Single-Molecule Study of RNA Hairpin and Pseudoknot Folding

by

Huizhong Xu

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

at

Department of Physics,

Carnegie Mellon University,

Pittsburgh, Pennsylvania

Committee in charge:

Dr. Maumita Mandal (advisor)

Dr. Stephen Garoff

Dr. Markus Deserno

Dr. Daniel Zuckerman

January, 2015

Abstract

Understanding the folding mechanism of RNAs is critical for studying their structure-function relationship. The folding energy landscape of RNA is often rugged, involving multiple folding routes and intermediates. Optical tweezers is a powerful tool to study the folding of RNA, as it enables direct investigation of single RNA molecules and their folding pathways. In this dissertation, I used optical tweezers to investigate the folding of two RNA structural motifs, a hairpin and a pseudoknot.

RNA hairpin, the A-form double helix capped with a terminal loop, is a basic RNA secondary structure. I examined a model RNA hairpin, P5ab, derived from the group I intron ribozyme. P5ab was only considered as a two-state folder in previous single-molecule experiments. Here, I have identified multiple intermediate states during the folding of P5ab. These intermediate states are located near the bulge, internal loop, and terminal loop-closing regions. The folding free energy of P5ab was computed from both equilibrium unfolding/folding kinetics and non-equilibrium work done using fluctuation theorems. The features of the energy barrier were explored by model-dependent fit of the unfolding force and kinetics, as well as by equilibrium sampling of the hopping transitions near the equilibrium force.

Pseudoknot is an RNA tertiary structure that contains base pairing between the hairpin loop and an unpaired region in the RNA. Here, I examined the class I preQ1 riboswitch aptamer, which folds into a pseudoknot upon binding to the ligand preQ1. Using mechanical unfolding and mutational studies, I elucidated the folding pathways of the aptamer. In the absence of ligand, only a hairpin state is observed, while in the presence of preQ1, the pseudoknot is formed through the hairpin intermediate, and shows a ligand concentration-dependent folding kinetics. This study also serves as the basis for future investigation of the riboswitch control mechanism involving both the aptamer and the expression platform.

Acknowledgements

First, I would like to thank my advisor Dr. Maumita Mandal. You have been a great mentor for me. Thank you for encouraging my research and for allowing me to grow as a scientist.

I would also like to thank my committee members, Dr. Stephen Garoff, Dr. Markus Deserno, and Dr. Daniel Zuckerman, for serving as my committee members. I also want to thank you for your valuable comments and suggestions during my annual reviews and thesis defense.

I am also grateful to many other lab members. I want to thank Muwen Kong and Guangtao Song for helpful discussions when I was learning molecular biology skills for making the single-molecule constructs. I also want to thank Sriram Kumar, Maverick Chen, and Jessica Zhu for assistance in data analysis in P5ab project. Besides, I want to thank Soumya Samanta and Corina Ramirez for helping me make the preQ1 aptamer and m1 mutant constructs.

I would like to thank the Dean Fred Gilman for financially supporting me for two years, which made me focus on research without financial worries.

Last but not least, a special thanks to my wife, Jing Hua, and two kids, Matthew and Astro. You brought great happiness to me and became my motivation to success.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Ribonucleic acid (RNA) plays diverse roles in the cell. Apart from being the carrier of genetic information (1, 2), RNA has been found to perform catalysis including splicing (3), processing of transfer RNA (4), and protein synthesis (5). Moreover, RNA regulates gene expression in riboswitches (6-9), microRNAs (10, 11), and small interfering RNAs (12-14). These biological functions are highly dependent on the structure and folding of the RNA. Misfolded RNAs often inhibit their functions and lead to diseases (15, 16). Thus, investigating the folding of RNAs is critical for understanding their functions.

1.1 RNA base pairs

Folded RNA molecules are stabilized by stacking and hydrogen bonding between bases. In RNA, the four types of bases, adenine (A), uracil (U), guanine (G) and cytosine (C) can form canonical Watson-Crick base pairs as A-U and G-C, which contain two and three hydrogen bonds respectively. Non-canonical base pair can also occur. Figure 1.1B shows some examples of non-canonical base pairs. The G·U wobble pair frequently occurs in ribosomal RNAs (17, 18) and ribozymes (19-21). The sheared G·A base pair is formed in hairpins closed by GNRA tetraloops (N is any base; R is A or G) (22-25). The G·U reverse wobble pair is formed between the first and the last bases in UUCG tetraloops (26-28). The protonated A^+·C base pair is often observed in hairpin stems and can contribute to RNA catalysis (29, 30). The G·G Hoogsteen base pair is found in G-quadruplex structures, where two or more layers of G-tetrad (four guanine bonded planar structure) stack on top of each other. The G-quadruplex structures are found in human telomeric RNA (31, 32) and 5'-untranslated regions of mRNAs (33).

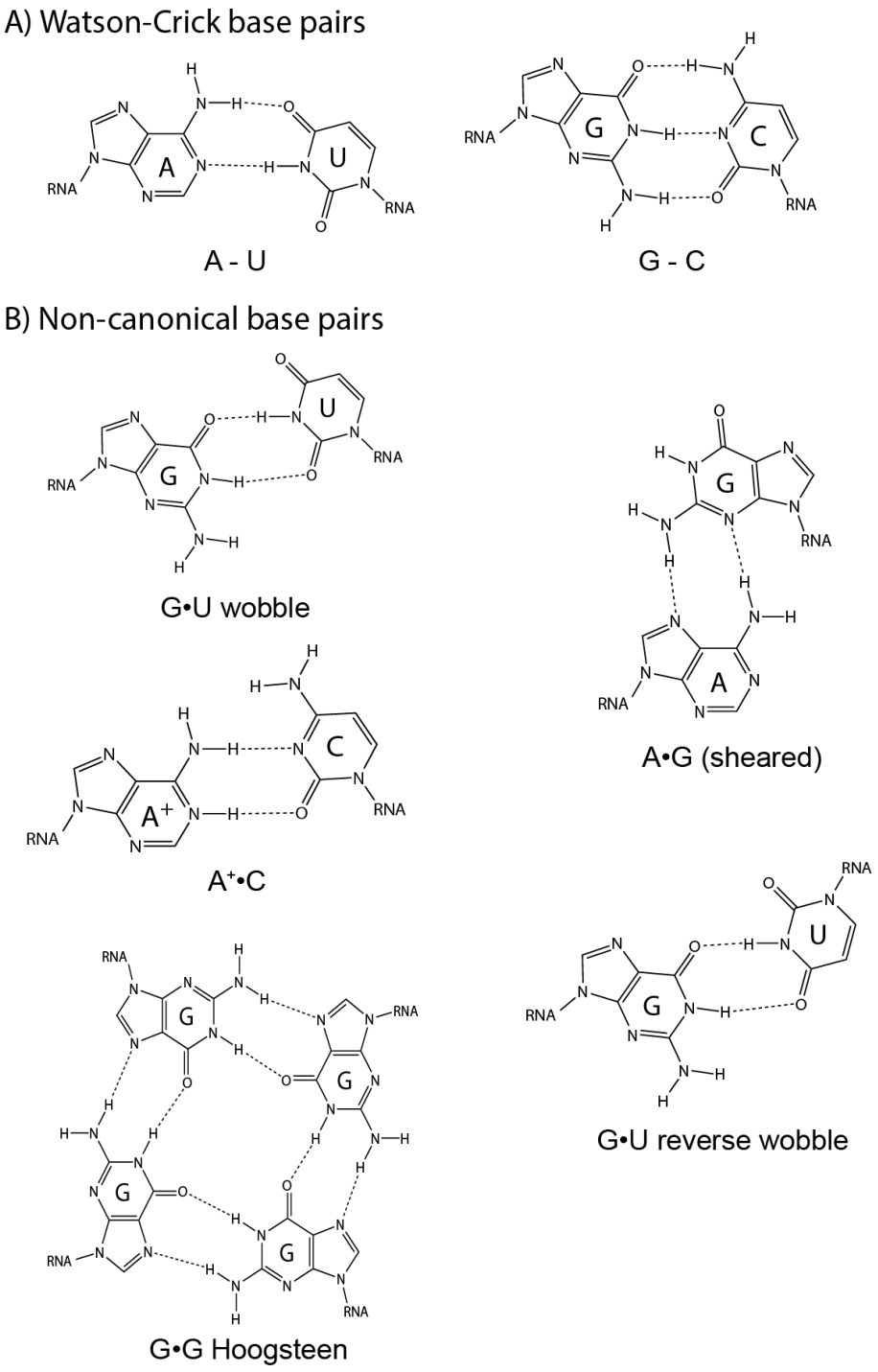


Figure 1.1 Canonical Watson-Crick and non-canonical base pairs between RNA bases.

(A) Canonical Watson-Crick base pairs.

(B) Examples of non-canonical base pairs.

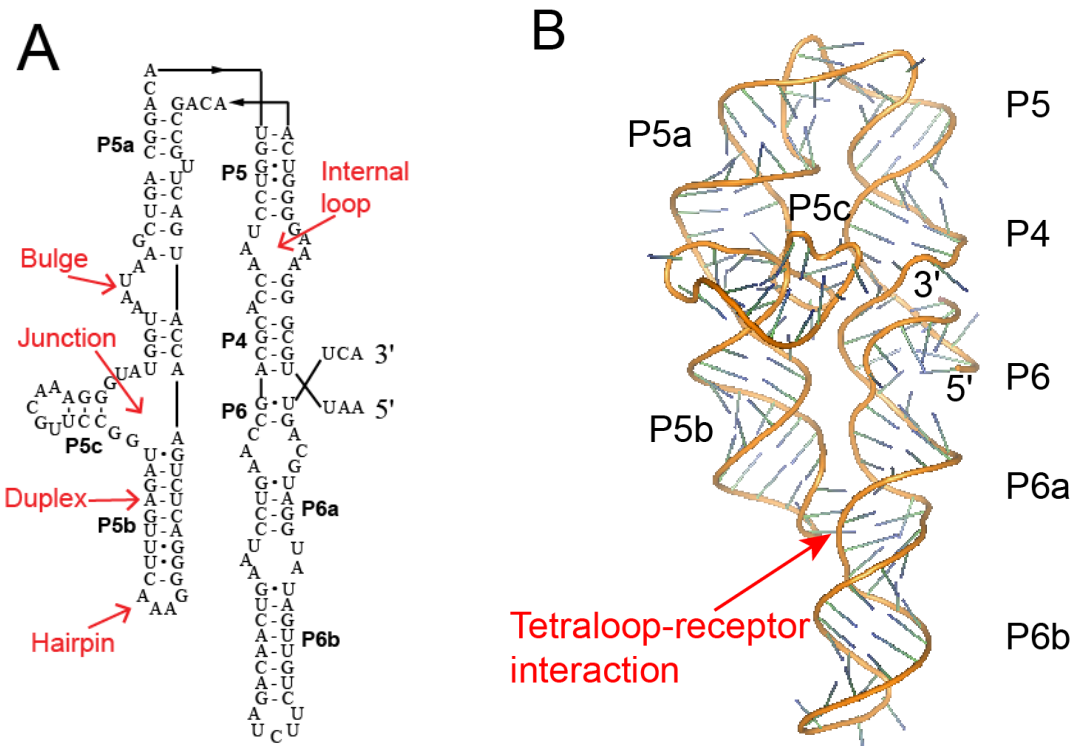


Figure 1.2 Secondary and tertiary structures of the P4-P6 domain in *Tetrahymena thermophila* group I ribozyme.

(A) The secondary structure of the P4-P6 domain in *T. thermophila* group I ribozyme (34). Within this structure, I have indicated different types of secondary structural motifs in red.

(B) Crystal structure of the same RNA (35), where the tetraloop-receptor tertiary interaction is indicated. The figure is generated by PyMol software (PDB id: 1GID).

1.2 RNA secondary structure and tertiary structure

The secondary structure of RNA refers to the base-pairing interactions in the molecule. There are five types of RNA secondary structures (36):

- Duplex: an A-form double helix formed by Watson-Crick base pairs.
- Hairpin: consists of double helix stem that ends with an unpaired loop
- Bulge: unpaired nucleotides occur on one strand of duplex
- Internal loop: unpaired nucleotides occur on both strands of duplex
- Junction: loops where three or more helices intersect

Examples of the above structures are indicated in Figure 1.2A, which shows the secondary structure of the P4-P6 domain in *Tetrahymena thermophila* group I intron RNA.

The tertiary structure of RNA is its unique three-dimensional structure, with different secondary structural elements interacting with each other. Some of the tertiary structural motifs include (37):

- Pseudoknot: long-range Watson-Crick base pairing between the loop nucleotides of a hairpin and an unpaired region of RNA. The pseudoknot structure performs important functions including translational repression (38-40), ribosomal frameshifting (41-44) and activity of telomerase (45).
- Kissing loop: long-range base pairings between two hairpin loops. The kissing loop motif is found in purine sensing riboswitches (46, 47). It also plays an important role in molecular recognition, as in dimerization of the HIV genomic RNA for packaging (48-51).
- A-minor motif: interactions of unpaired adenines with the minor grove of a distal helix. It is often found in clusters of consecutive adenines. For example, in the P4-P6 domain of group I intron, the A-minor motif is responsible for the interaction between the GAAA tetraloop and P4-P6 helix (35) (Figure 1.2B).

1.3 The RNA folding problem

It has been established that RNA folding is largely hierarchical, with secondary structures forming first, followed by tertiary structures (52). The RNA folding problem discusses how RNAs fold into their three-dimensional structures (52, 53). The following aspects are usually dealt with while discussing RNA folding.

- **The thermodynamics of folded RNA structure**

The three-dimensional structure of any RNA is determined by X-ray crystallography (54, 55), nuclear magnetic resonance (NMR) (56) and small-angle X-ray scattering (SAXS) (57) with spatial resolutions from a few Å to tens of nanometers. Similarly, the base pairing in the RNA is probed by various techniques including chemical probing (58-60), in-line probing (61), and nucleotide analog interference mapping (NAIM) (62). These studies together suggest the RNA conformation, as in the ribozyme and ribosomal RNA (58-60). The thermodynamics of RNA folding is mainly studied by thermal melting with ultraviolet spectroscopy (63), and by calorimetry (64), which is also used to study RNA-ligand binding energetics (65). The free energy and enthalpy parameters for secondary structural elements have been experimentally determined at a standard salt condition 1M NaCl, as explicitly used in Mfold prediction for RNA secondary structures (66-68).

- **The folding kinetics and pathways**

Traditionally, the RNA molecule to be studied is denatured or renatured through change of temperature or solvent. Recent development in single-molecule techniques allows RNA to be unfolded by force, which will be discussed in Section 1.4. In bulk ensemble experiments, the RNA structural dynamics are monitored by various techniques, as shown in Table 1.1. The folding kinetics of RNA has been studied at both the secondary and tertiary levels. The folding of secondary structural elements typically occurs in microseconds, as determined by temperature jump (T-jump) experiments with multiple probes (UV, Infrared, and fluorescence) (69-78). Recently, a series of T-jump experiments (72, 74, 75, 77, 78) show that RNA hairpins with only 2-4 base pairs in the stem involve multiple intermediate states along the folding pathway, suggesting a rugged folding energy landscape (for details, see Section 3.1). At the tertiary level,

however, folding may take from milliseconds to as long as several hours, as extensively investigated in the group I intron ribozyme (79-88) and other large RNAs (89, 90). Moreover, tertiary folding has been found to depend on Mg^{2+} concentration. As shown in Figure 1.3 for the group I intron ribozyme, at low ionic strength, the RNA is in extended conformations (U), which contain only secondary structure helices. Low Mg^{2+} concentration (0.02 mM) stabilized part of the tertiary contacts (mostly P3/P7 pseudoknot and assembly of core helices) and created an ensemble of intermediates (I_C). Higher Mg^{2+} concentration (1.7 mM) allowed formation of the native tertiary structure (N), which occurred in less than 50 ms (87).

Table 1.1 Experimental methods to study RNA folding kinetics

Denaturation method	**Methods to monitor folding**
Temperature	UV (69-71, 73, 74), Infrared (75), fluorescence (72, 76-78)
Solvent	SAXS (91, 92), NMR (93, 94) , footprinting (82, 95), FRET (96, 97)
Force	atomic force microscope (98-100), magnetic tweezers (99, 101, 102), optical tweezers (103-105)

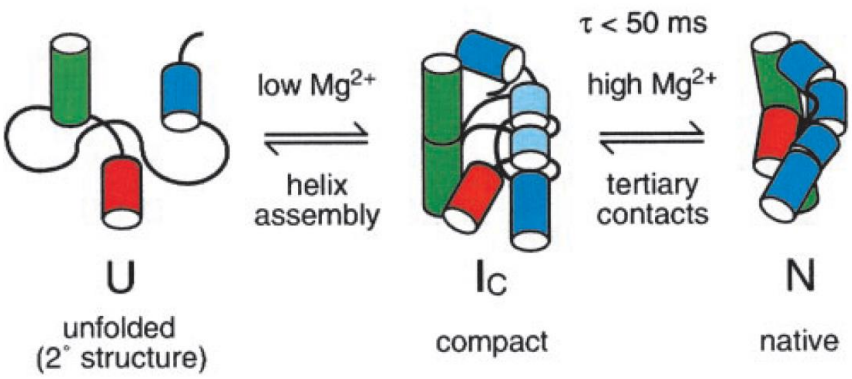


Figure 1.3 Mg^{2+} dependent folding of group I intron ribozyme from *Azoarcus.*

Time-resolved hydroxyl radical footprinting experiment on the *Azoarcus* group I ribozyme revealed that the extended conformation (U) folds into compact but non-native intermediates (I_C) at low Mg^{2+} concentration, and the native conformation (N) in high Mg^{2+} concentration. This figure is adapted from reference (87).

1.4 Single-molecule studies on RNA folding

In bulk ensemble experiments, large numbers of molecules are measured and the observed signal represent the average of the entire population. Consider, for example, a thermal melting experiment on 1 mL volume of 10 μM RNA, which contains 10^{15} RNA molecules. A UV spectroscopic measurement of the solution will depend on the properties of all the 10^{15} molecules. Adding to that problem is that every molecule might take different paths to unfold.

Single-molecule techniques, such as Fluorescence Resonance Energy Transfer (FRET), atomic force microscopy (AFM), magnetic tweezers and optical tweezers, are capable of providing detailed structural and kinetic information going beyond the ensemble averaged approaches. The latter three techniques allow one to investigate the effect of an external force upon structure and function. The RNA unfolding studies by mechanical loads are relevant in cellular context where molecular motors unwind RNA structures by localized forces. These single-molecule techniques are described below with examples of the studied RNA systems.

- Single-molecule FRET

Single-molecule FRET (106, 107) is used to study inter- and intra- molecular motions separated by a few nanometers. In this approach, donor and acceptor fluorophores are attached to the molecules of interest. The energy transfer efficiency, E, between donor and acceptor is related to their separation, R, as $E = 1/[1+(R/R_0)^6]$, where R_0 is the Förster distance, typically 2–8 nm, at which E = 0.5. Figure 1.4 shows a single-molecule FRET experiment on the loop-loop interactions in the adenine riboswitch aptamer domain (96), where the loop L3 was labeled with a Cy3 donor and loop L2 with a Cy5 acceptor. In the absence of magnesium, the molecules primarily exhibit a low FRET state (E = 0.25), indicating that the loops are not interacting. At magnesium concentrations of 0.05-0.5 mM, the RNA interconverts between low and high FRET states with the latter state (E = 0.9) assigned as the L2–L3 kissing interaction. Higher concentrations of magnesium cause the aptamer to become mostly locked in the high FRET state. These data demonstrates the dynamic nature of the aptamer tertiary architecture as well as the effect of magnesium on the formation of the kissing loop structure.

- Atomic force microscopy

In a typical AFM pulling experiment (98, 99), a single molecule is attached between the surface of a glass slide and the tip of a flexible cantilever (Figure 1.5A). Both the force (detected by cantilever deflection) and the tip-surface separation are recorded, and plotted as force-separation curves. Figure 1.5 shows an AFM experiment on the effect of RNA helicase on the unfolding of RNA hairpins (100). In the absence of any helicase, the RNA hairpin unfolds at 180 pN. In the presence of the helicase Ded1, the unfolding force is significantly reduced to below 100 pN. Thus, Ded1 acts as an unwinding factor on RNA substrates.

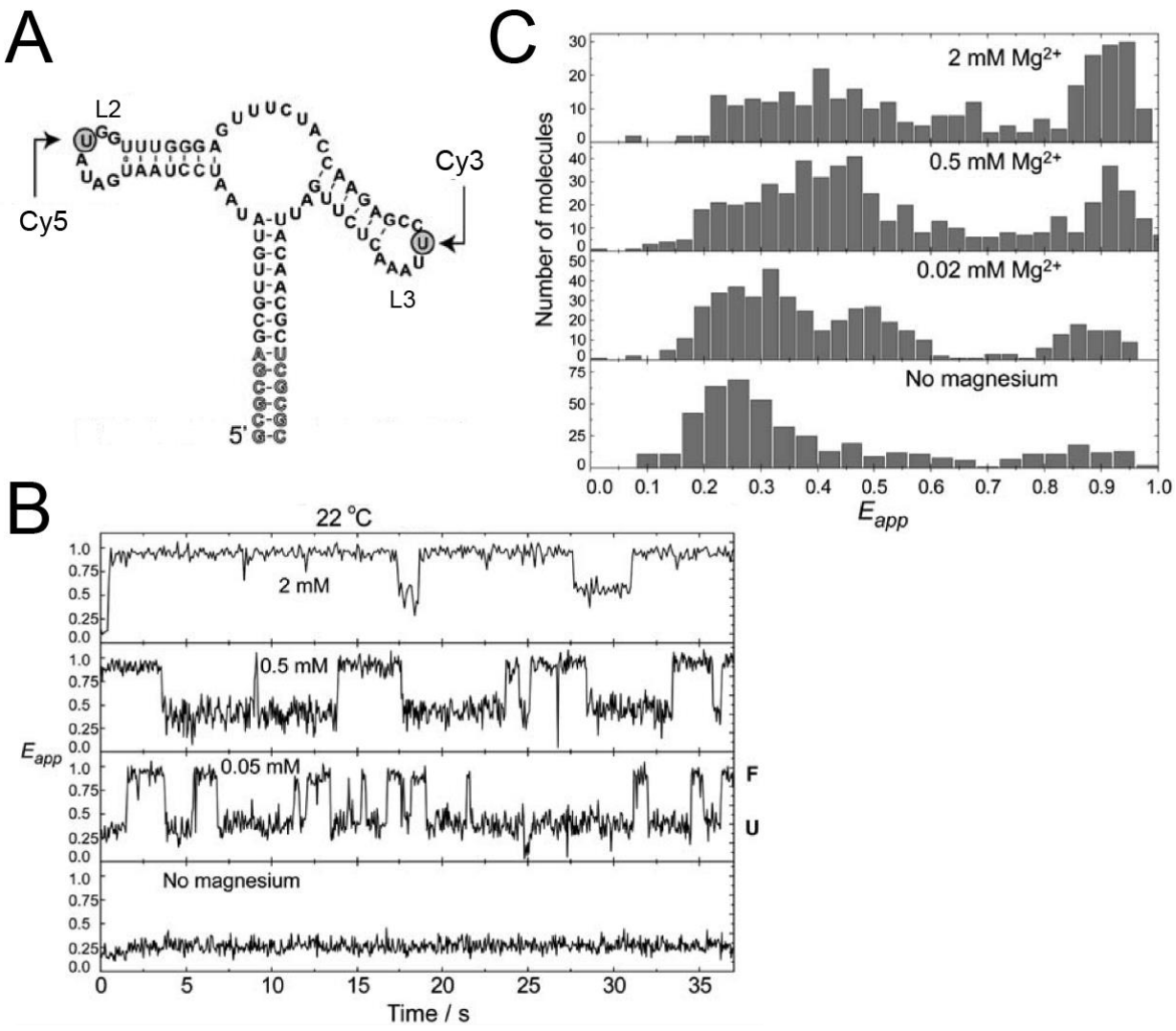


Figure 1.4 Loop-loop interaction in adenine riboswitch studied by single-molecule FRET.

(A) Secondary structure of the adenine aptamer with donor (Cy3) and acceptor (Cy5) fluorophores attached in the loops at the shaded positions by conjugation to 5'-aminoallyl uridine nucleotides.

(B) Time records of FRET efficiency (E_{app}) for single aptamer molecules at various magnesium ion concentrations. The unfolded state, U, shows $E_{app} = 0.25$, and the folded state, F, shows $E_{app} = 0.9$.

(C) Histograms of FRET efficiency summed over multiple single molecules at different magnesium ion concentrations.

This figure is adapted from reference (96).

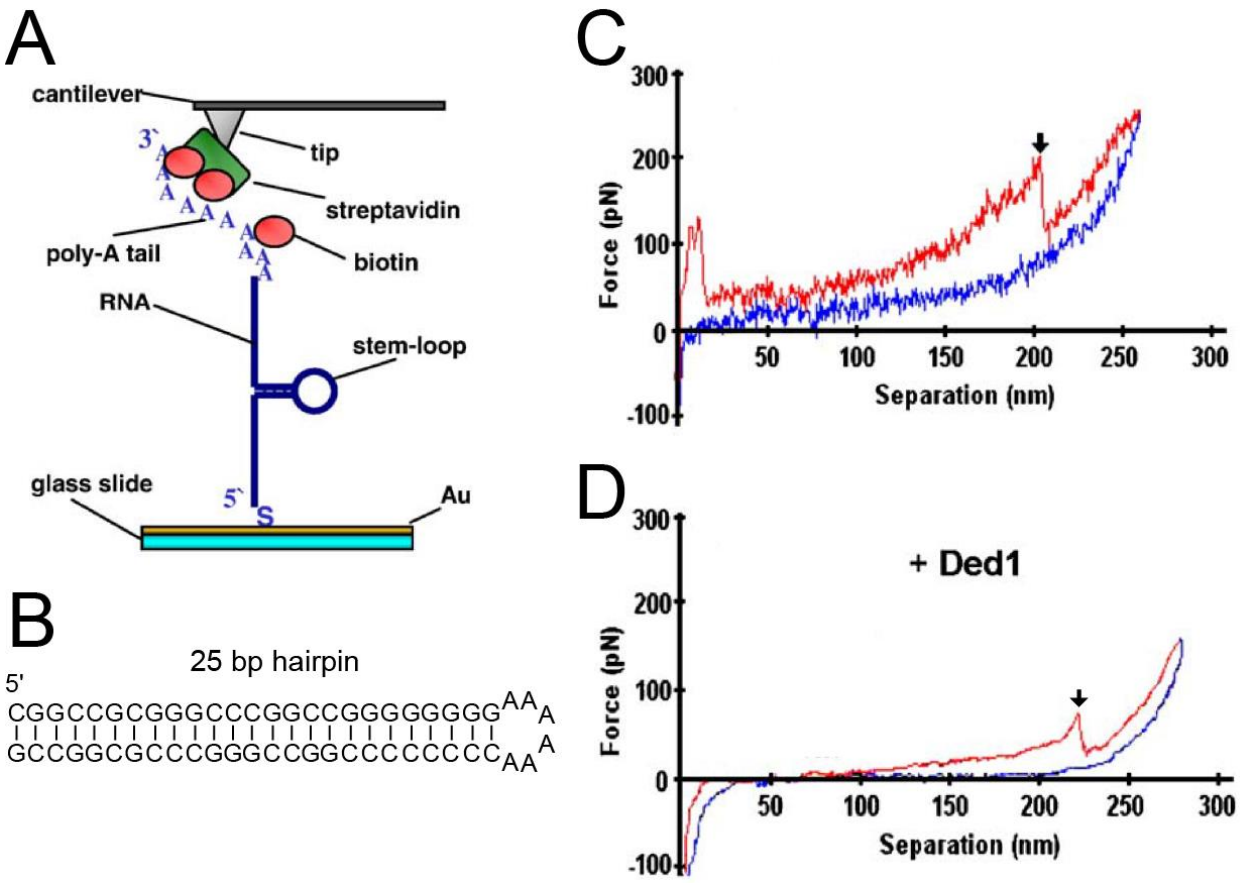


Figure 1.5 AFM study of single RNA hairpin unwinding by RNA helicase.

(A) Immobilization of the RNA molecule between the AFM tip and a gold-coated microscope slide.

(B) The GCN4 RNA hairpin, which contains 25 GC base pairs in the stem.

(C) Example of an AFM force-separation curve representing stretch (red) and relax (blue) of the GCN4 RNA hairpin. The rupture event is indicated by an arrow.

(D) Example of an AFM force-separation curve in the presence of 0.4 μM of the helicase Ded1. The rupture event is indicated by an arrow.

This figure is adapted from reference (100).

- Magnetic tweezers

In magnetic tweezers (99, 101), the molecule of interest is tethered between the coverslip and a magnetic bead that is trapped in a magnetic field. The magnetic field is generated by permanent magnets capable of applying force in excess of 200 pN. Figure 1.6 shows a magnetic tweezers experiment on the force-extension response of a 4.2 kbp dsRNA (102). Two types of dsRNA were studied, one torsionally relaxed (with nicks in RNA) and one torsionally constrained (without nicks). The stretching behavior of torsionally relaxed dsRNA in the presence of 100 mM monovalent salt is well fit by the extensible worm-like chain (WLC) model, with contour length L_C = 1.15 ± 0.02 μm, persistence length P = 57 ± 2 nm, and stretch modulus K = 350 ± 100 pN. The overstretching transition occurs at around 50 pN. In contrast, the torsionally constrained dsRNA shows no overstretching transition up to 75 pN.

- Optical tweezers

In the optical tweezers assay (103, 104), used in this dissertation, the RNA molecule is tethered between two micron-sized beads, one held on a micro-pipette and the other in the optical trap (Figure 1.7A). The force is measured by change in light momentum and molecular extension is determined from bead positions. Figure 1.7 shows an optical tweezers experiment on the folding of 52-nucleotide TAR RNA hairpin (105). The unfolding and refolding are almost reversible at small loading rate, but shows hysteresis at higher loading rate (Figure 1.7C). In the constant force experiment at 12.4 pN (Figure 1.7B), the molecule shows a hopping distance of 18 nm, indicating opening and closing of the hairpin. The kinetics of unfolding and folding are then determined from the folded and unfolded lifetimes.

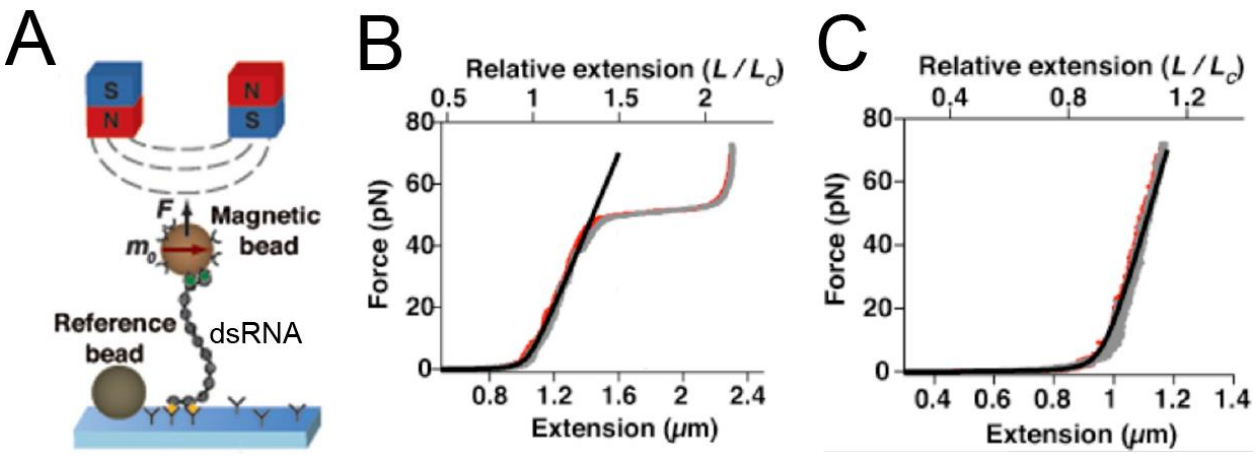


Figure 1.6 Single-molecule magnetic tweezers experiment on the stretching of double stranded RNA.

(A) A magnetic tweezers experimental setup on dsRNA. A streptavidin-coated magnetic bead is tethered to an anti-digoxigenin-coated surface by a dsRNA molecule. A surface-attached reference bead is tracked for drift correction. Permanent magnets (N: north pole, S: south pole) above the flow cell are used to exert a stretching force.

(B) Force-extension curve for stretching (red) and relaxing (grey) a torsionally relaxed dsRNA. The data below 40 pN is well fit by the extensible WLC model (black solid line).

(C) Force-extension curve for stretching (red) and relaxing (grey) a torsionally constrained dsRNA. The data is well fit by the extensible WLC model (black solid line). No overstretching transition was observed up to 75 pN.

This figure is adapted from reference (102).

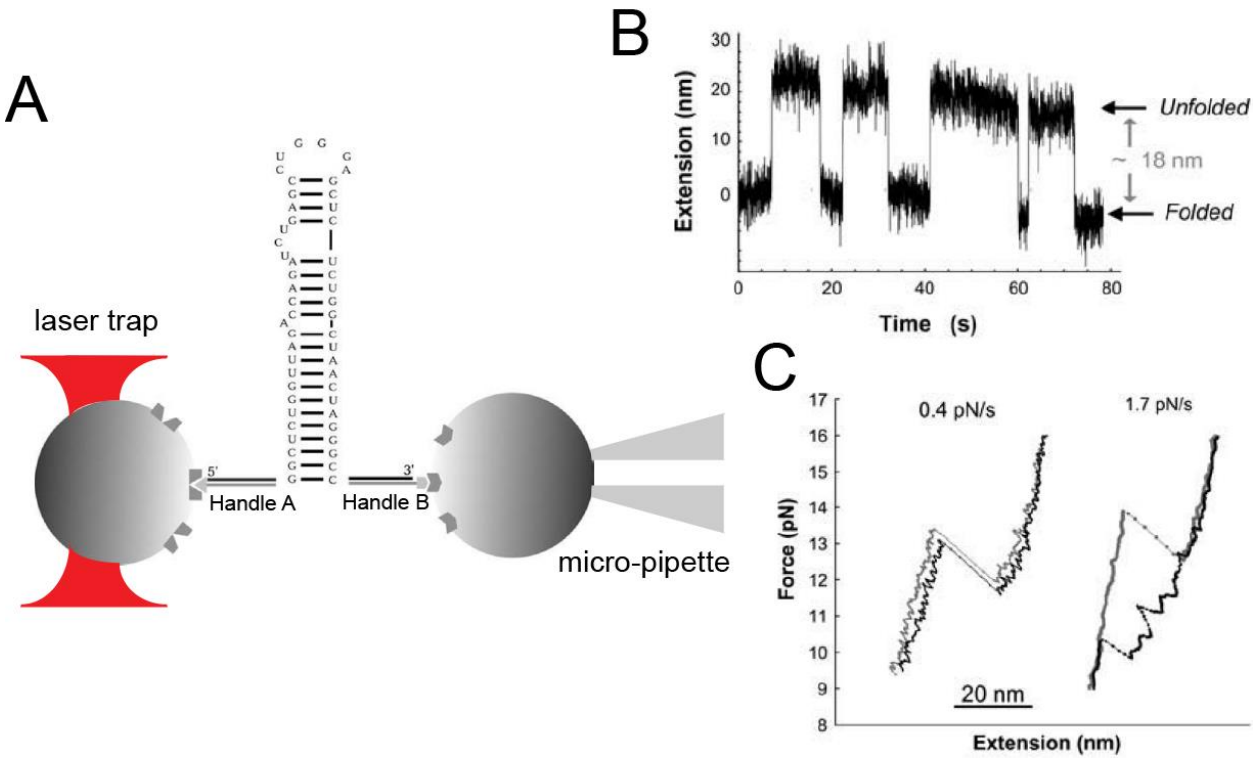


Figure 1.7 Single-molecule optical tweezers study of TAR RNA hairpin unfolding.

