Supporting Information

Lucks et al. 10.1073/pnas.1106501108

SI Materials and Methods.

Synthesis of *Bacillus subtilis* RNase P, pT181 sense RNAs. A DNA template for transcription of each RNA, inserted in the context of the 3′ flanking structure cassette (1), was generated by PCR [1 mL; containing 20 mM Tris (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 200 μM each dNTP, 500 nM each forward and reverse primer, 5 pM template, and 0.025 units/μL Taq polymerase; denaturation at 94 °C, 45 s; annealing 55 °C, 30 s; and elongation 72 °C, 1 min; 34 cycles]. The PCR product was recovered by ethanol precipitation and resuspended in 150 μL of TE [10 mM Tris (pH 8.0), 1 mM EDTA]. Transcription reactions (1.0 mL, 37 °C, 12–14 h) contained 40 mM Tris (pH 8.0), 20 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 0.01% (vol/vol) Triton X-100, 5 mM each NTP, 50 μL of PCR-generated template, 0.12 U/μL RNase Inhibitor (Promega), and 0.1 mg/mL of T7 RNA polymerase. The RNA products were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 32 W, 2 h), excised from the gel, and recovered by passive elution and ethanol precipitation. The purified RNA (approximately 3 nmol) was resuspended in 200 μL TE. All RNAs except for the "non-bar-coded" RNase P RNA contained a four nucleotide degenerate bar code between the reverse transcriptase primer binding site and the 3′ linker region within the 3′ flanking structure cassette (Table S1).

Structure-Selective RNA Modification. All RNAs were modified with 1M7 (6.5 mM, final) as described previously with minor modifications for the analysis by deep sequencing (2). For the deep sequencing samples, RNAs (50 pmol) were folded in a volume of 90 μL and modified in a total volume of 50 μL. Recovered modified RNA was resuspended in 10 μL of 1/2 × TE. Each library used a 100 pmol total of RNA [50 pmol for each the (+) and the (−) reactions]. For the RNase P library 50 pmols each of non-bar-coded RNA and bar-coded RNA was used (257 total different RNAs). For the more complex mixture library, 14.2 pmol of each RNA was used (seven total different RNAs).

SHAPE (Selective 2′-Hydroxyl Acylation Analyzed by Primer Extension) Analysis by Deep Sequencing. The general procedure for first strand cDNA synthesis was carried out following the primer extension protocol outlined elsewhere (3) except that reverse transcription primers were unlabeled and the 5′ end contained a sequence corresponding to A_adapter_t (4) (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) and a handle of RRRY (R = A,G; Y = C,T) or YYYR for (+) and (−) primer extension reactions, respectively. Other exceptions were the total amount of primer used in primer extension reactions was 9 pmol and a 95 °C heat step was added prior to primer annealing to reduce possible intermolecular interactions of the primer. After primer extension, RNA was hydrolyzed by adding NaOH (1 μL, 4 M) and incubating for 5 min at 95 °C. cDNAs were ethanol precipitated and resuspended in 35.5 μL of nuclease-free water. A_adapter b (4)

(AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG) modified with a 3′-C3 spacer group was ligated to each cDNA using a ssDNA ligase (CircLigase, Epicentre Biotechnologies) [100 μL; 50 mM MOPS (pH 7.5), 10 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.05 mM ATP, 2.5 mM $MnCl_2$, 5 μM A_adapter_b, and 200 U ligase] and incubating for 6 h at 68 °C in a thermal cycler. Separate ligation reactions were carried out for the (+) and (−) cDNA library pools. The ligation reactions were stopped by heating to 80 °C for 10 min, recovered by ethanol precipitation, and resuspended in 20 μL of nuclease-free water. Excess A_adapter_b was removed using Agencourt Ampure XP beads (A63880). Finally, 9 or 12 cycles of PCR amplification were performed on the library using primers that bind to the Illumina adapter sequences, followed by cleanup with Agencourt Ampure XP beads (5). No size selection was performed on the resulting adapter-ligated library. Libraries were assayed for quality (on an Agilent Bioanalyzer 2100 using a high-sensitivity DNA chip) by comparing 9- and 12-cycle amplification, looking for characteristic peaks and peak enrichment (Fig. S9). Libraries were sequenced on an Illumina Genome Analyzer IIx following the manufacturer's standard cluster generation and sequencing protocols, for 50 cycles of sequencing per paired-end read (6). Data available upon request.

