Published in final edited form as:

Mol Phys. 2006 April 20; 104(8): 1343–1352.

Modelling RNA folding under mechanical tension

JEFFREY R. VIEREGG^{†,§} and IGNACIO TINOCO JR^{‡,§,*}

[†] *Department of Physics, University of California, Berkeley, USA*

[‡] *Department of Chemistry, University of California, Berkeley, USA*

[§] *Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

Abstract

We investigate the thermodynamics and kinetics of RNA unfolding and refolding under mechanical tension. The hierarchical nature of RNA structure and the existence of thermodynamic parameters for base pair formation based on nearest-neighbour interactions allows modelling of sequence-dependent folding dynamics for any secondary structure. We calculate experimental observables such as the transition force for unfolding, the end-to-end distribution function and its variance, as well as kinetic information, for a representative RNA sequence and for a sequence containing two homopolymer segments: A.U and G.C.

1. Introduction

The biomolecular folding problem is one of the classic problems of biophysics. Proteins and nucleic acids are synthesized as linear polymer chains. They must then spontaneously and rapidly fold into their three-dimensional active forms. Misfolded molecules are inactive or even harmful, so the process must occur with high fidelity, even in the crowded, heterogeneous cellular environment. Understanding how this takes place is essential to designing molecules with new functions. Ideally, one would like to be able to predict both the final form of a molecule and the pathways it takes to reach that form given only the molecule's amino acid or nucleotide sequence.

1.1. The RNA folding problem

Ribonucleic acids (RNAs) are an attractive system in which to study the folding problem, and offer several distinct advantages relative to proteins. RNA plays many vital roles within the cell, including not only information transfer (messenger RNA) but also structural, regulatory (RNAi, riboswitches), and catalytic (ribozymes, ribosomes) functions, many of which were previously thought to be only performed by proteins. Folding is important for all of these functions. Despite this diversity of function, RNA is relatively simple compared to proteins. RNA is composed of only four monomer units with similar chemical properties. Even more importantly, RNA folding is hierarchical [1]. Watson–Crick pairing and base stacking interactions produce secondary structure elements (A-form helices and single-stranded loops, bulges, and junctions [2]— see figure 1) which are substantially more stable than the tertiary structure resulting from these elements' interactions with one another. This is in stark contrast to proteins, for which secondary and tertiary interactions are similar in magnitude, making it extremely difficult to predict structure from sequence alone. Finally, many tertiary interactions are mediated by divalent cations, which allows for experimental control over the two levels of organization. In this article, we focus on calculational models for predicting thermodynamics

*Corresponding author. Email: INTinoco@lbl.gov.

and kinetics of secondary structure formation. Other researchers are investigating the corresponding problem for RNA tertiary folding [3].

Understanding RNA secondary structure is aided considerably by the observation that thermodynamic parameters such as Gibbs free energies and melting temperatures can be predicted with remarkable accuracy by considering only nearest-neighbour stacking interactions and additive contributions from loops and bulges [4]. Together with a systematically compiled database of oligonucleotide melting data [5], this allows calculation of likely secondary structures for any RNA sequence using programs such as MFOLD [6]. These thermodynamic data, though coarse-grained over atomic-scale features, may also form a basis for predicting folding pathways and kinetics.

1.2. Probing folding with mechanical force

Traditionally, researchers have studied folding by using rapid temperature jumps or changes in the chemical environment (pH, ionic strength, etc.) to induce molecules to fold or unfold. Biological processes generally do not involve large changes in temperature or chemical environment. The structures formed by these methods can be different than those formed in the cell [7]. The folding and unfolding produced by these methods can also be difficult to characterize, as there is no obvious reaction coordinate or point of reference against which to track the molecules' behaviour.

Mechanical force offers several advantages over thermal or chemical denaturation. The force axis creates a natural reaction coordinate along which molecules' evolution can be followed in real time. In most experimental set-ups, it is also possible to simultaneously vary other parameters such as ionic strength and temperature to see how they affect the (un)folding process triggered by force. This is often impossible for thermal and chemical denaturation (as, for example, divalent cations catalyse backbone phosphodiester cleavage at elevated temperatures). Force is also important *in vivo*; the densely packed cellular matrix opposes motion (including folding) through hydrodynamic drag and numerous enzymes have been found to exert force on their substrates, including RNA polymerase [8] and HIV reverse transcriptase [9]. Force also readily lends itself to single-molecule techniques which are capable of extracting information not accessible by traditional methods. By following an individual molecule's trajectory in real time, one can detect and characterize rare and short-lived intermediate states which, although undetectable in bulk experiments, may play key roles in the folding process.

