PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

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

Yonghyun Song[a], D. Thirumalai[b,1], and Changbong Hyeon[a,1]

Edited by Joseph Puglisi, Stanford University School of Medicine, Stanford, CA; received May 31, 2022; accepted October 5, 2022

CYT-19 is a DEAD-box protein whose adenosine-triphosphate (ATP)-dependent helicase activity facilitates the folding of group I introns in precursor RNA (pre-RNA) of *Neurospora crassa* (*N. crassa*). In the process, they consume a substantial amount of ATP. While much of the mechanistic insight into CYT-19 activity has been gained through the studies on the folding of *Tetrahymena* group I intron ribozyme, the more biologically relevant issue, namely the effect of CYT-19 on the self-splicing of pre-RNA, remains largely unexplored. Here, we employ a kinetic network model, based on the generalized iterative annealing mechanism (IAM), to investigate the relation between CYT-19 activity, rate of ribozyme folding, and the kinetics of the self-splicing reaction. The network rate parameters are extracted by analyzing the recent biochemical data for CYT-19-facilitated folding of *Tetrahymena* ribozyme. We then build extended models to explore the metabolism of pre-RNA. We show that the timescales of chaperone-mediated folding of group I ribozyme and self-splicing reaction compete with each other. As a consequence, in order to maximize the self-splicing yield of group I introns in pre-RNA, the chaperone activity must be sufficiently large to unfold the misfolded structures, but not too large to unfold the native structures prior to the self-splicing event. We discover that despite the promiscuous action on structured RNAs, the helicase activity of CYT-19 on group I ribozyme gives rise to self-splicing yields that are close to the maximum.

iterative annealing mechanism | kinetic partitioning mechanism | self-splicing activity | ATP consumption | thermodynamic cost

Significance

In cells, RNA chaperones assist misfolding-prone ribozymes to fold correctly to carry out their biological function. CYT-19 is an ATP-consuming RNA chaperone that accelerates the production of native group I intron ribozyme by partially unfolding the kinetically trapped structures. Using the theoretical framework based on the iterative annealing mechanism, we establish that to maximize the processing of pre-RNA, an optimal balance should exist between the timescales of self-splicing activity and CYT-19-mediated production of the native ribozyme. Remarkably, the activity of CYT-19 has been optimized to unfold the misfolded structures but is not so high that it disrupts the native ribozyme, which ensures that the yield of the self-splicing reaction is maximized in a biologically relevant timescale.

The enzymatic functions that are carried out by structured RNAs (1–5) regulate a diverse array of cellular processes including translation, RNA processing, the maintenance of chromosome ends, and the regulation of gene expression (6–9). However, the homopolymeric nature of the interactions between the nucleotides and the multivalent ion-mediated tertiary interaction result in highly rugged RNA folding landscapes (4). Even RNA secondary structures are heterogeneous, making interconversions between them extremely slow (10–13). Subsequent transitions of RNA to a functional native structure occur via multiple pathways, usually with diverse timescales, giving rise to multiexponential kinetics, which is quantitatively explained by the kinetic partitioning mechanism (KPM) (4, 14, 15). Only a small fraction of the ensemble ($\Phi \ll 1$) folds directly to the native basin of attraction, whereas the remaining fraction ($1 - \Phi$) is trapped in the manifold of long-lived misfolded states that define the competing basins of attraction (Fig. 1) (4). Although spontaneous transitions from misfolded to native states, in principle, are not unattainable, they are kinetically inaccessible because the timescales for such processes are far too long compared to the typical cell-doubling time (16–19). RNA chaperones, which assist RNA molecules to escape from deep kinetic traps in the folding landscapes and facilitate their folding on a biologically relevant timescale, are the key players in cellular RNA metabolism (20–23).

CYT-19 in *Neurospora crassa* and its homologue Mss116 in yeast are the best-studied RNA chaperones (9, 24). An earlier work showed that unspliced group I introns accumulated in the mutant *N. crassa* that is devoid of CYT-19, a DEAD-box protein with a helicase domain whose ATP-dependent activity accelerates the folding and self-splicing of group I introns (25). Interestingly, although it is a protein derived from fungi, CYT-19 has a generic activity toward RNAs from different organisms (25, 26), as does the bacterial GroEL, the protein chaperone (27, 28).

Russell et al. have quantified the effect of CYT-19 on the folding of *Tetrahymena* ribozyme, a group I intron-derived catalytic RNA (29–32), over a wide range of experimental conditions (33–37). They have determined that CYT-19 is generally reactive toward surface-exposed RNA helices (33). Surprisingly, unlike protein chaperones (38), CYT-19 disrupts the structures of both the misfolded and native ribozymes (35, 36).

Author affiliations: [a]Korea Institute for Advanced Study, Seoul 02455, Korea; and [b]Department of Chemistry, The University of Texas at Austin, Austin, TX 78712

Author contributions: C.H. designed research; Y.S., D.T., and C.H. performed research; Y.S. and C.H. analyzed data; and Y.S., D.T., and C.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: hyeoncb@kias.re.kr or dave.thirumalai@gmail.com.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2209422119/-/DCSupplemental.

Published November 28, 2022.

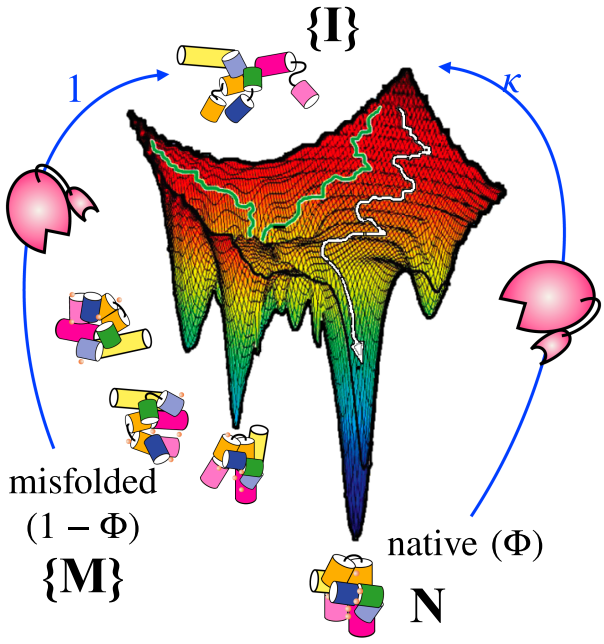

Fig. 1. Chaperone-facilitated folding of RNA on a rugged folding landscape. The fraction Φ among the ensemble of RNA molecules in the intermediate (**{I}**) states reaches the native (**N**) state rapidly, while $(1-\Phi)$ are trapped in the ensemble of misfolded (**{M}**) states. The repeated actions of RNA chaperone (depicted as a Pac-Man-like object) anneal the otherwise misfolding-prone population of RNA into the one with a high native yield as quantified in Eq. **1**.

We have previously provided a mechanistic explanation of the action of RNA chaperones, which consume ATP for their functions, by generalizing (39) the iterative annealing mechanism (IAM) (23, 38, 40–42). According to the IAM, molecular chaperones anneal the population of misfolded biopolymers by disrupting their structures and offering them multiple chances to fold. The theory based on the IAM predicts that the yield of the native state (N), after n iterations of folding transitions, $Y_{\rm N}(n)$, is given by, (23, 41, 42)

$$Y_{\rm N}(n) = \frac{\Phi - (1-\kappa)^n(1-\Phi)^n\Phi}{\kappa + (1-\kappa)\Phi} = \begin{cases} 1-(1-\Phi)^n & \text{for } \kappa = 0, \\ \Phi & \text{for } \kappa = 1, \end{cases} \quad [1]$$

where the recognition factor, κ, is the ratio of chaperone-mediated transition rates associated with reversing the spontaneous transitions to the native and misfolded states (see Fig. 1), respectively. For $\kappa \approx 0$, the unfolding action of the chaperone is restricted to the misfolded states, the condition that holds in the case of the bacterial GroEL-GroES chaperonin system (38, 40). On the other hand, $\kappa = 1$ corresponds to a hypothetical chaperone that promiscuously interacts with both the misfolded and native conformations. A chaperone with $\kappa = 0$ increases $Y_{\rm N}(n)$ from Φ ($n = 1$) to the unity ($n \gg 1$ – a proxy for time), whereas a chaperone with $\kappa = 1$ leaves the native yield unaltered ($Y_{\rm N}(n) = \Phi$ for all n). The latter process would correspond to a futile enzymatic cycle. In the case of CYT-19 acting on *T.* ribozyme, the value of κ is in the range $0 \lesssim \kappa \ll 1$ ($\kappa \approx 0.1$ (23)).

The generalized IAM (Eq. **1**), that was mapped onto the three-state kinetic model (23), showed that although the steady-state native yield of the ribozyme decreases with increasing CYT-19 concentration, the product of the chaperone-mediated folding rate and the steady-state yield of native substrate is an increasing function of chaperone concentration for both RNA and protein chaperones. This optimization condition suggests that molecular chaperones increase the production of functionally active substrates on a biologically relevant timescale, which appears to be a general principle governing the functions of a class of chaperones. Achieving the optimization comes at an expense (38). For instance, it has been argued that CYT-19-facilitated folding of the ribozyme is an energetically intensive process, resulting in the hydrolysis of as many as ∼100 ATP molecules in order to assist the folding of a single misfolded ribozyme (37). The optimization condition, requiring a substantial amount of ATP consumption, highlights the nonequilibrium nature of chaperone-facilitated dynamics of substrate molecules for a number of seemingly unrelated systems (42).

