RNA Folding Dynamics Studied with Structure-based Models

Michael Faber and Stefan Klumpp

Max Planck Institute of Colloids and Interfaces, Science Park Golm, 14424 Potsdam, Germany
E-mail: michael.faber@mpikg.mpg.de

RNA molecules form three-dimensional structures as complementary bases form bonds and the molecule coils. These structures determine the function and biochemical activity of the molecule. For example, the presence or absence of a specific RNA structure can invoke transcriptional pauses or terminate the transcription altogether. We have developed a structure-based model for studying the folding dynamics of RNA secondary structures. To simulate the dynamics, we use a Monte-Carlo method with Metropolis rates, where the basic steps are the closing or opening of one native contact. We apply this model to the folding and unfolding of simple RNA structures in the presence and absence of an external force.

1 Introduction

RNA is a linear polymer made out of four different bases: adenine (A), cytosine (C), guanine (G) and uracil (U). Two pieces of an RNA molecule can connect via hydrogen bonds between complementary bases (AU and GC), such that the RNA folds into a three dimensional structure. RNA structures are usually described as a hierarchy of structures: the sequence of bases in the molecule is called the primary structure, the set of all base pairings the secondary structure and the three-dimensional shape of the molecule including all other structural elements the tertiary structure. Typically, this hierarchical description reflects the hierarchy of the folding process where the primary structure determines the secondary which in turn determines the tertiary contacts[1]. Great effort has been done on understanding and predicting the secondary structure of RNA molecules[2]. Moreover, over the last decade single molecule experiments using optical tweezers were performed on a number of RNA structures to study their stability, their force dependence and their dynamics. In cells, RNA structures often fold while the RNA is transcribed. In such cases, the dynamics of folding is typically crucial for the function of the RNA. An example is the formation of hairpins during transcription, which can invoke transcriptional pauses or terminate the transcription altogether[3].

In the following we present a simple model which aims at describing the dynamics of RNA secondary structures. The model we have developed is a structure based model, i.e. we concentrate on the native contacts of a given RNA structure and study its dynamics. Structure-based models have been used extensively in studies of protein folding. They are based on the principle of minimal frustration that states that functional sequences have been selected to avoid energetic frustration to ensure rapid folding[4]. As a consequence, the dynamics of folding is expected to be governed by the same interactions that govern the folded state. The same arguments should also apply to the folding of structured RNAs, and indeed similar argument have occasionally been used for RNA[5]. Here we use a Monte Carlo method to simulate the folding dynamics. We will show results on the stability of

secondary structures and the distributions of folding and unfolding times as well as on force induced unfolding.

2 Model

The secondary structures we consider consists of five basic structural motifs that arise from base pairing: simple unconfined single stranded pieces of RNA, helical regions of subsequent paired bases, hairpin loops that form an end to a helical region, internal loops with more then one outgoing helical region and bulges. If we number the bases of an RNA molecule $\{1 \dots N\}$ the secondary structure can be described as a set of pairs (i, j) denoting the formed base pairs. Here we consider only structures without pseudo knots which is a common restriction in the prediction of secondary structures. Therefore two base pairs (i, j) and (i', j') must either fulfill $i < i' < j' < j$ or $i' < i < j < j'$. These conditions ensure that no base pair can form between a base in the region separated by the first base pair (i, j) and a base outside that region.

Our RNA model is structure based. We take the RNA to be a sequence of bases where only specific, predefined contacts can be made between bases of that sequence. These positions are defined by the native (folded) structure of the RNA molecule. Here we restrict ourselves to contacts that form by base pairing, but additional types of contacts could also be included. Then the dynamics of the RNA molecule are analyzed using a Monte-Carlo method with Metropolis rates. The basic steps are the closing and opening of contacts. This is done by choosing a base pair randomly from the list of possible base pairs. If the chosen base pair exists already, then it might open, and if it does not exist, it may close. The probabilities for the opening or closing moves are calculated from the free energy difference of the structure before and after the step. We assume that the free energy of a structure can be calculated as a sum of energy contributions from the different structural motifs. Forming a base-pair is energetically favorable, on the other hand the formation of a loop constrains the RNA molecule which is entropically costly

$$G_{\text{tot}} = \sum_{\text{all basepairs}} G_{\text{basepair}} + \sum_{\text{all loops}} G_{\text{loop}}. \tag{1}$$

Here we use a simple parametrization of the free energies: We chose each base pair to contribute $G_{\text{basepair}} = -2$ kcal/mol. Energy contributions of loops depend logarithmically on the loop length. For hairpin loops, which require $n \geq 3$ bases in the loop, we take $G_{\text{hairpin loop}}(n) = \big(5 + \ln(n/3)\big)$ kcal/mol, while internal loops and bulges are assigned $G_{\text{int loop}}(n) = \big(2 + ln(n)\big)$ kcal/mol.

