# Chain Length Determines the Folding Rates of RNA

Changbong Hyeon[†*] and D. Thirumalai[‡*]
[†]School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea; and [‡]Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland

ABSTRACT We show that the folding rates ($k_F$s) of RNA are determined by $N$, the number of nucleotides. By assuming that the distribution of free-energy barriers separating the folded and the unfolded states is Gaussian, which follows from central limit theorem arguments and polymer physics concepts, we show that $k_F \approx k_0 \exp(-\alpha N^{0.5})$. Remarkably, the theory fits experimental rates spanning over 7 orders of magnitude with $k_0 \sim 1.0(\mu s)^{-1}$. Our finding suggests that the speed limit of RNA folding is ~1 ms, just as it is in the folding of globular proteins.

Received for publication 30 November 2011 and in final form 5 January 2012.

*Correspondence: hyeoncb@kias.re.kr or thirum@umd.edu

RNA molecules are evolved biopolymers whose folding has attracted a great deal of attention (1–3) because of the crucial role they play in a number of cellular functions. The slightly branched polymeric nature of RNA implies that the shapes, relaxation dynamics, and even their folding rates must depend on $N$. In support of this assertion, it has been shown that the radius of gyration of the folded states, obtained with the use of data available in the Protein Data Bank, scales as $R_g \sim 5.5N^{\nu}$ Å, where the Flory exponent $\nu$ varies from 0.33 to 0.40 (4–6). Although this result is expected from the perspective of polymer physics, it is surprising from the viewpoint of structural biology because one might argue that the sequence and complexity of secondary and tertiary structure organization could lead to substantial deviations from the predictions based on Flory-like theory. Here, we show that the folding rates, $k_F$s, of RNA are also primarily determined by $N$, thus adding to the growing evidence that it is possible to understand RNA folding by using polymer physics principles.

## THEORETICAL CONSIDERATIONS

Theoretical arguments based on the dynamics of activated transitions in disordered systems suggest that

$$k_F = k_0 \exp\left(-\alpha N^{\beta}\right), \quad (1)$$

where $\beta$ should be 0.5 (7). The rationale for this finding hinges on the observation that favorable basepairing interactions and the hydrophobic nature of the bases tend to collapse RNA, whereas the charged phosphate residues are better accommodated by extended structures. Thus, the distribution of activation free energy, $\Delta G^{\ddagger}_{UF}/k_BT$, between the folded and unfolded states is a sum of favorable and unfavorable terms. We expect from central limit theorem that the distribution of $\Delta G^{\ddagger}_{UF}/k_BT$ should be roughly Gaussian with dispersion $\left\langle (\Delta G^{\ddagger}_{UF})^2 \right\rangle \sim N$. Thus, $\Delta G^{\ddagger}_{UF}/k_BT \sim N^{\beta}$ with $\beta = 1/2$.

We analyzed the available experimental data (see Table 1 for a list of RNA molecules) on RNA folding rates by assuming that $\Delta G^{\ddagger}_{UF}$ grows as $N^{\beta}$ with $\beta$ as a free parameter. The theoretical value for $\beta$ is 0.5. The folding rates of RNA spanning over 7 orders of magnitude is well fit using $\log k_F = \log k_0 - \alpha N^{\beta}$ with a correlation coefficient of 0.98 (Fig. 1). The fit yields $k_0^{-1} = 0.87$ $\mu$s, $\alpha = 0.91$, and $\beta \approx 0.46$. In the inset we show the fit obtained by fixing $\beta = 0.5$. Apart from the moderate differences in the $k_0^{-1}$ values, the theoretical prediction and the numerical fits are in agreement, which demonstrates that the major determining factor in determining RNA folding rates is $N$.

It is known that RNAs, such as *Tetrahymena* ribozyme, fold by multiple pathways that are succinctly described by the kinetic partitioning mechanism (8). According to this mechanism, a fraction, $\Phi$, of molecules reaches the native states rapidly and the remaining fraction is trapped in an ensemble of misfolded intermediates. For *Tetrahymena* ribozyme $\Phi \sim 0.1$ (9). The $N$ dependence given by Eq. 1 holds for the majority of molecules that fold to the native state from the compact intermediates, which form rapidly under folding conditions (10).

