ABSTRACT

Title of dissertation: STRETCHING BIOMOLECULES

Changbong Hyeon, Doctor of Philosophy, 2005

Dissertation directed by: Professor Devarajan Thirumalai
Institute for Physical Science and Technology,
Department of Chemistry and Biochemistry

Biomolecular self-assembly is the complicated processes characterized by broad and hierarchical structure of energy, length, and time scales. Various experimental tools have, for decades, been used to understand the principles governing the dynamics of biomolecular systems. The recent advent of single molecule force experiments has expanded our perspective on the energetics of biomolecules, explicitly showing the diverse traces of the individual molecules undergoing heterogeneous processes.

Based on theoretical arguments and Langevin dynamics simulations of coarse-grained models, I suggest that the following aspects of biomolecules can be elucidated through the force experiments. (1) The energy landscape roughness can be directly measured if the force experiment is performed in varying temperature. (2) The diverse nature of biomolecular energetics, reflecting the underlying complexity of energy

landscape, manifests itself when the molecule is subject to various conditions. As a simple example, thermodynamic and kinetic properties of two-state folding RNA hairpins are investigated at varying temperature and force. We show the thermal and the force induced unfolding/refolding dynamics are vastly different. (3) The free energy landscape of a molecule deforms its terrain differently depending on the nature of external control variable. Force plays a different role from temperature when it is exerted on the molecule. The explicit computations and comparisons of free energy profiles along the reaction coordinate find a Hammond behavior in force but not in temperature. (4) Native topology and polymeric nature of biomolecule determines the force-induced unfolding pathway. (5) There exist pulling speed-dependent unfolding pathways in biomolecules if the molecule consists of multiple subdomains.

When combined with theory, single molecule force experiments can unravel the rich nature of biomolecular free energy landscape. All of the theoretical predictions made in the present work are amenable to the future experiments.

STRETCHING BIOMOLECULES

by

Changbong Hyeon

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Commmittee:

Professor Devarajan Thirumalai, Chair/Advisor
Professor J. Robert Dorfman
Professor George H. Lorimer
Professor David Fushman
Professor Dorothy Beckett

PREFACE

This dissertation presents a collection of theoretical and computational works on mechanical stretching of biomolecules, a research project performed since the fall of 2001.

In Chapter 1, I presented the experimental and the theoretical backgrounds of force exepriments.

Chapter 2 suggests the experimental methodology of measuring the roughness of energy landscape of biomolecules using mechanical force. This suggestion was recently adopted by an experimental group (Ziv Reich) in the Weizmann Institute of Science to measure the energy landscape roughness of protein complex.

Chapter 3 and 4 discuss the mechanical unfolding of RNA hairpins by explicitly taking coarse-grained model of RNA for an example. Experimentally observed RNA hairpin dynamics under force are reproduced. Through phase diagram and free energy profiles obtained using the extensive sampling of RNA conformations with varying temperature and external force, we provide a number of new insights into RNA dynamics.

Lastly, in Chapter 5, I suggested a computational methodology to simulate a much bigger system than the 22-nt RNA discussed in Chapter 3 and 4. *Azoarcus* (195-nt) and *Tetrahymena* ribozyme (407-nt) are taken for examples. For multidomain proteins and RNAs, the pulling speed dependent unfolding pathways are manifested

At computationally minimized pulling speed, the unfolding pathway of the *Tetrahymena* ribozyme shows good agreement with that analyzed in force experiment by Bustamante *et. al.*

Chapter 2 and 3 are adapted from the paper previously published in *Proc. Natl. Acad. Sci.* C. Hyeon and D. Thirumalai **100** 10249 (2003) and *ibid* **102** 6789 (2005), respectively. Chapter 4 is adapted from the manuscript by C. Hyeon and D. Thirumalai (in preparation). Chapter 5 is adapted from the manuscript by C. Hyeon, R. I. Dima, and D. Thirumalai (in preparation).

ACKNOWLEDGMENTS

I should like to acknowledge the people who have helped me to complete my graduate work for last five years.

First of all I would like to thank my PhD advisor, Dave Thirumalai, for leading me to such a great opportunity to study this challenging field of scientific discipline. I learned from him what the important problems are in science and how I should ask questions in proper ways through the valuable conversations with him. The conversations include not only a discussion on specific research problems of mine in his office, but also lunch time chatting. In retrospect, I have to confess that not all conversations were useful especially when it comes to American politics, but I obtained many important ideas at flashing moments of chatting with him. I like his liberal style of understanding the complicated problem based on his broad knowledge and sharp intuition. I enjoyed the time spent with him.

I would like to thank Dmitri Klimov, now an assistant Professor in George Mason University, for patiently answering my questions regarding the biomolecular simulation in most straightforward way. I also would like to thank Ruxandra Dima, my longtime officemate who recently moved to Boston as an Assistant Professor in University of Massachusetts at Lowell. We worked together on extracting the stacking interaction parameter of RNA using statistical analysis of Protein Data Bank, which, I did not include in the current dissertation. This work has been published in *J.*

Mol. Biol **347** 53 (2005). I value all the scientific conversations with my colleagues, Ed O'brien, George Stan, Valeri Barsegov, David Pincus, Margaret Cheung, Greg Morrison, Vaithee Subramanian, and Jie Chen. I appreciate many useful discussions with a frequent visitor to our group, Mai Suan Li from the Polish Academy of Science.

Lastly, I would like to attribute the largest credit to Yumee, Hera, and all my family in Korea. Their love and dedication always gave me an enormous amount of power to overcome all the obstacles when I got exhausted.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Protein/RNA folding

Through longstanding evolutionary processes, biopolymers, represented by proteins and RNA, have been designed to organize themselves into the unique three dimensional native structure at physiological conditions. The acquisition of native structure is of utmost importance for the biological function and this process occurs in a fast and cooperative manner, which is the main feature that distinguishes biopolymers from random heteropolymers. In light of an astronomically large number of configurational states inherent to chain molecules composed of many atoms the ability of the molecule to find the unique configuration within such a short time scale ($< sec$) is remarkable. Levinthal [3] has described this extraordinary situation using a paradox. [1] However, RNA and protein are designed to avoid the Levinthal's scenario. The sequence of monomers composing the chain are nicely arranged so that the confluence of interactions is optimized for the fast search of the particular configuration. For biopolymers to reach the native state, complete sampling of phase space

[1] Suppose that each monomer can take maximally three different states for simplicity. If the chain size is $N = 100$ then there are $3^{100} \approx 10^{47}$ possible configurational states. Even at the search rate of $10^{13}/sec$, the search of one particular configuration, i.e., the native state can take more than 10^{27} years! [4]

is not required. The quest to understanding those evolutionarily selected collective interactions of biomolecules, conceptually called the energy landscape, has long been the subject of great interest to both theorists and experimentalists. What is the molecular mechanism underlying the folding process? How robust are biopolymers to mutations? What is the effect of solvent conditions on the folding? Various questions have been asked from different angles regarding biomolecular self-assembly. A number of experimental techniques traditionally used in the analysis of chemical compound have been adapted to resolve these questions.

1.2 Force as a variable in single molecule experiments

Recent progress in technology has led new methods to manipulate and probe a single molecule. For instance, the real-time dynamics of polymer can be monitored under various flows by attaching a fluorescence dye to the polymer [5, 6, 7, 8], the foldings of proteins [9] and RNA [10] are monitored using FRET (Förster resonant energy transfer), the magnetic tweezer can stretch and even twist double-stranded DNA to quantify the mechanical force in the biological system [11, 12, 13]. There is a fundamental advantage in probing a single molecule. Single molecule measurements can separate out the fluctuations of individual molecules from the ensemble average behavior observed in conventional experiments. Hence, one can elucidate the dynamical process, hidden from the perspective of ensemble measurement, using single molecule measurement technique [14]. One of the striking findings in single molecule experiments, especially on chain molecules, is that under the same external

condition the molecules with identical chemical components undergo an unexpectedly rich heterogeneous process. Time traces of molecules show large deviations from the ensemble averages. No single trajectory is identical to another. The energy landscape picture of biopolymers is glimpsed by tracing the motion of a large number of individual molecules under broadly varying external conditions.

Recent force experiments using the instruments such as the atomic force microscope (AFM) or laser optical tweezer (LOT) not only lie among those single molecule techniques but also introduced a new variable of interest to the study of biomolecular dynamics. Mechanical properties of single molecules have been studied for proteins, RNA, and DNA, providing glimpses of underlying energy landscapes. Following are the brief accounts of AFM and LOT, both of which are adapted to probe the mechanical response of biomolecules under force.

Atomic force microscope (AFM): AFM, initially invented to probe the surface, has been adapted to investigate the internal structure of single molecules. A molecule sitting in the solid support is attached to cantilever tip by biotin-streptavidin bond formation. As a result the two ends of the molecule are held between the experimental setup. The molecule is stretched by increasing the distance between the AFM cantilever and solid support. If the molecule resists the conformational change then the cantilever tip deflects from its equilibrium position, which is detected by the laser, and the corresponding force value is computed. In general, the resulting force versus extension curve is used for the molecular analysis. The effective spring constant of AFM cantilever tip is typically 1-100 pN/nm (see Fig.1.1-A).

Laser optical tweezer (LOT): Laser optical tweezers use the same principle

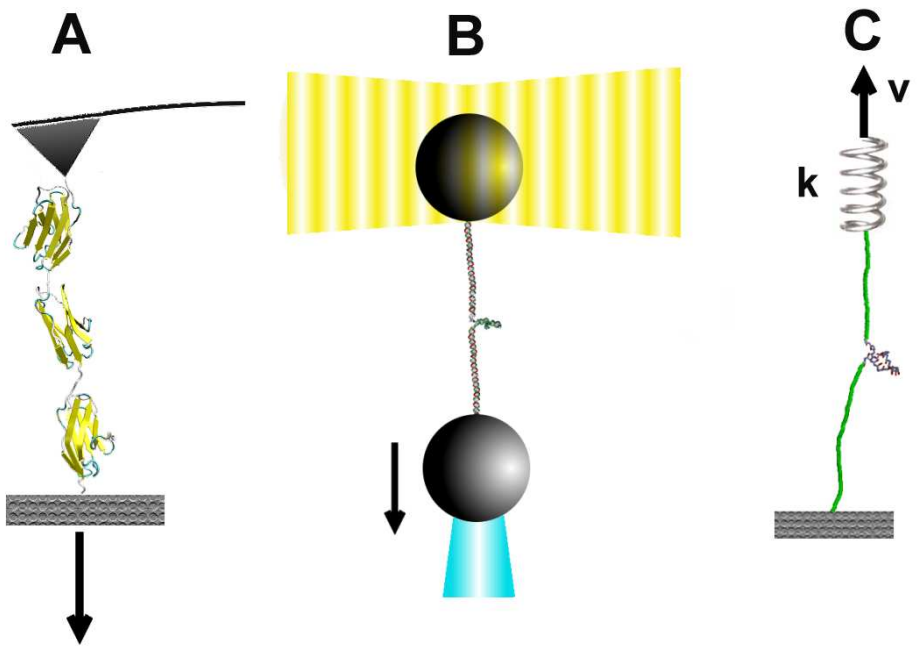

Figure 1.1: Illustrations of two major single molecule stretching techniques: Stretching of Ig27 domain of muscle titin using AFM (**A**) and P5abc domain of *Tetrahymena* RNA ribozyme using optical tweezer (**B**). Schematic representations of two experimental technique are similar with different spring constant k, which are depicted in **C**

as AFM. Instead of cantilever and solid support, two polystyrene beads are used to hold the molecule. One is optically trapped in the laser light and the other is sucked to the micropipette. The motion of micropipette is controlled by pietzoelectric actuator. Thus, as a function of extension the force value is recorded as proportional to the deviation of polystyrene bead position from the center of optical trap. The effective spring constant of optical trap is $0.001 - 0.1$ pN/nm (see Fig.1.1-B).

AFM and LOT are practically identical except that the effective spring constant (k) used to measure the force value is different from each setup (Fig.1.1-C). This difference in spring constant affect the precision of force or extension signal. Force is measured more precisely in LOT, but the insensitivity to the change of extension

can mask the minor unfolding event, a short-lived intermediate state for instance, of a target molecule. The typical saw-tooth pattern of titin unfolding observed in AFM is not seen in LOT (see [15]). Meanwhile, in AFM with a cantilever of hard spring constant, the large fluctuation in force value is inevitable.

1.3 Brief summary of force experiments on proteins and RNA

Studies of mechanical properties of single molecules by direct force application started in early the 1990s. The initial focus was on DNA because of its core function in all biological system [2]. DNA not only stores the genetic information of the organism but also shows remarkable conformational adaptation depending on conditions. The packed conformation as chromosome unfolds at the stage of transcription and replication. The volume change between folded and unfolded conformation of DNA is enormous. The surprising elasticity of DNA was investigated using optical tweezer. The elasticity coming from entropy fit to the wormlike chain model and gave a persistence length of 50 nm [16]. It turns out that ds-DNA at physiological conditions is very stiff in comparison to its diameter. Interestingly, under further stretching, DNA exhibited an additional sharp transition (B-S transition), leading to the discovery of another structure (S-DNA) [17].

The force experiments were extended to proteins in 1997. The muscle protein titin was an ideal subject for the experimentation because the mechanical properties of titin had direct relevance to its biological function. Both AFM and LOT were used to investigate the extension of titin as a function of applied force. Saw-tooth patterns of

force-extension curves and refolding upon stress relaxation suggest that titin protein is made of multiple domains (module) and that titin uses the multi-domain structure as a strategy (modularity) to obtain its large elasticity. More proteins have been studied including spectrin [18], T4 lysozyme [19], ubiquitin [20, 21], and RNase H [22].

The force experiments on RNAs have been conducted only recently. Bustamante and coworkers have presented their pioneering work on RNA by stretching P5ab and P5abc domain of *Tetrahymena thermophila* ribozyme using LOT [2, 23] which showed reversible transition between two discrete extensions at $f = 10 - 20\ pN$, and the ratio between residence times at two extension varied as a function of the force. They argued that the ergodicity of RNA hairpin dynamics was achieved by showing that the thermodynamic of single molecule was in good agreement with that of ensemble experiment. They also performed similar experiments on a whole *Tetrahymena* ribozyme molecule. Through the systematic comparison of mechanical response between the constituent domain and the whole molecule, the multiple force peaks on force-extension curve were identified in terms of the unfolding of each of particular structural domains in the molecule, so that the structural hierarchy of 400-nt ribozyme was successfully characterized using the force experiments.

The use of mechanical force to probe the biological system is not restricted to single molecules of RNA or a protein. The mechanical properties and dynamics of any biological systems, such as vesicle [24], cell membrane protein [25], DNA binding molecular motors [26, 27], virus [28], and etc, are the subjects of the investigation.

1.4 Theoretical background

One of the notable aspects in force experiments on single molecules is that end-to-end distance (R) is directly measurable or controlled by instrumentation. R becomes the natural reaction coordinate for describing the mechanical processes. Hence, the theory involving force experiments is testable and has been developed in a more practical manner whereas many theories on protein folding are rather elusive in practice because they use a difficult-to-measure reaction coordinate like a fraction of native contact.[2]

The theoretical framework to understand the role of force in experiment was first discussed in the context of cell-cell adhesion by Bell in 1978 [30]. Bell hypothesized that in the presence of external force the lifetime (τ) of a bond between cells attached to each other is written as Arrhenius type equation.

$$\tau = \tau_0 \exp\left[(E_0 - \gamma f)/k_B T\right] \tag{1.1}$$

where τ_0 is the inverse of the natural frequency of the atom in solids, E_0 is the bond energy, γ is the parameter to be empirically determined to account for the structure of the solids and its imperfection, and f is the applied force. Cell-cell adhesion characterized by bond energy E_0 is weakened by the amount of γf under constant force. Twenty years later, Bell's idea was revived with the development of

[2]Nowadays the number of native contact is not a totally elusive observable. Radical footprinting techniques [29] are used to probe the solvent accessibility of local region of nucleic acid structure, so that the time-resolved conformational dynamics can be monitored in terms of the number of native contact if this technique is combined with dynamics experiments.

dynamic force spectroscopy that was used for studying the ligand-receptor unbinding experiment. However, Bell's equation required some modification since Bell implied that Eq.(1.1) was based on traditional transition state theory in the gas phase [31] by assigning 10^{-13} *sec* to τ_0, which corresponds to $h/k_BT \approx 1.6 \times 10^{-13}$ *sec* (h is Planck's constant) the vibrational frequency of single bond. When a reaction takes place in condensed phase, Kramers showed that the reaction rate (k_R) should depend on the solvent viscosity (γ) and the angular frequency (curvature) at the reactant bottom (ω_0) and barrier top (ω_{ts}) of the effective reaction coordinate [32, 33].

$$k_R = \frac{\omega_0 \omega_{ts}}{2\pi\gamma} \exp\left(-E/k_BT\right) \tag{1.2}$$

where E is the activation free energy barrier. Usually, the prefactor gives $\frac{\omega_0\omega_{ts}}{2\pi\gamma} \sim 1\mu s$. [3] Biological reactions belong to Kramers' category. Evans combined Bell's idea with Kramers' equation [34], suggesting Kramers' equation under constant force.

$$k_R(f) = \frac{\omega_0 \omega_{ts}}{2\pi\gamma} \exp\left(-(E - f \cdot x_\beta)/k_BT\right). \tag{1.3}$$

In dynamic force spectroscopy the force, instead of being constant, is ramped with a constant loading rate ($r_f (= df/dt)$) and the most probable unbinding force (f^*) is

[3] There exists another fundamental difference between Eyring's rate equation and Kramers' equation other than the value of prefactor. *The prefactor in Kramers' rate contains the information on the shape of energy profile along the reaction coordinate* while the prefactor k_BT/h in Eyring's rate equation is simply a constant. It is plausible that the prefactor can vary depending on the external conditions. When mechanical force is acted on the system, the unfolding rate changes not only because of lowered energy barrier but also because of the deformation of energy profile. In principle, ω_0, ω_{ts}, and x_β should be functions of force. This point is discussed in Chapter 4 in the context of Hammond behavior under mechanical force.

obtained from the unbinding force distribution. Simple relation between f^* and r_f was established as

$$f^* = \frac{k_B T}{x_\beta} \log \frac{r_f x_\beta}{k_R k_B T}, \tag{1.4}$$

where x_β is the projected distance between bound state and transition state along the direction parallel to the applied force.[4] Meanwhile, there is an analogy between receptor-ligand unbinding and forced-unfolding of biomolecule. From the perspective of Kramers' theory, both unbinding and forced-unfolding are considered to be a noise-induced barrier crossing phenomena. Hence, Eq.(1.4) lent itself to the study of forced-unfolding experiments on proteins [35] and RNAs [2]. Structural information such as the distance between native conformation and transition state ensemble along the end-to-end distance is read by determining parameter x_β from the slope of f^* versus $\log r_f$ plot.

Recently, it has become possible to perform constant force experiments. Constant force experiments have shown the stepwise unzipping of ds-DNA and polyubiquitin, and the stochastic biphasic transition of the end-to-end distance of RNA molecule at the transition mid-force,

Accompanied with Eq.(1.4), the following force-extension relation for wormlike chain (WLC) model derived by Marko and Siggia are widely used in the study of force experiment on biopolymer.

$$f = \frac{k_B T}{l_p} \left(\frac{1}{4(1 - z/L)^2} - \frac{1}{4} + \frac{z}{L} \right) \tag{1.5}$$

where l_p is the persistence length and z is the extension. Eq.(1.5) was originally

[4] See Chapter 2 for the derivation.

used to obtain the persistence length of ds-DNA ($\sim$ 50 nm) [16], but because of its generality based on WLC model it has been used to extract both persistence length and contour length of proteins and RNA. One can find many examples of applying Eq.(1.5) in the fits to the force-extension curve with saw-tooth pattern that is obtained from the forced-unfolding of multidomain proteins [36, 18, 19]. The increase of contour length (ΔL) is quantified from one force peak to another, and the comparison of ΔL with the known structure can be used to decipher the structural change of the molecule under force.

1.5 Computational background

For the study of biological systems that are characterized by many levels of complexity, computer simulations serve as an excellent tool. The system of interest is prepared based on a well-defined energy Hamiltonian, and the evolution of the system is computed by solving the equation of motion under a certain external condition as a function of time. Under the assumption that the energy Hamiltonian properly captures the essence of the system it is expected that the resulting dynamics should reflect the corresponding dynamics of real system.

Deep insights into the protein folding problems in particular have been gained through a large number of simulations performed in last two decades [37, 38, 39]. The forced-unfolding process of biomolecule is the simple extension of the preexisting simulation strategies on the protein foldings. Stretching of biomolecules under both constant force and constant loading rate can be implemented by simply adding the

constraint on the one molecule at the end of the chain with fixing the position of first molecule in one place.

The steered molecular dynamics (SMD) that combines the stretching condition with the molecular dynamics was pioneered by Schulten and coworkers [40]. They simulated the forced-unfolding of titin immunoglobulin domains, I27 and I1, showing the atomic details of the molecular motion under force. The focus was on the rupture events of hydrogen bonds that stabilized the structures. The structural and energetic analysis enabled them to identify the origin of free energy barrier during the forced unfolding process. However, one has to notice that there is enormous difference between the simulation condition used in the SMD and real experiment. Typically SMD simulations at constant pulling speed use eight to ten orders of higher pulling speed, and one to two orders of larger spring constant than those of AFM experiments. Thus, effective force acting on the molecule at unit time is eight-orders larger. It is unlikely that the dynamics under such an extreme condition behaves the same way as the dynamics of real experiment. The large discrepancies of simulation condition from experiment stems from the broad time-scale spanned by molecular motions and the system size. The usual molecular dynamics can simulate the small-sized protein for up to tens of ns at most in realtime whereas meaningful chain dynamics of biopolymers takes place in at least μs. The time-scale for folding easily reaches the timescale of $> ms - s$. There exists an insurmountable barrier of time-scale between biomolecular dynamics and computation especially using the molecular dynamics simulations. To terminate the unfolding simulation within computationally accessible time SMD imposes an extreme condition onto the molecule. What the word "steered" in SMD

means that the simulation in SMD is indeed "steered" under extreme condition (large loading rate or high force).

The severe stretching condition in SMD simulations can be greatly alleviated by using a simplified model. First, all-atom structure is coarse-grained, the collisional effect of water is replaced by the random force of white noise, and Brownian dynamics algorithm takes care of the fluctuation-dissipation [41, 42, 43]. The basic strategy of simulational approaches used in this dissertation are based on the simplified model. Go potential[5] is adopted to optimize the native conformation. Although the coarse-grained models neglect the atomic details the chain dynamics under force is simulated in more reasonable conditions that those in SMD [46].

1.6 Overview

The following chapters present theoretical studies on various topics regarding the dynamics of biomolecules under force.

Chapter 2 provides the theoretical background for experimentally measuring the roughness of molecular energy landscape. In the presence of perturbative ruggedness ϵ on energy profile R. Zwanzig [47] showed that Kramers' rate equation is modified to have an additional $\langle \epsilon^2 \rangle / T^2$ temperature dependence term in the exponent. I rederived

[5]Go potential, which implements the idea of minimal frustration postulate in protein folding [44, 45], chooses the energy parameter in such a way that native contact is attractive and non-native contact is non-interactive or repulsive. This artificial choice of the parameter smooth the energetic ruggedness of energy landscape and prevents the molecule from being trapped in the local minima of $\epsilon > k_B T$

the Kramers' equation with ruggedness in the presence of mechanical force. Force experiment is ideal in measuring the energetics of the molecule because the extension of the molecule is directly measurable, from which the reaction coordinate of force experiment is simultaneously described. The value of roughness becomes the average roughness along the end-to-end distance coordinate. Performing the Monte Carlo simulation in roughness-imposed one-dimensional energy profile mimicking the force experiments I showed that $1/T^2$ dependence in both constant force and constant loading rate stretching should be in principle detectable with varying temperatures unless the signal-to-noise ratio is too small. Recently, Reich *et. al.* reported the first direct measurement of protein energy landscape roughness based on our suggestion. They used a protein complex consisting of the small GTPase Ran and the nuclear transport receptor importin-β and obtained $\epsilon \sim 5-6k_BT$ for the roughness amplitude [48].

In Chapter 3 and Chapter 4, I introduced the coarse-grained model of 22-nt RNA hairpin molecule to study the effect of mechanical force on RNA folding and unfolding dynamics. I modeled the nucleotides in RNA using three beads representing phosphate, ribose, and base. Go-potential was assigned on the beads for stabilizing the three dimensional native hairpin structure. By varying force and temperature, I monitored the thermodynamics and the kinetics of the RNA molecule. First, the thermodynamics was recapitulated on the force-temperature phase diagram, The co-operativity of folding transition, the deformation of energy profile under force, the forced unfolding and the force quench refolding in comparison to unfolding/refolding process in temperature variations were discussed in Chapter 3. Chapter 4 discusses

the practical issue of the role of linkers holding the molecule in force experiment. Especially, the two controversial issues of the energy profile deformation under force and the dynamics under force quench in Chapter 3 were revisited and were discussed in a thorough manner.

The last chapter is for the force simulation using the more simplified model than the one used in Chapter 3 and 4. The limitation of computational speed for simulating the larger system prompted me to devise *the minimal model* for the biopolymers. The basic strategy of the model is to retain only the chain connectivity and the interactions required for native topology of the molecule. I showed that these two factors were sufficient to simulate the various important features in forced unfolding dynamics. The new model (*topology model*) enabled me to simulate forced unfolding of 400-nt RNA, and gave a consistent result of unfolding pathway of the molecule with the recent force experiments. Additionally, the origin of pulling speed dependent unfolding pathways was clarified using the force propagation argument. I simulated the refolding of RNAs from stretched conformation, and showed that RNA folds from secondary to tertiary structure in a hierarchical manner. There was also a large time scale gap in the formation of secondary and tertiary contact. Despite the simplicity of neglecting most of the molecular details, topology model was able to qualitatively reproduce the main dynamics of RNA molecules.

Chapter 2

Measuring energy landscape roughness of proteins and RNA using mechanical unfolding experiments

Abstract

By considering temperature effects on the mechanical unfolding rates of proteins and RNA, whose energy landscape is rugged, the question posed in the title is answered in the affirmative. Adopting a theory of Zwanzig [Zwanzig, R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2029-2030] we show that, due to roughness that is characterized by an energy scale ϵ, the unfolding rate at constant force is retarded. Similarly, in non-equilibrium experiments done at constant loading rates the most probable unfolding force increases due to energy landscape roughness. The effects are dramatic at low temperatures. Our analysis suggests that, using temperature as a variable in mechanical unfolding experiments of proteins and RNA, the ruggedness energy scale ϵ can be directly measured.

2.1 Introduction

Visualizing folding of proteins [45, 38, 49] and RNA [38] in terms of the underlying energy landscape has been fruitful in anticipating diverse folding scenarios. The polymeric nature of these biomolecules and the presence of multiple energy scales associated with the building blocks of proteins and RNA make their underlying energy landscape rugged. By rugged we mean that the energy landscape consists of several minima separated by barriers of varying heights. To understand the diversity of the folding pathways it is necessary to determine the characteristics of the energy landscape experimentally.

It is difficult to represent the large dimensional space of the energy landscape in terms of a few parameters. However, the ruggedness of the energy landscape can be described in terms of an energy scale ϵ *regardless* of the precise nature of the underlying reaction coordinate. Within the energy landscape picture, it is clear that rapid folding to the native conformation on finite (biologically relevant) time scale is unlikely if $\epsilon/k_BT \gg 1$ (k_B is the Boltzmann constant and T is the temperature) [45]. Folding of proteins and RNA for which $\epsilon/k_BT \gg 1$ is, therefore, dominated by kinetic traps.

There are a number of proteins that fold kinetically in an apparent "two state" manner [50]. Similarly, t-RNA and independently folding subdomains of large RNA (for example, P5abc construct of *Tetrahymena thermophila* Group I intron) are also predicted to follow "two-state" kinetics [38]. However, due to the polymeric nature of proteins and RNA there is inherent incompatibility between local structures (α-

helical, β-strands in protein and helical "secondary" structures in RNA) and the global folds which results in topological frustration [38] even for "two-state" folders. This, in turn, can give rise to length scale-dependent ruggedness in the free energy landscape. For "two-state" folders the roughness arising from the presence of multiple energetic interactions is likely to be small i.e., $\epsilon/k_BT \sim O(1)$. By $O(1)$ we mean $0 < \epsilon/k_BT < 5$. Although no direct measurement of ϵ has been made for biomolecules, it has been inferred, using diffusion limited time scale for loop closure in model peptides, that $\epsilon/k_BT \approx 2$ [51].

The purpose of this chapter is to show that mechanical unfolding experiments can be used to extract ϵ. The stretching of proteins [15, 52, 53, 40, 54] and more recently RNA [2, 55, 23] already have begun to provide glimpses of their underlying energy landscape. In concert with theoretical studies [40, 56, 46] these measurements can be used to construct the free energy landscape of proteins and RNA. For example, it has been shown that, using dynamics of non-equilibrium stretching experiments, one can infer the distribution of free energy barriers in the *absence* of force [57]. This requires following time dependent events, at the single molecule level, when unfolding is induced by force. At present, laser optical tweezer and AFM methods can measure force(f)-extension(z) ($f - z$) curves either at constant f or as a function of loading rates. We show here that, using such experiments, ϵ can be directly measured provided the force-dependent unfolding rates can be measured over a range of temperatures.

