Supporting Information for “Standardization of RNA chemical mapping experiments”

Wipapat Kladwang¹, Thomas H. Mann¹, Alex Becka¹, Siqi Tian¹, Hanjoo Kim², Sungroh Yoon², and Rhiju Das¹,³,†

¹Department of Biochemistry, Stanford University, Stanford CA 94305
²Department of Electrical and Computer Engineering, Seoul National University, Seoul 151-744, Korea
³Department of Physics, Stanford University, Stanford CA 94305

This Supporting Information document contains the following sections:

Supporting derivation
Experimental methods
Supporting References
Supporting Table 1. Nucleic acid sequences and database IDs.
Supporting Table 2. Hotspot nucleotides.
Supporting Figure 1. Example of data averaging and error estimation.
Supporting Figure 2. Proposed standardization brings data taken with varying chemical modifier concentrations into concordance.
Supporting Figure 3. CE and MAP-seq data for unmodified tRNA(phe), S. cerevesiae.
Supporting Figure 4. CE and MAP-seq data for ligand-binding domain of adenine riboswitch, V. vulnificus.
Supporting Figure 5. CE and MAP-seq data for ligand-binding domain of cyclic-di-GMP riboswitch, V. cholerae
Supporting Figure 6. CE and MAP-seq data for 5S ribosomal RNA, E. coli.
Supporting Figure 7. CE and MAP-seq data for P4-P6 domain of the Tetrahymena ribozyme.
Supporting Figure 8. CE and MAP-seq data for ligand-binding domains of glycine riboswitch, F. nucleatum.
SUPPORTING DERIVATION

Relation between observed and actual product fractions

Determining chemical reactivity profiles for nucleic acids requires taking into account how a chemical modification internal to a fragment can lower the probability of observing longer fragments. For completeness, we derive the relation between observed and actual product fractions [main text eq. (1)] here. In the case of reverse transcription, let the first reverse transcribed nucleotide be \( N \) (i.e., the maximum product length), the total number of products be \( M \), and the fraction of chemical modified nucleotides at each position \( i = 1, \ldots N \) be \( r_i \). Then, the number of full length products \( F_0 \), corresponding to events with no modification at any internal site is:

\[
F_0 = (1-r_1)(1-r_2)\cdots(1-r_N)M,
\]  
(S1)

The number of products \( F_i \) corresponding to modification at each nucleotide \( i \) are:

\[
F_i = r_i (1-r_{i+1})(1-r_{i+2})\cdots(1-r_N)M
\]  
(S2)

See also refs.\(^{1,2}\). Partial summation of these values from the 5\(^{\prime}\) end gives:

\[
F_0 + F_1 = (1-r_2)(1-r_3)\cdots(1-r_N)M \\
F_0 + F_1 + F_2 = (1-r_3)\cdots(1-r_N)M
\]  
(S3)

Combining eqs. (S2) and (S3) gives the sought relation of \( r_i \) to the observed \( F_i \):

\[
r_i = \frac{F_i}{F_0 + F_1 + \ldots F_i},
\]

The equation corresponds to the fraction of reverse transcriptases that stopped at position \( i \), compared the total number of reverse transcriptases that reached position \( i \) (and either stopped or proceeded beyond). This derivation also holds for protocols that involve chemical or enzymatic cleavage of end-labeled nucleic acids instead of primer extension, with \( N \) marking the position of the 3\(^{\prime}\)-end label. For 5\(^{\prime}\)-end labels, the same relations hold but with indices \( i \) reversed in order; in all cases \( F_0 \) should correspond to unmodified nucleic acid. The above values \( r_i \) range from 0 to 1. The number of modification events \( R_i \) is a better estimate of chemical reactivity which scales linearly with modifier.
concentration; it can be estimated from the relation \( r_i = 1 - \exp(-R_i) \). However, \( R_i = r_i \) for \( r_i \ll 1 \), as was the case herein. Last, background subtraction and scaling based on internal standards (see main text) gives \( r_{i,\text{norm}} \), which can exceed 1.

