Time-resolved, single-molecule, correlated chemical probing of RNA

Jeffrey E. Ehrhardt and Kevin M. Weeks
Department of Chemistry
University of North Carolina

Supporting Information

Methods

TMO. Trimethyloxonium tetrafluoroborate (TMO) is a nonvolatile, hygroscopic salt. TMO stocks should be handled in a fume hood and stored in desiccator at -20 °C. Due to the high reactivity of TMO, stock solutions of TMO were prepared in 1:2 (v/v) nitromethane/sulfolane (NS), as follows: TMO (MilliporeSigma) was first fully dissolved in nitromethane at a concentration of 3.0 M; subsequently, the TMO solution was diluted with two volumes sulfolane. Final concentrations of TMO stock solutions used in this work (before dilution with RNA) were 1.0 M. Sulfolane is a solid at room temperature; it was warmed to 37 °C immediately prior to use. Due to the fast TMO reaction kinetics, the large volume of RNA solution was added to the smaller TMO volume and mixed by immediate rapid pipetting. Unlike DMS probing, no quench step is needed to stop the TMO reaction. TMO was identified from a small screen of strong alkylating reagents.

Safe handling notes. TMO should be stored in a desiccator at -20 °C; large volumes or solid salt should be handled in a fume hood. TMO is nonvolatile, making small volumes (1 mL) safe to handle outside of the fume hood once dissolved in NS solution. The salt and its hydrolysis products are corrosive: gloves, lab coat, and safety glasses should be used at all times during handling. TMO stock solutions and TMO waste can be quenched in the fume hood by addition of water before (acid) waste disposal.

Hydrolysis rate determination. Hydrolysis was monitored under the same solution conditions as used for structure probing by adding 200 µL of chemical probe mixture (1.0 M TMO or 1.7 M DMS in NS) to 1800 µL of reaction buffer (11.1 mM $MgCl_2$, 111 mM NaCl, 333 mM Bicine, pH 8.0) equilibrated at 37 °C. Mixing the components yielded final concentrations of 100 mM TMO or 170 mM DMS in 10 mM $MgCl_2$, 100 mM NaCl, 300 mM Bicine, pH 8.0, 3.3% (v/v) nitromethane, and 6.6% (v/v) sulfolane. Pseudo-first order rate constants were obtained by monitoring the proportional pH change over time (PASPORT pH sensor connected to a Pasco interface, sampling frequency of 2 Hz). pH values were normalized to values between 0 and 1, where pH_x is the *pH* at time x, $pH_{t=\infty}$ is the final *pH* value, and $pH_{t=0}$ is the initial *pH*:

$$pH_{norm} = \frac{(pH_x - pH_{t=\infty})}{(pH_{t=0} - pH_{t=\infty})}$$

After normalization, hydrolysis rates were determined by fitting a time series to a first-order rate equation; kinetic fits had $R > 0.9$.

RNA synthesis. DNA templates (IDT) encoding the *B. stearothermophilus* RNase P catalytic domain[1], inserted between 5' and 3' structure cassette flanking sequences[2] and preceded by a 5' T7 promoter sequence, were amplified by PCR. The DNA template was recovered (Omega Mag-Bind beads) and eluted in water. RNAs were transcribed *in vitro* (2.5 mM each NTP, 25 mM $MgCl_2$, 40 mM Tris, pH 8.0, 2.5 mM spermidine, 0.01% (wt/vol) Triton X-100, 10 mM DTT, 0.025 units pyrophosphatase, 2.5 µg T7 polymerase, 1.125 µg DNA template; in 50 µL H_2O; 37 °C; 6 h). RNA was recovered (Omega Mag-Bind beads) and resuspended in 50 µL of 0.1× TE (1 mM Tris, pH 8.0, 0.1 mM EDTA).

Equilibrium RNA structure probing. RNA (5 pmol) in 6 µL 0.1× TE was heated at 95 °C for 2 min, cooled on ice, and mixed with 3 µL of 3× folding buffer (33 mM $MgCl_2$, 333 mM NaCl, 1 M Bicine, pH 8.0). Reaction conditions use Bicine buffered to pH 8.0, conditions previously optimized to maximize the reactivity of DMS with all four ribonucleotides[3] and which proved well suited for TMO. Bicine should be pH 8.0 at the reaction temperature; buffer stocks titrated to pH 8.3 at room temperature will then have pH 8.0 at 37 °C due to the temperature dependence of Bicine pK_a. The resulting solution, containing 5 pmol of RNA in 110 mM NaCl, 333 mM Bicine, and 11 mM $MgCl_2$, was incubated at 37 °C for 20 min. The RNA solution was added to 1 µL of reagent (1.0 M TMO or 1.7 M DMS in NS). The TMO reaction is self-quenching; in contrast, DMS required quenching with 1 µL of 2-mercaptoethanol after 6 minutes of incubation at 37 °C. The no-reagent control contained 1 µL of NS. Modified RNA was recovered (Omega Mag-Bind beads) and resuspended in 20 µL of 0.1× TE.

Time-resolved RNA structure probing. Recently developed, optimized conditions (Bicine, pH 8.0) allow TMO to react with all four ribonucleotides (see Supporting Figure 3 and ref. 3). RNA (65 pmol) in 52 µL 0.1× TE was heated at 95 °C for 2 min, cooled on ice, and mixed with 39 µL of 3× folding buffer (333 mM NaCl, 1 M Bicine, pH 8.0). The resulting solution, containing 65 pmol of RNA in 110 mM NaCl and 333 mM Bicine, was incubated at 37 °C for 10 min. A Mg^{2+}-free sample was removed for addition to TMO reagent. Tertiary structure folding was initiated by adding 12 µL of 100 mM $MgCl_2$ (final Mg^{2+} concentration equal to 10 mM). After mixing, 9 µL of RNA solution was removed (at 5, 10, 15, 20, 30, 60, 120, 180, 300, 600, 900, and 1200 s) and added directly into 1 µL of 10× reagent (1.0 M TMO in NS). For the no-reagent control, the aliquot was added to 1 µL NS. Modified RNA was recovered with Omega Mag-Bind beads and resuspended in 20 µL of 0.1× TE.