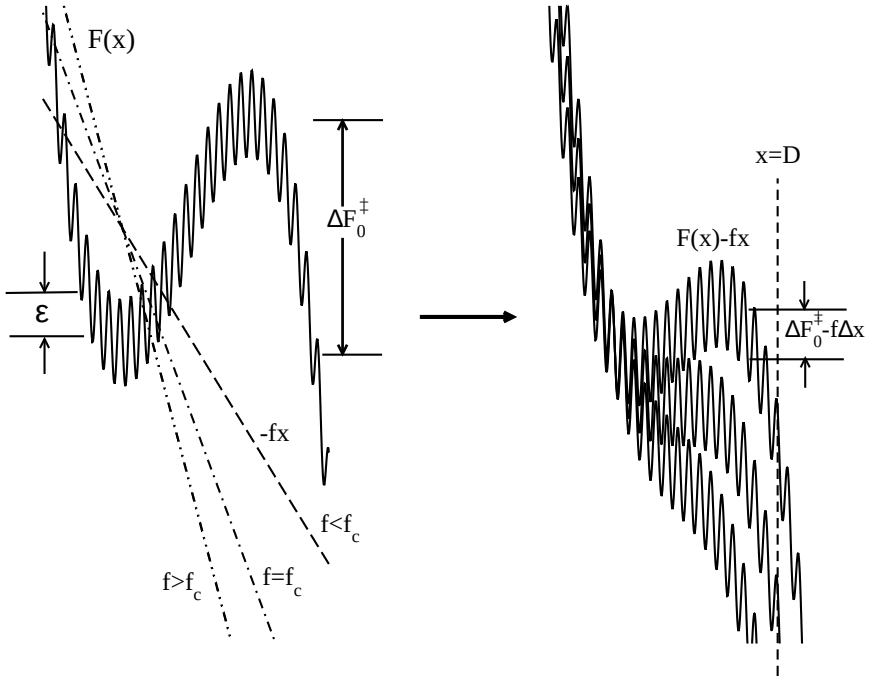

Figure 2.1: Caricature of the rough energy landscape of "two-state" proteins and RNA using x, the coordinate that couples to force f, as the reaction constant. For clarity we have represented the U state by an unbounded potential. The graph on the left side shows that on long length scale the landscape is smooth whereas on short length scales there there are a number of barriers. We used $F(x) = F_0(x) + F_1(x) = -Ax(x-p)(x+p) + \epsilon \cos 2\pi qx$ to generate the rugged energy landscape. The parameters in the potential are $A = 1pN/nm^2$, $p = 5nm$, $\epsilon = 3.0pN \cdot nm$, $q = 2.0(nm)^{-1}$. The random variable ϵ, q are assumed to be given by Gaussian distribution with dispersions $\sigma_q = 0.5(nm)^{-1}$ and $\sigma_\epsilon = 0.5pN \cdot nm$. Upon application of force(shown as -fx on left hand side) the landscape is deformed (see right hand side). The extent of tilting depends on the precise value of force. The tilted potential along with changes in the mean barriers are shown for three values, $f < f_c$, $f = f_c$ and $f > f_c$ where f_c the critical value of force is $25pN$ for the chosen parameters.

2.2 Caricature of energy landscape of RNA and proteins

We consider proteins and RNA that fold and unfold by an apparent "two-state" transition. Thus, the conformations of the biological molecules belong either to the native basin of attraction (NBA) or the unfolded basin of attraction (UBA). To describe the effect of force on RNA and proteins, we assume that the reaction $\mathbf{UBA} \rightleftharpoons \mathbf{NBA}$ may be described using a suitable low dimensional reaction coordinate. Although the

reduction of a multidimensional free energy landscape into one (or two) dimensions is a major simplification it has been shown that such a description gives nearly quantitative results for folding rate of minimally frustrated sequences [58]. When subject to force the one dimensional parameter, namely the extension of the molecule, is a natural reaction coordinate.

The free energy profile for **UBA** $\rightleftharpoons$ **NBA** in terms of the scalar reaction coordinate, x, is taken to be "smooth" on long [47] (greater than the persistence length of the polypeptide or polynucleotide) length scales. As stated in the Introduction, we expect that for "two-state" folders, due to topological frustration, there is a scale dependent roughness superimposed on the overall smooth free energy landscape. Following Zwanzig [47], we assume that the length scale for roughness is $\Delta x \ll R_g$(R_g is the dimension of RNA or protein of interest) and the overall energy scale of roughness is ϵ. Thus, on length scales comparable to Δx there are free energy barriers with typical barrier heights on the order of ϵ (Fig.2.1). Our calculations are only valid when $\epsilon/\Delta F^{\ddagger} < 1$ ($\Delta F^{\ddagger}$ is the average barrier height separating **UBA** from **NBA**). Zwanzig showed that diffusion in such a rough landscape can be extremely slow, especially at low temperature, even for modest values of ϵ. Here, we extend these calculations in the presence of applied force to show that ϵ can be measured using single molecule pulling experiments.

2.3 Unfolding at constant force

The energy landscape (see Fig.2.1 for a caricature) can be decomposed into a smooth part $F_0(x)$ on which is superimposed the roughness $F_1(x)$. If a suitable one dimensional reaction coordinate can describe mechanical unfolding of RNA and proteins then the force-dependent unfolding rate can be obtained by solving the Kramer's problem in the overdamped limit. Accordingly, we assume that diffusion in such an energy landscape is described by the Smoluchowski equation. The objective is to compute the mean first passage time (mfpt), $\langle t, x\rangle$, for transition from the NBA to UBA in the presence of force. The mfpt $\langle t, x\rangle$ is the mean time for reaching x starting from x_0. Assuming a reflecting barrier at $a(a < x_0 < x)$ Zwanzig showed (see also [59]) that

$$\langle t, x\rangle \cong \int_{x_0}^{x} dy e^{\beta F_0(y)+\psi^+(y)} \frac{1}{D} \int_a^y dz e^{-\beta F_0(z)+\psi^-(z)} \tag{2.1}$$

where D is the bare diffusion constant, $\beta = 1/k_BT$, and $e^{\psi^{\pm}(z)} \equiv \langle e^{\pm\beta F_1(z)}\rangle$ where $\langle \ldots \rangle$ is the spatial average of $F_1(z)$ over the length scale Δx. The structure of the equation for $\langle t, x\rangle$ suggests that there exists an effective Smoluchowski equation [47]

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x} D^*(x) e^{-\beta F^*(x)} \frac{\partial}{\partial x} e^{\beta F^*(x)} \rho(x,t) \tag{2.2}$$

that describes diffusion in a spatially averaged potential $F^*(x) = F_0(x) - k_BT\psi^-(x)$. In Eq.(2.2) $\rho(x,t)$ is the probability distribution function and $D^*(x) = e^{-\psi^+(x)} D e^{-\psi^-(x)}$ is the spatially averaged effective diffusion constant. Assuming that Eq.(2.2) is valid, we wish to solve the Kramer's problem [33] (computation of the unfolding rate of the native state) upon application of f for the energy landscape sketched in Fig.2.1. For the analysis to make sense $\epsilon/\Delta F^{\ddagger} < 1$ where $\Delta F^{\ddagger}$ is the value of the mean barrier

(Fig.2.1) that separates **UBA** from **NBA**. This will ensure that the biomolecule of interest would fold in an apparent two-state manner when $f = 0$. When $f \neq 0$ the effective potential is $F^*(x) = F_0(x) - fx - k_B T\psi^-$. By expanding this potential around the minimum x_0 and the transition state, x_{ts}, to quadratic order the unfolding rate can be calculated using Kramer's theory. Let ω_0 be the frequency of the minimum around the NBA and ω_{ts} be the frequency around saddle point (transition state). The unfolding rate is

$$k_R^{-1}(f) = \frac{2\pi\gamma}{\omega_0\omega_{ts}} \langle e^{-\beta F_1} \rangle \langle e^{\beta F_1} \rangle e^{\beta \Delta F_0^{\ddagger} - \beta f \cdot (x_{ts} - x_0)} \tag{2.3}$$

where $\Delta F_0^{\ddagger} = F_0(x_{ts}) - F_0(x_0)$ and γ is the friction coefficient. It follows from Zwanzig's analysis [47] that the effect of roughness manifests itself only in the renormalization of the effective diffusion constant. If we assume that the amplitude of roughness is independent of the position along the reaction coordinate and that it is distributed as a Gaussian then $\langle e^{-\beta F_1} \rangle = \langle e^{\beta F_1} \rangle$ so that

$$\log\left(\frac{k_R^{-1}(f)}{k_0^{-1}}\right) = \beta(\Delta F_0^{\ddagger} - f\Delta x) + \log \langle e^{\beta F_1} \rangle^2 \tag{2.4}$$

where $k_0^{-1} = \frac{2\pi\gamma}{\omega_0\omega_{ts}}$ and $\Delta x = x_{ts} - x_0$. The central result in Eq.(2.4) shows that roughness can lead to a non-Arrhenius temperature dependence in the mechanical unfolding rate. The result in Eq.(2.3) can also be extended to the case when there is memory in the friction i.e., when γ is time dependent. The Kramer's result for this case is $k_R(f) = [\frac{[\frac{\hat{\gamma}^2(\lambda_0^{ts})}{4} + \omega_{ts}^2]^{1/2} - \frac{\hat{\gamma}(\lambda_0^{ts})}{2}}{\omega_{ts}}] \frac{\omega_0}{2\pi} \langle e^{\beta F_1} \rangle^{-2} exp(-\beta(\Delta F_0^{\ddagger} - f\Delta x))$ where $\hat{\gamma}$ is Laplace transform of $\gamma(t)$ and λ_0^{ts} is the normal mode frequency defined at the TS [33]. For weak roughness $\beta F_1 < 1$ we get

$$\log\left(k_R^{-1}(f)/k_0^{-1}\right) = \beta(\Delta F_0^{\ddagger} - f\Delta x) + \beta^2 \langle F_1^2 \rangle \tag{2.5}$$

which might be more useful in analyzing experiments.

2.4 Stretching at constant loading rate

Mechanical unfolding experiments are typically performed far from equilibrium by stretching the ends of the molecule at a constant pulling speed. Large changes in the loading rate can be achieved by variations of the spring constant (related to the curvature of the optical trap potential in laser tweezer experiments) of the cantilever in AFM. To fully probe the energy landscape of proteins and RNA it is necessary to explore their response over a wide range of loading rates $r_f = df/dt = kv$ [34] where k is the spring constant of the cantilever and v is the pulling speed. By using a combination of laser optical tweezer, bioforce membrane probes, and AFM one can cover loading rates that vary by several orders of magnitude. In these experiments the histogram of forces required to unfold the biomolecule is measured by stretching one molecule at a time.

The effect of roughness on the dependence of the most probable unfolding force on the loading rate, r_f, can be calculated using Kramer's theory. Under constant loading rate, the curvatures of the barrier ω_{ts} as well as the barrier height, change as f increase with time. The unfolding rate changes dynamically as (see Eq.(2.3)).

$$k_R(f(t)) = \frac{\omega_0[f(t)]\omega_{ts}[f(t)]}{2\pi\gamma\langle e^{-\beta F_1}\rangle^2} e^{-\beta\{F_0(x_{ts}(f(t))-F_0(x_0(f(t)))\}+\beta f(t)(x_{ts}(f(t))-x_0(f(t)))} \tag{2.6}$$

The time dependent probability of unfolding is $P(t) = k(t)exp(-\int_0^t d\tau k(\tau))$. Using $r_f = df/dt$ the distribution of forces for overcoming the barrier is

$$P(f) = \frac{1}{r_f} k_R(f) \exp\left(-\int_0^f df' \frac{1}{r_f} k_R(f')\right) \tag{2.7}$$

The most probable threshold force that drives unfolding is obtained using $dP(f)/df|_{f=f^*} = 0$ which leads to

$$\begin{aligned} f^* &= \frac{k_BT}{\Delta x(f^*)}\{\log\left(\frac{r_f\Delta x(f^*)}{\nu_D(f^*)e^{-\beta\Delta F_0^{\ddagger}(f^*)}k_BT}\right) \\ &+ \log\left(1+f^*\frac{\Delta x'(f^*)}{\Delta x(f^*)}-\frac{\left(\Delta F_0^{\ddagger}\right)'(f^*)}{\Delta x(f^*)}+\frac{\nu_D'(f^*)}{\nu_D(f^*)}\frac{k_BT}{\Delta x(f^*)}\right) \\ &+ \log\langle e^{\beta F_1}\rangle^2\} \end{aligned} \tag{2.8}$$

where $\Delta F_0^{\ddagger}(f) \equiv F_0(x_{ts}(f)) - F_0(x_0(f))$, $'$ denotes differentiation with respect to the argument, $\Delta x(f) \equiv x_{ts}(f) - x_0(f)$ is distance between the transition state and the native state, and $\nu_D(f) \equiv \frac{\omega_0(f)\omega_{ts}(f)}{2\pi\gamma}$. If $\Delta F_0^{\ddagger}(f)$, $\Delta x(f)$, and the intrinsic angular frequency at the top and the bottom of barrier are relatively insensitive to the change of force the second term on the right hand side of Eq.(2.8) would vanish. This leads to $f^* \propto \frac{k_BT}{\Delta x}\log r_f$ [34]. However, if the loading rate spans a wide range, the resulting f^* can deviate substantially from the predicted logarithmic dependence on r_f.

It follows from Eq.(2.8) that the effect of roughness on f^* is similar to that shown in Eq.(2.4). As the extent of roughness increases the value of f^* can increase substantially. To use Eq.(2.8) to estimate $\langle e^{\beta F_1}\rangle^2$ shifts in f^* due to roughness must exceed the dispersions in $P(f)$ when $\epsilon \approx 0$.

2.5 Numerical results

The theory presented here shows that using constant force mechanical unfolding experiments the magnitude of the roughness can be measured provided temperature can be varied. Our central result has been derived by assuming that in the presence

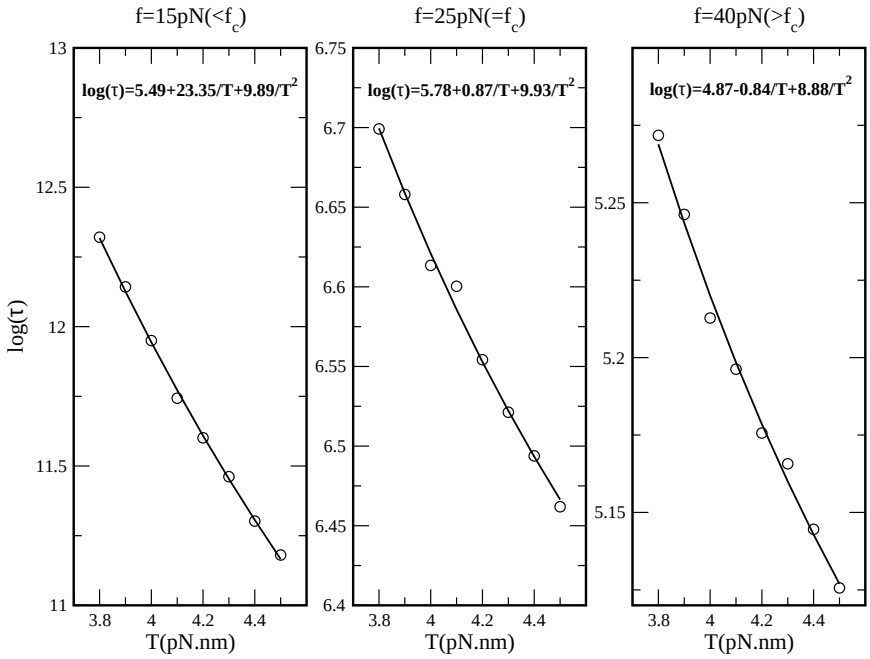

Figure 2.2: The logarithm of the mean first passage time, $\log\langle\tau\rangle$, as a function of temperature at three values of f. The symbols are numerical results and the lines are for using $\log\langle\tau\rangle = a + b/T + c/T^2$ (Eq.(2.5)).

of force the Smoluchowski equation (Eq.(2.2)) is valid for the roughness averaged potential. To test the validity of this assumption we have performed Monte Carlo simulations for a model potential $F(x) = F_0(x) + F_1(x) = -Ax(x-p)(x+p) - f(x+p/\sqrt{3}) + \epsilon\cos(2\pi qx)$. The strength of the random potential ϵ and its range q^{-1} are taken from a Gaussian distribution with zero mean. For this potential, with the parameters given in the caption to Fig.2.1, we calculated the mean unfolding time for a number of realizations of the random potential. The mean unfolding time is computed from the histogram of the first passage time which is the first time the particle reaches the boundary at $x = D = 1.5p$. For a given force, the mfpt is computed as a function of temperature in the range of $3.80pN \cdot nm \leq k_BT \leq 4.50pN \cdot nm$. The numerical simulations were performed using the smart Monte Carlo algorithm [60]

with parameters, $\gamma = 4.14 \times 10^{-8} pN \cdot sec/nm$, and time step, $dt = 2.0 \times 10^{-10} sec$.

In accord with our prediction(Eq.(2.4)) we find that the numerically computed logarithm of mfpt, $\log \tau$, as a function of $k_B T$ for the three values of f, can be fit as $\log \tau = a + b/k_B T + c/k_B^2 T^2$ (Fig.2.2). If the theoretical prediction is valid then we expect that the value of $c = \epsilon^2$ which for our model is $\epsilon^2 = 9.0 pN^2 \cdot nm^2$ independent of f. We have used Eq.(2.5) instead of Eq.(2.4) because the value of ϵ is less than the temperature used in the simulation i.e., $\epsilon/k_B T \sim O(1)$. The numerical fits give ϵ^2=9.9, 9.9, and 8.9 for $15pN(< f_c)$, $25pN(= f_c)$, $40pN(> f_c)$ respectively. The critical force f_c for our choice of parameters is $25pN$. The coefficient $b = \beta(\Delta F_0^{\ddagger} - f\Delta x)$ should be equal to 24.3, 0.0, 0.0 for $15pN(< f_c)$, $25pN(= f_c)$, $40pN(> f_c)$. The fits give b = 23.4, 0.9, -0.8, respectively. Thus, the numerical results confirm the validity of Eq.(2.5) in the limit when the roughness is small. If $\epsilon/k_B T \gg 1$ then the temperature dependence of force-induced unfolding would have significant curvature as can be seen by a cumulant expansion of $\log \langle e^{\beta F_1} \rangle^2$.(Eq.(2.4))

2.6 Proposed experiments

The theoretical considerations presented here allow us to propose mechanical unfolding experiments that can be used to measure a key statistical characteristic of the energy landscape of proteins and RNA, namely the roughness scale ϵ. Measurements of the unfolding rates as a function of temperature with the force held constant can be used to estimate ϵ (Eqs.(2.3) and (2.5)). There is a significant advantage in the proposed experiments to measure ϵ. If the theoretical predictions hold

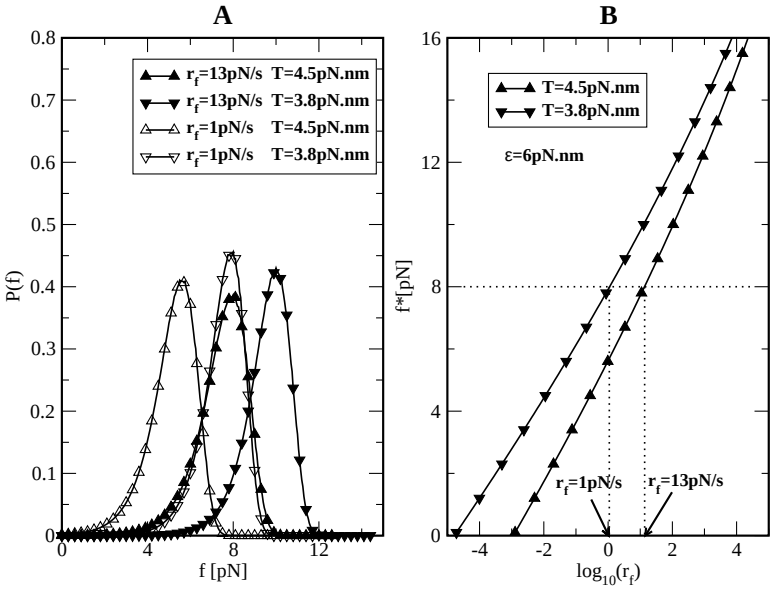

Figure 2.3: (A): Distribution of unfolding forces $P(f)$ at two temperatures and two loading rates for the potential described in the caption to Fig.2.1. The parameters are the same as in Fig.2.2. (B): Dependence of f^*, the most probable unfolding force, as a function of the loading rate r_f at two temperatures. The procedure to extract ϵ from these curves is described in the text.

good then measurements of ϵ using single molecule mechanical unfolding experiments would not involve any assumption about the underlying folding reaction coordinates. Moreover, no modeling of the dynamics of proteins and RNA is required to extract ϵ from mechanical unfolding experiments.

Using Eqs.(2.3) and (2.4) to measure ϵ is most efficacious if temperature variation of k_0^{-1} does not compromise the signal arising from the third term in Eq.(2.4) or Eq.(2.5). Because $k_0^{-1} \propto \eta$, the viscosity of the solution, corrections due to variations of η with respect to T have to be taken into account in using Eq.(2.3) to analyze experiments. The variable temperature range in mechanical unfolding experiments is likely to be between $5^\circ C < T < 50^\circ C$. Over this temperature range η for water varies as $\exp(A/T)$ [61]. As a result the coefficient of the $1/T^2$ term in $\log k_R^{-1}(f)$

can be estimated using mechanical unfolding experiments.

Let us consider how to use mechanical unfolding experiments at constant loading rate to obtain ϵ. A straightforward application of Eq.(2.8) to measure ϵ is difficult because a number of variables in Eq.(2.8) are functions of f^*. A plausible way of using Eq.(2.8) to estimate ϵ is illustrated in Fig.2.3, which shows $P(f)$ and f^* as function of loading rate at two different temperatures. Using such two dimensional experiments (done by varying r_f and T) one can obtain the values of r_f that give rise to identical f^* value at two temperatures. The measured r_f values at temperature T_1 and T_2 allows us to estimate of ϵ^2 (for small ϵ/k_BT) using

$$\epsilon^2 \approx \frac{k_BT_1 \times k_BT_2}{k_BT_1 - k_BT_2}\left(k_BT_1 \log \frac{r_f(T_1)\Delta x(f^*)}{\nu_D(f^*)k_BT_1} - k_BT_2 \log \frac{r_f(T_2)\Delta x(f^*)}{\nu_D(f^*)k_BT_2}\right). \qquad (2.9)$$

In obtaining Eq.(2.9) we have assumed that the second term in Eq.(2.8) is small. The numerical results (Fig.2.3) for the energy landscape in Fig.2.1 shows that the separation in f^* as a function of temperature is greater at smaller loading rates that are easily accessible in laser optical tweezer experiments. Moreover, the dispersion in $P(f)$ is also smaller at smaller loading rates. These calculations show that measurements of $f^*(r_f, T)$ using optical tweezer experiments can be used to measure ϵ^2 using Eq.(2.9). [1]

[1]Following our suggestion, Reich *et. al.* performed AFM unbinding experiment on the protein complex consisting of GTPase Ran and importin-β at three different temperatures (7, 20, 32^oC). Assuming Δx depends on the temperature but not on the force, they modified Eq.2.9 to

$$\begin{aligned}\epsilon^2 &\approx \frac{\Delta x(T_1)k_BT_1 \times \Delta x(T_2)k_BT_2}{\Delta x(T_1)k_BT_1 - \Delta x(T_2)k_BT_2} \\ &\times \left[\Delta F_0^{\ddagger}\left(\frac{1}{\Delta x(T_1)} - \frac{1}{\Delta x(T_2)}\right) + \frac{k_BT_1}{\Delta x(T_1)} \log \frac{r_f(T_1)\Delta x(f^*)}{\nu_D(f^*)k_BT_1} - \frac{k_BT_2}{\Delta x(T_2)} \log \frac{r_f(T_2)\Delta x(f^*)}{\nu_D(f^*)k_BT_2}\right],\end{aligned}$$

2.7 Conclusions

We have shown that including temperature as a variable in mechanical unfolding experiments can yield valuable quantitative measurements of the characteristics of the energy landscape of proteins and RNA without any assumption about their dynamics. In our previous studies [57] we had shown that the phase diagram in the (f, T) plane of protein can reveal many features about intermediates in the folding/unfolding problem. Previous theoretical studies [40, 56, 46] and the present work show that variable temperature mechanical unfolding experiments have great potential in probing the energy landscape of biomolecule at the single molecule level.

(2.10)

and calculated ϵ from the nine pairs of point in f^* vs $\log r_{f^*}$, all of which gave consistent values of $\epsilon \sim 5 - 6k_BT$.

Chapter 3

Mechanical unfolding of RNA hairpins

Abstract

Mechanical unfolding trajectories, generated by applying constant force in optical tweezer experiments, show that RNA hairpins and the P5abc subdomain of the group I intron unfold reversibly. We use coarse-grained Go-like models for RNA hairpins to explore forced-unfolding over a broad range of temperatures. A number of predictions that are amenable to experimental tests are made. At the critical force the hairpin jumps between folded and unfolded conformations without populating any discernible intermediates. The phase diagram in the force-temperature (f, T) plane shows that the hairpin unfolds by an all-or-none process. The cooperativity of the unfolding transition increases dramatically at low temperatures. Free energy of stability, obtained from time averages of mechanical unfolding trajectories, coincide with ensemble averages which establishes ergodicity. The hopping time between the the native basin of attraction (NBA) and the unfolded basin increases dramatically along the phase boundary. Thermal unfolding is stochastic whereas mechanical unfolding occurs in *quantized steps* with great variations in the step lengths. Refolding times, upon force quench, from stretched states to the NBA is *at least an order of magnitude* greater than folding times by temperature quench. Upon force quench from stretched

states the NBA is reached in at least three stages. In the initial stages the mean end-to-end distance decreases nearly continuously and only in the last stage there is a sudden transition to the NBA. Because of the generality of the results we propose that similar behavior should be observed in force quench refolding of proteins.

3.1 Introduction

Unraveling the complexity of the energy landscape of RNA molecules requires exploration of their assembly and unfolding over a wide range of external conditions. In the last decade a combination of experiments, theoretical arguments, and simulations have been used to decipher the folding mechanisms of RNA molecules [62, 63, 64]. These studies have shown that RNA folding depends critically on a number of factors including valence and shape of counterions [65], and temperature. Somewhat more surprisingly recent experiments have shown that the folding mechanisms depend sensitively on the initial folding conditions [66]. In conventional experiments the difficult-to-characterize unfolded conformations are typically generated by altering temperature or by lowering the counterion concentration. In contrast, well-defined and vastly different initial conditions can be realized by applying force. Indeed, in remarkable experiments Bustamante and coworkers [2, 55] have generated mechanical unfolding trajectories for RNA hairpins and *T. thermophila* ribozyme. These experiments, which use constant external force to denature folded RNA, show that unfolding involves multiple routes in which a number of kinetic intermediates are sampled in the transition from the folded state to a stretched conformation [2, 55]. The lifetimes of the intermediates vary considerably, which is indicative of the large dispersion in the unfolding pathways. Thus, force unfolding is a powerful method to probe, at the single molecule level, regions of the energy landscape that are *inaccessible in conventional folding experiments*. In addition, to the importance of these experiments to map the RNA folding landscape response of RNA to locally applied

force may also be relevant in understanding cellular processes such as mRNA translocation through ribosomes, viral replication, and enzymatic activity of RNA dependent RNA polymerases.

In the force-induced unfolding experiments mechanical force, f, was applied using optical tweezers either to a part or to the whole *Tetrahymena* ribozyme assembly in differing ionic conditions. In their first report Liphardt *et al.* [2] showed that a simple hairpin, three helix junction, and the P5abc subdomain of the *Tetrahymena* ribozyme can fold reversibly when subject to a constant force. At the transition force the systems hop between folded and unfolded states. Assuming that the system is ergodic the dynamics of the reversible folding was used to calculate force-dependent equilibrium properties of the RNA constructs. These experiments established that f −a new variable to initiate unfolding− is a viable way to measure free energy difference between folded and unfolded states and to locate transition states with the mean extension of the molecule as a reaction coordinate.

Mechanical unfolding experiments on RNA have already lead to a number of theoretical studies [67, 68, 69, 70] that have addressed different aspects of forced-unfolding. Inspired by these experiments and building on previous theoretical works we report here the results for forced-unfolding of a small RNA hairpin using coarse-grained off-lattice simulations under varying forces and temperatures. We choose small hairpins for the preliminary study because they undergo reversible folding under force and they represent a basic subunit of large RNA assemblies.

We address the following questions: (1) What are the forced-unfolding pathways and how they differ from thermal denaturation? (2) How do the diagram of

states change as T and f are varied? (3) What are the differences in the time scales and pathways in force quench refolding and thermal refolding? We find that, just as in proteins [71], forced-unfolding occurs in quantized steps whereas the thermal unfolding is stochastic. Even for the simple hairpin we find a well-defined equilibrium phase diagram in the (f, T) plane in which hairpin states are separated by a phase boundary from the unfolded states. Surprisingly, when refolding is initiated by quenching to zero force from high forces, the folding occurs in *multiple stages with the initial compaction being nearly continuous*. Remarkably, the refolding time under force quench is nearly an order of magnitude greater than thermal refolding time.

