PHYSICAL REVIEW E **70**, 040901(R) (2004)

Unfolding single RNA molecules by mechanical force: A stochastic kinetic method

Fei Liu[1,*] and Zhong-can Ou-Yang[1,2]
[1]*Center for Advanced Study, Tsinghua University, Beijing 100084, People's Republic of China*
[2]*Institute of Theoretical Physics, The Chinese Academy of Sciences, P. O. Box 2735, Beijing 100080, People's Republic of China*
(Received 25 April 2004; revised manuscript received 6 July 2004; published 14 October 2004)

Using simple polymer elastic theory and known RNA free energies, we study the single RNA folding and unfolding on the secondary structure level under mechanical constant force by stochastic kinetic simulation. As a primary application, this method is used to simulate the experiment performed by Liphardt *et al.* [Science **292**, 733 (2001)]. The extension-force curves in equilibrium and kinetic reaction rate constants for folding and unfolding are calculated. Our results show that the agreement between simulation and experimental measurements is satisfactory.

DOI: 10.1103/PhysRevE.70.040901 PACS number(s): 87.15.Aa, 87.15.He, 87.15.La

Folding processes and folding structures of protein and RNA are the result of physical forces generated in molecules, such as by hydrophobic or hydrophilic interaction, self-avoidance between molecular components, and electrostatic interactions, etc. Therefore, direct force measurement or exerting force to molecules is an intriguing way to study the molecular folding problem. The advent of single-molecule manipulations offers the opportunity to directly measure and manipulate protein or RNA molecules on a single-molecule level [1–8]. Compared to the general thermodynamics of protein or RNA under forces in equilibrium, single-molecule methods could be more interesting in the kinetic studies of biomolecules [9]. Unlike bulk folding research, the single-molecule experiments allow us to follow the actual folding or unfolding trajectories of single molecules with high resolution [5,6], which will shed light on the difficult kinetic folding problem.

Recently, enormous theoretical efforts have been devoted to understanding the folding and unfolding phenomena of proteins and RNA observed in single-molecule experiments. Diverse methods, including molecular dynamics [10,11], the Monte Carlo method [12,13], and other theoretical models [14,15], have been developed. However, we note that these studies mainly focused on the dynamical behavior of proteins under force, and few were concerned about RNA [8]. To fill this gap, in this work, we study the RNA kinetics under *constant force*. The theoretical model is sketched in Fig. 1. The essence of our method is a stochastic RNA folding algorithm. The algorithm is based on individual base pairs and was developed by Flamm *et al.* [16]. We show that the secondary structural RNA folding and unfolding kinetics under mechanical force can be traced by using a simple polymer elastic theory and known RNA free energies obtained from bulk experiments.

A secondary structure S of a RNA sequence $l=(x_1,x_2,\ldots,x_n), x_i \in \{A,U,C,G\}$ is a list of base pairs $[x_i,x_j]$, $i<j$; for any two base pairs $[x_i,x_j]$ and $[x_k,x_l]$ with $i<k$ holds $i<k<l<j$. All structures of the sequence l comprise a set $S(l)=\{S_0,S_1,\ldots,S_m,0\}$; here 0 denotes the completely open chain conformation. In order to construct a kinetic algorithm, a relation M must be defined to specify whether two structures are accessible from each other by an elementary "move." The move set we apply here consists of the removal and insertion of single base pairs; it is also the simplest move set on the level of secondary structure [16]. The structural set and the relation comprise a RNA secondary structural conformational space $C(l)=\{S(l),M\}$. Any trajectory of RNA folding and unfolding can be described by a succession of elementary steps in this space. We model the RNA folding and unfolding as a Markov process which satisfies the master equation,

$$\frac{dP_i(t)}{dt}=\sum_{j=0}[P_j(t)k_{ji}-P_i(t)k_{ij}], \tag{1}$$

where $P_i(t)$ is the probability of the RNA conformation being S_i at time t. k_{ij} is the transition rate from conformation S_i to

