# Modelling Force-Extension Relation for RNA Folding

Ying-Yu Ho*
*Department of Physics, Massachusetts Institute of Technology*
(Dated: May 12, 2011)

We present a mechanical description of RNA folding based on the *worm-like chain* model of RNA elasticity. The force-extension and the energy-extension relations are studied and applied to experimental data in order to check the accuracy of the theoretical model we obtain. The agreement is good in general. Some discrepancy possibly due to the use of equilibrium statistical mechanics and the ignorance of RNA structure are also discussed.

## 1. UNFOLDING RNA BY FORCE

The structure of a biomolecule is closely related to its function in an organism. For example, the stable double-helix structure of DNA is suitable for the information in DNA to be stored and duplicated. In contrast to DNA, its analogue RNA is usually single-stranded, allowing it to appear in a richer variety of structures. This flexibility reflects different roles of RNA in a biological system such as carrying genetic information (mRNA), delivering amino acids (tRNA), and linking amino acids together to form proteins (rRNA).

Recent investigations have further expanded our knowledge of capabilities of RNA in addition to those mentioned above. For instance, *riboswitches* were recently found to be parts of mRNA that regulate gene activity through binding small target molecules (*ligand binding*) [1]. Accordingly, many creative experiments have contributed to the structure determination of riboswitches. One experiment carried out by Greenleaf *et al.* employed optical tweezers to exert forces on riboswitches and thus unfold their secondary structures [2]. Their experiment was concerned about *pbuE* adenine riboswitch aptamers (oligonucleic acid with ligand-binding ability), which has a hierarchical folding consisting of two hairpins, as shown in Fig. 1 and Fig. 2. They analyzed the force-extension (FE) relation of aptamer, from which they identified the folding states and further constructed energy landscapes for aptamer folding.

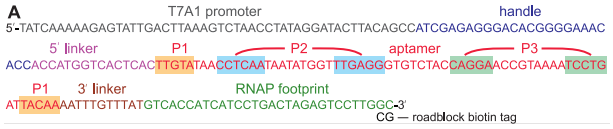


FIG. 1: The corresponding DNA sequence of the *pbuE* riboswitch aptamer. P2 and P3 represent the two hairpin structure (adopted from [2]).

The great work of Greenleaf *et al.* could not have been successful without a good theoretic model for of RNA folding, especially a model that explains the FE relation.

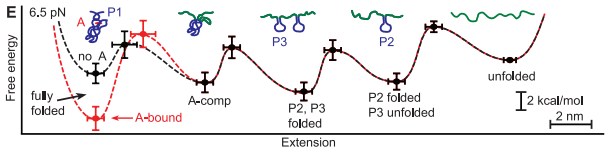


FIG. 2: Energy landscapes of aptamer at $F = 6.5$ pN. The region from "P2, P3 folded" to "unfolded" will be studied in this paper (adopted from [2]).

Although some details such as binding energy are best determined by experiments, the spirit of a phenomenological model can be well-understood in the frame of statistical mechanics, which will be discussed below.

## 2. RNA ELASTICITY

Since an RNA molecule always has some unfolded regions, we cannot understand the FE relation for a *pbuE* aptamer without knowing that for an unfolded RNA. A particularly useful model for RNA elasticity is the *worm-like chain* (WLC) model, which regards an RNA molecule as a chain that can bend but not stretch [3]. Since an RNA molecule is negatively charged, bending it costs energy, characterized by the coefficient of *bending rigidity* $\kappa$. On the other hand, a bent molecule is more disordered, which is preferred in high temperature. The balance of bending rigidity and thermal motions therefore introduces the *persistence length* $\xi_P = \kappa/k_BT$, which serves as the fundamental length scale of the molecule [4].

The WLC model is not analytically solvable in general. Nevertheless, an approximate formula that take account of the asymptotic behaviors at very large and very small forces can provide much insight into the WLC model. A simplified derivation is provided in this section, with some basic equations for the WLC model omitted. An interested reader might consult the references provided above. The application of WLC model on nucleic acids is also discussed in detail by [5].

