

“All we have to decide is what to do with the time that is given us.”

J.R.R. Tolkien

The Lord of the Rings (1954)

University of Alberta

**ENERGY LANDSCAPE EXPLORATION OF THE FOLDING PROCESSES OF
BIOLOGICAL MOLECULES**

by

Megan Clare Engel

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics

© Megan Clare Engel
Fall 2013
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

For decades, scientists from every discipline have struggled to understand the mechanism of biological self-assembly, which allows proteins and nucleic acids to fold reliably into functional three-dimensional structures. Such an understanding may hold the key to eliminating diseases such as Alzheimer's and Parkinson's and to effective protein engineering. The current best framework for describing biological folding processes is that of statistical mechanical energy landscape theory, and one of the most promising experimental techniques for exploring molecular energy landscapes is single molecule force spectroscopy (SMFS), in which molecules are mechanically denatured. Theoretical advances have enabled the extraction of complete energy landscape profiles from SMFS data. Here, SMFS experiments performed using laser optical tweezers are analyzed to yield the first ever full landscape profile for an RNA pseudoknot. Further, a promising novel landscape reconstruction technique is validated for the first time using experimental data from a DNA hairpin.

Acknowledgements

My supervisors, Michael Woodside and Kevin Beach, I acknowledge and deeply thank for all of their time, teaching, support, insight, and faith that somehow this thesis could be accomplished. I would also like to thank my committee members, Sharon Morsink and Mariusz Klobukowski, for their kind agreement to be part of my thesis committee under extenuating circumstances and with short notice. In particular, I would like to thank Sharon for encouraging me to finish my thesis and providing me with resources that enabled me to do so.

I would like to sincerely thank Dustin Ritchie for his work collecting the RNA pseudoknot data which made this thesis possible and for his contributions training me to use the optical tweezers.

I also would like to thank Krishna Neupane for the DNA hairpin data he collected and for his helpful explanations on many topics.

Amar Nath Gupta I acknowledge for his contributions helping me understand how to implement the Weierstrass code and for his explanations of the Jarzynski code.

The Woodside Group at large I would like to acknowledge for the time and effort they put into providing me with helpful suggestions and support.

Finally, I acknowledge my mom, dad, sister, and fiancé, Mitch, for their sustaining love and for the home they have always been to me.

Contents

1	Introduction	1
2	Energy Landscape Extraction Methods	10
2.1	Equilibrium Force Measurements	14
2.2	Non-Equilibrium Force Measurements	18
2.3	Model-Dependent Landscape Formalisms	25
3	Energy Landscape Extraction for RNA Pseudoknot ScYLV	31
3.1	RNA System	31
3.2	Experimental Setup	35
3.3	Results and Analysis	36
3.3.1	WLC Fits	36
3.3.2	Model-Dependent Analysis	39
3.3.3	Full Landscape Reconstruction	41
3.3.4	Estimates of Free-Energy Difference between Folded and Unfolded States	43
3.3.5	Comparison of Results	46
4	Method of Inverse Weierstrass Transform	53
4.1	Theory	54
4.2	Data Collection	59
4.3	Data Analysis and Results	60
4.3.1	DNA Hairpin	60

4.3.2 RNA Pseudoknot	65
--------------------------------	----

5 Future Work	68
----------------------	-----------

List of Tables

3.1	Results from Model-Dependent Fits to Pseudoknot Data	41
-----	----------------------------------------------------------------	----

List of Figures

1.1	Sample Energy Landscape	5
1.2	Different hypothetical landscape types	6
2.1	SMFS Methods	11
2.2	Landscape Projection	15
2.3	Equilibrium Measurements	17
2.4	Optical Tweezers Setup Illustrating the WHAM Method	21
2.5	FECs	22
2.6	Landscape Parameters	26
2.7	Force Rupture Distribution	27
3.1	Pseudoknot Sequence	33
3.2	RNA Pseudoknot	34
3.3	Experimental Set-up: Pseudoknot Measurements	35
3.4	WLC Fits	38
3.5	Force Rupture Distribution	40
3.6	Lifetime Distribution	41
3.7	Integrated Probability Plot	42
3.8	Pseudoknot Free-Energy Landscape at Zero Force	44
3.9	Tilted Free-Energy Landscapes	44
3.10	Average Free-Energy Landscape for Pseudoknot	45
3.11	WLC Integrals	47
3.12	Dissipated Work and Variance	50

3.13	Hysteresis in Pseudoknot Re-folding	52
4.1	Laser Tweezers Schematic	54
4.2	Landscape Extraction Flowcharts	55
4.3	FECs for Hairpin A	60
4.4	Landscape Comparison for Hairpin A, Various Methods	61
4.5	Landscape and laser trap curvatures for Hairpin A	63
4.6	Landscape and laser trap curvatures for Hairpin A, sharper feature	64
4.7	Landscape Comparison for RNA pseudoknot ScYLVWT, Various Methods	65
4.8	Landscape and laser trap curvatures for RNA pseudoknot ScYLVWT	66

List of Abbreviations

List of commonly used abbreviations

–1 PRF	Programmed –1 Ribosomal Frameshifting
AOD	Acousto-Optic Deflector
EOD	Electro-Optic Deflector
FEC	Force-Extension Curve
LOT	Laser Optical Tweezer
PMF	Potential of Mean Force
PSF	Point Spread Function
ScYLV	Sugar Cane Yellow Leaf Virus
SMFS	Single-Molecule Force Spectroscopy
WHAM	Weighted Histogram Analysis Method
WLC	Worm-Like Chain

Chapter 1

Introduction

Of all the specialized abilities evolution has cultivated in living organisms, biomolecular self-assembly is perhaps the most crucial and remarkable. Proteins and nucleic acids, molecules that carry out the host of functions that sustain life—from digesting food to generating nerve impulses to preserving and packaging genetic material for new generations—exist in a wide array of specialized three-dimensional structures which dictate their specific functions [1]. Biomolecules are synthesized as polymer chains—comprised of nucleotides and amino acids in the case of nucleic acids and proteins, respectively—that subsequently fold, or self-assemble, into three-dimensional structures. Problems surrounding understanding biological folding have plagued scientists for over forty years [2], and the possible applications of a more comprehensive understanding of biological folding are extensive, ranging from the design of novel proteins to accomplish specific tasks [3] to developing treatments for human diseases caused by malfunctions in protein folding; such diseases include Alzheimer’s and Parkinson’s disease [4] and Type II Diabetes [2]. Folding mechanisms thus continue to be an active area of research, and *Science* magazine listed prediction of how proteins fold as one of 100 major scientific questions in 2005 [5].

Only twenty amino acids and five varieties of nucleotide bases exist in nature, and all the information about the global structure of a biopolymer is contained in the particular sequence of its constituent monomers. The fact that proteins require

no additional, external biological mechanisms to fold, but rather contain all necessary information in their one-dimensional amino acid sequences, was demonstrated definitively by Anfinsen in 1961 [6]. The decades since have seen scientists across disciplines—physicists, chemists, biologists, and mathematicians—struggling to understand *how* folding is encoded.

Anfinsen’s seminal work not only demonstrated that protein structure is encoded by monomer sequence; it also established that protein folding is usually a reversible process, confirming the so-called “thermodynamic hypothesis” [7], which states that a protein folds to a native conformational state that minimizes the system’s free-energy under “proper” conditions [8]. The free-energy, not the internal energy, is the relevant quantity in protein folding, as folding is driven by the interplay between entropy, which favours the denatured, disordered state, and enthalpy, which favours the folded state. In light of the astronomical number of possible configurations of the residues comprising a protein, the method by which a protein reliably locates its native conformation on the timescales typical of protein folding, milliseconds to minutes [9], puzzled scientists in the wake of Anfinsen’s discovery. Noting that a random search through configuration space would lead to folding times in excess of the age of the universe [10], Cyrus Levinthal postulated the existence of specific “folding pathways” that would guide proteins through a series of well-defined intermediate states as they progressed to their native conformation [11]. Each intermediate state along the pathway represented a particular arrangement of the residues, or monomer subunits, of a protein in three-dimensional space. Levinthal’s argument created a “paradox”: on one hand, Anfinsen’s thermodynamic hypothesis implies that regardless of initial conformation, proteins find a unique native state corresponding to a global free-energy minimum; on the other hand, the existence of kinetic pathways facilitating fast folding would imply that the native state is not necessarily a global free-energy minimum, but rather the endpoint of a well-defined sequence of conformations along a pathway [12, 13].

Studies of protein folding subsequent to Levinthal’s proposal couched experiments

in the language of folding pathways and attempted to chart the various intermediate stages of protein folding [13]. Scientists hoped that characterizing intermediates would yield insights into how nature's conformational search algorithms function [2], so observed reaction rates were fit to various phenomenological schemes involving both "off-pathway" and "on-pathway" intermediate states [14, 15]. The folding discourse quickly grew confusing, as Levinthal's "paradox" remained unresolved and the question of how protein folding could be both pathway-independent—as implied by Anfinsen's thermodynamic hypothesis—and pathway-dependent—as necessitated by biologically relevant folding timescales—remained unexplained [13]. In the 1980s, a simplifying picture emerged that today forms the basic framework for scientific understanding of biological folding. Bryngelson and Wolynes pioneered the conceptual energy landscape formalism through several papers that cast the protein folding problem into the language of statistical mechanics [16, 17, 18].

In the statistical mechanical viewpoint, protein folding is not interpreted as a sequential movement through a series of well-defined conformations along a pathway; instead, it is seen as a diffusive process across an *energy landscape*. Essentially, an energy landscape is a hyper-dimensional surface representing the free-energy of a protein as a function of the values of its conformational degrees of freedom, such as the dihedral angles describing the orientation of each amino acid. If such a surface has a single minimum but is otherwise flat, as in Levinthal's picture, finding the native state via a random search—that is, via diffusive Brownian motion driving random conformational shifts—would not occur on the timescales observed for protein folding. This "paradox" disappears, however, when one introduces an energy bias favouring conformations nearer to the native state; Zwanzig demonstrated that such a bias would vastly accelerate the search and lead to adoption of the native conformation on biologically-relevant timescales [10]. If moves that produce a conformation closer to the native state are energetically favourable, they will preferentially be accepted. The energy landscape, then, is shaped like a funnel that guides the protein to its native conformation. Instead of a single, definite path from unfolded state to folded state,

there exist on the energy landscape a multiplicity of paths; the folding process is redefined in terms of ensembles: an ensemble of denatured states, an ensemble of transition states, and an ensemble of intermediate states [13].

The energy landscapes of proteins are not expected to be strictly smooth funnels: the interactions that drive protein folding—hydrogen bonding, the tendency of hydrophobic residues to cluster to reduce exposure to solvent, and electrostatic interactions, to name a few—are also apt to compete with one another and induce substructures that are inconsistent with the native fold [12]. In other words, a protein trying to fold is a *frustrated* system, and this frustration manifests itself in many local minima strewn throughout a protein’s energy landscape. Indeed, a random sequence of amino acids will not fold reliably to a native structure, in general [19]; the majority of naturally-occurring proteins have been selected by evolution to be *minimally* frustrated, and this “principle of minimal frustration” is what lends an overall funnel shape to the energy landscape. The complete picture of a landscape, therefore, is a multi-dimensional funnel riddled with local minima across which a protein diffuses. The “bumpiness” of the energy landscape directly affects the diffusive motion on the landscape, and thus, reaction rates [12]. Figure 1.1 illustrates a hypothetical energy landscape for a naturally-evolved protein, and Figure 1.2 presents a comparison between a number of the different types of landscapes alluded to above.

The rich new language of statistical mechanics and the energy landscape formalism eliminated Levinthal’s paradox and provided clarification of many experimental protein-folding results [13]; it can be used to explore aspects of the folding reaction, including transition states, folding rates, effects of sequence mutation on the folding process, protein phase diagrams, and the “misfolding” that results when proteins become trapped in non-native free-energy minima. Indeed, the energy landscape effectively contains *all* the useful information about the protein folding process [20]. In order to fully realize the potential of the formalism, however, more quantitative development is needed. At present, most of the applications of the theory are qualita-

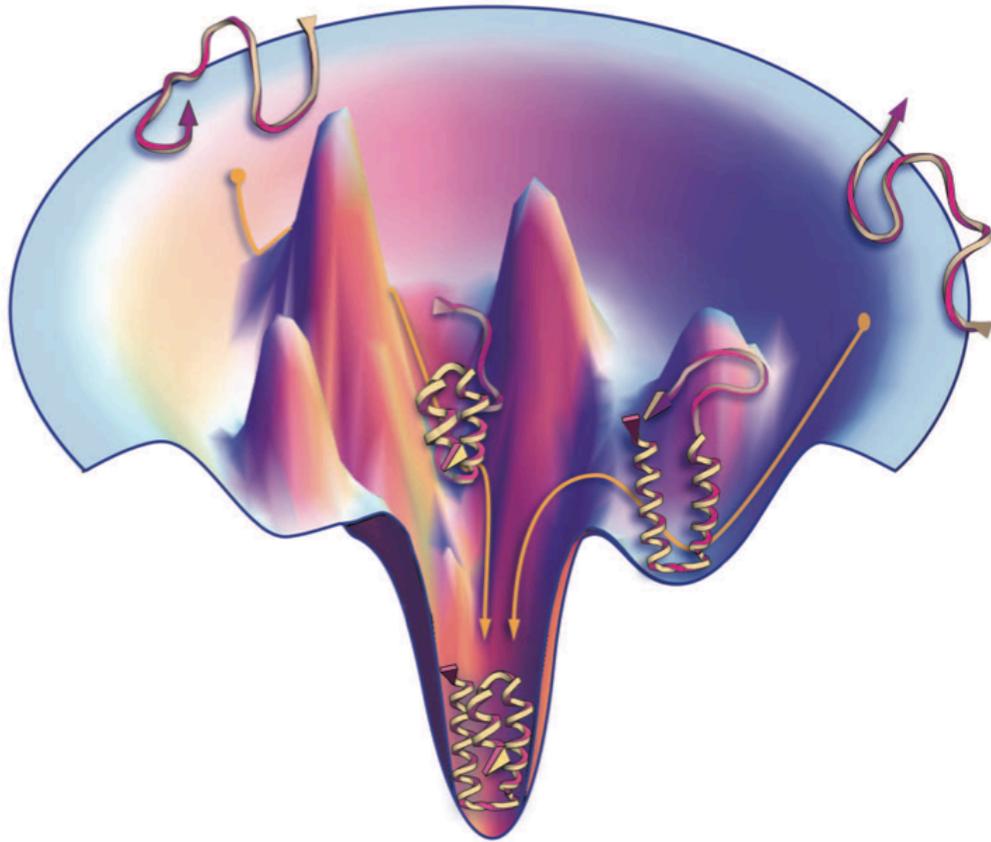


Figure 1.1: Hypothetical energy landscape for a protein. The vertical coordinate is free-energy, and the two horizontal coordinates are measures of protein conformation (in reality there are far more than two such coordinates). The rim of the funnel represents the ensemble of unfolded, or denatured, states of the protein, wherein it is an unstructured chain of amino acids. As the protein diffuses across the landscape and down the slope of the funnel, it acquires a structure increasingly closer to its native conformation and increasingly lower in free-energy. The native state of the protein is represented by the bottom of the funnel; it is a global minimum in the free-energy. The peaks and valleys riddled throughout the landscape are local free-energy minima and maxima, and represent barriers and metastable intermediate states for the protein, respectively. Figure taken from Dill and MacCallum, 2012 [2].

tive and non-comprehensive; more experimental data is needed before the theory has substantial predictive power [2]. Specifically, the principles underlying the “folding mechanism,” or way in which amino acid sequence shapes the landscape and propels the folding reaction, must be elucidated before quantitative predictions can be derived from the landscape framework [2]. To shed light on these principles, landscapes for

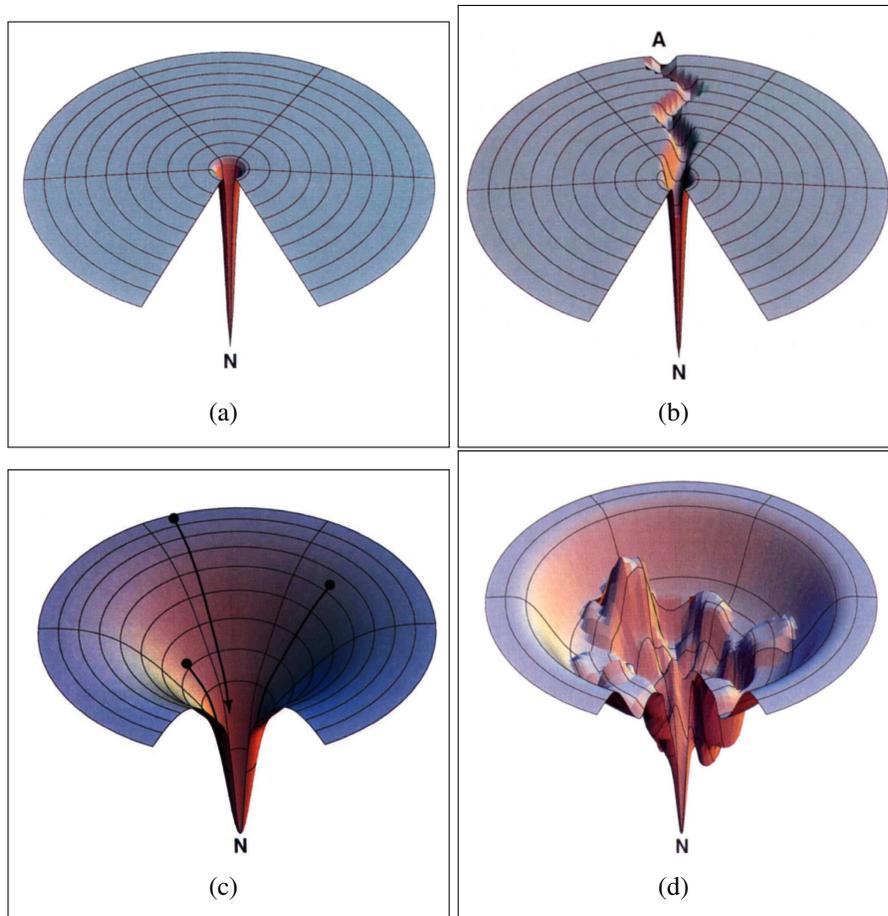


Figure 1.2: Illustration of the various types of landscapes discussed in the text; all images are hypothetical landscapes. Part (a) displays what the energy landscape would look like if the search for the native conformation were random and unbiased. Part (b) is a depiction of Levinthal’s “pathway” idea. Part (c) shows a perfectly smooth funnel; this would correspond to a totally unfrustrated system. Part (d) is representative of what a landscape for a naturally-evolved protein looks like: it is a minimally-frustrated system, so the landscape is “bumpy” and contains many local minima. Figure adapted from Dill and Chan, 1997. [13].

real proteins need to be measured. What is ideally required is a picosecond-scale measurement of the structures acquired by a protein as it traverses the entire pathway from the unfolded to the folded state. Unfortunately, the main experimental challenge in monitoring protein-folding reactions is achieving this very fine time resolution in structural measurements [21], and no current experimental techniques allow for this type of comprehensive measurement to be performed. Instead, complete landscapes

must be developed through computational and theoretical methods; the predictions of these methods can then be constrained by experiments.

Protein folding has been computationally explored since the 1960s [22]. Strategies vary, ranging from simplified lattice models [23] and Monte Carlo energy-minimization techniques that lack atomic detail, including the Gō model [24], to full atomistic molecular dynamics (MD) simulations [25]. The latter have improved vastly since their first implementation. Advances in computational power now allow folding simulations that are stable on millisecond timescales [2], and distributed-grid computing projects, which make use of the idle computer power of hundreds of thousands of users, have allowed for increased sampling of folding events [26]. Further, computational modelling of the interactions involved in folding, namely the force field surrounding the protein-solvent complex, has improved [26]. These advancements have enabled the successful prediction of stable states for a few small proteins [27, 28]. The theoretical energy landscape framework has developed in tandem with computational advancements; landscapes can be directly constructed from knowledge of entropic costs of bringing different regions of the protein together alongside the energies associated with the formation of native configurations [29]. Features on predicted landscapes such as the transition state ensemble (a collection of states that act as a folding bottleneck), and the predicted kinetics associated with diffusion on the landscape can be compared with experiments [30]. Current landscape theory is founded upon postulating funnelled landscapes; it has yet to explain proteins that violate the principle of minimal frustration—whose landscapes are not funnelled, but involve many minima—and still cannot describe exactly how funnelled landscapes are shaped by sequence [29]. Between computational and theoretical folding models there is much that is testable, however, particularly for the folding of small proteins [29].

Experimental techniques for studying protein folding provide crucial constraints on Theoretical and Computational studies. The main experimental challenge in monitoring protein-folding reactions is achieving very fine time resolution in struc-

tural measurements, since folding occurs on such short timescales [21]. The general method is to induce the folding or re-folding reaction and then attempt to chart its structural progress. To accomplish the former goal, chemical denaturants, sudden temperature changes, or mechanical force can be used [21, 31]. Probes of time-resolved structure include nuclear magnetic resonance, small angle X-ray scattering, fluorescence studies, and hydrogen exchange experiments, to name a few [21, 32]; however, such methods enable measurement of ensemble averages only and do not yield much information about the landscape itself. More detailed landscape data is required, as scientists still do not have a good understanding of the folding mechanism for many proteins [2]. The most promising candidates to date for providing measurements that approach the ideal mentioned above, the complete path of a protein on a landscape as it folds, are single molecule (SM) techniques. SM studies provide excellent resolution of the folding process [33] and provide insight into the rare trajectories, transient states, alternate conformations, and mis-folded minima that are obscured in bulk measurements. The advent of single-molecule folding studies thus brings much promise for further experimental elucidation of the *full* energy landscapes of folding proteins, necessary for comparison to, and refinement of, theory.

Monitoring the fluorescence emission from single molecules as they fold is one such single-molecule technique. Proteins are labelled with two sites, one that “donates” fluorescence to an “acceptor” site; because transfer efficiency depends on protein conformation, structural changes during folding can be monitored. This technique is known as Förster Resonance Energy Transfer (FRET), and it has proven a highly successful means to study single-molecule folding kinetics [34, 35]. Even more effective from a landscape study perspective, however, are single molecule force spectroscopy (SMFS) experiments [34, 36, 37], effected by laser optical tweezers (LOTs), magnetic tweezers, or atomic force microscopes (AFMs) [38]. In these experiments, a single molecule is tethered and manipulated; the forces applied act as a mechanical denaturant, inducing folding and re-folding. Recent theoretical advances have enabled the extraction of a full energy landscape profile—along a

single coordinate—from these experiments [39, 37, 40, 41, 42]. SMFS experiments thus constitute an exciting frontier in science of biopolymers, one that promises to yield a wealth of new data to complement developing landscape theory and computational investigation, as well as the first truly quantitative tests and applications of landscape theory.