Bioinformatic Analysis of Bar-Coded Sequencing Reads. Reads for reverse transcription (RT) fragments were first split into 1M7-treated and -untreated pools by examining the 4 nucleotide handle sequence on the 5′ end of the read generated from the 3′ end of each RNA probed in the experiment. Reads with an RRRY handle identified (+) fragments and those with YYYR identified (−) fragments. This handle was then trimmed from each read to allow alignment of the reads to probed RNAs. Reads were then trimmed for A_adapter_b and A_adapter_t using the FASTX toolkit [http://hannonlab.cshl.edu/fastx_toolkit/], because RT products shorter than the length of a sequencing read will produce reads with adapter at their 3′ ends.

Paired reads were optimally aligned to the probed RNAs using Bowtie 0.12.8 (7) with the parameters - -best -X 2000 –y –m 1, and allowing no mismatches in the alignments with parameters -v 0. The 3′ end of each fragment alignment (toward the 5′ end of the probed RNA) corresponds to the point at which RT stopped. Two counters (one for the 1M7 condition and one for the control condition) at each position of each probed RNA position were used to track RT stopping points. The (+) counter for a probed RNA at position i was incremented when a fragment from the (+) pool aligned to the RNA starting at position $(i + 1)$ and ending at the RNA's 3′ end. The (−) counter was incremented for (−) pool fragment alignments. These RT-stop counts were then used to calculate maximum likelihood reactivities.

Calculating Jensen–Shannon Divergences (JSD). Jensen–Shannon divergences between two sets of reactivities, Θ^1 and Θ^2, were calculated according to the formula

$$\mathrm{JSD}_{1,2} = H(\Theta^{\mathrm{avg}}) - \frac{1}{2}(H(\Theta^1) + H(\Theta^2)),$$

where

$$H(\Theta) = -\sum_i \theta_i \ln \theta_i.$$

θ_i is the reactivity at position i, and Θ^{avg} is the average of Θ^1 and Θ^2.

Secondary Structure Prediction Using SHAPE-Seq Reactivity Constraints. SHAPE intensities were converted into a pseudo-free-energy change term in the RNAstructure program (8):

$$\Delta G_{\mathrm{SHAPE}} = m^* \ln[\text{SHAPE-Seq reactivity} + 1.0] + b.$$

The intercept, b, is the free-energy bonus for formation of a base pair with zero or low SHAPE reactivity, whereas m, the slope, drives an increasing penalty for base pairing as the SHAPE reactivity increases. These parameters dictate the strength of the experimental contribution to the energy function. The b and m parameters used were −0.5 and 3.4 kcal/mol, respectively (9).

1. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127:4223–4231.
2. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129:4144–4145.
3. Mortimer SA, Weeks KM (2009) Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* 4:1413–1421.
4. Kozarewa I, et al. (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes. *Nat Methods* 6:291–295.
5. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223.
6. Bentley DR, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
7. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
8. Reuter JS, Mathews DH (2010) RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129.
9. Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52:150–158.

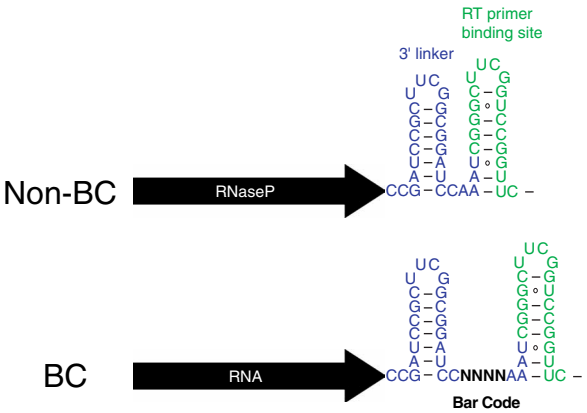

Fig. S1. The 3′ structure cassette used in this study. The structure cassette forms the 3′ end of the RNA and contains an RT primer binding site (green) and a linker region (blue) to separate this site from the RNA sequence under study. As originally designed (1), the structure cassette should fold into two hairpin structures independent of the RNA under study. The non-bar-coded (Non-BC) structure cassette is identical to that in ref. 1. The bar-coded (BC) structure cassette has a 4-nt bar-code sequence placed in between the two hairpins.