Reverse transcription. MaP was performed essentially as described[4,5]. Use of the MaP detection strategy is important, as MaP allows detection of multiple chemical modifications in RNA, including of low frequency modification events, from single RNA molecules[3,6] (see Supporting Figure 3). In brief, a 10 µL solution containing RNA, 200 nM gene specific primer (Table S1), and 2 mM premixed dNTPs was incubated at 65 °C for 5 min followed by incubation at 4 °C for 2 min. To the solution was added 9 µL of 2.22× MaP buffer (1× MaP buffer contains 1M betaine, 50 mM Tris, pH 8.0, 75 mM KCl, 10 mM DTT, 6 mM $MnCl_2$), and the mixture was incubated at room temperature for 2 min. SuperScript II Reverse Transcriptase (1 µL, Invitrogen) was added, and reverse transcription was performed according to the following temperature program: 25 °C for 10 min, 42 °C for 90 min, 10 × [50 °C for 2 min, 42 °C for 2 min],

72 °C for 10 min. cDNA was then purified (Illustra MicroSpin G-50 columns, GE Healthcare).

Library preparation and sequencing. Sequencing libraries were prepared from cDNA products using a two-step PCR approach (Table S1)[3]. In the first PCR step, a 5-µL aliquot of purified cDNA was amplified for Illumina sequencing with the following temperature program: 98 °C for 30 s, 15 × [98 °C for 5 s, 68 °C for 20 s, 72 °C for 20 s], 72 °C for 2 min. The PCR product was recovered (Omega Mag-Bind beads) and eluted in water. In the second PCR step, treatment-specific barcodes were added to the ends of amplicons, with the following temperature program: 98 °C for 30 s, 15 × [98 °C for 5 s, 68 °C for 20 s, 72 °C for 20 s], 72 °C for 2 min. PCR products were recovered (Omega Mag-Bind beads), pooled, and sequenced (Illumina MiSeq instrument; 500 cycle kit).

Sequence alignment, mutation parsing, and reactivity calculation. *ShapeMapper* (v2.1.4) was used to align reads to the reference sequence and to identify positions mutated during MaP[6]. The --output-parsed option was used to generate mutational files that serve as input to *RingMapper* and *PairMapper* correlation analysis software packages[3]. *ShapeMapper* outputs raw mutation rates (as shown in Figure 2A). Scaled equilibrium and per-timepoint reactivities for a given RNA profile were obtained by dividing by the mean reactivity of the top 10% of reactivities, after excluding reactivities above the 95th percentile (to remove outliers).

Rate determinations. Per-nucleotide reactivities were processed using a custom Python program. First, a time-series matrix of per-nucleotide reactivities was created from individual reactivity profiles, as output by *ShapeMapper*. Reactivities were normalized for each nucleotide position by computing reactivity ratios relative to the initial reactivity R_1 and the final reactivity R_2 across the time series:

$$N = \ abs\left(\frac{(R_x - R_2)}{(R_1 - R_2)}\right)$$

Normalized nucleotide reactivities were input to the *scipy* curve fitting algorithm as a function of time. Nucleotide reactivities were fit to a single exponential:

$$I = A + \ (1 - A)e^{-k_1 t}$$

where I is normalized reactivity, A is a reactivity amplitude, and k_1 is the rate constant. Nucleotides with curve fit correlations greater than 0.9 were analyzed further. Fitted kinetic curve plots were visually inspected for goodness of fit (JMP Pro 14). Roughly 40% of nucleotides in the RNase P RNA show significant time-dependent signals, with the remaining nucleotides showing no change over time.

Correlation analysis. *RingMapper* and *PairMapper* were used to compute correlated modification events[3]. Most of the analyses presented in this work used *RingMapper*; *PairMapper* was used to generate restraints used for secondary structure modeling (see Figure 2C). *RingMapper* and *PairMapper* were run with default settings, which invoke multiple quality control filters. Parameters critical to obtaining high quality data include the default *RingMapper* and *PairMapper* filters that ignore nucleotides with high background mutation rates, ignore

nucleotide pairs with high background correlations, and set cutoffs for minimum pairwise read depth. *RingMapper* outputs a table containing correlation position, read depth, and correlation significance. Negative correlations were removed. Those with positive correlations were required to have average product corrected (APC) G statistic strength above 100; if this criterion was not met, the events were removed. The average product corrected G statistic is a metric of correlation significance normalized relative to background correlation significance. Correlations with G statistic > 100 were plotted on secondary and tertiary structure models for the RNase P RNA. Correlation heatmaps were generated using a custom Python package; individual correlations were color coded according to their APC G statistic strength, with a minimum threshold of 100. *RingMapper* and *PairMapper* software packages are available for download at https://github.com/Weeks-UNC and from weekslab.com.

Secondary structure modeling. The PAIR-MaP framework[3] as implemented in *ShapeKnots* (from software package *RNAstructure*, v6.1)[7], was used to model the RNase P and TPP riboswitch RNA secondary structures (as shown in Figure 2D and Supporting Figure 2D). Required parameters were the RNase P catalytic domain[1] or TPP riboswitch[8] primary sequence and the output .ct file name. Pseudo-free energy change restraints were generated by calling the flag *-dmsnt* to use TMO- or DMS-normalized reactivities and the flag *-x* to incorporate *PairMapper* .bp files, thus introducing two TMO- or DMS-based bonuses (per-nucleotide and pairwise correlation) into the secondary structure calculation[3]. The *-m* 1 option was used to select only the minimum free energy structure, outputted as a .ct file. The unique features, implementation, and limitations of the PAIR-MaP framework have been outlined in prior work[3]. PAIR-MaP filters for correlations over three-nucleotide windows that are also compatible with canonical RNA base pairing, and adds both these pairing and per-nucleotide energy bonuses for RNA structure modeling. The key feature relevant to this work is that PAIR-MaP has been extensively validated for DMS, emphasizing that DMS-induced perturbations do not significantly disrupt RNA structure or negatively affect the ability of internucleotide correlations to allow accurate RNA structure modeling. In this work, by applying PAIR-MaP to two RNAs based on TMO reagent reactivities, we validate that, likewise, TMO does not induce large-scale disruption to RNA structure under the conditions employed here.