**EXPERIMENTAL METHODS**

**Data and software availability**

All analysis steps have been implemented in two freely available software packages HiTRACE (for capillary electrophoresis analysis; http://www.hitrace.org for HiTRACE-Web server or MATLAB software download) and MAPseeker (for MAP-seq deep sequencing analysis; https://github.com/MAPseeker for software download). See below for description of processing steps and implementations. Data have been deposited in the RNA Mapping Database\(^3\) (http://rmdb.stanford.edu) under accession codes listed in SI Table 1.

**Preparation of RNA**

RNA preparation procedures followed those in experiments described previously\(^4,5\), with small modifications noted here. Briefly, DNA templates were produced through PCR assembly of oligonucleotides of length 60 nucleotides or smaller (Integrated DNA Technologies) using Phusion polymerase (Finnzymes, MA). DNA templates were designed with the T7 RNA polymerase promoter (TTCTAATACGACTCACTATA) at their 5’ ends. A custom reverse transcription primer-binding site (AAAGAAACAAACAACAAAC) was included at the 3’ terminus of each template. See Table 1. Flanking sequences, here including referencing hairpin stems, were screened computationally to not interact with the target RNA sequence on the NUPACK server.\(^6\) (Data were consistent within error and scaling with prior measurements with different flanking sequences.\(^2,4,7\) ) RNA transcribed with T7 RNA polymerase (New England Biolabs) was purified using the RNA Clean & Concentrator 5 kit (Zymo Research).

**Chemical mapping experiments read out by capillary electrophoresis**
Chemical mapping procedures with capillary electrophoresis (CE) followed those in these experiments described previously.\textsuperscript{4,5} Briefly, modification reactions were performed in 20 µL reactions containing 1.2 pmol RNA, 50 mM Na-HEPES (pH 8.0), and 10 mM MgCl\textsubscript{2}. Ligand-binding RNAs were incubated with specified ligands at room temperature for 30 minutes prior to mapping. Chemical probes were used at the following final concentrations: DMS (0.125% v/v for P4P6-2HP, varied from 0.03125% to 0.5% where noted; added with 0.25% ethanol), CMCT in water (2.6 mg/mL standard; 0.66–10.5 mg/mL where noted), 1M7 (1.05 mg/mL standard; 0.2625–4.2 mg/mL where noted; stock prepared in anhydrous DMSO gave final DMSO concentration of 25%). Chemical probes were allowed to react for 15 minutes prior to quenching. The reaction quench for 1M7 and CMCT contained 5.0 µL of 0.5 M Na-MES (pH 6.0), 3 µL of 3 M NaCl, 1.5 µL of oligo-dT beads (poly(A) purist, Ambion), and 0.25 µL of a 0.25 µM 5´-rhodamine-green labeled primer (Table 1), complementary to the reverse transcription primer-binding site at the RNA 3´ ends. The reaction quench for DMS was identical except that the Na-MES component was replaced with 5 µL of 2-mercaptoethanol. This quench mixture allowed for purification and reverse transcription on magnetic beads. Chemically modified RNAs were reverse transcribed with Superscript III Reverse Transcriptase (Life Technologies). RNA was subsequently hydrolyzed for 3 minutes at 90 °C in 0.2 M NaOH. After pH neutralization and ethanol rinsing, cDNAs were eluted into 10 µL HiDi formamide (Life Technologies) and co-loaded with a ROX-350 standard ladder (Life Technologies) for electrophoresis on ABI 3130 or 3700 sequencers.

CE data were quantitated with HiTRACE\textsuperscript{8} to give observed product frequencies \( F_{\text{observed}} \). Fitting errors were estimated with the function \texttt{fit_to_gaussians} based on analytical computation of the sum of the squares of standard deviations of \( F_{\text{observed}} \) upon shifting all band positions by \( \pm 0.5 \) of the mean band-to-band spacing.

