Supplementary Information For:

RNase P specificity for 5' leaders of pre-tRNA is governed by energetic coupling between RNA and protein contacts

Courtney N. Niland[1], David R. Anderson[2], Eckhard Jankowsky[3], Michael E. Harris[1]

[1]Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106

[2]School of Business, CUNY Baruch College, New York, NY 10010

[3]Center for RNA Molecular Biology, Case Western Reserve University School of Medicine, Cleveland, OH 44106

To whom correspondence should be addressed: Michael E. Harris, Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland OH 44106. Phone: 216-368-4779; Email: meh2@cwru.edu

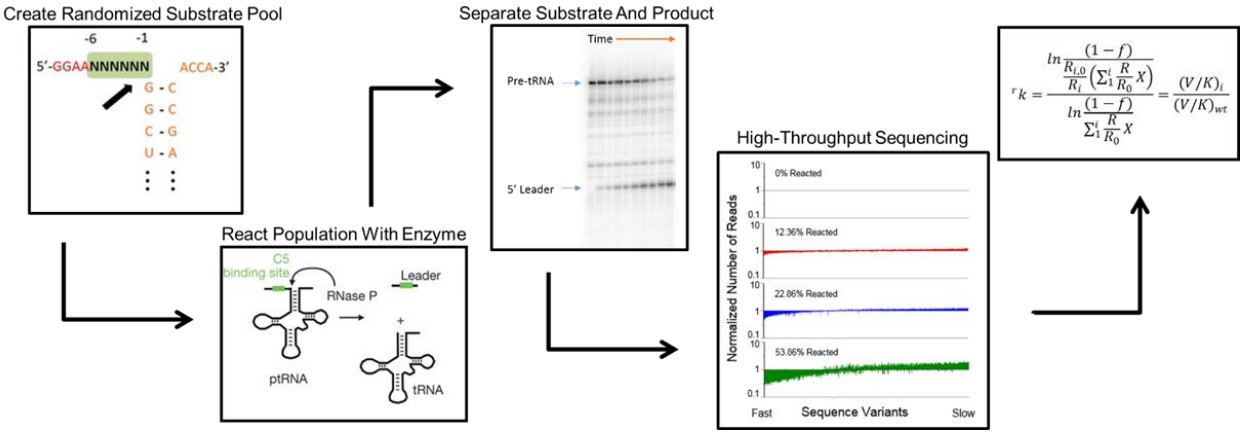

Figure S1: Schematic of the HTS-Kin procedure. First, the RNA is randomized in sequence at the region of interest, here at the -1 to -6 positions in the 5' leader yielding 4,096 pre-tRNA variants. Next, we react that substrate pool with RNase P, and thus high affinity substrates will be cleaved quickly, while the lower affinity pre-tRNAs will take longer to react. Then, we isolate the residual substrate population at different reaction times by gel purification and perform RT-PCR to analyze by Illumina sequencing. The number of high-throughput sequencing reads allows us to estimate the change in abundance of each pre-tRNA as a function of time. As shown, if we align sequences from fast to slow reacting, and compare the number of reads, we can see that as the reaction progresses, fast reacting sequences become depleted from the substrate population while slower substrates accumulate. We calculate for each substrate a relative rate constant, *or* k_{rel} *value*, (which is a ratio of its k_{cat}/K_m to that of the genomically encoded reference) using the total fraction of reaction, f, and the ratio of the number of reads of the sequence to the WT reference represented by R.