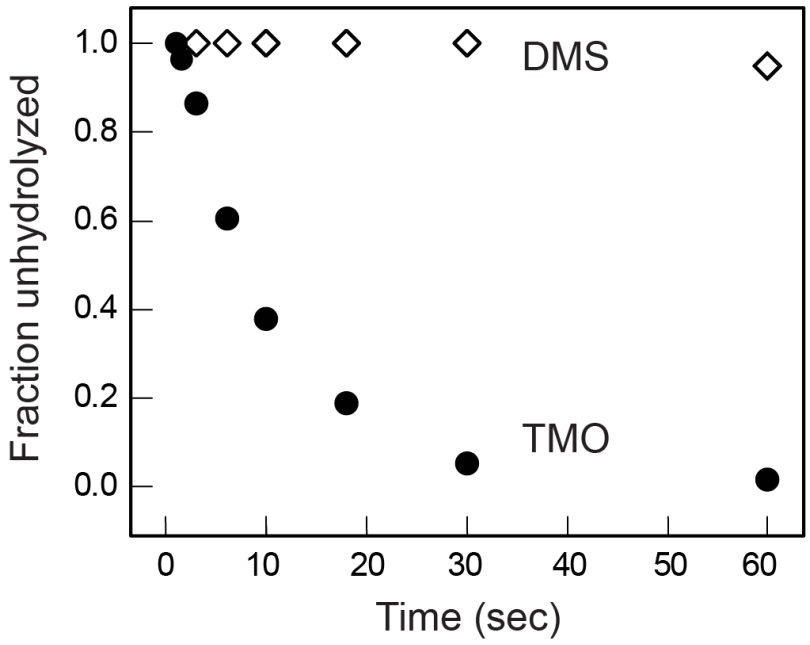
Table S1. Template sequences and primers used in this study

RNA	Sequence 5' → 3'
RNase P catalytic domain, native sequence	T7 transcription template: TAATACGACTCACTATAGGGCCTTCGGGCCAAGTTAATCATGCTCGG GTAATCGCTGCGGCCGGTTTCGGCCGTAGAGGAAAGTCCATGCTCG CACGGTGCTGAGATGCCCGTAGTGTTCGTGGAAACACGAGCGAGAA ACCCAAATGATGGTAGGGGCACCTTCCCGAAGGAAATGAACGGAGG GAAGGACAGGCGGCGCATGCAGCCTGTAGATAGATGATTACCGCCG GAGTACGAGGCGCAAAGCCGCTTGCAGTACGAAGGTACAGAACATG GCTTATAGAGCATGATTAACGTCTCGATCCGGTTCGCCGGATCCAAA TCGGGCTTCGGTCCGGTTC Forward PCR template primer: TAATACGACTCACTATAGGGCCTTCGGG Reverse PCR template primer: GAACCGGACCGAAGCCCG Reverse Transcription primer: *Same as Step 1 reverse primer Step 1 forward primer: CCCTACACGACGCTCTTCCGATCTNNNNNGGCCTTCGGGCCAAGGA Step 1 reverse primer: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNTTGAACCGG ACCGAAGCCCGATTT
RNase P catalytic domain, loop mutant	T7 transcription template: TAATACGACTCACTATAGGGCCTTCGGGCCAAGTTAATCATGCTCGG GTAATCGCTGCGGCCGGTTTCGGCCGTAGAGGAAAGTCCATGCTCG CACGGTGCTGAGATGCCCGTAGTGTTCGTGGAAACACGAGCGAGAA ACCCAAATGATGGTAGGGGCACCTTCCCGAAGGAAATGAACGGAGG GAAGGACAGGCGGCGCATGCAGCCTGTAGATAGATGATTACCGCCG GAGTACGAGGCGCAAAGCCGCTTGCAGTACGAAGGTACAGAACATG GCTTATAGAGCATGATTAACGTCTCGATCCGGTTCGCCGGATCCAAA TCGGGCTTCGGTCCGGTTC *Primers for the mutant RNase P are identical to native sequence primers.