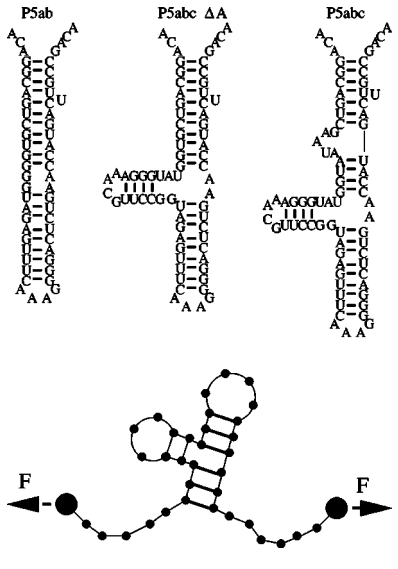

FIG. 1. Theoretical model and RNA sequences and their native structures studied in the present work. The structures are folded by the VIENNA RNA package 1.4. The equilibrium and kinetic behaviors of these three RNAs, p5ab, p5abcΔ*A*, and p5abc, have been studied in detail [6].

*Email address: liufei@tsinghua.edu.cn

conformation S_j; the nonzero rates are required to satisfy the detailed balance condition,

$$\frac{k_{ij}}{k_{ji}} = \exp[-\Delta E_{ij}(f)/k_B T], \quad (2)$$

where $\Delta E_{ij}(f)$ is the energy difference between two conformations under constant force f. On the coarse-grain level, the energy difference can be written as

$$\Delta E_{ij}(f) \approx \Delta G_{ij}(f) = \Delta G_j^0 - \Delta G_i^0 - (m_j - m_i) f d_s(f), \quad (3)$$

where ΔG_i^0 and ΔG_j^0 are the energies obtained from folding the sequence into the conformations S_i and S_j, respectively, while m_i and m_j are the numbers of exterior unpaired bases of corresponding conformations. The extension $d_s(f)$ per base is $d_s(f) = b_{ss}[\cot(u) - 1/u]$, here $u = lf/k_B T$, b_{ss} is the base distance, and l is the Kuhn length [17].

To improve the efficiency of the simulation, we assume a symmetric rule for the transition rates k_{ij} [18]

$$k_{ij} = \tau_o^{-1} \exp[-\Delta G_{ij}(f)/2k_B T], \quad (4)$$

where τ_o is used to scale the time axis of the unfolding process; it will be determined by experimental data. Although the form of the master equation looks relatively simple, it is mathematically intractable to solve analytically even for a simpler system such as the p5ab molecule below. We resort to a continuous time Monte Carlo simulation [16,19,20] to study the stochastic process described by Eq. (1).

To test the accuracy of our simulation, we first study the extension-force curves of three small RNA, p5ab, p5abΔA, and p5abc in equilibrium. These molecules represent major structural units of large RNA assemblies [6]: p5abc is a simple RNA hairpin; p5abcΔA has an additional helix and thus is a three-helix junction, and p5abc is comparatively complicated and contains an *A*-rich bulge; see Fig. 1. We perform the simulation at the experimental temperature $T = 298$ K, and use the single-strand DNA parameters for RNA because they have a similar chemical structure: $b = 0.56$ nm, $l = 1.5$ nm [17]. The free energy parameters for RNA secondary structures are selected from the VIENNA package 1.4 [21]. The salt concentrations are $[Na^+] = 1M$ and $[Mg^{2+}] = 0M$. Figure 2 shows the extension-force curves for the sequences. When these molecules are stretched by small forces, their extensions increase monotonically. However, when force increases to around 13.3 pN, the extension of p5ab is interrupted by an ~20 nm jump. Similar to p5ab, the extension of p5abcΔA also has a sharp jump at force ~11.3 pN. These curves show that the transitions of mechanical unfolding for p5ab and p5abcΔA are all or none. Because the native states of these two molecules are almost the same except an additional helix in p5abcΔA, we conclude that the presence of an the helix destabilizes the molecular structure. Compared to the simple all-or-none behaviors of p5ab and p5abcΔA, the extension-force curve of p5abc provides more features: the extension has a large jump (~17 nm) around force 8.0 pN, and an inflection is followed and the extension increases gradually to full length. The smaller unfolding force of p5abc shows that the presence of *A*-rich bulge makes the molecule more unstable. Because the jump is about two-thirds of the full extension, the transition represents an unfolding of the p5a helix and parts of the three-helix junction. All results obtained above are consistent with experiment [6].