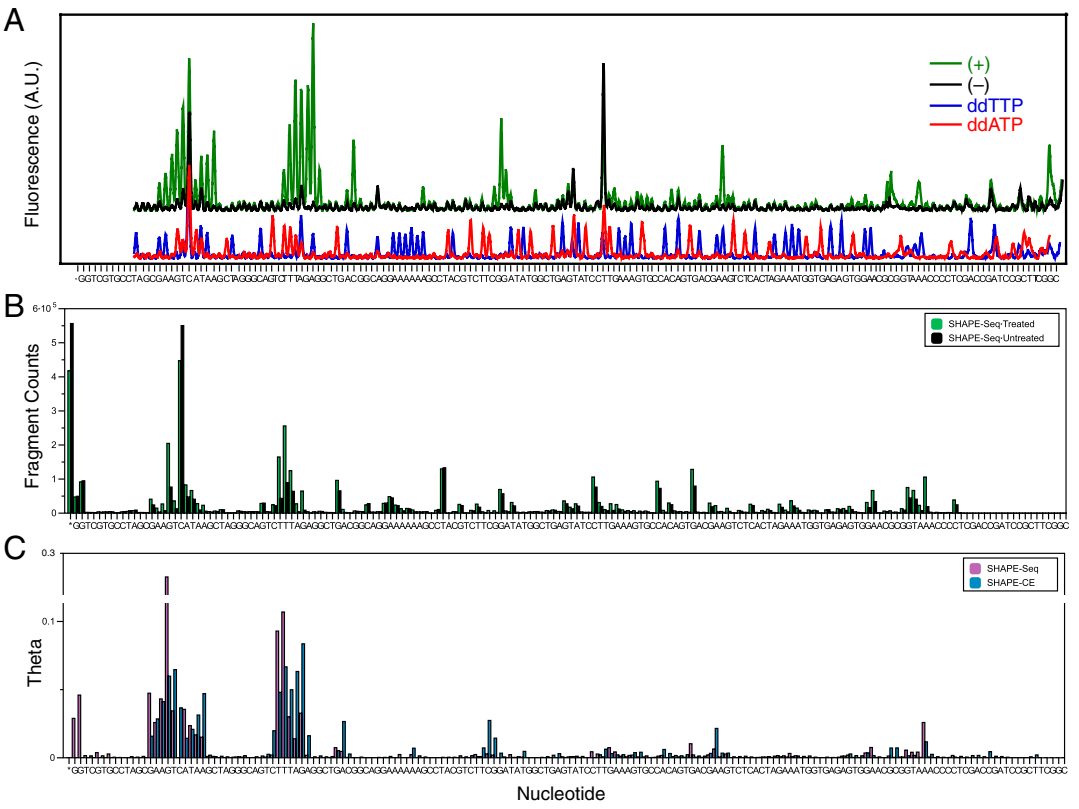

Fig. S2. WT RNaseP SHAPE data analyzed by capillary electrophoresis (SHAPE-CE) or direct sequencing (SHAPE-Seq). (*A*) Electropherograms for SHAPE-CE (+) and (–) reactions and two sequencing reactions. (*B*) Mapped fragment counts for the SHAPE-Seq (+) and (–) channels. The asterisk represents fragments mapped to the full-length molecule. Note that SHAPE-Seq is capable of mapping the 5′ end of the molecule, which is difficult in SHAPE-CE due to the limits of capillary electrophoresis detection. (*C*) Derived Θ from the data in *A* and *B*. Note that for most of the 3′ half of the molecule, the data in the (+) and (–) channel cancel to give very low Θ as expected. The two islands of high Θ on the 5′ portion of the molecule correspond to the P8 and P9 loops. It is important to note that there are a large proportion of counts in these regions in SHAPE-Seq (–) raw data compared to SHAPE-CE raw data, reflecting the natural bias due to RNA structure in the SHAPE-Seq (–) channel. Because the (–) channel represents an RNA-Seq experiment, this is direct evidence of these biases in that technique. The remaining nucleotides on the 3′ end of the molecule are part of the structure cassette (Fig. S1) and show no reactivities.