In this paper, we present calculations with a model for force-induced unfolding of RNA. By combining the advantages of mechanical force unfolding with the power and simplicity of RNA thermodynamics, we show that a surprising amount of insight into the folding process can be gained from relatively simple models. As further experiments are published, these models can be tested and refined. Hopefully, the combination of theory and experiments with force (un)folding will enable substantial progress towards understanding the complex behaviour of RNA and other biomolecules.

2. Sequential-step model

In modelling the force-induced unfolding and refolding of RNA secondary structure, we make the common assumption that base pairs open and close only at the boundary between the double-stranded region and the single-stranded ends to which force is applied. This is a reasonable assumption at room temperature and for forces above a few pN, as the energy required to bring two nucleotides close enough to bond is large compared to $k_B T$. This condition, which is not applicable to thermal or chemical denaturation, simplifies the problem dramatically. If bases are allowed to pair in any order, a sequence with N bases paired in the

fully folded state has $\sim 2^N$ possible states, versus only N states for the sequential model. A related assumption is that no significant misfolding occurs upon closure of loops and bulges. This assumption is harder to justify but can be tested experimentally.

2.1. Equilibrium calculations

We number the possible states by terminal closed base pair, with state 1 the completely base paired (folded) state and N the completely single-stranded (unfolded) state. The incremental change in Gibbs free energy upon opening the n th base pair of an RNA molecule stretched by mechanical force F can be written as

$$\Delta G_{\text{inc}}(n, F) = g_i(n) - g_s(n, F), \quad (1)$$

where g_i is the intrinsic free energy of breaking the base pair (i.e. the free energy change at zero force) and g_s is the free energy associated with stretching the newly single-stranded sections of the molecule. We choose to work in the force ensemble (rather than extension) as this more readily facilitates comparison between molecules. The intrinsic free energy is sequence dependent but the stretching term depends only on the number of bases released at each step (two for a helix section, three or more for a bulge or loop). The intrinsic free energies are assumed not to vary with force; this is a reasonable approximation, as both hydrogen bonds and inter-base stacking interactions are known to be short ranged.

The intrinsic free energies can be obtained from many sources, including microscopic theory [10] or experiment [11]. For a general sequence, however, the nearest-neighbour thermodynamic rules are a good choice, particularly as they are the basis for the MFOLD program [12] which is often used to calculate secondary structure. The stretching free energies are obtained using a suitable polymer model. Cocco *et al.*, whose kinetic model we explore, here, use a freelyjointed-chain-like (FJCL) model. The worm-like-chain model has been shown to reproduce nucleic acid polymer behaviour over a larger range of forces [13], but has the disadvantage that it must be numerically integrated to obtain g_s at each force. The two models agree within a few percent from 5–30 pN; we use the FJCL model here for simplicity.

With the incremental free energies from equation (1), it is a straightforward matter to calculate the energy of each state (relative to the fully folded state) and the partition function for the system:

$$\begin{aligned} G(m, F) &= \sum_{n=1}^m \Delta G_{\text{inc}}(n, F); \\ Z(F, T) &= \sum_{m=1}^N \exp\left(-\frac{G(m, F)}{k_B T}\right). \end{aligned} \quad (2)$$

With the partition function, any desired statistical mechanical quantity can be computed. In particular, we can calculate the free energy landscape for the molecule and the state probability distribution function (pdf) at any force, as shown in figure 2. The state pdf can be converted to an end-to-end distance distribution using one of the polymer models described above. This enables the calculation of observable equilibrium quantities such as the mean extension $\langle x \rangle$ and its variance σ^2 . The utility of these observables in interpreting experimental data will be discussed in section 3.3.