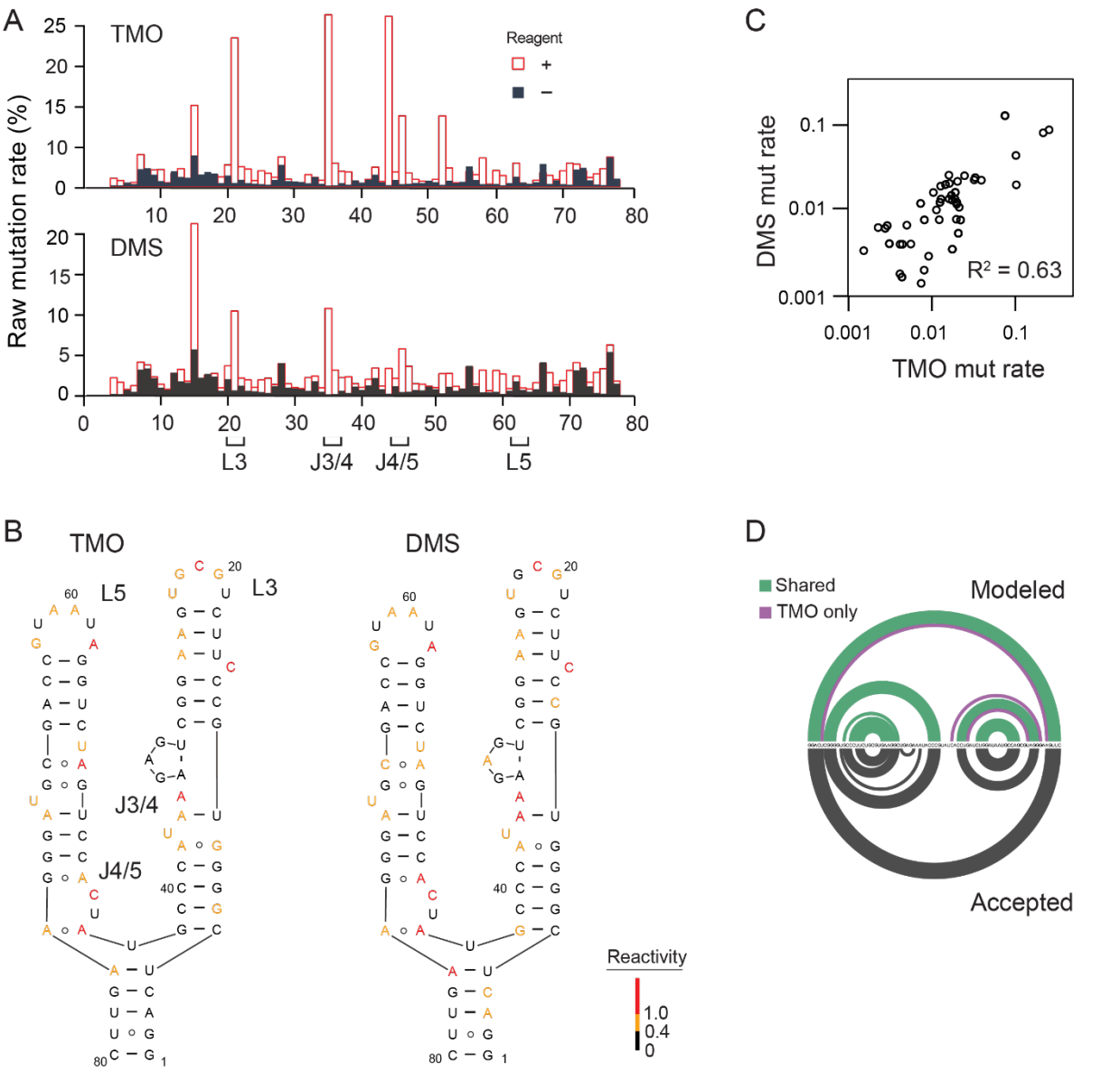
Table S2. Summary of sequencing data

Equilibrium structure probing, RNase P catalytic domain				
Sample identifier	Median read depth	5th percentile depth	Median modification rate	95th percentile modification rate
TMO Replicate 1	144,326	125,326	0.13	12.9
TMO Replicate 2	127,920	118,222	0.12	8.56
DMS Replicate 1	138,777	124,847	0.11	9.85
DMS Replicate 2	153,237	134,866	0.11	10.0
Control Replicate 1	154,898	143,281	0.07	1.28
Control Replicate 2	118,699	110,011	0.07	1.52
Time-resolved TMO probing, RNase P catalytic domain, replicate 1				
0 (Mg free)	376,889	350,172	0.19	6.60
5 sec	380,023	356,188	0.18	5.60
10	232,054	216,263	0.21	6.05
20	246,815	228,270	0.20	7.02
30	317,991	297,950	0.18	5.59
60	370,708	344,228	0.18	5.34
120	301,150	273,655	0.17	4.51
180	325,288	300,064	0.18	4.84
300	389,033	354,612	0.18	5.01
600	398,041	329,334	0.24	5.07
900	419,601	376,521	0.16	4.74
1200	325,303	287,225	0.18	5.02
Control	218,889	205,089	0.060	0.54
Time-resolved TMO probing, RNase P catalytic domain, replicate 2				
0 (Mg free)	724,359	612,245	0.14	12.1
5 sec	899,092	809,810	0.16	9.92
10	766,636	689,774	0.15	9.31
20	899,898	798,606	0.17	9.36
30	803,060	719,750	0.15	8.43
60	915,192	783,573	0.16	10.7
120	861,356	720,105	0.17	12.3
180	870,895	715,984	0.17	13.9
300	687,243	607,859	0.16	8.47
600	894,511	689,233	0.14	10.9
900	704,812	594,194	0.15	9.78
1200	769,678	599,215	0.12	8.03
Control	685,880	642,472	0.05	0.600

