

Sieg et al.

Supplementary Information

The Metabolome Weakens RNA Helix Stability and Increases RNA Chemical Stability

Jacob P. Sieg^{1,2}, Lauren McKinley^{1,2}, Melanie Huot^{3,4}, Neela H. Yennawar⁵,
and Philip C. Bevilacqua*^{1,2,3}

¹Department of Chemistry, Pennsylvania State University, University Park, PA 16802.

²Center for RNA Molecular Biology, Pennsylvania State University, University Park, PA 16802.

³Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802.

⁴Department of Biology, Pennsylvania State University, University Park, PA 16802.

⁵The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802.

Keywords

Magnesium ion, Metabolites, Chelated magnesium, RNA folding, RNA function, Near-cellular condition

ORCID:

Jacob P. Sieg: 0000-0001-5414-1667

Lauren McKinley: 0000-0002-3937-3811

Melanie Huot: 0000-0002-6669-2244

Neela H. Yennawar: 0000-0001-7278-659X

Philip C. Bevilacqua: 0000-0001-8074-3434

*Corresponding author

Contents

Supplementary information methods	3
Reagent preparation	3
Data analysis, transparency, and reproducibility	3
Apparent Mg ²⁺ binding constant determination with isothermal titration calorimetry	3
Determination of Mg ²⁺ speciation with HQS fluorescence	3
Apparent Mg ²⁺ binding constant determination with HQS	4
Determination of the total Mg ²⁺ required to have 2 mM free Mg ²⁺ in artificial cytoplasm	4
Artificial cytoplasm preparation	5
Single-binding-site statistical model for Mg ²⁺ speciation in artificial cytoplasm	5
Fluorescence binding isotherms	7
Fluorescence binding isotherm data analysis	7
MeltR concentration optimization algorithm	9
Helix folding energy error analysis	10
RNA transcription and purification	11
Small angle x-ray scattering (SAXS)	12
Supplementary information figures	14
SI figure 1	14
SI figure 2	15
SI figure 3	16
SI figure 4	16
SI figure 5	17
SI figure 6	17
SI figure 7	18
SI figure 8	19
SI figure 9	20
SI figure 10	21
SI figure 11	22
SI figure 12	23
SI figure 13	24
Supplementary information tables	25
SI table 1	25
SI table 2	26
SI table 3	26
SI table 4	26
SI table 5	27
SI 6able 6	28
SI table 7	29
SI table 9	30

Supplementary information methods

Reagent preparation

Buffers were prepared by dissolving salts in purified (18 MΩ*cm) water in volumetric flasks, at the purity specified in supplementary information (SI) table 1. All reagents contained a background of 20 mM MOPS, 0.01 mM EDTA, 0.001% (w/v) SDS, pH 7 buffer unless otherwise stated. Small amounts of EDTA were present to sequester trace divalent metal ion contamination. Small amounts of SDS were present to prevent RNA from sticking to tubes/pipette tips and to protect RNA from enzymatic degradation. To ensure accurate magnesium chloride concentrations, solutions were prepared by diluting standard 1.00 (± 0.01) M magnesium chloride solution from Sigma into buffer, or magnesium chloride hexahydrate was dissolved in buffer and then the concentration was determined with atomic absorption spectroscopy.

Data analysis, transparency, and reproducibility

Data analysis was performed in R (v4.1.2; R Core Team 2021) with the *tidyverse* package (v1.1.1; Wickham et al. 2019) for data wrangling and plotting. Unless otherwise stated, non-linear regression was performed with the *nls* function and linear regression was performed with the *lm* function in base R. The packages *MuMIn* (v1.46.0; Kamil Barton 2022), *ggpubr* (v0.4.0; Alboukadel Kassambara 2020), *viridis* (v0.6.2; Simon Garnier, Noam Ross, Robert Rudis, Antônio P. Camargo, Marco Sciaiani, and Cédric Scherer 2021), *ggbbeeswarm* (v0.6.0; Erik Clarke and Scott Sherrill-Mix 2017), and *cowplot* (v1.1.1; Claus O. Wilke 2020) were used to evaluate polynomial fits, display statistics, generate color-blind-friendly palettes, make beeswarm plots, and arrange figure panels, respectively. Secondary structures were drawn with R2easyR (<https://github.com/JPSieg/R2easyR>) and R2R. All raw data and analysis code is available with instructions for reconstitution on any console at (<https://github.com/JPSieg/JPSiegMetaboMetaloRNA>).

Apparent Mg²⁺ binding constant determination with isothermal titration calorimetry

Samples were degassed using a ThermoVac (MicroCal, LLC) degasser before loading into a VP-ITC MicroCalorimeter (MicroCal, LLC) according to the manufacturer's recommendations, with pure (18 MΩ*cm) water in the reference cell. Titration was performed with a 10 μcal/sec reference power and a stirring speed of 310 rpm. Twenty-nine injections (one 2 μL injection followed by twenty-eight 10 μL injections) were performed in total at an injection rate of 0.5 μL/sec with a 150 sec equilibration period following each injection. The first injection was not included in subsequent analysis. Data were analyzed using ITC parsing and non-linear regression fitting functions in MetaboMgITC (<https://github.com/JPSieg/MetaboMgITC>).¹

Determination of Mg²⁺ speciation with HQS fluorescence

8-Hydroxy-5-quinolinesulfonic (HQS) acid hydrate (98% purity) was purchased from Sigma Aldrich and recrystallized in water 10 times to remove trace metal contamination. Purified HQS was dissolved in buffer to a stock concentration of 100 mM (determined with UV absorbance using $\epsilon_{326\text{ nm}} = 2600 \text{ M}^{-1}\text{cm}^{-1}$). HQS was then diluted to a final concentration of 50 μM with variable concentrations of Mg²⁺ and metabolites with a final sample volume of 100 μL. Control samples contained 240 mM NaCl and 140 mM KCl.

HQS emission was measured with a Biotek Cytation 3 plate reader with an excitation of 355 nm and emission of 500 nm. Plates were sealed with transparent PCR film and were incubated at 37 °C for 10 min prior to reading to ensure thermal equilibration.

Fluorescence data in the absence of metabolite chelators was fit to Equation 1, in order to determine the inverse molar apparent association constant for HQS binding to Mg²⁺ (K_{HQS}).

$$\text{Emission} = (F_{\max} - F_{\min}) \left(\frac{K_{\text{HQS}}[\text{Mg}]}{1 + K_{\text{HQS}}[\text{Mg}]} \right) + F_{\min} \quad (1)$$

[Mg] is the free Mg²⁺ concentration in the sample, which is equal to the total concentration of magnesium chloride in the absence of chelators. F_{max} and F_{min} are the intensity of the Mg²⁺-HQS complex and the intensity of free HQS, respectively. The fluorescence for all samples was normalized using I_{max} and I_{min} to produce a normalized fluorescence intensity F_{norm}.

$$F_{\text{norm}} = \left(\frac{\text{Emission} - F_{\min}}{F_{\max} - F_{\min}} \right) \quad (2)$$

Free Mg²⁺ concentrations for each sample were calculated with Equation 3, using F_{norm} and K_{HQS}.

$$[\text{Mg}] = \frac{F_{\text{norm}}}{K_{\text{HQS}}(1 - F_{\text{norm}})} \quad (3)$$

Apparent Mg²⁺ binding constant determination with HQS

Metabolite concentrations are provided in the legend of SI figure 2 for HQS experiments to determine apparent binding constants, and extra NaCl or KCl was added to maintain a constant total monovalent metal ion concentration of 240 mM Na⁺ and 140 mM K⁺. Metabolite solutions were treated with chelex resin (Sigma) to remove divalent ion contamination. The solution was passed through a 0.2 µm filter and the pH was readjusted to 7.0. The free Mg²⁺ concentration in the presence of chelators versus the total Mg²⁺ was fit to Equation 4 to obtain K'.

$$[\text{Mg}] = [\text{Mg}]_T - \frac{([\text{Mg}]_T + [L]_T + \frac{1}{K'}) \pm \sqrt{([Mg]_T + [L]_T + \frac{1}{K'})^2 - 4[\text{Mg}]_T[L]_T}}{2} \quad (4)$$

Where [Mg]_T is the total Mg²⁺ concentration, [L]_T is the metabolite concentration, and K' is the apparent ligand binding constant for Mg²⁺, represented by Equation 5.

$$K' = \frac{[\text{MgL}]}{[\text{Mg}][\text{L}]} \quad (5)$$

Determination of the total Mg²⁺ required to have 2 mM free Mg²⁺ in artificial cytoplasm

The total Mg²⁺ concentration required to have 2 mM free Mg²⁺ in artificial cytoplasm was approximated using the free Mg²⁺ concentrations calculated using HQS data. The free Mg²⁺ concentration, which was measured as a function of the total Mg²⁺, was fit to 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th and 10th order polynomials in the form:

$$[Mg] = \sum_{i=1}^N A_i [Mg]_T^i + C \quad (6)$$

Where N is the order of the polynomial and A/C are constants determined by linear regression. Fit polynomials were evaluated by calculating the Akaike information criterion (AIC), in order to determine which polynomial best described the data using the minimum number of terms. The polynomial with the lowest AIC was then solved numerically to calculate the total Mg^{2+} required to generate 2 mM free Mg^{2+} in the presence of metabolites, with a tolerance of 0.01 mM. The fluorescence of control samples containing 10 mM EDTA was compared to the no added $MgCl_2$ condition to determine divalent metal ion contamination levels in artificial cytoplasm.

Artificial cytoplasm preparation

Artificial cytoplasm was prepared to maintain a final monovalent ion concentration of 240 mM Na^+ and 140 mM K^+ and a final pH of 7.0. First, metabolite salts or solutions were dissolved or diluted into water in volumetric flasks. Second, the pH of each solution was adjusted to 7.0 using 10 M NaOH. The amount of Na^+ or K^+ added with each component was recorded. Third, 5 M $NaCl$ and 2 M KCl were added for a final Na^+ and K^+ concentration of 480 mM and 280 mM, respectively. Then, the volumetric flask was filled with water to 10 mL. Lastly, a test quantity of the 2x solution was diluted to a 1x concentration and the pH was tested with a VWR Symphony pH probe. The 2x concentrated artificial cytoplasm was aliquoted and stored at -20 °C. A detailed recipe for the Eco80 artificial cytoplasm is available in SI Table 1. Periodically, the artificial cytoplasm was spot checked for NTP degradation using thin-layer chromatography on PEI-Cellulose F plates (Supelco) developed with a 0.3 M potassium phosphate pH 7.0 mobile phase.

Single-binding-site statistical model for Mg^{2+} speciation in artificial cytoplasm

We used a statistical model, that considers experimental errors in metabolite/ Mg^{2+} concentrations and uncertainties in apparent binding constant determination, to calculate the expected free Mg^{2+} concentration in artificial cytoplasm, assuming single-site binding (meaning one metabolite molecule binds one Mg^{2+} and vice-versa). Single site binding of a metabolite to Mg^{2+} is described by the apparent dissociation constant (K_D').

$$K_D' = \frac{[L][Mg]}{[MgL]} \quad (7)$$

Rearranging to determine the concentration of the metabolite-ligand complex $[MgL]$ as a function of the concentration of free Mg^{2+} , and the total ligand concentration ($[L]_T$), we get equation 8.

$$[MgL] = \frac{[L]_T[Mg]}{K_D' + [Mg]} \quad (8)$$

In a mixture of metabolites, the total Mg^{2+} will be the free Mg^{2+} , plus the Mg^{2+} bound by different metabolites.

$$[Mg]_T = [Mg] + \sum_{i=1}^N [MgL_i] \quad (9)$$

Commented [SJP1]: Phil: "Do you mean Cl-?"

Jacob: No. This is an interesting point that I don't have my mind wrapped around. But there is something profound here. Here are two things I am thinking about.

(1) Metabolites were ordered as either Na or K salts. It is not possible to get to 240 mM Na 140 mM K pH 7.0 using only Na salts for *E. coli* metabolite concentrations because there has to be electroneutrality. If you tried using only sodium, you would go beyond 240 mM Na .

(2) Eukaryotes, have much less NTPs, which come as a charge of -2 to -3. I have not tried, but according to math, it is possible to make something like a Mammalian imbk 80 artificial cytoplasm with solely 240 mM Na at pH 7. Likewise, Eukaryotes are generally thought to have much less K in the cell, like 10 mM.

(3) Essentially what we are doing here is replacing Cl^- anions with metabolite anions. And this is destabilizing to RNA secondary structure.