2.2. Kinetic behaviour

The kinetic model we examine in this paper was first proposed by Cocco *et al.* in 2001 for DNA unzipping [14] and later extended to RNA [15]. It has been shown to accurately reproduce the behaviour of a simple hairpin (P5ab) from the *T. thermophila* ribozyme. To calculate the

kinetic behaviour of the molecule, Cocco *et al.*'s model returns to the idea of sequential opening and closing steps. Transitionstate theory predicts that the rate for decay of a metastable state should have the general form $rate = A \exp(-\Delta G^\ddagger/k_B T)$ where $\Delta G^\ddagger$ is the free energy of the transition state relative to the initial state and A is a constant. The transition state for base pair opening is taken to be the state where the hydrogen bonds and stacking interactions have been disrupted but the opening nucleotides have not yet been stretched to their equilibrium length. $\Delta G^\ddagger$ can then be identified as $g_i(n)$ for opening one base pair and $g_s(n, F)$ for closing one base pair, giving the respective rates:

$$k_o = A \exp\left[-\frac{g_i(n)}{k_B T}\right]; \quad k_c = A \exp\left[-\frac{g_s(n, F)}{k_B T}\right]. \quad (3)$$

Determining an *a priori* value for the common prefactor A is difficult [16]. As a first approximation, one can set it equal to the inverse self-diffusion time for a nanometre-sized object in water ($\approx 5 \times 10^6 \text{ s}^{-1}$) [15]. This has been shown to at least provide the correct order of magnitude for the time scale of the problem. Practically speaking, A might be treated as a free parameter for fitting experimental data at a given ionic strength, pH and temperature. It should not, however, vary from molecule to molecule, which will provide a good test of this model as more experimental data are collected.

Notice that the opening rate has no force dependence. This is reasonable for the brittle interactions described above. The closing rate, on the other hand, depends on the applied force through $g_s(F)$. Comparison with equation (1) shows that detailed balance is satisfied, ensuring that the kinetic model is consistent with the thermodynamic relations of section 2.1.

Using the elementary rates from equation (3), we can construct the master equation for the molecule as a whole. Defining $\rho_n(t)$ as the probability for the molecule to be in state n at time t , we have

$$\begin{aligned} \rho_n(t) = & -[k_o(n) + k_c(n, F)]\rho_n(t) + k_o(n-1)\rho_{n-1}(t) \\ & + k_c(n+1, F)\rho_{n+1}(t) \end{aligned} \quad (4)$$

or, more compactly,

$$\rho_n(t) = - \sum_m T_{nm} \rho_m(t), \quad (5)$$

where the rate matrix T_{nm} is tridiagonal with non-zero elements $T_{m,m} = k_o(m) + k_c(m, F)$, $T_{m+1,m} = -k_o(m)$, and $T_{m-1,m} = -k_c(m, F)$. This rate matrix contains all the dynamic information about the molecule, including kinetic information such as folding and unfolding rates and also the equilibrium properties. In the following two sections, we apply this model to a typical biological hairpin and to a hypothetical hairpin of two linked homopolymers which, although unlikely to occur naturally, serve as limiting cases and display interesting behaviour in their own right.

3. A typical hairpin

To illustrate the utility of this approach for calculating the properties of biologically important RNA molecules, we first apply it to a RNA hairpin chosen arbitrarily from the mRNA which encodes the S3 ribosomal protein of the *E. coli* bacterium. We used the MFOLD program to calculate a probable secondary structure for the mRNA sequence, then selected a 29 nucleotide sequence (nt 186–214 counting from the AUG start codon) which is predicted to form a hairpin with an 8 nt loop and a 3 nt bulge located approximately halfway along the helical stem, as

shown in figure 1. There is nothing unusual about this hairpin in terms of structure or nucleotide content (62% base pairing, 56% base pairs G–C), nor is it thought to have any particular biological function; it rather serves as an example of the method's general applicability.

3.1. Equilibrium properties

Using the nearest-neighbour energy rules and equation (2), we can calculate the free energy landscape and state probabilities as functions of applied force (figure 2). Even a simple hairpin such as this shows interesting features. The Watson–Crick base pairs require energy input to break; they provide a substantial energetic barrier at zero force ($\Delta G = 7.5 \text{ kcal mol}^{-1} = 12.2 k_B T$ at 37°C) which prevents spontaneous unfolding. As force increases, the energy landscape tilts under the influence of g_s , until, at a force of 8.23 pN, the fully open and fully closed states have equal probabilities; this is referred to as the critical force or $F_{1/2}$. At still higher forces, the molecule exists predominantly in its unfolded form. Although the destabilizing effects of the bulge are apparent as a dip in the free energy landscape at $n=6$, there is no force at which the state with the bulge open, but not the terminal loop, has a significant probability of occupation. The thermodynamic model predicts this molecule should have only two significantly populated equilibrium states, folded and unfolded.