3.2 Methods

3.2.1 Hairpin sequence

We have studied the thermal and forced-unfolding of a 22-nucleotide hairpin, P5GA, that is similar to P5ab in the P5abc domain of group I intron. Both these structures have GA mismatches and are characterized by the presence of GAAA tetraloop. The sequence of P5GA is GGCGAAGUCGAAAGAUGGCGCC and its NMR structure has been determined [1] (PDB id:1eor).

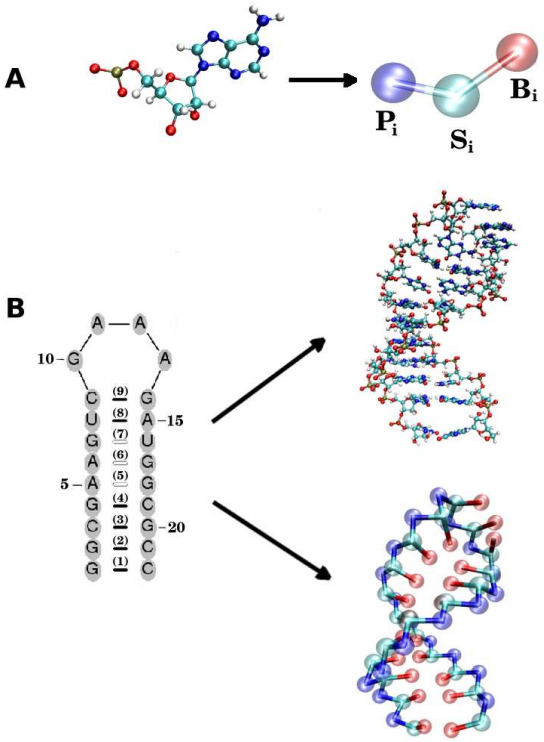

Figure 3.1: **A.** Coarse-grained representation of a nucleotide using three sites, namely, phosphate (P), sugar (S) and base (B) are given. **B.** The secondary structure of the 22-nt P5GA hairpin in which the bonds formed between base pairs are labeled from 1 to 9. The PDB structure [1] and the corresponding structure using the coarse grained model are shown on the right.

3.2.2 Model

Building on previous studies on proteins in our group [46] we introduce a coarse-grained off-lattice model of RNA by representing each nucleotide by three beads with interaction sites corresponding to phosphate group (P), ribose group (S), and the base (B) (see [72]). In this model the RNA backbone is reduced to the polymeric structure $[(P-S)_n]$ and the base is covalently linked to the ribose center. Thus, a RNA molecule with N nucleotides corresponds to $3N$ interaction centers. The potential energy of a conformation is written as $V_{TOT} = V_{BL} + V_{BA} + V_{DIH} + V_{STACK} + V_{NON} + V_{ELEC}$.

Harmonic potentials are used to maintain structural connectivity and rigidity on the distances and the angles formed between the covalently bonded beads. The connectivity between two beads (P_iS_i, S_iP_{i+1} and B_iS_i) is realized by the potential

$$V_{BL} = \sum_{i=1}^{2N-2} \frac{1}{2} k_r \{|\vec{r}_{(SP)_{i+1}} - \vec{r}_{(SP)_i}| - (R^o_{SP})_i\}^2 + \sum_{i=1}^{N} \frac{1}{2} k_r \{|\vec{r}_{B_i} - \vec{r}_{S_i}| - (R^o_{BS})_i\}^2 \quad (3.1)$$

where $k_r = 20kcal/(mol \cdot \mathring{A}^2)$. The notation $(SP)_i$, denoting i-th backbone bead S or P, is used for convenience. $(R^o_{SP})_i$ and $(R^o_{BS})_i$ are the distances between covalently bonded beads in PDB structure. In addition, the angle θ formed between three successive beads ($P_i - S_i - P_{i+1}$ or $S_{i-1} - P_i - S_i$) along sugar-phosphate backbone is subject to the bond-angle potential.

$$V_{BA} = \sum_{i=1}^{2N-3} \frac{1}{2} k_\theta (\theta_i - \theta^o_i)^2 \quad (3.2)$$

where $k_\theta = 20kcal/(mol \cdot rad^2)$. θ^o_i is the bond angle in the PDB structure.

V_{DIH} is the dihedral angle potential that describes the ease of rotation around the angle formed between four successive beads along the sugar-phosphate backbone ($S_{i-1}P_iS_iP_{i+1}$ or $P_iS_iP_{i+1}S_{i+1}$). The i-th dihedral angle ϕ_i is defined by $cos\phi_i = (\vec{r}_{i+1,i} \times \vec{r}_{i+1,i+2}) \cdot (\vec{r}_{i+2,i+1} \times \vec{r}_{i+2,i+3})$, which is the angle formed between the two planes defined by four successive beads i to $i+3$. At the level of the coarse-grained model the right-handed chirality of RNA is realized through this potential. Based on the angle determined from PDB structure (ϕ^o_i), one of the three types of dihedral potentials is assigned at each of four successive beads along the backbone : trans (t, $0 < \phi^o_i < 2\pi/3$), gauche(+) (g^+, $2\pi/3 < \phi^o_i < 4\pi/3$), gauche($-$) ($g^-$, $4\pi/3 < \phi^o_i <$

2π). The total dihedral potential of the hairpin is written as

$$\begin{aligned} V_{DIH} &= \sum_{i=1}^{2N-4} [A^{\eta}_{1i} + B^{\eta}_{1i} + C^{\eta}_{1i} \\ &+ A^{\eta}_{2i} cos(\phi_i - \phi^o_i + \phi^{\eta}_i) + B^{\eta}_{2i} cos3(\phi_i - \phi^o_i + \phi^{\eta}_i) + C^{\eta}_{2i} sin(\phi_i - \phi^o_i + \phi^{\eta}_i)] \end{aligned} \tag{3.3}$$

where the parameters are defined for three types of angles as

$A_{1i} = 1.0,\ A_{2i} = -1.0,\ B_{1i} = B_{2i} = 1.6,\ C_{1i} = 2.0,\ C_{2i} = -2.0\ (\eta = g^+)$.

$A_{1i} = 1.0,\ A_{2i} = -1.0,\ B_{1i} = B_{2i} = 1.6,\ C_{1i} = 2.0,\ C_{2i} = 2.0\ (\eta = g^-)$.

$A_{1i} = A_{2i} = 1.2,\ B_{1i} = B_{2i} = 1.2,\ C_{1i} = C_{2i} = 0.0\ (\eta = t)$.

where the unit of parameters are in $kcal/mol$. To facilitate the folding we reduce the dihedral angle barrier especially on the loop region. All the parameters are halved in $19 \leq i \leq 24$.

Simple RNA secondary structures are stabilized largely by stacking interactions whose context dependent values are known [73, 74]. In the native state the P5GA hairpin has nine hydrogen bonds between the base pairs including two GA mismatch pairs [1]. The stacking interactions that stabilize a hairpin are $V_{STACK} = \sum_{i=1}^{n_{max}} V_i$ where $n_{max} = 8$ in P5GA. The orientational dependent terms V_i is taken to be

$$\begin{aligned} V_i(\{\phi\},\{\psi\},\{r\};T) = \Delta G_i(T) &\times e^{-\alpha_{st}\{sin^2(\phi_{1i}-\phi^o_{1i})+sin^2(\phi_{2i}-\phi^o_{2i})+sin^2(\phi_{3i}-\phi^o_{3i})+sin^2(\phi_{4i}-\phi^o_{4i})\}} \\ &\times e^{-\beta_{st}\{(r_{ij}-r^o_{1i})^2+(r_{i+1j-1}-r^o_{2i})^2\}} \\ &\times e^{-\gamma_{st}\{sin^2(\psi_{1i}-\psi^o_{1i})+sin^2(\psi_{2i}-\psi^o_{2i})\}} \end{aligned} \tag{3.4}$$

where $\Delta G(T) = \Delta H - T\Delta S$, the bond angles $\{\phi\}$ are $\phi_{1i} \equiv \angle S_i B_i B_j$, $\phi_{2i} \equiv \angle B_i B_j S_j$, $\phi_{3i} \equiv \angle S_{i+1} B_{i+1} B_{j-1}$, $\phi_{4i} \equiv \angle B_{i+1} B_{j-1} S_{j-1}$, the distance between two paired bases $r_{ij} = |B_i - B_j|$, $r_{i+1j-1} = |B_{i+1} - B_{j-1}|$, and ψ_{1i} and ψ_{2i} are the dihedral angles formed

by the four beads $B_iS_iS_{i+1}B_{i+1}$ and $B_{j-1}S_{j-1}S_jB_j$, respectively. The superscript o refers to angles and distances in the PDB structure. The values of α_{st}, β_{st} and γ_{st} are 1.0, 0.3Å^{-2} and 1.0 respectively. We take ΔH and ΔS from Turner's thermodynamic data set [74, 73]. There are no estimates for GA related stacking interactions, which typically do not form a stable bond and hence is considered a mismatch. Because of the absence of stacking parameters for the GA pair, we use the energy associated with GU in place of GA.

To mimic the hydrophobicity of purine/pyrimidine group, we use the Lennard-Jones interactions between non-bonded interaction centers. The total nonbonded potential is

$$V_{NON} = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} V_{B_iB_j}(r) + \sum_{i=1}^{N}\sum_{m=1}^{2N-1}{}' V_{B_i(SP)_m}(r) + \sum_{m=1}^{2N-4}\sum_{n=m+3}^{2N-1} V_{(SP)_m(SP)_n}(r) \quad (3.5)$$

where $r = |\vec{r}_i - \vec{r}_j|$, the prime in the second term on the Eq.(3.5) denotes the condition $m \neq 2i-1$. A native contact exists between two non-covalently bound beads provided they are within a cut-off distance r_c (=7.0Å). Two beads beyond r_c are considered to be non-native. For a native contact,

$$V_{\xi_i\eta_j}(r) = C_h^{\xi_i\eta_j}[(\frac{r_{ij}^o}{r})^{12} - 2(\frac{r_{ij}^o}{r})^6] \quad (3.6)$$

where r_{ij}^o is the distance between beads in PDB structure and $C_h^{\xi_i\eta_j} = 1.5kcal/mol$ for all native contact pairs except for $B_{10}B_{13}$ base pair associated with the formation of the hairpin loop, for which $C_h^{B_{10}B_{13}} = 3.0kcal/mol$. The additional stability for the base pair associated with loop formation is similar to the Turner's thermodynamic rule for the free energy gain in the tetraloop region. For beads beyond r_c the interaction

is

$$V_{\xi_i\eta_j}(r) = C_R[(\frac{a}{r})^{12} + (\frac{a}{r})^6] \tag{3.7}$$

with $a = 3.4$Å and $C_R = 1kcal/mol$. The value of C_h has been chosen so that the hairpin undergoes a first order transition from unfolded states. Our results are not sensitive to minor variations in C_h. The electrostatic potential between the phosphate groups is assumed to be pairwise additive $V_{ELEC} = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} V_{P_iP_j}(r)$. For $V_{P_iP_j}(r)$ we assume Debye-Hückel interaction, which accounts for screening by condensed counterions and hydration effects, and is given by

$$V_{P_iP_j} = \frac{z_{P_i}z_{P_j}e^2}{4\pi\epsilon_0\epsilon_r r}e^{-r/l_D} \tag{3.8}$$

where $z_{P_i} = -1$ is the charge on the phosphate ion, $\epsilon_r = \epsilon/\epsilon_0$ and the Debye length $l_D = \sqrt{\frac{\epsilon_r k_B T}{8\pi k_{elec} e^2 I}}$ with $k_{elec} = \frac{1}{4\pi\epsilon_0} = 8.99\times10^9 JmC^{-2}$. To calculate the ionic strength $I = 1/2\sum_i z_i^2 c_i$, we use the value $c_i = 200mM\text{-}NaCl$ from the header of PDB file [1]. We use $\epsilon_r = 10$ in the simulation [75]. Because the Debye screening length $\sim\sqrt{T}$ the strength of electrostatic interaction between the phosphate group is temperature dependent even when we ignore the variations of ϵ with T. At room temperature ($T \sim 300K$) the electrostatic repulsion between the phosphate groups at $r \sim$5.8Å, which is the closest distance between phosphate groups, is $V_{P_iP_{i+1}} \sim 0.5kcal/mol$. It follows that V_{ELEC} between phosphate groups across the base pairing ($r = 16 \sim 18\mathring{A}$) is almost negligible.

3.2.3 Simulations

The dynamics of stretching is obtained by integrating the Langevin equation. Forced-unfolding simulations are performed by applying a constant force to the S bead at one end of the molecule. Using a typical value for the mass of a bead in a nucleotide (B_i, S_i or P_i), m, $100g/mol \sim 160g/mol$, the average distance between the adjacent beads $a = 4.6$Å, the energy scale $\epsilon_h = 1 \sim 2kcal/mol$, the natural time is $\tau_L = (\frac{ma^2}{\epsilon_h})^{1/2} = 1.6 \sim 2.8ps$. We use $\tau_L = 2.0ps$ to convert simulation times into real times. To estimate the time scale for thermal and mechanical unfolding dynamics we use a Brownian dynamics algorithm [76, 77] for which the natural time for the overdamped motion is $\tau_H = \frac{\zeta\epsilon_h}{T}\tau_L$. We used $\zeta = 50\tau_L^{-1}$ in the overdamped limit, that approximately corresponds to friction constant in water. At $T = 290K$, 10^6 time steps correspond to $3.5\mu s$. To probe the thermodynamics and kinetics of folding we used a number of physical quantities (end-to-end distance (R), fraction of native contacts (Q), structural overlap function (χ), number of hydrogen bonds n_{bond}, etc) to monitor the structural change in the hairpin.

3.2.4 Phase diagram and free energy profile

The free energy profiles and the phase diagram were obtained using an adaptation of multiple histogram method [78] for force unfolding of biomolecules. For example, the thermodynamic average of the fraction of native contact, Q, can be extrapolated to the arbitrary external condition of T and f if the conformational states are well sampled at several conditions of (T,f). The thermodynamic average Q is

given as

$$\langle Q(T,f)\rangle = \frac{\sum_{E,R,Q} Q e^{-(E-fR)/T} \frac{\sum_{k=1}^{K} h_k(E,R,Q)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}}{\sum_{E,R,Q} e^{-(E-fR)/T} \frac{\sum_{k=1}^{K} h_k(E,R,Q)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}} \tag{3.9}$$

where K is the number of histograms, $h_k(E,R,Q)$ is the number of states between $E \sim E+\delta E$, $R \sim R+\delta R$, $Q \sim Q+\delta Q$ in the k-th histogram, $n_k = \sum_{E,R,Q} h_k(E,R,Q)$, T_k is the temperature of simulations, and f_k is the force of simulations. F_k is the free energy to be calculated self-consistently by

$$e^{-F_r/T_r} = \sum_{E,R,Q} e^{-(E-f_rR)/T_r} \frac{\sum_{k=1}^{K} h_k(E,R,Q)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}. \tag{3.10}$$

Using the low friction Langevin dynamics, we sampled the conformational states along $\{0K < T < 500K, f = 0.0pN\}$ and $\{0.0pN < f < 20.0pN, T = 305K\}$ lines with the exhaustive samplings around the transition regions at $\{305K \leq T \leq 356K$, $f = 0.0pN\}$ and $\{5.0pN \leq f \leq 7.0pN$, $T = 305K\}$.

The knowledge of free energy profile at (T,f) along the reaction coordinate can provide the succinct insight into the dynamic process.

The free energy profile with Q as an order parameter is given by

$$F[Q(T,f)] = F^o(T,f) - k_B T \log P[Q(T,f)]. \tag{3.11}$$

where $F^o(T,f) = -k_B T \log Z(T,f)$, $Z(T,f) = \sum_{E,R,Q} e^{-(E-fR)/T} \frac{\sum_{k=1}^{K} h_k(E,R,Q)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}$ and $P[Q(T,f)]$ is defined in Eq.(3.9). The free energy profile $F(R)$ with R as a reaction coordinate can be obtained using a similar expression.

3.3 Results and Discussion

3.3.1 Determination of the Native state

Using a combination of multiple slow cooling, simulated annealing, and steepest descent quenches we determined the native structure of the hairpin. To ensure that there is no other structure with lower energy, the structure obtained from steepest descent method is reheated to $T = 100K$ and cooled down again. By repeating this process we obtained the computed native conformation which has a RMSD of 0.1Å with respect to the PDB structure.. The bulk of the contribution to the total energy, $V_{TOT} = -154kcal/mol$, of the native conformation arises from $V_{STACK} = -95.5kcal/mol$, $V_{NON} = -59.2kcal/mol$.

3.3.2 Force-temperature (f,T) phase diagram

The diagram of states in the (f, T) plane shows that P5GA hairpin behaves as a "two-state" folder (Fig.3.2). In the absence of force ($f = 0\ pN$) the folding/unfolding transition midpoint is at $T_m = 341K$ using $\langle Q \rangle$ as an order parameter. At $T = 290K$ the equilibrium force, required to unfold the P5GA is about $7pN$ (Fig.3.2), which is half the value for unfolding P5ab. The difference is, in all likelihood, due to the smaller length of P5GA. As force increases, T_F decreases monotonically, so that the transition midpoints (T_m, f_m) form a phase boundary separating the folded ($\langle Q \rangle > 0.5$ and $\langle R \rangle < 3nm$) and the unfolded states. The phase boundary is sharp at low T_m and large f_m, but is fuzzy when the force is weak. The locus of points separating the

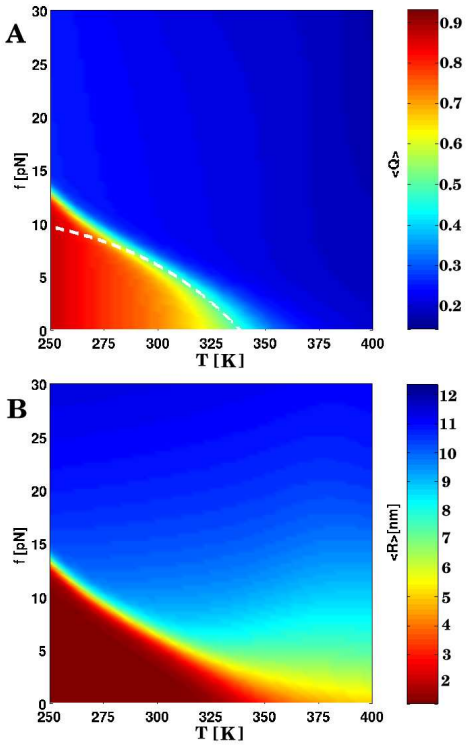

Figure 3.2: Phase diagram for the P5GA hairpin. **A.** This panel shows the diagram of states obtained using the fraction of native contacts as the order parameter. The values of the thermal average of the fraction of native contacts, $\langle Q \rangle$, are color coded as indicated on th e scale shown on the right. The dashed line is a fit using Eq.(3.12) to the locus of points in the (f, T) plane that separates the folded hairpin from the unfolded states. **B.** Plot of the phase diagram in the (f, T) plane using the mean end-to-end distance $\langle R \rangle$ as the order parameter. Although the diagram of states is qualitatively similar as in **A** there are quantitative differences in estimates of T_m at $f = 0$. However, estimates of threshold force values at $T < T_m$ are similar in both **A** and **B**.

unfolded and folded states can be fit using

$$f_c \sim f_o \left(1 - \left(\frac{T}{T_m}\right)^{\alpha}\right) \tag{3.12}$$

where f_o is the critical force at the low temperature and $\alpha(=6.4)$ is a sequence dependent exponent. The large value of α is indicative of a weak first order transition separating the hairpin and unfolded states [71].

3.3.3 Two state dynamics and equilibrium

We used the thermodynamic relation $\log K_{eq}(f) = -\Delta F_{UF}/k_BT + f \cdot \Delta x_{UF}/k_BT$ and the dependence of $\log K_{eq}$ (K_{eq} is computed as time averages of the traces in Fig.3.3-A) on f to estimate ΔF_{UF} and Δx_{UF} which is the equilibrium distance separating the native basin of attraction (NBA) and the basin corresponding to the ensemble of unfolded states (UBA). The transition midpoint ($K(f_m) = 1$) gives $f_m \approx 6pN$ in excellent agreement with the value obtained from the equilibrium phase diagram (Fig.3.2-A) which establishes ergodicity. From the slope, $\frac{\partial \log K_{eq}(f)}{\partial f} = 1.79pN^{-1}$, $\Delta x_{UF} \approx 7.5nm$ we found, by extrapolation to $f = 0$, that $\Delta F_{UF} \approx 6.2kcal/mol$ under the assumption that Δx_{UF} is *constant and independent of f*.

The independence of Δx_{UF} on f was also used by Liphardt *et al.* [2] to estimate ΔF_{UF}. To check the validity of this assumption we computed free energy profiles using the multiple histogram method with R as the progress variable. At $T = 305K$, we find, from the free energy profile $F(R)$, that $\Delta F_{UF} \approx 5.8kcal/mol$ and $\Delta x_{UF} \approx 5.2nm$. Although the change in ΔF_{UF} computed from estimate of $K_{eq}(f)$ based on hopping dynamics and the "exact" result is small ($\approx 7\%$) there is substantital difference in Δx_{UF}. The exact free energy profile (Fig. 3.3-C) clearly shows that Δx_{UF} *varies with* f because of large variations in the unfolded states. In general the assumption that Δx_{UF} is a constant leads to an overestimate of both ΔF_{UF} and Δx_{UF}.

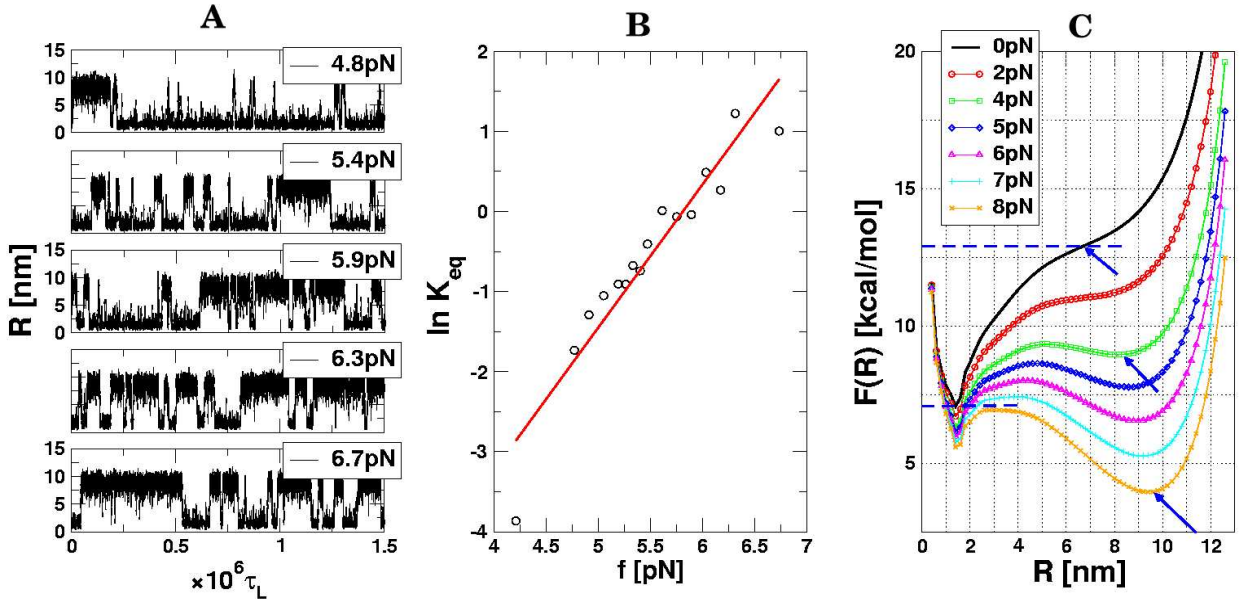

Figure 3.3: **A.** Time traces of R at various values of constant force at $T = 305K$. At $f = 4.8pN < f_m \approx 6pN$ $\langle R \rangle$ fluctuates around at low values which shows that the NBA is preferentially populated (first panel). As $f \sim f_m$ (third panel) the hairpin hops between the folded state (low R value) and unfolded states ($R \approx 10nm$). The transitions occur over a short time interval. These time traces are similar to that seen in Fig.2-C of [2]. **B.** Logarithm of the equilibrium constant K_{eq} (computed using the time traces in **A**) as a function of f. The red line is a fit with $\log K_{eq} = 10.4 + 1.79f$. **C.** Equilibrium free energy profiles $F(R)$ as a function of R at $T = 305K$. The colors represent different f values that are displayed in the inset. The arrows give the location of the unfolded basin of attraction.

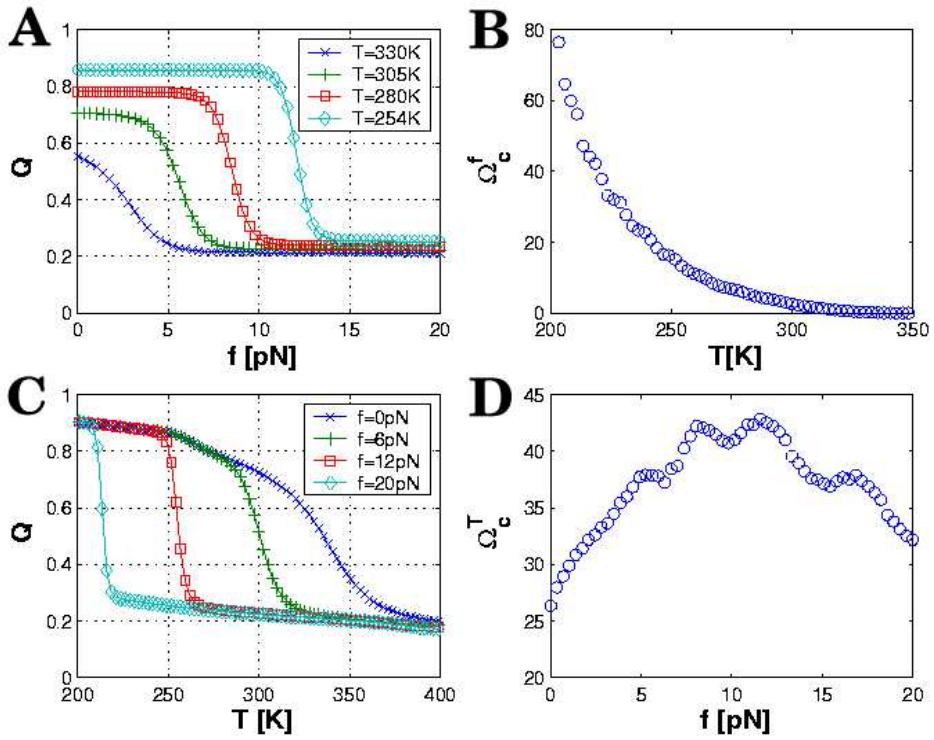

Figure 3.4: **A.** Dependence of $\langle Q(T,f)\rangle$ as a function of f at various temperatures. **B.** Values of $(\Omega_c^f)_T$ as a function of temperature. **C.** Variation of $\langle Q(T,f)\rangle$ as a function of T at various values of f. **D.** Dimensionless cooperativity measure $(\Omega_c^T)_f$ for $0 \leq f \leq 20$.

3.3.4 Cooperativity of unfolding depends on force

Slice of the phase diagram at either constant f or constant T shows the typical sigmoidal curves for $\langle Q\rangle$ as a function of either f or T (Fig.3.4). The cooperativity of the transition depends on whether T or f is held constant. We use the dimensionless cooperativity index $(\Omega_c^T)_f$ with respect to T [79].

$$(\Omega_c^T)_f = \frac{T_{max}^2}{\Delta T}\left|\frac{d\langle Q\rangle}{dT}\right|_{max} \tag{3.13}$$

where ΔT is the full width at the half maximum of $|(\frac{d\langle Q\rangle}{dT})_{max}|$ and T_{max} is the temperature at which $\frac{d\langle Q\rangle}{dT}$ has a maximum. Similarly, the dimensionless cooperativity index

with respect to f can be defined. The force dependent cooperativity index $(\Omega_c^T)_f$ has a maximum around $f \sim 10pN$, whereas $(\Omega_c^f)_T$ decreases monotonically to zero as T increases (Fig.3.4-B and 3.4-D). The difference between $(\Omega_c^T)_f$ and $(\Omega_c^f)_T$ arises because thermal denaturation at all forces is more stochastic while forced-unfolding disrupts RNA structures in steps.

3.3.5 Time scales of hopping transition

In the RNA pulling experiments [2] the time interval between the hopping transitions between folded and unfolded states at midpoint of force was measured at a single temperature. We have evaluated the dynamics along the phase boundary (T_m,f_m) (Fig.3.5) to evaluate the variations in the free energy profiles and the dynamics of transition from the NBA to UBA. Along the boundary (T_m,f_m) there are substantial changes in the free energy landscape (Fig.3.5-A). The free energy barrier $\Delta F^{\ddagger}$ increases dramatically at low T and high f. We predict that the weakly first order phase transition at $T \approx T_m$ and low f becomes increasingly stronger as we move along the (T_m,f_m) boundary to low T and high f.