This work aims to explore, apply, and refine techniques for extraction of energy landscapes from single molecule force spectroscopy experiments. Chapter 2 begins with a detailed treatment of how SMFS experiments are conducted and the ways in which they can recover energy landscapes, and kinetics on those landscapes, for biological systems. While originally developed to study protein folding, the theoretical framework and experimental tools aforementioned are applicable to other biological molecules that attain three-dimensional folds from one-dimensional sequences of monomers, such as nucleic acids: DNA and RNA [43]. Nucleic acid systems are much simpler than proteins, and thus provide a platform to test out landscape analysis ideas; this work focuses for that reason on nucleic acids rather than proteins. Chapter 3 explores the folding of a RNA pseudoknot that figures in viral propagation. The full energy landscape for this system, never before computed, is presented, alongside refinements for folding kinetics results presented by other authors [44]. Chapter 4 explores a novel theoretical approach to reconstructing energy landscape profiles from SMFS experiments, presenting the first experimental confirmation of its viability. The work closes with a section briefly examining the future steps that should be taken toward the advancement of the research presented herein.

Chapter 2

Methods for Energy Landscape

Extraction from Force Spectroscopy

Experiments

As described in the previous chapter, SMFS experiments offer an exciting glimpse into the energy landscapes of biomolecules. Such experiments afford control of which region of the energy landscape is explored; the mechanical control can cause a single folding pathway to be traversed repeatedly and studied [36]. Further, they allow for observation of molecular states that are obscured by ensemble measurements, such as rare alternate conformations and intermediates, and for characterization of transition states [36]. Mechanical denaturation also offers improved resolution for measurements of molecular motion, since applied force usually increases molecular stiffness and this in turn decreases positional fluctuations, as encapsulated by the equipartition theorem [33]. There are multiple experimental realizations of SMFS, compared in Figure 2.1 [45, 33].

Magnetic tweezers feature a superparamagnetic bead suspended in the field of an external magnet [45, 33]. The molecule of interest is attached at one end to the bead and at the other end to a surface; both connections are made via a “handle”—comprised, for instance, of DNA. The bead experiences a force proportional to the

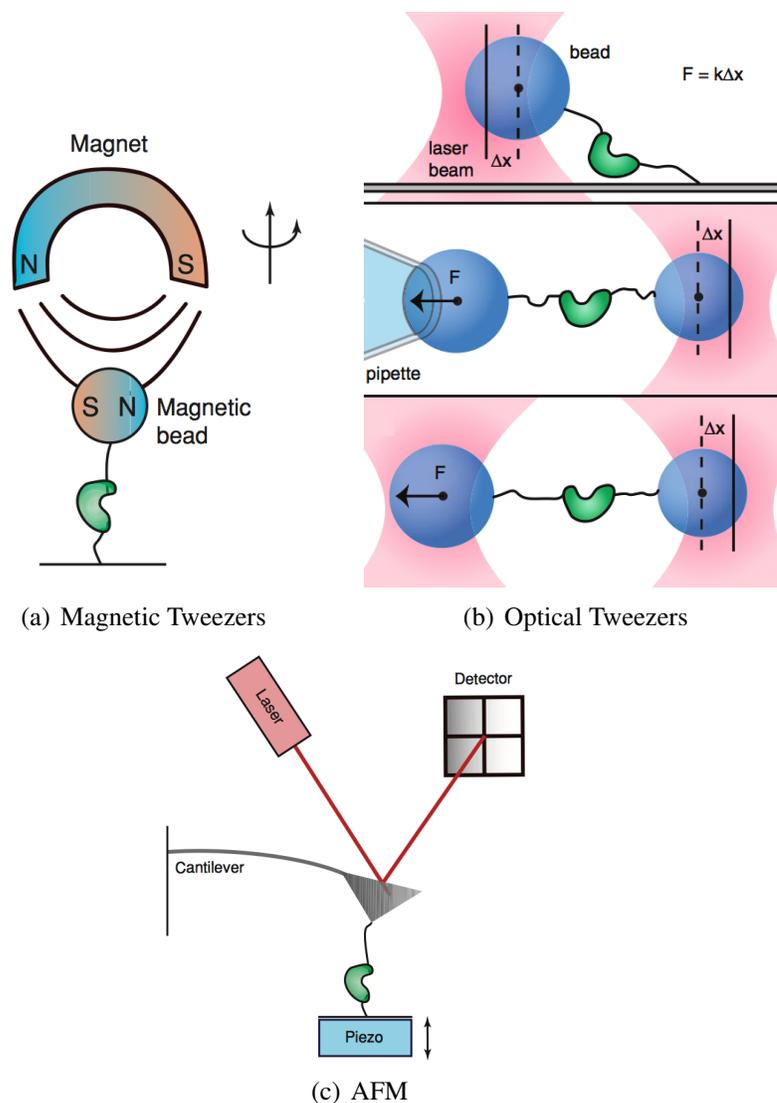


Figure 2.1: Illustration of three experimental realizations of SMFS. In all cases, the molecule of interest is normally attached to molecular linker handles (a) Magnetic tweezers feature a paramagnetic bead suspended in an external magnetic field, which can be manipulated to apply linear forces and torsion. Figure adapted from Greenleaf *et al.* 2007 [33]. (b) Optical tweezers feature one or more focused laser beams, which exert harmonic restoring forces on dielectric beads. The bead can be suspended between laser and a surface (top), laser and a micropipette tip (middle), or two lasers (bottom). Figure adapted from Woodside and Valentine 2009 [46]. (c) AFM experiment. The molecule of interest is attached at one end to a flexible cantilever which can exert harmonic restoring forces. The other end of the molecule is attached to a surface; in this figure, the surface is a piezoelectric moveable stage. The molecular extension is measured by deflecting a laser from the cantilever tip to a PSD. Figure adapted from Greenleaf *et al.* 2007 [33].

gradient of the magnetic field; this is approximately constant for magnetic tweezers, and they are thus used to perform *constant* force experiments. Because a dipole moment is induced in the bead by the external field, it is possible to apply torques to molecules using magnetic tweezers [45, 33]. Atomic Force Microscopes (AFM) can also be used to perform SMFS experiments. The biological molecule of interest is again tethered at both ends by handles, often short DNA or peptide chains, but in this case one end is affixed to a surface and the other to the AFM tip on a cantilever, which behaves as a spring [45, 33]. By moving the cantilever or the surface, the amount of force applied to the molecule can be adjusted. The applied force can be kept constant through the action of a feedback loop, constituting *constant-force* experiments, or it can be changed, as in *force-ramp* experiments [45, 33].

A third apparatus used for SMFS experiments is that of Laser Optical Tweezers (LOT), relevant to the current work. LOT employ one or more highly focused laser beams, in which dielectric beads are suspended. An electric dipole moment is induced in the beads by the electromagnetic field of the laser, and since an electric dipole experiences a force in a non-uniform electric field, the bead consequently experiences a force proportional to the intensity gradient of the laser light, *i.e.*,

$$F = \alpha \nabla I, \tag{2.1}$$

where α is a constant that depends on properties like the bead size and index of refraction and the wavelength of light used [47]. This force will tend to maintain the bead in an equilibrium position at the transverse focus of the laser beam, but slightly axially displaced from the focus due to scattering forces exerted by the light [47]. For small enough displacements of the bead from equilibrium, the laser traps behave as Hookean springs [47]. Several LOT setups are possible; in each, one end of the molecule is tethered via a DNA handle to a bead suspended in a laser beam. The other end can be tethered to a surface, to a micropipette, or to a bead in a second laser beam. The dual-beam setup, while the most complicated to implement, is advantageous

in that it avoids the mechanical noise generated by the coupling of the laser to a surface or a micropipette. The data included in the present work were collected on a dual-beam setup which has been described previously [44]. In this configuration, applied forces are manipulated either by adjusting the intensity of the laser light or by moving the laser traps and thus altering the displacement of the beads from their equilibrium position. Molecules can be held under constant tension via force clamp feedback loops or through a passive, all-optical clamping mechanism like that developed by Greenleaf *et al.* [48], which eliminates the time delay associated with feedback loops. Force-ramp experiments can also be performed: by moving the laser traps apart at a constant velocity, a steadily increasing force is applied to the molecule. In yet a third scenario, forces can be increased and decreased suddenly, or “jumped”; the molecule is allowed to equilibrate at one force, then the tension is discontinuously changed to induce denaturation or renaturation [42]. During LOT SMFS experiments, the separation between beads, which is the molecular extension plus handle length, is measured, as is the force exerted on the molecule. One method to achieve this is by deflecting laser light from the dielectric beads and using a position-sensitive diode (PSD) to record bead position as the experiment progresses [47].

Any attempts to study the multidimensional folding energy landscape using SMFS experiments are necessarily limited by the fact that the many degrees of freedom on which the folding landscape depends are not measurable in experiments. Consequently, in most single-molecule experimental studies, one-dimensional (1D) models are used in which the full energy landscape is projected onto one reaction coordinate, such as the molecular extension [49] (see Figure 2.2). Because this projection amounts to averaging over many other degrees of freedom, the 1D energy landscape profile is a potential of mean force (PMF) [50]. While landscape reconstructions from SMFS experiments assume folding dynamics can be captured by studying diffusion on the 1D PMF [37, 41, 51], it is not clear that a 1D projection along the reaction coordinate accurately recaptures the full folding landscape [49]. Indeed, some authors have already identified potential pitfalls of assuming 1D projections are reliable

representations of the full landscape, noting that the folding barrier can be obscured in 1D projections [49, 52]. The 1D landscape projection has successfully reproduced observed folding dynamics for a prion protein system, however [20], suggesting the validity of the approach in some cases. There is by no means a consensus on this topic, and choice of reaction coordinate and the effects of simplifying the multidimensional free-energy landscape are active areas of research. In this and subsequent chapters, this caveat must be kept in mind.

This chapter will explore techniques for extracting information about energy landscapes from SMFS experiments; it will couch these techniques in the language of dual-trap LOT studies, as these are of relevance for the current work. There are essentially three angles of approach when studying landscapes through LOT SMFS experiments. First, free-energy profiles can be reconstructed from equilibrium measurements; these include constant-force experiments, in which molecular extension at constant tension is monitored, and constant position measurements, in which a constant separation between the laser traps is maintained and extension fluctuations are measured. Second, free-energy profiles can be reconstructed from non-equilibrium measurements; these include force-ramp experiments, in which the laser traps are moved at a constant velocity to either increase or decrease applied tension, and force jump experiments, in which the force is discontinuously ramped up or down. Finally, landscape parameters can be extracted from SMFS experiments by assuming particular models for the shape of the free-energy landscape. The following sections examine these techniques in greater detail.

2.1 Landscape Reconstructions from Equilibrium Force Measurements

As explained above, equilibrium SMFS measurements can be carried out in two different ways: first, by maintaining a constant force and measuring the molecular extension as a function of time, and second, by maintaining a constant separation

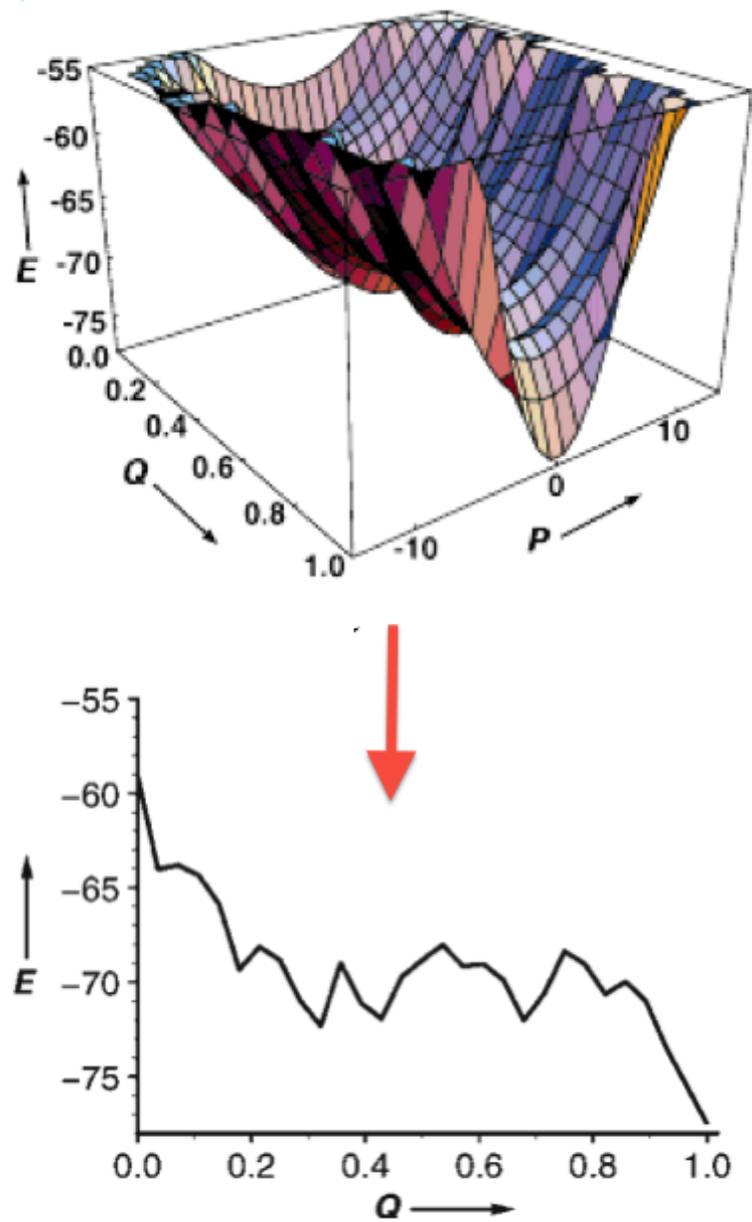


Figure 2.2: In SMFS experiments, wherein a single reaction coordinate is normally measured, the full multi-dimensional energy landscape is projected onto one axis, becoming a PMF; this is illustrated by the red arrow. Here, E represents free-energy, and Q and P are arbitrary degrees of freedom describing molecular conformation. The units are arbitrary. Figure adapted from Dobson *et al.* 1998 [21].

between laser traps and monitoring fluctuations in force. The former technique lends itself to straightforward extraction of landscape profiles along the reaction coordinate x , which is molecular extension, provided that the constant tension is maintained through passive clamping [48] rather than through a feedback loop. This is because active feedback loops have a finite inherent response time, so the force is not truly kept constant and the extension changes do not all occur at the same force value. To extract landscapes from passively-clamped constant-force data, extension-versus-time traces are collected and binned to compile probability histograms, $P(x)$. Because the molecules are in equilibrium, the extension distributions are expected to be Boltzmannian, and the free-energy $G(x)$ can be constructed by inverting:

$$P(x) = \exp\left(-\frac{G(x)}{k_B T}\right) \quad (2.2)$$

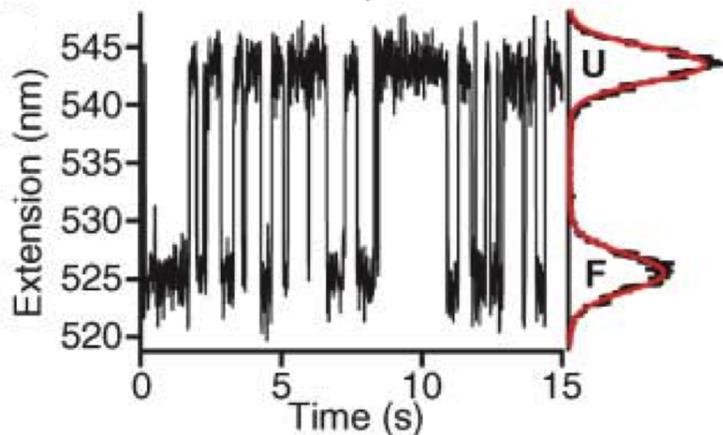
where k_B is the thermal energy. The landscape profile is thus given by

$$G(x) = -k_B T \ln P(x). \quad (2.3)$$

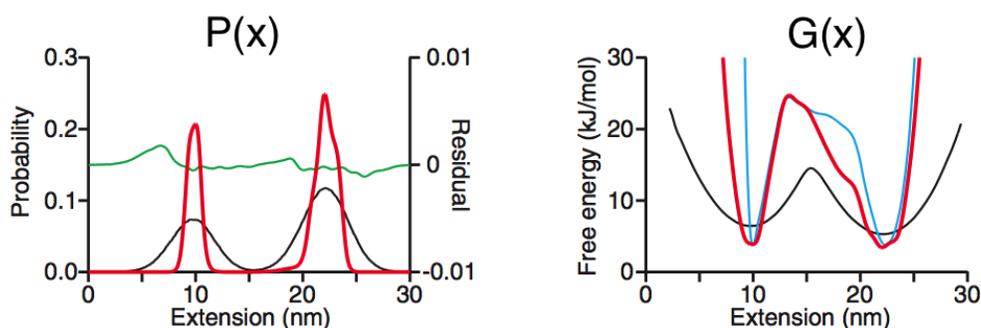
This method was successfully implemented by Woodside *et al.* on DNA hairpins measured with LOT [37]; Figure 2.3 summarizes the procedure. In 2.3(a) the extension-time traces are shown. The reconstructed landscape profile $G(x)$, along with the $P(x)$ histogram, is shown in 2.3(b).

A similar procedure can be applied to equilibrium measurements taken at constant probe position (fixed separation of the laser traps); Gebhardt *et al.* demonstrated this successfully for a protein system [40]. In this case, when the molecule undergoes the folding transition and its extension changes, the force on the molecule will vary in accordance with the changing displacement of the bead from its equilibrium position. Position histograms can still be constructed, but must be corrected for these changing forces before landscapes can be extracted [40].

One of the major limitations of any landscape reconstruction attempted using LOT experiments is instrumental distortion. The DNA handles are a source of additional



(a) Extension-Time Trace



(b) Probability Histogram and Reconstructed Landscape

Figure 2.3: Illustration of the landscape reconstruction method from constant-force data. (a) Molecular extension is monitored as the molecule is held at constant tension, resulting in an extension-time trace. The oscillation visible in the trace corresponds to the molecule unfolding and refolding. To the right of the trace is the experimental $P(x)$ histogram; U refers to the unfolded state, at larger extension, and F refers to the folded state. (b) The free-energy profile $G(x)$ is constructed from the experimental $P(x)$ histogram according to Equation (2.2). In black are the raw results; red represents results after deconvolution. Green is the residual from deconvolution, and blue represents a theoretical model to which the authors were comparing their landscape. Figure adapted from Woodside *et al.* 2006 [37].

compliance, as is the finite stiffness of the trapping lasers. Rotations of the beads in the traps can further complicate the matter. In effect, the instrument response introduces a point-spread function (PSF) that is convolved with the intrinsic response of the molecule. If the intrinsic molecular extension distribution is $p(x)$ and the instrumental PSF is $S(x)$, the measured extension distribution is $P(x) = (S * p)(x)$.

To overcome this issue, some form of deconvolution is required. Woodside *et al.* [37] implemented the non-linear deconvolution algorithm of Jansson [53], obtaining the instrumental PSF empirically by repeating measurement protocols with a molecule that remained folded in the relevant force range. Gebhardt *et al.* [40] employed a more complicated position-dependent empirical PSF to account for the changing force in their constant probe measurements. Another approach is to derive a theoretical model for the instrumental PSF; this was successfully implemented using experimental data by Hinczewski *et al.* [51]. There is a limit, however, to the landscape resolution obtainable where deconvolution is necessary [53]; to combat this, instrumental effects can also be mitigated by introducing shorter [54] or stiffer DNA handles, such as DNA origami beams [55].

2.2 Landscape Reconstructions from Non-Equilibrium Force Measurements

For many systems, equilibrium landscape reconstructions are unfeasible due to very slow folding dynamics, caused by a large barrier or slow diffusion or both. Complex RNA systems, such as the one examined in this work, provide an example of slow folding that is impractical to measure using equilibrium techniques [56]. An alternative, non-equilibrium approach must be employed for such systems: *viz.* force-ramp or force-jump experiments. Such experiments must be interpreted through the lens of non-equilibrium statistical mechanics.

Two decades ago, a theoretical framework relating measurements arbitrarily far from equilibrium to equilibrium thermodynamic quantities did not exist. In 1997, Jarzynski changed this with a revolutionary equality [57]. Previously, it was known that the work dissipated in a non-equilibrium process obeys

$$\overline{W} \geq \Delta G, \tag{2.4}$$

where ΔG is the free-energy difference between two states of a system and \overline{W} denotes the ensemble average of work required to effect the change in system state. The equality holds only for reversible processes. Jarzynski demonstrated that:

$$\overline{e^{-\beta W(z)}} = e^{-\beta \Delta G(z)}, \quad (2.5)$$

where the overbar indicates an average taken over the ensemble of all possible measurements of the work done in altering the state of the system by continuously changing the value of some control parameter, z . Here, $\beta = 1/k_{\text{B}}T$ is the reciprocal of the thermal energy. Essentially, this equality demonstrates that, while on average, the work done on a system over the course of a non-equilibrium trajectory will exceed the equilibrium work in accordance with Equation (2.4), occasionally, one “gets out more than one puts in” in the sense that the non-equilibrium work is *less* than the equilibrium value [58]! The exponential average of Equation (2.5) exactly balances these low-work trajectories with the excess work trajectories to recover the equilibrium free-energy difference. This remarkable result was verified by Liphardt *et al.* [59] for single-molecule pulling experiments carried out on an RNA molecule.

In its above form, Jarzynski’s equality cannot be used to extract intrinsic free-energy profiles from LOT experiments. The reason for this is that the free-energy difference given in the identity (2.5) is between two different values of an external control parameter, z , *not* between two different values of some internal reaction coordinate for the system, such as the pulling coordinate that is measured in LOT experiments. The control parameter in LOT experiments is the time, or equivalently, the separation between laser traps, which is increased monotonically in force-ramp experiments to induce fluctuations in the pulling coordinate: molecular extension. Hummer and Szabo [39] have addressed this issue, providing a formalism for extraction of $G(q)$ —the free-energy profile along molecular extension q —which is based on the Jarzynski equality.