N is the total number of metabolites in the solution. Equation 8 was plugged into Equation 9 to obtain the total Mg^{2+} as a function of the free Mg^{2+} , the apparent dissociation constant, and the metabolite concentration.

$$[Mg]_T = [Mg] + \sum_{i=1}^N \frac{[L_i]_T [Mg]}{K'_{D_i} + [Mg]} \quad (10)$$

The total Mg^{2+} and the total metabolite concentrations are set in the experiment, and the apparent dissociation constants were determined with ITC or HQS fluorescence. Thus, the expected free Mg^{2+} was determined to a precision of 0.01 mM by solving equation 10 numerically (possible because equation 10 has one real solution between 0 free Mg^{2+} and the total Mg^{2+} concentration in the solution).

Expected experimental error was calculated by propagating uncertainties in mass and volume during sample preparation. Briefly, equation 11 and 12 was used for propagation of uncertainty for addition and multiplication/division respectively.

$$\text{For } A = B + C \text{ or } B - C \left\{ dA = \sqrt{dB^2 + dC^2} \right\} \quad (11)$$

$$\text{For } A = B * C \text{ or } A = \frac{B}{C} \left\{ dA = A \sqrt{\left(\frac{dB}{B}\right)^2 + \left(\frac{dC}{C}\right)^2} \right\} \quad (12)$$

The percent impurities (supplied by the chemical vendor) were used as the uncertainty for mass measurements because the uncertainty from the impurity was greater than the precision of our analytical balance (0.1 mg). Volume uncertainties were provided by the manufactures of the pipettes and volumetric flasks. For metabolites, uncertainty in concentration was on average 9.1 % of the final concentration in artificial cytoplasm. Thus, we used 10% uncertainty for metabolite concentrations.

For the statistical model, random error was seeded into equation 10 using the *rnorm* function in base R, which creates random error based on a gaussian distribution with a standard deviation (σ_x) described by equation 13.

$$\sigma_x = \frac{\% \text{uncertainty}}{100\%} x \quad (13)$$

The standard deviation provided by the fit was used for apparent dissociation constants. Thus, Equation 14 can be used to generate a virtual artificial cytoplasm with realistic errors in reagent concentrations.

$$[Mg]_T + rnorm(\sigma_{[Mg]_T}) = [Mg] + \sum_{i=1}^N \frac{([L_i]_T + rnorm(\sigma_{[L_i]_T})) [Mg]}{K'_{D_i} + rnorm(\sigma_{K'_{D_i}}) + [Mg]} \quad (14)$$

One thousand virtual artificial cytoplasm were created with Equation 14 to fully explore possible variance and the results were compared with experimental measurements of free Mg²⁺ determined with HQS.

Fluorescence binding isotherms

5'-FAM and 3'-BHQ1 labeled RNA were ordered from Integrated DNA Technology (IDT) and HPLC purified by IDT or in house. For in house HPLC, RNA peaks were manually collected on a Water's AQUITY Arc UPLC system equipped with a Waters Xbridge C18 2.3 μm 4.6x150 mm column. 5 nmole of RNA was injected on a 60 °C preheated column and separated for 5 min with 100% 0.1 M aqueous triethylamine acetate (TEAA), followed by a 25 min transition to 100% 0.1 M TEAA 20% acetonitrile, followed by 5 minutes of 100% 0.1 M TEAA 20% acetonitrile, and then a 5 min transition back to 100% aqueous 0.1 M TEAA, at a flow rate of 1 mL/min. Samples were dried under vacuum and resuspended in buffer. Then, samples were dialyzed twice against 1 L of buffer in an 8 well Life Technologies Microdialysis System equipped with a 1 kDa dialysis membrane (SpectraPore) for 24 hours per liter of buffer. Dialyzed RNA concentrations were determined using the absorbance at 260 nm and extinction coefficients provided by IDT. Final samples were prepared in triplicate, where FAM-RNA were diluted to 200 nM and BHQ1-RNA were diluted to 0, 1, 10, 50, 100, 150, 200, 400, 800, and 1000 nM (SI figure 3). The 2x concentrated salts or artificial cytoplasm were diluted into the same samples to achieve a final 1x concentration with 0.5 mM free ROX as a passive reference dye. Samples were heated to 90 °C for 1.5 min, cooled at room temperature for 20 min, transferred into a 96 well qPCR plate, sealed with transparent qPCR film, and centrifuged for 2 min to remove air bubbles. Fluorescence emission of the entire plate as a function of temperature was measured using an Applied Biosystems Step 1 Plus qPCR instrument. Samples were heated from 20 to 80 °C at a ramp rate of 0.5 °C per minute, recording fluorescence at 0.5 °C increments. ROX-normalized FAM emission was exported from the Applied Biosystems software to a flat text file.

Fluorescence binding isotherm data analysis

Fluorescence binding isotherms were fit using the *meltR.F* program in the MeltR package V0.9. MeltR performs the following data preprocessing steps before fitting:

- 1.) The fluorophore labeled strand concentration is optimized using the concentration optimization algorithm, described in detail below.
- 2.) Isotherms are fit to Equation 15 to determine K_D and error in the K_D at each temperature using "nls" in base R.

$$F = F_{\max} + (F_{\min} - F_{\max}) \frac{(K_D + [F]_T + [Q]_T) - \sqrt{(K_D + [F]_T + [Q]_T)^2 - 4[F]_T[Q]_T}}{2[F]_T} \quad (15)$$

Where F is the observed fluorescence emission, F_{max} is the fluorescence emission of the unbound fluorophore labeled RNA strand, F_{min} is the fluorescence emission of the fluorophore labeled RNA strand bound to the quencher labeled RNA strand, [F]_T is the total fluorophore labeled RNA strand concentration, [Q]_T is the total quencher RNA strand concentration, and K_D is the dissociation constant given by Equation 16.

$$K_D = \frac{[F][Q]}{[Q]} \quad (16)$$

$[F]$ is the concentration of free fluorophore labeled RNA, $[Q]$ is the concentration for free quencher labeled RNA, and $[FQ]$ is the concentration of the helical complex. Initial values for F_{\max} and F_{\min} are estimated by taking the mean of the 20% highest readings in each isotherm and the 20% lowest readings respectively. Initial values for the K_D are provided by the user, by default 0.1 nM.

3.) K_D s are filtered by magnitude and error, according to user specifications, to determine which isotherms are most reliable. First, K_D s outside of a user specified range (10 to 1000 nM by default) are thrown out. Second, K_D s are ranked by the error in the K_D . K_D s that are below a user specified error quantile are thrown out. The default K_D error quantile is 0.25, meaning the algorithm will keep the 25% most accurate K_D s, after filtering by magnitude (Filtering is discussed in more detail below).

4.) The most reliable K_D s and temperatures are passed to Method 1 to make Van't Hoff plots and to calculate helix formation energies.

5.) Fluorescence data that produce reliable K_D s are passed to Method 2 for global fitting and to calculate helix formation energies.

Method 1 fits a Van't Hoff plot to Equation 17 to determine the enthalpy and entropy of folding.

$$\ln(K_D) = \frac{\Delta S'}{R} - \frac{\Delta H'}{RT} \quad (17)$$

Where $\Delta S'$ is the entropy of helix dissociation, $\Delta H'$ is the enthalpy of helix dissociation, R is the gas constant and T is the temperature in Kelvin.

Method 2 Globally fits raw fluorescence data to equation 18, produced by plugging equation 17 into equation 15 to determine the entropy and enthalpy of folding.

$$F = F_{\max}^i + (F_{\min}^i - F_{\max}^i) \cdot \frac{\left(e^{\frac{\Delta S'}{R} - \frac{\Delta H'}{RT}} + [F]_T + [Q]_T \right) - \sqrt{\left(e^{\frac{\Delta S'}{R} - \frac{\Delta H'}{RT}} + [F]_T + [Q]_T \right)^2 - 4[F]_T[Q]_T}}{2[F]_T} \quad (18)$$

" i " is the reading when the data was collected, meaning that the F_{\max} and F_{\min} can float between readings but the $\Delta S'$ and $\Delta H'$ are required to be the same.

Helix folding energies are traditionally reported in terms of the association constant.

$$K = \frac{1}{K_D} = \frac{[FQ]}{[F][Q]} \quad (19)$$

Thus, the entropy and enthalpy of helix association, ΔS and ΔH , are obtained by multiplying the $\Delta S'$ and $\Delta H'$ by negative 1.

MeltR reports standard error (SE) estimates provided by the non-linear model generated with the *nls* function in base R. The Gibbs free energy of helix association at 37 °C ($\Delta G^{37\text{°C}}$) is calculated with Equation 20.

$$\Delta G^{37\text{°C}} = \Delta H - 310.15\Delta S \quad (20)$$

Error in the $\Delta G^{37\text{°C}}$ of helix association was calculated with Equation 21.

$$SE_{\Delta G^{37\text{°C}}} = \sqrt{SE_{\Delta H}^2 + (310.15SE_{\Delta S})^2 - 620.3 \frac{\sigma_{\Delta H, \Delta S}}{\Delta H, \Delta S}} \quad (21)$$

Where $\sigma_{\Delta H, \Delta S}$ is the covariation between the ΔS and ΔH given by *nls* in base R. Additional error analysis is described below.

MeltR concentration optimization algorithm

We found that helix energies from fitting fluorescence binding isotherms were highly dependent on errors in the determination of RNA concentrations in stock solutions, which is propagated systematically during sample preparation. For example, if the FAM-RNA stock was actually 20% more concentrated than the estimate calculated using UV absorbance and extinction coefficients, the FAM-RNA concentration in the final samples would be consistently 20% higher because the same amount of master mix was added to each sample (SI Figure 3). This can bias the resulting helix folding energies.

To understand why, we modeled data assuming a ΔH , ΔS , and $\Delta G^{37\text{°C}}$ of -56.2 kcal/mol, -136.4 cal/mol/K, and -13.9 kcal/mol respectively, 5% random fluorescence error, and perfectly accurate determination of RNA concentrations in concentrated stocks (a 200 nM FAM-RNA concentration, and 0, 1, 10, 50, 100, 150, 200, 250, 400, 600, 800, and 1000 nM RNA-BHQ1 concentrations). We then used MeltR to fit the modeled data, resulting in a very accurate determination of $\Delta H = -56.0$ kcal/mol, $\Delta S = -135.7$ cal/mol/K, and $\Delta G^{37\text{°C}} = -13.9$ kcal/mol. We next considered how assuming incorrect RNA stock concentrations, thus incorporating a systematic error, could effect the accuracy of the fits. Systematic error (-50% to +50%) was seeded into virtual stock concentrations and the data were refit (SI figure 12A). We found that fit accuracy was highly dependent on error in stock concentrations. For example, if the FAM-RNA concentration in the stock was 25% lower than the estimate, and the RNA-BHQ1 concentration was 25% higher than the estimate, the $\Delta G^{37\text{°C}}$ that MeltR calculated was about 4 kcal/mol off. Likewise, if the FAM-RNA concentration in the stock was 25% higher than the estimate, and the RNA-BHQ1 concentration is 25% lower than the estimate, the $\Delta G^{37\text{°C}}$ that MeltR calculated was about 2.5 kcal/mol off. However, if the FAM-RNA error and the RNA-BHQ1 error compensated for each other, e.g. %RNA-BHQ1 error = %FAM-RNA error, the $\Delta G^{37\text{°C}}$ that MeltR calculates was less than 0.2 kcal/mol off, even with 50% inaccuracy in RNA stock.

We next considered a more realistic scenario, where the experiment assumed perfectly accurate determination of RNA concentrations but there was actually +20% FAM-RNA concentration error. Data was modeled using the same folding energies and RNA-BHQ1 concentrations, but with a 240

nM FAM-RNA concentration (+20% error). We then used MeltR to fit the modeled data, assuming a 200 nM FAM-RNA concentration, resulting in inaccurate determination of $\Delta H = -35.8$ kcal/mol, $\Delta S = -75.5$ cal/mol/K, and $\Delta G^{37^\circ C} = -12.3$ kcal/mol, a $\Delta G^{37^\circ C}$ error of 1.6 kcal/mol. Once again, we seeded systematic error (-50% to +50%) into virtual stock concentrations and the data were refit (SI figure 12B). Similar to SI figure 12A, we found that fit accuracy was highly dependent on errors in stock concentrations. However, the fits were most accurate according to Equation 22 instead of %RNA-BHQ1 error = %FAM-RNA error.

$$\%BHQ1_{error} = \frac{1}{X} \%FAM_{error} + \frac{100 - 100X}{X} \quad (22)$$

Where X is the actual FAM-RNA concentration divided by the estimated FAM-RNA concentration (240 nM/200 nmol = 1.2 in this example). Thus, MeltR does not need perfectly accurate concentrations, just an optimization algorithm that finds the FAM-RNA concentration correction factor X. Note, %RNA-BHQ1 error = %FAM-RNA error is a special case of Equation 22, where X = 1. To find X, MeltR selects an isotherm (usually the lowest temperature), where the K_D is more than 10 times less than the FAM-RNA labeled concentration (SI figure 12C). At this K_D range, the shape of the binding curve is independent of K_D , and MeltR uses the isotherm as a Job plot to determine X. MeltR then uses X to correct the FAM-RNA concentration.