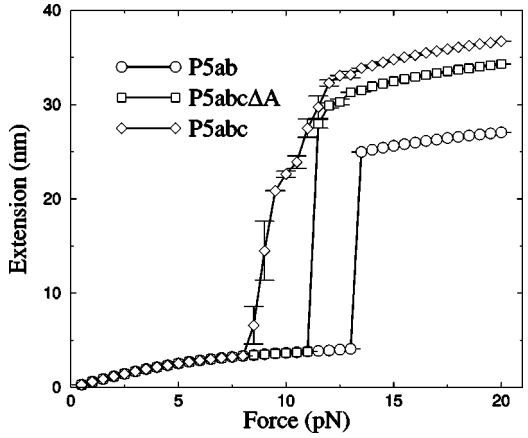

FIG. 2. The extension-force curves for the three RNA molecules. Our simulations show that the folding and unfoldings for p5ab and p5abcΔA under constant force are all or none around the transitions 13.3 pN and 11.3 pN, respectively, while the extension-force curve for p5abc has an intermediate phase between 8.0 and 11.0 pN.

The main advantage of our simulation is in studying RNA kinetics under force. In unfolding experiments, when the force was held constant at the transition within ~1 pN, p5ab and p5abΔA switched back and forth with time from folded hairpin (hp) to unfolded single strand (ss). A two-state kinetic description has been proposed to explain the intriguing phenomena. The researchers found that the rate constants for unfolding can be fit to an Arrhenius-like expression, $k_u(f) \propto \exp(f\Delta x_u^{\ddagger}/k_B T)$ very well, where $\Delta x_u^{\ddagger}$ is the thermally averaged distance between the hairpin state and the transition state along the direction of force. A similar expression holds for the folding rate $k_f(f)$ [6]. Apparently, this description cannot clarify the physics underlying the folding and unfolding [22].

In contrast to the two-state model, our simulation is completely based on microscopic interactions. We simulate the experimental process. The extension-time traces of RNA molecules at various forces in equilibrium are recorded. For example, three time traces of p5ab are shown in Fig. 3(a). We find that the extension of the molecule jumps between two values, ~5 nm and ~22 nm around the transition force. Because the jumps are extremely rapid with respect to the lifetimes of the two extension values, we simply classify the states whose extensions are larger than 15 nm as single stranded states, and the others as hairpin states. Around the transition the frequencies of different lifetimes at the single stranded state and the hairpin state are obtained by a long time simulation (the time for each running is $10^9 \tau_0$). Figure 3(b) gives the frequency distributions of a typical simulation at a force of 12.90 pN. These distributions can be fit to exponential functions $\propto \exp(-t/\langle \tau_i \rangle)$ very well, where $\langle \tau_i \rangle, i = u, f$ denotes force-dependent average lifetimes of the single-stranded and hairpin states, respectively. For example, the average lifetimes of the two states in this simulation are