The idea behind the procedure is that the ensemble of non-equilibrium trajectories

of a system behaves the same as the equilibrium ensemble provided each state is weighted with the exponential of the external, non-equilibrium work required to establish it from some reference state [60]. In their approach, Hummer and Szabo [39] use a derivation of the Jarzynski equality based on the Feynman-Kac theorem [61] to obtain the following expression for free-energy as a function of pulling coordinate q :

$$\beta G_0(q) = -\ln \langle \delta(q - q_t) e^{-\beta \Delta W_t} \rangle, \quad (2.6)$$

where δ is the Dirac delta function, q is the molecular extension coordinate, q_t is its value at a particular time, t , and $\beta = 1/k_B T$. Note that we must subtract from external work done on the system the instantaneous bias potential of the laser traps: $\Delta W_t = \int_0^t \frac{\partial H}{\partial t'}(x_{t'}, t') dt' - k\xi^2/2$, where k is the spring constant for the system. For force-ramp LOT experiments, ξ is the displacement of the beads from equilibrium position; in the setup illustrated by Figure 2.4, where the laser traps are moved apart at constant velocity v , this displacement is $\xi = z(t) - q(t) = z(0) + vt - q(t)$, where $z(0)$ indicates the initial trap separation, and $z(t)$ and $q(t)$ are the time-dependent trap and bead separations, respectively. The angled brackets in Equation (2.6) indicate an ensemble average over all possible pulling trajectories. The Hamiltonian describing the LOT system is given by

$$H(x, t) = H_0(x) + k \left(z(0) + vt - q(t) \right)^2 / 2, \quad (2.7)$$

where $H_0(x)$ is the Hamiltonian of the system in the absence of applied tension and $(k/2)(z(0) + vt - q(t))^2$ is the energy stored in the displacement of the beads from their equilibrium positions (k again is the spring constant for the system). Given this Hamiltonian, ΔW_t becomes after evaluating the integral:

$$\Delta W_t = kv \left[\frac{vt^2}{2} + z(0)t - \int_0^t q(t') dt' \right] - \frac{k}{2} \left(z(0) + vt - q(t) \right)^2 \quad (2.8)$$

One can now write Equation (2.6) as:

$$\beta G_0(q) = -\ln \left\langle \delta(q - q_t) e^{-\beta \left[kv \left[\frac{vt^2}{2} + z(0)t - \int_0^t q(t') dt' \right] - \frac{k}{2} \left(z(0) + vt - q(t) \right)^2 \right]} \right\rangle \quad (2.9)$$

In principle, the free-energy profile along q can be computed directly from the

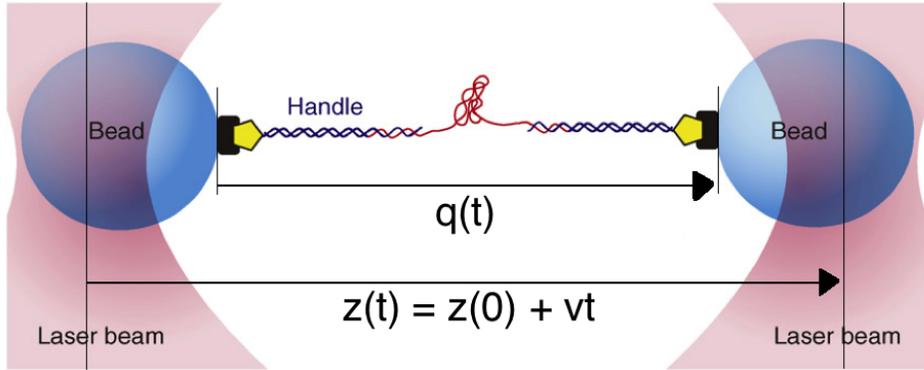


Figure 2.4: Setup for LOT measurements. When determining the external work done on the system, the displacement of the beads in the spring potential is considered. This displacement is $z(t) - q(t)$. Figure adapted from Woodside *et al.* 2008 [36].

ensemble average in Equation (2.9), taken at one particular moment in time t . In practice, however, given the finite number of trajectories sampled in an experiment, the full range of q values will not be sampled in one single time slice [39]. Indeed, most q values will cluster about the location of the equilibrium in spring potential, $z(t)$, at any given time t . To solve this problem, snapshots from *many* time slices must be combined; Hummer and Szabo [39, 62] adapted the Weighted Histogram Analysis Method (WHAM) of Ferrenberg and Swendsen [63] for this purpose. The situation is similar to that of umbrella sampling; an unbiased energy profile is derived by sampling many different biased profiles and combining these in a weighted average that minimizes statistical error [64]. In the end, the biases are removed from the weighted average. In the present case, each time slice acts as one umbrella sampling “trial” featuring a particular value of the time-dependent spring potential bias. When these trials are combined, the observations must be unbiased not only with respect to the spring potential, but also with respect to the non-equilibrium work [62]. This

results in the following expression for $G_0(q)$:

$$G_0(q) = \frac{\sum_t \frac{\langle \delta(q-q(t)) e^{-\beta W_t} \rangle}{\langle e^{-\beta W_t} \rangle}}{\sum_t \frac{e^{-\beta V(q,t)}}{\langle e^{-\beta W_t} \rangle}} \quad (2.10)$$

Here, the summations are over all time slices. $V(q,t)$ is the bias energy stored in the traps, and W_t is the non-equilibrium work. The angled brackets again denote ensemble averages. Since in real experiments a *finite* number of trajectories are available, the ensemble averages have to be approximated, necessitating further binning. A typical

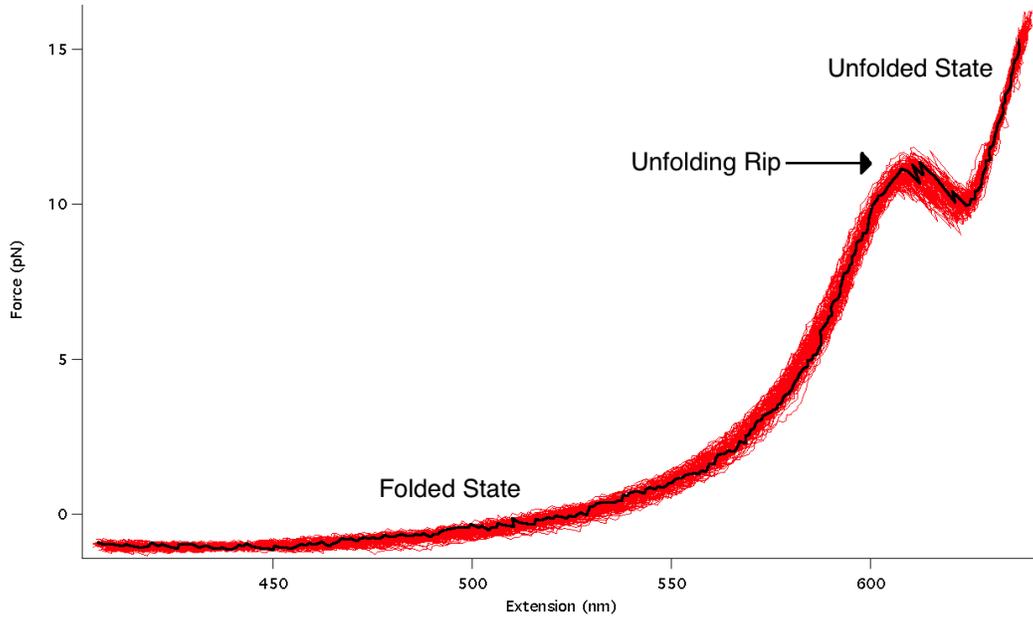


Figure 2.5: A typical data set from a LOT force-ramp experiment. The displacement of beads from equilibrium is measured and used to compute the tension and the total system extension. For folding experiments, FECs exhibit a characteristic sawtooth pattern; the ‘rip’ represents the sudden increase in extension and decrease in force consequent on molecular unfolding. These curves were collected from a DNA hairpin; see Chapter 4 for further details.

data set collected during LOT force-ramp experiments consists of Force-Extension Curves (FECs), depicted in Figure 2.5. As the laser traps are moved at a constant speed, the beads are deflected from their equilibrium position. By measuring the bead position, it is possible to derive values for system extension (molecule plus handles)

as well as force. To implement Equation (2.10), first, the non-equilibrium work for the i^{th} time bin and k^{th} pulling trajectory is computed via trapezoidal numerical integration:

$$W_{i,k} = kv \left[\frac{v_i^2}{2} - \sum_{j=1}^i (t_j - t_{j-1})(q_{j,k} + q_{j-1,k}) \right]. \quad (2.11)$$

Next, $\langle e^{-\beta W_i} \rangle$ is approximated for every time bin:

$$\langle e^{-\beta W_i} \rangle \approx \frac{1}{N} \sum_{k=1}^N e^{-\beta W_{i,k}}, \quad (2.12)$$

where N is the total number of pulling curves collected in the experiment and i indexes the relevant time bin. The top of the numerator in Equation (2.10) is then approximated for the l^{th} extension bin as follows:

$$\langle \delta(q_l - q(t_i)) e^{-\beta W_{i,k}} \rangle = \frac{1}{N} \sum_{k=1}^N e^{-\beta W_{i,k}} \Theta_l(q_{i,k}) \quad (2.13)$$

Here, N is again the total number of pulling trajectories collected. W_{ik} is the work corresponding to the i^{th} time bin and k^{th} pulling curve, l is an index referring to the extension bin, and Θ is a function that is 1 if $q_{i,k}$ falls into the l^{th} extension bin and 0 otherwise. The top denominator of Equation (2.10) is computed for each time bin using the *average* separation between laser traps at that time:

$$V(q, t) = \frac{k(\overline{z(t)} - q_l)^2}{2}, \quad (2.14)$$

where q_l is the midpoint of the l^{th} extension bin.

The WHAM method was validated on a system of DNA hairpins by Gupta *et al.* [41]. Free-energy profiles obtained via the above procedure were “tilted” to approximate their appearance at some constant force value, F ; this allows the landscape to be compared to one collected during a constant-force equilibrium experiment. The “tilting” is effected by subtracting a term representing the work done by the constant

force from each point on the profile:

$$G_F(q) = G_0(q) - Fq. \quad (2.15)$$

This tilted landscape was then compared to the results of an equilibrium landscape reconstruction performed for the same DNA system; the two landscapes agreed very well with one another [41].

The main difficulty with Jarzynski-based methods is that many trajectories are required to ensure adequate statistics. Equation (2.5) is exact only if the average is taken over the entire ensemble of possible trajectories, yet experiments can sample only a finite number of trajectories. The farther from equilibrium a measurement is taken, the greater the amount of data necessary to accurately reconstruct the free-energy [59]. This problem is exacerbated by the necessity to bin data for WHAM reconstructions; coarse binning is often required to produce adequate statistics for the work distribution in each bin, and this decreases the resolution of the landscape. Further, Jarzynski's equality is predicated upon all trajectories being drawn from an identical initial equilibrium distribution; instrumental drift over the course of many pulls may compromise the identity of initial conditions. Thus, the demand for many experimental trajectories likewise increases the instrumental stability requirements [59]. Often, it is not experimentally feasible to procure the volume of data required, and entire portions of the landscape remain unresolved, as illustrated by Harris *et al.* [65]. An alternative Jarzynski-based reconstruction method may somewhat alleviate these issues; Hummer and Szabo proposed a procedure based on a mathematical Weierstrass transform, which does not have the same stringent requirements on experimental statistics [66]. Chapter 4 provides a detailed treatment of this method.

The formalism outlined above is not applicable to the analysis of force-jump experiments, because Jarzynski's equality is valid only for processes wherein a control variable is *continuously* changed [57]. In force-jump experiments, the force is

discontinuously switched between two different values to induce folding or unfolding. Zhang *et al.* [67] developed a method for determining the free-energy profile by inverting the non-equilibrium stationary probability density from a set of force-jump trajectories. While this method has been used on experimental data [42], it has not yet been validated on a system for which the landscape is known. The force-jump method is not employed in the current work.

2.3 Model-Dependent Landscape Formalisms

Full reconstructions of the landscape profile are experimentally and analytically demanding. Alternative approaches aim to characterize a few parameters that dictate the shape of the landscape—such as barrier height and distance from the folded well to the barrier—by making certain assumptions. In these approaches, a form for the dependence of the folding reaction rate on force is first assumed. Next, distributions of observables implied by the assumed form are derived and fit to experimental data to extract landscape parameters, like the ones shown in Figure 2.6. The most basic such method is due to Evans and Ritchie [69], who expanded upon an expression for force-dependent rates originally derived by Bell [70] on the basis of earlier work done by Zhurkov [71]. In his treatment, Bell assumes that the only effect of an external force, such as that applied by an harmonic pulling spring, on a smooth, double-well energy landscape is the lowering of the barrier by an amount $F\Delta x^\ddagger$, where Δx^\ddagger is the distance from the well minimum to the barrier along the reaction coordinate x . Bell then inserts this effect on the barrier height into the phenomenological Arrhenius law to obtain the following force-dependent reaction rate:

$$k(F) = k_0 e^{\beta F \Delta x^\ddagger}, \quad (2.16)$$

where k_0 is the unperturbed reaction rate. Evans and Ritchie extend this treatment to derive expected force rupture distributions for SMFS unfolding experiments [69]. A

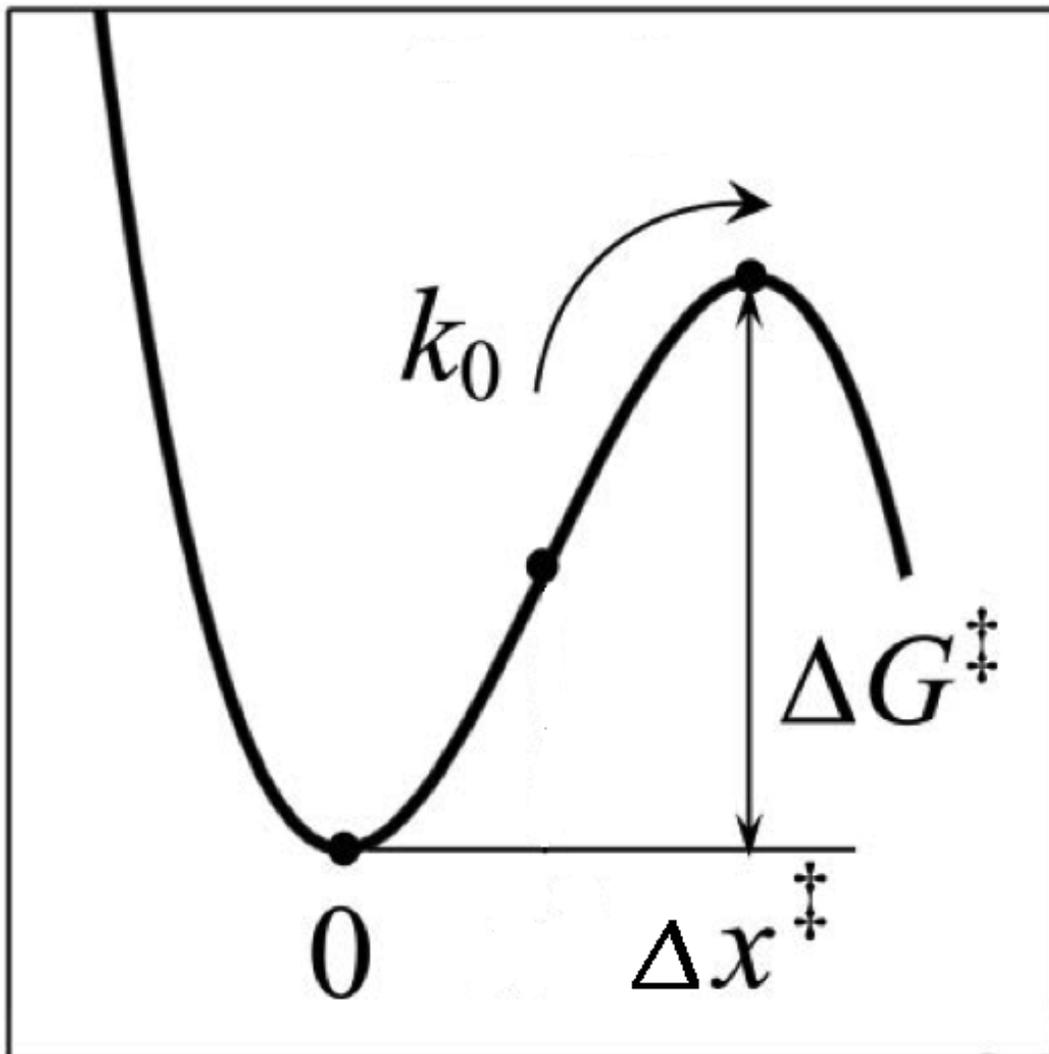


Figure 2.6: A hypothetical linear-cubic energy landscape. The ordinate is free-energy, and the abscissa is a reaction coordinate, like molecular extension. Several key parameters characterize the landscape: k_0 , the rate of escape from the ‘folded’ well; ΔG^\ddagger , the height of the barrier, and Δx^\ddagger , the distance to barrier. Figure adapted from Dudko *et al.* 2006 [68].

force rupture distribution is simply the collection of forces at which the characteristic “rip” in FECs occurs; this sudden change in extension and force represents the folding \Leftrightarrow unfolding transition. Because the process of escape over a barrier on the folding energy landscape is stochastic, the force at which this rip occurs varies from trajectory to trajectory; the total collection of rip forces from all trajectories in an experiment forms a distribution, like the one shown in Figure 2.7. While the Evans

and Ritchie approach has been widely and successfully applied to experiment [72], their assumption that the only effect of force is barrier height adjustment is invalid. Indeed, the larger the perturbing forces, the more significantly the distance between the folded well and the barrier, Δx^\ddagger , is altered [68].

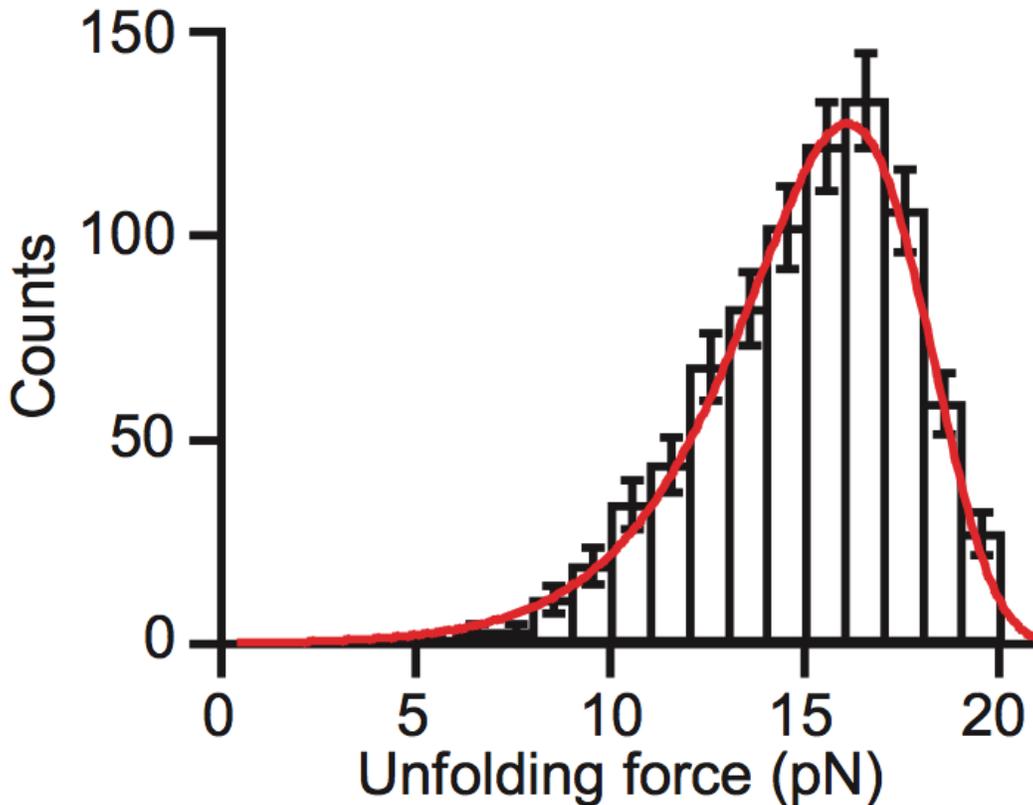


Figure 2.7: A sample rupture force distribution for 800 curves collected during a riboswitch aptamer experiment. The force at which the ‘rip’ in each FEC occurs is binned to compile a histogram. The red curve represents a fit using the theory of Dudko *et al.*; see Equation (2.19). Figure adapted from Greenleaf *et al.* 2008 [73].

Expanding upon the work of Garg [74], Dudko *et al.* [68] developed a framework that treats perturbations in barrier height *and* position by external forces. First, a shape for the landscape is assumed; Dudko *et al.* treat linear-cubic and cusp-like energy landscapes. Next, these profiles are substituted into Kramers’ expression for escape rate across an energy barrier [75] to yield the following force-dependent rate

expression:

$$k(F) = k_0 \left(1 - \frac{vF\Delta x^\ddagger}{\Delta G^\ddagger}\right)^{1/v-1} e^{\beta\Delta G^\ddagger} \left[1 - (1 - vF\Delta x^\ddagger/\Delta G^\ddagger)^{1/v}\right], \quad (2.17)$$

where k_0 is the escape rate in the absence of force; F is the applied force; Δx^\ddagger is the distance to the barrier on the unperturbed landscape; v parametrizes the shape of the landscape, and is 1/2 for a cusp-like landscape, 2/3 for a linear-cubic landscape, and 1 to recover the phenomenological Bell-Evans theory; and ΔG^\ddagger is the barrier height on the unperturbed landscape. Dudko *et al.* [68] derive a general expression for the distribution of rupture forces:

$$p(F) = \frac{k(F)}{\dot{F}} \exp\left(-\int_0^F [k(F')/\dot{F}]dF'\right), \quad (2.18)$$

where $k(F)$ is the force-dependent reaction rate and \dot{F} is the loading rate. In LOT experiments, the applied force is equal to $F(t) = kvt$, where k is the spring constant for the laser traps and v is the pulling speed, which implies $\dot{F} = kv^1$; in this context, and given the above expression for $k(F)$, the expected force rupture distribution is [68]:

$$p(F) = \frac{1}{kv} k(F) e^{k_0/\Delta x^\ddagger kv} e^{-[k_0/\Delta x^\ddagger kv] \left[1 - (vF\Delta x^\ddagger/\Delta G^\ddagger)\right]^{1-1/v}}. \quad (2.19)$$

This expression has been successfully fit to experimental data [73] and was actually verified against a full landscape reconstructed via the WHAM method by Yu *et al.* [20].

In order to fit Dudko *et al.*'s theory to force rupture distributions, it is necessary to bin the forces in extension due to limited experimental data (see Figure 2.7). An alternative approach is to *integrate* Dudko *et al.*'s expression for $p(F)$, Equation (2.19),

¹Actually, in the presence of the molecular linkers that normally tether the molecule of interest to the beads suspended in LOT experiments, this is not quite correct. Dudko *et al.* [76] provide an expression for determining the loading rate in the presence of molecular handles; however, it is experimentally unfeasible to implement a loading rate clamp, and in practice the average loading rate just before the rip in FECs is used.

and fit this integral to experimental data. Specifically, in a given experiment, the observed rupture forces can first be sorted in increasing order. The i^{th} point in this distribution is then assigned the value $\frac{i-0.5}{N}$, where N is the total number of points; this ensures a normalized distribution. When these assigned values are plotted against the corresponding force, the result is a sigmoidal curve, as shown in Figure 3.7 in Chapter 3. The integral of Equation (2.19) is given by:

$$\begin{aligned}
 P(F) &= \int_0^F p(F') dF' \\
 &= 1 - e^{\frac{1}{kv} \int_0^F k(F') dF'} \\
 &= 1 - e^{k_0/\Delta x^\ddagger kv} e^{-[k_0/\Delta x^\ddagger kv] [1 - (vF\Delta x^\ddagger/\Delta G^\ddagger)]^{1-1/v}}.
 \end{aligned} \tag{2.20}$$

Fitting the data to this form avoids the necessity of binning, which is an undesirable and somewhat arbitrary procedure, by assigning equal weight to each rupture force observed in the experiment. This particular application of Dudko *et al.*'s theory has not heretofore been implemented; the present work endeavours to remedy this.

