Supporting Information

Chakrabarti et al. 10.1073/pnas.1712962114

SI Solution of the Three-State Model

Denoting by $\vec{P}(t)$ a column vector with components $P(i,t)$, Eq. **2** can be cast into a matrix form: $\frac{d}{dt}\vec{P}(t) = W\vec{P}(t)$, where the rate matrix W is

$$W = \begin{pmatrix} -(k_{\mathrm{IN}} + k_{\mathrm{IM}}) & k_{\mathrm{NI}} & k_{\mathrm{MI}} \\ k_{\mathrm{IN}} & -(k_{\mathrm{NI}} + k_{\mathrm{NM}}) & k_{\mathrm{MN}} \\ k_{\mathrm{IM}} & k_{\mathrm{NM}} & -(k_{\mathrm{MI}} + k_{\mathrm{MN}}) \end{pmatrix} \quad \textbf{[S1]}$$

For the rate matrix in Eq. **S1**, $\vec{P}(t)$ is given by:

$$\vec{P}(t) = c_1\vec{u}_1 e^{\lambda_1 t} + c_2\vec{u}_2 e^{\lambda_2 t} + c_3\vec{u}_3 e^{\lambda_3 t} \quad \textbf{[S2]}$$

where λ_1, λ_2, and λ_3 are the eigenvalues of W and $\vec{u}_1$, $\vec{u}_2$, and $\vec{u}_3$ are the corresponding eigenvectors. The first eigenvalue λ_1 is equal to 0, while $\lambda_2, \lambda_3 < 0$ (1). Since $\lambda_1 = 0$ (and the other two eigenvalues are negative), $c_1\vec{u}_1$ represents the steady-state ($t \to \infty$) solution of Eq. **S2**. If $|\lambda_2| \ll |\lambda_3|$, Eq. **S2** effectively describes a single exponential relaxation to the steady state, with $|\lambda_2|$ the observed rate for the relaxation—that is, $|\lambda_2| \approx k_{obs}$. The coefficients c_1, c_2, and c_3 are constants determined from the initial conditions—the fraction of substrate in I, M, and N at time $t = 0$. In all of the experiments analyzed later, three types of initial conditions arise: All of the substrate begins in state M [P(M, 0) = 1], all of the substrate begins in state N [P(N, 0) = 1], and the substrate is in a mixture of states. The last initial condition was needed only for analyzing some of the ribozyme data (Figs. 3 and 4), and for these cases, P(N, 0) was obtained directly from the experimental data at $t = 0$ and P(M, 0) was set to $1 - P(\mathrm{N}, 0)$.

SI The Long-Time Steady State is Far from Equilibrium

To assess whether the long-term steady state of Eq. **S2** is an equilibrium or nonequilibrium solution, we calculate the probability current between any two states of the model. This current J is the same between any two states of our three-state model and is defined as:

$$J = k_{ij}P(i,\infty) - k_{ji}P(j,\infty), \quad \textbf{[S3]}$$

where $i, j =$ I, M, N. Using Eq. **S2** with $t \to \infty$, the current J is given by:

$$J = \frac{k_{\mathrm{IM}}k_{\mathrm{MN}}k_{\mathrm{NI}} - k_{\mathrm{IN}}k_{\mathrm{MI}}k_{\mathrm{NM}}}{(k_{\mathrm{IM}} + k_{\mathrm{MI}} + k_{\mathrm{MN}})k_{\mathrm{NI}} + k_{\mathrm{MI}}k_{\mathrm{NM}} + k_{\mathrm{IM}}(k_{\mathrm{MN}} + k_{\mathrm{NM}}) + k_{\mathrm{IN}}(k_{\mathrm{MI}} + k_{\mathrm{MN}} + k_{\mathrm{NM}})}. \quad \textbf{[S4]}$$

The steady state reached is out of equilibrium, as in the long time limit a nonzero (but constant) probability current J exists between any two states of the system. As is evident from Eq. **S4**, only when both $k_{\mathrm{NI}} = 0$ and $k_{\mathrm{MI}} = 0$, which is realized either when [C] = 0 or [T] = 0 or both, the current becomes zero, which is characteristic of equilibrium.