There are a plethora of studies that investigate the effects of CYT-19 on the folding of *T.* ribozyme. However, studies on the CYT-19 effect on the self-splicing of group I intron in pre-RNA are relatively scarce, except one that used a mutant ribozyme without the P5abc domain (25). The lifetime of the native state is finite and even shortened by the helicase activity of CYT-19. Given that self-splicing of group I ribozyme occurs in a finite time, we expect that there ought to be competition between the lifetime of the native ribozyme and the timescale for the self-splicing reaction. To our knowledge, there are no studies that explore the link between RNA chaperone-mediated folding and the splicing activity. Here, we fill this gap using theory. We first analyze the recent in vitro measurements of *T.* ribozyme folding (35–37) to extract the parameters involving chaperone activity for use in the generalized IAM-based three-state model (23). We then use our theory to make experimentally testable predictions about the role of RNA chaperones in the processing of pre-RNA in vivo.

Our study shows that there are multiple competing timescales in the processing of pre-RNA. The condition for maximizing the yield of the self-spliced product depends on the lifetimes of the misfolded, native state, the rate of self-splicing reaction, and the extent of CYT-19 activity. Our theory shows that the action of molecular chaperone on RNA should be strong enough to transform the misfolded ribozyme to the native state but should not be too strong to disrupt the native state before the self-splicing of pre-RNA occurs. The theory provides a quantitatively precise condition that gives rise to the optimal level of chaperone activity to maximize the yield of spliced pre-RNA in vivo.

Results

IAM-Based Theory Using the Three-State Kinetic Model for CYT-19-Assisted Folding of *Tetrahymena* Ribozyme. At the early stage of folding, under native conditions (determined by the solvent quality, temperature, and ion concentration), the RNA molecule rapidly collapses from an unfolded state and adopts secondary structures (Fig. 2*A*) (44), yielding an ensemble of intermediate states [I, represented in red (Fig. 2*B*)]. The three-dimensional structures that are assembled by forming tertiary contacts are divided into a nonfunctional ensemble of misfolded states (M, green) and the functionally competent native states (N, blue) (Fig. 2*B*). In the absence of RNA chaperone, the transitions from I to M (I → M with rate constant $k_{\rm IM}$) or from I to N (I → N with rate constant $k_{\rm IN}$) are effectively irreversible. The fraction of ribozymes that fold spontaneously to the N state is given by the partition factor, $\Phi \equiv k_{\rm IN}/(k_{\rm IN}+k_{\rm IM})$. For *T.* ribozyme, $\Phi \approx 0.1$ (19, 45–47). Transitions between the M

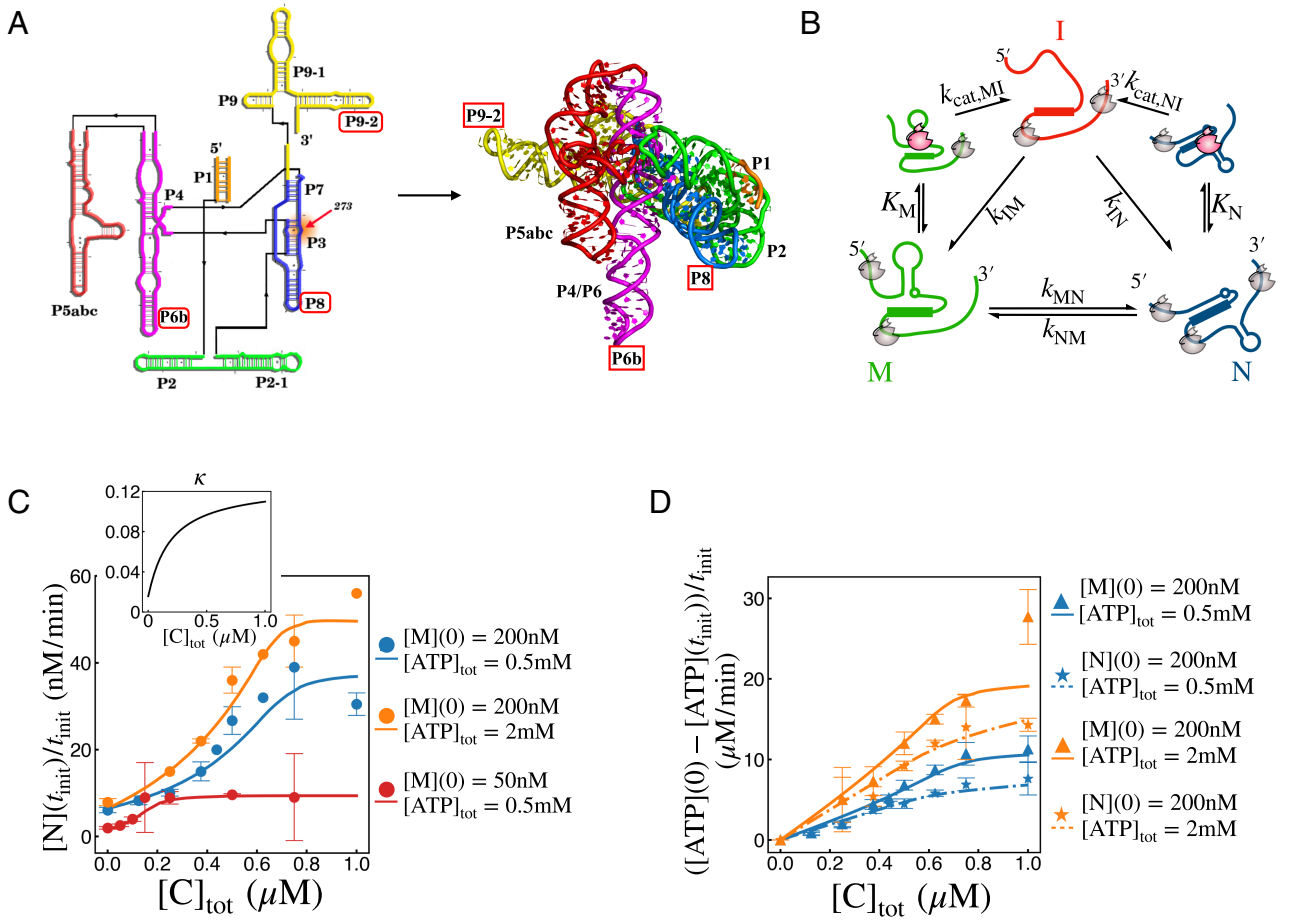

Fig. 2. RNA structure and chaperone-assisted folding. (*A*) Schematics of the folding of *T.* ribozyme. On the left is the secondary structure map, and the right side represents the three-dimensional native structure (PDB id: 7EZ0 (43)). (*B*) A kinetic scheme for the three-state model of CYT-19-assisted folding of *T.* ribozyme used in the theoretical analysis. RNA chaperones in productive and futile binding modes are depicted in red and gray, respectively. (*C*, *D*). Analysis of CYT-19-assisted folding of *T.* ribozyme. The curves represent the fits obtained using our model, and the symbols are the experimental data. The error bars denote two standard deviations from the mean, and the error bars are omitted for data points generated from single trials. The initial refolding rates of the ribozyme (*C*) and the ATP hydrolysis rates (*D*) measured over multiple CYT-19, ATP, and ribozyme concentrations. The initial rates of N state formation and ATP consumption were set to $[N](t_{init})/t_{init}$ and $([ATP](0) - [ATP](t_{init}))/t_{init}$, respectively (see *SI Appendix, Model Fit to Data* for the definition of the early time point t_{init}). All curves are computed with the best-fit parameters shown in Table 1. The experimental data points, obtained from (35–37), are provided in *SI Appendix*.

and N states could occur with rates k_{MN} and k_{NM}, but the timescale associated with the M → N transition is far too slow to be biologically relevant ($k_{MN} < 0.05\ min^{-1}$) (35, 36). Thus, the dynamic process of ribozyme folding in vitro is under kinetic control, which implies that a substantial portion ($(1-\Phi) \approx 0.9$) of the ribozyme, starting from the I state, is trapped in the M state.