3 Simulation Results

In the following we will use a contrived and simple hairpin structure to demonstrate key features that our model describes. Our model structure consists of a loop closed by consecutive identical base pairs. Despite its simplicity, this structure already shows some universal properties which one can expect to find in more complicated systems. First we look at the behavior of a free hairpin (Fig. 1). Starting simulations with a fully closed structure, i.e. with all possible base pairs formed, we observe base pairs to open, and after some time

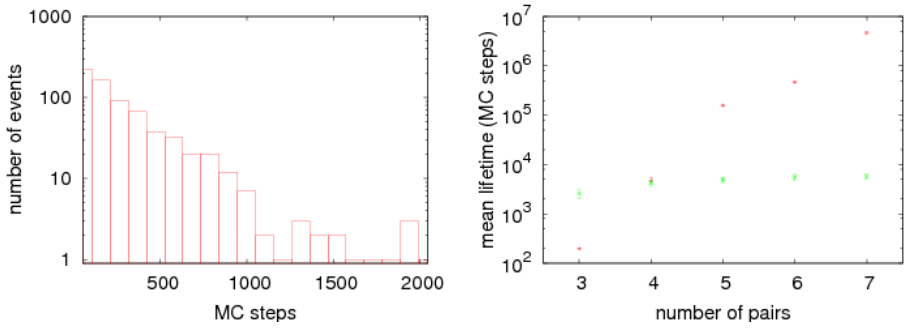

Figure 1. Lifetime (unfolding time) of a hairpin: left: Distribution of the lifetime of a hairpin with a 5 bp stem- The lifetime is defined as the time it takes to get from the state where all possible base pairs are closed to the state where all are open. right: Mean lifetime plotted as a function of the number of base pairs in the stem (red) and corresponding closing (folding) time (green).

all contacts are disconnected for the first time. We call this the lifetime or unfolding time of the hairpin. The distribution of unfolding times is exponential (Fig. 1(left)), a hallmark of two-state folding, as also indicated by experiments[1]. Likewise the distribution of folding times is also exponential. We then varied the length of the stem, i.e. the number of possible base pairs in the stem. The mean lifetime of the hairpin depends exponentially on the stem length, while the folding time, the time it takes to close a hairpin from a single stranded chain does not depend strongly on the stem length (Fig. 1(right)). This result is plausible since the limiting step of folding is the formation of the first bond, which is unfavorable due to the loss of entropy from the loop formation, while the other base pairs are closed very quickly once the first bond is formed.

Next, we use our model to simulate a hairpin under pulling forces. We introduce an additional energy term which goes with $F_{ext} * \Delta x$, where F_{ext} is a constant external force and Δx the relevant change in chain length arising from base pairing. We determine the equilibrium distribution between the folded and unfolded state as a function of the applied force. For a hairpin of length 5, we observe a sharp transition from mostly closed to mostly open at about 9 pN (Fig. 2), reminiscent of experimental observations for more complex hairpins[6].

4 Concluding Remarks

We have studied the folding and unfolding dynamics of simple RNA molecules with Monte Carlo simulations of a structure based model. With our model we are able to show the expected dynamic behavior of an RNA hairpin. As may be expected we find folding and unfolding times that are exponentially distributed. The folding of such a structure is mainly limited by the formation of the first base pair, while the dissolution strongly depends on the length of the stem which gives the stability of the folded state. Our model also allows us to introduce external forces on the RNA molecule. We see a typical force extension behavior where at a narrow force range the RNA changes from a folded to an unfolded formation. This model can also be extended towards a more detailed and more realistic, empirical

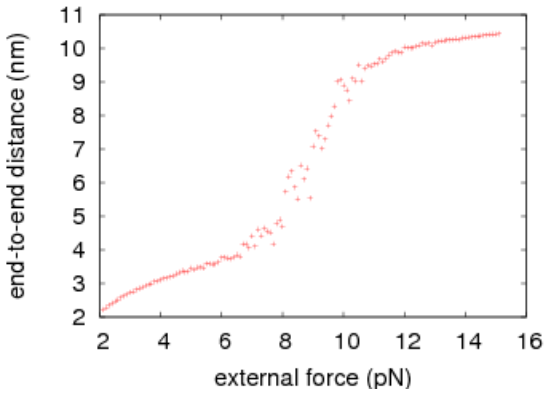

Figure 2. Mean end-to-end distance of an RNA molecule as a function of the external pulling force. Our model hairpin of length 5 bp unzips.

energy parametrization, similar to what is used in secondary structure prediction. With that parametrization, which is, of course, sequence dependent, quantitative agreement with the experimental data is obtained[7].

References

1. I Tinoco and C Bustamante, *How RNA folds.*, J Mol Biol, **293**, no. 2, 271–81, Oct. 1999.
2. M Zuker, *Mfold web server for nucleic acid folding and hybridization prediction.*, Nucleic Acids Res, **31**, no. 13, 3406–15, July 2003.
3. T Henkin and C Yanofsky, *Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions.*, Bioessays, **24**, no. 8, 700–7, Aug. 2002.
4. J N Onuchic, Z Luthey-Schulten, and P G Wolynes, *Theory of protein folding: the energy landscape perspective.*, Annu Rev Phys Chem, **48**, 545–600, 1997.
5. D Pincus, S Cho, C Hyeon, and D Thirumalai, *Minimal models for proteins and RNA from folding to function.*, Prog Mol Biol Transl Sci, **84**, 203–50, 2008.
6. J Liphardt, B Onoa, S B Smith, I Tinoco, and C Bustamante, *Reversible unfolding of single RNA molecules by mechanical force.*, Science, **292**, no. 5517, 733–7, Apr. 2001.
7. M Faber and S Klumpp, *Manuscript in preparation*, nn, 2011.