PNAS
www.pnas.org

Supplementary Information for

Moderate activity of RNA chaperone maximizes the yield of self-spliced pre-RNA *in vivo*

Yonghyun Song, D. Thirumalai, Changbong Hyeon

E-mail:hyeoncb@kias.re.kr

This PDF file includes:

Figs. S1 to S3
References for SI reference citations

Contents

1. Effective rate constants for CYT-19 activity

We derive an expression of the Michaelis-Menten type for the effective rate constant, $k_{\mathrm{MI}}^{\mathrm{eff}}$, that quantifies the CYT-19-mediated unfolding of the misfolded (M state) ribozyme. The unfolding reaction can be broken down into a series of molecular events: (i) the binding of CYT-19 to the ribozyme, (ii) the binding of ATP to the catalytic site of CYT-19, and (iii) the unfolding of the M state ribozyme catalyzed by CYT-19. With $k_{\mathrm{off}}^{\mathrm{M}}$ and $k_{\mathrm{on}}^{\mathrm{M}}[\mathrm{C}]$ denoting the unbinding and binding rates of CYT-19 to the M state ribozyme, and $k_{\mathrm{off}}^{\mathrm{ATP}}$ and $k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{ATP}]$ denoting the unbinding and binding rates of ATP to the CYT-19-ribozyme complex, the reaction scheme of CYT-19-mediated unfolding reads,

$$\mathrm{M} \underset{k_{\mathrm{off}}^{\mathrm{M}}}{\overset{k_{\mathrm{on}}^{\mathrm{M}}[\mathrm{C}]}{\rightleftharpoons}} \mathrm{CM} \underset{k_{\mathrm{off}}^{\mathrm{ATP}}}{\overset{k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{ATP}]}{\rightleftharpoons}} \mathrm{TCM} \xrightarrow{k_{\mathrm{cat,MI}}} \mathrm{I} + \mathrm{C} + \mathrm{D}, \qquad [S1]$$

where T and D denote ATP and ADP, and CM and TCM are the CYT-19-ribozyme and ATP-CYT-19-ribozyme complexes, respectively. Based on the structural and biochemical studies, which suggested that the CYT-19 catalyzed unwinding of RNA is limited to short RNA duplexes, we assumed that the CYT-19 activity is non-processive (1, 2).

If the ribozyme molecules are initially in the M state (i.e. $[\mathrm{M}](t=0) = [\mathrm{R}]_{\mathrm{tot}}$), and the concentrations [C] and [ATP] remain effectively constant in time, the time evolution of the probability vector, $\vec{P}(t) = ([\mathrm{M}](t), [\mathrm{CM}](t), [\mathrm{TCM}](t))^T/[\mathrm{R}]_{\mathrm{tot}}$ is described by the differential equation with an initial condition $\vec{P}(0) = (1,0,0)^T$,

$$\frac{d}{dt}\vec{P}(t) = \mathcal{K}\vec{P}(t), \qquad [S2]$$

where

$$\mathcal{K} = \begin{bmatrix} -k_{\mathrm{on}}^{\mathrm{M}}[\mathrm{C}] & k_{\mathrm{off}}^{\mathrm{M}} & 0 \\ k_{\mathrm{on}}^{\mathrm{M}}[\mathrm{C}] & -k_{\mathrm{off}}^{\mathrm{M}} - k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{ATP}] & k_{\mathrm{off}}^{\mathrm{ATP}} \\ 0 & k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{ATP}] & -k_{\mathrm{off}}^{\mathrm{ATP}} - k_{\mathrm{cat,MI}} \end{bmatrix}. \qquad [S3]$$

The accumulation of the I state can be quantified using $\frac{d}{dt}[\mathrm{I}](t) = k_{\mathrm{cat,MI}}[\mathrm{TCM}](t)$, and the mean first passage time for the RNA molecule to reach the I state is given by $\langle\tau\rangle = -\mathcal{K}^{-1}\vec{P}(0)$ (see the SI Appendix **Mean first passage time**). The effective rate of CYT-19-mediated unfolding of M state ribozyme is then given by $k_{\mathrm{MI}}^{\mathrm{eff}} = \langle\tau\rangle^{-1}$.

Provided that the binding and unbinding of ATP, CYT-19, and the ribozyme are pre-equilibrated prior to the final catalytic reaction (i.e. $k_{\mathrm{cat,MI}} \ll 1$), we obtain the expression

$$k_{\mathrm{MI}}^{\mathrm{eff}} \approx k_{\mathrm{cat,MI}} f([\mathrm{ATP}](t), [\mathrm{C}](t); K_{\mathrm{ATP}}, K_{\mathrm{M}}), \qquad [S4]$$

with

$$f([\mathrm{ATP}](t), [\mathrm{C}](t); K_{\mathrm{ATP}}, K_{\mathrm{M}}) \equiv \frac{\frac{[\mathrm{ATP}](t)}{K_{\mathrm{ATP}}}\frac{[\mathrm{C}](t)}{K_{\mathrm{M}}}}{1 + \frac{[\mathrm{C}](t)}{K_{\mathrm{M}}} + \frac{[\mathrm{ATP}](t)}{K_{\mathrm{ATP}}}\frac{[\mathrm{C}](t)}{K_{\mathrm{M}}}}, \qquad [S5]$$

where $K_{\rm M} \equiv k^{\rm M}_{\rm off}/k^{\rm M}_{\rm on}$ is the dissociation constant of CYT-19 from the M state ribozyme, and $K_{\rm ATP} \equiv k^{\rm ATP}_{\rm off}/k^{\rm ATP}_{\rm on}$ is the dissociation constant of ATP from the CYT-19-ribozyme complex. Here, the binding and unbinding of the chaperone to RNA and ATP to chaperone are assumed to be pre-equilibrated relative to the time scale of the other dynamics (3).

We note that the limit of slow unbinding of CYT-19 from the ribozyme (i.e. $k^{\rm M}_{\rm off} \ll 1$) leads to the expression used by Chakrabarti *et. al.,* (4)

$$k^{\rm eff}_{\rm MI} \approx \frac{k^{\rm M}_{\rm on}[{\rm C}]\frac{k_{\rm cat,MI}[{\rm ATP}]}{K_M^{\rm ATP}+[{\rm ATP}]}}{k^{\rm M}_{\rm on}[{\rm C}] + \frac{k_{\rm cat,MI}[{\rm ATP}]}{K_M^{\rm ATP}+[{\rm ATP}]}}, \qquad [S6]$$

where $K_M^{\rm ATP} \equiv (k_{\rm cat,MI} + k^{\rm ATP}_{\rm off})/k^{\rm ATP}_{\rm on}$. The effective rates of unfolding of native ribozyme ($k^{\rm eff}_{\rm NI}$) and ATP hydrolysis ($k^{\rm eff}_{{\rm ATP},\alpha}$) are obtained based on the same argument used in deriving $k^{\rm eff}_{\rm MI}$.