SI Parameter Estimates for RNA and Proteins

To analyze the fraction of native substrate as a function of time, we fit Eq. **S2** to experimental data using a custom-written nonlinear least squares method in Mathematica (2). The least squares method of minimizing χ^2 values is equivalent to maximizing a log-likelihood function, with the assumption that errors in the mean fraction of native substrate are Gaussian-distributed. As can be seen from Tables S1–S3, a number of parameters could not be uniquely identified from the fits. To find ranges of these parameters over which the fits did not appreciably change, we varied each parameter individually around the best fit value, keeping all other parameters fixed, and computed the change in the minimum χ^2. We reported parameter ranges that allowed the χ^2 value to increase by 1 from the minimum. Quantitatively, this means that a parameter resulting in an increase of 1 for the χ^2 is $\exp(-1/2) \sim 60\%$ as likely to be correct as the best fit parameter value.

SI Δ_{NE} for GroEL-Mediated Folding of MDH

Fig. S1 shows a plot of Δ_{NE} for the protein MDH as a function of GroEL concentration. Similar to Rubisco and *Tetrahymena* ribozyme analyzed in the main text, this is an increasing function of chaperone concentration, saturating at GroEL concentrations of approximately $2 - 3\mu M$. This finding and the observation that the in vivo GroEL concentration is approximately 5.2 μM (see *Maximization of the Finite-Time Yield by Iterative Annealing and in Vivo Regulation of Chaperone Concentration*) further support our prediction that GroEL maximizes Δ_{NE}.

SI Predictions for Possible Mutations

Having obtained the best fit parameters for the ribozymes and proteins (Table S1– S3), we can now modify the rates of some of the important parameters to predict the outcome of mutations that could conceivably be performed. The results, shown in Fig. S2 for the WT ribozyme and Fig. S3 for MDH, suggest that the most sensitive mutation would be one that changes the binding of the chaperone to native ribozyme or protein. Interestingly, our analysis predicts that other possible mutations (which would change ATP hydrolysis rates or binding rates of chaperone to the misfolded ribozyme/protein) would hardly change the final yield of N states.

1. van Kampen NG. (2007) Stochastic Processes in Chemistry and Physics. (Elsevier).

2. Wolfram Research Inc. (2017) Mathematica, Version 11.1. Available at wolfram.com. Accessed December 4, 2017.

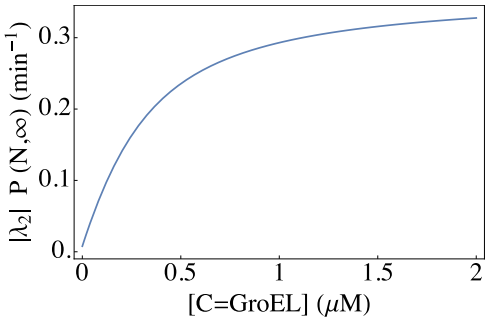

Fig. S1. Plot of $\Delta_{NE} = |\lambda_2|P(N,\infty)$ as a function of GroEL concentration (7- or 14-mers of GroEL). The concentration of MDH was 0.5μM.

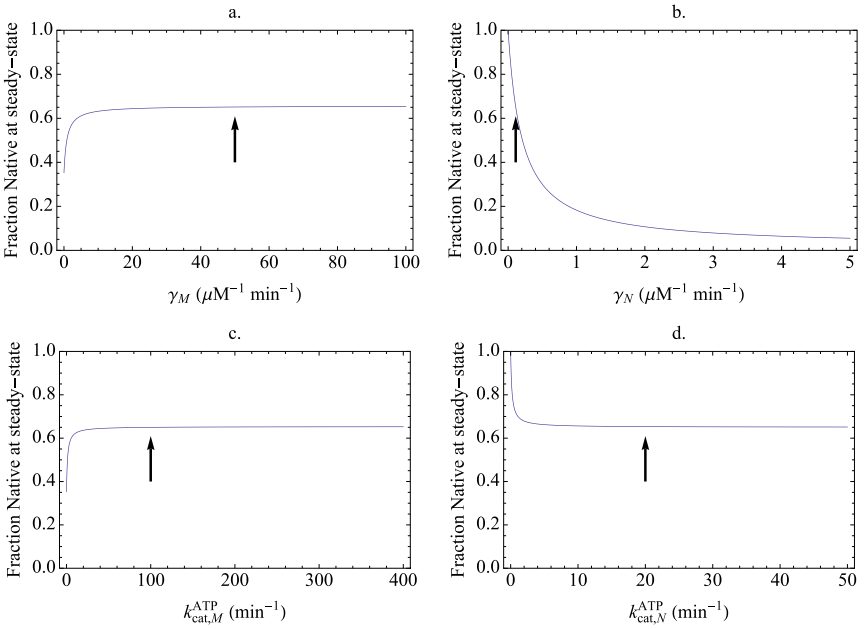

Fig. S2. Effect of possible mutations on the final yield of N states in the WT ribozyme–CYT-19 complex. Parameters were fixed to the best fit results with [C] = 1 μM, [T] = 2,000 μM, and only γ_M (*A*), γ_N (*B*), $k_{cat,M}^{ATP}$ (*C*), and $k_{cat,N}^{ATP}$ (*D*) were varied to observe the effect on the fraction of N states. Arrows indicate the position of the best fit value around which the parameter is being varied.