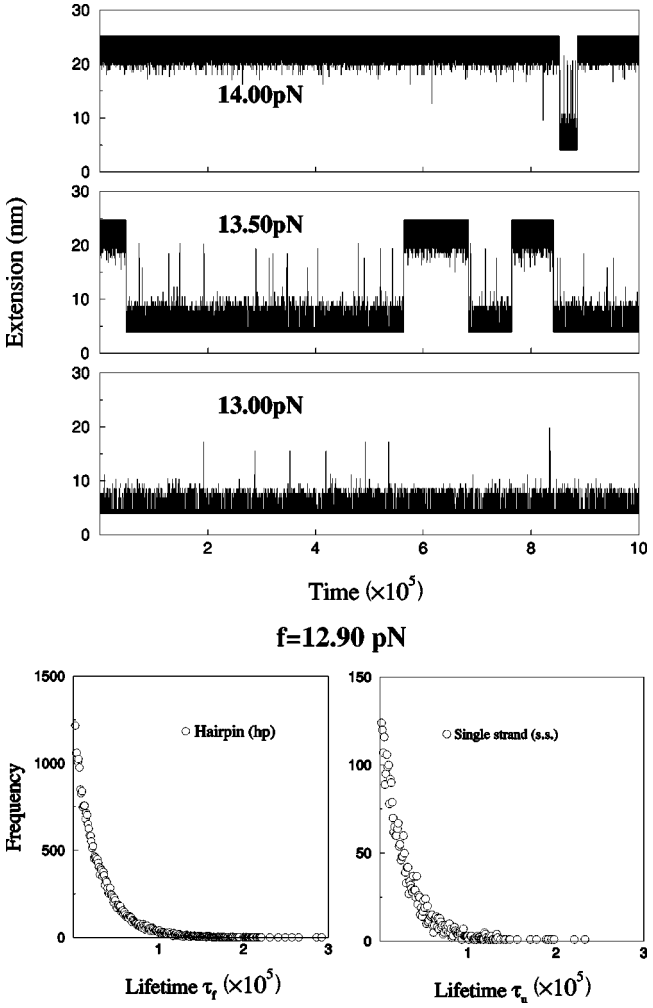

FIG. 3. Simulation for RNA p5ab kinetics. (a) The extension versus time traces of the molecule at different forces in equilibrium; here the unit of time is τ_o. (b) The frequency distributions of the lifetimes of the single stranded and hairpin states at the transition force 12.90 pN. The average lifetimes of these two states in this simulation can be obtained by fitting to exponential functions.

$\langle \tau_u \rangle \approx 3.4 \times 10^4 \tau_o$ and $\langle \tau_f \rangle \approx 3.1 \times 10^4 \tau_o$. We calculate all average lifetimes near the transition force of p5ab; their corresponding values with various forces are shown in Fig. 4. We find that the logarithms of the lifetimes for the two states are strikingly consistent with linear functions of force. Because the rate constants are inverse to the average lifetimes, we fit τ_o by making $\langle \tau_u \rangle (f^*) = \langle \tau_f \rangle (f^*)$ equal to experimental value $1/k^*$, where $k^* \equiv k_u = k_f$, and have $\tau_o^{-1} = 2.2 \times 10^5$ sec^{-1}. Using the same method, we also calculate the reaction rate constants of the p5abcΔA molecule. These results and experimental data are listed in Table I [23]. Because our simulation does not introduce additional fitting parameters, the striking consistency of our results with the experiment assures us that the RNA model under force proposed here has grasped the main physics.

It is noteworthy to compare our method with the recent work given by Cocco *et al.* [24]. They used a simple dynamical model to explain the slow folding and unfolding kinetics. There are several differences with the method proposed by us. First is the force energy formula. They suggested that the energy contributed by constant force should be $g_s(f) = k_B T b_s s / l \ln \sin(u)/u$. We have applied the same formula at the beginning of our work. Our simulation, however, showed that this energy could not obtain the results within the experimental errors, e.g., the unfolding forces larger than measurements about 3 pN. On the other hand, in serious simulations which include conformational transformations, the force energy is $\mathbf{f} \cdot \mathbf{r}$, where $\mathbf{f}$ is constant force vector and $\mathbf{r}$ is the conformational end-to-end vector at each simulation step [25]. To avoid complete thermodynamic or complex microscopic description, we propose Eq. (4). In addition to its accuracy in predicting experiments, the physical meaning of our formula is also apparent: the maximum probable conformation is important for the force energy. The second is the choice of the move set. Cocco *et al.* restricted formation or removal of the base pair just proceeding in the "fork" location. This restriction should be reasonable for the pairing of two distant bases is inhibited by a kinetic barrier. However, we have to point out that their method must preset the RNA native structure, while it is available only for simpler RNA

