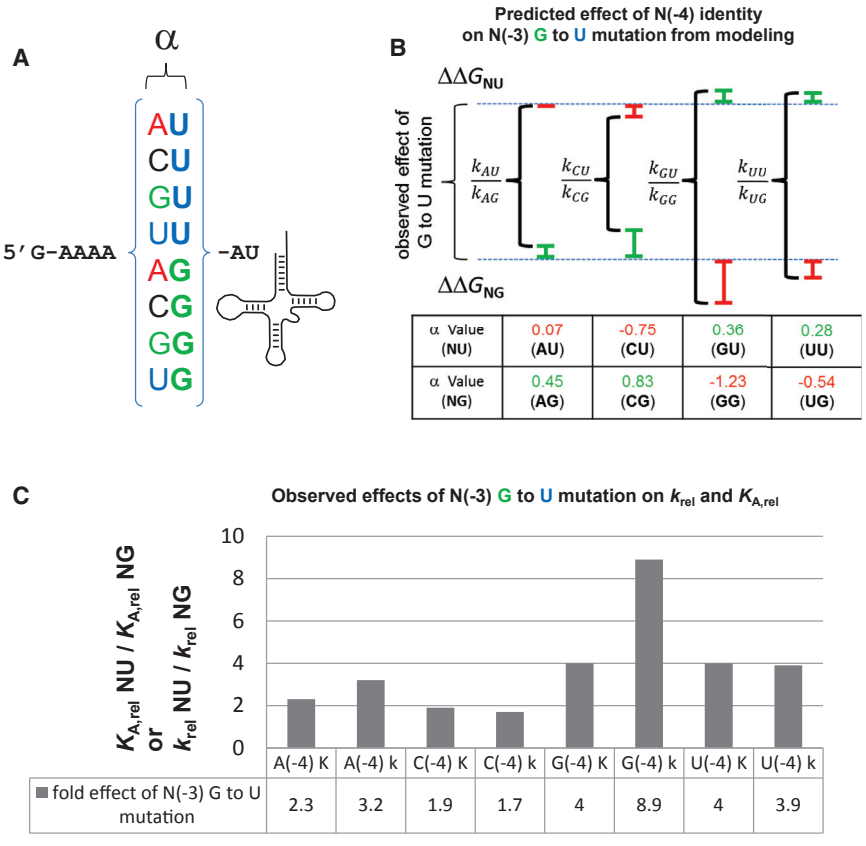


Figure 6. Validation of the Context-Dependent Effects of Sequence Variation in the C5 Binding Site using Single ptRNA Substrates

(A) The effect of a U to G substitution at position N(-3) is influenced by the base identity at position N(-4) as expressed by the α value from modeling as described in the text and shown in Figure 5D. The 16 different individual ptRNAs lacking the 21A extension with different nucleobases at N(-3) and N(-4) were synthesized and their k_{rel} and $K_{A,rel}$ values determined using individual substrate assays.

(B) The magnitude of the α values of U(-3) or G(-3) in the context of the other four nucleobases at N(-4) predicted from linear regression of the k_{rel} and $K_{A,rel}$ datasets.

(C) Experimentally observed effects of G to U substitution at N(-3) are influenced by base identity at N(-4). The bar graph shows the fold effect on the observed k_{rel} and $K_{A,rel}$ values measured for individual ptRNA substrates representing all four nucleobases at the N(-4) position.