(A) The 52 nt TAR RNA hairpin is flanked by two 500 base pair DNA/RNA handles shown with the RNA in black and the DNA in gray. The 3' end of the DNA handle A and the 5' end of the DNA handle B are labeled with biotin and digoxigenin respectively. The entire molecule is attached to two microspheres coated with streptavidin and anti-digoxigenin respectively. One bead is held in the laser trap, and the other is fixed on a micro-pipette.

(B) An extension versus time trace of TAR RNA hairpin at 12.4 pN. The two states show an extension difference of ~18 nm.

(C) Typical force-extension curves of TAR RNA hairpin collected at loading rates of 0.4 pN/s and 1.7 pN/s. Unfolding trajectories are shown in gray and refolding in black.

This figure is adapted from reference (105).

1.5 Outline of the thesis

In this thesis, I used single-molecule optical tweezers to investigate the folding of two RNA structural motifs, a hairpin and a pseudoknot, present in two different RNA molecules. In Chapter 2, the basic principle of optical trapping and the experimental setup are described. In Chapter 3, I investigate the folding of P5ab hairpin, where I identify multiple folding intermediates, which suggest a more complex folding pathway than initially thought. The folding free energy and energy barrier information are extracted from the force spectroscopy data. Chapter 4 describes my single-molecule study of pseudoknot folding in preQ1 riboswitch aptamer, where the folding pathway and ligand binding kinetics are elucidated.

Chapter 2 Experimental Setup

2.1 Optical trapping

Optical trapping of microscopic objects was first shown by Arthur Ashkin in 1970 (108). Since then, optical trapping has found many applications in physics, chemistry and biology (109-111). The mechanism of optical trapping of a dielectric microsphere involves two optical forces: the scattering force, F_{scat} and the gradient force, F_{grad}, as show in Figure 2.1 (112). The scattering force on the microsphere acts in the direction of the light beam. The gradient force, F_{grad}, is pointed in the direction of highest light intensity. In Figure 2.1A, the dielectric microsphere is situated off the center of a Gaussian laser beam. The optical paths of the two rays, a and b, are shown through the microsphere. Due to refraction at the microsphere interface, a force is exerted on the microsphere by the rays due to light momentum change. Since ray a represents a higher intensity than ray b, the corresponding force F_a is greater in magnitude than F_b. As a result, the vertical component of the total force, i.e. F_{grad}, acts as a restoring force, towards the direction of highest intensity. When the microsphere is placed in a focused laser beam (Figure 2.1B), the scattering force F_{scat} becomes a restoring force as well, and thus the particle can be stably trapped in the laser focus.

Using optical tweezers, forces of 0.1-100 pN and displacements from 1 nm to 100 μm can be accurately measured. Optical tweezers have been an excellent tool to study the forces and distances inherent in a broad range of biological processes, such as molecular motor (113, 114), biomolecule folding (115-119), and DNA packaging (120).

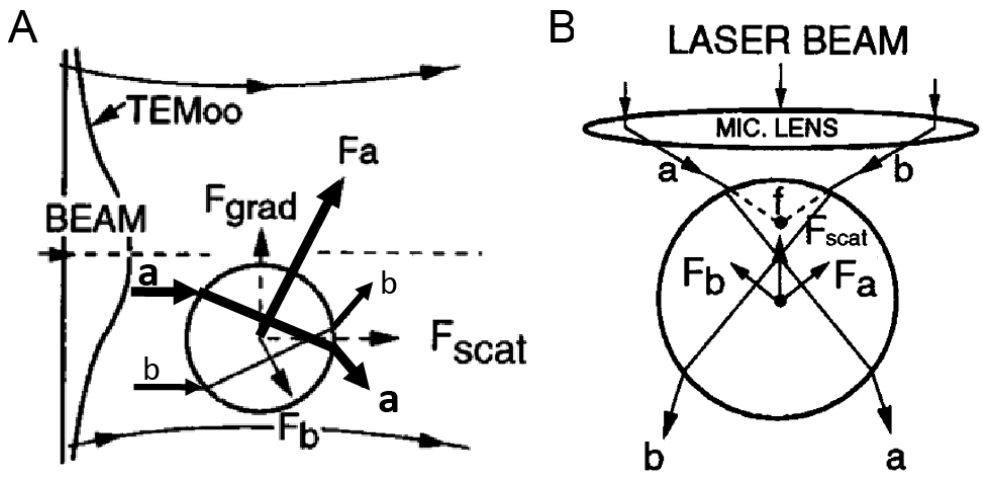


Figure 2.1 Ray optics illustration of the gradient force and scattering force for optical trapping.

(A) Ray optics illustration of the gradient force that is exerted on a dielectric microsphere by TEM00 mode (Gaussian) laser beam. (B) Origin of a restoring scattering force for a microsphere located below a microscope lens focus f. This figure is adapted from reference (112).

2.2 Optical tweezers setup

The optical tweezers instrument I used was built by Dr. Mandal in 2010. It consists of two counter-propagating laser beams (λ, 845nm), focused through objective lenses, forming a stable optical trap (103). The force on the bead is measured as the change in the light momentum from position-sensitive photo detectors (PSD). The position of the trap (x_{trap}) is monitored by a light lever in the instrument. In order to study the unfolding of RNA molecules in this thesis, the ends of single pieces of RNA were attached by two micron-sized beads. The titrated RNA molecules were first incubated with polystyrene beads (4.0-4.9 μm, Spherotech) coated with anti-digoxigenin (Roche). These beads attached with the molecules were then diluted in a buffer containing 10 mM Tris-HCl, pH 7.5 and 250 mM NaCl, flowed into a microfluidics chamber, and trapped by the lasers one at a time. Meanwhile streptavidin-coated polystyrene beads (1.5-1.9 μm, Spherotech) were also flowed in and held stationary via suction on a micropipette built in the chamber. Attachment of the biotin end of the molecule to the streptavidin bead was made by moving the trapped bead towards the stationary one (Figure 2.2).

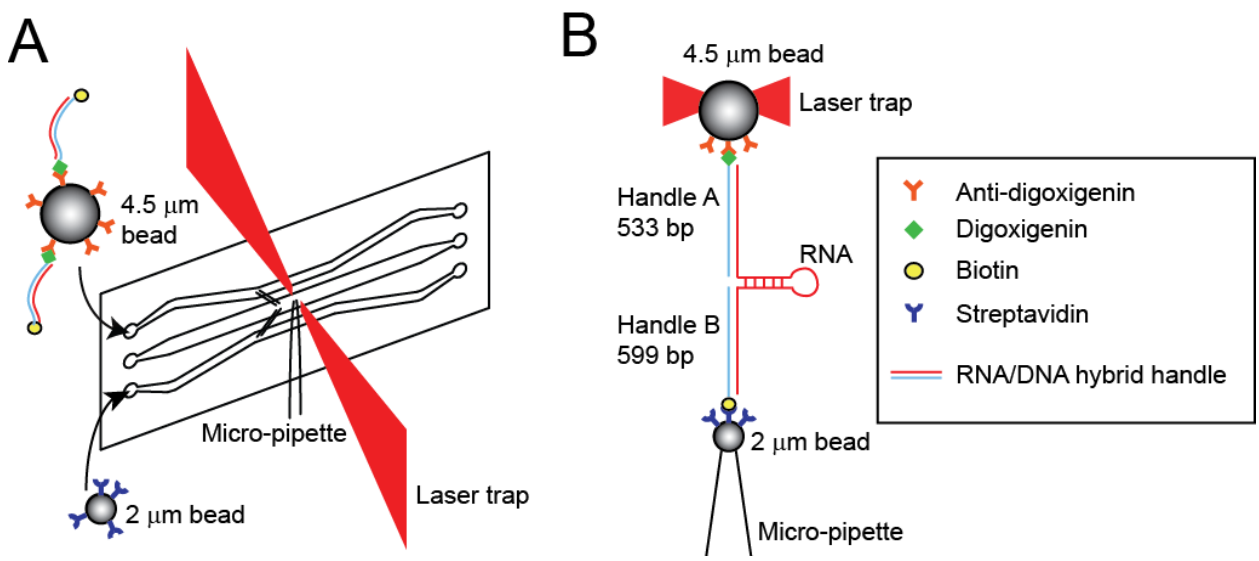


Figure 2.2 Experimental setup for single-molecule RNA pulling experiment.

(A) Fluidic chamber setup with two opposite laser beams propagating perpendicular to the plane of the chamber. The chamber has three channels, of which the top and bottom ones are used for dispensing beads into the central channel where the pulling study is performed. (B) A single molecule of RNA hairpin flanked by two ~550 bp RNA/DNA hybrid handles is held between two polystyrene beads, one bead (2 μm) fixed on a micropipette via suction and the other bead (4.5 μm) in an optical trap. The ends of Handle A and Handle B are labelled with digoxigenin and biotin respectively to facilitate attachments to the anti-digoxigenin and streptavidin coated beads respectively.

2.3 Calibration and accuracy

2.3.1 Force calibration

During my thesis, periodically I have calibrated the force experienced by the trapped bead by Stokes' law,

$$F = 6\pi\eta r v \qquad \text{(Eq. 2.1)}$$

in aqueous buffer, with known water viscosity $\eta = 1 \times 10^{-3}$ Pa · sec, bead radius r, and fluidic chamber velocity v. In order to ensure that the measured forces are similar to the published

literature, I measured the overstretching force of double stranded DNA molecules. The overstretching transition for a dsDNA is between 63-65 pN (118, 119), which is used as a standard for force accuracy in the instrument. In the lab, we used $\frac{1}{3}$ λ-phage DNA molecules in 10 mM Tris-HCl, pH 7.5 and 500 mM NaCl to measure the overstretching transition. The result is close to the reported value of ~65 pN (Figure 2.3).

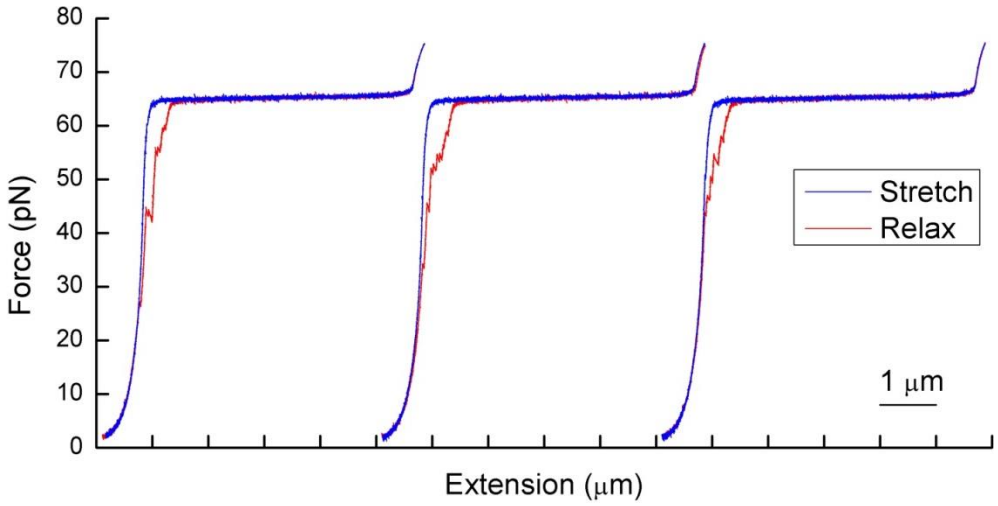


Figure 2.3 Pull-relax cycles of a 20 kB double stranded DNA molecule.

The 20 kB $\frac{1}{3}$ λ-phage DNA was pulled at a speed of 200 nm/s. The overstretching plateau was observed around force 65 pN.

2.3.2 Power spectrum analysis

Measuring the Brownian fluctuations of a bead in the optical trap provides a powerful tool to measure the stiffness of the optical trap. I performed power spectrum analysis (121, 122) on the trapped bead to calibrate the force fluctuations given by,

$$P(f) = \frac{2k_B T \gamma f_c^2}{\pi(f_c^2 + f^2)} \qquad \text{(Eq. 2.2)}$$

where the viscous drag coefficient, $\gamma = 6\pi\eta r = 3.8 \times 10^{-8}\ \mathrm{N \cdot s/m}$, as obtained from Stokes' law (Figure 2.4B) and the corner frequency, $f_c = \frac{\kappa}{2\pi\gamma}$, where κ is trap stiffness. By fitting Eq. 2.2 to the power spectrum obtained by Fourier transform of the force fluctuation data (Figure 2.4A), the corner frequency f_c and trap stiffness κ were calculated. The variations of the trap stiffness as a function of laser powers (PSD SUM) were fit well with a linear relation. Moreover, the trap stiffness also varies with different bead sizes (Figure 2.4C). In my single-molecule RNA folding studies, I used polystyrene beads with 4.5 μm in diameter and laser power around 150 mW (PSD SUM = 24000 for each laser), and the measured trap stiffness was 0.050 ± 0.002 pN/nm (mean ± standard deviation from 9 beads).

The position of the trapped bead (x_{bead}) was calculated as

$$x_{bead} = x_{trap} - \frac{F}{\kappa} \qquad \text{(Eq. 2.3)}$$

wherein *F* is the observed force.

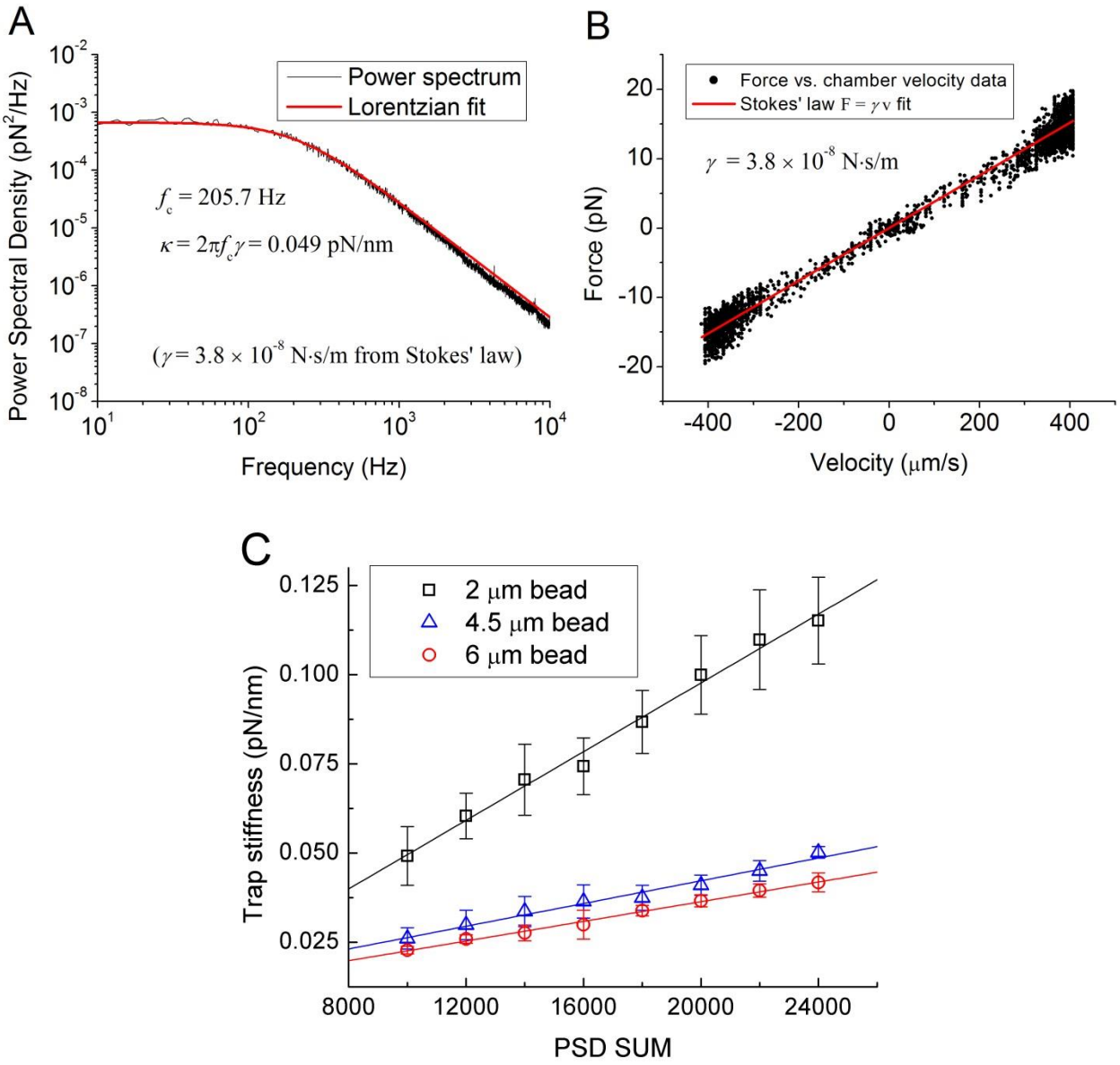


Figure 2.4 Measurement of trap stiffness from power spectrum analysis.

(A) Power spectrum analysis of a trapped bead with diameter ~ 4.5 μm (Spherotech Inc.). The data were recorded at an acquisition rate of 44 kHz. The power spectral density is fitted to Eq. 2.2 to obtain the corner frequency (f_c) and the trap stiffness (κ). (B) Stokes' law fit to the force vs. velocity data of a trapped bead, yielding the viscous drag coefficient γ. (C) Trap stiffness as a function of laser power (PSD SUM) for different bead sizes. Each dataset represents mean ± standard deviation measured from more than 3 beads.

2.3.3 Force feedback

Constant force mode is useful for measuring the kinetics of folding. When operating in constant-force-feedback mode, the feedback controls the position of the trap and hence the force on the bead and molecule. The feedback operation frequency is 4 kHz. An average force could be maintained at the preset value with a standard deviation of 0.1-0.15 pN (Figure 2.5).

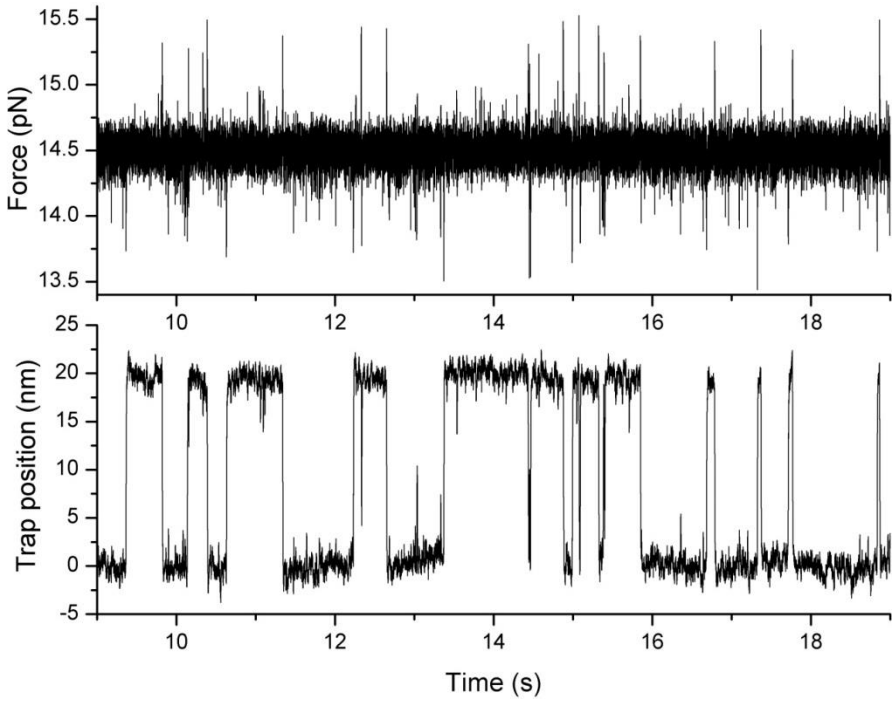


Figure 2.5 Force and trap position versus time for a constant-force-feedback experiment on a P5ab molecule.

Data was collected at 4 kHz. The force was maintained at the preset value of 14.5 pN, with a standard deviation of 0.13 pN. The brief spikes in the force trace represent the feedback response to the force changes caused by molecular transitions.

2.4 Types of experiments

The single-molecule RNA folding experiments were performed at room temperature (25 ± 0.5 °C). Once a tether is established between the beads, it is subjected to three different force modes: force ramp, constant force (CF), and passive mode (104, 123-125). All experimental data were recorded at 400 Hz – 4 kHz.

- Force ramp

In force ramp experiments, the tether between the beads was continuously stretched and relaxed between 1 pN and 30 pN by moving the trap at a constant speed (typically 200 nm/sec, which corresponds to a loading rate around 7.6 pN/sec). Both the force (F) and the trap position (x_{trap}) were monitored on a graphical user interface (GUI). By fitting the F vs x_{trap} pulling curves to the modified worm-like chain model (Eq. 2.5), I was able to determine the single molecule connectivity between a bead pair. The molecular extension x was calculated from Eq. 2.3 and plotted against force F as the force-extension curve (FEC), where structural transitions of the molecule are observed as sudden changes in force and extension.

- Constant force

In the constant force mode, the molecule was held at a pre-set force wherein the molecule hopped between different states.

- Passive mode

In the passive mode experiment, the trap was held stationary near the transition force without an active force feedback. Thus, as the molecule undergoes a folding transition, the force increases accordingly. State identification and kinetics can be obtained from both the constant force and passive mode settings.

2.5 Preparation of RNA constructs for single-molecule studies

The RNA constructs for single-molecule studies were generated as described earlier (115). The protocol contains the following steps, as schematically shown in Figure 2.6.

① The DNA coding for the RNA sequence of interest was synthesized as an EcoR1-HindIII fragment and digested by EcoR1 and HindIII restriction enzymes.

② The digested product was inserted into pBR322 plasmid vector.

③ The ligated vector was transformed into E. coli DH5α competent cells. The integrity of the cloned DNA was confirmed by sequencing (Genewiz Inc.).

④ An 1100 bp DNA fragment encompassing the insert and the flanking sequences (533 bp and 599 bp) was PCR amplified from the cloned vector using the following primers, where the T7 RNA polymerase promoter sequence is underlined.

5'-<u>TAATACGACTCACTATAGGG</u>ACTGGTGAGTACTCAACCAAGTCATTCTG

and 5'-TAGGAAGCAGCCCAGTAGTAGGTTGAGGCC

⑤ The DNA product was purified and *in vitro* transcribed into the required 1.1 kB-long transcript using T7 Ribomax kit (Promega).

⑥ The RNA was annealed with Handle A (533 bp, Digoxigenin labelled) and Handle B (599 bp, Biotin labelled). The final product is the RNA/DNA hybrid construct used for single-molecule experiments, as shown in the experimental setup in Figure 2.2B.

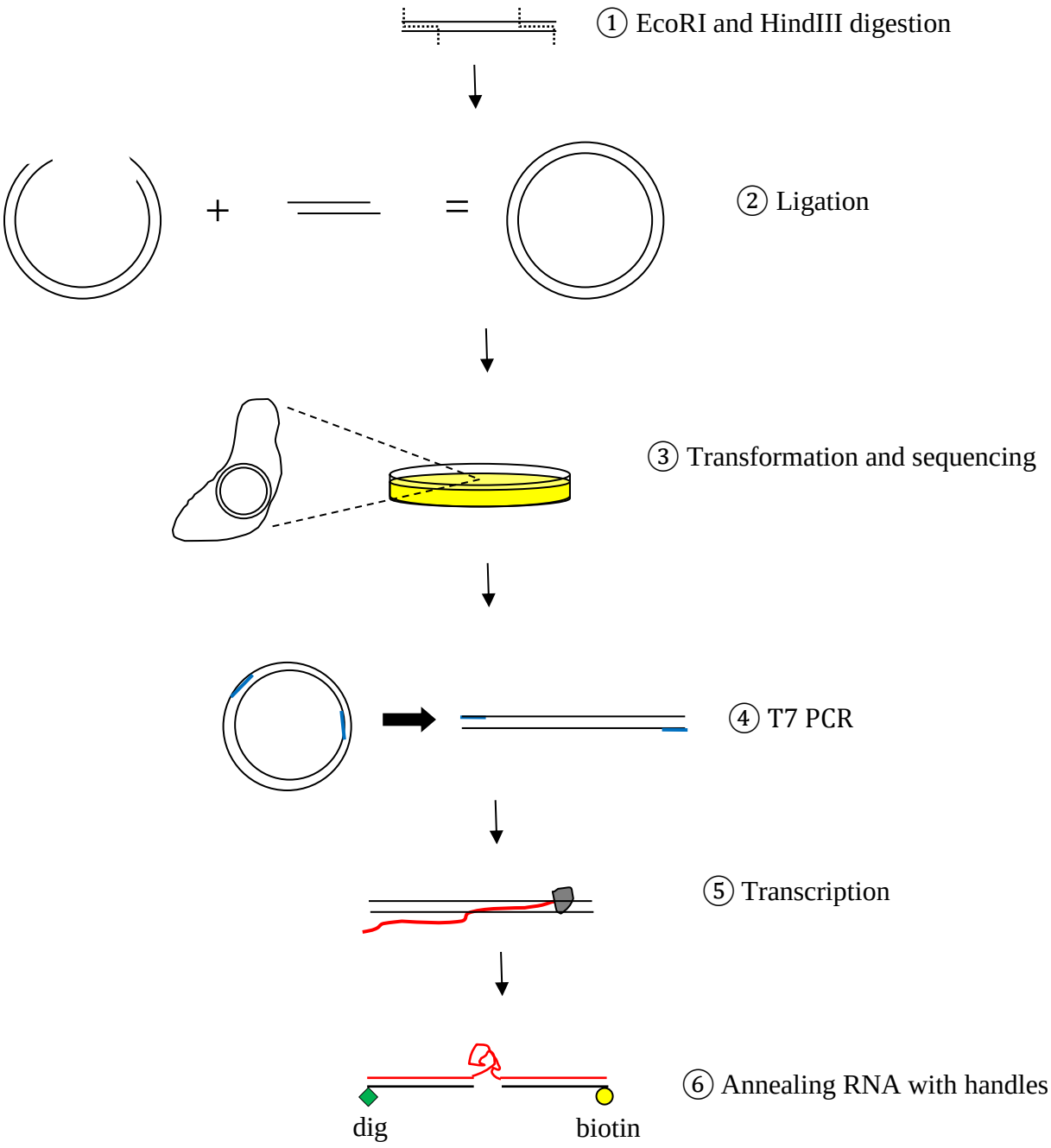


Figure 2.6 Schematic workflow for making RNA constructs for single-molecule study.

2.6 Data analysis

2.6.1 Modified WLC fit to force-extension curves (FECs)

The FECs obtained from our force-ramp data were fitted to the worm-like chain (WLC) model (126-128), which describes the entropic elasticity of semi-flexible polymers, such as DNA, RNA and polypeptide. The force (F) versus extension (x) relation in the WLC model was developed by Marko and Siggia (129, 130) as,

$$F = \frac{k_B T}{P}\left[\frac{1}{4}\left(1-\frac{x}{L}\right)^{-2} - \frac{1}{4} + \frac{x}{L}\right] \qquad \text{(Eq. 2.4)}$$

where k_B is Boltzmann's constant, T is temperature, P is the persistence length, and L is the contour length. Later, by adding the enthalpic contribution for stretching, the modified WLC (mWLC) equation (131) was proposed as

$$F = \frac{k_B T}{P}\left[\frac{1}{4}\left(1-\frac{x}{L}+\frac{F}{K}\right)^{-2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K}\right] \qquad \text{(Eq. 2.5)}$$

where K is the stretch modulus. In this study, the following fitting parameters for the handles and RNA were used: P_{handle} = 3-6 nm, L_{handle} = 0.25 nm/bp, K_{handle} = 500-800 pN, and P_{RNA} = 1 nm, L_{RNA} = 0.59 nm/base, K_{RNA} = 1600 pN. The persistence length and stretch modulus agree well with the previously reported values (124, 132).

2.6.2 Transition force and distance distributions from FECs

The extension change (Δx) of the molecule during the unfolding/refolding cycles as observed in FECs was determined as the extension difference of the molecule before and after the transition at a given force. All the observed transitions are shown as force and distance histogram plots. The number of nucleotides unfolded/refolded can be obtained by relating to mWLC model (Eq. 2.5). For example, since the mWLC predicted RNA inter-nucleotide distance is 0.43 nm at 14.5 pN, the expected unfolding distance for the 49-nt P5ab hairpin at this force is 21 nm.

Chapter 3 Single-Molecule Mechanical Unfolding Studies of P5ab Hairpin

3.1 Introduction

As mentioned in Chapter 1, RNA hairpin is the building block of more complex RNA structures, such as multi-branched junctions, pseudoknots, and kissing loops (36). Besides, RNA hairpin is an important structural motif involved in various biological processes, such as catalysis in ribozyme (133, 134), gene silencing in microRNA and small interfering RNA (12), termination of transcription (135, 136) and initiation of translation (137). Thus, elucidating the folding mechanisms of RNA hairpins can help us better understand their biological functions.

Over the last four decades, the folding dynamics of RNA hairpins has been experimentally measured by temperature jump (T-jump) method where rapid heating is achieved by electric discharge or laser radiation (time resolution ~100 ns). Table 3.1 shows a list of T-jump experiments on RNA hairpins to date (69-78). The RNA conformational changes are monitored by various probes, including UV absorbance, IR absorbance and fluorescence. Most studies have been focused on RNA hairpins with 2-6 base pairs in the stem. It is found that these small hairpins fold on timescales of 1-10 μs, except for large hairpin loops (>15 nt) (76). Moreover, multiphasic kinetics has been observed from the relaxation times (72, 74, 75, 77, 78), suggesting hairpin folding could involve more than two states. Menger et al. (72) studied the dynamics of GAAA tetraloop hairpin, where A^6 and A^7 are each substituted by a fluorescent probe, 2-aminopurine (2AP) (Table 3.1). They observed a relaxation process with a time constant of 22 μs in HP1 (with A^6 substituted by 2AP), and two relaxation processes with time constants of 5 μs and 41 μs for HP2 (with A^7 substituted by 2AP), as shown in Table 3.1. These results demonstrated the existence of more than one conformation state of the hairpin loop. Another tetraloop hairpin motif, UUCG, was studied by Ma et al. (74) using a combination of T-jump and UV absorbance. They observed three unfolding relaxation phases (fast, medium, and slow), which required a four-state model to fit the kinetics. This suggested the hairpin has a rugged energy landscape. Later, Stancik et al. (75) probed the relaxation kinetics of the same RNA sequence at two different infrared wavenumbers, 1574 cm^{-1} and 1669 cm^{-1}, which are indicators

of base stacking and hydrogen bonding respectively. With a single phase at 1574 cm^{-1} and bi-exponential behavior at 1669 cm^{-1}, they concluded that the base unstacking dynamics (~1 μs) occurs faster than the hydrogen-bond breakage (~10 μs). More recently, the Gruebele group used 2AP fluorescence to probe the stem and loop dynamics of the UUCG and UACG tetraloop hairpins (77, 78). Based on the multi-exponential behavior of the fluorescence decay profiles, as shown in Figure 3.1, they proposed a four-state free-energy landscape for the RNA hairpin. Combining the experimental data with molecular dynamics simulations, they interpreted the structures of the multiple states along the unfolding pathway of the hairpin: from the Native state, to a stem-fraying state (Frayed), to loss of hydrogen bonding and partial unstacking (Unfolded), and to additional unstacking of single strands (Unstacked), as shown in Figure 3.1C.

However, in the above ensemble experiments, the transition kinetics between the multiple states is not directly observed. Moreover, the folding of RNA hairpins with longer stems (>10 bp) have not been well studied, probably due to the difficulty in interpretation of the spectroscopic data. Single molecule force spectroscopy has the advantage of observing individual molecules throughout the course of events, and is thus free from the averaging effect as in bulk ensemble studies. It also allows RNA structures to denature at physiological temperature and buffer conditions, using pN scale forces. So far, single-molecule force spectroscopy has been used to study folding of several RNA hairpins, derived from ribozyme (115, 138), viral (105), and small interfering RNA (139) sequences. In these studies, the RNA hairpins are only treated as a two-state folder without any intermediate states. I studied P5ab as a model RNA hairpin for my thesis. P5ab forms part of the catalytic core domain of the group I ribozyme. The P5ab hairpin consists of one apical GAAA tetraloop and two stems, P5a and P5b, with an internal loop $A^{13}A^{14}/G^{38}G^{37}$ connected in between (Figure 3.2A). With high resolution optical tweezers instrument in our lab, I wanted to address the following two aims:

Aim 1: What is the folding pathway of P5ab hairpin? Are there any intermediate states?

Aim 2: What are the kinetics and energetics of the folding transitions?

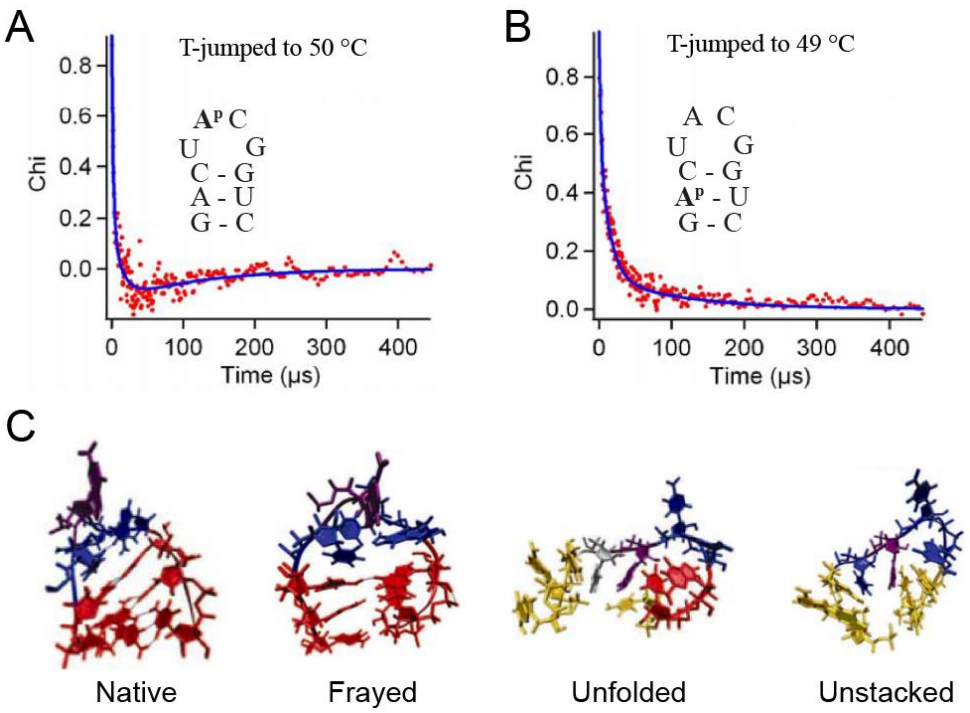


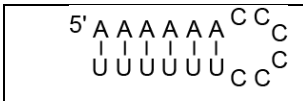
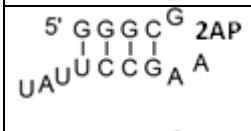
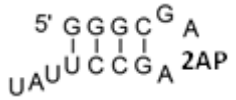
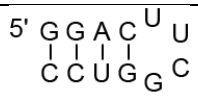
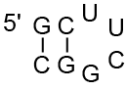
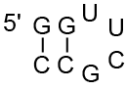
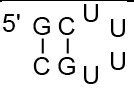
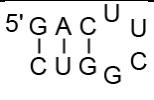
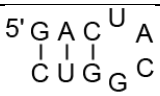
Figure 3.1 Hairpin folding intermediates detected by temperature jump with fluorescent probe.

(A)(B) Fluorescence decay profiles upon temperature jumps for the hairpin sequences indicated in the figures. **A^p** stands for 2-aminopurine used as a fluorescence probe. The data are fitted by sum of multiple exponentials (blue). (C) Four conformational states proposed by molecular dynamics simulation to explain the multiple phases observed in (A) and (B). This figure is adapted from reference (78).

Table 3.1 Temperature-jump experiments on RNA hairpin folding dynamics

The melting temperatures (T_m) are determined by thermal melting experiments on the RNA sequences. The relaxation times represent T-jump measurements at temperature near T_m, unless indicated otherwise.