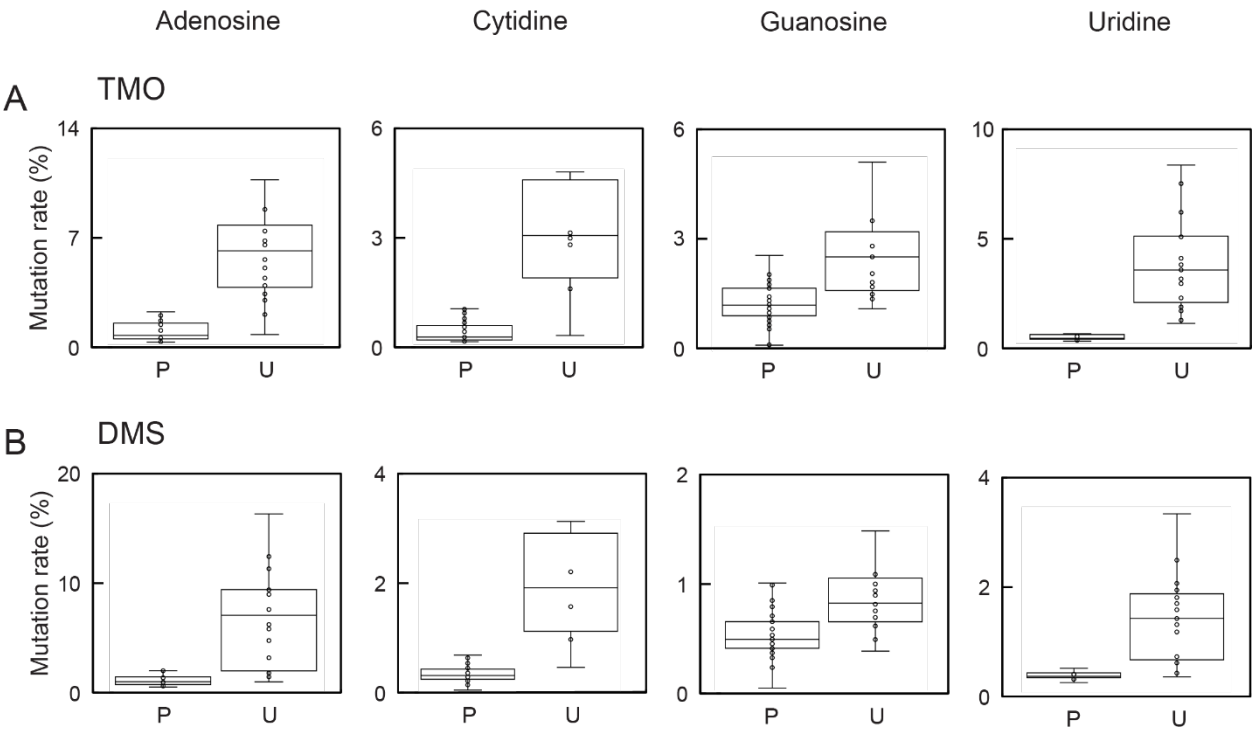
Time-resolved TMO probing, L15.1 RNase P mutant, replicate 1				
Sample identifier	Median read depth	5th percentile depth	Median modification rate	95th percentile modification rate
0	1,253,508	1,111,064	0.14	9.08
10	584,354	518,206	0.12	8.27
20	617,478	549,769	0.13	8.49
30	569,308	510,243	0.13	7.95
60	727,545	642,086	0.15	9.53
120	747,465	668,322	0.14	8.03
180	751,002	665,129	0.13	8.76
300	439,057	391,093	0.14	8.75
600	626,402	561,912	0.14	8.05
900	628,305	559,746	0.15	9.29
1200	193,228	176,717	0.14	9.6
Control	500,592	454,031	0.050	0.460
Time-resolved TMO probing, L15.1 RNase P, mutant replicate 2				
0	163,220	142,007	0.25	11.9
10	169,893	155,042	0.24	8.39
20	207,030	189,214	0.26	9.11
30	212,520	193,985	0.25	9.15
60	225,531	202,844	0.26	9.80
120	225,629	202,940	0.21	9.13
180	222,090	197,348	0.24	10.1
300	179,789	165,451	0.27	9.04
600	211,582	194,441	0.27	8.85
900	172,981	158,701	0.28	9.15
1200	211,710	194,732	0.26	9.01
Control	233,348	223,607	0.060	0.590



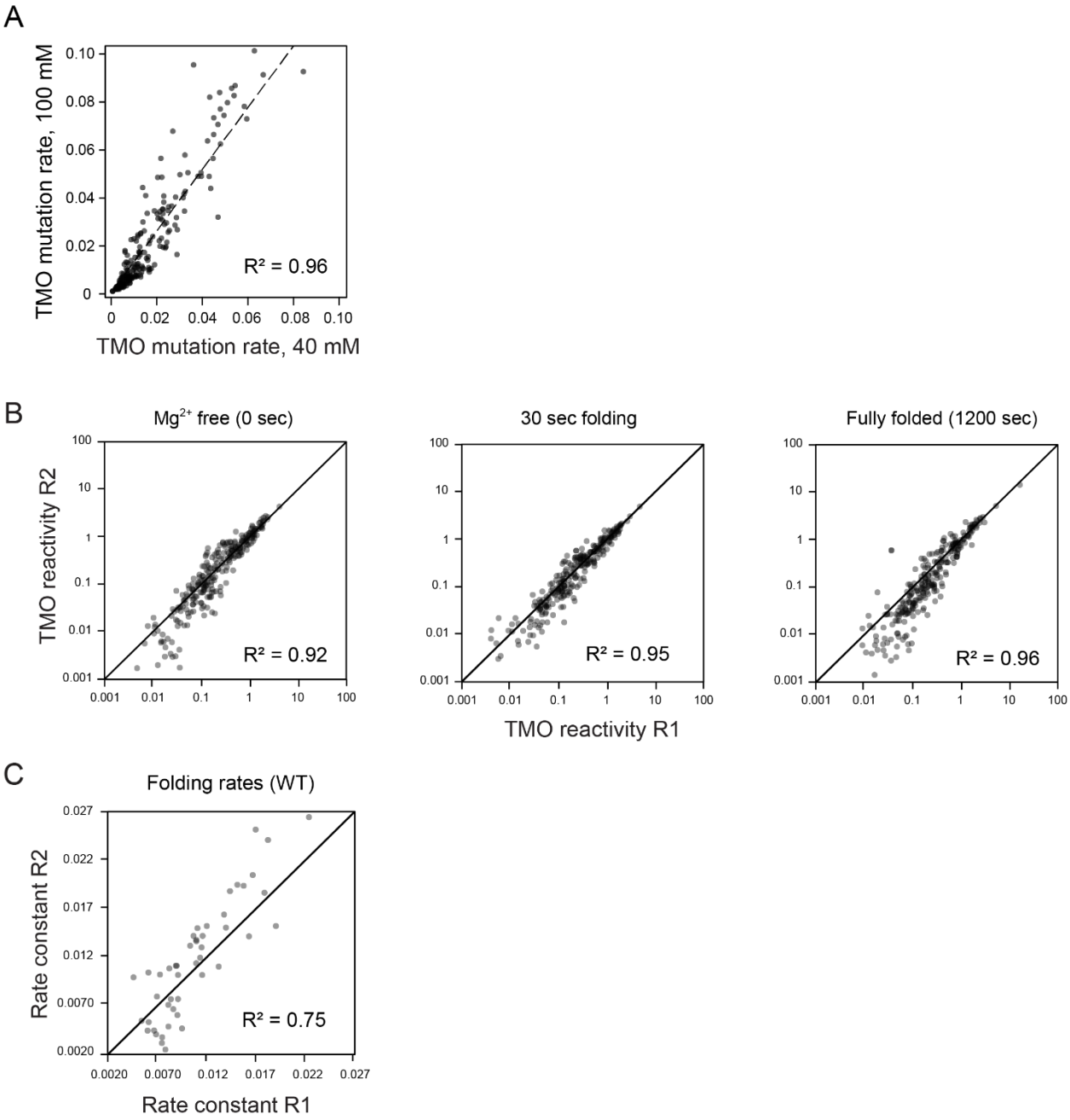
Supporting Figure 1. Hydrolysis time courses for TMO and DMS at 37 °C. Half-lives are 7.5 and 690 sec, respectively. Reagent hydrolysis was monitored by pH.