demonstrated by photocrosslinking, mutagenesis, and X-ray crystallography (Koutmou et al., 2010; Niranjanakumari et al., 2007; Reiter et al., 2010). The quantitative binding model predicts that the contribution of the nucleobase at N(-4) is dependent on the identity of the nucleobase at position N(-3). Specifically, the PWM and α values derived from data fitting predict that A(-4) or C(-4) enhances the contribution of G(-3) (α = +0.45 to +0.83), while a G(-4) or U(-4) suppresses the contribution of G(-3) (α = −1.23 to −0.54) (Figures 5D and 6B). The identity of N(-4) has a smaller, but opposite effect on the contribution of a U(-3). To test for the predicted context dependence we used reactions containing single ptRNA lacking additional 5′ leader sequences distal to the C5 binding site. The results validate that for k_{rel}, U is optimal at N(-4) and that a G(-3) to U(-3) substitution has only a 2- to 3-fold effect when N(-4) is A or C. However, this same change in RNA sequence has a 4- to 8-fold effect on the experimentally observed rate and equilibrium constants when the adjacent N(-4) is G or U (Figure 6C). Therefore, we conclude that the PWM values accurately reflect the intrinsic sequence specificity contributed by C5 to ptRNA recognition. Importantly, the coupling effects measured by α values, although small, together make large and significant contributions to the observed sequence specificity.

DISCUSSION

Because RBPs like C5 interact with their binding sites in both a sequence- and secondary structure-specific manner both factors must necessarily contribute to discrimination between alternative RNA binding sites. Several transcriptome-wide studies have now linked mRNA secondary structure to RBP binding and regulation of translation, mRNA stability, alternative splicing, and polyadenylation (Barrass et al., 2015; Ding et al., 2014; Gosai et al., 2015; Li et al., 2012). It has also been shown that sequestration of protein binding sites by RNA structures can impact RBP binding in vitro (Zhuang et al., 2012) and in vivo (Maenner et al., 2013). RNA structure can thus impact selectivity even for proteins that bind to unstructured sites. Exploring the contributions of sequence and structure to RBP specificity at a quantitative, mechanistic level is therefore important, and the studies presented here represent a necessary first step. We are able to comprehensively define for C5 how structure exerts an influence on specificity that is independent and separable from its intrinsic RNA sequence specificity. These properties of C5 in turn allow RNase P to respond to the differences in both ground state and transition state free energies of alternative ptRNA substrates.

To date, few studies have been aimed at determining RNA binding specificity by analyzing the kinetics of large numbers of RNA substrates. Analysis of the binding of the MS2 coat protein to a large set of variants of its cognate RNA hairpin (Buenrostro et al., 2014) revealed differences in substrate preferences due to variations in ground state structure as well as the strength of RNA-protein contacts. The specificity of MS2 primarily arises due to effects of sequence variation on association rate constants, with comparably small contributions from dissociation rate constants. Although ptRNA dissociation was not directly measured in the current study, the fact that we observe similar effects of sequence variation on rate and equilibrium constants suggests minimal effects on k_{off}. Variation in tRNA structure has a minimal effect on the dissociation rate constant for

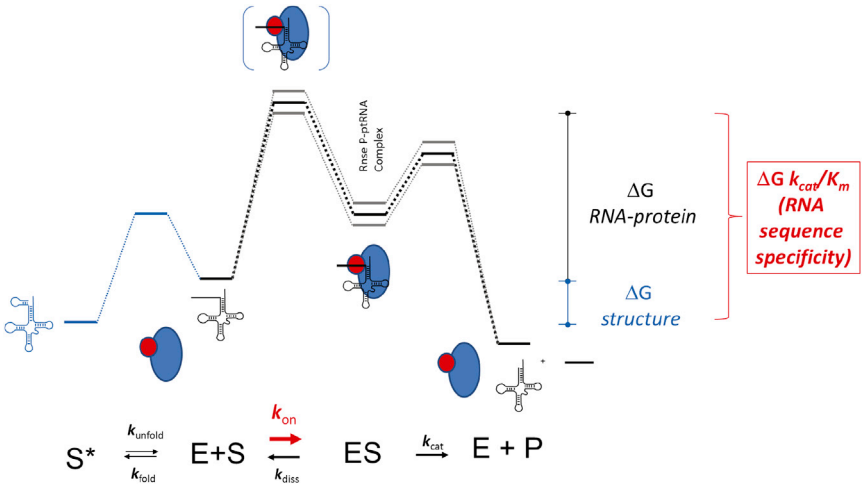


Figure 7. C5 Protein Specificity Contributions to RNase P Substrate Discrimination at Both the Ground State and Transition State for Association