Sequence	Method	T_m (°C)	Relaxation time (μs)	References
5' UGGGCC GCCCGC	T-jump UV absorbance	88	~10	(69)
5' AAAAGCC UUUUCC	T-jump UV absorbance	20	~10	(70)

	T-jump UV absorbance	20	20	(71)
 (HP1)  (HP2)	T-jump Fluorescence	66	22 (HP1 at 8.1 °C) 5, 41 (HP2 at 8.1 °C)	(72)
	T-jump UV absorbance	71.8	9.1	(73)
	T-jump UV absorbance	59.2	5.3	(73)
	T-jump UV absorbance	49.9	7.6	(73)
			Multiple phases	(74)
	T-jump IR absorbance	58	3.1 (1574 cm^{-1}) 3.7, 14.9 (1669 cm^{-1})	(75)
	T-jump UV absorbance	27.7	53	(73)
			Multiple phases	(74)
	T-jump UV absorbance	27	Multiple phases	(74)
 j = 0, 5, 15, 30	T-jump Fluorescence	58.3 55.0 41.3 36.2	10.3 36.3 334.2 404.4	(76)
	T-jump Fluorescence	60	Multiple phases	(77)
	T-jump Fluorescence	55	Multiple phases	(78)

3.2 Materials and Methods

3.2.1 RNA constructs and reagents

The RNA constructs (P5ab and mutants) for optical tweezers studies were generated as described in Chapter 2. All single-molecule folding/unfolding measurements were performed in a buffer containing 10 mM Tris-HCl, pH 7.5, 250 mM NaCl and 3 mM $MgCl_2$ at 25 °C. Tris buffer and sodium chloride (NaCl) were purchased from Sigma-Aldrich, and magnesium chloride ($MgCl_2$) was purchased from EMD. The anti-digoxigenin coated beads and streptavidin coated beads were purchased from Spherotech.

3.2.2 Force ramp, constant force and passive mode experiments

In the force ramp experiments, the tether between the beads was consecutively stretched and relaxed between 1 pN and 30 pN by moving the trap at a constant speed. Typically the pulling speed was 200 nm/sec, which corresponded to a loading rate of 7.6 ± 0.4 pN/sec (mean ± standard deviation, 163 traces in 5 molecules), measured from the linear fits to force vs. time plots. Other pulling rates (50 nm/sec, 400 nm/sec and 800 nm/sec) were also used, in order to investigate the effect of loading rates (1.8 pN/sec, 14.7 pN/sec, and 28.4 pN/sec) on the rupture force distributions and the non-equilibrium mechanical work done. All force ramp data was collected at 400 Hz.

In the constant force (CF) mode, the molecule was held at a pre-set force to investigate its hopping behavior. The force was stabilized through feedback control that operates at 4 kHz. The data acquisition rate for the CF data was initially 500 Hz, and then improved to 4 kHz.

In addition to the CF mode, I also measured the hopping of P5ab molecules in passive mode (PM) at 4 kHz, wherein the trap was held stationary near the transition force without an active force feedback, thus enabling the molecule to hop between various states as the force varied between the transitions. The effective stiffness κ_{eff} was calculated as κ_{eff} = 0.045 ± 0.001 pN/nm (mean ± standard deviation from 27 traces of 2 molecules).

3.2.3 State identification in CF traces

To identify states in single-molecule data, several methods have been established including the threshold method (140), hidden Markov model (140, 141), change-point method (142), and wavelet denoising (143, 144). Here I used the threshold method for state identification in the P5ab extension (x) vs. time CF traces, as described below and illustrated in Figure 3.6.

For three-state analysis:

Two threshold values x_1 =10 nm and x_2 =14 nm were selected such that the observed extension, x, for F and U corresponded to $x < x_1$ and $x > x_2$ respectively, and I in between x_1 and x_2.

For five-state analysis:

In order to separate the five states, i.e. F, I_1, I, I_2, and U, the following four thresholds were used: x_1= 2.5 nm, x_2= 8.5 nm, x_3=15 nm, x_4=17 nm, such that, $x < x_1$ corresponds to F state; $x_1 \leq x < x_2$ corresponds to I_1 state; $x_2 \leq x < x_3$ corresponds to I state; $x_3 \leq x < x_4$ corresponds to I_2 state; $x \geq x_4$ corresponds to U state.

The code for the threshold analysis was generated in the lab using Matlab software.

3.3 Results

3.3.1 Characterization of P5ab unfolding and refolding

In order to study its folding behavior, I subjected the P5ab RNA to multiple pull-relax cycles between 1 pN and 30 pN. In the force-extension curves (FECs), I observed ripping and zipping events which represent the unfolding and refolding of the molecule respectively (Figure 3.2B). The molecule unfolds at 15.0 ± 0.5 pN (mean ± standard deviation, Figure 3.2C) and refolds at 13.9 ± 0.6 pN (Figure 3.2D) with unfolding distance of 20.5 ± 1.3 nm (Figure 3.2E) and refolding distance of 19.9 ± 1.4 respectively (Figure 3.2F). With an extension of ~0.43 nm/nt as predicted by the mWLC model, the transition distances correspond to a contour length change of 48 ± 3 nt, indicating the opening or closing of the complete 49-nt hairpin structure.

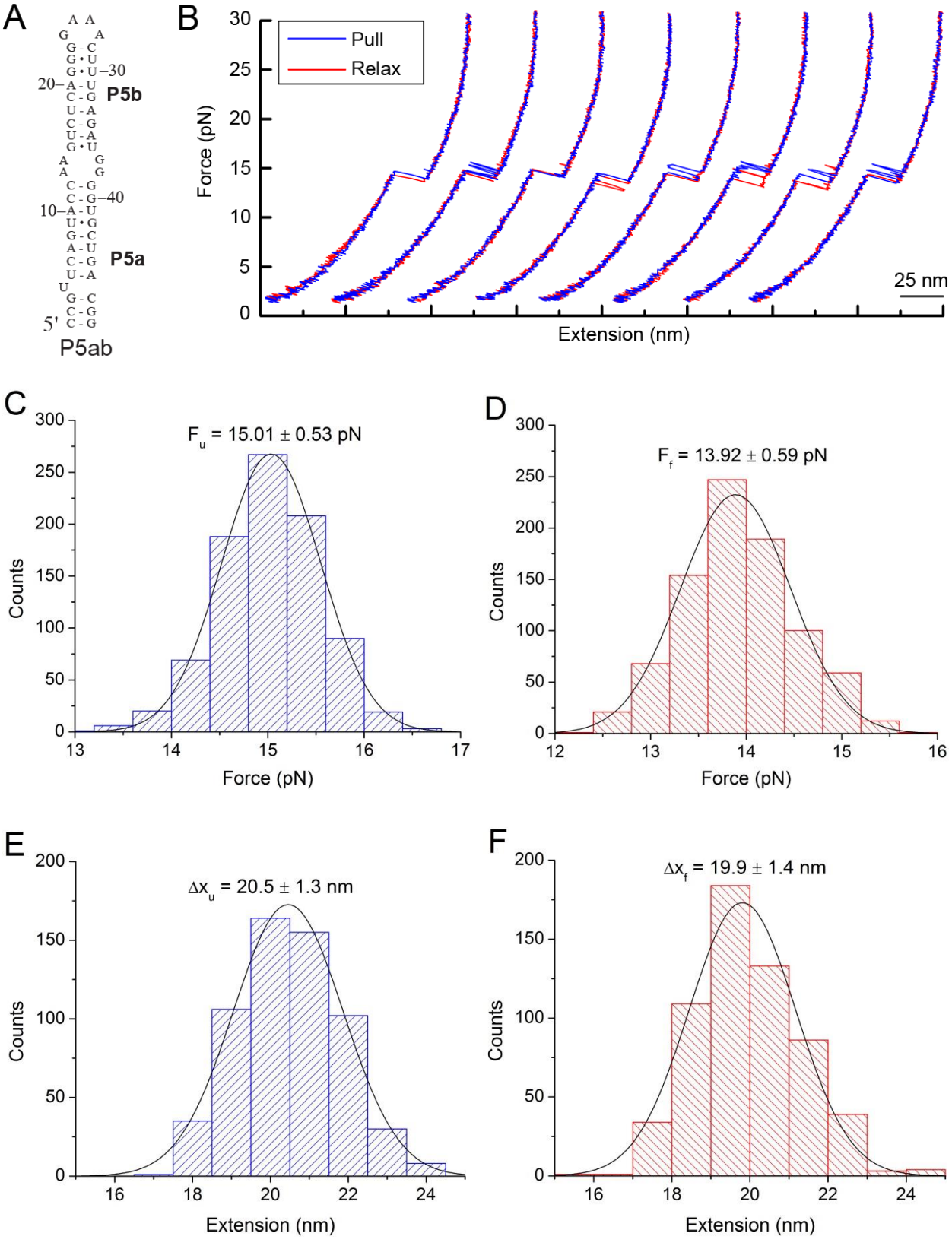

A
P5b
P5a
P5ab
B
Pull
Relax
Force (pN)
Extension (nm)
25 nm
C
F_u = 15.01 ± 0.53 pN
Counts
Force (pN)
D
F_f = 13.92 ± 0.59 pN
Counts
Force (pN)
E
Δx_u = 20.5 ± 1.3 nm
Counts
Extension (nm)
F
Δx_f = 19.9 ± 1.4 nm
Counts
Extension (nm)

Figure 3.2 Unfolding and refolding characteristics of P5ab in force ramp experiment.

(A) Sequence and secondary structure of P5ab RNA hairpin.

(B) Consecutive force-ramp cycles for P5ab, showing unfolding (blue) and refolding (red) transitions.

(C) Distribution of the unfolding force F_u for P5ab RNA measured from force ramp experiments. The counts are representative of 600 pulls on 22 molecules, and binned at 0.4 pN intervals. F_u shows a mean and standard deviation of 15.01 ± 0.53 pN. The Gaussian fitted mean is 15.04 ± 0.005 pN.

(D) Distribution of the refolding force F_f from the same experiments. F_f shows a mean and standard deviation of 13.92 ± 0.59 pN. The Gaussian fitted mean is 13.89 ± 0.03 pN.

(E) Distribution of the unfolding distance Δx_u from the same experiments. The mean and standard deviation of Δx_u is 20.5 ± 1.3 nm. The Gaussian fitted mean is 20.46 ± 0.04 nm.

(F) Distribution of the refolding distance Δx_f from the same experiments. The mean and standard deviation of Δx_f is 19.9 ± 1.4 nm. The Gaussian fitted mean is 19.81 ± 0.08 nm.

3.3.2 Identification of intermediate states

Besides the folded state F and the unfolded state U, other intermediate extensions were also observed on the FECs (Figure 3.3). These intermediate states (I_1, I, and I_2) can transit both to F and U, with transition forces and distances shown in Table 3.2.

Table 3.2 Transition distances and forces for the intermediates in P5ab force ramps

Transition	Δx (nm)	Force (pN)	Observed number of events
F↔I_1	4.06 ± 0.60	14.44 ± 0.54	158
F↔I	11.59 ± 0.66	14.34 ± 0.48	19
F↔I_2	15.71 ± 0.59	13.85 ± 0.53	5
U↔I_1	16.70 ± 0.63	15.44 ± 0.53	13
U↔I	8.73 ± 0.86	15.03 ± 0.54	35
U↔I_2	4.71 ± 0.70	14.85 ± 0.73	88

Data represents mean ± standard deviation, measured from 197 FECs of 7 molecules.

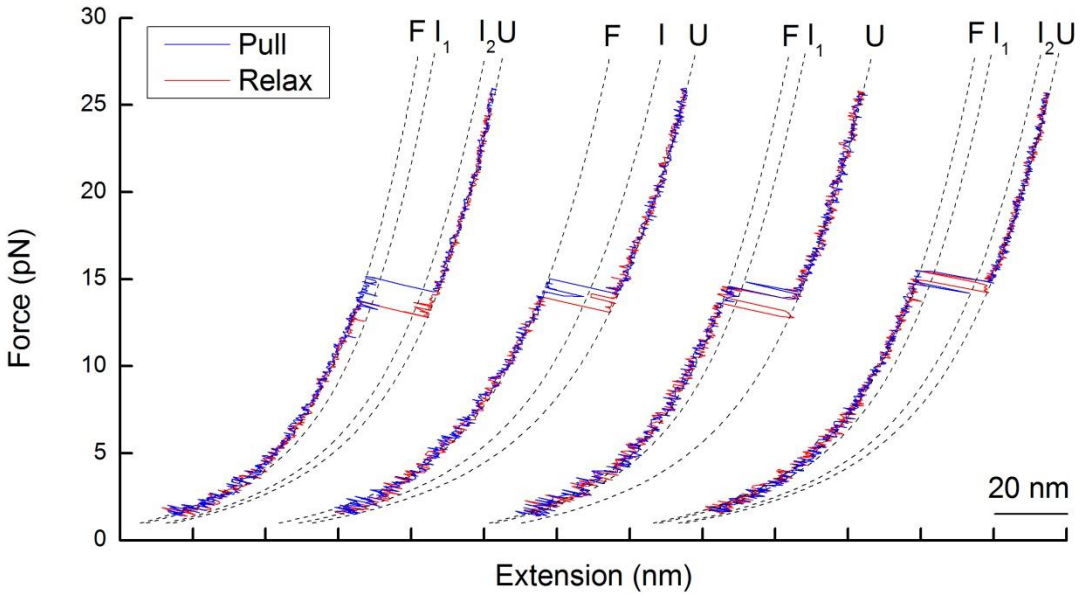


Figure 3.3 Intermediate states I_1, I and I_2 observed during the force ramp cycles of P5ab.

The FECs were fitted by the modified WLC model (black dashed lines) with the fitting parameters P_{handle} = 3-6 nm, L_{handle} = 0.25 nm/bp, K_{handle} = 500-800 pN, and P_{RNA} = 1 nm, L_{RNA} = 0.59 nm/base, K_{RNA} = 1600 pN.

In order to directly observe the unfolding and refolding dynamics among the various states, I held the molecule at a preset force where the molecule displays bi-stability between F and U. The force was maintained through feedback control. As shown in Figure 3.4, the extension of P5ab hopped back and forth by 20.0 ± 1.3 nm, signaling the repeated folding and unfolding of the hairpin. Moreover, the relative populations between F and U were shifted by the force: as the force was increased from 14.1 pN to 14.9 pN, the molecule spent more time in the unfolded state and less time in the folded state.

Consistent with the force ramp experiments, here I also observed intermediate states I_1, I, and I_2, as shown by the 14.5 pN traces in Figure 3.4. The extensions of these intermediates from the folded state were measured to be 4.1 ± 0.4 nm, 12.4 ± 0.6 nm, and 15.7 ± 0.4 nm (mean ± standard deviation from 4 molecules) respectively.

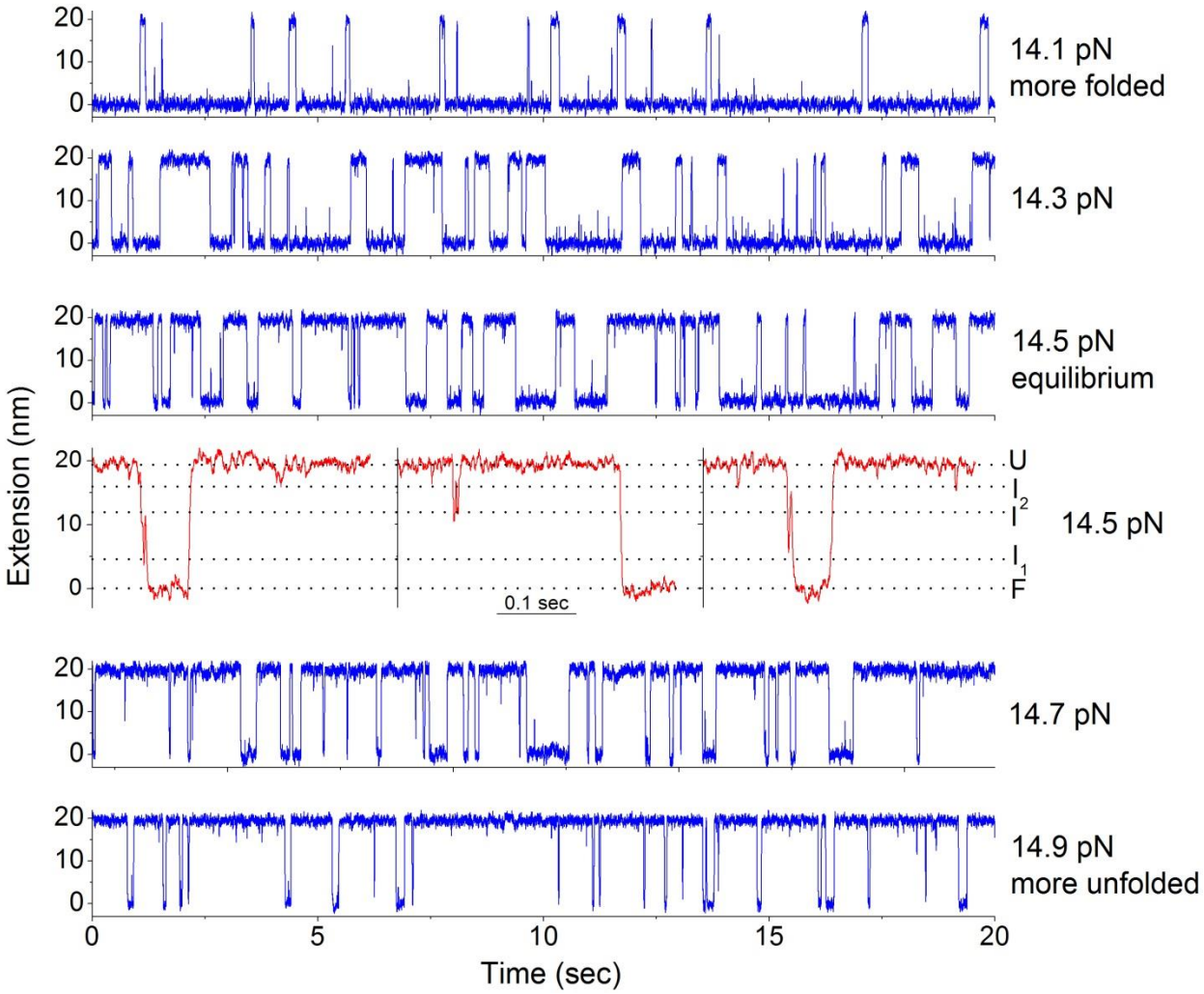


Figure 3.4 Extension versus time traces of P5ab measured at various constant forces.

Besides F and U, three intermediate states I_1, I, and I_2 were resolved, with extensions of 4.1 ± 0.4 nm, 12.4 ± 0.6 nm, and 15.7 ± 0.4 nm (mean ± standard deviation, 4 molecules) respectively. The transitions of the intermediate states are highlighted in the 14.5 pN CF traces (red).

These intermediate states were further confirmed in the passive mode experiment, wherein the trap position is held fixed and the force on the molecule fluctuates as it folds and unfolds. As shown in Figure 3.5, the force oscillates around 14.5 pN between various states. The force changes were then converted to extension changes by the effective stiffness κ_{eff} = 0.045 pN/nm. The unfolded distances for the intermediates I_1, I and I_2 are 4.45 ± 0.38 nm, 12.15 ± 0.38 nm, and 15.75 ± 0.47 nm respectively.

Based on the intermediate extensions measured from the above force ramp, constant force and passive mode experiments, I calculated the nucleotide positions of the intermediate states: I_1 corresponds to opening 9 nucleotides up to U^5, I up to the $A^{13}A^{14}/G^{38}G^{37}$ internal loop, and I_2 up to the base pair A^{20}-U^{31}.

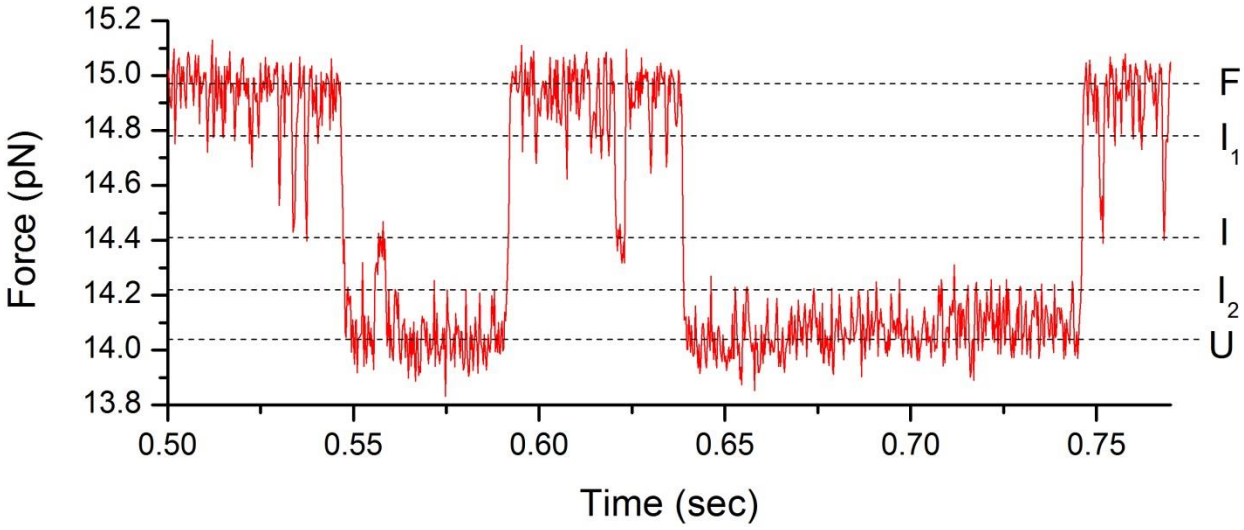


Figure 3.5 Passive mode measurement of P5ab.

The force versus time hopping trace for P5ab RNA in passive mode. The data is collected at 4 kHz. The intermediate states I_1, I, and I_2 show extensions of 4.45 ± 0.38 nm, 12.15 ± 0.38 nm and 15.75 ± 0.47 nm from the folded state (mean ± standard deviation from 5 passive mode traces).

3.3.3 Kinetics of P5ab unfolding and folding

From the transitions in the CF traces, I calculated the dwell times and kinetics between various states. Typically, at a given force if a molecule hops between states A and B, then the dwell time, $\tau_{A\to B}$ is the time the molecule spends in state A before transitioning to state B. The states on the extension vs. time traces were identified by the threshold method, as shown in Figure 3.6.

The mean of the dwell time $<\tau_{A\to B}>$ is inversely dependent on the rate constant, $k_{A\to B}$, which represents the frequency of a molecule showing a transition from A to B (104).

$$k_{A\to B} = \frac{1}{\langle \tau_{A\to B} \rangle} \qquad \text{(Eq. 3.1)}$$

The distribution of the dwell time τ can also be plotted as a normalized cumulative histogram or integrated probability p(τ) (Figure 3.7). According to the first-order reaction rate theory, p(τ) can be fitted by a single exponential function (117, 145):

$$p(\tau) = 1 - e^{-k\tau} \qquad \text{(Eq. 3.2)}$$

Both the direct (Eq. 3.1) and fitting (Eq. 3.2) methods yielded similar rate values for F→U and U→F transitions (Figure 3.7A and B). Similarly, the dwell time distribution for F→I transition was calculated and agreement between the two methods was also observed (Figure 3.7C). The rates measured at a series of constant forces are summarized in Table 3.3. At 14.5 pN, the forward and reverse rates between F and U are close to each other: $k_{F,U}$ = 6.15 ± 0.14 sec^{-1}, and $k_{U,F}$ = 3.66 ± 0.08 sec^{-1}, suggesting F and U reach an equilibrium. On the other hand, the rates coming from I state are two orders of magnitude faster: $k_{I,F}$ = 425.0 ± 3.6 sec^{-1} and $k_{I,U}$ = 313.7 ± 3.0 sec^{-1}. Similarly, the rates coming from I_1 and I_2 are greater than 200 sec^{-1}. These rates coming from the intermediates are one order of magnitude smaller than our data collection rate (4 kHz) and constant force feedback rate (4 kHz), suggesting they are not artifacts due to instrument time resolutions. Moreover, the intermediate extensions are much larger than the typical extension noise (± 0.6 nm, Figure 3.17) from beads and handles, which further supports that the intermediates are not noise fluctuations.

Thus, the folding pathway of P5ab is illustrated in Figure 3.8. During folding, the rate from U to I is slow ($k_{U,I}$ = 3.95 s^{-1}), whereas the step from I to F ($k_{I,F}$ = 425 s^{-1}) is very fast. Along the

unfolding direction, the rate from F to I ($k_{F,I}$ = 6.27 s^{-1}) is much slower than I to U ($k_{I,U}$ = 314 s^{-1}). This suggests the transition to I state is the rate-limiting step for folding and unfolding.

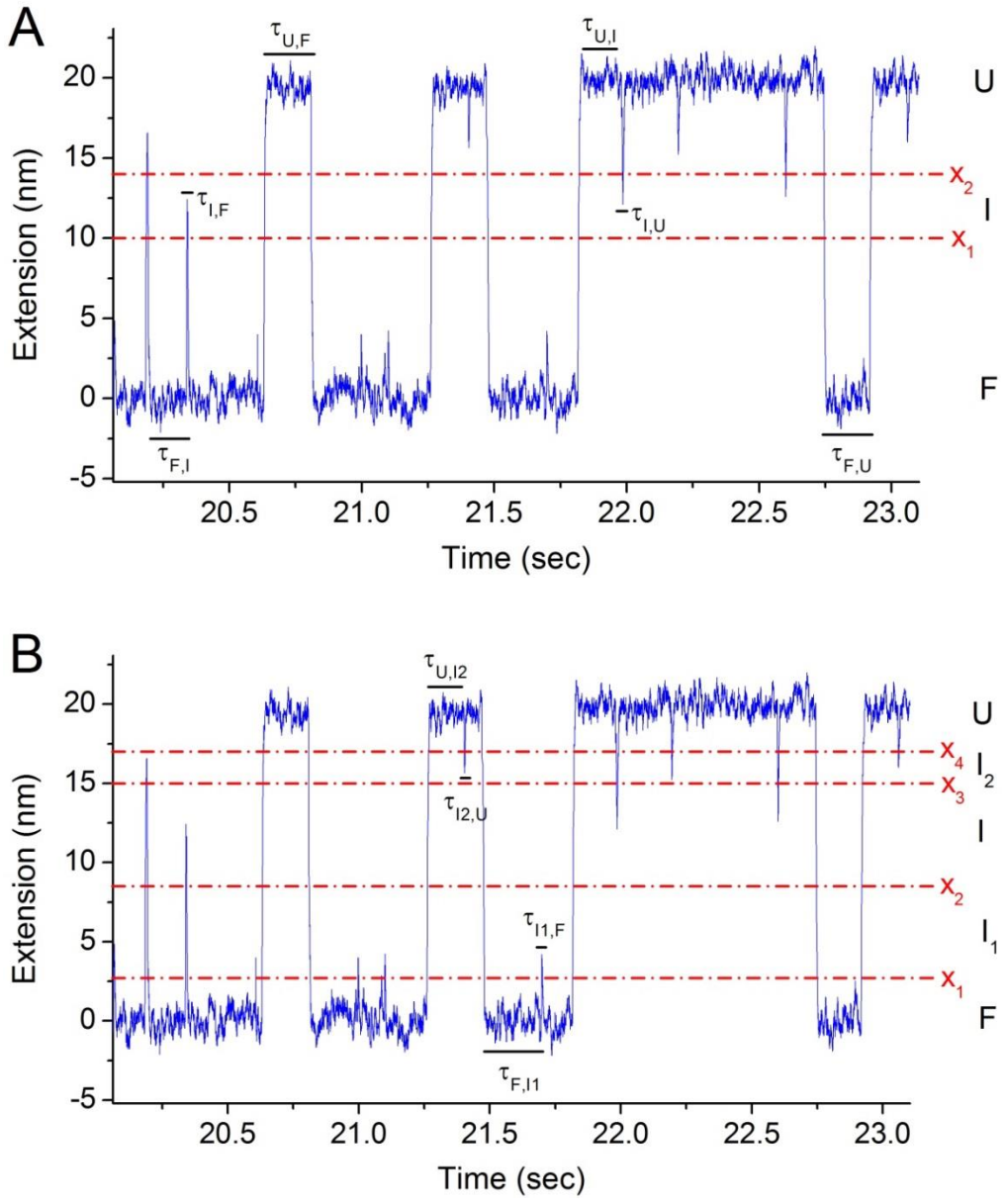
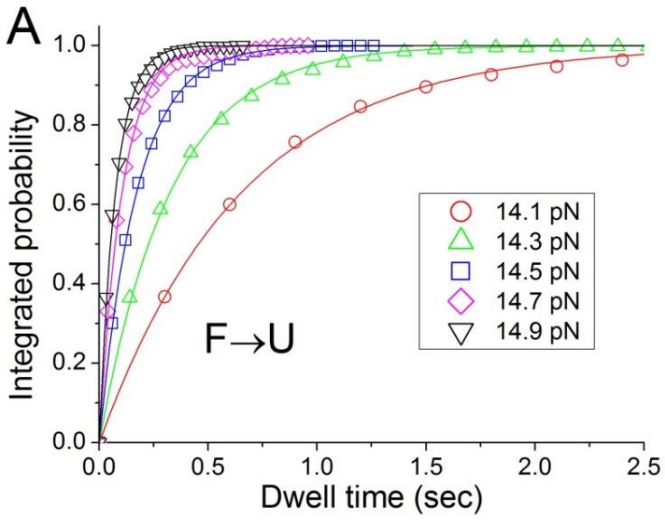


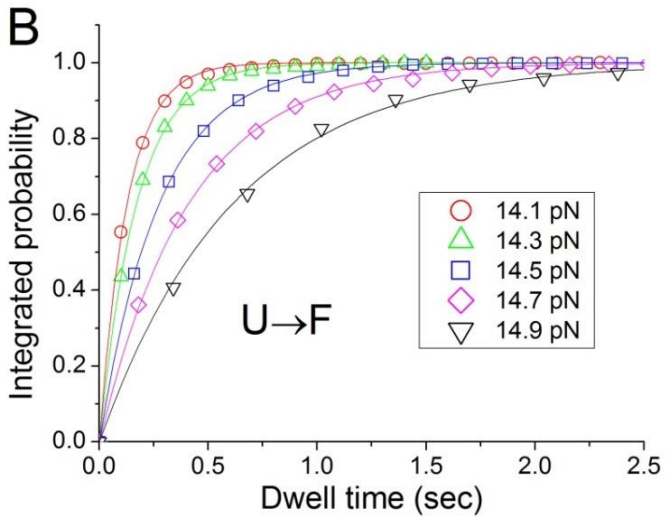
Figure 3.6 State identification in P5ab CF traces by the threshold method.

Threshold method applied to a 14.5 pN CF trace in order to identify (A) three states (F, I and U) or (B) five states (F, I_1, I, I_2 and U). The threshold extension values are mentioned in Materials and Methods section. Examples of the dwell times τ for various transitions are shown in the figure.



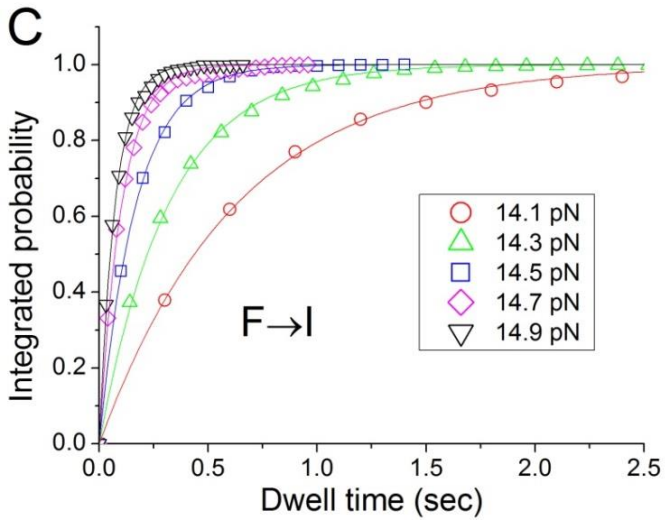
F→U

Force (pN)	k (s^{-1}) (exponential fit)	k (s^{-1}) ($1/\langle\tau\rangle$)
14.1	1.52±0.01	1.54±0.04
14.3	3.07±0.03	3.19±0.08
14.5	5.83±0.03	6.15±0.14
14.7	9.60±0.14	10.04±0.34
14.9	13.55±0.21	14.12±0.51



U→F

Force (pN)	k (s^{-1}) (exponential fit)	k (s^{-1}) ($1/\langle\tau\rangle$)
14.1	7.81±0.07	8.02±0.25
14.3	5.79±0.04	5.74±0.13
14.5	3.61±0.01	3.66±0.08
14.7	2.42±0.01	2.46±0.07
14.9	1.62±0.02	1.61±0.06



F→I

Force (pN)	k (s^{-1}) (exponential fit)	k (s^{-1}) ($1/\langle\tau\rangle$)
14.1	1.59±0.01	1.61±0.04
14.3	3.15±0.03	3.29±0.08
14.5	5.94±0.04	6.27±0.14
14.7	9.72±0.13	10.20±0.34
14.9	13.71±0.20	14.25±0.51

Figure 3.7 Dwell time distributions of P5ab at constant forces.

Dwell time distributions of (A) F to U, (B) U to F, and (C) F to I from 69 CF traces of 16 molecules. The plots are displayed as normalized cumulative counts, i.e. integrated probabilities. The distributions were fit by single exponentials which yield similar rate constants to the reciprocal of average dwell times, as shown by the tables on the right of each figure. The errors represent standard errors.

Table 3.3 P5ab transition rate constants measured in CF experiments

Three-state analysis						
Force (pN)	$k_{F,I}$ (s^{-1})	$k_{I,F}$ (s^{-1})	$k_{I,U}$ (s^{-1})	$k_{U,I}$ (s^{-1})	$k_{F,U}$ (s^{-1})	$k_{U,F}$ (s^{-1})
14.1	1.61 ± 0.04	421.2 ± 4.6	326.3 ± 3.8	8.20 ± 0.25	1.54 ± 0.04	8.02 ± 0.25
14.3	3.29 ± 0.08	416.0 ± 3.9	324.7 ± 3.1	6.00 ± 0.13	3.19 ± 0.07	5.74 ± 0.13
14.5	6.27 ± 0.14	425.0 ± 3.6	313.7 ± 3.0	3.95 ± 0.08	6.15 ± 0.14	3.66 ± 0.08
14.7	10.20 ± 0.34	432.4 ± 4.8	302.2 ± 3.6	2.80 ± 0.08	10.04 ± 0.34	2.46 ± 0.07
14.9	14.25 ± 0.51	422.5 ± 5.2	279.0 ± 4.0	1.94 ± 0.06	14.12 ± 0.51	1.61 ± 0.06
Five-state analysis						
Force (pN)	$k_{F,I1}$ (s^{-1})	$k_{I1,F}$ (s^{-1})	$k_{I2,U}$ (s^{-1})	$k_{U,I2}$ (s^{-1})		
14.1	3.24 ± 0.08	240.7 ± 4.4	250.5 ± 7.6	9.25 ± 0.29		
14.3	5.27 ± 0.16	241.2 ± 5.2	245.5 ± 8.3	5.85 ± 0.18		
14.5	8.87 ± 0.33	245.1 ± 7.4	227.0 ± 7.2	4.96 ± 0.16		
14.7	11.37 ± 0.66	247 ± 10	254.0 ± 8.6	3.71 ± 0.19		
14.9	20.59 ± 0.88	202.3 ± 9.1	240.4 ± 7.0	2.63 ± 0.08		

Data represents mean ± standard error, measured from 69 CF traces of 16 molecules.

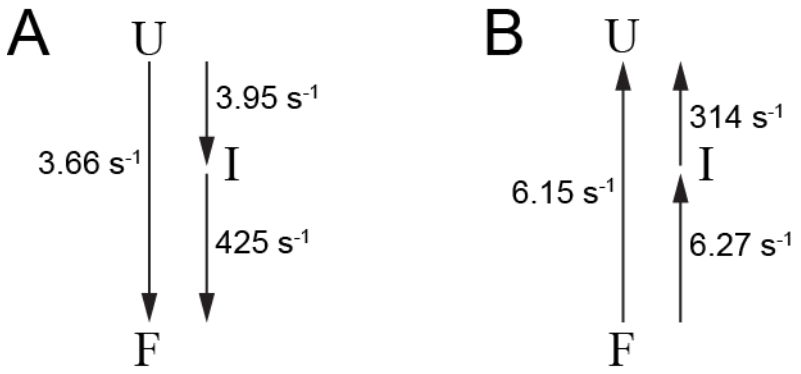


Figure 3.8 Folding pathways of P5ab.