The two basins of attraction (NBA and UBA) are separated by a free energy barrier whose height increases as force increases (or temperature decreases) along (T_m,f_m) (Fig.3.5-A). The hopping time τ_h along (T_m,f_m) is

$$\tau_h = \tau_0 \exp\left(\Delta F^{\ddagger}/k_B T\right). \tag{3.14}$$

To estimate the variations in τ_h along the (T_m,f_m) boundary, we performed three

very long overdamped Langevin simulations at $T_m = 305K$ and $f_m = 6pN$. The unfolding/refolding time is observed to be between 1 to $4ms$ (Fig.3.5-B). From the free energy profile (Fig.3.5-A) we find $\Delta F^{\ddagger}/T \sim 3$, so that $\tau_0 = 0.05$ to $0.2ms$. Consequently, τ_h at $T = 254K$ and $f = 12pN$ is estimated to be 1 to $4s$, which is *three orders of magnitude greater than at the higher* T_m *and lower* f_m.

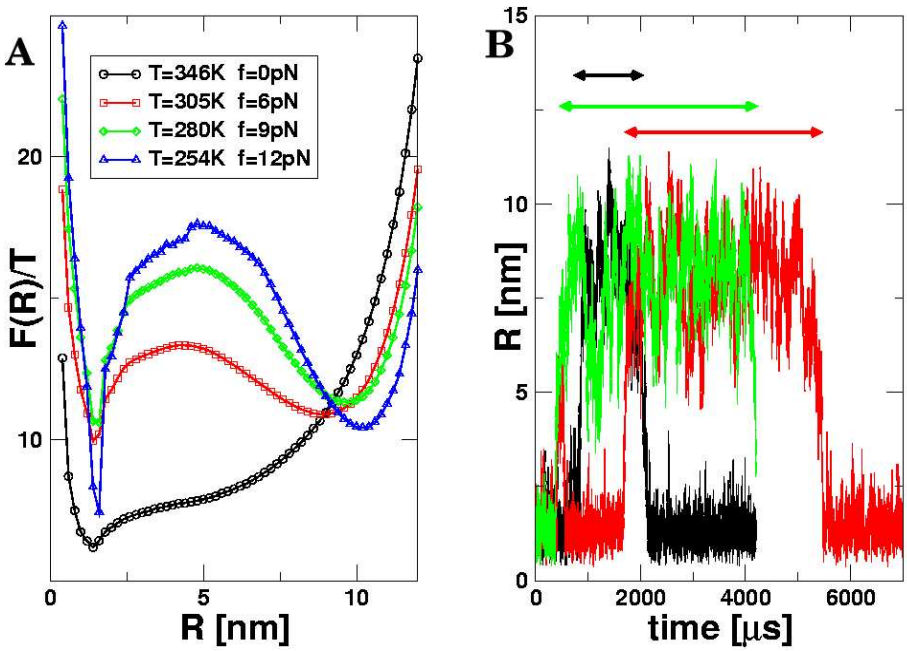

Figure 3.5: **A.** Free energy profiles $F(R)$ along the phase boundary (T_m,f_m) (see Fig.3.2). The barrier separating NBA and UBA increases at low T_m and high f_m values. **B.** Time traces of R obtained using Brownian dynamics simulations. The values of T and f are $305K$ and $6pN$, respectively. The arrows (black, red, green) indicate the residence times in the NBA for three trajectories.

Thermal refolding and unfolding: To induce thermal refolding we performed a temperature quench starting from a thermally equilibrated ensemble at $T = 510K$ to $T = 290K < T_m$. The approach to the folded RNA hairpin is monitored using the time dependence of Q, χ, and n_{bond}. A molecule is in the native state if $Q > 0.97$ and $n_{bond} = 9.0$. To confirm that the conformations with these values of

Q and n_{bond} are in the NBA we performed steepest descent simulations from states with $Q > 0.97$. Most of these conformations reach the native state with $\chi = 0.00$.

To calculate the folding time we performed temperature quench simulations for 100 different initially denatured conformations to obtain the distribution of the first passage time, i.e., the first time molecule i reaches the NBA. The initial population of unfolded molecules decays exponentially with the folding time $\tau_F^T \approx 12.4\mu s$. Nearly, 90% of the initially denatured molecules form folded structures in an "all-or-none" manner in which hairpin formation is initiated near the loop region with zipping of stabilizing contents progressing towards the end until the 5' and 3' contacts are established. In rare instance, the 5' and 3' ends meet first and zipping proceeds from the ends to the loop region (10%). Because of high entropy costs this process is less probable.

For comparison with mechanical unfolding we also performed simulations to monitor thermal unfolding. Equilibrated conformations at $T = 100K$ are heated to $T = 346K$ to initiate unfolding. Unlike in the thermal refolding, in which hairpin is formed by a zipping process, there is no characteristic disruption pathway. All of the nine bonds fluctuate independently until denaturation occurs. Thus, thermal unfolding is stochastic. Details of thermal unfolding and refolding will be published elsewhere (C. Hyeon and D. Thirumalai, Biophys. J. (to be submitted)).

3.3.6 Unfolding dynamics at constant force

To probe the structural transitions in the hairpin we performed steered Langevin dynamics simulations at constant force at $T = 254K$. From the phase diagram the equilibrium unfolding force at this temperature is $12pN$ (Fig.3.2-A). To monitor the complete unfolding of P5GA, in the time course of the simulations, we applied $f = 42pN$ to one end of the hairpin with the other end fixed. In contrast to thermal unfolding (or refolding) the initially closed hairpin unzips from the end to the loop region. The unzipping dynamics, monitored by the time dependence of R, shows *quantized staircase-like jumps* with great variations in step length, that depends on the initial conditions. The lifetimes associated with the "intermediates" vary greatly (Fig.3.6-A). The large dispersion reflects the heterogeneity of mechanical unfolding pathways. Approach to the stretched state that occurs in a stepwise "quantized" manner [71], which was first shown in lattice models of proteins, has recently been experimentally observed in the unzipping dynamics of DNA under constant force [80]. The presence of initial condition-dependent unfolding suggests that even in the small P5GA hairpin several distinct "metastable intermediates" are explored upon stretching.

3.3.7 Refolding under force quench

To monitor the dynamics of approach to the NBA we initiated refolding from extended conformations with $R = 13.5nm$, prepared by stretching at $T = 290K$ and

$f = 90pN$. Subsequently, we set $f = 0$ and the approach to the native state was monitored. From the distribution of first passage times the refolding kinetics follows exponential kinetics with the mean folding time of *about* $191\mu s$ *compared to* $12.4\mu s$ *in the temperature quench*. It is remarkable that, even though the final conditions ($T = 290K$ and $f = 0$) are the same as in thermal refolding, the time scale for hairpin formation $\tau_F^f \approx 15\tau_F^T$!.

The large difference in τ_F^T and τ_F^f arises because the molecules under the distinct initial conditions navigate entirely different regions of the energy landscape. The distribution of R in the thermally denatured conformations is $P(R)_{thermal} \propto e^{-\beta V_{tot}(R)/k_B T_H}$ (T_H is the initial temperature), while in the ensemble of the stretched conformation $P(R)_{stretch} \propto \delta(R - R_{ext}) e^{-\beta(V_{tot}(R) - \vec{f}\cdot\vec{R})/k_B T}$. The stretched conformations ($R_{ext} = 13.5nm$) do not overlap with the the accessible regions of the canonical ensemble of thermally denatured conformations (data not shown). As a consequence the regions of the free energy landscape from which folding commences in force jump folding are vastly different from those corresponding to the initial population of thermally equilibrated ensemble.

3.3.8 Force quench refolding occurs in multiple stages

The pathways explored by the hairpins en route to the NBA are heterogeneous (Fig.3.6-B). Different molecules reach the hairpin conformation by vastly different routes. Nevertheless, the time dependence of R shows that the approach to the na-

tive conformation occurs in stages (Fig.3.6-B). Upon release of force there is a rapid initial decrease in R that results in the collapse of the hairpin. Surprisingly, this process takes on an average several μs, which is much larger than expectations based on theories of collapse kinetics of polymer coils [81, 82]. In the second stage, the hairpin fluctuates in relatively compact state with R in the broad range (25-75)Å for prolonged time periods. On this greatly varying time scales, which varies considerably depending on the molecules, conformational search occurs among compact structures. The final stage is characterized by a further decrease in R that takes the molecules to the NBA. The last stage is most cooperative and sudden whereas the first two stages appear to much more continuous (Fig.3.6-B). Interestingly, similar relaxation patterns characterized by heterogeneous pathways and continuous collapse in the early stages has been observed in force quench refolding of ubiquitin [21]. The multistage approach to the native stage is reminiscent of the Camacho-Thirumalai proposal for protein refolding [41].

3.4 Conclusion

Use of constant force to unfold or initiate refolding (by force quench) provides glimpses of regions of the energy landscape of biomolecules that cannot be probed by conventional methods. In the mechanical unfolding experiments the molecules go from an initial low entropy state (folded) to another low entropy state (stretched). This is different from conventional experiments in which unfolding results in a transition

from a low entropy state to a high entropy state (unfolded). This difference results in vastly different mechanisms and time scales of folding and unfolding. Using novel coarse-grained models of RNA we have highlighted some of the major differences by considering temperature and force effects on unfolding RNA hairpins.

Our studies have lead to the following predictions, all of which are amenable to experimental tests: (1) The hairpin undergoes a first order transition from the folded to unfolded states at a critical value of f. The transition becomes strongly first order at low temperatures and high forces. Force unfolding, at a fixed f, is more cooperative than unfolding with T fixed and f being varied. (2) Unfolding of RNA occurs in steps with long pauses in a number of discrete intermediates that have a large dispersion in R values. (3) There are great variations in the hopping times between the NBA and the UBA along the locus of points in the (f, T) plane separating NBA and UBA. At low T_m and high f_m the hopping times are orders of magnitude greater than at $T \approx T_m$ and low f_m. (4) Remarkably, refolding times by force quench are much greater than folding initiated by temperature quench. The approach to the native state from stretched conformations occurs in several stages. The earliest events involve continuous changes in the progress variable that monitors folding rather being an "all-or-none" process.

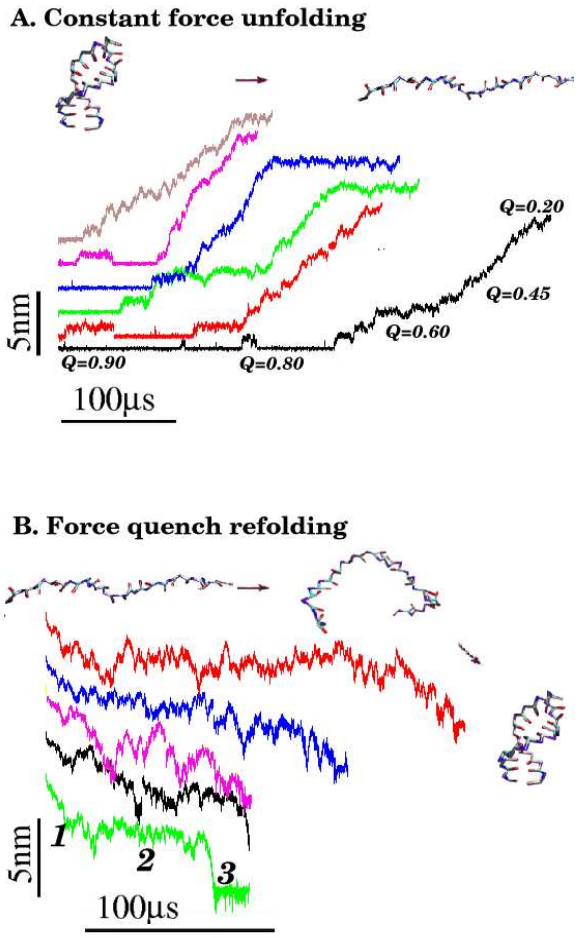

Figure 3.6: **A.** Time traces of unfolding of P5GA at a constant force f=42pN at T=254K monitored by the increase in R. The values of Q at different unfolding stages are given for the trajectory in black. **B.** Refolding is initiated by a force quench from the initial value f=90pN to f=0. The five time traces show great variations in the relaxation to the hairpin conformation. However, in all trajectories R decreases in at least three stages that are explicitly labeled for the trajectory in green. The trajectories in **A** and **B** are offset for clarity.

Chapter 4

Investigation on forced unfolding and refolding of RNA hairpins from force propagation and energy landscape perspective

Abstract

Single-molecule force spectroscopy is a novel technique for unraveling the complexity of biomolecular energy landscape on which the molecule of interest undergoes a large amount of heterogeneity in its dynamical pathways. Using a coarse-grained off-lattice model of RNA hairpin, we study the general characteristics of force experiments and biomolecular dynamics under mechanical force. We investigate the unfolding dynamics of RNA with both constant force and constant loading rate. The dynamics of linker polymer is appreciated as a force propagator in constant loading rate unfolding experiment. Coupling between force and energy landscape leads to the strong Hammond behavior, which we believe is the general feature in force experiments of proteins or RNAs. Force quench refolding dynamics from the fully stretched state are investigated in various angles. We find a hidden intermediate state which causes a substantially long refolding dynamics from fully stretched state.

4.1 Introduction

Folding/unfolding of RNA takes place on the energy landscape which is a conceptual representation of interactions between nucleotides and configurational entropy [45, 39]. The interactions between nucleotides are optimized under folding conditions for RNA to reach the native basin of attraction (NBA). Change of the experimental condition, such as counterions, denaturants, pH, and temperature, renormalizes the interactions between nucleotides, driving the RNA to other configurational states, or strengthening the bias towards the native states [64, 39]. Switching experimental conditions initiates the folding/unfolding of molecules and repopulates the ensemble of RNA molecules according to the modified Boltzmann weight. Traditional ways of unraveling the energy landscape have been utilizing the repopulation and relaxation technique on the ensemble of molecules by switching experimental condition.

The advent of single molecule manipulation technique has introduced mechanical force as a new control variable in biomolecular unfolding problem. Like other single molecule techniques, force experiments have enabled us to monitor the real time dynamics of biomolecules using the end-to-end distance as an order parameter [2, 80, 21]. The molecular individualism [83], which is suggestive of the complexity in underlying energy landscape, manifests itself through a large heterogeneity in unfolding pathways. In addition to the capability of force experiments that provide at single molecule level the detailed information for unraveling the folding landscape of the molecule, force experiments benefit from the simplicity in interpretation. From the energy landscape perspective the conversion of structure induced by force is un-

derstood as a rate process over the reduced free energy barrier. Constant force acting on the ends of a molecule changes the Boltzmann factor, like other control variables do but in more straightforward way, from $e^{-\beta\mathcal{H}}$ to $e^{-\beta(\mathcal{H}-f\cdot R)}$ [30]. Free energy of stability is obtained through the time average of unfolding/refolding trajectories [2, 72] Kinetic analysis with varying forces has been used to identify the position of transition states along the reaction coordinate [2, 21]. Free energy differences among the states and position of transition states being taken together, the force is suggested as a viable way of quantitatively probing the energy landscape along the end-to-end reaction coordinate.

In this chapter, combining the coarse-grained off-lattice simulation of RNA hairpin [72] and theoretical analysis, we answer the four major issues addressed in force experiments done on biomolecules.

(i) How does the presence of linker affect the experimental observation? Force-extension curve (FEC) is a typical experimental result under investigation. It is possible that the applied force at the end of linker molecule may not effectively be transferred to the target molecule depending on the linker length and flexibility. Due to the nature of nonequilibrium dynamics in constant loading rate pulling experiment, the unfolding force read from FEC can be affected.

(ii) Is the mechanism of forced unfolding different from temperature jump unfolding? There is a notion that the least stable part of molecule unfolds first because mechanical force acts as a perturbation like thermal impact [35]. However, the force is local and directed whereas thermal motion is stochastic in nature. Using RNA model, we will show the stark difference in unfolding pathways under two conditions.

(iii) Is the position of transition state static under varying force? The relative position of transition state along the end-to-end distance reaction coordinate at zero force is routinely obtained by extrapolating the result of kinetic analysis performed under varying force values [23]. However, force can affect not only the free energy differences between states but also the distance metric along the direction of force. We will verify this point using the free energy profile computed from histogram reweighting technique as well as using the analysis of force kinetics.

(iv) Where does the large time gap between force-quench refolding and thermal refolding stem from? Recent controversial force-quench experiment performed on polyubiquitin by Fernandez's group [21] showed that force-quench refolding occurs on a much larger time scale than thermal refolding. Our RNA hairpin model behaves phenomenologically in the similar manner. We will analyze the kinetic data and the trajectories of force-quench simulation from various angles, and elucidate the cause of retardation.

4.2 Methods

The dynamics of the molecule is computed in the same protocol discussed in Chapter 3. We also studied stretching at constant pulling speed and in the presence of linker.

4.2.1 Force simulations at constant pulling speed

Schematic representation of pulling simulation is shown in Fig.4.1-C, which provides the simplified view of the experimental setups for a single molecule stretching like atomic force microscope (AFM) (Fig.4.1-A) and laser optical tweezer (LOT) (Fig.4.1-B). AFM and LOT are essentially identical except for the difference in effective spring constant ($k = 1 - 10\ pN/nm$ for AFM, $k = 0.01 - 0.1\ pN/nm$ for LOT) when compared to the schematic representation. Mechanical unfolding of a molecule is simulated either by applying a constant force on 3'- end of the molecule with 5'-end being fixed (unfolding at constant force), or by pulling 3'- end through the combination of linker polymer and harmonic spring with constant speed in one direction (unfolding at constant loading rate). (i) Constant force simulations are implemented by subjecting the 3'- end to a constant force value f. The spring constant or the presence of linker is irrelevant issue in the constant force experiments. (ii) On the other hand, for unfolding at constant loading rate ($r_f \equiv k \times v$), where the force (f) is computed by multiplying the given spring constant (k) with the extension of spring (δz) as $f = k \times \delta z$, there are several factors that can affect the experiment: pulling speed (v), spring constant (k), linker length (L), and linker flexibility. We will study the effect of each factor on the resulting FEC, separately.

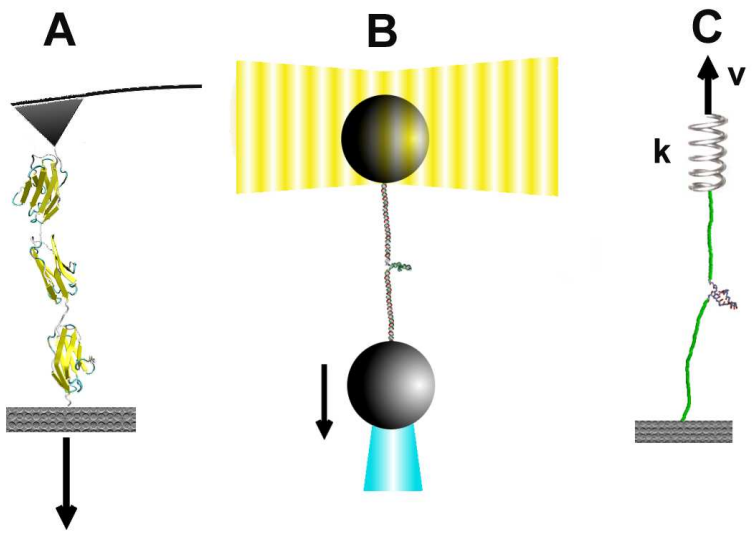

Figure 4.1: Illustrations of two major single molecule stretching techniques: The stretching of Ig27 domain of titin muscle protein was the first successful experiment that exhibited the power of AFM in single molecule manipulation. As the molecule is stretched the deflection of cantilever is detected by the instrumentation using laser (**A**). The saw-tooth pattern of force-extension curve is the striking evidence of single molecule unraveling under the mechanical force. Much more elegant and sensitive manipulation of single molecule has become available due to the laser optical tweezer (LOT). Single molecule is held between the two polystyrene beads via molecular handle with one of the polystyrene beads being optically trapped in the laser light. The other bead is sucked to the micropipette. The extension of the molecule through the molecular handles induces deviation in the position of the polystyrene bead held in the force-measuring optical trap (**B**). Stretching of Ig27 domain of muscle titin using AFM (**A**) and P5abc domain of *Tetrahymena* RNA ribozyme using optical tweezer (**B**). Schematic representations of two experimental technique are similar with different spring constant k, which are depicted in **C**

4.2.2 Force simulations in the presence of linker

The linker molecule is explicitly modeled as a polymer composed of N monomers using the sum of two potential terms, $V_B + V_A$. The bond potential with harmonic approximation and the angle potential are given as,

$$V_B + V_A = \sum_{i=1}^{N-1} \frac{k_B}{2}(r_{i,i+1} - b)^2 - \sum_{i=1}^{N-2} k_A \hat{r}_{i,i+1} \cdot \hat{r}_{i+1,i+2} \tag{4.1}$$

where $r_{i,i+1}$ and $\hat{r}_{i,i+1}$ are the distance and unit vector connecting i and $i+1$ residue, respectively. For bond potential we set $k_B = 20\ kcal/(mol \cdot \mathring{A}^2)$, $b = 5\ \mathring{A}$. We assign either $k_A = 80\ kcal/(mol \cdot rad^2)$ or $20\ kcal/(mol \cdot rad^2)$ to study the effect of linear flexibility. For computational efficiency, we did not include the excluded volume interaction between linker and linker or between linker and RNA. When the linkers are under tension the chains do not cross unless the thermal fluctuations are bigger than the tension. Our interest is also in the force propagation along the chain rather than the exact description of chain motion in the regime where thermal fluctuations dominate. To study the linker effect, we attach two linker polymers with the length $L/2$ to both ends like a pair of chopsticks, and stretch the molecule using the single pulling speed $0.86 \times 10^2\ \mu m/s$ with spring constant $k = 0.7\ pN/nm$. The linker length is varied from $L = (10 - 50)\ nm$ (see Fig.4.2). For simplicity, pulling speed (or loading rate) dependency is studied without linker polymer ($L = 0\ nm$).

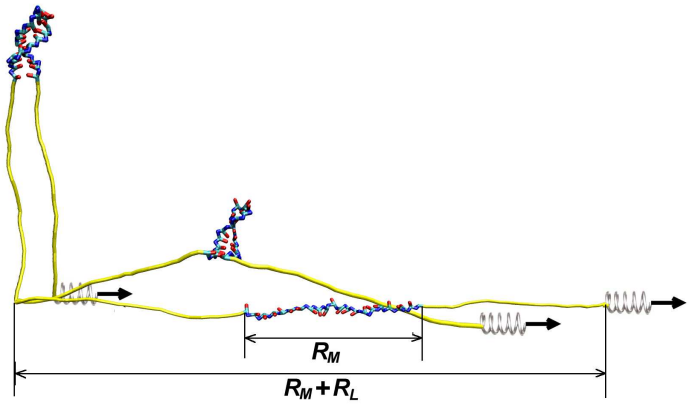

Figure 4.2: Three snapshots of RNA stretching simulations. Two linker polymers of total length 30 nm are attached to the both ends of RNA. Stretching is performed on one end with the other end being fixed.

4.3 Results and Discussion

4.3.1 Factors affecting stretching experiment

As mentioned earlier, the pulling speed (v), spring constant (k), and linker are characteristics of the linker that can affect the force experiments. There is a concise theoretical expression relating the loading rate with the unfolding force as $f^* \sim \frac{k_B T}{\Delta x_F^{TS}} \log r_f$ [34] where Δx_F^{TS} is the distance between native state and transition state. Pulling with high loading rate leads to a large unfolding force.

The effect of spring constant on the force-extension curves (FECs) is subtle. This is illustrated by solving the following Langevin equation numerically.

$$\zeta \dot{x} = k(vt - x) - \frac{k_B T}{l_p} \left(\frac{1}{4(1 - \frac{x}{L})^2} + \frac{x}{L} - \frac{1}{4} \right) + \eta(t) \tag{4.2}$$

where ζ is the friction coefficient, l_p is the persistence length, L is the contour length of chain, $\eta(t)$ is the magnitude of random force at time t, and x is the extension of the chain. The force is measured from the extension of the spring multiplied by spring

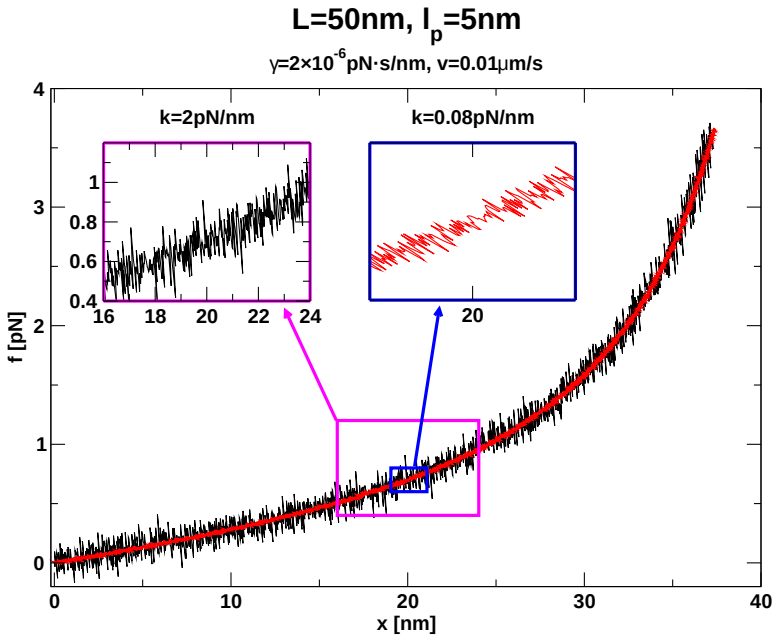

Figure 4.3: The force-extension curve of wormlike chain model at hard ($k = 2pN/nm$) and soft spring constant ($k = 0.01pN/nm$). Each pattern of fluctuation is typical for AFM and LOT, respectively. The precision of force-extension curve is greater in small spring constant.

constant as $f = k(vt - x)$. FEC corresponds to f plotted against x (Fig.4.3). The fluctuation of signal on the $f - x$ plane is "vertical" for hard spring ($k = 2pN/nm$) and is "horizontal" for soft spring ($k = 0.01pN/nm$). Such fluctuation pattern are seen in AFM and LOT experiments.

In the stretching experiments, the mechanical force is usually exerted on the target molecule through the linker polymers. The introduction of linker polymer is necessary for preventing the interference of signal from the instrumentation. In Ig27 pulling experiment using AFM, the first major peak in force-extension curve results from the nonspecific interaction between the molecule and the surface rather than the unfolding event in one of the multidomain protein. For the optical tweezer the polystyrene bead is coated with streptavidin protein, and the distance between the

beads should be large enough to ensure that only one bead is trapped by the laser. If the length of DNA/RNA handle is too short then it becomes difficult to discern the dynamics in RNA from FEC for the DNA-RNA-DNA construct. The isolation of the target molecule from the source of unnecessary interaction is fulfilled by the introduction of linker polymer.

FEC obtained by the simulations with varying linker lengths (Fig.4.4) shows the plateau region around 20-40 pN force values, which corresponds to hairpin opening. Depending on the definition of the extension one can plot either Fig.4.4-A or Fig.4.4-B for FEC. Fig.4.4-B, which plot the force versus RNA-only extension (R_M), shows the unfolding event unambiguously. The end-to-end distance of hairpin itself does not change until force reaches above 20 pN. The transition event expands the end-to-end distance of RNA from 1 nm to 12 nm in the narrow range of force ($\delta f < 5$ pN). However, one should note that experimentally measured end-to-end distance ($R_M + R_L$) is from the total system combining RNA molecule and linker polymers. Fig.4.4-A is the experimentally more relevant plot. (i) In Fig.4.4-A we can still observe the transition plateau, but the signal isolating the plateau region becomes unclear when the linker length reaches 50 nm. (ii) Another interesting observation, which is clear from Fig.4.4-B, is that unfolding force gradually increases with increasing linker length. (iii) Fig.4.4-C shows the influence of linker flexibility. The more flexible linker leads to a higher value of unfolding force. We below explain the cause of these three observations.