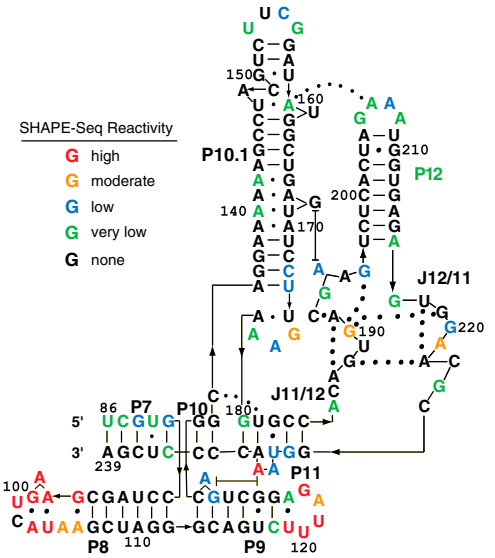

Fig. S3. Structure diagram representing the base pairing and tertiary interactions for the specificity domain of *Bacillus subtilis* RNase P color-coded according to SHAPE-Seq reactivities (see Fig. 2).

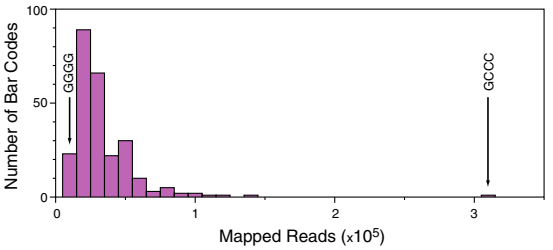

Fig. S4. RNaseP bar code counts distribution. The range of mapped reads for each bar code was from 10,488 (GGGG) to 308,978 (GCCC).

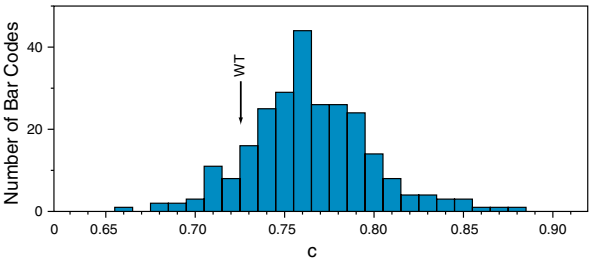

Fig. S5. Distribution of *c* over bar-coded RNase Ps. The average *c* over all bar codes is 0.76, which is close to the *c* for the WT (un-bar-coded) RNaseP sequence, 0.73.