Dudko *et al.* also demonstrated how the results of constant-force SMFS experiments can be related to the results of force-ramp experiments [76]. Equation (2.18) is first inverted to yield an expression for the expected lifetime of, for instance, the folded state if the molecule is held at some constant force F :

$$\tau(F) = \frac{1}{\dot{F}(F)p(F)} \int_F^\infty p(F') dF', \tag{2.21}$$

where $\tau(F)$ is the force dependent lifetime, $p(F)$ is the rupture force distribution, and \dot{F} is the loading rate. This relation can be used to transform data collected in force-ramp experiments into the lifetimes expected to be observed in constant-force experiments. Force-ramp data can then be plotted alongside lifetimes measured in constant-force experiments; both sets of data should lie along a master curve [76] fit

by the inversion of Equation (2.17):

$$\tau(F) = \tau_0 \left(1 - \frac{vF\Delta x^\ddagger}{\Delta G^\ddagger}\right)^{1-1/v} e^{-\beta\Delta G^\ddagger [1-(1-vF\Delta x^\ddagger/\Delta G^\ddagger)^{1/v}]}, \quad (2.22)$$

where τ_0 is the expected lifetime at zero force and the other parameters retain their above definitions. This relation was verified by Dudko *et al.* [76] for DNA hairpins and for a protein. The curvature of the master curve in log space illustrates that the dependence of rates on force is greater than exponential.

The formalism developed by Dudko *et al.* [68, 76], in addition to being easier to implement than full landscape extraction methods, may also have the advantage of being valid for cases in which the pulling coordinate in LOT experiments is not a good reaction coordinate [49]. By this, it is meant that the landscape projection onto the pulling coordinate, or the PMF, is not “good” in the sense that the dynamics on the full, multi-dimensional landscape are obscured in the projection. Dudko *et al.* demonstrated for simulated data that their formalisms for extracting Δx^\ddagger and ΔG^\ddagger yield correct values even when the folding barrier is hidden in the PMF projection [49]. Nonetheless, model-dependent methods are not as detailed or as general as the formalisms outlined in Sections 2.1 and 2.2, which do not assume *a priori* a shape for the landscape. In practice, therefore, multiple approaches should be implemented together to obtain the richest possible description of the energy landscape for a biomolecule. In Chapter 3, such a combination of approaches is employed to study a RNA pseudoknot.

Chapter 3

Realization of Free-Energy

Landscape Extraction for a RNA

Pseudoknot

3.1 RNA Pseudoknot

RNA is comprised of constituents similar to those of DNA; the only differences are the inclusion of an additional OH group in the pentose sugar and the substitution of a uracil nucleotide for thymine. In contrast to DNA, however, it normally occurs in the cell as a single strand, which typically forms more complex structures through various tertiary interactions. RNA performs a variety of functions in the cell, from transcribing and translating DNA to enzymatic activity, and is used by viruses to encode genetic material that goes on to infect host cells. One of the processes in which the structure of RNA is important is the phenomenon of programmed -1 ribosomal frameshifting (-1 PRF), which figures in viral propagation. During protein synthesis, the ribosome reads a strand of messenger RNA—or mRNA—in three-nucleotide segments called codons; each codon represents a particular amino acid [77]. Eventually, the ribosome reaches a particular set of three nucleotides that signals completion of the protein and reading ceases. In -1 PRF, the ribosome halts

and is forced backward one nucleotide; this alters the particular codon sequence originally being read and can cause the ribosome to bypass the stop signal [78]. A protein may thus be synthesized that is entirely distinct from the one that would have resulted from the original *reading frame* of (or pattern of codons interpreted by) the ribosome [78]. Efficient frameshifting maximizes the utility of a gene sequence, since a single strand of genetic material can encode multiple proteins. Many viruses, including HIV [79] and SARS [80], undergo tightly-regulated frameshifting that produces alternate protein products in a definite ratio [78, 81], which enables them to carry out myriad functions, such as replication of the viral genome [82]. While the mechanism of -1 PRF remains controversial [44], the basic ingredients are known to be (a) a sequence on the strand of mRNA that the ribosome can easily slide across and (b) a complex structure a few nucleotides upstream of this “slippery” sequence that promotes the stalling of the ribosome [78, 81]. Most often, this latter structure is a *pseudoknot*.

The simplest secondary structure a strand of RNA can generate is a hairpin, consisting of a stem of complementary base pairs and a loop. A pseudoknot is formed when a “hairpin loop base-pairs with a complementary sequence outside of that loop” [83], giving rise to a complex tertiary structure containing at least two stem regions and at least two loop regions. The structure’s name derives from the fact that, while it is complex, it is not a true knot in the topological sense (since it can be continuously deformed into a linear strand). It has been suggested that a pseudoknot’s mechanical resistance to unfolding is responsible for stimulating frameshifting; in this scenario, the ribosome slips backwards as it attempts to unravel the pseudoknot [84]. In their survey of pseudoknots from a variety of viruses, however, Ritchie *et al.* found that mechanical resistance is uncorrelated with frameshifting; instead, frameshifting appears to be related to a pseudoknot’s tendency to fold into alternate structures [44]. The specific mechanism of frameshifting remains nevertheless elusive, and Ritchie *et al.*’s results invite a more in-depth look at the complete folding landscapes of RNA pseudoknots.

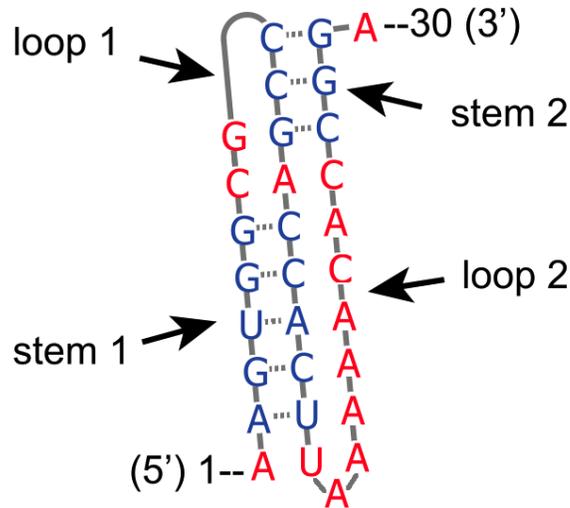
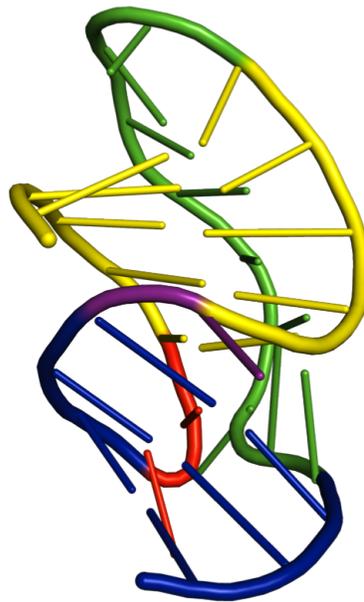
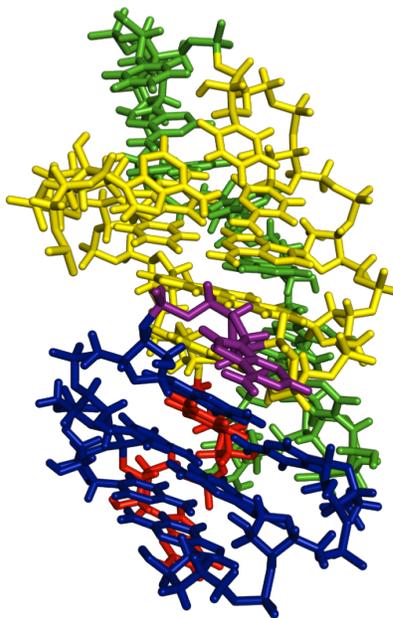


Figure 3.1: The sequence of RNA pseudoknot ScYLV. The pseudoknot contains 28 nucleotides and is divided into 2 stem structures (blue) and two loops (red). The solid grey line indicates the backbone of the RNA chain, whereas dotted grey lines denote base-pairing interactions. Some nucleotides are also involved in triplex bonds. Figure taken from Ritchie *et al.* 2012 [44].

A valuable strategy for studying frameshifting mechanisms, and particularly, how they may be related to energy landscapes, is to examine a system and a well-characterized mutant whose frameshifting efficiency is markedly different from that of the wild-type. The comparison of wild-type and mutated landscapes can then shed light on the determining factors of frameshifting efficiency. One candidate for such a study is the Sugar Cane Yellow Leaf viral pseudoknot (ScYLV), whose C27A mutant has been well-characterized, structurally and thermodynamically [85]. The two pseudoknots, both 28 nucleotides long, differ only by a single nucleotide in the sequence—which in the wild-type is cytosine and in the mutant is adenine—and are structurally similar [85]. Their frameshifting efficiencies are quite different, however: the wild-type has a 15% – 1 PRF efficiency, whereas the C27A mutant has 2% [44]. In the present work, a full landscape profile is constructed for the wild-type pseudoknot, shown in Figures 3.1 and 3.2. Future studies characterizing its mutant can build on the results detailed herein.



(a)



(b)

Figure 3.2: Structure of RNA pseudoknot ScYL V. The pseudoknot is 28 nucleotides in length and contains 2 stems (yellow and blue) and 2 loops (purple/red and green). In (a), the nucleotides are represented by sticks, while in (b), the ribose (5-carbon) sugar rings are explicitly depicted. Figures created using MacPyMOL.

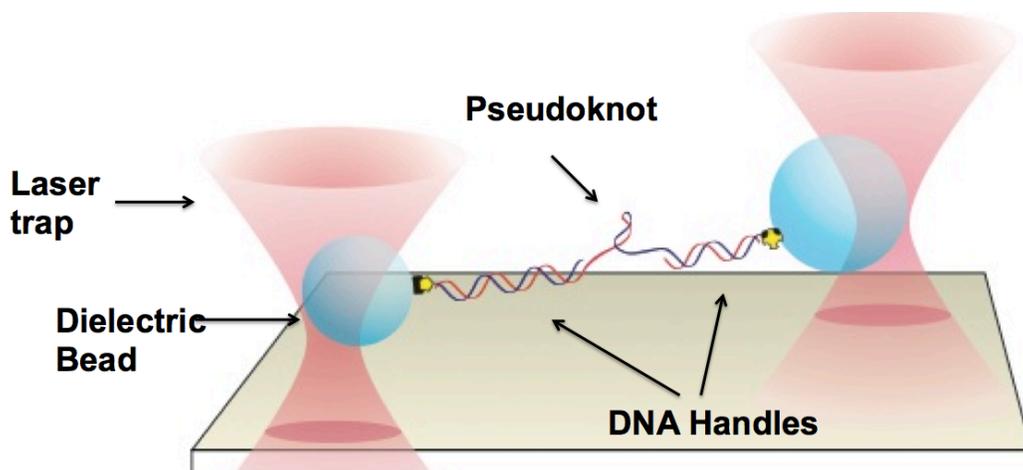


Figure 3.3: Simplified representation of the experimental setup. Pseudoknots were attached to long linker handles, of lengths ~ 840 and ~ 2200 nucleotides, which were then attached to dielectric beads of average sizes 600 nm and 820 nm. Figure courtesy of M.T. Woodside.

3.2 Experimental Methods

A simplified experimental schematic is shown in Figure 3.3. The apparatus used is similar to ones detailed previously [86, 41, 44]. It is a custom-built, dual-beam optical trap in which two laser traps are created from orthogonally-polarized 1064-nm beams. The intensity and position of each trap are manipulated independently with acousto-optic deflectors (AODs). Motion of the dielectric beads in the traps is detected by measuring changes in the amount of light scattered by the beads from an 830-nm detection laser onto PSDs. Long wavelengths are used for laser light to minimize damage to biological samples [46]. Data are sampled at 20 kHz, and an 8-pole Bessel filter is used to filter data online at the Nyquist frequency, 10 kHz. The trap is located in a low-vibration, isolated environment that is temperature-controlled to remain at 20.0°C. Trap stiffnesses are normally calibrated via three different methods: first, by measuring the power spectrum of the beads in the trap and fitting it to a Lorentzian function; second, by measuring the variance of the bead displacements and employing the equipartition theorem; and third, by measuring the bead response to fluid motion and fitting it to predictions based on Stokes drag [47]. Trap positions

are calibrated using the known motions of a piezoelectric stage; nanometer stage motions are translated into pixels on a CCD camera; then motion of the laser traps is used to translate AOD frequency changes into nanometers. Laser traps are moved apart at definite rates by adjusting the frequencies of the AODs to perform force ramp experiments. Further details regarding optical tweezers configuration and calibration are available in published reviews [47, 46].

The pseudoknot constructs were prepared and measured by Dustin Ritchie; see Ritchie *et al.* for details [44]. The pseudoknots were flanked by linker handles that were hybrids of ssDNA and RNA; the handle lengths were approximately 840 and 2200 nucleotides. The handles were attached to 600 nm and 820 nm beads to form dumbbells, which were placed in a buffer. The dumbbell solution was then deposited on a plasma-cleaned microscope slide for trapping.

In total, 15 separate sets of data were collected and analyzed, comprising 9 different molecules. The data were collected over the course of 17 months, from May 2011 to October 2012. Results from four of the data sets, collected in 2011, have been previously published [44], but the analysis was repeated for the current work. The trap stiffnesses for these data sets were 0.58–0.63 pN/nm for one trap and 0.43–0.45 pN/nm for the other. Most of the data sets contain 100–200 pulling curves, collected at pulling speeds ranging from 100 nm/s to 267 nm/s. Between each curve, the traps were held at near-zero force for 3–10 seconds to allow the pseudoknot to refold before mechanically unfolding it again.

3.3 Results and Analysis

3.3.1 WLC Fits

Typical pulling curves for the pseudoknot are shown in Figure 3.4. The FECs rise nonlinearly at low forces as the linker handles are stretched out. An abrupt increase in extension and decrease in force signals unfolding of the pseudoknot; this generates a characteristic sawtooth-shaped “rip”. The curved regions before and after this

unfolding rip reflect the elastic behaviour of the linker handles and are well-described by worm-like chain (WLC) polymer theory [87, 88]. In WLC theory, dsDNA is modelled as an elastic rod, which bends rather than kinking as a freely jointed chain would. WLCs are parametrized by their contour length L_c , or total length along the polymer backbone; their persistence length L_p , which describes the flexibility of the rod, or tendency for different segments to align; and their elastic modulus K , which measures the polymer's resistance to longitudinal stretching [88]. The interpolation formula generally used to fit FECs is [88, 46]:

$$F(x) = \frac{k_B T}{L_p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} + \frac{F}{K} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K} \right] \quad (3.1)$$

Here, $F(x)$ is the force as a function of end-to-end extension x , L_p and L_c are the persistence and contour length of the polymer, respectively, and K is the elastic modulus. At low forces, the polymer behaves as an entropic spring, while at high forces, the stretching of bonds in the polymer backbone is the dominant effect [46].

Individual FECs were first aligned to correct for the instrumental drift that occurs over the course of a measurement. The low force region of the FECs (below the unfolding rip) was then fit to Equation (3.1), thereby obtaining the WLC parameters describing the elasticity of the handles. An additional fitting parameter was included to account for any offset in force arising from the force calibration, since $F = 0$ at $x = 0$ is required physically. Typical values of the persistence length for dsDNA are 40–50 nm, and typical values for the elastic modulus are 1000–1200 pN [88]. The experimental fits yielded persistence lengths somewhat lower, at 14–30 nm. This is likely due to the fact that the linker handles are DNA/RNA hybrids, which have been known to have smaller L_p than dsDNA due likely to imperfect annealing [89]. We next fit the high-force region above the unfolding rip to two WLCs in series: one for the handles, using the same parameters found by the first fit, and the other for the unfolded RNA pseudoknot. In the second fit, the persistence length and elastic modulus of the unfolded RNA were fixed at typical values: 1 nm and 2000 pN,

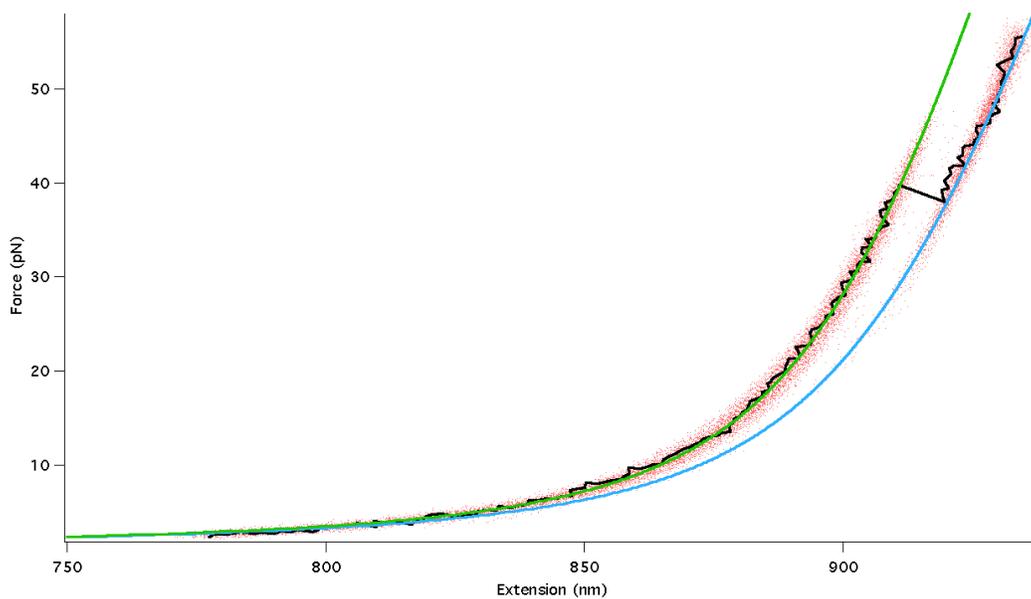


Figure 3.4: Raw FECs for a particular data set of 99 curves. The WLC fits represented in green (WLC with folded pseudoknot) and blue (WLC with unfolded pseudoknot) have a handle persistence length L_p of 27 nm, a handle contour length L_c of 919 nm, a handle elastic modulus of 1716 pN, a contour length change upon unfolding of $\Delta L_c = 11.8$ nm, and a force offset of 1.2 pN. This force offset is subtracted from all force values prior to further analysis. The persistence length and elastic modulus for the unfolded RNA were fixed at 1 nm and 2000 pN, respectively.

respectively [90, 91]. The unfolded RNA contour length was thus the only parameter free to vary in the fit, yielding the contour length change upon unfolding, ΔL_c . From fits like the ones shown in Figure 3.4 (green: RNA folded; blue: RNA unfolded), we found $\Delta L_c = 12.2 \pm 0.5$ nm, similar to the result found earlier (13 ± 1 nm) [44]. This value is slightly lower than what is expected from the high-resolution structure of this pseudoknot (13.4 nm), determined by subtracting the distance between the ends of the folded pseudoknot (1.92 nm) from the L_c of 28 nucleotides of RNA (16.52 nm), but it is still consistent within error.

3.3.2 Model-Dependent Analysis

After the FECs were aligned and adjusted using the force offset value derived from the WLC fits outlined in Section 3.4.3, the model-dependent methods of Dudko *et al.* [68, 76] were employed to study the folding landscape of the pseudoknot. First, force rupture histograms like the one in Figure 3.5 were compiled for each of the 15 data sets. The average unfolding force for each depends on the pulling speed, and faster pulling rates will induce folding at higher forces (see Equation (5) from reference [68]), but the shape of each distribution is similar. The distributions are then fit to Equation (2.19), allowing for the extraction of the barrier height ΔG^\ddagger , distance to barrier Δx^\ddagger , and escape rate from the folded well k_0 . In Figure 3.5, the red curve is a fit performed under the assumption of a linear-cubic landscape like the one in Figure 2.6, where $\nu = 2/3$, and the black curve is a fit performed under the assumption of a cusp-like barrier, where $\nu = 1/2$. For most distributions, there was little difference between these models and the parameters they produced, and stability for different ν values suggests the parameters returned by the fits are meaningful [76].

As an alternative way of looking at the same information, we transform the force-rupture distributions into force-dependent lifetimes according to Equation (2.21), and $\ln \tau(F)$ was fit in the least squares sense with multiple values of ν as per the suggestion of Dudko *et al.* [76]. Figure 3.6 illustrates a sample lifetime curve along with fits; the fit range excluded points on either end of the distribution corresponding

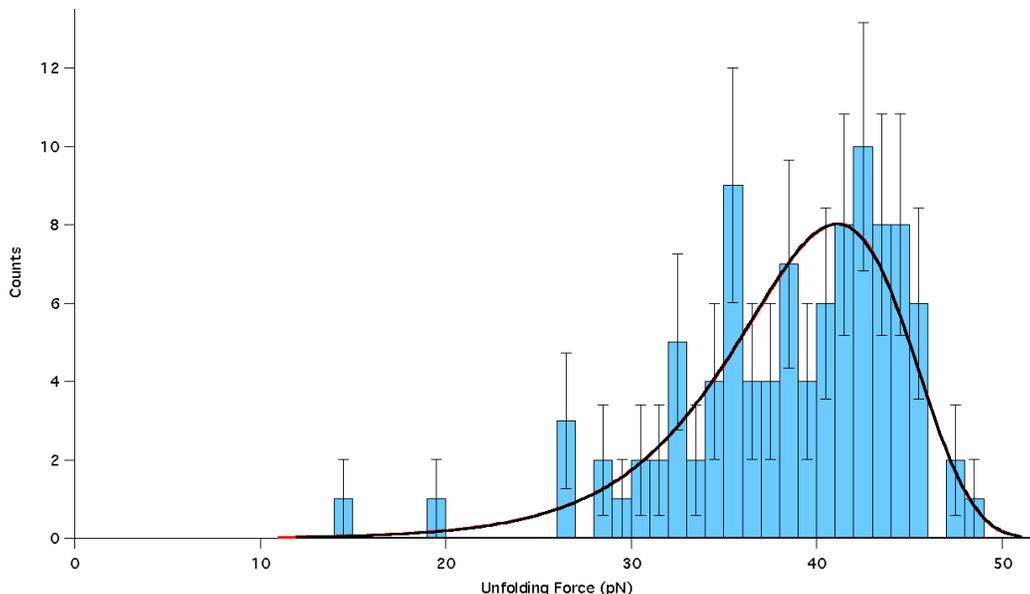


Figure 3.5: A force-rupture distribution representing 100 pulling curves for which the average force loading rate \dot{F} was 44.3 pN/s. The red curve is for a fit with $\nu = 2/3$; the landscape parameters it produced are $k_0 = 6 \times 10^{-5} \text{ s}^{-1}$, $\Delta G^\ddagger = 165 \text{ pN}\cdot\text{nm}$, and $\Delta x^\ddagger = 0.98 \text{ nm}$. The black curve is for a fit with $\nu = 1/2$; the landscape parameters it produced are $k_0 = 8 \times 10^{-4} \text{ s}^{-1}$, $\Delta G^\ddagger = 152 \text{ pN}\cdot\text{nm}$, and $\Delta x^\ddagger = 1.0 \text{ nm}$. The average unfolding force for this distribution was 39 pN. Errors are standard \sqrt{N} counting errors, but the fits are not weighted.

to force bins with very few counts.