Reactions without chemical modification gave estimates of backgrounds, \( F_{\text{background}} \). We note that the units of these measurements are arbitrary; final reactivity fractions \( r_i \) are calculated based on ratios of these observed products to sums of products measured within the same trace (see Supporting Derivation and next section).
Quantitative analysis of nucleotide reactivities (capillary electrophoresis)

Four data processing steps were carried out for analysis of chemical mapping experiments (see also main text Fig. 1a) from capillary electrophoresis experiments. The entire pipeline is available through an online workflow on the HiTRACE-Web server as well as from the MATLAB implementation via a single HiTRACE script `get_reactivities`.

1. Saturation correction from dilution samples

Applying eq. (1) of the main text (see also SI Appendix) requires accurate quantitation of observed products $F_i$. In most electrophoresis experiments, the full-length band ($F_0$) and occasionally other strong bands saturate the detector. In addition to our fully concentrated samples, we acquired capillary electrophoretic traces for 10-fold dilutions of each sample, by removing 1 µL of the reactions prepared for ABI sequencers into running buffer of 9 µL of HiDi formamide with the ROX-350 standard. For some experiments, it may be necessary to dilute further than 10-fold, although carrying out this additional dilution series was not necessary for the the linearity range of our CE detector and maximum band intensity observed in our experiments.

A scalefactor $\alpha$ was determined to match the resulting $F_{\text{diluted}}$ to the original $F_{\text{observed}}$.

$$\alpha = \frac{\sum_{i=\text{non-saturated}} F_{\text{observed}}}{\sum_{i=\text{non-saturated}} F_{\text{diluted}}}$$

where “non saturated” refers to the subset of nucleotides at which the diluted data matched the undiluted data. This subset was determined by beginning with the full set of measured nucleotides, determining $\alpha$, computing residuals $\delta_i = F_{\text{diluted}} - F_{\text{observed}}$, filtering out any nucleotides $i$ at which $\delta_i$ exceeded the mean of $\delta_i$ by more than 1.5 times the standard deviation of $\delta_i$, and iterating the procedure a total of three times. For each nucleotide $i$ that was filtered out as ‘saturated’ by this procedure, $F_{\text{observed}}$ and its estimated error was replaced with $\alpha F_{\text{diluted}}$ and its error (scaled by $\alpha$). The same procedure was used to correct for saturation in control measurements without chemical...
modification $F_i^{\text{background}}$. The procedure, automated in the script `unsaturate` in HiTRACE, returns an image showing saturated residues.

2. Over-modification correction
Main text equation 1 (see also Supporting Derivation) was applied to transform the (saturation corrected) observed product values $F_i^{\text{observed}}$ and $F_i^{\text{background}}$ to give modification fractions $r_i^{\text{observed}}$ and $r_i^{\text{background}}$. Relative errors on $r_i^{\text{observed}}$ were taken from errors on $F_i^{\text{observed}}$. [Additional relative errors due to summation across the denominator $F_0 + F_1 + \ldots F_i$ would give rise to correlated errors across the entire profile and were not modeled here due to their complexity and to their generally small contributions.] This procedure was automated in the script `correct_for_attenuation`.

3. Background subtraction
The modification fraction $r_i$ at each nucleotide $i$ due to chemical modification was given by $r_i^{\text{observed}} - r_i^{\text{background}}$, with error estimated by summing errors of components in quadrature. This procedure was automated in the script `subtract_array`.

4. Normalization to referencing segments
Inclusion of at least one referencing hairpin with pentaloop sequence GAGUA enabled normalization, giving final values $r_i^{\text{norm}}$ that were independent of the chosen modifier concentration and time. Modification fractions $r_i$ were scaled so that the underlined nucleotides gave mean reactivities of 1.0: $\text{GAGUA}$ (DMS), $\text{GAGUA}$ (CMCT), and $\text{GAGUA}$ (SHAPE). In constructs with two hairpins in both 5’- and 3’- flanking sequences, we used data for the 3’- GAGUA hairpin for normalization, due to high errors in background subtraction for GAGUA hairpins in 5’ flanking sequences. This procedure was automated in the script `apply_normalization`.  