The folding pathway of P5ab in the (A) folding and (B) unfolding direction. The rates between F, I and U are measured at 14.5 pN (Table 3.3).

3.3.4 Equilibrium folding free energy from CF measurements

At force *F*, the equilibrium constant K(F) between the folded state and unfolded state is defined as the ratio between their corresponding probabilities ($P_{unfolded}/P_{folded}$). Using detailed balance, K(F) can also be related to their rate constants as

$$K(F) = \frac{k_{F,U}}{k_{U,F}} \qquad \text{(Eq. 3.3)}$$

The corresponding free energy change at force *F* is

$$\Delta G(F) = -k_B T ln\, K(F) \qquad \text{(Eq. 3.4)}$$

The thermodynamic stability of the hairpin is represented as the free energy change at zero force, which is obtained from $\Delta G(F)$ by the following relation (146):

$$\Delta G(0) = F \cdot \Delta x(F) + \Delta G(F) - \int_0^{\Delta x} F dx_{ss} \qquad \text{(Eq. 3.5)}$$

where Δx is the extension change at force *F*. The third term is the free energy of stretching the single-stranded RNA to force *F*, calculated from the mWLC model taking T = 298 K, P = 1 nm, L = 0.59 nm/base and K = 1600 pN. $\Delta G(0)$ can be graphically represented as the hatched area in

Figure 3.9, since in Eq. 3.5 the first term $F \cdot \Delta x$ is the work done, the second term $\Delta G(F)$ equals to zero if transition occurs at equilibrium force, and the third term can be represented as the difference between the stretching free energies of the folded state ($\Delta G_F^{stretch}$) and unfolded state ($\Delta G_U^{stretch}$). Using the measured $k_{F,U}$ and $k_{U,F}$ at F = 14.5 pN (Table 3.3), I obtained $\Delta G(0)$ = 52.44 ± 0.03 k_BT.

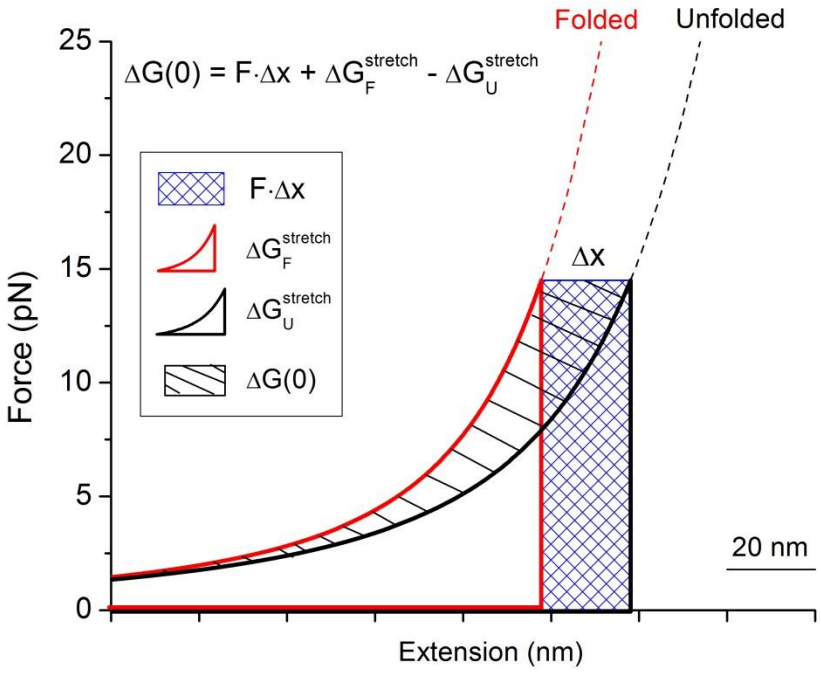


Figure 3.9 Graphical representation of folding free energy at zero force from Eq. 3.5.

The blue checkered area indicates the work done during unfolding. The red outlined triangle represents the stretching free energy of the handles. The black outlined triangle represents the stretching free energy of both the handles and unfolded RNA. Thus, the resulting ΔG(0) is the black hatched area. Note that the stretching free energy, especially $\Delta G_U^{stretch}$, can only be modelled by mWLC, but cannot be calculated from the experimental FEC, because the unfolded state does not exist at zero force.

3.3.5 Folding free energy from non-equilibrium work done

In the following section, the folding free energy of P5ab is examined from the mechanical work done during unfolding. Free energy is an important quantity used to describe many biological processes, such as folding, ligand-receptor binding, and energy transfer. For single-molecule experiments as temperature (T) and pressure (P) are constant, we use a modified Gibbs free energy to describe the system, such that the work done by the mechanical force is incorporated (145-147). Thus, the free energy change dG for a reversible reaction is

$$dG = - SdT + VdP + Fdx \qquad \text{(Eq. 3.6)}$$

where S is entropy, V is the volume, and dx is the molecular extension at a given force, F.

Equilibrium thermodynamics states that the free energy difference between state A and state B is equal to the reversible work needed to bring the system from A to B (148):

$$\Delta G_{AB} = W_{rev,\,AB} \qquad \text{(Eq. 3.7)}$$

Thus, A and B can be treated as the folded and the unfolded state in the force-extension curve. However, due to the finite loading rate on the molecule, the pulling process is non-equilibrium. As a result, the average work done in general is greater than the free energy change, i.e.

$$\langle W \rangle \geq \Delta G \qquad \text{(Eq. 3.8)}$$

Thanks to two recent non-equilibrium fluctuation relations, Jarzynski's Equality and Crooks Fluctuation Theorem (149, 150), it is now possible to deduce the equilibrium free energy from work values done under non-equilibrium conditions. I'll begin discussion of results by defining and explaining the following terms.

- Non-equilibrium fluctuation relations

Jarzynski's Equality (JE) (149) was developed in 1997. It relates the irreversible work W to the equilibrium free energy difference ΔG:

$$\exp(-\Delta G/kT) = \lim_{N\to\infty} \langle \exp(-W/kT) \rangle_N \qquad \text{(Eq. 3.9)}$$

where < > denotes averaged signal over N trajectories. In other words, ΔG can be determined by averaging Boltzmann-weighted work values from multiple measurements. However, JE is proven to be biased for finite N datasets and converge slowly. Therefore, if <W> - ΔG = 4 k_BT, it requires N ≈ 100 to obtain ΔG with estimation error within 1 k_BT; if <W> - ΔG = 8 k_BT, it requires N ≈ 1000 to obtain ΔG with estimation error within 1 k_BT (151, 152).

Later in 1999, the Crooks Fluctuation Theorem (CFT) (150) was proposed, which states that

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W-\Delta G}{kT}\right) \quad \text{(Eq. 3.10)}$$

where $P_F(W)$ is the distribution of the work performed along the forward direction, $P_R(-W)$ is the work distribution for the reverse direction, and ΔG is the Gibbs free energy difference between the initial and final state. Since CFT utilizes information from both the forward and reverse processes, it is converges faster and is more predictive than JE.

- Work done in single-molecule experiments

Consider the unfolding and folding of P5ab hairpin. In the forward process, the molecule unfolds as its extension increases from x_1 to x_2 (Figure 3.10A). In the reverse process, the molecule refolds as its extension decreases from x_2 to x_1. The work done is defined as the area under the corresponding FEC, i.e.

$$W = \int_{x_1}^{x_2} F dx \quad \text{(Eq. 3.11)}$$

The above integral can be numerically calculated by the trapezoid method:

$$W = \sum_i (F_i + F_{i+1}) \cdot (x_{i+1} - x_i)/2 \quad \text{(Eq. 3.12)}$$

where i and i+1 refer to corresponding data points on the FEC. In this report, the distribution of work done is obtained by calculating 171 traces where the extension boundaries x_1 to x_2 are chosen such that they cover all the unfolding and refolding transitions on the FECs (Figure 3.10A).

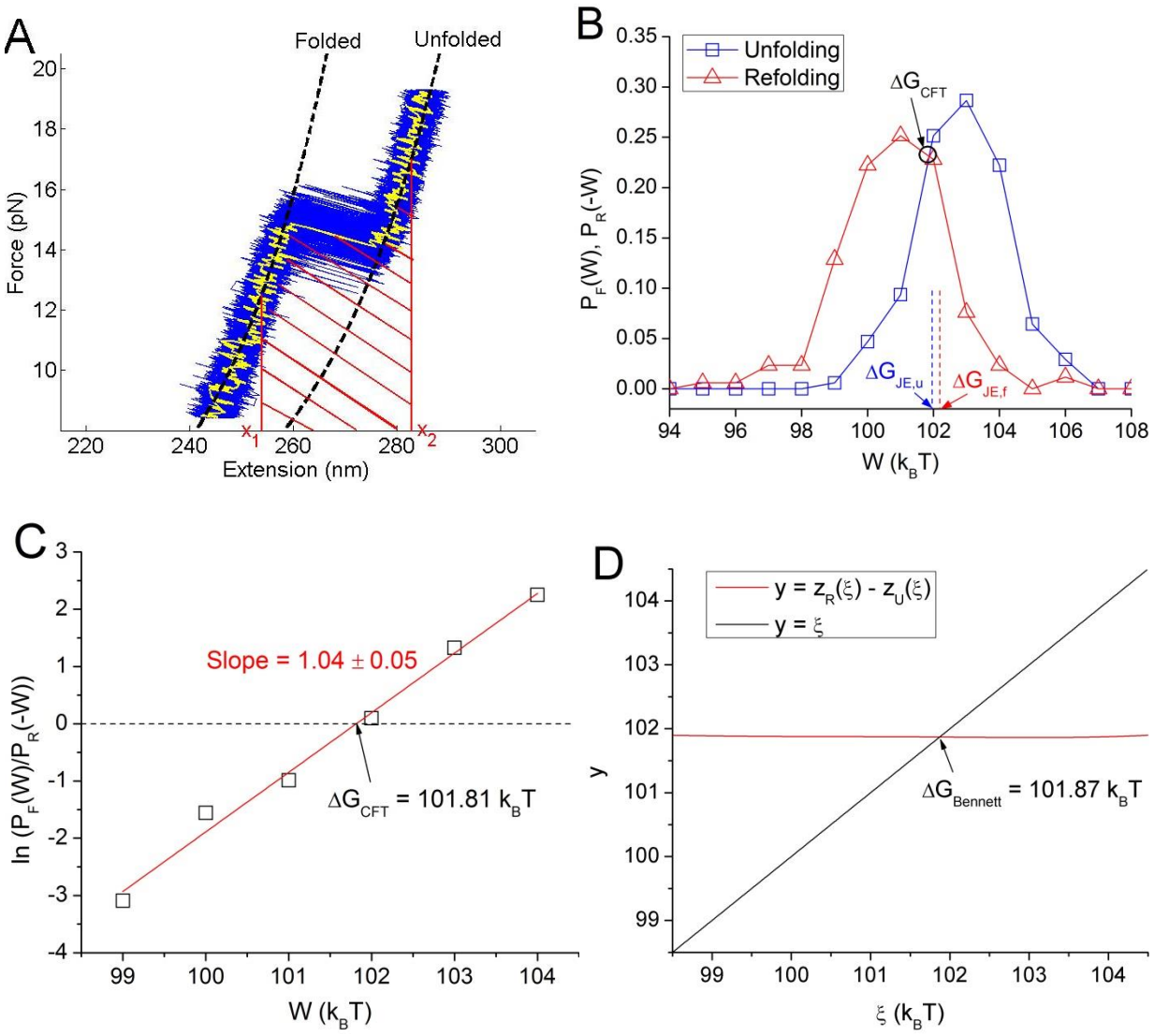


Figure 3.10 Determination of folding free energy of P5ab from non-equilibrium work done.

(A) Overlay of 171 unfolding FECs (blue) from one P5ab molecule, where the work done is defined as force integrated between two boundaries: x_1 = 253.9 nm and x_2 = 282.8 nm. The work done is shown as the red hatched area for one FEC (yellow). The work done for refolding is defined with similar boundaries. The dashed lines are mWLC fits to the folded state and unfolded state.

(B) The unfolding (square) and refolding (triangle) work distributions calculated from the 171 force ramp cycles from one P5ab molecule. The bin size is 1 k_BT. The unfolding and refolding free energies obtained from JE are indicated by blue and red dashed lines. The free energy from CFT lies at the intersection of the unfolding and refolding distributions, as shown by the circle.

(C) Fitting method based on the CFT equation (Eq. 3.10). From the work distributions in (B), $\ln\left(\frac{P_F(W)}{P_R(-W)}\right)$ is calculated, as shown by the squares. The fitted straight line (red) shows a slope of 1.04 ± 0.05, close to the theoretical value of 1. The x-intercept denotes the free energy ΔG_{CFT} = 101.81 k_BT.

(D) The Bennett acceptance ratio method applied to work values to extract ΔG.

- Estimation of ΔG using JE and CFT

Using JE (Eq. 3.9), the unfolding free energy is $\Delta G_{JE,u}$ = 101.96 k_BT, and refolding free energy is $\Delta G_{JE,f}$ = 102.19 k_BT. Using CFT (Eq. 3.10), the intersection point of $P_F(W)$ and $P_R(-W)$ gives the equilibrium free energy of ΔG_{CFT} = 101.87 k_BT (Figure 3.10B). Alternatively, a plot of $\ln\left(\frac{P_F(W)}{P_R(-W)}\right)$ against W is fitted by a straight line, the x-intercept gives ΔG_{CFT} = 101.81 k_BT (Figure 3.10C). The straight line fit shows a slope of 1.04 ± 0.05, close to the theoretical value of 1, which validates CFT. The third method for ΔG determination is the Bennett acceptance ratio method (139, 153), which rewrites Eq. 3.10 as follows:

$$\frac{\Delta G}{k_BT} = z_R(\xi) - z_F(\xi) \qquad \text{(Eq. 3.13)}$$

where $z_F(\xi) = \ln\left(\langle f_\xi(W)\exp\left(-\frac{W}{k_BT}\right)\rangle_F\right)$ and $z_R(\xi) = \ln\left(\langle f_\xi(W)\rangle_R\right)$, in which $< >_F$ and $< >_R$ denote averaging over the forward and reverse work distributions respectively; $f_\xi(W) = \frac{\exp\left(\frac{\xi}{2k_BT}\right)}{1+\exp\left(\frac{\xi-W}{k_BT}\right)}$ is a function of both W and ξ that has been proven to minimize the statistical error of ΔG estimation. The best estimate for ΔG is obtained from the intersection of the curves $y(\xi) = z_R(\xi) - z_F(\xi)$ and $y(\xi) = \xi$ (Figure 3.10D), which lies at $\Delta G_{Bennett}$ = 101.87 k_BT.

All of the above ΔG values agree with each other within their standard errors (estimated from Bootstrap method), as summarized in Table 3.4.

Table 3.4 Experimental ΔG values (unit: k_BT) from JE and CFT methods

$\Delta G_{JE,u}$	$\Delta G_{JE,f}$	ΔG_{CFT} intersection	ΔG_{CFT} linear fit	$\Delta G_{Bennett}$
101.96 ± 0.13	102.19 ± 0.32	101.87 ± 0.15	101.81 ± 0.12	101.87 ± 0.08

Thus, the free energy change at zero force is then calculated based on Eq. 3.5 and Figure 3.9. I obtained $\Delta G(0)$ = 50.94 ± 0.08 k_BT, which agrees well with the result from CF measurement, $\Delta G(0)$ = 52.44 ± 0.03 k_BT.

- Work distributions from different loading rates

Next, the effect of loading rate on the non-equilibrium work distributions was examined. Figure 3.11 shows the results from three different loading rates: 1.8 pN/sec, 7.6 pN/sec and 28.4 pN/sec. As loading rate is increased, the work values become more distributive (indicated by their standard deviations), and the hysteresis between unfolding and refolding ($\langle W_F \rangle$ - $\langle W_R \rangle$) gets progressively larger. However, the ΔG values, calculated from CFT, are similar for different loading rates.

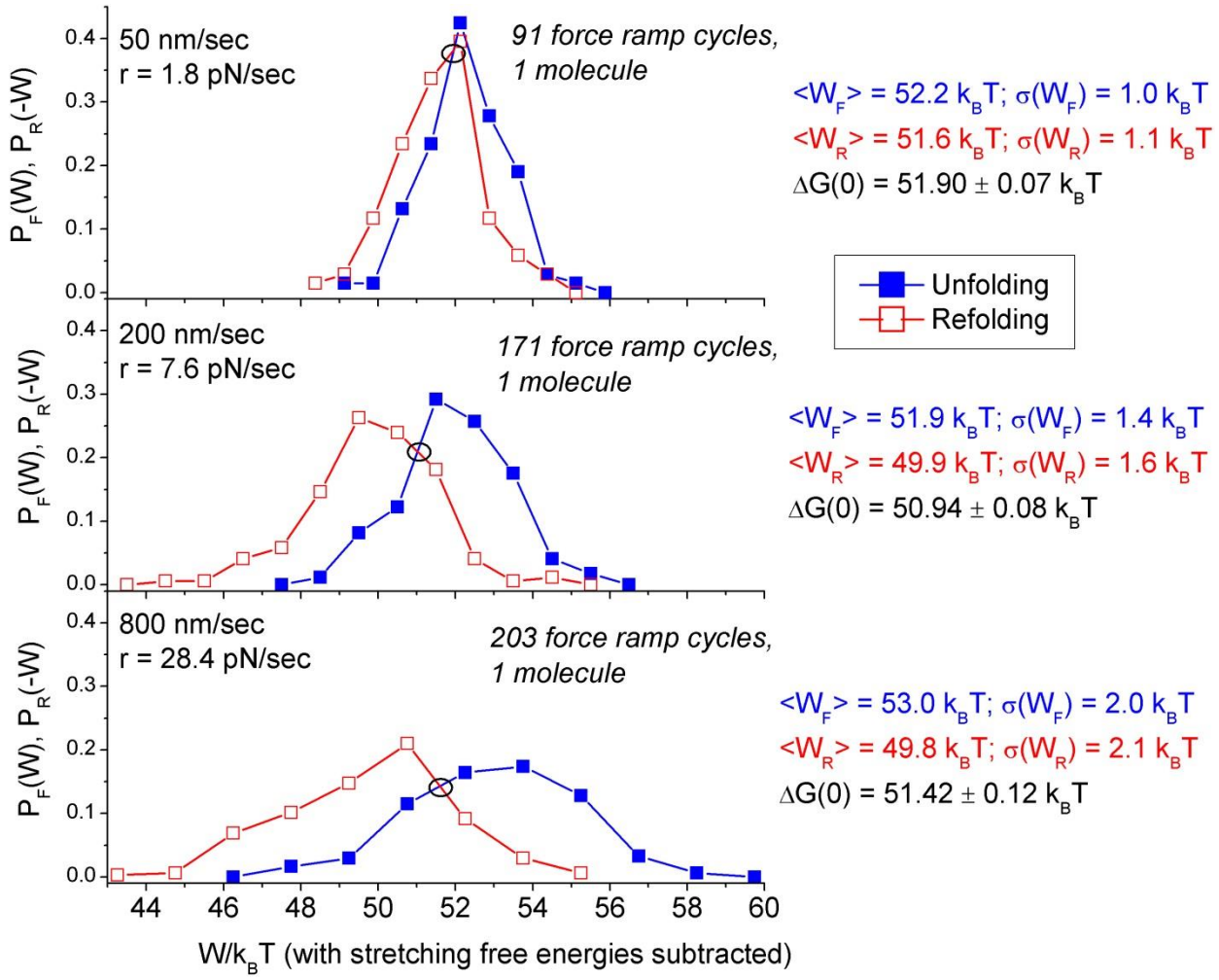


Figure 3.11 Unfolding and refolding work distributions at three different pulling speeds.

The mean and standard deviation of the work values are indicated next to each figure. The ΔG(0) values are calculated from CFT using the Bennett method.

3.3.6 Model-dependent fits of energy landscape parameters

Molecular unfolding under force can be viewed as crossing over a barrier on a free energy landscape. In the single-molecule experiments, the reaction coordinate is the molecular extension. The energetic parameters such as transition barrier positions and barrier heights can also be extracted from the force spectroscopy data. In the force ramp experiments, the molecular

unfolding can be described with the equation that relates the survival probability $P(F)$ [the probability of the molecule to remain folded at force F] to the unfolding rate constant $k_u(F)$,

$$\frac{dP(F)}{dt} = -k_u(F)P(F) \qquad \text{(Eq. 3.14)}$$

Since the applied force increases linearly with time, with $dF/dt = r$ defined as the loading rate, Eq. 3.14 becomes

$$\frac{dP(F)}{dF} = -\frac{k_u(F)}{r}P(F) \qquad \text{(Eq. 3.15)}$$

Using a proper model for $k_u(F)$, it is possible to obtain key features of the free energy landscape. Two well-accepted models are Bell's model and Dudko's model. Both the models assume that force linearly tilts the free energy landscape, which is an approximation that does not take into account the stretching free energy (see Eq. 3.5). However, both models can still fit the experimental data as long as the force range is narrow (1-3 pN in Figure 3.13) so that the change in stretching free energy is small.

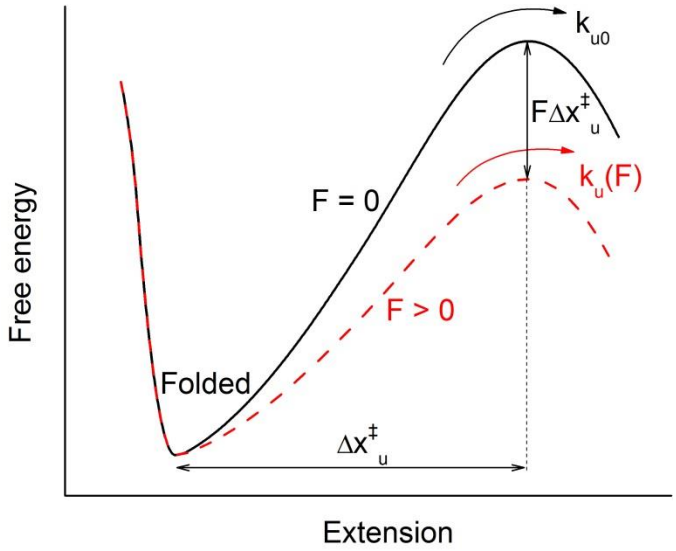


Figure 3.12 The effect of force on the free energy landscape according to Bell's model.

While the barrier position $\Delta x_u^{\ddagger}$ does not move with force, the barrier height is reduced by $F \cdot \Delta x_u^{\ddagger}$ upon the application of force. As a result, the unfolding rate, $k_u(F)$, is higher than that at zero force, k_{u0}.

- Bell's Model

In this model (154, 155), the effect of the force is simply to change the height of the barrier by an amount $F \cdot \Delta x_u^{\ddagger}$, where $\Delta x_u^{\ddagger}$ is the distance to the barrier from the initial state, illustrated in Figure 3.12. The Bell's model assumes that $\Delta x_u^{\ddagger}$ is independent of force. As a result, the unfolding rate varies exponentially with barrier height changes,

$$k_u(F) = k_{u0} \exp\left(\frac{F\Delta x_u^{\ddagger}}{k_B T}\right) \qquad \text{(Eq. 3.16)}$$

where k_{u0} is the unfolding rate at zero force, and $\Delta x_u^{\ddagger}$ is the distance from the folded state to the transition state. Combining Equation 3.16 with Equation 3.15 and solving the differential equation gives

$$P(F) = \exp\left(-\frac{k_{u0} k_B T}{r \Delta x_u^{\ddagger}}\left(\exp\left(\frac{F\Delta x_u^{\ddagger}}{k_B T}\right) - 1\right)\right) \qquad \text{(Eq. 3.17)}$$

By taking the derivative of P(F) respect to force F, the expression for the unfolding force distribution can be obtained as,

$$p_u(F) = \frac{k_{u0}}{r} \cdot \exp\left(\frac{F\Delta x_u^{\ddagger}}{k_B T}\right) \cdot \exp\left(-\frac{k_{u0} k_B T}{r \Delta x_u^{\ddagger}}\left(\exp\left(\frac{F\Delta x_u^{\ddagger}}{k_B T}\right) - 1\right)\right) \qquad \text{(Eq. 3.18)}$$

Eq. 3.18 was used to fit the experimental unfolding force distributions obtained at four loading rates, 1.8 pN/s, 7.6 pN/s, 14.7 pN/s, and 28.4 pN/s. As the loading rate gets higher, the average unfolding force increases, observed as 14.44 ± 0.04 pN, 14.88 ± 0.03 pN, 15.03 ± 0.04 pN, and 15.17 ± 0.04 pN respectively. Figure 3.13 shows the fitted curve and parameters k_{u0} and $\Delta x_u^{\ddagger}$ obtained from each distribution. Meanwhile, Eq. 3.16 was used to fit the unfolding rates at various constant forces, with the fitting parameters k_{u0} and $\Delta x_u^{\ddagger}$ as well (Figure 3.13E).

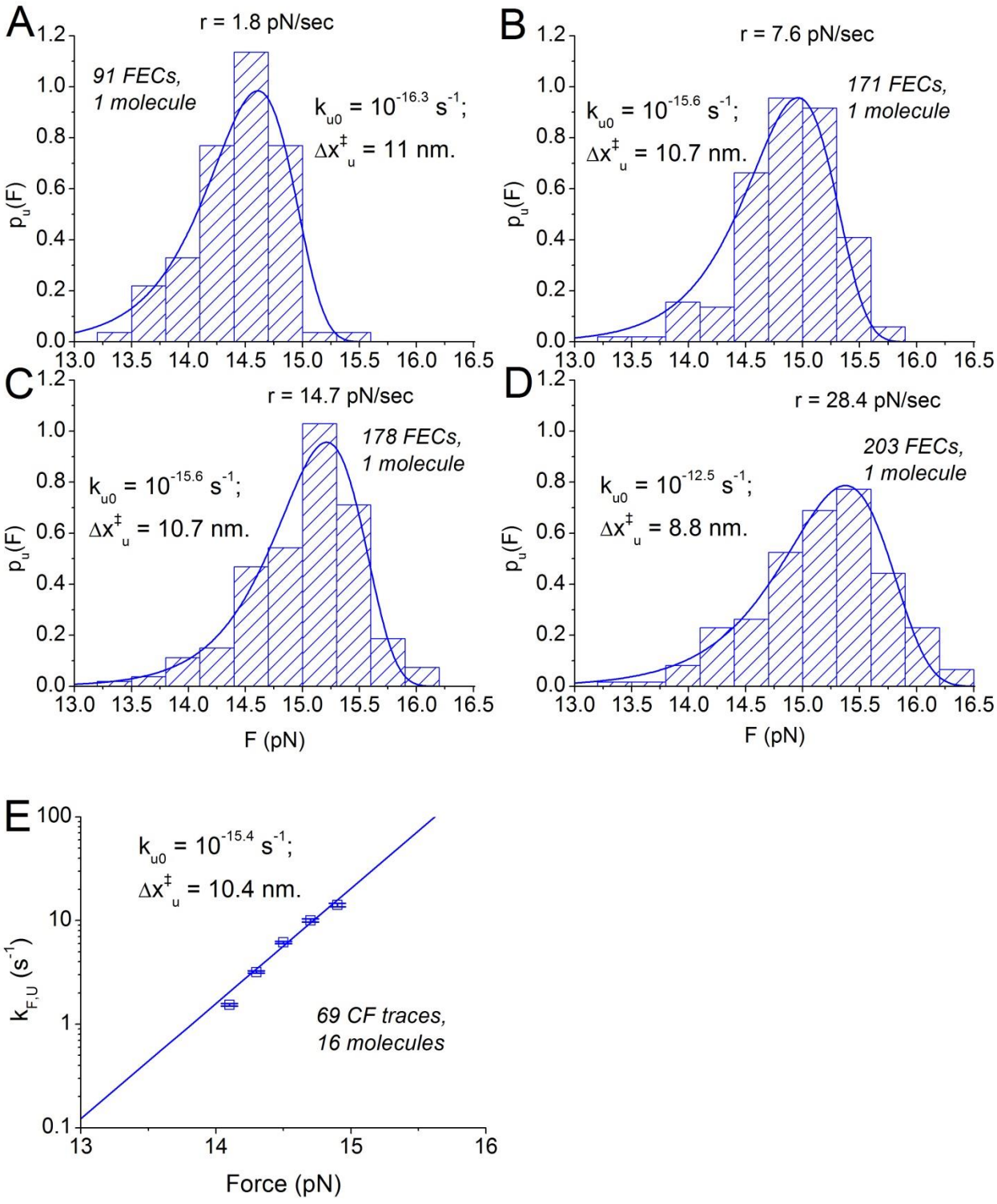


Figure 3.13 Bell's model fit to P5ab data.

(A)(B)(C)(D) Bell's model fit to P5ab unfolding force distributions $p_u(F)$ obtained by force ramps at various loading rates r. (E) Bell's model fit to unfolding rate $k_{F,U}(F)$ measured at constant forces for P5ab. The population size and the fitted parameters for each dataset are shown in each figure.

- Dudko's Model

Unlike Bell's model, the Dudko's model (156-158) considers the effect of force on both the barrier position and barrier height. Two specific shapes of energy landscape are investigated: the cusp shape $G_0(x) = \begin{cases} \Delta G_u^{\ddagger}\left(\frac{x}{\Delta x_u^{\ddagger}}\right)^2, & x < \Delta x_u^{\ddagger} \\ -\infty, & x \geq \Delta x_u^{\ddagger} \end{cases}$, and the linear-cubic shape $G_0(x) = \frac{3}{2}\Delta G_u^{\ddagger}\frac{x}{\Delta x_u^{\ddagger}} - 2\Delta G_u^{\ddagger}\left(\frac{x}{\Delta x_u^{\ddagger}}\right)^3$. In both landscapes, $\Delta x_u^{\ddagger}$ is the barrier position and $\Delta G_u^{\ddagger}$ is the barrier height, as shown in Figure 3.14A and 3.14C. Upon the application of force F, the landscape is tilted and becomes $G(x,F) = G_0(x) - Fx$. Based on the positions of minimum and maximum in $G(x,F)$, as illustrated in Figure 3.14B and 3.14D, the force dependent barrier position $\Delta x_u^{\ddagger}(F)$ and barrier height $\Delta G_u^{\ddagger}(F)$ are obtained.

For the cusp shape,

$$\Delta x_u^{\ddagger}(F) = \Delta x_u^{\ddagger}\left(1 - \frac{F\Delta x_u^{\ddagger}}{2\Delta G_u^{\ddagger}}\right) \qquad \text{(Eq. 3.19)}$$

$$\Delta G_u^{\ddagger}(F) = \Delta G_u^{\ddagger} - F\Delta x_u^{\ddagger} + \frac{\left(F\Delta x_u^{\ddagger}\right)^2}{4\Delta G_u^{\ddagger}} \qquad \text{(Eq. 3.20)}$$

For the linear-cubic shape,

$$\Delta x_u^{\ddagger}(F) = \Delta x_u^{\ddagger}\sqrt{1 - \frac{2F\Delta x_u^{\ddagger}}{3\Delta G_u^{\ddagger}}} \qquad \text{(Eq. 3.21)}$$

$$\Delta G_u^{\ddagger}(F) = \Delta G_u^{\ddagger}\left(1 - \frac{2F\Delta x_u^{\ddagger}}{3\Delta G_u^{\ddagger}}\right)^{3/2} \qquad \text{(Eq. 3.22)}$$

Thus, both $\Delta x_u^{\ddagger}(F)$ and $\Delta G_u^{\ddagger}(F)$ decrease as force increases. The Bell's model can be viewed as the extreme case when $\Delta G_u^{\ddagger}$ is infinitely high, resulting in a force-independent $\Delta x_u^{\ddagger}$ and $\Delta G_u^{\ddagger}(F) = \Delta G_u^{\ddagger} - F\Delta x_u^{\ddagger}$.

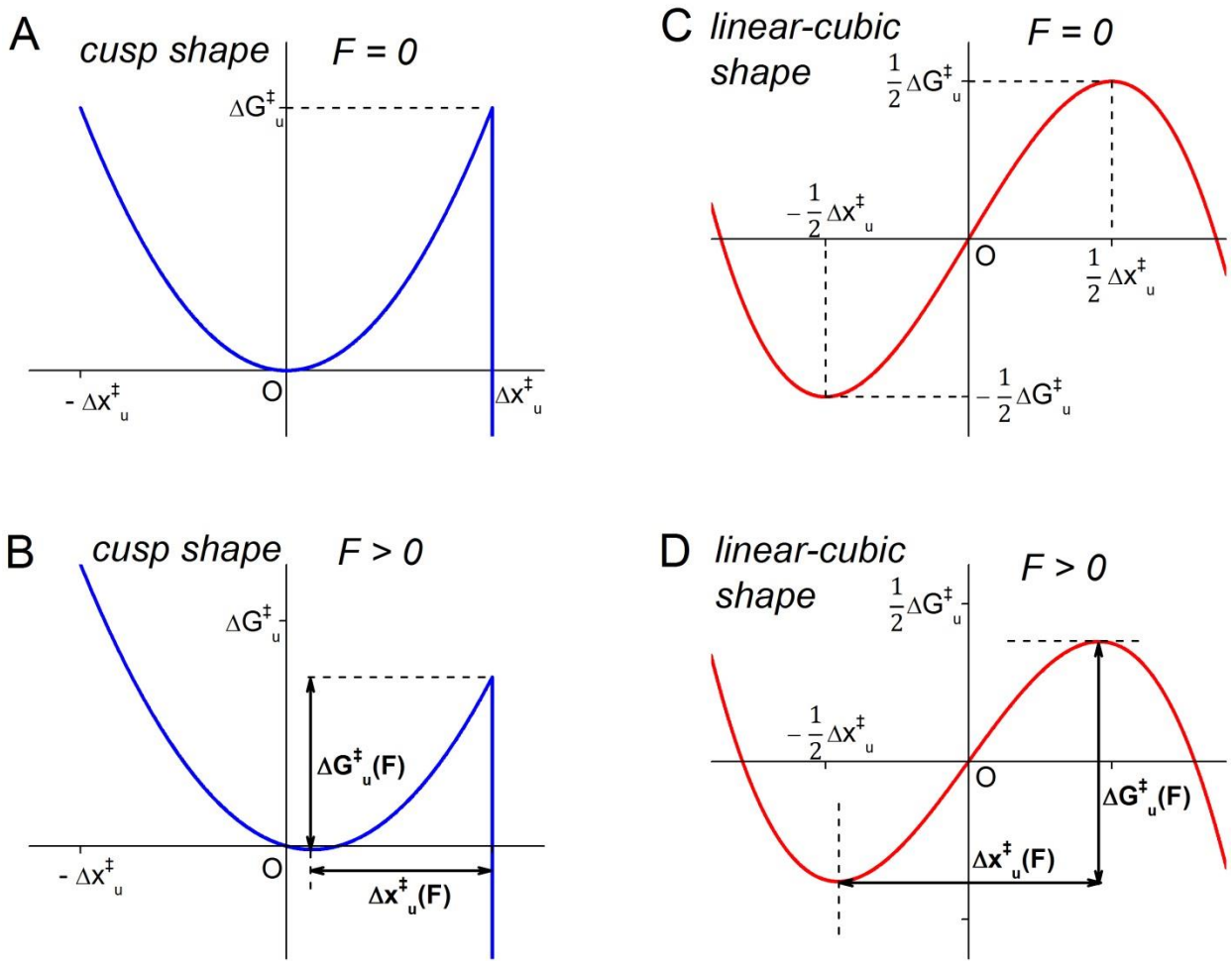


Figure 3.14 The cusp and linear-cubic landscape shapes and their changes with force.

The effect of force on (A)(B) cusp shape landscape and (C)(D) linear-cubic landscape. As force is applied, the landscapes are tilted by the term $-Fx$. Both $\Delta x_u^{\ddagger}(F)$ and $\Delta G_u^{\ddagger}(F)$ are reduced compared to those at zero force.