Finally, the data are examined from yet another angle by compiling plots representing the cumulative probability of finding the pseudoknot unfolded as a function of force, $P(F)$; this is simply the integral of the force-dependent unfolding probability distributions. $P(F)$ is computed according to the method outlined in Section 2.3, and the resulting sigmoidal curve is then fit using Equation (2.20). We finally averaged the results from the three fitting methods, to obtain our best estimate of the true values. Errors on final values are standard errors of the mean. Final average values are displayed in Table 3.1, along with the previously reported values of Ritchie *et al.* [44]. Values for both models, linear-cubic and cusp-like, are fairly close to one another, lending weight to the results. The average between the results for $\nu = 1/2$ and $\nu = 2/3$ is $\Delta G_{F=0}^\ddagger = 59 \pm 6 \text{ kJ/mol}$.

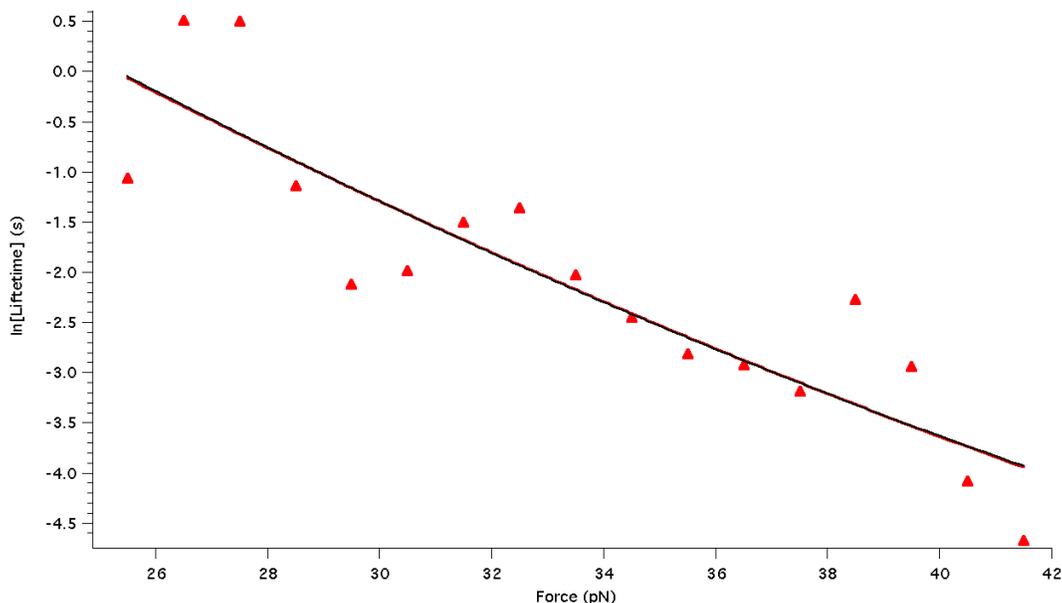


Figure 3.6: Force-dependent lifetimes, $\tau(F)$, from 99 pulling curves for which the average force loading rate \dot{F} was 54.4 pN/s. The red curve is for a fit with $\nu = 2/3$; the landscape parameters it produced are $\tau_0 = 1677$ s, $\Delta G^\ddagger = 68$ pN·nm, and $\Delta x^\ddagger = 1.5$ nm. The black curve is for a fit with $\nu = 1/2$; the landscape parameters it produced are $\tau_0 = 7120$ s, $\Delta G^\ddagger = 76$ pN·nm, and $\Delta x^\ddagger = 1.8$ nm. The average unfolding force for this distribution was 32.9 pN.

Table 3.1: Summary of results from model-dependent fits to the pseudoknot data. Parameters obtained by assuming both cusp-like ($\nu = 1/2$) and linear-cubic ($\nu=2/3$) landscape shapes are shown. Results from Ritchie *et al.* (marked Previous Value) are displayed for comparison [44].

Parameter	$\log k_{\text{off}} (\text{s}^{-1})$	$\Delta x^\ddagger (\text{nm})$	$\Delta G^\ddagger (\text{kJ/mol})$
$\nu = 1/2$	-4.5 ± 0.2	1.8 ± 0.1	64 ± 5
$\nu = 2/3$	-5 ± 1	1.7 ± 0.1	54 ± 4
Previous Value	-5.0 ± 0.3	1.7 ± 0.1	60 ± 4

3.3.3 Full Landscape Reconstruction

We next analyzed the pseudoknot FECs to reconstruct the shape of the landscape profile without assuming any particular model, by using the WHAM method described in Section 2.2. Landscape profiles were calculated using a custom C++ implementation of the WHAM procedure. Note that 4 of the 15 data sets had insufficient statistics for a reliable reconstruction: in order for the Jarzynski-based WHAM method to

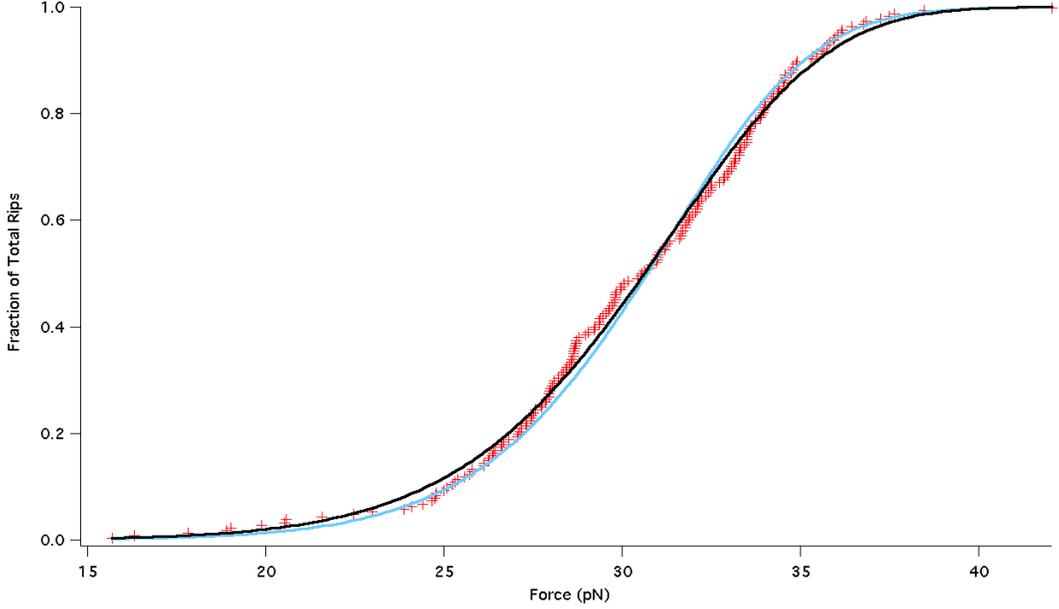


Figure 3.7: An integrated probability distribution representing 199 pulling curves for which the average force loading rate \dot{F} was 52.8 pN/s; the data are shown in red. The blue curve is for a fit with $\nu = 2/3$; the landscape parameters it produced are $k_0 = 7 \times 10^{-6} \text{ s}^{-1}$, $\Delta G^\ddagger = 67 \text{ pN}\cdot\text{nm}$, and $\Delta x^\ddagger = 2.4 \text{ nm}$. The black curve is for a fit with $\nu = 1/2$; the landscape parameters it produced are $k_0 = 8 \times 10^{-6} \text{ s}^{-1}$, $\Delta G^\ddagger = 70 \text{ pN}\cdot\text{nm}$, and $\Delta x^\ddagger = 2.5 \text{ nm}$. The average unfolding force for this distribution was 30.2 pN.

work properly, the rare trajectories with low unfolding forces that represent the tail of the work distribution must be adequately sampled. The landscape profile was reconstructed for 11 data sets, containing 1217 curves. The system stiffness k used in the calculation of work in Equation (2.11) was determined by combining the two laser trap stiffnesses k_0 and k_1 , treating them as springs in series:

$$k = \left(\frac{1}{k_0} + \frac{1}{k_1} \right)^{-1}. \quad (3.2)$$

The result of Equation (2.10) is a free-energy landscape profile at zero force, $G_0(q)$; a sample of such a landscape from one data set comprised of 199 curves is shown in Figure 3.8. As would be expected, the landscape is dominated by the energy of stretching the linker handles [41], and the unfolding barrier is only a small blip in the landscape around 890 nm. In order to more clearly see the details of the barrier, it is

desirable to “tilt” the zero-force landscape to a constant force value, simulating the result that would be obtained in a constant-force experiment:

$$G_F(x) = G_0(x) - Fx. \quad (3.3)$$

The effect of tilting the landscape to various forces is shown in Figure 3.9. The landscape was ultimately tilted to $F_{1/2}$, the force at which the folded and unfolded minima are at the same free-energy and the molecule has a 0.5 probability of being in either state. Landscapes from the 11 data sets were each tilted to their $F_{1/2}$ value, which varied from landscape to landscape, then combined to form an average landscape (black curve in Figure 3.10). The average $F_{1/2}$ was 30 ± 2 pN. The barrier height of the average reconstruction is 14 ± 2 kJ/mol and the distance between folded well and barrier on the average reconstruction is 5.0 ± 2 nm. There is a fairly wide spread of results in the barrier region of the profiles, similar to previous work applying the WHAM analysis to a protein molecule [20] and reflecting the variability of exponentially-weighted averages.

3.3.4 Estimates of Free-Energy Difference between Folded and Unfolded States

There are two methods for determining the intrinsic free-energy difference between the folded and unfolded states (this is *not* equivalent to the barrier on the free-energy landscape, which contains not only the intrinsic energy difference between states, but the reaction activation energy, as well) which were pursued for this work. Firstly, the WLC fits to the FECs can be integrated; since they are force-distance curves, the integral of the WLC fits corresponds to the free-energy invested in stretching the linker handles. The WLC integrals are then overlaid on the zero-force free-energy landscape; see Figure 3.11. Clearly, as aforementioned and noted in other work [41], the zero-force free-energy landscape contains contributions from the stretching of the linker handles. The difference between the y-intercepts of the two integral curves

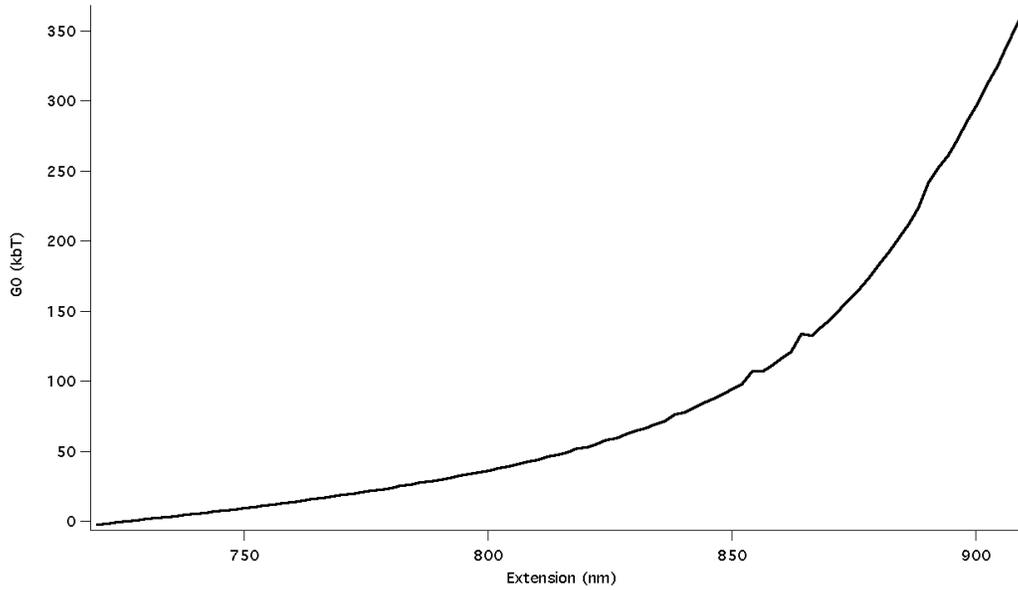


Figure 3.8: Free-energy landscape at zero force constructed using the WHAM method of Hummer and Szabo [39] for a data set comprising 199 FECs. A clear feature of the landscape is the energy of stretching the linker handles (regions of broad curvature) [41]; the barrier to unfolding is the small blip visible at 890 nm. The extension bin size is 2.0 nm and the time bin size is 0.008 seconds. There are an average of 358 points per extension bin.

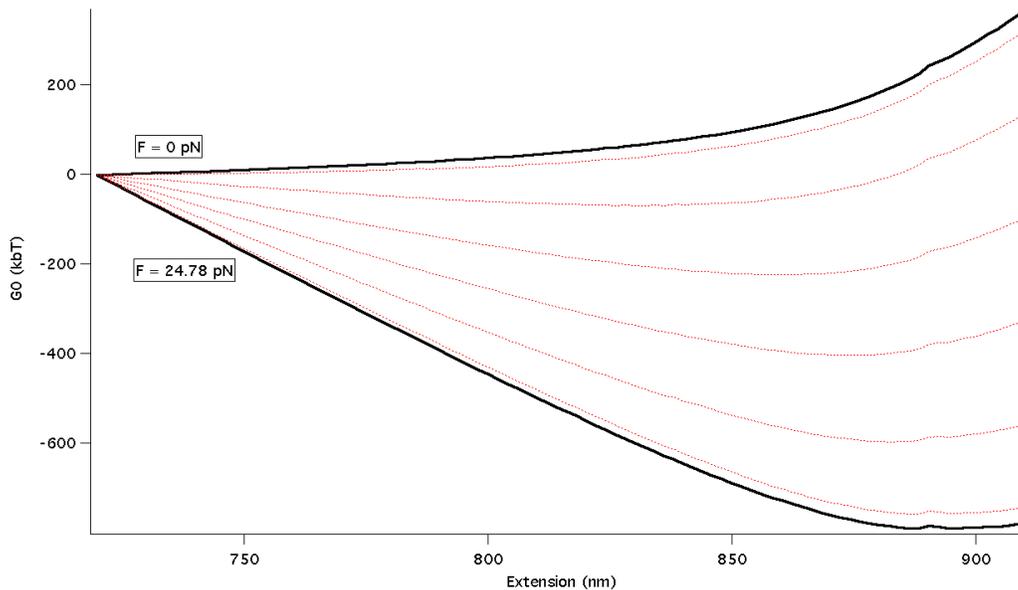


Figure 3.9: Landscape shown in Figure 3.8, smoothed for clarity and tilted to a variety of constant forces. The black curve at the bottom represents the landscape tilted to $F_{1/2}$. The barrier to unfolding is visible at 890 nm.

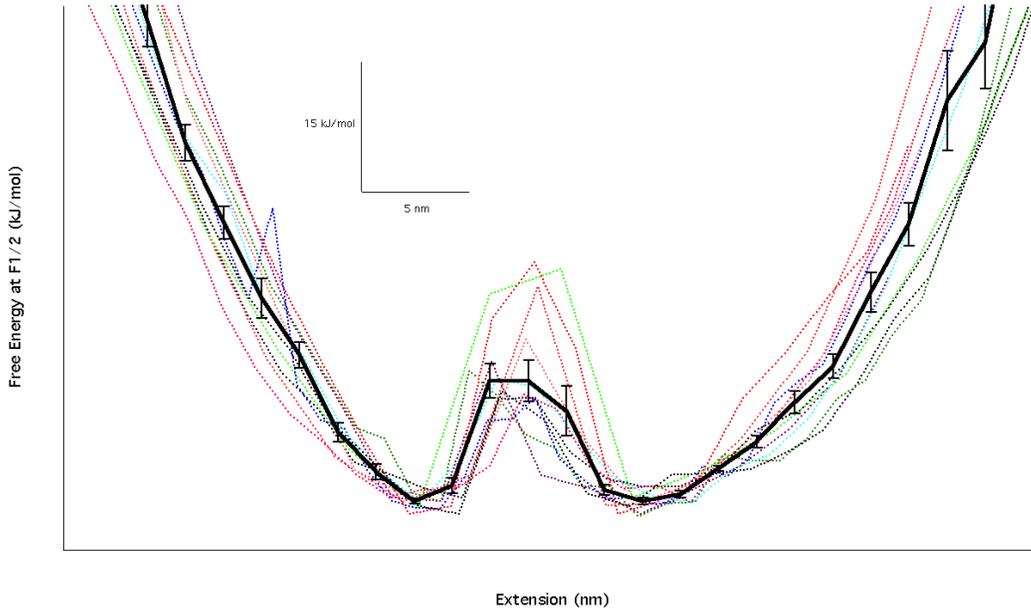


Figure 3.10: Final, average WHAM-reconstructed landscape for the ScYLV pseudoknot, tilted to an average force of 30 ± 2 pN. Errors are standard errors on the mean. Eleven data sets were combined, comprising a total of 1217 FECs. The bin size is 1.8 nm.

represents the difference in free-energy between the folded and unfolded states of the pseudoknot. WLC integration and overlaying onto the G_0 energy profile was pursued for 13 of the 15 sets of FECs; for the other two sets, the WLC integrals did not match the G_0 profile well. The free-energy differences thus obtained were averaged; the final result for free-energy difference between folded and unfolded states is $\Delta G_{\text{fold}} = 135 \pm 11$ kJ/mol.

Another estimate for the energy difference between for folded and unfolded states can be determined as follows:

1) Multiply the average $F_{1/2}$ for the tilted landscape by the distance between the unfolded well and the folded well to derive the total work required to bring the molecule from the folded to the unfolded state. This inter-well distance can be determined using the black curve in Figure 3.10 and is $11 \text{ nm} \pm 2 \text{ nm}$. The average $F_{1/2}$ is 30 ± 2 pN, so $F_{1/2} \times \Delta x = 330 \pm 64$ pN·nm.

2) Subtract from this value of work the portion of the energy that goes into

stretching the pseudoknot once it's unfolded. To determine what this is, a WLC corresponding to the unfolded strand of RNA is plotted and integrated out to the extension it would have at $F_{1/2}$. When this is done, the stretching energy is found to be 83 pN·nm. The error in the stretching energy can be estimated to be $\sim 10\%$, or 8 pN·nm.

3) Subtracting 2) from 1) yields an estimate for the free-energy difference between folded and unfolded states: 330 ± 64 pN·nm - 83 ± 8 pN·nm = 247 ± 64 pN·nm, or 146 ± 38 kJ/mol.

The two estimates, 135 ± 11 kJ/mol and 146 ± 38 kJ/mol, are in agreement with one another. Because these estimates are both derived from the landscape reconstruction for ScYLV, this agreement is encouraging; the landscape is self-consistent.

3.3.5 Comparison of Results

Several different characterizations of the free-energy landscape were detailed above, and it is valuable to compare the various parameter results reported to ensure consistency. In order to compare the zero-force ΔG^\ddagger value derived in Section 3.4.2 with the barrier height of the reconstructed landscape in Figure 3.10, we must first transform the former estimate to the constant-force regime. That is, given a barrier height of ΔG^\ddagger at $F = 0$, we would like to know what that barrier height is at $F = F_{1/2} = 30 \pm 2$ pN. To perform this transformation, it is necessary to (a) *add* to ΔG^\ddagger the energy required to stretch the portion of the pseudoknot that is unfolded at the midpoint of the energy barrier; and (b) *subtract* from ΔG^\ddagger the work done by the constant force $F_{1/2} = 30 \pm 2$ pN. For the estimates performed here, the average of the $v = 1/2$ and $v = 2/3$ results is used.

The stretching energy can be found by integrating a WLC with L_c equalling the amount of RNA unfolded at the midpoint of the barrier. We can crudely estimate this L_c as the one that yields an extension of 1.8 ± 0.1 nm—the barrier position from the model-dependent fits—at a force of $F_{1/2} = 30 \pm 2$ pN. This length is 2.2 nm. Integrat-

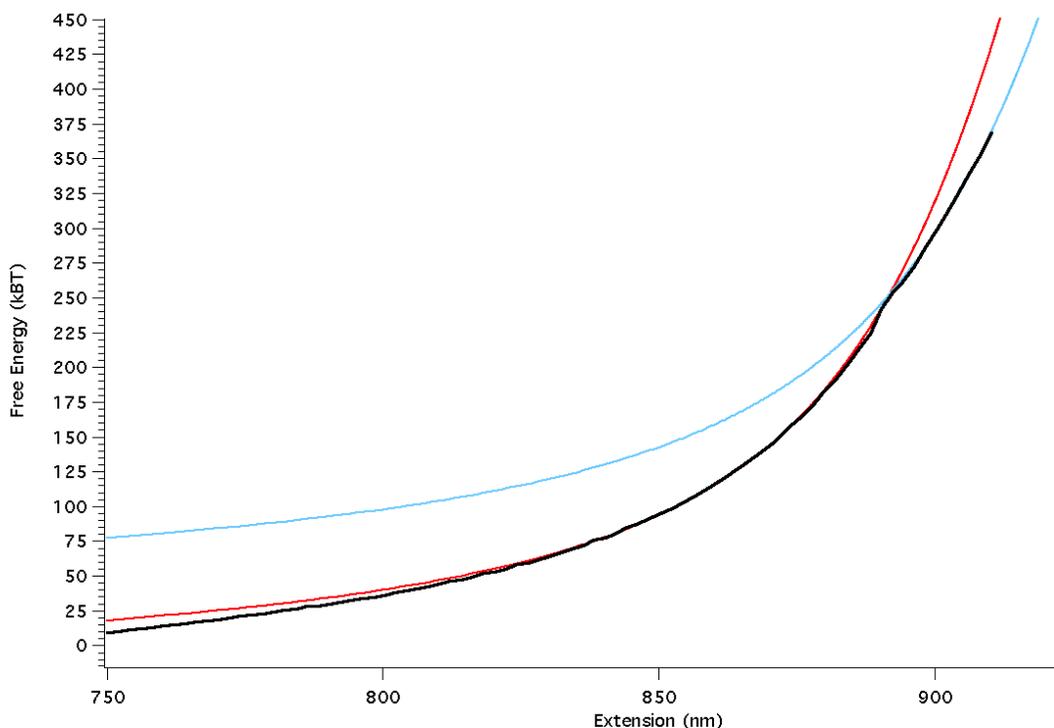


Figure 3.11: The WLC fits for the unfolded and folded states are integrated, and the results are overlaid on the zero-force free-energy reconstruction. The landscape profile contains contributions from the free-energy of the handles in both the folded (red) and unfolded (blue) states. The difference in y-intercept between the curves represents the change in free-energy due to the molecular transition from the folded to the unfolded state. For this data set, this difference is $62 k_B T$, or $149 kJ/mol$.

ing a WLC with this contour length out to a force of 30 pN yields a stretching energy of ~ 11 pN·nm, or ~ 6.5 kJ/mol. Again, we can estimate the error on this stretching energy to be about $\sim 10\%$: 6.5 ± 0.7 kJ/mol. The work done by the constant $F_{1/2}$ can be found by multiplying $F_{1/2}$ by Δx^\ddagger . Using the Δx^\ddagger value from Section 3.4.2 (1.8 ± 0.1 nm), this work is 54 ± 5 pN·nm, or 32 ± 3 kJ/mol. The model-dependent value for barrier height adjusted to constant force is thus:

$$\begin{aligned}
 \Delta G_{F_{1/2}}^\ddagger &= \Delta G_0^\ddagger - F_{1/2} x^\ddagger + E_{stretch} \\
 &= 59 \pm 6 \text{ kJ/mol} - 32 \pm 3 \text{ kJ/mol} + 6.5 \pm 0.7 \text{ kJ/mol} \quad (3.4) \\
 &= 34 \pm 7 \text{ kJ/mol}.
 \end{aligned}$$

Thus, we see that the result from the model-dependent fit, 34 kJ/mol, does not within

error match the barrier height on the reconstructed landscape of Figure 3.10, 14 ± 2 kJ/mol. One expects that the WHAM-reconstructed landscape will feature a barrier that is *lower* and *broader* than the true barrier; this is due to the convolution of the intrinsic landscape with the PSF of the beads, handles, and traps [51]. Indeed, Yu *et al.* present a reconstructed landscape for a protein with a barrier height that increases by a factor of ~ 3 upon deconvolution, bringing the value from the WHAM landscape into agreement with the model-dependent fit value [20]. Further, Woodside *et al.* [37] found that the barrier heights on deconvolved DNA hairpin landscapes were a factor of ~ 2 higher than the heights on the undeconvolved landscapes, consistent with discrepancy in the present work, which is slightly larger perhaps due to the longer handles used here (a larger $\frac{L_p}{L_c}$ ratio leads to more severe blurring [51]).