All steps above are included in HiTRACE software, and a step-by-step workflow is available as a default stage in the online HiTRACE-web server.\textsuperscript{9}

**Chemical mapping experiments read out by deep sequencing (MAP-seq)**

The detailed MAP-seq protocol has been presented in ref.\textsuperscript{10} and is briefly summarized here. Chemical modification reactions were carried out as in CE reactions in 10 mM MgCl\textsubscript{2} and 50 mM Na-HEPES, pH 8.0, but with 4-8 pmols of each RNA in 50 µL volumes. In one set of experiments, the six RNAs, along with other RNAs (including ligand-binding domains for an adenosyl-cobalamin and flavin mononucleotide riboswitch; data not shown), were subjected to modification at different concentrations of DMS (0.125% and 0.5% v/v final), 1M7 (1.05 and 4.24 mg/mL final), or no reagent, either without added ligand or with a ligand mixture (final concentrations of 5 mM adenine, 10 µM cyclic-diguanosine monophosphate, 10 mM glycine, 200 µM flavin mononucleotide, and adenosyl cobalamin, 70 µM) in 12 samples. In a second set of experiments, each of the six RNAs was modified by DMS (0.125% v/v final), 1M7 (1.05 mg/mL final), or no reagent, with ligand (5 mM adenine, 10 µM cGMP, or 10 mM glycine for the relevant riboswitches; a mixture of all three for the other RNAs) or no ligand, in 36 separate samples. In each case, modified RNA in 50 µL volumes was precipitated by addition of 10 µL 3 M Na-acetate and 330 µL cold ethanol; microcentrifugation; washed with 70 % ethanol; and resuspended in 9.7 µL deionized water. Reverse transcription reactions were carried out as in CE experiments except scaled up to 15 µL volumes and using 5´-FAM-labeled Illumina-Oligo C-containing primers with 12 nt barcodes; pulldowns were carried out with DynaBeads (Life Technologies) displaying the reverse complement to Oligo C (Oligo C´, TGTGTAGATCTCGGTGGTCGCCGTATCATTCTTTTTTTTTTTTTT3´-double-biotin) and cDNA was resuspended (with beads remaining) into 2 µL deionized water. Aliquots of these reactions were run by CE to confirm reverse transcription; for the second set of experiments, each sample involved a single RNA, and so these data could be analyzed by HiTRACE and gave data consistent with CE measurements above. To complete the MAP-seq protocol, sets of four samples were pooled into 8 µL volumes and cDNAs were ligated to the second Illumina adapter (1.25 µM 5´-phosphate-
AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGA
TTCTGCTTG-3’-phosphate) with CircLigase I (5 U µL) in 2.5 mM MnCl₂, 1x
CircLigase I buffer, 4% PEG 1500, and 50 µM ATP for 2 hours in 50 µL volumes at 68 °C, followed by 10 minutes at 80 °C for deactivation of the ligase. Samples were again pulled down by magnetic separation and washed, and ligated cDNA concentrations were quantitated by loading on ABI 3130 sequencer with FAM-labeled standards. Samples were eluted by addition of 4.5 µL 10 mM EDTA, 95% formamide, heating to 90 °C for 2 minutes, cooling to room temperature; and addition of 0.5 µL 5 nM PhiX control (Illumina). The sample was then prepared for loading onto Miseq v2 kits following manufacturer instructions.

Data processing was carried out with the quick_look_mapseeker routine in MAPseeker software, as described in ref.¹⁰. All MAPseeker steps requiring an internal control are automatically activated when the P4P6-2HP sequence with flanking GAGUA hairpins (SI Table 1) is included in the run and specified as one of the probed RNAs in RNA_sequences.fasta. In particular, the critical ligation bias correction term for the fully extended product (relative to the average sequence) was determined by applying an initial value of β = 1.0 to find the background-subtracted reactivities of the control P4P6-2HP sequence; testing for equality at GAGUA sequences at the 5´ and 3´ ends; and then optimizing β to give exact equality at the GAGUA sequences by numerical search with MATLAB’s fminbnd function. Over-modification correction, background subtraction, and normalization to GAGUA segments were carried out as described above for C.E.; these steps occur automatically in the MAPseeker workflow.