In order to derive the expression for the unfolding rate $k_u(F)$, the above results are combined with the Kramers' equation (159),

$$\frac{k_u(F)}{k_{u0}} = \frac{\int_{well} e^{-G_0(x)/k_BT}dx \int_{barrier} e^{G_0(x)/k_BT}dx}{\int_{well} e^{[-G_0(x)+Fx]/k_BT}dx \int_{barrier} e^{[G_0(x)-Fx]/k_BT}dx} \qquad \text{(Eq. 3.23)}$$

where k_{u0} is the unfolding rate at zero force and the integrals extend over the well and barrier regions respectively. Thus,

$$k_u(F) = k_{u0}\left(1-\frac{F\Delta x_u^{\ddagger}}{\Delta G_u^{\ddagger}}\upsilon\right)^{1/\nu-1}\exp\left\{\frac{\Delta G_u^{\ddagger}}{k_BT}\left[1-\left(1-\frac{F\Delta x_u^{\ddagger}}{\Delta G_u^{\ddagger}}\upsilon\right)^{1/\nu}\right]\right\} \quad \text{(Eq. 3.24)}$$

where ν parameterizes the shape of the landscape, with ν = 1/2 for a cusp shape, and ν = 2/3 for a linear-cubic shape. Combining Eq. 3.24 with Eq. 3.15, the unfolding force distribution is obtained as

$$p_u(F) = \frac{k_u(F)}{r}\cdot\exp\left\{\frac{k_{u0}k_BT}{\Delta x_u^{\ddagger}r}-\frac{k_u(F)k_BT}{\Delta x_u^{\ddagger}r}\left(1-\frac{F\Delta x_u^{\ddagger}}{\Delta G_u^{\ddagger}}\upsilon\right)^{1-1/\nu}\right\} \qquad \text{(Eq. 3.25)}$$

Using the experimental datasets of $p_u(F)$ and $k_u(F)$, three parameters, k_{u0}, $\Delta x_u^{\ddagger}$ and $\Delta G_u^{\ddagger}$, were fitted from Eq. 3.25 and Eq. 3.24 respectively (Figure 3.15 and Figure 3.16). Based on these results, the barrier position and height at the equilibrium force 14.5 pN were calculated using Eq. 3.19-3.22.

The fitting results of the three landscape models (Bell, cusp, and linear-cubic) are summarized in Table 3.5. All three models gave similar $\Delta x_u^{\ddagger}$ values of 10 nm at F = 14.5 pN.

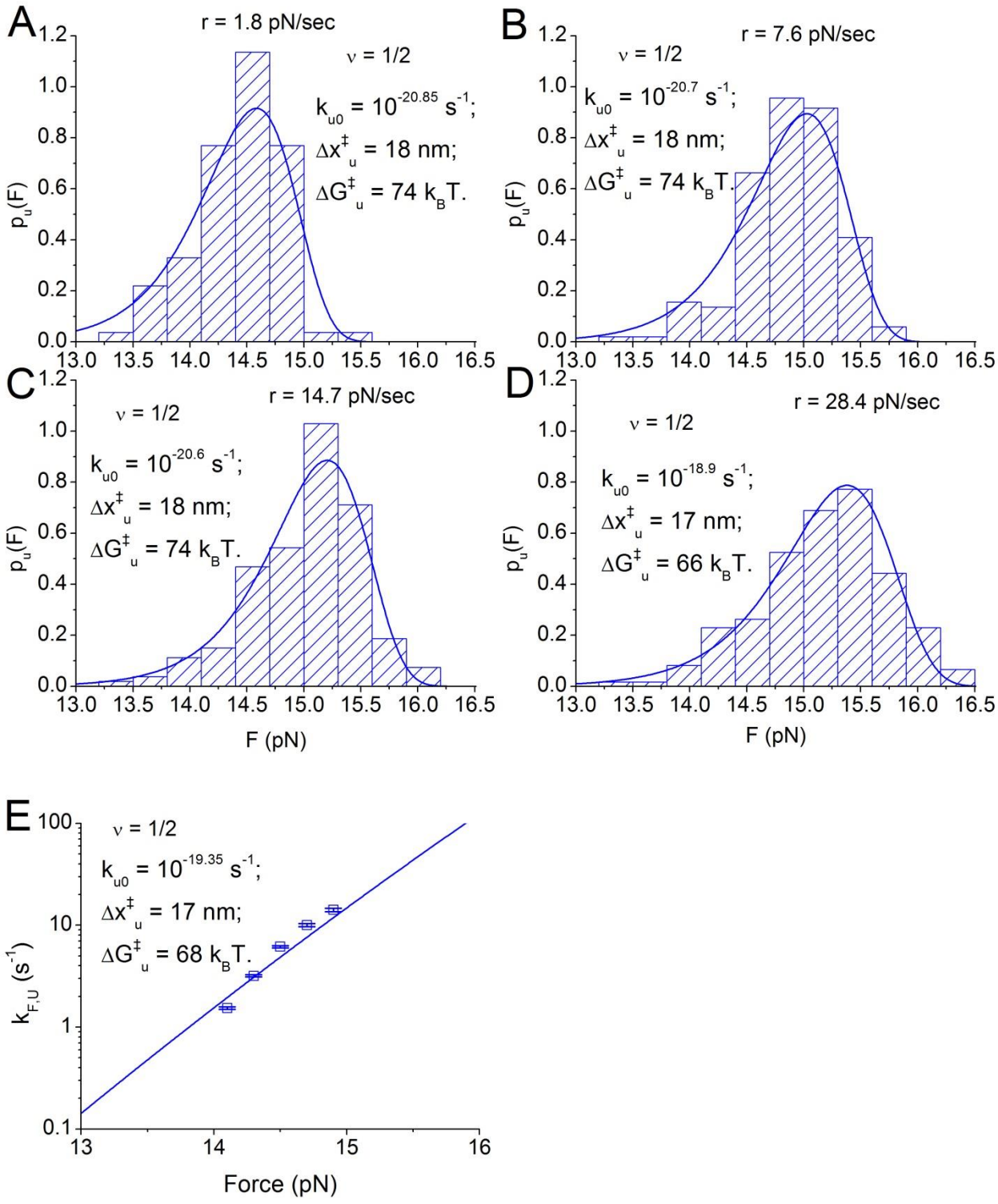


Figure 3.15 Fit of Dudko's cusp-shape model to P5ab data.

Dudko's model (ν=1/2; cusp-shape landscape) fit to the same $p_u(F)$ and $k_{F,U}(F)$ datasets as in Figure 3.13. The fitted parameters are shown in each figure.

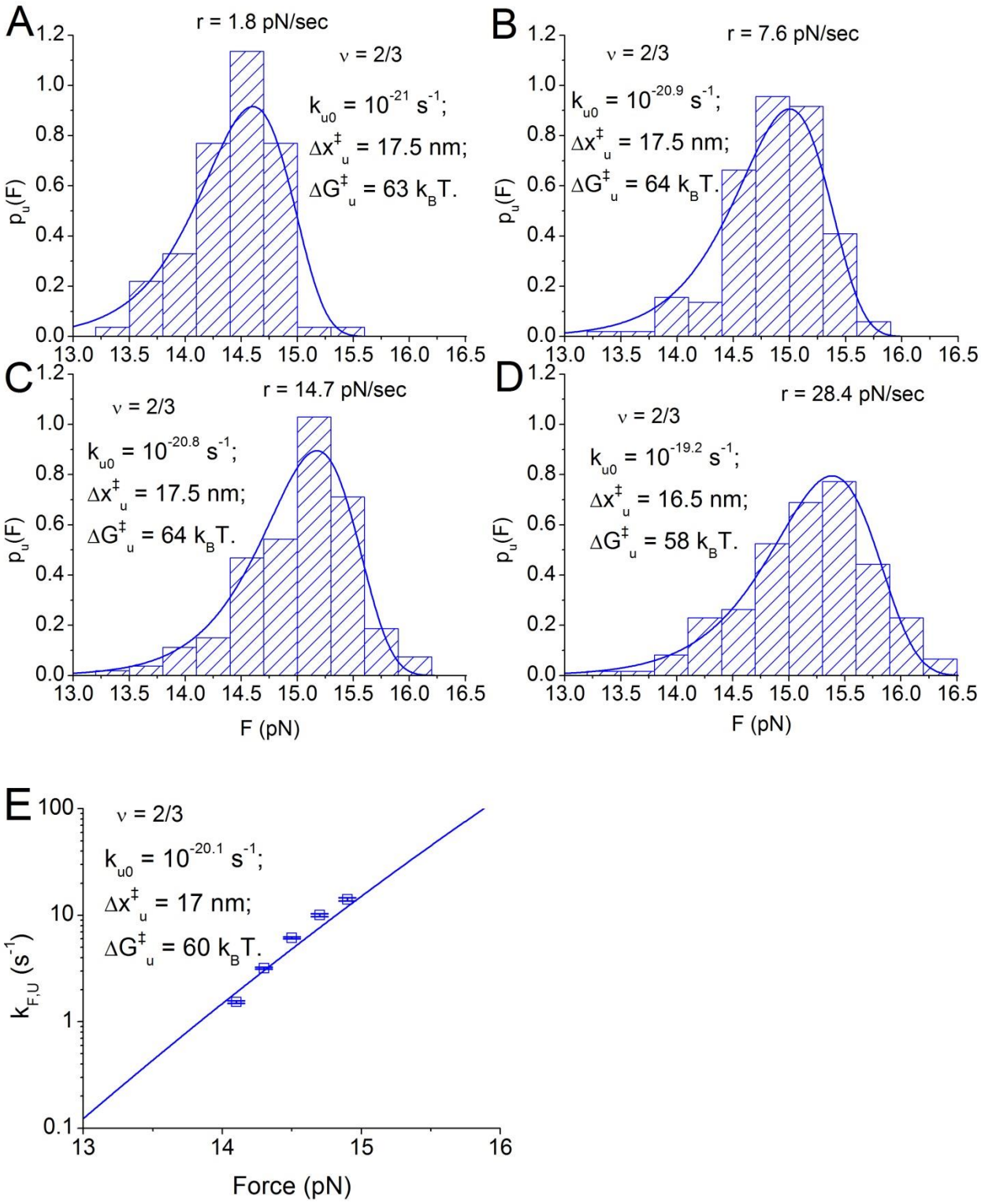


Figure 3.16 Fit of Dudko's linear-cubic model to P5ab data.

Dudko's model (ν=2/3; linear-cubic landscape) fit to the same $p_u(F)$ and $k_{F,U}(F)$ datasets as in Figure 3.13. The fitted parameters are shown in each figure.

Table 3.5 Model fitted free energy landscape parameters for P5ab unfolding

Energy landscape model	k_{u0} (s^{-1})	$\Delta x^{\ddagger}_u$ (nm)	$\Delta G^{\ddagger}_u$ (k_BT)	$\Delta x^{\ddagger}_u$ (F=14.5pN) (nm)	$\Delta G^{\ddagger}_u$ (F=14.5pN) (k_BT)
Bell	$10^{-15.1 \pm 1.5}$	10.3 ± 0.9	-	10.3 ± 0.9	-
Dudko, ν=1/2	$10^{-20.1 \pm 0.9}$	17.6 ± 0.5	71.2 ± 3.9	9.9 ± 0.4	22.7 ± 3.3
Dudko, ν=2/3	$10^{-20.4 \pm 0.8}$	17.2 ± 0.4	61.8 ± 2.7	10.1 ± 0.4	12.6 ± 2.3

Each fitted parameter is shown as the average ± standard deviation of the multiple fits for each model (Figure 3.13, 3.15, and 3.16).

3.3.7 Free energy landscape from equilibrium sampling

The complete free energy landscape profile can be independently calculated in a model-free manner. In principle, the free energy landscape $\Delta G(x, F)$, as a function of extension and force, is calculated by the Boltzmann relation

$$\Delta G(x,F) = -k_B T \ln\big(P(x,F)\big) \qquad \text{(Eq. 3.26)}$$

where $P(x, F)$ is the extension probability distribution of the RNA's extension x at force F. The extension probability distribution can be measured from equilibrium sampling in CF measurements. However, in the optical trapping assay, the exact RNA extension changes are blurred by the thermal fluctuations of beads and handles. This can be mathematically represented as the following convolution relation:

$$P^0(x,F) = PSF(x,F) \otimes P(x,F) \qquad \text{(Eq. 3.27)}$$

where $\otimes$ denotes the convolution operator. $P^0(x, F)$ is the experimentally measured extension distribution, $P(x, F)$ is the true RNA extension distribution, and $PSF(x, F)$ is the point-spread function characterizing the noise from the beads and handles. In order to recover $P(x, F)$, $P^0(x, F)$ needs to be deconvoluted with $PSF(x, F)$. Here $PSF(x, F)$ was measured from a control construct consisting of RNA/DNA handles only (Figure 3.17).

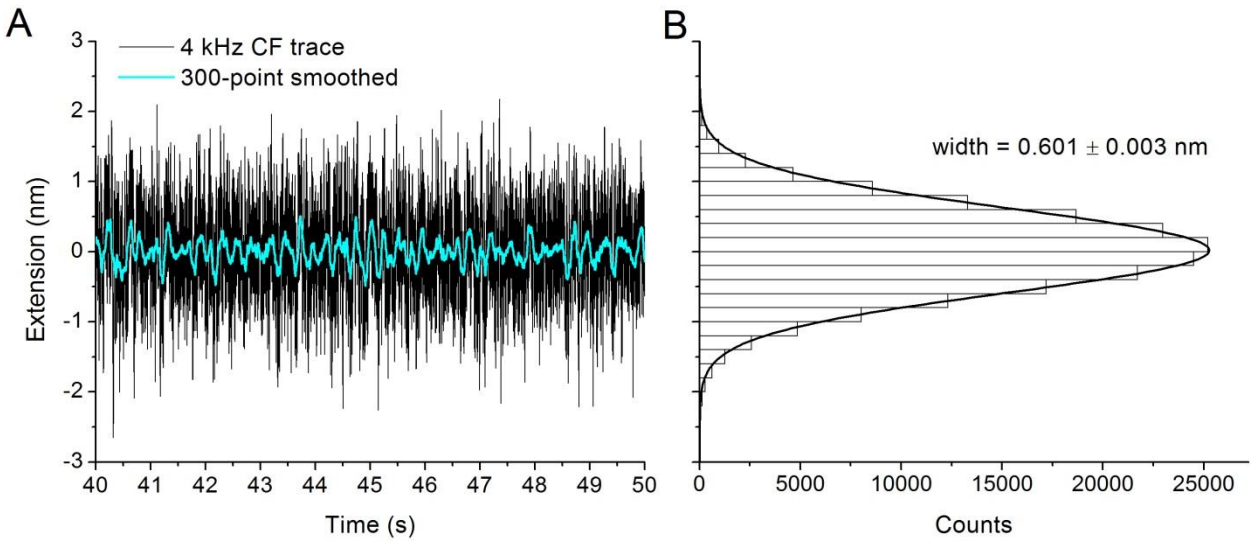


Figure 3.17 Measurement of the point-spread function (PSF) from a handle-only construct.

(A) A typical CF trace at 14.5 pN for a molecule consisting of RNA/DNA handles only. The data was collected for 1 minute at 4 kHz (black), with the 300-point smoothed trace shown in cyan. (B) The corresponding extension distribution is fit by a Gaussian with a width of 0.601 ± 0.003 nm. Measurement of three handle-only molecules gives a Gaussian width of 0.64 ± 0.05 nm. Such a Gaussian form of PSF is used for the deconvolution procedure.

The deconvolution procedure was done by the non-linear constrained iterative method (160), as used in several previous single-molecule studies (116, 117). The probability distribution at the (k+1)th iteration was given by

$$P^{k+1}(x) = P^k(x) + r\left(P^k(x)\right) \times [P^0(x) - PSF(x) \otimes P^k(x)] \qquad \text{(Eq. 3.28)}$$

where $r(P^k(x)) = r_0(1 - 2|P^k(x) - 0.5|)$ is the relaxation function which constrains the solution to remain within the physical boundary $0 \leq P^k(x) \leq 1$. Starting with the measured probability distribution $P^0(x)$, I used r_0=1 with n = 10000 iterations. To test the accuracy of the deconvolution procedure, the residual error was calculated as

$$R(x) = P^0(x) - PSF(x) \otimes P^n(x) \qquad \text{(Eq. 3.29)}$$

Once the final solution $P^n(x)$ was obtained (Figure 3.18A), its associated energy landscape $\Delta G^n(x)$ was calculated according to Eq. 3.26, as shown in Figure 3.18B.

At 14.5 pN, the deconvoluted energy landscape of P5ab shows two major populations F and U, with an energy barrier at $\Delta x^{\ddagger}_{F,U}$ = 8.93 ± 0.03 nm (Figure 3.18B), which agrees well with the model fitted barrier positions (Table 3.5). The measured barrier height, $\Delta G^{\ddagger}_{F,U}$ = 6.1 ± 0.3 k_BT, shows some deviation from the barrier height of 12.6 k_BT for the linear-cubic model fit; it differs much more from the barrier height of 22.7 k_BT for the cusp case. Meanwhile, the intermediate states I_1, I and I_2 exhibit as humps around the barrier top, indicating they are metastable compared to F and U.

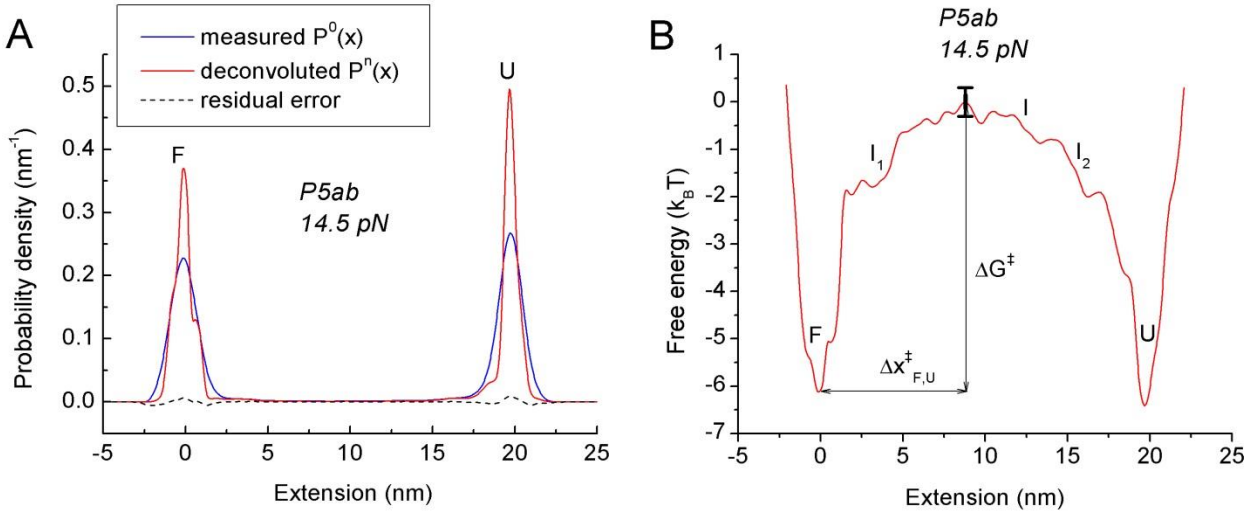


Figure 3.18 Free energy landscape of P5ab from deconvolution of probability distributions.

(A) Deconvolution of the extension probability density for P5ab measured in 4 kHz CF for 20 seconds. The experimentally observed probability (blue) was deconvoluted with the point-spread function (Figure 3.17). The resulting deconvoluted probability is shown in red and the residual error is plotted as the black dashed line. (B) The free energy landscape of P5ab calculated from the deconvoluted probability in (A). The error bar indicates the error in the barrier height due to the standard deviation in the Gaussian width for PSF.

3.3.8 Mutational analysis

Based on the P5ab sequence, I designed several mutant hairpins in order to prove the folding pathway and further examine the effect of loops and bulges on hairpin folding kinetics. The sequences of the mutants m1, m2, and m3 are shown in Figure 3.19. This work is in progress. The folding study of only m1 is shown below.

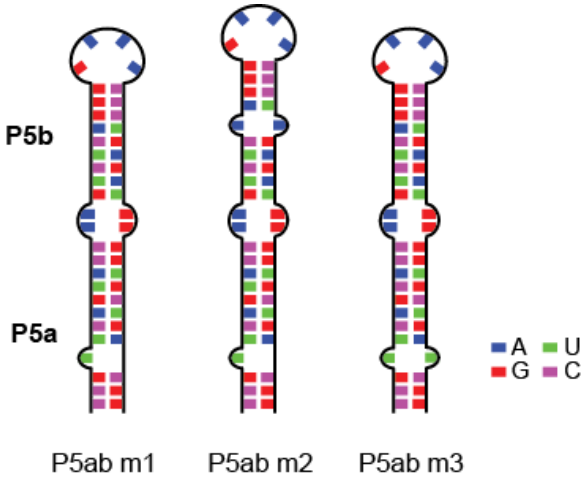


Figure 3.19 Sequences of P5ab mutant hairpins

In mutant m1, two G·U wobble pairs in the P5b stem are mutated into two stronger G-C base pairs. The expectation is that the intermediate state I will become more stable. The unfolding FECs for m1 showed two distinct steps (Figure 3.20): step *a* at 15.9 ± 0.7 pN with an extension of 12.2 ± 0.7 nm and step *b* at 18.9 ± 1.0 pN with another extension of 7.9 ± 0.5 nm. These extension values indicate that the intermediate I has now separated the unfolding transitions into two sub-steps: F → I (i.e. opening of P5a) and I → U (i.e. opening of P5b). Thus, the intermediate I is proven in mutant m1.

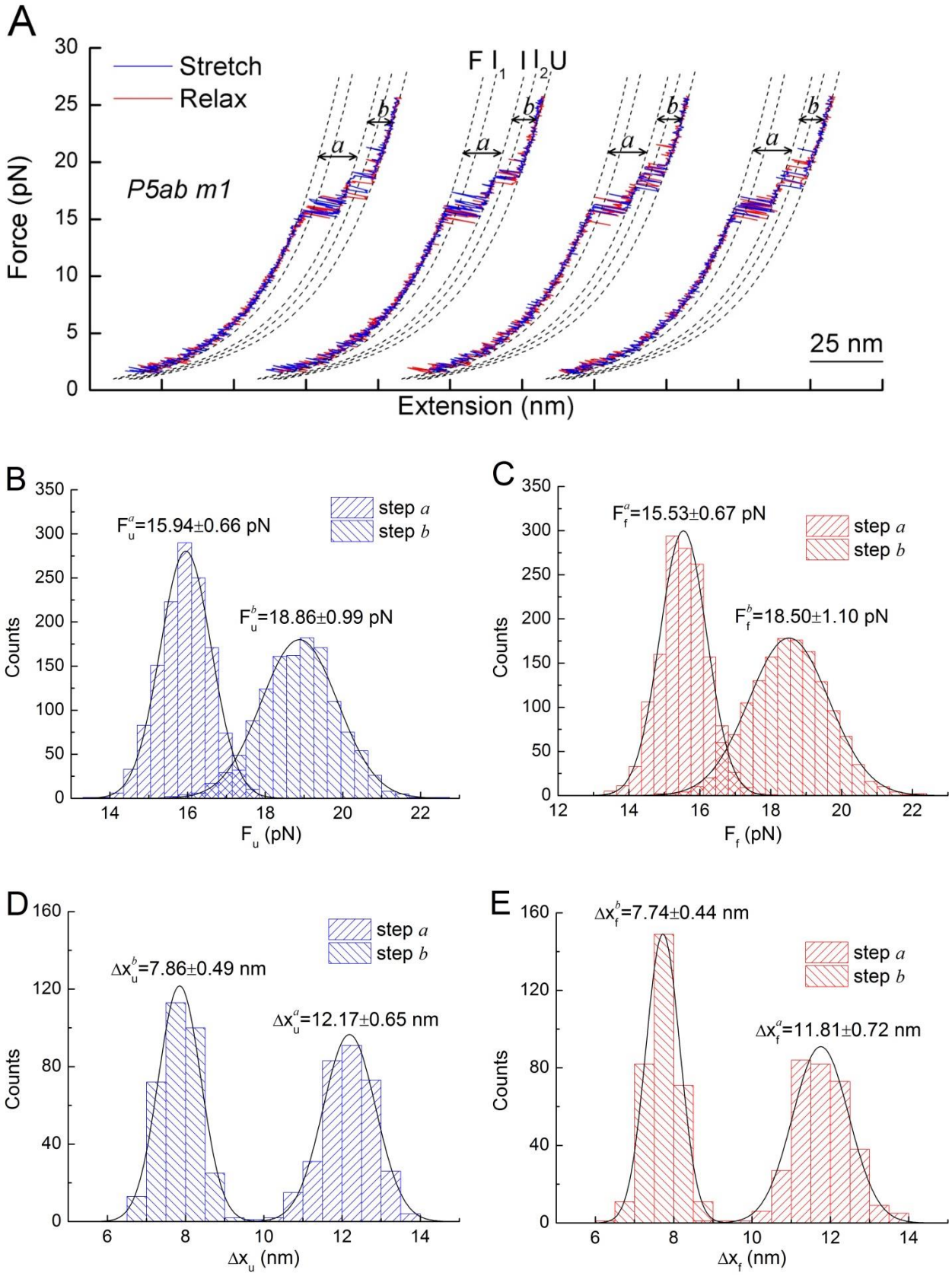

A
Stretch
Relax
P5ab m1
F I1 I1 I2 U
a
b
Force (pN)
Extension (nm)
25 nm
B
step a
step b
F_u^a=15.94±0.66 pN
F_u^b=18.86±0.99 pN
Counts
F_u (pN)
C
F_f^a=15.53±0.67 pN
F_f^b=18.50±1.10 pN
F_f (pN)
D
Δx_u^b=7.86±0.49 nm
Δx_u^a=12.17±0.65 nm
Δx_u (nm)
E
Δx_f^b=7.74±0.44 nm
Δx_f^a=11.81±0.72 nm
Δx_f (nm)

Figure 3.20 Unfolding and refolding characteristics of P5ab m1 in force ramp experiment.

(A) Four stretch-relax cycles on a P5ab m1 molecule. The unfolding and refolding exhibit two-step transitions (*a* and *b*). The dashed lines are the WLC fits to the five states: F, I_1, I, I_2, and U.

(B) Distribution of the unfolding forces F_u for P5ab m1. The data is representative of 330 pulls on 11 molecules. The bin size is 0.35 pN. The steps *a* and *b* correspond to the two-step transitions at lower and higher forces respectively. The means and standard deviations are indicated on the figure.

(C) Distribution of the refolding forces F_f for P5ab m1 from the same experiments.

(D) Distribution of the unfolding distance Δx_u from the same experiments. The bin size is 0.5 nm. The means and standard deviations are indicated on the figure.

(E) Distribution of the refolding distance Δx_f from the same experiments.

I then evaluated the folding free energies for each of the two unfolding steps in mutant m1. By selecting proper integration boundaries for the work done (Figure 3.21A), I obtained the work distributions from 126 pull-relax cycles at loading rate of 7.6 pN/sec (Figure 3.21B). The equilibrium free energies were then calculated using CFT Bennett method. The two-step transitions show $\Delta G_{F,I}(0) = 35.38 \pm 0.10\ k_BT$, $\Delta G_{I,U}(0) = 30.76 \pm 0.12\ k_BT$, and the total free energy change is $\Delta G_{F,U}(0) = 66.13 \pm 0.13\ k_BT$. Compared to $\Delta G_{F,U}(0) = 50.94\ k_BT$ for P5ab, m1 shows increased stability, as expected from the mutations.

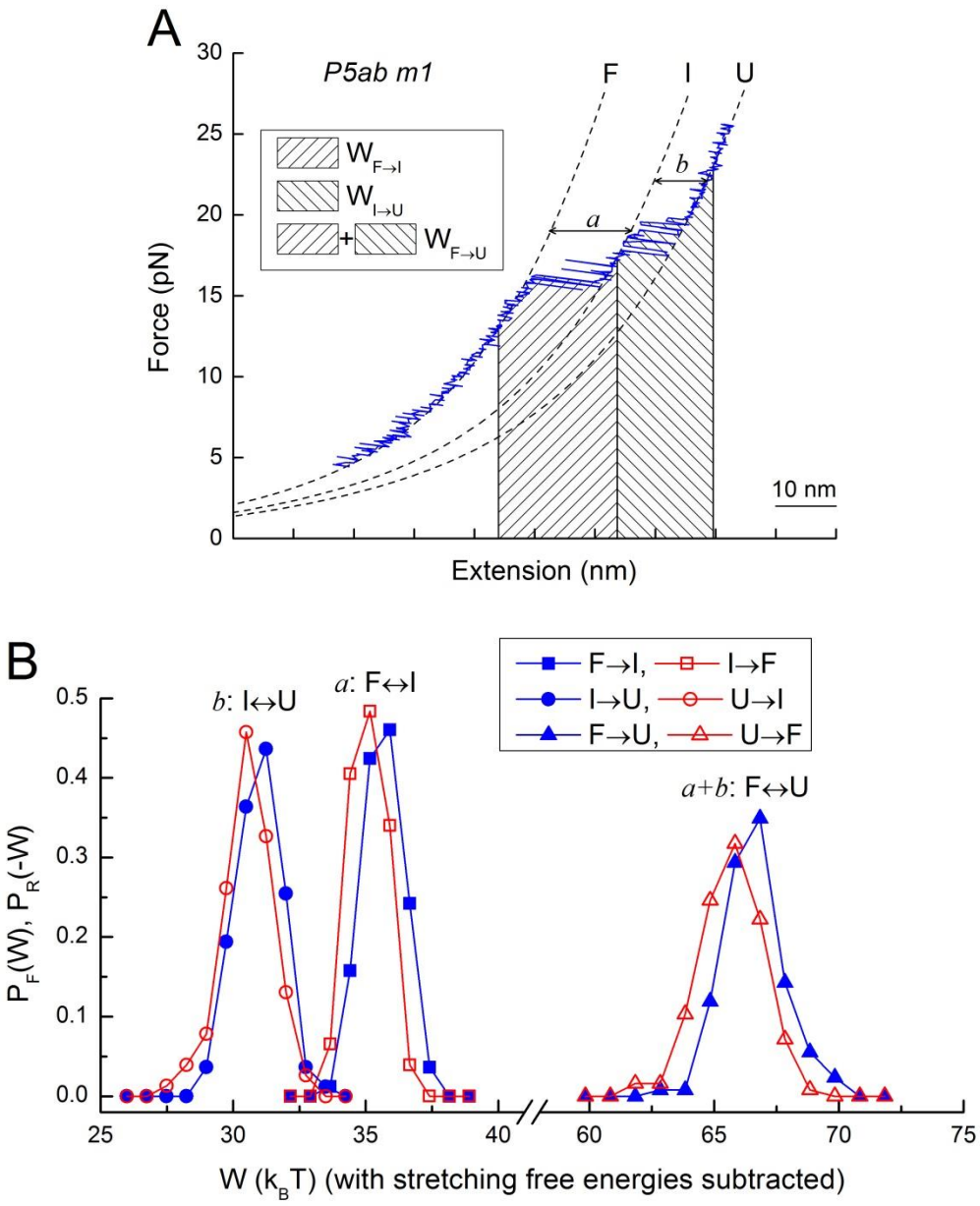


Figure 3.21 Determination of P5ab m1 folding free energies from non-equilibrium work done.

(A) Illustration of work done calculation on an unfolding FEC for P5ab m1. The work done for the two steps, *a* and *b*, is individually evaluated, and their sum represents the total work done from folded state to the unfolded state.

(B) Unfolding (blue) and refolding (red) work distributions from 126 pull-relax cycles on one P5ab m1 molecule. The loading rate is 7.6 pN/sec. The contributions from stretching handles and single-stranded RNA are already corrected from the work values. Thus, the extracted equilibrium free energies using CFT represent the $\Delta G(0)$ values for the transitions F,I; I,U and F,U respectively.

3.4 Discussion

3.4.1 Biological relevance of the hairpin folding intermediate states

Within in the cell, RNA folds spontaneously as it is synthesized, but it requires energy to unfold. Molecular motors such as RNA helicases, ribosomes and RNA-dependent RNA polymerases need to unwind RNA structures before they can perform their biological functions. Typically, these enzymes use chemical energies to mechanically separate the strands of double-stranded RNAs. Thus the hairpin intermediate states identified in this study can also occur *in vivo* as partially melted conformations under localized forces and be utilized by molecular motors to make genetic decisions.

While the hairpin defects such as bulges and internal loops often contribute to the folding intermediates, these unpaired regions also play other crucial biological roles. First, they can form tertiary contacts, i.e. pseudoknots, with other parts of the RNA, which is important for the global fold and function of large RNAs. Second, these deformable imperfections can affect RNA-protein interactions. For example, the bulges and internal loops enhance the cleavage efficiency in Drosha processing of pri-miRNAs (161). For the activation of protein kinase (PKR), bulges within dsRNA decrease the activation of PKR compared to perfectly paired dsRNA (162).

3.4.2 Comparison of energy barrier information from multiple methods

Although the barrier position $\Delta x^{\ddagger}$ shows a similar value (10 nm) from the landscape models and the deconvoluted landscape, the barrier height $\Delta G^{\ddagger}$ at 14.5 pN exhibits larger differences among them: 6.1 k_BT from deconvoluted landscape, 12.6 k_BT from linear-cubic model, and 22.7 k_BT from cusp model. These three results can be further scrutinized from the perspective of the Kramer's theory $k(F) = A_0 \exp\left(-\frac{\Delta G^{\ddagger}(F)}{k_B T}\right)$, where A_0 is the pre-exponential factor that depends mainly upon the viscous drag on the molecule and beads, as well as the shape of the landscape (159, 163). Combining the $\Delta G^{\ddagger}$ results with the measured rate constant $k_{F,U}$ = 6.15 sec^{-1} at 14.5 pN (Table 3.2), the pre-exponential factor A_0 can be determined. I obtained A_0 values of 2.7×10^3 s^{-1}, 1.8×10^6 s^{-1} and 4.4×10^{10} s^{-1} for the above three cases respectively. Since A_0 has been

estimated to fall into the range 10^3~10^7 sec^{-1} by other single-molecule experiments (117, 125, 164), the $\Delta G^{\ddagger}$ results from the deconvolution (6.1 k_BT) and linear-cubic model (12.6 k_BT) make more sense than the cusp model result (22.7 k_BT).

3.5 Conclusion

Here using single-molecule force spectroscopy, I have studied the kinetics and folding pathway of P5ab RNA hairpin, which involve multiple intermediate states I_1, I, and I_2. The intermediate state I is also proven in P5ab m1 mutant. More mutational studies will further probe the hairpin folding pathway and the effect of loops and bulges on hairpin folding.

I have calculated the free energy of folding for P5ab hairpin from mechanical work done on the molecule. Unlike the free energy measurements in bulk thermal melting experiment, which often assumes the denaturation of RNA as an equilibrium process, here the unfolding and refolding of RNA in force ramp is rigorously treated as a non-equilibrium process using JE and CFT and the effect of loading rate on free energy estimation is examined. In the future, it will be interesting to extend the applicability of these fluctuation theorems to study the thermodynamics of the intermediate states, for example by using the recently developed extended JE and CFT (165, 166) which allow the initial or final state to be equilibrium between multiple states, instead of one stable state.

I have also investigated the energy barrier in P5ab folding. The barrier crossing mechanism is found in many physical and chemical processes, such as chemical reactions (167-169), fracture of solids (170), antibody-antigen binding (171, 172), adhesion between cells (154, 173), and folding of biomolecules (174, 175). Here using the Bell's and Dudko's models, as well as the equilibrium sampling method, the barrier position and barrier height are determined. An important future direction is to extract the energy barrier information in more complex RNA molecules, such as P5ab m1 mutant, which exhibits a two-step unfolding, implying two barriers separating the folded state and the unfolded state. While the kinetic rate theories for energy landscapes involving intermediates have been recently developed (176, 177), a thorough test by experimental data remains to be done.