First we look at the small-force behavior. In the absence of external force, the correlation between tangent vectors along the RNA chain is

$$\langle \vec{t}(s) \cdot \vec{t}(u) \rangle = e^{-|s-u|/\xi_P}, \tag{1}$$

*Electronic address: yingyu@mit.edu

giving the mean squared end-to-end distance

$$\begin{aligned}\langle R^2\rangle &= \int ds \int du\ \langle \vec{t}(s)\cdot\vec{t}(u)\rangle \\ &\approx L\int_{-\infty}^{\infty} du\ e^{-|s-u|/\xi_P} = 2L\xi_P, \end{aligned} \tag{2}$$

where $L$ is the contour (total) length of RNA. Applying the *central limit theorem*, the probability distribution of the extension at small force $F$ is encoded in the *Gibbs partition function*

$$\begin{aligned}\mathcal{G}(F) &= \int dz \left[-\frac{z^2}{2\langle z^2\rangle} + \beta F z\right] \\ &= \sqrt{2\pi\langle z^2\rangle}\ \exp\left[\frac{1}{2}\beta^2F^2\langle z^2\rangle\right]. \end{aligned} \tag{3}$$

Differentiating Eq. (3) with respect of $F$ then gives us a linear F-E relation

$$\begin{aligned} l &= \langle z\rangle = k_BT\frac{\partial}{\partial F}\log\mathcal{G} \\ \Rightarrow\ F &= \frac{3k_BT}{2L\xi_P}\cdot l. \end{aligned} \tag{4}$$

To study the large-force behavior, we start from the Hamiltonian

$$\mathcal{H} = \int_0^L ds\left[\frac{1}{2}\kappa\left(\frac{d\vec{t}}{ds}\right)^2 - \vec{F}\cdot\vec{t}\right]. \tag{5}$$

The first term in the integrand is the elastic energy for small bending, while the second term takes care of the potential due to the external force. Remember that $\kappa = k_BT\xi_P$ is the coefficient of bending rigidity. At large force, the tangent vectors of the chain are almost in parallel with the force, so it is convenient to make a decomposition $\vec{t} = \vec{t}_\parallel + \vec{t}_\perp$. Since $|\vec{t}| = 1$ and $\vec{t}_\perp$ is small, $|\vec{t}_\parallel| \approx 1 - \vec{t}_\perp^2/2$. Discarding a constant, Eq. (5) can be approximated as

$$\mathcal{H} \approx \int_0^L ds\left[\frac{1}{2}\kappa\left(\frac{d\vec{t}_\perp}{ds}\right)^2 + \frac{1}{2}F\vec{t}_\perp^2\right]. \tag{6}$$

Following the traditional way of studying small vibration, we expand $\vec{t}$ in terms of normal modes

$$\vec{t}_\perp = \sum_{n=1}^{\infty}(a_n\hat{x} + b_n\hat{y})\cos\left(\frac{n\pi s}{L}\right). \tag{7}$$

Here Dirichlet boundary conditions are assumed, but other choices will not change the asymptotic behavior of $a_n$ and $b_n$. With this expansion, the Hamiltonian becomes a clean summation

$$\mathcal{H} = \sum_{n=1}^{\infty}\frac{L}{2}\left(a_n^2 + b_n^2\right)\left(\frac{1}{2}\frac{n^2\pi^2\kappa}{L^2} + \frac{1}{2}F\right). \tag{8}$$

Since $a_n$ and $b_n$ denote the amplitudes of independent harmonic oscillators, the *equipartition theorem* gives

$$a_n^2 = b_n^2 = k_BT\cdot\frac{1}{L}\left(\frac{1}{2}\frac{n^2\pi^2\kappa}{L^2} + \frac{1}{2}F\right)^{-1}. \tag{9}$$

Considering the second term in the integrand of Eq. (6), the difference between the contour length $L$ and the chain extension $l$ is

$$\begin{aligned} L - l &\approx k_BT\int_0^{\infty} du\left(\frac{\pi^2\kappa}{L^2}u^2 + F\right)^{-1} \\ &= \frac{k_BTL}{2\sqrt{\kappa F}}, \end{aligned} \tag{10}$$

giving the FE relation at large $F$

$$\frac{F\xi_p}{k_BT} = \frac{1}{4}\left(1 - \frac{l}{L}\right)^{-2}. \tag{11}$$

The results of Eq. (4) and Eq. (11) can be summarized in a single formula

$$\frac{F\xi_p}{k_BT} = \frac{l}{L} + \frac{1}{4}\left[\left(1 - \frac{l}{L}\right)^{-2} - 1\right]. \tag{12}$$

To check the small $l$ behavior of Eq. (13), we expand it at $l = 0$ to obtain

$$\frac{F\xi_p}{k_BT} \approx \frac{l}{L} + \frac{1}{4}\left[1 + \frac{2l}{L} - 1\right] = \frac{3l}{2L}, \tag{13}$$

which indeed agrees with the expected small $l$ behavior. Eq. (13) can be integrated to obtain the elastic potential

$$\frac{V\xi_p}{k_BTL} = \frac{1}{2}\left(\frac{l}{L}\right)^2 + \frac{1}{4}\left(1 - \frac{l}{L}\right)^{-1} - \frac{1}{4}\left(1 + \frac{l}{L}\right). \tag{14}$$

This potential will be used in the next section to model the FE relation of RNA with several folding states.