3.2. Kinetics

Applying Cocco *et al.*'s model, we can calculate the time-dependent behaviour of the molecule using equation (5). We have found two modes of calculation helpful. The first approach is analytic; we diagonalize the rate matrix T_{nm} and examine the eigenvalues and eigenvectors over a range of forces [17]. Each eigenvalue represents the rate (forward)backward of a transition with first-order kinetics; the corresponding eigenvector shows the flow of probability for that transition. Denoting the population (or occupation probability) of each state by the vector $\mathbf{P}(t)$ and the (column) matrix of eigenvectors by $\mathbf{E}$, the overall kinetics of the molecule can be written as

$$\mathbf{P}(t) = \sum_m^N C_m \Gamma_m e^{\lambda_m t} \quad (6)$$

with Γ_m and λ_m the m th eigenvector and eigenvalue and the weighting coefficients C_m determined by the initial conditions:

$$C_m = \sum_I \mathbf{E}_{mI}^{-1} \mathbf{P}(0)_I \quad (7)$$

Figure 3 shows an example of this type of calculation. Most of the transitions correspond to rapid interconversion among groups of states with similar configurations and are not helpful for understanding the mechanism of the (un)folding reaction. Transitions between substantially different states, such as closing of a loop or bulge, are slower, often by orders of magnitude. For molecules with simple behaviour (two or few-state switching), it is easy to calculate experimentally observable rates and to gain insight into a molecule's folding and unfolding pathways. This method has also been applied to folding after thermal denaturation [18], but the rapid scaling of the number of states with molecule length for that type of experiment can be a problem.

The eigenvalue spectrum for the S3 hairpin is shown in figure 4(a). There is also another eigenvalue, zero, corresponding to the equilibrium state. Its eigenvector is identical to the probability distribution computed using the thermodynamic model of section 3. There is a marked separation between the lowest two non-zero eigenvalues and the others at all forces. The eigenmodes with fast eigenvalues reflect fast interconversion of states within the $n = \{1,$

2, 3} and $n = \{6, 7, 8\}$ manifolds. The smallest eigenvalue and its eigenvector represent the rate-limiting step for the folding of the hairpin. As figure 4(b) (top plot) shows, this involves a transition between the unfolded and fully folded states without long-lived intermediates, consistent with the two-state behaviour predicted by equilibrium thermodynamics. A second distinguishable transition is predicted by the other distinct eigenvalue. Its eigenvector (figure 4(b), bottom) connects the bulge-opened $n = 6$ state to the folded and unfolded states. This reflects kinetic trapping by the dip in the free energy landscape caused by the bulge; molecules in this state reside there for a time before escaping to either the completely folded or unfolded state. The lifetime of this state (0.84 ms at the critical force, figure 3 upper inset) is given by the corresponding eigenvalue. If the bulge is deleted from the sequence, this slow transition disappears. These results show that even a simple kinetic model is capable of providing insight into the mechanisms of RNA secondary structure folding.

We can also predict single-molecule observables such as opening and closing rates using this model. For eigenmodes connecting only two states, such as the folding pathway of the S3 hairpin, the eigenvalue λ_l is equal to the sum of the forward and reverse rates $k_{ij} + k_{ji}$ for the transition $(n = i) \rightleftharpoons (n = j)$ represented by the corresponding eigenvector. Further, detailed balance requires that $k_{ij} \cdot P_i = k_{ji} \cdot P_j$, where P_i and P_j are the (equilibrium) probabilities for states i and j . The forward and reverse rates can then be calculated as follows:

$$k_{ij} = \frac{\lambda_l}{1 + P_i/P_j}; \quad k_{ji} = \frac{\lambda_l}{1 + P_j/P_i}. \quad (8)$$

For molecules with well-separated eigenvalues such as the S3 hairpin, these rates can be assigned to observable transitions. Figure 5 shows the predicted rates as functions of force for opening and closing of the hairpin. The two rates are equal at the critical force, again reflecting detailed balance. To the extent that a molecule behaves in a two-state manner, the slopes of the rates, $d(\ln k)/dF$, give the locations of the effective transition states. The critical force and these slopes are independent of the rate prefactor A . As noted earlier, Cocco *et al.* were able to reproduce both the magnitude and force dependence of the opening and closing rates for the P5ab hairpin reported by Liphardt *et al.* [19] with reasonable accuracy using this method. As more experimental data are published, we will be able to test the accuracy of the model more rigorously.