We next tested the MeltR concentration optimization algorithm. The modeled data, with an uncorrected +20% FAM-RNA concentration error, was fit using MeltR with the concentration optimization algorithm on, resulting in an accurate determination of $\Delta H = -51.9$ kcal/mol, $\Delta S = -123.5$ cal/mol/K, and $\Delta G^{37^\circ C} = -13.6$ kcal/mol, a $\Delta G^{37^\circ C}$ error = 0.3 kcal/mol. We then seeded additional error into the data set (-50% to +50%), refit the data using the MeltR concentration optimization algorithm, and found that MeltR calculates accurate folding energies (within 0.2 kcal/mol in terms of the $\Delta G^{37^\circ C}$ on average) even with large inaccuracies in reagent concentration estimates (SI Figure 12D).

MeltR then filters out isotherms that produce inaccurate K_D s, according to user specifications (see above). Both the MeltR K_D range, and K_D error threshold can be adjusted by the user in real time to refine fits by manually inspecting Van't Hoff plots to identify the most accurate temperature range to fit. For example, the MeltR fit of the data that was modeled using +20% FAM-RNA concentration error can be improved to $\Delta H = -56.4$ kcal/mol, $\Delta S = -136.9$ cal/mol/K, and $\Delta G^{37^\circ C} = -13.9$ kcal/mol using the concentration optimization algorithm and by selecting the most accurate temperature range by inspecting the Van't Hoff plot.

Helix folding energy error analysis

MeltR reports standard errors propagated from the fit, which reflect the precision of the parameters and can be treated statistically.² However, the accuracy of such parameters is also dependent on systematic errors from instrument calibration, concentration estimates, RNA impurities, etc. These are more difficult to control for and treat statistically. Precision estimated from the fit for method 1 were 1.7%, 2%, and 0.3 % uncertainty on average for the ΔH , ΔS , and $\Delta G^{37^\circ C}$ (SI table 5). Method 2 was less precise, at 21.7%, 26.5%, and 3.4 % uncertainty on average for the ΔH , ΔS , and $\Delta G^{37^\circ C}$. However, the difference in helix folding energies between the two methods was much smaller, at 1.2%, 1.4%, and 0.2% on average for the ΔH , ΔS , and $\Delta G^{37^\circ C}$. Given that the two methods provide similar helix folding energies, but different precisions, the discrepancy in standard error likely reflects differences in the number of parameters that must be estimated by the fit (two for method 1, ΔH and ΔS ; 2+2*number of raw isotherms for method 2, ΔH , ΔS , and a

F_{\max}/F_{\min} for each raw isotherm), more than it accuracy in helix folding energy estimation. Turner and colleagues estimated that an uncertainty of 12%, 13.5%, and 4% for the ΔH , ΔS , and $\Delta G^{37^\circ\text{C}}$, respectively, conservatively reflects systematic and random errors for absorbance melting curves.² Given that the discrepancy between method 1 and method 2 were smaller than the Turner uncertainty for all fluorescence binding isotherms in this study, and that we are determining differences between conditions on the same sequences collected in the same lab, the 4% uncertainty in terms of the $\Delta G^{37^\circ\text{C}}$ was too conservative. We determined that there was on average 0.2 kcal/mol, or 1.5%, error for MeltR fitting modeled fluorescence data using the concentration optimization algorithm (SI Figure 12). This deals with errors in RNA concentration estimates, and important source of systematic error. Thus, we reported the $\Delta G^{37^\circ\text{C}}$ of helix formation from method 1 with an uncertainty of 1.5% in Table 3 and Figure 2D.

RNA transcription and purification

Yeast tRNA^{Phe} was transcribed from 7 μM hemiduplex DNA and Guanine aptamer/CPEB3 ribozymes were transcribed from primer extended duplexes prepared with polymerase chain reaction (PCR) reactions (Supplementary table 8).³⁻⁵ PCR was performed with Taq polymerase (New England Biolabs) and cleaned up with a RNA clean and concentrate kit (Zymogen), according the manufacturer's instructions. For cleaning up large scale PCR to make template for SAXS experiments, polymerase was removed with an equal volume 25:24:1 phenol:chloroform:isoamyl alcohol extraction, and two subsequent equal volume 24:1 chloroform:isoamyl alcohol extractions. The dNTPs were removed using an ethanol precipitation with ~ 2 M ammonium acetate and $>75\%$ ethanol. Transcription reactions were performed with in-house prepared T7 polymerase at 37°C for 3 hr, with 4 mM ATP, 4 mM CTP, 4 mM UTP, 8 mM GTP, 25 mM MgCl₂, 0.1 mM spermidine, 2 mM dithioritol, and 20 mM Tris pH 8.0. NTP stocks were preadjusted to pH 8.0 with Tris base. Transcriptions were stopped with the addition of EDTA to a final concentration of 50 mM and cleaned up using phenol:chloroform:isoamyl alcohol extractions (as described for the PCR clean up) followed by ethanol precipitations containing ~ 2 M ammonium acetate and $>75\%$ ethanol. RNA was resuspended in 50% formamide with 25 mM EDTA (pH 8.0), heated to 90°C , and loaded hot onto a denaturing (8.3 M urea) 10% (g/mL) acrylamide gel containing 1x TBE. Full length RNA bands were identified with UV-shadowing, excised, and electroeluted (Elutrap, Whatman) according to the manufacturer's recommendation. Electroluted RNA was concentrated with an ethanol precipitation containing ~ 0.8 M LiCl and $>75\%$ ethanol, resuspended in buffer, and stored at -20°C .

For 5'-end labeling, the 5'-triphosphate on the transcription product was removed from about 10 pmol of RNA using shrimp alkaline phosphatase SAP (New England Biolabs). SAP was heat inactivated by incubating at 72°C for 5 min. RNA from the SAP reactions were 5'-³²P labeled with 0.5 to 3.5 μM [γ -³²P] ATP (PerkinElmer, 5 mCi) using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's recommendation. Kinase reactions were stopped with an equal volume of 90% formamide with 50 mM EDTA pH 8.0 loading dye containing trace xylene cyanol and bromophenol blue. Kinased products were excised from a denaturing (8.3 M urea) 10% (g/mL) acrylamide gel containing 1x TBE, crushed and soaked overnight in 1 mM EDTA, 250 mM NaCl, and 10 mM Tris pH 7.5 buffer at 4°C , ethanol precipitated, and resuspended in 100 μL of buffer. For internal (body) labeling, RNA was transcribed with an additional 2 μL of [α -³²P] ATP (Perkin Elmer, 250 μCi) and purified as described for the kinase reaction.

In-line probing chemical degradation assay. For in-line probing (ILP) degradation assays, 5 to 10 kcpm/ μ L of 5'-³²P labeled RNA was incubated in triplicate at 37 °C on a PCR block. Time points were taken at 0, 2, and 4 days by quenching 10 μ L of reaction in 10 μ L of 95% formamide containing 50 mM EDTA pH 8.0 buffer with trace xylene cyanol and bromophenol blue. Then, 3 μ L of sample was fractionated on a denaturing (8.3 M urea) 10% (g/mL) acrylamide gel containing 1x TBE. The gel was dried, exposed on a phosphoimager cassette, and scanned with a Typhoon Phosphoimager. Gels images were quantified with SAFA⁶ three times to capture variance between qualifications. Time points were corrected for background degradation at each band by subtracting out the band intensity at time 0. The change in counts with time was determined by a simple linear fit at each band ($y = mx + b$), where the slope "m" is the change in counts with time.

Small angle x-ray scattering (SAXS)

Transcribed and purified RNA was dialyzed twice against 1 L of buffer containing 2 mM MgCl₂ in an 8 well Life Technologies Microdialysis System equipped with a 1 kDa dialysis membrane (SpectraPore) for 24 hours per liter of buffer. RNA were renatured at 80 °C for 1.5 min and cooled to room temperature for 20 min. About 700 μ g of RNA was ran on an Agilent 1260 Infinity II HPLC system equipped with a 100 Å pore 5 μ m silica bead 7.8 x 300mm Wyatt SEC hydrophilic column at 4 °C to purify RNA in the monomer conformation, with the dialysis buffer as the mobile phase (0.5 mL/min). The correct molecular weight of monomers was confirmed using an in-line Wyatt Technology DAWN MALS and Wyatt Optilab Refractive Index (RI) detector. Monomers were then diluted, to 100 μ g/mL, in artificial cytoplasm with Mg²⁺. RNA was incubated at 37 °C for 1 hour before SAXS data acquisition.

SAXS data was collected at 1.54 Å using a Rigaku MM007 rotating anode and a BioSAXS200nano Kratky camera system. The system includes OptiSAXS confocal max-flux optics and a sensitive HyPix-3000 hybrid photon counting detector. The sample capillary-to-detector distance was 495.8 mm and was calibrated using silver behenate powder (The Gem Dugout, State College, PA). RNA samples were loaded using a autosampler onto a quartz capillary flow cell mounted on a stage cooled to 4°C. The sample cell and full X-ray flight path, including beam stop, were kept in vacuo (< 1 mtorr) to eliminate air scatter. Data reduction including image integration, normalization, and background buffer data subtraction were performed with SAXSLAB software. Six ten-minute images from RNA and buffer samples were collected and averaged after ensuring that no X-ray radiation damage had occurred. This was followed by reference buffer subtraction to get the raw SAXS scattering curve from only the RNA (SI figure 7A).

The Gunier analysis, distance distributions (SI figure 7B), and porod analysis were calculated using Primus in the ATSAS package (SI table 6).⁷ Ambiguity scores for Ab-initio shape determination were close to 2.5 (SI table 7), calculated with AMBIMETER in the ATSAS package, so great care was taken for Ab-initio shape reconstruction (SI figure 7 F-G). Electron density maps were reconstructed by averaging 80 individual reconstructions using DENSS.⁸ Bead models were created with the ATSAS package by generating 20 individual models using DAMMIF, averaged using DAMAVER, and the average was refined using DAMMIN. R_Gs were in good agreement between the P(r) plot, the electron density reconstructions, and the bead model and the bead model reconstruction statistics were close to ideal (SI table 9), indicating that *ab-initio* reconstructions were reasonable representations of the solution state. Electron density maps and bead models were

Commented [SJP2]: Phil: Why, It's math on the same data?

Jacob: SAFA performs a number of steps that are idiosyncratic, in particular the alignment step. Since my cleaved was so weak (pH 7.0 and low free Mg) I quantified the gel more than once to see how reproducible it was. I did see variance between quantifications so I quantified everything three times.

I do not think it is necessary for kinetics because the bands are strong and there are only two of them

overlaid on the crystal structure of the guanine aptamer bound to its ligand (PDB 4FE5) with solvent molecules modeled using WAXSiS.^{9,10}

CPEB3 self-cleaving assay. Internal (body) ³²P labeled pre-CPEB3 ribozyme was diluted into artificial cytoplasm to a final concentration of 250 cpm/µL. RNA was renatured at 95 °C for three minutes then cooled in a 37 °C heat block for 10 min. Reactions were initiated by adding a 1/10 volume of a 10x MgCl₂ stock, and time points were taken by quenching 4 µL of reaction in 4 µL of 50 mM EDTA pH 8.0, 90% formamide, with trace xylene cyanol and bromophenol blue. Then, 4 µL of sample were fractionated on a denaturing (8.3 M urea) 10% (g/mL) acrylamide gel containing 1x TBE. Gels were dried and imaged with a PhosphorImager (Typhoon 650; GE Healthcare) and quantified with ImageQuant (GE Life Sciences). The fraction cleaved (*f*) was calculated from Equation 23, normalizing raw band intensities (*I*) for the number of As in the cleaved CPEB3 and pre-CPEB3 bands.

$$f = \frac{\frac{I_{\text{CPEB3}}}{13}}{\frac{I_{\text{pre-CPEB3}}}{21} + \frac{I_{\text{CPEB3}}}{13}} \quad (23)$$

Rates of self-cleavage were analyzed by fitting the fraction cleaved to a single exponential curve (Equation 24).