We can also compare the barrier position result from the model-dependent fits, $\Delta x^\ddagger = 1.8 \pm 0.1$ nm, with the value from the landscape reconstruction, 5 ± 2 nm. Even though the former value is for $F = 0$, tilting the landscape with applied force will only move the barrier *closer* to the folded well (in accordance with expected Hammond behaviour [92]); thus 1.8 ± 0.1 nm is an upper limit on the Δx^\ddagger value at $F_{1/2}$. Convolution with PSF tends to “smear” barriers out horizontally as well, moving them more towards the central point between wells; this effect was observed by Woodside *et al.* [37] for several DNA hairpins. The above comparisons and the evidence of previous work thus highlight the need for deconvolution of the pseudoknot landscape. Such a deconvolution is not straightforward, however; because of the larger range of forces involved in pseudoknot unfolding (10–50 pN), deriving an empirical, position-dependent PSF following the treatment of Yu *et al.* [20], where the force range was only 8–12 pN, would be extremely difficult. More promising would be a theoretical PSF derivation in the spirit of Hinczewski *et al.* [51], but which is valid for arbitrary ranges of force.

The estimates for ΔG_{fold} from Section 3.4.4 can be interpreted in the context of the unfolding barrier height estimates discussed above. Our estimated ΔG_{fold} values (135 ± 11 kJ/mol and 146 ± 38 kJ/mol) can firstly be compared with independent estimates

for the stability of other type-H pseudoknots: the values derived by Cao and Chen [93] for pseudoknots similar to ScYLV are around ~ 40 kJ/mol at 20°C . This is rather too low for consistency with the ΔG_{fold} implied by the WHAM reconstruction; further, the unfolded state appears to lie above the barrier at zero force, $\Delta G^\ddagger = 59 \pm 6$ kJ/mol. While some have argued that barrierless folding at zero force is possible [94], it does not seem to be a common phenomenon, and is usually the case for folding that occurs much faster than is observed for pseudoknots. We consider, then, the possibility that the ΔG_{fold} values derived using the WHAM method are inaccurate. A possible source of this inaccuracy is the bias associated with the Jarzynski estimator.

Equation (2.5)—and by extension, Equation (2.10)—is only exact in the limit of an infinite set of pulling trajectories. For finite N , the Jarzynski estimator is a biased one; this is because the trajectories that contribute most to the exponential average are the rare ones with negative dissipated work found in the tail of the work distribution. The accuracy of Jarzynski-based methods is predicated upon efficient sampling of this region. Lack of sampling of the tail will lead to inflated estimates for equilibrium free-energy difference [59, 95]; for a given number of trajectories N , this overestimate is more severe the larger the amount of dissipated work. That is, the farther from equilibrium an experiment is carried out, the more trajectories will be required to adequately sample the tail of the work distribution; see Figure 3.12. Gore *et al.* [95] propose the following expression for the Jarzynski bias for Gaussian work distributions:

$$B(N) = \langle \Delta \hat{G}(N) \rangle - \Delta G = \frac{\overline{W_{\text{dis}}}}{N^\alpha}, \quad (3.5)$$

where $B(N)$ is the bias for N trials, $\Delta \hat{G}(N)$ is the Jarzynski estimate for free-energy difference, ΔG is the true free-energy difference, $\overline{W_{\text{dis}}}$ is the average dissipated work, the angled brackets indicate an average over all trajectories, and α is an exponent chosen according to a method outlined by Gore *et al.* [95]. Bias is thus more severe for experiments carried out far from equilibrium, where $\overline{W_{\text{dis}}}$ is large.

In previous studies of DNA hairpins, there was no noticeable disagreement between the free-energy difference obtained using the landscape reconstruction and

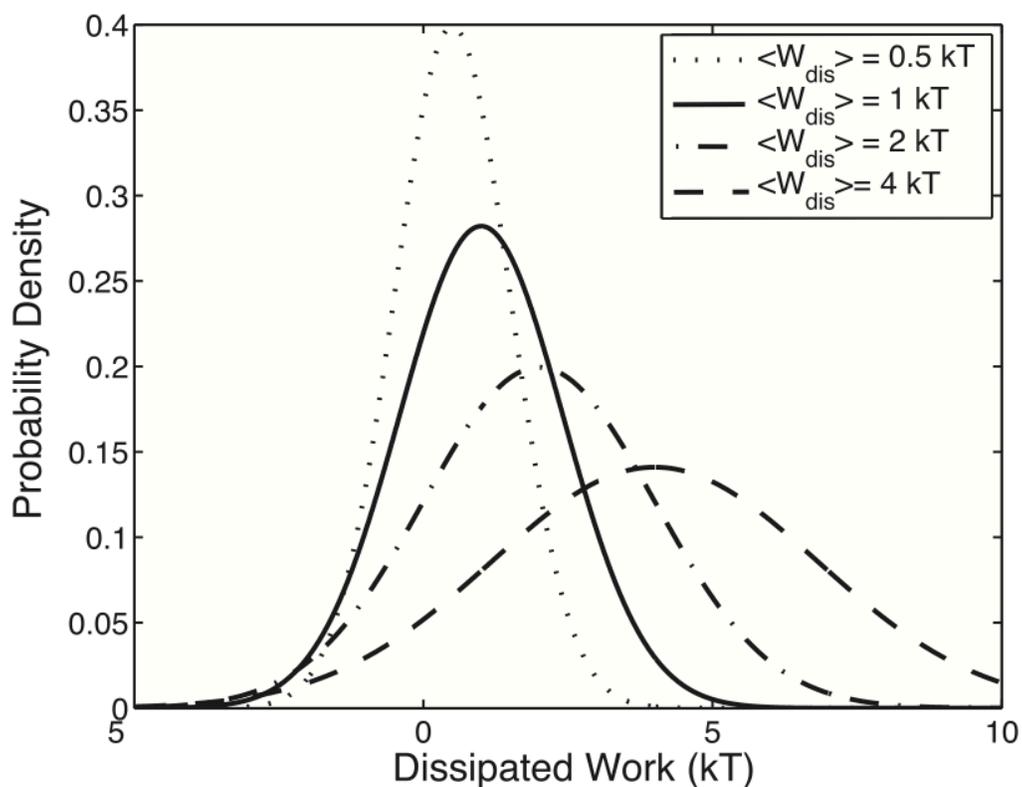


Figure 3.12: As the mean and variance of the dissipated work distribution increases—that is, as measurements become farther and farther from equilibrium—the area of the tail, or region below 0, decreases. This means that more trajectories are needed to efficiently sample the negative-work trajectories which are so heavily weighted by the Jarzynski average. Inadequate sampling leads to bias and overestimation of free-energy differences. Figure taken from Gore *et al.* 2003 [95]

independent estimates [41]; however, these hairpin measurements were carried out much closer to equilibrium, so the bias would correspondingly have been smaller and potentially negligible within error. This is not the case in the present work. Figure 3.13 shows a folding and refolding curve for the ScYLV pseudoknot. It is clear that the hysteresis is very severe; the pseudoknot appears to refold along a completely different path. The RNA system is thus far from equilibrium, and a large number of trials N would be required to bring estimates into the regime where statistical error, and not inherent bias, is the dominant source of noise [95]. If the landscape in Figure 3.10 is indeed biased, the bias would be most severe in the barrier region, where the folding reaction takes place. If all the points on the $G_0(q)$ landscape barrier of Figure 3.8 were lower, $F_{1/2}$ would be lower and the vertical distance between the WLC integrals in Figure 3.11 would decrease. This could bring the estimates of ΔG_{fold} into agreement with the value 59 ± 6 kJ/mol from the model-dependent fits and into the ballpark of the estimates of Cao and Chen [93].

A crude estimate of the bias in the WHAM results can be made as follows: first, $\overline{W}_{\text{dis}}$ can be estimated by multiplying the force midway between where re-folding occurs (~ 5 pN) and where unfolding occurs (~ 30 pN) with the distance between the folded well and the unfolded well (11 ± 2 nm): this yields $\overline{W}_{\text{dis}} \sim 135$ pN·nm. Next, an appropriate value for α taken from Gore *et al.* [95] (~ 0.1) can be plugged into Equation (3.5) to yield $B(N)$ for a typical sample of 100 curves: ~ 85 kJ/mol. Such a substantial bias could easily account for the seemingly large ΔG_{fold} values. The manifestation of bias in the theories of Hummer and Szabo employed here [39, 66] has not yet been theoretically explored; future work should be aimed at understanding and correcting for the influence of finite sample sizes in WHAM reconstructions.

In spite of the likely bias inherent in the Jarzynski-based landscape reconstruction pursued here and the need for deconvolution, the landscape in Figure 3.10 is notable in being the first complete landscape profile computed for an RNA pseudoknot. Further, it accurately recaptures the greater proximity of the folding barrier to the folded rather than the unfolded well, and it thus may be possible to glean useful

information from a comparison of the landscapes of this pseudoknot and its C27A mutant, whose frameshifting efficiency is much smaller.

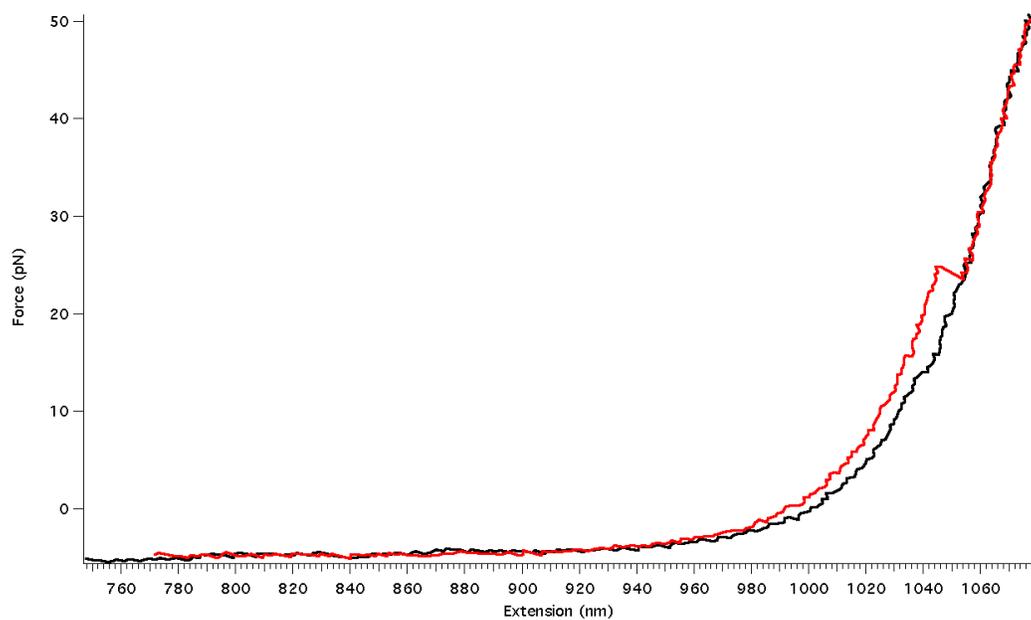


Figure 3.13: A folding curve (red) and a corresponding refolding curve (black) for ScYLV. Severe hysteresis is clear; this system is being driven far from equilibrium.

Chapter 4

Method of Inverse Weierstrass Transform

The Weighted Histogram Analysis Method (WHAM) of Hummer and Szabo [39] continues to be used to extract full free-energy landscape profiles from single-molecule force-ramp pulling experiments. Recently, however, Hummer and Szabo proposed an alternative framework for landscape reconstruction, predicated upon a mathematical integral transformation [66]. While their method accurately reproduces free-energy surfaces from simulated RNA unfolding optical tweezer experiments, up until now it has not been validated against experimental data. Here, such a validation is pursued, and the method is used to reconstruct free-energy landscapes for a DNA hairpin and a RNA pseudoknot. The landscapes for the systems studied have already been characterized using established analysis techniques, and the results of the Weierstrass analysis can be compared against these other reconstructions. As will be shown, the Weierstrass transform method, though much simpler to implement than the WHAM method, in practice only works for systems possessing a sufficiently stiff energy barrier in comparison to the stiffness of the pulling spring.

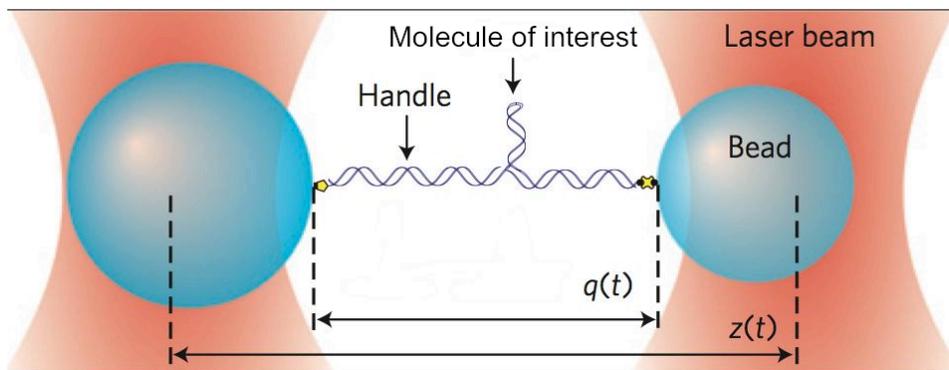


Figure 4.1: Optical tweezers set-up. Here, q represents the extension of the molecule of interest plus linker handles, while $z(t)$ represents an experimental control parameter, the trap separation. Figure adapted from Gupta *et al.* 2011 [41].

4.1 Theory

The goal of any landscape reconstruction procedure is to produce a free-energy profile as a function of a relevant reaction coordinate for the system of interest. In the LOT experiments considered here, this coordinate is the molecular extension of the system, which in practice includes the molecule of interest plus the linker handles attaching it to the apparatus applying force [39, 66, 51]. As mentioned in Section 2.2, the difficulty in using Jarzynski's equality to accomplish this aim is that it yields free-energy as a function of an experimental control parameter; in the LOT experiments carried out for this thesis, this is the inter-trap distance. In Figure 4.1, $z(t)$ is this control parameter, while $q(t)$ is the system extension, which is the desired abscissa in free-energy profiles. Hummer and Szabo have proposed two ways to transform the FECs obtained in LOT experiments to free-energy profiles as a function of system extension, $G(q)$: the first, the WHAM method, moves directly from FECs to $G(q)$ [39], while the second, the Weierstrass transform method, first uses Jarzynski's relation to derive $A(z)$, the free-energy profile as a function of the experimental control parameter, then transforms this into $G(q)$ [66]. Figure 4.2 establishes these two procedures pictorially. The explanation in the following paragraphs follows the work of Hummer and Szabo in their 2010 paper.

The first step in the Weierstrass transform procedure is obtaining $A(z)$, the free-

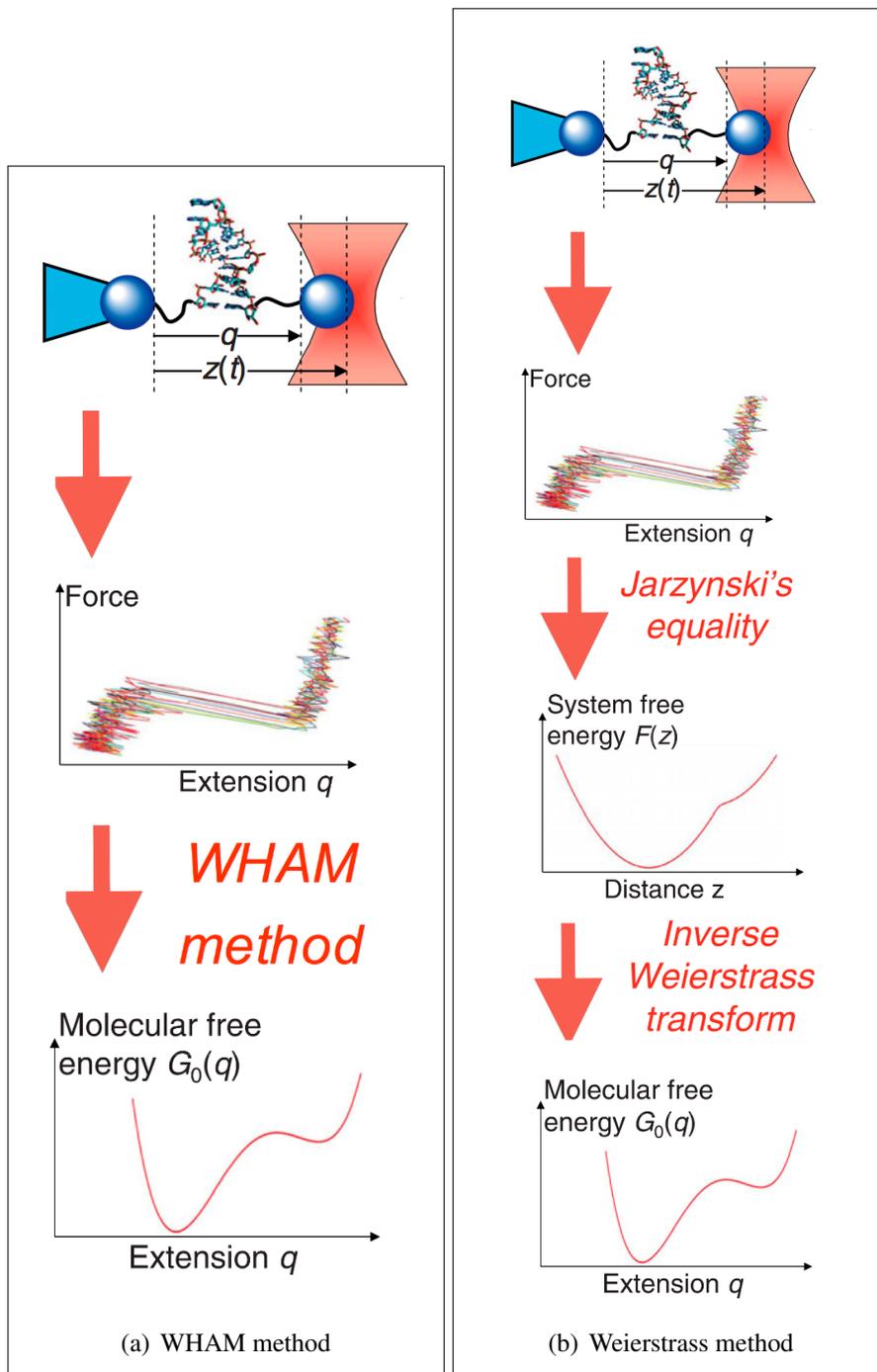


Figure 4.2: Illustration of two landscape extraction methods. The WHAM method is shown in (a); FECs are translated directly into the free-energy landscape as a function of molecular extension coordinate q : $G_0(q)$. The Weierstrass method, shown in (b), involves first constructing the free-energy as a function of the experimental control parameter, $A(z)$, then using an inverse Weierstrass transform to extract $G_0(q)$. Figure adapted from Hummer and Szabo 2010 [66].

energy profile along the experimental control coordinate. This is simply given by Jarzynski's equality:

$$e^{-\beta A(z)} = \langle e^{-\beta W(z)} \rangle, \quad (4.1)$$

where $W(z)$ is the experimental work performed on the system as a function of the trap separation; this work is given by:

$$W(z) = \int_{z(0)}^{z(t)} F dz. \quad (4.2)$$

The angled brackets in Equation (4.1), indicate an average over all pulling trajectories. Now, consider the definition of $A(z)$:

$$A(z) = -k_B T \ln Q(z). \quad (4.3)$$

This is the free-energy as a function of the distance between traps, where $\beta = \frac{1}{k_B T}$, where k_B is Boltzmann's constant, and T is the temperature, and Q is the partition function of the system. The partition function Q is defined by:

$$Q(z) = \int d\mathbf{x} e^{-\beta E(\mathbf{x})}. \quad (4.4)$$

Above, \mathbf{x} is a vector of coordinates in the phase space of the system; these could be, for instance, configurational degrees of freedom. $E(\mathbf{x})$ is the energy of the system in configuration \mathbf{x} ; the integral runs over all of phase space. For single-molecule pulling experiments, we consider the relevant phase space coordinate to be the reaction coordinate, molecular extension q . Now, consider the energy function for a molecule suspended in a harmonic laser trap as a function of molecular extension q and trap separation z :

$$E(q, z) = G_0(q) + V(q, z) = G_0(q) + \frac{1}{2}k(q - z)^2. \quad (4.5)$$

Where $G_0(q)$ is the intrinsic free-energy function for the molecule as a function of

its extension, and $V(q, z) = \frac{1}{2}k(q - z)^2$ is the spring potential energy of the system with spring stiffness k ; the displacement of the beads from the centre of the trap is $z - q$. Substituting this expression for energy into the partition function, noting that the phase space differential $d\mathbf{x}$ becomes dq , and exponentiating Equation (4.3), we have

$$e^{-\beta A(z)} = \int e^{-\beta G_0(q) - \beta \frac{1}{2}k(q-z)^2} dq. \quad (4.6)$$

Now, the definition of the generalized Weierstrass integral transform of a function $f(y)$ is [96]

$$F(x) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} f(y) e^{-\frac{(x-y)^2}{4t}} dy. \quad (4.7)$$

Here t is some positive number. The generalized Weierstrass transform is just the convolution of a function $f(y)$ with a Gaussian function of variance t . Looking again at Equation (4.6), we see that it has the same form as a generalized Weierstrass transform with $x \mapsto z$, $y \mapsto q$, Gaussian variance $t = \frac{1}{2\beta k}$, $F(x) \mapsto e^{-\beta A(z)}$, and $f(y) \mapsto \sqrt{8\pi\beta k} e^{-\beta G_0(q)}$. We can thus obtain an expression for $G_0(q)$, the desired free-energy as a function of molecular extension, by inverting this generalized Weierstrass transformation. Applying one such inversion formula [96] yields

$$e^{-\beta G_0(z)} = \frac{\beta k}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\beta k \xi^2}{2} - \beta A(z+i\xi)} d\xi, \quad (4.8)$$

where ξ is an auxiliary variable. Next, the argument of the exponential is Taylor-expanded to second order about a stationary point to render the integral tractable. This stationary point is $\xi_0 = -\frac{i\dot{A}(z+i\xi_0)}{k}$. An expansion about this point to second order leads to the integral

$$e^{-\beta G_0(z)} \approx \frac{\beta k}{2\pi} e^{-\frac{k\beta\xi_0^2}{2} - \beta A(z+i\xi_0)} \int_{-\infty}^{\infty} e^{-\frac{\xi^2}{2}(k\beta - \beta\ddot{A}(z+i\xi_0))} dx. \quad (4.9)$$