Here, we the provide expressions for the effective ATP turnover rates, $k^{\rm eff}_{\rm ATP,M}(t)$, $k^{\rm eff}_{\rm ATP,N}(t)$, and $k^{\rm eff}_{\rm ATP,H}(t)$, in Eq. 3 of the main text. $k^{\rm eff}_{\rm ATP,M}(t)$ and $k^{\rm eff}_{\rm ATP,N}(t)$ are, respectively, the effective ATP turnover rates catalyzed by CYT-19 bound productively to the M and N states, and are given as

$$k^{\rm eff}_{{\rm ATP},\alpha}(t) = k_{\rm cat,P} f([{\rm ATP}](t), [{\rm C}](t); K_{\rm ATP}, K_\alpha), \qquad [S7]$$

with α = M or N, where $k_{\rm cat,P}$ denotes the ATP hydrolysis rate of CYT-19 bound productively to the ribozyme. CYT-19 bound to the exposed helices consumes ATP with the effective rate constant $k^{\rm eff}_{\rm ATP,H}(t)$

$$k^{\rm eff}_{\rm ATP,H}(t) = k_{\rm cat,H} f([{\rm ATP}](t), [{\rm C}](t); K_{\rm ATP}, K_{\rm H}), \qquad [S8]$$

where $k_{\rm cat,H}$ is the ATP hydrolysis rate of the exposed helix-bound CYT-19. The dissociation constant of CYT-19 from the helices is given by $K_{\rm H}$. The mathematical form of $f([{\rm ATP}](t), [{\rm C}](t); K_{\rm ATP}, K_\alpha)$ with α = M, N, and H is given in Eq. S5, except that the arguments are different.

2. Mass balance of CYT-19

Unlike the total chaperone concentration, $[{\rm C}]_{\rm tot}$, that stays constant in time, the concentration of free chaperone, [C](t), varies with time. Specifically, we calculate [C](t) at each time point t by the mass balance relation,

$$\begin{aligned}[{\rm C}]_{\rm tot} &= [{\rm C}](t) + [{\rm CM}](t) + [{\rm TCM}](t) + [{\rm CN}](t) + [{\rm TCN}](t) + [{\rm CH}](t) + [{\rm TCH}](t) \\ &= [{\rm C}](t)\left\{1 + \left(1 + \frac{[{\rm ATP}](t)}{K_{\rm ATP}}\right)\left(\frac{[{\rm M}](t)}{K_{\rm M}} + \frac{[{\rm N}](t)}{K_{\rm N}} + \frac{[{\rm H}](t)}{K_{\rm H}}\right)\right\}, \end{aligned} \qquad [S9]$$

where $[{\rm CN}](t)$ and $[{\rm CH}](t)$ are the concentrations of CYT-19 complexed with the N state and the exposed helices, respectively. $[{\rm TCN}](t)$ and $[{\rm TCH}](t)$ are the concentrations of complexes where ATP molecules are bound. $[{\rm M}](t)$, $[{\rm N}](t)$, and $[{\rm H}](t)$ denote the concentrations of the *free* ribozymes in M and N states and *free* exposed helices without being complexed to chaperones. $K_{\rm N}$(= [C][N]/[CN]), $K_{\rm M}$(= [C][M]/[CM]), and $K_{\rm H}$(= [C][H]/[CH]) denote the respective dissociation constants from CYT-19. By expressing the free ribozyme concentration in each state in terms of $[{\rm C}](t)$,

$$\begin{aligned}[{\rm M}](t) &= [{\rm M}]_{\rm tot}\left(1 + \frac{[{\rm C}](t)}{K_{\rm M}} + \frac{[{\rm ATP}](t)}{K_{\rm ATP}}\frac{[{\rm C}](t)}{K_{\rm M}}\right)^{-1} \\ [{\rm N}](t) &= [{\rm N}]_{\rm tot}\left(1 + \frac{[{\rm C}](t)}{K_{\rm N}} + \frac{[{\rm ATP}](t)}{K_{\rm ATP}}\frac{[{\rm C}](t)}{K_{\rm N}}\right)^{-1} \\ [{\rm H}](t) &= [{\rm H}]_{\rm tot}\left(1 + \frac{[{\rm C}](t)}{K_{\rm H}} + \frac{[{\rm ATP}](t)}{K_{\rm ATP}}\frac{[{\rm C}](t)}{K_{\rm H}}\right)^{-1}, \end{aligned} \qquad [S10]$$

we solve for $[{\rm C}](t)$ in Eq. S9. For the model of CYT-19 activity on pre-RNA, the mass balance relation is given by

$$[{\rm C}]_{\rm tot} = [{\rm C}](t)\left\{1 + \left(1 + \frac{[{\rm ATP}](t)}{K_{\rm ATP}}\right)\left(\frac{[{\rm M_1}](t)}{K_{\rm M_1}} + \frac{[{\rm M_2}](t)}{K_{\rm M_2}} + \frac{[{\rm N}](t)}{K_{\rm N}}\right)\right\}, \qquad [S11]$$

where $[{\rm M_1}]$, $[{\rm M_2}]$ and [N] are the concentrations of free ribozyme in the M_1, M_2, N states, and $K_{\rm M_1}$, $K_{\rm M_2}$ and $K_{\rm N}$ denote the respective dissociation constants from CYT-19. Similarly to Eq. S10, the free ribozyme concentration in each state can be

expressed as

$$[\mathrm{M_1}](t) = [\mathrm{M_1}]_{\mathrm{tot}}\left(1 + \frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{M_1}}} + \frac{[\mathrm{ATP}]_{\mathrm{tot}}}{K_{\mathrm{ATP}}}\frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{M_1}}}\right)^{-1}$$
$$[\mathrm{M_2}](t) = [\mathrm{M_2}]_{\mathrm{tot}}\left(1 + \frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{M_2}}} + \frac{[\mathrm{ATP}]_{\mathrm{tot}}}{K_{\mathrm{ATP}}}\frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{M_2}}}\right)^{-1}$$
$$[\mathrm{N}](t) = [\mathrm{N}]_{\mathrm{tot}}\left(1 + \frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{N}}} + \frac{[\mathrm{ATP}]_{\mathrm{tot}}}{K_{\mathrm{ATP}}}\frac{[\mathrm{C}](\mathrm{t})}{K_{\mathrm{N}}}\right)^{-1}, \quad [S12]$$

where we keep the ATP concentration, $[\mathrm{ATP}]_{\mathrm{tot}}$, constant in time.

The IAM-based model in Ref. (4) was developed under the assumption that $[\mathrm{C}](\mathrm{t}) \approx [\mathrm{C}]_{\mathrm{tot}}$. This was feasible in part because the model was fit to the dynamics of RNA folding measured by Bhaskaran *et. al.* (5), in which the experiments were mostly performed with a high CYT-19 concentration relative to the ribozyme concentration; $[\mathrm{R}]_{\mathrm{tot}} = 200$ nM and $[\mathrm{C}]_{\mathrm{tot}}$ ranged from 0 to 3 μM. However, in more recent works by Jarmoskaite *et. al.* (6, 7), the folding of ribozyme was mostly measured in conditions where $[\mathrm{C}]_{\mathrm{tot}}$ was similar to, or smaller than, $[\mathrm{R}]_{\mathrm{tot}}$. Thus, to address all the available data on the folding of ribozyme from the three experimental studies (5–7), we take into account the mass balance relation in Eq. S9.