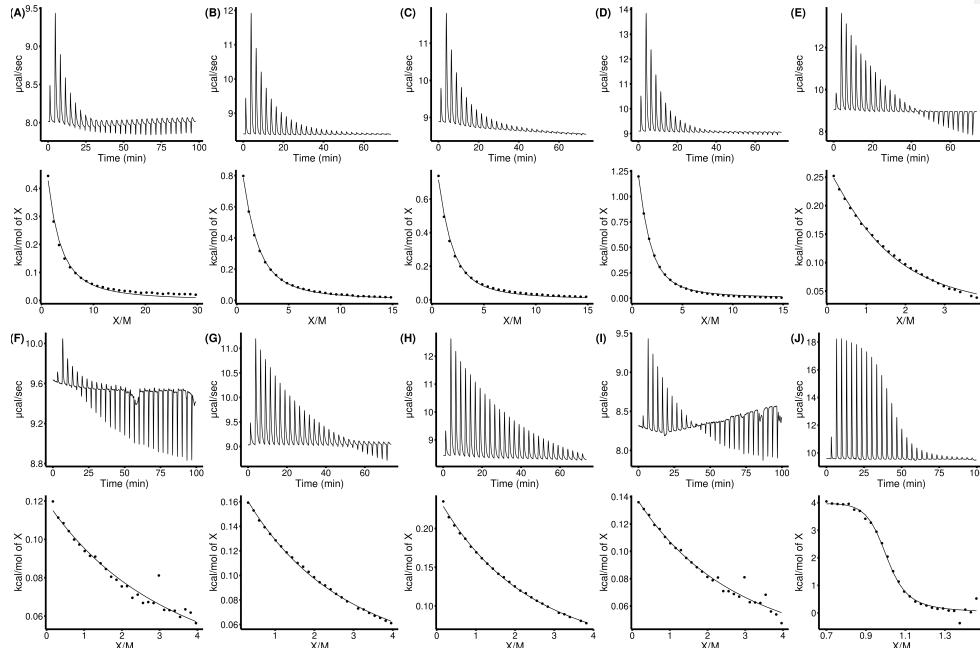
$$f = A + Be^{-kt} \quad (24)$$

Where *A* is the fraction of ribozyme cleaved at completion, -*B* is the amplitude of the observable phase, and *k* is the observed first-order rate constant for ribozymes cleaving in non-burst phase, and *t* is the time in minutes.

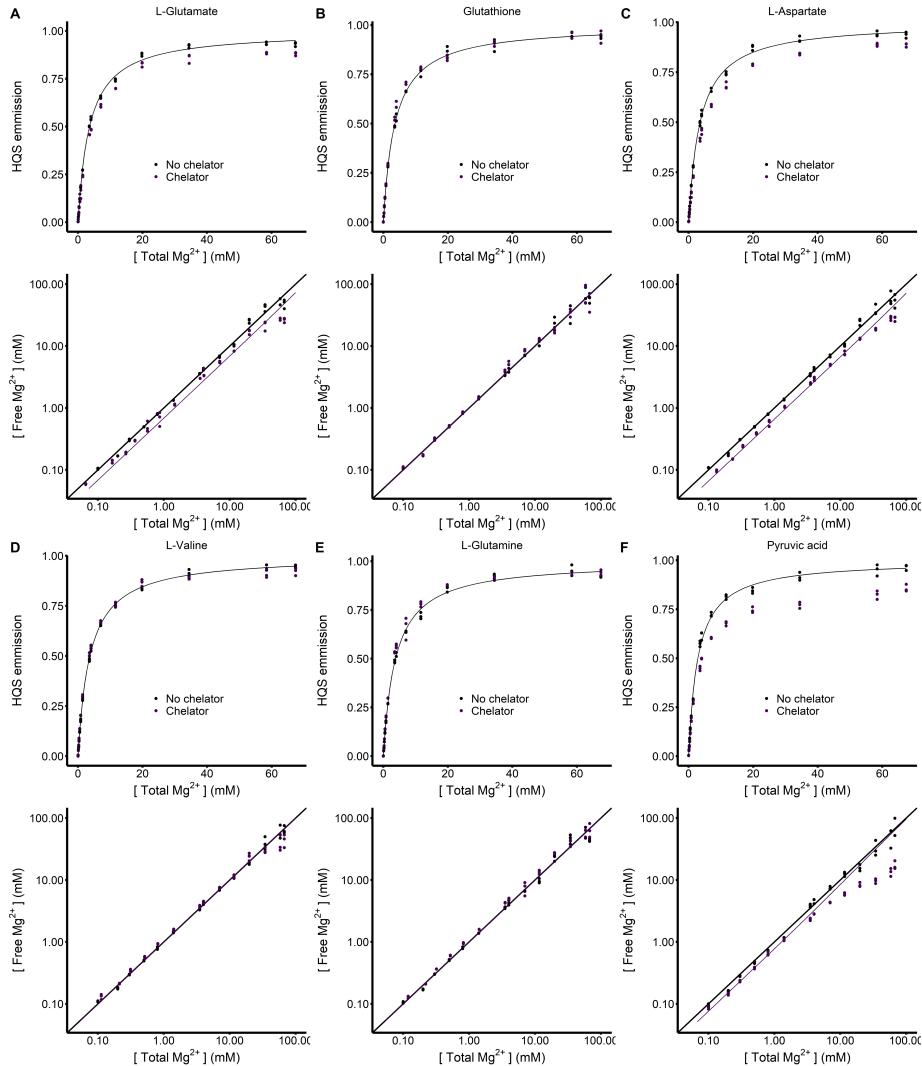
Commented [SJP3]: Phil: I believe A + B is the burst phase. Should there be one?

Jacob: I don't think a HDV burst phase has been observed but you would know better than me. But, there could be, because there is some magnesium in Eco80 and WMCM before the 10xMg is added because the purest dihydroxyacetone phosphate I could find from a reliable vendor was a hemi-magnesium salt. This should only be a small amount because the metabolites should sequester it. However, I looked for the burst phase. The analysis of the kinetics is preliminary but A + B is about 0 so I do not think there is a burst phase.

Supplementary information figures



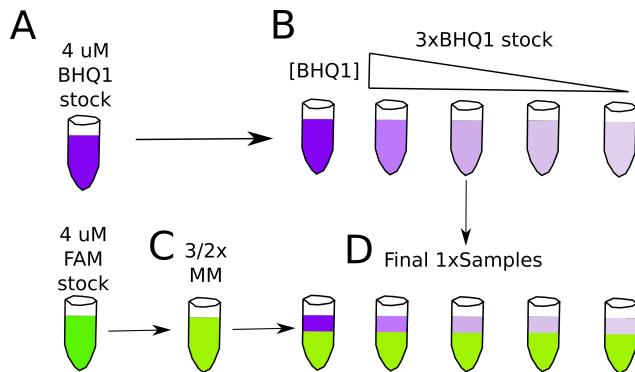
SI figure 1 Isothermal titration calorimetry (ITC) analysis of Mg^{2+} binding to metabolites in 240 mM NaCl, 140 mM KCl, and 10 mM HEPES pH 7.0 at 37 °C. $MgCl_2$ was titrated into metabolites and the power was monitored over time (Top panel). Heat of the injection was calculated by integrating the raw power curve, and the background heat of $MgCl_2$ dilution collected on buffer containing no metabolite. The background was subtracted to produce the isotherms in the bottom panels. Lines in bottom panels represent fits to the Weismann isotherm equation to determine apparent association constants. (A) Adenosine triphosphate (ATP). (B) Uridine triphosphate (UTP). (C) Guanosine triphosphate (GTP). (D) Deoxythymidine triphosphate (dTTP). (E) Fructose 1,6-bisphosphate. (F) Uridine diphosphate (UDP)-N-acetylglucosamine. (G) Glucose 6-phosphate. (H) 6-phosphogluconic acid. (I) Phosphoenol pyruvate. (J) Ethylene diamine-tetraacetic acid (EDTA). Thermodynamic values are found in SI table 2.



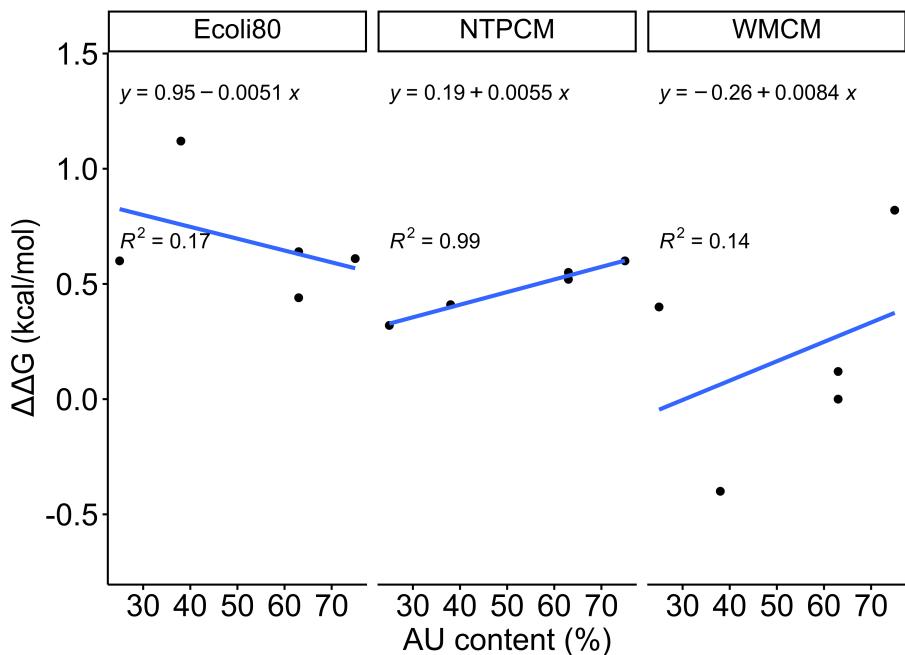
SI figure 2 HQS analysis of Mg^{2+} binding to metabolites in 240 mM NaCl 140 mM KCl 20 mM MOPS 0.01 mM EDTA 0.001% SDS pH 7.0. **(A-F Top panels)** Dependence of HQS emission on the total concentration of $MgCl_2$ in the presence and absence of a Mg^{2+} chelator. Black lines represent a fit to SI equation 1 to determine the F_{max} , F_{min} , and K_{HQS} . **(A-F Bottom panels)** Fits used to determine binding constants K in Table 1 and SI table 3. Dependence of the free Mg^{2+} concentration on the total concentration of $MgCl_2$ in the presence and absence of a Mg^{2+} chelators. Black lines are to the equation $y=x$ and represent where the free Mg^{2+} concentration equals the total $MgCl_2$ concentration. Purple lines represent a fit to SI equation 4 to determine the association constant between HQS and a chelator. **(A)** 240 mM L-glutamate. **(B)** 194 mM Glutathione. **(C)** 240 mM L-aspartate. **(D)** 240 mM L-valine. **(E)** 240 mM L-glutamine. **(F)** 5 mM pyruvic acid.

Commented [BPC4]: "and are the same data in all panels" True?