Following the IAM-based three-state model (Fig. 2*B*) for the chaperone-assisted folding (23), we assume that CYT-19 binds to M and N states of the ribozyme (35) with distinct binding affinities (dissociation constants) K_M and K_N. Subsequent binding of ATP to the CYT-19-ribozyme complexes, with the affinity K_{ATP}, and the ATP-dependent helicase activity of CYT-19, revert the M and N states to the I state with rates $k_{cat,MI}$ and $k_{cat,NI}$. The CYT-19 and ATP concentration-dependent chaperone activity-mediated transition rates from the M and N states to the I state, represented by the *effective* rates $k_{MI}^{eff}(t)$ and $k_{NI}^{eff}(t)$, which account for the sequential events of the CYT-19 binding to ribozyme (M or N) and the catalytic activity of CYT-19 powered by ATP consumption, are modeled using Michaelis–Menten type kinetics,

$$k_{\alpha I}^{eff}(t) = k_{cat,\alpha I} f([ATP](t), [C](t); K_{ATP}, K_\alpha) \qquad [2]$$

with

$$f([ATP](t), [C](t); K_{ATP}, K_\alpha) \equiv \frac{\frac{[ATP](t)}{K_{ATP}} \frac{[C](t)}{K_\alpha}}{1 + \frac{[C](t)}{K_\alpha} + \frac{[ATP](t)}{K_{ATP}} \frac{[C](t)}{K_\alpha}},$$

where α = M or N, $[C](t)$ denotes the time-dependent concentrations of free CYT-19, and $[ATP](t)$ is also modeled time dependent since the ATP concentration is not held constant in the experiments (35–37). For the derivation of Eq. **2**, see *SI Appendix, Effective Rate Constants for CYT-19 Activity*, and for further detail on the calculation of $[C](t)$, see *SI Appendix, Mass Balance of CYT-19*. The parameters $k_{cat,MI}$, $k_{cat,NI}$, K_M, K_N, and K_{ATP} are the catalytic rates and binding constants as shown in Fig. 2*B*. In the presence of chaperone activity with $[C](t)$, $[ATP](t) \neq 0$, the reactions between I, M, and N states are reversible. Implicitly shown in Fig. 2*B*, CYT-19 can strongly bind to surface-exposed, protruding helices, such as P6b, P9-2, and P8 (see Fig. 2*A*) and unwind them by futile consumption

of ATP (37). Taken together, in our theoretical framework, that implicitly takes into account the futile ATP consumption to unwind the protruding helices, ATP molecules are consumed via three pathways and obey the following rate equation:

$$-\frac{d}{dt}[\mathrm{ATP}](t) = k^{\mathrm{eff}}_{\mathrm{ATP,M}}(t)[\mathrm{M}](t) + k^{\mathrm{eff}}_{\mathrm{ATP,N}}(t)[\mathrm{N}](t) + k^{\mathrm{eff}}_{\mathrm{ATP,H}}(t)[\mathrm{H}]. \quad [3]$$

Here, $k^{\mathrm{eff}}_{\mathrm{ATP,M}}(t)$ and $k^{\mathrm{eff}}_{\mathrm{ATP,N}}(t)$ are the effective ATP turnover rates catalyzed by CYT-19 bound to the ribozyme in the M and N states, respectively, and are defined similarly to $k^{\mathrm{eff}}_{\mathrm{MI}}(t)$ and $k^{\mathrm{eff}}_{\mathrm{NI}}(t)$ in Eq. **2** (see Eq. **S7** for more precise definition). $k^{\mathrm{eff}}_{\mathrm{ATP,H}}(t)$ is the effective ATP turnover rate catalyzed by CYT-19 bound to the surface exposed helices. The total concentration of the surface exposed helices is given by $[\mathrm{H}] = n_{\mathrm{H}}[\mathrm{R}]_{\mathrm{tot}}$, where we assume that the number of secondary structure motifs, n_{H}, is identical in the I, M, and N states. To account for the constitutively exposed helices of *T.* ribozyme, P6b, P8, and P9-2 (see Fig. 2*A*), we set $n_{\mathrm{H}} = 3$ (36). It is noteworthy that Eq. **3**, *SI Appendix*, Eqs. **S7** and **S8** model the rate of ATP consumption to be proportional to the amount of the CYT-19-ribozyme complex, and not to the occurrence of unfolding reactions, so that they include the contribution from unproductive futile cycles, which do not result in unfolding the ribozyme. For further detail, see *SI Appendix, Effective Rate Constants for CYT-19 Activity*.

Finally, the kinetic and thermodynamic parameters for the model, given in Fig. 2*B*, are determined by simultaneously fitting the biochemical data collected from three reports (35–37) to the solution of the three-state kinetic model described by the following rate equation:

$$\frac{d}{dt}\vec{P}(t) = \mathcal{W}_o(t)\vec{P}(t), \quad [4]$$

where $\vec{P}(t) \equiv ([\mathrm{I}](t), [\mathrm{M}](t), [\mathrm{N}](t))^T/[\mathrm{R}]_{\mathrm{tot}} = (P_{\mathrm{I}}(t), P_{\mathrm{M}}(t), P_{\mathrm{N}}(t))^T$ with $[\mathrm{R}]_{\mathrm{tot}} = [\mathrm{I}] + [\mathrm{M}] + [\mathrm{N}]$ being the vector of the normalized concentrations (or probabilities) and the rate matrix $\mathcal{W}_o(t)$ being

$$\mathcal{W}_o(t) = \begin{bmatrix} -k_{\mathrm{IM}} - k_{\mathrm{IN}} & k^{\mathrm{eff}}_{\mathrm{MI}}(t) & k^{\mathrm{eff}}_{\mathrm{NI}}(t) \\ k_{\mathrm{IM}} & -k^{\mathrm{eff}}_{\mathrm{MI}}(t) - k_{\mathrm{MN}} & k_{\mathrm{NM}} \\ k_{\mathrm{IN}} & k_{\mathrm{MN}} & -k^{\mathrm{eff}}_{\mathrm{NI}}(t) - k_{\mathrm{NM}} \end{bmatrix}. \quad [5]$$

More specifically, the parameters in the theory were obtained by simultaneously fitting to 11 sets of data on the folding of *T.* ribozyme (*SI Appendix*, Fig. S1) and two sets of data on the folding of mutant *T.* ribozyme lacking two helices (Fig. 2*C*) and ATP consumption rate (Fig. 2*D*) with varying concentrations of CYT-19, ATP, and ribozyme, along with two constraints on the parameters: i) $\Phi = k_{\mathrm{IN}}/(k_{\mathrm{IN}} + k_{\mathrm{IM}}) = 0.1$ (45–47) and ii) $k_{\mathrm{MN}}/k_{\mathrm{NM}} = 10$ based on the observation that ~90% of the ribozyme is found in the N state at equilibrium (Fig. 2a in ref. (36)). We extracted the best-fit parameters by minimizing the weighted sum of squared residuals (χ^2-statistic, which is further explained in *SI Appendix, Model Fit to Data*). It should be emphasized that despite having about ten parameters, they were robustly obtained by simultaneously fitting to a large set of data. The model captures all the experimentally observed trends (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S1), and the best-fit parameters are consistent with experimental measurements, although some of the measurements were performed under different conditions of temperature and ion concentrations (Table 1).

From the parameters determined from the analysis, the following inferences may be made.

1. CYT-19 binds more strongly to the M state ($K_{\mathrm{M}} < K_{\mathrm{N}}$) and catalyzes the unfolding reaction more rapidly ($k_{\mathrm{cat,NI}} < k_{\mathrm{cat,MI}}$). The helicase activity of CYT-19 is more productive if the ribozyme is in the M state. The extracted parameters in Table 1 indicate $k_{\mathrm{cat,NI}}/K_{\mathrm{N}} < k_{\mathrm{cat,MI}}/K_{\mathrm{M}}$. As a result, CYT-19 displays a greater "specificity" toward the M state with the recognition factor satisfying $\kappa (\equiv k^{\mathrm{eff}}_{\mathrm{NI}}/k^{\mathrm{eff}}_{\mathrm{MI}}) \ll 1$. In the range of $[\mathrm{C}]_{\mathrm{tot}} = (1 - 20)\ \mu$M and $[\mathrm{ATP}]_{\mathrm{tot}} = 0.5$ mM, κ increases monotonically with $[\mathrm{C}]_{\mathrm{tot}}$, from $\kappa \approx 0.01$ to $\kappa \approx 0.1$. These results are consistent with a previous study (23).
2. The binding of CYT-19 to the surface-exposed, protruding helices (P6b, P9-2, and P8. See Fig. 2*A*) is stronger than that the ribozyme core ($K_{\mathrm{H}} < K_{\mathrm{M}}, K_{\mathrm{N}}$), which means that the helices effectively sequester CYT-19 and hamper the CYT-19-assisted folding of the ribozyme when the amount of free CYT-19 in solution is small. The effect of protruding helices manifests itself as a sigmoidal increase in the folding rate with increasing CYT-19 concentration (Fig. 2*C*). Note that the initial rates of ATP hydrolysis in the presence of M and N states are comparable. This is because a large portion of the ATP turnover rates is contributed by CYT-19 bound to the protruding helices ($K_H \ll K_M \ll K_N$. See Table 1), especially when the $[\mathrm{C}]_{\mathrm{tot}}$ concentration is low (Fig. 2*D*).

   Our model, in which the number of exposed helices per mutant ribozyme devoid of P9-2 and P6b is set to unity ($n_{\mathrm{H}} = 1$), qualitatively reproduces the experimentally observed suppression of the refolding rate of wild-type ribozyme in comparison with the mutant (*SI Appendix*, Fig. S1E) and the lower ATP consumption rate in the mutant (*SI Appendix*, Fig. S1F). We find that ATP catalysis occurs five times faster in the productive CYT-19 binding to the ribozyme than in the binding to the protruding helices ($k_{\mathrm{cat,P}} \approx 5k_{\mathrm{cat,H}}$).

Predictions for CYT-19-Facilitated Folding and Self-Splicing of pre-RNA. The aforementioned experiments (35–37) were concerned with the folding of *T.* group I intron ribozyme (35–37), which catalyzes its own excision from the precursor RNA (pre-RNA) (51). Here, we extend the IAM-based model to predict the effect of CYT-19 activity on the self-splicing of pre-RNA.