## 3. FORCE-EXTENSION RELATION

We should supply our theoretical model with some experimental data to figure out its performance. Without access to raw data, we cannot use regression analysis to find all the needed parameters. Nevertheless, we can based our calculation on some measured quantities, and then compare other data with theoretical predictions to check their consistency. Without special mentioning, all parameters in this section are taken from the work of Greenleaf *et al.* [2].

Eq. (14) is derived for an unfolded RNA. However, for hairpin structures shown as P2 and P3 in Fig. 1 and Fig. 2, we can regard the sequence outside the hairpin regions as unfolded and apply Eq. (14). Since different folding states can convert into each other, we need to consider these states simultaneously. For simplicity we

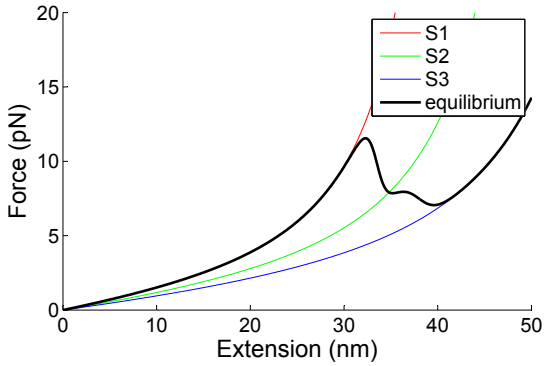


FIG. 3: Black: FE curve computed from Eq. (16). Colored: FE curve for individual stated S1 - S3.

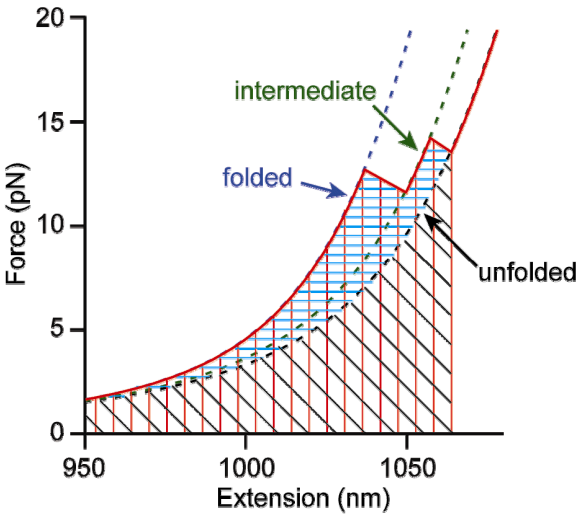


FIG. 4: Measured FE curve showing approximate state transitions (adopted from the supporting material of [2]).

study only the three states shown in the region from "P2, P3 folded" to "unfolded" (call them S1 - S3 from left to right) in Fig. 2.

We consider each state Si a chain with a contour length $L_i$ and a common persistence length $\xi_p$. Eq. (14) then becomes

$$\frac{V_i}{k_BT} = \frac{L_i}{\xi_p}\left[\frac{1}{2}x_i^2 + \frac{1}{4}(1-x_i)^{-1} - \frac{1}{4}(1+x_i)\right], \quad (15)$$

where $x_i = l/L_i$, $L_i = N_i a$ is the contour length of the $i$-th state, and $N_i$ is the length in terms of nucleotides. By counting the sequence in Fig. 1, we find $N_1 \approx 79, N_2 \approx 98$ and $N_3 \approx 119$. We use $T = 297$ K, $a = 0.59$ nm and $\xi_p = 1$ nm. The transition from S1 to S2 costs $\epsilon_{12} = 11.7\ k_BT$, and that from S2 to S3 costs $\epsilon_{23} = 8.5\ k_BT$. These are given by hairpin models and assumed to be independent of RNA extension or applied force. Now we can write down the all-important partition function

$$\begin{aligned}\mathcal{Z}(l) &= \theta(L_1 - l)\exp\left[-\frac{V_1(l)}{k_BT}\right] \\ &+ \theta(L_2 - l)\exp\left[-\frac{\epsilon_{12} + V_2(l)}{k_BT}\right] \\ &+ \theta(L_3 - l)\exp\left[-\frac{\epsilon_{12} + \epsilon_{23} + V_3(l)}{k_BT}\right], \quad (16)\end{aligned}$$

where step functions are added since $V_i(l) \to \infty$ as $l \to L_i$. The force-extension relation can be obtained via

$$\langle F \rangle = -k_BT\frac{\partial}{\partial l}\log \mathcal{Z}, \quad (17)$$

which is shown in Fig. 3.