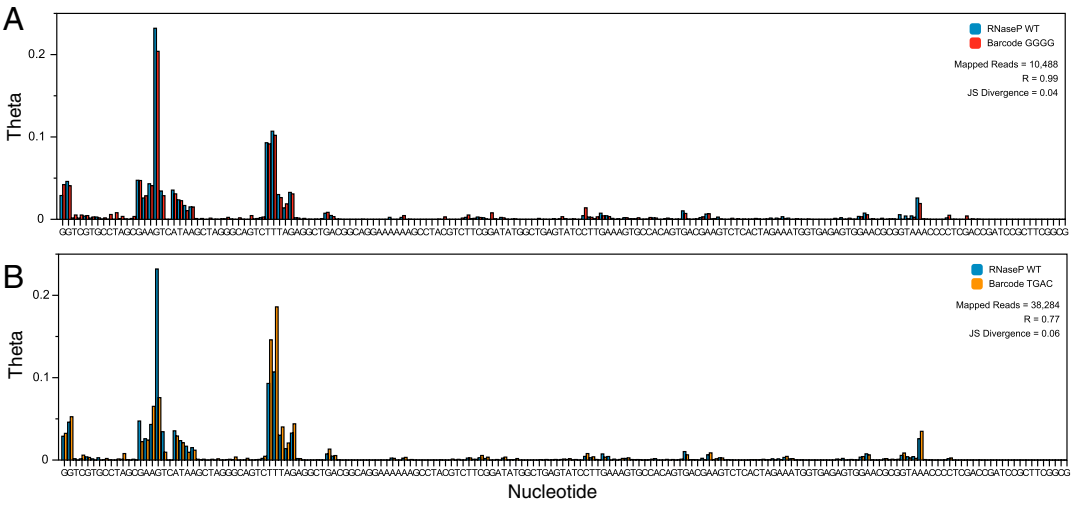

Fig. S6. Comparison of WT RNaseP with bar-coded RNase Ps. (*A*) Comparison with the GGGG bar code, which had the minimum number of fragments, mapped (Fig. S4). This number of fragments corresponded to approximately 0.1 pmol of starting RNA, but still yielded highly accurate SHAPE reactivities. (*B*) Comparison with the TGAC bar code, which gave the worst Jensen–Shannon divergence and R with the WT Θ (Fig. 2*C*). It is clear from the plot of Θ that the overall reactivities are accurately captured and that the high JSD and low R result from a few positions in the highly reactive loops having different absolute reactivities. We note that these positions would all be classified as being "highly reactive," and thus this reactivity spectrum would be interpreted exactly the same as the WT RNaseP spectrum. The WT RNaseP Θ were derived from over 7.5 million sequencing reads.