3. Mean first passage time

We provide additional context to the thermodynamic cost to refold a misfolded ribozyme to the native state, $\langle A_{\mathrm{M}\to\mathrm{N}}\rangle$ (Eq. 9 in the main text). The cost associated with refolding the native ribozyme into the misfolded state, $\langle A_{\mathrm{N}\to\mathrm{M}}\rangle$, can be obtained analogously. In order to evaluate the cost of forming the N state from the M state, we set the transition rates out of the N state to 0 (i.e. $k_{\mathrm{NM}} = k_{\mathrm{cat,NI}} = 0$ in Eq. 4). Furthermore, we assume that CYT-19 does not bind productively to the N state (i.e. $[\mathrm{C}](t)/K_{\mathrm{N}} \ll 1$). Then, the relative concentrations of ribozyme in the I and M states, $P_{\mathrm{I}}(t) = [\mathrm{I}](t)/[\mathrm{R}]_{\mathrm{tot}}$ and $P_{\mathrm{M}}(t) = [\mathrm{M}](t)/[\mathrm{R}]_{\mathrm{tot}}$, evolve in time, following the equation,

$$\frac{d}{dt}\begin{bmatrix} P_{\mathrm{I}}(t) \\ P_{\mathrm{M}}(t)\end{bmatrix} = \begin{bmatrix} -k_{\mathrm{IM}}-k_{\mathrm{IN}} & k_{\mathrm{MI}}^{\mathrm{eff}} \\ k_{\mathrm{IM}} & -k_{\mathrm{MI}}^{\mathrm{eff}}-k_{\mathrm{MN}}\end{bmatrix}\begin{bmatrix} P_{\mathrm{I}}(t) \\ P_{\mathrm{M}}(t)\end{bmatrix} = \mathcal{K}\begin{bmatrix} P_{\mathrm{I}}(t) \\ P_{\mathrm{M}}(t)\end{bmatrix}, \quad [S13]$$

where $\mathcal{K}$ is a 2×2 matrix, and the chaperone-mediated unfolding rate, $k_{\mathrm{MI}}^{\mathrm{eff}}$, is assumed to be independent of time. The ribozyme eventually transitions to the N state with the reaction rates k_{IN} and k_{MN}, so that the relative concentration of the N state, $P_{\mathrm{N}}(t) = [\mathrm{N}](t)/[\mathrm{R}]_{\mathrm{tot}}$, increases with rate $dP_{\mathrm{N}}(t)/dt = k_{\mathrm{IN}}P_{\mathrm{I}}(t) + k_{\mathrm{MN}}P_{\mathrm{M}}(t)$. With the initial condition $P_{\mathrm{M}}(0) = 1$, $P_{\mathrm{I}}(0) = P_{\mathrm{N}}(0) = 0$, the mean first passage time (MFPT) for the ribozyme to transition from the M state to the N state can be written as $\langle\tau\rangle = \int_0^\infty \left(t\frac{dP_{\mathrm{N}}(t)}{dt}\right)dt$, where $(dP_{\mathrm{N}}(t)/dt)\,|_{t=\tau}$ is the probability that the ribozyme transitions into the N state at time τ. By using the equality $P_{\mathrm{N}}(t) = 1 - P_{\mathrm{M}}(t) - P_{\mathrm{I}}(t)$ and integration by parts,

$$\begin{aligned}\langle\tau\rangle &= -\int_0^\infty \left[t\frac{d}{dt}\left(P_{\mathrm{M}}(t) + P_{\mathrm{I}}(t)\right)\right]dt \\ &= -t\left(P_{\mathrm{M}}(t) + P_{\mathrm{I}}(t)\right)\Big|_0^\infty + \int_0^\infty \left(P_{\mathrm{M}}(t) + P_{\mathrm{I}}(t)\right)dt \\ &= \int_0^\infty \left(P_{\mathrm{M}}(t) + P_{\mathrm{I}}(t)\right)dt. \end{aligned} \quad [S14]$$

The inverse of MFPT, $\langle\tau\rangle^{-1}$, is the rate of N state formation from the M state.

For a given duration T, $\frac{1}{T}\int_0^T P_{\mathrm{I}}(t)dt$, $\frac{1}{T}\int_0^T P_{\mathrm{M}}(t)dt$, and $\frac{1}{T}\int_0^T P_{\mathrm{N}}(t)dt$ are the time-averaged probabilities for the ribozyme to be found in the I, M and N states, respectively. Thus, the expression $\lim_{T\to\infty}\left(T^{-1}\int_0^T P_{\mathrm{M}}(t)dt\right)\Big/\left(T^{-1}\int_0^T P_{\mathrm{I}}(t)dt\right)$ quantifies the ratio between the time-averaged probabilities of finding the ribozyme in the M versus I state. In other words, $\langle\tau\rangle$ can be decomposed into two parts, $\langle\tau_{\mathrm{M}}\rangle = \int_0^\infty P_{\mathrm{M}}(t)dt$ and $\langle\tau_{\mathrm{I}}\rangle = \int_0^\infty P_{\mathrm{I}}(t)dt$, that represent the mean survival times in the M and I states before reaching the N state. With an assumption that only CYT-19 that is bound to the M state hydrolyzes ATP with the time-independent effective rate constant $k_{\mathrm{ATP,M}}^{\mathrm{eff}}$, the average ATP cost per refolded ribozyme is given by

$$\langle A_{\mathrm{M}\to\mathrm{N}}\rangle = k_{\mathrm{ATP,M}}^{\mathrm{eff}}\langle\tau_{\mathrm{M}}\rangle, \quad [S15]$$

which is equivalent to Eq. 9 in the main text.