To study the effect of CYT-19 on pre-RNA, we refer to the model suggested by Pan et al., who performed in vitro folding measurements (19). The model assumes that *T.* ribozyme containing pre-RNA undergoes transitions among the ensembles of intermediate states (I), two different misfolded states (M_1 and M_2), and the native state (N) (Fig. 3*A*). In addition, CYT-19 is assumed to bind to the M_1, M_2, and N states with dissociation constants K_{M_1}, K_{M_2}, and K_{N}, respectively. Subsequent CYT-19-mediated transition rates from the M_1, M_2, and N states to the I state are denoted by the effective rate constants $k^{\mathrm{eff}}_{\mathrm{M}_1\mathrm{I}}(t)$, $k^{\mathrm{eff}}_{\mathrm{M}_2\mathrm{I}}(t)$, and $k^{\mathrm{eff}}_{\mathrm{NI}}(t)$, which are given in the form,

$$k^{\mathrm{eff}}_{\alpha\mathrm{I}}(t) = k_{\mathrm{cat},\alpha\mathrm{I}} f([\mathrm{ATP}]_{\mathrm{tot}}(t), [\mathrm{C}](t); K_{\mathrm{ATP}}, K_\alpha), \quad [6]$$

with $\alpha = \mathrm{M}_1, \mathrm{M}_2$, or N. In the absence of experimental data, we set the concentration of the protruding helices to 0 in the calculation of $[\mathrm{C}](t)$ ($[\mathrm{H}]_{\mathrm{tot}} = 0$. See *SI Appendix*, Eq. **S11** *in Mass Balance of CYT-19* for further detail).

Table 1. Best-fit parameters extracted by fitting the experimental data at 25°C with 1 mM Mg^{2+} and 2 mM Mg^{2+} in ref. 35 and with 2 mM Mg^{2+} in refs. 36 and 37

	$k_{cat,MI}$ (min^{-1})	$k_{cat,NI}$ (min^{-1})	K_H (nM)	K_M (nM)	K_N (nM)	k_{NM} (min^{-1})		$^*k_{IM}$ (min^{-1})	$^*k_{IN}$ (min^{-1})	$^*k_{MN}$ (min^{-1})	K_{ATP} (μM)	$k_{cat,H}$ (min^{-1})	$k_{cat,P}$ (min^{-1})
Best-fit	11	1.4	7.3	50	5000	3.6×10^{-3}	Best-fit	4.7	0.53	0.037	900	27	140
							exp.	$^\dagger 1.5 \pm 0.3$	$^\dagger 0.37 \pm 0.10$	$^\ddagger 0.05$	$^\S 50-500$	$^\S 6-600$	$^\S 6-600$

* In the fitting procedure, we fix the ratios $k_{MN}/k_{NM} = 10$ and $\Phi = k_{IN}/(k_{IN} + k_{IM}) = 0.1$.

† Ref. 45, with 10 mM Mg^{2+}.

‡ Ref. 35, with 1 mM Mg^{2+}.

§ Refs. 48 and 49.

The relations among the best-fit parameters, $k_{cat,NI} < k_{cat,MI}$, $K_H < K_M < K_N$, and $k^{eff}_{ATP,H} < k_{ATP,P}$ are also confirmed in the distribution of good-fit parameters acquired from the exhaustive sampling using the Markov chain Monte Carlo (MCMC) method (*SI Appendix*, Fig. S2).

We consider the experimental setup in which $[R]_{tot}$ of pre-RNA is incubated with $[C]_{tot}$ of CYT-19 and $[ATP]_{tot}$ of ATP. The time evolution of the pre-RNA concentrations in the four states, $[I](t)$, $[M_1](t)$, $[M_2](t)$, and $[N](t)$, are described by

$$\frac{d}{dt}\vec{P}(t) = \mathcal{W}_o(t)\vec{P}(t), \qquad [7]$$

where $\vec{P}(t) \equiv ([I](t), [M_1](t), [M_2](t), [N](t))^T/[R]_{tot} = (P_I(t), P_{M_1}(t), P_{M_2}(t), P_N(t))^T$ is the vector of the normalized concentrations, and the transition matrix, $\mathcal{W}_o(t)$, is given by,

$$\mathcal{W}_o(t) = \begin{bmatrix} -k_{IM_1} - k_{IM_2} - k_{IN} & k^{eff}_{M_1I}(t) & k^{eff}_{M_2I}(t) & k^{eff}_{NI}(t) \\ k_{IM_1} & -k^{eff}_{M_1I}(t) - k_{M_1N} & 0 & k_{NM_1} \\ k_{IM_2} & 0 & -k^{eff}_{M_2I}(t) - k_{M_2N} & k_{NM_2} \\ k_{IN} & k_{M_1N} & k_{M_2N} & -k^{eff}_{NI}(t) - k_{NM_1} - k_{NM_2} - k_s \end{bmatrix}. \qquad [8]$$

The splicing rate is set to $k_s = 0\ min^{-1}$ when the self-splicing reaction is blocked, e.g., by omitting the required guanosine cofactor. But, in the presence of the guanosine cofactor, the splicing rate is set to $k_s = 2.5\ min^{-1}$.

The filled circles in blue in Fig. 3*B* are the time evolution data of native pre-RNA, $P_N(t)$, reported by Pan et al. (19), who measured $P_N(t)$ starting from the initial condition $P_I(0) = 1$, while blocking the self-splicing reaction. It was shown using theoretical arguments and experimental data, that a small fraction of pre-RNA ($\Phi \approx 0.1$) must fold rapidly to the N state with $k_{IN} = 60\ min^{-1}$, in accord with theoretical estimates (44, 52), and was observed subsequently in a single-molecule experiment (31). The remaining fraction transitions slowly into the N state in a biphasic manner (19), suggestive of two distinct ensembles of misfolded states (M_1 and M_2). In the absence of RNA chaperone, the pre-RNA in the M_1 and M_2 states transitions slowly to the N state with rates k_{M_1N} and k_{M_2N}. The values for the associated rate constants listed in Table 2 are obtained by fitting the measured time trace of the fraction of the native state to the solution of Eq. 7. Specifically, we minimize the sum of the squared residuals between the measurements and predicted curves, with the constraints $k_{IN} = 60\ min^{-1}$, $k_{IM_1} + k_{IM_2} = 540\ min^{-1}$, $k_{NM_1} = k_{M_1N}/10$, and $k_{NM_2} = k_{M_2N}/10$.

For the model shown in Fig. 3*A*, we assume that CYT-19, when present, binds to the M_1, M_2, and N states and induces unfolding transitions. For the parameters associated with CYT-19 dynamics, we used the values from Table 1, assuming that CYT-19-associated parameters for M_1 and M_2 are identical (for further detail, see Eq. **6** and Table 2). An increase in the CYT-19 activity facilitates the production of the N state at early times, but this trend does not continue to hold at later times (Fig. 3*B*). As a consequence, the steady-state yield of the native state (P^{ss}_N) depends nonmonotonically on $[C]_{tot}$. At low $[C]_{tot}$, when CYT-19 predominantly binds to the M state, P^{ss}_N increases with $[C]_{tot}$; however, at high $[C]_{tot}$, when substantial CYT-19-mediated unfolding of the N state ribozyme takes place, P^{ss}_N decreases with $[C]_{tot}$ (inset in Fig. 3*B*). This key prediction suggesting an optimal value of CYT-19 for folding of the ribozyme is an emergent property of our theory.

In the presence of the guanosine cofactor, pre-RNA in the N state self-splices irreversibly with the rate $k_s \approx 2.5\ min^{-1}$ (54).

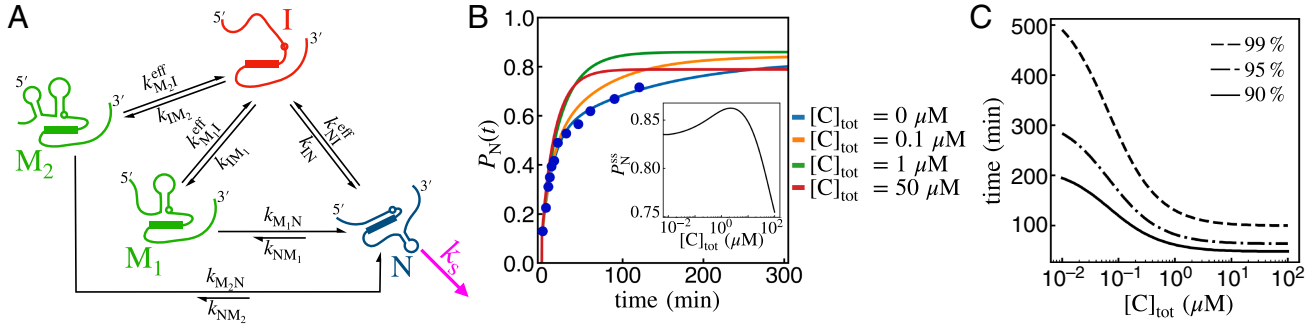

Fig. 3. CYT-19-assisted folding and self-splicing of pre-RNA. (*A*) Schematic network model for chaperone-assisted folding coupled to self-splicing of pre-RNA. The self-splicing reaction can occur at the splice site (open circle) and could be inhibited partly by the hairpin structure formed between the 5′ exon and the internal guide sequence (filled box) (50). *B* Fraction of pre-RNA in the N state, P_N, as a function of time, for varying concentrations of CYT-19. The inset in (*B*) shows the steady-state yield of native pre-RNA, P^{ss}_N, as a function of $[C]_{tot}$. (*C*) The first passage time t_{fpt} that satisfies **the probability of spliced state**, $P_{SP}(t_{fpt}) = 0.90$, 0.95, and 0.99, is shown as a function of $[C]_{tot}$. The values of t_{fpt} decrease sharply as $[C]_{tot}$ increases. All the curves are plotted using Eq. **7** with $[ATP]_{tot} = 2$ mM and $[R]_{tot} = 200$ nM, and the kinetic parameters are as described in Table 2. The data points in **B** (blue filled circles) are from ref. 19.