Equilibrium formation of a secondary structure in the 5′ leader sequence can interfere with C5 binding, which lowers the ground state ($\Delta G_{structure}$) for the reaction resulting in a slower k_{cat}/K_m. Formation of favorable interactions between C5 and its binding site stabilize the transition state for the association ($\Delta G_{RNA\text{-}protein}$) and the bound RNase P-ptRNA complex equivalently. Both contributions affect the magnitude of k_{cat}/K_m and therefore contribute to the discrimination by RNase P between alternative ptRNAs.

EF-Tu binding, but has a large effect on association rate constant that in turn determines its binding specificity (Schrader et al., 2009). Such a mode of "k_{on} specificity," may be common for RNA recognition given the ability of RNA to form alternative structures and the involvement of higher-order structure, or its lack, in the affinity of many RNA binding proteins.

For RNase P, a competing unfavorable structure lowers the free energy of the ground state, while favorable 5′ leader interactions with C5 lower the free energy of the transition state for association (Figure 7). RNA structure negatively affects binding presumably by competing directly with the formation of RNA-protein interactions that require single-stranded conformations (Niranjanakumari et al., 1998; Rueda et al., 2005). This adds a contribution to the activation energy ($\Delta G_{structure}$) resulting in a slower k_{cat}/K_m relative to a substrate in which structure is absent. The formation of favorable RNA-protein interactions with C5 (red circle in Figure 7) stabilizes the RNase P-ptRNA complex ($\Delta G_{RNA\text{-}protein}$) and is therefore observed as tighter binding. Thus, the k_{cat}/K_m is determined by both contributions from independent effects due to intrinsic sequence specificity and from unfavorable effects due to secondary structure. The contribution from unfavorable structure can clearly modulate the observed sequence specificity of an RBP to different extents contingent on whether the local sequences are complementary and available for pairing. Thus, in some circumstances the contribution from this effect arising from binding site context could dominate the observed specificity for alternative RNA substrates independent of effects on RNA-protein interactions.

These results raise the potential for significant contributions from local sequence context on RNase P specificity in vivo. However, little is known about how leader sequence structure and pairing interactions influence RNase P processing. The 5′ leader length is important since leaders lacking sufficient length to interact with the protein are processed more slowly both in vitro and in vivo (Crary et al., 1998; Fredrik Pettersson et al., 2005; Niranjanakumari et al., 1998). Endogenous ptRNAs occur in a variety of precursor RNA contexts. Of the 89 ptRNA genes in *E. coli* K12, 48 are polycistronic, 31 are monocistronic, and 10 occur in rRNA operons. There is a wide range of leader sequence lengths (from 6 to >100 nucleotides) (Fredrik Pettersson et al., 2005; Koutmou et al., 2010; Sun et al., 2006). Among the 69 unique ptRNA genes, 28 have leaders that are less than 15 nucleotides and may only form transient structures as predicted by MFOLD (J.Z., unpublished data). The remaining 35 leaders have the potential to form stable structure; however, for the majority (23) distal pairing interactions exclude the C5 binding site leaving it free. The remaining 12 are predicted to form stable pairing interactions involving N-3-8 that in some cases involves stable stem-loop structures predicted to prevent C5 binding. Thus, like other RBPs the local contexts of C5 binding sites appear to avoid the formation of interfering secondary structure, except in a few notable cases. These possible exceptions may point to instances where structure plays a more important role in modulating the rate of tRNA processing.