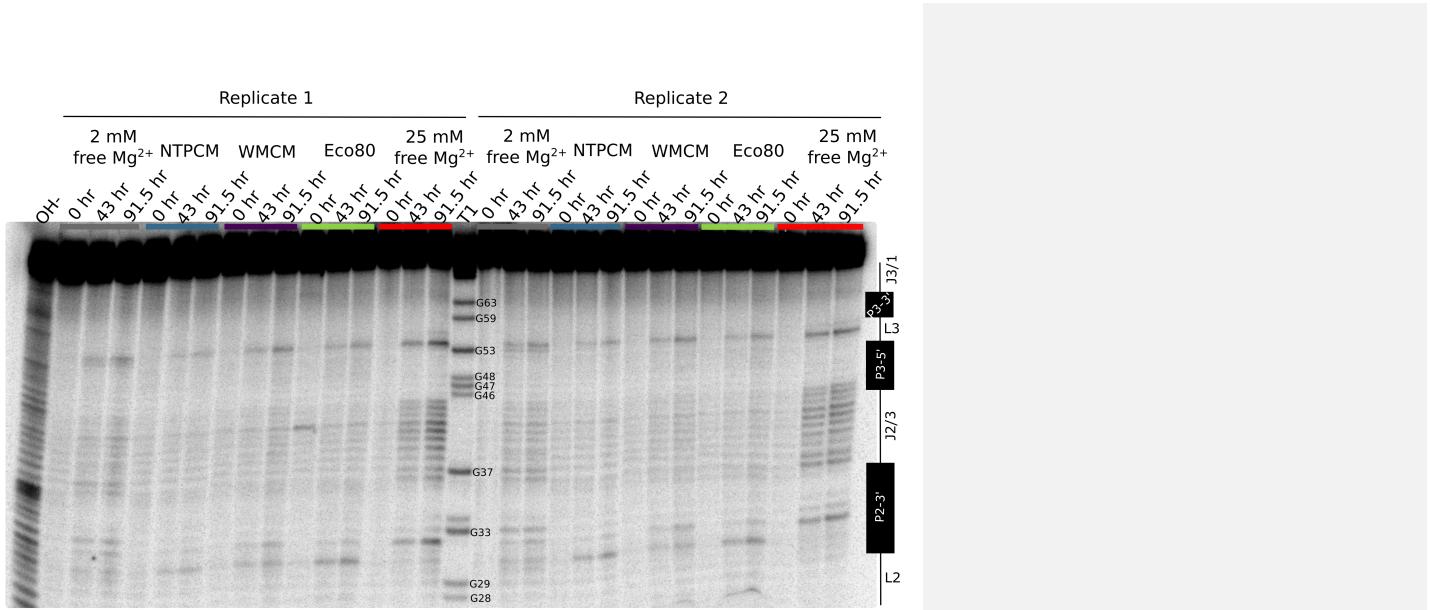
Commented [SJP5R4]: No, they are collected for every metabolite



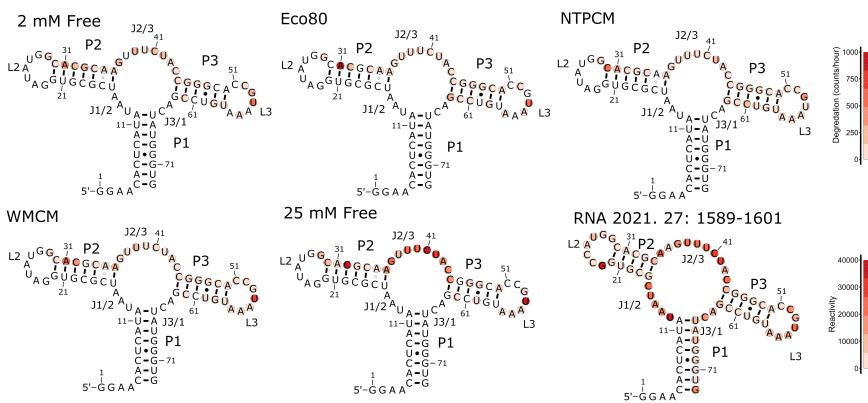
SI figure 3 Sample preparation for fluorescence binding isotherm analysis of RNA helix folding energies. Errors in the determination of concentrations of RNA stocks are systematically propagated during sample prep for fluorescence isotherm experiments. **(A)** FAM-RNA and RNA-BHQ1 stock concentrations are determined at low μM concentrations with UV-absorbance. In this work, FAM and BHQ1 stocks were prepared at the same concentration, 4 μM . **(B)** RNA-BHQ1 stocks are diluted to a 3x concentration from one stock. **(C)** FAM-RNA stocks were diluted to a 3/2x concentration into artificial cytoplasm to make a master mix (MM). **(D)** One volume of 3x RNA-BHQ1 stock is mixed with two volumes of 3/2x MM to prepare the final solution. Solutions appear layered in D to emphasize dilution, but they are mixed to form a homogeneous solution.



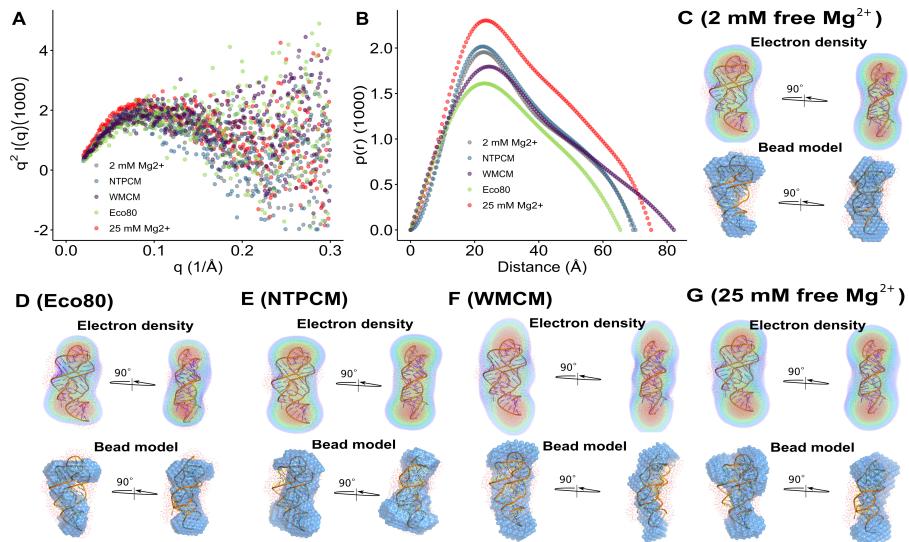
SI figure 4 Helix destabilization by artificial cytoplasm in comparison to the 2 mM free Mg^{2+} condition versus AU content.



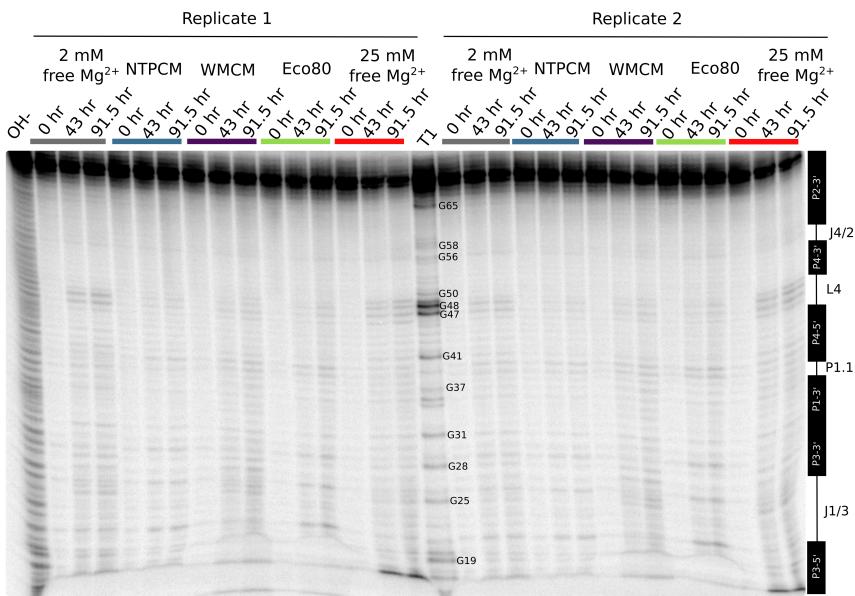
SI figure 5 Raw ILP gel image for the guanine riboswitch aptamer incubated in artificial cytoplasms at 37 °C and pH 7. The OH- lane contains a hydrolysis ladder which cleaves after every nucleotide and denaturing T1 contains the RNA treated with T1 ribonuclease which cleaves after every unfolded G residue. Enough Mg²⁺ was added to each artificial cytoplasm to have 2 mM free Mg²⁺ as determined in Table 2. Nucleotides 28 to 63 were quantified. Cleavage fragments shorter than 28 could not be quantified because of band smearing due to the high concentration of charged molecules in the samples and cleavage fragments longer than 63 could not be resolved.



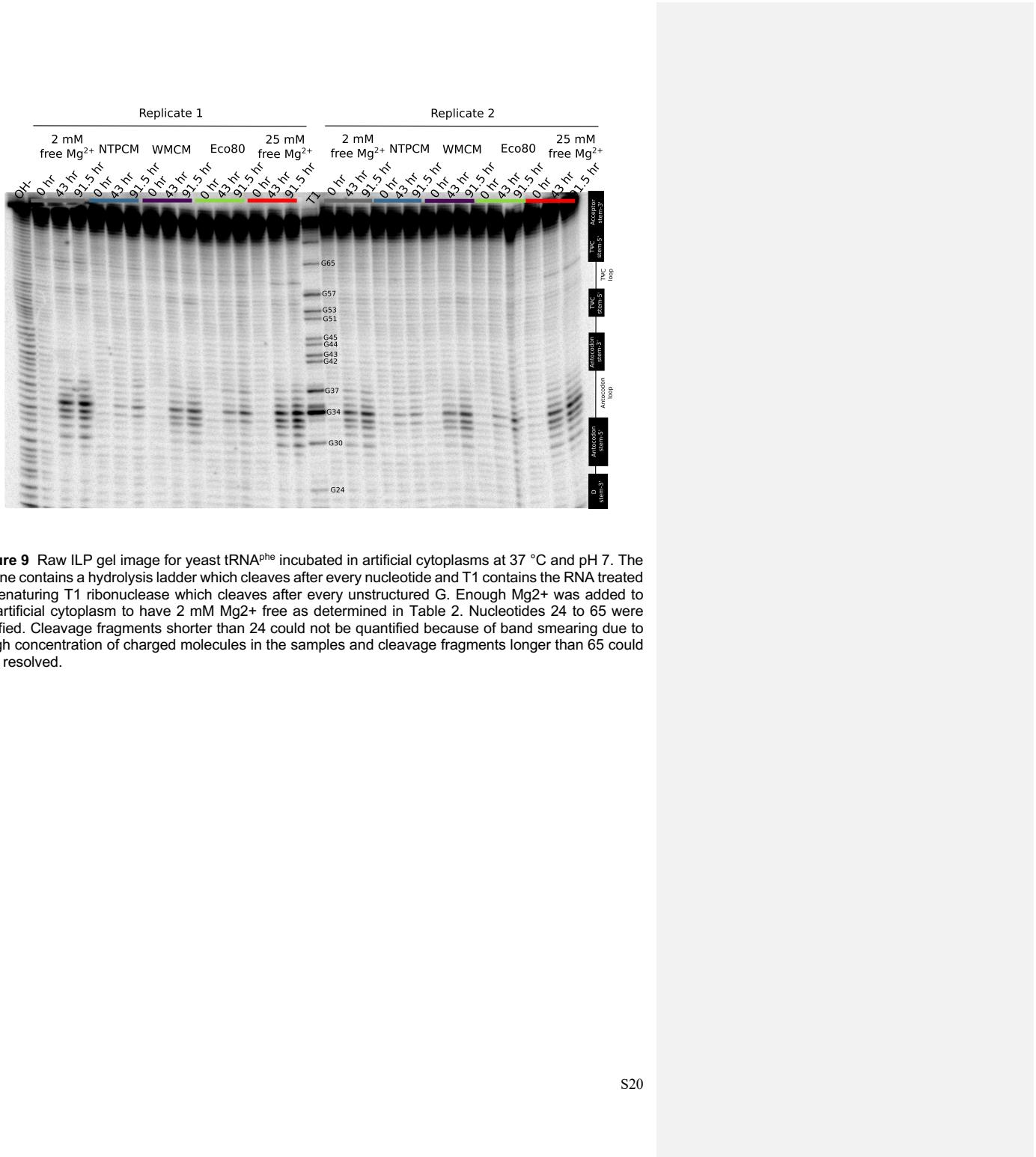
SI figure 6 In-line degradation rates mapped onto the secondary structure of the Guanine riboswitch aptamer.