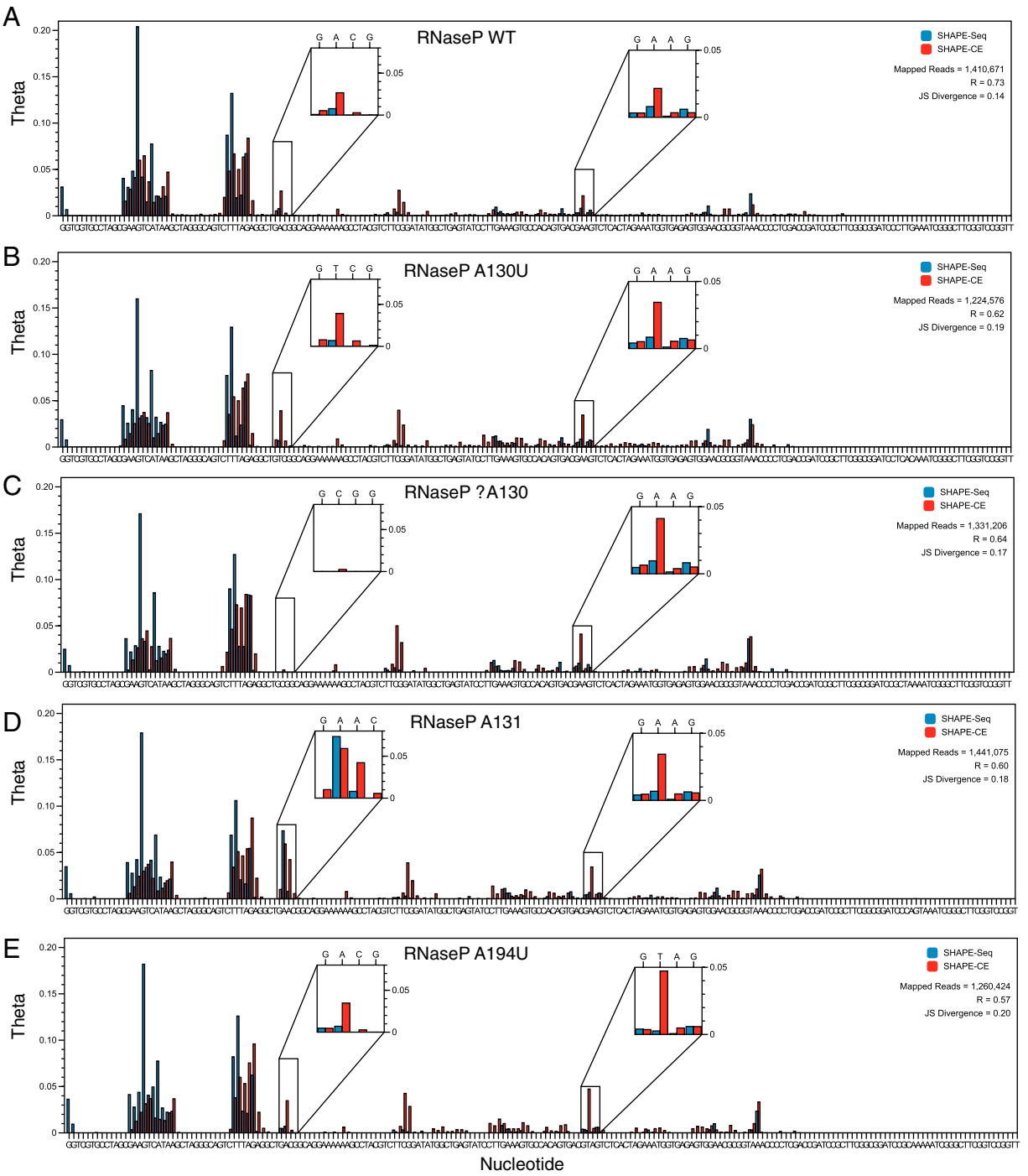

Fig. S7. SHAPE-Seq versus SHAPE-CE for the RNaseP variants shown in Fig. 3. The A130 and A194 mutation regions are highlighted by insets. R and Jensen–Shannon divergence were calculated over the subset of nucleotides for which CE data were available. Prior to the calculation, SHAPE-CE peak areas were normalized to sum to 1, and SHAPE-Seq Θ values were renormalized to sum to 1 in this range. The nucleotide range includes the 3′ structure cassette and bar code (Table S1) for each molecule.

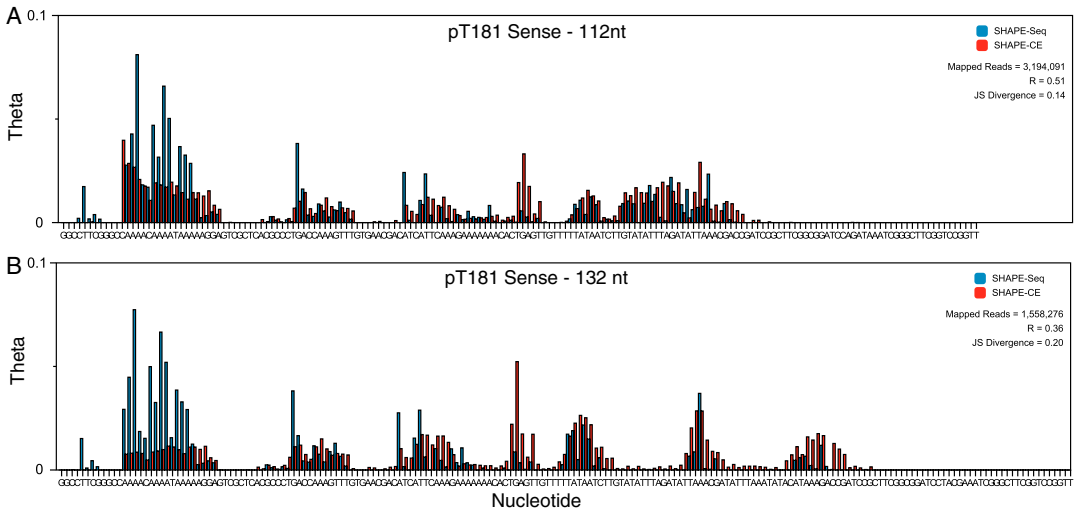

Fig. S8. SHAPE-Seq versus SHAPE-CE for the pT181 Sense variants shown in Fig. 4. The nucleotide range includes the 5′ and 3′ structure cassettes and bar code (Table S1) for each molecule. The observation that the sequence and structures of these two RNAs are identical on the 5′ end (1) is exactly recapitulated in the SHAPE-Seq data and is seen by comparing the reactivity profiles of these two RNAs in that region (blue bars at 5′ end). This is in contrast to the SHAPE-CE data, which displays variable amplitudes most likely due to the inability to apply a rigorous mathematical analysis which corrects for signal drop off that can vary due to RNA lengths (red bars at 5′ end).

1. Brantl S, Wagner EGH (2000) Antisense RNA-mediated transcriptional attenuation: An in vitro study of plasmid pT181. *Mol Microbiol* 35:1469–1482.