After integration, this becomes

$$e^{-\beta G_0(z)} \approx \frac{\sqrt{\beta k}}{\sqrt{2\pi(k - \ddot{A}(z+i\xi_0))}} e^{-\frac{k\beta\xi_0^2}{2} - \beta A(z+i\xi_0)}. \quad (4.10)$$

or, with $\xi_0 = -\frac{\dot{A}^2(z+i\xi_0)}{k^2}$,

$$e^{-\beta G_0(z)} \approx \frac{\sqrt{\beta}k}{\sqrt{2\pi(k - \ddot{A}(z+i\xi_0))}} e^{\frac{\beta \dot{A}^2(z+i\xi_0)}{2k} - \beta A(z+i\xi_0)}. \quad (4.11)$$

Now we have arrived at an expression for the desired quantity: the intrinsic free-energy A of the molecular system. We need only evaluate the free-energy of the system (trap plus molecule) and its derivatives at the point $z+i\xi_0$. To rid the expression of complex quantities, we write $z = z+i\xi_0 - i\xi_0 = z+i\xi_0 + \frac{\dot{A}(z+i\xi_0)}{k} = z' - \frac{\dot{A}(z')}{k}$. The argument of A and \dot{A} is then z' , and the argument of G_0 is $z' - \frac{\dot{A}(z')}{k}$:

$$e^{-\beta G_0(z' - \frac{\dot{A}(z')}{k})} \approx \frac{\sqrt{\beta}k}{\sqrt{2\pi(k - \ddot{A}(z'))}} e^{\frac{\beta \dot{A}^2(z')}{2k} - \beta A(z')}. \quad (4.12)$$

Taking the logarithm of both sides, dropping the primes, and further simplifying, we are left with

$$G_0(q = z - \frac{\dot{A}(z)}{k}) \approx A(z) - \frac{\dot{A}(z)^2}{2k} + \frac{1}{2\beta} \ln \left(1 - \frac{\ddot{A}(z)}{k} \right). \quad (4.13)$$

All that remains at this stage is to obtain $A(z)$ and its derivatives. As Hummer and Szabo point out, there are two ways to do this: directly from $A(z)$ via numerical differentiation or using work-weighted trajectory averages, as shown below:

$$\dot{A}(z) = -k \langle \langle q - z \rangle \rangle = \langle \langle F \rangle \rangle \quad (4.14)$$

$$1 - \frac{\ddot{A}(z)}{k} = \frac{\beta}{k} (\langle \langle F^2 \rangle \rangle - \langle \langle F \rangle \rangle^2). \quad (4.15)$$

Here, the double angle brackets $\langle \langle \dots \rangle \rangle$ represent $\langle (\dots) e^{-\beta W(z)} \rangle / \langle e^{-\beta W(z)} \rangle$, where the single angle brackets indicate that averages are taken over all trajectories. It is worth noting that in the implementation of the Weierstrass method pursued herein, the work

values $W(z)$ are computed by integrating in bias position z :

$$W(z) = \int_0^z F dz. \quad (4.16)$$

This is in contrast to the external work integration detailed in the description of the WHAM method found in Chapter 3; there, the independent variable was time, not bias. The method of using an inverse Weierstrass transform to extract the molecular free-energy landscape now detailed, it will be implemented on real datasets in the following sections. Two systems will be studied: first, a DNA hairpin, as it constitutes a simple model system that possesses only secondary structure, and second, an RNA pseudoknot system, more complex as it contains tertiary structure.

4.2 Data Collection

FECs were collected for a DNA hairpin, a simple structure containing a stem and a loop. The hairpin was attached to handles comprised of 1000 base pairs of double-stranded DNA. Such systems have been extensively studied in the past, and the landscape profiles for this particular hairpin have been measured by two different methods [37, 41]. The data were collected in force-ramp experiments, wherein the laser traps are pulled apart at constant velocity. Hairpin 30RS0T4 (henceforth Hairpin A) is 30 base pairs long [97]. Data for Hairpin A were collected at several different pulling speeds, ranging from 10 nm/s to 300 nm/s. The optical tweezer apparatus used to collect hairpin data is similar to the one outlined in a previous chapter, but uses electro-optic deflectors instead of acousto-optic deflectors to manipulate the laser traps. The data were collected between the months of February and May 2010, prior to the start of this thesis work; interested readers can consult reference [41] for further details regarding the collection of the hairpin data.

The pseudoknot data is a subset of the data described in Chapter 3, containing 300 curves taken at pulling speeds ranging from 240–250 nm/s. The data were collected

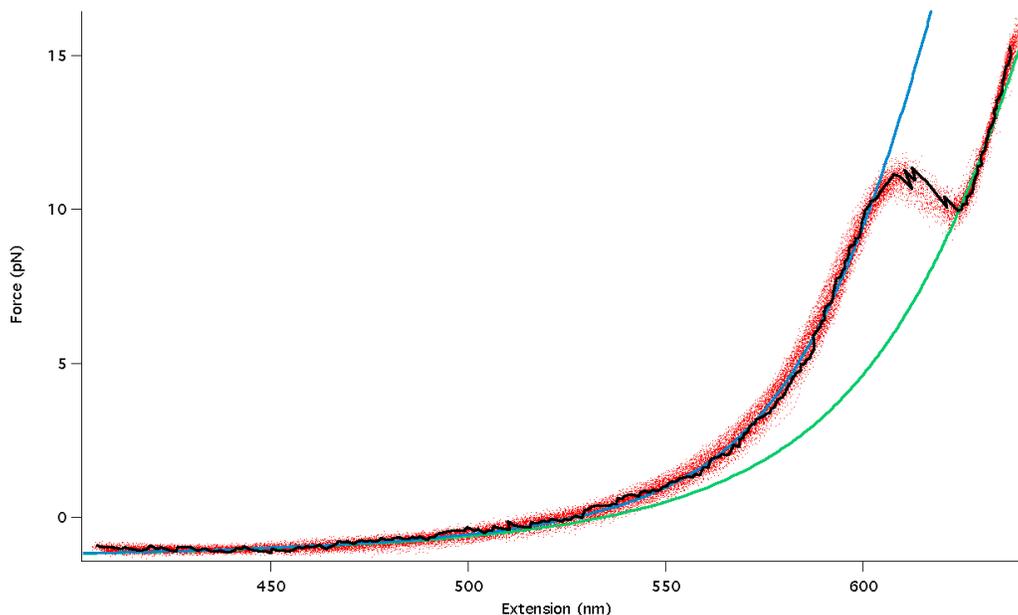


Figure 4.3: Representative FECs for Hairpin A, taken at a pulling speed of 10 nm/s, along with the WLC fits to the folded (blue) and unfolded (green) states. The black curve is a particular FEC, and the red dots represent the FECs of multiple curves.

between the months of July and September 2012 on the optical tweezer apparatus outlined in Chapter 3.

4.3 Data Analysis and Results

4.3.1 DNA Hairpin

The individual FECs for the hairpin was first aligned and fit to a WLC model in order to determine the force offset, as explained in Section 3.3.1. Representative FECs, along with the WLC fits to the folded and unfolded states, are shown in Figure 4.3. Next, the WHAM landscape reconstruction method, described in earlier chapters, and Weierstrass method, described above, were applied to the data. For this molecule, 10 data sets—comprising 894 curves—were analyzed using the Weierstrass method, and 9 separate data sets—comprising 818 curves—were analyzed with the WHAM method; one of the 10 data sets produced an unusable landscape when analyzed with

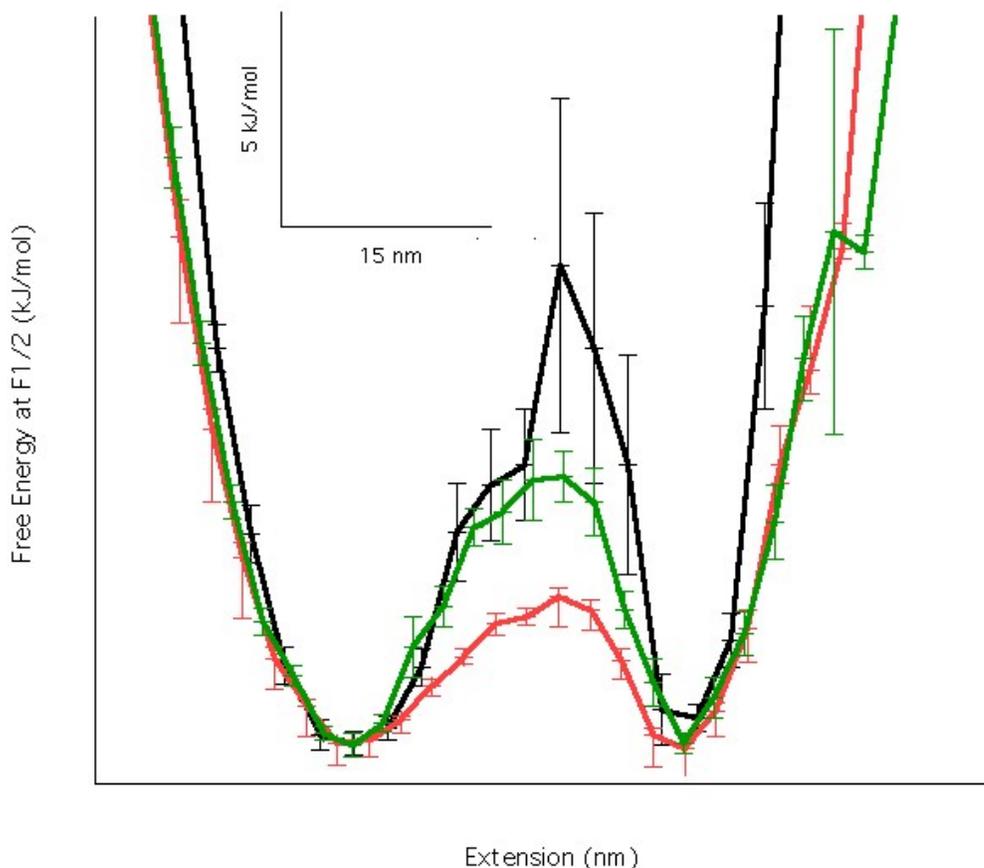


Figure 4.4: Average reconstructed landscapes for Hairpin A derived from the WHAM method (black), Weierstrass method using work-weighted averages (green) and finite differences (red) to compute $A(z)$ and its derivatives. The WHAM curve represents the average over 9 reconstructions, and the Weierstrass curves represent the average over 10 reconstructions. Errors are standard errors of the mean. The average $F_{1/2}$ to which the WHAM landscapes are tilted is 11 ± 1 pN; for the two Weierstrass landscapes, it is 11.8 ± 0.6 pN

the WHAM method, likely due to poor statistics in the low force rip region, and was rejected.

The Weierstrass method was implemented in two different ways: first, using the central difference method to calculate the derivatives of $A(z)$, and second, using work-weighted averages [Equations (4.14) and (4.15)]. Figure 4.4 shows the results of the various methods of landscape extraction; it contains three landscapes, each tilted to $F_{1/2}$, the force at which there is a 0.5 probability of the hairpin being folded or unfolded. The average $F_{1/2}$ for the WHAM landscapes is 11 ± 1 pN; for the

Weierstrass landscapes, it is 11.8 ± 0.6 pN. The discrepancy is likely due to subtle differences in the implementation of “tilting” the landscapes to a force. In the WHAM procedures, the tilt is applied at the very end of the analysis; a term Fx is subtracted from the free-energy at each point x . In the Weierstrass procedures, because the work is calculated directly from the force [see Equation (4.2)], it is possible to “tilt” the experiment to a new force value earlier in the procedure, when work is calculated, by subtracting the constant F from each force value. The difference amounts to subtracting a constant from an average in the former case, and in the latter case, subtracting the same constant from individual terms, then averaging these individual terms. The reason these formulations differ in the present work is that extensions are binned, and Fx_c , where the subscript c denotes “bin centre,” has a different value than $\frac{1}{N} \sum_{i=0}^N Fx_i$.

Qualitatively, it is clear that the Weierstrass method (green and red curves) is able to reproduce the barrier to unfolding, but the barrier is lower and less sharp than it is in the WHAM reconstruction. Further, the Weierstrass reconstructions accurately recapture the fact that the barrier is closer to the unfolded state than to the folded state, which is expected for this hairpin [41, 97]; this skew is also visible in the WHAM landscape reconstruction. The slope of the handle region in the Weierstrass reconstructions is not as high as in the WHAM reconstruction; this deviation is especially pronounced in the unfolded well, though given the error bars, it may not be statistically significant. Between the Weierstrass reconstructions, the work-weighted trajectory averages method for calculating $A(z)$ and its derivatives seems superior, and recaptures the height of the WHAM barrier more accurately.

In principle, the stiffer the trap in which the molecule is caught, the more accurately the Weierstrass method reproduces the free-energy landscape; in fact, Hummer and Szabo suggest that the Weierstrass method is incapable of resolving any landscape barriers whose curvature exceeds the stiffness of the trap [66]. Figure 4.5 illustrates the trap stiffness compared to the barrier curvature. In black is the WHAM landscape, tilted to an average of $F_{1/2}$ of 11 ± 1 pN. In violet is a quadratic function

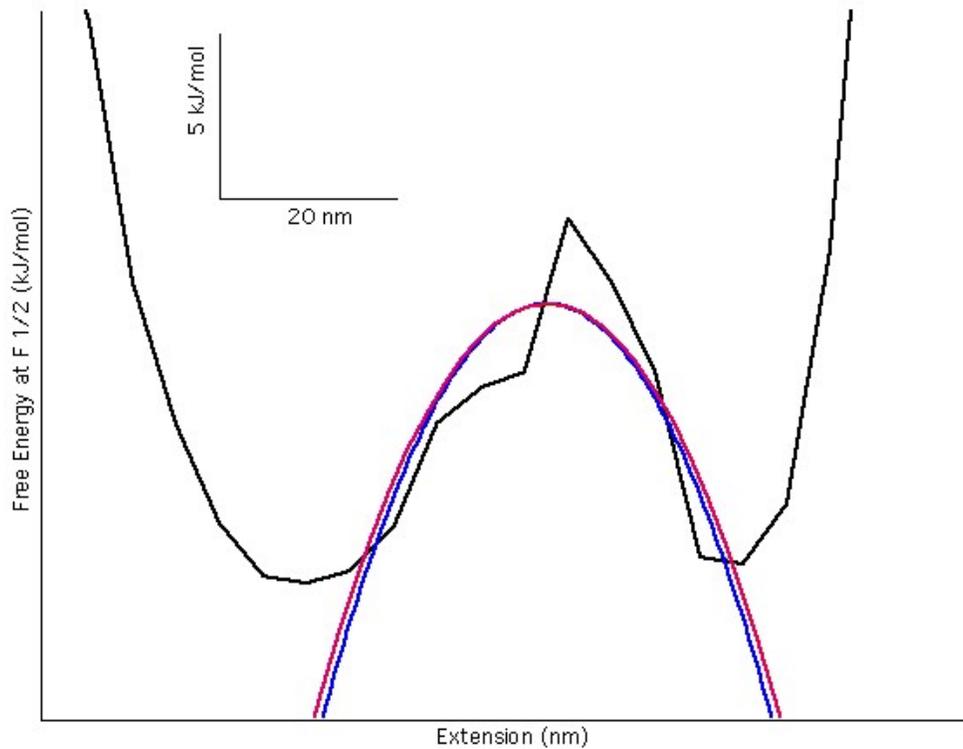


Figure 4.5: Average reconstructed landscapes for Hairpin A derived from the WHAM method (black) and tilted to an average $F_{1/2}$ of 11 ± 1 pN, along with a fit to the curvature of the barrier (blue) and a curve representing the stiffness of the laser trap (violet). The curvature of the barrier is 0.26 pN/nm, and the stiffness of the laser trap is 0.24 pN/nm.

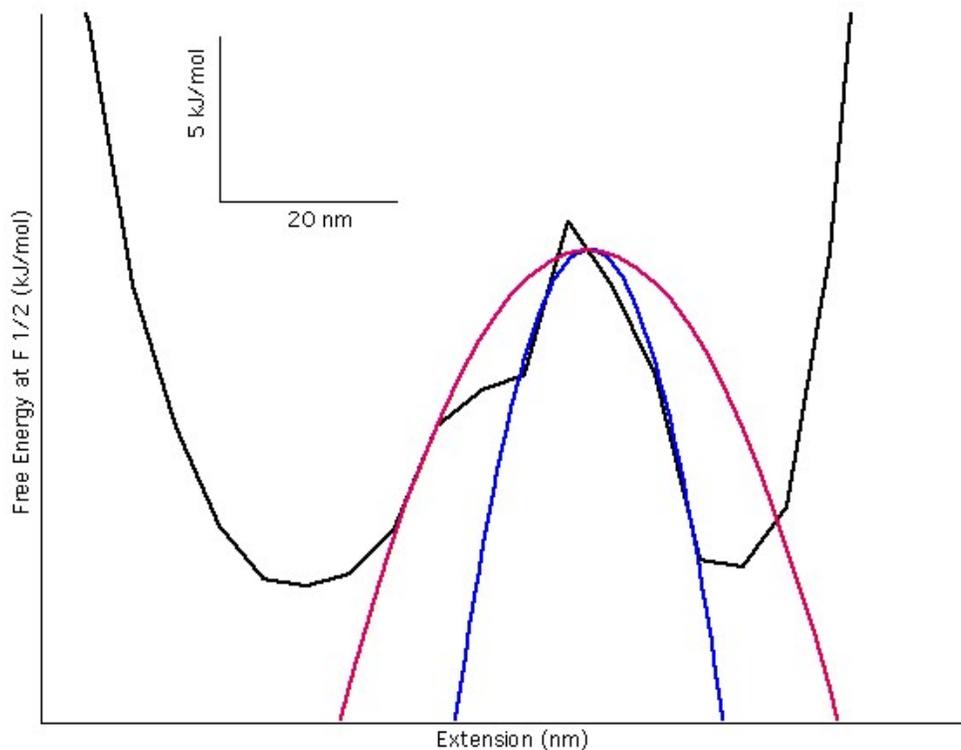


Figure 4.6: Average reconstructed landscapes for Hairpin A derived from the WHAM method (black) and tilted to an average $F_{1/2}$ of 11 ± 1 pN, along with a fit to the curvature of a sharp feature on the barrier (blue) and a curve representing the stiffness of the laser trap (purple). The curvature of this sharp barrier feature is 0.83 pN/nm, and the stiffness of the laser trap is 0.24 pN/nm.

with a curvature equal to the stiffness of the laser trap, 0.24 pN/nm. The blue curve is a quadratic fit to the barrier region; its curvature is 0.26 pN/nm. Because the trap stiffness is comparable to the barrier curvature, the Weierstrass method is able to capture the presence of the barrier; see Figure 4.4. However, Figure 4.6 illustrates a quadratic fit to the sharp barrier feature near the unfolded well. This feature has greater curvature than the entire barrier, at 0.83 pN/nm, and it is clear that the laser trap is far “softer” than this feature. As a result, the Weierstrass method is unable to reproduce this feature; see again Figure 4.4. As a caveat, the large standard error for the points along this sharp feature shed doubt on any statistically significant difference between the Weierstrass and WHAM reconstructions.

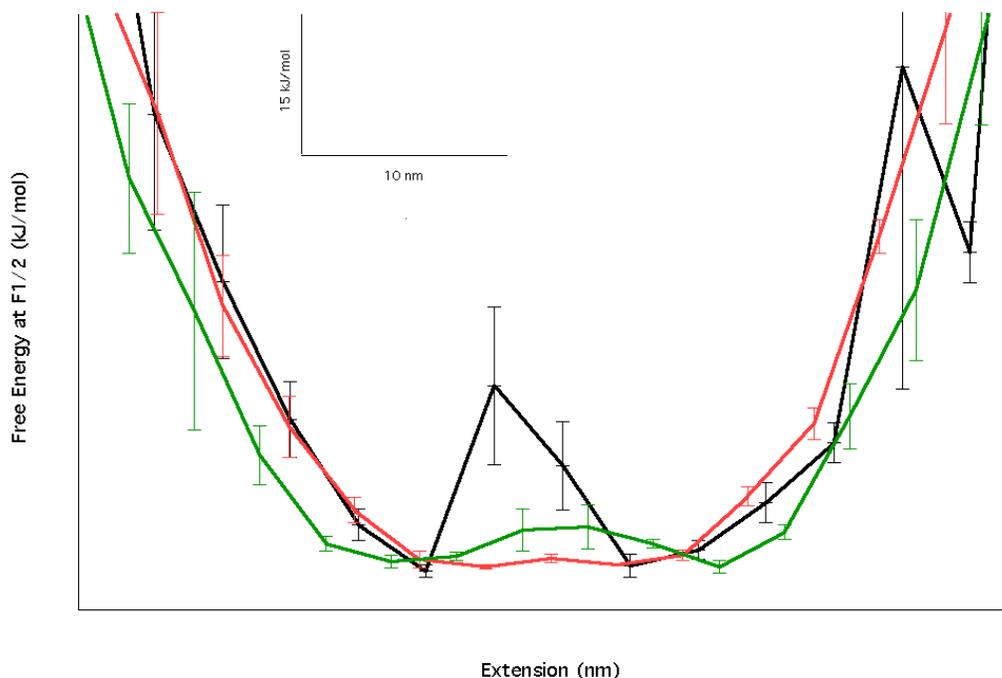


Figure 4.7: Average reconstructed landscapes for RNA pseudoknot derived from a subset of 300 curves from the data shown in Figure 3.10. The result of using the WHAM method is shown in black, the Weierstrass method using work-weighted averages in green, and the Weierstrass method using finite differences in red. Errors are standard errors of the mean. The average $F_{1/2}$ to which the WHAM landscapes are tilted is 34 ± 8 pN; for the two Weierstrass landscapes, it is 33 ± 8 pN.

4.3.2 RNA Pseudoknot

DNA hairpins exhibit secondary structure only; for the purposes of characterizing the usefulness of the Weierstrass analysis method, it is desirable to test it on a more complex structure. The RNA pseudoknot studied in Chapter 3 is such a system. The Weierstrass method was implemented on three data sets, comprised of 300 curves, collected for the analysis in Chapter 3. As can be seen in Figure 4.7, the Weierstrass method fails to reproduce the barrier evident in the WHAM landscape (black), regardless of whether the finite-difference method (red) or the work-weighted averages method (green) is used. The latter Weierstrass landscape does exhibit a low barrier, but it is clear this level of resolution is insufficient for practical applications: it gives a barrier that is in the wrong location and that is 5–10 times too low.