We observed that normalization of MAP-seq data at flanking GAGUA hairpins produced reactivities that were higher than CE measurements within the target of interest, presumably due to systematic ligation bias at those nucleotides relative to other sequences. For Fig. 1 and SI Figures 3-8, we therefore used the average MAP-seq reactivity value over nucleotides in the P4P6-2HP RNA, compared to their average value in CE data, to provide correction factors applied to all other MAP-seq data (0.55 and 0.40 for 1M7 and DMS, respectively). Note that for any future MAP-seq experiments
with the P4P6-2HP, the normalization scale will be automatically set by MAPseeker based on GAGUA hairpins, but will need to be re-scaled by the above factors if matching to CE experiments is required.

**Averaging and error estimations based on multiple replicates**

For CE experiments, we found that estimates of error due to peak fitting or from standard deviations within each experiment generally underestimated errors estimated from repeating experiments. For MAP-seeker experiments, estimates of error based on Poisson counting statistics also gave underestimates of error for MAP-seq runs with >100,000 counts for each RNA (this high-statistics limit was the case herein). For example, a previous averaging procedure, which estimated final reactivity errors based solely on propagating estimates from CE peak-fitting, did not capture the high tRNA(phe) DMS reactivity at A58 (SI Fig. 3).\(^2\)\(^,\)\(^7\) A user error assigned this strong band to an adjacent residue in some replicates; the deviation between the replicates should have been reflected in high errors at both positions, but the propagated CE errors, dominated by replicates with zero reactivity, showed low reactivities. The identification of this issue led to developments of an automated sequence assignment tool (available in HiTRACE and HiTRACE-web; SRY, HY, RD, in prep.), a visual display of averaging results, and a more conservative error estimation procedure.

The new averaging and error estimation procedure is encoded in the HiTRACE function `average_data_filter_outliers` (and a wrapper function for data formatted in the RDAT format\(^3\), `rdat_combine`). At each nucleotide position, the reactivity values from different measurements were averaged, weighted by the inverse of errors estimated from peak fitting (as is returned by HiTRACE CE fits) or Poisson error (as is returned by MAPseeker error). (To avoid extremely low error points from dominating this average, the error on each input measurements was set to 10% of the reactivity if it was estimated to be lower.) The standard error on this average value was taken as the standard deviation among measurements, divided by the square root of the number of observations. If any of the measurements gave a value more than five standard deviations from the original average, that entire measurement was automatically flagged as an outlier, and the average
was recalculated without this value. In rare cases, an entire replicate gave a mean discrepancy at all nucleotides more than 2.5 standard deviations from the replicate average; this measurement was also flagged as an outlier and not included in the final calculation. All measurements and averaging are displayed graphically in both heat-map and trace overlay formats to allow visual assessment of sequence assignments, automatically assigned outliers, and other variability across measurements. An example of this display of measurement averaging and error estimation is given in SI Fig. S1.

Data visualization

Data figures prepared in MATLAB (http://www.mathworks.com) and PyMol (http://www.pymol.org). 3D coloring scripts are freely available at https://github.com/DasLab/pymol_daslab.