Chapter 4 Single-Molecule Investigation of Ligand-dependent Folding of preQ1 Riboswitch Aptamer

4.1 Introduction

Riboswitches are natural biosensors present in the 5' untranslated region of mRNA elements that bind small metabolites to control gene expression (6-9). The metabolites for which a riboswitch has been discovered range from vitamins (178-182), nucleobases (183-188), amino acids (189-195), sugar (196), to metal ions (197, 198). A typical riboswitch consists of a metabolite-binding aptamer domain and an adjacent expression platform. Specific binding of a metabolite to the aptamer domain leads to structural rearrangements in the expression platform, which turns on or off the expression of downstream genes.

In bacteria, the class I preQ1 riboswitch (hereafter referred to as preQ1 riboswitch) is the smallest riboswitch known to date, comprising ~35 nucleotides in the aptamer domain (186). It regulates the gene expression involved in biosynthesis of queuosine, a hypermodified nucleoside found in the anticodon wobble position of certain transfer RNAs. The riboswitch binds to preQ1 (7-aminomethyl-7-deazaguanine), which is an intermediate in queuosine biosynthesis pathway.

X-ray crystallography and NMR spectroscopy have determined the structures of the preQ1 riboswitch aptamers from three different species, *Bacillus subtilis* (199, 200), *Fusobacteria nucleatum* (201, 202) and *Thermoanaerobacter tengcongensis* (203-205). In the preQ1-bound state, the aptamers from all three species adopt an H-type pseudoknot structure. Figure 4.1B and 4.1C show the NMR structure of the preQ1 bound aptamer from *Bacillus subtilis* (200). This pseudoknot structure contains two stems (5 base pair P1 and 3 base pair P2) and three loops (3 nt L1, 6 nt L2 and 8 nt L3). The preQ1 binding pocket is stabilized by the preQ1 base quadruple C18-preQ1-A31-U7, with a base triple C32-G12-A17 above and a base quadruple A29-G6-C19-A30 below (Figure 4.1D and E). The loop L3 also establishes tertiary contacts with the P1 stem via A24-G3 and A26-A4 hydrogen bonds. In the absence of ligand, however, the *Bacillus subtilis* and aptamer forms a stem-loop structure with an unorganized 3' adenine-rich tail (200-202) (Figure 4.1A).

The dynamics of preQ1-aptamer binding has been studied for the species *Fusobacteria nucleatum* and *Thermoanaerobacter tengcongensis* using fluorescence spectroscopy (202) and surface plasmon resonance (203) respectively. The measured binding affinity and ligand on and off rates are shown in Table 4.1. While the binding affinity for *Bacillus subtilis* preQ1 riboswitch has been determined by in-line probing assays (K_d = 50 nM) (186), its binding kinetics has not been examined. Moreover, the folding dynamics of the secondary and tertiary structures in the preQ1 aptamer has not been characterized in real time. To that end, I studied the ligand-dependent folding of the preQ1 riboswitch from *Bacillus subtilis* using optical tweezers force spectroscopy. Using the mechanical unfolding approach along with mutagenesis method, I addressed the following aims:

Aim 1: What is the folding pathway of the pseudoknot in the aptamer?

Aim 2: What are the kinetics and energetics of ligand binding in the aptamer folding?

Table 4.1 Binding affinity and kinetics for preQ1 riboswitch aptamers from different species

Species	K_d (nM)	k_{on} ($M^{-1}\cdot s^{-1}$)	k_{off} (s^{-1})
Fusobacteria nucleatum	283	6.02×10^5	0.17
Thermoanaerobacter tengcongensis	2.05	7.77×10^4	0.00015

The above results are from references (202, 203).

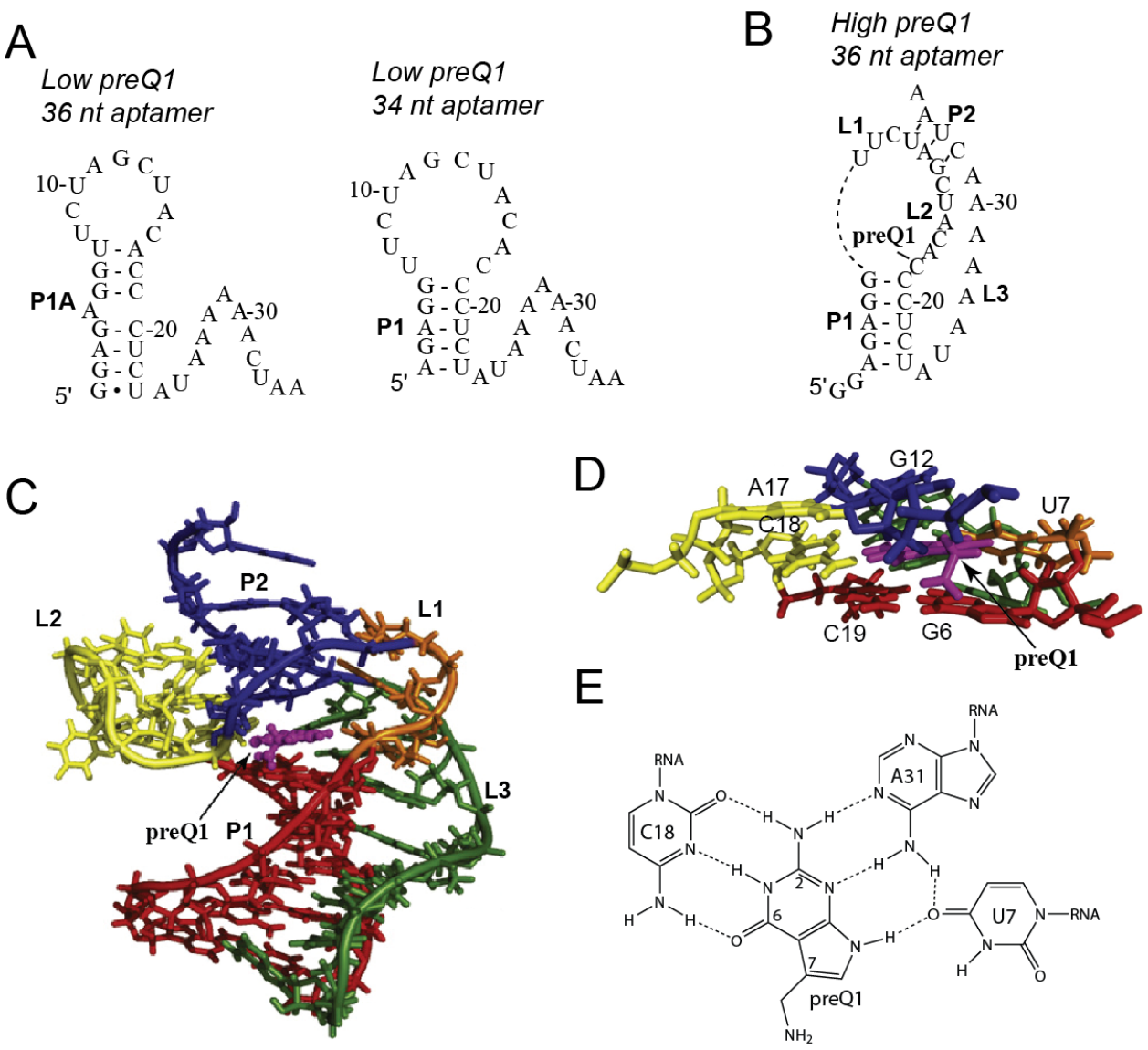


Figure 4.1 Structures of the preQ1 aptamer from *Bacillus subtilis* in the absence and presence of preQ1.

(A) Structure of the preQ1 aptamer in low preQ1 concentration. Using a 36-nt aptamer construct, NMR detected the P1A hairpin structure. However, using a 34-nt construct (with 2 fewer G's at the 5' end), the aptamer structure becomes the P1 hairpin. In our single-molecule study, we used a 35-nt construct with only 1 G at the 5' end. The numbering of the nucleotide sequence is based on our 35-nt construct.

(B) The pseudoknot structure formed in the 36-nt aptamer at high preQ1 concentration. The 34-nt aptamer forms a similar pseudoknot structure at the same condition.

(C) Stick representation of the preQ1 bound pseudoknot structure for the 36-nt aptamer, as determined by NMR. (D) Stick representation of the three layers of hydrogen bonding in the preQ1 binding pocket. The middle layer of hydrogen bond network, which involves the ligand preQ1, is shown in (E).

This figure is adapted from reference (200).

4.2 Materials and Methods

4.2.1 RNA constructs and reagents

The RNA constructs (preQ1 aptamer and mutants m1-m4) for single molecule studies were generated as described in Chapter 2. The ligand preQ1 was generously provided by Dr. Marc Greenberg lab in John Hopkins University. A stock solution of 2 mM preQ1 was made in H_2O. The ligand 2,6-diaminopurine (DAP) was purchased from Sigma Aldrich. A stock solution of 10 mM DAP was made in H_2O. Similar to Chapter 3, Tris buffer and NaCl were purchased from Sigma-Aldrich, and $MgCl_2$ was purchased from EMD. The beads were purchased from Spherotech.

4.2.2 Piecewise-constant analysis of constant-force traces

In order to study distinct states in the extension vs. time traces (Figure 4.7), I performed the piecewise-constant analysis that has been developed by an undergraduate researcher Ben Plaut in our lab. The algorithm consists of the following two steps:

(i) The background noise of the trajectory was reduced by applying the appropriate filter developed by Chung and Kennedy (206). This filter has the advantage of preserving the edges of abrupt transitions. The following parameters were used: windows sizes of 2, 4, 8, 16, 32 points for calculating the forward and backward predictors, M=10 points, weighting factor p=100, and the number of iterations is 3.

(ii) Transitions are assigned to the smoothed trace if the extension change is greater than a cutoff distance d = 0.8 nm. Each dwell time was assigned a constant extension value, calculated as the mean extension in that segment.

4.3 Results

4.3.1 Ligand-induced conformational changes of the aptamer

At first, I examined the unfolding and refolding characteristics of the preQ1 aptamer in the absence and presence of the ligand. In the optical trapping setup (see Figure 2.2 in Chapter 2), the 35-nucleotide preQ1 aptamer RNA is flanked by two DNA/RNA hybrid handles (each with ~550 base pairs) and held between a micropipette bead and a trapped bead. Experiments were carried out at 25 °C in a physiological buffer condition containing 10 mM Tris, pH 7.5, 250 mM NaCl, and 3 mM $MgCl_2$, with varying ligand concentrations. I subjected the aptamer RNA to multiple pull-relax force-ramp cycles between 1 pN and 30 pN. The unfolding patterns of the aptamer in the absence and presence of preQ1 were found to be different (Figure 4.2). In the absence of preQ1, the aptamer unfolds at force 9.30 ± 0.96 pN (mean ± standard deviation, same below) with an extension of 9.09 ± 0.51 nm. In the presence of 200 nM preQ1, the aptamer unfolds at force 21.6 ± 3.1 pN with an extension of 14.7 ± 1.2 nm (Figure 4.3). When the preQ1 concentration in the buffer was further increased to 1 μM preQ1, the aptamer unfolds at force 22.4 ± 2.9 pN with an extension of 15.2 ± 1.9 nm.

Using the mWLC model, the number of nucleotides opened upon unfolding are assigned to the transitions. In the absence of preQ1, the unfolding distance corresponds to 24 ± 2 nucleotides, suggesting opening of either the P1A or the P1 hairpin structure, both comprising 22 nucleotides (Figure 4.1A). These two hairpin structures have almost similar thermodynamic stability as evaluated by Mfold (ΔG_{P1A} = - 4.40 kcal/mol, and ΔG_{P1} = - 4.80 kcal/mol). In the 200 nM preQ1 and 1 μM preQ1 conditions, the unfolding distances correspond to 32 ± 3 nucleotides and 33 ± 4 nucleotides respectively, both suggesting opening of the 33-nucleotide pseudoknot structure (Figure 4.3B). Thus, in the presence of preQ1 the aptamer forms possibly a pseudoknot structure.

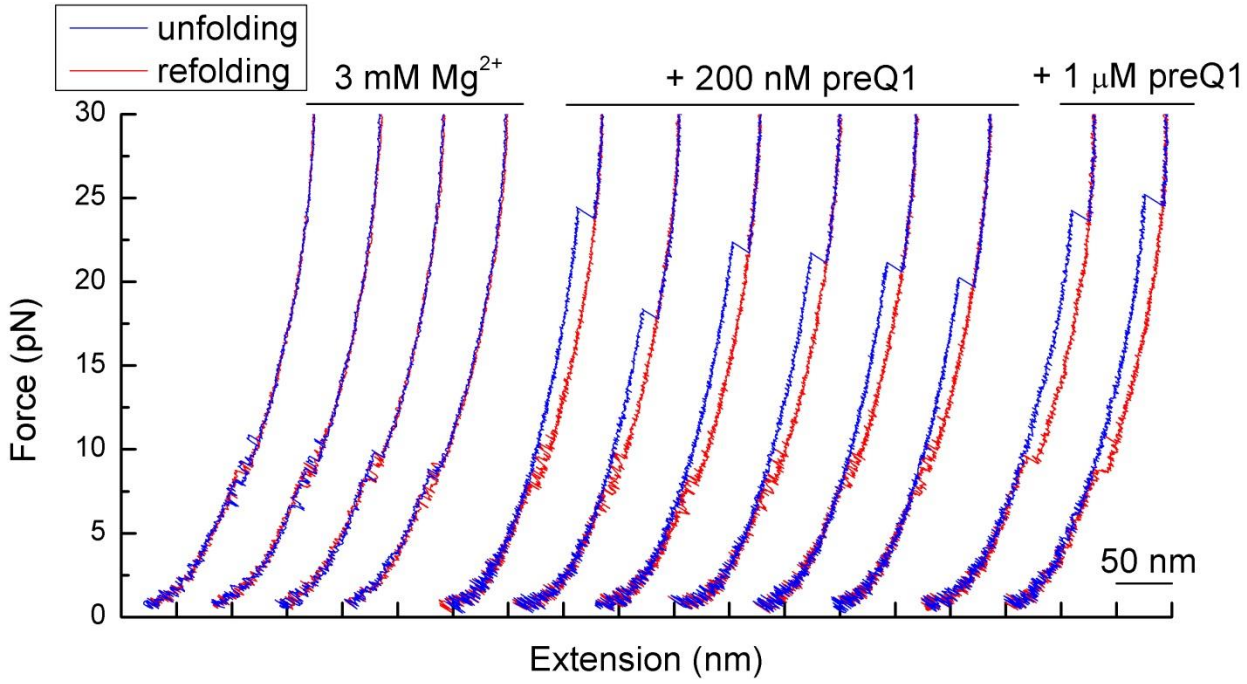


Figure 4.2 Pull-relax cycles for the preQ1 aptamer at various preQ1 concentrations.

(From left to right) Consecutive force-ramp cycles for the preQ1 aptamer construct in the presence of 0, 200 nM, and 1 μM preQ1. The RNA was subjected to multiple rounds of pull-relax cycles from 1- 30 pN with pulling speed of 200 nm/sec.

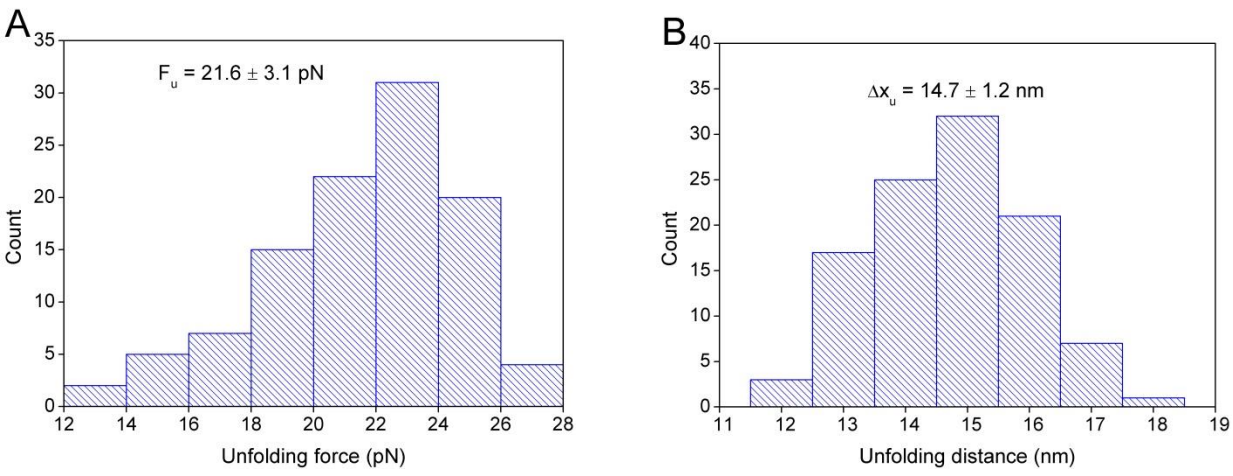


Figure 4.3 Unfolding force and distance distributions for preQ1 aptamer in 200 nM preQ1.

Distributions of (A) unfolding force and (B) unfolding distance observed for preQ1 aptamer pulled in 200 nM preQ1 in force ramp experiments. Data is representative of 110 pull-relax cycles on 14 independent molecules.

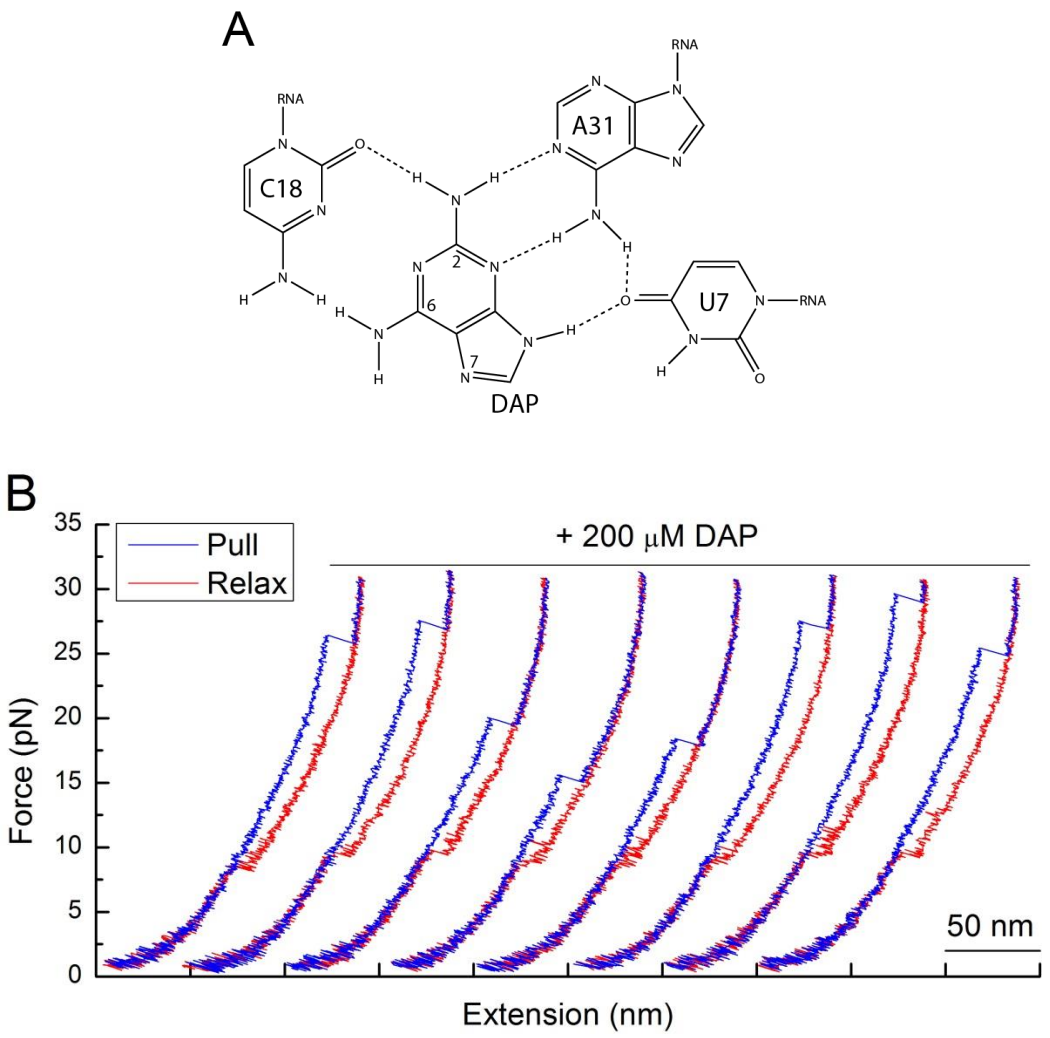


Figure 4.4 Effect of ligand DAP on preQ1 aptamer in pull-relax cycles.

(A) In the hydrogen bonding network of preQ1, replacement of preQ1 with an analog ligand DAP leads to impairment of two hydrogen bonds with C18. (B) Consecutive force-ramp cycles for the preQ1 aptamer construct in the presence 200 μM DAP.

In order to probe whether the aptamer shows a similar unfolding trajectory in the presence of related preQ1 analogs, I subjected the aptamer to the pulling assays in the presence of 2,6-diaminopurine (DAP). DAP differs from preQ1 at three positions: the absence of one hydrogen atom at position 1, the presence of one amine group at position 6, and the absence of one methylamine group at position 7 (Figure 4.4A). As a result, the replacement of DAP in the preQ1

base quadruple interaction C18-preQ1-A31-U7 leads to impairment of two hydrogen bonds between the ligand and C18. Previous equilibrium dialysis studies (186) showed that the dissociation constant K_d for DAP is > 100 μM, which is much higher than that of preQ1 (K_d=50 nM). Using 200 μM DAP in our single-molecule assay (Figure 4.4B), the preQ1 aptamer showed an unfolding force 24.1 ± 4.2 pN and a distance 14.83 ± 0.90 nm, similar to the preQ1 results, indicating that DAP also binds to the aptamer to form a pseudoknot structure.

4.3.2 Kinetics and pathways of ligand-dependent aptamer folding

To look into the folding pathways of the aptamer, I first examined the refolding cycles on the FECs (Figure 4.2). In the presence of 200 nM preQ1, the aptamer hops around 8.40 ± 0.98 pN until it folds to its initial state. However, when the concentration of preQ1 is increased to 1 μM, hopping is reduced (Figure 4.2), suggesting a ligand based concentration dependence of the hopping. In order to better resolve the folding transitions of the preQ1 aptamer, I carried out constant-force (CF) experiments. By monitoring the extension versus time traces, I can closely examine the equilibrium transitions between various conformational states of the aptamer.

- CF measurements in the absence of ligand

In the absence of ligand, when held at 9.5 pN, the aptamer RNA hops back and forth with a distance of 8.86 ± 0.88 nm (Figure 4.5A and B), corresponding to the opening and closing of the 22-nucleotide hairpin consistent with the force-ramp measurements. Thus the folding pathway proceeds from the unfolded (U) state to the hairpin (HP) state, i.e.

$$U \rightleftharpoons HP$$

To measure the folding and unfolding rates, I partitioned each trace into two states using the threshold method. The rate constants were calculated from the dwell times as follows:

$$k_{HP,U} = \frac{1}{\langle \tau_{HP,U} \rangle} \qquad \text{(Eq. 4.1)}$$

$$k_{U,HP} = \frac{1}{\langle \tau_{U,HP} \rangle} \qquad \text{(Eq. 4.2)}$$

As expected, the rate of hairpin unfolding decreases when force is lowered, as shown in Table 4.2.

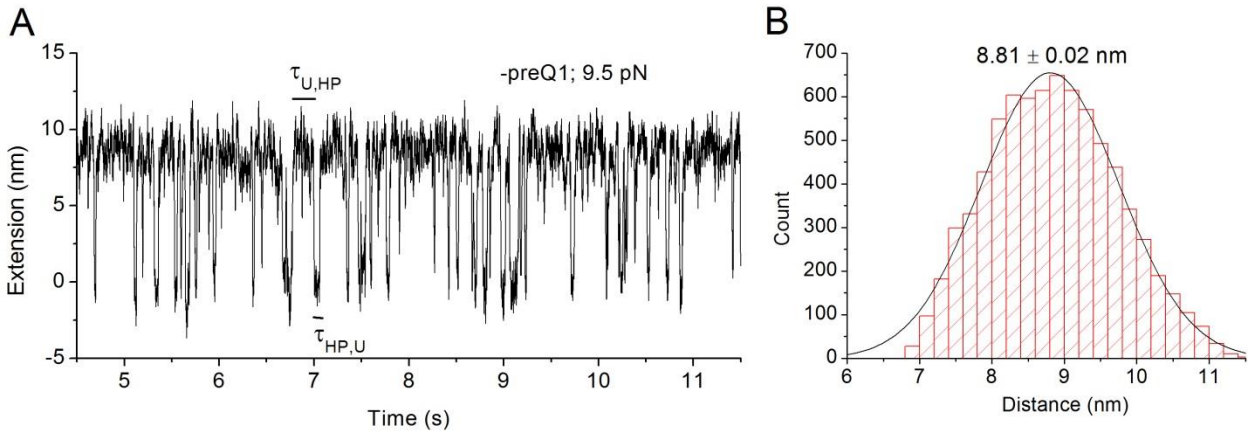


Figure 4.5 Constant force measurement of preQ1 aptamer in the absence of ligand.

(A) Constant force trace for the preQ1 aptamer at 9.5 pN in the absence of ligand. Examples of the dwell times $\tau_{HP,U}$ and $\tau_{U,HP}$ are indicated in the trace. (B) Distribution of the hopping distance, analyzed from a 40-second CF trace shown in (A), with mean and standard deviation of 8.86 ± 0.88 nm. The Gaussian fitted mean is 8.81 ± 0.02 nm.

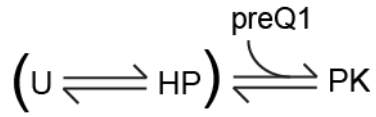
Table 4.2 preQ1 aptamer transition rates in the absence of preQ1

Force (pN)	$k_{unfolding}$ (HP→U) (s^{-1})	$k_{folding}$ (U→HP) (s^{-1})
10	48.5 ± 2.5	3.51 ± 0.21
9.5	39.2 ± 1.2	6.85 ± 0.27
9	25.77 ± 0.76	13.70 ± 0.40
8.5	18.13 ± 0.48	23.27 ± 0.60

Data represents mean ± standard error, measured from 12 CF traces of 3 molecules.

- CF measurements in the presence of preQ1

In the presence of 200 nM preQ1, the constant-force trace exhibits a combination of slow and fast kinetics at force 9.5 pN (Figure 4.6A). In the slow kinetics region, these rapid fluctuations are suppressed, suggesting the stabilization of the aptamer structure through preQ1 binding. In the presence of preQ1, the extension distribution from 6 CF traces in 6 molecules showed two major folding distances 8.06 nm and 11.31 nm (Figure 4.6B), corresponding to 22 nucleotides and 33 nucleotides, which suggest the folding of hairpin (HP) and pseudoknot (PK) structures respectively. The PK state can be converted from HP state, or directly from U state, as seen in Figure 4.6A. Thus, the folding pathway of the aptamer in the presence of preQ1 can be written as:



Based on the dwell time identifications as exemplified in Figure 4.6A, the hairpin and pseudoknot transition kinetics can be calculated. These rate constants measured at various constant forces are summarized in Table 4.3. At 9.5 pN, the hairpin folding rate $k_{U,HP}$ is 6.4 sec^{-1} and unfolding rate $k_{HP,U}$ is 41.7 sec^{-1} respectively, whereas the PK folding rate $k_{PK,f}$ is 0.04 sec^{-1} and unfolding rate $k_{PK,u}$ is 0.08 sec^{-1}. The hairpin folding and unfolding rates are similar to those in the absence of ligand (Table 4.2), suggesting the HP state is not much affected by the ligand preQ1. The unfolding of the pseudoknot state is indicative of ligand dissociation, as it bears a similar rate ($k_{PK,u}$ = 0.080 ± 0.014 sec^{-1}) to the off rate of preQ1, previously measured in the *F. nucleatum* preQ1 aptamer (k_{off} = 0.17 sec^{-1}, Table 4.1) using fluorescence spectroscopy (202).

To examine the concentration dependence of the binding kinetics, the aptamer was also pulled at 50 nM preQ1 and 1 μM preQ1. In the presence of 50 nM preQ1, only 56.8% of the force ramp traces show the ligand-bound unfolding signature. This percentage is much less than percentages for 200 nM preQ1 (91.3%) and 1 μM preQ1 (95.2%) conditions, suggesting higher preQ1 concentration promotes the ligand binding to the aptamer. Not surprisingly, in CF measurements, the PK folding rate $k_{PK,f}$ increases with the preQ1 concentration, whereas the PK unfolding rate $k_{PK,u}$ is relatively insensitive to the ligand concentration (Table 4.4).

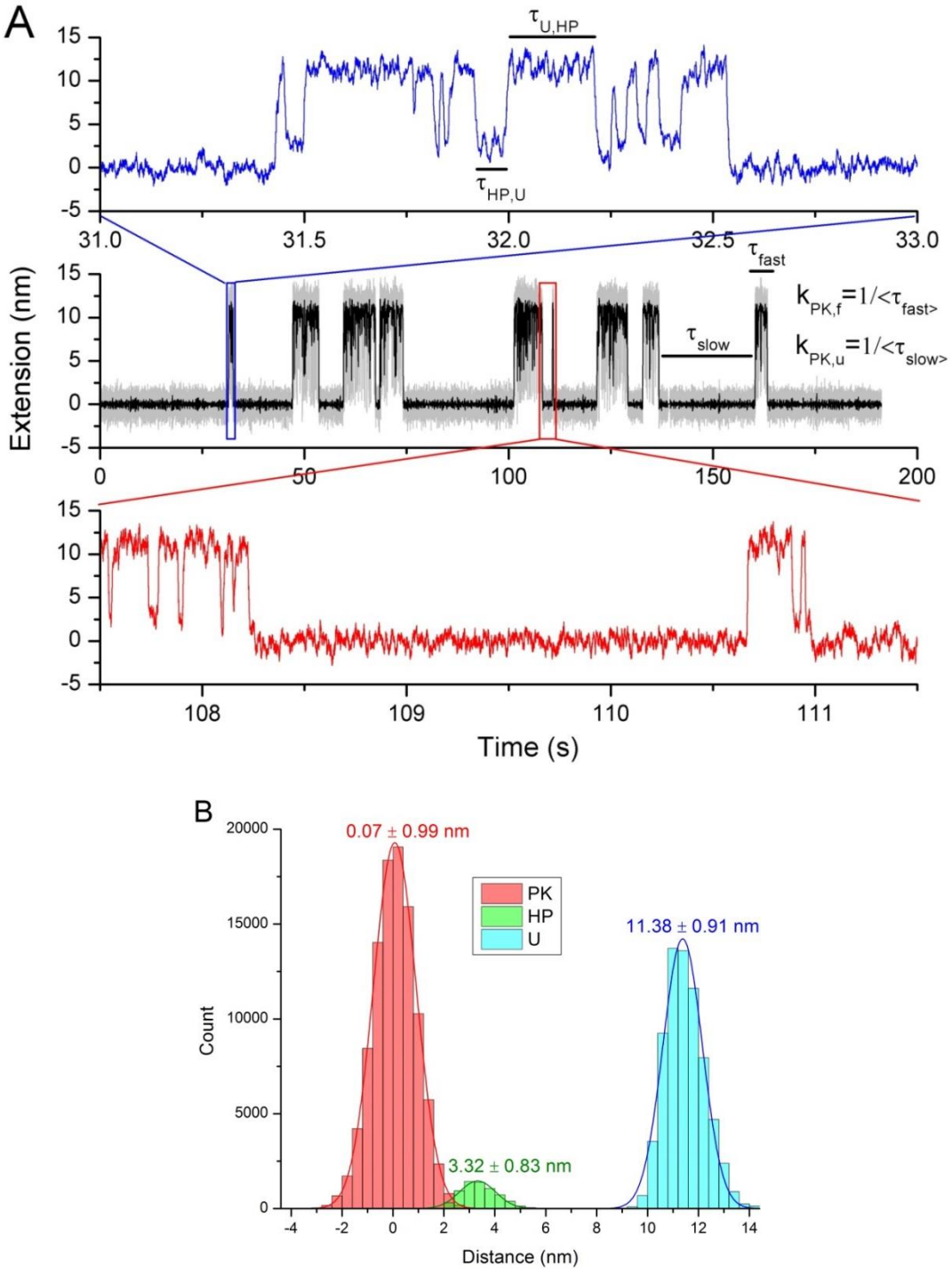


Figure 4.6 Constant force measurement of preQ1 aptamer in the presence of 200 nM preQ1.

(A) Constant force trace (gray, blue and red: 4 kHz trace; black: 300 point moving averaged) for the preQ1 aptamer at 9.5 pN in the presence of 200 nM preQ1. (B) Population histograms showing the extension values of three major states, i.e. unfolded state, hairpin, and pseudoknot. The distribution is representative of 6 CF traces at 9.5 pN from 6 molecules.

Table 4.3 preQ1 aptamer transition rates in the presence of 200 nM preQ1

Force (pN)	$k_{HP,U}$ (s^{-1})	$k_{U,HP}$ (s^{-1})	$k_{PK,u}$ (s^{-1})	$k_{PK,f}$ (s^{-1})
10	55.8 ± 1.2	3.26 ± 0.08	0.079 ± 0.017	0.038 ± 0.009
9.5	41.73 ± 0.85	6.42 ± 0.16	0.080 ± 0.014	0.130 ± 0.019
9	29.47 ± 0.65	13.20 ± 0.31	0.076 ± 0.011	0.249 ± 0.036
8.5	18.68 ± 0.59	24.20 ± 0.63	0.083 ± 0.011	0.540 ± 0.068

Data represents mean ± standard error, measured from 24 CF traces of 6 molecules.

Table 4.4 Ligand concentration dependence of pseudoknot unfolding and folding rates

Force (pN)	preQ1 concentration (nM)	$k_{PK,u}$ (s^{-1})	$k_{PK,f}$ (s^{-1})
9.5	50	0.100 ± 0.019	0.043 ± 0.014
	200	0.080 ± 0.014	0.130 ± 0.019
	1000	0.077 ± 0.012	0.223 ± 0.074
9	50	0.084 ± 0.013	0.100 ± 0.013
	200	0.076 ± 0.011	0.249 ± 0.036
	1000	0.075 ± 0.017	0.34 ± 0.15

Data represents mean ± standard error, measured from 6 molecules (50 nM preQ1), 6 molecules (200 nM preQ1) and 4 molecules (1 μM preQ1).

In the presence of 200 nM preQ1, close inspection of the constant-force traces revealed that the aptamer hops between more than U, HP and PK states (Figure 4.7). Specifically, a state with an additional folding distance of 1-2 nm from the HP state was observed, implying that the 3' A-rich tail forms tertiary contacts (TC) with the hairpin stem. This TC intermediate allows the aptamer to efficiently bind ligands since it acts as a precursor towards the PK state.