A second approach to understand the kinetics of a molecule is to numerically integrate the master kinetic equation (equation (5)) and perform a Monte Carlo type simulation of the molecule's dynamic behaviour. This is much slower than the analytic method described above, but it accesses the entire spectrum of states and transitions instead of focusing on the slowest eigenvalues. Such a simulation also generates single-molecule trajectories which can be compared to experimental data. Finally, this approach readily allows incorporation of experimental conditions such as force ramping [19, 20] and fluctuations due to the apparatus [21]. Section 4.2 illustrates the utility of this method for both S3 and an interesting artificial hairpin.

3.3. Experimental observables

The thermodynamic and kinetic models described above contain a great deal of information about the state of RNA molecules under mechanical tension, but the variable they track (the last closed base pair) can be difficult to directly observe. In order to compare calculations to experimental data, we must convert the state information into a measurable quantity. As much of the recent work on unfolding single molecules has been done with optical or magnetic tweezers, end-to-end extension is a natural choice. As discussed in section 2.1, several polymer models describe the behaviour of nucleic acids when stretched by mechanical tension. Strictly

speaking, these models are valid only in the limit of long polymers, but the flexibility of single-stranded nucleic acids (persistence length of ssRNA ≈ 1 nm) allows reasonable confidence in their accuracy for even a few nucleotides. Still, better modelling of oligonucleotide extension versus force would be helpful for more accurate comparisons.

Using the equilibrium thermodynamic model of section 2.1, we calculate two common experimental time-averaged observables, the mean extension and its variance, as functions of force applied to the S3 mRNA hairpin (figure 6). The FJCL model was used to calculate the end-to-end extension of the molecule's possible states.

The mean extension curve has a sigmoidal shape characteristic of two-state systems with no other inflection points to signal the presence of an intermediate state. Below about 7 pN, the extension stays constant at zero; the molecule is fully folded at these forces. Above 10 pN, the molecule's slow increase in extension is due solely to the entropic stretching of the unfolded RNA chain, as shown in the figure. The transition region is roughly symmetric around the critical force with a slope of about 7.0 nm pN^{-1} . Either this slope or the width of the transition region (in $\langle x \rangle$ or σ^2 ; section 4.1) can be used as a measure of the cooperativity of the (un)folding transition.

The variance in extension at different forces is another experimental observable that can be calculated using the thermodynamic model. In the case of the S3 hairpin, the folding transition is distinct and large enough (12 nm) to be observed directly by most experimental techniques, but for other molecules, particularly sequences with intermediates of similar extension, this may not be possible. In these cases, the increased fluctuation amplitude when multiple states are thermodynamically accessible may be a valuable experimental signature, particularly for experiments with significant low-frequency noise or drift. Of course, these fluctuations will only be observed if the energy barrier between the states is small enough that interconversion occurs on the experimental timescale. For this molecule, we see one peak in the variance whose position (8.27 pN) is in excellent agreement with the critical force calculated earlier. The amplitude of the peak (28.4 nm^2) is consistent with the molecule's extension fluctuating between zero and its full length. The full width at half max (FWHM) of the peak is 1.41 pN.

3.4. Alternate observable: fluorescence transfer

Another popular experimental technique for observing single molecule behaviour exploits resonant energy transfer from a donor to an acceptor fluorophore (FRET). The efficiency of energy transfer, ϵ , depends strongly on the distance between the two fluorophores and is expressed by the well-known Förster formula:

$$\epsilon = \frac{1}{1 + (R/R_0)^6}. \quad (9)$$

For typical dye pairs, R_0 is between 1 and 5 nm. This limits the range of distances which FRET can examine, but it also offers great specificity. To date, most FRET studies of RNA folding have focused on tertiary interactions [22], but the technique could also provide unique information about folding of secondary structures. FRET could determine how often, and potentially in what order, specific base pairs are formed. This would be particularly valuable for studying kinetically important processes like loop and bulge closure that may involve misfolded structures and conformational searching. The thermodynamic and kinetic models presented in this paper can predict FRET signals as well, since they specify the base pairing state of each nucleotide at arbitrary times and forces.