SUPPORTING REFERENCES

1. Aviran, S., Trapnell, C., Lucks, J. B., Mortimer, S. A., Luo, S., Schroth, G. P., Doudna, J. A., Arkin, A. P., and Pachter, L. (2011), Proc Natl Acad Sci U S A 108, 11069-11074.
2. Kladwang, W., Vanlang, C. C., Cordero, P., and Das, R. (2011), Biochemistry 50, 8049-8056.
3. Cordero, P., Lucks, J. B., and Das, R. (2012), Bioinformatics 28, 3006-3008.
4. Kladwang, W., VanLang, C. C., Cordero, P., and Das, R. (2011), Nat Chem 3, 954-962.
5. Seetin, M. G., Kladwang, W., Bida, J. P., and Das, R. (2013) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol, In RNA Folding (Methods in Molecular Biology) (Waldsich, C., Ed.), p in press.
6. Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011), Journal of computational chemistry 32, 170-173.
7. Cordero, P., Kladwang, W., VanLang, C. C., and Das, R. (2012), Biochemistry 51, 7037-7039.
8. Yoon, S., Kim, J., Hum, J., Kim, H., Park, S., Kladwang, W., and Das, R. (2011), Bioinformatics 27, 1798-1805.
9. Kim, H., Cordero, P., Das, R., and Yoon, S. (2013), Nucleic Acids Res 41, W492-498.
10. Seetin, M. G., Kladwang, W., Bida, J. P., and Das, R. (2014), Methods Mol Biol 1086, 95-117.
SUPPORTING TABLE 1. Nucleic acid sequences and database IDs. Sequences written from 5’ to 3’. Structured RNA domain of interest highlighted in blue. Reference GAGUA hairpins highlighted in red; some sequences have two references to test overmodification correction. Protein DataBank IDs give crystallographic models used to assess 3D environments, and accession IDs of data collected herein and deposited in the RNA Mapping Database are also given.

| Molecule | Sequence (conventional numbering) | PDB | RMDB IDs |
|----------|-----------------------------------|-----|----------|
| tRNA-1HP, unmodified tRNA (phe), S. cerevisiae | GGACAACAAACAAAACAGCGAUUUGCCAGACUGGAGAU CAGUUGGAGACGGCCAAGACUGGAUA CUGGAGGUCUGUUGCGAUCCACAGA AUUCGCAACAAACGUAAAGGAUAAGACU UAACCAAAGAAAACAAACACACACAC (–15 to 117) | 1EHZ | TRNAPH_DMS_0005 TRNAPH_CMC_0005 TRNAPH_1M7_0005 TRNAPH_DMS_0006 TRNAPH_CMC_0006 TRNAPH_1M7_0006 |
| ADD-2HP, adenine riboswitch, V. vulnificus | GGAGAACAAUUCGAGUAGAUAUGGAAA GGGAAAAGAACGCUGUCAUAUAAUCCUA AUGCUGUUGUUGGGAGUUCUAACAA GAGCCUUUAACUCUUGIUUAUGGAGUG AAAACAAAGUAAAGGAUAACUUAACAC AACAAAAACACACACACAC (–25 to 129) | 1Y26 | ADDRSW_DMS_0005 ADDRSW_CMC_0005 ADDRSW_1M7_0005 ADDRSW_DMS_0006 ADDRSW_CMC_0006 ADDRSW_1M7_0006 |
| cidGMP-2HP, Cyclic di-GMP riboswitch, V. cholerae | GGAGAAAGCAAUUCGACAGGGAACCC AUUGGGAAGGAGGCUGACGAAAGCUUGG GAGCCGGAUACUGGACUGGAACU ACACAAAAGGAAAGGAUAACUUAACAC AACAAAAACACACACACACAC (0 to 151) | 3MXH | CDIGMP_DMS_0005 CDIGMP_CMC_0005 CDIGMP_1M7_0005 CDIGMP_DMS_0006 CDIGMP_CMC_0006 CDIGMP_1M7_0006 |
| 5S-2HP, 5S rRNA, E. coli | GGAAGACAAUUCGAGUAGAUAUGGAAA GGGAAAAGAACGCUGUCAUAUAAUCCUA AUGCUGUUGUUGGGAGUUCUAACAA GAGCCGUAAACCCAGAAAGACUGGGAACU ACUCAGAAAGUAAACCGCAUUGGGAACUAU ACACAAAAGGAAAGGAUAACUUAACAC AACAAAAACACACACACACAC (–37 to 166) | 3OFC | 5SRRNA_DMS_0005 5SRRNA_CMC_0005 5SRRNA_1M7_0005 5SRRNA_DMS_0006 5SRRNA_CMC_0006 5SRRNA_1M7_0006 |
| P4P6-2HP, P4-P6 domain of Tetrahymena ribozyme | GGCCCAAAGGCGUCGAGUAGACCGCAAC ACGGAAAUUGCGGAAAGGGGGUGGGAAC CGCGUUGUCCACCUUGAACCCUAUGCCGA ACUCAGAAAGUAAACCGCAUUGGGAACUAU ACACAAAAGGAAAGGAUAACUUAACAC AACAAAAACACACACACACACAC (71 to 309) | 1GID | TRP4P6_DMS_0005 TRP4P6_CMC_0005 TRP4P6_1M7_0005 TRP4P6_DMS_0006 TRP4P6_CMC_0006 TRP4P6_1M7_0006 |
| FN-2HP, double glycine riboswitch, F. nucleatum | GCAAAUCUGAGUAGAUAUGGAGCAGAG GAUAUGAAGAAGAUAUUAUUAAUG AAACACCAAAGGAAUAAACCUUUGAG UAAAAACACGUAUAAUUGGAGCAACCU CGAGGAGCUUAACUAGAGAUAACAC CAGAAGGAGCAACAGCUAAUUAUGGACCU AACCUGAGUUAAGGACGGAAGAAA CACACAGUGAACCAGAGAACCAGAACG AAAACACACACACACACACAC (–27 to 204) | 3P49 | FNLGLYC_DMS_0005 FNLGLYC_CMC_0005 FNLGLYC_1M7_0005 FNLGLYC_DMS_0006 FNLGLYC_CMC_0006 FNLGLYC_1M7_0006 |