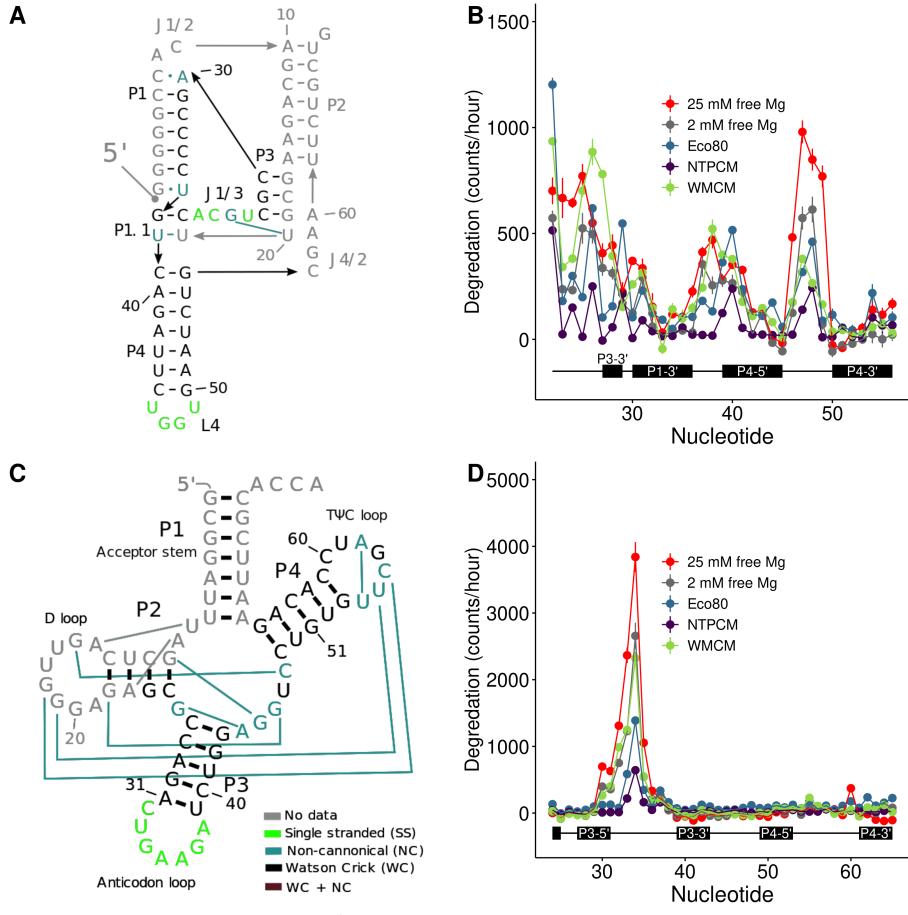


SI figure 7 Small angle x-ray scattering (SAXS) indicates that the guanine riboswitch aptamer adopts a similar structure in 2 mM free Mg²⁺, Eco80, NTPCM, WMCM and 25 mM free Mg²⁺. **(A)** Kratky scattering curves generated by the guanine aptamer in artificial cytoplasm. **(B)** P(r) plots generated with GNOM for the guanine riboswitch aptamer in artificial cytoplasm. The peak of the P(r) distribution is the radius of gyration (R_g). **(C-G)** Ab-initio shape reconstructions overlayed on the crystal structure of the guanine aptamer (PDB 4FE5) with a solvent shell modeled with WAXSiS. Electron density maps were generated with DENSS and bead models were generated with the ATSAS package.

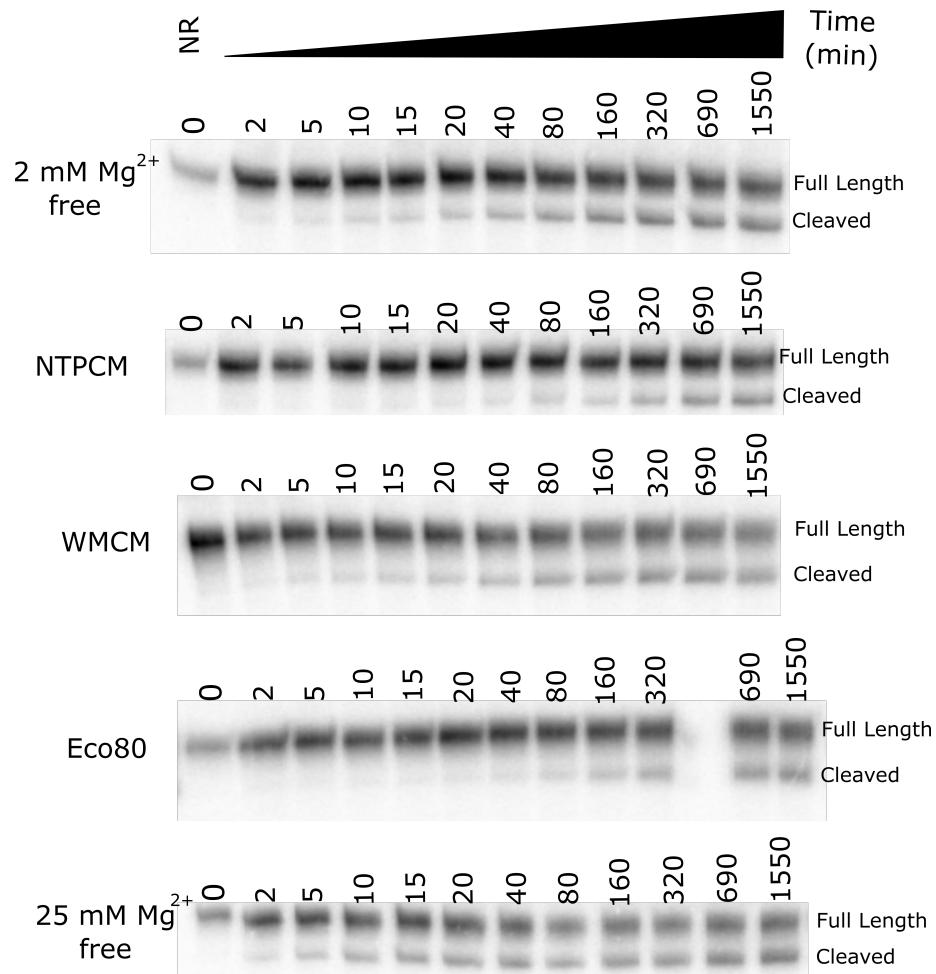


SI figure 8 Raw ILP gel image for the cleaved CPEB3 ribozyme incubated in artificial cytoplasmas at 37 °C and pH 7. The OH- lane contains a hydrolysis ladder which cleaves after every nucleotide and T1 contains the RNA treated with denaturing T1 ribonuclease which cleaves after every unstructured G. Enough Mg^{2+} was added to each artificial cytoplasm to have 2 mM Mg^{2+} free as determined in Table 2. Nucleotides 22 to 56 were quantified. Cleavage fragments shorter than 22 could not be quantified because of band smearing due to the high concentration of charged molecules in the samples and cleavage fragments longer than 56 could not be resolved.

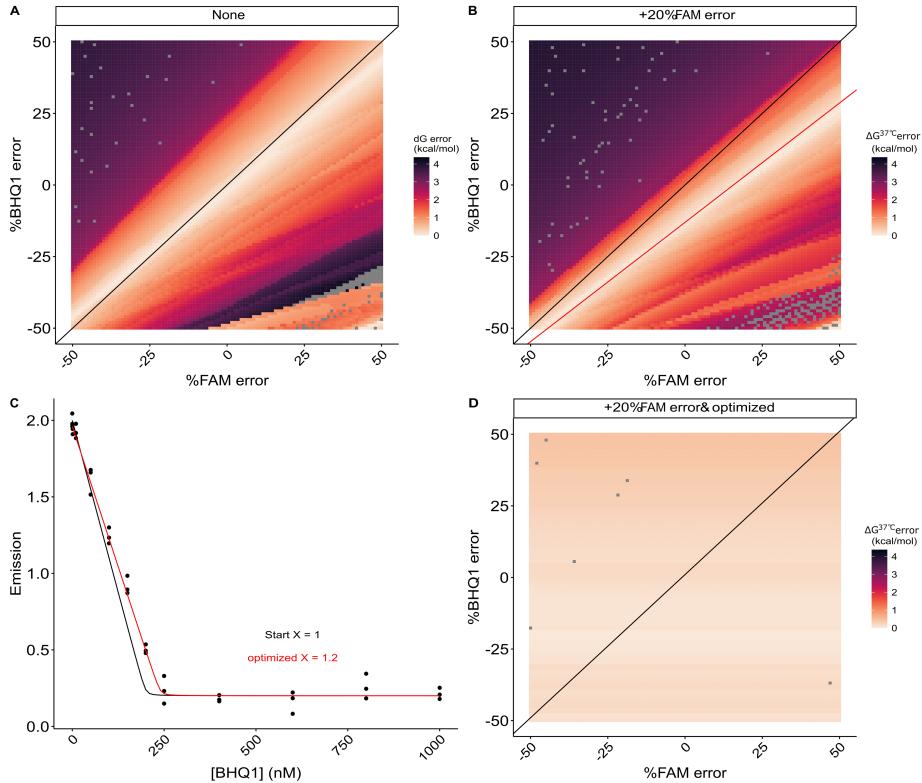




SI figure 10 *E. coli* metabolite and Mg^{2+} mixtures stabilize the chemical structure of RNA. **(A)** Secondary structure of the cleaved CPEB3 ribozyme with tertiary contacts represented by colored lines. **(B)** Degradation rate, by the increase in counts with time at each residue, in different solution conditions as a function of location in the cleaved CPEB3 ribozyme. Points represent the slope for the line of best fit of counts versus time and error bars represent the standard error in the slope estimated from the fit. **(C)** Secondary structure of yeast tRNA^{phe} with tertiary contacts represented by colored lines. **(D)** Degradation rate, by the increase in counts with time at each residue, in different solution conditions as a function of location in the tRNA^{phe}. Points represent the slope for the line of best fit of counts versus time and error bars represent the standard error in the slope estimated from the fit.

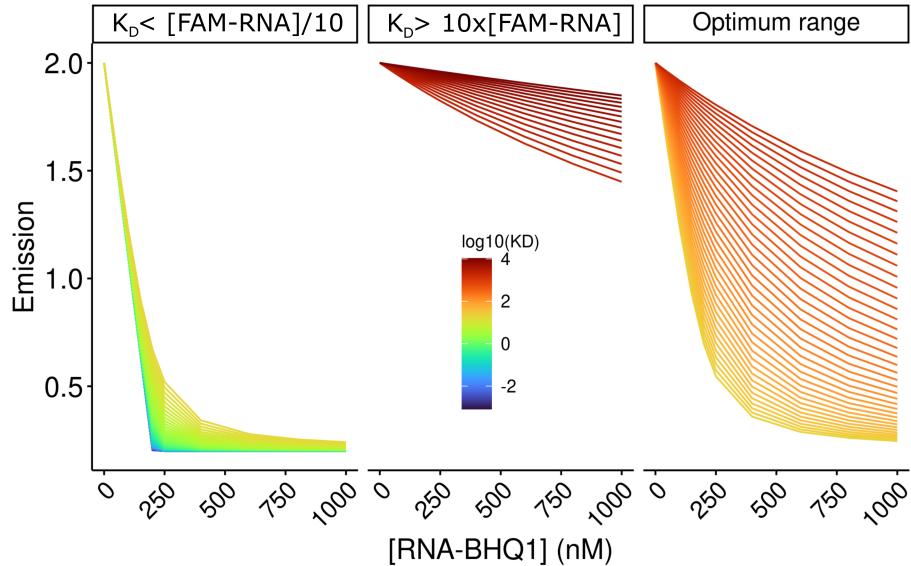


SI figure 11 CPEB3 ribozyme kinetics denaturing polyacrylamide gel images.



SI figure 12 The concentration optimization algorithm improves the accuracy of helix folding energies calculated with MeltR using data with inaccurate RNA concentration estimates. **(A)** Error in the Gibbs free energy ($\Delta G^{37^\circ\text{C}}$) calculated with MeltR on modeled data assuming perfectly accurate RNA concentration estimates is reduced when the %FAM-RNA concentration error and %RNA-BHQ1 concentration error compensates according to the black line, %FAM-RNA error = %RNA-BHQ1. Data were modeled assuming an association enthalpy (ΔH), entropy (ΔS), and Gibb's free energy at 37°C ($\Delta G^{37^\circ\text{C}}$) of -56.2 kcal/mol, -136.4 cal/mol/K, and -13.9 kcal/mol respectively, and 5% random fluorescence error. Concentration errors were seeded into the modeled data and the data was fit with MeltR. **(B)** Error in the Gibbs free energy (dG) calculated with MeltR on modeled data, assuming a +20% FAM-RNA concentration estimate error, is reduced when the %FAM-RNA concentration error and %RNA-BHQ1 concentration error compensate according to the red line, %RNA-BHQ1 = $X * \% \text{FAM-RNA error} + (100\% - 100\% * X) / X$, where X is the factor the MeltR concentration optimization algorithm multiplies the FAM-RNA concentration estimate by. Data was modeled assuming an association ΔH , ΔS , and $\Delta G^{37^\circ\text{C}}$ of -56.2 kcal/mol, -136.4 cal/mol/K, and -13.9 kcal/mol respectively, 5% random fluorescence error, and a +20% increase in the FAM-RNA concentration. Additional concentration errors were seeded into the modeled data and the data was fit with MeltR. **(C)** MeltR identifies the FAM-RNA concentration correction factor (X) using a low temperature fluorescence isotherm as a Job plot. Black data points represent modeled fluorescence data from B at 20°C . The black line represents the shape of the curve with a X of 1. The red line represents the shape of the curve with an optimized X of 1.2. **(D)** Error in the $\Delta G^{37^\circ\text{C}}$ calculated with MeltR using the concentration optimization algorithm and the data from B. On average, MeltR estimates the correct $\Delta G^{37^\circ\text{C}}$ within 0.2 kcal/mol, using the concentration optimization algorithm.