4. Poly(G · C) and Poly(A · U): an atypical hairpin

The S3 mRNA hairpin examined in section 3 is an example of a ‘typical’ biological RNA secondary structure. In this section, we examine the behaviour of an artificial hairpin which contains segments approximating the ideal homopolymers poly(A · U) and poly(G · C). They display marked differences in folding characteristics and serve as limiting cases for hairpin behaviour.

4.1. Cooperativity and transition behaviour

As shown in figure 1, the hairpin contains sequential segments of thirty A · U and thirty G · C base pairs, closed by a AAAU tetraloop. The weaker A · U base pairs ($\Delta G(T = 37^\circ\text{C}, \text{zero force}) = 0.90 \text{ kcal mol}^{-1}$) are at the open end of the hairpin, followed by the much stronger ($\delta G = 3.30 \text{ kcal mol}^{-1}$) G · C base pairs. Figure 7 shows the calculated mean extension and variance for the molecule as a function of force. There is a clear intermediate state at forces between 15–27 pN, in which the A · U base pairs are broken but the G · C section remains folded.

As discussed in section 3.3, we can determine the cooperativity of a folding transition by examining its behaviour versus force. The results of these calculations are shown in table 1. Of the two homopolymer sections, the G · C transition is much sharper than that of the A · U. If the three base bulge is removed from the S3 mRNA, its folding transition is predicted to be intermediate to these cases, as one might expect, since $\frac{AA}{UU}$ and $\frac{GG}{CC}$ are nearly the least and most stable nearest-neighbour pairs in the nearest neighbour scheme. Interestingly, the native S3 hairpin is (by the slope and σ^2 methods) less cooperative than the AU homopolymer sequence; this is most likely due to the dip in the free energy landscape caused by the destabilizing bulge.

4.2. Diffusion and fluctuations near the transition

In addition to being much less cooperative than the poly-GC transition, the poly-AU sequence exhibits interesting kinetic behaviour near the critical force. The FJCL model predicts that, at a force of 11.55 pN, g_s is equal to g_i for an 2(A · U) stack; there is no energy cost to break or form a base pair. The G · C base pairs which follow are still very stable at this force and effectively form a rigid barrier against further unfolding. The combination of these two effects should, according to Cocco *et al.*’s model, result in random motion of the unfolding fork within the AU region analogous to bounded one-dimensional diffusion. We can follow that motion by means of a Monte Carlo simulation as described in section 3.2. Figure 8(b) shows the location of the unfolding fork as a function of time. The time evolution of the unfolding fork is clearly different from the two-state switching of the S3 hairpin (figure 8(a)). From equation (3), D , the diffusion constant for the motion, should be $A/2 \approx 2.5 \times 10^6 \text{ bp}^2 \text{ s}^{-1}$, so individual steps will be very difficult to resolve experimentally. The increase in variance of the extension over time, however, may be measurable. An exact solution of the bounded one-dimensional diffusion problem (see, for example, [23]) predicts that the variance should approach a steady-state value of $L^2/12 = 75 \text{ bp}^2$ with a characteristic time of $L^2/D \approx 0.36 \text{ ms}$ for L , the number of base pairs in the flat energy region, equal to 30. Our simulation reproduces that behaviour, with a similar characteristic time and final variance for a 10 s Monte Carlo simulation of 75.5 bp^2 . If diffusion-like behaviour could be observed experimentally, it would help validate Cocco *et al.*’s model and provide a measurement of the prefactor A independent of the models used for g_i and g_s .

Figure 8 also illustrates the unique information afforded by single molecule studies. In these simulations and experimentally [19], it is possible to follow the folding process in real time.

This enables direct observation of intermediate states such as the bulge-opened $n = 6$ state in S3 which would be extremely difficult to detect and characterize in, for example, a bulk melting experiment. By tracking molecular trajectories in real time, we can go beyond ensemble averages and start to learn how folding actually happens.