To obtain the expression for $\langle \tau_M \rangle$, we first define the eigenvalues λ_1 and λ_2 along with the corresponding eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ of $\mathcal{K}$ in Eq. S13, so that $\lambda_1 \mathbf{v}_1 = \mathcal{K}\mathbf{v}_1$ and $\lambda_2 \mathbf{v}_2 = \mathcal{K}\mathbf{v}_2$. Next, we define the matrix of eigenvectors $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$. Then, the solution to Eq. S13 can be written as

$$\begin{bmatrix} P_I(t) \\ P_M(t) \end{bmatrix} = e^{t\mathcal{K}} \begin{bmatrix} P_I(0) \\ P_M(0) \end{bmatrix} = \mathbf{V} e^{t\Lambda} \mathbf{V}^{-1} \begin{bmatrix} P_I(0) \\ P_M(0) \end{bmatrix}, \tag{S16}$$

where $\Lambda \equiv \begin{bmatrix} \lambda_1, & 0 \\ 0, & \lambda_2 \end{bmatrix}$. Assuming that $\lambda_1, \lambda_2 < 0$, evaluating the integral leads to

$$\begin{aligned} \begin{bmatrix} \langle \tau_I \rangle \\ \langle \tau_M \rangle \end{bmatrix} = \int_0^\infty \begin{bmatrix} P_I(t) \\ P_M(t) \end{bmatrix} dt &= \int_0^\infty \mathbf{V} e^{t\Lambda} \mathbf{V}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} dt \\ &= -\mathbf{V} \Lambda^{-1} \mathbf{V}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= -\mathcal{K}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{k_{IN} k_{MI}^{eff} + k_{MN}(k_{IM} + k_{IN})} \begin{bmatrix} k_{MI}^{eff} \\ k_{IM} + k_{IN} \end{bmatrix}. \end{aligned} \tag{S17}$$

Thus, $\langle A_{M\to N} \rangle = k_{ATP,M}^{eff} \langle \tau_M \rangle = A_{M\to I}/(\Phi + k_{MN}/k_{MI}^{eff})$, where $\Phi \equiv k_{IN}/(k_{IN} + k_{IM})$ and $A_{M\to I} \equiv k_{ATP,M}^{eff}/k_{MI}^{eff}$.

4. Model fit to data

We provide more detail on how the theoretical model (Eq. 4) is fit to the experimental measurements from Refs. (5–7). The quality of fit of the 10-dimensional array of parameters, $\vec{\theta} \equiv [k_{cat,MI}, k_{cat,NI}, k_{cat,H}, k_{cat,P}, k_{MN}, k_{IM}, K_H, K_M, K_N, K_{ATP}]^T$ is evaluated by the corresponding χ^2 value,

$$\chi^2(\vec{\theta}) \equiv \sum_i \left(\frac{\mu_i^{model}(\vec{\theta}) - \mu_i}{\sigma_i} \right)^2, \tag{S18}$$

where i runs through each measured data point with mean μ_i and standard deviation σ_i. $\mu_i^{model}(\vec{\theta})$ is the simulated data point that matches μ_i. The remaining parameters, k_{IN} and k_{NM}, are set by the constraints $\Phi = 0.1 = k_{IN}/(k_{IN} + k_{IM})$ and $k_{MN}/k_{NM} = 10$.

To fit the initial rate constants reported in Refs. (6, 7), we consider the initial condition with [M](0) = 200 nM, [I](0) = [N](0) = 0 nM, $[C]_{tot}$ = 1 μM and $[ATP]_{tot}$ = 0.5 mM. First, Eq. 4 is numerically solved while assuming that ATP concentration remains constant in time. The proxy for the time required to reach steady state, t^{ss}, is defined as the time point at which the inequality $\{d \ln([N](t))/dt\}|_{t=t^{ss}} < 0.0025$ holds, where t is in minutes. Then, Eq. 4 is solved a second time from the initial condition until time t^{ss}. For this second run, ATP concentration, [ATP](t), is also modeled as a time dependent variable. The initial rate of N state formation is evaluated at time t_{init}, defined by $[N](t_{init}) = [N](t^{ss})/3$. Thus, the initial rate of N state formation is $[N](t_{init})/t_{init}$, and the initial rate of ATP consumption is $([ATP](0) - [ATP](t_{init}))/t_{init}$. For the experiments in which all the ribozyme is initially in the N state, t_{init} is analogously defined as $([N](t_{init}) - [N](0)) = ([N](t^{ss}) - [N](0))/3$.

For most of the measured initial rate constants, both the mean (μ) and standard deviation (σ) were reported. For the initial rate constants reported without replicate experiments, the standard deviation is set so that the ratio $\alpha = \mu/\sigma$ is equal to twice the average α value computed from the rest of the data points belonging to the same family of experiments. For the time course data, the numerical solutions are directly compared with the measurements of the fraction of N state from Ref. (5) (Fig. S1A). No replicate experiments were reported for the time course data, and the standard deviations are set to one-tenth the mean value, so that $\sigma = \mu/10$ (see additional Dataset (SIdata.pdf) for all values used).

5. Markov Chain Monte Carlo (MCMC) sampling

We performed MCMC sampling to characterize the distribution of parameters that yield a good fit. To generate the starting point for the MCMC algorithm, each element of the parameter vector is uniformly sampled from the log space with the following

bounds: 10^{-2} min^{-1} $< k_{\mathrm{cat,MI}}, k_{\mathrm{cat,NI}} < 10^{2}$ min^{-1}, 10^{0} min^{-1} $< k_{\mathrm{cat,H}}, k_{\mathrm{cat,P}} < 10^{3}$ min^{-1}, 10^{0} min^{-1} $< k_{\mathrm{MN}} < 10^{3}$ min^{-1}, 10^{-1} min^{-1} $< k_{\mathrm{IM}} < 10^{2}$ min^{-1}, 10^{-1} nM $< K_{\mathrm{H}}, K_{\mathrm{M}}, K_{\mathrm{N}} < 10^{4}$ nM, and 10^{2} μM $< K_{\mathrm{ATP}} < 10^{4}$ μM. The bounds for $k_{\mathrm{cat,H}}$, $k_{\mathrm{cat,P}}$, k_{MN}, k_{IM}, and K_{ATP} are based on the previous experimental measurements as further detailed in Table. 1. For $k_{\mathrm{cat,MI}}$ and $k_{\mathrm{cat,NI}}$, the bounds are set to reflect the range of measured rates for N state formation. For K_{H}, K_{M}, and K_{N}, the bounds reflect the range of CYT-19 concentrations assayed in the experiments. The remaining parameters are set by the constraints $\Phi = k_{\mathrm{IN}}/(k_{\mathrm{IN}} + k_{\mathrm{IM}}) = 0.1$ and $k_{\mathrm{MN}}/k_{\mathrm{NM}} = 10$. By repeated random sampling, we generate the parameter vector $\vec{\theta}_1$ with a sufficiently good fit, $\chi^2(\vec{\theta}_1) < 10^4$, which is used as the starting point of the Metropolis-Hastings algorithm.

For the i^{th} iteration of Metropolis-Hastings algorithm, we sample a candidate parameter vector $\vec{\theta}'$ from a multivariate Gaussian centered at $\vec{\theta}_i$, $g(\vec{\theta}'|\vec{\theta}_i)$. The covariance matrix of $g(\vec{\theta}'|\vec{\theta}_i)$, Σ, is a 10×10 diagonal matrix, in which each entry is proportional to the range of the corresponding upper and lower bounds used to sample $\vec{\theta}_1$. Specifically, for the j^{th} parameter $[\vec{\theta}]_j$, we set $[\Sigma]_{jj} = 0.002\left(\log_{10}(u_j) - \log_{10}(l_j)\right)$, where u_j and l_j are the upper and lower bounds. If the trial parameter yields a better fit (i.e. $\chi^2(\vec{\theta}') < \chi^2(\vec{\theta}_i)$), we accept it and set $\vec{\theta}_{i+1} = \vec{\theta}'$. Otherwise, we determine the acceptance by generating a random number $\beta \in [0, 1]$, and comparing it with $\exp\left(0.5\left[\chi^2(\vec{\theta}_i) - \chi^2(\vec{\theta}')\right]\right)$. We set $\vec{\theta}_{i+1} = \vec{\theta}'$ if $\beta < \exp\left(0.5\left[\chi^2(\vec{\theta}_i) - \chi^2(\vec{\theta}')\right]\right)$; otherwise $\vec{\theta}_{i+1} = \vec{\theta}_i$ and start the $(i+1)^{\mathrm{th}}$ iteration by resampling for the trial vector $\vec{\theta}'$.