Table 2. Parameters for folding and self-splicing of pre-RNA in the presence of CYT-19

	k_{M_1N} (min^{-1})	k_{M_2N} (min^{-1})	k_{NM_1} (min^{-1})	k_{NM_2} (min^{-1})	k_{IM_1} (min^{-1})	k_{IM_2} (min^{-1})	k_{IN} (min^{-1})	k_{cat,M_1I} (min^{-1})	k_{cat,M_2I} (min^{-1})	$k_{cat,NI}$ (min^{-1})	K_{ATP} (μM)	K_{M_1} (μM)	K_{M_2} (μM)	K_N (μM)
Best-fit	0.07	0.007	0.007	0.0007	264	276	60	$11/S$	$11/S$	$1.4/S$	900	$50S$	$50S$	$5000S$

Since the folding of pre-RNA was measured at $[Mg^{2+}] = 6$ mM (19), the turnover rates and binding affinities associated with CYT-19 activity are adjusted using the factor $S = e^{m([Mg^{2+}]-2mM)}$ with $m \approx 1\ mM^{-1}$ (53), such that k_{cat,M_1I}, k_{cat,M_2I}, $k_{cat,NI}$, K_{M_1}, and K_{M_2} are scaled to $K_N \rightarrow k_{cat,M_1I}/S, k_{cat,M_2I}/S, k_{cat,NI}/S, K_{M_1}S, K_{M_2}S, K_NS$.

The probability that pre-RNA has undergone self-splicing can be calculated using $P_{SP}(t) = \int_0^t d\tau k_s P_N(\tau)$. Since the self-splicing reaction is irreversible and the system is closed, $P_{SP}(t \gg 1)$ reaches 1 in the steady states, for all CYT-19 concentrations. It is important to note that unlike P_N^{ss} that displays nonmonotonic dependence on $[C]_{tot}$ (inset of Fig. 3*B*), $P_{SP}(t)$ increases monotonically. The first passage time, t_{fpt}, that satisfies the conditions, $P_{SP}(t_{fpt}) = 0.90$, 0.95, and 0.99, decreases monotonically with $[C]_{tot}$ (Fig. 3*C*). This is because most of the pre-RNA in the N state would undergo self-splicing before being unfolded by CYT-19, with $k_s = 2.5\ min^{-1} \gg k_{cat,NI} \approx 0.02\ min^{-1}$. These results show that functional requirements override the kinetic criterion for folding. Hence, the ribozyme has evolved to optimize the rate of the splicing reaction.

Discussion

We generalized the IAM to characterize the chaperone-assisted folding of RNA and the associated consumption of ATP by chaperone activity. Our theory is consistent with data on the folding and splicing of ribozymes (Figs. 2 and 3) and highlights the competition between the chaperone-mediated unfolding and self-splicing reactions in the native state (Fig. 3). In the following sections, we further explore the important implications of our theory, including the thermodynamic cost of CYT-19 activity, and the effect of chaperone activity in RNA metabolism in vivo, where the yield of the spliced product is determined by the self-splicing, folding, degradation, and chaperone-mediated unfolding reactions.

Thermodynamic Cost of CYT-19 Activity Needed to Fold a Single Misfolded Ribozyme. Chaperone activity generated by ATP hydrolysis, which reshapes the distribution of RNA conformations in the rugged folding landscapes, is a nonequilibrium phenomenon (23, 35, 55). A quantity of great relevance, which goes to the heart of the quantities that are optimized in vivo, is the thermodynamic cost of correcting the errors due to the formation of nonnative interactions in a single misfolded ribozyme to the native state, which we quantify using the average number of ATPs consumed by CYT-19, $\langle A_{M\rightarrow N}\rangle$. By calculating the thermodynamic costs precisely, it is possible to quantitatively assess how cells balance the energy costs to fold ribozymes in order to execute the self-splicing reaction.

Since the spontaneous transitions from N to I or M are rare, we set $k_{NM} = k_{NI}^{eff} = 0$ in Eq. **4**. Counting only the ATPs consumed by the productively bound CYT-19, and setting the rate of ATP processing in the M state by $k_{ATP,M}^{eff}(t)$ (Eq. **S7**), we calculated the average number of ATP molecules consumed in converting M→N, $\langle A_{M\rightarrow N}\rangle$ as follows (see Eq. **S17** in *SI Appendix, Mean First Passage Time* for the derivation):

$$\begin{aligned}\langle A_{M\rightarrow N}\rangle &= \int_0^\infty k_{ATP,M}^{eff}(t)P_M(t)dt \\ &\approx \int_0^\infty k_{ATP,M}^{eff}(0)P_M(t)dt \\ &\approx \frac{A_{M\rightarrow I}}{\Phi + k_{MN}/k_{MI}^{eff}(0)},\end{aligned} \quad [9]$$

where $A_{M\rightarrow I} \equiv k_{ATP,M}^{eff}(0)/k_{MI}^{eff}(0)$ corresponds to the number of ATPs catalyzed in a single event of chaperone-mediated unfolding (M → I). To obtain the simplified expression in the last line of Eq. **9**, we assume that chaperone-mediated unfolding rates only vary weakly with time, which implies that $k_{MI}^{eff}(t) \approx k_{MI}^{eff}(0)$ and $k_{ATP,M}^{eff}(t) \approx k_{ATP,M}^{eff}(0)$. These approximations are validated by the similarity between the two $\langle A_{M\rightarrow N}\rangle$s (solid and dashed lines in Fig. 4 *A* and *B*) calculated based on the expressions in the first and the last lines of Eq. **9**.

With the best-fit parameters listed in Table 1, we compute $\langle A_{M\rightarrow N}\rangle$ for varying concentrations of CYT-19 and ATP (Fig. 4 *A* and *B*). $\langle A_{M\rightarrow N}\rangle$ increases with CYT-19 and ATP concentrations and saturates to a constant value $A_{M\rightarrow I}/\Phi \approx 125$, where $A_{M\rightarrow I} \approx 12.5$ and $\Phi = 0.1$. Our prediction is in good agreement with the experimental estimate (37). Thus, RNA chaperone consumes ATP lavishly in order to ensure that the ribozyme folds, which is not unlike the bacterial chaperonin, GroEL (23).

If the rate of transition M → N is negligible ($k_{MN} \ll k_{MI}^{eff}(0)$), the number of ATPs required to refold the misfolded ribozyme to the native state is $\langle A_{M\rightarrow N}\rangle \approx A_{M\rightarrow I}/\Phi \approx 125$ (Eq. **9**, Fig. 4). In comparison, for RuBisCO with $\Phi = 0.02 - 0.05$ (38), M → I transition demands consumption of 3–4 ATPs per cycle ($A_{M\rightarrow I} \approx 3 - 4$) (56) in the case of the bacterial GroES-GroEL system (57). Thus, we estimate that about 60–200 ATPs are required for GroEL to anneal a single misfolded RuBisCO to its native form. It was argued elsewhere (38) that the large number of ATP molecules needed to fold RuBisCo (or other proteins) is a small price to pay compared to the free energy cost needed to synthesize the protein in the first place. Similar reasoning holds for RNA synthesis as well.

With the assumption that $k_{MN} = k_{cat.MI} = 0$, we also obtain an expression for ATP consumption involving the chaperone-mediated transition of N → M,

$$\begin{aligned}\langle A_{N\rightarrow M}\rangle &= \int_0^\infty k_{ATP,N}^{eff}(t)P_N(t)dt \\ &\approx \frac{A_{N\rightarrow I}}{(1-\Phi) + k_{NM}/k_{NI}^{eff}(0)},\end{aligned} \quad [10]$$

with $A_{N\rightarrow I} \equiv k_{ATP,N}^{eff}(0)/k_{NI}^{eff}(0)$.

