SUPPLEMENTAL MATERIAL

Evaluation of folding midpoints and Hill coefficients for divalent ion titrations

Hydroxyl radical footprinting data *d* summed over a subset of nucleotides (e.g., 180, 181, & 182) were fit to Hill isotherms by least-squares optimization in KaleidaGraph:

$$d = A + B\frac{\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^n}{1+\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^n} \qquad \text{(S1)}$$

Each fit returned four parameters – the "unfolded" reactivity *A,* the maximal change in reactivity *B*, the midpoint $[M^{2+}]_{1/2}$, and the apparent Hill coefficient *n* – along with their associated errors δA, δB, $\delta[M^{2+}]_{1/2}$, and δn. Independently fitted values for multiple subsets of residues (e.g., 176-177 and 180-182) were averaged, with appropriate weighting and assuming normally distributed errors, to yield the final values of $[M^{2+}]_{1/2}$ and *n* for each individual folding transition. Explicitly, given midpoints $\left([M^{2+}]_{1/2}\right)_i \pm \delta\left([M^{2+}]_{1/2}\right)_i$ obtained for data over different residue subsets *i*, the averaged midpoints and associated errors were calculated as:

$$[M^{2+}]_{1/2} = \frac{\sum_i \frac{\left([M^{2+}]_{1/2}\right)_i}{\delta\left([M^{2+}]_{1/2}\right)_i^2}}{\sum_i \frac{1}{\delta\left([M^{2+}]_{1/2}\right)_i^2}} \qquad \text{(S2)}$$

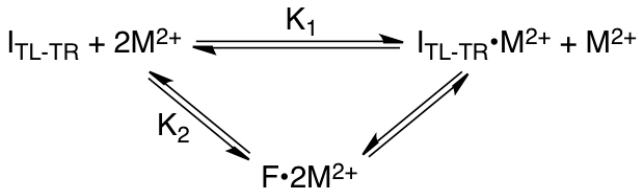
$$\delta[M^{2+}]_{1/2} = \left[\sum_i \frac{1}{\delta\left([M^{2+}]_{1/2}\right)_i^2}\right]^{-1/2} \qquad \text{(S3)}$$

An analogous relation was used to average the apparent Hill coefficients and obtain their associated errors.

Derivation of thermodynamic relations for models with non-constant Hill coefficients

Model 1. Two site-bound metal ions, derived from solution or from the atmosphere.

We derive the free energy relations and apparent Hill coefficients (Weiss 1997; Garcia et al. 2011) for the following model of the P4-P6 folding transition in a background of 2 M NaCl:



For simplicity, this scheme does not show additional diffusely associated 'background' divalent metal ions, which are assumed to be equal in number for each species. The scheme also does not show monovalent ions (Na^+ herein), since we are considering the dependence of folding on the divalent metal ion concentration only. Finally, the scheme neglects anions, since the divalent ion titrations minimally perturb the solution concentration of anions in the presence of 2 M NaCl. The Boltzmann weights for each state sum to the partition function for the system:

$$Z = P_1 + P_2 + P_3 = 1 + \frac{[M^{2+}]}{K_1} + \left(\frac{[M^{2+}]}{K_2}\right)^2 \qquad \text{(S4)}$$

Here, K_1 parameterizes the equilibrium between the first and second states (for associating a single divalent metal ion with the $I_{TL/TR}$ state), and K_2 parameterizes the equilibrium for folding between the first state and the third state. The free energy difference between the two $I_{TL/TR}$ states and F is then:

$$\Delta G = -RT \ln\left[\frac{P_3}{(P_1 + P_2)}\right] = -RT \ln\left[\frac{\left(\frac{[M^{2+}]}{K_2}\right)^2}{1 + \left(\frac{[M^{2+}]}{K_1}\right)}\right] \qquad \text{(S5)}$$

The apparent Hill coefficient is defined by the thermodynamic relation:

$$n = -\frac{1}{RT}\frac{\partial \Delta G}{\partial \ln[M^{2+}]} \qquad \text{(S6)}$$

Using equation S5 gives:

$$n = 2 - \frac{\left(\frac{[M^{2+}]}{K_1}\right)}{1+\left(\frac{[M^{2+}]}{K_1}\right)} \qquad \text{(S7)}$$