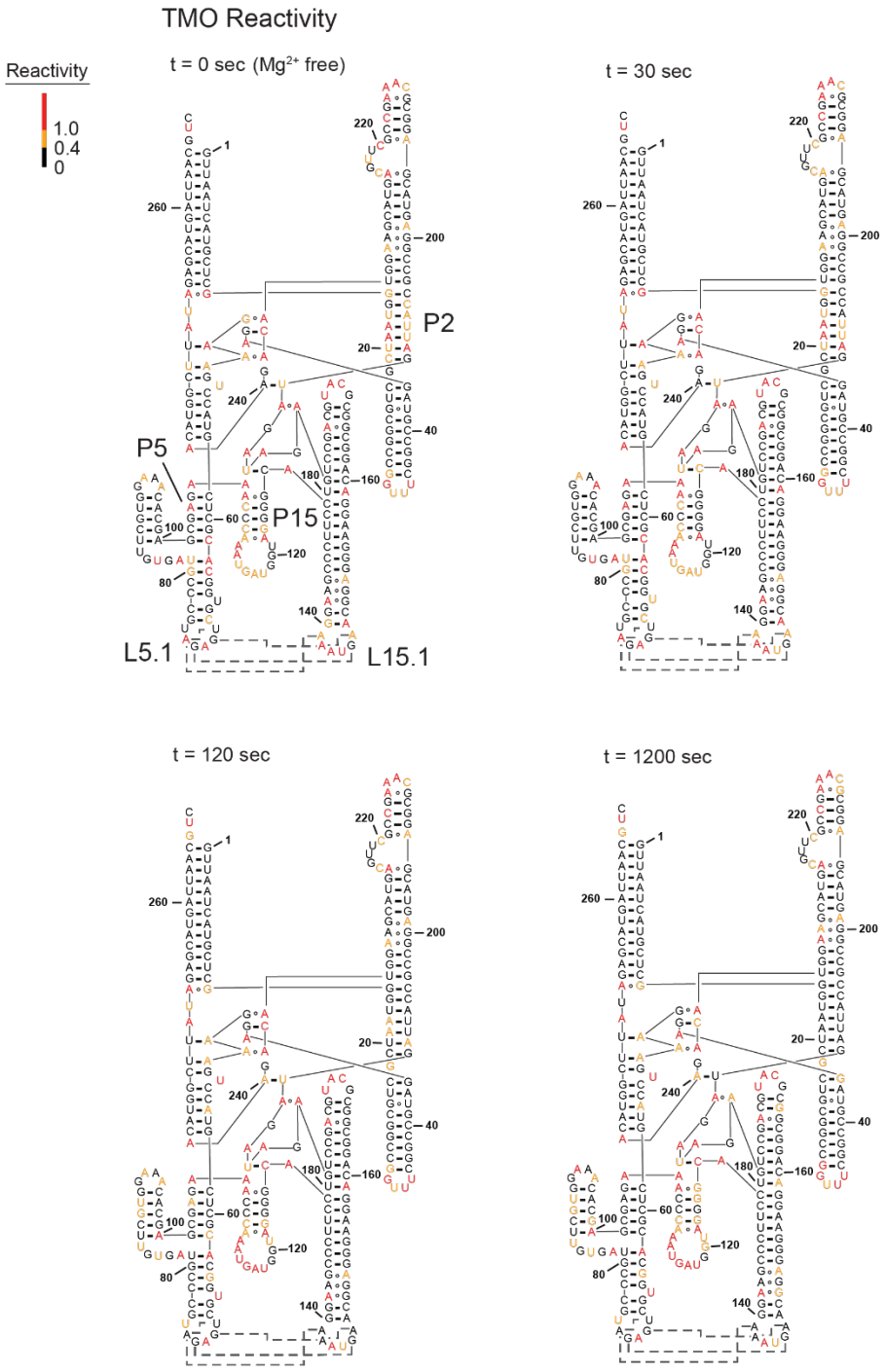
Supporting Figure 2. Comparison of TMO and DMS reactivities for the *E. coli* thiM riboswitch[8] (A) Raw mutation rates observed for TMO and DMS modification of the RNA as a function of nucleotide position. Red and grey bars indicate treated and no-reagent samples respectively. Loop and junction (L, J) landmarks are emphasized. (B) Chemical reactivities superimposed on the secondary structure. Red, yellow, and black nucleotides represent high, medium, and low normalized reactivities, respectively (see legend). Dashes and circles connecting nucleotides indicate Watson-Crick and non-canonical base pairing, respectively. (C) Reactivity correlation for TMO versus DMS; each point indicates an individual nucleotide. (D) Chemical probing-directed secondary structure models for the TPP riboswitch RNA based on TMO and DMS reactivities. Structures were modeled using the PAIR-MaP framework[3]. Arcs indicate base pairs. The accepted structure is gray. Green arcs represent base pairs predicted using both TMO and DMS reactivity data; purple arcs indicate the two additional base pairs returned by TMO-informed probing.