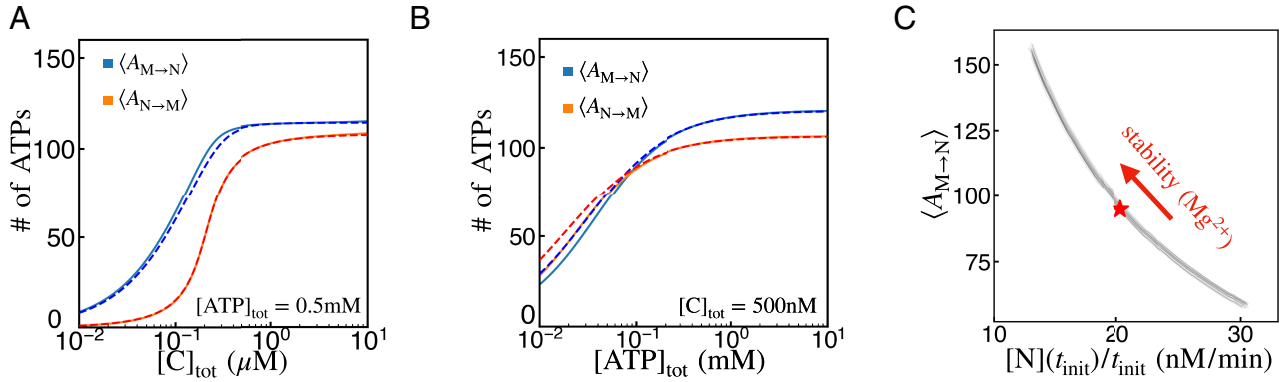

Fig. 4. ATP consumption during the CYT-19-assisted folding of ribozyme. (*A*, *B*) The average number of ATP molecules required to transform a single misfolded ribozyme to the native state ($\langle A_{M\to N}\rangle$), or a single native ribozyme to the misfolded state ($\langle A_{N\to M}\rangle$), as functions of (*A*) CYT-19 and (*B*) ATP concentrations (solid lines). The dashed lines in (*A*) and (*B*) are the approximate expressions in Eqs. **9** and **10**. (*C*) Plot of $\langle A_{M\to N}\rangle$ versus the initial rate of N state formation ($[N](t_{init})/t_{init}$), as functions of ribozyme stability. Ribozyme stability was modulated by the factor $S = e^{m\left([Mg^{2+}]-2\ mM\right)}$, with $m = 1\ mM^{-1}$ and $[Mg^{2+}]$ ranging from 1.5 to 2.5 mM. The star denotes the condition of $[Mg^{2+}] = 2$ mM ($S = 1$). The individual lines represent 20 of the parameter vectors sampled using MCMC, $\vec{\theta}$, with the constraint $\chi^2(\vec{\theta}_{min}) = 310.5 \le \chi^2(\vec{\theta}) < 313$, where $\vec{\theta}_{min}$ is the best-fit parameter vector shown in Table 1. For details on MCMC sampling, see *SI Appendix, MCMC Sampling*.

The large amount of ATP consumption by chaperones (Eqs. **9** and **10** and Fig. 4) raises a couple of points that are worthy of further discussion:

1. $A_{N\to I}(\approx 97) \gg A_{M\to I}(\approx 12.5)$ shows that more amount of ATPs is consumed for CYT-19 to disrupt the ribozyme in the N state than if the ribozyme is misfolded, which accords well with their stabilities. We also surmise that futile consumption of ATP must be prevalent in the helicase activity of CYT-19, which involves disruption of the stable secondary or tertiary motifs of ribozyme. The prediction that there are futile cycles is common to many molecular machines (42). It is also noteworthy that the ratio $A_{N\to I}/A_{M\to I}$ is similar to that between the concentrations of ribozyme in the N and M states at equilibrium, $[N]_{eq}/[M]_{eq} \approx 9$ (36).
2. Since $k_{MN}, k_{NM} \ll 1$, Eqs. **9** and **10** may be simplified as $\langle A_{M\to N}\rangle \approx A_{M\to I}/\Phi$ and $\langle A_{N\to M}\rangle \approx A_{N\to I}/(1-\Phi)$. From Figs. 4*A* and 4*B*, it follows that $\langle A_{M\to N}\rangle \approx \langle A_{N\to M}\rangle$. This shows that, although the cost of unfolding a ribozyme in the M state is smaller than unfolding a ribozyme in the N state, there are more ribozymes in the M state than in the N state ($\Phi \ll 1$). As a result, the average numbers of ATPs required by a chaperone to unfold a substrate molecule in the M and N states are comparable.

The more stable and compact an RNA is, the weaker should be chaperone binding to the ribozyme and less frequent would be the transitions between I, M, and N states. In experiments, RNA stability is controlled by changing the Mg^{2+} concentration ($[Mg^{2+}]$). Although $k_{ATP,P}$ and K_{ATP} do not change, Mg^{2+} affects the transition rates of RNA dynamics and binding affinities of CYT-19. The factor $S = e^{m\left([Mg^{2+}]-2\ mM\right)}$, where the m value ($m \approx 1\ mM^{-1}$) quantifies the effect of Mg^{2+} on the stability of ribozyme in the folded states (53), may be used to calibrate the transition rates and affinities of CYT-19 to ribozyme at Mg^{2+} concentrations other than at the physiological concentration of Mg^{2+}, $[Mg^{2+}] = 2$ mM. With increasing $[Mg^{2+}]$, $\langle A_{M\to N}\rangle$ increases, and the production rate of the N state measured at an early stage ($[N](t_{init})/t_{init}$) is reduced, which implies that the thermodynamic cost of creating RNA with enhanced stability is greater, resulting in the overall slowing down of the overall process (Fig. 4*C*). Our finding, summarized in Fig. 4*C*, is in good agreement with the trend observed in the experiment (Fig. 4A in ref. (37)).

IAM-Based Model For Self-Splicing of pre-RNA In Vivo. Upon folding to their respective native ribozyme structures, the introns undergo self-splicing reactions (25, 26, 58). If the folding and self-splicing processes are too slow, RNA can be removed either by degradation or dilution from cell growth (59, 60). In the following, we consider the effect of chaperone on the self-splicing "yield" of pre-RNA in vivo, which hinges on the competition between spontaneous folding, chaperone-mediated unfolding, degradation, and self-splicing reactions.

In our minimal model, upon transcription, pre-RNA rapidly reaches the I state, which is the entry point in the kinetic network that describes coupled folding and splicing (Fig. 3*A*). We assume that the pre-RNA transitions between the states I, M, N, and SP at the rates that were extracted through the analysis of the in vitro model and is degraded at a rate k_d from all the four states (Fig. 5*A*). The pre-RNA molecules evolve with time as

$$\frac{d\vec{c}(t)}{dt} = \mathcal{W}_s\vec{c}(t) + \vec{R}, \qquad [11]$$

where $\vec{c}(t) \equiv ([I](t), [M](t), [N](t), [SP](t))^T$,

$$\mathcal{W}_{o(t)} = \begin{bmatrix} -k_{IM}-k_{IN}-k_d & k_{MI}^{eff} & k_{NI}^{eff} & 0 \\ k_{IM} & -k_{MI}^{eff}-k_{MN}-k_d & k_{NM} & 0 \\ k_{IN} & k_{MN} & -k_{NI}^{eff}-k_{NM}-k_d-k_s & 0 \\ 0 & 0 & k_s & -k_d \end{bmatrix} \qquad [12]$$

and $\vec{R} = (r_{syn}, 0, 0, 0)^T$ is the input source current to the I state (see the cyan on the left in Fig. 5*A*) that represents the rate of pre-RNA being supplied upon transcription. The steady-state concentrations $[I]^{ss}$, $[M]^{ss}$, $[N]^{ss}$, and $[SP]^{ss}$ are obtained from $d\vec{c}(t)/dt = 0$ in Eq. **11**. The total concentration $[R]^{ss}_{tot} = [I]^{ss}+[M]^{ss}+[N]^{ss}+[SP]^{ss}$ can be equated as $[R]^{ss}_{tot} = r_{syn}/k_d$.

Experimental measurements (25, 59, 60) show that the self-splicing rate is greater than the folding (I → M or I → N) rates and that folding occurs more rapidly than the degradation ($k_d <$

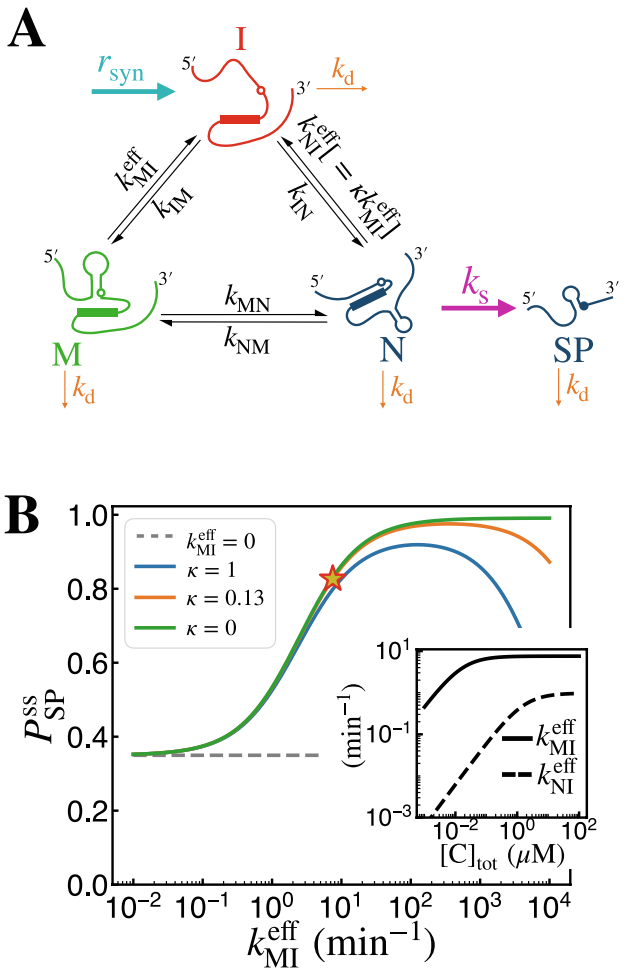

Fig. 5. Self-splicing in vivo. (*A*) Schematic for chaperone-mediated folding and self-splicing of pre-RNA in vivo. (*B*). Self-splicing yield (P_{SP}^{ss}) as a function of chaperone activity (k_{MI}^{eff}), at different values of the recognition factor κ. The parameters other than k_{MI}^{eff} and k_{NI}^{eff} $\left[= \kappa k_{MI}^{eff}\right]$ are set to $k_{IN} = 60\ min^{-1}$, $k_{IM} = 540\ min^{-1}$, $k_{MN} = 0.07\ min^{-1}$, $k_{NM} = 0.007\ min^{-1}$, $k_d = 0.18\ min^{-1}$, and $k_s = 30\ min^{-1}$. The inset in *B* shows the effective unfolding rates computed using Eq. **2**, where we assumed $[C] = [C]_{tot}$ and [ATP] = 2 mM and used the best-fit parameters in Table 1.