## CONCLUSIONS

Our findings indicate that the inverse of the prefactor, $k_0^{-1} = \tau_0 \approx 0.87$ $\mu$s, is almost 6 orders of magnitude larger than the transition-state theory estimate of $h / k_BT \approx 0.16$ ps. The value of $\tau_0$, which coincides with the typical basepairing time (11), is the speed limit for RNA folding. Of

Editor: Kathleen Hall.
© 2012 by the Biophysical Society
doi: 10.1016/j.bpj.2012.01.003

**TABLE 1 RNA length versus folding rate**

| RNA | $N$ | $k_f$ (sec$^{-1}$) | |
|---|---|---|---|
| GCUUCGGC (16) | 8 | $6.7 \times 10^4$ | Tetraloop hairpin |
| GCUUCGGC (16) | 8 | $27.2 \times 10^4$ | Tetraloop hairpin |
| GGUUCGCC (16) | 8 | $1.3 \times 10^4$ | Tetraloop hairpin |
| GGUUCGCC (16) | 8 | $4.7 \times 10^4$ | Tetraloop hairpin |
| GGACUUUUGUCC (16) | 12 | $6.1 \times 10^4$ | Tetraloop hairpin |
| GGACUUCGGUCC (16) | 12 | $4.5 \times 10^4$ | Tetraloop hairpin |
| $A_6C_6U_6$ (17) | 18 | $3.4 \times 10^4$ | Tetraloop hairpin |
| Extra arm of tRNA$^{ser}$ (yeast) (18) | 21 | $1 \times 10^5$ | tRNA |
| pG half of tRNA$^{Phe}$ (yeast) (18) | 36 | $9 \times 10^3$ | tRNA |
| CCA half of tRNA$^{Phe}$ (yeast) (18) | 39 | $8.5 \times 10^3$ | tRNA |
| CCA half of tRNA$^{Phe}$ (wheat) (18) | 39 | $8 \times 10^3$ | tRNA |
| tRNA$^{Phe}$ (yeast) (19) | 76 | $5.3 \times 10^2$ | tRNA |
| tRNA$^{Ala}$ (yeast) (18) | 77 | $9 \times 10^2$ | tRNA |
| $Y_4$ hairpin (20) | 14 | $5.75 \times 10^4$ | Hairpin ($5 \times 2 + 4$) |
| $Y_9$ hairpin (20) | 19 | $2.29 \times 10^4$ | Hairpin ($5 \times 2 + 9$) |
| $Y_{19}$ hairpin (20) | 29 | $8.70 \times 10^2$ | Hairpin ($5 \times 2 + 19$) |
| $Y_{34}$ hairpin (20) | 44 | $6.03 \times 10^2$ | Hairpin ($5 \times 2 + 34$) |
| VPK pseudoknot (21) | 34 | $9.09 \times 10^2$ | Pseudoknot |
| Hairpin ribozyme (four-way junction) (22,23) | 125 | 6 | Natural form of hairpin ribozyme |
| P5abc (24) | 72 | 50 | Group I intron *T.* ribozyme |
| P4-P6 domain(*Tetrahymena* ribozyme) (24) | 160 | 2 | Group I intron *T.* ribozyme |
| *Azoarcus* ribozyme (23,25) | 205 | 7 ~ 14 | |
| *B. subtilis* RNase P RNA catalytic domain (26) | 225 | 6.5 ± 0.2 | |
| Ca.L-11 ribozyme (27) | 368 | 0.03 | |
| *E. coli* RNase P RNA (28) | 377 | 0.011 ± 0.001 | |
| *B. subtilis* RNase P RNA (28) | 409 | 0.008 ± 0.002 | |
| *Tetrahymena* ribozyme (23,29) | 414 | 0.013 | Group I intron *T.* ribozyme |

interest, general arguments based on the kinetics of loop formation have been used to predict that the speed limit for protein folding is also ~1 ms (12–14). It remains to be ascertained whether the common folding speed limit for proteins and RNA is due to evolutionary pressure on the folding of evolved sequences. It is worth pointing out that Dill et al. (15) recently showed that the rates and stabilities of protein folding depend only on the number of amino acids, which in turn places strict constraints on their functions in the cellular context. Taken together, these studies show that despite the complexity of protein and RNA folding, it is possible that only a few variables determine their global properties. This suggests that certain simple principles may determine biological functions.

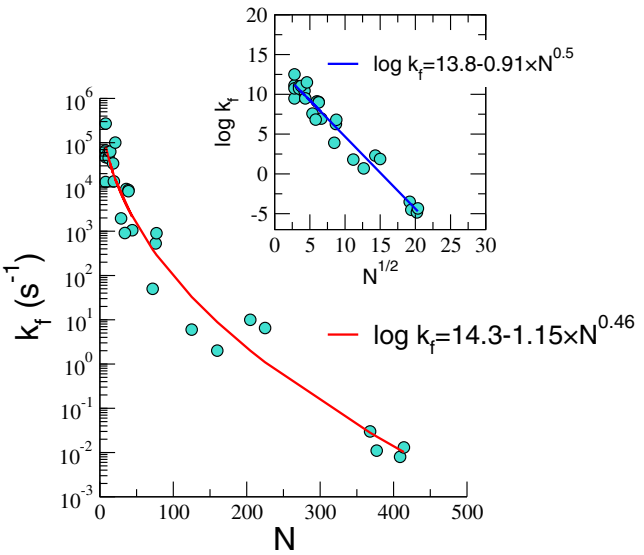


**FIGURE 1 Dependence of the folding rates of RNA on *N*. The circles are experimental data and the line is the fit obtained using $\log k_F = \log k_0 - \alpha N^{\beta}$, with $\beta$ used as an adjustable parameter. Inset shows the fit obtained by fixing $\beta$ to the predicted theoretical value of 0.5.**

## ACKNOWLEDGMENTS

This work was supported by grants from the Korea Research Foundation (2010-0000602) (C.H.) and the National Science Foundation (CHE 09-14033) (D.T.).