RMDB IDs ending in 0005 store CE data; those ending in 0006 store MAP-seq data.
**SUPPORTING TABLE 2. Hotspot nucleotides.** Nucleotides within structured RNAs that gave DMS, CMCT, or 1M7 reactivity above 1.5 (average from CE measurements).

| Molecule                  | Nucleotide | Base reactivity<sup>a</sup> | 2'-OH reactivity<sup>b</sup> | Feature               |
|---------------------------|------------|-----------------------------|-----------------------------|-----------------------|
| tRNA(phe)                 | U16        | 1.54                        | 0.52                        | extrahelical bulge    |
| tRNA(phe)                 | U17        | 1.81                        | 0.55                        | extrahelical bulge    |
| tRNA(phe)                 | U33        | 0.43                        | 1.69                        | apical loop           |
| tRNA(phe)                 | G34        | 0.37                        | 2.23                        | apical loop           |
| tRNA(phe)                 | A35        | 1.53                        | 1.85                        | apical loop           |
| tRNA(phe)                 | A36        | 1.21                        | 1.63                        | apical loop           |
| tRNA(phe)                 | A38        | 1.84                        | 1.55                        | apical loop           |
| tRNA(phe)                 | A58<sup>c</sup> | 5.34                        | 0.1                         | buried pocket         |
| Adenine riboswitch        | U36        | 1.7                         | 0.18                        | extrahelical bulge    |
| Adenine riboswitch        | U48        | 3.64                        | 1.58                        | extrahelical bulge    |
| Adenine riboswitch        | U62        | 1.61                        | 0.87                        | extrahelical bulge    |
| Cyclic-diGMP riboswitch   | A23        | 1.76                        | 0.03                        | buried pocket         |
| Cyclic-diGMP riboswitch   | A24        | 2.05                        | 0.06                        | buried pocket         |
| Cyclic-diGMP riboswitch   | A68        | 0.97                        | 1.78                        | apical loop           |
| P4-P6 RNA                 | A122       | 0.78                        | 2.46                        | 1-stack               |
| P4-P6 RNA                 | A125       | 0.86                        | 1.81                        | extrahelical bulge    |
| P4-P6 RNA                 | U168       | 0.18                        | 1.76                        | bulge (makes H-bond)  |
| P4-P6 RNA                 | A178       | 0.69                        | 2.65                        | 1-stack               |
| P4-P6 RNA                 | A198       | 4.55                        | 1.93                        | buried pocket         |
| P4-P6 RNA                 | U199       | 1.36                        | 2.26                        | extrahelical bulge    |
| P4-P6 RNA                 | A207       | 2.83                        | 0.49                        | buried pocket         |
| P4-P6 RNA                 | A219       | 2.06                        | 0.46                        | buried pocket         |
| P4-P6 RNA                 | A235       | 1.16                        | 1.9                         | apical loop           |
| P4-P6 RNA                 | U236       | 1.18                        | 2.18                        | apical loop           |
| Glycine riboswitch        | U21        | 0.92                        | 2.31                        | apical loop           |
| Glycine riboswitch        | A56        | 1.58                        | 0.15                        | extrahelical bulge    |
| Glycine riboswitch        | G73        | 0.11                        | 1.84                        | linker                |
| Glycine riboswitch        | A74        | 1.44                        | 1.70                        | linker                |
| Glycine riboswitch        | A77        | 1.84                        | 0.36                        | linker                |
| Glycine riboswitch        | A78        | 1.52                        | 0.27                        | linker                |
| Glycine riboswitch        | U96        | 0.14                        | 2.35                        | apical loop           |
| Glycine riboswitch        | A98        | 1.25                        | 1.64                        | apical loop           |
| Glycine riboswitch        | A125       | 1.52                        | 1.78                        | apical loop           |