Fluctuations of chain at equilibrium: From the partition function of linker

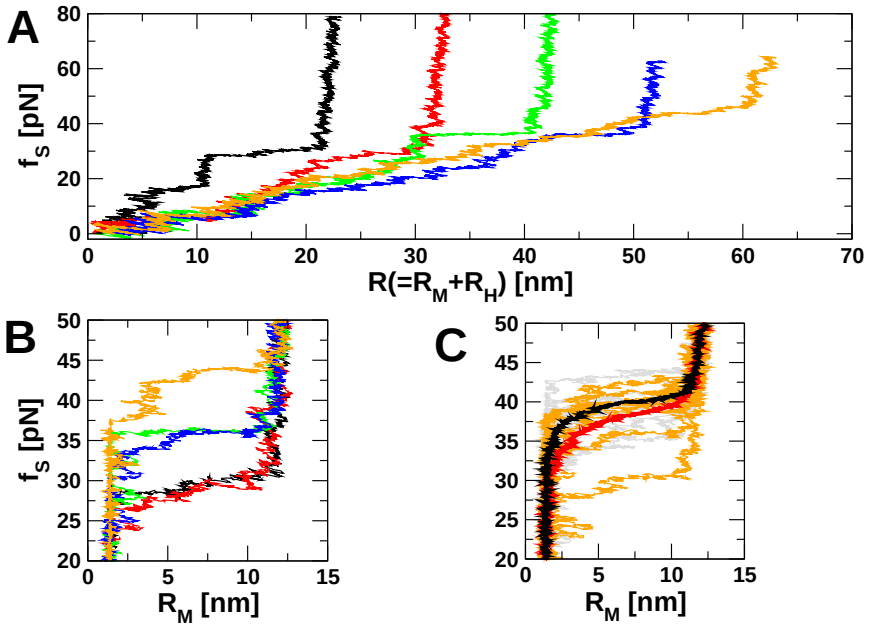

Figure 4.4: The force extension curve of RNA hairpin at constant pulling speed with varying linker lengths and flexibilities. The pulling speed, $v = 0.86 \times 10^2 \mu m/s$. The spring constant, $k = 0.7 pN/nm$. FECs from different linker lengths are plotted in black (10 nm), red (20 nm), green (30 nm), blue (40 nm), and orange (50 nm). FEC in (**A**) is the experimentally relevant plot. End-to-end distance measured in experiment is the full length of the molecule and the linkers. The signature for hairpin opening transition becomes hard to detect only from the FEC in (**A**). Actual FEC only for molecule is plotted in (**B**). The gradual increase of rupture force is observed with increasing linker length (**B**). The bending rigidity of linker polymer used in (**A**) and (**B**) is 80 $kcal/(mol \cdot \mathring{A})$. (**C**) shows the comparison between FECs with the same linker length (40nm) but different bending rigidities. Red curve is the average of 7 individual FECs in orange (80 $kcal/(mol \cdot \mathring{A})$), Black curve is the average of 9 individual FECs in grey (20 $kcal/(mol \cdot \mathring{A})$). Less stiff linker leads to slightly larger unfolding force.

polymer under force $Z = \int d\tau \exp\left(-\beta(\mathcal{H} - fR)\right)$, the fluctuation of the end-to-end distance in polymer is given by $(\delta R)^2 = \partial\langle R\rangle/\partial(\beta f)$. Depending on the force exerted on the ends, $\langle R\rangle$ scales differently. If the total system is large enough, $\langle R\rangle \sim (2Ll_p)^{1/2}$ for small f and $\langle R\rangle \sim L$ for large f. Therefore,

$$\delta R \sim \begin{cases} (Ll_p)^{1/4}(k_BT/f)^{1/2} & (f_BT) \\ L^{1/2}(k_BT/f)^{1/2} & (f \gg k_BT) \end{cases} \tag{4.3}$$

The fluctuation of spring extension is, using equipartition theorem, given by

$$\delta x \sim \sqrt{\frac{k_B T}{k}}. \tag{4.4}$$

In order for the transition signal to be unambiguously identifiable, the expansion of end-to-end distance of the molecule at transition should be larger than both fluctuation δR and δx. Since δR grows sublinearly with the linker length (Eq.(4.3)) the attachment of large linker polymer can mask the transition signal.

Linker relaxation time at constant loading rate experiment : Note again the snapshots of forced unfolding of RNA and linker molecules illustrated in Fig.4.2. The stretching is not initiated from the conformation where the linker molecules are aligned parallel to the force direction, so that the angle $\theta(t)$ is formed between the force direction and the orientation of the linker at time t. A finite amount of time (say τ_R) will be spent aligning the linker molecule along the force direction. Provided that the force acting on the end of linker is f and we assume for simplicity the linker is rod-like, the effective force exerted on the 3'-end of RNA molecule is $f\cos\theta(t)$. The unfolding condition is achieved when $f\cos\theta(t) = f_c$ where f_c is the unfolding force of RNA in the absence of linker. Since, in constant loading rate experiments, the force f is ramped at constant rate ($r_f = k \times v$) and the time τ_R, that satisfies $\cos\theta(\tau_R) \approx 1$ or $\theta(\tau_R) \approx 0$, is finite, unless $r_f\tau_R < f_c$ at time τ_R, the unfolding force f measured from the experimental setup including linker molecule becomes larger than f_c. Therefore, if the relaxation time of linker (τ_R) is longer than f_c/r_f the time at which the ramped force reaches the unfolding force, then the non-equilibrium effect

of the linker dynamics on the force-extension curve will manifest itself.

The relaxation time of linker to the parallel direction to force (τ_R) depends on the characteristics of linker molecules, which include the size and the flexibility.

Theory of force propagation : We have shown that the adaptation time of linker conformation in the presence of force is an important factor that affects the force-extension curves. In this subsection we try to provide insight into the force propagation problem in more systematic way by presenting the formal solution of force propagation in the semiflexible chain [84]. The free energy functional of semiflexible chain under tension at one end is

$$\mathcal{F}[\vec{r}(s)] = \frac{\kappa}{2}\int_0^L ds \left(\frac{\partial^2 \vec{r}}{\partial s^2}\right)^2 + \frac{1}{2}\int_0^L ds\lambda(s)\left(\frac{\partial \vec{r}}{\partial s}\right)^2 - \vec{f}\cdot\vec{r}(0). \tag{4.5}$$

where $\vec{r}(s)$ is the position vector at arc position s, the bending rigidity κ is set to constant throughout the chain contour, and the Lagrange multiplier, $\lambda(s)$, corresponding to the tension at the arc position s, is incorporated to ensure the constraint $\left(\frac{\partial \vec{r}}{\partial s}\right)^2 = 1$. In the overdamped condition, the equation of motion of chain is written as

$$\zeta\frac{\partial \vec{r}}{\partial t} = -\frac{\delta \mathcal{F}}{\delta \vec{r}} + \vec{\eta}(s,t) \tag{4.6}$$

where ζ is the friction coefficient and $\vec{\eta}(s,t)$ is the white noise acting on segment s at time t satisfying a relation $\langle\vec{\eta}(s,t)\cdot\vec{\eta}(s',t')\rangle = 6k_BT/\zeta\delta(s-s')\delta(t-t')$. The functional derivative with respect to the segment coordinate in the chain leads to

$$\zeta\frac{\partial \vec{r}}{\partial t} = -\kappa\partial_s^4\vec{r} + \partial_s(\lambda\dot{\vec{r}}) + [f + \lambda(s)\dot{\vec{r}}(s) - \kappa\partial_s^3\vec{r}]\delta(s) + \vec{\eta}(s,t) \tag{4.7}$$

and the condition $\partial_t \dot{\vec{r}}^2 = 0$ gives

$$\ddot{\lambda}(s) = \ddot{\vec{r}}(s)^2 \lambda(s) + \kappa \dot{\vec{r}} \partial_s^4 \dot{\vec{r}}. \tag{4.8}$$

Neglecting the bending rigidity term for simplicity (because of our interest in the stretched regime at high force) we encounter the coupled nonlinear differential equation with boundary condition $\lambda(0) = f$ and $\dot{\lambda}(L) = 0$. One of the simplest possible solutions for tension profile is when the whole polymer is aligned parallel to the direction of the applied force ($\vec{r}(s) \sim constant \times s$), i.e., polymer is stretched. In this case the tension profile is $\lambda(s) = f$ for all values of s. The tension propagation is rapid for the fully stretched polymer. This argument is valid regardless of the bending rigidity since the last term in Eq.(4.8) is also zero when the polymer is stretched. Starting from a given initial conformation of the polymer chain, the tension profile evolves with time adapting its new profile to the newly generated conformation. Eq.(4.7) and Eq.(4.8) should be iteratively solved. For the arbitrary conformation, the tension profile, $\lambda(s)$, monotonically decreases from the applied force value f at $s = 0$ to the smaller value at $s = L$. The wiggle in the chain conformation, if any, is straightened until the chain conformation reaches the form of $\vec{r}(s) \sim constant \times s$ [84] that corresponds to rod-like state.

To summarize, in constant speed stretching experiments, the linker polymer should be short and stiff in order not to compromise the signal from the molecule of interest.

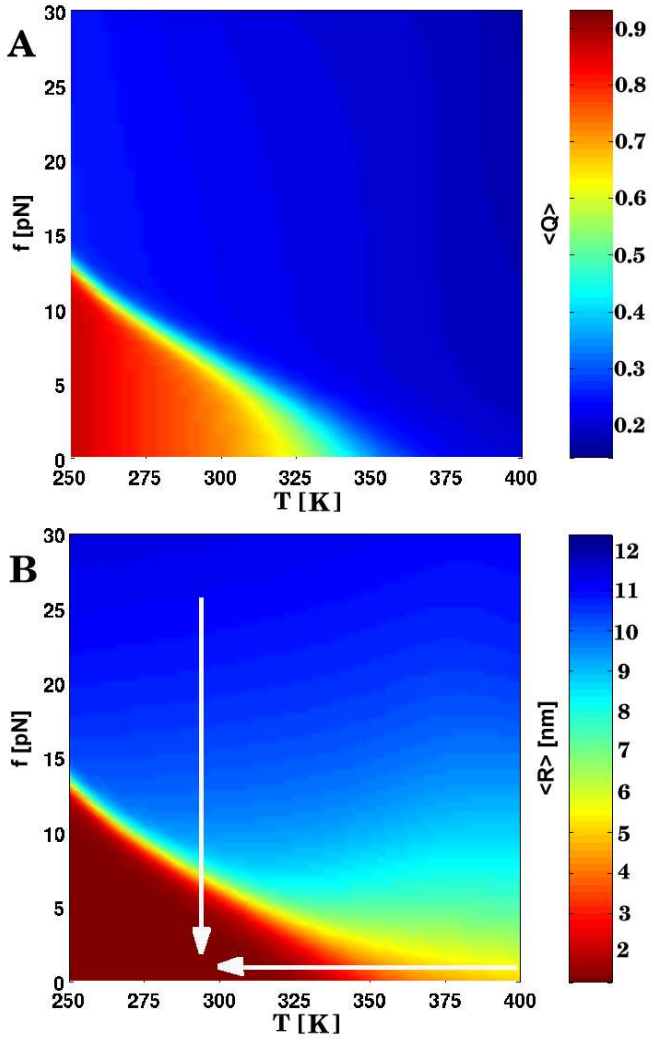

Figure 4.5: Phase diagram for the P5GA hairpin. **A.** This panel shows the diagram of states obtained using the fraction of native contacts as the order parameter. The value of $\langle Q \rangle$ are color coded as indicated on the scale shown on the right. **B.** Plot of the phase diagram in the (f, T) plane using the mean end-to-end distance $\langle R \rangle$ as the order parameter. Although the diagram of states is qualitatively similar as in **A** there are quantitative differences in estimates of T_m at $f = 0$. However, estimates of threshold force values at $T < T_m$ are similar in both **A** and **B**. The white arrows are drawn to emphasize that the hairpin undergoes different conformational changes under the force quench and temperature quench.

4.3.2 Thermal Unfolding and Forced Unfolding

We compared the reaction pathway between thermal unfolding and forced unfolding. For thermal unfolding, the equilibrated conformations at $T = 100K$ are heated to $T = 346K$ to initiate unfolding. Forced unfolding is achieved by applying the constant force $f = 42pN$ to the equilibrated conformations at $T = 254K$. The results in Fig.4.6 succinctly compares the difference between two unfolding pathways. For thermal unfolding, all of the nine bonds fluctuate independently until complete denaturation occurs (Fig.4.6-A). Whereas, a sequential unzipping from ends to loop occurs in forced unfolding (Fig.4.6-B).

To describe the thermal unfolding, the fraction of native contacts Q is a better reaction coordinate than R because R at high temperatures is subject to large fluctuations. The free energy profiles before and after switching the external condition are depicted in the Fig.4.6-A (right panel). The potential to the NBA centered around $Q \approx 0.9$ at low temperature ($T = 100K$) becomes unstable at high temperature ($T = 346K$). The conformations of the hairpin evolve to achieve the structures with small Q. A similar explanation can be made for forced unfolding using R instead of Q. Under a large force ($f = 42pN$), potential bias is formed in $R \approx 12nm$ where stretched states are favored. Once the activation barrier at $R \approx 1.5nm$ is overcome the hairpin starts unfolding (Fig.4.6-B, right panel).

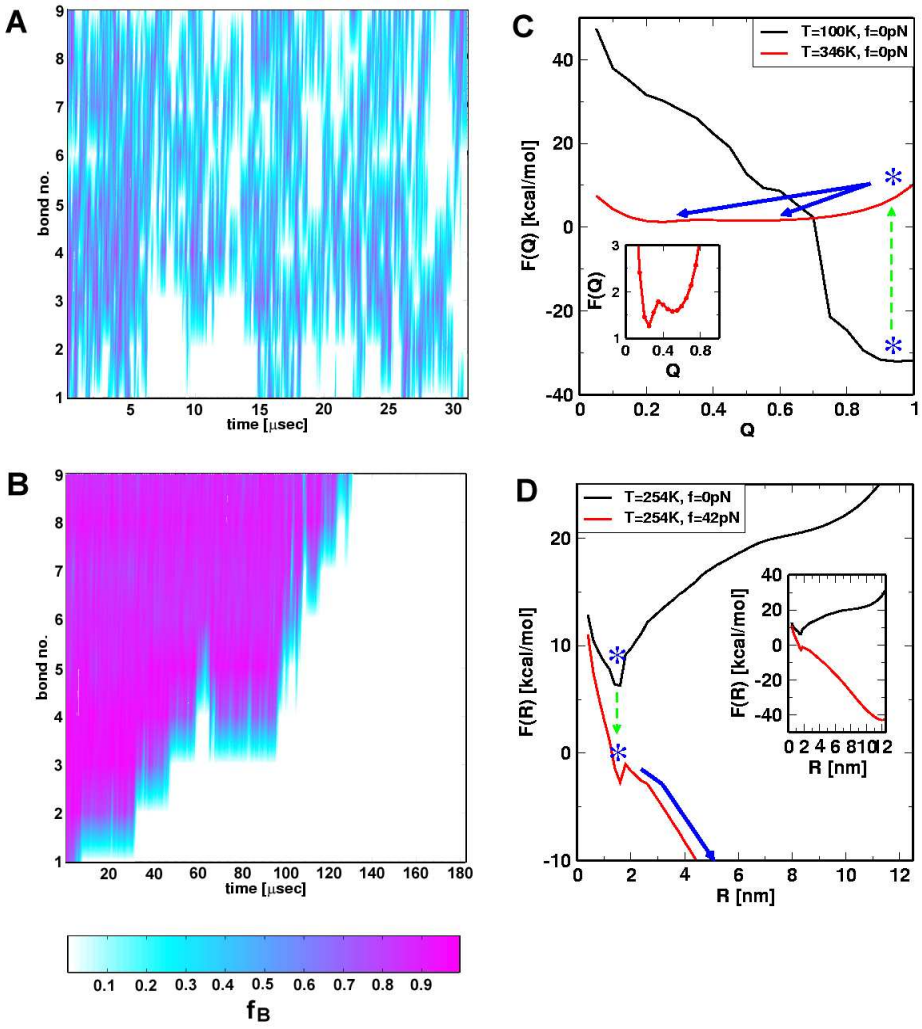

Figure 4.6: Unfolding pathways upon temperature and force jump. **A.** The bond rupture is monitored with time when the temperature is raised from $T(=100K) < T_m(\approx 341K)$ to $T(=346K) > T_m$. The nine bonds fluctuates stochastically until the hairpin unfolds. **C.** shows the change in free energy profile when temperature is jumped above T_m. The stable hairpin with $Q \approx 1$ at $T = 100K$ becomes unstable at $T = 346K$ (blue $*$), hence should relax to the lower free energy minima formed at $Q \approx 0.2$ where the denatured conformations are more stable. **B.** The bonds are ruptured from the ends to the loop regions, which is described by staircase pattern. Left panel shows the deformation of free energy profile of R as a reaction coordinate. The small transition barrier traps native conformations but UBA formed around $R = 12nm$ is far more stable than NBA under the force ($f = 42pN$). The native hairpin unfolds accordingly.

Forced unfolding, refolding kinetics and free energy profiles: The thermally averaged distance between the native state and the transition state, Δx_F^{TS}, is one of the estimates which have been routinely extracted from the steered molecular dynamics simulation [46] and the force experiments [36, 35]. In the dynamic force spectroscopy experiment Δx_F^{TS} is obtained using the relationship $f^* \sim k_B T/\Delta x_F^{TS} \log r_f$ where f^* is the most probable unfolding force and r_f is the loading rate defined by df/dt [34]. Similarly, in constant force unfolding experiments Δx_F^{TS} is obtained from the Arrhenius relation between the unfolding rate and the applied force, $\log k = \log k_0 + f\Delta x_F^{TS}/k_B T$ where k_0 is the unfolding rate in the absence of force. However, if nonzero curvature is intrinsic in these plots, one should check the validity of the extracted values such as Δx_F^{TS} or k_0 from the linear regression of the experimental data. In the presence of curvature, the extrapolation to the zero force is not always correct unless the data are obtained near $f \sim 0$. In the following two subsections we will present the result of the forced unfolding kinetics using both constant loading rate and constant force simulation. We found that the curvature was developed over the broad range of loading rate or force variations. The free energy profiles of R computed using the histogram reweighting technique are presented for the comparison. The equilibrium $F(R)$ curves explain the origin of curvature in the context of energy landscape.

Unfolding at constant loading rate: We first performed the forced unfolding simulation corresponding to the dynamic force spectroscopy. We varied both pulling speed v and spring constant k, spanning the broad range of loading rates. Just as

in dynamic force spectroscopy, the unfolding force, defined at which all the hydrogen bonds are ruptured, formed a distribution whose average and dispersion increase with growing loading rate (Fig.4.7-A). From the force distribution, we plot f^* as a function of $\log r_f$ (Fig.4.7-B). The curvature of the plot is observed "over the broad range of loading rates", which leads to the increasing tangent with increasing r_f. There are two reasons for the increasing tangent. One is the reduction of Δx_F^{TS}, which increases the value of $k_BT/\Delta x_F^{TS}$. The other is the increase of curvature at the transition state region, i.e., barrier top of free energy landscape. Whichever the case is, the standard way of estimating Δx_F^{TS} at a large loading rate results in very small value of Δx_F^{TS}. Our maximum estimate of Δx_F^{TS} from f^* versus $\log r_f$ plot is 4Å at $r_f \approx 10^5 pN/s$. In RNA hairpin model structure 4Å is even smaller than the distance between the neighboring beads. Because of the computational cost we did not further reduce the pulling speed. The minimum loading rate we could use is 10^5 times larger than the one used in Liphardt *et al.*'s RNA unfolding experiment [2]. Using the same loading rate with the one in experiment is computationally impossible. In full atom steered molecular dynamics (SMD) the situation is worse. The loading rate used in SMD is even about 10^{10} times larger than the real experiment! We, however, certainly expect that the slope of the plot will further decrease at smaller loading rate until the plot forms an intercept with the axis of loading rate. In principle, the slope of plot should be used near this intercept to extract Δx_F^{TS} of energy landscape in the absence of force.

Unfolding at constant force: Using the constant loading rate in force experiment

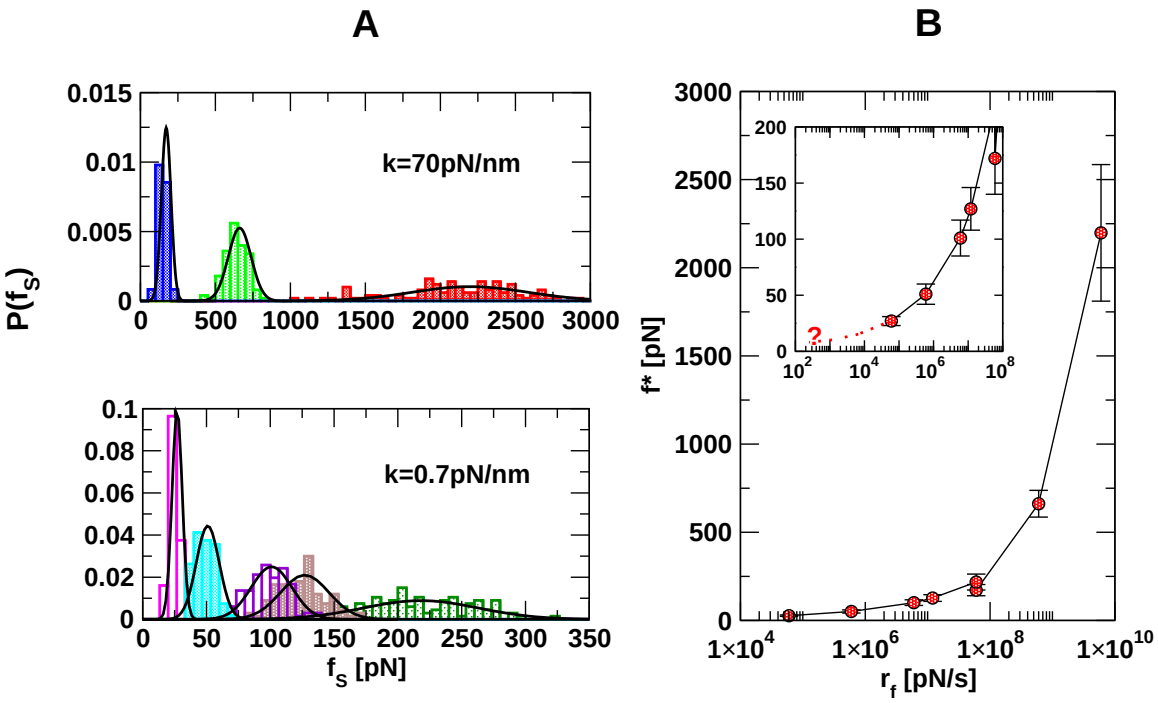

Figure 4.7: Constant loading rate force unfolding. **A.** The unfolding force distributions under varying pulling speeds with hard ($k = 70pN/nm$, up) and soft springs ($k = 0.7pN/nm$, down). For the hard spring, the pulling speeds are $v = 8.6 \times 10^4$, 8.6×10^3, $8.6 \times 10^2 \mu m/s$. For the soft spring, the pulling speeds are $v = 8.6 \times 10^4$, 4.3×10^3, 8.6×10^3, 8.6×10^2, $8.6 \times 10^1 \mu m/s$ from right to left force peaks. The distributions are fit to the Gaussian for convenience and the most probable force f^* is taken to be $\sim< f >$. **B.** The most probable forces from the distributions, f^*, are plotted with respect to the loading rates, r_f. The results from the hard spring and soft spring are combined using the loading rate instead of pulling speed. f^* versus $\log r_f$ plot shows the upward curvature, suggesting the transition state movement.

is a complicated process from the energy landscape perspective. The free energy landscape is progressively deformed with the increasing force, so that the stochastic unfolding events are coupled with the dynamic change of the landscape. Applying the ramping force on the system is analogous to increasing the concentration of denaturant with time. What is usually done, instead, in conventional kinetic experiments is to monitor the relaxation time of a system due to switching to the new constant value of temperature, denaturant or ionic concentration. Monitoring the system behavior

under constant force is more straightforward than constant loading rate experiment because, at least in our analysis, the static picture of free energy profile is available at each force value.

We performed the constant force unfolding simulation at $f > 20pN$ force range. The rate process at $f < 20pN$ is too slow for our computational time scale. For 50-100 trajectories generated at $20pN < f < 150pN$ the mean unfolding times are computed from the distribution of the first passage times that a molecule reaches $R = 5nm$. Fig.4.8-A shows the mean unfolding time (τ_U) at different force. Over the full range of force we used, curvature develops as in constant loading rate simulation. Although the linear regression can be done in the narrow range of force values, $\Delta x_F^{TS} \approx 4\mathring{A}$ is still too small to be a physically meaningful value at zero force. In fact, the free energy profiles of R at different f can reveal the origin of curvature and small Δx_F^{TS}. Fig.4.8-C shows how $F(R)$ changes with increasing force. The location of the transition state becomes more native-like as force increases and larger position shift is found at small force ($< 20pN$). The transition state resembles unfolded state more in the absence of force, so that $\Delta x_F^{TS} \approx 5.5nm$. $\Delta x_F^{TS} \approx 4\mathring{A}$ is from the deformed free energy profile at $20pN < f < 50pN$ not from the original free energy profile without force. Unless the transition barrier dividing NBA and UBA is sharp, the position of transition state is greatly affected by the force.

Refolding under force quench: Refolding is initiated from the extended conformations with $R = 13.5nm$, prepared by stretching at $T = 290K$. Instead of completely releasing the ends force is held at small but finite values ($0.5 \leq f \leq 4pN$).

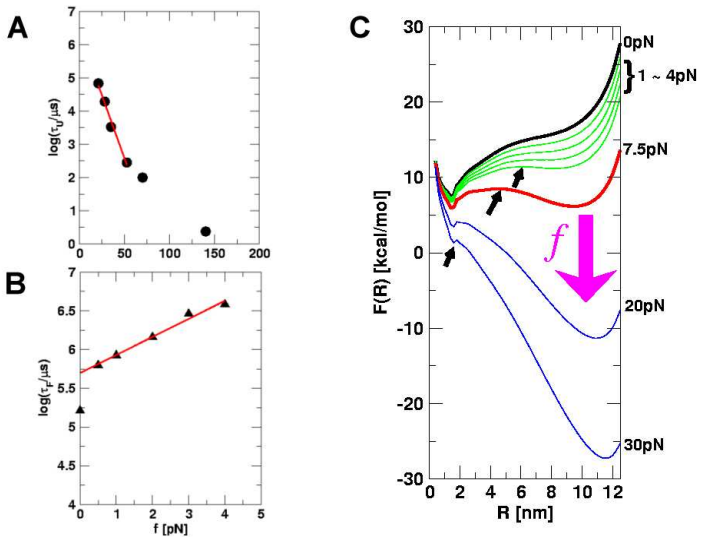

Figure 4.8: Kinetics under force. **A.** Unfolding time (τ_U) under varying force values. Arrhenius behavior manifests itself in the narrow range of force variation. The curvature is observed over the broad force range, suggesting the transition state movement. We estimated $\Delta x_U^{TS} \approx 4$Å using τ_Us at four smallest force values. **B.** Refolding time (τ_F) under finite force quench values. From the Arrhenius relation $\Delta x_F^{TS} \approx 1nm$. **C.** Free energy profiles with respect to R at $T = 290K$ under varying forces. For emphasis, the free energies at $f = 0$ (black) and at transition midpoint $f = 7.5pN$ (red) are drawn in thick lines.

Probabilities of remaining unfolded RNA at six quenched force values are plotted in Fig.4.15. The mean refolding time (τ_F) is computed from the distribution of the first passage time to reach the native state. Contrary to the forced unfolding, the perturbation due to force is minor and the time scale of mean refolding time ranges narrowly over the force range. In case of force quench refolding from stretched state, the refolding time is approximately given as

$$\tau_F = \tau_F^o \exp\left(f \Delta x_U^{TS} / k_B T\right) \tag{4.9}$$

where Δx_U^{TS} is the distance between UBA and transition state, τ_F^o is the refolding time in the absence of force. Linear regression leads to $\Delta x_U^{TS} \sim 1nm$, which is consistent with the results of free energy profile. The extrapolation using $0.5 \leq f \leq 4pN$ to

zero force gives $\tau_F^o \approx 270\mu s$ which reproduces the mean refolding time in the absence of force ($\sim 191\mu s$) with some error.

It is of particular interest to compare the refolding time from the stretched ensemble (τ_F^f) and the refolding time from the thermally denatured ensemble (τ_F^T). In a previous paper [72], we reported a big difference between τ_F^f and τ_F^T ($\tau_F^f = 15\tau_F^T$). The origin of the long refolding time under force quench will be investigated in below.

4.3.3 Hammond postulate for force-unfolding experiments

In physical organic chemistry Hammond postulate [85, 86] is widely used to qualitatively predict the nature of transition state in the chemical reaction of organic molecules. The original postulate by G. S. Hammond is stated as follows. *If two states, as for example, a transition state and an unstable intermediate, occur consecutively during a reaction process and have nearly the same energy content, their interconversion will involve only a small reorganization of the molecular structure* [85]. Although the folding/unfolding pathways of RNA/proteins are far more complex than the reaction pathways of simple organic reaction it would be interesting to see whether the same postulate can be applied. In the context of protein folding, Hammond postulate can be rephrased as *the position of transition state along the reaction coordinate is shifted towards the destabilized state, either folded or unfolded state, depending on the experimental perturbation.* Indeed, Hammond postulate on protein folding has been examined by Sánchez and Kiefhaber [87, 88]. They used the

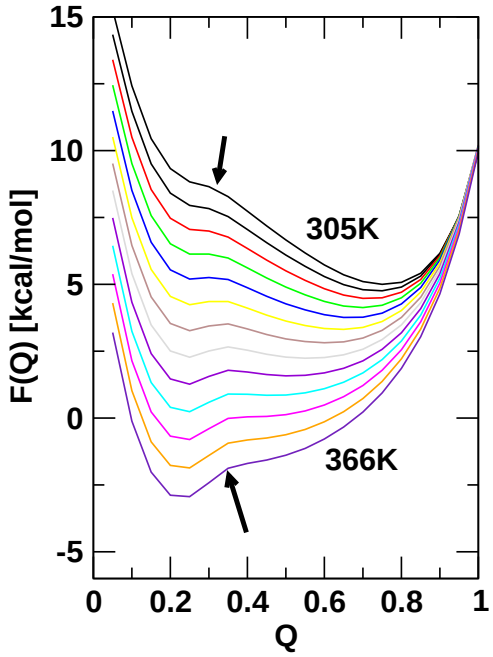

Figure 4.9: Free energy profiles of Q at various temperatures. Note that the positions of transition states over the temperature variation almost remain constant.

denaturant (urea/GdmCl) under instead of force for testing the Hammond behavior on 21 well-characterized proteins. The result is only one of the 21 proteins shows a small but coherent Hammond behavior, which suggests that at least for 20 proteins under study the structure of the transition state is well conserved independent of the denaturant concentration. We have already shown the Hammond behavior in forced unfolding simulation in Fig.4.8. The nonlinear Arrhenius plot suggested the movement of transition state and the free energy profile at increasing force (Fig.4.8-C) have clarified this. In comparison with force, it would be interesting to check the Hammond behavior of the RNA hairpin model due to denaturant as in protein experiment. However, because the denaturant effect is not directly tested using our hairpin model we alternatively examined Hammond behavior with respect to temperature change Fig.4.9 shows the free energy profiles at various temperature without

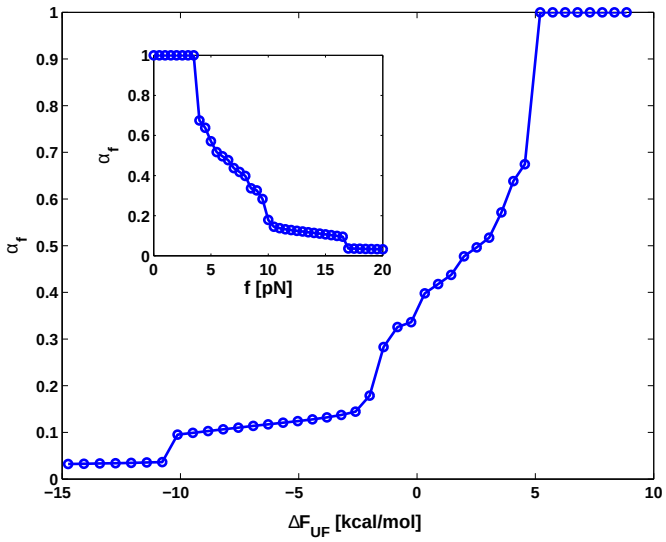

Figure 4.10: The movement of transition state. The structural nature of transition state is monitored by the free energy difference between NBA and UBA when the force is a external variable. The inset shows the variation of α with respect to force.

applying force. Interestingly, the location of transition state does not change when the temperature alters the stability of UBA or NBA. Applying force on the molecule is similar to increasing the denaturant concentration in that the unfolded state is preferentially stabilized, but there is intrinsic difference in the mechanism of unraveling the molecule by force.