## REFERENCES and FOOTNOTES

1. Thirumalai, D., and C. Hyeon. 2005. RNA and protein folding: common themes and variations. *Biochemistry.* 44:4957–4970.
2. Treiber, D. K., and J. R. Williamson. 1999. Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* 9:339–345.
3. Woodson, S. A. 2010. Compact intermediates in RNA folding. *Annu Rev Biophys.* 39:61–77.
4. Hyeon, C., R. I. Dima, and D. Thirumalai. 2006. Size, shape, and flexibility of RNA structures. *J. Chem. Phys.* 125:194905.
5. Yoffe, A. M., P. Prinsen, …, A. Ben-Shaul. 2008. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. USA.* 105:16153–16158.

6. Hajdin, C. E., F. Ding, …, K. M. Weeks. 2010. On the significance of an RNA tertiary structure prediction. *RNA.* 16:1340–1349.
7. Thirumalai, D. 1995. From minimal models to real proteins: time scales for protein folding kinetics. *J. Phys.* 5:1457–1467.
8. Guo, Z., and D. Thirumalai. 1995. Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers.* 36:83–102.
9. Pan, J., D. Thirumalai, and S. A. Woodson. 1997. Folding of RNA involves parallel pathways. *J. Mol. Biol.* 273:7–13.
10. Thirumalai, D., N. Lee, …, D. Klimov. 2001. Early events in RNA folding. *Annu. Rev. Phys. Chem.* 52:751–762.
11. Pörschke, D., O. Uhlenbeck, and F. Martin. 1973. Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. *Biopolymers.* 12:1313–1335.
12. Hagen, S. J., J. Hofrichter, …, W. A. Eaton. 1996. Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA.* 93:11615–11617.
13. Yang, W. Y., and M. Gruebele. 2003. Folding at the speed limit. *Nature.* 423:193–197.
14. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* 14:76–88.
15. Dill, K. A., K. Ghosh, and J. D. Schmit. 2011. Physical limits of cells and proteomes. *Proc. Natl. Acad. Sci. USA.* 108:17876–17882.
16. Proctor, D. J., H. Ma, …, P. C. Bevilacqua. 2004. Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry.* 43:14004–14014.
17. Porschke, D. 1974. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem.* 1:381–386.
18. Riesner, D., and R. Römer. 1973. Thermodynamics and kinetics of conformational transitions in oligonucleotides and tRNA. *In* Physico-Chemical Properties of Nucleic Acids, *Vol. 2*. J. Duchesne, editor. Academic Press, London, pp. 237–318.
19. Serebrov, V., R. J. Clarke, …, L. Kisselev. 2001. $Mg^{2+}$-induced tRNA folding. *Biochemistry.* 40:6688–6698.
20. Kuznetsov, S. V., C. C. Ren, …, A. Ansari. 2008. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* 36:1098–1112.
21. Narayanan, R., Y. Velmurugu, …, A. Ansari. 2011. Fast folding of RNA pseudoknots initiated by laser temperature-jump. *J. Am. Chem. Soc.* 133:18767–18774.
22. Tan, E., T. J. Wilson, …, T. Ha. 2003. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. USA.* 100:9308–9313.
23. Sosnick, T. R., and T. Pan. 2004. Reduced contact order and RNA folding rates. *J. Mol. Biol.* 342:1359–1365.
24. Deras, M. L., M. Brenowitz, …, S. A. Woodson. 2000. Folding mechanism of the *Tetrahymena* ribozyme P4-P6 domain. *Biochemistry.* 39:10975–10985.
25. Rangan, P., B. Masquida, …, S. A. Woodson. 2003. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. USA.* 100:1574–1579.
26. Fang, X., T. Pan, and T. R. Sosnick. 1999. Mg-dependent folding of a large ribozyme without kinetic traps. *Nat. Struct. Biol.* 12:1091–1095.
27. Xiao, M., M. J. Leibowitz, and Y. Zhang. 2003. Concerted folding of a *Candida* ribozyme into the catalytically active structure posterior to a rapid RNA compaction. *Nucleic Acids Res.* 31:3901–3908.
28. Zarrinkar, P. P., J. Wang, and J. R. Williamson. 1996. Slow folding kinetics of RNase P RNA. *RNA.* 2:564–573.
29. Sclavi, B., M. Sullivan, …, S. A. Woodson. 1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science.* 279:1940–1943.