SIGNIFICANCE

Understanding how RBPs select cognate RNA binding sites among the excess of non-cognate binding sites in the transcriptome remains a significant challenge. Modeling RNA-protein interaction networks requires accurate prediction of relative affinities of RBPs for alternative RNA binding sites. Overcoming these challenges requires consideration of the problem of RBP specificity comprehensively, and in chemical terms of how sequence variation alters the free energy landscape for RNA binding and processing reactions. Measuring and analyzing quantitative large-scale structure function datasets provides a powerful way to globally determine how RNA sequence variation affects in vitro reaction mechanisms. The methods used for comprehensive characterization of RBP specificity described here utilize standard in vitro RNA enzymologic methods and commercial HiTS and so are likely to be adaptable to a variety of systems. The complex interdependence of sequence and structure in RNA molecular recognition is widely appreciated in molecular and systems biology. However, predicting and understanding the biological consequences of this interdependence in chemical terms has been limited by a lack of quantitative descriptions of the contributions of competing alternative RNA structure RBP to specificity. Thus, the advances reported here help to establish fundamental principles of biological specificity, and provide a detailed

description of RNA discrimination by the essential protein subunit of a ribonucleoprotein enzyme central to RNA metabolism.

EXPERIMENTAL PROCEDURES

RNase P RNA, ptRNA, and randomized ptRNA pools were synthesized by in vitro transcription using plasmid or PCR template DNAs (see Supplemental Information). After PAGE purification RNAs were concentrated by precipitation and stored in buffered aqueous solution at −80°C. RNase P reaction kinetics and equilibrium binding reactions were performed under standard conditions of 50 mM Tris, 100 mM NaCl, 0.005% Triton. Single-turnover reactions contained an additional 17.5 mM $MgCl_2$ with limiting (<10 nM) ptRNA substrate and saturating concentrations (1–5 μM) of *E. coli* RNase P holoenzyme. Multiple-turnover reactions were performed using the same conditions with limiting RNase P (<10 nM) and substrate concentrations indicated in the text. The ptRNA substrates were radioactively labeled at the 5′ end with ^{32}P using standard methods. Unreacted ptRNA substrate and the tRNA and 5′ leader sequence products were resolved by denaturing PAGE and quantified by isotopic counting using a phosphorimager. In Figure 2 the data for single- and multiple-turnover reactions of randomized ptRNA are fit to a single exponential function to illustrate the difference in reaction kinetics compared with the substrate containing the genomically encoded leader sequence. For equilibrium binding reactions Mg^{2+} was replaced with Ca^{2+} to slow the cleavage step and bound ptRNA was resolved from the free population using EMSA.

HiTS-KIN simultaneously measures the k_{cat}/K_m value relative to an internal reference sequence (k_{rel}) for thousands of substrates in a single in vitro kinetic RNA processing reaction. Measurement of k_{rel} values for the ptRNA^{Met82}(N-3-8)21A population were performed as described (Guenther et al., 2013). In brief, the ptRNA population containing six randomized positions N(-3) to N(-8) was reacted in vitro with RNase P. The unreacted ptRNA was purified from individual reaction time points, and the time-dependent change in the distribution of individual ptRNA sequence variants was determined by Illumina sequencing. Internal competition kinetics were used to calculate k_{rel} using the genomically encoded ptRNA^{Met82} 5′ leader sequence as a reference. HiTS-EQ binding analysis is a complementary method that measures the K_A relative to an internal reference sequence ($K_{A,rel}$) for thousands of substrates in a single in vitro RNA equilibrium binding reaction. The overall workflow for HiTS-EQ is similar to HiTS-KIN (see Figure S1). The same randomized ptRNA^{Met82}(N-3-8) populations were used in equilibrium binding reactions, and the free and bound fractions were resolved and purified by EMSA. The free and bound populations were visualized by staining with ethidium bromide and excised using standard methods. The recovered RNA was converted into cDNA and PCR amplified as described in more detail in Supplemental Information. PCR primers contained appropriate adaptors for Illumina sequencing and the individual samples were sequenced using standard 50 bp single-end reads on an Illumina HiSeq 2500 by the CWRU Genomics Core. The concentration-dependent changes in the distribution of sequences in the free and bound populations were analyzed using a simple competitive binding model to calculate the relative association constant ($K_{A,rel}$).