In total, 400 independent trials of Metropolis-Hastings algorithm were performed, with 3×10^5 iterations each. Out of the 400 trials that started at random initial points, 202 converged to near-optimal χ^2 values, and generated $\sim 10^7$ instances of $\vec{\theta}$ with $\chi^2(\vec{\theta}) < 350$. The best fit-parameter array $\vec{\theta}_{\mathrm{min}}$, for which $\chi^2(\vec{\theta}_{\mathrm{min}}) \approx 310.5$, is given in Table 1. The distribution of parameters that yielded relatively small χ^2 values, as well as the pairwise correlations between the parameters, are shown in Fig. S2.

6. Self-splicing yield

The minimal model of self-splicing *in vivo* in the main text describes the time evolutions of the concentrations of pre-RNA in the I, M, N, and SP states, denoted by the column vector $\vec{c}(t) = ([\mathrm{I}](t), [\mathrm{M}](t), [\mathrm{N}](t), [\mathrm{SP}](t))^T$. Here, in a slightly more general model than that described in Eq. 11, we assume that the pre-RNA in the intermediate states I, M and N are degraded at the rate $k_{\mathrm{d,i}}$, and pre-RNA in the spliced state SP are degraded at the rate $k_{\mathrm{d,s}}$. Then, the pre-RNA molecules evolve with time as

$$\frac{d\vec{c}(t)}{dt} = \mathcal{W}_s \vec{c}(t) + \vec{R} \qquad \text{[S19]}$$

where $\mathcal{W}_s$ is

$$\mathcal{W}_s = \begin{bmatrix} -k_{\mathrm{IM}} - k_{\mathrm{IN}} - k_{\mathrm{d,i}} & k_{\mathrm{MI}}^{\mathrm{eff}} & k_{\mathrm{NI}}^{\mathrm{eff}} & 0 \\ k_{\mathrm{IM}} & -k_{\mathrm{MI}}^{\mathrm{eff}} - k_{\mathrm{MN}} - k_{\mathrm{d,i}} & k_{\mathrm{NM}} & 0 \\ k_{\mathrm{IN}} & k_{\mathrm{MN}} & -k_{\mathrm{NI}}^{\mathrm{eff}} - k_{\mathrm{NM}} - k_{\mathrm{d,i}} - k_{\mathrm{s}} & 0 \\ 0 & 0 & k_{\mathrm{s}} & -k_{\mathrm{d,s}} \end{bmatrix} \qquad \text{[S20]}$$

By solving Eq. S19, we obtain the array of steady state concentrations, $\vec{c}^{\,\mathrm{ss}} = ([\mathrm{I}]^{\mathrm{ss}}, [\mathrm{M}]^{\mathrm{ss}}, [\mathrm{N}]^{\mathrm{ss}}, [\mathrm{SP}]^{\mathrm{ss}})^T = -\mathcal{W}_s^{-1}\vec{R}$. Then, with the total concentration $[\mathrm{R}]_{\mathrm{tot}}^{\mathrm{ss}} = [\mathrm{I}]^{\mathrm{ss}} + [\mathrm{M}]^{\mathrm{ss}} + [\mathrm{N}]^{\mathrm{ss}} + [\mathrm{SP}]^{\mathrm{ss}}$, the self-splicing yield is given by

$$P_{\mathrm{SP}}^{\mathrm{ss}} = \frac{[\mathrm{SP}]^{\mathrm{ss}}}{[\mathrm{R}]_{\mathrm{tot}}^{\mathrm{ss}}} = \frac{\mathcal{M}}{\mathcal{N}}, \qquad \text{[S21]}$$

where

$$\mathcal{M} = k_{\mathrm{s}}\left\{k_{\mathrm{MN}}\left(k_{\mathrm{IM}} + k_{\mathrm{IN}}\right) + k_{\mathrm{IN}}\left(k_{\mathrm{d,i}} + k_{\mathrm{MI}}^{\mathrm{eff}}\right)\right\}$$

$$\mathcal{N} = k_{\mathrm{s}}\left\{k_{\mathrm{IM}}^{\mathrm{eff}} k_{\mathrm{MN}} + k_{\mathrm{IN}}^{\mathrm{eff}}\left(k_{\mathrm{d,i}} + k_{\mathrm{MI}}^{\mathrm{eff}} + k_{\mathrm{MN}}\right)\right\}$$
$$+ k_{\mathrm{d,s}}\left\{k_{\mathrm{d,i}}^2 + k_{\mathrm{MI}}^{\mathrm{eff}}(k_{\mathrm{IN}} - k_{\mathrm{MN}}) + (k_{\mathrm{IM}} + k_{\mathrm{IN}} + k_{\mathrm{MI}}^{\mathrm{eff}})(k_{\mathrm{MN}} + k_{\mathrm{NM}}) + (k_{\mathrm{IM}} + k_{\mathrm{MI}} + k_{\mathrm{MN}})(\kappa k_{\mathrm{MI}}^{\mathrm{eff}} + k_{\mathrm{s}}) + k_{\mathrm{d,i}}\left(k_{\mathrm{IM}} + k_{\mathrm{IN}} + k_{\mathrm{MI}}^{\mathrm{eff}} + \kappa k_{\mathrm{MI}}^{\mathrm{eff}} + k_{\mathrm{MN}} + k_{\mathrm{NM}} + k_{\mathrm{s}}\right)\right\}.$$

When direct transitions between the M and N states are negligible ($k_{\mathrm{MN}}, k_{\mathrm{NM}} \ll 1$), $P_{\mathrm{SP}}^{\mathrm{ss}}$ is approximated to

$$P_{\mathrm{SP}}^{\mathrm{ss}} \approx \frac{k_{\mathrm{s}}}{k_{\mathrm{s}} + \frac{k_{\mathrm{d,s}}}{k_{\mathrm{IN}}}\left\{k_{\mathrm{d,i}} + k_{\mathrm{IN}} + \kappa k_{\mathrm{MI}}^{\mathrm{eff}} + k_{\mathrm{s}} + \frac{k_{\mathrm{IM}}(\kappa k_{\mathrm{MI}}^{\mathrm{eff}} + k_{\mathrm{s}}) + k_{\mathrm{d,i}} k_{\mathrm{IM}}}{k_{\mathrm{d,i}} + k_{\mathrm{MI}}^{\mathrm{eff}}}\right\}}. \qquad \text{[S22]}$$