The fraction folded is given by:

$$f = \frac{P_3}{Z} = \frac{\left(\frac{[M^{2+}]}{K_2}\right)^2}{1+\left(\frac{[M^{2+}]}{K_1}\right)+\left(\frac{[M^{2+}]}{K_2}\right)^2} \qquad \text{(S8)}$$

At the folding midpoint $[M^{2+}]_{1/2}$, where by definition $\frac{P_3}{P_1+P_2}=1$, the following relation holds (setting equation S8 equal to 0.5):

$$K_2 = [M^{2+}]_{1/2}\left(1+\frac{[M^{2+}]_{1/2}}{K_1}\right)^{-1/2} \qquad \text{(S9)}$$

Equation S10 below can be derived by taking a Taylor expansion of $\ln[f/(1-f)]$ in equation S8 with respect to $\ln[M^{2+}]$ around the midpoint $[M^{2+}]_{1/2}$, and is a good approximation near the folding midpoint up to corrections logarithmic in $[M^{2+}]$.

$$f = \frac{\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^{n\left([M^{2+}]_{1/2}\right)}}{1+\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^{n\left([M^{2+}]_{1/2}\right)}} \qquad \text{(S10)}$$

Using $n\left([M^{2+}]_{1/2}\right)$ given by equation S7, we can evaluate equation S10 at the midpoint to give:

$$n\left([M^{2+}]_{1/2}\right) = 2 - \frac{[M^{2+}]_{1/2}}{K_1+[M^{2+}]_{1/2}} \qquad \text{(S11)}$$

Indeed, equation S10 is the standard Hill form used for least-squares fits (see equation S1) to estimate n_j and $\left([M^{2+}]_{1/2}\right)_j$, and associated errors δn_j and $\delta\left([M^{2+}]_{1/2}\right)_j$, for each RNA variant *j*. These values were then fit to equation S11 by χ^2 minimization (KaleidaGraph) to give the

parameter K_1 with standard errors δK_1, as shown in main text Figure 5. Finally, for comparing each modified variant to the wild type RNA, the midpoints were substituted into equation S5; applying equation S9 gives

$$\Delta\Delta G_j = 2RT \ln\left[\frac{\left([M^{2+}]_{1/2}\right)_j}{\left([M^{2+}]_{1/2}\right)_0}\right] - RT \ln\left[\frac{1+\frac{\left([M^{2+}]_{1/2}\right)_j}{K_1}}{1+\frac{\left([M^{2+}]_{1/2}\right)_0}{K_1}}\right] \qquad \text{(S12a)}$$

$$= 2RT \ln\left[\frac{\left(K_2\right)_j}{\left(K_2\right)_0}\right] \qquad \text{(S12b)}$$

where $\left([M^{2+}]_{1/2}\right)_0$ denotes the folding midpoint of the unmodified RNA, and $(K_2)_0$ and $(K_2)_j$ are the K_2 parameters fitted for the unmodified and variant RNAs, respectively. Note that, as desired, this expression is independent of the divalent ion concentration at which the free energy difference between the unmodified and modified RNAs is evaluated. The error on $\Delta\Delta G_j$ was dominated by uncertainty in K_1; it was evaluated by recalculating $\Delta\Delta G_j$ with $K_1 \pm \delta K_1$.

Model 2. Linear expansion of the apparent Hill coefficient.

This model makes no assumptions about the structural ensembles or their associated ion atmospheres and site-bound ions, although we continue to assume that the apparent Hill coefficient for all variants is the same at a given metal ion concentration $[M^{2+}]$ (but varies across metal ion concentrations). This model assumes that the apparent Hill coefficient can be expanded as a linear function of the logarithm of the divalent ion concentration, as is commonly seen in calculations as well as empirical fits to RNA folding isotherms. Explicitly, we write the expansion near the folding midpoint $\left(\left[M^{2+}\right]_{1/2}\right)_0$ of the unmodified RNA:

$$n_{\text{Hill}}\left(\left[M^{2+}\right]\right) = n_0 + \alpha_n \ln\left(\frac{\left[M^{2+}\right]}{\left(\left[M^{2+}\right]_{1/2}\right)_0}\right) \qquad \text{(S13)}$$