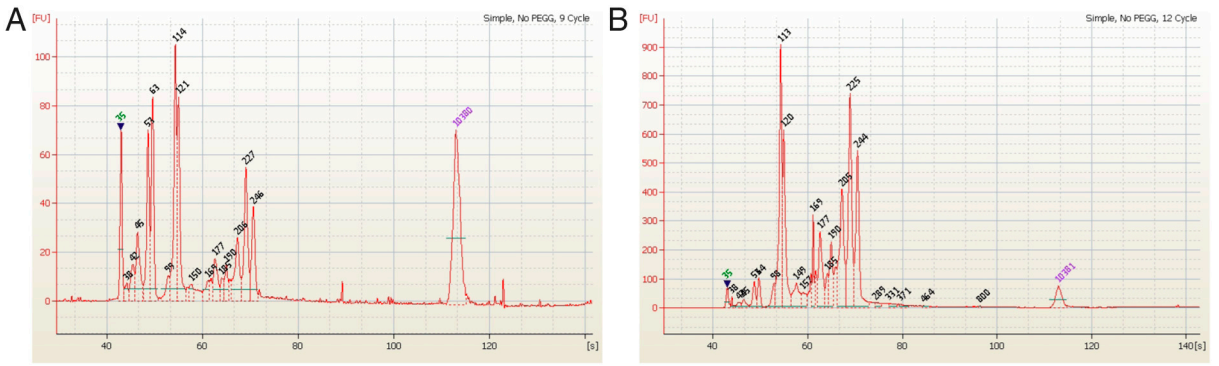

Fig. S9. Bioanalyzer traces for the bar-coded RNaseP library showing (*A*) 9 and (*B*) 12 cycles of PCR amplification. Library quality is assessed by looking for distinct peaks that are enriched from 9 to 12 cycles and that lie in approximately the expected size range. Approximate sizes of fragments are noted on the top of each peak. Enrichment can be observed by the change in intensity scale from *A* to *B*.

Table S1. Bar-coded RNA sequences used in this study

RNA	Bar code	Sequence
RNaseP WT	CTTG/NNNN	GGUCGUGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGACGGCAGGAAAAAAGCCUACGU CUUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGAAGUCUCACUAGAAAUGGUGAGAGU GGAACGCGGUAAACCCCUCGACCGAU<u>CCGCUUCGGCGGAUCC</u>**CUUG**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
RNaseP A130U	TCAC	GGUCGUGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGUCGGCAGGAAAAAAGCCUACGU CUUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGAAGUCUCACUAGAAAUGGUGAGAGU GGAACGCGGUAAACCCCUCGACCGAU<u>CCGCUUCGGCGGAUCC</u>**UCAC**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
RNaseP ΔA130	GCTA	GGUCGUGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGCGGCAGGAAAAAAGCCUACGUC UUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGAAGUCUCACUAGAAAUGGUGAGAGU GGAACGCGGUAAACCCCUCGACCGAU<u>CCGCUUCGGCGGAUCC</u>**GCUA**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
RNaseP A131	CAGT	GGUCGUGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGAACGGCAGGAAAAAAGCCUACG UCUUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGAAGUCUCACUAGAAAUGGUGAGAG UGGAACGCGGUAAACCCCUCGACCGAU<u>CCGCUUCGGCGGAUCC</u>**CAGU**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
RNaseP A194U	GCAA	GGUCGUGCCUAGCGAAGUCAUAAGCUAGGGCAGUCUUUAGAGGCUGACGGCAGGAAAAAAGCCUACGU CUUCGGAUAUGGCUGAGUAUCCUUGAAAGUGCCACAGUGACGUAGUCUCACUAGAAAUGGUGAGAGU GGAACGCGGUAAACCCCUCGACCGAU<u>CCGCUUCGGCGGAUCC</u>**GCAA**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
pT181 Sense 112	AGAT	GGCCUUCGGGCCAAAACAAAAUAAAAAGGAGUCGCUCACGCCCUGACCAAAGUUUGUGAACGACAUCAU UCAAAGAAAAAAACACUGAGUUGUUUUUAUAAUCUUGUAUAUUUAGAUAUUAAACGA <u>CCGAUCCGCUUCGGCGGAUCC</u>**AGAU**<u>AAAUCGGGCUUCGGUCCGGUUC</u>
pT181 Sense 132	TACG	GGCCUUCGGGCCAAAACAAAAUAAAAAGGAGUCGCUCACGCCCUGACCAAAGUUUGUGAACGACAUCAU UCAAAGAAAAAAACACUGAGUUGUUUUUAUAAUCUUGUAUAUUUAGAUAUUAAACGAUAUUUAAAU AUACAUAAAGA<u>CCGAUCCGCUUCGGCGGAUCC</u>**UACG**<u>AAAUCGGGCUUCGGUCCGGUUC</u>

Bar codes are highlighted in bold and are embedded in the 3′ structure cassette, which is underlined.. The two G residues on the 5′ end of each RNA are added by T7 RNA Polymerase during in vitro transcription of the RNA (see "Synthesis of *Bacillus subtilis* RNase P, pT181 sense RNAs").