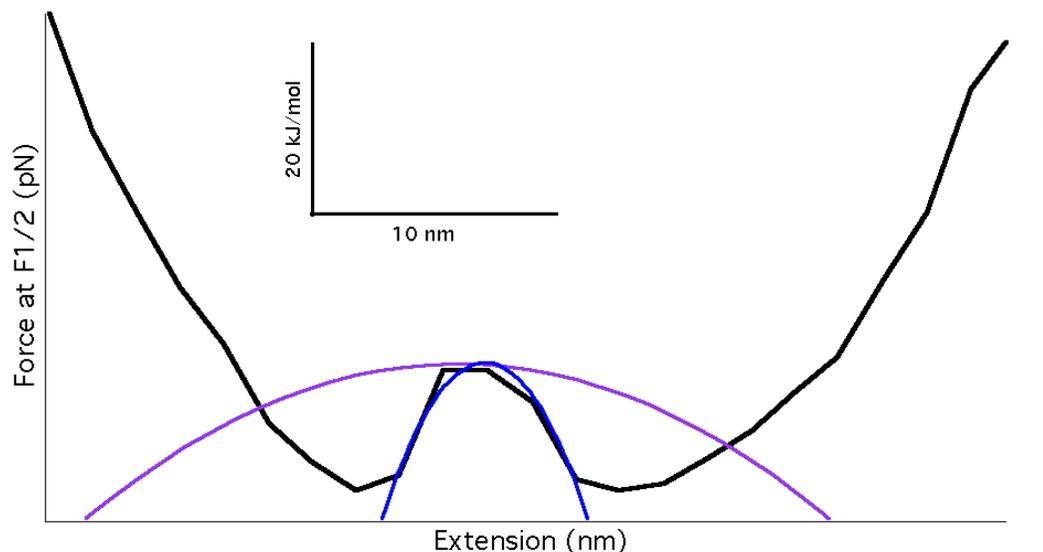


Figure 4.8: Average reconstructed landscapes for RNA pseudoknot ScYLVWT derived from the WHAM method (black) and tilted to an average $F_{1/2}$ of 30 ± 2 pN, along with a fit to the curvature of the barrier (violet) and a curve representing the stiffness of the laser trap (red). The curvature of the barrier is 2.33 pN/nm, and the stiffness of the laser trap is 0.26 pN/nm.

The failure of the Weierstrass method to resolve the barrier in the case of the pseudoknot landscape can be understood in terms of the effects of the probe stiffness on the reconstruction, as described in the previous section. The barrier curvature for the pseudoknot is 2.33 pN/nm, while the stiffness of the laser trap used to collect these data is 0.26 pN/nm. The probe stiffness is nearly ten times lower than that of the barrier, and because the Weierstrass method is incapable of resolving any features with curvature exceeding the stiffness of the trap in which the data is taken, one does not expect the pseudoknot landscape, with its sharp barrier, to be reproducible using the Weierstrass analysis. These notions are illustrated in Figure 4.8.

Based on the above results, it can be concluded that the Weierstrass method is indeed capable of reproducing energy landscapes for real systems provided the trap stiffness is equal to—or, preferably, greater than—the system’s barrier curvature. A great advantage of the Weierstrass method is that it does not demand as extensive a volume of data as the WHAM method does, since the WHAM method necessitates

binning data multiple times [98]. The Weierstrass method is thus a valuable analysis tool that is most appropriate for data collected using stiff force probes, such as AFM microscopes [66], which tend to have cantilever stiffnesses in the range 10–100 pN/nm [33]. Indeed, the WHAM landscape reconstruction method has been problematic when used with AFM data [65], and the Weierstrass could provide a valuable alternative. The method is also applicable to data collected in laser optical tweezers apparatus for systems with “soft” landscape barriers.

Chapter 5

Future Work

The advent of SMFS studies has allowed exploration of the statistical mechanical energy landscape of biomolecules in unprecedented detail. This work presents an application of energy landscape theory to a never-before characterized system—the ScYLV pseudoknot—and contributes to the growing body of knowledge surrounding biological folding processes. It verifies and demonstrates the viability of a promising new technique for extracting energy landscapes of biomolecules from SMFS based on an inverse Weierstrass transform. As always, however, much work remains to be done.

The apparent overestimates of pseudoknot stability presented in Chapter 3 invite further analysis. To shed light on this issue, other stability measurements can be calculated and compared to the values from the WLC integrals; for instance, another value for ΔG_{fold} can be derived directly from the Jarzynski equality:

$$\Delta G(x) = -k_B T \ln \left\langle e^{-\frac{W(x)}{k_B T}} \right\rangle \quad (5.1)$$

Here, $W(x)$ is the total non-equilibrium work performed in a particular trial and angled brackets indicate an average over all trajectories. This characterization of stability will also be inflated, however, if the bias of the Jarzynski estimator is an issue for the pseudoknot as we suspect. Independent bulk experiments aimed at

characterizing the stability of ScYLV should therefore be performed. While the issue of bias needs to be addressed, the precise manifestation of bias in the WHAM method [39] has not yet been established, and theoretical work is needed before possible bias correction schemes in the spirit of Gore *et al.* [95] or Palassini *et al.* [99] can be applied to WHAM reconstructions.

The need for deconvolution of the pseudoknot landscape was highlighted by the comparisons between the landscape parameters and the results of model-dependent fits. An empirical deconvolution would be difficult, however, considering the wide range of forces over which a PSF would need to be measured. One possible approach to the deconvolution problem is to derive a theoretical PSF for the traps in the spirit of Hinczewski *et al.* [51], who present an approach applicable to constant-force data. An adaptation of their method valid for wide ranges of forces would be very useful for systems for which constant-force studies are not ideal and for which empirical deconvolution is impractical due to broad force distributions, as in the present work. Future work, therefore, should concentrate on expanding the promising methods of theoretical correction for instrument distortion.

The landscape analysis performed here for ScYLV is novel and interesting, but should be applied to the C27A pseudoknot as well to explore the source of the disparity in the frameshifting efficiencies between wild type and mutant. Some force-ramp data has already been collected for the mutant [44], but more data will be needed to acquire sufficient statistics. It is likely that for such a comparative study to be useful, refinements of the pseudoknot landscape presented in Chapter 3 are needed. For example, the error in each bin can be more accurately characterized using a bootstrapping analysis similar to the treatment of Gupta *et al.* [41]. Additionally, landscape statistics for the pseudoknot could be improved by tweaking the WHAM technique of Hummer and Szabo [39] such that FECs collected at different pulling rates can be combined and analyzed as a single data set. Minh *et al.* [100] detail such an approach, which replaces time bins with bias bins. It would be relatively straightforward to adapt the WHAM tools developed for this work to this strategy.

A comparison between the landscape reconstruction method presented here and the Boltzmann transform method of landscape reconstruction has not been done for pseudoknot systems as it has been for DNA hairpins [41] (indeed, this work constitutes the first-ever landscape reconstruction for an RNA pseudoknot). To attempt this comparison, constant-force experiments should be performed on the pseudoknot. A potential problem with this tactic is the pseudoknot's high $F_{1/2}$; the system will not survive long at such a high tension, and a lengthy observation would be required given the slow folding dynamics stemming from the system's high free-energy barrier [44]. Nevertheless, such studies could illuminate transition times (which could be compared to results derived from FECs using the approach of Yu *et al.* [20]). Further, while other work has demonstrated that extension is a reasonable reaction coordinate for DNA hairpins [101], it has not been tested for a complex tertiary structure like a pseudoknot. The goodness of extension as a reaction coordinate for pseudoknots and other systems possessing tertiary structure could be explored with constant-force experiments by testing whether the splitting probability is indeed 0.5 at the top of the energy barrier [102].

The promising new landscape reconstruction method of Hummer and Szabo [66] validated experimentally here for the first time is vastly simpler than the WHAM method to implement. In the immediate future, landscapes for other systems can be constructed with the Weierstrass method and validated against earlier results. A good starting place is Hairpin B studied by Gupta *et al.* [41], since large amounts of FEC data have already been collected for this system and since an independent landscape characterization exists. The free-energy barrier for Hairpin B is somewhat sharper than that of Hairpin A [41], so it may however be necessary to increase the stiffness of the LOT system—by increasing the laser power, for instance—and collect more data. The Weierstrass method can also be widely employed to reconstruct landscapes from SMFS AFM studies, since AFMs exhibit high stiffness. Future work can explore the possibility of increasing the number of terms retained in the expansion used in the derivation of the Weierstrass method; this would relax the constraints

on system stiffness. These higher-order terms include higher derivatives of $A(z)$, however, which may not be recoverable from experimental data [66]. Nevertheless, the potential benefits of increasing the applicability of the Weierstrass method to “soft” systems render such expansions worthy of further study.

In the long term, questions about the goodness of reaction coordinates and 1D energy landscape projections, multiple barriers on energy landscapes, and the alteration of landscape barriers with force invite investigation. The study of energy landscapes of biomolecules is a broad and multifaceted field that promises to remain dynamic in years to come.

Bibliography

- [1] Berg, J.M., Tymoczko, J.L., and Stryer, L. *Biochemistry*. New York: W H Freeman, 5th edition, 2002.
- [2] Dill, K.A. and MacCallum, J.L. The Protein-Folding Problem, 50 Years On. *Science*, 338:1042–1046, 2012.
- [3] Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., and Baker, D. Principles for designing ideal protein structures. *Nature*, 491: 222–229, 2012.
- [4] Chiti, F. and Dobson, C.M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.
- [5] So much more to know... *Science*, 309(5731):78–102, 2005.
- [6] Anfinsen, C.B., Haber, E., Sela, M., and White Jr., F.H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. USA.*, 47(9):1309–1314, 1961.
- [7] Anfinsen, C.B. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.
- [8] Ben-Naim, A. Pitfalls in Anfinsen’s thermodynamic hypothesis. *Chem. Phys. Lett.*, 511:126–128, 2011.
- [9] Wales, D.J. *Energy Landscapes*. Cambridge University Press, 1st edition, 2003.

- [10] Zwanzig, R., Szabo, A., and Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. USA.*, 89:20–22, 1992.
- [11] Levinthal, C. Are there pathways for protein folding? *J. Chem. Phys.*, 65: 44–45, 1968.
- [12] Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *PROTEINS: Structure, Function, and Genetics*, 21:167–195, 1995.
- [13] Dill, K.A. and Chan, H.S. From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 4(1):167–195, 1997.
- [14] Ikai, A. and Tanford, C. Kinetic Evidence for Incorrectly Folded Intermediate States in the Refolding of Denatured Proteins. *Nature*, 230:100–102, 1971.
- [15] Tsong, T.Y., Baldwin, R.L., and Elson, E.L. The Sequential Unfolding of Ribonuclease A: Detection of a Fast Initial Phase in the Kinetics of Unfolding. *Proc. Natl. Acad. Sci. USA.*, 68(11):2712–2715, 1971.
- [16] 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.
- [17] Bryngelson, J.D. and Wolynes, P.G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, 93:6902–6915, 1989.
- [18] Bryngelson, J.D. and Wolynes, P.G. A simple statistical field theory of heteropolymer collapse with application to protein folding. *Biopolymers*, 30: 177–188, 1990.
- [19] Onuchic, J.N., Luthey-Schulten, Z., and Wolynes, P.G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Chem.*, 48:545–600, 1997.

- [20] Yu, H., Gupta, A.N., Liu, X., Neupane, K., Brigley, A.M., Sosova, I., and Woodside, M.T. Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc. Natl. Acad. Sci. USA.*, 109(36):14452–7, 2012.
- [21] Dobson, C.M., Sali, A., and Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem. Int. Ed.*, 37:868–893, 1998.
- [22] Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. The Protein Folding Problem. *Annu. Rev. Biophys.*, 37:289–316, 2008.
- [23] Lau, K.F. and Dill, K.A. Lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22:3986–3997, 1989.
- [24] Ueda, Y., Taketomi, H., and Gō, N. Studies on Protein Folding, Unfolding, and Fluctuations by Computer Simulation. 11. A Three-Dimensional Lattice Model of Lysozyme. *Biopolymers*, 17:1531–1548, 1978.
- [25] Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.*, 168:595–617, 1983.
- [26] Best, R.B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struc. Biol.*, 22:52–61, 2012.
- [27] Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., and Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330:341–346, 2010.
- [28] Lindorff-Larsen K., Piana, S., Dror, R.O., and Shaw, D.E. How fast-folding proteins fold. *Science*, 334:517–520, 2011.
- [29] Wolynes, P.G. Recent successes of the energy landscape theory of protein folding and function. *Quart. Rev. Biophys.*, 38(4):405–410, 2005.

- [30] Oliveberg, M. and Wolynes, P.G. The experimental survey of protein-folding energy landscapes. *Quart. Rev. Biophys.*, 38(3):245–288, 2005.
- [31] Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Sosova, I., Woodside, M.T. Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc. Natl. Acad. Sci. USA.*, 109(14):5283–5288, 2012.
- [32] Various authors. Various chapters. In Buchner, J. and Kiefhaber, T., editor, *Protein Folding Handbook*. WILEY-VCH Verlag GmbH Co., 2005.
- [33] Greenleaf, W.J., Woodside, M.T., and Block, S.M. High-Resolution, Single-Molecule Measurements of Biomolecular Motion. *Annu. Rev. Biophys. Biomol. Struct.*, 36:171–190, 2007.
- [34] Borgia, A., Williams, P.M., and Clarke, J. Single-Molecule Studies of Protein Folding. *Annu. Rev. Biochem.*, 77:101–125, 2008.
- [35] Ha, T. Single-molecule fluorescence resonance energy transfer. *Methods*, 25:78–86, 2001.
- [36] Woodside, M.T., Garcia-Garcia, C., and Block, S.M. Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.*, 12:640–646, 2008.
- [37] Woodside, M.T., Anthony, P.C., Behnke-Parks, W.M., Larizadeh, K., Herschlag, D., Block, S.M. Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science*, 314:1001–1004, 2006.
- [38] Zoldak, G. and Rief, M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.*, 23:48–57, 2013.
- [39] Hummer, G. and Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA.*, 98:3658–3661, 2001.

- [40] Gebhardt, J.C.M., Bornschlogl, T., and Rief, M. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. USA.*, 107(5):2013–2018, 2010.
- [41] Gupta, A.N., Vincent, A., Neupane, K., Yu, H., Wang, F., and Woodside, M.T. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single- molecule force spectroscopy measurements. *Nature Phys*, 7:631–634, 2011.
- [42] Lannon, H., Haghpanah, J.S., Montclare, J.K., Vanden-Eijnden, E., and Brucic, J. Force-clamp experiments reveal the free-energy profile and diffusion coefficient of the collapse of protein molecules. *Phys. Rev. Lett.*, 110(12): 8301–8305, 2013.
- [43] Thirumalai, D. and Hyeon, C. RNA and Protein Folding: Common Themes and Variations. *Biochemistry*, 44(13):15755–15760, 2005.
- [44] Ritchie, D.B., Foster, D.A., and Woodside, M.T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. USA.*, 109(40): 16167–16172, 2012.
- [45] Neuman, K.C. and Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5 (6):491–505, 2008.
- [46] Woodside, M.T. and Valentine, M.T. Single-Molecule Manipulation Using Optical Traps. In Hinterdorfer, P. and Oijen, A., editor, *Handbook of Single-Molecule Biophysics*, pages 341–370. Springer US, 2009.
- [47] Neuman, K.C. and Block, S.M. Optical trapping. *Rev. Sci. Inst.*, 75(9): 2787–2809, 2004.

- [48] Greenleaf, W.J., Woodside, M.T., Abbondanzieri, E.A., and Block, S.M. Passive all-optical force clamp for high-precision laser trapping. *Phys. Rev. Lett.*, 95:208102, 2005.
- [49] Dudko, O.K. Graham, T.G.W., and Best, R.B. Locating the Barrier for Folding of Single Molecules under an External Force. *Phys. Rev. Lett.*, 107:208301, 2011.
- [50] Chandler, D. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1st edition, 1987.
- [51] Hinczewski, M., Gebhardt, J.C.M., Rief, M., and Thirumalai, D. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl. Acad. Sci. USA.*, 110:4500–4505, 2012.
- [52] Berkovich R., Garcia-Manyes, S., Klafter, J., Urbakh, M., and Fernández, J.M. Hopping around an entropic barrier created by force. *Biochem. Biophys. Res. Comm.*, 405:133–137, 2010.
- [53] Jansson, P.A., Ed. *Deconvolution of Images and Spectra*. Academic Press, New York, 2nd edition, 1997.
- [54] Forns, N., de Lorenzo, S., Manosas, M., Hayashi, K., Huguet, J.M., and Ritort, F. Improving Signal/Noise Resolution in Single-Molecule Experiments Using Molecular Constructs with Short Handles. *Biophys. J.*, 100:1765–1774, 2011.
- [55] Pfitzner, E., Wachauf, C., Kilchherr, F., Pelz, B., Shih, W.M., Rief, M., and Dietz, H. Rigid DNA Beams for High-Resolution Single-Molecule Mechanics. *Angew. Chem.*, 52(30):7766–7771, 2013.
- [56] Onoa, B. and Tinoco, I.Jr. RNA folding and unfolding. *Curr. Opin. Struct. Biol.*, 14(3):374–379, 2004.
- [57] Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14):374–379, 1997.

- [58] Wang, G.M., Sevick, E.M., Mittag, E., Searles, D.J., and Evans, D.J. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.*, 89:050601, 2002.
- [59] Liphardt, J., Dumont, S., Smith, S.B., Tinoco, I. Jr., and Bustamante, C. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science.*, 296:1832–1835, 2002.
- [60] Jarzynski, C. How does a system respond when driven away from thermal equilibrium? *Proc. Natl. Acad. Sci. USA.*, 98(7):3636–3638, 2001.
- [61] Schuss, Z. *Theory and Applications of Stochastic Differential Equations*. Wiley, New York, 1st edition, 1980.
- [62] Hummer, G. and Szabo, A. Free Energy Surfaces from Single-Molecule Force Spectroscopy. *Acc. Chem. Res.*, 38:504–513, 2005.
- [63] Ferrenberg, A.M. and Swendsen, R.H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.
- [64] Roux, B. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91:275–282, 1995.
- [65] Harris, N.C., Song, Y., and Ching-Hwa, K. Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy using Jarzynski's Equality. *Phys. Rev. Lett.*, 99(6):068101, 2007.
- [66] Hummer, G. and Szabo, A. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA.*, 107:21441–21446, 2010.
- [67] Zhang, Q., Brujic, J., and Vanden-Eijnden, E. Reconstructing Free Energy Profiles from Nonequilibrium Relaxation Trajectories. *J. Stat. Phys.*, 144: 344–366, 2011.

- [68] Dudko, O.K., Hummer, G., and Szabo, A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96:108101, 2006.
- [69] Evans, E. and Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541, 1997.
- [70] Bell, G.I. Models for the specific adhesion of cells to cells. *Science*, 200:618, 1978.
- [71] Zhurkov, S.N. Kinetic concept of the strength of solids. *Int. J. Fract. Mech.*, 1:311, 1965.
- [72] Strunz, T., Oroszlan, K., Schafer, R., and Guntherodt, H.J. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. USA.*, 96(1):11277–1182, 1999.
- [73] Greenleaf, W.J., Frieda, K.L., Foster, D.A.N., Woodside, M.T., and Block, S.M. Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science*, 319:630–633, 2008.
- [74] Garg, A. Escape-field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Phys. Rev. Lett.*, 51:15592, 1995.
- [75] Kramers, H.A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica.*, 7:284, 1940.
- [76] Dudko, O.K., Hummer, G., and Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. USA.*, 105(41):15755–15760, 2008.
- [77] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Energy Landscapes*. Garland Science New York, 4th edition, 2002.

- [78] Brierley, I., Gilbert, R.J.C., and Pennell, S. Pseudoknot-dependent programmed -1 ribosomal frameshifting: Structures, mechanisms and models. *Nucleic Acids Mol. Biol.*, 24:149–174, 2010.
- [79] Dulude, D., Berchiche, Y.A., Gendron, K., Brakier-Gingras, L., and Heveker, N. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology*, 345: 127–136, 2006.
- [80] Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F., and Howard, M.T. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, 332:498–510, 2005.
- [81] Farabaugh, P.J. Programmed translational frameshifting. *Microbiol. Rev.*, 60: 103–134, 1996.
- [82] Dinman, J.D. and Wickner, R.B. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.*, 66:3669–3676, 1992.
- [83] ten Dam, E., Pleij, K., and Draper, D. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *Biochemistry.*, 31(47):11665–11676, 1992.
- [84] Namy, O., Moran, S.J., Gilbert, R.J., and Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature.*, 441:244–247, 2006.
- [85] Cornish, P.V., Stammer, S.N., Giedroc, D.P. The global structures of a wild-type and poorly functional plant luteoviral mRNA pseudoknot are essentially identical. *RNA.*, 12:1959–1969, 2006.

- [86] Neupane, K., Yu, H., Foster, D.A., Wang, F., and Woodside, M.T. Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res.*, 39:7677–7687, 2011.
- [87] Marko, J.F. and Siggia, E.D. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.
- [88] Wang, M.D., Yin, H., Landick, R., Gelles, J., and Block, S.M. Stretching DNA with Optical Tweezers. *Biophys. J.*, 72:1335–1346, 1997.
- [89] Liphardt, J., Onoa, B., Smith, S.B., Tinoco Jr., I., and Bustamante, C. Reversible Unfolding of Single RNA Molecules by Mechanical Force. *Science*, 292:733–737, 2001.
- [90] Seol, Y., Skinner, G.M., and Visscher K. Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys. Rev. Lett.*, 93:118102, 2004.
- [91] Seol, Y., Skinner, G.M., Buhot, A., Halperin, A., and Visscher, K. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys. Rev. Lett.*, 98:158103, 2007.
- [92] Hammond, G.S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.*, 77: 334–338, 2011.
- [93] Munoz, V. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.*, 34(9):2634–2652, 2006.
- [94] Munoz, V. Conformational Dynamics and Ensembles in Protein Folding. *Ann. Rev. Biophys. Biom. Struct.*, 36:395–412, 2007.
- [95] Gore, J., Ritort, F., and Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. USA.*, 100(22):12564–12569, 2003.

- [96] Zayed, A.I. *Handbook of Function and Generalized Function Transformations*. CRC Press, 1st edition, 1996.
- [97] Woodside, M.T., Behnke-Parks, W.M., Larizadeh, K., Travers, K., Herschlag, D., and Block, S.M. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. USA.*, 103(16):6190–6195, 2006.
- [98] Hummer, G. and Szabo, A. Theory and Evaluation of Single-Molecule Signals. In Eli Barkai, Frank L. H. Brown, Michel Orrit, Haw Yang, editor, *Thermodynamics and Kinetics from Single-Molecule Force Spectroscopy*, chapter 5, pages 139–180. World Scientific Publishing Co. Pte. Ltd., 2008.
- [99] Palassini, M. and Ritort, F. Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment. *Proc. Natl. Acad. Sci. USA.*, 107:12564–12569, 2011.
- [100] Minh, D.D.L. Free-energy reconstruction from experiments performed under different biasing programs. *Phys. Rev. E.*, 74:061120, 2006.
- [101] Morrison, G., Hyeon, C., Hinczewski, M., and Thirumalai, D. Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments. *Phys. Rev. Lett.*, 106:138102, 2011.
- [102] Chodera, J.D. and Pande, V.S. Splitting probabilities as a test of reaction coordinate choice in single-molecule experiments. *Phys. Rev. Lett.*, 107:098102, 2011.