Secondary structures and minimum folding energies for 5′ leader sequences in the ptRNA^{Met82}21A and 21B populations were calculated by the minimum free energy algorithm approach (Zuker and Stiegler, 1981) using the ViennaRNA Package (Lorenz et al., 2011). Quantitative analysis to calculate PWM and a values from the k_{rel} and $K_{A,rel}$ distributions was performed as described. First, the subsets of sequences in the 21A and 21B populations that represent sequences that have minimal effects due to unfavorable secondary structure and therefore reflect intrinsic C5 specificity were selected. Data points were binned that spanned values that were ±0.5 reflective to a line with a slope of m = 0.5 drawn through the reference sequence (at $\ln k_{rel} = \ln K_{A,rel} = 0$) in plots of $\ln k_{rel}$ versus $\ln K_{A,rel}$ as described in the text. The resulting datasets were fit as described in Supplemental Information to

$$\ln({}^{r}K_a) \sim \sum_{i=3}^{8}(a_iA_i + c_iC_i + g_iG_i + u_iU_i) + \sum_{i=1}^{n}\alpha_n I_n$$

which contains terms for specificity, PWM values, at individual nucleotide positions (a_i, c_i, g_i, and u_i), as well as interaction terms, α values, that adjust the PWM score for an individual position depending on the sequence identity at other positions in the binding site. The PWM values express the contribution of individual positions in the binding site to the rate or equilibrium constant distributions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2016.09.002.

AUTHOR CONTRIBUTIONS

Conceptualization, H.C.L., M.E.H., and E.J.; Methodology, H.C.L., B.T., and M.E.H.; Investigation, H.C.L., J.Z., C.N.N., and B.T.; Writing – Original Draft, H.C.L. and M.E.H.; Writing – Review & Editing, H.C.L., C.N.N., M.E.H., and E.J.; Funding Acquisition, M.E.H. and E.J.

ACKNOWLEDGMENTS

Funding was provided by NIH grant RO1 GM056740 (to M.E.H.), R35 GM118011 (to E.J.), C.N.N. was supported by NIH predoctoral training grant GM008056, and B.T. was supported in part by a Provost Summer Undergraduate Research Grant from Case Western Reserve University.

Received: March 2, 2016
Revised: August 5, 2016
Accepted: September 2, 2016
Published: September 29, 2016

REFERENCES

Anderson, V.E. (2015). Multiple alternative substrate kinetics. Biochim. Biophys. Acta *1854*, 1729–1736.

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012). Identification of RNA-protein interaction networks using PAR-CLIP. Wiley Interdiscip. Rev. RNA *3*, 159–177.

Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. *16*, 282.

Buenrostro, J.D., Araya, C.L., Chircus, L.M., Layton, C.J., Chang, H.Y., Snyder, M.P., and Greenleaf, W.J. (2014). Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat. Biotechnol. *32*, 562–568.

Campbell, Z.T., and Wickens, M. (2015). Probing RNA-protein networks: biochemistry meets genomics. Trends Biochem. Sci. *40*, 157–164.

Cook, K.B., Hughes, T.R., and Morris, Q.D. (2015). High-throughput characterization of protein-RNA interactions. Brief. Funct. Genomics *14*, 74–89.

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-tRNAAsp. Biochemistry *37*, 9409–9416.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Farcasiu, D. (1975). The use and misuse of the Hammond Postulate. J. Chem. Educ. *52*, 76.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. *224*, 771–782.

Fredrik Pettersson, B.M., Ardell, D.H., and Kirsebom, L.A. (2005). The length of the 5' leader of *Escherichia coli* tRNA precursors influences bacterial growth. J. Mol. Biol. *351*, 9–15.

Gosai, S.J., Foley, S.W., Wang, D., Silverman, I.M., Selamoglu, N., Nelson, A.D., Beilstein, M.A., Daldal, F., Deal, R.B., and Gregory, B.D. (2015). Global

analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. Mol. Cell *57*, 376–388.

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.