Here, n_0 is the apparent Hill coefficient of the unmodified RNA near its midpoint. For simplification, we let $\gamma = \frac{\left[M^{2+}\right]}{\left(\left[M^{2+}\right]_{1/2}\right)_0}$, the ratio of the metal ion concentration to the folding midpoint of the unmodified RNA. Then equation S13 becomes:

$$n_{\text{Hill}}\left(\left[M^{2+}\right]\right) = n_0 + \alpha_n \ln\gamma \tag{S14}$$

Integrating equation S6 gives the free energy for any variant *j* as:

$$\Delta G_j(\gamma) = -n_0 RT \ln\gamma - \frac{1}{2}\alpha_n RT(\ln\gamma)^2 + const \tag{S15}$$

The constant of integration (*const*) is readily determined by evaluating equation S15 at the folding midpoint of the unmodified RNA (γ = 1), where the free energy of a mutant is equal to $\Delta\Delta G_j$:

$$\Delta G_j(\gamma) = -n_0 RT \ln\gamma - \frac{1}{2}\alpha_n RT(\ln\gamma)^2 + \Delta\Delta G_j \tag{S16}$$

Finally, setting this expression equal to zero corresponds to evaluating $\Delta G_j(\gamma)$ at the folding midpoint $\left([M^{2+}]_{1/2}\right)_j$ of the modified variant *j*:

$$\Delta\Delta G_j = n_0 RT \ln\left[\frac{\left([M^{2+}]_{1/2}\right)_j}{\left([M^{2+}]_{1/2}\right)_0}\right]\left(n_0 + \frac{1}{2}\ln\left[\frac{\left([M^{2+}]_{1/2}\right)_j}{\left([M^{2+}]_{1/2}\right)_0}\right]\right) \tag{S17}$$

Noting that the apparent Hill coefficient for the mutant at its midpoint is $n_j = n_0 + \ln\left[\frac{\left([M^{2+}]_{1/2}\right)_j}{\left([M^{2+}]_{1/2}\right)_0}\right]$ gives a compact expression for the free energy difference $\Delta\Delta G_j$ associated with the modification:

$$\Delta\Delta G_j = \frac{n_0 + n_j}{2} RT \ln\left[\frac{\left([M^{2+}]_{1/2}\right)_j}{\left([M^{2+}]_{1/2}\right)_0}\right] \tag{S18}$$

In practice, we carried out independent least-squares Hill fits (see equation S1) to titrations for the unmodified and modified variants to yield n_j and $\left([M^{2+}]_{1/2}\right)_j$. Using equation S13, we then

carried out a least-squared fit to obtain n_0 and α_n, and their associated errors δn_0 and $\delta\alpha_n$. The midpoint values $\left([M^{2+}]_{1/2}\right)_j$ and the best-fit parameters for n_j (equation S13) were then substituted into equation S18 to give $\Delta\Delta G_j$ values. The error on $\Delta\Delta G_j$ was dominated by uncertainties in n_0 and α_n, and was estimated by recalculating $\Delta\Delta G_j$ with $n_0 \pm \delta n_0$ and $\alpha_n \pm \delta\alpha_n$, and summing the observed deviations in quadrature.

Calculation of P values

Values for $\Delta\Delta\Delta G$ were calculated by subtracting the $\Delta\Delta G$ obtained for Mn^{2+}-induced folding from that obtained for Mg^{2+}-induced folding. The associated *P* values were calculated using the Gauss error function (erf) in Microsoft Excel, according to

$$P = \frac{1 - \mathrm{erf}\left(\frac{\Delta\Delta\Delta G}{\delta\Delta\Delta\Delta G\sqrt{2}}\right)}{2} \qquad \text{(S19)}$$

where $\delta\Delta\Delta\Delta G$ is the standard error of $\Delta\Delta\Delta G$.

Complete likelihood-based inference of metal ion rescue from footprinting gels

The analysis given above permits the inference of $\Delta\Delta G$ and $\Delta\Delta\Delta G$ values, as well as their errors, for metal ion rescue experiments using common tools (Excel & KaleidaGraph), but makes assumptions about the Gaussianity of the errors to propagate the errors. Consequently, we have also carried out a complete likelihood-based analysis of the data without invoking such assumptions, using MATLAB scripts. This analysis gives consistent results. For completeness, we describe this analysis herein, presenting detailed equations for the high salt (2 M NaCl background) conditions.