Hammond behavior can also be quantified using the Leffler's proportionality constant α_x which measures the energetic sensitivity of the transition state relative to the native states in respect to a perturbation x [86, 88]. In case the perturbation is due to force,

$$\alpha_f = \frac{\partial \Delta F^{\ddagger}(R)/\partial f}{\partial \Delta F_{UF}(R)/\partial f}$$

which reduces to

$$\alpha_f = \frac{\Delta x_F^{TS}}{\Delta x_{UF}}. \tag{4.10}$$

Using the free energy profile we computed α_f as a function of ΔF_{UF} or f (Fig.4.10). The shift of the transition state quantified by α_f ranges from 0 to 1 and the shift rate (or self-interaction parameter, $p_f \equiv \partial\alpha_f/\partial\Delta F_{UF}$ [88]) has maximum at $4 < f < 10pN$. The transition state movement under force is very sensitive to the shape of the transition barrier. If the potential is expanded near barrier ($x \sim x_{ts}$) as $F(x) \sim F(x_{ts}) - \frac{1}{2}F''(x_{ts})(x - x_{ts})^2 + \cdots$ and is tilted by amount of $-f \cdot x$, the new barrier position (x_{ts}^{NEW}) will be formed at $x_{ts}^{NEW} \approx x_{ts} - f/F''(x_{ts})$. In case of a sharp transition barrier ($F''(x_{ts}) \gg f$), the force will not affect the position of transition state ($x_{ts}^{NEW} \approx x_{ts}$). Whereas, the transition barrier is broadly distributed, like the unzipping pathway of RNA hairpin, the structure of the transition state ensemble progressively changes depending on the magnitude of force, and upward curvature in kinetic plot notifies this change. In general forced unfolding experiment, the scenarios of sequential [89] and/or parallel transition barriers [90, 91] are also possible. The careful inspection of nonlinearity in Arrhenius plot and α_f will elucidate the energetics of folding problem. We want to emphasize again that force is a special control variable distinguished from other experimental variables.

4.3.4 Refolding dynamics from stretched state

In refolding kinetics from the stretched state we found that the refolding time due to force quench (τ_F^f) is much greater than refolding

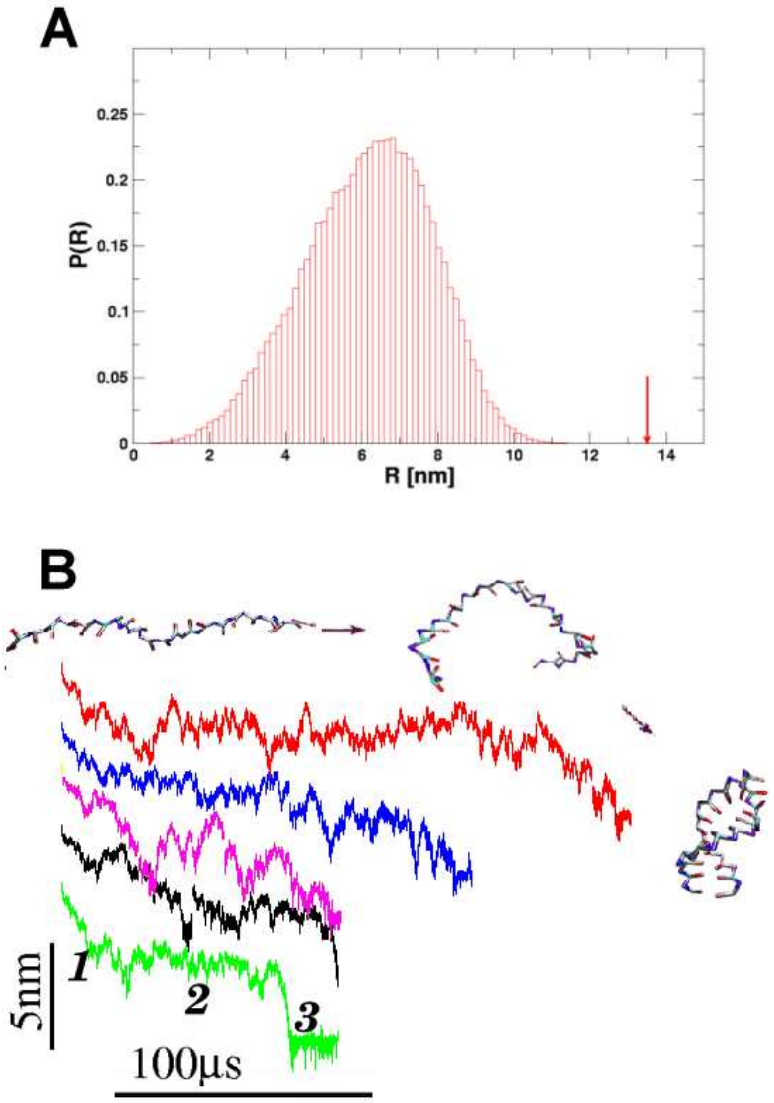

Figure 4.11: **A.** The equilibrium distribution of end-to-end distance at extremely high temperature ($T = 1500K$). Even at this elevated temperature the fully stretched conformations of $R = 13.5nm$ (arrow) is not found in the ensemble of thermally denatured conformations. **B.** Refolding is initiated by a force quench from the initial value f=90pN to f=0. The five time traces show great variations in the relaxation to the hairpin conformation. However, in all trajectories R decreases in at least three distinct stages that are explicitly labeled for the trajectory in green.

time from the thermally denatured states (τ_F^T). One of the general arguments for the large difference in refolding times is that the molecules under distinct initial conditions navigate entirely different regions of the energy landscape [92, 72]. The fully stretched conformations with $R = 13.5nm$ is very unlikely to occur by heating the molecules. The canonical distribution of thermally denatured conformations even at $T = 1500K$ shows no overlap with the conformations of $R = 13.5nm$ (Fig.4.11-A). Trajectories of refolding from extended conformations, $R(t)$, exhibit the broad range fluctuations (25-75Å) for long time periods, followed by the cooperative reduction in R at the final stage. Recently, similar behavior of long refolding times (0.1-10 sec) has been experimentally observed for the force quench refolding of ubiquitin [21]. Because the collapse processes of the single ubiquitin molecule in solution studies are known to occur in less than a millisecond, the observation of long refolding times immediately raised debates [93, 92]. Since we have observed similar behavior from the RNA hairpin model, we scrutinize the origin of the long refolding dynamics.

In terms of the fraction of native contact no difference is found between the thermally denatured conformations and the mechanically stretched conformations. Whereas, the end-to-end distance differentiates the two group of ensembles. Force and temperature quench refolding follow along vertical and horizontal lines in $\langle R \rangle(T, f)$ of Fig.4.5-B, respectively. The specific question being asked is what kind of interactions creates the big difference between two pathways. First, the inspection of dihedral angle shows that the severe distortion is required in one of the dihedral angle in the native state. The 19-th dihedral angle along the sugar-phosphate backbone is g^+ form in native state, but when the hairpin is stretched $g^+ \rightarrow t$ transition occurs. The

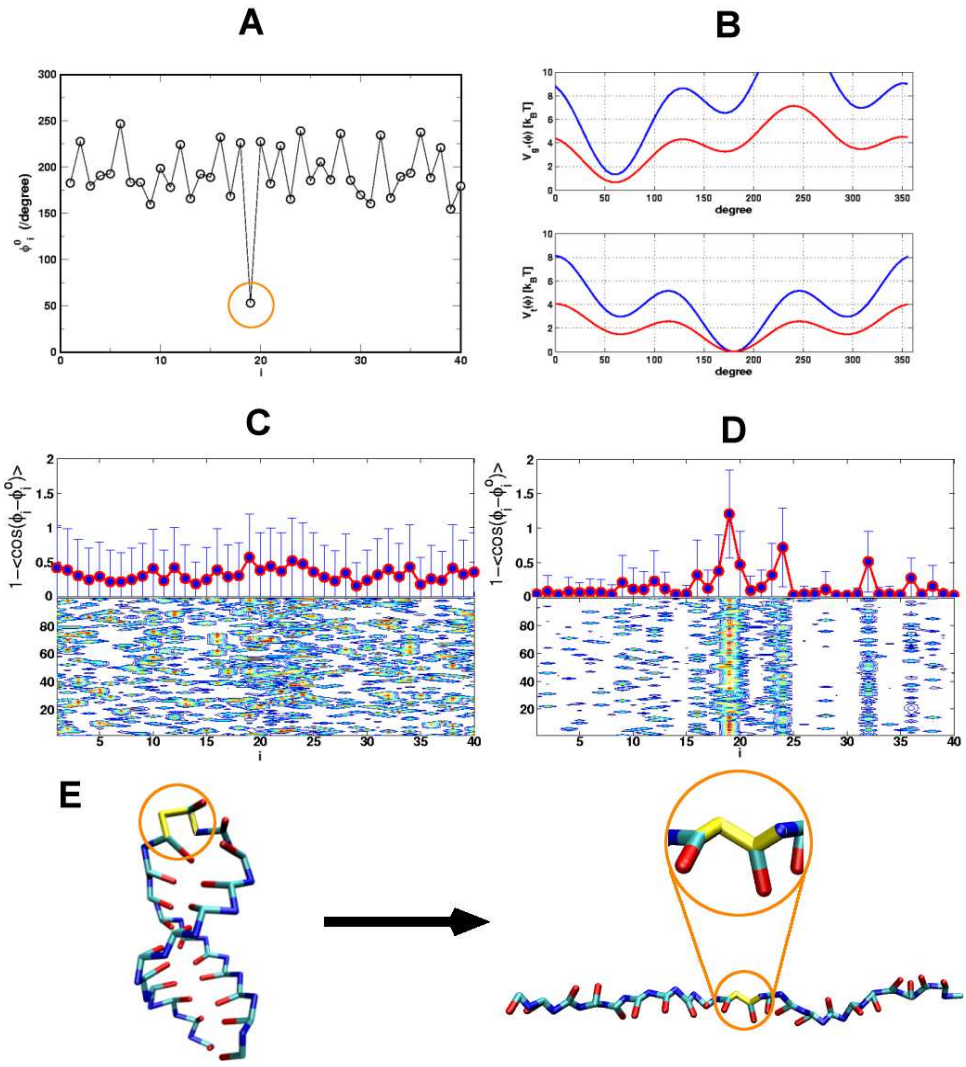

Figure 4.12: **A.** The dihedral angles of P5GA hairpin in native states. All the dihedral angles are in trans form except 19-th position of dihedral angle which is in gauche(+) (indicated by orange circle). **B.** The dihedral angle potentials for trans (up) and gauche(+) form (down) are plotted using Eq.(3.3). The red lines are for the loop region. **C, D.** The average deviation of the i-th dihedral angle relative to the one in native state is computed using the 100 different structures generated by high temperature ($T = 500K$) (**C**) and by force ($R = 13.5nm$) (**D**). To express the angle deviations from the dihedral angle of native state we used $1 - \cos(\phi_i - \phi_i^o)$ for i-th dihedral angle ϕ_i where ϕ_i^o is the i-th dihedral angle of the native state. **E.** The snapshot of fully stretched hairpin. Note the transition in 19-th dihedral angle from g^+ to t form when hairpin is stretched.

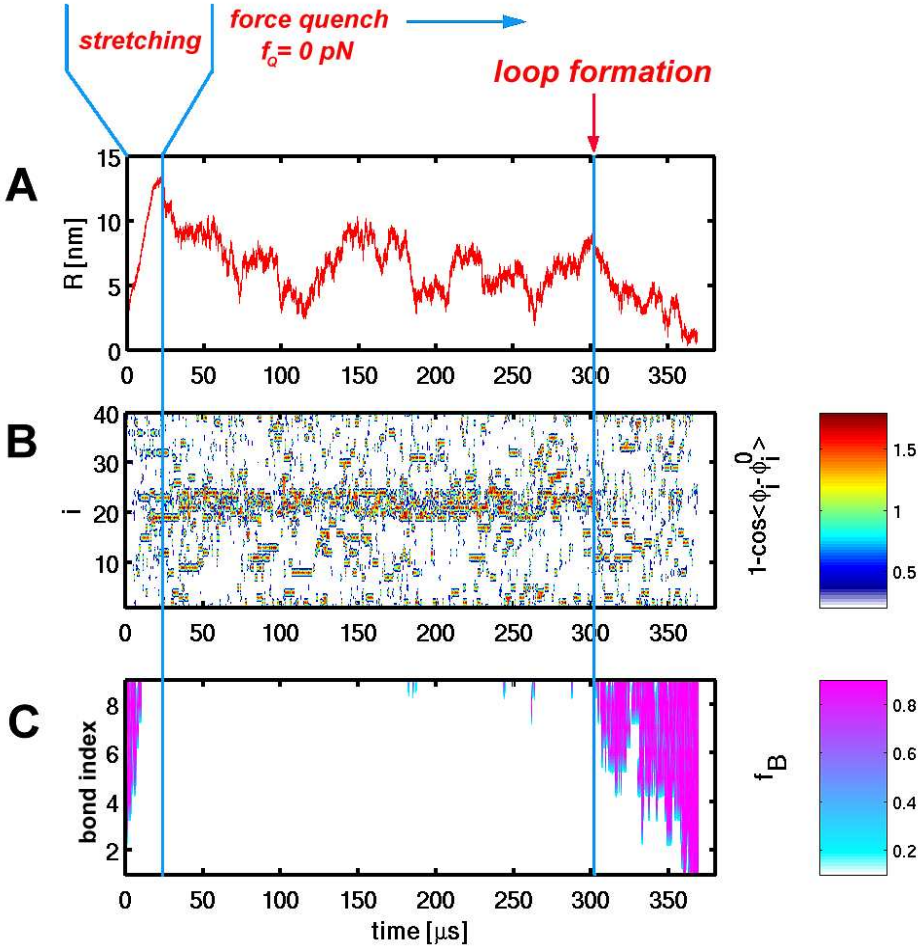

Figure 4.13: Trajectory of refolding dynamics from the stretched state. The hairpin was initially unfolded to fully stretched state and quenched to zero force at $t \approx 20\mu s$. **A.** End-to-end distance monitored with time. **B.** The deviation of dihedral angles with time. The large deviation of dihedral angle in loop region can be seen in red strip. Note that this strip disappears around $t \approx 300\mu s$, which coincides the formation of bonds shown in **C.**. f_B is the fraction of bond with pink color meaning that bond is formed.

ensemble of stretched hairpin has the incorrect t form of dihedral angle at $i = 19$ in common. In terms of potential energy, this particular incorrect dihedral angle costs $\sim 2k_BT$ enthalpy penalty. Although significant fluctuations are found in thermally denatured ensemble at $T = 500K$, the fluctuations are not large enough to produce the incorrect dihedral angles. The thermally denatured conformations still retain similarity in terms of backbone structure with the native conformation. Next, we monitored the time evolution of all the dihedral angles. As we have designed the flexible bond rotation of loop region, a large amount of fluctuations due to $g^+ \leftrightarrow t \leftrightarrow g^-$ transitions were observed in the dihedral angles of loop regions ($i = 19 - 24$). For the specific trajectory shown in Fig.4.13 we observe the persistence of incorrect dihedral angle on the loop region until $\sim 300\mu s$. However, the incorrect dihedral angles on the loop region disappear all of sudden at $\sim 300\mu s$, after which the formation and propagation of bonds take place. The phenomena is reminiscent of the nucleation process. The loop formation is the rate limiting step in the refolding kinetics from the extended states. The observation that the zipping of the hairpin takes place under the synchronization in the formation of all correct dihedral angles suggests the involvement of high entropic barrier.

We adopted the idea of the simple model description of protein folding by Zwanzig [94] to estimate the time scale involved in the random search of conformational space under some biases. Suppose that the energy difference between incorrect and correct dihedral angle is ϵ and that there are N correct dihedral angles required for forming the loop contact whose stabilization energy is given by ϵ_{loop}. The potential energy, E_n, with n-incorrect dihedral angles is given by $E_n = n\epsilon - \delta_{n0}\epsilon_{loop}$.

Assuming that the dihedral angle is the discrete variable with $1+\nu$ states, where ν is incorrect value of dihedral angle, one can write the partition function of N-dihedral angle system as

$$Z_N = \sum_{n=0}^{N} \binom{N}{n} \nu^n e^{-\beta(n\epsilon - \delta_{n0}\epsilon_{loop})} = e^{\beta\epsilon_{loop}} + (1+\nu e^{-\beta\epsilon})^N - 1 \tag{4.11}$$

The thermal probability of any configuration with n incorrect dihedral angles is

$$P_n(eq) = \frac{\binom{N}{n}\nu^n e^{-\beta(n\epsilon - \delta_{n0}\epsilon_{loop})}}{Z_N}. \tag{4.12}$$

The free energy profile as a function of n is $\Delta F(n) = n\epsilon - \delta_{n0}\epsilon_{loop} - k_B T \log \nu^n \binom{N}{n}$. For the present problem $N = 6$, $\epsilon \approx 2k_BT$, and $\epsilon_{loop} = 7.6k_BT$. Fig.4.14 plots the free energy profile with transition barrier for three different ν. The free energy barrier of $\sim 5k_BT$, which gives a factor of 100 when substituted into the Boltzmann factor, is obtained when $\nu = 20$. Because the angle is a continuous variable, we use ν as an undetermined parameter. The value of $\sim 5k_BT$ is significantly larger than the free energy barrier $\Delta F_U^{\ddagger}$ in the absence of force. The large fluctuations of the end-to-end distance is in fact caused by the metastable conformation with incorrect dihedral angles in the loop regions. As can be noticed the finding of all-correct dihedral angles is significantly affected by the entropic cost. Only by application of large f can such states be populated. Once the loop is formed the zipping process quickly stabilizes the hairpin so that barrier crossing in reverse direction is unlikely to happen in the absence of force.

To further address the origin of long refolding times upon force-quench, we performed refolding simulations from the ensemble of structures that are stretched to the conformations with only single bond remaining before the loop region. The

refolding kinetics exhibit single exponential decay with mean refolding time of $\sim 33\mu s$, which is 2-3 times longer than the refolding time of thermally denatured states, but still shorter than the refolding time from fully stretched states ($> 191\mu s$, see Fig.4.8-B).

The presence of the entropic barrier due to variations in the local angular degrees of freedom is not reflected in the phase diagram or in $F(R)$ or $F(Q)$. Indeed, Q cannot even distinguish the stretched states from thermally denatured states. There is little correlation between the formation of all the correct loop dihedral angle and the end-to-end distance. The distinction between the conformational states beyond and within the entropic barrier can properly be made by neither Q nor R. The entropic barrier against forming the loop from stretched states is *hidden* along the reaction coordinate of Q or R.

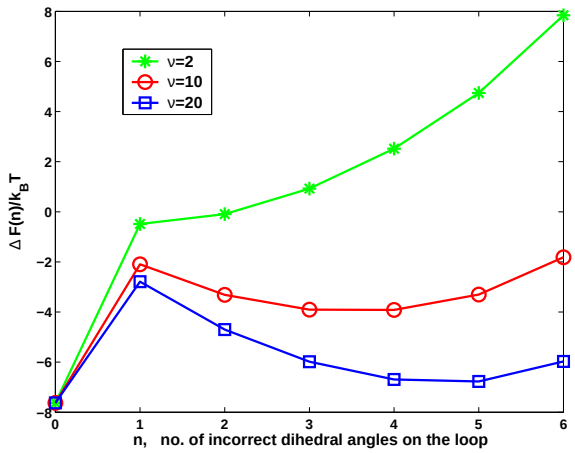

Figure 4.14: The free energy profile computed using $\Delta F(n) = n\epsilon - \delta_{n0}\epsilon_{loop} - k_B T \log \nu^n \binom{N}{n}$ (see text).

The implication of intermediate state along the reaction coordinate of refolding from stretched state can also be found at $P_U(t)$ in Fig.4.15. Parallel kinetic process

characterized by multiexponential function ($\sum_i \rho_i e^{-t/\tau_i}$ with $\sum_i \rho_i = 1$) does not seem to be appropriate to explain the lag phase in $P_U(t)$. The lag kinetics becomes more conspicuous at high quench force. We suggest the preliminary kinetics scheme based on the presence of intermediate state as $S \xrightarrow{\tau_1} I \xrightarrow{\tau_2} F$ where S denotes the stretched state, I intermediate state, and F folded state. Setting $P_U(t) \equiv P_S(t) + P_I(t) = 1 - P_F(t)$, we obtain

$$P_U(t) = \frac{1}{\tau_2 - \tau_1} \left(e^{-t/\tau_2} - e^{-t/\tau_1} \right). \tag{4.13}$$

Fig.4.15 has the fits of $P_U(t)$ to Eq.4.13, determining a pair of parameters (τ_1, τ_2) at each quench force.

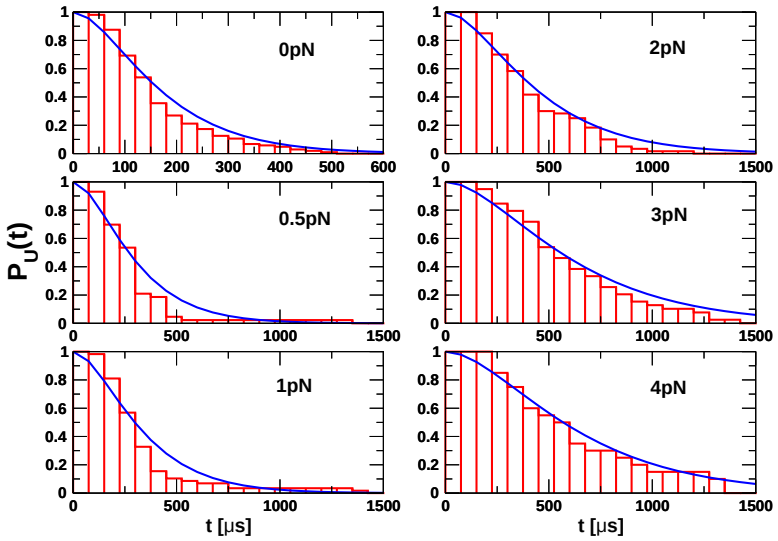

Figure 4.15: Probability of unfolded RNA molecule at different quench force. Fits are made on Eq.4.13. The pair of time constant (τ_1,τ_2) at each force is found as follows. (81.2, 101.3) at $f = 0pN$, (159.5, 160.8) at $f = 0.5pN$, (180.0, 174.8) at $f = 1pN$, (237.6, 240.5) at $f = 2pN$, (326.8, 335.6) at $f = 3pN$, and (347.7, 329.7) at $f = 4pN$ with time units being all in μs.

Force-quench refolding in the presence of linker : There is an additional issue regarding the experimental observation of refolding event. Fig.4.16 shows time evo-

lutions of $R_M + R_L$ and R_M of Fig.4.2 when the molecule is refolded from stretched state under final quench force of 1 pN. As is obvious from the plot, $R_M + R_L$ and R_M are not synchronized. It is very unlikely to obtain reliable signals of target molecule dynamics if the linker molecule is long and flexible [95].

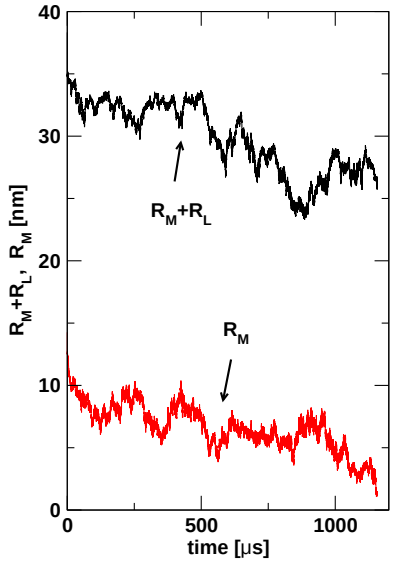

Figure 4.16: Refolding trajectory monitored through linker polymer The total linker molecule is 30 nm long and has bending rigidity of 80 $kcal/(mol \cdot rad^2)$. Initial force stretching RNA hairpin to 13.5 nm is quenched to 1 pN. The refolding trajectory is monitored in two different ways, $R_M + R_L$ and R_M.

4.4 Conclusions

Here we recapitulate the lessons learned from our study on the general role of mechanical force in unfolding/refolding process using RNA hairpin model.

First, the non-equilibrium force propagation is an important issue in force experiment [84, 96, 97]. The RNA or linker molecules adopt its conformation accordingly with its conformational relaxation time (say Rouse relaxation time). However, if

the variation of force (pulling speed or loading rate) is faster than molecular relaxation time, even the steady state picture that is underlying assumption of Kramers' equation is questionable. From this aspect, a careful study from polymer physics is required for the force experiment on chain molecule. Prior to discussing the energetics of target molecule, it is necessary to take the dynamics of linker molecule into account. In order to minimize the interference from the linker molecule, one should use a short and rigid linker molecule.

Second, from the mechanistic point of view the changes of the external control parameters (counterion, denaturant, temperature, pressure, pH, force, and etc.) affect the interactions between the nucleotides in totally different manners; The increase of counterion concentration decreases the electrostatic repulsion between the phosphates and facilitates the folding. The denaturant impedes the formation of hydrogen bond. The increase of temperature raises the entropic contributions ($\Delta G = \Delta H - T\Delta S$), favoring the thermally denatured states. The force acting on the ends of RNA unzips the hydrogen bonds between the base pairs, which propagates along the networks of bonds and eventually produces the low entropic stretched states. There is no reason to believe that a molecule undergoes the same dynamic process when the different experimental conditions are imposed. Indeed, the results in this chapter show the dramatic differences between thermal and force unfolding/refolding pathways.

Third, force can severely affect the position of the transition state ensemble depending on the specific shape of energy profile along the reaction coordinate. Theory [98, 99] and experimental data analysis are based on Arrhenius type Bell's model [30]. However, the assumption of static transition state leads to the erroneous conclusion

that unfolding is triggered by a small extension because the transition state is very native-like [35]. We showed this conclusion could be misleading if force or loading rate used in experiment is large enough to shift the position of transition state to more native-like. In addition, the systematic experimental investigation similar to Sánchez and Kiefhaber's study on Hammond behavior of RNA and proteins under mechanical force would be very interesting.

Lastly, the origin of long refolding dynamics in the RNA hairpin model was attributed to the dynamical difficulty in forming a initial loop. The large degrees of freedom involved with dihedral angles in loop region lead to an entropic bottleneck. The argument based on "incorrect dihedral angles" could be model-dependent in terms of simulation. However, we provided the specific example that there exist a large amount of heterogeneity and energetic ruggedness in the unfolded conformations of even simple RNA hairpins. Thermally denatured conformations already contain the residual structure which enhances the folding rate. From the standpoint of conformational space, thermally denatured conformations are more proximal to the native basin of attraction than rod-like conformations. In contrast to other non-specific experimental parameters, mechanical force is a very special control variable that populates the rod-like conformation among the large spectrum of heterogeneous unfolded states.

Chapter 5

Predicting force-induced unfolding pathways of biomolecules using a topology-based model

Abstract

In recent years single molecule force experiment has been used to unravel the structure and the dynamics of biopolymers. While the response of a biopolymer under force, which is probed in the form of force vs extension curve, provides a glimpse of the underlying energy landscape, the extraction of unfolding pathways at molecular level is difficult. Use of a minimal model suitable for simulating the forced dynamics allows us to visualize the chain motion and provide a molecular interpretation of the force/extension curve. Forced unfolding pathways predicted using the minimal model for *Tetrahymena* ribozyme, which is a challenging system to simulate by conventional methods, show an excellent agreement with those analyzed in recent force experiment. We find that molecules follow distinct unfolding pathways depending on the pulling speed. We show that the interplay between tension propagation and molecular relaxation time in pulling experiments determines force-induced unfolding mechanism. Refolding simulations from the stretched conformation also provide valuable insights into the self-assembly of large biomolecules.