$k_{IN}, k_{IM} < k_s$). The chaperone activity is quantified by the effective rate constants, k_{MI}^{eff} and k_{NI}^{eff}, with implicit dependence on the ATP and chaperone concentrations. Here, we set the splicing rate to $k_s = 30\ min^{-1}$ that was measured in the cellular environment (61). With $k_d = 0.18\ min^{-1}$ (60), and the folding rates of pre-RNA, k_{IN}, k_{IM}, k_{MN}, and k_{NM} obtained from the fit to in vitro data (Fig. 3*B*), we calculate $P_{SP}^{ss}(= [SP]_{ss}/[R]_{tot}^{ss})$ for *Tetrahymena* pre-RNA as a function of k_{MI}^{eff} (Fig. 5*B*).

Our calculation shows that the chaperone activity can increase the steady-state yield of the spliced ribozyme, P_{SP}^{ss}, up to a certain level by facilitating ribozyme folding. However, when the chaperone activity exceeds a critical value, it leads to unfolding of the native ribozyme and reduces the P_{SP}^{ss}, especially when $\kappa > 0$, which engenders a nonmonotonic variation of P_{SP}^{ss} when the chaperone activity increases (Fig. 5*B*).

Under the condition of $k_{MN}, k_{NM} \ll 1$ and $k_d \ll k_{IM}, k_s$, the complicated expression for P_{SP}^{ss} simplifies greatly. For the full expression of P_{SP}^{ss} and the derivation of Eq. **13**, see *SI Appendix*, Eqs. S21 and S23 in *Self-splicing Yield*. It can be shown that P_{SP}^{ss} is indeed a nonmonotonic function of chaperone activity (k_{MI}^{eff}), which is maximized at

$$\left(k_{MI}^{eff}\right)^* \approx \sqrt{\frac{k_{IM}k_s}{\kappa}}. \quad [13]$$

Eq. **13** can be rearranged as $k_{MI}^{eff}k_{NI}^{eff} \approx k_{IM}k_s$, which offers greater insights into the conditions for the chaperone to maximize P_{SP}^{ss}. On the one hand, the chaperone-induced unfolding rate of the misfolded state should be sufficiently large compared to the rate of folding ($k_{MI}^{eff} \gg k_{IM}$), so that the pre-RNA can escape from the misfolded state. On the other hand, the unfolding rate of the native state (k_{NI}^{eff}) should not overwhelm the rate of self-splicing (k_s), which would cause the pre-RNA in the N state to unfold before splicing. The interplay between these opposing factors determines the optimal condition for function.

More generally, we can consider a pre-RNA under weakly stabilizing conditions, in which spontaneous unfolding from N or M to I state is significant, and chaperones are not required. Even in this case, our formulation based on the kinetic scheme in Fig. 5*A* and the optimal chaperone activity summarized in Eq. **13** are applicable by considering that k_{MI}^{eff} and k_{NI}^{eff} represent the sum of the spontaneous unfolding and chaperone-mediated unfolding rates of the pre-RNA ribozyme. Therefore, we surmise that "moderate" activity of CYT-19 is required to maximize the self-splicing yield of the pre-RNA, which is the condition that ought to be met for other downstream processes.

Although the value of k_{MI}^{eff} is varied over a broad range in Fig. 5*B*, k_{MI}^{eff} is upperbounded to $k_{cat,MI}$ (Eq. **2**). At [ATP] = 2 mM, it saturates to $k_{MI}^{eff} \approx 7.5\ min^{-1}$ (Fig. 5*B* inset). Along with k_{NI}^{eff}, we find that $\kappa(= k_{NI}^{eff}/k_{MI}^{eff}) = 0.13$ at the saturating CYT-19 concentration. Notably, the self-splicing yield calculated from $k_{MI}^{eff} \approx 7.5\ min^{-1}$ with $\kappa = 0.13$ is $P_{SP}^{ss} \simeq 0.83$ (star symbol in Fig. 5*B*), which is close to the maximum, $\left(P_{SP}^{ss}\right)_{max} = 0.97$ (the maximum point of the orange line in Fig. 5*B*). The discrepancy between the two numbers, albeit small, requires some discussion. It may simply reflect that $k_{cat,MI}$, the speed limit of k_{MI}^{eff}, is smaller than $(k_{MI}^{eff})^*$ in the case of CYT-19 acting on group I intron of pre-RNA. Otherwise, given the promiscuity of CTY-19 action on surface-exposed helices of structured RNAs, CYT-19 activity cannot be fully optimized for all the structured RNAs in terms of P_{SP}^{ss}. More specifically, for RNAs that fold correctly on their own with $\Phi \sim 1$, which is tantamount to $k_{IM} \sim 0$, the optimal chaperone activity, $(k_{MI}^{eff})^*$ (Eq. **13**), approaches 0. In this case, CYT-19 activity (k_{MI}^{eff} increased beyond $(k_{MI}^{eff})^*$) will lead to a reduction in the self-splicing yield while catalyzing the futile cycles of ATP hydrolysis. Thus, it may be argued that the extent of CYT-19 activity (star symbol in Fig. 5*B*) reflects the balance between facilitating the folding of RNA with Φ close to 0 and limiting the futile unfolding reactions on RNA with Φ close to 1.

Our theory that incorporates the splicing reaction in the folding network clarifies the effect of the recognition factor κ on the self-splicing yield, P_{SP}^{ss}. We find that the differential effect of κ on P_{SP}^{ss} manifests itself only at *hypothetical* values of k_{MI}^{eff} that exceed the turnover rate of CYT-19, $k_{cat,MI}$. An RNA chaperone with a smaller κ engenders a larger P_{SP}^{ss} for *Tetrahymena* pre-RNA if such a chaperone exists in nature (Fig. 5*B*, and *SI Appendix, Self-splicing Yield*). Alternatively, for a given chaperone, a ribozyme with smaller k_s and k_{IM} could

reduce the value of $\left(k_{\mathrm{MI}}^{\mathrm{eff}}\right)^{*}$ (Eq. **13**), rendering the differential effect of κ more pronounced. It is noteworthy that the product between the chaperone-mediated folding (relaxation) rate of the RNA molecule (λ) and the steady-state yield of the native state ($P_{\mathrm{N}}^{\mathrm{ss}}$), $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$, which was previously highlighted as the key quantity for optimization in the chaperone-facilitated folding, in principle, depends on κ in a self-consistent manner. It satisfies $\lambda P_{\mathrm{N}}^{\mathrm{ss}} = k_{\mathrm{IM}} k_{\mathrm{MI}}^{\mathrm{eff}}/(k_{\mathrm{IM}} + k_{\mathrm{MI}}^{\mathrm{eff}}) \times \left[\Phi + f(\kappa)\Phi^2 + \cdots\right]$ (see *SI Appendix*, Eq. **S28** and Fig. S3); however, when Φ is small, the dependence of $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ on κ is in effect negligible (see *SI Appendix, Dependence of* $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$ *on* κ for more details). In vitro measurements are consistent with $\Phi \approx 0.1$, and hence, the Φ^2 term is, for all practical purposes, negligible. For stringent protein substrates, Φ is on the order of 0.02 (38), which even more justifies the neglect of Φ^2 for protein chaperones. Hence, we posit that both RNA and protein chaperones evolved to maximize $\lambda P_{\mathrm{N}}^{\mathrm{ss}}$.

Finally, the effect of exon sequences on the folding kinetics of pre-RNA can potentially be significant (18, 62, 63) but was not considered in this study. However, their effects may be included by suitably modifying the appropriate rates. Nevertheless, our analyses and predictions, presented in Figs. 3, 4 and 5, are amenable to experimental test. The IAM-based framework presented here illuminates the interplay between chaperone-facilitated dynamics of ribozyme and self-splicing and could be further extended to analyze experimental data in more complex network models, which are likely required in cases involving an interplay between substrate folding, multiple chaperones, and function.

Summary

We developed a theory, based on kinetic network models that couples folding and catalysis, to determine the interplay between chaperone-mediated folding of pre-RNA and the downstream self-splicing reaction. We discovered a tension between chaperone activity and the functional requirement for sufficient yield of the self-spliced RNA. The chaperone activity must be large enough to unfold the kinetically trapped structure but not be too large to disrupt the native state because the latter is needed for the self-splicing reaction. Because the unwinding propensities of the misfolded and folded states of the pre-RNA are determined by their (kinetic) stabilities, it follows that the lifetime of the native state must be greater than the time needed for the splicing reaction to go to completion. The theory, which elucidates these ideas quantitatively, shows that RNA chaperones have evolved to satisfy the dual requirement of converting the misfolded structures to functionally competent folded states without disrupting the native state structures.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by the KIAS individual grants CG067102 (Y.S.) and CG035003 (C.H.) at Korea Institute for Advanced Study. D.T. thanks the National Science Foundation (CHE-1900033) and the Welch Foundation through the Collie-Welch Chair (F-0019) for supporting this research. We thank the Center for Advanced Computation in KIAS for providing computing resources.