5. Conclusions

Understanding how biological polymers such as nucleic acids and proteins fold into their three-dimensional active states is a key problem in modern biophysics. Recent years have seen tremendous progress made on both experimental and theoretical fronts. New experimental techniques allow researchers to follow folding of single molecules with improved distance (nm-scale) and time (ms) resolution. Constructing comprehensive theories with monomeric resolution has continued to prove difficult, however, due both to the numerical complexity of the problem and the difficulty of finding a proper reaction coordinate against which to describe the folding process.

For RNA molecules unfolded by mechanical force, both of these issues may be resolved. The application of force creates a ready-made reaction coordinate, and an extensive database of thermodynamic data allows accurate computation of secondary structure free energies based on nearest-neighbour and additive loop and bulge interactions. The hierarchical nature of RNA folding means that larger-scale tertiary interactions can then be added as a perturbation, allowing computation of even complex RNA folds on a reasonable time scale.

In this paper, we have tried to demonstrate the power that this affords by calculating thermodynamic and kinetic folding properties for both a typical mRNA sequence and for a hypothetical sequence of homopolymers which illustrates extreme behaviour. Using simple models, we can calculate free energies of folding, the presence or absence of long-lived intermediate states, and rates of both folding and unfolding for any RNA whose secondary structure can be specified. This provides a powerful tool for understanding experimental results and predicting new directions for exploration. As more experimental results come in, these simple models can be tested and refined. This is a very exciting time for the field of RNA folding.

Acknowledgements

One of us (IT) has been a colleague and friend of Bob Harris for over 40 years. He is the person who has always been willing to explain quantum mechanics, statistical mechanics, and the meaning of life. Bob is a rare and valuable treasure.

Also, we would like to thank Laurent Lacroix and Sophie Dumont for their critical reading of the manuscript. This work was supported by NIH Grant GM-10840. JV was partially funded by a NSF Graduate Research Fellowship.

References

1. Tinoco I Jr, Bustamante C. *J Molec Bio* 1999;293:271. [PubMed: 10550208]
2. V. A. Bloomfield, D. M. Crothers, and I. Tinoco Jr, *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA, 1999).
3. Isambert H, Siggia ED. *Proc Natl Acad Sci USA* 2000;97:6515. [PubMed: 10823910]
4. M. E. Burkard, D. H. Turner, and I. Tinoco Jr, in *RNA World*, 2nd ed., edited by R. F. Gesteland, T. R. Cech, and J. F. Atkins (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1999), pp. 233–264.
5. Mathews DH, Sabina J, Zuker M, Turner DH. *J Molec Bio* 1999;288:911. [PubMed: 10329189]
6. Zuker M. *Nucleic Acids Res* 2003;31:3406. [PubMed: 12824337]
7. Uhlenbeck OC. *RNA* 1995;1:4. [PubMed: 7489487]
8. Wang MD, Schnitzer MJ, Yin H, Landick R, Gelles J, Block SM. *Science* 1998;282:902. [PubMed: 9794753]

9. Lu H, Macosko J, Habel-Rodriguez D, Keller RW, Brozik JA, Keller DJ. J Biol Chem 2004;279:54529. [PubMed: 15385563]
10. Cocco S, Monasson R. Phys Rev Lett 1999;83:5178.
11. Aalberts DP, Parman JM, Goddard NL. Biophys J 2003;84:3212. [PubMed: 12719250]
12. Zuker M. Nucleic Acids Res 2003;31:3406. [PubMed: 12824337]
13. Bustamante C, Marko JF, Siggia ED, Smith SB. Science 1994;265:1599. [PubMed: 8079175]
14. Cocco S, Monasson R, Marko J. Phys Rev E 2002;65:041907.
15. Cocco S, Monasson R, Marko J. Euro Phys J E 2003;10:153.
16. Hänggi P, Talkner P, Borkovec M. Rev Mod Phys 1990;62:251.
17. Cieplak M, Henkel M, Karbowski J, Banavar JR. Phys Rev Lett 1998;80:3654.
18. Zhang W, Chen SJ. J Chem Phys 2003;118:3413.
19. Liphardt J, Onoa B, Smith SB, Tinoco I Jr, Bustamante C. Science 2001;292:733. [PubMed: 11326101]
20. Onoa B, Dumont S, Liphardt J, Smith SB, Tinoco I Jr, Bustamante C. Science 2003;299:1892. [PubMed: 12649482]
21. Manosas M, Ritort F. Biophys J 2005;88:3224. [PubMed: 15764661]
22. Zhuang X, Rief M. Curr Opin Struct Bio 2003;13:88. [PubMed: 12581665]
23. M. Deserno, One-dimensional diffusion on a finite region. Available online at: www.mpip-mainz.mpg.de/~deserno/science_notes/diffusion/diffusion.pdf (accessed 17 May 2005).