Likelihood form

We make the following assumptions for computing the likelihood that the cleavage intensities read from a footprinting gel are consistent with a thermodynamic model:

1. The folding is a two-state process, as described, for example, in equation S8. The cleavage patterns of the unfolded and folded states are not known *a priori* but are to be estimated through the likelihood analysis.
2. The amount of sample loaded in each lane varies due to pipetting errors. The exact amount of this lane loading variation is a number to be estimated on a gel-by-gel basis.
3. Further deviations of peak intensities from the prediction are due to statistical errors (shot noise), random errors in the SAFA analysis procedure, nuclease contaminants, and variations in background cleavage. The latter issues in particular can affect residues in somewhat unpredictable ways. We assume that the range of deviations differs for each residue and needs to be estimated through the likelihood analysis. (For residues with large nuclease contaminants or background cleavage, we would hope that the analysis recognizes the large deviations from the model favored by the majority of residues and effectively downweights the contributions of anomalous residues.)
4. We assume that the error at each residue is *at least* 10% of the mean cleavage intensity. Allowing the likelihood fits to assume smaller errors appears to lead to overfitting, in which some residues are assigned too much confidence.

   Mathematically, the likelihood model is given by

$$L \propto \prod_{i,j} \frac{1}{\sigma_j} \exp\left[\frac{-\left(\alpha_i \left(C_j^{\text{unfold}} + \Delta C_j f_i^{\text{fold}} \right) - \text{DATA}_{ij} \right)^2}{2\sigma_j^2} \right] \prod_i \frac{1}{\sigma_\alpha} \exp\left[\frac{-(\alpha_i - 1)^2}{2\sigma_\alpha^2} \right] \prod_j \exp\left[-\frac{s_j^2}{2\sigma_j^2} \right] \qquad \text{(S20)}$$

Here, i is an index over lanes, and j is an index over residues. The parameters are as follows: α_i are the lane normalization parameters; σ_a is the lane loading error; C_j^{unfold} is the cleavage intensity of the unfolded state; ΔC_j is the change in cleavage intensity between the unfolded and folded states; σ_j are the errors associated with each residue; and s_j are the minimum errors assigned to each residue (set equal to 0.1 times the mean of DATA_{ij} across all lanes).

This model has been implemented in MATLAB. Given a prediction for the folding isotherm f_{fold}, maximum likelihood estimates for C_j^{unfold}, ΔC_j, α_i, σ_j, and s_a are determined by iteration. It is first assumed that $a_i = 0$, s_j = constant, and the maximum likelihood cleavage patterns of the unfolded and folded states are estimated. With these initial estimates, the s_j are updated, then the a_i, then the s_a; and the procedure is iterated five times (empirically, after three iterations, the procedure appears to converge within 0.1 log likelihood units).

For the full likelihood analysis, a range of K_1 and K_2 were explored, varying both parameters from 10^{-2} mM to 10^2 mM and computing the likelihood in equation S20. Contours of these likelihood fits are given in Figure S3; it is apparent that for each construct, a wide range of K_1 and K_2 are consistent with the data, though the two parameters are strongly correlated. In particular, for low K_1, a range of models with $(K_2)^2/K_1$ = *constant* gives similar predictions for the fraction folded (equation S8) with similar likelihoods.

Comparing likelihoods of thermodynamic models across different gels

One of the advantages of likelihood analysis is that replicates are not formally necessary to estimate errors like the σ_j of equation S20; these errors, and the corresponding likelihoods, are effectively read out for each gel from the intrinsic scatter of the data. Nevertheless, it is still important to compare the likelihood analyses made with independent gels for consistency. Figure S3 shows likelihood contours (with log-likelihood within 2 of the maximum likelihood, corresponding to approximately ~95% confidence) for the experiments presented in the main text, with replicates shown as differently colored contours. Replicates for each construct with each metal ion are in excellent agreement; their data can be combined (by multiplying likelihoods from replicates) with confidence and lead to tighter likelihood contours (black contours in Figure S3).