<sup>a</sup>DMS reactivity for A, C; CMCT reactivity for G, U. (1.0 corresponds to A’s and U in GAGUA, respectively.)

<sup>b</sup>1M7 reactivity (1.0 corresponds to average reactivity over the five nucleotides of GAGUA).

<sup>c</sup>Not observed in prior work; see SI Methods “Averaging and error estimations based on multiple replicates”.
SUPPORTING FIGURE 1. Example of data averaging and error estimation. Output from `rdat_combine` function, available in HiTrace software. **Top panel:** heat-map representation of fourteen 1M7 measurements for the P4P6-2HP RNA enables rapid visualization of agreement and any sequence assignment errors; red-boxed residues are automatically determined outliers. **Bottom panel:** individual reactivity estimates (different colors) and averaged reactivity with error bars (black). Automatically determined outliers are marked with red circles.
SUPPORTING FIGURE 2. Proposed standardization brings data taken with varying chemical modifier concentrations into concordance. Modifiers tested were DMS (A-B), 1M7 (C-D), and CMCT (E-F) on the P4-P6-2HP RNA. (A,C,E) HiTRACE peak fits to capillary electrophoresis traces give ‘raw’ peak areas with modified concentrations noted; black curves show background (no modifier) data. (B,D,F) Normalized reactivities after standardization (saturation correction, over-modification correction, background subtraction, and normalization to GAGUA reference hairpin at nucleotides 278-282).
SUPPORTING FIGURE 3. CE and MAP-seq data for unmodified tRNA(phe), *S. cerevisiae*. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.
SUPPORTING FIGURE 4. CE and MAP-seq data for ligand-binding domain of adenine riboswitch, *V. vulnificus*. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.
SUPPORTING FIGURE 5. CE and MAP-seq data for ligand-binding domain of cyclic-di-GMP riboswitch, *V. cholerae*. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.
SUPPORTING FIGURE 6. CE and MAP-seq data for 5S ribosomal RNA, *E. coli*. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.
SUPPORTING FIGURE 7. CE and MAP-seq data for P4-P6 domain of the *Tetrahymena* ribozyme. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.
SUPPORTING FIGURE 8. CE and MAP-seq data for ligand-binding domains of the glycine riboswitch, *F. nucleatum*. In top panels, sequence of interest is between vertical black bars; GAGUA reference sequences marked in magenta.