5.1 Introduction

Numerous efforts have been recently made to understand the biological process through computer simulations [100]. All-atom molecular dynamics (MD) simulations using an atomic model for water is a computational strategy typically adopted in the hope that one can study the biological process in the most accurate way. However, for practical purposes, the reductionist approach using all-atom MD is not suitable for studying large scale motions because the insurmountable computational cost results from wide timescale gap between atomic and global motion leading to the self-assembly. All-atom simulation of mechanically-driven dynamics has the same problem. The steered molecular dynamics, designed for studying forced unfolding of biomolecule at the atomic level, typically uses a $\sim 10^8$ times larger pulling speed ($1\mathring{A}/ps$) than the one in experiments [101]. It is necessary to develop a simple and fast computational method that can provide us with insights on the gross dynamics of biopolymer without distorting the real dynamics. The energetics of biomolecular self-assembly processes are characterized by broad levels of hierarchical structure. We expect that the overall mechanism does not change as long as the gross structure of energy landscape is retained. One can effectively study the phenomenon of interest by properly choosing a minimal set of coarse-grained dynamical variables and interaction potential among them. One of such implementations is the off-lattice Brownian dynamics simulation using coarse-grained model. Small-sized proteins or RNA have been simulated, providing useful insights on folding/unfolding dynamics [102, 103, 72]. Nevertheless, this method also encounters timescale and size problems

for molecules of large size. Given the phenomena of interest the key task is to find the *minimal representation* of the biomolecules so that several predictions can be made without unduly compromising reality.

The elastic network model (ENM) [104, 105, 106] has been remarkably successful in predicting the equilibrium dynamics of biopolymer in native conformation. This representation of structure is not hampered by the timescale or size problem (see [107, 108]). The basis of ENM is established on the following fundamental physical principles. (i) Biologically relevant global motions are determined by the first several low frequency normal modes. (ii) As long as the overall topology is maintained, details of structure or details of interactions should not affect the global motion qualitatively. (iii) The success of ENM in describing global motion shows "network connectivity" or "topology" is the major factor that determines the overall dynamics. The role of topology has been both directly and indirectly appreciated in the context of proteins and RNA folding/unfolding problem. First, the implementation of Gõ interaction [109] is connected with the idea that folding dynamics takes place mainly by regaining the native contact topologies. Many examples using Gõ interaction for self-assembly simulation are found in the literature [110, 111, 112]. Second, a study of the force-induced unfolding dynamics of model proteins has led to the conclusion that the unraveling pathways of a protein strand is determined by the stability of strand and the topology in the native state [46]. Lastly, there has been efforts to relate the topology with protein folding kinetics. "Relative contact order" is one of the attempts to characterize the topology of protein folds [113, 114].

Here we, combining the lessons from ENM and from the role of topology gleaned

in literatures, propose the minimal model for simulating the folding/unfolding of biomolecules with emphasis on chain dynamics. System Hamiltonian of biomolecules can effectively be represented by only two terms with appropriate parameters: chain connectivity with FENE potential, Gõ interaction with Lennard-Jones potential (See Method). We name this model "topology model". Force-induced unfolding and refolding dynamics of RNAs will be investigated.

5.2 Method

5.2.1 Topology model

Supposing one wants to probe the disruption of native contact during the process of force-induced unfolding dynamics, what would be the first order potential suitable for this purpose? Consider the composition of force field in MD or typical energy potential in off-lattice simulation. The covalent bonds along the backbone and the attractive potential stabilizing the native contacts are the two major components to maintain the native topology. In addition, no interaction is greater than the excluded volume interaction while chain undergoes a large scale dynamics. Meanwhile, the bond-angle and the dihedral-angle potentials are rather local interactions that do not necessarily contribute to the large scale motion. If our interest is in the overall dynamics of strands rather than in the detailed motion of individual residues, we can safely neglect the local interaction contributions. Under physiological conditions, counterion concentration is large enough to effectively screen the electrostatic repul-

sion between the phosphate groups. To this end, we put forward a simplified energy potential which retains the chain connectivity and Lennard-Jones type Gõ-like interaction for the stability of the native state. Coarse-graining of RNA is done such that the center of mass of a nucleotide becomes an interaction site.

For the present purpose all the parameters given below are for RNA molecules. This method of coarse-graining can be easily adopted to describe proteins as well. Corresponding parameters for proteins are found in Table.5.1. The total energy potential is given as

$$\begin{aligned} V_{tot} &= V_{FENE} + V_{nb}^{(att)} + V_{nb}^{(rep)} \\ &= -\sum_{i=1}^{N-1} \frac{k}{2} R_0^2 \log(1 - \frac{(r_{i,i+1} - r_{i,i+1}^o)^2}{R_0^2}) \\ &+ \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \epsilon_h [(\frac{r_{ij}^o}{r_{ij}})^{12} - 2(\frac{r_{ij}^o}{r_{ij}})^6] \Delta_{ij} + \sum_{i<j} \epsilon_l \phi (\frac{\sigma}{r_{ij}})^6 (1 - \Delta_{ij}). \end{aligned} \tag{5.1}$$

The first term is for the chain connectivity. The finite extensible nonlinear elastic (FENE) potential [115] is used with $k = 20 kcal/(mol \cdot \text{Å}^2)$, $R_0 = 8\text{Å}$, and $r_{i,i+1}$ is the distance between neighboring beads i and $i+1$, $r_{i,i+1}^o$ is the distance in the native structure. FENE potential is a better way to represent bond stretching potential than harmonic potential especially when it comes to the issue of stretching because the finite extension of bond. Second, the Lennard-Jones potential is used for stabilizing the native topology. Native contact is defined for the pair of reaction centers whose distance is less than $R_C = 12\text{Å}$ in native state. If i and j reaction centers are in contact in native state, $\Delta_{ij} = 1$, otherwise $\Delta_{ij} = 0$. $\epsilon_h = 1\ kcal/mol$ in the potential for native pairs, $\epsilon_l = 1\ kcal/mol$ and $\phi = 1$. To ensure that chain do not self cross, we set $\sigma = 7\text{Å}$. With $\phi = 0.2$ we additionally assigned repulsion between the neighboring

beads along the backbone because V_{FENE} potential does not prevent the volume overlap. In actual computations, the time spent to calculate Lennard-Jone force is scaled as $\mathcal{O}(N^2)$, which is time-consuming. The drastic savings in computational time (approximately half) can be achieved by truncating Lennard-Jones potential between pairs with $r_{ij} > 3r_{ij}^o$, 3σ beyond which the force is effectively zero. We will name this model as "topology model" because it only uses the topological constraint of the native structure.

Using the topology model, we simulated the mechanical unfolding of various RNAs from a simple hairpin to a large ribozyme (number of nucleotides $N \approx 400$). Since a typical value for the mass of a nucleotide, $m \sim 300g/mol - 400g/mol$, the average distance between the adjacent beads $a = 5$ Å, the energy scale $\epsilon = 1 \sim 2kcal/mol$, the natural time is $\tau_L = (\frac{ma^2}{\epsilon})^{1/2} = 3 \sim 5\ ps$. We use $\tau_L = 4.0ps$ to convert simulation times into real times. To estimate the time scale for mechanical unfolding dynamics we use a Brownian dynamics algorithm [76, 77] for which the natural time for the overdamped motion is $\tau_H = \frac{\zeta\epsilon}{T}\tau_L$. We used $\zeta = 100\tau_L^{-1}$ in the overdamped limit, that approximately corresponds to friction constant in water. All the force simulations are performed at $T = 300\ K$. For the integration time step $h = 0.05\tau_L$, 10^6 integration time steps correspond to 33 μs. The harmonic spring with spring constant k_s attached to the 3' end of molecule is pulled with constant speed. The length of entire system defined by RNA and spring is subject to increase along the force direction by δx with every 10^4 integration time steps. For the increment $\delta x = 0.03$Å with every 10^4 integration time steps, $v = 8.8\mu m/s$ for $h = 0.05\tau_L$

Table 5.1: Parameters for topology model of RNA and proteins.

	RNA	protein
R_0	8Å	8Å
k	$20kcal/(mol \cdot \mathring{A}^2)$	$20kcal/(mol \cdot \mathring{A}^2)$
R_C	12Å	8Å
ϵ_h	$1kcal/mol$	$2kcal/mol$
ϵ_l	$1kcal/mol$	$1kcal/mol$
σ	7Å	3Å
ζ	$100\tau_L^{-1}$	$50\tau_L^{-1}$
τ_L	$4ps$	$3ps$

5.3 Results and Discussions

5.3.1 Stretching *Tetrahymena thermophila ribozyme*

Tetrahymena thermophila ribozyme (*T.* ribozyme) is one of the most intensively studied RNAs (Fig.5.1-A) whose three dimensional structure is well identified [116, 117]. Bustamante *et al* have recently investigated the mechanical response of L-21 sca construct of the ribozyme using laser optical tweezers [2, 23]. In the presence of Mg^{2+} the force/extension curve of *T.* ribozyme exhibits six transitions corresponding to successive unfolding events. Correlation between the force peaks and unfolding of domains or subdomain depicted in Fig.5.1-A is elucidated by a rather laborious experimental procedure. Force/extension curves of individual domain are studied separately, followed by the investigation on the mechanical response of the larger

multidomain complex assembled from individual domain [23]. Despite the structural complexity of *T. ribozyme*, the analysis is straightforward thanks to the hierarchical nature of the RNA structure, i.e., the notion that the energy scale between secondary interaction and tertiary interaction in RNA is well separated. However, it is questionable whether a similar experimental procedure for force peak identification is applicable to proteins where the separation between secondary and tertiary structure is not so clear as in RNA.

We computed force/extension curves using topology model of *T. ribozyme* at varying pulling speeds. For the complete simulation of the entire structure from P1 to P9 domain, native structure of *T. ribozyme* is taken from Westhof's theoretical model [118]. Mg^{2+} dependent interaction with subsequent unzipping of secondary structure is implicitly taken into account by properly selecting the energy parameters (See Method). As a result force/extension curves with multiple force peaks characterizing the unfolding event of domains are obtained by unfolding the topology model at varying pulling speeds (Fig.5.1-B). The assignment of force peaks using the simulation result of topology model is much simpler than the one in experiment. Comparison of force, extension, and rupture history (Fig.5.1-C) as a function of time leads us to identify the force peaks for each trajectory (See Appendix A). According to Bustamante *et al* the unfolding pathway at loading rates of 3-5 pN/s is P9.2-P9.1-[P9,P9.1a]-[P2,P2.1]-[P7,P3,P8]-[P6,P4]-P5-[P5a,P5b,P5c], where opening of peripheral region precedes the core region.

We tried to reduce the loading rate as much as possible ($v = 3.8\ \mu m/s$, $k_s = 3.5\ pN/nm$), but a wide gap from the experimental loading rate ($v \sim 0.1\ \mu m/s$, $k_s \sim 0.01$

pN/nm, which lead to $r_f \sim 1\ pN/s$) was unavoidable due to large system size of *T.* ribozyme. Despite the disparity between the loading rates in experiments and our simulations, we surprisingly retrieved the same unfolding sequence extracted by Bustamante *et al.* The magnitude of unfolding force is larger than the one in experiment because of the higher loading rate (note that unfolding force increases logarithmically with loading rate [34]), but the correct unfolding sequence of *T. ribozyme* suggests that simulation is performed in the time regime where the molecule has enough time to rearrange its conformation. Note that as the loading rate increases the unfolding pathways start deviating from the one in small loading rate, i.e., unfolding pathway are r_f-dependent. At four fold larger loading rate the first noticeable change is between P2 and P378 domain where P378 domain opening precedes P2. At even higher loading rate, *T. ribozyme* gradually unfolds from the P9 domain and unfolding ends at P2 domain. The bigger the system size, the smaller loading rate is required to establish a uniform tension along the chain. We clarify this point using smaller-sized *Azoarcus* ribozyme.

5.3.2 Pulling speed dependent unfolding pathways

We observe the pulling speed dependent unfolding pathways in *Azoarcus* ribozyme for which the crystal structure has recently been determined [119]. The *Azoarcus* ribozyme structure is similar to the catalytic core of *T.* ribozyme, characterized by its pseudoknot. Since the size is the half of *T.* ribozyme, the computational

cost is reduced. We generated multiple unfolding trajectories at three different loading rates around which the pulling speed dependent unfolding pathways are well characterized. The average force-extension curves are depicted at three different loading rates in Fig.5.2-B. In force/extension curve, we observe four conspicuous force peaks at large loading rate ($v = 30\mu m/s$, $k_s = 28pN/nm$), whereas only three peaks are discerned at medium ($v = 9\ \mu m/s$, $k_s = 28\ pN/nm$) and small loading rates ($v = 3.8\ \mu m/s$, $k_s = 3.5\ pN/nm$). Identification of the underlying conformational change only through the given force/extension curve is, if not impossible, a challenging problem. Reference to a model simulation which can visualize the major dynamical event would be of great help. The origin of changing number of major peaks can be explained when the contact rupture history is inspected (Fig.5.2-C). At medium and small loading rates, the events of P1 and P6 domain opening are synchronous. The first broad peak in force-extension curve at small loading rate is, indeed, a merged peak from the two.

As is a typical case in both experimental setups like AFM and LOT, the force is acted to one end with the other end being fixed. Unlike other perturbations such as temperature, denaturant, and pH change, the mechanical force does not perturb the system uniformly. Strong but local perturbation acting on one end should be propagated through the bond network of the molecule to whole system. Rate of tension propagation depends on chain stiffness, friction coefficient, and chain conformation at the moment tension is applied [84]. When a time average of bond segment vector correlates with force direction (strictly speaking, the bond segment vector cannot become completely parallel to force direction because of thermal fluc-

tuation. Assuming freely jointed chain molecule, we can relate angle and force as $\langle \cos\theta_i \rangle = \coth\left(fr^o_{i,i+1}/k_BT\right) - k_BT/fr^o_{i,i+1}$), tension acting on that particular segment converges to the applied force at the end. To visualize the force propagation we compute the angles between bond segment vectors and force direction (Fig.5.2-D,E,F) during the forced unfolding process. The signature of finite tension propagation rate can be gleaned from the angle gradient captured with dark to light red color ($\langle \cos\theta \rangle \leq 1$) especially at the large loading rate (Fig.5.2-E, top panel). Direct comparison of the first (θ_1) and the last alignment angle (θ_{N-1}) illuminates the discrepancy between force values at 3' and 5' end. Note how different these two alignment angles are at varying loading rates (Fig.5.2-F).

Fig.5.3 recapitulates the origin of pulling speed dependent unfolding pathway variation. Forced unfolding does not necessarily initiate from the less stable domain unless the force acting on each domain is uniform. There could be a certain loading rate, $r_f = r_f^c$, at which effective free energy barriers in both pathways are similar. In this case we expect kinetic partitioning of forced unfolding. In fact, in pulling simulations of *Azoarcus* ribozyme at high loading rate ($v = 30\,\mu m/s$, $k_s = 28\,pN/nm$) we observe a few trajectories follow the same pathway as in medium loading rate. In force/extension curves we observe only three force peaks instead of four. The inspection of contact rupture history clarifies the heterogeneity of unfolding pathways (See Appendix B).

5.3.3 Refolding

The use of topology model should not be limited to the study of forced unfolding dynamics. The underlying principle adopted in topology model, "native topology determines the force-induced unfolding pathway", can be applied for folding dynamics. It is of particular interest to see the refolding mechanism that quenching the force to zero value drives the stretched state of molecule close to native state. We generated 20 refolding trajectories of P4-P6 domain (The domains enclosed by blue rectangle in Fig.5.1-A. Crystal structure with PDB code 1GID is used for simulation) of *T.* ribozyme for finite time. Fig.5.4-A shows one of refolding trajectories monitored by the fraction of native contact ($Q(t)$). Native structure formation is initiated from the P5b and P6b hairpin loop and the zipping of secondary structure proceeds within a reltively short time ($< 5\ ms$). Loop formation dynamics is a key event in early stage of folding from stretched conformation. Conformational search for the tertiary contact generates a long pause with respect to $Q(t)$ and $R(t)$. Because 3' and 5' end are also the part of hairpin stack and close in distance in the native state, upon formation of secondary structure involved with 3' and 5', $R(t)$ does not reveal any further dynamical information about the structure. Meanwhile, the radius of gyration, $R_g(t)$, exhibits the size fluctuation throughout the whole process. Following another structural transition of $Q(t)$ from 0.6 to 0.7, tertiary contacts P5b-P6b and P5a/P5c-P4 are observed in structure as well as in contact formation history (Fig.5.4-B). The folding mechanism of RNA exhibits the hierarchical nature of RNA structure.

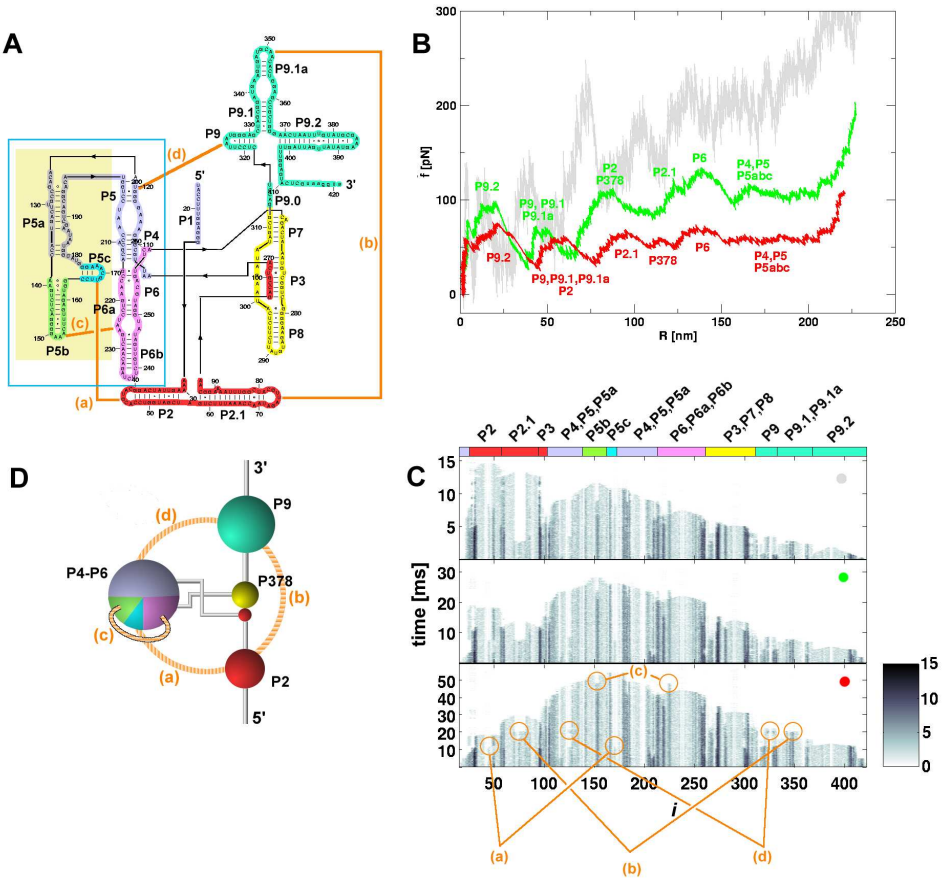

Figure 5.1: **(A)** Secondary structure of *Tetrahymena thermophila* ribozyme. Each subdomain is specified by paired helices P1 to P9. Four major tertiary contacts are identified with orange line as (a) P2-P5c (b) P2.1-P9.1a (c) P5b-P6a (d) P5-P9. **(B)** Force-extension curves of *Tetrahymena* ribozyme computed through forced unfolding simulation at varying loading rates ($v = 15\ \mu m/s$, $k_s = 70\ pN/nm$ in gray, $v = 7.5\ \mu m/s$, $k_s = 7.0\ pN/nm$ in green, and $v = 3.8\ \mu m/s$, $k_s = 3.5\ pN/nm$ in red). Each force peak is identified as the opening event of a subdomain through the comparison of R(t), f(t) with contact rupture history. Note that the sequence of subdomain opening differs at different loading rate. **(C)** The history of contact rupture at three different loading rate is shown. The colors on the top correspond to the domains. The signature of tertiary contact rupture can explicitly be identified. **(D)** Simplified structural scheme of *T.* ribozyme. Each bead in color corresponds to a specific domain in **A**. It is possible to guess the time sequence of rupture event simply by inspecting this topologically simplified model. Upon application of force to the 3' end P9 and/or P2 are affected. Hence, tertiary contact (a), (b), (d) rupture at early stage. (c) is intact unless interaction between P378 and P2 (pseudoknot) disrupts.

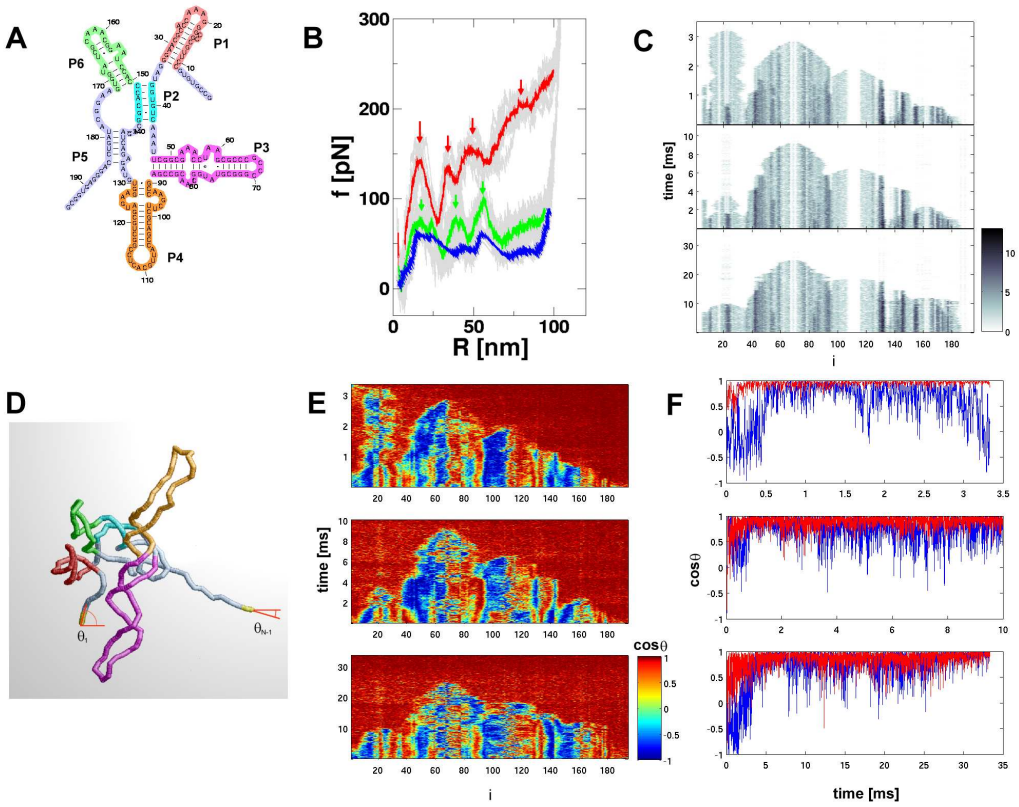

Figure 5.2: **(A)** Secondary structure of *Azoarcus* ribozyme. **(B)** Force-extension curves of *Tetrahymena* ribozyme computed through forced unfolding simulation at varying loading rates ($v = 30\ \mu m/s$, $k_s = 28\ pN/nm$ in red, $v = 9\ \mu m/s$, $k_s = 28\ pN/nm$ in green, and $v = 3.8\ \mu m/s$, $k_s = 3.5\ pN/nm$ in blue). Each of colored force/extension curve is averaged on multiple (9-10) trajectories. **(C)** The history of contact rupture at three different loading rate. **(D)** Snapshot of topology model of *Azoarcus*. The first and the last alignment angle between bond-vector and force direction is specified. **(E)** Time evolutions of $\cos\theta_i$ $(i = 1, \cdots N-1)$ under pulling dynamics at three different loading rate are depicted. The values of $\cos\theta_i$ are color-coded as indicated on the scale shown on the right of bottom panel. **(F)** Discrepancy in force values at opposite ends is greater at high loading rate. This is shown by comparing time evolutions of $\cos\theta_1$ (blue) and $\cos\theta_{N-1}$ (red) for three loading rates.

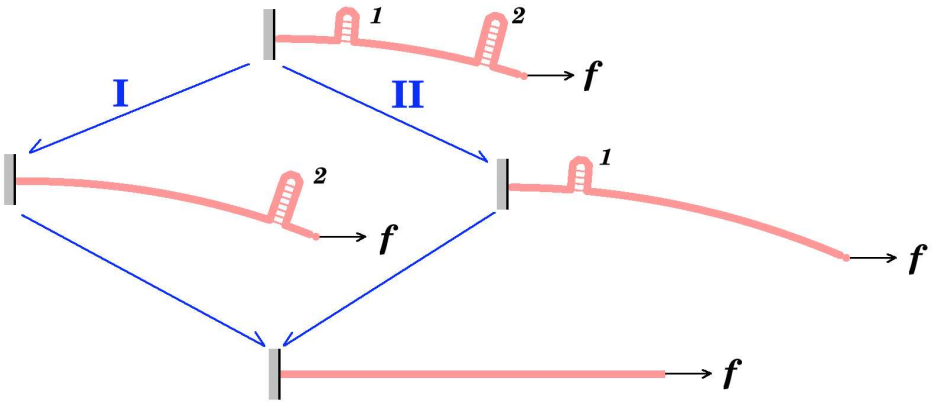

Figure 5.3: Illustration of pulling speed dependent unfolding pathways for a double hairpin loop RNA. Depending on the pulling speed, the magnitude of effective force acting on each hairpin loop (1 and 2) varies. Let these force be f_1 and f_2. Suppose unfolding free energy barrier for each hairpin loop 1 and 2 in the absence of force is $\Delta G_1^{\ddagger}$ and $\Delta G_2^{\ddagger}$ with $\Delta G_1^{\ddagger} < \Delta G_2^{\ddagger}$, then the effective barrier of hairpin loop under force becomes $\Delta G_1^{\ddagger} - f_1 \times \delta z$ and $\Delta G_2^{\ddagger} - f_2 \times \delta z$, where δz is the distance between native state and transition state along the force direction ($\delta z_1 = \delta z_2 = \delta z$ is assumed for simplicity). At slow pulling speed, $f_1 \approx f_2$, $(\Delta G_1^{\ddagger} - f_1 \times \delta z) < (\Delta G_2^{\ddagger} - f_2 \times \delta z)$, thus the dynamics along pathway I will dominate. On the other hand, at high pulling speed, since $f_1 < f_2$, the relation $(\Delta G_1^{\ddagger} - f_1 \times \delta z) > (\Delta G_2^{\ddagger} - f_2 \times \delta z)$ becomes plausible. In this case the unfolding dynamics will take place along pathway II.

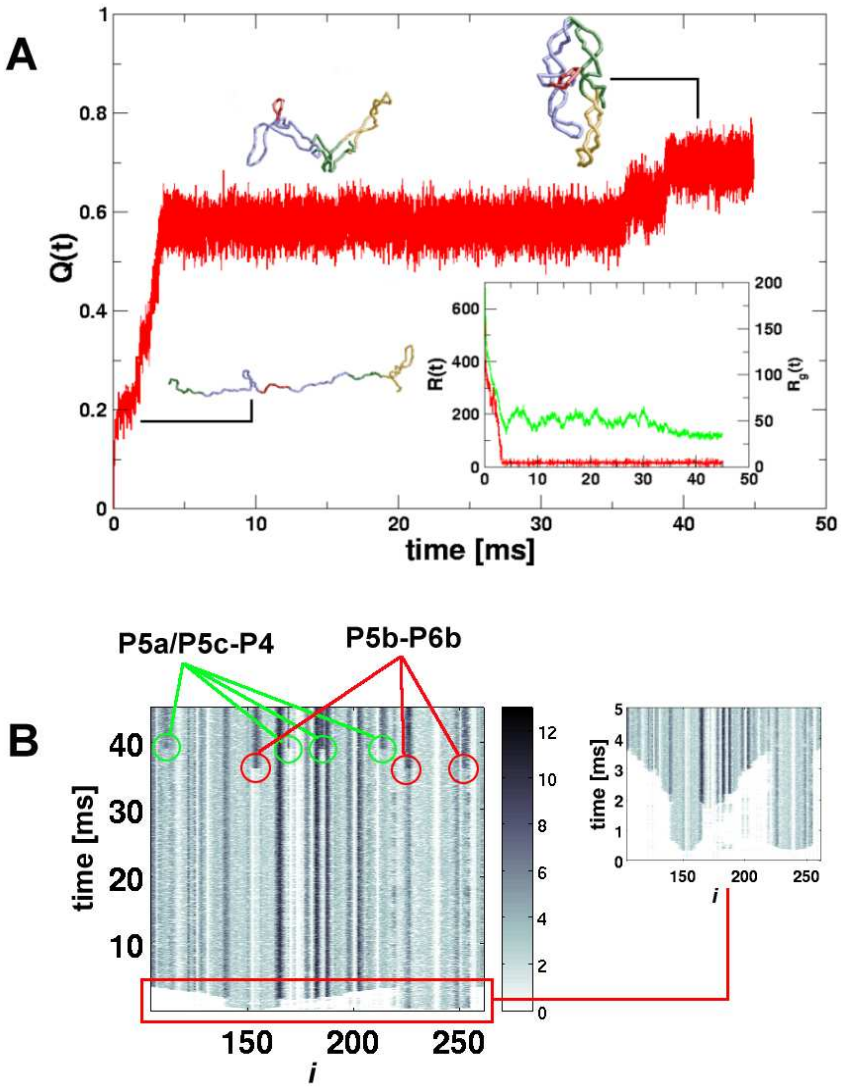

Figure 5.4: **(A)** Refolding dynamics of P4-P6 domain of *T.* ribozyme monitored by $Q(t)$. Dynamics by $R(t)$ (red) and $R_g(t)$ (green) is in the inset. **(B)** Contact formation history corresponding to the dynamics in **A**. Early stage of folding related to the hairpin loop formation enclosed in red box is separately shown in the right panel. Zipping of secondary structure is propagated from around P5b Formation of tertiary contacts ($\sim 40ms$) is enclosed by the circles.