Comparing with the measured FE curve Fig. 4, we see that the critical force for the S1 $\to$ S2 transition is correctly predicted, while the predicted force for the S2 $\to$ S3 is too small. Note that the "extension" shown in Fig. 4 includes the length of a handle DNA used to link the RNA for measurement. Therefore the change in extension is larger due to DNA elasticity. Since the measurement of FE curve is a non-equilibrium process, we can expect that the total work spent to pull the RNA is typically larger than the difference in free energy between the initial and the final state, which might cause the discrepancy in FE curves.

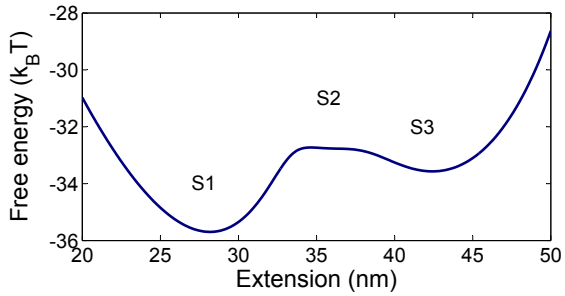


FIG. 5: Free energy versus extension computed from Eq. (18). The zero point of free energy is unimportant here.

Because Greenleaf *et al.* also measured the energy landscapes for aptamer folding, which are equilibrium quantities, it would be helpful to compare them with our theoretical predictions. Including the effect of external force the free energy in our model is

$$A(l) = -k_BT\log \mathcal{Z}(l) - Fl, \quad (18)$$

with the partition function $\mathcal{Z}$ given by Eq. (16). With $F = 8$ pN corresponding to the state transition in Fig. 3, we plot the free energy $A(l)$ in Fig. 5. Looking at the plot, we find that S2 is never a stable (and barely metastable) state, which is not true according to the landscapes shown in Fig. 2. This suggests that our model for aptamer structure assuming only unfolded and hairpin regions could be inaccurate.

In fact, Greenleaf *et al.* also found the unfolding force for S2 → S3 predicted by hairpin models smaller than their experimental result. As pointed out by them, the hairpin P2 in state S2 could be stabilized by intraloop interactions, which is not consider by the hairpin model.

## 4. CONCLUSION

Our phenomenological model for the FE relation of *pbuE* riboswitch aptamers based on the WLC model of RNA elasticity and the hairpin model is consistent in describing the transition S1 → S2 but not in describing S2 → S3. It may seems that our model is not so perfect and that a simpler model that ignores RNA elasticity could be as successful (or unsuccessful) as ours. Indeed, we can build a model that identifies the RNA extension where state transition occurs as the contour length of each state. State transition is thus forced when one state reaches its maximum length. However, such model would require empirical parameters, and these parameters are not constant as experimental conditions such as the length of the handle DNA change. On the other hand, the number of parameters needed by our model is optimized in some sense. For the WLC model, the only parameters are the persistent length $\xi_p$ and the phosphate-to-phosphate distance of RNA $a$. For the hairpin model, we need the folding energy $\epsilon_{12}$ and $\epsilon_{23}$ as inputs, which can be predicted from established models (although they are somewhat complicated) [6].

The formulas Eq. (12) and Eq. (14) we used to compute RNA elasticity are only approximate, and there are revised formulas that include corrections in the form of series expansions [5]. Nevertheless, the comparison between our theoretical predictions and the experimental results of Greenleaf *et al.* does not suggest that the approximate formulas are unsuitable. Instead, the hairpin model providing folding energy is more likely to be wrong. We can easily make amendments by giving different folding energy, which can also be determined by regression if there are sufficient data.


[1] W. C. Winkler and R. R. Breaker, Annual Review of Microbiology **59**, 487 (2005), URL http://www.annualreviews.org/doi/abs/10.1146/annurev.micro.59.030804.121336.

[2] W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, and S. M. Block, Science **319**, 630 (2008), URL http://www.sciencemag.org/content/319/5863/630.abstract.

[3] P. Nelson, *Biological Physics* (W. H. Freeman, 2008), updated 1st ed.

[4] M. Kardar and L. Mirny, *8.592 Lecture Notes* (2011).

[5] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995), URL http://pubs.acs.org/doi/abs/10.1021/ma00130a008.

[6] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, Journal of Molecular Biology **288**, 911 (1999).



### Acknowledgments

The author acknowledges Prof. Mehran Kardar and Prof. Leonid Mirny for offering the subject 8.592, which is an exciting adventure in the boundary of statistical physics and biology. The author also appreciates their generosity and patience for giving him a extension.