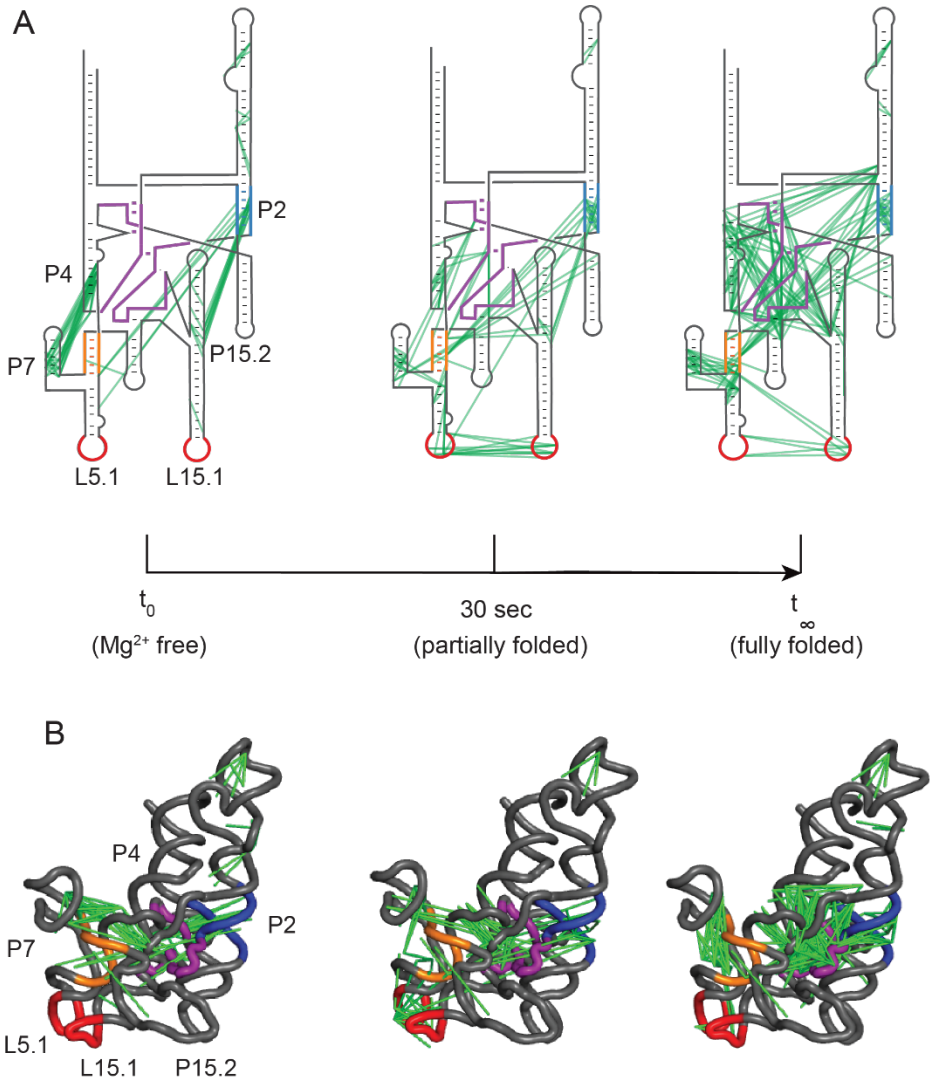
Supporting Figure 3. Distribution of mutation rates at paired (**P**) and unpaired (**U**) nucleotides in the RNase P RNA. Experiments performed with (A) TMO and (B) DMS. Reagent concentrations were 100 and 170 mM, respectively. Noncanonical nucleotides were excluded. Box plots enclose the central 50% of the data; whiskers extend to 1.5 times the interquartile range; medians are shown with the centered horizontal line.