Estimating rescue factors

For the ΔG to be well-defined in Model 1 with respect to $[M^{2+}]$, the variation in free energy of the unfolded state needs to be the same across different constructs (but can be different for Mn^{2+} versus Mg^{2+}), and thus K_1 is a constant across constructs. The posterior distributions for the relevant K_1 parameter for the different constructs are given in Figure S7. These posterior distributions are obtained by integrating the likelihood estimates over $\ln(K_2)$. The distributions do not strongly constrain K_1 and are consistent with each other. Our best estimate for K_1 is given by the product of the posterior probabilities for the different constructs (black curve in Figure S4).

The posterior distributions of K_2 are similarly broad. However, to estimate $\Delta\Delta G$, we are interested in the *ratios* of K_2 (equation S12b) between the unmodified constructs and each phosphorothioate construct, which are better defined. Carrying out the integrations over $K_1(Mn^{2+})$ and over $K_1(Mg^{2+})$, and convolving the posterior probability distributions for $\Delta\Delta G$, gives the final distributions shown on the right hand side of Figure S5. We note that the final distributions are symmetric and indistinguishable from Gaussian curves, although this was not assumed in the analysis. Indeed, the underlying likelihood contours [Figures S3 and S5 (left)] are strikingly non-Gaussian.

Tables of the 95% confidence intervals for free energy values in equation S22 are given in Table S3. It is apparent that only the A184 R_P, A184 S_P, and A184 PS_2 constructs are consistent with metal ion rescue effects. The experiments for G188 R_P and G163 R_P do not give evidence for significant rescue. The data and their errors are in excellent agreement with Table 1 in the main text, which used simpler fits.

A fully analogous likelihood-based analysis was carried out to implement fits to Model 2 (linear expansion of the apparent Hill coefficient). Again, the resulting values and confidence

intervals (Table S4) are indistinguishable within error from the simpler fits in Table 4 in the main text.

SUPPLEMENTAL MATERIAL FIGURE LEGENDS

Figure S1: Strategies for constructing ΔC209 P4-P6 RNAs with site-specific phosphorothioate substitutions. We used HPLC to separate synthetic phosphorothioate oligonucleotide diastereomers (Hougland et al. 2005; Frederiksen and Piccirilli 2009) and incorporated them into full-length ΔC209 P4-P6 RNAs via successive splinted enzymatic ligations (Moore and Sharp 1992; Silverman and Cech 1999). **(A)** Splint ligation schemes for RNA construction. **(B)** Schematic of phosphorothioate linkages used in this work. **(C)** Typical anion exchange HPLC trace showing the separation between R_P and S_P single-phosphorothioate diastereomers.

Figure S2: A model using a constant Hill coefficient does not adequately fit the footprinting data. The plots show Mg^{2+}- and Mn^{2+}-dependent (open versus closed symbols) folding of unmodified ΔC209 P4-P6 (top) and the A184 S_P phosphorothioate variant in a background of 2 M NaCl. The data are fit to a Hill equation in which n_{Hill} either varies (solid lines) or is set equal to 2 (dashed lines).

Figure S3: Likelihood analyses from independent gels give consistent results. For each construct and metal ion, contours are shown marking values of K_1 and K_2 that have log-likelihood within 2 of the maximum likelihood point for different experiments. The black curve gives the analogous contour for the combined data, i.e., summing the likelihoods over the different experiments.

Figure S4: Posterior distributions over ln K_1 (the equilibrium constant for binding the "hidden" metal ion) do not exhibit a strong preference for particular values, and are consistent among different constructs. Black curves give final posterior distributions for $K_1(Mg^{2+})$ and $K_1(Mn^{2+})$

assuming that these values are the same across all constructs (obtained by multiplying posterior distributions for all constructs).

Figure S5: Estimating rescue factors for the different constructs. Left panels show log-likelihood contours (within 10 of maximum likelihood point) for the construct in Mn^{2+} (blue) and Mg^{2+} (red); and for unmodified ΔC209 in Mn^{2+} (light blue) and in Mg^{2+} (light red). Right panel shows posterior probability for rescue factor for K_2 (monitoring equilibrium constant between no metal ion and two metal ion state), after integration of the likelihood over ln K_1, and making the assumption that $K_1(Mg^{2+})$ and $K_1(Mn^{2+})$ are the same for all constructs, with distribution shown in Figure S4. Note that the rescue factor, when defined here in terms of ratios of K_2, is related to ΔΔΔG by eq. S12b.