Commented [BPC6]: I want to go over this and the SI tables together



SI figure 13 K_D s calculated using MeltR are most accurate between the FAM-RNA concentration divided by 10 and 10 times the FAM-RNA concentration. At K_D s that are more than 10 fold lower than the FAM-RNA concentration, the shape of the curve is independent of the K_D . At K_D s more than 10 fold higher than the FAM-RNA concentration, there is very little dependence of FAM-RNA emission on RNA-BHQ1 concentration. However, the curve shape is highly dependent on K_D within the optimum range, so MeltR allows the user to specify an optimum K_D range to calculate helix folding energies.

Supplementary information tables

SI table 1 Recipe for a 2x stock solution of the Eco80 artificial cytoplasm. Prepared in a 10 mL volumetric flask.

Step 1 mass reagents into a 10 mL volumetric flask							
Reagent	Vendor (product #)	Purity	Molar mass (g/mL)	2x[Ligand] (mM)	Mass or volume	Na ⁺ added (mM)	K ⁺ added (mM)
Adenine triphosphate disodium salt hydrate	Sigma-Aldrich A2383	≥99%	551.14	19.26	525 µL of 367 mM	38.52	
Uridine triphosphate trisodium salt	Alpha-Azar J63427	≥98%	550.09	16.58	448 µL of 370 mM	49.74	
Guanosine triphosphate trisodium salt	Sigma-Aldrich G8877	≥95%	523.18	9.74	380 µL of 256 mM	29.22	
Deoxythymidine triphosphate disodium salt	Thermo R0171	≥99%	solution	9.24	924 µL of 100 mM	18.48	
L-glutamic acid potassium salt monohydrate	Millipore Sigma G1501	≥99%	203.23	192	0.3902 g		192
Glutathione, reduced, free acid	Cal biochem 3541	≥98%	307.32	33.2	0.102 g		
D-Fructose-1,6-bisphosphate trisodium salt hydrate	Sigma-Aldrich F6803	≥98%	406.06	30.4	217 µL of 1400 mM	91.2	
Uridine 5'-diphospho-N-acetylglucosamine disodium salt	Sigma-Aldrich U4375	≥98%	651.32	18.48	313 µL of 591 mM	36.96	
D-Glucose 6-phosphate dipotassium salt hydrate	Sigma-Aldrich G7375	≥98%	336.32	15.76	157 µL of 1001 mM		31.52
L-aspartic acid dipotassium salt	Millipore Sigma A9256	≥98%	133.10	8.46	66 µL of 1276 mM		16.92
L-valine monopotassium salt	Sigma Aldrich V0500	≥98%	117.15	8.04	80 µL of 1000 mM		16.08
L-glutamine monopotassium salt	Sigma G3126	≥99%	146.14	7.62	43 µL of 1759 mM		7.62
6-Phosphogluconic acid trisodium salt	Sigma-Aldrich P6888	≥95%	342.08	7.54	90 µL of 839 mM	22.62	
Sodium pyruvate	Sigma P8574	≥99%	110.04	7.32	121 µL of 606 mM		7.32
dihydroxyacetone phosphate hemi magnesium salt	Sigma 51296	≥95%	180.19	6.12	54 µL of 1128 mM		
MOPS	Sigma-Aldrich M3183	≥99.5	209.26	20	41.9 g		
EDTA	IBI scientific IB70182	≥99%	372.24	0.01	1 µL of 100 mM tetrasodium solution		0.04
SDS	JT Baker 4095-04	≥99%	288.38	0.035 (0.001%)	10 µL of 35 mM (1%)		0.035

Step 2 Add acid or base to pH 7.0

Reagent	Vendor (product #)	Purity	Molar mass (g/mL)	2x[Ligand] (mM)	Mass or volume	Na ⁺ added (mM)	K ⁺ added (mM)
Sodium hydroxide	Millipore Sigma Sx0593	≥97%	40.00	NA	100 µL of 10 M	100	

Step 3 Calculate how much Na⁺ and K⁺ was added

Sum	Na ⁺ (mM)	K ⁺ (mM)
Total monovalent added (mM)	394.135	264.14

Step 4 Add NaCl and KCl for a total final monovalent concentration of 480 mM Na⁺ and 280 mM K⁺

Reagent	Vendor (product #)	Purity	Molar mass (g/mL)	2x[Ligand] (mM)	Mass or volume	Na ⁺ added (mM)	K ⁺ added (mM)
Sodium chloride	Dot Scientific DSS23020	≥99%	58.4		171.7 µL of 5 M	85.865	
Potassium chloride	Millipore Sigma P9541	≥99%	74.55		79.3 µL of 2 M		15.86

Step 5 Fill to 10 mL in a volumetric flask and 2 µm filter	
Step 6 Check 1xConcentrated pH	
Final pH	6.99

SI table 2 Apparent binding constants determined with Isothermal titration calorimetry (ITC).

Metabolite	Syringe (mM)	Cell (mM)	ΔH (kcal/mol)	K' (M ⁻¹)	K _{D'} (mM ⁻¹)
ATP	15 mM MgCl ₂ ^a	0.1 mM ATP ^a	1.83 (0.04)	3600 (200)	0.28 (0.01)
UTP	15 mM MgCl ₂ ^a	0.2 mM UTP ^a	1.70 (0.01)	4200 (70)	0.248 (0.004)
GTP	15 mM MgCl ₂ ^a	0.2 mM GTP ^a	1.43 (0.02)	5000 (200)	0.201 (0.007)
dTTP	15 mM MgCl ₂ ^a	0.2 mM dTTP ^a	2.19 (0.02)	6300 (300)	0.160 (0.003)
Fructose 1,6-BP	100 mM MgCl ₂ ^a	5.0 mM Fructose 1,6-BP ^a	0.414 (0.004)	169 (4)	5.9 (0.1)
UDP-GlcNAC	100 mM MgCl ₂ ^a	5.0 mM UDP-GlcNAC	0.57 (0.02)	34 (2)	29 (2)
Glucose 6-P	100 mM MgCl ₂ ^a	5.0 mM Glucose 6-P ^a	0.555 (0.003)	57.9 (0.7)	17.3 (0.2)
6-P-gluconic acid	100 mM MgCl ₂ ^a	5.0 mM 6-P-gluconic acid ^a	0.662 (0.005)	70 (1)	14.4 (0.2)
Dihydroxyacetone phosphate	100 mM MgCl ₂ ^a	5.0 mM dihydroxy-acetone phosphate ^a	0.50 (0.01)	51 (3)	20 (1)
EDTA	6 mM MgCl ₂ ^a	1.5 mM EDTA 1.0 mM MgCl ₂ ^{a,b}	2.85 (0.04)	220,000 (30,000)	0.0045 (0.0006)

^a240 mM NaCl 140 mM KCl 10 mM HEPES pH 7.0 at 37 °C ^bMg²⁺ and EDTA were incorporated into the cell in order to sequester tight binding metal ions and thereby negate their contribution to ITC signal.

SI table 3 Apparent metabolite-Mg²⁺ binding constants determined with HQS emission. F_{max}, F_{min}, and K_{HQS} are determined by fitting HQS emission in the absence of chelators and used to calculate the free Mg²⁺ concentration in each sample. Metabolite binding constants, K' and K_{D'} are determined by fitting the relationship between the free Mg²⁺ concentration and the total MgCl₂ concentration using SI Equation 4.

Metabolite	F _{max}	F _{min}	K _{HQS} (mM ⁻¹)	K' (mM ⁻¹)	K _{D'} (mM ⁻¹)
L-Glutamic acid	187,000 (1000)	0 (810)	0.281 (0.008)	0.0019 (0.0002)	520 (50)
Glutathione	182,000 (1000)	592 (750)	0.279 (0.007)	NA ^c	NA ^c
L-Aspartic acid	196,000 (1000)	0 (820)	0.283 (0.007)	0.0021 (0.0001)	465 (12)
L-Valine	188,200 (800)	495 (580)	0.274 (0.005)	NA ^c	NA ^c
L-Glutamine	190,000 (1400)	516 (110)	0.27 (0.01)	NA ^c	NA ^c
Pyruvic acid	188,000 (1500)	0 (1300)	0.35 (0.01)	0.063 (0.003)	15.8 (0.9)

^cNo binding observed as per SI Figure 2

SI table 4 HQS fits in the absence of chelators, used to determine free Mg²⁺ concentrations in artificial cytoplasm. F_{max}, F_{min}, and K_{HQS} are determined by fitting HQS emission in the absence of chelators and used to calculate the free Mg²⁺ concentration in each sample.

Artificial cytoplasm	F _{max}	F _{min}	K _{HQS} (mM ⁻¹)
Eco80	185,100 (800)	124 (1000)	0.239 (0.005)
NTPCM	187,000 (1500)	436 (1000)	0.26 (0.01)
WMCM	179,000 (1600)	0 (1400)	0.32 (0.01)

SI table 5 Stability of RNA helices in *E. coli* metabolite mixtures. Helix energy was calculated by fitting raw fluorescence data with MeltR. Standard errors are estimated from fits. Extra significant digits are included to avoid propagating rounding errors.

Sequence ^a	Condition ^b	X ^c	Method 1 VH plot ΔH kcal/mol	Method 1 VH plot ΔS cal/mol/K	Method 1 VH plot ΔG kcal/mol	Method 2 Global fit ΔH kcal/mol	Method 2 Global fit ΔS cal/mol/K	Method 2 Global fit ΔG kcal/mol	%Diff. ^d ΔG	%Diff. ^d ΔS	%Diff. ^d ΔG
1: 5'CGGAUGGC3' 3'GCCUACCG5'	2 mM free	1.1	-71.1 (0.8)	-179 (2)	-15.6 (0.06)	-71.3 (15.6)	-179.6 (46.6)	-15.6 (1.1)	0.3%	0.3%	0%
	NTPCM	1.2	-70.4 (0.6)	-177 (2)	-15.28 (0.05)	-70.5 (15.1)	-178.0 (45.2)	-15.3 (1.0)	0.1%	0.6%	0.1%
	WMCM	1.1	-65.5 (2)	-162 (7)	-15.2 (0.2)	-65.3 (12.2)	-161.6 (36.4)	-15.2 (0.9)	0.3%	0.2%	0%
	Ecoli80	1.0	-69.7 (0.8)	-176 (3)	-15.0 (0.1)	-69.6 (11.4)	-176.3 (34.2)	-15.0 (0.8)	0.1%	0.2%	0.2%
2: 5'CCCAUCCU3' 3'GCGUAGGA5'	2 mM free	1.3	-55.9 (0.2)	-136.0 (0.7)	-13.82 (0.01)	-56.0 (8.9)	-135.9 (27.0)	-13.8 (0.5)	0.2%	0.1%	0.1
	NTPCM	2	-52.2 (0.4)	-125 (1)	-13.41 (0.02)	-52.4 (7.6)	-125.6 (23.2)	-13.4 (0.4)	0.4%	0.5%	0.1%
	WMCM	1.0	-61.4 (0.8)	-152 (2)	-14.22 (0.05)	-61.1 (14.0)	-151.2 (42.5)	-14.2 (0.9)	0.5%	0.5%	0.1%
	Ecoli80	1.5	-44.5 (0.7)	-102 (2)	-12.70 (0.04)	-44.4 (7.6)	-102.1 (23.3)	-12.7 (0.4)	0.2%	0.1%	0.0%
3: 5'CGUAUAGUA3' 3'GCAUACAU5'	2 mM free	0.8	-63.2 (0.9)	-169 (3)	-10.85 (0.02)	-62.3 (7.3)	-165.9 (23.0)	-10.8 (0.1)	1.4%	1.9%	0.5%
	NTPCM	0.9	-59 (1)	-157 (4)	-10.30 (0.01)	-58.5 (7.6)	-155.5 (24.3)	-10.3 (0.1)	0.9%	1.0%	0.0%
	WMCM	1.0	-67 (1)	-180 (3)	-10.85 (0.02)	-66.1 (10.5)	-178.1 (33.2)	-10.8 (0.2)	1.4%	1.1%	0.1%
	Ecoli80	1.0	-61 (1)	-164 (3)	-10.41 (0.01)	-60.9 (4.09)	-162.9 (13.1)	-10.4 (0.06)	0.2%	0.7%	0.2%
4: 5'CCAUAUCA3' 3'GGUUAUAGU5'	2 mM free	0.9	-53.4 (1.0)	-133 (3)	-12.02 (0.04)	-52.3 (14.9)	-129.9 (43.2)	-12.0 (0.5)	2.1%	2.4%	0.2%
	NTPCM	1.3	-42.9 (0.5)	-101 (1)	-11.50 (0.02)	-42.4 (20.4)	-99.6 (63)	-11.5 (0.8)	1.2%	1.4%	0.0%
	WMCM	1.0	-53 (2)	-132 (7)	-11.9 (0.1)	-51.6 (5.3)	-128.0 (1.6)	-11.9 (0.2)	2.7%	3.1%	0.0%
	Ecoli80	0.9	-57 (2)	-146 (5)	-11.38 (0.05)	-54.0 (13.1)	-137.7 (-41.1)	-11.3 (0.4)	5.4%	5.9%	0.7%
5: 5'CCAUAUUA3' 3'GGUAAA5'	2 mM free	0.9	-53.5 (0.4)	-137 (1)	-10.76 (0.01)	-53.2 (8.4)	-136.7 (27.1)	-10.8 (0.1)	0.6%	0.2%	0.4%
	NTPCM	1.0	-45.0 (0.2)	-112.5 (0.5)	-10.158 (0.002)	-45.0 (8.0)	-112.2 (25.6)	-10.2 (0.1)	0.0%	0.3%	0.4%
	WMCM	0.8	-43 (2)	-107 (5)	-9.94 (0.02)	-40.5 (9.3)	-98.4 (29.6)	-9.9 (0.1)	6.0%	8.4%	0.4%
	Ecoli80	1.2	-41.3 (0.2)	-100.4 (0.7)	-10.15 (0.01)	-41.2 (9.2)	-100.3 (29.4)	-10.2 (0.2)	0.2%	0.1%	0.5%
Average %error			1.7%	2.0%	0.3%	21.7%	26.5%	3.4%	1.2%	1.4%	0.2%

^aThe first sequence was 5'-FAM labeled and the second sequence was 3'-BHQ1 labeled. ^bAll solutions contain 2 mM Free Mg, 240 Na⁺ 140 mM K⁺. ^cConcentration optimization factor used to correct FAM-RNA concentrations by MeltR. ^dPercent difference between Method 1 and Method 2.