$P_{\mathrm{SP}}^{\mathrm{ss}}$ in Eq. S22 is maximized with respect to $k_{\mathrm{MI}}^{\mathrm{eff}}$:

$$\left(k_{\mathrm{MI}}^{\mathrm{eff}}\right)^* = \sqrt{\frac{k_{\mathrm{IM}}\left\{k_{\mathrm{s}} + (1-\kappa)k_{\mathrm{d,i}}\right\}}{\kappa}} - k_{\mathrm{d,i}}$$
$$\approx \sqrt{\frac{k_{\mathrm{IM}} k_{\mathrm{s}}}{\kappa}}, \qquad \text{[S23]}$$

where the approximation holds if $k_{\mathrm{d,i}} \ll k_{\mathrm{IM}}, k_{\mathrm{s}}$. When $\kappa = 0$, $P_{\mathrm{SP}}^{\mathrm{ss}}$ is a hyperbolic function of $k_{\mathrm{MI}}^{\mathrm{eff}}$, saturating to its maximum value. Qualitatively similar conclusions are drawn for the case with finite values of k_{MN} and k_{NM} (Fig. 5B).

7. Dependence of $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ on κ

We discuss the κ-dependence of $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ that was used to quantify the steady state yield of native ribozyme over the folding time in Ref.(4). The probability of ribozyme in the I, M, and N states, denoted by the column vector $\vec{P}(t) \equiv ([\mathrm{I}](t), [\mathrm{M}](t), [\mathrm{N}](t))^T / [\mathrm{R}]_{\mathrm{tot}}$, evolve in time following $\frac{d}{dt}\vec{P}(t) = \mathcal{W}_o \vec{P}(t)$, where $\mathcal{W}_o$ is

$$\mathcal{W}_o = \begin{bmatrix} -k_{\mathrm{IM}} - k_{\mathrm{IN}} & k_{\mathrm{MI}}^{\mathrm{eff}} & k_{\mathrm{NI}}^{\mathrm{eff}} \\ k_{\mathrm{IM}} & -k_{\mathrm{MI}}^{\mathrm{eff}} - k_{\mathrm{MN}} & k_{\mathrm{NM}} \\ k_{\mathrm{IN}} & k_{\mathrm{MN}} & -k_{\mathrm{NI}}^{\mathrm{eff}} - k_{\mathrm{NM}} \end{bmatrix}. \quad [S24]$$

When all the rates are constant in time, $\vec{P}(t) = c_1\vec{u}_1 e^{\lambda_1 t} + c_2\vec{u}_2 e^{\lambda_2 t} + c_3\vec{u}_3 e^{\lambda_3 t}$, where $\lambda_1 = 0 > \lambda_2 > \lambda_3$ are the eigenvalues of $\mathcal{W}_o$, and u_1, u_2, and u_3 are the corresponding eigenvectors. Thus, if $|\lambda_2| \ll |\lambda_3|$, $\vec{P}(t)$ exponentially relaxes to the steady state, with the relaxation rate $\lambda = |\lambda_2|$. With the initial condition $P_{\mathrm{N}}(0) = 0$, $P_{\mathrm{N}}(t)$ is approximated as

$$P_{\mathrm{N}}(t) \approx P_{\mathrm{N}}^{\mathrm{ss}}\left(1 - e^{-\lambda t}\right). \quad [S25]$$

Provided that the transitions between the M and N states are negligible ($k_{\mathrm{MN}}, k_{\mathrm{NM}} \ll 1$), λ and $P_{\mathrm{N}}^{\mathrm{ss}}$ are expressed as follows with $\kappa(= k_{\mathrm{NI}}^{\mathrm{eff}}/k_{\mathrm{MI}}^{\mathrm{eff}})$ and $\Phi(= k_{\mathrm{IN}}/(k_{\mathrm{IN}} + k_{\mathrm{IM}}))$,

$$\lambda = \frac{\left(\left[k_{\mathrm{IM}} + (1+\kappa)(1-\Phi)k_{\mathrm{MI}}^{\mathrm{eff}}\right] - \sqrt{\left[k_{\mathrm{IM}} + (1+\kappa)(1-\Phi)k_{\mathrm{MI}}^{\mathrm{eff}}\right]^2 - 4k_{\mathrm{MI}}^{\mathrm{eff}}(1-\Phi)\left[\left(k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}\right)(1-\Phi)\kappa + k_{\mathrm{IM}}\Phi\right]}\right)}{2(1-\Phi)}, \quad [S26]$$

$$P_{\mathrm{N}}^{\mathrm{ss}} = \frac{k_{\mathrm{IM}}\Phi}{\left(k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}\right)(1-\Phi)\kappa + k_{\mathrm{IM}}\Phi}, \quad [S27]$$

For $0 \lesssim \Phi \ll 1$ and $0 \ll \Phi \lesssim 1$, $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ can be expanded as a Taylor series of Φ and $1-\Phi$, respectively: For $0 \lesssim \Phi \ll 1$,

$$\lambda P_{\mathrm{N}}^{\mathrm{ss}} = \frac{k_{\mathrm{IM}}k_{\mathrm{MI}}^{\mathrm{eff}}}{k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}}\left[\Phi + \frac{k_{\mathrm{MI}}^{\mathrm{eff}}}{k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}}\frac{(2-\kappa)k_{\mathrm{IM}} + (1-\kappa)k_{\mathrm{MI}}^{\mathrm{eff}}}{k_{\mathrm{IM}} + (1-\kappa)k_{\mathrm{MI}}^{\mathrm{eff}}}\Phi^2 + O(\Phi^3)\right], \quad [S28]$$

and for $0 \ll \Phi \lesssim 1$,

$$\lambda P_{\mathrm{N}}^{\mathrm{ss}} = k_{\mathrm{MI}}^{\mathrm{eff}} + \frac{\kappa(k_{\mathrm{MI}}^{\mathrm{eff}})^2}{k_{\mathrm{IM}}}(1-\Phi) + O[(1-\Phi)^2]. \quad [S29]$$

For $0 \lesssim \Phi \ll 1$, the relative difference between the values of $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ with $\kappa = 0$ and $\kappa = 1$ is given as

$$\frac{\left(\lambda P_{\mathrm{N}}^{\mathrm{ss}}\right)_{\kappa=0} - \left(\lambda P_{\mathrm{N}}^{\mathrm{ss}}\right)_{\kappa=1}}{\left(\lambda P_{\mathrm{N}}^{\mathrm{ss}}\right)_{\kappa=1}} = \frac{k_{\mathrm{IM}}k_{\mathrm{MI}}^{\mathrm{eff}}}{\left(k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}\right)^2}\Phi + O\left[\Phi^2\right]. \quad [S30]$$

Therefore, the difference between $(\lambda P_{\mathrm{N}}^{\mathrm{ss}})_{\kappa=0}$ and $(\lambda P_{\mathrm{N}}^{\mathrm{ss}})_{\kappa=1}$ is negligible when $0 \lesssim \Phi \ll 1$. Qualitatively similar conclusions for the case with finite values of k_{MN} and k_{NM} are shown in Fig. S3, in which we use the rate constants in Table 1.