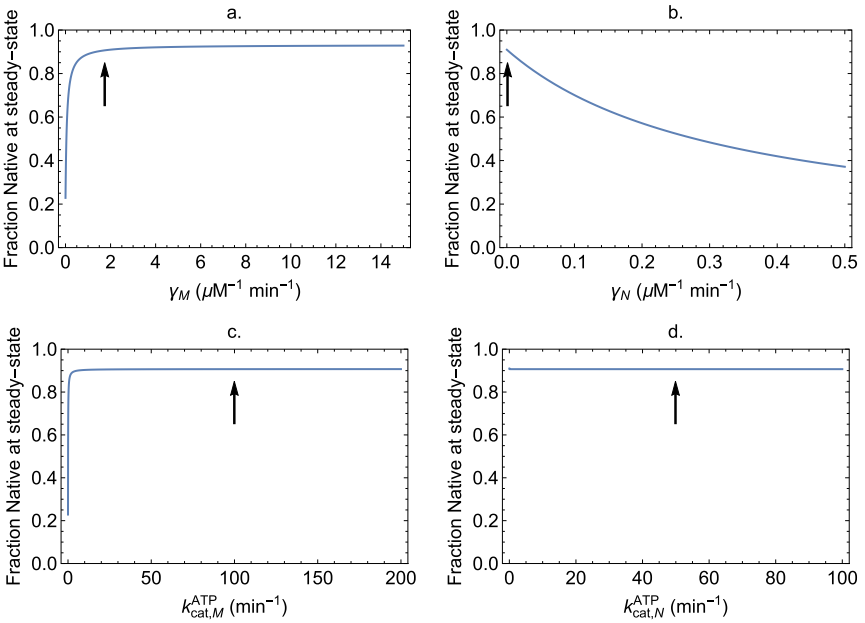

Fig. S3. Effect of possible mutations on the final yield of N states in the MDH–GroEL complex. Parameters were fixed to the best fit results with [C] = 1 μM, [T] = 2,000 μM, and only γ_M (*A*), γ_N (*B*), $k^{ATP}_{cat,M}$ (*C*), and $k^{ATP}_{cat,N}$ (*D*) were varied to observe the effect on the fraction of N states. Arrows indicate the position of the best fit value around which the parameter is being varied.

Table S1. Best fit parameters extracted by fitting Eq. S2 to the experimental data (Eq. S3) obtained at 25 °C and 1 mM (WT) and 5 mM (P5a and $E^{\Delta P5abc}$) Mg^{2+} ion concentrations

Best fit	k_{IN}, min^{-1}	k_{IM}, min^{-1}	γ_N, M^{-1}min^{-1}	γ_M, M^{-1}min^{-1}	$k^{ATP}_{cat,N}$, min^{-1}	$k^{ATP}_{cat,M}$, min^{-1}	$K^{ATP}_{m,N}$, μM	$K^{ATP}_{m,M}$, μM	k_{NM}, min^{-1}	k_{MN}, min^{-1}
WT, fit	0.2*	1.6*	1.1×10^5	$\geq 50 \times 10^6$	> 10	> 10	> 1	> 1	10^{-6}	0.05
WT, exp	0.37 ± 0.10 (1)	1.5 ± 0.3 (1), 2.4†				300 to 600 (2)		50 to 500 (2)	10^{-9} (3)‡	0.05 (4)§
P5a var. (fit)	0.7	5.11 to 6.1	13.6×10^5	$\geq 15 \times 10^7$	7 to 30	100 to 400	> 3000	700 to 900	0.002	0.032
P5a var., exp	4.75¶ (4)	51.13†								
$E^{\Delta P5abc}$, fit	1.8*	11.07*	6.8×10^6	$\geq 20 \times 10^6$	> 2	> 2	> 800	> 800	0.0615	0.0614
$E^{\Delta P5abc}$, exp	0.28 (3)	1.7†							0.008 (3)‡	0.011 (3)‡

For comparison, we list the corresponding rates from direct experimental measurements (in bold) and the indirect (details in the footnotes below) estimates. The experimental rates were determined at 25 °C but with different Mg^{2+} ion concentrations, 10 mM in ref. 1 and 10 to 50 mM in ref. 3.