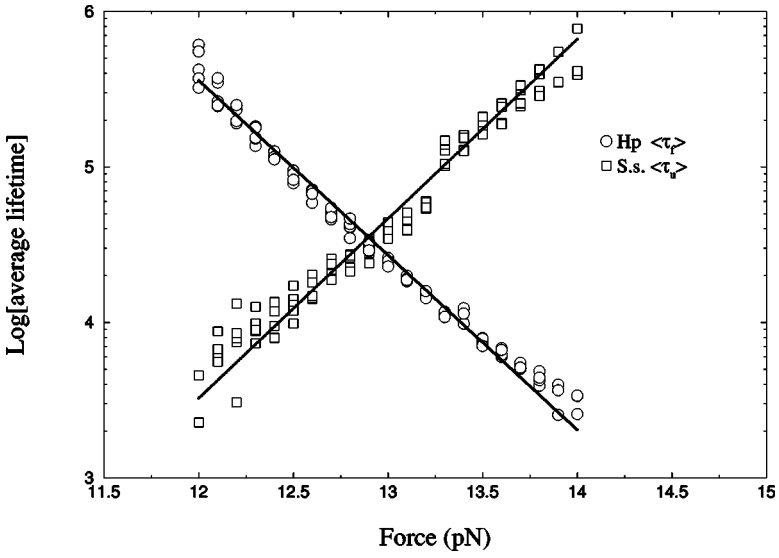

FIG. 4. Logarithm of the average lifetimes of single stranded and hairpin states for the p5ab molecule at difference forces around the transition 12.9 pN. The time is in unit τ_o, which can be obtained by fitting to experimental data. Note the slopes of $\ln\langle \tau_f \rangle$ and $\ln\langle \tau_u \rangle$ are independent of the fitting value τ_o.

TABLE I. Simulation results for the p5ab and p5abcΔA molecules compared to experimental values from Ref. [6] (in bold).

Molecule	$\langle \Delta x \rangle$(nm)	f^*(pN)	$\ln k_f(f)(s^{-1})$	$\ln k_u(f)(s^{-1})$
p5ab, Mg^{+2}	19±2	14.5±1	$41\pm1.9-(2.8\pm0.1)f$	$-39\pm2.3+(2.9\pm0.2)f$
p5ab, by Marko		15.1	$27.5-2.74f$	$-42.9+1.93f$
p5ab, EDTA	18±2	13.3±1	$37\pm4.0-(2.7\pm0.3)f$	$-32\pm4.8+(2.6\pm0.4)f$
p5ab, by us	20.0	12.9±0.5	$36.1\pm1.4-(2.7\pm0.1)f$	$-31.5\pm0.7+(2.6\pm0.1)f$
p5abcΔA, Mg^{+2}	22±4	12.7±0.3	$58\pm7.5-(4.2\pm0.5)f$	$-39\pm9.3+(2.7\pm0.7)f$
p5abcΔA, by Marko		12.9	$9.4-2.05f$	$-43.8+2.06f$
p5abcΔA, EDTA	23±2	11.4±0.5	$31\pm6.0-(2.6\pm0.5)f$	$-31\pm11+(2.5\pm0.3)f$
p5abcΔA, by us	25.0	11.3±0.6	$36.2\pm1.8-(3.4\pm0.2)f$	$-27.8\pm0.8+(2.3\pm0.1)f$

hairpins or double-stranded DNA. Therefore, their model is not suitable to study general RNA unfolding by force. In contrast, because our model is based on a general RNA folding algorithm, the kinetic inhibition effects are naturally taken into account by the entropy term in RNA free energy [26,27]. Finally, we require the rate constants to satisfy the detailed balance; the symmetric rule used here is only for the improvements of simulation efficiency rather than an ad hoc physical assumption having to be made in Ref. [24].