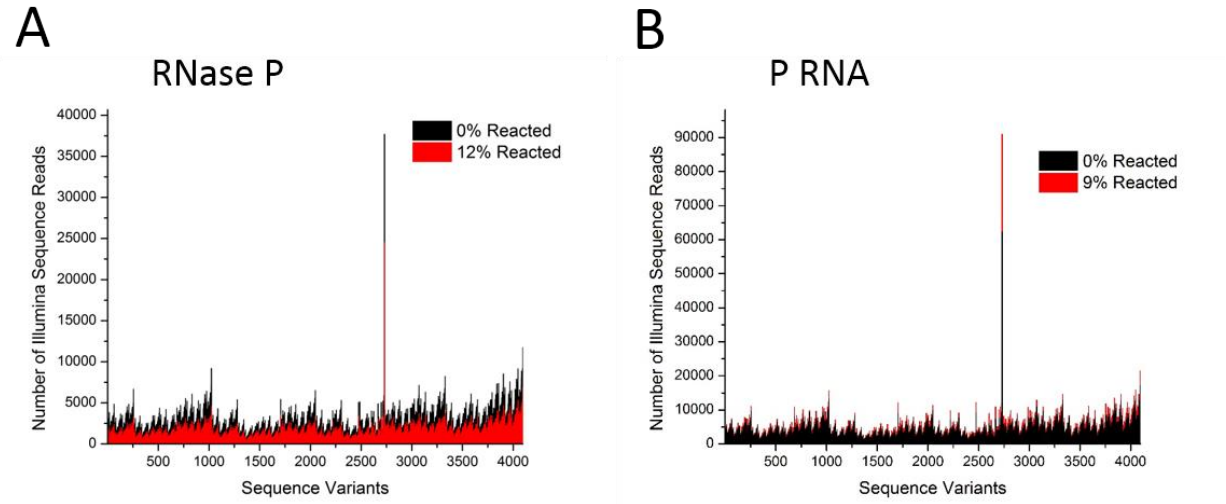

Figure S2: Representative histograms of the change in Illumina sequence reads for each of the 4,096 pre-tRNA 5' leader sequence variants. A) Plot of sequence reads for a reaction of the randomized substrate pool with RNase P. A noticeable decrease in sequence reads is apparent for most substrates as the reaction progresses. B) Plot of the raw number of sequence reads from Ribozyme reaction with the randomized substrate pool. A small change in the number of reads is observed and several sequences show increased reads over time consistent with their much slower processing rate constants. The outlier on both plots showing a large number of reads contains the 5' leader sequence 5'-GGGGGG-3' from -6 to -1.

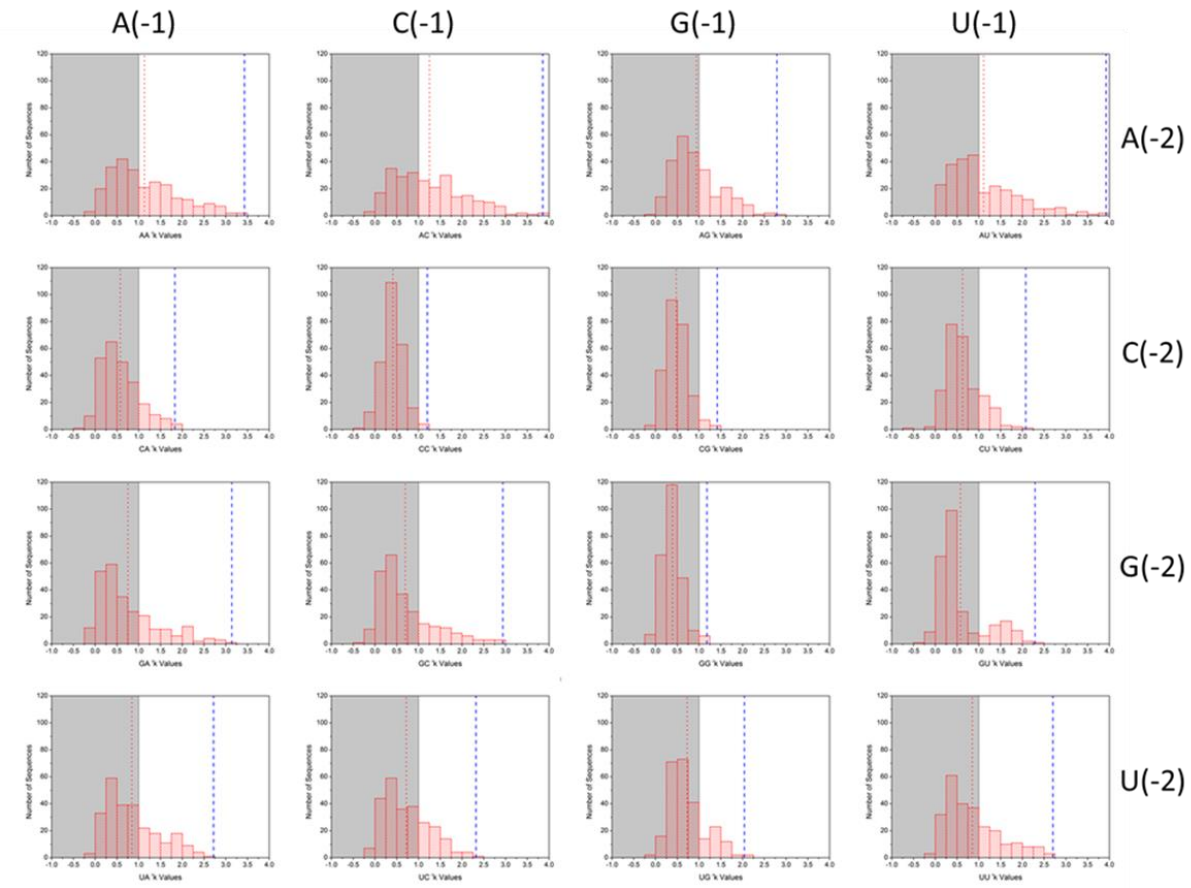

Figure S3: Dividing the holoenzyme HTS-Kin data into subsets reveals extensive coupling between the sequence identity in the RNA binding site on sequence specificity of the protein binding site in the 5' leader of pre-tRNA. The HTS-Kin data was divided into 16 subsets based on the identity of nucleotides at N(-2) and N(-1) in the RNA binding site. Across each row symbolizes a change in nucleotide identity at the N(-1) position while down a column follows a change at N(-2). Each of the 16 boxes represents the affinity distribution of substrate variants with a common set of nucleotides at N(-2) and N(-1) in the 5' leader in which only nucleotides contacting the C5 protein are allowed to vary (N(-6) to N(-3)). The dashed red line indicates the average k_{rel} of this population and the dashed blue line shows the highest k_{rel} in the data subset.