*Because of insufficient data for extracting unique parameter values, we constrained the ratio of k_{IM} and k_{IN} for the WT and $E^{\Delta P5abc}$ while fitting to reproduce experimental results for Φ (1, 5–7).

†These values of k_{IM} have been estimated from the experimentally determined values of k_{IN} and Φ.

‡The rates $k_{MN} = 10^{-4}\ min^{-1}$ and $k_{NM} = 10^{-9}\ min^{-1}$ were obtained in experiments without the chaperone CYT-19. See footnote d for more details.

§This value of k_{MN} obtained in ref. 4 is higher than what was found in experiments in the absence of chaperone (3). The reason, as discussed in supplementary figure 2 of ref. 4, is the presence of a KCl buffer in the preparation of CYT-19.

¶This rate was not obtained directly at 5 mM Mg^{2+} ion concentration but obtained by linear extrapolation from higher Mg^{2+} concentrations (4).

1. Russell R, Herschlag D (1999) New pathways in folding of the Tetrahymena group I RNA enzyme. *J Mol Biol* 291:1155–1167.
2. Russell R, Jarmoskaite I, Lambowitz AM (2013) Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol* 10:44–55.
3. Russell R, Tijerina P, Chadee AB, Bhaskaran H (2007) Deletion of the p5abc peripheral element accelerates early and late folding steps of the Tetrahymena group i ribozyme. *Biochemistry* 46:4951–4961.
4. Bhaskaran H, Russell R (2007) Kinetic redistribution of native and misfolded RNAs by dead-box chaperone. *Nature* 449:1014–1018.
5. Pan J, Thirumalai D, Woodson SA (1997) Folding of RNA involves parallel pathways. *J Mol Biol* 273:7–13.
6. Zhuang X, et al. (2000) A single-molecule study of RNA catalysis and folding. *Science* 288:2048–2051.
7. Thirumalai D, Lee N, Woodson SA, Klimov DK (2001) Early events in RNA folding. *Annu Rev Phys Chem* 52:751–762.

Table S2. Best fit parameters determined by fitting Eq. S2 to experimental data on GroEL-mediated folding of Rubisco (1)

Fit	k_{IN}, min^{-1}	k_{IM}, min^{-1}	γ_N, $M^{-1}min^{-1}$	γ_M, $M^{-1}min^{-1}$	$k^{ATP}_{cat,N}$, min^{-1}	$k^{ATP}_{cat,M}$, min^{-1}	$K^{ATP}_{m,N}$, μM	$K^{ATP}_{m,M}$, μM	k_{NM}, min^{-1}	k_{MN}, min^{-1}
Value	0.25	20.36	27×10^5	389×10^6	0.02	$48 - 52$	< 40	$60 - 145$	0.0094 - 0.01	$10^{-6} - 2 \times 10^{-4}$

1. Todd MJ, Lorimer GH, Thirumalai D (1996) Chaperonin-facilitated protein folding: Optimization of rate and yield by an iterative annealing mechanism. *Proc Natl Acad Sci USA* 93:4030–4035.

Table S3. Best fit parameters determined by fitting Eq. S2 to experimental data on GroEL-mediated folding of MDH

Fit	k_{IN}, min^{-1}	k_{IM}, min^{-1}	γ_N, $M^{-1}min^{-1}$	γ_M, $M^{-1}min^{-1}$	$k^{ATP}_{cat,N}$, min^{-1}	$k^{ATP}_{cat,M}$, min^{-1}	$K^{ATP}_{m,N}$, μM	$K^{ATP}_{m,M}$, μM	k_{NM}, min^{-1}	k_{MN}, min^{-1}
Value	0.366	0.37	$< 10^{3*}$	$1.7 \times 10^{6*}$	$> 0.1^{\dagger}$	$> 20^{\dagger}$	$< 2000^{\ddagger}$	$< 2000^{\ddagger}$	0.025	7.78×10^{-3}

*Because of insufficient data for extracting unique parameter values, the additional constraint of $\gamma_M > \gamma_N$ was maintained while fitting to data. This is physically reasonable since the chaperone is expected to bind more efficiently to misfolded protein rather than native protein and was found to be the case for the unconstrained fit for Rubisco shown in Table S2.

†Because of insufficient data for extracting unique parameter values, the k_{cat} values were constrained to be less than 5000 min^{-1}, since most enzymes fall in this range (1).

‡Because of insufficient data for extracting unique parameter values, the K_m values were constrained to be less than 10000 μM, since most enzymes fall in this range (1).

1. Bar-Even A, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–4410.