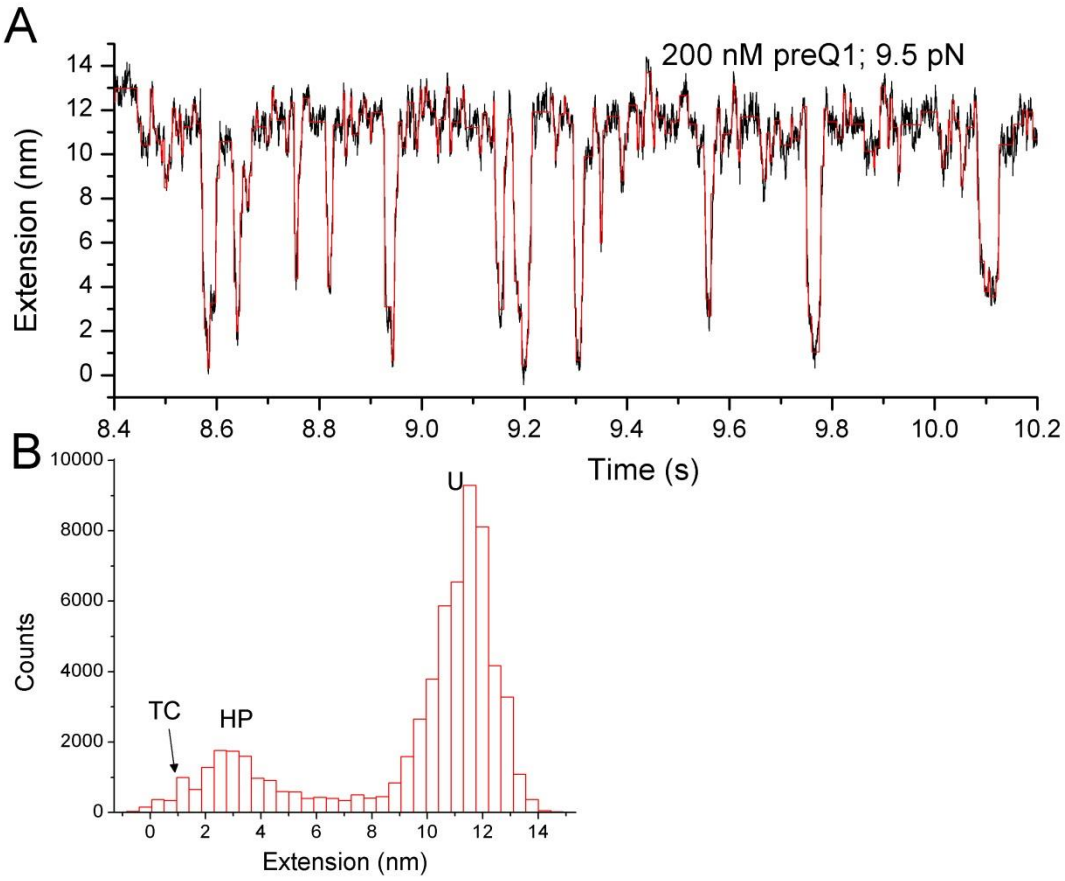


Figure 4.7 Piecewise-constant analysis of preQ1 aptamer CF trace at 9.5 pN.

(A) Piecewise-constant analysis (red) of a typical constant-force trace (black) at 9.5 pN and (B) the corresponding extension histogram. Besides the major populations of U (11.5 nm) and HP (3 nm), a state with 1-2 nm folding distance from HP is observed, indicating tertiary contacts between HP and the 3' A-rich tail of the aptamer. The PK state is omitted in this histogram.

- CF measurements in the presence of DAP

To further test the folding pathway, I also investigated the aptamer folding in the presence of the ligand DAP under constant forces. As shown in Figure 4.8A, with 200 μM DAP present, the aptamer hops between the folded and the unfolded states at 11 pN with a distance of 12.0 ± 1.0 nm (Figure 4.8B), corresponding to 30 ± 3 nt, which suggests the formation of the PK structure.

Further inspection revealed another folding distance ~8 nm, indicating the HP state (Figure 4.8C). Thus the aptamer undergoes a ligand-dependent folding pathway in the presence of DAP.

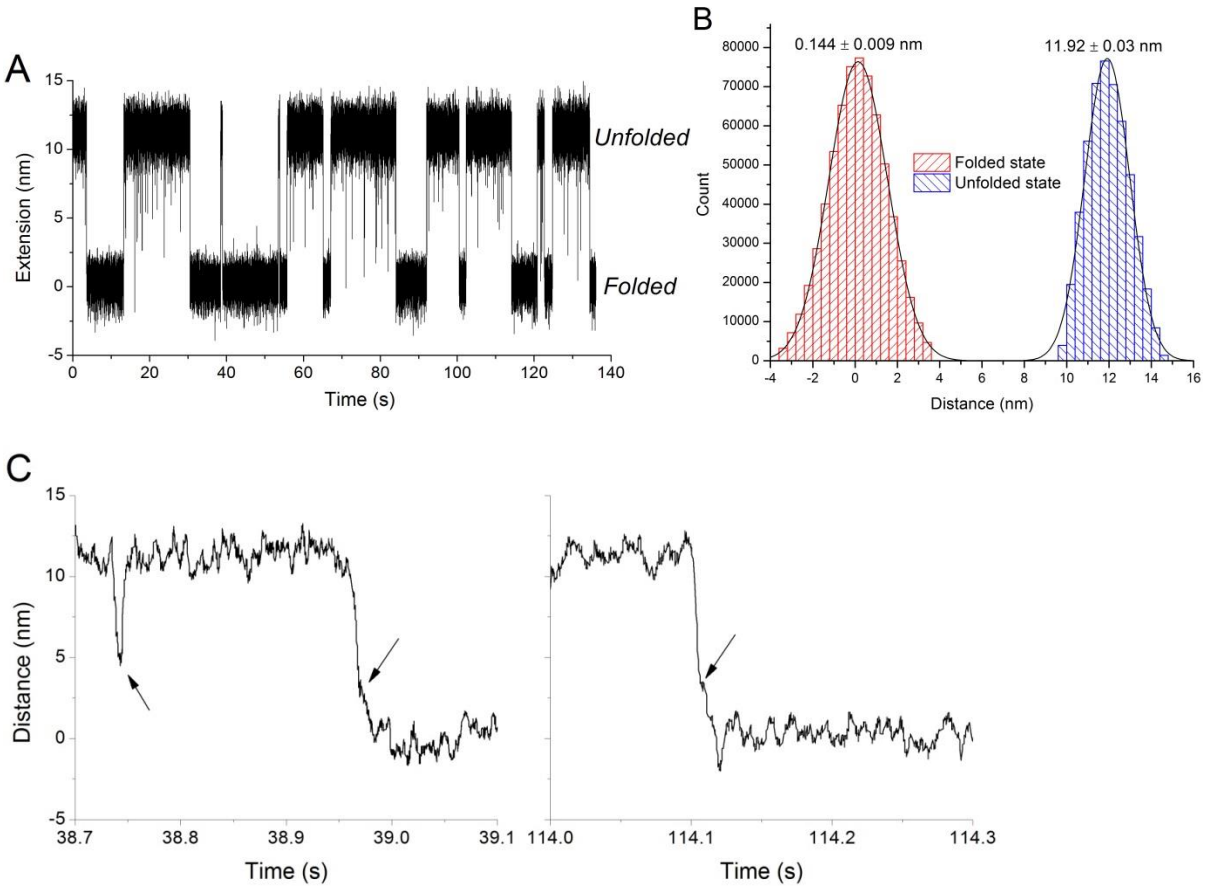


Figure 4.8 Constant-force measurement of preQ1 aptamer in the presence of 200 μM DAP.

(A) A single preQ1 aptamer is held at a constant force of 11 pN, with the molecule hopping between the folded and the unfolded states. (B) Extension distribution for the folded (red) and the unfolded (blue) states. The peak centers on the Gaussian fitted curves for the folded and unfolded states are 0.144 ± 0.009 nm and 11.92 ± 0.03 nm as indicated. The data is representative of 6 molecules, with the bin size of 0.4 nm and a mean unfolding distance of 12.0 ± 1.0 nm, thus suggesting the unfolding of the pseudoknot structure at 11.0 pN in the presence of 200 μM DAP concentrations. (C) Two close-up views of the trace in Figure 4.8A. The arrows indicate the folding events with a folding distance around 8 nm, suggesting formation of the hairpin structure.

4.3.3 Mutational analysis of the preQ1 aptamer

Next, I designed four aptamer mutants in order to investigate the folding pathways of the pseudoknot structure.

In mutant m1, I replaced the A2-U21 base pair in the P1A stem with a G-C base pair, thus making P1A hairpin more stable. This is to test whether P1A hairpin alone is capable of ligand binding. Thus, the mutation stabilizes P1A while destabilizing the P1 hairpin. The theoretical stability for P1A by Mfold becomes ΔG_{P1A} = - 6.50 kcal/mol. The P1 hairpin becomes unstable due to A4-C21 mismatch and G2-U23 wobble pair (Figure 4.9A). As C18 base pairs with G5 in the P1A hairpin, the ligand is unable to bind. Hence, we did not observe binding in the FECs (Figure 4.9B). At 9.5 pN, the hopping distance is 8.65 ± 0.72 nm at a saturating 1 μM preQ1 concentration (Figure 4.9D), suggesting the formation of only the hairpin (P1A) structure. Thus the results show that P1A hairpin is incapable to bind preQ1.

In mutant m2, I replaced C18 with a uracil. This mutation was investigated in previous in-line probing studies to test the role of C18 in preQ1 binding specificity (186). The hydrogen bonding between preQ1 and U18 in the base quadruple interaction is disrupted (Figure 4.10C). Hence, m2 aptamer does not show a binding signature at 1 μM preQ1 (Figure 4.10B).The characteristic hopping at 9.5 pN showed a distance of 8.70 ± 0.59 nm (Figure 4.10E), suggesting the hairpin structure. On the other hand, the ligand DAP binds the m2 aptamer at 200 μM concentration as evidenced by higher unfolding force (Figure 4.10B) and longer force hopping distance of 11.41 ± 0.57 nm at 9.5 pN, confirming the pseudoknot state is formed (Figure 4.10F). The binding capability of DAP is probably due to its potential of forming three hydrogen bonds with U18 (Figure 4.10D). A similar observation was made in the in-line probing assays wherein, preQ1 binding was abolished and DAP showed a binding with the aptamer with K_d > 300 μM (186).

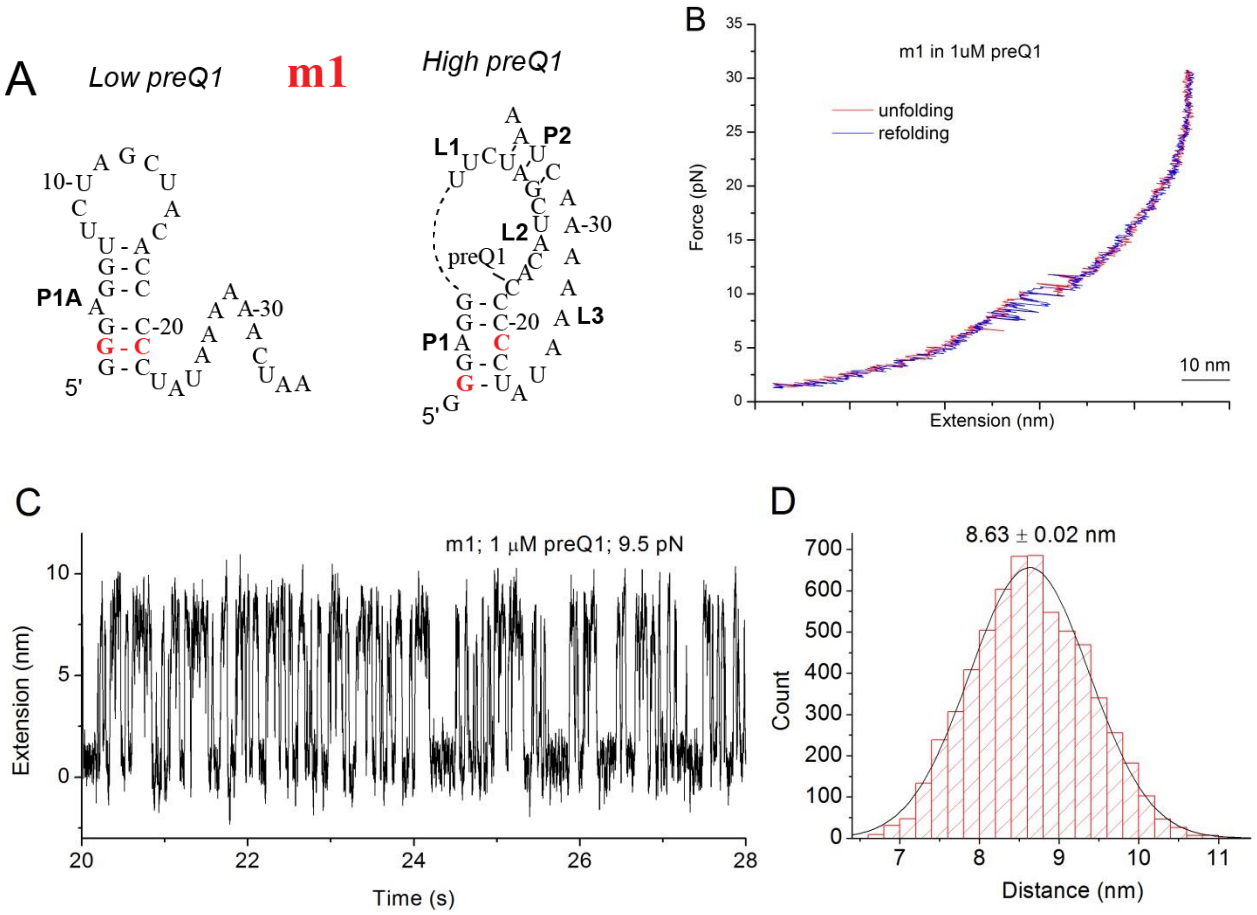


Figure 4.9 Force ramp and constant force measurements on m1 mutant.

(A) Mutated bases in m1 mutant. (B) Force ramp trace of m1 in 1 μM preQ1 condition. (C) Constant force trace and (D) distance distribution of m1 in 1 μM preQ1 at 9.5 pN. The mean distance is 8.65 ± 0.72 nm, with peak center at 8.63 ± 0.02 nm.

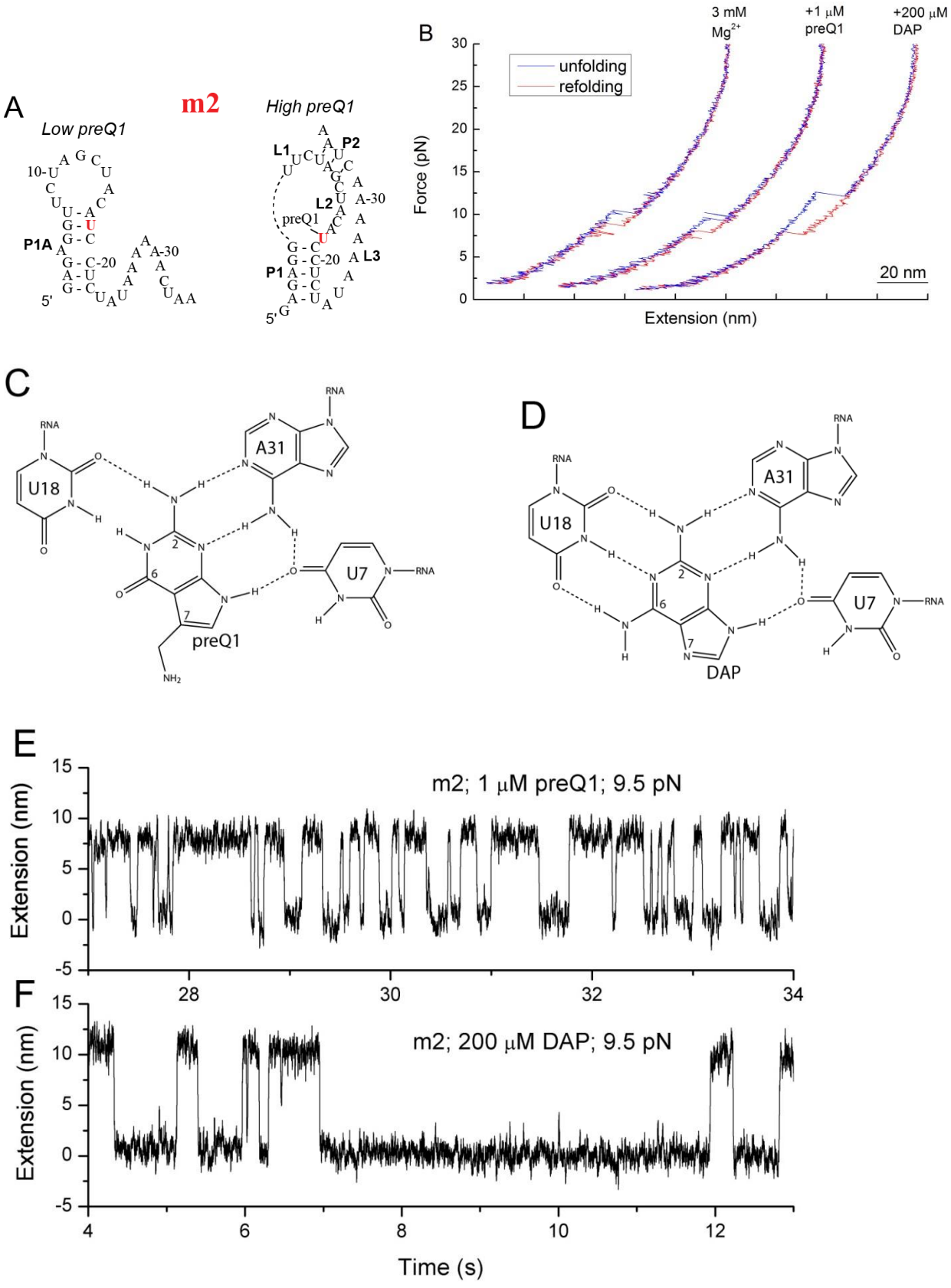

A
m2
Low preQ1
High preQ1
P1A
P1
L1
L2
L3
P2
preQ1
10-U
A-30
5'
B
3 mM Mg2+
+1 μM preQ1
+200 μM DAP
unfolding
refolding
Force (pN)
Extension (nm)
20 nm
C
RNA
A31
U18
U7
preQ1
NH2
D
DAP
E
m2; 1 μM preQ1; 9.5 pN
Extension (nm)
F
m2; 200 μM DAP; 9.5 pN
Time (s)

Figure 4.10 Force ramp and constant force measurements on m2 mutant.

(A) Mutated bases in m2 mutant. (B) Typical force ramp traces of m2 in various ligand conditions. (C) Impairment of two hydrogen bonds between preQ1 and U18 in mutant m2. (D) Restoration of three hydrogen bonds between DAP and U18 in mutant m2. (E) Constant force trace of m2 in the presence of 1 μM preQ1 at 9.5 pN. (F) Constant force trace of m2 in the presence of 200 μM DAP at 9.5 pN.

Next, in mutant m3, I replaced A4-U21 base pair in the P1 stem with a G-C base pair (Figure 4.11A). This stabilized the P1 hairpin (ΔG_{P1} = - 6.90 kcal/mol as predicted by Mfold) while maintaining the ligand binding capability of the aptamer. We expected to observe higher unfolding force for the hairpin and slower hairpin hopping kinetics, due to the stable hairpin in m3.

In the absence of ligand, the unfolding and refolding cycles of m3 are almost reversible (Figure 4.11B). The average unfolding force is 11.9 ± 1.3 pN (98 FECs, 9 molecules). Upon addition of 200 nM preQ1 in the solution, no obvious change of the unfolding pattern was observed (Figure 4.11B). M3 aptamer only starts to show a preQ1 binding phenotype at concentrations greater than 1 μM. When 10 μM preQ1 was supplied in the buffer, a much higher unfolding force 20.2 ± 4.6 pN (94 FECs, 7 molecules) was observed (Figure 4.11B), indicating a ligand bound conformation. This also suggests that m3 is a poor binder, compared to the wild type aptamer which shows binding in 200 nM preQ1. Such a change in binding affinity is possibly due to the alterations in the P1-L3 tertiary interaction.

When held at constant force of 12 pN in the absence of ligand, m3 hops back and forth with a distance of 8.58 ± 0.59 nm, suggesting the formation of the hairpin structure (Figure 4.11C and D). In the presence of 10 μM preQ1 at 12 pN, m3 showed folding events of both hairpin and pseudoknot (Figure 4.11E and F). In the CF trace hopping is restricted within 8.19 ± 0.60 nm, which corresponds to the opening and closing of the hairpin. An extra gain of folding distance 4.20 ± 0.60 nm was observed, indicating the ligand induced formation of the pseudoknot structure (Figure 4.11G). Thus, the mutant m3 proves the proposed folding pathways for the aptamer, with kinetic rates shown in Table 4.5.

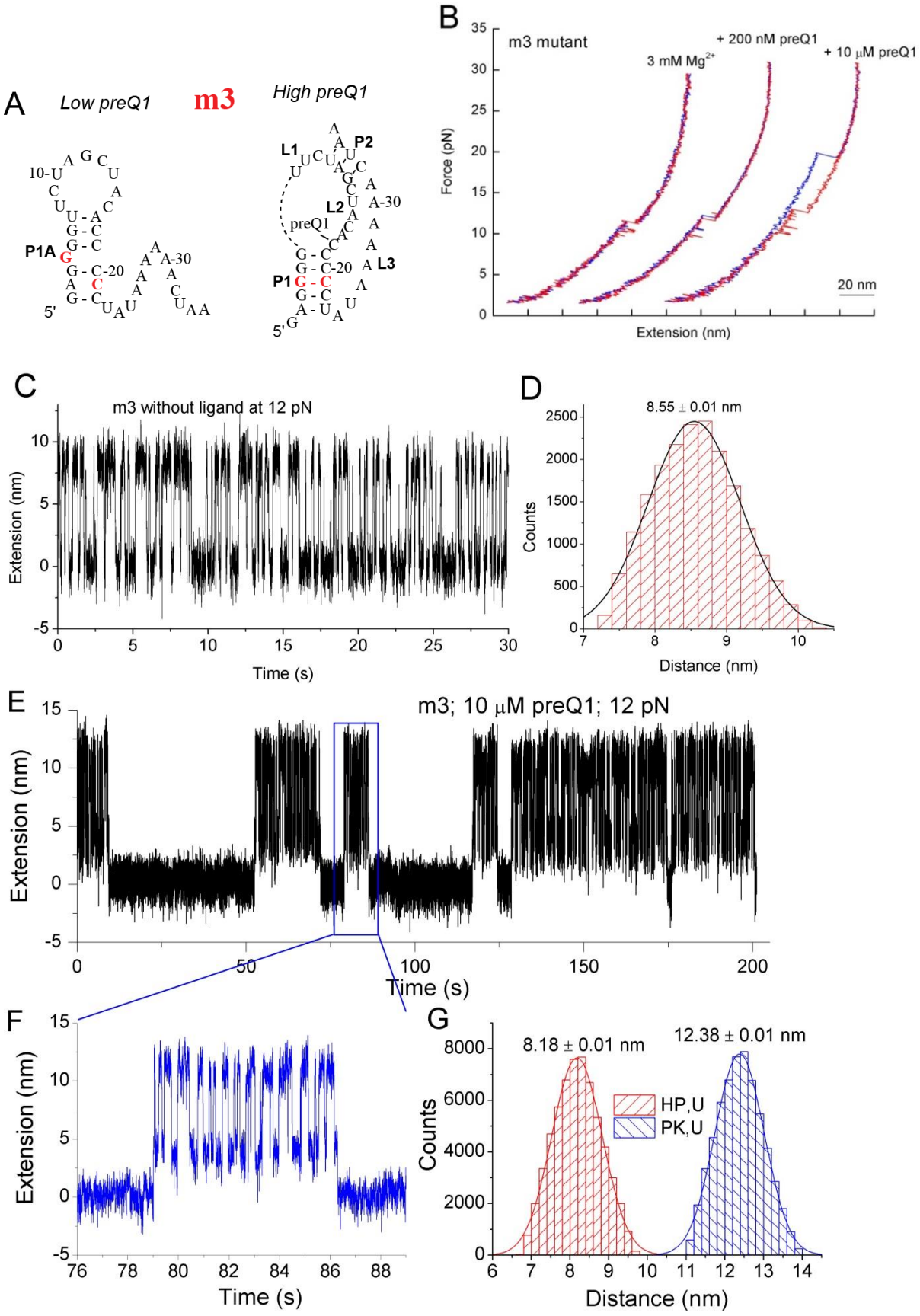

A
Low preQ1
m3
High preQ1
P1A
L1
P2
L2
L3
P1
preQ1
B
m3 mutant
3 mM Mg2+
+ 200 nM preQ1
+ 10 μM preQ1
20 nm
Force (pN)
Extension (nm)
C
m3 without ligand at 12 pN
Extension (nm)
Time (s)
D
8.55 ± 0.01 nm
Counts
Distance (nm)
E
m3; 10 μM preQ1; 12 pN
Extension (nm)
Time (s)
F
Extension (nm)
Time (s)
G
8.18 ± 0.01 nm
12.38 ± 0.01 nm
HP,U
PK,U
Counts
Distance (nm)

Figure 4.11 Force ramp and constant force measurements on m3 mutant.

(A) Mutated bases in m3 mutant.

(B) (From left to right) Typical FECs for m3 in the absence of ligand, in 200 nM preQ1 and in 10 μM preQ1. The ligand binds at 10 μM preQ1, as evidenced by the higher unfolding force.

(C) Constant force trace of m3 at 12 pN in the absence of preQ1. The corresponding hopping distance is shown in (D). The Gaussian fitted mean distance is 8.55 nm.

(E)(F) Constant force trace of m3 at 12 pN in the presence of 10 μM preQ1. The corresponding hopping distance distributions are shown in (G), where the Gaussian fitted mean distances for HP and PK are 8.18 nm and 12.38 nm respectively.

Table 4.5 m3 mutant transition rates in the presence of 10 μM preQ1

Force (pN)	$k_{HP,U}$ (s^{-1})	$k_{U,HP}$ (s^{-1})	$k_{PK,u}$ (s^{-1})	$k_{PK,f}$ (s^{-1})
12.5	8.09 ± 0.30	2.68 ± 0.10	0.072 ± 0.025	0.023 ± 0.009
12	6.12 ± 0.11	5.22 ± 0.10	0.078 ± 0.012	0.034 ± 0.006
11.5	4.72 ± 0.10	9.05 ± 0.19	0.075 ± 0.012	0.044 ± 0.008
11	3.51 ± 0.14	15.87 ± 0.55	0.098 ± 0.018	0.097 ± 0.019

Data represents mean ± standard error, measured from 21 CF traces of 6 molecules.

Finally, in mutant m4, C32 and C33 were changed to two A's (Figure 4.12A). As a result, the P2 stem and its associated base triple C32-G12-A17 in the binding pocket are disrupted. As shown in Figure 4.12B, mutant m4 does not show a binding signature in the unfolding FEC in 1 μM preQ1 containing buffer. The mutant m4 hops at 9.5 pN with a distance of 8.58 ± 0.48 nm (Figure 4.12C and D), similar to the wild type in ligand-free condition. This mutant proved that the base pair formation in P2 stem is a prerequisite for preQ1 binding and pseudoknot folding of the aptamer, which is consistent with the NMR and mutagenesis study by (201).

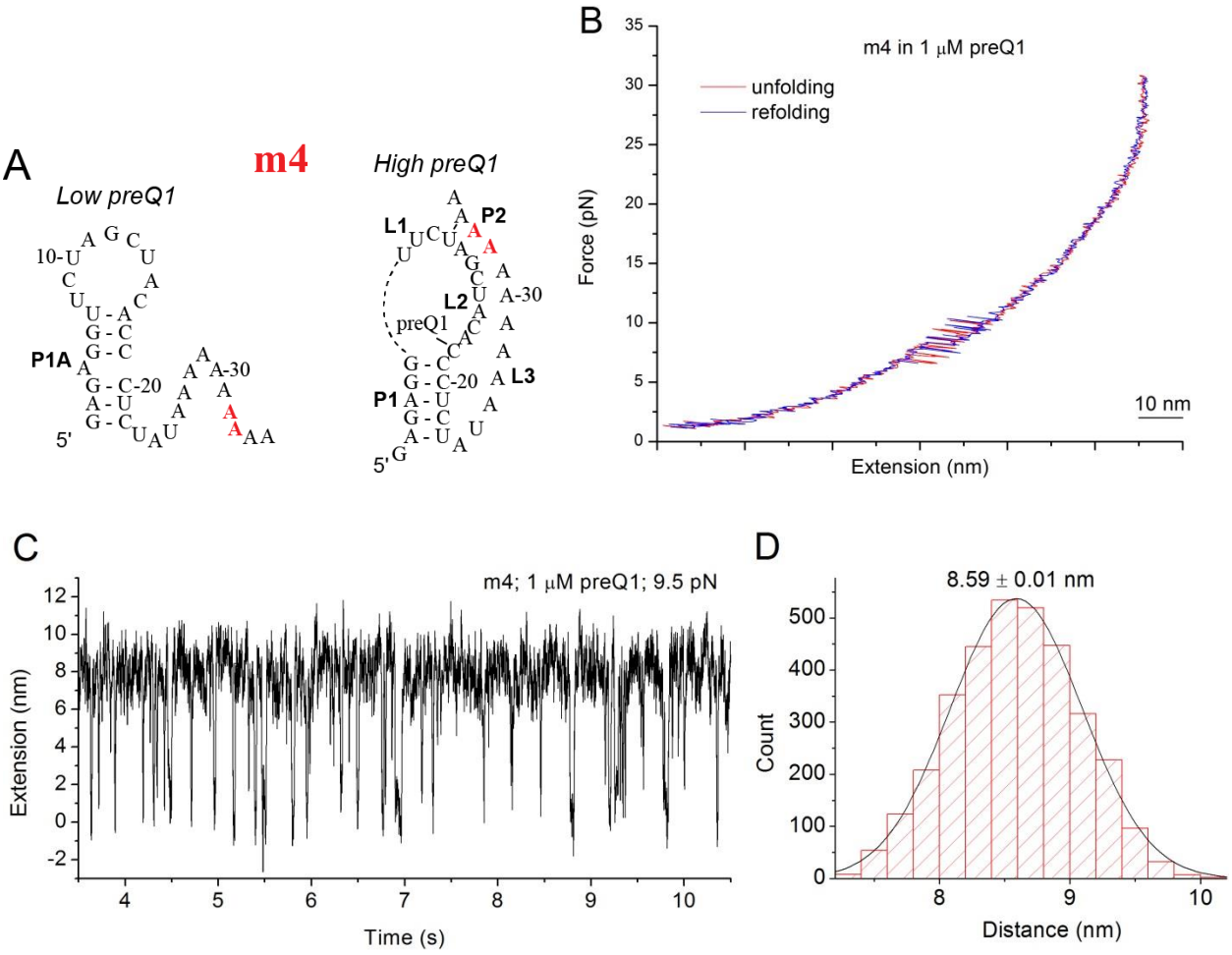


Figure 4.12 Force ramp and constant force measurements on m4 mutant.

(A) Mutated bases in m4 mutant. (B) Force ramp trace of m4 in saturating preQ1 condition. (C) Constant force trace and (D) distance distribution of m4 in 1 μM preQ1 at 9.5 pN. The mean distance is 8.58 ± 0.48 nm, with peak center at 8.59 ± 0.01 nm.

4.3.4 Free energy landscapes of ligand-dependent aptamer folding

Having characterized the folding pathway and kinetics of the preQ1 aptamer, I then looked into the energetics of the ligand-dependent aptamer folding to develop a free energy landscape using the equilibrium sampling method. As observed in Chapter 3, a free energy landscape developed from single-molecule assays contains detailed information on energy barriers between the different conformational states.

The energy landscape for the wild type aptamer was constructed at 9.5 pN in the absence and presence of ligand. In the absence of ligand, the extension probability distribution shows two major populations (U and HP), while in the presence of 200 nM preQ1 the plot shows three populations namely U, HP, and PK (Figure 4.13A and C). The resulting energy landscapes are shown in Figure 4.13B and D, which represent the averaged energy surfaces obtained from the same molecule at three different forces (10 pN, 9.5 pN and 9 pN). The standard deviation between the energy surfaces is around 1 k_BT. In the absence of ligand, the HP↔U transition shows a barrier of $\Delta G^{\ddagger}_{HP,U}$ = 2 k_BT. In the presence of ligand, the barrier between HP and U is similar $\Delta G^{\ddagger}_{HP,U}$ = 3 k_BT. The folding from HP to PK, which depends on ligand binding, exhibits an energy barrier of $\Delta G^{\ddagger}_{HP,PK}$ = 3.2 k_BT. The energy landscapes for m3 (Figure 4.13F and H) were computed at 12 pN, at which the HP and U are close to equilibrium. In both the ligand-free and 10 μM preQ1 conditions, the barrier height between HP and U is around 6 k_BT, twice of that of the wild type aptamer, indicating that a single G-C base pair replacement results in a stronger stem-loop conformation and a higher barrier for unfolding. The folding barrier from HP to PK, $\Delta G^{\ddagger}_{HP,PK}$ = 7.4 k_BT, was also larger than that of the wild type. This suggests that ligand association is slower in m3 and thus m3 is a poorer binder than wild type.

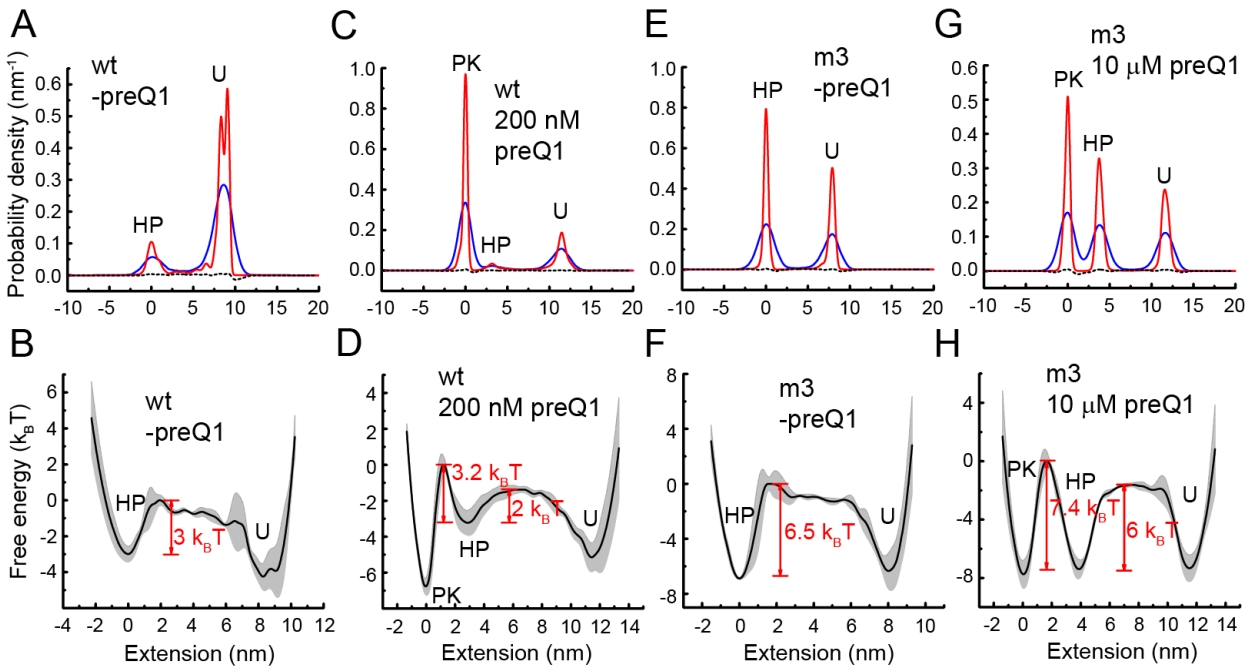


Figure 4.13 Free energy landscapes of preQ1 wild type aptamer and m3 mutant.

(A) Extension probability deconvolution for preQ1 aptamer at 9.5 pN in the absence of ligand. The measured extension distribution (blue) is deconvoluted (red) to remove the effects of thermal fluctuations of beads and handles. The residual error is plotted as the black dashed line.

(B) Free energy landscape for preQ1 aptamer at 9.5 pN in the absence of ligand. The solid curve represent the averaged landscape from three force measurements (10 pN, 9.5 pN and 9 pN), while the shaded area indicates ± one standard deviation margin. The barrier top is aligned at 0 k_BT.

(C) (D) Same as (A)(B) except that they represent results for wild type aptamer in the presence of 200 nM preQ1.

(E) Extension probability deconvolution for m3 mutant at 12 pN in the absence of preQ1.

(F) Free energy landscape for m3 at 12 pN in the absence of preQ1. The solid curve represent the averaged landscape from three force measurements (12.5 pN, 12 pN, and 11.5 pN), while the shaded area indicates the margin of one standard deviation. The barrier top is aligned at 0 k_BT.

(G)(H) Same as (E)(F) except that they represent results for m3 in the presence of 10 μM preQ1.