Figure 1.

(a) Predicted structure for a hairpin from the *E. coli* S3 ribosomal protein mRNA. Several types of secondary structures occur: two helical stems (nucleotides 1–5, 9–12), a three base bulge (nt 6–8), and the eight nt terminal loop. (b) Hypothetical hairpin with linked A.U and G.C homopolymer sections. This hairpin consists of one long helical stem and a terminal loop.

Figure 2.

(a) Free energy landscape (total free energy for each state) for the S3 hairpin at 0 pN (solid line), 8.23 pN (dotted) and 12 pN (dashed). The dip at $n=6$ is due to the unfolding of the three base bulge. (b) State occupation probabilities at the same forces. Increased force favours extended (unfolded) states over compact (folded) states.

Figure 3.

Time scales of RNA folding. Main figure shows the evolution of state occupation probability for the S3 hairpin from the fully folded state to equilibrium at the critical force. Insets: kinetic trapping. Molecules in the $n=6$ state escape (upper inset; initial condition $P(n=6)$ set to 1) with a characteristic time which is short compared to folding but long compared to equilibration of neighbouring states (lower inset, same initial condition as main plot).

Figure 4.

(a) Eigenvalue spectrum (log scale) versus force for the S3 hairpin. The two smallest non-zero eigenvalues are separated by at least a factor of 50 at all forces. They in turn are separated by several orders of magnitude from the rest. (b) Eigenvectors corresponding to the folding pathway (top plot) and kinetic trapping (bottom) at the critical force.

Figure 5.

Folding (dashed) and unfolding (solid) rates for the S3 hairpin. The unfolding (folding) rate increases (decreases) with force, reflecting the free energy contribution from stretching. $d(\ln k)/dF$ changes with force, indicating a change in the location of the effective transition state.

Figure 6. Predicting equilibrium observables: (a) mean extension (solid) for the S3 hairpin and FJCL extension (dashed) of the fully unfolded molecule versus force. The molecule unfolds completely from 7–10 pN. (b) Calculated extension variance versus force. The maximum occurs at the critical force of the transition.

Figure 7. Linked homopolymers: (a) mean extension for the AUGC hairpin versus force. The poly(G · C) section of the molecule unfolds at a much higher force, and over a narrower range. (b) Calculated extension variance versus force. The widths of the peaks indicate the relative cooperativity of the two transitions.

Figure 8. Tracking time evolution: (a) two-state switching of a typical hairpin (S3) at the critical force. Except for occasional stops in the bulge kinetic trap, S3 switches between the $n = \{1, 2, 3\}$ folded state manifold and the $n = 10$ fully-unfolded state. (b) Sample trace from a Monte Carlo simulation of the linked homopolymer hairpin. The unfolding fork diffuses over the flat potential surface.

Table 1
Predicting folding cooperativity. All values normalized by number of base pairs (un)folded. Larger values correspond to greater cooperativity.

Transition	$\Delta G(\text{kcal mol}^{-1}, F = 0)$	$X(F)$ Width ^a	$X(F)$ Slope ^b	$\sigma^2(F)$ FWHM ^c
A · U homopolymer	26.1	0.05	0.19	0.28
G · C h.p. + loop	94.3	1.35	1.45	1.80
S3 (no bulge)	13.3	0.82	0.26	0.31
S3 (native)	7.5	0.51	0.15	0.20

^a $F_{1/2}$ /Force range for 10–90% of the change in extension.

^b dX/dF at $F_{1/2}/(\Delta X/F_{1/2})\delta X$ is the total extension change.

^c $F_{1/2}$ /FWHM of variance peak.