SI 6table 6 SAXS analysis

Condition	Guinier R_g^a (Å)	$P(r) R_g^a$ (Å)	D_{max}^a (Å)	Porod volume ^a (Å ³)
2 mM free	24.5 (0.2)	25.0 (0.2)	69.4	42200
Eco80	23.9 (0.4)	23.4 (0.3)	65.4	31897
NTPCM	24.5 (0.4)	26.0 (1)	70.15	52905
WMCM	26.9 (0.5)	27.2 (0.4)	82.1	40335
25 mM free	27.1 (0.2)	27.0 (0.2)	75.0	44299

^aDetermined with Primus in ATSAS⁷

Commented [NHY7]: Include Guinier $R(g)$ as from
ATSAS → Analysis → Distance distribution → Guinier Rg

SI table 7 Structural classification of each nucleotide by manually inspecting crystal structures.

Guanine riboswitch aptamer			CPEB3 ribozyme			tRNA ^{phe}					
N	Class ^a	PDB	N	Class ^a	PDB	N	Class ^a	PDB			
29	G	WC	PDB 4FE5: G38	22	C	WC	PDB 3NKB: C23	24	G	WC	PDB 1EHZ: G24
30	C	WC	PDB 4FE5: C39	23	A	SS	PDB 3NKB: U24	25	C	WC	PDB 1EHZ: C25
31	A	WC	PDB 4FE5: A40	24	C	SS	PDB 3NKB: C25	26	G	NC	PDB 1EHZ: G26
32	C	WC	PDB 4FE5: C41	25	G	NC	PDB 3NKB: G26	27	C	WC	PDB 1EHZ: C27
33	G	WC	PDB 4FE5: G42	26	U	SS	PDB 3NKB: C27	28	C	WC	PDB 1EHZ: C28
34	C	WC	PDB 4FE5: C43	27	C	WC	PDB 3NKB: G28	29	A	WC	PDB 1EHZ: A29
35	A	WC	PDB 4FE5: A44	28	G	WC	PDB 3NKB: G29	30	G	WC	PDB 1EHZ: G30
36	A	WC	PDB 4FE5: A45	29	C	WC	PDB 3NKB: C30	31	A	WC	PDB 1EHZ: A31
37	G	WC	PDB 4FE5: G46	30	A	NC	PDB 3NKB: G31	32	C	SS	PDB 1EHZ: C32
38	U	SS	PDB 4FE5: U47	31	G	WC	PDB 3NKB: C32	33	U	SS	PDB 1EHZ: U33
39	U	SS	PDB 4FE5: U48	32	C	WC	PDB 3NKB: A33	34	G	SS	PDB 1EHZ: G34
40	U	NC	PDB 4FE5: U49	33	C	WC	PDB 3NKB: A34	35	A	SS	PDB 1EHZ: A35
41	C	NC	PDB 4FE5: C50	34	C	WC	PDB 3NKB: G35	36	A	SS	PDB 1EHZ: A36
42	U	SS	PDB 4FE5: U51	35	C	WC	PDB 3NKB: C36	37	G	SS	PDB 1EHZ: G37
43	A	WC	PDB 4FE5: A52	36	U	NC	PDB 3NKB: U37	38	A	SS	PDB 1EHZ: A38
44	C	WC	PDB 4FE5: C53	37	G	WC	PDB 3NKB: G38	39	U	WC	PDB 1EHZ: U39
45	C	WC	PDB 4FE5: C54	38	U	NC	PDB 3NKB: G39	40	C	WC	PDB 1EHZ: C40
46	G	WC	PDB 4FE5: G55	39	C	WC	PDB 3NKB: A43	41	U	WC	PDB 1EHZ: U41
47	G	WC	PDB 4FE5: G56	40	A	WC	PDB 3NKB: C44	42	G	WC	PDB 1EHZ: G42
48	G	WC	PDB 4FE5: G57	41	G	WC	PDB 3NKB: A45	43	G	WC	PDB 1EHZ: G43
49	C	WC	PDB 4FE5: C58	42	A	WC	PDB 3NKB: U46	44	A	NC	PDB 1EHZ: A44
50	A	WC	PDB 4FE5: A59	43	U	WC	PDB 3NKB: U47	45	G	NC	PDB 1EHZ: G45
51	C	WC	PDB 4FE5: C60	44	U	WC	PDB 3NKB: C48	46	G	NC	PDB 1EHZ: G46
52	C	WC	PDB 4FE5: C61	45	C	WC	PDB 3NKB: C49	47	U	SS	PDB 1EHZ: U47
53	G	NC	PDB 4FE5: G62	46	U	SS	PDB 3NKB: G50	48	C	NC	PDB 1EHZ: C48
54	U	NC	PDB 4FE5: U63	47	G	SS	PDB 3NKB: A51	49	C	WC	PDB 1EHZ: C49
55	A	NC	PDB 4FE5: A64	48	G	SS	PDB 3NKB: A52	50	U	WC	PDB 1EHZ: U50
56	A	NC	PDB 4FE5: A65	49	U	SS	PDB 3NKB: A53	51	G	WC	PDB 1EHZ: G51
57	A	NC	PDB 4FE5: A66	50	G	WC	PDB 3NKB: G54	52	U	WC	PDB 1EHZ: U52
58	U	WC	PDB 4FE5: U67	51	A	WC	PDB 3NKB: G55	53	G	WC	PDB 1EHZ: G53
59	G	WC	PDB 4FE5: G68	52	A	WC	PDB 3NKB: A57	54	U	NC	PDB 1EHZ: U54
60	U	WC	PDB 4FE5: U69	53	U	WC	PDB 3NKB: A58	55	U	NC	PDB 1EHZ: U55
61	C	WC	PDB 4FE5: C70	54	C	WC	PDB 3NKB: U59	56	C	NC	PDB 1EHZ: C56
62	C	WC	PDB 4FE5: C71	55	U	WC	PDB 3NKB: G60	57	G	SS	PDB 1EHZ: G57
63	G	WC	PDB 4FE5: G72	56	G	WC	PDB 3NKB: G61	58	A	NC	PDB 1EHZ: A58
						59	U	SS	PDB 1EHZ: U59		
						60	C	SS	PDB 1EHZ: C60		
						61	C	WC	PDB 1EHZ: C61		
						62	A	WC	PDB 1EHZ: A62		
						63	C	WC	PDB 1EHZ: C63		
						64	A	WC	PDB 1EHZ: A64		
						65	G	WC	PDB 1EHZ: G65		

^aGroupings are based on analysis of crystal structures. SS: Single stranded, the base was not participating in hydrogen bonding interactions with other residues. NC: non-canonical, the base was forming non-canonical hydrogen bonding interactions in the tertiary structure. WC: Watson-Crick, the base was in a helix composed mostly of Watson-Crick base pairs.

SI table 8 DNA sequences used for RNA transcription

RNA	Template type	DNA sequences
Guanine riboswitch aptamer	Primer extended duplex	Forward primer: CGCGTAATACGACTCACTATAGGAACACTCATATAATCGCGTGGATATGGCAC-GCAAGTT Reverse primer: CACCCATAGTCGGACATTACGGTGCCCCGTAGAAACTT-GCGTGCATATCCACCGCATT
CPEB3 ribozyme	Primer extended duplex	Forward primer: GCGAAATTAAATACGACTCACTATAGGATCAAGGGGA-TAACAGGGGCCACAGCAGAAGCGTTCACGTCGCAGCC Reverse primer: CAGCAGAATTTCGAGAATTACCCAGAATCTGACAGGGCTGCGACGTGAACTCTCTG
tRNAPhe	Hemi-duplex	Forward primer: TAATACGACTCACTATA Template: TGGTGCGAATTCTGTGGATCGAACACAGGACCTCCAGATCTTCAG-TCTGGCGCTCTCCAACTGAGCT AAATCCGCTATAGTGAGTCGTATTA

SI table 9 Ab-initio bead model reconstruction from SAXS data

Condition	Ambiguity score	Normalized spatial discrepancy (NDS)	Models rejected	Clusters	Refined X ²
2 mM free	2.455	1.0 (0.3)	0/20	1	1.054
Eco80	2.948	1.0 (0.3)	0/20	1	1.115
NTPCM	2.318	0.7 (0.4)	0/20	1	0.99
WMCM	2.761	1.1 (0.3)	0/20	1	1.032
25 mM free	2.803	0.8 (0.3)	0/20	1	1.083
Ideal	< 2.5 (Preferable 1.5)	< 1.0	0/20	1	< 1.5

Supplementary References

- (1) Yamagami, R.; Sieg, J. P.; Bevilacqua, P. C. Functional Roles of Chelated Magnesium Ions in RNA Folding and Function. *Biochemistry* **2021**, *60* (31), 2374–2386.
<https://doi.org/10.1021/acs.biochem.1c00012>.
- (2) Xia, T.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson–Crick Base Pairs †. *Biochemistry* **1998**, *37* (42), 14719–14735.
<https://doi.org/10.1021/bi9809425>.
- (3) Leamy, K. A.; Yennawar, N. H.; Bevilacqua, P. C. Cooperative RNA Folding under Cellular Conditions Arises From Both Tertiary Structure Stabilization and Secondary Structure Destabilization. *Biochemistry* **2017**, *56* (27), 3422–3433.
<https://doi.org/10.1021/acs.biochem.7b00325>.
- (4) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in *Bacillus Subtilis* and Other Bacteria. *Cell* **2003**, *113* (5), 577–586. [https://doi.org/10.1016/S0092-8674\(03\)00391-X](https://doi.org/10.1016/S0092-8674(03)00391-X).
- (5) Yamagami, R.; Bingaman, J. L.; Frankel, E. A.; Bevilacqua, P. C. Cellular Conditions of Weakly Chelated Magnesium Ions Strongly Promote RNA Stability and Catalysis. *Nat. Commun.* **2018**, *9* (1), 2149. <https://doi.org/10.1038/s41467-018-04415-1>.
- (6) Laederach, A.; Das, R.; Vicens, Q.; Pearlman, S. M.; Brenowitz, M.; Herschlag, D.; Altman, R. B. Semiautomated and Rapid Quantification of Nucleic Acid Footprinting and Structure Mapping Experiments. *Nat. Protoc.* **2008**, *3* (9), 1395–1401. <https://doi.org/10.1038/nprot.2008.134>.
- (7) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2012**, *45* (Pt 2), 342–350.
<https://doi.org/10.1107/S0021889812007662>.
- (8) Grant, T. D. Ab Initio Electron Density Determination Directly from Solution Scattering Data. *Nat. Methods* **2018**, *15* (3), 191–193. <https://doi.org/10.1038/nmeth.4581>.
- (9) Batey, R. T.; Gilbert, S. D.; Montange, R. K. Structure of a Natural Guanine-Responsive Riboswitch Complexed with the Metabolite Hypoxanthine. *Nature* **2004**, *432* (7015), 411.
<https://doi.org/10.1038/nature03037>.
- (10) Knight, C. J.; Hub, J. S. WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Res.* **2015**, *43* (W1), W225–W230.
<https://doi.org/10.1093/nar/gkv309>.