Hsieh, J., and Fierke, C.A. (2009). Conformational change in the *Bacillus subtilis* RNase P holoenzyme–pre-tRNA complex enhances substrate affinity and limits cleavage rate. RNA *15*, 1565–1577.

Hsieh, J., Andrews, A.J., and Fierke, C.A. (2004). Roles of protein subunits in RNA-protein complexes: lessons from ribonuclease P. Biopolymers *73*, 79–89.

Iadevaia, V., and Gerber, A.P. (2015). Combinatorial control of mRNA fates by RNA-binding proteins and non-coding RNAs. Biomolecules *5*, 2207–2222.

Jankowsky, E., and Harris, M.E. (2015). Specificity and non-specificity in RNA binding. Nat. Rev. Mol. Cell Biol. *16*, 533–544.

Jencks, W.P. (1985). A primer for the Bema Hapothle. An empirical approach to the characterization of changing transition-state structures. Chem. Rev. *85*, 511–527.

Kirby, A.J., and Nome, F. (2015). Fundamentals of phosphate transfer. Acct Chem. Res. *48*, 1806–1814.

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.

Kurz, J.C., Niranjanakumari, S., and Fierke, C.A. (1998). Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNAAsp. Biochemistry *37*, 2393–2400.

Lassila, J.K., Zalatan, J.G., and Herschlag, D. (2011). Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. Annu. Rev. Biochem. *80*, 669–702.

Li, F., Zheng, Q., Vandivier, L.E., Willmann, M.R., Chen, Y., and Gregory, B.D. (2012). Regulatory impact of RNA secondary structure across the *Arabidopsis transcriptome*. Plant Cell *24*, 4346–4359.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

Mackereth, C.D., and Sattler, M. (2012). Dynamics in multi-domain protein recognition of RNA. Curr. Opin. Struct. Biol. *22*, 287–296.

Maenner, S., Müller, M., Fröhlich, J., Langer, D., and Becker, P.B. (2013). ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. Mol. Cell *51*, 174–184.

Matouschek, A., and Fersht, A.R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. Proc. Natl. Acad. Sci. USA *90*, 7814–7818.

Mitchell, S.F., and Parker, R. (2014). Principles and properties of eukaryotic mRNPs. Mol. Cell *54*, 547–558.

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. USA *95*, 15212–15217.

Niranjanakumari, S., Day-Storms, J.J., Ahmed, M., Hsieh, J., Zahler, N.H., Venters, R.A., and Fierke, C.A. (2007). Probing the architecture of the *B. subtilis* RNase P holoenzyme active site by cross-linking and affinity cleavage. RNA *13*, 521–535.

Ray, W.J., Jr. (1983). Rate-limiting step: a quantitative definition. Application to steady-state enzymic reactions. Biochemistry *22*, 4625–4637.

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.

Rueda, D., Hsieh, J., Day-Storms, J.J., Fierke, C.A., and Walter, N.G. (2005). The 5' leader of precursor tRNAAsp bound to the *Bacillus subtilis* RNase P holoenzyme has an extended conformation. Biochemistry *44*, 16130–16139.

Schneider, T.D., Stormo, G.D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. J. Mol. Biol. *188*, 415–431.

Schrader, J.M., Chapman, S.J., and Uhlenbeck, O.C. (2009). Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis. J. Mol. Biol. *386*, 1255–1264.

Shi, Z., and Barna, M. (2015). Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. Ann. Rev. Cell Dev. Biol. *31*, 31–54.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The clothes make the mRNA: past and present trends in mRNP fashion. Annu. Rev. Biochem. *84*, 325–354.

Sosnick, T.R. (2008). Kinetic barriers and the role of topology in protein and RNA folding. Protein Sci. *17*, 1308–1318.

Stormo, G.D. (2013). Modeling the specificity of protein-DNA interactions. Quant Biol. *1*, 115–130.

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.

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.

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., and Steenackers, H.P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. Front. Microbiol. *6*, 141.

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.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell *40*, 939–953.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. Nucleic Acids Res. *40*, e54.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. *9*, 133–148.