1. T. R. Cech, A. J. Zaug, P. J. Grabowski, In vitro splicing of the ribosomal-RNA precursor of Tetrahymena-involvement of a quianosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496 (1981).
2. K. Kruger *et al.*, Self-splicing RNA - Auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).
3. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, The RNA moiety of Ribonuclease-P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
4. D. Thirumalai, C. Hyeon, RNA and Protein folding: Common Themes and Variations. *Biochemistry* **44**, 4957–4970 (2005).
5. S. A. Woodson, Taming free energy landscapes with RNA chaperones. *RNA Biol.* **7**, 677–686 (2010).
6. A. Serganov, D. J. Patel, Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).
7. B. S. Strunk, K. Karbstein, Powering through ribosome assembly. *RNA* **15**, 2083–2104 (2009).
8. E. D. Egan, K. Collins, Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–1759 (2012).
9. I. Jarmoskaite, R. Russell, RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* **83**, 697–725 (2014).
10. H. M. Al-Hashimi, N. G. Walter, RNA dynamics: It is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).
11. C. Hyeon, J. Lee, J. Yoon, S. Hohng, D. Thirumalai, Hidden complexity in the isomerization dynamics of holliday junctions. *Nat. Chem.* **4**, 907–914 (2012).
12. M. Hinczewski, C. Hyeon, D. Thirumalai, Directly Measuring Single Molecule Heterogeneity using Force Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3852–E3861 (2016).
13. D. Kowerko *et al.*, Cation-induced kinetic heterogeneity of the intron-exon recognition in single group ii introns. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3403–3408 (2015).
14. Z. Guo, D. Thirumalai, Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* **36**, 83–102 (1995).
15. D. Thirumalai, D. K. Klimov, S. A. Woodson, Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* **96**, 14–22 (1997).
16. B. L. Bass, T. R. Cech, Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: Implications for biological catalysis by RNA. *Nature* **308**, 820–826 (1984).
17. V. L. Emerick, S. A. Woodson, Self-Splicing of the Tetrahymena Pre-rRNA Is Decreased by Misfolding during Transcription. *Biochemistry* **32**, 14062–14067 (1993).
18. V. L. Emerick, J. Pan, S. A. Woodson, Analysis of rate-determining conformational changes during self-splicing of the Tetrahymena intron. *Biochemistry* **35**, 13469–13477 (1996).
19. J. Pan, D. Thirumalai, S. A. Woodson, Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7–13 (1997).
20. D. Herschlag, RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).
21. D. K. Treiber, J. R. Williamson, Beyond kinetic traps in RNA folding. *Curr. Op. Struct. Bio.* **11**, 309–314 (2001).
22. S. A. Woodson, Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).
23. S. Chakrabarti, C. Hyeon, X. Ye, G. H. Lorimer, D. Thirumalai, Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10919–E10927 (2017).
24. B. Turcq, K. F. Dobinson, N. Serizawa, A. M. Lambowitz, A protein required for rna processing and splicing in neurospora mitochondria is related to gene products involved in cell cycle protein phosphatase functions. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1676–1680 (1992).
25. S. Mohr, J. M. Stryker, A. M. Lambowitz, A dead-box protein functions as an atp-dependent rna chaperone in group i intron splicing. *Cell* **109**, 769–779 (2002).
26. H. R. Huang *et al.*, The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl. Acad. Sci. USA* **102**, 163–168 (2005).
27. P. Gologuinoff, A. A. Gatenby, G. H. Lorimer, GroEL heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia-Coli. *Nature* **337**, 44–47 (1989).
28. P. V. Viitanen, A. A. Gatenby, G. H. Lorimer, Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. *Protein Sci.* **1**, 363–369 (1992).
29. J. H. Cate, E. A. Doherty, J. A. Doudna, Crystal structure of A group I ribozyme domain reveals principles of higher order RNA folding. *Science* **273**, 1678–1685 (1996).
30. W. D. Downs, T. R. Cech, Kinetic pathway for folding of the Tetrahymena ribozyme revealed by three UV-inducible crosslinks. *RNA* **2**, 718–732 (1996).
31. X. Zhuang *et al.*, A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048–2051 (2000).
32. R. Russell *et al.*, Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA* **99**, 155–160 (2002).
33. P. Tijerina, H. Bhaskaran, R. Russell, Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc. Natl. Acad. Sci. USA* **103**, 16698–16703 (2006).
34. J. K. Grohman *et al.*, Probing the mechanisms of DEAD-box proteins as general RNA chaperones: The C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* **46**, 3013–3022 (2007).
35. H. Bhaskaran, R. Russell, Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* **449**, 1014–1018 (2007).
36. I. Jarmoskaite, H. Bhaskaran, S. Seifert, R. Russell, DEAD-box protein CYT-19 is activated by exposed helices in a group I intron RNA. *Proc. Natl. Acad. Sci. USA* **111**, E2928–E2936 (2014).
37. I. Jarmoskaite, P. Tijerina, R. Russell, ATP utilization by a DEAD-box protein during refolding of a misfolded group I intron ribozyme. *J. Biol. Chem.* **296**, 100132. (2021).
38. M. J. Todd, G. H. Lorimer, D. Thirumalai, Chaperonin-facilitated protein folding: Optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl. Acad. Sci. USA* **93**, 4030–4035 (1996).
39. C. Hyeon, D. Thirumalai, Generalized iterative annealing model for the action of RNA chaperones. *J. Chem. Phys* **139**, 121924. (2013).
40. D. Thirumalai, G. H. Lorimer, Chaperonin-Mediated Protein Folding. *Annu. Rev. Biophys.* **30**, 245–269 (2001).
41. D. Thirumalai, G. H. Lorimer, C. Hyeon, Iterative annealing mechanism explains the functions of the GroEL and RNA chaperones. *Prot. Sci.* **29**, 360–377 (2020).
42. M. L. Mugnai, C. Hyeon, M. Hinczewski, D. Thirumalai, Theoretical perspectives on biological machines. *Rev. Mod. Phys.* **92**, 025001 (2020).
43. Z. Su *et al.*, Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature* **596**, 603–607 (2021).

44. D. Thirumalai, N. Lee, S. A. Woodson, D. K. Klimov, Early Events in RNA Folding. *Annu. Rev. Phys. Chem.* **52**, 751–762 (2001).
45. R. Russell, D. Herschlag, New pathways in folding of the Tetrahymena group I RNA enzyme. *J. Mol. Biol.* **291**, 1155–1167 (1999).
46. R. Russell, D. Herschlag, Probing the folding landscape of the Tetrahymena ribozyme: Commitment to form the native conformation is late in the folding pathway. *J. Mol. Biol.* **308**, 839–851 (2001).
47. Y. Wan, H. Suh, R. Russell, D. Herschlag, Multiple Unfolding Events during Native Folding of the Tetrahymena Group I Ribozyme. *J. Mol. Biol.* **400**, 1067–1077 (2010).
48. R. Russell, I. Jarmoskaite, A. M. Lambowitz, Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol.* **10**, 44–55 (2013).
49. A. Henn, M. J. Bradley, E. M. De La Cruz, ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annu. Rev. Biophys.* **41**, 247–267 (2012).
50. S. A. Woodson, Exon sequences distant from the splice junction are required for efficient self-splicing of the Tetrahymena IVS. *Nuc. Acids. Res.* **20**, 4027–4032 (1992).
51. T. R. Cech, Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**, 543–568 (1990).
52. D. Thirumalai, From minimal models to real proteins: Time scales for protein folding kinetics. *J. Phys. I (Fr.)* **5**, 1457–1467 (1995).
53. J. Pan, D. Thirumalai, S. A. Woodson, Magnesium-dependent folding of self-splicing RNA: Exploring the link between cooperativity, thermodynamics, and kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6149–6154 (1999).
54. V. L. Emerick, J. Pan, S. A. Woodson, Analysis of rate-determining conformational changes during self-splicing of the Tetrahymena intron. *Biochemistry* **35**, 13469–13477 (1996).
55. P. Goloubinoff, A. S. Sassi, B. Fauvet, A. Barducci, P. D. Rios, Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. *Nat. Chem. Biol.* **14**, 388–394 (2018).
56. M. Hayer-Hartl, A. Bracher, F. U. Hartl, The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *TiBS* **41**, 62–76 (2016).
57. D. Balchin, M. Hayer-Hartl, F. U. Hartl, Recent advances in understanding catalysis of protein folding by molecular chaperones. *FEBS Lett.* **594**, 2770–2781 (2020).
58. J. P. Potratz, M. Del Campo, R. Z. Wolf, A. M. Lambowitz, R. Russell, ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing in vitro and in vivo. *J. Mol. Biol.* **411**, 661–679 (2011).
59. S. P. Koduvayur, S. A. Woodson, Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice. *RNA* **10**, 1526–1532 (2004).
60. S. A. Jackson, S. Koduvayur, S. A. Woodson, Self-splicing of a group I intron reveals partitioning of native and misfolded RNA populations in yeast. *RNA* **12**, 2149–2159 (2006).
61. S. L. Brehm, T. R. Cech, Fate of an intervening sequence ribonucleic acid: Excision and cyclization of the tetrahymena ribosomal ribonucleic acid intervening sequence in vivo. *Biochemistry* **22**, 2390–2397 (1983).
62. J. Pan, S. A. Woodson, Folding intermediates of a self-splicing RNA: Mispairing of the catalytic core. *J. Mol. Biol.* **280**, 597–609 (1998).
63. T. Nikolcheva, S. A. Woodson, Facilitation of group I splicing in vivo: Misfolding of the Tetrahymena IVS and the role of ribosomal RNA exons. *J. Mol. Biol.* **292**, 557–567 (1999).