5.4 Conclusions

The success of topology model simulation in predicting the unfolding pathway is the consequence of the notion that "native topology determines the force-induced unfolding pathway" [46]. The movies of unfolding process shows acrobatic chain dynamics that the molecule reorganizes its conformation in such a way that volume crash or entanglement is avoided. Especially, when the dynamics of molecule is driven externally, it is likely that the interaction due to volume exclusion becomes the major issue. We believe that the chain dynamics of biomolecule under force is well represented by our model. There are many examples associated with experimental study on folding dynamics of proteins and RNA molecule whose size and time scale is far beyond the scope of conventional simulation techniques. The unfolding dynamics of proteins as well as RNA with size up to 250 residue can be conveniently simulated on a PC in a few days with the simulation condition used in this paper. Topology model can serve as a useful tool that provides a quick insight into the folding/unfolding process of topologically involved large macromolecules within the current computational cost. The applicability of topology model is expandable beyond monomer dynamics. Since the diffusional time scale, $> \mu s$, is easily accessible in topology model, the dynamics of multimer formation from the denatured monomer would be a good subject of investigation. It would also be of great interest to study the folding/unfolding pathways of topologically distinct protein folds using topology model.

Appendix A

Identification of force peak through comparison between force and contact rupture as a function of time

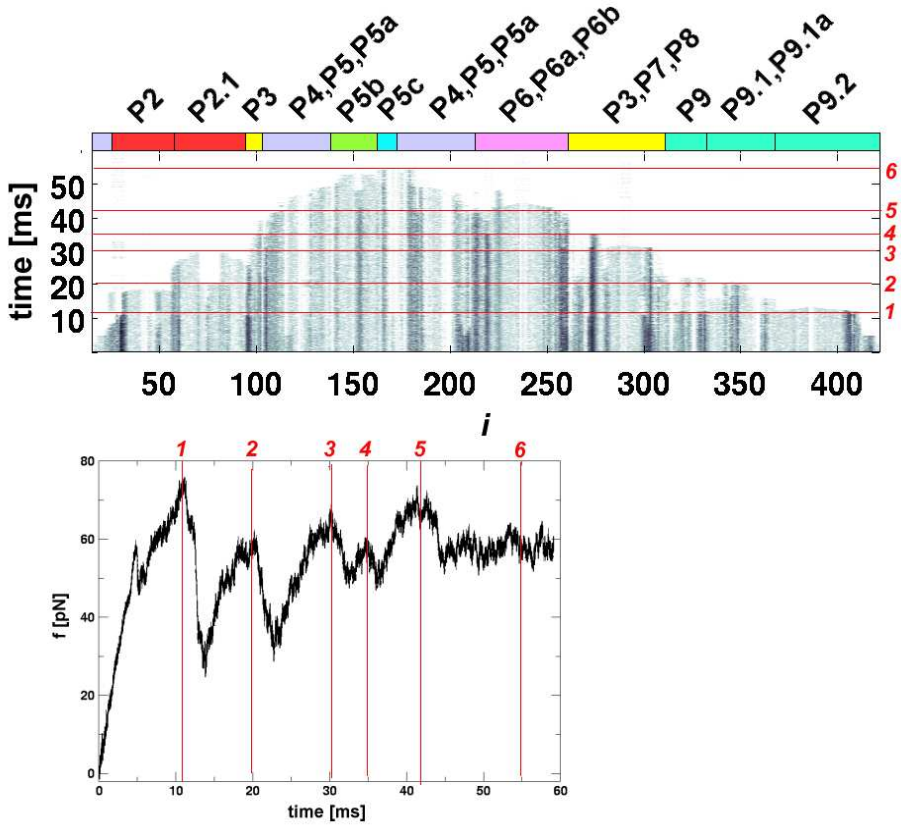

Figure A.1: Force peaks in the force/extension curve are assigned by the comparison between $f(t)$ (bottom panel) and contact rupture history (top panel). The time lines labeled from 1 to 6 on the plot is used to identify the corresponding events on both panels.

Appendix B

Contact rupture history of Azoarcus ribozyme at varying loading rates

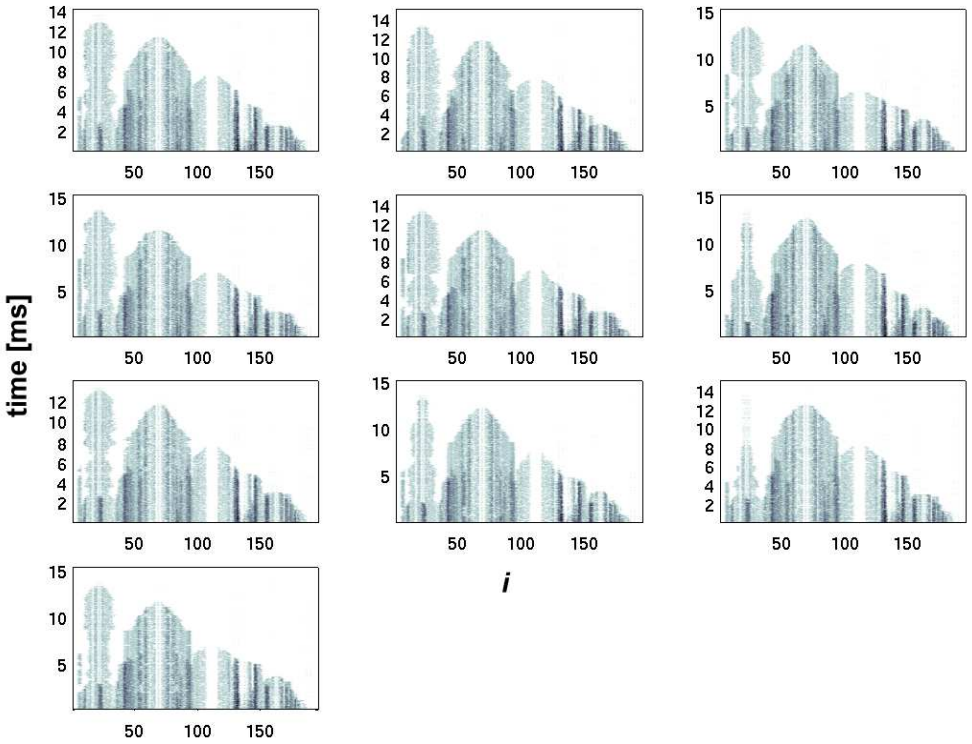

Figure B.1: Large loading rate, $v = 30\mu m/s$, $k_s = 28pN/nm$, $r_f = 8.4 \times 10^5 pN/s$

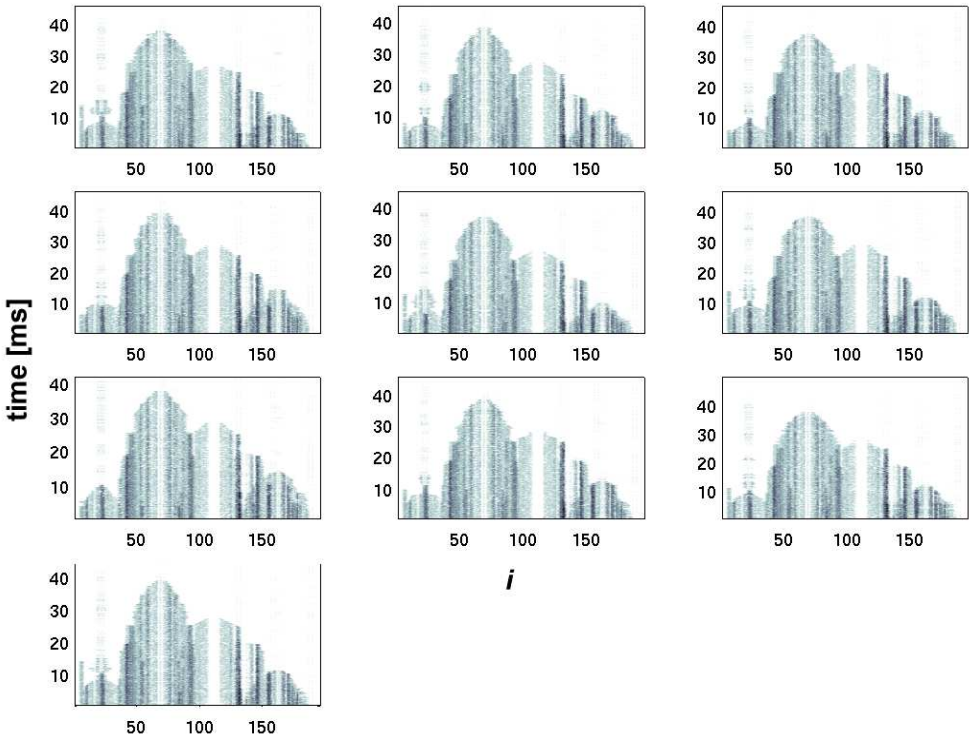

Figure B.2: Intermediate loading rate, $v = 9\mu m/s$, $k_s = 28pN/nm$, $r_f = 2.52 \times 10^5 pN/s$

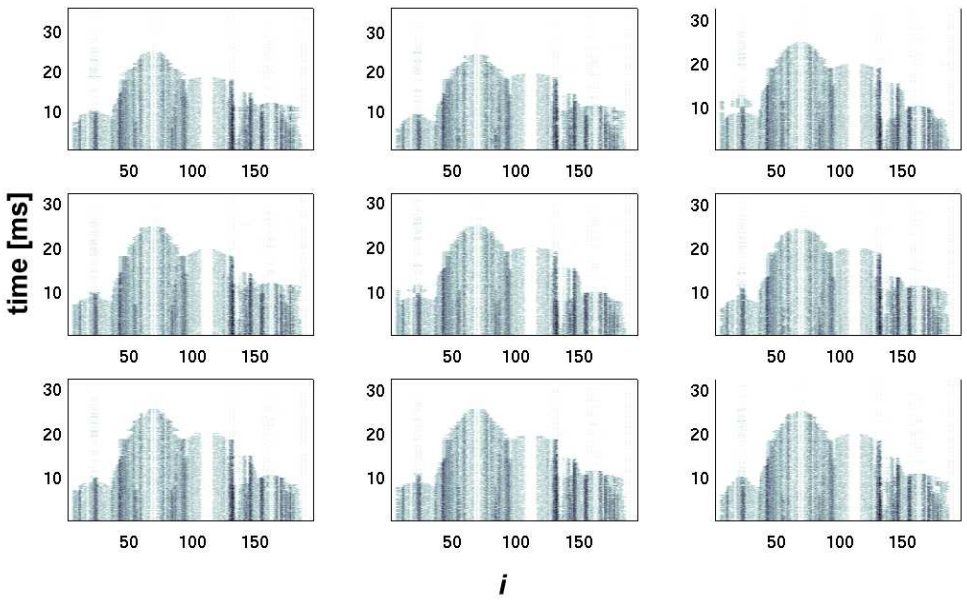

Figure B.3: Small loading rate, $v = 3.8\mu m/s$, $k_s = 3.5pN/nm$, $r_f = 1.33 \times 10^4 pN/s$

BIBLIOGRAPHY

1. Rudisser, S. and Tinoco, Jr., I. Solution Structure of Cobalt(III)Hexammine Complexed to the GAAA Tetraloop, and Metal-ion Binding to GA Mismatches. *J. Mol. Biol.* **295**, 1211–1223 (2000).

2. Liphardt, J., Onoa, B., Smith, S., Tinoco, Jr., I., and Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292**, 733–737 (2001).

3. Levinthal, C. Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44 (1968).

4. Zwanzig, R., Szabo, A., and Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci.* **89**, 20–22 (1992).

5. Perkins, T. T., Smith, D. E., and Chu, S. Single polymer dynamics in an elongation flow. *Science* **276**, 2016–2021 (1997).

6. Smith, D. E. and Chu, S. Response of flexible polymers to a sudden elongational flow. *Science* **281**, 1335–1340 (1998).

7. Smith, D. E., Babcock, H. P., and Chu, S. Single-Polymer Dynamics in Steady Shear Flow. *Science* **283**, 1724–1727 (1999).

8. Schroeder, C. M., Babcock, H. P., Shaqfeh, E. S. G., and Chu, S. Observation of polymer conformation hysteresis in extensional flow. *Science* **301**, 1515–1519 (2003).

9. Rhoades, E., Gussakovsky, E., and Haran, G. Watching proteins fold one molecule at a time. *Proc. Natl. Acad. Sci.* **100**, 3197–3202 (2003).

10. Zhuang, X., Bartley, L., Babcock, A., Russell, R., Ha, T., Hershlag, D., and Chu, S. A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048–2051 (2000).

11. Smith, S. B., Finzi, L., and Bustamante, C. Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads. *Science* **258**, 1122–1126 (1992).

12. Maier, B., Bensimon, D., and Croquette, V. Replication by a single dna polymerase of a stretched single-stranded dna. *Proc. Natl. Acad. Sci.* **97**, 12002–12007 (2000).

13. Bryant, Z., Stone, M. D., Gore, J., Smith, S. B., Cozzarelli, N. R., and Bustamante, C. Structural transitions and elasticity from torque measurements on dna. *Nature* **424**, 338–341 (2003).

14. Barkai, E., Jung, Y., and Silbey, R. Theory of single-molecule spectroscopy: Beyond the ensemble average. *Annu. Rev. Phys. Chem.* **55**, 457–507 (2004).

15. Kellermeyer, M. Z., Smith, S. B., Granzier, H. L., and Bustamante, C. Folding-Unfolding Transitions in Single Titin Molecules Characterized by Force-Measuring Laser Tweezers. *Science* **276**, 1112–1116 (1997).

16. Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. Entropic elasticity of λ-phase DNA. *Science* **265**(5178), 1599–1600 (1994).

17. Smith, S. B., Cui, Y., and Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **271**, 759–799 (1996).

18. Reif, M., Pascual, J., Saraste, M., and Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.* **286**, 553–561 (1999).

19. Yang, G., Cecconi, C., Baase, W. A., Vetter, I. R., Breyer, W. A., Haack, J. A., Matthews, B. W., Dahlquist, F. W., and Bustamante, C. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci.* **97**, 139–144 (2000).

20. Chyan, C.-L., Lin, F.-C., Peng, H., Yuan, J.-M., Chang, C.-H., Lin, S.-H., and Yang, G. Reversible mechanical unfolding of single ubiquitin molecules. *Biophys. J.* **87**, 3995–4006 (2004).

21. Fernandez, J. and Li, H. Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science* **303**(5664), 1674–1678 (2004).

22. Cecconi, C., Shank, E. A., Bustamante, C., and Marqusee, S. Direct Observation of Three-State Folding of a Single Protein Molecule. *Science* **309**, 2057–2060 (2005).

23. Onoa, B., Dumont, S., Liphardt, J., Smith, S., Tinoco, Jr., I., and Bustamante, C. Identifying Kinetic Barriers to Mechanical Unfolding of the *T. thermophila* Ribozyme. *Science* **299**, 1892–1895 (2003).

24. Discher, B. M., Won, Y. Y., Ege, D. S., Lee, J. C.-M., Bates, F. S., Discher, D. E., and Hammer, D. A. Polymersomes: Tough vesicles made from diblock copolymers. *Science* **284**, 1143–1146 (1999).

25. Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H. E., and Müller, D. J. Unfolding pathways of individual bacteriorhodopsins. *Science* **288**, 143–146 (2000).

26. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R., and Block, S. M. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684–687 (2003).

27. Pease, P. J., Levy, O., Cost, G. J., Gore, J., Ptacin, J. L., Sherratt, D., Bustamante, C., and Cozzarelli, N. Sequence-directed dna translocation by purified ftsk. *Science* **307**, 586–590 (2005).

28. Ivanovska, I. L., de Pablo, P. J., Ibarra, B., Sgalari, G., MacKintosh, F. C., Carrascosa, J. L., Schmidt, C. F., and Wuite, G. J. L. Bacteriophage capsids: Tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci.* **101**, 7600–7605 (2004).

29. Brenowitz, M., Chance, M. R., Dhavan, G., and Takamoto, K. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical 'footprinting'. *Curr. Opin. Struct. Biol.* **12**, 648–653 (2002).

30. Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).

31. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).

32. Kramers, H. A. Brownian motion a field of force and the diffusion model of chemical reaction. *Physica* **7**, 284–304 (1940).

33. Hanggi, P., Talkner, P., and Borkovec, M. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).

34. Evans, E. and Ritchie, K. Dynamic Strength of Molecular Adhesion Bonds. *Biophys. J.* **72**, 1541–1555 (1997).

35. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci.* **96**, 3694–3699 (1999).

36. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science* **276**, 1109–1111 (1997).

37. Onuchic, J. N. and Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).

38. Thirumalai, D. and Woodson, S. A. Kinetics of Folding of Proteins and RNA. *Acc. Chem. Res.* **29**, 433–439 (1996).

39. Thirumalai, D. and Hyeon, C. RNA and Protein folding: Common Themes and Variations. *Biochemistry* **44**(13), 4957–4970 (2005).

40. Isralewitz, B., Gao, M., and Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol* **11**, 224–230 (2001).

41. Camacho, C. and Thirumalai, D. Kinetics and Thermodynamics of Folding in Model Proteins. *Proc. Natl. Acad. Sci.* **90**, 6369–6372 (1993).

42. Guo, Z. and Thirumalai, D. Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* **36**, 83–102 (1995).

43. Guo, Z. and Thirumalai, D. Kinetics and Thermodynamics of Folding of a de Novo Designed Four-Helix Bundle Protein. *J. Mol. Biol.* **263**, 323–343 (1996).

44. Bryngelson, J. D. and Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding,. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528 (1987).

45. Bryngelson, J., Onuchic, J., Socci, N., and Wolynes, P. Funnels, pathways, and the energy landscape of protein-folding - A synthesis. *Proteins: Structure, Function and Genetics* **21**, 167–195 (1995).

46. Klimov, D. and Thirumalai, D. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. USA* **97**, 7254–7259 (2000).

47. Zwanzig, R. Diffusion in rough potential. *Proc. Natl. Acad. Sci. USA* **85**, 2029–2030 (1988).

48. Nevo, R., Brumfeld, V., Kapon, R., Hinterdorfer, P., and Reich, Z. Direct measurement of protein energy landscape roughness. *EMBO reports* **6**(5), 482 (2005).

49. Dill, K. A. and Chan, H. S. From Levinthal to Pathways to Funnels. *Nature Struct. Biol.* **4**, 10–19 (1997).

50. Jackson, S. E. How do proteins fold? *Folding Des.* **3**, R81–R91 (1998).

51. Lapidus, L., Eaton, W., and Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA* **97**, 7220–7225 (2000).

52. Tskhovrebova, L., Trinick, J., Sleep, J., and Simmons, R. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).

53. Reif, M., Gautel, H., Oesterhelt, F., Fernandez, J., and Gaub, H. Reversible unfolding of individual titin immunoglobulin domains by AFM . *Science* **276**, 1109–1112 (1997).

54. Zhuang, X. and Reif, M. Single-molecule folding. *Curr. Opin. Struct. Biol.* **13**, 86–97 (2003).

55. Liphardt, J., Dumont, S., Smith, S., Tinoco, Jr., I., and Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **296**, 1832–1835 (2002).

56. Hummer, G. and Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **98**, 3658–3661 (2001).

57. Klimov, D. and Thirumalai, D. Lattice model studies of force-induced unfolding of protein. *J. Phys. Chem. B* **105**, 6648–6654 (2001).

58. Nymeyer, H., Socci, N., and Onuchic, J. Landscape approaches for determining the ensemble of folding transition states : Success and failure hinge on the degree of frustration. *Proc. Natl. Acad. Sci. USA* **97**, 634–639 (2000).

59. Lifson, S. and Jackson, J. On self-diffusion of ions in a polyelectrolyte solution. *J. Chem. Phys.* **36**, 2410–2414 (1962).

60. Rossky, P., Doll, J., and Friedman, H. Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.* **69**, 4628–4633 (1978).

61. Lide, D. *Handbook of Chemistry and Physics.* CRC press, Boca Raton, FL, 76th edition, (1995).

62. Onoa, B. and Tinoco, Jr, I. RNA folding and unfolding. *Curr. Opin. Struct. Biol.* **14**(3), 374–379 (2004).

63. Treiber, D. and Williamson, J. Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **11**, 309–314 (2001).

64. Thirumalai, D., Lee, N., Woodson, S. A., and Klimov, D. K. Early Events in RNA Folding. *Annu. Rev. Phys. Chem.* **52**, 751–762 (2001).

65. Koculi, E., Lee, N., Thirumalai, D., and Woodson, S. A. Folding of the *Tetrahymena* Ribozyme by Polyamines: Importance of Counterion Valence and Size. *J. Mol. Biol.* **341**(1), 27–36 (2004).

66. Russell, R., Zhuang, X., Babcock, H., Millett, I., Doniach, S., Chu, S., and Herschlag, D. Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci.* **99**(1), 155–160 (2002).

67. Mueller, M., Krzakala, F., and Mezard, M. The secondary structure of RNA under tension. *Eur. Phys. J. E* **9**, 67–77 (2002).

68. Cocco, S., Marko, J., and Monasson, R. Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur. Phys. J. E* **10**, 153–161 (2003).

69. Gerland, U., Bundschuh, R., and Hwa, T. Force-Induced Denaturation of RNA. *Biophys. J.* **81**(3), 1324–1332 (2001).

70. Gerland, U., Bundschuh, R., and Hwa, T. Mechanically Probing the Folding Pathway of Single RNA Molecules. *Biophys. J.* **84**, 2831–2840 (2003).

71. Klimov, D. and Thirumalai, D. Stretching single-domain proteins: Phase diagram and kinetics of force-induced unfolding. *Proc. Natl. Acad. Sci.* **96**(11), 6166–6170 (1999).

72. Hyeon, C. and Thirumalai, D. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci.* **102**, 6789–6794 (2005).

73. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* **91**, 9218–9222 (1994).

74. Mathews, D., Sabina, J., Zuker, M., and Turner, D. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **288**, 911–940 (1999).

75. Misra, V. and Draper, D. A thermodynamic framework for Mg^{2+} binding to RNA. *Proc. Natl. Acad. Sci.* **98**(22), 12456–12461 (2001).

76. Ermack, D. and McCammon, J. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **69**, 1352–1369 (1978).

77. Klimov, D., Betancourt, M., and Thirumalai, D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha helics: effect on stability, cooperativity and kinetics. *Folding Des.* **3**, 481–498 (1998).

78. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. The weighted histogram analysis method for free-energy calculation on biomolecules. I. The method. *J. Comp. Chem.* **13**(8), 1011–1021 (1992).

79. Klimov, D. and Thirumalai, D. Cooperativity in Protein Folding: From Lattice Models with Side Chanins to Real Proteins. *Folding Des.* **3**, 127–139 (1998).

80. Danilowicz, C., Coljee, V. W., Bouzigues, C., Lubensky, D. K., Nelson, D. R., and Prentiss, M. DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc. Natl. Acad. Sci. USA* **100**(4), 1694–1699 (2003).

81. Thirumalai, D. From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I (Fr.)* **5**, 1457–1467 (1995).

82. Pitard, E. and Orland, H. Dynamics of swelling or collapse of a homopolymer. *Europhys. Lett.* **41**(4), 467–472 (1998).

83. de Gennes, P. G. Molecular individualism. *Science* **276**, 1999–2000 (1997).

84. Seifert, U., Wintz, W., and Nelson, P. Straightening of thermal fluctuation in semiflexible polymers by applied tension. *Phys. Rev. Lett.* **77**(27), 5389–5392 (1996).

85. Hammond, G. S. A correlation of reaction rates. *J. Am. Chem. Soc.* **77**, 334–338 (1953).

86. Leffler, J. E. Parameters for the description of transition states. *Science* **117**, 340–341 (1953).

87. Sánchez, I. E. and Kiefhaber, T. Evidence for Sequential Barriers and Obligatory Intermediates in Apparent Two-state Protein Folding. *J. Mol. Biol.* **325**, 367–376 (2003).

88. Sánchez, I. E. and Kiefhaber, T. Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded States. *J. Mol. Biol.* **327**, 867–884 (2003).

89. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., and Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50–53 (1999).

90. Nevo, R., Stroh, C., Kienberger, F., Kaftan, D., Brumfeld, V., Elbaum, M., Reich, Z., and Hinterdorfer, P. A molecular switch between alternative conformational states in the complex of Ran and importin β 1. *Nature. Struct. Biol.* **10**, 553–557 (2003).

91. Barsegov, V. and Thirumalai, D. Dynamics of unbinding of cell adhesion molecules: Transition from catch to slip bonds. *Proc. Natl. Acad. Sci.* **102**, 1835–1839 (2005).

92. Fernandez, J. M., Li, H., and Brujic, J. Response to Comment on "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein". *Science* **306**, 411c (2004).

93. Sosnick, T. R. Comment on "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein". *Science* **306**, 411b (2004).

94. Zwanzig, R. Simple model of protein folding kinetics. *Proc. Natl. Acad. Sci.* **92**, 9801–9804 (1995).

95. Best, R. B. and Hummer, G. Comment on "Force-Clamp Spectroscopy Monitors the Folding Trajectories of a Single Protein". *Science* **308**, 498b (2005).

96. Brochard-Wyart, F., Buguin, A., and de Gennes, P. G. Dynamics of taut dna chains. *Europhys. Lett.* **47**, 171–174 (1999).

97. Lee, N. and Thirumalai, D. Pulling-speed-dependent force-extension profiles for semiflexible chains. *Biophys. J.* **86**, 2641–2649 (2004).

98. Rief, M., Fernandez, J. M., and Gaub, H. E. Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. *Phys. Rev. Lett.* **81**(21), 4764–4767 (1998).

99. Hummer, G. and Szabo, A. Kinetics from Nonequilibrium Single-Molecule Pulling Experiments. *Biophys. J.* **85**, 5–15 (2003).

100. Elber, R. Long-timescale simulation methods. *Curr. Opin. Struct. Biol.* **15**, 151–156 (2005).

101. Lu, H. and Schulten, K. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* **79**, 51–65 (2000).

102. Thirumalai, D. and Klimov, D. K. Deciphering the Time Scales and Mechanisms of Protein Folding Using Minimal Off-Lattice Models. *Curr. Opin. Struct. Biol.* **9**, 197–207 (1999).

103. Klimov, D. K. and Thirumalai, D. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci.* **97**, 2544–2549 (2000).

104. Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**(9), 1905–1908 (1996).

105. Haliloglu, T., Bahar, I., and Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **79**(16), 3090–3093 (1997).

106. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **80**, 505–515 (2001).

107. Tama, F., Valle, M., Frank, J., and Brooks III, C. L. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci.* **100**, 9319–9323 (2003).

108. Rader, A. J., Vlad, D. H., and Bahar, I. Maturation Dynamics of Bacteriophage HK97 Capsid. *Structure* **13**, 413–421 (2004).

109. Ueeda, Y., Taketomi, H., and Go, N. Studies on protein folding, unfolding and fluctuations by computer simulation. a three-dimensional lattice model o lysozyme. *Biopolymers* **17**, 1905–1908 (1978).

110. Sorin, E. J., Nakatani, B. J., Rhee, Y. M., Jayachandran, G., Vishal, V., and Pande, V. S. Does Native State Topology Determine the RNA Folding Mechanism? *J. Mol. Biol.* **337**, 789–797 (2004).

111. Nivón, L. G. and Shakhnovich, E. I. All-atom Monte Carlo Simulation of GCAA RNA Folding. *J. Mol. Biol.* **344**, 29–45 (2004).

112. Yang, S., Cho, S. S., Levy, Y., Cheung, M. S., Levine, H., Wolynes, P. G., and Onuchic, J. N. Domain swapping is a consequence of minimal frustration. *Proc. Natl. Acad. Sci.* **101**(38), 13786–13791 (2004).

113. Plaxco, K. W., Simons, K. T., and Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).

114. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).

115. Kremer, K. and Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990).

116. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. A preorganized active site in the crystal structure of the tetrahymena ribozyme. *Science* **282**, 259–264 (1998).

117. Guo, F., Gooding, A. R., and Cech, T. R. Structure of the *Tetrahymena* Ribozyme : Base Triple Sandwich and Metal Ion at the Active Site . *Mol. Cell* **16**(5), 351–362 (2004).

118. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the *Tetrahymena* thermophila ribozyme. *Chemistry & Biology* **3**, 993–1009 (1996).

119. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).