Several improvements can be added in the algorithm. Firstly, because mechanical unfolding experiments are typically performed by stretching the ends of the molecule at a constant pulling speed, inclusion of the loading rate is essential to correctly understand the RNA unfolding kinetics under forces, e.g., the influence of the apparatus on the actual kinetics of molecules [22]. Secondly, a complete model should include the effects of Mg^{+2} [6]. The tertiary structures induced by the ions cannot be described in our simulation now. In addition, how to extend our algorithm to include pseudoknots [28] should be considered in future studies [8].

We thank Dr. Flamm for providing us his computer program KINFOLD. It is also a pleasure to acknowledge Professor H.-W. Peng for many helpful discussions on this work.

[1] M. Z. Kellermeyer *et al.*, Science **276**, 1112 (1997).
[2] L. Tskhovrebova *et al.*, Nature (London) **387**, 308 (1997).
[3] M. Rief *et al.*, Science **276**, 1109 (1997).
[4] A. F. Oberhauser *et al.*, Nature (London) **393**, 181 (1998).
[5] J. M. Fernandez and H.-B. Li, Science **303**, 1674 (2004).
[6] J. Liphardt *et al.*, Science **292**, 733 (2001).
[7] Onoa *et al.*, Science **299**, 1892 (2003).
[8] S. Harlepp *et al.*, Eur. Phys. J. E **12**, 605 (2003).
[9] I. Tinoco, Jr. and C. Bustamante, Biophys. Chem. **101-102**, 513 (2003).
[10] H. Lu *et al.*, Biophys. J. **75**, 662 (1998).
[11] Z. Bryant, V. S. Pande, and D. S. Rookhsar, Biophys. J. **78**, 584 (2000).
[12] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **96**, 6166 (1999).
[13] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **96**, 2031 (1999).
[14] U. Gerland, R. Bundshuh, and T. Hwa, Biophys. J. **81**, 1324 (2001).
[15] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).
[16] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster, RNA **6**, 325 (2000).
[17] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[18] K. Kawasaki, Phys. Rev. **145**, 224 (1966).
[19] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[20] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[21] I. L. Hofacker, Nucleic Acids Res. **31**, 3429 (2003).
[22] J. M. Fernandez, S. Chu, and A. F. Oberhauser, Science **292**, 653 (2001).
[23] One of the main goals of the experiments carried out by Liphardt *et al.* is to investigate the role of Mg^{2+} in RNA folding and unfolding under mechanical forces. They found that the presence of the ions increases the average force of folding and unfolding when the RNA molecules only form secondary structures, e.g., removal of the ions lowers the average force from 14.5 to 13.3 pN, which has been recovered by our simulation. Cocco *et al.* have studied the rate constants of folding and unfolding reactions in the presence of Mg^{+2}. Unfortunately, they did not given any results in the absence of Mg^{+2}. As a comparsion, we list their results in Table I.
[24] S. Cocco, J. F. Marko, and R. Monasson, Eur. Phys. J. E **10**, 153 (2003).
[25] Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **81**, 1133 (2001).
[26] J. S. McCaskill, Biopolymers **29**, 1105 (1990).
[27] Considering that quantitative comparison of the two force energy formules and the two dynamic models needs more space, we do not show these results in the present work.
[28] E. Rivas and S. R. Eddy, J. Math. Biol. **285**, 2053 (1999).