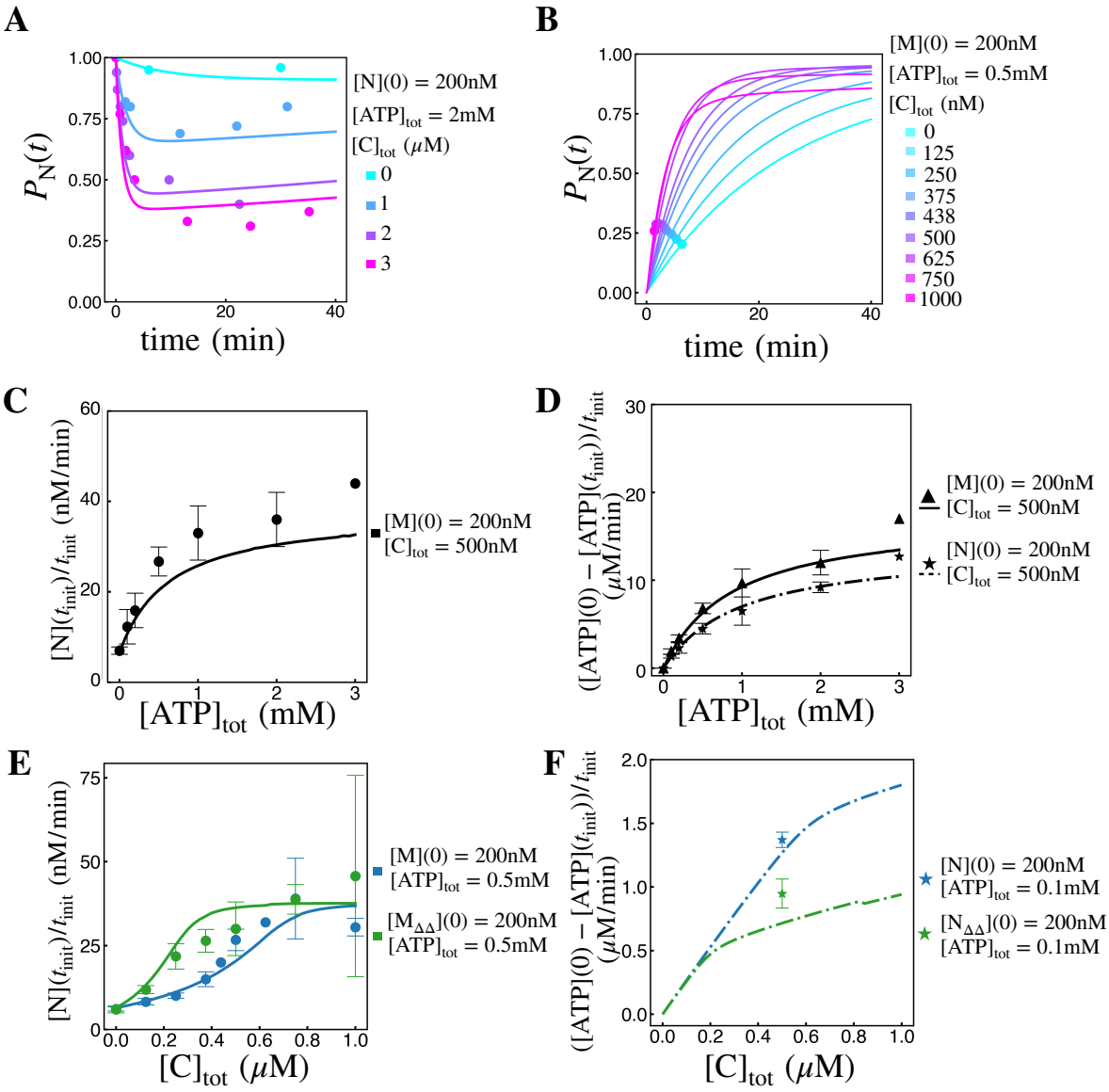

Fig. S1. Analysis of CYT-19-facilitated folding of *T.* ribozyme. The curves represent the fitted model, and the dots (except in panel (B)) represent the experimental data. The error bars associated with the dots denote two standard deviations from the mean, and the error bars are omitted for data points generated from single trials. All curves are computed with the best-fit parameters shown in Table 1. **A**. The fraction of N state ribozyme in time ($P_N(t)$) at different concentrations of CYT-19, with the initial condition $[N](0) = [R]_{tot} = 200$ nM. **B**. The fraction of N state ribozyme ($P_N(t)$) at different concentrations of CYT-19, with the initial condition $[M](0) = [R]_{tot} = 200$ nM. The filled circles represent the initial time points (t_{init}) from which the initial rates $[N](t_{init})/t_{init}$ plotted in blue in Fig. 2C are derived. **CD**. The initial refolding rate (C) and ATP hydrolysis rate (D) measured at different concentrations of ATP. **EF**. The initial refolding rate (E) and the initial ATP hydrolysis rate (F) of the mutant ribozyme devoid of P9-2 and P6b (denoted by $M_{\Delta\Delta}$ and $N_{\Delta\Delta}$) are shown in green. The Mg^{2+} concentration in each experiment was 2 mM for all but the condition $[C]_{tot} = 0$ in panel A.