4.4 Discussion

In this work, I have applied single-molecule optical tweezers assays to characterize pseudoknot folding in preQ1 riboswitch. The folding of a pseudoknot, in general, is an intriguing subject under extensive study. Besides its functional importance in translational repression (38-40), ribosomal frameshifting (41-44), activity of ribozyme (207, 208) and telomerase (45), and riboswitches of preQ1 (186), class II S-adenosylmethionine (195), and S-adenosylhomocysteine (209), it also provides an excellent model system for studying complex RNA structures involving the interplay between secondary and tertiary interactions, non-canonical base pairs, triplexes, coaxial stacking, and sharp turns. Although different topologies of pseudoknots exist (210), the most common and best characterized is the H-type, in which the bases in the hairpin loop base-pairs with the single-stranded region outside of the stem.

The folding thermodynamics of the H-type pseudoknot RNAs in human telomerase, retroviruses bacteriophages, and bacterial operons has been studied by thermal melting, chemical and enzymatic probing, and calorimetry (45, 211-219). Recently, developments in single-molecule techniques have further facilitated the study of unfolding/folding dynamics of biomolecules. Mechanical unfolding studies of pseudoknots by optical tweezers found that the frameshifting efficiency could be related to the mechanical stability or formation of alternative structures (220-223). Mg^{2+} ion has been found to stabilize pseudoknot structure, as observed in optical tweezers (223) and FRET experiment (224). Optical tweezers experiment on the pseudoknot RNA in human telomerase indicated a hairpin intermediate during its folding at force below 10 pN (225). Many molecular dynamics (MD) simulation studies have also shed light on the folding pathways of pseudoknot (226-230). A recent MD study proposed that the relative stabilities of the helical stems (P1 and P2) are the main determinant of the folding pathways of H-type pseudoknots (228). If there is sufficient difference in the stability of the two stems, the relatively more stable one forms first. Otherwise, parallel folding pathways could exist for both stems. In the preQ1 aptamer, the 5-bp P1 stem is much more stable than the 3-bp P2 ($\Delta\Delta G$ = -6.5 kcal/mol predicted by Mfold) (Figure 4.14). Thus, it is not surprising that P1 folds before P2. It is notable that in the preQ1 riboswitch, the binding of ligand adds another dimension to the folding of pseudoknot. Thus, there is a ligand concentration dependence on the PK folding kinetics (Table 4.4).

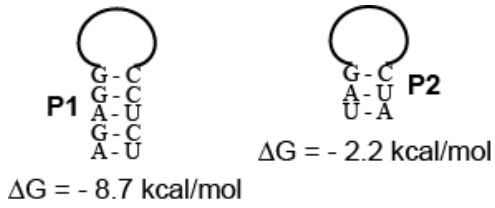


Figure 4.14 Comparison of the stabilities of P1 stem and P2 stem.

Both stems are connected by a GCAA tetraloop, and then evaluated by Mfold.

4.5 Conclusion

In this chapter, I have dissected the folding pathways of the aptamer domain in preQ1 riboswitch by single-molecule force spectroscopy and mutational methods. The pseudoknot structure of the aptamer is found to fold through a hairpin intermediate and require the binding of ligand, with kinetics dependent on ligand concentrations. In the future, the thermodynamics of ligand binding will be more extensively investigated. As shown in Chapter 3, using fluctuation theorems such as JE and CFT, the free energy change can be computed from work done along FECs. Here, by computing the work done for ligand-free and ligand-bound FECs (Figure 4.2), I can calculate the ligand binding free energy as the difference between the two cases. This binding free energy can also be compared with results measured from bulk experiments such as isothermal titration calorimetry.

Although the kinetics for aptamer folding and ligand binding are determined in this study, the interplay between the aptamer and the downstream expression platform needs to be further examined in order to elucidate the riboswitch control mechanism. To that end, the current single-molecule study needs to be extended to the full-length riboswitch. Previous phylogenic analysis of the full-length preQ1 riboswitch sequence predicted the structures of an anti-terminator and a terminator in the expression platform (186, 199, 200) (Figure 4.15). The single-molecule study will measure the real-time transitions of the full-length riboswitch between these alternative structures and help understand the switching mechanism.

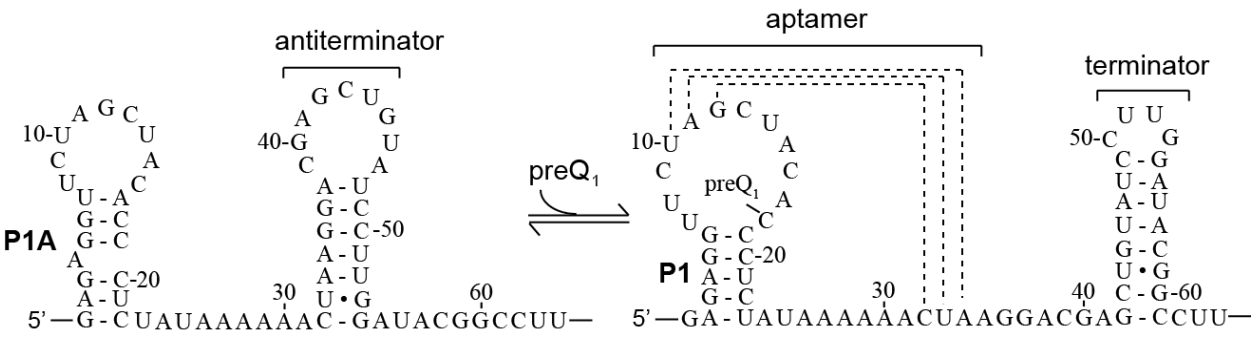


Figure 4.15 Phylogenic analysis predicts the anti-terminator and the terminator structures in the full-length *B. subtilis* preQ1 riboswitch. This figure is adapted from reference (200).

In this thesis, using optical tweezers, which enables measurement of RNA folding dynamics in real time and at the single-molecule level, I have characterized the folding of an RNA hairpin and a pseudoknot, which suggest that the RNA folding pathway is complex, involving multiple intermediates even for the most basic secondary structures. An important future direction is to study the folding of RNA during transcription, which is more relevant to *in vivo* conditions. Single-molecule study of co-transcriptional folding of structured RNAs, such as ribozymes and riboswitches, will be critical for understanding their functions.

References

1. Geiduschek E. P., Haselkorn R. (1969) Messenger RNA. *Annu. Rev. Biochem.* 38:647-676.
2. Schweet R., Heintz R. (1966) Protein synthesis. *Annu. Rev. Biochem.* 35:723-758.
3. Cech T. R. (1990) Self-splicing of group I introns. *Annu. Rev. Biochem.* 59:543-568.
4. Frank D. N., Pace N. R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* 67:153-180.
5. Noller H. F., Hoffarth V., Zimniak L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256(5062):1416-1419.
6. Montange R. K., Batey R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* 37:117-133.
7. Serganov A. (2009) The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* 19(3):251-259.
8. Roth A., Breaker R. R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* 78:305-334.
9. Serganov A., Nudler E. (2013) A decade of riboswitches. *Cell* 152(1-2):17-24.
10. Ambros V. (2004) The functions of animal microRNAs. *Nature* 431(7006):350-355.
11. Aksel T., Majumdar A., Barrick D. (2011) The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure* 19(3):349-360.
12. Carthew R. W., Sontheimer E. J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4):642-655.
13. Hamilton A. J., Baulcombe D. C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286(5441):950-952.
14. Elbashir S. M., *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411(6836):494-498.
15. Heinicke L. A., Bevilacqua P. C. (2012) Activation of PKR by RNA misfolding: HDV ribozyme dimers activate PKR. *RNA* 18(12):2157-2165.
16. Stahlhut Espinosa C. E., Slack F. J. (2006) The role of microRNAs in cancer. *Yale J. Biol. Med.* 79(3-4):131-140.
17. Woese C. R., Gutell R., Gupta R., Noller H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* 47(4):621-669.
18. Gutell R. R., Larsen N., Woese C. R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* 58(1):10-26.
19. Barfod E. T., Cech T. R. (1989) The conserved U.G pair in the 5' splice site duplex of a group I intron is required in the first but not the second step of self-splicing. *Mol. Cell. Biol.* 9(9):3657-3666.
20. Michel F., Umesono K., Ozeki H. (1989) Comparative and functional anatomy of group II catalytic introns--a review. *Gene* 82(1):5-30.
21. Strobel S. A., Cech T. R. (1995) Minor groove recognition of the conserved G.U pair at the Tetrahymena ribozyme reaction site. *Science* 267(5198):675-679.
22. Jucker F. M., Heus H. A., Yip P. F., Moors E. H., Pardi A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* 264(5):968-980.

23. Orita M., *et al.* (1993) High-resolution NMR study of a synthetic oligoribonucleotide with a tetranucleotide GAGA loop that is a substrate for the cytotoxic protein, ricin. *Nucleic Acids Res.* 21(24):5670-5678.
24. Szewczak A. A., Moore P. B. (1995) The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* 247(1):81-98.
25. Correll C. C., *et al.* (1998) Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. U. S. A.* 95(23):13436-13441.
26. Cheong C., Varani G., Tinoco I., Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* 346(6285):680-682.
27. Sakata T., *et al.* (1990) Studies on the structure and stabilizing factor of the CUUCGG hairpin RNA using chemically synthesized oligonucleotides. *Nucleic Acids Res.* 18(13):3831-3839.
28. Varani G., Cheong C., Tinoco I., Jr. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry* 30(13):3280-3289.
29. Korostelev A., Trakhanov S., Laurberg M., Noller H. F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126(6):1065-1077.
30. Bevilacqua P. C., Brown T. S., Nakano S., Yajima R. (2004) Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers* 73(1):90-109.
31. Azzalin C. M., Reichenbach P., Khoriauli L., Giulotto E., Lingner J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318(5851):798-801.
32. Benetti R., *et al.* (2008) A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol.* 15(9):998.
33. Bugaut A., Balasubramanian S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* 40(11):4727-4741.
34. Cech T. R., Damberger S. H., Gutell R. R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.* 1(5):273-280.
35. Cate J. H., *et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282):1678-1685.
36. Bloomfield V. A., Crothers D. M., Tinoco I. (2000) *Nucleic acids : structures, properties, and functions* (University Science Books, Sausalito, Calif.) pp x, 794 p.
37. Butcher S. E., Pyle A. M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* 44(12):1302-1311.
38. Spedding G., Draper D. E. (1993) Allosteric mechanism for translational repression in the Escherichia coli alpha operon. *Proc. Natl. Acad. Sci. U. S. A.* 90(10):4399-4403.
39. Chiaruttini C., Milet M., Springer M. (1997) Translational coupling by modulation of feedback repression in the IF3 operon of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 94(17):9208-9213.
40. Shamoo Y., Tam A., Konigsberg W. H., Williams K. R. (1993) Translational repression by the bacteriophage T4 gene 32 protein involves specific recognition of an RNA pseudoknot structure. *J. Mol. Biol.* 232(1):89-104.
41. Giedroc D. P., Cornish P. V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* 139(2):193-208.

42. Michiels P. J. A., *et al.* (2001) Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* 310(5):1109-1123.
43. Nixon P. L., *et al.* (2002) Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* 322(3):621-633.
44. Shen L. X., Tinoco I. (1995) The Structure of an Rna Pseudoknot That Causes Efficient Frameshifting in Mouse Mammary-Tumor Virus. *J. Mol. Biol.* 247(5):963-978.
45. Theimer C. A., Blois C. A., Feigon J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* 17(5):671-682.
46. Batey R. T., Gilbert S. D., Montange R. K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432(7015):411-415.
47. Serganov A., *et al.* (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* 11(12):1729-1741.
48. Paillart J. C., Marquet R., Skripkin E., Ehresmann C., Ehresmann B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* 78(7):639-653.
49. Jossinet F., *et al.* (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA* 5(9):1222-1234.
50. Ennifar E., Walter P., Ehresmann B., Ehresmann C., Dumas P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* 8(12):1064-1068.
51. Rist M. J., Marino J. P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* 41(50):14762-14770.
52. Tinoco I., Bustamante C. (1999) How RNA folds. *J. Mol. Biol.* 293(2):271-281.
53. Draper D. E. (1992) The Rna-Folding Problem. *Acc. Chem. Res.* 25(4):201-207.
54. Holbrook S. R., Kim S. H. (1997) RNA crystallography. *Biopolymers* 44(1):3-21.
55. Reyes F. E., Garst A. D., Batey R. T. (2009) Strategies in RNA crystallography. *Methods Enzymol.* 469:119-139.
56. Latham M. P., Brown D. J., McCallum S. A., Pardi A. (2005) NMR methods for studying the structure and dynamics of RNA. *Chembiochem* 6(9):1492-1505.
57. Yang S. C., Parisien M., Major F., Roux B. (2010) RNA Structure Determination Using SAXS Data. *J. Phys. Chem. B* 114(31):10039-10048.
58. Tullius T. D., Dombroski B. A. (1986) Hydroxyl Radical Footprinting - High-Resolution Information About DNA Protein Contacts and Application to Lambda-Repressor and Cro Protein. *Proc. Natl. Acad. Sci. U. S. A.* 83(15):5469-5473.
59. Tijerina P., Mohr S., Russell R. (2007) DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.* 2(10):2608-2623.
60. Mortimer S. A., Weeks K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129(14):4144-4145.
61. Gopinath S. C. (2009) Mapping of RNA-protein interactions. *Anal. Chim. Acta* 636(2):117-128.
62. Ryder S. P., Strobel S. A. (1999) Nucleotide analog interference mapping. *Methods* 18(1):38-50.

63. Tinoco I. (1960) Hypochromism in Polynucleotides. *J. Am. Chem. Soc.* 82(18):4785-4790.
64. Privalov P. L., Dragan A. I. (2007) Microcalorimetry of biological macromolecules. *Biophys. Chem.* 126(1-3):16-24.
65. Gilbert S. D., Love C. E., Edwards A. L., Batey R. T. (2007) Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry* 46(46):13297-13309.
66. Mathews D. H., Sabina J., Zuker M., Turner D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288(5):911-940.
67. Serra M. J., Turner D. H. (1995) Predicting thermodynamic properties of RNA. *Methods Enzymol.* 259:242-261.
68. Zuker M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13):3406-3415.
69. Coutts S. M. (1971) Thermodynamics and kinetics of G-C base pairing in the isolated extra arm of serine-specific transfer RNA from yeast. *Biochim. Biophys. Acta* 232(1):94-106.
70. Gralla J., Crothers D. M. (1973) Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J. Mol. Biol.* 73(4):497-511.
71. Porschke D. (1974) Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem.* 1(5):381-386.
72. Menger M., Eckstein F., Porschke D. (2000) Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry* 39(15):4500-4507.
73. Proctor D. J., *et al.* (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(44):14004-14014.
74. Ma H., *et al.* (2006) Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.* 128(5):1523-1530.
75. Stancik A. L., Brauns E. B. (2008) Rearrangement of partially ordered stacked conformations contributes to the rugged energy landscape of a small RNA hairpin. *Biochemistry* 47(41):10834-10840.
76. Kuznetsov S. V., Ren C. C., Woodson S. A., Ansari A. (2008) Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* 36(4):1098-1112.
77. Sarkar K., Meister K., Sethi A., Gruebele M. (2009) Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe. *Biophys. J.* 97(5):1418-1427.
78. Sarkar K., Nguyen D. A., Gruebele M. (2010) Loop and stem dynamics during RNA hairpin folding and unfolding. *RNA* 16(12):2427-2434.
79. Russell R., Herschlag D. (2001) Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J. Mol. Biol.* 308(5):839-851.
80. Zarrinkar P. P., Williamson J. R. (1994) Kinetic intermediates in RNA folding. *Science* 265(5174):918-924.
81. Zarrinkar P. P., Williamson J. R. (1996) The kinetic folding pathway of the Tetrahymena ribozyme reveals possible similarities between RNA and protein folding. *Nat. Struct. Biol.* 3(5):432-438.

82. Sclavi B., Sullivan M., Chance M. R., Brenowitz M., Woodson S. A. (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279(5358):1940-1943.
83. Pan T., Sosnick T. R. (1997) Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat. Struct. Biol.* 4(11):931-938.
84. Thirumalai D., Woodson S. A. (2000) Maximizing RNA folding rates: a balancing act. *RNA* 6(6):790-794.
85. Wan Y., Suh H., Russell R., Herschlag D. (2010) Multiple unfolding events during native folding of the Tetrahymena group I ribozyme. *J. Mol. Biol.* 400(5):1067-1077.
86. Rook M. S., Treiber D. K., Williamson J. R. (1998) Fast folding mutants of the Tetrahymena group I ribozyme reveal a rugged folding energy landscape. *J. Mol. Biol.* 281(4):609-620.
87. Rangan P., Masquida B., Westhof E., Woodson S. A. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* 100(4):1574-1579.
88. Behrouzi R., Roh J. H., Kilburn D., Briber R. M., Woodson S. A. (2012) Cooperative tertiary interaction network guides RNA folding. *Cell* 149(2):348-357.
89. Fang X. W., Pan T., Sosnick T. R. (1999) Mg2+-dependent folding of a large ribozyme without kinetic traps. *Nat. Struct. Biol.* 6(12):1091-1095.
90. Woodson S. A. (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc. Chem. Res.* 44(12):1312-1319.
91. Russell R., *et al.* (2002) Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci. U. S. A.* 99(7):4266-4271.
92. Kwok L. W., *et al.* (2006) Concordant exploration of the kinetics of RNA folding from global and local perspectives. *J. Mol. Biol.* 355(2):282-293.
93. Buck J., Furtig B., Noeske J., Wohnert J., Schwalbe H. (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* 104(40):15699-15704.
94. Lee M. K., Gal M., Frydman L., Varani G. (2010) Real-time multidimensional NMR follows RNA folding with second resolution. *Proc. Natl. Acad. Sci. U. S. A.* 107(20):9192-9197.
95. Laederach A., Shcherbakova I., Liang M. P., Brenowitz M., Altman R. B. (2006) Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule. *J. Mol. Biol.* 358(4):1179-1190.
96. Lemay J. F., Penedo J. C., Tremblay R., Lilley D. M., Lafontaine D. A. (2006) Folding of the adenine riboswitch. *Chem. Biol.* 13(8):857-868.
97. Brenner M. D., Scanlan M. S., Nahas M. K., Ha T., Silverman S. K. (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry* 49(8):1596-1605.
98. Zlatanova J., Lindsay S. M., Leuba S. H. (2000) Single molecule force spectroscopy in biology using the atomic force microscope. *Prog. Biophys. Mol. Biol.* 74(1-2):37-61.
99. Neuman K. C., Nagy A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* 5(6):491-505.

100. Marsden S., Nardelli M., Linder P., McCarthy J. E. (2006) Unwinding single RNA molecules using helicases involved in eukaryotic translation initiation. *J. Mol. Biol.* 361(2):327-335.
101. Abels J. A., Moreno-Herrero F., van der Heijden T., Dekker C., Dekker N. H. (2005) Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* 88(4):2737-2744.
102. Lipfert J., *et al.* (2014) Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* 111(43):15408-15413.
103. Smith S. B., Cui Y., Bustamante C. (2003) Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* 361:134-162.
104. Tinoco I., Jr., Li P. T., Bustamante C. (2006) Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.* 39(4):325-360.
105. Li P. T., Collin D., Smith S. B., Bustamante C., Tinoco I., Jr. (2006) Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* 90(1):250-260.
106. Ha T., *et al.* (1996) Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U. S. A.* 93(13):6264-6268.
107. Roy R., Hohng S., Ha T. (2008) A practical guide to single-molecule FRET. *Nat. Methods* 5(6):507-516.
108. Ashkin A. (1970) Acceleration and Trapping of Particles by Radiation Pressure. *Phys. Rev. Lett.* 24(4):156-&.
109. Chu S., Bjorkholm J. E., Ashkin A., Cable A. (1986) Experimental observation of optically trapped atoms. *Phys. Rev. Lett.* 57(3):314-317.
110. Neuman K. C., Block S. M. (2004) Optical trapping. *Rev. Sci. Instrum.* 75(9):2787-2809.
111. Moffitt J. R., Chemla Y. R., Smith S. B., Bustamante C. (2008) Recent advances in optical tweezers. *Annu. Rev. Biochem.* 77:205-228.
112. Ashkin A. (1997) Optical trapping and manipulation of neutral particles using lasers. *Proc. Natl. Acad. Sci. U. S. A.* 94(10):4853-4860.
113. Herbert K. M., Greenleaf W. J., Block S. M. (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* 77:149-176.
114. Svoboda K., Schmidt C. F., Schnapp B. J., Block S. M. (1993) Direct observation of kinesin stepping by optical trapping interferometry. *Nature* 365(6448):721-727.
115. Liphardt J., Onoa B., Smith S. B., Tinoco I., Jr., Bustamante C. (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science* 292(5517):733-737.
116. Woodside M. T., *et al.* (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001-1004.
117. Gebhardt J. C., Bornschlogl T., Rief M. (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. U. S. A.* 107(5):2013-2018.
118. Cluzel P., *et al.* (1996) DNA: an extensible molecule. *Science* 271(5250):792-794.
119. Smith S. B., Cui Y., Bustamante C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271(5250):795-799.

120. Smith D. E., *et al.* (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* 413(6857):748-752.
121. Smith S. B., Cui Y., Bustamante C. (2003) Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* 361:134-162.
122. Berg-Sorensen K., Flyvbjerg H. (2004) Power spectrum analysis for optical tweezers. *Rev. Sci. Instrum.* 75(3):594-612.
123. Li P. T., Collin D., Smith S. B., Bustamante C., Tinoco I., Jr. (2006) Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* 90(1):250-260.
124. Wen J. D., *et al.* (2007) Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys. J.* 92(9):2996-3009.
125. Woodside M. T., *et al.* (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U. S. A.* 103(16):6190-6195.
126. Kratky O., Porod G. (1949) Rontgenuntersuchung Geloster Fadenmolekule. *Recl. Trav. Chim. Pays-Bas* 68(12):1106-1122.
127. Fixman M., Kovac J. (1973) Polymer Conformational Statistics .3. Modified Gaussian Models of Stiff Chains. *J. Chem. Phys.* 58(4):1564-1568.
128. Kovac J., Crabb C. C. (1982) Modified Gaussian Model for Rubber Elasticity .2. The Wormlike Chain. *Macromolecules* 15(2):537-541.
129. Marko J. F., Siggia E. D. (1995) Stretching DNA. *Macromolecules* 28(26):8759-8770.
130. Bustamante C., Marko J. F., Siggia E. D., Smith S. (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599-1600.
131. Wang M. D., Yin H., Landick R., Gelles J., Block S. M. (1997) Stretching DNA with optical tweezers. *Biophys. J.* 72(3):1335-1346.
132. Seol Y., Skinner G. M., Visscher K. (2004) Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys. Rev. Lett.* 93(11):118102.
133. Chowrira B. M., Burke J. M. (1991) Binding and cleavage of nucleic acids by the "hairpin" ribozyme. *Biochemistry* 30(35):8518-8522.
134. Wilson T. J., Nahas M., Ha T., Lilley D. M. (2005) Folding and catalysis of the hairpin ribozyme. *Biochem. Soc. Trans.* 33(Pt 3):461-465.
135. Yakhnin A. V., Yakhnin H., Babitzke P. (2008) Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. *Proc. Natl. Acad. Sci. U. S. A.* 105(42):16131-16136.
136. Wilson K. S., Vonhippel P. H. (1995) Transcription Termination at Intrinsic Terminators - the Role of the Rna Hairpin. *Proc. Natl. Acad. Sci. U. S. A.* 92(19):8793-8797.
137. Malys N., McCarthy J. E. G. (2011) Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* 68(6):991-1003.
138. Wen J. D., *et al.* (2007) Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. *Biophys. J.* 92(9):2996-3009.
139. Collin D., *et al.* (2005) Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* 437(7056):231-234.
140. Blanco M., Walter N. G. (2010) Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol.* 472:153-178.
141. McKinney S. A., Joo C., Ha T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* 91(5):1941-1951.

142. Watkins L. P., Yang H. (2005) Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* 109(1):617-628.
143. Taylor J. N., Makarov D. E., Landes C. F. (2010) Denoising single-molecule FRET trajectories with wavelets and Bayesian inference. *Biophys. J.* 98(1):164-173.
144. Taylor J. N., Landes C. F. (2011) Improved resolution of complex single-molecule FRET systems via wavelet shrinkage. *J. Phys. Chem. B* 115(5):1105-1114.
145. Tinoco I. (2004) Force as a useful variable in reactions: Unfolding RNA. *Annu. Rev. Biophys. Biomol. Struct.* 33:363-385.
146. Tinoco I., Jr., Bustamante C. (2002) The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys. Chem.* 101-102:513-533.
147. Tinoco I., Jr., Collin D., Li P. T. (2004) The effect of force on thermodynamics and kinetics: unfolding single RNA molecules. *Biochem. Soc. Trans.* 32(Pt 5):757-760.
148. Perrot P. (1998) *A to Z of thermodynamics* (Oxford University Press, Oxford ; New York) pp vi, 329 p.
149. Jarzynski C. (1997) Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78(14):2690-2693.
150. Crooks G. E. (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* 60(3):2721-2726.
151. Zuckerman D. M., Woolf T. B. (2002) Theory of a systematic computational error in free energy differences. *Phys. Rev. Lett.* 89(18):180602.
152. Gore J., Ritort F., Bustamante C. (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U. S. A.* 100(22):12564-12569.
153. Bennett C. H. (1976) Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* 22(2):245-268.
154. Bell G. I. (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342):618-627.
155. Evans E., Ritchie K. (1997) Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72(4):1541-1555.
156. Dudko O. K., Hummer G., Szabo A. (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* 96(10).
157. Dudko O. K., Mathe J., Szabo A., Meller A., Hummer G. (2007) Extracting kinetics from single-molecule force spectroscopy: nanopore unzipping of DNA hairpins. *Biophys. J.* 92(12):4188-4195.
158. Dudko O. K., Hummer G., Szabo A. (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.* 105(41):15755-15760.
159. Kramers H. A. (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7:284-304.
160. Jansson P. A. (1997) *Deconvolution of images and spectra* (Academic Press, San Diego) 2nd Ed pp xiv, 514 p.
161. Quarles K. A., *et al.* (2013) Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. *Biochemistry* 52(5):795-807.
162. Heinicke L. A., Nallagatla S. R., Hull C. M., Bevilacqua P. C. (2011) RNA helical imperfections regulate activation of the protein kinase PKR: effects of bulge position, size, and geometry. *RNA* 17(5):957-966.

163. Hanggi P., Talkner P., Borkovec M. (1990) Reaction-Rate Theory - 50 Years after Kramers. *Rev. Mod. Phys.* 62(2):251-341.
164. Schuler B., Lipman E. A., Eaton W. A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419(6908):743-747.
165. Junier I., Mossa A., Manosas M., Ritort F. (2009) Recovery of free energy branches in single molecule experiments. *Phys. Rev. Lett.* 102(7):070602.
166. Alemany A., Mossa A., Junier I., Ritort F. (2012) Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nat. Phys.* 8(9):688-694.
167. Eyring H. (1935) The activated complex in chemical reactions. *J. Chem. Phys.* 3(2):107-115.
168. Evans M. G., Polanyi M. (1935) Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* 31(1):0875-0893.
169. Grote R. F., Hynes J. T. (1980) The Stable States Picture of Chemical-Reactions .2. Rate Constants for Condensed and Gas-Phase Reaction Models. *J. Chem. Phys.* 73(6):2715-2732.
170. Curran D. R., Seaman L., Shockey D. A. (1987) Dynamic Failure of Solids. *Phys. Rep.* 147(5-6):253-388.
171. Schwesinger F., *et al.* (2000) Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proc. Natl. Acad. Sci. U. S. A.* 97(18):9972-9977.
172. Alemany A., Sanvicens N., de Lorenzo S., Marco M. P., Ritort F. (2013) Bond Elasticity Controls Molecular Recognition Specificity in Antibody-Antigen Binding. *Nano Lett.* 13(11):5197-5202.
173. Evans E. A., Calderwood D. A. (2007) Forces and bond dynamics in cell adhesion. *Science* 316(5828):1148-1153.
174. Bryngelson J. D., Wolynes P. G. (1989) Intermediates and Barrier Crossing in a Random Energy-Model (with Applications to Protein Folding). *J. Phys. Chem.* 93(19):6902-6915.
175. Sosnick T. R. (2008) Kinetic barriers and the role of topology in protein and RNA folding. *Protein Sci.* 17(8):1308-1318.
176. Zhang Y., Dudko O. K. (2013) A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc. Natl. Acad. Sci. U. S. A.* 110(41):16432-16437.
177. Garai A., Zhang Y., Dudko O. K. (2014) Conformational dynamics through an intermediate. *J. Chem. Phys.* 140(13):135101.
178. Nahvi A., *et al.* (2002) Genetic control by a metabolite binding mRNA. *Chem. Biol.* 9(9):1043.
179. Borovok I., Gorovitz B., Schreiber R., Aharonowitz Y., Cohen G. (2006) Coenzyme B12 controls transcription of the Streptomyces class Ia ribonucleotide reductase nrdABS operon via a riboswitch mechanism. *J. Bacteriol.* 188(7):2512-2520.
180. Warner D. F., Savvi S., Mizrahi V., Dawes S. S. (2007) A riboswitch regulates expression of the coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis: implications for differential methionine synthase function in strains H37Rv and CDC1551. *J. Bacteriol.* 189(9):3655-3659.

181. Winkler W. C., Cohen-Chalamish S., Breaker R. R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U. S. A.* 99(25):15908-15913.
182. Winkler W., Nahvi A., Breaker R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419(6910):952-956.
183. Mandal M., Boese B., Barrick J. E., Winkler W. C., Breaker R. R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* 113(5):577-586.
184. Johansen L. E., Nygaard P., Lassen C., Agerso Y., Saxild H. H. (2003) Definition of a second Bacillus subtilis pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL). *J. Bacteriol.* 185(17):5200-5209.
185. Mandal M., Breaker R. R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* 11(1):29-35.
186. Roth A., *et al.* (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* 14(4):308-317.
187. Meyer M. M., Roth A., Chervin S. M., Garcia G. A., Breaker R. R. (2008) Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA* 14(4):685-695.
188. Sudarsan N., *et al.* (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321(5887):411-413.
189. Mandal M., *et al.* (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694):275-279.
190. Grundy F. J., Lehman S. C., Henkin T. M. (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl. Acad. Sci. U. S. A.* 100(21):12057-12062.
191. Epshtein V., Mironov A. S., Nudler E. (2003) The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 100(9):5052-5056.
192. McDaniel B. A., Grundy F. J., Artsimovitch I., Henkin T. M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U. S. A.* 100(6):3083-3088.
193. Winkler W. C., Nahvi A., Sudarsan N., Barrick J. E., Breaker R. R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* 10(9):701-707.
194. Fuchs R. T., Grundy F. J., Henkin T. M. (2006) The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* 13(3):226-233.
195. Corbino K. A., *et al.* (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.* 6(8):R70.
196. Winkler W. C., Nahvi A., Roth A., Collins J. A., Breaker R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428(6980):281-286.
197. Cromie M. J., Shi Y., Latifi T., Groisman E. A. (2006) An RNA sensor for intracellular Mg(2+). *Cell* 125(1):71-84.
198. Baker J. L., *et al.* (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* 335(6065):233-235.

199. Klein D. J., Edwards T. E., Ferre-D'Amare A. R. (2009) Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* 16(3):343-344.
200. Kang M., Peterson R., Feigon J. (2009) Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell* 33(6):784-790.
201. Rieder U., Lang K., Kreutz C., Polacek N., Micura R. (2009) Evidence for pseudoknot formation of class I preQ1 riboswitch aptamers. *Chembiochem* 10(7):1141-1144.
202. Rieder U., Kreutz C., Micura R. (2010) Folding of a transcriptionally acting preQ1 riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* 107(24):10804-10809.
203. Jenkins J. L., Krucinska J., McCarty R. M., Bandarian V., Wedekind J. E. (2011) Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* 286(28):24626-24637.
204. Banas P., Sklenovsky P., Wedekind J. E., Sponer J., Otyepka M. (2012) Molecular mechanism of preQ1 riboswitch action: a molecular dynamics study. *J. Phys. Chem. B* 116(42):12721-12734.
205. Spitale R. C., Torelli A. T., Krucinska J., Bandarian V., Wedekind J. E. (2009) The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem.* 284(17):11012-11016.
206. Chung S. H., Kennedy R. A. (1991) Forward-backward non-linear filtering technique for extracting small biological signals from noise. *J. Neurosci. Methods* 40(1):71-86.
207. Ke A. L., Zhou K. H., Ding F., Cate J. H. D., Doudna J. A. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* 429(6988):201-205.
208. Rastogi T., Beattie T. L., Olive J. E., Collins R. A. (1996) A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.* 15(11):2820-2825.
209. Wang J. X., Lee E. R., Morales D. R., Lim J., Breaker R. R. (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol. Cell* 29(6):691-702.
210. Han K., Byun Y. (2003) PSEUDOVIEWER2: Visualization of RNA pseudoknots of any type. *Nucleic Acids Res.* 31(13):3432-3440.
211. Theimer C. A., Finger L. D., Trantirek L., Feigon J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* 100(2):449-454.
212. Theimer C. A., Giedroc D. P. (1999) Equilibrium unfolding pathway of an H-type RNA pseudoknot which promotes programmed -1 ribosomal frameshifting. *J. Mol. Biol.* 289(5):1283-1299.
213. Theimer C. A., Wang Y., Hoffman D. W., Krisch H. M., Giedroc D. P. (1998) Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. *J. Mol. Biol.* 279(3):545-564.
214. Nixon P. L., Giedroc D. P. (1998) Equilibrium unfolding (folding) pathway of a model H-type pseudoknotted RNA: the role of magnesium ions in stability. *Biochemistry* 37(46):16116-16129.
215. Qiu H., Kaluarachchi K., Du Z., Hoffman D. W., Giedroc D. P. (1996) Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA. *Biochemistry* 35(13):4176-4186.

216. Gluick T. C., Wills N. M., Gesteland R. F., Draper D. E. (1997) Folding of an mRNA pseudoknot required for stop codon readthrough: effects of mono- and divalent ions on stability. *Biochemistry* 36(51):16173-16186.
217. Gluick T. C., Gerstner R. B., Draper D. E. (1997) Effects of Mg2+, K+, and H+ on an equilibrium between alternative conformations of an RNA pseudoknot. *J. Mol. Biol.* 270(3):451-463.
218. Gluick T. C., Draper D. E. (1994) Thermodynamics of folding a pseudoknotted mRNA fragment. *J. Mol. Biol.* 241(2):246-262.
219. Puglisi J. D., Wyatt J. R., Tinoco I., Jr. (1988) A pseudoknotted RNA oligonucleotide. *Nature* 331(6153):283-286.
220. Ritchie D. B., Foster D. A., Woodside M. T. (2012) Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. U. S. A.* 109(40):16167-16172.
221. Chen G., Chang K. Y., Chou M. Y., Bustamante C., Tinoco I., Jr. (2009) Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* 106(31):12706-12711.
222. Hansen T. M., Reihani S. N. S., Oddershede L. B., Sorensen M. A. (2007) Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* 104(14):5830-5835.
223. Green L., Kim C. H., Bustamante C., Tinoco I. (2008) Characterization of the mechanical unfolding of RNA pseudoknots. *J. Mol. Biol.* 375(2):511-528.
224. Hengesbach M., Kim N. K., Feigon J., Stone M. D. (2012) Single-molecule FRET reveals the folding dynamics of the human telomerase RNA pseudoknot domain. *Angew. Chem. Int. Ed. Engl.* 51(24):5876-5879.
225. Chen G., Wen J. D., Tinoco I., Jr. (2007) Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *RNA* 13(12):2175-2188.
226. Cao S., Chen S. J. (2007) Biphasic folding kinetics of RNA pseudoknots and telomerase RNA activity. *J. Mol. Biol.* 367(3):909-924.
227. Xu X., Chen S. J. (2012) Kinetic mechanism of conformational switch between bistable RNA hairpins. *J. Am. Chem. Soc.* 134(30):12499-12507.
228. Cho S. S., Pincus D. L., Thirumalai D. (2009) Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proc. Natl. Acad. Sci. U. S. A.* 106(41):17349-17354.
229. Zhang Y., Zhang J., Wang W. (2011) Atomistic analysis of pseudoknotted RNA unfolding. *J. Am. Chem. Soc.* 133(18):6882-6885.
230. Biyun S., Cho S. S., Thirumalai D. (2011) Folding of human telomerase RNA pseudoknot using ion-jump and temperature-quench simulations. *J. Am. Chem. Soc.* 133(50):20634-20643.