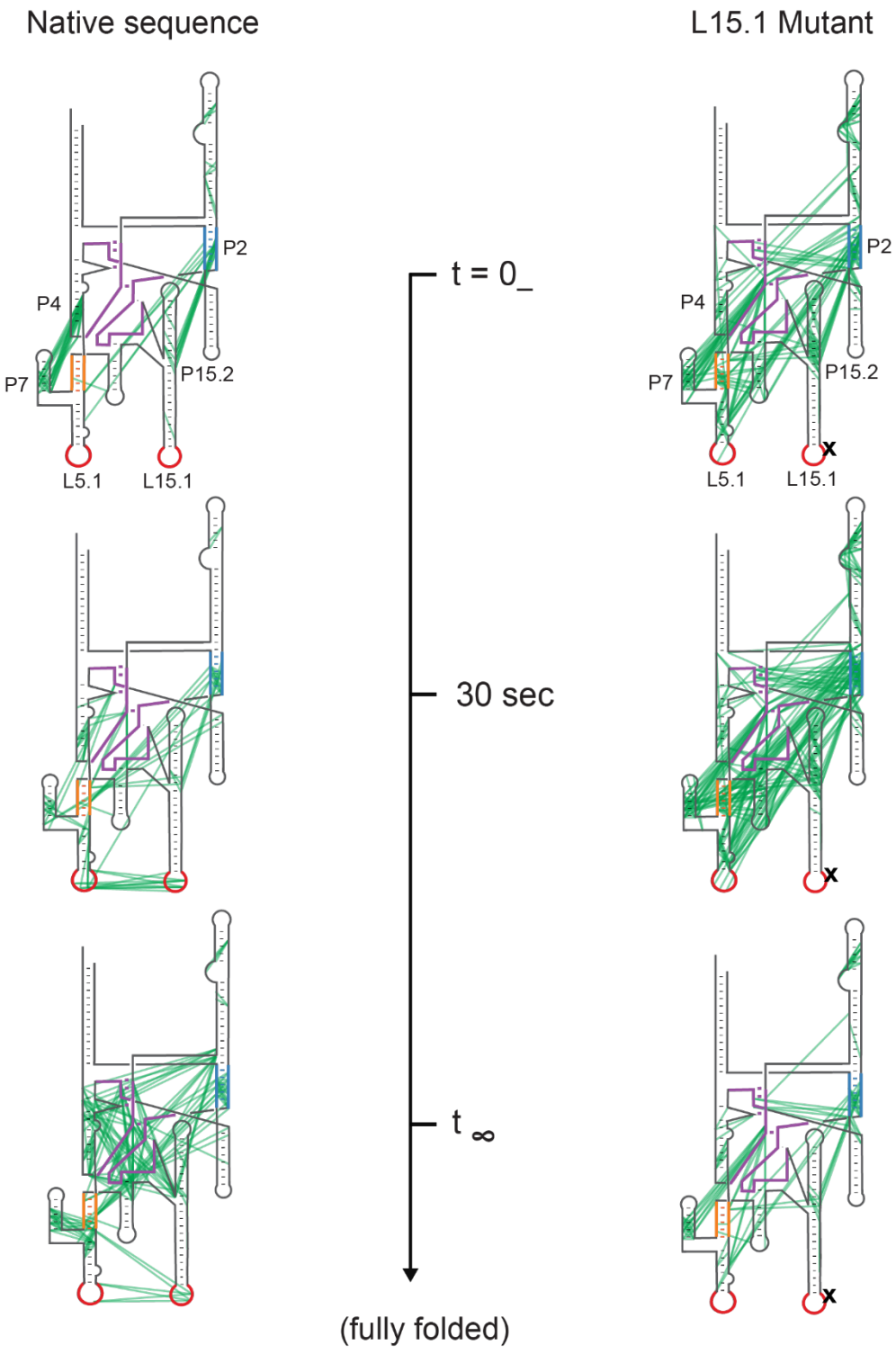
Supporting Figure 4. Correlations between TMO reactivities (A) as a function of reagent concentration, and (B, C) for full experimental replicates, performed on the RNase P RNA. (A) Experiments performed at 40 and 100 mM TMO reagent. Points indicate individual nucleotides; slope is 1.34. Correlation shows that differing levels of TMO yield similar relative modifications across the RNA, indicating that high-level modification does not induce global destabilization of RNA structure. (B) TMO reactivity correlations for replicate experiments; each point indicates an individual nucleotide. (C) Correlations in reaction rate measurements; each point indicates a nucleotide fit to a first-order rate process, with corresponding rate constants (sec^{-1}) on the x and y axis for each replicate. R1 and R2, replicate datasets.



Supporting Figure 5. Single-nucleotide TMO reactivity data as a function of position on the secondary structure of the RNase P RNA. TMO reactivities at time points 0 (Mg^{2+} free), 30, 120, and 1200 (fully folded) sec are shown on the final, equilibrium structure of the RNase P catalytic domain. Red, yellow, and black nucleotides represent high, medium, and low TMO reactivities, respectively (see legend). Dashes and circles connecting nucleotides indicate Watson-Crick and non-canonical base pairing, respectively. Tertiary interactions between the L5.1 and L15.1 loops are shown with gray dashed lines.



Supporting Figure 6. Time-dependent, RING single molecule, correlated chemical probing of the native-sequence RNase P RNA. (A) Visualization of through-space correlations (green), superimposed on a secondary structure model. Structural landmarks are shown with colored backbone: pseudoknot helices P2 and P5, blue and orange; catalytic core, purple; L5.1-L15.1, red. (B) Visualization of correlations from panel A superimposed on the 3-dimensional structure[1] of the fully folded RNase P RNA. Images at 0 and 30 sec serve to illustrate changes in through-space correlations; the solution RNA structure will be different.



Supporting Figure 7. Comparison of time-dependent, single molecule, correlated chemical probing of native-sequence and loop mutant RNase P RNAs. Mutant RNA is described in Figure 4; position of 2-nucleotide mutation is shown with **x** (at right) Correlations (green lines) are superimposed on the RNase P secondary structure. Domains with notable folding behavior are emphasized in color: pseudoknot helices P2 and P5, blue and orange; catalytic core, purple; L5.1-L15.1, red. Data in left-hand column are reproduced from Supporting Figure 6.

Supporting References

(1) Kazantsev, A. V; Krivenko, A. A.; Pace, N. R. Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **2009**, *15*, 266–276.

(2) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231.

(3) Mustoe, A. M.; Lama, N. N.; Irving, P. S.; Olson, S. W.; Weeks, K. M. RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. U. S. A*. **2019**, *116*, 24574–24582.

(4) Homan, P. J.; Favorov, O. V; Lavender, C. A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N. V; Weeks, K. M. Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13858–13863.

(5) Siegfried, N. A.; Busan, S.; Rice, G. M.; Nelson, J. A. E.; Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11*, 959–965.

(6) Busan, S.; Weeks, K. M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **2018**, *24*, 143–148.

(7) Reuter, J. S.; Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **2010**, *11*, 129.

(8) Serganov, A.; Polonskaia, A.; Phan, A. T.; Breaker, R. R.; Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **2006**, *441,* 1167-1171.