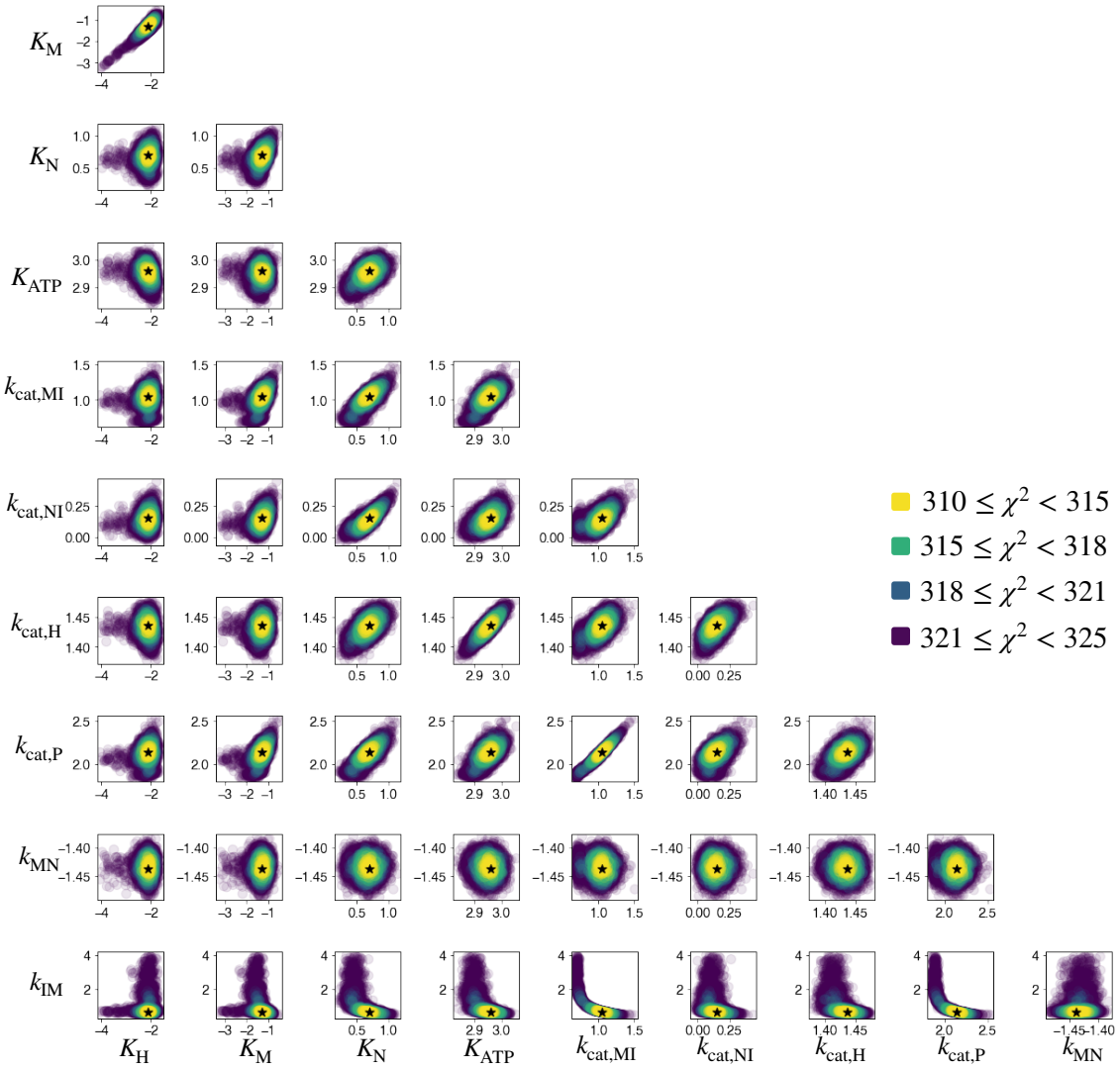

Fig. S2. The distribution of parameters that fit experimental data, generated by Monte-Carlo-Markov-Chain (MCMC) method. 400 independent trials of Metropolis-Hastings algorithm were performed to sample the parameters in log space. The pair-wise correlations between all parameters are shown, with the color representing the goodness of fit, $\chi^2(\vec{\theta})$ (Eq. S18). The black stars indicate the best-fit parameter vector $\vec{\theta}_{\min}$. k_{MN}, k_{IM}, $k_{cat,MI}$, $k_{cat,NI}$, $k_{cat,H}$, and $k_{cat,P}$ are in units of $\min^{-1}$, and K_H, K_M, K_N, and K_{ATP} are in units of μM. All axes are shown in $\log_{10}$ scale.

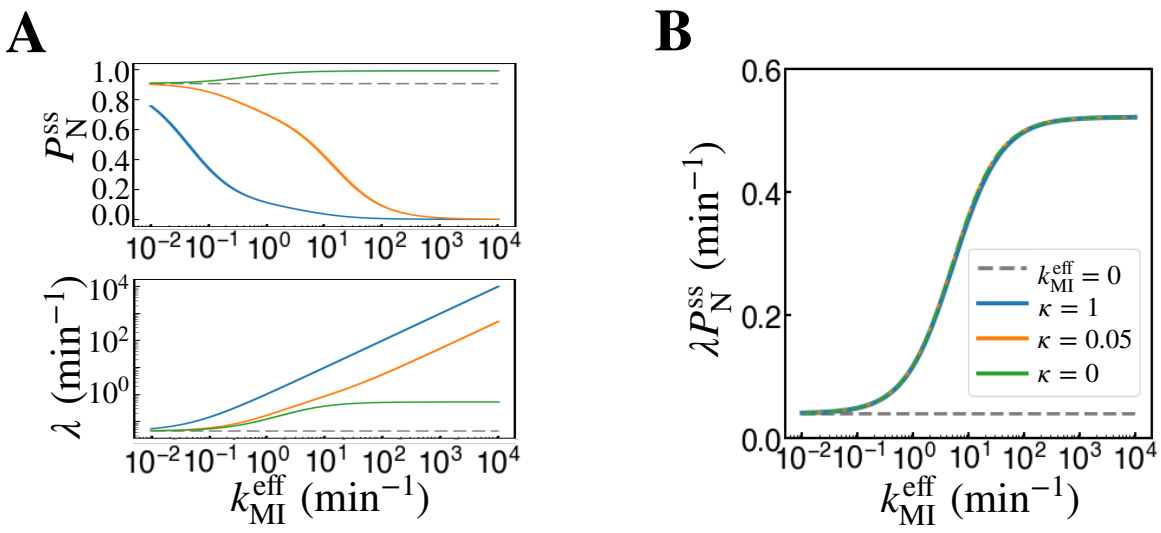

Fig. S3. Dependence of λP_N^{ss} on κ. **A** (top) P_N^{ss} and (bottom) λ, **B** λP_N^{ss} as functions of k_{MI}^{eff}, plotted with different κ values. The panels **A** and **B** are color-coded identically. The parameters other than k_{MI}^{eff} and $k_{NI}^{eff}\left[=\kappa k_{MI}^{eff}\right]$ are set to $k_{IM} = 4.7\ \text{min}^{-1}$, $k_{IN} = 0.53\ \text{min}^{-1}$, $k_{MN} = 0.04\ \text{min}^{-1}$, and $k_{NM} = 0.004\ \text{min}^{-1}$.

References

1. Yang Q, Del Campo M, Lambowitz AM, Jankowsky E (2007) DEAD-box proteins unwind duplexes by local strand separation. *Mol. Cell* 28(2):253–263.
2. Mallam AL, Del Campo M, Gilman B, Sidote DJ, Lambowitz AM (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490(7418):121–125.
3. Cao W, et al. (2011) Mechanism of Mss116 ATPase reveals functional diversity of DEAD-Box proteins. *J. Mol.Biol.* 409(3):399–414.
4. Chakrabarti S, Hyeon C, Ye X, Lorimer GH, Thirumalai D (2017) Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc. Natl. Acad. Sci. USA* 114(51):E10919–E10927.
5. Bhaskaran H, Russell R (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* 449(7165):1014–1018.
6. Jarmoskaite I, Bhaskaran H, Seifert S, Russell R (2014) DEAD-box protein CYT-19 is activated by exposed helices in a group I intron RNA. *Proc. Natl. Acad. Sci. USA* 111(29):E2928–E2936.
7. Jarmoskaite I, Tijerina P, Russell R (2021) ATP utilization by a DEAD-box protein during refolding of a misfolded group I intron ribozyme. *J. Biol. Chem.* 296:100132.